

Reviewed Preprint

v1 • April 20, 2026

Not revised

✉ For correspondence:

xiang_shine@foxmail.com**Competing interests:** No competing interests declared**Funding:** See [page 13](#)**Reviewing editor:** Martin Graña, Institut Pasteur de Montevideo, Uruguay

© 2026, Zhu et al. This article is distributed under the terms of the

[Creative Commons Attribution](#)[License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

New insights into the evolution of spider silk proteins illuminated by long-read transcriptomes

Kesen Zhu^{1,2}, Shiyi Zhou¹, Mo Lyu¹, Jiahao Xiang¹, Shaohan Niu¹, Yongping Huang³, Lei Gao¹, Anjiang Tan⁴, Hui Xiang^{1,2} ✉¹Guangdong Provincial Key Laboratory of Insect Developmental Biology and Applied Technology, Guangzhou Key Laboratory of Insect Development Regulation and Application Research, Institute of Insect Science and Technology, School of Life Sciences, South China Normal University, Guangzhou, China • ²Laboratory for Lingnan Modern Agriculture, Guangzhou, China • ³School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai, China • ⁴Jiangsu University of Science and Technology, Chinese Academy of Agricultural Sciences, Jiangsu, China

eLife Assessment

This **valuable** study by Zhu et al. offers a high-resolution evolutionary framework for spider silk proteins (spidroins) through long-read transcriptomics across a broad phylogenetic range, with theoretical implications for protein family evolution, biomaterials, and silk biology. By identifying putative ancestral spidroin templates in early-diverging spiders, the authors make a significant contribution to understanding genetic innovations underlying silk diversification. The long-read sequencing approach is well-suited to these highly repetitive genes. However, the support is **incomplete**: key claims regarding direct ancestry between silk protein families, the independent origin of certain silk types, and the co-option of flagelliform spidroins in non-web-building spiders rely on absence-based inferences and indirect phylogenetic reasoning that the data cannot yet fully substantiate, and some gene family assignments overreach the available molecular evidence.

<https://doi.org/10.7554/eLife.110512.1.sa3>

Abstract

Spider silk has long fascinated scientists because of its exceptional mechanical properties, yet the evolutionary origins and diversification of spider silk proteins (spidroins) remain incompletely understood. By analyzing long-read spidroin transcripts from spiders representing key evolutionary lineages, we identify two ancestral proteins present in basal spiders: an alanine–serine-rich (AS-type) protein and a glycine–serine-rich (GS-type) protein. These ancestral proteins likely served as primary evolutionary templates for the diversification of modern spider silks. We show that the AS-type ancestral spidroin remained relatively conserved and ultimately gave rise to the specialized tubuliform spidroin (TuSp) used in eggcase silk. In contrast, the GS-type protein underwent extensive functional radiation, evolving into differentiated minor ampullate spidroins (MiSp) through the acquisition of distinct terminal domains and β -sheet–associated structural innovations. Our results further suggest that major ampullate spidroins (MaSp) may have originated from MiSp, with the GS motif representing an evolutionarily favorable genetic substrate for the emergence of high-performance silk proteins. Finally, we propose a revised evolutionary trajectory for flagelliform spidroins (Flag), suggesting that they were co-opted as components of ampullate silk during functional degeneration in modern non-web-weaving RTA clade spiders. Together, these findings provide a high-resolution framework for understanding the genetic innovations that drove the diversification of spider silk.

Introduction

Beyond Lepidopteran silk, the dragline silk of spiders (order Araneae) has attracted considerable attention because of its exceptional tensile strength and toughness (Blamires et al., 2017). Unlike Lepidopterans, which typically produce a single silk type, spiders use specialized silk glands to spin a wide diversity of fibers. At the genetic level, this functional diversity reflects the evolutionary radiation of spider silk proteins (spidroins), which are encoded by a multigene family. In web-weaving spiders, dragline silk consists primarily of major ampullate (MA) silk, which is composed of major ampullate spidroins (MaSp).

In a typical orb web, MA silk forms the frame and radial threads (Dugger et al., 2020), whereas its counterpart, minor ampullate (MI) silk (Papadopoulos et al., 2009), provides additional structural reinforcement (Chen et al., 2012). MaSp and MiSp (particularly MaSp1 and MiSp) exhibit striking similarities in amino acid composition and repetitive motif organization (Nakamura et al., 2023; Papadopoulos et al., 2009). They are distinguished from other spidroins by their relatively short and variable repetitive units, in contrast to the long and highly stereotyped repeats characteristic of other silk types (Hayashi et al., 2004; Zhou et al., 2021). These include flagelliform spidroin (Flag) and aggregate spidroin (AgSp), which together form the “wet” spiral capture thread of araneoid webs (Stellwagen & Renberg, 2019), as well as Flag and cribellar spidroins (CrSp) in the “dry” capture threads of deinopoid webs (J. E. Garb et al., 2006). Additional specialized spidroins include the eggcase-associated tubuliform (TuSp) and aciniform (AcSp) spidroins (Garb & Hayashi, 2005), and pyriform spidroin (PySp), which is used to produce attachment discs (Chaw et al., 2017). Chemically, ampullate spidroins are enriched in alanine (A) and glycine (G), forming poly-A, poly-GA, or GGX motifs. In contrast, TuSp, AcSp, and PySp contain lower proportions of glycine but are enriched in serine (S) within their stereotyped repeats (Hayashi et al., 2004). Ampullate spidroins are further distinguished from Flag by C-terminal domain substitutions that facilitate the liquid-to-solid structural transitions required for fiber assembly (Rat et al., 2023).

The divergence of these specialized spidroins raises fundamental questions about how ancestral genetic material in basal spider lineages evolved into the diverse spidroins observed in modern spiders (Arakawa et al., 2022; Garb et al., 2010). Phylogenetically, Araneae comprise the basal Mesothelae, the subsequently diverged Mygalomorphae, and the highly diverse Araneomorphae (Craig, 2004). Araneomorphae include the orb-weaving Araneoidea and Deinopoidea, as well as the “RTA” (retrolateral tibial apophysis) clade, which represents the most successful non-web-building lineage derived from orb-weaving ancestors (Arakawa et al., 2022). Although MaSp is thought to have emerged during the initial radiation of Araneomorphae (Craig, 2004), the evolutionary origin of GA-rich ampullate spidroins remains unresolved (Fan et al., 2023). Efforts to identify ancestral forms in basal lineages have been limited by incomplete genomic resources. Although GA- and poly-A-rich motifs occur in Mygalomorphae, their relatively low glycine content suggests that they are unlikely to represent direct precursors of modern ampullate spidroins (Garb et al., 2007). In the basal Mesothelae, reported spidroin sequences are largely restricted to CySp-like and AcSp-like proteins (Arakawa et al., 2022; Starrett et al., 2012), further obscuring the nature of ancestral spidroins. In addition, the apparent loss of Flag spidroins in the RTA clade, which produces diverse silks for draglines and egg protection but lacks foraging webs (Arakawa et al., 2022), poses an intriguing evolutionary question regarding how silk components are lost or repurposed.

In this study, we utilized long-read transcriptomics to analyze 12 spider species representing both the most ancient and the most derived lineages of Araneae. We identify two primordial spidroins in basal Mesothelae that likely served as ancestral templates for modern AS-rich and GA-rich spidroins. Our analyses reveal contrasting evolutionary trajectories for these proteins, including strong conservation of the eggcase spidroin TuSp and extensive innovation of MiSp. We further propose that MiSp-like proteins provided the genetic foundation for the emergence of MaSp and

identify poly-GS motifs as a versatile substrate for silk protein evolution. Finally, we present evidence that Flag spidroins were not simply lost in RTA spiders but were co-opted as auxiliary components of ampullate silk during their functional degeneration.

Results

Identification of spidroin resources across representative nodes in spider evolution

Using long-read transcriptome sequencing corrected with short-read RNA-seq data, we identified a total of 60 spidroin sequences across 12 representative spider (Araneae) species, among which 30 were full-length (figure supplements 3 and 4, Dataset 1). These species span the major evolutionary lineages of spiders, including Mesothelae, Mygalomorphae, Haplogynae, and the later-diverging Araneoidea and RTA groups within Araneomorphae. We constructed a phylogenomic tree based on single-copy orthologous proteins to resolve lineage relationships and estimate divergence times (Figure 1). The resulting topology and divergence estimates are generally consistent with the established phylogenetic backbone of spiders (Fernandez et al., 2018; Garrison et al., 2015). Notably, species from the families Eresidae and Araneoidea clustered together into a single clade that is sister to the RTA clade (Figure 1—figure supplements 1 and 2). We then mapped the identified spidroins onto the species tree (Figure 1). Possibly due to the limitation of whole-body sampling, we were not able to obtain the complete repertoire of spidroins for each spider. However, we were able to reconstruct a consistent evolutionary landscape, showing that basal spiders have fewer and more divergent spidroins, while the emergence of Araneomorphae is accompanied by diversification of spidroins.

