Evidence from Multiple Species that Spider Silk Glue Component ASG2 is a Spidroin

Matthew A. Collin1, Thomas H. Clarke III1,2, Nadia A. Ayoub2 & Cheryl Y. Hayashi1

Spiders in the superfamily Araneoidea produce viscous glue from aggregate silk glands. Aggregate glue coats prey-capture threads and hampers the escape of prey from webs, thereby increasing the foraging success of spiders. cDNAs for Aggregate Spider Glue 1 (ASG1) and 2 (ASG2) have been previously described from the golden orb-weaver, Nephila clavipes, and Western black widow, Latrodectus hesperus. To further investigate aggregate glues, we assembled ASG1 and ASG2 from genomic target capture libraries constructed from three species of cob-web weavers and three species of orb-web weavers, all araneoids. We show that ASG1 is unlikely to be a glue, but rather is part of a widespread arthropod gene family, the peritrophic matrix proteins. For ASG2, we demonstrate its remarkable architectural and sequence similarities to spider silk fibroins, indicating that ASG2 is a member of the spidroin gene family. Thus, spidroins have diversified into glues in addition to task-specific, high performance fibers.

Spiders have evolved numerous types of webs for prey capture that vary in geometry, mechanism, and constituent proteins. One of the most well-studied capture webs is the wagon-wheel shaped orb-web spun by spiders in the superfamily Araneoidea. Dominant architectural elements in these webs are the frame, radii, and the sticky capture spiral. Spider silks are synthesized in abdominal silk glands, and the silk glands of araneoid spiders are variously specialized to produce a particular type of silk. For example, major ampullate silk glands are the source of major ampullate fibers that are used in the frame and radii. Similarly, flagelliform silk glands produce the filament of the capture spiral while aggregate silk glands secrete sticky aggregate glue.

Each silk gland is connected to its own extruding spigot on the exterior of the spider’s spinnerets. The spigots for flagelliform and aggregate silk glands are adjacent to each other on the posterior lateral spinnerets, and a spider secretes both silks simultaneously when spinning the capture spiral1. Aggregate silk beads into evenly spaced droplets along the length of the flagelliform silk fibers2. When an insect impacts a web, sticky droplets behave as viscoelastic solids that not only adhere to the prey, but also elongate to maintain a physical connection between the struggling prey and the web3,4. For example, a single viscous droplet from the orb-weaving spider, Larinioides cornutus, can withstand up to 400 μN of pull off force and extend up to 800 μM3. However, the adhesive capability of spider webs is even greater as the close spacing of the droplets along the length of the capture spiral filament ensures that multiple droplets contact the prey. Each additional contacting viscous droplet increases the force necessary for an insect to pull free from the web5.

Compound light microscopy images of the capture spiral reveal three “layers” of viscid silk within a viscous droplet: a core, an aqueous layer, and an outer layer. The inner core is a fibrous nodule of aggregate gland secretions that encase flagelliform silk fibers3,6. It is thought that the nodule transfers the force of the captured prey to the axial fibers1. The aqueous middle layer is a sticky globe of glycoproteins, low molecular weight ionic compounds, and neurotransmitters (e.g. choline, isethionic acid, GABamide), collectively called “salts,” that coats the nodule2. The hygroscopic salts absorb ambient humidity, which allows the glycoproteins to remain mobile, thereby enhancing the adhesive properties of droplets2,7. Additionally, the collected water keeps the flagelliform fibers pliable8. The outer layer is characterized as being more fluid than the inner glycoprotein layer and contains fatty acids that may attract prey3.

Aggregate silk gland secretions have been greatly understudied at the genetic level. In fact, aggregate silk gland complementary (c) DNAs and proteins have been characterized from only two species, the golden orb-weaver Nephila clavipes and Western black widow Latrodectus hesperus. For ASG2, we demonstrate its remarkable architectural and sequence similarities to spider silk fibroins, indicating that ASG2 is a member of the spidroin gene family. Thus, spidroins have diversified into glues in addition to task-specific, high performance fibers.

1University of California, Riverside, Department of Biology, Riverside, California 92521, United States. 2Washington and Lee University, Department of Biology, Lexington, Virginia 24450, United States. Correspondence and requests for materials should be addressed to M.A.C. (email: matthew.collin@ucr.edu)
orb-weaver *Nephila clavipes* and the Western black widow *Latrodectus hesperus*, which builds three-dimensional cob webs.\(^{10,11}\) The first reported aggregate gland cDNAs from *N. clavipes* were considered to be full-length and proposed to encode sticky components of the viscous glue droplets, aggregate spider glue 1 (ASG1) and 2 (ASG2).\(^7\) Both the ASG1 and ASG2 amino acid sequences are predicted to have several glycosylation sites, consistent with the observation that *N. clavipes* viscous glue droplets contain glycosylated proteins.\(^9\)

An unusual feature of the *N. clavipes* ASG1 and ASG2 transcripts is that they contain a 353 base pair (bp) repetitive region that is present in ASG1 in one orientation and present in ASG2 in the reverse complement orientation.\(^5\) The repetitive region of ASG1 encodes iterations of an amino acid sequence motif that is dominated by glutamic acid, threonine and proline. This protein region is analogous to vertebrate mucin domains, which are glycosylated.\(^9,12\) In ASG2, the reverse complement of the 353 bp repetitive region translates into iterated motifs that are rich in serine, glycine and valine. The related sequences in ASG1 and ASG2 were interpreted as evidence for, "two proteins expressed from opposite strands of the same DNA sequence".\(^9\) This suggests that the 353 bp region could be derived from a common exon that is transcribed in opposite directions (i.e., overlapping genes).\(^13\)

