Tarantulas and Social Spiders:

A Tale of Sex and Silk

by

Jonathan Bull BSc (Hons) MSc ICL

Thesis

Presented to the Institute of Biology of

The University of Nottingham

in Partial Fulfilment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Nottingham

May 2012

DEDICATION

To my parents…

…because they both said to dedicate it to the other…

I dedicate it to both

ACKNOWLEDGEMENTS

ABSTRACT

Studies of spider silks indicate that they may outperform virtually all synthetic fibres in terms of strength, elasticity and toughness. To date, most silks studied come from only a select few species and likely underrepresent the immense diversity of the clades. Here, protein and mRNA sequence analyses were used to study silk from two types of spider. The first approach used ESI tandem mass spectrometry to sequence peptide fragments of a silk from a tarantula (Mygalomorphae, Theraphosidae), a hitherto neglected family. The results confirm that the common silk types found in araneomorph spiders, Spidroin 1 and Spidroin 2, are also found in mygalomorphs. A putative N-terminal domain that bears a striking similarity to the N-terminus of araneomorph pyriform silk was isolated. If correctly identified, this would be the first ever recorded N-terminal domain for a mygalomorph. The second approach taken was to construct a cDNA library from theraphosid silk glands and adjacent tissue. Sequencing identified a significant number of uniquely truncated rRNAs. These may be the result of specific 'fragile sites' within these transcripts, which would explain the discrete classes of length polymorphisms found. The cDNA library sequences also provided evidence consistent with RNA editing and furthermore identified the presence of both transcribed nuclear pseudogenes and transposable elements. These may reflect past evolutionary horizontal gene transfer events within the spider genome. Similar analysis of next generation sequencing data from the transcriptomes of three *Stegodyphus* spp. (Araneomorphae) reveal a range of apparent silk types with similarity to major ampullate, minor ampullate and pyriform silks. These were identified by searching for comparative sequence homologies using Microsoft Office Word. No flagelliform silk or recognisable sticky silks were identified, which is consistent with the biology of *Stegodyphus* species. In addition to studies of silk, previous common conceptions of dimensional morphologies were examined to see if they could adequately sex theraphosid spiders, including the species that was the subject of the silk study already described. An independent samples *t*-test was conducted to compare morphologies of particular leg hairs and statistical analysis demonstrated that there were significant differences between males and females ($t(70) = 9.445$, $p < .001$). This technique may be important in future evolutionary and ecological studies of theraphosids.

Keywords: Silk, major ampullate spidroin, dragline, tarantula, transcriptomics, proteomics, spines, cDNA library

TABLE OF CONTENTS

LIST OF FIGURES

X

LIST OF TABLES

GLOSSARY

Word	Definition
Abdomen	The posterior division of the spider body.
Accessory Claws	Serrated, thickened hairs near the true claws in some spiders.
Aciniform Gland	Produces the swathing silk.
Aggregate Gland	Produces the glue droplets coating the web.
Ampullate Glands	Non-sticky dragline silk. Silk from the minor ampullate gland is only half as strong as that from the major gland.
Anal Tubercle	A small projection, dorsal to the spinnerets, carrying the anal opening.
Annulations	Rings of pigmentation around leg segments.
Antennae	The segmented sensory organs on the heads of insects, Crustacea, etc, but missing in all arachnids.
Anterior	Nearer the front or head end.
Apomorphic	Distinguished by characteristics.
Apophysis	An outgrowth.
Appendage	Parts or organs (such as the legs, spinnerets, chelicerae) that are attached to the body.
Arachnida	A principal division, or class, of the air-breathing arthropods, which includes the scorpions, mites, spiders, harvestmen, etc.
Arachnologist	One who studies arachnids.
Araneae	The ordinal name of all spiders; same as Araneida.
Araneology	The study of spiders.
Arthropoda	The phylum including all 'jointed-legged' species including insects, arachnids and Crustacea.
Attachment Disc	The series of tiny lines that serve to anchor the draglines of spiders.
Autophagy	The eating of an appendage shed from the body by autotomy or otherwise.
Autospasy	The loss of appendages by breaking them at a predetermined locus of weakness when pulled by an outside form; frequent in spiders and arachnids.
Autotomy	The act of reflex self-mutilation by dropping appendages; unknown in the arachnids.
Ballooning	Aeronautical dispersal by means of air currents acting on strands of silk.
Book Lungs	The respiratory pouches of the arachnids, filled with closely packed sheets or folds to provide maximum surface for aeration; believed to be modified insunk gills.
Branchial Operculum	A sclerotised hairless plate overlying the book lung.
Calamistrum	The more or less extensive row of curved hairs on the hind metatarsi, used to comb the silk from the cribellum.
Carapace	The hard dorsal covering of the cephalothorax in the Arachnida.
Cephalothorax	The united head and thorax of Arachnida and Crustacea.
Chelicerae	The pincer-like first pair of appendages of the arachnids; in spiders two-segmented, the distal portion or fang used to inject venom from enclosed glands into the prey.
Chitin	A linear homopolysaccharide found as the characteristic component of the cuticle of arthropods.
Claw Tuft	A bunch of hairs at the tip of the leg tarsus in spiders with only two claws.
Clypeus	The area between the anterior row of eyes and the anterior edge of the carapace.

Colulus	The slender or pointed appendage immediately in front of the spinnerets of some spiders.
Coxa	The segment of leg nearest the body; modified in the palp to form the maxilla.
Cribellum	A sieve-like, transverse plate, usually divided by a delicate keel into two equal parts, located in front of the spinnerets of many spiders.
Cuticle	The hard outer covering of an arthropod.
Cylindriform Gland	Produces egg sac silk.
Cymbium	The broadened, hollowed-out tarsus of the male palp within which the palpal organs are attached.
Distal	Pertaining to the outer end, furthest away from the body or point of attachment.
Diving Bell	A silken construct used by underwater spiders to capture air bubbles.
Dope	The pre-extruded silk.
Dorsal	Pertaining to the upper surface.
Dorsum	The upper surface.
Ecdysis	Moulting; the periodic casting off of the cuticle.
Embolus	The structure in the male palp, containing the terminal part of the ejaculatory duct and its opening.
Entelegyne	The group of spiders in which the females have an epigyne.
Epigastric Fold	A fold and groove separating the anterior part of the ventral abdomen (with epigyne and book lungs) from the posterior part.
Epigynum	The more or less complicated apparatus for storing the spermatozoa, immediately in front of the opening of the internal reproductive organs of female spiders.
Exoskeleton	The hard external, supportive covering found in all arthropods.
Exuviae	The parts of cuticle cast off during moulting.
Fang	The claw-like part of each chelicera; the poison duct opens near its tip.
Fecundity	The ability to reproduce.
Femur	The thigh; usually the stoutest segment of the spider's leg.
Fibril	A smaller unit of larger fibres.
Fibroin	The structural component of silk.
Fibrous	Resembling fibres.
Flagelliform Gland	Produces the core fibres of sticky silk.
Genitalia	All the genital structures.
Glandula Coronatae	Produces the adhesive threads.
Glandula Aciniformes	Produces threads for the encapsulation of prey.
Glandula Aggregata	Produces the sticky material for the threads.
Glandula Ampulleceae	Major and minor are used for the silk of the walking thread.
Glandula Pyriformes	Produces the attachment threads.
Glandula Tubuliformes	Produces thread for cocoons.
Gossamer	A light film of silk threads or groups of these floating through the air.
Haemolymph	The circulatory fluid in arthropods.
Haplogyne	The group of spiders in which the females have no epigyne.
Head	The part of the carapace carrying the eyes which is separated from the thorax by a shallow groove.
Hemimetabolous	Undergoing incomplete metamorphosis.

Holometabolous	Undergoing complete metamorphosis.
Labium	The lip, under the mouth opening and between the maxillae, attached to the front of the sternum.
Lateral	Pertaining to the side.
Lyriform Organ	A sensory organ near the distal end of limb segments formed of a group of parallel slit organs.
Malpighian Tubules	Glands surrounding the mouthparts of some arthropods.
Maxilla	The mouthparts on each side of the labium which are the modified coxae of the palps.
Median	In the midline or middle.
Median Apophysis	A sclerite arising from the middle division of the male palpal organs.
Metamorphosis	An abrupt change in morphology during development in arthropods.
Metatarsus	The sixth segment of the leg, counting from the body.
Mygalomorphae	An infraorder of spiders including tarantulas and their kin.
Nanocrystal	Any nanomaterial with at least one dimension ≤ 100 nm and that is single-crystalline.
Orb-Web	A two-dimensional web, roughly circular in design and, strictly speaking, a misnomer.
Palp/Pedipalp	Leg-like appendages on the cephalothorax of arachnids.
Palpal Organs	The more or less complex structures found in the terminal part of the adult male palp.
Paracymbium	A structure in the male palp branching from, or loosely attached to, the cymbium.
Patella	The fourth segment of the leg or palp, counting from the body.
Pedicel	The narrow stalk connecting the cephalothorax and the abdomen.
Pheromone	A chemical secreted by an animal in minute amounts which brings about a behavioural response in another, often of the opposite sex.
Phylogenetic	Pertaining to evolutionary relationships between and within groups.
Posterior	Near the rear end.
Process	A projection from the main structure.
Procurved	Curved as an arc having its ends ahead of its centre.
Prolateral	Projecting from, or on, the side facing forwards.
Proximal	Pertaining to the inner end; closest to the body or point of attachment.
Pseudoflagelliform	Silk similar in nature to flagelliform silk but with $GPQ(X)_n$ motifs rather than $GPG(X)_n$ motifs found in flagelliform silk.
Pyriform Gland	The gland that produces the attachment threads - attachment discs are made to anchor a thread to a surface or another thread.
Recurved	Curved as an arc having its ends behind its centre.
Reticulated	Like network.
Retrolateral	Projecting from, or on, the side facing backwards.
Scape	A finger, tongue, or lip-like projection from the midline of the female epigyne.
Sclerite	Any separate sclerotised structure connected to other structures by membranes.
Sclerotised	Hardened or horny; not flexible or membranous.
Scopula	A brush of hairs on the underside of the tarsus and metatarsus in some spiders.
Serrated	Saw-toothed.
Sexual Dimorphism	A difference in form, colour, size, etc., between sexes of the same species.
Slit Organ	A stress receptor in the exoskeleton.
Sperm Induction	The process of transferring the spermatozoa from the genital orifice beneath the base of the abdomen into the receptacle in the male palpus.

Sperm Web	A web of few or many threads on which male spiders deposit the semen prior to taking it into the palpus.
Spermathecae	The sacs or cavities in female spiders which receive and store semen.
Spiderling	A tiny immature spider, usually the form just emerged from the egg sac.
Spidroin	Spider silk, derived from the word 'fibroin'.
Spigot	A nozzle in the spinnerets from where silk is extruded.
Spine	A thick, stiff hair or bristle.
Spinnerets	The finger-like abdominal appendages of spiders through which the silk is spun.
Spinners	Paired appendages at the rear end of the abdomen, below the anal tubercle, from the spigots of which silk strands are extruded.
Spiracle	A breathing pore or orifice leading to tracheae or book lungs.
Stadium	The interval between the moults of arthropods; instar; a period in the development of an arthropod.
Sternum	A sclerotised plate between the coxae marking the floor of the cephalothorax.
Sub-adult	Almost adult; the last instar before maturity.
Synapomorphy	A character or a trait that is shared by two or more taxa and their most recent common ancestor.
Tarantula	A common name for members of the Theraphosidae.
Tarsus	The foot; the most distal segment of the legs, which bears the claws at its tip.
Tartipore	A cuticular scar that results after ecdysis forming in the exoskeleton.
Taxon	Any taxonomic unit (e.g. family, genus, species).
Taxonomy	The theory and practice of classifying organisms, part of systematics, the study of the kinds and diversity of organisms.
Tergites	Dorsal sclerites on the body; the hard plates on the abdomen of the atypical tarantulas that indicate the segmentation.
Thorax	The second region of the body of insects that bears the legs; in spiders, fused with the head to form the cephalothorax.
Tibia	The fifth division of the spider leg, between the patella and metatarsus.
Tracheae	The air tubes in insects; in spiders, tubular respiratory organs of different origin; by many thought to be modified book lungs.
Trichobothrium	(pl. Trichobothria) a long, fine hair rising almost vertically from a socket on the leg. Trichobothria detect air vibrations and currents.
Trochanter	The second segment of the leg or palp, counting from the body.
Ventral	Pertaining to the underside. The second segment of the leg or palp, counting from the body.

Adapted and modified from the Collins field guide (Roberts 1995) and American spiders (Gertsch 1979)

Table 1 IUPAC nucleotide codes for single nucleotides and their combinations

IUPAC nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base
. or -	gap

Table 2 IUPAC amino acid codes with single letter and triple letter abbreviations

IUPAC amino acid code	Three letter code	Amino acid
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Table 3 DNA codon table. Essentially the same as an RNA codon table except that thymine (T) is replaced by uracil (U).

Reproduced from Cusack *et al*. (2011)

1 INTRODUCTION

1.1 THE ORIGINS OF SPIDER SILK

Silk, perhaps one of the strongest of nature's fibrous proteins, has been utilised by the Arthropoda since the Early Devonian Period (~386 million years ago) by the most ancient ancestors of the true spiders found in the extinct Order: Uraraneida (formerly: Trigonotarbida, Shear *et al*., 1987). Previously thought to be the world's oldest spider was *Attercopus fimbriunguis* (Selden 2008), which lacked distinguishable spinnerets and had a segmented abdomen similar to today's whip-scorpions (Order: Thelyphonida as of 2005 (formerly Uropygi)), but instead produced silk from a series of spigots located on the ventral plates of its abdomen. The oldest of what can be considered the 'true spiders' developed approximately 100 million years later within the Carboniferous Period with those capable of weaving recognisable orb-webs or gum-footed webs arising much later within the Cretaceous Period, ~130 million years ago (Selden 1989).

Within the Class: Arachnida, there are several other individuals capable of producing silks, such as the pseudoscorpions (Order: Pseudoscorpionida) (Del-Claro and Tizo-Pedroso 2009) and mites (Hazan *et al*., 1975), though spiders use them for a far greater range of functions.

1.2 THE HISTORY OF SILKWORM SILK

The potential of silk obtained from the silkworm (*Bombyx mori*) was first recognised around 3000 BC (Hyde 1984) by Lady His-Ling-Shih, the wife of the 'Yellow Emperor', who then went on to invent the loom. Silk subsequently turned out to be such a valuable commodity that anyone found guilty of revealing its secrets or caught smuggling any live silkworms out of China was punishable by death. According to legend, China's domination over the silk market ceased when silkworm eggs and cocoons were smuggled out in the headdress of a Chinese princess betrothed to the then king of Khotan around the early 1st century AD (Hill 2009). Since then, silk has slowly spread across the world although, due to the climate required to maintain the silkworm, the main stronghold of the silk industry still resides in South-East Asia.

Silk is incredibly strong and resilient to most atmospheric conditions, enabling moth larvae to thrive (Tsukada *et al*., 1985). However, whilst *B. mori* silk has been tailored to suit the requirements of the moth, i.e. it is a tough material capable of protecting the larva during its pupation, the lepidopterans are not the only arthropod capable of producing silk (Zhou *et al*., 2000).

1.3   OTHER INSECTS THAT PRODUCE SILK

It is known that a number of other insects produce silk. These include: honey bees (*Apis mellifera*) (Craig 1997) and other Hymenoptera (Yamada 2004; Sutherland 2007), Trichoptera (Yonemura 2006), Diptera (Wieslander 1994) and Hemiptera (Chang *et al*., 2005), amongst others. These silks have highly variable interspecies compositions of amino acids (Figure 1.1). Some of these may not be considered typical silk variants because they are products created in the gut, or more specifically the Malpighian tubules (Zurovec *et al*., 1998). Insect silks can currently be grouped into 23 likely dependent lineages with no obvious linkages between the silk structure, the glands or the function (Sutherland *et al*., 2010). All of these lineages must have had common convergently evolved structures and properties relating to silk production. Silk manufacture occurs in three stages: glandular secretion, storage and extrusion (pultrusion). The protein, high in alanine, serine and/or glycine (ampullate silk) or proline (flagelliform silk) folds independently into a dominant secondary structure (Sutherland *et al*., 2010). The storage sac is responsible for accumulating and concentrating the silk dope prior to extrusion. The duct controls the secondary structure conformational changes prior to extrusion through terminal spigots. Unlike the spiders, which have developed a wide assortment of silks tailored to specific mechanical and environmental needs, insect silks tend to be far weaker in terms of tensile strength and elasticity (Collin *et al*., 2009). This is perhaps due to the presence of a wider variety of amino acid residues as opposed to the commonly found alanine, glycine, serine and proline within spider silk (Figure 1.1).
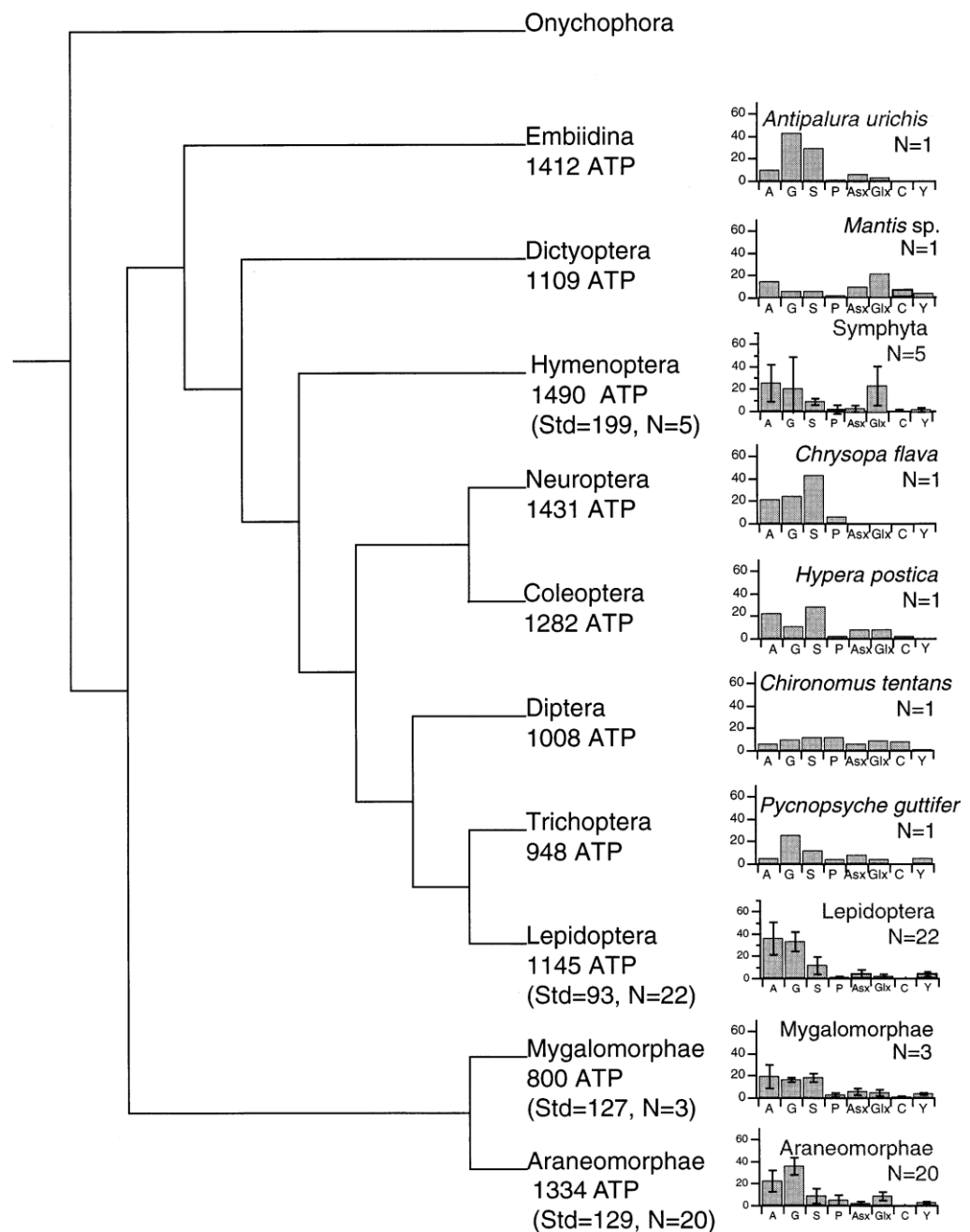
Figure 1.1 Phylogenetic relationships among silk-producing orders of arthropods for which amino acid data are available (Craig *et al.*, 1999). Structural tensile superiority seems to be conferred by the presence of a high relative abundance of alanyl and glycyl residues while elasticity is achieved with high prolyl residue content.

1.4   SPIDER DISTRIBUTION AND THE USE OF SILK FOR AERIAL DISPERSAL

Spiders are found on every continent except Antarctica although dispersal studies indicate that they probably traverse this region frequently. Unlike many insects, spiders do not migrate between these continents using wings. Instead, they employ a unique strategy termed ballooning/mechanical kiting (Figure 1.2), a method as yet understood to be used only by spiders (Bell *et al*., 2005; Woolley *et al*., 2007). This demonstrates one method of utilising their silk, as the method involves a single thread produced from their abdomen which is then caught by the wind. When the upward pull is great enough, they release themselves into the air. This strategy is employed by a wide range of species within the Araneomorphae, although it has also been observed (albeit, less frequently) in the Mygalomorphae (Enock 1885; Bristowe 1939; Coyle 1983). This discrepancy is perhaps due to the mygalomorph spiderlings' greater mass when compared to the average araneomorph. Adult linyphiids (Araneae) are even capable of ballooning as they are lighter than the typical mygalomorph spiderling.
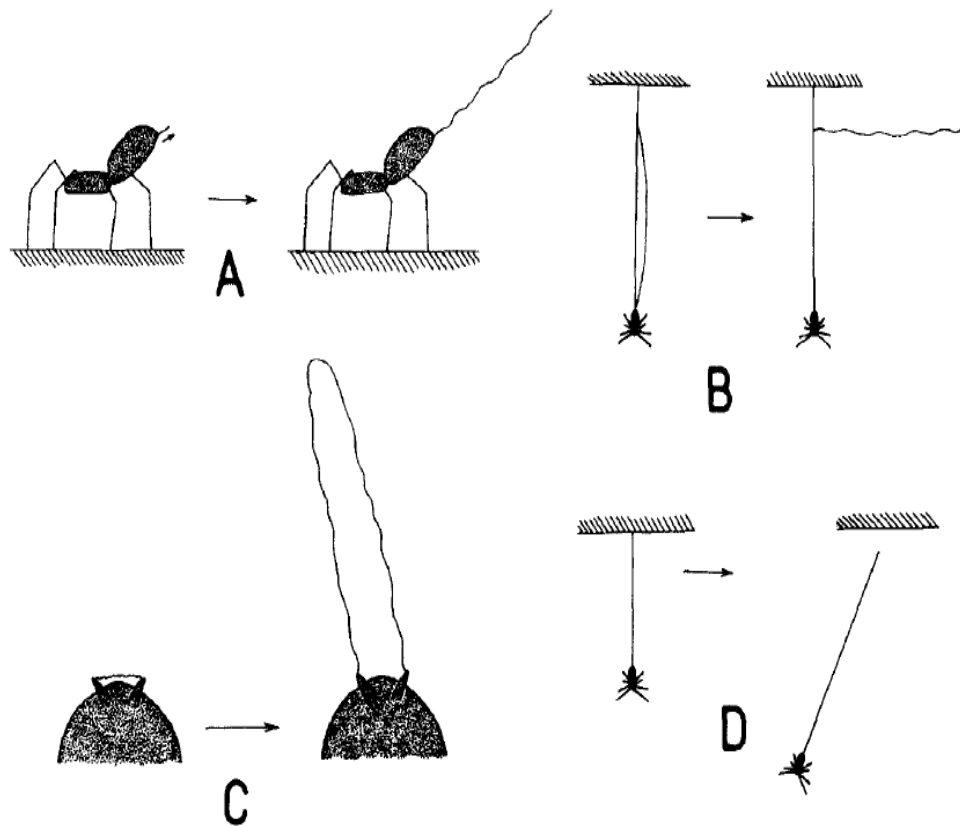
Figure 1.2 A diagrammatic representation of mechanisms of initiation of airborne line production. A, Bristowe (1939) single line method; B, Savory (1952) two-line break method; C, Blackwell (in McCook 1889) loop method; D, Coyle (1983, 1985) and Braendegaard (1938) suspension method. Adapted from: Eberhard 1987. This technique facilitates efficient colonisation and dispersal and explains how spiders are found on all the main inhabited continents.

1.5 SPIDER SILK

Spiders are tenacious creatures and have successfully adapted to a wide variety of climates, including tropical rainforests, arid deserts and mountainous regions. Arguably, many of these successful adaptations would not have been possible without the ability to produce silk.

Spiders use silk for many purposes such as capturing prey (Zschokke 1996), the preparation/swathing of prey (La Mattina *et al*., 2008), egg sac production and transportation (Partridge 1978), shelter (Opell 1984), as a guide-rope (Garrido 2002) and, in the case of *Argyroneta aquatica* (Clerck), the construction of a diving bell enabling the spider to live under water (Schütz *et al*., 2007).

Many of the more common uses of silk by spiders are believed to be apomorphic. It is commonly thought that the initial purposes of silk were for egg protection (Zhou *et al*., 2005), prey capture or shelter building (Decae 1984). However, it has been argued that the former two were unlikely as similar compounds are used by annelids of the genus *Diopatra* to plug their burrows (Brenchley 1976). Spider silk is known to provide information on the number of settlers in a patch (Giraldeau 1997; Wagner and Danchin 2003). This was elaborated more fully by Bernard and Krafft (2002) using a spider (*Anelosimus eximius*) and a T-maze. It was concluded that spiders preferred localities laden with conspecific silk. This does not appear to be a behaviour unique to *Anelosimus*, as demonstrated with other species (Leborgne and Pasquet 1987; Schuck-Paimand and Jimenez Alonso 2001). Surprisingly, despite their far greater mass, even mygalomorphs have been known to construct rudimentary aerial capture webs (Coyle 1986).

7

The use of silk by araneomorph spiders can roughly be divided into two groups: the orb weavers (orbicularians) and others that use a web to ensnare prey and the 'RTA clade' (Coddington and Levi 1991). RTA refers to a probable synapomorphy of a retrolateral tibial apophysis on the male palpal tibia. Early popular conception hypothesised that the function of the RTA is to be merely a palpal positioner prior to haematodochal expansion (Gerhardt 1921, Gerhardt 1923 and Bristowe 1929; Coddington 2005). Recent studies have suggested it is used to secure the male pedipalp within the female epigyne in order to facilitate the intromission of the sperm-transferring embolus (Eberhard and Huber 2010; Huber 1995). The RTA clade encompasses 39 entelegyne families including, the funnel-web spiders (Agelenidae), tangled nest spiders (Amaurobiidae), pond water spiders (Cybaeidae), mesh-web weavers (Dictynidae), dwarf sheet spiders (Hahniidae), wolf spiders (Lycosidae), lynx spiders (Oxyopidae), nursery web spiders (Pisauridae), jumping spiders (Salticidae), ground/ant spiders (Zodariidae) and some groups of crab spider (Thomisidae). Research has indicated that members of the RTA clade average 23% higher fecundity than their ancestors, whilst orb spiders average 123% higher. This supports a link between the adaptive escape from cribellate silk (a fine, woolly silk, explained later) and increased resource allocation to reproduction in spiders (Blackledge *et al*., 2009a). Despite this, many families such as the Deinopoidea still use cribellate silk (or have reverted back), where it is utilised as a composite material along with pseudo-flagelliform fibres (flagelliform fibres being the most elastic of the silks). Silk of this nature can stretch to the initial fracture of the pseudo-flagelliform fibres

at which point the cribellate fibrils, working in tandem, start to take the strain

(Blackledge and Hayashi 2006).

1.6 SILK PROTEIN MACROSTRUCTURE

Silks are long, fibrous, filamentous proteins that have similar structural

properties to the keratins, collagens and elastins found in mammals. In fact, the

similarities to some of the clotting cascade proteins such as the fibrins are such

that they are even found circulating around lepidopteran haemolymph,

presumably serving a similar role (Korayem *et al*., 2007). No similar studies

have thus far been conducted within the Arachnida. Silk within the Insecta is

produced in a variety of glands and from a range of structures (Figure 1.3). Silk

within the spiders is produced within silk glands and extruded through

spinnerets. Some spiders have only two types of silk gland (Mygalomorphae)

but there are known to be up to seven (Vollrath and Knight 2001), although no

single spider has been found with all seven.

The silk proteins produced by spiders are often termed 'spidroins'[1].

Research on the structure of spidroins is still at a relatively early stage. Most

research to date has been carried out on *Nephila clavipes* because these spiders

weave large webs with high silk strength inferred by the ability to capture

small birds. This would therefore be a sensible candidate spidroin if tensile

strength is the physical property of interest for biotechnological applications.

The silk itself is composed of a hydrophilic (Becker *et al*., 2003) humidity

sensitive flagelliform protein (Vollrath and Edmonds 1989; Edmonds and

Vollrath 1992; Adams *et al*., 1994; Liu *et al*., 2005) with a wide range of

_____

[1] Derived from the word 'fibroin' used to describe the silk protein produced by the silkworm, although
frequently fibroin is seen to be used interchangeably.

mechanical properties (Porter and Vollrath 2007; Emile *et al*., 2006). The size of the proteins themselves appear to be around 275 kDa (several times larger than human keratin (40 kDa) (Eckert 1988)) or 190 kDa when reduced with β-mercaptoethanol and run on a polyacrylamide gel (Mello *et al*., 1994).

Common name of insect group [higher classification]	Purpose of silk	Life stage/gland
Mayflies [Ephemeroptera: family Polymitarcyidae]	Lining for U-shaped tunnels in submerged wood	Larvae/apparently in Malpighian tubules
Webspinners [Embiidina]	Tunnels and egg coatings	All stages/Type III secretory units in prothoracic tarsomeres
Crickets [Orthoptera: Stenopelmatoidea in the families Gryllacrididae and Anostostomatidae]	Binding leaves together for construction of cocoon-like nests, linings for sand burrows	All stages/labial glands
Water beetles [Coleoptera: family Hydrophilidae]	Silken rafts to support eggs	Adult female/colleterial glands
Lacewings [Neuroptera, found within four of the sixsuperfamilies]	Egg stalks or egg coverings	Adult females/colleterial glands
Sawflies and parasitic wasps [Hymenoptera]	Cocoons, nests, and webs	Larvae/labial gland
Bees, ants, and wasps [Hymenoptera: Apoidea and Vespoidea]	Nests and cocoons	Larvae/labial gland
Fleas [Siphonaptera]	Cocoons	Larvae/labial glands
Dance flies [Diptera: family Empididae in the subfamily Empidinae]	Silk-wrapped nuptial gifts	Adult males/Type III secretory units in prothoracic basal tarsomeres
Glowworms [Diptera: family Keroplatidae]	Nests/prey capture threads	Larvae/labial glands
Butterflies, moths, caddisflies [Lepidoptera, Trichoptera]	Cocoons (aquatic and terrestrial), tunnels, retreats, communal webs, prey capture nets	Larvae/labial gland

Figure 1.3 Summary of the function of a range of insect silks (Sutherland *et al*., 2010).

Silk proteins characterised from spiders thus far appear highly conserved, particularly at the termini (although, as stated above, the range of species studied is limited). When the NCMAG2 (*Nephila clavipes* major ampullate gland) (Sponner *et al*., 2005a) and ADMAG1 (*Araneus diadematus* major ampullate gland) spidroins were compared, there was 99% identity (Hayashi and Lewis 1998). Spider silks appear to be conserved only at an amino acid level; the nucleotides appear to be extremely varied with no

apparent codon bias. Upon inspection of the silk sequence, the number of repeats is immediately noticeable (Xu and Lewis 1990 and Hinman and Lewis 1992, Figure 1.4), making cloning (insertion into a vector) particularly difficult because the sequence similarity results in a high chance of recombination. Recombination occurs when homologous repetitive regions within the repeats align, resulting in either a deletion or insertion.

Research has shown that the poly-alanine regions of silk predominantly form β-sheets while the glycine rich regions form the disorderly regions that consist of 3^{10}-like helices, spirals, β-turns (Keten and Buehler 2010) and spacers of unknown function (Teulé et al., 2009, Figure 1.8a/b). This no doubt accounts for a soft, entropically elastic section and a hard damageable fraction, which was examined in more detail by De Tommasi et al. (2010). These antiparallel β-sheet crystals at the nanoscale consist of highly conserved poly-(glycine-alanine) and poly-alanine domains (Hayashi et al., 1999). Modifications of these X-glycine-glycine regions, with X being alanine, tyrosine, glutamine or leucine, resulted in β-sheet structure formation in all examples except leucine. Tyrosine, glutamine and alanine therefore contribute to the formation of the glycine-rich β-sheet structure as shown by FT-IR spectroscopy (Fukushima 2000). This is surprising as the key molecular interactions within β-sheets are hydrogen bonds (Keten and Buehler 2008 a,b; Figure 1.6), one of the weakest bonds, but these β-sheet nanocrystals (Heim et al., 2010) which are confined to only a few nanometres, achieve a much higher stiffness than larger crystals (Keten et al., 2010).

More than 50% of the total silk is comprised of β-sheet structure (Grubb and Jelinski, 1997; Rousseau et al., 2004; Du et al., 2006). This results

in protein with extraordinary mechanical properties despite relying on individually weak hydrogen bonds (Keten *et al*., 2010; Qin and Buehler 2010). Stretching these β-sheet nanocrystals reinforces the macromolecular chains by interlocking, which transfers the load between the chains (Lefevre *et al*., 2007; Brockwell *et al*., 2003; Buehler and Yung 2009).

Silk	Ensemble Repeats
MaSp1	GGAGQGGYGRGGAGQGGAGAAAAAAAA Poly(A) blocks, (GA)$_n$ and GGX
MaSp2	GGAGPGRQQGYGPGSSGAAAAAAA GGX, GPGXX, (GA)$_n$ and poly(A) blocks
MiSp1	GAGAGAGAAAGAGAGAGGAGYGGQGGYGAGAGAGAAAAGAGAGGAGGYGR GGX, GA$_n$, poly(A) blocks and spacer
MiSp2	GAGVGAGAAAGFAAGAGGAGGYR GGX, GA$_n$, poly(A) blocks and spacer
Flag	ISEELTIGGAGAGGVGPGGSGPGGVGPGGSGPGGVGPGGSGPGGVGSGGSGPGGVGPGGSGPG GVGSGGFGPGGIGPGGSGPGGVGPGGVGGPYGPGGSGPGGAGGAGGSYGPGGPYGPGGSGGP GGAGGPYGPGGAGGPYGPGGPYGPGGAGGPGGEGPGGAGGPYGPGGPGGAGPGGYGPGGAGP GGYGPGGAGPGGYGPGGAGSGGYGPGGAGPGGYGPGGPGPGGYGPGGAGPGGYGPGGTGPGG AAPGGAGPGGAGPGGYGPGGSGPGGYGPGGGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAG PGGAGPGGAGPGGAGPGGAGPGGAGPGGVGTGGLGRGGAGRGGAGRGGAGRGGAGRGGAG RGGTGGVGGAGGAGGAGGVGGAGGSGGTTVIEDLDITIDGADGPIT GPGGGX$_n$, GGX and spacer
AcSp1	GSAGPQGGFGATGGASAGLISRVANALANTSTLRTVLRTGVSQQIASSVVQRAAQSLASTLGVD GNNLARFAVQAVSRLPAGSDTSAYAQAFSSALFNAGVLNASNIDTLGSRVLSALLNGVSSAAQG LGINVDSGSVQSDISSSSSFLSTSSSSASYSQASASSTSGAGYTGPSGPSTGPSGYPGPLGGGAPFGQ SGFG Poly(S) blocks and GGX
TuSp1	RTVGVGASPFQYANAVSNAFGQLLGGQGILTQENAAGLASSVSSAISSAASSVAAQAASAAQSSA FAQSQAAAQAFSQAASRSASQSAAQAGSSSTSTTTTTSQAASQAASQSASSSYSAASQSAFSQASS SALASSSSFSSAFSSASSASAVGQVGYQIGLNAAQTLGISNAPAFADSVSQAV Poly(S), GX, AAQAASAA, AAAQA, AASQAA and SQ$_n$
ECP-1	AGVGNNARFINGAGNNWSVSSMSGAGAFSGRRNSVYSGSSAGAAAGAHAASGGRAGAVAGAG AGASARAGAGARAAAGAGAGASAEAGAGARAAAGAGAGTGAGSGAGAAAGAGAAATSNAQA GAAVGSRGRASAGSRARAASFSEANTLAGAGASSNARAASFSGANALAGAGSRAGAEAQAGAR AGAGAASEASAAASAEARAGARAGAGAASEASAAASAEANAGARAVAGAGASAGAESNAGAK AVTRGRARAAAGAGATASSSASSLASSLSEAASSSSSSSSSSSSSSSSSFGESLASSAASAASALGA Short poly(A) blocks and GA repeats
ECP-2	GAGATAGAEAGAASGAAAGAGASSGAGAGAGASSGAGAGAGAGASSGAGARAGAGAGAGAG AGASSGADANAGAVASSGAGANAGAGASSGAGANAGVGAGAGAGANAGAGAGAGANAGAGA GAGANAGAGAGAGAGASSGAGAGAGAAASAGAEA Short poly(A) blocks and GA repeats

Figure 1.4 Comparison of single internal core ensemble repeats of orb or cob-weavers. Underlined sequences in AcSp1 and TuSp1 represent SQ and/or poly (S) modules. Areas coloured red indicate poly-alanine or GA repeats; blue letters denote GGX repeats; orange lettering represents GPGXX motifs; and pink-coloured regions denote GX iterations. Flag silk is from *Nephila madagascariensis* (AAF36091); AcSp1 is from *Argiope trifasciata* (AAR83925); MiSp1 (AAC14589) and MiSp2 (AAC14591) are from *Nephila clavipes*. ECP-1 and ECP-2 are from C-termini rather than ensemble repeats [*sic*] (Hu *et al*., 2006).

*Abbreviations explained in the glossary.

12
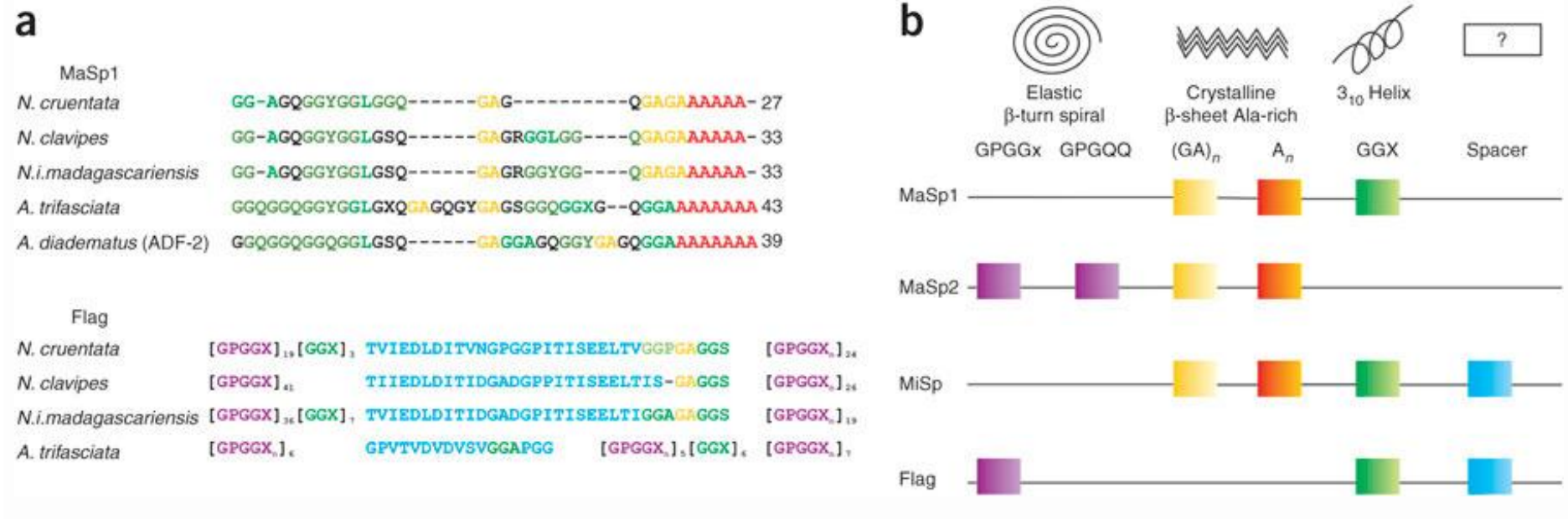
Figure 1.5 Alignment of the consensus repetitive sequences of (a) major ampullate (MaSp 1) and flagelliform (Flag) silk proteins. Structural amino acid motifs found consensus repeats of spider silk proteins (b). The square-coloured boxes indicate that the structural motif is part of the silk protein (Teulé *et al*., 2009).
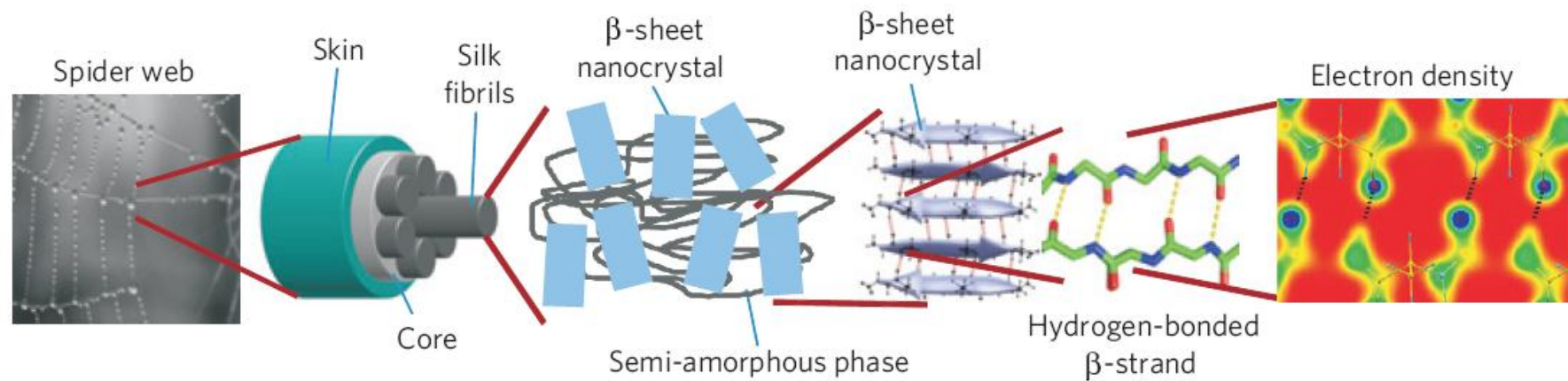
Figure 1.6 Hierarchical structure of spider silk. Key structural features of silk, including the electron density at the Ångström scale, hydrogen bonded β-strands, β-sheet nanocrystals embedded in a softer semi-amorphous phase (Termonia 1994) and silk fibrils, which assemble into macroscopic silk fibres (Keten 2010).

1.7 SILK EVOLUTION

The earliest known terrestrial arachnids appeared from the late Silurian to the early Permian and include the now extinct Trigonotarbida, along with Araneae, Amblypygi and Acari specimens as shown by fossil records found near Gilboa, New York (Shear *et al*., 1984, Shear *et al*., 1987, Shear *et al*., 1989 Dunlop 2010). Morphologically, these were highly dissimilar to the spiders we know today. Platnick and Gertsch (1976) divided the order Araneae into two suborders: the primitive Mesothelae and the Opisthothelae. The Mesothelae are known for various plesiomorphies including a segmented opisthosoma and an unusual sense organ between the tibia and metatarsi on the legs (Platnick and Goloboff 1985). The Opisthothelae on the other hand are what are usually considered as 'spiders' and, unlike the Mesothelae, have an unsegmented opisthosoma. This suborder is divided into the Araneomorphae (true spiders) and Mygalomorphae ('tarantulas'), with the divergence between these two suborders having occurred during the Triassic era around at least 240 million years ago (Selden and Gall 1992). The Araneomorphae are characterised by having one pair of book lungs, (although the primitive hypochilids still retain a second pair (Forster *et al*., 1987)) and chelicerae that point towards one another while the Mygalomorphae possess two pairs of book lungs and downward pointing chelicerae. In practice it is often the size that is used to distinguish these species as the Mygalomorphae regularly attain leg spans greater than 10 centimetres, with a maximum of around 30 centimetres in the case of *Theraphosa apophysis* (Tinter 1991).

The oldest known spinneret (and therefore earliest known record of spider silk) was found in middle Devonian rocks around 385 - 380 million years ago (Shear *et al.*, 1989). According to Shear, "the Devonian spinneret resemble those of members of the living suborder Mesothelae, but the number of spigots and their distribution are like those of members of the suborder Opisthothelae, infraorder Mygalomorphae." Spiders alone produce silk derived from opisthosomal (abdominal) glands through spigots located on an anterior reduced abdominal appendage now known as the spinnerets. It can be presumed that all of the known silks originate from one common ancestor such as an ancient *Liphistius* (Table 1.1). This species is generally thought to be one of the most basal genera. This is reflected in the morphology as, like the class Insecta, the species still has a segmented abdomen. *Liphistius* has a unique morphology with characteristics unlike the more derived species. It has highly differentiated spinnerets unlike the Araneomorphae and, although it has "single-articled posterior median spinnerets with numerous spigots" clustered near the tip like the Mygalomorphae, it lacks tartipores (cuticular structures) found in both the Mygalomorphae and Araneomorphae (Coddington 1989).

As it appears logical that silk is plesiomorphic, it would also seem sensible to assume that all spider silks share a similar nucleotide/amino acid sequence. As yet, this appears to hold but is biased towards species presumed to have commercially and structurally 'valuable' silks. Challis *et al.* (2006) compared a large number of these sequences and found conserved C-termini motifs, although whether these sequences are retained over the other infraorders remains to be seen.

It is not clear whether the spider evolved around the properties of silk or whether silk adapted to the needs of the spider. There is evidence to suggest that not only did spiders develop around this new structurally-superior protein but that there is also a sexual function (at least in respect of the Araneomorphae). A biomechanical model using available data has been drawn (Rodríguez-Gironés *et al.* 2010) which suggests that because Araneomorphae traverse suspended silk lines, there is a negative relationship between body size and traversing ability. This would favour a sexual size dimorphism because males use wind-caught bridging lines to find a mate. Web orientation (Opell *et al.*, 2006) and viscid silk (glue silk) (Bond and Opell 1998) have both been implicated as innovations relating to species diversification. Lower energetic costs (Opell 1996), UV reflectance (Craig *et al.*, 1994) and better adhesiveness and extensibilities may have all been responsible for the enhanced fitness pertaining to the viscid spiders as opposed to their cribellate deinopid (fuzzy, non-sticky webbed) counterparts (Kawamoto 2008).

Despite the development with regards to certain aspects of silk, some cribellate (species with a cribellum) representatives can have increased fitness due to UV-reflectant (Li *et al.*, 2004; Watanabe 1999) and vertical webs (Lubin 1986) employing more extensible (Opell and Bond 2000) and adhesive (Opell 1996) silks to attract and capture prey.

Table 1.1 Comparison between the silk-spinning apparatus of the primitive Mesothelae, an extinct Devonian fossil and a typical representative of the Mygalomorphae (Adapted from Shear *et al*., 1989).

Character	*Liphistius* (Mesothelae) posterior median spinneret	Devonian fossil Spinneret	Mygalomorph posterior median spinneret
Spigot arrangement	Single apical spigot	On mesal side of spinneret, not ranked, clustered at tip	Numerous on mesal side, not ranked, clustered at tip
Spigot types	One	One	Rarely one, usually two
Cuticle texture	Scaly	Less pronounced, usually scaly	Rarely one, usually two, slightly scaly
Shaft sculpture	Absent	Apparently absent	Present on at least distal third
Shaft-base union	Smoothly graded	Smoothly graded	Collar-like articulation
Tartipores	Absent	Absent	Present

1.8   PROTEIN CONSTRAINTS

It can be argued that web structure, strength and elasticity could reflect the fitness/health of an individual spider. Silk is subject to the same constraints as other proteins within an organism, its production is highly dependent on diet, and starved spiders would be predicted to weave either less silk or silk with inferior mechanical properties. The web density aspect was tested and proven with *Lactrodectus hesperus* (Salomon 2007). Boutry and Blackledge (2008) proposed two hypotheses with regards to diet: silk production could be dependent on physical condition, size or health or the production of silk could be modified according to the prey the spider expects to catch or has been catching.

A study conducted on *Nephila pilipes* (Tso *et al*., 2007) shows that the diameter and stiffness of the major ampullate (MA) silk varied according to the 'struggle' expected from the prey, such that those expecting crickets produced a thicker, stiffer silk as opposed to those fed exclusively on flies which were presented with a less stiff (i.e. more elastic) silk. Here it could also be argued that the variation in the silk's physical properties is solely dependent on diet as reflected in a study on *Argiope keyserlingi* (Craig *et al*., 2000). The energy rich diet of the lepidopterans provides a higher proportion of alanine and glycine as these can be created from carbohydrates but correspondingly less of the other amino acids. This is in contrast to spiders, which have a diet that is more diverse in amino acids but less energy-rich. (Craig 1999). A logical hypothesis was also made by Sutcliff (1963) who asserted that hemimetabolous insects (i.e. those that undergo incomplete metamorphosis) would have fewer free

amino acids circulating within the haemolymph, whereas the holometabolous species would have more. Again, this is open to critique as some of the hemimetabolous insects are indeed herbivores and would consume a less diverse supply of amino acids resulting in a downstream nutritional disadvantage to the spider.

Interestingly, there was also a dramatic change found in the percentage of amino acids according to spatial variation. A survey of *N. pilipes* by Tso *et al*. (2005) found similar results to Craig (1999) but on a spatial scale, in that habitat variation often came with a corresponding fauna change which led to higher amounts of proline and glutamine-containing β-turns and a lower percentage of alanine-containing β-sheet structures due to more energetic prey. It also appears that it is not just the spidroin that changes composition as different low molecular weight organic compounds are induced with a varied diet (Higgins *et al*., 2001). Even the pigment appears to be changeable according to the intensity and spectral composition within the environment (Craig *et al*., 1996). This is likely to have an effect on the visibility of the web according to the prey's visual acuity.

A paper by Madsen and Vollrath (2000) highlights an interesting twist on the idea that spidroin composition is altered in response to the environment. Their data show a diameter increase consistently and predictably ~18% post anaesthesia. This is accompanied by reductions in breaking strain, breaking energy and by an increase in initial modulus (Sirichaisit *et al*., 1999). Subsequent predictions were made on whether this is a result of relaxation of the extrusion die (sphincter relaxation) or conditions along the ducts. The inferences from previous studies might be altered in light of these findings,

particularly those that use the standard protocol of forcibly silking according to Work and Emerson (1982) such as Frische *et al*. (1997) and Shao *et al*. (1999a).

1.9 GLANDS

As mentioned above, spiders have the potential to make up to seven types of silk originating from up to seven different glands (Figure 1.7)

Each gland is responsible for a different type of silk and each silk has its own unique uses and properties. The aciniform glands are used for making silk to swathe prey. In many species these are accompanied by both a cribellum and a calamistrum. The cribellum is a flat plate on the ventro-posterior side of the female abdomen. It is essentially a modified pair of spinnerets arranged into thousands of microscopic nozzles out of which a very fine silk is extruded usually at a thickness of 10 - 100 μm (Okada 2008). One example of cribellar silk use is by net-casting spiders (*Deinopis* spp.) which weave a web utilising the spider's leg span. The net that is woven is made almost exclusively from cribellar silk, which is effectively used by the spider to lunge at passing insects. Other spiders use the cribellar silk for its woolly texture as many insects have spiny legs and cribellar silk is ideal for their capture. Once extruded, the calamistrum, which is effectively a comb on the fourth pair of legs in cribellate spiders is used to pull out and untangle or in many cases actually tangle the silk that has been pulled out.
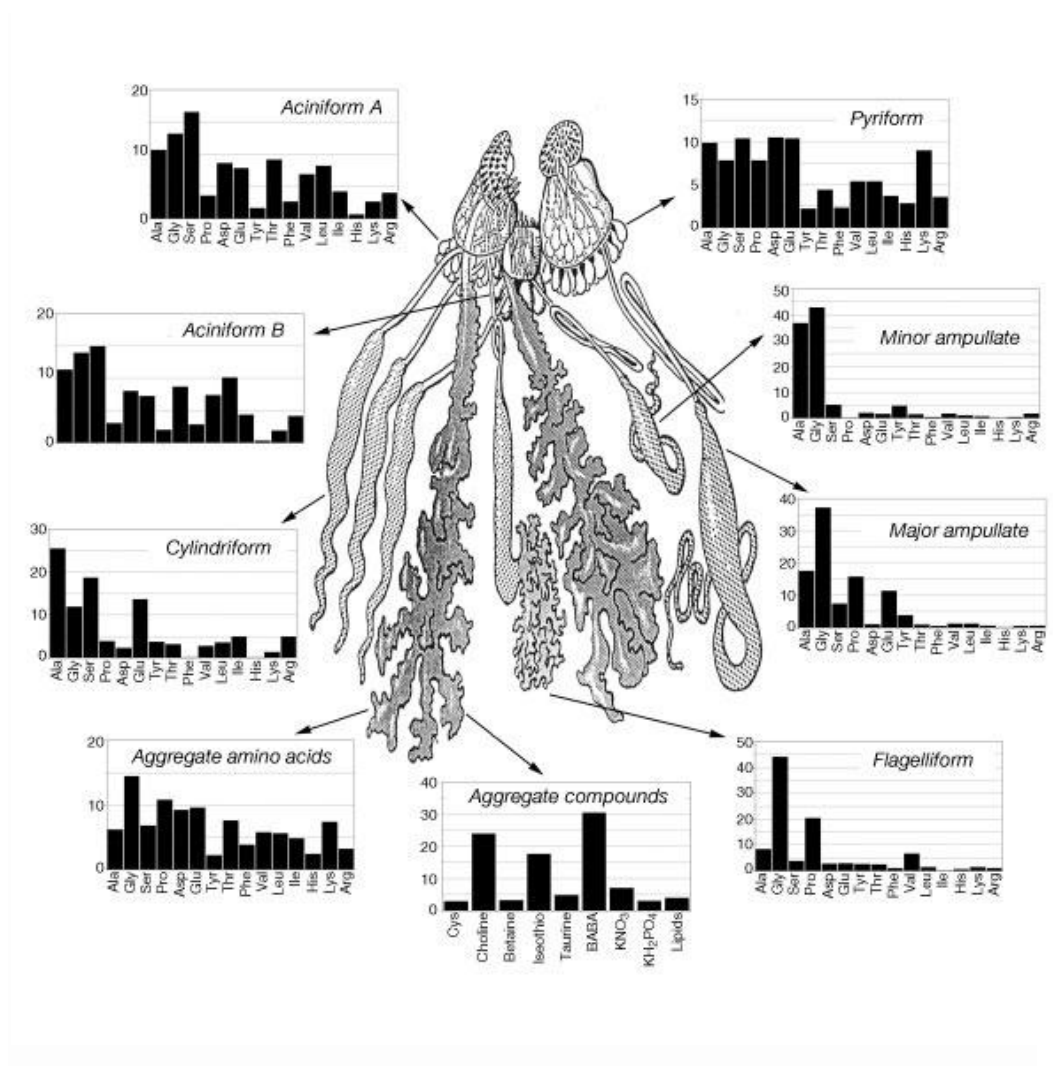
Figure 1.7 An illustration of the different silk and silk accessory protein producing glands in a typical spider. Minor glands provide threads that can be added to any structural thread (Vollrath and Knight 2001).

1.9.1 Cribellum and calamistrum

The cribellum and calamistrum have historically been used to classify araneomorph spiders into groups that have a cribellum (cribellate spiders) and those that do not (ecribellate). However, this distinction between families is now arbitrary as many families include both cribellate and ecribellate members. It is now believed that the cribellum was present in the earliest of spider ancestors and was lost in the apomorphic spiders (Coddington and Levi 1991). It is known that the cribellum is actually a homologue of a pair of

22

spinnerets in the Mesothelae and Mygalomorphae, both of which lack the cribellum. Instead, a similar structure has been integrated into the spinneret itself, which serves to separate the fibres (hence why tarantulas do not weave a single thread but rather a sheet of silk - author observation). Cribellar thread is mainly spun by the Araneomorphae, which comprise more than 95% of the ~42 000+ living spiders (Platnick 2012). Cribellate species are known from 22 families of araneomorph spiders: Agelenidae, Amaurobiidae, Amphinectidae, Austrochilidae, Ctenidae, Deinopidae, Desidae, Dictynidae, Eresidae, Filistatidae, Gradungulidae, Hypochilidae, Miturgidae, Neolanidae, Nicodamidae, Oecobiidae, Psechridae, Stiphidiidae, Tengellidae, Titanoecidae, Uloboridae and Zoropsidae although not all of them are exclusively cribellate, some containing both cribellar and ecribellar members (Griswold *et al*., 1999).

Although many spiders have retained the cribellum, the more common method of prey capture appears to involve the use of sticky droplets. These have more enhanced adhesive properties than that of cribellar silk, which has a tendency to peel away from a surface once a threshold limit has been exceeded at the edges (Opell and Hendricks 2007). Similarly, when compared with the cribellar thread of the Deinopoidea clade, viscous threads performed more uniformly over a range of insect hosts including fly abdomens, wings and smooth beetle elytra (Opell and Schwend 2007). The replacement of cribellar threads is also more efficient from an energy consumption perspective. The cost of producing sticky droplet thread that is recycled is estimated to be 66% less than that of cribellar thread (Opell 1998).

1.9.2 CYLINDRIFORM GLANDS

The cylindriform glands make the tubuliform silk. It is most often used to cover the egg sac (Hu *et al*., 2005; Tian and Lewis 2005) and is similar in amino acid composition to ampullate silk (Barghout 1999).

1.9.3 PYRIFORM GLANDS

The pyriform glands are responsible for the attachment discs that are used to fasten the major ampullate silk to surfaces (Perry *et al*., 2010) and for the attachment of flagelliform threads to major ampullate scaffolding threads (Kovoor and Zylberberg 1980). The composite produced here has the largest percentage of polar residues compared to other silks with fewer small side-chain amino acids, the rest comprising charged residues (Andersen 1970).

1.9.4 AMPULLATE GLANDS (MAJOR AND MINOR)

1.9.4.1 Major ampullate glands

The ampullate glands are composed of a major and minor set. The major ampullate gland makes the dragline silk which is used for descending from the web, escaping from danger and when a strong silk is required, such as for the web frame. It has an extremely high tensile strength and a limited amount of elasticity; i.e. it is not comparable to flagelliform silk (Gosline 1984). Once secreted, the solution is coated in a spidroin-like protein, a glycoprotein and then a lipid layer (Hardy *et al*., 2008 and Vollrath and Tillinghast 1991). There has also been a suggestion after working on the spider

Nephila senegalensis that a peroxidase gene (NsPox) has some role to play in the formation and/or processing of both the major and minor ampullate silk (MaSp/MiSp) (Pouchkina *et al*., 2003) but the precise nature of this potential interaction is not yet known.

1.9.4.2 Minor ampullate glands

The smaller ampullate gland is responsible for similar functions to the major ampullate glands but it produces a lower volume of silk and the fibre is less elastic than its major ampullate counterpart. This is in part due to a far lower proline and glutamine content (Andersen 1970). This silk is used to build the temporary spiral (a frame on which the spider can move so as to build a more permanent, stronger spiral) and also stabilises the web during construction. It is now known that this spiral is consumed once the capture spiral has been constructed, and therefore it is reasonable to assume that this is evolutionarily beneficial from an energy conservation perspective.

1.9.4.3 Dragline silk

Dragline silk has been the most extensively studied of the seven silks that spiders can produce due to its extremely high tensile strength and toughness. Its strength and elongation at breaking makes it superior to the best man-made fibres (Jelinski *et al*., 1999) and per unit weight far surpasses high tensile steel (Heslot 1998). It is composed of three layers (Knight and Vollrath 2001a): the core in which the canaliculi are found, a skin, thought to be made of glycoprotein and a thin filamentous layer sheathing the whole structure (Figure 1.8).

schematic top view of a silk thread
(sketch not to scale)
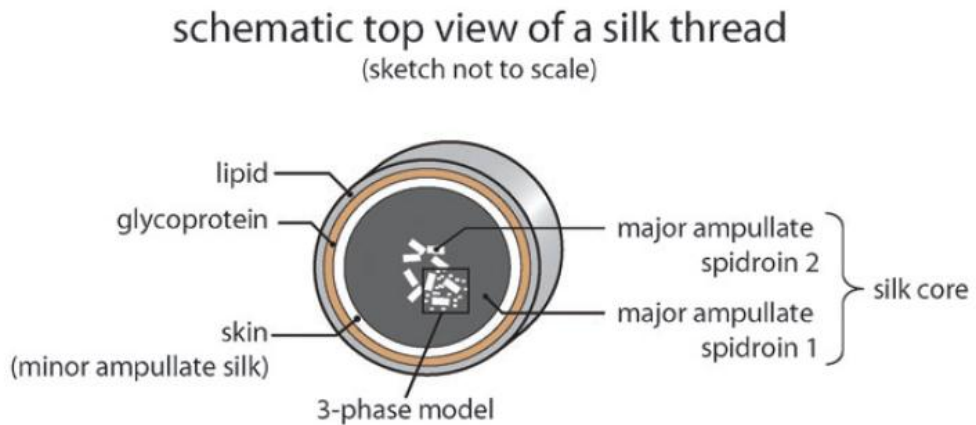
Figure 1.8 Core-skin structure of a silk thread. The proteinaceous core is subdivided into two distinct parts. A skin made of MIS surrounds a region where inhomogenously distributed MaSp2 clusters are embedded within the homogenous MaSp1 phase. The core is covered by a glycoprotein and a lipid-like layer (Heim *et al*., 2010).

1.9.5 FLAGELLIFORM GLANDS

It is the flagelliform glands that produce the silk that form the capture spiral of orb-webs. The threads are extremely elastic (Becker *et al*., 2003) and as one might expect they have a proportionately higher number of proline residues (Andersen 1970). Working in tandem with the unique amino acid motifs are structures termed windlasses (Blackledge *et al*., 2005). These are droplets of adhesive at the ends of the flagelliform filaments into which flagelliform silk is extruded. The purpose is to feed out additional silk upon impact from the prey onto the web. These three properties: the plasticisation of the web, the windlasses and the unique amino acid composition all play an essential role in the capture spiral's extensibility (Blackledge *et al*., 2005). The elastic capture spiral works in tandem with the aggregate glands along with several other glands, which coat it in sticky droplets of glycoproteins (Sahni *et al*., 2010). Along with the high elasticity of the flagelliform silk, this aids in arresting the momentum of flying insects.

1.9.6 ACINIFORM GLANDS

Aciniform silk is used for multiple purposes, including lining egg sacs and as padding to protect the developing spiderlings. It is also mixed in with the pyriform silk which acts as a cement to thicken it (Hayashi *et al*., 2004). These glands also work alongside the cribellum and the calamistrum to produce fuzzy silk of the stabilimenta, which are thought to act as warning signs to birds that might inadvertently fly into the web and destroy it (Blackledge and Wenzel 1999; Lubin 1975).

The orbicularian stabilimentum silk has been found to reflect slightly more UV light than white light and probably acts as a predatory defense as opposed to a prey attractant (Zschokke 2002). Peters (1993) suggested that silk production within the *glandulae aciniformes* and subsequent stabilimenta abundance was inversely proportional to egg sac lining. Walter *et al.* (2008) expanded upon these data by revealing a positive feedback mechanism in which swathing of multiple prey items induced an enhancement of glandular secretion, the excess of which would be expended as suggested by Peters (1993).

1.9.7 AGGREGATE GLANDS

The aggregate gland is not strictly a silk gland but rather a chemical factory making the accessory proteins responsible for optimal web performance. Numerous organic compounds, glycoproteins and salts have been identified as being produced by the gland (Vollrath and Tillinghast 1991). These compounds are used to aid in the maintenance of the web and presumably performing functions like UV protection, water absorption,

temperature regulation or to serve as antimicrobials/antifungals. This gland is

responsible for producing the droplets that glue prey items to the web. One

paper has described two protein components of these droplets, which

reportedly contain domains that are encoded by opposite strands of the same

piece of DNA (Choresh *et al*., 2009). When flattened on a microscope slide

these droplets are seen to contain a small opaque granule, which is thought to

be the glycoprotein glue. Opell and Hendricks (2010) tested this hypothesis and

found a negative correlation to thread stickiness. They subsequently

hypothesised that these granules serve to anchor transparent glycoproteins and

to generate adhesion to the thread to prevent slippage. A summary of the types

of silks made by spiders, their uses and properties is shown in Table 1.2.

Table 1.2 Summary of silks, their functions and glands of origin in the golden orb weaver *Nephila clavipes* (Foo and Kaplan 2002).

Dragline	Major Ampullate	Anterior	Orb-web frame, radii, safety line
Viscid	Flagelliform	Posterior	Prey capture, core fibres of adhesive spiral
Glue-like	Aggregate	Anterior and posterior	Prey capture, adhesive silk of spiral
Minor ampullate	Minor ampullate	Medial	Orb-web frame reinforcement
Cocoon	Tubuliform	Posterior	Reproduction
Wrapping	Aciniform	Anterior	Wrapping captured prey, inner egg sac
Attachment	Pyriform	Anterior	Attachment disc and joining fibres

1.10 NUCLEOTIDE/GENETIC STRUCTURE OF SILK GENES

Due to the abundance of silk that is being constantly produced and/or stored within a spider, large quantities of mRNA can easily be isolated. Cloning the cDNA on the other hand poses numerous challenges due to the particular characteristics of the spider silk genes. For one, while the protein can be anything up to 0.5 MDa, (typically found to be in the ranges of 260 and 320 kDa; Sponner *et al.*, 2004) clones containing MaSp1 and MaSp2 have been found to be up to 34 046 bp and 37 092 bp respectively for *Lactrodectus* species (Ayoub *et al.*, 2007). These may not be representative of cDNA lengths within the spider silks as until Gatesy (2001) sequenced a further five genera, only representatives from two genera *Araneus diadematus* (Guerette *et al.*, 1996) and *Nephila clavipes* (Xu and Lewis 1990) had been characterised. To this day, there are still comparatively few full-length spidroin cDNAs, although there are many partial C-terminal transcripts. Silks of the Mygalomorphae were only later to be sequenced (Garb *et al.*, 2007) with the first C-termini from theraphosid silks only very recently identified (Bittencourt *et al.*, 2010).

Secondary structure is abundant throughout mRNAs due to the prevalence of cytosine-rich codons for alanine, glycine and proline (Andersen 1970). Spiders appear to reduce the strain on their tRNA pools by balancing out these glycine and alanine codons ((G/C)(G/C)N) (Candelas *et al.*, 1990) with adenine or thymine residues in the third position, favouring them at almost 85% and 90% for MaSp1 and MaSp2 respectively in *Lactrodectus* (Ayoub *et al.*, 2007). Attempts have been made to avoid this codon issue by selecting for less prevalent codons to accommodate better to the bacterial host's tRNA pools when designing synthetic genes (Capello *et al.*, 1990;

Cantor 1994). This high secondary structure (three-dimensional configuration of local segments of DNA/RNA due to intramolecular hydrogen bonding) inhibits cDNA synthesis, which make sequencing especially difficult (Hayashi and Lewis 2001). Silk is also highly repetitive even at the nucleotide level and that repetitive nature is thought to promote errors from slippage during replication while the strings of glycine, alanine and proline are thought to create recombination hot-spots (Mita *et al*., 1994; Beckwitt *et al*., 1998). These iterations of highly homogenised repeats (~98 - 100%) identical at the nucleotide level are generally explained as a result of non-reciprocal recombination or unequal crossing over (Beckwitt *et al*., 1998; Hayashi *et al*., 2004). The combination of the GC-rich regions (~70%), secondary structure and recombination events make these proteins difficult to express (Xia *et al*., 2010). Amino acid substitutions appear to occur in both the repetitive and non-repetitive regions with far more occurring in the former. Suggestions have been made that the MaSp1 of the Araneomorphae may be exhibited in several forms either as a result of multiple genes or a single gene with a multiple intron/exon organisation (Tai *et al*., 2004).

Scheibel (2004) used *N. clavipes*, (which at the time was the species from which most data had been collected) and found that from the ~15.5 kb mRNA transcript originating from a 30 kb flagelliform silk locus the coding sequence was divided into 13 exons. Each of these exons encoded exactly one repeating unit with the N/C-terminal exons additionally containing a unique sequence. The first characterised cDNA was that of a flagelliform silk from the spider *N. clavipes*, chosen no doubt due to its large body size (2.5 cm) as well as the morphologically distinct flagelliform glands present within.

1.11 SILK PRODUCING CONDITIONS IN THE SPIDER

Spider silk is spun under normal atmospheric conditions, i.e. ambient temperature and at moderate pressure (Vollrath and Knight 1999). This transition from liquid to solid is far from understood but it is thought to occur in a two-stage process of nucleation (seeding) and aggregation (Li *et al*., 2001). According to research (Zhou *et al*., 2001) the former is far less thermodynamically favourable than the latter. This naturally seems sensible as aggregation is only favourable when the spider needs to perform a function with the silk but how the spider prevents premature aggregation of spidroin before self-assembly is not yet understood.

Evidence suggests that a monotonic drop in pH from 7.2 - 6.3 resulting in a change from random coil to a β-sheet rich conformation (Dicko *et al*., 2004), forces (Rammensee *et al*., 2008) and ionic gradients (Knight and Vollrath 2001b; Chen *et al*., 2004; Chen *et al*., 2008) are responsible for the conversion to solid silk. It also appears that the optimal pH for conformation transition occurs at 4.35, which is remarkably similar to its isoelectric point (pI, 4.22) (Dicko *et al*., 2004). There is evidence to suggest that metal ions such as potassium and copper play a part during both the nucleation phase and the transportation of the silk along the internal glands (Kerkam *et al*., 1991; Sehnal and Zurovec 2004; Chen *et al*., 2004). The cuticle that lines the duct has an advanced "hollow fibre dialysis membrane" which appears to facilitate the rapid removal of water through an ionic pump. This is followed by a structure termed the 'valve', which is thought to repair accidentally internally ruptured thread (Vollrath and Knight 1999). Once the chains have passed through the

spinneret they become stretched and hydrogen bonds form between them. These are then folded back on themselves giving rise to the β-sheet structure which accounts for approximately 60% of the resulting fibre (Carboni 1952). It is believed that a relay-like mechanism involving the N-terminal domain is responsible for inhibiting precocious aggregation (Askarieh *et al*., 2010). The spider itself can modify the diameter of its own silk at will prior to extrusion (Calvert 1998).

Barghusen *et al.* (1997) found that there was an optimal temperature range for web construction with *Achaearanea tepidariorum*. They were seen favouring 20°C and avoiding temperatures above 25°C. This suggests that there is either a preference by the spider (being ectothermic) to be within that range as they produce web more efficiently at that temperature or conversely they produce a more efficient web.

1.12 *IN VITRO* PRODUCTION OF SYNTHETIC SILK

When silk is being artificially extruded, phase transformation has been achieved by stretching (Hiroaki and Tetsuo 1990), extrusion in methanol (Jun and Yoshiko 1981) and by heating to remove the water (Jun *et al*., 1977). Artificial extrusion may be an efficient way of producing a fibre but the quality is often not comparable to the natural fibre. A similar problem is encountered when the fibre is dissolved in a neutral salt solution. After the fibre had been recovered, Zou *et al.* (2006) found that the mechanical properties were subsequently inferior and the biodegradability increased. Whether the loss of

tensile strength was due to breakage of the silk or to accessory proteins therein becoming unbound in the solvent is not known.

Silk is extraordinarily resilient to degradation over time and has even shown an increase in performance a year after extrusion (Agnarsson *et al*., 2008). The diameter has been shown to decrease over time (presumably due to molecular reorganisation) resulting in a much stiffer and higher yielding silk. The silk also retains its tensile strength for up to 4 years (Agnarsson *et al*., 2008). Work and Emerson (1982) devised a way in which to forcibly silk spiders, a method still of practical use today.

The fundamental problem with making synthetic silk is that although it has been sequenced and the tertiary structure, involving possible disulphide bridges, hydrophobic and ionic interactions, has been modelled, encouraging correct folding is far from easy. It is relatively easy to produce an artificial silk construct and clone it into an expression vector (Case and Thornton 1999). Fahnestock *et al.* (2000) have efficiently produced spider dragline silk analogues up to 100 kDa in size using microbial expression systems such as *Escherichia coli* and *Pichia pastoris*. They also stated that the advantage of these two systems is the ability to secrete into extracellular media circumventing the issue of overwhelming the available intracellular volume and isolation from the host's proteins.

The *P. pastoris* expression system is not limited by truncated synthesis and therefore has an advantage over *E. coli* in which truncated synthesis is prevalent, presumably as a result of ribosome termination errors. Here, ribosomes are found to pause due to an absence of cognate aminoacyl-tRNA, which often leads to termination of synthesis (Rosenberg *et al*., 1993). The

problem is not producing an identical sequence to the spiders but the extrusion method and post-translational modifications, water exclusion and alignment issues. The spider's glands do all these automatically and at a rate that makes current artificial methods seem primitive (Teulé *et al*., 2009).

There are numerous physical steps required to turn this liquid silk into a strong fibrous polymer. It is commonly assumed that the transition is achieved on contact with air. However, the production of silk by the diving bell spider (*Argyroneta aquatica*) should be considered. This spider spends most of its life under water, breathing from a diving bell constructed from silk extruded under water (Schutz 2007). Silk's secondary/tertiary structure appears to be achieved through numerous modifications including disulphide bond formation, cation interactions, glycosylations and many other theorised steps involving both chemical and physical means (Kaplan *et al*., 1992a, b).

Consideration should also be put into using other invertebrates as silk models. *In vitro* production of spider silk is not without certain inherent problems relating to the size and repetitive nature of the spidroins. In contrast, the silk of honey bees, which is composed of four small and non-repetitive proteins yields a substantial 0.2 - 2.5 g/l in *E. coli* and even self assembles into the native coiled structure replicating the tensile strength of the native protein (Weisman *et al*., 2010).

1.13 OTHER PROPERTIES OF SILK

Aside from the aforementioned tensile strength, silk also has a range of other interesting physical and chemical properties. Due to the nanofibrilar structure, energy is dissipated evenly along the silk thread (Poza *et al.*, 2002) (Figure 1.9). This contributes to the strength of the fibre and may be the result of the β-sheet packing, which relies on numerous hydrophobic interactions between those crystalline regions (Hayashi *et al.*, 1999). Silk is regarded as a non-biodegradable material as it takes longer than 60 days to degrade *in vivo* (Altman *et al.*, 2003).

Silk is also able to recover upon simple rest after being unloaded due to the reforming of previously broken hydrogen bonds (Denny 1976; Vehoff *et al.*, 2007). There was also a theory regarding the 'electron lucent domains' along the thread that contended that they acted as stress concentration points, cracking at those precise points when the fibre is stretched. These localised areas would take up and dissipate the energy of the load. This theory was tested by Shao *et al.*, (1999a) and distinct cracks were observed between these domains. These canaliculi are also presumed to contain fluid that ensures hydration and proper silk plasticisation (Work 1984; Vollrath and Edmonds 1989). These longitudinally aligned canaliculi could act as force distributors acting laterally to the plane like a fluid-filled shock absorber or as areas of lubrication 'reducing inter-fibrillar friction' (Shao *et al.*, 1999a). Osaki (1999) found that the elastic limit and breaking strength increased proportionally to the weight of the spider. Twice the spider's weight corresponded to the elastic limit and six times its weight to the breaking strength of drag-lines. This means that a single dragline containing two fibres can more than adequately support

36

the weight of a spider but should one of those break, the remaining line can take the strain. Osaki then suggests that a spider's morphology is limited by the dimensions and physical properties of the silk as opposed to the silk optimising around the spider. Osaki's figures however do not take into account the momentum of a falling spider. Research (Work 1978) found that spiders are able to convert the kinetic energy of descent into strain energy in the dragline by drawing silk. Surprisingly, the strain energy capacity was still insufficient to absorb the potential energy and the spider also dissipates energy by using its own inertia to pull out more drag-line silk from the spinnerets (Brandwood 1985).

A common preconception is that *Nephila* silk is the stiffest. However, work conducted by Madsen *et al*. (1999) demonstrates that the dragline silk of *Euprosthenops* sp. is comparably stronger than that of *Nephila edulis*, *Araneus diadematus* or *Latrodectus mactans*. This is probably due to higher poly-alanine content contributing to its remarkable mechanical properties (Pouchkina-Stantcheva and McQueen-Mason 2004). Such research indicates that a greater range of taxa should be surveyed rather than making assumptions based on limited field observations. The elastic properties of silk are highly anisotropic and alter significantly for both compressional and tensile strain in the presence of water (Schäfer *et al*., 2008).

The orbicularian silk studied thus far has been found to reflect slightly more UV light than white light. However, despite popular citations silks that make up the stabilimenta appear to be highly variable. Based on this knowledge it would seem likely that these stabilimenta act as a predatory defence as opposed to a prey attractant (Zschokke 2002).

