

Research



Cite this article: Kono N, Ohtoshi R, Malay AD, Mori M, Masunaga H, Yoshida Y, Nakamura H, Numata K, Arakawa K. 2021 Darwin's bark spider shares a spidroin repertoire with *Caerostris extrusa* but achieves extraordinary silk toughness through gene expression. *Open Biol.* **11**: 210242.

<https://doi.org/10.1098/rsob.210242>

Received: 17 August 2021

Accepted: 23 November 2021

Subject Area:

genomics/molecular biology

Keywords:

Darwin's bark spider, spidroin, spider, Madagascar

Author for correspondence:

Kazuharu Arakawa

e-mail: gaou@sfc.keio.ac.jp

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5752440>.

Darwin's bark spider shares a spidroin repertoire with *Caerostris extrusa* but achieves extraordinary silk toughness through gene expression

Nobuaki Kono^{1,2}, Rintaro Ohtoshi³, Ali D. Malay³, Masaru Mori^{1,2}, Hiroyasu Masunaga^{4,5}, Yuki Yoshida^{1,2}, Hiroyuki Nakamura⁶, Keiji Numata^{3,7} and Kazuharu Arakawa^{1,2}

¹Institute for Advanced Biosciences, Keio University, 403-1 Nihonkoku, Daihouji, Tsuruoka, Yamagata 997-0017, Japan

²Systems Biology Program, Graduate School of Media and Governance, Keio University, 5322 Endo, Fujisawa, Kanagawa 252-0882, Japan

³RIKEN Center for Sustainable Resource Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

⁴Japan Synchrotron Radiation Research Institute, 1-1-1, Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5198, Japan

⁵RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5198, Japan

⁶Spiber Inc., 234-1 Mizukami, Kakuganji, Tsuruoka, Yamagata 997-0052, Japan

⁷Department of Material Chemistry, Kyoto University, Kyotodaigaku-Katsura, Nishikyo-ku, Kyoto 615-8510, Japan

NK, 0000-0001-5960-7956; KA, 0000-0002-2893-4919

Spider silk is a protein-based material whose toughness suggests possible novel applications. A particularly fascinating example of silk toughness is provided by Darwin's bark spider (*Caerostris darwini*) found in Madagascar. This spider produces extraordinarily tough silk, with an average toughness of 350 MJ m⁻¹ and over 50% extensibility, and can build river-bridging webs with a size of 2.8 m². Recent studies have suggested that specific spidroins expressed in *C. darwini* are responsible for the mechanical properties of its silk. Therefore, a more comprehensive investigation of spidroin sequences, silk thread protein contents and phylogenetic conservation among closely related species is required. Here, we conducted genomic, transcriptomic and proteomic analyses of *C. darwini* and its close relative *Caerostris extrusa*. A variety of spidroins and low-molecular-weight proteins were found in the dragline silk of these species; all of the genes encoding these proteins were conserved in both genomes, but their genes were more expressed in *C. darwini*. The potential to produce very tough silk is common in the genus *Caerostris*, and our results may suggest the existence of plasticity allowing silk mechanical properties to be changed by optimizing related gene expression in response to the environment.

1. Introduction

Since its discovery in Madagascar in 2010, Darwin's bark spider, *Caerostris darwini* (Araneae: Araneidae), has fascinated the world because of its unique behaviour, ecology and biomaterial production ability, and has become a model organism for addressing the evolution of environmental adaptation in spiders [1]. Darwin's bark spider is an orb-weaving spider capable of creating large river-bridging webs with anchor threads as long as 25 m and sizes of up to 2.8 m² [1,2]. The size of the webs built by two other species, *Nephilengys borbonica* and *Trichonephila inaurata*, producing relatively large orb webs, reaches only approximately 0.4 m², indicating the exceptional size of the river-bridging webs of *C. darwini* [3,4]. It is thought that its extremely large orb web remains stable and unbroken because it consists of tough dragline silk spun from a relatively long

spinning duct of the major ampullate gland [5]. Spider silk is a typical high-performance protein material and with the attractive properties of high toughness, strength and extensibility [6]. The silk of Darwin's bark spider, in particular, shows an extraordinarily high level of toughness. A previous study on the mechanical properties of Darwin's bark spider silk reported tensile strength, extensibility and toughness values of 1.6 GPa, 52% and 354 MJ m⁻³, respectively [4]. Other orb-weaving spiders, such as *Trichonephila clavipes* or *Nephila pilipes*, show silk toughness ranging from only 100 to 150 MJ m⁻³ on average [7]. Some novel specific amino acid motifs in spidroins have been suggested as candidate factors responsible for the excellent mechanical properties of *C. darwini* silk.

A previous transcriptome analysis of Darwin's bark spider revealed that major ampullate spidroin (MaSp) families 4 (MaSp4) and 5 (MaSp5) harbour unique motifs [5,8]. Among these homologous groups, MaSp is a main component of dragline silk and is encoded by at least five gene families. MaSp families 1 and 2 (MaSp1 and MaSp2) have been well studied over a long period, whereas MaSp family 3 (MaSp3) was revealed by recent genomic and transcriptomic analyses [9–11]. MaSp4 and MaSp5 are new homologues that have so far only been found in Darwin's bark spider [5]. MaSp4 and MaSp5 contain 'VSVVSTTVS' and 'GGLGGSG' motifs, respectively, specific to Darwin's bark spider in their repetitive domains, and these proteins have been reported as possible factors contributing to silk toughness [5]. However, the sequence architectures of MaSp4 and MaSp5 have been only partially identified, and their detailed variation and conservation and even their actual role in dragline silks remain unclear. Knowledge of gene sizes, motif patterns and phylogenetic conservation is essential for discussing the evolution of silk mechanical properties. In addition, recent molecular biology research on spider silk has revealed that it is composed not only of spidroin, but also of protein materials of various sizes. Spider silk-constituting elements (SpiCEs) are non-spidroin low-molecular-weight (LMW) proteins of unknown function that have been widely found in the family Araneidae and were recently investigated to determine their effects on the formation and mechanical properties of spider silk [9,12–15].

Here, to enable comparative analysis between closely related species, we present two bark spider draft genomes and conduct a multiomics analysis. As a close relative of Darwin's bark spider, we use *Caerostris extrusa*. *Caerostris extrusa* is clearly defined as a different species even though it lives in the same region of Madagascar because the genetic distances between individuals of these species are much larger than those within the species inferred from DNA barcodes [16]. Using draft genomes prepared by applying hybrid sequencing technology, we conducted a multiomics analysis to curate a highly accurate spidroin catalogue, conduct phylogenetic searches of spidroins and profile their protein and mRNA expression. Based on these analyses, we describe candidate elements contributing to the differences in silk mechanical properties according to the identified genes or their expression patterns.

2. Methods

2.1. Sample collection

Adult female spiders (*C. darwini* and *C. extrusa*) were collected from Andasibe, eastern Madagascar (18°56'49.3" S,

48°25'09.1" E) and Parc National Andasibe Mantadia, eastern Madagascar (18°56'08.3" S, 48°24'53.2" E). The spiders were identified based on morphological characteristics and cytochrome c oxidase subunit 1 (*COI*) sequences in the Barcode of Life Data System (BOLD; <http://www.barcodinglife.org>). The natural dragline silks used for all experiments were sampled directly from adult female bark spiders restrained with two sponge pieces and locked with rubber bands. Silk reeling was performed at a constant speed (1.28 m m⁻¹ for 1 h) with a reeling device developed by Spiber Inc. The specimens collected according to a previously established field sampling method [17] were transported to the laboratory, immersed in liquid nitrogen (LN2) and stored at -80°C until subsequent processing. gDNA and total RNA were extracted from the samples.

