

A molecular atlas reveals the tri-sectional spinning mechanism of spider dragline silk

Received: 19 July 2022

Accepted: 7 February 2023

Published online: 15 February 2023

 Check for updates

Wenbo Hu^{1,3}, Anqiang Jia^{1,3}, Sanyuan Ma¹, Guoqing Zhang¹, Zhaoyuan Wei¹, Fang Lu¹, Yongjiang Luo¹, Zhisheng Zhang², Jiahe Sun¹, Tianfang Yang¹, TingTing Xia¹, Qinhui Li¹, Ting Yao¹, Jiangyu Zheng¹, Zijie Jiang¹, Zehui Xu¹, Qingyou Xia¹✉ & Yi Wang¹✉

The process of natural silk production in the spider major ampullate (Ma) gland endows dragline silk with extraordinary mechanical properties and the potential for biomimetic applications. However, the precise genetic roles of the Ma gland during this process remain unknown. Here, we performed a systematic molecular atlas of dragline silk production through a high-quality genome assembly for the golden orb-weaving spider *Trichonephila clavata* and a multiomics approach to defining the Ma gland tri-sectional architecture: Tail, Sac, and Duct. We uncovered a hierarchical biosynthesis of spidroins, organic acids, lipids, and chitin in the sectionalized Ma gland dedicated to fine silk constitution. The ordered secretion of spidroins was achieved by the synergetic regulation of epigenetic and ceRNA signatures for genomic group-distributed spidroin genes. Single-cellular and spatial RNA profiling identified ten cell types with partitioned functional division determining the tri-sectional organization of the Ma gland. Convergence analysis and genetic manipulation further validated that this tri-sectional architecture of the silk gland was analogous across Arthropoda and inextricably linked with silk formation. Collectively, our study provides multidimensional data that significantly expand the knowledge of spider dragline silk generation and ultimately benefit innovation in spider-inspired fibers.

Spiders (Order Araneae) are abundant generalist arthropod predators including more than fifty thousand extant species^{1,2}. All spiders produce silks, which are natural high-performance proteinaceous fibers that are crucial for spider survival and reproduction^{3,4}. Silk production is a fascinating spider-salient trait of particular economic interest, mainly due to the exceptional properties of these fibers, including high tensile strength and toughness, low density, self-powered rotational actuation, and biocompatibility^{5–7}. Numerous researchers have tried to emulate natural spidroin (the main proteins of spider silk) production and spinning processes for the biomimetic generation of artificial materials with spider silk-like properties^{8–11}; however, much of our

current understanding of spider silk formation is based on physical and material studies that have provided only a partial picture of its nature^{12,13}. Thus, the elucidation of the molecular biological mechanisms involved in the natural silk production system will be valuable for gaining an in-depth understanding of spider silk^{14–17}.

Although orb-web spiders have multiple silk-producing glands, the major ampullate (Ma) gland is often used as a model system in silk production research due to its relatively large size and, especially, the impressive properties of its product, dragline silk¹⁸. Accordingly, most attempts at innovating dragline silk-inspired fibers have generally involved the silk proteins and microenvironment produced by the Ma

¹State Key Laboratory of Silkworm Genome Biology, Biological Science Research Center, Southwest University, Chongqing 400715, China. ²School of Life Sciences, Southwest University, Chongqing 400715, China. ³These authors contributed equally: Wenbo Hu, Anqiang Jia. ✉ e-mail: xiacy@swu.edu.cn; yiwang28@swu.edu.cn

gland^{11,12,19}. The Ma gland can be divided into three macroscopic segments, the Tail, the Sac, and the Duct, which are characterized by gradients of pH values, ion concentrations, and shear forces^{20–22}. Liquid silk protein is synthesized and stored at a very high concentration in the Tail and Sac and transformed into insoluble fiber via the Duct^{23,24}.

In this context, recombinant major ampullate spidroins (MaSps) have been constructed to achieve specific physical properties of silk, including strength, extensibility, and stickiness^{25–27}. Spider silk-constituting elements (SpiCEs), which are nonspidroin proteins, have been utilized to increase the tensile strength in the case of composite silk films^{28,29}. In addition, a microfluidic device designed to closely simulate natural ionic and pH conditions allowed fibers to be directly pulled from the outlet and then reeled in air, as in natural spinning^{30–32}. It is becoming apparent that the detailed mechanisms underlying dragline silk production, including the cellular architecture and molecular function of the Ma gland as well as the biocomposition and formation of dragline silk, are fundamental to advanced fiber innovation^{11,33}.

To shed light on these mechanisms, we herein present a high-quality chromosome-scale reference genome for the golden orb-web spider *Trichonephila clavata*, which exhibits a colorful body and constructs a large and impressive orb web (Fig. 1a). By multiomic analysis of the Ma gland and dragline silk, we traced the origins of dragline silk components from the Ma gland segments (Tail, Sac, and Duct), elucidated the epigenetic and post-transcriptomic regulatory features of spidroin genes, and built a single-cell spatial architecture of the tri-sectional Ma gland, which provides the first detailed cytological definitions of the spider Ma gland based exclusively on segment-specific, cell-type-specific, space-specific, and dragline silk-related gene expression classification. These results allowed us to generate a comprehensive molecular atlas of natural silk production within the tri-sectional Ma gland, thereby elucidating the generation mechanism of dragline silk. The data were further extended to reveal the convergent evolution of silk production in the silkworm *Bombyx mori*, a model species from another arthropod group, and exhibited the shared molecular characteristics of the spider and silkworm silk glands. Our multiomics datasets are accessible in SpiderDB (<https://spider.bioinfotoolkits.net>) and will be valuable for future explorations of the evolutionary origins of silk production strategies and the creation of biomimetic spider silks.

Results

Chromosomal-scale genome assembly and full spidroin gene set of *T. clavata*

To explore dragline silk production in *T. clavata*, we sought to assemble a high-quality genome of this species. Thus, we first performed a cytogenetic analysis of *T. clavata* captured from the wild in Dali City, Yunnan Province, China, and found a chromosomal complement of $2n = 26$ in females and $2n = 24$ in males, comprising eleven pairs of autosomal elements and unpaired sex chromosomes ($X_1X_1X_2X_2$ in females and X_1X_2 in males) (Fig. 1a). Then, DNA from adult *T. clavata* was used to generate long-read (Oxford Nanopore Technologies (ONT)), short-read (Illumina), and Hi-C data (Supplementary Data 1). A total of 349.95 Gb of Nanopore reads, 199.55 Gb of Illumina reads, and -438.41 Gb of Hi-C raw data were generated. Our sequential assembly approach (Supplementary Fig. 1c) resulted in a 2.63 Gb genome with a scaffold N50 of 202.09 Mb and a Benchmarking Universal Single-Copy Ortholog (BUSCO) genome completeness score of 93.70% (Table 1; Supplementary Data 3). Finally, the genome was assembled into 13 pseudo-chromosomes. Sex-specific Pool-Seq analysis of spiders indicated that Chr12 and Chr13 were sex chromosomes (Fig. 1b; Supplementary Fig. 2). Based on the MAKER2 pipeline³⁴ (Supplementary Fig. 1e), we annotated 37,607 protein-encoding gene models and predicted repetitive elements with a collective length of 1.42 Gb, accounting for 53.94% of the genome.

To identify *T. clavata* spidroin genes, we searched the annotated gene models for sequences similar to 443 published spidroins (Supplementary Data 6) and performed a phylogenetic analysis of the putative spidroin sequences for classification (Supplementary Fig. 12a). Based on the knowledge that a typical spidroin gene consists of a long repeat domain sandwiched between the nonrepetitive N/C-terminal domains¹⁶, 128 nonrepetitive hits were primarily identified. These candidates were further validated and reconstructed using full-length transcript isoform sequencing (Iso-seq) and transcriptome sequencing (RNA-seq) data. We thus identified 28 spidroin genes, among which 26 were full-length (Supplementary Fig. 11a), including 9 *MaSps*, 5 *minor ampullate spidroins (MiSps)*, 2 *flagelliform spidroins (FlSps)*, 1 *tubuliform spidroin (TuSp)*, 2 *aggregate spidroins (AgSp)*, 1 *aciniform spidroin (AcSp)*, 1 *pyriform spidroin (PySp)*, and 5 other *spidroins*. This full set of spidroin genes was located across nine of the 13 *T. clavata* chromosomes. Interestingly, we found that the *MaSp1a-c* & *MaSp2e*, *MaSp2a-d*, and *MiSp-a-e* genes were distributed in three independent groups on chromosomes 4, 7, and 6, respectively (Fig. 1c). Notably, using the genomic data of another orb-weaving spider species, *Trichonephila antipodiana*³⁵, we identified homologous group distributions of spidroin genes on *T. antipodiana* chromosomes (Fig. 1d), which indicated the reliability of the grouping results of our study. When we compared the spidroin gene catalog of *T. clavata* and those of five other orb-web spider species with genomic data^{28,29,36,37}, we found that *T. clavata* and *Trichonephila clavipes* possessed the largest number of spidroin genes (28 genes in both species; Fig. 1e).

To further explore the expression of spidroin genes in different glands, all morphologically distinct glands (major and minor ampullate- (Ma and Mi), flagelliform- (Fl), tubuliform- (Tu), and aggregate (Ag) glands) were cleanly and separately dissected from adult female *T. clavata* spiders except for the aciniform and pyriform glands, which could not be cleanly separated because of their proximal anatomical locations and were therefore treated as a combined sample (aciniform & pyriform gland (Ac & Py)). After RNA sequencing of these silk glands, we performed expression clustering analysis of transcriptomic data and found that the Ma and Mi glands showed the closest relationship in terms of both morphological structure (Fig. 1g) and gene expression (Fig. 1f, h). We noted that the expression profiles of spidroin genes were largely consistent with their putative roles in the corresponding morphologically distinct silk glands; for example, *MaSp* expression was found in the Ma gland (Fig. 1h). However, some spidroin transcripts, such as *MiSps* and *TuSp*, were expressed in several silk glands (Fig. 1h). Unclassified spidroin genes, such as *Sp-GP-rich*, did not appear to show gland-specific expression (Fig. 1h).

In summary, the chromosomal-scale genome of *T. clavata* allowed us to obtain detailed structural and location information for all spidroin genes of this species. We also found a relatively diverse set of spidroin genes and a grouped distribution of *MaSps* and *MiSps* in *T. clavata*.

Dragline silk origin and the functional character of the Ma gland segments

To further evaluate the detailed molecular characteristics of the Ma gland-mediated secretion of dragline silk, we performed integrated analyses of the transcriptomes of the three *T. clavata* Ma gland segments and the proteome and metabolome of *T. clavata* dragline silk (Fig. 2a). Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis of dragline silk mainly showed a thick band above 240 kDa, suggesting a relatively small variety of total proteins (Fig. 2b). Subsequent liquid chromatography–mass spectrometry (LC–MS) analysis identified 28 proteins, including ten spidroins (nine *MaSps* and one *MiSp*) and 18 nonspidroin proteins (one glucose dehydrogenase (GDH), one mucin-19, one venom protein, and 15 SpiCEs of dragline silk (SpiCE-DS)) (Fig. 2b; Supplementary Data 10). Among these proteins, we found that the core protein components of dragline

