ELSEVIER

International Journal of Biological Macromolecules 40 (2007) 201–216

INTERNATIONAL JOURNAL OF
Biological
Macromolecules
STRUCTURE, FUNCTION AND INTERACTIONS

www.elsevier.com/locate/ijbiomac

Patterns of variation among distinct alleles of the *Flag* silk gene from *Nephila clavipes*

Linden E. Higgins [a,*], Sheryl White [b], Juan Nuñez-Farfán [c], Jesus Vargas [c]

[a] *Department of Biology, University of Vermont, 109 Carrigan Drive, Marsh Life Sciences 120A, Burlington, VT 05405, United States*
[b] *Department of Anatomy and Neurobiology, University of Vermont, Burlington, VT 05405, United States*
[c] *Instituto de Ecología, Universidad Nactional Autonoma de México, México, D.F., Mexico*

Received 14 March 2006; received in revised form 22 July 2006; accepted 22 July 2006
Available online 28 July 2006

Abstract

Spider silk proteins and their genes are very attractive to researchers in a wide range of disciplines because they permit linking many levels of organization. However, hypotheses of silk gene evolution have been built primarily upon single sequences of each gene each species, and little is known about allelic variation within a species. Silk genes are known for their repeat structure with high levels of homogenization of nucleotide and amino acid sequence among repeated units. One common explanation for this homogeneity is gene convergence. To test this model, we sequenced multiple alleles of one intron–exon segment from the *Flag* gene from four populations of the spider *Nephila clavipes* and compared the new sequences to a published sequence. Our analysis revealed very high levels of heterozygosity in this gene, with no pattern of population differentiation. There was no evidence of gene convergence within any of these alleles, with high levels of nucleotide and amino acid substitution among the repeating motifs. Our data suggest that minimally, there is relaxed selection on mutations in this gene and that there may actually be positive selection for heterozygosity.

Keywords: Silk gene; Flagelliform silk; Nephila; Tetragnathidae; Allelic variation

1. Introduction

Silk proteins and their genes are very attractive to researchers in a wide range of disciplines including materials scientists [1,2], and evolutionary ecologists [3]. Spiders are a particularly attractive model system for the study of silk genes, as each individual produces multiple distinct silk types, each from a particular set of glands (reviewed in Ref. [3]). These different silks are used by the spiders in diverse functions, ranging from prey capture to nest building and egg-sac protection. Recent advances in sequencing of many silk genes from spiders have allowed researchers to begin to link the molecular structure to physical properties and to ecological function. These comparative studies are further enhanced by phylogenetic reconstructions that provide an evolutionary framework for understanding the evolution of these genes. Thus far, the consensus is that these diverse genes form a gene family, that evolved through gene duplication and subsequent divergence and specialization [4].

The gene coding for the spider flagelliform silk protein, *Flag*, is one of the more recently characterized silk genes [5,6], and codes for a silk found in a relatively derived clade of spiders, the araneoids (orb-weaving spiders and their relatives [7]). Flagelliform silk forms the core of the viscid spiral, and is remarkable for its toughness and elasticity [1]. Like many silk genes in spiders and other arthropods, the *Flag* gene has a nested structure. Each exon includes many small motif repeats, in this case that code for the amino acid motif GPGGX. These series of repeating amino acid motifs are interspersed with non-repeating sequences forming a higher-level repeating unit. In flagelliform silk protein, the larger motifs are ensemble repeats of up to 61 GPGGX motifs forming high-glycine regions that flank non-repetitive glycine-poor spacers [6]. Each ensemble repeat corresponds to an individual exon. An estimated 13 exons are separated from one another by introns which are themselves very similar to each other across the *Flag* gene. The GPG portion of the motif is believed to form a β-helix, which may act like a spring in this exceedingly elastic silk [5].

* Corresponding author. Tel.: +1 802 656 0454.
E-mail address: Linden.Higgins@uvm.edu (L.E. Higgins).

0141-8130/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijbiomac.2006.07.007

Observed patterns of variation among homologous genes across species, and, in repetitive proteins, among repeat motifs within a sequence, have been used as the basis for various models of evolutionary changes in gene and protein structure (i.e. [3]). Some authors have speculated that the repeating motifs of silk genes may function like "minisatellite" DNA, with misalignment during recombination generating variation upon which selection can act [3,6]. Another common model evoked to explain patterns of variation among motifs is that high frequencies of recombination increase the chances of motif duplication and deletion. "Chi-like" sequences (gctggag [8]) have been reported in other organisms to increase the frequency of recombination, and such sequences have been found at several locations in silk genes [3]. In contrast to these models predicting increased variation, the low levels of variation among motifs within a sequence is hypothesized to be due to gene conversion [6,8].

However, most of these models of molecular mechanisms of evolution in spider silks are based upon comparisons across repeated regions within a single sequence or upon comparisons among species, with only a single sequence for a particular species (e.g. [4], however, see [9,10]). To truly understand the patterns of mutational change in microevolution of a particular protein, one needs to study allelic variation within a species. The patterns found among species of web-building spiders and among genes within a species reflect evolutionary history, but the grist of natural selection is variation among individuals, which can only be studied by sequencing the same gene repeatedly within and among populations and examining allelic variation. If, for instance, misalignment during recombination leads to variation in number of motifs within an exon [3], this pattern should be visible among sequences within a species as high levels of allelic variation in number of motifs within a particular exon among alleles. Likewise, the hypothesis that "chi-like" sequences increase the likelihood of duplication and deletion is best tested by investigating within a species whether alleles are more variable near such sequences.

To better understand the patterns of variation in a gene with repeating motifs, we compared 17 sequences of *Flag* from 10 individual *Nephila clavipes* (Linnaeus) (Araneae: Tetragnathidae) from five disjunct populations, one in Florida [6] and four in Mexico. The sequences were all from the last intron–exon pair and included the initial portion of the carboxyl terminus [6]. Overall, we found nearly identical patterns of insertion and deletion of entire motifs in alleles from Mexican populations compared to Florida. Among the Mexican sequences, we found no patterns that suggesting that homogenization across motifs is occurring and indeed in the repeating region of the gene there is a very high frequency of single nucleotide substitutions and amino acid substitutions. In contrast, the region coding the upstream portion of the C terminus had much lower frequency of single nucleotide polymorphisms (snps) and amino acid substitution, which could reflect the hypothesized organizational function of this portion of the protein [9,11]. We conclude that high rates of mutation and amino acid substitution are at least tolerated in the repeating region of this exon, perhaps because this is a secreted protein, and that there may actually be positive selection for heterozygosity in this gene.

2. Methods

2.1. Sampling

Spiders were sampled in Mexico in July 2002. We choose populations that are geographically disjunct and inhabit very different habitats (Table 1). Data from population genetics studies of these populations indicate that there is little gene flow among them (Nuñez Farán and Vargas, pers. commun.). Within each site, we sampled haphazardly, by finding large juvenile or mature females and pinching a leg to cause automization, which we then placed in 100% ethanol, one specimen per vial.

2.2. DNA extraction

DNA was extracted from whole legs through salting-out, in a protocol modified from Sunnucks and Hales [12] for the larger mass of the spider legs. We dried specimens 2 h before extraction, to insure complete evaporation of the ethanol, and incubated the legs with TNES and proteinase K for 18 h at room temperature. After precipitating the proteins with NaCl, we precipitated the DNA with cold 100% ethanol. In some cases, DNA visibly precipitated and was removed, dried and resuspended in water. If DNA did not visibly precipitate, or after removing any precipitated DNA, we centrifuged the sample (14,000 rpm 5 min), to collect the remaining DNA. These DNA specimens were air dried, resuspended in water (PCR quality, 100 μl if tibia + patella length was greater than 0.75 cm, otherwise 500 μl), and then further cleaned with a phenol/chloroform extraction, reprecipitated with ethanol and sodium acetate, and resuspended in water. All DNA specimens were stored in 50 μl water at −80 °C.

