Conservation of a pH-sensitive structure in the C-terminal region of spider silk extends across the entire silk gene family

Michelle Strickland[1]*, Victor Tudorica[2], Milan Řezáč[3], Neil R. Thomas[2], Sara L. Goodacre[1]

[1]School of Life Sciences, University of Nottingham, NG7 2RD, UK

[2]Centre for Biomolecular Sciences, School of Chemistry, University of Nottingham, NG7 2RD, UK

[3] Biodiversity Lab, Crop Research Institute, Drnovská 507, CZ-16106 Prague 6 - Ruzyně, Czechia

*corresponding author michelle.strickland@nottingham.ac.uk

rezac@vurv.cz

neil.thomas@nottingham.ac.uk

sara.goodacre@nottingham.ac.uk

Word count: 2493

Abstract

Spiders produce multiple silks with different physical properties that allow them to occupy a diverse range of ecological niches, including the underwater environment. Despite this functional diversity, past molecular analyses show a high degree of amino acid sequence similarity between C-terminal regions of silk genes that appear to be independent of the physical properties of the resulting silks; instead, this domain is crucial to the formation of silk fibres.

Here we present an analysis of the C-terminal domain of all known types of spider silk and include silk sequences from the spider *Argyroneta aquatica*, which spins the majority of its silk underwater. Our work indicates that spiders have retained a highly conserved mechanism of silk assembly, despite the extraordinary diversification of species, silk types and applications of silk over 350 million years. Sequence analysis of the silk C-terminal domain across the entire gene family shows the conservation of two uncommon amino acids that are implicated in the formation of a salt bridge, a functional bond essential to protein assembly. This conservation extends to the novel sequences isolated from *A. aquatica*.

This finding is relevant to research regarding the artificial synthesis of spider silk, suggesting that synthesis of all silk types will be possible using a single process.

Key Words: Spider silk, Synthetic silk, pH-bridge, *Argyroneta aquatica*

Introduction

Spider silk proteins form multiple types of materials, including fibres and glues, each with specific mechanical properties and functions (Blamires *et al*, 2017; Eisoldt *et al*, 2011; Garb, 2013; Hormiga and Griswold, 2014). Gene duplication, recombination and diversification are thought to have generated this huge diversity of different silk types, each having a different ecological function (Clarke *et al*, 2015). These functions include safety lines, the structural frameworks of webs, external layers of protection around egg sacs, and securing prey (Hinman *et al*, 2000; Rising and Johansson, 2015).

In the case of *Argyroneta aquatica* (Clerck, 1757) (Araneae: Cybaeidae (Catalog, 2016)), silk is used in a typical fashion to construct a web in which the spider resides and performs a number of actions, from feeding to mating and storing eggs (Schütz and Taborsky, 2003; Schütz and Taborsky, 2005; Schütz and Taborsky, 2011; Schutz *et al*, 2007; Seymour and Hetz, 2011). What is unusual about this spider is that it is the only known species to spin silk whilst submersed in water. The subsequent sheet web is then inflated with air drawn down from the surface and is utilised as an air reservoir. The silken “diving bell” allows for oxygen diffusion to occur, allowing the spider to avoid surfacing for extended periods of time (Seymour and Hetz, 2011). *Cybaeus angustiarum*, a fellow cybaeid, is a terrestrial species found in dense forests (often on north facing, scree slopes) in areas under stones or in decaying wood with humidity close to 100%.

In all spiders, liquid silk dope is passed through specialised, elongated glands within the body of the spider and extruded through spinnerets (Askarieh *et al*, 2010; Rising and Johansson, 2015; Vollrath and Knight, 2001).

All silk genes have three components; a type-specific repetitive region flanked by conserved N-terminal (Motriuk-Smith *et al*, 2005; Rising *et al*, 2006) and C-terminal (Challis *et al*, 2006; Collin *et al*, 2016; Gnesa *et al*, 2012; Hagn *et al*, 2010) domains. Whilst the repetitive region determines the mechanical properties of each silk product (Hayashi and Lewis, 1998; Hayashi *et al*, 1999), the terminal domains work together ensuring that individual proteins assemble correctly and that the fibre forms at the correct stage of the spinning process (Andersson *et al*, 2014; Andersson *et al*, 2017; Rising and Johansson, 2015). The N-terminal domain restricts the formation of silk fibres to a precise point in the silk duct, preventing silk proteins stored in the silk gland from agglutinating (Askarieh *et al*, 2010). The C-terminal domain drives spontaneous fibre formation, likely through use of a pH-sensitive 'salt bridge' (Ittah *et al*, 2006; Stark *et al*, 2007), where noncovalent interactions between one basic and one acidic residue are disrupted at low pH because the acidic residue becomes protonated and is no longer charged.

