Full Length Research Paper

Expression analysis and characteristics of hypothetical protein BmLOC778500 from silkworm, *Bombyx mori*

Wenping Zhang[1], Hanglian Lan[2], Tao Wu[1], Caiying Jiang[1], Zhengbing Lv[1], Zuoming Nie[1], Xiangfu Wu[1,3] and Yaozhou Zhang[1]*

[1]Institute of Biochemistry, College of Life Sciences, Zhejiang Science and Technology University, Hangzhou 310018, China.
[2]Institute of Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou 310018, China.
[3]Institute of Biochemistry and Cell Biology, Chinese Academy of Science, Shanghai 200031, China.

We cloned a cDNA from silkworm pupal cDNA library and found that, it encodes LOC778500 protein (a "hypothetical protein" in the National Center for Biotechnology Information (NCBI) database). Thus, we named it *BmLOC778500* (*Bombyx mori LOC778500*). It contains an open reading frame of 489 bp encoding a polypeptide of 162 amino acid residues with a predicted molecular mass of 19.4 kDa. The gene codes for a protein homologous to hypothetical proteins of *Tribolium castaneum*, *Pediculus humcorporisanus* and *Drosophila erecta* among others. Purified BmLOC778500 was used to generate anti-BmLOC778500 polyclonal antibody for determining tissue distribution and sub-cellular localization of BmLOC778500. Western blotting analysis revealed that, BmLOC778500 proteins existed in Malpighian tubule of fifth-instar larva of *B. mori*. Immunostaining indicated that, BmLOC778500 proteins are mainly located in cytosol and nucleus as well as in not-dividing cells, whereas are distributed in cell membrane in dividing cells. These results have provided important information for further function studies of BmLOC778500 protein.

Key words: BmLOC778500, *Bombyx mori*, expression analysis, tissue distribution, sub-cellular localization.

INTRODUCTION

The domesticated silkworm (*Bombyx mori*) has long been used as a model system for basic studies because of its large body size, ease of rearing in the laboratory and economic importance in sericulture (Mita et al., 2004). With the development of genomic and biotechnological tools, *B. mori* has also become an important bioreactor for production of various recombinant proteins of bio-medical interest (Wang et al., 2005; Zhang et al., 2006; Shimomura et al., 2009; Zhang et al., 2009). A 3× coverage draft sequence (Mita et al., 2004) and a draft sequence with 5.9× coverage of the silkworm genome (Xia et al., 2004) were independently reported in 2004. The size of the genome is 428.7 Mb and the gene numbers are 21,302 (Yu et al., 2009). Comparative genomics shows that, a substantial fraction of the genes in sequenced genomes encodes 'conserved hypothetical' proteins, that is, those that are found in organisms from several phylogenetic lineages but have not been functionally characterized (Galperin and Koonin, 2004). Gene annotation indicates that, 'conserved hypothetical proteins' often represent more than half of the potential protein-coding regions of a genome (Galperin and Koonin, 2004; Roberts, 2004; Sivashankari and Shanmughavel, 2006). Many bioinformatics methods to predicate the function of genes in any given genome have difficulties, which shows more than 70% prediction accuracy (Bork, 2000). Roberts (2004) proposed to establish a database for those

---

Abbreviations: ORFs, Open reading frame; FCS, fetal calf serum; BLAST, Basic Local Alignment Search Tool; PCR, polymerase chain reaction; IPTG, isopropylthiogalactoside; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; ELISA, enzyme-linked immunosorbent assay; PVDF, polyvinylidene fluoride; BSA/PBS, bovine serum albumin/phosphate buffered saline; DAPI, 4'-6-diamidino-2-phenylindole; EST, expressed sequenced tag.

conserved hypothetical and non-hypothetical genes and then to offer those proteins to the biochemical community as potential targets for function research. This phenomenon also exists in the field of silkworm genome research. Two databases about silkworm genome were created by Zhang and his colleagues (Zhang et al., 2009). The first was based on a 9×draft sequence of the silkworm genome and contained 14,632 putative proteins. The second was based on a *B. mori* pupal cDNA library containing 3,187 putative proteins of at least 30 amino acid residues in length (Zhang et al., 2009). In order to elucidate how many genes are expressed and what function of those genes are involved in silkworm metamorphosis, 1,659 cDNAs were screened from the cDNA library of silkworm pupa constructed by our library (Zhang et al., 2007), of which 837 were entered into Unigene database (National Center for Biotechnology Information, NCBI) (Yu et al., 2009).

The deduced amino acids matched to BmLOC778500 gene was referred to 'hypothetical protein LOC778500' from *B. mori* in the NCBI RefSeq database, thus, we named it BmLOC778500. No protein encoded by 'BmLOC778500 gene' has been reported yet. However, four cDNA clones for this gene are obtained from four different silkworm cDNA libraries, respectively, such as pupal (GenBank accession: CK516718), testis (GenBank accession: CK535437), orary (GenBank accession: BY922303) and Malpighian tubule (GenBank accession: BY915773). It is likely that, all these four cDNA encode same protein. It will be very helpful to determine whether the open reading frame (ORFs) of BmLOC778500 encode a functional protein and what the normal function of the encoded protein is in the silkworm.

In this study, we identified a silkworm BmLOC778500 complementary DNA from a pupal cDNA library (GenBank accession: DN236938). Using bioinformatics analysis, we determined the evolutionary relationship and degree of conservation between BmLOC778500 and other orthologs. Western blotting analysis was used to determine tissue distribution of BmLOC778500 in the different tissues of the fifth instar larvae. Subcellular localization of BmLOC778500 in Bm5 cells (silkworm ovary cell line) was analyzed by immunocytochemistry. Our results lay the foundation for further studies on the functional roles played by BmLOC778500 in the silkworm.