Later, a study of *L. hesperus* aggregate glands described cDNAs for aggregate silk factors 1 (AgSF1) and 2 (AgSF2), which function to cement web fibers and wrap prey.\(^10\) Unlike ASG1 and ASG2, AgSF1 and AgSF2 are not thought to form glue droplets. Subsequently, a transcriptome study of *L. hesperus* silk glands identified homologs to the *N. clavipes* ASG1 and ASG2.\(^11\) The *L. hesperus* ASG1 transcript has sequence that corresponds to the *N. clavipes* 353 bp region, but the *L. hesperus* ASG2 transcript is devoid of any matching sequence to that region, in any orientation. The lack of shared sequence between *L. hesperus* ASG1 and ASG2 transcripts, as described for the *N. clavipes* ASG1 and ASG2 cDNAs, may be due to alternative splicing, misassembled transcripts for either *N. clavipes* or *L. hesperus*, or a taxon specific insertion/deletion event.

To assess the aggregate glue genes of cob-web and orb-web weaving spiders, we characterized ASG1 and ASG2 from the genomes of multiple species. Furthermore, gene sequences can reveal exon-intron structures, particularly in relation to the 353 bp region that is present in opposite orientations in the published *N. clavipes* ASG1 and ASG2 cDNAs, but absent from *L. hesperus* ASG1 and ASG2.\(^11\) To identify homologs of ASG1 and ASG2, we constructed genomic target capture next generation libraries. We then conducted comparative analyses to identify key features of spider silk glue.

**Results and Discussion**

**ASG1 has Introns.** Multiple (4–11) contigs that match regions of published ASG1 were identified from each target capture species (Table 1). For *N. clavipes*, none of the contigs contains an entire ASG1 coding region. Instead, the multiple contigs represent exons and portions of flanking introns (Fig. 1a). These contigs are predicted to have exon acceptor and/or donor splice junctions with probabilities greater than 60%. Genomic contigs of exons two through six match the cDNA with 94% sequence identity along the 1,184 bp length of concatenated exons, indicating that the cDNA and genomic contigs are correctly assembled. Furthermore, given the correspondence to our genomic contigs, the cDNA available for *N. clavipes* ASG1 contains only 15 bp of what we now know is exon 1 and no 5' UTR. The longest *N. clavipes* ASG1 genomic contig corresponds to the sixth exon, which is the final and longest exon, coding for about half of the mRNA transcript. This exon includes a chitin binding domain and the 353 bp repetitive region that was reported to be shared by ASG2, but in the reverse complement orientation.\(^5\) However, our genomic contigs show that the 353 bp region is not isolated in a separate exon. Furthermore, if reverse complemented, the coding region of the final exon of ASG1 extends beyond the ASG2 repetitive region and codes for an additional 34 amino acids that is not present in the ASG2 cDNA. Thus *N. clavipes* ASG1 and ASG2 cannot be overlapping genes.

For *L. hesperus*, contigs representing all six exons were recovered. When concatenated, these exons support the full-length cDNA transcript (Fig. 1b; 92% sequence identity over 1,152 bp). The final exon in *L. hesperus* is 603 bp, representing about 50% of the mRNA transcript length. This exon is shorter than the corresponding *N. clavipes* exon because it has fewer iterated repeats in the repetitive region.

**Figure 1.** Comparison of ASG1 gene models to cDNAs for *N. clavipes* (a) and *L. hesperus* (b). Boxes represent exons and horizontal lines represent introns. Exons are numbered. Lines show the correspondence of exons to cDNA regions. Slashes indicate unknown sequence between exons. Inverted triangles represent the mucin-like domain, and purple bars represent chitin binding domain 2 (see text). Exon 1 of *N. clavipes* ASG1 was not captured, but is predicted from the cDNA. GenBank accession numbers are in Table 1.
**Table 1. Description of ASG1 sequences. Entries in blue are from this study.**

| Abbreviated Sequence Name | Species            | Description                  | Accession     |
|---------------------------|--------------------|------------------------------|---------------|
| A. dia var. 1             | Araneus diadematus | variant 1, final exon, complete cds | KU132336      |
| A. dia var. 2             | Araneus diadematus | variant 2, partial cds      | KU132337      |
| A. arg var. 1             | Aranea argentata   | variant 1                  | KU132338      |
| A. arg var. 2             | Aranea argentata   | variant 2, final exon, complete cds | KU132339      |
| A. arg var. 3             | Aranea argentata   | variant 3, final exon, complete cds | KU132340      |
| L. geo 3f004239g1u        | Latrodectus geometricus | cDNA | GBJM01027103 |
| L. geo 3f004239g9u        | Latrodectus geometricus | cDNA | GBJM01128153 |
| L. hes var.1              | Latrodectus hesperus | variant 1, complete cds | KU132335      |
| L. hes var.2              | Latrodectus hesperus | variant 2, partial cds       | KU132341      |
| L. hes var.3              | Latrodectus hesperus | variant 3, final exon, complete cds | KU132342      |
| L. hes contig 682         | Latrodectus hesperus | cDNA | GBCS01010755 |
| L. hes contig 21628       | Latrodectus hesperus | cDNA | GBCS01018793 |
| N. cla var. 1             | Nephila clavipes   | variant 1, exons 2–6, partial cds | KU132334      |
| N. cla var. 2             | Nephila clavipes   | variant 2, final exon, complete cds | KU132343      |
| N. cla var. 3             | Nephila clavipes   | variant 3, partial cds       | KU132344      |
| N. cla cDNA               | Nephila clavipes   | cDNA | EU780014      |
| P. tep var. 1             | Parasteatoda tepidariorum | variant 1, final exon, complete cds | KU132345      |
| P. tep var. 2             | Parasteatoda tepidariorum | variant 2, final exon, complete cds | KU132346      |
| S. gro var. 1             | Steatoda grossa    | variant 1, final exon, complete cds | KU132347      |
| S. gro var. 2             | Steatoda grossa    | variant 2, final exon, complete cds | KU132348      |
| S. gro var. 3             | Steatoda grossa    | variant 3, partial cds       | KU132349      |
| S. gro 3f004239g6u        | Steatoda grossa    | cDNA | GBJQ01082579  |
| S. gro 3f004239g1u        | Steatoda grossa    | cDNA | GBQ01148537   |
| S. mim contig 83103       | Stegodyphus mimosarum | Genomic contig | AZAQ01083103  |
| S. mim contig 57853       | Stegodyphus mimosarum | Genomic contig | AZAQ01057853  |
| S. mim contig 120194      | Stegodyphus mimosarum | Genomic contig | AZAQ01120194  |
| S. mim contig 16388       | Stegodyphus mimosarum | Genomic contig | AZAQ01016388  |