Salt bridges, either individual or paired, have been proposed to explain the dimeric bundling of 4 or 5 alpha helices in the C-termini of one particular silk, the major ampullate (Hagn *et al*, 2010; Sponner *et al*, 2004; Sponner *et al*, 2005). Whilst the miniature spidroin 4RepCT, formed of 4 copies of a MaSp repetitive region and one C-terminus, has been sufficient to produce self-assembling silk

fibres (Stark *et al*, 2007), it has recently been shown in the minispidroin NT2RepCT that including the N-terminal region achieves greater efficiency in the production of synthetic fibres (Andersson *et al*, 2017). However, studies have shown that the level of solubility demonstrated by the terminal regions differs between spider species, leading to further questions around their function and the suitability of individual terminal regions for use in silk protein synthesis (Andersson *et al*, 2014; Andersson *et al*, 2017; Askarieh *et al*, 2010).

A degree of amino acid sequence conservation has been observed from studies of small numbers of different silks (Beckwitt and Arcidiacono, 1994; Challis *et al*, 2006; Collin *et al*, 2016; Gnesa *et al*, 2012; Hagn *et al*, 2010; Sponner *et al*, 2004; Sponner *et al*, 2005). What is not clear is the extent to which this conservation is maintained, particularly where silk proteins exhibit vastly different properties (e.g. the glue-like aggregate and piriform silk versus the superior strength of major ampullate and aciniform silks). Additionally, the diving bell spider *Argyroneta aquatica* spins silk whilst completely submersed in fresh water; given how silk proteins are dehydrated as part of the spinning process, does this “extreme” environment necessitate variation within the silk protein and spinning process?

Here, we analyse the genetic sequences silks of all known types from spiders of as many groups as are available in GenBank. We investigate whether the salt bridge structure is conserved across all species and silk types and ask if changes in biophysical properties or the utilisation of silk in an “extreme” environment necessitates a change in how silk proteins are formed.

Materials and Methods

Transcriptome assembly

RNA was extracted from the silk glands and whole abdomen of adult females and sequenced on an Illumina NextSeq500 (DeepSeq, University of Nottingham). Transcriptomes were trimmed using Scythe (Buffalo, 2014) and Sickle (Joshi and Fass, 2011) and assembled using Trinity (Grabherr *et al*, 2011). Protein sequences were predicted using TransDecoder (Haas *et al*, 2013).

Identification of silk genes

Custom blast databases of silk genes downloaded from GenBank and the *Nephila clavipes* (Babb *et al*, 2017) and *Stegodyphus mimosarum* genomes (Sanggaard *et al*, 2014) were generated to screen transcriptomes for silk sequences, which were manually examined and blasted against GenBank for confirmation (e-value <0.0005).

Phylogenetic analysis

Silk sequences from GenBank, the *N. clavipes* and *S. mimosarum* genomes and newly-identified spidroins were filtered to only include C-terminal regions, identified by the "QALLE" motif (Challis *et al*, 2006). Duplicate sequences were discarded, leaving the longest sequence for analysis. Two sequences were considered identical if their nucleotide sequences had only a small number of variations which did not significantly alter the amino acid composition (e.g. variation causing a polar residue to change to acidic was considered significant,

whereas polar to polar was not). Selected sequences were manually trimmed leaving only the C-terminal region (maximum length: 360bp (Hagn *et al*, 2010; Rising and Johansson, 2015)).

Selected nucleotide sequences were aligned in Geneious v8.1.8 (Kearse *et al*, 2012) (Clustal W algorithm (Larkin *et al*, 2007)) and subsequently translated in order to allow refined alignment by eye. The refined untranslated nucleotide sequence alignment was imported into MEGA6.0 (Tamura *et al*, 2013) and used to construct a maximum-likelihood tree (1000 bootstrap replicates; see supplementary file 2).