MATERIALS AND METHODS

The *B. mori* strain used in this study is the progeny of Jingsong×Haoyue. Silkworms were reared on mulberry leaves under standard conditions. Heads, silk glands, midguts, fat bodis, haemolymph, Malpighian tubules, trachea and epidermises from fifth instar larvae were dissected, frozen immediately in liquid nitrogen and stored at –80°C. Bm5 cells (a gift from Prof. Z. F. Zhang) were cultured in TC-100 medium (Sigma) supplemented with 10% fetal calf serum (FCS, Gibco BRL) at 27°C. The *Escherichia coli* strains TG1 and BL21 (DE3) were grown at 37°C in

LB medium, pH 7.5 (5 g of yeast extract, 10 g of tryptone and 10 g of NaCl per liter). General reagents were of analytical grade.

Bioinformatics analysis

The BmLOC778500 gene was identified from the *B. mori* pupa cDNA library (Zhang et al., 2007). Genomic structure for this gene was analyzed by comparing the cDNA sequence with corresponding genomic DNA sequences downloaded from the Beijing Genomics Institute (http://silkworm.genomics.org.cn). Analysis of the similarity of nucleotide and protein sequences was performed using the Basic Local Alignment Search Tool (BLAST) (McGinnis and Madden, 2004) algorithm from NCBI (http://www.ncbi.nlm.nih.gov/). The characteristics of BmLOC778500 gene were analyzed using DNAstar software (DNAStar, Madison, Wis., USA) (Song and Goodman, 2002). The orthologous sequences used for multiple sequence alignments were obtained from NCBI. Sequence alignments were done in BioEdit version 7.0.1 (Hall, 1999) by use of ClustalW multiple alignment (Thompson et al., 1994). The software phyml version 2.4.4 was used to construct a phylogenetic tree by applying results determined by multiple sequence alignments using a maximum-likelihood method; the bootstrap was set to 1,000 (Guindon and Gascuel, 2003).

Expression and purification of recombinant BmLOC778500

The entire sequence of a cDNA containing the complete ORF of BmLOC778500 was amplified by polymerase chain reaction (PCR) using the primer pair 5'-CGCGGATCCAAAATGGGAAATCTTAAG-3'/5'-CCGCTCGAGCTACATCAAATTTTTC-3' (the restriction endonuclease sites of *Bam*H I and XhoI are respectively underlined). The resulting PCR product was purified using the PCR rapid purification kit (BioDev-Tech, China). After digestion with *Bam*H I and *Xho* I, the purified product was subcloned into the expression vector pET28a (Invitrogen, Carlsbad, CA, USA) using T4 DNA ligase (Promega, USA) and transformed into *E. coli* TG1 cells for screening purposes. The recombinant plasmid was named as pET-28a-*BmLOC778500*. A positive clone was identified by PCR and double digestion of the purified plasmid followed by electrophoresis on a 1% agarose gel. The construct was subsequently verified by sequencing with the BigDye terminator method on an ABI PRISM 3130XL/A automated sequencer (Applied Biosystems, USA).

The recombinant plasmid pET-28a-*BmLOC778500* was transformed into *E. coli* cells (strain BL21(DE3)) and expression of the His-tagged fusion protein was induced using isopropylthio-galactoside (IPTG). The recombinant fusion protein His-BmLOC778500 was extracted from the bacteria and purified using Ni NTA superflow affinity chromatography (QIAGEN, Japan) according to the manufacturer's instructions. The purified recombinant protein was determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie brilliant blue R-250 staining.

Antibody preparation and western blotting analysis

A male New Zealand white rabbit was injected with 500 μg of purified recombinant protein His-BmLOC778500 emulsified with Freund's complete adjuvant (Sigma) (Bomford, 1980a, b). Three booster doses were given at intervals of 7 days with the same dose of antigen and Freund's incomplete adjuvant (Sigma) (Bomford, 1980a, b). On the 7th day after the last immunization, blood was collected and the serum fraction was isolated. The antiserum was purified using a HiTrap Protein A HP column (GE Lifesciences, USA) following the manufacturer's instructions. An enzyme-linked

```
 -24                                                 CTTTATTTTACCATTATAATAAAA
   1     ATGGGAAATCTTAAGAAAAATTTAGAGAAACTAAAACACCCTAACAGTAGAAAAACTATG
   1      M  G  N  L  K  K  N  L  E  K  L  K  H  P  N  S  R  K  T  M
  61     ACATTAGCTAAGAAATTGAAGAAATCCGAAAAAAAGGATCAAAGCAAACTGGGTACTCAT
  21      T  L  A  K  K  L  K  K  S  E  K  K  D  Q  S  K  L  G  T  H
 121     ATTAAGCAAAATTTAATCGGCGAAAAAATAATGTGGTTCAAAGAGAGAATACCAGAGGAA
  41      I  K  Q  N  L  I  G  E  K  I  M  W  F  K  E  R  I  P  E  E
 181     TGTGAAGTCTTTGATAAAGAACAGACACTGAAATTAATCGAAAGCTATTTAAGTCGTTTC
  61      C  E  V  F  D  K  E  Q  T  L  K  L  I  E  S  Y  L  S  R  F
 241     GATGAAGAACTAGAACAGATCGCATTAAAAAATTCAATTGGGCAAAGGAAAAACCGTCAA
  81      D  E  E  L  E  Q  I  A  L  K  N  S  I  G  Q  R  K  N  R  Q
 301     CACGCGAGTAGAGAAGACGTTATTAATATGACGAGGAAAAAGGAACAGGACGAATTTGAG
 101      H  A  S  R  E  D  V  I  N  M  T  R  K  K  E  Q  D  E  F  E
 361     ACTTGCGGTATAGAAATACCGGATTTAATGGACCCACAACAAATGAACGTGGTGAGGAAT
 121      T  C  G  I  E  I  P  D  L  M  D  P  Q  Q  M  N  V  V  R  N
 421     TGGAATGGTGAACTGAGATTTCTTCACCATTTTAAGTTAAAAAGATTTACTAAGAAAAAT
 141      W  N  G  E  L  R  F  L  H  H  F  K  L  K  R  F  T  K  K  N
 481     TTGATGTAGTTTCTTTAAGGTTCTAACTCTTCTTTAGAAACCCAGGTGTAAGAGATTAGA
 161      L  M  *
 541     GATCCCTTATTTGAGTAAAATAGTGTCAGCTACAACCTCAAGAACAACAAAGTCAGCCCC
 601     AACATATTATAGAGCTATATAAGCTATATTAAAAAATAAATAGAACCTTTGTAGACGTTA
 661     ACTAACAACAAACAAATATTGAAACGGCATTGTTATTGAACTGTATATTGTTATTGAACT
 721     GTATATTCTTATTGAACTGTATATTGTTATTGTATGCTTAAAATGTTAAATTAAAATTT
 781     ATATGCTTAAGGGTCTTAAAAAAAAATAAATTTTACTATTTTGTTTTAAAAAAAAAAAAAA
```

