Int. J. Indust. Entomol. Vol. 5, No. 1, 2002, pp. 93~102

Research Article

International Journal of *Industrial Entomology*

Molecular Cloning and Characterization of the Fibroin Light Chain Gene from the Silkworm Baekok-Jam (*Bombyx mori*)

Kwang Ho Choi[1, 2], Tae Won Goo[2], Eun Young Yun[2], Jae Sam Hwang[2], Seok Woo Kang[2], Sang Mong Lee[3], Hung Dae Sohn[1] and Byung Rae Jin[1,*]

[1]*College of Natural Resources and Life Science, Dong-A University, Busan 604-714, Korea.*
[2]*Department of Sericulture and Entomology, National Institute of Agricultural Science and Technology, RDA, Suwon 441-744, Korea.*
[3]*Department of Sericultural and Entomological Biology, Miryang National University, Miryang 627-130, Korea.*

(Received 19 July 2002; Accepted 22 August 2002)

We have cloned and characterized the complete fibroin light chain gene from the silkworm Baekok-Jam, *Bombyx mori*, a recommended variety in Korea. It consists of seven exons and six introns. It consists of 14,663 nucleotides long with an open reading frame of 786 nucleotides that encodes a protein of 262 amino acid residues with a molecular mass of approximately 26,000 dalton. The amino acid alignment revealed that the Baekok-Jam fibroin light chain had 98.5% protein sequence identity to J139 strain: differed at four amino acid positions (11, 46, 80 and 123). The Northern hybridization analysis showed that the Baekok-Jam fibroin light chain gene was specifically expressed in the posterior silk gland.

Key words: Silkworm, *Bombyx mori*, Fibroin, Light chain gene, Genomic DNA

Introduction

Silk fibroin of the silkworm, *Bombyx mori*, is synthesized within the cells of a pair of posterior silk glands (PSGs) during the larval fifth instar, secreted into the lumen of PSG, and transported through the middle silk gland, where heterogeneous molecules of sericin are added, and further toward the anterior part of the silk gland, where the silk fiber is formed and spun (Inoue *et al.*, 2000). Silk fibroin of *B. mori* has a molecular complex consisting of the heavy chain (H-chain; about 350 kDa), the light chain (L-chain; 26 kDa) and glycoprotein P25 (about 30 kDa) (Tanaka *et al.*, 1993, 1999).

Fibroin H- and L-chains secreted from the posterior silk gland are linked by a disulfide bond (Tanaka *et al.*, 1993; Yamaguchi *et al.*, 1989). The disulfide linkage of fibroin H- and L-chains is essential for the efficient secretion of silk fibroin as demonstrated in the analysis of naked pupae mutations, Nd-s^D and Nd-s (Mori *et al.*, 1995; Takei *et al.*, 1987; Yamaguchi *et al.*, 1989). Glycoprotein P25 is associated with the H-L chain complex by non-covalent interactions (Tanaka *et al.*, 1999). Inoue *et al.* (2000) found that molar ratios of the H-chain, L-chain and P25 are 6:6:1 by using quantitative enzyme-linked immunosorbent assay. H-chain is located on the 25th chromosome and L-chain is on the 14th chromosome but their expression seems to be regulated coordinately in tissue and molting-stage-specific manners (Hui *et al.*, 1990a; 1990b; Kikuchi *et al.*, 1992; Kimura *et al.*, 1985).

Previously, the primary structure of fibroin L-chain of *B. mori* J139, a Japanese breed, was determined by nucleotide sequencing analysis of fibroin L-cDNA clones and genomic DNA, and by amino acid sequence analysis of fibroin L-chain peptides (Kikuchi *et al.*, 1992; Yamaguchi *et al.*, 1989). Recently, Zhou *et al.* (2000) reported the fine organization of *B. mori* fibroin heavy chain gene.

In the present study, we have cloned and sequenced fibroin L-chain gene from the silkworm Baekok-Jam, a recommended variety in Korea. The complete nucleotide sequence, exon-intron structure and characterization of fibroin L-chain gene from the silkworm Beakok-Jam, *B. mori*, are discussed.

*To whom correspondence should be addressed.
College of Natural Resources and Life Science, Dong-A University, Busan 604-714, Korea. Tel: +82-51-200-7594; Fax: +82-51-200-7594; E-mail: brjin@mail.donga.ac.kr

Materials and Methods

Silkworm

The silkworm Baekok-Jam, *Bombyx mori*, a recommended variety in Korea was used in this study. The silkworm Baekok-Jam is a variety generated by single cross F_1 hybrid between Japanese race Jam 123 and Chinese race Jam 124.

PCR amplification, cloning and nucleotide sequencing

The genomic DNA of Baekok-Jam was extracted from posterior silk gland of the fifth instar larvae. PCR primers were synthesized for the 8.5 kb front region and the 7.2 kb rear region of fibroin light chain gene according to the published sequence data for the L-chain gene, GeneBank accession number M76430 (Kikuchi *et al.*, 1992). The primers synthesized for amplification of front region fragment were: primer 1 as a forward primer (5'-TAACAAAGTGGTGCCTATC CCAC-3') and primer 2 as a reverse primer (5'-GGACTGACAAGAGGTCCATCAGT-3'). The primers synthesized for amplification of rear region fragment were: primer 3 as a forward primer (5'-GTACCGGGATTGCGCGTGTCAGGATGACTCAGA-3') and primer 4 as a reverse primer (5'-CCATGTAC-CCACTGTCCAATCCACCGTCTTTGGG-3').

