ELSEVIER

Gene 223 (1998) 347–354

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Structure and polymorphism of the *Chironomus thummi* gene encoding special lobe-specific silk protein, ssp160[1]

Eugene Berezikov [a], Alexander G. Blinov [a], Svetlana Scherbik [a], Carol K. Cox [b], Steven T. Case [b,*]

[a] *Laboratory of Cell Biology, Institute of Cytology and Genetics, 630090 Novosibirsk, Russia*
[b] *Department of Biochemistry, The University of Mississippi Medical Center, Jackson, MS 39216-4505, USA*

Received 8 December 1997; received in revised form 6 March 1998; accepted 6 March 1998; Received by J. Wild

Abstract

cDNA encoding *Chironomus thummi* ssp160 was used to isolate a genomic clone that hybridized in situ to band A2b on polytene chromosome IV, the site of the *ssp160* gene. DNA sequencing, primer extension and gene/cDNA nucleotide sequence alignment revealed the gene contains six exons and five introns; 70% of ssp160 is encoded in exon 3. Variations between cDNA and gene sequences led to the design of a polymerase chain reaction, restriction fragment length polymorphism assay that was subsequently used to demonstrate the existence of polymorphic alleles whose distribution varied between geographically separated populations of larvae. The polymorphism is associated with codon deletions in a six-amino-acid repeat containing an N-linked glycosylation motif. These deletions may have resulted from slipped-strand mispairing during DNA replication.

Keywords: In-situ hybridization; Transcription start point; Introns; Exons; Restriction fragment length polymorphism; Nucleotide sequence variants; Repeated glycosylation motifs

1. Introduction

Multigene families are common among eukaryotes. Structural comparison of family members is often instructive for recognizing putative *cis*-acting sequences required for transcriptional (Travers, 1996; Kamakaka, 1997; Smale, 1997) and post-transcriptional (Senapathy et al., 1990) regulation of gene expression, protein domains partitioned in exons (Dorit and Gilbert, 1991) and mechanisms that influenced the evolutionary history of the family (Ohta, 1991; Meisler and Ting, 1993; Fryxell, 1996).

The midge, *Chironomus tentans*, contains a multigene family that encodes secreted silk proteins that can be divided into small (<100 kDa)-size, intermediate (100–240 kDa)-size and large (about 1000 kDa)-size classes (Case and Wieslander, 1992). Genes from all three classes have been sequenced and exhibit developmental stage (larval)- and tissue (salivary gland)-specific expression (Wieslander, 1994). Most intermediate- and large-size proteins contain a major central domain composed of an extensive array of tandemly repeated amino acid (aa) sequences encoded in one exon separate from smaller terminal exons that encode non-repeated sequences. These genes appear to have evolved by duplication and divergence of ancestral sequences with Cys- and/or Pro-containing motifs found in genes encoding the small-size class proteins. Modern-day repeats exhibit sequence homogenization as well as expansion and contraction presumably due to unequal crossing over (Wieslander, 1994). The gene encoding sp185 is an exception; it contains numerous exons encoding Cys-containing motifs whose spacing (22–26 aa) is more

* Corresponding author. Tel: +1 601 9841518; Fax: +1 601 9841501; e-mail: stcase@fiona.umsmed.edu

[1] Published in conjunction with A Wisconsin Gathering Honoring Waclaw Szybalski on the occasion of his 75th year and 20 years of Editorship-in-Chief of *Gene*, 10–11 August 1997, University of Wisconsin, Madison, WI, USA.

Abbreviations: aa, amino acid(s); *B.*, *Bombyx*; bp, base pair(s); *C.*, *Chironomus*; *D.*, *Drosophila*; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism; ssp160, special lobe-specific silk protein with an apparent molecular mass of 160 kDa; *tsp*, transcription start point(s); *UTR*, untranslated region; X, any aa.

0378-1119/98/$ – see front matter
PII: S0378-1119(98)00165-6

conserved than the surrounding sequences (Paulsson et al., 1990; Case et al., 1997).

C. thummi larvae synthesize one silk protein that *C. tentans* larvae fail to make; ssp160 (special lobe-specific silk protein with an apparent molecular mass of 160 kDa) (Kolesnikov et al., 1981; Kiknadze et al., 1990; Hoffman et al., 1996). Three features make ssp160 novel among *Chironomus* silk proteins. First, half of its apparent mass may be due to extensive *N*- and *O*-linked glycosylation. Second, though three regions of repeated *N*-linked glycosylation motifs occur, it lacks the extensive arrays of tandem repeats and Cys- and Pro-containing motifs that typify other intermediate- and large-size silk proteins. Third, ssp160 synthesis is limited to four cells that comprise the special lobe of the salivary gland. Whereas the *ssp160* gene promoter is expected to share a subset of *cis*-acting regulatory sites found in other silk protein-encoding genes (i.e., sites for stage- and tissue-specific transcription by RNA polymerase II), it presumably contains additional sites that convey special lobe-specific regulation of transcription.

