Multicomponent nature underlies the extraordinary mechanical properties of spider dragline silk

Nobuaki Kono, Hiroyuki Nakamura, Masaru Mori, Yuki Yoshida, Rintaro Ohtoshi, Ali D. Malay, Daniel A. Pedrazzoli Moran, Masaru Tomita, Keiji Numata, and Kazuharu Arakawa

Dragline silk of golden orb-weaver spiders (Nephilinae) is noted for its unsurpassed toughness, combining extraordinary extensibility and tensile strength, suggesting industrial application as a sustainable biopolymer material. To pinpoint the molecular composition of dragline silk and the roles of its constituents in achieving its mechanical properties, we report a multiteomics approach, combining high-quality genome sequencing and assembly, silk gland transcriptomics, and dragline silk proteomics of four Nephilinae spiders. We observed the consistent presence of the MaSp3B spidroin unique to this subfamily as well as several nonspidroin SpiCE proteins. Artificial synthesis and the combination of these components in vitro showed that the multicomponent nature of dragline silk, including MaSp3B and SpiCE, along with MaSp1 and MaSp2, is essential to realize the mechanical properties of spider dragline silk.

Spider silk is a typical natural high-performance structural protein and has potential for numerous applications as a protein biopolymer with biodegradability and biocompatibility (1). Spider silk has a tensile strength superior to that of steel, yet it is highly elastic, showing greater toughness than aramid fibers such as Kevlar (2); thus, it has received interest for use in industrial applications (1, 3). Orb-weaver spiders, especially those belonging to the family Araneidae, are often used as models in natural spider silk research. The average mechanical properties of their dragline silks reach ~1 GPa for breaking strength, 30% for breaking strain, and 130 MJ/m² for toughness (4). Numerous works have reported the use of recombinant protein and artificial fiber spinning to produce artificial spider silk in genetically optimized organisms (5–10), but it remains challenging to fully reproduce and rival the mechanical properties of natural silks (6, 11–13). The difficulties are multifactorial; the unusually large protein size and the repetitive nature of its sequence are inherent challenges for synthesis, and the natural conditions of spinning are only beginning to be fully uncovered (13–15). However, one primary reason that full reproduction has not been successful is that the previous recombinant approaches employed only MaSp1, only MaSp2, or at most a combination of the two (11, 16–18). In fact, it is often suggested that dragline silk is composed primarily of two components, MaSp1 and MaSp2 (5, 19–23). However, recent proteome analyses suggest the existence of additional components in spider silks, such as a cysteine-rich protein (CRP) in black widows (24–26). Genome and transcriptome analyses have identified many MaSp families (27, 28), and in the genus Araneus, it is known that dragline silk contains nearly equal amounts of MaSp3 and MaSp1/2 (28). Furthermore, low-molecular-weight (LMW) nonspidroin proteins, such as silkworm silk constituting element (SpiCE), have been found by transcriptomic and proteomic analyses. SpiCE is a protein of unknown function that is commonly highly expressed in the silk gland and in spider silk. It is becoming apparent that dragline silk is a complex multicomponent material containing much more than MaSp1 and MaSp2.

To pinpoint the protein constituents of dragline silk through quantitative proteomics, a high-quality reference genome and full-length coding sequence annotation are essential to allow the correct and comprehensive identification of the proteins corresponding to the detected peptide fragments. Since spider fibroin genes are extremely long (~10 kbp) and consist almost entirely of repeat sequences (29, 30), genome sequencing using PCR-free long reads is critical, and in order to eliminate false-positive annotations, predicted coding sequences need to be confirmed on the bases of conservation in closely related species and actual mRNA expression in the silk gland. Hence, we took a multiteomics approach to quantitatively identify the dragline protein constituents and the genomes of four golden silk orb-weavers (subfamily Nephilinae): Trichonephila clavipes, Trichonephila clavipes, Trichonephila inaurata madagascariensis, and Nephila pilipes. These spiders are reported to produce high-performance dragline silk with average toughness values of 169, 131, 285, and 292 MJ/m³, respectively (SI Appendix, Fig. S1). T. clavipes genome data have already been reported (27), but we chose to construct ab initio assemblies, including for this species, since the existing genome is based on PCR-amplified sequencing for fibroins, and some fibroin gene sequences remain incomplete. Moreover, the existing T. clavipes assembly is suspected to be substantially contaminated, as the entirety of the longest scaffold has been identified as bacterial in origin (31).

**Significance**

Artificial synthesis of spider silk has been actively pursued. However, until now, the natural mechanical properties of spider silk have been largely un reproducible. We thoroughly investigated the genomes and transcripts of four related species of orb-weaver spiders as well as the proteins in their silk threads. Then, in addition to spidroin, we found several low-molecular-weight proteins in common. Interestingly, the low-molecular-weight protein component of spider dragline silk doubled the tensile strength of artificial silk–based material. This discovery will greatly advance the industry and research on the use of protein-based materials.

Author contributions: N.K. and K.A. designed research; N.K., H.N., M.M., R.O., A.D.M., D.A.P.M., K.N., and K.A. performed research; N.K., M.T., and K.A. contributed new reagents/analytic tools; N.K., H.N., M.M., Y.Y., A.D.M., K.N., and K.A. analyzed data; and N.K. and K.A. wrote the paper.

Competing interest statement: H.N., R.O., and D.A.P.M. are employees of Spiber Inc., a venture company selling artificial spider silk products. However, all study design decisions were made by N.K. and K.A. of Keio University, and Spiber Inc. had no role in the study design.

This article is a PNAS Direct Submission. Published under the PNAS license.

1To whom correspondence may be addressed. Email: gaou@fsk.keio.ac.jp.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2107065118/-/DCSupplemental.

