Conservation of Essential Design Features in Coiled Coil Silks

Tara D. Sutherland,* Sarah Weisman,* Holly E. Trueman,* Alagacone Sriskantha,* John W. H. Trueman,† and Victoria S. Haritos*

*CSIRO Entomology, Canberra, ACT, Australia; and †School of Botany and Zoology, Australian National University, Canberra, ACT, Australia

Silks are strong protein fibers produced by a broad array of spiders and insects. The vast majority of known silks are large, repetitive proteins assembled into extended β-sheet structures. Honeybees, however, have found a radically different evolutionary solution to the need for a building material. The 4 fibrous proteins of honeybee silk are small (~30 kDa each) and nonrepetitive and adopt a coiled coil structure. We examined silks from the 3 superfamilies of the Aculeata (Hymenoptera: Apocrita) by infrared spectroscopy and found coiled coil structure in bees (Apoidea) and in ants (Vespoidea) but not in parasitic wasps of the Chrysidoidea. We subsequently identified and sequenced the silk genes of bumblebees, bulldog ants, and weaver ants and compared these with honeybee silk genes. Each species produced orthologues of the 4 small fibroin proteins identified in honeybee silk. Each fibroin contained a continuous predicted coiled coil region of around 210 residues, flanked by 23–160 residue length N- and C-termini. The cores of the coiled coils were unusually rich in alanine. There was extensive sequence divergence among the bee and ant silk genes (<50% similarity between the alignable regions of bee and ant sequences), consistent with constant and equivalent divergence since the bee/ant split (estimated to be 155 Myr). Despite a high background level of sequence diversity, we have identified conserved design elements that we propose are essential to the assembly and function of coiled coil silks.

Introduction

Silks are integral to invertebrate biology, providing building material for structures ranging from prey capture webs to protective cocoons for vulnerable juvenile stages. Silks have also attracted much attention in recent years due to their excellent mechanical properties as new generation biomaterials (Vollrath 2000; Scheller et al. 2001; Vollrath and Knight 2001; Lazaris et al. 2002). The best studied silks are the cocoon fibers of silkworms and related moths and the webs spun by orb-weaving spiders. Despite the evolutionary distance between these groups, their silk proteins share a number of convergent design principles (Bini et al. 2004). The architecture of the large silk proteins (>150 kDa) is an alternation of long hydrophobic blocks and short hydrophilic spacers, flanked by hydrophilic termini. The proteins are highly soluble in the gland due to their hydrophilic moieties and then are processed by hydrophobic aggregation into insoluble fibers with β-crystalline structure. The primary sequences of the silk proteins excluding their termini consist of arrays of highly conserved repeat motifs.

It has long been known that several species of Hymenoptera, including honeybees, hornets, and solitary wasps, produce silk with an entirely different molecular structure to the β-sheet silks described above. The cocoon silk of these species consists of coiled coils that run parallel to the fiber axis (Rudall 1962; Atkins 1967). Coiled coils are a protein-structural arrangement where multiple α-helices wind around each other. Their protein sequences are characterized by a 7-residue periodicity where large, nonpolar residues generally occupy the first and fourth position of each heptad (Woolfson 2005). This coiled coil silk structure is apparently confined to aculeate species with the silks of the nonaculeate Hymenoptera (parasitic wasps and sawflies) generally composed of proteins in predominantly a β-sheet conformation (Rudall 1962). The paraphyly of parasitic wasps and sawflies with respect to the aculeate Hymenoptera suggests that coiled coil silks represent the derived condition within this insect order with the β-sheet silks being plesiomorphic.

Recently, honeybee silk has been shown to consist of 4 small (~30 kDa) proteins that contain extensive coiled coil motifs (Sutherland et al. 2006). These proteins are paralogues but their sequence has diverged extensively so that they retain only around 30% identity. Within the proteins there are no primary sequence repeats. Clearly, the honeybee silks are based on completely different design principles to moth and spider silks. We analyzed silks produced by a range of members of the Aculeata to determine the distribution of coiled coil silks. Among the Aculeata producing coiled coil silk, we selected exemplar species separated by the greatest evolutionary distance. We obtained the silk genes of these species and compared the sequences of the encoded silk in order to identify conserved features that may be necessary for the assembly and function of coiled coil silk.

Materials and Methods

Insects and Tissues

Honeybee (*Apis mellifera*) larvae and silk were obtained from a commercial hive. Bumblebee (*Bombus terrestris*) larvae and silk were obtained from Biobees (New Zealand). Weaver ant (*Oecophylla smargdina*) larvae and silk were collected from Darwin, Northern Territory, Australia. Bulldog ant (*Myrmecia forficata*) larvae and silk were collected from the Tinderry Nature Reserve outside Canberra, Australia. *Goniozus* sp. wasp silk was kindly provided by Santhosh S. Nair, University of Calicut, and *Prosierola* sp. wasp silk was kindly provided by Robert Wharton, Texas A&M University. The labial gland from larvae in the final instar was dissected under phosphate-buffered saline, immediately transferred to RNAlater (Ambion, Austin, TX), and subsequently stored at 4 °C.

Key words: silk, coiled coils, bees, ants, social insects.

E-mail: tara.sutherland@csiro.au.

Fourier Transform Infrared Spectroscopy Methods

The bee silks were washed extensively in chloroform to remove wax. All silks were washed extensively in warm water containing 2% sodium dodecyl sulfate to remove water-soluble contaminants and then rinsed in distilled water. Cleaned and dried silk samples were examined using a Bruker Tensor 37 Fourier transform infrared spectrometer with a Pike Miracle diamond attenuated total reflection accessory. The amide I region of the spectra was analyzed to identify the protein secondary structure. Spectra from samples that were degummed (boiled for 30 min in 0.05% sodium carbonate solution, a standard procedure for degumming silkworm silk) were unchanged.

cDNA Library Construction

Total RNA was isolated from 4 bumblebee (2 μg), 4 bulldog ant (3 μg), and ~100 weaver ant (0.4 μg) late-larval labial glands using the RNAqueous4PCR kit from Ambion. mRNA was isolated from the total RNA using the Micro-FastTrack 2.0 mRNA Isolation kit from Invitrogen (Carlsbad, CA) into a final volume of 10 μl water. cDNA libraries were constructed from the mRNA using the CloneMiner cDNA kit of Invitrogen with amounts of reagents reduced from that described in the standard protocol to accommodate the small amounts of starting material (as described in Sutherland et al. 2006). The cDNA libraries comprised ~2×10^7 (bumblebee), 5×10^7 (bulldog ant), and 5,000 (weaver ant) colony-forming units with less than 1% original vector for the bulldog ant and bumblebee libraries and greater than 80% original vector in the weaver ant library. The average insert size within the libraries was 1.3 kbp.