2.2. High-molecular-weight gDNA extraction and genome sequencing

2.2.1. Extraction, purification and quality quantification

High-molecular-weight (HMW) gDNA was extracted from the legs of flash-frozen spiders using Genomic-tips 20/G (Qiagen) based on previous studies [9]. The specimens were gently and quickly homogenized using a BioMasher II homogenizer (Funakoshi) and mixed with 2 ml of Buffer G2 (Qiagen), including 200 µg ml⁻¹ RNase A and 50 µl proteinase K (20 mg ml⁻¹). After incubation at 50°C for 12 h on a shaker (300 r.p.m.), the mixed lysate was centrifuged at 5000g for 5 min at 4°C, and the aqueous phase was loaded onto a pre-equilibrated QIAGEN Genomic-tip 20/G by gravity flow and washed three times. The DNA was eluted with a high-salt buffer (Buffer QF) (Qiagen), desalted and concentrated using isopropanol precipitation and resuspended in 10 mM Tris-HCl (pH 8.5). The extracted gDNA was qualified using a TapeStation 2200 instrument with genomic DNA Screen Tape (Agilent Technologies) and quantified using a Qubit Broad Range dsDNA assay (Life Technologies). The purified gDNA was size-selected (greater than 10 kb) with a BluePippin with High Pass Plus Gel Cassette (Sage Science).

2.2.2. Library preparation and sequencing

Nanopore library preparation was implemented following the 1D library protocol (SQK-LSK109, Oxford Nanopore Technologies). The quality of the prepared library was calculated by the TapeStation 2200 system with D1000 Screen Tape (Agilent Technologies). Sequencing was performed using a GridION instrument with a SpotOn Flow Cell Rev D (FLO-MIN106D, Oxford Nanopore Technologies). Base calling was performed after the runs with Guppy base calling software (v. 3.2.10+aabd4ec). For 10X GemCode library preparation, purified gDNA fragments longer than 60 kb (10 ng) were used to prepare the library with a Chromium instrument and Genome Reagent Kit v2 (10X Genomics) following the manufacturer's protocol. 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).

2.3. RNA extraction and cDNA sequencing

RNA extraction was implemented based on a spider transcriptome protocol [17]. Flash-frozen dissected abdomen tissue was immersed in 1 ml TRIzol Reagent (Invitrogen) along with a metal cone and homogenized with a Multi-Beads Shocker (Yasui Kikai). After the addition of chloroform, the upper aqueous phase containing RNA was automatically purified with an RNeasy Plus Mini Kit (QIAGEN) on a QIAcube instrument (QIAGEN). The quantity and quality of the purified total RNA were calculated with a Qubit Broad Range RNA assay (Life Technologies) and a NanoDrop 2000 system (Thermo Fisher Scientific). The RNA integrity number (RIN) was estimated by electrophoresis using a TapeStation 2200 instrument with RNA ScreenTape (Agilent Technologies). mRNA was selected from the total RNA using oligo d(T). cDNA was synthesized from mRNA isolated from 100 µg of total RNA by NEBNext Oligo d(T)25 beads (skipping the Tris buffer wash step). First- and second-strand cDNA were synthesized using ProtoScript II Reverse Transcriptase and NEBNext Second Strand Synthesis Enzyme Mix. cDNA library preparation was performed 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 under the following conditions: 20 µl cDNA, 2.5 µl Index Primer, 2.5 µl Universal PCR Primer and 25 µl NEBNext Q5 Hot Start HiFi PCR Master Mix 2X; 98°C for 30 s; 12 cycles of 98°C for 10 s and 65°C for 75 s; and 65°C for 5 min. The cDNA library was sequenced with a NextSeq 500 instrument (Illumina) using 150 bp paired-end reads with a NextSeq 500 High Output Kit (300 cycles).

2.4. Genome assembly and contaminant elimination

We used a hybrid assembly strategy for *Cerostris* genomes based on previous studies [9,15]. Natural long reads were produced by Nanopore technology. The synthetic long reads were generated using a combination of Illumina and 10X Genomics technologies. The contaminants in the genome assemblies were eliminated based on BlobTools analysis [18]. The detail methods are described in electronic supplementary material.

2.5. Gene prediction and annotation

Genes were predicted using a gene model created from cDNA-seq mapping data. The cDNA-seq reads were mapped to the reference genome with HISAT2 (v.2.1.0) [19]. Repeat sequences were detected by RepeatModeler (1.0.11) and soft-masked by RepeatMasker (v. 4.0.7) (<http://www.repeatmasker.org>). The soft-masked genome was subjected to gene prediction with BRAKER (v. 2.1.4, `-softmasking -gff3`) [19,20]. The numbers of predicted protein-coding genes were initially 56 145 and 82 821 for *C. darwini* and *C. extrusa*, respectively. The predicted genes were annotated by Diamond BLASTP searches against public databases (UniProt TrEMBL, UniProt Swiss-Prot). Redundant genes were eliminated by CD-HIT-EST [21] clustering with a nucleotide identity of 97%. Furthermore, we collected the genes with

an expression level of more than 0.1 and annotated them to obtain functional gene sets. Finally, functional protein-coding gene sets of 16 907 and 18 655 genes were obtained. BUSCO (v. 4.0.5) was used to determine the quality of our functional gene set using the eukaryote lineage.

2.6. Spidroin catalogue curation

Spidroin genes identified in the bark spiders were curated based on a Spidroin Motif Collection (SMoC) algorithm [9,15,22–25]. This spidroin curation algorithm was implemented using the hybrid assembly of short and long reads. The de Bruijn graph assembly of Illumina short reads was used for N/C-terminal domain searching by homology searches. The obtained terminal domains were used as seeds for screening the short reads harbouring exact matches of extremely large k-mers extending to the 5'-end, and the short reads were aligned on the 3'-side of the matching k-mer to build a position weight matrix (PWM). Based on stringent thresholds, the terminal domains were extended until the next repeats appeared. Finally, the collected full-length subsets of the repeat units were mapped onto error-corrected Nanopore long reads. The data on spidroin gene length or architecture were curated manually based on the mapped long reads. Therefore, the full length could be obtained within one long-length read.

2.7. Phylogenetic tree

The spidroin phylogenetic trees were constructed by MEGAX [26] based on the first 100 N-terminal amino acid residues of the corresponding available spidroin sequences in the family Araneidae. The collected spidroin genes were aligned with MUSCLE, and the phylogenetic relationships were calculated using NJ. FigTree v. 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) was used as the viewer for the trees.

2.8. Proteome analysis

The proteome analysis of dragline silks was performed with nanoElute and timsTOF Pro (Bruker Daltonics, Bremen, Germany) based on a spider silk proteome protocol [15]. The details of the LC-MS analysis are described in electronic supplementary material.

2.9. Gene expression analysis

mRNA expression profiling was conducted from the cDNA-seq data. Gene expression levels were quantified and normalized as transcripts per million (TPM) values by mapping the processed reads to our assembled draft genome references with Kallisto v. 0.42.1 [27].