37
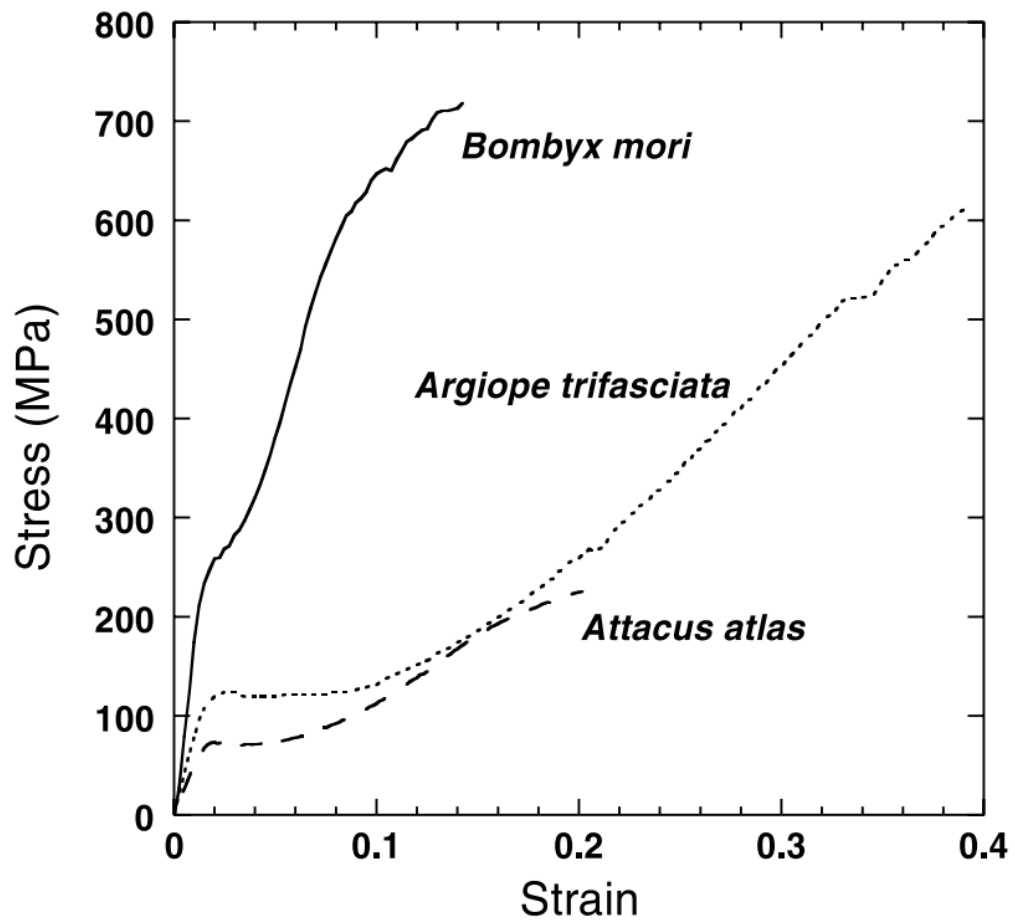
Figure 1.9 Representative stress–strain curves of *A. atlas***,** *B. mori* **and** *A. trifasciata***. Values of** *A. atlas* **are an estimate due the experimental difficulties in measuring accurately the fibre cross-section (Poza** *et al.***, 2002).**

1.14 INDUSTRIAL APPLICATIONS AND PROBLEMS WITH SILK

One problem with silk for engineers is a process termed supercontraction (Work 1981). When silk is immersed in water there is a significant decrease in its overall length (Shao and Vollrath 1999) and substantial forces develop within restrained fibres (Bell *et al.*, 2002). This property is believed to be exploited by spiders to facilitate web tautness

(Guinea *et al*., 2003). When in water, the contraction can be such that the fibre can shrink to half its original length and swell to double the width (Shao *et al*., 1999b). This is reversible and through stretching and subsequent removal of water, the fibre will regain its original properties (Van Beek *et al*., 1999). The discovery was first noticed in major ampullate silk fibres which shrink to half their original length when hydrated, although this phenomenon does not occur in the minor ampullate fibres (Work 1977a/b, 1981). This wetting/drying relaxation-contracting response is cyclic, however upon exposure to > 70% humidity the silk becomes permanently taut. Even after drying, the silk mass has increased ~1% (Blackledge *et al*., 2009b). This cyclic contraction can repeatedly generate work 50 times greater than a comparable mass of human muscle. This finding has indicated new possibilities for designing lightweight and compact actuators and biomimetic silk muscle fibres for robots (Agnarsson *et al*., 2009a/b; Bland 2009; Blackledge *et al*., 2009b). Silk also shows a reduction with fracture strain upon increase in exposure time to acid rain. However, this was only apparent with a pH lower than 4 (Kitagawa and Kitayam 1997).

Radiation also has a detrimental factor on silk's toughness. Beta irradiated silk shows a "reduction in strength, toughness and in maximum extension before failure" (Pogozelski *et al*., 2008). This also applies to UV radiation in that daily sunlight is approximately equal to 1 MJ m^{-2} (as calculated by Kitagawa and Kitayama (1997) from Hayakawa's work (1989)). They predicted this to be the equivalent of the three hours under a UV generator that they subjected it to, resulting in brittleness. They proposed that this is a reason why some spiders rebuild their webs on a daily basis.

It has been hypothesised that supercontraction is a result of particular motifs within the silk protein (Jelinkski *et al*., 1999; Yang *et al*., 2000) whereas man-made fibres undergo a similar process but as a result of thermal influence (Wilson 1974). This was expanded upon (Pérez-Rigueiro *et al.* 2010) with work on Oxyopidae (Lynx spiders). They concluded that supercontraction may have preceded the advent of capture webs and that proline is not involved in this process. This was due to the absence of proline in the Oxyopidae dragline silk, a silk which still has the ability to supercontract.
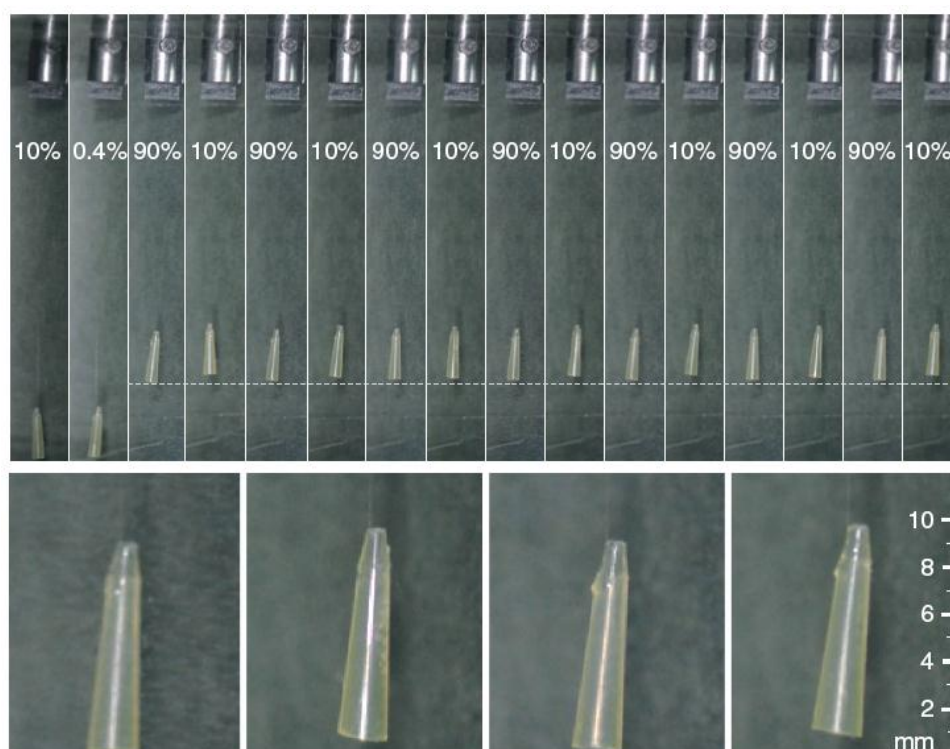
Figure 1.10 Lifting performed by spider dragline silk during repeated cycles of wetting and drying. A plastic weight is suspended from a single dragline silk thread and subjected to repeated changes in humidity. The relative humidity is indicated in each frame. The average displacement during each contraction was 0.65 mm or 1.7% of the thread's total post-supercontraction length. Enlarged views for two cycles are shown at the bottom (Agnarsson 2009).

1.15 USES AND POTENTIAL USES OF SILK

Due to silk's remarkable extensibility, toughness and biodegradability, it has remarkable potential in today's society. With regards to its tensile strength it has the scope to become a replacement for synthetic fibres such as nylon. This would be of use for climbing ropes, bow strings, arresting cables on aircraft carriers and anywhere else that ropes are used. In addition, there are numerous other potential uses. Due to its extreme toughness and resistance to compression (Cunniff *et al*., 1994a/b) silk could be used in bullet-proof vests and high-stress applications such as socks. Prof. Masao Nakagaki at Shinshu University in Japan is credited with introducing a spider silk gene into the silkworm to produce a protein consisting of roughly 10% spider silk. Okamoto, a Japanese company, had planned to have socks from this fibre on the market by 2012. Ironically, the bullet-proof vest (invented by Casimir Zeglen) was originally made from silk and later replaced by Kevlar®.

Another use of silk is as a biomaterial. A biomaterial is "a substance that has been engineered to take a form which, alone or as part of a complex system, is used to direct, by control of interactions with components of living systems, the course of any therapeutic or diagnostic procedure" (Journal of Biomaterials: Williams 2006; Williams 2009). According to research, silk has proven to be a very dynamic biomaterial. Silk is permeable to many drugs as well as oxygen (Lazaris 2002), is resistant to enzymatic degradation and retains many of its mechanical properties when wet, making it an excellent candidate (Minoura *et al*., 1995). It can also be used as a hydrogel (Kim *et al*., 2004), for cell culture (Chiarini *et al*., 2003) and as a collagen substitute for sports

41

injuries and replacement of auto-immune-ravaged skeletal systems present in severe cases of rheumatoid arthritis (Inouye *et al*., 1998).

There has also been increasing work into biofilms. These are usually self-produced matrices of extracellular polymeric substances (EPS) onto which cells adhere. Cells readily adhere to silk fibroin nets and even bridge gaps between individual fibres forming tissue-like materials. Thus far they appear to be highly human-cell compatible and readily support endothelial cells required for vascularisation of new tissue (Unger *et al*., 2004). Electrospun silk fibroin scaffolds also show compatibility with vascular cells. Coupled with the high mechanical properties and slow degradability, there is a lot of scope for fibroin-based tissue-engineered blood vessels (Zhang *et al*., 2008) and similarly, industrial capillary vessels (Lovett *et al*., 2008).

Similar to biofilms are microbeads. An engineered spider silk protein ADF4(C16) mimics the natural dragline silk protein ADF4 of *Araneus diadematus* and can easily be synthesised using *E. coli* as a host. Results have shown that these have a high potential as delivery systems for hydrophobic pharmaceuticals and as storage compounds (Liebmann *et al*., 2008).

Silk can be applied as a bandage as demonstrated by its use for many centuries as a wound dressing by certain tribes. This is due not only to the obvious properties of it being sterile, strong and oxygen permeable but also to the fact that it can be made to be transparent. Research has proven that in wound healing the recombinant spider silk protein based around an arginine-glycine-alanine repeat, pNSR-16 and pNSR-32 was better than the collagen control group ($p < .01$) (Baoyong *et al*., 2010).

With respect to current silk research perhaps the most famous silk-related advance is the use of goats to express the silk genes in their milk (University of British Columbia/UBC). With the herd reared by UBC, it is possible to produce 1.5 mg of silk per goat per day. This was scaled up significantly by Nexia Biotechnologies Inc., though even on the scale used (1500 goats), the amount of silk is sufficient only for medical rather than industrial use.

The ideal solution would be to replace the fibroin gene (*B. mori*) with a spidroin (such as *N. edulis*) using a baculovirus (Yamao 1999). A degree of success has also been achieved using recombinant silk proteins in *E. coli*. Xia *et al*. (2010) have successfully produced a 284.8 kDa recombinant protein from *Nephila clavipes*. The silk, which has an extremely high glycine content (44.9%), was efficiently expressed in a glycyl-tRNA saturated media. It was also found that those proteins of a lower molecular weight yielded inferior fibre properties.

The fundamental flaw with the silks that can be artificially produced at the present time still appears to be that supercontraction cannot be controlled and it is not always a desirable trait.

1.16 ORB-WEBS

In addition to flagelliform spirals and deposits of the aggregate glands, many orb-weaving spiders add bands of conspicuous silken opaque masses to their webs: the stabilimenta. These are hypothesised to be either prey attractants or predatory defences (Zschokke 2002) though neither theory has

been confirmed conclusively. Presumably, if their function is to defend against predation they would be altered in response to predation pressure. Li and Lee (2004) tested this hypothesis by exposing the St Andrew Cross Spider, *Argiope versicolor* to predatory odour cues from the jumping spider *Portia labiata*. They found that *A. versicolor*, not to be confused with *Avicularia versicolor* (a mygalomorph, used later), not only retained the stabilimentum with respect to both its area and fibre incorporation but also decreased the frequency at which these were constructed.

In addition to the stabilimenta, detritus is also often found in varying positions throughout the web. In one experiment, stabilimenta from several members of the genus *Cyclosa* were placed in artificial webs along with spider models then prey-interception and predator interest was recorded. It was found that there was no significant tendency to be attracted towards webs containing stabilimenta. However, the model spiders in the control webs (without detritus) suffered a higher attack rate. This casts doubt over the prey attraction hypothesis but is possibly due to confusion between detritus and the spider's outline, which could aid in reducing predation (Gonzaga and Vasconellos-Neto 2005).

Assuming stabilimenta serve as prey attractants, starved spiders should invest more energy into constructing them than well-fed spiders would have to. Conversely, should they be for predator defence, starved spiders should invest less effort in making them. Blackledge (1998) found that starved spiders reduced stabilimentum area whereas well-fed spiders increased it, thus supporting the predator defence hypothesis.

In respect of prey capture, the spider *Cyclosa conica* (Pallas 1772) was found to capture as much as 150% more prey using decorated webs than undecorated ones even when web diameters and locations were taken into consideration (Tso 1998).

The shape of the stabilimentum also plays a role in prey attraction. Cruciate stabilimenta were significantly more attractive to prey than linear forms whether silken or dummies and dummy forms attracted significantly more prey than vertical and horizontal linear forms (Cheng *et al*., 2010). This would indicate that spiders may take advantage of the 'directional indicators' visible in the ultraviolet end of the spectrum which are also present in flowers and used to guide pollinating insects to the nectaries (Thompson *et al*., 1972). Combining these two findings, Li *et al*. (2004) found that discoid stabilimenta, but not spiral, radial or junction silk reflected UV light and that significantly more *Drosophila* were intercepted in webs with these decorations. Interestingly, stabilimenta are also present in non-orb 'resting webs.' This goes against both previous hypotheses and introduces the possibility that they might additionally be used to aid spider camouflage or to warn off potential web-destroyers such as flying birds (Eberhard 2006).

Likewise, the colour of the web is highly variable. Using *N. clavipes* it was found that the colour could suddenly change up to 38% of the time from white to yellow with no apparent explanation and with no loss of tensile strength (Putthanarata *et al*., 2004).

In reality, it appears to be the case that stabilimenta are present to encompass all these properties and as Seah and Li (2002) found, they are tailored to suit multiple functions throughout the spider's life.

The orientation of the web is also important. Zschokke *et al*. (2006) found that vertical orb-webs retained prey longer than horizontal webs and more active prey escaped more quickly than less active individuals. In addition, the webs with the shortest retention time had owners that were the fastest to capture prey. Sheet-webs however, were the least efficient but this is presumed to be compensated by the lower maintenance required.

1.17 TARANTULA SILK

As explained previously, the Araneae can be divided into two Suborders: Mesothelae and Opisthothelae, the latter of which is further split into the two Infraorders: Mygalomorphae and Araneomorphae. Araneomorphae represent the vast majority of the Araneae and due to the prevalence of both spiders and their remarkable capture devices, the majority of research to date has been conducted on this group. The Mygalomorphae are composed of relatively large sedentary individuals that tend to live in silk-lined burrows (Hedin and Bond 2006) and use silk for far less 'elaborate' uses such as egg sac production and prey capture (Coyle 1986; Shultz 1987) drawing on a smaller number of generalised silks (Garb *et al*., 2007). Prior to the work of Fritz Vollrath who is most credited for his analyses of silk glands, Apstein (1889) noted that there were at least five distinct glands in the aranaeids. He also gives a brief description of a theraphosid, "*Lasiodora Erichsonii*, of the family Avicularii", which according to his findings, only had pyriform glands. With regard to spinnerets, only the Mesothelae have the full complement of eight (Haupt 2003), while the Theraphosidae have just two pairs. The most

complex silk glands appear to be in the Araneomorphae with representatives of the Mygalomorphae having simpler glands and spigot morphologies. *Antrodiaetus unicolor* (Mygalomorphae) females have only one type of silk gland and spigot, the most simple production system described histochemically (Palmer *et al*., 1982).

The first mygalomorph silk sequenced was that of *Euagrus chisoseus* (Gatesy 2001), a species likely chosen in part due to the morphological detail described in Palmer (1985). They found the typical GA, GGX and (A_n) motifs and comment that although *Plectreurys* (Haplogynae) and *Euagrus* (Mygalomorphae) fibroins have internal repeats, the repeats from these basal taxa are unlike any previously described because all these primitive fibroins contain poly-serine tracts. Subsequent research has demonstrated that GGX, GA and poly-A motifs were present in all Araneomorphae and Mygalomorphae species examined comparing both cribellate and ecribellate representatives (Tai *et al*., 2004). Likewise, the non-repetitive N-terminal domain appears ubiquitous throughout spidroin proteins (Garb *et al*., 2010).

In 2007, Garb *et al*., using EST screening and probing of silk gland libraries, found six new mygalomorph spidroins. They comment that the intragenic homogeneity of the mygalomorph spidroins is consistent with the araneomorph examples suggesting that modular architecture and maintenance thereof were present before the infraorder split (240 MYA) and that they have persisted since then (Vollrath and Selden 2007).

Recently, the N-termini of spidroins of a true theraphosid (*Avicularia juruensis*) were sequenced (Bittencourt *et al*., 2010). What was found was a mygalomorph-like (akin to the *Euagrus* spidroin, for example) 'Spidroin 1'

(3154 bp) containing repetitive sequences similar to the tubuliform silk protein 1 (BAE54450) from *N. clavata*. Three orthologous transcripts for the Spidroin 1 gene were found (Spidroin 1A - 2 clones, 1B - 9 clones and 1C - 17 clones) with translation showing high similarity between both their repetitive regions and C-terminal domains with few nucleotide substitutions. Spidroin 1 was the most abundant transcript but in addition, a second silk was also found. This contradicts the previously held view that the theraphosids produced only a single type of silk. This contained the typical $(GA)_n$, poly-A and GPGXX motifs of Spidroin 2 usually present in flagelliform silks with phylogenetic analyses placing this gene within the orbicularian MaSp2 clade.

Although not yet proven, it has been proposed that there is another set of silk glands present on the second abdominal segments of male spiders. These are referred to as the epiandrous (epigastric) glands and it is proposed that they play a role in sperm web production (Figure 1.11). It has been suggested that these may be serially homologous with median spinnerets which, rather than being appendicular in origin, would be modifications of ventral glandular structures (Marples 1967; Lopez 1988; Lopez and Emerit 1988). Author observation has shown a male theraphosid depositing an additional patch of this substance from the inter-spinneret region onto a sperm web (Bull, unpublished). This is usually followed by deposition of the sperm packet onto the patch, onto which the embolus of the palp is positioned to subsequently retrieve it.
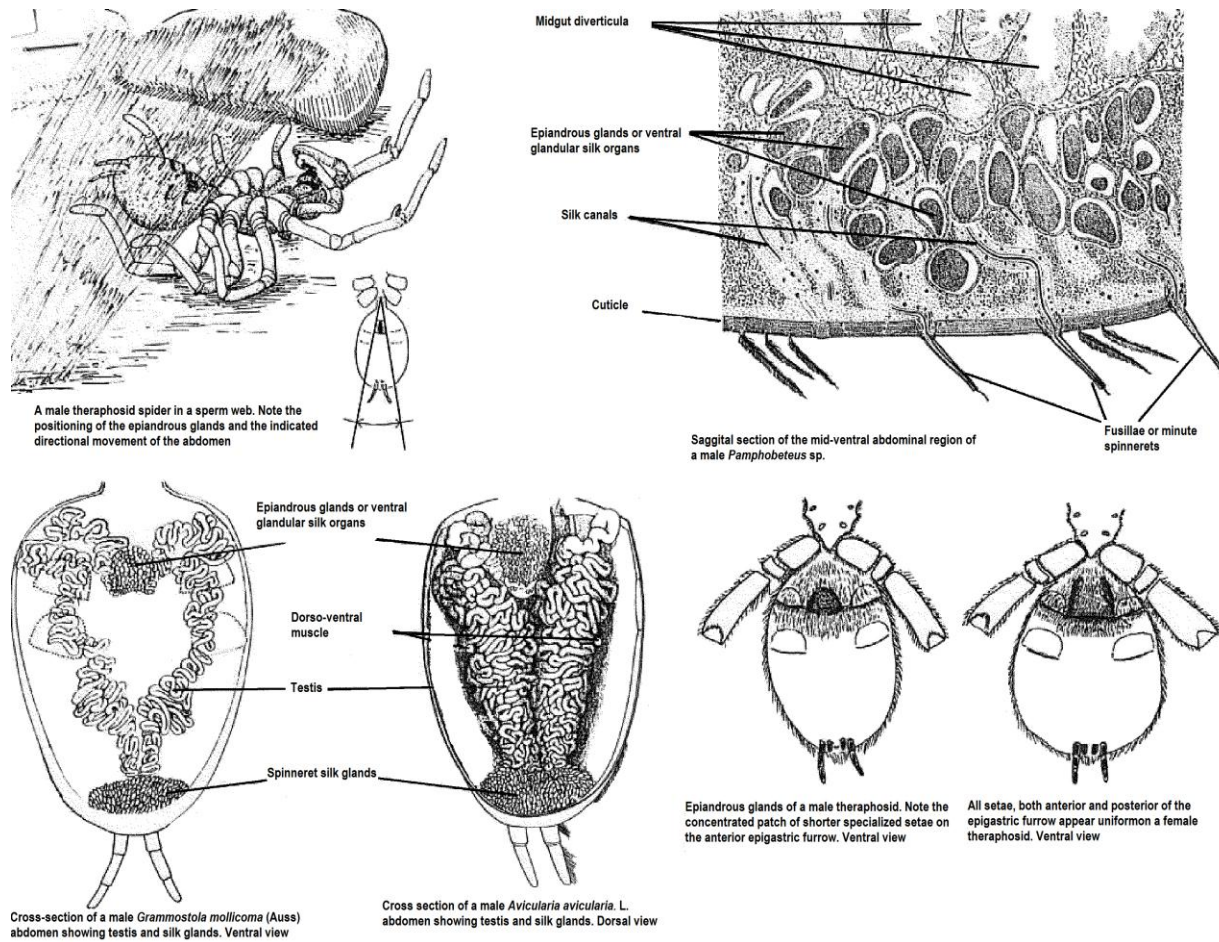
Figure 1.11 Image of the epiandrous fusillade (Melchers 1964).

The following labels appear in the figure:

- Midgut diverticula
- Epiandrous glands or ventral glandular silk organs
- Silk canals
- Cuticle
- Fusillae or minute spinnerets

Saggital section of the mid-ventral abdominal region of a male *Pamphobeteus* sp.

A male theraphosid spider in a sperm web. Note the positioning of the epiandrous glands and the indicated directional movement of the abdomen

- Epiandrous glands or ventral glandular silk organs
- Dorso-ventral muscle
- Testis
- Spinneret silk glands

Cross-section of a male *Grammostola mollicoma* (Auss) abdomen showing testis and silk glands. Ventral view

Cross section of a male *Avicularia avicularia*. L. abdomen showing testis and silk glands. Dorsal view

Epiandrous glands of a male theraphosid. Note the concentrated patch of shorter specialized setae on the anterior epigastric furrow. Ventral view

All setae, both anterior and posterior of the epigastric furrow appear uniform on a female theraphosid. Ventral view

49

1.18 TARSAL SILK

It has been proposed that spiders secrete silk from their feet, although there is evidence that contradicts this theory. In 2003, Arzt *et al*. found that spiders have a dry attachment system relying on van der Waals forces generated by many thousands of spatulate hairs similar to those seen on gecko feet (Autumn 2002; Lee *et al*., 2007). Subsequent observations led to a hypothesis that this may not be the only mechanism involved in adhesion to vertical surfaces. This resulted in Gorb *et al*. (2006) using *Aphonopelma seemanni* (a theraphosid) as a model for tarsal adhesion on glass. According to this paper, the authors induced *A. seemanni* to walk up a vertical glass plane and made the observation that "as it started to slip down the glass silk produced by the tarsal spigots on all four pairs of legs arrested the spider's descent."

However more recently, Pérez-Miles *et al*. (2009) found that upon placing *A. seemanni* on 40 microscope slides both vertical and horizontal, on average half of them received silk threads in addition to urticating hairs. To eliminate the possibility that these were somehow deposited from the posterior spinnerets, they sealed said spinnerets with wax and repeated the experiment. They also made transverse cuts of the tarsi to make a histological assessment. It was found that once the spinnerets were sealed, no silk was deposited and likewise, no glandular structures were apparent within the tarsi. Their conclusion was that these 'spigots' "are very similar in morphology and size to fragments of tarsal thermosensory setae reported for other tarantulas" (Raven 2005) and are not necessarily silk-related. It has been suggested in response to this by Gorb *et al*. (2009) that the parallel tracks of silk are inconsistent with

50

those expected from posterior spinneret transfer. They also make reference to a

fluid appearing at the end of the fibre they also deem unlikely to have

originated from this structure. The suggestion that these spigots are indeed

sensory structures is then defended relating to a suggestion that silks in web-

spinner insects are derived from sensory cells.
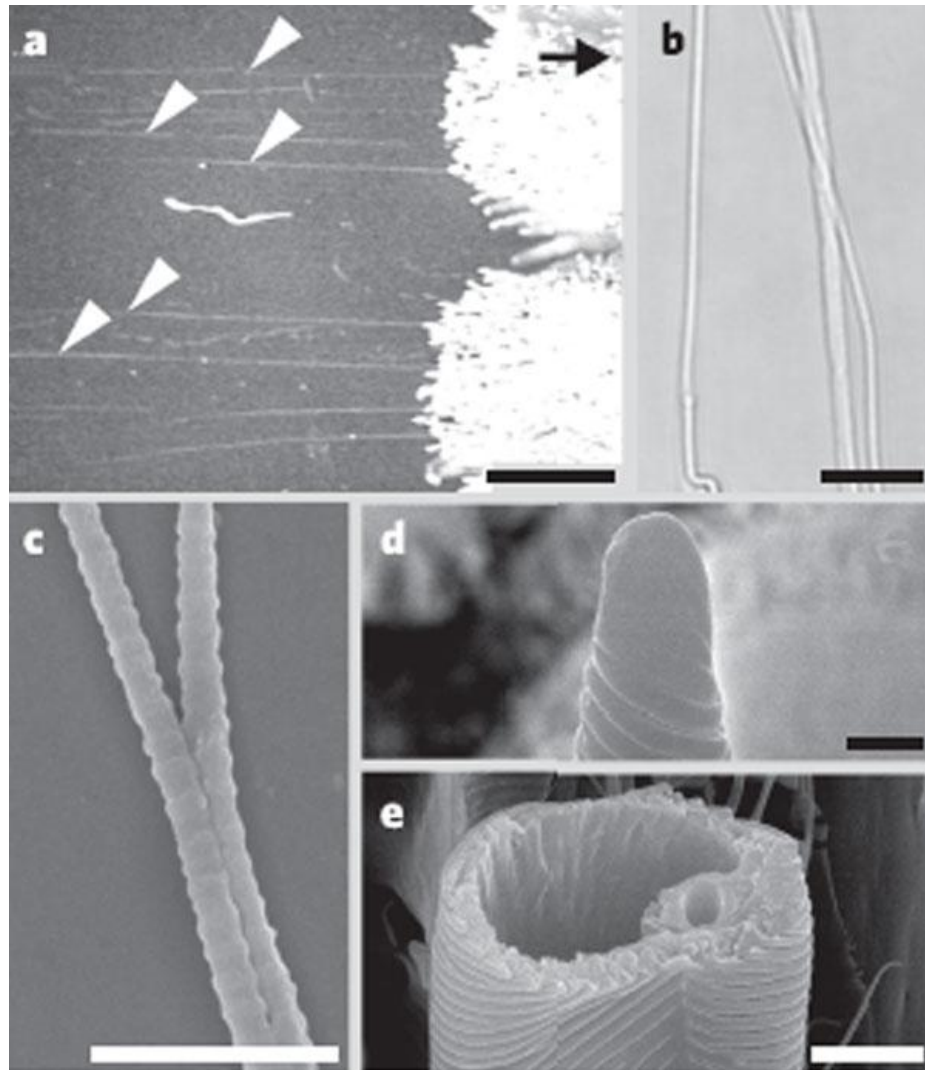
Figure 1.12 a, Fibres left behind (arrowheads) by a spider sliding down a vertical glass surface. Black arrow indicates direction of sliding. The spherical structures are the distal part of the tarsus (scopula), covered with hairs and spigots. b, Traces left by the tarsus of a spider walking on a cover slip. c, Single fibres observed by cryoscanning electron microscopy. d, Tip of a tarsal spigot with the opening obstructed by silk. e, Tarsal spigot broken near the base. Scale bars: a, 500 μm; b, 10 μm; c, 1 μm; d, 2 μm; e, 5 μm (Adapted from Gorb _et al_., 2006).

1.19 AIMS

The aim of this thesis is to further analyse the silks and morphologies of spiders with particular emphasis on the theraphosidae (tarantulas) and *Stegodyphus*. Silks were studied using a combination of transcriptomics, bioinformatics and proteomics. The measurement of particular morphological features (leg hairs), was also investigated for the purpose of sexing individuals before maturity. The efforts resulted in the production of a cDNA library that was aligned with the NCBI BLAST database to identify several silk-related proteins as well as an array of novel proteins with their corresponding database homologues. The aim here was to discover new silk proteins including those which might could be responsible for the protein-folding mechanism. Additional research was associated with the solubilisation and mass spectrometric analysis of tarantula silk proteins and the examination of those peptides' masses determined by MALDI-MS. The initial intent of this was to sequence native spun silks to identify sequence homologies with previously described silks but instead yielded results that suggest that there are more silks produced than previously thought by the tested species as well as revealing a novel method for N-terminal sequencing. Finally, behavioural studies were conducted to assess variation between the sexes and upon closer inspection, provided a method of sexing at earlier instars. The theraphosids are typically sexually monomorphic until later instars permit the close examination of exuviae and/or 'by eye' inspection of morphological features. This technique provides a statistical assessment of the likelihood of correct sex assignment of unknown individuals across a range of taxa than previously described methods.

2 MATERIALS AND METHODS

This chapter provides a complete list of all materials and methods used in the subsequent chapters. Each experimental chapter begins with an overall summary of the methods used to obtain the data therein.

2.1 MATERIALS

2.1.1 CHEMICALS AND REAGENTS

Analytical or Molecular grade chemicals were obtained from Abcam, Alpha labs, Ambion, Anachem, Bioline, BioRad, Fermentas, Fisher, Geneflow, Helena, Invitrogen, Melford, New England Biolabs, PAA, Promega, Qiagen, Roche, Sigma, Stratagene, VWR and Web Scientific.'Grades' are as outlined in the British Pharmacopoeia 2008, meaning they have been tested to the specifications (country dependent) therein. Analytical grades are defined as those with impurities often at 10 - 1000 ppm levels of a substance that would otherwise interfere with the activity of the chemical or organism in analytical or biological procedures by competing/degrading/bonding with the analyte, the specifications of which are indicated by the manufacturer or chemical institutions.

2.2 ENZYMES AND KITS

Restriction endonucleases, DNA-modifying enzymes, DNA polymerases, Deoxynucleotide Solution Sets, DNA ladders and RNA equivalents of the above were obtained from New England Biolabs; TRIzol® Reagent or TRI Reagent® (interchangeable) from Invitrogen or Sigma respectively; pGEM®-T Easy Vector System I and Trypsin Gold Mass Spectrometry Grade from Promega; QIAprep Spin Miniprep Kits and QIAquick Gel Extraction Kits from Qiagen (interchangeable with Wizard® Genomic DNA Purification Kits).

2.3 OLIGONUCLEOTIDE PRIMERS

Vector-specific and gene-specific oligonucleotide primers (GSPs) were designed manually or with the use of primer design programs such as Primer3 (http://frodo.wi.mit.edu/) and custom synthesised by Invitrogen (Life Technologies). Primers were then reconstituted in sterile distilled water (SDW) usually to a stock solution of 1 mg/ml and the concentration verified at 260 nm with a Thermo Scientific NanoDrop™ 1000 Spectrophotometer. The primer concentration was calculated using the formula proposed by Breslauer *et al*. (1986) (see section 2.10.4.5.1). All primers, oligonucleotides, deoxynucleotides and plasmid concentrations were calculated using said method and stored at -20°C.

2.4 CLONING AND EXPRESSION VECTORS

Table 2.1 Cloning and expression vectors

Vector	Backbone	Purpose	Primers	Primer sequence 5' - 3'
pGEM-T Easy	pGEM$^®$ - 5Zf(+)	Cloning	T7 Promoter	TAATACGACTCACTATAGGG
			SP6 Promoter	TATTTAGGTGACACTATAG
pBluescript II SK (+)	pBluescript SK +/-	Cloning	T7 Promoter	TAATACGACTCACTATAGGG
			T3 Promoter	GCAATTAACCCTCACTAAAGGGA

2.5 *E. COLI* STRAINS

Table 2.2 *Escherichia coli* strains

Strain	Bacteria	Purpose
XL10 Gold$^®$ *Ultracompetent*	*Escherichia coli*	Library construction
TOP10	*Escherichia coli*	Cloning from ligations
JM109	*Escherichia coli*	Cloning from ligations
DH5α	*Escherichia coli*	Standard cloning
BL21	*Escherichia coli*	Protein expression

2.6 ARACHNID SPECIES

'Tarantulas' (large arachnids belonging to the family: Theraphosidae) used in this study were purchased from 'pet-trade' dealers or were wild-caught. Those species belonging to the Araneomorphae (also called the Labidognatha), were all wild-caught. The following mygalomorph species from the family Theraphosidae were used: *Grammostola rosea*, *Avicularia avicularia*, *Lasiodora parahybana*, *Psalmopoeus cambridgei* and *Brachypelma smithi* and were chosen based upon their inclusion in previously conducted studies (Petersen *et al*., 2007), Old World (OW)/New World (NW) status, habitat, behaviour and silk production capability. Attempts were made to breed said species to obtain sufficient numbers for behavioural experiments and statistically significant analyses and success was achieved with the following species: *Avicularia avicularia* (NW), *Grammostola rosea* (NW), *Lasiodora parahybana* (NW) and *Psalmopoeus cambridgei* (OW).

2.7 METHODS

2.7.1 SPIDER REARING

2.7.1.1 Introduction to husbandry techniques

Despite the environment from which the species derived, most of the Mygalomorphae can be kept in much the same way in terms of the essentials: temperature, humidity, hygiene, feeding frequency and variable diet. Most will happily survive at a temperature of around 25°C and humidity can be regulated by means of regular spraying for species originating from tropical climates. All spiders were maintained in standard daylight cycles. Precise conditions for each individual species were slightly different, for example a higher relative humidity (80%) was used for *Avicularia avicularia*, *Lasiodora parahybana* and *Psalmopoeus cambridgei* compared to ambient conditions for the others.

2.7.1.2 Substrates

A 50:50 mix of vermiculite and coir (coconut fibre) was the preferred substrate of choice as it had a more natural look, was cheap and retained moisture well. Vermiculite on its own was used to raise spiderlings as it has both a neutral pH and is inert as well as being an unfavourable environment for microorganisms. For tropical species, humidity was maintained by damping of the soil to a constantly moist state. Despite the animals apparently being content with drinking directly from the soil, a water bowl was nevertheless provided in the form of a Petri dish filled with water. For species requiring a dryer habitat, the substrate was the same, but instead the water dish was their sole source of moisture.

2.7.1.3 Hygiene

Due to the nature of obligate burrowers pulling down prey items into the burrow and leaving the husks there, both mould and fungus gnats (family: Sciaridae) can easily become a problem. Uneaten prey items can encourage mould and so were removed, however obligate burrowers are more prone to biting and so were restrained by blocking the burrow to remove the detritus to reduce the chance of an infestation. Likewise, the only way of knowing whether an obligate burrower has died is the onset of an infestation, by which time neighbouring tanks have also acquired the pests. Control would have preferentially been achieved by means of predatory mites such as *Hypoaspis miles*; however these are not compatible with the *Drosophila* cultures sharing the same laboratory. Therefore, all soil was sterilised in an oven prior to use to minimise endogenous pests and was changed on a quarterly basis.

2.7.1.4 Containers and enclosures

Small 60 ml screw-top pots were used to house spiderlings (< 30 mm leg-span), square 80 mm x 80 mm x 120 mm pots were used to house small juveniles (30 - 60 mm leg-span), small faunariums 230 x 155 x 170 mm (Exo-Terra: item no. PT2255) used to house large juveniles (60 mm - 130 mm) and for most spiders greater than this size (< 200 mm) larger tanks (Wilko Fish Tank/Vivarium 11 litre) were used. Ventilation was provided by flaming a large darning needle and puncturing the lids. Larger tanks came with ventilation grilles. For arboreal species, hides/climbing apparatus were provided in the form of strips of expanded polystyrene packing materials cut into strips. Burrowing species were provided with deeper substrate.

2.7.1.5 Diet

Species were fed prey items that were as a general rule of a size comparable with the size of their abdomens. Diet was varied to ensure adequate nutrition was being provided except for spiderlings where there was a prevalent usage of wax-worms (*Galleria*), due to the high fat content, which appeared to encourage faster growth. Standard prey items included black crickets (*Gryllus bimaculatus*) chosen preferentially over domestic crickets (*Acheta domesticus*) due to their greater mass (despite the chirping of the males, which in some instances seemed to discourage predation), hissing cockroach adults (*Gromphadorhina portentosa*) for the largest spiders (> 20 cm leg span) and a colony of the orange spotted cockroach (*Blaptica dubia*) was maintained and utilised during periods where *Gryllus bimaculatus* was not available. A single prey item was offered at a time and removed within 12 hours if not taken. Removal of uneaten prey was important as when a spider moults it is extremely vulnerable and omnivorous insects like crickets have been known to nibble immobilised spiders and kill them. For most species with the exception of the North-American varieties (author's personal observations), spiders will generally take whatever prey they are offered on a weekly basis. Spiders not taking prey within a week of their last feed were presumed to be in pre-moult and were not fed until a suitable period post-moulting.

2.7.1.6 Handling

All tarantulas can bite and many release urticating hairs. When handled, nitrile gloves were worn to minimise urticaria from the New World species. Old World species included in this study were not handled directly due to their aggressive nature. Care sheets and safety information were collated about each species and were distributed as a safety precaution (Appendix 1).

2.7.1.7 Breeding

Spiderlings represent convenient 'packets' of DNA for extraction and may also enable behavioural experiments and growth studies. Breeding was attempted with adult individuals older than three months post final moult during which time both males and females were fed to the point of food rejection and until the male had produced a sperm web. At this stage the male was introduced into the female's tank and retrieved quickly if the female exhibited any aggressive behaviour.

2.8 LEG AUTOTOMY

Spiders have the ability to autotomise (remove) their legs when damaged. This is potentially life-saving for a spider because they will bleed haemolymph indefinitely unless the wound is blocked externally by a scab of crystallised proteins and cell detritus. If a theraphosid spider senses its leg is injured beyond that drying stage, it jerks its coxa (Figure 2.1) up away from the trochanter of the leg causing a pressure tear in the adjoining membrane. A new limb will then regenerate from this site, often to its full size after the next moult, but only if lost within the first quarter of the intermoult period (Bonnet 1930). During this study, it was also found that autotomy was a voluntary process and could not be achieved with anaesthetised spiders. Likewise, not all species are capable of autotomy, for example the Aranaeidae (Foelix 1996; Uetz *et al*., 1996) and widows (Theridiidae). Although autotomy can occur, regeneration does not (Randall 1981). Randall (1981) also demonstrated that total autotomy can occur (all eight legs) using the Pisauridae, although total regeneration (i.e. all legs recovered to the same dimensions as the primary legs) required three successive moults.

This process was utilised effectively in non-anaesthetised spiders for tissue acquisition for DNA extraction with no permanent harm to the spider as per Longhorn *et al*. (2007).
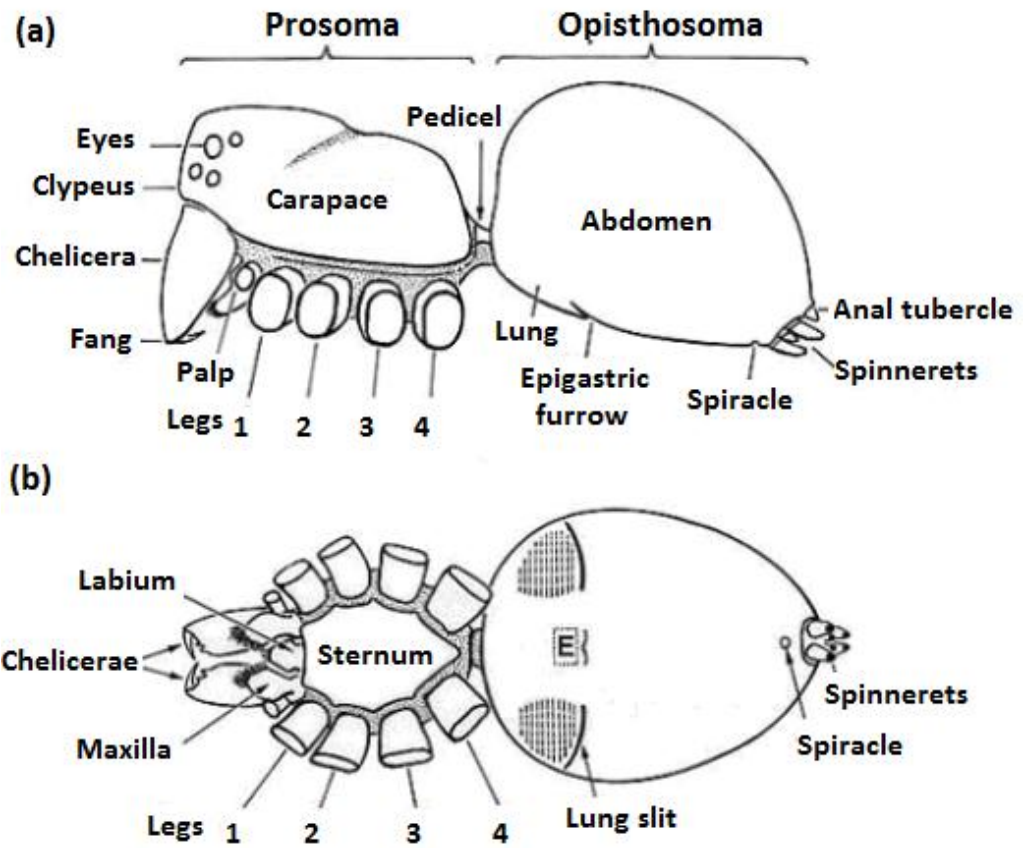
Figure 2.1 Spider anatomy from Foelix (1996). (a) side view, (b) ventral view. E = epigynum (in adult females).

2.9 HARVESTING OF SPIDER MATERIAL

Adult and juvenile spiders were not induced to lethargy by use of gaseous nitrogen or carbon dioxide as performed in previous studies. Instead, according to Work (1976) and in response to Madsen and Vollrath (1999), they were directly subjected to submersion in liquid nitrogen. The frozen tissue was generally used immediately to preserve RNA integrity.

2.10 MOLECULAR BIOLOGY

2.10.1 ISOLATION OF TOTAL RNA

After treating bench-top, instruments and mortar and pestle with Sigma-RNaseZAP [®], whole spiders were dropped into sufficient liquid nitrogen to submerge them. Due to the nature of the hairs coating the entirety of the exoskeleton, it was often necessary to hold it down to prevent floatation and incomplete/localised freezing. At this stage, the legs were snapped off and a femur retained in 70% ethanol for subsequent DNA extraction with an additional leg frozen directly at -20°C. The chelicerae were snapped from the cephalothorax and discarded and the abdomen separated in a similar fashion. Both the cephalothorax and abdomen were treated individually in a similar fashion according to the following:

Tissue was ground to a fine powder under liquid nitrogen and quickly transferred, often with liquid nitrogen still maintained on its surface into TRIzol[®] Reagent (volume according to manufacturer's instructions and approximated weight of the aforementioned tissue). The mixture was quickly

shaken to disperse powder and TRIzol® before it solidified. Agitation was maintained until a homogeneous liquid resulted. Samples were centrifuged for 30 minutes at 4 000 x g at 4°C to remove insoluble material. The clear supernatant was transferred to a new tube. In the case of the cephalothorax, a large amount of fat accumulated on the surface. This was pipetted through and the aqueous layer below was transferred to a new tube. The fat layer was discarded along with the chitinous cellular debris.

The samples were then left to stand for five minutes at room temperature after which 0.2 ml of chloroform were added per ml of supernatant. The sample was then vortexed for 15 seconds and allowed to stand for another 15 minutes at room temperature. The mixture was then spun at 4 000 x g for half an hour. Centrifugation times were scaled up to accommodate the larger volumes used. The upper aqueous phase was removed and an additional 0.2 ml of chloroform per ml of supernatant was added. Again the mixture was vortexed, spun and the upper layer removed as above. This step was found to be necessary to remove additional 'hairs' retained in the solution and to eliminate contaminating coloured compounds (likely proteins). To this solution, 0.5 ml of isopropanol were added and left to stand for 10 minutes at room temperature before being spun at 4 000 x g for an hour at 4°C. The isopropanol was then decanted and one volume of 70% ethanol per ml of TRIzol® was added, vortexed and spun for a further 30 minutes at 4 000 x g at 4°C. The ethanol was then decanted again, re-spun to collect residual ethanol, which was removed by pipetting. The pellet was then left for 10 minutes at room temperature in a fume hood and resuspended in an appropriate volume of distilled de-ionised water (ddH$_2$O) and stored at -80°C. The pellet consistently

had a pinkish tinge or in some species, a blue colour, which appeared to have arisen from the colour of the spider itself. This pigment could not be removed by phenol/chloroform extraction (as per method below) and was thus concluded to be water soluble. It did not appear to interfere with downstream reactions.

2.10.2 ISOLATION OF mRNA USING DYNABEADS® mRNA PROTOCOL

mRNA was isolated from total RNA as per the Dynabeads® protocol. This step was performed to remove contaminating ribosomal RNA, tRNA, miRNA, siRNA, non-poly (A) RNA and pre-processed RNA, resulting in higher transformation efficiency downstream. It also served to remove the pigment.

2.10.3 LIBRARY cDNA SYNTHESIS

2.10.3.1 First strand synthesis

Reverse transcription was performed using oligo (dT) primers to generate first-strand cDNA from the aforementioned RNA templates. This RNA-cDNA heteroduplex was then entered into a second strand synthesis to obtain double-stranded cDNA (dscDNA) using a modified protocol from Simon Dawson (SD, Nottingham University) which was found to give far greater yields than the recommended SuperScript® III protocol provided with this enzyme. mRNA isolated using the Dynabeads® protocol was first heated to 70°C for 10 min to denature templates and then snap-chilled on ice. To a clean,

sterile, RNase-free autoclaved Eppendorf tube the following were added on ice: 8 µl 5 x SuperScript RTase buffer, 8 µl 10 mM methyl dNTPs, 4 µl 100 mM DTT, 2 µl oligo dT primer (2 µg), 1 µl RNase Inhibitor, 6.5 µl H_2O, 10 µl mRNA and 0.5 µl SuperScript$^®$ III Reverse Transcriptase (~100 units). The solution was mixed by gentle pipetting and incubated at 37°C for two hours and the reaction stopped by chilling on ice.

2.10.3.2 Second strand synthesis (SD)

To this 35 µl reaction, on ice, the following were added in this order: 40 µl 10 x DNA Polymerase Buffer (Buffer 1, NEB), 15 µl 100 mM DTT, 12 µl 10 mM dNTPs, 293 µl distilled de-ionised water (ddH_2O), 1 µl RNase H (0.9 units), 1 µl *E. coli* DNA ligase (10 units) and 2 µl DNA Polymerase I (20 units). These were mixed by gentle pipetting and placed on top of an ice bucket (~10°C) for one hour followed by incubation at room temperature (~24°C) for a further hour. To this, 200 µl of phenol (1 g/ml w/v) and 200 µl of chloroform were added, the mixture vortexed and spun for one minute at 14 000 x g. The supernatant was removed, 200 µl of chloroform was added and the mixture vortexed and spun again. The supernatant was removed and the cDNA precipitated by the addition of 40 µl 3 M sodium acetate, 100 mM magnesium acetate and 1 ml absolute ethanol, inverted several times to mix and left overnight at -80°C. The dscDNA was pelleted by centrifugation for an hour at 14 000 x g at 4°C, washed with 70% ethanol, re-spun for 15 minutes at 14 000 x g at 4°C, air dried and resuspended in an appropriate volume of distilled de-ionised water (ddH_2O) and stored at -20°C.

2.10.4 DNA EXTRACTION

2.10.4.1 (a) Isolation of genomic DNA from invertebrate tissues

The femur (mentioned above) was first washed three times in 1 ml ice-cold STE (0.1 M NaCl, 10 mM Tris-HCl (pH 8.0) 1 mM EDTA (pH 8.0)) by submersion and vortexing to remove residual traces of ethanol. The sample was then refrozen for one hour at -80°C or submerged in liquid nitrogen, ground to a fine powder and resuspended in buffer. The resulting homogenate was divided between two microcentrifuge tubes and made up to 1.2 ml with buffer, 15 µl proteinase K added, mixed by inversion and incubated overnight at 50°C. The tubes were removed from the incubator and 5 µl RNase A (10 mg/ml) was added, the samples mixed by inversion and incubated for one hour at 37°C followed by 30 min at 70°C. Tubes were spun for 30 minutes at 14 000 x g at 4°C to remove large chitinous debris and the supernatant transferred to a new microfuge tube. To precipitate residual proteins from solution, an equal volume of 5 M sodium chloride was added, the sample was mixed by inversion and centrifuged in a microcentrifuge for 30 minutes at 14 000 x g. The supernatant was again removed and divided between microfuge tubes to allow 2.5 volumes of 100% ice-cold, absolute ethanol to be added to each tube. Samples were left overnight at -80°C and centrifuged for one hour at 14 000 x g. All but ~20 µl ethanol was carefully removed using a P1000 Gilson pipette so as not to disturb the DNA pellet, 1 ml 70% ethanol was added, vortexed and re-spun for 15 minutes at 14 000 x g. Ethanol was completely removed and residual traces evaporated off in a fume hood for 10 minutes. The pellet was then resuspended in an appropriate volume of distilled de-ionised water (ddH$_2$O).

2.10.4.2 (b) Isolation of plasmid DNA

For small amounts of pDNA or for 'valuable' samples (i.e. cDNA library stocks), the Qiaprep® Spin Miniprep Kit (Qiagen) or the Wizard® Genomic DNA Purification Kits were used. These were favoured when the downstream applications were sequencing or ligations due to the purity required but larger amounts and DNA required for less stringent applications were purified using the alkaline lysis method of Sambrook *et al*. (1989).

2.10.4.3 DNA purification from excised agarose gel slices

The DNA band of interest was excised from the gel using a clean sterile scalpel and placed in a 1.5 ml Eppendorf tube and 100 µl of distilled de-ionised water (ddH$_2$O) was added. The gel slice was gently crushed with a pipette tip, vortexed and frozen either in liquid nitrogen for 10 seconds, a dry ice/ethanol bath for 10 minutes or at -80°C for an hour. The tube was removed and held for 10 seconds to warm it up slightly before being spun at 14 000 x g (at ambient temperature) for 30 minutes. The supernatant was transferred to a new Eppendorf and 200 µl ddH$_2$O was added to the gel slice before being vortexed, refrozen, spun and the supernatant transferred as above. To this pooled supernatant, 100 µl of phenol and 100 µl chloroform: isoamyl alcohol (24:1) were added, the mixture vortexed for 30 seconds, spun again at 14 000 x g and the supernatant removed to a clean fresh Eppendorf. It should be noted here that this can be a three step purification with a phenol step, a phenol: chloroform: isoamyl alcohol step followed by a chloroform step but with careful pipetting this was found to be excessive and instead was reduced to the 100 µl of phenol and 100 µl chloroform: isoamyl alcohol (24:1) treatment as

described above. If the collected volume was > 40 µl, then the solution was made up to 1.5 ml with isopropanol, vortexed and incubated for an hour before being spun at 14 000 x g at 4°C for an hour. If the volume was < 400 µl, then a ml of absolute ethanol was added along with 40 µl 3 M sodium acetate (pH 5.2), vortexed and incubated as above but for three hours. Upon removal, the supernatant was decanted off and discarded and 1 ml 70% ethanol was added, the solution vortexed and centrifuged at 4 000 x g for 15 minutes. The supernatant was once again removed and the pellet left to air dry for 15 minutes before being resuspended in an appropriate volume of water.

2.10.4.4 Agarose gel electrophoresis

An appropriate volume of 6 x loading buffer (10 mM Tris-HCl (pH 7.6), 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol, 60 mM EDTA) was added to the sample and electrophoresed through 0.7% - 3% agarose gels stained in 1 x TAE buffer (40 mM Tris acetate; 1 mM EDTA) according to Sambrook *et al*. (1989) following standard protocols. DNA size ladders were used according to the expected size of the DNA.

2.10.4.5 Polymerase chain reaction (PCR)

2.10.4.5.1 Standard PCR

PCR was performed generally in 20 µl reactions for diagnostic purposes or 50 µl when larger amounts of product were required for downstream applications. It was found that standard *Taq* polymerase could not handle the high-GC content of many templates and so Phusion® High-Fidelity

DNA Polymerase, or Phire® Hot Start DNA Polymerase were used (interchangeably). For some awkward templates (high GC and secondary structure) KOD Hot Start DNA Polymerase was used and for longer amplicons (5 kb+) LongAmp® *Taq* DNA Polymerase was used (the latter two using the manufacturers protocols). For Phusion® and Phire®, reactions were composed of template DNA (gDNA: 100 pg – 100 ng. cDNA: 1 ng – 200 ng. Plasmid DNA: 10 pg – 1 ng), 10 x buffer (supplied with enzyme), dNTPs (0.1 mM each, usually made up to a stock solution incorporating all four: adenine thymine, guanine and cytosine), primers (both forward and reverse, made up to a final concentration of 0.5 µM) and the corresponding DNA Polymerase and made up to 20 or 50 µl with distilled de-ionised water (ddH$_2$O). Usually this was made up as a master mix and pipetted into the PCR tubes minus primers and template where applicable. No observable difference was observed between a master mix containing the polymerase and a hot-started reaction mix into which the polymerase was added after, so a master mix was favoured to reduce wastage due to pipetting errors. The thermocycler (PCR machine) was first preheated to 98°C (both lid and heat block) and the tubes placed inside. The cycle was generally as follows unless otherwise stated: initial denaturation 'hot start' at 98°C for 30 seconds and cycling parameters were as follows: 98°C for 15 s (denaturation), 45 - 72°C for 15 s (annealing) and 72°C for 15 - 30 s per kb (extension) for 35 cycles and a final extension step at 72°C for 10 min. The annealing temperature was calculated as 5°C below the 'nearest neighbour method thermodynamic' of the lowest annealing temperature of both primers (Breslauer *et al.*, 1986; Sugimoto *et al.*, 1996) as given by the following formula (oligocalc):

$$T = \frac{\Delta H - 3.4 \frac{kcal}{{}^{\circ}K\,mole}}{\Delta S + R\ln\left(\frac{1}{[primer]}\right)} + 16.6\log_{10}\left(\left[Na^{+}\right]\right)$$

The nearest neighbour and thermodynamic calculations are done essentially as described by Breslauer *et al*. (1986) (http://www.basic.northwestern.edu/biotools/oligocalc.html) and using the final salt concentrations [Na^{+}] recommended corresponding to each polymerase. Amplicons were presumed to be correct in most cases if estimated sizes were within 20 bp of the expected value when assessed by agarose gel electrophoresis.

2.10.4.5.2 Touchdown PCR

In situations where non-specific products were produced (a smear or multiple bands), touchdown PCR was utilised. Here, the same parameters were used as in a standard PCR but during each successive cycle, the annealing temperature was reduced by approximately 0.2°C. Assuming the primers only had one perfect binding site – i.e. one where every single nucleotide in the primer found a complementary base pair in the target, in theory, at the very highest temperature that the primers bind, they should only bind to a perfectly complementary sequence – the sequence of interest. Here, the number of cycles was increased to 40 and the annealing temperature range was calculated as the temperature two degrees above the 'nearest neighbour thermodynamic theory' to 5°C below it.

2.10.4.5.3 Gradient PCR

Gradient PCR was used as a way of determining the optimal conditions for a PCR as well as a convenience if multiple reactions are being conducted – for example if one set of reactions requires an annealing temperature several degrees higher/lower than another set. The conditions here were the same as a standard PCR except the machine was set to provide a range of annealing temperatures across the PCR machine heat block.

2.10.4.5.4 A-tailing using *Taq* DNA Polymerase

A-tailing is in this instance a by-product of a standard elongation by *Taq* polymerase. A single adenine nucleotide base is added on to a dsDNA at the 5' end of each strand, both as a continuation from said strand's synthesis (standard PCR) or onto clean blunt-ended digested DNA. This overhang can be used to more efficiently ligate to T-vectors (cut plasmids with thymine residues overhanging the 3' ends) such as pGEMT-EASY as described below. 17 µl of cleaned PCR product or recently synthesised, cleaned 2° cDNA is added to a PCR tube along with 2 µl 10 x *Taq* PCR buffer, 0.5 µl dATP (100 uM) and 0.5 µl *Taq* polymerase. This is then incubated at 72° for 20 - 30 minutes.

2.10.5 CLONING OF DNA

2.10.5.1 Restriction digests

Vector and insert were always digested separately with the appropriate restriction enzymes according to manufacturer's protocols and scaled up/down as required. Where a double digest was being performed, the NEB Double

Digest Finder was utilised and incubation times calculated accordingly. This tool simply gave recommendations on enzyme-buffer compatibilities so two digests could be performed in a single step with one buffer rather than two steps with purification.

2.10.5.2 Alkaline phosphatase treatment

Particularly for blunt-ended ligations, an alkaline phosphatase was used to reduce the occurrence of backbone re-ligations. Antarctic phosphatase was the preferred enzyme of choice because it is completely deactivated after 5 minutes at 65°C (NEB literature). Subsequent to a digest, the phosphatase buffer along with the phosphatase was added to the required concentration and incubated accordingly. The phosphatase was then inactivated at 65°C for 10 minutes.

2.10.5.3 Recovery and purification of DNA

Digests were usually separated by agarose gel electrophoresis, the band excised from the gel and the DNA purified using either a phenol/chloroform method (in which the gel slice was frozen and centrifuged for an hour and the resulting supernatant removed and treated as a protein contaminated sample) or via the Promega/Qiagen gel purification kits taking into account the limitations of both where applicable. Here, despite a markedly improved purity of DNA resulting from the kits, the yield was lower compared to the phenol/chloroform method so the purification protocol was chosen according to a yield/purity preference.

2.10.5.4 Ligation of vector and insert

Ligations were usually performed in 10 µl reactions (usually with an insert/vector ratio of 3:1) but on rare occasions a 20 µl reaction utilising the same amount of enzyme and the same 2 x reaction buffer (132 mM Tris-HCl, 20 mM $MgCl_2$, 2 mM dithiothreitol, 2 mM ATP, 15% Polyethylene glycol (PEG 6000) at pH 7.6 and 25°C) yielded better results. Often the digest mixture following dephosphorylation was used directly without any further clean-up with satisfactory results. Ligations were constructed using T4 DNA ligase (20 units/µl) and incubated for 15 minutes at room temperature followed by an overnight incubation at 4°C. Reactions were often heated to 65°C for 15 minutes and snap chilled on ice, which seemed to improve transformation efficiency. Habitually, the majority of PCR products were designed with restriction sites and digested/ligated into pBluescript or A-tailed for T-vector ligation to serve as consistent backups for future downstream applications.

2.10.5.5 Making competent *E. coli* (DH5α)

A seed culture was inoculated and grown overnight in 2.5 ml of LB media in a shaking incubator at 37°C. This overnight culture was used to inoculate a 250 ml culture and again grown (usually for 4 - 6 hours) as above until the A_{600} was between 0.4 and 0.6. The cells were pelleted in sterile Falcon tubes (50 ml) at 4 000 x g at 4°C. The following steps were performed using chilled equipment. The cells were then gently resuspended in 100 ml of ice cold TBF1 (30 mM of potassium acetate, 100 mM potassium chloride, 10 mM calcium chloride, 50 mM manganese chloride, 15% (v/v) glycerol) and incubated for half an hour. The cells were pelleted as before and then

74

resuspended in 10 ml ice cold TBF2 (10 mM MOPS, 75 mM calcium chloride, 10 mM potassium chloride, 15% glycerol) and then incubated on ice for another half hour. Aliquots of this mixture were then transferred to 200 μl Eppendorf tubes, snap frozen in liquid nitrogen and stored at -80°C until use.

2.10.5.6 Transformations using competent *E. coli* cells

Transformations were performed using the above laboratory stocks of competent DH5α *E. coli.* For more delicate and important transformations, either XL10 - Gold® ultracompetent *E. coli* (Stratagene) or *E. coli* strain JM109 (Promega) were used.

An aliquot of competent cells was removed from the -80°C freezer and thawed on ice for 20 minutes. During this time, 1.5 ml Eppendorfs were placed on ice and 2 μl of the ligation mixture or plasmid prep was pipetted into the bottom. Likewise, an aliquot (500 μl per transformation) of Lysogeny Broth (LB) was allowed to warm up to 42°C in a water bath. Once the cells had thawed sufficiently, 50 μl was removed using chilled pipette tips and gently pipetted into each Eppendorf (tube number and competent cell mixture scaled up/down accordingly) and gently mixed with the pipette. These were left for 30 - 60 minutes after which they were heat-shocked at 42°C for 60 s and then returned to ice for 2 min. The warmed LB was then removed from the water bath and 500 μl aliquots were pipetted into each transformation mixture. These were then incubated at 37°C in a shaking incubator for 30 - 60 minutes. For non-insert control plasmids, 100 μl of the transformation mixture was plated out onto appropriate pre-warmed (37°C) antibiotic LB agar plates.

2.10.5.7 Identification of colonies that contained inserts

For identification of colonies with plasmids containing inserts within the LacZα MCS, blue–white screening was performed. Here, 100 µl of the transformation mixture was plated along with 40 µl 2% X-gal (5-bromo-4-chloro-indolyl-β-D-galactopyranoside) and 7 µl 20% IPTG (isopropyl ß-D-thiogalactoside) and incubated overnight at 37°C for 12 − 16 h. Plates containing transformants were then stored at 4°C.

2.10.6 AUTOMATED DNA SEQUENCING

2.10.6.1 (a) Sequencing reactions

As there was a potentially high volume of samples to be sent off for sequencing, Macrogen Inc. (Seoul, Korea; later Amsterdam, the Netherlands) (ABI3730XL machine) was chosen as the preferred sequencing service due to the reduced costs. The subsequent quality of reads and turnaround time resulted in the utilisation of a more local company, GeneService (now Source BioScience LifeSciences) (ABI 3730XL machine). Sequences came back as a compressed Zip file containing both the AB.1 file (Figure 4.1) (Applied Biosystem's Sequencing Analysis software) and a FASTA format notepad file.

2.10.6.2 (b) Sequence data analysis

VecScreen (www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html) alignments were performed using ClustalW v1.4 (Higgins *et al*., 1996), assembly and sequence editing were performed with the CAP3 Sequence Assembly Program (Huang and Madan 1999) http://pbil.univ-lyon1.fr/cap3.php, DNA and amino acid manipulations were performed with the ExPASy translate tool (http://web.expasy.org/translate/) and BLAST (Altschul *et al*., 1997) was used for similarity searches.

2.10.6.3 DNA Minipreps from *E. coli* cultures

Sterilins (15 ml vials) were used to grow up 5 ml cultures overnight in a shaking incubator at 37°C. The cultures were then spun for 5 minutes at 14 000 x g and resuspended in 100 µl Solution 1 (10 mM Tris pH 8, 10 mM EDTA). The mixture was then transferred to a clean sterile Eppendorf tube and the cells lysed by adding 200 µl Solution 2 (0.2 M NaOH, 1% SDS). The tube was gently inverted several times and left for three minutes. The proteins were precipitated by adding 150 µl Solution 3 (7.5 M ammonium acetate) and inverted and left for three minutes as above before being spun at 14 000 x g for 10 minutes. The supernatant was then transferred to a new clean Eppendorf containing 900 µl absolute ethanol and left for half an hour before being spun at 14 000 x g for 30 minutes. The supernatant was then decanted and the pellet washed with 70% ethanol before being spun again for 10 minutes at 14 000 x g. The supernatant was removed and the last drops of ethanol were blotted away with a KimWipe® and the pellet was left for 15 minutes at rt. The pellet was then resuspended in an appropriate volume of de-ionised water (ddH$_2$O).

2.10.6.4 DNA Maxipreps from *E. coli* cultures

This protocol was adapted to suit a 50 ml Falcon tube but could also be scaled up to a 250 ml centrifuge vial. A 1 ml culture picked from a plate was first grown up overnight with the appropriate selective antibiotic (determined by the plasmid used). This was transferred to a 300 ml flask of Lysogeny Broth and left to grow in a shaking incubator at 37°C for four hours. This was then spun down in a 50 ml Falcon tube in 50 ml increments (removing the media each time) until the LB had been completely removed. To the large remaining pellet, 10 ml solution 1 (as above) was added and the pellet was resuspended by vigorous vortexing. Freshly-made lysis solution, 10 ml (as above) was added and mixed by gentle inversion before an additional 10 µl 20 mg/ml RNase A was added followed once more with gentle inversion. Proteins were then precipitated using 15 ml 2.5 M potassium acetate, 2 M acetic acid before being spun for an hour at 14 000 x g. Centrifugation for 15 minutes was usually sufficient but required filtering through a Whatman 3MM. The longer centrifugation was used because this was found to remove all protein, usually leaving a large pellet and a white, flocculent surface layer. The supernatant could then be divided between two new Falcon tubes, separating these two protein layers by gentle decanting. To these two pellets, 0.5 volumes isopropanol was added and incubated for half an hour on ice before being spun at 14 000 x g for an hour, or for a higher yield, absolute ethanol/sodium acetate was used as per the DNA precipitation protocol below (2.10.6.5). The pellet was then resuspended in 1.5 ml de-ionised water (ddH$_2$O) after an hour of air drying at room temperature. To this solution, 5 µl 20 mg/ml RNase A was again added and incubated for an hour at 37°C. The solution was then split

between two Eppendorfs and to each, 0.5 volumes phenol/chloroform was added, the mixture vortexed, spun for one minute and the supernatant transferred to a clean Eppendorf in each case. This step was repeated until there was no white precipitate at the interphase. The DNA was then precipitated/resuspended once again as per the DNA precipitation protocol.

2.10.6.5 Precipitation of nucleic acids using alcohols

Choice of precipitation method was determined by the desired yield and volume of the initial solution. Nucleic acids were precipitated from aqueous solution by the addition of 2.5 volumes of ethanol or 0.5 volumes of isopropanol and 0.1 volumes 3 M sodium acetate (pH 5.2). Solutions were mixed and incubated at -80°C for 30 minutes followed by centrifugation at 14 000 x g for an hour, decanting off the supernatant and a wash with 70% ethanol. The supernatant was removed again and the pellet left to air-dry for 15 minutes. For higher yields, ethanol was chosen, the incubation was left overnight and the centrifugation step was an hour. For approximately a 70% yield (shown by comparative studies), isopropanol was used, the incubation was shortened to 20 minutes, centrifugation reduced to 15 minutes and the pellet heated quickly to 65°C for 3 minutes. The pellet was then dissolved in an appropriate volume of water.

2.10.6.6 DNA concentration

DNA concentration was estimated by measuring absorbance at 260 nm (A_{260}), adjusting for turbidity (measured by A_{320}), taking into account the dilution factor, where A_{260} of 1.0 = 50 µg/ml pure DNA. [2]

Therefore:

Concentration (µg/ml) = (A_{260} reading − A_{320} reading) × dilution factor × 50 µg/ml.

2.10.6.7 Total yield

Total yield was obtained by:

DNA yield (µg) = DNA concentration × total sample volume (ml).

2.10.6.8 DNA purity

Purity was estimated from the A_{260}/A_{280} ratio. Pure DNA was considered to have a $A_{260/280}$ of ~1.8 and pure RNA a $A_{260/280}$ of ~2. The ratio is calculated thus after correcting for turbidity (absorbance at 320 nm).

DNA Purity (A_{260}/A_{280}) = (A_{260} reading − A_{320} reading) ÷ (A_{280} reading − A_{320} reading)

Note: The spectrophotometer used was considered to be accurate when within the range of 10 - 4000 ng/µl.

---

[2] Adapted from: http://www.promega.com/enotes/faqspeak/fq0059.htm

3 NATIVE THERAPHOSID SILK PROTEIN ANALYSIS

3.1 INTRODUCTION

The family Theraphosidae comprises a diverse group of large terrestrial spiders with over 100 genera and now probably close to 1000 identified species. The genus *Avicularia* includes over 50 species, all originating from South America and categorised by having distinctive pink foot-pads from whence the common name 'Pink-Toe' is derived. In the present study, a proteomic and biological characterisation of the silks of *A. avicularia* (L) (Araneae: Theraphosidae) was performed to gain a deeper insight into the true nature of these theraphosid silks and gain knowledge about their composition and probable accessory proteins produced *in situ*. Spider silks possess physical properties of tensile strength (Griffiths and Salanitri 1980), elasticity (Liu *et al*., 2008) and toughness (Vollrath and Knight 2001) rivalling most known natural fibres. One group of fibres known as major ampullate or dragline silks are transcribed from genes anything up to 40 Kb in size with the translated protein reaching almost 0.5 MDa.

The protein is stored as a liquid dope that is extruded into the posterior spinnerets (Work *et al*., 1977a/b) and hardens through, amongst others, a combination of pressure gradients, pH gradients (Dicko *et al*., 2004), dehydration and mechanical stresses (Rising *et al*., 2011) before emerging as the complete composite fibre. All known silks can be divided into three domains: a highly conserved N-terminal domain, a highly repetitive and (typically) hydrophobic middle domain, usually comprising over 80% of the

81

fibre, and a highly conserved but unique C-terminal domain. Due to the extreme length of silk mRNAs, a large number of C-terminal domains have been characterised along with the repetitive domains but less is known about N-terminal domains, which are thought to aid in the pH aspect of fibril formation (Hedhammer *et al*., 2008; Sponner *et al*., 2005b). Using *A. juruensis* Bittencourt *et al*. (2010) identified two distinct mygalomorph spidroins. Following their work, but employing a sister species (*A. avicularia*), proteomic analyses were conducted.

After trypsin digestion, the peptides were subjected to LC-ESI-tandem mass spectrometry on a Q-TOFII mass spectrometer. Using multiple sequence alignments and contig assembly, the majority of the N-terminus has been reconstructed providing confirmation of the family from whence the theraphosid silks derived. These data appear to be the first use of ESI-tandem mass spectrometry (MS) to characterise novel native mygalomorph silk proteins. The two spidroins identified by Bittencourt *et al*. (2010) were characterised to the relative ratios found through their cloning procedures. However, an interesting finding was that the ESI-MS seemed to select for the N-terminal domains from which the majority of contigs seem to have been derived. This is somewhat surprising as the N-terminal domain comprises < 1% of the total transcript length for other silks (Ayoub et al., 2007). Furthermore, this sequence does not appear to have originated from a respective/predicted Spidroin 1 or Spidroin 2 and clearly falls within the N-terminal identity of pyriform silks.

3.2 MATERIALS AND METHODS

3.2.1 RAW SILK COLLECTION

An adult female Pink-Toe Tarantula (*Avicularia avicularia*) was transferred to an ethanol-cleaned, dust-free ventilated box with approximate dimensions 20 cm x 10 cm x 10 cm after being power-fed on a mixed diet of black crickets (*Gryllus bimaculatus*), orange-spotted/Guyana cockroaches (*Blaptica dubia*) and giant mealworms (*Tenebrio molitor*) for the month prior to isolation. These sterile conditions were uncomfortable for the spider and encouraged deposition of fresh, clean silk. Silk was harvested approximately two weeks later using a sterile drinking straw, which took advantage of the electrostatic attraction between the silk and the plastic, allowing easier collection. Soiled regions of the silk were cut around with a border of approximately 1 cm and discarded. The native silk was stored in a clean, sterile Eppendorf tube and stored at room temperature until use (~one week).

3.2.2 DENATURATING PROTEIN SAMPLES

The raw silk was first solubilised using 10 M lithium thiocyanate and incubated at 40°C over 3 days with occasional vortexing. It was observed at this stage that even under these extremely chaotropic conditions, the inner 'core' of the silk appeared to remain undissolved. The silk fibre precipitates out when dialysed or if the solution is transferred to standard buffers (confirming observations by Xu and Lewis 1990). Two aliquots were made at this point. Samples for SDS polyacrylamide gel electrophoresis were boiled for 2 minutes

in 2 x sample buffer (100 mM Tris pH 6.8, 20% glycerol, 8% SDS, 2% β-mercaptoethanol, 2 mM EDTA, 0.2% bromophenol blue, 0.2% xylene cyanol) (Laemmli 1970) to completely denature and unfold the polypeptide chain. After boiling, the sample was allowed to cool to 60°C and 1 μl of a 20% (w/v) aqueous iodoacetamide was added to the mixture and incubated for a further 30 minutes at room temperature. This step, although not always necessary produced sharper bands and removed artefacts and lines across the gel.

3.2.3 SODIUM DODECYL SULPHATE POLYACRYLAMIDE GEL ELECTROPHORESIS

Glass plates and combs were first cleaned with ethanol before setting up the cassette. The gel was prepared (Appendix 2) without both the APS and TEMED, which were added prior to pouring the gel after gentle inversion. Degassing of the TEMED was usually unnecessary. Due to silk's consistently high molecular weight, a 6% gel was chosen. The solution was poured until the level reached ~1 cm below the bottom of the comb that will be used within the cassette. A small amount of alcohol (usually propanol) was then poured on and was used to determine when the gel had set, at which point the propanol was poured away and the remainder blotted off with KimWipes® or blotting paper.

The stacking gel (Appendix 2) was then prepared, again without the APS and TEMED, which were added immediately prior to pouring into the cassette as per the resolving gel and the comb inserted quickly thereafter. In this instance, a 4% stacking gel was used. The gel was then left overnight at 22°C to set completely in a sealed plastic bag so as to retain moisture. The gel can be stored in cling film for several days like this if need be. The comb was

then carefully removed and residual gel fragments removed with a razor blade and air bubbles/excess liquid with a syringe. The gel was then placed in the gel tank that was filled with SDS buffer (250 mM Tris HCl, 1.92 M glycine, 1% SDS, pH 8.3) and again, any obstructions in the wells removed. Samples were boiled (see above) in 2 x loading buffer and 20 µl loaded along with a suitable protein standard ladder. The power pack was connected and 80 mA was passed through the gel until the sample had passed beyond the stacking gel (usually half an hour) at which point the power was increased to ~150 mA. The power was turned off and the gel removed once the dye front had reached the bottom of the gel. For expectedly larger protein bands (silk), the gel was often left to run for a further hour, by which time, the band had rarely travelled more than halfway down the gel.

3.2.4 COOMASSIE BLUE STAINING PROTOCOL

The gel was washed in ddH_2O three times for 5 minutes on a rocker to remove the SDS. The gel was then fixed in a solution comprising 40% (v/v) methanol, 10% (v/v) acetic acid and 50% H_2O (v/v) for one hour and rinsed again twice more in ddH_2O as before. Enough Coomassie Blue staining solution (10% acetic acid, 90% ddH_2O, 0.0006% Coomassie Blue) was added to just cover the gel, which was then gently shaken for an hour. Washing was performed twice more as above in ddH_2O and then the gel destained in a solution of 40% (v/v) methanol, 10% (v/v) acetic acid and 50% ddH_2O (v/v), three times for half an hour. Two more wash steps were performed with ddH_2O before images were acquired using a scanner with the gel wrapped in cling film. The gel was usually stored in a small sandwich bag in 1% acetic acid.

3.2.5   SILVER STAINING PROTOCOL

For silver staining, all solutions were pre-filtered to remove protein contaminants, gloves were worn and exposed steps carried out in a laminar flow cabinet.