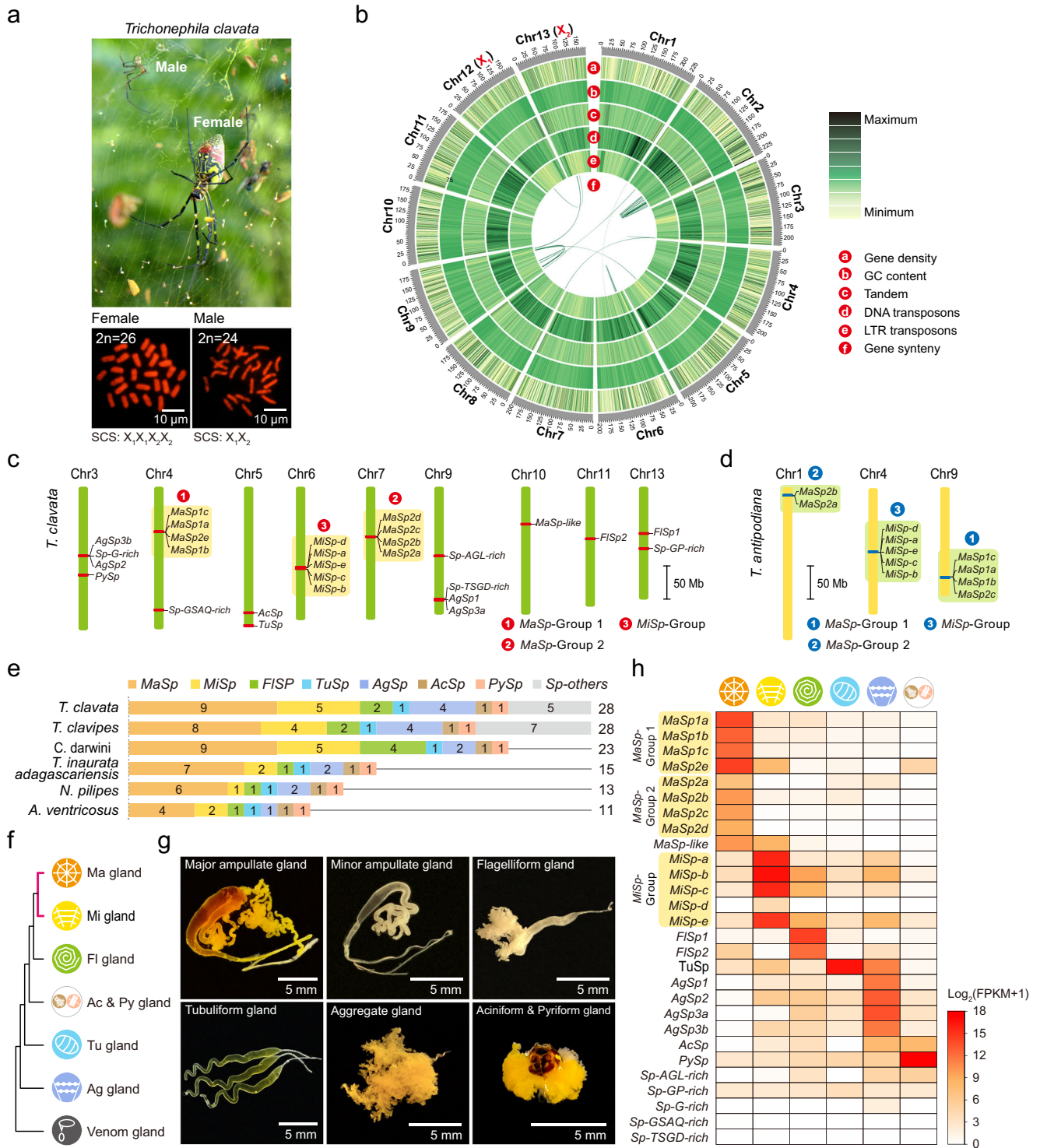


Fig. 1 | Chromosomal-scale genome assembly and full spidroin gene set of *T. clavata*. **a** Photograph of *T. clavata* showing an adult female and an adult male on the golden orb-web (above) and the female and male karyotypes (below). SCS, sex chromosome system. **b** Circular diagram depicting the genomic landscape of the 13 pseudochromosomes (Chr1–13 on an Mb scale). **c** Twenty-eight *T. clavata* spidroin genes anchored on chromosomes. **d** Spidroin gene groups of another orb-web spider, *T. antipodiana*. The published genomic data of *T. antipodiana*³⁵ was analyzed to identify the location information of spidroin genes. **e** Spidroin gene catalog

of six orb-web spider species. **f** Expression clustering of silk glands (major and minor ampullate (Ma and Mi), flagelliform- (FI), tubuliform- (Tu), aggregate- (Ag), and aciniform & pyriform (Ac & Py) glands) and venom glands. The pink line shows the closest relationship between the Ma and Mi glands. **g** Morphology of *T. clavata* silk glands. Similar results were obtained in three independent experiments and summarized in Source data. **h** Expression patterns of 28 spidroin genes in different types of silk glands. Source data are provided as a Source Data file.

Table 1 | Characteristics of the *T. clavata* genome assembly

Estimated genome size (bp) ^a	2,722,533,452
Heterozygosity (%)	1.44
Length of genome assemble (bp)	2,730,479,360
Total length of scaffolds (bp)	2,732,859,360
N50 of scaffolds (bp)	202,090,426
N90 of scaffolds (bp)	173,619,284
Total length of contigs (bp)	2,730,479,360
N50 of contigs (bp)	1,788,282
GC content (%)	31.80
Complete BUSCOs ^b	93.70
Total length of pseudomolecules (bp)	2,629,098,020
Number of genes	37,607
Repeat content (%)	53.94

^aGenome size was estimated based on k-mer (GenomeScope).

^bBUSCO analysis was based on metazoa lineage of protein-coding genes.

silk in order of intensity-based absolute quantification (iBAQ) percentages were MaSp1c (37.7%), MaSp1b (12.2%), SpiCE-DS1 (11.9%, also referred to as SpiCE-NMa1 in a previous study²⁸), MaSp1a (10.4%), and MaSp-like (7.2%), accounting for approximately 80% of the total protein abundance in dragline silk (Fig. 2b). These results revealed potential protein components that might be highly correlated with the excellent strength and toughness of dragline silk.

To evaluate the composition of *T. clavata* dragline silk, we then assessed its metabolite composition and identified a total of 180 components (Supplementary Data 12). Among the metabolites, 109 were classified into ten categories: 34 organic acids, 22 organoheterocyclic compounds, 16 lipids, 13 benzenoids, 5 organic nitrogens, 8 organic oxygens, 5 nucleosides, 3 organooxygens, 2 phenylpropanoids and polyketides, and 1 alkaloid (Fig. 2c; Supplementary Data 13). We noted that xanthurenic acid (XA, a yellow pigment³⁸) was the most abundant pigment (Fig. 2d), while other yellow pigments (such as carotenoids and flavonoids^{39,40}) were not detected in our analysis, implying that XA is the major pigment providing *T. clavata* dragline silk with its golden coloration. The presence of XA was further confirmed by LC-MS analysis (Fig. 2e), consistent with a recent report⁴¹.

To explore the origin of dragline silk components from the tri-sectional Ma gland, we focused on the transcriptomic features of the Tail, Sac, and Duct. We determined that the gene expression profiles of the Tail and Sac were more highly correlated with each other than with that of the Duct (Fig. 2f, g; Supplementary Fig. 13c), implying that the Tail and Sac have similar molecular functions. Furthermore, combined transcriptomic and proteomic analyses revealed the expression patterns of the 28 dragline silk protein transcripts in the Tail, Sac, and Duct (Fig. 2h; Supplementary Data 10). Notably, *MaSp1a-c* & *MaSp2e* (*MaSp*-Group1) were highly coexpressed in the Tail and Sac, and *MaSp2a-d* (*MaSp*-Group2) were highly coexpressed in only the Sac, while neither of these groups of proteins was highly coexpressed in the duct, indicating that the Tail and Sac are major silk-secreting segments.

We then used the tri-sectional Ma gland datasets to trace the source of the metabolite XA. We found that the genes encoding key enzymes involved in the XA biosynthesis (tryptophan metabolism) pathway were activated in all three Ma gland segments (Fig. 2i); in particular, a kynurenine aminotransferase gene (*KAT*, *Tc09G169510*) encoding the primary enzymes catalyzing the transamination of 3-hydroxy-L-kynurenine (3-HK) to XA showed this pattern. These findings suggested that XA is secreted by the Tail, Sac, and Duct.

To characterize the specific biological functions of the Tail, Sac, and Duct related to dragline silk production, we next assigned Gene Ontology (GO) terms to classify the functions of Ma gland segment-specific genes (Supplementary Data 14). We found that the GO terms

that were significantly enriched in the Tail (relative to the Duct and Sac) were mainly related to the synthesis of organic acids (the largest group in the silk metabolome), those in the Sac (relative to the Duct and Tail) were mainly related to the synthesis of lipids (the third-largest group in the silk metabolome), and those in the Duct (relative to the Sac and Tail) were related to ion (Ca²⁺ and H⁺) exchange and chitin synthesis (Fig. 2j). Thus, a segmental division of biological functions was revealed.

Taken together, our results demonstrate a tri-sectional generation process of dragline silk in the Ma gland. Thus, we have established a genetic relationship between dragline silk components and the Tail, Sac, and Duct glands.

Comprehensive epigenetic features and ceRNA network of the Ma gland tri-section

Based on the Ma gland RNA-seq data, we found that the total fragments per kilobase of transcript per million mapped reads (FPKM) of the dragline silk genes accounted for 47.49% and 34.33% of the FPKM values of the Tail and Sac, respectively; however, in the Duct, these genes accounted for only 0.76% of the FPKM values (Supplementary Data 11), indicating that the transcription of dragline silk genes was incredibly efficient in the first two segments. In particular, the *MaSps* within the two groups were highly coexpressed in the specific segments of the Ma gland (Fig. 2h). These findings revealed a segment-specific expression pattern of dragline silk genes.

To better understand the transcriptional regulatory mode of these genes, we first investigated genome-wide chromatin accessibility (CA) in the Tail, Sac, and Duct using the assay for transposase-accessible chromatin with sequencing (ATAC-seq). A total of 702,037 (Tail), 767,517 (Sac), and 653,361 (Duct) significant ATAC peaks (RPKM > 2) were identified in the 2 kb regions upstream and downstream of genes, and 10,501,151 (Tail), 11,356,55 (Sac), and 9,778,368 (Duct) significant ATAC peaks (RPKM > 2) were identified at the whole-genome level. The Tail (mean RPKM: 1.78) and Sac (mean RPKM: 2.04) plots showed genes with more accessible chromatin than the Duct (mean RPKM: 1.59) plots (Fig. 3a). We then analyzed the genome-wide DNA methylation level in the Tail, Sac, and Duct. We found the highest levels of DNA methylation in the CG context (beta value: 0.12 in Tail, 0.13 in Sac, and 0.10 in Duct) and only a small amount in the CHH (beta value: 0.04 in Tail, 0.05 in Sac, and 0.03 in Duct) and CHG (beta value: 0.04 in Tail, 0.05 in Sac, and 0.04 in Duct) contexts (Fig. 3b). Overall, there was no significant difference in methylation levels among the Tail, Sac, and Duct. Taken together, our results suggest a potential regulatory role of CA rather than DNA methylation in the transcription of dragline silk genes.

Next, the visualization of ATAC-seq and methylation datasets of the two *MaSp* groups in a genome browser revealed a reverse trend of peak signals (Fig. 3c, d; Supplementary Fig. 16). We analyzed potential TF motifs among the ATAC-seq peak sets in the 2 kb regions upstream of the transcriptional start sites (TSSs). We identified nine Tail- and Sac-specific TF motifs for *MaSp1b* (in *MaSp*-Group 1) and 13 Sac-specific TF motifs for *MaSp2b* (in *MaSp*-Group 2) (Supplementary Data 15; Supplementary Fig. 17a, b). Interestingly, we noted that the TF motifs closest to the TSSs, such as MYB and homeobox motifs for *MaSp*-Group 1 and two C2H2 motifs for *MaSp*-Group 2, were shared within each *MaSp* group (Fig. 3c, d). However, the Venn network of TF motifs between *MaSp*-Group1 and *MaSp*-Group2 showed little commonality among the Tail, Sac, and Duct (Fig. 3e). Therefore, we concluded that there was a common regulatory pattern within each *MaSp* group but a differentiated regulatory pattern between the two *MaSp* groups.

