UNIVERSITY OF CALIFORNIA
RIVERSIDE

Characterization and Evolutionary Analyses of Silk Fibroins From Two Insect Orders: Embioptera and Lepidoptera

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Genomics and Bioinformatics

by

Matthew Aric Collin

June 2010

Dissertation Committee:
Dr. Cheryl Y. Hayashi, Chairperson
Dr. David N. Reznick
Dr. Julia N. Bailey-Serres

The Dissertation of Matthew Aric Collin is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

The text of this dissertation, in part or in full, is a reprint of the material as it appears in Insect Biochemistry and Molecular Biology, February 2009, and *Biomacromolecules*, August 2009. Co-author Edina Camama listed in the Biomacromolecules publication provided research assistance. Co-author Janice S. Edgerly listed in those publications provided embiopteran cultures and technical expertise. Co-author Cheryl Y. Hayashi listed in those publications directed and supervised the research that forms the basis for this dissertation. Co-authors Jessica E. Garb and Brook O. Swanson listed in those publications provided technical expertise.

Justin Bastow, Daniel Gruner, John Mclaughlin, May Roberts, and Don Strong provided larval *Hepialus californicus.* I thank Edina Camama, Shou-Wei Ding, Samer Elkashef, Stuart Le, and Arthur Omura for laboratory assistance; Dan Borchardt, Richard Debus, and Bohdan Schatschneider for assistance with FTIR techniques; and Shun-ichi Sasanuma for the technical part of *H. californicus* EST sequencing. Nadia Ayoub, Dan Borchardt, Merri Lynn Casem, Richard Debus, Janice Edgerly, Jessica Garb, John Gatesy, Michael McGowen, Morris Maduro, Frantisek Sehnal, James Starrett, Brook Swanson, and anonymous reviewers contributed helpful comments. Research was supported by NSF awards DEB-0515868 to Cheryl Y. Hayashi, DEB-0515865 to Janice S. Edgerly, Czech grant ME32 to Frantisek Sehnal. *H. californicus* EST sequencing was supported by a grant from Japan's Ministry of Agriculture, Forestry and Fisheries to KM. I thank UCR GGB and UCR for financial support through Teaching and Graduate Student Researcher Assistantships and a College Fellowship. Additional Graduate

Student Researcher support was provided by NSF award DEB-0515868 to Cheryl Y. Hayashi.

ABSTRACT OF THE DISSERTATION

Characterization and Evolutionary Analyses of Silk Fibroins From Two Insect Orders: Embioptera and Lepidoptera

by

Matthew Aric Collin

Doctor of Philosophy, Graduate Program in Genetics, Genomics and Bioinformatics
University of California, Riverside, June 2010
Dr. Cheryl Y. Hayashi, Chairperson

Silk production has independently evolved in several insect lineages, such as Hymenoptera (bees, ants and wasps), Siphonaptera (fleas), and Archeognatha (bristletails and silverfish). Primarily composed of proteins, silks are used for functions, such as protection, prey capture, dispersal, and reproduction. Because of the ecological significance of silks, natural selection can act on silk fibers and thus the underlying silk proteins (fibroins). Lepidoptera (moths and butterflies) and Embioptera (webspinners) are two insect orders that utilize silk mostly for protection. Characterizing fibroins from distantly related insect lineages may reveal convergent evolutionary patterns and provide insight into the functional elements of their fibroins. Within a lineage, conserved characteristics may yield a better understanding of silk evolution and fiber formation.

I obtained fibroin sequence data through silk gland cDNA libraries of Lepidoptera and Embioptera. This was accomplished through construction of species-specific silk

gland libraries for the lepidopteran *Hepialus californicus* (ghost moth), and five embiopteran species: *Antipaluria urichi*, *Archembia* n. sp., parthenogenic *Haploembia solieri*, *Oligotoma nigra*, and *Saussurembia davisi*. After screening the libraries, transcripts for fibroins were identified for each species. *Hepialus californicus* heavy chain fibroin and light chain fibroin were compared to other lepidopteran and trichopteran heavy and light chain fibroins. Analyses revealed that both heavy and light chain fibroins have historically experienced stabilizing selection and light chain fibroin has potential as a phylogenetic marker. For the embiopteran fibroins, many similar features were found to be shared across species, such as glycine/serine rich, repetitive motifs, and a short, non-repetitive carboxyl-terminal region. Conservation of these elements suggests that these regions are important for fiber formation and mechanical properties. Furthermore, the conserved carboxyl-terminal regions imply that embiopteran fibroins have undergone stabilizing selection. Additional characterizations were performed on embiopteran silks to determine amino acid composition, secondary structural conformations, and mechanical properties. These features were found to have a high degree of similarity across taxa, suggesting that embiopteran silk has remained largely unchanged since it arose over 100 million years ago.

Table of Contents

List of Figures

List of Tables

Introduction

Background

My dissertation research focuses on the evolution of silks spun by Embioptera (webspinner insects) and Lepidoptera (moths and butterflies). With Embioptera, my goals are to characterize gene transcripts, secondary structural profiles, and mechanical properties of silks produced by multiple species. With Lepidoptera, my goals are to describe the silk gene transcripts from a ghost moth species, then perform comparative analyses with the silk molecules of other moths and Trichoptera (caddis flies, the sister-group to Lepidoptera). To place my research in the context of previous silk studies, I briefly review silk use and production in arthropods.

Silk is a fine, proteinaceous fiber that has independently arisen numerous times in the history of life, such as in Lepidoptera (moths, butterflies), Hymenoptera (bees, ants, wasps), Diptera (flies), and Araneae (spiders). Typically, silk is important for foraging, dispersal, defense, reproduction, and other vital tasks (Craig, 1997). Because of the importance of structures built with silk (e.g., capture webs, pupal cocoons, protective domiciles) to organismal survival and reproduction, silk should be subject to considerable natural selection (Fedič *et al.*, 2003; Swanson *et al.*, 2006). Thus, characterizing silks from divergent taxa will yield insight into functionally significant features by discerning which elements of fiber design have convergently evolved. Comparisons of silks produced by closely related organisms, however, is also informative, especially to determine how silks are fine-tuned for specific purposes and environments.

Most silks are polymers constructed from structural proteins (fibroins) that are held together largely by hydrogen bonding (Keten *et al.*, 2010). For example, spider

major ampullate (dragline) silk is assembled from a combination of major ampullate spidroin (contraction of spider fibroin) 1 and 2 (Hinman and Lewis, 1992; Bini *et al.*, 2004). In most lepidopteran silks, the basic polymer unit consists of three types of proteins, heavy chain fibroin, light chain fibroin, and P25 (Tanaka *et al.*, 1999). These three proteins form a 13-mer assembled from six heavy chain fibroins, six light chain fibroins, and one P25.

Silk fibroins are typically high molecular weight proteins constructed from non-essential amino acids (Rodriguez and Candelas, 1995). Specifically, most fibroins are dominated by glycine, alanine, and serine (Lewis, 1992; Gatesy *et al.*, 2001). These three amino acids in combination with a small number of other residues form motifs that are repeated throughout the fibroin sequence. For instance, major ampullate spidroins have numerous repeats of poly-alanine, glycine-alanine, glycine-proline-glycine, and/or glycine-glycine-X, where X represents a residue from a subset of amino acids (Hayashi *et al.*, 1999). These repeats form series of secondary structures that contribute to the mechanical properties of dragline silk. Additionally, some spidroins possess a higher level of organization with the amino acid motifs recurring in particular patterns that are called ensemble repeats (Hayashi and Lewis, 2001; Garb and Hayashi, 2005).

Fibroins are synthesized by silk glands and stored within them as liquid dope (Craig, 1997). As the dope flows through the glandular lumen, it is slowly concentrated and begins crystallizing (Vollrath and Knight, 2001). In moths, the fibroins aggregate through an acidification process in the glandular lumen and crystallization occurs just prior to spinning near the end of the gland duct (Wong Po Foo *et al.*, 2006). Upon being

drawn or sheared through a small orifice, commonly called a spigot, the liquid self-organizes into a fiber, which is then woven into capture webs, pupal cocoons, egg sacs, etc. (Craig 1997).

The morphology and developmental origin of silk glands vastly differ across arthropod lineages, providing evidence for the numerous origins of silk production rather than a single origin (Sutherland *et al.*, 2010). In silk-producing insects, silk glands seem to never evolve as *de novo* structures, but instead are derived from a diversity of existing organ systems that are of ectodermal origin (Sehnal and Akai, 1990; Sutherland *et al.*, 2010). For example, lepidopteran and trichopteran silks are produced in modified salivary glands, hymenopteran silk is manufactured in accessory genital glands, and dipteran silk arises from Malpighian tubules that have a primary function of excretion (Sehnal and Akai, 1990). Silk glands in Embioptera (webspinner insects) are derived from sensilla, which are sensory organs such as gustatory or olfactory receptors (Merritt, 2007).

Silk production within specialized, secretory glands often coincides with a specific life stage. For instance, lepidopteran larvae use silk to make escape lines for predator avoidance and build protective cocoons for pupation. Silk is spun into underwater snares and for stitching pebbles and other detritus together into protective domiciles by larval trichopterans (Brown *et al.*, 2004). Larval wasps utilize silk to create caps that provide cover and help with thermoregulation during pupation (Ishay *et al.*, 2002). In Hilarini (dance flies), adult males ball algae with silk for presentation to females as nuptial gifts (Young and Merritt, 2003). Unlike these insect examples, spiders employ silk throughout their lives for a variety of functions. Both female and male

spiders use silk for reproduction; females form egg sacs from silk to swaddle their eggs until the spiderlings emerge, and males rely on silk to transfer sperm from their gonads to their copulatory palps. Spiders also create silk capture snares to trap prey and some spiderlings disperse by ballooning with silk lines. Like spiders and unlike other insects, Embioptera use silk throughout their lifetime to spin silken galleries and tunnels in which they live, forage, and reproduce.

There are several ways that characterizing silk fibers and the underlying fibroins in detail will provide a better understanding of the function and evolution of these vital materials. First, silk mechanical properties can be quantified to not only compare and contrast different silks, but also to formulate hypotheses about how organisms use their silks. Second, deciphering the amino acid sequences of fibroins will yield information to develop models of fiber assembly mechanisms and the relationship of fibroin sequence elements to fiber performance. Third, silk fibroin secondary structural profiles can be determined to clarify the relationships between primary fibroin sequence and the observed physical characteristics of silk.

Characterization techniques

Measuring mechanical properties of biological materials not only quantifies their attributes, but also helps elucidate their importance to the organisms that use them. One method of mechanical characterization that is frequently applied to silk is tensile testing. Tensile tests provide information about the strength, extensibility, stiffness, and toughness of silk fibers. For example, Stauffer *et al.* (1994) report that egg case silk of

Araneus gemmoides has high tensile strength and is quite stiff. The stiffness may help protect the eggs during incubation because the ease of breaking or puncturing a sheet of fibers is related to initial stiffness (Termonia, 2006).

Tensile testing has provided insights on the different silks that an individual spider can produce. It has been shown that the various types of silk fibers that compose an aerial orb–web have different mechanical properties, which work in concert to ensnare prey (Gosline *et al.*, 1999). Dragline silk provides the structural support for the aerial web and exhibits high tensile strength and stiffness (Swanson *et al.*, 2006). By contrast, capture spiral silk (flagelliform) has lower tensile strength but demonstrates extreme extensibility (greater than 200%) that enables the web to trap prey by absorbing the energy of a flying insect, rather than letting the insect escape by ricocheting off or breaking through the web (Blackledge *et al.*, 2005). An orb web constructed entirely of major ampullate or capture spiral silk may not effectively capture an insect flying at high speeds, yet when used together, these silks help spiders efficiently capture food.

Studies on moth and spider fibroins have shown that specific secondary structures contribute to the observed mechanical properties of silk fibers. For example, β-sheets form β-crystalline regions that provide strength and stiffness to the fiber, while β-spirals form spring-like structures that expand when the fiber is stretched and contribute to extensibility (Savage and Gosline, 2009). The combination of these secondary structures leads to the impressive mechanical properties of spider dragline silk (Hayashi *et al.*, 1999).

A direct method of characterizing protein secondary structures in native silk fibers is Fourier transformed infrared spectroscopy (Dong *et al.*, 1991; Chang *et al.*, 2006). In this technique, the silk sample is pulsed with infrared radiation and the absorbance of the radiation is measured. The absorbance spectrum has peaks where the molecular bonds of the silk fibroins vibrated due to the radiation pulse. Correspondence is then made between the sample peaks and peak locations determined from reference proteins with known secondary structures (Cai and Singh, 2000). Fourier transformed infrared spectroscopy has been utilized to determine the secondary structures of caterpillar, leafhopper, and spider silks (Dong *et al.*, 1991; Chang *et al.*, 2006). Within one type of silk, the relative proportion of a variety of secondary structures can be deduced, including α-helices, β-sheets, and β-turns (Chang *et al.*, 2006).

In combination with tensile testing and Fourier transformed infrared spectroscopy, characterizing fibroin amino acid sequences allows inferences to be made about the relationship of specific amino acid motifs to secondary structures and their effect on mechanical properties. Additionally, comparisons of fibroin sequences within a lineage may reveal adaptive and neutral variation that provide clues regarding how fibers form and how these particular silk molecules have evolved. For example, all lepidopteran light chain fibroins have a similarly positioned cysteine residue for intermolecular bonding with heavy chain fibroin (Tanaka *et al.*, 1999). Deletion of this cysteine reduces fiber formation efficiency, resulting in the lack of production of a pupal cocoon (Mori *et al.*, 1995).

Comparisons among distantly related, silk-producing lineages could also be illuminating because functionally significant traits often evolve convergently. These analogous traits may represent optimal or near optimal solutions to common challenges of producing silk. As an example, many previously studied fibroins have carboxyl-terminal regions that are distinctly different from the motif-filled, repetitive region. The convergent evolution of non-repetitive carboxyl-terminal regions and features within those regions has been shown to impact the efficiency of fiber formation (Ittah *et al.*, 2006).

Determination of previously unknown protein sequences can be achieved through a variety of biochemical and molecular genetic approaches. The method that has been used most often to characterize silk fibroins is the construction of complementary DNA (cDNA) libraries from silk glands, the tissue in which the fibroin genes are abundantly expressed (Garb and Hayashi 2005; Yonemura and Sehnal, 2006). Identification of fibroin transcripts within the expression libraries occurs by screening and sequencing clones. The conceptual translations of the silk-positive cDNA will reveal the primary structure of the fibroins, including the repetitive amino acid motifs and their arrangement into ensemble repeats.

Study organisms

Embioptera is an order of hemimetabolous insects. Embiopterans are commonly referred to as embiids or webspinner insects. They have a worldwide distribution, with most of the ~360 named species occurring in the tropics (Engel and Grimaldi, 2006).

However, embiids are reclusive and often missed by general insect collecting techniques. Thus, their actual species richness is estimated to be several times greater, probably in the thousands (Ross, 2000).

Embiids create galleries from silk attached to a substrate, such as a tree or rock surface (Ross, 2000). Some tropical species construct galleries that can cover large patches of a tree's surface, at times more than one square meter (Edgerly, 1987). Within the gallery, webspinners develop, feed, hide from predators, and reproduce. Additionally, females spin protective silk coverings over their clutches of eggs (Edgerly, 1987; Ross, 2000).

The biology of Embioptera is little known and even less is known about their silk. Characterization of the silks from multiple species will allow comparisons among embiopteran silks and to the silks of other arthropods that may provide insights into how these insects manufacture and use silks. Because of their heavy reliance on silk, each species could have selective pressures unique to their ecology and environment that may lead to species-specific differences in silk mechanical properties, secondary structures, and fibroin genes. Furthermore, comparing embiopteran silk properties and genes to other arthropod silks will identify common features that are of potential functional and evolutionary significance.

In addition to characterizations of embiopteran silk, my dissertation research involves studies of lepidopteran silk. Lepidoptera (moths and butterflies) is a widely distributed order of holometabolous insects with over 150,000 described species. Caterpillars typically use silk to spin pupal cocoons and escape lines to avoid predators.

Lepidopteran silk glands synthesize the fibroins (heavy chain fibroin, light chain fibroin, P25) in the posterior section of the gland and the sericins, which compose the sticky coating of the silk fiber, are manufactured in the middle section.

Homologs of the three silk fibroins have been identified in several other lepidopterans (Tanaka and Mizuno, 2001; Žurovec and Sehnal, 2002; Yonemura and Sehnal, 2006). However, Lepidoptera is one of the most species-rich insect orders and characterizations of silk fibroins have been limited to a single major clade, Ditrysia. Therefore, I describe the fibroins of the ghost moth, *Hepialus californicus*, a non-ditrysian, and compare them to the fibroins of other lepidopterans and Trichoptera. Studying *Hepialus californicus* will identify which lepidopteran fibroins are present outside of Ditrysia and allow insight into fibroin adaptations in Lepidoptera and Trichoptera.

References

Bini, E., Knight, D.P., Kaplan, D.L., 2004. Mapping domain structures in silks from insects and spiders related to protein assembly. Journal of Molecular Biology. 335, 27-40.

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

Brown, S.A., Ruxton, G.D., Humphries, S.J.N., 2004. Physical properties of *Hydropsyche siltalai* (Trichoptera) net silk. Journal of the North American Benthological Society. 23, 771-779.

Cai, S. W., Singh, B. R. 1999. Identification of beta-turn and random coil amide III infrared bands for secondary structure estimation of proteins. Biophysical Chemistry. 80, 7-20.

Chang, J. C., Gurr, G. M., Fletcher, M. J., Gilbert, R. G. 2006. Structure-property and structure-function relations of leafhopper (*Kahaono montana*) silk. Australian Journal of Chemistry. 59, 579-585.

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

Dong, Z., Lewis, R.V., Middaugh, C. R., 1999. Molecular mechanism of spider silk elasticity. Archives of Biochemistry and Biophysics. 284, 53-57.

Edgerly, J.S., 1987a. Colony composition and some costs and benefits of facultatively communal behavior in a Trinidadian webspinner, *Clothoda urichi* (Embiidina: Clothodidae). Annals of the Entomological Society of America. 80, 29-34.

Engel, M.S., Grimaldi, D.A., 2006. The earliest webspinners (Insecta: Embiodea). American Museum Novitates. 3514, 1-15.

Fedič, R., Žurovec, M., Sehnal. F., 2003. Correlation between fibroin amino acid sequence and physical silk properties. Journal of Biological Chemistry. 278, 35255-35264.

Garb, J.E., Hayashi, C.Y., 2005. Modular evolution of egg case silk genes across orb-weaving spider superfamilies. Proceedings of the National Academy of Sciences of the United States of America. 102, 11379-11384.

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

Gosline, J., Guerette, P.A., Ortlepp, C.S., Savage, K.N., 1999. The mechanical design of spider silks: from fibroin sequence to mechanical function. Journal of Experimental Biology. 202, 3295-3303.

Hayashi, C.Y., Lewis, R.V., 2001. Spider flagelliform silk: lessons in protein design, gene structure, and molecular evolution. Bioessays. 23, 750-756.

Hayashi, C.Y., Shipley, N.H., Lewis, R.V., 1999. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. International Journal of Biological Macromolecules. 24, 271-275.

Hinman, M. B., Lewis, R. V., 1992. Isolation of a clone encoding a 2nd dragline silk fibroin - *Nephila clavipes* dragline silk is a 2-protein fiber. Journal of Biological Chemistry 267:19320-19324.

Ishay, J. S., Litinetsky, L., Pertsis, V., Linsky, D., Lusternik, V., Voronel, A., 2002. Hornet silk: thermophysical properties. Journal of Thermal Biology. 27, 7-15.

Ittah, S., Cohen, S., Garty, S., Cohn, D., Gat. U., 2006. An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. Biomacromolecules. 7, 1790-1795.

Keten, S., Xu, Z., Ihle, B., Buehler, M. J., 2010. Nanoconfinement controls stiffness, strength and mechanical toughness of beta-sheet crystals in silk. Nature Materials. 9, 359-367.

Lewis, R. V., 1992. Spider silk - the unraveling of a mystery. Accounts of Chemical Research. 25, 392-398.

Merritt, D.J., 2007. The organule concept of insect sense organs: Sensory transduction and organule evolution. Advances in Insect Physiology. 33, 192-241.

Mori, K., Tanaka, K., Kikuchi, Y., Waga, M., Waga, S., Mizuno, S., 1995. Production of a chimeric fibroin light-chain polypeptide in a fibroin secretion deficient naked pupa mutant of the silkworm *Bombyx mori*. Journal of Molecular Biology. 251, 217-228.

Rodriguez, R., Candelas, G. C., 1995. Flagelliform or coronata glands of *Nephila clavipes*. Journal of Experimental Zoology. 272, 275-280.

Ross, E.S., 2000. Contributions to the biosystematics of the insect order Embiidina. Part 1. Origin, relationships and integumental anatomy of the insect order Embiidina. Occasional Papers California Academy of Sciences. 149, 1-53.

Savage, K. N., Gosline, J. M., 2008. The effect of proline on the network structure of major ampullate silks as inferred from their mechanical and optical properties. Journal of Experimental Biology. 211, 1937-1947.

Sehnal, F., Akai, H., 1990. Insect silk glands - Their types, development and function, and effects of environmental factors and morphogenetic hormones on them. International Journal of Insect Morphology and Embryology. 19, 79-132.

Stauffer, S. L., Coguill, S.L., Lewis, R. V., 1994. Comparison of physical properties of three silks from *Nephila clavipes* and *Araneus gemmoides*. The Journal of Arachnology. 22, 5-11.

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, 171-188.

Swanson, B. O., Blackledge, T. A., Summers, A. P., Hayashi, C. Y., 2006b. Spider dragline silk: Correlated and mosaic evolution in high-performance biological materials. Evolution. 60, 2539-2551.

Tanaka, K., Kajiyama, N., Ishikura, K., Waga, S., Kikuchi, A., Ohtomo, K., Takagi, T., Mizuno, S., 1999. Determination of the site of disulfide linkage between heavy and light chains of the silk produced by *Bombyx mori*. Biochimica et Biophysica Acta. 1432, 92-103.

Tanaka, K., Mizuno, S., 2001. Homologues of fibroin L-chain and P25 of *Bombyx mori* are present in *Dendrolimus spectabilis* and *Papilio xuthus* but not detectable in *Antheraea yamamai*. Insect Biochemistry and Molecular Biology. 31, 665-677.

Termonia, Y., 2006. Puncture resistance of fibrous structures. International Journal of Impact Engineering. 32, 1512-1520.

Vollrath, F., Knight, D.P., 2001, Liquid crystalline spinning of spider silk. Nature. 410, 541-548.

Wong Po Foo, C., Bini, E., Hensman, J., Knight, D.P., Lewis, R.V., Kaplan, D.L., 2006. Role of pH and charge on silk protein assembly in insects and spiders. Applied Physics A-Materials Science and Processing. A82, 223-233.

Yonemura, N., Sehnal, F., 2006. The design of silk fiber composition in moths has been conserved for more than 150 million years. Journal of Molecular Evolution. 63, 42-53.

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

Žurovec, M., Sehnal, F., 2002. Unique molecular architecture of silk fibroin in the waxmoth, *Galleria mellonella*. Journal of Biological Chemistry. 277, 22639-22647.

Chapter 1

Characterization of silk spun by the embiopteran, *Antipaluria urichi*

Abstract

Silks are renowned for being lightweight materials with impressive mechanical properties. Though moth and spider silks have received the most study, silk production has evolved in many other arthropods. One insect group that has been little investigated is Embioptera (webspinners). Embiopterans produce silk from unique tarsal spinning structures during all life stages. We characterize the molecular and mechanical properties of *Antipaluria urichi* (Embioptera) silk through multiple approaches. First, we quantify the number of silk secretory structures on their forelimb and the tensile properties of *Antipaluria* silk. Second, we present silk protein (fibroin) transcripts from an embiopteran forelimb protarsomere cDNA library. We describe a fibroin that shares several features with other arthropod silks, including a subrepetitive core region, a non-repetitive carboxyl-terminal sequence, and a composition rich in glycine, alanine, and serine. Despite these shared attributes, embiopteran silk has several different tensile properties compared to previously measured silks. For example, tensile strength of *Antipaluria* silk is much lower than that of *Bombyx mori* silk. We discuss the observed mechanical properties in relation to the fibroin sequence, spinning system, and embiopteran silk use.

Keywords: biomechanics; cDNA; fibroin; silk; silk glands; webspinner

Introduction

Embioptera, commonly called webspinners or embiids, is an order of insects within Polyneoptera that is most closely related to Phasmatodea, the walking sticks (Terry and Whiting, 2005). There are ~360 named species of embiids (Engel and Grimaldi, 2006) with an estimated diversity of ~2000 species (Ross, 2000). Although webspinners have a worldwide distribution, the majority of species are located within the tropics. Based on the earliest definitive embiopteran fossil, a specimen encased in Burmese amber, the order dates to the mid–Cretaceous. However, because of the derived characters of this fossil, the order is likely to be much older (Engel and Grimaldi, 2006).

Embiopterans are distinctive in their ability to spin silk from unique tarsal organs (Fig. 1.1A). The forelimb protarsomeres of embiids are greatly enlarged and ultrastructural analysis reveals that these tarsi contain spherical secretory glands. Unlike lepidopterans, which have one pair of silk glands per individual, Ross (2000) estimates that the forelimb tarsi of the webspinner *Oligotoma nigra* have over 300 silk glands (150 per tarsus). The lumen of each embiid silk gland feeds into its own reservoir that has an individual duct leading to a single, hollow, setae-like cuticular process called a silk ejector (Alberti and Storch, 1976; Nagashima *et al.*, 1991). With over one hundred silk ejectors per tarsus, embiids can quickly spin extensive networks of silk (Fig. 1.1B).

Embiopterans use highly choreographed silk spinning behaviors to construct galleries in which they live, forage and reproduce (Edgerly *et al.*, 2002). Galleries consist of dense patches of silken tubes attached to a substrate, such as a tree (Fig. 1.1) or within leaf litter on the ground. Impressively, galleries that spread over one square meter have

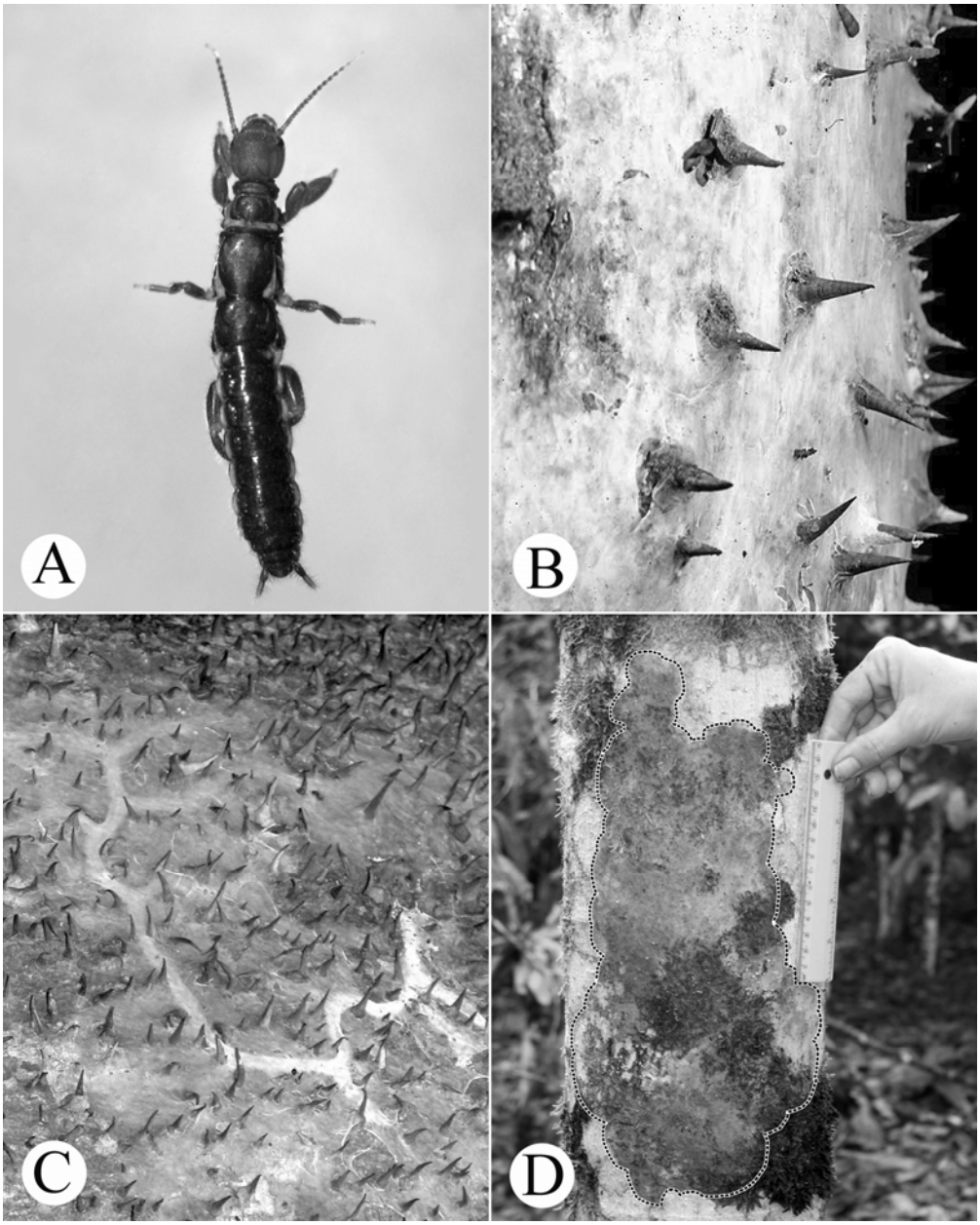


Fig. 1.1. Photographs of *Antipaluria urichi* (Clothodidae) and silk in the field. (A) Adult *Antipaluria* female (body length = 1.5 cm), showing typical embiid traits: elongate, juvenile form and enlarged front tarsi housing silk glands, (B) close-up of a mature colony of *Antipaluria* in Trinidad on thorny bark of *Hura crepitans* (Euphorbiaceae) showing the sheet-like quality of their silk, (C) tube-like galleries spun by *Antipaluria* extending their silk in search of food (epiphytic algae or lichens), and (D) camouflaged sheet-like silk of closely related *Clothoda longicauda* (Clothodidae) of Ecuador (ruler = 15 cm). Silk gallery, circumscribed by the dotted line, is coated with pulverized materials, pieces of moss, and other scraps gathered and incorporated into the silk by the embiids. White patches on the bark are lichens, partially consumed by *Clothoda*. (photographs by J. S. Edgerly and E. C. Rooks).

been documented on trees in the tropics, where embiids graze on lichens and epiphytic algae (Edgerly, 1987a). As fresh food sources are needed, embiids extend the foraging tunnels (Fig. 1.1C) that radiate outward from a more heavily-silked domicile where embiids typically hide during the day. Investment in silk production by adult females is a significant component of their maternal behavior, a trait that characterizes the order Embioptera. Within a gallery, females lay their eggs and either stitch the eggs into the silk or spin protective silk coverings over the egg clutches (Edgerly, 1987b). Upon emergence, the nymphs can spin silk and continue to do so throughout their lifetime (Ross, 2000).