Table 1
Field sites

Site	Initials	Co-ordinates	Habitat type
Xalapa, Veracruz	Xal	19°30′45N, 96°52′78W	Mid-altitude temperate, coffee plantation
Fortin de las Flores, Veracruz	FF	18°54′N, 96°59′56W	Mid-altitude temperate, private garden
Los Tuxtlas region, Veracruz	LT, Nan	18°27′40N, 95°3′96W	Lowland tropical, forest preserves
Mateos de Romero, Oaxaca	Mat	16°53′N, 95°02′W	Lowland, semi-deciduous tropical, private ranch

Los Tuxtlas (LT) and Nanciyaga (Nan) are two sampling sites in the same region separated by approximately 10 km; Xalapa and Fortin are separated from each other by approximately 100 km of desert uninhabited by *N. clavipes*, and these mid-altitude sites are separated from the coastal sites by cane fields where *N. clavipes* is not found.

2.3. PCR

PCR was performed using the Failsafe PCR system (Epicentre Technologies). Spider DNA (2 μg) was combined with 1 μM each of forward and reverse primer designed from the GenBank sequence for *N. clavipes Flag* (GenBank no. AF218621.S2 DNA; 3′Flag_forward: 5′-gcaaccgcctcatcgtcatttcgtac-3′, 3′Flag_reverse: 5′-gcgaacattcttcctacaga-3′), Failsafe enzyme mix and PCR premix I. Reactions were then put in the thermal cycler (Techne) and cycled for 40 cycles under the following conditions: 30 s at 96 °C, 1 min at 55 °C, and 2.5 min at 72 °C.

2.4. Cloning and isolation

We purified all PCR products using gel purification (Qiagen) and purified products were then cloned into TOPO vectors (Invitrogen 2.1-TOPO TA cloning kit with One-shot *E. coli*), following the manufacturers' protocols. After extracting plasmid DNA (Qiagen miniprep) from recombinant colonies, we did an *Eco*R1 (Invitrogen) digest to verify the presence of the full size insert within purified plasmids. Known positive clones were sent to the Vermont Cancer Center sequencing facility for sequencing.

2.5. Sequencing and alignment

The sequencing was done using three primers, as the target DNA sequence was 2.3 Kb long. The end primers we used were the universal primers, M13 forward and reverse, and a unique middle primer (Flagmid: 5′-tgcaggtgtaggacctgatggaagtg-3′). The three contiguous sequences were aligned in MacClade (version 4 [13]) by hand, as the repeating nature of the DNA sequence resulted in errors in automated alignment (i.e., sequencher reversed the middle sequence to align the 3′-end over the 3′-end of the upstream sequence). Only sequences that overlapped by at least 20 bp were included to assure proper alignment of the contiguous sequences. We sequenced 2–4 clones from each individual and utilized the cleanest (lowest number of ambiguous readings and longest overlap between contiguous sections) in alignment. In most cases, we obtained two clearly distinct alleles from each individual, and thus include two sequences. The merged sequences were then aligned to each other and then relative to the published downstream sequence from a Florida population ([6] GenBank no. AF218621.S2 DNA). Alignment with a cDNA sequence from the same Florida population ([6] GenBank no. AF027973 cDNA) was used to determine where the exon started. To verify that all the sequences came from the same section of the *Flag* gene, all were run through BLAST. The best match for all the sequences was the same sequence that we had used to align the sequences (AF218621.S2).

To verify the accuracy of hand-alignment of these sequences, the Mexican sequences were also aligned to each other using both muscle [14] and T-coffee [15]. SinicView [16] was then used to compare the three alignments (hand, muscle, and T-coffee) of each individual sequence. All of the alignments were in agreement, with 98% or greater concordance.

3. Results

The sequences we obtained showed very high levels of variation with an overall snp frequency of 123/2180 or 5.64%, but all corresponded to the same region of the Florida sequence (BLAST results show ≥95% correspondence for positions 676–1590 and positions 2047–2790 of the GenBank sequence AF218621.S2). Among the new sequences, the single nucleotide substitutions were most often silent (third codon) indicating that most were not generated during the amplification and cloning procedure, which would be blind to codon position. The amino acid substitutions are not evenly distributed across the entire exon, but are more common in the high-glycine (repeating motif) section than in other sections. However, we did find that one portion of the high-glycine region had exceptionally low amino acid substitution rates. Finally, there were few instances of motif insertion or deletion among the Mexican sequences but alignment of these Mexican sequences with the sequence from Florida show a series of insertions and deletions of entire motifs.

The aligned nucleotide sequences are presented in Appendix A and a schematic is presented in Fig. 1. Comparison of multiple sequences of the same clone showed only three nucleotide substitutions apparently due to the sequencing reaction ($N = 11$ clones or 22 sequences, 384–877 nucleotides sequenced, for a total of 12,722 bases or an error rate of 1/4240 bases sequenced). However, we suspected that the preparation of this complex gene might be error prone. To determine the error rate during PCR and cloning, we sequenced multiple clones for six individuals and matched sequences that appeared to be from the same allele, primarily using patterns of insertions and deletions of motifs but also matching according to the smaller number of single nucleotide differences between sequences. The apparent error rate during PCR and cloning is higher than the error rate during sequencing, ranging from a low of 1 in 2075 nucleotides to a high of 14 in 2089 nucleotides. The average rate including all duplicate clones was 33/8786 nucleotides sequenced (1/265.6) but removing the one high point (14 differences) it drops to 1/352.5 nucleotides. The variation among specimens in error rate presumably reflects the quality of the template DNA. Because of this apparently high error rate, for each of these repeatedly sequenced alleles we choose to use the sequence that had the lowest number of differences (snps) compared to the other Mexican sequences in the remaining analyses, presuming it to be the least error prone.

Overall, 5.22% of the nucleotide positions were variable. Most of the snps appear to be due to real differences among alleles, not errors generated during sequence preparation. This assumption is supported by the observation that a significant fraction of snps in the exon are in the third codon position (38/79 or 48.1%; $\chi^2 = 8.226$, d.f. = 2, $p < 0.05$), a bias that would not be selected for during sequence preparation.

There are some differences in frequency of nucleotide polymorphisms among the four regions of the sequenced portion of this gene. In the intron, there are 40 polymorphic sites over 725 base pairs (5.51%), in the repeating region of the exon (which includes one short non-repeating region), there were 62 poly-