Quantitative Reverse Transcriptase–Polymerase Chain Reaction Analysis

The cDNA library ligation mix was analyzed by quantitative reverse transcriptase–polymerase chain reaction (RT-PCR) in the Applied Biosystems 7,000 sequencing detection machine using iTaq plus ROX master mix (BioRad, Hercules, CA) according to manufacturer's instructions using the following primer pairs (shown in 5′–3′ direction)—honeybee: AmelF1 TCGCCGAGATTGATAAAAAG, GCGGCACGCGCATTA; AmelF2 TAACGCCGATTCAATCATCAC, TTGCCGCCTTAACTTGAGCTT; AmelF3 GGGATGGAGCGATTATAGGACTT, CGCCTGTGCAAGCAACTG; AmelF4 TGGCAGGTGAGATGAAACCA, TCTTCCTTCGTTGACGTTTTCC; bumblebee: BBF1 CGCATCCGCGATGTCAT, TGATTCTTGGCCGTCTTTTCA; BBF2 CGGCCGCCGGTTTC, GTTTAGTGGCCTCGGATTCG; BBF3 CAACGCAGGCACAGTTGCT, CAGTTGTTGCTCAGATTGAGTTT; BBF4 ACCGCAGTGAACGCCATT, TGCTACACCAGCGGTGTTCTC; bulldog ant: BAF1 CTCGTGCAGGCCAATGC, GCTCTAGACTTCGAATCCAATTCG; BAF2 CAGGCCGCAAGTCAGATACA, GCGGTGTTGGTCTCTGATTGT; BAF3 AATCTTCACGTGTTGCCATTACC, TTGATGTCGAACACCGACTTG; BAF4 AACGCCGATCTGTCCAAGAG, CTGATTTTCTTGACTTCCTGTACTTGA; and weaver ant: WAF1 CAATCAGGAGATCGTTGAAATCG, TGTGCCGGTCGAGATCACT; WAF2 CGACCAGGGCGTTGCT, TTTCTGGGTAAGTTCGGCTTGA; WAF3 CGTTAGCGTCCACCGTCAA, CAGAAGAACTGGCGGTATTTCC; WAF4 AGCTACCGAAGCCGCTTCTAC, GGATTGGTGATCTCGCCTTTC. Primers were designed using the Primer Express software of Applied Biosystems and chosen based on software scores and relative position on the gene. Relative levels of cDNA were determined by comparing a threshold cycle for amplification for the 4 primer pairs at different library concentrations (1/1,000–1/100,000 dilutions) within the same experiment.

Silk Preparation, Amino Acid Analysis and Liquid Chromatography/Mass Spectrometry Methods

The amino acid compositions of clean silks were determined after 24 h gas phase hydrolysis at 110 °C using the Waters AccQTag chemistry by the Australian Proteome Analysis Facility Ltd. (Macquarie University, Sydney). The proteins present in the clean silks were identified by liquid chromatography followed by tandem mass spectrometry (MS/MS) as described previously (Sutherland et al. 2006). Briefly, silk was digested with trypsin and the resultant peptides were analyzed by MS/MS. The experimentally observed sets of peptide mass fragments were compared with predicted proteins encoded by cDNAs from the silk gland libraries.

Cladistic analysis and domain predictions

Signal peptides were predicted by SignalP 3.0 (Bendtsen et al. 2004), and coiled coil regions were predicted by MARCOIL (Delorenzi and Speed 2002; http://www.isrec.isb-sib.ch/webmarcoil/webmarcoilC1.html). The intrinsically unstructured nature of the flanking regions was determined using the protein prediction program DisEMBLE (Linding et al. 2003). Protein alignments of the 210 residue coiled coil region were performed with MUSCLE (Edgar 2004).

Plots of the pairwise distances against transition/transversion ratios in the nucleotides encoding the coiled coil regions showed no decline in transition/transversion ratio with pairwise distances other than for ant–ant and bee–bee comparisons within each fibroin, indicating that the third codon positions are saturated except between the species pairs *Apis* and *Bombus*, and *Oecophylla* and *Myrmecia*. Likewise, the first and second codon positions are saturated in transitions and in synonymous changes (unpublished data). Thus, phylogenetic information is restricted to the pattern of nonsynonymous mutations that was analyzed using equally weighted protein parsimony, and support for the tree was analyzed by doing 1,000 bootstrap runs using PAUP* version 4.0b10 (Swofford 2002).

Distances within heptad positions among the 4 taxa were calculated for the 4 fibroin groups using the Tajima-Nei nucleotide model (DNA distance) excluding the third codon position. The analysis assumed uniform rates of
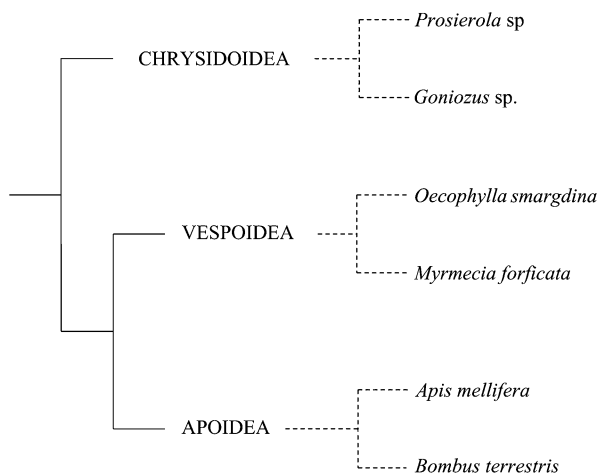
FIG. 1.—Cladogram of the Hymenopteran infrasuborder Aculeata. The monophyly of the Aculeata and the radiation of the 3 superfamilies Chrysidoidea, Vespoidea, and Apoidea are well supported (Grimaldi and Engel 2005).
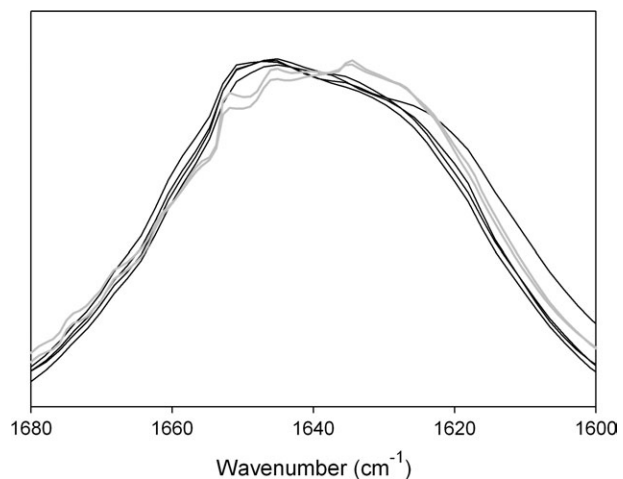

FIG. 2.—The Fourier transform infrared spectra of Aculeata silks. Spectra of honeybee, bumblebee, weaver ant, and bulldog ant silks are shown by black lines, and spectra of *Goniozus* and *Prosierola* wasp silks are shown by gray lines.

changes among sequences. These molecular evolutionary analyses were conducted using MEGA version 3.1 (Kumar et al. 2004).