To investigate the impact of competing endogenous RNAs (ceRNAs: a post-transcriptional regulatory system implemented by miRNA and lncRNA⁴²) corresponding to the regulation of dragline silk genes, we performed a whole-transcriptomic analysis of the tri-sectional Ma gland and identified a total of 527 miRNAs (179 in the Tail, 167 in the

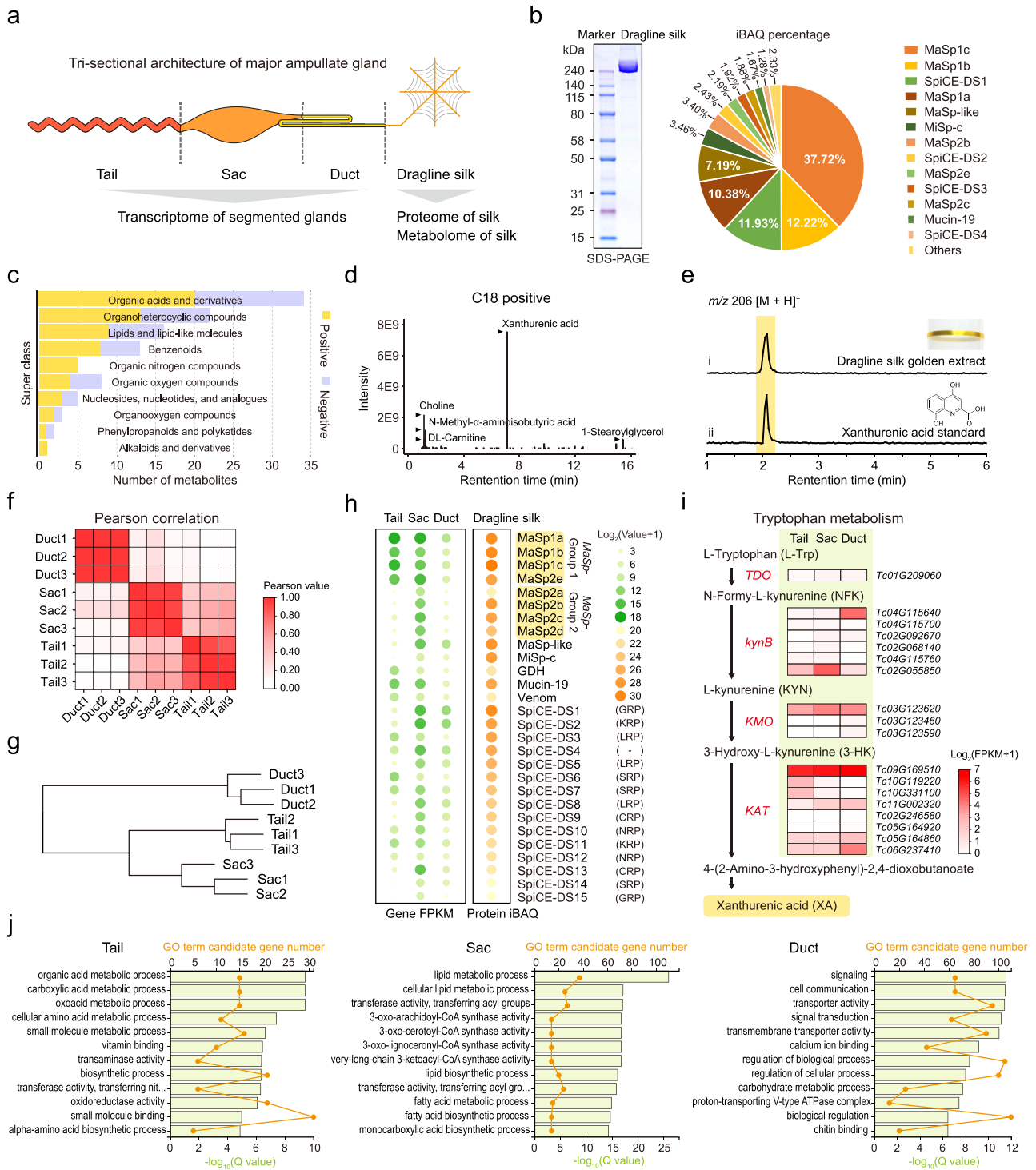


Fig. 2 | Dragline silk origin and the functional character of the Ma gland segments. **a** Schematic illustration of Ma gland segmentation. **b** Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) (left) and LC–MS (right) analyses of dragline silk protein. iBAQ, intensity-based absolute quantification. Similar results were obtained in three independent experiments and summarized in Source data. **c** Classification of the identified metabolites in dragline silk. **d** LC–MS analyses of the metabolites. **e** LC–MS analyses of the golden extract from *T. clavata* dragline silk. The golden pigment was extracted with 80% methanol. The extracted ion chromatograms (EICs) showed a peak at m/z 206 $[M + H]^+$ for xanthurenic acid. **f** Pearson correlation of different Ma gland segments (Tail, Sac, and Duct). **g** Expression clustering of the Tail, Sac, and Duct. The transcriptomic data were clustered according to the hierarchical clustering (HC) method. **h** Combinational

analysis of the transcriptome and proteome showing the expression profile of the dragline silk genes in the Tail, Sac, and Duct. **i** Concise biosynthetic pathway of xanthurenic acid (tryptophan metabolism) in the *T. clavata* Ma gland. Gene expression levels mapped to tryptophan metabolism are shown in three segments of the Ma gland. Enzymes involved in the pathway are indicated in red, and the genes encoding the enzymes are shown beside them. **j** Gene Ontology (GO) enrichment analysis of Ma gland segment-specific genes indicating the biological functions of the Tail, Sac, and Duct. The top 12 significantly enriched GO terms are shown for each segment of the Ma gland. A P -value < 0.05 was set as the criterion for screening significantly enriched GO terms. Source data are provided as a Source Data file.

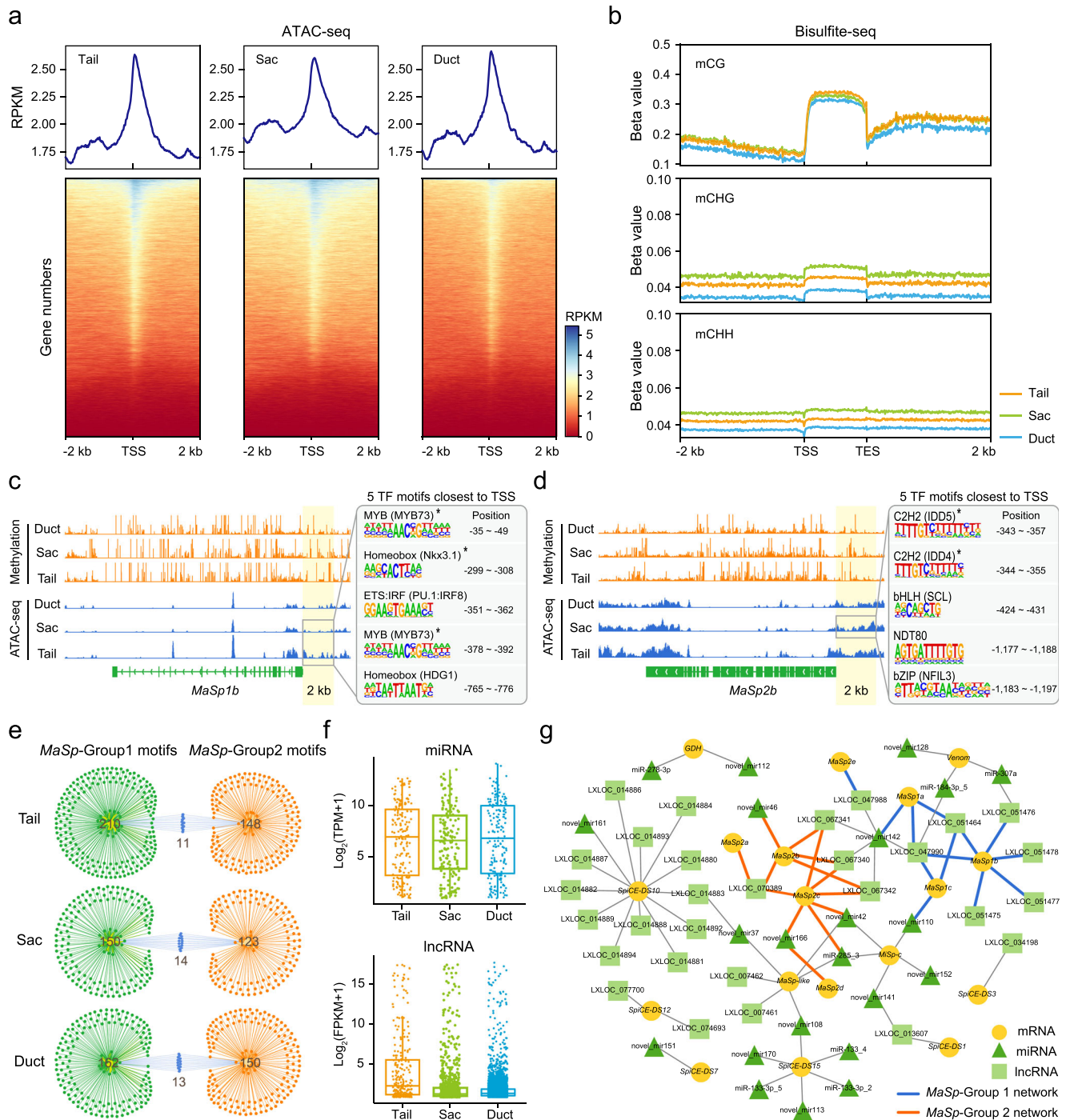


Fig. 3 | Comprehensive epigenetic features and ceRNA network of the tri-sectional Ma gland. **a** Metagenome plot of ATAC-seq signals and heatmap of the ATAC-seq read densities in the Tail, Sac, and Duct. The chromatin accessibility was indicated by the mean RPKM value (upper) and the blue region (bottom). **b** Metagenome plot of DNA methylation levels in CG/CHG/CHH contexts in the Tail, Sac, and Duct. **(c, d)** Screenshots of the methylation and ATAC-seq tracks of the *MaSp1b* (c) and *MaSp2b* (d) genes within the Tail, Sac, and Duct. The potential TF

motifs (E -value $< 1e^{-10}$) in the indicated peak set (2 kb upstream of the TSS) are listed to the right and sorted by position. Asterisks represent the shared TF motif within the corresponding *MaSp* group. **e** Venn network of TF motifs between *MaSp*-Group1 and *MaSp*-Group2. **f** Expression levels of miRNAs and lncRNAs in the Tail, Sac, and Duct. Data are presented as mean \pm SD ($n = 3$ for each Ma segment). Box plots show minimum to maximum (whiskers), 25–75% (box), median (band inside) with all data points. **g** ceRNA network of the dragline silk genes.

Sac, 181 in the Duct) and 10,110 lncRNAs (240 in the Tail, 982 in the Sac, and 4808 in the Duct) (Fig. 3f). From these data, we constructed a potential lncRNA–miRNA–mRNA interaction pairs by using the miRanda⁴³ and RNAhybrid⁴⁴ algorithms to identify the potential binding site between miRNA and lncRNA/mRNA, and then visualized the interaction networks by using Cytoscape software⁴⁵. As shown in Fig. 3g, the ceRNA network of dragline silk genes consisted of 28

lncRNAs, 21 miRNAs, and 13 mRNAs. Remarkably, we noted that the ceRNA networks of *MaSp1a–c* & *MaSp2e* (*MaSp*-Group 1) were tightly clustered, as were those of *MaSp2a–d* (*MaSp*-Group 2); three lncRNAs (LXLOC_047988, LXLOC_047990, and LXLOC_051464) in the *MaSp*-Group 1 network were highly expressed in the Ma gland (Supplementary Fig. 18); one lncRNA (LXLOC_070389) and four miRNAs (novel_mir42, novel_mir46, novel_mir166, and miR-285_3) in the *MaSp*-Group

2 network were highly expressed in the Ma gland (Supplementary Fig. 18); in addition, the ceRNA networks of the two *MaSp* groups were independent of each other (Fig. 3g; Supplementary Fig. 18). These results further revealed potential post-transcriptional networks and the differentiated coregulatory pattern of the genes in the two *MaSp* groups in the Ma gland.

In summary, we observed an abundance of epigenetic and ceRNA signatures associated with the efficient and segment-specific transcription of dragline silk genes. Our data suggested the existence of differential regulation strategies in the three segments of the Ma gland dedicated to achieving the hierarchical gene expression of the *MaSps*.

Single-cell spatial architecture at the whole Ma gland scale

To further explore the cytological basis related to the hierarchical organization of the Ma gland, we generated single-cell and spatial patterns of gene expression in the *T. clavata* whole Ma gland. A total of 9349 high-quality single cells (SCs) were obtained after quality control, and they were then split into ten clusters through uniform manifold approximation and projection (UMAP) clustering (Fig. 4a). Based on the GO analysis of cluster-specific marker genes combined with the expression profiles of the segment-specific genes in each cluster (Supplementary Fig. 21 and Supplementary Note “Cell type annotation”), we carried out the fine annotations of ten SC clusters, namely, cluster 1: Ma gland origin cell (MaGO), cluster 2: MaSp-Group synthesis cell (MGIS), cluster 3: Chitin synthesis cell (CS), cluster 4: Unknown cell, cluster 5: Ampullate lumen skeleton cell (ALS), cluster 6: Ion transport cell (IT), cluster 7: Lipid synthesis cell (LS), cluster 8: pH adjustment cell (PA), cluster 9: MaSp-Group 2 synthesis cell I (MG2S I), and cluster 10: MaSp-Group 2 synthesis cell II (MG2S II), and delineated their sources from the Tail (clusters 1 and 2), Sac (clusters 1, 2, 5, 7, 9, 10), and Duct (clusters 1, 3, 4, 6, 8) (Fig. 4a).

