Int. J. Indust. Entomol. Vol. 9, No. 2, 2004, pp. 167~171

International Journal of
Industrial Entomology

Characterization of Expressed Sequence Tags (ESTs) Generated from the *Bombyx mandarina* Whole Larvae and Molecular Cloning of Serine Protease Homologue Gene

Jae Sam Hwang*, Eun Young Yun, Tae Won Goo, Iksoo Kim, Kwang Ho Choi, Su Il Seong[1], Keun Young Kim, Sang Mong Lee[2] and Seok Woo Kang

Department of Agricultural Biology, National Institute of Agricultural Science and Technology, Suwon 441-100, Korea.
[1]Department of Life Science, Collage of Natural Science, The University of Suwon, Hwaseong 445-743, Korea.
[2]Department of Sericultural and Entomology, Faculty of Agriculture, Miryang National University, Miryang 627-130, Korea.

(Received 13 July 2004; Accepted 10 September 2004)

We constructed an oligo-d(T) primed directional cDNA library from the *Bombyx mandarina* whole larvae. In an effort to isolate genes expressed in the *B. mandarina*, 227 expressed sequence tags (ESTs) were generated by single-pass sequencing from the cDNA library. Sequence analysis showed that 107 clones (47.1%) were classified into known genes and 120 clones (52.9%) were novel transcripts, which are unknown for their function. Of the 107 known genes, the most abundant gene was found to be actin and followed by serine protease in the expression profile. Among these clones, a serine protease homolog (BmSP) which is a class of proteolytic enzymes isolated. Full-length sequence of the BmSP cDNA clone was 922 bp in length and has an open reading frame of 276 amino acids. The conserved histidine, aspatic acid and serine residues forming the catalytic center as well as cysteine residues contributing to three disulphide bonds also were found in Bmsp gene. mRNA expression analysis revealed a high and specific expression of the gene only in midgut tissue, suggesting that BmSP gene is closely associated with the expression of digestive enzyme.

Key words: ESTs, *Bombyx mandarina*, Serine protease, Midgut

*To whom correspondence should be addressed.
Department of Agricultural Biology, National Institute of Agricultural Science and Technology, Suwon 441-100, Korea. Tel: +82- 31- 290-8573; E-mail: hwangjs@rda.go.kr

Introduction

The cDNA approaches toward genome studies have provided the comprehensive information on the expression pattern of genes in a variety of tissues. Thus, the expressed sequence tags (ESTs) generated by single-pass sequencing of cDNAs have been used as a starting point for the functional and structural analyses of the entire genome of organisms (Adams *et al.*, 1995; Hiller *et al.*, 1996; Kirkness *et al.*, 1996). The recovery of single-passed sequences from randomly selected cDNA clones has been pursued as a relatively inexpensive and rapid means to access many of the expressed genes of an organism (Milner and Sutcliffe, 1983; Putney *et al.*, 1983). With the advent of high-throughput sequencing technology and an increased interest in genome-wide study, it became clear that ESTs could be generated in sufficient numbers to provide a rapid means of gene discovery (Khan *et al.*, 1992; Newman 1994; Sasaki *et al.*, 1994; Adams *et al.*, 1995; Houlgatte *et al.*, 1995).

In this paper, we described the gene expressed profile of the wild silkworm *B. mandarina* whole larvae by analyzing 227 random cDNA clones. This work is hoped not only to accelerate the functional studies of silkworm genome but also to make a contribution to understand the functions of a variety of ESTs that have been isolated from other organisms including *Drosophila*. In addition, we reporte the full-length cDNA cloning, sequencing and the identification of the tissue specific, putative digestive serine protease from *B. mandarina*. We also discussed the difference between our data and those obtained from other insects.

Materials and Methods

Construction of cDNA library

Total RNA from *Bombyx mandarina* larvae whole body was extracted by the guanidium isothiocyanate procedures (Sambrook *et al.*, 1989) and poly(A)$^+$ RNAs were purified from the total RNA using poly(A)$^+$ Quick Isolation Kit (Stratagene). The cDNA synthesis, ligation into Uni-ZAP XR vector and *in vitro* packaging were performed using the unamplified cDNA library construction kit (Stratagene) according to the manufacturer's instructions. An aliquot of the cDNA library was *in vivo* excised with the ExAssist helper phage and transfected into *E. coli* XL1-Blue. The transfected cells were plated on LB plates containing ampicillin, X-gal and IPTG.