Results

Coiled Coil Silks Are Found in Bees and Ants but Not in Aculeate Parasitic Wasps

We used Fourier transform infrared spectroscopy (FTIR) to identify the silk structure for species within each of the 3 superfamilies of the infrasuborder Aculeata (Hymenoptera) including bumblebees (*B. terrestris*; Apoidea), weaver ants (*O. smargdina*; Vespoidea), Australian bulldog ants (*M. forficata*; Vespoidea), and 2 species of aculeate parasitic wasps (*Goniozus* and *Prosierola* sp.; Chrysidoidea). A schematic depicting the relationship between these species is shown in figure 1. The silks of bumblebees, weaver ants, and bulldog ants all have FTIR spectra very similar to that of honeybee (*A. mellifera*; Apoidea) silk (fig. 2). We, therefore, expect these silks to have coiled coil structure, as does honeybee silk (Atkins 1967). Consistent with this, the spectral maxima in these 4 FTIR spectra (black lines, fig. 2) are at 1645–1646 cm^{-1}, shifted $\sim$10 cm^{-1} lower than a classical α-helical signal and broadened, as is typical of coiled coil proteins (Heimburg et al. 1999). In contrast, the silks of *Goniozus* and *Prosierola* wasps had spectra (gray lines, fig. 2) with maxima at $\sim$1635 cm^{-1}, suggesting predominantly β-sheet structure.

Bee and Ant Silks Are Composed of 4 Fibroins

Sequence data were obtained from 117, 131, and 23 random clones from cDNA libraries constructed from the silk glands of bumblebee, bulldog ant, and weaver ant larvae, respectively (table 1). As we obtained low levels of mRNA from the silk glands of the tiny weaver ants, the library we constructed from this species contained a high proportion of vector-only sequence and only 23 nonvector

sequences were present in the approximately 500 clones examined. We identified the silk genes from the bumblebee, bulldog ant, and weaver ant by matching mass spectral data obtained from trypsin-digested silk to the predicted proteins encoded by the silk gland cDNAs. The Spectrum Mill software used (Agilent) generated scores for the quality of each match, and the sequence matches reported in this

Table 1
Properties of Proteins Identified in the Silks of Honeybee, Bumblebee, Bulldog Ant, and Weaver Ant

Species	Protein Name	Number of Amino Acids	Percent of cDNA Library Clones[a]	Normalized Expression Levels[b]	LC/MS Identification Score[c]
Fibrous proteins
Bumblebee	BBF1	327	4	1.0 ± 0.16	180
	BBF2	313	14	1.1 ± 0.22	100
	BBF3	332	20	1.3 ± 0.19	218
	BBF4	357	32	1.0 ± 0.15	137
Bulldog ant	BAF1	422	16	1.0 ± 0.09	99
	BAF2	411	30	2.3 ± 0.41	90
	BAF3	394	26	1.6 ± 0.26	88
	BAF4	441	24	1.7 ± 0.24	116
Weaver ant	WAF1	391	35	2.4 ± 0.18	228
	WAF2	400	22	1.7 ± 0.11	191
	WAF3	395	13	1.9 ± 0.20	156
	WAF4	443	17	1.0 ± 0.07	148
Honeybee	AmelF1	333	6[d]	1.0 ± 0.05	52[d]
	AmelF2	309	7[d]	1.3 ± 0.06	51[d]
	AmelF3	335	11[d]	1.2 ± 0.07	107[d]
	AmelF4	342	7[d]	1.5 ± 0.14	88[d]
Glue-like proteins
Bumblebee	BBSA1	>501	3	Not done	138
Honeybee	AmelSA1	578	13[d]	Not done	40[d]

[a] Total number of cDNA sequenced: honeybee 82, bumblebee 117, bulldog ant 131, weaver ant 23.

[b] Normalized ratio of expression levels determined by quantitative RT-PCR as described in the Materials and Methods.

[c] The default score in Spectrum Mill software required for confident identification of a protein greater than 20.

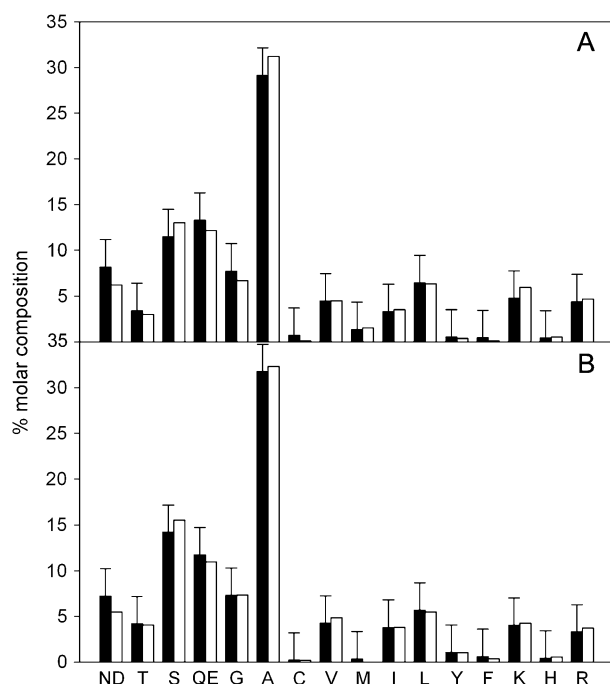[d] Data from Sutherland et al. (2006).

FIG. 3.—Measured and predicted amino acid composition of the silk proteins of (A) bulldog ant cocoons and (B) weaver ant nests. The mean measured amino acid content is shown by filled bars, and predicted amino acid composition assuming equimolar amounts of each of the 4 silk fibroins is shown by open bars.

FIG. 4.—Common architecture of the 16 silk proteins. The proteins contain a signal peptide (shaded bar) and a coiled coil region containing around 30 heptad repeats flanked by regions of variable length (length range indicated by the dotted line).

The mature proteins each contain an approximately 210-amino acid coiled coil region (predicted by MARCOIL; Delorenzi and Speed 2002). The properties of the coiled coil protein sequences are described in detail below. The coiled coil regions are flanked by regions of variable length (N-terminal region, 33–105 amino acids; C-terminal region, 23–125 amino acids) that contain high levels of charged amino acids (average 20.1%). Generally, the flanking regions are longer in the ant fibroins (average 90 residues) than in the bee fibroins (average 50 residues) in part due to terminally located serine-rich regions of 15–23 residues that are intrinsically unstructured (predicted by DisEMBLE; Linding et al. 2003). Peptides from protease-digested mature silk matched to the N- and C-terminal diverged regions of all bee and ant fibroins, confirming that these regions are retained in the processed silk (unpublished data).

