Characterization of black fly (Diptera: Simuliidae) silk proteins

by

Patrick R. Viel, B.Sc.

A Thesis

submitted to the Department of Biological Sciences

in partial fulfillment of the requirements

for the degree of

Master of Science

July, 2014

Brock University

St. Catharines, Ontario

Abstract

Black fly (Simuliidae) silk is produced by the larvae and pharate pupae and is used for anchorage and cocoon production. There exists limited information on simuliid silks, including protein composition and genetic sequences encoding such proteins. The present study aimed to expand what is known about simuliid silks by examining the silks of several simuliid species and by making comparisons to the silk of non-biting midges (Chironomidae).

Silk glands were dissected out of larval and pupal simuliids, and protein contents were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and visualized with silver stain. Protein contents were compared by mass in kilodaltons (kDa) between life stages and among species. Polymerase chain reaction (PCR) was used to expand upon known gene sequence information, and to determine the presence of genes homologous to chironomid silk. SDS-PAGE of cocoons revealed the presence of a 56 kDa and a 67 kDa protein.

Silk gland contained as many as 28 different proteins ranging from 319 kDa to 8 kDa. Protein profiles vary among species, and group into large (>200), intermediate(>100), and small (<100) protein classes as is found in chironomids.

It is likely that silk evolved in a common ancestor of simuliids and chironomids

Acknowledgements

Firstly, I would like to thank my supervisors Dr. Alan Castle and Dr. Fiona Hunter for their guidance and encouragement throughout my studies. I would also like to thank my Dr. Bruce and Dr. Haj-Ahmad for their valuable input, suggestions, and service on my supervisory committee.

I would like to extend my thanks to all of my lab mates for all of their help, and for helping to make the lab a second home.

Lastly I'd like to thank my family and friends for all of their support through my graduate studies.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Insect silks in general have properties which lend themselves to many practical applications, and black fly silks specifically may possess more unique qualities yet to be discovered. Currently there is little information about the composition of black fly silks at the molecular level (the number, size, proportions, and properties of these proteins). This project seeks to provide a more complete understanding of the proteins present in the glands, which may help to guide future work into understanding the composition of black fly silk as well as its properties.

This thesis will attempt to determine the number and size of the proteins in the silk glands of *Simulium vittatum*, as well as to compare the silk protein profiles among simuliid species. This thesis will also aim to use previously determined gene sequence information and newly designed primers based on silk genes from chironomids (insects within the same superfamily) to determine the sequence of the silk gene(s) of *S. vittatum*.

1.1. Silk in Arthropods

The term silk is used to describe functional fibers that are spun from aqueous solutions and produced in a variety of glands in several arthropod species (Sutherland *et al*., 2010). Silk fibers are composed of silk fibroin (SF), the structural fibrous proteins, and silk sericin (SS), proteins that glue SFs together (Wotton, 2011; Jin & Kaplan, 2003)

Arthropods produce silk in three different types of glands: Malpighian, dermal, and labial. The majority of silk producing glands are labial, with Malpighian tubules being the least common silk producing gland (Sutherland *et al*., 2010). In spite of gland homology, it is not

necessary that silk proteins produced in homologous glands will themselves be homologous (Sutherland *et al*., 2010), i.e., not all homologous glands will produce homologous proteins. However, gland homology can be a useful criterion in the determination of protein homology (Sutherland *et al*., 2010). Labial glands, epidermal glands, and Malpighian tubules are present in a large variety of insects, and are not all responsible for producing silks. It is more likely that silk production evolved in existing insect glands rather than the formation of a novel silk production gland.

Silk has evolved independently within insects, with evidence suggesting that it has evolved up to 23 times in 17 different Orders (Sutherland *et al*., 2010; Sehnal & Sutherland, 2008). There is no apparent correlation between gland of production, ordered structure, and function of silk (Sutherland *et al*., 2010). Because of this lack of correlation between gland-structure-function it is highly unlikely that one major type of silk would evolve into another (Sutherland *et al*., 2010). It is more likely that a new silk type would evolve from a structurally similar non-silk gene. For example, it has been proposed that silk has evolved from cuticle proteins, which contain repeats of short amino acid motifs (Sehnal & Sutherland, 2008). Of those 23 evolutionary events, silk has evolved 3 times in Diptera: Culicomorpha, Bibionidae, and Brachycera (Sutherland *et al*., 2010). This literature review will focus on common aspects of insect silks and their potential application.

All insect silks share common features including: specialized glands that produce silk proteins, accumulation of silk proteins at high concentrations in the glands, proteins rich in alanine, serine and/or glycine, proteins that fold into secondary structure with high crystallinity, and mechanically strong and/or tough silk fibers (Sutherland *et al*., 2010). It has been suggested that aquatic silk is the most adsorptive biological material known (Wotton, 2011; Brereton,

1999). Aquatic silk tubes will adsorb detritus and become brown with adhesion of mineral fragments and organic matter (Wotton, 2011; Wotton & Hirabayashi, 1999).

Silk protein structures are largely dependent on amino acid sequences (Sutherland *et al*., 2010). Non-essential amino acids are used to allow for higher levels of silk production without having to compete with other physiological processes for resources (Sutherland *et al*., 2010). The majority of silks contain high proportions of alanine, serine, and glycine - likely due to the non-essentiality and partial hydrophobicity of these amino acids. Alanine, serine, and glycine are able to form hydrogen bonds with water and lend to the protein's solubility in the gland lumen; however, their solubility does not hinder the folding and eventual insolubility of the crystallized proteinaceous fiber (Sutherland *et al*., 2010).

Insect silk is semi-crystalline, and there are 5 different structures (Figure 1.1.) that have been identified; coiled coil, extended beta sheet, cross-beta sheet, collagen triple helix, and polyglycine II (Sutherland *et al*., 2010). In general, silk alpha helices have a 7 residue periodicity, with the first and fourth residues being hydrophobic, and other residues being polar or charged. Proline and glycine are generally not present. Alpha-helical secondary structures give rise to coiled coil tertiary structures (Sutherland *et al*., 2010).

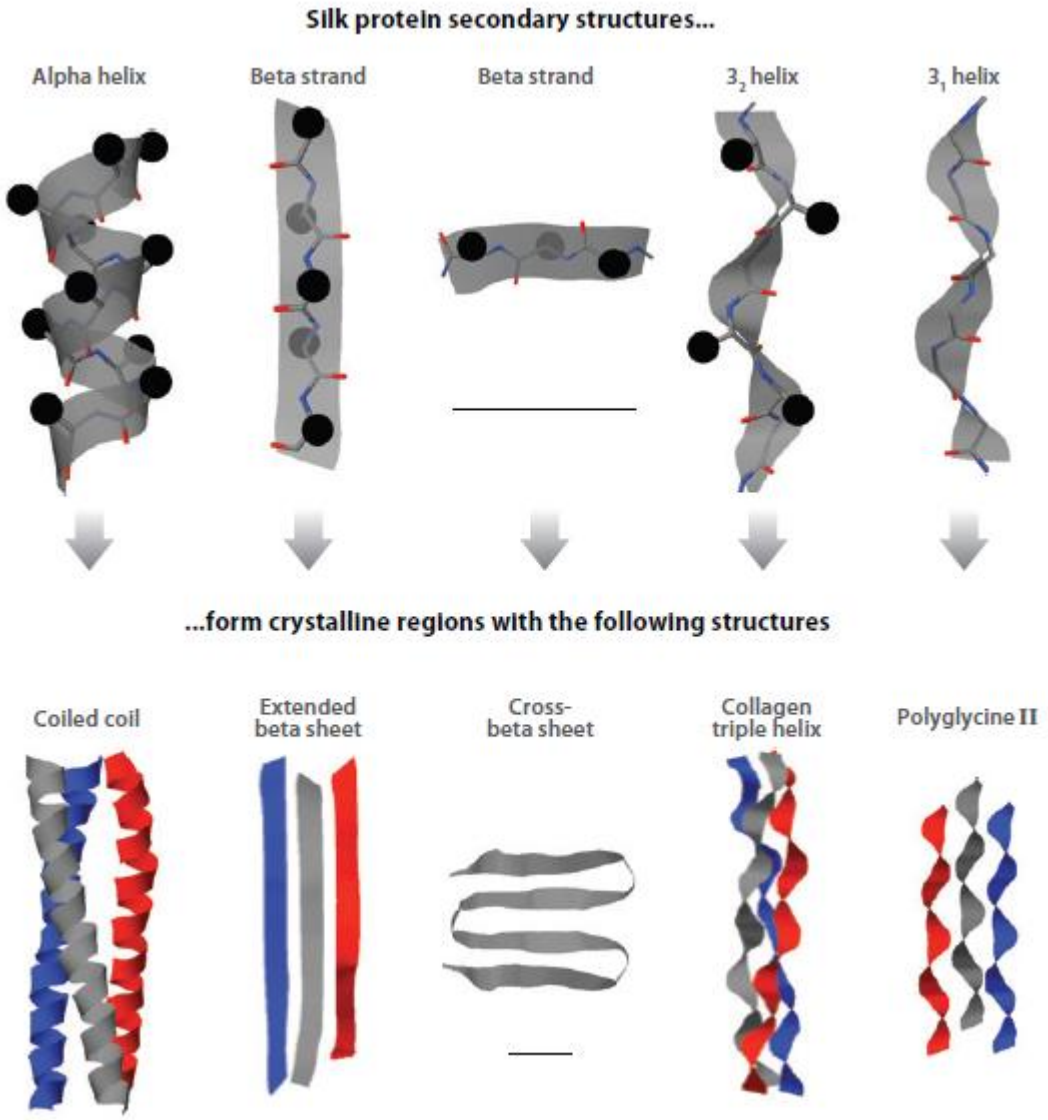

Figure 1.1. The five silk protein secondary structures from which the crystalline regions of the silk proteins are formed (Figure adapted from Sutherland et al., 2010)

Beta strands generally have a 2 residue periodicity, with alternating residues oriented on opposite sides of the strands, with glycine found at regular intervals where the protein folds back on itself. Beta strand secondary structures can give rise to either extended beta sheet or cross-beta sheet tertiary structures (Sutherland *et al*., 2010).

A collagen triple helix is formed by three individual protein chains, each forming a left-handed helix with three amino acids per turn. These three protein chains come together to form a

collagen triple helix tertiary structure. As the core cannot accommodate side chains larger than hydrogen, every third residue in the chain must be glycine (Sutherland *et al*., 2010).

Finally, the polyglycine II tertiary structure occurs when multiple protein chains formed of right-handed helices with three amino acids per turn come together. This structure cannot accommodate any side chains larger than glycine (Sutherland *et al*., 2010).

As it is unlikely that silks from one structure category have gradually evolved into another, it is theorized that silks of different structures have evolved independently (Sutherland *et al*., 2010). To transition from one structure to another, the intermediate structures would be non-functional. For that reason it is more likely that silks evolved from similarly structured proteins present in the insect instead of changing from one structure to another.

Spinning is the process of extruding an insoluble silk fiber from the aqueous protein solution inside the silk gland. Silk fibers are generally spun from aqueous solutions with high concentrations of silk proteins (Sutherland *et al*., 2010). During the spinning process proteins are aligned using shear force. As the solution is pulled apart the proteins organize themselves into fibers, until eventually the proteins are in a stable orientation, at which point they harden and form the final silk structure (Sutherland *et al*., 2010). Jin and Kaplan (2003) suggested that the proteins form micelles, with the hydrophilic regions forming the outer edges of the micelle. The formation of the micelles seems to be self-directed by the fibroins themselves. During the final steps of spinning, the physical shear elongates and aligns the globule-containing micelles and initiates the final structural transition to the insoluble crystalline beta-sheet.

1.1.1. Production of silk in non-aquatic arthropods

As previously mentioned, silk has evolved independently many times in insects. Although there are many general properties, each of the 23 silks evolutions also have resulted in unique properties. Some of these properties are examined below.

1.1.1.1. Empididae

Male dance flies (Diptera: Empididae) present silk-wrapped nuptial gifts to females, copulating with the female as she consumes the gift (Young & Merritt, 2003; Sutherland *et al*., 2007; Sutherland, 2008). The silk is produced in a dermal gland found in the tarsus, making the dance fly the only holometabolan to produce silk in a dermal gland (Sutherland *et al*., 2007). It is thought that the original function of the silk was to control the struggling prey, but now the silk has evolved into a stimulus role (Young & Merritt, 2003). Hilarine silk, named after the silk producing genus *Hilaria*, is produced from an independently evolved gene (Sutherland *et al.*, 2007). Like other silks, Hilarine silk is coded by a highly repetitive gene. The silk proteins are rich in asparagine (21.9%) and glycine (13.9%), with all other amino acids below 10%. Non-essential amino acids account for 67% of the protein. The gene for Hilarine silk is expected to be 6 kb, and the protein is approximately 220 kDa. The repeat blocks (short amino acid sequences repeated in tandem) are shorter than in other silks; however, this may be due to the fact that this silk does not need to be very strong (Sutherland *et al.*, 2007).

1.1.1.2. Lepidoptera

The silk worm *Bombyx mori* (Lepidoptera: Bombycidae) is the major supplier of the world's textile silks. Larval lepidopterans produce silk for shelter, pads for support during moulting, and durable cocoons for pupation (Sehnal & Sutherland, 2008). Lepidopteran silk glands terminate in structures specialized for silk secretion, including an internal silk press and an external spigot (Sutherland *et al*., 2010). The silk fibers are highly organized structures composed of two strands of a polymer of 3 major proteins, H-Fibroin (350 kDa), L-Fibroin (26 kDa), and P25 (30 kDa) (Inoue, 2000). *B. mori* silks are very large (fibroin is >300 kDa) and have highly repetitive sequences which fold into extended beta sheet crystalline structures (Zhou *et al.*, 2000; Sutherland *et al.*, 2007).

The proteins are secreted into the gland lumen at high concentrations, where they are enveloped in sericins (Sehnal & Sutherland, 2008). As has been previously mentioned, fibroin is the main component of silk fibers, while sericins act as the glue that holds the fibroin strands together (Tamada *et al*., 2004). Sericin constitutes 25-30% of silk fibers. The molecular weight of sericins ranges from 10 to over 300 kDa, with high serine (33.4%) and aspartic acid (16.7%) content (Zhang, 2002). There are as many as 5 different sericins, all of which act to either hold fibers together or to glue fibers to each other (Sehnal & Sutherland, 2008). The *ser1* gene has 9 exons, and 4-5 mRNAs. The central exon encodes 60 copies of a 38 amino acid repeat (Sehnal & Sutherland, 2008; Tamada *et al*., 2004). Sericin is removed during the degumming process for commercial silk production (Tamada *et al*., 2004).

Although all required proteins are present in the glands, the crystallization only takes place as a result of large shear stress occurring at the spinneret portion of the *B. mori* gland (Ki *et al*., 2007). The random coiled structure of the silks is more stable than the beta-sheet structure. Furthermore, the beta sheets, when stacked together, are more stable than the lone beta sheets.

The shear stress from the silk dope, the silk solution found within the gland, passing through the spinneret causes conformational changes in the proteins (from random coil to beta sheet). The beta sheets stack and form the nucleus upon which more beta sheets begin to aggregate (Zhang, 2002). The spinning process is slow because of the need to force the proteins through the spinneret to ensure the change from coils to sheets (Zhang, 2002). When viscosity is extremely high, flow slows down and can cause the extrusion to cease. This is why it is important to control viscosity (Ki *et al*., 2007).

Industrial silk processing requires that silks undergo degumming (removal of sericins) before the fibers are regenerated in order to improve the look and texture. Ki *et al.* (2007) showed that in the presence of sericins the regenerated fibers had 50% higher breaking strength, as well as improved tensile strength. It has also been shown that in the presence of sericins, fibroins crystallized at a slower rate (Lee, 2004). Fibroins transition from random coils to beta sheets when treated with methanol (Lee, 2004).

1.1.1.3. Hymenoptera

Hymenoptera produce silk for cocoons in their labial glands (Sehnal & Sutherland, 2008). Hymenopteran silks have beta sheet structures which have evolved in the common ancestor, while coiled coil, polyglycine and collagen structures are thought to have evolved later in the phylogeny (Sehnal & Sutherland, 2008). There are many different types of silks found among hymenopterans, and they differ in size and function among taxa. For example, parasitic wasp silk proteins are large and highly repetitive (Yamada *et al.*, 2004; Sutherland *et al.*, 2007), whereas honeybees have small (~30 kDa) non-repetitive proteins (Sutherland *et al.*, 2006; Sutherland *et*

al., 2007). Unlike other hymenopteran silks, honeybee silk proteins adopt a coiled structure (Rudall, 1962).

1.1.2. Production of silk in aquatic insects

1.1.2.1. Trichoptera

Caddisfly larvae (Order Trichoptera) live in aquatic environments. They use silk to stick together pieces of plants and other materials into protective cases, and to close the case at the end of larval development (Sehnal & Sutherland, 2008). Variations in case structure and building components are characteristic of different genera (Wotton, 2011). Caddisfly larvae may also use their silk to create catching nets, hiding tubes and small domes from sand grains (Sehnal & Sutherland, 2008). Caddisfly silk glands are similar to lepidopteran silk glands in their organization (Sehnal & Sutherland, 2008). Much like lepidopterans, caddisfly silks are composed of H- and L-fibroin homologues. There has also been evidence of sericin-like materials in caddisfly silk. However the caddisfly silks differ from lepidopteran silk in that they do not contain the P25 protein (Sehnal & Sutherland, 2008). The proteins are all highly conserved with regards to the spacing of residue properties such as hydrophobicity and charge, as well as the spacing of cysteines (Sehnal & Sutherland, 2008).

1.1.2.2. Chironomidae

Chironomid larvae live in flowing water and use silk to construct larval tubes which they use for both protection and food capture, and eventually the silks are used in the construction of a

pupal cocoon (Case & Weislander, 1992). The secreted polypeptides (silk) are initially soluble in the gland, but as they pass through the salivary duct, they undergo a dramatic transformation that results in their polymerization into a long, elastic fiber (Kao & Case, 1985).

The silk secretory proteins are encoded by a multigene family known as the Balbiani Ring multigene family (Case & Weislander, 1992). The labial gland secretory cells contain giant polytene chromosomes, which contain Balbiani Rings (large chromosomal puffs in areas of high transcriptional activity) after which the gene family is named (Case & Weislander, 1992). The multigene family can be separated into 3 classes based on protein size: large (>500 kDa), intermediate (between 100 and 500 kDa), and small (<100 kDa) (Rylander & Edstrom, 1980; Hertner *et al.*, 1980; Case & Weislander, 1992).

The large size class is composed of 4 genes known as the spI proteins, spIa, spIb, spIc, and spId (Rylander & Edstrom, 1980; Hertner *et al.*, 1980; Kao & Case, 1985; Case & Weislander, 1992). The spIs are composed of tandem copies of distinctive core repeats, each core repeat containing 60-90 amino acids within two distinct regions, the constant region (C) and the subrepeat region (SR).

The C region has a core repeat of 30 residues with invariant cysteine (4 residues), methionine (1 residue), and phenylalanine (1 residue) at specific locations. The repeat is predicted to give 18-20% alpha-helix and 7% beta-turn (Case & Weislander, 1992)

The SR region is composed of a series of smaller repeats (subrepeats), and can be variant in length. The SR region contains a “+ Proline -” (Basic-Proline-Acidic) tripeptide motif, in which the basic residue tends to be Arginine or Lysine, and the acidic residue tends to be Glutamic acid, phospho-Serine or phospho-Tyrosine. This region is predicted to yield a collagen-like structure (Case & Weislander, 1992).

All four spIs share a huge central domain of 130-150 core repeats flanked by two globular domains. The amino termini are all about approximately 200 residues long and fold into a globular domain. The carboxyl termini are each about 211 amino acids, and are more well conserved than the central core repeats (Case & Weislander, 1992).

Although there are 4 spIs, larvae typically only synthesize two spIs, spIa and spIb (Rylander & Edstrom, 1980; Hertner *et al.*, 1980; Kao & Case, 1985). spIa and spIb are known to be glycosylated (Hertner *et al.*, 1983) and phosphorylated almost exclusively at serine residues (Galler *et al*., 1984; Kao & Case, 1985). The addition of galactose or certain other monosaccharides to the larval diet stimulates the synthesis of spIc and inhibits spIb and spId synthesis through a mechanism involving the levels of cellular phosphate (Edström *et al.*, 1980; Edström *et al.*, 1982; Kao & Case, 1985). It was therefore hypothesized that spIb and spId are within close proximity to each other on the chromosome (Kao & Case, 1985).