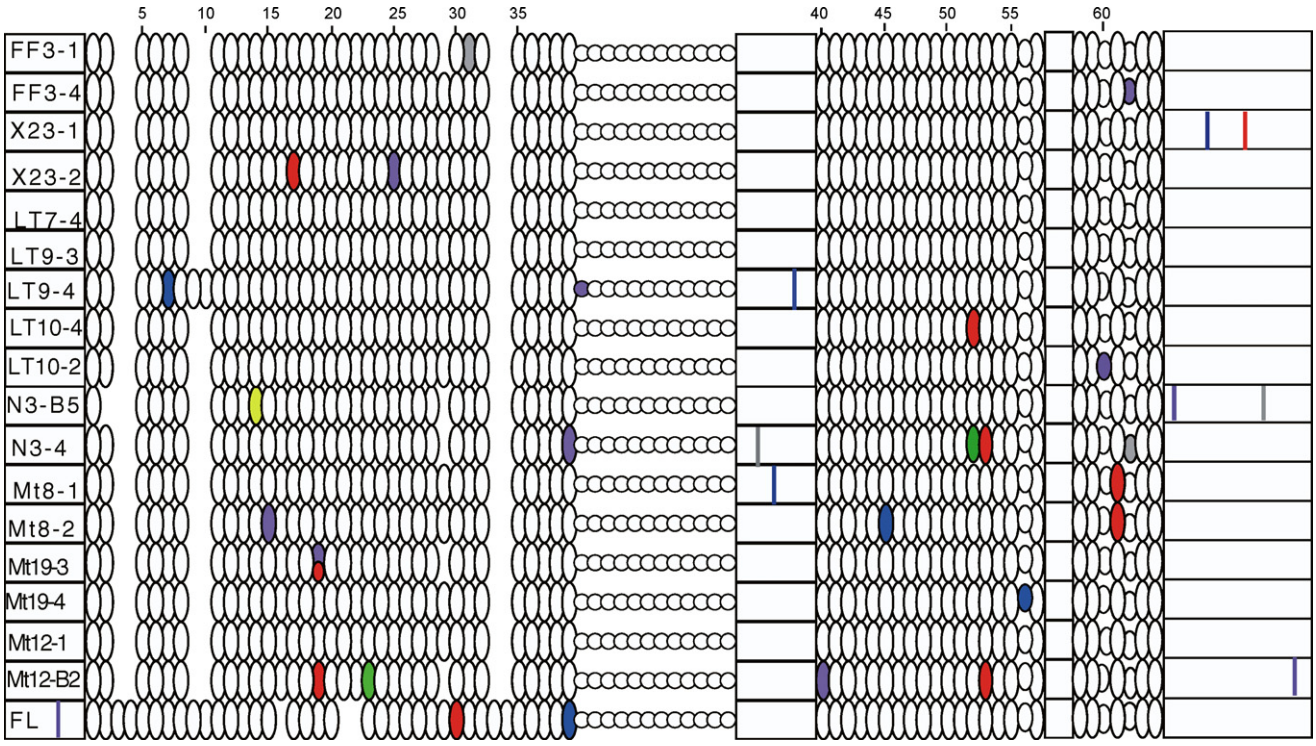

Fig. 1. Graphical representation of the pattern of indels and amino acid polymorphisms among the Mexican sequences and the two GenBank sequences from Florida. The first letter refer to the site of collection (Table 1), the first number refers to the individual specimen, and the number after the dash is the sequence identifier for that individual (when two sequences are presented, they are assumed to be separate alleles for a heterozygous individual). The different repeat motifs are identified by the following symbols: large oval, XGPGG; medium oval, XGGY; circles, XGG; box, low glycine (non-repetitive) spacer. A colored oval or a bar in the spacer indicates an amino acid substitution, where only one amino acid substitution was found in all motifs except number 19. The nature of the substitution is indicated by the color: grey, same group; purple, exchange of glycine and long chain residue; blue, exchange of hydrophobic and hydrophilic residue; green, exchange of aromatic and straight chain residue; yellow, addition or deletion of cysteine; red, addition or deletion of proline. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

morphic sites/1111 bp (5.6%), in the spacer there were seven polymorphic sites/78 bp (8.97%) and in the C terminus region there were 10 polymorphic sites/264 bp (3.79%). Several of these polymorphic sites have more than two character states. The differences in snp frequency among the repeating, spacer, and C terminus regions of the exon are not significant when the differences in sequence lengths are taken into account ($\chi^2 = 3.183$, d.f. = 2, $p > 0.05$).

Examining the presumed protein sequence (obtained from MacClade translation using the standard genetic code, Appendix B, Fig. 1), the differences in frequency of amino acid polymorphic sites among the different regions of the protein are very apparent. There are three different motifs in the high-glycine section: XGPGG, XGGY, and XGG (note that we have rearranged these motif codings from [6] to accommodate the most common break point for insertions and deletions). Among the variable motifs, all but one vary at only one amino acid position (i.e., only one substitution per motif) hence our ability to color code the substitutions in Fig. 1. The XGGY motif is most prone to variation, with three variable residues or approximately 25% of the 12 amino acids in the three repetitions (which are, notably, scattered across the exon). Twenty-two of the 61 XGPGG motifs are variable (7.2% of the amino acid positions), and most of these substitutions were at the "X" position (Appendix B). The least variable motif is the XGG section, which has 12 contiguous motifs or 36 amino acids, and only 1 (2.7%) polymorphic amino acid position. Six of the 12 XGG motifs are AGG, a purported hotspot for recombination in silk [3]. Across the entire exon, there are eight AGG (three coded by gctggag), however there is only one amino acid substitution among them, and no other evidence of increased recombination (unless the homogeneity in XGG reflects gene conversion through unequal recombination).

Most of the 35 amino acid polymorphisms involved changes from one functional group to another (Fig. 1). Eleven involved gain or loss of a proline (seven were losses of the proline from the XGPGG motif or the variants on this motif near the carboxyl end), 10 were exchanges between glycine and an amino acid with a long side chain (six involved loss of a glycine from the XGPGG motif), and seven involved the interchange of a hydrophobic and hydrophilic residue.

The most striking difference between all the Mexican sequences and the Florida sequences is the insertion or deletion of entire motifs in the upstream high-glycine (repetitive) region of the exon (Fig. 1). Compared to the Florida sequence, there are six motif-long gaps and three motif-long insertions in nearly all of the Mexican sequences, for a net difference of three motifs in length. Some sequences have additional gaps of one or more motifs (i.e., Nanciyaga N3-B5). All of these indels occur in the part of the exon that is upstream from the low-glycine spacer.

4. Discussion

The *Flag* gene is composed of an estimated 13 nearly identical exons interspersed with nearly identical introns [6]. The *Flag* gene section that we sequenced, the downstream intron and exon pair including a portion of the carboxyl end, appears to be more variable than most structural genes that have been repeatedly sequenced, without any indication of a loss of function of the protein that forms the core of the viscid spiral in the prey-capture web. Indeed, all the spiders we sampled were on normal-appearing orb webs and most were large juveniles and sexually mature females, indicating that the viscid silk functioned properly in prey capture. Although the frequency of polymerase error found through comparison of multiple sequences from the same allele is higher than is generally reported (10^{-5} for *Taq* polymerase and 10^{-6} for proofreading polymerases; reviewed in Ref. [17]), the unequal distributions across codon positions and the bias in amino acid substitution patterns support our conclusion that most of the observed variation reflects real genetic differences among these sequences and between the Mexican and Florida sequences.

The high polymerase error rate deduced by comparisons of sequences of the same clone may reflect the complex structure of the *Flag* gene, although the repeats themselves did not appear to increase error rate. We did not observe a significant increased snp frequency in the region coding for the XGPGG region of the exon compared to the regions of the gene coding for the spacer or coding for the carboxyl end. The consistency of insertion and deletion patterns across most of the Mexican sequences leads us to conclude that the process we used to contig the three separate sequences from each allele was reliable, and therefore the high error rate does not reflect misalignment during this step. It must be noted that there remains the possibility that these comparisons of what we presumed to be repeated, slightly different, sequences of the same allele could possibly reflect the existence of multiple copies of this gene in these individuals. We do not, at this time, have the ability to distinguish between sequencing error and multiple copies.

Biologically, there are two patterns of variation that appear particularly interesting in light of our understanding of the structure of the Flag protein. One of the most common types of amino acid substitution involved the proline residues (Fig. 1). The substitution of other amino acids for the proline in the XGPGG motif could disrupt the β-helix region of the protein [18,19], possibly altering the extraordinary extensibility of the protein [1]. The other interesting variable aspect of these sequences is the number of XGPGG motifs in the high-glycine section of the exon. The Mexican sequences are on average three motifs shorter than the Florida sequence. If the Hinman and Lewis model [18] is correct, this could result in a shorter "spring" in the protein synthesized by these spiders, particularly if this change is echoed in a majority of the 13 exons. Whether these changes have functional significance cannot be discerned without biomechanical studies, and differences (if any) may be below detection by current technology (Hayashi, pers. commun.).