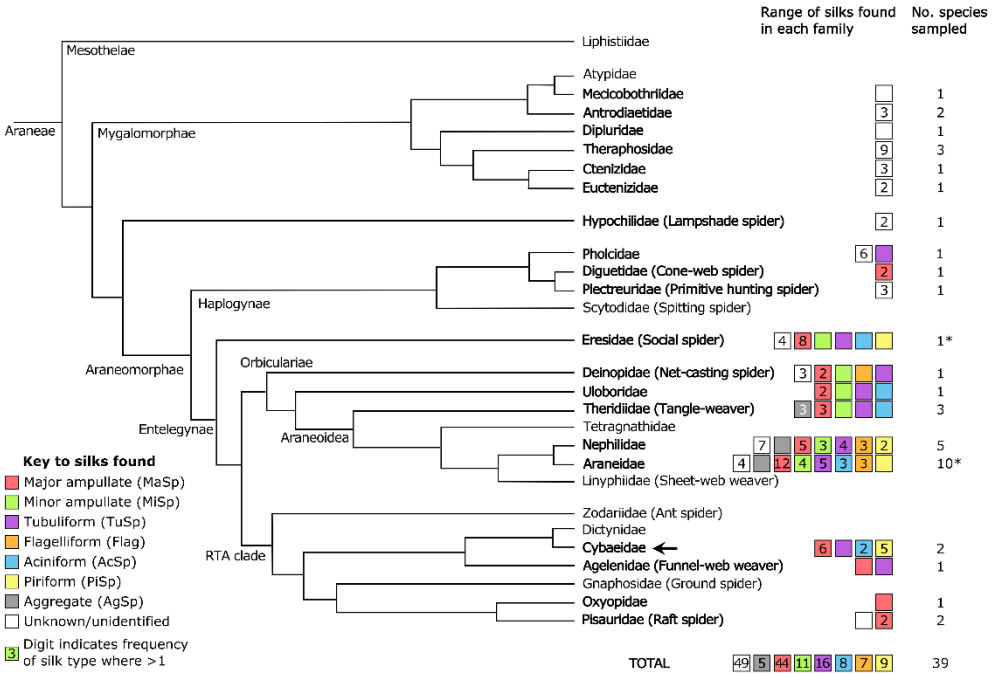
Results

Diversity of silk C-termini

44 new silk sequences isolated from transcriptomes of the spiders *A. aquatica*, *C. angustiarum* and *Pholcus phalangioides* included a C-terminal domain (this study; GenBank Accession Numbers MG744694 – MG744714). Once refined (see methods), 21 unique sequences remained for analysis. Three hundred and thirty-four silk sequences were retrieved from GenBank and the *N. clavipes* and *S. mimosarum* genomes, of which 150 sequences were both unique and contained a C-terminal region. Specialised silks were mostly found in the Orbiculariae and *S. mimosarum* genome, with the exception of one tubuliform and a small number of major ampullate sequences (Babb *et al*, 2017; Garb *et al*, 2010; Perez-Rigueiro *et al*, 2010; Rising *et al*, 2007; Sanggaard *et al*, 2014; Stark *et al*, 2007) (Fig. 1). An unrooted maximum-likelihood tree of nucleotide sequences (1000 bootstraps)

147 shows the clustering together of sequences that are identified as belonging to the

148 same silk type by BLAST (see supplementary figure 2).



149

150 *Fig. 1*

151 *Familial origins of silk sequences used in this study.*

152 *Phylogenetic tree from Nentwig (Nentwig et al, 2013), modified to show positions*

153 *of the different spider families and the silks previously described as being of a*

154 *particular type. Location of* Argyroneta aquatica *and* Cybaeus angustiarum

155 *(Cybaeidae) marked with an arrow.*

156 *Each coloured box represents a silk type; uncoloured boxes represent unclassified*

157 *published spidroins. Numbers in each box indicate the total number of occurrences*

158 *of each silk type. The total number of species sampled per family is detailed at the*

159 *end of each row; * indicates where sequences have been retrieved from a genome.*

160

Conservation of a salt bridge structure across spider silk genes

Analysis of the predicted secondary structures of sequences representing each silk type with JPred4 (Drozdetskiy *et al*, 2015) suggests the secondary structure of the C-terminus is conserved (data not shown), containing up to five α-helices, supporting previous studies (Hagn *et al*, 2010; Ittah *et al*, 2007). However, some predictions suggest Helix 1 may be absent, including in the *N. clavipes* MaSp-a sequence analysed here (Fig.2).