Published July 26, 2021.
Results

Draft Genomes of Four Nephilinae Spiders. Spider genomes are large and complex, and de novo sequencing is challenging. We extracted genomic DNA from dissected leg and cephalothorax tissue of adult female individuals and conducted hybrid sequencing with short- and long-read sequence technologies (32) (see Materials and Methods). The 10X Genomics linked sequencing produced 702.57 M reads (106.09 Gb), 1,090 M reads (164.60 Gb), 1,010.42 M reads (152.57 Gb), and 900.42 M reads (135.96 Gb) from T. clavata, T. clavipes, T. inaurata madagascariensis, and N. pilipes, respectively (SI Appendix, Dataset S1). PacBio sequencing for the T. clavata genome yielded 74.81 Gb of data with an average length of 5.24 kb and an N50 of 11.1 kb (SI Appendix, Dataset S1). Furthermore, 2.19 M reads (11.03 Gb), 4.70 M reads (51.43 Gb), 5.88 M reads (24.33 Gb), and 2.92 M reads (8.48 Gb) were generated by Nanopore sequencing of gDNA from T. clavata, T. clavipes, T. inaurata madagascariensis, and N. pilipes, respectively (SI Appendix, Dataset S1). Total coverage was ~80x. Obtained sequence reads were assembled, polished, corrected, and contaminant eliminated (see Materials and Methods); the resulting the number of scaffolds and N50 length in the T. clavata, T. clavipes, T. inaurata madagascariensis, and N. pilipes draft genomes were 39,792 (N50 of 112,957 base pairs [bp]), 21,438 (N50 of 73,925 bp), 28,263 (N50 of 231,352 bp), and 132,899 (N50 of 292,147 bp), respectively (Table 1 and SI Appendix, Figs. S2 and S3). The completeness of each assembled result was assessed by BUSCO (33). The completeness scores were 90.6%, 93.0%, 89.8%, and 92.9% for the eukaryote lineage, indicating the high quality of the obtained draft genomes. The tRNAs and ribosomal RNAs were annotated using tRNAscan-SE version 2.0 (34) and Barrnap (https://github.com/tseemann/barrnap) with default parameters. In total, 8,008, 11,433, 8,679, and 1,837 tRNAs and ribosomal RNAs were identified in T. clavata, T. clavipes, T. inaurata madagascariensis, and N. pilipes, respectively (SI Appendix, Fig. S4 and Supplementary Text).

The number of protein-coding genes predicted by BRAKER was initially 70,418, 517,373, 51,581, and 72,271. Redundant genes were eliminated by CD-HIT-EST (35) clustering with a nucleotide identity of 97% (Fig. S5 for details). Seven orthologous groups (MaSp: major ampullate spidroin, MiSp: minor ampullate spidroin, AcSp: aciniform spidroin, Flag: flagelliform spidroin, AgSp: aggregate spidroin, PySp: pyrimid spidroin, and CySp: cylindrical/tubuliform spidroin) are known as the typical spidroins in Araneoid orb-weaving spiders. We obtained all seven groups contain 2 to 3 families or subfamilies from four Nephilinae draft genomes, five of which were full length (MaSp, MiSp, AcSp, CySp, and PySp) (see Fig. 1 and SI Appendix, Fig. S5 for details).

The spidroin genes were well conserved among Nephilinae spiders and followed the established phylogenetic relationship (SI Appendix, Figs. S6 and S7). Interestingly, the spidroin catalog revealed a group of major ampullate spidroins distinct from MaSp families 1 and 2 that form a unique clade within Nephilinae. The N-terminal phylogeny suggests that this group forms a clade with the MaSp family 3 sequence of Araneidae (SI Appendix, Fig. S8), reflected by the characteristic DGGR motif, yet the N terminus conservation is sufficiently low to define this as a unique group; thus, we named this gene MaSp3B based on a spidroin nomenclature (SI Appendix, Fig. S9 and Supplementary Text). Previously identified partial sequences, Sp-907 and Sp-74867 (27), revised as MaSp3_A and MaSp3_B (39), belong to this group.

Expression Profile of Spidroin and SpiCE. With this high-quality reference of full-length spidroins, the existence of a large proportion of MaSp3B in the forcibly reeled dragline silk of T. clavata was immediately observable as a distinct band in the high-molecular-weight (HMW) fraction on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 2A). The abundance ratios of the proteins were then quantified by intensity-based absolute quantification (iBAQ) (40) with shotgun proteomics of selectively reeled dragline silks (SI Appendix, Fig. S10). These analyses showed consistent composition not only within the replicates but also among the four species analyzed (Fig. 2B). Among the spidroins, only MaSp family members were detected; MaSp1, 2, and 3 accounted for ~90 wt% of the total in any spider species (Fig. 2B). Transcriptome analysis of silk glands confirmed the specific expression of MaSp3B in the major ampullate silk gland, similar to that of MaSp1/2 (Fig. 2C and SI Appendix, Fig. S11 and Datasets S2–S4). Both of these omics analyses also clearly demonstrated the consistent presence of nonspidroins, which was also confirmed by SDS-PAGE of the LMW fraction (Fig. 2C). Across the four Nephilinae species, these genes are consistently expressed as mRNAs at a level equivalent to that of MaSp genes in the major ampullate silk gland and at lower levels in the minor ampullate silk gland (Fig. 2C and SI Appendix, Fig. S11). Three such proteins highly expressed in both silk (as protein) and the silk gland (as mRNA) were annotated as SpiCEs, following the identification of nonspidroin constituents in A. ventricosus dragline silk (28). These SpiCEs are conserved in Nephilinae (SI Appendix, Fig. S12), but their sequences are considerably nonhomologous to those of A. ventricosus; therefore, we named them SpiCE-NMa1, 2, and 4 (SpiCE Nephilinae Major Ampululate). Among them, SpiCE-NMa2 and SpiCE-NMa4 were CRP type with high cysteine content. The most abundant nonspidroin protein among the dragline silks was SpiCE-NMa1 at ~1 wt% (Fig. 2B).
SpiCE Doubles the Tensile Strength of Artificial Silk–Based Material.
To investigate the role of these proteins in the expression of the mechanical properties of Nephilinae dragline silks, we first produced composite films of recombinant MaSp family proteins. Recombinant MaSp was synthesized based on shortened amino acid sequences of ∼50 kDa (SI Appendix, Figs. S13 and S14). The transparency of silk-based films did not drastically decrease even when the composite films were prepared from two or three spider silks, suggesting that heterogeneous interactions among MaSp family proteins did not induce specific aggregation (Fig. 3B). Wide-angle X-ray scattering (WAXS) profiles demonstrated (210) and (040) reflections; these are indicative of the antiparallel β-sheet structure (41, 42). Thus, MaSp1, MaSp2, and MaSp3 predominantly form β-sheet structures, similar to polyalanine (Fig. 3B). The composite films of MaSpS showed similar β-sheet structures, even with different protein compositions, that is, with MaSp1 used alone, with the mix of three families, and with the natural composition (Fig. 3B). We further tested the impact of SpiCE inclusion by producing composite films of recombinant MaSp and SpiCE-NMa1, which was ∼5 times more abundant than other SpiCEs (Fig. 2B). Recombinant SpiCE-NMa1 was added in the range of 2%–5% based on proteome quantification to produce composite silk films with recombinant MaSp and SpiCE-NMa1. Composite silk fibers with 2% SpiCE-NMa1 showed an increase in elongation at break as well as a decrease in tensile strength (Fig. 4C). These changes in mechanical properties indicate that SpiCE-NMa1 disturbed the strain-induced crystallization of silk molecules and plasticized the amorphous region. This can be because the amorphous region of silk fibers is relatively more oriented and aligned along the fiber axis than that of silk film (43), and hence, SpiCE-NMa1 cannot interact with silk molecules in the amorphous region of silk fibers, in contrast to the amorphous silk in film. A small percentage (≤2%) of SpiCE-NMa1, which interacts with the amorphous silk, in artificial silks could potentially be the molecular tool to modify and tune their mechanical properties.

