© 2015. Published by The Company of Biologists Ltd | Journal of Experimental Biology (2015) 218, 3138-3149 doi:10.1242/jeb.128306

RESEARCH ARTICLE

Identification and classification of silks using infrared spectroscopy

Maxime Boulet-Audet[1,2,*], Fritz Vollrath[2] and Chris Holland[3]

ABSTRACT

Lepidopteran silks number in the thousands and display a vast diversity of structures, properties and industrial potential. To map this remarkable biochemical diversity, we present an identification and screening method based on the infrared spectra of native silk feedstock and cocoons. Multivariate analysis of over 1214 infrared spectra obtained from 35 species allowed us to group silks into distinct hierarchies and a classification that agrees well with current phylogenetic data and taxonomies. This approach also provides information on the relative content of sericin, calcium oxalate, phenolic compounds, poly-alanine and poly(alanine-glycine) β-sheets. It emerged that the domesticated mulberry silkmoth *Bombyx mori* represents an outlier compared with other silkmoth taxa in terms of spectral properties. Interestingly, *Epiphora bauhiniae* was found to contain the highest amount of β-sheets reported to date for any wild silkmoth. We conclude that our approach provides a new route to determine cocoon chemical composition and in turn a novel, biological as well as material, classification of silks.

KEY WORDS: Silkworm, Cocoon, Lepidopteran, Multivariate analysis, Phylogenetic

INTRODUCTION

Silkworm silk is a high-value agricultural product offering sustainable harvesting that directly contributes to poverty alleviation in rural communities (Astudillo et al., 2014; Dooley, 2004). Yet, it also has growing technical applications (Borkner et al., 2014; Omenetto and Kaplan, 2010). Developments in mulberry sericulture and the increasing use of fibres from 'wild' silkworms provide the backdrop for increased interest in understanding the diversity of all silks. Not surprisingly, millions of years of divergent evolution have resulted in a rich biodiversity of silks (Scoble, 1999). Typically used in cocoons, this class of materials consists of a silk fibroin protein thread of up to 1 km long coated with sericin proteins acting as a resin/matrix glue (Chen et al., 2010b). This non-woven composite structure (Chen et al., 2010a) regulates gas flow and humidity (Danks, 2004; Horrocks et al., 2013; Roy et al., 2012), as well as protecting the encased pupae from predation (Ishii et al., 1984), micro-organisms (Franceschi and Nakata, 2005) and the environment (Chen et al., 2012b). Silkworms produce cocoons with a broad variety of morphologies and architectures, ranging in porosity from loose meshes to full shells, with or without an exit opening (Chen et al., 2012c). Cocoons may also incorporate extraneous materials as well, such as integrated leaves for camouflage or an internally applied calcium oxalate solution that hardens the cocoon and may impart toxicity (Arnott and Webb, 2000; Chen et al., 2012c; Franceschi and Nakata, 2005; Gheysens et al., 2011; Takahashi et al., 1969; Teigler and Arnott, 1972b). The diversity of lepidopteran silk materials includes a molecular dimension, with amino acid analysis showing widely varying chemical compositions of silkworm silk (Hwang et al., 2001). However, while more advanced biochemical methods can inform on protein size (Hwang et al., 2001; Inoue et al., 2000; Mita et al., 1994), amino acid residue patterns (Navarro et al., 2008) and propensity to fold (Dicko et al., 2008), they are often labour intensive and expensive. Hence, only a handful of fibroin proteins have been sequenced to date (Tanaka and Mizuno, 2001). Furthermore, these methods are often focused on one specific molecular component of the cocoon and are unable to account for the other compounds present.

An alternative approach to achieve a broader assessment of chemical diversity is to employ complementary spectroscopic and scattering techniques (Gheysens et al., 2011; Warwicker, 1954). For example, the use of attenuated total reflection infrared spectroscopy (ATR-IR) is particularly well suited to studying silks in all forms as it is capable of measuring rough and deformable solids (Chen et al., 2012a; Gheysens et al., 2011), as well as turbid and concentrated protein solutions (Boulet-Audet et al., 2011). Requiring only minimal sample preparation, ATR-IR can selectively probe the inside or outside surface of silk cocoons, providing information on the local chemical composition (Boulet-Audet et al., 2014; Chen et al., 2012a, 2007). This spectroscopic method can determine (i) the level of protein crystallinity (Boulet-Audet et al., 2008), (ii) secondary structure (Goormaghtigh et al., 2006), and (iii) specific protein components such as sericin (Barth, 2000; Teramoto and Miyazawa, 2003). Infrared spectra are also indicative of (iv) non-protein molecules present in silk, such as the amount of water (Boulet-Audet et al., 2011), calcium oxalate (Chen et al., 2012b; Gheysens et al., 2011) and carbohydrate (Lu et al., 2011; Schulz and Baranska, 2007). In addition, the multivariate analysis of infrared spectra can (v) discriminate and classify samples based on their degree of relatedness. This infrared-based classification approach can even discriminate bacterial species (Kansiz et al., 1999; Preisner et al., 2007), types of human hairs (Panayiotou and Kokot, 1999), and coffee bean varieties (Briandet et al., 1996), as well as providing information for the construction of taxonomic trees (Zhao et al., 2006).

In this study, we analysed unspun native silk feedstock from six species across the Saturniini and Attacini tribes, and spun silks from 35 species across the Lepidoptera and Arachnida. Multivariate and hierarchical clustering analysis performed on over 1000 individual spectra allowed us to build taxonomic trees and compare them with trees based on protein-coding nuclear genes (Chen et al., 2012c;

[1]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK. [2]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. [3]Department of Materials Science and Engineering, University of Sheffield, Sheffield S1 3JD, UK.

*Author for correspondence (m.boulet-audet@imperial.ac.uk)

Received 16 July 2015; Accepted 6 August 2015

Journal of Experimental Biology

Journal of Experimental Biology (2015) 218, 3138-3149 doi:10.1242/jeb.128306

Regier et al., 2005, 2008a,b, 1998, 2002). As we demonstrate below, we identified several interesting outlier species that produce silk with very different chemical compositions and provide a hypothesis as to their origin. These newly characterised silks could even have greater potential for use in industrial and biomedical applications than those currently employed today.

RESULTS

Native feedstock spectral features

To evaluate the chemical variability of unspun silk without exogenous material, we used infrared spectroscopy to compare the native feedstocks of key species from the Lepidoptera genera *Actias*, *Attacus*, *Bombyx* and *Saturnia*, with the spider *Nephila edulis* as the outgroup. Bombycidae feedstocks such as *Bombyx mori* silk comprise heavy and light chain fibroins as well as P25 linked with disulphide bonds (Chevillard et al., 1986; Mita et al., 1994) in a 6:6:1 ratio (Inoue et al., 2000). In contrast, feedstocks of the Saturniidae such as *Antheraea yamamai*, *Actias luna*, *Attacus atlas* and *Saturnia pavonia* (Tanaka and Mizuno, 2001) comprise a homodimer (double heavy chain, H–H) protein mixture. As arthropods, spiders are very distantly related to the silkworms, yet by all accounts evolved silk production independently around 400 million years ago (Craig, 1997). Yet, the similar flow properties of their feedstocks thus represent an excellent example of convergent evolution (Craig, 1997; Holland et al., 2006).

Fig. 1 illustrates the distinctive features of native silk feedstock infrared spectra between 900 and 1500 cm^{-1} for silk from a variety of species. Table 1 indexes band assignments. Peaks between 1340 and 1456 cm^{-1} are commonly assigned to the vibration mode of residues (Barth, 2000). The strong 1383 cm^{-1} band associated with CH_2 bending for wild silks (the top four curves in Fig. 1) suggests a higher proportion of long-chain residues in feedstocks from *B. mori* and *N. edulis* major ampullate silk glands. Another important distinction for wild silk feedstocks is the presence of the well-resolved 1308 cm^{-1} peak in the amide III region. Monitored by Rheo-IR (Boulet-Audet et al., 2014), this band vanishes under shear-induced denaturation (see supplementary material Fig. S1) and is absent from cocoon spectra. We have assigned this component to β-turns that are precursors to β-sheets formed after spinning (Bandekar and Krimm, 1980; Cai and Singh, 2004; Rousseau et al., 2006). The arginine–glycine–aspartic acid (RGD) residue pattern (Sukopp et al., 2002) is believed to procure a greater fibroblast proliferation rate to the wild silkworm *Antheraea mylitta* compared with domesticated *B. mori* silk (Minoura et al., 1995; Navarro et al., 2008). Hence, we hypothesized that RDG patterns might contribute to the wild silk-specific vibration mode at 1308 cm^{-1}. The amide III shoulder at 1270 cm^{-1} results from α-helices (Cai and Singh, 2004; Krimm and Bandekar, 1986), and also appears stronger in wild silkworm silk feedstock. The neighbouring peak at 1245 cm^{-1} is commonly assigned to random coil secondary structures (Cai and Singh, 2004; Shao et al., 2005; Taddei and Monti, 2005; Yoshimizu and Asakura, 1990), and is strongest in the non-wild silks of *B. mori* and *N. edulis*. While present for all silk feedstocks, the peak at 1165 cm^{-1}, associated with the stretching of the N–C_α bond, is clearly broader for non-wild species, suggesting a wider distribution of conformations.

Actias luna is the only species probed that shows a well-resolved peak at 1144 cm^{-1}. We speculatively assigned this distinct band to the C–O stretching of sericin-like components used as a binding resin/matrix, as this species produces a cocoon with low porosity and high density (Chen et al., 2012c). While this study focused on unprocessed silk, extracting the sericin for further analysis could help to clarify this speculative assignment.

