

Containment of Extended Length Polymorphisms in Silk Proteins

Alberto Chinali · Wolfram Vater · Baerbel Rudakoff ·
Alexander Sponner · Eberhard Unger · Frank Grosse ·
Karl-Heinz Guehrs · Klaus Weisshart

Received: 21 May 2009 / Accepted: 10 February 2010 / Published online: 27 March 2010
© Springer Science+Business Media, LLC 2010

Abstract The spider silk gene family to the current date has been developed by gene duplication and homogenization events as well as conservation of crucial sequence parts. These evolutionary processes have created an amazing diversity of silk types each associated with specific properties and functions. In addition, they have led to allelic and gene variants within a species as exemplified by the major ampullate spidroin 1 gene of *Nephila clavipes*. Due to limited numbers of individuals screened to date little is known about the extent of these heterogeneities and how they are finally manifested in the proteins. Using expanded sample sizes, we show that sequence variations expressed as deletions or insertions of tri-nucleotides lead to different sized and structured repetitive units throughout

a silk protein. Moreover, major ampullate spidroins 1 can quite dramatically differ in their overall lengths; however, extreme variants do not spread widely in a spider population. This suggests that a certain size range stabilized by purifying selection is important for spidroin 1 gene integrity and protein function. More than one locus for spidroin 1 genes possibly exist within one individual genome, which are homogenized in size, are differentially expressed and give a spider a certain degree of adaptation on silk's composition and properties. Such mechanisms are shared to a lesser extent by the second major ampullate spidroin gene.

Keywords Spider silk evolution · Length polymorphism · Spidroin structure–function relationships · Genus *Nephila*

Electronic supplementary material The online version of this article (doi:10.1007/s00239-010-9326-2) contains supplementary material, which is available to authorized users.

A. Chinali · W. Vater · B. Rudakoff · A. Sponner · E. Unger ·
F. Grosse · K.-H. Guehrs · K. Weisshart
Leibniz Institute for Age Research-Fritz Lipmann Institute
(former Institute of Molecular Biotechnology),
Beutenbergstrasse 11, 07745 Jena, Germany

Present Address:

A. Chinali
Invitrogen Ltd., 3 Fountain Drive, PA4 9RF Paisley, UK

Present Address:

A. Sponner
Institute of Entomology, Branisovska 31, 37005 Ceske
Budejovice, Czech Republic

K. Weisshart (✉)
Carl Zeiss MicroImaging GmbH, Carl-Zeiss-Promenade 10,
07745 Jena, Germany
e-mail: weisshart@zeiss.de

Introduction

The divergence of the Araneomorphae (order Araneae) from their sister taxa Mygalomorphae occurred in the middle Triassic 250 million years ago (Selden and Gall 1992). This split was accompanied by the evolution of one new type of silk-producing gland, the major ampullate (MA) gland (Craig 2004). At least two types of silk fibroins or spidroins, the protein components of spider silks, have co-evolved with the MA gland. They have been designated as major ampullate spidroins (MaSp) 1 and 2 (Hinman and Lewis 1992; Xu and Lewis 1990). Like other spidroins they possess conserved N- and C-termini of ~100–150 amino acids (aa) in length and long (>2,900 aa) repetitive parts (Ayoub et al. 2007). The repetitions are largely attributable to four amino acid motifs (A_n , GA, GGX, and GPGX_n with X representing a subset of aa) that vary in appearance and proportion in the different spidroin types, but are shared

across distantly related species (Gatesy et al. 2001; Hayashi et al. 1999).

Repetitive sequences are characteristic for their paralog spidroins, which have been derived by gene duplication followed by sequence divergence prior to species diversification (Challis et al. 2006; Garb et al. 2006). These spidroins are mostly expressed in specialized glands within a species. For example, derived orb-weaving spiders express as many as seven different paralog spidroins including those derived from the major ampullate, minor ampullate, flagelliform, tubuliform, aggregatae, aciniform, and pyriform glands. Ortholog spidroins, which are expressed in evolutionary-related glands of different species, share similarities in their repetitive parts in regard to the sequence motifs employed. Repetitive sequences have been further homogenized by homologous recombination and gene conversion between allelic forms and genetic loci after species diversification (Beckwitt et al. 1998; Hayashi 2002; Ayoub et al. 2007). Hence, they are more similar to each other along the same gene than to repetitions at the same positions of ortholog genes. Not all repetitions, however, seemed to have undergone such a concerted evolution process. For example in the case of major ampullate spidroins, the last four repetitive units juxtaposed to the conserved C-termini are more similar between ortholog genes than to the other repetitions of the same gene (Hayashi and Lewis 2000).

Both the major ampullate spidroins share poly-alanine (A_n) blocks and to a certain extent the GGX ($X = A, L, Q$ or Y) motif, which in case of MaSp 2 is invariantly embedded within the GPGXX motif and not tandemly repeated. The two other non poly-alanine sequences are more characteristic for one spidroin type: GA for MaSp 1 and GPGXX ($XX = QQ$ or GY) for MaSp 2, which renders MaSp 2 proline rich (Gosline et al. 1999). The poly-alanine (4–10 aa) and GA (2–4 aa) regions adopt β -sheet structures and compose the crystal forming domains of MaSp 1 and 2, which are believed to confer tensile strength to the fiber (Gosline et al. 1999; Grubb and Jelinski 1997; Riekel et al. 1999; Simmons et al. 1996; Thiel and Viney 1995). These crystalline regions are interspersed by glycine rich domains of 10–30 aa composing the non-crystalline regions of the proteins. GGX motifs are believed to form 3_1 - and glycine II helical structures involved in alignment among adjacent protein chains, whereas $GPGX_n$ sequences are predicted to organize into β -turn spirals and type II β -turns, all contributing to the extensibility of a fiber (Dong et al. 1991; Gosline et al. 2002; Hayashi and Lewis 2000; Hayashi et al. 1999; Rising et al. 2005; Simmons et al. 1996; van Beek et al. 1999). Despite these sequence and secondary structure differences there is evidence that MaSp 2 genes have been developed from MaSp 1 duplicates by recombination between genetic loci followed by sequence

convergence to their current appearance (Ayoub and Hayashi 2008).

The more pronounced variability in the non poly-alanine sequences within a silk gene suggests that diversification in silks' properties and hence functions reside in the majority in the non-crystalline areas (Craig et al. 2000). Poly-alanine stretches are followed by a mostly invariant glycine (G) duplet in MaSp 1, creating the sequence AGG (Xu and Lewis 1990). These three amino acids are often encoded by triplets encompassing the crossover hotspot instigator (Chi)-sequence ($5'$ -GCT GGT GG- $3'$) that is known to be a DNA hotspot of homologous recombination initiating rearrangements of a gene (Beckwitt et al. 1998; Jeffreys et al. 1985; Sezutsu and Yukuhiro 2000). A Chi-like sequence ($5'$ -CCT GGA GG- $3'$) that is part of trinucleotides coding for PGG could be a potential hot spot of recombination, but is rare in MaSp 2 (Craig 2003; Hinman and Lewis 1992). The modular architecture of spider genes and the sequels of nucleotides coding for short amino acid sequence motifs could additionally promote replication slippage potentially resulting in the length variations between repeat units observed within and between ampullate spidroin 1 genes (Beckwitt et al. 1998).

If the heterogeneities in the spidroin gene sequences are transformed into length variations of silk proteins alike the situation in the silkworm *Bombyx mori* is currently unknown (Lizardi et al. 1979). No comparative analysis is available for gene sizes and only one complete gene sequence of each of the major ampullate spidroins from the spider *Latrodectus hesperus* have become available (Ayoub et al. 2007). The gene sizes roughly correspond to published transcript lengths, which exceed 10 kb for both spidroin types (Lewis 2006). The translated proteins are of considerable sizes with molecular weights reported in the range of 200–350 kDa (Candelas et al. 1989; Hayashi et al. 1999; Mello et al. 1994; Sponner et al. 2005a). Within the species *Nephila clavipes*, a cluster of closely running polypeptides of similar size (260–320 kDa) has been observed by gel electrophoresis, but these variations might be rather due to protein modifications or posttranslational processing than being due to bona fide protein length polymorphisms (Augsten et al. 2000; Sponner et al. 2005a). Molecular weights reported in the range of 725 kDa for spidroins are likely due to dimerization of MaSps through disulfide bond formation (Jackson and O'Brien 1995; Mello et al. 1994; Sponner et al. 2005a).