Discussion
MaSp3 is widely conserved in the large-web-forming group of the family Araneidae and has been found by proteomics to be a major component of dragline silk in A. ventricosus. The conservation of MaSp3 has been thought to be limited to closely related species of the genus Araneus (28), and it was believed that Nephilinae spiders do not have MaSp3 even though they also build large webs. However, our high-quality genome analyses demonstrate that MaSp3B, a distinct subtype of MaSp3 in the genus Trichonephila, is conserved in the genera Trichonephila and Nephila. Further, our analyses suggest that MaSp3B is one of the major components of dragline silk and may associate with other MaSp family proteins, such as MaSp4, for the purpose of producing araneoid spider silk, which has particularly strong mechanical properties.

Table 1. Assembly and annotation statistics of four spider draft genomes (T. clavata, T. clavipes, T. inaurata madagascariensis, and N. pilipes)

| Statistics                  | T. clavata | T. clavipes | T. inaurata madagascariensis | N. pilipes |
|-----------------------------|------------|-------------|-----------------------------|------------|
| Genotype                    | v4         | v4          | v4                          | v4         |
| Estimated genome size (Gbp)*| 2.46       | 2.72        | 2.32                        | 2.12       |
| Assembly size (Gbp)         | 2.50       | 2.87        | 2.51                        | 2.69       |
| Scaffold number             | 39,792     | 21,438      | 28,263                      | 132,899    |
| Average scaffold length (bp)| 62,773     | 134,077     | 88,704                      | 20,274     |
| Longest scaffold length (bp)| 1,797,358  | 4,645,559   | 2,667,260                   | 4,220,928  |
| Shortest scaffold length (bp)| 1,010      | 150         | 1,150                       | 451        |
| N50 (bp) (# of scaffolds in N50) | 112,957 (#5,596) | 738,925 (#1,112) | 231,352 (#2,910) | 292,147 (#1,975) |
| N90 (bp) (# of scaffolds in N90) | 22,952 (#25,683) | 104,003 (#4,898) | 40,466 (#12,765) | 6,691 (#34,939) |
| BUSCO version 4.0.5 completeness (%)† | C:90.6%[S:85.5%,D:5.1%], F:5.5%,M:3.9%,n:255 | C:93.0%[S:92.2%,D:0.8%], F:2.7%,M:4.3%,n:255 | C:89.8%[S:86.7%,D:3.1%], F:2.7%,M:7.5%,n:255 | C:92.9%[S:89.8%,D:3.1%], F:3.5%,M:3.6%,n:255 |
| Number of protein-coding genes | 18,822     | 45,304      | 16,899                      | 17,127     |
| tRNAs                       | 8,008      | 11,433      | 8,679                       | 1,837      |
| rRNAs                       | 248        | 106         | 128                         | 167        |
| BUSCO version 4.0.5 completeness (%)† | C:91.4%[S:89.0%,D:2.4%], F:5.9%,M:2.7%,n:255 | C:87.5%[S:85.5%,D:2.0%], F:8.6%,M:3.9%,n:255 | C:91.0%[S:89.8%,D:1.2%], F:4.3%,M:4.7%,n:255 | C:93.4%[S:91.0%,D:2.4%], F:5.1%,M:1.5%,n:255 |

*Genome size was estimated based on k-mer (GenomeScope).
†BUSCO analysis was based on eukaryota lineage of protein-coding genes.

Kono et al.
Multicomponent nature underlies the extraordinary mechanical properties of spider dragline silk
https://doi.org/10.1073/pnas.2107065118
the multicomponent nature of the mechanical property of dragline silk partly explains the robustness of the compositional percentages of spidroins previously reported for *N. pilipes* (44). Our proteome analysis showed compositional variability in line with the previous report, but the existence of MaSp3 and SpiCE prescribes the performance in these silks. Although the specific role of MaSp3 is still unclear, the key residues contributing to the N-terminal domain dimerization, such as D40, K65, E79, and E119, are conserved in MaSp1-3 (SI Appendix, Fig. S18). The N-terminal domain of MaSp changes its structure in a pH-dependent manner to form dimers, which are believed to be responsible for stable filament formation in ducts (45, 46). Each conserved amino acid site is associated with this dimerization (47). Therefore, it is likely that MaSp3, like other MaSp1 and MaSp2, functions in pH-sensitive dimerization of the N-terminal domain and contributes to silk formation.

*N. pilipes* shows higher composition compared to the other three species, which confirms the higher proline content documented in this species (48). Therefore, we analyzed the relationship between the amino acid composition of dragline silk and its mechanical properties. We reconfirmed the correlation between the mechanical properties and amino acid composition, such as glycine or proline, discussed in previous studies (48) (SI Appendix, Fig. S19). This result supports the previous claim that the evolutionary bias in spidroin may differ between the terminal region and the repeat region and that this variation in the amino acid composition may have contributed to some of the variability in physical properties among species.

The contribution of the LMW weight component of the spider dragline silk, SpiCE, to its mechanical properties was also confirmed in vitro, in that inclusion of SpiCE doubled the tensile strength of the artificial film, but the exact mechanism responsible