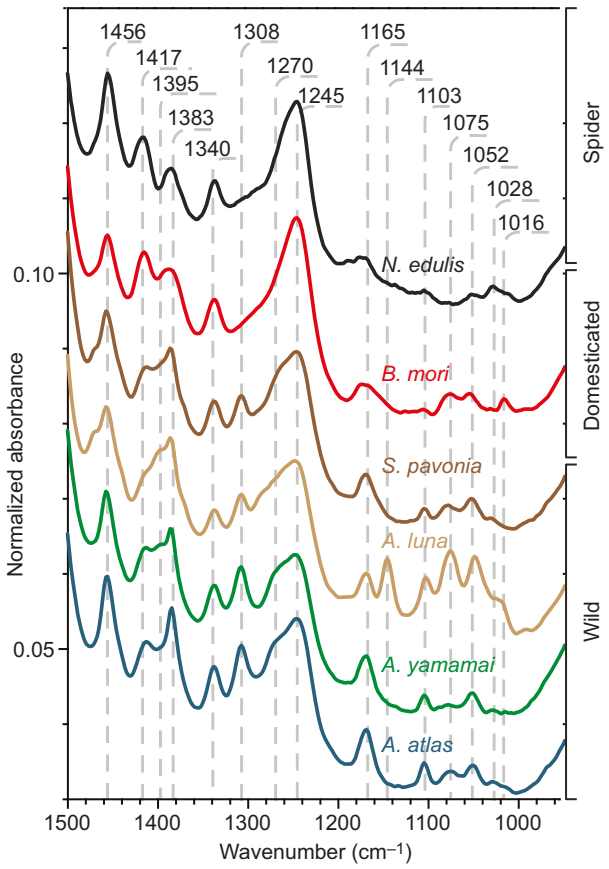

Fig. 1. Infrared spectra of unspun native silk feedstock. Data are for domesticated silkworm silk (*Bombyx mori*), wild silkworm silk (*Attacus atlas*, *Antheraea yamamai*, *Actias luna*, *Saturnia pavonia*) and spider silk feedstock (*Nephila edulis* major ampullate). The 1700–1500 cm^{-1} region is not shown as little difference between species was observed. Infrared spectra were collected from feedstocks extracted directly from the animal and kept at a native concentration (~22% dry weight).

The 1103 cm^{-1} band appeared on all spectra collected, although it was stronger in wild silkworm feedstocks. In the skeletal vibration region, this peak is likely to be caused by the C–C stretching of tyrosine aromatic rings, tryptophan or phenolic compounds (Andrus, 2006; Barth, 2000; Taddei and Monti, 2005). The adjacent component at 1075 cm^{-1} is present in all silk feedstocks (except sericin-free spider silk) and is also observed in pure sericin spectra, but is strongest in *A. luna*, thus reinforcing our previous assignment of the 1144 cm^{-1} peak for this species (Anghileri et al., 2007; Barth, 2000; Gupta et al., 1997). We also assigned the band at 1052 cm^{-1} to sericin C–O stretching, which is well resolved in most silkworm silks (Gupta et al., 1997; Taddei and Monti, 2005; Teramoto and Miyazawa, 2003).

Cocoon silk spectral features

Our findings (above) show that silk feedstocks have clear spectral differences between species. Therefore, we must assume that the cocoons produced from these feedstocks would also show variability. Moreover, we would also expect this diversity to increase as construction introduces variables such as the larva's spinning behaviour, other silkworm secretions such as faeces, and exogenous materials such as tannins diffusing from leaves. Previous work has shown that the properties of silk cocoons vary between the innermost and outermost layers (Chen et al., 2012a). Thus, to examine these sources of chemical diversity, we compared the infrared spectra of the

Journal of Experimental Biology (2015) 218, 3138-3149 doi:10.1242/jeb.128306

Table 1. Assignment of the main bands present in silk between 1700 and 1315 cm^{-1}

Position (cm^{-1})	Assignment	Observation	References
1733	υ(C=O)O	Tanned cocoon silks **O. eucalypti* **A. edwardsii*	Silverstein et al., 1981
1699	Amide I, β-sheets/β-turns	All spun silks **E. bauhiniae*	Bandekar and Krimm, 1979; Garside et al., 2005; Miyazawa and Blout, 1961; Moore and Krimm, 1976; Teramoto and Miyazawa, 2005
1642	Amide I, unordered	All silks	Boulet-Audet et al., 2008; Goormaghtigh et al., 2006; Jeong et al., 2006; Venyaminov and Kalnin, 1990
1620	Amide I, β-sheets	All spun silks **E. bauhiniae*	Boulet-Audet et al., 2008; Moore and Krimm, 1976; Sonoyama and Nakano, 2000
1547	Amide II, unordered	All silks	Boulet-Audet et al., 2008; Goormaghtigh et al., 2006; Jeong et al., 2006; Venyaminov and Kalnin, 1990
1516	Tyr–OH	All silks **A. yamamai*	Chirgadze et al., 1975
1508	Amide II, β-sheets	All spun silks **E. bauhiniae* **A. panda*	Boulet-Audet et al., 2008; Moore and Krimm, 1976; Sonoyama and Nakano, 2000
1456	δ_{as} (CH_3), Ala, Val	All unspun silks **N. clavipes* major	Barth, 2000; Boulet-Audet et al., 2008; Colthup, 1964
1443	δ_{as} (CH_3), β-sheets, $(AG)_n$, $(A)_n$	All spun silks **N. clavipes* **A. panda*	Barth, 2000; Moore and Krimm, 1976
1417	δ_s (CH_3), Ala, Val	All unspun silks **N. clavipes* major	Barth, 2000; Moore and Krimm, 1976
1403	δ_s (CH_3), Ala, Val	All spun silks **E. bauhiniae*	Barth, 2000; Moore and Krimm, 1976
1395	δ(CH_2, OH), Ser	Silkworm silk *Outermost layer	Anghileri et al., 2007; Barth, 2000; Teramoto and Miyazawa, 2003
1383	δ(CH_2), $(AG)_n$	All unspun silks **A. attacus* **S. pavonia*	Barth, 2000; Moore and Krimm, 1976
1370	δ(CH_2), $(AG)_n$	All spun silks **E. bauhiniae*	Barth, 2000; Moore and Krimm, 1976
1340	δ(CH) or *w*(CH_2)	All unspun silks	Barth, 2000; Colthup, 1964
1315	υ_s (OCO^-) Calcium oxalate	*G. postica* outer cocoon *Antheraea* outer cocoon	Chen et al., 2012c; Gheysens et al., 2011; Sargut et al., 2010; Silverstein et al., 1981

δ, bending; υ, stretching; *w*, wagging; $(A)_n$, polyalanine; $(AG)_n$, polyalanine glycine. Asterisks indicate the strongest observation.

innermost and outermost layers of cocoons from 34 species of silkworm alongside the spectra of *N. edulis* dragline silk. Because of silk's molecular alignment, spectra will vary depending on the orientation of the fibres relative to the beam path (Boulet-Audet et al., 2008; Papadopoulos et al., 2007). For a fair comparison against cocoons without preferential orientation (Chen et al., 2012b), spider silk filaments were arranged in a similar random orientation order.

Fig. 2A shows spectra acquired from the innermost part of the cocoons from selected distinctive species. While the primary constituent of these cocoons is still silk proteins, the cocoons' infrared signature differed substantially from that of their respective feedstocks. We assigned these differences to a number of causes: the water content is lower in cocoons, reducing the ratio of amide I to amide II height (1642/1508 cm^{-1}); and precursor helical structures and random coils present in the feedstocks are converted via spinning into β-sheets, resulting in decreasing absorbance at 1642, 1547, 1308 and 1245 cm^{-1} and rising absorbance at 1699, 1620, 1508, 998 and 961 cm^{-1} (see Tables 1, 2). The relative absorbance of these β-sheet peaks can serve as an indicator of protein crystallinity. The peaks at 1699 and 1620 cm^{-1} in the amide I region are commonly used to determine the anti-parallel β-sheet content, but this overlaps with adjacent components from the fibroin and other compounds present in cocoons. In contrast, the low-frequency component at 961 cm^{-1} assigned to poly-alanine $(A)_n$ is much better resolved (Moore and Krimm, 1976; Papadopoulos et al., 2007; Taddei and Monti, 2005). This band also appears in the *N. edulis* dragline and most spectra of wild silk cocoons, particularly that of *Epiphora bauhiniae*. In contrast, some species like *B. mori* and *Anaphe panda* have two weaker peaks at 998 and 975 cm^{-1} as their β-sheets are constituted instead of poly(alanine–glycine) segments (Moore and Krimm, 1976; Taddei and Monti, 2005).

Fig. 2B demonstrates that the distinctive spectral features observed in the innermost layer of the cocoons are even more prominent in the outermost layer. For example, there is a higher relative absorbance of bands between 1395 and 1058 cm^{-1}, which is consistent with the greater amount of sericin in the outermost layer for species like *B. mori.* For comparison, a pure spectrum of sericin is included in Fig. 2A (Chen et al., 2012a). Another clear difference between the two layers is the amount of calcium oxalate [$Ca(COO)_2$] crystals found embedded in the outermost layer of some species (Freddi et al., 1994, 1993; Gheysens et al., 2011; Takahashi et al., 1969; Teigler and Arnott, 1972b). Calcium oxalate vibration modes at 1315 and 779 cm^{-1} dominate the outermost layer of *Gonometa postica* cocoons yet are much weaker in the innermost layer. Another spectral contrast between layers was found in *Opodipthera eucalypti*, where the shoulder at 1733 cm^{-1} can be assigned to the carboxylic acid and the polyphenol hydroxyls around 1000 cm^{-1} (Andrus, 2006; Silverstein et al., 1981).

Journal of Experimental Biology (2015) 218, 3138-3149 doi:10.1242/jeb.128306

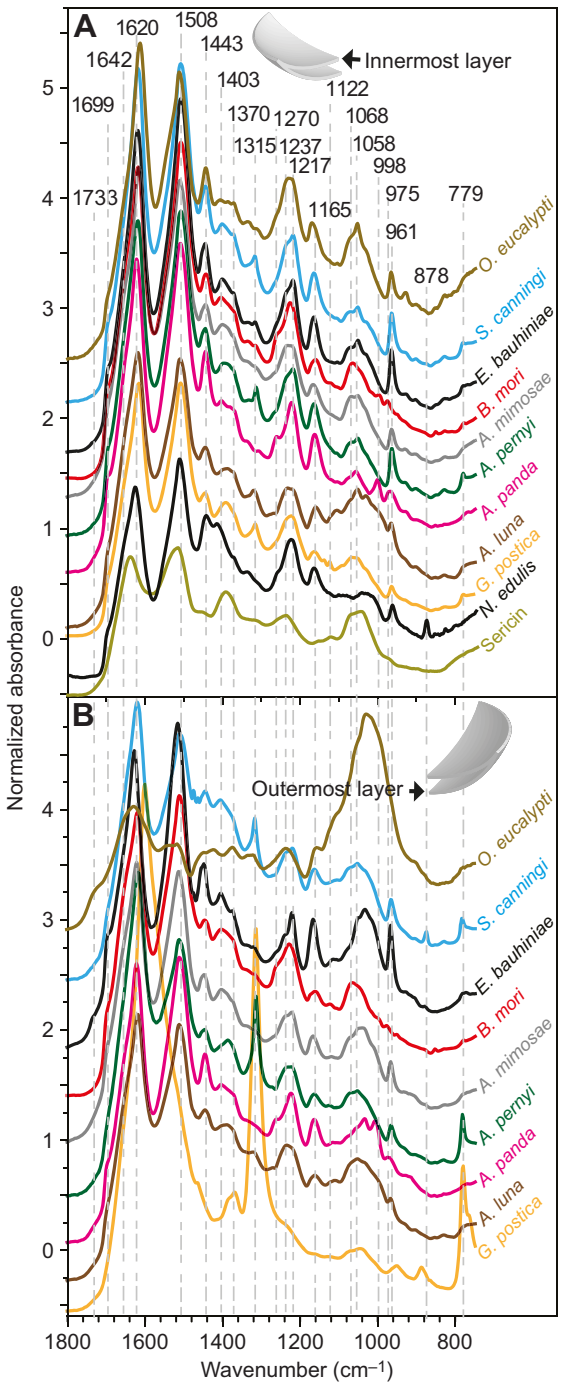

Fig. 2. Infrared spectra of the innermost and outermost layer of the cocoon of selected species. (A) Infrared spectra of the innermost cocoon layer of key species, and *B. mori* sericin spectrum for comparison. (B) Outermost cocoon layer spectra.