2.10. Measurement of bark spider silk properties

The surface morphology of the dragline silks was observed by SEM (JCM 6000, JEOL Ltd, Tokyo Japan). Samples were mounted on an aluminium stub with conductive tape backing and sputter-coated with gold for 1 min using a Smart Coater (JEOL) prior to SEM visualization at 5 kV. At least eight individual mechanical stretching tests were performed for each dragline silk. The experimental set-up was similar to those reported previously [28]. Each fibre was attached

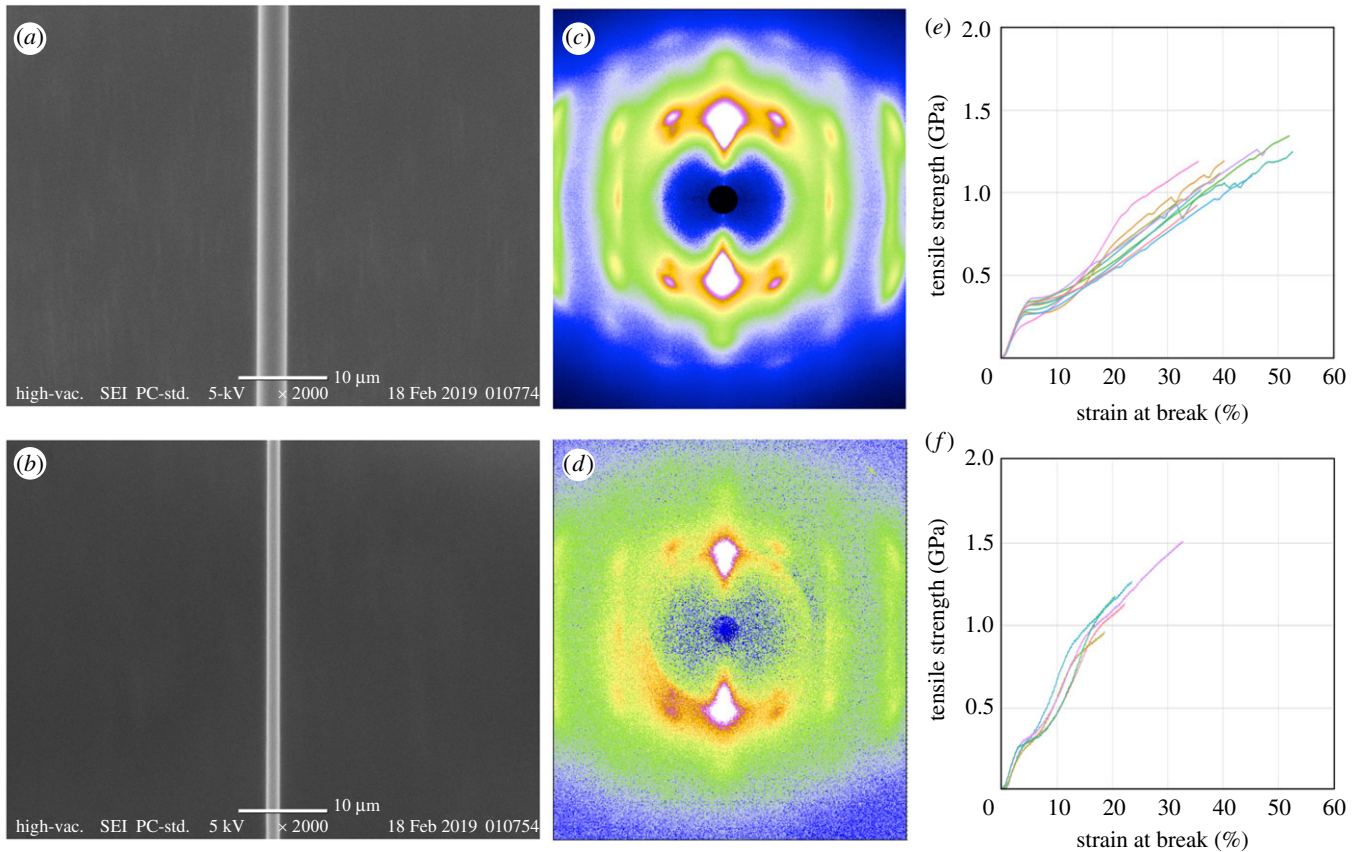


Figure 1. Structures and mechanical properties of dragline silks. (a,b) Scanning electron micrographs (SEMs), (c,d) 2D WAXS pattern images and (e,f) stress–strain (S–S) curves of *C. darwini* (a,c,e) and *C. extrusa* (b,d,f). Scale bar in (a) and (b) is 10 μm. S–S curves are the results of tensile tests of dragline silks from *C. darwini* (toughness: $284 \pm 80 \text{ MJ m}^{-3}$; tensile strength $1.17 \pm 0.14 \text{ GPa}$; strain at break: $42.8 \pm 6.93\%$) and *C. extrusa* (toughness: $172 \pm 76 \text{ MJ m}^{-3}$; tensile strength: $1.31 \pm 0.19 \text{ GPa}$; strain at break: $23.5 \pm 5.5\%$).

to a rectangular piece of cardboard with a 5 mm aperture using 95% cyanoacrylate. The tensile properties of the fibres were measured using an EZ-LX universal tester (Shimadzu, Kyoto, Japan) with a 1 N load cell at a strain rate of 10 mm min^{-1} (0.033 s^{-1}) at 25°C and 48% relative humidity. For each tensile test, the cross-sectional area of an adjacent section of the fibre was calculated based on the SEM images.

2.11. Measurement of wide-angle X-ray scattering

The crystalline state of the dragline silks was measured by synchrotron wide-angle X-ray scattering (WAXS) analysis in the SPring-8 BL05XU beamline, Harima, Japan, according to a previous report [29]. The X-ray energy was 12.4 keV at a wavelength of 0.1 nm. The sample-to-detector distance for the WAXS measurements was approximately 257 mm. The exposure time for each diffraction pattern was 10 s. The resultant data were converted into one-dimensional radial integration profiles using Fit2D software [30]. The resultant data were corrected by subtracting the background scattering. The degree of crystallinity was evaluated from the area of the crystal peaks divided by the total area of the crystal peaks and the amorphous halo by fitting the Gaussian function using Igor Pro 6.3.

3. Results

3.1. Silk comparison between bark spiders

To understand the uniqueness of Darwin's bark spider (*C. darwini*) silk is among known spider silks, we reeled

dragline silks from two bark spiders (*C. darwini* and *C. extrusa*) and measured the mechanical properties and WAXS profiles of the reeled dragline silks. The diameter of *C. darwini* dragline silk was on average 2.3 times larger than that of *C. extrusa*, and the crystallinity of *C. darwini* silk was 30%, compared to the 23% crystallinity of *C. extrusa* silk (figure 1a–d). Tensile tests confirmed the extraordinary toughness of Darwin's bark spider silk, as reported in a previous study [4]. The toughness of *C. darwini* silk was approximately twice that of *C. extrusa* silk, and the parameter that contributed most to this difference was extensibility (figure 1e,f). The extensibility of *C. extrusa* dragline silk was approximately 23.5%, which is an average value within the family Araneidae [31], while that of *C. darwini* was up to 49.73% (table 1). It is particularly interesting that this dragline silk can maintain such high extensibility while maintaining a tensile strength above 1 GPa. A comparative multiomics analysis was then carried out to investigate how *C. darwini* could produce such extraordinarily tough silk.

3.2. Genome sequences of bark spiders

We present the draft genome sequences of two bark spiders (*C. darwini* and *C. extrusa*) (figure 2a,b). The de novo sequencing of these large, complex spider genomes is challenging, and we sequenced the bark spider genomes via hybrid sequencing with a combination of Nanopore, 10x GemCode and Illumina technologies. Genomic DNA (gDNA) was extracted from dissected legs of adult female individuals.

Table 1. Mechanical properties of orb-weaving spider silks.

organism	Young's modulus (GPa)	tensile strength (GPa)	strain at break (%)	toughness (MJ m ⁻³)	reference
<i>C. darwini</i>	8.73 ± 0.76	1.17 ± 0.14	42.8 ± 6.93	284 ± 8.0	this study
<i>C. extrusa</i>	8.91 ± 1.29	1.31 ± 0.19	23.5 ± 5.50	172 ± 7.6	this study
<i>N. pilipes</i>	—	1.03 ± 0.18	29.0 ± 2.00	149 ± 25	[31]
<i>T. clavipes</i>	13.8 ± 0.76	1.00 ± 0.00	20.0 ± 1.10	111 ± 6.4	[7]
<i>Araneus gemmoides</i>	8.30 ± 0.54	1.06 ± 0.01	29.0 ± 2.40	141 ± 0.8	[7]

The Nanopore gDNA sequencing of the produced 10.23 million long reads with an N50 length of over 5.34 kb. In *C. darwini*, in addition to long reads, 990 million GemCode-barcoded 150 bp paired-end reads were sequenced by using Illumina technology. These gDNA sequenced reads were assembled, and the numbers of scaffolds (and N50 lengths) in the *C. darwini* and *C. extrusa* draft genomes were 15 733 (and 440 877 bp) and 21 729 (and 98 474 bp), respectively (figure 2c,d and table 2). The genome size of *C. darwini* was estimated with GenomeScop [32] based on the k-mer distribution to be 1.58 Gb, which is relatively small relative to other spider genomes, which average over 2.50 Gb in size [33]. The assembled scaffolds were assessed by Benchmarking Universal Single-Copy Ortholog (BUSCO) analysis [34], and the completeness scores of *C. darwini* and *C. extrusa* were 92.6% and 82.7%, respectively (figure 2c,d and table 2).