PCR amplification was carried out with the long template PCR system (Boehringer Mannheim, Germany) with 10 pmol of each primer and 500 ng of genomic DNA as template. After heating the reaction mixture (50 μl) at 94°C for 10 min, amplification was carried out for 30 cycles of denaturation (30 sec at 94°C) and annealing extension (10 min at 68°C). A final 12 min step at 72°C was performed at the complementation of these cycles. And the fragments were inserted into pGem-T easy vector (Promega). These pGem-L-chian clones were sequenced using T3 and T7 primers by an automatic sequencer (Perkin Elmer, Watsonville, CA, ABI 377). Complete sequence analysis of the fibroin L-chain was performed by using primer walking with synthetic primers. Each purified DNA sample (300-500 ng) was mixed with L-chain specific primer (3.2 pmol) and Termination Reaction Mix (Perkin Elmer), and sequenced following 25 cycles of PCR condition (30 sec at 96°C, 15 sec at 50°C, and 4 min at 60°C). The resulting PCR products were separated on 4.5% denatured polyacrylamide gel and analyzed by DNA Sequencing Analysis Software (Perkin Elmer).

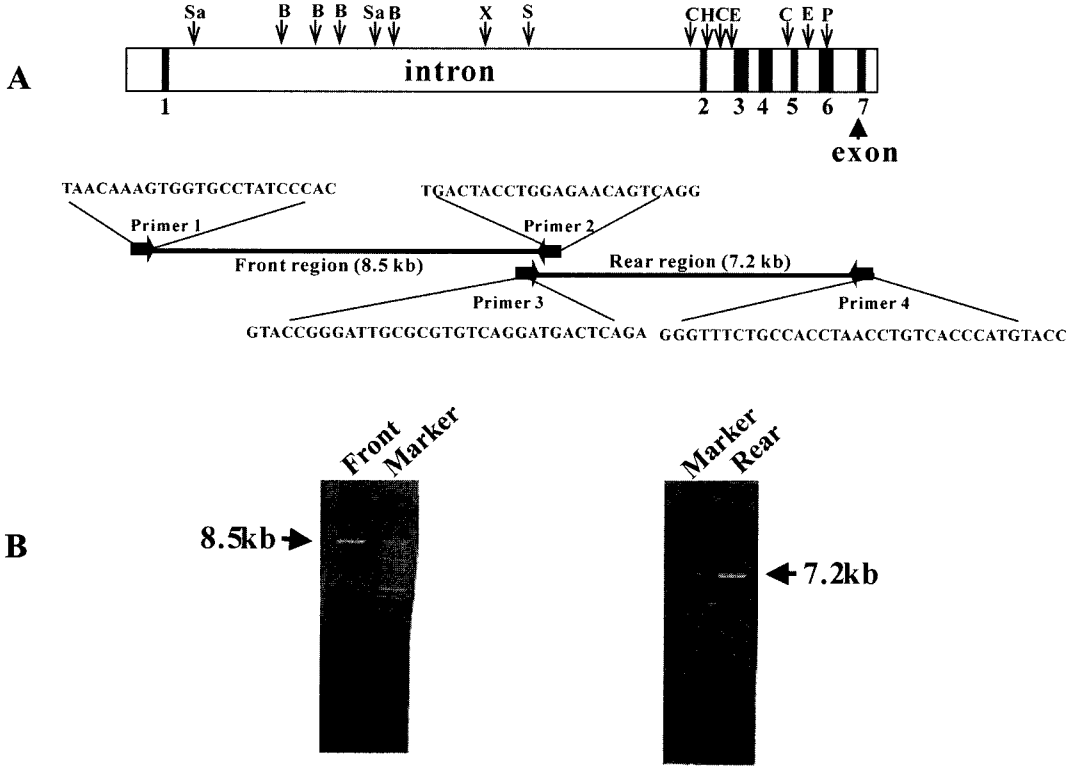

Fig. 1. Genomic DNA structure (A) and PCR products (B) of the silkworm Baekok-Jam fibroin L-chain gene. The Baekok-Jam fibroin L-chain gene was composed of seven exons. Restriction enzymes are abbreviated as follows. B, *Bam*HI; C, *Cla* I; E, *Eco*RV; H, *Hind*III; P, *Pvu* II; Sa, *Sal* I; S, *Sph* I; and X, *Xho* I. The PCR primers (primer 1 to 4) are indicated. The amplified fibroin L-chain gene was analyzed by 0.8% agarose gel electrophoresis.

```
   1 Gaattcaaataacaaagtggtgcctatcccacttttttgatccagacaaagaaaataagtgttttcgg
  71 gagctgaaaaattaatttcaggaaacaacaaaaataatgacgcaaaagtacaccggagtgaaaataaac
 141 ctaagaaagtaatcgctaaaaattattcatctcgtgaattgattgagcgcgataataacgcagtactat
 211 ggagagattctatgtttaatatattaatgatatgatataaaaaagggtgcgtgtacttatgtacgcgcg
 281 aagaagttatactttattttcattaaatttatttctttttttttatttcaattttaatcaatttgaaaa
 351 aaatcgaataaacaacatcctcaaacatgcatattggacatccctttttcttgacatcgtataaattcgg
 421 aattctcggtacggttcgtaaagttcacctgcggctatattccgactcgccaagttacgtcagtcgtat
 491 gtaatgagcgatttagtgggcaacttcattctgttaattttgtgtcacggtgcgcgcgcatcgtaaaac
 561 tcactctcatagatttttcataacgcgcctaaagaagtataacttcaataatttaaatttaaaaaaaaa
 631 atgcatagaataattatatgaattatttaaaatgtcatttaccgacattgacataacagacgacgttaa
 701 actacaaaacatttttaattccacattgttacatattcaacagttaaatttgcgttaattctcgatgcga
 771 caaatataagaacaatcggatcaattagatcgctttgtttcgaacaacacttagtttaactagaggcgt
 841 cacctcaagaaatcatcttcattagaaactaaaccttaaaatcgcaataataaagcatagtcaatttta
 911 ctgaaatgcaaagtcttttgaacgttagatgctgtcagcgttcgttggtacagttgtttgatatttatt
 981 taattgtctttttatatataaatagtgaaacattaatcacggaatcctgtatagtatataccgattggtc EXON 1
1051 acataacagaccactaaaatgaagcctatatttttggtattactcgtcgttacagtaagtaattttaac
                     M  K  P  I  F  L  V  L  L  V  V  T
1121 Ttttcaaatttatttacaatatatactcaccacgttttcaaacaatttttacaattctagaactttcg
1191 tctgtttttttttatttgttcaacttgtgtatttgtggattccctacaatatggaaaagatgttttcgt
1261 tatattttcacagataaaagaaaaaacaattgtctagctagagcgtagacaattaattctcattaatta
1331 tttttgactatgacataatatatagttacatttgcttttatttagtattatttacgatttttataagg
1401 agtgtaattatttaaaaaaatattttactgcattgtgaaattgttaataaaaaaatattctatacgtga
1471 tcgcgctcgaagcgatttcaaaaattaagacgttaattaagtaagaaaaaataaaaataaaattatgcg
1541 ggaatggccggcattcacggtatcgcaatcactacatagtataaaacaaagtcgctttctctgtcccta
1611 atctgtccctatgtatgcttaaatctttaaaactacacaacggattttgatgcggtttttttaataga
1681 agagtgattgaagaggaaagtttatatgtataataacatccattaaataggggagaaatcaataataaa
1751 tacagtttccgaagcgaagcgagggcgtttcgctagttgtaacaataacttccaaagtcgacagatatg
1821 ttcctacttttttatttcattttgttcattgcccgagcccacaactatcacaatgtagggcaccaagtgg
1891 tataccaatattgtaatgcttggttctattagaccagattttgaatataacgaaattatataacagtta
1961 agcgggtttgatttttattacacgattttattctttcatcgtggaagatagaaaacttttgttaagtac
2031 tattttatcagaaaaattggtatctgcctgcgggattcgaacatcgatacagtcgcatcactcgatacg
2101 atgcactggacgtcttatccgttaggccacgacgactacaaaagtctgcgtgtcttgtgtctgcctgcg
2171 gattcgaacatcggtacagtcgcatcgcttgatacgaatgcactggacgtctcatcctttaggacacaa
2241 gacttctaagctctttttatccctacctattcactagtaggctaaagagctattccagcttcacgtgaa