Whereas sequence variations at the transcript levels are well established for major ampullate spidroins, there is currently only indirect evidence as to which extent they are encoded by their genes. Since a spider's genome is diploid in nature, two different sequences can be due to two different allelic variants (Araujo et al. 2005). *Euprostheno*s cDNA screens have, however, shown more than two

sequence variants within one individual (Rising et al. 2007). They could be potentially expressed by more than the two allelic forms, which would imply different genetic loci and hence more than one gene for the major ampullate spidroins. However, since these studies were based on transcript analysis mechanisms like alternative splicing and RNA editing cannot be ruled out as adequate scenarios to create such sequence diversity. More compelling evidence for different genetic loci of major ampullate spidroins genes have been provided by single nucleotide polymorphisms within the genes of *N. clavipes* and analysis of genomic clones from individuals from *Latrodectus* both revealing more than two sequence variants on the genomic level within an individual (Ayoub et al. 2007; Gaines and Marcotte 2008). However, no cytogenetic analysis has been presented so far supporting the presence of multiple spidroin gene loci.

Environmental and spinning conditions have been shown to influence the higher order structure of spun silk (Thiel et al. 1997; Vollrath et al. 2001). In addition, a spider's diet impinges on the amino acid composition of silk and affects a fiber's characteristic features (Craig et al. 2000; Lewis 1992). The mechanisms that control these alterations could potentially again involve transcript as well as protein processing or the usage of a cluster of different spidroin genes on which the spider can draw, all of which could eventually lead to the expression of different-sized silk fibroins (Ayoub et al. 2007; Gaines and Marcotte 2008; Rising et al. 2007). Yet it is not clear, if spidroin length will have a profound impact on the mechanical performance like the primary sequence and polypeptide chain organization and might therefore be under evolutionary constraints (Vollrath and Knight 2001; Vollrath et al. 2001). A comparative study on the DNA, RNA, and protein level should, therefore, be instructive how silk properties are controlled (Vollrath 1999). A high conservation at the gene, transcript, and protein levels indicative of a selective pressure would stress the impact of the spinning conditions on a silk's characteristics. Extended polymorphisms, on the other hand, would render such a contribution less dominant. Nature, extent, and influence of polymorphisms on production and performance of silks would therefore give us additional insights about their evolution and regulation.

Unfortunately, our knowledge about how sequence variations influence silk appearance and properties are still basic (Beckwitt et al. 1998; Hayashi 2002; Rising et al. 2007). It was, therefore, the motivation of the present study to look for sequence and length polymorphisms in a more comprehensive and systematic way to better understand their mode of generation and their impact on silk characteristic features. We show here that within a spider population and also within an individual multiple allelic variants

encoded by possibly more than one spidroin 1 gene can exist. These variations are not only limited to sequence rearrangements within the repetitive parts but also express themselves in length variations of entire spidroin 1 genes and silk proteins. Albeit length polymorphisms can be extensive in an individual, a selective pressure discriminates against its spread within a population. Interestingly, sizes of different spidroin 1 genes are homogenized within an individual. The different abundance of cDNAs of some spidroin genes indicates that their expression levels might be differentially controlled. This suggests that a certain length range is beneficial for a spider enabling it for example to adapt gene expression to environmental conditions, but that extreme over- or undersized spidroins are likely detrimental for the polymerization process as well as mechanical properties of silk and might lead to instabilities of their genes. The proneness for homologous recombination with unequal crossing over as the most likely mechanism to explain the heterogeneities found in the MaSp 1 gene seems to be reduced in its MaSp 2 paralog, which seems to evolve rather by local nucleotide insertions and deletions.

Materials and Methods

Spiders

Spiders of the species *N. clavipes* were collected in Florida and obtained from Mascarino Tarantulas (Los Angeles, CA). They were grouped in this study into different cohorts in dependence on their collection time and territory. Each cohort comprised ~100 individuals and was regarded to be part of the same population since all individuals were collected in a restricted area. The animals were kept at 80% humidity in little cages to discourage web building. For the experiments, adult female animals at the 6th to 9th instar states were used.

Determination of sequence variations on the genomic, transcript and protein level

In order to look for sequence variations of major ampullate (MA) spidroins within a spider population, an MA-gland specific cDNA library was generated using mRNA isolated from the glands' epithelia of a total of 50 *N. clavipes* spiders from cohort 1. 20 spidroin 1 and 2 positive clones (from >100) of the library were chosen and sequenced. As long as sequences matched up starting from the 3' end over their entire available lengths to each other disregarding single nucleotide polymorphisms (SNPs), they were regarded as identical. The sequences from the longest of the different clones were subsequently used for sequence comparison with published sequences using the MegAlign program from the Lasergene v7.1 software package (supplementary alignments 1 and 2). cDNAs were also

generated in a similar fashion from 50 individual spiders of the same cohort and selected clones were sequenced.

In order to see if sequence variations on the transcript levels were due to gene rearrangements or posttranscriptional processing, other tissues of the same spiders used to construct the MA gland specific bulk cDNA library were used to prepare genomic DNA. From the constructed gDNA library numerous clones (>100) were positive against MaSp 1 and 2 repetitive sequences. In contrast, we were able to retrieve only three MaSp 1 clones and one MaSp 2 clone positive for 3' specific sequences of the genes. The latter sequences were compared to the ones obtained from the cDNA library.

In order to look how gene sequence variations were transmitted to the protein level, cDNA open reading frames were translated with the EditSeq program from Lasergene v7.1 (supplementary Alignments 3 and 4). Alignments of sequences were done with MegAlign from the Lasergene v7.1 followed by visual inspection to evaluate exchanges, gaps, and inserts between the different clones.

Determination of length variations

Sizes of proteins were analyzed by polyacrylamide gel electrophoresis and western blotting. Transcripts were pretreated with either DNase or RNase, respectively, to discriminate against DNA contamination, prior to agarose gel electrophoresis followed by northern blotting. Genomic DNA was treated with restriction enzymes EcoR I, which is rare and for spidroin genes null cutting and the more frequently cutting enzyme Sau3A I. Fragments were analyzed by agarose gel electrophoresis followed by Southern blotting. Detection was accomplished with radioactively labeled spidroin sequences for DNA (Southern blots) and RNA (northern blots) or spidroin specific antibodies for proteins (western blots).

Determination of polymerization efficiencies and material properties

In order to assess if size would impinge on stability and properties of silk proteins, the influence of spidroin lengths on polymerization capabilities was tested with recombinant proteins expressing C-terminal sequences along with different extensions in their repetitive parts, with silk that was specifically degraded into a certain size range in chaotropic solutions or with solubilized silk that was composed of different length spidroins. Silk proteins aggregate in neutral buffers and the extent of this process was followed by atomic force microscopy. A link between spidroin size and mechanical features was established in tensile measurements on commercial available instrumentation. To this end, 10 independent fiber sections of five dragline samples obtained from one individual were analyzed for characteristic features like tensile strength, E modulus, extensibility, and energy-to-break. Diameters of fibers and fractured ends were examined by transmission electron microscopy (TEM). Statistical analysis was done by ANOVA testing using Origin 6.1 software (OriginLab Corporation).

Detailed protocols can be found in the supplementary material.

Results

Sequence variations are a general feature in spidroins within spider populations

From a cDNA library representing a population of 50 spiders from cohort 1, we retrieved clones showing up to five different sequences for each spidroin class (Table 1). Alignments of those to published ortholog sequences of the

Table 1 Sequence variations in spidroin genes

Type	Variant (gene)	Acc. No.	cDNA	gDNA	C1S1	C1S2	C1S3	C1S4	C1S5
1	var 1	AY654288	1		1				
1	var 2	AY654289	4		3		7 (SNP)		
1	var 3	AY654290	5		4 (SNP)		9 (SNP)		
1	var 4 (MaSp 1B)*	AY654291	6	2		7 (SNP)		6 (SNP)	4 (SNP)
1	var 5 (MaSp 1A)*	AY654292	4	1		8		5 (SNP)	4 (SNP)
1	var 6	EU617388					2		
2	var 1	AY654293	3		6				
2	var 2	AY654294	6		6	7		4	6
2	var 3	AY654295	5				5		
2	var 4	AY654296	3	1		5		4	5
2	var 5	AY654297	3				4		

The table lists the frequency a certain sequence was found within the cDNA and gDNA primary libraries from an ensemble preparation as well as from individuals. The presence of single nucleotide polymorphisms (SNP) between cDNA clones isolated from the designated individuals is indicated. Details can be obtained from supplementary Tables 1 and 2. *Gene designations taken from literature (Gaines and Marcotte 2008)

same species showed that within the same spidroin type the 3'-sequences coding for the non-repetitive parts were completely identical with only rare nucleotide exchanges occurring in the case of spidroin 1 (supplementary Alignments 1 and 2). This perfect match continued only to a limited extent into the sequences coding for the repetitive parts before sequence deviations between the clones and the published sequences became obvious (Beckwitt and Arcidiacono 1994; Hinman and Lewis 1992; Xu and Lewis 1990). The differences were manifested as deletions and insertions of multiples of trinucleotides (Fig. 1a, b).

