Int. J. Indust. Entomol. Vol. 17, No. 2, 2008, pp. 229~234

[Note]

International Journal of
Industrial Entomology

Comparative Study of Gene Expression Profiles in Posterior Silk Glands of the Silkworm, *Bombyx mori* L.

Kwang Ho Choi*, Tae Won Goo, Seok Woo Kang, Min Uk Kang, Eun Young Yun and Jae Sam Hwang
Department of Agricultural Biology, NAAS, RDA, Suwon 441-100, Korea

(Received 12 November 2008; Accepted 12 December 2008)

We used serial analysis of gene expression (SAGE) approach to derive a profile of expressed genes of the posterior silk glands (PSG) and to create a reference for understanding gene cluster related to the mechanism of silk protein synthesis in the silkworm, *Bombyx mori*. We constructed a 3' SAGE library from the PSG of the fifth instar larvae of the silkworm. In total we obtained 2,406 SAGE tags, of which 682 were unique tags. Sorted by tag count number, 27 (4%) unique tags were significantly more abundant genes (ten or more times), whereas 445 (65%) unique tags were detected as single copies. The annotation of 682 unique SAGE tags revealed that 462 (68%) of the SAGE tag sequences represented known genes, whereas 220 (32%) of the tag sequences had no matches in SAGE map and silkworm EST databases. Of the 682 SAGE tags, the most abundant tag sequences were that of the fibroin light chain gene and the silk protein P25. In addition, we compared two relative abundance results of the SAGE and the EST approaches to verify whether their transcript quantitative aspects are significant or not. The comparative results of relative abundances of the fibroin H-, L- chain and P25 glycoprotein genes indicated that the quantitative approach based on SAGE tags is effective for quantitative cataloging and comparison of expressed genes in same organs. The SAGE tag information reported in this study would be useful for researchers in the field to analyze genes associated with silk processing mechanisms of insects.

Key words: EST, Gene, Posterior silk glands SAGE Silkworm.

*To whom the correspondence addressed
Department of Agricultural Biology, National Academy of Agricultural Science, Rural Development Administration, Suwon 441-100, Korea. Tel: +82-31-290-8533; E-mail: ckh@rda.go.kr

Introduction

Many approaches have been employed to study the expressed gene profiling on various species. Expressed sequence tag (EST) methodology is the most popular tool used to search a novel gene (David, 2003). Currently more than 480,000 silkworm ESTs from independent projects are available in public database. The EST approach is a valuable tool for gene discovery, but like RNA blotting, ribonuclease protection, and reverse transcriptase polymerase chain reaction (RT-PCR) analysis, EST evaluates only a limited number of genes at a time. Serial analysis of gene expression (SAGE), by contrast, is one of the versatile methods for genome-level studies of gene expression (Velculescu *et al*., 1995). The SAGE analysis technique supplied valuable data on gene expression in the form of a table of counts. Particularly, mRNA transcripts are converted to complement DNA (cDNA) and then processed so as to generate tags (10~14 bp). Ideally, these tags should provide enough information to uniquely identify the source mRNA transcript. The frequency of each tag in the SAGE tag library is an accurate estimate of the abundance of its corresponding transcript (Keith *et al*., 2003). Most methods currently advanced for assessing differential expression in SAGE address the one library is contrasted with another, assuming a null hypothesis that there is no difference between the libraries being compared (Sun *et al*., 2004). In silkworm, the SAGE technology has been used for finding the differentiated transcripts among several SAGE libraries (Funaguma *et al*., 2007; Huang *et al*., 2005; Zhang *et al*., 2007).

The domesticated silkworm, *Bombyx mori*, has been used as a model for physiological basic studies, and provides a number of mutants and genetically improved strains (Goldsmith *et al*., 2005). Above all, silkworm itself

has considerable importance as a producer of recombinant fibrous silk proteins from transgenic silkworms (Imamura *et al*., 2003; Tomita *et al*., 2003). The silkworm larvae have a pair of silk glands, each divided into three anatomically distinct regions: anterior (ASG), middle (MSG) and posterior (PSG). The three silk gland compartments display immense degree of functional differentiation. Especially, the PSG exclusively synthesize the silk proteins such as fibroin heavy chain (fib H-chain) and fibroin light chain (fib L-chain), and fibrohexamerin P25 (Inoue *et al*., 2000; Tanaka *et al*., 1999).