Figure 1. Nucleotide sequence and deduced amino acid sequence of the silkworm BmLOC778500 cDNA. The deduced amino acid sequence is shown in single-letter code below the second base of each codon; stop codon (TAG) is indicated by asterisk (*). The classical polyadenylation signal is underlined.

immunosorbent assay (ELISA) was used to determine the titer of the polyclonal antibody. The specificity of the polyclonal antibody was determined by western blotting analysis: the purified polyclonal antibody was diluted 1: 500 and incubated with the immunoblot at room temperature for 2 h; the goat anti-rabbit secondary antibody (LI-COR, USA) was diluted 1:1000. This purified antibody was used for immunoblotting, immunohistochemistry and immunofluorescence.

Tissue distribution analysis of BmLOC778500 by western blotting

For western blotting analysis, different tissues from fifth instar larvae of *B. mori* were ground to a fine powder in liquid nitrogen. Powders were suspended in protein extraction lysate buffer (50 mM Tris pH 8.0, 100 mg/l PMSF, 5 µg/ml Aprotin, 0.15 M NaCl, 5mM EDTA, 0.5 %NP-40(v/v), 1 mM dithiothreitol, 5 g/l deoxycholate sodium). After 30 min on ice, homogenates were centrifuged at 12,000 g for 15 min at 4°C. Protein concentrations were determined using the Bradford method. Equal amounts of proteins were separated by SDS-PAGE with 10% polyacrylamide gel and electro-transferred onto polyvinylidene fluoride (PVDF) membranes (Millipore). Membranes were blocked overnight at 4°C in 3% bovine serum albumin/phosphate buffered saline (BSA/PBS) and simulta-neously incubated with anti BmLOC778500 polyclonal antibody (1:500 dilutions) for 2 h at room temperature with gentle shaking. Then, horseradish peroxidase-conjugated anti-rabbit total IgG from goat (Sigma-Aldrich, USA) secondary antibodies were added and allowed to incubate for 2 h at room temperature. Immunoblotting bands were detected with dialkyldimethyl-ammonium bromide (DAB).

Subcellular localization analysis of BmLOC778500 by immunocytochemistry

Bm5 cells were cultured overnight on glass cover dishes which were used specifically for confocal microscopy. After the culture medium had been removed, cells were washed for 10 min three times in PBS and fixed (PBS, pH 7.2, 4% poly-formaldehyde, 0.1%

Triton X-100) at room temperature for 15 min. The fixed cells were blocked in 3% BSA in PBS at room temperature for 2 h followed by three 10 min washes in PBST (0.05% Tween-20 in PBS). Cells were then incubated with anti-BmLOC778500 polyclonal antibody (diluted 1:500 in blocking buffer) at 4°C overnight; cells were simultaneously incubated with control serum. The control serum was obtained from the rabbit before immunization with the antigen (BmLOC778500). After three 10 min washes in PBST, cells were incubated with Cy3-labeled goat anti-rabbit IgG (diluted 1:500; Promega) at 37°C for 2 h and were then washed three times for 10 min in PBST. Cells were then incubated with 4′-6-diamidino-2-phenylindole (DAPI) (1 g/ml in PBS) at room temperature for 10 min. After washing once with PBST, cells were examined under a Nikon ECLIPSE TE2000-E Confocal laser scanning microscopy (Nikon, Japan). Images were analyzed using EZ-C1 software (Nikon, Japan).