Gene prediction was conducted using a gene model generated from cDNA-seq mapping data. The cDNA library was constructed using mRNA extracted from abdomen samples of each species, and Illumina sequencing produced approximately 35 million reads per sample. The numbers of protein-coding genes initially predicted were 56 046 and 82 821 for *C. darwini* and *C. extrusa*, respectively. Redundant genes were eliminated based on identity, expression levels and annotation, and 16 907 and 18 655 functional protein-coding gene sets with BUSCO completeness scores of 94.1% and 82.8% were obtained from *C. darwini* and *C. extrusa*, respectively (table 2). The number of orthologous genes identified by bidirectional best hits (BBH, 1.0×10^{-5}) between expressed gene groups (TPM greater than 1.0) was 12 188 (figure 2f).

3.3. Full spidroin catalogues

In addition to the known classical spidroins, various families or subfamilies specific to the genus *Caerostris* were all found to be shared between the two newly obtained genomes (figure 3). Each spidroin was assigned a name according to the previously reported nomenclature [5,10,11,15]. Five families of MaSp, the main component of dragline silk, were observed, and nine genes (including paralogues) encoding these proteins were found. The full-length sequences were determined for almost all MaSp genes so that each gene type could be strictly distinguished. The repetitive units and gene lengths varied from 4 to 10 kb, while the terminal domains were well conserved.

MaSp4 and MaSp5, reported in a previous study [5], were found in both bark spiders. The GPGPQ repetitive motifs of MaSp4 were unique and differed from those of the spidroins reported thus far. However, a phylogenetic analysis showed

that all of the MaSp4 s belonged to the same clade as the MaSp2 proteins based on their terminal domains, suggesting that they may constitute a lineage of MaSp2 (figure 3a; electronic supplementary material, figure S1a,b). This result supported the previous study [5]. In general, GPGQQ is considered the MaSp2-specific motif in the family Araneidae [35], but the bark spider MaSp2 s had GPGSQ motifs in which the diglutamine (QQ) was replaced by SQ (table 3; electronic supplementary material, figure S1c). Therefore, the GPGPQ motif of MaSp4 can be considered an alternative to GPGQQ. In addition, MaSp4 did not contain a poly-A sequence, like MaSp2, but instead contained a VSVVSTTVS sequence, composed of neutral amino acids other than alanine (table 3; electronic supplementary material, figure S1c). Hence, if SQ or PQ and poly-X (X = neutral amino acids) sequences are viable alternatives to QQ and poly-A sequences, it may be reasonable to consider MaSp4 as a new subfamily of MaSp2.

The full-length of MaSp5 was also revealed. The size of MaSp5 was 3.5 kb, which is relatively small among spidroins. The MaSp5 proteins did not contain a tail portion of the repetitive unit (such as the poly-A sequence present in other MaSp families) and only showed tandemly arranged GGLGGSG or GSGGR motifs (table 3). The MaSp5 N-terminal regions domain did not cluster into the same clade as those of any other MaSp family; thus, MaSp5 seems to be a new family of MaSp, as suggested by previous studies [5]. In addition, we confirmed the presence of MaSp family 3 (MaSp3) in the two bark spider genomes by phylogenetic analysis (figure 3a). The MaSp3 proteins of the Araneidae family contain a DGGRGGY motif [9,10], but the MaSp3 found in the bark spiders contained a SGGRGGY motif, in which aspartic acid (D) was replaced by serine (S) (table 3).

No significant differences in spidroin other than MaSp were observed between the two species. A total of five minor ampullate spidroins (MiSps) were identified and classified into three subfamilies (MiSp1B, MiSp1C and MiSp1D). MiSp1B and MiSp1D had typical MiSp motif tails (poly-A/GA), but MiSp1C had an alternative poly-X tail, as observed in MaSp4. Multiple spidroin paralogues were also observed for flagelliform spidroin (Flag), the main component of the core fibre of the prey capture thread, including five paralogues in the *C. darwini* genome and six in the *C. extrusa* genome (figure 3b; electronic supplementary material, figure S2).

These spidroin characteristics and types were conserved in the two bark spiders without exception. Therefore, since omics approaches beyond the genome level were required to explain the toughness of *C. darwini* dragline silk, we performed protein and mRNA expression profiling.