Two highly divergent putative spidroins in a basal species belonging to Mesothelae

In the basal Mesothelae genus *Liphistius*, which represents the earliest-diverging spider lineage (Wheeler et al., 2016; Xin et al., 2015), we identified two highly divergent spidroin-like transcripts encoding putative spidroin-like proteins, *Liphistius* sp._6705 and *Liphistius* sp._5400 (Figure 2A). As to *Liphistius* sp._6705, we recovered 12 repetitive units along with the C-terminal sequence. Each repetitive unit contains poly-alanine (polyA) and poly-threonine (polyT) motifs, while the remaining non-polymeric regions are dominated by alanine (A) and serine (S). We termed this the AS-type spidroin. The second transcript, *Liphistius* sp._5400, which we term the GS-type spidroin, was recovered as a full-length sequence and is characterized by a high frequency of glycine-serine (GS) repetitive motifs. The Poly-GS motifs are known β -sheet-forming elements and are also commonly observed in silk-producing insects, including silkworms (Mita & James, 1994), earwigs (Sutherland et al., 2010) and bagworm moths (Kono et al., 2019). In this protein, poly-GS motifs are interspersed with a small number of glycine-alanine (GA) sequences and terminate in a spacer region of unknown function, together forming a repetitive unit. We identified four such repetitive units together with the C-terminal sequence in this partial protein (Figure 2A). Notably, neither the N-terminal nor the C-terminal sequence of *Liphistius* sp._5400 shows homology to known spidroins (Figure 2B), suggesting that this spidroin-like protein may not assemble into typical silk fibers (Li et al., 2022), despite being substantially expressed (figure supplement 3).

Phylogenetic trees based on different domains elucidated the evolutionary pattern of spidroins

Phylogenetic analyses based on different spidroin domains revealed distinct evolutionary patterns. In the N-terminal phylogeny, spidroins from the basal Mesothelae genus *Liphistius* (Xin et al., 2015) occupied the most basal position with robust support (Figure 3A). AgSp and PySp then diverged as a sister group to the remaining spidroins, suggesting that their terminal domains retain relatively basal characteristics. N-terminal sequences from the later-diverging Mesothelae

Figure 1. Phylogene and identified spidroins of the spiders in this study.

The phylogenetic status of the spiders was marked with different colors. the green line represents Mesothelae, the blue line represents Mygalomorphae, and the red and purple line represents Araneomorphae, specific species are marked at the end of the branch. An asterisk (*) indicates that an event with a significant impact is likely to occur, the ruler represents when the event occurred, the specific time scale is shown on the X axis in millions of years. Each colored box represents a silk type, the gray boxes represent unidentified spidroins, and numbers in each box indicate the total number of occurrences of each silk type.

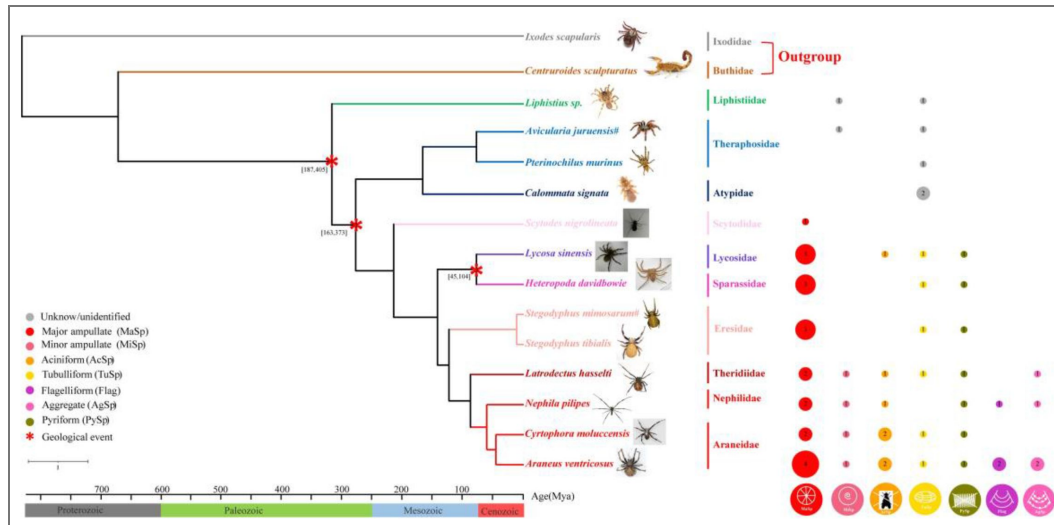
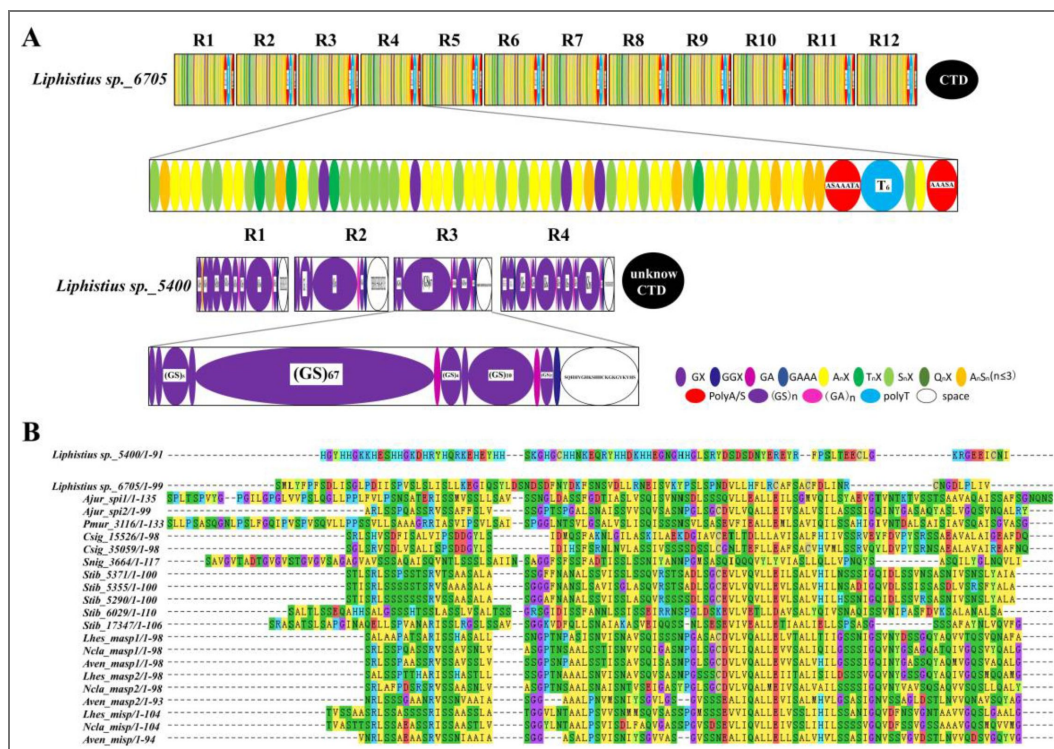


Figure 2. Analysis of silk protein structure of *Liphistius sp.*

(A) Structural model analysis of the silk protein of *Liphistius sp.* from Thailand. (B) Alignment of amino acids in the C-terminal domain of spider silk protein sequences in multiple spider species, red arrows indicate the location of two ancient spider silk proteins. Ajur, *A. juruensis*; Pmur, *P. murinus*; Csig, *C. signata*; Snig, *S. nigrolineata*; Stib, *S. tibialis*; Lhes, *L. hesperus*; Ncla, *N. clavipes*; Aven, *A. ventricosus*.



genera *Heptathela* and *Ryuthela* (Xin et al., 2015) clustered with cribellar spidroin (CrSp), and subsequently with Flag (Figure 3A), indicating that spidroin differentiation may have occurred early in the earliest diverged spider lineage, Mesothelae. N-terminals from two Mygalomorphae species clustered with the two eggcase spidroins (AcSp and TuSp) and were parallel to the ampullate spidroin clade (Figure 3A). The N-terminal of a basal Haplogynae species within Araneomorphae is positioned as a sister lineage to the ampullate spidroin clade (Figure 3A). In the C-terminal phylogeny, *Liphistius* spidroins cluster with AgSp to form the basal clade (Figure 3B), further supporting the basal status of AgSp. In contrast, spidroins from Mygalomorphae are distributed across multiple clades, including TuSp, Flag, and the ampullate spidroins (Figure 3B).