To discern how the Tail, Sac, and Duct develop within the Ma gland, we ordered all single cells according to pseudotime and constructed a developmental trajectory. This resulted in a continuum of cells with three distinct branch points. We found that a set of cells from cluster 1 (ubiquitous SC cluster) assembled at the beginning of the pseudotime period and gradually bifurcated into four end-points representing two segments (the Sac and Duct) of the Ma gland, with the clusters arranged at different branch sites (Fig. 4b). We further investigated the developmental trajectories of the Tail, Sac, and Duct separately. As expected, most cells from cluster 1 were assembled at the beginning of the pseudotime period, while the Sac cells (clusters 2, 5, 7, 9, and 10) and the Duct cells (clusters 3, 4, 6, 8) were grouped into different branches (Fig. 4b; Supplementary Fig. 23). These results identified SC cluster 1 as the Ma gland origin cell and provided insights into the differentiation trajectories of Ma gland cells during cell state transitions.

We next assessed the spatial organization of cell populations in the Ma gland sections. This dataset contained gene expression information from a resource generated across 579 spatial transcriptomic (ST) spots within Ma gland sections (Fig. 4c). After analyzing the transcriptional signatures of ST spots, we identified seven spot clusters (one in the Tail, four in the Sac, and two in the Duct). As a first demonstration of the single-cellular and spatial gene expression patterns in the Ma gland, we visualized the expression of the *MaSp*-Group1 and *MaSp*-Group2 genes in UMAP and ST feature plots to better localize silk protein-secretion cells within our captured cell populations. Interestingly, we found that the *MaSp*-Group1 genes were prominently expressed in the Tail and Sac clusters (SC clusters 1, 2, 5, 7, 9, and 10 and ST clusters “a–f”), while the *MaSp*-Group2 genes were predominantly expressed in the Sac clusters (SC clusters 9–10 and ST cluster “d”) (Fig. 4d; Supplementary Fig. 22a). Thus, the genes exhibited cellular and spatial cluster-specific patterns according to both the scRNA-seq and ST results that were consistent with the results of the expression and regulation analyses (Figs. 2h, 3c, d, g).

To characterize the identified SC and ST clusters related to the molecular function of the Ma gland, we selected 35 of the segment-specific genes identified in the segment transcriptomes to perform expression analyses based on the scRNA-seq and ST datasets. We found that SC cluster 2 and ST cluster “a” were major sets with the functions of organic acid metabolic process and oxidoreductase activity; SC cluster 7 and ST cluster “e” were major sets with the functions of lipid metabolic process and transferase activity; SC cluster 6 and ST cluster “f” were major sets with the function of calcium ion binding; SC cluster 3/8 and ST cluster “c/g” were major sets with the function of proton-transporting V-type ATPase complex; and SC cluster 3 and ST cluster “f” were major sets with the function of chitin binding (Fig. 4e; Supplementary Figs. 25, 26).

In summary, the detailed anatomic and molecular description of the Ma gland revealed a single cell type within the Tail and multiple cell types within the Sac and Duct, highlighting the developmental and functional differentiation of the Ma gland in the tri-section.

Convergent evolution of the tri-sectional silk gland between *T. clavata* and *Bombyx mori*

To extend our investigation to an established model organism that has also evolved specialized glands for spinning silk, we chose the silkworm, *Bombyx mori*, which has a distinct phylogenetic position from spiders in the phylum Arthropoda⁴⁶. Through anatomical observations, we noted a morphological convergence of the silk-spinning gland between the *T. clavata* Ma gland (Tail, Sac, and Duct) and the *B. mori* silk gland (posterior silk gland (PSG), middle silk gland (MSG), and anterior silk gland (ASG)), indicating a one-to-one correspondence of the tri-sectional architecture (Fig. 5a, b). We also found a similar number of silk gland cell types between *T. clavata* and *B. mori*⁴⁷ but differentiated annotations except for the chitin-related process in the Duct/ASG (Fig. 5b; Supplementary Fig. 29). Our previous studies revealed defects in silk spinning by silkworms caused by structural deficiency of the PSG and ASG^{48,49}. However, the genetic manipulation system for spiders has not yet been established. To examine whether the remaining segment of the silkworm silk gland is required for proper silk production, we used *ser1* promoter-driven transgenic overexpression (OE) of a butterfly cytotoxin (pierisin-1A, P1A⁵⁰) to generate an MSG-deficient silkworm (P1A-OE) (Fig. 5c). We found that the MSG of the P1A-OE strain was successfully truncated and that these silkworms failed to spin a cocoon (Fig. 5c), consistent with the phenotypes of PSG- and ASG-deficient silkworms^{48,49}. Our results further demonstrated that the tri-sectional architecture of the silk gland is essential for silk spinning.

To further investigate the convergence between these two species, we performed whole-genome blastp alignment, and we identified 9593 and 7355 orthologous genes in *T. clavata* and *B. mori*, respectively. From these assigned gene ortholog pairs, we selected the *Hsp20* pair (*Tc04G175L20* and *BMSK0007630*) because of the high sequence identity of 87.1% between *T. clavata* and *B. mori* (Fig. 5d); in addition, *Hsp20* was expressed in the silk glands of both *T. clavata* and *B. mori* and simultaneously served as a marker gene of SC cluster 5 of the *T. clavata* Ma gland (Supplementary Data 17), which encodes a small heat shock protein that acts as a protein chaperone to protect other proteins against misfolding and aggregation⁵¹. We next performed CRISPR/Cas9-based knockout (KO) of *Hsp20* in silkworm and successfully identified indels (96.46%) at the target site of the *Hsp20* sgRNA (Fig. 5d). Interestingly, we found that the silk production (cocoon weight and cocoon layer rate) of the *Hsp20*-KO strain was significantly lower than that of the wild type (Fig. 5e). From these results, we concluded that the orthologous *Hsp20* played a positive role in silk production in both the investigated spider and the silkworm.

To further investigate the molecular signatures of silk spinning for evidence of convergent evolution in the spider and silkworm, we next

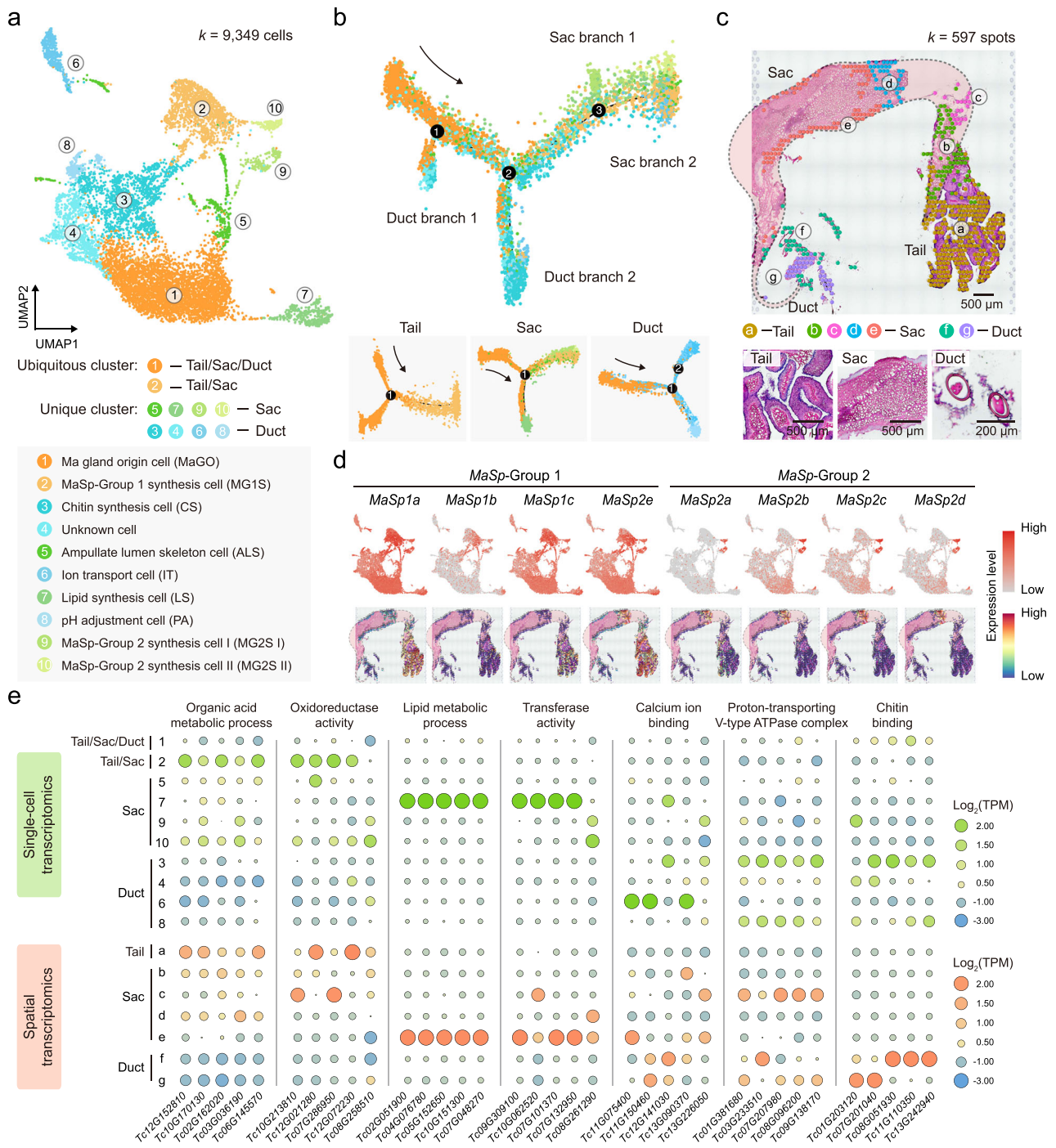


Fig. 4 | Single-cell spatial architecture at the whole-Ma-gland scale. a Uniform manifold approximation and projection (UMAP) analysis of cell types in the Ma gland and their grouping into ten cell clusters. The numbers in white-filled circles indicate cell clusters. The ubiquitous clusters are shown in the yellow series, the Sac clusters in the green series, and the Duct clusters in the blue series. **b** Pseudotime trajectory of all 9349 Ma gland cells. Each dot indicates a single cell, color-coded by the cluster as in (a). The numbers in black-filled circles indicate branch sites. The

black arrows indicate the start of the trajectory. **c** Hematoxylin and eosin staining of Ma gland sections and unbiased clustering of spatial transcriptomic (ST) spots. Dotted lines depict the outline of the Ma gland. Similar results were obtained in three independent experiments and summarized in Source data. **d** UMAP and ST feature plots of the expression of genes in the *MaSp* group. **e** Heatmap showing the expression of Ma gland segment-specific genes in each cell type and each ST cluster along with corresponding GO terms. Source data are provided as a Source Data file.

performed comparative transcriptomic, proteomic, and metabolomic analyses of the tri-sectional silk glands and silks of *T. clavata* and *B. mori*. Shared GO terms were identified between each corresponding silk gland region of *T. clavata* and *B. mori*. We found that three GO terms (calcium ion binding, chitin binding, and signal transduction) were commonly enriched in the Duct and ASG and that one GO term (fatty acid metabolic process) was commonly enriched in the Sac and

MSG (Fig. 5f). These shared GO terms were silk gland-specific and not identified in other tissue types (hemocyte and ovary) (Supplementary Fig. 30). Next, we performed unique gene screening for each silk gland segment using a customized pipeline (Supplementary Fig. 28a). We found that 42, 6, and 2 pairs of orthologous genes were coexpressed in the Duct and ASG, Sac and MSG, and Tail and PSG, respectively (Fig. 5g; Supplementary Data 20). Interestingly, we found that seven ortholog