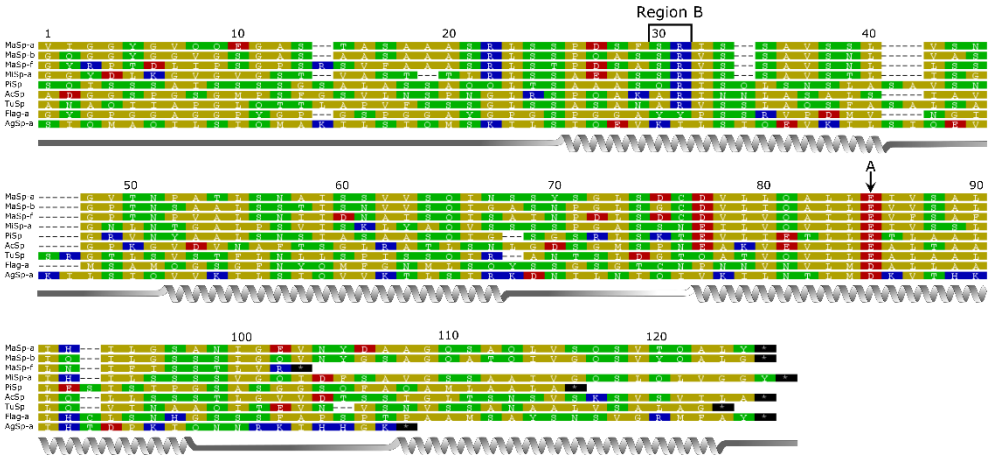


Fig. 2

Alignment of Nephila clavipes *C-terminal sequences representing characterised silk types, with predicted helical secondary structure of MaSp-a below.*

Amino acid residues implicated in the formation of a salt bridge are marked A (acidic residue; glutamic or aspartic acid) and found within region B (basic residue; arginine or lysine). Residues are highlighted to show whether acidic (red), basic (blue), hydrophilic (green) or hydrophobic (yellow).

In all sequences analysed in this study, one acidic residue and one basic residue pair has been conserved. A conserved basic residue is consistently found in Helix 2 (position 30-31 Fig. 2, 23- 37 supplementary figure 1), whilst an acidic base is always found towards the centre of Helix 4 (position 85 in Fig. 2; 88 in supplementary figure 1). Additionally, in major ampullate, minor ampullate, aciniform and piriform sequences a second acidic and basic residue appears to be partially conserved (supplementary figure 1) as identified by Sponner et al. (Sponner *et al*, 2004; Sponner *et al*, 2005). These pairs correspond with those already shown to form a salt bridge in the secondary protein structure in previous studies (Hagn *et al*, 2010; Ittah *et al*, 2006).

Content of the C-terminal domain

The majority of the C-terminus is composed of hydrophobic and hydrophilic amino acids (average 53.5% and 37.0% respectively) including serine, alanine and leucine (average 21.2%, 13.2%, and 10.8% respectively) in an alternating arrangement. These residues promote the formation of alpha-helices in the protein structure, with the hydrophilic residues exposed, promoting solubility (Hagn *et al*, 2010).

Overall DNA sequence identity is highest in the acidic residue found in the "QALLE" motif (Challis *et al*, 2006), where the first two nucleotides (guanine, adenine) are completely conserved and the final nucleotide determines whether it is a glutamic or aspartic acid residue (89.3% and 10.7% respectively of all sequences in this study; see supplementary figure 1). The basic residues involved

199 in the formation of the salt bridge are usually arginine, although some are lysine
200 or histidine (94.7%, 3.3% and 2.0% respectively).

201 Overall, the charged residue content of the C-terminus is typically less than 10%
202 (5.2% acidic, 4.1% basic; average across all sequences in this study) although this
203 varies depending on the sequence (aggregate spidroins average 13% acidic, 8%
204 basic). However, one acidic and basic residue pair (A and region B, respectively,
205 Fig. 2 and supplementary figure 1) are present in each sequence examined,
206 irrespective of species or silk type.

207 Additionally, we find a raised percentage of the amino acids surrounding the basic
208 residue within spider silk are hydrophilic (43.0%, region B, supplementary figure
209 1), reducing its pKa and hence promoting its protonation, whereas a majority of
210 those surrounding the acidic residue are hydrophobic (68.8%, region A,
211 supplementary figure 1), increasing its pKa and again promoting protonation.

212 Discussion

213 This analysis encompasses 19 families from within the Araneae and shows
214 conservation of a pH-sensitive salt bridge, typically composed of an arginine-
215 glutamic acid pairing, within all the silk types and species examined. Our finding of
216 this degree of conservation confirms an essential role for this feature in the
217 correct assembly of silk fibres. The extended coverage in terms of species and
218 phylogenetic diversity suggests this feature has been conserved for the entire
219 evolutionary history of the group - some 360 million years. Moreover, this
220 conservation has persisted despite the diversification of other spider traits. For

example, terrestrial spiders such as the mygalomorphs produce two to three silk proteins from a large, undifferentiated silk gland whereas the Orbiculariae have evolved several specialised glands from which different types of silk are extruded (Blackledge and Hayashi, 2006; Blamires *et al*, 2017; Clarke *et al*, 2017; Garb, 2013; Garb *et al*, 2010; Perez-Rigueiro *et al*, 2010; Rising *et al*, 2007; Stark *et al*, 2007).