---

**Fig. 1.** Catalog of spidroins in four Nephilinae spiders. This summary panel shows the spidroin sequence characteristics and structures obtained from four Nephilinae draft genomes. The icons in the first column represent spidroin groups. The second column represents the motif variety in the domains, as follows: β-sheet [An, (GA)n, and AS], blue; β-turn (XQQ, GPGQQ, and GPGXX), yellow; spacer (SSX and TTX), green; 310 helix (GGX), orange; and MaSp motif (DGGR, GGYGGRF, and GGYGL), red. The spidroin structure columns show the N/C-terminal (yellow and blue box) and repeat domains, and each structure is drawn to scale. The number, width, and color of stripes reflect the number, size, and motif identity of the repeats. Assembly gaps are represented by “N” (SI Appendix, Fig. S5 for details). *T. clavata, T. clavipes, and N. pilipes* images credit: Akio Tanikawa (University of Tokyo, Tokyo, Japan).
for this phenomenon is still an area of future work. Accessory components of structural proteins are often used to stabilize the structure; for example, elastin, another known structural protein, binds with fibulin proteins to produce elastic fibers (49). Likewise, while a self-assembly system involving pH-induced fibrillation of the N-terminal domain and liquid-liquid phase separation (LLPS) of the C-terminal domain (14) is known to participate in spider silk fibril formation, SpiCE could serve to support the interaction between repeat regions rather than the terminal domains. According to SI Appendix, Fig. S17, the β-sheet structure and crystallinity of the films were not changed significantly. Overall, the SpiCE-NMa1 preferentially interacts with silk molecules in the amorphous region rather than those in the crystalline state, namely, β-sheet crystal structures. The intermolecular interactions of amorphous silk molecules via SpiCE-NMa1 decrease the molecular weight between crosslinking points, resulting in the enhancement of the strength and modulus of the composite films. Further observation of the interaction between SpiCE and spidroin with detailed biophysical analysis, such as NMR spectroscopy or atomic force microscopy scanning, is desirable to elucidate the functions of SpiCE. The evolutionary origin of SpiCE also requires further analyses, as these proteins seem to be specific to a narrow range of spider clades. Unlike MaSp3, the conservation of SpiCE has been confirmed only within very closely related groups (genus or subfamily at most), and CRP seems to be used instead of SpiCE in black widow spiders (26). The MaSps retain homology across different spider clades, but there is no sequence homology among SpiCEs, suggesting that MaSps and SpiCEs did not coevolve. Multiple lineage-specific duplications of MaSp paralogs are observed even among orb-weaving spiders (27, 28, 50–52), and the evolution of SpiCE seems to follow this lineage-specific evolutionary pattern.

Considering the contributions of MaSp3 and SpiCE to the mechanical performance of spider silk, it is essential to consider silk as not just a product of MaSp1 and MaSp2 but rather a multicomponent material that also includes MaSp3 and SpiCE. The contribution of multiple paralogs of MaSp1 and MaSp2, which is not covered in this work, may also contribute to the specific ecological adaptation of Nephilinae spiders. The availability of four closely related Nephilinae genomes and their gland transcriptomes as well as the dragline silk proteomes presented in this work will contribute to the research in this direction, as well as other phylogenomic work of arachnids.

Materials and Methods

Spider Sample and Rearing Environment. T. clavata and N. pilipes samples were collected in Japan, and T. inaurata madagascariensis was collected in Madagascar. T. clavipes samples were purchased from Spider Pharm Inc. All the spiders used in this study were adult females. The identification of spider
specimens was performed based on morphological characteristics and sequence identification of cytochrome oxidase subunit 1 (COX1) in the barcode of Life Data System (BOLD; http://barcodinglife.org). The spiders used for dissection or silk reeling were kept in a PET cup (129 × 97 mm, PAPM340; RISUPACK CO., LTD) in a climate-controlled laboratory with 25.1 °C, 57.8% humidity, and 12 to 12 h light–dark cycles. The spiders were fed every two days with crickets purchased from Mito-koriogi Farm, and water was given daily by softly spraying inside the cup. The legs and cephalothoraxes were used for gDNA extraction, and abdomens and silk glands were used for RNA extraction. All of the tissues were dissected from adult female spiders. Dissected samples were immersed in liquid nitrogen (LN2) and stored at −80 °C until the next process.

HMW Genomic DNA Extraction. Spider genomic DNA was extracted from dissected leg and cephalothorax tissue from adult female individuals using Genomic-tip 20G (QIAGEN) following the manufacturer’s protocol. To obtain HMW genomic DNA, all steps were conducted as gently as possible. The legs and cephalothorax were separated from flash-frozen spider specimens (−80 °C) and then homogenized with a BioMasher II (Funakoshi) and mixed with 2 mL of Buffer G2 (QIAGEN) containing 200 μg/mL RNase A. After the addition of 50 μL Proteinase K (20 mg/mL), the lysate was incubated at 50 °C for 12 h on a shaker (300 rpm) and centrifuged at 5,000 μg/mL RNase A. After the addition of chloroform, and the upper aqueous phase containing RNA was purified automatically with an RNeasy Plus Mini Kit (QIAGEN) in a QIAcube instrument (QIAGEN). The RNA was quantified and checked for quality using a Qubit Broad Range RNA assay (Life Technologies) and NanoDrop 2000 (Thermo Fisher Scientific). The integrity was estimated by electrophoresis using a TapeStation 2200 instrument with RNA ScreenTape (Agilent Technologies). mRNA selection was performed using oligo d(T)7. For direct RNA sequencing, mRNA was prepared using the NucleoTrap mRNA Mini Kit (Clontech) from 500 ng of total RNA. cDNA was synthesized from mRNA isolated from 100 μg of total RNA by NEBNext Oligo d(T)12 beads (skipping the Tris buffer wash step). The first- and second-strand cDNA were synthesized using ProtoScript II Reverse Transcriptase and NEBNext Second Strand Synthesis Enzyme Mix.

Library Preparation and Sequencing. For cDNA, sequencing was carried out on the abdomens of T. clavata (n = 14), T. clavipes (n = 2), T. inaurata mada-gascariensis (n = 1), and N. pilipes (n = 14), in the major ampullate silk glands of T. clavata (n = 6), T. clavipes (n = 7), and N. pilipes (n = 5), and in the minor ampullate silk glands of T. clavata (n = 6), T. clavipes (n = 8), and N. pilipes (n = 6). All cDNA libraries were constructed according to the standard protocol of the NEBNext Ultra RNA Library Prep Kit for Illumina (New England BioLabs). The synthesized double-stranded cDNA was end-repaired using NEBNext End Prep Enzyme Mix before ligation with NEBNext Adaptor for Illumina. After USER enzyme treatment, cDNA was amplified by PCR with the following conditions: 20 μL cDNA, 2.5 μL Index Primer, 2.5 μL Universal PCR Primer, 25 μL NEBNext Q5 Hot Start HiFi PCR Master Mix 2X; 98 °C for 30 s and 12 cycles each of 98 °C for 10 s, 65 °C for 35 s, 63 °C for 5 min. cDNA sequencing was conducted with a NextSeq 500 instrument (Illumina) using 150-bp paired-end reads with a NextSeq 500 High Output Kit (300 cycles).

The synthetic long-read sequencing using 10X Genomics was carried out in all four Nephilinae spiders. Purified genomic DNA fragments longer than 60 kb (10 ng) were used to prepare the 10X GemCode library with the Chromium
instrument and Genome Reagent Kit version 2 (10X Genomics) following the manufacturer’s protocol. The 10X GemCode library sequencing was conducted with a NextSeq 500 instrument (Illumina) using 150-bp paired-end reads with a NextSeq 500 High Output Kit (300 cycles).

For PacBio sequencing, the library preparation was performed with gDNA fragments of more than 10 kb in size (selected by BluePippin) and sequenced in PacBio RSII using P6-C4 chemistry. PacBio library construction and sequencing were conducted at Genomic Information Research Center, Osaka University.