As a first step toward understanding the structure and transcriptional regulation of the *C. thummi ssp160* gene and providing a molecular explanation as to why *C. tentans* fails to synthesize an ssp160 homolog, whereas its close relative, *C. pallidivittatus*, does (Beermann, 1961), we report here its nucleotide (nt) sequence and exon/intron organization. During the course of this study, we discovered that geographically separated larval populations exhibit a RFLP within one exon. This RFLP occurs near a dynamic site whose evolutionary history includes codon deletion.

2. Materials and methods

2.1. Libraries and clones

Routine molecular biological techniques were used unless otherwise noted. Dr. R.T. Hoffman constructed a λgt22 cDNA library made from salivary gland poly(A)$^+$RNA extracted from a Siberian strain of *C. thummi* and isolated four *ssp160* clones using a special lobe-specific subtracted hybridization probe (Hoffman, 1995). The sequence of near-full-length cDNA in λ160.1 was reported (Hoffman et al., 1996); λ160.2–160.4 had similar size inserts and restriction patterns. Professor J. Niessing provided a λEMBL4 genomic DNA library made from a German strain of *C. thummi* by partial cleavage with *Sau*3A. The *ssp160* gene was isolated by plaque hybridization with *ssp160* cDNA amplified from λ160.1 by polymerase chain reaction (PCR). Two clones were selected and characterized by Southern blotting. λC2 had a 17.6-kb insert; hybridization of *ssp160* cDNA was limited to a 7.6-kb *Eco*RI fragment that was subcloned into pUC19, creating plasmid pC2A.

2.2. Hybridization

Plaque lifts and Southern blots were hybridized with amplified cDNA radiolabeled by nick-translation using [α-^{32}P]dATP. Squashed preparations of larval salivary gland polytene chromosomes were hybridized with pC2A labeled with biotin-14-dATP (BioNick™ Labeling System, GIBCO-BRL) and detected with streptavidin-alkaline phosphatase conjugate (In Situ Hybridization and Detection System, GIBCO-BRL).

2.3. DNA sequencing

Southern blotting indicated that *ssp160* cDNA hybridized the distal 3.6-kb *Hin*dIII–*Eco*RI fragment of the pC2A. Smaller restriction fragments (*Bam*HI, *Hin*dIII, *Sph*I, *Taq*I, *Sca*I, *Eco*RI, *Eco*RV and *Nsi*I) were subcloned into pBluescript and pUC129 and sequenced (dsDNA Cycle Sequencing System, GIBCO-BRL). The final sequence was compiled from overlapping sequences from both strands of subcloned restriction fragments. The *ssp160* gene was located, and the exon/intron structure was deduced by alignment with the cDNA sequence using the program MACAW (Schuler et al., 1991). The transcription start point (*tsp*) was mapped by reverse transcriptase-dependent extension of end-labeled primer, 160.RT (5′-CAAGGCGCACACTAAGATCACT, the reverse complement of gene nt 246–267), annealed to salivary gland RNA (Paulsson et al., 1990). This product was run in parallel to a DNA sequencing ladder produced with the same primer and pC2A. Direct comparison of a polymorphic region of the cDNA and gene was made using an infra-red-labeled primer, 160.MUT (5′-TACGGACGCGCTTCTTCATTGATC, reverse complement of gene nt 1807–1830), in parallel chain-termination cycle sequencing reactions using SequiTherm DNA polymerase (Epicenter Technologies). Products were resolved on a LI-COR Model 4000L OL automated DNA sequencer (Middendorf et al., 1992). This same primer was used to sequence one strand of genomic PCR products.

2.4. PCR

Reactions were carried out in 100 μl as described (Hoffman, 1995) except for the use of 250 μM each deoxyribonucleoside triphosphate, either 1 ng (phage or plasmid) or 200–400 ng (genomic) template DNA and 1 μM each primer (λ160.1-Forward, 5′-TCATCTC-CCTCTTGTTCTCC, gene nt 644–663; λ160.1-Reverse, 5′-TGGAACTGTTACGGATCC, the reverse complement of nt 3060–3077). These primers amplified an 1886-bp fragment of *ssp160* cDNA from λ160.1 (templates λ160.2–4 yielded a similar size fragment) and a 2434-bp fragment of the *ssp160* gene from pC2A (the

larger size of this fragment compared to cDNA is due to introns). Genomic DNAs yielded a product similar in size to pC2A. PCR products were purified and subjected to agarose gel electrophoresis with and without cleavage by *Hin*dIII and cycle sequencing with primer 160.MUT.

3. Results and discussion

3.1. Clone selection

PCR-amplified cDNA from λ160.1 (Hoffman et al., 1996) was used as a plaque hybridization probe to screen a λEMBL4 library of *C. thummi* genomic DNA. A 7.6-kb *Eco*RI fragment in clone λC2 hybridized the cDNA and was subcloned into pUC19 creating plasmid, pC2A. pC2A hybridized in situ to a single site, band A2b, on salivary gland polytene chromosome IV (Fig. 1). Since this band is the site of hybridization *ssp160* cDNA (Hoffman et al., 1996) and forms special lobe-specific BRa (Kiknadze et al., 1990), we concluded that the *ssp160* gene was within this clone and proceeded to sequence part of it.

3.2. Gene structure

The distal half of the pC2A insert (a 3.6-kb *Hin*dIII–*Eco*RI fragment) was sequenced and aligned with *ssp160* cDNA (Fig. 2). This revealed that the *ssp160* gene consists of six exons separated by five introns demarcated by dinucleotides (5′-GT....AG-3′) typical of 5′- and 3′-splice junctions (Senapathy et al., 1990).