Conservation of C-terminus in all *A. aquatica* silk genes

Argyroneta aquatica predominantly spins its silk in water, although will occasionally leave a dragline when walking in a terrestrial environment. Comparatively, individual *Cybaeus* did not appear to produce a dragline but given time would spin silken webs in which they resided, which appear visually similar to the dehydrated diving bells of *A. aquatica*. In sequences obtained from both these species, and also from the first fully-annotated spider genome of *Nephila clavipes* (Babb *et al*, 2017) in which a number of new silk sequences were identified, the residue pairing within the C-terminal domain has been conserved. This suggests that the formation of silk fibres by *A. aquatica* occurs using the same method as in terrestrial spiders. This leads to questions around how this spider is adapted to spinning silk underwater and the subsequent mechanical properties of *A. aquatica* silk in such an “extreme” environment. This is particularly relevant as studies of silk in humid conditions suggest the structure may be temporarily affected by the level of humidity; the degree to which major ampullate silk supercontracts depends on the species, whereas minor ampullate silk does not

supercontract (Agnarsson *et al*, 2009; Blackledge *et al*, 2009; Boutry and Blackledge, 2010).

Conservation of a physical structure

The residues responsible for the formation of the pH-sensitive salt bridge(s) are both uncommon and conserved within the restricted number of sequences studied thus far (Challis *et al*, 2006; Hagn *et al*, 2010). Analysis of the expanded range of sequences used in this study shows this trait is maintained, with >90% of the C-terminal domain typically composed of hydrophobic and hydrophilic residues. Of the charged residues present, at least one acidic and one basic residue is conserved in all sequences and as such inferred to be crucial to the formation of a salt bridge and therefore the correct folding of silk proteins into a fibre. The conservation of an arginine – glutamic acid pair of residues suggests this may be the optimum pairing for a salt bridge in silk, but the presence of lysine and aspartic acid in some sequences implies this is not always essential, although the effect of these substitutions on the final protein structure and its physical properties is currently unknown.

The higher level of hydrophilic residues around the basic residue and hydrophobic residues around the acidic residue may be necessary to ensure the correct formation of the salt bridge during the protein folding stage and allowing its regulation by smaller changes in pH due to the local environment created by the protein. Where two pairs of residues are conserved in major ampullate, minor ampullate, piriform and some aciniform sequences it is likely that there are two

salt bridges present in the C-terminus, although further structural analysis would confirm that the second is not merely sequence duplication or an evolutionary artefact suggesting the presence of a second bridge.

Conclusion

In silk research, the C-terminus is a key component of the minimal sequence used for recombinant spider silk production, as without this region fibres will not form (Stark *et al*, 2007). What we find is an overall model for all silk types irrespective of physical traits that illustrates the range of environments in which this single protein family may be utilised.

Our results allow us to identify conserved amino acid residues essential to the correct formation of silk proteins, thereby enabling the identification of less essential residues that may be chemically functionalised in artificially synthesised silk using techniques such as click-chemistry (Harvey *et al*, 2017). This study will aid researchers in selecting suitable sequences without repetitive testing *in vivo* whilst predicting sites which may be suitable for modification, as seen in Harvey *et al*. (Harvey *et al*, 2017), and build a repository of spidroin parts that could be combined to achieve novel and custom characteristics.

Summary

Here we compare C-terminal sequences representing all known silk types, and from all the major clades within the Araneae. We illustrate how the salt bridge structure is conserved throughout the gene family as a whole, and show by inclusion of new data that this conservation extends to silks found in all the major

287 divisions within the Araneae. This degree of conservation includes groups that are
288 highly diverged in terms of their basic morphology, such as their silk spinning
289 apparatus (Coddington, 2005). We include new data on silk of the Cybaeid
290 *Argyroneta aquatica*, the only extant species adapted to spin silk in an
291 underwater environment, and a terrestrial Cybaeid for comparison, *Cybaeus*
292 *angustiarum*. The insight gained allows us to characterise amino acid residues
293 essential to correct protein assembly and aids the identification of residues that
294 might be substituted to chemically functionalised alternatives for use in artificially
295 synthesised silk proteins (Harvey *et al*, 2017).