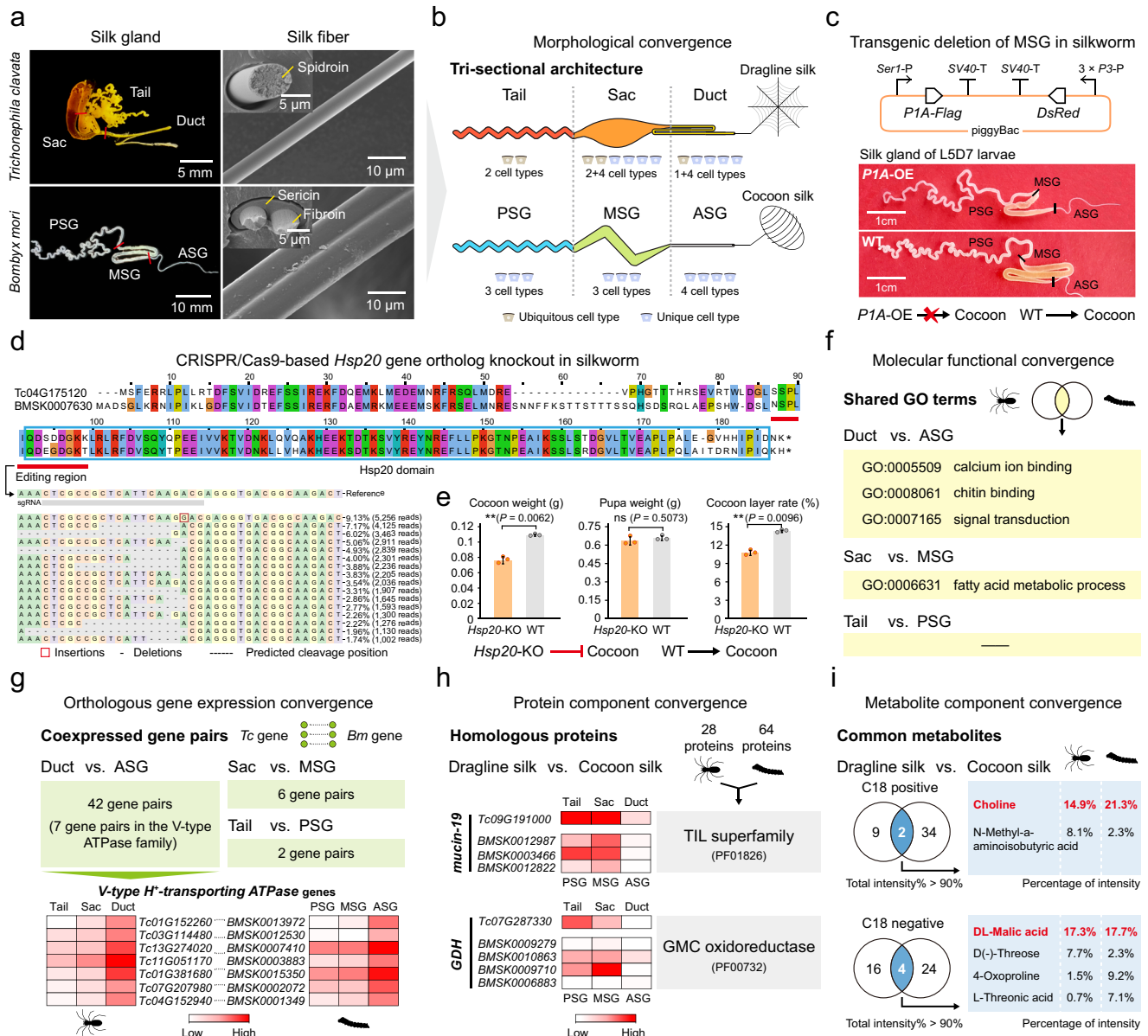


Fig. 5 | Convergent evolution of the tri-sectional silk gland between *T. clavata* and *B. mori*. **a** Morphology of the *T. clavata* Ma gland, dragline silk, *B. mori* silk gland, and cocoon silk. Insets show cross-sections of silk threads. Similar results were obtained in three independent experiments and summarized in Source data. **b** Schematic illustration showing the morphological convergence of the *T. clavata* Ma gland, with the Tail, Sac, and Duct indicated (above), and the *B. mori* silk gland, with the PSG, MSG, and ASG indicated (below). **c** Construction of the piggyBac transgenic vector (above) and silk gland phenotypes of the *P1A*-OE and wild-type (WT) strains in silkworm (below). *Ser1*-P, *Ser1* promoter. $3 \times P3$ -P, $3 \times P3$ promoter. *SV40*-T, *SV40* terminator. LSD7 represents the 7th day of the fifth instar. The arrows indicate the silk production process, and a red cross indicates that the process was blocked. **d** Sequence alignment of the orthologous *Hsp20* proteins (Tc04G175120 and BMSK0007630) in spider and silkworm. The red line indicates the editing region. The arrow shows the distribution of identified alleles around the cleavage

site of the sgRNA. The top 16 sequences with a high percentage were exhibited. **e** Cocoon weight, pupa weight, and cocoon layer rate performance of the CRISPR/Cas9-based *Hsp20* knockout silkworm strain. Data were presented as mean \pm SD ($n = 3$). Statistical comparisons were made using two-tailed Student's *t* test. ns indicates non-significant. **P*-value < 0.01. The arrow indicates the silk production process. The red line ending in a crossbar indicates that the process was suppressed. **f** Molecular functional convergence of the *T. clavata* Ma gland and the *B. mori* silk gland according to GO term analysis. A *P*-value < 0.05 was set as the criterion for screening significantly enriched GO terms. **g** Orthologous gene expression convergence between the *T. clavata* Ma gland and the *B. mori* silk gland. **h**, **i** Component convergence of protein (**h**) and metabolite (**i**) between silks produced by *T. clavata* and *B. mori*. The metabolites with a total intensity percentage above 90% were analyzed. The major metabolites in *T. clavata* dragline silk and *B. mori* cocoon silk are indicated in red. Source data are provided as a Source Data file.

gene pairs involved in the *V*-type *ATPase* family, encoding key factors involved in regulating silk fibrillogenesis⁵², were significantly upregulated in both the Duct and ASG (Fig. 5g). Our results indicated a higher consistency between the Duct and ASG in molecular functions than between the Sac and MSG or the Tail and PSG.

We then explored whether the components of spider and silkworm silks showed convergence by performing Venn analyses of the

silk proteomes and metabolomes across the two species (Supplementary Fig. 28b). We found that mucin-19 and GDH were the only two proteins existing in both *T. clavata* dragline silk and *B. mori* cocoon silk, and more interestingly, these proteins were mainly synthesized and secreted by Tail/PSG and Sac/MSG (Fig. 5h). We also identified six common metabolites in these two silks: choline, N-methyl- α -aminoisobutyric acid, DL-malic acid, D(-)-threose, 4-oxoproline, and

L-threonic acid (Fig. 5i). Among these metabolites, it is worth noting that choline, a component of phospholipids in cell membranes⁵³, and DL-malic acid, which act as a preservative or pH-adjuster⁵⁴, were both major metabolites of dragline silk and cocoon silk (Fig. 5i). We therefore speculated that silk secretion from gland cells to the lumen is accompanied by choline release from the cell membrane and that, in natural silk, anti-rot and anti-bacteria characteristics are conferred by DL-malic acid.

In summary, the shared molecular characteristics of the tri-sectional silk glands of a spider and silkworm indicated convergent evolution under similar cases related to silk spinning and provided rich insights into silk gland function and silk biosynthesis.

Discussion

Herein, we report the first chromosome-scale reference genome assembly of *T. clavata*. This high-quality genome combined with multiomics analyses enabled the elucidation of silk biosynthesis processes in the tri-sectional Ma gland. Our findings led us to hypothesize that the Tail performs the primary function in MaSps secretion, while the Sac, acting as a storage site, secretes a wide range of proteins (MaSps and nonspidroin proteins), and the Duct plays a limited role in protein secretion but a crucial role in protein structural transition^{20,21,24}.

We propose a molecular model comprehensively describing the mechanism underlying dragline silk biosynthesis in the *T. clavata* Ma gland (Fig. 6): (i) The Tail, consisting of two ubiquitous cell types, is the major site of the secretion of organic acids and 13 dragline silk proteins constituting the inner layer of dragline silk. The presence of organic acids can affect protein solubility in water by altering conformational states⁵⁵. In the Tail, these 13 dragline silk genes are highly activated, among which the transcripts encoding *MaSp*-Group 1 proteins are the

dominant components. Relatively high CA in this segment, along with mCG methylation, might contribute to this abundant gene expression. (ii) The Sac, consisting of two ubiquitous and four unique cell types, is the major site of the secretion of lipids and 28 dragline silk proteins constituting the middle layer of dragline silk. The lipids act as a coat that wraps around the dragline silk to regulate its water content⁵⁵. These 28 dragline silk genes are also highly activated in the Sac, among which *MaSp*-Group1, *MaSp*-Group2, *MaSp*-like, *SpiCE-DS1*, *SpiCE-DS2*, *SpiCE-DS4*, and *SpiCE-DS13* are the dominant components. Relatively high CA and mCG methylation might also contribute to the high gene expression observed in this segment. (iii) The Duct, consisting of one ubiquitous and four unique cell types, is the major site of the secretion of chitin, cuticular proteins, ions (Ca^{2+} and H^+), and 13 dragline silk proteins. Chitin and cuticular proteins form the cuticular intima, where shear forces are generated^{21,56}. The ions are responsible for a multi-dimensional state of flux, including ion exchange, pH gradient formation, and dehydration^{11,20}, and the 13 dragline silk proteins constitute the outer layer of dragline silk. The transcription of dragline silk genes is less active in the Duct than in the Tail and Sac, and only 0.76% of the transcripts come from these genes. Relatively low CA in this segment, along with mCG methylation, might contribute to this decrease in expression. Overall, our results not only define the proteomic and metabolic components of dragline silk but also trace their origins from the tri-sectional Ma gland. The numbers of cell types in the Tail, Sac, and Duct are positively correlated with the diversity of their functions.

Our results provide genomic clues for the hierarchically ordered biosynthesis of spidroins. We documented that the *MaSp1a-c* & *MaSp2e*, *MaSp2a-d*, and *MiSp-a-e* genes are distributed in three distinct groups. In addition, we demonstrated that the *MaSps* within each

Molecular basis of tri-sectional dragline silk generation in *T. clavata* major ampullate gland

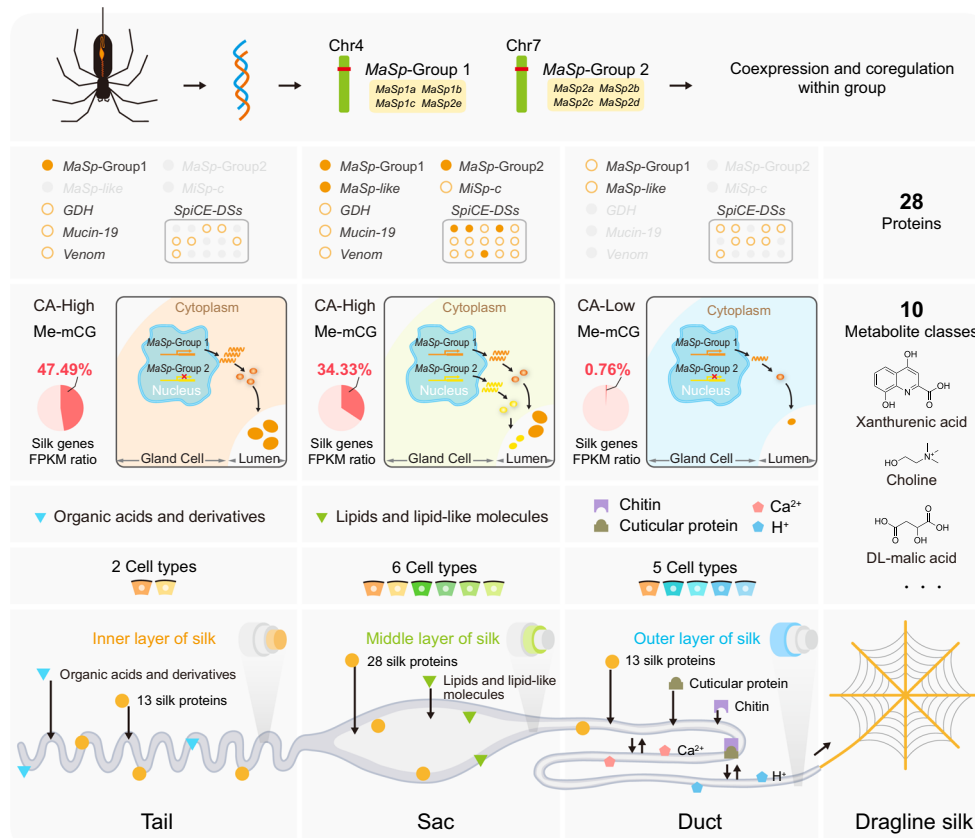


Fig. 6 | Molecular basis of tri-sectional dragline silk generation in the *T. clavata* Ma gland. Orange solid circles represent the genes with an FPKM > 10,000, and orange hollow circles represent the genes with an FPKM < 10,000, CA chromatin accessibility, Me methylation. Adobe Illustrator 2020 was used to create the image.

of these groups exhibited concerted SC and ST expression profiles in the tri-sectional Ma gland. We also identified the group-specific common TF motifs at the epigenetic level and constructed group-specific lncRNA-miRNA-mRNA networks at the ceRNA level. Such results revealed novel structural, expressional, and regulatory characteristics of spidroin genes that have not been reported in other spider genomes. Spidroins are thought to exert important influences on the mechanical properties of dragline silk^{16,57,58}. As the group distributions of spidroin genes have already been identified in the high-quality spider genomes of *T. antipodiana*³⁵ and *Latrodectus elegans*⁵⁹ with cataloged spidroin proteins, our data fill an important information gap regarding the arrangement of spidroin genes on whole chromosomes and provide a general entry point for the mechanical differentiation of silk and the further study of spider genome evolution.