Sequencing and database search

The insert sizes of randomly selected cDNA clones were estimated by the *EcoR*and *Xho*digestion of plasmid DNAs that were purified using the Wizard plasmid purification kit (Promega). Plasmid DNAs longer than 500 bp were used for single-pass sequencing. These clones (DNA 250 – 500 ng) were mixed with primer (3.2 pmol) and BigDye terminator ready reaction mix (Perkin Elmer) and was reacted by polymerase chain reaction (PCR). PCR product was electrophoresed on 4.5% polyacryamide gel and analyzed by the DNA Sequencing Analysis Softerware (Perkin Elmer). The similarity analyses were performed by the BLAST search.

Northern blot analysis

Total RNA from each tissue and stage was extracted using the SV total RNA isolation system. Total RNA (30 μg) from each tissue was denatured in formamide-containing buffer and subjected to denaturing agarose gel electrophoresis as described by Shambrook *et al.* (1989). After samples were blotted onto the nylon membrane, the blot was prehybridized in 0.5 M $NaPO_4$ (pH 7.4) / 1 mM EDTA / 1% BSA, and 7% SDS for 1 hr at 65 and hybridized in the same solution containing 2×10^6 cpm/μ*l* of probes labeled with [α-^{32}P]dCTP (Amersham) by nick translation. The membrane was then washed twice in $0.1 \times$ standard saline citrate and 0.1% SDS at room temperature for 10 min and twice at 42°C for 10 min.

Results and Discussion

We have constructed an oligo d(T)-primed cDNA library from the *Bombyx mandarina* whole larvae, from which 227 useful ESTs have been generated by single-pass sequencing of cDNAs. The cDNA clones were randomly selected to extract plasmid DNAs. Evaluation of the ESTs revealed that an average insert size was approximately 1.2 kb and the length of sequenced ESTs was ranged in 200 to 800 bp. Sequence analysis showed that 107 clones (47.1%) were classified into known genes while 120 clones (52.9%) were novel transcript and unknown function genes (Table 1). Of the 107 known genes, in the expression profile, the most abundant genes were found to be actin, followed by serine protease (data not shown).

Table 1. General classification of the random cDNA clones

Classification	No. of clones	Frequency (%)
Known genes	107	47.1
Insect matched	81	35.7
Bombyx mori	32	14.1
Drosophila melanogaster	15	6.6
Maduca sexta	19	8.4
Others	15	6.6
Mammalian matched	15	6.6
Others	11	4.8
Novel transcript genes	120	52.9
Total	227	100