In this paper, we describe the tensile properties and encoding cDNAs of silk from the embiopteran, *Antipaluria urichi* (Saussure) (Embioptera, Clothodidae) (Fig. 1.1A). We then compare *Antipaluria* silk to other arthropod silks to identify which features of webspinner silk are putatively unique or convergent. For example, some spider and moth silks have sequences that are rich in glycine and alanine (Gatesy *et al.*, 2001). These fibroins can be highly repetitive in amino acid sequence, often being composed of numerous iterations of short subrepeats (Gatesy *et al.*, 2001). Additionally, arthropod fibroins frequently have conserved regions at both the amino- and carboxyl-termini. In spiders, the termini are thought to assist in solubilizing the fibroins within the glandular lumen (Sponner *et al.*, 2005) and in moths, the termini bind other proteins involved in fiber formation (Inoue *et al.*, 2000). Finally, both moth and spider fibers have been shown to be remarkably strong and yet considerably extensible (Gosline *et al.*, 2002). Given the extreme reliance of embiids on their silk, molecular and mechanical characterization of

Antipaluria silk deepens our understanding of the structure-function relationships of arthropod fibroins.

Materials and Methods

Insect rearing

Embiids were cultured in plastic terrariums containing oak leaves (*Quercus* sp.) and fed romaine or red leaf lettuce. Terrariums were covered with fine mesh nylon (250 μm X 250 μm weave density) to prevent small nymphs from escaping. Cultures were watered two or three time a week with a spray bottle to keep leaf litter moist and maintain humidity.

Ejector number determination

Adult females were anesthetized with CO_2 gas and placed in 1.5 mL microfuge tubes filled with 0.15 M sodium chloride, 0.015 M sodium citrate (SSC) buffer. The tubes were placed in a sonication bath to clean debris from the insects. Specimens were then fixed in 2% gluteraldahyde in SSC buffer for two hours, followed by five washes in fresh SCC buffer. The insects were then placed in 1% osmium tetroxide in SSC for 2 hours. After rinsing with fresh SSC buffer, the specimens were dehydrated in increasing concentrations (30 – 100%) of ethanol and stored in 100% ethanol.

Fixed specimens were critical point dried in a Balzers CPD0202 (BAL-TEC AG, Liechtenstein) and then coated with gold palladium in a Cressington 108 AUTO (Cressington Scientific Instruments Ltd, Watford, UK) sputter coater. A Phillips XL30 FEG (FEI Co., Hillsboro, OR, USA) scanning electron microscope was used to image

tarsi. Pictures were taken of the whole plantar tarsal surface (57x, Fig. 1.2A) and at increased magnification (449x, Fig. 1.2B).

The number of ejectors was estimated from the SEM images using Image J software (Abramoff *et al.*, 2004). Starting with a single, high quality 449x image (Fig. 1.2B), the number of individual ejectors was counted. The area of the 449x image was then computed by multiplying the scaled length and width of the photo. The total tarsal surface area was determined from the whole tarsus image (57x) by using the Image J polygon area tool to trace around the edges of the tarsus. By dividing the total tarsal surface area by the 449x image area, the number of 449x images that would fit within the total tarsal area was calculated. This number was multiplied by the ejector count from the 449X image to estimate the total number of ejectors per tarsus.

Silk collection

Individual embiids were anesthetized using CO_2 gas and placed on their dorsal side atop a dissection microscope stage. C-shaped cards with 0.5 cm gaps (distance between the two arms) were dragged across a webspinner's tarsus to obtain silk fibers. Small drops of cyanoacrylate were dabbed on each arm of the c-shaped card to affix the silk sample. Each sample card was inspected with a compound light microscope at 100x and 1000x magnification, to determine the number of fibers spanning the gap. Cards with more than four fibers were discarded because of the intractability of distinguishing individual fibers for diameter measurements.

Tensile testing

Silk sample diameters were measured using polarized light microscopy (Blackledge *et al.*, 2005). Accuracy of the polarized light microscope measurements was confirmed using scanning electron microscopy on exemplar fibers (data not shown). Tensile tests were performed on a Nano Bionix® tensile tester (MTS Oakridge, TN, USA). Silk sample cards were mounted to the clamps of the tensile tester and connecting card material was removed such that only silk spanned the gap between the clamps. Fibers were axially stretched at a strain rate of 1% per second until failure. Tests were conducted at room temperature, which ranged from 23–26^{o}C, and relative humidity of 26–32%. Measurements collected included: tensile strength (True stress), the force needed to stretch a fiber to breaking; extensibility (True strain), the maximum elongation at breaking; stiffness, the material's resistance to stretching; and toughness, the amount of energy a given volume of material can absorb before breaking.

Silk cDNA characterization

Twenty adult female embiids were used for tissue collection. Live insects were anesthetized with CO_2 gas prior to removal of whole forelimb tarsi. The tarsi were excised at the joint between the first and second segments such that only the silk producing part of the forelimb was removed. The tarsi were immediately flash frozen in liquid N_2 and stored at –80°C.

Frozen tarsi were homogenized in TRIzol reagent (Invitrogen, Carlsbad, CA, USA) for RNA extraction. Total RNA was purified using the RNeasy mini kit (Qiagen, Valencia, CA, USA) and mRNA was pooled from total RNA using oligo $(dT)_{25}$ magnetic

beads (Invitrogen, Carlsbad, CA, USA). cDNA synthesis followed the Superscript II protocol (Invitrogen, Carlsbad, CA, USA) using an anchored oligo $(dT)_{18}V$ primer. The resulting cDNA were then size selected with a ChromaSpin 1000 (Clontech, Mountain View, CA USA) column and ligated into pZErO-2 plasmid (Invitrogen, Carlsbad, CA, USA) at the EcoRV site. The plasmids were electroporated into TOP10 *Escherichia coli* cells (Invitrogen, Carlsbad, CA, USA). Approximately 1800 recombinant colonies were arrayed into a cDNA library. The library was replicated onto nylon membranes. One third of the library was assayed to determine insert size (Beuken *et al.*, 1998). Clones with inserts greater than ~600 bp were chosen for sequencing. The entire library was screened with a γ-^{32}P radiolabeled probe (5'-CCAGAKCCTGAKCCTGCACC-3') designed from a putative fibroin cDNA discovered amongst the size-selected clone sequences.

Size-selected and probe positive clones were sequenced with the universal primers, T7 and Sp6, at the Core Instrumentation Facility at the University of California, Riverside. DNA sequences were translated using Sequencher 4.2 (Gene Codes, Ann Arbor, MI, USA) and submitted for BLASTX searches against the nr database (www.ncbi.nlm.nih.gov/BLAST). Predicted translations and codon usage analyses were performed in MacVector 7.0 (Oxford Molecular Group, Oxford, UK).

Amino acid analysis

Embiids were placed in petri dishes and allowed to spin silk. Silk samples were collected and sent to the Molecular Structure Facility at the University of California, Davis. The samples were hydrolyzed in 6N HCl, 0.1% phenol at 110^{o}C for 24 hours. Molar fractions were determined on a Hitachi L-8800 (Hitachi Ltd., Tokyo, Japan) amino

acid analyzer using ion exchange chromatography followed by ninhydrin reaction detection.

Northern blot hybridization

Total RNA was extracted using the methods described above, from the forelimb tarsi of six *Antipaluria* females. Total RNA was also extracted from a female that was intact except that her forelimb tarsi had been removed. 10 μL of total RNA was loaded into a 1.2% agrose gel with 3% formaldehyde. Equal loading of RNA samples and Ambion RNA Millennium markers (Ambion, Austin, TX, USA) were visualized using methylene blue staining. Northern blot procedure followed Ding *et al.*, (1995) and the membrane was probed with the same oligonucleotide as the cDNA library (see above). The membrane was exposed to a phosphor screen overnight and imaged using an Amersham Typhoon 9410 scanner (Amersham Biosciences, Piscataway, NJ, USA).

Results and Discussion

Ejector number determination

Silk producing organs vary widely across arthropod lineages. For example, spiders spin silk from abdominal glands, moths produce silk from modified salivary glands, and beetle larvae secrete silk from malpighian tubules (Craig, 1997). To visualize the details of the embiopteran silk production apparatus, we examined the *Antipaluria* silk spinning organ with SEM. Numerous silk ejectors were observed in the SEM image of the ventral tarsal surface (Fig.1.2A). The total tarsal surface area determined from Fig. 1.2A was 0.775 mm^2. From Fig. 1.2B, the number of ejectors was estimated by counting the longer,

setae-like projections on the tarsal surface. Given the total area of Fig. 1.2B (50,647.8 μm^2) and the number of ejectors in that area (15), there are approximately 230 ejectors per tarsus. This results in an ejector density for the ventral tarsal surface of roughly 1 ejector per 2900 μm^2. Assuming that each ejector is fed by its own silk gland, then *Antipaluria* has ~230 silk glands per tarsus. In comparison to the ~150 silk glands per *Oligotoma nigra* tarsus (Ross, 2000), *Antipaluria* has 1.5x more glands. This suggests that the number of silk glands in embiids may roughly scale with body length because *Antipaluria* is 1.6x longer than *Oligotoma* (Ross, 1957; Edgerly, 2002). Other factors that may contribute to the difference in ejector number could be the relationship of tarsal dimension to body size and variation in ejector density among species.

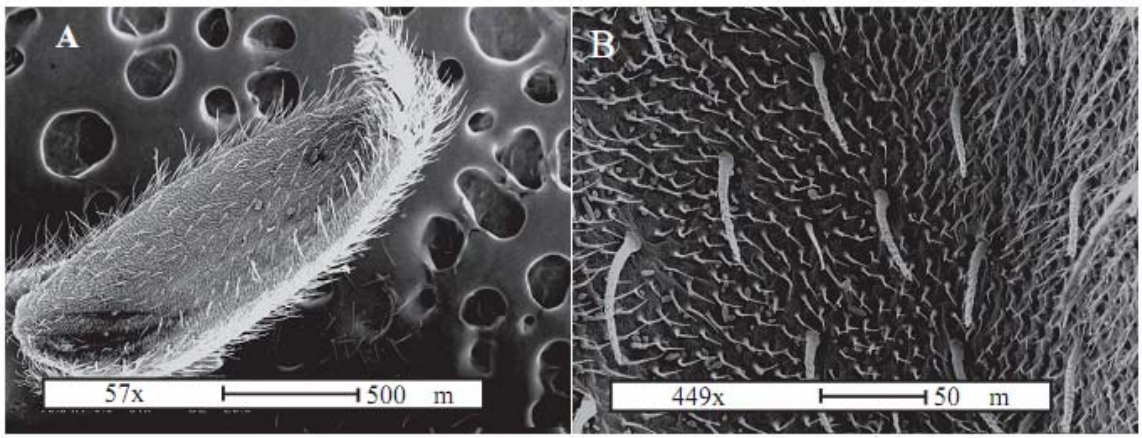


Fig. 1.2. SEM images of *Antipaluria* tarsus. (A) Whole *Antipaluria* tarsus and (B) higher magnification image of tarsal ventral surface showing prominent silk ejectors.

We determined that the mean diameter of an *Antipaluria* silk fiber is 0.805 μm ± 0.324 μm (s.d.), which is towards the lower end of the reported 0.1-11.5 μm range of arthropod silk diameters (Pérez-Rigueiro *et al.*, 1998; Young and Merritt, 2003; Blackledge and Hayashi, 2006b). Despite the minute size of its fibers, *Antipaluria* can readily spin large quantities of silk due to the impressive number of ejectors on their tarsi (~460 total per individual).

Tarsal silk production has independently evolved in tarantulas (Gorb *et al.*, 2006) and dance flies (Young and Merritt, 2003). In addition to their abdominally produced silks, the Costa Rican zebra tarantula (*Aphonopelma seemanni*) secretes silk through tarsal ejectors that assist these spiders during locomotion (Gorb *et al.*, 2006). Tarantula tarsal silk is secreted as a liquid and then solidifies quickly to prevent the animal from falling while climbing a vertical surface (Gorb *et al.*, 2006). In contrast, embiid silk appears to be secreted as a dry fiber and is not used in an adhesive fashion (Beutel and Gorb, 2001).

Silk production is sexually dimorphic in dance flies (Empidinae). Adult males use tarsal silk to ball algae and prey insects into packages, and then present these bundles to females as a prelude to mating (Young and Merritt, 2003). The male's enlarged basitarsus has twelve pairs of glands with each ejector producing silk fibers up to 3 μm in diameter (Young and Merritt, 2003). The spinning apparatuses of dance flies and embiids may be of similar developmental origin. For example, Sutherland *et al.* (2007) note that both embiopterans and dance flies produce silk through class III dermal glands. Additionally, in both taxa, tarsal silk glands have been hypothesized to be evolutionarily derived from

modified sensory organs (Merritt, 2007). However, the enlarged tarsi of a dance fly contain few silk glands surrounded by large volumes of haemolymph (Young and Merritt, 2003), while webspinner forelimb tarsi are entirely filled with over a hundred silk producing glands (Nagashima *et al.*, 1991).

Tensile properties

The impressive and diverse mechanical properties of arthropod silks have generated great interest in exploring their potential to inspire new types of high-performance biomimetic materials (Fedič *et al.*, 2003; Chang *et al.*, 2005; Blackledge and Hayashi, 2006b; Swanson *et al.*, 2006; Sutherland *et al.*, 2007). For example, some spider silks are tougher than steel or Kevlar, while others can extend over 200% (Gosline *et al.*, 2002). Independently evolved arthropod silks are likely to have different mechanical properties because of lineage-specific secretory mechanisms, molecular compositions, spinning behaviors, and ecological functions (e.g., pupation, dispersal, prey capture, predator avoidance, reproduction). Given the importance of silk to organismal fitness, it is expected that silk performance has been optimized through natural selection.

We tensile tested twelve *Antipaluria urichi* silk samples. The resulting stress-strain curves are characteristic of viscoelastic materials, such as linear polymers and tendons, that at first resist applied forces like a solid (initial slope of the stress-strain curve) and then reach a yield point exhibited by a change in slope (Fig. 1.3). After the yield point, the silk fiber behaves more like a viscous material and stretches until breaking.

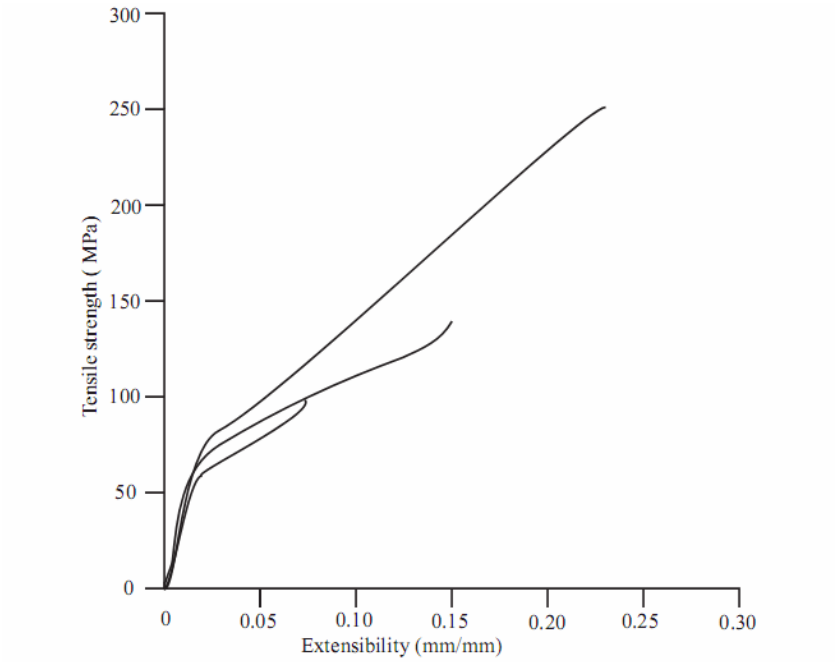


Fig. 1.3. Three representative stress-strain curves of *Antipaluria* silk tensile tests that depict the maximum, median, and minimum true stress values.

The extensibility of webspinner silk is comparable to that of *Bombyx* (silkworm) and *Kahaono* (leafhopper) silks (Table 1.1). These distantly related insects all employ their silks to build protective structures. Besides extensibility, silk stiffness may serve an important role in webspinner defense. Breaching a woven silk sheet by puncturing or pulling is dependent on the stiffness of the material (Termonia, 2006). Since many organisms that feed upon webspinners tear the silk sheets to obtain their prey, silk stiffness may be critical to impede predators and provide time for the embiids to flee. Leafhoppers and silkworms also use their silk to create protective structures, hence we hypothesize that similarity of function has led to roughly equivalent stiffness values (Table 1.1).

Table 1.1 Tensile properties of *Antipaluria* silk and other arthropod silks (*Hydropsyche*-Brown *et al.*, 2004; *Kahaono*-Chang *et al.*, 2005; *Bombyx*-Gosline *et al.*, 1999; *Kukulcania* and *Araneus* draglines-Swanson *et al.*, 2006). For *Antipaluria*, N=number of individuals from which silks were tested and n=number of tests.

	Tensile strength (MPa)	**Extensibility (ln mm/mm)**	**Stiffness (GPa)**	**Toughness (MJ/m^3)**
Antipaluria N=6, n=12	158.4 ± 43.3 (s.d.)	0.134 ± 0.08 (s.d.)	5.9 ± 2.4 (s.d.)	13.9 ± 8.5 (s.d)
Hydropsyche (caddisfly)	221 ± 22 (s.d.)	1.16 ± 0.10 (s.d.)	–	–
Kahaono (leafhopper)	280 ± 72	0.069 ± 0.01	4.4 ± 1.6	–
Bombyx (silkworm)	600	0.18	7	70
Kukulcania (house spider)	830 ± 6 (s.e.m)	0.26 ± 0.014 (s.e.m)	22.2 ± 1.52 (s.e.m)	132.2 ± 7.53 (s.e.m)
Araneus (orb-weaving spider)	1050 ± 5 (s.e.m)	0.29 ± 0.014 (s.e.m)	8.3 ± 0.54 (s.e.m)	141.2 ± 0.77 (s.e.m)

While the extensibility and stiffness of *Antipaluria* silk is comparable to other arthropod silks, the tensile strength of webspinner silk is markedly lower (Table 1). Because toughness is a characteristic dependent on extensibility, stiffness, and strength (quantified by the area under the stress-strain curve, Fig. 3), the weak strength of embiid silk results in modest toughness. Embiids simultaneously spin hundreds of fibers to construct their galleries and, to our knowledge, never suspend themselves with silk. Spiders and moths however, use paired filaments for high tensile, energy absorbing applications such as orb-weaving spiders capturing flying insects with aerial webs (Blackledge and Hayashi, 2006a) or caterpillars releasing an escape line to drop quickly from predators (Sugiura and Yamazaki, 2006). Thus, in contrast to these arthropod silks, it is likely that embiid silk is not under selection for high tensile strength or toughness.

Amino Acid Sequence

We found six clones in our cDNA library that encode the same putative silk protein. The longest clone had 753 nucleotides of coding sequence and included the stop codon. These clones were identified based on sequence similarity to other arthropod silks. Specifically, as with a number of known silks, the predicted translation of the *Antipaluria* fibroin cDNA was primarily composed of glycine, serine and alanine (Fig.1.4A; **Accession FJ361212**).

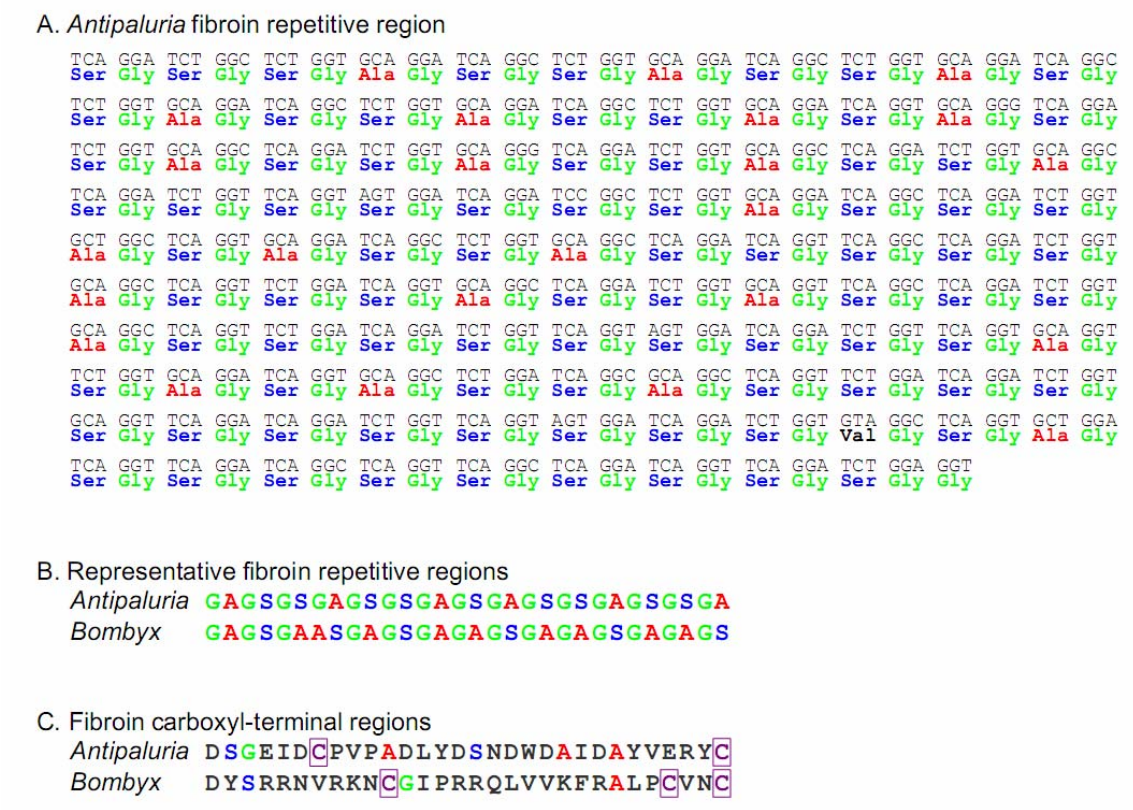


Fig. 1.4. *Antipaluria* fibroin partial-length cDNA sequence and translation (**Accession # FJ361212**). (A) Nucleotides (one-letter abbreviations) grouped into codons with amino acid (aa) translation (three-letter abbreviations) beneath. (B) Exemplar repetitive regions of *Antipaluria* fibroin and *Bombyx mori* heavy-chain fibroin (**Accession # NM_001113262**) shown by one-letter aa abbreviations. (C) Fibroin carboxyl-terminal regions shown by one-letter aa abbreviations. Glycine, serine, and alanine are shown in green, blue, and red, respectively. Cysteine is purple and boxed.

The predicted amino acid composition of the putative fibroin closely matched the empirically determined amino acid composition of spun silk (Fig. 1.5). The slight discrepancy between the molar percentages of spun silk and the fibroin cDNA translation is possibly due to the cDNA being partial length and lacking amino-terminal coding sequence. Additionally, the presence of other proteins in spun silk may partially explain the difference. Moth silks, for example, are known to be composed of accessory proteins as well as fibroins (Žurovec *et al.*, 1998; Nirmala *et al.*, 2001).

Examining the *Antipaluria* fibroin cDNA in more detail, it is evident that strings of codons appear in highly repetitive groupings (Fig. 1.4A). A particular group of codons (GGT GCA GGA TCA GGC TCT) appears six times within the sequence. Dominating the fibroin sequence are 21 variants of this codon motif, 17 of which differ by two nucleotides or less. This codon motif encodes Gly-Ala-Gly-Ser-Gly-Ser, the most prevalent peptide motif in the fibroin.

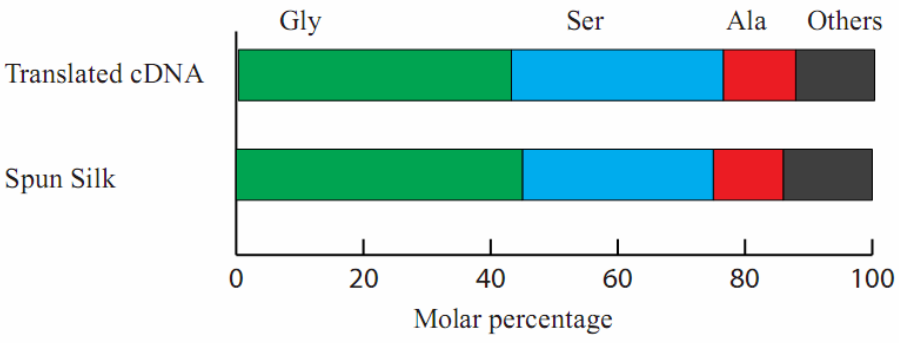


Fig. 1.5. Predicted amino acid composition of the translated *Antipaluria* fibroin cDNA in comparison to the empirically determined composition of *Antipaluria* spun silk. Molar percentages of glycine, serine, alanine, and all other amino acids are shown in green, blue, red, and grey, respectively.

Table 1.2. Molar percentage and codon usage for the most common amino acids (aa) in the predicted translation of the *Antipaluria* fibroin cDNA.

aa	% aa	codon	% codon
Gly	44	GGA	36
		GGC	23
		GGG	2
		GGT	39
Ser	34	AGC	1
		AGT	5
		TCA	58
		TCC	1
		TCG	0
		TCT	35
Ala	12	GCA	86
		GCC	0
		GCG	0
		GCT	14

Codon usage within the fibroin cDNA is highly skewed (Table 1.2). There is a strong preference for glycine, alanine, and serine codons to have thymine (T) and adenosine (A) located in the third nucleotide position. This same pattern has been observed in convergently evolved silks (Ayoub *et al.*, 2007; Sutherland *et al.*, 2007). Some causes suggested for codon usage bias include adaptation to tRNA frequencies within the cytoplasm (Rocha, 2004) and mutational bias of the organism towards particular transitions and transversions (Murray *et al.*, 1989). Another reason would be to reduce frame shift mutations. Given the repeated codon motifs in the *Antipaluria* fibroin cDNA described above, the embiid silk gene has attributes similar to microsatellites. Changes in microsatellite repeat number often occur because of mutational events, such as slip-strand mispairing (Taylor and Breden, 2000). In coding sequence, slip-strand mispairing could result in a deleterious frameshift. Given the importance of silk to embiid fitness, mechanisms that reduce frameshift mutations within the sequence would be

beneficial. The observed codon usage bias (Table 1.2) might act as a mechanism to reduce frameshifts in repetitive fibroin genes. Without the interspersed Ts and As in the glycine, alanine, and serine codons, the gene sequence would be dominated by cytosine (C) and guanine (G) bases. A G/C rich sequence is expected to experience frequent frameshifts because Gs and Cs can partner anywhere along the length of the sequence. Because of codon usage bias, the *Antipaluria* fibroin transcript is not G/C rich and the distribution of Ts and As may help the sequence remain in frame after slip-strand mispairing or unequal crossover events. If such mutations occurred within a silk sequence, then variation in the length of silk genes would be observed among individuals. In fact, allelic variants of spider and silkworm fibroin genes have been observed that are akin to the length variations seen in microsatellite loci (Ueda *et al.*, 1985; Higgins *et al.*, 2007). Similar allelic variation is expected in the webspinner silk gene.

We observed similarities between the repetitive regions of the webspinner fibroin and the *Bombyx mori* heavy chain (HC) fibroin repetitive region (**Accession # NM_001113262**). Both silks are composed of short repeats of glycine, serine, and alanine (Fig 1.4B). Iterations of glycine-alanine couplets and poly-alanine repeats are thought to form β-crystalline structures, which contribute to tensile strength (Warwicker, 1960; Hayashi *et al.*, 1999; Fedič *et al.*, 2003; Lawrence *et al.*, 2004, Trancik *et al.*, 2006). Craig (1997) hypothesized that webspinner silk contains crystallites based on sheering forces that occur during spinning. Our sequence results are consistent with this hypothesis. Despite the similarity in their protein sequences, *Antipaluria* silk is weaker than *Bombyx* silk. This performance difference may result from disparities in the

percentage of β-sheet secondary structure, crystallite size (Termonia, 1994), or molecular weight (Termonia *et al.*, 1985; see below).

The carboxyl-terminal region of the embiid fibroin has features observed in many other silks (Fig. 1.4C; Sehnal and Žurovec, 2004; Garb *et al.*, 2006; Ittah *et al.*, 2006; Ayoub *et al.*, 2007). One such feature is the non-repetitive and hydrophilic nature of the carboxyl-terminal region (Fig. 1.4C). It has been hypothesized that the hydrophilic amino acids in this region aid in the solubility of silk proteins in the glandular lumen (Sponner *et al.*, 2005). Another possibly convergent feature of the webspinner fibroin carboxyl-terminal region is the presence of cysteine residues (purple and boxed in Fig. 1.4C). In *Bombyx* and other lepidopteran HC fibroins, similarly positioned cysteine residues form disulfide bonds that join silk monomers (Inoue *et al.*, 2000; Yonemura and Sehnal, 2006). Additionally, spider major ampullate silk has a conserved carboxyl-terminal cysteine that has been demonstrated to be important in protein-protein interactions leading to fiber formation (Ittah et al 2006). Given the importance of the carboxyl-terminal region in silk fiber formation and the observed similarities across divergent taxa, we hypothesize that these features can be explained by convergence. We suggest that due to the analogous placement of cysteines in the *Antipaluria* fibroin, these residues may play a key role in fiber formation. As more webspinner fibroins are characterized, it is expected that their carboxyl-terminal regions will contain the highly conserved cysteine residues.

Northern Blot

The transcript size of the *Antipaluria* silk fibroin is in the range of 1.5-2.5 kb (Fig. 1.6). The lack of a narrow, sharp band may be due to overexposure because the probe, which was specific to the repetitive coding region, hybridizes numerous times within each fibroin transcript. Alternatively, the indistinct band on the blot image may result from the necessity to use multiple individuals to acquire enough total RNA. Individuals may vary in transcript length, as would be consistent with the hypothesis that fibroin genes can readily change length because of their microsatellite-like sequence characteristics (Garb and Hayashi, 2005; Ayoub *et al.*, 2007).

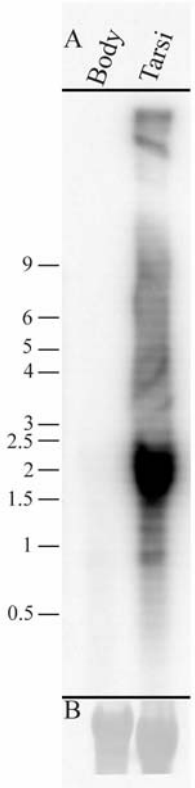


Fig. 1.6. Northern blot analysis of the *Antipaluria* fibroin transcript. (A) Tarsal-specific expression of transcript with sizes (in kb) indicated on the left. (B) Verification of equal loading of RNA in each lane using ribosomal bands visualized with methylene blue stain, scanned in black and white.