The bee and ant silk genes have diverged extensively as we were unable to align the nucleotide sequences. We were also unsuccessful in aligning the protein sequences of the N- and C-flanking regions. Although sequence identity was detected between some regions in species in the same superfamily, primary sequence identity in these regions between ant and bee fibroins, or within a species, was unconvincing. We were able to align the 210-amino acid section of predicted coiled coil protein sequence from the 16 fibroins (fig. 5) using the protein alignment algorithm MUSCLE (Edgar 2004). As this alignment did not fragment the coiled coil heptad motifs predicted by MARCOIL (Delorenzi and Speed 2002) and agreed with manual alignments of the honeybee silk genes described in Sutherland et al. (2006), it was used for all subsequent analysis. Pairwise protein identity was low, ranging from 21% to 68% identity between the ant species, 25% to 71% between the bee species, and 20% to 45% between fibroins from the different superfamilies.

Unusual Composition of Coiled Coil Regions in Fibroins

Coiled coil protein sequences are characterized by a 7-residue $[(abcdefg)_n]$ repeating motif usually containing hydrophobic residues in the a and d positions and hydrophilic residues in the remaining positions (Woolfson 2005). The MARCOIL algorithm predicted that the bee and ant silk fibroins contain approximately 30 contiguous heptad repeats. The amino acid composition at the predicted heptad positions of the silks is very high in alanine in comparison to classic coiled coils especially in the a and d core positions (fig. 6; average composition—a: 65%, b: 35%, c: 37%, d: 51%, e: 39%, f: 33%, g: 30%). The core positions were also surprisingly abundant in small polar residues (serine and threonine) and low in large hydrophobic residues (leucine, isoleucine, and methionine) (fig. 6).

paper scored between 88 and 228 (table 1), considerably over the default cutoff for confident acceptance of valid matches (>20).

We found that bumblebee silk comprised 5 proteins: 4 fibroins (BBF1–4) and a glue-like protein (BBSA1), homologous to those found in the honeybee (Sutherland et al. 2006). In the bulldog ant and weaver ant silks, we found 4 proteins (BAF1–4, WAF1–4; table 1) homologous to the fibroin proteins of the honeybee, but there were no homologs of the bee glue-like proteins. The measured amino acid compositions of the ant silks are consistent with the silk comprising only the 4 fibroin proteins in equimolar ratios (fig. 3). With the exception of the honeybee fibroins, the bumblebee and ant fibroins are unlike any other published sequences.

In all species, the silk proteins corresponded to the most abundant cDNAs in the silk gland libraries. Individual silk cDNAs accounted for 4–35% of the message in the libraries, and combined, the silk proteins accounted for 73% (bumblebee), 87% (weaver ant), and 99% (bulldog ant) of the silk gland cDNAs (table 1). We used quantitative PCR to show that in each species, including honeybees, the 4 silk genes were expressed in approximately equimolar amounts within the silk gland (table 1).