In our *B. mandarina* whole larvae ESTs, the abundant *B. mori* proteins are serine protease (n = 8), low molecular proteins (n = 4), actin (n = 2), prophenoloxidase (n = 2), lysozyme (n = 2), silk protein p25, silk fibroin light chain (n = 2) and cystein protease (n = 2). Further investigation for the abundant genes of our *B. mandarina* whole larvae ESTs would provide very important information not only on the physiological state of the *B. mandarina* but also on its gene expression and regulation. Among such clones, the nucleotide sequence of the BmSP cDNA encoding serine protease homolog of *B. mandarina* was determined (Fig. 1). There is a 5'-untranslated region of 37 nucleotides followed by an initiating ATG codon and the TAA termination codon occurs at nucleotide position 865. The translation of the sequence from nucleotide 38 to 865 produces a 276-residue protein with a calculated molecular mass of 28.949 kD. The AATAAA consensus polyadenylation signal is present at 24 nucleotides downstream from the stop codon. The poly(A) tail is 15 nucleotides downstream from the polyadenylation signal sequence, which agrees with the fact that the signals are most often present 11-30 nucleotide upstream from the poly(A) tail (Fitzgerald and Shenk, 1981). There are no N-linked glycosylation sites (Asn-X-Thr/Ser), indicating that BmSP is not a glycoprotein. The deduced amino acid sequence of the BmSP cDNA was aligned with other serine protease family of *B. mori* (96% protein identity), *D. erecta* (21%), *Dermatophagoides farinae* (20%), *Sus scrofa* (21%), *Sar-*

```
1   CTGCAGGAATTCGGCACGAGCTTCGCAGCAGTACTGATGGCGTTGGCGGCCGTGGTCGTG
                                              M  A  L  A  A  V  V  V
61  GCAGAAGAGGCCATCGAACTTGACTACCACAACAAGATCGGTATCCCCCGGGCCGAGAGT
    A  E  E  A  I  E  L  D  Y  H  N  K  I  G  I  P  R  A  E  S
121 CTTAGACGCGCCGAGGAAGCCGCTGACTTCGACGGTACCAGGATTGTGGGTGGTTCTGCC
    L  R  R  A  E  E  A  A  D  F  D  G  T  R  I  V  G  G  S  A
181 GCCAACGCTGGTGCTCACCCCCATCTTGCTGGACTTGTGATCGCACTGACGAATGGCAGA
    A  N  A  G  A  H  P  H  L  A  G  L  V  I  A  L  T  N  G  R
241 ACTTCCATCTGCGGAGCTTCCTTACTGACCAACACCCGCTCCGTGACCGCCGCTCATTGC
    T  S  I  C  G  A  S  L  L  T  N  T  R  S  V  T  A  A  H  C
301 TGGAGGACCAGGAGAGCCCAGGCTCGTCAGTTCACCCTCGCTCTTGGCACAGCTAACATC
    W  R  T  R  R  A  Q  A  R  Q  F  T  L  A  L  G  T  A  N  I
361 TTCTCCGGAGGCACCAGGGTCACCACCTCCAATGTCCAGATGCACGGCAGCTACAACATG
    F  S  G  G  T  R  V  T  T  S  N  V  Q  M  H  G  S  Y  N  M
421 GACACCCTCCACAACGACGTCTCCATCATCAACCACAACCATGTTGGCTTCACCAACAAC
    D  T  L  H  N  D  V  S  I  I  N  H  N  H  V  G  F  T  N  N
481 ATCCAGCGCATCAACATAGCCAGTGGAAGCAACAACTTTGCTGGTACTTGGGCCTGGGCT
    I  Q  R  I  N  I  A  S  G  S  N  N  F  A  G  T  W  A  W  A
541 GCCGGCTTCCGAAGGACCTCCGATGCTGCTTCGGGAGCCAACAACCAACAAAAACGCCAA
    A  G  F  R  R  T  S  D  A  A  S  G  A  N  N  Q  Q  K  R  Q
601 GTGAGCTTGCAGGTCATTACCAACGCCGTGTGCGCCCGCACGTTTGGAAACAATGTGATC
    V  S  L  Q  V  I  T  N  A  V  C  A  R  T  F  G  N  N  V  I
661 ATTGCCTCCACCCTCTGTGTTGACGGCTCTAACGGTCGCAGCACCTGCAGCGGAGACTCC
    I  A  S  T  L  C  V  D  G  S  N  G  R  S  T  C  S  G  D  S
721 GGCGGCCCTCTCACCATCGGCAGCGGCGGAAGCCGTCAACTGATCGGTATCACATCGTTC
    G  G  P  L  T  I  G  S  G  G  S  R  Q  L  I  G  I  T  S  F
781 GGATCAGCTCAAGGTTGCCAGAGAGGCCACCCTGCCGGCTTCGCCAGAGTCACATCCTTC
    G  S  A  Q  G  C  Q  R  G  H  P  A  G  F  A  R  V  T  S  F
841 AACTCCTGGATCCGGGCTAGAATTTAACTGACGAATGACCAATGATTCAATAATAAAATA
    N  S  W  I  R  A  R  I  * 865
901 ATTTCATTGACCAAAAAAAAAA
```

Fig. 1. Nucleotide and deduced amino acid sequences of the BmSP . The open reading frame is indicated in uppercase letters. The predicted amino acid sequence (single-letter abbreviation) is shown below the nucleotide within the open reading frame. Codons for initiation, termination, polyadenylation and poly(A) tail are shaded, respectively.

cophaga bullata (23%), *Lucilia cuprina* (20%), *Stomoxys calcitrans* (22%), *Stomoxys calcitrans* (17%), *Anopheles gambiae* (19%), *Lucilia cuprina* (20%) and *Hypoderma lineatum* (17%) (Fig. 2). BmSP contains histidine, aspatic acid and serine residues forming the catalytic center. Six cysteines that are thought to contribute to three disulphide bonds commonly found in invertebrate serine proteases are also found at conserved sites and there are many other completely conserved amino acids.