RESULTS

Bioinformatics analysis

The cDNA sequence of BmLOC778500 (accession number DN236938) contained an ORF of 489 bp encoding a polypeptide of 162 amino acids with a predicted molecular weight of 19.4 kDa and a theoretical isoelectric point of 9.58 (Figure 1). Structure analysis of the cDNA indicates that, the genomic sequence of BmLOC778500 contain one exon. When submitted to BLAST, the deduced amino acids sequence displayed some homology to the hypothetical proteins and their functions were unknown. By sequence alignment, BmLOC778500 was shown to have 56.4% identity with the related hypothetical protein from *Tribolium castaneum*, 51% identity with *Pediculus humcorporisanus* and 46.6% identity with *Drosophila erecta,* respectively (Figure 2). To determine biological
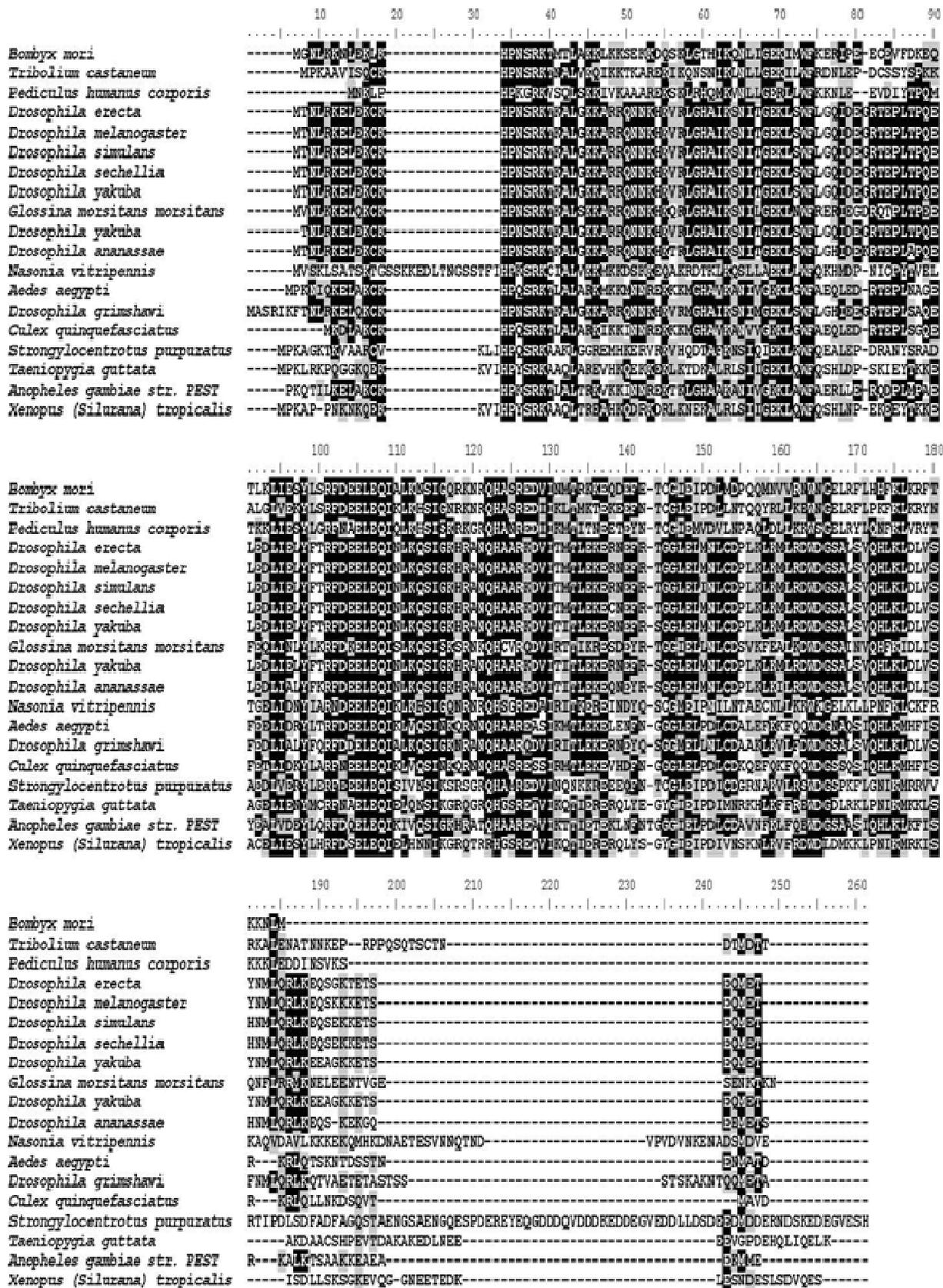
Figure 2. Multiple sequence alignment of BmLOC778500. Black-shaded amino acids indicate identical residues; gray-shaded amino acids indicate similar residues.
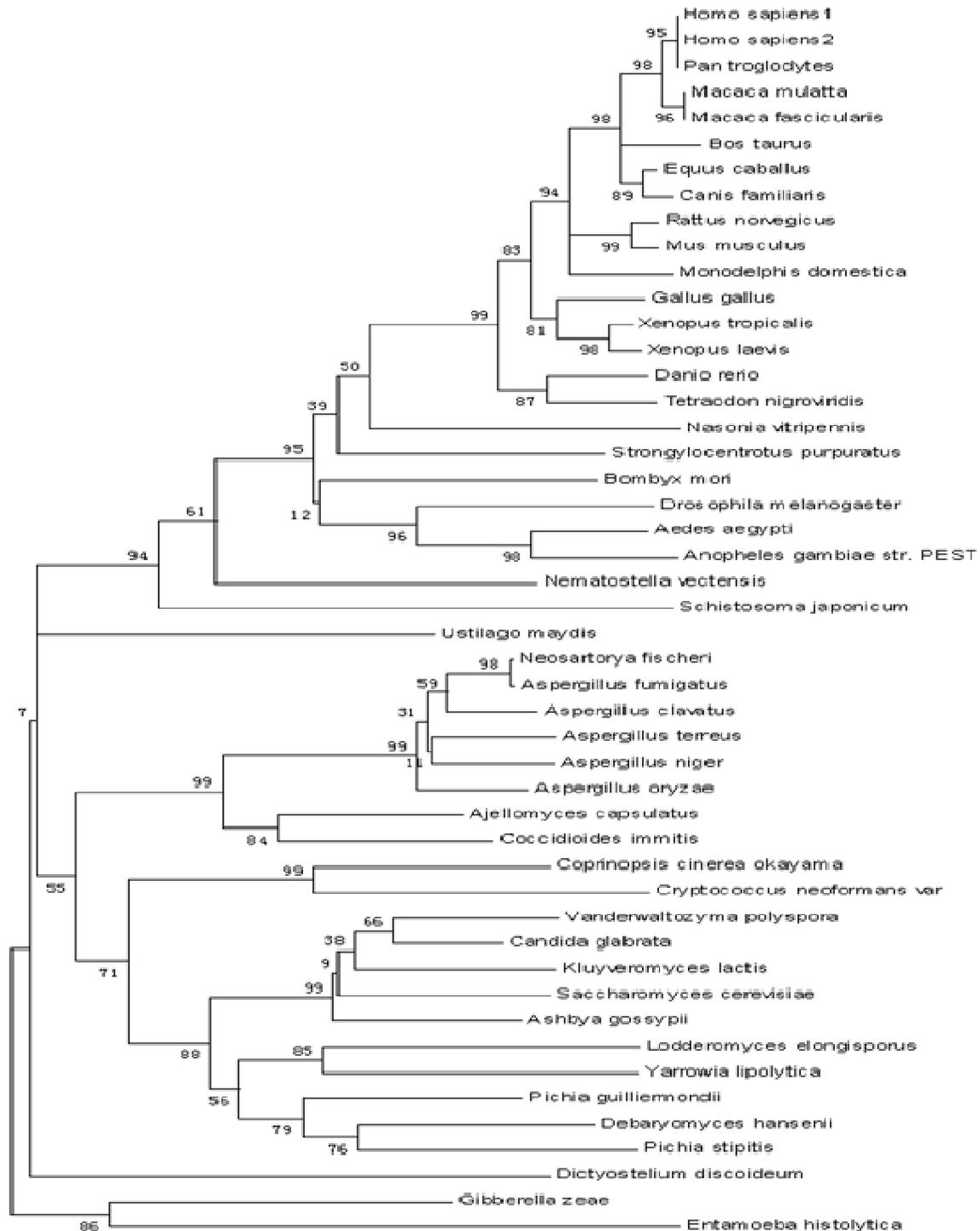
Figure 3. Phylogenetic tree of BmLOC778500.