In addition to the above genomic findings, our results provide cellular clues for the tri-sectional organization and functional division of the Ma gland. While previous studies have shown that there are two or three cell types that secrete spidroins in the Ma gland of orb-web spiders^{21,60}, the number of cell types in the Ma gland has been a subject of debate. The divergent findings concerning these cell types have suggested that they may vary between species but probably also reflect technological difficulties in the capture of intact cells after sample preparation for cellular and morphological studies^{20,21,61}. Our study provides well-verified scRNA-seq and ST data indicating the existence of ten cell types in the whole Ma gland, thus contributing to answering this contentious question. The fascinating single-cell spatial architecture of the gland further indicates the differentiation trajectory of Ma gland cells during cell state transitions, suggesting that the Sac and Duct are derived from the early cell type present in the Tail via cell differentiation.

Our genomic, proteomic, and metabolomic analyses further reveal details about the generation of silk in silk-producing glands. Physical and material studies have shown that liquid silk dope is transformed into insoluble fiber through the combined effects of pH and ion gradients as well as extensional and shear forces as it migrates through the silk gland^{3,62}. Our work provides biological evidence of this role of the silk gland and further demonstrates high convergence of several molecular functions in the spider Ma gland Duct and the silkworm ASG, including calcium ion binding, chitin binding, signal transduction, and *V-type H⁺-transporting ATPase* expression. Such overlaps were silk gland-specific and not identified in other tissue types (hemocyte and ovary) (Supplementary Fig. 30). Furthermore, we identified convergent silk components in spider and silkworm silks, including two proteins (mucin-19 and GDH) and two major metabolites (choline and DL-malic acid). These analogous aspects indicated critical processes and components essential to silk formation, which contribute to understanding the evolution of silk spinning and can be used as references for the optimization of Duct-based artificial spinning devices and the design of silk protein solutions in artificial spinning.

In conclusion, our current work comprehensively illustrates the biological basis of dragline silk formation in an orb-weaving spider. To our knowledge, this study is the first to reveal the biological mechanism of silk spinning in spiders in such detail. We believe that the chromosome-scale reference genome of the golden orb-web spider and the molecular atlas of the tri-sectional Ma gland presented here will facilitate the understanding of the spider silk-spinning process in spiders and further serve as a powerful platform for evolutionary studies of silk-spinning organisms. More importantly, the significant molecular characteristics revealed by our results and the generated datasets should ultimately pave the way for producing optimized synthetic silks through genetic and biomimetic manipulations.

Methods

Karyotyping

T. clavata spiders were captured from the wild in Dali City, Yunnan Province, China. The captured adult female *T. clavata* spiders were

cultured indoors at 25 °C, and their eggs were sampled on the third day after spawning for DNA karyotyping. Approximately 20 eggs were mashed with tweezers, mixed with 1.5 mL of 0.05% colchicine for three hours, immobilized with 0.075 mol/L KCl solution (20 min), and fixed in methanol:acetic acid (3:1) solution (30 min). The cell suspension was dropped onto precooled glass slides, flame-dried, stained with 5% Giemsa solution (50 min), rinsed with running water, and air-dried. Chromosome karyotypes were observed by using an Olympus microscope with 40× magnification.

Short- and long-read transcriptome sequencing

Short-read transcriptome sequencing. To more comprehensively capture the gene set of *T. clavata*, the 18 following fresh tissues (Supplementary Data 1) were dissected in PBS solution: female body (TcF), male body (TcM), hemolymph (Hem), pedipalps (Ped), legs (Leg), epidermis (Epi), fat body (Fat), ovary (Ova), venom gland (Ven), whole major ampullate gland (Ma), tail of the Ma gland (Tail), sac of the Ma gland (Sac), duct of the Ma gland (Duct), minor ampullate gland (Mi), tubuliform gland (Tu), flagelliform gland (Fl), aggregate gland (Ag), and aciniform and pyriform glands (Ac & Py); and three independent biological replicates were performed for each sample. Then, total RNA was extracted using TRIzol reagent and sequenced on the DNBSEQ platform.

Long-read transcriptome sequencing. To improve the quality of spidroin gene annotation, we also performed full-length transcriptome sequencing. Seven silk gland tissues (Ma, Mi, Tu, Fl, Ag, Ac, and Py glands) were mixed to construct libraries for Isoform sequencing (Iso-seq) on the PacBio platform. To reduce fragment bias, we constructed two libraries: one ranging from 0–10 kb and the other from 5–10 kb.

Genome assembly

For the de novo assembly of the *T. clavata* genome (Supplementary Fig. 1c), Nanopore long reads were initially corrected using Canu v2.2⁶³ with the default parameters, and SMARTdenovo (<https://github.com/ruanjue/smartdenovo>) was subsequently employed to produce the primary contigs. Then, three rounds of contig correction were performed using Racon v1.5.0⁶⁴. To improve the accuracy of the assembly, Pilon v1.24⁶⁵ was utilized for contig polishing based on Illumina short reads. The paired Hi-C short reads were primarily trimmed by using Hicpro v2.10⁶⁶, and the optimized reads were mapped to the draft contig with Juicer v1.6.2⁶⁷ using 3D-DNA v180419⁶⁸ for pseudochromosome construction and the Juicebox v1.9⁶⁷ tool for manual corrections.

Genome annotation

Repeat annotation. A custom *T. clavata* repeat database was de novo identified by using RepeatModeler v2⁶⁹ and LTR_FINDER v1.06⁷⁰. Then, the obtained repeat sequences were imported into RepeatMasker v4.05⁷¹ to search for transposable elements (TEs) in the genome. Tandem Repeats Finder v4.09⁷² was used to find tandem repeats.

Gene annotation. Gene structure prediction was based on homology data and de novo prediction. The data files included RNA-seq data (RNA-seq of 17 tissues), Iso-seq data, and protein sequences of closely related species (four species: *A. ventricosus*, *A. bruennichi*, *T. antipodiana*, and *T. clavipes*). AUGUSTUS v3.2.3⁷³ (<http://bioinf.uni-greifswald.de/>) software was used for de novo gene prediction. Evidence-based gene annotations were produced in the MAKER2³⁴ pipeline with the default parameters. Then, we integrated the MAKER and de novo results based on the following screening criteria: the proportion of repeat sequences was less than 50%, the protein sequence length was greater than 50 aa, the FPKM value was greater than 0.1 in at least ten samples, and the sequence was supported by Iso-seq evidence (identity > 95%) or a corresponding sequence from at least one closely related species (*E*-value > 1e⁻⁵, identity > 50%).

Spidroin analysis

Spidroin identification. A total of 443 redundant spidroin sequences (Supplementary Data 6) were downloaded from the NCBI protein database (<https://www.ncbi.nlm.nih.gov/protein>). All protein sequences of *T. clavata* were used to align these spidroins by blastp (E -value $< 1e^{-10}$), and 128 nonrepetitive hits were primarily identified. To further reliably identify spidroins, we double-checked that each sequence included typical repeats and N- or C-termini, and the final 28 putative spidroins were finally identified.

Artificial full-length spidroin assembly. Briefly, (1) 79 complete spidroin sequences, 48 N-terminal sequences, and 22 C-terminal sequences (Supplementary Data 7) were mapped to the *T. clavata* genome to generate the GFF track files using the genBlast v138⁷⁴ tool; (2) the BAM files of both the silk gland Iso-seq data and the silk gland RNA-seq data were extracted as expression data tracks; (3) the GFF3 files of the 28 spidroins were extracted as one track; (4) the track files from the previous three steps were imported into IGV v2.9.4⁷⁵, and the integrity of each spidroin was manually checked; and (5) the DNA sequence of the incomplete spidroin on the corresponding genome regions were truncated, then repaired based on frameshift translation and prediction using the online tool AUGUSTUS v3.2.3⁷³ (<http://Augustus.gobics.de/>).

Spidroin motif analysis. Amino acid content and motifs were quantified using a Perl script. The motifs included $(GA)_n$, $(A)_n$, GGX, XQQ, and GPGXX. O-glycosylation was predicted with the NetOGlyc v4.0⁷⁶ tool.

Proteomics

The dragline silk of the spider *T. clavata* and cocoon silk of the silkworm *B. mori* were cut with scissors (Supplementary Data 1). First, silk samples ($n = 3$, 20 mg of each sample) were dissolved in 500 μ L of 9 M LiSCN, and the supernatant was collected after centrifugation (10,000 \times g, 10 min). Protein concentrations were measured via the Bradford protein assay (Beyotime, China). Then, the samples were diluted 10-fold in an 8 M urea solution, mixed with 5 \times SDS-PAGE buffer, and separated using a NuPAGE 4–12% Bis-Tris protein gel (Thermo Fisher Scientific, USA). The samples were subsequently digested with trypsin (1/50 μ g protein) for 20 h at 37 °C. The digested samples were lyophilized and resuspended in 0.1% formic acid (FA) and then measured via LC-MS/MS on a Q Exactive HF-X column (Thermo Fisher Scientific, USA). The gradient was set as 5–95% acetonitrile in 0.1% formic acid. The flow rate and analysis time were 600 nL/min and 70 min, respectively. The instrument parameters were as follows: full MS scan ranged from m/z 350 to 1500 with a resolution of 60,000 (at m/z 200); the automatic gain control (AGC) target value was 3×10^6 and the maximum ion injection time was 20 ms; the top 40 precursors of the highest abundance in the full scan were selected and fragmented by higher energy collisional dissociation (HCD) and analyzed in MS/MS, where the resolution was 15,000 (at m/z 200), the AGC target value was 1×10^5 , the maximum ion injection time was 45 ms, a normalized collision energy was set as 27%, an intensity threshold was 2.2×10^4 , and the dynamic exclusion parameter was 20 s. The resulting raw MS data were analyzed with MaxQuant v1.3.0.5⁷⁷. The MaxQuant searches were performed against the SpiderDB (<https://spider.bioinfotoolkits.net>) database containing 37,607 protein sequences. The search parameters were as follows: methionine oxidation and N-terminal acetylation were set as variable modifications, and carbamidomethyl cysteine was set as the fixed modification; the minimal peptide length was set to seven amino acids and a maximum of two miscleavages were allowed; both peptide and protein identifications were filtered at a 1% false discovery rate; the identified protein required at least one unique peptide and all common reverse hits and contaminants were removed. Based on the sum of peak intensities, the intensity-based absolute quantification (iBAQ) algorithm in MaxQuant was used to calculate the protein abundances.