Thus, in summary, we established that it is possible to use structural and chemical markers to determine the type of crystallinity, the presence of sericin and calcium oxalate, and the polyphenol content in the measured cocoons.

Between-species comparison of the chemical composition of cocoons

For each spectrum collected, the interesting peaks identified above (see Fig. 2A,B) were integrated to estimate the relative content of calcium oxalate (1315 and 779 cm^{-1}), β-sheet crystallinity (1699, 1620, 1508, 998 and 961 cm^{-1}), the presence of tannin/phenolic compounds (1000 and 1733 cm^{-1}) and the sericin resin/matrix (1395 and 1058 cm^{-1}) (Fig. 3).

Calcium oxalate mineral crystals

Calcium oxalate, also called raphide, forms highly toxic needle-like crystals, which can tear soft tissues and are thought to represent a plant defence mechanism (Arnott and Webb, 2000). Because no known metabolic pathways process calcium oxalate in silkworms, we assume that calcium oxalate presence in the cocoon is a result of the ingestion of leaves containing the compound and resultant excretion by the silkworm. While this may be the case for wild silkworms, it appears that artificial selection has changed the behaviour of the *B. mori* silkworm to prevent this excretion into the cocoon. Fig. 3A shows the relative intensity of the band at 779 cm^{-1} achieved by integrating the absorbance between 740 and 800 cm^{-1}. This well-resolved band was used as a relative indicator of the amount of microscopic calcium oxalate monohydrate [$Ca(COO)_2$] mineral crystals present in the cocoons (Chen et al., 2012b; Gheysens et al., 2011). Our ATR-IR results identified cocoons from *G. postica* as having the highest calcium oxalate content. The host plant of *G. postica*, *Acacia*, is also rich in calcium oxalate as a means to detoxify calcium ions (Franceschi and Nakata, 2005; Martin et al., 2012; Teigler and Arnott, 1972a). The high calcium oxalate content of *G. postica* and *Antheraea* genera cocoons measured is in agreement with previous reports and electron microscopy observations on these cocoons (Chen et al., 2012b; Gheysens et al., 2011). Cocoons from *Samia*, *Hylophora* and *Attacus* species also indicate the presence of calcium oxalate, but in lower proportions while other species measured have only minute amounts in their cocoons.

The presence of calcium oxalate in the cocoon is known to complicate the industrial reeling as it prevents the extraction of long lengths of fibre (Gheysens et al., 2011). Calcium oxalate is notoriously toxic to humans and responsible for kidney stone formation (Evan et al., 2007). The commonplace edetic acid (EDTA) treatment for dissolving kidney stones was found to be equally effective at demineralizing wild silk cocoons containing calcium oxalate crystals and enabling industrial processing (Gheysens et al., 2011). Thus, the ability to detect and quantify the amount of calcium oxalate present in a cocoon prior to processing may have industrial advantages in minimizing reagent use or in selecting low mineral content cocoons in the first place.

β-Sheet crystallinity

X-ray scattering initially showed the presence of β-sheet nanocrystals inside silk fibres (Warwicker, 1954). Conveniently, polyalanine $(A)_n$ and polyalanine glycine $(AG)_n$ β-sheet structures also give distinctive peaks in silk infrared spectra, indicative of the degree of crystallinity present, and by extension may relate to mechanical properties (Boulet-Audet et al., 2008; Moore and Krimm, 1976; Porter et al., 2005; Sonoyama and Nakano, 2000). Using the integrated absorbance of $(A)_n$ antiparallel β-sheets peaking at 931–983 cm^{-1}, Fig. 3B shows the relative $(A)_n$ β-sheet content across the species tested. Our results suggest that *E. bauhiniae* has the highest degree of crystallinity amongst all the $(A)_n$-containing silks measured, followed by species from the *Samia*, *Antheraea* and *Attacus* genera. For most species, the $(A)_n$ β-sheet content appears greater in the innermost layer, probably due to non-fibroin compounds contributing to the infrared signal more on the outermost layer. Spider silk dragline from *N. edulis* appears to have a comparable $(A)_n$ β-sheet crystallinity to that of most

Journal of Experimental Biology (2015) 218, 3138-3149 doi:10.1242/jeb.128306

Table 2. Assignment of the main bands present in silk between 1308 and 779 cm^{-1}

Position (cm^{-1})	Assignment	Observation	References
1308	Amide III, β-turns	Some unspun silks except *B. mori* and *N. clavipes*	Bandekar and Krimm, 1980; Krimm and Bandekar, 1986; Rousseau et al., 2006
1270	Amide III, α-helices	All silks	Cai and Singh, 2004; Krimm and Bandekar, 1986
1237–45	Amide III, random coil	All silks	Cai and Singh, 2004; Shao et al., 2005; Taddei and Monti, 2005; Yoshimizu and Asakura, 1990
1217	Amide III, β-sheets	All spun silks **E. bauhiniae*	Cai and Singh, 2004; Moore and Krimm, 1976; Shao et al., 2005; Taddei and Monti, 2005; Yoshimizu and Asakura, 1990
1165	$\upsilon_s(N–C_\alpha)$	All silks **E. bauhiniae*	Barth, 2000; Moore and Krimm, 1976
1130	CH_2OH, polyphenols	Tanned cocoon silks **O. eucalypti* **A. edwardsii*	Lu et al., 2011; Schulz and Baranska, 2007
1103	υ(C–O), υ(C–C), polyphenols or tyrosine	Tanned cocoon silks Most wild silkworm silks **A. atlas*	Andrus, 2006; Barth, 2000
1068–75	$\upsilon(C–O)$, $\upsilon(N–C_\alpha)$, Ser	Silkworm cocoon *Outermost	Anghileri et al., 2007; Barth, 2000; Gupta et al., 1997
1052–58	$\upsilon(C_\beta–O)$, υ(C–OH), Ser	Most silkworm silks **A. luna*	Gupta et al., 1997; Taddei and Monti, 2005; Teramoto and Miyazawa, 2003
1028	$r(CH_2)$, $(A)_n$, random coil	Unspun *N. edulis* and wild silks	Moore and Krimm, 1976; Taddei et al., 2006
1016	$r(CH_2)$, $(AG)_n$, random coil	Unspun *B. mori* only	Moore and Krimm, 1976; Taddei and Monti, 2005
998	$r(CH_2)$, $(AG)_n$, β-sheets	**A. panda* **Bombyx*	Moore and Krimm, 1976; Taddei and Monti, 2005
975	$r(CH_2)$, $(AG)_n$, β-sheets	**A. panda* and *Bombyx*	Chen et al., 2012a; Moore and Krimm, 1976
961	$r(CH_2)$, $(A)_n$, β-sheets	Wild silks	Moore and Krimm, 1976; Papadopoulos et al., 2007
779	$\delta(OCO^-)$, calcium oxalate	**G. postica* Antheraea outer	Chen et al., 2012c; Gheysens et al., 2011; Sargut et al., 2010; Silverstein et al., 1981

δ, bending; υ, stretching; *r*, rocking.

silkworm silks measured. *Gonometa*, *Argema* and *Caligula* genera seem to have the lowest $(A)_n$ β-sheet content amongst all the species studied. Integrating the region between 984 and 1006 cm^{-1} quantified the contribution of the $(AG)_n$ peaks at 975 and 998 cm^{-1} while excluding the $(A)_n$ β-sheet peak at 961 cm^{-1}. Only three species appear to have $(AG)_n$ repetitive segments, *B. mori*, *Bombyx mandarina* and *A. panda* (see Fig. 3C). Unlike Bombycidae and Noctuidae families, none of the Saturniidae cocoons displayed peaks associated with the $(AG)_n$ structure (Moore and Krimm, 1976; Taddei and Monti, 2005). This fundamental distinction could be related to their appurtenance to different taxonomic families (see below).

Tannins and phenolic compounds

Wild silkworms naturally secrete some phenolic compounds in their silk (Brunet and Coles, 1974), but our results suggest that additional hydroxyl-containing compounds, such as polyphenols, could come from exogenous sources. By integrating the absorbance between 1035 and 1094 cm^{-1}, the relative amount of these molecules can be estimated. Fig. 3D shows that a few species had phenolic compounds, located mainly on the outside of the cocoon, including *O. eucalypti*, *Saturnia pyri*, *Hyalophora gloveri*, *Attacus edwardsii*, *Antheraea polyphemus* and *A. luna*. This finding agrees with the hypothesis that leaves incorporated by the silkworm into the cocoon structure leech water-soluble plant polyphenols when wet. In contrast, species that do not integrate leaves into their cocoons, such as *A. mylitta* and *A. atlas*, showed low phenolic compound parameter scores.

Sericin protein gum

Sericin proteins are essential to cocoon construction as they are used to bond fibres together (Chen et al., 2012a). The amount of sericin present in a cocoon can be inferred from the absorption bands between 1384 and 1403 cm^{-1} associated with the amino acid serine, which is present in high quantities in sericin but not in fibroin (Teramoto and Miyazawa, 2003). Fig. 3E suggests that *Bombyx* genus silks have the most sericin along with *Actias*, *Antheraea*, *Saturnia* and *Samia* genera silks. Our results indicate that there is less sericin in the coats of the high-porosity cocoons of species such as *Cricula trifenestrata*, *Graellsia isabellae* and *Loepa katinka* (Chen et al., 2012b). Differences in sericin abundance between the innermost and outermost layers of the cocoons tested indicate that *B. mori* cocoons have more sericin in the outermost layer, consistent with previous findings (Chen et al., 2012a). However, because of the additional mineral and phenolic components of the wild silks, it is challenging to interpret the distribution reliably in the other silks tested.