The intermediate size class contains 4 proteins: sp195, sp185, sp140, and sp115 (Case & Weislander, 1992). Protein sp195 is composed of about 50 tandemly arranged copies of a 25-residue repeat, with 2 conserved cysteine residues at specific locations, and one copy of the + Proline - motif (Case & Weislander, 1992). Protein sp185 is composed of 50-60 tandem copies of highly diverged repeats of hydrophilic amino acids, with a conserved Cys-X-Cys-Y-Cys motif every 22-26 residues, and sporadic copies of the "+ Proline -" motif (Case & Weislander, 1992). Unlike previously mentioned secreted proteins, sp140 is not associated with the Balbiani Ring. It contains 65 copies of a 14-residue repeat rich in Lysine and Glycine residues, as well as the "+ Proline -" (Case & Weislander, 1992).

The small size class contains 5 proteins: sp65, sp55, sp40, sp39, and sp38 (Case & Weislander, 1992). In both sp40 and sp38, the central portion contains six Cysteine residues with

spacing similar to that of the C region in the spIs. The two proteins are very similar with regards to the composition of their repeating regions that it is likely the genes diverged from a duplication event (Case & Weislander, 1992).

The length of repeating units can vary considerably, from 14 to 90 residues, with larger proteins tending to have longer repeats (Case & Weislander, 1992). There are two expression classes, the larval class and the prepupal class, which are used to classify which proteins are present during each life stage. Most proteins are present during both, however sp40 is only present in the larval stage, and sp195, sp140, and sp38 are only present in the gland during the prepupal stage (Case & Weislander, 1992). Purified spIs form complexes similar to secretions, making it likely that spIs are the main component of a self-assembling fibrous backbone (Case & Weislander, 1992).

Cysteine residues within the spIs represent a site for propagation of both covalent intraprotein and protein-protein interaction; the C regions lack the heptad repeat which is characteristic of alpha-helical coiled coils (Case & Weislander, 1992). The proteins within the gland likely form a seemingly random web of proteins while at rest in the gland. However, when the proteins are pushed toward the duct for secretion, the proteins are likely to experience shear stress causing alignment into their final fiber structure (Case & Weislander, 1992). The roles of the intermediate and small proteins are still unknown; however, they may play a role in the crosslinking of the spIs once they have reached their proper alignment (Case & Weislander, 1992). These proteins may act as sericins do in the lepidopteran silk, functioning to both coat the fiber, but also to retard the crystallization of the silk in order to allow the larva to manipulate the silk into its final shape. With respect to the other gland secreted proteins, spIs have unique 5’ and 3’ regions (Sehnal & Sutherland, 2008). The C region is predicted to fold into alpha-helices and

the SR region is predicted to fold into an extended poly-glycine-II type helix (Sehnal & Sutherland, 2008).

In 1998, Berezikov *et al.* described a novel protein, ssp160, which was exclusively expressed in the special lobe of the silk gland of *Chironomus thummi*. The silk has an apparent molecular mass of 160 kDa; however, over half of its apparent mass is due to extensive N- and O-linked glycosylation. The protein lacks the extensive array of tandem repeats and Cys- and Pro-containing motifs. The gene consists of 6 exons separated by 5 introns, with 70% of ssp160 being encoded in exon 3. The gene sequence is not overtly similar to the other chironomid silk genes. However the exon/intron organization is very similar to genes which encode extensive arrays of tandem repeats.

A common feature of all secretory proteins is the presence of either a conserved pattern of Cysteine residues or "+ Proline -" motif, or both (Case & Weislander, 1992). The spI genes are all very closely related to each other and have likely evolved by translocative duplications followed by changes in the repeats within each gene (Case & Weislander, 1992). The non-large size proteins share common structural units. For example, non-large proteins either contain multiple repeats, or they represent a prototype for a core repeat (Case & Weislander, 1992). Case & Weislander (1992) speculate that all genes encoding secretory proteins may have evolved from a common ancestor, which would have been a small protein which would have contained at least one cysteine residue and the + Pro - motif. Reduplication events would have amplified parts or all of the ancestor sequence, resulting in repeats (Case & Weislander, 1992).

Based on their phylogenetic proximity, Sutherland *et al.* (2010) speculated that all silks produced by organisms in the Chironomoidea superfamily (specifically Chironomidae and Simuliidae) evolved from a common ancestor.

1.1.2.3. Simuliidae

Black fly (Family Simuliidae) larvae produce silk for purposes including safety lines, anchorage pads, and pupal cocoons. Simuliid larvae produce large amounts of silk in a large pair of labial glands.

Simuliid larvae can use their thoracic proleg to pull silk from the mouthparts and apply it to the substrate, where it hardens into a silk pad. The proleg can then be used to temporarily attach to the pad while the abdomen is brought forward and attaches to the pad (Barr, 1984). Hooks in the thoracic proleg are smoother and are meant to be inserted into the silk as a temporary means, while hooks in the abdominal proleg have tubercles which will help to keep the hooks inserted in the silk. This allows the hooks to remain embedded into the silk even when it loses its stickiness and elasticity (Barr, 1984). Larvae spend little to no energy maintaining attachment to silk, as can be seen when dead larvae and exuvium remain attached (Barr, 1984). The larvae have to spend energy to contract the abdominal circlet in order to lift the tubercles to allow the hooks to slide out (Barr, 1984).

Safety lines produced by black flies are only 3 micrometers thick, and therefore must be exceptionally strong in order to support against the forces pushing the larva downstream (Wotton, 1986). Black fly silk safety lines, pads, and cocoons are made from the same labial gland (Kiel & Röder, 2002), but the protein composition of the secretions from the glands likely changes as the organism transitions into different life stages

Black fly silks are highly adsorptive, which leads to the quick colonization by bacteria. This bacterial colonization could have been used in trapping and processing nutrients, similarly to a process observed in chironomids and trichopterans (Ings, 2010; Wotton, 2011). In his thesis,

Barr (1982) found that *S. vittatum* silk was composed of 16 amino acids, with a high proportion of glycine (34%). These results, however, do not seem to fit the current model of silk composition. Silks in general are known to have a high proportion of glycine, serine, and alanine, as well as well-conserved cysteine residues. As these results (Barr, 1982) show 0% cysteine in the silk, it has to be questioned whether Barr's composition analysis was accurate.

Three proteins were found to be major constituents of the secretion, with masses 70, 40 and 20 kDa (Kiel & Röder, 2002). The cocoons were found to be composed of the same 3 proteins as the larval silks (Kiel & Röder, 2002).

1.2. Practical Applications

1.2.1. Biomaterials

Biodegradable materials can be made by making a copolymer with sericins and other resins. High mass sericin peptides are used in degradable biomaterials, compound polymers, functional biomembranes, hydrogels, and functional fibers (Zhang, 2002). Since silk sericin is susceptible to action of proteolytic enzymes, this makes sericin and its polymers digestible and biodegradable (Padamwar & Pawar, 2004). Sericin polymers can be polymerized with other resins to create hydrogels as well as membranes. For example, polyvinyl alcohol (PVA) and sericin can yield cross-linked hydrogels (Padamwar & Pawar, 2004) or durable functional membranes that can withstand heat and strain (Zhang, 2002).

Teramoto *et al.* (2005) described sericin hydrogels as flexible materials that can be easily prepared without chemical crosslinking or irradiation, making it suitable as a drug carrier or scaffold for tissue engineering. They also found that using denatured sericin resulted in

formation of a weaker hydrogel as compared to hydrogels made of non-denatured sericins (Teramoto *et al.*, 2005). Sericin is also sensitive to chemical modification (Teramoto *et al.*, 2004), making it possible to modify sericins to add specific properties. It was also discovered that the hydrogels can be dehydrated and rehydrated, and that prepared hydrogels can be stored in ethanol for months (Teramoto *et al.*, 2005). A copolymer material made of silk sericin and polyethylene glycol diglycidyl ether (PEG-DE) can be frozen and then freeze dried to create a porous material (Tao *et al.*, 2005).

Cast films of pure silk fibroin were stable in water and swelled to form a hydrogel. Furthermore, upon stretching the hydrogel (previously hydrated from a film), the water was expelled and the gel became a film (Jin and Kaplan, 2003). Silk films have adequate flexibility and tensile strength, as well as infection resistant properties (Padamwar & Pawar, 2004).

1.2.2. Medical Applications

Lower mass sericin peptides are used in cosmetics, health products and medications (Zhang, 2002). The high serine content in sericin gives it good skin moisturizing properties (Aramwit & Sangcakul, 2007). The gelling properties, moisture retention, and skin adhesion of silk fibroins and sericins are all qualities that contribute to its potential applications (Padamwar & Pawar, 2004).

Sericins have properties that lend themselves to many biomedical applications including inhibition of lipid peroxidation and tyrosinase activity (Kato, 1998), high cryoprotective effects for cells and proteins (Tsujimoto, 2001), and mouse skin tumorgenesis suppression by protection against oxidative stress, inhibition of inflammatory responses, and inhibition of tumor necrosis

factor-alpha (Zhaorigetu, 2003). Sasaki *et al.* (2004) have also shown silk sericins to suppress colon carcinogenesis in mice.

After undergoing sulfation, sericin has anticoagulant properties (Tamada *et al*., 2004), with higher mass sericins having stronger anticoagulant properties than those with lower mass (Tamada *et al*., 2004).

When treated with an 8% sericin cream, mouse wounds were found to heal faster than the control group (Aramwit & Sangcakul, 2007). In addition, the sericin cream showed no ulceration and no inflammatory responses. In contrast, the cream base control treatment (cream without the addition of sericin) showed some ulceration and acute inflammatory responses (Aramwit & Sangcakul, 2007). Sericins show great promise as membrane components as they have been shown to induce accelerated proliferation of mammalian cells (Terada *et al.*, 2002; Teramoto *et al.*, 2004; Tsubouchi, 2005). Membranes composed of sericin and fibroin proved to be effective substrates for the proliferation of adherent animal cells (Zhang, 2002). For example, films of pure sericin proteins permitted cell attachment and growth comparable to that of collagen films (Zhang, 2002). Inflammatory response of cells can be stronger on collagen compared to silk (Meinel *et al*., 2005). In addition, sericin films have excellent oxygen permeability (Zhang, 2002). In light of their properties these sericin/fibroin films could be ideal for use in the production of products such as contact lenses, highly elastic artificial blood vessels, and other such prostheses (Zhang, 2002).

One of silk's notable properties is its ability to adsorb organic compounds. In 2007, Bayçin *et al*. were able to show that *B. mori* silk was able to selectively adsorb the polyphenol antioxidants from olive leaf extracts. The adsorbed polyphenols retained their antioxidant and antimicrobial properties. *B. mori* silk fibroin on its own also showed antioxidant activities

(Bayçin *et al*., 2007). Given its ability to adsorb antioxidant polyphenols, these silks could be used to help supplement patients with natural antioxidants rather than having to rely on synthetic sources.

1.2.3. Filter Applications

The ability to adsorb organic compounds could be useful in a number of applications. One such application is the use of silk membranes in water filtration systems. Silk sericins can be made into membranes by polymerization with many different materials that can be used to filter and separate compounds from water (Zhang, 2002). Black fly larval silk has been shown to be highly adsorptive to organic compounds, showing a high level of adsorption of organophosphorous pesticides (Brereton *et al*., 1999). This high level of adsorption is more likely attributed to properties of the silk as opposed to the hydrophobic nature of the organic compounds (Brereton *et al*., 1999). A polymer made of sericin and urethane could be used to make a reusable membrane that can remove water from a water/alcohol mixture (Mizoguchi *et al.*, 1991; Zhang, 2002). These adsorptive properties could make silk a useful component in water filtering and water polishing processes, helping to bind organic compounds and remove them from the water.

1.2.4. Phylogenetics

Phylogenetic analysis of genes involves the study of genetic divergence of one of more specific genes among species. The two main criteria for a useful molecular marker are that the gene must be well conserved enough to compare to other species, and conversely there must be

enough variability within the gene to distinguish among species. In many insects the mitochondrial gene cytochrome c oxidase subunit 1 (COI) is used for DNA barcoding. However, the phylogenies done using mitochondrial DNA (mtDNA) might only reflect the evolution of mtDNA (Makarevich, 2000). Makarevich *et al.* (2000) used two nuclear genes to create phylogenies, one gene being the ssp160 silk gene found in chironomids. It was found that the ssp160 sequences differ from 1.7-22.8%, which would make these silks genes useful tools for the construction of molecular phylogenies. Given that silk proteins are necessary for survival, the genes encoding silk proteins are likely to be well conserved, which could make them useful in the construction of molecular phylogenies.

2.1. Aims of the thesis

This thesis aims to determine the number and size of the silk proteins in *S. vittatum* silk glands. These data will help to better understand the protein composition of black fly silk, which may in turn help in future work to understand the properties of black fly silks.

This thesis will also examine the difference between larval and pupal silk proteins of *S. vittatum* in order to determine any differential protein expression. This could give insight about which proteins are used in the production of stage specific silks such as pupal cocoon silks.

In addition, the silk gland protein profiles of different species of black fly will be examined in order to find any differences among species. These differences may give some explanations about differences among genera.

Lastly, this thesis will attempt to determine more genetic information about *S. vittatum* silk. These gene sequences will give insight to the protein sequences that could be valuable for determining protein properties such as folding and conserved domains. Furthermore, this genetic

information would allow for the comparison of black fly silk with other silks at the gene and protein sequence level.

Chapter 2: Proteins present in simuliid silk glands

2.1. Introduction

As previously discussed in Chapter 1, silks possess properties that can lend themselves to many practical applications, and black fly silks in particular remain an untapped resource in this respect. In order to be able to apply these properties, we must first have a better understanding of what these silks are composed of. This chapter will attempt to identify properties of these silk proteins, including size, number, and modifications of the proteins involved.

Phylogenetically, simuliids have been placed in the superfamily Chironomoidea (Figure 2.1.1.). Within the superfamily Chironomoidea there are 4 families; Thaumaleidae, Simuliidae, Chironomidae, and Ceratopogonidae. Silk production is a common feature among chironomids, simuliids and thaumaleids, and thus the production is thought to have been present in a common ancestor (Sutherland *et al*., 2010). Apparently, the ceratopogonids have lost the ability to produce silks.

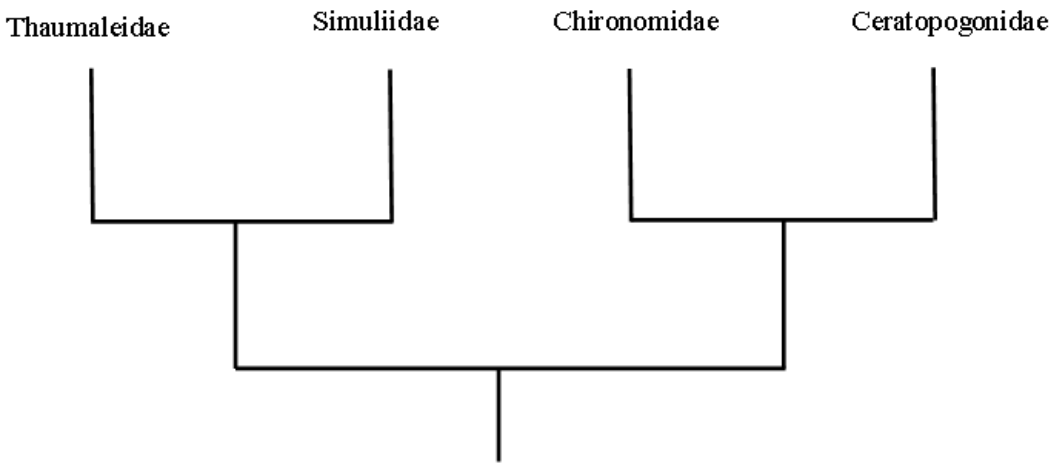

Figure 2.1.1. Phylogeny of the families within the superfamily Chironomoidea (adapted from Craig *et al*., 2012; Wood and Borkent, 1989).

Given that silk is essential for the survival of larval black flies, it is likely that the corresponding genes are well conserved across the family. Examining silk protein composition across different genera (Figure 2.1.2.) may give insight into the divergence and conservation of these proteins. This may also show common elements of black fly silk that could suggest which proteins are essential to the composition of silk.

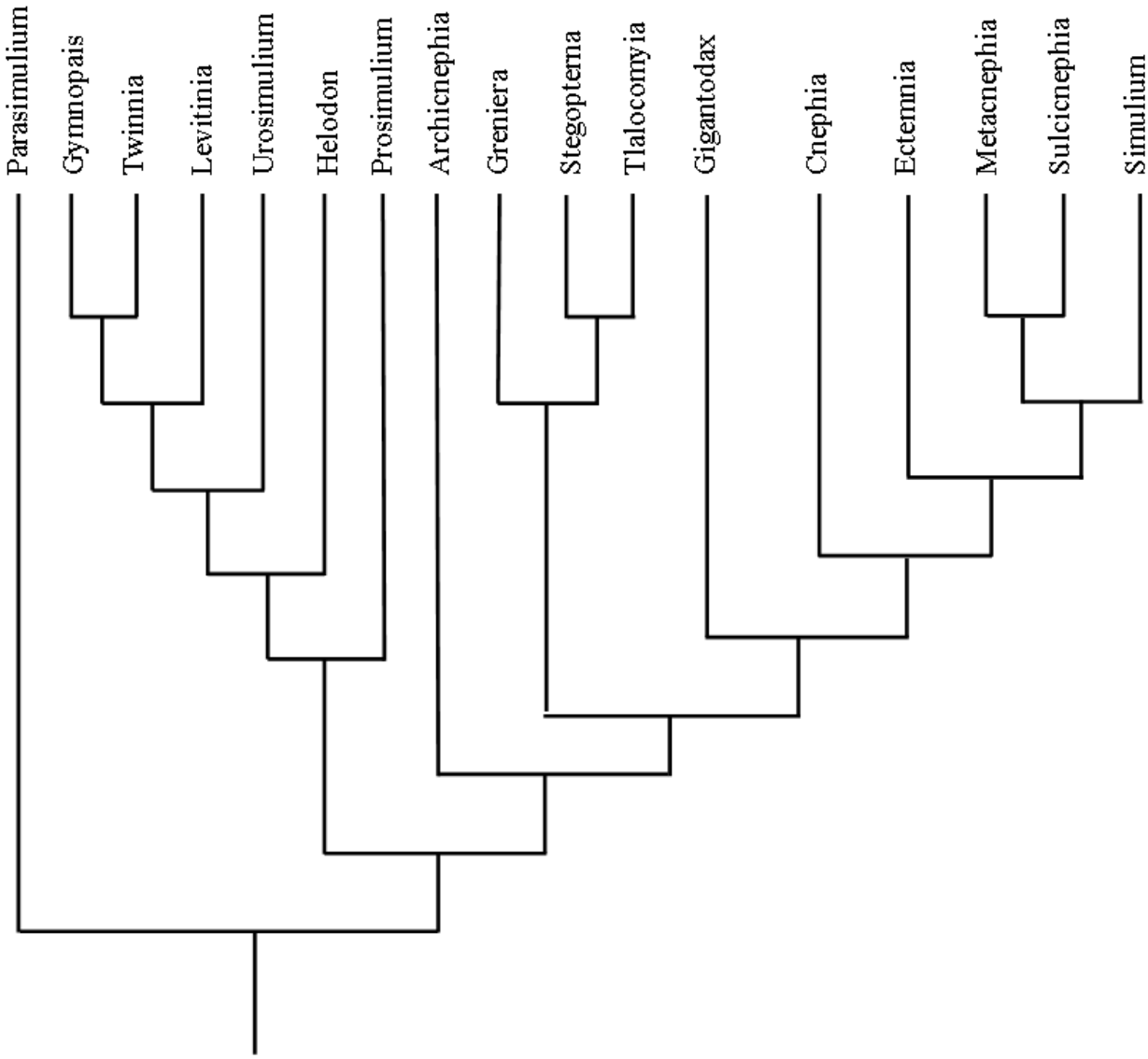

Figure 2.1.2. Phylogeny of the genera of black flies in North America (Adapted from Adler *et al*., 2004)

The aims of this chapter are to determine the protein composition of silks by examining both cocoons as well as the contents of the silk glands. Furthermore, larval and pupal protein profiles

will be compared to determine any differential protein expression between the two life stages. Additionally, protein profiles from different species across genera will be examined to determine common proteins within the family. These findings will help determine the composition of black fly silks, which will lend to future research about the mechanisms and applications of black fly silk protein properties.

2.2. Materials and Methods

2.2.1. Cocoon protein preparation

Dried empty *S. vittatum* cocoons were obtained from Dr. Fiona Hunter's lab. Dried cocoons were weighed and 10mg of cocoon material was transferred into a clean 1.5 mL microfuge tube and boiled for 2 min in 2X reducing protein sample buffer (100 mM Tris HCl, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200mM beta-mercaptoethanol). Cocoons were then rinsed with milli-Q water 5-6 times (until the Bromophenol blue had washed away). 0.5 mL of 99% formic acid was added to the cocoons and incubated for 1.5 h at 4°C. The samples were added to 40 volumes of water and lyophilized. The protein powder was reconstituted in 50 µl 100mM Tris pH 8 (Hunter & Bidochka, 2003).