The gDNAs of *T. clavata*, *T. clavipes*, *T. inaurata madagascariensis*, and *N. pilipes* were also sequenced by Nanopore technology. Library preparation was completed following the 1D library protocol (SQK-LSK109, Oxford Nanopore Technologies). The quality of the libraries was estimated by TapeStation 2200 with D1000 Screen Tape (Agilent Technologies). Direct RNA sequencing was performed for *T. clavata*, *T. clavipes*, and *N. pilipes*. Libraries constructed from 500 ng of total RNA were prepared following the direct RNA sequencing protocol (SQK-RNA001, Oxford Nanopore Technologies). Sequencing was performed using a GridION instrument with Spot On Flow Cell Rev D (FLO-MIN106D, Oxford Nanopore Technologies). The base calls were performed by Guppy basecalling software (version 3.2.10+aabd4ec).

**Genome Assembly.** We adopted a hybrid assembly strategy for spider draft genomes that combined the various reads produced by each sequencing technology. Natural long reads were produced by direct sequencing of HMW genomic DNA with Nanopore or PacBio technology. The synthetic long reads were generated using a combination of Illumina and 10X Genomics technologies. The details of the assembly methods used for each of the four spider genomes are described in SI Appendix, Supplementary Methods.

**Contaminant Elimination and Genome Size Estimation.** To detect possible contaminants in the genome, we submitted each genome assembly to BlobTools analysis (54). The genome sequence was submitted to a Diamond (55) BLASTX search (–sensitive–max-target-seqs. 1–evaluate 1e-25) against the UniProt Reference proteome database (downloaded November 2018). Each contig was classified into a lineage with BlobTools taxify. We then mapped DNA-Seq reads to the genome by BWA MEM and conducted SAM to BAM conversion.

Fig. 4. SpiCEd film and silk. (A) Composite films mixing MaSp2 and SpiCE-NMa1. The stress-strain curve represents the typical tensile engineering stress-strain in MaSp2 and SpiCE composite films under different SpiCE concentrations (0, 1, 3, 5 wt%). The transparency (%) of these films is shown as T in each picture (SI Appendix, Figs. S14–S16 for details). (B) Recombinant spider silks mixing MaSp2 and SpiCE-NMa1. SEM images show the cylindrical and smooth surface and cross-section of each recombinant spider silk. (C) Stress-strain curve of the recombinant spider silks mixing MaSp2 and SpiCE-NMa1.
with SAMTools. These taxonomic annotation and coverage data were visualized by BlobTools following the protocol. Contigs that were classified as from bacteria, plasmids, or fungal DNA were filtered out before assembly, and the filtered genome assembly was assayed by Benchmarking Universal Single-Copy Orthologs (BUSCO) version 4.0.5 (33) (eukaryote lineage) to validate genome completeness. Genome heterozygosity, repeat content, and size were estimated based on the k-mer distribution with Jellyfish and GenomeScope (56).

**Gene Prediction and Annotation.** Gene prediction was performed using a gene model created by cDNA-seq mapping data with HISAT2 and BRAKER (57, 58). cDNA-seq reads were mapped to the genome with HISAT2 (version 2.1.0) and the resulting SAM file was converted, sorted, and indexed by SAMtools (version 1.4). Repeat sequences were detected by RepeatModeler (1.0.11) and soft-masked by RepeatMasker (version 4.0.7). The soft-masked genome was submitted to gene prediction with BRAKER (version 2.1.4, softmasking–gfif3). The amino acid sequences were submitted to BLASTP or DIAMOND BLASTP searches against public databases (UniProt TrEMBL, UniProt Swiss-Prot). Redundant genes were eliminated by CD-HIT-EST (35) clustering with a nucleotide identity of 97%. To obtain a functional gene set, we removed the genes with an expression level of less than 0.1 and unannotated genes. BUSCO (version 4.0.5) was used to determine the quality of our functional gene set using the eukaryote lineage. The rRNAs were annotated using TRINAscan-SE version 2.0 (34) with default parameters. Ribosomal RNAs (rRNAs) were predicted using Barnmap (https://github.com/tseemann/barnmap).

**Spidroin Gene Catalog.** We used a hybrid method that combines short- and long-read sequencing to catalog the spidroin diversity in four Nephilinae spiders. The short reads were obtained from Illumina sequencing of cDNA or gDNA libraries, and the long reads were obtained by Nanopore or PacBio sequencing of the gDNA library. The typically used de Bruijn graph algorithm is not suitable for the assembly of long repeat domains, so the SMoC Spidroin Motif Collection algorithm (28), developed based on the OLC (Overlap-Layout-Consensus) algorithm, was used. The SMoC algorithm collects as many patterns of repeat motifs as possible, scaffolds these repeat motifs based on long reads, and then provides the full-length spidroin gene sequence. SMoC first finds the N-terminal region (nonrepetitive region) contigs from the assembled contigs with a homology search. These terminal fragments were used as seeds for screening of the short reads obtained by an illumina sequencer harboring an exact match of extremely large k-mers (~100) up to the 5′-end. The obtained short reads were aligned on the seed sequence to construct a PWM (position weight matrix) on the 3′-side of the matching k-mers, and the seed sequence was extended based on the PWM until there was a split in the graph using the resulting PWM until there was a split in the graph using the comprehensive collection of repeat motif patterns, scaffolding was performed based on the k-mer distribution with Jellyfish and GenomeScope (56).

**Phylogenetic Tree of Spidroin.** The phylogenetic tree in SI Appendix, Fig. S5 was constructed based on a core orthology gene set obtained from BUSCO ortho- podaodb9 (33) using four Nephilinae draft genomes (this study) and the A. ventricosus draft genome (Ave. 3.0) (28). The phylogenetic tree of spidroin (SI Appendix, Fig. S7) was constructed by FastTree (version 2.1.10) (59) as an approximately maximum-likelihood phylogenetic tree from aligned and trimmed N-terminal sequences (150 residues from the start codon) with MAFFFT (version 7.273) (60) and trimal (1.2rev93) (61). The accession numbers for known spidroin sequences are described in SI Appendix, Supplementary Methods.

**Gene Expression Analysis.** Gene expression profiling was conducted with multiple biological replicates from mRNA extracted from the abdomen or silk glands. The gene expression levels were quantified and normalized as transcripts per million by mapping processed reads to our assembled draft genome references with Kallisto version 0.42.1 (62) (SI Appendix, Datasets 52–54).