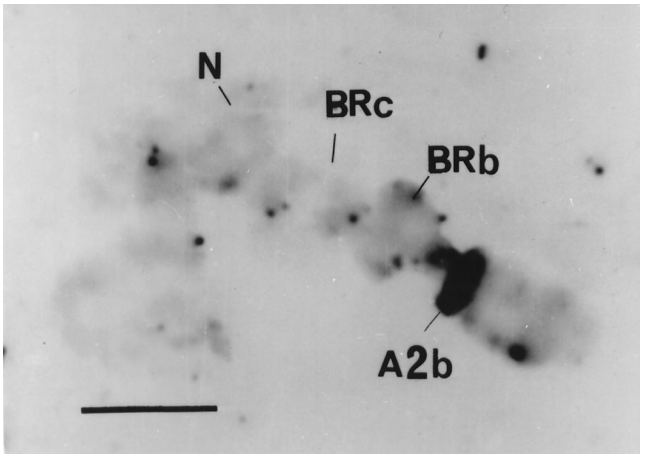

Fig. 1. Hybridization in situ of biotinylated pC2A to salivary gland polytene chromosome IV. Photomicrograph of brightfield image showing detection of streptavidin-alkaline phosphatase conjugate (see Section 2.2 for details) on distal side of the constriction that occurs on the telomeric side of BRb. Phase contrast images of chromosomes with suitable morphology (data not shown) confirm that this site is band A2b. The positions of this band (A2b), Balbiani rings (BRb, BRc) and a nucleolus (N) are indicated. Bar = 15 μm.

The *tsp* was mapped by primer extension of mRNA (Fig. 3). Reverse transcription using a cDNA-specific 22-nt primer produced one minor and three major products that were 39, 66, 67 and 70 nt in length. No attempt was made to characterize these products; however, they imply that reverse transcriptase in vitro may often terminate 39 nt short of the 5′-end of mRNA transcribed by RNA polymerase II in vivo using several closely spaced *tsp*. In the absence of an assay for *Chironomus* promoters, we used the size of the largest abundant extension product to assign the *tsp* to nt 198 (Fig. 2). This extends the 5′-end of previously reported near full-length cDNA sequence (Hoffman et al., 1996) by 21 nt.

The sequence upstream of the *ssp160* gene was compared to promoters for genes encoding other salivary and silk gland proteins from *C. tentans*, *B. mori*, and *D. melanogaster*, including two *ssp160* alleles from *C. pallidivittatus* (S.T.C., in preparation). Within 130 bp upstream of the *tsp*, all three *ssp160* promoters share, at identical positions, one imperfect copy of a TATA- (nt 169) and CAAT-box (nt 84) box and an SGF-1 (nt 95) and SGFB (nt 69) binding site. The latter two sites are required for tissue-specific transcription in *B. mori* silk (Grzelak, 1995; Nony et al., 1995) and *D. melanogaster* salivary (Mach et al., 1996) glands; however, the significance of these, or any other, sequences with regard to special lobe-specific transcription in *Chironomus* salivary glands awaits a suitable promoter assay.

Exon 1 contains a 37-bp 5′-untranslated region (*UTR*), the translation start codon and an 18-aa hydrophobic leader sequence that targets this protein for secretion into the endoplasmic reticulum. Exons 2, 3 and 4 are exclusively protein-coding, and 70% of ssp160 is encoded in exon 3. Exon 5 contains a 128-bp 3′-*UTR* and several imperfect putative recognition sequences for cleavage/polyadenylation (Sheets et al., 1990; Wahle and Keller, 1992; Wahle and Kuhn, 1997). Whereas ssp160 has no overt sequence similarity to other *Chironomus* silk proteins, the exon/intron organization of its gene is most similar to those encoding extensive arrays of tandem repeats (Wieslander, 1994).

3.3. Gene polymorphism

Closer inspection revealed a number of striking differences between the gene and cDNA sequences (Fig. 2). There are 38 nt substitutions, five of which are in the 3′-*UTR*; however, 33 are in protein-coding sequences, and 29 of these encode synonymous codons. In addition, the cDNA appears to lack an 18-bp in-frame sequence found in the gene (nt 1464–1482). Sequence differences were confirmed by reinspection of original