**ASG1 is Multi-locus.** For the other target capture species, only contigs corresponding to ASG1 exon 6 could be definitively identified. All contigs with a complete sixth exon began with the 3′ portion of the preceding intron, followed by a predicted exon acceptor splice junction, coding region, and 3′ UTR sequence. Since the sixth exon is several hundred bases long, numerous target capture probes covered this region, consistent with the recovery of this exon from all species.

Two or three ASG1 exon 6 homologs were recovered per target capture species and additional homologs were identified from NCBI databases (translations shown in Fig. 2). Considering the number and sequence divergence of exon 6 variants within a species, this means that ASG1 is a member of a diversified gene family. For example, *N. clavipes* has three known exon 6 variants, and they share 52–95% pairwise nucleotide sequence identity within an individual spider genome. Because spiders are diploid, there must be at least two ASG1 loci underlying these variants.

A phylogenetic analysis of the conserved carboxyl-terminal region alignment (Fig. 2) reveals the relationships of ASG1 variants within a genome to each other and to previously published sequences (Fig. 3). Focusing on the *N. clavipes* sequences, the ASG1 variants belong to two separate clades. One clade includes the pairing of the published *N. clavipes* cDNA and our *N. clavipes* variant 1 (Figs. 1a and 3 yellow box), as well as *N. clavipes* variant 2. The translated cDNA and genomic variant 1 differ by only a single amino acid replacement and we interpret them as alleles at the same locus. Genomic variants 1 and 2 are 93% identical at the amino acid level (79 out of 85 amino acids, Fig. 2) and have 84% nucleotide identity in the 3′ UTR. Given these similarities, variants 1 and 2 may also be alleles at the same locus. *N. clavipes* variant 3, however, is distantly related to *N. clavipes* variants 1 and 2, sharing a scant 47% average amino acid identity with them (Fig. 2) and 27.9% average nucleotide identity in 3′ UTR. Thus, variant 3 is attributed to a second locus.

**ASG1 is Not Specific to Aggregate Silk Glands.** Homologs of ASG1 were found in the *Stegodyphus mimosarum* genome (Fig. 3, red). *S. mimosarum*, a social velvet spider, is in the family Eresidae, which is distantly related to Araneoidea. Aggregate silk glands are restricted to Araneoidea, thus *S. mimosarum* spiders have neither aggregate silk glands nor are known to produce viscous glue. Instead *S. mimosarum* spiders use silk nanofibrils that are Velcro-like for prey capture. Given their lack of aggregate silk glands and lack of viscous glue use, *S. mimosarum* is not expected to have genes for aggregate silk glues. Nevertheless, ASG1 homologs were identified from the *S. mimosarum* genome and are interspersed with the araneoid ASG1 homologs (Fig. 3). The presence of ASG1 homologs in *S. mimosarum* suggests that ASG1 does not have a function specific to the production of aggregate glue.
Several of our theridiid ASG1 target capture contigs match published cDNA transcripts from *L. hesperus*, *S. grossa*, and *L. geometricus* (Fig. 2)\(^{11,15}\). Each of these cDNAs unequivocally clusters with their corresponding...
Potential Function of ASG1 as a Peritrophic Matrix Protein. To infer the function of ASG1, we examined the alignment of the translated exon 6 from the various homologs (Fig. 2). Invariant amino acids were found across all homologs (highlighted in red, Fig. 2). These amino acids correspond to a conserved chitin binding domain type 2 (ChtBD2). As part of ChtBD2, the six cysteine (C) residues in each sequence are predicted to form three sets of disulfide bonds that link β-sheets within the domain into a chitin binding pocket. Similarly, the conserved aromatic residues (phenylalanine (F) and tryptophan (W) depicted in red, Fig. 2) are predicted to be important for binding chitin.

Immediately upstream of ChtBD2, 21 of the 27 ASG1 homologs have a complete or partial-length region of iterated repeat units dominated by proline (P) and threonine (T) residues, referred to as the mucin-like domain (triangles, Figs. 1 and 2). The mucin-like repeats include other polar amino acids, such as serine, glutamine, or glutamic acid (S, Q, E, respectively, Fig. 2 left panel). The β-turns resulting from the proline residues in mucin-like domains could expose the glycosylated threonine and serine residues by extending them outward, away from bound chitin.