**Silk Collection.** The natural spider silks were sampled directly from adult female T. clavata restrained using two pieces of sponge and locked with rubber bands. Silk reeling from the spider was performed at a constant speed (1.28 m/min for 1 h) with a reeling machine developed by Spiber Inc. with six biological replicates and three technical replicates each. Since the spinnerets are very close together and many silks can be spun at the same time, simple silk reeling may involve minor amouple silk as well. Therefore, to identify the proteins that are significantly present in minor amouple silk and remove them as background, we separated the dragline and minor amouple silks clearly from their spinnerets using a stereomicroscope (Leica S4E, Leica Microsystems GmbH) (SI Appendix, Fig. S10).

**SDS-PAGE Analysis.** SDS-PAGE analysis of dragline silk was conducted at IDEA Consultants, Inc. The dragline silk reeled from T. clavata was dissolved in 2 mL ofionic liquid (1-buty1-3-methylimidazoliazole) acetate) per mg. After vortexing for 2 min, the sample was incubated at 100 °C for 15 min. The ionic liquid solution was dialyzed into lysis buffer (6 M urea, 2 M thiourea, 2% CHAPS, 1% DTT) using a 3k MWCO cellulose membrane (Amicon Ultra, Millipore).

The HMW fraction of 0.33 mg dragline silk was dissolved in the ionic liquid and filtered through a 50k MWCO membrane (Amicon Ultra, Millipore), collected in 100 μL of lysis buffer, and incubated at 95 °C for 5 min. Then, 2.5 μL and 5 μL samples were applied to a 4% SDS-polyacrylamide gel. Electrophoresis was performed at 20 mA for 80 min in running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS), and a HiMark Unstained Standard (LC5688, Life Technologies) was used as a molecular marker. After electrophoresis, the SDS gel was fixed in fixing solution (30% MeOH, 5% acetic acid), stained with SYPRO Ruby overnight, and washed with a wash solution (10% MeOH, 7% acetic acid). Silver staining was conducted after fixation with 50% MeOH and sensitization with 0.005% sodium thiosulfate (Fig. 2A).

A solution of 0.37 mg dragline silk dissolved in 80 μL Tris HCl (pH 8.6), 1% Triton-X, and 2% SDS was vortexed, incubated for 1 h on ice, and freeze-dried. After resuspension, the total volume was applied to a 12.5% SDS-polyacrylamide gel. Electrophoresis was performed at 20 mA for 80 min, with Broad Range Protein Molecular Weight Markers (Y8491, Promega) used as molecular markers. After electrophoresis, the SDS gel was fixed in fixing solution (30% MeOH, 5% acetic acid), stained with SYPRO Ruby overnight, and washed with a wash solution (10% MeOH, 7% acetic acid). Silver staining was conducted after fixation with 50% MeOH and sensitization with 0.005% sodium thiosulfate (Fig. 2A).

**Liquid Chromatography Mass Spectrometry Analysis.** Approximately 1.0 mg of dragline silk was used for the liquid chromatography mass spectrometry (LC-MS) analysis. Using the iBAQ score, we identified the proteins specifically expressed in the minor amouple silk based on significance (false discovery rate < 0.01) and removed them from the result of the dragline silk proteome analysis as contaminants. The percentage of each protein in silk was calculated based on the iBAQ values multiplied by the amino acid length. The details of the LC-MS analysis used for each spider silk are described in SI Appendix, Supplementary Methods.

**Recombinant Protein Expression and Purification.** Sequence-optimized recombinant T. clavata MaSp family and Spic-NMa1 proteins were used for the composite film and silk. Targeting a size of ~50 kDa, we reduced the number of repeat units to 14 for MaSp1, 9 for MaSp2, and 8 for MaSp3 during the keeping of the N- and C-terminal domains (SI Appendix, Fig. S13). The recombinant spidroin domains are composed of a 6x His tag (MHHHHH), a linker (SsgsS), an HRV 3c protease recognition site (LEVLFQGP), an N-terminal domain,repeat units, and a C-terminal domain. The Spic-NMa1 gene is also composed of a 6x His tag, a linker (SsgsS), an HRV 3c protease recognition site, and a T. clavata comp1999 protein sequence. Each signal peptide was removed based on the tBLR (DE3). Nucleotide sequences were adjusted to the codon usage of Escherichia coli. Fragmented sequences were chemically synthesized by Fasmac Co., Ltd. The synthesized sequences were assembled with overlap extension PCR (64). Assembled sequences were cloned into pet-22b(+), and transformed into E. coli BLR (DE3). A 100-mL bacterial culture (5.0 g/L glucose, 4.0 g/L KH2PO4, 6.0 g/L yeast extract, and 0.1 g/L ampicillin) was grown at 30 °C to an optical density at 600nm (OD600) of 5.0 and used for inoculation. The cells were cultured in a 10 L jar fermenter (Takasugi Seisakusho Co., Ltd.) containing 5.7 L of medium (12.0 g/L glucose, 9.0 g/L KH2PO4, 15 g/L yeast extract, 0.04 g/L FeSO4·7H2O, 0.04 g/MnSO4·SH2O, 0.04 g/CAcL2·H2O, GD-113, Nof Corporation) at 37 °C. The initial OD600 was 0.05. After the initial glucose was consumed, a feeding solution containing 65.9 wt/vol% glucose was pumped into the fermenter at a feeding rate of 50 mL/h. The dissolved oxygen concentration was kept at 20% air saturation. The pH was kept at pH 6.9. After 24 h, the protein expression was induced by adding isopropyl-β-thiogalactoside (IPTG) to a final concentration of 0.1 mM. The cells were harvested by centrifugation at 11,000 × g for 15 min at 24 h after induction. For the recombinant spidroins, the harvested cells were washed out with 20 mM Tris-HCl (pH 7.4) and centrifuged at 11,000 × g, and the supernatant was discarded. The cells were resuspended in 20 mM Tris HCl (pH 7.4), 100 mM NaCl (pH 8.0) with 0.9 g/L NaCl as the weight (g), 82 μg of lysozyme per wet cell weight (g), and 0.2 mL of phenylmethylsulfonyl fluoride (PMSF) per wet cell weight (g). Cells were agitated overnight at 37 °C. The cells were resuspended in buffer solution (50 mM Tris-HCl, 100 mM NaCl [pH 8.0]) with 3 wt/vol% SDS, and the cell
Ethanol precipitation was performed at room temperature. The mixture was centrifuged at 11,000 rpm for 10 min. The precipitated ethanol was washed with reverse-osmosis purified (RO) water three times, followed by lyophilization. For SpiCE-NA1, the harvested cells were suspended in sodium phosphate buffer (50 mM sodium phosphate, 300 mM NaCl [pH 7]) with 7.5 M urea, PMSF, and Triton-X100 (pH 7.4). The dialyzed sample was loaded onto a 5-mL Bio-Scale Macro-Prep High Q Column (Bio-Rad). The resin was washed with 20 mM Tris-HCl buffer containing 7.5 M urea, 80 mM NaCl, 1 mM DTT, and Triton-X100 (pH 7.4) and then eluted using a linear gradient of NaCl concentration from 80 mM to 1 M. The eluents were collected and loaded onto Nuvia IMAC resin again and eluted using a linear gradient of imidazole concentration from 15 to 500 mM. The eluents were dialyzed against RO water and then lyophilized. For SpiCE-NA1, the resin was washed with 100 mM sodium phosphate buffer with 7.5 M urea, and loaded onto a column packed with Nuvia IMAC (Bio-Rad). After loading, the resin was washed with 100 mM sodium phosphate buffer with 7.5 M urea, DTT, Triton-X100, and 15 mM imidazole and then eluted using a linear gradient of imidazole concentrations from 15 to 500 mM. The eluents were collected and dialyzed against buffer A (20 mM Tris-HCl, 7.5 M urea, 50 mM NaCl, DTT and Triton-X100 [pH 7.4]). The dialyzed sample was loaded onto a 5-mL Bio-Scale Macro-Prep High Q Column (Bio-Rad). The resin was washed with 20 mM Tris-HCl buffer containing 7.5 M urea, 80 mM NaCl, 1 mM DTT, and Triton-X100 (pH 7.4) and then eluted using a linear gradient of NaCl concentration from 80 mM to 1 M. The eluents were collected and loaded onto Nuvia IMAC resin again and eluted using a linear gradient of imidazole concentration from 15 to 500 mM. The eluents were dialyzed against RO water and then lyophilized.