2311 gggtaggtgagatcacgggttcaacttgagagattttgctaatacaagccctagcaagagtactgcttc
2381 cagaatctactaccggatcggaatcgggacctactgagaagatgcggaaactcagtgggctatgtctat
2451 gattaagttcgctcgtcaaactcttaatcgactagttcgacgagaacggtgaccgactcttgaagatcc
2521 gaaatcaccgagagtagatccgcaggattcgacaagacatgtttcgggaaaaggtggctttgccttgct
2591 cgtcgcctaagtcggataggtaattagcggtgatcactggttttactggtggtaggacctcttgtgagt
2661 cgcacgggtaggtaccaccaccctgcctatttctgccgtgaagcagtaatgcgtttcggtttgaagggt
2731 gggcagtcgttgtaactataattgagaccttagaacttatatctcaaggtgggtggcgcatttacgttg
2801 agatgtctatgggctccagtaaccacttaacagcaggtgggctgtgagctcgtccactcacctaagcaa
```

Fig. 2. Complete nucleotide and deduce amino acid sequences of the silkworm Baekok-Jam fibroin L-chain gene. The exons of fibroin L-chain gene are solid boxed. CAAT and TATA box sequences are open boxed and underlined, respectively. The polyadenylation signal is double underlined. The deduced amino acid sequences are indicated under the exons.

```
2871 aaaaaaaaaaaaaaaaaaaaaagatcataagatcatgggagcggacatcttatgagtccgcacgggta
2941 gtaccaccaccctgcctatttctgccgtgaagcagtaatgcagtagtttgttttaatataaacatttga
3011 tatactgttttattcatttatttattacacttcatgtaatatctacatgggctgatttaatgcttaagg
3081 attccctaccagtcaaccaaaactagggcagaatgaataatggtaggtgcattaaaagatacgttttat
3151 atttttaataatttttataaaatattatttatgaaattattttatatacatatttataaaatagttttaa
3221 attattttcaatattccctagctcgttcggtcgttaaataaatttttttgaaccatgataagtttcgcgg
3291 aaactgatagtatatctcaattcgcacgaaaaaattaaaaacacttatttctcagaactttattagcca
3361 ttaggattaattatttaaatcgtgtattaaaaacaatgaaatatttactcaaattggacacaataataa
3431 atctaggacacggtccatggctggagcggtgtggccacgtttttttttttttgcttagatgggtagacta
3501 ctcacagcccacctgatgataagtggttactggagcccatagacatctacaacgtaaatgggccaccca
3571 cttgagatataagttctaaggtctcaagtatagatacatgtgaattcaaataattatagtagaataaat
3641 aaatgtgatgtaaataataatacattgacgtaagagttccatgtcgagttcgacactagttagtggaaa
3711 atttgggaacagtaatcagtatcggtatcagtatcaggatcagtactgtatcaagatcagtgtcagtat
3781 agtgtcagtttcagtatcagtatcagtatcagtatcagtaacagtaacagtatcagtatcagtatcagt
3851 tcagtatcagtaacagtaacagtaacagtatcagtatcagtatcagtatcagtatcagtatcagtatca
3921 tatcagtatcagtatcagtatcagtaacagtaacagtatcagtatcagtatcagtatcagtatcagtaa
3991 agtatcagtgtcagtatctgtgtcagtatctgtgtcagtatctgtgtcagtatcagattagtattagta
4061 tagtattagtattagtattagtattagtattagtattagtattagtatcagtatcagtatcagtatcag
4131 ttcaaagggtttgcgtacatgcgatctagcccgtagaaaacccactgagcttttcaccggattttcgca
4201 tggatcccgatttcgatctgatggtagattctgcgaagcactgctcctgctagagccggtgttagcaaa
4271 tctctcagcttagcccgtgagctcaccgtcccgtccgcacgtataaaaataatagcaccttaggctccc
4341 gcgaataggtaggagataaaaagaattatattattagtttctataaaattttattataatttttatttt
4411 atttttttattgcttagatgggtggaggagctcacagcccacctggtgttaagtggtgactagagctca
4481 agacatctacaacgtaaacgcgccacccaccttgagatataagttctaaggtctcagtatagttacaac
4551 gctgccccacccttcaaaccgaaacgcattactgcttcacggcagaaataggcagggcgggggtaacta
4621 ccgcgcggactcacaacaggtcctaccaccagtaatatattacgtactaaacgagtcccttcaaaataa
4691 taggtcgtaaatcgcaaaacataaacattaaaaccatccctaaacccgtcattaagtaatctgacagca
4761 attttatatgaacaacgtgttgcgtgttttgaatatataaatattatccgaagcgtgcggtgaaacttt
4831 caaattggaaatactccgacgtaaatacttaggacgcgttagtattatgccctacgacttgtttcgctt
4901 tcttctgtcttattatatcacatccaattgcttatctatacttctatactaatattataaagaggaaag
4971 tttgtttgtttgtttgtattaaataggctccgaaactgctgaaccgatttgaaaaattctttcactgtt
5041 ggagctacactattcccgagtgcataggctataatatttttgaaaaaattagggatccttactaaaac
5111 ccaataatgtaaccaaaagtgtaaaaaaattacctaaaatactctttacatcgcgtgccctgcgaaaac
5181 attgatgatagaacaaaataatgtactacgactttgtagaacacattgttatttacagaaagtgtcgcg
5251 cagcatatgtttaactattatagttatgccgcacgaagtgttcttttatttttttaaataaaacaactt
5321 aaatatcgttgaaatttttgttaaagacccgagcggatccggagccggccgctagttaaatataaactg
5391 ttttacagtttgatctgacatttcattatttttttacttataaggagcctagtgtaatcgtatcgtctga
5461 gcgactatgcccctattcaaattcaaggatgcttaaagacgcccactcttttctcaataaataattggt
5531 actaagacagtataatttttcacaatgaaagaattacaacacgtaataaaaatcaaaccaccaaatttttt
5601 aatctttgcgttaacagtgtgtgcacaaattgtaagccgctacaggtaggtatcatagttttgcctatt
5671 cagtcataccacgattcgaagagtgaggtagttttgaaacgattgctacctggacaaggtactttcaag
5741 tttttgcgttcatgccgtctataagctacggtaacggcttagctgaggcgagccataaactcacatggc
```