Classification of silk species

Our results show that the integration of infrared spectra bands assigned to individual compounds can provide select windows into a silk cocoon's chemical composition. However, single variable analysis exploits only a small fraction of the information contained within the spectra with thousands of data points. In contrast, multivariable analysis is far more powerful for classifying and discriminating samples. Hence, we first performed a principal

Journal of Experimental Biology (2015) 218, 3138-3149 doi:10.1242/jeb.128306

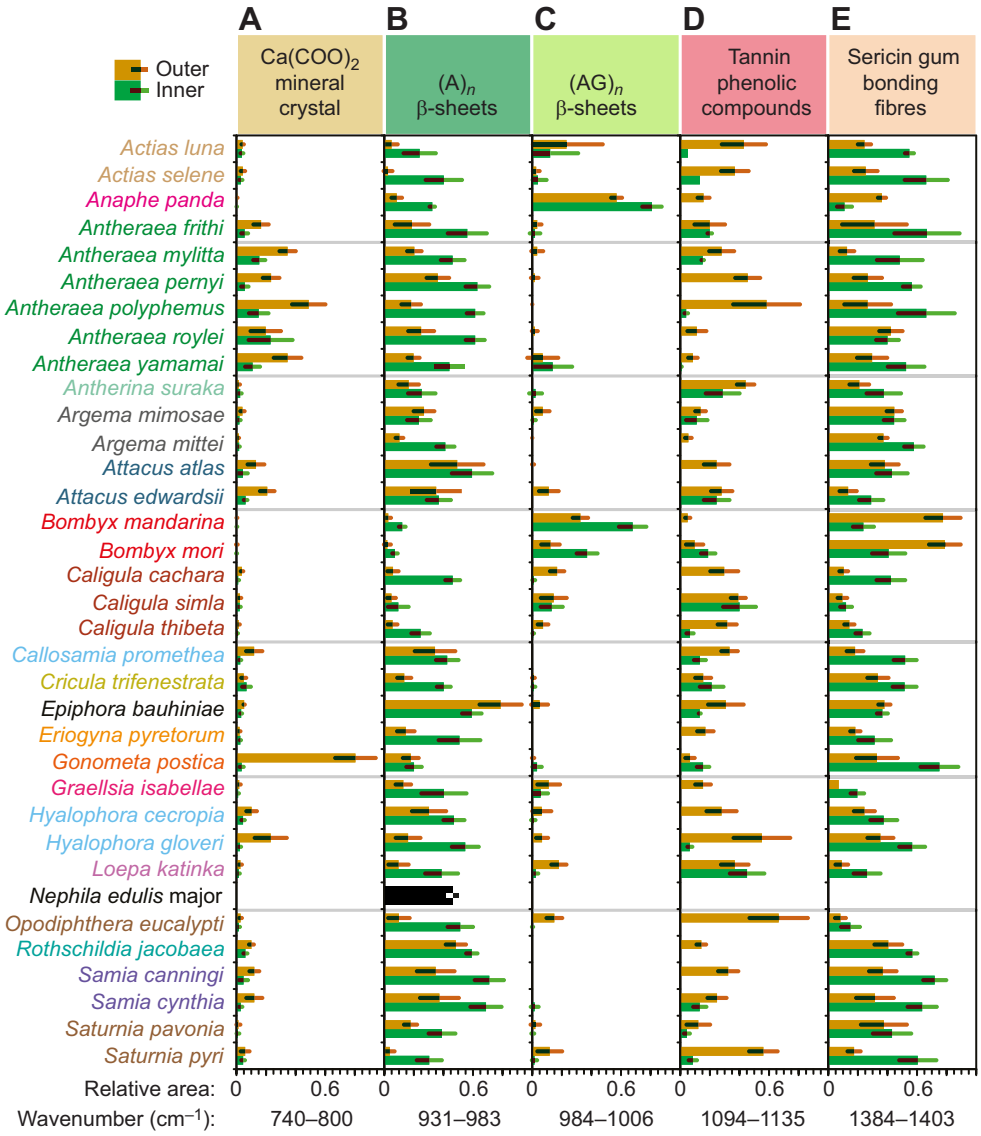

Fig. 3. Composition of cocoon outer and inner layers. (A) Relative area of a band assigned to calcium oxalate (740–800 cm^{-1}). (B) Relative area of the band associated with $(A)_n$ β-sheets (931–983 cm^{-1}). (C) Relative area of the band assigned to $(AG)_n$ β-sheet (984–1006 cm^{-1}). (D) Relative area of a band associated with tannins (1094–1135 cm^{-1}). (E) Relative area of a sericin marker band (1384–1403 cm^{-1}). The outermost and innermost layer values are shown (see key). The error bars represent the s.d. of the different observations (N>10). A value of 1 represents the highest area calculated and 0 represents the minimum measured.

component analysis (PCA) (Pearson, 1901) to reduce the number of variables while retaining most of the variability. The first principal component (PC) expresses the largest variance between samples. The PC scores, indicating the relative importance of these PC for each spectrum, were subsequently used for the linear discrimination analysis (LDA) to model the differences between species with a set of factor coefficients and scores.

The LDA scores were able to discriminate broadly between wild and domesticated silks as well as spider silk. Of the 25 measurements selected randomly for validation, the method identified the correct species for 100% of the 'unknown spectra' (see Materials and methods). Supplementary material Fig. S2 shows the tree generated from the LDA scores using hierarchical clustering analysis (HCA). However, as previously noted, once the silk has been spun into the cocoon structure, even more variables are introduced, and thus our multivariate approach becomes even more powerful. The multivariate analysis had an identification hit rate of 70% for species and 75% for genus, tested using the randomly selected validation group of 200 'unknown spectra' (see Materials and methods).

Our initial multivariate analysis of cocoon diversity is summarized in Fig. 4, which highlights the values of the first and second factor scores calculated from the cocoon spectra. The primary cluster encompasses most silks from wild silkworm species with *Antheraea* silks near its centroid (green markers). *Antherina suraka*, *L. katinka*, *E. bauhiniae* and *Samia cynthia* silks appear in the periphery of the cluster, suggesting a greater dissimilarity with the average of the measured silks. Clearly discriminated species outside this cluster such as *A. panda*, *B. mori* and *B. mandarina* appear as outliers. The *N. edulis* spider dragline silk is also outside the primary cluster, and easily discriminated from silkworm cocoons with the second factor scores. Notably, our LDA implies that *E. bauhiniae* is the silkworm species producing the closest silk to the *N. edulis* dragline. However, more species from other families would need to be studied to identify which of the thousands of silkworm species spins 'spider silk'. To develop this analysis further and begin to draw quantitative links between species, our HCA used the scores of the 10 most important factors to group these species according to their similarity.

Group 1: *Caligula*, *Saturnia* and *Actias*

Group 1 (see Fig. 5B) encompasses *Caligula*, *Saturnia* and *Actias* genera together with *O. eucalypti* and *C. trifenestrata*. As most species from this group had high absorbance between 1094 and

Journal of Experimental Biology (2015) 218, 3138-3149 doi:10.1242/jeb.128306

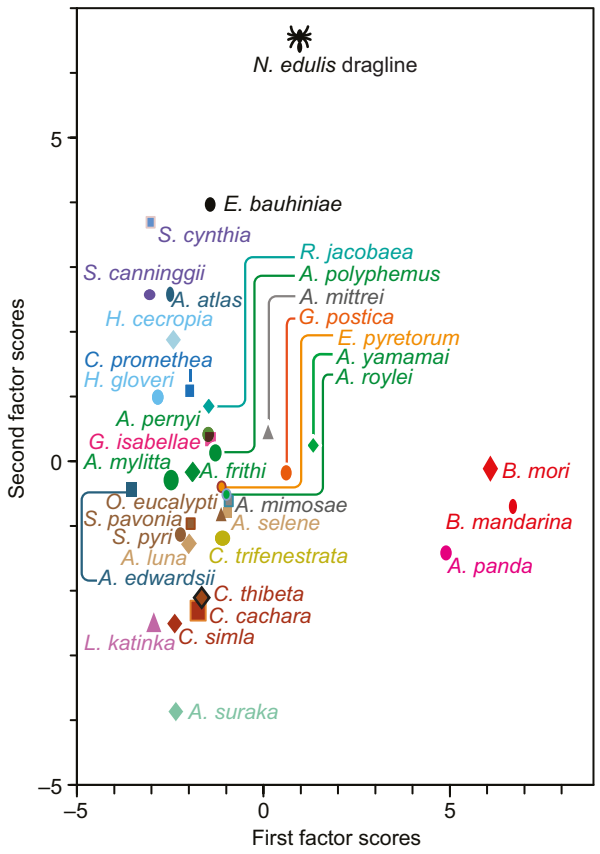

Fig. 4. Factor scores of the cocoon spectra. The first and second factor scores contribute to 62% of species discrimination of the linear discriminant analysis.

1135 cm^{-1} (Fig. 3D), this result suggests that these species were grouped together partly based on their high phenolic content. Except for *C. trifenestrata*, these species' cocoons appeared substantially tanned with a dark brown coloration (Chen et al., 2012c). Also, these species do not present calcium oxalate crystals on their surface (Chen et al., 2012c), as confirmed in Fig. 3A. Group 1 also appears to have a lower β-sheet content than the other groups.

Group 2: *Argema*

Group 2 contains the *Argema* genus. Unlike cocoons from group 1, they do not appear to have a high phenolic content or calcium oxalate, and although largely comparable to neighbouring groups these two factors could explain the large Euclidean distance from group 1.

Group 3: *Antheraea*

From our classification, it appears that *Antheraea* silks all have small Euclidean distances relative to one another and as such they were all grouped together with *A. suraka* in group 3. Previous studies based on morphological feature classification argued that *A. suraka* could be more closely related to the African Bunaeini tribe than other species of the Saturniini (Oberprieler, 1997). The comparable Euclidean distance between *A. suraka* and *Antheraea frithi* weakens this hypothesis. As *Antheraea* is the genus with the most calcium oxalate (Fig. 3A), the LDA method could have regrouped these silks mainly on mineral content. However, *A. suraka* and *A. frithi* show less absorption between 740 and 800 cm^{-1} (calcium oxalate) and are more distant from the other species of this group. In addition, group 3 has weaker phenolic compound bands than group 1 and yet has an average amount of sericin. The next closest species to these groups are *L. katinka* and *G. isabellae*, which both present a low sericin content and high porosity according to other studies (Chen et al., 2012b,c).

Group 4: *Attacus* and *Samia*

Much more distant are the species classified in group 4, including *Samia*, *Hyalophora* and *Attacus* genera along with *Callosamia promethea*. When compared with previous reports, the morphology of the cocoons classified together in group 4 appears characteristic as the innermost layers are much more compact than their outer layers (Chen et al., 2012c). This morphological difference could explain the higher amount of sericin measured in the innermost layer (Fig. 3E). Adding to the composition variation between the innermost and outermost layers, these cocoons have an intermediate content of $(A)_n$ β-sheets, sericin and tannin when compared with those from groups 1 and 3. Branching from group 4, *E. bauhiniae* has the largest amount of $(A)_n$ β-sheets, little phenolic compounds, no calcium oxalate and little sericin.