The Silk Fibroins Have Similar Architecture but Low Primary Sequence Identity

The architecture of the bee and ant fibroin proteins is very similar (fig. 4). The silk fibroins are all small (309–443 amino acids) and do not contain primary sequence repeats.

```
         abcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabc
BBF1    AVASQKDATLQADASAAAAAAARASADQSASLAQQSASLQSKAAARAKSAEESAAATAKAELQAESIAASASS
AmelF1  ALASQKDAVLQAQAAASAASEARAAADLTAKLSQESASVQSQAAAKGKETEEAAVGQARAGLESVSMAASATS
BAF1    TAEEQREALEMLTLSADKNAEARILADDTAVLVQGSAEAQSVAAAKTVAVEEESASLDAAAVEAEVAAATSKS
WAF1    TVEEQREALQLLTASAEKNAEARSLADDAAVLVQGAAEAQSVAAAKTVAVEQGSNSLDAAAAEAEAAAAASRV
BBF2    ARAAAAAATKQAKDTEAAEAGAAAAIAIAIAKREEAIKASELASKLLTAAAGSSEAAVSATVRAAQLTAAASA
AmelF2  ARAASVAAAAQAKNTEAAEAGANAALAAAIAKREEAIKASEIANQLLTNAAKAAEATVSATKRAAQLTAAAKE
BAF2    ANEAEILAEAQAATSAQAGAVANSAAERAIAAMEMADRTEYIAALVTTKAAKAAEATMAATARATAAASASKI
WAF2    GNSAESLARAQASSSASAGAKANALIKQSIAAIEITEKAEYLASIVATKAAKAAEATAAATARATAVAEAAKV
BBF3    SAKVKASALALAEAFLRASAAFAAASAKAAAAVKEATQAQLLAQEKALIALKTQSEQQAASARADAAAAAAVS
AmelF3  AAVVKASALALAEAYLRASALSAAASAKAAAALKNAQQAQLNAQEKSLAALKAQSEEEAASARANAATAATQS
BAF3    LAMIKKIAMARSSAYVQAAWASAQASADALASARVAQASQEAAEAKGRAASEALSRAIEASSRADAAAAATLD
WAF3    LAIIKKIVALLISAYVRAAEAAARASAEALATVRAAEQAQKIAEAKGRAAAEALSELVEASQKADAAAAGTTD
BBF4    AAEEAQAAAAAQAKSAAAAAANAGESSNSAAALVAAAAAAQGKAAAAAAAATKASLEAADAAEEAESAVALARA
AmelF4  AAEESQALAASQSKTAANAAIGASELTNKVAALVAGATGAQARATAASSSALKASLATEEAAEEAEAAVADAKA
BAF4    AAEEAQAAAAGQAEVAAKSCELADKTTAKAVAMVEAAAEAEIEVANQEVAAVKLSTWAAKAARIVEEDSAAVRA
WAF4    AAEEAQAAAAGQAAIAAKSCALAAKSTAQAVALVEKVARAEVDLAESARKATRLSAEAAKAAAEVEKDLVGLRG


         defgabcdefgabcdefgabcdefgabcdefgabcdefgabcdef  gabcdefgabcdefgabcdefgabcd
BBF1    NAREAAASAKASASAMSSAAVQAKLAEKTAKNQALASEEAKLKAA~~AAASAAAAASAAAEAALKAERIAEEA
AmelF1  AAKEASTAAKAAASALSTAVVQAKIAERAAKAEAVASDEAKAKAI~~AAANLAAEASVAAEAALKAEKVAEEA
BAF1    SAGQALQSAQTAASALRTSARSALTALKLARLQGAASSNAARMME~~KALAATQDANAAAQQAMAAESAAAEA
WAF1    SAQQALQAAQTSAAAIQTAAGSALTALKLKARKQEAESNNAAEQAN~~KALALSRAASAATQRAVAAQNAAAAS
BBF2    AAKASASASEASAEAQVRANAEANIAKKASAAEAKAAAEAQVKAE~~LAKKAAAGFLAKARLAASAESEATKL
AmelF2  ATRASAAAAEAATEAQVKANADSIITKRAAIAEAQAAAEAQVKAA~~IARKSAANFLAKAQIAAAAESEATKL
BAF2    SSQESAASAANAANAEAKANAASIIANKANAVLAEAAAVLAATAA~~KAKESAMKSLSAAQAAAKAQARNAEA
WAF2    SSEQFAAEARAAADAEAKANAASIIANKANAVLAEAATGLSASAG~~KAQQSATRALQAARAAAKAQAELTQK
BBF3    ALERAQASSRAATTAQDISSSDLEKRVATSAAAEAGATLRAEQSAA~~QSKWSAALAAQTAAAAAAIEAKATAS
AmelF3  ALERAQASSRLATVAQNVASDLQKRTSTKAAAEAAATLRQLQDAE~~RTKWSANAALEVSAAAAAAETKTTAS
BAF3    AMDRTMENARAANAAQTQASGQAENANRSAAAILAALLRIAEASA~~LNNEAAVNAAAAAAAASALQAKANAA