Fig. 2. Contined.

5811 ttctttttttttgtgtttgaaagattactggtagcccgaaggcctttccagtttcgccagaacaggtggg
5881 gagtaaaggctcagccagaaggggtgggatttgcttacaactgcccgagcgcctccaaaggagacctaa
5951 aactcaagagcaattgcttcgcgaatgaatctactaccggatcggaatcgcgacccgctcagaaaatcc
6021 gcgagaaactcagcgggctaatgcatgggttaggttgcacgtcgacctcttggtcgagttcgacgagta
6091 ggttaccggggtccttaagcctgctcctagtattagagctgaaggtgtctaatgcaaaggttattggat
6161 tgatggatccgtaaggacgtgacttggcctatttgcctatctgcctataaaaatgtattccgtgaaatc
6231 aaattcaaatttaacgtcaaatttatatgtgaattacatatgtatgaggaatgcgtgatcgtgtgaagt
6301 cggaccgtggaaaattgaaattggctttgaggatgacgggtctgattaggtgcctgaagcgcgttcatt
6371 cgtagaattgaaatgttcggacttcaattggatgttcaatggatataatgtaaaaaaatattattttca
6441 gattcaatcttgtttttttttacatcgttggtgatcaagaaaactgtagagtgattgaaattatcttgt
6511 aataaaaccagcgagattaaaccttaaaattaaagatcaaaattaagttattttacatatgtctataat
6581 tttttaattgtttttttattgcttatatcggtggacgagctcacagcccacctggtgttaagtggttac
6651 ggagcacatagacatctacaacgtaaatgcgccacccccctcgagatataagttaaatatagttacaac
6721 gctgccccactctttaaaccgaaacgcattactgcttcacggcagaaatgggcggggtggtggtaccta
6791 ccgtgcggactcacaagaggtcctaccaccagtaattatgcaaattataatttgcgggtttgattttt
6861 ttaactacacgatgttattccttcaccgtggaagtcaatcgtgaagatttcttatgtacgtatttcatt
6931 gaaaaattggtacctgcctgcgggattccaacaccgatgcaccgctagataccaatgcccggacgtctt
7001 tcttttaggccacgacgacttttaaaaccaaatagataatgaggtatcgagagagtacatcgcacccca
7071 agcgattcccacttcctctacattcaaatgcgtatatatcgtaataatgctacacaaatcggttaaaat
7141 aaattcgcagaagacactaaatacaaaatcatcatcggattgtgtcttactgcgaacgagctgacacac
7211 agtcctaaagcaaaacaaccgtgcattttattcggtacacaatggaaatatttatttctctcgccttgt
7281 attcctcaagatggcttcatacgtcaaacgcaggatatacggtacatcgcggcgaacggtttcgtaccg
7351 gattgcgcgtgtcaggatactcagatttgcagagcgaacacaaaaatactgacggtttttcttggcggg
7421 aaattagcttggaatcatgagtgaattatttgatatttcttcagtagcgaatggcaattaaataacata
7491 acataagagataacgaatctatatattaatacgtgaagcaaaaactttgtatccctttttacgaaaatt
7561 cgcggacggaggagtatgaaatttttccacacttatagagaatatagagaagaagtgcacaatgctaata
7631 ttttttaaataatgcataaaagatacattaaatcaataaagaaaacattacacagactacataccatat
7701 tttgacgcacacacgcatgcatagtgcatactatttattgtcaaacttttgttcttgacgtctgttgtc
7771 aattgagaatagattaaatattgtttgtcattgataatatttttttatagtgtagtcttggcaaaatttc
7841 tggcgaagtataaaatacaatcataatagtgtacaaacttacaattccaattaattatagtcgaatttc
7911 actactgcgggacctctagttcagataattctgcgatataatctcaatgtgctacattaagatatgatt
7981 atattaaaaatactgttactgttttatttcacgttttgaactgtcattatattatttcctgtttttcga
8051 tactgaacggtcgaacgactacttgaacggtcgaaggtcacttgaacatcttatataagaagggtcaat
8121 cgttacttttttcgtctttttgtatgtctgtcttttcagaagtcaaaaaatgtactgccgtgccgctggt
8191 acgtacgccctcgttccatttatttttcgaattattttcaaatctctgtcttagtttgtctatgtctagt
8261 attgtaggtcttatgtcctgtcaaaaaggtcacacgccactacaaatactttagcgttttcgtgaacac
8331 gacgatttaattactatcaaattctctaggaaaacatttctaaagattcttaaggttcgcttagaacgt
8401 tgtttgttcgtttggtaaaaacaaaaggagcacacaaaacataaaaattttatctttgtcatttctcgc
8471 atccccttttttttacggaaacacctaccatgcgtttacgaactaactttactgatggacctcttgtcag
8541 ccacatgggtaggtaccaccaccctgcctatttttgacgtgaagcagtaatgcgtttcggcttgaaggg
8611 ggggcagcacttttaactatacttgggaccttagaacttatatctcaagatgggtggtgcatttacatt
8681 tagatgtctatgggcttcagtaaccacttaacaccaggtgggctgtcatctcgtacaatcatctaagca