2.2.2. SDS-PAGE

Glass plates were assembled in the casting clamp apparatus (BioRad Mini-PROTEAN). Plates were marked 1cm below the comb to indicate when to stop pouring the resolving gel. Resolving gel (Table 2.2.2.1) was poured between plates to the previously made mark, overlaid with 2-propanol to create an equal surface, and allowed to polymerize for 1 h. Alcohol was washed away with water, and the gel surface was dried with filter paper. Stacking gel (Table 2.2.2.1) was poured on top of the resolving gel and the comb was inserted. The gel was allowed to polymerize for 45 min. The comb was removed and the gels were assembled into the running apparatus and placed in the electrophoresis tank (BioRad Mini-PROTEAN Tetra Cell).

Electrophoresis buffer (10X) (250 mM Tris, 1.920 M Glycine, 1% SDS) was diluted to 1X and added to the core between the gel plates to a volume that came between the lower and upper plates. Electrophoresis buffer was also added to the gel tank to submerge the bottom of the gel.

Protein sample (13µL) was mixed with 1 volume of 2X reducing protein sample buffer and samples were heated to 95°C for 5 min. Samples were loaded into the individual wells and the gels were run at 200V until the loading dye reached the bottom of the gel. Gel was removed from the sandwich apparatus and the gel was stained (Simpson *et al*., 2009).

Table 2.2.2.1. Stacking and resolving 4%, 7%, and 12% polyacrylamide gels.

	Stacking Gel	Resolving	
	4%	7.5%	12%
30% Acrylamide	1.33	2.5 mL	4 mL
0.5M Tris pH 6.8	2.5mL	-	-
1.5M Tris pH 8.8	-	2.5mL	2.5mL
10% SDS	100µL	100µL	100µL
Milli-Q Water	6.1mL	4.85	3.35mL
TEMED	10µL	5µL	5µL
10% APS	50µL	50µL	50µL
Total Volume	10mL	10mL	10mL

2.2.3. Coomassie Brilliant Blue Stain

Gels were gently agitated in Coomassie blue stain (0.1% (w/v) Coomassie Brilliant Blue, 50% (v/v) methanol, 10% (v/v) acetic acid) for 30-45 min, followed by agitation in destain (40% (v/v) methanol, 10% (v/v) acetic acid) until desired background was achieved.

2.2.4. Periodic acid Schiff Stain

Gel was gently agitated overnight in 40% ethanol/5% acetic acid, followed by 2 h in periodic acid solution (0.7% (w/v) periodic acid, 5% (v/v) acetic acid). The gel was washed twice in 5% acetic acid, for ten minutes each wash. The gel was gently agitated for 2 h in 200 mL of sodium metabisulphite solution 1 (0.2% (w/v) sodium metabisulphite, 5% (v/v) acetic acid), followed by two ten minute washes in 5% (v/v) acetic acid. The gel was incubated in Schiff's reagent until colour developed, followed by a 90 min incubation at 55°C in 200 mL of metabisulphite solution 2 (0.2% (w/v) sodium metabisulphite, 40% (v/v) ethanol, 5% (v/v) acetic acid). The gel was gently agitated in 200 mL of destain (40% (v/v) ethanol, 5% (v/v) acetic acid). Methods adapted from Segrest and Jackson (1972).

2.2.5. Sample Collection

Simulium vittatum larvae and pupae were frozen in water and shipped on ice from the black fly colony at University of Georgia (Athens, Georgia) courtesy of Elmer Gray. Specimens were stored at -20°C, and thawed before dissection.

Species were collected from sites with flowing water (See Table 2.2.5.1) across Algonquin Park between May 4th and May 18th 2013. Organisms were identified to species using the keys found in The Black Flies (Simuliidae) of North America (Adler *et al*., 2004), and sorted into larvae (absence of histoblasts) and pharate pupae (presence of histoblasts).

Table 2.2.5.1. Sites visited in and near Algonquin Provincial Park, and species found at each site.

Site Name	Site Description	Species found
Davie's Bog	Fast rocky culvert outlet	*Simulium venustum/verecundum*
Madawaska River	Slow flow, grassy areas	*S. venustum/verecundum* *Simulium tuberosum*
Highland Trail Outlet	Slow beaver dam	*Simulium croxtoni*
Broadwing Creek	Slow creek	*Stegopterna diplomutata*
Pasquale Creek	Quick creek	*Stegopterna mutata* *S. venustum/verecundum* *Prosimulium fontanum*
Highway 11 – Gryffin Lodge Road[1]	Trickle	*Greniera abditoides*
Costello Creek	Quick rocky creek	*Cnephia dacotensis*

2.2.6. Silk Gland Dissection

Larvae/pupae were placed under a dissection microscope (Leica MS 5) and handled with fine forceps. While holding the head capsule, a small superficial incision was made just below the head capsule. Holding the head capsule firmly with the fine forceps, the abdomen was grasped with a second pair of fine forceps and the body was pulled away from the head. The glands and midgut remained attached to the head as the abdomen was pulled away. Silk glands were quickly separated from any extra materials (midgut, fat body, and cuticle) and placed in

[1] This site is outside the Park.

100µL extraction buffer in a 1.5mL microfuge tubule; the pair of glands from a single larva were placed in each tubule.

2.2.7. Silk Gland Protein Extraction

The protein extraction protocol was designed based on the protocol in Kao *et al.* (1985). Two denaturants were used to make up protein extraction buffers: guanidine HCl extraction buffer (6 M Guanidine Hydrochloride, 40 mM Tris-HCl (pH 8.8), 20 mM EDTA, 100 mM β-mercaptoethanol) and urea extraction buffer (8 M urea, 40 mM Tris-HCl (pH 8.8), 20 mM EDTA, 100 mM β-mercaptoethanol). Samples were incubated on ice for 2 h, with occasional vortexing. Samples were centrifuged for 5 min at 12,000 *g*, and the supernatant was transferred to a new 1.5 mL microfuge tube.

Nine volumes of 80% acetone was added to the buffer, followed by 1 h of incubation at -20°C, centrifugation for 3 min at 12 000 *g*, and resuspension in 10 mM Tris-HCl (pH 7.2). Early observation indicated the resulting pellet from the centrifugation to be insoluble in the Tris buffer. As a result, the proteins were not precipitated using acetone and protein samples used remained in the urea buffer.

2.2.8. SDS-PAGE of Precast Gradient Gels

Protein samples were mixed with 2X protein sample loading buffer in a 1:1 ratio, and samples were then heated to 95°C for 5 min. After cooling back to room temperature, samples were loaded into 4-20% Mini-Protean TGX Precast Gels (Bio-Rad Catalogue # 456-1096). Electrophoresis conditions were as follow: 200V were applied to the system until the smallest

prestained ladder protein was a few millimeters from the bottom edge of the gel, at which point electrophoresis was stopped and gels were ready for staining.

2.2.9. Silver Staining

Gels were silver stained using the Bio-Rad Silver Stain Plus Kit (catalogue # 161-0449). The manufacturer's protocol was followed and gels were stained until colour developed. Development was stopped by gentle agitation in 5% acetic acid for 20 min. The gels were then gently agitated for 10 min in distilled water to prepare for imaging.

2.2.10. Glycoprotein Staining

In order to save time, gels were stained for glycosylation using the Pierce Glycoprotein Staining Kit (#24562). The manufacturer's protocol was followed; however, it was found that when incubation times were doubled, the staining was more visible.

2.2.11. Protein Gel Imaging and Band analysis

Stained gels were imaged using the Bio-Rad Gel Doc EZ Imager and images were imported into ImageLab 4.1 software. Using the software bands were auto-detected using a custom sensitivity of 50. Detected band boundaries were manually focused to the edges of each band. Molecular mass was measured in kiloDaltons (kDa) using the Thermo Scientific Spectra Multicolor High Range Protein Ladder (catalogue # 26625) as a standard, extrapolating from the standard curve to determine the mass of large proteins. Proteins were named based on the

position of the bands in order to aid in matching of similar proteins and to help with calculating mean and standard deviation.

Mean masses were calculated by taking the average of the molecular mass of each protein given the same label based on position.

2.3. Results

In order to determine silk protein composition, silk samples from cocoons, larval glands and pupal glands were dissolved and separated using SDS-PAGE.

2.3.1. Simulium vittatum cocoon proteins

In order to determine the composition of cocoon silk proteins, cocoons were dissolved using either formic acid or urea. Separation using SDS-PAGE (Figure 2.3.1.1.) showed the presence of 2 proteins among all 3 samples, with masses of 64 and 56 kDa.

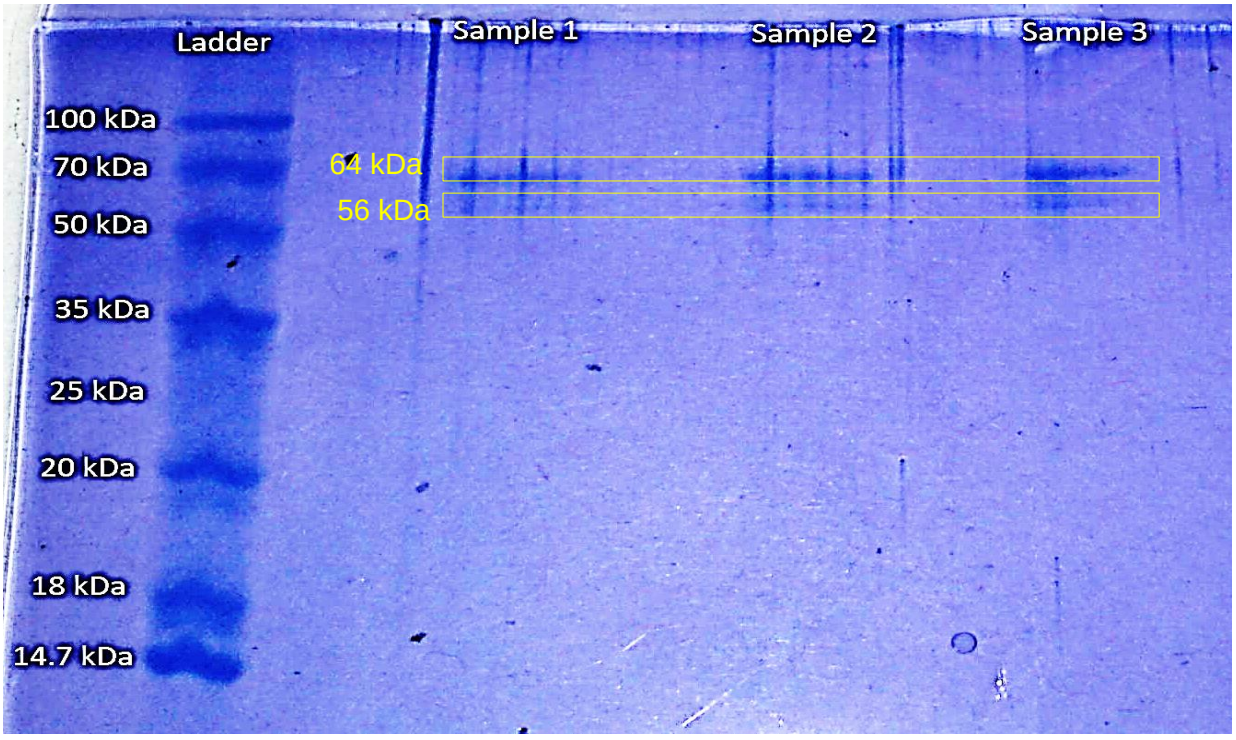

Figure 2.3.1.1. Coomassie blue stained 12% polyacrylamide gel of *S. vittatum* cocoon proteins. Ladder used was Norgen ProteoLadder 150. Samples 1 and 2 were from cocoon proteins dissolved using formic acid. Sample 3 was from cocoons dissolved using 8M urea. All 3 samples had 56 kDa and 64 kDa proteins.

2.3.2. Simulium vittatum larval silk gland SDS-PAGE

To determine the composition of larval silk proteins, larval silk gland proteins were extracted with urea and Guanidine HCl and separated by SDS-PAGE. Although the extraction buffer using Guanidine HCl yielded much higher quantities of proteins, the bands in the gels were warped and each run (n=4) resulted in inconsistent banding patterns. The urea extraction buffer resulted in focused, thinner, more reproducible banding patterns. For this reason urea buffer was used for all subsequent work.

The silver stained 4-20% polyacrylamide gel (Figure 2.3.2.1) with 8 individual larval silk gland samples was analyzed using the ImageLab 4.1 software. Bands were generated and annotated based on their position relative to other proteins within the same and different lanes (Figure 2.3.2.2).

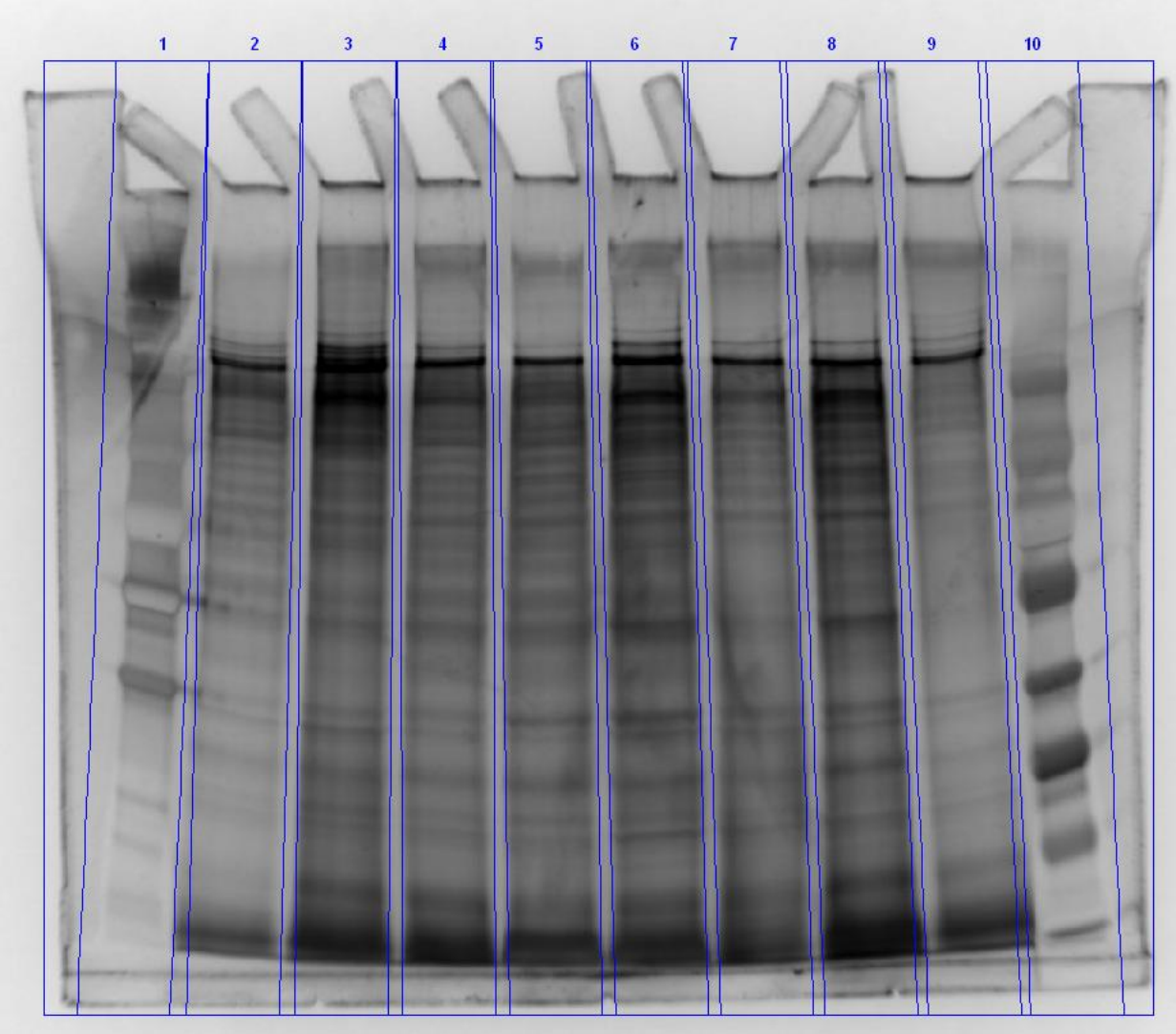

Figure 2.3.2.1. Silver stained 4-20% SDS-PAGE of S. vittatum larval silk gland proteins. Lane 1 is the Spectra Multicolor High Range Protein Ladder (Thermo Scientific). Lanes 2-9 are larval silk gland proteins. Lane 10 is the Spectra Multicolor Page Ruler Protein Ladder (Thermo Scientific).

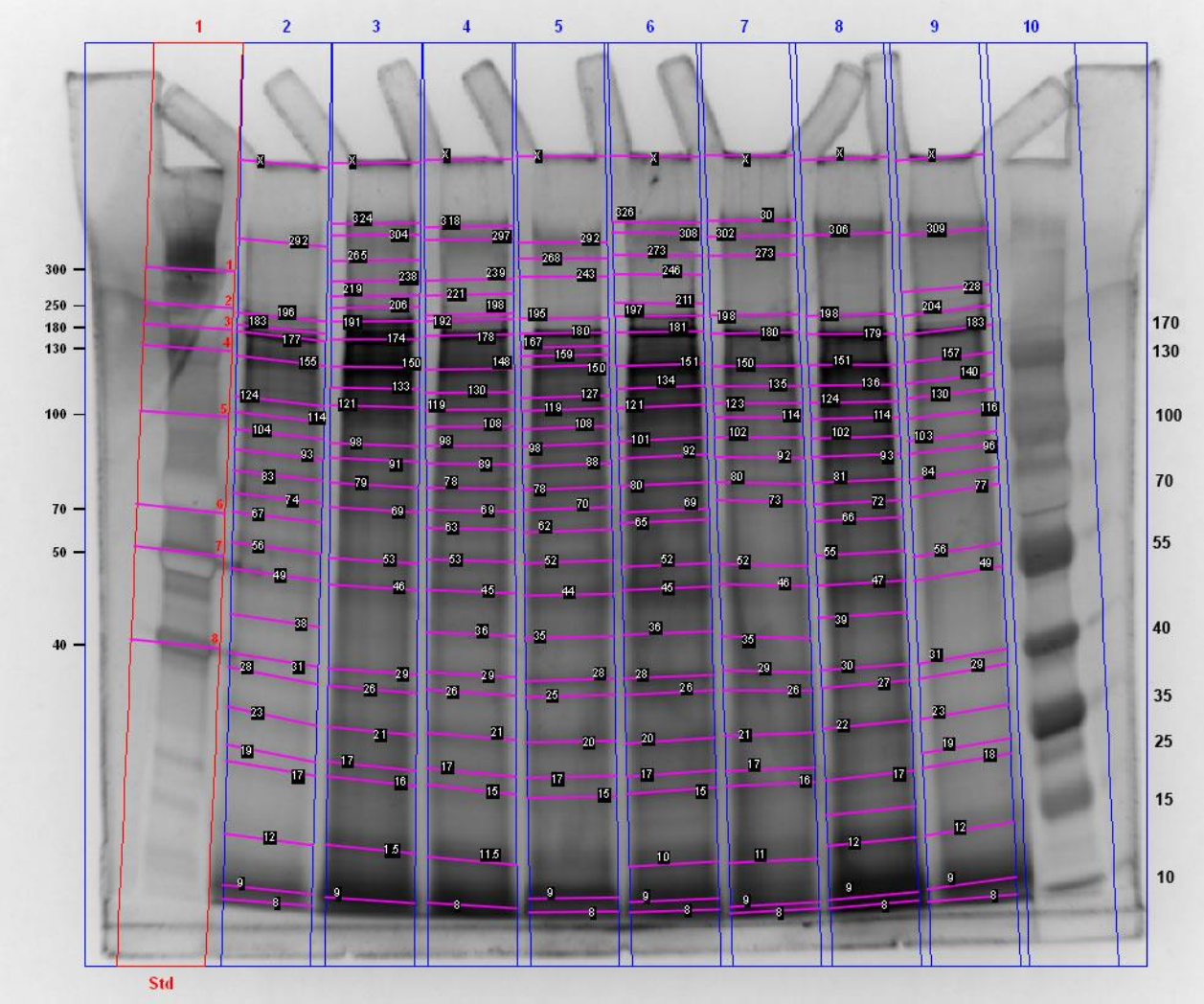

Figure 2.3.2.2. Annotated 4-20% SDS-PAGE of *S. vittatum* larval silk gland proteins. Lane 1 is the Spectra Multicolor High Range Protein Ladder (Thermo Scientific). Lanes 2-9 are larval silk gland proteins. Lane 10 is the Spectra Multicolor Page Ruler Protein Ladder (Thermo Scientific).