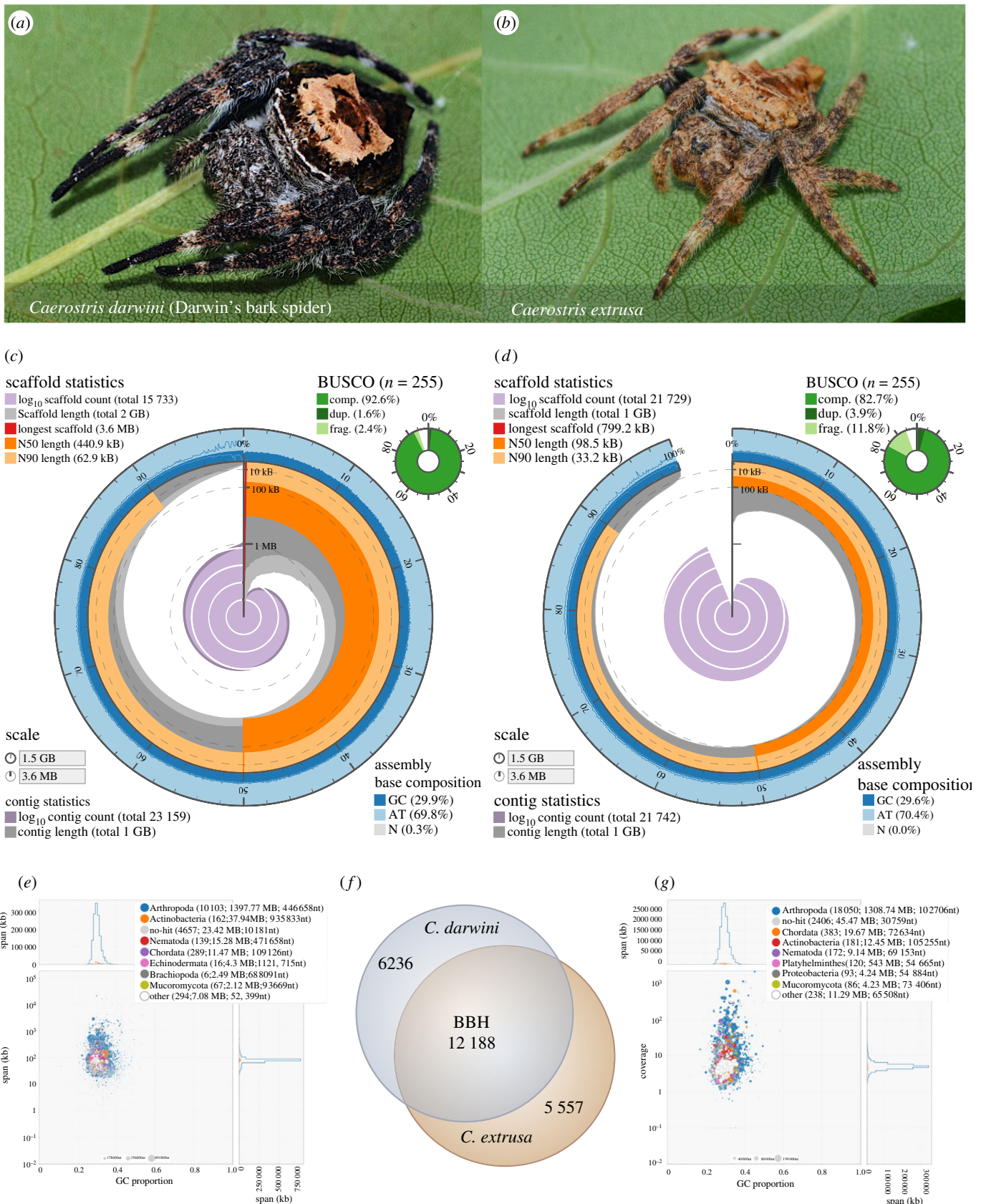


Figure 2. Overview of two bark spiders. Spider images and genome statistics of *C. darwini* (a,c,e) and *C. extrusa* (b,d,g). (c,d) The assembly and genome statistics with BUSCO completeness. (e,g) The results from the contamination check using BlobTools. These plots show the taxonomic affiliation at the phylum rank level, distributed according to GC% and coverage. (f) The number of orthologous genes identified by BBH (1.0×10^{-5}) between expressed gene groups (TPM > 1.0).

3.4. Expression profiling of proteins in dragline silk and mRNAs in the spider body

Proteome analysis was performed with nanoElute and timS-TOF using dragline silks reeled from adult female spiders.

The obtained spectra were annotated based on our draft genome database, and dozens of proteins were detected. The dragline silk contained all MaSp families 1–5, with MaSps alone accounting for more than 80% of the total proteins (figure 4a). The proteome analysis of the dragline silk

Table 2. Genome statistics.

	<i>Caerostris darwini</i>	<i>Caerostris extrusa</i>
genome	v4.0	v4.0
assembly size (Gb)	1.50	1.42
scaffold number	15 733	21 729
average scaffold length (bp)	95 463	65 380
longest scaffold length (bp)	3 567 188	799 240
N50 (bp) (no. of scaffolds in N50)	440 877 (918)	98 474 (4432)
N90 (bp) (no. of scaffolds in N90)	62 931 (4211)	33 245 (14 032)
BUSCO v4.0.5 completeness (%) ^a	92.6	82.7
genes		
number of protein-coding genes	16 907	18 655
tRNAs	2146	1608
BUSCO v4.0.5 completeness (%) ^a	94.1	82.8

^aBUSCO analysis based on eukaryotic lineages of protein-coding genes.

also demonstrated the presence of a small percentage of non-spidroins of unknown function. We defined the top four non-spidroin genes whose expression was confirmed by transcriptome analysis of the genus *Caerostris* dragline silk as SpiCEs, which were designated SpiCE-CMa1 to 4. SpiCEs are defined as non-spidroin LMW proteins associated with spider silk, of unknown function and showing both mRNA and protein expression [9]. Among these proteins, those with a high cysteine content are known as cysteine-rich proteins (CRPs) [14]. SpiCE-CMa4 contained 10% cysteine, suggesting that it was a CRP member (figure 4a). It was also notable sequence feature that SpiCE-CMa3 is rich in glycine and other hydrophobic amino acids. These SpiCE gene sequences were well conserved in the genomes of the two bark spiders, as were their spidroin sequences (figure 4b). However, a partial deletion of the disordered region was observed only in the gene sequence of *C. extrusa* SpiCE-CMa3. On average, the four SpiCEs contained 1–5% of dragline silk and SpiCE-CMa3 was the most abundant protein. Furthermore, the expression levels of several genes in dragline silk were observed to differ between the two spiders. According to the results of mRNA expression profiling by transcriptome analysis, the *MaSp4* and *MaSp5* genes, which were found to be *Caerostris* specific, were expressed at levels 2–6 times higher in *C. darwini* than in *C. extrusa*. In addition, the expression levels of almost all SpiCEs were more than twofold higher in *C. darwini* than in *C. extrusa* (figure 4c).

4. Discussion

Genome information allows comprehensive gene prediction and annotation and aids in large-scale transcriptomic, proteomic and phylogenomic analyses. However, in spider research, the available molecular information is limited in many cases. One of the main reasons for this is the large genome sizes of spiders. We have solved this problem by using hybrid sequence technology and obtained two bark spider genomes. The genome sizes of the bark spiders were smaller than those of other Araneidae spiders (approx. 3 Gb), but they were still more than 1 Gb, in line with the size of the *Argiope bruennichi*

genome [36]. The Nanopore long reads contributed significantly to the construction of a highly accurate spidroin catalogue.

Our catalogue confirmed the existence of *Caerostris*-specific MaSp paralogues, namely, MaSp families 4 and 5, reported in a previous study [5]. According to the full-length sequence of MaSp4 identified in this work, the gene previously reported as MaSp4 was clearly shown to cluster with the MaSp2 family based on the sequence similarity of terminal amino acid motifs (figure 3). The repetitive motifs also showed resemblance to the MaSp2 family. The previously identified QQ of *Caerostris* MaSp2 was found to be substitutable with SQ in the genus *Caerostris* (table 3); PQ was likewise used substitutably in MaSp4. The polyalanine region was replaced with other neutral amino acids (S, V, I or T). Similar replacement of polyalanine motifs with other neutral amino acids was observed in the MaSp family 1 (MaSp1), in which the tail portion of MaSp1E or MaSp1C repetitive unit contained a mixture of poly-A and SSSVAISL or ISISAAAS motifs, respectively (table 3). Overall, extreme specialization of the amino acid motifs of MaSp paralogues seems to have occurred in the *Caerostris* genus. Previous studies of the genomes of *Araneus ventricosus* and *Trichonephila clavata* identified the existence of the Araneoidea-specific MaSp3 family of spidroins as well as the clade-specific non-spidroin component SpiCE proteins, both of which are essential for the high mechanical performance of silks in these species [9,15]. *Caerostris*, classified in the family Araneidae, also possesses MaSp3 paralogues as well as clade-specific SpiCEs. Intriguingly, all of these genes were shown to be conserved in both *C. darwini* and *C. extrusa*, and these two species showed little difference in terms of the silk gene repertoire, although the mechanical performance of the silks of these two species differed significantly.

What then, are the elements contributing the toughness of Darwin's bark spider silk? Our expression analysis showed differences in the deployment of the gene repertoire. *Caerostris extrusa* showed a typical expression profile of Araneoidea, with high expression levels of MaSp3 and canonical MaSp1 and MaSp2 paralogues [5,9,11,15,23]. Conversely, the expression of the *Caerostris*-specific paralogues *MaSp4* and *MaSp5* genes