We further investigated the potential function of ASG1 with DELTA-BLAST. DELTA-BLAST search of the nr database with the translated N. clavipes ASG1 cDNA (EU780014) corroborated the previous identification of three chitin binding domains. The top hit after the query itself was insect peritrophic matrix intestinal mucin from the diamondback moth Plutella xylostella (AAN63949; 84% coverage, e-value 7e-49). Peritrophic matrix intestinal mucin proteins bind to chitin to coat anatomical structures, such as the walls and linings of midgut epithelium and glands in insects. Thirty-one copies of peritrophic matrix intestinal mucin genes have been identified within the Tribolium castaneum (red flour beetle) genome, and multiple copies have also been found in other insect genomes. Thus, peritrophic matrix genes comprise a multi-gene family and homologs have been identified within several arthropod lineages outside of insects, including Chelicerae

Insect intestinal mucins, a subtype of peritrophic matrix protein, characterized from Mamestra configurata (Bertha armyworm) contain two to five ChtBD2s and some have mucin domains high in threonine (up to 68.7%), serine (13.5%) and proline (18.1%)%. ASG1 also has ChtBD2s and mucin-like domains that are high in threonine, serine, and proline. Indeed, CDART, which searches the NCBI database based on placement and number of conserved protein domains, identifies the entire translated products of ASG1 cDNAs from both L. hesperus (GBCS01010755) and N. clavipes (EU780014) as mucin-like proteins.

ASG1 has ChtBD2 domains, implying that there is an interaction with chitin. Chitin is a common component of arthropod bodies. In fact, chitin has been detected within the spider Nephila edulis where it has been implicated in stiffening the distal portion of the major ampullate silk gland duct to create shear forces for fiber formation and to reinforce the duct walls as internal pressure increases. Histochemical staining and Fourier transformed infrared spectroscopy of N. edulis silk gland ducts and the hindgut revealed the same chitin signatures. Furthermore, chitin was observed in the extracellular matrix within the distal portion of L. hesperus major ampullate silk gland. Chitin use in silk glands is not limited to spiders, chitin is present in the convergently evolved silk gland ducts of the domesticated silkworm Bombyx mori. Thus ASG1 likely binds chitin in the silk gland wall and ducts, explaining the presence of ASG1 transcripts in N. clavipes aggregate silk gland mRNA. However, given the higher expression of ASG1 in non-silk tissues than silk-glands of L. hesperus, we posit that ASG1 is not a silk glue but rather is a homolog of an insect intestinal mucin that functions in multiple locations throughout a spider.
Evolutionarily conserved elements of ASG2. Contigs matching ASG2 were identified from all six target capture species. Within each species, the ASG2 contigs shared > 97% nucleotide identity and thus we report only the longest contig assembled from each species (Table 2). Across species, no splice junctions were detected and despite the substantial 1,874–3,130 bp lengths of the contigs, no contig had a putative start codon. Instead, contigs were either entirely protein coding sequence or protein coding with a stop codon and 3′ UTR. All ASG2 capture contigs are consistent with derivation from a large exon at the 3′ end of the gene or a large single exon gene.
percent of the aligned amino acids (21 out of 123 positions) are invariant and another 19% (23/123) share a single locus gene. Within Theridiidae mirrored phylogenetic relationships based on morphology25 and nuclear and mitochondrial ancestor of Araneoidea. mimosarum for Latrodectines (Latrodectus sequences26. Specifically, 97% bootstrap support for a the ASG2 contigs from the biological replicates shared ASG2 aligned well with the others (Fig. 6). Within species, the A. argentata P. tepidariorum species except and ASG2 carboxyl-terminal region. Juxtaposition of conserved sequence implies a potential structural or functional significance, such as separating (Fig. 5). Despite the length differences, 40 residues are 100% conserved and many occur in adjacent positions. considerably across species, making them more difficult to align with each other than the exemplar repeats. The various turn-coil-turn structures and β-sheets map identically across species on the amino acid alignment (Fig. 5b green and blue shading). Similar to fibrous spider silks, the periodic structure of the repeats may contribute to the elasticity and “suspension bridge” properties observed for viscous, aggregate silk droplets5,8. Furthermore, the abundant threonine residues (composing 9–15% of the repeat, depending on species) may provide O-glycosylation sites, which could explain the glycosylated proteins detected within the N. clavipes viscous droplets8. Following the repetitive region is a non-repetitive linker region. The length of the linker region varies across species, making them more difficult to align with each other than the exemplar repeats (Fig. 5). Despite the length differences, 40 residues are 100% conserved and many occur in adjacent positions. Juxtaposition of conserved sequence implies a potential structural or functional significance, such as separating the carboxyl-terminal region from the repetitive region. 

Conserved ASG2 carboxyl-terminal region. The ASG2 carboxyl-terminal region was captured from all species except A. argentata and P. tepidariorum. Both A. argentata and P. tepidariorum ASG2 contigs were truncated towards the end of the linker, before the start of the carboxyl-terminal region (Fig. 5c, black arrow and question marks). After removing the 423 bp insertion from the N. clavipes ASG2 cDNA, its carboxyl-terminal region aligned well with the others (Fig. 6). Within species, the ASG2 contigs from the biological replicates shared > 97% nucleotide identity, suggesting that unlike ASG1, ASG2 is a single locus. 