The determination of the *Antipaluria* fibroin transcript size provides insight regarding the observed tensile strength. Theoretical and empirical studies of synthetic polymers have shown that strength is greatly affected by the molecular weight of the monomers (Termonia *et al.*, 1985; Hallam *et al.*, 1986). Larger monomeric subunits increase tensile strength through greater molecular interaction via hydrogen bonding (Termonia *et al.*, 1985). In *Bombyx*, the transcript size is 15 kb (Ohshima and Suzuki, 1977), which is approximately an order of magnitude larger than the *Antipaluria* fibroin transcript. This suggests that the much smaller protein size of the *Antipaluria* fibroin may explain the lower tensile strength of *Antipaluria* silk compared to *Bombyx* silk.

In summary, we characterized the molecular and mechanical properties of *Antipaluria urichi* silk. This silk has remarkably convergent similarities and surprising differences in comparison to other arthropod silks. Like many silks (e.g., Chang *et al.*, 2005; Ayoub *et al.*, 2007), embiid silk is dominated by glycine, alanine, and serine (Fig. 1.5) and is encoded by a transcript with a highly skewed codon usage (Table 1.2). We suggest that this codon usage bias in fibroin genes is important for minimizing the possibility of frame shift mutations within these highly repetitive sequences. Also, the *Antipaluria* carboxyl-terminal region has paired cysteine residues that have analogs, which are necessary for fiber formation, in the *Bombyx mori* HC fibroin (Fig. 1.4C; Inoue *et al.*, 2000). Despite these similarities, *Antipaluria* silk has lower tensile strength and toughness compared to convergently evolved silks (Gosline *et al.*, 2002; Chang *et al.*, 2005; Swanson *et al.*, 2006). These differences in tensile properties may be largely explained by the effects of molecular weight on tensile strength. By investigating both the

tensile properties, encoding transcript and transcript size of *Antipaluria* silk, we gain insight into the functional significance of the embiid fibroin sequence elements. Webspinners are vitally dependent on their silk and they possess a radically different spinning system from the more frequently studied moths and spiders. Examining convergently evolved silks highlights recurring themes in the evolution of these adaptive molecules.

References

Abramoff, M.D., Magelhaes, P.J., Ram, S.J., 2004. Image processing with ImageJ. Biophotonics International. 11, 36-42.

Alberti, G., Storch, V., 1976. Ultrastructural investigations on silk glands of Embioptera. Zoologicscher Anzeigr. 197, 179-186.

Ayoub, N.A., Garb, J.E., Tinghitella, R.M., Collin, M.A., Hayashi, C.Y., 2007. Blueprint for a high-performance biomaterial: Full-length spider dragline silk genes. PLoS ONE. 2, e514.

Beuken, E., Vink, C., Bruggeman, C.A., 1998. One-step procedure for screening recombinant plasmids by size. Biotechniques. 24, 748-750.

Beutel, R.G., Gorb, S. N., 2001. Ultrastructure and attachment specializations of hexapods (Arthopoda): evolutuionary patterns inferred from a revised ordinal phylogeny. Journal of Zoological Systematics and Evolutionary Research. 39, 177-207.

Blackledge, T.A., Cardullo, R.A., Hayashi, C.Y., 2005. Polarized light microscopy, variability in spider silk diameters, and the mechanical characterization of spider silk. Invertebrate Biology. 124, 165-173.

Blackledge, T.A., Hayashi, C.Y., 2006a. Silken toolkits: biomechanics of silk fibers spun by the orb web spider *Argiope argentata* (Fabricius 1775). Journal of Experimental Biology. 209, 2452-2461.

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

Brown, S.A., Ruxton, G.D., Humphries, S.J.N., 2004. Physical properties of *Hydropsyche siltalai* (Trichoptera) net silk. Journal of the North American Benthological Society. 23, 771-779.

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

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

Ding, S-W., Li, W-X., Symons, R. H., 1995. A novel naturally occurring hybrid gene encoded by a plant RNA virus facilitates long distance virus movement. EMBO Journal. 14, .5762-5772.

Edgerly, J.S., 1987a. Colony composition and some costs and benefits of facultatively communal behavior in a Trinidadian webspinner, *Clothoda urichi* (Embiidina: Clothodidae). Annals of the Entomological Society of America. 80, 29-34.

Edgerly, J.S., 1987b. Maternal behavior of a webspinner (Order Embiidina). Ecological Entomology. 12: 1-11.

Edgerly, J.S., Davilla, J.A., Schoenfeld, N., 2002. Silk spinning behavior and domicile construction in webspinners. Journal of Insect Behavior. 15, 219-242.

Engel, M.S., Grimaldi, D.A., 2006. The earliest webspinners (Insecta: Embiodea). American Museum Novitates. 3514, 1-15.

Fedič, R., Žurovec, M., Sehnal, F., 2003. Correlation between fibroin amino acid sequence and physical silk properties. Journal of Biological Chemistry. 278, 35255-35264.

Garb, J.E., Hayashi, C.Y., 2005. Modular evolution of egg case silk genes across orb-weaving spider superfamilies. Proceedings of the National Academy of Sciences of the United States of America. 102, 11379-11384.

Garb, J.E., DiMauro, T., Vo, V., Hayashi, C.Y., 2006. Silk genes support the single origin of orb webs. Science. 312, 1762-1762.

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

Gorb, S.N., Niederegger, S., Hayashi, C.Y., Summers, A.P., Vöetsch, W., Walther, P., 2006. Silk-like secretion from tarantula feet. Nature. 443, 407-407.

Gosline, J., Guerettte, P.A., Ortlepp, C.S., Savage, K.N., 1999. The mechanical design of spider silks: from fibroin sequence to mechanical function. Journal of Experimental Biology. 202, 3295-3303.

Gosline, J., Lillie, M., Carrington, E., Guerette, P., Ortlepp, C., Savage, K., 2002. Elastic proteins: biological roles and mechanical properties. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. 357, 121-132.

Hallam, M. A., Cansfield, D.L.M., Ward, I.M., Pollard, G., 1986. A study of the effect of molecular weight on the tensile strength of ultra-high modulus polyethylenes. Journal of Materials Science. 21, 4199-4205.

Hayashi, C.Y., Shipley, N.H., Lewis, R.V., 1999. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. International Journal of Biological Macromolecules. 24, 271-275.

Higgins, L.E., White, S., Nuñez-Farfán, J., Vargas, J., 2007. Patterns of variation among distinct alleles of the *Flag* silk gene from *Nephila clavipes*. International Journal of Biological Macromolecules. 40, 201-216.

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. 275:40517-40528.

Ittah, S., Cohen, S., Garty, S., Cohn, D., Gat, U., 2006. An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. Biomacromolecules. 7, 1790-1795.

Lawrence, B.A., Vierra, C.A., Moore, A.M.F., 2004. Molecular and mechanical properties of major ampullate silk of the black widow spider, *Latrodectus hesperus*. Biomacromolecules. 5, 689-695.

Merritt, D.J., 2007. The organule concept of insect sense organs: Sensory transduction and organule evolution. Advances in Insect Physiology. 33, 192-241.

Murray, E.E., Lotzer, J., Eberle, M., 1989. Codon usage in plant genes. Nucleic Acids Research. 17, 477-498.

Nagashima, T., Niwa, N., Okajima, S., Nonaka, T., 1991. Ultrastructure of silk gland of webspinners *Oligotoma japonica* (Insecta, Embioptera). Cytologia. 56, 679-685.

Nirmala, X., Mita, K., Vanisree, V., Žurovec, M., Sehnal, F., 2001. Identification of four small molecular mass proteins in the silk of *Bombyx mori*. Insect Molecular Biology. 10, 437-445.

Ohshima, Y., Suzuki, Y., 1977.Cloning of the silk fibroin gene and its flanking sequences. Proceedings of the National Academy of Sciences of the United States of America. 74, 5363-5367.

Pérez -Rigueiro, J., Viney, C., Llorca, J., Elices, M., 1998. Silkworm silk as an engineering material. Journal of Applied Polymer Science. 70, 2439-2447.

Rocha, E.P.C., 2004. Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. Genome Research. 14, 2510-2510.

Ross, E.S., 1957. The Embioptera of California. Bulletin of the California Insect Survey. 6, 51-57.

Ross, E.S., 2000. Contributions to the biosystematics of the insect order Embiidina. Part 1. Origin, relationships and integumental anatomy of the insect order Embiidina. Occasional Papers California Academy of Sciences. 149, 1-53.

Sehnal, F., Žurovec, M., 2004. Construction of silk fiber core in Lepidoptera. Biomacromolecules. 5, 666-674.

Sponner, A., Vater, W., Rommerskirch, W., Vollrath, F., Unger, E., Grosse, F.,Wiesshart, K., 2005. The conserved C-termini contribute to the properties of spider silk fibroins. Biochemical and Biophyscial Research Communications. 338, 897-902.

Sugiura, S., Yamazaki, K., 2006. The role of silk threads as lifelines for caterpillars: pattern and significance of lifeline-climbing behaviour. Ecological Entomology. 31, 52-57.

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, 1036-1043.

Swanson, B.O., Blackledge, T.A., Beltran, J., Hayashi, C.Y., 2006. Variation in the material properties of spider dragline silk across species. Applied Physics a-Materials Science & Processing. 82, 213-218.

Taylor, J.S., Breden, F., 2000. Slipped-strand mispairing at noncontiguous repeats in *Poecilia reticulata*: A model for minisatellite birth. Genetics. 155, 1313-1320.

Terry, M.D., Whiting, M.F., 2005. Mantophasmatodea and phylogeny of the lower neopterous insects. Cladistics. 21, 240-257.

Trancik, J.E., Czernuszka, J.T., Bell, F.I., Viney, C., 2006. Nanostructural features of a spider dragline silk as revealed by electron and X-ray diffraction studies. Polymer. 47, 5633-5642.

Termonia, Y., 1994. Molecular modeling of spider silk elasticity. Macromolecules. 27, 7378-7381.

Termonia, Y., 2006. Puncture resistance of fibrous structures. International Journal of Impact Engineering. 32, 1512-1520.

Termonia, Y., Meakin, P., Smith, P., 1985. Theoretical study of the influence of molecular weight on the maximum tensile strength of polymer fibers. Macromolecules. 18, 2246-2252.

Ueda, H., Mizuno, S., Shimura, K., 1985. Sequence polymorphisms around the 5'-end of the silkworm fibroin H-chain gene suggesting the occurrence of crossing-over between heteromorphic alleles. Gene. 34, 351-355.

Warwicker, J.O., 1960. Comparative studies of fibroins 2. Crystal structures of various fibroins. Journal of Molecular Biology. 2, 350-362.

Yonemura, N., Sehnal, F., 2006. The design of silk fiber composition in moths has been conserved for more than 150 million years. Journal of Molecular Evolution. 63, 42-53.

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

Žurovec, M., Kodrik, D., Yang, C., Sehnal, F., Scheller, K., 1998. The P25 component of *Galleria* silk. Molecular & General Genetics. 257, 264-270.

Chapter 2

Comparison of embiopteran silks reveals tensile and structural similarities across taxa

Abstract

Embioptera is a little studied order of widely distributed, but rarely seen insects. Members of this group, also called embiids or webspinners, all heavily rely on silken tunnels in which they live and reproduce. However, embiids vary in their substrate preferences and these differences may result in divergent silk mechanical properties. Here, we present diameter measurements, tensile tests, and protein secondary structural analyses of silks spun by several embiid species. Despite their diverse habitats and phylogenetic relationships, these species have remarkably similar silk diameters and ultimate stress values. Yet, ultimate strain, Young's modulus, and toughness vary considerably. To better understand these tensile properties, Fourier transformed infrared spectroscopy was used to quantify secondary structural components. Compared to other arthropod silks, embiid silks are shown to have consistent secondary structures, suggesting that commonality of amino acid sequence motifs and small differences in structural composition can lead to significant changes in tensile properties.

Keywords: biomechanics; silk; webspinner; Fourier transformed infrared spectroscopy; fibroin

Introduction

Silk has evolved multiple times and in all cases appears to be of key ecological importance. For example, spiders rely heavily on silk for prey capture, dispersal, reproduction, and protection (Foelix, 1996). Similarly, lepidopteran larvae utilize silk for pupation cocoons and escape lines. Other silk-producing arthropods include caddisfly larvae that construct underwater capture nets and protective domiciles with silk, (Brown *et al.*, 2004) and mites that require silk for egg attachment, dispersal, and nest construction (Jung and Croft, 2001; Mori and Saito, 2005).

Embioptera (webspinners or embiids) are insects that also vitally depend on silk for survival. Embiids produce silk from specialized glands located within their forelimb tarsi and silk is spun by rapidly moving their forelimbs in a highly choreographed manner (Edgerly *et al.*, 2002). The most extraordinary aspect of webspinner ecology is that they are the only insects able to produce silk throughout their entire life span (Edgerly, 1987). Webspinners spin their silk into galleries consisting of sheets or tubes that provide protection from predators and abiotic threats such as heavy rainfall. These galleries are domiciles in which embiids live together in gregarious colonies composed of reproductive females and their offspring. Webspinner galleries are found in a variety of habitats, most typically on the bark of trees, or within leaf litter, beneath rocks, and underground (Ross, 2000). Embiids feed on detritus, lichens, or epiphytic algae, which they cover with silk as they graze, expanding their foraging zones outward from their domiciles.

Embiid silk is used in fibrous form. The mechanical performance of such fibers can be quantified by tensile testing. Through this type of characterization, particular spider and lepidopteran silks have been shown to possess remarkably high tensile strength or extensibility (Gosline *et al.*, 2002; Swanson *et al.*, 2006a). For example, orb-weaving spiders use dragline silk to frame their webs. The tensile strength of this silk enables the web to withstand impact by flying insects. In contrast, the capture spiral silk of orb-webs is exceptionally stretchy, a property essential to slow and stop flying insects rather than allow prey to escape by ricocheting off or breaking through an overly stiff web. The different properties of these silks are thought to have evolved through selection for divergent functions (Blackledge and Hayashi, 2006; Swanson *et al.*, 2006b; Swanson *et al.*, 2007). Lepidopterans, unlike spiders with their multitude of task-specific silk glands, have just one pair of spinning organs. Thus lepidopterans produce a single silk type that in some species is used by several larval instars. This multipurpose silk has a combination of high tensile strength (but less so than orb-weaver dragline silk) to support the larva during escape line use and high stiffness to protect the larva during pupation (Pérez-Rigueiro *et al.*, 2001; Sugiura and Yamazaki, 2006).

Silk fibers are primarily composed of fibroins, structural proteins characteristically rich in glycine, alanine, and serine (Gatesy *et al.*, 2001). Fibroins typically contain highly repetitive amino acid sequences (called motifs) and an individual fibroin may consist largely of runs of a single motif, such as *Bombyx* silkworm heavy chain fibroin (Mita *et al.*, 1994). Alternatively, fibroins may possess multiple types of motifs, as found in black widow spider dragline silk proteins (Ayoub *et al.*, 2007). The

amino acid motifs in fibroins have been hypothesized to result in secondary structures that contribute to the mechanical properties of silk fibers (Denny, 1980; Gosline *et al.*, 1986; Hayashi *et al.*, 1999). For example, poly-alanine (A_n) and glycine-alanine (GA) couplets form β-crystalline structures that underlie the impressive tensile strength of silks. Additionally, the motifs containing glycine-proline-glycine in spider dragline silk result in β-spirals that are predicted to provide elastin-like properties (Hayashi *et al.*, 1999). Therefore, determining the type and proportion of secondary structures in silk fibers can yield insights into the basis for tensile differences among silks.

In this paper, we present performance and structural quantification of embiopteran silk. We use tensile testing to characterize the mechanical properties of silk from six species of Embioptera. Additionally, we determine silk secondary structure through Fourier transformed infrared spectroscopy (FTIR). With these complementary techniques, we explore the relationship between the observed tensile properties and the secondary structure of embiopteran silk. These characteristics will be further discussed in the context of other arthropod silks.

Methods

Taxonomic sampling and insect rearing

We reared *Antipaluria urichi* (Clothodidae), *Archembia* nova species (Archembiidae), *Australembia incompta* (Australembiidae), *Oligotoma nigra* (Oligotomidae), parthenogenetic *Haploembia solieri* (Oligotomidae), and *Saussurembia davisi* (Anisembiidae). Average body lengths of adult females are, respectively, 1.53,

1.26, 1.18, 0.99, 0.92 and 0.95 cm. These six species represent five of the nine named embiopteran families and were chosen for broad phylogenetic sampling across the order (Fig. 2.1; Szumik et al., 2008). In particular, *Antipaluria urichi* was selected because it is a member of Clothodidae, the sister group to all other embiid families. Vouchers for *Archembia* n. sp. and parthenogenetic *Haploembia solieri* are deposited at Museum of Southwestern Biology, Albuquerque, NM. For brevity, taxa are henceforth referred to by genus name throughout this article and in Supplementary Figures.

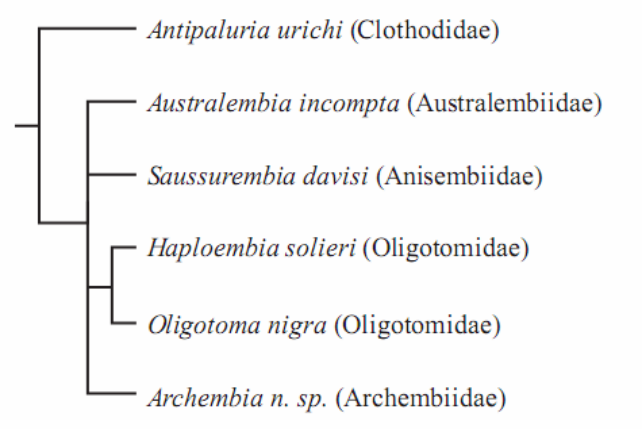


Fig. 2.1. Phylogenetic relationships of species in this study based on Szumik *et al.* (2008).

Embiids were cultured in plastic terrariums containing oak leaves (*Quercus* sp.) and fed romaine or red leaf lettuce (*Lactuca sativa*). Terrariums were covered with fine mesh nylon (250 μm X 250 μm weave density) to prevent small nymphs from escaping. Colonies were misted with water two or three times a week to moisten leaf litter and maintain humidity.

Silk collection

Silk was obtained from *Haploembia* adult females by allowing individuals to lay silk while walking in clean petri dishes. The perimeter of each dish was lined with approximately sixteen C-shaped pieces of poster board. Each C-shaped card had a 1 cm gap between the arms. Embiids were introduced into the dish and left to spin on the cards overnight. The next morning, each card was examined for silk fibers under a dissection microscope. When silk was detected, the ends of the fibers were secured to the card with small drops of cyanoacrylate. After allowing the glue to dry, excess fibers were carefully cleared away to ensure that only a small number of fibers stretched across the gap. Ideally, samples were single fibers. Inspection with light microscopy (100x and 1000x magnification) was then done to verify the number of fibers on each sample card.

To obtain silk fibers from *Antipaluria*, *Saussurembia*, *Oligotoma*, *Australembia*, and *Archembia*, a manual silking technique was developed. It was not possible to use the same collection method as with *Haploembia* because these five species tended to remain motionless when introduced into petri dishes. Hence, they did not naturally lay down silk onto sample cards. The manual silking was done by anesthetizing adult female embiids with carbon dioxide gas and placing them on their dorsal sides atop a dissection microscope stage. C-shaped cards with a gap of 0.5 cm between the arms were dragged across the forelimb tarsi to obtain silk. Fibers were secured to the card and inspected following the procedures described above for *Haploembia.*

Tensile testing

The diameter of each silk sample was measured using polarized light microscopy (Blackledge *et al.*, 2005). Accuracy of the polarized light microscopy measurements was confirmed by comparison to scanning electron microscopy measurements of exemplar fibers. Tensile tests were performed on a Nano Bionix® tensile tester (MTS Systems Corp., Oakridge, TN). Tests were conducted at room temperature, which varied between 22–26^{o} C, and relative humidity of 30–65%. Each silk sample card was mounted to the clamps of the tensile tester and then most of the card material was removed such that only the silk spanned the gap between each clamp. The fiber was stretched at a strain rate of 1% per second until failure. The tensile test system software (TestWorks version 4.09.70 by MTS Systems Corp., Eden Prairie, MN) was used to calculate ultimate stress, the force needed to stretch a fiber to breaking; ultimate strain, the maximum elongation at breaking; stiffness, the material's resistance to stretching; and toughness, the amount of energy a given volume of material can absorb before breaking. Multivariate analysis of variance (MANOVA) was used to compare ultimate stress, ultimate strain, Young's modulus, and toughness across the six silk types. Tukey's HSD was used for *post hoc* comparison between pairs. Statistical analyses of the tensile data were performed using SPSS 13.0 (release 13.0.01.09.2004 for Windows by SPSS Inc., Chicago, IL).

Fourier transformed infrared spectroscopy

Spectroscopy analyses were done on the silks of *Antipaluria, Haploembia, Saussurembia* and *Archembia*. For each species, silk samples were harvested from an anesthetized adult female that was placed on her dorsal side atop a dissection microscope

stage. From each insect, numerous fibers were collected to completely fill the 5 mm gap of a C-shaped card. This dense sampling of silk was necessary to ensure that there was enough material to obtain clean spectrograms. Each silk sample was scanned 256 times in five different areas using a Bruker Equinox 55 FTIR (Bruker Optik, Billerica, MA) with microscope (15X objective). The five spectrograms from each sample were compared for consistency. Peak positions within the amide III region were identified with second derivative analyses. Through iterative curve fitting analysis, a Gaussian fit of each original spectrogram was used to determine component bands and secondary structures were assigned according to Cai and Singh (Cai and Singh, 1999). Spectrographic analyses were performed using OPUS software (version 6.0.72 by Bruker Optik, Billerica, MA).

Results and Discussion

Diameter measurements

The diameters of the silk fibers produced by the six embiid species in our study are similar (420-594 nm; Table 2.1) despite a 9-15 mm span of body lengths. These diameters are towards the lower end of the 0.1-11.5 μm range of previously reported arthropod silk sizes (exemplars in Table 2.1; Pérez-Rigueiro et al., 1998; Young and Merritt, 2003; Brown et al., 2004; Chang et al., 2005; Blackledge and Hayashi, 2006; Okada et al., 2008). Webspinner silks are thin in comparison to these other silks, especially considering organism sizes. For example, *Hilarempis* dance flies and *Kahaono* leafhoppers are about 4mm long, yet spin thicker silks than all our embiids (Evans, 1966;

Sutherland *et al.*, 2007). The small silk diameter reported for embiopterans may reflect the unique aspects of the embiid spinning apparatus and production method. Embiids spin silk from hundreds of silk ejectors on their tarsi rather than from paired spigots on the abdominal or oral region, as in spiders, lepidopterans, and leafhoppers (Ross, 2000; Collin *et al.*, 2009).

Table 2.1. Mean embiopteran silk diameters (± standard deviation) measured in this study* compared to other arthropod silks. Silk collected was from (N) individuals with a total of (n) silk samples. Diameter measurements (*Aposthonia*-Okada *et al.*, 2008; *Kahaono*-Chang *et al.*, 2005; *Hilarempis*-Young and Merritt, 2003, 2003; *Hydropsyche*-Brown *et al.*, 2004; *Bombyx*-Pérez-Rigueiro *et al.*, 2001; *Argiope*-Blackledge and Hayashi, 2006. Body length measurements (*Aposthonia*-Davis, 1938; *Kahaono*-Evans, 1966; *Hilarempis*-Sutherland *et al.*, 2007; *Hydropsyche*-Wiggins, 1977; *Bombyx*-Peterson, 1948; *Argiope*-Kaston, 1978).

Taxon	Silk fiber diameter (nm)	Typical body length (mm)
Oligotoma, N=6, n=15	420 ± 69*	10*
Archembia, N= 7, n=12	491 ± 135*	13*
Haploembia, N=7, n=12	583 ± 93*	9*
Saussurembia, N=12, n=32	460 ± 77*	10*
Australembia, N=5, n=10	469 ± 38*	12*
Antipaluria, N=6, n=12	594 ± 91*	15*
Aposthonia, webspinner	87 ± 10	9
Kahaono, leafhopper	700-1000	4
Hilarempis, dance fly	3,000	4
Hydrospyche, caddis fly	11,000	16
Bombyx, silkworm	8,200	40
Argiope, spider (dragline)	3,500	20

Besides the six species in our study, *Aposthonia gurneyi* (Oligotomidae) is the only other embiid with a reported silk diameter. Okada *et al.* (2008) used scanning electron microscopy on a patch of silk removed from an existing *in situ* gallery to calculate a mean fiber width of 87 nm (with 10-20 nm of this diameter attributed to the gold sputter coating). While *Aposthonia* may indeed spin silks that are much thinner than the fibers spun by the embiids in our study (87 nm vs. 420-594 nm; Table 2.1), the difference could be due to sampling a patch of fibers from a silk gallery rather than fibers

from adult females only. *In situ* silk galleries are expected to contain silks of varying diameters because webspinners produce silk during all life-stages; even newly hatched nymphs can spin fibers (Ross, 2000). Presumably, tiny nymphs spin smaller silk fibers than adults. Thus, it is likely that the patch of silk collected from the *Aposthonia* colony by Okada *et al.* (2008) contained very thin fibers spun by immature embiids.

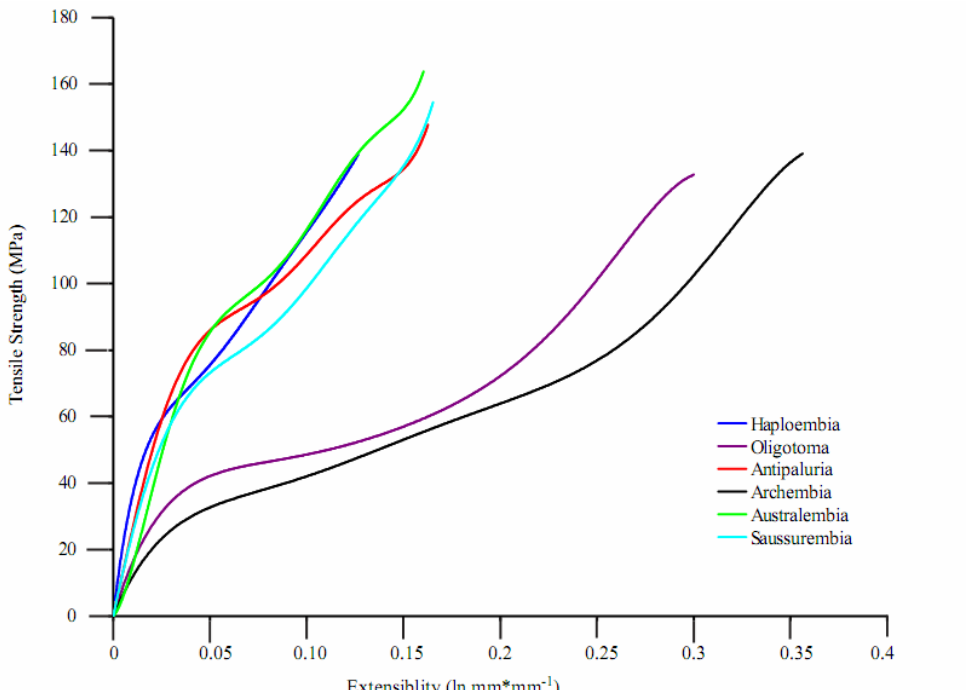


Fig. 2.2. Representative stress-strain curves for silk fibers spun by *Haploembia* (dark blue), *Oligotoma* (purple), *Antipaluria* (red), *Archembia* (black), *Austalembia* (green) and *Saussurembia* (cyan).

Tensile properties

Tensile testing was performed on silk collected exclusively from adult females to reduce confounding sources of variation, such as fibers too small to measure accurately via light microscopy. Silk from all six species had tensile characteristics of a visco-elastic material (Fig. 2.2). As the fibers were elongated, they initially resisted the force like a

solid until a yield point (mean yield point for each species ranged from 0.015-0.039 ln*(mm/mm), data not shown) and then stretched more compliantly until failure. The MANOVA revealed significant differences in tensile properties among the six types of embiopteran silks overall. However, between-property comparisons showed that this significance was driven by three of the four tensile properties. Mean ultimate stress (tensile strength) varied from 129-169 MPa (Fig. 2.3a) with no significant differences across species despite the tested *Haploembia* silk being naturally spun fibers while the other tested fibers were obtained by forcible silking, which has been shown to impact tensile strength (Ortlepp and Gosline, 2004). By contrast, there was more variation in ultimate strain, Young's modulus, and toughness, with each property differing significantly across species (Fig. 2.3 b,c,d). The differences in toughness among embiid silks are expected given the variation in ultimate strain and Young's modulus, two properties that impact the shape of the stress-strain curve, which is used to calculate toughness (Fig. 2.2).

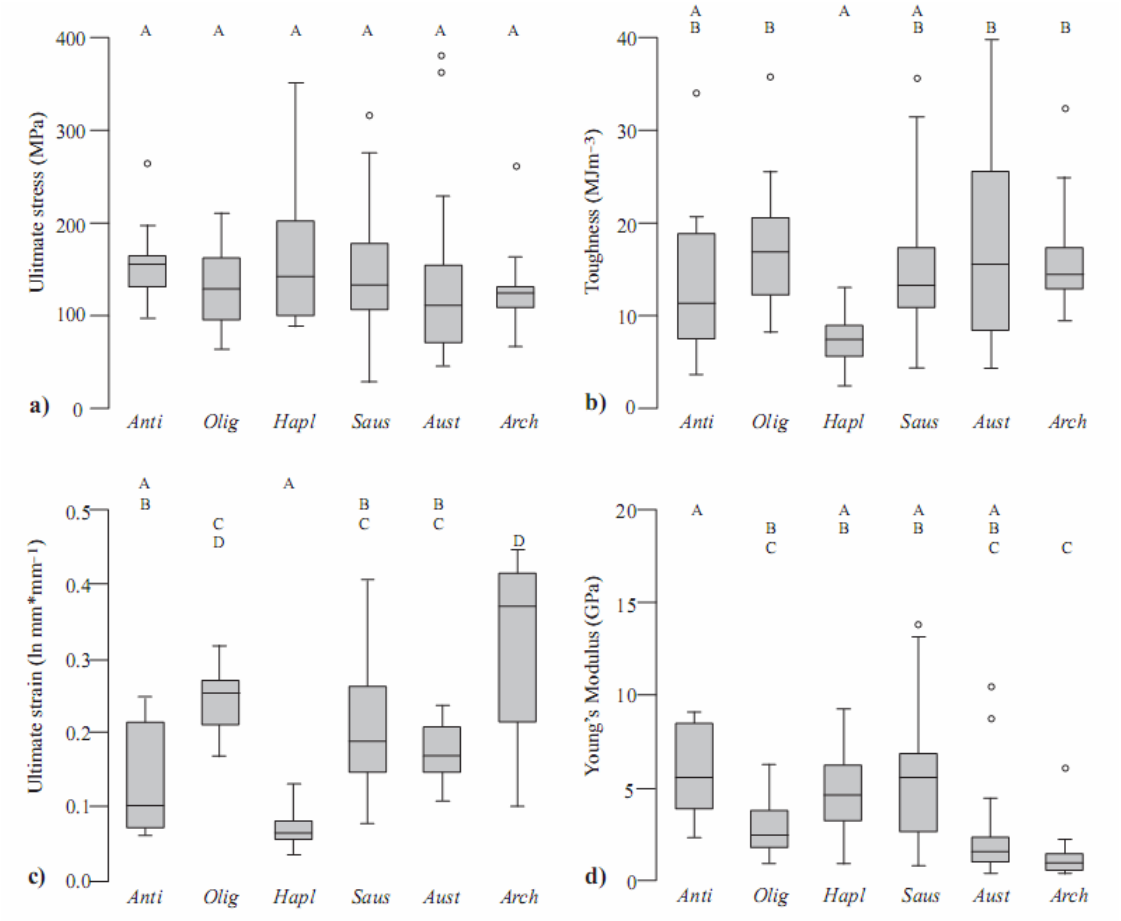


Fig. 2.3. Box plots comparing tensile properties across taxa. Lines denote the median, boxes enclose the upper and lower quartiles and whiskers show range. Open circles are outliers that exceed the range of the interquartile distance by 1.5 times. Upper-case letters above each set of box plots denote significantly different means calculated by Tukey's HSD where $p < 0.05$. (a) ultimate stress, (b) toughness, (c) ultimate strain, and (d) Young's modulus. Taxonomic abbreviations: *Antipaluria* (*Anti*), *Oligotoma* (*Olig*), *Haploembia* (*Hapl*), *Saussurembia* (*Saus*), *Australembia* (*Aust*), and *Archembia (Arch)*.