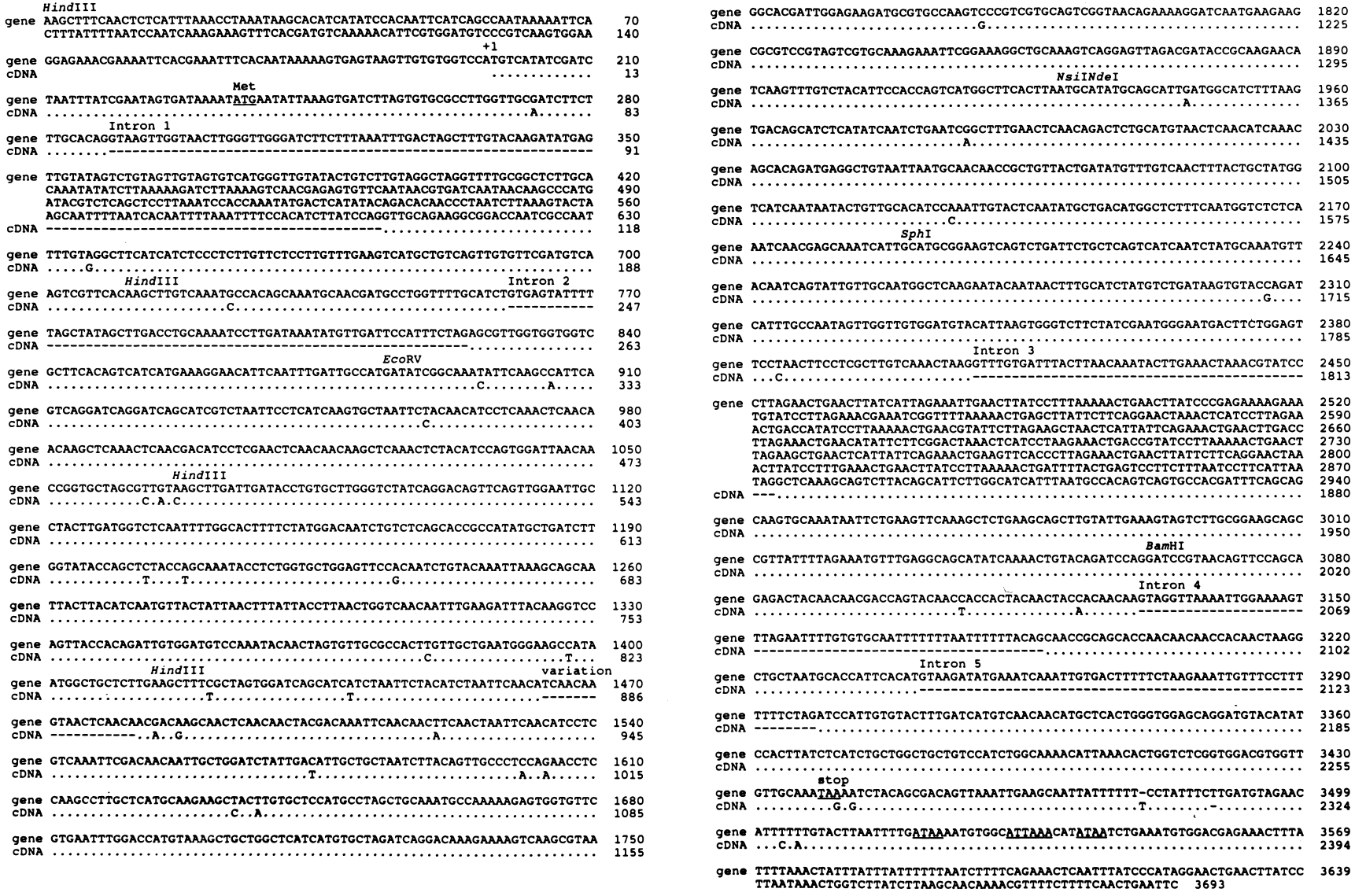

Fig. 2. Comparison of gene and cDNA nt sequences encoding *C. thummi* ssp160. The nt sequence reported here (GenBank Accession No. AF036895) was aligned with cDNA [No. U24265 minus 14 3′-end-A's presumably derived from the poly(A) tail] whose *tsp* was extended 21 nt upstream (Fig. 3). Base substitutions (A, T, G and C), identities (.) and voids (–) in cDNA are indicated below the gene nt sequence. Labels above the sequence indicate the location of restriction sites used to subclone fragments for DNA sequencing, the *tsp* (+1), start (Met) and termination (stop) codons for protein synthesis, introns and the 18-bp sequence (variation) that is present in the genomic clone but absent from the cDNA clone.Putative sites for cleavage/polyadenylation are underlined.

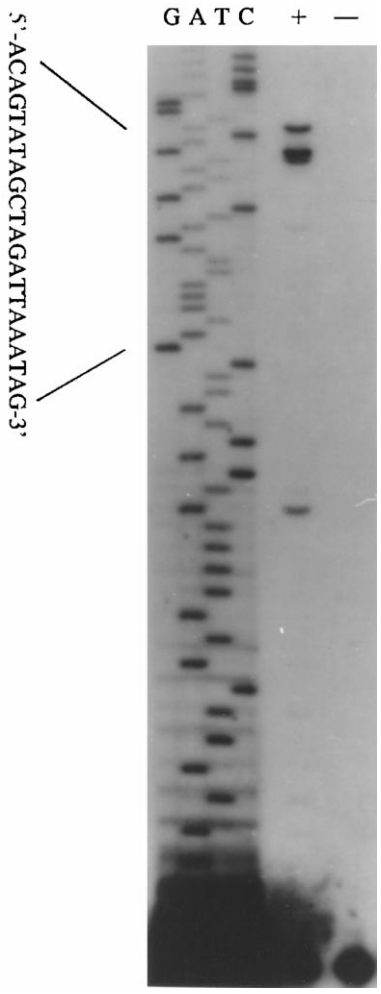

Fig. 3. Mapping the *tsp* by primer extension of *ssp160* mRNA. Autoradiogram of denaturing 6% polyacrylamide gel displaying products obtained when primer 160.RT was used in parallel DNA sequencing reactions with pC2A (G,A,T, and C) and reverse transcription reactions with (+) and without (−) 20 μg of salivary gland RNA (the reaction lacking salivary gland RNA contained 20 μg of tRNA). See Section 2.3. The sequence (left) shows the resulting 21-nt extension of previously reported near full-length cDNA (Hoffman et al., 1996) to the *tsp*.

data and simultaneous parallel sequencing of cDNA and genomic clones in the region of the 18-bp variation.