Hayashi and Lewis [6] postulated that the high degree of similarity among the introns and exons within their *Flag* sequences might reflect gene conversion. Their model of concerted silk gene evolution allows us to make two predictions from our sequences. First, the consistently lower number of motifs in the Mexican sequences suggests that the shorter "spring" may be functionally advantageous. Second, if Hayashi and Lewis [6] are correct about the process of gene conversion in *Flag* and other silk genes, the remaining exons in *Flag* among Mexico should present similarly shorter repeat sections (although the terminal exon may not be homogenized with the interior repeats).

By obtaining multiple sequences of *Flag* from different individuals in several populations, we observed a very high degree of allelic variation. The rarity of this observation may be in part an artifact of the logistical difficulties of sequencing silk genes, which has resulted in most publications and sequence analyses being based upon a single sequence for each gene from each species (i.e. [8,9]). Beckwitt et al. [20] and Tai et al. [10] have done the only other within-species comparisons we are aware of. Beckwitt and colleagues obtained three sequences of major ampulate spid2 gene from *N. clavipes* and, as in our results, each of the three sequences is unique, again suggesting high levels of heterozygosity. Tai and colleagues obtained multiple sequences of *MaSp* from several individual *N. pilipes*, and in each case they found two alleles (i.e., among the five sequences obtained from cDNA of one individual, they found four identical sequences and one distinct one; [10, Section 3.5]). Although they interpret these results as indicating multiple copies of this gene, their results are also consistent with these individuals being heterozygotes for a single gene. As in our study of *Flag*, Tai et al.'s *MaSp* sequences include a portion of the carboxyl terminus and are therefore unlikely to be different exons within the gene. Both of these groups found that the carboxyl end of the silk gene was relatively conserved, a finding that has been repeated with our study of *Flag*, among species, and among silk genes [11,21]. The conservation of the carboxyl end of the protein is possibly related to its function in maintaining protein solubility prior to extrusion [11].

There are relatively few comparable data sets of multiple within-species sequences of structural genes. Most studies repeatedly sequencing genes in a single species involve genes of medical interest. The genes most analogous to silk genes in structure are the collagen-like genes which, like silk genes, consist of repeated motifs of amino acids (GXY [22]). Unlike silk genes, the collagen-like proteins are internally expressed. Perhaps reflecting the internal function of these proteins, most allelic variation in the collagen-like genes strongly affects the function of the protein and is related to disease ([22], Boot-Handford, pers. commun.). A closer analogy may exist between silk genes and a collagen-like bacterial gene coding for a protein that is extruded and forms part of the spore capsule of *Bacillus anthracis*. Sylvestre et al. [23] found nine alleles with different numbers of the GXY repeating motif. Bacteria with distinct alleles produced proteins of different lengths, which Sylvestre et al. believe reflects adaptive differences among strains from different ecological backgrounds.

The high amount of allelic variation and presumed protein sequence variation may be tolerated in *Flag* because of the function of silk proteins: they are extruded, and have no known

internal function to the spider. After synthesis, silk proteins are maintained in a liquid state inside the silk glands and do not take final quaternary structure until spun. That silk is an extruded protein leads us to postulate that the high rate of non-silent snps at minimum reflects relaxed selection, and based upon this conclusion we would expect to find similarly high levels of amino acid substitution among sequences within and among populations for all of the silk genes. So long as the protein maintains its integrity, there may be little functional affect of occasional amino acid substitution. It is also possible that there is positive selection for genetic diversity. Nearly all the individuals we examined were heterozygotes, suggesting that there may be a functional advantage of heterozygosity in this gene. This is all the more remarkable given the very low levels of genetic diversity in two mitochondrial markers sequenced for these same populations (COH, NADH, Higgins, pers. obs.).

In contrast to patterns of amino acid substitution, the consistency of differences in motif number suggests that the addition and deletion of entire motifs may be under more stringent selection, although this could also reflect historical accident such as a bottle-neck event in one or both populations. Certainly, our initial hope that repeat number may be hyper-variable was not met, and this gene is not appropriate for population genetics work as a "mini satellite". However, the patterns of variation indicate that population-level work is a fruitful source for improved understanding of the evolution of these genes and, when biophysical studies are done, the possible adaptive significance of the differences in gene structure. Because the genotype of an individual and the resultant protein produced from the gene can both be obtained without destructive sampling, research can concurrently include genetic and structural studies, making this a particularly strong system for investigating the interaction between genetic variation and protein function.

Acknowledgements

This work was supported by NSF INT-0233440 (LEH) and Semarnat-CONACyT #0355 (JNF). Collecting in private properties was permitted by C. Rodriguez (Nanciyaga), the family of L. Forbes (Fortin de las Flores), and the manager of "Rancho La Esperanza" (Mateos de Romero). Logistical support was provided by the Insituto de Ecologia, UNAM, and laboratory space graciously provided by Charles Goodnight (UVM). The laboratory work was aided by Thomm Buttolph (UVM). Conversations with W. Kilpatrick and P. O'Grady at UVM were vital to interpreting these results, and R. Barrantes (Cell and Molecular Biology, UVM) was very helpful in testing our alignments and comparing across different alignment options. Comments from reviewers were very helpful in polishing the presentation of these data. Lastly, C. Hayashi first provided the impetus and positive controls for the PCR, has been encouraging and helpful throughout, and provided thoughtful comments on an earlier version of the manuscript.

Appendix A

Nucleotide alignment of the Mexican sequences against two GenBank sequences from Florida, one cDNA. The first refers to the site of collection (Table 1), the first number refers to the individual specimen and the second number (after the dash) refers to the sequence identifier for that individual. The initiation of the 13th exon, where the cDNA sequence starts, is indicated by a vertical line. Dots indicate no difference from the first sequence, dashes indicate a deletion or no data.

```
FF3-1      CTTGCAACCGCCTCATCGTCATTTCGTACATTTGCCTTTTTTCACACGATAGAGAAAGACTTTGAGAAATTAGCATTTTTGCTAGATTAGTACGAATGCT
FF3-4      ..........................................................................A.........C.................
Xal23-1    ..........................................................................A.........C.................
Xal23-2    ..........................................................................A.........C.................
LT7-4      ..........................................................................A.........C.................
LT9-3      ..........................................................................A.........C.................
LT9-4      ..........................................................................A.........C.................
LT10-2     ..........................................................................A.........C.................
LT10-4     ..........................................................................A.........C.................
Nan3-B5    ..........................................................................A.........C.................
Nan3-4     ..................................................A.......................A.........C.................
Mat8-1     ..........................................................C.........T.........C.................
Mat8-2     ..........................................................................GT........C.................
Mat19-3    ..........................................................................A.........C.................
Mat19-4    ..........................................................................A.........C.................
Mat12-1    ............G.............................................................A.........C.................
Mat12-B2   ..........................................................................A.........C.................
FL         ....................C.....................................................A.........C.................
FL cDNA    ----------------------------------------------------------------------------------------------------

FF3-1      TTTTTAGGAACGTTCATACTTCCATTTTCTTATATGGATAGAAATAGAGAAATAAATTCTAATTTTTGTTCGTTTCGAAACCGTTTGGAGTTGTAGTAAT
FF3-4      ....................................................................................................
Xal23-1    ....................................................................................................
Xal23-2    ....................................................................................................
LT7-4      ....................................................................................................
LT9-3      ....................................................................................................
LT9-4      ....................................................................................................
LT10-2     ....................................................................................................
LT10-4     ....................................................................................................
Nan3-B5    ....................................................................................................
Nan3-4     ....................................................................................................
Mat8-1    ..........................C.........................................................................
Mat8-2    ....................................................................................................
Mat19-3   ..........................................................................C.........................
Mat19-4   ..........................................................................C.........................
Mat12-1   ....................................................................................................
Mat12-B2  ....................................................................................................
FL         ....................................................................................................
FL cDNA    ----------------------------------------------------------------------------------------------------

FF3-1      ATGAAATTTCTTAATTATATATATGTTAAAATTTTTTTTAATTCTATCTCATATTGTAAACTGGGTAAATCTTTACAATTTTGGTCCCATTTAACGAAGA
FF3-4      ....................................................................................................
Xal23-1    ............................A.......................................................................
Xal23-2    ............................A.....................................................................G......
LT7-4      ....................................................................................................
LT9-3      ........................................T...........................................................
LT9-4      ............................A.....................................................C..............
LT10-2     ....................................................................................................
LT10-4     ..................................................C.................................................
Nan3-B5    ............................A.....................................................G..A............
Nan3-4     ............................A.......................................................................
Mat8-1     ....................................................................................................
Mat8-2     ....................................................................................................
Mat19-3    ............................A.......................................................................
Mat19-4    ............................A.......................................................................
Mat12-1    ............................A.......................................................................
Mat12-B2   ............................A.......................................................................
FL         ............................A.....................T.................................................
FL cDNA    ----------------------------------------------------------------------------------------------------

FF3-1      AACAAGTACTCTCTTAAATATTAAAAGGTAAAAAATAATTTTTGCATTGAAGAATGTGTGGTTCCAAGAATATATAACAATTTTAGTTACAGATCTGATC
FF3-4      ....................................................................................................
Xal23-1    ....................................................................................................
Xal23-2    ....................................................................................................
LT7-4      ....................................................................................................
```