As was done with ASG1, we used BLAST with our target capture contigs to search the NCBI databases for ASG2 homologs. From the P. tepidariorum whole genome shotgun assembly, we located contig 63868, which possessed the ASG2 carboxyl-terminal region that was missing from our P. tepidariorum target capture contig (BLASTN e-value 0.0, 97% identity over 1,874 bp). BLASTN searches against the transcriptome shotgun assemblies for L. hesperus and S. grossa tissues, which included silk glands, also successfully yielded cDNAs that included all of the sequence from the ASG2 contig. However, there were no matches to ASG2 in the S. mimosarum genome (TBLASTX search, no matches with e-value < e-01). Although this absence may be due to incompleteness of the S. mimosarum genome assembly, absence of ASG2 is also consistent with the lack of aggregate glands in S. mimosarum. Aggregate glands and ASG2 are likely to be evolutionary innovations that arose in a recent common ancestor of Araneoidea. 

Phylogenetic analysis of the ASG2 carboxyl-terminal region resulted in one tree (Fig. 7). Relationships within Theridiidae mirrored phylogenetic relationships based on morphology28 and nuclear and mitochondrial sequences26. Specifically, 97% bootstrap support for a Latrodectus clade (L. geometricus, L. hesperus), 98% support for Latrodectines (Latrodectus, S. grossa), and 100% for Theridiidae (Latrodectines, P. tepidariorum). The strong clade support values and congruence of the gene tree with species-level relationships are further evidence that ASG2 is a single locus gene. 

There are numerous features of the carboxyl-terminal region that are conserved across species. Seventeen percent of the aligned amino acids (21 out of 123 positions) are invariant and another 19% (23/123) share

| Abbreviated Sequence Name | Species       | Accession                |
|--------------------------|---------------|--------------------------|
| A. dia                   | Araneus diadematus | KU132358                |
| A. arg                   | Araneus argentia | KU132351                |
| L. geo cDNA              | Latrodectus geometricus | GBM01131277             |
| L. hes cDNA              | Latrodectus hesperus | GBCS01005615             |
| L. hes                   | Latrodectus hesperus | KU132352                |
| N. cla cDNA              | Nephila clavipes | EU780015 excluding bases 1612–2035 |
| N. cla cDNA edited       | Nephila clavipes | KU132353                |
| P. tep contig 63868      | Parasteatoida tepidariorum | ACMJ01063868          |
| P. tep                   | Parasteatoida tepidariorum | KU132354                |
| S. gro cDNA              | Steatoda grossa | GBQ01029471              |
| S. gro                  | Steatoda grossa | KU132355                |

Table 2. Description of AgSp1/ASG2 sequences. Entries in blue are from this study.
physiochemical conservation (Fig. 6 red and blue, respectively), Given an estimated age of ~155 million years for Araneoidea27, purifying selection appears to have been maintaining these aspects of ASG2. The conserved residues are likely of structural and/or functional importance for prey capture. For example, the conserved cysteine (Fig. 6 orange box) could form a disulfide bridge with other ASG2 monomers, chemically linking them into polymers facilitating droplet extension to "suspension bridge" conformation5.

ASG2 is a spidroin (spider fibroin). Given the previously unrecognized carboxyl-terminal sequence similarities across species (Fig. 6), we investigated putative homologs for ASG2. We performed NCBI conserved domain searches with our ASG2 capture contigs and the edited N. clavipes ASG2 cDNA as queries, and found that the ASG2 carboxyl-terminal region matched the carboxyl-terminal regions of spidroins (Spidroin_MaSp pfam11260, e-values > 3.07e-3). DELTA-BLAST search with N. clavipes ASG2 resulted in a hit of e-value 2e-17 to major ampullate spidroin 1 (MaSp1) from Argiope amoena (AAP88232). Spidroins are a family of proteins that compose spider silks 28. Spidroin gene family members tend to have coding regions > 10 kb that translate into > 200 kDa proteins29,30. Spidroin paralogs share a similar architecture that consists of an ~150 amino acid non-repetitive amino-terminal region, an extensive repetitive region, and an ~100 amino acid non-repetitive carboxyl-terminal region30. The amino acid sequence attributes of spidroin repetitive regions are associated with protein structures that underlie the impressive mechanical properties of spider silk fibers31.

The conserved domain and DELTA-BLAST searches with our ASG2 sequences resulted in hits to MaSp1 (see above). However, ASG2 repeat units (Fig. 5b) lack the distinguishing characteristics of MaSp1 paralogs (poly-alanine, glycine-alanine couplets, and glycine-glycine-X triplets)32. Instead, ASG2 is a newly discovered member of the spidroin gene family. We propose that ASG2 should be renamed aggregate spidroin 1 (AgSp1). As with other spidroins, given its extremely conserved architecture and sequence elements, the repetitive region of AgSp1 (ASG2; Fig. 5b) is also likely to be critical for the mechanical properties of viscous silk glues.

The AgSp1 (ASG2) repetitive region is followed by a linker region that varies in length across species (Fig. 5c), as has been observed in some spidroin family members33,34. Unlike the linker regions, spidroin carboxyl-terminal repeats are linked by disulfide bridges, whereas ASG2 linker regions are not linked.
regions are conserved in length and structure. The AgSp1 (ASG2) carboxyl-terminal region is no exception, with a series of conserved \( \alpha \)-helices (Fig. 6, yellow highlights) that in other spidroins are described as forming a barrel-like structure. In aciniform silk, interlocking spidroin dimers are stabilized by two salt bridges and a disulfide bond. Consistent with this model, the AgSp1 (ASG2) monomers may be linked with salt bridges between the physiochemically conserved R/K and D/E residues (Fig. 6 asterisks) and disulfide bonds between the conserved cysteine residues (Fig. 6 orange box).