Recombinant Films. HFIP was used as the solvent for the film synthesis. Recombinant MaSp2 proteins were mixed in 8 mL HFIP to a total of 100 mg and then dried at room temperature for 16 h in a dish (SI Appendix, Fig. 514). Composite films containing SpiCE-NA1 were produced by mixing the MaSp2 solution with recombinant SpiCE-NA1 at 1, 3, or 5 wt% each.

Recombinant Silks. Recombinant MaSp2 protein of 22 wt% (1.66 g) was dissolved in 5.89 g DMSO (4 wt% LiCl) and incubated at 90 °C for 8 h. For composite films containing SpiCE-NA1, recombinant MaSp2 protein of 22 wt% and recombinant SpiCE-NA1 of 0.012 g were dissolved in 4.2 g DMSO. The dope was aspirated with an N2 pump and D = 0.1 mm needle at 70 °C and spun directly into baths containing MeOH/DMSO 30 vol%/vol% (5 °C), MeOH (25 °C), and water (25 °C). Only MaSp2/SpiCE-NA1 composite silk could be produced because of the low yield of recombinant MaSp1 and MaSp3 proteins.

WAXS Measurement. The crystalline samples were characterized by a synchrotron WAXS measurement at the BL45XU beamline of SPring-8, Harima, Japan (43). The X-ray energy was set at 12.4 keV (wavelength: 0.1 nm), and the distance between the sample and the detector was ~257 mm. For the diffraction patterns, an exposure time of 10 s was used. Using Fit2D software (65), the obtained data were converted into one-dimensional radial integration profiles by subtracting the background scattering. The degree of crystallinity was calculated from the area of the crystal peaks divided by the combined area of the crystal peaks and the amorphous halo by fitting the Gaussian function using Igor Pro 6.3.

Measurement of Mechanical Properties. The tensile properties of the fibers were measured using an EZ-LX universal tester (Shimazu) with a 1 N load cell at a strain rate of 10 mm/min (0.016 s⁻¹) at 25 °C and 45% relative humidity. For each tensile test, the cross-sectional area of an adjacent section of the fiber was calculated based on the SEM images. The transparency of the film was measured at selected wavelengths between 350 and 600 nm using a UV instrument. Each fiber was attached to a rectangular piece of cardboard with a 5-mm aperture using 95% cyanoacrylate.

Tensile tests of the recombinant silk fibers were performed using a mechanical testing apparatus (Automatic Single-Fiber Test System FAVIMAT+, Textecno H. Stein GmbH & Co. KG) with a 210 cN load cell at a strain rate of 10 mm/min (0.016 s⁻¹) at 20 °C and RH 65%. From each fiber sample, diameters were measured at different sites (n = 6) by microscopic observation (Nikon Eclipse LV100ND, lens 150 x) to calculate the cross-sectional area of an adjacent section of the fibers.

Data Availability. Genome assembly, transcriptome sequencing, and proteome analysis data have been deposited in DNA Data Bank of Japan (DDBJ) and JPOST. All raw reads and assembled sequence data have been uploaded to DDBJ under BioProject nos. PRJDB10007, PRJDB10126, PRJDB10127, PRJDB10128, PRJDB7399, and PRJDB11252. The whole genome sequence is available at the Whole-Genome Shotgun database in DDBJ under accession nos. BMAW10000001–BMAW100000010, BMAW100000011–BMAW100000013 (T. clavipes), BMAW10000001–BMAW100028263 (T. inaurata madagascariensis), and BMAW10000001–BMAW1001132898 (N. pilipes). The PM 5 data raw reads and files have been uploaded to the ProteomeXchange Consortium from the JPOST partner repository under accession nos. PXD024371, PXD024372, PXD024374, and PXD024376.