At the trypsin substrate specificity sites, this gene contains the serine residue at position 222. It thus appears that BmSP gene is close to chymotrypsin-like protease than

```
fm  ----------MLRFIAVFALVNCALAGTL-------------PNDLDG-RIVNGVDTTIE 60
hm  ----------MLRFVILFALVSTSLAGVS-------------RNDFYG-RIVNGVATTIE 60
cm  --------MHGLVVLLVCLAVGSAFAGTIGVS-------NADPFEREG-RIVGGEDTTIR 60
gm  ----------MLKFVVLLCAISCALGAAVPEG---------MVPQLDG-RIVGGVATTIS 60
km  ----------MLKFVVLLCAISCALGAAVPEG---------MVPQLDG-RIVGGVATTIS 60
lm  ----------MLKFVILLCSIAYVFGAVVPLG---------MLSQSDG-RIVGGVESKIE 60
dm  --------------------------------------------------IVGGVKAKAG 60
em  --------------------------------------------------IVGGYTCAAN 60
am  --------MALAAVVVAEEAIELDYHNKIGIPRAESLRRAEEAADFDGTRIVGGSAANAG 60
bm  MKVFAAVLMALAAVVVAEEPIELDYHIKIGIPRAESLRRAEEAADFDGTRIVGGSAANAG 60
jm  ----------MLRKVFAVVSVLLVVSAAKVPK--------LVLDDNYVNRVVGGVVAKNC 60
im  ----------MKLFVAIAALVIACASAASLDG-------IARPGFPEG-RIINGLPATKG 60

fm  AHPYQVPLQN---AALSHFCGGSIISEDLVVTAAHCMQS--YTASQIKVRLGS-TIYNEG 120
hm  EHPYQVSLQG---LSGSHFCGGSIISEDIVVTAAHCMQS--HSASEFKVRLGS-TQYNTG 120
cm  AHPYQVSLQN---KKGSHFCGGSLINEDTVVTAAHCLVG--KKIAKVFVRLGS-TLYNEG 120
gm  SFPWQISLQ----RSGSHSCGGSVYNSRIIVTAAHCLQS--VSTSVLKVRRGS-SYWNSG 120
km  SFPWQISLQ----RSGSHSCGGSVYNSRIIVTAAHCLQS--VSTSVLKVRRGS-SYWNSG 120
lm  DFPWQISLQ----RDGRHYCGGSIYSKNVIITAAHCLRN--VVAEELRVRVGS-SYWEHG 120
dm  DCPYQISLQ-----SSSHFCGGSILDEYWILTAAHCVNG--QSAKKLSIRYNT-LKHASG 120
em  SIPYQVSLN-----SGSHFCGGSLINSQWVVSAAHCY----KSRIQVRLGEHN-IDVLEG 120
am  AHPHLAGLVIALTNGRTSICGASLLTNTRSVTAAHCWRTRRAQARQFTLALGT-ANIFSG 120
bm  AHPHLAGLVIALTNGRTSICGASLLTNTRSVTAAHCWRTRRAQARQFTLALGT-ANIFSG 120
jm  SAPYQVSLQVP---GWGHNCGGSLLNDRWVLTAAHCLVG--YEPSDLMVLVGT-NSLKEG 120
im  QAPFIVSLK-----SGSHFCGGSIIDEHWVLTAAHCLT---KSQFQLVAGLYERSDESDV 120

                           0           +0
fm  GE-LVSVKAFKF-HEGYNPKT--MVNDVALIKLATPVRESS-KIRY-IRLAD----RTPP 180
hm  GE-LVEVKAFKF-HENYNSGT--MKNDVAVIKLARPVKESA-TIRF-VKLAD----KTPA 180
cm  GI-VVAVRALTY-NADYSSKT--MENDVGILKLAEKVKETD-DIRY-IELAT----ETPP 180
gm  GV-VVSVAAFKN-HEGYNPKT--MVNDIAVIRLSSSLTMSS-TIKA-IALTT----AAPA 180
km  GV-VVSVAAFKN-HEGYNPKT--MVNDIAVIRLSSSLTMSS-TIKA-IALTT----AAPA 180
lm  GS-LRNISKFQI-HESYVEPT--KEYDVALLKLDSDLSFNS-TIKA-IELTN----EIPP 180
dm  GE-KIQVAEIYQ-HENYDSMT--IDNDVALIKLKTPMTLDQ-TNAKPVPLPPQ--GSDVK 180
em  NEQFINAAKIIT-HPNFNGNT--LDNDIMLIKLSSPATLNS-RVAT-VSLPR----SCAA 180
am  GT-RVTTSNVQM-HGSYNMETDTLHNDVSIINHN-HVGFTN-NIQR-INIAS---GSNNF 180
bm  GT-RVTTSNVQM-HGSYNMDT--LHNDVAIINHN-HVGFTN-NIQR-INLAS---GSNNF 180
jm  GE-LLKVDKLLY-HSRYNRPQ--FHNDIGLMRLEQPVQFSE-LVQS-VEYLE----KAVP 180
im  QIRNVNGKQFIFTHEIYGGNV--GPHDIGLSIVEEAFDLNACLVMDLLPLPALTCLLANM 180
                          +
```