These sequence variations implied that either the genome contained two similar yet different copies of the *ssp160* gene or the gene was polymorphic. To distinguish between these possibilities, the following PCR/RFLP assay was designed based on the initial assumption that the absence of the 18 bp (nt 1464–1482) in cDNA coincides with the A→C substitution that eliminates the *Hin*dIII site at nt 1067 (Fig. 2). Two opposing primers were used to amplify a 2.4-kb segment of the gene by PCR (Fig. 4). A gene similar to the sequence reported for pC2A (variant 1) would yield a 2434-bp PCR product which, when cleaved with *Hin*dIII, would result in four fragments of (5′→3′): 68, 356, 347 and 1663 bp. A gene whose sequence matched the cDNA (variant 2) would yield a 2416-bp product, which *Hin*dIII would cleave into three fragments: 68, 703 (356 and 347-bp bands combined due to mutated *Hin*dIII site) and 1645 bp (this band is smaller due to the 18-bp deletion). Similarly, correcting for the lack of 65-bp intron 2 in the 356- and 703-bp bands and lack of 465-bp intron 3 in the 1663- and 1645-bp bands (see Fig. 4), cDNA from variant 1 would yield four bands (68, 291, 347 and 1198 bp), whereas cDNA from variant 2 would yield three (68, 638 and 1180 bp).

Using this PCR/RFLP assay, pC2A gave the four-band pattern predicted for variant 1, whereas four independently isolated cDNA clones (λ160.1–4) all gave the three-band pattern predicted for variant 2 (S.T.C., unpublished data). Whereas this initially suggested that one variant was expressed more than the other, we subsequently realized that the libraries were of a different origin: the cDNA was made from Siberian larvae, whereas the genomic library was made from German larvae. This prompted us to apply the PCR/RFLP assay to genomic DNA acquired from geographically separated larval populations. PCR amplification of genomic DNA from American, German and Siberian larvae all

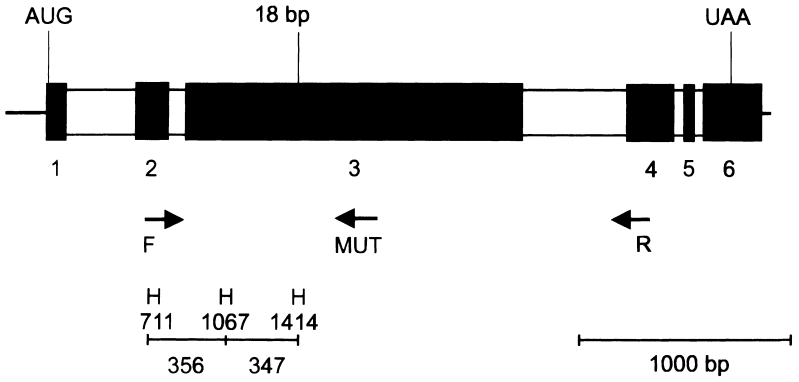

Fig. 4. Structural organization of the *ssp160* gene. Diagrammatic summary of exons (numbered black boxes) and introns (open boxes) deduced from gene and cDNA nt sequence alignment (Fig. 2), including the location of the start (AUG) and stop (UAA) codons and the polymorphic region (18 bp). Arrows below represent the position and direction of primers used for PCR amplification (F and R) and sequencing (MUT) of the polymorphic region. Bars below indicate the position of *Hin*dIII sites (H at nt 711, 1067 and 1414) critical to the PCR/RFLP assay and the scale (1000 bp) for the entire diagram.

produced a band similar in size to the 2444-bp band made from pC2A; however, upon digestion with *Hin*dIII, RFLPs were noted (Fig. 5). Densitometry indicated that the ratio of variant 1 (356+347-bp bands) to variant 2 (703-bp band) was 80:20 in American larvae; 22:78 in German larvae and 5:95 in Siberian larvae. If the variants were due to gene duplication, then their frequency should be 50:50; instead, the ratios observed suggest that the variants are polymorphic alleles whose frequency varies in these populations.