Appendix A (*Continued*)

```
LT9-3     ..........................................................................................
LT9-4     ..............................C.....................C.....................................
LT10-2    ......C...................................................................................
LT10-4    ......C.....................................................C.............................
Nan3-B5   .......T.......G...........A.....................................................................
Nan3-4    ..........................................................................................
Mat8-1    ..........................................A.....................................................
Mat8-2    ..........................................A.....................................................
Mat19-3   ..........................................................................................
Mat19-4   ..........................................................................................
Mat12-1   ..................................................................G.........................
Mat12-B2  ..........................................................................................
FL        ..........................................................................................
FL cDNA   ------------------------------------------------------------------------------------------
```

```
FF3-1     AAACTTAGGTTATTGAGAGGTGCATGTAATTCGAAGAGGGATTCTTGTATATTTCAATAAACCGTGACATTTTTCCTCCTTTCTAAGCAATCGAATATTT
FF3-4     ..........................................................................................
Xal23-1   ..........................................................................................
Xal23-2   ..................................................C.....................................
LT7-4     ..........................................................................................
LT9-3     ..................................................A.....................................
LT9-4     ..........................................................................................
LT10-2    ..........................................................................................
LT10-4    ..........................................................................................
Nan3-B5   ..................................G.....................................................
Nan3-4    ..............C.....T.....................................................................
Mat8-1    ..........................................................................................
Mat8-2    ..........................................................................................
Mat19-3   ..........................................................................................
Mat19-4   ..........................................................................................
Mat12-1   ..........................................................................................
Mat12-B2  ..........................................................................................
FL        ..................................................T.....................................
FL cDNA   ------------------------------------------------------------------------------------------
```

```
FF3-1     CCTTCTCTCATGAGCGGGCTAATATAAGAAAGAAAAGAACTGTTCTTCTGCCAGTGTCTGAAAAAACCTAAATTCAATTCTATCGATGGAGTTGAATTGA
FF3-4     ..........................................................................................
Xal23-1   ..........................................................................................
Xal23-2   ............................T.............................................................
LT7-4     ..........................................................................................
LT9-3     ..........................................................................................
LT9-4     ..........................................................................................
LT10-2    ..........................................................................................
LT10-4    ..........................................................................................
Nan3-B5   ..........................................................................................
Nan3-4    ..........................................................................................
Mat8-1    ..........................................................................................
Mat8-2    ..................................................................G.....................
Mat19-3   ..........................................................................................
Mat19-4   ..........................................................................................
Mat12-1   ..........................................C.................................................
Mat12-B2  ..........................................................................................
FL        ..........................................................................................
FL cDNA   ------------------------------------------------------------------------------------------
```

```
FF3-1     AGGTATAATTCCAAATTTCCCCTTGGACGATTTCTTTTGCTTCACTCGATTGCCTAATGATTTGTCATTGCGCATATCTCTTGTTCTTCTGTGTTGAACC
FF3-4     ..........................................................................................
Xal23-1   ..........................................................................................
Xal23-2   ..........................................................................................
LT7-4     ..........................................................................................
LT9-3     ........................................C.................................................
LT9-4     ........................................G.....................................................?..
LT10-2    ..........................................................................................
LT10-4    ............................C.....................................................G.............
Nan3-B5   ..................................................................G..C.....................
Nan3-4    ..........................................................................................
Mat8-1    ..........................................................................................
Mat8-2    ..................................................................C.......................
Mat19-3   ..........................................................................................
Mat19-4   ..................................................................G.......................
Mat12-1   ..........................................................................................
Mat12-B2  ..........................................................................................
FL        ..........................................................................................
```

Appendix A (*Continued*)

```
FL cDNA     ----------------------------------------------------------------------------------------------------

FF3-1       TGAATTCGAATTTTCTTGGGTTTGC|AGGTGTAGGACCTGATGGAAGTGGACCTGGAGGTTATGGACCTGGTGGA-------------------------
FF3-4       ........................|...G.....................................................-------------------------
Xal23-1     ........................|.........................................................-------------------------
Xal23-2     ........................|.........................................................-------------------------
LT7-4       ........................|...G.....................................................-------------------------
LT9-3       ...........G............|...G.....................................................-------------------------
LT9-4       ........................|.........................................................-------------------------
LT10-2      ........................|...G.....................................................-------------------------
LT10-4      ........................|...G.....................................................-------------------------
Nan3-B5     ..............?.........|.........................................-------------------------------------
Nan3-4      ........................|.........................................................-------------------------
Mat8-1      ........................|...G.....................................................-------------------------
Mat8-2      ........................|...G.....................................................-------------------------
Mat19-3     ........................|.........................................................-------------------------
Mat19-4     ........................|.........................................................-------------------------
Mat12-1     ........................|.........................................................-------------------------
Mat12-B2    ........................|.............................G...........-------------------------
FL          ........................|............G.....................................AGTGGACCTGGAGGTTATGGACCTGG
FL cDNA     ------------------------|............g.....................................agtggacctggaggttatggacctgg

FF3-1       ----GCTGGACCTGGAGGTTACGGACCTGGTGGTTCTGGTCCAGGTGGATACGGACCCGGTGGT-----------------------------TCCGGA
FF3-4       ----.....................................................................-----------------------------......
Xal23-1     ----.....................................................................-----------------------------......
Xal23-2     ----.....................................................................-----------------------------......
LT7-4       ----.....................................................................-----------------------------......
LT9-3       ----.....................................G...................-----------------------------......
LT9-4       ----.................T...........AG....A..T..A..T........T......TCTGGTCCAGGTGGATACGGACCCGGTGGT......
LT10-2      ----.....................................................................-----------------------------......
LT10-4      ----.....................................................................-----------------------------......
Nan3-B5     ------------.............................................................-----------------------------......
Nan3-4      ----.....................................................................-----------------------------......
Mat8-1      ----.....................................................................-----------------------------......
Mat8-2      ----.....................................................................-----------------------------......
Mat19-3     ----.....................................................................-----------------------------......
Mat19-4     ----.....................................................................-----------------------------......
Mat12-1     ----.....................................................................-----------------------------......
Mat12-B2    ----.....................................................................-----------------------------......
FL          TGGA.....................................................TCTGGTCCAGGTGGATACGGACCCGGTGGT......
FL cDNA     tgga.....................................................tctggtccaggtggatacggacccggtggt......

FF3-1       CCAGGAGGATACGGACCTGGCGGTTCTGGACCTGGTGGATACGGACCTGGCGGTGCTGGACCTGGTGGATACGGATCTGCTGGATACGGACCTGGCGGTT
FF3-4       ...........................................................................................................
Xal23-1     ...........................................................................................................
Xal23-2     ...........................................................................................T.......
LT7-4       ...........................................................................................................
LT9-3       ...........................................................................................................
LT9-4       ...........................................................................................................
LT10-2      ...........................................................................................................
LT10-4      ...........................................................................C.......................
Nan3-B5     .....................................G.....................................................................
Nan3-4      ..................................K........G.....................................................................
Mat8-1      ...........................................................................................................
Mat8-2      ...................................................C.........G.....................................
Mat19-3     ...........................................................................................T....
Mat19-4     ...........................................................................................T....
Mat12-1     ...........................................................................................................
Mat12-B2    ............................C..............................................................T....
FL          ..........................................T.....................----------------.........
FL cDNA     ..........................................t.....................----------------.........

FF3-1       CTGGACCTGGTGGATACGGACCTGGCGGTTCTGGACCTGGTGGATACGGTCCTGGAGGTTCTGGACCTGGTGGTTATGGACCTGGTGGTACTGGACCTGG
FF3-4       ...................................................................................C.....................
Xal23-1     ...............................................................G...........................................
Xal23-2     ...........................................................................................................
LT7-4       ...........................................................................................................
LT9-3       ...........................................................................................................
LT9-4       ...........................................................................................................
LT10-2      ...........................................................................................................
LT10-4      ...........................................................................................................
Nan3-B5     ...........................................................................................................
```