As in the Coomassie Blue protocol, the gel was first washed twice in ddH$_2$O for five minutes. The gel was then fixed in a 40% (v/v) methanol, 10% (v/v) acetic acid solution twice for half an hour. Sensitising was carried out in a 30% (v/v) methanol, 0.2% (w/v) sodium thiosulphate pentahydrate, 0.5 M sodium acetate solution twice for 30 minutes and washed three times more in ddH$_2$O as above. The gel was then impregnated with 0.1% silver nitrate (w/v) for 20 minutes and washed twice for one minute in ddH$_2$O. Using pre-chilled developer (2.5% w/v sodium carbonate, 0.1% v/v (37% w/v) formaldehyde) the bands were visualised and the process stopped 30 - 60 seconds before optimum staining had been achieved through decanting the developer and immersing in 1% (w/v) acetic acid.

3.2.6   EXCISION AND DESTAINING OF BANDS FROM COOMASSIE BLUE GEL

Protein samples were processed using the ProteomeWorks MassPREP robotic liquid handling station (Waters, Ltd). Using a sterile scalpel, bands were excised from the gel with as little gel-border as possible so as to minimise artefacts on mass spec and diced into cubes (~1 mm^3). Samples were first incubated in destaining solution (50 mM ammonium bicarbonate, 50% acetonitrile) three times for 10 minutes at room temperature and then dehydrated in 50 µl of acetonitrile for 5 minutes. Residual acetonitrile was

allowed to evaporate off and 50 µl of reducing solution (10 mM dithiothreitol, 100 mM ammonium bicarbonate) was added and incubated for a further 30 minutes. This was followed by 30 minutes in a solution of 55 mM iodoacetamide and 100 mM ammonium bicarbonate, a wash for 10 minutes in 50 µl of 100 mM ammonium bicarbonate and two more washes in 50 µl of acetonitrile for 5 minutes, after which the solvent was evaporated off.

3.2.7 DESTAINING OF BANDS FROM A SILVER STAINED GEL

Bands were excised and diced into cubes (~1 mm^3) as with the Coomassie Blue bands and 50 - 100 µl destaining solution (1:1 mixture of 30 mM potassium ferricyanide and 100 mM sodium thiosulphate) was added to each sample and incubated until the colour was completely removed. The solution was removed and gel pieces washed with 50 µl 200 mM ammonium bicarbonate for 20 min (scaled up accordingly if the gel slice is larger). The supernatant was removed and washed twice in 150 µl of HPLC water for 15 min, after which the water was removed (Gharahdaghi 1999).

3.2.8 TRYPSIN DIGESTS

Trypsin digests and subsequent tandem MS were performed courtesy of Susan Liddell (University of Nottingham). Gel microtitre plates were pre-cooled to 6°C for 10 minutes and then 25 µl Trypsin Gold, Mass Spectrometry Grade (Promega), diluted to a concentration of 10 ng/µl with trypsin digestion buffer (50 mM ammonium bicarbonate) was added to each sample and incubated for 20 minutes. The plate was maintained at 6°C for a further 20

minutes followed by four hours at 40°C and then stored at 4°C until MS analysis.

Digested samples were passed through a P10 C18 zip-tip reverse phase (Millipore) to concentrate and desalt the peptides and eluted into 10 µl of 50% methanol, 0.1% formic acid. They were then loaded into borosilicate nanospray needles (Waters, Ltd) and inserted into a Q-TOFII mass spectrometer with a nanoflow ESI (electrospray ionisation) source (Waters, Ltd). This was operated at a capillary voltage of 900 – 1200 V in positive ion mode, using argon as the collision gas.

3.2.9   TANDEM MS – MANUAL ACQUISITIONS

Manual acquisitions were performed as in Rodriguez-Martin *et al.* (2010) and carried out thus:

Survey scans were performed with the sampling cone set at 45 – 50 V and data typically acquired from 400 – 2000 *m/z* over a scan time of 2.4 seconds. Peptide mass spectra results shown were from typically 5 – 15 minutes of data acquisition. Selection of the candidate multiply-charged peptide ions of the survey spectra was performed visually. Tandem MS fragmentation spectra were selected usually from 50 to 1600 *m/z* and deconvoluted into singly charged, mono-isotopic masses with the assistance of MaxENT3 maximum entropy software (Waters, Ltd), while manual interpretation of the peptide sequence was conducted using the PepSeq software, a component of the MassLynx package (Waters, Ltd).

BLASTp (http://www.ncbi.nih.gov) using the "short, nearly exact matches" parameters, was used to compare *de novo* sequences across the

database. Analyses on the resulting peptides were carried out by ESI-MS/MS after online separation on a PepMap C18 reverse phase, 75 μm i.d., 15 cm column (LC Packings). This was performed on a CapLC system attached to a Q-TOF2 mass spectrometer equipped with a NanoLockSpray source (Waters, Ltd) utilising MassLynx Version 4.0 acquisition software.

Automated data-dependent switching between the MS and MS/MS scanning based upon ion intensity, mass and charge state were used to acquire the tandem MS data. Here, a method was created in the MassLynx 4.0 software in which charge state recognition was used to select doubly, triply and quadruply charged precursor peptide ions for fragmentation. For tandem MS acquisition up to four precursor masses at a time were chosen. A collision energy parameter was selected based on charge and mass of each precursor and varied from 15 to 55 eV. ProteinLynxGlobalServer version 2.0 (Waters, Ltd) was used to analyse the raw MS data into peak list (pkl) files and searched against all entries in the Swissprot and/or NCBInr databases (as of 2010) using MASCOT MS/MS ions search tool (http://www.matrixscience.com: Perkins *et al.*, 1999). Cysteine carbamidomethylation and methionine oxidation were set as variable modifications. A single missed cleavage by trypsin was accepted. Other than file type (Micromass pkl) and instrument type (ESI-QUAD-TOF), all other search values were left as their defaults. Only protein identifications with probability-based MOWSE scores above a biologically statistical threshold of $p < .05$ were accepted.

3.2.10 TANDEM MS – DATA DEPENDENT ACQUISITIONS (DDA)

DDAs were performed as in Rodriguez-Martin *et al*. (2010) and carried out thus:

Data directed analysis (DDA^{TM}) was used to acquire tandem MS data using an automated data-dependent switching between the MS and MS/MS scanning based upon ion intensity, mass and charge state. The data were searched against the public databases using MS/MSIONS search on the MASCOT web site (http://www.matrixscience.com/search_form_select.html) with standard default settings. Using the MassLynx 4.0 software, a method was created in which charge state recognition was used to select doubly, triply and quadruply charged precursor peptide ions for fragmentation. A precursor mass was chosen one at a time for tandem MS acquisition. Using charge and mass of each precursor which varied from 15 to 55 eV, the collision energy was automatically selected. Uninterpreted MS data were analysed as above.

3.2.11 *DE NOVO* SEQUENCE INTERPRETATION OF PEPTIDES

De novo sequences were interpreted using the PepSeq tool of the MassLynx™ 4.0 software package (Waters, Ltd). Tandem MS spectra from the DDA LC-tandemMS runs were sorted into singly charged, mono-isotopic masses using MaxEnt 3 maximum entropy software (Waters, Ltd). The resulting spectra were examined in the PepSeq window using a combination of automated and manual direction to elucidate each peptide sequence. *De novo* peptide sequences were compared to databases using BLASTp (at: http://www.ncbi.nih.gov) using the parameters: "short, nearly exact matches."

3.2.12 *DE NOVO* INTERPRETATION AND CAVEATS

Results obtained by ESI tandem MS with low energy collision induced dissociation (CID), are open to interpretation even with the aid of specialised software used for analyses. Even if the results are clear and the spectrum is clean, there can be problems in distinguishing between certain residues. Here it is essential to call upon additional resources and techniques to verify the resultant spectra and amino acid sequences.

In various literature where two residues are together in brackets (XY), the order is interchangeable. Likewise with two residues separated by a forward slash X/Y and similarly with two residues underlined XY.

Leucine (L) and isoleucine (I) are isobaric (they are constitutional, i.e. structural isomers of each other) and therefore have a mono-isotopic mass value of 113.08406, and consequently cannot be differentiated using this type of MS. This means that wherever I/L are present in a *de novo* sequence, the potential possibilities branch out by a factor of two at this point.

Some residues only differ by a fraction of their molecular weight such as glutamine (Q) and lysine (K) which differ by only 0.036 u. Likewise, phenylalanine (F) and oxidised methionine (M^{SO}) only differ by 0.033 u. So although there is a slightly higher likelihood of accuracy in calling a particular residue, care should be taken in calling one definitively over the other. Further details and explanations of these complications are given in Appendix 3.

Dipeptide fragments also yield potential complications as they increase the chance of molecular weight isomerisms. For example, with two consecutive glycine-glycine residues or likewise with a glycing-alanine fragment, cleavage rarely yields an abundant ion.

When ions of the same molecular weight are recorded, again, differentiation is virtually impossible. For example Gly-Gly is isomeric with asparagine (Asn) (N), Gly-Ala is isomeric with glutamine (Gln) (Q) (as above) and isobaric with lysine (Lys) (K) (same molecular formula, different arrangement of atoms). Therefore, again, where a single Asn is called, it could either be Gln/Lys, Gly-Gly or Gly-Ala, respectively.

The above issue is extremely important in this study as spidroins are known to have extremely abundant amounts of glycine, alanine and relatively frequent occurrences of leucine and isoleucine.

To obtain reliable spectra and therefore reliable interpretation of the data, fragmentation at every peptide bond is required, which is not always achieved and is particularly infrequent around proline residues. Likewise the two terminal amino acids are rarely cleaved and so at best, only a combined molecular ion can be obtained – a significant issue if they happen to be those isomeric/isobaric species mentioned above.

3.3 RESULTS

3.3.1 MASS SPECTROMETRIC SEQUENCE ANALYSIS OF THE TRYPTIC DIGESTED PROTEIN

To solubilise the deposited silk fibres, saturated (~37 M) solutions of lithium thiocyanate (LiSCN) were prepared as reported by Chen *et al*. (2009). Attempts were made to solubilise said proteins using other means (100% DMSO, 9 M LiBr, 10 M GdnHCl) and saturated urea solutions (~10 M) but with limited success or apparent deterioration of the silk proteins. This treatment however was only partially successful with the outer fibres dissolving readily but the inner core fibres (unknown constitution) remaining out of solution despite heating to 90°C and constant agitation. SDS-PAGE analysis, Coomassie Blue and silver staining of the solubilised avicularian silk revealed at least two distinct bands (Figure 3.1), both with an apparent molecular mass ~300 kDa. Coomassie Blue staining was insufficient to visualise the proteins, which solubilised poorly in all attempts so silver staining was utilised (Figure 3.2). Previous studies on another species within this genus have shown that there are two distinct proteins; Spidroin 1 with (at least) three putatively orthologous transcripts and Spidroin 2, a structurally unique silk, highly under-expressed relative to Spidroin 1 (Bittencourt *et al*., 2010). Initially, the MALDI-MS analysis for the silk analysed in the present study was unsuccessful on these separate bands (Figure 3.1) so further SDS-PAGE analyses were conducted. Gel resolutions were not sufficient, presenting as a smear (Figure 3.3) and therefore the exact size of each silk likely corresponding to those found in Figure 3.1 is yet to be established.
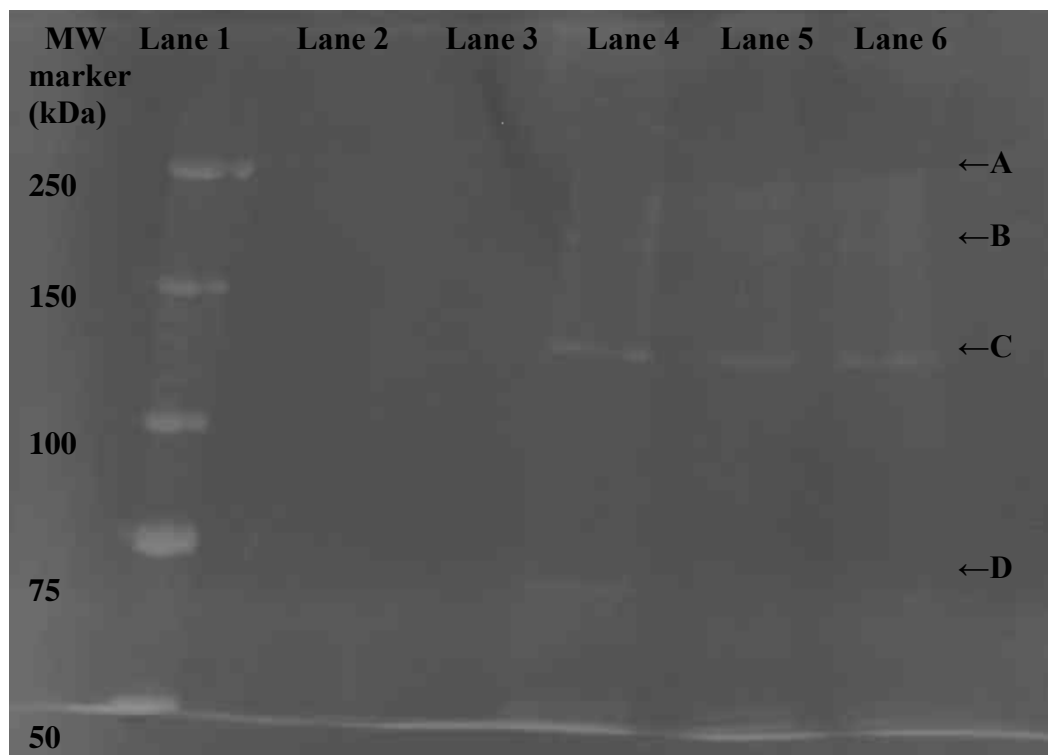
Figure 3.1 Analysis of proteins deposited in the enclosure of *Avicularia avicularia*. Freshly deposited silk was dissolved in ~37 M LiSCN and the proteins resolved by SDS-PAGE analysis followed by Coomassie Blue staining. Lane 1 - molecular weight markers (All-Blue Bio-Rad), lanes 4, 5 and 6 identical loadings of 5% w/v silk/LiSCN. Four distinct bands can be seen (A, B, C and D). Negative image was presented and modified using Photoshop cs5 to improve image clarity.
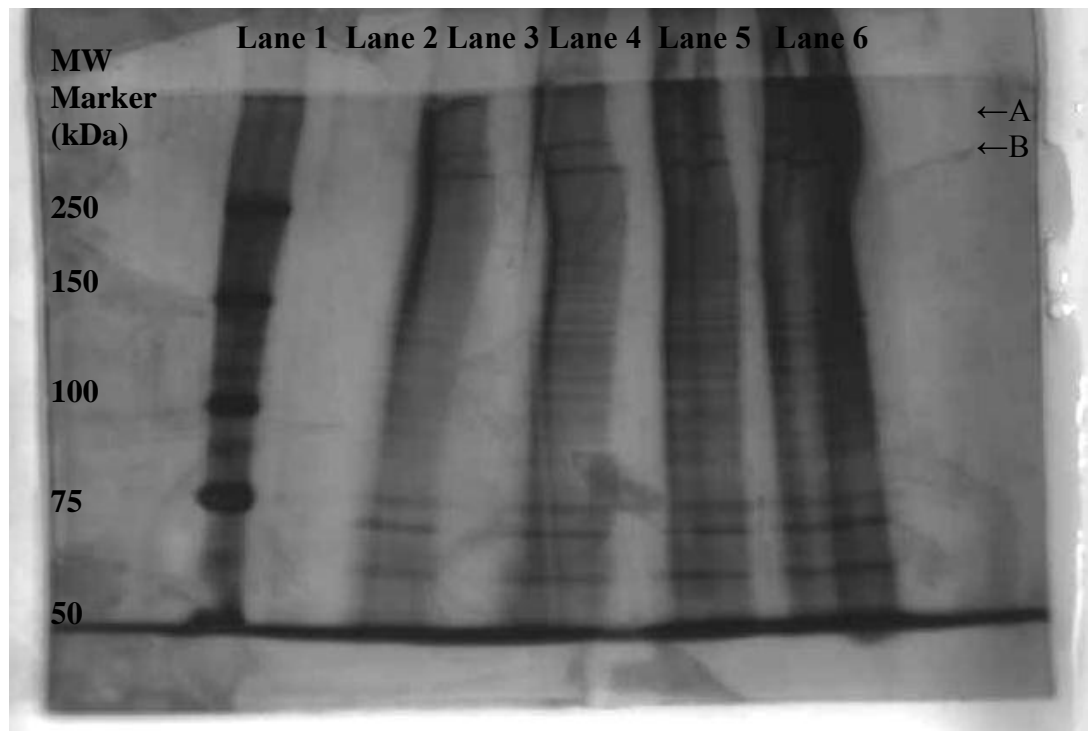
Figure 3.2 Analysis of proteins deposited in the enclosure of *Avicularia avicularia* (increased sensitivity). Freshly deposited silk was dissolved in LiSCN and the proteins resolved by SDS-PAGE analysis, followed by silver staining. The main constituents of the fibres appear at ~300 kDa as a doublet (A and B), possibly corresponding to Spidroin 1 and Spidroin 2 (Bittencourt *et al*., 2010) and were later removed with a clean sterile scalpel for MS analysis. Lane 1 - molecular weight markers (All-Blue Bio-Rad), lanes 3, 4, 5 and 6 were loaded with 30% w/v silk: LiSCN, 10, 20, 30 and 37 M LiSCN solutions respectively.
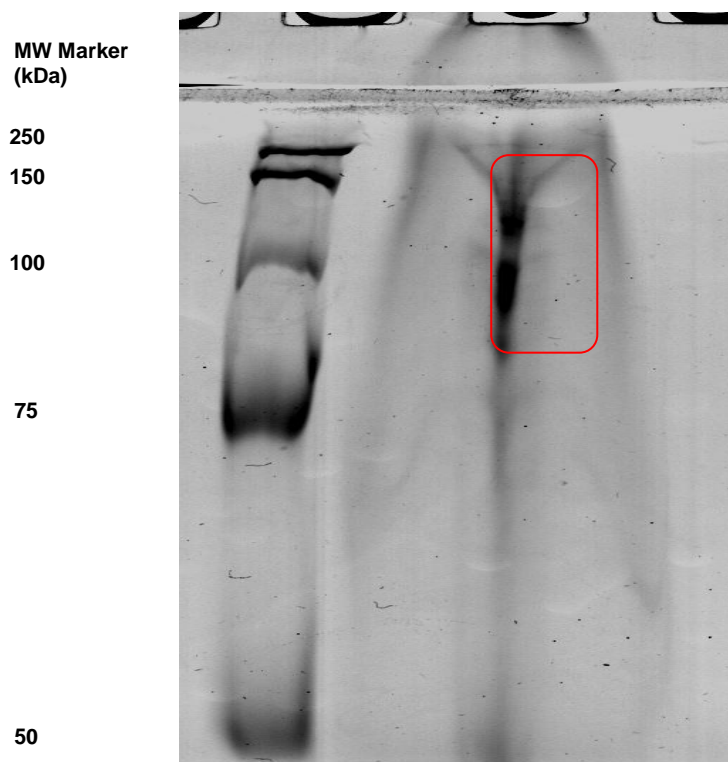
Figure 3.3 Analysis of proteins deposited in the enclosure of *Avicularia avicularia* (20 M LiSCN). Freshly deposited silk was dissolved in ~20 M LiSCN and the proteins resolved by SDS-PAGE analysis, followed by silver staining. Contrast of image was modified using Photoshop cs5 to improve image clarity. Here, the main constituents of the fibres appear at ~100 and 130 kDa but as a smear (highlighted). Lane 1 - molecular weight markers (All-Blue Bio-Rad), while lane 2 was loaded with 30% w/v silk.

3.3.1.1 Band selection and segregation

The 300 kDa bands as noted above do not appear to be the most abundant component of the loaded native silk, either that or they are far less readily dissolved under the chaotropic denaturing conditions, thereby decreasing their presence on the gel. Based upon previous studies (Xia *et al.*, 2010), it was presumed that the silks would be the higher molecular weight (~300 kDa) bands on the gel and so these were analysed preferentially over the lower molecular weight fractions. Tryptic digests were nevertheless performed on all fractions and those peptides' masses determined by MALDI-MS analysis (Table 3.1).

3.3.1.2 Silk progenitor peptides

Nineteen *de novo* peptides were found that appeared definitively to be from a silk progenitor (determined by threshold ($p < .05$) matches to database sequences), corresponding to mass-to-charge ratios (*m/z*) of: 590.00, 609.77, 618.00, 669.80, 692.41, 696.36, 811.36, 859.51, 873.45, 885.47, 975.95, 989.97, 989.98, 1021.84, 1050.51, 1133.11, 1133.60, 1161.00 and 1316.57. The product ion spectra of the peptides 873.45, 989.98 and 1133.11 are shown in Figure 3.4, Figure 3.5 and Figure 3.6 respectively as these have high significance due to their pronounced similarity to identified silk sequences. The peptide of *m/z* 989.97 is a confirmatory ion relating to *m/z* 989.98, giving a higher degree of certainty to the amino acid constitution of the latter. A further 25 fragments with mass to charge ratios 514.80, 557.00, 584.35, 600.34, 603.00, 603.79, 607.29, 613.71, 618.74, 642.28, 646.31, 648.34, 651.84, 686.36, 730.89, 755.71, 781.33, 788.33, 1002.50, 1004.40, 1021.84, 1037.33, 1057.04, 1066.41, 1146.65 and 1354.00 were deemed to be either too short or of insufficient homology to database sequences to be reliably assigned, retrieving highly positive *E*-values or no matches at all. A further three relating to *m/z* ratios of 779.35, 791.83, 1324.00 were sufficiently long to obtain viable alignments with database sequences but 'contaminations' of the database were prevalent, i.e. along with viable silk proteins, the BLASTx search (default parameters) also pulled up keratins. Due to silk having regions being homologous to keratins, these could potentially be derived from a silk progenitor but could only be conclusively determined through future studies (i.e. Edman degradation (Edman 1950) or 454 sequencing of the DNA).

Table 3.1 Assignment of generated peptides to protein families by collision-induced fragmentation by ESI-MS/MS of selected peptide ions from in-gel digested protein bands (separated by SDS-PAGE as in Figure 3.1, Figure 3.2 and Figure 3.3).

Peptide family	Isotope averaged molecular mass	Peptide ion m/z	z	Sequence	Comments and most probable sequence based on *E*-values
Silk		590		--GGGSGGGTAR	fibroin 2 [*Deinopis spinosa*]
		609.77		ssavsgeggpag/qr	major ampullate Spidroin 2 [*Latrodectus hesperus*]
		609.77		SSAVSGSEGGGPAR	major ampullate Spidroin 2 [*Latrodectus hesperus*]
		618		NAGGGGGDFGQGSR	fibroin 2 [*Dolomedes tenebrosus*]
		669.8		GG/ N GGGSGSGGGFG---	major ampullate Spidroin 2-2 [*Kukulcania hibernalis*]
		692.41		vaLASVLAyfe---	fibroin 1 [*Aliatypus plutonis*]
		696.36		dt/clleaslaesslr	tubuliform spidroin-like protein [*Nephilengys cruentata*]
	1620.7	811.36	2	RSGAGSGAGEGSGSGAPFL	major ampullate spidroin [*Agelenopsis aperta*]
	858.5	859.51	1	ELTDLLR	tubuliform Spidroin 1 [*Argiope argentata*]
	2617.3	873.45	3	teAVSEALTAAFLHTTQV-----R	ampullate Spidroin 1 locus 3 [*Latrodectus geometricus*]
	884.5	885.47	1	QQPPMLR	[*Plectreurys tristis*] spidroin
		885.47		QQPPFLR	[*Plectreurys tristis*] spidroin
	1949.88	975.95	2	**VEEFNVLEDTGASQTVGR**	major ampullate Spidroin 2 [*Latrodectus geometricus*]
		989.97		---sLADLVASE----	major ampullate Spidroin 1 [*Latrodectus mactans*]
	1977.9	989.98	2	AFAASLADiVASEGGGGSLSQK	major ampullate Spidroin 1 [*Latrodectus mactans*]
	1977.9	989.98	2	AFAASLADLVASEGGFLVLK	major ampullate Spidroin 1 [*Latrodectus mactans*]
	1282.7	1021.84	1	**LAASVLAGALLER**	fibroin 3 [*Bothriocyrtum californicum*]
	2099	1050.51	2	**AASSAASSEFKQYLV--**	fibroin 3 [*Plectreurys tristis*]
	2264.2	1133.11	2	DDLQSLSESLLSTLSLLRFK	egg Case Silk [*Nephila Antipodiana*]
	2265.2	1133.6	2	QSLSESLLSTLSLLSTSR	egg Case Silk [*Nephila Antipodiana*]
		1161		**-NQGGGGGSGGAGSGNL--**	silk protein [*Nephila inaurata madagascariensis*]
		1316.57		--GNGSGGGFK	major ampullate Spidroin 2-1 [*Kukulcania hibernalis*]

Unknown	1027.6	514.8	2	------DllR	x
	1027.6	514.8	2	FPSLakplR	x
		557		VGQVR	major ampullate Spidroin 2 [*Latrodectus hesperus*]
	1166.7	584.35	2	VLLESLAAVRP	pyriform spidroin [*Argiope trifasciata*]
	1198.7	600.34	2	NGFPRPSLRR	x
	1198.7	600.34	2	GNFPPDMYCR	x
	1198.7	600.34	2	RREADMYCR	x
		603		ALDGGGDGFVGNK	x
	1205.6	603.79	2	QVLDFLNMAR	x
	1212.6	607.29	2	-tSSTAALFA--	fibroin 1 [*Bothriocyrtum californicum*]
	1223.4	613.71	2	--ddvv-----R	x
		618.74		QNGGGGDFGQSGR	Spidroin 1 [*Argiope trifasciata*]
		642.28		NNFESLDECR	Kunitz domain?
		646.31		PEQEDLEFER	paramyosin [*Latrodectus hesperus*]
		648.34		NPMSALN-----R	x
		651.84		SLDLDLSLAEVK	x
		651.84		VTDLDLSLAEVK	x
		686.36		NLTEDVVPEAER	x
		686.36		NLTEDVV-----LR	x
		730.89		VDLLGGqEL-----K	egg case fibroin [*Latrodectus hesperus*]
		755.71		----mlstlsllstsr	x
		781.33		LAGADLET------	x
		788.33		EEQAEPDQEMLEK	x
		1002.5		-- pgALVEAVpgae	elongation factor?
		1004.4		---QLVEAVpg--TR	elongation factor?
	2041.7	1021.84	2	---------EVLTLGNER	actin?
	2041.7	1021.84	2	--- LPNGEVLTLGNER	actin?

99

	2072.6	1037.33	2	---pyAPLLDELADE--	x
	2072.6	1037.33	2	---meAPLMPELADE---	x
	2112.1	1057.04	2	----QLESMQ------K	x
	2130.8	1066.41	2	--FQLVEAP/v----R	x
		1066.41		------vLMLTR	x
	2291.9	1146.65	2	--eeP---- DEDDEVLDR	x
		1354		-GGFGKGSGGGFGGASGGGGFGKGGG--	flagelliform silk protein [*Argiope trifasciata*]
Indeterminate		779.35		NDLEGQVQELQER	spidroin or keratin
		791.83		gLNFGSGGGGG---	spidroin or keratin
		1324		-gNGPGSGG---GSR	spidroin or keratin
Structural	975.4	488	2	K/AGFAGDDAP/R	actin
	975.4	488.72	2	AGFAGDDAPR	actin
		516		R/GYSFVTTAER/E	keratin
	1030.6	516.31	2	VLDELTVSR	keratin
		530		K/TLLDIDNTR/M	keratin
	1059.6	531.29	2	TLLDLDNTR	actin
	1129.5	565.77	2	GYSFVTTAER	actin
		589.27		EITALAPSTMsoK	actin
		589.27		ELTALAPSTMsoK	keratin
		616		R/SGGGGGGGGLGSGGSIR/R	keratin
		616		SGGGGGGGGLGSGGLSR	keratin
		616.79		SGGGGGGGGLGSGGSIR	keratin
		616.79		SGGGGGGGGLGSGGVEK	keratin
		616.79		SGGGGGGGGLGSGGDLK	keratin
Trypsin		680.14		DNDLEWLGAQAK	trypsin
		743		vatvslpr	trypsin

100

Observed MW: 2617.3311   Precursor ion charge state: 1
M/z tolerance: 0.30   Intensity threshold: 2 (0.750%)

	Thr	Glu	Ala	Val	Ser	Glu	Ala	Leu	Thr	Ala	Ala	Phe	Leu	His	Thr	Thr	Gln	Val
a	74.06	203.10	274.14	373.21	460.24	589.28	660.32	773.40	874.45	945.49	1016.53	1163.59	1276.68	1413.74	1514.79	1615.83	1743.89	1842.96
	---	-0.01	---	0.00	---	-0.08	-0.05	-0.01	-0.00	0.04	---	---	---	-0.03	-0.22	---	---	---
b	102.06	231.10	302.14	401.20	488.24	617.28	688.32	801.40	902.45	973.48	1044.52	1191.59	1304.67	1441.73	1542.78	1643.83	1771.89	1870.96
	-0.01	-0.01	-0.00	-0.03	---	---	---	-0.01	---	-0.11	---	---	---	---	---	---	---	---
y	2618.34	2517.29	2388.25	2317.21	2218.14	2131.11	2002.07	1931.03	1817.95	1716.90	1645.86	1574.83	1427.76	1314.67	1177.61	1076.57	975.52	847.46
	---	---	-0.09	-0.08	-0.10	-0.11	-0.06	-0.07	-0.05	-0.08	-0.05	-0.06	-0.05	-0.04	-0.05	-0.05	-0.05	-0.05
z	2601.31	2500.26	2371.22	2300.18	2201.11	2114.08	1985.04	1914.00	1800.92	1699.87	1628.83	1557.80	1410.73	1297.64	1160.58	1059.54	958.49	830.43
	---	-0.03	---	---	---	---	-0.04	---	---	0.01	---	---	-0.02	-0.04	-0.06	-0.09	---	---

a   2572.33
    ---
b   2600.33

    729.42

y   748.39
    -0.05
z   731.36
    -0.06

Figure 3.4 Raw product ion spectra for the peptide 873.45 (MW: 2517.3311). Note: the y-series reads from the right to left.

Observed MW: 1977.9376   Precursor ion charge state: 1
M/z tolerance: 0.30   Intensity threshold: 4 (0.350%)

a	44.05	191.12	262.16	333.19	420.22	533.31	604.35	719.37	832.46	931.53	1002.56	1089.59	1218.64	1275.66	1332.68	1389.70	1476.73	1589.82
	---	-0.01	---	-0.01	-0.02	-0.02	0.00	0.02	-0.02	0.11	0.08	---	---	0.04	---	---	---	---
b	72.04	219.11	290.15	361.19	448.22	561.30	632.34	747.37	860.45	959.52	1030.56	1117.59	1246.63	1303.65	1360.67	1417.70	1504.73	1617.81
	---	-0.01	-0.01	-0.01	-0.01	-0.02	-0.02	-0.01	-0.02	-0.02	0.00	---	---	---	-0.22	---	---
	Ala	Phe	Ala	Ala	Ser	Leu	Ala	Asp	Leu	Val	Ala	Ser	Glu	Gly	Gly	Gly	Ser	Leu
y	1979.01	1907.97	1760.90	1689.87	1618.83	1531.80	1418.71	1347.68	1232.65	1119.56	1020.50	949.46	862.43	733.38	676.36	619.34	562.32	475.29
	---	---	-0.10	-0.04	-0.01	-0.02	0.00	-0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02
z	1961.98	1890.94	1743.87	1672.84	1601.80	1514.77	1401.68	1330.65	1215.62	1102.53	1003.47	932.43	845.40	716.35	659.33	602.31	545.29	458.26
	---	---	---	---	---	---	---	-0.07	---	-0.03	-0.08	-0.03	0.01	0.03	0.02	-0.01	0.03	0.02

a	1676.85	1804.91	1933.00
	---	---	---
b	1704.84	1832.90	1961.00
	---	---	---
	Ser	Gln	Lys
y	362.20	275.17	147.11
	0.02	-0.01	-0.00
z	345.17	258.14	130.08
	0.02	-0.01	-0.01

spider silk in gel digest pooled
IG_2A_MP1710_23July10 MaxEnt 3 259 [Ev-124226,It50,En1] (0.050,200.00,0.200,1400.00,2)

2: TOF MSMS 989.98ES+

K — Q — S — L — S — G — G — G — E — S — A — V — L — D — A — L — S — A — A — FA   yMax

120.08;F
191.12 a2
219.12 b2
290.16 b3
325.20
401.23
461.25
472.27
543.31
619.32 y6
676.34 y7
733.37 y8
747.38 b8
844.41
862.42 y9
931.42 a10
949.45 y10
1020.49 y11
1101.54
1119.56 y12
1214.64
1232.64 y13
1329.66
1347.68 y14
1418.71 y15
1513.80
1531.81 y16
1618.84 y17
1689.91 y18
1828.09
1898.23
1963.90

Figure 3.5 Raw product ion spectra for the peptide 989.98 (MW: 1977.9376). Note: the y-series reads from the right to left.

Observed MW: 2264.2122   Precursor ion charge state: 1
M/z tolerance: 0.30   Intensity threshold: 1 (0.750%)

a	88.04	203.07	316.15	444.21	531.24	644.33	731.36	860.40	947.43	1060.52	1173.60	1260.63	1361.68	1474.76	1561.80	1674.88	1787.96	1944.07
	---	-0.00	-0.01	-0.02	-0.09	---	-0.05	-0.05	---	---	0.02	---	---	---	---	---	---	---
b	116.03	231.06	344.15	472.20	559.24	672.32	759.35	888.40	975.43	1088.51	1201.60	1288.63	1389.68	1502.76	1589.79	1702.88	1815.96	1972.06
	---	-0.01	-0.01	-0.02	-0.02	-0.01	-0.06	---	-0.04	-0.04	---	---	---	---	---	---	---	---
	Asp	Asp	Leu	Gln	Ser	Leu	Ser	Glu	Ser	Leu	Leu	Ser	Thr	Leu	Ser	Leu	Leu	Arg
	56	56	56	48	55	100	100	100	100	100	100	100	100	100	100	100	100	97
y	2265.20	2150.17	2035.14	1922.06	1794.00	1706.97	1593.89	1506.85	1377.81	1290.78	1177.69	1064.61	977.58	876.53	763.45	676.41	563.33	450.25
	---	---	---	-0.04	-0.03	-0.00	-0.02	-0.02	-0.03	-0.05	-0.02	-0.01	-0.02	-0.00	-0.01	-0.02	0.00	-0.00
z	2248.17	2133.14	2018.11	1905.03	1776.97	1689.94	1576.85	1489.82	1360.78	1273.75	1160.66	1047.58	960.55	859.50	746.42	659.38	546.30	433.22
	---	---	---	---	---	-0.06	---	---	---	---	---	---	0.03	-0.05	0.23	---	0.03	-0.01

a	2091.13	2219.19
	---	---
b	2119.13	2247.19
	---	---
	Phe	Gln
	95	84
y	294.15	147.08
	0.01	-0.01
z	277.12	130.05
	---	---

spider silk in gel digest pooled
1G_2A_MP1710_23July10 MaxEnt 3 420 [Ev281922,It50,En1] (0.050,200.00,0.200,1400.00,2)   2: TOF MSMS 1133.11ES+

Figure 3.6 Raw product ion spectra for the peptide 1133.11 (MW: 2264.2122). Note: the y-series reads from the right to left.

103

3.3.1.3 Keratin contaminants and trypsin fragments

Despite 791.83 aligning very well with a flagelliform silk protein (*Argiope trifasciata*), it likewise aligns (albeit to a far lower '*E*-value') with the Lk-a protein (keratin) (*Xenopus laevis*), somewhat expected due to silks homologies to keratin and often being classed as such. The task of accurately determining the usefulness of peptides in this instance is confounded by the prevalence of amino acids most MS is least adept at identifying, in this case glycine chains, leucines and isoleucines (Bishop *et al*., 2007). Finally, a further 10 sequences were identified corresponding to structural proteins (keratin and actin), with *m/z* ratios of 488.00, 488.72, 516.00, 516.31, 530.00, 531.29, 565.77, 589.27, 616.00 and 616.79 and for completeness, two more fragments with *m/z* ratios corresponding to trypsin (*m/z* = 680.14 and 743.00). Trypsin fragments are remnants from the digestion prior to mass spec and it is reasonable to assume that the keratin peptides resulted from contamination (possibly from airborne human skin/dust settling on the SDS-PAGE gels); however the actin readings are unexpected as these are not known to be components of typical silks.

3.3.2   ANALYSES

3.3.2.1 MASCOT and BLAST searches and interpretations

MASCOT searches were performed on the raw, uninterpreted data but
this yielded no significant hits, or non-silk proteins (trypsin, keratin). All
sequences were subjected to a BLASTx database search (default parameters) to
identify closest sequence homologies with particular attention paid to
occurrences of keratin or human structural proteins. Sequences were then
subjected to a BLASTx search with the Organism parameter set to "arachnids
(taxid:6854)." These BLAST E-value cores were expectedly low due to the
short peptide inputs so were assessed relative to other pulled sequences.
Peptides were classified into categories according to their presence on the
BLASTx searches: silk hits only, silk and structural protein hits, structural
protein hits and 'unknown.' The 'unknown' protein searches usually yielded
results but are classed in this category because either the sequence was too
short to get a significant result (E-value) from the search, produced a "no
significant similarity found" result or produced a result that contained
significant numbers of hypothetical proteins or implausible sequences.

3.4 DISCUSSION

3.4.1 CHARACTERISATION AND ALIGNMENTS OF *DE NOVO* 'SILKS'

It was noted that the silk sequences fell into three categories: N-terminal fragments, unassigned but characteristically silk fragments and repetitive region fragments. It was surprising that the latter yielded very few peptide fragments with only one corresponding to what could be considered as a member of the Spidroin 1 (MaSp1) family and one other belonging to the Spidroin 2 (MaSp2) family (Bittencourt *et al*., 2010). Alignments of the MaSp1 putatively orthologous transcripts (repetitive region and C-terminal region) and MaSp2 alongside the MaSp1 transcripts (C-terminal end) are shown in Figure 3.7, Figure 3.8 and Figure 3.9. It should be noted that while there is a high similarity/identity between both the repeats and C-terminal ends of the orthologous transcripts of Spidroin 1 (1.1, 2.1 and 3.1), with very few insertions, deletions or substitutions; this similarity drops considerably when aligned with Spidroin 2, in both the repetitive region and the C-terminus. There are however conservations amongst the typically resilient residues, for example the QALLE motif (Challis *et al*., 2006), Q/E, N/D and P.

```
1.1        AFAFASAFSQVLSNYGLLNINNAYSLASSIASAASSSASSAAAAAASSSSSAAAGAAAASG  220
2.1        AFAFASGVSQVLSNYGLINLSNALFLASSIANAASASASSAAAAASSSS-AATGAAAALG  169
3.1        AFAFASAFSQVLSNYGLLNISNAYSLASSIANAASASASSAAAAAASSSSSAAAGAAAASG  420
           *****..********:*:.**  ****** .***:**********:*** **:***** *

1.1        TAASAAATSTTTTTSTSRAAAAAS--AAAAASASGAADAAGAASAASAASASSSLQQSLG  278
2.1        GAGSAAATSTTTITSTSTAVAAAS--------GSGAARAAQTASAASAASASSSLAQSLG  221
3.1        AAGSVAATSTTTTASTSTAAAAASAAAAAAASASGAARAAGASSAASAASASSSLQQSLG  480
            *.*.******* :*** *.****       .**** ** ::************ ****

1.1        SALAQSSSFAAAFAQANSAASAAAIAYALAQTVANQIGFSSYSSAFASAASSAVSSLGGF  338
2.1        SALAQSSSFAAAFDQGNSAASAAAIAYVLAQSAANKVGLSSYSAAISNAASAAVESVGGY  281
3.1        SALAQSSSFAAAFAQANSAASAAAIAYALAQTVANQIGFSSYSSAFASAASSAVYSLGSF  540
           ************ *.************ .***:. **::*:****:*::.***:** *:*.:

1.1        ASASAYAFAFASAFSQVLSNYGLLNINNAYSLASSIASAASSSASSAAAAASYSFSATG-  397
2.1        ASASAHAFAFASAVSQVLSNYGLINLSNALSLASSIANAVSASASSAAAVSS-AAAATG-  339
3.1        ASASAYAFAFASAFSQVLSNYGLLNINNAYSLASSIANAASASASSAAAAAASSSSAAAG  600
           *****:*******.**********:*:.** *******.*.*:********.:: : :*:.
```

Figure 3.7 ClustalW alignments of consensus repeats for Spidroin 1 from *A. juruensis*. Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)), blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q.

```
1.1    ---------------SSATGVGGSVSSGASPASAGTATGGGISFLPVQTQRGFGLVPSP 465
2.1    ---------------SSATGVGETVACATSPASTGTAAGGGISSLPVQTQPGFGFLLSP 402
3.1    ATAAGAASTSVPATSLSSATRVGGSLSSAVSPASARTATGDGTTYLPVQIQPGIGFVPSL 840
                       ****  **  :::...****: **:*.*  :  **** *  *:*:: *

1.1    SGNIGANFPGSGEFGPSPLTSP------------VYGPGILGPGLVVPSLQGLLPPLFV 512
2.1    SGNIGPSVSGSGGFGPSPLPSPASDGFSPSPLPSQVYGPGILGPGLVAPSLEGLLPPLSI 462
3.1    SGDIGPNVPGSGGFGSPALPSP------------VYGPAILGPGLVAPALANLLPPLSV 887
       **:**....*** **...*.**         ****.*******.*:*  .***** :

1.1    LPSNSATERISSMVSSLLSAVSSNGLDASSFGDTIASLVSQISVNNSDLSSSQVLLEALL 572
2.1    LPSDSANERISSVVSSLLAAVSSNGLDASSLGDNLASLVSQISANNADLSSSQVMVEALL 522
3.1    LPSDSANERISSVVSSLLSAISSNGLDASSLGGTIASLVSQISVSNAKLSSSQVFLEALL 947
       ***.** .*****:*****:*:*********:*  ..:********  .*:.******::****

1.1    EILSGMVQILSYAEVGTVNTKTVSSTSAAVAQAISSAFSGNQNS-AA---RFFFIKYF-N 627
2.1    EVLSGIVQILSYAEVGAVNTETVSSTSSAVAQAISSAVLG-SKFLSCLMKLFFF-QIF-K 579
3.1    EVLSGMVQILSYAEVGAVNTDTVISTSSAVAQAISSAVSG---------TVFFFNECL-N 997
       *:***:**********:***.** ***:*********. *         *** : : :
```

Figure 3.8 ClustalW alignment of the C-terminal domain for Spidroin 1 from *A. juruensis* showing a high degree of identity for acidic, prolyl, alanyl and acidic residues. Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)), blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q.

```
1.1          ----------------SSATGVGGSVSSGASPASAGTATGGGISFLPVQTQRGFGLVPSP 465
2.1          ----------------SSATGVGETVACATSPASTGTAAGGGISSLPVQTQPGFGFLLSP 402
3.1          ATAAGAASTSVPATSLSSATRVGGSLSSAVSPASARTATGDGTTYLPVQIQPGIGFVPSL 840
Spidroin2    ----------------YGKKLGFEFGTPAAAAVT-LGPGAGQQGPGGAGQQGPGGQGPY 334
                      ..  :*    ..    .:.* :  ..* *       * * *      .

1.1          SGNIGANFPGSGEFGPSPLTSP-------------VYGPGILGPGLVVPSLQGLLPPLFV 512
2.1          SGNIGPSVSGSGGFGPSPLPSPASDGFSPSPLPSQVVYGPGILGPGLVAPSLEGLLPPLSI 462
3.1          SGDIGPNVPGSGGFGSPALPSP-------------VYGPAILGPGLVAPALANLLPPLSV 887
Spidroin2    GPVASAAAVAGGYGPGALPQG---------------PARQGPSGPVSSAPVASAAAAR 378
              .     ..   . :* :*. .*..               *.   **.  ..:       ..

1.1          LPSNSATERISSMVSSLLSAVSSNGLDASSFGDTIASLVSQISVNNSDLSSSQVLLEALL 572
2.1          LPSDSANERISSVVSSLLAAVSSNGLDASSLGDNLASLVSQISANNADLSSSQVMVEALL 522
3.1          LPSDSANERISSVVSSLLSAISSNGLDASSLGGTIASLVSQISVSNAKLSSSQVFLEALL 947
Spidroin2    LSSPQASSRVS---SAFFSLVSSGPTSPGALSNAISSVVSQVSASNPGLSGCDVLVQALL 435
              *.*  .*..*:*    *::::: :**.    ...::.. ::*:***:*..*. **..:*:::***

1.1          EILSGMVQILSYAEVGTVNTKTVSSTSAAVAQAISSAFSGNQNS 616
2.1          EVLSGIVQILSYAEVGAVNTETVSSTSSAVAQAISSAVLG---- 562
3.1          EVLSGMVQILSYAEVGAVNTDTVISTSSAVAQAISSAVSG---- 987
Spidroin2    EIVSALVSILASSSIGQINYGASAQYASLVGQSVNQALRY---- 475
              *::*.:*.**: :.:* :*   :   . :: *.*::..*.
```

Figure 3.9 ClustalW alignment of the C-terminal domains of Spidroin 1 aligned with that of Spidroin 2 from *A. juruensis*. Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)), blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q.

Kyte-Doolittle structural analyses illustrate the significant differences between the Spidroin 1 and Spidroin 2 repetitive region (Figure 3.10 and Figure 3.11 respectively) and the Spidroin 1 and Spidroin 2 C-terminal domains (Figure 3.12 and Figure 3.13 respectively), (Kyte and Doolittle 1982). The analyses predict a predominantly helical N-terminus in the case of *Nephila clavipes* (Figure 3.14) in agreement with the Engelman and Steitz (1981) helical hairpin hypothesis. This hypothesis states that the leader peptide of a helical structure is followed by a turn (coil) and then another helical region. An example of helical motifs spread over five different species, using 12 different algorithms is shown in Appendix 4. This helix-turn-helix conformation is also present in the contig peptide generated from the MS data (Figure 3.17).
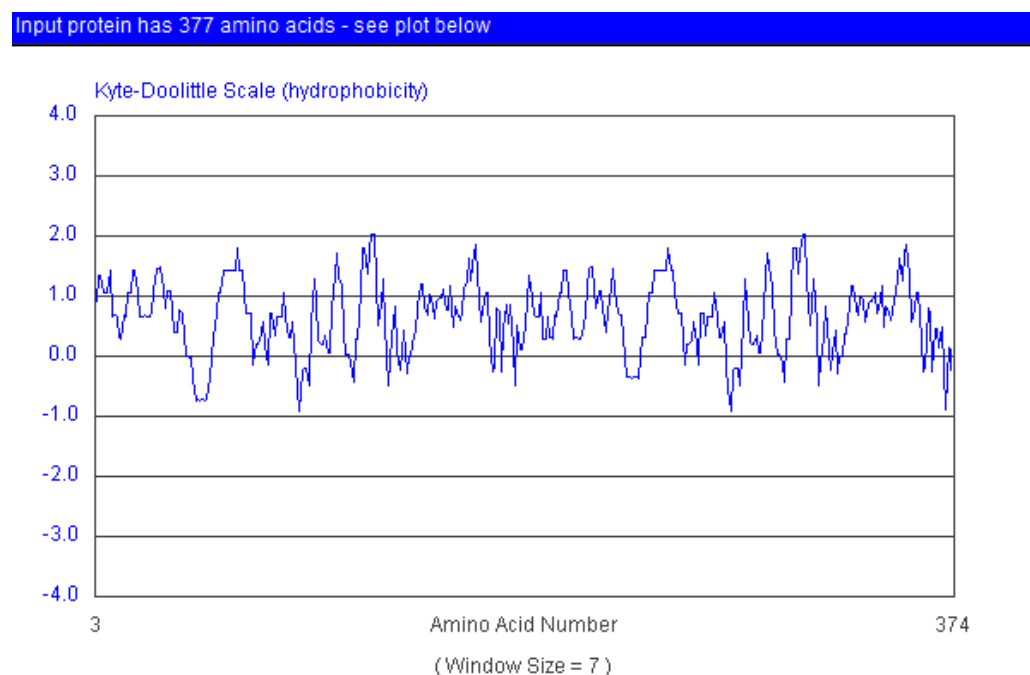
Figure 3.10 Kyte-Doolittle plot of two consensus repeats for Spidroin 1 (3.1) from Bittencourt *et al.* (2010). The Kyte-Doolittle hydropathy prediction method was used to identify obvious motifs. Negative points on the axis indicate the degree of hydrophilicity, while positive indicate hydrophobicity.

Kyte-Doolittle Scale (hydrophobicity)

4.0

3.0

2.0

1.0

0.0

-1.0

-2.0

-3.0

-4.0

3 Amino Acid Number 91

(Window Size = 7)
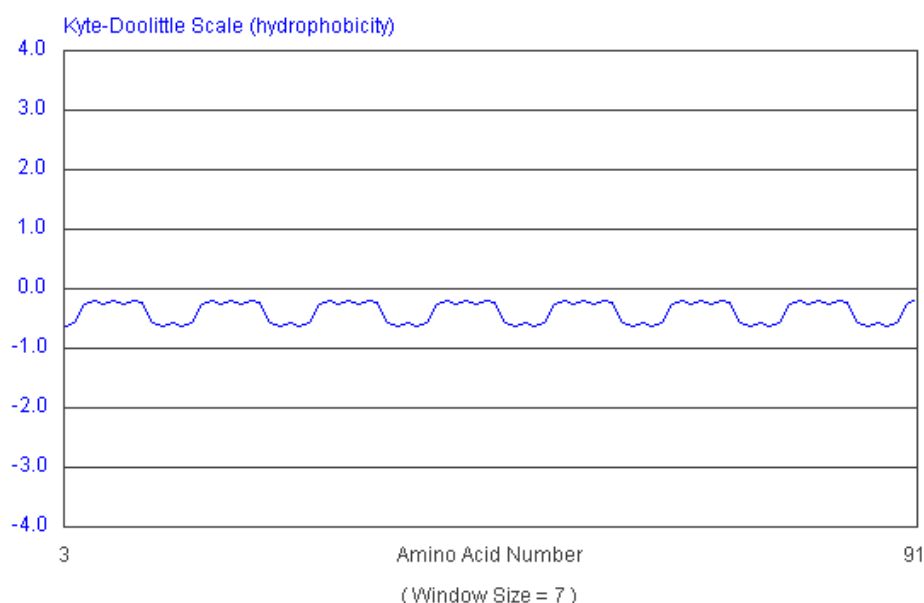
Figure 3.11 Kyte-Doolittle plot of seven consensus repeats for Spidroin 2 from Bittencourt *et al.* (2010). The Kyte-Doolittle hydropathy prediction method was used to identify obvious motifs. Negative points on the axis indicate the degree of hydrophilicity, while positive indicate hydrophobicity.

Kyte-Doolittle Scale (hydrophobicity)

4.0

3.0

2.0

1.0

0.0

-1.0

-2.0

-3.0

-4.0

3 Amino Acid Number 116

(Window Size = 7)
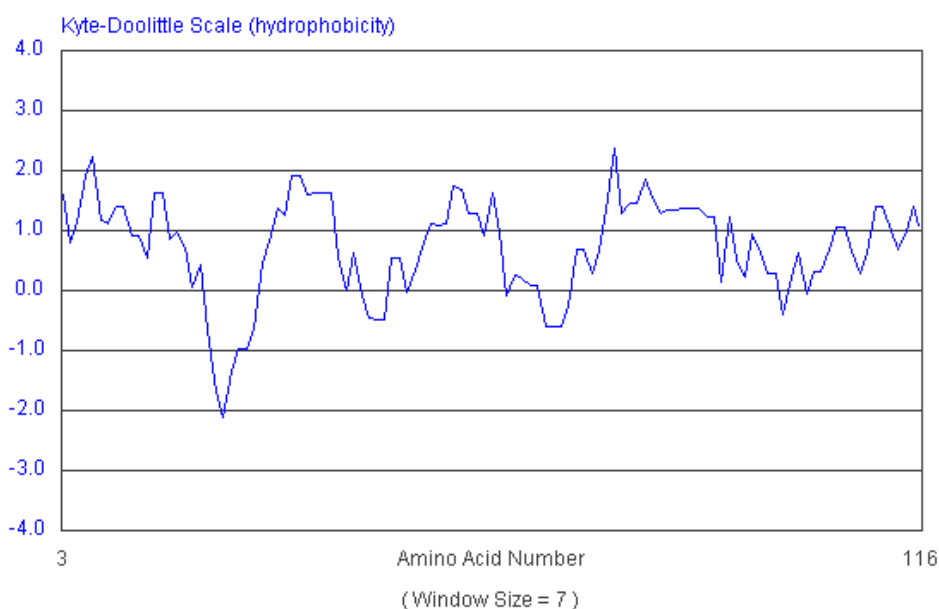
Figure 3.12 Kyte-Doolittle plot of the C-terminal domain for Spidroin 1 from Bittencourt *et al.* (2010). The Kyte-Doolittle hydropathy prediction method was used to identify obvious motifs. Negative points on the axis indicate the degree of hydrophilicity, while positive indicate hydrophobicity.

111
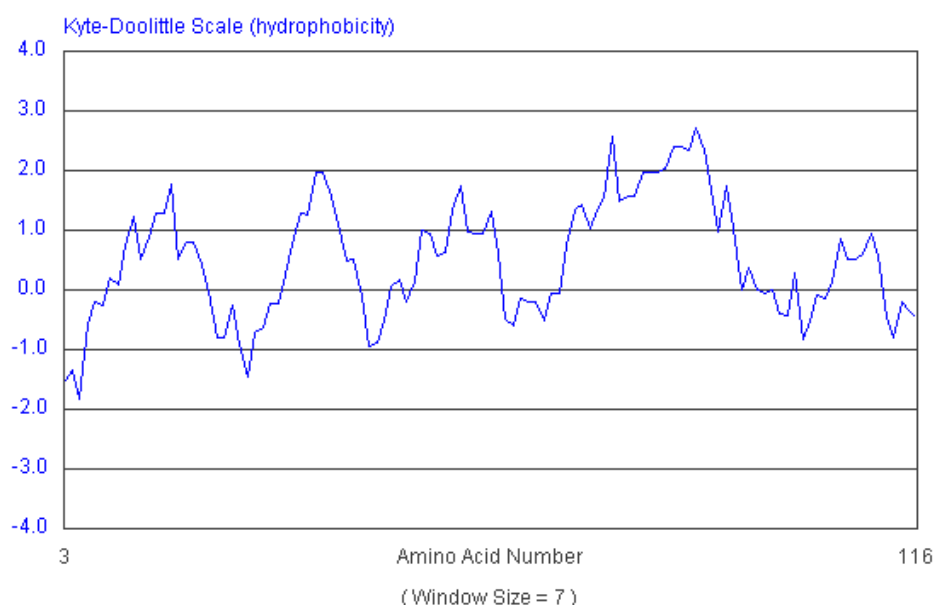
Figure 3.13 Kyte-Doolittle plot of the C-terminal domain for Spidroin 2 from Bittencourt *et al.* (2010). The Kyte-Doolittle hydropathy prediction method was used to identify obvious motifs. Negative points on the axis indicate the degree of hydrophilicity, while positive indicate hydrophobicity.
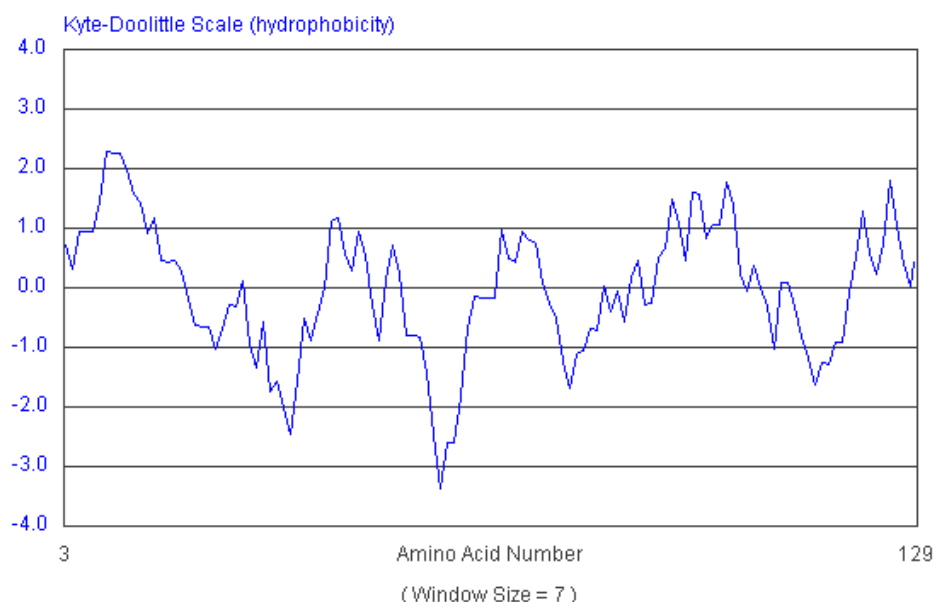
Figure 3.14 Kyte-Doolittle plot of the N-terminal domain for the major ampullate spidroin 1A precursor. The Kyte-Doolittle hydropathy prediction method was used to identify obvious motifs. Negative points on the axis indicate the degree of hydrophilicity, while positive indicate hydrophobicity.

112

3.4.2 Contig generation of major ampullate silk proteins

The C-terminal domain is a short, non-repetitive region that recent findings suggest is involved in silk processing (Jin and Kaplan 2003) and is retained in both the silk dope as well as the native silk fibres (Sponner *et al.*, 2004). The N-terminal regions are likewise retained in the native fibres but little is known about their function. Multiple alignments of N-termini show particular conservation in methionine residues and identities become increasingly more prevalent nearer to the repetitive region. Methionine is rarely found within the repetitive region and the C-termini of silk proteins (data not shown), however within the N-terminal domains, it is found in relatively high abundance. Here, it is theorised to create additional translation start sites (Motriuk-Smith *et al.*, 2005) and in the event of a deletion of the first methionine, translation can be initiated downstream, albeit with a shorter transcript.

It is not known whether these shorter isoforms are actually intentionally utilised. The loss of a short sequence within the N-terminus does not appear to significantly alter the structure/properties of silk as proven when purely the repetitive region is used to produce artificial homologues in expression vectors (Xia *et al.*, 2010), for example with the motif:

$$[SGRGGLGGQGAGMAAAAAMGGAGQGGYGGLGSQGT]_n.$$

Of the peptides sequenced in this study, three peptide fragments clearly fell into the N-terminal domain classification (m/z = 989.98: AFAASLADiVASEGGGSLSQK, 873.45: teAVSEALTAAFLHTTQV-----R and 692.41 vaLASVLAyfe---) and as such were loaded into a multiple sequence alignment package (ClustalW2) to develop a contig assembly (Figure 3.15). When aligned with the nearest phylogenetic relative subsequent to a BLASTp search (*Euprosthenops australis*) assuming most likely candidate amino acids where applicable (I/L, Q/K), there was a 46% identity and an 82% similarity between the contig assembly and its nearest neighbour. Likewise, this contig included amino acids frequently found in the N-terminal domain such as E, F, L, M and T, which rarely appear in the repetitive region, but conversely lacked its typical hydrophobic A,G, I, L and V residues found therein and show no typical 'motifs'. A multiple sequence alignment with other species more adeptly illustrates the amino acid conservation (Figure 3.16) and helical regions (Figure 3.17). Helicality predictions of this MS peptide against corresponding sections of representative spidroins are given in Table 3.2.

```
MASP1prec.E.australis      GFTASQLDDMSTIAQSMVQSIQSLAAQGRTSPNKLQALNMAFASSMAEIA 100
m/z 989.98                 ------------------------------------------AFAASLADIV 10
                                                                     ***:*:*:*.

MASP1prec.E.australis      ASEEGGGSLSTKTSSIASAMSNAFLQTTGVVNQPFINEITQLVSMFAQAG 150
m/z 873.45 and 692.41      ASE--GGFLVLKTEAVSEALTAAFLHTTQV-----V-AIASVLAYFE--- 11
                           ***   ** *   **.::.*:: ***:** *      :   *:.::: *

MASP1prec.E.australis      MNDVSASASAGASAAASAGAPGYSPAPSYSSGGYASSAASAAAAGQGGP 200
m/z 609.77                 --------------------------------SSAVS-----GEGGP 10
                                                           ***.*       *:***
```

```
989.98/873.45/692.41 contig: AFAASLADIVASEXXGGFLVLKTEAVSEALTAAFLHTTQVXXXXXV-AIASVLAYFE
```

Figure 3.15 ClustalW alignment of the N-terminal domains of the MaSp1 precursor of *E. australis* as a backbone for which to map on the peptide fragments *m/z* 989.98, 873.45 and 692.41 for contig assembly. *M/z* 609.77 is also shown (line 3) and illustrates a poor alignment relating to the repetitive elements. Here, the Q and E discrepancy could be explained by a deamination reaction (see Discussion). The contig generated from the three peptides: 989.98/873.45/692.41 is shown underneath.

```
Latrodectus.hesperus      MTWSTRLALSFLFVLCTQSLYALAQANTPWSSKANADAFINSFISAASNT 50
Latrodectus.geometricus   MTWSTRLALSVLLVLCTQSIYALAQANTPWSSKANADAFINSFISSAQNT 50
Argiope.trifasciata       MNWSIRLALLGFVVLSTQTVFSAGQGATPWENSQLAESFISRFLRFIGQS 50
Euprosthenops.australis   MSWTARLALLLLFVACQGSS-SLASHTTPWTNPGLAENFMNSFMQGLSSM 49
Agelenopsis.aperta        MTWTVRLAIPILILILQGSK-CLGQSN-PWTDTATAESFISSVMSSVANQ 48
                          * .*:  ***:   :.:      :   . ...   **  .   *: *:. .:      .

                                                               AFAASLADI
Latrodectus.hesperus      GSFSQDQMEDMSLIGNTLMAAMDNMG--GRITPSKLQALDMAFASSVAEI 98
Latrodectus.geometricus   GSFSQDQMDDMSLIGNTLMTAMDNMG--GRITPSKLQALDMAFASSVAEI 98
Argiope.trifasciata       GAFSPNQLDDMSSIGDTLKTAIEKMAQSRKSSKSKLQALNMAFASSMAEI 100
Euprosthenops.australis   PGFTASQLDDMSTIAQSMVQSIQSLAAQGRTSPNKLQALNMAFASSMAEI 99
Agelenopsis.aperta        GCLSYDQIDDMQAVGDTMLATMDNLVRSGKSSSHMLKAMNMAMGTSIAEI 98
                          ::  .*::**. :...:   ::::.:       : :   *.*::**..:*.*.***

Contig                    VASE--GGFLVLKTEAVSEALTAAFLVTTQV-----V-AIASVIAYFE--
Latrodectus.hesperus      AASEGG--DLGVTTNAIADALTSAFYQTTGVVNSRFISEIRSLIGMFAQA 146
Latrodectus.geometricus   AASEGG--DLGVTTNAIADALTSAFYQTTGVVNNRFISEIRSLISMFAQA 146
Argiope.trifasciata       AVAEQGGLSLEAKTNAIASALSAAFLETTGYVNQQFVNEIKTLIFMIAQA 150
Euprosthenops.australis   AASEEGGGSLSTKTSSIASAMSNAFLQTTGVVNQPFINEITQLVSMFAQA 149
Agelenopsis.aperta        VADGGG--NLGSKVSCISNALSSAFLQTTGSVNTQFVNEIVSLISMFAQA 146
                          ..  .    *    .*   ....*:.*:. **  :*** **   *:.**   *: *:***
```

Figure 3.16 ClustalW alignment of the N-terminal domains of 5 araneid spider species aligned with that of the *de novo* contig (bold/underlined) generated in Figure 3.15. Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)),blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q. GenBank accessions (from top to bottom): (contig), ABY67423.1, ABY67427.1, AAZ15371.1, CAJ90517.1 and ADM14324.1

```
Conf:    Confidence       0=low   9=high
Pred:    Predicted secondary structure  H=helix        E=strand       C=coil
AA:      Target sequence

>gi|303307750|gb|ADM14313.1| fibroin 1 [Bothriocyrtum californicum]
Conf: 93589999999999998756543114544677889365899999999866528999935775599999879999987420379990899999999989889999998469875677899999999997875319
Pred: CCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHC
  AA: MGWLTHITLSLFFVSAQFSISSARTKVTNVPWTDEAKGKKFLSTFLDYALDHGLFPQQERDDLEAISQNLIPVFRKTMDSGGNAAAKMKALNMAFASSIAEIAVQEGGAGSIEEKTQAVSEALAHAFLQTTG
             10        20        30        40        50        60        70        80        90       100       110       120       130

>gi|193506891|gb|ACF19411.1| major ampullate spidroin 1A precursor [Nephila clavipes]
Conf: 915889999999999998644432147999999787999999999996521899996677448899999999998734238890469999899889879999998608863565699999999987865309
Pred: CCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHC
  AA: MTWTARLALSFLAVICTQSLFAQGQNTPWSSTELADAFINAFMNEAGRTGAFTADQLDDMSTIGDTIKTAMDKMARSNKSSKGKLQALNMAFASSMAEIAAVEQGGLSVDAKTNAIADSLNSAFYQTTG
             10        20        30        40        50        60        70        80        90       100       110       120       130

>gi|164709232|gb|ABY67421.1| major ampullate spidroin 1 locus 1 [Latrodectus hesperus]
Conf: 925889999999999998766666413688899967599999999999753289999577866999999999999863299995258988789879879999975088831149999999998787875409
Pred: CCHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHC
  AA: MTWSTRLALSFLFVLCTQSLYALAQANTPWSSKANADAFINSFISAASNTGSFSQDQMEDMSLIGNTLMAAMDNMGGRITPSKLQALDMAFASSVAEIAASEGGDLGVTTNAIADALTSAFYQTTG
             10        20        30        40        50        60        70        80        90       100       110       120

>gi|303307754|gb|ADM14315.1| major ampullate spidroin [Diguetia canities]
Conf: 925889999999999987777640244566557899985779999999999953079999946776589999999999997400026684999999999998997999998751676446469999999999998976309
Pred: CCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCHHHHHHHHHHHHCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHC
  AA: MGCVVRLVFLAFATTQCLLAGQGVHGAMANTPFANKIMAEDFMNKFTNQLANSPYFSSQQKEDMSSIKDELISVIESMDSAHKSSAAKLQAMNMAFASAIADIAATEAYGADISLETSAIANALSEAFLQTTG
             10        20        30        40        50        60        70        80        90       100       110       120       130

MS contig peptide
Conf: 956779899875078445423799999999988766627999999864786643409
Pred: CHHHHHHHHHHHHHCCCEEEEHHHHHHHHHHHHHHHCCCCCCCCHHHHHHHHCC
AA:   AFAASLADLVASEGGFLVLKTEAVSEALTAAFLHTTQVXXXXXVXALASVLAYFE
             10        20        30        40        50
```

Figure 3.17 Secondary structure prediction using PSIpred for the N-terminal domains of four spider species, including the mygalomorph: *Bothriocyrtum californicum* **alongside the contig generated in Figure 3.15. Contig helical domains agree with those of the four other species. The underlined A (alanine) residue indicates the start position of the contig relative to the other four sequences.**

Table 3.2 Summary of the predicted helical, turn and % helicality in representative sequences and MS peptide (Figure 3.15). MS peptide starting from ~aa 80 within those species is likewise predicted.

N-terminus sequence source (Short isoform)	Helix (amino acid range)	Coil (turn) (amino acid range)	Helix %. (H/total)
B. californicum fib1	3-23,35-51,57-77,84-105,112-130	1-2,24-34,52-56,78-83,106-111,131+	103/132 = 78
N. clavipes MaSp1a prec	3-22,32-46,54-75,82-103,109-128	1-2,23-31,47-53,76-81,104-108,129+	101/129 = 78
L. hesperus MaSp1	3-24,33-47,55-74,81-101,107-125	1-2,25-32,48-54,75-80,102-106,126+	98/126 = 78
D. canities MaSp	3-21,36-51,58-77,85-108,113-132	1-2,22-35,52-57,78-84,109-112,133+	99/133 = 74
MS peptide (+~80)	2-13,21-36,45-53	1,14-20,37-44,54+	47/55 = 85

Three other peptides (m/z = 609.77: SSAVSGSEGGGPAR (Figure 3.15), 669.80: GG/N GGGSGSGGGFG--- and 590.00: --GGGSGGGTAR) aligned well with MaSp1 of the same species but were not contiguous (data not shown). It is known that silk protein repeats fall into four categories: poly-alanine (A), GGX (where X denotes one of usually five or so amino acids), GA (glycine and alanine) and GPGX(X)$_n$ (P - proline). Of these, it would seem that the GGC motif (as illustrated by peptides 669.80: GG/N GGGSGSGGGFG--- and 590.00: --GGGSGGGTAR) is the most prevalent, although based upon these alone, it would appear a GGGX motif is also probable. However, given how few representatives of the repeat region have been obtained, this would only be speculation. Unlike the aforementioned spidroins obtained from *A. juruensis*, very few poly-threonine motifs were encountered and according to simple counts (data not shown), threonine appears no more abundant than any other amino acid.

One peptide (m/z = 811.36) RSGAGSGAGEGSGSGAPFL aligned with Spidroin 2 (Bittencourt *et al*., 2010), 87% identity (Figure 3.18). The authors here describe how the number of clones obtained for one silk (MaSp1), (n = 28) far outnumbered that of those for Spidroin 2 (MaSp2), (n = 1). One would expect that likewise, the number of peptide fragments would also correlate to the relative expression levels. Here, the relative number of peptides correlating with MaSp1 significantly outnumber those of MaSp2, supporting those findings. The terminal amino acids, in this example proline, phenylalanine and leucine are usually the least reliably identified, which may explain the discrepancy.

One other peptide (m/z = 1133.60) DDLQSLSESLLSTLSLLRFKR did not match anything previously reported. After amino acid manipulation (the indistinguishable amino acids (I/L, Q/K) were exchanged for their most favourable counterparts; see section 3.2.12) a BLASTp probe of the database suggests this is likely to be a fragment of pyriform silk (egg case silk) and aligned with a 55% identity and an 89% similarity to the *Nephila clavipes* pyriform silk (Figure 3.19). As the silk was taken from an adult female specimen, pyriform silk is an extreme possibility. The inconsistency however, lies in the fact that this sequence does not match anything previously reported for the sister species *Avicularia juruensis* (Bittencourt *et al*., 2010).

After alignments with the entirety of Spidroin 1 (3.1) (Figure 3.20a) and Spidroin 2 of *A. juruensis* (Figure 3.20b) and N-termini of related species (Figure 3.20c), it becomes apparent this sequence is definitely not of MaSp 1 or MaSp2 origin. However, it becomes indeterminate when aligned with tubuliform and cylindriform silks (Figure 3.20d). It is therefore curious why this was not identified from the 34 positive clones in the Bittencourt *et al.* (2010) cDNA library.

>gb|ACF71410.1| spidroin 2 [Avicularia juruensis]
Length=475
Score = 34.6 bits (74)
Expect = 1e-05
Identities = 13/15 (87%)
Positives = 13/15 (87%)
Gaps = 0/15 (0%)

811.36   SGAGSGAGEGSGSGA
            SGAGSG G GSGSGA
A.jur Sp2   SGAGSGSGSGSGSGA

Figure 3.18 Fragment *m/z* = 811.36 aligned with a region of Spidroin 2 from *A. juruensis*.

>gb|ADN39426.1| piriform spidroin [Nephila clavipes
Length=726
Score = 29.9 bits (63)
Expect = 6e-06
Identities = 10/12 (83%)
Positives = 10/12 (83%)
Gaps = 0/12 (0%)

113.60   QSLSESILSTLS
            QSLS SILS LS
N.cla pirs QSLSSSILSSLS

Figure 3.19 Fragment *m/z* = 1133.60 aligned with a region of *Nephila clavipes* pyriform silk.

```
Pyriform spidroin [N. clavipes]   LVQQSGTVSAGQEQSISQSLSSSILSSLSQVVAQRPLPVPAPRPLPAP 250
a                                 -------------DDI-QSLSESILSTLSLLRFQR------------ 21
                                  :.* ****.****:** :   **

Spidroin 1 (3.1)                  GLDASSLGGTIASLVSQISVSNAKLSSSQVFLEALLEVLSGMVQILSYA 960
b                                 --------------------DDLQS---LSESLLSTLS----LLRFK 20
                                  .*.*    : *:**..**     :* :

Spidroin 2                        SPGALSNAISSVVSQVSASNPGLSGCDVLVQALLEIVSALVSILASSSI 450
c                                 --------------------------DDLQSLSESLLSTLS---L 16
                                  : * .: .:*:* *:   :

Peptide 1133                                   DDLQSLSESLLSTL SLLRF K
Peptide 1133                                   QSLSESLLSTL SLLSTSR
TuSp1_Nephila antipodiana     SNAQIISLGLQTTLAPVLSSSGLSSASASARVSSLAQSLASALSTSRGTLSLSTFLNLLS
Tusp1_C                       SNAQIISLGLQTTLAPVLSSSGLSSASASARVSSLAQSLASALSTSRGTLSLSTFLNLLS
Cylindrical silk protein      SNAQIISPSLQTTLAPVLSSSGLSSASASARVGSLAQSLASALSTSRGTLSLSTFLNLLS
d                             ****** .************************.************************
```

Figure 3.20a/b/c/d ClustalW alignment of peptide 1133 alongside the *Nephila clavipes* pyriform spidroin (a), Spidroin 1 (3.1) from Bittencourt *et al*. (2010) (b), Spidroin 2 from Bittencourt *et al*. (2010) (c) and tubuliform and cylindriform silks from *Nephila antipodiana* (d). Due to the lysine residues found in the latter half, alignments favoured the QALLE motif (Challis *et al*., 2006). Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)), blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q.

To improve alignment scores, the 1130.60 was modified to make it more pyriform-like, i.e. once homologous sequences were found (in this case, pyriform silk), the indistinguishable isobaric amino acids (I/L, Q/K) were exchanged for their most favourable counterparts (i.e. most resembling those relating to the database sequences), which resulted in an *E*-value of 0.023. However, the second most homologous sequence was indeed a tubuliform silk, which was similar to findings by Altschul *et al*., 1997, who found that the Spidroin 1 repetitive sequence bears homology to the tubuliform silk protein 1 (BAE54450) from the spider *N. clavata*. Upon manipulation for more favourable amino acids, this hypothesis was tested resulting in an *E*-value of 0.042 (Figure 3.21). As these are within the same order of magnitude, only speculations can be made as to the true origin of this peptide fragment but upon alignment with its nearest BLASTp retrieval and the resulting gaps introduced into the sequence when aligned using ClustalW, pyriform silk would seem the most logical progenitor.

```
gb|HM752576.1|  Agelenopsis aperta clone AE177F tubuliform spidroin 1 mRNA, partial
cds
Length=1629
Score = 19.6 bits (39)
Expect = 0.044,
Identities = 9/11 (82%)
Positives = 10/11 (91%)
Gaps = 0/11 (0%)
Frame = +1


1133        DIQSISESILS  12
            DIQSI+ SILS
A.ape tubs  DIQSIASSILS  258



A.ape       GGIQASPAFPRQEQADIQSIASSILSAGNTATKSKAIEQALSTALASSLAEIVITESGGQ 120
1133        --------------DDIQSISESILST------------LS------------------- 15
                          *****:.****:                **

A.ape       DYSKQITDLNGILSNCFIQTTGVENKRFVNSIQNLIRLLAESAVSETTNSIQIGPYASTS 180
1133        -------------------------------LLRFQR---------------------- 21
                                           *:*:
```

Figure 3.21a BLASTp database result showing similarity of 1133 to a tubuliform silk. ClustalW alignment of 1133 with the corresponding peptide shows a high degree of identity towards the N-terminal region but is subsequently scattered towards the latter half. Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)), blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q.

124

Despite the apparent presence of only one undifferentiated silk gland in *Avicularia avicularia*, similar to the *Avicularia juruensis* used by Bittencourt *et al*. in 2010, it now appears that they produce at least three distinct spidroins. This is in contrast to the previously suggested two MaSp sequences found within *Avicularia avicularia* (Bittencourt *et al*., 2010). It is not uncommon for several different types of silk to be produced by a single silk gland as was reported when MaSp silk was found in the tubuliform glands of *A. diadematus* and *L. hesperus* (Guerette *et al*., 1996; Garb and Hayashi 2005). This is not to say that the silks are distinct and utilised individually as Xu and Lewis (1990) demonstrated with their observation that dragline silk is a dimer composed of MaSp1 and MaSp2. An obvious conundrum here is that if there is indeed a reservoir of silks, how are they separated or indeed aggregated together into the correct quaternary structured macromolecule? A possible explanation was implied via dissections upon *Antrodiaetus unicolor,* the folding-door spider (Mygalomorphae) which revealed a segregation of two distinct secretory products between the proximal and distal hemispheres of the spherical silk glands as well as clustering of these glands into two sets of lateral and two sets of central silk gland bunches (Palmer *et al.*, 1982). As the morphologies of these glands are reminiscent of those within the theraphosids, it is possible that the products, one described as rich in basic amino acids and sulphydryl groups and the other as acidic with significant numbers of C-terminal carboxyl groups, could play a vital role in both segregation and aggregation of these spidroins (Palmer *et al.*, 1982). Additionally, the clusters of glands, which are often pooled for RNA extraction, could be discrete glands whose similar morphologies have led to a misconception of identity.