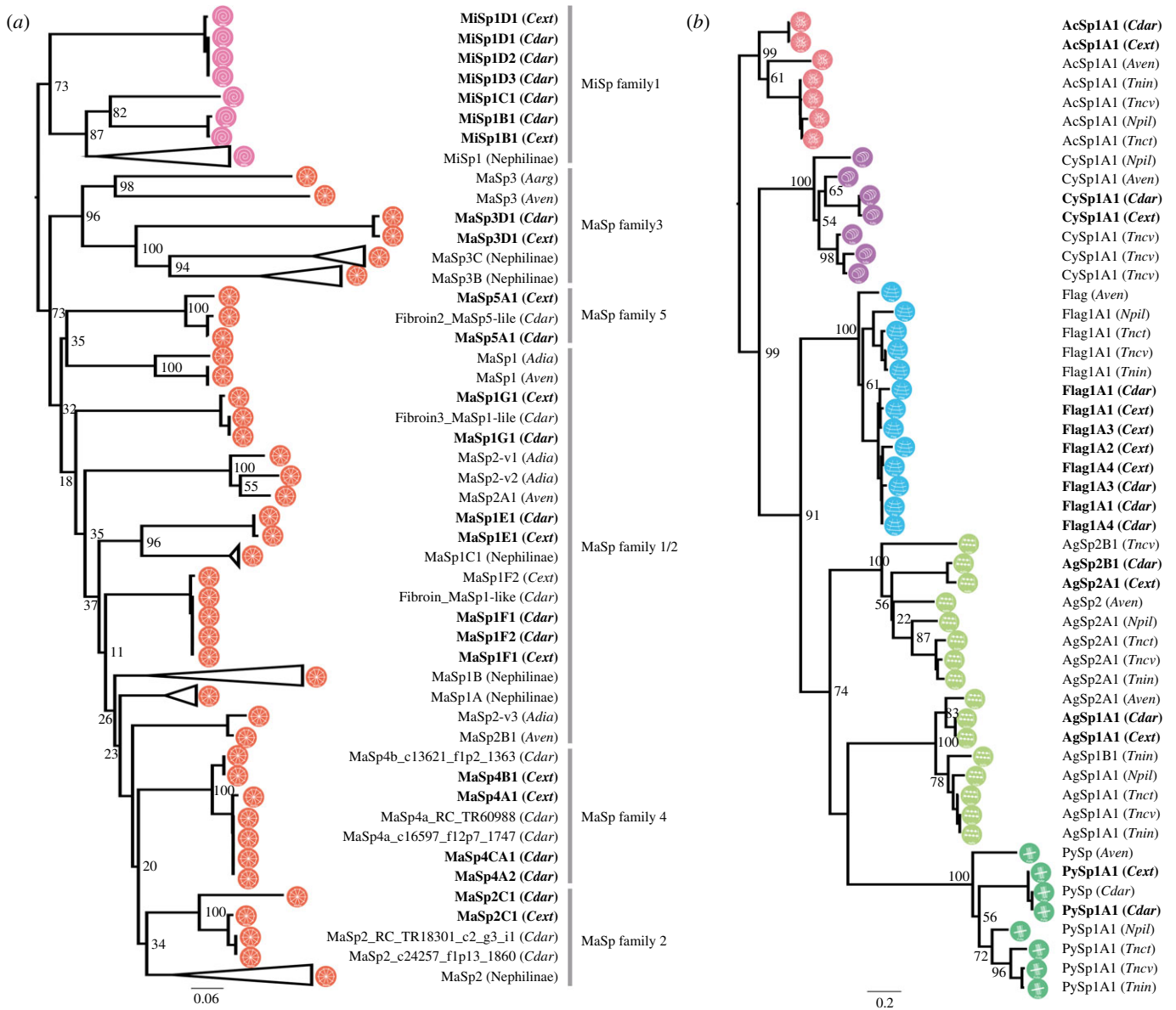


Figure 3. Phylogenetic tree of orb-weaving spider spideroins. Phylogenetic analysis of all spideroins ((a) ampullate spideroins and (b) other spideroins) with 100 aa N-terminal regions (*Cdar*, *C. darwini*; *Cext*, *C. extrusa*; *Aarg*, *Argiope argentata*; *Aven*, *A. ventricosus*; *Adia*, *Araneus diadematus*; *Tnct*, *T. clavata*; *Tncv*, *T. clavipes*; *Tnin*, *T. inaurata madagascariensis*; *Npil*, *N. pilipes*). Genus *Caerostris* spideroins are indicated by bold font. Branch labels are bootstrap support values.



Figure 4. Expression profiles of protein and mRNA. (a) The results of the proteome analysis of five dragline silk samples. Spideroins and SpiCEs are indicated by grey and orange, respectively. (b) Sequence alignment of SpiCEs from *C. darwini* and *C. extrusa*. (c) Gene expression ratio between *C. darwini* and *C. extrusa*. Spideroins and SpiCEs are indicated by grey and orange, respectively.

Table 3. Repeat motifs in *Caerostris* MiSp and MaSp.

	spidroin	motif		linker (aa)	length (bp)
		body	tail		
MiSp1B	<i>C. darwini</i>	GGYGGGA	(GA) _n (A) _n	130	10659
	<i>C. extrusa</i>	GGYGGGA	(GA) _n (A) _n	130	10584
MiSp1C	<i>C. darwini</i>	GGYGGGA	[(A) _n ISISAAAAS]	–	7434
	<i>C. extrusa</i>	–	–	–	–
MiSp1D	<i>C. darwini</i>	GGYGGG[Y Q]GG	(GA) _n (A) _n	–	8967, 2211, 7053
	<i>C. extrusa</i>	GGYGGG[Y Q]GG	(GA) _n (A) _n	–	8628
MaSp1E	<i>C. darwini</i>	GYPGQ	[(A) _n SSSVAISL]	–	3744
	<i>C. extrusa</i>	GYPGQ	[(A) _n SSSVAISL]	–	5739
MaSp1F	<i>C. darwini</i>	GQGGYGS	(A) _n	–	8958, 9423
	<i>C. extrusa</i>	GQGGYGS	(A) _n	–	11967, 9111
MaSp1G	<i>C. darwini</i>	GSGGYGGR	(A) _n	–	8652
	<i>C. extrusa</i>	GSGGYGGR	(A) _n	–	8880
MaSp2C	<i>C. darwini</i>	GPGSQ	(A) _n	–	11913 <
	<i>C. extrusa</i>	GPGSQ	(A) _n	–	12711
MaSp3D	<i>C. darwini</i>	SGGRGGY	(A) _n	–	6987
	<i>C. extrusa</i>	SGGRGGY	(A) _n	–	7224
MaSp4A	<i>C. darwini</i>	GPGPQ	[V I]SVVS[A T]VS	–	8205, 5460
	<i>C. extrusa</i>	GPGPQ	VSVVS[A T]VS	–	5142, 6435
MaSp5A	<i>C. darwini</i>	GGLGGSG GSGGR	–	–	3510
	<i>C. extrusa</i>	GGLGGSG GSGGR	–	–	4458

was greatly increased by 2–6 times in *C. darwini*. Furthermore, the expression of SpiCE-CMa3 was almost doubled in *C. darwini*, and the length of SpiCE-CMa3 in *C. extrusa* was approximately half that in *C. darwini* (figure 4a; electronic supplementary material, figure S3a–e). These changes together may provide one explanation for the significant differences in the mechanical performance of the draglines of these two species.

Previous proteomics analyses of *A. ventricosus* and *T. clavata* silks showed a predominance of MaSp3 in the silk composition of these species [9,15]. However, the proteome analysis of *C. darwini* dragline silk demonstrated that the proteins of the

MaSp2 class (MaSp2 and MaSp4) were its major components, accounting for approximately 50% of the total proteins content. As shown in figure 1, the high toughness of *C. darwini* silk is mainly accounted for by enhanced extensibility. The presence of spidroins with many prolines in the repetitive domain (as found MaSp2) is known to increase the extensibility of spider silk [37–42], and the predominance of the MaSp2 class in the dragline silk of *C. darwini* therefore seemed to be in accord with its high extensibility (figure 4), explaining the long stretch beyond the yield point in particular (figure 1e). The initial stretching of the spider silk reaches the yield point with the rupture of the hydrogen bonds in semiamorphous regions (helices

and β -turns), and subsequent deformation involves the β -sheets in the crystal domain [43]. The spidroin β -sheets are composed of poly-A motifs [39,44–47]. Johansson & Rising [48] have proposed a hypothesis about the implications of the replacement of poly-A sequences with other amino acids for silk engineering; for example, alanine polymers are proposed to be replaceable by polymers of valine (poly-V) or isoleucine (poly-I) to form stiff sheets [49]. The bark spider sequences do not contain valine and isoleucine as homopolymers, but they are included in combination with other amino acids, as in the ISVVSTTVS motif (table 3). This semiconservative mode of replacement, accompanied by expression level regulation, may improve β -stacking after the yield point to a greater extent than is found in other silks.