Fig. 2. Comparison of deduced amino acid sequences for BmSP and other serine protease sequences. The full-length sequences of BmSP were used to search GenBank using BLASTP. Multiple alinments were performed using Clustal W on selected sequences. The amino acid residues forming catalytic triad are indicate by + and six highly conserved cysteine residues by O. am, *Bombyx mandarina*; bm, *Bombyx mori* (AB00367); cm, *Drosophila erecta* trypsin (p54628); dm, *Dermatophagoides farinae* Der f3 mite allergen (U54781); em, *Sus scrofa* porcine beta trypsin (G494360); fm, *Sarcophaga bullata* trypsin (P51588); gm, *Lucilia cuprina* trypsin (p35044), hm, *Stomoxys calcitrans* (AF074956); im, *Stomoxys calcitrans* (AF074955); jm, *Anopheles gambiae* (S44184); km, *Lucilia cuprina* (L15632); lm, *Hypoderma lineatum* (P35587).

```
fm   TGTPAVVTGWGTKCFLTCVSLPKTLQEVEVDIVDQKACASNEFKYGSQIQDTMVCAYALK 240
hm   TGTPAVVTGWGTTCFMACNTLPKTLQKVVVDIVDEKTCASSEYKYGSKIKPTMVCAYAED 240
cm   TGTTAVVTGWGSKCYFWCMTLPKTLQAVYVNIVDWKTCASDEYKYGEVIYDTMVCAYEKK 240
gm   NGAAATVSGWGTTSSGG--SIPAQLRYVDLKIVGRTQCASSTYGYGSQIKPSMICAYTVG 240
km   NGAAATVSGWGTTSSGG--SIPAQLRYVDLKIVGRTQCASSTYGYGSQIKPSMICAYTVG 240
lm   EYADAIVSGWGETLVPPP-GIPDQLRSVDVKIIHREKCASRNFGYGSNIKASMICAYAIG 240
dm   VGDKIRVSGWGYLQEGSY-SLPSELQRVDIDVVSREQCDQLYSKAGADVSENMICGGDVA 240
em   AGTECLISGWGNTKSSGS-SYPSLLQCLKAPVLSNSSCKSSYP---GQITGNMICVGFLQ 240
am   AGTWAWAAGFRRTSDAASGANNQQKRQVSLQVITNAVCARTFG--NNVIIASTLCVDGSN 240
bm   AGTWAWAAGFGRTSDAASGANNQQKRQVSLQVITNAVCARTFG--NNVIIASTLCVDGSN 240
jm   VNATVRLTGWGRTSTNG--NVPTLLQSLNVVTLSNEDCKAKMGN-PENVDLGHVCTLTKA 240
im   KALAVVNSGWGRDNSGS---LPNTLQTLEVDIIGYTECKAAVP-LDAPLADVNICSYTAG 240
                                       0          0
fm   ---KD-ACQGDSGGPLVANN----QLVGIVSWG--SGCARVGYPGVFCDVPSVRSWIEKT 300
hm   ---KD-ACQGDSGGPLVAGG----KLVGVVSWG--KGCALPAIPGVYADVPSLRTWIEKT 300
cm   ---KD-ACQGDSGGPLAIGN----TLVGIVSWG--YACASNLLPGVYSDVPALRKWILNA 300
gm   ---KD-SCQGDSGGPLVSGG----RLVGVVSWG--YGCAFANYPGVYADVAALRTWVVSA 300
km   ---KD-SCQGDSGGPLVSGG----RLVGVVSWG--YGCAFANYPGVYADVAALRTWVVSA 300
lm   ---KD-SCQGDSGGPLVVNN----LLVGVVSWG--IDCARPSYPGVYVDVSHVRSWIVSN 300
dm   NGGVD-SCQGDSGGPVVDIAT--KQIVGIVSWG--YGCARKGYPGVYTRVGNFVDWIESK 300
em   -GGKD-SCQGDSGGPVVCNG----QLQGIVSWG--YGCAQKNKPGVYTKVCNYVNWIQQT 300
am   --GRS-TCSGDSGGPLTIGSGGSRQLIGITSFGSAQGCQR-GHPAGFARVTSFNSWIRAR 300
bm   --GRS-TCSGDSGGPLTIGSGGSRQLIGITSFGSAQGCQR-GHPAGFARVTSFNSWIRAR 300
jm   --GEG-ACNGDSGGPLVYEG----KLVGVVNFG--VPCGR-GFPDGFARVSYYHEWVRTT 300
im   --TKDGACNGDSGGPLVKNTKGGYELVGLVSWG-YVGCASTQMPSIYTSVASYKQWIADT 300
               0  +                  0
fm   AKEL-- 336
hm   AKEL-- 336
cm   SQTL-- 336
gm   ASSV-- 336
km   ASSV-- 336
lm   AESI-- 336
dm   RSQ--- 336
em   IAAN-- 336
am   I----- 336
bm   I----- 336
jm   MANNS- 336
im   IAAYKN 336
```