Together, these phylogenetic trees indicate that spidroin terminal domains from *Liphistius* represent the most ancestral forms, with subsequent differentiation accompanying Mesothelae diversification. The emergence of Mygalomorphae coincides with substantial expansion and diversification of spidroins, while ampullate spidroins exhibit the greatest divergence in terminal domains. In contrast, phylogenetic analysis of repetitive units reveals less diversification in Mesothelae and Mygalomorphae compared with terminal regions (Figure 3), suggesting that innovation in repetitive sequences lagged behind the evolution of silk gland types. The repetitive unit phylogeny identifies a basal clade containing AS-type spidroins from Mesothelae together with multiple spidroins from Mygalomorphae (Figure 3). PySp and subsequently TuSp diverge from this group, forming basal clades relative to other spidroins (Figure 3). In contrast, the GS-type spidroin from *Liphistius* clusters with a spidroin from a Mygalomorphae species within the derived MiSp clade (Figure 4), implying that MiSp-like motif represents the basal form of ampullate spidroins. The repetitive unit tree further shows that MiSp clusters tightly with MaSp1, suggesting that MaSp1 may represent the ancestral copy within the MaSp family (Figure 4). Analyses of amino acid composition within repetitive motifs further clarify the divergence of the two major spidroin groups derived from the two types of Mesothelae spidroins, i.e., the AS-rich and GA-rich groups, respectively (Figure 4). The AS-rich group includes undifferentiated basal spidroins as well as the derived TuSp and PySp, whereas the GA-rich group comprises ampullate spidroins, particularly MiSp and MaSp1, together with the GS-type spidroin identified in *Liphistius*.

Fairly conservation between the AS-rich ancestral spidroins and the egg-case TuSp

The repetitive units of AS-rich spidroins in Mesothelae and Mygalomorphae spiders were highly conserved, characterized by polyT and polyA motifs interspersed with other amino acids, as well as similar composition of additional motifs, such as AnX, SnX, and AnSn (where “X” represents any amino acid and “n” represents the number 1, 2, or 3) (Figure 5A). These features show substantial similarity with TuSp (Figure 5B), as all TuSp sequences also preserve polyT motifs, and one TuSp in RTA spiders additionally retains a polyA motif. TuSp proteins also show strong similarity to ancestral AS-rich spidroins in the composition of other repetitive motifs (Figure 5B). In contrast, polyA and polyT motifs are partially lost in the PySp of RTA spiders, which also feature novel spacer motifs in later-diverging spiders (Figure 5B), indicating greater divergence of PySp relative to TuSp.

The flagelliform-like spidroin in a basal RTA spider possesses MaSp-like terminal domains

Flagelliform spidroin (Flag) is a key component of prey-capture silk in typical orb webs and is commonly found in araneoid and cribellate orb-weaving spiders (Correa-Garhwal et al., 2022). Notably, we identified one Flag-like spidroin (*H.dav_6495*) in a basal RTA spider *Heteropoda davidbowie* (Wheeler et al., 2016) with highly homology to the Flag of two araneoid orb-web spiders, exhibiting substantial “GPGGX” and “GPG” motifs, which are typical features of Flag (Figure 6A). Transcriptome data confirmed its expression (figure supplements 5). This finding extends the existence of Flag to the non-web-weaving RTA clade. *Heteropoda* spiders are relatively close to cribellate orb-web spiders compared to other later-derived RTA spiders (Wheeler et al.,

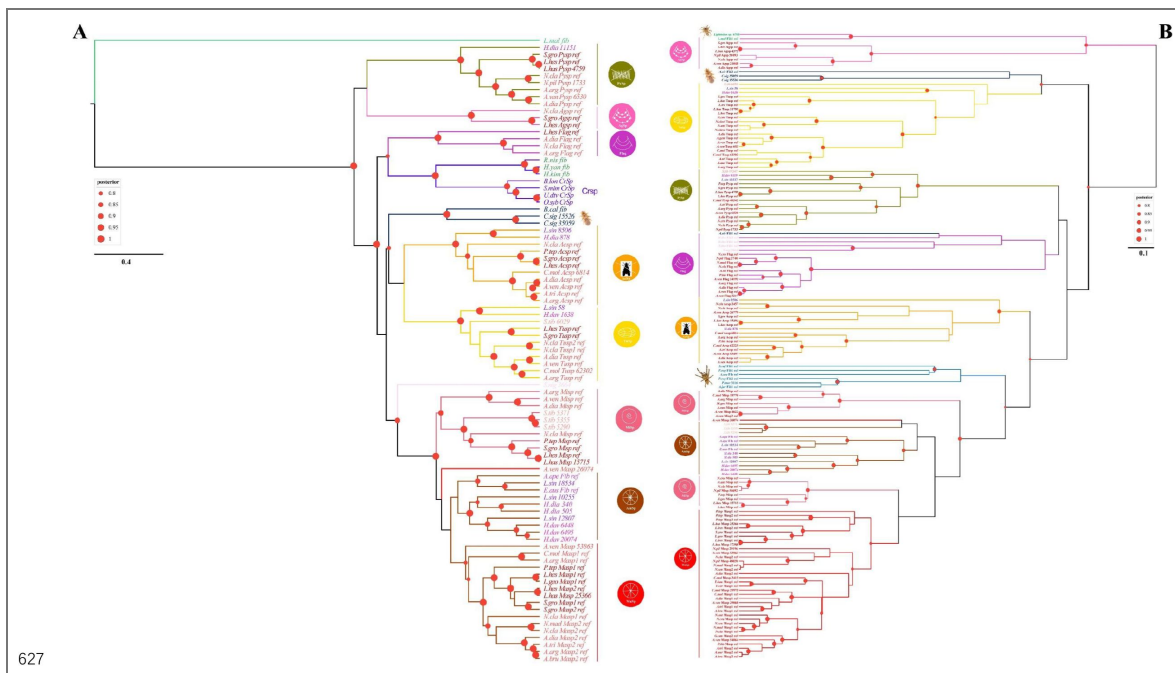


Figure 3. Bayesian Inference tree of spidroin amino-terminal regions.

The colored icons represent different spidroin type, and the specific colour is used for each type. Colored icons surround spidroins categorized as flagelliform spidroin (Flag); aggregate spidroin (AgSp); pyriform spidroin (PySp); tubuliform spidroin (TuSp); aciniform spidroin (AcSp); major ampullate spidroin (MaSp); or minor ampullate spidroin (MiSp). At the end of the gene tree are various spider spidroin, which are represented by different colors and can be roughly divided into three categories: the green line represents Mesothelae, the blue line represents Mygalomorphae, the red and purple line represents Araneomorphae. Red dots indicate nodes with bootstrap values > 80%, Scale bar is substitutions per site. The primitive group Mesothelae and Mygalomorphae sequenced in this study are marked with photos. **(A)** Bayesian tree of spidroins N-terminal regions. **(B)** Bayesian tree of spidroins C-terminal regions.

Figure 4. Bayesian Inference tree of spidroin repeat unit regions.

The colored icons represent different spidroin type, and the specific colour is used for each type. Colored icons surround spidroins categorized as flagelliform spidroin (Flag); aggregate spidroin (AgSp); pyriform spidroin (PySp); tubuliform spidroin (TuSp); aciniform spidroin (AcSp); major ampullate spidroin (MaSp); or minor ampullate spidroin (MiSp). At the end of the gene tree are various spider spidroin, which are represented by different colors and can be roughly divided into three categories: the green line represents Mesothelae, the blue line represents Mygalomorphae, the red and purple line represents Araneomorphae. Red dots indicate nodes with bootstrap values > 80%, Scale bar is substitutions per site. The primitive group Mesothelae and Mygalomorphae sequenced in this study are marked with photos. Calculating the amino acid frequency from the repeating domain in spidroins and displaying it in a heatmap, the size of the circle and the depth of the color represent the frequency.