To determine whether the absence of the *Hin*dIII site at nt 1067 coincided with absence of the 18 bp in the cDNA, a primer (160.MUT) was used to simultaneously acquire single-strand sequence from PCR products made from pC2A (gene), λ160.1 (cDNA) and genomic DNAs described above. Whereas PCR products from cloned templates produced unambiguous sequences, genomic PCR products repeatedly displayed ambiguous bases at particular positions (Fig. 6A). These ambiguities likely reflect sequence heterogeneity in the population at sites most tolerant of variations. None the less, all sequences readily aligned spanning the divergent (nt 1067) and

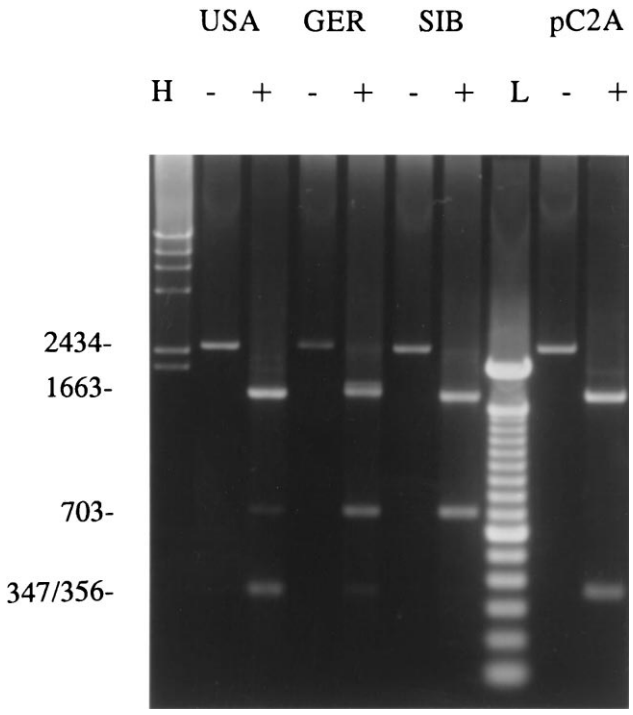

Fig. 5. PCR/RFLP assay for the *C. thummi ssp160* gene. Genomic DNA isolated from American (USA), German (GER) and Siberian (SIB) larvae and plasmid DNA (pC2A) were used as templates for PCR with F and R primers (Fig. 4). Aliquots of PCR products were incubated with (+) and without (−) addition of *Hin*dIII and fractionated by electrophoresis in a 1.5% agarose gel in parallel to λ *Hin*dIII fragments (H) and 100-bp ladder (L) size markers. Templates yielded a 2416–2434-bp fragment that was cleaved into a 1645–1663-bp and either a 703- or 347- and 356-bp fragments, depending upon the presence of a *Hin*dIII site at nt 1067 (Fig. 4). The frequency of this site varied in geographically separated larval populations.

one conserved (nt 1414) *Hin*dIII site and the site of the 18-bp variation (nt 1464). DNA from American larvae contained the same 18-bp sequence found in pC2A. Similar to the cDNA, this sequence was undetectable in DNA amplified from Siberian larvae; however, in this case, its apparent absence presumably reflects the relatively low abundance of this variant in the population. In other words, whereas the cDNA clone is representative of the most abundant allele in the Siberian larvae from which it was acquired (95% lack both the 18-bp sequence and *Hin*dIII site at nt 1067), the genomic clone represents the less prevalent allele in the German larvae from which it was acquired (only 22% have the 18-bp sequence and *Hin*dIII site) (Fig. 5). The most abundant allele in German larvae has a 15-bp deletion at the corresponding site (Fig. 6A).

3.4. Dynamic repeat region

Exon 3 contains the only repetitive element apparent in the cDNA; three regions composed of six or seven tandem copies of a 6- to 11-aa sequence ending with the *N*-linked glycosylation motif, N–X–(S/T) (single-letter aa abbreviations where X is any aa residue) (Hoffman et al., 1996). Proceeding from the *N*- to *C*-terminal end of the protein, consensus repeats are increasingly diverged in sequence and length: region I (aa 105–140 in Fig. 6B) is TSSNST, region II (aa 274–307 in Fig. 6B) is T(S/T)SNST and region III (aa 452–501 in Fig. 4 of Hoffman et al., 1996) is X_{0-5}XXXNST.

The evolutionary history of repeat region II appears dynamic when compared to other regions of the gene where only aa substitutions have been detected. For example, one 6-aa repeat in this region of the cDNA has two codons deleted (Hoffman et al., 1996). This region is also the site of the 18-bp variation between the cDNA and gene, and this sequence encodes another copy of a similar 6-aa repeat, STSNST (Fig. 6B). Furthermore, the sequence of genomic PCR products from geographically separated populations of *C. thummi* demonstrates that genes encoding both variants exist, and German larvae exhibit yet another variant; the corresponding site has a 5-aa deletion flanked by substitutions not detected in other species (Fig. 6B). Whereas unequal crossing-over cannot be eliminated, a more plausible explanation for the genetic polymorphisms observed in region II is that interspersed copies of CAA repeats provide targets for slipped-strand mispairing during DNA replication resulting in deletion of partial or full-length 6-aa repeats (Levinson and Gutman, 1987). However, it remains unclear as to why region II appears more dynamic than regions I and III where repeats are, respectively, more and less homogeneous.