2.3.3. Simulium vittatum larval silk gland proteins

Using ImageLab 4.1 software, quantitative protein mass data were generated from the gel (Figure 2.3.2.1.). Protein bands not found in more than 2 lanes were treated as misstaining and excluded from data analysis. In order to correct for the curve present in the gel, the mass values were only taken from the sample directly to the right of the ladder (lane 1). Few proteins were not found in lane 2 but were found in other lanes in the gel. The mass of protein bands not found in lane 2 were corrected by dividing by a correction factor. Correction factor was calculated by

dividing the mass of a protein band in lane 3 by the mass of the same protein band in lane 2. In the case that a protein was found in lane 3 but not in lane 2, the average correction factor of the protein band above and below was calculated and applied to the protein.

Table 2.3.3.1. Mass of proteins present in *S. vittatum* larval silk glands. The contents of eight gland pairs were separated by SDS-PAGE and bands were analyzed using ImageLab 4.1.

Protein Mass (kDa)
317
292
262
236
217
204
196
176
155
138
124
114
104
93
83
74
67
56
49
38
31
28
23
19
17
12
9
8

A total of 28 unique protein bands were found to be present in larval *S. vittatum* silk glands, with mean mass ranging from 8 to 317 kDa. A larger band was observed at the edge of the well indicating that there were larger proteins that were not resolved on the gel.

2.3.4. Simulium vittatum pupal silk gland proteins

In order to determine if there are any silk proteins which are exclusive to either the larval or pupal life stages, pupal and larval gland samples were separated using SDS-PAGE.

Silver stained 4-20% polyacrylamide gel (Figure 2.3.4.1.) with larval and pupal silk gland protein samples in alternating lanes was analyzed using the ImageLab 4.1 software. The software was used to identify bands that were then annotated based on their position relative to other proteins in the gel (Figure 2.3.4.2.).

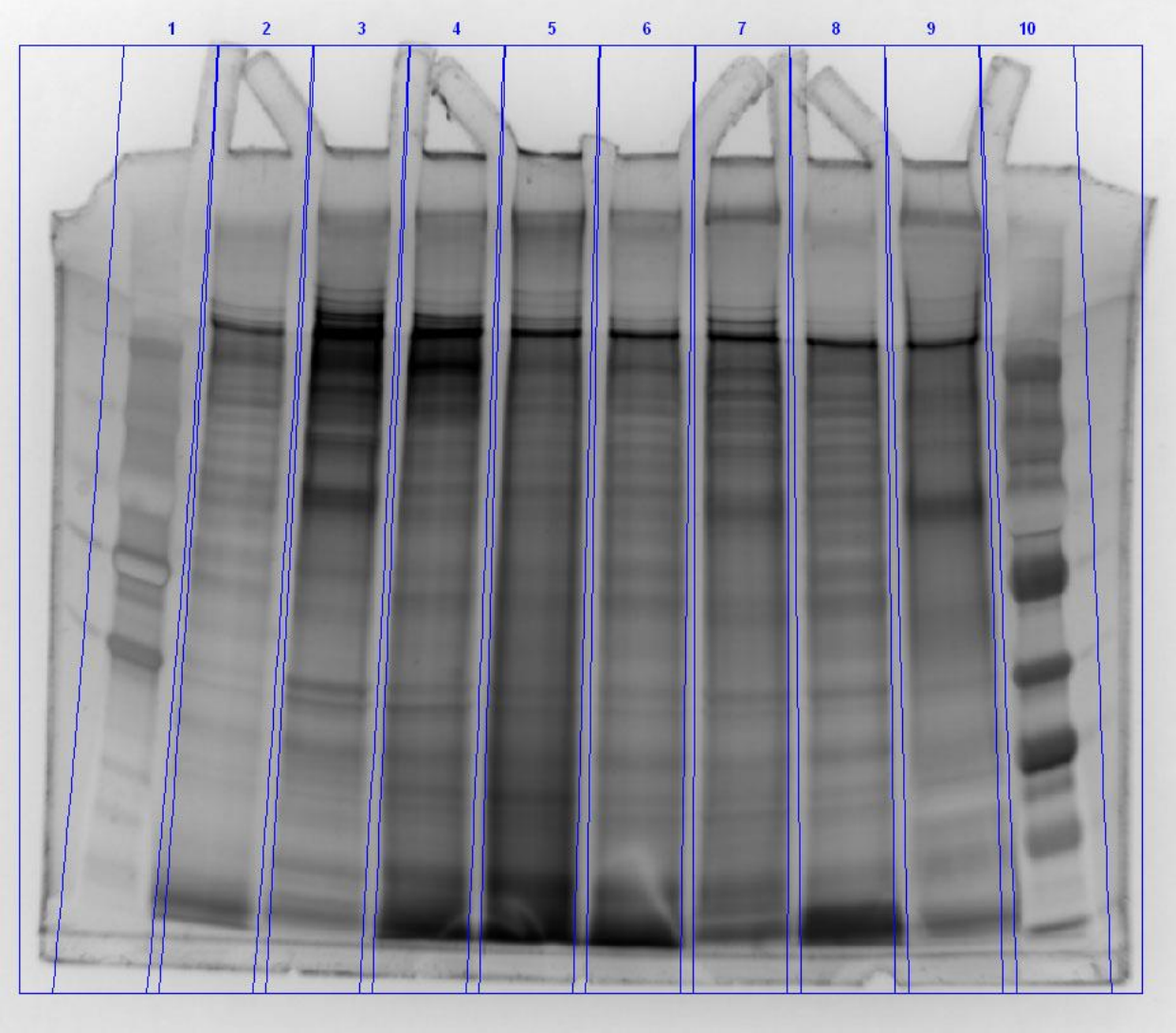

Figure 2.3.4.2. Silver stained 4-20% SDS-PAGE of S. vittatum larval and pupal silk gland proteins. Lane 1 is the Spectra Multicolor High Range Protein Ladder (Thermo Scientific). Lanes 2, 4, 6, and 8 are larval silk gland proteins. Lanes 3, 5, 7, and 9 are pupal silk gland proteins. Lane 10 is the Spectra Multicolor Page Ruler Protein Ladder (Thermo Scientific).

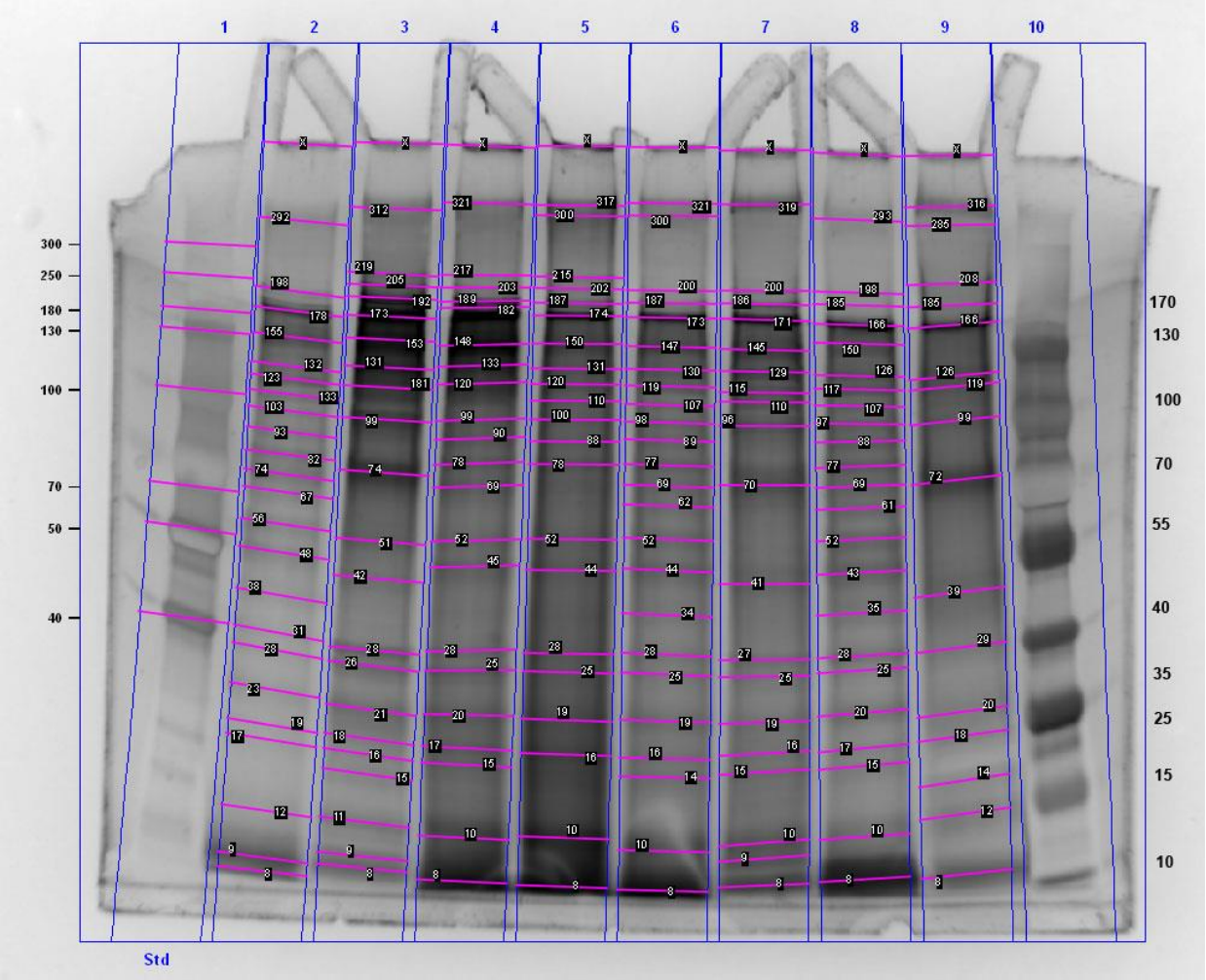

Figure 2.3.4.2. Annotated 4-20% SDS-PAGE of *S. vittatum* larval and pupal silk gland proteins. Lane 1 is the Spectra Multicolor High Range Protein Ladder (Thermo Scientific). Lanes 2, 4, 6, and 8 are larval silk gland proteins. Lanes 3, 5, 7, and 9 are pupal silk gland proteins. Lane 10 is the Spectra Multicolor Page Ruler Protein Ladder (Thermo Scientific).

Using the ImageLab 4.1 software, the protein mass data were quantified. Protein bands not found in more than 2 lanes were treated as misstaining and excluded from data analysis. In order to correct for the curve present in the gel, the mass values were only taken from the sample directly to the right of the ladder (lane 1). Few proteins were not found in lane 2 but were found in other lanes in the gel. The mass of protein bands not found in lane 2 were corrected by dividing by a correction factor. Correction factor was calculated by dividing the mass of a protein band in lane 3 by the mass of the same protein band in lane 2. In the case that a protein

was found in lane 3 but not in lane 2, the average correction factor of the protein band above and below was calculated and applied to the protein.

Table 2.3.4.1. Mass of proteins present in the pupal silk glands of *S. vittatum*. The protein content of four gland pairs were separated by SDS-PAGE, silver stained, and proteins were analyzed using ImageLab 4.1.

Protein Mass (kDa)
331
319
222
208
199
178
155
132
124
114
104
93
82
74
56
48
46
38
31
28
23
19
17
16
12
9
9

A total of 27 unique protein bands were found to be present in pupal *S. vittatum* silk glands, with mean mass ranging from 9 to 331 kDa.

2.3.5. Larval and pupal protein comparison

For ease of comparison larval and pupal proteins were matched based on average mass (Table 2.3.5.1.). The 63 kDa protein was found exclusively in the larvae and was put into the category of larval proteins. In contrast, the 41 kDa and 14 kDa proteins were found exclusively in pupae and were thus put into the pupal category. The majority of the proteins in the glands were found to be present in both life stages.

Table 2.3.5.1. Silk gland proteins by name and their presence or absence in the larval or pupal glands of *S. vittatum*. Secreted proteins (sp) were named based on their mass in kDa. Two proteins were found exclusively in the pupal silk glands with a mass of 41 kDa and 14 kDa, and one protein was found exclusively in the larval glands with a mass of 63 kDa.

Name	Present in Larva	Present in Pupa	Category
sp320	Yes	Yes	General
sp292	Yes	Yes	General
sp262	Yes	Yes	General
sp236	Yes	Yes	General
sp220	Yes	Yes	General
sp205	Yes	Yes	General
sp196	Yes	Yes	General
sp176	Yes	Yes	General
sp155	Yes	Yes	General
sp135	Yes	Yes	General
sp124	Yes	Yes	General
sp114	Yes	Yes	General
sp104	Yes	Yes	General
sp96	Yes	Yes	General
sp83	Yes	Yes	General
sp74	Yes	Yes	General
sp67	Yes	No	Larval
sp56	Yes	Yes	General
sp48	Yes	Yes	General
sp46	No	Yes	Pupal
sp38	Yes	Yes	General
sp31	Yes	Yes	General
sp28	Yes	Yes	General
sp23	Yes	Yes	General
sp19	Yes	Yes	General
sp17	Yes	Yes	General
sp16	No	Yes	Pupal
sp12	Yes	Yes	General
sp9	Yes	Yes	General
sp8	Yes	Yes	General

2.3.6. Silk gland proteins of other simuliid species

The silk gland proteins of larvae and pupae of 8 different species were analyzed in order to determine any differences and similarities among different species within the family Simuliidae.

Silk glands were dissected out of larvae and pupae of various simuliid species and extracted as described in section 2.2.7. Proteins were separated via SDS-PAGE and data were quantified using the ImageLab 4.1 software. Average mass and standard deviation were determined for proteins found in the glands of *Prosimulium fontanum* (Table A1), *Greniera abditoides* (Table A2), *Stegopterna diplomutata* (Table A3), *Stegopterna mutata* (Table A4), *Cnephia dacotensis* (Table A5), *Simulium croxtoni* (Table A6), *Simulium venustum* (Table A7), and *Simulium tuberosum* (Table A8).

Proteins present from the various species analyzed were compared to the proteins identified in *S. vittatum* (Table 2.3.6.1). It is of interest to note that the appearance of proteins does not seem to have any specific patterns from older to younger species. The larger proteins are distributed across the family with no apparent phylogenetic pattern. In contrast, the smaller proteins are more common throughout all species. For example, all proteins smaller than 20 kDa are found in 7 out of 8 species examined.

Table 2.3.6.1. Proteins by size (kDa) present in the silk glands of *S. vittatum* and their presence or absence in the glands of *S. tuberosum* larvae, *S. venustum/verecumdum* larvae, *S. croxtoni* larvae, *C. dacotensis* pupae, *St. mutata* larvae, *St. diplomutata* larvae, *G. abditoides* pupae, and *P. fontanum* pupae. Presence is indicated by an X (with the actual size of the protein for that species in brackets), and absence is indicated by a dash. L – Larvae; P - Pupae

S. vittatum	***S. tuberosum* (L)**	***S. ven/ver* (L)**	***S. croxtoni* (L)**	***C. dacotensis* (P)**	***St. mutata* (L)**	***St. diplomutata* (L)**	***G. abditoides* (P)**	***P. fontanum* (P)**
sp320	-	-	-	-	-	-	-	-
sp292	X (294)	-	X (296)	X (292)	-	-	-	-
sp262	X (260)	X (265)	-	-	X (269)	-	-	-
sp236	X (235)	-	-	-	-	-	-	-
sp220	-	-	-	-	-	-	-	-
sp205	-	-	X (205)	-	X (207)	-	-	-
sp196	X (196)	-	-	X (196)	-	-	-	-
sp176	X (172)	-	-	-	-	-	-	X (172)
sp155	-	-	-	X (155)	X (153)	-	X (160)	X (155)
sp135	X (132)	-	X (133)	-	X (139)	-	-	-
sp124	-	-	-	-	-	X (122)	X (128)	-
sp114	X (110)	-	-	X (114)	X (116)	-	X (112)	-
sp104	-	-	X (102)	X (106)	-	-	-	X (108)
sp96	-	X (93)	-	X (96)	-	X (91)	X (91)	-
sp83	-	-	-	X (88)	X (88)	-	-	-
sp74	X (78)	-	-	X (71)	X (75)	-	X (77)	X (76)
sp67	-	X (64)	-	-	-	-	-	-
sp56	X (57)	-	-	X (56)	X (53)	-	X (58)	-
sp48	-	-	-	X (51)	-	-	-	-
sp46	-	-	X (44)	-	-	-	X (43)	-
sp38	X (37)	X (39)	X (38)	X (36)	X (35)	X (37)	X (36)	X (35)
sp31	X (31)	X (32)	X (33)	X (31)	-	-	-	-
sp28	X (26)	X (28)	X (27)	-	X (28)	X (29)	-	X (28)
sp23	X (21)	X (22)	X (21)	X (21)	-	-	-	X (22)
sp19	-	-	X (18)	-	-	-	X (20)	X (19)
sp17	X (17)	X (18)	X (17)	X (17)	-	-	X (17)	-
sp16	X (16)	X (15)	X (15)	X (15)	X (15)	X (16)	X (15)	X (16)
sp12	X (12)	X (11)	X (12)	X (13)	X (12)	X (13)	X (12)	X (11)
sp9	X (9)	X (9)	X (9)	X (9)	X (9)	X (9)	X (8)	-
sp8	X (7)	X (7)	X (8)	X (7)	X (7)	X (7)	X (7)	X (7)

2.3.7. Glycosylation of silk gland proteins

To determine the presence of protein modifications that could give insight into additional protein properties, silk gland proteins were separated via SDS-PAGE and stained for glycosylation using the Periodic acid Schiff stain (PAS).

PAS stain showed staining of at least 4 unique silk gland protein bands (Figure 2.3.7.1). Staining did not show in each lane as protein concentrations varied for each sample and not all samples had enough protein to stain. Three proteins were stained at 517, 474 and 434 kDa (Table 2.3.7.1). These specific proteins were not found to stain consistently using silver staining. As with all previous gels, there was staining at the edge of the well.

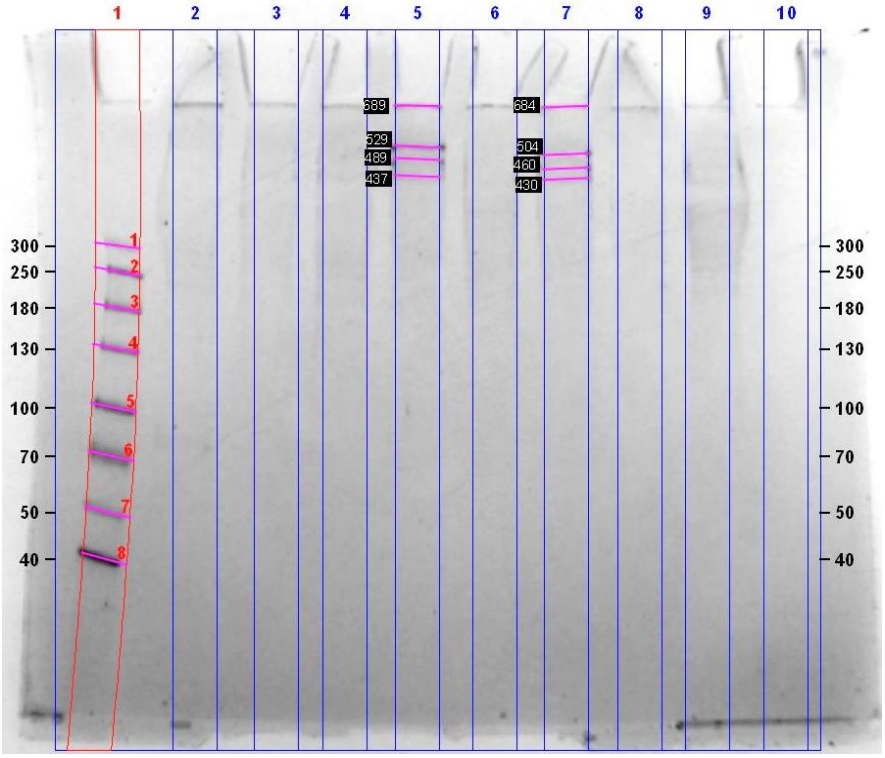

Figure 2.3.7.1. Periodic acid Schiff stained 4-20% SDS-PAGE of *S. vittatum* silk gland proteins. Lane 1 is the Spectra Multicolor High Range Protein Ladder (Thermo Scientific). Lanes 2-10 are *S. vittatum* silk gland proteins.