Metabolomics

Shredded dragline silk and cocoon silk samples ($n = 6$, 20 mg of each sample) were used for metabolite extraction (Supplementary Data 1). Each silk sample was placed in a tube, homogenized in 500 μ L of 80% methanol, kept on ice for 10 min, and centrifuged at 15,000 \times g for 20 min at 4 °C. The supernatant (250 μ L per sample) was collected for LC-MS/MS experiments and injected onto a Hypesil Goldcolumn (C18, 100 \times 2.1 mm, 1.9 μ m) using a 17 min linear gradient at a flow rate of 0.2 mL/min. The eluents for the positive polarity mode were eluent A (0.1% FA in Water) and eluent B (Methanol). The eluents for the negative polarity mode were eluent A (5 mM ammonium acetate, pH 9.0) and eluent B (Methanol). The solvent gradient was set as follows: 2% B, 1.5 min; 2–100% B, 3 min; 100% B, 10 min; 100–2% B, 10.1 min; 2% B, 12 min. Q Exactive HF mass spectrometer (Thermo Fisher Scientific, USA) was operated in positive/negative polarity mode with a spray voltage of 3.5 kV, capillary temperature of 320 °C, sheath gas flow rate of 35 psi, and aux gas flow rate of 10 L/min, S-lens RF level of 60, Aux gas heater temperature of 350 °C. Full scan ranged from m/z 350 to 1500 with a resolution of 60,000 (at m/z 200). The AGC target value was 3×10^6 and the maximum ion injection time was 20 ms. The top 40 precursors of the highest abundant in the full scan were selected and fragmented by HCD and analyzed in MS/MS, where the resolution was 45,000 (at m/z 200) for 10 plex, the AGC target value was 5×10^4 , the maximum ion injection time was 86 ms, a normalized collision energy was set as 32%, an intensity threshold was 1.2×10^5 , and the dynamic exclusion parameter was 20 s. Metabolite identities and quantities were confirmed by Compound Discoverer 3.1 (CD 3.1) software analysis, including alignment against the mzCloud (<https://www.mzcloud.org/>), mzVault (mzCloud Offline for mzVault 2.3_2020A.db), and Mass List (Endogenous Metabolite-Animal-POS/NEG-20191029) databases (retrieval date: 6/15/2020). Furthermore, the KEGG⁷⁸ (<https://www.kegg.jp/>), HMDB⁷⁹ (<https://hmdb.ca/metabolites>), and LIPIDMaps⁸⁰ (<http://www.lipidmaps.org/>) databases (retrieval date: 7/17/2020) were utilized for metabolite annotation.

Methylation analysis

Whole-genome bisulfite sequencing (WGBS). The separated Tail, Sac, and Duct of the Ma gland were subjected to WGBS. First, genomic DNA was extracted and fragmented to a mean size of 250 bp by sonication using a Bioruptor (Diagenode, Belgium), followed by the addition of an “A” base to the 3' terminal ends of the DNA fragments and the ligation of methylated adaptors to both ends of DNA fragments. The EZ DNA Methylation-Gold Kit (ZYMO) was used for the bisulfite conversion of ligated DNA. Products were purified using a QIAquick Gel Extraction Kit (Qiagen, USA) and amplified by PCR. Finally, libraries were sequenced on the Illumina HiSeq 2500 platform.

Calculation of methylation levels. Clean reads were mapped to the reference genome by using BSMAP 2.90 software (<https://code.google.com/archive/p/bsmap/>) with the parameters (-v 8 -z 33 -p 4 -n 0 -w 20 -s 16 -f 10 -L 100) after removing the adaptors and low-quality reads ($N > 10\%$ and $Q < 20$). Methylation levels were estimated by determining the corresponding beta values: $\beta = M/(M + U)$, where M and U represent methylated and unmethylated signal values, respectively. The BED file of methylation data was converted to BigWig format by using the “bedGraphToBigWig” command line and visualized in IGV⁷⁵.

Assay for transposase-accessible chromatin (ATAC)

ATAC-seq library preparation and sequencing. The separated Tail, Sac, and Duct of the Ma gland were subjected to ATAC sequencing. Samples were ground individually in liquid nitrogen and transferred to precooled cell lysate solution at 4 °C for 10 min. After centrifugation (500 \times g, 4 °C, and 10 min), the supernatant was removed, and the cell pellets were retained. The pellets were washed with precooled buffer

solution and centrifuged (500 × g, 4 °C, and 10 min) to collect cell nuclei. The cell nucleus solution was mixed with Tn5-transposase and incubated for 30 min at 37 °C, followed by DNA fragment collection, PCR amplification, and sequencing on the Illumina PE150 platform.

Chromatin accessibility evaluation. Clean reads were mapped to the genome using Bowtie2⁸¹ software, the data were converted from .sam format to .bam format, and the reads were filtered by using SAMtools v0.1.19⁸² with the parameters (view -b -f 2 -q 30). The peaks of chromatin accessible regions were identified in each sample separately using MACS2⁸³ (-shift -100-extsize 200-nomodel -B -q 0.05). The “intersect” command of BEDTools v2.26.0⁸⁴ was used to identify shared peaks between replicate samples. Heatmaps were generated using the “plotHeatmap” function of deepTools v3.5.1⁸⁵. Significantly enriched motifs were identified by the HOMER⁸⁶ tool.

Whole-transcriptome analysis

Total RNA was extracted from the Tail, Sac, and Duct of the Ma gland with TRIzol reagent, followed by lncRNA and miRNA library construction, as described below.

lncRNA sequencing and analysis. Total RNA samples were incubated with DNase I to digest DNA fragments and with RNase H to remove rRNA. After RNA purification, cDNA was synthesized with an “A” base added to the 3' end, and adapter ligation was performed, followed by UDG digestion of the second strand and PCR amplification. The constructed library was sequenced on the DNBSEQ platform. The clean reads were mapped to the reference genome using HISAT2⁸⁷ (-phred64-sensitive-no-discordant-no-mixed -I 1 -X 1000), and the transcripts were assembled using StringTie⁸⁸ (-f 0.3 -j 3 -c 5 -g 100 -s 10000 -p 8) and compared with known mRNAs using the CuffCompare (-p 12) program of Cufflinks⁸⁹. Then, the cooperative data classification (CPC), txCdsPredict, and coding noncoding index (CNCI) methods were used to distinguish mRNAs and lncRNAs. Cis and trans methods were used to evaluate the target mRNAs of lncRNAs. Cis-lncRNAs were defined as those lncRNAs located 10 kb upstream or 20 kb downstream of the target mRNA. The binding energy (BE) of the lncRNAs and mRNAs was analyzed by RNAplex⁹⁰, and if the BE was < -30, the lncRNA was considered to be a trans-lncRNA.

Small-RNA sequencing and analysis. Total RNA was purified and isolated using PAGE electrophoresis and was then excised from the gel to obtain 18-30 nt RNA fragments. The 3' end of each small RNA fragment was ligated to a 5'-adenylated and 3'-blocked adaptor, and a unique molecular identifiers (UMI)-labeled primer was added for hybridization with the 3' adaptor. Then, the 5' adaptor was hybridized with the 5' end of the UMI label. The cDNA strand was synthesized with UMI-labeled primers, and highly sensitive polymerase was used to amplify the product. We used Bowtie2⁸¹ (-q -L 16 -phred64 -p 6) to align the clean reads to the reference genome and the Rfam small-RNA database⁹¹ (using cmsearch with the parameters -cpu 6 -noali). To identify miRNAs, which were annotated and predicted with the miRBase⁹² database (<https://www.mirbase.org/>) and miRDeep2⁹³ tool, targets were predicted between miRNAs and lncRNAs/mRNAs using miRanda⁴³ (-en -20 -strict) and RNAhybrid⁴⁴ (-b 100 -c -f 2,8 -m 100000 -v 3 -u 3 -e -20 -p 1).

Single-cell transcriptomics

Sample processing and sequencing. Freshly dissected Ma glands from three independent spiders were washed using culture medium and cut into small pieces, and tissue pieces were digested in a constant-temperature incubator for 45 min. After digestion, the medium was filtered through a 40 μm cell sieve and washed at least three times via centrifugation for 5 min at 300 × g and 4 °C. Finally, the cell suspensions with activity greater than 85% and agglomeration rates less than

5% were used for library construction and sequencing. The library was constructed according to the manufacturer's instructions (10 × Genomics) on the 10 × Chromium Single Cell 3' Platform.

Data analysis. The raw data were imported into the cellranger v4.0.0 pipeline to obtain the expression matrices of each cell and each gene. To remove low-quality cells, cell numbers and gene UMI counts within the range of the median ± 2-fold the median absolute deviation (MAD) were retained for further analysis. Additionally, doublets (≥ 2 cells in one oil droplet) were removed using DoubletFinder v2.0.3⁹⁴. A total of 9349 high-quality cells were used to analyze clustering and expression based on the Seurat v4.0.6⁹⁵ package. The developmental trajectories of all cell clusters were predicted with the Monocle v2.22.0⁹⁶ package.

Spatial transcriptomics

Sample processing and sequencing. Fresh Ma glands were frozen with isopentane in liquid nitrogen, embedded with OCT, and stored in an airtight container at -80 °C. Embedded frozen tissue was sectioned, and sections with acceptable RNA quality (RIN > 7.0) were used for further experiments. Then, the sections were placed on Visium Spatial slides, fixed with methanol, stained with hematoxylin and eosin, and observed by a BX53 microscope (Olympus, Japan). The polyadenylated mRNA was captured with the primers employed for spots. RT Master Mix reverse transcription reagent was subsequently added to permeate tissue sections, and second-strand cDNA synthesis was performed on the slide by adding the Second Strand Mix. The obtained cDNA was denatured from each capture zone and transferred to the corresponding tube for amplification, library construction, and sequencing on the Illumina NovaSeq600 platform.

Data analysis. After removing adapters and low-quality (N > 3, Q < 5, and base ratio ≥ 20%) reads, the spaceranger v1.3.1 mkref and count programs were used for data preprocessing. Then, spot clustering was analyzed using the Seurat v4.0.6⁹⁵ package.

Convergent evolution analysis

To evaluate the occurrence of convergent evolution between the silkworm silk gland and the spider Ma gland, we compared five aspects (see Supplementary note for more details): (1) the morphological structure of the silk glands; (2) the microstructures of dragline silk and cocoon silk observed under a SU3500 scanning electron microscopy (SEM) (Hitachi, Japan) with an accelerating voltage of 5 kV; (3) the expression convergence of silk gland genes, (4) the homologous protein components of dragline silk and cocoon silk; and (5) the metabolite components of dragline silk and cocoon silk. The orthologous genes shared by the silkworm and spider were identified using the blastp⁹⁷ program (*E*-value < 1e⁻⁵).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All high-throughput sequencing raw data in this project were deposited into the Genome Sequence Archive (GSA) of the National Genomics Data Center (NGDC, <https://ngdc.cncb.ac.cn/>) and are available through BioProject ID PRJCA014503. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org/cgi/GetDataset>) via the iProX partner repository⁹⁸ with the dataset identifier PXD038734. The metabolomics data have been deposited to the NGDC (<https://ngdc.cncb.ac.cn/omix/release/OMIX002862>, <https://ngdc.cncb.ac.cn/omix/release/OMIX002863>). The single-cell and spatial metrics have been deposited in FigShare (https://figshare.com/articles/dataset/Single_cell_matrices_zip/20399475, https://figshare.com/articles/dataset/Spatial_matrices_zip/20399475).

Transcriptomics_zip/20399580). The genome assembly, gene annotation, and multiomics analysis results of *Trichonephila clavata* are also available on SpiderDB (<https://spider.bioinfotoolkits.net>). Source data are provided for Figs. 1, 2, 4, 5. Source data are provided with this paper.

Code availability

Custom scripts and workflows for data analysis are available at FigShare (https://figshare.com/articles/dataset/SpiderGenomeAnalysis_code_zip/20399652, https://figshare.com/articles/dataset/SpiderGenomeAnalysis_code1_zip/20399706).