Group 5: *Bombyx*

The LDA placed *B. mori* and *B. mandarina* silks into a distant group. Even though *B. mandarina* appears to have more $(AG)_n$ β-sheets than *B. mori* (Fig. 3C), the difference between their spectra is subtle in comparison with the other species presented. This result suggests that the artificial selection of *B. mori* might have played a lesser role than natural selection in differentiating this species from other Lepidoptera families.

Silk from other superfamilies

While still very distant, *A. panda* had the smallest Euclidean distance to *Bombyx*. Our results suggest that *A. panda* also has $(AG)_n$ β-sheets, sericin and no calcium oxalate. Testing more silks from these two families would confirm whether these silks share the same spectral features. Social spinning behaviour is another interesting characteristic of *A. panda*, which partners with many other worms to build a communal cocoon nest (Mbahin et al., 2007). The cocoon's structure does not depend on the silk quality of a single individual, resulting in different natural selection constraints. Also from another superfamily (Lasocampiadae), *G. postica* silk is very distinct from all other species studied. The innermost layer appears closer to Saturniidae silks (Fig. 2), whereas the major difference between their outer layer is likely to come from the large amount of calcium oxalate present.

Comparison of ATR-IR and phylogenetic trees

The ultrametric tree generated from the infrared spectra (see Fig. 5A) was compared with the phylogenetic tree built from the sequencing of a few protein-coding nuclear genes by Regier et al. (2008a,b, 2002; see also Chen et al., 2012c). The genes selected to construct this phylogeny produce proteins other than silk, with various enzymatic functions such as carbamoylphosphate synthetase, aspartate transcarbamylase, dihydroorotase (Moulton and Wiegmann, 2004), dopa decarboxylase (Fang et al., 1997), enolase (Farrell et al., 2001) and wingless (Brower and DeSalle, 1998). Although a quantitative comparison between an ultrametric and a unitless tree is not possible, they are strikingly similar, except for few species.

DISCUSSION

By assessing the diversity of wild silks, this study compared the biochemical composition of native silk feedstock from six species and silk cocoons from 34 species using infrared spectroscopy and

Journal of Experimental Biology (2015) 218, 3138-3149 doi:10.1242/jeb.128306

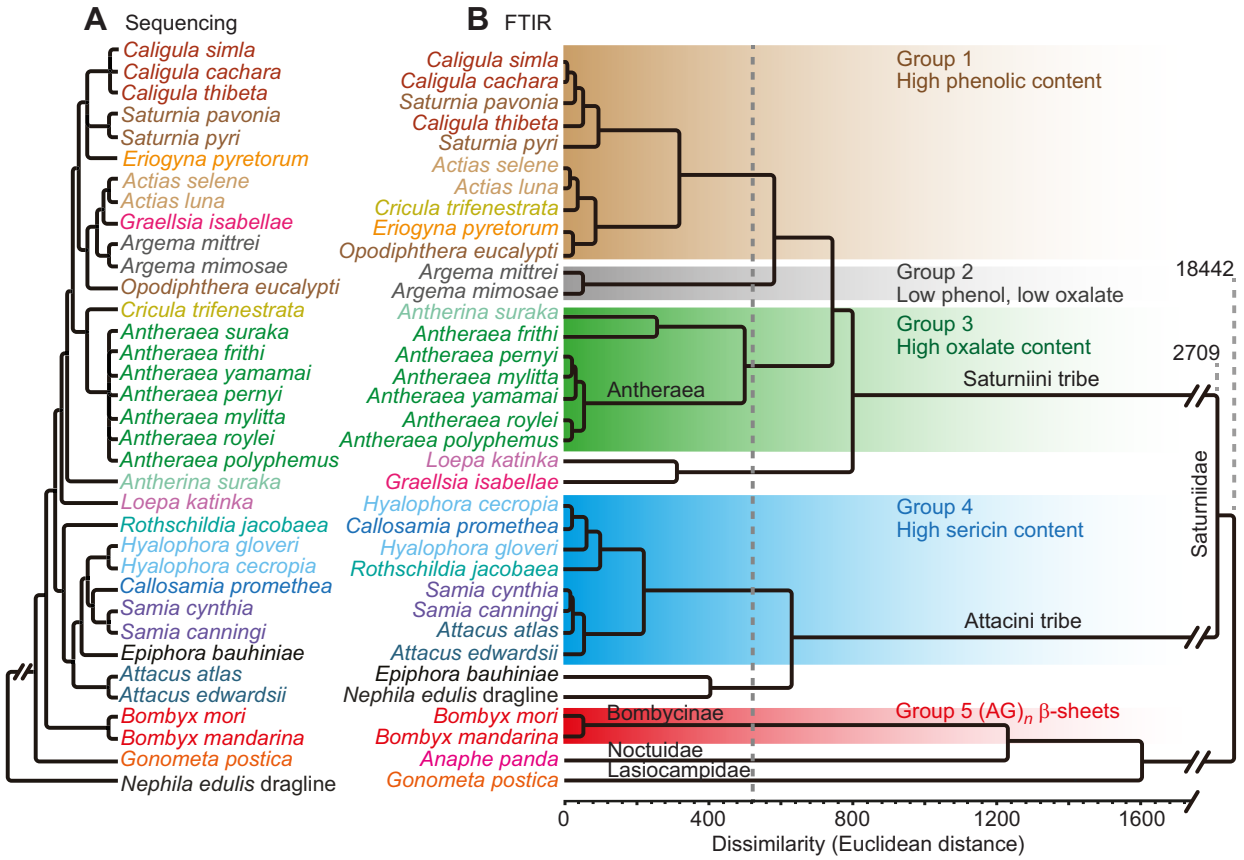

Fig. 5. Classification of silk species. (A) Cladogram generated from the phylogenetic analysis of Regier et al. (2005, 2008a,b, 2002; see also Chen et al., 2012c). (B) Ultrametric tree generated from the hierarchical clustering analysis of cocoon infrared spectra LDA factor scores. Species with a Euclidean distance smaller than 525 were grouped together. FTIR, Fourier transform infrared spectroscopy.

multivariate analysis. For unspun native silk feedstocks, we identified new spectral markers unique to wild silkworm silks, which we assigned to β-turn secondary structures. The hierarchical clustering of the feedstocks also profiled the dissimilarity of Saturniidae silks to the silks of Bombycidae and spiders.

Collecting spectra from silkworm cocoons provided information not only on the spun fibre but also on the non-protein chemical content and distribution across the layers. The specific infrared bands revealed the relative content of sericin, calcium oxalate, phenolic compounds, $(A)_n$ and $(AG)_n$ β-sheets. The multivariate analysis also permitted the hierarchical classification of 35 species (including one spider silk) into groups based on their chemical composition. This analysis revealed the presence of interesting outlier species with very dissimilar spectra, which could manifest as distinctive mechanical or chemical properties. Amongst these outliers were *G. postica* cocoons, which had the highest calcium oxalate of all species measured. Furthermore, the species with the most β-sheets, *E. bauhiniae*, also appeared to have the closest chemical composition to *N. edulis* spider silk dragline. The *Bombyx* genus stood out from all other species measured, representing an outlier group. Consequently, using *B. mori* as the model species for silk studies could lead to conclusions that are not applicable to all types of silks. Although our sampling had a bias towards Saturniidae silk, *Antheraea* silks were found to have median PC scores, suggesting that *Antheraea* silks are more representative of silk biodiversity. Not only did the multivariate analysis have a species identification hit rate of 70% but also the ultrametric trees were created from the infrared spectra.

Our analysis thus suggests a relationship between non-silk coding nuclear genes selected by Regier et al. and the silkworm cocoon's overall biochemical composition (Regier et al., 2005, 2008a,b, 2002). Such a link implies that infrared spectra could be used as a proxy for the phylogenetic classification of species. Despite huge similarities between these trees, a few silk species were classified differently under these two approaches. This difference could be the result of non-protein-based variation such as temperature or humidity or the incorporation of exogenous material into the cocoons. For instance, *C. trifenestrata* was expected to be closer to the *Antheraea* silks rather than classified into group 1. The difference could be due to the fact that *C. trifenestrata* lives in an environment with a warm climate, requiring more ventilation than *Antheraea* silk cocoons found in colder regions (Kakati and Chutia, 2009). Interestingly, *G. isabellae* silk should have been very similar to *Actias* silks, but was classified by our analysis outside group 1 along with *L. katinka*. Their separate classification could result from the high concentration of tannins measured in the cocoons of these species. Oberprieler and Nassig (1994) suggested that these species might have been misclassified, and our study strengthens the hypothesis that *Graellsia* and *Actias* are two distinct genera. As expected, *E. bauhiniae* was classified in the Attacini tribe but is rather distant from the other species of group 4, most likely because of its higher $(A)_n$ β-sheet crystallinity content. In summary, despite minor differences in the classifications, our method represents a powerful but straightforward hierarchical classification tool to help resolve some of the ambiguity in the relationships of Lepidoptera species.

Journal of Experimental Biology (2015) 218, 3138-3149 doi:10.1242/jeb.128306

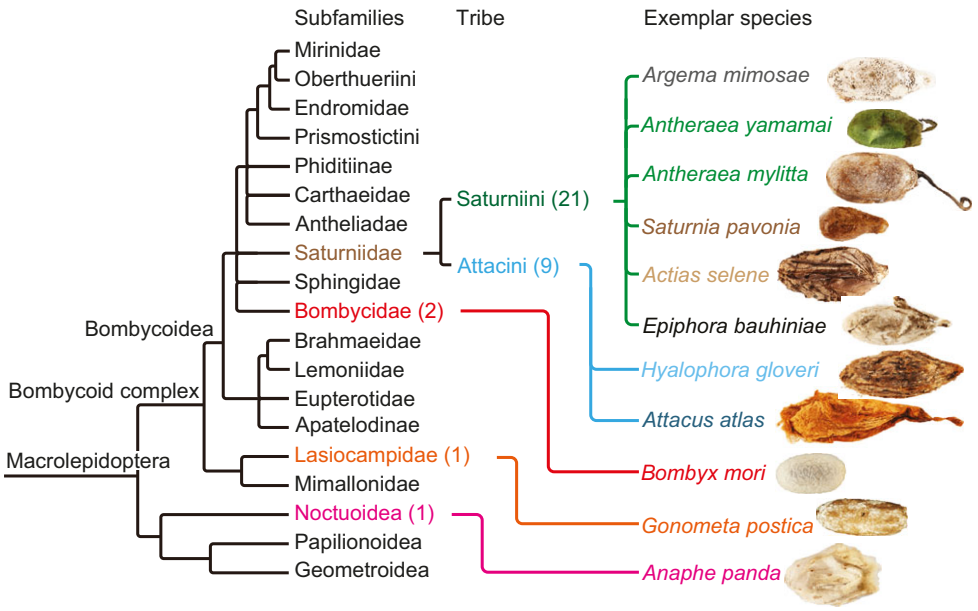

Fig. 6. Summary of higher-level relationships of the subfamilies related to species studies adapted from Regier et al. (2008a). Numbers in parentheses represent the number of species measured in the subfamilies. The images on the right are the cocoons of the species measured.