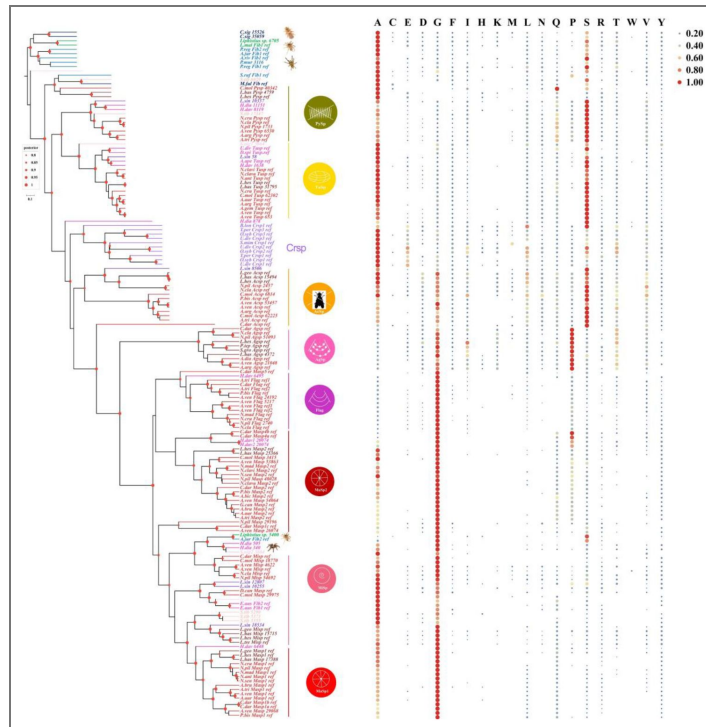
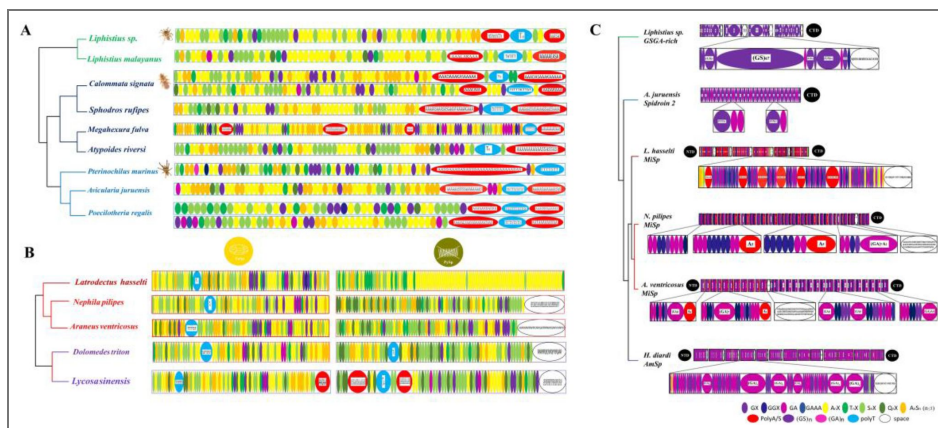


Figure 5. Evolutionary patterns of AS-rich and GSGA-rich ancient spidroin.

Amino acid sequence pattern structure of spider silk is depicted, with different colored circles representing different mechanical properties of motif, the specific type motif has been annotated in the lower left corner of the figure. (A) The AS-rich spider silk sequence structure of Mesothelae and Mygalomorphae, with photos marking the primitive groups Mesothelae and Mygalomorphae sequenced in this study. (B) Tubuliform silk protein and pyriform silk protein patterns in Araneomorphae. (C) The pattern structure of GSGA-rich spidroin.



2016). Consistent with this, we did not identify Flag transcripts in the more derived RTA jumping spider *Hyllus diardi* (figure supplements 6). Genomic synteny analyses further indicate that Flag is absent from the araneoid outgroup *Stegodyphus dumicola* (Eresidae) as well as from the derived RTA species *Dolomedes plantarius* (Figure 6B). These results suggest that Flag may have originated independently in araneoid and cribellate orb-weaving lineages, arguing against a single origin of orb webs (Garb et al., 2006), and that Flag was subsequently lost during the evolution of RTA spiders.

Importantly, both the N- and C-terminals of this Flag-like spidroin (*H.dav_6495*) are highly homologous to those of ampullate spidroins rather than typical Flag (Figure 3 and Figure 6C, D). The N- and C-terminals function in maintaining high conservation of soluble spidroins in the corresponding silk glands and in the rapid assembly of spidroins (Andersson et al., 2014; Eisoldt et al., 2012). Specifically, the C-terminal of ampullate spidroins differs from that of Flag in that it undergoes structural transitions that contribute to greater toughness (Rat et al., 2023). Therefore, this switch of terminals from Flag-type to MaSp-type in the basal RTA spider implies that the Flag-like spidroin may have evolved to be adaptive to the microenvironment of the ampullate silk gland, adopting a MaSp-like assembly mechanism.

Discussion

Supported by long-read transcript sequences combined with published data, our study provides a novel opportunity to decipher the evolutionary scenario of spider silk protein (spidroin) genes. Notably, in basal spider lineages, the N- and C-terminals of undifferentiated spidroins show greater diversification than the repetitive units (Figure 3 and 4). Both N- and C-terminals are important for silk storage and assembly, responding to the microenvironment of the silk gland (Andersson et al., 2014). Therefore, the phylogenetic relationships of spidroin terminal sequences may reflect the evolution of the corresponding silk gland microenvironments. The diverse phylogenetic positions of terminal sequences in these undifferentiated spidroins suggest that differentiation of silk gland microenvironments may have occurred earlier than genetic innovations in repetitive sequences. Supporting this idea, Mygalomorphae spiders possess morphologically complex silk-spinning apparatuses (Ferretti et al., 2017). Future in-depth molecular and genomic studies of silk glands in basal spiders will provide additional insight into these evolutionary dynamics.

In the basal-most lineage, i.e. *Liphistius sp* of Mesothelae (Wheeler et al., 2016; Xin et al., 2015), we discovered two genes with completely distinct characteristics: the AS-type spidroin and the GS-type spidroin-like gene. The coexistence of these two genes highlights the different evolutionary routes of dragline silk proteins (ampullate spidroins) and AS-rich spidroins (Fan et al., 2023). Notably, in the subsequently diverged basal lineage of Mygalomorphae, AS-type spidroins were well conserved, whereas the ancestral GS-type protein evolved into a typical spidroin. This divergent evolution may have occurred prior to the split between Mygalomorphae and Mesothelae.

The AS-type spidroins, rich in alanine and serine with mosaic poly-threonine and poly-alanine motifs, likely represent the earliest spidroins. This preliminary form of AS-rich spidroins was already used for silk production at the emergence of spiders. Sequence analysis suggests that the AS-rich TuSp is highly conserved with the AS-type spidroin in *Liphistius sp*, extending the proposed conservation of TuSp to the most basal spider lineages (Garb et al., 2007). Broad conservation of TuSp has also been reported in *Tengella perfuga*, a cribellate spider from the RTA clade (Correa-Garhwal et al., 2018), further supporting its evolutionary conservation. TuSp is a pivotal protein in spider silk evolution, constituting protective silk that forms the outer layer of the spider egg sac, which is crucial for offspring survival. Mesothelae and Mygalomorphae spiders also produce eggcase silk (Palmer et al., 2010). Strong selective constraints on this AS-rich spidroin may contribute to its functional stability.

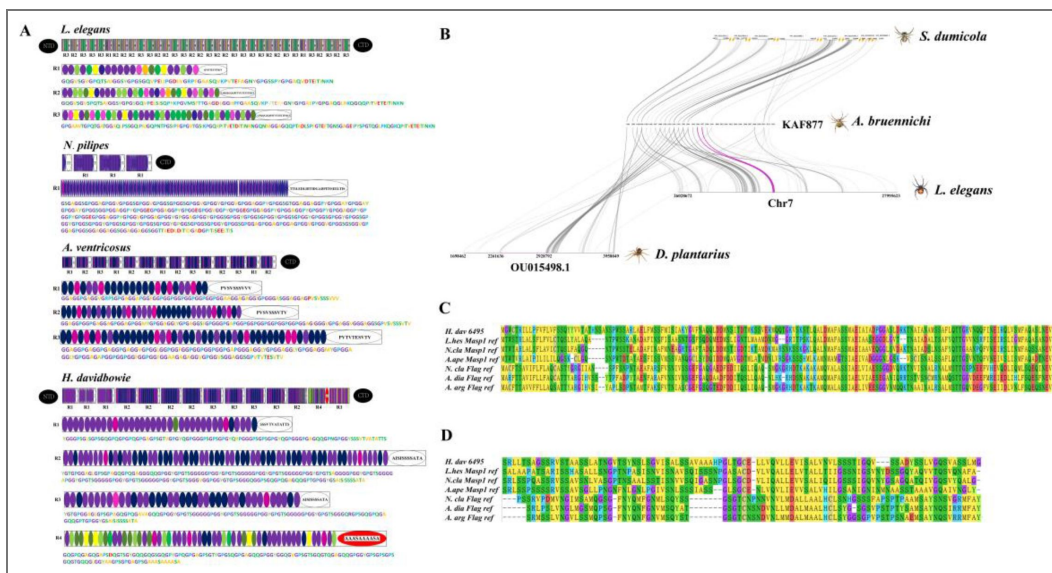


Figure 6. Analysis of important capture silk protein Flag sequence features.