characteristics of BmLOC778500 protein in evolution, a phylogenetic tree was constructed. As shown in Figure 3, *B. mori* and *Drosophila* belong to the same branch, with close genetic relationship during evolution. The results showed that, the protein BmLOC778500 is highly conserved in vertebrates and invertebrates.

Prokaryotic expression of BmLOC778500

The recombinant fusion protein His-BmLOC778500 was detected in bacterial lysates by SDS-PAGE. The results showed that, His-BmLOC778500 was mainly located in the lysate supernatant fraction of the cells. The molecular
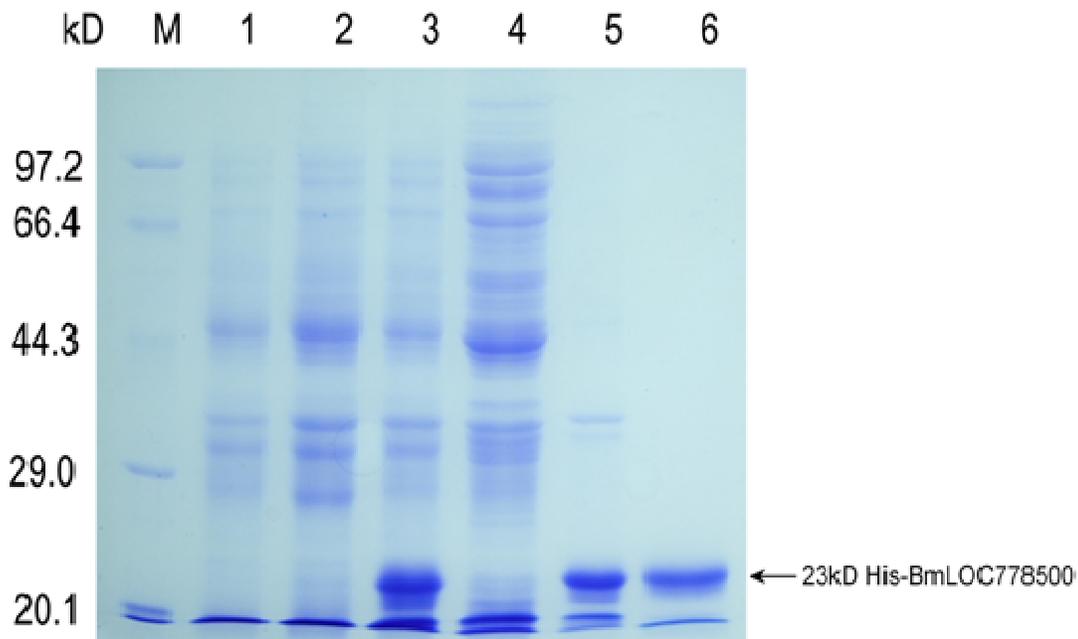
Figure 4. Expression and purification of the fusion protein His-BmLOC778500. The fusion protein His-BmLOC778500 purified by Ni-NTA were analyzed by SDS-PAGE and visualized by Coomassie brilliant blue R-250 staining. M, Molecular weight markers with the masses shown, in kilodalton, to the left; 1, negative control; 2, the total protein of lysate from uninduced BL21(DE3)-pET-28a-*BmLOC778500*; 3, the total protein of lysate from BL21(DE3)-pET-28a-*BmLOC778500* after inducing with IPTG for 5 h; 4, the lysate precipitation from BL21(DE3)-pET-28a-*BmLOC778500* after inducing with IPTG for 5 h; 5, the lysate supernatant from BL21(DE3)-pET-28a-*BmLOC778500* after inducing with IPTG for 5 h; 6, the purified proteins His-BmLOC778500.

weight of the fusion protein, His-BmLOC778500 is 23 kDa as determined by electrophoresis, which is consis-tent with the calculated value (BmLOC778500: 19.4 kDa and His tag: 3.6 kDa) (Figure 4).

Specificity of anti-BmLOC778500 polyclonal antibody

The anti-BmLOC778500 polyclonal antibody was gene-rated in an immunized New Zealand rabbit and was purified by Protein A HP column chromatography (Amersham). The anti-BmLOC778500 polyclonal anti-body titer, as determined by ELISA, was 1:32000 and western blotting analysis indicated that, the antibody reacted specifically with His-BmLOC778500 (Figure 5).