WAF3    AIERTYQDARAATSAQTKASGEAENANRNAAATLAAVLSIAKAASGQGGTRAAVDAAAAAAAAAALHAKANAV
BBF4    ASAKAEALASTAAAANTRAALQAEKSNELAQEAAAAAEAQAKAA~~AAAKATQLALKVAETAVKTEADAAAA
AmelF4  AAEKAESLAKNLASASARAALSSERANELAQAESLSAAAAEAQAKTA~~AAAKAAEIALKVAEIAVKAEADAAAA
BAF4    AAGKLLLAARAAAAAERRANEESEAANELAQASSAAAAEAEAKAN~~AGREAAAAALAIAEAAVAIEQEAVIL
WAF4    AAGKLNLAARAGSKAQERANEDSIEANELAQATAAAGAEAEAKAN~~AAQEAGASALAIAQAALNIEQETVKL


         efgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefg
BBF1    IAKAAAAKAAARAAAAALNSAKEAATSSARSAAEAEAKSEVAILISELDKKSREVAASASAKARAA
AmelF1  IARAASAKAAARAAAAALASSKEAATASARNAAESEARNEVAVLIAEIDKKSREIDAASSLNARAA
BAF1    AAIAAAKQSEARDAGAEAKAAMAALITAQRNLVQANARAEMASEEAELDSKSRASDAKVNAVARAA
WAF1    AASAGAAQAEARNAYAKAKAAIAALTAAQRNYAAAKASASAGSVVAEQDAQSRAADAEVNAVAQAA
BBF2    AAEEAEVALAKARVAVDQSQSAQATATAQAATAVQLQSQAANAEASAVAQAETLLVTAEAVSAAEAE
AmelF2  AAEEAVVALTNAEVAVNQARNAQANASTQASMAVRVDSQAANAEAAAVAQAETLLVTAEAVAAAEAE
BAF2    SAEEAQIKLSQARAAVARAAADQAVCSSQAQAASQIQSRASASESAASAQSETNTAAAEAVATADAE
WAF2    AAQEILVLIAEAKAAVSRASADQSVCTSQAQAASQIQSRASAAEASAASAQSEANTIAAEAVARADAE
BBF3    SESTAAATSKAAVLTADTSSAEAAAAAEAQSASRIAGTAATEGSANWASENSRTAQLEASASAKAT
AmelF3  SEAANAAAKKAAAIASDADGAERSASTEAQSAAKIESVAAAEGSANSASEDSRAAQLEASTAARAN
BAF3    SQATARAAGQASTAAEEAQSAQEAADKNAELTTVMLEKASADQQAASARADYYTASTEAEEAAAQAS
WAF3    SQATSKAAAEARVAAEEAASAQASASASAQLTAQLEEKVSADQQAASASTDTSAAIAEAEAAAALAS
BBF4    AVAAAKARAVADAAASRATAVNAIAEAEERDSAQAENTAGVAQAALAAAEAQDSCIGAAATPRHSS
AmelF4  AVAAAKARAVADAAAARAAAVNAIAKAEEEASAQAENAAGVLQAAASAAAESRAAAAAAAATSEAA
BAF4    ARKAQDARLNAEAAAAAAMNARVIASAESEASEDLENRASVARASAAGAAEAKAIATDAGATAEIA
WAF4    TRQAQNTRLRSENILAAASNARAIASAEAEASSDLNNRANAARSNARAAAETRAVATEAASTAEIA
```
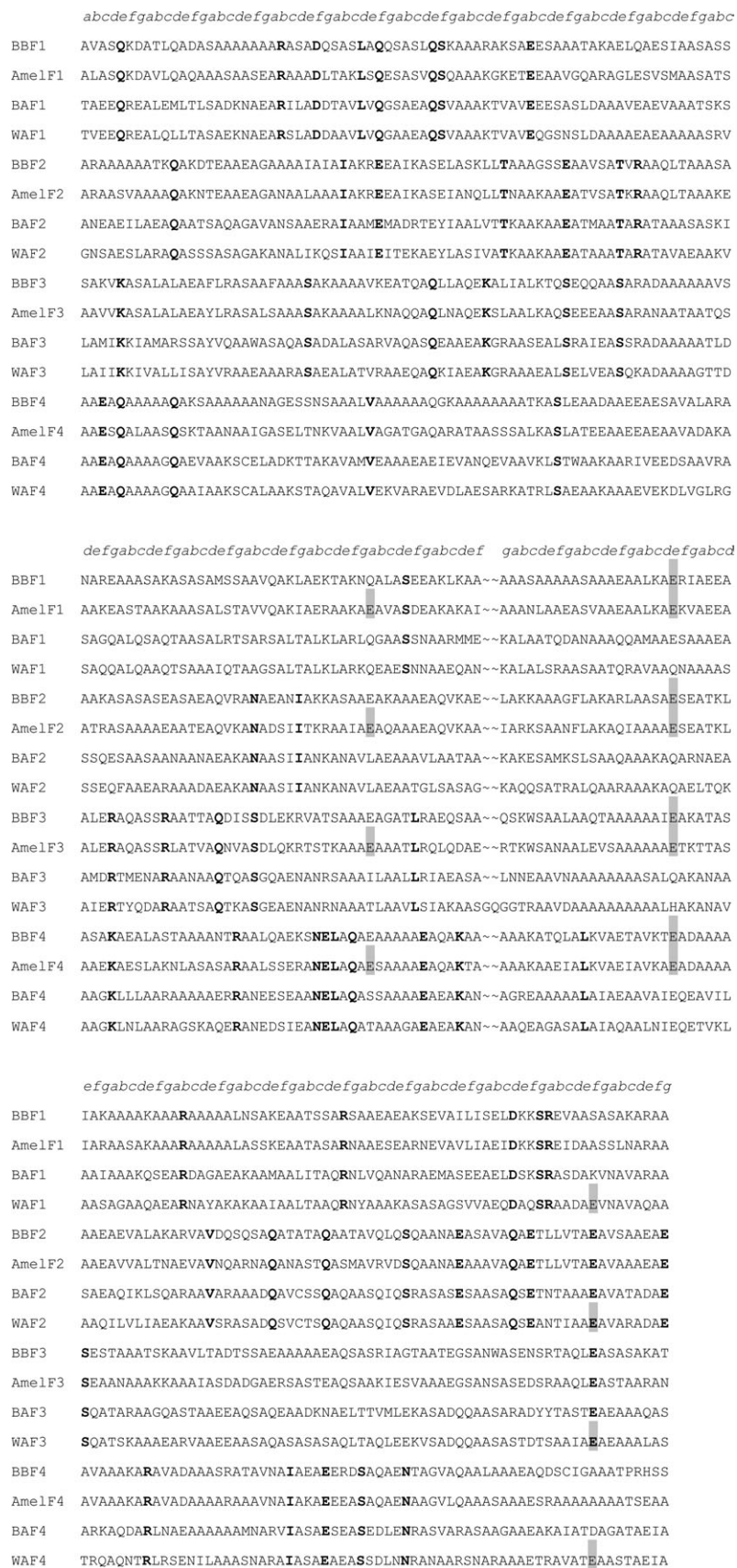
Fig. 5.—Amino acid sequence alignment of the coiled coil regions of the silk proteins. HB, honeybee; BB, bumblebee; BA, bulldog ant; WA, weaver ant; F1–4, silk fibroins 1–4. Residues conserved between species within a fibroin set are indicted in bold typeface and residues conserved in all fibroins within a species have shaded boxes, excluding alanine conservation in any position, which is likely to arise through chance.
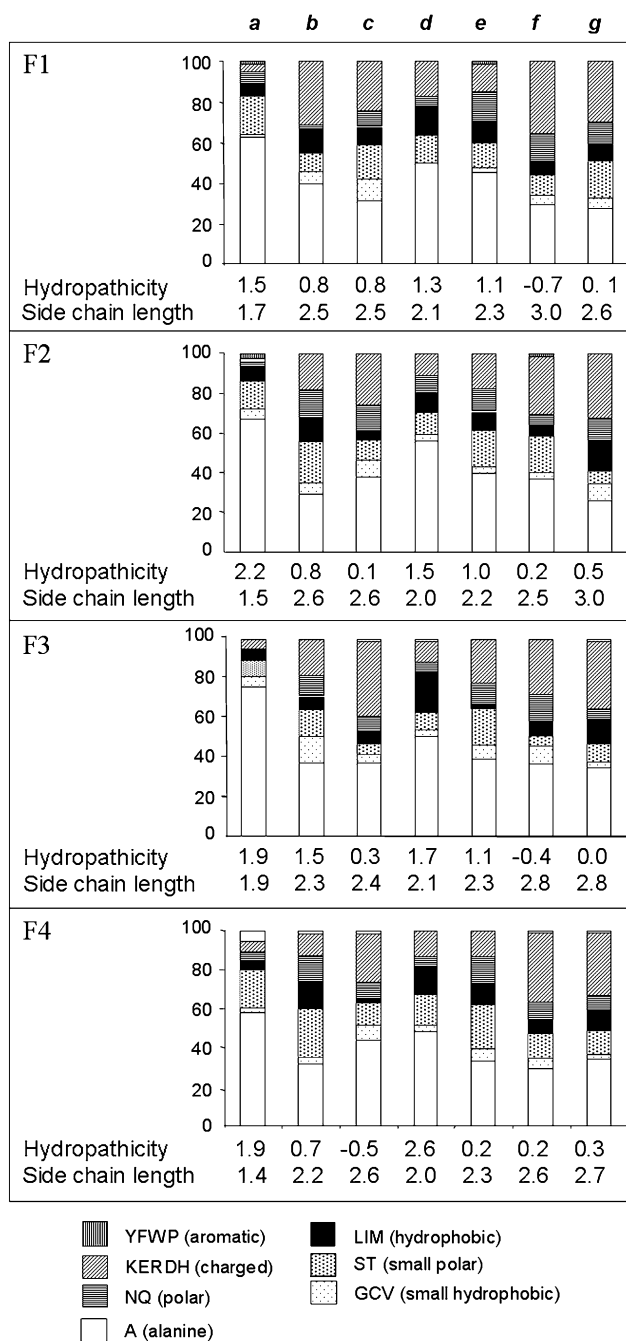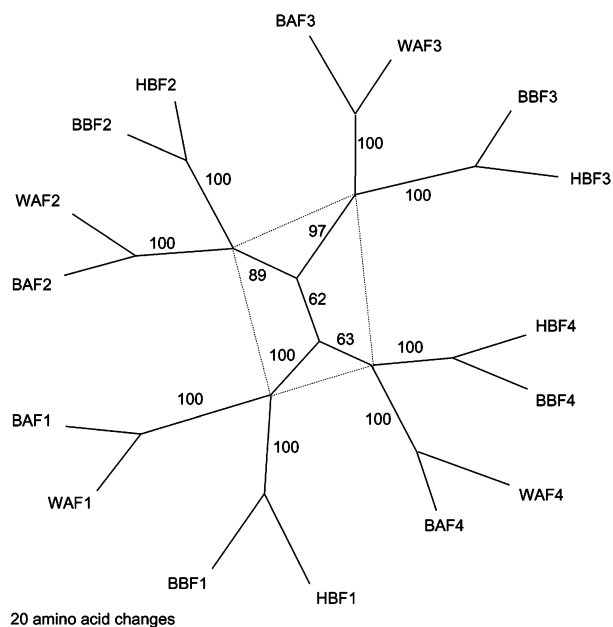
FIG. 6.—Summary of the properties of amino acids occupying each of the predicted heptad positions (a–g) in the coiled coil fibroins that comprise the silk of ants and bees. The graph shows the distribution of amino acids populating each heptad position for each fibroin (F1–F4). Beneath the graph is the average hydrophobicity (kJ/mol) and average side-chain length (atoms) of the residues occupying each position.