Protein data were quantified using the ImageLab 4.1 software. Average mass and standard deviation were calculated using Microsoft Excel (Microsoft Office 2010), and organized into Table 2.3.7.1.

Table 2.3.7.1. Mean mass of glycosylated proteins present in the silk glands of *S. vittatum*. The protein contents of eight gland pairs were separated by SDS-PAGE, stained with periodic acid Schiff stain, and proteins were analyzed using ImageLab 4.1. n is equal to the number of lanes which contained a band at the specified mass.

Mean Mass (kDa)	Standard Deviation	n
517	17.88	2
474	20.51	2
434	5.63	2

2.4. Discussion

2.4.1. Simulium vittatum cocoon proteins

Proteins from the *S. vittatum* cocoons were extracted and run on polyacrylamide gels. Two proteins were stained and identified, at 64 kDa and 56 kDa. These two proteins are similar in size to the sp70 and sp52 found in the *S. vittatum* salivary glands. The extraction methods attempted were not very efficient with regards to dissolving the cocoon, as most of the cocoon remained in the tube after centrifugation. In addition to the lack of dissolution of the cocoons, the sample itself did not yield consistent results between gels. The same problem was observed with silk glands that had been extracted from larvae that were stored in ethanol. Kiel *et al.* (2001) also found that the use of older cocoons yielded less dissolved protein than use of fresh cocoons, and much less than using silk directly from the silk gland itself. One possible explanation as to why these cocoons did not dissolve well could be due to the strong bonding of the silk proteins. Silk proteins change conformation after extrusion from the gland, and this conformational change yields crystallization of the proteins in the fiber. Silks are thought to form many disulfide bridges

during extrusion (Weislander, 1994), which likely contribute to the stronger bonds than the original conformation. These crystallized proteins may be much more resistant to being solubilized due to the stronger bonding together than the original conformation. This may explain why using the proteins in the silk gland, which would have non-crystalized proteins, would dissolve more readily than the crystalized cocoon silks.

2.4.2. Simulium vittatum larval silk gland proteins

It was found that the proteins in the *S. vittatum* silk glands are much more numerous than previously thought (Hunter and Bidochka, 2003; Kiel *et al*., 2002). These proteins range from a few proteins larger than 200 kDa to many proteins under 100 kDa. This supports the current model (Weislander, 1994) in which it is thought that silk fibers are composed of a number of different proteins produced in the silk glands.

Although most of the proteins in the gel were able to be given a molecular mass value, it should be noted when looking at the photos of the gels that there is one band present in all lanes which has not been given a value. These bands may correspond to a number of different proteins which are too large to be resolved through SDS-PAGE, and could account for any number of larger proteins. It is possible that these unresolved proteins at the edge of the well of the gel are as large 1400 kDa given the current information on silk protein in chironomid glands. Another such problem with resolution occurred between 300 kDa and 250 kDa, in which protein bands would resolve inconsistently. At times the bands would stain as one large smear, at other times a few bands would stain. This introduced much difficulty in the process of determining the actual values of the bands, and only the bands that appeared consistently were included in the data.

The banding patterns of the proteins in simuliid and chironomid silk glands are similar in their distribution into 3 size groups, however, there seems to be a large variation between the

actual sizes of the proteins between these two families. This variation in sizes could be attributed to the fact that in chironomids the proteins themselves have numerous internally repeating regions that are thought to have originated through duplication events within the genes (Weislander, 1994). If these genes were present in a common ancestor, the reason why these genes may not align together perfectly could be that some duplication events happened differently in simuliids after divergence from the common ancestor.

Kiel *et al.* (2001) resolved 6 silk proteins weighing 23, 42, 47, 68, 110, and >200 kDa from *S. ornatum* silk glands. It can be seen that most of these proteins are of similar size to the proteins described in this study (Table 2.3.5.1.), where *S. vittatum* proteins sp41, sp45, sp70, and sp99 are of very similar size to those found in *S. ornatum*. In addition to the work of Kiel *et al*. (2001), 25 new proteins have been added to the list of black fly silk proteins.

The use of silk glands from larvae was much simpler than the use of cocoons as the proteins were readily dissolved in the extraction buffers. One problem, however, was that the small size of the larvae and their silk glands sometimes made it difficult to tell whether there would be enough protein. Some samples had much more protein than others simply due to the larger amount of silk in the glands, and protein concentrations varied from individual to individual.

Given the information at hand, it is very difficult to comment about the function of any of the proteins present in the glands. However, I will speculate on their functions based on the functions of similar protein bands in chironomids. Although there is not one major protein present in the gland, it is likely (given information on other silk producing insects) that the largest proteins are the main backbones of the silk fibers. Additionally, it may be postulated that the other proteins in the glands serve as sericin-like proteins which may inhibit premature fiber

formation within the gland, coat the fiber during extrusion of the silk, raise the concentration of protein within the silk gland to increase the viscosity, or perhaps any combination of those roles.

2.4.3. Simulium vittatum pupal proteins

The large majority of proteins found in *S. vittatum* pupal silk glands were identical in size to proteins found in the larval silk glands. There were, however, 3 proteins that were found to be exclusive to specific life stages; the sp63 protein was found exclusively in the larval stage and the proteins sp41 and sp14 were present exclusively in the pupal glands. These proteins could be of interest as they may be important in any differences in the larval and pupal silk. The current study has suggested that sp41 and sp14 are prepupal proteins; however, sp41 was only found in 2 of the 4 pupal samples. The next step would be to analyse more pupae to confirm whether this protein is a pupal exclusive protein, or simply a degradation artifact. Similarly, sp14 is only found in 2 of the 4 pupal samples, however there are proteins of a similar size present in 7 of the 8 other *simuliid* species analyzed. These proteins may be pupal exclusive, although more work would need to be done to determine the validity of the results.

The pupal proteins could be involved in many processes. They could be cocoon specific proteins that begin to accumulate before the formation of the pupal casing. These could contribute properties that are more important to the cocoon such as longevity. For example, lepidopteran cocoons are known to have serine protease inhibitor proteins (Nirmala *et al*., 2001; Kurioka *et al*., 1999). Larvae can replace their silk pads by extruding new silk; however pupae are not able to renew their cocoons.

2.4.4. Silk gland proteins of other simuliid species

The proteins found in other species varied enormously from each other. The general trend observed in all species was that there were proteins ranging from above 200 to approximately 7 kDa. There was a large amount of difficulty with aligning the proteins among species, as there was a lot of variation between sizes and banding patterns. This may suggest that, due to these size variations, looking at silk with regards to protein size may not be the best approach. Alternatively, examining these silks at either amino acid or DNA sequences may give more information about the evolution of these genes in the common ancestor to Simuliids and Chironomids.

It was observed in *S. vittatum* that there were some discrepancies between samples and the use of multiple samples allowed for a better understanding of the actual protein profile for the species. In the section concerning other species only 2 samples were analyzed due to limiting factors. Although these data provide an understanding of the proteins and their relationship among species, this small sample size could lead to some error. It would be of benefit to analyze a larger set of samples for each species and further for each species at different life stages, as this would help to eliminate error such as misstaining or lack of stained proteins.

2.4.5. Glycosylation of silk gland proteins

Glycosylation of proteins can affect the way they are folded and the way they interact with each other (Lodish *et al*., 2007). Glycosylation is often found in secreted proteins or proteins found in the extracellular matrix, and it is shown that glycoproteins degrade at a slower rate than their non-glycosylated counterparts (Lodish *et al*., 2007). Glycosylated proteins also

tend to be slippery or have a slimy feel to them. This study has found 3 large proteins to be glycosylated, sp517, sp474, and sp434, which were not previously found using silver staining. In addition to these larger proteins, staining of glycoproteins at the edge of the wells is also observed. This indicates that there may be larger unresolved glycoproteins within the gland.

Proteins that are glycosylated are likely to be unaffected by trypsin, as the glycans can block access to the positive residues which trypsin finds and cleaves. This makes it difficult to the send proteins for sequencing using LC/MS as they must be cleaved first. In order to remove glycosylation from the protein it would first need to be isolated, de-glycosylated, and then digested and sequenced.

In conclusion, this study has found that black fly silk proteins are composed of multiple proteins that vary largely in size. Furthermore we have shown evidence to support the idea that simuliid silk originated in a common ancestor to chironomids (Sutherland *et al*., 2010). These findings may aid in future research on black fly silk properties that may lend themselves to many practical applications.

Chapter 3: *Simulium vittatum* silk gene sequencing

3.1. Introduction

The previous chapter showed that there are multiple silk proteins found in simuliid silk glands, which may be coded for by several genes. In order to better compare black fly silks to other silks in insects, attempts were made to determine the gene sequences of a few silk genes, as well as expand information on current gene sequence data. These data can help to provide information about protein sequences and domains, which can be used to better understand the evolution of black fly silks and their properties.

3.1.1. Previous work

In 2003, Hunter and Bidochka isolated a cocoon silk protein from the black fly *Simulium vittatum* by SDS-PAGE and had the N-terminus of the protein sequenced (GVAPKKYRKGHYVGGYGKKY). mRNA was isolated and using reverse transcriptase, cDNA was created. Using this amino acid sequence information, degenerate primers were designed and were used with a poly-T primer to amplify cDNA. The product was sequenced (Figure 3.1.1.2) and sequence was patented (Hunter & Bidochka, 2003). Given the size of the silk protein analyzed (56 kDa, as seen in Figure 2.3.1.1.), it is likely that there is more gene sequence information that is missing from the complete gene sequence. In this chapter, two different methods are used to determine more gene sequence information upstream of the gene: Inverse PCR, and SiteFinding-PCR.

```
          M  V  D  L  Q  A  A  A  L  V  I  G  V  A  P  K
1     CATATGGTCGACCTGCAGGCGGCCGCACTAGTGATTGGAGTTGCTCCAAA  50
        K  Y  R  K  G  H  Y  V  G  G  Y  G  K  K  Y  R
51    GAAGTACCGCAAGGGACACTATGTCGGGGGTTACGGGAAGAAGTATCGTA  100
      I  F  D  S  N  C  A  M  N  N  A  N  C  Q  N  P  N
101   TTTTTGACAGCAATTGTGCTATGAACAACGCCAACTGTCAGAATCCAAAC  150
       E  S  A  F  A  E  V  D  F  T  L  C  N  D  I  K  C
151   GAATCCGCCTTCGCCGAAGTTGATTTCACGCTGTGCAATGATATCAAATG  200
        P  R  K  C  D  K  K  L  D  P  V  C  A  F  D  G
201   TCCTAGGAAATGCGATAAAAAACTAGACCCGGTTTGTGCTTTTGATGGGA  250
      K  T  Y  R  Q  F  N  N  K  C  L  L  Q  E  F  N  D
251   AAACGTACAGACAATTTAACAACAAATGTCTGCTGCAAGAATTCAATGAT  300
       C  D  Q  N  V  F  Q  Y  F  N  A  V  T  N  K  K  M
301   TGCGATCAAAATGTGTTTCAATATTTCAACGCTGTGACTAACAAAAAAAT  350
        C  V  V  E  K  P  K  C  P  T  I  C  P  A  I  Y
351   GTGCGTGGTTGAGAAGCCAAAATGCCCGACCATTTGTCCAGCAATTTATG  400
      A  P  V  C  G  R  N  A  K  G  D  Y  K  S  F  A  S
401   CTCCCGTTTGTGGTCGAAATGCCAAAGGGGATTACAAAGTTTGCGAGT  450
       E  C  N  Q  S  A  F  N  C  L  I  S  K  N  Q  Y  T
451   GAATGCAACCAATCCGCATTCAACTGCTTGATTTCTAAGAATCAATATAC  500
        G  K  Y  D  L  S  F  C  D  I  E  F  P
501   GGGCAAGTATGATTTGAGTTTTTGCGACATCGAGTTCCCTTAAGCATGAC  550
551   GTTGTAACGTTTTTTCTCTGGATGTGCAAAACATAAATTACAAGCACTGG  600
601   ATTGAATGGTGTTTTATTAAATTTCCTTGTGACCTTTTTTCCATTATTCT  650
651   TTCCGGCCTTTAACAAGTAATCAATATTGATATCGGTCGTTTTTGTAAAG  700
701   ATTTTTTTTCAGTAAAAATATCCATCTCATTTTCACAAAAAAAAAAAAAA  750
751   AAAAAAAAAAAAAAAAAGCTTGTACAAAAAATCCCGCGGCCATGGCGGCC  800
801   GGGAGCATGCGACGTCGGGCCCA                              823
```

Figure 3.1.1.2. Cocoon silk sequence derived from cDNA of *S. vittatum* using degenerate primers designed from N-terminal sequence of cocoon silk protein and predicted amino acid sequence. Alternating triplets are highlighted (Hunter & Bidochka, 2003).

3.1.2. Inverse PCR

Inverse PCR (Ochman *et al*., 1988) is a method of genome walking which provides sequence information both upstream and downstream of a known sequence. Inverse PCR is carried out by digesting genomic DNA, ligating the digested genomic DNA into circularized DNA, and finally PCR on the self-ligated genomic DNA.

Primers are designed within the known sequence of the DNA; however, the forward primers are designed downstream of the reverse primers. During PCR the forward reaction will continue downstream beyond the known sequence until it reaches the ligated cut site and will continue until it reaches back to the region where the reverse primer binds. Thus the amplicon will contain known downstream information followed by unknown downstream sequence information, the cut site, unknown upstream information, and finally the known upstream information. Amplicons can then be purified and sent for Sanger sequencing.

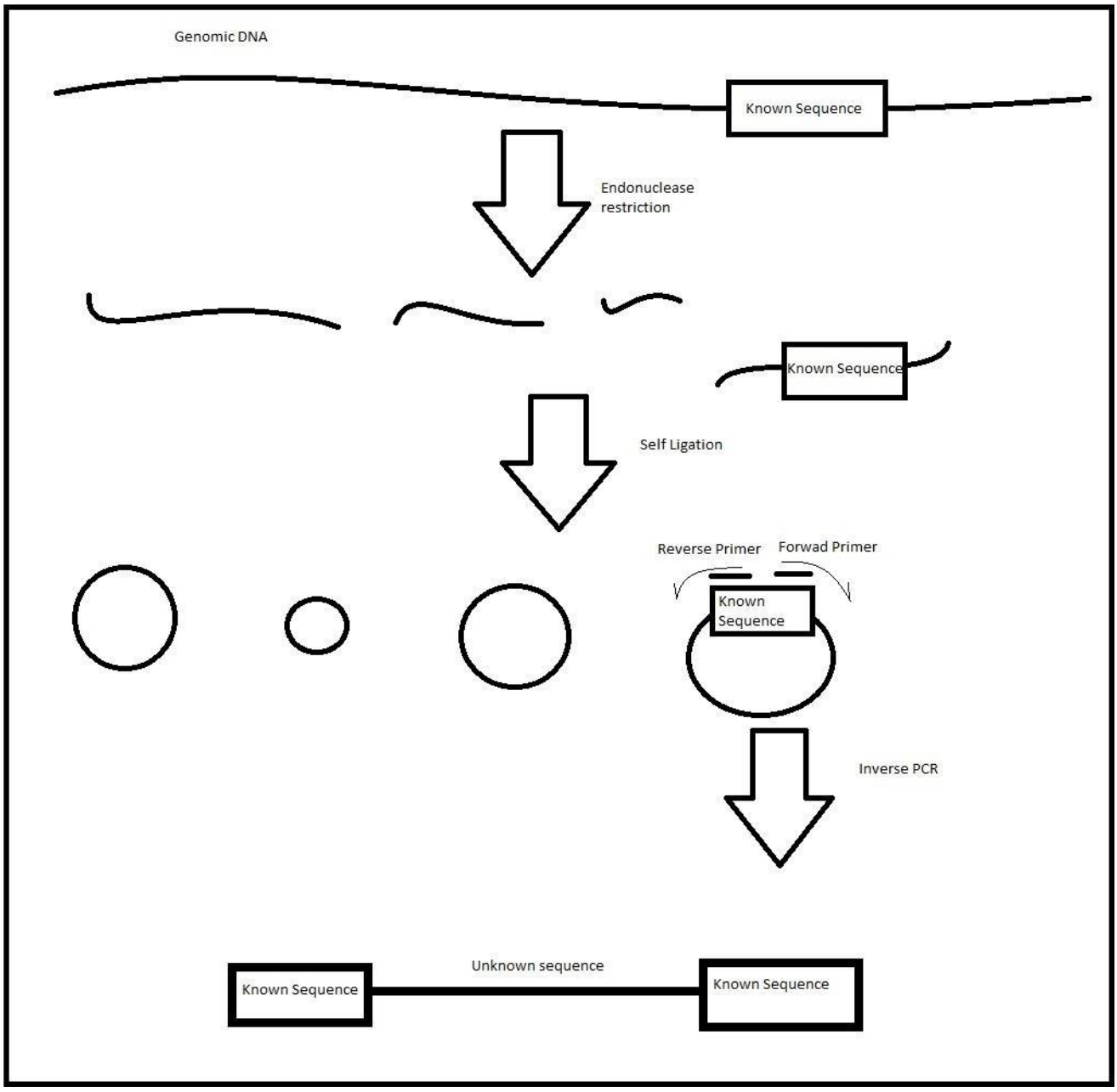

Figure 3.1.2.1. Summary of inverse PCR. Genomic DNA is digested with a restriction endonuclease and undergoes self-ligation. Gene specific primers are used to amplify unknown gene sequence upstream and downstream of the known sequence (Green & Sambrook, 2001).

3.1.3. SiteFinding-PCR

SiteFinding-PCR (Tan *et al*., 2005) is another method of genome walking which uses large non-specific primers to bind to genomic DNA in a variety of areas, and then relies on

nested PCR to amplify an unknown sequence upstream of gene specific reverse primers (Figure 3.1.3.1.).

In the first step the SiteFinding primer is allowed to bind randomly to the genomic DNA and *Taq* polymerase is used to amplify the regions bound by the primers. In the second step a nested primer which is identical to a portion of the large primer is used along with a gene specific primer to amplify a region upstream of the known sequence. The large primer has a stem-loop structure to ensure that the primers do not amplify regions flanked by another large primer. In the third step, a second nested PCR is performed specifically to amplify any amplicons from the first nested PCR. Amplicons can then be purified and sent for Sanger sequencing.

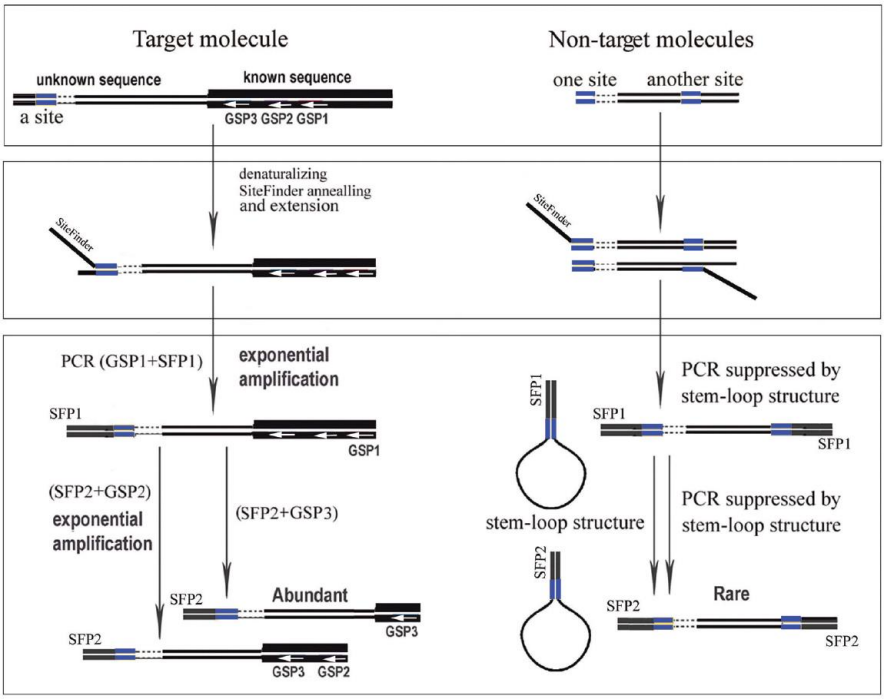

Figure 3.1.3.1. A summary of SiteFinding-PCR. A large non-specific primer binds randomly to genomic DNA. Nested primers and gene specific primers are used to amplify DNA fragments containing unknown sequence downstream of the large primer. (Figure adapted from Tan *et al.*, 2005).