References

1. Natural History Museum Bern. *World Spider Catalog*. Version 23.5. <http://wsc.nmbe.ch> (accessed 17 July 2022).
2. Pennisi, E. Untangling spider biology. *Science* **358**, 288–291 (2017).
3. Vollrath, F. & Knight, D. P. Liquid crystalline spinning of spider silk. *Nature* **410**, 541–548 (2001).
4. Garrison, N. L. et al. Spider phylogenomics: untangling the spider tree of life. *PeerJ* **4**, e1719 (2016).
5. Roberts, A. D. et al. Synthetic biology for fibres, adhesives and active camouflage materials in protection and aerospace. *MRS Commun.* **9**, 486–504 (2019).
6. Whittall, D. R., Baker, K. V., Breitling, R. & Takano, E. Host systems for the production of recombinant spider silk. *Trends Biotechnol.* **39**, 560–573 (2021).
7. Liu, D. et al. Spider dragline silk as torsional actuator driven by humidity. *Sci. Adv.* **5**, eaau9183 (2019).
8. Rising, A. & Johansson, J. Toward spinning artificial spider silk. *Nat. Chem. Biol.* **11**, 309–315 (2015).
9. Vollrath, F. & Porter, D. Silks as ancient models for modern polymers. *Polymer* **50**, 5623–5632 (2009).
10. Vollrath, F., Porter, D. & Holland, C. There are many more lessons still to be learned from spider silks. *Soft Matter* **7**, 9595–9600 (2011).
11. Malay, A. D., Craig, H. C., Chen, J., Oktaviani, N. A. & Numata, K. Complexity of spider dragline silk. *Biomacromolecules* **23**, 1827–1840 (2022).
12. Li, J. et al. Spider silk-inspired artificial fibers. *Adv. Sci.* **9**, e2103965 (2022).
13. Blamires, S. J., Blackledge, T. A. & Tso, I. M. Physicochemical property variation in spider silk: ecology, evolution, and synthetic production. *Annu. Rev. Entomol.* **62**, 443–460 (2017).
14. Lewis, R. V. Spider silk: ancient ideas for new biomaterials. *Chem. Rev.* **106**, 3762–3774 (2006).
15. Garb, J. E., Sharma, P. P. & Ayoub, N. A. Recent progress and prospects for advancing arachnid genomics. *Curr. Opin. Insect Sci.* **25**, 51–57 (2018).
16. Gatesy, J., Hayashi, C., Motriuk, D., Woods, J. & Lewis, R. Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science* **291**, 2603–2605 (2001).
17. Chaw, R. C., Correa-Garhwal, S. M., Clarke, T. H., Ayoub, N. A. & Hayashi, C. Y. Proteomic evidence for components of spider silk synthesis from black widow silk glands and fibers. *J. Proteome Res.* **14**, 4223–4231 (2015).
18. Vollrath, F. & Selden, P. The role of behavior in the evolution of spiders, silks, and webs. *Annu. Rev. Ecol. Evol. S.* **38**, 819–846 (2007).
19. Ramezaniaghdam, M., Nahdi, N. D. & Reski, R. Recombinant spider silk: promises and bottlenecks. *Front. Bioeng. Biotechnol.* **10**, 835637 (2022).
20. Andersson, M. et al. Carbonic anhydrase generates CO₂ and H⁺ that drive spider silk formation via opposite effects on the terminal domains. *PLoS Biol.* **12**, e1001921 (2014).
21. Andersson, M., Holm, L., Ridderstrale, Y., Johansson, J. & Rising, A. Morphology and composition of the spider major ampullate gland and dragline silk. *Biomacromolecules* **14**, 2945–2952 (2013).
22. Knight, D. P. & Vollrath, F. Changes in element composition along the spinning duct in a *Nephila* spider. *Naturwissenschaften* **88**, 179–182 (2001).
23. Andersson, M., Johansson, J. & Rising, A. Silk spinning in silkworms and spiders. *Int. J. Mol. Sci.* **17**, 1290 (2016).
24. Foo, C. W. P. et al. Role of pH and charge on silk protein assembly in insects and spiders. *Appl. Phys. A-Mater.* **82**, 223–233 (2006).
25. Teule, F. et al. Silkworms transformed with chimeric silkworm/spider silk genes spin composite silk fibers with improved mechanical properties. *Proc. Natl Acad. Sci. USA.* **109**, 923–928 (2012).
26. Xu, J. et al. Mass spider silk production through targeted gene replacement in *Bombyx mori*. *Proc. Natl Acad. Sci. USA.* **115**, 8757–8762 (2018).
27. Zhang, X. L. et al. CRISPR/Cas9 initiated transgenic silkworms as a natural spinner of spider silk. *Biomacromolecules* **20**, 2252–2264 (2019).
28. Kono, N. et al. Multicomponent nature underlies the extraordinary mechanical properties of spider dragline silk. *Proc. Natl Acad. Sci. USA* **118**, e2107065118 (2021).
29. Kono, N. et al. Orb-weaving spider *Araneus ventricosus* genome elucidates the spider genome catalogue. *Sci. Rep.* **9**, 8380 (2019).
30. Luo, J. et al. Tough silk fibers prepared in air using a biomimetic microfluidic chip. *Int. J. Biol. Macromol.* **66**, 319–324 (2014).
31. Luken, A. et al. Biocompatible micron-scale silk fibers fabricated by microfluidic wet spinning. *Adv. Healthc. Mater.* **10**, e2100898 (2021).
32. Kinahan, M. E. et al. Tunable silk: using microfluidics to fabricate silk fibers with controllable properties. *Biomacromolecules* **12**, 1504–1511 (2011).
33. Ramezaniaghdam, M., Nahdi, N. D. & Reski, R. Recombinant spider silk: promises and bottlenecks. *Front. Bioeng. Biotechnol.* **10**, 835637 (2022).
34. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinforma.* **12**, 491 (2011).
35. Fan, Z. et al. A chromosome-level genome of the spider *Trichonephila antipodiana* reveals the genetic basis of its polyphagy and evidence of an ancient whole-genome duplication event. *Giga-science* **10**, giab016 (2021).
36. Babb, P. L. et al. The *Nephila clavipes* genome highlights the diversity of spider silk genes and their complex expression. *Nat. Genet.* **49**, 895–903 (2017).
37. Kono, N. et al. Darwin's bark spider shares a spider genome with *Caerostris extrusa* but achieves extraordinary silk toughness through gene expression. *Open Biol.* **11**, 210242 (2021).
38. Li, J., Beerntsen, B. T. & James, A. A. Oxidation of 3-hydroxykynurenine to produce xanthommatin for eye pigmentation: a major branch pathway of tryptophan catabolism during pupal development in the yellow fever mosquito, *Aedes aegypti*. *Insect Biochem. Molec.* **29**, 329–338 (1999).
39. Sakudoh, T. et al. Carotenoid silk coloration is controlled by a carotenoid-binding protein, a product of the *Yellow blood* gene. *Proc. Natl Acad. Sci. USA.* **104**, 8941–8946 (2007).
40. Daimon, T. et al. The silkworm *Green b* locus encodes a quercetin 5-O-glucosyltransferase that produces green cocoons with UV-shielding properties. *Proc. Natl Acad. Sci. USA.* **107**, 11471–11476 (2010).
41. Fujiwara, M. et al. Xanthurenic acid is the main pigment of *Trichonephila clavata* gold dragline silk. *Biomolecules* **11**, 563 (2021).
42. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell* **146**, 353–358 (2011).
43. Kuhn, D. E. et al. Experimental validation of miRNA targets. *Methods* **44**, 47–54 (2008).

44. Kruger, J. & Rehmsmeier, M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* **34**, W451–W454 (2006).
45. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
46. Xia, Q. Y., Li, S. & Feng, Q. L. Advances in silkworm studies accelerated by the genome sequencing of *Bombyx mori*. *Annu. Rev. Entomol.* **59**, 513–536 (2014).
47. Ma, Y. et al. A single-cell transcriptomic atlas characterizes the silk-producing organ in the silkworm. *Nat. Commun.* **13**, 3316 (2022).
48. Hu, W., Chen, Y., Lin, Y. & Xia, Q. Developmental and transcriptomic features characterize defects of silk gland growth and silk production in silkworm *naked pupa* mutant. *Insect Biochem. Molec.* **111**, 103175 (2019).
49. Ma, S. Y. et al. Targeted activation of BmCyclinE in *Bombyx mori* using designer TAL effectors. *Insect Sci.* **26**, 1055–1058 (2019).
50. Otsuki, R. et al. Bioengineered silkworms with butterfly cytotoxin-modified silk glands produce sericin cocoons with a utility for a new biomaterial. *Proc. Natl Acad. Sci. USA.* **114**, 6740–6745 (2017).
51. Buchberger, A., Bukau, B. & Sommer, T. Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol. Cell* **40**, 238–252 (2010).
52. Wang, X. et al. Fiber formation and mechanical properties of *Bombyx mori* silk are regulated by vacuolar-type ATPase. *ACS Biomater. Sci. Eng.* **7**, 5532–5540 (2021).
53. Tayebati, S. K., Marucci, G., Santinelli, C., Buccioni, M. & Amenta, F. Choline-containing phospholipids: structure-activity relationships versus therapeutic applications. *Curr. Med. Chem.* **22**, 4328–4340 (2015).
54. Lee, B. et al. dl-Malic acid as a component of alpha-hydroxy acids: effect on 2,4-dinitrochlorobenzene-induced inflammation in atopic dermatitis-like skin lesions in vitro and in vivo. *Immunopharm. Immunot.* **41**, 614–621 (2019).
55. Hardy, J. G., Romer, L. M. & Scheibel, T. R. Polymeric materials based on silk proteins. *Polymer* **49**, 4309–4327 (2008).
56. Wang, X. et al. Chitin and cuticle proteins form the cuticular layer in the spinning duct of silkworm. *Acta Biomater.* **145**, 260–271 (2022).
57. Vollrath, F. Biology of spider silk. *Int. J. Biol. Macromol.* **24**, 81–88 (1999).
58. Hayashi, C. Y. & Lewis, R. V. Molecular architecture and evolution of a modular spider silk protein gene. *Science* **287**, 1477–1479 (2000).
59. Wang, Z. et al. Chromosome-level genome assembly of the black widow spider *Latrodectus elegans* illuminates composition and evolution of venom and silk proteins. *Gigascience* **11**, giac049 (2022).
60. Dicko, C., Vollrath, F. & Kenney, J. M. Spider silk protein refolding is controlled by changing pH. *Biomacromolecules* **5**, 704–710 (2004).
61. Bell, A. L. & Peakall, D. B. Changes in fine structure during silk protein production in the ampullate gland of the spider *Araneus sericatus*. *J. Cell Biol.* **42**, 284–295 (1969).
62. Jin, H. J. & Kaplan, D. L. Mechanism of silk processing in insects and spiders. *Nature* **424**, 1057–1061 (2003).
63. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
64. Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
65. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
66. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
67. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
68. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
69. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA.* **117**, 9451–9457 (2020).
70. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
71. Chen, N. in *Using RepeatMasker to identify repetitive elements in genomic sequences*. Unit 4 10 Ch. 4, (Curr. Protoc. Bioinformatics, 2004).
72. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
73. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
74. She, R. et al. genBlastG: using BLAST searches to build homologous gene models. *Bioinformatics* **27**, 2141–2143 (2011).
75. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
76. Steentoft, C. et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–1488 (2013).
77. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
78. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
79. Wishart, D. S. et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
80. Fahy, E., Sud, M., Cotter, D. & Subramaniam, S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* **35**, W606–W612 (2007).
81. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
82. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
83. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
84. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
85. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
86. Tu, J. C. et al. Homer binds a novel proline-rich motif and links group 1 metabotropic glutamate receptors with IP3 receptors. *Neuron* **21**, 717–726 (1998).
87. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
88. Perteau, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
89. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
90. Tafer, H. & Hofacker, I. L. RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics* **24**, 2657–2663 (2008).
91. Kalvari, I. et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200 (2021).

92. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).
93. Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* **40**, 37–52 (2012).
94. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
95. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
96. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
97. Mount, D. W. Using the basic local alignment search tool (BLAST). *CSH Protoc.* **2007**, pdb top17 (2007).
98. Chen, T. et al. iProX in 2021: connecting proteomics data sharing with big data. *Nucleic Acids Res.* **50**, D1522–D1527 (2022).

Acknowledgements

We thank Daojun Cheng at Southwest University and Yi Zou at Southwest University for critical reading of the manuscript. We thank Yazhou Li at Sichuan Agricultural University for providing help in karyotyping. We also thank BGI, Novogene, Frasergen, OE Biotech, and Berry Genomics companies for providing multiomics sequencing. This work was supported by the National Natural Science Foundation of China [U21A20248], National Natural Science Foundation of China [32000340], and Fundamental Research Funds for the Central Universities [XDJK2019TJ003].

Author contributions

W.H., A.J., Q.X., and Y.W. conceptualized the project. Y.W. supervised the project. W.H., A.J., Z.Z., J.S., T.Y., T.X., Q.L., T.Y., J.Z., and Y.W. collected samples for multiomics sequencing. A.J. designed and implemented the genome assembly pipeline and performed data analysis. W.H. and A.J. drafted the manuscript. F.L., Y.L., Z.J., and Z.X. generated the data and built the database. W.H. and S.M. conducted the

experiments. Y.W. finalized the manuscript with input from all authors. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-36545-6>.

Correspondence and requests for materials should be addressed to Qingyou Xia or Yi Wang.

Peer review information *Nature Communications* thanks Nadia Ayoub, Shuqiang Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023