ACKNOWLEDGMENTS. We thank Akio Tanikawa for morphological identification of spiders, helpful comments about phylogenetic discussion, and providing the photographs of T. clavata, T. clavipes, and N. pilipes, and we thank the Malagasy Institute for the Conservation of Tropical Environments, the Ministry of Environment and Sustainable Development of Madagascar, the Menton Zoologie et Biologie Animale, Universite d’Antananarivo, and Madagascar National Parks for their cooperation and permission to conduct sampling in Madagascar. We also thank Yuki Takai, Nozomi Abe, Yuki Onozawa, Sumiko Ohnuma, and Naoko Ishi for technical support in the sequencing and proteomic analysis; Hongphuc Pham, Haruka Funayama, Kaori Yaosaka, Ryota Sato, and Hirohori Yamamoto for technical support in the fiber spinning and analysis of fibers; Hiroshi Kano, Ryoko Sato, and Hideki Nishijima for the development of reeling machines; and Hiroyasu Masunaga for his technical support in SPring-8 measurements. H.N., M.T., K.N., and A.K. are supported by the IMPACT Program of the Council for Science, Technology, and Innovation (Cabinet Office, Government of Japan). N.K. is supported by a Nakatsuji Foresight Foundation Research Grant, the Sumitomo Foundation (190426), The Uehara Memorial Foundation, Grant-in-Aid for Scientific Research (B), N.K., M.K., M.T., and A.K. are supported by research funds from the Yamagata Prefectural Government and Tsuruoka City, Japan. Y.Y. is supported by Japan Society for the Promotion of Science KAKENHI Grant Number JP18J11155, and K.N. is financially supported by a Grant-in-Aid for Transformative Research Areas (B).
25. C. Larracas et al., Comprehensive proteomic analysis of spider dragline silk from black widows: A recipe to build synthetic silky fibers. Int. J. Mol. Sci. 17, 1537 (2016).
26. T. Pham et al., Dragline silk: A fiber assembled with low-molecular-weight cysteine-rich proteins. Biomacromolecules 15, 4073–4081 (2014).
27. P. L. Babb et al., The Neaphila clavipes genome highlights the diversity of spider silk genes and their complex expression. Nat. Genet. 48, 895–903 (2017).
28. N. Kono et al., Orb-weaving spider Araeus ventricosus genome elucidates the spidroin gene catalogue. Sci. Rep. 9, 8380 (2019).
29. J. Gatesy, C. Hayashi, D. Motriuk, J. Woods, R. Lewis, Extreme diversity, conservation, and convergence of spider silk fibroin sequences. Science 291, 2603–2605 (2001).
30. C. Y. Hayashi, R. V. Lewis, Molecular architecture and evolution of a modular spider silk protein gene. Science 287, 1477–1479 (2000).
31. M. Steinegger, S. L. Salzberg, Terminating contamination: Large-scale search identifies more than 2,000,000 contaminated entries in GenBank. Genome Biol. 21, 115 (2020).
32. N. Kono, K. Arakawa, Nanopore sequencing: Review of potential applications in functional genomics. Dev. Growth Diff. 61, 316–326 (2019).
33. P. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212 (2015).
34. T. M. Lowe, S. R. Eddy, tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964 (1997).
35. W. Li, A. Godzik, Dahl: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659 (2006).
36. N. A. Youbi, J. E. Garb, R. M. Tinglehitla, M. A. Collin, C. Y. Hayashi, Blueprint for a high-performance biomaterial: Full-length spider dragline silk genes. PLoS One 2, e514 (2007).
37. C. Y. Hayashi, T. A. Blackledge, R. V. Lewis, 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 (2004).
38. N. Kono et al., The bagworm genome reveals a unique fibroin gene that provides high tensile strength. Commun. Biol. 2, 148 (2019).
39. S. M. Correa-Garhwal, P. L. Babb, B. F. Voight, C. Y. Hayashi, Golden orb-weaving spider (Trichonephila clavipes) silk genes with sex-biased expression and atypical arachnida architecture. G3 (Bethesda) 11, gkaa039 (2021).
40. B. Schwannhäuser et al., Global quantification of mammalian gene expression control. Nature 473, 337–342 (2011).
41. C. Riekel, F. Vollrath, Spider silk fibre extrusion: Combined wide- and small-angle X-ray microdiffraction experiments. Int. J. Biol. Macromol. 29, 203–210 (2001).
42. K. Numata et al., Use of extension-deformation-based crystallisation of silk fibres to differentiate their functions in nature. Soft Matter 11, 6335–6342 (2015).
43. K. Yazawa, K. Ishida, H. Masunaga, T. Hakima, K. Numata, Influence of water content on the β-sheet formation, thermal stability, water removal, and mechanical properties of silk materials. Biomacromolecules 17, 1057–1066 (2016).
44. S. J. Blamires et al., Mechanical performance of spider silk is robust to nutrient-mediated changes in protein composition. Biomacromolecules 16, 1218–1225 (2015).
45. W. A. Gaines, M. G. Sehorn, W. R. Marciotte Jr, Spidroin N-terminal domain promotes a pH-dependent association of silk proteins during self-assembly. J. Biol. Chem. 285, 40745–40753 (2010).
46. G. Askari et al., Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. Nature 465, 236–238 (2010).
47. S. Schwarze, F. U. Zweitler, C. M. Johnson, H. Neuweiler, The N-terminal domains of spider silk proteins assemble ultrafast and protected from charge screening. Nat. Commun. 4, 2815 (2013).
48. H. C. Craig, D. Pirkovski, S. Nakagawa, M. M. Kasumovic, S. J. Blamires, Meta-analysis reveals matteromic relationships in major amouple silk across the spider phylogeny. J. R. Soc. Interface 17, 20200471 (2020).
49. J. E. Wagenseil, R. P. Mecham, New insights into elastic fiber assembly. Birth Defects Res. C Embryo Today 81, 229–240 (2007).
50. N. Kono, H. Nakamura, M. Mori, M. Tomita, K. Arakawa, Spidroin profiling of cribellate spiders provides insight into the evolution of spider prey capture strategies. Sci. Rep. 10, 15721 (2020).
51. J. E. Garb et al., The transcriptome of Darwin’s bark spider silk glands predicts proteins contributing to dragline silk toughness. Commun. Biol. 2, 275 (2019).
52. M. N. Sheffer et al., Chromosome-level reference genome of the European wasp spider Argiope bruennichi: A resource for studies on range expansion and evolutionary adaptation. Gigascience 10, giaa148 (2021).
53. N. Kono, H. Nakamura, Y. Ito, M. Tomita, K. Arakawa, Evaluation of the impact of RNA preservation methods of spiders for de novo transcriptome assembly. Mol. Ecol. Res. 16, 662–672 (2016).
54. D. Lettsch, M. Blaxter, BlobTools: Interrogation of genome assemblies. F1000 Res. 6, 1287 (2017).
55. B. Buchfink, C. Xie, H. D. Huson, Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60 (2015).
56. G. W. Vurture et al., GenomeScope: Fast reference-free genome profiling from short reads. Bioinformatics 33, 2202–2204 (2017).
57. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915 (2019).
58. K. J. Hoff, S. Lange, A. Lomsadze, M. Borodovsky, M. Stanke, BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32, 767–769 (2016).
59. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490 (2010).
60. K. Kato, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780 (2013).
61. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1915–1920 (2009).
62. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527 (2016).
63. H. Nielsen, K. D. Tsirigos, S. Brunak, G. von Heijne, A brief history of protein sorting trees for large alignments. Bioinformatics 35, 1659–1669 (2019).
64. G. V. Brykšin, I. Matsumura, Overlay extension PCR cloning: A simple and reliable way to create recombinant plasmids. Biotechniques 48, 463–465 (2010).
65. J. Chang, S. S. Mageres, C. C. Ling, Evaluation of rapid dose map acquisition of a scanning liquid-filled ionization chamber electronic portal imaging device. Int. J. Radiat. Oncol. Biol. Phys. 55, 1432–1445 (2003).