Because of the intense selection pressure on this vital biological structure, we believe silk cocoons represent a model for the phylogenetic analysis of all silkmoth species. This untapped proxy method not only adds to more traditional gene and protein sequencing but is also less time consuming and cheaper than, for example, protein sequencing. As silk cocoons are commonly part of entomology collections spanning hundreds of years of sampling, they can be readily sourced and rapidly tested in a non-destructive manner. Such powerful longitudinal studies could shed light on silkmoth evolution and ecology by helping to resolve some of the relationship ambiguities of Lepidoptera species. Furthermore, with the advent of affordable handheld IR instruments, our approach could also allow such analysis to take place in the field. Thus, combining ATR-IR with multivariate analysis could aid in unravelling the evolution and biodiversity of silk-producing species as well as inform us regarding which species is best suited to a particular industrial application.

MATERIALS AND METHODS

Native silk feedstock preparation

All wild silkworm eggs were purchased from Worldwide Butterflies (WWB, Dorset, UK). *Actias luna*, *A. yamamai*, *A. atlas* and *S. pavionia* were fed with walnut (*Juglans regia*), hawthorn (*Crataegus monogyna*), privet (*Ligustrum vulgare*) and hawthorn (*C. monogyna*), respectively. When larvae started spinning their cocoon, native silk feedstocks were extracted from the silk glands of last instar silkworms. Final instar *B. mori* worms were fed with white mulberry leaves (*Morus alba*). *Nephila edulis* major ampullate glands and dragline were extracted from mature female spiders fed with *Drosophila* spp. and *Caliphora* spp. and reared in-house under controlled temperature and humidity as described elsewhere (Holland et al., 2006).

Silk cocoon preparation

We analysed cocoons from 34 species across the Lepidoptera. The superfamilies Saturniini and Attacini are highlighted in Fig. 6. The International Centre of Insect Physiology and Ecology (icipe; African Insect Science for Food and Health) in Kenya provided *G. postica* silkworm cocoons. The other cocoon species were purchased from WWB; the species chosen cover four Lepidopteran families. At least four, 3.5 mm round cocoon discs were cut from each cocoon using a plier punch for analysis by infrared spectroscopy.

Spectral acquisition and treatment

A Golden Gate single bounce diamond ATR accessory (Specac Ltd, London, UK) coupled to a Nicolet 6700 FTIR spectrometer equipped with a MCT nitrogen cooled detector (Thermo Scientific, Madison, WI, USA) was used for spectra collection. Spectra acquisition was performed at a 4 cm^{-1} resolution from 500 to 6000 cm^{-1}, averaging 32–64 scans at a 5.06 cm s^{-1} mirror speed. Although the fibres in a cocoon sample are randomly oriented, spectra were collected with the infrared beam polarized perpendicular to the plane of incidence (*s*) with a zinc selenide holographic wire grid polarizer (Thermo Scientific). The ATR diamond's internal reflection element (IRE) had a refractive index of 2.417 and an angle of incidence of 45 deg. For this configuration, the evanescent wave emerging out of the IRE could probe around 1.2 µm deep into the sample, the penetration varying with the wavelength (Harrick, 1967). The liquid state of native feedstock spectra ensured a good contact with the IRE for data collection. As the cocoons have an inherent roughness greater than tens of micrometres, an anvil was used to press on the cocoon discs to ensure a good contact with the IRE. For acquisition consistency, the pressure applied on cocoon discs was kept to the minimum necessary to obtain an absorbance of 0.1 for the amide II band. By aiming to keep the absolute absorbance consistent, the anomalous dispersion of the refractive index was therefore comparable for each spectrum collected (Boulet-Audet et al., 2010). Before each measurement, the crystal was cleaned with tissue and demineralized water before a new background was acquired. This method helped to compensate for the detector's signal fluctuations as well as preventing contamination between measurements. The innermost and outermost layers of these discs were measured by collecting at least 18 distinct spectra from each species for a total of 1185 spectra across the 35 species studied. This spectroscopic study thus encompasses the largest number of wild silk types to date.

Data pre-processing

Spectral operations were performed using OMNIC 7.3 (Thermo Scientific) using a custom-written VBA code. An offset was first subtracted from all spectra as calculated from the average of the region from 1950 to 1900 cm^{-1}. Spectra were then normalized using the integrated absorbance from 1900 to 800 cm^{-1} to compensate for absolute signal variations incurred by differing cocoon contact with the IRE. For the single component analysis, the relative area of each peak integrated was calculated by subtracting a linear baseline between the interval limits from the integrated absorbance.

Multivariate analysis and dendrogram generation

Despite high spectral reproducibility, the absolute absorbance values vary between measurements depending on the contact between the porous

cocoon and the IRE. To discriminate silks based on their spectral line shape and peak position rather than absolute absorbance values, the multivariate analysis was performed on the first derivative of the spectra. The second derivative is as effective, but enhances the noise further (Kansiz et al., 1999). Mid-infrared spectra contained 2853 variables, but were not all interdependent as a single compound spectral line often contributed in different regions simultaneously. To reduce the number of variables while preserving most of the dataset variability, a Pearson PCA (Pearson, 1901) was performed using XLSTAT (Addinsoft, Paris, France). Keeping only 10 PCs still preserved 90% of native silk feedstock spectral variability, while selecting the first 40 PCs still accounted for 86% of cocoon spectral variability.

A LDA (Fisher, 1936; Yang et al., 2005) was performed on the PC scores to find a linear combination of features that separate infrared spectra from silks of different species; 27 of 52 native silk spectra and 962 of the 1162 cocoon spectra were randomly selected for the training (estimation) group to construct the discrimination function. The remaining 25 native silk and 200 cocoon spectra were used to validate the discrimination function. The LDA achieved a hit rate of 100% for native silk and 70% for cocoon spectra while assigning 75% to the correct genus.

The LDA factor centroid scores were subsequently used to calculate the Euclidean distance between each species for Ward's HCA (Mariey et al., 2001; Ward, 1963). This method minimizes the total variance within clusters starting from singleton clusters (one species per cluster) in a top-down approach. The resulting HCA dendrograms were subsequently compared with the phylogenic tree dendrogram built from genetic data (Chen et al., 2012c; Regier et al., 2005, 2008a,b, 2002).

Supplementary material Fig. S1 shows the infrared spectra before and after shear-induced denaturation of *B. mori* and *A. atlas* along with their corresponding difference spectrum. Supplementary material Fig. S2 shows the ultrametric tree generated from the infrared spectra of native feedstock primary canonical functions.

Supplementary material Fig. S2 shows the tree generated from the infrared spectra of native feedstock main canonical functions using HCA. Sharing common spectral features such as the 961, 1103 and 1308 cm^{-1}, feedstocks from *A. luna*, *S. pavonia* and *A. atlas* are more closely related. This result corroborates the fact that these four species are from the same Saturniidae arthropod superfamily. Although much more distant, the closest to these species is the silkworm silk *B. mori* feedstock as it is also a silkworm silk feedstock containing sericin proteins. With fewer types of fibroins and no sericin, spider silk feedstock infrared spectra are very distinct. Relative to the silkworm feedstock's dissimilarity, the spider silk feedstock tested has a much greater Euclidean distance.

Acknowledgements
The authors wish to thank Dr Fujia Chen, Dr Ann Terry, Dr David Porter and Dr Beth Mortimer for their critical comments on the manuscript. The authors also wish to thank Dr Fujia Chen, Julia Van Campen, Alexander Greenhalgh and Addis Akebede for assisting with data collection.

Competing interests
The authors declare no competing or financial interests.

Author contributions
The manuscript was written with contributions from all authors. All authors gave approval for the final version of the manuscript.

Funding
This study was supported by the European Research Council (SP2-GA-2008-233409), Magdalen College Oxford, the UK Engineering and Physical Sciences Research Council (EPSRC; P/K005693/1; EP/G068224/1), the Canadian Natural Sciences and Engineering Research Council (NSERC; PGS 3D/6799-379132-2009) and the US Air Force Office of Scientific Research (AFOSR F49620-03-1-0111).

Supplementary material
Supplementary material available online at
http://jeb.biologists.org/lookup/suppl/doi:10.1242/jeb.128306/-/DC1