Appendix A (*Continued*)

```
Nan3-4      ..........................................................................................
Mat8-1      ..........................................................................................
Mat8-2      ..........................................................................................
Mat19-3     ................A..A......................................................................
Mat19-4     ................A..A......................................................................
Mat12-1     ................G.........................................................................
Mat12-B2    ..................A.....................................................A...................
FL          ..................T.....A.............------------------------------.....................
FL cDNA     ..................t.....a.............------------------------------.....................

FF3-1       TGGTACTGGACCTGGTGGTTCTGGACCTGGCGGATACGGACCTGGTGGTTCTGGACCTGGCGGTTCTGGA---------------TCTGGTGGTTTCGGA
FF3-4       ..........................................................................CCTGGCGGTTCTGGA........A.A....
Xal23-1     ..........................................................................CCTGGCGGTTCTGGA........A.A....
Xal23-2     ..............A...........................................................CCTGGCGGTTCTGGA........A.A....
LT7-4       ..........................................................................CCTGGCGGTTCTGGA........A.A....
LT9-3       ..........................................................................CCTGGCGGTTCTGGA........A.A....
LT9-4       ..........................................................................CCTGGCGGTTCTGGA........A.A....
LT10-2      ..........................................................................CCTGGCGGTCCTGGA........A.A....
LT10-4      ..........................................................................CCTGGCGGTCCTGGA........A.A....
Nan3-B5     ..........................................................................---------------........A.A....
Nan3-4      ..........................................................................---------------........A.A....
Mat8-1      A.........................................................................CCTGGCGGTTCTGGA........A.A....
Mat8-2      ..........................................................................CCTGGCGGTTCTAGA........A.A....
Mat19-3     ..........................................................................---------------........A.A....
Mat19-4     ..........................................................................---------------........A.A....
Mat12-1     ..........................................................................CCTGGCGGTTCTGGA........A.A....
Mat12-B2    ..........................................................................---------------........A.A....
FL          ..........................................................................CCTGGCGGTTCTGGAC.......A.A....
FL cDNA     ..........................................................................cctggcggttctggac.......a.a....

FF3-1       CCTAGTGGTTCTGGACCTGGCGGATAC----------------------------GGTCCTGGCGGTTCTGGACCTGGTGGATACGGACCGGGTGGCT
FF3-4       ........................----------------------------.....................................
Xal23-1     ........................----------------------------.....................................
Xal23-2     ........................----------------------------.....................................
LT7-4       ........................----------------------------.....................................
LT9-3       ........................----------------------------.....................................
LT9-4       ........................----------------------------.....................................
LT10-2      ........................----------------------------.....G...............................
LT10-4      ........................----------------------------.....................................
Nan3-B5     ........................----------------------------.....................................
Nan3-4      ........................----------------------------.....................................
Mat8-1      ........................----------------------------.....................................
Mat8-2      ........................----------------------------.....................................
Mat19-3     ........................----------------------------.....................................
Mat19-4     ........................----------------------------.....................................
Mat12-1     ........................----------------------------.....................................
Mat12-B2    ........................----------------------------.....................................
FL          ..........G........T......GGACCTAGTGGTTCTGGACCTGGCGGATAC.......................G.............
FL cDNA     ..........g........t......ggacctagtggttctggacctggcggatac.....................................

FF3-1       CTGGAGCCGGTGGTGCTGGACCTGGTGGCGCTGGAGGAGCAGGCGGAGCAGGAGGTTCAGGTGGAGCAGGAGGTTCAGGTGGTGCAGGAGGTTCGGGTGG
FF3-4       ..........................................................................................A.....
Xal23-1     ..........................................................................................A.....
Xal23-2     ..........................................................................................G.....A.....
LT7-4       ..........................................................................................A.....
LT9-3       ..............................A...........................................................A.....
LT9-4       ..........................................................................................A.....
LT10-2      ..........................................................................................A.....
LT10-4      ..........................................................................................A.....
Nan3-B5     ..........................................................................................A.....
Nan3-4      ................T.........................................................................A.....
Mat8-1      ..........................................................................................A.....
Mat8-2      .....?....................................................................................A.....
Mat19-3     ..........................................................................................A.....
Mat19-4     ..........................................................................................A.....
Mat12-1     ..........................................................................................A.....
Mat12-B2    ..........................................................................................A.....
FL          ............A.............................................................................A.....
FL cDNA     ............a.............................................................................a.....

FF3-1       AGCAGGAGGTTCAGGTGGAGTAGGAGGATCCGGCGGTACAACAATCACCGAAGACTTGGATATCACAATTGATGGCGCAGATGGCCCGATAACGATTTCA
FF3-4       ..........................................................T...............................C...
```