Attempts to clone the cocoon silk sequence and upstream elements (Hunter & Bidochka, 2003) from genomic DNA are described. Furthermore, attempts were made to determine if the silk genes in simuliids are similar to the genes in chironomids by designing primers for chironomid silk genes. These gene sequences may also provide information that could help with future applications including the foreign expression of silk proteins for a number of applications.

3.2. Material and Methods

3.2.1. DNA extraction

DNA was extracted from individual *Simulium vittatum* larvae using a QiaAmp DNA minikit (Qiagen Catalogue number 51306). Individual larvae were placed in the clean microfuge tubes provided in the kit and crushed in 180 µL ATL buffer. DNA was extracted following the manufacturer's protocol for DNA purification from tissues.

3.2.2. Cocoon Silk Sequence PCR

Primers were designed based on the cocoon silk DNA sequence (Table 3.2.2.1). BFSintF and BFSexoF are forward primers that bind upstream and within the predicted gene, respectively. BFSR is the reverse primer which binds near the end of the predicted gene region. BFCSilkF is BFSintF designed with a *Sac*I restriction site at the 5' region of the primer, as well as an additional 3 nucleotides in order to bring Tm closer to that of BFCsilkR. BFCSilkR is BFSR designed with a *Bam*HI restriction site at the 5' region of the primer.

Table 3.2.2.1. Primers designed for PCR of the cocoon silk sequence. Bold text indicates a *Sac*I restriction site. Italicized text indicates a *Bam*HI restriction site.

Name	Sequence (5' → 3')	Tm (°C)
BFSintF	TCCCATATGGTCGACCTGC	65.9
BFSexoF	GGGTTACGGGAAGAAGTATCG	63.3
BFSR	GCTTAAGGGAACTCGATGTCG	64.7
BFCSilkF	TA**GAGCTC**CTCTCCCATATGGTCGACCTGC	76.4
BFCSilkR	TA*GGATCC*GCTTAAGGGAACTCGATGTCG	75.2

```
          BFSintF
1    CATATGGTCGACCTGCAGGCGGCCGCACTAGTGATTGGAGTTGCTCCAAA  50
                                          BFSexoF
51   GAAGTACCGCAAGGGACACTATGTCGGGGGTTACGGGAAGAAGTATCGTA  100
101  TTTTTGACAGCAATTGTGCTATGAACAACGCCAACTGTCAGAATCCAAAC  150
151  GAATCCGCCTTCGCCGAAGTTGATTTCACGCTGTGCAATGATATCAAATG  200
201  TCCTAGGAAATGCGATAAAAAACTAGACCCGGTTTGTGCTTTTGATGGGA  250
251  AAACGTACAGACAATTTAACAACAAATGTCTGCTGCAAGAATTCAATGAT  300
301  TGCGATCAAAATGTGTTTCAATATTTCAACGCTGTGACTAACAAAAAAAT  350
351  GTGCGTGGTTGAGAAGCCAAAATGCCCGACCATTTGTCCAGCAATTTATG  400
401  CTCCCGTTTGTGGTCGAAATGCCAAAGGGGATTACAAAAGTTTTGCGAGT  450
451  GAATGCAACCAATCCGCATTCAACTGCTTGATTTCTAAGAATCAATATAC  500
                                      BFSR
501  GGGCAAGTATGATTTGAGTTTTTGCGACATCGAGTTCCCTTAAGCATGAC  550
551  GTTGTAACGTTTTTTCTCTGGATGTGCAAAACATAAATTACAAGCACTGG  600
601  ATTGAATGGTGTTTTATTAAATTTCCTTGTGACCTTTTTTCCATTATTCT  650
651  TTCCGGCCTTTAACAAGTAATCAATATTGATATCGGTCGTTTTTGTAAAG  700
701  ATTTTTTTTCAGTAAAAATATCCATCTCATTTTCACAAAAAAAAAAAAAA  750
751  AAAAAAAAAAAAAAAAAGCTTGTACAAAAAATCCCGCGGCCATGGCGGCC  800
801  GGGAGCATGCGACGTCGGGCCCA                              823
```

Figure 3.2.2.1. Primer binding sites of BFSintF, BFSexoF and BFSR on the cocoon silk gene sequence.

PCR (Table 3.2.2.2) was carried out on the genomic DNA extracted from *S. vittatum* and both combinations of forward and reverse primers BFSintF, BFSexoF, and BFSR. PCR was carried out under the following parameters: initial denaturation step of 95°C for 5 min, followed by 35 cycles of 95°C for 1 min, 55°C for 1 min, 72°C for 2 min, and a final extension step at 72°C for 10 min. Elongation time was set at 2 minutes to allow for a fragment as large as 2000 bp. Annealing temperature was set as low as 55°C in response to an initial lack of amplification.

Table 3.2.2.2. Initial reaction conditions for PCR of cocoon silk sequence on *S. vittatum* genomic DNA.

Component	Final concentration
Buffer	1x
$MgCl_2$	3 mM
dNTPs	400 µM
Forward Primer (BFSintF or BFSexoF)	0.6 nM
Reverse Primer (BFSR)	0.6 nM
Genomic DNA	200 ng
Taq DNA polymerase	1.25 units
Milli-Q Water	to 25 µL

3.2.3. Cocoon Silk Sequence Inverse PCR

Inverse PCR was attempted in order to obtain more sequence information on the cocoon silk sequence. Primers (Table 3.2.3.1) were designed with the forward primers being downstream of the reverse primers.

Table 3.2.3.1. Primers used for inverse PCR based on the known cocoon silk DNA sequence of *S. vittatum*.

Name	Sequence (5’ → 3’)	Tm (°C)
iPCRF1	TGAGTTTTTGCGACATCGAG	63.9
iPCRF2	ACGGGCAAGTATGATTTGAG	61.5
iPCRR1	TTCGACCACAAACGGGAG	64.8
iPCRR2	GATTGGTTGCATTCACTCG	61.7

```
1    CATATGGTCGACCTGCAGGCGGCCGCACTAGTGATTGGAGTTGCTCCAAA  50
51   GAAGTACCGCAAGGGACACTATGTCGGGGGTTACGGGAAGAAGTATCGTA  100
101  TTTTTGACAGCAATTGTGCTATGAACAACGCCAACTGTCAGAATCCAAAC  150
151  GAATCCGCCTTCGCCGAAGTTGATTTCACGCTGTGCAATGATATCAAATG  200
201  TCCTAGGAAATGCGATAAAAAACTAGACCCGGTTTGTGCTTTTGATGGGA  250
251  AAACGTACAGACAATTTAACAACAAATGTCTGCTGCAAGAATTCAATGAT  300
301  TGCGATCAAAATGTGTTTCAATATTTCAACGCTGTGACTAACAAAAAAAT  350
351  GTGCGTGGTTGAGAAGCCAAAATGCCCGACCATTTGTCCAGCAATTTATG  400
         iPCRR1
401  CTCCCGTTTGTGGTCGAAATGCCAAAGGGGATTACAAAAGTTTTGCGAGT  450

        iPCRR2
451  GAATGCAACCAATCCGCATTCAACTGCTTGATTTCTAAGAATCAATATAC  500
        iPCRF2            iPCRF1
501  GGGCAAGTATGATTTGAGTTTTTGCGACATCGAGTTCCCTTAAGCATGAC  550
551  GTTGTAACGTTTTTTCTCTGGATGTGCAAAACATAAATTACAAGCACTGG  600
601  ATTGAATGGTGTTTTATTAAATTTCCTTGTGACCTTTTTTCCATTATTCT  650
651  TTCCGGCCTTTAACAAGTAATCAATATTGATATCGGTCGTTTTTGTAAAG  700
701  ATTTTTTTTCAGTAAAAATATCCATCTCATTTTCACAAAAAAAAAAAAAA  750
751  AAAAAAAAAAAAAAAAAGCTTGTACAAAAAATCCCGCGGCCATGGCGGCC  800
801  GGGAGCATGCGACGTCGGGCCCA
```

Figure 3.2.3.1. Primer binding sites of iPCRF1, iPCRF2, iPCRR1, and iPCRR2, on the cocoon silk gene sequence. Note iPCRF1 primer is bolded to show overlap with iPCRF2 primer.

One µg of *S. vittatum* genomic DNA was digested (Table 3.2.3.2) by *Aat* II, *Ava* I, or *Nco* I for 1 h at 37°C, followed by heat inactivation of endonucleases at 80°C for 20 min.

Table 3.2.3.2. Reaction conditions for the digestion of *S. vittatum* genomic DNA.

Component	Final Concentration
10X Endonuclease Buffer	1X
DNA	1 µg
Endonuclease	10 units
Milli-Q Water	To 25 µL

Varying amounts (20, 40 and 80 ng) of digested DNAs were added to ligation buffer (Table 3.2.3.3) and allowed to self-ligate overnight at room temperature with T4 DNA Ligase (NEB Catalogue # M0202S). Samples were incubated at 65°C for 20 min in order to heat inactivate the T4 DNA ligase.

Table 3.2.3.3. Reaction conditions for the self-ligation of enzyme digested *S. vittatum* genomic DNA.

Component	Final Concentration
10X T4 DNA Ligase Buffer (with 10 mM ATP)	2 µL
Digested genomic DNA	1-2 ng
T4 DNA Ligase	1 µL (400 units)
Milli-Q Water	to 20 µL

Self-ligated genomic DNA was used as a template for inverse PCR (Table 3.2.3.4.), using all four possible combinations of forward and reverse primers (Table 3.2.3.1.). PCR was carried out under the following parameters: initial denaturation step of 95°C for 5 min, followed by 35 cycles of 95°C for 1 min, 55°C for 1 min, 72°C for 2.5 min, and a final extension step at 72°C for 10 min. Elongation time was set at 2.5 minute to allow for the amplification of a fragment as large as 2500 bp. Annealing temperature was set at 55°C in response to an initial lack of amplification.

Table 3.2.3.4. Initial reaction conditions for inverse PCR on self-ligated *S. vittatum* genomic DNA.

Component	Final concentration
Buffer	1x
$MgCl_2$	3 mM
dNTPs	400 µM
Forward Primer	0.6 nM
Reverse Primer	0.6 nM
Self-Ligated genomic DNA	0.2 - 1 ng
Taq DNA polymerase	1.25 units
Milli-Q Water	to 25 µL

3.2.4. Cocoon Silk Sequence SiteFinding-PCR

In an attempt to acquire more upstream sequence information, primers were designed (Table 3.2.4.1.) to assist in SiteFinding-PCR.

Table 3.2.4.1. Cocoon silk sequence gene specific primers for SiteFinding-PCR. TSP3 is upstream of TSP2, which is upstream of TSP1, allowing for nested PCR.

Primer Name	Sequence (5' → 3')	Tm (°C)
VSP57 TSP1	CGTTACAACGTCATGCTTAAGG	63.2
VSP57 TSP2	AATGCGGATTGGTTGCATTCACTC	71.5
VSP57 TSP3	GGCATTTCGACCACAAACGGGAGC	78.2

```
1    CATATGGTCGACCTGCAGGCGGCCGCACTAGTGATTGGAGTTGCTCCAAA  50
51   GAAGTACCGCAAGGGACACTATGTCGGGGGTTACGGGAAGAAGTATCGTA  100
101  TTTTTGACAGCAATTGTGCTATGAACAACGCCAACTGTCAGAATCCAAAC  150
151  GAATCCGCCTTCGCCGAAGTTGATTTCACGCTGTGCAATGATATCAAATG  200
201  TCCTAGGAAATGCGATAAAAAACTAGACCCGGTTTGTGCTTTTGATGGGA  250
251  AAACGTACAGACAATTTAACAACAAATGTCTGCTGCAAGAATTCAATGAT  300
301  TGCGATCAAAATGTGTTTCAATATTTCAACGCTGTGACTAACAAAAAAAT  350
351  GTGCGTGGTTGAGAAGCCAAAATGCCCGACCATTTGTCCAGCAATTTATG  400
         TSP3
401  CTCCCGTTTGTGGTCGAAATGCCAAAGGGGATTACAAAAGTTTTGCGAGT  450
              TSP2
451  GAATGCAACCAATCCGCATTCAACTGCTTGATTTCTAAGAATCAATATAC  500
                                                  TSP1
501  GGGCAAGTATGATTTGAGTTTTTGCGACATCGAGTTCCCTTAAGCATGAC  550
551  GTTGTAACGTTTTTTTCTCTGGATGTGCAAAACATAAATTACAAGCACTGG  600
601  ATTGAATGGTGTTTTATTAAATTTCCTTGTGACCTTTTTTCCATTATTCT  650
651  TTCCGGCCTTTAACAAGTAATCAATATTGATATCGGTCGTTTTTGTAAAG  700
701  ATTTTTTTTCAGTAAAAATATCCATCTCATTTTCACAAAAAAAAAAAAAA  750
751  AAAAAAAAAAAAAAAAAGCTTGTACAAAAAATCCCGCGGCCATGGCGGCC  800
801  GGGAGCATGCGACGTCGGGCCCA
```

Figure 3.2.4.1. Primer binding sites of TSP1, TSP2, TSP3, on the cocoon silk gene sequence.

Figure 3.2.4.2. Large non-specific site-finding primers used for Site-finding PCR. (Figure adapted from Tan *et al.*, 2005)

As described in Tan *et al*. (2005), two site-finding reactions were set up (Table 3.3.4.2) in which *S. vittatum* genomic DNA was used as a template for PCR using either site-finding primer SiteFinder1 or SiteFinder2. SiteFinding reaction was carried out as follows: an initial denaturation step at 95°C for 2 min, followed by 2 cycles of 95°C for 1 min, 25°C for 1 min, ramp to 72°C over 3 min, 72°C for 10 min, and a final extension step of 72°C for 10 min.

Table 3.3.4.2. Reaction conditions for initial SiteFinding reaction on genomic *S. vittatum* genomic DNA.

Component	Final Concentration
Buffer	1X
$MgCl_2$	2 mM
dNTPs	400 µM
SiteFinder1 or SiteFinder2	
Genomic DNA	200 ng
Taq DNA Polymerase	2.5 units
Milli-Q Water	to 20 µL

Product from previous site-finding reactions was used as a template for the first nested PCR (Table 3.3.4.3). Nested PCR was carried out under the following parameters: initial denaturation step of 95°C for 5 min, followed by 35 cycles of 95°C for 30 min, 55°C for 1 min, 72°C for 5 min, and a final extension step at 72°C for 10 min.

Table 3.3.4.3. Reaction conditions for first nested PCR after initial site-finding reaction on *S. vittatum* genomic DNA.

Component	Final Concentration
Buffer	1X
$MgCl_2$	2 mM
dNTPs	400 µM
SFP1	10 pmol
TSP1	200 nM
Site-Finder primed genomic DNA	5 µL
Taq DNA Polymerase	2.5 units
Milli-Q Water	to 20 µL

Products from the first nested PCR were used as a template for the second nested PCR (Table 3.3.4.4). Nested PCR was carried out under the following parameters: initial denaturation step of 95°C for 5 min, followed by 35 cycles of 95°C for 30 min, 55°C for 1 min, 72°C for 5 min, and a final extension step at 72°C for 10 min.

Table 3.3.4.4. Reaction conditions for second nested PCR after initial SiteFinding reaction on *S. vittatum* genomic DNA.

Component	Final Concentration
Buffer	1X
$MgCl_2$	2 mM
dNTPs	400 µM
SFP2	0.2 µM
TSP2 or TSP3	200 nM
First nested PCR product	1 µL
Taq DNA Polymerase	2.5 units
Milli-Q Water	to 20 µL

3.2.5. Chironomid 3' region primer design

The NCBI database was accessed to search for chironomid Balbiani Ring (BR) gene sequences. Four unique silk genes were selected: *Chironomus tentans* BR1 gene 3' end (GenBank X03490.1), *Chironomus tentans* Balbiani Ring mRNA BR 2.1 3'-end (GenBank:

X07701.1), *Chironomus tentans* Balbiani Ring mRNA BR2.2 3'-end (GenBank: X07702.1), and *Chironomus pallidivittatus* BR1 gene for giant secretory protein (GenBank: X05621.1). These genes were selected because they contain a well-conserved 3' region. A multiple sequence alignment was performed in A plasmid Editor (ApE).

BRcon (Balbiani Ring Consensus sequence) primers were designed in ApE based on the consensus sequence of the multiple sequence alignment. Using ApE, each primer was confirmed to work on each individual sequence allowing for 3 nucleotide errors, as long as the nucleotide errors were not in the last 7 nucleotides from the 3' end of the primers, though most of the errors were in the first 8 bases from the 5' end.

BRcte (Balbiani Ring *Chironomus tentans*) forward primers were designed based on *Chironomus tentans* BR1 (Balbiani Ring) gene 3' end sequence. There were no additional BRcte reverse primers designed because the BRcon reverse primers were identical to designed BRcte reverse primers.

Table 3.2.5.1. Primers designed based on *Chironomus tentans* BR1 gene (BRcte) and the consensus sequence (BRcon) of BR1, BR 2.1, and BR2.2 genes from *C. tentans* and *C. pallidivittatus*.

Name	Sequence (5' → 3')	Tm (°C)
BRcteF1	TCAGCTGGTAAACCAAGCAAATCTG	68.8
BRcteF2	GCAAATCTGAACCAAGAACTGAAAG ACC	70.0
BRconF1	CAATTGCGATGGTGCCAAATTCC	72.6
BRconF2	GGCTAAGCTCTTCAAACCACAAGG	67.7
BRconR1	CCAAGCCTTTGACAGCATCTTGG	70.6
BRconR2	GAGTGCCTTGCCAAACTTCTCC	68.1

PCR was conducted (Table 3.2.5.2.) on genomic *S. vittatum* using chironomid BR primers using all 8 combinations of forward and reverse primers. PCR was carried out under the

following parameters: initial denaturation step of 95°C for 5 min, followed by 35 cycles of 95°C for 30 seconds, 55°C for 45 seconds, 72°C for 2 min, and a final extension step at 72°C for 10 min. Annealing temperatures were set as low as 55°C in response to an initial lack of amplification.

Table 3.2.5.2. Initial reaction conditions for PCR using chironomid primers on *S. vittatum* genomic DNA.

Component	Final concentration
Buffer	1x
$MgCl_2$	3 mM
dNTPs	400 µM
Forward Primer	0.6 nM
Reverse Primer	0.6 nM
Self-Ligated genomic DNA	200 ng
Taq DNA polymerase	1.25 units
Milli-Q Water	to 25 µL

3.2.6. PCR Products and Sequencing

PCR products were visualized by agarose gel electrophoresis with ethidium bromide. Products were cleaned using Sigma GeneElute PCR Clean-Up kit (NA1020) and sent for sequencing.

In reactions that resulted in multiple bands, the individual bands were excised from the gel using a clean scalpel and DNA was extracted using QIAquick Gel Extraction Kit (#28704) following manufacturer's instructions.

Amplicons were sent to The Centre for Applied Genomics at the Hospital for Sick Children in Toronto, Ontario. Samples were prepared by adding 50 ng of the PCR product in 7 µL of milli-Q water with an addition 0.7 µL of milli-Q water containing 50 ng of one primer. Both a forward and a reverse sample were sent for sequencing.

Forward and Reverse sequencing results were aligned to each other using (ApE) to verify agreement, as well as a BLAST search (http://blast.ncbi.nlm.nih.gov) for comparison to the database sequences. Finally sequence information was aligned with chironomid BR genes to check for homology.

3.3. Results

In order to determine more sequence information about cocoon silk, inverse PCR and SiteFinder PCR were performed to obtain upstream gene sequence. In addition, primers for chironomid silk genes were designed and optimized in order to determine any homologous gene sequences in simuliids.

3.3.1. Cocoon Silk Sequence Primer Optimization

Primer combination BFSintF and BFSR (Table 3.2.2.1) was used in the initial PCR conditions (Table 3.2.2.2); however, agarose gel electrophoresis showed no amplification of the desired product, although the COI positive control showed amplification. A series of reactions were performed in order to optimize magnesium ion concentration (1mM - 4mM), primer concentration (0.2 µM - 1.2 µM), and annealing temperature (50°C - 65°C). Optimization of each parameter yielded no amplification.

The same optimization effort was made for the primer combination BFSexoF and BFSR (Table 3.2.2.1), with the same result of no amplification.

3.3.2. Cocoon Silk Sequence Inverse PCR Optimization

The inverse PCR method was attempted in order to obtain more upstream information for the cocoon silk sequence.

The initial trial of the inverse PCR yielded large bright smears of DNA when visualized on agarose gel (data not shown). Annealing temperature and magnesium ion concentration were optimized; however, these resulted in the same bright smear of DNA. The amount of self-ligated DNA, as well as the concentration of DNA in the initial ligation reaction, were varied with the same result.

3.3.3. Cocoon Silk Sequence SiteFinding-PCR Optimization

As an alternative to inverse PCR, SiteFinding-PCR (Tan *et al.*, 2005) was attempted in order to find more upstream information about the cocoon silk sequence.