The mechanical properties we measured here are concordant with the previous characterization of an embiid silk (Collin *et al.*, 2009). This earlier study compared embiid silk tensile properties to other arthropod silks. Specifically, embiid silk stiffness and extensibility were shown to be quite similar to values reported for the silks spun by silkworms and some spiders (Pérez-Rigueiro *et al.*, 2001; Blackledge and Hayashi,

2006). These correspondences may be due to functions common to these silk producing organisms, such as protection from biotic and abiotic factors. Embiid silk toughness and ultimate stress values, however, are considerably weaker (by an order of magnitude) than either silkworm or spider silks. Toughness, being a measurement of the amount of energy required to break a fiber, is determined by the total area under the stress-strain curve (Fig. 2.2) and therefore is affected by stiffness, yield point, ultimate strain, and ultimate stress. Hence, the lower toughness of embiopteran silks compared to spider and silkworm silks is largely attributable to dramatic tensile strength differences (e.g. 158 MPa for *Antipaluria* silk vs. 1495 MPa for *Argiope* dragline; Blackledge and Hayashi, 2006).

Embiid silks have lower tensile strengths than have been reported for other arthropod silks (Collin *et al.*, 2009). Given the network of ejectors on their forelimb tarsi, embiids spin multiple fibers at a time and weave these strands into silken meshes. Since embiopteran fibers are used as sheets, single fibers do not need to have the high tensile strengths that are characteristic of lepidopteran escape lines or spider draglines. The similarity of ultimate stress values across the six webspinner taxa (Fig. 2.3a) may reflect that all species use silk for the same basic function, the construction of tunnels and galleries.

The fairly uniform tensile strengths may be explained, in part, if all embiid fibroins have similar molecular weights. Molecular size has been hypothesized to affect tensile strength (Termonia *et al.*, 1985). Specifically, fibers composed of high molecular weight monomers tend to have higher tensile strengths than fibers built from smaller monomers due to the increased molecular bonding that is possible between larger

monomers. This correlation has been empirically shown with synthetic nylon fibers (Hallam *et al.*, 1986). Recently, Brooks *et al.* experimentally tested this hypothesis by producing silk fibers from recombinant silk proteins that varied in length (Brooks *et al.*, 2008). It was observed that fibers spun from proteins with a higher number of iterated motifs had greater tensile strength than fibers with fewer motifs. Accordingly, the relatively low tensile strength of embiid silks may be due to embiid fibroins having smaller monomers than other arthropod silks.

Using Northern blot analysis, Collin *et al.* (2009) showed that *Antipaluria* silk mRNA transcripts range between 1.5-2.5 kilobases. These sizes are considerably smaller than the >15 kilobase transcript lengths reported for some spider and moth fibroins (Ohshima and Suzuki, 1977; Hayashi *et al.*, 1999). Since silk production presumably evolved once within Embioptera, all embiid silk fibroins are considered homologous and potentially share characteristics, such as transcript size, retained from an embiopteran ancestor. Given that the basal split in Embioptera is between Clothodidae (represented here by *Antipaluria)* and all other families (Szumik *et al.*, 2008), if the *Antipaluria* silk mRNA is typical for the order, then other embiid species may also have fibroins of similarly small molecular weight.

The variation observed for ultimate strain (7-36% across species; Fig. 2.3c) may indicate differences in fibroin amino acid sequences that alter secondary structure. Secondary structures, such as β-spirals, have been hypothesized to increase extensibility (Savage and Gosline, 2008). These secondary structures allow the fiber to stretch as hydrogen bonds of the β-spirals are broken before the fiber ruptures. Additionally, these

same structures may also affect Young's modulus (stiffness). Presumably, a fiber with greater percentages of β-spirals and less β-sheets should be more pliable (lower Young's modulus) than fibers with proportionally more β-sheets. This difference is exemplified by the mechanical and molecular properties of dragline silk versus capture spiral produced by orb-weaving spiders (Guerette *et al.*, 1996; Hayashi and Lewis, 1998). Dragline silk is a stiff, high tensile material used to frame orb webs, while capture spiral silk is a strong but pliable silk able to slow and stop flying insects. Molecularly, capture spiral and dragline share some common motif elements such as glycine-glycine-x and glycine-proline-glycine (Hayashi and Lewis, 1998). However, dragline silk proteins contain a poly-alanine repeats that form β-crystalline structures to provide strength and stiffness. Savage and Gosline (2008) performed a more detailed comparison of silk mechanical properties in their investigation of draglines spun by two spider species. These draglines are composed of the same structural elements, but differ in the relative proportion of each element. Through tensile testing, it was shown that the dragline with a greater frequency of motifs associated with extensibility could stretch 5% more before breaking (Savage and Gosline, 2008). A similar amount of variation in extensibility was observed with the embiid silks, such as between *Saussurembia* and *Australembia* fibers (Fig. 2.3c). Based on studies of other arthropod silks, the interspecific variation in the ultimate strain and Young's moduli of embiid silks may result from shifts in the proportions of secondary structural elements common to all the fiber types.

FTIR measurement

FTIR is an analytical technique that can be used to determine vibrational frequencies of atomic bonds or groups of bonds in proteins. These molecular vibrations are known to be contingent on hydrogen bonding and secondary structure (Sutherland, 1952). By empirically associating these modes of vibration with known structural conformations, FTIR has been used to assess the secondary structures in a variety of previously uncharacterized proteins (Chang *et al.*, 2006; Weisman *et al.*, 2008). In our study, the similarity of the complete FTIR spectrograms for the different embiid silks (supplementary Figure 1 in the Supporting Information) indicates that they have similar secondary structures. Our data also resemble the spectrogram for the embiopteran, *Aposthonia* (Okada *et al.*, 2008).

The consistency in the FTIR spectra for multiple embiid silks is in contrast to the breadth of structural variation noted in examinations of other arthropod silks (Lucas *et al.*, 1960; Warwicker, 1960; Sirichaisit *et al.*, 2003). Warwicker (1960) used X-ray diffraction to categorize spider and insect silks into five different groups that later researchers expanded to eight (Rudall and Kenchington, 1971). These eight silk types were sampled from diverse taxa and represent multiple evolutionary origins of silk and silks that vary in their ecological function. Given these disparate parameters, lineage-specific evolution and selection for divergent functions are expected to have resulted in a range of physical properties and hence, different structural compositions. Our study on embiid silks is on a much finer scale because our silks are spun by species within a single order, represent a single origin of silk, and perform comparable ecological functions.

To elucidate the subtle differences between embiopteran silk spectrograms, we performed curve-fitting analyses to determine secondary structures within the silk. In particular, the amide I and amide III regions of the FTIR spectrograms are often used for protein structural studies (Fu *et al.*, 1994; Cai and Singh, 1999; Chang *et al.*, 2006). Although weaker in intensity than amide I, spectral bands in amide III are more resolved, which aids in band assignment (Singh, 2000). This difference in resolution can be seen with the embiopteran silks. The amide I region is a broad peak composed of multiple overlapping bands (Supplementary Fig. 2.1) while the amide III region possesses more distinct peaks (Fig. 2.4). The amide III region primarily results from C-N stretching, N-H bending, and C-O stretching (Socrates, 2001). By analyzing the amide III spectral bands, secondary structures such as α-helices, β-turns, and β-sheets can be deduced (Cai and Singh, 1999; Chang *et al.*, 2006). These conformations are of importance because they have been hypothesized to impart specific attributes to silk mechanical performance (Denny, 1980; Hayashi *et al.*, 1999).

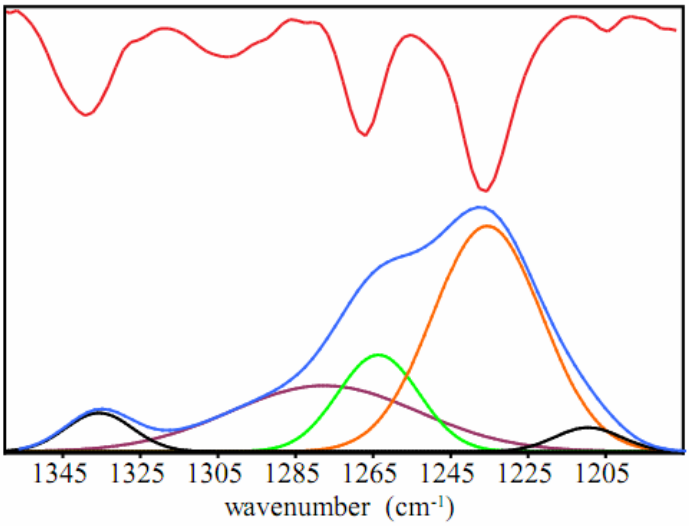


Fig. 2.4. Amide III region from *Saussurembia* original FTIR spectrogram (blue) with second derivative (red) above and component bands obtained through curve-fitting analysis underneath at 1235 cm^{-1} (orange), 1263 cm^{-1} (green), and 1277 cm^{-1}(purple). Black curves correspond to bands that fall outside the range of assigned secondary structures.

The amide III regions of the embiopteran silk spectra all have a broad curve with highest intensity at ~1235 cm^{-1} indicating a high proportion of β-sheet structure (Supplementary Fig. 2.2). By examining the second derivative of the original spectrogram, other bands within the amide III region were identified (Fig. 2.4). Through curve fitting analysis of the amide III region, it was determined that all the embiopteran silks in this study are primarily composed of β-sheets, random coils, and β-turns, with *Antipaluria, Haploembia,* and *Archembia* silk also containing a small percentage of α-helices (Table 2.2). The relative proportion of each structural element varies by species, however all the silks have >45% β-sheet conformation (Fig. 2.5).

Table 2.2 Assignments and relative proportions of amide III components.

Species	Frequency (cm^{-1})	Assignment	Proportion (%)
Haploembia	1315	α-helix	4
	1287	β-turn	13
	1265	random coil	20
	1238	β-sheet	63
Antipaluria	1314	α-helix	2
	1292	β-turn	14
	1263	random coil	29
	1254	random coil	8
	1232	β-sheet	47
Saussurembia	1277	β-turn	28
	1263	random coil	17
	1235	β-sheet	55
Archembia	1323	α-helix	5
	1288	β-turn	14
	1264	random coil	30
	1237	β-sheet	51

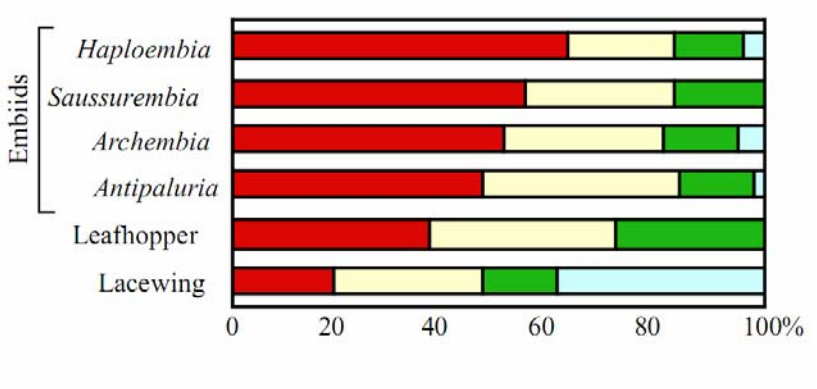


Fig. 2.5. Secondary structural component percentages from FTIR amide curve fitting analysis. Proportions of β-sheet, random coil, β-turns, and α-helix are shown in red, yellow, green, and light blue, respectively. Leafhopper-Chang *et al.*, 2006; Lacewing-Weisman *et al.*, 2008.

The high proportion of β-sheet conformation and consistent spectrograms across species likely reflects conservation of fibroin amino acid sequence within Embioptera. For example, the repetitive region of the *Antipaluria* silk fibroin contains numerous glycine-serine couplets (Collin *et al.*, 2009). Besides *Antipaluria* (Clothodidae), the other reported embiid fibroin sequence is from the distantly related *Aposthonia* (Oligotomidae), which also has a repetitive region consisting of glycine-serine couplets (Okada *et al.*, 2008). While embiid fibroin sequences have only recently been characterized, comparison can be made to the well-studied *Bombyx* silkworm heavy-chain fibroin. In *Bombyx* silk, the glycine-serine couplets have been shown to participate in the formation of β-crystalline structure (Marsh *et al.*, 1955). Given the similarity between *Antipaluria* and *Aposthonia* fibroin repetitive motifs, sequence conservation throughout Embioptera may underlie the similar silk spectrograms and proportions of β-sheet conformation observed through FTIR.

Despite the overall similarity of secondary structural elements determined by FTIR, there are differences that may account for the variation observed in ultimate strain. For example, *Haploembia* has the highest percentage of β-sheet and least extensibility of the embiid silks examined (Table 2.2; Fig. 2.3c). In other silks, β-sheets are known to form crystalline regions that reduce extensibility (Termonia, 1994; Trancik *et al.*, 2006). This relationship suggests that the larger β-sheet component of *Haploembia* silk fibers compared to the other embiid silks may correspond to a higher crystallite percentage, thus increasing rigidity and hence decreasing the ability to stretch.

In contrast to β-sheets, β-spirals are components that can increase extensibility. For example, the greater extensibility of *Saussurembia* fibers compared to *Haploembia* fibers (7% vs. 20%; Figure 2.3c) may be accounted for by *Saussurembia* silk having a greater β-turn percentage than that of *Haploembia* (Table 2.2). Because a β-spiral is a series of concatenated β-turns that form an ordered spiral structure, any β-spirals in the silk fiber would be part of the β-turn component. Although some β-turns are necessary for β-crystalline formation, presumably some percentage of β-turns is organized into β-spirals that provide extensibility. Mechanistically, silk extensibility is thought to result from the stretching of the highly ordered β-spiral structures (Becker *et al.*, 2003).

Additional factors might impact the observed differences in ultimate strain of embiid silks. A key element is crystallite size, which has been theoretically calculated to affect silk tensile properties such that smaller crystallites (2 x 6 x 21 nm) increase tensile strength more than larger crystallites (Termonia, 1994). Crystallites of varying dimension can be found within the same fiber, such as the black widow spider dragline with its

bimodal population of crystallite sizes. Each size category makes a distinct contribution to the observed tensile properties of the dragline (Trancik *et al.*, 2006). Besides crystallite dimension, the arrangement of structural components relative to each other has been shown to affect silk performance (Papadopoulos *et al.*, 2009).

Surprisingly, the secondary structural percentages measured in silkworm silk are remarkably similar to our embiopteran silk measurements (Lefèvre *et al.*, 2007). Lefèvre *et al.* (2007) analyzed *Bombyx mori* silks via Raman spectroscopy and determined that *Bombyx* silk secondary structural percentages are: 8% random coil, 14% α-helix, 50% β-sheet, and 28% β-turns. This β-sheet percentage is almost identical to *Archembia* silk (Table 2.2), yet *Bombyx* silk is at least an order of magnitude stronger and tougher than *Archembia* silk. If *Archembia* silk protein size is similar to the ~67 kDa estimate of *Antipalura* silk (extrapolated from mRNA length; Collin et al., 2009), then the greater tensile strength of *Bombyx* silk may be due to the much bigger (350 kDa) molecular weight of its heavy-chain fibroin (Yamada *et al.*, 2001). With this massive size, *Bombyx* fibroin has numerous β-sheet forming regions, which may enable *Bombyx* silk to have more strength-providing β-crystallites than embiopteran silks (Termonia, 1994). Additionally, sequence motif differences may affect tensile strength. For example, *Bombyx* silk is predominately composed of glycine-alanine repeats whereas the two characterized embiopteran fibroins consist largely of glycine-serine repeats (Okada *et al.*, 2008; Collin *et al.*, 2009). Though glycine-alanine repeats and glycine-serine repeats both form β-sheet secondary structures, the occurrence of serine instead of alanine in these motifs may contribute to the lower tensile strength of embiid silks. Serine rich β-sheets

may have slightly more inter-sheet spacing than alanine rich β-sheets due to the larger size and greater polarity (from the hydroxyl side-chain) of serine compared to alanine.

The secondary structures of only a few arthropod silks have been analyzed with methods similar to those presented here. For direct evaluation of FTIR generated results, we compare embiid silk secondary structural components to silks of *Kahaono montana* leafhoppers (Cicadellidae, Hemiptera; Chang *et al.*, 2006) and larval *Mallada signata* lacewings (Chrysopidae, Neuroptera; Weisman *et al.*, 2008). These insect silks vary substantially from each other in the proportions of secondary structural elements. Both leafhopper and lacewing cocoon silk have lower percentages of β-sheets than embiid silk. The lower relative amount of β-sheet is compensated by a higher proportion of β-turns in leafhopper silk and a higher α-helical component in lacewing cocoon silk compared to embiid silk.

A number of factors could explain these differences in secondary structure among insect silks. Certainly contributing to the variation is that insect silks have evolved many times and Embioptera, Hemiptera and Neuroptera are just three of the independent derivations (Craig, 1997). Embiid silks are produced in secretory cells packed into the forelimb tarsus (Nagashima *et al.*, 1991), while leafhopper and lacewing cocoon silk are produced in malpighian tubules (Chang *et al.*, 2005; Weisman *et al.*, 2008). Given the differences in how silk proteins are secreted and aggregate into fibers in each type of spinning organ, some variation in silk secondary structures is expected. In addition, each insect clade experiences a host of lineage-specific selection pressures that should

uniquely shape characteristics of their silk, such as repetitive motifs and other protein sequence elements (Okada *et al.*, 2008; Weisman *et al.*, 2008; Collin *et al.*, 2009).

In summary, we quantified the mechanical properties of embiopteran silks through tensile testing and structural components with FTIR. We show that mature females spin silk with surprisingly consistent diameters (~0.5 μm; Table 2.1), despite a 60% difference in body lengths and extensive phylogenetic divergence. Across species, the silks have similar tensile strengths (Fig. 2.3a). Among species, however, there is significant variation in stiffness, maximum extensibility, and toughness (Fig. 2.3b-d). Some of these performance differences can be attributed to differences in secondary structure, as revealed by FTIR (Table 2.2). Specifically, the relative amounts of crystallite-forming β-sheets versus flex-enabling β-spirals correspond to variation in fiber stiffness and extensibility. Because crystallite dimension has also been shown to be a predictor of fiber tensile properties (Termonia, 1994; Trancik *et al.*, 2006), measuring embiid crystallite size via X-ray crystallography or transmission electron microscopy would provide more in depth understanding of the relationship between mechanical properties and silk nanostructure.

Enhanced insight into the structure and function of embiid silks can be gained from amino acid sequence characterization of their constituent fibroins. The data available thus far indicate that embiid fibroins are mostly composed of repeated glycine-serine couplets, which are known to form β-crystalline structures in silkworm silk (Marsh *et al.*, 1955). The prevalence of glycine-serine couplets in embiid fibroins corresponds well with the silks being >45% β-sheet, as indicated by our FTIR analyses (Fig. 2.5).

Based on the transcript size of *Antipaluria* fibroin monomer (Collin *et al.*, 2009), embiid fibroins appear to be only a fraction the length of other arthropod fibroins. The shorter monomer size means that embiid silks have fewer possible intramolecular and intermolecular interactions, and thus lower tensile strength, than silks spun by silkworms and spiders. Additional analyses of independently evolved silks (such as those of Araneae, Lepidoptera, Hemiptera), combined with fine-scale studies of silks within the order, will further uncover the functional basis and evolutionary history of embiopteran silks.

References

Ayoub, N. A., Garb, J. E., Tinghitella, R. M., Collin, M. A., Hayashi, C. Y. 2007. Genetic blueprint of spider silk uncovered. PLoS ONE. 2.

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

Blackledge, T. A., Cardullo, R. A., Hayashi, C. Y. 2005. Polarized light microscopy, variability in spider silk diameters, and the mechanical characterization of spider silk. Invertebrate Biology. 124, 165-173.

Blackledge, T. A., Hayashi, C. Y. 2006. Silken toolkits: biomechanics of silk fibers spun by the orb web spider *Argiope argentata* (Fabricius 1775). Journal of Experimental Biology. 209, 2452-2461.

Brooks, A. E., Stricker, S. M., Joshi, S. B., Kamerzell, T. J., Middaugh, C. R., Lewis, R. V. 2008. Properties of synthetic spider silk fibers based on *Argiope aurantia* MaSp2. Biomacromolecules. 9, 1506-1510.

Brown, S. A., Ruxton, G. D., Humphries, S. 2004. Physical properties of *Hydropsyche siltalai* (Trichoptera) net silk. Journal of the North American Benthological Society. 23, 771-779.

Cai, S. W., Singh, B. R. 1999. Identification of beta-turn and random coil amide III infrared bands for secondary structure estimation of proteins. Biophysical Chemistry. 80, 7-20.

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

Chang, J. C., Gurr, G. M., Fletcher, M. J., Gilbert, R. G. 2006. Structure-property and structure-function relations of leafhopper (*Kahaono montana*) silk. Australian Journal of Chemistry. 59, 579-585.

Collin, M. A., Garb, J. E., Edgerly, J. S., Hayashi, C. Y. 2009. Characterization of silk spun by the embiopteran, *Antipaluria urichi*. Insect Biochemistry and Molecular Biology. 39, 75-82.

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

Davis, C. 1938. Studies in Australian Embioptera. Part III. Proceedings of the Linnean Society of New South. Wales, Sydney 63, pp. 226-272.

Denny, M. W. 1980. Silks-Their properties and functions. In: The mechanical properties of biological materials.Vincent J.V.F., C., J.D. (Ed.) Society of Experimental Biology, Cambridge Eng., pp. 247.

Edgerly, J. S. 1987. Colony composition and some costs and Benefits of facultatively communal Behavior in a trinidadian webspinner, *Clothoda urichi* (Embiidina, Clothodidae). Annals of the Entomological Society of America. 80, 29-34.

Edgerly, J. S., Davilla, J. A., Schoenfeld, N. 2002. Silk spinning behavior and domicile construction in webspinners. Journal of Insect Behavior. 15, 219-242.

Evans, J. W. 1966. The Leafhoppers and Froghoppers of Australia and New Zealand (Homoptera: Cicadelloidea and Cercopoidea). Australian Museum, Sydney.

Foelix, R. F. 1996. The Biology of Spiders. Oxford University Press, New York.

Fu, F. N., Deoliveira, D. B., Trumble, W. R., Sarkar, H. K., Singh, B. R. 1994. Secondary structure estimation of proteins using the amide-III region of Fourier-transform infrared spectroscopy application to analyze Calcium binding induced structural changes in Calsequestrin. Applied Spectroscopy. 48, 1432-1441.

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

Gosline, J., Lillie, M., Carrington, E., Guerette, P., Ortlepp, C., Savage, K. 2002. Elastic proteins: biological roles and mechanical properties. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. 357, 121-132.

Gosline, J. M., Demont, M. E., Denny, M. W. 1986. The structure and properties of spider silk. Endeavour. 10, 37-43.

Guerette, P. A., Ginzinger, D. G., Weber, B. H. F., Gosline, J. M. 1996. Silk properties determined by gland-specific expression of a spider fibroin gene family. Science. 272, 112-115.

Hallam, M. A., Cansfield, D. L. M., Ward, I. M., Pollard, G. 1986. A study of the effect of molecular weight on the tensile strength of ultrahigh modulus polyethylenes. Journal of Materials Science. 21, 4199-4205.

Hayashi, C. Y., Lewis, R. V. 1998. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. Journal of Molecular Biology. 275, 773-784.

Hayashi, C. Y., Shipley, N. H., Lewis, R. V. 1999. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. International Journal of Biological Macromolecules. 24, 271-275.

Jung, C. L., Croft, B. A. 2001. Ambulatory and aerial dispersal among specialist and generalist predatory mites (Acari : Phytoseiidae). Environmental Entomology. 30, 1112-1118.

Kaston, B. J. 1978. How to Know the Spiders. W. C. Brown Co., Dubuque, IA.

Lefèvre, T., Rousseau, M. E., Pézolet, M. 2007. Protein secondary structure and orientation in silk as revealed by Raman spectromicroscopy. Biophysical Journal. 92, 2885-2895.

Lucas, F., Shaw, J. T. B., Smith, S. G. 1960. Comparative studies of fibroins .1. Amino acid composition of various fibroins and its significance in relation to their crystal structure and taxonomy. Journal of Molecular Biology. 2, 339-349.

Marsh, R. E., Corey, R. B., Pauling, L. 1955. An investigation of the structure of silk fibroin. Biochimica Et Biophysica Acta. 16, 1-34.

Mita, K., Ichimura, S., James, T. C. 1994. Highly repetitive structure and its organization of the silk fibroin gene. Journal of Molecular Evolution. 38, 583-592.

Mori, K., Saito, Y. 2005. Variation in social behavior within a spider mite genus, Stigmaeopsis (Acari : Tetranychidae). Behavioral Ecology. 16, 232-238.

Nagashima, T., Niwa, N., Okajima, S., Nonaka, T. 1991. Ultrastructure of silk gland of webspinners *Oligotoma japonica* insecta embioptera. Cytologia. (Tokyo) 56, 679-685.

Ohshima, Y., Suzuki, Y. 1977. Cloning of silk fibroin gene and its flanking sequences. Proceedings of the National Academy of Sciences of the United States of America. 74, 5363-5367.

Okada, S., Weisman, S., Trueman, H. E., Mudie, S. T., Haritos, V. S., Sutherland, T. D. 2008. An Australian webspinner species makes the finest known insect silk fibers. International Journal of Biological Macromolecules. 43, 271-275.

Ortlepp, C. S., Gosline, J. M. 2004. Consequences of forced silking. Biomacromolecules 5, 727-731.

Papadopoulos, P., Sölter, J., Kremer, F. 2009. Hierarchies in the structural organization of spider silk-a quantitative model. Colloid and Polymer Science. 287, 231-236.

Pérez-Rigueiro, J., Elices, M., Llorca, J., Viney, C. 2001. Tensile properties of silkworm silk obtained by forced silking. Journal of Applied Polymer Science. 82, 1928-1935.

Pérez-Rigueiro, J., Viney, C., Llorca, J., Elices, M. 1998. Silkworm silk as an engineering material. Journal of Applied Polymer Science. 70, 2439-2447.

Peterson, A. 1948. Larvae of Insects An Introduction to Nearctic Species. Edward Brothers Inc, Ann Arbor MI.

Ross, E. S. 2000. Contributions to the biosystematics of the insect order Embiidina. Part 1. Origin, relationships and integumental anatomy of the insect order Embiidina. Occasional Papers California Academy of Sciences. 149, 1-53.

Rudall, K. M., Kenchington, W. 1971. Arthropod silks problem of fibrous proteins in animal tissues. Annual Review of Entomology. 16, 73-96.

Savage, K. N., Gosline, J. M. 2008. The effect of proline on the network structure of major ampullate silks as inferred from their mechanical and optical properties. Journal of Experimental Biology. 211, 1937-1947.

Singh, B. R. 2000. Basic Aspects of the Technique and Application of Infrared Spectroscopy of Peptides and Proteins In: Singh, B. R. (Ed.) Infrared Analysis of Peptides and Proteins: Principles and Applications. ACS, Washington, D.C., pp. 2-37.

Sirichaisit, J., Brookes, V. L., Young, R. J., Vollrath, F. 2003. Analysis of structure/property relationships in silkworm (*Bombyx mori*) and spider dragline (*Nephila edulis*) silks using Raman Spectroscopy. Biomacromolecules. 4, 387-394.

Socrates, G. 2001. Infrared and Raman Characteristic Group Frequencies: Tables and Charts. John Wiley, Chichester, UK.

Sugiura, S., Yamazaki, K. 2006. The role of silk threads as lifelines for caterpillars: pattern and significance of lifeline-climbing behaviour. Ecological Entomology. 31, 52-57.

Sutherland, G. B. B. M. 1952. Infrared Analysis of the structure of Amino Acids, Polypeptides and Proteins. Advances in Protein Chemistry. 7, 291-318.

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, 1036-1043.

Swanson, B. O., Blackledge, T. A., Beltrán, J., Hayashi, C. Y. 2006a. Variation in the material properties of spider dragline silk across species. Applied Physics a-Materials Science & Processing. 82, 213-218.

Swanson, B. O., Blackledge, T. A., Hayashi, C. Y. 2007. Spider capture silk: Performance implications of variation in an exceptional biomaterial. Journal of Experimental Zoology Part a-Ecological Genetics and Physiology. 308A, 654-666.

Swanson, B. O., Blackledge, T. A., Summers, A. P., Hayashi, C. Y. 2006b. Spider dragline silk: Correlated and mosaic evolution in high-performance biological materials. Evolution. 60, 2539-2551.

Szumik, C., Edgerly, J. S., Hayashi, C. Y. 2008. Phylogeny of embiopterans (Insecta). Cladistics. 24, 993-1005.

Termonia, Y. 1994. Molecular modeling of spider silk elasticity. Macromolecules. 27, 7378-7381.

Termonia, Y., Meakin, P., Smith, P. 1985. Theoretical-study of the influence of the molecular-weight on the maximum tensile-strength of polymer fibers. Macromolecules. 18, 2246-2252.