Fig. 2. Contined.

```
 8751 taaaatataaaaaataaaaataactccctcccaaaaagaatatcgtcgcgtatgagtgaaataatttat
 8821 gcgagctcgcacgaatatcctccctatgcctattactgcggcgaaataatacactttccaagatgaaac
 8891 atataataatatgtactcagactacaacctcatgtctcaaggtctcatgtgaatttgttcgcctccgag
 8961 gctacttaacgtgagatcgaacgcgaactccatctagaactataatataataatcattagaatggagta
 9031 gtcgtctgaaattcgattctccgctataacccgttaatctaacctgaccattaagtaaataaaaattag
 9101 tctagatgagctcccggcgtacctggtcgcattaacacatcgatcggtggagcgacaaagctcttcgct
 9171 taagcctcaaaagaacaatcccgagaaaacaatttgttgtgtataatttacaccaaaacccgaatttaa
 9241 ttttcgcagagcgcctacgctgcaccatcggtgaccatcaatcaatacagtgataatgaaattccacgc  EXON 2
                S  A  Y  A  A  P  S  V  T  I  N  Q  Y  S  D  N  E  I  P  R
 9311 acattgatgatggaaaagctaggtaattatgtaagctttaactaatattaatataggtcttatataatc
       D  I  D  D  G  K  A
 9381 ggcagaggcaaatgcaccgagcgacttccgcaatagaccacgacgacttcgactaaaatatcttcatta
 9451 ttttttattacttttaaaaaatatgaaaatacggtccattgaatggtcaattatcaccgtcacagatc
 9521 tcacatcagcagtaccaaagacacatagactaactgacgttatactattttccaacctcgttcgctcga
 9591 ctgcgccattgagggtggatcatttcacgatcatacggcaaaaatgtaaaccgaaaatgcactgtaaat
 9661 atcgtaagtggtttcaagaagtagtagtgatctcttctccagtggtgggtggtgtgattttcctcaatg
 9731 caaattcttcagaacactgcccatggaacatacagtaagaaatacaaagggaatcaaagtcccaccgta
 9801 ccccagaggctcttacttatccttctgtatgaagtcaaacagaactatctcgaaattggatatagaagt
 9871 ccctacgtcaaatagaaagtggaagtcttgatattaaagcggagtgaaagccttacacccaaaatccct
 9941 gacccttgagggcttgatgaccggtgagtgatcgatatctgaccgtccttgaccaaaacagttactaa
10011 taaactgagacctatcaagttaggccaagttctctaggaacagcatattgataaaccactctatcatct
10081 ggaaagcagagtactcatagttatcgtataggccttaactcatagtcagttacgtcattttatttggga
10151 gagaatggacgtgagtattctgtcacaaagacgaaatcctattgccatgtgtctttgaatctaatttgg
10221 tatcggaaaatctgggagttgtacggcgtatacatcgtttaaaagcagactatataatacgaatcgcaa
10291 tatgaagccaatcatcacagtgcatctagtgataatagacatcacgacgtcaatggcaccgactatctt
10361 agactagaagttaaactgtctgttgcattgaaatctattacaccgaaatcgttttgctgcttcaaaact
10431 taacagacaaagctcacagtatccttttctttaaactatgtttttattgcacaaaacttttcaatcatgg
10501 tcaagtgttttcaataattatatatttaataattattataatttagttccgtaatctcacgtgcatgg  EXON 3
                                                S  S  V  I  S  R  A  W
10571 actacgtcgatgacaccgacaaaagcatcgccatcctcaacgttcaagagatcttgaaggacatggcca
       D  Y  V  D  D  T  D  K  S  I  A  I  L  N  V  Q  E  I  L  K  D  M  A
10641 ccagggcgactatgcaagtcaagcaccagcggtggcccaaaccgccggaattatcgcccatctatctgc
       S  Q  G  D  Y  A  S  Q  A  P  A  V  A  Q  T  A  G  I  A  H  L  S  A  G
10711 ggtatccccggtgatgcctgtgcagccgctaacgtaagtagatgccgctgttagaaacaataaaacagt
         I  P  G  D  A  C  A  A  A  N  V
10781 cgtctaaatatttattacccgaaactagaagaagcagtgtcacttttgcgagactttagaaagaaaaat
10851 ggtaccgtagattggaggaaatcgggacccctttctaaatccgacacaaattttcaatggttagggtaat
10921 accaataaaatagaggttaaaacgtttccggagggttaaattttaatatacggtgcgttattatacata
10991 ttcattacaaaatctacacaaaaactgaaaaaactagcttgacccgattcaccttggctttggggtgaa
11061 caggtttcgtatggggacaatagagtttttcgataaggctggtactatatttagtagttaggtaggcac
11131 taactttaggtaattaacttttttaatatctctaatgagttacaaaactctgatgactaaattaaactgt
      taatatttttaaacactaaataaaagtcattttcagttcaacaaagttaaatttaacaactcaacaaaa
```