296

297 Declarations

298 Availability of data and materials

299 The novel sequences generated in this study are deposited in the GenBank
300 repository, Accession Numbers MG744694 – MG744714
301 (https://www.ncbi.nlm.nih.gov/genbank/).

302 Electronic supplementary material

303 The online version of this article (https://doi.org/10.1038/s41437-018-0050-9)
304 contains supplementary material, which is available to authorized users.

Acknowledgements

Spider collection and dissection by MR. RNA extracted by VT and MS, transcriptomics and analysis by MS, with genomics assistance by James Fowkes. All living authors read and approved the final manuscript.

Sequencing and data quality control by Sunir Malla (DeepSeq facility, University of Nottingham). In memory of Victor Tudorica (1990 – 2016).

Funding

This work was supported by the Royal Society [Grant Number IE141413 to SLG and MR], Biotechnology and Biological Sciences Research Council doctoral training studentships to MS and VT [Grant Number BB/M008770/1] and the Ministry of Agriculture of the Czech Republic [Grant Number MZe Ro0415 to MR].

Competing interests

The authors declare that they have no conflict of interest.

References

Agnarsson I, Boutry C, Wong SC, Baji A, Dhinojwala A, Sensenig AT *et al* (2009). Supercontraction forces in spider dragline silk depend on hydration rate. *Zoology* **112**(5)**:** 325-331.

Andersson M, Chen G, Otikovs M, Landreh M, Nordling K, Kronqvist N *et al* (2014). Carbonic Anhydrase Generates CO(2) and H(+) That Drive Spider Silk Formation Via Opposite Effects on the Terminal Domains. *PLoS Biology* **12**(8)**:** e1001921.

Andersson M, Jia Q, Abella A, Lee X-Y, Landreh M, Purhonen P *et al* (2017). Biomimetic spinning of artificial spider silk from a chimeric minispidroin. *Nature Chemical Biology* **13:** 262.

Askarieh G, Hedhammar M, Nordling K, Saenz A, Casals C, Rising A *et al* (2010). Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature* **465**(7295)**:** 236-238.

Babb PL, Lahens NF, Correa-Garhwal SM, Nicholson DN, Kim EJ, Hogenesch JB *et al* (2017). The Nephila clavipes genome highlights the diversity of spider silk genes and their complex expression. *Nature Genetics* **49:** 895.

Beckwitt R, Arcidiacono S (1994). Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae). *Journal of Biological Chemistry* **269**(9)**:** 6661-6663.

Blackledge TA, Boutry C, Wong SC, Baji A, Dhinojwala A, Sahni V *et al* (2009). How super is supercontraction? Persistent versus cyclic responses to humidity in spider dragline silk. *J Exp Biol* **212**(Pt 13)**:** 1981-1989.

Blackledge TA, Hayashi CY (2006). Silken toolkits: biomechanics of silk fibers spun by the orb web spider Argiope argentata (Fabricius 1775). *Journal of experimental biology* **209 (Pt 13):** 2452-2461.

Blamires SJ, Blackledge TA, Tso I-M (2017). Physicochemical Property Variation in Spider Silk: Ecology, Evolution, and Synthetic Production. *Annual Review of Entomology* **62**(1)**:** 443-460.

Boutry C, Blackledge TA (2010). Evolution of supercontraction in spider silk: structure-function relationship from tarantulas to orb-weavers. *J Exp Biol* **213**(Pt 20)**:** 3505-3514.

Buffalo V (2014). Scythe - A Bayesian adapter trimmer (version 0.994 BETA) [Software].

Catalog WS (2016). World Spider Catalog, version 17.

Challis RJ, Goodacre SL, Hewitt GM (2006). Evolution of spider silks: conservation and diversification of the C-terminus. *Insect Mol Biol* **15**(1)**:** 45-56.

368 Clarke TH, Garb JE, Haney RA, Chaw RC, Hayashi CY, Ayoub NA (2017).
369 Evolutionary shifts in gene expression decoupled from gene duplication across
370 functionally distinct spider silk glands. *Scientific Reports* **7**(1)**:** 8393.

371
372 Clarke TH, Garb JE, Hayashi CY, Arensburger P, Ayoub NA (2015). Spider
373 Transcriptomes Identify Ancient Large-Scale Gene Duplication Event Potentially
374 Important in Silk Gland Evolution. *Genome biology and evolution* **7**(7)**:** 1856-1870.