Trancik, J. E., Czernuszka, J. T., Bell, F. I., Viney, C. 2006. Nanostructural features of a spider dragline silk as revealed by electron and X-ray diffraction studies. Polymer. 47, 5633-5642.

Warwicker, J. O. 1960. Comparative studies of fibroins .2. Crystal structures of various fibroins. Journal of Molecular Biology. 2, 350-&.

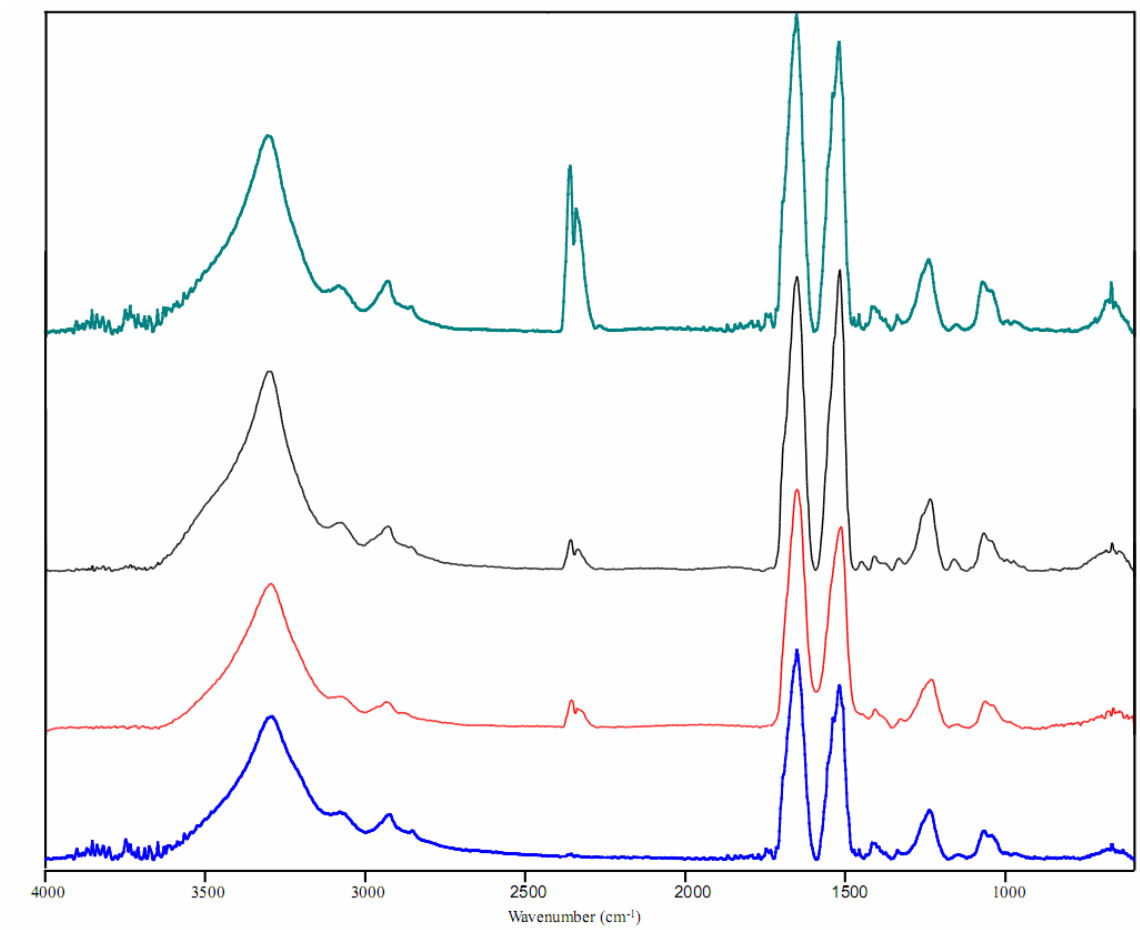
Weisman, S., Trueman, H. E., Mudie, S. T., Church, J. S., Sutherland, T. D., Haritos, V. S. 2008. An Unlikely Silk: The Composite Material of Green Lacewing Cocoons. Biomacromolecules. 9, 3065-3069.

Wiggins, G. B. 1977. Larvae of North American Caddisfly Genera (Trichoptera). University of Toronto Press, Toronto, ON CAN.

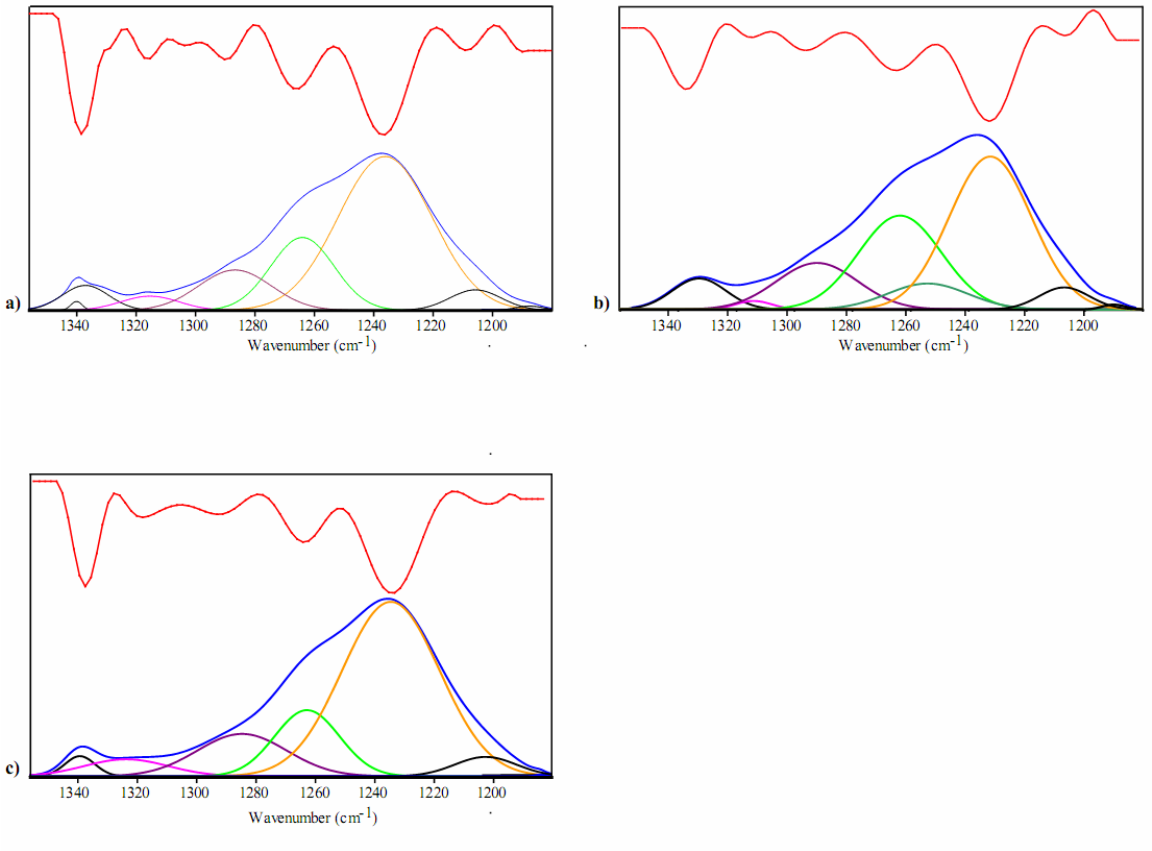
Yamada, H., Nakao, H., Takasu, Y., Tsubouchi, K. 2001. Preparation of undegraded native molecular fibroin solution from silkworm cocoons. Materials Science & Engineering C-Biomimetic and Supramolecular Systems. 14, 41-46.

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

Supplemental Figures



Supplemental Fig. 2.1. Complete original FTIR spectrograms of embiopteran silks spun by *Haploembia* (green), *Saussurembia* (black), *Antipaluria* (red), and *Archembia* (blue).



Supplemental Fig. 2.2. Amide III regions of (a) *Haploembia*, (b) *Antipaluria*, and (c) *Archembia* from the original FTIR spectrograms (blue) with second derivatives (red) above and component bands obtained through curve-fitting analyses underneath. Component bands have been assigned to α-helix (pink), β-turn (purple), random coil (light green, dark green), and β-sheet (orange). See Table 2 for wavenumber frequencies. Black curves correspond to bands that fall outside the range of assigned secondary structures.

Chapter 3

Conservation of molecular properties among embiopteran fibroins

Abstract

Embiopterans (webspinner insects) are renowned for their prolific use of silk. These organisms spin silk to construct elaborate networks of tubes in which they live, forage, and reproduce**.** The silken galleries are essential for protecting these soft-bodied insects from predators, parasitoids, and other environmental hazards. Despite the ecological importance of embiopteran silk, very little is known about its constituent proteins. Thus, obtaining molecular data is vital to understanding the functional characteristics and evolutionary patterns of these adaptive molecules. In this paper, we characterize the silk protein cDNAs from four embiopteran species. Webspinner fibroins (silk proteins) are highly repetitive in sequence and share several characteristics across species despite differences in habitat preferences. The most striking similarities are in the codon usage biases of the silk genes and the repetitive motifs, as well as sequence conservation of the carboxyl-terminal regions of the fibroins. Based on analyses of the silk genes, we propose hypotheses regarding codon bias and its effect on transcription, translation, and replication of these highly repetitive gene sequences. Furthermore, we discuss the implications of the protein sequences to the mechanical and structural characteristics of the silk fibers. Lastly, conservation of the fibroin carboxyl-terminal regions suggests potential mechanisms of fiber formation. From these results, we gain insight into the tempo and mode of evolution that has shaped embiopteran fibroins.

Introduction

Silk production has independently evolved in numerous arthropod clades (Sutherland *et al.*, 2010). Silk fibers are primarily composed of protein and are typically involved in ecologically significant functions. For example, lacewing insects construct silk egg stalks, which protect broods against predation and parasitism, and caddisflies spin silk into underwater snares for prey capture (Duelli, 1986; Yonemura *et al.*, 2009). These direct interactions with environmental and physical stressors allow evolutionary mechanisms to act on these vital proteins (Fedič *et al.*, 2003; Swanson *et al.*, 2006). Comparison of silk proteins (fibroins) from independent origins can reveal features important for fiber formation and performance that arose through convergent evolution (Bini *et al.*, 2004). Within clades, similar silk use coincides with conservation of molecular and physical properties.

The most well studied silks are those produced by moths (Lepidoptera) and spiders (Araneae). Caterpillar cocoons have been commercially important for thousands of years while spider silks are of great interest because of their exceptional strength and toughness. In spiders, the amino- and carboxyl-terminal regions of fibroins are conserved in sequence and experimental work has suggested that this is due to their relevance to fiber formation (Bini *et al.*, 2004; Ittah *et al.*, 2006). Similarly in moths, many features of the fibroins (heavy chain, light chain, and P25) that compose lepidopteran silk fibers have been maintained for at least 150 million years, emphasizing the crucial role of these proteins in fiber production (Yonemura and Sehnal, 2006). Comparative studies can reveal critical functional characteristics, such as the essential cysteine residue in the

carboxyl-terminal region of lepidopteran heavy chain fibroin that forms a disulfide bond with light chain fibroin, thus linking the two molecules (Yonemura *et al.*, 2009).

Another clade that relies on silk is Embioptera, commonly referred to as embiids or webspinners. Unlike other silk producing insects, embiids spin silk throughout their lifespan (Ross, 2000). Embioptera synthesize silk in unicellular glands located in their forelimb tarsi and silk fibers issue forth from silk ejectors on the ventral tarsal surface. Webspinners spin silk into sheets to form galleries, which are foraging tunnels that lead to food resources and protective domiciles (Edgerly, 1987a). Galleries can cover large areas (e.g., > 1 square meter), especially compared to the size of an embiid (e.g., < 1.5 cm in length). Within galleries, females also use silk and substrate material to wrap eggs for safeguarding during incubation (Edgerly, 1987b).

To date, the silks of only two webspinners species have been characterized at the molecular level (Okada *et al.*, 2008; Collin *et al.*, 2009b). As with other arthropod silks, embiid fibroins are largely composed of glycine, alanine, and serine (Gatesy *et al.*, 2001; Okada *et al.*, 2008; Collin *et al.*, 2009a). Additionally, webspinner fibroin structure is analogous to other known fibroins in that they contain a repetitive region with numerous iterations of simple amino acid motifs that is followed by a non-repetitive carboxyl-terminal region (Collin *et al.*, 2009b). Repetition is also seen at even the nucleotide level. For example, the silk gene of the webspinner *Antipaluria urichi* has a highly repetitive internal structure that resembles an 18 nucleotide minisatellite (Collin *et al.*, 2009b). Variants of this 18-mer (GGT GCA GGA TCA GGC TCT) encode Gly-Ala-Gly-Ser-Gly-Ser and appear at least 33 times in the *Antipaluria* fibroin gene. This repeat motif

most likely forms the β-sheets that make up over 45% of the secondary structure of *Antipaluria* silk (Collin *et al.*, 2009a). A multi-species comparison of silk fibroins may reveal if these findings are consistent across the order Embioptera. Additionally, because embiid species use their silks for essentially the same ecological functions, the range of molecular variation will be assessed.

In this paper, we present the silk fibroins of four webspinner species inferred from forelimb tarsal cDNA libraries. By analyzing these new sequences with the two previously published embiopteran cDNAs, we generate hypotheses about the evolution of embiopteran silk genes. We then compare the fibroin amino acid sequences across species to determine the conservation of sequence characteristics and relate primary sequences to secondary structures of embiopteran silks. Finally, through comparison of embiopteran fibroins to other arthropod silk fibroins, a better understanding is gained of embiid silk evolution and fiber formation.

Materials and Methods

Insect rearing

We reared *Archembia* nova species (Archembiidae), *Oligotoma nigra* (Oligotomidae), parthenogenetic *Haploembia solieri* (Oligotomidae), and *Saussurembia davisi* (Anisembiidae). Average body lengths of adult females are, respectively, 1.26, 0.99, 0.92 and 0.95 cm. Figure 1 depicts the phylogenetic relationships and family affiliations for the four embiopteran species in this study and two previously studied webspinner species, *Antipaluria urichi* (Clothodidae) and *Aposthonia gurneyi*

(Oligotomidae). Vouchers for *Archembia* n. sp. and parthenogenetic *Haploembia solieri* are deposited at the Museum of Southwestern Biology, Albuquerque, NM. Henceforth, species will be referred to by their genus name for brevity.

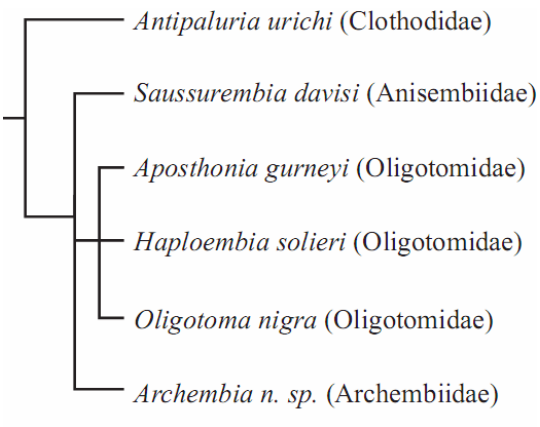


Fig. 3.1. Phylogenetic relationships of species in this study based on Szumik *et al.* (2008).

Embiids were reared in plastic terrariums containing oak leaves (*Quercus* sp.) and fed romaine or red leaf lettuce. To prevent small nymphs from escaping, terrariums were covered with fine mesh nylon (250 µm X 250 µm weave density). Colonies were misted with water two or three time a week to keep leaf litter moist and maintain humidity.

Tissue Collection

Tissue was collected from adult female embiids. For *Haploembia*, *Oligotoma*, and *Saussurembia*, 300 females per species were dissected. Because of their larger body size, only 40 females of *Archembia* were needed. Live insects were anesthetized with CO_2 gas prior to removal of whole forelimb tarsi. Only the silk producing part of the forelimb was removed by excising the tarsi at the joint between the first and second segments. The tarsi were immediately flash frozen in liquid N_2 and stored at –80°C.

Silk cDNA characterization

For each species, frozen tarsi were homogenized in TRIzol reagent (Invitrogen, Carlsbad, CA, USA) for RNA extraction. Total RNA was purified using the RNeasy mini kit (Qiagen, Valencia, CA, USA) and mRNA was pooled from total RNA using oligo $(dT)_{25}$ magnetic beads (Invitrogen, Carlsbad, CA, USA). cDNA synthesis followed the SuperScript III protocol (Invitrogen, Carlsbad, CA, USA) using an anchored oligo $(dT)_{18}V$ primer. The resulting cDNA were then size selected with a ChromaSpin 1000 column (Clontech, Mountain View, CA USA) and ligated into pZErO-2 plasmids (Invitrogen, Carlsbad, CA, USA) at the EcoRV site. The plasmids were electroporated into TOP10 *Escherichia coli* cells (Invitrogen, Carlsbad, CA, USA). ~1500 recombinant colonies from each species were arrayed into cDNA libraries and replicated onto nylon membranes. At least one third of each library was assayed to determine insert sizes and clones with inserts > ~600 bp were chosen for sequencing (Beuken *et al.*, 1998). The entire *Saussurembia* library was screened via size selection. The other three libraries were additionally screened with γ-^{32}P radiolabeled probes designed from putative fibroin cDNAs discovered amongst the size-selected clone sequences (*Archembia* probes 5'-GAWCCWGAWCCWGAWCCWGA-3', GCAAGGATCWGGWTCWGGACAAG; *Oligotoma* probes CCACTKCCKGAKCCRGAACC, CCNGCACCRCTTCCNGCACC; *Haploembia* probes TCCACTACCATCTCCGGAACC, CCACTTTCACTCTTKGAACC). Size-selected and probe positive clones were sequenced with the universal primers, T7 and Sp6, at the Core Instrumentation Facility at the University of California, Riverside. DNA sequences were translated using

Sequencher 4.2 (Gene Codes, Ann Arbor, MI, USA) and submitted for BLASTX searches against the nr database with the low complexity filter disabled (www.ncbi.nlm.nih.gov/BLAST). Additionally, the embiopteran fibroin sequences of *Antipaluria urichi* (accession # FJ361212) and *Aposthonia gurneyi* (accession # EU170437) were obtained from the NCBI database. Predicted translations, base composition, and codon usage analyses were performed with MacVector 7.0 (Oxford Molecular Group, Oxford, UK). Amino acid alignments were conducted with Clustal X (Thompson *et al.*, 1994) and refined by eye. Protparam (ExPASy tools: Gasteiger *et al.*, 2005) was used to compute amino acid composition and isoelectric point (pI).

Amino acid analysis

Embiids were placed in new petri dishes and allowed to spin silk. Collected silk samples were sent to the Molecular Structure Facility at the University of California, Davis. The samples were hydrolyzed in 6N HCl, 0.1% phenol at 110^{o}C for 24 hours. Molar fractions were determined on a Hitachi L-8800 (Hitachi Ltd., Tokyo, Japan) amino acid analyzer using ion exchange chromatography followed by ninhydrin reaction detection.

Results and Discussion

Silk cDNAs

Each cDNA library contained transcripts representing genes expressed within the forelimb tarsi of embiopterans, such as cellular maintenance and secretory genes. Due to the lack of genetic data on webspinners, several transcripts found in the libraries could

not be annotated by comparison to the NCBI database. However, every library had transcripts that encoded highly repetitive proteins. The top BLASTX hit for each of these transcripts was either the previously published embiid fibroin sequence from *Aposthonia gurneyi* (Accession # EU170437; Okada *et al.*, 2008) or *Antipaluria urichi* (Accession # FJ361212; Collin *et al.*, 2009b). Through size-screening and hybridization with probes, 3-8 fibroin cDNA clones were found per library. As with the previously reported webspinner fibroin sequences, all fibroin cDNAs were partial length and most represented the 3' transcript region. In addition to eight 3' fragments, the *Oligotoma* library had one clone that contained a 5' gene fragment. The predicted translation of the longest clone in each species-specific cDNA library is presented in Figure 3.2.

The fibroin cDNAs were computationally translated and in general, embiopteran fibroin organization is similar across species. Each silk fibroin has a short, (~ 30 amino acids), non-repetitive carboxyl-terminal region, which is largely polar and acidic (see below; Fig. 3.2, underlined) and a repetitive region (Fig. 3.2, non-underlined) that extended to the 5' end of each clone. The repetitive region is composed of short, repeated amino acid motifs with glycine-X couplets (where X is most often serine or alanine). Another motif that appears within the fibroins, but less frequently, is glycine-X-glycine (where X is often a polar or bulky amino acid such as glutamine or tyrosine).

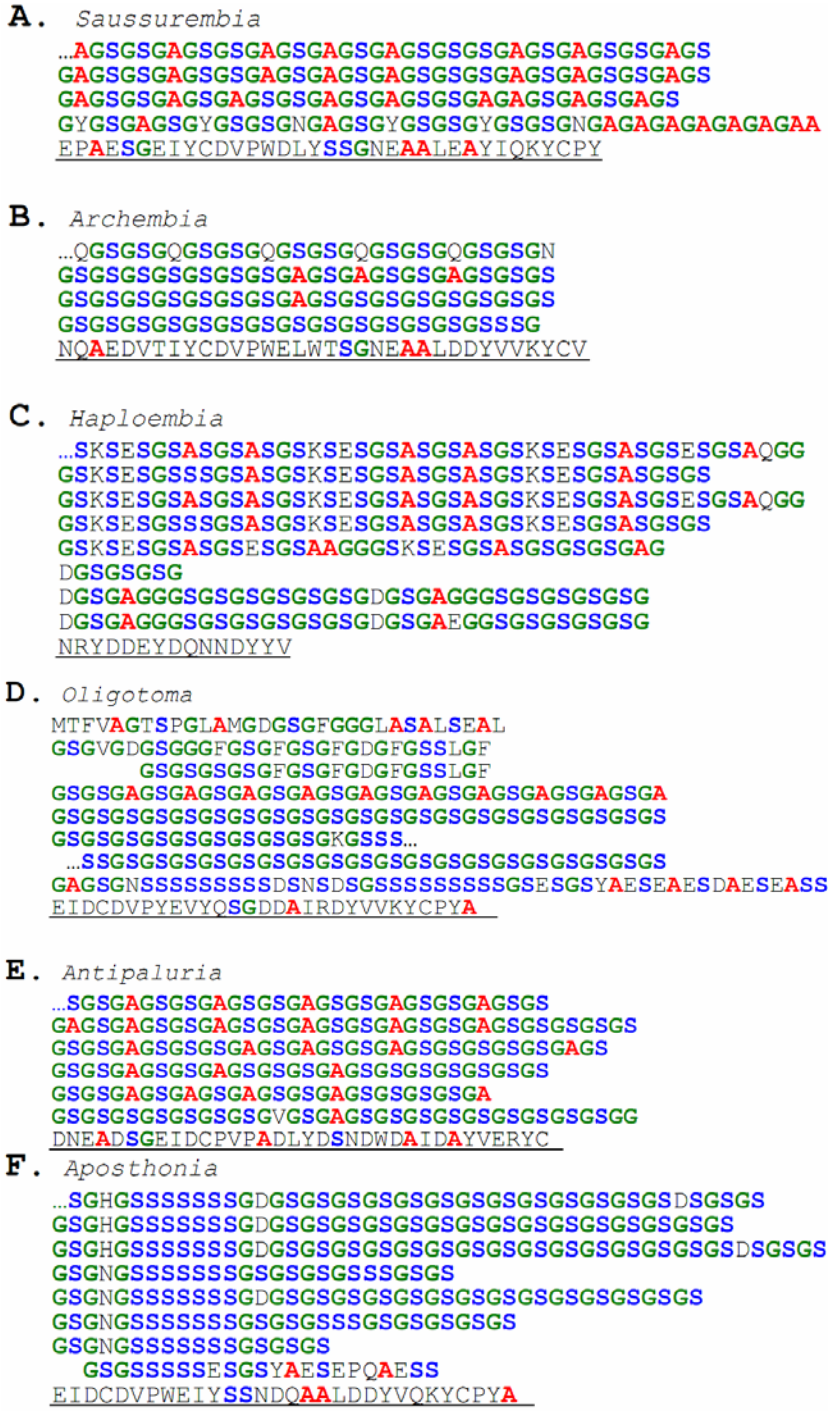


Fig. 3.2. Predicted amino acid translation of embiopteran fibroin transcripts depicted with single letter amino acid notation. Glycine, serine and alanine in bold and colored green, blue, and red, respectively. Non-repetitive carboxyl-terminal regions underlined. Genbank accession numbers: *Antipaluria* (FJ361212), *Aposthonia* (EU170437), *Archembia* (HM189215), *Haploembia* (HM189216), *Oligotoma* (HM189217,HM189218), *Saussurembia* (HM189214).

Within each species-specific library, the differences among the fibroin clones could represent allelic and/or gene copy variants. Most of the variation was within the repetitive region, which made alignment difficult. The localized length differences and sequence rearrangements may be due to the elevated rates of recombination experienced by highly repetitive sequences (Taylor and Breden, 2000; Žurovec and Sehnal, 2002). Alignment of the non-repetitive regions of intraspecific variants was trivial because there was neither length nor amino acid changes (only limited synonymous variation), which may indicate stabilizing selection of the carboxyl-terminal region.

Examining the embiopteran cDNAs more closely, each fibroin transcript contains short repeats similar to minisatellites (Taylor and Breden, 2000; Collin *et al.*, 2009b). For example, in the *Saussurembia* sequence, a particular group of thirty nucleotides (ten codons) appears six times in near perfect identity (maximum of one base difference) within the 624 basepair coding region of the cDNA (Fig. 3.3). Additionally, length variants that miss up to 12 nucleotides and have few (1-3) synonymous substitutions occur nine times within the sequence. Combined, the near perfect and imperfect repeats constitute 60% of the *Saussurembia* cDNA. The fibroin transcripts from the other embiid species are similarly structured with the repetitive regions largely composed of short repeats that are four to ten codons in length.

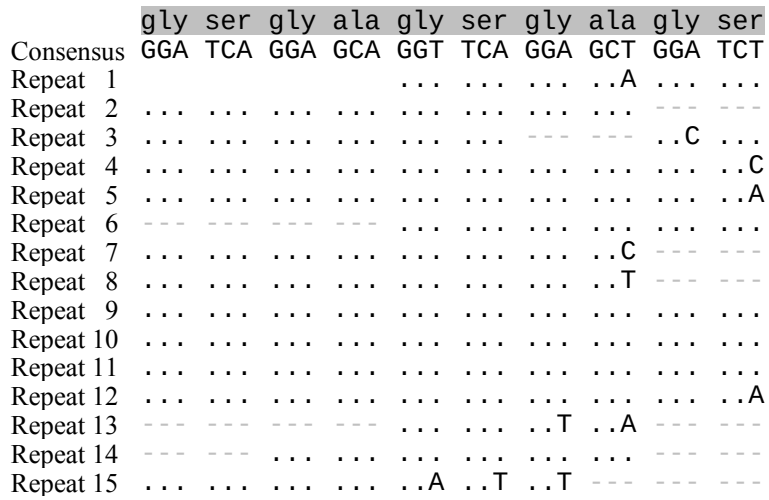


Fig. 3.3. Nucleotide (single, upper case letter abbreviations) alignment of repeat units from the *Saussurembia* fibroin transcript. Dots and dashes indicate nucleotide matches and gaps, respectively, to the consensus sequence. Encoded amino acids (shaded, three letter, lower case abbreviations) shown at top.

Based on the cDNAs, a large portion of each webspinner fibroin is composed of the repetitive region (Fig. 3.2). This finding agrees with the amino acid composition analyses of spun embiopteran silk. The silk fibers are dominated by just three amino acids. Specifically, glycine, alanine, and serine make up almost 85% of the total molar composition (Fig. 3.4). By far, the most abundant amino acid is glycine, which is over 45% of the spun silk, followed by serine at ~30%. The differences in amino acid composition across species can be explained by the species-specific repetitive motifs that were identified from the fibroin cDNAs (Fig. 3.2). For example, the *Archembia* fibroin sequence has very few alanine residues, consistent with the 5% alanine that was empirically determined from spun *Archembia* silk. By contrast, the *Saussurembia* fibroin sequence has many alanines, which corresponds to the 17% alanine found in the composition of *Saussurembia* spun silk.

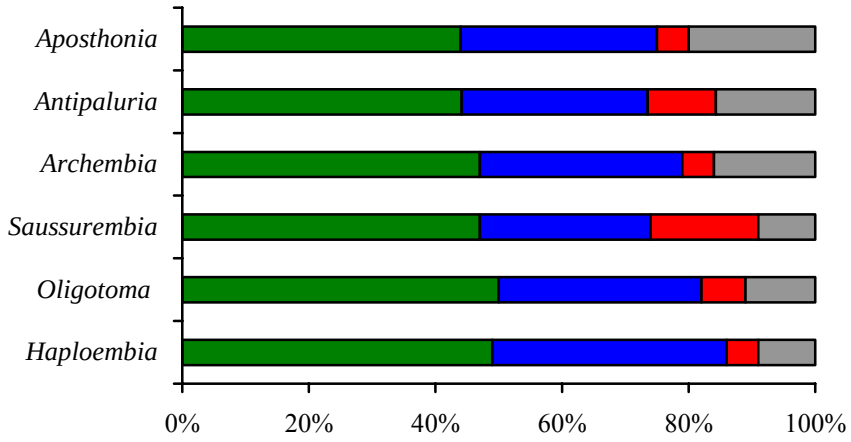


Fig. 3.4. Amino acid composition of embiopteran spun fibers determined by amino acid analysis. Data for *Aposthonia* (Okada *et al.*, 2008) and *Antipaluria* (Collin *et al.*, 2009b) included for comparison. Molar percentages of glycine, serine, alanine, and the sum of all other amino acids presented in green, blue, red and grey respectively.

Codon Bias

Embiopteran fibers have a high glycine content (over 45%) and, to a lesser extent, alanine proportion (5-15%). Because the codons for glycine and alanine are guanine/cytosine (GC) rich (GGN and GCN, respectively), it might be expected that the fibroin genes would have a high GC content. However, the highest GC content observed for embiopteran fibroin coding sequence is the 56% of *Aposthonia* and *Antipaluria*. The fibroin genes for the other species have GC contents of 49-54%. One of the reasons for the approximately equal guanine/cytosine to adenosine/thymine ratio in embiid fibroin genes is codon usage bias.

Codon bias is the differential use of synonymous codons within a genome or a specific gene. Codon bias has been shown to be conserved within taxonomic lineages. For example, mammalian genomes tend to have an overabundance of codons with

guanine or cytosine in the third nucleotide position (Grantham, 1980). Although mutational bias, the uneven mutation rate between bases, may result in synonymous changes within a gene, selection on codon bias may favor codons with more abundant tRNAs (Hershberg and Petrov, 2008). The genomes of numerous organisms are thought to have codon bias to improve transcription, control translational efficiency and modulate translation, and mRNA stability (Grantham, 1980). In *Bombyx mori* heavy chain fibroin mRNA, the pattern of codon usage throughout the 15 kb transcript stabilizes hairpins, which may be necessary for the functional state of the mRNA. The tRNA population in the silk (salivary) gland cells of *B. mori* at the time of pupation is strongly correlated to the codon usage in the heavy chain fibroin mRNA (Mita *et al.*, 1988).