Fig. 2. Contined.

```
11201  taatatttttaaacactaaataaaagtcattttcagttcaacaaagttaaatttaacaactcaacaaaa
11271  tattgtgacatcacagatatatccctatactaactcttagtctttattcccccaatctactgtatccaa
11341  attattctttaaacatccaatattacctgttgtttcagtaaaaggaaatggaaaatctaattaaaatct
11411  ctttgcaggtcattaactcttacacagacggcgtcaggtccggaaacttcgccggcttcatacaatctc  EXON 4
                I  N  S  Y  T  D  G  V  R  S  G  N  F  A  G  F  I  Q  S  L
11481  cggtcccttcttcggacacgtgggacaaaacttgaatcttatcaatcaactcgtcaccaaccctggtca
       L  G  P  F  F  G  H  V  G  Q  N  L  N  L  I  N  Q  L  V  I  N  P  G  Q
11551  ctccgatactctgtaagtactcagcaattgttaccgaagagttacatacttgaatcatggtcggggctt
          L  R  Y  S
11621  aatattcgatttatggattagagaaatctgcactgaggtcaatatagcggtatgaggttactaccctat
11691  gctttttattctgaagaaaatagaatataaaagaatacaaaaaaggtcaattaccaaacctcttgtctc
11761  tgccaaacggcaaaatatattctgtgacgtccatgggctccgataacgacttagtgggaggcacttgat
11831  cttatcctgggtaaaatgatcgtttcatcacgggatactatcaccatccaagctaaatagtttagtaga
11901  atctactaatatatgcacgacggtttttcgtaggagggccgtgatccaatgcaatcgataattcgactt
11971  atgcctcaagttaggtagtggaattcaattctatggattcctaaccgtcctaggatgaacttaagggaa
12041  acgagcgctaattgtgtacgataatttaaaagtcacgggtattttaaattttgtattcacacttctagt
12111  taaatattaaactattgtttttcgagaaagatgataaaccacactagtccaaatcgagtatgttcacca
12181  cgggtaagaataacaagtgttgttttttttagGTCGGACCAGCCCTGGGTTGTGCCGGAGGTGGAAGAA  EXON 5
                                   V  G  P  A  L  G  C  A  G  G  G  R
12251  CTATGACTTCGAAGCCGCTTGGGATGCAATCTTAGCCAGCAGTGACTCTAGgtaagcatattgagcttt
       I  Y  D  F  E  A  A  W  D  A  I  L  A  S  S  D  S
12321  caaataagaatgtaaactggcacacaaggaatcaatcactaagaatttatctctagtattttattagag
12391  cattcattcttttgcagatccttcttatcttaatctttagaagcagcaaatagaagacgtttgcaccct
12461  accatccttgtgtctccccgtctttaactcctcacctccaaacgcagatcgcaggagcttttagctaat
12531  acacacgctaataagattttaagcatgtgcacccatttaaatgagtgtgatcttatctgttctaatata
12601  cacgaattacacgtgctttacgctagtttatgttttactgtaactgttctggaccaaaattcttcttct
12671  tcatctagcatcttttctttagcaatatgtggctctatagcattgtataacaagggtgtgttatgatct
12741  gctcgattagcacagattggtaggatagtggtgcctatttttgcggacttccgatacgccctagctcca
12811  aatatatatttcaactagcagtcgtggccacacgttgctcttgctaaggttttcgtttctaaataacac
12881  tggccggctatgctaataccaaaaattaacaaagtaagatcaattggatttcagtgtaatgcgtttact
12951  cgaaataatgttcacttatagtttagcctcaacaacacgtggtggttatgcttttaaattgtagcttat
13021  tgacacatttatagccatctattgattacttactcaatccagtcaacagatatcgccatctgttaaaat
13091  gtttgaagttaacagataattgtgaatgtcaaataatagataaataatttgcaataaaataatattgcg
13161  ctataaattaagatgtaagctatcctatctttcaagatggatcaaactgcacacggtgtgcaaatttaa
13231  ttaaaatcggttaagtagtttaggagtccatcgcggacaaaacaacgtgacacgtgatttatatatatt
13301  agatatgtaccgacgattgaattgtattttactttccttcaattcgattacagTTTCTTAAATGAAGAG  EXON 6
                                                     S  F  L  N  E  E
13371  actgcatcgtcaagagattgtacaactctcgcaacagccaaagcaacaacatcgctgcctatataaccg
       Y  C  I  V  K  R  L  Y  N  S  R  N  S  Q  S  N  N  I  A  A  Y  I  T
13441  tcacttacttccaccagttgctcaagtgttccaccaatcagctggatcaatcacagacgtaagttacgt
       A  H  L  L  P  P  V  A  Q  V  F  H  Q  S  A  G  S  I  T  D
```