(A) Schematic diagrams of cob weaver (including repeat motif from *L. elegans*), two orb weavers (including repeat motif from *N. pilipes* and *A. ventricosus*) and RTA specialized clade spider (*H. davidbowie*). Each repeat is composed of different motifs which are shown as oval module labeled with different colors. Spacer, NTD, N-terminal domain. CTD, C-terminal domain. (B) The collinearity of the upstream and downstream of *A. bruennichi* Flag gene in the ancient species of Araneomorphae (*S. dumicola*), the cobweb spider (*L. elegans*) and RTA specialized clade spider (*D. plantarius*) genome, the gray lines represent the genomic regions of other species compared to the Flag gene, the purple line represents the alignment position of the Flag gene. (C) N-terminal sequence alignment of MaSp1, Flag from 3 spiders of Araneoidea and *H. dav* spiders. (D) C-terminal sequence alignment of MaSp1, Flag from 3 spiders of Araneoidea and *H. dav* spiders.

The polyGS-bearing protein (GS-type) is suggested to be the prototype of GA-rich ampullate spidroins. It appears not to have been initially used for silk production, given its non-homologous N- and C-terminal sequences relative to canonical spidroins (Figure 2). In the subsequently diverged Mygalomorphae lineages, a novel spidroin evolved from this genetic material through adoption of canonical spidroin terminal sequences. We also demonstrate that the spidroin in the mygalomorph spider *Avicularia juruensis* is MiSp-like instead of MaSp2-like (Bittencourt et al., 2010). Fibers composed of these polyGS-bearing proteins are relatively weak in tensile strength (Collin et al., 2009). Along spider evolution, increases in GA motifs and further recruitment of polyA motifs in these silk proteins enhanced mechanical properties, as observed in the evolved MiSp of web-weaving and RTA spiders [3, 5]. Phylogenies of spidroin terminal sequences indicate a derived position of MiSp and MaSp, with MaSp appearing to have differentiated later than MiSp (Figure 3). Combining this with repetitive sequence analyses, we propose that MaSp (the dragline spidroin) originated from MiSp, potentially before the emergence of the basal Araneomorphae lampshade spider *Hypochilus thorelli*, which possesses MaSp (Starrett et al., 2012).

GS-rich and derived GA-rich silk proteins are also widely distributed in silk-producing insects. For instance, in the basal insect lineage Embioptera (webspinners), polyGS motifs are used in silk proteins (Collin et al., 2011). Later-diverged Lepidoptera caterpillars also spin MiSp- or MaSp-like silks in early larval stages (figure supplements 7), and their silk glands show similarities to spider ampullate silk glands (Hu et al., 2023; Kono et al., 2019). Although more evidence is needed to clarify potential convergent or parallel evolution between spider ampullate spidroins and Lepidopteran silk fibroins, GS-rich genetic material appears to be evolutionarily favorable for the formation of modern silk proteins.

Interestingly, we identified a novel full-length spidroin gene in the basal Araneomorphae species *Scytodes nigrolineata* (*S. nig_3664*) (figure supplements 8A). This highly expressed spidroin contains six repetitive units, each comprising two tightly connected motifs: one AS-rich and one GA-rich, representing the two types of spidroin composition (Figure 4—figure supplements 8). The AS-rich motif is highly similar to TuSp, whereas the GA-rich motif resembles MiSp (figure supplements 8). This finding highlights vigorous rearrangement of spidroins during this evolutionary period and underscores the complexity of silk protein evolution. Further genomic evidence will help clarify the nature of these rearrangements. The same confusion may be the case of the Flagelliform spidroin (Flag). Its origin could not be traced through sequence analysis, but its evolutionary scale was extended to the non-web-weaving RTA lineage. The origin of the RTA clade from the ancestor of Deinopoidea orb-web spiders (Arakawa et al., 2022; Yu et al., 2022) coincided with degradation of web-weaving capacity. As a typical web component, flagelliform silk functions as capture thread. Flag is a high-cost spidroin and thus poses a selective burden during adaptive evolution of RTA spiders. In contrast, ampullate silks and spidroins are universally essential as dragline and lifeline silk for spiders and even some caterpillars. By examining the Flag-like spidroin in basal RTA spiders, we observed that its terminals are possibly adaptive to the ampullate silk gland. Given that later RTA spiders lost Flag, we propose that it may have initially been incorporated into ampullate silks in early RTA spiders, followed by progressive degradation and eventual loss of the corresponding gene in derived RTA species.

Materials and methods

Spider samples

Through field investigations, we initially collected eight species of spiders based on morphological characteristics in four geographically distinct regions in China (Sichuan, Yunnan, Xinjiang, Guangdong). These species include: *Calommata signata* Karsch (Mianyang, Sichuan: 31.30°N, 104.42°E), *Scytodes nigrolineata* (Liangshan, Sichuan: 27.89°N, 102.26°E), *Stegodyphus tibialis* (Dali, Yunnan: 25.34°N, 100.13°E), *Latrodectus hasselti* (Liangshan, Sichuan: 27.89°N, 102.26°E), *Lycosa sinensis* (Yili, Xinjiang: 43.55°N, 81.20°E), *Hyllus diardi* (*Xishuangbanna*, Yunnan: 22.01°N, 100.80°E), *Nephila pilipes* (Luoping, Yunnan: 24.884626°N, 104.31°E), and *Cyrtophora moluccensis* (Liannan,

Guangdong: 24.77°N, 112.28°E). Additionally, through donations from spider enthusiasts, we collected three species: *Liphistius sp.*, *Pterinochilus murinus*, *Heteropoda davidbowie*. All spider samples were collected and photographed, as shown in [Figure 1](#). Spiders used for high-throughput sequencing were starved for 2 days to prevent exogenous contamination and then stored at -80°C after rapid freezing in liquid nitrogen.

RNA extraction, long read transcriptome sequencing (ISO-Seq) and Illumine RNA-seq

Spider samples were sent to Novogene Co, Ltd for RNA extraction and transcriptome sequencing. Briefly, the whole body of spider samples was ground in liquid nitrogen, and total RNA was extracted using TRIzol reagent (Invitrogen) on dry ice. RNA concentration was accurately quantified using Qubit (Invitrogen Life Technologies, USA), while the integrity of the RNA was detected using an Agilent 2100 Bioanalyzer (Agilent Technology, USA). For ISO-Seq, libraries were prepared using the PacBio Iso-Seq library preparation method, followed by sequencing on the Pacific Biosciences Sequel II platform at the University of Maryland Institute for Genome Sciences. For RNA-Seq, libraries were prepared using the Illumina RNA-seq library preparation method and sequenced on the Illumina HiSeq™ 2000 platform.

Raw data processing

Raw data were filtered and processed using SMRTlink v6.0 software with the following parameters: --minLength 50, --maxLength 15000, --minPasses 2, --min_seq_len 200. Subreads sequences were obtained, and circular consensus sequences (CCS) were generated by correcting subreads sequences. CCS containing 5'-primer, 3'-primer, and polyA were detected. CCS were classified into full-length non-chimeric (FLNC) sequences and non-full-length (nFL) sequences. FLNC sequences of the same transcript were clustered using the ICE algorithm to obtain consensus sequences, which were then corrected by nFL sequences using Arrow software. Sequences were further corrected using Illumina RNA-seq reads with LoRDEC v0.7 (Salmela & Rivals, 2014) software, followed by de-redundancy processing using CD-HIT-EST v4.6.8 (Li & Godzik, 2006) (-c 0.95). The resulting non-redundant transcripts were used for subsequent analyses. Long-read transcriptome data (SRR12208061) and Illumina RNA-seq data (SRR11905576) of *A. ventricosus* were downloaded from the NCBI SRA database and analyzed using the same pipeline. The average length of sequences and contig N50 size were calculated (figure supplements 1 and 2).

Annotation of genes

The genes were functionally annotated by homology-based searches against the following database: NCBI Non-redundant Protein (NR), NCBI Non-redundant Nucleotide (Nt), SwissProt, Kyoto Encyclopedia of Genes and Genomes (KEGG), euKaryotic Ortholog Groups (KOG) and Gene Ontology (GO).

Construction of pseudo-coding genome

Non-redundant transcripts were assembled to identify full-length genes (CDSgenome) using Cogent v3.2 software. Each transcript family was folded into one or several unique transcript models (UniTransModels), and redundant isoforms were further collapsed using minimap2 (Li, 2018) alignment and cDNA_Cupcake v5.3. The constructed pseudogenome was analyzed with ANGEL v2.4 (Shimizu et al., 2006) software (--min_aa_length 100) for Open Reading Frame (ORF) prediction analysis, and BUSCO v5.0 (Simão et al., 2015) software was used to evaluate data integrity after processing.

Orthologous gene identification

Genome information files of the spider *Stegodyphus mimosarum* (GCA_000611955.2) and two outgroup species *Ixodes scapularis* (GCA_016920785.2) and *Centruroides sculpturatus* (GCA_000671375.2) were developed. Single copy orthologous proteins were identified by reciprocal blast (Mount, 2007) (evalue cutoff: 10⁻⁵), with the gene set of *S. mimosarum* as a reference,

against the annotated gene set of each of the 14 species. All single-copy orthologs of each species were aligned using Mafft v7.407 (Kato et al., 2002) software and concatenated into one super gene for each species.