Embiid silk genes are biased against having guanine or cytosine in the third position of codons for the three most common amino acids (Figure 3.5). The bias against third position guanine is much stronger than third position cytosine based on percentages observed from cDNA sequences (supplemental Table 3.1). The same codon bias is conserved across all embiid fibroin genes. Furthermore, the same bias against third codon position guanines and cytosines has been observed in both moth heavy chain fibroin and spider fibroin genes (Mita *et al.*, 1988; Xu and Lewis, 1990; Hayashi *et al.*, 2004). There are several possible reasons for codon bias in fibroin genes, such as translational efficiency and preventing frameshift mutations during DNA replication (Grantham, 1980).

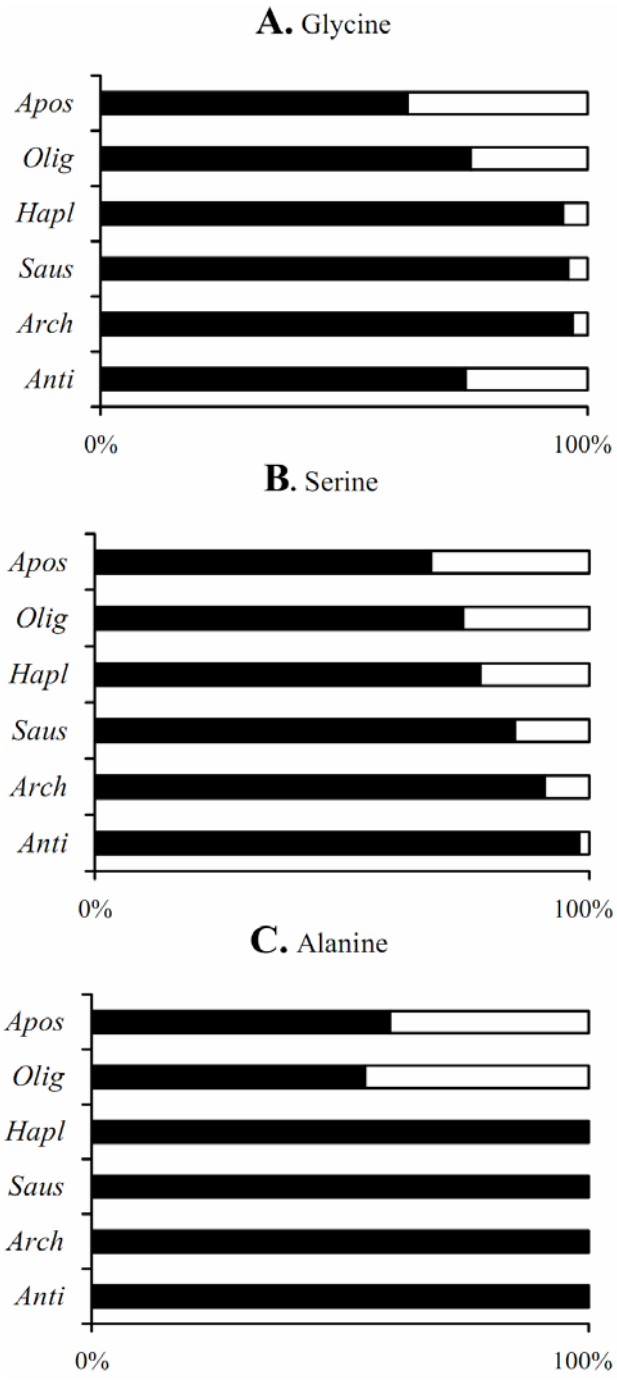


Fig. 3.5. Codon usage for the most common amino acids in embiopteran fibroins. Frequencies of synonymous codons with third position adenine or thymine (black) versus third position cytosine or guanine (white). A. Glycine codon usage (frequencies of GGA+GGT versus GGC+GGG). B. Serine codon usage (frequencies of AGT+TCA+TCT versus AGC+TCC+TCG). C. Alanine codon usage (frequencies of GCA+GCT versus GCC+GCG). Taxon abbreviations: *Antipaluria* (*Anti*), *Archembia* (*Arch*), *Saussurembia* (*Saus*), *Haploembia* (*Hapl*), *Oligotoma* (*Olig*), *Aposthonia* (*Apos*).

Codon bias in embiopteran fibroin genes may help prevent frameshift mutations. Highly repetitive sequences such as minisatellites often have high recombination rates due to unequal crossing over or slip strand mispairing (Žurovec and Sehnal, 2002). Similarly, repetitive nucleotide sequences underlie the recurring amino acid motifs in the embiopteran fibroins (Fig. 3.2). These fibroins have a high proportion of glycine and alanine, which are encoded by the four-fold degenerate codons, GGN and GCN, respectively. Hence, codon bias against third position guanines and cytosines prevents the sequences from containing stretches of poly-guanines (e.g., GGGG) or guanine-cytosine couplets (e.g., GCGC). If such sequences were present in fibroin genes, during recombination events, these regions are expected to have a higher frequency of frameshift mutations compared to regions with runs of codons that are punctuated by third position adenines and thymines. Selection against third codon position guanine and cytosine may help to prevent nonsense or truncated protein products that would be deleterious to silk production and thus to the embiid.

Codon bias in embiopteran fibroin genes may function in another manner to enhance silk production. Because webspinners produce large quantities of silk on a daily basis, numerous fibroin mRNA transcripts are likely to be present at all times within silk gland cells (Edgerly *et al.*, 2006). mRNAs can and do form secondary structures that are determined by the nucleotide sequences (Mita *et al.*, 1988). Some of these secondary structures can inhibit translation such as the G-quadruplex conformation, which is created by hydrogen bonding among four inter- or intra-molecular G-tracks (four guanines in a row (Williamson, 1993; Lipps and Rhodes, 2009). These very stable structures can have

melting temperatures that are 20-30°C higher than double stranded DNA and may inhibit translation by limiting initiation via affecting scanning of the 43S pre-initiation complex or ribosome translocation. (Lipps and Rhodes, 2009).

In embiid fibroin genes, the codon bias against third position guanines is reflected in the near absence of stretches of four or more contiguous guanines, thereby preventing the formation of G-quadruplexes. Across all six species, there are only two instances of G-tracks in fibroin cDNAs. In both cases, the G-tracks result from the GGG codon for glycine followed by a codon with a first position G (in *Oligotoma*, GGG GAT; in *Saussurembia*, GGG GCA). Through codon bias in embiopteran fibroin silk genes, multiple G-tracks are prevented, thereby potentially improving the translational efficiency of these highly repetitive genes.

Fibroin conservation

An embiid encased in Burmese amber establishes mid-Cretaceous (~ 100 million years ago) as the minimum age of the order Embioptera (Engel and Grimaldi, 2006a). Since all extant embiopterans use silk, tarsal silk production is considered a synapomorphy (shared derived trait) of the order. Hence, the conservative estimate for the evolution of silk production in the embiid ancestor is the mid-Cretaceous. But the order is presumably older than this estimate because the fossil embiid has morphological characters that identify it as belonging among derived lineages. An earlier origination of Embioptera is supported by a sclerogibbid wasp that is entombed in Lebanese amber dated to the early Cretaceous (Engel and Grimaldi, 2006b). Extant sclerogibbid wasps are obligate embiid parasites and thus if this amber wasp also had an embiid host, then the

minimum age of Embioptera and tarsal silk production is pushed to ~140 millions years ago.

All known webspinners utilize silk to construct protective tubes and egg coverings (Ross, 2000). Because embiids rely on silk for essentially the same vital functions, it is expected that there should be persistent natural selection on silk production in all species. Another prediction is that selection may act on fibroins to conserve characteristics crucial for silk fiber formation and performance. In caterpillar and spider silks, the repetitive region motifs provide insight into silk mechanical properties and features of the carboxyl-terminal regions are thought necessary for fiber formation (Bini *et al.*, 2004; Wong Po Foo *et al.*, 2006; Ittah *et al.*, 2007). Similarly, in embiopteran fibroins, these regions may also impart information about silk fibroin processing and fiber mechanical performance.

Serine participates in glycine-serine couplets and is the second most abundant amino acid in embiopteran silk fibroins and spun silk (see above; Figs. 3.2 and 3.4). Indeed, embiopteran silks are particularly rich in serine compared to spider dragline and moth silks (6% and 12% respectively; Chang *et al.*, 2005). Because serine is polar and therefore hydrophilic, it may play a role in fibroin solubility within embiopteran silk glands. In spider dragline fibroins and moth heavy chain fibroins, the hydrophobic repetitive regions are interspersed with hydrophilic motifs that aid protein solubility within the glandular lumen (Bini *et al.*, 2004). By contrast, the embiopteran fibroin sequences do not have these obvious hydrophilic motifs. Instead, the greater abundance of serine residues in embiopteran fibroins may serve an analogous function to the

hydrophilic regions of moth and spider fibroins. Furthermore, serine residues may also aid in β-sheet formation (Minor and Kim, 1994).

Glycine-serine and glycine-alanine couplets dominate the repetitive region of all embiid fibroins (Fig 3.2). These couplets form β-sheets that compose over 45% of embiopteran silk fibers (Collin *et al.*, 2009a). β-sheets stack into β-crystalline structures that provide strength and stiffness, which help protect the embiids in their galleries. Glycine-serine couplets outnumber glycine-alanine couplets in most fibroins, based on the deduced fibroin sequences (Fig. 3.2). Ratios of glycine-serine versus glycine-alanine couplets are skewed from the moderate 46:33 in *Saussurembia* to the dramatic 94:0 in *Aposthonia*. These serine-rich regions may enhance stiffness due to the greater propensity of serine to form β-sheets than alanine or glycine (Minor and Kim, 1994). Furthermore, the hydroxyl groups of the serine residues likely form additional hydrogen bonds within the β-crystalline structure, possibly enhancing tensile strength (Takahashi *et al.*, 1999). Conservation of these repetitive motifs may explain the similar tensile strengths of embiid silks. For example, mean tensile strengths for webspinner silks differ by only 30 MPa and are not significantly different across taxa (Collin *et al.*, 2009a).

Sequence conservation is not limited to the repetitive region, but is also apparent in the carboxyl-terminal region of webspinner fibroins (Fig. 3.6). In moth heavy chain and spider dragline fibroins, the carboxyl-terminal regions are conserved (within each taxonomic group) and critical for fiber formation (Bini *et al.*, 2004; Wong Po Foo *et al.*, 2006). For example, in spider dragline fibroins, the carboxyl-terminal region is ~75% identical across species (Ittah *et al.*, 2007). Similarly, conservation of the webspinner

carboxyl-terminal region suggests that it, too, has functional significance to fiber formation. Across five of six species, embiid fibroin carboxyl-terminal regions share specific attributes, such as a low pI (the pH at which a molecule has zero net charge) and placements of cysteine residues, aromatic residues, and amino acid chemical types (Fig. 3.6). Although the carboxyl-terminal region of *Haploembia* is markedly different from that of the other five species, it does share some of these common features, particularly a low pI and conserved aromatic residues.

Fig. 3.6. Alignment of embiopteran fibroin carboxyl-terminal regions and their pIs, on the right. Glycine, alanine, and serine colored as in Fig. 3.2. Acidic amino acids are in orange and cysteines in purple. Shading shows >50% conservation of chemical type: acidic (D, E), hydrophobic (A, G, I, L, V), amine (N, Q), aromatic (F, W, Y), basic (K, R, H), hydroxyl (S, T), proline (P), sulfur-containing (C, M). The *Haploembia* fibroin carboxyl-terminal region is separated from the other carboxyl-terminal regions due its lower sequence identity.

In spider dragline fibroins, the carboxyl-terminal region forms a barrel arrangement of four α-helices with a loop projecting perpendicularly from the top (Ittah *et al.*, 2007). A conserved cysteine residue in these fibroins is located within the loop for optimal positioning of disulfide bond formation. Additionally, next to this cysteine is an aspartic acid residue that is suggested to have a stabilizing effect on the loop through the formation of a salt bridge (Ittah *et al.*, 2007). Analogously, the conservation of amino acid sequence and chemical type in webspinner fibroin carboxyl-terminal regions (except

in *Haploembia*) may also form a secondary structure that is important for fiber formation. This carboxyl-terminal secondary structure must be smaller than the structure in spiders due to differences in carboxyl-terminal lengths (~30 amino acids in webspinner fibroins vs. ~90 in spider dragline fibroins; Ittah *et al.*, 2007).The embiid carboxyl-terminal regions may have a secondary structure dictated by the disulfide bonding of the paired cysteine residues. Notably, except in *Haploembia*, the embiid fibroins also have a conserved aspartic acid residue positioned adjacent to the cysteine residue at -27 amino acids from the carboxyl-terminus, reminiscent of the cysteine-aspartic acid of spider dragline fibroins.

Conservation of aromatic amino acids among the embiopteran fibroins suggests a functional significance. The secondary structure of the carboxyl-terminal region in embiopteran fibroins may be stabilized by aromatic-aromatic interactions. Aromatic side chains separated by a minimum of seven amino acids can interact at dihedral angles between 50° and 90° (Burley and Petsko, 1985). Five conserved aromatic residues are located within the webspinner fibroin carboxyl-terminal regions (Fig. 3.6). The aromatic residue at -23 or -20 may interact with residues at -9 or -5 from the carboxyl-terminus to stabilize a secondary structure in this region (mean distance between aromatic pairs is 14.5 residues). Notably, lepidopteran and trichopteran light chain fibroins have several aromatic residues that are consistently positioned throughout the protein. These amino acids have been hypothesized to enhance the assembly of the micellar formation in lepidopteran fibroins (Chapter 4).

As with lepidopteran light chain fibroins, embiid fibroins have conservation of a low pI (Fig. 3.6). In *B. mori* (silkworm) silk glands, the pH decreases from 6.9 in the posterior portion to 4.8 in the anterior section (Wong Po Foo *et al.*, 2006). The reduction of pH along the path of the silk gland in combination with the low overall pI (4.39) of the heavy chain fibroin-light chain fibroin heterodimer may indicate charge suppression by acidification to assist in fiber formation (Wong Po Foo *et al.*, 2006). At a neutral pH, molecular aggregation may be inhibited by the repulsion of the negatively charged fibroins. Based on the conservation of low pIs across embiopteran fibroins, an acidification process similar to what occurs in lepidopterans may be taking place, thus assisting molecular aggregation and fiber formation. This possible convergence is striking given the differences between the embiopteran and lepidopteran silk gland architectures (i.e., tiny, spheroid glands with short ducts versus large, elongated glands with long ducts).

Most of the embiid carboxyl-terminal sequences have two conserved cysteine residues, which may help in polymerization by linking monomers (see above). These cysteines are similarly positioned to the cysteines in the carboxyl-terminal regions of lepidopteran heavy chain fibroins (Collin *et al.*, 2009b). Within most lepidopteran heavy chain fibroins, the cysteine residues form inter- and intra-molecular disulfide bonds that link the heavy chain and light chain fibroins into heterodimers, which then aggregate with other heterodimers (Mori *et al.*, 1995). Although there may be only one type of fibroin within embiid silk, the two cysteines in the fibroin carboxyl-terminal region may interlink fibroin monomers through disulfide bonding, thereby facilitating fiber formation.

In *Haploembia* however, the fibroin carboxyl-terminal region conspicuously lacks the two conserved cysteine residues, among other autapomorphic features, compared to the other embiid sequences (Fig 3.6). All five *Haploembia* fibroin clones that were discovered are missing these cysteines. *Oligotoma* and *Aposthonia*, which are in the same family as *Haploembia* (Oligotomidae; Fig. 3.1), both have these two conserved cysteines, as do species outside of Oligotomidae. Therefore, the lack of the cysteines in *Haploembia* is explainable as secondary lost. Absence of the cysteine residues may have implications for *Haploembia* silk use that stem from reduced fiber formation efficiency. Diminished fiber formation ability resulting from a missing cysteine residue is supported by experiments done with recombinant silk. Ittah *et al.* (2007) engineered a spider dragline silk construct so that one of the conserved cysteine residues was replaced with serine. Although fiber formation still occurred in their expression system, the fibers were only 50% the usual length.

Haploembia may similarly have a reduced capacity to produce silk due to their lack of cysteine residues in the carboxyl-terminal region of their fibroin. In lab cultures and in the field, *Haploembia* does not spin as much visible silk as other species (pers. obs.; Edgerly, pers. comm.). *Haploembia* silk sheets are often wispy and sheer, unlike the typical opaque sheets produced by other embiids. *Haploembia* also prefer to live within substrates (e.g., within tree bark crevasses) or beneath substrates (e.g., under rocks, or underground) rather than on tree surfaces. Perhaps the greater use of substrate by *Haploembia* is associated with the reduced silk production that may be caused by the molecular changes in their fibroin carboxyl-terminal region.

In summary, we revealed several similarities among embiopteran fibroins through comparative analyses. The critical significance of silk to webspinner ecology suggests stabilizing selection has acted on these adaptive molecules and the sequence conservation that we observe may have persisted for over 140 million years. Embiopteran silk genes show codon biases (Fig 3.5) that could likely prevent frameshift mutations and ensure efficient translation of these highly repetitive genes (Fig. 3.3). Among embiopteran species, conservation of fibroin sequence elements, particularly the repetitive motifs (Fig. 3.2) and the carboxyl-terminal region (Fig 3.6), suggest that secondary structures and modes of fiber formation have also remained largely unchanged. This is corroborated by the uniform tensile strengths and structural profiles of embiopteran silks across taxa (Collin *et al.*, 2009a). Common features of the carboxyl-terminal regions such as conserved cysteine and aromatic residues, and low pIs (Fig. 3.6), signify their potentially key roles in embiopteran fiber formation. Through a multi-species comparative approach, we have developed a more comprehensive understanding of embiid silk structure, function, and evolution.

References

Beuken, E., Vink, C., Bruggeman, C.A., 1998. One-step procedure for screening recombinant plasmids by size. Biotechniques. 24, 748-750.

Bini, E., Knight, D. P., Kaplan, D. L., 2004. Mapping domain structures in silks from insects and spiders related to protein assembly. Journal of Molecular Biology. 335, 27-40.

Burley, S. K., Petsko, G. A., 1985. Aromatic-aromatic interaction - a mechanism of protein-structure stabilization. Science. 229, 23-28.

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

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

Collin, M. A., Garb, J. E., Edgerly, J. S., Hayashi, C. Y., 2009b. Characterization of silk spun by the embiopteran, *Antipaluria urichi*. Insect Biochemistry and Molecular Biology. 39, 75-82.

Duelli, P., 1986. A missing link in the evolution of the egg pedicel in lacewings. Experientia. 42, 624-624.

Edgerly, J. S., 1987a. Colony composition and some costs and benefits of facultatively communal behavior in a Trinidadian webspinner, *Clothoda urichi* (Embiidina, Clothodidae). Annals of the Entomological Society of America. 80, 29-34.

Edgerly, J. S., 1987b. Maternal behavior of a webspinner (Order Embiidina). Ecological Entomology. 12, 1-11.

Edgerly, J. S., Shenoy, S. M., Werner, V. G., 2006. Relating the cost of spinning silk to the tendency to share it for three embiids with different lifestyles (Order Embiidina: Clothodidae, Notoligotomidae, and Australembiidae). Environmental Entomology. 35, 448-457.

Engel, M. S., Grimaldi, D. A., 2006a. The earliest webspinners (Insecta: Embiodea). American Museum Novitates. 3514, 1-15.

Engel, M. S., Grimaldi, D. A., 2006b. The first Cretaceous sclerogibbid wasp (Hymenoptera: Sclerogibbidae). American Museum Novitates. 3515, 1-7.

Fedič, R., Žurovec, M., Sehnal, F., 2003. Correlation between fibroin amino acid sequence and physical silk properties. Journal of Biological Chemistry. 278, 35255-35264.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In: Walker, J. M. (Ed.) The Proteomics Protocols Handbook. Humana Press.

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

Grantham, R., 1980. Workings of the genetic code. Trends in Biochemical Sciences. 5, 327-331.

Hayashi, C. Y., Blackledge, T. A., Lewis, R. V., 2004. Molecular and mechanical characterization of aciniform silk: Uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. Molecular Biology and Evolution. 21, 1950-1959.

Hershberg, R., Petrov, D. A., 2008. Selection on codon bias. Annual Review of Genetics. 42, 287-299.

Ittah, S., Cohen, S., Garty, S., Cohn, D., Gat, U., 2006. An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. Biomacromolecules. 7, 1790-1795.

Ittah, S., Michaeli, A., Goldblum, A., Gat, U., 2007. A model for the structure of the C-terminal domain of dragline spider silk and the role of its conserved cysteine. Biomacromolecules. 8, 2768-2773.

Lipps, H. J., Rhodes, D., 2009. G-quadruplex structures: in vivo evidence and function. Trends in Cell Biology. 19, 414-422.

Minor, D. L., Kim, P. S., 1994. Context is a major determinant of beta-sheet propensity. Nature. 371, 264-267.

Mita, K., Ichimura, S., Zama, M., James, T. C., 1988. Specific codon usage pattern and its implications on the secondary structure of silk fibroin messenger-RNA. Journal of Molecular Biology. 203, 917-925.

Mori, K., Tanaka, K., Kikuchi, Y., Waga, M., Waga, S., Mizuno, S., 1995. Production of a chimeric fibroin light-chain polypeptide in a fibroin secretion-deficient naked pupa mutant of the silkworm *Bombyx mori*. Journal of Molecular Biology. 251, 217-228.

Okada, S., Weisman, S., Trueman, H. E., Mudie, S. T., Haritos, V. S., Sutherland, T. D., 2008. An Australian webspinner species makes the finest known insect silk fibers. International Journal of Biological Macromolecules. 43, 271-275.

Ross, E. S., 2000. Contributions to the biosystematics of the insect order Embiidina. Part 1. Origin, relationships and integumental anatomy of the insect order Embiidina. Occasional Papers California Academy of Sciences. 149, 1-53.

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, 171-188.

Swanson, B. O., Blackledge, T. A., Summers, A. P., Hayashi, C. Y., 2006. Spider dragline silk: Correlated and mosaic evolution in high-performance biological materials. Evolution. 60, 2539-2551.

Szumik, C., Edgerly, J. S., Hayashi, C. Y., 2008. Phylogeny of embiopterans (Insecta). Cladistics. 24, 993-1005.

Takahashi, Y., Gehoh, M., Yuzuriha, K., 1999. Structure refinement and diffuse streak scattering of silk (*Bombyx mori*). International Journal of Biological Macromolecules. 24, 127-138.

Taylor, J. S., Breden, F., 2000. Slipped-strand mispairing at noncontiguous repeats in *Poecilia reticulata*: A model for minisatellite birth. Genetics. 155, 1313-1320.

Thompson, J. D., Higgins, D. G., Gibson, T. J., 1994. Clustal-W improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research. 22, 4673-4680.

Williamson, J. R., 1993. Guanine quartets. Current Opinion in Structural Biology. 3, 357-362.

Wong Po Foo, C., Bini, E., Hensman, J., Knight, D. P., Lewis, R. V., Kaplan, D. L., 2006. Role of pH and charge on silk protein assembly in insects and spiders. Applied Physics A (Materials Science Processing). A82, 223-233.

Xu, M., Lewis, R. V., 1990. Structure of a protein superfiber - spider dragline silk. Proceedings of the National Academy of Sciences of the United States of America. 87, 7120-7124.

Yonemura, N., Mita, K., Tamura, T., Sehnal, F., 2009. Conservation of silk genes in Trichoptera and Lepidoptera. Journal of Molecular Evolution. 68, 641-653.

Yonemura, N., Sehnal, F., 2006. The design of silk fiber composition in moths has been conserved for more than 150 million years. Journal of Molecular Evolution. 63, 42-53.

Žurovec, M., Sehnal, F., 2002. Unique molecular architecture of silk fibroin in the waxmoth, *Galleria mellonella*. Journal of Biological Chemistry. 277, 22639-22647.

Supplemental Table

Supplemental Table 3.1. Codon usage in the fibroin cDNAs for the most common amino acids (aa) of embiopteran silk proteins. Taxon abbreviations: *Antipaluria* (*Anti*), *Archembia* (*Arch*), *Saussurembia* (*Saus*), *Haploembia* (*Hapl*), *Oligotoma* (*Olig*), *Aposthonia* (*Apos*).

		Anti	*Arch*	*Saus*	*Hapl*	*Olig*	*Apos*
aa	codon	% codon	% codon	% codon	% codon	% codon	% codon
Gly	GGA	36	75	56	51	45	34
	GGC	23	3	3	5	20	36
	GGG	2	0	1	0	4	1
	GGT	39	22	40	44	31	29
Ser	AGC	1	9	15	11	15	21
	AGT	5	7	2	60	39	18
	TCA	58	62	66	14	15	14
	TCC	1	0	0	11	12	11
	TCG	0	0	0	0	0	0
	TCT	35	22	17	4	19	36
Ala	GCA	86	86	60	97	32	0
	GCC	0	0	0	0	36	40
	GCG	0	0	0	0	9	0
	GCT	14	14	40	3	23	60

Chapter 4

Molecular evolution of lepidopteran silk proteins:

insights from the ghost moth, *Hepialus californicus*

Abstract:

Silk production has independently evolved in numerous arthropod lineages, such as Lepidoptera, the moths and butterflies. Lepidopteran larvae (caterpillars) synthesize silk proteins in modified salivary glands and spin silk fibers into protective tunnels, escape lines, and pupation cocoons. Molecular sequence data for these proteins are necessary to determine critical features of their function and evolution. To this end, we constructed an expression library from the silk glands of the ghost moth, *Hepialus californicus*, and characterized *light chain fibroin* and *heavy chain fibroin* gene transcripts. The predicted *H. californicus* silk fibroins share many elements with other lepidopteran and trichopteran fibroins, such as conserved placements of cysteine, aromatic, and polar amino acid residues. Further comparative analyses were performed to determine site-specific signatures of selection and to assess whether fibroin genes are informative as phylogenetic markers. We found that purifying selection has constrained mutation within the fibroins and that light chain fibroin is a promising molecular marker. Thus by characterizing the *H. californicus* fibroins, we identified key functional amino acids and gained insight into the evolutionary processes that have shaped these adaptive molecules.

Keywords: Lepidoptera, Trichoptera, silk, fibroin, cDNA, gene tree, purifying selection

Introduction

Silks are ideal for the study of adaptive evolution because they have independently arisen in numerous arthropod lineages. Typically, silks are primarily made of proteins that are dominated by non-essential amino acids, such as glycine, alanine, and serine (Gatesy *et al.*, 2001). Arthropods use silk for a broad range of ecological functions, such as prey capture, protection, reproduction, and dispersal (Craig, 1997). Silk proteins that perform similar tasks tend to possess particular attributes that have either been maintained over long periods of evolutionary time or arose convergently. These conserved protein traits can range from the uniform positioning of a single cysteine residue, to more complex features, such as long regions of high sequence identity. For spider silks, studying these conserved features has helped determine important sequence elements and molecular evolutionary patterns of silk proteins (Guerette *et al.*, 1996; Garb and Hayashi, 2005; Ittah *et al.*, 2006).

Like spiders, Trichoptera (caddisflies) is a lineage with silk proteins that have highly conserved elements (Yonemura *et al.*, 2009). Caddisfly larvae synthesize silk proteins in paired labial glands, whose original function was the secretion of saliva. The larvae spin silk underwater to construct aquatic capture nets, domiciles, and pupation cocoons. Silk proteins have been studied in species sampled from each of the three trichopteran suborders (Yonemura *et al.*, 2006; Yonemura *et al.*, 2009). In these exemplars, the core of the larval silk fiber is formed from two proteins, heavy chain fibroin (H-fibroin) and light chain fibroin (L-fibroin). Across trichopteran suborders,

there is ~50% sequence identity within the L-fibroin sequence and 56% sequence identity in the amino terminal region of H-fibroin (Yonemura *et al.*, 2009).

H-fibroin and L-fibroin genes and proteins are also conserved in Lepidoptera (moths and butterflies). Larval lepidopterans (caterpillars) use silk for diverse functions that in most species include constructing pupal cocoons. Lepidoptera is the sister group to Trichoptera, and together they form the supraorder Amphiesmenoptera (Kristensen, 1999; Whiting, 2002). Unlike caddisfly larvae that spin their silk underwater, caterpillars spin silk in terrestrial environments. Except for this difference, Lepidoptera produce silk in a homologous manner to Trichoptera. Specifically, silk proteins are secreted in paired labial glands (Grimaldi and Engel, 2005). The two insect orders share other aspects of silk production. Regions of sequence similarity have been observed throughout their L-fibroins and in the terminal regions of their H-fibroins (Sehnal and Žurovec, 2004; Yonemura and Sehnal, 2006; Yonemura *et al.*, 2009). The use of L- and H-fibroins in amphiesmenopteran silk production has been hypothesized to be conserved for over 250 million years (Yonemura *et al.*, 2009). However, unlike trichopteran silk, most previously studied moth silks include an additional protein, P25 (Inoue *et al.*, 2000; Sehnal and Žurovec, 2004). L-fibroin, H-fibroin, and P25 are thought to have formed the core of lepidopteran silk fibers for over 150 million years (Yonemura and Sehnal, 2006).

To date, all silk studies of Lepidoptera have focused on members of Ditrysia, which includes most moth and butterfly species (Mita et al., 1994; Žurovec and Sehnal, 2002; Yonemura and Sehnal, 2006). However, Ditrysia is a highly derived clade (Friedlander et al., 1996) and so to provide better insight into the composition and

evolution of lepidopteran silk, we focused on the ghost moth, *Hepialus californicus*. *Hepialus* is a member of Exoporia, a basal lepidopteran lineage of moths with primitive appearances (Nielsen et al., 2000, Wiegmann et al., 2002). *H. californicus* larvae feed on the roots of lupines (Strong et al., 1996) and live in silk lined subterranean tunnels that they build around the host plant's roots (Nielsen et al., 2000).

In this study we describe H-fibroin and L-fibroin transcripts in a cDNA library constructed from the silk glands of *H. californicus*. We compare the fibroin sequences to those known in other Lepidoptera and Trichoptera. We then determine the rate of molecular evolution at each amino acid site and assess the potential phylogenetic utility of the silk genes across Amphiesmenoptera. One of our aims is to reconstruct the silk groundplan for Lepidoptera by ascertaining whether *H. californicus* utilizes three core silk proteins like previously studied Lepidoptera, or only two silk proteins as found in Trichoptera. In addition, by comparing the sequences of *H. californicus* silk to other known fibroins, we shed light on the evolution and function of these molecules of adaptive significance.