Multiple sequence variations are expressed in individuals to different extents

Sequences obtained from cDNA clones of individual spiders of the same cohort used for the population study (C1S1–C1S5) had been recovered already from the bulk preparation with the exception of C1S3 (Table 1). Dependent on the numbers of sequenced clones available up to three or two different sequences were retrieved for MaSp 1 and 2, respectively, if single nucleotide polymorphisms (SNPs) were disregarded. SNPs were obvious for

some of the MaSp 1 sequences showing that a spider's genome can be multi allelic and heterozygous for a spidroin 1 gene and that more than one genetic locus possibly exists within some individuals (supplementary Tables 1 and 2). No SNPs were found for MaSp 2 clones. It should be noted that the numbers of different clones per individual were similar in the case of MaSp 2, but some MaSp 1 sequences represented by var 1 and var 6 in individuals C1S1 and C1S3, respectively, were significantly under-represented compared to the others. They in all probability are representative for the abundance of the corresponding transcripts and hence the level of gene expression (Table 1).

Sequence variations in transcripts are already manifested in the genes

From a gDNA library constructed from different tissues of the same 50 spiders used to construct the MA gland epithelium specific cDNA library we were able to retrieve three MaSp 1 clones and one MaSp 2 clone positive for 3' specific sequences of the genes. All their sequences found a match in one of the 5 different cDNA sequences over the sequenced lengths available for each class suggesting that variations occur already in the genes and are maintained in the transcripts for the analyzed samples (Table 1).

Sequence variations in silk genes express themselves on the protein level as primary structure differences of spidroin repetitive units

If protein sequences obtained from open reading frames of the cDNA clones from the bulk preparation were aligned their C-terminal ends were completely identical as expected with only rare amino acid exchanges in the case of MaSp 1 and this conservation continued for a certain length into the repetitive parts (supplementary Alignments 3 and 4; Fig. 2a, b). If sequences were annotated with ensemble repeats aligned, all MaSp 2 clones showed identical sequences at the same positions up to the fourth repetitive unit with only one slight deviation in clone AY654293 (supplementary Fig. 2). In contrast, a complete match continued up to the fourth repetitive unit until individual sequences at the same position differed from each other for only a few of the spidroin 1 clones (supplementary Fig. 1). Sequences AY64592 and EU617388 deviated already between the first repeat units located adjacent to the conserved part to a more or less extent and in AY64589 the repeat unit second to the conserved C-terminus was altered. The variations expressed themselves as different numbers of sequence motifs within ensemble repeats and differences in their lengths. This kind of variations was also present in the repetitive units that were in continuation to the conserved N-terminal sequences in both spidroins (supplementary Figs. 1, 2). Sequences of repetitive units along the same gene were also more heterogeneous for MaSp 1 than MaSp 2.

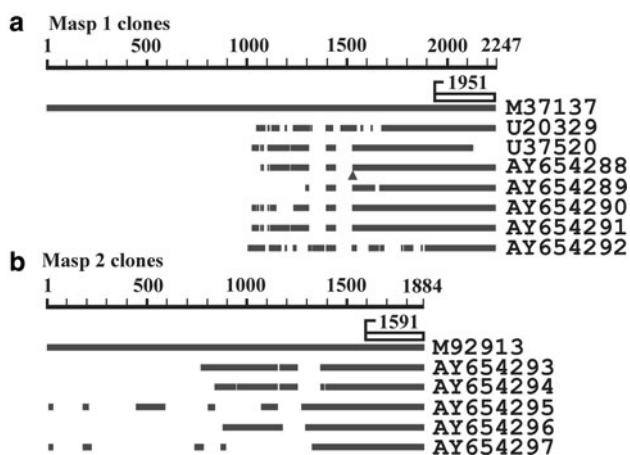


Fig. 1 Sequence comparison of spidroin cDNA clones from cohort 1. The MaSp 1 (panel a) and MaSp 2 (panel b) gene sequences are schematically presented on top and numbered in nucleotides according to the published sequences (Beckwitt and Arcidiacono 1994; Hinman and Lewis 1992; Xu and Lewis 1990). Note that only the 3'-ends of the sequences are available. The parts coding for the non-repetitive C-termini are indicated as open boxes with the start nucleotide indicated. Sequences from different clones are represented schematically by the dark shaded boxes. Accession numbers are given for each clone (AY numbers, this study; all others taken from GenBank). Deletions and insertions compared to published sequences are represented by gaps and triangles, respectively, and were always in multiples of trinucleotides. Note that due to the repetitiveness of the 5'-sequences other than the shown alignments would be possible. The alignment presented here was obtained by starting the alignment from the 3'-end. Exact boundaries have been omitted for simplicity, but can be retrieved from the supplementary Alignments 1 and 2. The sequence U37520 represents the only genomic clone, and lacks the very 3'-end of the gene

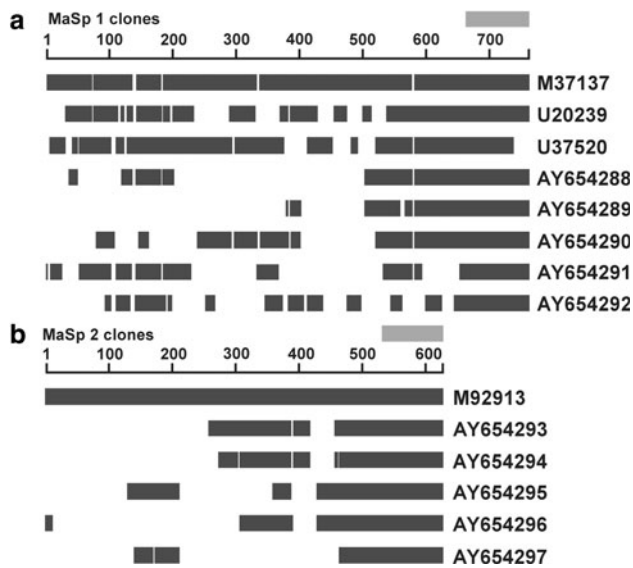


Fig. 2 Comparison of spidroin protein sequences from cohort 1 The major ampullate spidroin 1 (*panel a*) and 2 (*panel b*) protein sequences are schematically presented on top and numbered in amino acids according to the published sequences (Beckwitt and Arcidiacono 1994; Hinman and Lewis 1992; Xu and Lewis 1990). The locations of the non-repetitive C-termini are indicated as *light shaded boxes*. Sequences from different clones are represented by *dark shaded boxes*. The sketch depicts schematically how the different sequences can be registered to each other. Note that due to the repetitiveness of the N-terminal sequences other than the shown alignments would be possible. The alignment presented here was obtained by starting alignment from the C-terminal end. Accession numbers are given for each clone (AY numbers, this study; all others taken from GenBank). Exact boundaries have been omitted for simplicity, but can be retrieved from the supplementary Alignments 3 and 4

Length polymorphisms in transcripts can be extensive

Major transcripts from individuals C1S1–C1S3 were identical in their sizes and amounted to ~12.5 kb for MaSp 1 and ~12.0 kb for MaSp 2 despite the differences in their sequences (supplementary Fig. 3B). In the case of individual C1S3, a second less abundant transcript of ~13.5 kb for MaSp 1 was evident. This suggests that rearrangements within repetitive units alone do not lead to extensive length variations.