3.4.3 ACTIN

Four other peptides identified showed homology to actin. These were: *m/z* = 488.00/488.72 (K/AGFAGDDAP/R), 565.77 (GYSFVTTAER), 589.27 (EITALAPSTMsoK) and 1021.84 (LPNGEVITIGNER). These were mapped onto a multiple sequence alignment alongside the actins of the King Baboon tarantula (*Citharischius crawshayi*), another arachnid, the tick (*Hyalomma asiaticum*), a member of a sister group to the Arthropoda, the water bear (*Hypsibius klebelsbergi*) and other eukaryotes including humans (Table 3.3) in the order: 488.00/488.72 (Figure 3.22a), 565.77 (Figure 3.23b), 589.27 (Figure 3.24c) and 1021.84 (Figure 3.25d).

Table 3.3 Multiple sequence alignment candidates

Representative	Phylum	Common name
Bos taurus	Chordata	Domestic cow
Citharischius crawshayi	Arthropoda	King Baboon tarantula
Haliangium ochraceum	Tardigrada	Water bear
*Homo sapiens*1	Chordata	Human
*Homo sapiens*2	Chordata	Human
Hyalomma asiaticum	Arthropoda	Ixodid tick
Macaca mulatta	Chordata	Rhesus macaque
Pongo abelii	Chordata	Sumatran orang-utan

```
                                                              488.72    AGFAGDDAPR
Bos_taurus                 -----------------------MCEE-E-TTALVCDNGSGLCKAGFAGDDAPRAVFPS
Citharischius_crawshayi    -----------------------MCDD-D-IAALVVDNGSGMCKAGFAGDDAPRAVFPS
Haliangium_ochraceum       -----------------------MSDSYVFSSPIIIHPGSDTLQAGLADEEHPGSIFPN
Homo_sapiens1              -----------------------MCDEDE-TTALVCDNGSGLVKAGFAGDDAPRAVFPS
Homo_sapiens2              -----------------------MCDEDV-TTALVCDNGSGLVKAGFAGDDAPRAVFPS
Hyalomma_asiaticum]        -----------------------MCDD-E-VAALVVDNGSGMCKAGFAGDDAPRAVFPS
Hypsibius_klebelsbergi]    ---------------------M-ED-E-VAALVVDNGSGMCKAGFAGDDAPRAVFPS
Macaca_mulatta             MRPAPTPLCLSPGLRPKSCEAAPAMCEEED-STALVCDNGSGLCKAGFAGDDAPRAVFPS
Pongo_abelii               -----------------------MCEE-E-TTALVCDNGSGLCKAGFAGDDAPRAVFPS
                            *  :.        :.::  .  **.   :**:*.::  *  ::**.
```

Figure 3.22a ClustalW alignment of 488.72 with the actin from nine representative species. Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)), blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q. Bold residues show identity between peptide fragment and representative species.

```
                                                                      565.77        GYSFVTTAER
Bos_taurus               GIVLDSGDGVTHNVPIYEGYALPHAIMRLDLAGRDLTDYLMKILTERGYSFVTTAEREIV
Citharischius_crawshayi  GIVLDSGDSVSHTVPIYEGYALPHAILRLDLAGRDLTDNLMKILTERGYSFVTTAEREIV
Haliangium_ochraceum     GLVVSLGDFVSYVAPVHRGAIVDAGLTFLEPDGRSITEYLSRLLLERGHVFTSPEALRLV
Homo_sapiens1            --------------VPIYEGYALPHAIMRLDLAGRDLTDYLMKILTERGYSFVTTAEREIV
Homo_sapiens2            GIVLDSGDGVTHNVPIYEGYALPHAIMRLDLAGRDLTDYLMKILTERGYSFVTTAEREIV
Hyalomma_asiaticum]      GIVLDSGDGVSHTVPIYEGYALPHAILRLDLAGRDLTDYLMKILTERGYSFTTTADREIL
Hypsibius_klebelsbergi]  GIVLDSGDGVSHTVPIYEGYALPHAILRLDLAGRDLTDYLMKILTERGYSFVTTAEREIV
Macaca_mulatta          GIVLDSGDGVTHNVPIYEGYALPHAIMRLDLAGRDLTDYLMKILTERGYSFVTTAEREIV
Pongo_abelii            GIVLDSGDGVTHNVPIYEGYALPHAIMRLDLAGRDLTDYLMKILTERGYSFVTTAEREIV
                         .*::.*   :   .:  *: **.:*: * ::* ***: *.:.    .::
```

Figure 3.23b ClustalW alignment of 565.77 with the actin from nine representative species. Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)), blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q. Bold residues show identity between peptide fragment and representative species.

```
                                              589.27 EITALAPSTMK
Bos_taurus                ESAGIHETTYNSIMKCDIDIRKDLYANNVLSGGTTMYPGIADRMQKEITALAPSTMKIKI
Citharischius_crawshayi   ------------------------------------------------------------
Haliangium_ochraceum      ESPGLTDAVCNAIMKCDPSLQAELFGNIVVTGGGSLFPGLSERLQRELEQRAPAEAPVHL
Homo_sapiens1             ESAGIHETTYNSIMKCDIDIRKDLYANNVMSGGTTMYPGIADRMQKEITALAPSTMKIKI
Homo_sapiens2             ESAGIHETTYNSIMKCDIDIRKDLYANNVMSGGTTMYPGIADRMQKEITALAPSTMKIKI
Hyalomma_asiaticum]       ESCGIHETTYNSIMKCDVDIRKDLYANTVLSGGTTMYPGIADRMQKEITALAPSTMKIKI
Hypsibius_klebelsbergi]   ESCGIHETTYNSIMKCDIDIRKDLYANTVLSGGTTMYPGIADRMQKEITALAPSTMKIKI
Macaca_mulatta            ESAGIHETTYNSIMKCDIDIRKDLYANNVLSGGTTMYPGIADRMQKEITALAPSTMKIKI
Pongo_abelii              ESAGIHETTYNSIMKCDIDIRKDLYANNVLSGGTTMYPGIADRMQKEITALAPSTMKIKI
```

Figure 3.24c ClustalW alignment of 589.27 with the actin from nine representative species. Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)), blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q. Bold residues show identity between peptide fragment and representative species.

```
                                       1021.84      LPNGEVITIGNER
Bos_taurus              RDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVITIGNERFRCPETLFQPSFIGM
Citharischius_crawshayi RDIKEKLCYVALDFEQEMAIAASSSTVEKSYESPDGQVITIGNERFRCPETLF-------
Haliangium_ochraceum    RDIKETLCYVADDVAKEAARNA--DSVEATYLLPNGETLVLGNERFRCPEVLFHPDLLGW
Homo_sapiens1           RDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVITIGNERFRCPETLFQPSFIGM
Homo_sapiens2           RDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVITIGNERFRCPETLFQPSFIGM
Hyalomma_asiaticum]     RDIKEKLCYVALDFEQEMATAASSSSPEKSYELPDGQVITIGNERFRCPEALFQPSFLGM
Hypsibius_klebelsbergi] RDIKEKLCYVALDFEQEMATAAASSSLEKSYELPDGQVITIGNERFRCPEALFQPSFIGM
Macaca_mulatta          RDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVITIGNERFRCPETLFQPSFIGM
Pongo_abelii            RDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVITIGNERFRCPETLFQPSFIGM

                        ****.***** *. :* * *  .: * :*  *:*:.:.:**********.*
```

Figure 3.25d ClustalW alignment of 1021.84 with the actin from nine representative species. Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)), blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q. In this instance, there are no sequences with complete identity.

Each peptide pulls up specifically actin from BLASTp searches albeit with minor ambiguities in the sequences. One would expect the closest similarity between these sequences to be with the King Baboon tarantula actin found on the database or after BLASTp analysis, at least another arachnid; but here the database fails as this actin is incomplete resulting in the most similar match (amongst those used for msa) to be *Haliangium ochraceum* (Figure 3.26). Three of the four peptides are in good sequence agreement with all (shown) versions of actin, the fourth (1021), has two discrepant residues resulting in it falling in line with the unusual *H. ochraceum*. However, these potential variations would not appear to allow a differentiation of the source of the actin between spiders and mammals, for example. A multiple sequence alignment with the most 'complete' spider-derived actin, a grass spider (*Agelena silvatica*) provided an approximate order to the peptides along the actin backbone (Khaitlina 2001). This crude contig:

AGFAGDDAPR_**GYSFVTTAER_LPNGEVITIGNER_EITALAPSTMK**

was used in a BLASTp search against the database to identify the species utilising an actin with the nearest homology, which pulled up the snow crab (*Chionoecetes opilio*, data not shown). This discrepancy, (an arthropod nevertheless, but not an arachnid) could be explained in a number of ways. It is indeed possible the *de novo* peptide is simply incorrect, which is entirely possible given that there is only about 0.98 Da between N-D and Q-E but this seems implausible, as centrally distributed amino acids within a *de novo* sequence tend to be the most reliable. Database searches do occasionally yield E residues rather than the apparently well-conserved Q (peptide 1021).

```
                1021.84           LPNGEVITIGNER
Bos_taurus                        LPDGQVITIGNER 10/13
Citharischius_crawshayi           SPDGQVITIGNER 10/13
Haliangium_ochraceum              LPNGETLVLGNER 11/13
Homo_sapiens1                     LPDGQVITIGNER 10/13
Homo_sapiens2                     LPDGQVITIGNER 10/13
Hyalomma_asiaticum]               LPDGQVITIGNER 10/13
Hypsibius_klebelsbergi]           LPDGQVITIGNER 10/13
Macaca_mulatta                    LPDGQVITIGNER 10/13
Pongo_abelii                      LPDGQVITIGNER 10/13
```

Figure 3.26 Eye alignment of 1021.84 with the actin from nine representative species. Amino acids are indicated by one-letter abbreviations. Colours indicate general amino acid properties: red - Small aa (small + hydrophobic (incl. aromatic -Y)), blue - Acidic aa, magenta - Basic aa, green - Hydroxyl + Amine + Basic – Q. Total scores of identity are shown at the end of each row. Closest species on a simple identity count is *Haliangium ochraceum* with 11/13.

These however, tend to be more unusual organisms (i.e. single-celled foraminifera and d-proteobacteria – data not shown), although this could be the result of poor quality DNA sequencing. The most likely explanation is the deamidation of asparagine (N) to aspartic acid (D) which is a naturally occurring post-translational modification of proteins (Teshima 2000) resulting in the conversion of an asparagine residue to an isoaspartate and aspartate. Glutamine deamidation is likewise said to occur albeit at a much slower rate. Equally possible is that this deamidation occurred post-sampling as an artefact during trypsin digestion so the second residue E, which differs from the conserved residue Q, may really be glutamine that has been deamidated to glutamic acid. Assuming it is indeed an artefact, the sequence still differs significantly from human actin so candidacy as a contaminant should be disregarded. Previous studies have not identified actin amongst silk and as this is the first time MS has been utilised to sequence spider silk in this way, this possibility should not be ruled out. Likewise however, it is equally likely to be a contaminant from the organism itself (faecal matter, prey detritus, urticating hairs…) but this would be problematic to investigate due to the necessity for sufficiently large amounts of silk required to be resolved on an SDS-PAGE gel. Spooling silk from aranaeid spiders has indeed been performed on many occasions and this is certainly a possibility for future acquisition of mygalomorph spidroins but the author is unaware this has been attempted and would undeniably pose some technical challenges.

3.5 CONCLUSIONS

This chapter has demonstrated how ESI tandem mass spectrometry can be utilised to selectively target specific sequence fragments of a large macromolecular protein, in this instance silk, which would otherwise be virtually unobtainable by other conventional means. The aim of this study was to ideally acquire large enough fragments of the aforementioned Spidroin 1 and Spidroin 2 isoforms to identify sequence discrepancies and homologies to previous work. As MS is effectively a sequencing of random fragments, a choice of domain preference (in this case, the N-terminal domain) could not be made. The present study confirms previous findings and contributes additional evidence that Spidroin 1 and Spidroin 2 are both utilised in everyday silks of the *Avicularia* spp., moreover in agreement with the suggested ratios (from the repetitive domains, MaSp1: n = 28; MaSp2: n = 1) (Bittencourt *et al.*, 2010) implied therein.

However, these data have shown that the vast majority of useable peptides did actually originate from within the N-terminal domain, so many in fact that a contig was able to be constructed accounting for over 40% of the predicted size of this region. In addition to this, a peptide not corresponding to any silk previously sequenced from a mygalomorph was identified, which bore a striking similarity to a pyriform silk. Pyriform silk has been found to be the main constituent of attachment discs with which a spider anchors the mechanically active silks (e.g. dragline silks) to surfaces. This is somewhat surprising as there would appear to be no necessity for a theraphosid spider to actively attach silk to the substrate, or rather, not with sufficient adhesiveness to warrant specialised glue. Granted, this representative is indeed an arboreal

species but personal observations have noted the electrostatic interaction of the silk to the frame on which it is deposited to be more than sufficient for future deposition and reinforcement. However, field observations and personal correspondences have suggested that of all species, *Avicularia* spp. are more prone to being 'clumsy' and falling compared with considerably larger but equally arboreal species such as *Poecilotheria* spp. This would motivate the necessity of a more adhesive silken frame with which to rely on for support. Despite this, spiders like most organisms, tend to walk forwards and so any silk is deposited astern, offering no tarsal support. Additionally, the Mygalomorphae do not anchor their silk. This is in contrast to the araneid spiders, which constantly deposit silk anchor points that the spider is suspended from should it fall.

The findings of this study imply that MS can be utilised for future sequencing of regions of silk. Far more importantly, these data suggest that the N-terminus is sequenced preferentially over the other domains. Future work in this instance should be to utilise this possibility to acquire further N-terminal domains, which up until now have been extremely rarely characterised due to the methods with which silk cDNA is currently being sequenced. cDNA libraries, albeit the most conventional approach, are by far the least likely to yield complete silks due to their long transcript length.

The current study has only examined however the collective silks of one individual of a single species. Naturally, a larger sample size (perhaps a representative from all the 12 subfamilies) and range of species would be necessary for future evaluations of theraphosid spidroins. The findings of this report are of course subject to the limitations of the techniques used, namely

the difficulty in distinguishing between the isobaric residues like isoleucine and leucine. Future work should encourage the use of more precise analytical methods such as Edman degradation or the use of more sophisticated MS. The chance of acquiring solely the N-terminus is highly improbable statistically (~3% chance) and one would predict the ratio of peptides generated from the N-terminal domain, repetitive domain and the C-terminal domain to correspond to the relative space each occupies within the whole. These data not only demonstrate that the N-terminal domain was pulled out in a far greater than expected quantity, but that also, the quality of sequencing of said peptides was far more reliable.

4 ANALYSIS OF RNA ISOLATED FROM SILK GLANDS AND SURROUNDING TISSUE IN *GRAMMOSTOLA ROSEA* BY cDNA SEQUENCING

4.1 INTRODUCTION

The tarantula *Grammostola rosea* is perhaps the most common species of tarantula kept in captivity. It is exported in large numbers from the Atacama Desert region of Chile from where it originates to be sold in the pet trade. The spiders generally have a life expectancy of 15 - 20 years (for females, compared to the males that live just a few years post maturity). To date, mygalomorph/theraphosid studies have mostly focused on the venoms/toxins produced by the more medically significant representatives with only very recently, two on mygalomorph silks (Gatesy *et al*., 2001; Bittencourt *et al*., 2010). Araneoid spiders produce up to seven types of silk (Foelix 1996) and over the last decade, cDNAs from a large number of C-terminal domains from representatives of each class have been identified (Beckwitt *et al*., 1994; Beckwitt *et al*., 1998; Colgin and Lewis 1998; Hayashi and Lewis 1998; Hu *et al*., 2005; Hayashi *et al*., 2004). All known silk proteins can be divided into three domains: a highly conserved N-terminal domain, a highly repetitive and (typically) hydrophobic middle domain, usually composing over 80% of the

fibre and a highly conserved but unique C-terminal domain (see Chapter 3). Due to the large body size of an adult tarantula and indistinguishable internal morphology, which upon dissection would result in a large amount of RNA degradation, a total cDNA library of an adult female *G. rosea* was constructed. Sequencing of randomly selected cDNAs has been used as a tool to study relative levels of gene expression. In this study, multiple cDNA clones are described from the *G. rosea* library, highlighting abundant ESTs and *de novo* sequences. The aims of this study were to build the first cDNA library of an adult female tarantula suitable for the investigation of expressed sequence tags (ESTs) and to contribute towards the vastly underrepresented spider EST resource currently available. In addition, this resource was used to uncover *de novo* silk sequences as well as potentially contributory structural components, chaperone proteins and new leads as to how silk is produced and engineered within the spider.

4.2 MATERIALS AND METHODS

4.2.1 MATERIALS AND RNA ISOLATION

A single adult female *G. rosea* was obtained from a captive bred source. Attempts to obtain voucher specimens were undertaken but obtaining export permits and satisfactory identification down to species level was considered impractical. The individual was submerged in liquid nitrogen and the abdomen snapped off, ground up using a pestle and mortar and subjected to the TRIzol® reagent protocol (Invitrogen) to isolate total RNA. The concentration and purity of RNA was examined using a Thermo Scientific Nanodrop 1000 UV-Vis Microfluid Spectrophotometer and found to be in excess of 3 µg/µl in all trials conducted. This value was confirmed by conducting serial dilutions until the Nanodrop reading was within its optimal range.

4.2.2 cDNA SYNTHESIS

mRNA was isolated from the total RNA by means of Dynabeads® using the recommended protocol. Final mRNA concentration was found to be typically in excess of 250 µg/µl usually eluted from the beads with 20 µl of distilled de-ionised water (ddH$_2$O). This was used to synthesise second-strand cDNA using DNA polymerase I. Synthesis of sscDNA was conducted using the SD method (see: Methods) preferentially over the SuperScript® III Reverse Transcriptase protocol provided with the enzyme. Synthesis of dscDNA was conducted as per section 2.10.3.2.

4.2.3 CONSTRUCTION OF THE CDNA LIBRARY

The cDNA library was constructed using the pGEM®-T Vector Systems by ligating the A-tailed dscDNA into the vector and transforming those constructs into XL10-Gold®* ultracompetent cells. Colonies containing cDNA inserts within the LacZα MCS were identified using blue-white screening. A colony-pick PCR was performed using Phusion® DNA Polymerase and the applicable primers for pGEMT-Easy (T7F: 5'-TAA-TAC-GAC-TCA-CTA-TAG-GG-3' and SP6:5'-ATT-TAG-GTG-ACA-CTA-TAG-AA-3') using a Tm of 45°C. The PCR conditions for this were as follows: 95°C for 3 min, 35 cycles (95°C for 20 s, 45°C for 20 s and 72°C for 180 s) and finally 72°C for 300 s. This extended PCR was excessive and was designed to capture all inserts rather than selecting for those of a restricted size. Phusion® DNA Polymerase has an extension capability of approximately 1 kb in 20 s and therefore an extension time of 180 s are predicted to identify cDNAs of up to 9 kb. The products were analysed on a 1% agarose gel to identify conclusive positive transformants and to determine the size and frequency of cDNAs. Favourable transformants were grown at 37°C overnight in 5 ml Lysogeny Broth (LB) with ampicillin added to a final concentration of 100 μg/ml. Minipreps were performed of each of these cultures using the Wizard® Plus Minipreps DNA Purification System, which seemed to give higher yields and better quality DNA than the QIAprep Spin MiniPrep Kits.

4.2.4 SEQUENCING OF THE CDNA LIBRARY

PCR products were sequenced on an ABI 3730XL machine by, Macrogen Inc. (Seoul, Korea; later Amsterdam, the Netherlands) or by GeneService (now Source BioScience LifeSciences). Sequences came back as a compressed Zip file containing both the AB.1 file (Figure 4.1) (Applied Biosystem's Sequencing Analysis software) and a FASTA format notepad file.

A number in excess of 150 clones was randomly selected from the primary library and inserts sequenced from 5' end using the SP6 primer (5'-ATT-TAG-GTG-ACA-CTA-TAG-3') compatible with the pGEMT-Easy vector (Figure 4.2). Superfluous vector sequence was excised by eye using the adjoining EcoRI sites bordering the inserts or with VecScreen (NCBI) if adjoining sequences were missing due to substantial length or ambiguities. Assembly and sequence editing were performed with the CAP3 Sequence Assembly Program (Huang and Madan 1999) http://pbil.univ-lyon1.fr/cap3.php and the ExPASy translate tool (http://web.expasy.org/translate/). BLAST (Altschul *et al*., 1997) was used for the similarity searches. PHRED/PHRAP/CONSED software (http://www.phrap.org) was employed to assist with editing and sequence assembly. ESTs were subjected to downstream analyses only if they were < 200 bases with a phred quality value < 20.

Figure 4.1 Example output as displayed in BioEdit (Hall 1999) of a selected sequenced library clone and annotations thereof.

XmnI 2009
ScaI 1890
NaeI 2707
f1 ori
Amp^r
pGEM®-T Easy Vector (3015bp)
lacZ
ori

T7 ↓	1 start
ApaI	14
AatII	20
SphI	26
BstZI	31
NcoI	37
BstZI	43
NotI	43
SacII	49
EcoRI	52

SpeI	64
EcoRI	70
NotI	77
BstZI	77
PstI	88
SalI	90
NdeI	97
SacI	109
BstXI	118
NsiI	127
↑ SP6	141

Sequence and Multi-Cloning Site of the pGEM®-T Easy Vector

The pGEM®-T Easy Vector has been linearized at base 60 with EcoRV and a T added to both 3′-ends. The EcoRV site will not be recovered upon ligation of the vector and insert.

T7 Transcription Start

5′...TGTAA TACGA CTCAC TATAG GGCGA ATTGG GCCCG ACGTC GCATG CTCCC GGCCG CCATG
3′...ACATT ATGCT GAGTG ATATC CCGCT TAACC CGGGC TGCAG CGTAC GAGGG CCGGC GGTAC

T7 Promoter   ApaI AatII SphI BstZI NcoI

GCGGC CGCGG GAATT CGATT3′(cloned insert) ATCAC TAGTG AATTC GCGGC CGCCT GCAGG TCGAC
CGCCG GCGCC CTTAA GCTA   3′TTAGTG ATCAC TTAAG CGCCG GCGGA CGTCC AGCTG

NotI SacII EcoRI   Spel EcoRI NotI PstI SalI
BstZI   BstZI

SP6 Transcription Start

CATAT GGGA GAGCT CCCAA CGCGT TGGAT GCATA GCTTG AGTAT TCTAT AGTGT CACCT AAAT...3′
GTATA CCCT CTCGA GGGTT GCGCA ACCTA CGTAT CGAAC TCATA AGATA TCACA GTGGA TTTA...5′

NdeI SacI BstXI NsiI   SP6 Promoter

Figure 4.2 pGEM®-T Easy vector map and sequence reference points with its respective Multiple-Cloning Site.

4.2.5 PHRED QUALITY SCORES

Phred quality scores Q are defined as a property that is logarithmically related to the base-calling error probabilities P (Ewing and Green 1998).

$$Q = -10 \log_{10} P$$

or

$$P = \frac{-Q}{10}$$

Table 4.1 Phred quality scores

Phred quality score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

For example, a Phred Quality Score of 20 (as used) would indicate a relative certainty of said base being accurate as 0.99.

BLASTX (Altschul *et al.*, 1997) was used to perform similarity searches of the edited sequences against the GenBank non-redundant protein sequences (nr) database at the National Center for Biotechnology Information (NCBI).

4.2.6 EST BIOINFORMATIC ANALYSIS OF RETURNED SEQUENCES

Returned sequences were compared to the GenBank databases using different sets of parameters. BLASTn (entire database) was chosen to identify immediate and global EST families comparing nucleotide with nucleotide. A similar search was also performed but changing the Organism parameter to "arthropods (taxid:6656)." Both searches were performed on the Nucleotide collection (nr/nt) database, optimising for highly similar sequences (megablast) and adjusting the filters to allow low complexity regions and removing masks (Figure 4.3).

The two remaining BLAST searches were tBLASTx searches, as translated queries vs. translated database sequences (tBLASTx) are particularly useful for identifying novel genes in nucleotide query sequences that are error prone (NCBI). Again the searches were performed on the Nucleotide collection (nr/nt) database but entering under the Organism parameter "arthropods (taxid:6656)" (as per before) and narrowing the search further to "spiders (taxid:6893)." Again, there were no filters on the low complexity regions and the word size this time was reduced to 2. The E-value scores lower than the recommended (NCBI) values of 10^{-5} were considered to be significant although limitations of this approach are discussed later. These BLAST searches were also used to relate relativity and function to other taxa and individuals throughout the collected databases.

Figure 4.3 BLASTn search showing all parameters as utilised.

4.3 RESULTS

Total RNA was isolated from the abdomen of an adult female *G. rosea* and used to construct a cDNA expression library. The total RNA was found to have a concentration far in excess of the threshold range for the Micro-Volume Full-Spectrum Fluorospectrometer (NanoDrop) but yielded suitably clean RNA once sufficiently serially diluted (Figure 4.4). Over several trials of cDNA library construction, total RNA integrity was analysed by formaldehyde denaturation agarose gel electrophoresis and suitably clear bands of the expected 28S, 18S and 5S ribosomal RNA could be observed with a clear (by eye) 2:1 ratio of intensities, verifying RNA integrity (data not shown). The majority of corresponding cDNA observed via PCR and subsequently from sequencing results were found to visually average approximately 300 bp in length although ranged from ~50 base pairs to 1500 base pairs in length (Table 4.2). The latter is a surprising result as sequencing usually stopped at around 1 kb. Examples of the PCR gels for the library are shown in Figure 4.5.
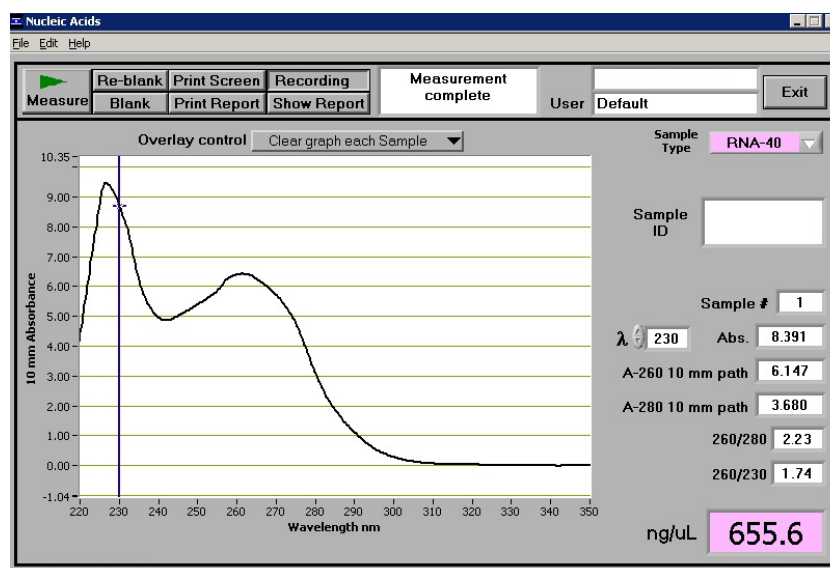
Figure 4.4 Typical output as expected on a micro-volume full-spectrum fluorospectrometer (NanoDrop) of an RNA sample diluted 1/10 post Tri-Reagent protocol.

Table 4.2 Descriptive statistics table of the EST library detailing standard distributions and fit of the data.

			Statistic	Std. Error
EST Lengths	Mean		271.08	20.78
	95% Confidence Interval for Mean	Lower Bound	229.76	
		Upper Bound	312.40	
	5% Trimmed Mean		245.98	
	Median		223.50	
	Variance		37145.04	
	Std. Deviation		192.73	
	Minimum		61	
	Maximum		1376	
	Range		1315	
	Interquartile Range		161.25	
	Skewness		3.03	.260
	Kurtosis		12.98	.514

(a)
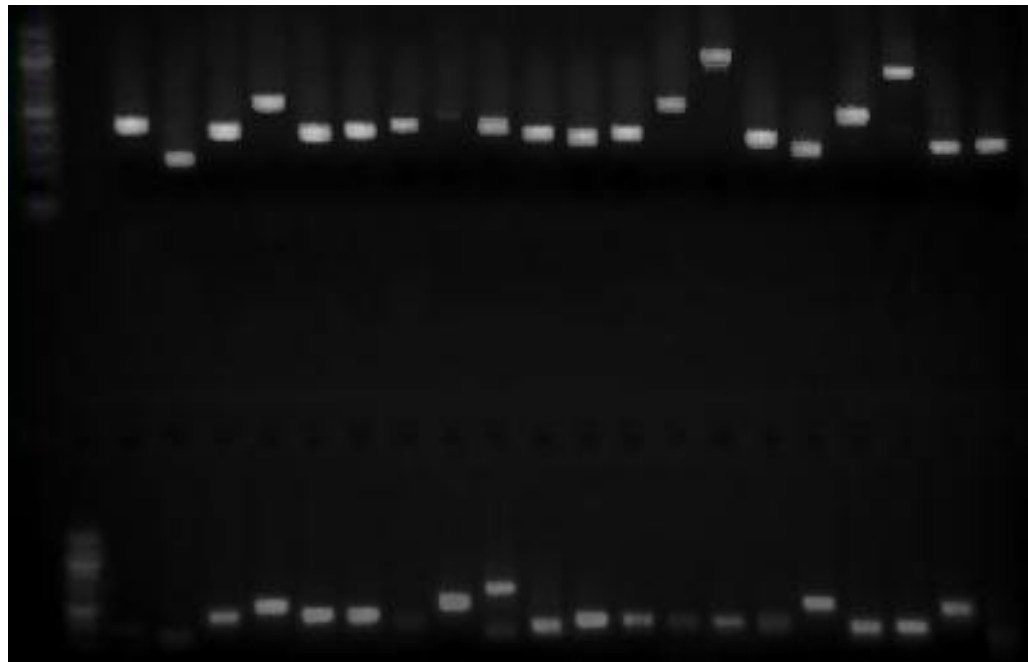
(b)

Figure 4.5 PCR amplification of randomly selected clones from the EST cDNA library within pGEMT-Easy. (a) Primary library with a fresh batch of T4-DNA ligase alongside DNA ladder (with respective sizes shown). (b) Randomly selected clones from 'secondary' library.

149

4.3.1 Library titre

The primary library titre was calculated as 2.5 x 10^4 cfu/ml and 3 x 10^8 cfu/ml after and was found to be approximately 84% recombinant as assessed by counting the number of blue/white colonies on a division of the plate.

Over 150 transformants were chosen and minipreps performed. Of those, approximately 70 arrived with unsatisfactory sequencing reads and 86 with sufficient quality (Figure 4.1). Raw data were manipulated to remove vector sequences and the invariably poor terminal ends containing ambiguities (signified by N's) using a combination of search functions in MS Word, or online programs such as VecScreen (www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html).

The sequences ranged from 61 bp to 1376 bp with a mean length of 271 bp (Table 4.2). The median length was found to be 223 bp, lower than the mean most likely due to the few but significantly high outliers (as illustrated by a box and whisker plot (Figure 4.6)). To test for normality of distribution, a normal Q-Q plot of EST lengths was drawn using SPSS (Figure 4.7) and both Kolmogorov-Smirnov and Shapiro-Wilk tests were conducted (Table 4.3). Both found the distribution to be highly skewed ($p < .001$) as shown by the Shapiro-Wilk test and as illustrated by a categorised histogram (Figure 4.8).

Table 4.3 Kolmogorov-Smirnov and Shapiro-Wilk tests for normality with their respective degrees of statistical significance

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
EST lengths	.198	86	.000	.719	86	.000

a. Lilliefors Significance Correction

150
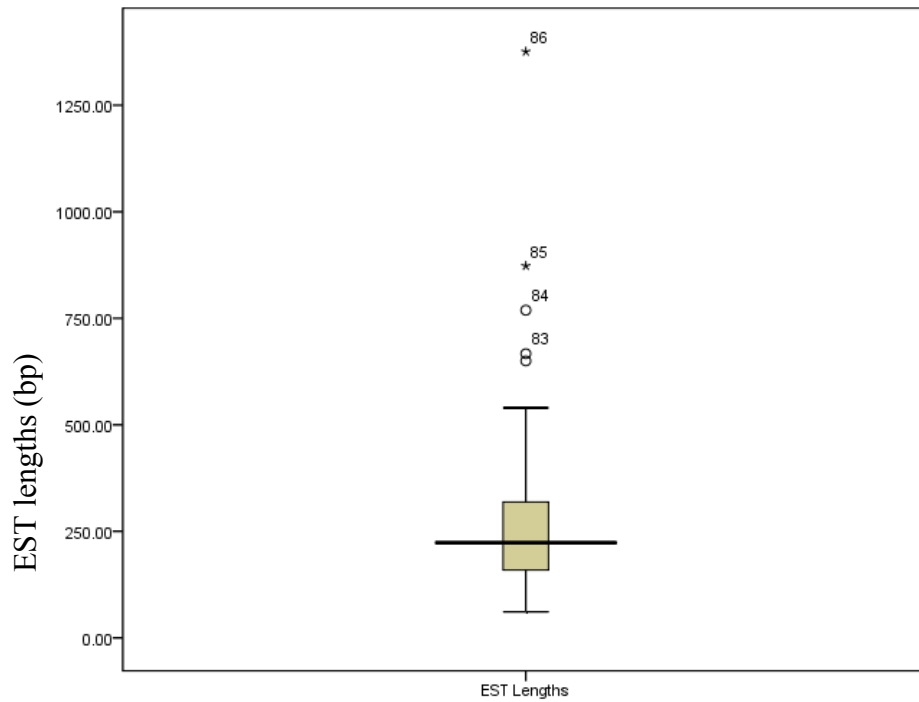
Figure 4.6 Box and whisker plot showing the median and the distribution of EST lengths from a random sequencing of library clones from *G. rosea*.
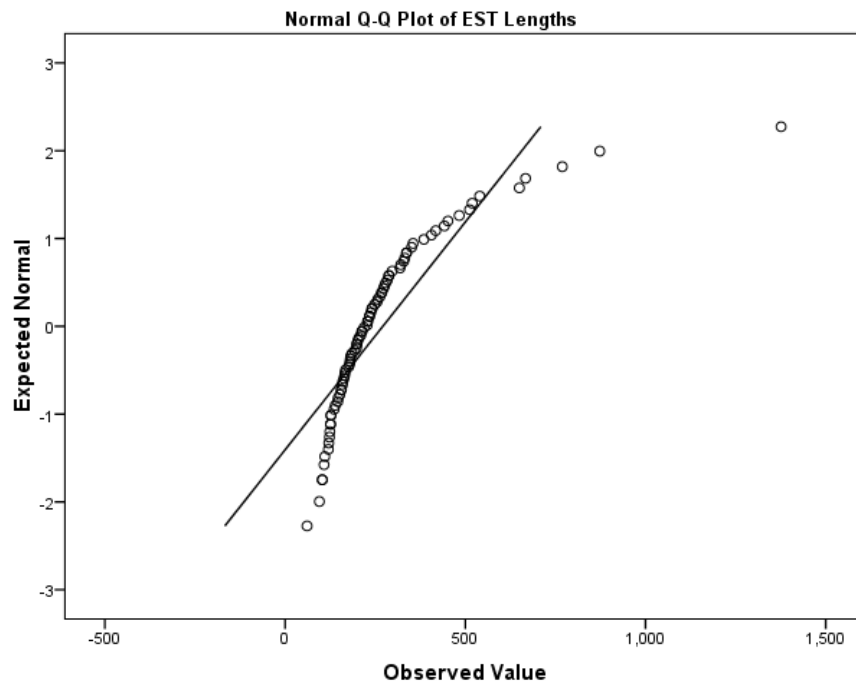
Figure 4.7 Normal Q-Q plot detailing the distribution of EST length values with respect to an expected normal distribution line of best fit.
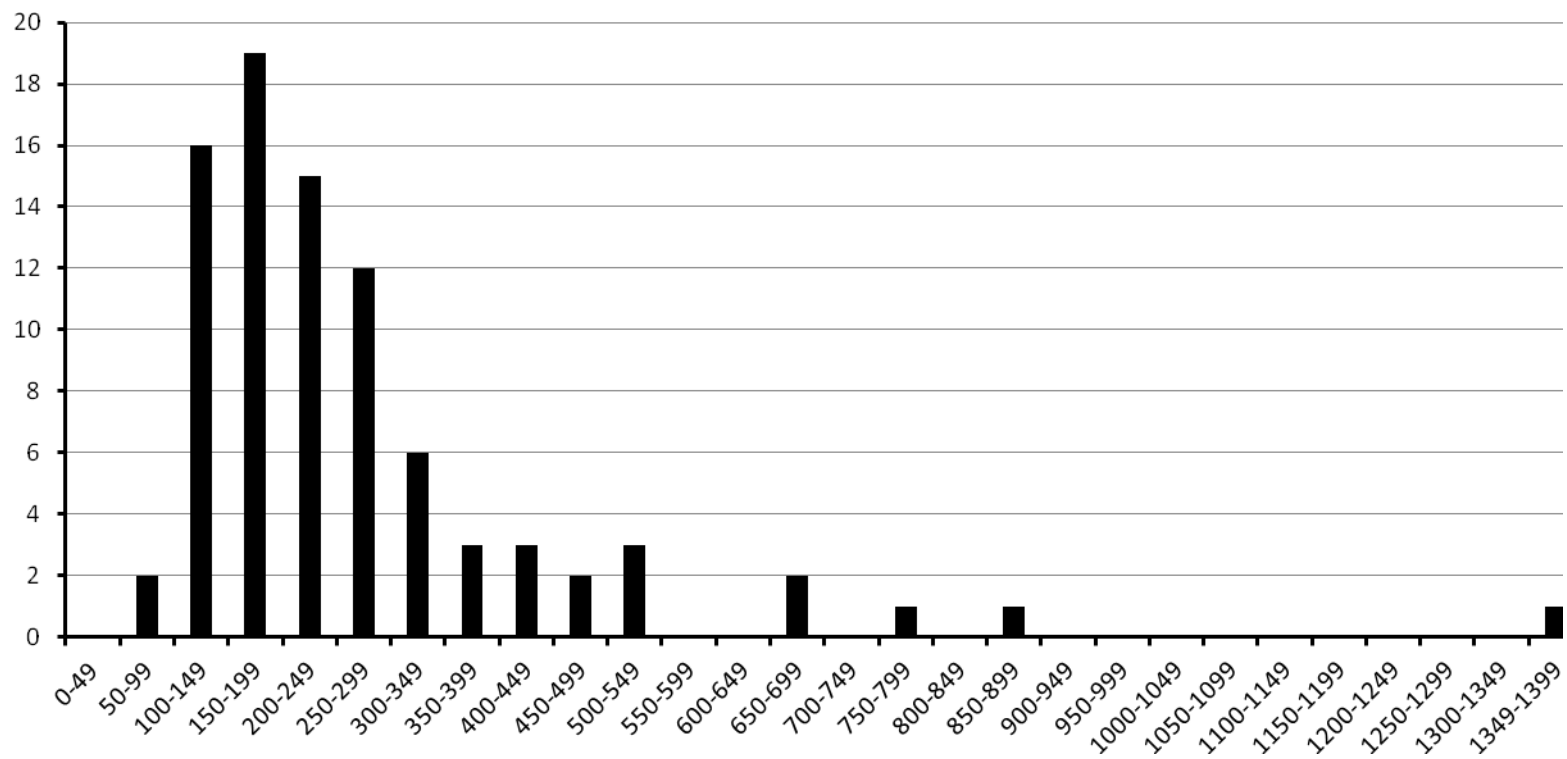
Figure 4.8 Histogram of the length distribution of 86 ESTs of *G. rosea*.
X-axis showing the length of the ESTs with the Y-axis showing the relative frequency. Skewed data could indicate typical transcript lengths used within this species, or be indicative of optimal lengths for insert ligations.

GenBank searches were used to determine the most likely progenitor genes from whence the RNAs had derived and sequences transformed to extrapolate plausible ORFs. Of those 86 high-quality ESTs encompassing a range of rRNA and mitochondrial RNA, BLAST searches identified 45 significant gene sequences (E-values $< 10^{-5}$), 12 contigs (deriving from 25 sequences in total, not all yielding definitive database results) and 64 singletons (Table 4.4). The high relative amounts of rRNA sequences are typical of this method as total RNA is utilised. Unique sequences are pending for deposition into the GenBank database.

Table 4.4 Cluster analysis summary

Description	Number
Clones sequenced	> 100
Eligible sequences	86
Total EST valid length (nt)	23313
Average ESTs length (nt)	271
Number of contigs	12
Number of singletons	64
Unique genes	35
GC (%)	39.3

Nt: nucleotides.

4.3.2 BLAST ANALYSES AND SEQUENCE ORGANISATION

BLAST analysis of these remaining (contig-optimised) unique genes (35) were subjected to the aforementioned BLASTn (entire database), BLASTn (arthropods), tBLASTx (arthropods) and tBLASTx (arachnids) program, hereon in simply referred to as 'BLASTnN' (eNtire), 'BLASTnR' (aRthropods), 'tBLASTxR' (aRthropods) and 'tBLASTxS' (Spiders) respectively, against the nr databases (Table 4.5). These 'partnered' ESTs (i.e. they were found to match an EST in at least one of the database searches to a significance of E-value $< 10^{-5}$) were assigned into three categories: probable genes, putative genes and unknown genes (Table 4.6). Probable genes were those that shared a significant homology with known genes, often with E-values $> 10^{-10}$; although some were found with values $E = 10^{-5} - 10^{-9}$ (some of which were included as these were found to be part of contigs). This dataset included 19 unique genes (54.3% of matched sequences). Putative genes were those that shared a high degree of similarity to those genes (or rather pulled up those genes in the BLAST searches) that were referred to in the EST databases as either 'putative' or 'like'. This dataset included 7 sequences (20.0% of matched sequences). Included in this group are the transposons/mariners. The final group included 9 'unknown' genes (25.7% of matched sequences), not to be confused with t he sequences that resulted in E-values $> 10^{-5}$. All of these brought about 'hypothetical protein' results and despite some very significant homologies with other sequences (as the name suggests), definitive functions of these ESTs can only be hypothesised (explained later).

Table 4.5 List of identified ESTs from female *G. rosea*. The letter 'y' denotes a positive hit in that particular search.

Seq ID	bn.all	bn.r	tx.r	tx.s	Putative Function	Species matched	Common name	*E*-value	Valid ID
GRF-5	-	-	y	y	Chitinase	*Araneus ventricosus*	Ghost spider	1.00E-22	gb|AY120879.1|
GRF-6	-	y	y	y	Ribosomal protein L8	*Aplysia californica*	California sea slug	9.00E-39	gb|AF481057.1|
GRF-7	y	-	y	y	Sphingomyelinase D - like protein	*Loxosceles arizonica*	Brown recluse spider	3.00E-48	gb|AF512954.1|
GRF-8	-	-	y	-	Thrombin inhibitor haemalin	*Haemaphysalis longicornis*	New Zealand cattle tick	2.00E-11	dbj|AB440203.1|
GRF-9	-	-	y	y	Putative serine proteinase inhibitor	*Latrodectus hesperus*	Western black widow spider	4.00E-12	gb|HQ005987.1|
GRF-14	-	-	y	-	Phospholipase B-like 2-like	*Acyrthosiphon pisum*	Pea aphid	2.00E-05	ref|XM_001948827.2|
GRF-17	y	y	y	y	16S ribosomal RNA gene	*Ornithoctonus huwena*	Chinese earth tiger tarantula	7.00E-40	gb|EU979519.1|
GRF-20	y	y	-	-	Hypothetical protein	*'Citharischius crawshayi'*	King baboon tarantula	3.00E-12	gb|GU170900.1|
GRF-21	-	-	y	y	Hypothetical protein	*Latrodectus hesperus*	Western black widow spider	8.00E-13	gb|HQ006016.1|
GRF-24	-	-	y	y	18S ribosomal RNA	*Spinileberis quadriaculeata*	~Crustacean~	7.00E-07	dbj|AB076638.1|
GRF-30	-	-	y	-	18S ribosomal RNA	*Hyperia galba*	~Crustacean~	2.00E-07	gb|DQ378046.1|
GRF-31	-	-	y	y	18S ribosomal RNA	*Oncodamus bidens*	~Spider~	2.00E-08	gb|EU003360.1|
GRF-32	y	y	y	y	18S ribosomal RNA	*Hyperia galba*	~Crustacean~	2.00E-157	gb|DQ378046.1|
GRF-33	y	y	y	y	18S small ribosomal subunit	Uncultured marine eukaryote	~~~	8.00E-34	gb|GU370021.1|
GRF-34	y	y	y	y	18S ribosomal RNA	*Hyperia galba*	~Crustacean~	0	gb|DQ378046.1|
GRF-40	-	-	y	-	Mariner transposase pseudogene	*Andrena erigenia*	Mining Bee	2.00E-05	gb|U91345.1|
GRF-41	-	-	y	y	Flagelliform silk protein	*Nephila clavipes*	Golden orb-web spider	5.00E-11	gb|AF218621.1|
GRF-43	-	-	y	y	Sphingomyelinase D-like protein	*Loxosceles arizonica*	Brown recluse spider	1.00E-24	gb|AF512954.1|
GRF-45	y	y	y	-	Transfer RNAs (K/N)	*Hyposoter didymator*	Ichneumonid wasp	3.00E-26	gb|GQ923582.1|
GRF-46	-	-	y	-	Transposon mariner-like element	*Helicoverpa armigera*	Cotton bollworm	2.00E-11	gb|HM807611.1|
GRF-49	-	-	y	-	Hypothetical protein	*Ixodes scapularis*	Deer tick	3.00E-12	ref|XM_002404400.1|
GRF-50	-	y	y	-	Lysozyme	*Macrobrachium rosenbergii*	Giant river prawn	3.00E-12	gb|AY257549.2|
GRF-56	y	y	-	-	16S ribosomal RNA	*Parantica sita niphonica*	Chestnut tiger butterfly	8.00E-06	gb|GU372440.1|
GRF-57	y	y	y	y	Putative toxin mRNA	*'Citharischius crawshayi'*	King baboon tarantula	3.00E-35	gb|GU170876.1|

GRF-58	y	y	y	y	16S ribosomal RNA	*Brachypelma vagans*	Mexican red-rump tarantula	6.00E-65	emb|AJ585408.1|
GRF-59	-	-	y	y	Venom protein	*Aedes aegypti*	Yellow fever mosquito	9.00E-08	ref|XM_001655448.1|
GRF-61	y	y	y	y	16S ribosomal RNA-like mRNA	*Ornithoctonus huwena*	Chinese earth tiger tarantula	5.00E-38	gb|EU979519.1|
GRF-62	-	-	y	y	Hypothetical protein	*Latrodectus hesperus*	Western black widow spider	2.00E-15	gb|HQ006063.1|
GRF-63	-	-	y	-	Valyl-tRNA synthetase	*Acyrthosiphon pisum*	Pea aphid	3.00E-06	ref|XR_119136.1|
GRF-64	y	y	y	y	16S ribosomal RNA	*Brachypelma vagans*	Mexican red-rump tarantula	1.00E-71	emb|AJ585394.1|
GRF-66	-	-	y	-	Cyclophilin A	*Mythimna separata*	Oriental armyworm	4.00E-06	gb|HM113489.1|
GRF-67	-	-	y	y	Uhu transposon (lambda-Het3)	*Drosophila heteroneura*	Pomace fly	2.00E-13	emb|X63028.1|
GRF-69	-	-	y	-	Hypothetical protein	*Pediculus humanus corporis*	Human body louse	5.00E-08	ref|XM_002429137.1|
GRF-75	y	y	y	y	16S ribosomal RNA	*Brachypelma vagans*	Mexican red-rump tarantula	1.00E-71	emb|AJ585394.1|
GRF-76	y	y	-	-	16S ribosomal RNA	*Parantica sita niphonica*	Chestnut tiger butterfly	8.00E-06	gb|GU372440.1|
GRF-77	y	y	y	y	Translation initiation factor 5A	*Ornithoctonus huwena*	Chinese earth tiger tarantula	2.00E-67	gb|EU979495.1|
GRF-78	-	-	y	y	Hypothetical protein	*Latrodectus hesperus*	Western black widow spider	5.00E-12	gb|HQ006051.1|
GRF-79	y	y	y	y	Casein kinase II	*Xenopus laevis*	African clawed frog	7.00E-10	ref|NM_001090657.1|
GRF-80	y	y	y	-	Crog-evp-516-318 transport protein	*Caligus rogercresseyi*	Sea louse	3.00E-12	gb|BT076872.1|
GRF-83	y	-	y	-	Cyclophilin-like protein	*Tribolium castaneum*	Confused flour beetle	8.00E-13	ref|XM_961215.2|
GRF-84	-	-	y	y	Hypothetical protein	*Latrodectus hesperus*	Western black widow spider	1.00E-28	gb|HQ006016.1|
GRF-85	y	y	y	y	18S small ribosomal subunit	*Uncultured marine eukaryote*	~~~	5.00E-34	gb|GU370021.1|
GRF-86	y	y	y	y	18S ribosomal RNA	*Hyperia galba*	~Crustacean~	0	gb|DQ378046.1|

===

Contigs of note:

CT-09x51	y	Putative serine proteinase inhibitor	*Haemaphysalis longicornis*	New Zealand cattle tick	3.00E-12	dbj|AB440203.1|
CT-17x58x64x75	y	16S ribosomal RNA	*Ornithoctonus huwena*	Chinese earth tiger tarantula	3.00E-28	gb|EU979519.1|
CT-34x86	y	18S ribosomal RNA	*Hyperia galba*	~Crustacean~	2.00E-83	gb|DQ378046.1|

All other contigs are given in Appendix 5. Note: *‘Citharischius crawshayi’* as of 2010 is now: *Pelinobius muticus*.

Table 4.6 The results of the EST tBLASTxR analysis against the NCBI nr protein database.

Characterisation	No. of clones
Total number of unigenes or ESTs for tBLASTxR analysis	86
Known genes (E-values $< 10^{-5}$)	19
Putative genes (E-values $< 10^{-5}$)	7
Unknown genes (E-values $< 10^{-5}$)	9
Novel genes	51

There was a high homology between this library's unknown genes and translated (tBLASTxR and tBLASTxS) queries within the database. The remaining completely unidentifiable sequences (E-value $> 10^{-5}$), 44/86 (51.2%), while in some searches sharing significant homology with database ESTs, did not break the significance threshold and thus are ignored from further analyses. It should be noted at this stage that a proportion of the ESTs, as expected, came with a poly-A tail from the mRNA parent. The regions encompassing these tails were favoured due to their simplicity and the number of similar sequences in the database resulting in many ambiguous results post-BLAST search. Therefore, these and the adjoining 20 or so nucleotides were removed to effectively BLAST the gene rather than the UTR and the ubiquitous stretch of A's. These 3' UTR's however did have some conservation (Figure 4.9), particularly with the adenine and thymine residues in which they were abundant, with 120:37:40:123 residues for A:C:G:T respectively (Table 4.7). However, due to the high A/T content of this region (75% of total, ignoring the A-tail), these matches could easily have arisen by chance.

Table 4.7 Nucleotide proportions

Nucleotides	Number	% Total
Total	23313	100.00
A	7148	30.66
C	4493	19.27
G	4674	20.05
T	6646	28.51
N	352	1.51
Unambiguous nucleotides	22961	98.49
A/T	13794	59.17
G/C	9167	39.32
Purine	11822	50.71
Pyrimidine	11139	94.22

```
ATTTATATAGATGTAACCAATNTTCCTTGTTGGACAAATAAATAAAAAATNTAATAAATAAAAAAAAAAAA
ACATATGTAACTGTAAATTACAATTGCATTGCATTATAAAAAATGAATTTTTATTGCATAAAAAAAAAAAA
AGATCTGCTTTTTATAAGTAAAACTGCCTTATACAAAATAAATTTGATTTTGAACTGGTAAAAAAAAAAAA
AGATTTGTCATTCTACTATTAATGTTCACTAAACTGGAATATCATTAAAGATGAACTTGAAAAAAAAAAAA
TTTTGTCACCTTTGTATAAGTCATGTTCCAGCATGAGATGTAATTAAATGTTTAAGTATAAAAAAAAAAAA
                    TGCTTAATAAACAAAAATCTGTTCTGTAAAAAAAAAAAA
:--*-*-----*---:-------------:-------*--::----*----:------:***********
```

Figure 4.9 By-eye alignment of the 3' regions of the six ESTs with obvious A-tails. Bold (*) residues indicate 100% identity while greyed (:) residues indicate > 80% identity.

4.3.3 PUTATIVE FUNCTIONS

Of those 86 high-quality ESTs, 35 (40.7%) bore significant homology to database genes and have been grouped into five individual cluster categories. Of those remaining, 18 out of the 35 (representing 20.1% of the 86, but 51.4% of the gene-matched ESTs) corresponded with ribosomal proteins of one kind or other, 7 (8.1%:20.0%) were considered to be housekeeping genes for example translation initiation factor 5A or cyclophilin, 5 (5.8%:14.3%) were either digestive or ecdysis-related genes, e.g. chitinase (explained later), 2 (2.3%:5.7%) were for defence/immunity e.g. the serine protease inhibitors (serpins), 2 (2.3%:5.7%) for cell-communication e.g. casein kinase and finally one (1.1%:2.9%) represented a silk (Figure 4.10). There is also no evidence of peroxidases, which Pouchkina (2003) implied might be responsible for the formation and/or processing of both the major and minor ampullate silk.
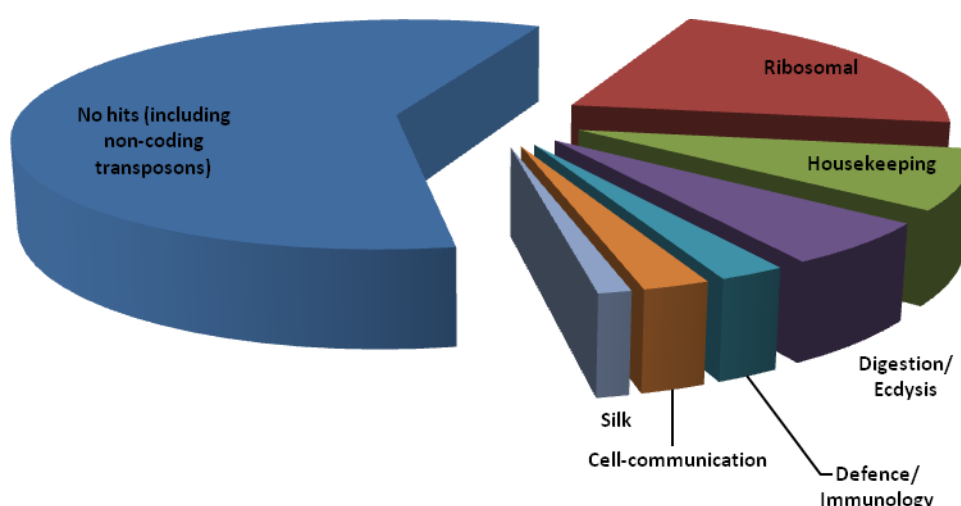
Figure 4.10 Pie chart illustrating the classification of ESTs obtained from the cDNA library on the basis of their putative functions.

159

4.4 DISCUSSION

Construction of a cDNA library proved to be an efficient and useful tool in examining the expression patterns of *A. avicularia*. The initial intent to sequence full-length silk transcripts was deemed implausible using this method but post-dissection studies of the specimens suggested a significant part of the opisthosoma was utilised in the production of silk and the relative probabilities of acquiring at least a partial transcript was reasonably high. Partially sequencing selected cDNA clones is becoming a rapidly growing area in the generation of ESTs for genomic research (Ko *et al.*, 2000). Here, EST analysis was used to explain a functional genomic examination of the spider *Grammostola rosea* but this was limited by the quality of the library due to the complexity, diversity and relative abundance of cDNA clones.

4.4.1 LIBRARY ANALYSIS

4.4.1.1 Discrepancies between the searches

Table 4.8 illustrates the relative disproportion of sequences found within the databases. EST GRF-7 for example yielded 'positive' results ($E > 10^{-5}$) for the BLASTnN (entire database) search as well as the tBLASTxR (arthropods) and tBLASTxS searches but failed to retrieve significant sequence homology with the BLASTnR (arthropods). Conversely, GRF-17 only scored hits in the tBLASTxR search with the others far short of that level of significance. The latter example can possibly be explained by the severe lack of arachnid sequences within the database with the exception of frequent occurrences of *Ixodes scapularis*, a result of the *Ixodes scapularis* Genome Project (IGP). As a result of this there are many such hits (GRF-14, 30, 40, 46, 49, 66 and 69) in which only those subjected to the general arthropod search (tBLASTxR) produced any significant matches while 'spiders' as a search term failed. The former example however, where a broad database search (BLASTnN) found hits while a more precise search failed to do so (GRF-7 only, checked and rechecked) can only be explained by a database optimisation error in which subcategories (specific taxon) of the database are utilised preferentially over others. Instances where (x)BLAST**x**(x) were used, which resulted in hits while the (x)BLAST**n**(x) searches did not can easily be explained by codon bias; where at a nucleotide level, the sequences vary significantly but once translated, there is sufficient similarity to flag a suitable match.

161

Table 4.8 Hypothetical proteins and their corresponding 'most likely' proposed putative functions

Sequence ID	Proposed Putative Function	*E*-value
GRF-20	Cytochrome c oxidase subunit IV	2.7
GRF-21	Heat shock protein 90 (HSP90)	2.4
GRF-40	Mariner transposase pseudogene	2.00E-05
GRF-46	Transposon mariner-like element	2.00E-11
GRF-49	Putative phospholipase B-like	1.00E-07
GRF-62	5' nucleotidase	0.66
GRF-69	N6-adenosine-methyltransferase	7.00E-08
GRF-78	Nidogen and related basement membrane protein	3.00E-09
GRF-84	Inhibitor of nuclear factor k-B	1.7

4.4.2 STRENGTH OF THE LIBRARY/ANOMALOUS RETRIEVALS

Irrespective of the statistical significance of the corresponding *E*-values, particular sequences are undeniably wrong when judged by eye. A string of 20 identical nucleotides would be hard to refute unless this was a poly-adenine stretch pertaining to the poly-A tail of mRNA, in which case a specific contributing EST would be hard to allocate, but some EST matches here have been shown to be unlikely due to their obviously divergent progenitor. In this instance, GRF-41, while there is undeniably a high degree of similarity between the sequences, the disparity between amino acids make this unlikely to be a silk. This is reaffirmed by the frame to which it is aligned and the number of stop codons therein as when aligned with the correct frame, the homology falls to non-significant. Nevertheless, perhaps it should not be ruled out entirely.

```
>gb|AF218621.1|AF218621S1 Nephila clavipes flagelliform silk protein (Flag)
gene, upstream
partial cds
Length=14999
Score = 63.4 bits (132),
Expect = 5e-11
Identities = 33/87 (38%),
Positives = 43/87 (49%),
Gaps = 0/87 (0%)
Frame = +1/-1

Query   10   YQAFGACSQLVLDVRFEQLGLTLRSGNCGHVLQTSFR*FYTRTSNWKAGRRTTNNISSMS   189
             ++ F +C QL+  V+  Q      R+ N  HV  TS R*FY   + WK  RR   NN
Sbjct   914  FKGFDSCRQLLFQVKVLQERFNRRT*NFWHVFPTSHR*FYAWPNYWKDRRRAKNNRCCQG   735

Query   190  ILHRPQFCF*ALGSVSNKGNVYRRKRG   270
             + HR Q CF A+  + N  NV    RG
Sbjct   734  VRHRSQRCFTAVKVI*NYWNV**TARG   654


Score = 63.4 bits (132),
Expect = 5e-11
Identities = 29/59 (49%),
Positives = 39/59 (66%),
Gaps = 0/59 (0%)
Frame = +2/-2

Query   80   DLEIAVMSCRHHLDNFTRGRVIGKLEGGRLIIYLP*AFYIAHSFASRLWAAFQTKVTCT   256
             +L+I  MS RH +D+F RGR+IGK+E GR I  +   F IAHS SRL +F+T   C+
Sbjct   844  ELKIFGMSSRHRIDDFMRGRIIGKIEEGRKITDVAREFDIAHSVVSRL*KSFKTTGMCS   668
```

Figure 4.11 GRF-41 aligned with *Nephila clavipes* flag gene illustrating the numerous stop codons and identity disparity.

4.4.3 CONTIGS

In total 12 contigs were aligned/generated from the following 25 singletons: GRF-02-03-70, 09-51, 14-59, 17-58-61-64-75, 34-86, 37-39, 42-47, 52-74, 55-65, 56-76 and 66-83 but only three (CT-09x51, CT-17x58x64x75 and CT-34x86) resulted in any significant database retrievals with a putative serpin, a 16S ribosomal RNA and 18S ribosomal RNA respectively, where the notation 'CT' represents an assembled contig whilst GRF prefixes identify singlets. The additional length allowed from the new contig from GRF-09-GRF-51, upgraded GRF-51 from being an unknown to having an *E*-value of 10^{-12} allowing it to share significant homology with the *Latrodectus hesperus* (Western black widow spider) putative serine protease inhibitor. Interestingly, CT-17-58-61-64-75 while surprising in its own right for encompassing the

163

exact same sequence five times (suggesting a highly expressed transcript, as would be expected with ribosomal RNAs), has fractured at exactly the same site in four instances. Had these transcripts been identical in length, it would imply a replication of the parent recombinant plasmid, perhaps from either a transference contamination or excessive time permitted for the ampR/blaTEM1 gene to transcribe the β-lactamase during the recovery phase of transformation. However, as the other terminus has been truncated at different residues, this would suggest five different progenitor plasmids. It has been reasonably well documented that DNA can include these so-called 'fragile sites' (Casper *et al*., 2002), but to the author's knowledge, no such sites have been documented on RNA.

Another interesting point to note is that upon aligning this contig with its nearest database reference sequence (*Brachypelma vagans* mitochondrial nd1 gene (partial), 16S rRNA gene (partial) and tRNA-Leu gene, isolate pooks8, GI: 53124977), at this truncation point, there is no fluid alignment with the *B. vagans* nd1 gene. The real alignment starts some 14 nts downstream of this locus. Under normal circumstances the sequence quality might be questioned but this time, there are five identical confirmatory transcripts. One can only speculate as to whether this is indicative of a precursor tRNA and this extraneous region of 14 nts is a pre-excised intron. Likewise, to the author's knowledge, there have never been any publications documenting spider introns to any great detail, let alone tRNA introns. tRNA introns do not appear to exist within the human genome but are found in other eukaryotes (Bernardi 1978) such as yeast (Hebbar *et al*., 1992) and protists (Gray *et al*., 1998) such as *Dictyostelium discoideum* (Gray *et al*., 2004) albeit

infrequently. It is conceivably more likely to be a highly expressed nuclear pseudogene, but given the data, this is perhaps all that can be speculated.

4.4.4   HYPOTHETICAL PROTEINS

Despite a large number of ESTs having found no homology to anything in the database, a significant number (25.7%) returned matches to hypothetical proteins (GRF-20, 21, (40, 46) 49, 62, 69, 78, 84) (Table 4.8). Initially, putative sequences were chosen due to identity matches, for example GRF-49 has an E-value of 3 x 10^{-12} (Figure 4.12) but is only described as a hypothetical protein, whereas the nearest 'true' protein match found for it within the database (phospholipase B-like 2-like) while less significant, still has a 'significant' statistical E-value of 1 x 10^{-7} (Figure 4.13). Using this method of collating the most frequent mention of terms and/or highest E-value matches pertaining to a previously categorised 'putative' protein, three 'new' putative ESTs have been categorised. These are GRF-49 (Figure 4.12 and Figure 4.13), GRF-69 (N6-adenosine-methyltransferase) and GRF-78 (nidogen and related basement membrane protein).

```
>ref|XM_002404400.1| Ixodes scapularis hypothetical protein, mRNA
Length=180

 GENE ID: 8031142 IscW_ISCW007955 | hypothetical protein [Ixodes scapularis]


Score = 47.8 bits (98),
Expect = 3e-12
Identities = 16/25 (64%),
Positives = 20/25 (80%),
Gaps = 0/25 (0%)
Frame = +3/+1

Query  12   FKNLEFVAIGGPTYDPLPPFKWSDS   86
            F +L F A+ GPTY+PLPPF+WS S
Sbjct  25   FADLMFTAVAGPTYNPLPPFRWSTS   99
```

Figure 4.12 GRF-49 aligned with a tick (*Ixodes scapularis*) hypothetical protein showing 80% homology (3 E^{-12}).

```
>ref|XM_001948827.2| PREDICTED: Acyrthosiphon pisum putative phospholipase B-
like 2-like (LOC100165900), mRNA
Length=1821

 GENE ID: 100165900 LOC100165900 | putative phospholipase B-like 2-like
[Acyrthosiphon pisum]

Score = 53.8 bits (111),
Expect = 1e-07
Identities = 17/33 (52%),
Positives = 22/33 (67%),
Gaps = 0/33 (0%)
Frame = +3/+2

Query  60     LPPFKWSDSDFGSTIPHEGHPDLWKFMPIVHKW   158
              L PF WS SDF  + H GHPD + F P++H+W
Sbjct  1586   LGPFCWSKSDFNDKVSHLGHPDCFNFKPVLHQW   1684


 Score = 22.6 bits (43),  Expect = 1e-07
 Identities = 8/13 (62%), Positives = 10/13 (77%), Gaps = 0/13 (0%)
 Frame = +3/+2

Query  12     FKNLEFVAIGGPT   50
              F+ L+F AI GPT
Sbjct  1529   FQQLQFKAIAGPT   1567
```

Figure 4.13 GRF-49 aligned with the pea aphid (*Acyrthosiphum pisum*) Phospholipase B-like 2-like sequence still showing significant (1 E^{-7}) homology.

4.5 CONCLUSIONS

The study described in this chapter has succeeded in its initial aim of uncovering unique sequences and additionally, a plausible silk candidate from a random sampling of cDNA clones. EST analysis has yielded successful identification of partial gene sequences that may be of special significance in the quest for understanding how silk secondary and tertiary structures are formed as well as numerous previously undescribed homologues to database theraphosid toxins and ribosomal RNAs. As ribosomal RNA (rRNA) comprises more than 80% of total RNA, (18S and 28S in mammalian systems), RNA sample integrity was verified. Incomplete rRNAs could be attributed to the fragility corresponding to the complex secondary structures formed, the harsh protocols involving the phenolic reagents or the 'hidden break' hypothesis. If the breakage (particularly within the 28S), was not induced during manipulation, there exists an AU-rich sequence called the 'hidden break' (coined by Ishikawa and Newburgh (1972) after Gould (1967)) which has a higher tendency to fracture. This bears resemblance to the proposed 'fragile sites' (Casper *et al*., 2002), which could explain consistent length polymorphisms. The single silk 'transcript' uncovered should not be considered a silk (perhaps a pseudogene), despite its high homologies due to the frame shift, which when corrected for, yielded no significant match to previously described silks and even then, none in frame. There is also a high prevalence of transposable elements (McClintock 1948/1950), nuclear pseudogenes and evidence of RNA intron editing, the former probably a result of evolutionary horizontal gene transfer (HGT) or an undescribed discrete form of heterologous recombination. These mobile genetic elements (MGEs) (Frost

et al., 2005) can serve as a method of tracking genome evolution as within

higher eukaryotes they comprise up to 40% or more of the total sequence

(Curcio and Derbyshire 2003). Likewise, pseudogenes can be utilised in much

the same way. These are remnants of functional genes lost through time in a

similar fashion to how single nucleotide polymorphisms (SNPs) are used to

trace ancestry of a genome.

5 *STEGODYPHUS* TRANSCRIPTOME MINING FOR *DE NOVO* SILK SEQUENCES

5.1 INTRODUCTION

Stegodyphus belongs to the araneomorph family Eresidae, which includes the now rare ladybird spider, *Eresus cinnaberinus* (Olivier 1789). Members of this family, first characterised by Simon (1892-1903), are commonly referred to as subsocial spiders despite only two (*S. dumicola* (Henschel 1998) and *S. mimosarum* (Ward 1985; Schneider *et al.,* 2001)) being truly social in terms of co-habiting with conspecifics. *Stegodyphus* can be found in Europe, Africa and Asia with two species, *S. manaus* and *S. annulipes* (Kraus and Kraus 1992), found in South America. Distribution of this species is further aided by their ability to balloon and has been observed in several representatives of the family, e.g. *S. mimosarum* (Wickler and Seibt 1986) and *S. dumicola* (Schneider *et al.,* 2001). Even adult females (Kraus and Kraus 1988) have been observed tiptoeing on the highest strand of the web, letting out a long strand of silk that separates into thousands of micro-strands and releasing themselves into the breeze (Schneider *et al.,* 2001).

Stegodyphus nests are either small tubular structures composed of cribellate silk (Ward and Lubin 1993; Johannesen and Lubin 1999) for *S. lineatus*, or large clumps in trees and bushes comprising of scores of pre-social, semelparous individuals (Crouch and Lubin 2000; Seibt and Wickler 1988) for *S. mimosarum*. Males and females look morphologically similar with the male's patterning having more colour contrast and a more pronounced

cephalothorax (Bellmann 1997). Females are polyandrous (Maklakov and Lubin 2006) but this can be costly for both sexes (Maklakov *et al.,* 2005) so the female usually tries to defend against further mates (Schneider and Lubin 1996). Post assisted eclosion, the female regurgitates pre-digested material for the young spiderlings (Kullmann and Zimmermann 1974). Soon after, the matriphagous offspring consume her.

Here, an analysis of a draft-grade genome sequenced by next-generation Roche 454 technology of three Eresid *Stegodyphus* spiders has been conducted (courtesy of Bilde *et al*., unpublished). To date there are about 36 000 expressed sequence tags (ESTs) generated from the Arachnida (spiders and their kin) within the NCBI database (http://www.ncbi.nlm.nih.gov), most of them encompassing those used for molecular phylogenetics and DNA barcoding (e.g. mitochondrial genes). These days there is great demand for high-throughput technologies such as next-generation sequencing and hybridisation-based microarray (Forrest and Carninci 2009; Peatman and Liu 2007). Pyrosequencing provides transcriptomic analysis of whole organisms, tissues or cells but is therefore spatially and temporally constrained. Roche 454 based sequencing surpasses capillary based sequencing in its capacity for sequence depth and contig numbers of ESTs and post normalisation of cDNA pools provides a far more representative sampling of transcripts. *Stegodyphu*s spidroins have never before been genetically characterised and so these transcriptomic analyses will add further insight into the variability of these structurally conserved but highly sequence-divergent proteins. Several classes of putative spidroins are identified and predictions of their structures and functions are made.

5.2 MATERIALS AND METHODS

5.2.1 METHOD BACKGROUND AND JUSTIFICATION

5.2.1.1 UNIX-based operating systems

Access to the software and databases can be achieved through various free and open-sourced terminal emulator applications such as PuTTY. This application can act as a client for Telnet, rlogin and Secure Shell computing protocols or likewise as a client for a serial console, which is for system administration as a text and display device. Secure Shell (SSH) allows for secure data communication between two networked computers over an insecure network whereby both server and client are running SSH. The UNIX-based operating system consists of a master control program called the kernel, which essentially does the system's housekeeping, handling the file system and avoiding programming conflicts between users.

Here, programs such as BLAST can be utilised to search DNA or protein databases as well as genome-wide searches with NCBI and Ensembl. Once sequences have been downloaded, usually in FASTA format (using a > symbol as the identifier of the sequence), numerous commands can be exploited to adapt the sequences and conduct various searches within them. The fundamental issue with UNIX is without a doubt the unfamiliar user interface and confusing commands required.

5.2.1.2 Searching for silk motifs using Microsoft Office Word

A far more recognisable and perhaps just as versatile a program is Microsoft Office Word (Word 95 - present). Sequences are usually downloaded in the aforementioned FASTA format into Notepad, a text-only (plain text) editor found on most Windows Operating Systems. Here they can be easily copied across to Microsoft Office Word where the alterations and searches can be performed. The wildcard feature of Microsoft Office Word 2007 is an extremely powerful tool on par with that of UNIX-based operating systems frequently used for sequence analyses. It utilises a set of punctuation characters ([ ] { } < > ( ) - @ ? ! * \) as a method of searching for others within a document by substituting them for any other character or characters in a string.

The Find/Replace wildcard feature is activated by turning on the Find/Replace dialog under 'Edit' in the Quick Access Toolbar or by holding down the 'Control' key while pressing the letter *F*. The wildcards option is then activated by pressing the 'More > >' tab ('Alt' key plus the letter *M*) and checking the box 'Use wildcards' (Alt key again, followed by the letter *U*). While using the 'Find' option, all search options (Match case, Find whole words only etc) are highlighted; however, during a wildcard search, these options are removed/greyed out. This option can also be activated by setting up a macro (set. Find.MatchWildcards = True) which achieves the same result. As wildcards are case sensitive, the whole document must be selected and the case changed to a uniform setting. This can be an issue when lower case characters have been used to define ambiguities or repetition but these are easily identified by aligning any identified sequences with the original.

Wildcards, like the 'grep' command in UNIX, identify strings of text while ignoring everything else. As some searches can be performed using a variety of wildcard characters, the choice of which to use is at the user's discretion. For example, in a DNA-based file in which all nucleotides are represented by the standard adenine (A), cytosine (C), guanine (G) and thymine (T) residues, using the command [ACGT] to find any one of them would achieve the same result as simply using the question mark character '?' or even the asterisk character '*'. The search results can then be uploaded individually or en masse into a six-frame translator to identify matches. Likewise, the search can be performed twice utilising the reverse complement wildcards. For example a Frame - 1 sequence corresponding to the amino acid sequence 'QALLE' would be: ca[ag]gc?[ct]t?[ct]t?ga[ag] but the sequence [ct]tc?a[ag]?a[ag]?gc[ct]tg could likewise be used to search for the Frame - 4 (reverse complement) equivalent.

ca[ag]gt?[ct]t?[ct]t?ga[ag]

Sequence 1:
Frame 1:   gta caa gtc tta ctg gaa gtg
Peptide: **vQVLLEv**

Sequence 2:
Frame 1:   ata gca agt ctt act gga atg a
Peptide: IASLTGM

Frame 2: a tag caa gtc tta ctg gaa tga
Peptide: -**QVLLE**-

An exhaustive list of all Microsoft Word wildcards is shown in Table 5.1.

Table 5.1 Wildcard symbols available for use in Microsoft Office Word (2007)

Character	Name	Searches for…	Will find	But will NOT find
*	Asterisk	A range of characters	B*T … will find BAT, BET, BULLET…	BaT, Bet, BulleTs…
?	Question mark	A single character	B?T … will find BAT, BET, BIT…	BaT, BeT, BiT…
@	'at' sign	Multiple occurrences of the preceding character	A@ … will find A, AA, AAA…	a, Aa, AaA…
< >	Angle brackets	The start and end of a word respectively	<B*T> … will find BIT, BOOT…	BITTER, BOOTS, BooT…
[ ]	Square brackets	Ranges of characters	[A-D] or [ABCD] … will find A or B or C or D…	a or b or c or d…
-	Dash	Illustrates those ranges		
\	Back slash	Wildcard characters	[*] … will find '*'	\a, \A or *…
!	Exclamation mark	Everything except the following character(s)	[!0-9] … will find a,b,c… A, B, C… %, £, #…	0, 1, 2, 3…
{ }	Curly brackets	Numbers of occurrences of the preceding character	A{6} … will find 'AAAAAA'	aaaaaa
^	Caret	Special characters	^% … will find the section mark '§'	^%, ^, % …
( )	Round brackets	Replaces the find in a different order	(University) (Nottingham) **replaced by:** \2 \1	**Will result in:** Nottingham University

5.2.2 DATA ACQUISITION

Genomic sequences of the three eresid spiders *Stegodyphus lineatus*, *Stegodyphus mimosarum* and *Stegodyphus tentoriicola* were acquired by Bilde *et al.* by 454 pyrosequencing (Aarhus University, Denmark) and relayed in FASTA format, pre-publication (Bilde *et al.,* submitted).

5.2.3 SEQUENCE MANIPULATION SOFTWARE

Sequences were sorted, identified and initial manipulation performed in Microsoft Office Word 2007. Quality control runs were conducted using the online ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) Translate program (Gasteiger *et al.,* 2003). Reverse complement sequences were generated using an online converter on http://www.bioinformatics.org/sms/rev_comp.html (Stothard 2000). Sequences were aligned using primarily ClustalW: http://www.ebi.ac.uk/Tools/msa/clustalw2/ (Thompson *et al.,* 1994) with more precise manipulations being conducted in BioEdit v. 7.1.3 (Hall 1999).

5.2.4 MOLECULAR MODELLING

Illustrations of molecular tertiary structures were performed using The PyMOL Molecular Graphics System, Version 1.5, Schrödinger, LLC and Open RasMol Molecular Graphics Visualisation Tool (Sayle and Milner-White 1995; Bernstein 2000).

5.2.5 TRANSLATION INTO PUTATIVE PROTEINS

Putative proteins were predicted by means of translating the raw data contigs using the (**Ex**pert **P**rotein **A**nalysis **Sy**stem) Translate program and visually inspecting for open reading frames (ORFs). When particular silk-like

motifs were recognised within a reading frame (A_n), $(GA)_n$, $(GPGGN)_n$, QALLE etc) and there were no obvious stop codons within a visually determined cut off point of 20 nucleotides in both upstream and downstream directions, then sequences were retained for subsequent analysis.