FIG. 7.—Equally weighted protein parsimony analysis showing the relationship of the coiled coil regions of the honeybee (HB), bumblebee (BB), bulldog ant (BA), and weaver ant (WA) silk proteins (F1–F4). The area bound by the dotted line indicates divergence that occurred before the split of the ants and wasps (Vespoidea) from the bees (Apoidea) in the Late Jurassic (155 Myr; Grimaldi and Engel 2005).

We calculated the average side-chain hydrophobicity of residues at each of the heptad positions for each fibroin (F1–F4) using the free energy of transfer between octanol and water for each amino acid relative to glycine (Eisenberg and McLachlan 1986). As expected, in all fibroins, the most hydrophobic positions were the classic core positions a and d (fig. 6). On average, the e and b positions were more hydrophobic than the remaining positions. We also compared the average side-chain length of residues at each of the fibroin heptad positions, calculated by counting the maximum number of linearly bonded atoms, excluding hydrogens. Average residue size was smallest in the a and d positions in all fibroin groups (fig. 6). Overall, residues in the e and b positions were also smaller than in the remaining positions.

Bees and Ants Have Retained Single Copies of 4 Ancestral Silk Genes

The pairwise distances between the nucleotides encoding the coiled coil regions of the 16 fibroin proteins were compared with the transition/transversion ratios in the same region. The result showed that the third codon positions of each fibroin gene were saturated except between the species pairs *Apis* and *Bombus*, and *Oecophylla* and *Myrmecia*. Likewise, the first and second codon positions were saturated in transitions and in synonymous changes (unpublished data). Phylogenetic information was therefore restricted to the pattern of nonsynonymous mutations in the coiled coil region and was analyzed using equally weighted protein parsimony (fig. 7).

Cladistic analysis of the coiled coil regions of the silk fibroins revealed that each species contained a single copy of each of the 4 fibroin proteins found in the honeybee (fig. 7). The presence of orthologs in each species suggests that after initial duplication events more than 155 MYA (Grimaldi and Engel 2005) in the common ancestor of the Apoidea and Vespoidea, there have been no further gene duplications or deletions in any of the species. At least

since the split of the ants and bees, the regions of the fibroin genes encoding the coiled coil motif have been diverging in an approximately equivalent manner (fig. 7).

Discussion

We have established that coiled coil silks are produced by members of the Apoidea and the Vespoidea, but not by insects of the Chrysidoidea, the oldest lineage of the Aculeata. Thus, this radically different building material replaced the plesiomorphic β-sheet silk over 155 MYA in the period after the split of the Chrysidoidea from the social Hymenopteran lineages and before the split of the ants and bees.

In order to understand the important elements governing the assembly and function of coiled coil silks, we identified and sequenced the silk genes from bumblebees (Apoidae) and 2 species from Vespoidea. Along with the previously described honeybee silk sequences (Sutherland et al. 2006), this allows us to compare 2 representative sets of sequences that are separated by the greatest possible evolutionary distance. In addition, these silks have diverse applications: honeybees build silk/wax composite hives, bulldog ant larvae produce cocoons for protection during pupation, bumblebee larvae spin cocoons within waxy hives and then reuse the cocoons to store pollen and honey, and silk from weaver ant larvae is used to fasten fresh plant leaves together to form large communal nests.

We found that the silk from each species was made of 4 coiled coil proteins (fibroins), each an ortholog of a silk protein identified in honeybee silk. The bee silks also contained an additional protein but, as this protein could be washed out of the silk while the silk retained its fibrous nature, it is not considered to be an integral component of the silken fibers. The fibroin proteins contained extensive coiled coil regions of conserved length flanked by largely unstructured termini. The hydrophilicity of the termini is similar to the terminal regions of β-sheet silks (Bini et al. 2004) and probably serves the same function of improving solubility in the gland.

The fibroin proteins had no primary sequence repeats and had diverged considerably between species. Only the protein sequence from the predicted coiled coil regions had sufficient identity to be aligned and even within this region pairwise similarity was low, at less than 50% between ant and bee fibroins. This degree of sequence divergence contrasts with the extreme conservation of primary sequence motifs seen between, and within, silk fibroins with β-sheet structure, such as among some orb-weaving spider fibroins, which have been separated for 125 Myr of evolution (Gatesy et al. 2001). We attribute the sequence diversity within the coiled coil silks to the nature of their structure. For coiled coils, the character of the residues in particular heptad positions is important rather than the retention of identical sequence repeats. Hence, there is great latitude for amino acid substitution in the sequence while retaining the coiled coil structure.

The amino acid composition of the coiled coil regions of the silks is unusual, with high levels of alanine, a small hydrophobic amino acid, in the *a* and *d* core positions (fig. 6). Most known coiled coils contain predominantly large
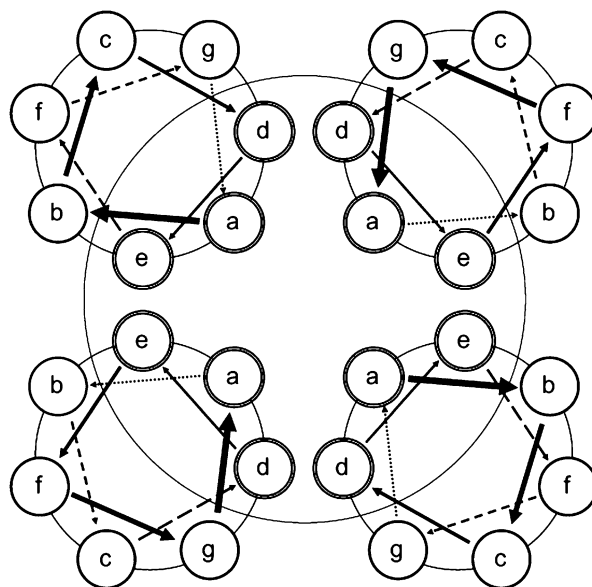

Fig. 8.—A structural model for coiled coil silks. The α-helical strands corresponding to each of the fibroins are arranged in an antiparallel tetrameric configuration (direction indicated by arrows). Three residues (*a*, *d*, *e*) from each heptad repeat are buried in the core.

hydrophobic residues at these positions to maximize the hydrophobic forces stabilizing the core (Woolfson 2005). We ascribe the atypical composition of the coiled coils in the bee and ant silks to metabolic constraints. Silks must be produced in large quantities over a short period and are not recycled, so insects may be constrained to using nonessential amino acids for silk production. Of residues able to form alpha-helices, alanine is the most hydrophobic of the nonessential amino acids for bees (DeGroot 1953).

In physiological conditions, alanine is insufficiently hydrophobic to hold together a classic dimeric or trimeric coiled coil with a 2-residue (*a* and *d* positions) core (Liu and Lu 2002). X-ray diffraction data indicate that honeybee silk comprises tetrameric coiled coils (Atkins 1967), which are stabilized by a wider hydrophobic interface. Fibroin sequence analysis shows that residues at heptad positions *a*, *d*, and *e* are generally smaller and more hydrophobic than are those at the other positions (fig. 6). We propose that all 3 of these positions are included within the core of the coiled coil silks, as shown in figure 8. In fibroins 3 and 4, the *b* heptad position is also small and hydrophobic, which is consistent with its proposed partially submerged position in the core. Three-residue cores are expected to dictate a highly stable, antiparallel, tetrameric variant of the coiled coil structure (Deng et al. 2006).