The parameters for the reactions laid out in the paper by Tan *et al.* (2005) were followed almost exactly. The only changes made were to the initial SiteFinding reaction, where two SiteFinding cycles took place instead of just one. This was in order to increase the initial transcripts available for the site-specific primers to bind to.

The reactions yielded no bands (data not shown).

3.3.4. Chironomid BR 3' region

Initial PCR reactions with different primer combinations yielded multiple amplicons of varying sizes.

One primer set, BRconF1 and BRconR2, was taken and optimized for magnesium chloride concentration and annealing temperature. The optimum annealing temperature for this primer set was found to be 62°C, with 3mM of magnesium chloride.

Although product was brighter in the trial with 2mM $MgCl_2$, the results were only reproducible with 3mM $MgCl_2$. Even with optimized annealing temperature and magnesium ion concentration, there were multiple PCR products in the lanes. PCR product of the optimum condition was sequenced (Figure 3.3.4.1)

5’ gaaagcgcttcatgtttccgctsaacagcttkataagttcaaargaacaatattggaaatggtccatcraatcaggaatttggca ccatcgcaattgacct 3’

Figure 3.3.4.1. Trimmed sequence read for the PCR product of BRconF1 and BRconR2 amplified from *S. vittatum* genomic DNA.

A BLAST search of the 101 nucleotide sequence read shows a 23 nucleotide identity with *Chironomus tentans* BR1 3’ end, as well as 21 nucleotide identity with the BR 2.1 and BR 2.2 3’ ends of the same species, which all correspond with the same sequence as the BRconF1 primer.

3.4. Discussion

3.4.1. Cocoon Silk Sequence Primer Optimization.

Optimization efforts for the primers designed for the cocoon silk sequence yielded no results. Although there was ample room for primer design downstream of the predicted gene, the amount of gene sequence upstream of the predicted start codon made it difficult to design good

primers. In order to design better primers, more sequence information was sought out using genome walking methods.

3.4.2. Cocoon Silk Sequence Inverse PCR

The expected result for the inverse PCRs was to have one single band where the primers fit on the self-ligated genomic DNA. Another possible outcome of inverse PCR could be multiple bands of different sizes, which would have been the result of incomplete genomic digestion. In response to the possibility that the smearing was due to multiple sizes of amplicons, the genomic DNA was completely digested overnight, but the same results were obtained.

3.4.3. Cocoon Silk Sequence SiteFinding-PCR

SiteFinding-PCR allowed for the use of the whole gene sequence for the design of the 3 nested reverse primers. The attempted PCRs did not yield any amplicons, which could be due to a number of reasons. With these multiple step genome walking procedures, there are a number of possible errors that could have come about, however even with optimization efforts there were no amplification results. It is possible that the design of new primers may result in successful amplification.

Upon protein BLAST analysis of the protein translation of the cocoon gene sequence we can see that there is a putative conserved domain from the kazal-like serine protease inhibitor. Proteins such as these could contribute to the cocoon's longevity as they could inhibit proteolytic activity of bacteria and fungi that could contribute to accelerated degradation of the pupal cocoon (Nirmala *et al.*, 2001).

3.4.4. Chironomid homologous gene sequences

With an abundance of chironomid information, the next step was to try to confirm that simuliid silk genes were similar to chironomid silk genes. Through the alignment of multiple genes over 2 species of chironomids, primers which amplify genes in simuliids were successfully designed. The successful amplification using chironomid primers indicates that the genes in simuliids are likely to be homologous to the silk genes in chironomids. In their aquatic habitat, silk is used by larval chironomids and simuliids to assure successful feeding and by pupal chironomids and simuliids for protection; without such genes these organisms would likely not survive. The important nature of these gene products has lead to their conservation in simuliids and chironomids from their common ancestor.

The 3' ends of these silk genes are very similar to one another. As with any genes which are part of a greater gene family, this could be the cause of multiple amplifications during PCR. Primers designed for one gene might amplify other genes due to the highly similar sequence of the different genes. In simuliids this similarity is likely to be the cause of multiple amplicons during PCR with these primers. The 3' regions are so similar to each other that during PCR the primer will find several different genes with the same sequences and amplify those regions.

It should be mentioned that during this protocol whole larvae were used when performing DNA extractions. This could lead to the amplification of non-simuliid genes that were present in the contents of the midgut. In order to verify that these genes are present in the actual simuliid it would be beneficial to dissect out the glands themselves and perform PCR on the DNA extracted from the gland.

Chapter 4: General Conclusions

Previous to this thesis, there had been very little information on the actual proteins present in simuliid silk glands. Kiel *et al.* (2003) had done some work to determine the presence of proteins in the labial glands; however, the results only focused on a few proteins, all of which were in the small category. A substantial amount of research has been done in chironomids in the past, and this thesis is novel because it examined simuliid silk genes and proteins in the context of their homology to chironomids.

Firstly, this thesis has established a general profile of the proteins produced by simuliid silk glands. The proteins present in the gland at any time are numerous and range from very large proteins (>300 kDa) to very small proteins (<10 kDa). While there are a number of different proteins present, there does not seem to be any one major protein that is in higher abundance within the gland. However, based on the current theory that the larger proteins in chironomids act as the backbone for the silk fiber, it follows that the same proteins are likely to serve the same functions in simuliids.

Similar to chironomids, the majority of the proteins found in the larval silk glands are the same as those found in the pupal glands. The larvae of *S. vittatum* differ from chironomid larvae in their stage specific proteins. *S. vittatum* has one larval silk protein not present in the pupae, sp63 kDa protein. This differs from *Chironomus tentans* which also has 1 larval specific protein, sp40, which weighs approximately 40 kDa. Alternatively, pupal *S. vittatum* has 2 stage specific proteins, a 41 kDa protein and a 14 kDa protein, while in *C. tentans* there are 4 pupal specific proteins, sp195, sp140, sp115 and sp40 (Weislander, 1994). Although there are stage specific proteins in both *S. vittatum* and *C. tentans*, only the sp41 protein from *S. vittatum* and the sp40 protein from *C. tentans* are nearly matching based on protein size.

The proteins between species and genera show quite a bit of similarity in their general profile, such as the proteins ranging from large to small; however the mass of these proteins tends to differ between species. Taking into account that many chironomid silk proteins are composed mostly of internal tandem repeats which are thought to have originated from duplication events, these size variations may be due to differences in the number of tandem repeats. For example, after divergence one species may have undergone multiple duplication events leaving it with a homologous protein that is very different in size from a different species which has not undergone as many duplication events.

Whatever the reason may be, the proteins in the silk glands between species appear to be similar in their distribution by size; however, the exact size of individual proteins in each species seems to differ.

In future work it may be beneficial to create antibodies for a specific protein and attempt to use these to label homologous proteins. This would result in a better estimation of homologous proteins among species rather than assuming proteins of the same size are homologous.

Although chironomid gene primers work for simuliid silk genes, there still exists much information to be added to the current sequence information. The chironomid silks have been extensively annotated and much sequence information exists for multiple genes in the Balbiani Ring gene family.

In chironomids, a number of silk genes are found on the chromosome puffs known as the Balbiani Rings of the giant polytene chromosomes. A future project should be the design of fluorescent probes specific for the silk gene sequences, which can be used for fluorescent in situ hybridization (FISH) to simuliid chromosomes. This could be used to confirm the idea that the

silk genes are found in the chromosomal puffs found in the giant polytene chromosomes of simuliid silk gland cells.

Future work should include the construction and annotation of an EST library for the transcriptome of *S. vittatum* larval and pupal silk glands. This would provide enough information to note the major changes in simuliid silk genes, as well as give insights as to the similarities of the genes in chironomids, allowing us to build a model of which genes were present in a common ancestor and which proteins evolved after divergence.

While there exist many potential applications for the use of simuliid silks, first a better understanding of their composition must be achieved. More genetic, proteomic, and biochemical information remains to be learned before we can begin to use this silk's incredible properties to create new and revolutionary biomaterials.

References

Adler, P. H., Currie, D. C., & Wood, D. M. (2004). The black flies (Simuliidae) of North America. Ithaca, New York: Cornell University Press.

Aramwit, P., & Sangcakul, A. (2007). The effects of sericin cream on wound healing in rats. Bioscience, Biotechnology, and Biochemistry, (0), 0709070553.

Barr, W. B. (1984). Prolegs and attachment of *Simulium vittatum* (sibling IS-7) (Diptera: Simuliidae) larvae. Canadian Journal of Zoology, 62(7), 1355-1362.

Bayçin, D., Altiok, E., Ülkü, S., & Bayraktar, O. (2007). Adsorption of olive leaf (*Olea europaea L.*) antioxidants on silk fibroin. Journal of Agricultural and Food Chemistry, 55(4), 1227-1236.

Berezikov, E., Blinov, A. G., Scherbik, S., Cox, C. K., & Case, S. T. (1998). Structure and polymorphism of the *Chironomus thummi* gene encoding special lobe-specific silk protein, ssp160. Gene, 223(1), 347-354.

Brereton, C., House, W., Armitage, P., & Wotton, R. (1999). Sorption of pesticides to novel materials: Snail pedal mucus and blackfly silk. Environmental Pollution, 105(1), 55-65.

Case, S. T. W., L. (1992). Secretory proteins of *Chironomus* salivary glands: Structural motifs and assembly characteristics of a novel biopolymer. Results and Problems in Cell Differentiation, 19, 187-226.

Craig, D. A., Craig, R. E. G., Crosby, T. K. (2012). Simuliidae (Insecta: Diptera). Fauna of New Zealand, 68, 336.

Edström, J., Sierakowska, H., & Burvall, K. (1982). Dependence of Balbiani ring induction in *Chironomus* salivary glands on inorganic phosphate. Developmental Biology, 91(1), 131-137.

Edström, J., Rydlander, L., & Francke, C. (1980). Concomitant induction of a Balbiani ring and a giant secretory protein in *Chironomus* salivary glands. Chromosoma, 81(1), 115-124.

Galler, R., Rydlander, L., Riedel, N., Kluding, H., & Edström, J. (1984). Balbiani ring induction in phosphate metabolism. Proceedings of the National Academy of Sciences, 81(5), 1448-1452.

Green, M. R., & Sambrook, J. (2001). Molecular cloning: A laboratory manual / Michael R. green, joseph Sambrook (3rd Ed.). Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Hertner, T., Eppenberger, H. M., & Lezzi, M. (1983). The giant secretory proteins of *Chironomus tentans* salivary glands: The organization of their primary structure, their amino acid and carbohydrate composition. Chromosoma, 88(3), 194-200.

Hertner, T., Meyer, B., Eppenberger, H. M., & Mähr, R. (1980). The secretion proteins in *Chironomus tentans* salivary glands: Electrophoretic characterization and molecular weight estimation. Wilhelm Roux's Archives of Developmental Biology, 189(1), 69-72.

Hunter, F., & Bidochka, M. (2003). "Isolated cocoon silk protein from *Simulium vittatum* and nucleic acids encoding such protein." US patent # US20030032089 A1.

Ings, N. L., Hildrew, A. G., & Grey, J. (2010). Gardening by the psychomyiid caddisfly *Tinodes waeneri*: Evidence from stable isotopes. Oecologia, 163(1), 127-139.

Inoue, S., Tanaka, K., Arisaka, F., Kimura, S., Ohtomo, K., & Mizuno, S. (2000). Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6: 6: 1 molar ratio. Journal of Biological Chemistry, 275(51), 40517-40528.

Jin, H., & Kaplan, D. L. (2003). Mechanism of silk processing in insects and spiders. Nature, 424(6952), 1057-1061.

Johnson, M. F., Reid, I., Rice, S. P., & Wood, P. J. (2009). Stabilization of fine gravels by net-spinning caddisfly larvae. Earth Surface Processes and Landforms, 34(3), 413-423.

Kao, W. Y., Case, S. T. (1985). A novel giant secretion polypeptide in *Chironomus* salivary glands: Implications for another Balbiani ring gene. The Journal of Cell Biology, 101(3), 1044.

Kato, N., Sato, S., Yamanaka, A., Yamada, H., FUWA, N., & NOMURA, M. (1998). Silk protein, sericin, inhibits lipid peroxidation and tyrosinase activity. Bioscience, Biotechnology, and Biochemistry, 62(1), 145-147.

Ki, C. S., Kim, J. W., Oh, H. J., Lee, K. H., & Park, Y. H. (2007). The effect of residual silk sericin on the structure and mechanical property of regenerated silk filament. International Journal of Biological Macromolecules, 41(3), 346-353.

Kiel, E. (1997). Durability of simuliid silk pads (Simuliidae, Diptera). Aquatic Insects, 19(1), 15-22.

Kiel, E., & Röder, T. (2002). Gelelectrophoretic studies on labial gland secretions of immature blackflies (Simuliidae, Diptera). Limnologica-Ecology and Management of Inland Waters, 32(3), 201-205.

Kurioka, A., Yamazaki, M., Hirano, H. (1999). Primary structure and possible functions of a trypsin inhibitor of *Bombyx mori*. European Journal of Biochemistry, 259, 120-126.

Lee, K. H. (2004). Silk sericin retards the crystallization of silk fibroin. Macromolecular Rapid Communications, 25(20), 1792-1796.

Lodish, H. F., Berk, A., Kaiser, C. A., Krieger, M., Scott, M. P., Bretscher, A., Ploegh, H., Matsudaira, P. (2007). Molecular cell biology (6th Ed.). New York: W.H. Freeman and Company.

Makarevich, I., Berezikov, E., Guryev, V., & Blinov, A. (2000). Molecular phylogeny of the *Chironomus* genus deduced from nucleotide sequences of two nuclear genes, ssp 160 and the globin 2b gene. Molecular Biology, 34(4), 606-612.

Meinel, L., Hofmann, S., Karageorgiou, V., Kirker-Head, C., McCool, J., Gronowicz, G., *et al*. (2005). The inflammatory responses to silk films *in vitro* and *in vivo*. Biomaterials, 26(2), 147-155.

Merritt, R. W., & Cummins, K. W. (1996). An introduction to the aquatic insects of North America. Kendall Hunt.

Mizoguchi, K., Iwatsubo, T., & Aisaka, N. (1991). Separating membrane made of cross-linked thin film of sericin and production thereof. Japan Patent.

Nirmala, X., Kodrik, D., Zurovec, M., & Sehnal, F. (2001). Insect silk contains both a kunitz-type and a unique kazal-type proteinase inhibitor. European Journal of Biochemistry, 268(7), 2064-2073.

Ochman, H., Gerber, A. S., & Hartl, D. L. (1988). Genetic applications of an inverse polymerase chain reaction. Genetics, 120(3), 621.

Padamwar, M., & Pawar, A. (2004). Silk sericin and its applications: A review. Journal of Scientific and Industrial Research, 63(4), 323-329.

Sasaki, M., Kato, N., Watanabe, H., & Yamada, H. (2000). Silk protein, sericin, suppresses colon carcinogenesis induced by 1, 2-dimethylhydrazine in mice. Oncology Reports, 7(5), 1049-1101.

Sasaki, M., Yamada, H., & Kato, N. (2000). A resistant protein, sericin improves atropine-induced constipation in rats. Food Science and Technology Research, 6(4), 280-283.

Segrest, J. P., & Jackson, R. L. (1972). Molecular weight determination of glycoproteins by polyacrylamide gel electrophoresis in sodium dodecyl sulfate. Methods in Enzymology, 28, 54-63.

Sehnal, F., & Sutherland, T. (2008). Silks produced by insect labial glands. Prion, 2(4), 145-153.

Simpson, R. J., Adams, P. D., & Golemis, E. (2009). Basic methods in protein purification and analysis: A laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sutherland, T. D., Campbell, P. M., Weisman, S., Trueman, H. E., Sriskantha, A., Wanjura, W. J., *et al*. (2006). A highly divergent gene cluster in honey bees encodes a novel silk family. Genome Research, 16(11), 1414-1421.

Sutherland, T. D., Young, J. H., Sriskantha, A., Weisman, S., Okada, S., & Haritos, V. S. (2007). An independently evolved dipteran silk with features common to lepidopteran silks. Insect Biochemistry and Molecular Biology, 37(10), 1036-1043.

Sutherland, T. D., Young, J. H., Weisman, S., Hayashi, C. Y., Merritt, D. J. (2010). Insect silk: One name, many materials. Annual Review of Entomology, 55, 171.

Tamada, Y., Sano, M., Niwa, K., Imai, T., & Yoshino, G. (2004). Sulfation of silk sericin and anticoagulant activity of sulfated sericin. Journal of Biomaterials Science, Polymer Edition, 15(8), 971-980.

Tan, G., Gao, Y., Shi, M., Zhang, X., He, S., Chen, Z., & An, C. (2005). SiteFinding-PCR: A simple and efficient PCR method for chromosome walking. Nucleic Acids Research, 33(13), e122.

Tanaka, T. (2001). Antifrosting method, antifrosting agent and snow melting agent. Japan Patent,

Tao, W., Li, M., & Xie, R. (2005). Preparation and structure of porous silk sericin materials. Macromolecular Materials and Engineering, 290(3), 188-194.

Terada, S., Nishimura, T., Sasaki, M., Yamada, H., & Miki, M. (2002). Sericin, a protein derived from silkworms, accelerates the proliferation of several mammalian cell lines including a hybridoma. Cytotechnology, 40(1-3), 3-12.

Teramoto, H., Nakajima, K., & Takabayashi, C. (2004). Chemical modification of silk sericin in lithium chloride/dimethyl sulfoxide solvent with 4-cyanophenyl isocyanate. Biomacromolecules, 5(4), 1392-1398.

Teramoto, H., Nakajima, K., & Takabayashi, C. (2005). Preparation of elastic silk sericin hydrogel. Bioscience, Biotechnology, and Biochemistry, 69(4), 845-847.

Tsubouchi, K., Igarashi, Y., Takasu, Y., & Yamada, H. (2005). Sericin enhances attachment of cultured human skin fibroblasts. Bioscience, Biotechnology, and Biochemistry, 69(2), 403-405.

Tsujimoto, K., Takagi, H., Takahashi, M., Yamada, H., & Nakamori, S. (2001). Cryoprotective effect of the serine-rich repetitive sequence in silk protein sericin. Journal of Biochemistry, 129(6), 979-986.

Weislander, 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.

Wood, D. M., & Borkent, A. (1989). Phylogeny and classification of the Nematocera. In J. F. McAlpine (Ed.), Manual of Nearctic Diptera (pp. 1359). Hull, Quebec: Canadian Government Publishing Centre.

Wotton, R. S. (2011). EPS (extracellular polymeric substances), silk, and chitin: Vitally important exudates in aquatic ecosystems. JNABS Journal, 30(3), 762-769.

Wotton, R. S., & Hirabayashi, K. (1999). Midge larvae (Diptera: Chironomidae) as engineers in slow sand filter beds. Water Research, 33(6), 1509-1515.

Wotton, R. (1986). The use of silk life-lines by larvae of *Simulium noelleri* (Diptera). Aquatic Insects, 8(4), 255-261.

Yonemura, N., Sehnal, F., Mita, K., & Tamura, T. (2006). Protein composition of silk filaments spun under water by caddisfly larvae. Biomacromolecules, 7(12), 3370-3378.

Young, J. H., & Merritt, D. J. (2003). The ultrastructure and function of the silk-producing basitarsus in the hilarini (Diptera: Empididae). Arthropod Structure & Development, 32(2), 157-165.

Zhang, Y. (2002). Applications of natural silk protein sericin in biomaterials. Biotechnology Advances, 20(2), 91-100.

Zhaorigetu, S., Yanaka, N., Sasaki, M., Watanabe, H., & Kato, N. (2003). Silk protein, sericin, suppresses DMBA-TPA-induced mouse skin tumorigenesis by reducing oxidative stress, inflammatory responses and endogenous tumor promoter TNF-α. Oncology Reports, 10(3), 537-543.