Methods

Tissue collection

Hepialus californicus (Lepidoptera: Hepialidae) larvae were collected from the Bodega Marine Reserve (38°32' N 123°07' W) in Sonoma County, CA, USA, and placed in individual culture trays (3.5 cm diameter) with moistened filter paper and organic carrot chunks. The larvae were kept at ambient room temperature until dissection. At that

time, eighteen final instar larvae were anesthetized with CO_2 gas and then individually placed in a dissection dish filled with 0.15 M sodium chloride, 0.015 M sodium citrate buffer. Once submerged within the solution, the head was gently disconnected from the thorax so as not to rupture the paired labial silk glands. For each larva, the silk gland pair was carefully separated from the head by delicately pulling on the anterior gland regions, placed in a 1.5 ml microfuge tube, immediately flash frozen in liquid nitrogen, and stored at –80°C.

cDNA library construction and sequencing

Frozen silk glands were pulverized under a small volume of liquid nitrogen with mortar and pestle and the resulting powder was added to TRIzol reagent (Invitrogen, Carlsbad, CA, USA) for RNA extraction. Total RNA was purified using the RNeasy mini kit (Qiagen, Valencia, CA, USA). After precipitation with isopropanol, the total RNA was dissolved in 0.5 % SDS containing 20 mM sodium acetate (pH 5.3) and sent to Takara Bio (Shiga, Japan) for cDNA library construction. cDNA was made with the Stratagene (La Jolla, CA, USA) cDNA Synthesis Kit then directionally cloned into pBluescript II SK(+) (Stratagene, La Jolla, CA, USA) that had been digested with EcoRI and XhoI. The resulting cDNA library size was 1.1×10^7 cfu, and the average insert size was 1.58 kb, which was estimated from 16 randomly chosen clones. Nucleotide sequences were determined from the 5' (T3) and 3' (T7) ends of 4,000 randomly chosen clones using an ABI3730XL (Applied Biosystems, Foster City, CA, USA).

Sequence characterization

Vector sequences were trimmed away and short reads of less than 100 bp were removed from the data set using Sequencher (Gene Codes, Ann Arbor, MI, USA). Batch BLASTX and BLASTN searches were performed on all sequences against the nr database (www.ncbi.nlm.nih.gov/BLAST). Both BLAST searches were performed twice, once with the low complexity filter enabled (default) and once with the filter disabled. The filter disabled searches were necessary to identify the *H-fibroin* gene, which is almost entirely highly repetitive (i.e., low complexity) sequence. All BLAST results were compared and all hits were visually inspected. Contiguous sequences (contigs) were generated using Phrap (Green 1994) employing the default settings with the exception of minimum base match, which was set to 60 instead of 20 to increase stringency. Contig assemblies were refined and checked for chimeras with Lasergene's SeqMan (DNASTAR, Madison, WI, USA). To verify the accuracy of the H-fibroin and L-fibroin contigs, a second set of these contigs was assembled using Sequencher 4.2 (settings were 90% minimum mismatch, 60 bp minimum overlap) from the sequences annotated as H-fibroin or L-fibroin based on BLAST searches.

Amino acid translations were estimated from cDNA transcripts using SeqMan (DNASTAR, Madison, WI, USA). Protparam (ExPASy tools: Gasteiger *et al.*, 2005) was used to compute amino acid composition and isoelectric point (pI).

Comparative analyses

The *H. californicus* H-fibroin and L-fibroin transcripts were compared to lepidopteran and trichopteran sequences obtained from the NCBI database. Specifically,

coding sequences for the H-fibroins of *Ephestia kuehniella* (accession number AY253535), *Antheraea yamamai* (AF410906), *Bombyx mandarina* (DQ459410), *Bombyx mori* (NM_001113262), *Galleria mellonella* (AF095240), *Hydropsyche angustipennis* (AB214507), *Limnephilus decipiens* (AB214509), *Rhyacophila obliterata* (AB354588), and *Yponomeuta evonymellus* (AB195979); and the L-fibroins of *Bombyx mandarina* (AB001820), *Bombyx mori* (X17291), *Dendrolimus spectabilis* (AB001822), *Galleria mellonella* (S77817), *Hydropsyche angustipennis* (AB214508), *Limnephilus decipiens* (AB214510), *Papilio xuthus* (AB001824), *Rhyacophila obliterata* (AB354590), and *Yponomeuta evonymellus* (AB195977) were downloaded. Sequence alignments were performed with ClustalW (Thompson *et al.*, 1994) and refined by eye.

Signatures of selection on the protein coding regions of L-fibroin and the carboxyl-terminal region of H-fibroin were examined through maximum likelihood analyses using the codeml package of PAML (Yang 2007). PAML determined ω, the ratio of non-synonymous substitutions per non–synonymous site to synonymous substitutions per synonymous site, for each codon. We used an input tree based on recent studies of higher-level lepidopteran relationships (Kjer *et al.*, 2002; Whiting, 2002; Wiegmann *et al.*, 2002) and analyzed the data with four different models: M1a, neutral model; M2a, selection model; M7, beta distribution neutral model; and M8, beta distribution with selection. Models M1a versus M2a and M7 versus M8 were then compared using likelihood ratio tests (Yang *et al.* 2005).

Phylogenetic analyses were conducted in PAUP* v.4.0b10 (Swofford, 1998). In parsimony analyses, all character transformations were weighted equally. In maximum

likelihood analyses, models of DNA evolution were chosen by the Akaike information criterion implemented in MODELTEST v.3.6 (Posada and Crandall, 1998) and PAUP*. Parsimony searches utilized the branch and bound method and maximum likelihood searches were heuristic. Both parsimony and maximum likelihood searches were performed with random taxon addition and tree-bisection and reconnection. Nodal support values were assessed using bootstrap analyses (Felsenstein, 1985). Bootstrap replicates were heuristic with tree-bisection and reconnection branch swapping. For parsimony searches, 10,000 iterations with 100 random taxon additions were done and 1000 iterations with 10 random taxon additions were executed in maximum likelihood searches.

Results and Discussion

Identification of silk cDNAs

The *H. californicus* silk gland cDNA library was constructed and 4000 clones were sequenced from both ends. From these reads, 600 contigs that contained between two to 80 sequences were assembled with Phrap (Green 1994). In addition to contigs of cellular and tissue maintenance genes, we identified several contigs of two major silk genes, *H-fibroin* and *L- fibroin*.

There were 27 clones that encoded long repetitive regions. The top BLAST hit for each of these clones was the H-fibroin from the Japanese oak silkmoth, *Antheraea yamamai* (Genbank AF410906). The similarity was largely due to the presence of frequent stretches of poly-alanine in the repetitive region. A representative clone containing 305

codons for a portion of the repetitive region of *H. californicus* H-fibroin (Genbank GU144521) is shown in Figure 4.1A. The longest clone with both repetitive region and a complete carboxyl-terminal region (Genbank GU144520) encoded 457 amino acid residues (Fig. 4.1B).

A.

B.

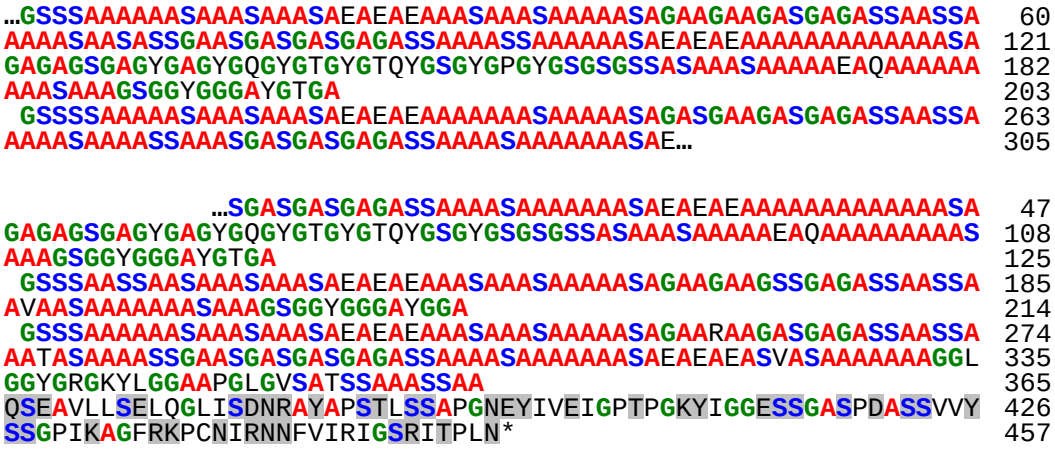


Fig. 4.1. *Hepialus californicus* heavy chain fibroin. (A) Middle segment clone depicting a full-length and partial-length ensemble repeat (GU144521). (B) Carboxyl-terminal clone with upstream non-repetitive region (GU144520). Single letter abbreviation for amino acids, residue position number on the right, and ellipsis (…) for additional, undetermined sequence. Highlights for glycine (green), serine (blue), alanine (red), acidic motifs in the repetitive region (underlined), and polar residues in the non-repetitive region (grey shading).

In addition to the 27 H-fibroin transcripts, 537 clones of our cDNA library contained L-fibroin transcripts. Many of these clones were full-length, as determined from the bidirectional sequencing reads (Genbank GU180666-GU180675; Fig. 4.2). The top BLAST hit for all these clones was the L-fibroin of the greater wax moth, *Galleria mellonella* (Genbank S77817). L-fibroin was by far the most numerous transcript identified within the cDNA library, reflecting a high expression level of the *L-fibroin* gene in the silk glands.

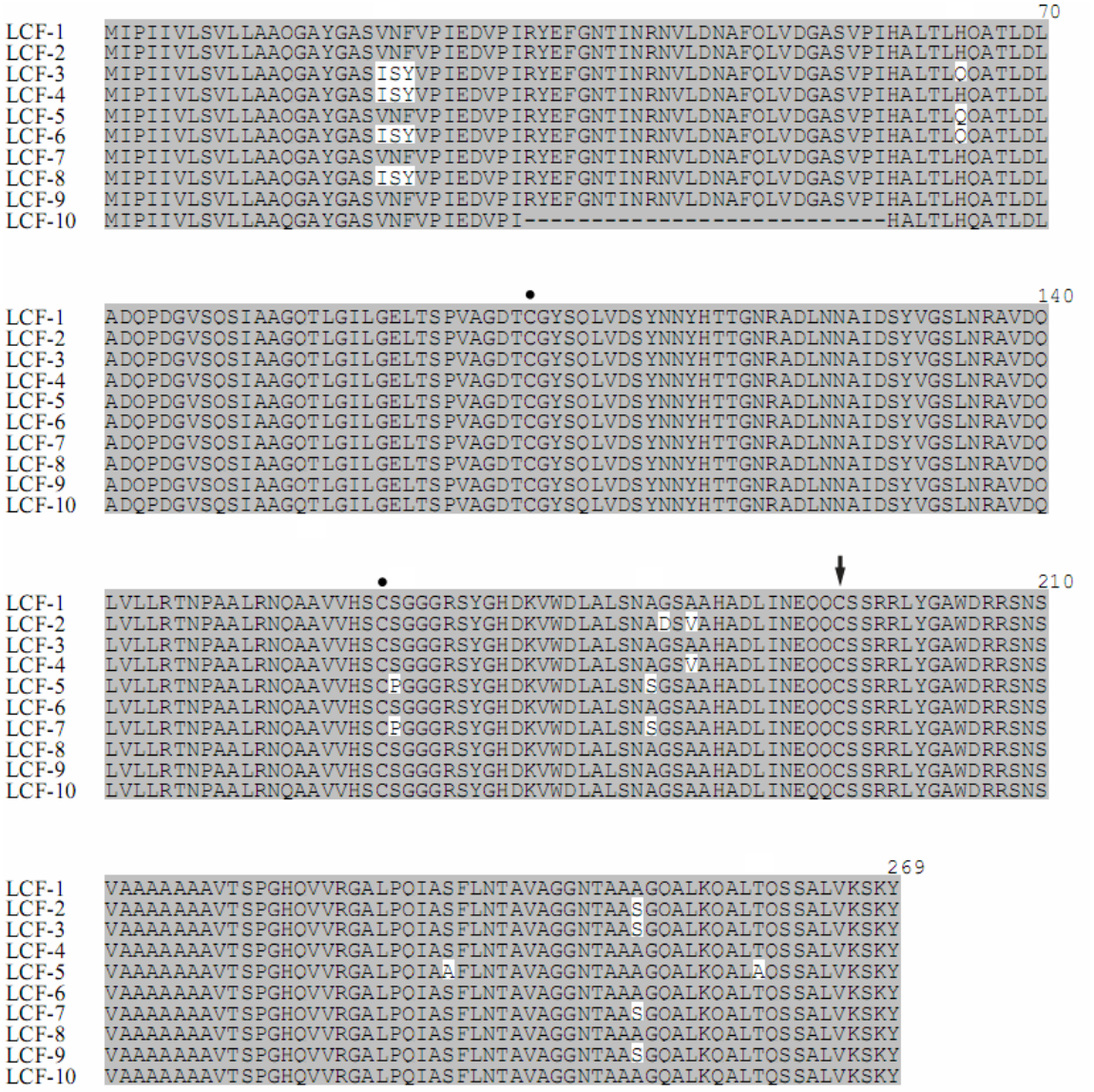


Fig. 4.2. Alignment of *Hepialus californicus* light chain fibroin (LCF) alleles (GU180666-GU180675). Shading indicates 100% amino acid identity. Dots and arrow denote cysteines involved in bonding within L-fibroin and with H-fibroin, respectively. Dashes are alignment gaps Position number appears above each block.

Unlike previous lepidopteran silk studies (Inoue *et al.*, 2000; Tanaka and Mizuno, 2001; Sehnal and Žurovec, 2004; Yonemura and Sehnal, 2006), we did not detect a *H. californicus* transcript that corresponded to the P25 protein. This absence may indicate

that the P25 transcript was present in much lower abundance than those of *L-fibroin*, *H-fibroin*, and the ~300 other genes that were identified in the cDNA library. Such rarity would have made P25 transcripts unlikely to be sampled from our cDNA library. However, it would be unexpected for an integral component of silk to be so scarce in an active silk gland. Hence, it is more likely that *H. californicus* lacks P25, which has only been identified within the lepidopteran clade Ditrysia (Inoue *et al.*, 2000; Sehnal and Žurovec, 2004; Yonemura and Sehnal, 2006). P25 is also not known from Trichoptera, despite characterization of silk gland cDNA and N-terminal sequencing of silk gland proteins close to the expected size (Yonemura *et al.*, 2006; Yonemura *et al.*, 2009). Thus, P25 may restricted to Ditrysia. Genomic characterization of *H. californicus* would provide more definitive evidence as to whether the lack of P25 in the silk gland cDNA library was due to absence or non-expression of a P25 gene.

A third explanation for the absence of P25 in our silk gland expression library is that *H. californicus* may have secondarily lost P25. There is precedence for losing this silk component. P25 has been lost along with L-fibroin in Saturniidae, a family within the Ditrysia. Silk fibers of these silkmoths are constructed solely from H-fibroins (Tamura *et al.*, 1987; Tanaka and Mizuno, 2001). Secondary loss of P25 in *H. californicus*, however, is not the most parsimonious given the absence of P25 in Trichoptera. Analysis of silks in more basal lepidopteran clades, such as Micropterygidae, is needed to more definitively address this possibility.

Heavy chain fibroin

From the 27 *H. californicus* H-fibroin clones, we obtained 12 middle fragments that encompassed only the repetitive region (described below) and 20 end fragments that included repetitive region and the carboxyl-terminal region. No nucleotide variation was found in the 3' sequences that encoded the 92 amino acid long, non-repetitive carboxyl-terminal region (Fig. 4.1B). In the repetitive region, however, the sequences from both middle (Fig. 4.1A) and end fragments (Fig. 4.1B) varied in subrepeat length and arrangement. These differences were likely due to recombination and insertion/deletion mutations because highly iterated sequences, such as the genes for moth H-fibroins are prone to slip strand mispairing and unequal crossover (Ueda *et al.*, 1985).

Like other lepidopteran H-fibroins and some spider silk proteins (Gatesy *et al.*, 2001), the *H. californicus* H-fibroin is highly repetitive in sequence and dominated by alanine, glycine, and serine (48%, 17%, 17%, respectively, of amino acid composition; Fig. 4.1). This composition resembles the 43% alanine, 27% glycine, and 11% serine of the H-fibroin of the saturniid silkmoths, *A. yamamai* and *A. pernyi* (Sezutsu and Yukuhiro, 2000; Hwang *et al.*, 2001). In other moth H-fibroins, alanine is typically much lower (e.g., 26% in the H-fibroin of *Y. evonymellus*; Yonemura and Sehnal, 2006). The H-fibroins of caddisflies contain dramatically less alanine; in *Limnephilus decipiens,* the repetitive region of H-fibroin is completely devoid of alanine and instead is composed of glycine (25%), serine (17%), and non-polar residues (Yonemura *et al.*, 2006).

H. californicus H-fibroin resembles moth H-fibroins in that it consists of a series of repetitive motifs that form a higher-level repeat unit known as an ensemble (Fig. 4.1).

The ensemble repeat in H-fibroin of *H. californicus* is ~203 amino acids in length and includes three distinct motifs, poly-alanine, glycine-X (where X is usually alanine, serine, or tyrosine), and alanine-glutamic acid. Each of these motifs is likely to form sequence-specific secondary structures that contribute to the formation of the silk filament. The poly-alanine, glycine-alanine, and glycine-serine repeats are known to form crystalline β-sheets that provide strength and stiffness to the silk fiber (Gosline *et al.*, 1986; Guerette *et al.*, 1996). *H. californicus* H-fibroin also contains glycine-glycine-X (where X is either alanine, serine, or threonine), which is part of the α-helical structures thought to reorient the amino acid chain in many spider fibroins (Hayashi *et al.*, 1999; Hayashi and Lewis, 2001; Ashida *et al.*, 2003).

Polar amino acids are interspersed among the prevalent glycine and alanine residues in *H. californicus* H-fibroin. The acidic amino acids, such as glutamic acid, increase the hydrophilicity and decrease the pI of the H-fibroin, both of which may be important for stability of the silk dope and subsequent filament formation (Wong Po Foo *et al.*, 2006). The overall pI (the pH at which a molecule has zero net charge) of the *H. californicus* repetitive region (Fig. 4.1A) is 3.98, which is close to the 4.03 of *B. mori* H-fibroin (Wong Po Foo *et al.*, 2006). The pI of the carboxyl-terminal region (last 92 residues) of the *H. californicus* H-fibroin (Fig. 4.1B) is 8.10, and in *B. mori* it is 10.53. The high pI of this region is offset by the low pI of the L-fibroin (see below).

The large number of poly-alanine repeats within *H. californicus* H-fibroin may be required to provide structural stiffness to the silk fibers. *H. californicus* larvae live underground in silk lined tunnels and feed on the root systems of plants. In preparation

for metamorphosis, the larvae burrow into the main tap root and migrate into the plant stalk, where they spin pupation cocoons (Strong *et al.*, 1996; Nielsen et al., 2000). Because the larvae live underground, high tensile strength, which is associated with prey capture or escape line functions, is unnecessary. Instead, the stiffness provided by the poly-alanine repeats may improve tunnel stability. Keeping tunnels clear and open could facilitate the maintenance of easy escape routes and access to food resources. Stiff silk may also prevent tunnel failure in sandy or sandy loam soil, such as at the Bodega Bay Marine reserve where our *H. californicus* were collected.

While the repetitive region is important for the structural characteristics of H-fibroin, the carboxyl-terminal region is implicated in fiber formation. In *H. californicus*, this region contains 42 polar amino acids (indicated by gray shading, Fig. 4.1B). In silk proteins it is common to have a higher proportion of polar amino acids in the carboxyl-terminal region than in the repetitive region (Bini *et al.*, 2004; Sehnal and Žurovec, 2004). The polarity of these amino acids increases hydrophilicity and thus can aid solubility of H-fibroin within the glandular lumen (Bini *et al.*, 2004). Also potentially contributing to solubility is the longer length of the *H. californicus* H-fibroin carboxyl-terminal region (92 amino acids) compared to the homologous region in other lepidopteran silks. For example, in *B. mori* the equivalent region is 50 amino acids in length and in *G. mellonella* it is 60 amino acids long (Bini *et al.*, 2004). By having a larger number of hydrophilic polar amino acids, the longer non-repetitive region in *H. californicus* H-fibroin may increase silk protein solubility, and possibly compensate for lacking a homolog to the hygroscopic glycoprotein P25.

Besides affecting solubility, the carboxyl-terminal region of H-fibroin enables intermolecular bonds that are necessary for proper fiber formation (Mori *et al.*, 1995). Comparing the H-fibroins of Lepidoptera and Trichoptera reveals several amino acids that are highly conserved across taxa (Fig. 4.3). Notably, all the H-fibroins have a cysteine at ~20 residues before the protein end (marked by arrow, Fig. 4.3). This cysteine forms a disulfide bridge with the cysteine at position 190 of L-fibroin (designated by arrow, Fig. 4.2; Tanaka *et al.*, 1999). This disulfide bond is necessary for transport of the H- and L-fibroin heterodimer into the glandular lumen (Inoue *et al.*, 2004). The consistent locations of these cysteines across lepidopteran and trichopteran H- and L-fibroins underscore their importance to fiber formation. For example, lacking one of these cysteines (position 190 of L-fibroin) prevents the formation of H- and L-fibroin heterodimers, thus rendering *B. mori* incapable of spinning a cocoon (the naked pupa mutation; Mori *et al.* 1995). A similarly positioned cysteine in the carboxyl-terminal region of spider major ampullate (dragline silk) fibroin has also been implicated in silk filament formation (Ittah *et al.* 2006).

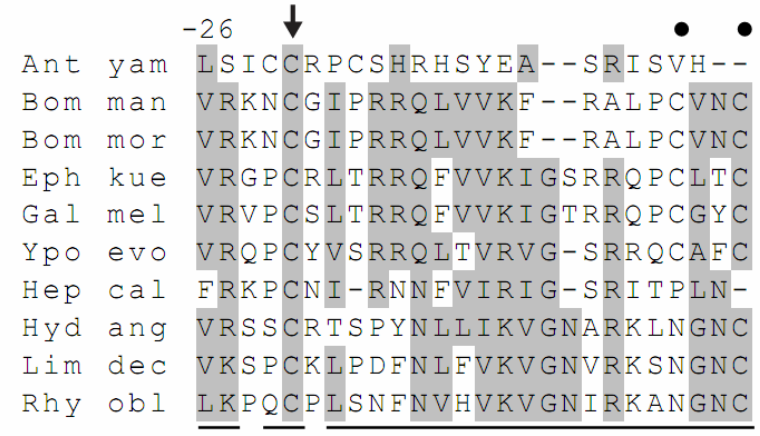


Fig. 4.3. Alignment of amphiesmenopteran heavy chain fibroin carboxyl-terminal regions. Shading shows >50% conservation of chemical type: acidic (D, E), hydrophobic (A, G, I, L, V), amine (N, Q), aromatic (F, W, Y), basic (K, R, H), hydroxyl (S, T), proline (P), sulfur (C, M). Dots and arrow denote cysteines involved in bonding within H-fibroin and with L-fibroin, respectively. Underlines at the bottom of columns mark positions that have experienced purifying selection. Dashes are alignment gaps. Positions are numbered in relation to the protein end. Taxon abbreviations and Genbank accessions: *Antheraea yamamai* (Ant yam; AF410906), *Bombyx mandarina* (Bom man; DQ459410), *Bombyx mori* (Bom mor; NM_001113262), *Ephestia kuehniella* (Eph kue; AY253535), *Galleria mellonella* (Gal mel; AF095240), *Yponomeuta evonymellus* (Ypo evo; AB195979), *Hepialus californicus* (Hep cal; GU144520), *Hydropsyche angustipennis* (Hyd ang; AB214507), *Limnephilus decipiens* (Lim dec; AB214509) and *Rhyacophila obliterata* (Rhy obl; AB354588).

While the *H. californicus* H-fibroin has the conserved cysteine at -22 (marked by arrow, Fig. 4.3), it lacks the conserved cysteines that occur in most other lepidopteran H-fibroins at positions -1 and -4 (marked by dots, Fig. 4.3). In *B. mori*, it has been shown that these latter two cysteines form an intramolecular bond within the H-fibroin molecule (Tanaka *et al.* 1999). *H. californicus*, however, is not unique in missing these cysteines. The lack of these paired cysteines in the H-fibroin of *A. yamamai* (Saturniidae) is associated with silk filament formation based on H-fibroin homodimers rather than H- and L-fibroin heterodimers (Tamura *et al.*, 1987). Similarly, their absence in the H-fibroin of *H. californicus* indicates yet another variation of silk fiber formation within lepidopterans.

Light chain fibroin

Ten contigs representing unique allelic variants were assembled from the 537 L-fibroin clones (Fig. 4.2). Nine contigs (LCF-1 to LCF-9) differ from each other by one to five non-synonymous changes. Strikingly, six of the eleven non-synonymous mutations occur at sites that are evolving more rapidly than 80% of the codon sites (see below). The tenth contig, LCF-10, is distinct from the other nine alleles by a sizeable gap that spans 27 amino acids and is likely a deletion mutant. The L-fibroin gene has six highly conserved exon/intron boundaries in the caddisflies *H. angustipennis* and *R. obliterata* and also in the moth *B. mori*, except that in the last species the fifth exon is split into two (Kikuchi *et al.*, 1992; Yonemura *et al.*, 2009). If the *H. californicus* L-fibroin gene also has this conserved exon arrangement, then a deletion of 81 nucleotides from the central part of exon three would account for LCF-10. Despite the differences present among the ten *L-fibroin* alleles, large stretches of amino acids remain constant, suggesting functional importance of the specific sequence in these regions.

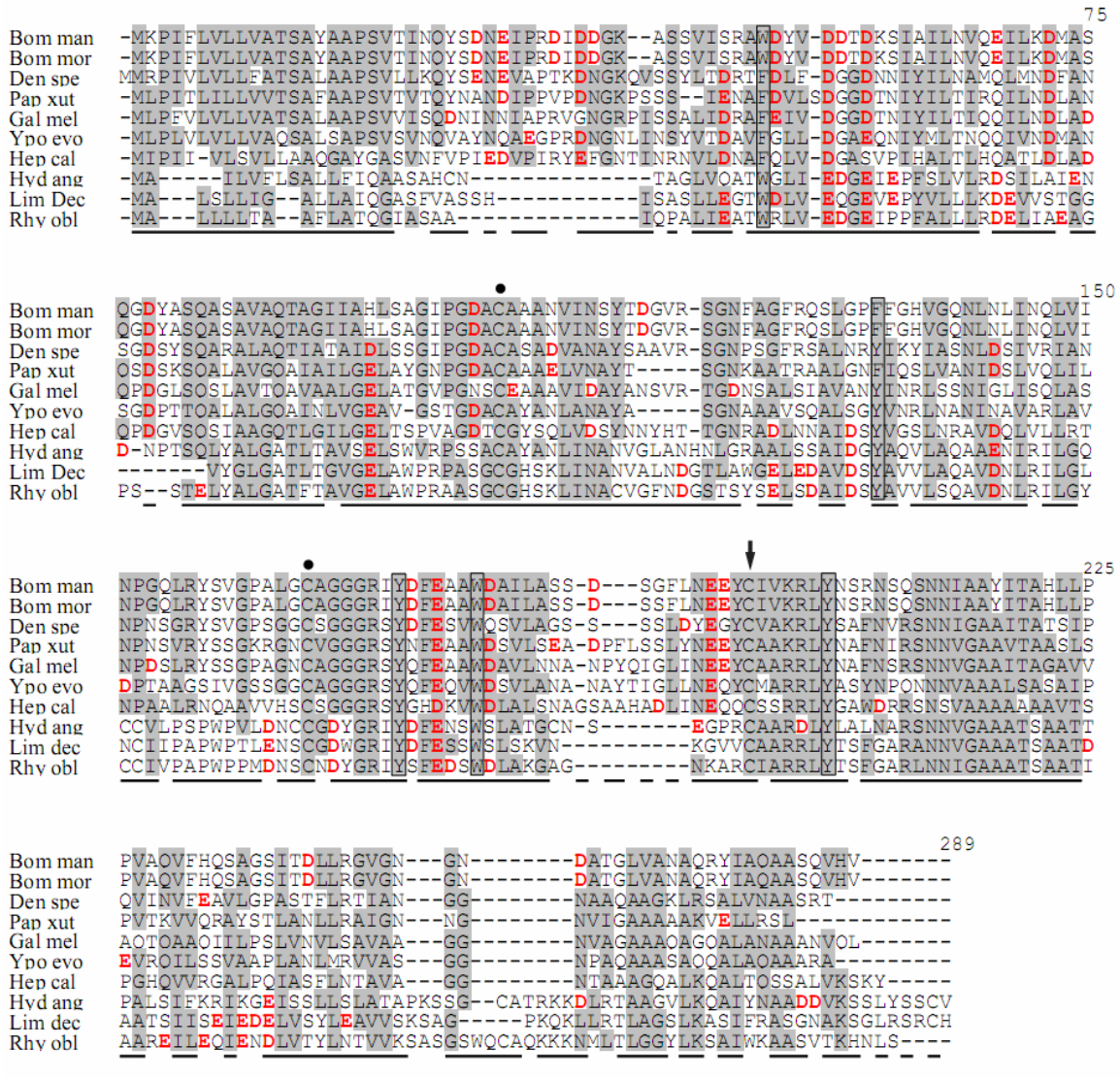


Fig. 4.4. Alignment of amphiesmenopteran light chain fibroins. Shading shows >50% conservation of chemical type as described in Fig. 4.3. Dots and arrow denote cysteines involved in bonding within L-fibroin and with H-fibroin, respectively. Underlines at the bottom of columns mark positions that have experienced purifying selection. Dashes are alignment gaps. Conserved aromatic residues are boxed and acidic residues are red and bold. Position number appears above each block. Taxon abbreviations and Genbank accessions: *Bombyx mandarina* (Bom man; AB001820), *Bombyx mori* (Bom mor; X17291), *Dendrolimus spectabilis* (Den spe; AB001822), *Papilio xuthus* (Pap xut; AB001824), *Galleria mellonella* (Gal mel; S77817), *Yponomeuta evonymellus* (Ypo evo; AB195977), *Hepialus californicus* (Hep cal; GU180666), *Hydropsyche angustipennis* (Hyd ang; AB214508), *Limnephilus decipiens* (Lim dec; AB214510) and *Rhyacophila obliterata* (Rhy obl; AB354590).

Comparison of the *H. californicus* L-fibroin to the other lepidopteran and trichopteran L-fibroins shows extensive conservation of amino acid sequence and biochemical properties (shading, Fig. 4.4). The conserved residues include aspartic acid and glutamic acid (red and bold, Fig. 4.4) that render the L-fibroin pI acidic. For example, the pI of *B. mori* L-fibroin is 5.06 and that of *H. californicus* is 6.13. Acidic pI values are hypothesized to be important for the conversion of silk proteins from a gel state to a liquid crystalline solution prior to spinning fibers (Wong Po Foo *et al.*, 2006). The negatively charged L-fibroin balances the charge of the carboxyl-terminal region of H-fibroin, which is often basic (see above).