Appendix A (*Continued*)

```
Xal23-1    ..........................................................T.....................................
Xal23-2    ..........................................................T.....?...............................
LT7-4      ..........................................................T.....................................
LT9-3      ..........................................................T.....................................
LT9-4      ..........................................................T.....................................
LT10-2     ..........................................................T.....................................
LT10-4     ..........................................................T.....................................C...
Nan3-B5    ..........................................................T.....................................
Nan3-4     ..................................................A.......T.....................................
Mat8-1     ..........................................................T.....................................C....
Mat8-2     ..........................................................T.....................................
Mat19-3    ..........................................................T.....................................
Mat19-4    ..........................................................T.....................................
Mat12-1    ..........................................................T.....................................
Mat12-B2   ..........................................................T.....................................
FL         ..........................................................T.....................................
FL cDNA    ..........................................................t.....................................

FF3-1      GAAGAATTAACAATTAGTGGTGCTGGAGGTTCTGGACCCGGTGGTGCTGGACCAGGTGGTGTAGGGCCTGGTGGCTCTGGACCAGGAGGTGTAGGACCTG
FF3-4      ..................................................................................................
Xal23-1    ..................................................................................................
Xal23-2    ..................................................................................................
LT7-4      ..................................................................................................
LT9-3      ............G.....................................................................................
LT9-4      ....T.............................................................................................
LT10-2     ..................................................................................................
LT10-4     ..................................................................................................
Nan3-B5    ..................................................................................................
Nan3-4     ..................................................................................................
Mat8-1     ..................................................................................................
Mat8-2     ..................................................................................................A.....
Mat19-3    ..................................................................................................
Mat19-4    ..................................................................................................
Mat12-1    ..................................................................................................
Mat12-B2   ............................T.....................................................................
FL         ..................................................................................................
FL cDNA    ..................................................................................................

FF3-1      GAGTCTCTGGACCAGGAGGCGTAGGACCTGGTGGTTCTGGACCAGGAGGCGTAGGTTCTGGTGGTTCTGGACCAGGAGGCGTAGGACCTGGTGGTTACGG
FF3-4      ..................................................................................................
Xal23-1    ..........................................................................T.........................
Xal23-2    ..................................................G...............................................
LT7-4      ..................................................................................................
LT9-3      ..................................................................................................
LT9-4      ..................................................................................................
LT10-2     ..................................................................................................
LT10-4     ..................................................................................................
Nan3-B5    ..................................................................................................
Nan3-4     ..................................................................................................C...
Mat8-1     ..................................................................................................
Mat8-2     ..................................................................................................
Mat19-3    ..................................................................................................
Mat19-4    ..................................................................................................
Mat12-1    ..................................................................................................
Mat12-B2   ..................................................................................................
FL         ..................................................................................................
FL cDNA    ..................................................................................................

FF3-1      ACCTGGAGGTTCTGGATCAGGAGGCGTAGGACCTGGTGGTTACGGACCTGGAGGTTCAGGAGGATTTTACGGACCTGGAGGTTCAGAAGGACCTTATGGA
FF3-4      ..................................................................................................
Xal23-1    ..................................................................................................
Xal23-2    ..................................................................................................
LT7-4      ..................................................................................................
LT9-3      ..............................................................................................A.....
LT9-4      ..................................................................................................
LT10-2     ..................................................................................................
LT10-4     ..A...............................................................................................
Nan3-B5    ............................G.................................G.....................................
Nan3-4     ..........C.........................................................................................
Mat8-1     ..................................................................................................G
Mat8-2     ..................................................................................................
Mat19-3    ..................................................................................................
Mat19-4    ..................................................................................................
```

Appendix A (*Continued*)

```
Mat12-1     ...............................................................................................
Mat12-B2    ..........C....................................................................................
FL          ...............................................................................................
FL cDNA     ...............................................................................................

FF3-1       CCTAGTGGAACTTATGGTTCTGGAGGAGGATATGGTCCTGGTGGTGCTGGAGGACCATATGGACCTGGAAGTCCTGGAGGAGCTTATGGACCTGGAAGCC
FF3-4       ..............................................................................A........G.............
Xal23-1     ...............................................................................................
Xal23-2     ...............................................................................................
LT7-4       ...............................................................................................
LT9-3       ...............................................................................................
LT9-4       ...............................................................................................
LT10-2      ...................................................A...........................................
LT10-4      ...............................................................................................
Nan3-B5     ...............................................................................................
Nan3-4      ...................................................................................C...........
Mat8-1      ...........................................................A...................................
Mat8-2      ..................................................................T............................
Mat19-3     ...............................................................................................
Mat19-4     ...............................................................................................
Mat12-1     ...............................................................................................
Mat12-B2    ...............................................................................................
FL          ..........C....................................................................................
FL cDNA     ...............................................................................................

FF3-1       CTGGAGGAGCTTATTATCCTAGCTCGCGTGTTCCCGATATGGTGAATGGTATAATGAGTGCTATGCAAGGATCTGGTTTTAACTACCAAATGTTTGGTAA
FF3-4       ...............................................................................................
Xal23-1     ...............................................................................................
Xal23-2     ...............................................................................................
LT7-4       ...............................................................................................
LT9-3       ...............................................................................................
LT9-4       ...................................................A...........................................
LT10-2      ...............................................................................................
LT10-4      ...............................................................................................
Nan3-B5     ..................G.........................A...................................................
Nan3-4      ...............................................................................................
Mat8-1      ...............................................................................................
Mat8-2      ...............................................................C...............................
Mat19-3     ...............................................................................................
Mat19-4     ...............................................................................................
Mat12-1     ...............................................................................................
Mat12-B2    ...............................................................................................
FL          ...............................................................................................
FL cDNA     ...............................................................................................

FF3-1       TATGCTATCACAATATTCGTCTGGTTCAGGAACATGCAATCCAAATAATGTTAATGTTTTGATGGATGCTTTGTTAGCTGCTTTGCACTGTCTAAGTAAC
FF3-4       ...............................................................................................
Xal23-1     ..C............................................................................................
Xal23-2     ...............................................................................................
LT7-4       ...............................................................................................
LT9-3       ...............................................................................................
LT9-4       ...............................................................................................
LT10-2      ...............................................................................................
LT10-4      ........................G......................................................................
Nan3-B5     ...............................................................................................
Nan3-4      ...............................................................................................
Mat8-1      ...............................................................................................
Mat8-2      ...............................................................................................
Mat19-3     ...............................................................................................
Mat19-4     ...............................................................................................
Mat12-1     ...............................................................................................
Mat12-B2    ...............................................................................................
FL          ...............................................................................................
FL cDNA     ...............................................................................................

FF3-1       CACGGATCATCATCTTTTGCACCTTCTCCAACTCCGGCTGCTATGAGTGCGTATTCTAATTCTGTAGGAAGAATGTTCGC---
FF3-4       ...............................................................................----
Xal23-1     ......C.....................................................T.........................---
Xal23-2     ...............................................................................---
LT7-4       ...............................................................................---
LT9-3       ..............................................................----------------------
LT9-4       ...............................................................................----
```

Appendix A (*Continued*)

```
LT10-2     ...............................................................................---
LT10-4     ...............................................................................---
Nan3-B5    ................A..............................................................---
Nan3-4     ...............................................................................---
Mat8-1     ...............................................................................----
Mat8-2     ...............................................................................---
Mat19-3    ...............................................................................---
Mat19-4    ...............................................................................---
Mat12-1    ...............................................................................---
Mat12-B2   ..........................................G....................................---
FL         ..........................................................-----------------------
FL cDNA    ...............................................................................---
```

Appendix B

Protein sequence alignment of the Mexican sequences against two GenBank sequences from Florida. Dots indicate no difference from the first sequence, dashes indicate no data. Specimen identifiers as in Fig. 1 and Appendix A.

```
FF3-1     GVGPDGSGPGGYGPGG----------AGPGGYGPGGSGPGGYGPGG----------
FF3-4     SGPGGYGPGGSGPGGYGP..............----------..................----------
Xal23-1   ..................................----------..................----------
Xal23-2   ..................................----------..................----------
LT7-4     ..................................----------..................----------
LT9-3     ..................................----------..................----------
LT9-4     ..................................----------
LT10-2    ..........A.........SGPGGYGPGG.................................----------
LT10-4    ....................----------.................................----------
Nan3-B5   ....................----------.................................----------
Nan3-4    ....................----------.............C..................----------
Mat8-1    ....................----------.................................----------
Mat8-2    ....................----------.................................----------
Mat19-3   ....................----------.................................----------
Mat19-4   ....................----------.................................----------
Mat12-1   ....................----------.................................----------
Mat12-B2  ....................----------.................................----------
FLFL      ....................----------
cDNA      ......................G...........SGPGGYGPGG.....................SGPGGYGPGG...
          ..............S....G...........SGPGGYGPGG.....................SGPGGYGPGG......
          ...........S

FF3-1     GGAGPGGYGSAGYGPGGSGPGGYGPGGSGPGGYGPGGSGPGGYGPGGTGPGGTGPGGSGPGGYGPGGSGPGGSG...
FF3-4     ..............................................................................
Xal23-1   ..............................................................................
Xal23-2   .....L.....................................S..................................
LT7-4     ..............................................................................
LT9-3     ..............................................................................
LT9-4     ..............................................................................
LT10-2    ..............................................................................
LT10-4    ..............................................................................
Nan3-B5   ..............................................................................
Nan3-4    ..............................................................................
Mat8-1    ..................................................A.............................
Mat8-2    ..........................................................ST..................
Mat19-3   ..................................................................ST..........
Mat19-4   ..............................................................................
Mat12-1   ..........................................................T..............N.....
Mat12-B2  ....................................-----..............----------
FLFL      ..........................................-----..............----------
cDNA      ................................

FF3-1     -----SGGFGPSGSGPGGY----------
FF3-4     GPGGSGPGGYGPGGSGAGGAGPGGAGGAGGAGGSGGAGGSGGAGGPGGSG...Y.........----------
Xal23-1   ..........................................PGGSG...Y.........----------
Xal23-2   ..........................................PGGSG...Y.........----------
LT7-4     ..........................................PGGSG...Y.........----------
LT9-3     ..........................................PGGSG...Y.........----------
LT9-4     ......................S...................PGGSG...Y.........----------
LT10-2    ..........................................PGGSG...Y.........----------
LT10-4    ..........................................PGGSG...Y.........----------
Nan3-B5   ..................................------...Y.........----------
Nan3-4    ..................................------...Y.........----------
Mat8-1    ..................W.....................PGGSG...Y.........----------
Mat8-2    ..........................................PGGSS...Y.........----------
Mat19-3   ..............?.....................------...Y.........----------
Mat19-4   ..................................------...Y.........----------
Mat12-1   ..........................................PGGSG...Y.........----------
Mat12-B2  ..................................-----...Y.........----------
FLFL      ..........................................PGGSGP..Y.........GPSGSGPGGY...
cDNA      ..............T.....................PGGSGP..Y.........GPSGSGPGGY......
          ............T.....................

FF3-1     SGGAGGSGGVGGSGGTTITEDLDITIDGADGPMatISEELTISGAGGSGPGGAGPGGVGPGGSGPGGVGPGVSGP..
```