Other design elements of the fibroins are expected to contribute to the stability of the coiled coil silks. Alanine is a strongly α-helix–favoring residue (Pace and Scholtz 1998), so the overall high alanine content of the proteins would stabilize helices and facilitate coiled coil formation. The predicted coiled coil regions of the fibroins are long and uninterrupted, features known to increase stability (Su et al. 1994; Brown et al. 1996). In addition, the uniform small size of the residues in the core (fig. 6) allows tight packing of the core and strong van der Waals interactions.

The conservation of 4 orthologous fibroin genes in each species for 155 Myr suggests that each gene product performs a nonredundant function in the silk. In addition to the absolute number of fibroin genes being retained, there is strict conservation of approximately 10% of amino acid residues in the coiled coil region between orthologous proteins from the different species, despite a background of extreme sequence divergence (fig. 5). This conservation suggests that the function of each orthologue has been retained in each species. It is most likely that each protein is exclusively located in 1 of the 4 strands of the coiled coil, forming a hetero-tetrameric structure. Our quantitative PCR data indicate that the fibroin genes are expressed at equivalent levels in the silk gland suggesting that the gene products are required in equimolar amounts, as expected for a hetero-tetrameric structure. Within orthologous fibroin groups (i.e., F1 from each species), there is complete conservation of between 14 and 21 residues (fig. 5), not including alanine residues (which are highly abundant and likely to occur in the same position by chance). This is significant given the high levels of sequence diversity observed in the remaining protein sequence, and we propose that these conserved residues are involved in specific polar interactions with neighboring strands, as are often observed in coiled coil proteins (Woolfson 2005). The conserved residues may dictate either partner selection within the coiled coil or intercoil interactions that facilitate fiber synthesis.

It is unclear how coiled coil silk proteins could arise from β-sheet structured silk proteins, which are large and repetitive with completely different functional constraints. Homology between the coiled coil fibroin genes suggests that they evolved from a single gene that emerged in the common ancestor of the Apoidea and Vespoidea, after the split from the Chrysidoidea (fig. 7). As coiled coils are a common and versatile motif, present in an estimated 10% of all eukaryote proteins (Liu and Rost 2001) and in many viral coat proteins, there are many templates that this original protein may have evolved from. The simplicity of the structure also opens the possibility that they evolved de novo. Why coiled coil silks arose in Apoidea and Vespoidea in place of β-sheet silks is not known, but honeybee silk has different functional properties to β-sheet silks that have been examined. Honeybee silk is tougher, primarily as a function of its greater extensibility, and unlike silkworm silk, it retains its properties when wet (Hepburn et al. 1979).

The social Aculeata are a spectacularly successful group; they dominate the terrestrial ecosystem in terms of animal biomass and energy consumption. Availability of an efficient building material is a prerequisite for the ability to build nests and protect juveniles, which is intrinsically linked to sociality (Wilson 1971). We propose that the development of a tough coiled coil silk has contributed to the huge evolutionary success of these insects.

Acknowledgments

Literature Cited

Atkins EDT. 1967. A four-strand coiled coil model for some insect fibrous proteins. J Mol Biol. 24:139–141.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 340:783–795.

Bini E, Knight DP, Kaplan DL. 2004. Mapping domain structures in silks from insects and spiders related to protein assembly. J Mol Biol. 335:27–40.

Brown JH, Cohen C, Parry DAD. 1996. Heptad breaks in α-helical coiled coils: stutters and stammers. Proteins. 26:134–145.

DeGroot AP. 1953. Protein and amino acid requirements of the honeybee (Apis mellifera). Physiologia Comparata et D'Ecologia. 3:197–285.

Delorenzi M, Speed T. 2002. An HMM model for coiled coil domains and a comparison with PSSM-based predictions. Bioinformatics. 18:617–625.

Deng Y, Liu J, Zheng Q, Eliezer D, Kallenbach NR, Lu M. 2006. Antiparallel four-stranded coiled coil specified by a 3-3-1 hydrophobic heptad repeat. Structure. 14:247–255.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. Nature. 319:199–203.

Gatesy J, Hayashi C, Motriuk D, Woods J, Lewis R. 2001. Extreme diversity, conservation, and convergence of spider silk fibroin sequences. Science. 291:2603–2605.

Grimaldi D, Engel MS. 2005. Evolution of the insects. Cambridge: Cambridge University Press.

Heimburg T, Schünemann J, Weber K, Geisler N. 1999. FTIR-spectroscopy of multistranded coiled coil proteins. Biochemistry. 38:12727–12734.

Hepburn HR, Chandler HD, Davidoff MR. 1979. Extensometric properties of insect fibroins: the green lacewing cross-β, honeybee α-helical and greater waxmoth parallel-β conformations. Insect Biochem. 9:69–77.

Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5:150–163.

Lazaris A, Arcidiacono S, Huang Y, Zhou JF, Duguay F, Chretien N, Welsh EA, Soares JW, Karatzas CN. 2002. Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. Science. 295:472–476.

Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. 2003. Protein disorder prediction: implications for structural proteomics. Structure. 11:1453–1459.

Liu J, Lu M. 2002. An alanine-zipper structure determined by long range intermolecular interactions. J Biol Chem. 277:48708–48713.

Liu J, Rost B. 2001. Comparing function and structure between entire proteomes. Protein Sci. 10:1970–1979.

Pace CN, Scholtz JM. 1998. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J. 75:422–427.

Rudall KM. 1962. Silk and other cocoon proteins. In: Florkin M, Mason HS, editors. Comparative biochemistry Vol. IV. New York: Academic Press. p. 397–433.

Scheller J, Gührs KH, Grosse F, Conrad U. 2001. Production of spider silk proteins in tobacco and potato. Nat Biotechnol. 19:573–577.

Su JY, Hodges RS, Kay CM. 1994. Effect of chain length on the formation and stability of synthetic α-helical coiled coils. Biochemistry. 33:15501–15510.

Sutherland TD, Campbell PM, Weisman S, Trueman HE, Sriskantha A, Wanjura WJ, Haritos VS. 2006. A highly divergent gene cluster in honeybees encodes a novel silk family. Genome Res. 16:1414–1421.

Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (*and Other Methods). Sunderland(MA): Sinauer Associates.

Vollrath F. 2000. Strength and structure of spider silks. Rev Mol Biotechnol. 74:67–83.

Vollrath F, Knight DP. 2001. Liquid crystalline spinning of spider silk. Nature. 410:541–548.

Wilson EO. 1971. The insect societies. Cambridge(MA): Harvard University Press.

Woolfson DN. 2005. The design of coiled coil structures and assemblies. In: Parry DAD, Squire JM, editors. Fibrous proteins: coiled-coils, collagen and elastomers. San Diego (CA): Elsevier. p. 79–112.