Fig. 2. Contined.

```
13511  aaattaccgagcgtctttatgtaattttgttattagtaatgattggacaaatgagcttatcgccgattc
13581  tcacaatcactcgtcactgatgtcacaaaggtcacaaagtgaacgtgttttggagttgatagatataga
13651  actccttattcaaaaaatatatatacatagacaacaattacaataggtgtcttacaaaaacgtgaaaaa
13721  atgtctgtcggttcgaggagtaaaacgtaaaacaaaataacaaaactgtgattttgataaaaaaaatct
13791  aagcatccacataatgaattctaatttataaaagtgtgttagtatgttatttcgtaactgcctgcactc
13861  gcgatttaaagattggactttgctaccaagaactttaaattatatctacgcgaccatcactatgagact
13931  aagctgaaagtatttctttcaaaacacgcttcatagatttaaagctgcccttcaatccagacatataag
14001  gctacgaatcagacttaggccagcaaggtgtccatcttgtttttaccacctaaattgggactactatat
14071  ttaattcacgttttaattatccacagctcctgagaggcgttggcaacggtaatgacgcgaccggctta  EXON 7
                                  L  L  R  G  V  G  N  G  N  D  A  T  G  L
14141  ttgctaatgctcaaagatatattgcacaagcagccagccaggttcacgtctaaataagaactgtaaata
       V  A  N  A  Q  R  Y  I  A  Q  A  A  S  Q  V  H  V  *
14211  tgtatatatataattatataaaagatatatataaccatatacaaacatatatatcattataagacaatc
14281  acctatataaaaacagactaaaattaataattatgtatactttaattgtgtttaggacattttatgcaa
14351  ttgtgtttgcgttaggatttttttggaagtttttagattatttatgaatatataaataaatatacgt
14421  aatataatatatattatataaatcaacgacacggcttttcattttggtgatgatcaatcttattgttct
14491  ctaattgatttttttgtacaataaagatgtatccagttttccagataaagaatttagtttgttatttct
14561  gccccattaaaataagtacggtattcgacaataccacatagtatatacccaaagacggtggattggaca
14631  tgggtgcatggatttcggtactgttgtcatgct
```

Fig. 2. Contined.

RNA isolation and Northern blot analysis

Total RNAs were isolated from the anterior, middle and posterial silk gland of the fifth instar larvae of Baekok-Jam by using the Total RNA Extraction Kit (Promega). Total RNAs (10 μg/lane) were denatured by glyoxalation (McMaster and Carmichael, 1977), transferred onto a nylon blotting membrane (Schleicher & Schuell, Dassel, Germany) and hybridized at 42°C with a probe in a buffer containing 2 × PIPES, 50% formamide, 1% SDS and blocking agent (Boehringer Mannheim). The probe used to detect the L-chain gene transcripts was a 412 bp including exon 7 gene labeled with [α-^{32}P] dCTP using the Prime-It II Random Primer Labeling Kit (Stratagene). After hybridization, the membrane filter was washed three times for 30 min each in 0.1% SDS and 0.2 × SSC (1 × SSC is 0.15 M NaCl and 0.015 M sodium citrate) at 65°C, and then exposed to X-ray film.

Results and Discussion

In the present work, we determined the complete sequence of fibroin L-chain gene from the silkworm Baekok-Jam, *B. mori* (Korean breed). The expected PCR products of Baekok-Jam fibroin L-chain gene were amplified in genomic DNA from the posterior silk gland of fifth instar larvae. The molecular sizes of PCR products were 8.5 kb (front region) and 7.2 kb (rear region), respectively. The genomic DNA structure, PCR primers and amplified PCR products of Baekok-Jam fibroin L-chain gene are shown in Fig. 1.

Fig. 2 showed the complete nucleotide sequence and exon-intron structure of the silkworm Baekok-Jam fibroin L-chain gene. The Baekok-Jam fibroin L-chain gene was composed of exon 1 that encodes 11 amino acid residues, exon 2 of 27 amino acid residues, exon 3 of 67 amino acid residues, exon 4 of 47 amino acid residues, exon 5 of 29 amino acid residues, exon 6 of 49 amino acid residues and exon 7 of 31 amino acid residues. The first exon contained a 41 bp noncoding sequence and a 36 bp coding sequence. The first intron was the largest with 8,145 bp long, which occupied approximately 60% of the fibroin L-chain gene. In the 5-flanking region, there were a canonical TATA box sequence (Corden *et al.*, 1980) between nucleotide −35 and −25, and a 5'-GTCAATTT-3' sequence starting at nucleotide −128, which might serve as promoters for the transcription of fibroin L-chain. In the 3'-noncoding region, there was a 5'-AATAAA-3' sequence starting at nucleotide 14,510 which seemed to serve as a polyadenylation signal (Proudfoot and Brownlee, 1976). As

```
J139      MKPIFLVLLVATSAYAAPSVTINQYSDNEIPRDIDDGKASSVISRRWDYVDDTDKSIAILNVQEIL   66
BaekOk    ----------V----------------------------------A--------------------

J139      KDMASQGDYASQASAVAQTAGIIAHLSAGIPGDACAAANVINSYTDGVRSGNFAGFRQSLGPFFGH  132
BaekOk    -------------P-----------------------------------------I----------

J139      VGQNLNLINQLVINPGQLRYSVGPALGCAGGGRIYDFEAAWDAILASSDSSFLNEEYCIVKRLYNS  198
BaekOk    ------------------------------------------------------------------

J139      RNSQSNNIAAYITAHLLPPVAQVFHQSAGSITDLLRGVGNGNDATGLVANAQRYIAQAASQVHV*   262
BaekOk    ------------------------------------------------------------------
```

Fig. 3. Alignment of the amino acid sequences of the silkworm Baekok-Jam and J139 fibroin L-chain genes. Identical amino acid sequences are indicated by dots.

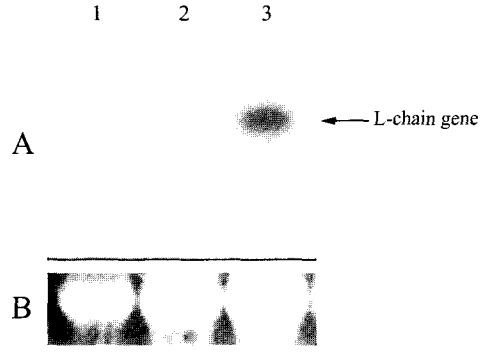

Fig. 4. Northern blot analysis of the silkworm Baekok-Jam fibroin L-chain gene. Total RNAs were extracted from anterior (lane 1), middle (lane 2) and posterior (lane 3) silk gland of the fifth instar larvae. The RNAs were separated by 1% formaldehyde agarose gel electrophoresis (B), transferred on to a nylon membrane, and hybridized with the probe (A). Transcripts of the fibroin L-chain gene are indicated by arrow.

shown in Fig. 2, splice-junctions for all the exons conformed to the GT/AG rule (Breathnach *et al.*, 1978). Sequences around the 5'- and the 3'- splice sites matched partially with the consensus sequences for those of invertebrates (Shapiro and Senapathy, 1987). In this result, these compositions were almost identical to the Japanese breed J139 fibroin L-chain gene (Kikuchi *et al.*, 1992).