5.2.6 HOMOLOGY SEARCHES

All identified (post-translational) contigs and singletons identified via the Word-Wildcard method were subjected to a BLAST search (tBLASTx) with spiders (taxid:6893) under the Organism parameter and disabling the filter for low complexity regions. Searches were performed against the Nucleotide Collection (nr/nt) database. No distinction was made between short and long contigs and all were tested for homologies to silk proteins. Likewise there were no bit-score cut off ranges as despite frequent high bit scores, silk proteins were in many cases definitely not the most homologous. *E*-value scores lower than the recommended (NCBI) *E*-values of 10^{-5} were considered to be significant, while all others were discarded; however there were numerous incidences where even an order of magnitude higher were deemed to be database alignment errors.

5.2.7 SEARCHING FOR SPECIFIC GENES

The sequences pertaining to silk were retained and Clustal analyses were performed between both inter- and intraspecific sequences to determine potential progenitors. In many instances, a distinct silk could be identified but when short or incomplete contigs were encountered and when many silks share a similar C-terminus for example; clear-cut distinctions could not be made between homologous and non-homologous transcripts.

5.3 RESULTS

5.3.1 BASIC SEQUENCE ANALYSIS

The raw data provided resulted in a total number of 24475 reads with an average sequence length of 939, 935 and 956 nucleotides for *S. tentoriicola*, *S. mimosarum* and *S. lineatus* respectively. An overview of the nucleotide content for each species is outlined in Table 5.2, Table 5.3 and Table 5.4 including the GC content and percentage of each individual base. The length of the sequences ranges from 5 - 6475, 30 - 7495 and 63 - 7584 for *S. tentoriicola*, *S. mimosarum* and *S. lineatus* respectively.

Each species' raw data were composed of no more than 10 million nucleotides, substantially less than the database haploid nuclear DNA content as illustrated by Figure 5.1 in which each species' genome size is in excess of 900 MB as calculated by the formula from Dolezel *et al.* (2003). It is unclear how the samples were acquired apart from their transcriptomic origin or with what stringency the reads were taken i.e. was it an intentionally limited read or were the samples modified and read to target specific sequences. Naturally, given a transcriptomic-derived dataset, the genes sequenced represent only those expressed at a given time; here, approximately 0.3% of the total average of genome sizes (Figure 5.1) for *S. tentoriicola* and *S. mimosarum* and 0.5% for *S. lineatus*. It is unclear why there is 33% more expression in *S. lineatus*.
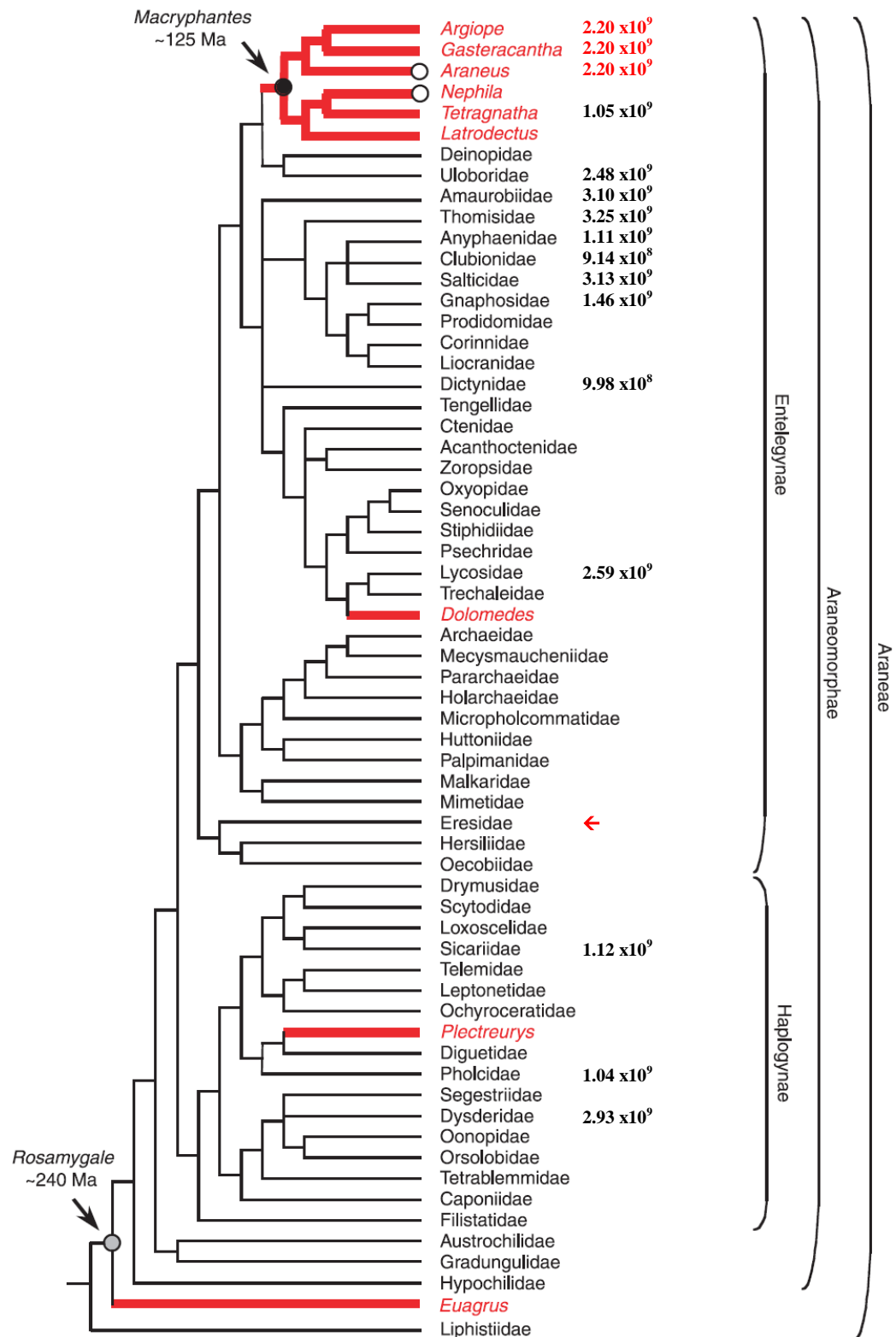
Figure 5.1 Hypothesised phylogenetic relationships of Araneae based on morphological evidence according to Gatesy *et al.* (2001). Previously published spider fibroin sequences are marked by white circles and those by Gatesy *et al.* (2001) in red. Numbers indicate additional approximations of haploid genome calculated from Gregory and Shorthouse (2003) according to Dolezel *et al.* (2003) as per the Animal Genome Size Database. http://www.genomesize.com, Gregory (2012). Unpublished data for Sicariidae (Rasch). Eresidae (incl. *Stegodyphus*)indicated by a red arrow.

5.3.2 HOMOLOGY SEARCHES

Preliminary analyses show that sequences with fewer than 200 nucleotides rarely resulted in any significant find from the BLAST searches, perhaps due to a threshold length being required for recognition or fragile sites occurring less frequently in conserved regions. Initial wildcard findings are shown in Table 5.5. According to the literature (Challis *et al.,* 2006), a QALLE motif, associated with the C-terminus of the majority of silks sequenced prior to said publication, was used as an initial search term. This search yielded 37 unique sequences results all with this motif, identical to a previously conducted UNIX search (Appendix 6).

There was a predictable loss of around two thirds of these when searched against the Genbank database using BLAST, due to their being out of frame. The resulting nine sequences returned silk as their closest matches from the BLAST search. Interestingly five of these nine aligned virtually perfectly with each other but when examined more closely, the BLAST retrieval was not recognising the (Q/E)ALLE motif, which had been the original search string but was matching a coincidentally GA-rich region to the silks within the database. The other (Q/E)ALLE 'positive' hits mirrored this trait, finding a string of cysteines and a silk terminus-like homology but spattered with stop codons.

Further searches utilising other common motifs such as the $(GA)_n$, $(A)_n$ and $GPGG(X)_n$ motifs found commonly in the more heavily researched MaSp and flagelliform spidroins (Hayashi and Lewis 2000), appeared to catch the remaining cryptic silks. These initial findings, particularly with the $(GA)_n$ motifs, illustrated a preference of *Stegodyphus* for the QVLLE motif, which is

179

rare but not unheard of. By exploiting this as a search-term, modifying the wildcards accordingly and incorporating the aforementioned $(GA)_n$, $(A)_n$ and $GPGG(X)_n$ hits, a total of 326 silk-positives were mined. Once frame implausibles were eliminated, a total of 29 apparent MiSps, 8 MaSps and 4 pyriform silk sequences were extracted as well as one N-terminus. Flagelliform proteins were difficult to distinguish from other silk-like hits due to a high proline content (explained later) so these were examined more thoroughly. A quick search of aggregate proteins was performed too using ten different search parameters based upon data by Choresh *et al.* (2009) in which there was conjecture surrounding the possibility of a silk being encoded by opposite strands of the same DNA sequence. Despite numerous hopeful attempts, exploiting as much of the repetitive domain as possible, no such sequences were found.

Table 5.2 Base composition of the received *S. tentoriicola* genome

Nucleotides	Number	% Total
Total	6,757,223	100
A	2,181,046	32.28
C	1,177,880	17.43
G	1,201,981	17.79
T	2,196,316	32.50
A/T	4,377,362	64.78
G/C	2,379,861	35.22
Purines	3,383,027	50.07
Pyrimidines	3,374,196	49.93

Table 5.3 Base composition of the received *S. mimosarum* genome

Nucleotides	Number	% Total
Total	6,367,977	100
A	2,036,067	31.97
C	1,120,797	17.60
G	1,173,793	18.43
T	2,037,320	31.99
A/T	4,073,387	63.97
G/C	2,294,590	36.03
Purines	3,209,860	50.41
Pyrimidines	3,158,117	49.59

Table 5.4 Base composition of the received *S. lineatus* genome

Nucleotides	Number	% Total
Total	9,944,393	100
A	3,254,058	32.72
C	1,664,765	16.74
G	1,711,558	17.21
T	3,314,012	33.33
A/T	6,568,070	66.05
G/C	3,376,323	33.95
Purines	4,965,616	49.93
Pyrimidines	4,978,777	50.07

Table 5.5 Motifs searched for within the genomes with their respective finds, IUPAC nucleotide codes and wildcard character matches

Motif	Initial Finds	IUPAC nucleotide code	Wildcard character search
QALLE	37	cargcnytnytngar	ca[ag]gc?[ct]t?[ct]t?ga[ag]
EALLE	31	gargcnytnytngar	ga[ag]gc?[ct]t?[ct]t?ga[ag]
QVLLE	60	cargtnytnytngar	ca[ag]gt?[ct]t?[ct]t?ga[ag]
GPGG(X)$_n$^1	119	ggnccnggnggn	gg?cc?gg?gg?
GPGG(X)$_n$^2	0	ggnccnggnggnggnccnggnggn	gg?cc?gg?gg?gg?cc?gg?gg?
GPGQQ	0	ggnccnggncarcar	gg?cc?gg?ca[ag]ca[ag]
(GA)^1	26669	ggngcn	gg?gc?
(GA)^2	181	ggngcnggngcn	gg?gc?gg?gc?
(GA)^3	21	ggngcnggngcnggngcn…	gg?gc?gg?gc?gg?gc?...
(GA)^4	5	…	…
(GA)^5	1	…	…
(GA)^6	1	…	…
(GA)^7	0	…	…
MAFASS (N)	0	atggcnttygcnwsnwsn	atggc?tt[ct]gc?[at][gc]?[at][gc]?
MAFAS (N)	3	atggcnttygcnwsn	atggc?tt[ct]gc?[at][gc]?
KLQAL (N)	61	aarytncargcnytn	aa[r][ct]t?ca[ag]gc?[ct]t?
A^1	805901	gcn	gc?
A^2	43139	gcngcn	gc?gc?
A^3	3676	gcngcngcn…	gc?gc?gc?...
A^4	470	…	…
A^5	112	…	…
A^6	64	…	…
A^7	32	…	…
A^8	20	…	…
A^9	19	…	…
A^10	15	…	…
A^11	15	…	…
A^12	12	…	…
A^13	7	…	…
A^14	7	…	…
A^15	3	…	…
A^16	0	…	…

5.4 DISCUSSION

The utilisation of 454 for genomic analysis has become increasingly more popular. To date this is, to the author's knowledge, the only araneomorph spider transcriptome to have been sequenced. The potential for comparative genomics amongst the lesser-studied organisms is now being exploited given the popularity of 454 pyrosequencing over the classic Sanger method. In this study, the limitations of classic techniques are described using three partial genomes of *Stegodyphus* spp. The silks presented here illustrate a unique and tangible model for visualising the mechanics of genomic recombination. Despite having found multiple examples of MaSp silks spread across the species, there is remarkable variation between them despite a probable common deployment within one type of silk. The variations between 14 different identified C-terminal silk contigs are shown in Figure 5.2. On a nucleotide level, the differences are not obvious except for a disparity between the variable amino acids in each of the reading frames. In position one of each triplet, there are 27 incidences of variability compared to just 13 for position two but a substantially larger 55 for position three, i.e. where there is a lack of conservation between all 14 sequences. As expected, the third position is more flexible due to degeneracy (redundancy/flexibility of the genetic code providing multiple codons for each amino acid). Particular conservation is seen around the 3' end of the DNA sequence despite the 3rd position degeneracy. Examining codon bias at this level is inconclusive and can only be determined upon inspection of a larger dataset. When an overall homology is taken into account, i.e. where there is greater than 50% identity between the 14 sequences at each respective base, the mutation scale can be examined. When there is a

majority of one particular base within a particular position (8 - 13 out of the total 14 bases), the remaining bases can be identified as either transition or transversion events. The transitions (purine to purine or pyrimidine to pyrimidine) with their given frequencies (in **bold**) are thus: A→G: **16**, G→A: **12**, C→T: **13** and T→C: **29.** The transversions (purine to pyrimidine and *vice versa*) with their given frequencies (in **bold**) are thus: A→C: **5**, A→T: **3**, C→A: **8**, C→G: **3**, G→C: **2**, G→T: **9**, T→A: **17** and T→G: **7**. So the relative approximate transition to transversion summation ratio is **70:54**. Here it would seem there is a bias towards transitions (purine to purine or pyrimidine to pyrimidine) which is as expected. One would expect that despite there being twice as many possible transversion prospects because of the intrinsic molecular mechanisms (amino-imino forms), transition mutations tend to occur at a relatively higher frequency compared to transversions (Freese 1959). Likewise, transitions result in fewer amino acid substitutions, the wobble hypothesis (Crick 1966) and tend to persist in populations as silent mutations as single nucleotide polymorphisms (SNPs). Note, the aforementioned ratio is not the transition/transversion ratio (R) which is found by comparing two sequences. Similarly with the transition and transversion rate (k = a/b) which is the number of non-synonymous substitutions per non-synonymous site (K_a) to the synonymous substitutions per synonymous site (K_s), an indicator of selective pressure on a gene. Likewise, this kind of comparison will only work on homologous sequences.

```
isotig01583  GGTTCAACTGTGTATTCTACTATTTCGCGATTGTCGTCGTCTTCTTCTTC
isotig05050  GGGTCAACTGTATATTCGACTATTTCGCGATTGTCGTCGTCTTCTTCTTC
isotig06627  GGTTCAACAGTTTATTCCACTATTTCACGATTGTCATCAGCTTCGTCTTC
isotig01555  GGTTCAACTGTATATTCAACTATTTCACGATTGTCATCGTCTTCGTCTTC
isotig03968  GGTTCTACCGTATATTCCACTCTGTCGCGACTGTCGTCACCTTCGTCTAC
isotig04284  AGTTCAACTGTATATTCAACTGTTTTACGATTGTCATCGGCTCCTTCTTC
isotig04370  GGTTCAGCTATATATTCTACTATTTCGCGATTGTCATCTTCTTCATCATC
isotig00411  GGTTCAACTGTGTATTCTACTATTTCGCGGTTGTCATCATCTTCTTCTTC
isotig00412  GGTTCAACTGTGTATTCTACTATTTCGCGGTTGTCATCATCTTCTTCTTC
isotig05058  CGTTCAACGGTGTATTCTACTATTTCACGTTTGTCAAGTGCTTCTTCTTC
isotig10167  AGTTCAACTGTATATTCGACTATTCCAAGATTGTCATCGTCTTCCTCTTC
isotig00671  AGTTCCACTCTCTATTCCACTATTTCACGATTGTCCTCATCTTCGTCATC
isotig00672  GGTTCAACTGTCTATTCCACTATTTCACGATTGTCATCATCTTCGTCATC
isotig01984  GGTTCAACTGTGTATTCTACTATTTCTCGATTGTCATCTTCTTCTTCTTC
              *  **   *   * ***** *** *   ****    ** *  **  *
             GGTTCAACTGTxTATTCxACTATTTCxCGATTGTCATCxTCTTCxTCTTC
             123123123123123123123123123123123123123123123123 12

isotig01583  GTCTAGAGTTTCTGCAGCAGCTTCTGCCCTGGCATCTGGTGGTGCCTTTA
isotig05050  GTCTAGAGTTTCTTCTGCAGCTTCTGCTCTTGCTTCTGGTGGTGCATTCA
isotig06627  GTCTAGAGTTTCTTCTGCTGCTTCTGCTCTCGCATCGGGTGGTTATTTCA
isotig01555  GTCTAGAGTTTCTTCTGCTGCTTCTGCTCTCGCATCTGGCGGTTACTTCA
isotig03968  GTCTAGAGTTACATCTGCTGCTTCTGCTCTCGCATCTAGCGGTTATTTCA
isotig04284  GTCAAGAGTTTCTGCTGCAGCTTCTGCTTTTGGCTTCTGGTGGTTCATTCA
isotig04370  ATCTAGAATTTCTTCTGCTGCTTCTGTTCTGGCCTCTGGTGGTACCTTCA
isotig00411  GTCTAGAGTTTCTGCAGCAGCTTCTGCCCTAGCTTCTGGTGGTGCCTTTA
isotig00412  GTCTAGAGTTTCTGCAGCAGCTTCTGCCCTAGCTTCTGGTGGTGCCTTTA
isotig05058  ATCTAGAATATCTTCCGCTGCTTCTGCTTTGGTATCTGACGGTTCTTTGA
isotig10167  GTCTAGAGTCTCTTCTGCCGCTTCTGCTTTGGCTTCTGGTGGTGCGTTCA
isotig00671  GTCTAGAGTTTCGTCTGCTGCTTCTGCTCTCGCATCTGGTGGTTTCTTTA
isotig00672  GTCTAGAGTTTCGTCTGCTGCTTCTGCTCTCGCATCTGGTGGTTTCTTTA
isotig01984  GTCTAGAGTTTCTGCAGCAGCTTCTGCTCTCGCTTCTGGTGGTGCCTTTA
              ** *** *    *   * ** *******     * * **    ***   ** *
             GTCTAGAGTTTCTTCTGCxGCTTCTGCTCTxGCxTCTGGTGGTxCCTTxA
             312312312312312312312312312312312312312312312312 1

isotig01583  ATTCTAATTCCTTGTCTACAGTAATTTCTGCTTTGGCTTCTCAAGTTCGC
isotig05050  ATGCTAACGCATTGTCTTCAGTTATTTCAAGTCTGTCGTCTCAAGTTCGT
isotig06627  ATGCTAACGCATTGTCTTCAGTTATTTCTAGTTTGTCGTCCCAAGTTCGC
isotig01555  ATGCTAACGCCTTGTCTTCAGTTATTTCTAGTCTGGCGTCTCAAGTTCGC
isotig03968  ATGCCAATGCGTTGTCGTCGGTTATTTCAAGTCTCTCGTCTCAAGTCCGC
isotig04284  ATGCTAATGCATTGTCTTCAGTTATTTCAAGTCTGGCTTCTCAAGTCCGC
isotig04370  ATGCTAATGCATTGTCTTCAGTTATTTCAGATCTGGCGGCTCAGGTTCGC
isotig00411  ATGGTAATTCCTTGTCTGCAGTTATATCTGGTTTGGCTTCTCAAGTTCGC
isotig00412  ATGGTAATTCCTTGTCTGCAGTTATATCTGGTTTGGCTTCTCAAGTTCGC
isotig05058  ACACTAATGCATTGCCGTCTGTTATTTCAAATTTGGCATCTCAAATTCGA
isotig10167  ATGCTAACGCCTTGTCTTCAGTTATTTCAAATCTGGCATACCAAGTTCGC
isotig00671  ATGCTAATGCATTGTCCTCGGTTATTTCTAGTATGGCGTCTCAAGTCCGC
isotig00672  ATGCTAATGCATTGTCCTCGGTTATTTCTAGTATGGCGTCTCAAGTCCGC
isotig01984  ATTCCGGTTCTTTGTCTTCCGTTATTTCAAGTTTGGCTTCTCAAGTTCGT
             *        * *** *   * **  ** ** **     * *  *  **  * **
             ATGCTAATGCxTTGTCTTCAGTTATTTCxAGTxTGGCxTCTCAAGTTCGC
             231231231231231231231231231231231231231231231231 23

isotig01583  TCCACATCTGCAGATCTTTCCGGATGTGAAGTCCTTGTT|CAAGTTCTTT
isotig05050  TCTTCGTCTTCAGATCTCTCCGGATGTGAAGTGCTCGTC|CAGGTTCTGT
isotig06627  TCCACATCTGCAGATCTCTCCGGGTGTGAAGTTCTCGTC|CAAGTGCTCT
isotig01555  TCCACATCTGCGGATCTGTCTGGGTGTGAAGTTCTCGTT|CAAGTGCTCT
isotig03968  TCCACGTCTGCAGATCTGTCCGAATGTGAAGTTCTTGTT|CAAGTGCTCT
isotig04284  TCCTCGTCTTCGGATATGTCCGGATGTGAAGTCCTCGTC|CAGGTTCTTT
isotig04370  TCCACGTCCGATACACTGTCTGGATGTGAAGTCCTTGTT|CAGGTTCTCT
isotig00411  TCCACGTCTGCCGATCTTTCCGGATGTGAAGTCCTTGTT|CAAGTTCTTT
isotig00412  TCCACGTCTGCCGATCTTTCCGGATGCGAAGTCCTTGTT|CAAGTTCTTT
isotig05058  TCCACCTCATCGGATCTTTCTGGATGCGAAATTCTTGTT|CAAGTTCTAC
isotig10167  TCCACATCTTCGGATCTGTCTGGATGTGAAGTCCTCGTC|CAGGTTCTTT
isotig00671  TCAACATCTGCTGATCTCTCTGGGTGTGAAGTTCTCGTT|CAAGTGCTCT
isotig00672  TCAACATCTGCTGATCTCTCTGGGTGTGAAGTTCTCGTT|CAAGTGCTCT
isotig01984  TCGACATCTGAAGATCTTTCCGGTTGTGAAGTCCTTGTT|CAAGTTCTTT
              ** *  **     * ***   ** *** * ** ** ** ** ** **
             TCCACxTCTGCxGATCTxTCCGGATGTGAAGTCCTxGTTCAAGTTCTxT
             123123123123123123123123123123123123123123123 1
```

Figure 5.2 Clustal alignment of 14 identified MaSp C-terminal regions from genomic contigs. Top three = *S. tentoriicola*, middle six = *S. mimosarum*, bottom five = *S. lineatus*. Below, identities are indicated with asterisks, consensus sequences and codon positions with numbers 1 - 3. QALLE motif position indicated by red box.

185

The difference becomes more apparent upon translation of the sequence. Despite obvious homologies between sequences, there are numerous single nucleotide polymorphisms, both interspecifically and intraspecifically resulting in not one sequence being identical to another. Even the sequences with the closest identity still have poly-alanine polymorphisms (Figure 5.3). It is entirely feasible that these differences are due to each of these contigs being derived from different types of silk but as they all contain a poly-alanine stretch at the 5' end, these are all clearly fragments of the MaSp/MiSp families. An interesting correlation has also been made with that of the previous chapter in that out of the whole ~24 000 contigs, there are multiple examples of termini, but very few representatives of the repetitive domains. Maybe this is merely a result of the sequencing technology or the inability of the software to recompile the data into a coherent string, repetitive elements tend to be an issue for all types of sequencing.

```
isotig00411    AAGAGAGSGGYGGDSGAAAAAAAAAAAAAAAGGRVGYGGSGGYGSGSSSAASSSASSSVDS 120
isotig00412    AAGAGAGSGGYGGDSG--AAAAAAAAAAAAAGGRVGYGGSGGYGSGSSSAASSSASSSVDS 118
               ***************   ******************************************

isotig00411    STVYSTISRLSSSSSSSRVSAAASALASGGAFNGNSLSAVISGLASQVRSTSADLSGCEV 180
isotig00412    STVYSTISRLSSSSSSSRVSAAASALASGGAFNGNSLSAVISGLASQVRSTSADLSGCEV 178
               ***********************************************************

isotig00411    LVQVL 185
isotig00412    LVQVL 183
               *****
```

```
isotig00671    YGARGGYGRGAGAGAAAASAAGAGAGQQQGQD--------HGAAAAAAAQGYGAGRGYG- 111
isotig00672    YGARGGYGRGAGAGAAAASAAGAGAGQQQGQDQLAAAAAAAAAAAAAAAQGYGAGRGYGR 120
               *******************************          .****************

isotig00671    ------------------------------------------------SGAGAGAAAA 121
isotig00672    GAGAGSAAASAAGAGAGQQQGQDQGAAAAAAATAAAAAAAQGYGARGGYGSGAGAGAAAA 180
                                                               *********

isotig00671    SAADSGVRQQGRSYDFIADAAALASAAASAFGSGGYDRRRYGGGSSAAAASSAASSSSVD 181
isotig00672    TAAVSGVGQQGRSYDFIADAAALASAAASAFGSGGYDRRGYGGGSSAAAASSAASSSSVD 240
                **  *** ******************************** *******************

isotig00671    SSTLYSTISRLSSSSSSSRVSSAASALASGGFFNANALSSVISSMASQVRSTSADLSGCE 241
isotig00672    SSTVYSTISRLSSSSSSSRVSSAASALASGGFFNANALSSVISSMASQVRSTSADLSGCE 300
               ***:*******************************************************
```

Figure 5.3 Clustal alignments of most similar contigs. Identities are indicated by an asterisk. Poly-alanine stretches (red/bold) are common sources of sequence variability.

Despite 454 pyrosequencing reportedly being able to characterise viral quasispecies variations up to four times more efficiently than capillary based Sanger fluorescent dideoxy termination sequencing (Liang *et al.,* 2011), the published per-base accuracy of a Roche GS20 (pyrosequencer) is only 96% (Margulies *et al.,* 2005). Published data suggests that this could be Roche being overly pessimistic by giving results at the lower end of its feasibility scale however Huse *et al.* (2007) indicate this could be as high as 99.5% in unassembled sequences and upon removal of all reads containing one or more Ns were able to reduce this error rate from about 0.5% to about 0.25%. Furthermore, by excluding just 1% of the reads whose lengths lie outside the main distribution in addition to those with inexact matches to the primer, this error rate for the V6-tag data was reduced to less than 0.2%.

Interestingly, given the peptide sequences of the aforementioned 14 contigs (Figure 5.4), the number of varying amino acids (excluding gaps) was found to be 105 out of a total of 2100, which works out to be exactly 0.05 (or an identity of 95%). This is perhaps merely just a coincidence, but could indeed be attributed to sequencing error as stipulated by Margulies *et al.* (2005) as it is on par with their calculated error rates. Moreover, somewhat as expected, despite the motif differing from the most common motif QALLE, to being the lesser-found QVLLE, this sequence is found uniformly and unchanged throughout these 14 transcripts (barring the two truncated contigs). This conservation is perhaps due to the α-helix predicted (Challis *et al.,* 2006) and later proven (Berman *et al.,* 2000, Hagn *et al.,* 2010) to exist at the core of a homeodimer complex found at the C-terminal of the peptides (Figure 5.6).

```
isotig01583  YAAAAASFAAGR-GGYGGRGGYVAGASS-AASSSTSSSVDSSTVYSTISRLSSSSSSSRV 123
isotig05050  ALASAAASAFGS-GGYGGREYG-TSSSAAASSAASSS-VDRSTVYSTISRLSSSSSSSRV 114
isotig06627  ALASAAASAFGA-GGYDRRGPAAASSAAAASSAASSS-SVSSTVYSTISRLSSASSSSRV 98
isotig01555  ALASAAASAFGA-GGFDRRGSGGASSGAAASSAASSS-VDSSTVYSTISRLSSSSSSSRV 153
isotig03968  AAAAAAAAVSAA-QGYGSR--SGFESSAAASSAGVSSSVDISTVYSTLSRLSSPSSTSRV 123
isotig04284  ALASAAASAFGS-GGYGPRGYG----TSSSASAASSS-VDSSTVYSTVLRLSSAPSSSRV 172
isotig04370  ALASAAASAFDLGSGRYRAG--------SRAAAASSSSIDTSAIYSTISRLSSSSSSSRI 184
isotig00411  AAAAAAAAAGGR-VGYGGSGGYGSGSSS-AASSASSSVDSSTVYSTISRLSSSSSSSRV 139
isotig00412  AAAAAAAAAGGR-VGYGGSGGYGSGSSS-AASSSASSSVDSSTVYSTISRLSSSSSSSRV 137
isotig05058  ALASAAASAFDS-TGYDRHGPAIGYEDISSTSASSATSSISSTVYSTISRLSSASSSSRI 80
isotig10167  ALASAAASAFGS-GGYGRTGYGPSSSAATASSAASSS-IDVSTVYSTIPRLSSSSSSSRV 144
isotig00671  ALASAAASAFGS-GGYDRRRYGGGSSAAAASSAASSSSVDSSTLYSTISRLSSSSSSSRV 201
isotig00672  ALASAAASAFGS-GGYDRRGYGGGSSAAAASSAASSSSVDSSTVYSTISRLSSSSSSSRV 260
isotig01984  AAAAAAAAASGL-GGYGGGSGYGSGSSSTSVSSSASSSLDSSTVYSTISRLSSSSSSSRV 132
             *:.**:  .      *                 ::. ::   *::***: ****..*:*:

isotig01583  SAAASALASGGAFNSNSLSTVISALASQVRSTSADLSGCEVLVQVLLEVLSALVHILNSA 183
isotig05050  SSAASALASGGAFNANALSSVISSLSSQVRSSSSDLSGCEVLVQVLLEVLSALVHILNSS 174
isotig06627  SSAASALASGGYFNANALSSVISSLSSQVRSTSADLSGCEVLVQVLLEILSALVHILNSS 158
isotig01555  SSAASALASGGYFNANALSSVISSLASQVRSTSADLSGCEVLVQVLLEILSALVHILNSS 213
isotig03968  TSAASALASSGYFNANALSSVISSLSSQVRSTSADLSECEVLVQVLLEILSALVHILNSS 183
isotig04284  SAAASALASGGSFNANALSSVISSLASQVRSSSSDMSGCEVLVQVLLEVLSALVHILNSS 232
isotig04370  SSAASVLASGGTFNANALSSVISDLAAQVRSTSDTLSGCEVLVQVLLEVSALVHILNSS 244
isotig00411  SAAASALASGGAFNGNSLSAVISGLASQVRSTSADLSGCEVLVQVL--------------- 185
isotig00412  SAAASALASGGAFNGNSLSAVISGLASQVRSTSADLSGCEVLVQVL--------------- 183
isotig05058  SSAASALVSDGSLNTNALPSVISNLASQIRSTSSDLSGCEILVQVLLEVSALVHILNSS 140
isotig10167  SSAASALASGGAFNANALSSVISNLAYQVRSTSSDLSGCEVLVQVLLEVLSALVHILNSS 204
isotig00671  SSAASALASGGFFNANALSSVISSMASQVRSTSADLSGCEVLVQVLLEILSALVHILNSS 261
isotig00672  SSAASALASGGFFNANALSSVISSMASQVRSTSADLSGCEVLVQVLLEILSALVHILNSS 320
isotig01984  SAAASALASGGAFNSGSLSSVISSLASQVRSTSEDLSGCEVLVQVLLEVLSAVVHILNSS 192
             ::***.*.*.* :* .:*.:*** :: *:**:*   :* **:******::**:******:

isotig01583  DIGQVDLKSISSASDLVSRSFYALA-N-----------------------VAC-IFEF 216
isotig05050  NIGPIDLSSVTSASNIVSNSLYALA-SKVC---------------------NLIFSER 210
isotig06627  NIGQVDLSSISSASNLVSNSLYALA-R-FL---------------------SICLSY-C 193
isotig01555  NIGQVDLSSISSASNIVSNSLYALA-------------------------SICLSC-C 245
isotig03968  SFGQIDLSSVNSASNIVSNSLYAIA-------------------------NFFDYIIL 216
isotig04284  NVGQIDLSSVSSASNIVSNSLYALA-SEVC---------------------NLIFSEL 268
isotig04370  NIGQIDLTSVNSASDTVSRSLFALA-------------------------NKLHYLFL 277
isotig00411  ------------------------------------------------------------
isotig00412  ------------------------------------------------------------
isotig05058  NVGPIDLSSINSSSNVISNSLYSLV-------------------------IRFIKFSI 173
isotig10167  NIGQIDMNSVSSASSIVSNSLYALA-SKIC---------------------ILIFSER 240
isotig00671  SIGQVDLSSISSASNIVSNSLYALA-MAFY-IFVYLVNP-VFWK-RYFLHCIILSVSFLC 317
isotig00672  SIGQVDLSSISSASNIVSNSLYALA-MAFY-IFVYLVNP-VFWK-RYFLHCIILSVSFLC 376
isotig01984  DIGQIDLSSVSSAADLVSRSLYSLA-N-----------------------ILD-IFEI 225
             ..* :*:.*:.*::. :*.*::::.
```

Figure 5.4 Clustal alignments of the 14 translated MaSp contigs. Species divisions as per Figure 5.3. Amino acid substitutions different from the majority are indicated by a dark-greyed box. Equal identity indicated by a light-grey box. QVLLE motif (akin to the QALLE motif), highlighted in red. Asterisks indicate identity while colons and dots represent decreasing identity respectively.

(A)

```
vGkTChrCnvTnTATTCnACTvTkyydmGdyTGTCvwsdbCTyCnTCwwCrTCwAGArThwCdkChGChG
CTTCTGyyyTnGyhTCkrryGGTdhnTTbAAydsyrrykCnTTGyCbdCnGTwATwTCwrvThTskCdkm
yCArrTyCGhTCnwCvTChkmnrmwmTbTCyGrdTGyGAArTbCTyGTbCArGTkCTny
```

(B)

```
GSTxYxTIxRLSxSxSSSRVSSxASAxxSGGxxNANxLSSVIxSxxSQVRSxSxDxSGCEVxVQVx
```

Figure 5.5 Nucleotide (A) and amino acid (B) consensus sequences as per the 14 untranslated and translated contigs. Colour coding is the same as that in Figure 5.2. Lower case letters signify ambiguities.

(A)

(B)
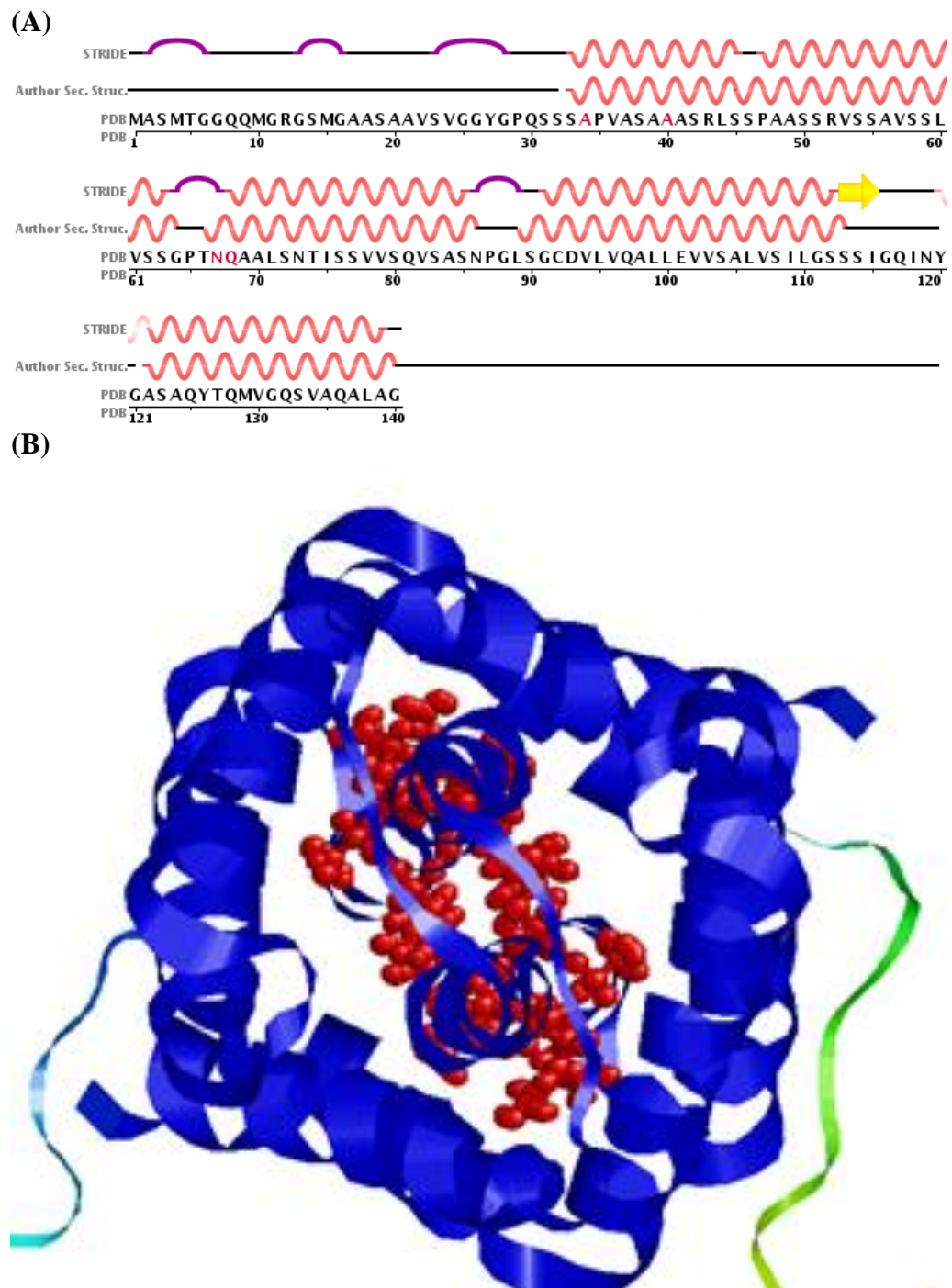
Figure 5.6 (A) Secondary structure predictions according to STRIDE (Frishman and Argos 1995) and the authors (Berman *et al.*, 2000) of the 'QALLE motif' as described by Challis *et al.* (2006). Colours indicate secondary structure with red representing α-helices, purple for β-turns and yellow for strands.

(B) A ribbon diagram of the C-terminus of ADF - 3 (PDB ID 2KHM) from the garden spider (*Araneus diadematus*) as discovered by Hagn *et al.* (2010). The red molecular structure indicates the α-helices of the respective 'QALLE domain' with the surrounding α-helices as blue ribbons.

189

5.4.1 N-TERMINAL DOMAIN

Using apparent motifs seen in Motriuk-Smith *et al.* (2005), N-terminal searches were performed to test if this region could be found within the genomes (Figure 5.7). The initial search of MAFASS (atggcnttygcnwsnwsn) failed to retrieve any results so the sequence was truncated by a single serine at the C-terminal end. This new search retrieved three hits, one of which was the only N-terminal sequence amongst the three genomes. This sequence, found within *S. mimosarum*, is definitely recognisable as an N-terminal domain but would appear to be an intermediate between some of the more radically divergent sequences. It still has conservation in 'key' residues such as all the phenylalanines (F), the MAFAS region and the final threonine-threonine-glycine motif towards the C-terminal end, but despite retaining many of the acidic residues, seems to lack many of the others, or has exchanged them for another (i.e. aspartic acid for glutamic acid). It is not clear what function these particular 'sparsely distributed' amino acids serve, but nevertheless, they are conserved and so must serve some sort of purpose. Only with structural analysis using X-ray crystallography will this question perhaps be answered.

(A)

```
Lg.MaSp1   PTKLQALDMAFASSVAEIAAVE--GQNIGVTTNAISDALTS  76
At.MaSp2   KSKLQALNMAFASSMAEIAVAEQGGLSLEAKTNAIASALSA 123
Nim.MaSp2  KSKLQALNMAFASSMAEIAAVEQGGQSMDVKTNAIANALDS 122
```

(B)

```
gi|70913273 | MaSp2  [A.trifasciata]      QLAESFISRFLRF-IGQSGAFSPNQLDDMSSIGDTLKTAI-EKMAQSRK-  80
gi|150416778| MaSp2  [L.hesperus]         ENADAFIGAFMNA-ASQSGAFSSDQIDDMSVISNTLMAAM-DNMG--GR-  79
gi|193506891| MaSp1A [N.clavipes]         ELADAFINAFMNE-AGRTGAFTADQLDDMSTIGDTIKTAM-DKMARSNK-  79
gi|115635734| MaSp1  [E.australis]        GLAENFMNSFMQGLSSMPG-FTASQLDDMSTIAQSMVQSI-QSLAAQGR-  79
gi|303307750| Fib1   [B.californicum]     AKGKKFLSTFLDYALD-HGLFPQQERDDLEAISQNLIPVFRKTMD-SGG-  82
gi|303307752| MaSp1  [K.hibernalis]       KTAEIFISKFISAILD-SNAFTREQKEDMMSIGETIIPAM-EKMSGSSK-  80
gi|164709244| MaSp1  [L.geometricus]      ANADAFINSFISS-AQNTGSFSQDQMDDMSLIGNTLMTAM-DNMG--GR-  78
gi|70913024 | MaSp2  [N.madagascariensis] ATADAFIQNFLGAVSG-SGAFTPDQLDDMSTVGDTIMSAM-DKMARSNK-  79
gi|303307754| MaSp   [D.canities]          IMAEDFMNKFTNQLAN-SPYFSSQQKEDMSSIKDELISVI-ESMDSAHK-  83
gi|303307772| MaSp   [A.aperta]            ATAESFISSVMSSVANQ-GCLSYDQIDDMQAVGDTMLATM-DNLVRSGK-  78
T.mimosarum                              NSAQTFATSFVNYIVA-SGVFPEQQEEEDMKEFIETLSMAV-TSLT-NNKW  91
                                           .. *   .            :. .: :*:   . : :    .  .:

gi|70913273 | MaSp2  [A.trifasciata]      SSKSKLQALNMAFASSMAEIAVAEQ-GGLSLE--AKTNAIASALSAAFLE 127
gi|150416778| MaSp2  [L.hesperus]         ITPSKLQALDMAFASSVAEIAVAD--GQ-NVG--AATNAISDALRSAFYQ 124
gi|193506891| MaSp1A [N.clavipes]         SSKGKLQALNMAFASSMAEIAAVEQ-GGLSVD--AKTNAIADSLNSAFYQ 126
gi|115635734| MaSp1  [E.australis]        TSPNKLQALNMAFASSMAEIAASEE-GGGSLS--TKTSSIASAMSNAFLQ 126
gi|303307750| Fibn1  [B.californicum]     NAAAKMKALNMAFASSIAEIAVQEG-GAGSIE--EKTQAVSEALAHAFLQ 129
gi|303307752| MaSp1  [K.hibernalis]       SIHAKLTALNMAFASSVAEIAVVEE-GGSDIN--EKTYAIVAALNQAFLD 127
gi|164709244| MaSp1  [L.geometricus]      ITPSKLQALDMAFASSVAEIAASE--GG-DLG--VTTNAIADALTSAFYQ 123
gi|70913024 | MaSp2  [N.madagascariensis] SSKSKLQALNMAFASSMAEIAAVEQ-GGQSMD--VKTNAIANALDSAFYM 126
gi|303307754| MaSp   [D.canities]        SSAAKLQAMNMAFASAIADIAATEA-YGADIS--LETSAIANALSEAFLQ 130
gi|303307772| MaSp   [A.aperta]          SSSHMLKAMNMAMGTSIAEIVA--D-GGGNLG--SKVSCISNALSSAFLQ 123
T.mimosarum                              ASRAKIEALSMAFASAMAELIVIEDDDGENVSTDVKVKVISDGLGQAFKE 141
                                          : *:.**:.::*::  .        .:    .  :  .:  **

gi|70913273 | MaSp2  [A.trifasciata]      TTGYVNQQFVNEIKTLIFMIAQ---AS--SNEISG----------SAAA- 161
gi|150416778| MaSp2  [L.hesperus]         TTGVVNNQFITEISSLIGMFAQ---VS--ANEVSY-----------TS- 156
gi|193506891| MaSp1A [N.clavipes]        TTGAANPQFVNEIRSLINMFAQ---SS--ANEVSYGGGYGGQSAGAAAS- 170
gi|115635734| MaSp1  [E.australis]       TTGVVNQQFINEITQLVSMFAQ---AG--MNDVSA----------SAS- 159
gi|303307750| Fib1   [B.californicum]    TTGSVNIQFIKEIRALITLFAKEGQDNETENEIPT----------QQAY 168
gi|303307752| MaSp1  [K.hibernalis]      TTGKVNKQFIAEIRDLVKMFAS---AN-EENEIGA----------ALS- 161
gi|164709244| MaSp1  [L.geometricus]    TTGVVNNRFISEIRSLISMFAQ---AS--ANDV-Y-----------AS- 154
gi|70913024 | MaSp2  [N.madagascariensis]TTGSTNQQFVNEMRSLINMLSA---AA--VNEVSYGG---GASA-AAAT- 166
gi|303307754| MaSp   [D.canities]        TTGVVNKRFISEIQELIYMFAQD--ASVQSNEIAS----------SSS- 166
gi|303307772| MaSp   [A.aperta]          TTGSVNTQFVNEIVSLISMFAQ---AD--TNEVGVG---------SGSG 158
T.mimosarum                             TTGSVNEGFIEEIQELLGMFAH---AV--VSGINE----------ADST 175
                                        *** .*  *: *: *: :::          . :              :
```

Figure 5.7 (A) Alignment of N-terminal domains from Motriuk-Smith *et al.* (2005). (B) Corresponding domains of most homologous domains from ten other species with gi numbers. *T.mimosarum* translation below in bold.

191

5.4.2 P-REGION

These regions, which the author has termed 'p-regions' are ambiguous stretches of nucleotides that yielded significant bit-matches from BLAST searches (when the $(GA)_n$ motif was mined) but which should still be considered as suspect. The nomenclature is justifiable in that when the sequences are translated, the database searches result in significantly high (e.g. $4 E^{-51}$) matches to database sequences highly rich in prolyl residues. However, and this is where the initial positive result is perhaps nullified, the sequence only yields a positive result when aligned with a backwards frame of the corresponding silk protein and subsequently provides a much lower similarity significance (*E*-value). An example is illustrated in Figure 5.8.

```
gb|DQ399326.1| Deinopis spinosa clone DM8 fibroin 1a mRNA, partial cds
Length = 2628

Score = 198 bits (427),
Expect = 4-51
Identities = 83/126 (66%),
Positives = 83/126 (66%),
Gaps = 0/126 (0%)
Frame = -2/-2

Query  379   PTPAPAPMPTPAPAPAPTPTPAPAPAPTPAPASKPAPAPTPKPAPAPTPAPAPTPAPALT   200
             PTPAPAP P P PAP P P PAP PAPTPAP   PAPAPTP PAPAPTPAPAP P PA T
Sbjct  1949  PTPAPAPAPAPTPAPTPAPKPAPTPAPTPAPEPAPAPAPTPAPAPAPTPAPAPAPKPAPT  1770


Score =  176 bits (378),
Expect = 2e-44
Identities = 83/126 (66%),
Positives = 86/126 (68%),
Gaps = 0/126 (0%)
Frame = +3/+1

Query  3     AGAGGGAGTGAGAGAGAGVGAGAGAAIDAGVGTEVDVAAEAEAEAEAEAEAEAEAEAEAE  182
             AGAG GAG GAGAG+GAG+GAGAGA   AG G  V    A A   A A A    A A A
Sbjct  1573  AGAGYGAGAGAGAGSGAGLGAGAGAGFGAGAGAGVSAGAGAGIGAGAGAGVGAGAGAGFG  1752
```

Figure 5.8 Reverse and forward frame alignments of translated contigs with closest homologous database retrieval (*Deinopis spinosa*).

Twelve such contigs spread across the three species yielded a result as described above and as expected, due to proline having the codon CCN, the cytosine content was correspondingly higher, e.g. in one instance the ratio was 576:1118:479:455 for A, C, G and T respectively. Naturally this is for the reverse complement of the actual sequence (455:479:1118:576). Another curious observation is the one hit that is always the highest: gb|DQ399327.1| *Deinopis spinosa* clone DS19 fibroin 1b mRNA, partial cds. For all 12 similar sequences, this exact database hit (*D. spinosa* clone) always ranked highest. The question is, is it the proline motifs that makes it more flagelliform-like within the target sequence, or is it the 'forward-frame' repetitive $(GA)_n$ motif, which albeit less homologous to database sequences, is more recognised as a silk motif found in many MaSp and MiSp genes (Figure 5.9)?

```
gi| 89113995| D.spinosa   GYG-GGAG-YGSGAGAGSGAGAGAGYGAGAGSGTG   480
gi|149929453| N.cruentata GAGVGGAGGYGRGAGAGAGAAAGAGAGAAAGAGAG   618
gi| 89114007| U.diversus  AASSAGAG-YGGQAGYGQGAGASAGA-AAAGAGAG   494
gi|  2605797| N.clavipes  GA--GGAGGYGRGAGAGAGAAAGAGAGAAAGAGAG   564
                          .   .*** **  ** * **.*.**  *.**:*:*
```
Figure 5.9 Clustal alignment of four silk $(GA)_n$ motifs. Asterisks indicate identity while colons and dots represent decreasing identity respectively.

One other interesting factor to note is the ubiquitous inverted repeats similar to those found in transposons. When translated in a forward frame, not only does the repetitive $(GA)_n$ motif occur as expected, but it is always interrupted by an alanine-glutamic acid $((AE)_n)$ stretch (Figure 5.10).

AGAGGGAGTGAGAGAGAGVGAGAGAAIDAGVGTEVDVA<u>AEAEAEAEAEAE AEAEAEAEAEAEAEAEA</u>GVSAGAGVGAGAGVGAGAGLGVGAGAGLEAGAGV GAGAGAGVGVGAGAGAGVGIGAGAGVGGATFGIGPGGSLE

Figure 5.10 Translated forward frame of proximal P-region contig. $(AE)_n$ stretch is underlined.

A complete example of this contig, along with neighbouring regions is shown below in Figure 5.11.

```
>isotig01034  gene = isogroup00208  length = 812
gtgctggagcaggaggcggtgctggaacaggagctggagctggagctggagctggtgttggtgctggagctggcgcagc
aatcgacgctggcgttggaacagaagttgatgttgcagcagaagcagaagctgaagcagaagctgaagcagaagctgaa
gcagaagctgaagcagaagctgaagcagaagccgaagctggtgttagtgctggggcaggagttggtgctggagctggag
taggagctggtgctggcttaggagtaggagctggtgctggcttagaagcaggagctggagttggtgccggtgcaggagc
tggagtaggagttggagctggagctggagcaggagttggcataggtgcaggggctggagttggtggtgctacttttgga
atagggcctggcggcagcttagaaagttttttgtagacagttgattggaaattcataacggaattttacgtctcttgtgc
tatattattacgctgtattttgtataaattttataatattttcattattatgatttggatattattcgaatgttctatt
atattttttaaatcaagaaatgcttattagtctataatgtcatactatttcccgatatatttgttccaaattcttattt
cttacattttatacaatgtttttattttagaggtaaaactatttgaaatgcactttaatgtaaggttaaattacagctaa
atatctttattttgtaaaaattgtcatatgcagaataaagtctattattttgcgtctttgtctgaatttcaatgtaatt
tctttgatacactcaataaaac
```

Figure 5.11 Complete isotig as per raw data. Black indicates the $(GA)_n/(AE)_n$ stretch while red is the non-repetitive region.

This unusual $(GA)_n - (AE)_n$ region in the forward frame is also flanked by another unusual element. This region (shown in red in Figure 5.12), is a non-repetitive stretch of amino acids, which when translated produces nothing visibly coding (Figure 5.12).

VFVDS*LEIHNGILRLLCYIITLYFV*IL*YFHYYDLDIIRMFYYIF*IK KCLLVYNVILFPDIFVPNSYFLHFIQCFILEVKLFEMHFNVRLNYS*I SLFCKNCHMQNKVYYFASLSEFQCNFFDTLNK

Figure 5.12 Translated non-repetitive region from isotig01034 as per Figure 5.11.

What it does produce is a region far more homologous to the other P-region contigs than that of the $(GA)_n - (AE)_n$. An alignment of the $(GA)_n - (AE)_n$ motifs with other P-region contigs results in a sparsely homologous 'island' patterning with lots of gaps inserted to make the alignments stick. The non-repetitive flanking region has virtually identical sister sequences when compared to the other P-region contigs (Figure 5.13). Likewise, the A:C:G:T ratios are vastly different with the $(GA)_n - (AE)_n$ motif being high in guanine, whereas the NR-flanking region has an adenine/thymine (A/T) content almost

three times that of guanine/cytosine (GC). Here it would be negligent not to relate this set of contigs to flagelliform genes. The reverse-complement of the P-contigs have a proline content of 35% and up to 55% within the flag gene (Figure 5.14). Both are interspersed with non-repetitive elements. The glycine content in the P-contigs is 43% while in the flag gene it is 41%. It is still not clear whether this family of contigs indeed codes for silks but the evidence is intriguing. At the very least, there are interesting similarities to transposable/repetitive elements that warrant further investigation.

```
isotig01034    AAGCAGAAGCCGAAGCTGGTGTTAGTGCTGGGGCAGGAGTTGGTGCTGGAGCTGGAGTAG 240
isotig01035    AAGCAGAAGCCGAAGCTGGTGTTAGTGCTGGGGCAGGAGTTGGTGCTGG----------- 181
isotig01432    GAGCAG--G----AGCTGGTG------CTGGAGCAGGAATTGGTGC-------------- 126
isotig03305    --GCAG--G----AGCTGG-----------AGCAGGAATTGGTGC-------------- 72
               ****  *    ******            ****** *******
```

```
isotig01034     GTTTTTGTAGACAGTTGATTGGAAATTCAT-AACGGAA-TTT-TACGTCTCTTGTGCTAT 477
isotig01035     GTTTTTGTAGACAGTTGATTGGAAATTCAT-AACGGAA-TTT-TACGTCTCTTGTGCTAT 405
isotig01432     GTTATTGTAGACAGTTGATAGGAAATTTTTGAA-GGAAACTTGTATGTCTCTTCTGTTAT 262
isotig03305     GTTATTGTAGACAGTTGATAGGAAATTTTTGAA-GGA-CTTGTATGTCTCTGGTGTTAT 201
                *** *************** *******  * ** ****  ** ** ******  ** ***
isotig01034     AT-TATTACGCTGTATTTT-GTATAAATTTTATAATATTTTCATTATTATGATTTGGATA 535
isotig01035     AT-TATTACGCTGTATTTT-GTATAAATTTTATAATATTTTCATTATTATGATTTGGATA 463
isotig01432     AA-TCTTATGATG-ATTTTTGTATAAATTTTATAAT-TTTT--TT--TATGATTTCTATA 315
isotig03305     AAATATTATGATG-ATTTTTTTATAAATTTGATAAT-TTTT--TC--T--GATTTGTATA 253
                *  * *** * ** *****  ********* ***** ****  *   *  *****  ***
```

A/C/G/T ratio = **95/64/182/80 black** A/C/G/T ratio = **124/45/48/179 red**

Figure 5.13 Clustal alignments of proximal (GA)$_n$ - (AE)$_n$ region in black and non-repetitive region in red. Respective ACGT ratios are given below. Asterisks indicate identity while colons and dots represent decreasing identity respectively.

```
MGKGRHDTKAKAKAMQVALASSIAELVIAESSGGDVQRKTNVISNALRNALMSTTGSPNEEFVHEVQDLI
QMLSQEQINEVDTSGPGQYYRSSSSGGGGGGGGGGPVITETLTVTVGGSGAGQPSGAGPSGTGGYAPTGYA
PSGSGPGGVRPSASGPSGSGPSGSRPSSSGSSGTRPSANAAGGSSPGGIAPGGSSPGGAGVSGATGGPAS
SGSYGSGTTGGAYGPGGGSEPFGPGAAGGQYGPGGAGPGGAGAYGPGGVGPGGAGPGGYGPGGAGPGGYG
PGGAGPGGYGPGGAGPGGYGPGGAGPGGYGPGGAGPGGYGPGGAGPGGYGPGGSTGGAGPGGYTPGGAG
PGGYGPGGYGPGGSGPGGAGSGGVGPGGYGPGGAGPGGAGPGGAGPGGAGPSGAGPGGAGTGGAGTGGAG
PGGAGPGGAGPGGAGRGGAGRGGAGRGGAGRGGAGRGGAGGAGGAGGAGGAGGAGGAGGS
GSTTIIEDLDITIDGADGPITISEELTIGGAGAGGSGPGGAGPGGVGPGR
```

Figure 5.14 Translated gb|AAF36091.1| flagelliform silk protein (*Nephila inaurata madagascariensis*) Length = 1884. Red indicates N-terminal domain, orange - an intermediary domain and the repetitive domain in green. Both repetitive elements are underlined (GPGG(X)$_n$) and NR nucleotide stretch.

5.5 CONCLUSIONS

In this study a total of 29 apparent MiSps, 8 MaSps and 4 pyriform silk sequences were extracted across the three 454 pyrosequencing transcriptomes of *S. tentoriicola*, *S. mimosarum* and *S. lineatus*. These contigs, which almost certainly pertain to silk genes are still confounded by both the search criteria utilised and the database limitations. Searches, like those in this instance using Microsoft Office Word package employ the use of comparative sequence homologies to identify motifs. The robustness of this approach is therefore dependent on being able to correctly identify those motifs and being able to align them correctly to database voucher models. As illustrated here, there is still yet insufficient known about the range of amino acid variability within silk motifs but likewise, the databases are so poorly stocked that invariably a string of false positives will result. In a significant number of resulting alignments, not only did an obviously 'forward frame' silk yield a higher bit score in the frame 4 - 6 orientations, but they were yielding secondary results compared to non-silk proteins due to high incidences of 'silk-like' amino acids such as alanine, glycine and proline (Figure 5.15).

(A)

```
gb|AY174110.1|  Araneus ventricosus major ampullate gland dragline silk
protein-1
(F1) mRNA, partial cds
Length = 1744
Score = 56.0 bits (119),
Expect = 6e-15
Identities = 30/81 (37%),
Positives = 33/81 (41%),
Gaps = 0/81 (0%)
Frame = -2/-3

Query  275  PRPILPPPQGLIITPIPPPGPPPPHILAMPRPPMPPVIGPPGTSFVPPMHPIAPPPPPQQ  96
            P P  PPP      P PPP P PP     P PP PP    P  +  PP  P APP P
Sbjct  260  PAPPAPPPTPPAPGPTPPPIPAPPGPPGAPGPPGPPGPPGPPGAPGPPPGPKAPPAPGP*  81

Query  95   TQTLVPGKTAAAAPSEDEPAP  33
            +    PG      P+   P P
Sbjct  80   SPPAPPGP*IPPGPAPPGPTP  18

--------------------------
```

196

(B)

```
emb|AJ508925.1|  Acanthoscurria gomesiana mRNA for acanthoscurrin 1 precursor
(acantho1 gene)
Length = 882
Score = 65.8 bits (141),
Expect = 6e-16
Identities = 35/93 (38%),
Positives = 37/93 (40%),
Gaps = 0/93 (0%)
Frame = -2/-2

Query  386  QPPLPTTATPTVTVIQQRPPLPPIPPIPMPPVSAILPPRPILPPPQGLIITPIPPPGPPP  207
            +PP P    P+       PP P PP P PP    P P   PP     P PPP  PP
Sbjct  434  KPPPPRPPPPSPFPPPSPPPPRPPPPSPPPPSPPPPSPFPPPSPPPPRPPPPSPPPPSPP  255
```

Figure 5.15 Comparison of two BLASTp retrievals for a single p-region contig.

In this example, the same sequence in one instance resulted in two hits of very similar bit-scores solely due to each individual's proline content. These hits, albeit both significant ($x > E^{-5}$) are probably both wrong and are identifying the only sequences available within the database for the given organism. This example also illustrates another of the P-region sequences' confounding issues in that it retrieves sequences out of frame to the respective peptide of origin, in this instance, for Figure 5.15 (A), a silk whose $(GA)_n$ motif and correspondingly high glycine content (G = GGN, A = GCN), resulted in a high probability of reverse-prolyl incidences.

The absence of both flagelliform and recognisable glues is consistent with the biology of the stegodyphid species. Unlike most orb-weavers, which spin viscid capture threads composed of a flagelliform core surrounded by aqueous glue droplets, *Stegodyphus* spiders weave tubular webs composed of cribellate silk. This silk relies on hydrostatic interactions and entanglement as opposed to the molecular nanosprings hydrated by aqueous glues of flagelliform fibres (Blackledge and Hayashi 2006). This is reaffirmed by the pseudo-communal behaviour of this species as entangled prey would probably

require more support from conspecifics than those relying solely on glues and radial webs, which appear to be far more effective in aiding prey subduing.

Perhaps one of, if not the most confounding issues with current database searches pertaining to silks and no doubt numerous other incidences within the field of bioinformatics is the lack of representative sequences with which to compare a query sequence to. Prior to these data, the spider database was vastly under-nourished with other spider sequences and has until recently been added to with a very restricted range of proteins. Two thirds of the meagre 18 000 sequences deposited therein are related to phylogenetics while the remaining ~4 000 are divided almost equally between venoms, silks and histone-related proteins (Figure 5.16). A small fraction of these ('Rest') are artificial constructs based on original species.

In conclusion, a more 'sourced' database on spiders and likewise, any taxon representative is necessary if a BLAST search is to be at all effective. The uses if achieved are many; from evolutionary and comparative studies to selecting interesting *de novo* genes for further functional analyses. Due to insufficient database sequences prior to these, in-depth analyses of these genomes and correct assignment of each individual contig is key to aid in the annotation of future arachnid genomes.
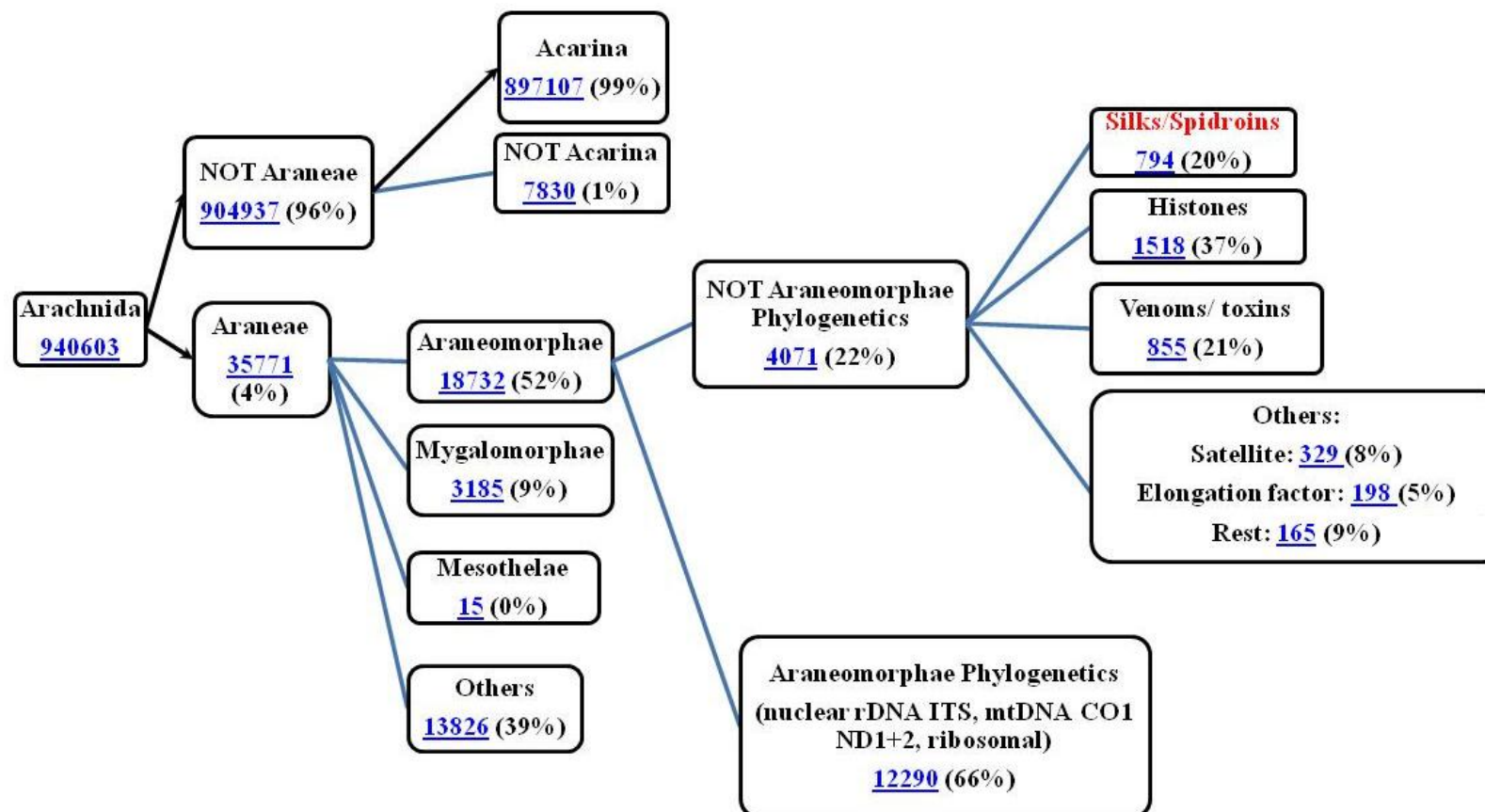
Figure 5.16 (NCBI) Nucleotide: Core subset of nucleotide sequence records for the Arachnida with respective allocations to different taxa/protein representatives. Blue numbers indicate total hits with that given search term while percentage is that of the previous tree root's total hits (as of Feb 2012).

199

6   A NOVEL METHOD FOR SEXING THERAPOSID SPIDERS USING TIBIAL SPINES

6.1   INTRODUCTION

The infraorder Mygalomorphae (Orthognatha) contains taxa with a hugely untapped potential for behavioural and genetic research due to the vast numbers of species available, interestingly, much of the accumulated data concerning their care, habitats and toxicological importance has come from the myriad amateur enthusiasts. Assessing primary sex ratios would be useful to gain insight into areas including population demography, adaptive sex ratio modification and sex-biased developmental mortality.

Sexing of adult mygalomorphs is relatively easy due to sexual dimorphism (palps and tibial hooks) which arise after the penultimate moult, however sexing of juveniles poses difficulties because there is no discernable sexual dimorphism in most species prior to their final moult. The relative sex ratios of individuals lend themselves to a variety of studies including population phenology, sex specific growth patterns, interspecies dimorphism and sex ratio modification. Sex ratio modifications have been studied in a limited number of species (Nager *et al.*, 1999; Austad and Sunquist 1986; Gunnarsson and Andersson 1992; Uhl and Gunnarsson 2001). The non-social species, *Pityohyphantes phrygianus* (Koch 1836) for example requires two years to grow to a size of 10 mm, twice the time for *Pterinochilus murinus* (Theraphosidae) to reach 12 cm. Karyotyping has proven to be effective at

sexing spiders (Aviles and Maddison 1991) but the precise technique is species-specific while morphological features are subjective at best particularly within the Theraphosidae where there is an apparent monomorphy during the early instars, unlike many of the Araneomorphae (Bristowe 1929; Arak 1988; Downes 1981). The most common method of sexing theraphosid spiders (detailed below) is the microscopic examination of exuviae (shed skins) (Hancock and Hancock 1994) but even this is troublesome as it relies on obtaining an intact exuvium, which is not easy in cases of burrowing, aggressive or destructive species. Ideally, a method is needed that can accurately sex spiders of all ages/sizes without the need of sophisticated/ expensive equipment, exuviae or dissection.

The objective of this study was to determine whether there was indeed a sexual dimorphism with regards to simple physical dimensions as previously and universally hypothesised and to examine the hypothesis that the tibial spurs on leg - 4 are sexually dimorphic with relation to the overall dimensions of the spider (Figure 6.1). This hypothesis was formulated based upon the author's personal visual observations of a range of male and female specimens in his care. Samples were examined in the Natural History Museum, London encompassing a variety of taxa. Here it is reported that these dimensions can indeed be used to sex individuals and are statistically significant down to $p <$ .001 with regards to simple dimensions and even more statistically significant when utilising the spur.

6.1.1 SEXING OF MATURE SPIDERS

By far the easiest period in a theraphosid's development in which to sex an individual is when it is an adult due to many species having a pronounced sexual dimorphism. In males, the most obvious distinguishing features are the pedipalps (palps). These are leg-like appendages in which the coxa has evolved to form the maxilla (the mouthparts on either side of the labium) (Figure 6.1). The palps are used in both courtship (Stoltz *et al*., 2009) and sperm transfer (Bukowski and Christensen 1997) and although in some species they can be recognised as being sexually dimorphic prior to maturity (Mahmoudi *et al*., 2008), in theraphosids at least, thus far no statistical data have been collected to support these observations.

The end segments of the palps are modified into an intricate structure resembling a tear drop in species such as *Brachypelma smithi* to a flagelliform structure in *Iridopelma seladonium* (Smith 1993) called the embolus. This acts very much like a syringe and is used to pick up and 'inject' sperm into the epigastric furrow of a female. The reservoir of the embolus is called the bulb. This is attached to the rest of the palp by a limber articulation joint enabling directed movement. These structures fit into a snug groove called the alveolus, which lies within a modified tarsus that as a whole is often referred to as the cymbium. The entire structure is only found in males and in large species such as theraphosids can readily be recognised at a distance making identification straightforward.
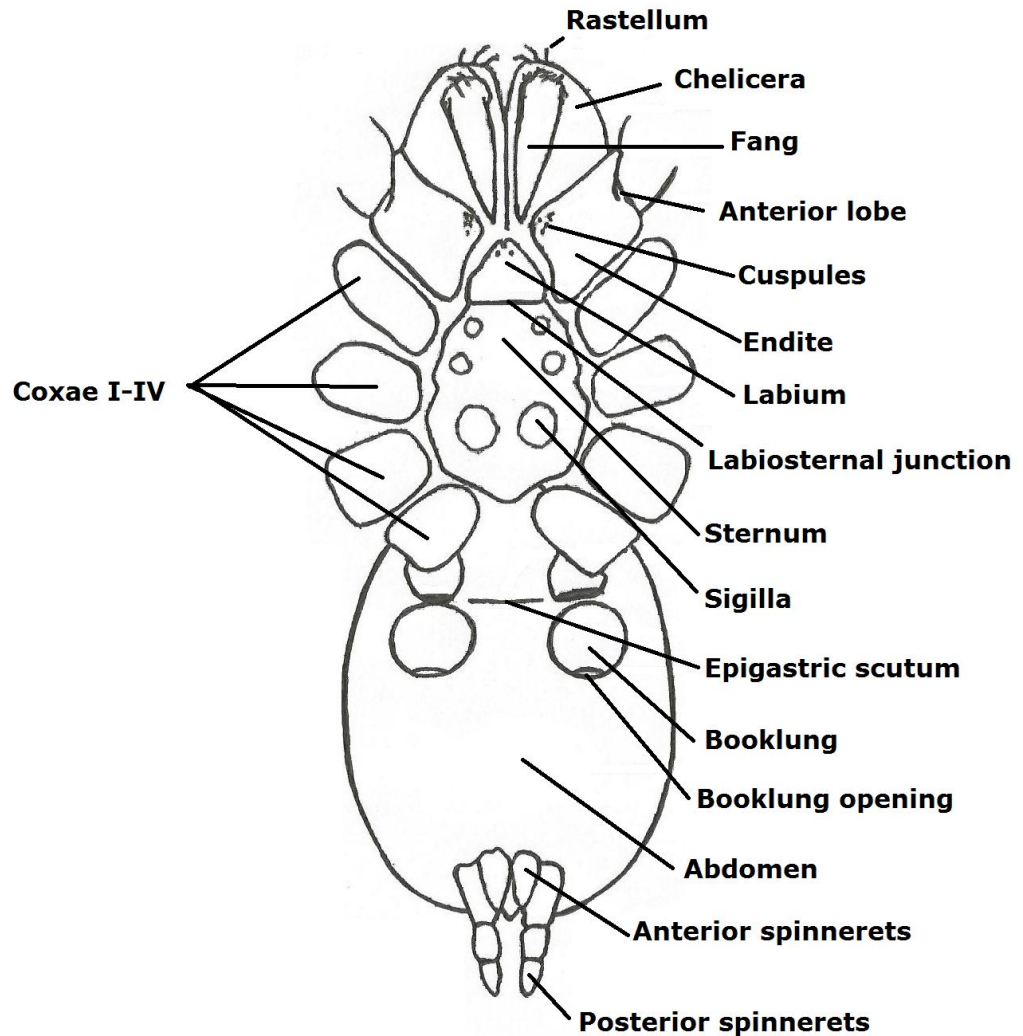
Figure 6.1 External anatomy of a theraphosid. Reproduced from Dippenaar-Schoeman and Jocqué (1997). Length of the prosoma was measured on the dorsal side, ignoring front protruding chelicerae to the pedicel (prosomal-opisthosomal junction), see Figure 2.1.

In addition to the palps, another characteristic possessed by many genera (e.g. *Avicularia, Brachypelma and Pterinochilus*) but absent in others (e.g. *Poecilotheria* and *Theraphosa*) are tibial spurs (Perez-Miles 1996). These are hook-like structures found on the tibia of mature males and are used for manoeuvring the female by interlocking them with her chelicerae during mating (personal observation; Costa and Perez-Miles 1992). It is unclear why these are only present in certain species and there appears to be no Old-

203

World/New-World divide between them either as they are present in the

Brachypelma (New World) and *Pterinochilus* (Old World). There are also

multiple reports in which some individuals (most likely due to a mutation)

possess two hooks per tibia (Patrick Mumford, private hobbyist, personal

communication). Despite not being a common occurrence by any means, this

morphological mutation has been witnessed enough to suggest perhaps an

autosomal dominant mutation in a single gene similar to polydactyly in

humans. The high frequency (i.e. it has been seen enough to have been

recorded) could likewise indicate a relative sexual advantage for males as it

perhaps increases the chance of restraining the female's chelicerae, thus

improving the male's chance of survival.

Figure 6.2 Dual tibial hooks (*Aphonopelma chalcodes*). Ollie Meidinger, American Tarantula Society

6.1.2 SEXING OF IMMATURE MYGALOMORPHS

6.1.2.1 Phenotypic variation

The above definitive characteristics are found solely in adult males and can be utilised to identify sex without comparison with another individual. However, there are no observable features on adult females that distinguish them from sub-adult males or even sub-adult females despite easily being used for sex determination for immature araneomorphs (Jocqué 1981).

6.1.2.1.1 Exuviae

Immediately after moulting, the exuvium (shed skin) is pliable and allows for manipulation for another method of sexing. This is because although much of the exoskeleton is dissolved and resorbed prior to moulting some internal non-digestable parts such as the spermatheca (sperm storage organ) and part of the female cuticle are not (Galiano 1984; Stradling 1978). The spermatheca is located between the anterior pair of book lungs and internally is often intricately shaped (Coyle *et al*., 1983). The males have no such structures and thus this sexual dimorphism lends itself to use for sexual identification. In some species the spermatheca is shallow making identification difficult but it is sometimes possible from around the fifth or sixth moult by examining the exuvium under a stereomicroscope (personal observation). Externally, the structure looks quite similar for both sexes and is composed of the epigyne (the slit into which the sperm are transferred) and a groove posterior to this called the epigastric furrow (Foelix 1996).

Unlike the aforementioned methods of sexing relying on external morphology, this can only be performed on a shed skin (or a dissection).

Figure 6.3 *Brachypelma boehmei* cast skin showing fused spermathecae of female (a) absent in the male (b) between the anterior pair of book lungs. Reproduced with permission (Guy Tansley).

6.1.2.1.2 Epiandrous fusillae

A less frequently utilised method of sexing makes use of the epiandrous fusillae (spinnerets, Marples 1967, Figure 1.11). All male mygalomorph spiders (tarantulas and their kin included) have a spinning apparatus with microscopic spigots (fusules) located on the underside of their abdomen (see Marples 1967). To the author's knowledge, these do not seem to appear in araneid spiders. These epiandrous fusillae are thought to be employed during the construction of the sperm web, the location of which was more accurately depicted by Melchers (1964) as being located anterior to the epigastric furrow (in front of the first pair of book lungs). Under light magnification, these hair-like structures have a semicircular or even triangular shape and in some species can be the same colour as surrounding hairs. Here, although reports have stated that the epiandrous fusillae have been observed on the exuviae and accurately used to sex spiderlings as small as 1 cm leg span (Hart and West 1997/personal communication; http://www.birdspiders.com/faq_sex.php), suitable manipulation of a live specimen of this size makes assessment virtually impossible. Practically, this technique can only realistically be utilised on specimens from non-aggressive species that are large enough to be manipulated (~4 - 5 cm leg span).