3.5. Conclusions

The *C. thummi ssp160* gene consists of six exons and five introns, and 70% of the protein is encoded in intron

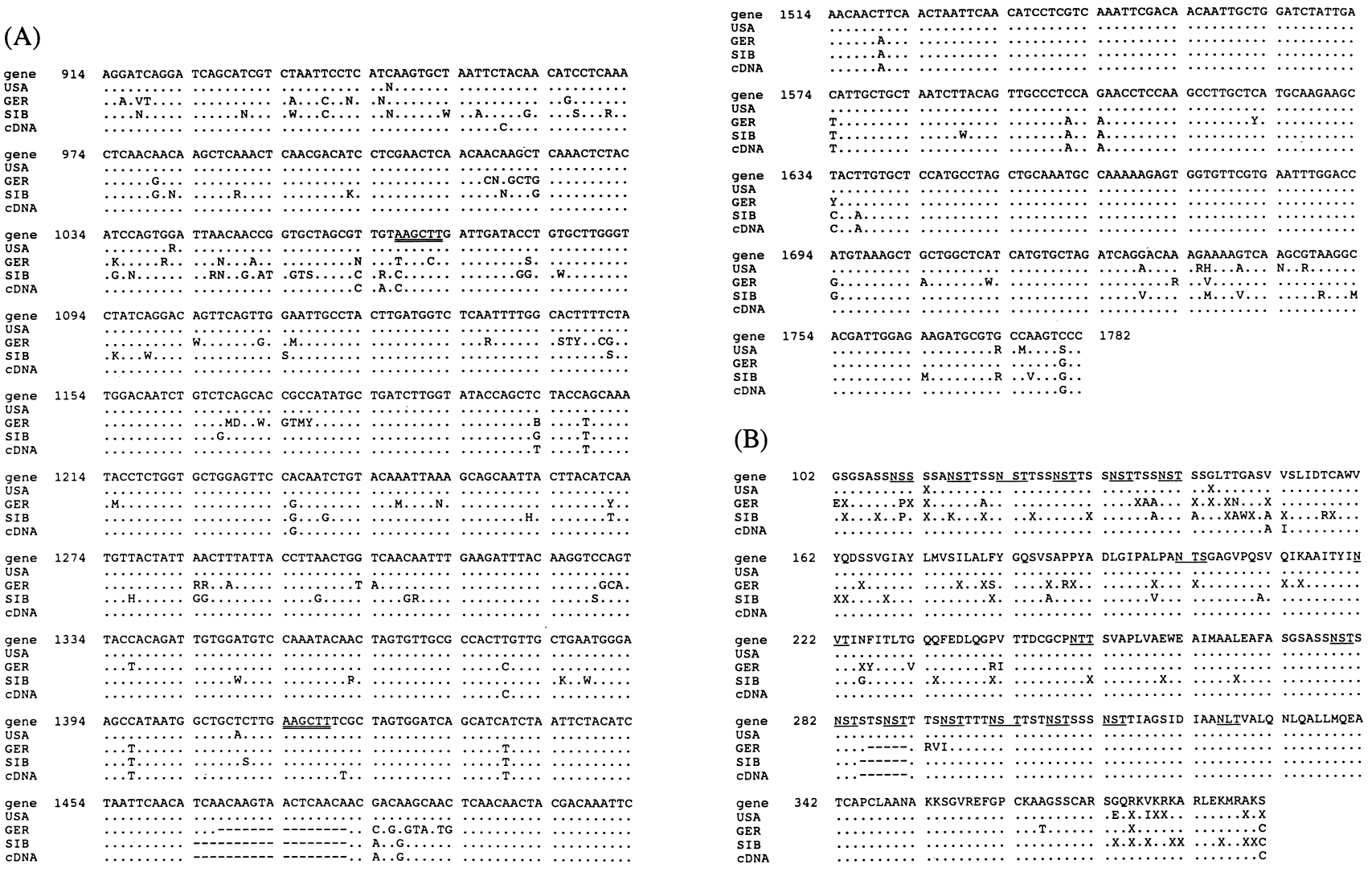

Fig. 6. Sequence alignments. (**A**) Unedited nt sequences acquired from PCR products obtained from pC2A (gene), λ160.1 (cDNA) and genomic DNA from American (USA), German (GER) and Siberian (SIB) larvae using primer 160.MUT (Fig. 4) were aligned. Numbering of the gene sequence corresponds to Fig. 2. Substitutions (G, A, T, and C), ambiguities (N, K, W, S, R), identities (.) and voids (–) are shown relative to the gene sequence. The variant (nt 1067) and one conserved (nt 1414) *Hin*dIII sites are double-underlined. The locations of these *Hin*dIII sites and the 18-bp void in the cDNA within exon 3 are shown in Fig. 4. (**B**) Amino acid sequences deduced from the nt sequences in (**A**). Numbering corresponds to the aa sequence deduced from near full-length cDNA (Hoffman et al., 1996). N–X–(S/T) glycosylation motifs (where X is any aa) are underlined.

3. Polymorphic alleles exist that can be identified by a PCR/RFLP assay. The polymorphism may be due to slipped-strand mispairing during DNA replication resulting in variations (0-, 5- and 6-aa deletions) in one of several 6-aa repeats containing a conserved *N*-linked glycosylation motif. The distribution of these alleles varies among geographically separated populations of larvae.

Acknowledgements

Dedicated to Professor Waclaw Szybalski in honor of 75 years of robust living, 50 years of imaginative research and 20 years of rigorous leadership as Editor-in-Chief of *Gene*.