References
Andrus, P. G. (2006). Cancer monitoring by FTIR spectroscopy. *Technol. Cancer Res. Treat.* **5**, 157-167.
Anghileri, A., Lantto, R., Kruus, K., Arosio, C. and Freddi, G. (2007). Tyrosinase-catalyzed grafting of sericin peptides onto chitosan and production of protein-polysaccharide bioconjugates. *J. Biotechnol.* **127**, 508-519.
Arnott, H. J. and Webb, M. A. (2000). Twinned raphides of calcium oxalate in grape (Vitis): implications for crystal stability and function. *Int. J. Plant Sci.* **161**, 133-142.
Astudillo, M. F., Thalwitz, G. and Vollrath, F. (2014). Life cycle assessment of Indian silk. *J. Clean. Prod.* **81**, 158-167.
Bandekar, J. and Krimm, S. (1979). Vibrational analysis of peptides, polypeptides, and proteins: characteristic amide bands of beta-turns. *Proc. Natl. Acad. Sci. USA* **76**, 774-777.
Bandekar, J. and Krimm, S. (1980). Vibrational analysis of peptides, polypeptides, and proteins. VI. Assignment of beta-turn modes in insulin and other proteins. *Biopolymers* **19**, 31-36.
Barth, A. (2000). The infrared absorption of amino acid side chains. *Prog. Biophys. Mol. Biol.* **74**, 141-173.
Borkner, C. B., Elsner, M. B. and Scheibel, T. (2014). Coatings and films made of silk proteins. *ACS Appl. Mater Interfaces* **6**, 15611-15625.
Boulet-Audet, M., Lefèvre, T., Buffeteau, T. and Pézolet, M. (2008). Attenuated total reflection infrared spectroscopy: an efficient technique to quantitatively determine the orientation and conformation of proteins in single silk fibers. *Appl. Spectrosc.* **62**, 956-962.
Boulet-Audet, M., Buffeteau, T., Boudreault, S., Daugey, N. and Pézolet, M. (2010). Quantitative determination of band distortions in diamond attenuated total reflectance infrared spectra. *J. Phys. Chem. B* **114**, 8255-8261.
Boulet-Audet, M., Vollrath, F. and Holland, C. (2011). Rheo-attenuated total reflectance infrared spectroscopy: a new tool to study biopolymers. *Phys. Chem. Chem. Phys.* **13**, 3979-3984.
Boulet-Audet, M., Terry, A. E., Vollrath, F. and Holland, C. (2014). Silk protein aggregation kinetics revealed by Rheo-IR. *Acta Biomater.* **10**, 776-784.
Briandet, R., Kemsley, E. K. and Wilson, R. H. (1996). Discrimination of *arabica* and *robusta* in instant coffee by Fourier transform infrared spectroscopy and chemometrics. *J. Agric. Food Chem.* **44**, 170-174.
Brower, A. V. Z. and DeSalle, R. (1998). Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. *Insect Mol. Biol.* **7**, 73-82.
Brunet, P. C. J. and Coles, B. C. (1974). Tanned silks. *Proc. R. Soc. Lond. B Biol. Sci.* **187**, 133-170.
Cai, S. and Singh, B. R. (2004). A distinct utility of the amide III infrared band for secondary structure estimation of aqueous protein solutions using partial least squares methods. *Biochemistry* **43**, 2541-2549.
Chen, X., Shao, Z., Knight, D. P. and Vollrath, F. (2007). Conformation transition kinetics of *Bombyx mori* silk protein. *Proteins Struct. Funct. Bioinform.* **68**, 223-231.
Chen, F., Porter, D. and Vollrath, F. (2010a). Silkworm cocoons inspire models for random fiber and particulate composites. *Phys. Rev. E* **82**, 041911.
Chen, Z., Hao, X. and Fan, K. (2010b). Preparation of polyvinyl alcohol film inlaid with silk fibroin peptide nano-scale particles and evaluation of its function to promote cell growth. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi* **27**, 1292-1297.
Chen, F., Porter, D. and Vollrath, F. (2012a). Silk cocoon (*Bombyx mori*): multi-layer structure and mechanical properties. *Acta Biomater.* **8**, 2620-2627.
Chen, F., Porter, D. and Vollrath, F. (2012b). Structure and physical properties of silkworm cocoons. *J. R. Soc. Interface* **9**, 2299-2308.
Chen, F., Porter, D. and Vollrath, F. (2012c). Morphology and structure of silkworm cocoons. *Mater. Sci. Eng. C Mater. Biol. Appl.* **32**, 772-778.
Chevillard, M., Couble, P. and Prudhomme, J.-C. (1986). Complete nucleotide sequence of the gene encoding the *Bombyx mori* silk protein P25 and predicted amino acid sequence of the protein. *Nucleic Acids Res.* **14**, 6341-6342.
Chirgadze, Y. N., Fedorov, O. V. and Trushina, N. P. (1975). Estimation of amino acid residue side-chain absorption in the infrared spectra of protein solutions in heavy water. *Biopolymers* **14**, 679-694.
Colthup, N. B. (1964). *Introduction to Infrared and Raman Spectroscopy*. New York: Academic Press.
Craig, C. L. (1997). Evolution of arthropod silks. *Annu. Rev. Entomol.* **42**, 231-267.
Danks, H. V. (2004). The roles of insect cocoons in cold conditions. *Eur. J. Entomol.* **101**, 433-437.
Dicko, C., Porter, D., Bond, J., Kenney, J. M. and Vollratht, F. (2008). Structural disorder in silk proteins reveals the emergence of elastomericity. *Biomacromolecules* **9**, 216-221.
Dooley, D. P. J. (2004). *The Social Costs of Underemployment: Inadequate Employment as Disguised Unemployment*. Cambridge, NY, USA: Cambridge University Press.
Evan, A. P., Coe, F. L., Lingeman, J. E., Shao, Y., Sommer, A. J., Bledsoe, S. B., Anderson, J. C. and Worcester, E. M. (2007). Mechanism of formation of human calcium oxalate renal stones on Randall's plaque. *Anat. Rec. Adv. Integr. Anat. Evol. Biol.* **290**, 1315-1323.
Fang, Q. Q., Cho, S., Regier, J. C., Mitter, C., Matthews, M., Poole, R. W., Friedlander, T. P. and Zhao, S. (1997). A new nuclear gene for insect phylogenetics: Dopa decarboxylase is informative of relationships within heliothinae (Lepidoptera: Noctuidae). *Syst. Biol.* **46**, 269-283.

Farrell, B. D., Sequeira, A. S., O'Meara, B. C., Normark, B. B., Chung, J. H. and Jordal, B. H. (2001). The evolution of agriculture in beetles (Curculionidae: Scolytinae and Platypodinae). *Evolution* **55**, 2011-2027.
Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugenics* **7**, 179-188.
Franceschi, V. R. and Nakata, P. A. (2005). Calcium oxalate in plants: formation and function. *Annu. Rev. Plant Biol.* **56**, 41-71.
Freddi, G., Svilokos, A. B., Ishikawa, H. and Tsukada, M. (1993). Chemical composition and physical properties of gonometa rufobrunnae silk. *J. Appl. Polym. Sci.* **48**, 99-106.
Freddi, G., Gotoh, Y., Mori, T., Tsutsui, I. and Tsukada, M. (1994). Chemical structure and physical properties of *Antheraea assama* silk. *J. Appl. Polym. Sci.* **52**, 775-781.
Garside, P., Lahlil, S. and Wyeth, P. (2005). Characterization of historic silk by polarized attenuated total reflectance Fourier transform infrared spectroscopy for informed conservation. *Appl. Spectrosc.* **59**, 1242-1247.
Gheysens, T., Collins, A., Raina, S., Vollrath, F. and Knight, D. P. (2011). Demineralization enables reeling of wild silkmoth cocoons. *Biomacromolecules* **12**, 2257-2266.
Goormaghtigh, E., Ruysschaert, J.-M. and Raussens, V. (2006). Evaluation of the information content in infrared spectra for protein secondary structure determination. *Biophys. J.* **90**, 2946-2957.
Gupta, A., Tandon, P., Gupta, V. D. and Rastogi, S. (1997). Vibrational dynamics and heat capacity of beta-poly(L-serine). *Polymer* **38**, 2389-2397.
Harrick, N. J. (1967). *Internal Reflection Spectroscopy*. New York: John Wiley & Sons.
Holland, C., Terry, A. E., Porter, D. and Vollrath, F. (2006). Comparing the rheology of native spider and silkworm spinning dope. *Nat. Mater.* **5**, 870-874.
Horrocks, N. P. C., Vollrath, F. and Dicko, C. (2013). The silkmoth cocoon as humidity trap and waterproof barrier. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **164**, 645-652.
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. et al. (2001). Cloning of the fibroin gene from the oak silkworm, Antheraea yamamai and its complete sequence. *Biotechnol. Lett.* **23**, 1321-1326.
Inoue, S., Tanaka, K., Arisaka, F., Kimura, S., Ohtomo, K. and 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. *J. Biol. Chem.* **275**, 40517-40528.
Ishii, S., Inokuchi, T., Kanazawa, J. and Tomizawa, C. (1984). Studies on the cocoon of the oriental moth, *Monema* (*Cnidocampa*) *flavescens*, (Lepidoptera: Limacodidae). III. Structure and composition of the cocoon in relation to hardness. *Jpn. J. Appl. Entomol. Zool.* **28**, 269-273.
Jeong, L., Lee, K. Y., Liu, J. W. and Park, W. H. (2006). Time-resolved structural investigation of regenerated silk fibroin nanofibers treated with solvent vapor. *Int. J. Biol. Macromol.* **38**, 140-144.
Kakati, L. N. and Chutia, B. C. (2009). Diversity and ecology of wild sericigenous insects in Nagaland, India. *Trop. Ecol.* **50**, 137-146.
Kansiz, M., Heraud, P., Wood, B., Burden, F., Beardall, J. and McNaughton, D. (1999). Fourier Transform Infrared microspectroscopy and chemometrics as a tool for the discrimination of cyanobacterial strains. *Phytochemistry* **52**, 407-417.
Krimm, S. and Bandekar, J. (1986). Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv. Protein Chem.* **38**, 181-364.
Lu, X., Wang, J., Al-Qadiri, H. M., Ross, C. F., Powers, J. R., Tang, J. and Rasco, B. A. (2011). Determination of total phenolic content and antioxidant capacity of onion (*Allium cepa*) and shallot (*Allium oschaninii*) using infrared spectroscopy. *Food Chem.* **129**, 637-644.
Mariey, L., Signolle, J. P., Amiel, C. and Travert, J. (2001). Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics. *Vibr. Spectrosc.* **26**, 151-159.
Martin, G., Guggiari, M., Bravo, D., Zopfi, J., Cailleau, G., Aragno, M., Job, D., Verrecchia, E. and Junier, P. (2012). Fungi, bacteria and soil pH: the oxalate-carbonate pathway as a model for metabolic interaction. *Environ. Microbiol.* **14**, 2960-2970.
Mbahin, N., Raina, S. K., Kioko, E. N. and Mueke, J. M. (2007). Spatial distribution of cocoon nests and egg clusters of the silkmoth *Anaphe panda* (Lepidoptera: Thaumetopoeidae) and its host plant *Bridelia micrantha* (Euphorbiaceae) in the Kakamega Forest of western Kenya. *Int. J. Trop. Insect Sci.* **27**, 138-144.
Minoura, N., Aiba, S.-I., Gotoh, Y., Tsukada, M. and Imai, Y. (1995). Attachment and growth of cultured fibroblast cells on silk protein matrices. *J. Biomed. Mater. Res.* **29**, 1215-1221.
Mita, K., Ichimura, S. and James, T. C. (1994). Highly repetitive structure and its organization of the silk fibroin gene. *J. Mol. Evol.* **38**, 583-592.
Miyazawa, T. and Blout, E. R. (1961). Infrared spectra of polypeptides in various conformations - amide I and II bands. *J. Am. Chem. Soc.* **83**, 712-719.
Moore, W. H. and Krimm, S. (1976). Vibrational analysis of peptides, polypeptides, and proteins. II. Beta-Poly(L-alanine) and Beta-Poly(L-alanylglycine). *Biopolymers* **15**, 2465-2483.
Moulton, J. K. and Wiegmann, B. M. (2004). Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoic-aged Eremoneuran Diptera (Insecta). *Mol. Phylogenet. Evol.* **31**, 363-378.
Navarro, M., Benetti, E. M., Zapotoczny, S., Planell, J. A. and Vancso, G. J. (2008). Buried, covalently attached RGD peptide motifs in poly(methacrylic acid) brush layers: the effect of brush structure on cell adhesion. *Langmuir* **24**, 10996-11002.
Oberprieler, R. G. (1997). Classification of the African Saturniidae (Lepidoptera) - the quest for natural groups and relationships. *Metamorphosis Occas. Suppl.* **3**, 142-155.
Oberprieler, R. G. and Nassig, W. A. (1994). Tarn - oder Warntrachten - ein Vergleich larvaler und imaginaler Strategien bei Saturniinen (Lepidoptera: Saturniidae). *Nachr. Entomol. Ver. Apollo* **15**, 267-303.
Omenetto, F. and Kaplan, D. (2010). From silk cocoon to medical miracle. *Sci. Am.* **303**, 76-77.
Panayiotou, H. and Kokot, S. (1999). Matching and discrimination of single human-scalp hairs by FT-IR micro-spectroscopy and chemometrics. *Anal. Chim. Acta* **392**, 223-235.
Papadopoulos, P., Sölter, J. and Kremer, F. (2007). Structure-property relationships in major ampullate spider silk as deduced from polarized FTIR spectroscopy. *Eur. Phys. J. E* **24**, 193-199.
Pearson, K. L.III. (1901). On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **2**, 559-572.
Porter, D., Vollrath, F. and Shao, Z. (2005). Predicting the mechanical properties of spider silk as a model nanostructured polymer. *Eur. Phys. J. E* **16**, 199-206.
Preisner, O., Lopes, J. A., Guiomar, R., Machado, J. and Menezes, J. C. (2007). Fourier transform infrared (FT-IR) spectroscopy in bacteriology: towards a reference method for bacteria discrimination. *Anal. Bioanal. Chem.* **387**, 1739-1748.
Regier, J. C., Fang, Q. Q., Mitter, C., Peigler, R. S., Friedlander, T. P. and Solis, M. A. (1998). Evolution and phylogenetic utility of the period gene in Lepidoptera. *Mol. Biol. Evol.* **15**, 1172-1182.
Regier, J. C., Mitter, C., Peigler, R. S. and Friedlander, T. P. (2002). Monophyly, composition, and relationships within Saturniinae (Lepidoptera: Saturniidae): evidence from two nuclear genes. *Insect Syst. Evol.* **33**, 9-21.
Regier, J. C., Paukstadt, U., Paukstadt, L. H., Mitter, C. and Peigler, R. S. (2005). Phylogenetics of eggshell morphogenesis in Antheraea (Lepidoptera: Saturniidae): unique origin and repeated reduction of the aeropyle crown. *Syst. Biol.* **54**, 254-267.
Regier, J. C., Cook, C. P., Mitter, C. and Hussey, A. (2008a). A phylogenetic study of the 'bombycoid complex' (Lepidoptera) using five protein-coding nuclear genes, with comments on the problem of macrolepidopteran phylogeny. *Syst. Entomol.* **33**, 175-189.
Regier, J. C., Grant, M. C., Mitter, C., Cook, C. P., Peigler, R. S. and Rougerie, R. (2008b). Phylogenetic relationships of wild silkmoths (Lepidoptera: Saturniidae) inferred from four protein-coding nuclear genes. *Syst. Entomol.* **33**, 219-228.
Rousseau, M.-È., Beaulieu, L., Lefèvre, T., Paradis, J., Asakura, T. and Pézolet, M. (2006). Characterization by Raman Microspectroscopy of the strain-induced conformational transition in fibroin fibers from the silkworm *Samia cynthia ricini*. *Biomacromolecules* **7**, 2512-2521.
Roy, M., Meena, S. K., Kusurkar, T. S., Singh, S. K., Sethy, N. K., Bhargava, K., Sarkar, S. and Das, M. (2012). Carbondioxide gating in silk cocoon. *Biointerphases* **7**, 45.
Sargut, S. T., Sayan, P. and Kiran, B. (2010). Influence of essential and non-essential amino acids on calcium oxalate crystallization. *Crystal Res. Technol.* **45**, 31-38.
Schulz, H. and Baranska, M. (2007). Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vibr. Spectrosc.* **43**, 13-25.
Scoble, M. J. (1999). *Geometrid Moths of the World*. Collingwood, Victoria: CSIRO Publishing.
Shao, J., Zheng, J., Liu, J. and Carr, C. M. (2005). Fourier transform Raman and Fourier transform infrared spectroscopy studies of silk fibroin. *J. Appl. Polym. Sci.* **96**, 1999-2004.
Silverstein, R. M., Bassler, G. C. and Morrill, T. C. (1981). *Spectrometric Identification of Organic Compounds*. New York: Wiley.
Sonoyama, M. and Nakano, T. (2000). Infrared rheo-optics of *Bombyx mori* fibroin film by dynamic step-scan FT-IR spectroscopy combined with digital signal processing. *Appl. Spectrosc.* **54**, 968-973.
Sukopp, M., Marinelli, L., Heller, M., Brandl, T., Goodman, S. L., Hoffman, R. W. and Kessler, H. (2002). Designed beta-turn mimic based on the allylic-strain concept: evaluation of structural and biological features by incorporation into a cyclic RGD peptide (cyclo(-L-arginylglcyl-L-alpha-aspartyl-)). *Helv. Chim. Acta* **85**, 4442-4452.
Taddei, P. and Monti, P. (2005). Vibrational infrared conformational studies of model peptides representing the semicrystalline domains of *Bombyx mori* silk fibroin. *Biopolymers* **78**, 249-258.
Taddei, P., Arai, T., Boschi, A., Monti, P., Tsukada, M. and Freddi, G. (2006). In vitro study of the proteolytic degradation of *Antheraea pernyi* silk fibroin. *Biomacromolecules* **7**, 259-267.