Phylogeny reconstruction and divergence time estimation

The well-aligned single-copy orthologs were subjected to Maximum likelihood-based phylogenetic analysis using RAxML v8.2.12 (Stamatakis et al., 2005) with the following parameters: -f a -T 12 -p 12345 -x 12345 -m PROTGAMMAAUTO -# 1000. Divergence time was calculated using the MCMCtree program in the PAML v4.8 (Yang, 2007) package, and fossil records obtained from the TimeTree database (Kumar et al., 2022) were used for result calibration.

Identification and preliminary analyses of spidroins

Released spidroin sequences in the NCBI NR database were collected, and representative complete spider silk protein sequences were used as reference sequences for blastp (Mount, 2007) (e-value cutoff: 10⁻⁵) against the protein set of each of the 12 spider species. Fitted sequences were manually checked for a typical spider silk protein characteristics, shorter sequences or missense mutations. Error-free long spider silk sequences were verified by blastp against the NR database. Identified spidroin sequences were analyzed for signal peptide prediction using SignalP v5.0 and hydrophobicity prediction using ProtScale and ProtParam tools in the Expasy platform (<https://www.expasy.org/>). Selected error-free spidroin sequences were divided into three regions: N-terminal, repeat region, and C-terminal. Amino acid frequency of repetitive sequences was calculated and a heatmap was generated using Tbtools v2.056 (Chen et al., 2020) software.

Phylogenetic analysis of spidroins

The lack of homology of the terminals Lera_5400 to any other spidroin supports the ancient status of Liphistius spidroins. We excluded these sequences and investigated the phylogenetic relationship of amino acid sequences of the rest spidroins N-terminal, C-terminal sequences, and repetitive motifs, respectively, with some publicly available sequences added in (figure supplements 5). The N-terminal, repeat region, and C-terminal of spidroins were aligned separately using MEGA v7 (Kumar et al., 2018). Aligned sequences of each region were utilized to construct gene trees using beast v1.10 (Suchard et al., 2018) software with the Bayesian method and the following parameter settings: chain length of 20,000,000, logEvery of 1000, and burnin of 0.1.

Calculation and visualization of spider gene expression

The Minimap2 (Li, 2018) software was used to compare the redundant transcript data of the full-length transcriptome to the pseudo-coding genome constructed by Cogent v3.2 software, the bowtie2 (Langmead & Salzberg, 2012) software was used to compare the raw data of second-generation transcriptome sequencing to the pseudo-coding genome. Use the Samtools v0.1.19 (Li et al., 2009) tool to convert the sam format generated after the comparison into a bam format file, the bam files are sorted and indexed, and then the bam files are imported into the IGV software for visual analysis.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All raw data was deposited at the NCBI in the sequence read archive (SRA) under accession number (BioProject Number: PRJNA754522).

Acknowledgements

We thank Prof. Wen Wang for help in revised the manuscript and give suggestions for this work. We thank Prof. Sheng Li for insightful comments. We thank Ms. Jiajia Zhang for her help in handling the spiders.

Additional information

CRedit authorship contribution statement

H. X. conceived and designed the investigation. K.S.Z., G.L., M.L and J.H.X. collected and prepared the samples. K.S. Z. analyzed the data. K.S. Z. and H. X. wrote the paper. A.J. T., H. X, and Y.P. H. revised the manuscript. All the authors read and approved the final manuscript.

Funding

This work was supported from a grand of the National Natural Science Foundation of China (Grant No. 32070411, 32270458), the Laboratory of Lingnan Modern Agriculture Project (NZ2021019), and a grant from National Natural Science Foundation of China (2023A1515010657).

Funding

Funder	Grant reference number	Author
MOST National Natural Science Foundation of China (NSFC)	32070411	Hui Xiang
MOST National Natural Science Foundation of China (NSFC)	32270458	Hui Xiang
MOST National Natural Science Foundation of China (NSFC)	2023A1515010657	Hui Xiang
the Laboratory of Lingnan Modern Agriculture Project	NZ2021019	Hui Xiang

Author ORCID iDs

Lei Gao: <https://orcid.org/0000-0001-5708-029X>

Additional files

[Supplementary file 1.](#) 

References

- Andersson M., Chen G., Otkovs M., Landreh M., Rising A. (2014) Carbonic Anhydrase Generates CO₂ and H⁺ That Drive Spider Silk Formation Via Opposite Effects on the Terminal Domains. *Plos Biology* **12**:e1001921 <https://doi.org/10.1371/journal.pbio.1001921>
- Arakawa K., Kono N., Malay A. D., Tateishi A., Ifuku N., Masunaga H., Sato R., Tsuchiya K., Ohtoshi R., Pedrazzoli D., et al. (2022) 1000 spider silkomes: Linking sequences to silk physical properties. *Sci Adv* **8**:eabo6043 <https://doi.org/10.1126/sciadv.abo6043>
- Bittencourt D., Dittmar K., Lewis R. V., Rech E. L. (2010) A MaSp2-like gene found in the Amazon mygalomorph spider *Avicularia juruensis*. *Comp Biochem Physiol B Biochem Mol Biol* **155**:419-426 <https://doi.org/10.1016/j.cbpb.2010.01.005>
- Blamires S. J., Blackledge T. A., Tso I. M. (2017) Physicochemical Property Variation in Spider Silk: Ecology, Evolution, and Synthetic Production. *Annu Rev Entomol* **62**:443-460 <https://doi.org/10.1146/annurev-ento-031616-035615>
- Chaw R. C., Sasaki C. A., Hayashi C. Y. (2017) Complete gene sequence of spider attachment silk protein (PySp1) reveals novel linker regions and extreme repeat homogenization. *Insect Biochemistry and Molecular Biology* **81**:80-90 <https://doi.org/10.1016/j.ibmb.2017.01.002>
- Chen C., Chen H., Zhang Y., Thomas H. R., Frank M. H., He Y., Xia R. (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular plant* **13**:1194-1202 <https://doi.org/10.1016/j.molp.2020.06.009>

- Chen G., Liu X., Zhang Y., Lin S., Yang Z., Johansson J., Rising A., Meng Q. (2012) Full-length minor ampullate spidroin gene sequence. *PLoS one* **7**:e52293 <https://doi.org/10.1371/journal.pone.0052293>
- Collin M. A., Edgerly J. S., Hayashi C. Y. (2011) Comparison of fibroin cDNAs from web-spinning insects: insight into silk formation and function. *Zoology (Jena)* **114**:239-246 <https://doi.org/10.1016/j.zool.2011.01.004>
- Collin M. A., Garb J. E., Edgerly J. S., Hayashi C. Y. (2009) Characterization of silk spun by the embiopteran, *Antipaluria urichi*. *Insect Biochem Mol Biol* **39**:75-82 <https://doi.org/10.1016/j.ibmb.2008.10.004>
- Correa-Garhwal S. M., Baker R. H., Clarke T. H., Ayoub N. A., Hayashi C. Y. (2022) The evolutionary history of cribellate orb-weaver capture thread spidroins. *BMC Ecology and Evolution* **22**:1-16 <https://doi.org/10.1186/s12862-022-02042-5>
- Correa-Garhwal S. M., Chaw R. C., Clarke T. H., Alaniz L. G., Chan F. S., Alfaro R. E., Hayashi C. Y., Heneberg P. (2018) Silk genes and silk gene expression in the spider *Tengella perfuga* (Zoropsidae), including a potential cribellar spidroin (CrSp). *Plos One* **13** <https://doi.org/10.1371/journal.pone.0203563>
- Craig C. L. (2004) Broad patterns of speciation are correlated with the evolution of new silk proteins in spiders but not in the lepidoptera. *Biomacromolecules* **5**:739-743 <https://doi.org/10.1021/bm034351s>
- Dugger T. W., Sarkar S., Correa-Garhwal S. M., Zhernenkov M., Zhang Y. G., Kolhatkar G., Mohan R., Cruz L., Lubio A. D., Ruediger A., et al. (2020) Ultrastructures and Mechanics of Annealed Major Ampullate Silk. *Biomacromolecules* **21**:1186-1194 <https://doi.org/10.1021/acs.biomac.9b01615>
- Eisoldt L., Thamm C., Scheibel T. (2012) The role of terminal domains during storage and assembly of spider silk proteins. *Biopolymers* **97**:355-361 <https://doi.org/10.1002/bip.22006>
- Fan Z., Wang L. Y., Xiao L., Tan B., Luo B., Ren T. Y., Liu N., Zhang Z. S., Bai M. (2023) Lampshade web spider *Ectatosticta davidi* chromosome-level genome assembly provides evidence for its phylogenetic position. *Commun Biol* **6**:748 <https://doi.org/10.1038/s42003-023-05129-x>
- Fernandez R., Kallal R. J., Dimitrov D., Ballesteros J. A., Arnedo M. A., Giribet G., Hormiga G. (2018) Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. *Curr Biol* **28**:2190-2193 <https://doi.org/10.1016/j.cub.2018.06.018>
- Ferretti N., Pompozzi G., Copperi S., Wehitt A., Galindez E., Gonzalez A., Perez-Miles F. (2017) A comparative morphological study of the epiandrous apparatus in mygalomorph spiders (Araneae, Mygalomorphae). *Micron* **93**:9-19 <https://doi.org/10.1016/j.micron.2016.11.005>
- Garb J. E., DiMauro T., Vo V., Hayashi C. Y. (2006) Silk Genes Support the Single Origin of Orb Webs. *Science* <https://doi.org/10.1126/science.1127946>
- Garb J. E., Ayoub N. A., Hayashi C. Y. (2010) Untangling spider silk evolution with spidroin terminal domains. *BMC Evol Biol* **10**:243 <https://doi.org/10.1186/1471-2148-10-243>
- Garb J. E., DiMauro T., Lewis R. V., Hayashi C. Y. (2007) Expansion and intragenic homogenization of spider silk genes since the Triassic: evidence from Mygalomorphae (tarantulas and their kin) spidroins. *Mol Biol Evol* **24**:2454-2464 <https://doi.org/10.1093/molbev/msm179>
- Garb J. E., DiMauro T., Vo V., Hayashi C. Y. (2006) Silk genes support the single origin of orb webs. *Science* **312**:1762 <https://doi.org/10.1126/science.1127946>
- Garb J. E., Hayashi C. Y. (2005) Modular evolution of egg case silk genes across orb-weaving spider superfamilies. *Proceedings of the National Academy of Sciences of the United States of America* **102**:11379-11384 <https://doi.org/10.1073/pnas.0502473102>
- Garrison N. L., Rodriguez J., Agnarsson I., Coddington J. A., Bond J. E. (2015) Spider phylogenomics: untangling the Spider Tree of Life. *PeerJ* <https://doi.org/10.7717/peerj.1719>