An average size of ~12.5 kb was also evident for most of the investigated 100 spiders from cohort 2 as exemplified by individuals C2S1 and C2S2 (Fig. 3a). A longer run for individual C2S1 revealed two closely running distinct transcripts of ~12.5 and ~13.0 kb (Fig. 3a, inset). Two of the clones from cohort 2, in contrast, showed sizes of exclusively ~18 kb (C2S3) and ~17.5 kb (C2S4) clearly exceeding the average 12.5 kb of the other clones. These bands represented bona fide RNA as the material was degraded by RNase (lanes R), but virtually unaffected by DNase (lanes D). Since rearrangements in individual repetitive units seemed to be too small to explain such huge

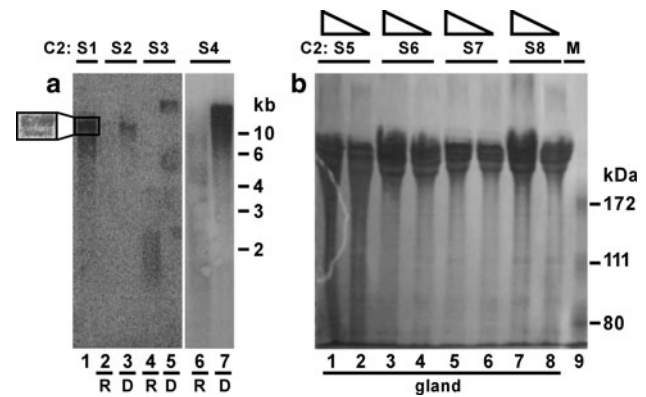


Fig. 3 Length polymorphism of transcripts and proteins of cohort 2 *Panel a* Northern blot analysis of transcripts. Total RNAs of individual spiders were hybridized to a spidroin 1 specific probe. The positions of size markers in kilo bases (kb) are listed. Were indicated, samples had been treated with RNase (R) or DNase (D). For the sample of lane 1 the result of a second experiment where RNA was left to separate longer is shown in the *box*. *Panel b*: Protein analysis of gland luminal extracts. Major ampullate gland contents of individual spiders were solubilized and analyzed by gel electrophoresis. Proteins were visualized by Coomassie staining. Marker proteins (M, lane 14, 10 kDa ladder) were run in parallel and their sizes are indicated in kilo Dalton (kDa). Two different concentrations of the same material were loaded. Cohort and spiders are identified by the letters “C” and “S” preceding numbers

size deviations, we conclude, that additional repetitions must account for them.

In order to get estimations on the abundance and range of size variations, additional cohorts were screened. Within 100 spiders investigated from each of cohorts 3 and 4 of which the results of eight (C3S1–C3S8) and three (C4S1–C4S3) representatives are displayed only three showed noticeable size deviations from the average 12.5 kb for other individuals (Fig. 4b and supplementary Fig. 4B). These are exemplified by individuals C3S6, C4S3, and C4S2 showing transcript sizes of ~10, ~10.5, and ~17.5 kb, respectively. In addition, three of the samples (C3S4, C3S5, and C3S8) showed a fainter second higher signal of ~13.5 kb. We want to point out that in every case studied all transcripts from one individual ran closely together in gels and were, therefore, of similar sizes independent of their overall lengths.

Different-sized transcripts are encoded by different sized genes

Restriction of genomic DNA with an enzyme that does not cut within the spidroin gene sequences resulted in a spidroin 1 positive band of ~13.5 kb for most individuals of cohort 3 and 4 in Southern blots demonstrating similar-sized genes (Fig. 4a and supplementary Fig. 4A). With non-transcribed 5' and 3' control regions in mind, this size would roughly correspond to the average size of ~12.5 kb of the major transcripts. CpG methylation sensitive

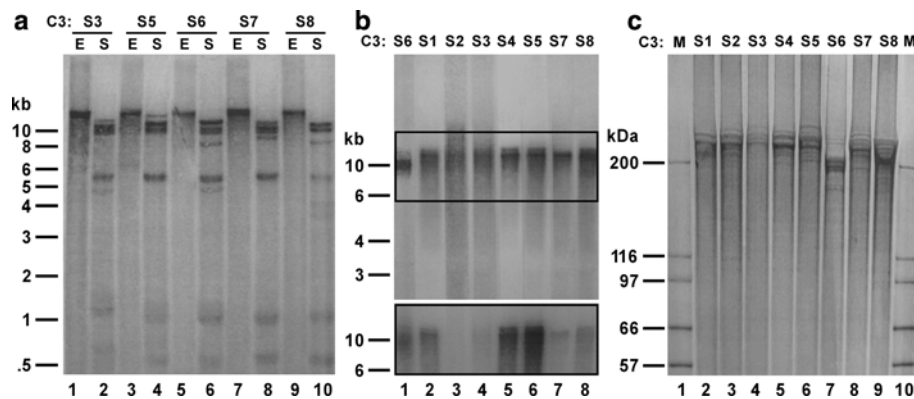


Fig. 4 Length polymorphism in spidroin genes, transcripts and proteins from cohort 3 *Panel a*: Southern blot analysis of genomic DNA. Genomic DNA from individual spiders were digested with either EcoRI (E, odd numbered lanes) or Sau3AI (S, even numbered lanes) and hybridized to a spidroin 1 specific probe. The positions of size markers in kilo bases (kb) are indicated. *Panel b* Northern blot analysis of total RNA. Total RNA of individual spiders were hybridized to a spidroin 1 specific probe. The positions of size markers in kilo bases (kb) are indicated. Note that individual lanes

were exposed for different times and corrected for the background signal to maximize signal visibility. The boxed area is also displayed as an original phosphor imager scan with one exposure time in the lower part. *Panel c* Protein analysis of silk fibers. Fibers of individual spiders were solubilized and analyzed by gel electrophoresis. A Coomassie stained gel is shown. Markers (M) were run in parallel and their sizes are indicated in kilo Dalton (kDa). Cohort and spiders are identified by the letters “C” and “S” preceding numbers

enzymes failed or only weakly digested the DNA sample demonstrating high CpG methylation of spider DNA (data not shown). Using a more frequently cutting restriction enzyme, slight fragment size polymorphisms became visible that might account for the small size variations of the corresponding transcripts (compare panels a and b of Fig. 4).

For smaller-sized transcripts in the range of ~10–10.5 kb represented by individuals C3S6 and C4S3, gene sizes were reduced to ~13 kb. For the individual C4S2 with a transcript size of ~17.5 kb, the gene size was also markedly increased to ~18.5 kb (supplementary Fig. 4A). In all but one case one gene fragment was evident (Fig. 4a and supplementary Fig. 4A). The only exception was individual C1S3 for which two closely running fragments of ~13.5 and ~15 kb were detected that roughly matched the two transcripts of ~12.5 and ~14 kb (supplementary Fig. 3A). This demonstrates similar-sized allelic or gene variants within an individual genome independent on their overall length.

Transcript lengths and protein sizes are collinear

For the most part, spider fibroins showed a size range of 220–250 kDa in protein gels (Fig. 3b). This molecular weight was invariantly linked to spiders showing an average length of ~12.5 kb for their major spidroin 1 transcripts (Fig. 4c; supplementary Figs. 3C and 4C). Lower and higher sized transcripts, on the other hand, resulted in lower and higher molecular weight distributions of the majority of silk proteins. For example, protein sizes varied between 190–205 kDa for C3S6 (transcript size ~10 kb), 150–200 kDa for C4S3 (transcript size ~10.5 kb) and 250–350 kDa for C4S2 (transcript size ~17.5 kb) for the most prominent polypeptides. The fact that the slightly

smaller transcript of C3S6 resulted in slightly larger proteins compared to individual C4S3 is in all probability due to the imprecision of molecular weight estimations since the standard markers employed did not cover the extreme high molecular weights seen with spidroin genes, transcripts, and proteins.

Extended length polymorphisms are not evident for the second spidroin

In all the examined cases, noticeable variations in transcript, gene, and protein sizes were restricted to MaSp 1 and MaSp 2 were not affected at all. All three spidroin 2 specific mRNAs from individuals C4S1, C4S2, and C4S3, for example, showed the same size of ~12.0 kb and their genes mapped to a ~13 kb restriction fragment in Southern blots (Fig. 5a, b). The respective proteins reactive to MaSp 2 specific sera ran exclusively between 220 and 250 kDa in each case, whereas those specific for MaSp 1 showed the expected size difference (Fig. 5c).

Spidroin length impinges on mechanical properties

The observation that extreme length variations were rarely encountered made us wonder if too highly over- or undersized proteins would display unfavorable characteristic features. Comparing mechanical properties of fibers from individuals C4S1–C4S3, C2S2–C2S4, and C3S5–C3S6 revealed that an increase in the molecular weight of the constituting spidroins resulted in a reduction in E modulus, extensibility and energy-to-break, whereas the tensile strength was elevated (shown for C4S1–C4S3 in Table 2 and Fig. 6). Diameters of the filaments varied only slightly with an increase for longer spidroins. The differences were not due to sequence variations between the clones since fibers composed of spidroins with different

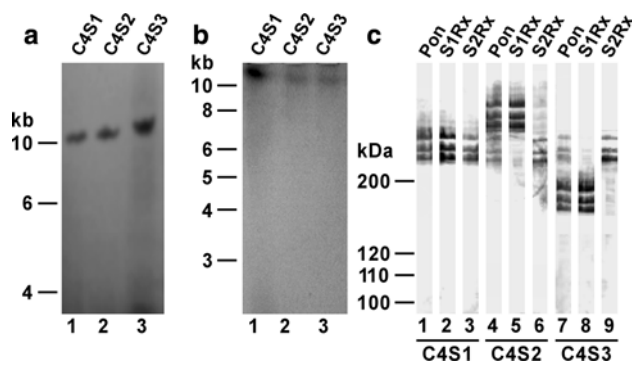


Fig. 5 Size distribution of individual spidroin 2 transcripts, genes and proteins Transcript and gene sizes were determined for individuals C4S1, C4S2 and C4S3 in Northern (a) and Southern (b) blots using a MaSp 2 specific probe. Markers in kilobases (kb) are indicated. Please consider the smiling effect of the gel in the Northern blot for the size estimations. Western blots (c) of spidroins were stained with Ponceau S (Pon) to visualize proteins or treated with MaSp 1 (S1Rx) and 2 (S2Rx) specific sera. Detection was with enhanced chemiluminescence (ECL). Molecular weights of selected 10 kDa-ladder marker proteins are indicated in kilo Dalton (kDa)

primary structures but of the same length did not show pronounced differences in their mechanical properties (supplementary Fig. 5 and supplementary Table 3). The fractured ends of filaments looked identical in each case and hinted to ductile materials (supplementary Fig. 5, inset).