Journal of Experimental Biology

Takahashi, S. Y., Suzuki, G. and Ohnishi, E. (1969). Origin of oxalic acid in Ca oxalate crystals in the Malpighian tubes of the tent caterpillar, *Malacosoma neustria testacea*. *J. Insect Physiol.* **15**, 403-407.
Tanaka, K. and 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 Biochem. Mol. Biol.* **31**, 665-677.
Teigler, D. J. and Arnott, H. J. (1972a). Crystal development in the Malpighian tubules of *Bombyx mori* (L.). *Tissue Cell* **4**, 173-185.
Teigler, D. J. and Arnott, H. J. (1972b). X-ray diffraction and fine structural studies of crystals in the Malpighian tubules of silkworms. *Nature* **235**, 166-167.
Teramoto, H. and Miyazawa, M. (2003). Analysis of structural properties and formation of sericin fiber by infrared spectroscopy. *J. Insect Biotechnol. Sericol.* **72**, 157-162.
Teramoto, H. and Miyazawa, M. (2005). Molecular orientation behavior of silk sericin film as revealed by ATR infrared spectroscopy. *Biomacromolecules* **6**, 2049-2057.
Venyaminov, S. Y. and Kalnin, N. N. (1990). Quantitative IR spectrophotometry of peptide compounds in water (H_2O) solutions. I. Spectral parameters of amino acid residue absorption bands. *Biopolymers* **30**, 1243-1257.
Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236.
Warwicker, J. O. (1954). The crystal structure of silk fibroin. *Acta Crystallogr.* **7**, 565.
Yang, H., Irudayaraj, J. and Paradkar, M. M. (2005). Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. *Food Chem.* **93**, 25-32.
Yoshimizu, H. and Asakura, T. (1990). The structure of *Bombyx mori* silk fibroin membrane swollen by water studied with Esr, 13C-Nmr, and FT-IR spectroscopies. *J. Appl. Polym. Sci.* **40**, 1745-1756.
Zhao, H., Parry, R. L., Ellis, D. I., Griffith, G. W. and Goodacre, R. (2006). The rapid differentiation of Streptomyces isolates using Fourier transform infrared spectroscopy. *Vibr. Spectrosc.* **40**, 213-218.

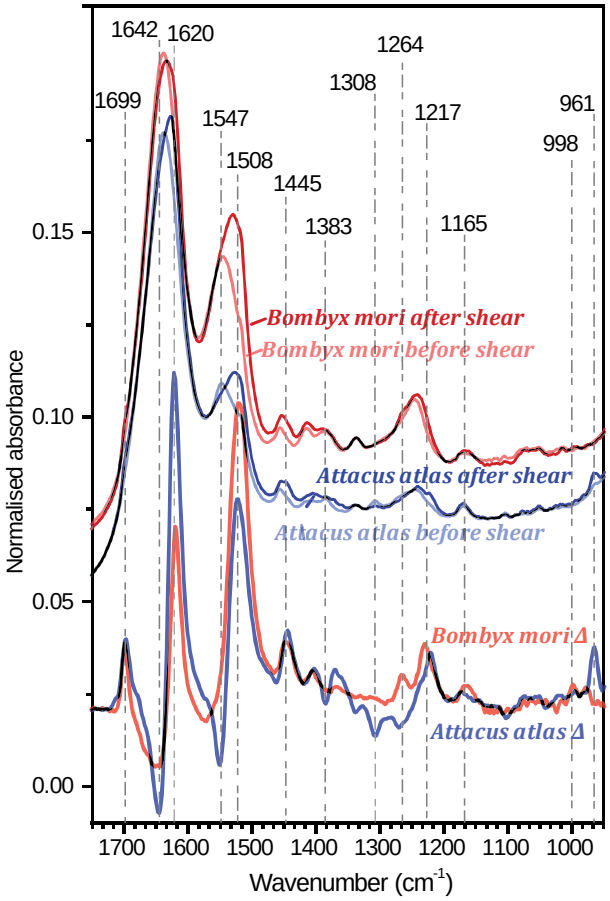

Fig. S1. Infrared spectra before and after shear induced denaturation of *Bombyx mori* and *Attacus atlas* along with their corresponding difference spectrum.

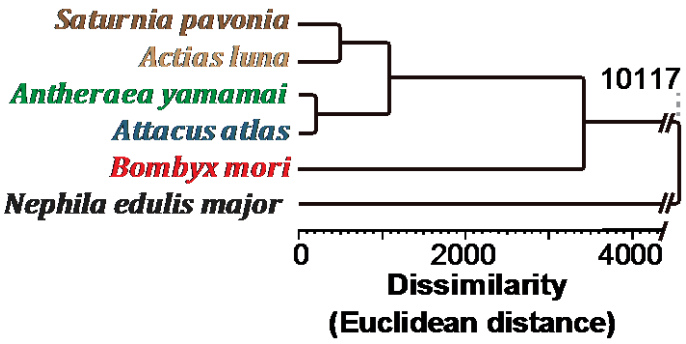

Fig. S2. Ultrametric tree generated from the infrared of native feedstock LDA scores.