- 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 <https://doi.org/10.1093/molbev/msh204>
- Hu W., Jia A., Ma S., Zhang G., Wei Z., Lu F., Luo Y., Zhang Z., Sun J., Yang T., *et al.* (2023) A molecular atlas reveals the tri-sectional spinning mechanism of spider dragline silk. *Nat Commun* **14**:837 <https://doi.org/10.1038/s41467-023-36545-6>
- Katoh K., Misawa K., Kuma K. i., Miyata T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**:3059-3066 <https://doi.org/10.1093/nar/gkf436>
- Kono N., Nakamura H., Ohtoshi R., Tomita M., Numata K., Arakawa K. (2019) The bagworm genome reveals a unique fibroin gene that provides high tensile strength. *Communications Biology* **2** <https://doi.org/10.1038/s42003-019-0412-8>
- Kumar S., Stecher G., Li M., Knyaz C., Tamura K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution* **35**:1547-1549 <https://doi.org/10.1093/molbev/msy096>
- Kumar S., Suleski M., Craig J. M., Kaspruwicz A. E., Sanderford M., Li M., Stecher G., Hedges S. B. (2022) TimeTree 5: an expanded resource for species divergence times. *Molecular biology and evolution* **39**:msac174 <https://doi.org/10.1093/molbev/msac174>
- Langmead B., Salzberg S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**:357-359 <https://doi.org/10.1038/nmeth.1923>
- Li H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094-3100 <https://doi.org/10.1093/bioinformatics/bty191>
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., Subgroup G. P. D. P. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078-2079 <https://doi.org/10.1093/bioinformatics/btp352>
- Li W., Godzik A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**:1658-1659 <https://doi.org/10.1093/bioinformatics/btl158>
- Li X., Fan J. S., Shi M., Lai C. C., Li J., Meng Q., Yang D. (2022) C-Terminal Domains of Spider Silk Proteins Having Divergent Structures but Conserved Functional Roles. *Biomacromolecules* **23**:1643-1651 <https://doi.org/10.1021/acs.biomac.1c01513>
- Mita K., James I. T. C. (1994) Highly repetitive structure and its organization of the silk fibroin gene. *Journal of Molecular Evolution* <https://doi.org/10.1007/BF00175878>
- Mount D. W. (2007) Using the basic local alignment search tool (BLAST). *Cold spring harbor Protocols* **2007** <https://doi.org/10.1101/pdb.top17>
- Nakamura H., Kono N., Mori M., Masunaga H., Numata K., Arakawa K. (2023) Composition of Minor Ampullate Silk Makes Its Properties Different from Those of Major Ampullate Silk. *Biomacromolecules* **24**:2042-2051 <https://doi.org/10.1021/acs.biomac.2c01474>
- Palmer J. M., Coyle F. A., Harrison F. W. (2010) Structure and cytochemistry of the silk glands of the mygalomorph spider *Antrodiaetus unicolor* (araneae, antrodiaetidae). *Journal of Morphology* **174** <https://doi.org/10.1002/jmor.1051740303>
- Papadopoulos P., Ene R., Weidner I., Kremer F. (2009) Similarities in the structural organization of major and minor ampullate spider silk. *Macromol Rapid Commun* **30**:851-857 <https://doi.org/10.1002/marc.200900018>
- Rat C., Heindl C., Neuweiler H. (2023) Domain swap facilitates structural transitions of spider silk protein C-terminal domains. *Protein Sci* **32**:e4783 <https://doi.org/10.1002/pro.4783>
- Salmela L., Rivals E. (2014) LoRDEC: accurate and efficient long read error correction. *Bioinformatics* **30**:3506-3514 <https://doi.org/10.1093/bioinformatics/btu538>

- Shimizu K., Adachi J., Muraoka Y. (2006) ANGLE: a sequencing errors resistant program for predicting protein coding regions in unfinished cDNA. *Journal of Bioinformatics and Computational Biology* **4**:649-664 <https://doi.org/10.1142/s0219720006002260>
- Simão F. A., Waterhouse R. M., Ioannidis P., Kriventseva E. V., Zdobnov E. M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210-3212 <https://doi.org/10.1093/bioinformatics/btv351>
- Stamatakis A., Ludwig T., Meier H. (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**:456-463 <https://doi.org/10.1093/bioinformatics/bti191>
- Starrett J., Garb J. E., Kuelbs A., Azubuie U. O., Hayashi C. Y. (2012) Early events in the evolution of spider silk genes. *Plos One* **7**:e38084 <https://doi.org/10.1371/journal.pone.0038084>
- Stellwagen S. D., Renberg R. L. (2019) Toward Spider Glue: Long Read Scaffolding for Extreme Length and Repetitious Silk Family Genes AgSp1 and AgSp2 with Insights into Functional Adaptation. *G3-Genes Genomes Genetics* **9**:1909-1919 <https://doi.org/10.1534/g3.119.400065>
- Suchard M. A., Lemey P., Baele G., Ayres D. L., Drummond A. J., Rambaut A. (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus evolution* **4**:vey016 <https://doi.org/10.1093/ve/vey016>
- Sutherland T. D., Young J. H., Weisman S., Hayashi C. Y., Merritt D. J. (2010) Insect silk: one name, many materials. *Annual Review of Entomology* **55**:171 <https://doi.org/10.1146/annurev-ento-112408-085401>
- Wheeler W., Coddington J., Crowley L., Zhang J. (2016) The spider tree of life: Phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling. *Cladistics* <https://doi.org/10.1111/cla.12182>
- Xin X., Liu F., Chen J., Ono H., Li D., Kuntner M. (2015) A genus-level taxonomic review of primitively segmented spiders (Mesothelae, Liphistiidae). *ZooKeys* <https://doi.org/10.3897/zookeys.488.8726>
- Yang Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**:1586-1591 <https://doi.org/10.1093/molbev/msm088>
- Yu N., Yang Z., Fan Z., Liu Z. (2022) Classification and functional characterization of spidroin genes in a wandering spider, *Pardosa pseudoannulata*. *Insect Biochem Mol Biol* **151**:e103862 <https://doi.org/10.1016/j.ibmb.2022.103862>
- Zhou S. Y., Dong Q. L., Zhu K. S., Gao L., Chen X., Xiang H. (2021) Long-read transcriptomic analysis of orb-weaving spider *Araneus ventricosus* indicates transcriptional diversity of spidroins. *Int J Biol Macromol* **168**:395-402 <https://doi.org/10.1016/j.ijbiomac.2020.11.182>
- Zhu K (2024) 11 full-length transcriptomes or second-generation transcriptomes spider data: Mesothelae, Mygalomorphae and Araneomorphae Raw sequence reads. NCBI Sequence Read Archive. ID PRJNA754522 <https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA754522>

Peer reviews

Reviewer #1 (Public review):

Summary:

In this study, Zhu et al. address spider silk spidroin evolution using long-read transcriptomics across 12 spider species. The study provides a novel evolutionary framework for spidroin diversification, proposing the existence of two ancient ancestral templates, i.e., AS and GS, and tracing how these templates diversified into major spidroin classes observed in radiated spiders. The manuscript further focused on the evolutionary history of multiple known spidroin proteins, with some previous hypotheses being revised.