The recognition that ASG2 is a spidroin (AgSp1) integrates disparate observations of silk dope chemistry, aggregate silk gland morphology, and viscous glue droplet properties. For example, chemical composition of the viscous droplets resembles unprocessed silk dope of non-aggregate silk glands in terms of salts and up to 50% water content. In the major ampullate silk glands, which produce dry, fibrous silk, salt and water molecules are removed during processing. The removal occurs as silk dope passes from the glandular lumen to the spigot via long and narrow ducts.

Figure 6. Alignment of ASG2 carboxyl-terminal regions. Predicted \( \alpha \)-helical regions are highlighted in yellow, and red arrow shows location of removed sequence from \( N. clavipes \) ASG2 cDNA (Fig. 4). Amino acids are depicted by single letter IUPAC abbreviations. Red indicates amino acids conserved across species and blue physiochemical conservation. Box encloses the conserved cysteines, and asterisks above columns indicate residues predicted to form salt bridges. Abbreviations of species names and GenBank accession numbers are in Table 2.

![Figure 6](image)

regions are conserved in length and structure. The AgSp1 (ASG2) carboxyl-terminal region is no exception, with a series of conserved \( \alpha \)-helices (Fig. 6, yellow highlights) that in other spidroins are described as forming a barrel-like structure. In aciniform silk, interlocking spidroin dimers are stabilized by two salt bridges and a disulfide bond. Consistent with this model, the AgSp1 (ASG2) monomers may be linked with salt bridges between the physiochemically conserved R/K and D/E residues (Fig. 6 asterisks) and disulfide bonds between the conserved cysteine residues (Fig. 6 orange box).

The recognition that ASG2 is a spidroin (AgSp1) integrates disparate observations of silk dope chemistry, aggregate silk gland morphology, and viscous glue droplet properties. For example, chemical composition of the viscous droplets resembles unprocessed silk dope of non-aggregate silk glands in terms of salts and up to 50% water content. In the major ampullate silk glands, which produce dry, fibrous silk, salt and water molecules are removed during processing. The removal occurs as silk dope passes from the glandular lumen to the spigot via long and narrow ducts. However, aggregate silk glands have distinctively short ducts connected to spigots with
wide openings. This means that silk dope could be secreted with minimal processing. The hypothesis that viscous droplets are composed of spidroin-based dope is supported by observations of droplet extension upon draw, crystallization upon dehydration, and a fibrous core to the viscous droplet.

**Summary.** By sequencing target capture genomic libraries, we report sequences for ASG1 and ASG2 from six spider species. We describe the gene structure of ASG1 and recovered multiple gene variants in each individual, indicating that ASG1 is multi-locus. Comparative analyses with homologs from the literature identified our ASG1 sequences as members of the peritrophic matrix intestinal mucin family. In insects, these mucins bind chitin in an extracellular matrix. Furthermore, in the Western black widow, ASG1 is more highly expressed in non-silk gland tissue than in silk glands. Thus, ASG1 is not a silk glue, but may have a role in providing structural reinforcement within silk glands, as well as other spider tissues.

Based on our target capture genomic contigs, we conclude that the previously reported *N. clavipes* ASG2 cDNA is a chimeric sequence, and that the region with identity to ASG1 appears to be a cloning artifact (Fig. 4b orange sequence). Because the artifactual sequence was inserted in the carboxyl-terminal region, the affinity of ASG2 to the spidroin gene family has been obscured until now. AgSp1 (aggregate spidroin 1) is a more appropriate name for ASG2, as ASG2 possesses the spidroin domain architecture of a repetitive region with tandem-arrayed repeat units, followed by a labile linker that leads to the conserved carboxyl-terminal region. Thus, aggregate silk glands produce viscous glue with a spidroin component. Realization that there is an aggregate spidroin deepens the mechanistic understanding of viscous glue droplet properties, such as formation of fibrous cores and viscoelastic-solid behaviour. The comparison of spidroin glue sequences from cob-web weavers and orb-web weavers revealed extensive evolutionary conservation of amino acid sequences and motifs. These biochemical features are relevant to research on bioinspired glues. Moreover, the discovery of a spidroin family member that is specialized to be a glue inspires future studies on the evolutionary reshaping of silk glands for production of dry fibers and wet adhesives.

**Methods**

**Target Capture Probes.** cDNA sequences for all annotated aggregate spider glues from *N. clavipes* (GenBank accessions EU780015, EU780014) and *L. hesperus* (JX262189, GBCS01010755, GBCS01018793, GBCS01005651, GBCS01005654) were collected from NCBI in June 2014. The nucleotide sequences were aligned and identical sequences were consolidated with Sequencher 4.2 (Gene Codes, Ann Arbor MI). The resulting set of ASG1 and ASG2 reference sequences were used to design 120-mer oligonucleotide target capture probes with 5X tiling by Agilent Inc. (Santa Clara, CA).

**Genomic DNA extraction.** We obtained live *N. clavipes* (Nephilidae) from Gainesville, Florida; *Argiope argentata* (Araneidae) from Encinitas, California; *Araneus diadematus* (Araneidae) from Berkeley, California; *L. hesperus* (Theridiidae) from Riverside, California; *Steatoda grossa* (Theridiidae) from Yarnell, Arizona; and *Parasteatoda tepidariorum* (Theridiidae) from Lexington, Virginia. In each species, we sampled two individuals per species. Each spider was euthanized and its cephalothorax was placed in a microfuge tube, flash frozen in liquid nitrogen, and stored at 80°C. Genomic DNA was extracted from each cephalothorax with the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA) and treated with RNase A. DNA concentration was determined with a Qubit fluorometer (Life Technologies, Carlsbad, CA).