The mechanical properties of spider silk sampled directly from a spider are affected by various factors such as humidity, temperature or reeling speed [29,50]. Therefore, it is difficult to directly compare several previous studies, and data analysis derived from the same test is essential to consider the molecular level effects in particular. In this study, we independently measured and compared the mechanical properties of silk under identical spinning parameters. Thus, the differences in the mechanical properties of silks between the two studied bark spiders were attributed to differences in the predominant spidroins. Although the existence of the clade-specific MaSp paralogues MaSp4 and MaSp5 in *C. darwini*, previously identified by transcriptome analysis, was confirmed by our genome analysis, these genes are also conserved in *C. extrusa*, whose silk does not show comparable performance. However, differences in the deployment of the shared repertoire of silk genes in these two species may be the key to the differences in their silk, as the predominant use of MaSp2 family proteins, including MaSp4s, in *C. darwini* is in accord with the significant enhancement of elasticity in its dragline. Since structural differences can also contribute to mechanical properties, research on the detailed contributions of these components will help to further explain the molecular mechanism underlying the extraordinary toughness of *C. darwini* silk.

Data accessibility. All raw reads and assembled sequence data have been uploaded to DDBJ under BioProject no. PRJDB11546 and PRJDB11547. The whole-genome sequence is available at the Whole-Genome

Shotgun (WGS) database of DDBJ under accession numbers BPLQ01000001–BPLQ01015733 (*C. darwini*) and BPLR01000001–BPLR01021729 (*C. extrusa*). The raw MS data and analysis files have been uploaded to the ProteomeXchange Consortium from the jPOST partner repository [51] under accession no. PXD026224.

The data are provided in electronic supplementary material [52].

Authors' contributions. N.K.: data curation, formal analysis, investigation, methodology, validation, visualization, writing—original draft, writing—review and editing; R.O.: formal analysis, investigation, methodology, resources, writing—review and editing; A.D.M.: formal analysis, investigation, methodology, resources, writing—review and editing; M.M.: formal analysis, investigation, methodology, writing—review and editing; H.M.: formal analysis, investigation, methodology, writing—review and editing; Y.Y.: data curation, formal analysis, investigation, methodology, writing—review and editing; H.N.: formal analysis, investigation, methodology, resources, writing—review and editing; K.N.: formal analysis, investigation, methodology, resources, writing—review and editing; K.A.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. H.N. is an employee of Spiber Inc., a venture company selling brewed protein products. However, the design of all study procedures was conducted by K.A. of Keio University, and Spiber Inc. had no role in the study design, data analysis or data interpretation. N.K., R.O., A.D.M., M.M., H.M., Y.Y., K.N. and K.A. declare no conflict of interests.

Funding. The work was supported by a Nakatsuji Foresight Foundation Research Grant, the Sumitomo Foundation (190426) and a Grant-in-Aid for Scientific Research (B) (21H02210) to N.K., a grant from the ImPACT Program of the Council for Science, Technology and Innovation (Cabinet Office, Government of Japan) to H.N., K.N. and K.A., and a grant from the Yamagata Prefectural Government and Tsuruoka City, Japan, to N.K., M.M. and K.A.

Acknowledgements. The authors thank Akio Tanikawa for the morphological identification of spiders and the Malagasy Institute for the Conservation of Tropical Environments (MICET) (especially Tiana Vololontiana), the Ministry of Environment and Sustainable Development of Madagascar (Ministère de l'Environnement de l'Ecologie et des Forêts at that time), and the Mention Zoologie et Biologie Animale (MZBA), Université d'Antananarivo and Madagascar National Parks (MNP) for their cooperation and permission to conduct sampling in Madagascar. The authors also thank Yuki Takai and Naoko Ishii for technical support in the sequencing analysis and Sumiko Ohnuma for technical support in proteome analysis.

References

- Kuntner M, Agnarsson I. 2010 Web gigantism in Darwin's bark spider, a new species from Madagascar (Araneidae: Caerostris). *J. Arachnol.* **38**, 346–356. (doi:10.1636/B09-113.1)
- Gregoric M, Agnarsson I, Blackledge TA, Kuntner M. 2011 How did the spider cross the river? Behavioral adaptations for river-bridging webs in *Caerostris darwini* (Araneae: Araneidae). *PLoS ONE* **6**, e26847. (doi:10.1371/journal.pone.0026847)
- Sensenig A, Agnarsson I, Blackledge TA. 2010 Behavioural and biomaterial coevolution in spider orb webs. *J. Evol. Biol.* **23**, 1839–1856. (doi:10.1111/j.1420-9101.2010.02048.x)
- Agnarsson I, Kuntner M, Blackledge TA. 2010 Bioprospecting finds the toughest biological material: extraordinary silk from a giant riverine orb spider. *PLoS ONE* **5**, e11234. (doi:10.1371/journal.pone.0011234)
- Garb JE, Haney RA, Schwager EE, Gregoric M, Kuntner M, Agnarsson I, Blackledge TA. 2019 The transcriptome of Darwin's bark spider silk glands predicts proteins contributing to dragline silk toughness. *Commun. Biol.* **2**, 275. (doi:10.1038/s42003-019-0496-1)
- Omenetto FG, Kaplan DL. 2010 New opportunities for an ancient material. *Science* **329**, 528–531. (doi:10.1126/science.1188936)
- Swanson BO, Blackledge TA, Beltrán J, Hayashi CY. 2006 Variation in the material properties of spider dragline silk across species. *Appl. Phys. A* **82**, 213–218. (doi:10.1007/s00339-005-3427-6)
- Htut KZ *et al.* 2021 Correlation between protein secondary structure and mechanical performance for the ultra-tough dragline silk of Darwin's bark spider. *J. R. Soc. Interface* **18**, 20210320. (doi:10.1098/rsif.2021.0320)
- Kono N *et al.* 2019 Orb-weaving spider *Araneus ventricosus* genome elucidates the spidroin gene catalogue. *Sci. Rep.* **9**, 8380. (doi:10.1038/s41598-019-44775-2)
- Collin MA, Clarke III TH, Ayoub NA, Hayashi CY. 2018 Genomic perspectives of spider silk genes through target capture sequencing: conservation of stabilization mechanisms and homology-based structural models of spidroin terminal regions. *Int. J. Biol. Macromol.* **113**, 829–840. (doi:10.1016/j.ijbiomac.2018.02.032)
- Babb PL *et al.* 2017 The *Nephila clavipes* genome highlights the diversity of spider silk genes and