Appendix A - Silk gland proteins of various black fly species

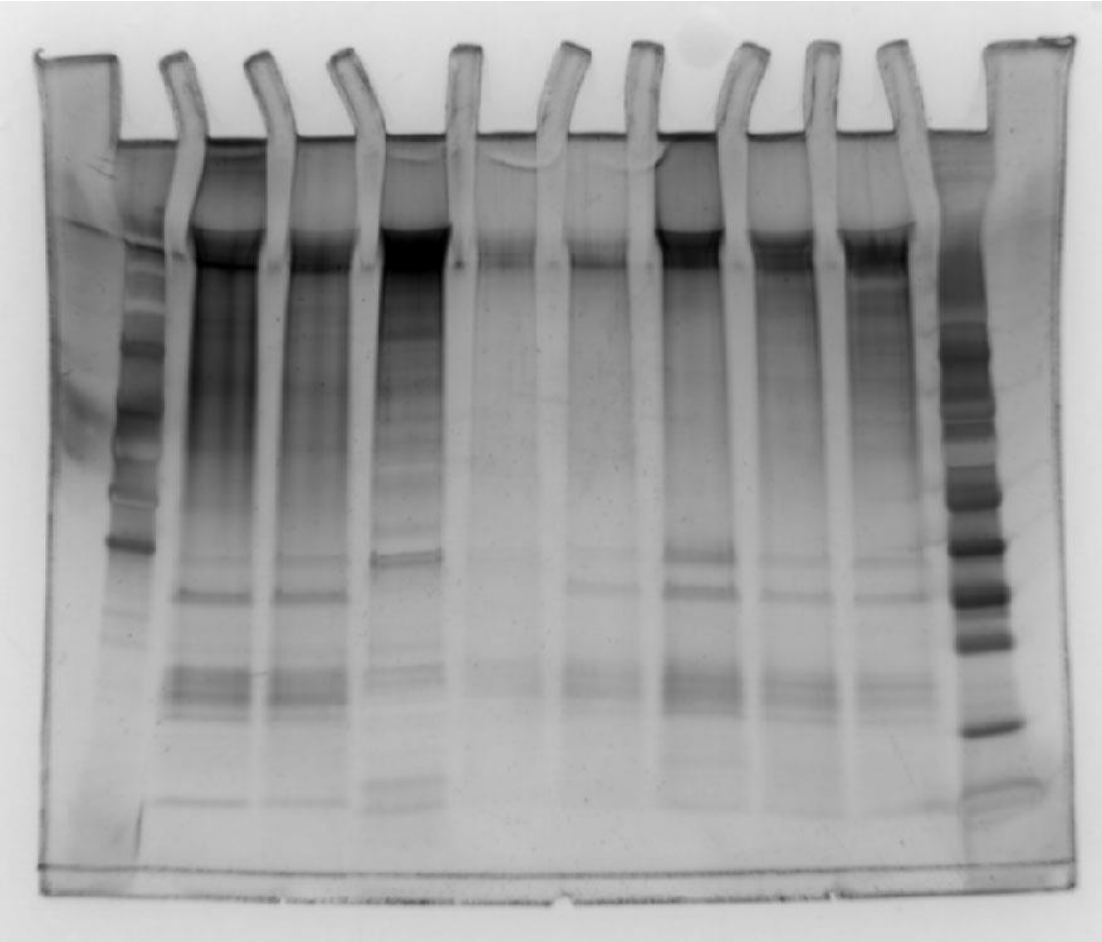

Figure A1. Silver stained 4-20% SDS-PAGE of simuliid silk gland proteins. Lane 1 is the Spectra Multicolor High Range Protein Ladder (Thermo Scientific). Lanes 2-3 are *P. fontanum* pupal silk gland proteins. Lanes 4-5 are *G. abditoides* pupal silk gland proteins. Lanes 6-7 are *St. diplomutata* larval silk gland proteins. Lanes 8-9 are *St. mutata* larval silk gland proteins. Lane 10 is the Spectra Multicolor Page Ruler Protein Ladder (Thermo Scientific).

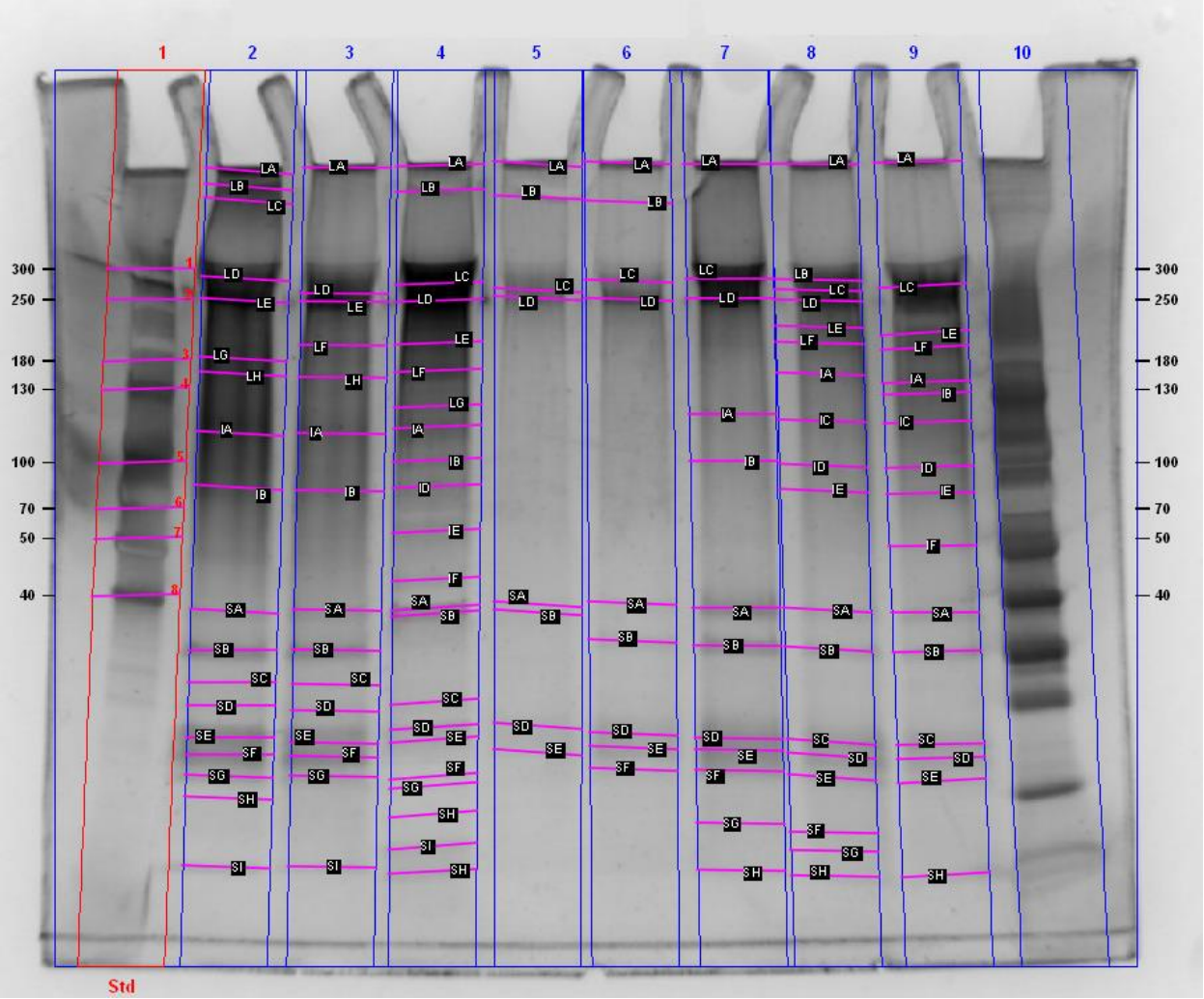

Figure A2. Annotated 4-20% SDS-PAGE of simuliid silk gland proteins. Lane 1 is the Spectra Multicolor High Range Protein Ladder (Thermo Scientific). Lanes 2-3 are *P. fontanum* pupal silk gland proteins. Lanes 4-5 are *G. abditoides* pupal silk gland proteins. Lanes 6-7 are *St. diplomutata* larval silk gland proteins. Lanes 8-9 are *St. mutata* larval silk gland proteins. Lane 10 is the Spectra Multicolor Page Ruler Protein Ladder (Thermo Scientific).
C. dacotensis (P), S. croxtoni (L), S. ven/ver (L), S. tuberosum (L)

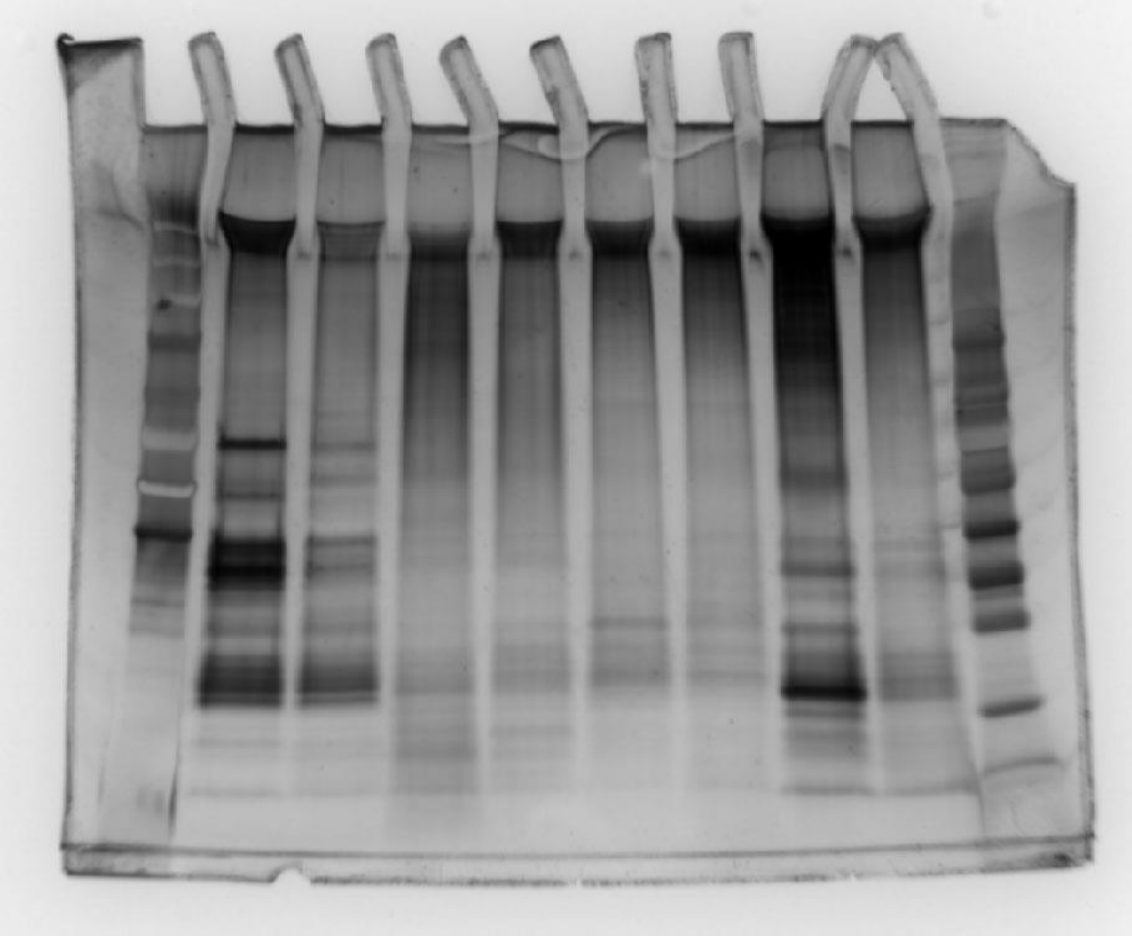

Figure A3. Silver stained 4-20% SDS-PAGE of simuliid silk gland proteins. Lane 1 is the Spectra Multicolor High Range Protein Ladder (Thermo Scientific). Lanes 2-3 are *C. dacotensis* pupal silk gland proteins. Lanes 4-5 are *S. croxtoni* larval silk gland proteins. Lanes 6-7 are *S. venustum/verecundum* larval silk gland proteins. Lanes 8-9 are *S. tuberosum* larval silk gland proteins. Lane 10 is the Spectra Multicolor Page Ruler Protein Ladder (Thermo Scientific).

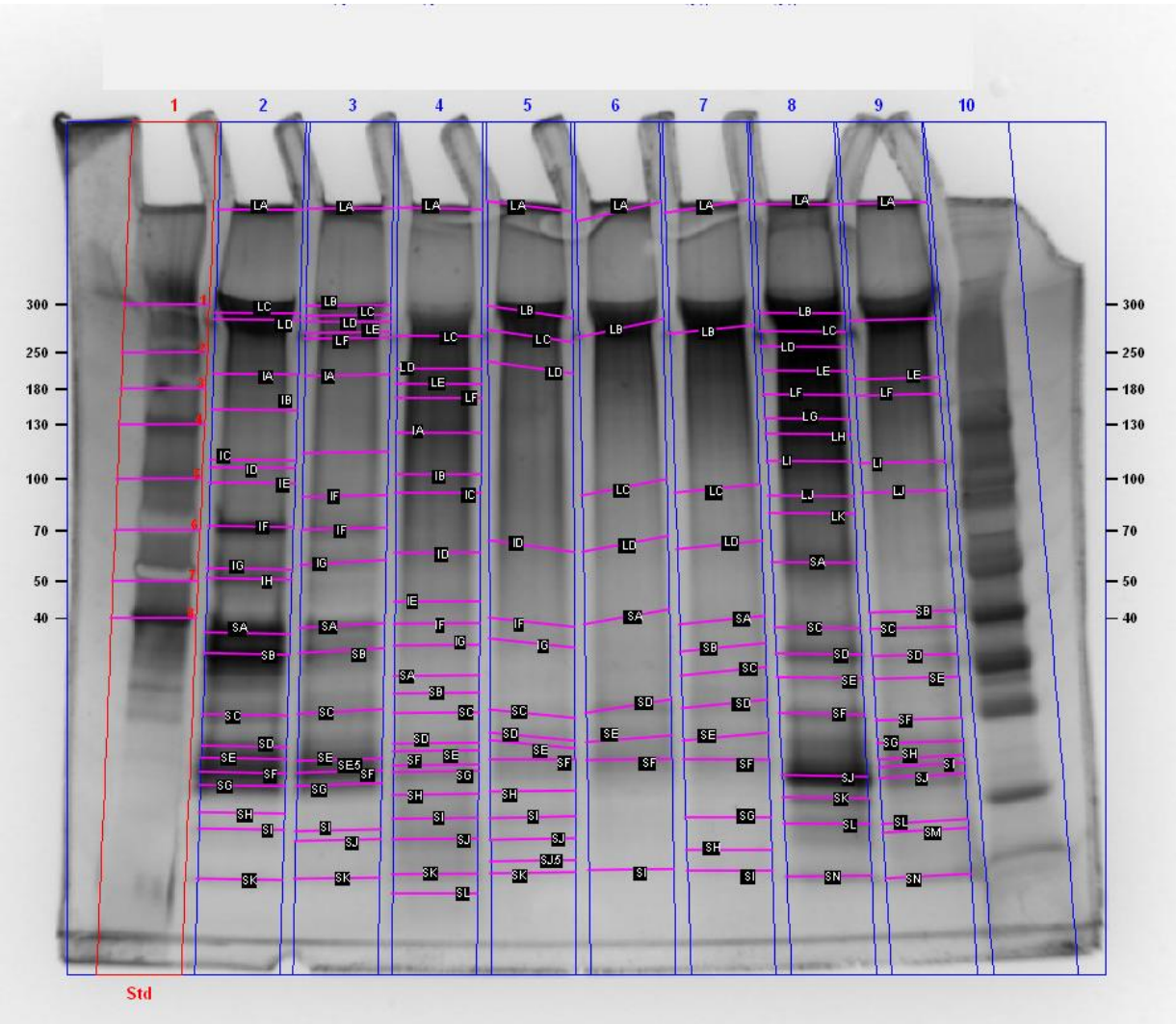

Figure A4. Annotated 4-20% SDS-PAGE of simuliid silk gland proteins. Lane 1 is the Spectra Multicolor High Range Protein Ladder (Thermo Scientific). Lanes 2-3 are *C. dacotensis* pupal silk gland proteins. Lanes 4-5 are *S. croxtoni* larval silk gland proteins. Lanes 6-7 are *S. venustum/verecundum* larval silk gland proteins. Lanes 8-9 are *S. tuberosum* larval silk gland proteins. Lane 10 is the Spectra Multicolor Page Ruler Protein Ladder (Thermo Scientific).

Table A1. Mass of proteins present in *P. fontanum* pupal silk glands.

Mean Mass (kDa)	SD	n
506	-	1
464	-	1
272	17.97	2
247	1.26	2
187	-	1
172	-	1
155	2.36	2
108	0.00	2
76	1.16	2
35	0.18	2
28	0.00	2
22	0.23	2
19	0.49	2
16	0.40	2
14	0.15	2
13	0.00	2
11	-	1
7	0.00	2

Table A2. Mass of proteins present in *G. abditoides* pupal silk glands.

Mean Mass (kDa)	SD	n
485	14.79	2
271	8.26	2
250	2.54	2
190	-	1
160	-	1
128	-	1
112	-	1
91	-	1
77	-	1
58	-	1
43	-	1
36	0.56	2
35	0.18	2
20	-	1
17	0.09	2
15	0.85	2
13	-	1
12	-	1
10	-	1
8	-	1
7	-	1

Table A3. Mass of proteins present in *St. diplomutata* larval silk glands.

Mean Mass (kDa)	SD	n
461	-	1
283	2.87	2
251	1.28	2
122	-	1
91	-	1
37	0.75	2
29	0.59	2
16	0.33	2
15	0.15	2
13	0.07	2
9	-	1
7	-	1

Table A4. Mass of proteins present in *St. mutata* larval silk glands.

Mean Mass (kDa)	SD	n
283	-	1
269	5.47	2
248	-	1
207	5.26	2
188	3.82	2
153	5.44	2
139	-	1
116	0.59	2
88	0.45	2
75	0.76	2
53	-	1
35	0.36	2
28	0.28	2
15	0.16	2
14	0.14	2
12	0.19	2
9	-	1
8	-	1
7	0.03	2

Table A5. Mass of proteins present in *C. dacotensis* pupal silk glands.

Mean Mass (kDa)	SD	n
309	-	1
292	2.97	2
280	2.84	2
259	-	1
249	-	1
196	1.00	2
155	-	1
114	4.07	2
106	-	1
96	-	1
88	-	1
71	0.73	2
56	1.69	2
51	-	1
36	1.30	2
31	0.48	2
21	0.32	2
17	-	1
15	0.09	2
14	0.07	2
13	0.07	2
11	-	1
10	0.05	2
9	-	1
7	0.04	2

Table A6. Mass of proteins present in *S. croxtoni* larval silk glands.

Mean Mass (kDa)	SD	n
296	-	1
253	0.00	2
205	1.04	2
184	-	1
168	-	1
133	-	1
102	-	1
90	-	1
62	1.89	2
44	-	1
38	0.19	2
33	0.34	2
27	-	1
24	-	1
21	0.11	2
18	0.54	2
17	0.51	2
15	0.39	2
14	-	1
12	0.19	2
11	0.05	2
9	0.00	2
8	-	1
7	0.04	2
6	-	1

Table A7. Mass of proteins present in *S. venustum* larval silk glands.

Mean Mass (kDa)	SD	n
265	1.35	2
93	0.47	2
64	0.32	2
39	0.60	2
32	-	1
28	-	1
22	0.22	2
18	0.18	2
15	0.00	2
11	-	1
9	-	1
7	0.04	2

Table A8. Mass of proteins present in *S. tuberosum* larval silk glands.

Mean Mass (kDa)	SD	n
534	-	1
294	-	1
282	-	1
260	-	1
235	-	1
196	6.97	2
172	0.00	2
146	-	1
132	-	1
110	0.56	2
89	1.81	2
78	-	1
57	-	1
41	-	1
37	0.00	2
31	0.16	2
26	0.00	2
21	0.52	2
17	-	1
16	-	1
15	-	1
14	0.00	2
12	-	1
10	0.10	2
9	-	1
7	0.00	2

Appendix B - Cocoon silk protein BLAST

```
  1 MVDLQAAALVIGVAPKKYRKGHYVGGYGKKYRIFDSNCAMNNANCQNPNESAFAEVDFTL 60
 61 CNDIKCPRKCDKKLDPVCAFDGKTYRQFNNKCLLQEFNDCDQNVFQYFNAVTNKKMCVVE 120
121 KPKCPTICPAIYAPVCGRNAKGDYKSFASECNQSAFNCLISKNQYTGKYDLSFCDIEFP  179
```

Figure B1. Amino acid sequence derived from cocoon silk gene. The highlighted section is the putative conserved domain sequence.

Cd Length: 46 **Bit Score:** 32.65 **E-value:** 6.11e-03

```
                        10        20        30
                  ....*....|....*....|....*....|
Cocoon silk     66 CPRKCDKKLDPVCAFDGKTYrqfNNKCLLQ 95
Kazal domain     2 CPEACPREYDPVCGSDGVTY---SNECHLC 28
```

Figure B2. BLAST output of the alignment of cocoon silk protein putative conserved domain to Conserved Domain Database Kazal type serine protease inhibitor (smart00280). Red text indicates high level of conservation, blue text indicates low level of conservation, grey indicates gaps.