We thank R.T. Hoffman for kindly providing ssp160 cDNA prior to publication and J. Niessing for providing the genomic library. Gerald Bergtrom and Thomas Hankeln generously provided genomic DNA from American and German larvae, respectively. This research was supported in part by National Science Foundation (Grant #DMB9204837) and US Army Research Office (Grant #DAAL04-95-1-0129) grants to S.T.C.

References

Beermann, W., 1961. Ein Balbiani-ring als locus einer speicheldrusen-mutation. Chromosoma 12, 1–25.

Case, S.T., Wieslander, L., 1992. Secretory proteins of *Chironomus* salivary glands: structural motifs and assembly characteristics of a novel biopolymer. Res. Prob. Cell Diff. 19, 187–226.

Case, S.T., Cox, C., Bell, W., Hoffman, R.T., Martin, J., Hamilton, R., 1997. Extraordinary conservation of cysteines among homologous *Chironomus* silk proteins sp185 and sp220. J. Mol. Evol. 44, 452–462.

Dorit, R.L., Gilbert, W., 1991. The limited universe of exons. Curr. Opin. Genet. Dev. 1, 464–469.

Fryxell, K.J., 1996. The co-evolution of gene family trees. Trends Genet. 12, 364–369.

Grzelak, C., 1995. Control of expression of silk protein genes. Comp. Biochem. Physiol. 110B, 671–681.

Hoffman, R.T., 1995. cDNA for a cell-specific protein from *Chironomus thummi*. PhD thesis, University of Mississippi Medical Center, Jackson, MS.

Hoffman, R.T., Schmidt, E.R., Case, S.T., 1996. A cell-specific glycosylated silk protein from *Chironomus thummi* salivary glands. J. Biol. Chem. 271, 9809–9815.

Kamakaka, R.T., 1997. Silencers and locus control regions: opposite sides of the same coin. Trends Biochem. Sci. 22, 124–128.

Kiknadze, I.I., Lopatin, O.E., Kolesnikov, N.N., Grunderina, L.I., 1990. The midge, *Chironomus thummi*. In: Detlaff, A., Vassetzky, S.G. (Eds.), Animal Species for Developmental Studies. Vol. 1. Invertebrates. Consultants Bureau, New York, pp. 133–178.

Kolesnikov, N.N., Karakin, E.I., Sebeleva, T.E., Meyer, L., Serfling, E., 1981. Cell specific synthesis and glycosylation of secretory proteins in larval salivary glands of *Chironomus thummi*. Chromosoma 83, 661–677.

Levinson, G., Gutman, G.A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4, 203–221.

Mach, V., Ohno, K., Kokubo, H., Suzuki, Y., 1996. The *Drosophila* fork head factor directly controls larval salivary gland-specific expression of the glue protein gene Sgs3. Nucleic Acids Res. 24, 2387–2394.

Meisler, M.H., Ting, C.N., 1993. The remarkable evolutionary history of the human amylase genes. Crit. Rev. Oral Biol. Med. 4, 503–509.

Middendorf, L.R., Bruce, J.C., Bruce, R.C., Eckles, R.D., Grone, D.L., Roemer, S.C., Sloniker, G.D., Steffens, D.L., Sutter, S.L., Brumbaugh, J.A., Patony, G., 1992. Continuous on-line DNA sequencing using a versatile infrared laser scanner/electrophoresis apparatus. Electrophoresis 13, 487–494.

Nony, P., Prudhomme, J.-C., Couble, P., 1995. Regulation of the P25 gene transcription in the silk gland of *Bombyx*. Biol. Cell 84, 43–52.

Ohta, T., 1991. Multigene families and the evolution of complexity. J. Mol. Evol. 33, 34–41.

Paulsson, G., Lendahl, U., Galli, J., Ericsson, C., Wieslander, L., 1990. The Balbiani ring 3 gene in *Chironomus tentans* has a diverged repetitive structure split by many introns. J. Mol. Biol. 211, 331–349.

Schuler, G.D., Altschul, S.F., Lipman, D.J., 1991. A workbench for multiple alignment construction and analysis. Proteins 9, 180–190.

Senapathy, P., Shapiro, M.B., Harris, N.L., 1990. Splice junctions, branch point sites and exons: sequence statistics, identification and applications to genome project. Meth. Enzymol. 183, 252–278.

Sheets, M.D., Ogg, S.C., Wickens, M.P., 1990. Point mutations in AAUAAA and the poly(A) addition sites: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. Nucleic Acids Res. 18, 5799–5805.

Smale, S.T., 1997. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochim. Biophys. Acta 1351, 73–88.

Travers, A., 1996. Transcription: building an initiation machine. Curr. Biol. 6, 401–403.

Wahle, E., Keller, W., 1992. The biochemistry of 3′-end cleavage and polyadenylation of RNA precursors. Annu. Rev. Biochem. 61, 419–440.

Wahle, E., Kuhn, U., 1997. The mechanism of 3′-cleavage and polyadenylation of eukaryotic pre-mRNA. Prog. Nucleic Acids Res. Mol. Biol. 57, 41–71.

Wieslander, L., 1994. The Balbiani ring multigene family: coding repetitive sequences and evolution of a tissue-specific cell function. Prog. Nucleic Acids Res. Mol. Biol. 48, 275–313.