**Capture Library Construction.** For each species, two target capture genomic libraries were constructed, for a total of twelve libraries. Each library represented the genome of an individual spider. The biological replicates were done to facilitate distinguishing allelic variation from multi-copy genes. Prior to library construction, 5 μg of DNA was sheared into ~500 bp fragments with an S220 focused-ultrasonicator (Covaris, Woburn, MA). The fragmented DNA was treated with PreCR Repair Mix (New England Biolabs, Ipswich, MA) and then purified with Agencourt AMPure XP beads (Beckman Coulter, Brea, CA). Library construction and target selection were performed with the SureSelectXT Reagent Kit (Agilent), using the target capture probes described above. The libraries were sequenced bidirectionally (2 × 300bp) in multiplexed sets of four on a MiSeq instrument (Illumina, San Diego, CA) at the University of California, Riverside Genomics Core Facility.

**Contig Assembly.** Prior to assembly, sequencing reads were filtered to minimize incorporation of low-quality bases into the results. Using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), each of the 24 fastq files (12 libraries, 2 directions each) was filtered separately. First, library construction adapters and indexing barcodes were removed from each read, and reads containing any ambiguous bases (basecalls other than A, C, G, T) were entirely discarded. For each fastq file, quality scores and base composition at each position were plotted. These plots were made with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) implemented on the Galaxy webserver, usegalaxy.org[40–42]. Based on these plots, the first 6–10 bases were removed from the 5′ end of each read with FASTX-toolkit. These initial bases had lower quality and skewed base composition, attributed to artifacts from the library construction steps (e.g., A-tailing). Scanning from 5′ to 3′, reads were also trimmed from the earliest position with a first-quartile quality score below Q25. Trimmed reads with > 15% of their bases below quality score Q25 were removed (FASTX-toolkit). The filtered reads were sorted to retain only paired-end reads (Prinseq[43]. Paired-end reads were interleaved into a single file before contig assembly, for a total of 12 quality-filtered fastq files (rack1)44. The Trinity assembler was used to construct contigs from the quality-filtered fastq files45. Each library was assembled separately, using default parameters and paired-end reads.

**Sequence Analyses.** Contigs representing homologs to ASG1 or ASG2 were identified by nucleotide similarity to the reference sequences (described above) using Geneious v. 6.18 (http://www.geneious.com)46. Only
Sequences were translated and carboxyl-terminal regions aligned in Geneious using the Clustal W algorithm\(^50\), confirmed through BLASTX searches against the NCBI nr database. Contigs representing homologs to ASG1 or ASG2 were translated in silico. Potential exon splice sites were annotated with the Splice Site Prediction tool at the Berkeley Drosophila Genome Project website using the “Drosophila” standard settings (http://www.fruitfly.org/seq_tools/splice.html)\(^{48}\). Amino acid compositions were calculated in Geneious and peptide secondary structures were predicted with the EMBOSS Geneious plugin\(^{23,24}\). TBLASTN searches against whole genome (\(P\). tepidariorum and \(S\). mimosarum) and transcriptome (\(L\). hesperus, \(L\). latrodectus geometricus and \(S\). grossa) shotgun assembly databases on NCBI were conducted to identify additional relevant sequences for ASG1 (Table 1) and for ASG2 (Table 2). CDD (conserved domain database)\(^{49}\), DELTA-BLAST (domain enhanced lookup time accelerated BLAST)\(^{16}\), and CDART (conserved domain architecture retrieval tool)\(^{50}\) implemented on the NCBI website were used to annotate putative protein function. Sequences were translated and carboxyl-terminal regions aligned in Geneious using the Clustal W algorithm\(^50\), then refined by eye. Amino acid model testing and maximum likelihood gene tree construction with 10,000 bootstrap replicates were done with RAXML v8.2.X\(^{51}\). Amino acid substitution models FLU and JTT were used for ASG1 and ASG2 tree constructions, respectively. Resulting trees were mid-point rooted and visualized with FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