375
376 Clerck CA (1757). Svenska Spindlar uti sina hufvud-slågter indelte samt under
377 några och sextio särskildte arter beskrefne och med illuminerade figurer uplyste /
378 Aranei Svecici, descriptionibus et figuris æneis illustrati, ad genera subalterna
379 redacti, speciebus ultra LX determinati.

380
381 Coddington JA (2005). Phylogeny and Classification of Spiders. In: Ubick D, Paquin
382 P, Cushing PE and Roth V (eds) *Spiders of North America: an identification manual*.
383 American Arachnological Society: Poughkeepsie, New York, USA.

384
385 Collin MA, Clarke TH, Ayoub NA, Hayashi CY (2016). Evidence from Multiple
386 Species that Spider Silk Glue Component ASG2 is a Spidroin. *Sci Rep* **6:** 21589.

387
388 Drozdetskiy A, Cole C, Procter J, Barton GJ (2015). JPred4: a protein secondary
389 structure prediction server. *Nucleic acids research* **43**(W1)**:** W389-W394.

390
391 Eisoldt L, Smith A, Scheibel T (2011). Decoding the secrets of spider silk. *Materials*
392 *Today* **14**(3)**:** 80-86.

393
394 Garb JE (2013). Spider silk: an ancient biomaterial for the 21st century. In: Penney
395 D (ed) *Spider research in the 21st century: trends and perspectives*. Siri Scientific
396 Press: Manchester, UK, pp 252-281.

397
398 Garb JE, Ayoub NA, Hayashi CY (2010). Untangling spider silk evolution with
399 spidroin terminal domains. *BMC evolutionary biology* **10:** 243.

400
401 Gnesa E, Hsia Y, Yarger JL, Weber W, Lin-Cereghino J, Lin-Cereghino G *et al* (2012).
402 Conserved C-terminal domain of spider tubuliform spidroin 1 contributes to
403 extensibility in synthetic fibers. *Biomacromolecules* **13**(2)**:** 304-312.

404

405 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I *et al* (2011). Full-
406 length transcriptome assembly from RNA-Seq data without a reference genome.
407 *Nat Biotech* **29**(7)**:** 644-652.

408
409 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J *et al* (2013).
410 De novo transcript sequence reconstruction from RNA-Seq: reference generation
411 and analysis with Trinity. *Nature protocols* **8**(8)**:** 10.1038/nprot.2013.1084.

412
413 Hagn F, Eisoldt L, Hardy JG, Vendrely C, Coles M, Scheibel T *et al* (2010). A
414 conserved spider silk domain acts as a molecular switch that controls fibre
415 assembly. *Nature* **465**(7295)**:** 239-242.

416
417 Harvey D, Bardelang P, Goodacre SL, Cockayne A, Thomas NR (2017). Antibiotic
418 Spider Silk: Site-Specific Functionalization of Recombinant Spider Silk Using "Click"
419 Chemistry. *Advanced materials* **29**(10)**:** 1604245-n/a.

420
421 Hayashi CY, Lewis RV (1998). Evidence from flagelliform silk cDNA for the
422 structural basis of elasticity and modular nature of spider silks1. *Journal of*
423 *molecular biology* **275**(5)**:** 773-784.

424
425 Hayashi CY, Shipley NH, Lewis RV (1999). Hypotheses that correlate the sequence,
426 structure, and mechanical properties of spider silk proteins. *International Journal*
427 *of Biological Macromolecules* **24**(2–3)**:** 271-275.

428
429 Hinman MB, Jones JA, Lewis RV (2000). Synthetic spider silk: a modular fiber.
430 *Trends in biotechnology* **18**(9)**:** 374-379.

431
432 Hormiga G, Griswold CE (2014). Systematics, Phylogeny, and Evolution of Orb-
433 Weaving Spiders. *Annual Review of Entomology* **59**(1)**:** 487-512.

434
435 Ittah S, Cohen S, Garty S, Cohn D, Gat U (2006). An Essential Role for the C-
436 Terminal Domain of A Dragline Spider Silk Protein in Directing Fiber Formation.
437 *Biomacromolecules* **7**(6)**:** 1790-1795.

438
439 Ittah S, Michaeli A, Goldblum A, Gat U (2007). A Model for the Structure of the C-
440 Terminal Domain of Dragline Spider Silk and the Role of Its Conserved Cysteine.
441 *Biomacromolecules* **8**(9)**:** 2768-2773.

442

443 Joshi N, Fass J (2011). Sickle: A sliding-window, adaptive, quality-based trimming
444 tool for FastQ files (Version 1.33) [Software].