Three cysteine residues in *H. californicus* L-fibroin are present in similar locations in all known L-fibroins of Amphiesmenoptera (Fig. 4.2 and 4.4). These conserved residues form the intra- and intermolecular bonds required for protein transport and fiber formation (Takei *et al.*, 1987; Tanaka *et al.*, 1999). Specifically, the first two cysteines (marked by dots, Fig. 4.2 and 4.4) are known in *B. mori* to form an intramolecular bond thought to be important for proper protein conformation (Tanaka *et al.* 1999). Also shown through studies of *B. mori*, the third conserved cysteine (denoted by an arrow, Fig. 4.2 and 4.4) forms a disulfide bond with the cysteine located in position -22 of H-fibroin (marked by an arrow, Fig. 4.3). Assembly of the L-fibroin-H-fibroin complex is essential for the transport of both components out of the endoplasmic reticulum and into the Golgi complex (Takei *et al.*, 1987).

In addition to the conserved cysteine residues, there are five sites with aromatic amino acids within all of the lepidopteran and trichopteran L-fibroins (boxed, Fig. 4.4).

There are also aromatic sites that while not conserved across Amphiesmenoptera, are consistent within Ditrysia (positions 114, 174) or Trichoptera (positions 84, 98, 158, 168 and 274). These aromatic amino acid sites may be important for forming aromatic-aromatic interactions that stabilize the configuration of L-fibroin. Burley and Petsko (1985) noted that many globular proteins contain amino acids with aromatic side chains. Using high resolution protein structures, they showed that aromatic side chains interact at dihedral angles between 50° and 90° and that these residues are separated by a mean of 38.6 residues (minimum of seven). Within L-fibroin, the mean interval between aromatic amino acid sites present across all Amphiesmenoptera is 42 and the Ditrysia-specific and Trichoptera-specific means are 60 and 45.8, respectively, with every individual interval much greater than seven. The stabilization resulting from these aromatic-aromatic interactions may enhance the micellar and globular conformations, which are required for the solubility of silk proteins within the glandular lumen prior to silk spinning (Jin and Kaplan, 2003).

Molecular evolution

Measuring the ratio of the non-synonymous substitution (replacement) rate versus the synonymous substitution (silent change) rate in coding sequences can provide insights into the evolution and function of a specific protein (Yang and Bielawski, 2000). The ratio of replacement substitutions (d_N) over silent substitutions (d_S) is ω. Fixation of non-synonymous substitutions may reflect functional adaptive change within a molecule. For example, Gatesy and Swanson (2007) analyzed the *ACR* gene, which encodes a fertilization protein within mammals. Within *ACR*, they detected 21 rapidly evolving

codon sites that may have adaptive significance. In contrast, conservation of amino acids over evolutionary time is thought to reflect a molecule that is undergoing purifying selection, indicating that change of amino acid residue is deleterious to the molecule's function.

If $\omega<1$ at a specific site, then that site may have undergone purifying selection (i.e. a slow rate of change or no change over time). If $\omega=1$ at a given site, then the fixation of changes has occurred at the neutral rate of evolution, and, finally, if $\omega>1$ at any site, then that site may have been subject to adaptive selection. Calculating ω at each site may thus provide insights into the function of a particular amino acid within a protein. For example, Sawyer *et al.* (2005) determined that adaptive evolution took place in the primate *TRIM5α*, a gene encoding a protein responsible for limiting retroviral infection. Through their analyses, they identified a 13 residue long stretch in the protein that is necessary for resisting retroviruses. Similarly, measuring the rate of selection on the silk fibroins may reveal additional areas of conservation or areas undergoing positive selection.

Given the importance of silk to the ecology of Trichoptera and Lepidoptera, we measured the rate of molecular evolution at each codon site for L-fibroin and last 26 amino acids of the carboxyl-terminal region of H-fibroin (Fig. 4.3). The repetitive regions were not analyzed because of the sequence divergence across taxa and the high variability of repeat motifs even within an individual fibroin molecule. For the carboxyl-terminal region, there was no significant difference between log likelihood scores of the M7 (neutral) and M8 (selection) models, but there was a significant difference (p value >

0.0001) in log likelihood values of M1a (nearly neutral; -713.81) and M2a (selection; -736.55). Based on these comparisons we accepted M1a, the nearly neutral model.

Although model M1a (nearly neutral) implies that no sites have rapidly evolved under positive selection, the ω values for individual sites also indicate that the carboxyl-terminal region has a history of purifying selection. All but two of the 26 sites had ω values less than 0.3, suggesting a slow rate of evolution within the carboxyl-terminal region and most were below 0.1 (underlined positions, Fig. 4.3). These very low values signal a slow rate of evolution within the H-fibroin carboxyl-terminal region, consistent with mutations in this region may being deleterious to silk solubility and fiber formation across Lepidoptera and Trichoptera.

We analyzed the rate of evolution at all 289 sites of the L-fibroin (Fig. 4.4). There was no significant difference between models M1a (-8393.39) vs. M2a (-8393.39) or models M7 (-8354.73) vs. M8 (-8354.20). Therefore, we chose M1a (nearly neutral model) as the best fit to the data. Most (80%) of the sites within L-fibroin are evolving had ω values < 0.5, indicating that much of the molecule has undergone purifying selection (underlined positions, Fig. 4.4). For example, the conserved aromatic and cysteine sites (Fig. 4.4) had ω values of 0.15 or lower. These very low ratios can be contrasted with the 20% of amino acid sites with ω values greater than 0.5 that are interspersed throughout the protein. These more labile sites may have lower functional significance to the fibroin. The overall slow rate of change over 250 million years within the coding region of L-fibroin highlights the critical role of this molecule in silk protein

transport and fiber formation, despite not constituting the bulk of the silk fiber (Yonemura *et al.*, 2009).

Phylogenetic signal

Amphiesmenoptera, the taxonomic grouping of Lepidoptera and Trichoptera, is a well established clade based on morphological and molecular data (Kristensen, 1999; Whiting, 2002). However, some higher-level groupings within Lepidoptera and Trichoptera have yet to be resolved. For example, phylogenetic relationships within the large lepidopteran subclades Ditrysia and Heteroneura are unclear (Wiegmann *et al.*, 2000; Wiegmann et al., 2002) and the position and monophyly of the trichopteran suborder Spicipalpia is uncertain (Morse, 1997; Kjer *et al.*, 2002). Given these issues, we test the phylogenetic utility of L-fibroin and the H-fibroin carboxyl-terminal region to resolve evolutionary relationships within Amphiesmenoptera.

We utilized the 26 codon carboxyl-terminal alignment plus the stop codon to construct a gene tree of the H-fibroin carboxyl-terminal region. The reconstructed H-fibroin gene tree does not include a monophyletic Lepidoptera or Trichoptera (Fig. 4.5). This is due to the placement of *A. yamamai* outside of Lepidoptera because of its divergent carboxyl-terminal sequence, which most likely reflects its unusual silk filament composition of only H-fibroin dimers (see above). If the *A. yamamai* sequence is removed from the data matrix, Trichoptera and Lepidoptera are each recovered as monophyletic, although there is no resolution of individual clades within Lepidoptera. The carboxyl-terminal alignment of H-fibroin does not provide many phylogenetically informative characters due to its short length and sparse taxon sampling.

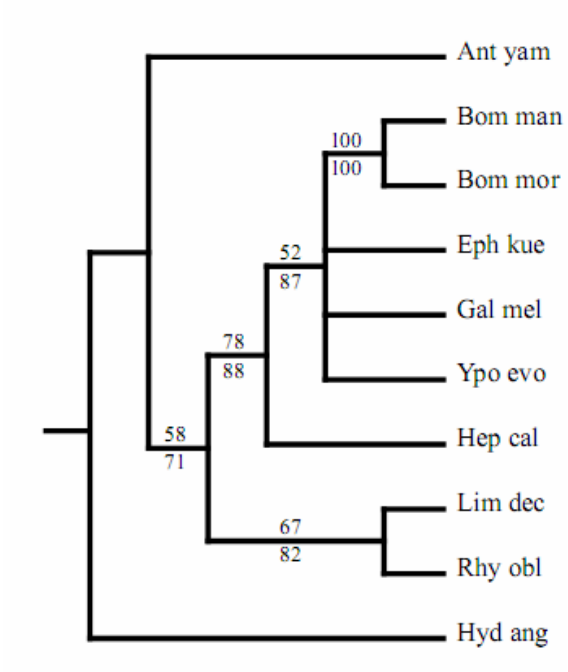


Fig. 4.5. 50% majority rule parsimony gene tree based on the carboxyl-terminal region of heavy chain fibroin. >50% bootstrap support values shown above (parsimony) and below (maximum likelihood) nodes. Taxon abbreviations and Genbank accessions as in Fig. 4.3.

L-fibroin was more promising. The L-fibroin gene tree was constructed from the full-length coding region transcripts. Including gaps, the total alignment length was 867 nucleotides. The gene tree (Fig. 4.6) recovers monophyletic Trichoptera, monophyletic Lepidoptera, and is concordant with previous lepidopteran molecular phylogenies (Regier *et al.* 2008). Both parsimony and maximum likelihood trees agree at all nodes with moderate to high bootstrap support values for a number of clades within Amphiesmenoptera. Several clades are resolved within Lepidoptera, including Bombycoidea (66% parsimony bootstrap), Macrolepidoptera (63% parsimony bootstrap), Obtectomera (91% parsimony bootstrap), and Ditrysia (100% parsimony bootstrap). In Trichoptera, *R. obliterata* groups with *L. decipiens*, which is consistent with previous

molecular studies (Kjer *et al.*, 2002). Given the size and numerous introns of the L-fibroin gene, the best method to obtain sequence data would be from mRNA, using methodology such as in Regier *et al.* (2008). Ideally, mRNA should be isolated from last instar larvae to ensure an abundance of *L-fibroin* transcripts. Based on the resolved gene tree and bootstrap support values, L-fibroin is a promising molecular marker for reconstructing the higher-level relationships within Amphiesmenoptera.

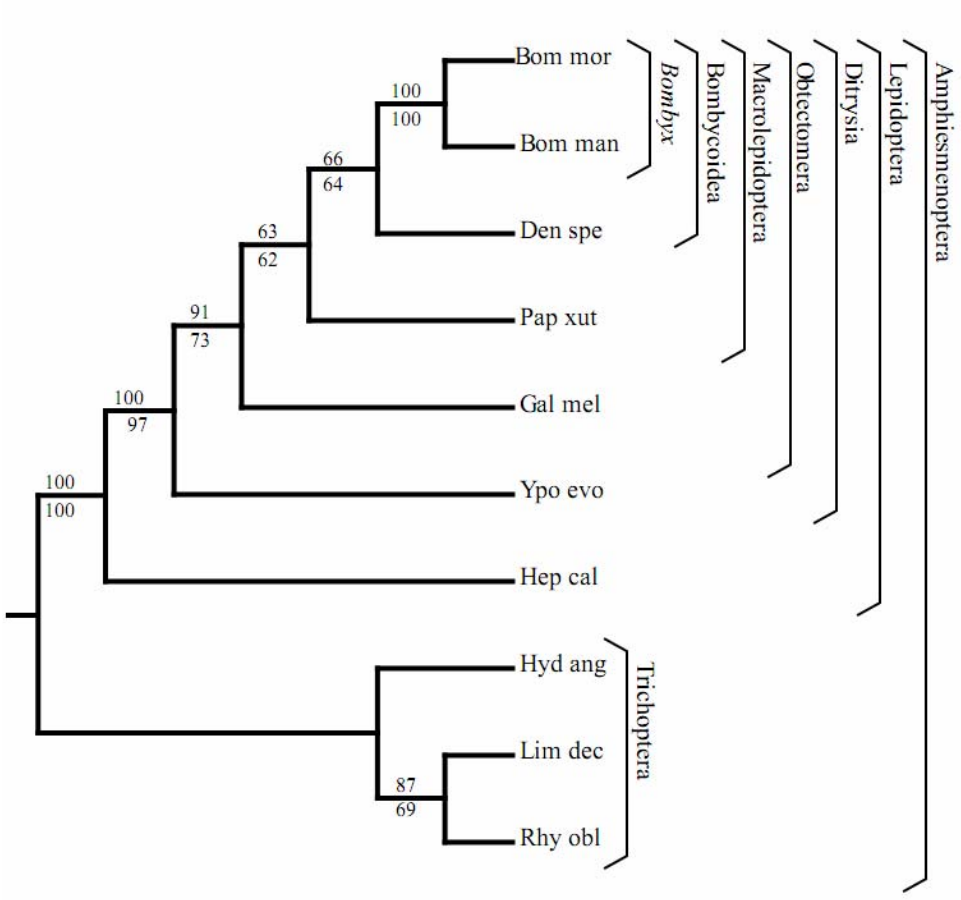


Fig 4.6. Maximum likelihood and parsimony gene tree based on light chain fibroin. >50% bootstrap support values shown above (parsimony) and below (maximum likelihood) nodes. Brackets indicate recognized clades (Weller and Pashley 1995; Wiegmann *et al.* 2000; Wiegmann *et al.* 2002). Taxon abbreviations and Genbank accessions as Fig. 4.4.

In summary, we characterized the L- and H-fibroins of the ghost moth, *H. californicus*. By involving H-fibroin and L-fibroin but apparently not P25, silk fiber formation in *H. californicus* is similar to that of Trichoptera (Yonemura et al. 2006; Yonemura et al. 2009), despite major differences in the spinning environment and ecological function of subterranean ghost moth silk versus aquatic caddisfly silk. The *H. californicus* L-fibroin had numerous attributes, such as an acidic pI and homologous placement of cysteine and aromatic residues that were maintained via purifying selection since the common ancestor of Lepidoptera and Trichoptera (Fig. 4.4). Similarly, the carboxyl-terminal region of the *H. californicus* H-fibroin contained several conserved amino acids, such as a cysteine located 22 residues upstream of the protein end (Fig. 4.3). The slow pace of substitutions suggests that mutations within L-fibroin and the H-fibroin carboxyl-terminal region may are deleterious to silk fiber formation. In contrast, the repetitive region of H-fibroin is divergent across taxa and prone to rearrangements. Despite its more rapid rate of evolution, the repetitive region has constraints on its sequence. Like the silk proteins of other insects and spiders, the *H. californicus* H-fibroin is internally repetitive (tandem arrayed ensemble repeats) and dominated by glycine, alanine, and serine (Fig. 4.1; Gatesy *et al.*, 2001).

While they had many evolutionarily conserved characteristics, the ghost moth silk proteins exhibited key differences from other amphiesmenopteran silk proteins. Specifically, the *H. californicus* H-fibroin carboxyl-terminal region lacked the two cysteines present in most other lepidopteran H-fibroins at positions -1 and -4 (Fig. 4.3). These missing cysteines in conjunction with lack of evidence for P25, indicate that the

ghost moth has an alternative method of fiber formation compared to more derived lepidopterans, as typified by *Bombyx mori* (Inoue et al., 2000). Thus, by filling in the substantial phylogenetic gap between caddisflies and more derived moths, *H. californicus* provides insight into the early evolution of silk in Lepidoptera.

References

Ashida, J., Ohgo, K., Komatsu, K., Kubota, A., Asakura, T., 2003. Determination of the torsion angles of alanine and glycine residues of model compounds of spider silk (AGG)(10) using solid-state NMR methods. Journal of Biomolecular NMR. 25, 91-103.

Bini, E., Knight, D.P., Kaplan, D.L., 2004. Mapping domain structures in silks from insects and spiders related to protein assembly. Journal of Molecular Biology. 335, 27-40.

Burley, S.K., Petsko, G.A., 1985. Aromatic-aromatic interaction a mechanism of protein-structure stabilization. Science. 229, 23-28.

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

Felsenstein, J., 1985. Confidence-limits on phylogenies an approach using the bootstrap. Evolution. 39, 783-791.

Friedlander, T.P., Regier, J.C., Mitter, C., Wagner, D.L., 1996. A nuclear gene for higher level phylogenetics: phosphoenolpyruvate carboxykinase tracks Mesozoic-age divergences within Lepidoptera (Insecta). Molecular Biology Evolution. 13, 594-604.

Garb, J.E., Hayashi, C.Y., 2005. Modular evolution of egg case silk genes across orb-weaving spider superfamilies. Proceedings of the National Academy of Sciences of the United States of America. 102, 11379-11384.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In: Walker, J. M. (Ed.) The Proteomics Protocols Handbook. Humana Press

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

Gatesy, J., Swanson, W.J., 2007. Adaptive evolution and phylogenetic utility of *Acr* (acrosin), a rapidly evolving mammalian fertilization gene. Journal of Mammalogy. 88, 32-42.

Gosline, J.M., Demont, M.E., Denny, M.W., 1986. The structure and properties of spider silk. Endeavour. 10, 37-43.

Green, P., 1994 Phrap
http://www.genome.washington.edu/UWGC/analysistools/phrap.htm.

Grimaldi, D.A., Engel, M.S., 2005. Evolution of the insects. Cambridge, UK. Cambridge University Press

Guerette, P.A., Ginzinger, D.G., Weber, B.H.F., Gosline, J.M., 1996. Silk properties determined by gland-specific expression of a spider fibroin gene family. Science. 272, 112-115.

Hayashi, C.Y., Lewis, R.V., 2001. Spider flagelliform silk: lessons in protein design, gene structure, and molecular evolution. Bioessays. 23, 750-756.

Hayashi, C.Y., Shipley, N.H., Lewis, R.V., 1999. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. International Journal of Biological Macromolecules. 24, 271-275.

Hwang, J-S., Lee, J-S., Goo, T-W., Yun, E-Y., Lee, K-S., Kim, Y-S., Jin, B-R., Lee, S-M., Kim, K-Y., Kang, S-W., Suh, D-S., 2001. Cloning of the fibroin gene from the oak silkworm, *Antheraea yamamai* and its complete sequence. Biotechnology Letters. 23, 1321-1326.

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, 40517-40528.

Inoue, S., Tanaka, K., Tanaka, H., Ohtomo, K., Kanda, T., Imamura, M., Quan, G-X., Kojima, K., Yamashita, T., Nakajima, T., Taira, H., Tamura, T., Mizuno, S., 2004. Assembly of the silk fibroin elementary unit in the endoplasmic reticulum and a role of L-chain for protection of α1, 2-mannose residues in N-linked oligosaccharide chains of fibrohexamerin/P25. European Journal of Biochemistry. 271, 356-366.

Ittah, S., Cohen, S., Garty, S., Cohn, D., Gat. U., 2006. An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. Biomacromolecules. 7, 1790-1795.

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

Kikuchi, Y., Mori, K., Suzuki, S., Yamaguchi, K., Mizuno, S., 1992. Structure of the *Bombyx mori* fibroin light-chain-encoding gene upstream sequence elements common to the light and heavy-chain. Gene. 110, 151-158.

Kjer, K.M., Blahnik, R.J., Holzenthal, R.W., 2002. Phylogeny of caddisflies (Insecta, Trichoptera). Zoologica Scripta. 31, 83-91.

Kristensen, N.P., 1999. Phylogeny of endopterygote insects, the most successful lineage of living organisms. European Journal of Entomology. 96, 237-253.

Mita, K., Ichimura, S., James, T.C., 1994. Highly repetitive structure and its organization of the silk fibroin gene. Journal of Molecular Evolution. 38, 583-592.

Mori, K., Tanaka, K., Kikuchi, Y., Waga, M., Waga, S., Mizuno, S., 1995. Production of a chimeric fibroin light-chain polypeptide in a fibroin secretion deficient naked pupa mutant of the silkworm *Bombyx mori*. Journal of Molecular Biology. 251, 217-228.

Morse, J.C., 1997. Phylogeny of Trichoptera. Annual Review of Entomology. 42, 427-450.

Nielsen, E.S., Robinson, G.S., Wagner, D.L., 2000. Ghost-moths of the world: a global inventory and bibliography of the Exoporia (Mnesarchaeoidea and Hepialoidea) (Lepidoptera). Journal of Natural History. 34, 822-878.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics. 14, 817-818.

Regier, J.C., Cook, C.P., Mitter, C., Hussey, A., 2008, A phylogenetic study of the 'bombycoid complex' (Lepidoptera) using five protein-coding nuclear genes, with comments on the problem of macrolepidopteran phylogeny. Systematic Entomology. 33, 175-189.

Sawyer, S.L., Wu, L.I., Emerman, M., Malik, H.S., 2005. Positive selection of primate TRIM5 alpha identifies a critical species-specific retroviral restriction domain. Proceedings of the National Academy of Sciences of the United States of America. 102, 2832-2837.

Sehnal, F., Žurovec, M., 2004. Construction of silk fiber core in Lepidoptera. Biomacromolecules. 5, 666-674.

Sezutsu, H., Yukuhiro, K., 2000. Dynamic rearrangement within the *Antheraea pernyi* silk fibroin gene is associated with four types of repetitive units. Journal of Molecular Evolution. 51, 329-338.

Strong, D.R., Kaya, H.K., Whipple, A.V., Child, A.L., Kraig, S., Bondonno, M., Dyer, K., Maron, J.L., 1996. Entomopathogenic nematodes: Natural enemies of root-feeding caterpillars on bush lupine. Oecologia. 108, 167-173.

Swofford, D.L., 1998. PAUP*: phylogenetic analysis using parsimony (* and other models). Sinauer Associates. Inc., Publishers, Sunderland, Massachusetts.

Takei, F., Kikuchi, Y., Kikuchi, A., Mizuno, S., Shimura, K., 1987. Further evidence for importance of the subunit combination of silk fibroin in its efficient secretion from the posterior silk gland-cells. Journal of Cell Biology. 105, 175-180.

Tamura, T., Inoue, H., Suzuki, Y., 1987. The fibroin genes of *Antheraea yamamai* and *Bombyx mori* are different in their core regions but reveal a striking sequence similarity in their 5' ends and 5' flanking regions. Molecular and General Genetics. 207, 189-195.

Tanaka, K., Kajiyama, N., Ishikura, K., Waga, S., Kikuchi, A., Ohtomo, K., Takagi, T., Mizuno, S., 1999. Determination of the site of disulfide linkage between heavy and light chains of the silk produced by *Bombyx mori*. Biochimica et Biophysica Acta. 1432, 92-103.

Tanaka, K., Mizuno, S., 2001. Homologues of fibroin L-chain and P25 of *Bombyx mori* are present in *Dendrolimus spectabilis* and *Papilio xuthus* but not detectable in *Antheraea yamamai*. Insect Biochemistry and Molecular Biology. 31, 665-677.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal W improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research. 22, 4673-4680.

Ueda, H., Mizuno, S., Shimura, K., 1985. Sequence polymorphisms around the 5'-end of the silkworm fibroin H-chain gene suggesting the occurrence of crossing-over between heteromorphic alleles. Gene. 34, 351-355.

Weller, S.J., Pashley, D.P., 1995. In search of butterfly origins. Molecular Phylogenetics and Evolution. 4, 235-246.

Whiting, M.F., 2002. Phylogeny of the holometabolous insect orders: molecular evidence. Zoologica Scripta. 31, 3-15.

Wiegmann, B.M., Mitter, C., Regier, J.C., Friedlander, T.P., Wagner, D.M., Nielsen, E.S., 2000. Nuclear genes resolve Mesozoic-aged divergences in the insect order Lepidoptera. Molecular Phylogenetics and Evolution. 15, 242-259.

Wiegmann, B.M., Regier, J.C., Mitter, C., 2002. Combined molecular and morphological evidence on the phylogeny of the earliest lepidopteran lineages. Zoologica Scripta. 31, 67-81

Wong Po Foo, C., Bini, E., Hensman, J., Knight, D.P., Lewis, R.V., Kaplan, D.L. 2006. Role of pH and charge on silk protein assembly in insects and spiders. Applied Physics A-Materials Science & Processing. A82, 223-233.

Yang, Z., 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 24, 1586-1591.

Yang, Z., Bielawski, J.P., 2000. Statistical methods for detecting molecular selection. Trends in Ecology and Evolution. 15, 496-503.

Yang, Z., Wong, W.S.W., Nielson, R., 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Molecular Biology and Evolution. 22, 1107-1118.

Yonemura, N., Mita, K., Tamura, T., Sehnal, F., 2009. Conservation of silk genes in Trichoptera and Lepidoptera. Journal Molecular Evolution. 68, 641-653.

Yonemura, N., Sehnal, F., 2006. The design of silk fiber composition in moths has been conserved for more than 150 million years. Journal of Molecular Evolution. 63, 42-53.

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

Žurovec, M., Sehnal, F., 2002. Unique molecular architecture of silk fibroin in the waxmoth, *Galleria mellonella*. Journal of Biological Chemistry. 277, 22639-22647.

Summary

Overview

To better understand the evolution of arthropod silks, I chose to study silks for which virtually nothing was previously known. My first objective was the characterization of gene transcripts, secondary structural profiles, and mechanical properties of silks produced by embiopteran species. Toward this end, I tensile tested the silk of six embiid species, determined secondary structural profiles for the silks of four species, and cloned cDNAs for the silk proteins (fibroins) of five species. My second aim was to describe the fibroin transcripts from a ghost moth, then compare these new sequences to the silk molecules of other moths and caddisflies. I discovered that the ghost moth has transcripts for a subset of the fibroins found in other moths. Specifically, I characterized heavy and light chain fibroins and used a variety of comparative analyses to infer how these fibroins have been shaped by natural selection.

Embioptera

I found that across species, webspinner insect silks are remarkably similar in tensile strength and percentage of β-sheet structure. For instance, mean tensile strengths varied by only 30 MPa across taxa (range 130-160) and β-sheets composed over 45% of the secondary structural profile for each silk. These similarities are consistent with the common usage of silk by these species for gallery construction and egg swaddling. Nevertheless, there were statistically significant differences in stiffness and extensibility. These differences corresponded well with the secondary structures that I quantified. The most extensible silks, produced by *Archembia* and *Saussurembia*, have a greater percentage of β-turns than that of the stiffest silk, which is spun by *Haploembia*.

Variation in stiffness and extensibility may result from species-specific microhabitat preferences, such as living in leaf litter versus on tree surfaces.

Embiopteran silk fibroins have several conserved elements that provide insight into silk fiber formation and the evolution of these adaptive molecules. Like many arthropod fibroins, embiopteran fibroins are primarily composed of glycine, serine, and alanine. However, webspinner fibroins are distinctive in containing much larger percentages of serine (over 30%). Serine, with its polar hydroxyl group, may enhance fiber formation by aiding in water solubility and increasing the extent of hydrogen bonding. Furthermore, the carboxyl-terminal region has a low pH, which suggests that fibroin aggregation within the glandular lumen occurs through an acidification process, as has been described in other silk spinning systems. Closer inspection of the embiid fibroin carboxyl-terminal regions reveals that many amino acid residues are conserved between distantly related species, evidence that this domain may have an important role. Selection may have maintained these sequences to ensure structural stability, positioning of cysteine residues, and maintaining molecular charges for fiber formation and polymer cross-linking through disulfide bonds.

My dissertation chapters on embiopteran silks begin the description of these previously uncharacterized silks. Further research on embiopteran fibroins could focus on clarifying silk production methods, genetic structure, and ecological differences that may influence silk properties. For example, analysis of pH along the length of the secretion duct to the ejector tip would help determine if acidification causes aggregation of fibroins. Quantification of the transcript lengths and/or molecular weights for webspinner

fibroins could corroborate the hypothesis that small protein monomer size (and thus a limited number of intermolecular bonds) results in the modest tensile strength observed for embiid silks. Furthermore, detailed ecological studies on several embiopteran species may reveal subtle silk usage differences that influence silk tensile properties. These dissertation chapters will hopefully inspire more research on these often overlooked insects.

Lepidoptera

Two fibroins were identified in the *Hepialus californicus* (ghost moth) silk gland cDNA library: light chain fibroin and heavy chain fibroin. Molecular evolutionary analyses indicated that light chain fibroin and the carboxyl terminal region of heavy chain fibroin have experienced purifying (negative) selection. This finding suggests that there are functional constraints on these sequences, most likely due to their importance in fiber formation. Additionally, light chain fibroin is shown to be a promising phylogenetic marker for Lepidoptera. For most key elements, such as the placement of cysteine and aromatic residues in light chain fibroin and the pH of both light chain and heavy chain fibroin, *Hepialus* fibroins are similar to their counterparts in divergent lepidopteran species. The *Hepialus* silk fiber, however, apparently lacks the silk glycoprotein P25, suggesting that *Hepialus* fiber formation occurs differently than in otherLepidotera. Indeed, from the composition of *Hepialus* silk, it can be inferred that the groundplan for silk fibers in Lepidoptera more closely resembles the core fiber design identified in Trichoptera rather than the specialized ditrysian moths.

My work on *Hepialus* presents the first characterization of silk fibroins from a non-ditrysian Lepidoptera. Future research on *Hepialus* silk fibroins could focus on tensile testing silk fibers, structural characterization of light chain fibroin, and genomic screening for a P25 homolog. Tensile testing *Hepialus* silk may help to validate predictions of high stiffness and tensile strength based on fibroin amino acid sequence and how the caterpillars use silk. The amino acid sequence of *Hepialus* light chain fibroin shares numerous conserved amino acids with other amphiesmenopteran light chain fibroins. While the function of the conserved cysteine residues has been previously identified, I suggest the first hypotheses for the role of the conserved aromatic residues. Structural characterization may confirm the stabilizing capacity of these conserved aromatic amino acids.

Broader Impacts

Silk production has arisen numerous times in Arthropoda. Studying convergently evolved silks yields insight into the multiple solutions for manufacturing and assembling fibers from proteinaceous solutions. For example, lepidopteran and embiopteran silks differ in their amino acid composition, body part in which they are produced, and type of glands used to manufacture silk. Yet, both groups form silk fibers that are utilized for protective structures. Identifying analogous features, such as a low pH and conserved cysteine residues, reveals necessary traits for fiber formation in both groups. Divergent features, such as high serine content in embiopteran fibroins or multiple types of silk fibroins in Lepidoptera, provide clues about evolutionary adaptation and functional constraints for silk production within each lineage. Understanding convergent and

divergent elements, as well as the extent of allowable variation, provides inspiration to engineers, chemists, and biochemists for the synthesis of silk-like materials.

Besides contributing to the development of biomimetic materials, my dissertation focuses on understudied taxa. When I started the project, nothing was known about the silks produced by Embioptera and *Hepialus* and very few (less than a dozen) ecological and/or taxonomic studies had been performed on any of my target species. For these fairly common but reclusive taxa, my research is among the first investigations of their silks. Working on these organisms has been difficult due to the lack of background knowledge that is available for better-studied species. For example, there are no genome or expression databases that I could draw on, no published protocols for collecting and characterizing their silk fibers, and little information regarding silk use. Despite these challenges, I find it immensely rewarding to have contributed to fields such as silk biology, entomology, biomechanics, and molecular evolution. I hope that my research fosters more studies that expand the ecological and genetic knowledge of these little studied, but fascinating, insects.