Spidroin length influences silk polymerization

We also investigated if differences in sizes of spidroins would alter their polymerization efficiencies. Solubilized silks of individuals C4S1–C4S3, C2S2–C2S4 and C3S5–C3S6 were therefore tested in their ability to aggregate in neutral buffers. Polymers formed less efficiently with a marked decrease in lateral and axial extensions for those solubilized silks that contained shorter spidroin peptides (shown for C4S1–C4S3 in Table 3 and Fig. 6 inset). This trend continued for even smaller spidroins that were obtained by fractionated degradation or by recombinant expressed spidroin sequences of a defined length (supplementary Fig. 6 and supplementary Table 4).

Table 2 Material properties of dragline filaments of cohort 4

	150–200 kDa (C4S3)	220–250 kDa (C4S1)	250–300 kDa (C4S2)	<i>P</i> value
Radius r [μm]	1.72 ± 0.31	1.75 ± 0.25	1.81 ± 0.23	<0.05
Linear density μ_F [dtex]	0.70 ± 0.16	0.72 ± 0.18	0.76 ± 0.18	0.11
Specific density ρ_F [g/cm^3]	1.29 ± 0.31	1.31 ± 0.35	1.34 ± 0.25	0.08
Yield strength σ_y [GPa]	0.21 ± 0.08	0.20 ± 0.07	0.17 ± 0.06	0.06
E modulus E [GPa]	9.1 ± 2.8	8.1 ± 2.1	6.9 ± 1.9	<0.03
Extensibility ε_F [%]	45 ± 6	40 ± 8	35 ± 5	<0.05
Tensile strength σ_u [GPa]	0.75 ± 0.25	0.80 ± 0.31	0.83 ± 0.34	<0.01
Energy-to-break τ [kJ/kg]	145 ± 20	110 ± 40	95 ± 20	<0.01

Shown are mechanical characteristic average values from 10 different sections of 5 samples each of dragline filaments composed of spidroin 1 of different lengths. Significant differences between the means are defined as *P* values below 0.05 and are indicated in bold

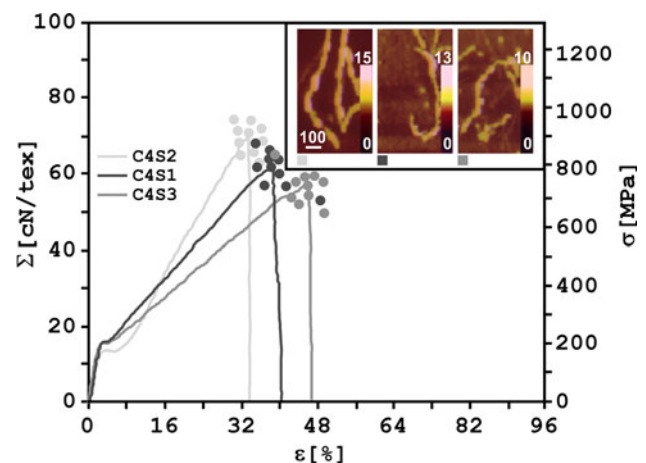


Fig. 6 Mechanical properties and polymerization capabilities of different sized spidroins Characteristic features of fibers from individuals C4S1, C4S2 and C4S3 were obtained in tensile measurements. Dots correspond to endpoints of ten measurements of different sections from the same dragline. The inset shows aggregation experiments of solubilized silk analyzed by atomic force microscopy. Length and height extensions in nanometers are indicated

Discussion

Homologous recombination with unequal cross over drives sequence diversities in the repetitive units of spidroin 1, but seems to be rare in spidroin 2

Phylogenetic studies of silk genes have concentrated on their non-repetitive C- and N-terminal parts. Phylogenetic tree analysis including the C-termini of silk proteins synthesized in various glands of orb-weaving spiders showed that these cluster according to the silk type rather than within a species (Challis et al. 2006; Garb et al. 2006). This indicates that silk genes are paralogs, which have originated by gene duplication prior to species diversification and that recombination had been sparse (Baldauf 2003; Koonin 2005). The repetitive parts, on the other hand, have further been subjected to concerted evolution that is they

Table 3 Polymerization efficiencies of spidroins of different lengths

	Length [nm] (axial)	Width [nm] (radial)	Height [nm] (radial)
C4S1 (220–250 kDa)	850 ± 250	10 ± 5	12.5 ± 2.5
C4S2 (250–350 kDa)	750 ± 200	7.5 ± 5	10 ± 5.0
C4S3 (150–200 kDa)	700 ± 250	5 ± 2.5	5 ± 2.5
<i>P</i> value	<0.05	<0.03	<0.01

The table lists average values obtained by 10 individual experiments of three solubilized spidroin samples each that were aggregated into buffer solution. Significant differences between the means are defined as *P* values below 0.05 and are indicated in bold

have undergone sequence homogenization by homologous recombination with unequal crossing over and gene conversion events (Beckwitt et al. 1998; Hayashi 2002). As a result, repetitive units evolved coordinately within a species and are more similar to each other along the same gene than between ortholog genes rendering them unusable to extract exact phylogenetic relationships (Hayashi and Lewis 2000; Liao et al. 1997). Homogenization is thought to rapidly spread advantageous patterns or repeats throughout a gene and applies in particular also for silkworm silk genes (Mita et al. 1994; Schlotterer and Tautz 1994).

Homologous recombination in the silkworm gene of the moth *B. mori* is triggered by mini satellite or Chi sequences, which are contained in the code for the amino acid sequel AGG of the silk's heavy chain gene and which are genetic hotspots that induce crossover events (Jeffreys et al. 1985; Lam et al. 1974; Mita et al. 1994; Paulsson et al. 1992; Sezutsu and Yukuhiro 2000). Similar mechanisms act also in the fly *Chironomus tetans* and they would appear to apply for spider genes as well (Galli and Wieslander 1993; Manning and Gage 1980; Mita et al. 1994; Paulsson et al. 1992). Indeed, Chi sequences coding for AGG are often found in MaSp 1 after sequences coding for polyalanine stretches (supplementary Alignment 1) (Craig 2003). For the spidroin 1 sequences available to date the last four repeat units before the conserved C-terminus group by position (Hayashi and Lewis 2000). Therefore, they are more similar between ortholog genes than to their corresponding repeat units on the same gene, suggesting that they evolved primarily by single point mutations and localized insertion/deletion events and are exempted from concerted evolution (Beckwitt et al. 1998). However, more extended variations were obvious for some of our clones (Fig. 1 and supplementary Fig. 1). In particular, for clone AY654292 already the repeat unit immediately juxtaposed to the C-terminal sequences showed extensive deletions compared to positional equivalent repeat units of other clones, suggesting that homologous recombination with unequal cross over will act on all repeat units of spidroin 1. In respect to the abundance of Chi sequences, the last repeat units indeed do not represent special cases low in

these. Variations are not only evident for repetitive units close to the C-terminus, but also for those in continuation to the N-terminal end (supplementary Figs. 1, 2) (Gaines and Marcotte 2008). This suggests that heterogeneities are likely to be manifested throughout the repetitive sequence parts. Since sequence and length polymorphisms were collinear between DNA, RNA, and protein in some of our examples, we can rule out RNA editing and splicing as general mechanisms of their generations.