Tissue distribution of BmLOC778500

Rabbit antiserum was raised against purified recombinant protein and the anti-BmLOC778500 antibody was purified by protein A chromatography. Purified anti-BmLOC-778500 antibody was used to confirm the expression of BmLOC778500 in different tissues (Heads, silk glands, midguts, fat bodis, haemolymph, Malpighian tubules, trachea and epidermises) of fifth-instar larvae of *B. mori*

by immunoblotting (Figure 6). Positive signals were only found in Malpighian tubules.

Subcellular localization of BmLOC778500

The treated Bm5 cells were examined under a Nikon ECLIPSE TE2000-E Confocal Microscope and images were analyzed using EZ-C1 software. Cy3-labeled goat anti-rabbit IgG shows red fluorescence under a wavelength of 550 nm and DAPI-stained nuclei emit blue fluorescence when stimulated with light having a wave-length of 353 nm. The results indicate that, BmLOC-778500 proteins are located in both the cytoplasm and nucleus of Bm5 cells but are primarily found in the cytoplasm when cells are not dividing (Figure 7). In contrast, when cells are dividing, most of BmLOC778500 distribute in the cell membrane. Subcellular localization of BmLOC778500 will offer a clue for biological function analysis of BmLOC778500 protein.

DISCUSSION

BmLOC778500 gene was a novel gene screened out from the silkworm pupal cDNA library. There was only
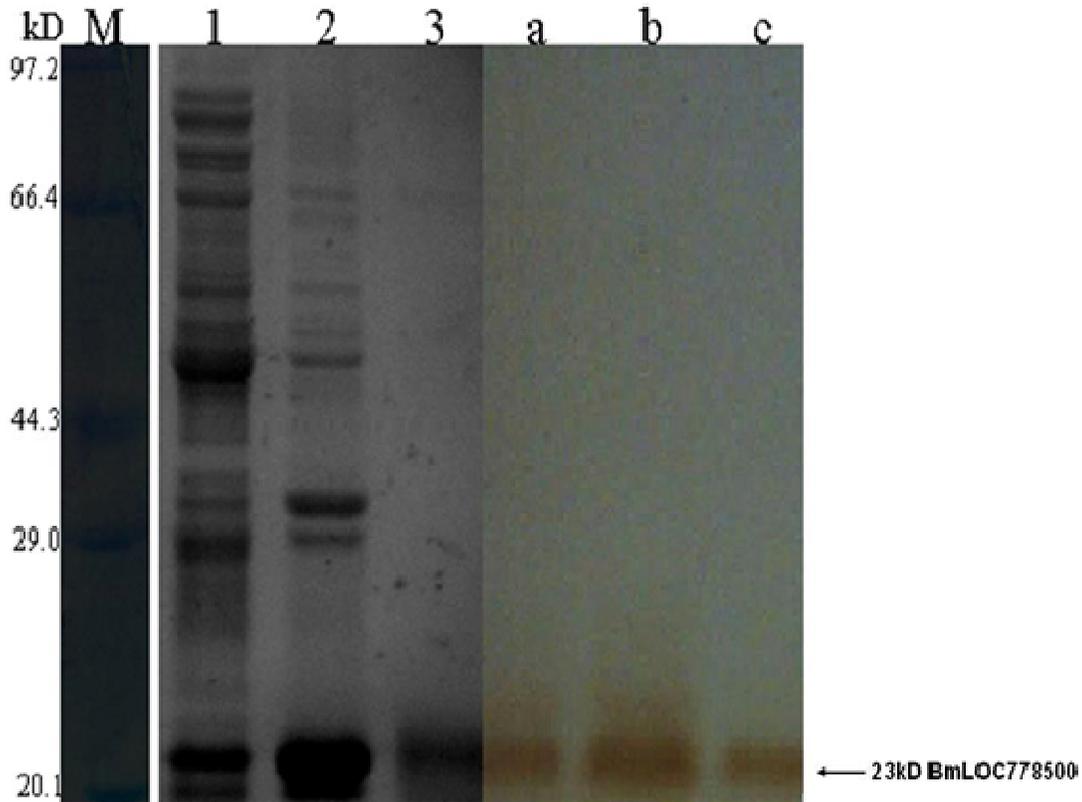
Figure 5. Western blotting analysis of the specificity of anti-BmLOC778500 polyclonal antibody. M, Molecular weight markers with the masses shown, in kilodalton, to the left; 1, the lysate precipitation from BL21(DE3)-pET-28a-*BmLOC778500* after inducing with IPTG for 5 h; 2, the lysate supernatant from BL21(DE3)-pET-28a-*BmLOC778500* after inducing with IPTG for 5 h; 3, the purified proteins His-BmLOC778500; A, western blotting of 1; B, western blotting of 2; C, western blotting of 3.
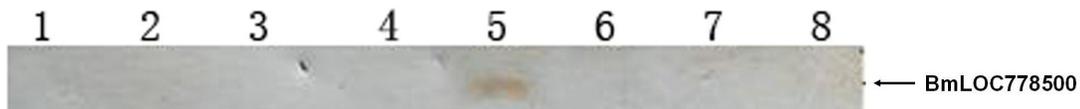

Figure 6. Western blotting analysis of the expression of BmLOC778500 in different tissues of fifth instar larva of *B. mori*. Lane 1, Haemolymph; lane 2, trachea; lane 3, fat bodis; lane 4, midguts; lane 5, Malpighian tubules; lane 6, silk glands; lane 7, epidermises; lane 8 heads. The arrow indicated the position of BmLOC778500.

experimental evidence for mRNA level but no protein(s) level up to now. Therefore, BmLOC778500 protein remained to refer as a "hypothetical protein". By BLAST analysis, the deduced amino acids sequence of BmLOC-778500 displayed homology to the related hypothetical proteins from other species including *T. castaneum*, *P. humcorporisanus* and *D. erecta* (Figure 2). The phylogenetic tree based on BmLOC778500 protein showed that, *B. mori* and *Drosophila* belong to the same branch with close genetic relationship during evolution (Figure 3). So we speculated that, BmLOC778500 gene was insect-specific gene.