Strengths:

A major challenge in silk biology, the highly repetitive content, was well addressed in this study by full-length transcriptome sequencing. Also, the authors performed very detailed analyses on sequence features across a wide range of species. I therefore think the study is supported by sound levels of sampling, technology, and analysis.

Weaknesses:

The manuscript presents a lot of detail regarding various sequence features and derived claims, but these features are sometimes not friendly to an audience not working with spider silks. Also, the current figures are not very helpful for understanding those described patterns. I found many colorful, trivial elements in almost every figure, but how their organization supported the corresponding statement was often unclear to me. I recommend that the authors further improve the figure design, including presenting a schematic evolutionary history for those spider silk proteins.

<https://doi.org/10.7554/eLife.110512.1.sa2>

Reviewer #2 (Public review):

Summary:

This paper utilizes long-read transcriptomics across 12 representative spider species to propose a new evolutionary framework for spider silk proteins (spidroins). By identifying ancestral templates in the most basal spider lineages, the authors trace how simple genetic materials diversified into the high-performance fibers used by modern spiders.

Strengths:

- (1) The authors utilized PacBio ISO-Seq (long-read transcriptomics), which is essential for resolving the massive, highly repetitive sequences of spidroin genes that often cause gaps in traditional short-read assemblies.
- (2) The researchers sampled 12 species representing the major nodes of spider evolution, including the basal Mesothelae, the Mygalomorphae (tarantulas), and the highly diverse Araneomorphae.
- (3) The study successfully identified two distinct primordial spidroins in basal spiders: the AS-type (alanine-serine-rich) and the GS-type (glycine-serine-rich) proteins.

Weaknesses:

- (1) The GS-Type "Base Gene" Paradox

The paper proposes that the GS-type gene (Liphistius sp._5400) in Liphistius (the most ancient spider lineage) is the prototype for all modern dragline silk. However, the data presented significantly undermines this conclusion.

Every functional spider silk protein requires N-terminal and C-terminal domains to control fiber assembly. The authors admit that neither the N- nor the C-terminal of this GS-type protein shows homology to any known spidroins. Because it lacks these domains, the authors explicitly state that it "may not assemble into typical silk fibers". The authors are identifying this as a "base gene" solely because it contains poly-GS motifs. Their logic is that because GS motifs are found in modern silk and other silk-producing insects, this must be the ancestor.

In the same spider, the AS-type gene (Liphistius sp._6705) does have recognizable C-terminal sequences and motifs similar to modern eggcase silk. This proves that "real" spidroins existed in Liphistius, making the claim that the non-homologous GS-type is a "spidroin ancestor" look like a misidentification of a general repetitive protein.

(2) Overstated Classification of FLAG in RTA Spiders

The authors identified a transcript in the RTA spider *Heteropoda davidbowie* (H.dav_6495) and labeled it a "Flag-like spidroin". This label is based on the repetitive internal motifs, which contain "GPGGX" and "GPG"-the classic building blocks of flagelliform capture silk. However, both the N- and C-termini of this gene are highly homologous to ampullate spidroins (MaSp), not typical Flag proteins. By calling it a "Flag-like spidroin" rather than a "MaSp with GPG motifs," the authors are forcing an evolutionary narrative. It is equally possible that this is simply a divergent Major Ampullate spidroin that evolved capture-like motifs, rather than a capture silk gene that "moved" into the ampullate gland.

The authors explicitly state, "Its origin could not be traced through sequence analysis". This admission directly contradicts the confidence with which they propose a "revised evolutionary trajectory".

Appraisal and Impact

This study provides a high-resolution map of spider silk evolution by utilizing long-read transcriptomics to bridge the gap between basal and derived lineages. By identifying the earliest known genetic templates for silk, the paper offers a significant leap forward in understanding how complex biological materials originate, though it raises critical questions about the functional definition of a "spidroin".

<https://doi.org/10.7554/eLife.110512.1.sa1>

Reviewer #3 (Public review):

Summary:

In this study, Zhu et al. use long-read transcriptomes, with correction using short-read RNA-seq, from 12 spider species that span the major evolutionary lineages to investigate the diversification of spider silk proteins (spidroins). Here, they identify 60 spidroin sequences and propose that two highly divergent sequences found in the basal *Liphistius* sp., where one is an alanine-serine-rich (AS-type), and one is a glycine-serine-rich (GS-type), represent ancestral templates from which all major spidroin families diversified. Using separate phylogenetic analyses for N-terminal domains, C-terminal domains, and repetitive domains, the authors argue that the AS-type lineage remained relatively conserved and gave rise to tubuliform spidroins (TuSp) used in eggcase silk, while the GS-type lineage evolved into minor ampullate spidroins (MiSp) and may have provided the substrate for major ampullate spidroins (MaSp). In addition, they describe a specific flagelliform-like (flag) transcript in a basal clade spider, with MaSp-like terminal domains, and propose that Flag was co-opted into ampullate silk glands before being progressively lost in more derived retrolateral tibial apophysis (RTA) lineages.

Strengths:

The taxon sampling is a strength of this study, covering representative species at key nodes across spider evolution, from the earliest-diverging Mesothelae through Mygalomorphae and into the most derived Araneomorphae lineages, which enables the authors to make comparative inferences about ancestral states. Also, the use of long-read sequencing is well-suited to the problem since spidroin genes contain highly repetitive coding sequences that would be very hard to resolve by short-read assembly alone. Thus, retrieving 30 full-length sequences in this context is notable, and the assembly quality appears reasonable for transcriptomic resources, with BUSCO completeness values reported between 85% and 93% across species.

The decision to analyse N-terminal, C-terminal, and repetitive domains in separate phylogenetic trees is methodologically sound and yields a biologically interesting result: terminal domains show greater diversification in basal lineages than repetitive regions, suggesting that specialisation of silk gland microenvironments preceded compositional innovation in the repetitive sequences.

Weaknesses:

While the paper has strengths in providing a useful comparative resource and generating interesting hypotheses, several of the central evolutionary conclusions are not directly supported by the current data. There are three main elements that require further attention:

(1) The GS-type *Liphistius* sequence (*Liphistius* sp._5400) is central to the manuscript's model for the origin of GA-rich ampullate spidroins, but the authors describe it as a spidroin-like transcript whose N- and C-terminal regions lack homology to known spidroins and may not support typical silk-fiber assembly. Since its terminal domains are excluded from the phylogenetic analyses, the proposed scenario, GS-type to MiSp to MaSp, rests largely on repeat-region similarity. Supplementary materials provided in this study further indicate no predicted signal peptide, although this feature alone is not unique among the annotated silk proteins. The manuscript should therefore either provide a stronger justification for treating *Liphistius* sp._5400 as an ancestral spidroin or more consistently frame it as a spidroin-like, repeat-based intermediate. The distinction between repeat-region clustering and full functional homology should also be made explicit.

(2) The whole-body transcriptome approach is an important sampling limitation that is acknowledged here, where the authors note that they were unable to recover complete spidroin repertoires for each species. Because the newly generated data are not silk-gland-specific, the absence of a transcript in a given species should be interpreted with caution and not equated directly with gene absence. This is particularly relevant to the manuscript's proposed loss of Flag during RTA evolution. In the focal taxa, the inference combines one positive transcript in *H. davidbowie* with non-detection in *H. diardi*, while broader support comes from limited synteny-based absence in a small number of external genomes. Therefore, while the Flag-loss scenario could be plausible, it remains suggestive rather than conclusive without more targeted silk-gland sampling or broader genomic validation.

(3) The Flag co-option model is interesting, but as presented now, it is based on limited evidence: a single Flag-like transcript in *H. davidbowie*, the absence of detection in *H. diardi*, restricted synteny comparisons, and terminal-domain similarity to ampullate spidroins. The manuscript does not present proteomic evidence that this Flag-like protein is incorporated into ampullate silk fibers, nor does it show a series of pseudogenized or truncated Flag loci across derived RTA lineages. This is a plausible and interesting scenario, but it should be framed more consistently as a testable hypothesis rather than as an established evolutionary pathway.

<https://doi.org/10.7554/eLife.110512.1.sa0>