In contrast, the last four repeat units in MaSp 2 were conserved between individual proteins and sequence variations were in general less pronounced between repeats within one protein (Fig. 2 and supplementary Fig. 2). Repeat units of this spidroin type seem, therefore, to originate primarily by confined deletion/insertion events and not by homologous recombination with unequal cross over. The spidroin 2 gene has likely been evolved from a spidroin 1 gene duplicate followed by sequence divergence, and the loss in potential recombination sites as the amino acid sequel PGG is rarely encoded by a Chi-like sequence might have led to gene stabilization (supplementary Alignment 2) (Ayoub et al. 2007). We cannot, however, conclusively rule out homologous recombination with equal cross over that also would sustain a more similar characteristic between the repeat units of MaSp 2.

Multiple allelic variants originating from potential different spidroin 1 genes can exist within an individual genome, which can be differentially regulated

Multiple loci of major ampullate spidroin genes have been suggested so far across taxa within the genera *Euprosthenoops* (Pisauridae), *Latrodectus* (Theridiidae) and *Nephila* (Nephilidae) (Ayoub et al. 2007; Rising et al. 2007; Gaines and Marcotte 2008). As a matter of fact, two of the sequences reported here, var 4 (AY654391) and var 5 (AY654392), are encoded by two potential different genes, MaSp 1B and MaSp 1A, respectively, with different allelic variants identified for the latter by single nucleotide polymorphisms (SNPs) (Gaines and Marcotte 2008). Both genes have been reported to be commonly expressed and in three of our screened individuals (C1S2, C1S4, and C1S5) this was also the case. The similar abundance of cDNA clones suggested that expression levels of both genes were

fairly comparable. (Table 1). Both sequences showed single nucleotide polymorphisms (SNPs) in the C-terminal parts suggesting that they are encoded by different genes and not merely by alleles (supplementary Tables 1, 2). The reason behind this assumption is the fact that one gene in a diploid genome can only show a maximum of two different sequences from the two allelic forms. Intriguingly, C1S1 and C1S3 expressed besides two major and equally expressed variants a third underrepresented gene version (Table 1). Our expression analysis, therefore, give independent evidence that a minimum of possibly three different gene copies can exist for spidroin 1 in some *Nephila* individuals. Given the diploid genome character of this spider genus duplication and multiplicity of spidroin 1 genes would be a likely scenario to account for these variants (Araujo et al. 2005; Gaines and Marcotte 2008). Two of the allelic MaSp 1 genes in *Euprosthenoops australis*, termed MaSp 1a and MaSp 1b, were also found to be different in their abundance in cDNA screens, albeit the possibility that Masp 1b represents a pseudo-gene could not be completely ruled out (Rising et al. 2007). For *N. pilipes*, two C-terminal variants of spidroin 1 were found to be commonly expressed as well (Tai et al. 2004). However, they might be rather allelic variants than being encoded by independent genes as there is no support for different genetic loci so far. Independent cytogenetic studies are required to corroborate and unambiguously demonstrate the existence of different genetic loci.

Multiple genes within a genome would have the benefit to enhance transcript levels when being commonly expressed (Kondrashov and Kondrashov 2006). Hence, they will provide means for the spider to cope with a huge demand for silk (Ayoub and Hayashi 2008; Gaines and Marcotte 2008; Rasch and Connelly 2005). On the other hand, differential expression sets the stage for functional diversities of the associated gene products (Li et al. 2005). These modes of operation must not be mutually exclusive and could potentially work jointly in the case of the species *Nephila*. The different abundance of distinct transcripts and cDNAs from one individual leaves the possibility that several gene copies are commonly expressed, but to different extents (Table 1; Figs. 3, 4; supplementary Figs. 3, 4). This implies variations in the genes' control regions, which are indeed evident in the case of the *Latrodectus* spidroin 1 genes (Ayoub et al. 2007). In this way, a spider would be able to produce large quantities of major ampullate silk and might partly tailor its properties toward its specialized use as for example the orb web's frame threads or the lifeline (Foelix 1996). Differential expression between the juvenile and adult state could be another benefit of multiple genes and occurs in the case of flagelliform silk employed for prey capture to adapt their properties to the preferred prey (Higgins et al. 2007). Silk

composition can vary in dependence of a spider's diet (Craig et al. 2000; Tso et al. 2005). However, this cannot be totally accounted for by the differential expression of the *Nephila* gene variants given their similarities in amino acid composition over their sequenced length (supplementary Fig. 1).

Multiple genes can either be created by gene duplication or by endopolyploidy, the latter leading to polytene chromosomes (Gregory and Hebert 1999). Endopolyploidy is well established for silk gland cells of the silkworm *B. mori* and give rise to their characteristic giant nuclei (Gregory and Shorthouse 2003; Perdrix-Gillot 1979; Rasch and Connelly 2005). Endonuclear DNA replication does also occur in somatic tissue cells of spiders and genomic amplification is evident in silk and poison glands of the spider *Pholcus phalangoides* (Gregory and Shorthouse 2003; Perdrix-Gillot 1979; Rasch and Connelly 2005). Polyploidy can lead to enhanced homologous recombination events, for example during mitosis, as genes on different DNA strands in polytene chromosomes line up to each other (Shao et al. 1999; Storchova et al. 2006). The contribution of gene amplification by endopolyploidy in the case of *N. clavipes* has to await, however, further cytogenetic analysis.

Two N-terminal and three C-terminal MaSp 2 variants have been identified so far in *L. hesperus* and *Argiope amoena*, but there is no indication if these are encoded by different genes (Ayoub et al. 2007; Pan et al. 2007). We also find multiple C-terminal forms of MaSp 2 in *N. clavipes* (Table 1). However, since not more than two different variants were commonly expressed in all screened individuals and no SNPs were detected, we have also no evidence for more than one MaSp 2 gene within the spider's genome and we never observed more than one band in gene and transcript blots (Table 1; Fig. 5). Both allelic copies seemed to be expressed to the same extent, suggesting that they are commonly regulated. The lack of establishment and maintenance of multiple spidroin 2 gene copies might reflect their more recent occurrence within orb-weaving spiders and their relatives (Orbiculariae) as well as their lower abundance and more confined and specialized functionality of their protein products (Ayoub et al. 2007; Hinman and Lewis 1992; Spöner et al. 2005a, b; Guehrs et al. 2008).

Extensive length polymorphisms of spidroin 1 genes can occur within an individual, but are contained within a population

Rearrangements in the *B. mori* silk gene can transform into size differences of the proteins in the range of 10–15% (Galli and Wieslander 1993; Lizardi et al. 1979; Manning and Gage 1980). Most of the size variations of spidroin 1 transcripts and their translated proteins stayed also approximately within these limits, but deviations of nearly as much as 20–50% below and above the average were also

observed (Fig. 3; supplementary Figs. 3, 4). The heterogeneities within the repeat units expressed as indels of sequence motifs cannot account for such huge deviations and instead the proteins must consist of less or additional repeat units. The extreme polymorphisms are likely to arise by intra-chromosomal homologous recombination with unequal crossing over during meiosis (Cromie and Smith 2007). Huge expansions of genetic loci are not without precedent and do occur for example in the case of the human U2 snRNA gene (Liao et al. 1997). The low numbers of genetic hotspots in the spidroin 2 gene would predict that such extended polymorphisms should be more limited in this type and we have as a matter of fact not detected any in our screens (Fig. 5).

The occurrence of huge length deviations for spidroin 1 was, on the other hand, only on the order of $\sim 1\%$ within the screened number of individuals and hence rare within a population. Extension of sequences coding for the repetitive part of spidroin 1 should be subject to an increased probability of recombination rates leading to gene instabilities and loss of functional genes. Such mechanisms would fall into the birth-and-death model of gene evolution and are supported by the existence of spidroin 1 pseudo-genes in *Latrodectus*, where concerted evolution is likely to act jointly with duplication and loss of spidroin 1 genes (Ayoub et al. 2007; Nei and Rooney 2005).

Purifying selection against unfavorable characteristics of extreme over- or undersized spidroins might be coupled to the birth-and-death mechanism (Nei et al. 2000; Piontkivska et al. 2002). Major ampullate silk has to serve two purposes with opposing requirements on its mechanical characteristics: as the lifeline tensile strength is wanted, as the frame thread extensibility (Lewis 1992; Vollrath and Knight 2001, p. 264). Albeit characteristics can partly be influenced by the spinning speed, the primary structure will also have a profound impact and is likely to represent the best compromise to serve both functions (Thiel et al. 1997; Vollrath 1999). Increased spidroin sizes were associated with higher tensile strength but lowered extensibility and vice versa smaller spidroins gave rise to silk biased for better extensibility but at the cost of reduced tensile strength (Fig. 6; supplementary Fig. 5; Table 2; supplementary Table 3). Extreme deviations could, therefore, lead to a loss in sufficient adaptability limiting either the capability of the silk to support the body weight of a falling spider or to dissipate the kinetic energy of prey impact on the orb web (Gosline et al. 2003). The observed increase in diameters of fibers composed of longer spidroins can partly compensate for their lower toughness to keep the total energy-to-break of the fiber higher; however, there will be physiological limitations for such an adjustment (Swanson et al. 2007; Vollrath and Knight 2001). Ecological studies will be required to support these assumptions and more

samples will be needed to increase the statistical significance of these results.