Deduced amino acid sequences of Baekok-Jam fibroin L-chain gene are presented in Fig. 3. Its 14,663 nucleotides long with an open reading frame of 786 nucleotides that encodes a protein of 262 amino acid residues with a molecular mass of approximately 26,000 dalton. The Baekok-Jam fibroin L-chain had 98.5% protein sequence identity to J139. However, the deduced protein sequence of Baekok-Jam fibroin L-chain gene was different from amino acid residues at four positions, 11, 46, 80 and 123 in J139. Two cysteine residues in fibroin L-chain were detected from deduced protein sequences at two positions, 160 and 190, suggesting that these cysteine residues would induced H-L fibroin structure by forming a disulfide bond (Mori *et al.*, 1995; Takei *et al.*, 1987).

Silk protein of silkworm, *B. mori* is consists of fibroin H-chain, L-chain and glycoprotein P25. In these elements, the fibroin L-chain secreted into the lumen of posterior silk gland from the surrounding posterior silk gland cells as a molecular complex consisting of an H-chain (Tanaka *et al.*, 1993). In the present study, we analyzed the tissue-specific expression of the fibroin L-chain gene from the silk gland of fifth instar larvae. Northern blot analysis was carried out using the 412 bp including exon 7 gene amplified by PCR in this study as a probe. A hybridization signal was detected as a single band in mRNA from the posterior silk gland, but not detected in the anterior and middle silk gland (Fig. 4). This result is in good agreement with previous finding in that the fibroin L-chain gene is tissue-specifically expressed in the posterior silk gland (Takei *et al.*, 1987).

In conclusion, we report the complete nucleotide sequence and exon-intron structure of the fibroin L-chain gene from the silkworm Baekok-Jam, a recommended variety in Korea. This gene is identical in its protein sequence and length to a Japanese breed J139 fibroin L-chain gene, except for amino acid residues at four positions, 11, 46, 80 and 123. Molecular characterization of the silkworm Baekok-Jam fibroin L-chain gene will expand information on the silkworm fibroin genes.

Acknowledgements

This work was supported by grant for the G-7 project from the Ministry of Science and Technology, Korea.

References

Breathnach, R., C. Benoist, K. OHara, F. Gannon and P. Chambon (1978) Ovalbumin gene: evidence for a leader sequence in the mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA* **75**, 4853-4857.

Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger and P. Chambon (1980) Promoter sequences of eukaryotic protein-coding genes. *Science* **209**, 1406-1414.

Hui, C. C., K. Matsuno and Y. Suzuki (1990a) Fibroin gene promoter contains a cluster of homeodomain binding sites that interact with three silk gland factors. *J. Mol. Biol.* **20**, 651-670.

Hui, C. C., Y. Suzuki, Y. Kikuchi and S. Mizuno (1990b) Homeodomain binding sites in the 5' flanking region of the *Bombyx mori* silk fibroin light-chain gene. *J. Mol. Biol.* **5**, 395-398.

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

Kikuchi, Y., K. Mori, S. Suzuki, K. Yamaguchi and S. Mizuno (1992) Structure of the *Bombyx mori* fibroin light-chain-encoding gene: upstream sequence elements common to the light and heavy chain. *Gene* **15**, 151-158.

Kimura, K., F. Oyama, H. Ueda, S. Mizuno and K. Shimura (1985) Molecular cloning of the fibroin light chain complementary DNA and its use in the study of the expression of the light chain gene in the posterior silk gland of *Bombyx mori*. *Experientia* **41**, 1167-1171.

McMaster, G. K. and G. G. Carmichael (1977) Analysis of single- and double-stranded nucleic acids on polyacryamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA* **74**, 4835-4838.

Mori, H., M. Yamao, H. Nakazawa, Y. Sugahara, N. Shirai, F. Matsubara, M. Sumida and T. Imamura (1995) Transovarian transmission of a foreign gene in the silkworm, *Bombyx mori*, By *Autographa californica* nuclear polyhedrosis virus. *Bio-Tchnology* **13**, 1005-1007.

Proudfoot, N. J. and G. G. Brownlee (1976) Non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**, 211-214.

Shapiro, M. B. and P. Senapathy (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155-7174.

Takei, F., Y. Kikuchi, A. Kikuchi, S. Mizuno and K. Shimura (1987) Further evidence for impotence of the subunit combination of silk fibroin polypeptides in its efficient secretion from the posterior silk gland cells. *J. Cell Biol.* **105**, 175-180.

Tanaka, K., K. Mori and S. Mizuno (1993) Immunological identification of the major disulfide-linked light component of silk fibroin. *J. Biochem.* **114**, 1-4.

Tanaka, K., S. Inoue and S. Mizuno (1999) Hydrophobic interaction of P25, containing Asn-linked oligosaccharide chains, with the H-L complex of silk fibroin produced by *Bombyx mori*. *Insect Biochem. Mol. Biol.* **29**, 269-276.

Yamaguchi, K., Y. Kikuchi, T. Takagi, A. Kikuchi, F. Oyama, K. Shimura and S. Mizuno (1989) Primary structure of the silk fibroin light chain determined by cDNA sequencing and peptide analysis. *J. Mol. Biol.* **210**, 127-139.

Zhou, C. Z., F. Confalonieri, N. Medina, Y. Zivanovic, C. Esnault, T. Yang, M. Jacquet, J. Janin, M. Duguet, R. Perasso. and Z. G. Li (2000) Fine organization of *Bombyx mori* fibroin heavy chain gene. *Nucleic Acids Res.* **28**, 2413-2419.