References

1. Coddington, J. A. Spinereret silk spigot morphology: evidence for the monophyly of orb-weaving spiders, Cyrtophorinae (Araneae), and the group Theridiidae plus Nesticidae. J. Arachnol. 17, 71–95 (1999).
2. Vollrath, F. et al. Compounds in the droplets of the orb spider’s viscid spiral. Nature. 345, 526–528 (1990).
3. Opell, B. D. & Hendricks, M. L. The role of granules within viscous capture threads of orb-weaving spiders. J. Exp. Biol. 213, 339–346 (2010).
4. Sahni, V., Blackledge, T. A. & Dhinojwala, A. Viscelastic solids explain spider web stickiness. Nat. Commun. 1, 19 (2010).
5. Opell, B. D. & Hendricks, M. L. Adhesive recruitment by the viscous capture threads of araneoid orb-weaving spiders. J. Exp. Biol. 210, 553–560 (2007).
6. Sahni, V., Blackledge, T. A. & Dhinojwala, A. Changes in the adhesive properties of spider aggregate glue during the evolution of cobwebs. Sci. Rep. 1, (2011).
7. Sahni, V. et al. Direct solvation of glycoproteins by salts in spider silk glues enhances adhesion and helps to explain the evolution of modern spider orb webs. Biomacromolecules 15, 1225–1232 (2014).
8. Amarupuri, G., Chaurasia, V., Jain, D., Blackledge, T. A. & Dhinojwala, A. Ubiquitous distribution of salts and proteins in spider glue enhances spider silk adhesion. Sci. Rep. 5, 9030 (2015).
9. Chorea, O., Bayarmagni, B. & Lewis, R. V. Spider web glue: two proteins expressed from opposite strands of the same DNA sequence. Biomacromolecules 10, 2852–2856 (2009).
10. Vasanthavada, K. et al. Spider glue proteins have distinct architectures compared with traditional spidroin family members. J. Biol. Chem. 287, 35986–35999 (2012).
11. Clarke, T. H. et al. Multi-tissue transcriptomics of the black widow spider reveals expansions, co-options, and functional processes of the silk gland gene toolkit. BMC Genomics 15, 365 (2014).
12. Hegedus, D., Erlandson, M., Gillett, C. & Topрак, U. New insights into peritrophic matrix synthesis, architecture, and function. Annu. Rev. Entomol. 54, 285–302 (2009).
13. Kumar, A. An overview of nested genes in eukaryotic genomes. Eukaryot. Cell 8, 1321–1329 (2009).
14. Haworth, A. C. & Opell, B. D. Evolution of adhesive mechanisms in cribellar spider prey capture thread: evidence for van der Waals and hygroscopic forces. Biol. Linn. Soc. Lond. 77, 1–8 (2002).
15. Clarke, T. H., Garb, J. E., Hayashi, C. Y., Arensburger, P. & Ayoub, N. A. Spider transcriptomes identify ancient large-scale gene duplication event potentially important in silk gland evolution. Genome Biol. Evol. 7, 1656–1670 (2015).
16. Boratyn, G. M. et al. Domain enhanced lookup time accelerated BLAST. Biol. Direct 7, 12 (2012).
17. Topрак, U., Baldwin, D., Erlandson, M., Gillett, C. & Hegedus, D. D. Insect intestinal mucins and serine proteases associated with the peritrophic matrix from feeding, starved and moulting Mamestra configurata larvae. Insect Mol. Biol. 19, 163–175 (2010).
18. Jasrapuria, S. et al. Genes encoding proteins with peritrophic A-type chitin-binding domains in Tribolium castaneum are grouped into three distinct families based on phylogeny, expression and function. Insect Biochem. Mol. Biol. 40, 214–227 (2010).
19. Jasrapuria, S., Specht, C. A., Kramer, K. J., Beenman, R. W. & Muthukrishnan, S. Gene families of cuticular proteins analogous to peritrophins (CPAPs) in Tribolium castaneum have diverse functions. PLoS One 7, e94844 (2012).
20. Greer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. CDART: protein homology by domain architecture. Genome Res. 12, 1619–1623 (2002).
21. Davies, G. J. G., Knight, D. P. & Vollrath, F. Chitin in the silk gland ducts of the spider Nephila edulis and the silkworm Bombyx mori. PLoS One 8, e73225 (2013).
22. Casem, M. L., Tran, L. P. P. & Moore, A. M. F. Ultrastructure of the major amputable gland of the black widow spider. Latrodectus hesperus. Tissue Cell 34, 427–436 (2002).
23. Garnier, J., Osguthorpe, D. J. & Robson, B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins, J. Mol. Biol. 120, 97–120 (1978).
24. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277 (2000).
25. Agranovski, I. Morphological phylogeny of cobweb spiders and their relatives (Araneae, Araneoidea, Theridiidae). Zool. J. Linn. Soc. 141, 447–626 (2004).
26. Arnedo, M. A., Coddington, J., Agranovski, I. & Gillespie, R. G. From a comb to a tree: phylogenetic relationships of the comb-footed spiders (Araneae, Theridiidae) inferred from nuclear and mitochondrial genes. Mol. Phylogenet. Evol. 31, 225–245 (2004).
27. Bond, J. E. et al. Phylogenomics resolves a spider backbone phylogeny and rejects a prevailing paradigm for orb web evolution. Curr. Biol. 24, 1765–1771 (2014).
28. Guerette, P. A., Ginzinger, D. G., Weber, B. H. & Gosline, J. M. Silk properties determined by gland-specific expression of a spider fibroin gene family. Science 272, 112–115 (1996).
29. Ayoub, N. A., Garb, J. E., Tinghitella, R. M., Collin, M. A. & Hayashi, C. Y. Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. PLoS One 2, e514 (2007).
30. Garb, J. E., Ayoub, N. A. & Hayashi, C. Y. Untangling spider silk evolution with spidroin terminal domains. BMC Evol. Biol. 10, 243 (2010).
31. Gosline, J. M., Guerette, P. A., Ortlepp, C. S. & Savage, K. N. The mechanical design of spider silks: from fibroin sequence to mechanical function. J. Exp. Biol. 202, 3295–3303 (1999).
Evidence from Multiple Species that Spider Silk Glue Component ASG2 is a Spidroin.

How to cite this article: Collin, M. A. et al. Evidence from Multiple Species that Spider Silk Glue Component ASG2 is a Spidroin. Sci. Rep. 6, 21589; doi: 10.1038/srep21589 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

Acknowledgements
The authors thank T. Blackledge, J. Garb, and M. Stowe for collecting spiders. Comments from R. Crystal Chaw, Sandra Correa-Garhwal, John Gatesy, Jason Stajich, and James Starrett improved this manuscript. This research was supported by the Army Research Office (W911NF-11-1-0299, W911NF-14-1-0145 to CYH) and the National Science Foundation (IOS-0951061 to CYH, IOS-0951886 to NAA).

Author Contributions
C.Y.H. and M.A.C. designed the study and constructed the capture libraries. M.A.C. and T.H.C. assembled the contigs and analyzed the results. M.A.C. and C.Y.H. prepared the manuscript. M.A.C., T.H.C., N.A.A. and C.Y.H. discussed the results and commented on the manuscript.

Additional Information
Competing financial interests: The authors declare no competing financial interests.