6.1.2.1.3 Relative dimensions of body parts

There are also several other methods that are reported as being indicators of the sex of a spider prior to maturity. One such indicator is the relative size of the animal. Very often the females are heavier with a much larger abdomen, several times the length and width of the male's whereas the males appear far thinner and the legs more wiry. The size discrepancy can be quite extreme in species such as *Theraphosa blondi* where an engorged/gravid female can have an abdomen the size of a chicken egg while the male's is the size of a grape. Size is influenced by how much the spider has been fed and watered or by the number of moults that it has gone through.

Figure 6.4 Sexual dimorphism in the adult tarantula *Poecilotheria formosa* - female left, male right. (Guy Tansley, reproduced with permission)

6.1.2.1.4 Chelicerae size

This is a trait only observable upon comparison to another individual of the opposite sex. The females tend to have more robust/bulky chelicerae while the males' are shorter and thinner (apparent in Figure 6.4) but this statistic has yet to be verified scientifically.

6.1.2.1.5 Colouration

In a small number of cases this is a very useful identification method as some species are extremely sexually dichromatic (e.g. *Haplopelma lividum* and *Poecilotheria metallica*), often with the females being vivid colours while the males tend to be more pastel/earth-toned. Usually these traits only become apparent during later instars. The author is unaware of any sexual dichromatism amongst spiderlings of any theraphosid, although it does occur rarely in araneid spiders, for example the hammerjawed jumper (*Zygoballus rufipes*) which becomes sexually dimorphic by the 3rd instar (Faber 1994).

6.1.2.1.6 Growth rate

Another indicator is growth rate. In general, males grow quicker than females during controlled conditions such as when the animal is fed *ad libitum* (given as much prey as it will take, all the time, J. Bull, personal observation). There is extreme variation in the time to maturity, which can take from as little as a year (*Pterinochilus murinus*) to many years (*Citharischius crawshayi,* (now: *Pelinobius muticus*)) which is well known to remain at one size for many months without moulting even as a spiderling.

6.1.2.2 Behavioural sexing

Males are usually far more active post final moult as they will be actively searching for a mate, while the females will become more sedentary in order to secure a good retreat and prepare for reproduction. Likewise the male will often be seen making a sperm web and charging his palps. Naturally these are only behaviours observed once the individuals are mature but in general, male theraphosids are throughout their life, comparably more docile and less inclined to release urticating hairs in the case of New World species (J. Bull, personal observation). Males are also known to present nuptial gifts and travel great distances for reproduction (Andersen *et al*., 2007; Hoefler *et al*., 2010; Albo and Costa 2010).

Table 6.1 Systematised signs and their features for male and females

Characteristic	Female	Male
Form of the epigastric furrow	More curved in contrast to the male	Practically horizontal and straight between the internal corners of the booklung
Distance between anterior pair of booklungs	Wider apart	Closer together in contrast to the female
Angle formed from the lower edge of the booklungs to the median line of the body of tarantula	Booklungs more angled from the horizontal than males, angle not less than 20 degrees	Booklungs more horizontal than in females, angle ~5 degrees

6.1.2.3 Aims

The aim of this study was to determine whether tibial spine dimensions can provide a straightforward, non-destructive sex-determination method across the instars. Here, the 72 available specimens spread over nine different genera were examined and tests made for the accuracy of sexing based on these spine dimensions when differences related to overall body size had been taken into account.

210

6.2   MATERIALS AND METHODS

Specimens were examined at the Natural History Museum, London courtesy of Janet Beccaloni (Curator of Arachnida and Myriapoda) within the Arachnology Department. Every available specimen was inspected (n = 72) spread over 13 different species from 10 genera: *Acanthoscurria*, (*Avicularia*), *Brachypelma*, *Harpactira*, *Lasiodora*, *Nhandu*, *Pamphobeteus*, *Paraphysa*, *Phormictopus* and *Theraphosa*. Specimens were recorded using their Latin name and categorised according to their condition upon inspection (presently in 75% ethanol, previously in 75% ethanol but now dry and those dried and stored from field-collection). Taxonomic identification had been previously verified by museum staff and external taxonomists. Measurements > 20 mm were made using a 2928 - 0104 Electronic Vernier Caliper (Accuracy: + 0.02 mm, (< 100 mm), + 0.03 mm, (> 100 − 150 mm)) and for measurements < 20 mm, a Series 293 Mitutoyo digital micrometer (Accuracy: ±1 μm).

Measurements are given in millimetres. Cephalothorax width was measured at its widest point (between leg 2 and leg 3) and cephalothorax length from the anterior margin of the chelicerae to the cephalothorax-abdomen juncture. Femur length was measured dorsally from the base of the trochanter to the top of the patella on leg 2 while the tibial spines were recorded on leg 4. As the spines are conical (for all intents and purposes), measurements were taken as far down the spine as possible, i.e. at the base of the spine. Leg 2 was chosen for measuring femur length because many specimens were not intact and lacked leg 1. Leg 4 was chosen for spine measurement because spines here appeared most dimorphic to the naked eye (personal observation). Due to the fragility of the specimens, the number of intact spines and manipulative

restrictions of many specimens, measurements from only one spine per sample were recorded. Spines were measured three times (data not shown) and an average was taken. Preliminary data showed no distinction between anterior (upper) and posterior (lower) spine dimensions, nevertheless the uppermost spine was always chosen.
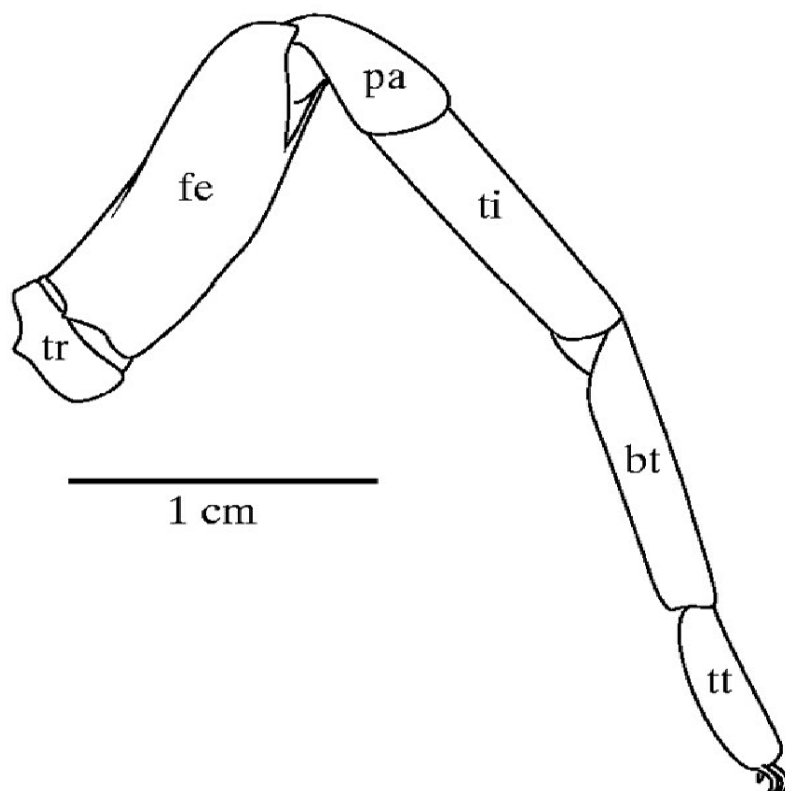
Figure 6.5 A typical theraphosid (*Aphonopelma* spp.) leg, posterior perspective. Spines occur down the retrolateral side of the tibia and occasionally above the joints. Note: on L1 of males, the mating hooks would occur approximately 2/3 down the length of the tibia. bf, basifemur; bt, basitarsus; es, elastic sclerite; ex, extensor muscle; fe, femur; pa, patella; tf, telofemur; ti, tibia; tr, trochanter; tt, telotarsus. Adapted from Sensenig and Shultz (2003).

trichobothrium for sensing air movement and low-frequency "sound"

slits (lyriform organ) for sensing vibrations and limb positions

pit (tarsal organ) for sensing moisture

taste sensitive hair

tarsal claws

Colourful scale-like hair

claw tuft hairs

touch sensitive hair
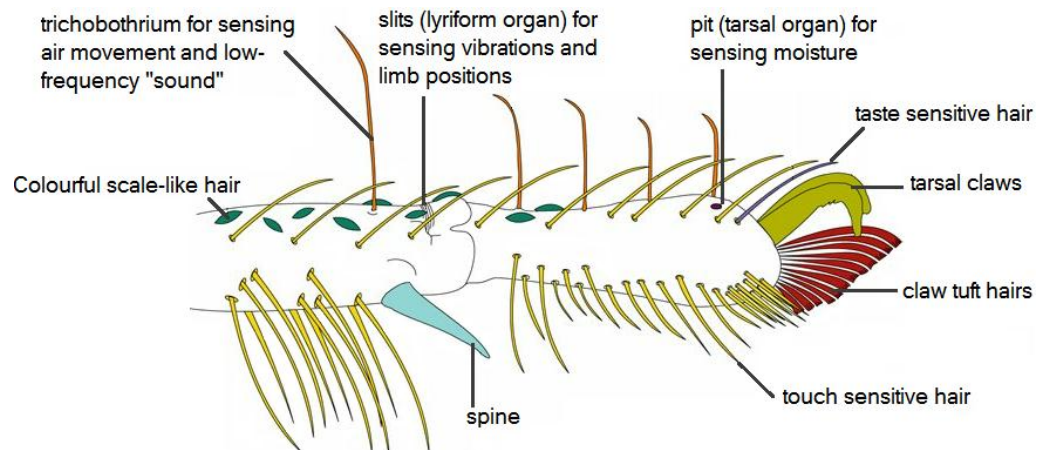
spine

Figure 6.6 A typical mygalomorph metatarsus and tarsus including spines and trichobothria (sensory hairs) (Grassé 1949). Spines (as indicated) occur down the retrolateral side of the tibia and occasionally above the joints.

Due to the range of morphological differences between both the species and sexes collected, a way of correlating the spine diameter to the variation in morphology was required. Here, merely comparing the diameters of the tibial spines would not have been sufficient to draw a sensible conclusion because in some species the male is of a comparable size to the female, while in others, the female is much larger. Likewise, intraspecific variation is also a factor because although maturity (the period after the final moult where reproduction is first possible), is dependent on the number of moults, overall size can be influenced by food acquisition over the course of a spider's life. This means that one individual that has acquired a larger number of nutritious prey will grow correspondingly larger than one that has not, despite moulting an equal number of times. Furthermore, females (unlike males) are capable of moulting

after they are sexually mature. Females will then continue to moult approximately every year of their 15 or so year lifespan and each time they will increase in size. Thus it is necessary to relate the size of each individual using standard 'tested' dimensions (cephalothorax length, cephalothorax width and femur length) to accurately gauge the relative sizes of the spines in order to make fair comparisons between the sexes. These measurements should in theory not change according to the individual spider's nourishment at the time of dimensional acquisition (pre/post-mortem). Six distinct analyses were conducted to determine the usefulness of each dimension and combination of dimensions in determining sex, shown below:

$$\frac{\text{Spine diameter}}{\text{Cephalothorax length}}$$ Abbreviation: SD/L

$$\frac{\text{Spine diameter}}{\text{Cephalothorax width}}$$ Abbreviation: SD/W

$$\frac{\text{Spine diameter}}{\text{Femur length}}$$ Abbreviation: SD/F

$$\frac{\text{Spine diameter}}{\text{C.length} + \text{C.width}}$$ Abbreviation: SD/LW

$$\frac{\text{Spine diameter}}{\text{C.length} + \text{C.width} + \text{F.length}}$$ Abbreviation: SD/LWF

$$\frac{\text{Femur length}}{\text{C.length} + \text{C.width}}$$ Abbreviation: F/LW

It should be noted at this stage that the sum of values for measurements such as LW was used rather than being multiplied to provide the surface area. The reason for this is that it would have introduced units of different factors when compared to one another.

The statistics SD/L, SD/W and SD/F are used to determine individual influences of each dimension on sex determination, while SD/LW and SD/LWF are combinatorial. The statistic F/LW is the control and is used to determine whether spine diameter is indeed necessary at all to determine sex.

6.3 RESULTS

The raw data for the 72 samples are shown in Table 6.2. A further 10 specimens from the species *Avicularia avicularia* that were devoid of any discernible tibial spines were also measured for morphometric analyses and used as a morphological control. Statistics were calculated from the 36 male and 36 female specimens sampled from 9 species. Independent samples *t*-tests were conducted using the six aforementioned test variables and sex (male vs. female) as the grouping variable. In all six cases, the male measurements produced a statistically significant *F*-statistic and were significantly larger than the females ($p < .001$), including the control F/LW test variable, which was still highly significantly different ($p < .001$). These results indicate that when the standard dimensions of the cephalothorax and the femur are compared, there is a significant difference between the sexes and this increases further when correlated with the spine diameter.

A univariate analysis of variance (ANOVA) was conducted to discern the between-subjects effects of both sex and species and it was found that sex had a significant effect with relation to SD/LWF ($p < .001$). Species had no significant effect ($p = .454$) and there was no significant interaction between sex and species ($p = .967$). The normality of each of the six test variables distributions' were evaluated visually with both Q-Q plots and the Shapiro-Wilk test where it was found that all exceeded the Shapiro-Wilk critical value at the 5% level, confirming that the data follow a normal distribution.

Table 6.2 Raw data for the 36 male:36 female specimens covering nine different species and physical dimensions thereof.

Species	C.thorax L	C.thorax W	Femur 2 L	Sex	Spine diameter
A. brocklehursti	28.67	24.44	19.23	F	0.209
	15.57	12.36	8.78	F	0.087
	21.05	18.88	15.28	F	0.19
	21	18.3	16.16	F	0.152
	20.77	17.87	13.81	F	0.199
	17.11	15.26	14.4	M	0.188
	14.11	13.15	12.36	M	0.19
	17.1	15.88	13.72	M	0.184
	13.2	12.02	11.76	M	0.153
	18.57	14.58	12.72	M	0.167
	14.67	13.83	12.46	M	0.174
A. geniculata	27.2	25.55	23.54	M	0.243
	22.2	20.41	19.35	M	0.246
	24.07	22.44	21.04	M	0.281
	24.98	23.65	18.54	F	0.206
B. vagans	16.15	17.21	16.52	M	0.232
Harpactira spp.	23.33	20.4	14.42	F	0.205
	16.9	14.8	14.26	M	0.185
N. vulpinus	22.16	21.18	18.53	M	0.244
	21.53	20.75	19.93	M	0.277
	24.56	22.06	17.26	F	0.193
	23.96	23.79	17.88	F	0.222
	15.34	13.78	12.95	F	0.14
	21.23	21.17	18.9	M	0.216
	18.08	16.6	15.73	F	0.171
	22.09	21.55	17.74	F	0.224
	22.15	21.36	17.17	F	0.123
	20.96	21.73	18.21	F	0.204
	27.09	24.44	18.62	F	0.236
	20.28	20.1	19.99	M	0.265
P. antinous	31.11	26.96	23.73	M	0.29
	25.31	24.3	23.08	M	0.306
	27.68	25.52	22.92	M	0.295
	27.09	25.41	22.19	M	0.295
P. insignis	22.2	19.93	19.42	M	0.244
	26.11	22.38	18.08	F	0.241
	26.88	26.36	19.21	F	0.233
P. scrofa	17.94	17.84	15.71	M	0.214
	18.29	14.4	12.42	F	0.154
	16.92	16.92	16.51	M	0.211

	19.8	16.63	15.15	M	0.246
	21.32	18.82	17.3	M	0.277
P. cancerides	22.52	21.48	18.72	M	0.229
	22.55	20.23	22.33	M	0.232
	21.99	20.69	19.47	F	0.198
	20.34	19.03	16.9	M	0.233
	20.93	19.37	17.75	M	0.213
	20.57	22.04	19.39	M	0.254
	18.52	19.58	17.85	M	0.23
T. blondi	37.44	35.95	25.75	F	0.338
	26.24	26.8	21.75	F	0.285
	38.15	36.42	24.22	F	0.317
	29.36	33.62	25.15	M	0.329
	30.95	30.32	21.41	F	0.238
	32.41	31.36	23.27	F	0.175
	33.12	32.83	22.73	F	0.306
	27.26	27.01	16.41	F	0.178
	36.2	36.89	27.94	M	0.406
T. apophysis	32.06	34.21	26.69	M	0.308
	27.19	29.85	22.1	F	0.198
	20.41	28.68	23.47	F	0.264
L. klugi	25.38	25.23	22.41	M	0.267
	20.57	20.54	20.66	F	0.200
	22.89	21.67	17.74	F	0.178
	23.32	22.47	18.1	F	0.188
	23.17	22.72	17.87	F	0.233
	19.83	18.82	16.02	F	0.137
	24.5	23.31	21.69	M	0.241
	29.06	28.38	20.51	F	0.191
	28.08	26.06	20.57	F	0.221
	25.06	22.58	18.11	F	0.142
	24.38	22.21	22.68	M	0.276
A. avicularia	18.72	17.97	14.42	F	N/A
	19.02	17.84	13.59	F	N/A
	19.71	19.99	16.44	F	N/A
	17.86	18.77	15.2	M	N/A
	20.27	19.58	15.67	M	N/A
	14.42	15.58	12.37	F	N/A
	18.69	21.63	14.87	F	N/A
	15.05	16.32	13.99	M	N/A
	14.89	15.57	11.31	F	N/A
	14.12	16.74	15.08	M	N/A

6.4 DISCUSSION

6.4.1 DIMENSIONAL ANALYSES

Traditional morphometric analyses have consistently been utilised (albeit visually) often without definitive statistical support as a way of predicting the sex of immature mygalomorph spiders. Granted, some species such as *Haplopelma lividum* are highly sexually dimorphic and even during early instars there is a clear morphological size discrepancy, however, this is rare in other tarantulas. The statistical evidence here confirms previous hypotheses about differences in size between the sexes and adds a further measurement that increases the statistical probability of accurate sexual determination.

Here, distinct descriptors (SD/L, SD/W, SD/F, SD/LW, SD, LWF and F/LW) have been utilised and tested to determine which, if any, can be used to accurately gauge the sex of an individual. In this instance, even the poorest descriptor was found to be well within the range of statistical significance with $p < 10^{-6}$ (SD/F) with the most significant descriptor having $p < 4.1 \times 10^{-14}$. The regression analyses also suggest these methods (even applying SD/F, the lowest statistically significant descriptor) are accurate at predicting sex across a range of species and despite interspecies variation and instar level.

6.4.1.1 Spine Diameter/Cephalothorax Length (SD/L)

An independent samples *t*-test for this statistic yielded a *t*-value of 8.614 where $p < .001$ with 1.35×10^{-12} (Table 6.3). The full *p*-value has been included as an ease of comparison to the successive statistics shown below. A Pearson's correlation statistic of 0.625 (Figure 6.7) is significant at the 0.001

level indicating a high level of correlation between the size of the tibial spines and the cephalothorax length.

6.4.1.2 Spine Diameter/Cephalothorax Width (SD/W)

The *t*-test for this statistic yielded a *t*-value of 9.189 where $p < .001$ with 1.19×10^{-13} (Table 6.4). The *p*-value in this instance is higher than that of SD/L and the means are comparable, however the standard deviation is a little higher suggesting a greater spread of the values. In each of the above tests, both the *F*-value and the significance values under the Levene's Test for Equality of Variances are above the 0.05 critical value and thus it can be concluded that the variance is equal. A Pearson's correlation statistic of 0.672 (Figure 6.8) is significant at the 0.001 level indicating again a high level of correlation between the sizes of the tibial spines and the cephalothorax width.

6.4.1.3 Spine Diameter/Femur Length (SD/F)

Here, the independent samples *t*-test for this statistic yielded a *t*-value of 4.961 where $p < .001$ with 5.68×10^{-6} (Table 6.5). This value was used as the Levene's Test for Equality of Variances calculated value is below the critical value of 0.05 and therefore equal variances are not assumed, which indicates a difference between the variances of the population. As above, *p* is still extremely significant albeit to a lesser extent compared to the previous data. Here, the Pearson's correlation statistic was 0.791 (Figure 6.9), so although the statistic yields less significance than the cephalothorax dimensions, the data follows a more linear relationship with the size of the tibial spines.

Table 6.3 Independent samples *t*-test for SD/L for the 72 individual specimens taken from the Natural History Museum, London. The *t*-values according to *t*-test for each variable are as shown in the table along with the significance level (*p*) calculated from Levene's Test for Equality of Variances.

Group Statistics

	sex	*N*	Mean	Std. Deviation	Std. Error Mean
SD/L	M	36	.01128	.001262	.000210
	F	36	.00836	.001599	.000267

Independent Samples Test

		Levene's Test for Equality of Variances		*t*-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper	
SD/L	Equal variances assumed	1.353	.249	8.614	70	.000000000001352	.002924	.000339	.002247	.003601	
	Equal variances not assumed			8.614	66.401	.000000000002019	.002924	.000339	.002247	.003602	
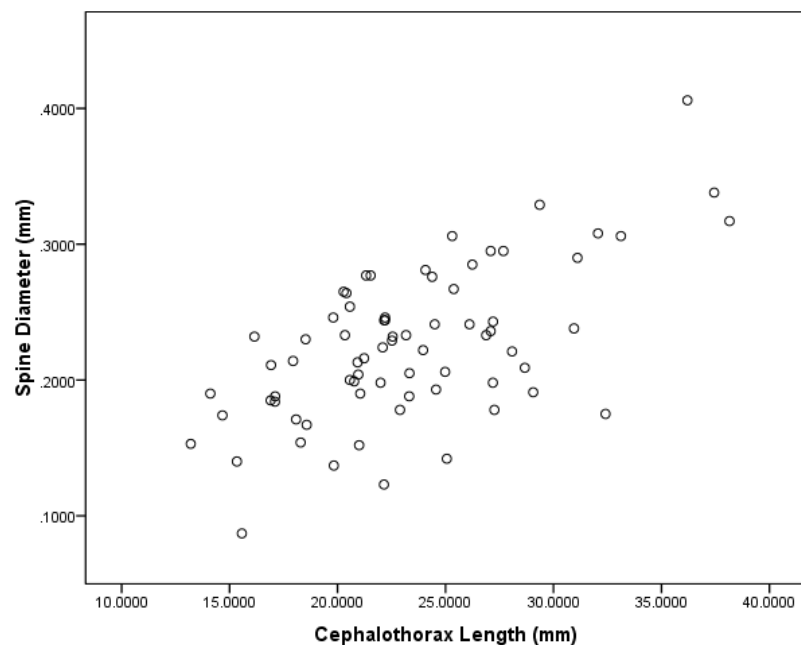
Figure 6.7 Graphical representation of the correlation between SD and L alongside the true numerical Pearson's correlation statistic.

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
L	23.34972	5.387593	72
SD	.22524	.055724	72

Correlations

		L	SD
L	Pearson Correlation	1	.625[**]
	Sig. (2-tailed)		.000
	N	72	72
SD	Pearson Correlation	.625[**]	1
	Sig. (2-tailed)	.000	
	N	72	72

**. Correlation is significant at the 0.01 level (2-tailed).

Table 6.4 Independent samples *t*-test for SD/W for the 72 individual specimens taken from the Natural History Museum, London. The *t*-values according to *t*-test for each variable are as shown in the table along with the significance level (*p*) calculated from Levene's Test for Equality of Variances.

Group Statistics

	sex	*N*	Mean	Std. Deviation	Std. Error Mean
SD/W	M	36	.01189	.001347	.000225
	F	36	.00880	.001501	.000250

Independent Samples Test

		Levene's Test for Equality of Variances		*t*-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper	
SD/W	Equal variances assumed	.716	.400	9.189	70	.000000000000119	.003090	.000336	.002419	.003760	
	Equal variances not assumed			9.189	69.196	.000000000000132	.003090	.000336	.002419	.003760	
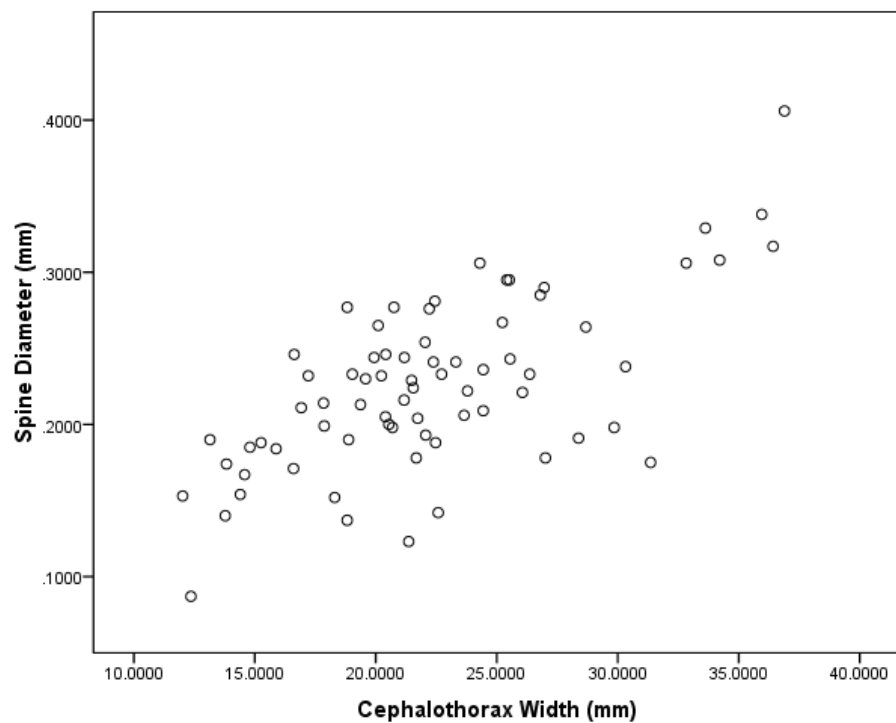
Figure 6.8 Graphical representation of the correlation between SD and W alongside the true numerical Pearson's correlation statistic.

Correlations

Descriptive Statistics

	Mean	Std. Deviation	*N*
W	22.29667	5.799099	72
SD	.22524	.055724	72

Correlations

		SD	L
W	Pearson Correlation	.672**	1
	Sig. (2-tailed)	.000	
	N	72	72
S	Pearson Correlation	1	.672**
	Sig. (2-tailed)	.000	
	N	72	72

**. Correlation is significant at the 0.01 level (2-tailed).

Table 6.5 Independent samples *t*-test for SD/F for the 72 individual specimens taken from the Natural History Museum, London. The *t*-values according to *t*-test for each variable are as shown in the table along with the significance level (*p*) calculated from Levene's Test for Equality of Variances.

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
SD/F	M	36	.01302	.001293	.000215
	F	36	.01115	.001855	.000309

Independent Samples Test

		Levene's Test for Equality of Variances		*t*-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SD/F	Equal variances assumed	5.812	.019	4.961	70	.000004721612351	.001870	.000377	.001118	.002621
	Equal variances not assumed			4.961	62.503	.000005676944719	.001870	.000377	.001117	.002623

225

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
F	18.70333	3.856730	72
SD	.22524	.055724	72

Correlations

		SD	F
F	Pearson Correlation	.791**	1
	Sig. (2-tailed)	.000	
	N	72	72
SD	Pearson Correlation	1	.791**
	Sig. (2-tailed)	.000	
	N	72	72

**. Correlation is significant at the 0.01 level (2-tailed).
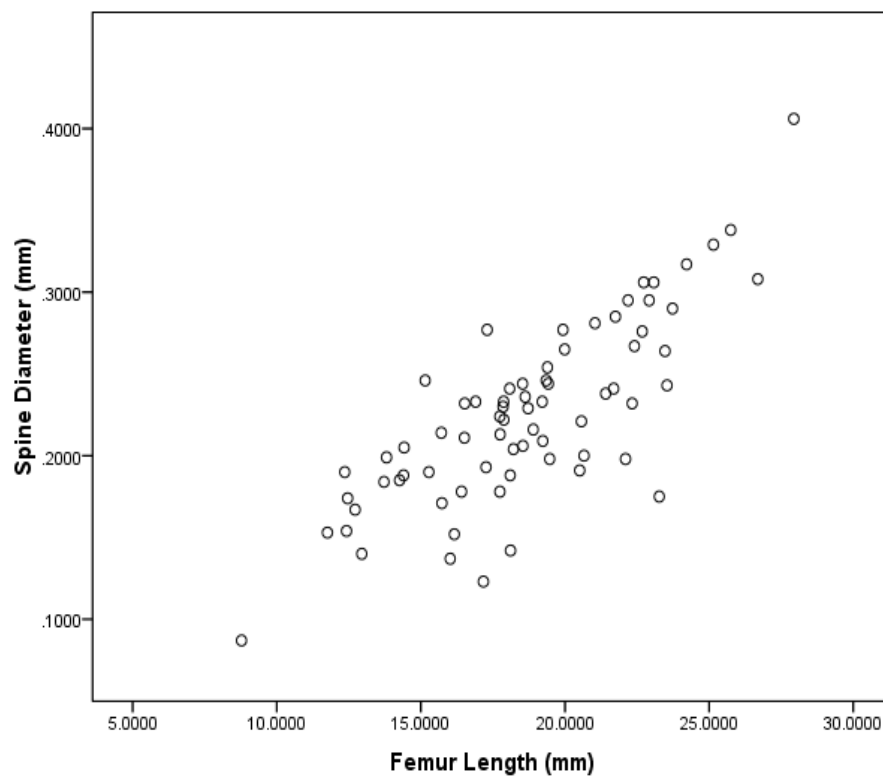
Figure 6.9 Graphical representation of the correlation between SD and L alongside the true numerical Pearson's correlation statistic.

All of the above dimensions were themselves tested to examine whether there were correlations between them. The Pearson's correlation statistics were 0.945 (Table 6.6), 0.817 (Table 6.7) and 0.833 (Table 6.8) for the correlation between L-W, L-F and W-F respectively, in all cases $p < .001$.

6.4.1.4 Spine Diameter/(C. Length + C. Width) (SD/LW)

The independent samples t-test for this statistic yielded a t-value of 9.445 with $p < 4.1 \times 10^{-14}$ (Table 6.9). Again, the significance value is above the critical value for Levene's Test for Equality of Variances so equal variance can be concluded. This is by far the most statistically significant p-value and thus has been deemed perhaps the most valuable for use in future analyses. Subsequent correlations will use this value to illustrate proof of principle. The Pearson's correlation statistic here was 0.659 ($p < .001$, data not shown), less than the aforementioned statistics as a summation of the spread of the data has resulted albeit to a seemingly negligible level.

6.4.1.5 Spine Diameter/(Femur Length + C. Length + C. Width) (SD/LWF)

The t-value for this test statistic was 8.612 with $p < .001$ with 1.36×10^{-12} (Table 6.10). Pearson's correlation statistic was 0.712 ($p < .001$). Despite utilising more variables, there is no significant difference in the usefulness of this statistic. The conclusion here is that while femur length does have a bearing, it is not required as a dimension.

6.4.1.6 Femur Length/(C. Length + C. Width) (F/LW)

Finally, the control analysis excluding the spine diameter measurements (to evaluate the latter's importance) yielded a *t*-value of 6.632 where $p < .001$ with 5.81×10^{-9} (Table 6.11). The Pearson's correlation statistic in this instance was 0.863 ($p < .001$), producing by far the most linearly correlated values. Despite these dimensions being perhaps the most easily accessible, and with a significant difference between the means of the sexes ($p < .001$), there is significant overlap and similarity between the means and standard deviations. This means that to predict sex based on solely these dimensions would yield a more ambiguous conclusion (explained later). Thus, incorporating spine diameter here adds significant weight to this method of sexing.

6.4.2   UNIVARIATE ANALYSIS OF VARIANCE

To determine the between-subjects effects, a univariate analysis of variance was conducted. Between-subjects effects are those whose values change in between-subjects but remain the same on a single subject. In this instance, while the dimensions of the animals change, the sex and species remain relatively unchanged with respect to one another. Here it is shown that while sex has a highly significant impact on the relationship between body dimensions and spine diameter ($p < .001$), there is little impact from species differences ($p < .454$) and practically zero impact from the interaction between sex and species ($p < .967$) (Table 6.12). In other words, this method (at least for the species examined), works across taxa.

228

Table (6.6), (6.7) and (6.8) Correlation statistics as calculated from Pearson's formula comparing L-W (6.6), L-F (6.7) and W-F (6.8). Significance is given at $p \leq 0.05$.

(6.6)

Descriptive Statistics

	Mean	Std. Deviation	N
L	23.34972	5.387593	72
W	22.29667	5.799099	72

Correlations

		L	W
L	Pearson Correlation	1	.945[**]
	Sig. (2-tailed)		.000
	N	72	72
W	Pearson Correlation	.945[**]	1
	Sig. (2-tailed)	.000	
	N	72	72

**. Correlation is significant at the 0.01 level (2-tailed).

(6.7)

Descriptive Statistics

	Mean	Std. Deviation	N
L	23.34972	5.387593	72
F	18.70333	3.856730	72

Correlations

		L	F
L	Pearson Correlation	1	.817[**]
	Sig. (2-tailed)		.000
	N	72	72
F	Pearson Correlation	.817[**]	1
	Sig. (2-tailed)	.000	
	N	72	72

**. Correlation is significant at the 0.01 level (2-tailed).

(6.8)

Descriptive Statistics

	Mean	Std. Deviation	N
W	22.29667	5.799099	72
F	18.70333	3.856730	72

Correlations

		W	F
W	Pearson Correlation	1	.883[**]
	Sig. (2-tailed)		.000
	N	72	72
F	Pearson Correlation	.883[**]	1
	Sig. (2-tailed)	.000	
	N	72	72

**. Correlation is significant at the 0.01 level (2-tailed).

Table 6.9 Independent samples *t*-test for SD/LW for the 72 individual specimens taken from the Natural History Museum, London. The *t*-values according to *t*-test for each variable are as shown in the table along with the significance level (*p*) calculated from Levene's Test for Equality of Variances.

Group Statistics

	sex	*N*	Mean	Std. Deviation	Std. Error Mean
SD/LW	M	36	.00578	.000617	.000103
	F	36	.00427	.000731	.000122

Independent Samples Test

		Levene's Test for Equality of Variances		*t*-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SD/LW	Equal variances assumed	1.740	.191	9.445	70	.000000000000041	.001506	.000159	.001188	.001824
	Equal variances not assumed			9.445	68.063	.000000000000053	.001506	.000159	.001188	.001824

Table 6.10 Independent samples *t*-test for SD/LWF for the 72 individual specimens taken from the Natural History Museum, London. The *t*-values according to *t*-test for each variable are as shown in the table along with the significance level (*p*) calculated from Levene's Test for Equality of Variances.

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
SD/LWF	M	36	.00400	.000398	.000066
	F	36	.00308	.000499	.000083

Independent Samples Test

		Levene's Test for Equality of Variances		*t*-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SD/LWF	Equal variances assumed	3.215	.077	8.612	70	.00000000001365	.000917	.000106	.000704	.001129
	Equal variances not assumed			8.612	66.719	.000000000001965	.000917	.000106	.000704	.001129

Table 6.11 Independent samples *t*-test for F/LW for the 72 individual specimens taken from the Natural History Museum, London. The *t*-values according to *t*-test for each variable are as shown in the table along with the significance level (*p*) calculated from Levene's Test for Equality of Variances.

Group Statistics

	sex	*N*	Mean	Std. Deviation	Std. Error Mean
F/LW	M	36	.44471	.031092	.005182
	F	36	.38528	.043859	.007310

Independent Samples Test

		Levene's Test for Equality of Variances		*t*-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
F/LW	Equal variances assumed	2.476	.120	6.632	70	.000000005813282	.059425	.008960	.041554	.077296
	Equal variances not assumed			6.632	63.085	.000000008667294	.059425	.008960	.041520	.077330

Table 6.12 (a) Univariate analysis of variance illustrating the between-subject factors (sex vs. species) and (b) how much influence these factors have on the dependent variable: size.

(a) Univariate Analysis of Variance

Between-Subjects Factors

		N
Sex	F	36
	M	36
Species	1	11
	2	4
	3	1
	4	2
	5	12
	6	4
	7	3
	8	5
	9	7
	10	9
	11	3
	12	11

(b)

Tests of Between-Subjects Effects

Dependent Variable: size

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	18.783[a]	21	.894	4.211	.000
Intercept	501.879	1	501.879	2.363E3	.000
Sex	5.190	1	5.190	24.435	.000
Species	2.352	11	.214	1.007	.454
Sex * Species	.598	9	.066	.313	.967
Error	10.619	50	.212		
Total	932.116	72			
Corrected Total	29.402	71			

R^2 = .639, Adjusted R^2 = .487

233

6.4.3 REGRESSION ANALYSES

Linear regression was employed to determine the relationship between sex and the physical dimensional variables to develop a rudimentary tool for sex prediction (Table 6.13). Here, the model summary provides an R^2 value of 0.587 (Table 6.13a) meaning that approximately 60% of the variation in sex can be explained by the dimensions recorded in this instance. The remaining 40% are a result of lurking variables or interspecies variability. A prediction of the statistical shrinkage of the model is provided by the adjusted R^2 score of 0.563 suggesting that the efficacy of the model will be reduced by approximately 4% when used on new data. Naturally there is an issue here with over-fitting the data and as the numbers of factors involved increases, so does the likelihood of the curse of dimensionality i.e. correlation here does not necessarily imply causation.

The high F-statistic (23.824) and the associated p-value ($p < .001$) from the analysis of variance (Table 6.13b) indicate that the regression equation (below) is explaining a statistically significant portion of the variability in sex as a direct result of the dimensional variables.

Table 6.13 A regression analysis for L, W, F and SD . (a) R and R^2 values using the predictors and dependent variables. (b) ANOVA, (c) coefficient calculations and accompanying logistic regression equation and (d) residuals.

(a)

Model Summary[b]

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
1	.766[a]	.587	.563	.333

a. Predictors: (Constant), SD, L, W, F

b. Dependent Variable: Sex

(b)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10.569	4	2.642	23.824	.000[a]
	Residual	7.431	67	.111		
	Total	18.000	71			

a. Predictors: (Constant), SD, L, W, F

b. Dependent Variable: Sex

(c)

Coefficients[a]

Model		Unstandardised Coefficients		Standardised Coefficients		
		B	Std. Error	B	t	Sig.
1	(Constant)	1.042	.222		4.686	.000
	L	-.021	.023	-.228	-.944	.348
	W	-.087	.026	-1.000	-3.367	.001
	F	.076	.027	.582	2.854	.006
	SD	6.524	1.165	.722	5.601	.000

a. Dependent Variable: Sex

$$Sex = 1.042 - (0.021L) - (0.087W) + (0.076F) + (6.524SD)$$

(d)

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.52	2.14	1.50	.386	72
Residual	-.696	.573	.000	.324	72
Std. Predicted Value	-2.527	1.669	.000	1.000	72
Std. Residual	-2.090	1.721	.000	.971	72

a. Dependent Variable: Sex

6.4.3.1.1 Logistic regression equation

For sex, the logistic regression equation encompassing the species (Table 6.13c) was:

Sex = 1.042 - (0.021L) – (0.087W) + (0.076F) + (6.524SD)

Where a final value of 1 = Female   and   2 = Male   (1.5 = equal likelihood)

The *t*-statistics and their associated 2-tailed *p*-values all indicate statistically significant ($p < .001$) factors to the sex-determination logistic regression equation with the exception of the cephalothorax length ($p < .348$). The residual statistics given in Table 6.13d are provided for reference only and will not be discussed.

6.4.3.1.2 Dimension-based sex prediction success probability

The probability of a given sex corresponding to a particular dimension (e.g. L, W, F…) in an experiment in which there is an equally large number of equally likely independent trials is approximated by the normal probability density function:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\left(\frac{x-\mu^2}{2\sigma^2}\right)}$$

Where:

π = 3.14159265 (8 d.p.)

σ = Standard Deviation

μ = Mean

e = 2.71828183 (8 d.p.)

x = Calculated value for the given dimension

The above formula is excessive as the constant cancels when coupling the

males-females, leaving:

$$\frac{e^{-\left(\frac{x-\mu^2}{\sigma}\right)}}{\sigma}$$

Where:

σ = Standard Deviation

μ = Mean

e = 2.71828183 (8 d.p.)

x = Calculated value for the given dimension

This rudimentary z-value or z score describes the divergence of the experimental result x from the most likely result μ (the mean) in the form of the number of standard deviations σ. Large values of 'z', indicate a lower probability that the experimental result has arisen due to chance. Doing this for both sexes over a range of experimental results ($x_{\{e.g.\ 1.1,\ 1.2,\ 1.3...\}}$) and dividing those of the male by those of the female produces the likelihood ratio 'LR' which is the probability of obtaining that value for x if the spider is male. For example, a LR of 700 means, given the corresponding value of x, the specimen is 700 times more likely to be a male over being female. To convert this to a probability, the link function of $LR/(1+LR)$ can be used to predict the likelihood of said individual being male. The probabilities of males, for the range of dimensional statistics were calculated (SD/L, SD/W…) but those representing the variable SD/L are shown in Table 6.14 (all others are shown in Appendix 7) while the probability predictions are shown in Figure 6.10 - Figure 6.15.

Table 6.14 Sex prediction table calculated from the logistic regression equation (Z) using the statistic SD/L.

x	LR	LR/(1+LR)
5	0	0
5.5	0	0
6	0	0
6.5	0	0
7	0.01	0
7.5	0.02	0.01
8	0.04	0.02
8.5	0.11	0.04
9	0.27	0.1
9.5	0.6	0.21
10	1.28	0.38
10.5	2.57	0.56
11	4.85	0.72
11.5	8.62	0.83
12	14.44	0.9
12.5	22.8	0.94
13	33.93	0.96
13.5	47.58	0.97
14	62.87	0.98
14.5	78.3	0.98
15	91.89	0.99
15.5	101.63	0.99
16	105.93	0.99

The data in Table 6.14 illustrate that even a very small change in the value of x (calculated in this instance by SD/L) causes a dramatic increase in the likelihood of said individual being identified as belonging to one sex over another. Here for example, a calculated x-value of 10 provides a *LR* of 1.28:1 chance of male:female; in effect, near equal chance. This means there is a 56% chance of a male and 44% chance of a female. An increase of just one unit ($x =$ 11) makes the individual almost 5 times more likely to be male than female, i.e. there is an 83% chance of it being male.

The statistics SD/L, SD/W, SD/F, SD/LW, SD, LWF and F/LW represented by Figure 6.10-Figure 6.15 respectively, are all fully capable of being used to sex individuals based on their dimensions to varying degrees as illustrated above by SD/L. SD/L and SD/W are both practically useful but SD/W would be more favoured due to it having a narrower spread of data. SD/F has a large spread of data for the female in particular and a shallow gradient resulting in a far greater proportion of overlapping data points makes sex determination far more unreliable. This case is true for the 'control' statistic F/LW with bulbous normal distributions for both sexes and a shallow probability curve. This is also supported and illustrated by a box and whisker plot comparing F/LW and SD/LWF (Figure 6.16). The two 'remaining' curves SD/LW and SD/LWF both appear equally useful in sex determination within which the differences in accuracy can only be truly discerned mathematically. Without using complex integrations and going purely on the *t*-values (above) and the practicality of only taking three measurements (SD, L and W), the statistic SD/LW would seem to be the most useful for future determination of the sexes but one cannot deny the apparent distinction between the sexes afforded by simply using the dimensions F vs. L or W (Figure 6.17a and b).
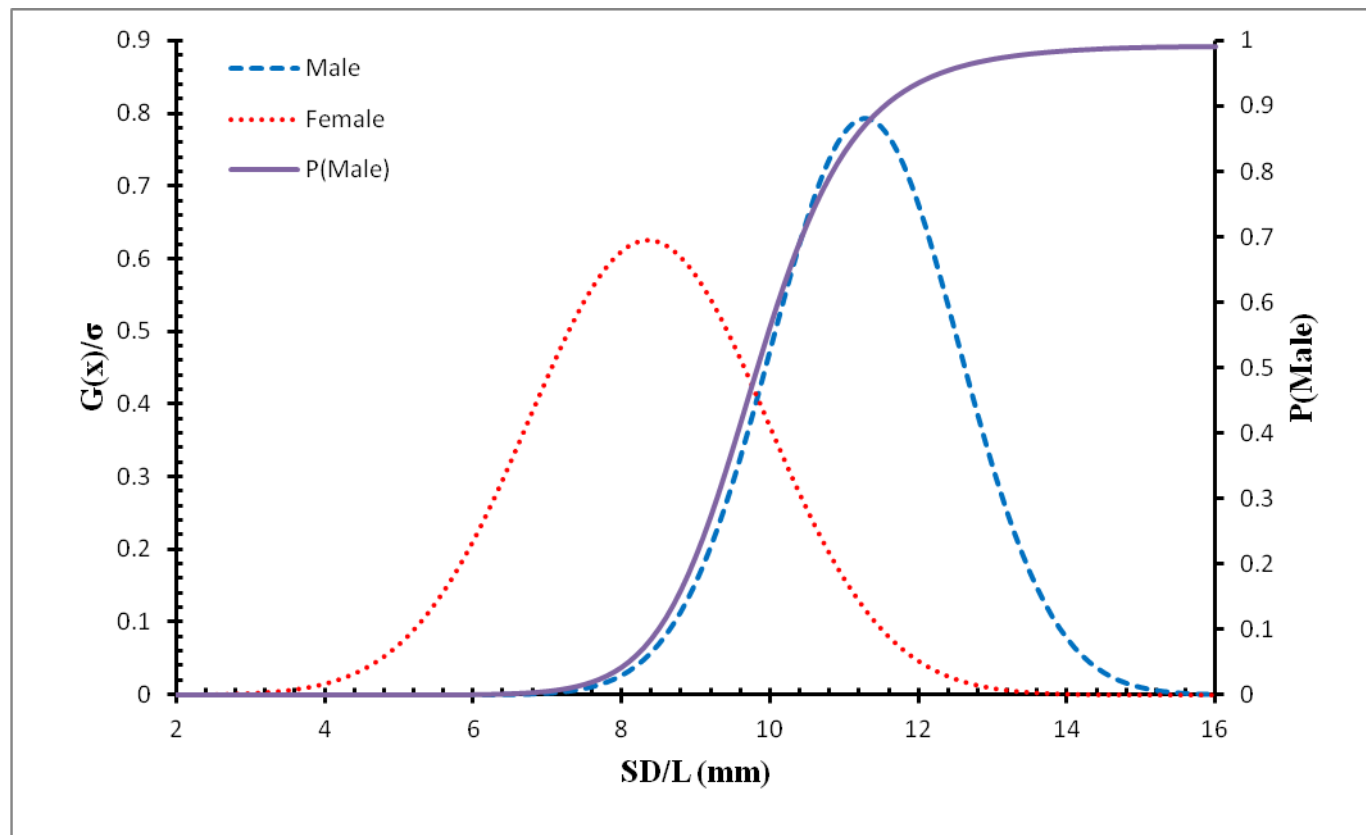
Figure 6.10 Predicted probability of the sex (sigmoidal curve) overlaid on the normal distribution curves for males and females using the statistic SD/L.
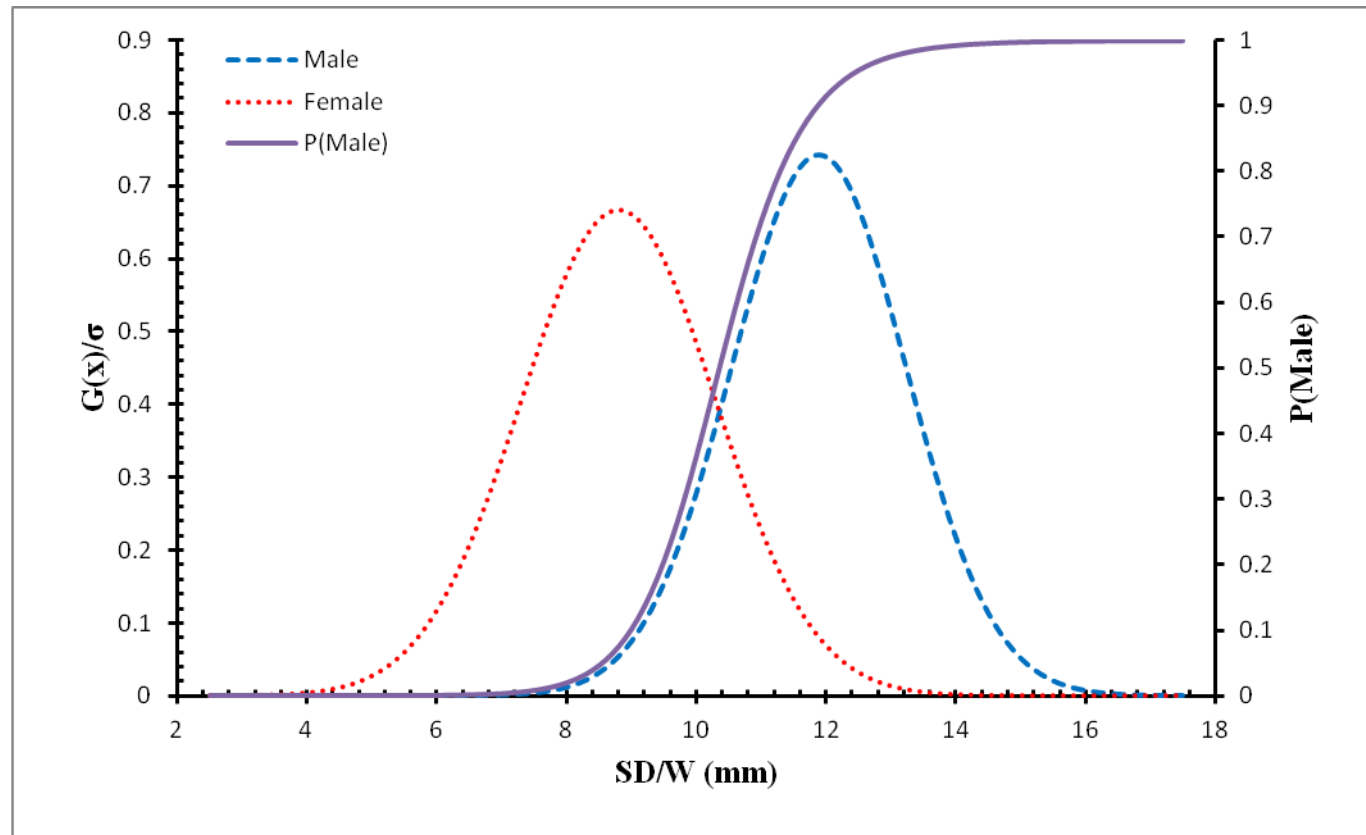
Figure 6.11 Predicted probability of the sex (sigmoidal curve) overlaid on the normal distribution curves for males and females using the statistic SD/W.
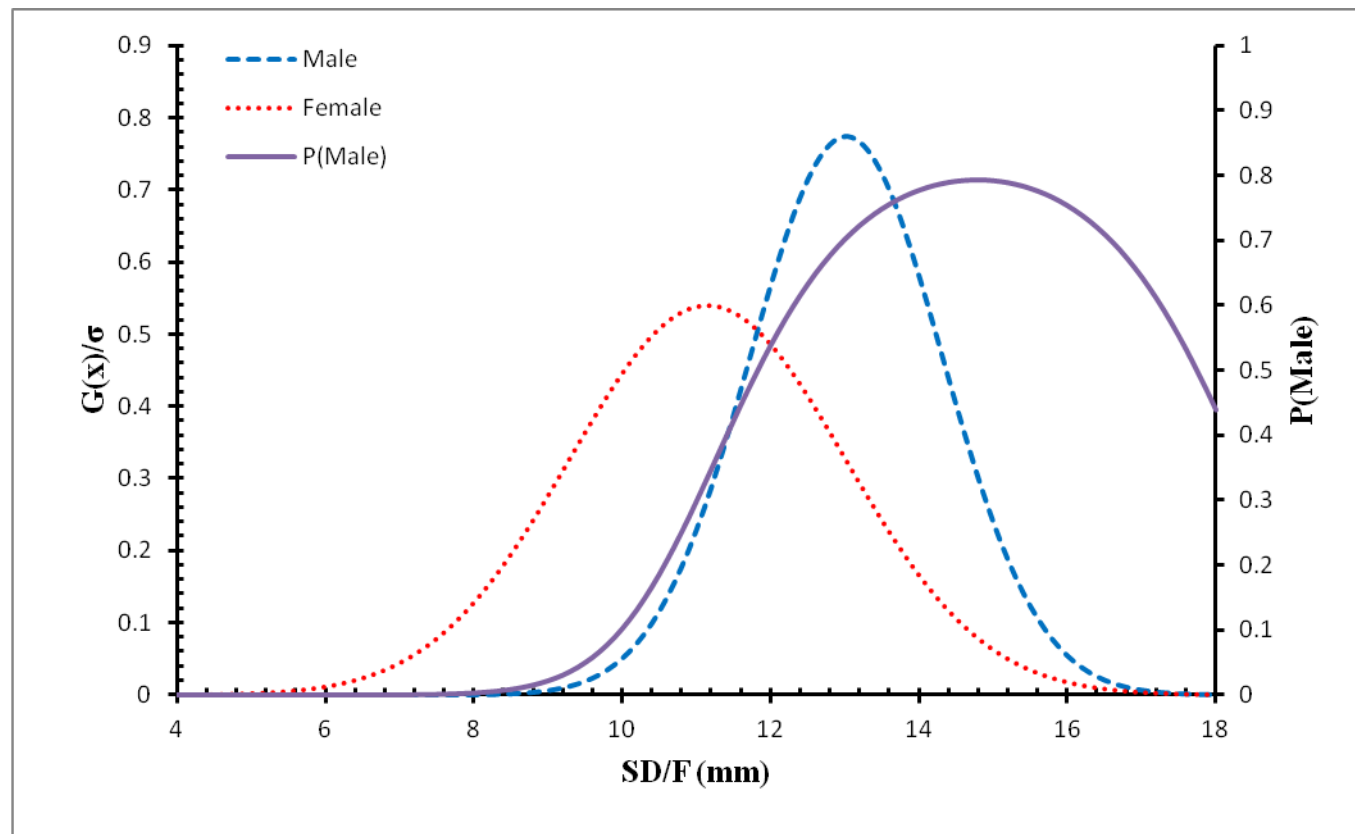
Figure 6.12 Predicted probability of the sex (sigmoidal curve) overlaid on the normal distribution curves for males and females using the statistic SD/F.
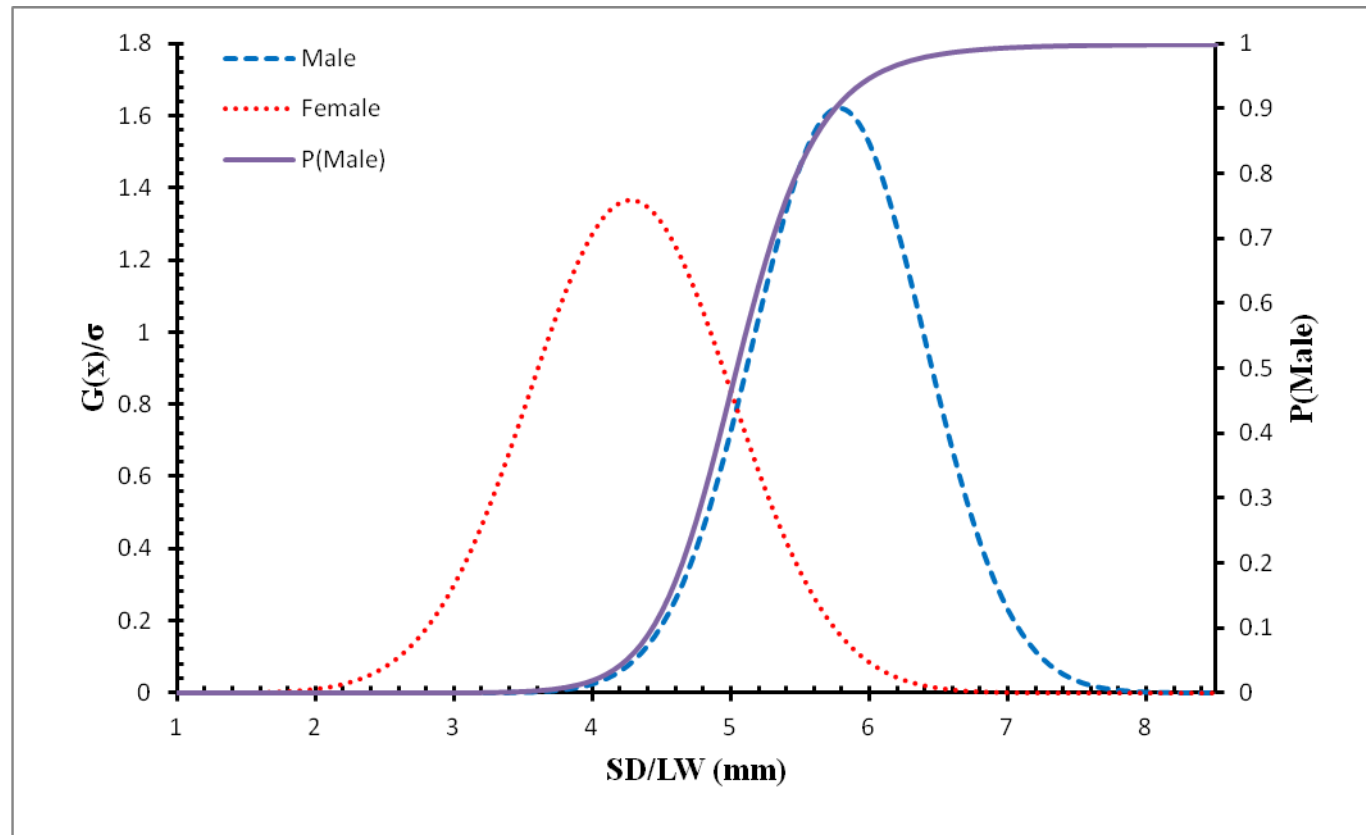
Figure 6.13 Predicted probability of the sex (sigmoidal curve) overlaid on the normal distribution curves for males and females using the statistic SD/LW.
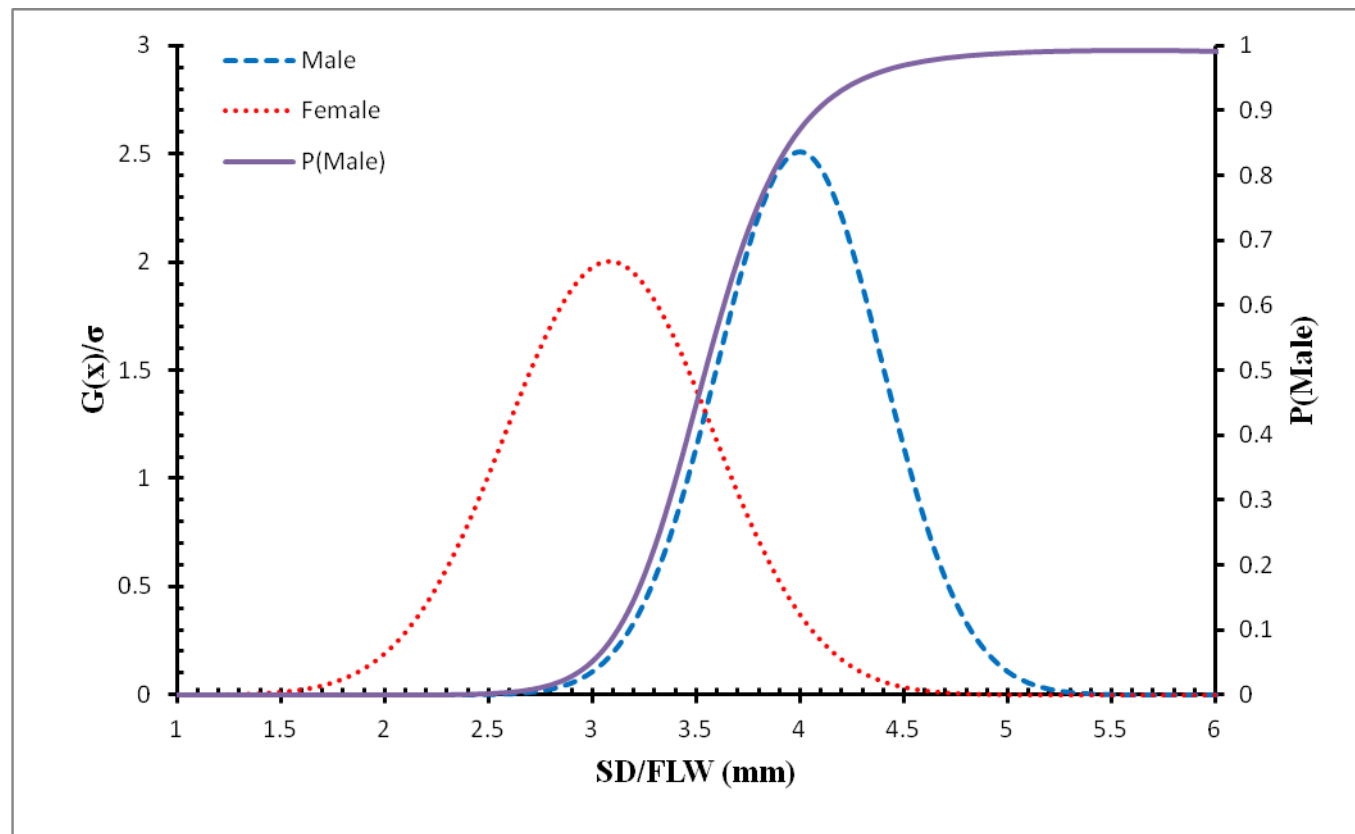
Figure 6.14 Predicted probability of the sex (sigmoidal curve) overlaid on the normal distribution curves for males and females using the statistic SD/FLW.
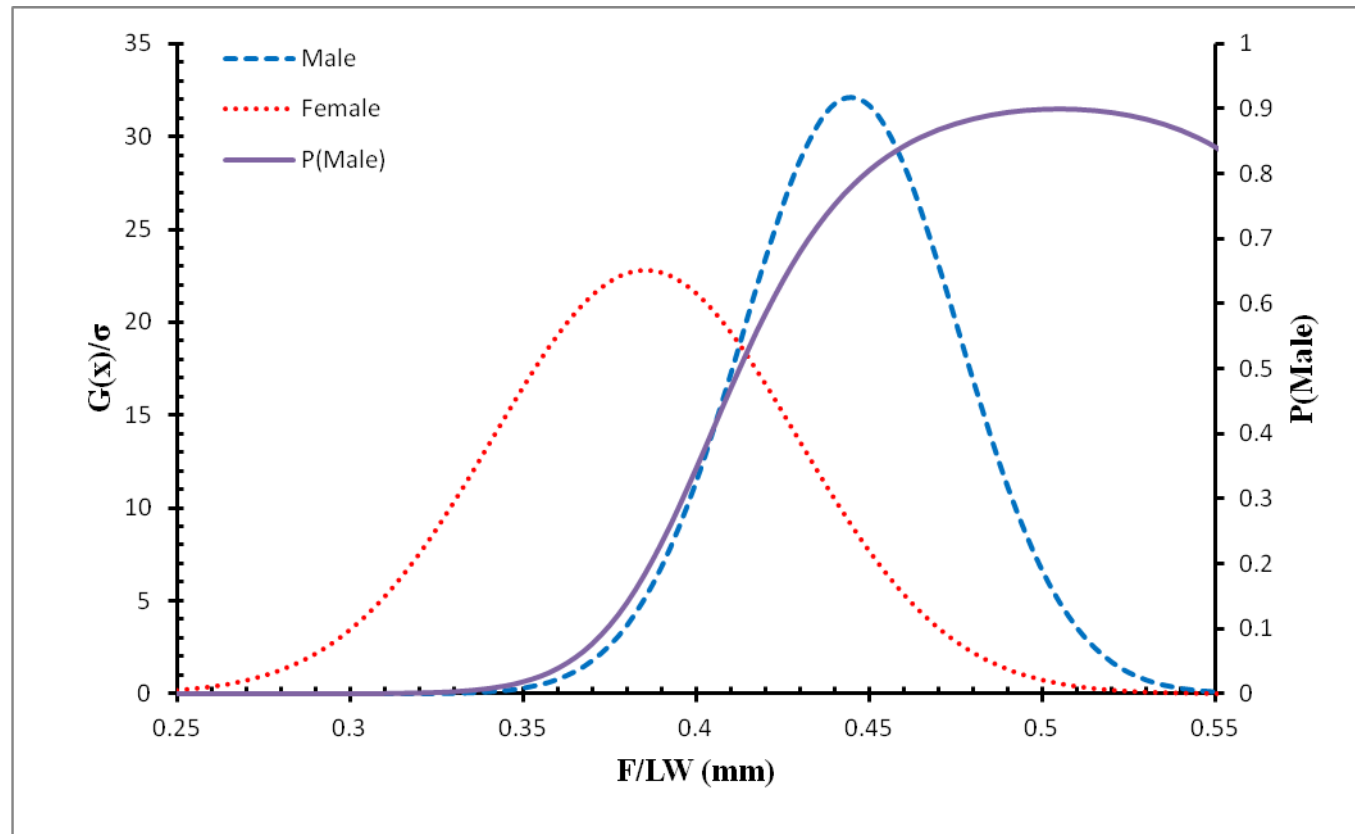
Figure 6.15 Predicted probability of the sex (sigmoidal curve) overlaid on the normal distribution curves for males and females using the statistic F/LW.
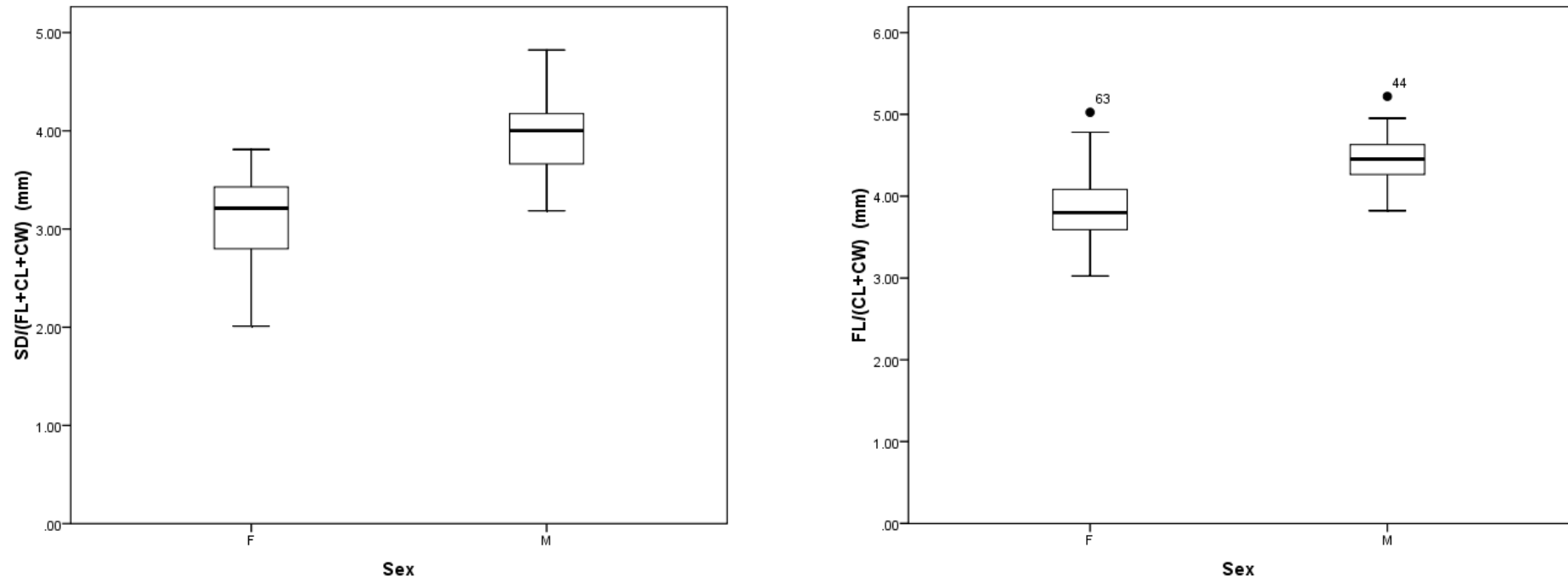
Figure 6.16 Box and whisker plot depicting the overlap between the sexes depending on which statistic (SD/LWF…or… F/LW) is used.
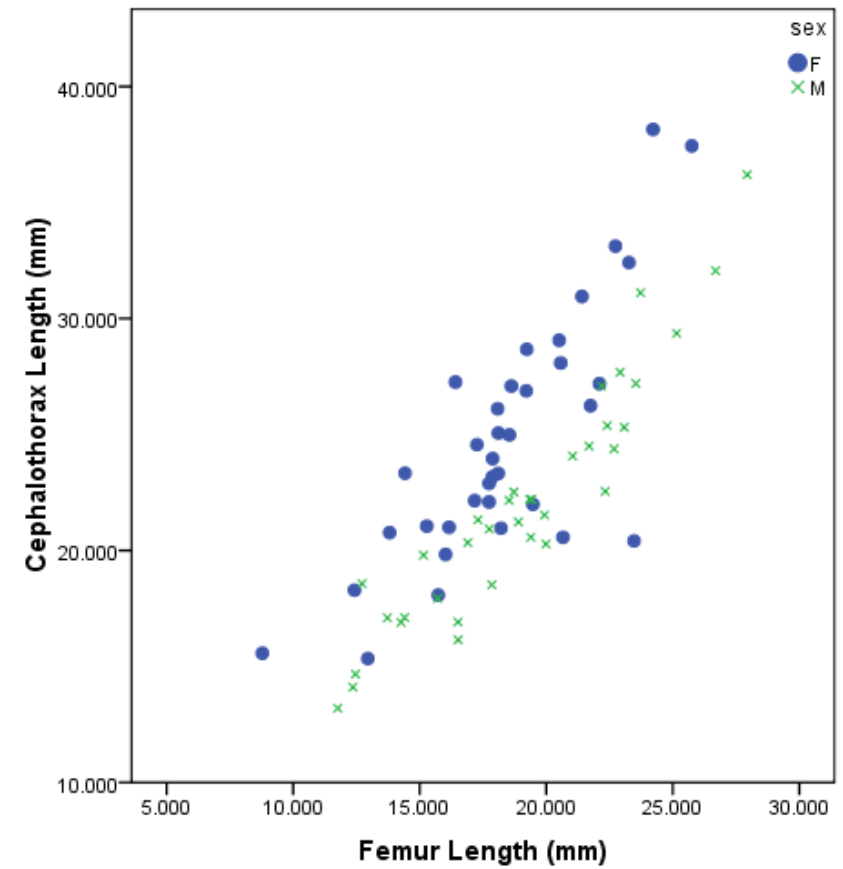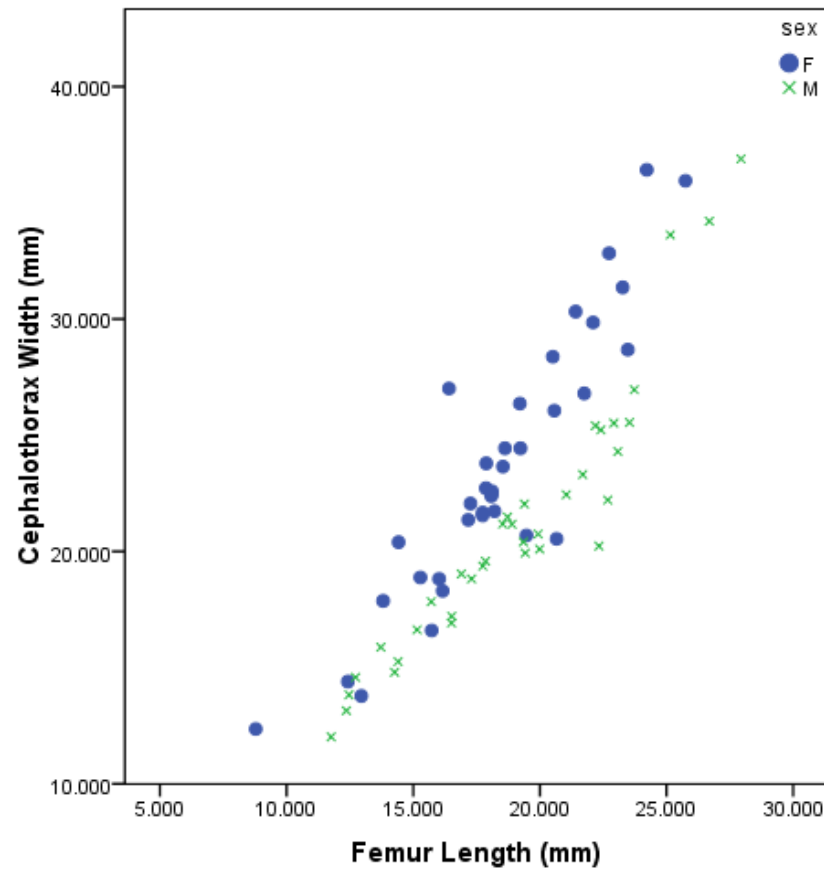
Figure 6.17 a. and b. Graphical representation of the correlation L-F and W-F with relation to sex.
Females are given by blue filled circles while males are green crosses. In both instances there are distinct special regions occupied by each sex with few overlapping individuals.

6.5   CONCLUSIONS/FUTURE WORK

For many decades there has been a preconception that subtle morphological variation between the sexes might be used to accurately sex individuals. The purpose of this study was to investigate whether one such previously unexamined feature, the spines on the posterior tibia of spiders, could be utilised for such a purpose. One of the most significant findings to emerge from this study is that the relative sizes of those spines to the rest of the body display extremely significant differences between the sexes. In doing so, comparisons of these other morphological dimensions were made and these on their own demonstrated a strong sexual divergence. The results of this research support the idea that simple morphometric analyses can be utilised to sex immature spiders to a high degree of accuracy. The implications of this study would allow simple immediate sexing to be carried out with basic tools, providing a mathematical probability of an individual's sex. Likewise, the user-friendly aspect of this technique, i.e. being able to immediately identify the likehood of a particular sex just by looking at a graph, or inserting calculated values into a formula make this an ideal tool for the casual hobbyist. These measurements can be performed *in situ* with minimal stress to the animal. Even if the spines are required to be separated, they can be snapped off quickly and safely with minimal effort (and with regards to safety, at a distance), the spider of course regrowing them post next successive moult. The examination can be performed at least five or six moults before the penultimate moult and advantageously without requiring said moult. However, it can also

be performed on a cast exuvium if available. Sexing would otherwise be possible only upon comparison to an individual of known sex.

There are however certain significant limitations to the data and the subsequent usefulness of this practical technique. Firstly, the data only came from 72 specimens since these were all that were available, hence weakening the power of the test. Despite this, the data were highly statistically significant suggesting the relative size difference between the sexes is large. Perhaps more investigation is required. Secondly, as not all species have these spines e.g. *Haplopelma* spp., this method is not viable for all specimens. Thirdly, it only takes into account a handful of morphometric data thus further studies should be conducted to investigate other potential, unconsidered variations between the sexes.

The function(s) of these spines are as yet unknown. One hypothesis is that they serve the same sensory role as those on insect legs as described by Richards and Richards (1979) who also state the mobility of these spines may be due to a non-sclerotised ring at the base as with the spurs on the legs. If they are sensory, what advantage would they offer a species that has them over one that does not? Conversely, why would a particular genus have evolved to remove them? In the author's personal opinion, the removal of these spines would only serve to agitate the spider and would not incur any deleterious health problems or result in any observable behavioural differences. Therefore, using these to help sex individuals should be considered non-destructive, but perhaps this is scope for further research.

7 SUMMARY OF CONTRIBUTIONS

The aim of this thesis was to try to enlarge upon existing knowledge and approaches in the field of arachnology and proteomics. It has demonstrated how simple methodologies can be utilised to tackle common obstacles such as the recognition of specific motifs within a pool of homologous transcripts/proteins and sexing spiders using simple morphometrics. This research has illustrated the problems of traditional sequencing data analyses, sequencing techniques and morphometric methodologies to sex individual species. Each chapter provides new challenges to universally held ideas and techniques and suggests ways to improve dramatically upon their efficacy and reliability.

Chapter three demonstrated how ESI tandem mass spectrometry can be utilised to sequence fragments of a large macromolecular protein, using silk as a model. The study confirmed previous findings and contributes additional evidence that Spidroin 1 and Spidroin 2 are both utilised in everyday silks of the *Avicularia* spp., moreover in agreement with the suggested ratios (from the repetitive domains, MaSp1 n = 28; MaSp2: n = 1) (Bittencourt *et al.*, 2010) implied therein. These data have definitively shown that the vast majority of useable peptides preferentially originated from within the N-terminal domain, a region that, up until now, has been remarkably challenging to map without whole genome sequencing or massive library screens. Tandem mass spectrometry is very simple yet provides highly informative data with regards to sequence information and motifs as opposed to the commonly utilised cDNA cloning strategies, which are extremely limited in their range of

sequence acquisition capabilities. Here, a contig was able to be constructed

accounting for over 40% of the predicted size of the N-terminal domain, the

first ever recorded for a mygalomorph, which additionally, bore a striking

similarity to a pyriform silk. These data identify previously unidentified silks

within the Mygalomorphae that will hopefully be the subject of future studies.