Purifying selection against off-average sized spidroins might also act against suboptimal polymerization. Spinning is an intricate process where premature precipitation of silk proteins in the highly concentrated spinning dope has to be avoided without compromising efficient crystallization upon exit from the spinneret (Bini et al. 2004; Jin and Kaplan 2003; Kaplan et al. 1991; Vollrath and Knight 2001). Such a balance might be perturbed by under- or oversized spidroins, since they form more readily aggregates as their length increase and are impeded in efficient polymerization the smaller they become (Fig. 6; supplementary Fig. 6; Table 3; supplementary Table 4). Since sample availability for off-average sized spidroins was limited, we cannot substantiate the findings in a statistical significant manner. However, we also resorted to genetically and chemically generated spidroins of different lengths, which essentially showed the same polymerization efficiencies in dependence of their size substantiating our finding with naturally occurring variants.

Spidroin genes are homogenized in their lengths within one genome

Unexpectedly, allelic spidroin gene variants and those potentially encoded by different loci within one genome, no matter if of average size or a deviation from it, were of similar length, since only one or closely sized transcripts and gene fragments were discovered in all studied cases (Figs. 3, 4; supplementary Figs. 3, 4). Length homogenization within an individual could involve mitotic unequal intra-chromosomal homologous recombination or gene conversion within gland precursor cells between different spidroin genetic loci and allelic genes including available pseudo-genes (Balakirev and Ayala 2003; Petes and Hill 1988). Alternatively, chromatin has been described to be organized into specific territories and it is a possibility that spider gene alleles and loci share their territory coming within sufficient proximity to each other to recombine (Cremer et al. 1996). Recombination between multiple MaSp 1 loci has already been observed in the case of *Latrodectus* (Ayoub and Hayashi 2008). Within a population meiotic intra-chromosomal homologous recombination with unequal cross over could also act for homogenization toward the average length, dependent on where crossing over points will occur. Through such mechanisms length variations could not only be created but also efficiently restricted within individuals and populations.

Conclusions

The major ampullate spidroin 1 gene structure is well suited to enhance protein evolution as evidenced by the

occurrence of a second spidroin gene to which it gave rise to; yet, the organization of its protein product acts as a constraint (Ayoub and Hayashi 2008; Craig 2003). Purifying selection is the most likely mechanism that results in the maintenance of gene size within a tolerable range and that guarantees a balance of opposing requirements for the protein. In this context, primary sequence and length variations are limited, and a spider has to rely additionally on its sophisticated spinning apparatus to fine tune properties of its silk by varying the spinning conditions (Vollrath et al. 2001). The major ampullate spidroin 2 genes, on the other hand, have approached nearer to a climax state with the loss of recombination prone sequences and a more restricted function of the protein. In the extreme case represented in the genus *Araneus*, a MaSp 1 homolog seems to have been replaced completely by a second MaSp 2 paralog that has, however, maintained the majority of the former's properties (Guerette et al. 1996; Huemmerich et al. 2004a; Huemmerich et al. 2004b). Such a scenario might currently take place in *Latrodectus* as one of its MaSp 1 loci shows sequence characteristics of MaSp 2 genes (Ayoub and Hayashi 2008).

Acknowledgments The authors want to thank Ursula Stephan, Irmgard Tiroke, and Kristina Hartung for excellent technical assistance. The help of Hans-Georg Schmidt (Thueringisches Institut fuer Textil- und Kunststoff-Forschung e. V., Schwarza, Germany) for the tensile measurements is appreciated. This study was supported by Bundesministerium fuer Forschung und Bildung (BMBF FKZ 031110), Bundesministerium fuer Landwirtschaft (BML FKZ 98NR049) and Thueringer Ministerium fuer Wissenschaft, Forschung und Kultur (TMWFK B307-99-001).

References

- Araujo D, Cella MD, Brescovit DA (2005) Cytogenetic analysis of the neotropical spider *Nephilengys cruentata* (Araneomorphae, Tetragnathidae): standard staining norms, C-bands and base-specific fluorochromes. *Braz J Biol* 65:193–202
- Augsten K, Muehlig P, Hermann C (2000) Glycoproteins and skin-core structure in *Nephila clavipes* spider silk observed by light and electron microscopy. *Scanning* 22:12–15
- Ayoub NA, Hayashi CY (2008) Multiple recombining loci encode MaSp 1, the primary constituent of dragline silk, in widow spiders (*Latrodectus*: Theridiidae). *Mol Biol Evol* 25:277–286
- Ayoub NA, Garb JE, Tinghitella RM, Collin MA, Hayashi CY (2007) Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS ONE* 2:e514
- Balakirev ES, Ayala FJ (2003) Pseudogenes: are they “junk” or functional DNA? *Annu Rev Genet* 37:123–151
- Baldauf SL (2003) Phylogeny for the faint of heart: a tutorial. *Trends Genet* 19:345–351
- Beckwitt R, Arcidiacono S, Stote R (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
- Beckwitt R, Arcidiacono S, Stote R (1998) Evolution of repetitive proteins: spider silks from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae). *Insect Biochem Mol Biol* 28:121–130
- Bini E, Knight DP, Kaplan DL (2004) Mapping domain structures in silks from insects and spiders related to protein assembly. *J Mol Biol* 335:27–40
- Candelas GC, Ortiz A, Ortiz N (1989) Features of the cell-free translation of a spider fibroin mRNA. *Biochem Cell Biol* 67:173–176
- Challis RJ, Goodacre SL, Hewitt GM (2006) Evolution of spider silks: conservation and diversification of the C-terminus. *Insect Mol Biol* 15:45–56
- Craig CC (2003) Spider web and silks. Tracing evolution from molecules to genes to phenotypes. Oxford University Press, Oxford, New York
- Craig CL (2004) Broad pattern of speciation are correlated with the evolution of new silk protein in spiders but not in Leptidoptera. *Biomacromolecules* 5:739–743
- Craig CL, Riekel C, Herberstein ME, Weber RS, Kaplan D, Pierce NE (2000) Evidence for diet effects on the composition of silk proteins produced by spiders. *Mol Biol Evol* 17:1904–1913
- Cremer C, Munkel C, Granzow M, Jauch A, Dietzel S, Eils R, Guan XY, Meltzer PS, Trent JM, Langowski J, Cremer T (1996) Nuclear architecture and the induction of chromosomal aberrations. *Mutat Res* 366:97–116
- Cromie GA, Smith GR (2007) Branching out: meiotic recombination and its regulation. *Trends Cell Biol* 17:448–455
- Dong Z, Lewis RV, Middaugh CR (1991) Molecular mechanism of spider silk elasticity. *Arch Biochem Biophys* 284:53–57
- Foelix RF (1996) Biology of spiders. Oxford University Press, New York
- Galli J, Wieslander L (1993) A repetitive secretory protein gene of a novel type in *Chironimos tetans* is specifically expressed in the salivary glands and exhibits extensive length polymorphism. *J Biol Chem* 268:11888–11893
- Garb JE, Dimauro T, Vo V, Hayashi CY (2006) Silk genes support the single origin of orb webs. *Science* 312:1762
- Gatesy J, Hayashi C, Motriuk D, Woods J, Lewis R (2001) Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science* 291:2603–2605
- Gosline JM, Guerette PA, Ortlepp CS, Savage KN (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
- Gosline JM, Lillie M, Carrington E, Guerette P, Ortlepp C, Savage K (2003) Elastomeric proteins: structures, biomedical properties, and biological roles. In: Sherry PR, Tatham AS, Bailey AJ (eds) Elastic proteins: biological roles and mechanical properties. Cambridge University Press, Cambridge, UK, pp 15–38
- Gregory TR, Hebert PD (1999) The modulation of DNA content: proximate causes and ultimate consequences. *Genome Res* 9:317–324
- Gregory TR, Shorthouse DP (2003) Genome sizes of spiders. *J Hered* 94:285–290
- Grubb DT, Jelinski LW (1997) Fiber morphology of spider silk: the effects of tensile deformation. *Macromolecules* 30:2860–2867
- Guehrs KH, Schlott B, Grosse F, Weisshart K (2008) Environmental conditions impinge on dragline silk protein composition. *Insect Mol Biol* 17:553–564
- Guerette PA, Ginzinger DG, Weber BH, Gosline JM (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science* 272:112–115