Fig. 2. Continued.

trypsin -like protease. It is well known that serine proteases are important factors for the fate of the disease agents ingested by the haematophagous insects. The serine proteases are a class of proteolytic enzymes characterized by the presence of a uniquely reactive serine side chain. All active serine proteases contain an invariant catalytic triad consisting of His-Asp-Ser. They are of ubiquitous occurrence and diverse function, playing important regulatory roles in protein processing and degradation (Neurath, 1989) in eukaryotes and prokaryotes. Principal functions in digestion (Barillas-Mury *et al.*, 1991; Muller *et al.*, 1993), hormone activation (Douglass *et al.*, 1984), embryogenesis and development (Delotto and Spierer, 1986), blood coagulation (Davie, 1986) and activation pathway of the complement system (Cooper, 1985) have all been well documented. A number of serine protease from arthropods has been biochemically purified or cloned and sequenced. These include several digestive trypsins and chymotrysins from mosquitoes and hornets (Jany *et al.*, 1983; Graf and Briegel, 1985; Kalhok *et al.*, 1993; Muller *et al.*, 1993).

In order to examine tissue-specific expression of the BmSP, we have performed a Northern blot analysis using BmSP cDNA fragment as a probe (posterior and middle divisions of silkgland, midgut, fat body and malphighian tubes). The mRNA expression of BmSP was detected at a high level in midgut (Fig. 3). The detected mRNA was found to have 0.9 kb in size, which corresponds to the size of the complete BmSP cDNA sequence as described above. This transcript is similar to trypsin and chymotrypsins previously characterized in mosquitoes, which range from 900 to 1,000 bp. The BmSP gene was highly and specifically expressed only in midgut tissue among several tissues. This result suggests that the BmSP gene is closely associated with the expression of digestive enzyme. Further characterization of gene expression and function will be necessary to determine the precise role of this gene.

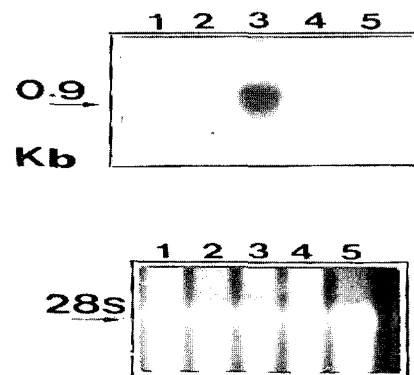

Fig. 3. Tissue-specific expression of the BmSP. Total RNAs (30 μg per lane) isolated from the following drent tissues of 1: posterior division of silkgland, 2: middle division of silkgland, 3: midgut, 4: fat body, 5: malpighian tubes was separated by agarose-formaldehyde gel electrophosis and transferred onto a nylon membrane (below). BmSP mRNA was detected commonly at 0.9 kb indicated by arrow head. Hybridization with BmSP cDNA demonstrates the existence a transcript of 0.9 kb in each tissue indicated by arrow head.

Acknowledgment

This work was partly supported by a grant from Biogreen 21 program, Rural Development Administration, Republic of Korea.

References

Adams, M. D., A. R. Kerlavage, R. D. Fleischmann, R. A. Fuldner, C. J. Bult, N. H. Lee, E. F. Kirkness, K. G. Weinstock, J. D. Gocayne, O. White, G. Sutton and J. C. Venter (1995) Initial assessment of human gene diversity and expressed patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377**, 3-7.