Appendix B (*Continued*)

```
FF3-4     ..........................................................................
Xal23-1   ..........................................................................
Xal23-2   ..................I?M.....................................................
LT7-4     ..........................................................................
LT9-3     ..........................................................................
LT9-4     ........................V.................................................
LT10-2    ..........................................................................
LT10-4    ..........................................................................
Nan3-B5   ..........................................................................
Nan3-4    .M........................................................................
Mat8-1    ............T.............................................................
Mat8-2    ........................................E.................................
Mat19-3   ..........................................................................
Mat19-4   ..........................................................................
Mat12-1   ..........................................................................
Mat12-B2  ............V.............................................................
FLFL      ..........................................................................
cDNA      ....................................
```

```
FF3-1     GGVGPGGSGPGGVGSGGSGPGGVGPGGYGPGGSGSGGVGPGGYGPGGSGGFYGPGGSEGPYGPSGTYGSGGGYG...
FF3-4     ..........................................................................
Xal23-1   ..........................................................................
Xal23-2   ..........................................................................
LT7-4     ..........................................................................
LT9-3     ....................................S.....................................
LT9-4     ..........................................................................
LT10-2    ..........................................................................
LT10-4    .....H....................................................................
Nan3-B5   ..........................................................................S..
Nan3-4    ..P.......................................................................
Mat8-1    ..........................................................................
Mat8-2    ..........................................................................
Mat19-3   ..........................................................................
Mat19-4   ..........................................................................
Mat12-1   ..........................................................P.................
Mat12-B2  ..........................................................................
FLFL      ............P.............................................................
cDNA      ..................
```

```
FF3-1     PGGAGGPYGPGSPGGAYGPGSPGGAYYPSSRVPDMVNGMMSAMQGSGFNYQMFGNMLSQYSSGSGTCNPNNVNV...
FF3-4     ..........E..C.....................................................................
Xal23-1   ..........................................T.........................................
Xal23-2   ..........................................................................
LT7-4     ..........................................................................
LT9-3     ..........................................................................
LT9-4     ..........................................................S.................
LT10-2    ..........................................................................
LT10-4    ..........................................................................
Nan3-B5   .G........D.....................................................H.............
Nan3-4    ..................................................T.........................
Mat8-1    ..................................................L.........................
Mat8-2    ..........................................................................
Mat19-3   ..........................................................................
Mat19-4   ..........................................................................
Mat12-1   ..........................................................................
Mat12-B2  ..........................................................................
FLFL      ..........................................................................
cDNA      ..................
```

```
FF3-1     LMDALLAALHCLSNHGSSSFAPSPTPAAMSAYSNSVGSMF..................................
FF3-4     ..................P.............F.....................................
Xal23-1   ..........................................................................
Xal23-2   ....?----
LT7-4     ..........................................................................
LT9-3     ..........................................................Y...............
LT9-4     ..........................................................................
LT10-2    ..........................................................................
```

Appendix B (*Continued*)

LT10-4	.....
Nan3-B5	.....G.....
Nan3-4	.....------.....
Mat8-1	
Mat8-2	
Mat19-3	
Mat19-4	
Mat12-1	
Mat12-B2	
FLFL	
cDNA	

References

[1] J.M. Gosline, M.E. DeMont, M.W. Denny, Endeavour 10 (1986) 37–42.
[2] P.A. Guerette, D.G. Ginzinger, B.H.F. Weber, J.M. Gosline, Science 272 (1996) 112–115.
[3] C. Craig, Spiderwebs and Silk: Tracing Evolution from Molecules to Genes to Phenotypes., Oxford University Press, Oxford, 2003, 230 pp.
[4] J. Gatsey, C. Hayashi, D. Motriuk, J. Woos, R. Lewis, Science 291 (2001) 2603–2605.
[5] C.Y. Hayashi, R.V. Lewis, J. Mol. Biol. 275 (1998) 773–784.
[6] C.Y. Hayashi, R.V. Lewis, Science 287 (2000) 1477–1479.
[7] J. Coddington, H.W. Levi, Ann. Rev. Ecol. Syst. 22 (1991) 565–592.
[8] H. Sezutsu, K. Yukuhiro, J. Mol. Evol. 51 (2000) 329–338.
[9] R. Beckwitt, S. Arcidiacono, J. Biol. Chem. 269 (1994) 6661–6663.
[10] Tai, et al., Int. J. Biol. Macromol 34 (5) (2004) 237–243.
[11] E. Bini, D.P. Knigh, D.L. Kaplan, J. Mol. Biol. 335 (2004) 27–40.
[12] P. Sunnucks, D.F. Hales, Mol. Biol. Evol. 13 (1996) 510–524.
[13] D.R. Maddison, W.P. Maddison, MacClade 4 (release version 4.06 for OS X), Sinauer Associates Inc., Sunderland, MA, 2003.
[14] R.C. Edgar, Nucl. Acids Res. 32 (2004) 1792–1797.
[15] C. Notredame, D.G. Higgins, J. Heringa, J. Mol. Biol. 302 (2000) 205–217.
[16] C.C. Shih, D.T. Lee, L. Lin, C.L. Peng, S.H. Chen, Y.W. Wu, C.Y. Wong, M.Y. Chou, T.C. Shiao, M.F. Hsieh, BMC Bioinform. 7 (2006) 103.
[17] J. Smith, P. Modrich, Proc. Natl. Acad. Sci. 94 (1997) 6847–6850.
[18] M.B. Hinman, R.V. Lewis, J. Biol. Chem. 267 (1992) 19320–19324.
[19] C.Y. Hayashi, N.H. Shipley, R.V. Lewis, Int. J. Biol. Macromol. 24 (1999) 271–275.
[20] R. Beckwitt, S. Arcidiacono, R. Stote, Insect Biochem. Mol. Biol. 28 (1998) 121–130.
[21] C.Y. Hayashi, T.A. Blackledge, R.V. Lewis, Mol. Biol. Evol. 21 (2004) 1950–1959.
[22] R.P. Boot-Handford, D.S. Tuckwell, Bioessays 25 (2003) 142–151.
[23] P. Sylvestre, E. Couture-Tosi, M. Mock, J. Bacteriol. 185 (2003) 1555–1563.