In this study, we performed some experiments on BmLOC778500 to analyze its characteristics. Proteins from heads, silk glands, fat bodis, haemolymph, Malpighian tubules, trachea, midguts and epidermises from the fifth-instar larvae of *B. mori* were extracted and analyzed by western blotting. The positive signals were only detected in larva's Malpighian tubules. No signals were detected in other tissues. One possibility is that, the expression of BmLOC778500 in these tissues is too little to be detected or there is no expression in these tissues. However, it might be sure that BmLOC778500 can be expressed in Malpighian tubules of the fifth-instar larvae
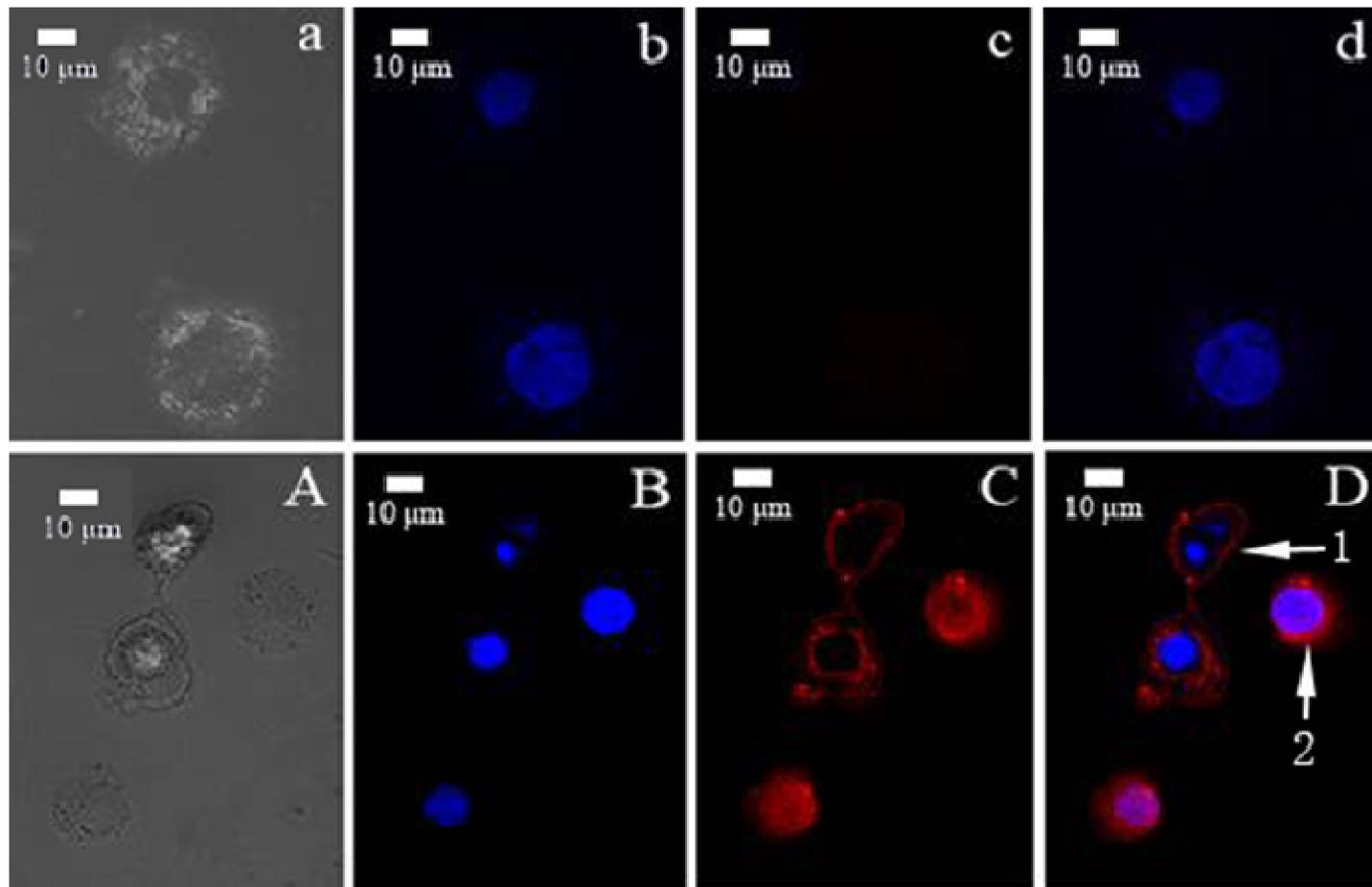
Figure 7. Subcellular localization of BmLOC778500 with Cy3-labled goat ant-rabbit IgG and DAPI. a to d: Negative control group, using negative rabbit serum; a: Bm5 cells at the transmission light; b: DAPI staining; c: Cy3-labeled secondary antibody; d: Merged image of b and c. A to D: Experimental group, using anti-BmLOC778500 polyclonal antibody; A: Bm5 cells at the transmission light; B: DAPI staining; C: BmLOC778500 subcellular localization as indicated by Cy3-labeled secondary antibody; D: Merged image of B and C. Arrow 1 indicated that, BmLOC778500 proteins were mostly located in the cell membrane when Bm5 cells were dividing. Arrow 2 indicated that, BmLOC778500 proteins existed both in the cytoplasm and nucleus when Bm5 cells were not dividing.

of *B. mori*. This is consistent to the result of searching expressed sequenced tag (EST) database of *B. mori* from NCBI. We found an EST (accession number BY915773) from Malpighian tubules of the fifth-instar larvae of *B. mori* is one of 7 sequences matched to BmLOC778500.