Barillas-Mury, C., R. Graf, Hagedorn and M. A. Wells (1991) cDNA and deduced amino acid sequence of a blood mealinduced trypsin from the mosquito, *Aedes aegypti. Insect Biochem.* **21**, 825-831.

Cooper, N. R. (1985) The classical complement pathway: activation and regulation of the first complement component. *Adv. Lmmunol.* **37**, 151-216.

Davie, E. W. (1986) Introduction to the blood coagulation cascade and the cloning of blood coagulation factors. *J. Protein Chem.* **5**, 247-253.

Delotto, R. and P. Spierer (1986) Agene required for the specification of dorsal- ventral pattern in *Drosophila* appears to encode a serine protease. *Nature* **323**, 688-692.

Douglass, J., O. Civelli and E. Herbert (1984) Polyprotein gene expression: generation of diversity of neuroendocrine peptides. *Annu. Rev. Biochem.* **53**, 665- 715.

Fitzgerald, M. and V. Shenk (1981) The sequence 5-AAUAAA-3 forms parts of the recognition site for polyadenyllation of late SV40 mRNAs. *Cell* **24**, 251-260.

Graf, R. and H. Briegel (1985) Isolation of trypsin isozymes from the mosquito *Aedes aegypti* (L.). *Insect Biochem.* **15**, 611-618.

Hillier, L. D., G. Lennon, M. Becker, M. F. Bonaldo, B. Chiapelli, S. Chissoe, N. Dietrich, T. DuBuque, A. Favello, W. Gish, M. Hawkins, M. Hultman, T. Kucaba, M. Lacy, E. Mardis, B. Moore, M. Morris, J. Parsons, C. Prange, L. Rifkin, T. Rohlfing, K. Schellenberg, M. B. Soares, F. Tan, J. Thierry-Meg, E. Trevaskis, K. Underwood, P. Wohldman, R. Waterston, R. Wilson and M. Marra (1996) Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* **6**, 807-828.

Houlgatte, R., R. Mariage-Samson, S. Duprat, A. Tessier, S. Bentolila, B. Lamy and C. Auffray (1995) The Genexpress Index: A resource for gene discovery and the gentic map of the human genome. *Genome Res.* **5**, 272-304

Jany, K. D., G. Beckler, G. Pfeiderer and J. lshay (1983) Amino and sequence of an insect chymotrypsin from the larvae of the hornet. *Vespa orientalis. Biochem. Biophys. Res. Commun.* **110**, 1-7.

Kalhok, S. E., L. M. Tabak, D. E. Prosser, W. Brook, A. E. R. Downe and B. N. White (1993) Isolation, sequencing and characterization of two cDNA clones coding for trypsin-like enzymes from the midgut of *Aedes aegypti. Insect Mol. Biol.* **2**, 71-79.

Khan, A. S., A. S. Wilcox and M. H. Polymeropoulos (1992) Single pass sequencing and physical and genetic mapping of human brain cDNAs. *Nature Genet.* **2**, 180-185.

Kirkness, E. F. (1996) Assessment of human gene diversity and expression pattern using expressed sequence tags. *Essays Biochem.* **31**, 1-9.

Milner, R. J. and J. G. Sutcliffe (1983). Gene expression in rat brain. *Nucleic Acids Res.* **11**, 5497-54520.

Muller, H. M., J. M. Crampton, A. D. Torre, R. Sinden and A. Crisanti (1993) Members of a trypsin gene family in *Anopheles gambiae* are induced in the gut by blood meal. *EMBO J.* **12**, 2891-2900.

Neurath, H. (1989) Proteolytic processing and physiological regulation. *Trends Biochem. Sci.* **14**, 268-271.

Newman, T. (1994) Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous aravidopsis cDNA clones. *Plant Physiol.* **106**, 1241-1225.

Putney, S. D., W. C. Herlihy and P. Schimmel (1983) A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. *Nature* **302**, 718-721.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular cloning: A laboratory manual, 2nd edition. pp. 25-88, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.

Sasaki, T., J. Song, Y. Koga-Ban, E. Matsui, F. Fang, H. Higo, H. Nagasaki, M. Hori, M. Miya and E. Murauama-Kayano (1994) Toward catalogung all rice genes: Large-scale sequencing of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA* **91**, 9228-9232.