Chapter four succeeded in its initial aim of uncovering unique

sequences and a probable silk candidate from a random sampling of cDNA

clones. EST analysis successfully identified partial gene sequences that may be

of special significance in the quest for understanding how silk secondary and

tertiary structures are formed as well as numerous previously undescribed

homologues to database theraphosid toxins and ribosomal RNAs. The presence

of incomplete rRNAs was attributed to the fragility corresponding to the

complex secondary structures formed, the harsh protocols involving the

phenolic reagents and/or to the 'hidden break' hypothesis. The evidence here

implies the existence of AU-rich sequences akin to the 'hidden break' regions

of Ishikawa and Newburgh (1972) and to the proposed 'fragile sites' (Casper *et

al*., 2002), which could explain consistent length polymorphisms due to higher

fracture tendencies. There is a high prevalence of nuclear pseudogenes,

evidence of RNA intron editing and transposable elements (McClintock

1948/1950) probably as a result of evolutionary horizontal gene transfer (HGT)

or an undescribed discrete form of heterologous recombination. These data

provide clues to the manipulation and artificial reproduction of protein folding

mechanisms such as that of the silks.

Chapter five details how 29 apparent MiSps, 8 MaSps and 4 pyriform silk sequences were recovered from within three 454 pyrosequencing transcriptomes of the spiders: *S. tentoriicola*, *S. mimosarum* and *S. lineatus*. These contigs, which almost certainly pertain to silk genes were identified through a range of comparative sequence homologies using Microsoft Office Word to identify motifs. The robustness of this approach was hindered by the inability to align them correctly to database voucher models. It illustrates a huge gap in the database, which is so poorly stocked that invariably a string of false positives will always result. The absence of both flagelliform and recognisable glues is consistent with the biology of the stegodyphid species. This is reaffirmed by the pseudo-communal behaviour of this species as entangled prey would probably require more support from conspecifics than those relying solely on glues and radial webs, which appear to be far more effective in aiding prey subduing. Hopefully these transcriptomic data provided by Bilde *et al.* (unpublished) will be the foundation of a new spider database to which definitive retrievals can be made. The implications and scope of such a database from an evolutionary and a comparative studies perspective are vast and will open up a whole host of further studies, which until now have been limited by inadequacies of the current tools.

Finally, **Chapter six** investigated whether the previous common conceptions of dimensional morphologies could be utilised to adequately sex theraphosid individuals. It adds an additional and previously unexamined feature, the spines on the posterior tibia, which likewise, can also be utilised for such a purpose. One of the most significant findings to emerge from this

study is that the relative sizes of those spines to the rest of the body display extremely significant differences between the sexes. Alongside this investigation, comparisons of these other morphological dimensions were made and these on their own demonstrated a strong sexual divergence. The implications of this study would allow simple immediate sexing to be carried out with basic tools, providing a mathematical probability of an individual's sex. These measurements can be performed *in situ* with minimal stress to the animal. This examination can be performed at least five or six moults prior to the penultimate moult, but likewise can be acquired from a moult, which is a means often used to sex spiders if it can be obtained intact. These analyses therefore have highly practical implications as the exuviae virtually always retain these dimensions and features regardless of mechanical and environmental damage post-moult. Despite this, the function(s) of these spines are still as yet unknown but would be an interesting target of future studies. One would hypothesis that they serve the same sensory role as those on insect legs as described by Richards and Richards (1979) but their articulated nature and ability to be moved apparently at will in a manner similar to the pilomotor reflex (goose bumps when cold) suggest a mechanical function which is as yet, not obviously apparent.

8   REFERENCES

Adams, W. W., Farmer, B. and Viney, C. (1994). Silk Polymers: materials science and biotechnology. American Chemical Society Symposium Series. 544:234-251.

Agnarsson, I., Boutry, C. and Blackledge, T. A. (2008). Spider silk aging: initial improvement in a high performance material followed by slow degradation. Journal of Experimental Zoology. Part A, Ecological genetics and physiology. 309(8):494-504.

Agnarsson, I., Boutry, C., Wong, S. C., Baji, A., Dhinojwala, A., Sensenig, A. T. and Blackledge, T. A. (2009a). Supercontraction forces in spider dragline silk depend on hydration rate. Zoology. 112(5):325-331.

Agnarsson, I., Dhinojwala, A., Sahni, V. and Blackledge, T. A. (2009b). Spider silk as a novel high performance biomimetic muscle driven by humidity. Journal of Experimental Biology. 212:1990-1994.

Albo, M. J. and Costa, F. G. (2010). Nuptial gift-giving behaviour and male mating effort in the Neotropical spider *Paratrechalea ornata* (Trechaleidae). Animal Behaviour. 79(5):1031-1036.

Altman, G. H., Diaz, F., Jakuba, C., Calabro, T., Horan, R. L., Chen, J., Lu, H., Richmond. J. and Kaplan, D. L. (2003). Silk-based biomaterials. Biomaterials. 24:401-416.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25:3389-3402.

Andersen, S. O. (1970). Amino acid composition of spider silks. Comparative Biochemistry and Physiology. 35:705-711.

Andersen, T., Bollerup, K., Toft, S. and Bilde, T. (2008). Why do males of the spider *Pisaura mirabilis* wrap their nuptial gifts in silk: female preference or male control? Ethology. 114:775-781.

Apstein, C. (1889). Bau und function der spinndrasen der Araneida. Inaugural Dissertation. Berlin.

Arak, A. (1988). Sexual dimorphism in body size: a model and a test. Evolution. 42:820-825.

Arzt, E., Gorb, S. and Spolenak, R. (2003). From micro to nano contacts in biological attachment devices. Proceedings of the National Academy of Sciences. 100(19):10603-10606.

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

Austad, S. N. and Sunquist, M. E. (1986). Sex-ratio manipulation in the common opossum. Nature. 324:58-60.

Autumn, K., Sitti, M., Liang, Y. A., Peattie, A.M., Hansen, W. R., Sponberg, S., Kenny, T. W., Fearing, R., Israelachvili, J. N. and Full, R. J. (2002). Evidence for van der Waals adhesion in gecko setae. Proceedings of the National Academy of Sciences. 99(19):12252-12256.

Aviles, L. and Maddison, W. (1991). When is the sex-ratio biased in social spiders – chromosome studies of embryos and male meiosis in *Anelosimus* species (Araneae, Theridiidae). Journal of Arachnology. 19:126-135.

Ayoub, N. A., Garb, J. E., Tinghitella, R. M., Collin, M. A. and Hayashi, C. Y. (2007). Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. Public Library of Science ONE. 2(6):e514.

Baoyong, L., Jian, Z., Denglong, C. and Min, L. (2010). Evaluation of a new type of wound dressing made from recombinant spider silk protein using rat models. Burns. 36(6):891-896.

Barghout, J. Y. J., Thiel, B. L. and Viney, C. (1999). Spider (*Araneus diadematus*), cocoon silk: a case of non-periodic lattice crystals with a twist? International Journal of Biological Macromolecules. 24:211-217.

Barghusen, L., Claussen, D., Anderson, M. and Bailer, A. (1997). The effects of temperature on the web-building behaviour of the common house spider, *Achaearanea tepidariorum*. Functional Ecology. 11(1):4-10.

Becker, N., Oroudjev, E., Mutz, S., Cleveland, J. P., Hansma, P. K., Hayashi, C. Y., Makarov, D. E. and Hansma, H. G. (2003). Molecular nanosprings in spider capture-silk threads. Natural Materials. 2:278-283.

Beckwitt, R. and 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. 269:6661-6663.

Beckwitt, R., Arcidiacono, S. and Stote, R. (1998). Evolution of repetitive proteins: spider silks from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae). Insect Biochemical Molecular Biology. 28:121-130.

Bell, F. I., McEwen, I. J. and Viney, C. (2002). Fibre science: supercontraction stress in wet spider dragline. Nature. 416, 37.

Bell, J. R., Bohan, D.A., Shaw, E. M. and Weyman, G. S. (2005). Ballooning dispersal using silk: world fauna, phylogenies, genetics and models. Bulletin of Entomological Research. 95:69-114.

Bellmann, H. (1997). Kosmos-atlas spinnentiere Europas. Kosmos. ISBN 3440107469 (Arachnids of Europe).

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). The Protein Data Bank. Nucleic Acids Research. 28(1):235-242.

Bernard, A. and Krafft, B. (2002). Silk attraction: base of group cohesion and collective behaviours in social spiders. Comptes Rendus Biologies. 325:1153-1157.

Bernardi, G. (1978). Intervening sequences in the mitochondrial genome. Nature. 276(5688):558-559.

Bernstein, H. J. (2000). Recent changes to RasMol, recombining the variants. Trends in Biochemical Sciences. 25(9):453-455.

Bilde, T., Lubin, Y., Smith, D., Schneider, J. M. and Maklakov, A. A. (2005).The transition to social inbred mating systems in spiders: role of inbreeding tolerance in a social predecessor. Evolution. 59:160-174.

Bishop, M. J., Crow, B., Norton, D., Kovalcik, K., George, J. and Bralley, J. A. (2007). A simple and selective method for the measurement of leucine and isoleucine from plasma using electrospray ionization tandem mass spectrometry. Rapid Communications in Mass Spectrometry. 21:1920-1924.

Bittencourt, D., Dittmar, K., Lewis, R. V. and Rech, E. L. (2010). A MaSp2-like gene found in the Amazon mygalomorph spider *Avicularia juruensis*. Comparative Biochemistry and Physiology Part B. 155(4):419-426.

Blackledge, T. A. (1998). Stabilimentum variation and foraging success in *Argiope aurantia* and *Argiope trifasciata* (Araneae, Araneidae). Journal of Zoology. 246:21-27.

Blackledge, T. A. and Wenzel, J. W. (1999). Do stabilimentum in orb-webs attract prey or defend spiders? Behavioural Ecology. 10(4):372-376.

Blackledge, T. A., Boutry, C., Wong, S. C., Baji, A., Dhinojwala, A., Sahni, V. and Agnarsson, I. (2009a). How super is supercontraction? Persistent versus cyclic responses to humidity in spider dragline silk. Journal of Experimental Biology. 212:1981-1989.

Blackledge, T. A., Coddington, J. A. and Agnarsson, I. (2009b). Fecundity increase supports adaptive radiation hypothesis in spider web evolution. Communicative and Integrative Biology. 2:459-463.

Blackledge, T. A. and Hayashi, C. Y. (2006). Unraveling the mechanical properties of composite silk threads spun by cribellate orb-weaving spiders. Journal of Experimental Biology. 209(16):3131-3140.

Blackledge, T. A., Summers, A. P. and Hayashi, C. Y. (2005). Gumfooted lines in black widow cobwebs and the mechanical properties of spider capture silk. Zoology. 108:41-46.

Bland, E. (2009). Spider silk used as artificial muscle. Discovery News [Internet].

Bond, J. E. and Opell, B. D. (1998). Testing adaptive radiation and key innovation hypotheses in spiders. Evolution. 52:403-414.

Bonnet, P. (1930). La mue, l'autotomie et la regeneration chez les Araignées, avec une étude des Dolomèdes d'Europe. Bulletin de la Société d'Histoire Naturelle de Toulouse. 59:237-700.

Boutry, C. and Blackledge, T. A. (2008). The common house spider alters the material and mechanical properties of cobweb silk in response to different prey. Journal of Experimental Biology. 309A:542-552.

Braendegaard, J. (1938). Aeronautic spiders in the arctic. Meddelelser om Grønland. 119(5):1-9.

Brandwood, A. (1985). Mechanical properties and factors of safety of spider drag-lines. Journal of Experimental Biology. 116:141-151.

Brenchley, G. A. (1976). Predator detection and avoidance: ornamentation of tube-caps of *Diopatra* spp. (Polychaeta: Onuphidae). Marine Biology. 38:179-188.

Breslauer, K. J., Frank, R., Blocker, H. and Marky, L. A. (1986). Predicting DNA duplex stability from the base sequence. Proceedings of the National Academy of Sciences. 83:3746-50.

Bristowe, W. S. (1929). The mating habits of spiders, with special reference to the problems surrounding sex dimorphism. Proceedings of the Zoological Society. 21:309-357.

Bristowe, W. S. (1939). The comity of spiders. London: Ray Society.

Brockwell, D. J., Paci, E., Zinober, R. C., Beddard, G. S., Olmsted, P. D., Smith, D. A., Perham, R. N. and Radford, S. E. (2003). Pulling geometry defines the mechanical resistance of a beta-sheet protein. Nature Structural Biology. 10(9):731-737.

Buehler, M. J. and Yung, Y. C. (2009). Deformation and failure of protein materials in physiologically extreme conditions and disease. Nature Materials. 8:175-188.

Bukowski, T. and Christenson, T. (1997). Determinants of sperm release and storage in a spiny orb-weaving spider. Animal Behaviour. 53:381-395.

Calvert, P. (1998). Silk and sequence. Nature. 393:309-310.

Candelas, G. C., Arroyo, G., Carrasco, C. and Dompenciel, R. (1990). Spider silk glands contain a tissue-specific alanine tRNA that accumulates *in vitro* in response to the stimulus for silk protein synthesis. Developmental Biology. 140:215-220.

Cantor, E. J., Creel, H. S., Deguchi, Y., Dougherty, M. J., Kothakota, S., Krejchi, M. T., Matsuki, K., McGrath, K. P., Parkhe, A. D., Atkins, E. D. T., Fournier, M. J., Mason, T. L. and Tirrell, D. A. (1994). *In vivo* synthesis and structural analysis of alanylglycine-rich artificial proteins. In Silk Polymers: Materials Science and Biotechnology (Kaplan, D. *et al*., eds), 98-103, American Chemical Society, Washington, DC, USA.

Cappello, J., Crissman, J., Dorman, M., Mikolajczak, M., Textor, G., Marquet, M. and Ferrari, F. (1990). Genetic engineering of structural protein polymers. Biotechnology Progress. 6:198-202.

Carboni, P. (1952). Silk: biology, chemistry, technology. Translation by Karl Walter. (Chapman and Hall, London).

Case, S. T. and Thornton, J. R. (1999). High molecular mass complexes of aquatic silk proteins. International Journal of Biological Macromolecules. 24:89-101.

Casper, A. M., Nghiem, P., Arlt, M. F. and Glover, T. W. (2002). ATR regulates fragile site stability. Cell. 111:779-789.

Challis, R. J., Goodacre, S. L. and Hewitt, G. M. (2006). Evolution of spider silks: conservation and diversification of the C-terminus. Insect Molecular Biology. 15:45-56.

Chang, J. C., Fletcher, M. J., Gurr, G. M., Kent, D. S. and Gilbert, R. G. (2005). A new silk: mechanical, compositional and morphological characterization of leafhopper (Kahaono, Montana) silk. Polymer. 46(19):7909-7917.

Chen, H., Suk, K., Chi-Young, P., Hun-Sik, K., In-Joo, C. and Hyoung-Joon, J. (2008). pH-triggered transition of silk fibroin from spherical micelles to nanofibrils in water. Macromolecular Research. 16:539-543.

Chen, W. Q., Priewalder, H., John, J. P. and Lubec, G. (2009). Silk cocoon of *Bombyx mori*: proteins and posttranslational modifications: heavy phosphorylation and evidence for lysine-mediated cross links. Proteomics. 10:369-379.

Chen, X., Huang, Y. F., Shao, Z. Z., Huang, Y., Zhou, P., Knight, D. P. and Vollrath, F. (2004). Function of potassium in spinning process of spider *Nephila*. Chemical Journal of Chinese Universities. 25(6):1160-1163.

Cheng, R. C., Yang, E. C., Lin, C. P., Herberstein, M. E. and Tso, I. M. (2010). Insect form vision as one potential shaping force of spider web decoration design. The Journal of Experimental Biology. 213(5):759-768.

Chiarini, A., Petrini, P., Bozzini, S., Pra, I. D. and Armato, U. (2003). Silk fibroin/poly(carbonate)-urethane as a substrate for cell growth: *in vitro* interactions with human cells. Biomaterials. 24:789–99.

Choresh, O., Bayarmagnai, B. and Lewis, R. V. (2009). Spider web glue: two proteins expressed from opposite strands of the same DNA sequence. Biomacromolecules. 10(10):2852–2856.

Coddington, J. A. (1989). Spinneret silk spigot morphology, evidence for the monophyly of orb-weaving spiders, Cyrtophorinae (Araneidae) and the group Theridiidae and Nesticidae. Journal of Arachnology. 17(I):71-95.

Coddington, J. A. (2005). Phylogeny and classification of spiders. Chapter 2 in: Spiders of North America: an identification manual eds. Ubick, D., Paquin, P., Cushing, P. E. and Roth, V. American Arachnological Society. 18-24.

Coddington, J. A. and Levi, H. W. (1991). Systematics and evolution of spiders (Araneae). Annual Review of Ecology and Systematics. 22:565-592.

Colgin, M. A. and Lewis, R. V. (1998). Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like spacer regions. Protein Science. 7:667–672.

Collin, M., Camama, E., Swanson, B. O., Edgerly, J. S. and Hayashi, C. Y. (2009). Comparison of embiopteran silks reveals tensile and structural similarities across taxa. Biomacromolecules. 10(8):2268-2274.

Costa, F. G. and Pérez-Miles, F. (1992). Notes on mating and reproductive success of *Ceropelma longisternalis* (Araneae, Theraphosidae) in captivity. Journal of Arachnology. 20:129-133.

Coyle, F. A. (1983). Aerial dispersal by mygalomorph spiderlings (Araneae, Mygalomorphae). Journal of Arachnology. 11:283-286.

Coyle, F. A. (1985). Ballooning behavior of *Ummidia* spiderlings (Araneae, Ctenizidae) . Journal of Arachnology. 13:137-138.

Coyle, F. A. (1986). The role of silk in prey capture by nonaraneomorph spiders. In: Shear WA, editor. Spiders: webs, behavior and evolution. Palo Alto. (CA). Stanford University Press. 319-363.

Coyle, F. A., Harrison, F. W., MacGimsey, W. C. and Palmer, J. M. (1983). Observations of the structure and function of spermathecae in haplogyne spiders. Transactions of the American Microscopical Society. 102(3):272-280.

Craig, C. L. (1997). Evolution of arthropod silks. Annual Review of Entomology. 42:231-267.

Craig, C. L., Bernard, G. D. and Coddington, J. A. (1994). Evolutionary shifts in the spectral properties of spider silk. Evolution. 48:287-296.

Craig, C. L., Hsu, M., Kaplan, D. and Pierce, N. E. (1999). A comparison of the composition of silk proteins produced by spiders and insects. International Journal of Biological Macromolecules. 24:109-118.

Craig, C. L., Riekel, C., Heberstein, M. E., Weber, R. W., Kaplan, D. and Pierce, N. (2000). Evidence for diet effects on the composition of silk proteins produced by spiders. Molecular Biology and Evolution. 17:1904-1913.

Craig, C. L., Weber, R. S. and Bernard, G. D. (1996). Evolution of predator–prey systems: spider foraging plasticity response to the visual ecology of prey. American Society of Naturalists. 147:205-229.

Crick, F. (1966). Codon–anticodon pairing: the wobble hypothesis. Journal of Molecular Biology. 19(2):548–55.

Crouch, T. and Lubin, Y. (2000). Effects of climate and prey availability on foraging in a social spider, *Stegodyphus mimosarum* (Araneae: Eresidae). Journal of Arachnology. 28:158-168.

Cunniff, P. M., Fossey, S. A., Auerbach, M. A. and Song, J. W. (1994a). Mechanical properties of major ampullate gland silk fibers extracted from *Nephila clavipes* spiders. Silk Polymers. 544:234-251.

Cunniff, P. M., Fossey, S. A., Auerbach, M. A., Song, J. W., Kaplan, D. L., Adams, W. W., Eby, R. K., Mahoney, D. and Vezie, D. L. (1994b). Mechanical and thermal properties of dragline silk from the spider *Nephila clavipes*. Polymers for Advanced Technologies. 5:401-410.

Curcio, M. J. and Derbyshire, K. M. (2003). The outs and ins of transposition: from mu to kangaroo. Nature Reviews Molecular Cell Biology. 4:865-877.

Cusack, B. P., Arndt, P. F., Duret, L. and Crollius, H. R. (2011). Preventing dangerous nonsense: selection for robustness to transcriptional error in human genes. Public Library of Science Genetics. 7:e1002276.

De Tommasi, D., Marzano, S. and Saccomandi, G. (2010). Localization and stability in damageable amorphous solids. Continuum Mechanics and Thermodynamics. 22(1):47-62.

Decae, A. E. (1984). A theory on the origin of spiders and the primitive function of spider silk. Journal of Arachnology. 12:21-28.

Del-Claro, K. and Tizo-Pedroso, E. (2009). Ecological and evolutionary pathways of social behavior in Pseudoscorpions (Arachnida: Pseudoscorpiones). Acta Ethologica. 12:13-22.

Denny, M. (1976). The physical properties of spider's silk and their role in the design of orb-webs. Journal of Experimental Biology. 65:483-506.

Dicko, C., Vollrath, F. and Kenney, J. M. (2004). Spider silk protein refolding is controlled by changing pH. Biomacromolecules. 5:704-710.

Dippenaar-Schoeman, A. S. and Jocqué, R. (1997). African Spiders. An identification manual. Plant Protection Research Institute. Handbook no.9. 392.

Doležel, J., Bartoš, J., Voglmayr, H. and Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. Cytometry. 51A:127-128.

Downes, M. F. (1981). Sexual dimorphism in *Latrodectus* (Araneae, Theridiidae). Australian Journal of Ecology. 6:289-290.

Du, N., Liu, X. Y., Narayanan, J., Li, L. A., Lim, M. L. M. and Li, D. Q. (2006). Design of superior spider silk: from nanostructure to mechanical properties. Biophysical Journal. 91:4528-4535.

Duguet, M., Perasso, R. and Li, Z. G. (2000). Fine organization of *Bombyx mori* fibroin heavy chain gene. Nucleic Acids Research. 28:2413-2419.

Dunlop, J. A. (2010). Geological history and phylogeny of Chelicerata. Arthropod Structure and Development. 39:124-142.

Eberhard, W. G. (1987). How spiders initiate airborne lines. Journal of Arachnology. 15(1):1-10.

Eberhard, W. G. (2006). Stabilimenta of *Philoponella vicina* (Araneae: Uloboridae) and *Gasteracantha cancriformis* (Araneae: Araneidae): evidence against a prey attractant function. Biotropica. 39(2):216-220.

Eberhard, W. G. and Huber, B. A. (2010). Spider genitalia: precise maneuvers with a numb structure in a complex lock. Pp. 249-284. In: J. L. Leonard and A. Córdoba-Aguilar (eds.) Evolution of primary sexual characters in animals. Oxford University Press, Oxford, U.K.

Eckert, R. L. (1988). Sequence of the human 40-kDa keratin reveals an unusual structure with very high sequence identity to the corresponding bovine keratin. Proceedings of the National Academy of Sciences. 85:1114-1118.

Edman, P. J. (1950). Method for determination of the amino acid sequence in peptides. Acta Chemica Scandinavica. 4:283-293.

Edmonds, D. and Vollrath, F. (1992). The contribution of atmospheric water vapour to the formation and efficiency of a spider's web. Proceedings of the Royal Society B. 248:145-148.

Emile, O., Le Floch, A. and Vollrath, F. (2006). Biopolymers: shape memory in spider draglines. Nature. 440:621.

Engelman, D. M. and Steitz, T. A. (1981). The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. Cell. 23(2):411-22.

Enock, F. (1885). The life history of *Atypus piceus* Sulz. Transactions of the Entomological Society of London. 389-420.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred II. Error probabilities. Genome Research. 8(3):186-194.

Faber, D. B. (1994). Przibram's rule and male body-size dimorphism in *Zygoballus rufipes* (Araneae: Salticidae). Journal of Zoology, London. 232:191-198.

Fahnestock, S. R., Yao, Z. and Bedzyk, L. A. (2000). Microbial production of spider silk proteins. Journal of Biotechnology. 74(2):105-19.

Foelix, R. (1996). Biology of Spiders (Oxford Univ. Press, Oxford).

Foo, C. W. P. and Kaplan, D. L. (2002). Genetic engineering of fibrous proteins: spider dragline silk and collagen. Advanced Drug Delivery Reviews. 54:8(18):1131-1143.

Forrest, A. R. and Carninci, P. (2009). Whole genome transcriptome analysis. RNA Biology. 6:107-112.

Forster, R. R., Platnick, N. I. and Gray, M. R. (1987). A review of the spider superfamilies Hypochiloidea and Austrochiloidea (Araneae, Araneomorphae). Bulletin of the AMNH. 185(1):1-116.

Freese, E. (1959). The specific mutagenic effect of base analogues on phage T4. Journal of Molecular Biology. 1(2):87-105.

Frische, S., Maunsbach, A. and Vollrath, F. (1997). Elongate cavities and skin-core structure in *Nephila* spider silk observed by electron microscopy. Journal of Microscopy. 189:64-70.

Frishman, D. and Argos, P. (1995). Knowledge-based protein secondary structure assignment. Proteins. 23:566-579.

Frost, L. S., Leplae, R., Summers, A. O. and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. Nature Reviews Microbiology. 3:722-732.

Fukushima, Y. (2000). Secondary structural analysis in the solid state for analogous sequential polypeptides of glycine-rich sequence of spider dragline silk. Polymer Bulletin. 45:237-244.

Galiano, M. E. (1984). Datos adicionales sobre el ciclo vital de *Acanthoscurria sternalis*. Revista de la Sociedad Entomologica Argentina. 43(1-4):45-55.

Garb, J. E. and Hayashi C. Y. (2005). Modular evolution of egg case silk genes across orb-weaving spider superfamilies. Proceedings of the National Academy of Sciences. 102(32):11379-11384.

Garb, J. E., Ayoub, N. A. and Hayashi, C. Y. (2010). Untangling spider silk evolution with spidroin N-terminal domains. BMC Evolutionary Biology. 10(1):243.

Garb, J. E., DiMauro, T., Lewis, R. and Hayashi, C. Y. (2007). Expansion and intragenic homogenization of spider silk genes since the Triassic: evidence from Mygalomorphae (tarantulas and their kin) spidroins. Molecular Biology and Evolution. 24:2454-2464.

Garrido, M. A., Elices, M., Viney, C. and Pérez-Rigueiro, J. (2002). Active control of spider silk strength: comparison of drag line spun on vertical and horizontal surfaces. Polymer. 43:1537-1540.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research. 31:3784-3788.

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

Gerhardt, U. (1921). Vergleichende Studien über die Morphologie des männlichen Tasters und die Biologie der Kopulation der Spinnen. Archiv fur Naturgeschichte. 87(A,4):78-247. pls.1-3.

Gerhardt, U. (1923). Weitere sexualbiologische Untersuchung an Spinnen. Archiv fur Naturgeschichte. 89(A,10):1-225. pls.1-3.

Gertsch, W. J. (1979). American Spiders, 2nd edition. New York: Van Nostrand Reinhold Company.

Gharahdaghi, F. (1999). Mass spectrometric identification of proteins from silverstained polyacrylamide gel: a method for the removal of silver ions to enhance sensitivity. Electrophoresis. 20:601-605.

Giraldeau, L. A. (1997). The ecology of information use. In: Krebs, J. and Davies, N. B. (Eds.), Behavioural ecology. An evolutionary approach. Blackwell, Oxford. 42-58.

Gonzaga, M. O. and Vasconellos-Neto, J. (2005). Testing the functions of detritus stabilimenta in webs of *Cyclosa fililineata* and *Cyclosa morretes* (Araneae:

Araneidae): do they attract prey or reduce the risk of predation? Ethology. 111:479-491.

Gorb, S. N., Niederegger, S., Hayashi, C. Y., Summers, A. P., Vötsch, W. and Walther, P. (2006). Silk-like secretion from tarantula feet. Nature. 443:407.

Gorb, S. N., Niederegger, S., Hayashi, C. Y., Summers, A. P., Vötsch, W. and Walther, P. (2009). Silk production from tarantula feet questioned reply. Nature. 461:9-10.

Gosline, J. M., Denny, M. W. and Demont, M. E. (1984). Spider silk as rubber. Nature. 309:551-552.

Gould, H. J. (1967). The nature of high molecular weight fragments of ribosomal RNA. Journal of Molecular Biology. 29(2):307-312.

Grassé, P. (1949). Treaty of Zoology. Anatomy, Systematics, Biology ed. published. Volume 6.

Gray, M. W., Lang, B. F. and Burger, G. (2004). Mitochondria of protists. Annual Review of Genetics. 38:477-524.

Gray, M. W., Lang, B. F., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T. G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, M. and Burger, G. (1998). Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Research. 26:865-878.

Gregory, T. R. (2012). Animal Genome Size Database. http://www.genomesize.com.

Gregory, T. R. and Shorthouse, D. P. (2003). Genome sizes of spiders. Journal of Heredity. 94:285-290.

Griffiths, J. R. and Salinatri, V. R. (1980). The strength of spider silk. Journal of Materials Science. 15:491-496.

Griswold, C. E., Coddington, J. A., Platnick, N. I. and Forster, R. R. (1999). Towards a phylogeny of entelegyne spiders (Araneae, Araneomorphae, Entelegynae). Journal of Arachnology. 27:53-63.

Grubb, D. T. and Jelinski, L. W. (1997). Fiber morphology of spider silk: the effects of tensile deformation. Macromolecules. 30:2860-2867.

Guerette, P., Ginzinger, D., Weber, B. and Gosline, J. M. (1996). Silk properties determined by gland-specific expression of a spider fibroin gene family. Science. 272:112-115.

Guinea, G. V., Elices, M., Perez-Rigueiro, J. and Plaza, G. (2003). Self-tightening of spider silk fibers induced by moisture. Polymer. 44:5785-5788.

Gunnarsson, B. and Andersson, A. (1992). Skewed primary sex-ratio in the solitary spider *Pityohyphantes phrygianus*. Evolution. 46:841-845.

Hagn, F., Eisoldt, L., Hardy, J. G., Vendrely, C. and Coles, M. (2010). A conserved spider silk domain acts as a molecular switch that controls fibre assembly. Nature. 465:239-242.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41:95-98.

Hancock, K. and Hancock, J. (1994). Sex determination of immature theraphosid spiders from their cast skins. R and A Publishing Ltd., Somerset, England.

Hardy, J. G., Romer, L. M. and Scheibel, T. R. (2008). Polymeric materials based on silk proteins. Polymer. 49:4309-4327.

Haupt, J. (2003). The Mesothelae-a monograph of an exceptional group of spiders (Araneae: Mesothelae). Zoologica. 154:1-102.

Hayakawa, J. (1989). Life and its prediction of high polymers (I.P.C., Tokyo).

Hayashi, C. Y. and Lewis, R. V. (1998). Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. Journal of Molecular Biology. 275(5):773-784.

Hayashi, C. Y. and Lewis, R. V. (2000). Molecular architecture and evolution of a modular spider silk protein gene. Science. 287:1477-1479.

Hayashi, C. Y. and Lewis, R. V. (2001). Spider flagelliform silk: lessons in protein design, gene structure and molecular evolution. BioEssays. 23:750-756.

Hayashi, C. Y., Blackledge, T. A. and Lewis, R. V. (2004). Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. Molecular Biology and Evolution. 21(10):1950-1959.

Hayashi, C. Y., Shipley, N. H. and Lewis, R. V. (1999). Hypotheses that correlate the sequence, structure and mechanical properties of spider silk proteins. International Journal of Biological Macromolecules. 24:271-275.

Hazan, A., Gertler, A., Tahori, A. S. and Gerson, U. (1975). Spider mite webbing - III. Solubilization and amino acid composition of the silk protein. Comparative Physiology and Biochemistry B. 51:457-462.

Hebbar, S. K., Belcher, S. M. and Perlman, P. S. (1992). A maturase-encoding group IIA intron of yeast mitochondria self-splices *in vitro*. Nucleic Acids Research. 20(7):1747-54.

Hedhammar, M., Rising, A., Grip, S., Saenz Martinez, A., Nordling, K., Casals, C., Stark, M. and Johansson, J. (2008). Structural properties of recombinant nonrepetitive and repetitive parts of major ampullate Spidroin 1 from *Euprosthenops australis*: implications for fiber formation. Biochemistry. 47:3407-3417.

Hedin, M. and Bond, J. E. (2006). Molecular phylogenetics of the spider infraorder Mygalomorphae using nuclear rRNA genes (18S and 28S): conflict and agreement with the current system of classification. Molecular Phylogenetics and Evolution. 41:454-471.

Heim, M., Römer, L. and Scheibel. T. (2010). Hierarchical structures made of proteins. The complex architecture of spider webs and their constituent silk proteins. Journal of Structural Biology. 170(2):420-425.

Henschel, J. R. (1998). Predation on social and solitary individuals of the spider *Stegodyphus dumicola* (Araneae, Eresidae). Journal of Arachnology. 26:61-69.

Heslot, H. (1998). Artificial fibrous proteins: a review. Biochimie. 80:19-31.

Higgins, D. G., Thompson, J. D. and Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. Methods in Enzymology. 266:383-402.

Higgins, L. E., Townley, M. A., Tillinghast, E. K. and Rankin, M. A. (2001). Variation in the chemical composition of orb-webs built by the spider *Nephila clavipes* (Araneae, Tetragnathidae). Journal of Arachnology. 29:82-94.

Hill, J. E. (2009). Appendix A: introduction of silk cultivation to Khotan in the 1st Century CE. 466-467.

Hinman, M. B. and Lewis, R. V. (1992). Isolation of a clone encoding a second dragline silk fibroin. *Nephila clavipes* dragline silk is a two-protein fiber. Journal of Biological Chemistry. 267:19320-19324.

Hiroaki, Y. and Tetsuo, A. (1990). The structure of *Bombyx mori* silk fibroin membrane. Journal of Applied Polymer Science. 40:1745-56.

Hoefler, C. D., Keiser, C. N. and Rypstra, A. (2010). Sex differences in early instar behavior in *Pholcus phalangioides* (Araneae: Pholcidae). The Journal of Arachnology. 38:581-583.

Hu, X., Kohler, K., Falick, A. M., Moore, A. M., Jones, P. R., Sparkman, O. D. and Vierra, C. A. (2005). Egg case protein-1. A new class of silk proteins with fibroin-like properties from the spider *Latrodectus hesperus*. The Journal of Biological Chemistry. 280:21220-21230.

Hu, X., Lawrence, B., Kohler, K., Falick, A. M., Moore, A. M., McMullen, E., Jones, P. R. and Vierra, C. A. (2005). Araneoid egg case silk: a fibroin with novel ensemble repeat unit from the black widow spider, *Latrodectus hesperus*. Biochemistry. 44:10020-10027.

Hu, X., Vasanthavada, K., Kohler, K., McNary, S., Moore, A. M. F. and Vierra, C. A. (2006). Molecular mechanisms of spider silk. Cellular and Molecular Life Sciences. 63(17):1986-1999.

Huang, X. and Madan, A. (1999). CAP3: a DNA sequence assembly program. Genome Research. 9:868-877.

Huber, B. A. (1995). The retrolateral tibial apophysis in spiders - shaped by sexual selection? Zoological journal of the Linnean Society. 113:151-163.

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L. and Welch, D. M. (2007). Genome Biology. 8(7):R143.

Hyde, N. (1984). The queen of textiles. National Geographic. 165:3-49.

Inouye, K., Kurokawa, M., Nishikawa, S. and Tsukada, M. (1998). Use of *Bombyx mori* silk fibroin as a substratum for cultivation of animal cells. Journal of Biochemical and Biophysical Methods. 37:159-164.

Ishikawa, H. and Newburgh, R. W. (1972). Studies of the thermal conversion of 28S RNA of *Galleria mellonella* (L.) to an 18S product. Journal of Molecular Biology. 64(1):135-144.

Jelinski, L. W., Blye, A., Liivak, O., Michal, C., LaVerde, G., Seidel, A., Shah, N. and Yang, Z. (1999). Orientation, structure, wet-spinning and molecular basis for supercontraction of spider dragline silk.International Journal of Biological Macromolecules. 24:197-201.

Jin, H. J. and Kaplan, D. L. (2003). Mechanism of silk processing in insects and spiders. Nature. 424:1057-1061.

Jocqué, R. (1981). On reduced size in spiders from marginal habitats. Oecologia. 49:404-408.

Johannesen, J. and Lubin, Y. (1999). Group founding and breeding structure in the subsocial spider *Stegodyphus lineatus* (Eresidae). Heredity. 82:677-686.

Jun, M. and Yoshiko, M. (1981). Journal of Polymer Science, Polymer Physics Edition. 19:185.

Jun, M., Yoshiko, M., Shigeo, N., Nobutami, K. and Masao, K. J. (1977). Journal of Polymer Science, Polymer Physics Edition. 15:1675.

Kaplan, D. L., Fossey, S., Mello, C. M., Arcidiacono, S., Senecal, K., Muller, W., Stockwell, S., Beckwitt, R., Viney, C. and Kerkam, K. (1992a). Biosynthesis and processing of silk proteins. Materials Research Society Bulletin. 10:41-47.

Kaplan, D. L., Fossey, S., Viney, C. and Muller, W. (1992b). Self-organization-assembly in biosynthesis of silk fibers - a hierarchical problem. In: Aksay, I. A., Baer, E., Sarikaya, M., Farmer, B. and Viney, C. Eds., Silk Polymers: Materials Science and Biotechnology. American Chemical Society Symposium Series. 544:67-79.

Kawamoto, T. H. and Japyassu, H. F. (2008). Tenacity and silk investment of two orb weavers: considerations about diversification of the Araneoidea. Journal of Arachnology. 36:418-424.

Kerkam, K., Viney, C., Kaplan, D. and Lombard, S. (1991). Liquid crystallinity of natural silk secretions. Nature. 349:596-598.

Keten, S. and Buehler, M. J. (2008a). Asymptotic strength limit of hydrogen bond assemblies in proteins at vanishing pulling rates. Physical Review Letters. 100:198301.

Keten, S. and Buehler, M. J. (2008b). Geometric confinement governs the rupture strength of H-bond assemblies at a critical length scale. Nano Letters. 8:743-748.

Keten, S., Xu. Z., Ihle, B. and Buehler, M. J. (2010). Nanoconfinement controls stiffness, strength and mechanical toughness of β-sheet crystals in silk. Nature Materials. 9:359-367.

Khaitlina, S. Y. (2001). Functional specificity of actin isoforms. International Review of Cytology. 202:35-98.

Kim, U. J., Park, J., Li, C., Jin, H. J., Valluzzi, R. and Kaplan, D. L. (2004). Structure and properties of silk hydrogels. Biomacromolecules. 5:786-92.

Kitagawa, M. and Kitayama, T. (1997). Mechanical properties of dragline and capture thread for the spider *Nephila clavata*. Journal of Materials Science. 32:2005-2012.

Knight, D. P and Vollrath, F. (2001a). Comparison of the spinning of selachian egg case ply sheets and orb-web spider dragline filaments. Biomacromolecules. 2:323-334.

Knight, D. P. and Vollrath, F. (2001b). Changes in element composition along the spinning duct in a *Nephila* spider. Naturwissenschaften. 88(4):179-182.

Ko, M. S. H., Kitchen, J. R., Wang, X., Threat, T. A., Hasegawa, A., Sun, T., Grahovac, M. J., Kargul, G. J., Lim, M. K., Cui, Y., Sano, Y., Tanaka, T., Liang, Y., Mason, S., Paonessa, P. D., Sauls, A. D., DePalma, G. E., Sharara, R., Rowe, L. B., Eppig, J., Morrell, C. and Doi, H. (2000). Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. Development. 127:1737-1749.

Korayem, A. M., Hauling, T., Lesch, C., Fabbri, M., Lindgren, M., Loseva, O., Schmidt, O., Dushay, M. S. and Theopold, U. (2007). Evidence for an immune function of lepidopteran silk proteins. Biochemical and Biophysical Research Communications. 352:317-322.

Kovoor, J. and Zylberberg, L. (1980). Fine structural aspects of silk secretion in a spider (*Araneus diadematus*). Elaboration in the pyriform glands. Tissue and Cell. 12(3):547-556.

Kraus, O. and Kraus, M. (1992). Eresid spiders in the neotropics: *Stegodyphus manaus* n.sp. (Arachnida, Araneae, Eresidae). Verhandlungen des Naturwissenschaftlichen. 33:15-19.

Kraus, O. and Kraus. M. (1988). The genus *Stegodyphus* (Arachnida, Araneae). Sibling species. Species groups and parallel origin of social living. Verhandlungen des naturwissenschaftlichen. 30:151-254.

Kullmann, E. J. and Zimmermann, W. (1974). Regurgitationsfütterung als Bestandteil der Brutfürsorge bei Haubennetz und Röhrenspinnen (Araneae, Theridiidae und Eresidae). Proceedings of the International Arachnological Congress. 6:1125-1146.

Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. Journal of Molecular Biology. 157:105-142.

La Mattina, C., Reza, R., Hu, X., Falick, A. M., Vasanthavada, K., McNary, S., Yee, R. and Vierra, C. A. (2008). Spider minor ampullate silk proteins are

constituents of prey wrapping silk in the cob weaver *Latrodectus hesperus*. Biochemistry. 47:4692-4700.

Laemmli, U. K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature. 227:680-685.

Lazaris, A., Arcidiacono, S., Huang, Y., Zhou, J., Duguay, F., Chretien, N., Welsh, E. A., Soares, J. W. and Karatzas, C. N. (2002). Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. Science. 295:472-476.

Leborgne, B. and Pasquet, A. (1987). Influence of conspecific silk-structures on the choice of a web-site by the spider *Zygielle x-notata* (Clerck). Revue Arachnologique. 7:85-90.

Lee, H., Lee, B. P. and Messersmith, P. B. (2007). A reversible wet/dry adhesive inspired by mussels and geckos. Nature. 448:338–341.

Lefevre, T., Rousseau, M. E. and Pezolet, M. (2007). Protein secondary structure and orientation in silk as revealed by Raman spectromicroscopy. Biophysical Journal. 92:2885-2895.

Li, D. and Lee, W. S. (2004). Predator-induced plasticity in web-building behaviour. Animal Behaviour. 67:209-318.

Li, D., Lim, M. L. M., Seah, W. K. and Tay, S. L. (2004). Prey attraction as a possible function of discoid stabilimenta of juvenile orb-spinning spiders. Animal Behaviour. 68:629-635.

Li, G., Zhou, P., Shao, Z., Xie, X., Chen, X., Wang, H., Chunyu, L. and Yu, T. (2001). The natural silk spinning process. A nucleation-dependent aggregation mechanism. European Journal of Biochemistry. 268:6600-6606.

Liang, B., Luo, M., Scott-Herridge, J., Semeniuk, C., Mendoza, M., Capina, R., Sheardown, B., Ji, H., Kimani, J., Ball, B. T., Van Domselaar, G., Graham, M., Tyler, S., Jones, S. J. and Plummer, F. A. (2011). A comparison of parallel pyrosequencing and Sanger clone-based sequencing and its impact on the characterization of the genetic diversity of HIV-1. Public Library of Science ONE. 6(10):e26745.

Liebmann, B., Huemmerich, D., Scheibel, T. and Fehr, M. (2008). Formulation of poorly water-soluble substances using self-assembling spider silk protein. Colloids and Surfaces A. 331(1-2), 126-132.

Liu, Y., Shao, Z. and Vollrath, F. (2005). Relationships between supercontraction and mechanical properties of spider silk. Nature Materials. 4:901-905.

Liu, Y., Sponner, A., Porter, D. and Vollrath, F. (2008). Proline and processing of spider silks. Biomacromolecules. 9(1):116-121.

Longhorn, S. J., Nicholas, M., Chuter, J. and Volger A. P. (2007). The utility of molecular markers from non-lethal DNA samples of the CITES II protected ''tarantula'' *Brachypelma vagans* (Araneae, Theraphosidae). Journal of Arachnology. 35:278-292.

Lopez, A. (1988). Glandes épigastrique et classification des araneides. Bulletin of the Associations of Anatomists, Paris. 58:899-904.

Lopez, A. and Emerit, M. (1988). New data on the epigastric apparatus of male spiders. Bulletin of the British Arachnological Society.7:220-224.

Lovett, M., Cannizzaro, C., Vunjak-Novakovic, G. and Kaplan, D. L. (2008). Gel spinning of silk tubes for tissue engineering. Biomaterials 29:4650-4657.

Lubin, Y. D. (1975). Stabilimenta and barrier webs in the orb-webs of *Argiope argentata* (Araneae: Araneidae) on Daphne and Santa Cruz Islands, Galapagos. Journal of Arachnology. 2:119-126.

Lubin, Y. D. (1986). Web building and prey capture in the Uloboridae. 132-171. In Spiders: Webs, Behavior and Evolution. (W.A. Shear, ed.). Stanford University Press, Stanford, California.

Madsen, B. and Vollrath, F. (2000). Mechanics and morphology of silk drawn from anesthetized spiders. Naturwissenschaften. 87(3):148-153.

Madsen, B., Shao, Z. Z. and Vollrath, F. (1999). Variability in the mechanical properties of spider silks on three levels: interspecific, intraspecific and intraindividual. International Journal of Biological Macromolecules. 24:301-306.

Mahmoudi, N., Modanu, M., Brandt, Y. and Andrade, M. C. B. (2008). Subtle pedipalp dimorphism: a reliable method for sexing juvenile spiders. Journal of Arachnology. 36:513-517.

Maklakov, A. A. and Lubin, Y. (2006). Indirect genetic benefits in a spider with direct costs of mating. Behavioural Ecology and Sociobiology. 61:31-38.

Maklakov, A. A., Bilde, T. and Lubin, Y. (2005). Sexual conflict in the wild: elevated mating rate reduces female lifetime reproductive success. American Naturalist. 165:38-45.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W. and Srinivasan, M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature. 437:376-380.

Marples, B. J. (1967). The spinnerets and epiandrous glands of spiders. Zoological Journal of the Linnean Society. 46:209-222.

McClintock, B. (1948). Mutable loci in maize. Carnegie Institute of Washington Year Book 47:155-169.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences. 36:344-55.

McCook, H. C. (1889). American Spiders and their Spinning Work. Volume 1. Published by the author, Philadelphia.

Melchers, Z. (1964). 'Zur Biologie Vogelspinnen (Fam. Avicullariidae)', Morphology. Ökologie der Tiere 53.

Mello, C. M., Senecal, K., Yeung, B., Vouros, P. and Kaplan, D. L. (1994). Initial characterization of *Nephila clavipes* dragline protein. In: Kaplan, D. L., Adams, W. W., Farmer, B. and Viney, C., Editors, (1994). Silk polymers: materials science and biotechnology. American Chemical Society Symposium Series. 544:67-79.

Minoura, N., Aiba, S., Higuchi, M., Gotoh, Y., Tsukada, M. and Imai, Y. (1995). Attachment and growth of fibroblast cells on silk fibroin. Biochemical and Biophysical Research Communications. 208:5115-16.

Mita, K., Ichimura, S. and James, T. C. (1994). Highly repetitive structure and its organization of the silk fibroin gene. Journal of Molecular Evolution. 38:583-592.

Motriuk-Smith, D., Smith, A., Hayashi, C. Y. and Lewis, R. V. (2005). Analysis of the conserved N-terminal domains in major ampullate spider silk proteins. Biomacromolecules. 6:3152-3159.

Nager, R. G., Monaghan, P., Griffiths, P., Houston, D. C. and Dawson. R. (1999). Experimental demonstration that offspring sex ratio varies with maternal condition. Proceedings of the National Academy of Sciences. 96:570-573.

Okada, S., Weisman, S., Trueman, H. E., Mudie, S. T., Haritos, V. S. and Sutherland, T. D. (2008). An Australian webspinner species makes the finest known insect silk fibers. International Journal of Biological Macromolecules. 1:43(3):271-275.

Opell, B. D. (1984). A simple method for measuring desiccation resistance of spider egg sacs. Journal of Arachnology. 12:245.

Opell, B. D. (1996). Functional similarities of spider webs with diverse architectures. American Naturalist. 148:630-648.

Opell, B. D. (1998). Economics of spider orb-webs: the benefits of producing adhesive capture thread and of recycling silk. Functional Ecology. 12:613-624.

Opell, B. D. and Bond, J. E. (2000). Capture thread extensibility of orb-weaving spiders: testing punctuated and associative explanations of character evolution. Biological Journal of the Linnean Society. 70:107-120.

Opell, B. D. and Hendricks, M. L. (2007). Adhesive recruitment by the viscous capture threads of araneoid orb-weaving spiders. Journal of Experimental Biology. 210:553-560.

Opell, B. D. and Hendricks, M. L. (2010). The role of granules within viscous capture threads of orb-weaving spiders. Journal of Experimental Biology. 213:339-346.

Opell, B. D. and Schwend, H. S. (2007). The effect of insect surface features on the adhesion of viscous capture threads spun by orb-weaving spiders. Journal of Experimental Biology. 210:2352 -2360.

Opell, B. D., Bond, J. E. and Warner, D. A. (2006). The effects of capture spiral composition and orb-web orientation on prey interception. Zoology. 109:339-345.

Osaki, S. (1999). Is the mechanical strength of spider's drag-lines reasonable as lifeline? International Journal of Biological Macromolecules. 24:283-287.

Palmer, J. M. (1985). The silk and silk production system of the funnel web mygalomorph spider *Euagrus*. (Araneae: Dipluridae). Journal of Morphology. 186:195-207.

Palmer, J. M., Coyle, F. A. and Harrison, F. W. (1982). Structure and cytochemistry of the silk glands of the mygalomorph spider *Antrodiaetus unicolor* (Araneae, Antrodiaetidae). Journal of Morphology. 174:269-274.

Partridge, L. (1978). Habitat selection. In: Krebs, J. R. and Davies, N. B. (Eds.), Behavioural ecology: an evolutionary approach (pp. 351-76), Sunderland, M.A. Sinauer Associates.

Peatman, E. and Liu, Z. (2007). Microarray fundamentals: basic principles and application in aquaculture. Aquaculture Genome Technologies, Blackwell Publishing. 355-368.

Pérez-Miles, F., Lucas, S. M., da Silva, P. and Bertani, R. (1996). Systematic revision and cladistic analysis of Theraphosinae (Araneae: Theraphosidae). Mygalomorph. 1:33-68.

Perez-Miles, F., Panzera, A., Ortiz-villatoro, D. and Perdomo, C. (2009). Silk production from tarantula feet questioned. Nature. 461;7267: E9.

Pérez-Rigueiro, J., Plaza, G. R., Torres, F. G., Hijar, A., Hayashi, C., Perea, G. B., Elices, M. and Guinea. G.V. (2010). Supercontraction of dragline silk spun by lynx spiders (Oxyopidae). International Journal of Biological Macromolecules. 46(5):555-557.

Perkins, D. N., Pappin, D. J., Creasy, D. M. and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 20(18):3551-3567.

Perry, D. J., Bittencourt, D., Siltberg-Liberles, J., Rech, E. L. and Lewis, R. V. (2010). Piriform spider silk sequences reveal unique repetitive elements. Biomacromolecules. 11:3000-3006.

Peters, H. M. (1993). Über das problem der stabilimente in spinnennetzen. Zoologische Jahrbucher, Abteilung für allgemeine Zoologie und Physiologie der Tiere. 97:245-264.

Petersen, S. D., Mason, T., Akber, S., West, R., White, B. and Wilson, P. (2007). Species identification of tarantulas using exuviae for international wildlife law enforcement. Conservation Genetics. 8(2):497-502.

Platnick, N. I . and Gertsch, W. J. (1976). The suborders of spiders: a cladistic analysis (Arachnida: Araneae). American Museum Novitates. 2607:1-1 5.

Platnick, N. I. (2012). The world spider catalog, version 12.5. American Museum of Natural History, online at http://research.amnh.org/iz/spiders/catalog. DOI:10.5531/db.iz.0001.

Platnick, N. I. and Goloboff, P. A. (1985). On the monophyly of the spider suborder Mesothelae (Arachnida: Araneae). Journal of the New York Entomological Society. 93:1265-1270.

Pogozelski, E., Abramo, D., Papasergi, L., See, B., Kieffer, C. and Padalino, S. (2008). Impact of beta radiation on the strength of *Steatoda triangulosa* spider silk. Journal of Materials Science. 43(17):5891-5897.

Porter, D. and Vollrath, F. (2007). Nanoscale toughness of spider silk. Nano Today. 2(3):6.

Pouchkina, N. N., Stanchev, B. S. and McQueen-Mason, S. J. (2003). From EST sequence to spider silk spinning: identification and molecular characterisation of *Nephila senegalensis* major ampullate gland peroxidase NsPox. Insect Biochemistry and Molecular Biology. 33:229-238.

Pouchkina, N. N., Stanchev, B. S. and McQueen-Mason, S. J. (2004). Molecular studies of a novel dragline silk from a nursery web spider *Euprosthenops* sp. (Pisauridae). Comparative Biochemistry and Physiology B. 138:371-376.

Poza, P., Perez-Rigueiro, J., Elices, M. and Llorca, J. (2002). Fractographic analysis of silkworm and spider silk. Engineering Fracture Mechanics. 69(9):1035-1048.

Putthanarat, S., Tapadia, P., Zarkoob, S., Miller, L. D., Eby, R. K. and Adams, W. W. (2004). The color of dragline silk produced in captivity by the spider *Nephila clavipes*. Polymer. 45(6):1933-1937.

Qin, Z. and Buehler, M. J. (2010). Cooperative deformation of hydrogen bonds in beta-strands and beta-sheet nanocrystals. Physical Review E. 82(6):061906.

Rammensee, S., Slotta, U., Scheibel, T. and Bausch, A. R. (2008). Assembly mechanism of recombinant spider silk proteins. Proceedings of the National Academy of Sciences. 105:6590-6595.

Randall, J. B. (1981). Regeneration and autotomy exhibited by the black widow spider, *Latrodectus variolus* Walckenaer. W. Roux's Archives of Developmental Biology. 190:230-232.

Raven, R. J. (2005). A new tarantula species from northern Australia (Araneae, Theraphosidae). Zootaxa. 1004:15-28.

Richards, G. A. and Richards, P. A. (1979). The cuticular protuberances of insects. International Journal of Insect Morphology and Embryology. 8:143-157.

Rising, A., Widhe, M., Johansson, J. and Hedhammar, M. (2011). Spider silk proteins: recent advances in recombinant production, structure-function relationships and biomedical applications. Cellular and Molecular Life Sciences. 68(2):169-184.

Roberts, M. J. (1995). Collins Field Guide: Spiders of Britain and Northern Europe. Harper Collins, London.

Rodríguez-Gironés, M. A., Corcobado, G. and Moya-Laraño J. (2010). Silk elasticity as a potential constraint on spider body size. Journal of Theoretical Biology. 266:430-435.

Rodriguez-Martin, A., Acosta, R., Liddell, S., Nunez, F., Benito, M. J. and Asensio, M. A. (2010). Characterization of the novel antifungal chitosanase PgChP and the encoding gene from *Penicillium chrysogenum*. Applied Microbiology and Biotechnology. 88(2):519-528.

Rosenberg, A. H., Goldman, E., Dunn, J. J., Studier, F. W. and Zubay, G. (1993). Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system. Journal of Bacteriology. 175:716-722.

Rousseau, M. E., Lefevre, T., Beaulieu, L., Asakura, T. and Pezolet, M. (2004). Study of protein conformation and orientation in silkworm and spider silk fibers using Raman microspectroscopy. Biomacromolecules. 5:2247-2257.

Rozen, S. and Skaletsky, H. J. (2000). Primer3 on the WWW. (http://frodo.wi.mit.edu/) for general users and for biologist programmers. In: Krawetz, S. and Misener, S. (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ. 365-386.

Sahni, V., Blackledge, T. A. and Dhinojwala, A. (2010). Viscoelastic solids explain spider web stickiness. Nature Communications. 1(19):1-4.

Salomon, M. (2007). Western black widow spiders express state-dependent web-building strategies tailored to the presence of neighbours. Animal Behaviour. 73:865-875.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Cold Spring Harbor (Eds.), Molecular Cloning: A Laboratory Manual, (2nd ed.): ISBN 0-87969-309-6.

Savory, T. H. (1952). The Spider's Web. Warne and Co. London and New York.

Sayle, R. and James Milner-White, E. (1995). RasMol: biomolecular graphics for all. Trends in Biochemical Sciences. 20(9):374.

Schafer, A., Vehoff, T., Glisovic, A. and Salditt, T. (2008). Spider silk softening by water uptake: an AFM study. European Biophysics Journal with Biophysics Letters. 37:197-204.

Scheibel, T. (2004). Spider silks: recombinant synthesis, assembly, spinning and engineering of sythetic proteins. Microbial Cell Factories. 3:14.

Schneider, J. M. and Lubin, Y. (1996). Infanticidal male eresid spiders. Nature. 381:655-656.

Schneider, J. M., Roos, J., Lubin, Y. and Henschel, J. R. (2001). Dispersal of *Stegodyphus dumicola* (Araneae, Eresidae): they do balloon after all! Journal of Arachnology. 29:114-116.

Schuck-Paim, C. and Jimenez Alonso, W. (2001). Deciding where to settle: conspecific attraction and web site selection in the orb-web spider *Nephilengys cruentata*. Animal Behaviour. 62:1007-1012.

Schutz, D., Taborsky, M. and Drapela, T. (2007). Air bells of water spiders are an extended phenotype modified in response to gas composition. Journal of Experimental Zoology. 307A:549-555.

Shultz, J. W. (1987). The origin of the spinning apparatus in spiders. Biological Reviews. 62:89-113.

Seah, W. K. and Li, D. (2002). Stabilimentum variations in *Argiope versicolor* (Araneae: Araneidae) from Singapore. Journal of Zoology, London. 258:531-540.

Sehnal, F. and Zurovec, M. (2004). Construction of silk fiber core in Lepidoptera. Biomacromolecules. 5:666-674.

Seibt, U. and Wickler, W. (1988). Bionomics and social structure of the genus *Stegodyphus*, with special reference to the African species *S. dumicola* and *S. mimosarum* (Araneida, Eresidae). Verhandlungen des Naturwissenschaftlichen Vereins zu Hamburg. 30:255-303.

Selden, P. A. (1989). Orb-web weaving spiders in the early Cretaceous. Nature. 340:711-713.

Selden, P. A. and Gall, J. C. (1992). A Triassic mygalomorph spider from the northern Vosges, France. Palaeontology. 35:257-285.

Selden, P. A., Shear, W. A. and Sutton, M. D. (2008). Fossil evidence for the origin of spider spinnerets and a proposed arachnid order. Proceedings of the National Academy of Sciences. 105(52):20781-20785.

Sensenig, A. T. and Shultz. J. W. (2003). Mechanics of cuticular elastic energy storage in leg joints lacking extensor muscles in arachnids. Journal of Experimental Biology. 206:771-784.

Shao, Z. and Vollrath, F. (1999). The effect of solvents on the contraction and mechanical properties of spider silk. Polymer. 40:1799-1806.

Shao, Z., Wen Hu, X., Frische, S. and Vollrath, F. (1999a). Heterogeneous morphology in *Nephila edulis* spider silk and its significance for mechanical properties. Polymers. 40:4709-4711.

Shao, Z., Young, R. J. and Vollrath, F. (1999b). The effect of solvents on spider silk studied by mechanical testing and single-fibre Raman spectroscopy. International Journal of Biological Macromolecules. 24:295-300.

Shear, W. A., Bonamo, P. M., Grierson, J. D., Rolfe, W. D. I., Smith, E. L. and Norton, R. A. (1984). Early land animals in North America: evidence from Devonian age arthropods from Gilboa, New York. Science. 224:492-494.

Shear, W. A., Palmer, J. M., Coddington, J. A. and Bonamo, P. M. (1989). A Devonian spinneret: early evidence of spiders and silk use. Science. 246:479-481.

Shear, W. A., Selden, P. A., Rolfe, W. D. I., Bonamo, P. M. and Grierson, J. D. (1987). New terrestrial arachnids from the Devonian of Gilboa, New York (Arachnida: Trigonotarbida). American Museum Novitates. 2901:1-74.

Simon, E. (1892-1903). Histoire naturelle des araignees. Paris (Roret). 2(3):381-668.

Sirichaisit, J., Young, R. J. and Vollrath, F. (1999). Molecular deformation in spider dragline silk subjected to stress. Polymers. 41:1223-1227.

Smith, A. M. (1993). Taxonomy focus. *Iridopelma* Journal of the British Tarantula Society. 9(1):13-18.

Sponner, A., Schlott, B.,Vollrath, F., Unger, E., Grosse, F. and Weisshart, K. (2005a). Characterization of the protein components of *Nephila clavipes* dragline silk. Biochemistry. 44:4727-4736.

Sponner, A., Unger, E., Grosse, F. and Weisshart, K. (2004). Conserved C-termini of spidroins are secreted by the major ampullate glands and retained in the silk thread. Biomacromolecules. 5:840-845.

Sponner, A., Vater, W., Rommerskirch, W., Vollrath, F., Unger, E., Grosse, F. and Weisshart, K. (2005b). The conserved C-termini contribute to the properties of spider silk fibroins. Biochemical and Biophysical Research Communications. 338:897-902.

Stoltz, J. A., Elias, D. O. and Andrade, M. C. B. (2009). Male courtship effort determines female response to competing rivals in redback spiders. Animal Behaviour. 77:79-85.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques. 28:1102-1104.

Stradling, D. J. (1978). The growth and maturation of the ''tarantula'', *Avicularia avicularia* L. Zoological Journal of the Linnean Society. 62:291-303.

Sugimoto, N., Nakano, S., Yoneyama, M. and Honda, K. (1996). Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. Nucleic Acids Research. 24:4501-4505.

Sutcliffe, D. W. (1963). The chemical composition of haemolymph in insects and some other arthropods in relation to their phylogeny. Comparative Physiology and Biochemistry. 9:121-123.

Sutherland, T. D., Weisman, S., Trueman, H. E., Sriskantha, A., Trueman, J. W. H. and Haritos, V. S. (2007). Conservation of essential design features in coiled coil silks. Molecular Biology and Evolution. 24:2424-2432.

Sutherland, T. D., Young, J. H., Weisman, S., Hayashi, C. Y. and Merritt, D. J. (2010). Insect silk: one name, many materials. Annual Review of Entomology. 55:171-188.

Tai, P. L., Hwang, G. Y. and Tso I. M. (2004). Inter-specific sequence conservation and intra-individual sequence variation in a spider silk gene. International Journal of Biological Macromolecules. 34:295-301.

Termonia, Y. (1994). Molecular modeling of spider silk elasticity. Macromolecules. 27:7378-7381.

Teshima, G. (2000). Deamidation in proteins and peptides. Proteomics, Proteome Digest [Internet].

Teulé, F., Cooper, A. R., Furin, W. A., Bittencourt, D., Rech, E. L., Brooks, A. and Lewis, R. V. (2009). A protocol for the production of recombinant spider silk-like proteins for artificial fiber spinning. Nature Protocols. 4(3):341-55.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research. 22:4673-4680.

Thompson, W. R., Meinwald, J., Aneshansley, D. and Eisner, T. (1972). Flavonols: pigments responsible for ultraviolet absorption in nectar guide of flower. Science. 177:528-530.

Tian, M. and Lewis, R. V. (2005). Molecular characterization and evolutionary study of spider tubuliform (egg case) silk protein. Biochemistry. 44:8006-8012.

Tian, M., Liu, T. C. and Lewis, R. V. (2004). Analysis of major ampullate silk cDNAs from two non-orb-weaving spiders. Biomacromolecules. 5:657-660.

Tinter, A. (1991). Eine neue Vogelspinne aus Venezuela *Pseudotheraphosa apophysis* n. gen. n. sp. (Araneae: Theraphosidae: Theraphosinae). Arachnologischer Anzeiger. 16:6-10.

Tso, I. M. (1998). Stabilimentum-decorated webs spun by *Cyclosa conica* (Araneae, Araneidae) trapped more insects than undecorated webs. Journal of Arachnology. 26:101-105.

Tso, I. M., Jiang, S. Y. and Blackledge, T. A. (2007). Does the giant wood spider *Nephila pilipes* respond to prey variation by altering web or silk properties? Ethology. 113:324-333.

Tso, I. M., Wu, H. C. and Hwang, I. R. (2005). Giant wood spider *Nephila pilipes* alters silk protein in response to prey variation. The Journal of Experimental Biology. 208(6):1053-1061.

Tsukada, M., Nagura, M. and Ishikawa, H. (1985). Relationship between the fine structure of silk sericin and its thermal decomposition. Journal of Sericultural Science of Japan. 54(1):43-47.

Uetz, G. W., McClintock, W. J., Miller, D., Smith, E. I. and Cook, K. K. (1996). Limb regeneration and subsequent asymmetry in a male secondary sexual character influences sexual selection in wolf spiders. Behavioral Ecology and Sociobiology. 38:321-326.

Uhl, G. and Gunnarsson, B. (2001). Female genitalia in *Pityohyphantes phrygianus*, a spider with a skewed sex ratio. Journal of Zoology, London. 255:367-376.

Unger, R. E., Wolf, M., Peters, K., Motta, A., Migliaresi, C. and Kirkpatrick, C. J. (2004). Growth of human cells on a non-woven silk fibroin net: a potential for use in tissue engineering. Biomaterials. 25:1069-1075.

Van Beek, J. D., Kummerlen, J., Vollrath, F. and Meier, B. H. (1999). Supercontracted spider dragline silk: a solid-state NMR study of the local structure. International Journal of Biological Macromolecules. 24:173-178.

Vehoff, T., Glisovic, A. and Salditt, T. (2007). Mechanical properties of spider dragline silk: humidity, hysteresis and relaxation. Biophysical Journal. 93:4425-4432.

Vollrath, F. and Edmonds, D. (1989). Modulation of the mechanical properties of spider silk by coating with water. Nature. 340:305-307.

Vollrath, F. and Knight, D. P. (1999). Structure and function of the silk production pathway in the spider *Nephila edulis*. International Journal of Biological Macromolecules. 24:243-249.

Vollrath, F. and Knight, D. P. (2001). Liquid crystalline spinning of spider silk. Nature. 410 (6828):541-548.

Vollrath, F. and Selden, P. (2007). The role of behavior in the evolution of spiders, silks and webs. Annual Review of Ecology, Evolution and Systematics. 38:819-46.

Vollrath, F. and Tillinghast, E. K. (1991). Glycoprotein glue beneath a spider web's aqueous coat. Naturwissenschaften. 78:557-559.

Wagner, R. H. and E. Danchin. (2003). Conspecific copying: a general mechanism of social aggregation. Animal Behaviour. 65:405-408.

Walter, A., Elgar, M. A., Bliss, P. and Moritz, R. F. A. (2008). 'Wrap attack` activates web decorating behavior in *Argiope* spiders. Behavioral Ecology. 19(4):799-804.

Ward, D. and Lubin, Y. (1993). Habitat selection and the life history of a desert spider, *Stegodyphus lineatus* (Eresidae). Journal of Animal Ecology. 62:353-363.

Ward, P. (1985). Prey availability increases less quickly than nest size in the social spider, *Stegodyphus mimosarum*. Behaviour. 97:213-225.

Watanabe, T. (1999). Prey attraction as a possible function of the silk decoration of the uloborid spider *Octonoba sybotides*. Behavioral Ecology. 10:607-611.

Weisman, S., Haritos, V. S., Church, J. S., Huson, M. G., Mudie, S. T., Rodgers, A. J., Dumsday, G. J. and Sutherland, T. D. (2010). Honeybee silk: recombinant protein production, assembly and fiber spinning. Biomaterials. 31(9):2695-2700.

Wickler, W. and Seibt, U. (1986). Aerial dispersal by ballooning in adult *Stegodyphus mimosarum*. Naturwissenschaften. 73:628-629.

Wieslander, L. (1994). The Balbiani ring multigene family: coding repetitive sequences and evolution of a tissue-specific cell function. Progress in Nucleic Acid Research and Molecular Biology. 48:275-313.

Williams, D. F. (2006). To engineer is to create; the link between engineering and regeneration. Trends in Biotechnology. 24:4-8.

Williams, D.F. (2009). On the nature of biomaterials. Biomaterials. 30:5897-5909.

Wilson, M. P. W. (1974). Shrinkage and chain folding in drawn poly(ethylene terephthalate) fibers. Polymer. 15:277-282.

Woolley, C., Thomas, C. F. G., Hutchings, L., Goodacre, S., Hewitt, G. M. and Brooks, S. P. (2007). A novel trap to capture ballooning spiders. Journal of Arachnology. 35(2).

Work, R. W. (1976). The force-elongation behavior of web fibers and silks forcibly obtained from orb-web-spinning spiders. Textile Research Journal. 46:485-492.

Work, R. W. (1977a). Dimensions, birefringences and force-elongation behavior of major and minor ampullate silk fibers from orb-web-spinning spiders — The effects of wetting on these properties. Research Journal. 47:650-662.

Work, R. W. (1977b). Mechanisms of major ampullate silk fiber formation by orb-web-spinning spiders. Transactions of the American Microscopical Society. 96:170-189.

Work, R. W. (1978). Mechanisms for the deceleration and support of spiders on draglines. Transactions of the American Microscopical Society. 97:180-191.

Work, R. W. (1981). A comparative study of the supercontraction of major ampullate silk fibers of orb-web-building spiders (Araneae). Journal of Arachnology. 9:299-308.

Work, R. W. (1984). Duality in major ampullate silk and precursive material from orb-web building spiders (Araneae). The American Society for Microbiology. 103:113-121.

Work, R. W. and Emerson, P. D. (1982). An apparatus and technique for the forcible silking of spiders. Journal of Arachnology. 10:1-10.

Xia, X. X., Qian, Z. G., Ki, C. S., Park, Y. H., Kaplan, D. L. and Lee, S. Y. (2010). Native-sized recombinant spider silk protein produced in metabolically engineered *Escherichia coli* results in a strong fiber. Proceedings of the National Academy of Sciences. 107(32):14059-14063.

Xu, M. and Lewis, R. V. (1990). Structure of a protein superfiber - spider dragline silk. Proceedings of the National Academy of Sciences. 87:7120-7124.

Yamada, H., Shigesada, K., Igarashi, Y., Takasu, Y., Tsubouchi, K. and Kato, Y. (2004). A novel asparagine-rich fibrous protein (Xenofibron) from the cocoons

of the parasitic wasp *Cotesia* (= *Apanteles*) *glomerata*. International Journal of Wild Silkmoths and Silk. 9:61-66.

Yamao, M., Katayama, N., Nakazawa, H., Yamakawa, M., Hayashi, Y., Hara, S., Kamei, K. and Mori, H. (1999).Gene targeting in the silkworm by use of a baculovirus. Genes and Development. 13:511-516.

Yang, Z., Liivak, O., Seidel, A., LaVerde, G., Zax, D. and Jelinski, L. W. (2000). Supercontraction and backbone dynamics in spider silk: 13C and 2H NMR studies. Journal of the American Chemical Society. 122:9019-9025.

Yonemura, N., Sehnal, F., Mita, K. and Tamura, T. (2006). Protein composition of silk filaments spun under water by caddisfly larvae. Biomacromolecules. 7:3370-3378.

Zhang, X., Baughman, C. B. and Kaplan, D. L. (2008). *In vitro* evaluation of electrospun silk fibroin scaffolds for vascular cell growth. Biomaterials. 29:2217-2227.

Zhao, A., Zhao, T., Sima, Y., Zhang, Y., Nakagaki, K., Miao, Y., Shiomi, K., Kajiura, Z., Nagata, Y. and Nakagaki, M. (2005). Unique molecular architecture of egg case silk protein in a spider, *Nephila clavata*. Journal of Biochemistry. 138:593-604.

Zhou, C. Z., Confalonieri, F., Jacquet, M., Perasso, R., Li, Z. G. and Janin, J. (2001). Silk fibroin: structural implications of a remarkable amino acid sequence. Proteins. 44:119-22.

Zhou, C. Z., Confalonieri, F., Medina, N., Zivanovic, Y., Esnault, C., Yang, T., Jacquet, M., Janin, J., Perasso, R. and Li, Z. G. (2000). Fine organization of *Bombyx mori* fibroin heavy chain gene (2000). Nucleic Acid Research. 28:2413-2419.

Zou, B., Dai, L. and Wu, Z. (2006). Analysis of structure and properties of biodegradable regenerated silk fibroin fibers. Journal of Materials Science. 41:3357-3361.

Zschokke, S., Hénaut Y., Benjamin, S. P. and García-Ballinas, J. A.. (2006). Prey-capture strategies in sympatric web-building spiders. Canadian Journal of Zoology. 84:964-973.

Zschokke, S. (1996). Early stages of web construction in *Araneus diadematus* Clerck. Revue Suisse de Zoologie Hors. 2:709-720.

Zschokke, S. (2002). Ultraviolet reflectance of spiders and their webs. Journal of Arachnology. 30:246-254.

Zurovec, M., Yang, C., Kodrik, D. and Sehnal, F. (1998). Identification of a novel type silk protein and regulation of its expression. Journal of Biological Chemistry. 273:15423-15428.

9 APPENDICES

Appendix 1 Example of a typical tarantula caresheet

***Pterinochilus murinus*/Mombasa golden starburst**
Chosen due to its high availability, fast growth, short gestation period and copious webbing tendencies. This species can be used as a representative of the tarantulas in terms of safety procedures as it is highly aggressive and so safety protocols related to it are applicable to the other available tarantulas.

Feeding: Live crickets, locusts, cockroaches, mealworms, maggots or fruit flies (smaller instars). Food detritus is removed with 30 cm tongs to maximise distance between specimen and keeper.
No physical contact is to be made between specimen and keeper.
Specimen is transferred using plastic boxes placed over the specimen and a plastic/cardboard sheet underneath to contain it.

Reported bite reactions:
Reaction(s): Tarantula bites are reported to be often no worse than a bee sting but more sensitive individuals may experience the following: immediate and intense burning pain lasting about 16 hours before gradually subsiding, localised moderate swelling and milder swelling in surrounding areas.
Swelling lasts approximately 1 week. The bitten region can feel sore and arthritic for weeks with painful and persistent muscle cramping (mostly in legs, chest and back) and can last for about 3 days.
The symptoms can apparently be relieved by antihistamines.

Common name: Starburst Baboon.

Range: Scrubland areas of Kenya, Tanzania and Zambia

Size: Medium sized species reaching up to 120 mm legspan.

Habitat: Terrestrial burrowing species that lay down copious amounts of silk in captivity**.**

Temp/humidity: 70 - 80°F / 60 - 70%

Housing: A typical terrestrial set-up will suffice if providing a cork bark retreat and slightly moist substrate. *P. murinus* is an opportunistic burrower and will sometimes fill the entire container with thick layers of silk.

Comments: Captive bred stock is desirable as this species adapts well to captivity and is easy to breed. Males possess tibial hooks and mating is straight forward given a receptive female. After mating the female should be fed as much as she will take and after approximately eight to ten weeks the egg sac is produced. A typical *P. murinus* egg sac is suspended in a hammock-like web surrounded by thick layers of silk. The female seals herself inside this web and should not be disturbed during the incubation period of around six to eight weeks. The spiderlings emerge fully mobile and number around 150. They can be left with the female for several weeks but should then be separated to prevent cannibalism. *P. murinus* is attractively marked with a golden starburst pattern on the carapace and a symmetrically spotted and striped abdomen. The overall colour is russet brown. This species is particularly defensive and will readily bite. Not recommended for the beginner but a hardy species, living approximately eight to ten years in captivity. A second egg sac is sometimes produced from a single pairing but this is usually smaller than the first and contains less young. The tank can be allowed to dry out occasionally but it is recommended that humidity is increased during stressful times such as moulting and egg sac production. Spiderlings grow rapidly and can reach maturity in under two years. There are several colour forms of *P. murinus* available and this includes the red colour form (RCF or Usambara). These variants are regional colour forms and care should be taken not to hybridise them.

Collated from: http://giantspiders.com/Pterinochilus_species.html amongst other reputable arachnology websites.

Appendix 2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) gel compositions

Resolving Gel (100 ml)

Gel %	4	6	8	10	12	14	16	18	20	22
ProtoGel 30% ml	13	20	26	33	40	46	53	59	66	73
1.5 M Tris-HCl, (pH 6.8) ml	25	25	25	25	25	25	25	25	25	25
10% SDS ml	1	1	1	1	1	1	1	1	1	1
Deionised H_2O ml	61	54	48	41	34	28	21	15	8	1
10% Ammonium Persulphate μl	50	50	50	50	50	50	50	50	50	50
TEMED μl	10	10	10	10	10	10	10	10	10	10

Stacking Gel (100ml)

Gel%	2.5	3	3.5	4
ProtoGel 30% ml	8.3	10.0	11.7	13.3
0.25 M Tris-HCl, (pH 6.8) ml	50	50	50	50
10% SDS ml	1	1	1	1
Deionized H_2O ml	40.7	39.0	37.3	35.7
10% Ammonium Persulphate μl	50	50	50	50
TEMED μl	10	10	10	10

Appendix 3 Q-ToF2 residue assignment complications

A complication that arises with the Q-ToF2 with particular peptides is that certain species have the same molecular weight as others. In the example of silks, glycine repeats (poly-G tracts) pose a problem but in a different way that repetitive elements would in DNA sequencing. Here, two adjacent G residues have the same molecular weight as that of an asparagine (N). In places where there are two successive glycines (or indeed, sometimes with a glycine-alanine couplet), there is often poor cleavage between them and so an assignment of N is given rather than the correct G-G.Likewise, G and A have the same molecular weight as glutamine (Q) and to complicate matters even further, Q is isobaric with lysine (K).

For example, one of the sequenced peptides could read:
QNGGGGDFGQSGR

F = oxidised methionine.

Underscoring in the *de novo* sequence records means that any order of the residues underlined is possible.