- Hayashi CY (2002) Evolution of spider silk proteins: insight from phylogenetic analyses. In: DeSalle R, Giribet G, Wheeler WC (eds) Molecular systematics and evolution: theory and practice. Birkhaeuser, Cambridge, pp 209–223
- Hayashi CY, Lewis RV (2000) Molecular architecture and evolution of a modular spider silk protein gene. *Science* 287:1477–1479
- Hayashi CY, Shipley NH, Lewis RV (1999) Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *Int J Biol Macromol* 24:271–275
- Higgins LE, White S, Nunez-Farfan J, Vargas J (2007) Patterns of variation among distinct alleles of the flag silk gene from *Nephila clavipes*. *Int J Biol Macromol* 40:201–216
- Hinman MB, Lewis RV (1992) Isolation of a clone encoding a second dragline silk fibroin. *Nephila clavipes* dragline silk is a two-protein fiber. *J Biol Chem* 267:19320–19324
- Huemmerich D, Helsen CW, Quedzuweit S, Oschmann J, Rudolph R, Scheibel T (2004a) Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry* 43:13604–13612
- Huemmerich D, Scheibel T, Vollrath F, Cohen S, Gat U, Ittah S (2004b) Novel assembly properties of recombinant spider dragline silk proteins. *Curr Biol* 14:2070–2074
- Jackson C, O'Brien JP (1995) Molecular weight distribution of *Nephila clavipes* dragline silk. *Macromolecules* 28:5975–5977
- Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* 314:67–73
- Jin HJ, Kaplan DL (2003) Mechanism of silk processing in insects and spiders. *Nature* 424:1057–1061
- Kaplan DL, Lombardi SJ, Muller WS, Fossey SA (1991) Silk. In: Byrom D (ed) Biomaterial—Novel materials from biological sources. Macmillan Publishers Ltd and ICI Biological Products Business, New York, p 1
- Kondrashov FA, Kondrashov AS (2006) Role of selection in fixation of gene duplications. *J Theor Biol* 239:141–151
- Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet* 39:309–338
- Lam S, Stahl M, McMilin K, Stahl F (1974) Rec-mediate recombination hot spot activity in bacteriophage lambda II: a mutation which causes hot spot activity. *Genetics* 77:425–433
- Lewis RV (1992) Spider silk: the unraveling of a mystery. *Acc Chem Res* 25:392–398
- Lewis RV (2006) Spider silk: ancient ideas for new biomaterials. *Chem Rev* 106:3762–3774
- Li WH, Yang J, Gu X (2005) Expression divergence between duplicate genes. *Trends Genet* 21:602–807
- Liao D, Pavelitz T, Kidd JR, Kidd KK, Weiner AM (1997) Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *EMBO J* 16:588–598
- Lizardi PM, Mahdavi V, Shileds D, Candelas GC (1979) Discontinuous translation of silk fibroin in a reticulocyte cell free system and in intact glands. *Proc Natl Acad Sci USA* 76:6211–6215
- Manning RF, Gage LP (1980) Internal structure of the silk fibroin gene of *Bombyx mori*. II. Remarkable polymorphism of the organization of crystalline and amorphous coding sequences. *J Biol Chem* 255:9451–9457
- Mello C, Senecal K, Yeung B, Vourus P, Kaplan D (1994) Initial characterization of *Nephila clavipes* dragline protein. In: Kaplan WA D, Farmer W, Viney C (eds) Silk polymers: material science and biotechnology. American Chemical Society, Washington, pp 67–79
- Mita K, Ichimura S, James TC (1994) Highly repetitive structure and its organization of the silk fibroin gene. *J Mol Evol* 38:583–592
- Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39:121–152
- Nei M, Rogozin IB, Piontkivska H (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci USA* 97:10866–10871
- Pan HC, Song DX, Zhou KY, Zhu GP (2007) Cloning and prokaryotic expression of major ampullate spidroin gene of spider. *Sheng Wu Gong Cheng Xue Bao* 23:446–451
- Paulsson G, Hoog C, Bernholm K, Weislander L (1992) Balbiani ring I gene in *Chironimos tetans*: sequence organization and dynamics of a coding minisatellite. *J Mol Biol* 225:349–361
- Perdrix-Gillot S (1979) DNA synthesis and endomitoses in the giant nuclei of the silk gland of *Bombyx mori*. *Biochimie* 61:171–204
- Petes TD, Hill CW (1988) Recombination between repeated genes in microorganisms. *Annu Rev Genet* 22:147–168
- Piontkivska H, Rooney AP, Nei M (2002) Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol Biol Evol* 19:689–697
- Rasch EM, Connelly BA (2005) Genome size and endonuclear DNA replication in spiders. *J Morphol* 265:209–214
- Riekell C, Branden C, Craig C, Ferrero C, Heidelberg F, Muller M (1999) Aspects of X-ray diffraction on single spider fibers. *Int J Biol Macromol* 24:179–186
- Rising A, Nimmervoll H, Grip S, Fernandez-Arias A, Storckenfeldt E, Knight DP, Vollrath F, Engstrom W (2005) Spider silk proteins—mechanical property and gene sequence. *Zool Sci* 22:273–281
- Rising A, Johansson J, Larson G, Bongcam-Rudloff E, Engstrom W, Hjalmar G (2007) Major ampullate spidroins from *Euprosthenois australis*: multiplicity at protein, mRNA and gene levels. *Insect Mol Biol* 16:551–561
- Schlotterer C, Tautz D (1994) Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr Biol* 4:777–783
- Selden PA, Gall JC (1992) A Triassic Mygalomorph spider from the Northern Vosges, France. *Palaeontology* 35:257–285
- Sezutsu H, Yukuhiro K (2000) Dynamic rearrangement within the *Antheraea pernyi* silk fibroin gene is associated with four types of repetitive units. *J Mol Evol* 51:329–338
- Shao C, Deng L, Henegariu O, Liang L, Raikwar N, Sahota A, Stambrook PJ, Tischfield JA (1999) Mitotic recombination produces the majority of recessive fibroblast variants in heterozygous mice. *Proc Natl Acad Sci USA* 96:9230–9235
- Simmons AH, Michal CA, Jelinski LW (1996) Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science* 271:84–87
- Sponner A, Schlott B, Vollrath F, Unger E, Grosse F, Weisshart K (2005a) Characterization of the protein components of *Nephila clavipes* Dragline silk. *Biochemistry* 44:4727–4736
- Sponner A, Unger E, Grosse F, Weisshart K (2005b) Differential polymerization of the two main protein components of dragline silk during fibre spinning. *Nat Mater* 4:772–775
- Storchova Z, Breneman A, Cande J, Dunn J, Burbank K, O'Toole E, Pellman D (2006) Genome-wide genetic analysis of polyploidy in yeast. *Nature* 443:541–547
- Swanson BO, Blackledge TA, Hayashi CY (2007) Spider capture silk: performance implications of variation in an exceptional biomaterial. *J Exp Zool Part A Ecol Genet Physiol* 307:654–666
- Tai PL, Hwang GY, Tso IM (2004) Inter-specific sequence conservation and intra-individual sequence variation in a spider silk gene. *Int J Biol Macromol* 34:295–301
- Thiel B, Viney C (1995) A nonperiodic lattice model for crystals in *Nephila clavipes* major ampullate silk. *Mater Res Bull* 20:52–56
- Thiel BL, Guess KB, Viney C (1997) Spider major ampullate silk (drag line): smart composite processing based on imperfect crystals. *J Microsc* 185:179–187

- Tso IM, Wu HC, Hwang IR (2005) Giant wood spider *Nephila pilipes* alters silk protein in response to prey variation. *J Exp Biol* 208:1053–1061
- van Beek JD, Kuemmerlen J, Vollrath F, Meier BH (1999) Supercontracted spider dragline silk: a solid-state NMR study of the local structure. *Int J Biol Macromol* 24:173–178
- Vollrath F (1999) Biology of spider silk. *Int J Biol Macromol* 24:81–88
- Vollrath F, Knight DP (2001) Liquid crystalline spinning of spider silk. *Nature* 410:541–548
- Vollrath F, Madsen B, Shao Z (2001) The effect of spinning conditions on the mechanics of a spider's dragline silk. *Proc R Soc Lond B Biol Sci* 268:2339–2346
- Gaines WA, Marcotte WR Jr (2008) Identification and characterization of multiple Spidroin 1 genes encoding major ampullate silk proteins in *Nephila clavipes*. *Insect Mol Biol* 17:465–474
- Xu M, Lewis RV (1990) Structure of a protein superfiber: spider dragline silk. *Proc Natl Acad Sci USA* 87:7120–7124