Finally, we carried out subcellular localization of BmLOC778500 in Bm5 cells. An interesting phenomenon was found under a Nikon ECLIPSE TE2000-E Confocal microscope. BmLOC778500 proteins were mostly located in the cell membrane when Bm5 cells were dividing (as shown by arrow 1 in Figure 7). In contrast, BmLOC-778500 proteins existed both in the cytoplasm and nucleus when Bm5 cells were not dividing (as shown by arrow 2 in Figure 7). The results suggested that, BmLOC778500 might participate in the activity of cell cycle. However, the actual physiological function and regulatory mechanism of BmLOC778500 in the silkworm needs further determination.

REFERENCES

Bomford R (1980a). The comparative selectivity of adjuvants for humoral and cell-mediated immunity. I. Effect on the antibody response to bovine serum albumin and sheep red blood cells of Freund's incomplete and complete adjuvants, alhydrogel, Corynebacterium parvum, Bordetella pertussis, muramyl dipeptide and saponin. Clin. Exp. Immunol. 39(2): 426-434.

Bomford R (1980b). The comparative selectivity of adjuvants for humoral and cell-mediated immunity. II. Effect on delayed-type hypersensitivity in the mouse and guinea pig, and cell-mediated

immunity to tumour antigens in the mouse of Freund's incomplete and complete adjuvants, alhydrogel, Corynebacterium parvum, Bordetella pertussis, muramyl dipeptide and saponin. Clin. Exp. Immunol. 39(2): 435-441.

Bork P (2000). Powers and pitfalls in sequence analysis: the 70% hurdle. Genome Res. 10(4): 398-400.

Galperin MY,Koonin EV (2004). 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res. 32(18): 5452-5463.

Guindon S, Gascuel O (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52(5): 696-704.

Hall T (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41: 95-98.

McGinnis S, Madden TL (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32(Web Server issue): W20-25.

Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-Okuda K, Yamamoto K, Ajimura M, Ravikumar G, Shimomura M, Nagamura Y, Shin-I T, Abe H, Shimada T, Morishita S, Sasaki T (2004). The genome sequence of silkworm, *Bombyx mori*. DNA Res. 11(1): 27-35.

Roberts RJ (2004). Identifying protein function-a call for community action. PLoS Biol. 2(3): E42.

Shimomura M, Minami H, Suetsugu Y, Ohyanagi H, Satoh C, Antonio B, Nagamura Y, Kadono-Okuda K, Kajiwara H, Sezutsu H, Nagaraju J, Goldsmith MR, Xia Q, Yamamoto K, Mita K (2009). KAIKObase: an integrated silkworm genome database and data mining tool. BMC Genomics. 10: p. 486.

Sivashankari S, Shanmughavel P (2006). Functional annotation of hypothetical proteins - A review. Bioinformation, 1(8): 335-338.

Song F, Goodman RM (2002). OsBIMK1, a rice MAP kinase gene involved in disease resistance responses. Planta. 215(6): 997-1005.

Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22): 4673-4680.

Wang Y, Wu X, Liu G,Cao C,Huang H,Xu Z,Liu J (2005). Expression of porcine lactoferrin by using recombinant baculovirus in silkworm, *Bombyx mori* L., and its purification and characterization. Appl. Microbiol. Biotechnol. 69(4): 385-389.

Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C, Pan G, Xu J, Liu C, Lin Y, Qian J, Hou Y, Wu Z, Li G, Pan M, Li C,Shen Y, Lan X, Yuan L, Li T, Xu H, Yang G, Wan Y, Zhu Y, Yu M, Shen W, Wu D, Xiang Z, Yu J, Wang J, Li R, Shi J, Li H, Li G, Su J, Wang X, Li G, Zhang Z, Wu Q, Li J, Zhang Q, Wei N, Xu J, Sun H, Dong L, Liu D, Zhao S, Zhao X, Meng Q, Lan F, Huang X, Li Y, Fang L, Li C, Li D,Sun Y, Zhang Z, Yang Z, Huang Y, Xi Y, Qi Q, He D, Huang H, Zhang X, Wang Z, Li W, Cao Y, Yu Y, Yu H, Li J, Ye J, Chen H, Zhou Y, Liu B,Wang J, Ye J, Ji H, Li S, Ni P, Zhang J, Zhang Y, Zheng H, Mao B,Wang W, Ye C, Li S, Wang J, Wong GK, Yang H (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science, 306(5703): 1937-1940.

Yu W, Tan A, Zhang Y (2009). Expression analysis and characteristics of Bm-LOC778477 gene, a hypothetical protein from silkworm pupae, *Bombyx mori*. Afr. J. Biotechnol. 8(14): 3192-3198.

Zhang W, Lv Z, Nie Z, Chen G, Chen J, Sheng Q, Yu W, Jin Y, Wu X, Zhang Y (2009). Bioavailability of orally administered rhGM-CSF: a single-dose, randomized, open-label, two-period crossover trial. PLoS One. 4(5): e5353.

Zhang Y, Chen J, Lv Z, Nie Z,Zhang X, Wu X (2006). Can 29kDa rhGM-CSF expressed by silkworm pupae bioreactor bring into effect as active cytokine through orally administration? Eur. J. Pharm. Sci. 28(3): 212-223.

Zhang Y, Xia Q,Xu J,Chen J,Nie Z,Wang D,Zhang W,Chen J,Zheng Q, Chen Q (2009). Aligning the proteome and genome of the silkworm, *Bombyx mori*. Funct Integr Genomics. 9(4): 447-454.

Zhang Y, Chen J, Nie Z, Lv Z, Wang D, Jiang C, He P, Liu L, Lou Y, Song L, Wu X (2007). Expression of open reading frames in silkworm pupal cDNA library. Appl. Biochem. Biotechnol. 136: 327-343.