This means the various versions we could have of this peptide include:-
qnGGGGDFGQgsR
nqGGGGDFGQsgR
qnGGGGDFGQgsR
qnGGGGDFGQsgR
QNnGGDFGQSGR
NQGnGDFGQSGR
QNGGnDFGQGSR
QNnGGDFGgaSGR
NQGnGDFGagSGR
qnGGGGDmGQgsR
nqGGGGDmGQsgR
qnGGGGDmGQgsR
qnGGGGDmGQsgR
etc
etc
.....
In silks where a poly-G tract is expected, particular sequences would be chosen over others preferentially, narrowing down the search as it were.

Appendix 4 Secondary structure predictions of N-terminal domains

Euprosthenops australis secondary structure prediction

```
                10        20        30        40        50        60
        ....:....x....:....x....:....x....:....x....:....x....:....x
        MSWTARLALLLLFVACQGSSSLASHTTPWTNPGLAENFMNSFMQGLSSMPGFTASQLDDM
        Euprosthenops australis
        CCCCCHHHHHHHHCCCCCCCCCEEEECCCCCCCHHHHHCCCCCCCCCCCCCHHHHHHH        BPS
        CCEEHHHHHHEEEEHECCCCCHHCCCCCCCCCCCHHHHHCEEHHCECCCCCCCHHHHCCH      D_R
        CCHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHCCCCCCCHHHHHHH       DSC
        CCCCCHHHHHHHHHCCCCCCCEECCCCCCCCCCCHHHHHHHHHHCCCCCCCCCCCCCCH      GGR
        HHHHHHHHHHHEEEHCCCCCEEEEECCCECCCCCCHCHHHHHHCCCCCCCCCECCCCCCC     GOR
        CCCHHHHHHHEEHHCCCCCCCEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC     H_K
        CCCCCCCEEEEECCCCCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCCCCCCCCCCCCCC     K_S
        CCCCHHHHHHHEECHCCCCCCCCECCCCCCCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCH    JOI

                70        80        90        100       110       120
        ....:....x....:....x....:....x....:....x....:....x....:....x
        STIAQSMVQSIQSLAAQGRTSPNKLQALNMAFASSMAEIAASEEGGGSLSTKTSSIASAM
        Euprosthenops australis
        HHHHHCCHCCCHCCCCCCCCCHHHHHHHHHHCCCCHHHHCCCCCCCCCCEEEEEHHHHH      BPS
        CEEHHEEEEEEHHHHHCCCCCCHHHHHHHHHHHHHHHHCCCCCCCCCCECEEHHHH         D_R
        HHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHCCCCCCCCCHHHHHHH         DSC
        HHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCHHHHH        GGR
        CEEHHCEEEECEEEECCCCCCCCCHHHHHHHCCHCCEECCCCCCCEEEEEEEECEEE        GOR
        CCCCCCCCEECEECCCCCCCCCCCHHHHHHHHHHCHHHHCCCCCCCCCCCCCCCEECCCCC    H_K
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCH    K_S
        CCCHHCCCCCCCHCCCCCCCCCHHHHHHHHHHCHHHHHHCCCCCCCCCCCCCECHHHH       JOI

                130       140       150       160       170       180
        ....:....x....:....x....:....x....:....x....:....x....:....x
        SNAFLQTTGVVNQPFINEITQLVSMFAQAGMNDVSASASAGASAAASAGAPGYSPAPSYS
        Euprosthenops australis
        HHHHHEEEEEEEEEECHCEHHCCHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC      BPS
        CHHHHHEEEEECECEEEHEEEEEEEHHHHCHCCECHHHCCCHHHHHHHCCCCCCCCCCCC     D_R
        HHHHHCCCCCCCHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHCCCCCCCCCCCC          DSC
        HHHHHCCCCCCCCCCCHHHHHHHHHHCCCCHHHHHHHHHHHHHHCCCCCCCCCCCC         GGR
        ECCCEEECEEECCCCCCHHHHHHHHHHHCCCCCCECCCCCCCEEEECCCCCCCCCCCEEE     GOR
        CCCCEECCCEECCCCCCHHHHCHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEE    H_K
        HHHHHHCCHHHCCCCHCCHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC      K_S
        HHHHHHCCCEECCCCCCHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC      JOI

                190       200       210       220       230       240
        ....:....x....:....x....:....x....:....x....:....x....:....x
        SGGYASSAASAAAAAGQGGPGGYGPAPNQGASSAAAAAAGSGQGPSGPYGTSYQISTQYT
        Euprosthenops australis
        CCCCHCHHHHHHCCCCCCCCCCCCCCCCCCCCHHHHCCCCCCCCCCCCCCCCCEEEEEEEEEE  BPS
        CCCCCCCHHHHHHHHCCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCCCCCCCEEEEEEEE    D_R
        CCCCHHHHHHHHHHHHCCCCCCCCCCCCCCCCHHHHHHHHHHCCCCCCCCCCCCEEEEEEE    DSC
        CCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCHHHHHHHHHHCCCCCCCCCCCCCEEEEEECEE  GGR
        CCCEEEEECCCEEEECECCCCCCCCCCCCCCCCCEEEEECCCCCCCCCCEEEEEEEEEEEE    GOR
        CCCCEEHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEEEEEEEE     H_K
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    K_S
        CCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCHHCCCCCCCCCCCCCCCCCCCEEEEEEEEE  JOI

                250       260       270       280       290       300
        ....:....x....:....x....:....x....:....x....:....x....:....x
        QTTTSQGQGYGSSSAGAAAAGAAGAGQGGYGGQGQGGYGQGAGGAAAAAAAA
        Euprosthenops australis
        EEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHCCCC           BPS
        EEECCCCCCCCCCCCCCHHHHCHCCCCCCCCCCCCCCCCCCCCCCCHHHHHCC            D_R
        CCEEECCCCCCHHHHHHHHHHHHHHCCCCCCCCCCHHHHHHHHHHHHH                 DSC
        CCEEECCCCCCCHHHHHHHHHCCCCCCCCCCCCCCHHHHHHHHHH                    GGR
        EEEEEEEEEEEEECCCEEEEEEEEEEEEEEEEEEEEEECCCCCCCCHCH                GOR
        EEEEEEEEEECCCCCHHHHCCCCCCCCCEEEECEEEEEECCCHHHHHHHHC             H_K
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC              K_S
        EEEEECCCCCCCCCCHHHHCCCCCCCCCCCCCCCCCCCCCCCCHHHHHCC              JOI
```

284

Latrodectus hesperus secondary structure prediction

```
                10        20        30        40        50        60
        ....:....x....:....x....:....x....:....x....:....x....:....x
        MTWSTRLALSFLFVLCTQSLYALAQANTPWSSKANADAFINSFISAASNTGSFSQDQMED
        Latrodectus hesperus
        CCCCCHHHHHHEEEEEHCHHHHHHHCCCCCHHHHHHEEEEECCCCCCCCCCCCCCCCC    BPS
        CCEEEHHHHEEEEEEEEEEHHHHHCCCCCCHCHHHHEEEEEEHHCCCCCCCHHHHHH    D_R
        CCCHHHHHHHEEEEHHHHHHHHHCCCCCHHCCCHHHHHHHHHCCCCCCCCCHHHHHH    DSC
        CCCCCCHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHCCCCCCCCCCHHHHH     GGR
        EEHHHHHHHHEEEHHHHHHCEECCCCCCCCCCCCCHHEEEEEECCCCCCCCCCHHHH    GOR
        CCCCHHHHHCHEEHHCCCCCCCCCCCCCCCCHHHHHCEEECCCCCCCCCCCCCCCCC    H_K
        CHHHHHHHHHHHHHHCHHHHHHCCCCCCCCCCHHHHHHHHHHHCCCCCCCCCCCCCC    K_S
        CCCCHHHHHHCHEEHHHHHHHHHHCCCCCCCCCCCCHHHHHHCEECCCCCCCCCCCCCHHHH JOI

                70        80        90       100       110       120
        ....:....x....:....x....:....x....:....x....:....x....:....x
        MSLIGNTLMAAMDNMGGRITPSKLQALDMAFASSVAEIAASEGGDLGVTTNAIADALTSA
        Latrodectus hesperus
        CCHEEHCHHHHHHCCCCCCCCCHCCCCCCCCHHCCCCCCCCCCCCCCCCCCCHHHHHHHH  BPS
        HEEECCEHHHHHHCCCCCECCCCHHHHHHHHHHEEHHHHHCCCCCCEEEEHHHHHHHEH   D_R
        HHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHCCCCCCCHHHHHHHHHHHH         DSC
        HHHHHHHHHHHHHCCCCCCCHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHH     GGR
        HHHHCCHHHCHCCCCCCCCCCCCHHHHHHHHHCCCCEEECCCCCCCEEEEECCCCCEECC  GOR
        CHHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHCHHHHHHCCCCCCCCCCCCCCCCCCCCC  H_K
        CHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCCHHCCHHHHHHHHH   K_S
        HHHHHHHHHHHHHCCCCCCCCCCHHHHHHHHHHHHHHHHCCCCCCCCCCHHHHHHHHH    JOI

               130       140       150       160       170       180
        ....:....x....:....x....:....x....:....x....:....x....:....x
        FYQTTGVVNSRFISEIRSLIGMFAQASANDVYASAGSSGGGGYGASSASAAASASAAAPSG
        Latrodectus hesperus
        EEEEEEEEEEEEECCCCCCCCCCHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC   BPS
        EEEEEEEEEEEEEHEEEEEEHHHHHHHCCEEHCCCCCCCCCCCCCHCHHHHHHHHCCC    D_R
        HCCCCCEEECHHHHHHHHHHHHHHHHHHEEEECCCCCCCCCCCHHHHHHHHHCCHHH     DSC
        HHCCCCCCCCCHHHHHHHHHHHHHHHHHEEECCCCCCCCCCCCHHHHHHHHHHCCCC     GGR
        CEEEEEEEEEEEEEHEEHHEEEECCCCCCEEEEECECCCEEEEEEECCCECCCCCCCC    GOR
        CEEECCEEECCCHHHHHHHHHHHHCCCCCEEEECCCCCEEEEEHHHHHHCCCCCCC      H_K
        HHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCHHHHCCCCCCCCCCCCCCCC         K_S
        CEEECCEEECCCHHHHHHHHHHHHHHHCEEECCCCCCCCCCCCCCCHCHHHHCCCCCCC   JOI

               190       200       210       220       230       240
        ....:....x....:....x....:....x....:....x....:....x....:....x
        VAYQAPAQAQISFTLRGQQPVSYGQGGASAASGAEAGQGGAGPGGAGAAAAAAAAAGGAG
        Latrodectus hesperus
        CCCHHHHHHHHHCECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHCCCCCCC BPS
        EHCHHCHHHHHEEEEEHCCCCECCCCCCCCCHHCCHHCCCCCCCCCCCCHHHHHHHHHCCCC D_R
        HHHCCCCCEEEEEECCCCCCCCCCCCCHHHHHHHHHCCCCCHHHHHHHHHHHHHH        DSC
        HHHCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHCCCCCCHHHHHHHHHHHHHHCCC     GGR
        CEECCCCCHEEEEEECCCCECEEEECCEEEEECEEEEECCCCCCCCCCCCCCCCCCCCCC   GOR
        CCCCCCCCCCEECCEEECCCCEEECCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHCCCCC   H_K
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC   K_S
        CCCCCCCHCECCEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCCC     JOI

               250       260       270       280       290       300
        ....:....x....:....x....:....x....:....x....:....x....:....x
        QGGQG
        Latrodectus hesperus
        CCCCC                                                         BPS
        CCCCC                                                         D_R
        CCCCC                                                         DSC
        CEEEC                                                         GGR
        CCCCE                                                         GOR
        CCCCC                                                         H_K
        CCCCC                                                         K_S
        CCCCC                                                         JOI
```

Agelenopsis aperta secondary structure prediction

```
                10        20        30        40        50        60
        ....:....x....:....x....:....x....:....x....:....x....:....x
        MTWTVRLAIPILILILQGSKCLGQSNPWTDTATAESFISSVMSSVANQGCLSYDQIDDMQ
        Agelenopsis aperta
        CCCCCEECCEECCCCCCCCCCCCCCCCCCCCCHCCCCCCCCCCCCCCCCCECCCCHHHHCCC    BPS
        CCEEEEEEEEEEEEEEECCCCCCCCCCCCCCHCHHEEEEEEEEEECCCCCCCCCCECCHH     D_R
        CCEEECCCCEEEEECCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCCCCCCCCCHHHHH      DSC
        CCCEEECCCHHHHHHCCCCEEECCCCCCCCCCHHHHHHHHHCCCCCCCCCCCCCCCHH       GGR
        HEEHHHHHCHHHHEHHCCCCECCCCCCCCCEEECCEEEEEEEEEECCCCCCCCCCCCCCHH    GOR
        CCCHHHHHCCCCEEEECCCCCCCCCCCCCCCCCCEECEECCCCCCCCCCCCCCCCCHHHHH    H_K
        CCCCCCCCCCEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    K_S
        CCCCCCCCCCECEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHH    JOI


                70        80        90       100       110       120
        ....:....x....:....x....:....x....:....x....:....x....:....x
        AVGDTMLATMDNLVRSGKSSSHMLKAMNMAMGTSIAEIVADGGGNLGSKVSCISNALSSA
        Agelenopsis aperta
        CEHHCHHHHHHHHCCCCCCCCCHHHHHHHHHEHHECCCCCCCCCCCCCCEEEEEEEHHHHH    BPS
        HECCEEHHCHCCEEECCCCCCCHHHHHHHHHCCCEEHEEHCCCCCCCCCEEEECCHHCCH     D_R
        HHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHEECCCCCCCCCEEEEEHHHHHHHH        DSC
        HHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHCHHHHEEEECCCCCCCCCEEEECCCCCHH     GGR
        HHCCHHHHHHCCEEECCCCCHHHHHCCHHCCCCEEEEECCCCCCEEEEEECCCCEEE       GOR
        HCCCHHHHCCCCCCCCCCHHHHHHHHHHHCCCCCEEEECCCCCCCEEEEECCCCCCC       H_K
        CCCCCHHCHHHHHHHHHCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHH     K_S
        HCCCHHHHHHHHHCCCCCCHHHHHHHHHHHCCCCCEECCCCCCCCCCEEEEECCCHCHH     JOI


               130       140       150       160       170       180
        ....:....x....:....x....:....x....:....x....:....x....:....x
        FLQTTGSVNTQFVNEIVSLISMFAQADTNEVGVGSGSGAGAGSGAGAGARYSASAVFSTG
        Agelenopsis aperta
        ECCEEEEEEECCEECCCCHHHHHECCCCCCCCCCCCCCCCCCCHCCHHECCECCC         BPS
        EEECCCCCECEEEEEEEEEEEEEEHHHHCCCCECECCCCCCCCCCCCCCCCHCCHEHEECC   D_R
        HEECCCCCCCCHHHHHHHHHHHHHCCCCEEEEECCCCCCCCCCCCCCCEEEEEEEEECC     DSC
        HHHCCCCCCCCEECHHHHHHHHHHHCCCCEEEEECCCCCCCCCCCCCCCHHHHCEEEEECC   GGR
        EEEEECCCCEEEEEEHHHHHHHCCCCCEEEEEEECCEEEEEEEEEEEEEEEEEEEEEE      GOR
        EEEECCCCCEEECCCCHHHHHHHCCCCEEEEECCCCCCCCCCCCCCCHHHHEEEEEEEEC    H_K
        HHHHHHHHCCCHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCCC K_S
        EEECCCCCCCCEECCHHHHHHHHHHCCCCEEEEECCCCCCCCCCCCCCHHCCCEEEEEECC   JOI


               190       200       210       220       230       240
        ....:....x....:....x....:....x....:....x....:....x....:....x
        SGASAGSGSGSGAGSGAGAGAGSGAGFGRTAVLRAGAGIGSGAGAGSGAGAGSGAGAGAG
        Agelenopsis aperta
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    BPS
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEHHCCCCCCCCCCCCCCCCCCCCCCCCC    D_R
        CCCEEECCCCCCCCCCCCCCCCCCCCHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCEEEEC   DSC
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCC    GGR
        CEEEEEEEEEEEEEEEEEEEEEECEEEEEEEEEEECEEEEEEEEEEEEEEEEEEEEEEEE    GOR
        CCCEEECCCCCCEEECCCCCCCCCCCCEEECCCCCCCCCCCCCCCCCCCCCCCCCCCEEE    H_K
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEECCCCCCCCCCCCCCCCCCCCCCCCCCCC    K_S
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    JOI


               250       260       270       280       290       300
        ....:....x....:....x....:....x....:....x....:....x....:....x
        SGAGSGSGAGAGRGTGLGGLAAGLGAGVGTGAG
        Agelenopsis aperta
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    BPS
        CCCCCCCCCCCCCCCCCCCHCCCCCCCCCCCCC    D_R
        CCCCCCCCCCCCCCCHHHHHHEEECCCCCCCCC    DSC
        CCCCCCCCCCCCCCCCHHHHHCCCCEEEEEC     GGR
        EEEEEEEEEEEEEEEEEEEEEEEEEEEEEECC    GOR
        ECCCCCCCCCCCCCCCCCCEEECCCCCCCCCC    H_K
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    K_S
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    JOI
```

Argiope trifasciata secondary structure prediction

```
                10        20        30        40        50        60
        ....:....x....:....x....:....x....:....x....:....x....:....x
        MNWSIRLALLGFVVLSTQTVFSAGQGATPWENSQLAESFISRFLRFIGQSGAFSPNQLDD
        Argiope trifasciata
        CCCCHHHHCCEEEEEEEEEEEECCCCCCCCCHHHHHHHHHHCCCCCCCCCCCCCCCCHCCC    BPS
        CCEEEHHHHEEEEEEEEEEEEEECCCCCCCCCCHHHHHEEEEEEEEECCCCCCCCCCHCC    D_R
        CCCHHHHHHHEEEEECCCEEEECCCCCHHHHHHHHHHHHHHHHHCCCCCCCCCCHHH       DSC
        CCCCEEEEEECCCEECCCEEEECCCCCCHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCC     GGR
        HHHHHHHHEEEEEECEEEEEECCCCCCCHCCHHHHHHHHEEEEEEECCCCCCCCCCC       GOR
        CCCHHHHHHEEECCCEEEEECCCCCCCCCCCCCCCHHHHHHHEECCCCCCCCCCCCCC      H_K
        CCCCCCEEEEEEEECCCCCCCCCCCCCCCCCCCCHHHHHHHHHCCCCCCCCCCCCCCC      K_S
        CCCCHHHHHEEEEEECCEEEEECCCCCCCCHCHHHHHHHHHHHCCCCCCCCCCCCCCC      JOI

                70        80        90       100       110       120
        ....:....x....:....x....:....x....:....x....:....x....:....x
        MSSIGDTLKTAIEKMAQSRKSSKSKLQALNMAFASSMAEIAVAEQGGLSLEAKTNAIASA
        Argiope trifasciata
        CCCCCCCHHHHHHCCCCCCCCCHHHHHHHHHHHCCHHHHHHCCCCHHHHHHHHHHHHHH     BPS
        HCCECCCHHHHHHHHHHCCCCCCCCHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHH       D_R
        HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCHHHHHHHHHHHHH         DSC
        CCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCHHHHHHHHHHHH         GGR
        HCHHCHHHHCHHHHHHHHHHCHHHHHHHHHHHHHHCHHHHEHHHHCCCHHHHHCHHHHHH    GOR
        CCCCCCCCCCHHHHHHHHHHCHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHH       H_K
        CCCCCHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC     K_S
        CCCCCCHHHHHHHHHHHHCCHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHH       JOI

               130       140       150       160       170       180
        ....:....x....:....x....:....x....:....x....:....x....:....x
        LSAAFLETTGYVNQQFVNEIKTLIFMIAQASSNEISGSAAAAGGSSGGGGGGSGQGGYGQG
        Argiope trifasciata
        HHHHHHHEEEEEEEHCECCCEEHCCCHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    BPS
        HHHHHHHEECEECEEEEHHEEEEEEEEHHHCCCCECCCHHHHCCCCCCCCCCCCCCCC     D_R
        HHHHHHHCCCCCCHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC     DSC
        HHHHHHHCCCCCCCCCCHHHHHHHHHHHCCCHHHHHHHHHCCCCCCCCCCCCCCCCCC     GGR
        HCCCEEEECCCCHHEEHHHHHHHHHEEHHECCCCCEEEEEEEEEEEECEEEEEEEEEEEE    GOR
        HHHHHHCCCCCCCCCHHHHHHHHEEHHCCCCCCCCCEEEECCEECCCEEEEEEEEEEE     H_K
        HHHHHHHCHHHHHHHHCCCCCEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    K_S
        HHHHHHHCCCCCCCCCHHHHHHHEEHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC     JOI

               190       200       210       220       230       240
        ....:....x....:....x....:....x....:....x....:....x....:....x
        AYASASAAAAYGSAPQGTGGPASQGPSQQGPVSQPSYGPSATVAVTAVGGRPQGPSAPRQ
        Argiope trifasciata
        HCHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEECEEEEECCCCCCCCCCC   BPS
        CHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEECCCCCCCCCCCC       D_R
        HHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEECCCCCCCCCCCC      DSC
        HHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEECCCCCCCCCCCC      GGR
        EEEEECCCEEEECCCCCCCCCCCCCCCCCCEECECCCCCCEEEEEEEECCCCCCCCCCCCE   GOR
        CHHHHHHHHCCCCCCCCCCCCEEECCCCCEEEECCCCCEEEEEEEECCCCCCCCCCCC      H_K
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC     K_S
        CHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEECCCCCCCCCCCC     JOI

               250       260       270       280       290       300
        ....:....x....:....x....:....x....:....x....:....x....:....x
        QGPSQQGPGQQGPGGRGPYGPSAAAAAAAAGGYGPGAGQQGQQAGQGSGQQGPG
        Argiope trifasciata
        CCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCC        BPS
        CCCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCC       D_R
        CCCCCCCCCCCCCCCCCCCCHHHHHHHHCCCCCCCHHHHHCCCCCCCCCC          DSC
        CCCCCCCCCCCCCCCCCCCCHHHHHHHHHCCCCCCCCCHHHCCCCCEEEC          GGR
        CCCCCCCCCECCCCCCCCCCCCCCEECCCCCCCCCEEEEEEEECCCCCCCCCCC       GOR
        CCCCCCCEEECCCCCCCCCCCHHHHHCCCCCCCCCEEEEEECCCCCEECCC         H_K
        CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC          K_S
        CCCCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCCCCCCCCCCCCCCCCCC         JOI
```

287

Latrodectus geometricus secondary structure prediction

```
           10        20        30        40        50        60
....:....x....:....x....:....x....:....x....:....x....:....x
MTWSTRLALSVLLVLCTQSIYALAQANTPWSSKANADAFINSFISSAQNTGSFSQDQMDD
Latrodectus geometricus
CCCCCHHHHHECEEEEEEEEEHHHHHCCCCCHHHHHHEEEEECCCCCCCCCCCCCCCC    BPS
CCEEEHHHHHEEEEEEEEEEEHHHHHCCCCCCHCHHHHEEEEECHCCCCCCCHHHHCH    D_R
CCCHHHHHHEEEEECCCHHHHHHHCCCCCCHHCCCHHHHHHHHCCCCCCCCCHHHHHH    DSC
CCCCCCHHHHHHHCHHHHHHHCCCCCCCCCCCCHHHHHHHCCCCCCCCCCCCCHHHH    GGR
EEHHHHCHEEEEEHHCCHHEEEECCCCCCCCCCCHCEEEEEEECCCCCCCCCCCCHHH    GOR
CCCCHHHHHEEEEHHCCCCCCCCCCCCCCCCCCCCCEECCEEECCCCCCCCCCCCCCC    H_K
CCCCCCCCCCEEEEECCCCCCCCCCCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCCC    K_S
CCCCCHHHHCEEEECCCCCCHHCCCCCCCCCCCCCHHHECCEECCCCCCCCCCCCHCH    JOI
```

```
           70        80        90       100       110       120
....:....x....:....x....:....x....:....x....:....x....:....x
MSLIGNTLMTAMDNMGGRITPSKLQALDMAFASSVAEIAASEGGDLGVTTNAIADALTSA
Latrodectus geometricus
CCCEEEEHHHHCCCCCCCCCCCHCCCCCCCHHCCCCCCCCCCCCCCCCCCHHHHHHHH    BPS
HCEECCEEHHHHCCCCCCECCCCHHHHHHHHHHEEHHHHHCCCCCCEEEEHHHHHHHEH    D_R
HHHHHHHHHHHHHCCCCCCHHHHHHHHHHHHHHHHCCCCCEEEHHHHHHHH    DSC
HHHHHHHHHHHHHCCCCCCCHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHH    GGR
HHHHCCCCEECECCCCCCCCCCCHHHHHHHHHCCCCEEECCCCCCCEEEEECCCCEECC    GOR
CCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCHHHHHHCCCCCCCCCCCCCCCCCCCC    H_K
CCCCCCCEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHH    K_S
HCCCCCCCHHHHCCCCCCCCCCHHHHHHHHHHCCHHHHHCCCCCCCCCCCHHHHHHHHH    JOI
```

```
          130       140       150       160       170       180
....:....x....:....x....:....x....:....x....:....x....:....x
FYQTTGVVNNRFISEIRSLISMFAQASANDVYASAGSSGGGGYGAASSSASAAAPSGVTY
Latrodectus geometricus
EEEEEEEEEEEEEECCCCCCCCCCHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCECCC    BPS
EEEEEEEECEEEEEHEEEEEEHHHHHHHCCEEHCCCCCCCCCCCCCCCCHHHHHCCCEEE    D_R
HHCCCCCCCCHHHHHHHHHHHHHHHHCCHHEEEECCCCCCCCCCHHHHHHCCCCCCCC    DSC
HHCCCCCCCCCCHHHHHHHHHHHHHHHHHEEECCCCCCCCCCHHHHHHCCCCCEE    GGR
CEEEEEEEEECEEEEHHHEHEEHEHCCCCCEEEEECCCCCEEEEEEEEEEECCCCEEE    GOR
CEEECCCEECCCHHHHHHHHHHHHHHCCCCCEEEECCCCCEEEEECCCCCCCCCCCCEE    H_K
HHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    K_S
CEEECCCECCCCHHHHHHHHHHHHHHCCCCEEECCCCCCCCCCCCCCCCCCCCCCCCEE    JOI
```

```
          190       200       210       220       230       240
....:....x....:....x....:....x....:....x....:....x....:....x
QAPSQAQISFSMRGQQPNNYGQSGASAGSAAAGGAGQAGYGQRGQGQGAAAAAAASAAGG
Latrodectus geometricus
HHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCCC    BPS
CCCCHHHEEEEHHCCCCCCCCCCCCCCCHHHCCCCCCCCCCCCCCCHHHHHHHHHHCC    D_R
CCCCCCEEEEEECCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHH    DSC
CCCCCHHHHHHHCCCCCCCCCCCHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHHHCC    GGR
ECCCCCEEEEEEECCCCCCCEEEEEEEEEECEEEEEEEEEECCCCCCCCCCCCCCCCC    GOR
CCCCCCEECCEECCCCCCCCCCEECCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCC    H_K
CCCCCCCCCHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC    K_S
CCCCCCCECCECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCC    JOI
```

```
          250       260       270       280       290       300
....:....x....:....x....:....x....:....x....:....x....:....x
AGQGGQG
Latrodectus geometricus
CCCCCCC    BPS
CCCCCCC    D_R
HCCCCCC    DSC
CCCEEEC    GGR
CCECCCE    GOR
CCCCCCC    H_K
CCCCCCC    K_S
CCCCCCC    JOI
```

Secondary Structure prediction Legend and Citation

Legend:
>   H - α-helices
>   E - β-strands
>   C - Coil
>   T - β-turns

Citation (as per author's website)

Algorithm Citation:

BPS : Burgess, A. W., Ponnuswamy, P. K. and Sheraga, H. A. (1974). Analysis of conformations of amino acid residues and prediction of backbone topography in proteins. Israel Journal of Chemistry. 12:239-286.

D_R : Deleage, G. and Roux, B. (1987). An algorithm for secondary structure prediction based on class prediction. Protein Engineering. 1(4):289-294.

DSC : King, R. D. and Sternberg, M. J. E. (1996). Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. Protein Science. 5:2298-2310.

GGR : Garnier, Gibrat and Robson (1996). R.F. Doolittle ed. Methods in Enzymology. 266:97-120.

GOR : Garnier, J., Osguthorpe, D. J. and Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. Journal of Molecular Biology. 120:97-120.

G_G : Gascuel, O. and Golmard, J. L. (1998). A simple method for predicting the secondary structure of globular proteins: implications and accuracy. Computer Applications in the Biosciences. 4:357-365.

H_K : Holley, H. L. and Karplus, M. (1989). Protein secondary structure prediction with a neural network. Proceedings of the National Academy of Sciences. 86:152-156.

K_S : King, R. D. and Sternberg, M. J. E. (1990). A machine learning approach for the prediction of protein secondary structure. Journal of Molecular Biology. 216:441-457.

L_G : Levin, J. M. and Garnier, J. (1988). Improvements in a secondary structure prediction method based on a search for local sequence homologies and its use as a model building tool. Biochimica et Biophysica Acta. 955(3):283-295.

Q_S : Quinn, N. and Sejnowski, T. J. (1988). Predicting the secondary structure of globular proteins using neural network models. Journal of Molecular Biology. 202:865-884.

JOI Joint prediction - Prediction made by the program that assigns the structure using a "winner takes all" procedure for each amino acid prediction using the other methods.

Program Citation:
Questions or comments on this program may be mailed to: Georgios J. Pappas Jr.. (gpappas@cysteine.ncsa.uiuc.edu)

Appendix 5 Contigs formed from cDNA library sequencing alignments

Number of segment pairs = 7310; number of pairwise comparisons = 30
'+' means given segment; '-' means reverse complement

Overlaps/Containments/No. of Constraints Supporting Overlap

```
****************** Contig 1 ******************
03-
                  70+ is in 03-
                  02+ is in 03-
****************** Contig 2 ******************
09+
                  51+ is in 09+
****************** Contig 3 ******************
14+
49+
****************** Contig 4 ******************
17+
                  61+ is in 17+
58+
75+
64+
****************** Contig 5 ******************
33+
                  85+ is in 33+
****************** Contig 6 ******************
34+
                  86+ is in 34+
****************** Contig 7 ******************
37-
                  39+ is in 37-
****************** Contig 8 ******************
42+
                  47+ is in 42+
****************** Contig 9 ******************
52+
                  74+ is in 52+
****************** Contig 10 ******************
55+
                  65+ is in 55+
****************** Contig 11 ******************
56+
                  76+ is in 56+
****************** Contig 12 ******************
83-
                  66+ is in 83-
```

DETAILED DISPLAY OF CONTIGS

```
******************* Contig 1 ********************
                      .    :    .    :    .    :    .    :    .    :    .    :
03-             TAGAAACATCTGTTTGGATTCCATTGGAAAGTGTCGGGGCCCCTACGCTGAAGGAAAAGA
02+                                TGGAAAGTGTCGGGGCCCCTACGCTGAAGGAAAAGA
                _____
consensus       TAGAAACATCTGTTTGGATTCCATTGGAAAGTGTCGGGGCCCCTACGCTGAAGGAAAAGA


                      .    :    .    :    .    :    .    :    .    :    .    :
03-             AGACAGTCTAATGTTCTCACCAATGGCGGTCAAGAAACCAGTAATGCTGCAACCCTAATG
70+               CAGTCTAATGTTCTCACCAATGGCGGTCAAGAAACCAGTAATGCTGCAACCCTAATG
02+             AGACAGTCTAATGTTCTCACCAATGGCGGTCAAGAAACCANTAATGCTGCAACCCTAATG
                _____
consensus       AGACAGTCTAATGTTCTCACCAATGGCGGTCAAGAAACCAGTAATGCTGCAACCCTAATG


                      .    :    .    :    .    :    .    :    .    :    .    :
03-             CGGAGATAGTGAATGATGTATTAGATCTTGGGAGAGCGTTACAAAGCTGTAGTTGATGAC
70+             CGGAGATAGTGAATGATGTATTAGATCTTGGGAGAGCGTTACAAAGCTGTAGTTGATGAC
02+             CGGAGATAGTGAATGATGTATTAGATCTTGGGAGAGCGTTACAAAGCTGTAGTTGATGAC
                _____
consensus       CGGAGATAGTGAATGATGTATTAGATCTTGGGAGAGCGTTACAAAGCTGTAGTTGATGAC


                      .    :    .    :    .    :    .    :    .    :    .    :
03-             TGAAGTTTGATCTTGTAATTAAAACAACGAACAATTCGGATCTAAAGTTTTGTGCAATAT
70+             TGAAGTTTGATCTTGTAATTAAAACAACGAACAATTCGGATCTAAAGTTTTGTGCAATAT
02+             TGAAGTTTGATCTTGTAATTAAAACAACGAACAATTCGGATCTAAAGTTTTGTGCAATAT
                _____
consensus       TGAAGTTTGATCTTGTAATTAAAACAACGAACAATTCGGATCTAAAGTTTTGTGCAATAT


                      .    :    .    :    .    :    .    :    .    :    .    :
03-             GTGTGTGATTTACCCATGCATTCACGATGATTATTTCTCTGTAATAAGAAACAATTTCGC
70+             GTGTGTGATTTACCCATGCATTCACGATGATTATTTCTCTGTAATAAGAAACAATTTCGC
02+             GTGTGTGATTTACCCATGCATTCACGATGATTATTTCTCTGTAATAAGAAACAATTTCGC
                _____
consensus       GTGTGTGATTTACCCATGCATTCACGATGATTATTTCTCTGTAATAAGAAACAATTTCGC


                      .    :    .    :    .    :    .    :    .    .    .    :
03-             AAAATAAAGATTAGATAGCACTAACAAAAAAA
70+             AAAATAAAGATTAGATAGCA
02+             AAAATAAAGATTAGATAGCA
                _____
consensus       AAAATAAAGATTAGATAGCACTAACAAAAAAA
```

```
******************* Contig 2 ********************
                      .    :    .    :    .    :    .    :    .    :    .    :
09+             GTTCGTTTNTATTNCGACGTAGAAGCTGGAGAANGCAAAACTTTCGTNTATGGNGGATGC
                _____
consensus       GTTCGTTTNTATTNCGACGTAGAAGCTGGAGAANGCAAAACTTTCGTNTATGGNGGATGC


                      .    :    .    :    .    :    .    :    .    :    .    :
09+             GGTGGCAATGAAAACAACTTNGAGACTAAAGAGGAATGTGAGGAATCTTGTTCCGAATAA
51+               TGGCAATGAAAACAACTTCGAGACTAAAGAGGAATGTGAGGAATCTTGTTCCGAATAA
                _____
consensus       GGTGGCAATGAAAACAACTTCGAGACTAAAGAGGAATGTGAGGAATCTTGTTCCGAATAA


                      .    :    .    :    .    :    .    :    .    :    .    :
09+             TCTGTGTTCAGCTCCGAAGCAAATGTTCAGNGAGTGAGTTTCATNTAGTCAATAAAATTG
51+             TCTGTGTTCAGCTCCGAAGCAAATGTTCAGCGAGTGAGTTTCATCTAGTCAATAAAATTG
                _____
consensus       TCTGTGTTCAGCTCCGAAGCAAATGTTCAGCGAGTGAGTTTCATCTAGTCAATAAAATTG


                      .    :    .    :    .    :    .    :    .    :    .    :
09+             CTGGTTTCTGTATATAATATGCTCATTTATATAGATGTAACCAATNTTCCTTGTTGGACA
51+             CTGGTTTCTGTATATAATATGCTCATTTATATAGATGTAACCAATCTTCCTTGTTGGACA
                _____
consensus       CTGGTTTCTGTATATAATATGCTCATTTATATAGATGTAACCAATCTTCCTTGTTGGACA


                      .    :    .    :    .    :    .    :    .    :    .    :
09+             AATAAATAAAAAATNTAATAAATAAAAAAAAAAAAAAAAAAAAAAAAAA
51+             AATAAATAAAAAATCTAATAAATA
                _____
consensus       AATAAATAAAAAATCTAATAAATAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
******************* Contig 3 *******************
                        .    :    .    :    .    :    .    :    .    :    .    :
14+             TCANCTTCATTATTNAAGAATTTAGAATTCGTCGCTATTGGAGGGCCAACGTACGATCCG
49+                        TATTTAAGAATTTAGAATTCGTCGCTATTGGAGGGCCAACGTACGATCCG
                _____
consensus       TCANCTTCATTATTTAAGAATTTAGAATTCGTCGCTATTGGAGGGCCAACGTACGATCCG

                        .    :    .    :    .    :    .    :    .    :    .    :
14+             CTTCCACCGTTTAAGTGGAGTGACTCGNACTTCGGATCGACCATACCTCATGAAGGACAT
49+             CTTCCACCGTTTAAGTGGAGTGACTCGGACTTCGGATCGACCATACCTCATGAAGGACAT
                _____
consensus       CTTCCACCGTTTAAGTGGAGTGACTCGGACTTCGGATCGACCATACCTCATGAAGGACAT

                        .    :    .    :    .    :    .    :    .    :    .    :
14+             CCAGATCTGTGGAAATTTATGCCCATCGTACATAAATGGTTGCAATGATGCCTCAAACAT
49+             CCAGATCTGTGGAAATTTATGCCCATCGTACATAAATGGTTGCAATGATGCCTCAAACAT
                _____
consensus       CCAGATCTGTGGAAATTTATGCCCATCGTACATAAATGGTTGCAATGATGCCTCAAACAT

                        .    :    .    :    .    :    .    :    .    :    .    :
14+             CACATGTAGTTTTGGAAATAGCATTGNAGTTATCTATTATCCAATAAAAATTTCTAAAAA
49+             CACATGTAGTTTTGGAAATAGCATTGTAGTTATCTATTATCCAATAAAAATTTCTAAAAA
                _____
consensus       CACATGTAGTTTTGGAAATAGCATTGTAGTTATCTATTATCCAATAAAAATTTCTAAAAA

                        .    :    .    :    .    :    .    :    .    :    .    :
14+             C
49+             CA
                _____
consensus       CAY

******************* Contig 4 *******************
                        .    :    .    :    .    :    .    :    .    :    .    :
17+             TGAAAACATTAAATTTGGTCCTTTCGTACTAAAATTTAAAAAGAGAAGATAGAAACCGAC
61+             TGAAAACATTAAATTTGGTCCTTTCGTACTAAAATTTAAAAAGAGAAGATAGAAACCGAC
58+             TGAAAACATTAAATTTGGTCCTTTCGTACTAAAATTTAAAAAGAGAAGATAGAAACCGAC
75+             TGAAAACATTAAATTTGGTCCTTTCGTACTAAAATTTAAAAAGAGAAGATAGAAACCGAC
64+                     TAAATTTGGTCCTTTCGTACTAAAATTTAAAAAGAGAAGATAGAAACCGAC
                _____
consensus       TGAAAACATTAAATTTGGTCCTTTCGTACTAAAATTTAAAAAGAGAAGATAGAAACCGAC

                        .    :    .    :    .    :    .    :    .    :    .    :
17+             CTGGCTTACGCCGGTCTGAACTCAAATCATGTAAATTATTAAAAGTCGAACAGACTTTCT
61+             CTGGCTTACGCCGGTCTGAACTCAAATCATGTAAATTATTAAAAGTCGAACAGACTTTCT
58+             CTGGCTTACGCCGGTCTGAACTCAAATCATGTAAATTATTAAAAGTCGAACAGACTTTCT
75+             CTGGCTTACGCCGGTCTGAACTCAAATCATGTAAATTATTAAAAGTCGAACAGACTTTCT
64+             CTGGCTTACGCCGGTCTGAACTCAAATCATGTAAATTATTAAAAGTCGAACAGACTTTCT
                _____
consensus       CTGGCTTACGCCGGTCTGAACTCAAATCATGTAAATTATTAAAAGTCGAACAGACTTTCT

                        .    :    .    :    .    :    .    :    .    :    .    :
17+             ATTCTTT
61+             ATTCT
58+             ATTCTTTAATTTTGCGTAAAGGAGATTTTTAATTCAACATCGAGGTCATAATCTTTTTTT
75+             ATTCTTTAATTTTGCGTAAAGGAGATTTTTAATTCAACATCGAGGTCATAATCTTTTTTT
64+             ATTCTTTAATTTTGCGTAAAGGAGATTTTTAATTCAACATCGAGGTCATAATCTTTTTTT
                _____
consensus       ATTCTTTAATTTTGCGTAAAGGAGATTTTTAATTCAACATCGAGGTCATAATCTTTTTTT

                        .    :    .    :    .    :    .    :    .    :    .    :
58+             TTGATAAGATCTCTAAAAAAAAATTGTGC
75+             TTGATAAGATCTCTAAAAAAAAATTGTGCTGTTATCCCTATAGTAACTTGATTTATTA
64+             TTGATAAGATCTCTAAAAAAAAATTGTGCTGTTATCCCTATAGTAACTTGATTTATTA
                _____
consensus       TTGATAAGATCTCTAAAAAAAAATTGTGCTGTTATCCCTATAGTAACTTGATTTATTA
```

292

```
****************** Contig 5 *******************
              .    :    .    :    .    :    .    :    .    :    .    :
33+           AGGACACCCAACTTTGTTTNCCCGGGGGCTGCCGGGCGAGACATTGAAGGANNAGNCGNG
              _____
consensus     AGGACACCCAACTTTGTTTNCCCGGGGGCTGCCGGGCGAGACATTGAAGGANNAGNCGNG


              .    :    .    :    .    :    .    :    .    :    .    :
33+           NANNNNNCTTGGTTGATCTGGGTCGGAGGTGGACCTCTTCTGATCGTCGTCGAACCTCTG
85+                      CTTGGTTGATCTGGGTCGGAGGTGGACCTCTTCTGATCGTCGTCGAACCTCTG
              _____
consensus     NANNNNNCTTGGTTGATCTGGGTCGGAGGTGGACCTCTTCTGATCGTCGTCGAACCTCTG


              .    :    .    :    .    :    .    :    .    :    .    :
33+           ACTTTCGTTCTTGACTAATGAAAACCTGCTTGGCACATGCTTTCGCAGTATTTCGTCCTA
85+           ACTTTCGTTCTTGACTAATGAAAACCTGCTTGGCACATGCTTTCGCAGTATTTCGTCCTA
              _____
consensus     ACTTTCGTTCTTGACTAATGAAAACCTGCTTGGCACATGCTTTCGCAGTATTTCGTCCTA


              .    :    .    :    .    :    .    :    .    .    .    :
33+           CGGTGATCCGAGATTTTCACCGCTGAACCCGTA
85+           CGGTGATCCGAGATTTTCACCGCTGAACCCGTA
              _____
consensus     CGGTGATCCGAGATTTTCACCGCTGAACCCGTA




****************** Contig 6 *******************
              .    :    .    :    .    :    .    :    .    :    .    :
34+           TTATGGGGGTTTGCGAGCACATTTGACTCAACACGGGAAATCGCACCAGANCCGAACATC
              _____
consensus     TTATGGGGGTTTGCGAGCACATTTGACTCAACACGGGAAATCGCACCAGANCCGAACATC

              .    :    .    :    .    :    .    :    .    :    .    :
34+           TCTAAGGATTGAAAGGAAAGCTCTTTCTTGATGAAGTGGATGGTGGTGCATGGTTCTTCA
86+                            AGCTCTTTCTTGATGAAGTGGATGGTGGTGCATGGTTCTTCA
              _____
consensus     TCTAAGGATTGAAAGGAAAGCTCTTTCTTGATGAAGTGGATGGTGGTGCATGGTTCTTCA

              .    :    .    :    .    :    .    :    .    :    .    :
34+           TAGTTGGGGGAGTGATCTGTCTGGTTAATTCCTATAACAAACGAGACTCTCCCCTGCTAA
86+           TAGTTGGGGGAGTGATCTGTCTGGTTAATTCCTATAACAAACGAGACTCTCCCCTGCTAA
              _____
consensus     TAGTTGGGGGAGTGATCTGTCTGGTTAATTCCTATAACAAACGAGACTCTCCCCTGCTAA

              .    :    .    :    .    :    .    :    .    :    .    :
34+           CGGACCTTAGTGTTTTTTTAGACGGTGCGATTCTGATAAAGGGACCATGGGTGTAAGCCC
86+           CGGACCTTAGTGTTTTTTTAGACGGTGCGATTCTGATAAAGGGACCATGGGTGTAAGCCC
              _____
consensus     CGGACCTTAGTGTTTTTTTAGACGGTGCGATTCTGATAAAGGGACCATGGGTGTAAGCCC

              .    :    .    :    .    :    .    :    .    :    .    :
34+           CGGTCTCAGACAACGTCCACTGATCCATGGTACGTCGGTGTTCTCTCTGGAGTCTGGGCG
86+           CGGTCTCAGACAACGTCCACTGATCCATGGTACGTCGGTGTTCTCTCTGGAGTCTGGGCG
              _____
consensus     CGGTCTCAGACAACGTCCACTGATCCATGGTACGTCGGTGTTCTCTCTGGAGTCTGGGCG

              .    :    .    :    .    :    .    :    .    :    .    :
34+           CTCGTGTGTGGGCGCCATGTCTCTCTCTTGTCAAGGCCCGTGGCAAAGCGTTGTTGCAGC
86+           CTCGTGTGTGGGCGCCATGTCTCTCTCTTGTCAAGGCCCGTGGCAAAGCGTTGTTGCAGC
              _____
consensus     CTCGTGTGTGGGCGCCATGTCTCTCTCTTGTCAAGGCCCGTGGCAAAGCGTTGTTGCAGC

              .    :    .    :    .    :    .    :    .    :    .    :
34+           GAGGGTGTGGGCTCATCTCATGCGGGGCGTCTGTATCTCTCGGGGGGATCATTGACAAAG
86+           GAGGGTGTGGGCTCATCTCATGCGGGGCGTCTGTATCTCTCGGGGGGATCATTGACAAAG
              _____
consensus     GAGGGTGTGGGCTCATCTCATGCGGGGCGTCTGTATCTCTCGGGGGGATCATTGACAAAG

              .    :    .    :    .    :    .    :    .    .    .    :
34+           TGCCTCGGAGTACGTTTCTTCTTAGAGGGATTGACCACTCATAAGTCGTAATAAACAGGG
86+           TGCCTCGGAGTACGTTTCTTCTTAGAGGGATTGACCACTCATAAGTCGTAATAAACAGGG
              _____
consensus     TGCCTCGGAGTACGTTTCTTCTTAGAGGGATTGACCACTCATAAGTCGTAATAAACAGGG

              .    :    .    :    .    :    .    :    .    .    .    :
34+           CGCAATAACAGGTCTGTGATGCCCTTAGATGT
86+           CGCAATAACAGGTCTGTGATGCCCTTAGATGT
              _____
consensus     CGCAATAACAGGTCTGTGATGCCCTTAGATGT
```

293

```
******************* Contig 7 ********************
                      .   :   .   :   .   :   .   :   .   :   .   :
37-          TGGCGCATTCCTGCGTTTTGACGACAAAAGGTGCCCGCTGGAGAATTATTAAATGAAGGA
39+                                      TGGAGAATAATTAAATGAAGGA
             _____
consensus    TGGCGCATTCCTGCGTTTTGACGACAAAAGGTGCCCGCTGGAGAATAATTAAATGAAGGA

                      .   :   .   :   .   :   .   :   .   :   .   :
37-          GGGTTAAATTTTATTGCATTTTCTTGGATGCTTTTGCGAGCTGAATCCACGCCGCACGCC
39+          GGGTTAAATTTTATTGCATTTTCTTGGATGCTTTTGCGAGCTGAATCCACGCCGCACGCC
             _____
consensus    GGGTTAAATTTTATTGCATTTTCTTGGATGCTTTTGCGAGCTGAATCCACGCCGCACGCC

                      .   :   .   :   .   :   .   :   .   :   .   :
37-          AGCGGCTCGCTATCTATATGCGTTCAGGCTAAAATCGCAACGGTATCTGAAAAGCGGTGG
39+          AGCGGCTCGCTATCTATATGCGTTCAGGCTAAAATCGCAACGGTATCTGAAAAGCGGTGG
             _____
consensus    AGCGGCTCGCTATCTATATGCGTTCAGGCTAAAATCGCAACGGTATCTGAAAAGCGGTGG

                      .   :   .   :   .   :   .   :   .   :   .   :
37-          TCGGAAACGCCGAAATTCCTGCGTTCTGACGACAAAACCTACCCGCTGGAGAGTTATAAA
39+          TCGGAAACGCCGAAATTCCTGCGTTCTGACGACAAAACCTACCCGCTGCAGAGTGATAAA
             _____
consensus    TCGGAAACGCCGAAATTCCTGCGTTCTGACGACAAAACCTACCCGCTGCAGAGTGATAAA

                      .   :   .   :   .   :   .   :   .   :   .   :
37-          ATGAAGGCGGGTTAACATTTATTGCATTTTCTTGGATGCTTTTGCGAGCTGGATTCATAC
39+          ATGAAGGCGGGTTAACTTTTATTGCATTTTCTTGGATGCTTTTGCGAGCTGGATTCATAC
             _____
consensus    ATGAAGGCGGGTTAACATTTATTGCATTTTCTTGGATGCTTTTGCGAGCTGGATTCATAC

                      .   :   .   :   .   :   .   :   .   :   .   :
37-          CGCATGCGAGCGGCTCGCTATCTCTGTGCGTTCAGGCTGAAATCGCAACGGTATCGGAAA
39+          CGCATGCGAGCGGCTCGCTATCTCTGTACGTTCAGGCTGAAATCCCAACGGTATCGGAAA
             _____
consensus    CGCATGCGAGCGGCTCGCTATCTCTGTACGTTCAGGCTGAAATCCCAACGGTATCGGAAA

                      .   :   .   :   .   :   .   :   .   :   .   :
37-          AGCCGTGGTCGAAAACGTCAAATTTCGAAATAACGCTAGAAACACAAATCCTGCTTTTTG
39+          AGCCGTGGTCGAAAACGTCAAATTTCGAAATAACGCTAGAAACACAAATCCTGCATTTTG
             _____
consensus    AGCCGTGGTCGAAAACGTCAAATTTCGAAATAACGCTAGAAACACAAATCCTGCATTTTG

                      .   :   .   :   .   :   .   :   .   :   .   :
37-          TTGTGTGGTTATGGATCACTAATATATTGCACATCAAATGCACTGGGTGCCGTGTTTTTT
39+          TTGTGTGGTTATGGATCACTAATATATTGCACATCAAATGCACTGGGTGCCGTGTTTTTT
             _____
consensus    TTGTGTGGTTATGGATCACTAATATATTGCACATCAAATGCACTGGGTGCCGTGTTTTTT

                      .   :   .   :   .   :   .   :   .   :   .   :
37-          CCGCTTTAAAATCGGCACCGAGTGTTTGGGGTCAGCGTTTTGTCAGCGAAAATTTAGCAG
39+          CCGCTTTAAAATCGGCACCGAGTATTTGGGGTCAGCGTTTTGTCAGCGAAAATTTAGCA
             _____
consensus    CCGCTTTAAAATCGGCACCGAGTATTTGGGGTCAGCGTTTTGTCAGCGAAAATTTAGCAG




******************* Contig 8 ********************
                      .   :   .   :   .   :   .   :   .   :   .   :
42+          GTACAACATCTGCCAAATTCCCACATTCTTACTCATTAAGAATAACTTAGTAAGCTTTGG
47+          GTACAACATCTGCCAAATTCCCACATTCTTACTCATTAAGAATAACTTAGTAAGCTTTGG
             _____
consensus    GTACAACATCTGCCAAATTCCCACATTCTTACTCATTAAGAATAACTTAGTAAGCTTTGG

                      .   :   .   :   .   :   .   :   .   :   .   :
42+          GATTCACCGCTATAAGTATAGTAACAAACACACACACGTATAA
47+          GATTCACCGCTATAAGTATAGTAACAAACACACACACGTATAA
             _____
consensus    GATTCACCGCTATAAGTATAGTAACAAACACACACACGTATAA
```

294

```
******************* Contig 9 *******************
               .    :    .    :    .    :    .    :    .    :    .    :
52+          GAAGACAGGAAGAGGCTTACTGTTTAAGCCAGGAGGAAAATATAAACTTACTGTACATAA
74+          GAAGACAGGAAGAGGCTTACTGTTTAAGCCAGGAGGAAAATATAAACTTACTGTACATAA
             _____
consensus    GAAGACAGGAAGAGGCTTACTGTTTAAGCCAGGAGGAAAATATAAACTTACTGTACATAA

               .    :    .    :    .    :    .    :    .    :    .    :
52+          ATCACTGTAAAGCAATAAACAAACCAACTTTTCAAAACTTACTCGAAACATATTCTGAAC
74+          ATCACTGTAAAGCAATAAACAAACCAACTTTTCAAAACTTACTCGAAACATATTCTGAAC
             _____
consensus    ATCACTGTAAAGCAATAAACAAACCAACTTTTCAAAACTTACTCGAAACATATTCTGAAC

               .    :    .    :    .    :    .    :    .    :    .    :
52+          TATTTGAAAAAATTATGTGAAAAAAAAAAAATTGTGGGAAAAAGTTGTCTCAAAATGGA
74+          TATTTGAAAAAATTATGTGAAAAAAAAAAAATTGTGGGAAAAAGTTGTCTCAAAATGGA
             _____
consensus    TATTTGAAAAAATTATGTGAAAAAAAAAAAATTGTGGGAAAAAGTTGTCTCAAAATGGA

               .    :    .    :    .    :    .    :    .    :    .    :
52+          TAGTTCGTATGATATGAAAGCTCGCAAACCACTGCTCATTAGTTTTTGCTCAATTTGGTC
74+          TAGTTCGTATGATATGAAAGCTCGCAAACCACTGCTCATTAGTTTTTGCTCAATTTGGTC
             _____
consensus    TAGTTCGTATGATATGAAAGCTCGCAAACCACTGCTCATTAGTTTTTGCTCAATTTGGTC

               .    :    .    :    .    :    .    :    .    :    .    :
52+          ATTTTTGGTGGAAATAAAGACCTAAACA
74+          ATTTTTGGTGGAAATAAAGACCTAAACA
             _____
consensus    ATTTTTGGTGGAAATAAAGACCTAAACAY




******************* Contig 10 *******************
               .    :    .    :    .    :    .    :    .    :    .    :
55+          AAAAGAGAAACAATTTATTAAATTCTAAGTATCTTCGTCGTCTTGGTCGTAATCCATTTG
65+          AAAAGAGAAACAATTTATTAAATTCTAAGTATCTTCGTCGTCTTGGTCGTAATCCATTTG
             _____
consensus    AAAAGAGAAACAATTTATTAAATTCTAAGTATCTTCGTCGTCTTGGTCGTAATCCATTTG

               .    :    .    :    .    :    .    :    .    :    .    :
55+          GAAAAAAAA-GATCCTACAAGAATTTGAAAAACCTAGGAGAATTTGAAGAACAGTCTGTT
65+          GAAAAAAAACGATCCTACAAGAATTTGAAAAACCTAGGAGAATTTGAAGAACAGTCTGTT
             _____
consensus    GAAAAAAAACGATCCTACAAGAATTTGAAAAACCTAGGAGAATTTGAAGAACAGTCTGTT

               .    :    .    :    .    :    .    :    .    :    .    :
55+          CACTAACTTGACACAGAGAAAATAAGAAGAAAGTTAGTCCTTTGATTTGAAATTCGGAAT
65+          CACTAACTTGACACAGAGAAAATAAGAAGAAAGTTAGTCCTTTGATTTGAAATTCGGAAT
             _____
consensus    CACTAACTTGACACAGAGAAAATAAGAAGAAAGTTAGTCCTTTGATTTGAAATTCGGAAT

               .    :    .    :    .    :    .    :    .    :    .    :
55+          GTGATAACTAAACTGAACA
65+          GTGATAACTAAACTGAACA
             _____
consensus    GTGATAACTAAACTGAACAY
```

295

```
****************** Contig 11 ********************
               .    :    .    :    .    :    .    :    .    :    .    :
56+            TTTTNNNTGATTGAAACCAAGAAATATGTTTATTTAAATCTGTTAAAATTTTTTTAAAAA
76+                   TGATTGAAACCAAGAAATATGTTTATTTAAATCTGTTAAAATTTTTTTAAAAA
               _____
consensus      TTTTNNNTGATTGAAACCAAGAAATATGTTTATTTAAATCTGTTAAAATTTTTTTAAAAA

               .    :    .    :    .    :    .    :    .    :    .    :
56+            TTAAATTTTTTAAACAATTTAAAATTTTTACTACAAATTTTAAGAAAAAGGTTAAATTTG
76+            TTAAATTTTTTAAACAATTTAAAATTTTTACTACAAATTTTAAGAAAAAGGTTAAATTTG
               _____
consensus      TTAAATTTTTTAAACAATTTAAAATTTTTACTACAAATTTTAAGAAAAAGGTTAAATTTG

               .    :    .    :    .    :    .    :    .    :    .    :
56+            CAGAAATTTTTTTATTTGGAAAATTCATAATTGAAATACCTTTATTATTTTAATGCAAAA
76+            CAGAAATTTTTTTATTTGGAAAATTCATAATTGAAATACCTTTATTATTTTAATGCAAAA
               _____
consensus      CAGAAATTTTTTTATTTGGAAAATTCATAATTGAAATACCTTTATTATTTTAATGCAAAA

               .    :    .    :    .    :    .    :    .    :    .    :
56+            AAAAACAAAAAAAAAACGAGACATGTTTTCCTGATTAAAGTTAAACTGACTTTTCAGTTA
76+            AAAAACAAAAAAAAAACGAGACATGTTTTCCTGATTAAAGTTAAACTGACTTTTCAGTTA
               _____
consensus      AAAAACAAAAAAAAAACGAGACATGTTTTCCTGATTAAAGTTAAACTGACTTTTCAGTTA

               .    :    .    :    .    :    .    :    .    :    .    :
56+            AAACCATTATAAAAAAAGAATTAATACGAGCATGATTATGAAAAAACATTAGTTTCTGTT
76+            AAACCATTATAAAAAAAGAATTAATACGAGCATGATTATGAAAAAACATTAGTTTCTGTT
               _____
consensus      AAACCATTATAAAAAAAGAATTAATACGAGCATGATTATGAAAAAACATTAGTTTCTGTT

               .    :    .    :    .    :    .    :    .    :    .    :
56+            TGCAAACATTTGCATTTGATTTCAATGACAGCTGCGA
76+            TGCAAACATTTGCATTTGATTTCAATGACAGCTGCGA
               _____
consensus      TGCAAACATTTGCATTTGATTTCAATGACAGCTGCGAY




****************** Contig 12 ********************
               .    :    .    :    .    :    .    :    .    :    .    :
83-            GTTTTTGGTCAAGTAACAGATGGTCTGGATGTTGTGAAGAAGATTGAAACCTTTGGTAGC
66+                                              AGATTGAAACCTTTGGTAGC
               _____
consensus      GTTTTTGGTCAAGTAACAGATGGTCTGGATGTTGTGAAGAAGATTGAAACCTTTGGTAGC

               .    :    .    :    .    :    .    :    .    :    .    :
83-            CAGAGTGGGAAGACAAGCAAAAGAATTGTTGTTGCAAACTGTGGTCAACTTTCTTAACTT
66+            CAGAGTGGGAAGACAAGCAAAAGAATTGTTGTTGCAAACTGTGGTCAACTTTCTTAACTT
               _____
consensus      CAGAGTGGGAAGACAAGCAAAAGAATTGTTGTTGCAAACTGTGGTCAACTTTCTTAACTT

               .    :    .    :    .    :    .    :    .    :    .    :
83-            CATGCTGTGTGTTATTCACTGCAATGTTTAATAATTTGGGTTGTACATTTTATATGTACA
66+            CATGCTGTGTGTTATTCACTGCAATGTTTAATAATTTGGGTTGTACATTTTATATGTACA
               _____
consensus      CATGCTGTGTGTTATTCACTGCAATGTTTAATAATTTGGGTTGTACATTTTATATGTACA

               .    :    .    :    .    :    .    :    .    :    .    :
83-            TAATGTTTTGCAATAAACA
66+            TAATGTTTTGCA
               _____
consensus      TAATGTTTTGCAATAAACA
```

Contig sequences

```
>02x03x70
TAGAAACATCTGTTTGGATTCCATTGGAAAGTGTCGGGGCCCCTACGCTGAAGGAAAAGAAGACAGTCTAATGTTCTCA
CCAATGGCGGTCAAGAAACCAGTAATGCTGCAACCCTAATGCGGAGATAGTGAATGATGTATTAGATCTTGGGAGAGCG
TTACAAAGCTGTAGTTGATGACTGAAGTTTGATCTTGTAATTAAAACAACGAACAATTCGGATCTAAAGTTTTGTGCAA
TATGTGTGTGATTTACCCATGCATTCACGATGATTATTTCTCTGTAATAAGAAACAATTTCGCAAAATAAAGATTAGAT
AGCACTAACAAAAAAA
>09x51
GTTCGTTTNTATTNCGACGTAGAAGCTGGAGAANGCAAAACTTTCGTNTATGGNGGATGCGGTGGCAATGAAAACAACT
TCGAGACTAAAGAGGAATGTGAGGAATCTTGTTCCGAATAATCTGTGTTCAGCTCCGAAGCAAATGTTCAGCGAGTGAG
TTTCATCTAGTCAATAAAATTGCTGGTTTCTGTATATAATATGCTCATTTATATAGATGTAACCAATCTTCCTTGTTGG
ACAAATAAATAAAAAATCTAATAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>14x59
TCANCTTCATTATTTAAGAATTTAGAATTCGTCGCTATTGGAGGGCCAACGTACGATCCGCTTCCACCGTTTAAGTGGA
GTGACTCGGACTTCGGATCGACCATACCTCATGAAGGACATCCAGATCTGTGGAAATTTATGCCCATCGTACATAAATG
GTTGCAATGATGCCTCAAACATCACATGTAGTTTTGGAAATAGCATTGTAGTTATCTATTATCCAATAAAAATTTCTAA
AAACAY
>17x58x64x75
TGAAAACATTAAATTTGGTCCTTTCGTACTAAAATTTAAAAAGAGAAGATAGAAACCGACCTGGCTTACGCCGGTCTGA
ACTCAAATCATGTAAATTATTAAAAGTCGAACAGACTTTCTATTCTTTAATTTTGCGTAAAGGAGATTTTTAATTCAAC
ATCGAGGTCATAATCTTTTTTTTTTGATAAGATCTCTAAAAAAAAATTGTGCTGTTATCCCTATAGTAACTTGATTTATT
A
>33x85
AGGACACCCAACTTTGTTTNCCCGGGGGCTGCCGGGCGAGACATTGAAGGANNAGNCGNGNANNNNNCTTGGTTGATCT
GGGTCGGAGGTGGACCTCTTCTGATCGTCGTCGAACCTCTGACTTTCGTTCTTGACTAATGAAAACCTGCTTGGCACAT
GCTTTCGCAGTATTTCGTCCTACGGTGATCCGAGATTTTCACCGCTGAACCCGTA
>34x86
TTATGGGGGTTTGCGAGCACATTTGACTCAACACGGGAAATCGCACCAGANCCGAACATCTCTAAGGATTGAAAGGAAA
GCTCTTTCTTGATGAAGTGGATGGTGGTGCATGGTTCTTCATAGTTGGGGGAGTGATCTGTCTGGTTAATTCCTATAAC
AAACGAGACTCTCCCCTGCTAACGGACCTTAGTGTTTTTTTAGACGGTGCGATTCTGATAAAGGGACCATGGGTGTAAG
CCCCGGTCTCAGACAACGTCCACTGATCCATGGTACGTCGGTGTTCTCTCTGGAGTCTGGGCGCTCGTGTGTGGGCGCC
ATGTCTCTCTCTTGTCAAGGCCCGTGGCAAAGCGTTGTTGCAGCGAGGGTGTGGGCTCATCTCATGCGGGGCGTCTGTA
TCTCTCGGGGGGATCATTGACAAAGTGCCTCGGAGTACGTTTCTTCTTAGAGGGATTGACCACTCATAAGTCGTAATAA
ACAGGGCGCAATAACAGGTCTGTGATGCCCTTAGATGT
>37x39
TGGCGCATTCCTGCGTTTTGACGACAAAAGGTGCCCGCTGGAGAATAATTAAATGAAGGAGGGTTAAATTTTATTGCAT
TTTCTTGGATGCTTTTGCGAGCTGAATCCACGCCGCACGCCAGCGGCTCGCTATCTATATGCGTTCAGGCTAAAATCGC
AACGGTATCTGAAAAGCGGTGGTCGGAAACGCCGAAATTCCTGCGTTCTGACGACAAAACCTACCCGCTGCAGAGTGAT
AAAATGAAGGCGGGTTAACATTTATTGCATTTTCTTGGATGCTTTTGCGAGCTGGATTCATACCGCATGCGAGCGGCTC
GCTATCTCTGTACGTTCAGGCTGAAATCCCAACGGTATCGGAAAAGCCGTGGTCGAAAACGTCAAATTTCGAAATAACG
CTAGAAACACAAATCCTGCATTTTGTTGTGTGGTTATGGATCACTAATATATTGCACATCAAATGCACTGGGTGCCGTG
TTTTTTCCGCTTTAAAATCGGCACCGAGTATTTGGGGTCAGCGTTTTGTCAGCGAAAATTTAGCAG
>42x47
GTACAACATCTGCCAAATTCCCACATTCTTACTCATTAAGAATAACTTAGTAAGCTTTGGGATTCACCGCTATAAGTAT
AGTAACAAACACACACGTATAA
>52x74
GAAGACAGGAAGAGGCTTACTGTTTAAGCCAGGAGGAAAATATAAACTTACTGTACATAAATCACTGTAAAGCAATAAA
CAAACCAACTTTTCAAAACTTACTCGAAACATATTCTGAACTATTTGAAAAAATTATGTGAAAAAAAAAAAAATTGTGGG
AAAAAAGTTGTCTCAAAATGGATAGTTCGTATGATATGAAAGCTCGCAAACCACTGCTCATTAGTTTTTGCTCAATTTG
GTCATTTTTGGTGGAAATAAAGACCTAAACAY
>55x65
AAAAGAGAAACAATTTATTAAATTCTAAGTATCTTCGTCGTCTTGGTCGTAATCCATTTGGAAAAAAAACGATCCTACA
AGAATTTGAAAAACCTAGGAGAATTTGAAGAACAGTCTGTTCACTAACTTGACACAGAGAAATAAGAAGAAAGTTAGT
CCTTTGATTTGAAATTCGGAATGTGATAACTAAACTGAACAY
>56x76
TTTTNNNTGATTGAAACCAAGAAATATGTTTATTTAAATCTGTTAAAATTTTTTTAAAAATTAAATTTTTTAAACAATT
TAAAATTTTTACTACAAATTTTAAGAAAAAGGTTAAATTTGCAGAAATTTTTTTATTTGGAAAATTCATAATTGAAATA
CCTTTATTATTTTAATGCAAAAAAAAACAAAAAAAAAACGAGACATGTTTTCCTGATTAAAGTTAAACTGACTTTTCAG
TTAAAACCATTATAAAAAAAAGAATTAATACGAGCATGATTATGAAAAAACATTAGTTTCTGTTTGCAAACATTTGCATT
TGATTTCAATGACAGCTGCGAY
>66x83
GTTTTTGGTCAAGTAACAGATGGTCTGGATGTTGTGAAGAAGATTGAAACCTTTGGTAGCCAGAGTGGGAAGACAAGCA
AAAGAATTGTTGTTGCAAACTGTGGTCAACTTTCTTAACTTCATGCTGTGTGTTATTCACTGCAATGTTTAATAATTTG
GGTTGTACATTTTATATGTACATAATGTTTTGCAATAAACA
```

Appendix 6 *Stegodyphus* UNIX preliminary results

1. Sequence: isotig01322_2   from: 1 to: 380 1 Pattern: [qe]alle x
2. Sequence: isotig01884_1   from: 1 to: 165 1 Pattern: [qe]alle -
3. Sequence: isotig02339_2   from: 1 to: 182 1 Pattern: [qe]alle x
4. Sequence: isotig03074_1   from: 1 to: 437 1 Pattern: [qe]alle -
5. Sequence: isotig03639_2   from: 1 to: 221 1 Pattern: [qe]alle -
6. Sequence: isotig04633_3   from: 1 to: 532 1 Pattern: [qe]alle -
7. Sequence: isotig04827_2   from: 1 to: 293 1 Pattern: [qe]alle -
8. Sequence: isotig05124_3   from: 1 to: 339 1 Pattern: [qe]alle -
9. Sequence: isotig06029_1   from: 1 to: 510 1 Pattern: [qe]alle -
10. Sequence: isotig00559_3   from: 1 to: 203 1 Pattern: [qe]alle x
11. Sequence: isotig00560_3   from: 1 to: 198 1 Pattern: [qe]alle x
12. Sequence: isotig01054_2   from: 1 to: 287 1 Pattern: [qe]alle -
13. Sequence: isotig01251_3   from: 1 to: 553 1 Pattern: [qe]alle -
14. Sequence: isotig01413_2   from: 1 to: 283 1 Pattern: [qe]alle x
15. Sequence: isotig01414_2   from: 1 to: 285 1 Pattern: [qe]alle x
16. Sequence: isotig02406_2   from: 1 to: 222 1 Pattern: [qe]alle x
17. Sequence: isotig03093_3   from: 1 to: 736 1 Pattern: [qe]alle x
18. Sequence: isotig03171_1   from: 1 to: 873 1 Pattern: [qe]alle -
19. Sequence: isotig05141_3   from: 1 to: 172 1 Pattern: [qe]alle -
20. Sequence: isotig05380_3   from: 1 to: 157 1 Pattern: [qe]alle -
21. Sequence: isotig06426_2   from: 1 to: 238 1 Pattern: [qe]alle -
22. Sequence: isotig06543_1   from: 1 to: 165 1 Pattern: [qe]alle -
23. Sequence: isotig06784_3   from: 1 to: 357 1 Pattern: [qe]alle -
24. Sequence: isotig06954_3   from: 1 to: 214 1 Pattern: [qe]alle -
25. Sequence: isotig07544_2   from: 1 to: 543 1 Pattern: [qe]alle -
26. Sequence: isotig07676_2   from: 1 to: 291 1 Pattern: [qe]alle -
27. Sequence: isotig09145_2   from: 1 to: 697 1 Pattern: [qe]alle -
28. Sequence: isotig09351_1   from: 1 to: 576 1 Pattern: [qe]alle -
29. Sequence: isotig09770_3   from: 1 to: 326 1 Pattern: [qe]alle -
30. Sequence: isotig01162_3   from: 1 to: 171 1 Pattern: [qe]alle -
31. Sequence: isotig01651_1   from: 1 to: 423 1 Pattern: [qe]alle x
32. Sequence: isotig01651_2   from: 1 to: 422 1 Pattern: [qe]alle ?
33. Sequence: isotig01914_3   from: 1 to: 151 1 Pattern: [qe]alle -
34. Sequence: isotig01968_3   from: 1 to: 229 1 Pattern: [qe]alle -
35. Sequence: isotig03905_3   from: 1 to: 167 1 Pattern: [qe]alle -
36. Sequence: isotig04426_3   from: 1 to: 247 1 Pattern: [qe]alle -
37. Sequence: isotig04916_3   from: 1 to: 325 1 Pattern: [qe]alle -
38. Sequence: isotig04987_2   from: 1 to: 242 1 Pattern: [qe]alle -
39. Sequence: isotig06587_2   from: 1 to: 451 1 Pattern: [qe]alle -
40. Sequence: isotig06934_1   from: 1 to: 366 1 Pattern: [qe]alle -
41. Sequence: isotig06965_1   from: 1 to: 332 1 Pattern: [qe]alle -

Appendix 7 Sex prediction tables calculated from the logistic regression equation

Sex prediction table calculated from the logistic regression equation (Z) using the statistic SD/L.

	SD/L			
X	M	F	LR	LR/(1+LR)
5.0	0	0.07	0	0
5.5	0	0.13	0	0
6.0	0	0.21	0	0
6.5	0	0.32	0	0
7.0	0	0.44	0.01	0.01
7.5	0.01	0.54	0.02	0.02
8.0	0.03	0.61	0.04	0.04
8.5	0.07	0.62	0.11	0.1
9.0	0.15	0.58	0.27	0.21
9.5	0.29	0.48	0.6	0.38
10.0	0.47	0.37	1.28	0.56
10.5	0.65	0.25	2.57	0.72
11.0	0.77	0.16	4.85	0.83
11.5	0.78	0.09	8.62	0.9
12.0	0.67	0.05	14.44	0.94
12.5	0.5	0.02	22.8	0.96
13.0	0.31	0.01	33.93	0.97
13.5	0.17	0	47.58	0.98
14.0	0.08	0	62.87	0.98
14.5	0.03	0	78.3	0.99
15.0	0.01	0	91.89	0.99
15.5	0	0	101.63	0.99
16.0	0	0	105.93	0.99

Sex prediction table calculated from the logistic regression equation (Z) using the statistic SD/W.

| | SD/W | | | |
X	M	F	LR	LR/(1+LR)
4.0	0	0	0	0
4.5	0	0.01	0	0
5.0	0	0.03	0	0
5.5	0	0.06	0	0
6.0	0	0.12	0	0
6.5	0	0.21	0	0
7.0	0	0.33	0	0
7.5	0	0.46	0.01	0.01
8.0	0.01	0.58	0.02	0.02
8.5	0.03	0.65	0.05	0.05
9.0	0.07	0.66	0.11	0.1
9.5	0.15	0.6	0.26	0.21
10.0	0.28	0.48	0.58	0.37
10.5	0.44	0.35	1.25	0.55
11.0	0.6	0.23	2.63	0.72
11.5	0.71	0.13	5.4	0.84
12.0	0.74	0.07	10.79	0.92
12.5	0.67	0.03	21	0.95
13.0	0.53	0.01	39.79	0.98
13.5	0.36	0	73.39	0.99
14.0	0.22	0	131.8	0.99

Sex prediction table calculated from the logistic regression equation (Z) using the statistic SD/F.

	SD/F			
x	M	F	LR	LR/(1+LR)
5.0	0	0	0	0
5.5	0	0.01	0	0
6.0	0	0.01	0	0
6.5	0	0.02	0	0
7.0	0	0.04	0	0
7.5	0	0.08	0	0
8.0	0	0.13	0	0
8.5	0	0.19	0.01	0.01
9.0	0.01	0.28	0.02	0.02
9.5	0.02	0.36	0.05	0.05
10.0	0.05	0.44	0.11	0.1
10.5	0.12	0.51	0.23	0.19
11.0	0.23	0.54	0.42	0.3
11.5	0.39	0.53	0.73	0.42
12.0	0.57	0.49	1.17	0.54
12.5	0.71	0.41	1.73	0.63
13.0	0.77	0.33	2.36	0.7
13.5	0.72	0.24	2.99	0.75
14.0	0.58	0.17	3.5	0.78
14.5	0.4	0.11	3.8	0.79
15.0	0.24	0.06	3.82	0.79

Sex prediction table calculated from the logistic regression equation (Z) using the statistic SD/LW.

		SD/LW		
x	M	F	LR	LR/(1+LR)
2.0	0	0.01	0	0
2.5	0	0.07	0	0
3.0	0	0.3	0	0
3.5	0	0.78	0	0
4.0	0.03	1.27	0.02	0.02
4.5	0.19	1.3	0.14	0.13
5.0	0.73	0.84	0.87	0.47
5.5	1.46	0.34	4.35	0.81
6.0	1.52	0.08	18	0.95
6.5	0.82	0.01	61.59	0.98
7.0	0.23	0	174.3	0.99

Sex prediction table calculated from the logistic regression equation (Z) using the statistic SD/LWF.

		SD/LWF		
x	M	F	LR	LR/(1+LR)
0	0	0	0	0
0.5	0	0	0	0
1.0	0	0	0	0
1.5	0	0.01	0	0
2.0	0	0.19	0	0
2.5	0	1.01	0	0
3.0	0.11	1.98	0.05	0.05
3.5	1.14	1.41	0.81	0.45
4.0	2.51	0.37	6.78	0.87
4.5	1.14	0.04	32.05	0.97
5.0	0.11	0	85.46	0.99
5.5	0	0	128.65	0.99
6.0	0	0	109.34	0.99

Sex prediction table calculated from the logistic regression equation (Z) using the statistic F/LW.

F/LW

x	M	F	LR	1/(1+LR)
0.33	0.04	10.3	0	0
0.34	0.11	13.38	0.01	0.01
0.35	0.31	16.5	0.02	0.02
0.36	0.79	19.31	0.04	0.04
0.37	1.79	21.46	0.08	0.08
0.38	3.69	22.64	0.16	0.14
0.39	6.84	22.67	0.3	0.23
0.40	11.44	21.55	0.53	0.35
0.41	17.25	19.45	0.89	0.47
0.42	23.45	16.67	1.41	0.58
0.43	28.76	13.56	2.12	0.68
0.44	31.8	10.47	3.04	0.75
0.45	31.7	7.68	4.13	0.81
0.46	28.5	5.34	5.33	0.84
0.47	23.1	3.53	6.54	0.87
0.48	16.89	2.21	7.63	0.88
0.49	11.13	1.32	8.44	0.89
0.50	6.62	0.75	8.88	0.9