- their complex expression. *Nat. Genet.* **49**, 895–903. (doi:10.1038/ng.3852)
12. Larracas C, Hekman R, Dyrness S, Arata A, Williams C, Crawford T, Vierra CA. 2016 Comprehensive proteomic analysis of spider dragline silk from black widows: a recipe to build synthetic silk fibers. *Int. J. Mol. Sci.* **17**, 1537. (doi:10.3390/ijms17091537)
 13. Chaw RC, Correa-Garhwal SM, Clarke TH, Ayoub NA, Hayashi CY. 2015 Proteomic evidence for components of spider silk synthesis from black widow silk glands and fibers. *J. Proteome Res.* **14**, 4223–4231. (doi:10.1021/acs.jproteome.5b00353)
 14. Pham T *et al.* 2014 Dragline silk: a fiber assembled with low-molecular-weight cysteine-rich proteins. *Biomacromolecules* **15**, 4073–4081. (doi:10.1021/bm5011239)
 15. Kono N *et al.* 2021 Multicomponent nature underlies the extraordinary mechanical properties of spider dragline silk. *Proc. Natl Acad. Sci. USA* **118**, e2107065118. (doi:10.1073/pnas.2107065118)
 16. Gregorič M, Blackledge TA, Agnarsson I, Kuntner M. 2015 A molecular phylogeny of bark spiders reveals new species from Africa and Madagascar (Araneae: Araneidae: Caerostris). *J. Arachnol.* **43**, 293–312. 220. (doi:10.1636/0161-8202-43.3.293)
 17. Kono N, Nakamura H, Ito Y, Tomita M, Arakawa K. 2016 Evaluation of the impact of RNA preservation methods of spiders for de novo transcriptome assembly. *Mol. Ecol. Resour.* **16**, 662–672. (doi:10.1111/1755-0998.12485)
 18. Laetsch D, Blaxter M. 2017 BlobTools: interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research* **6**, 1287. (doi:10.12688/f1000research.12232.1)
 19. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019 Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915. (doi:10.1038/s41587-019-0201-4)
 20. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016 BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769. (doi:10.1093/bioinformatics/btv661)
 21. Li W, Godzik A. 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659. (doi:10.1093/bioinformatics/btl158)
 22. Kono N, Arakawa K. 2019 Nanopore sequencing: review of potential applications in functional genomics. *Dev. Growth Differ.* **61**, 316–326. (doi:10.1111/dgd.12608)
 23. Kono N, Nakamura H, Mori M, Tomita M, Arakawa K. 2020 Spidroin profiling of cribellate spiders provides insight into the evolution of spider prey capture strategies. *Sci. Rep.* **10**, 15721. (doi:10.1038/s41598-020-72888-6)
 24. Kono N, Nakamura H, Ohtoshi R, Tomita M, Numata K, Arakawa K. 2019 The bagworm genome reveals a unique fibroin gene that provides high tensile strength. *Commun. Biol.* **2**, 148. (doi:10.1038/s42003-019-0412-8)
 25. Kono N, Nakamura H, Tateishi A, Numata K, Arakawa K. 2021 The balance of crystalline and amorphous regions in the fibroin structure underpins the tensile strength of bagworm silk. *Zool. Lett.* **7**, 11. (doi:10.1186/s40851-021-00179-7)
 26. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018 MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. (doi:10.1093/molbev/msy096)
 27. Bray NL, Pimentel H, Melsted P, Pachter L. 2016 Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527. (doi:10.1038/nbt.3519)
 28. Malay AD *et al.* 2016 Relationships between physical properties and sequence in silkworm silks. *Sci. Rep.* **6**, 27573. (doi:10.1038/srep27573)
 29. Yazawa K, Ishida K, Masunaga H, Hikima T, Numata K. 2016 Influence of water content on the beta-sheet formation, thermal stability, water removal, and mechanical properties of silk materials. *Biomacromolecules* **17**, 1057–1066. (doi:10.1021/acs.biomac.5b01685)
 30. Hammersley AP, Svensson SO, Hanfland M, Fitch AN, Hausermann D. 1996 Two-dimensional detector software: From real detector to idealised image or two-theta scan. *High Pressure Res.* **14**, 235–248. (doi:10.1080/08957959608201408)
 31. Kerr GG, Nahrung HF, Wiegand A, Kristoffersen J, Killen P, Brown C, Macdonald J. 2018 Mechanical properties of silk of the Australian golden orb weavers *Nephila pilipes* and *Nephila plumipes*. *Biol. Open* **7**, bio029249. (doi:10.1242/bio.029249)
 32. Vurture GW, Sedlaczek FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017 GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204. (doi:10.1093/bioinformatics/btx153)
 33. Fan Z, Yuan T, Liu P, Wang LY, Jin JF, Zhang F, Zhang ZS. 2021 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. *Gigascience* **10**, giab016. (doi:10.1093/gigascience/giab016)
 34. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212. (doi:10.1093/bioinformatics/btv351)
 35. Malay AD, Arakawa K, Numata K. 2017 Analysis of repetitive amino acid motifs reveals the essential features of spider dragline silk proteins. *PLoS ONE* **12**, e0183397. (doi:10.1371/journal.pone.0183397)
 36. Sheffer MM *et al.* 2021 Chromosome-level reference genome of the European wasp spider *Argiope bruennichi*: a resource for studies on range expansion and evolutionary adaptation. *Gigascience* **10**, giae148. (doi:10.1093/gigascience/giae148)
 37. Xu M, Lewis RV. 1990 Structure of a protein superfiber: spider dragline silk. *Proc. Natl Acad. Sci. USA* **87**, 7120–7124. (doi:10.1073/pnas.87.18.7120)
 38. Hinman MB, Lewis RV. 1992 Isolation of a clone encoding a second dragline silk fibroin. *Nephila clavipes* dragline silk is a two-protein fiber. *J. Biol. Chem.* **267**, 19 320–19 324. (doi:10.1016/S0021-9258(18)41777-2)
 39. Hayashi CY, Shipley NH, Lewis RV. 1999 Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *Int. J. Biol. Macromol.* **24**, 271–275. (doi:10.1016/S0141-8130(98)00089-0)
 40. Blamires SJ, Wu CL, Tso IM. 2012 Variation in protein intake induces variation in spider silk expression. *PLoS ONE* **7**, e31626. (doi:10.1371/journal.pone.0031626)
 41. Blamires SJ, Wu CL, Blackledge TA, Tso IM. 2012 Post-secretion processing influences spider silk performance. *J. R. Soc. Interface* **9**, 2479–2487. (doi:10.1098/rsif.2012.0277)
 42. Ayoub NA, Garb JE, Tinghitella RM, Collin MA, Hayashi CY. 2007 Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS ONE* **2**, e514. (doi:10.1371/journal.pone.0000514)
 43. Keten S, Buehler MJ. 2010 Nanostructure and molecular mechanics of spider dragline silk protein assemblies. *J. R. Soc. Interface* **7**, 1709–1721. (doi:10.1098/rsif.2010.0149)
 44. Simmons A, Ray E, Jelinski LW. 1994 Solid-state ¹³C NMR of *Nephila clavipes* dragline silk establishes structure and identity of crystalline regions. *Macromolecules* **27**, 5235–5237. (doi:10.1021/ma00096a060)
 45. Simmons AH, Michal CA, Jelinski LW. 1996 Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science* **271**, 84–87. (doi:10.1126/science.271.5245.84)
 46. van Beek JD, Beaulieu L, Schafer H, Demura M, Asakura T, Meier BH. 2000 Solid-state NMR determination of the secondary structure of *Samia cynthia ricini* silk. *Nature* **405**, 1077–1079. (doi:10.1038/35016625)
 47. Scheibel T. 2004 Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. *Microb. Cell Fact.* **3**, 14. (doi:10.1186/1475-2859-3-14)
 48. Johansson J, Rising A. 2021 Doing what spiders cannot—a road map to supreme artificial silk fibers. *ACS Nano* **15**, 1952–1959. (doi:10.1021/acsnano.0c08933)
 49. Goldschmidt L, Teng PK, Riek R, Eisenberg D. 2010 Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* **107**, 3487–3492. (doi:10.1073/pnas.0915166107)
 50. Vollrath F, Madsen B, Shao Z. 2001 The effect of spinning conditions on the mechanics of a spider's dragline silk. *Proc. R. Soc. Lond. B* **268**, 2339–2346. (doi:10.1098/rspb.2001.1590)
 51. Moriya Y *et al.* 2019 The jPOST environment: an integrated proteomics data repository and database. *Nucleic Acids Res.* **47**, D1218–D1224. (doi:10.1093/nar/gky899)
 52. Kono N, Ohtoshi R, Malay AD, Mori M, Masunaga H, Yoshida Y, Nakamura H, Numata K, Arakawa K. 2021 Darwin's bark spider shares a spidroin repertoire with *Caerostris extrusa* but achieves extraordinary silk toughness through gene expression. Figshare.