445
446 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S *et al* (2012).
447 Geneious Basic: An integrated and extendable desktop software platform for the
448 organization and analysis of sequence data. *Bioinformatics* **28**(12)**:** 1647-1649.

449
450 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al*
451 (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21)**:** 2947-2948.

452
453 Motriuk-Smith D, Smith A, Hayashi CY, Lewis RV (2005). Analysis of the conserved
454 N-terminal domains in major ampullate spider silk proteins. *Biomacromolecules*
455 **6**(6)**:** 3152-3159.

456
457 Nentwig W, Arnedo MA, Coddington JA, Hormiga G, Jocqué R, Kropf C *et al* (2013).
458 Appendix. Spider Phylogeny. In: Nentwig W (ed) *Spider Ecophysiology*. Springer-
459 Verlag: Berlin.

460
461 Perez-Rigueiro J, Plaza GR, Torres FG, Hijar A, Hayashi C, Perea GB *et al* (2010).
462 Supercontraction of dragline silk spun by lynx spiders (Oxyopidae). *Int J Biol*
463 *Macromol* **46**(5)**:** 555-557.

464
465 Rising A, Hjalm G, Engstrom W, Johansson J (2006). N-terminal nonrepetitive
466 domain common to dragline, flagelliform, and cylindriform spider silk proteins.
467 *Biomacromolecules* **7**(11)**:** 3120-3124.

468
469 Rising A, Johansson J (2015). Toward spinning artificial spider silk. *Nat Chem Biol*
470 **11**(5)**:** 309-315.

471
472 Rising A, Johansson J, Larson G, Bongcam-Rudloff E, Engström W, Hjälm G (2007).
473 Major ampullate spidroins from Euprosthenops australis: multiplicity at protein,
474 mRNA and gene levels. *Insect Molecular Biology* **16**(5)**:** 551-561.

475
476 Sanggaard KW, Bechsgaard JS, Fang X, Duan J, Dyrlund TF, Gupta V *et al* (2014).
477 Spider genomes provide insight into composition and evolution of venom and silk.
478 *Nature Communications* **5:** 3765.

479

480 Schütz D, Taborsky M (2003). Adaptations to an aquatic life may be responsible
481 for the reversed sexual size dimorphism in the water spider, Argyroneta aquatica.
482 *Evolutionary Ecology Research* **5:** 105 - 117.

483
484 Schütz D, Taborsky M (2005). MATE CHOICE AND SEXUAL CONFLICT IN THE SIZE
485 DIMORPHIC WATER SPIDER ARGYRONETA AQUATICA (ARANEAE,
486 ARGYRONETIDAE). *Journal of Arachnology* **33**(3)**:** 767-775.

487
488 Schütz D, Taborsky M (2011). Sexual Selection in the Water Spider: Female Choice
489 and Male-Male Competition. *Ethology* **117**(12)**:** 1101-1110.

490
491 Schutz D, Taborsky M, Drapela T (2007). Air bells of water spiders are an extended
492 phenotype modified in response to gas composition. *Journal of experimental*
493 *zoology Part A, Ecological genetics and physiology* **307**(10)**:** 549-555.

494
495 Seymour RS, Hetz SK (2011). The diving bell and the spider: the physical gill of
496 Argyroneta aquatica. *The Journal of Experimental Biology* **214**(13)**:** 2175-2181.

497
498 Sponner A, Unger E, Grosse F, Weisshart K (2004). Conserved C-Termini of
499 Spidroins Are Secreted by the Major Ampullate Glands and Retained in the Silk
500 Thread. *Biomacromolecules* **5**(3)**:** 840-845.

501
502 Sponner A, Vater W, Rommerskirch W, Vollrath F, Unger E, Grosse F *et al* (2005).
503 The conserved C-termini contribute to the properties of spider silk fibroins.
504 *Biochemical and Biophysical Research Communications* **338**(2)**:** 897-902.

505
506 Stark M, Grip S, Rising A, Hedhammar M, Engström W, Hjälm G *et al* (2007).
507 Macroscopic Fibers Self-Assembled from Recombinant Miniature Spider Silk
508 Proteins. *Biomacromolecules* **8**(5)**:** 1695-1701.

509
510 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: molecular
511 evolutionary genetics analysis version 6.0. *Molecular biology and evolution* **30**(12)**:**
512 2725-2729.

513
514 Vollrath F, Knight DP (2001). Liquid crystalline spinning of spider silk. *Nature*
515 **410**(6828)**:** 541-548.

516

517