In our previous study, we reported a valuable data of EST profile of the PSG for finding a useful or novel gene related with the biosynthetic mechanisms of silk proteins of the silkworm (Choi *et al*., 2007). In the psgEST, 554 unique ESTs were generated including some well known genes for silk protein synthesis of such as fib H-chain, fib L-chain, and P25. As a result of former examination, it is necessary to verify a precision quantitative aspect of mRNA transcripts in the PSG of the silkworm. Therefore, we have applied the SAGE method to examine an extensive aspect of the expressed gene's quantitative pattern of the PSG of the silkworm. In addition, to verify an agreement of relative transcript abundances using by SAGE and EST approaches, the relative abundances on some silk protein genes were compared between the two databases.

Materials and Methods

Silkworm and mRNA preparation

The inbred silkworm strain Jam124, *Bombyx mori* L. that is maintained at the National Academy of Agricultural Science (NAAS) of Suwon, Korea was used in this study. Posterior silk glands (PSGs) were dissected out from four-day-old fifth instar larvae in cold PBS (0.1M phosphate and 0.15M NaCl, pH 7.0). One hundred milligrams of the PSG immediately frozen in liquid nitrogen were ground into fine powder. The total RNA was isolated from PSG homogenates using a RNA isolation kit (QIAGEN, Valencia, CA) and mRNA was selected using a Micro Fast-Track 2.0 mRNA isolation kit (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol.

SAGE procedures

3' SAGE tag library, dubbed "psgSAGE", was constructed using 2.5 μg mRNA and was converted to cDNA with a I-SAGE kit version E (Invitrogen) following the instruction of the supplier. The cDNA was cleaved with the restriction enzyme *Nla*III, and the 3'-terminal cDNA fragments were bound to streptavidin-coated magnetic beads (Dynal, Oslo, Norway). After capturing of 3' cDNA fragments, the bound cDNA was divided into two pools, and one of the following 5' biotinylated oligonucleotide linkers containing recognition sites for *Bsm*F1 was ligated to each pool: linker 1, 5'-tttggatttgctggtgcagtacaactaggcttaata gggacatg-3', 5'-tccctattaagcctagttgtactgcaccagcaaatcc[Amino mod. C7]-3', linker 2, 5'-tttctgctcgaattca agcttctaacgatgtacggg-gacatg-3', 5'-tccccgtacatcgttagaagcttgaattcgagcag [Amino mod. C7]-3'. Because *Bsm*F1 cleaves 14 bp away from its recognition site and the *Nla*III site overlaps the *Bsm*F1 site by 1 bp, a 15 bp SAGE tag was released with *Bsm*F1. SAGE tag overhangs were filled-in with Klenow flagment (Invitrogen), and tags from the two pools were ligated to each other another to generate ditags flanked by linkers 1 and 2. Ditags were diluted and then amplified with PCR with 5'-ggatttgctggtgcagtaca-3' and 5'-ctgctcgaattcaagct-tct-3' as primers. The PCR amplification was analyzed by 4.5% agarose gel electrophoresis, and the PCR product containing two tags ligated tail to tail (ditag) was excised. The PCR product was then cleaved with *Nla*III, and the band containing the ditags was excised and self-ligated. After ligation, the concatenated products were separated by 8% polyacrylamide gel electrophoresis (PAGE) and products between 300 bp and 1,000 bp were excised. These products were cloned into the *sph*I site of pZero-1 (Invitrogen), and then transformed into *E. coli* Top10F' competent cells (Invitrogen). Colonies were screened for inserts by PCR with Sp6 and T7 primers. PCR products containing inserts of greater than 300 bp were sequenced with the BigDye Terminator Cycle Sequencing Ready Reaction kit version 3.1 with AmpliTaq DNA polymerase (PE Applied Biosystems), and analyzed using an ABI prism 377 DNA analyzer (PE Applied Biosystems).

SAGE tag data analysis

The SAGE tag sequence data were analyzed using the SAGE 2000 software version 4.5 (http://www.invitrogen.com/sage), which extracts tags from the sequences and creates a report of each tag and its abundance level (Marguiles and Innis, 2000). Annotation of the psgSAGE tag sequences was performed by matching tags sequences to reference sequences on the SAGEmap database (http://www.ncbi.nlm.nih.gov/SAGE) and the *B. mori* psgEST database (Choi *et al*., 2007).

Results and Discussion

Many people have made efforts to understand the silk protein processing in silk glands of the silkworm (Huang *et al*, 2005; Jin and David, 2003). How many genes are related to silkworm silk protein synthesis? This is important question for the accurate understanding of silkworm

Table 1. Summary of SAGE tag analysis generated from posterior silk glands of the silkworm

Total number of concatemers sequenced	Number of reliable concatemers data	Total number of extracted tags	Number of unique tags	Matches to databases	Unknown or novel
200	142 (71%)[1)]	2,406	682 (28%)[2)]	462 (68%)	220 (32%)

[1)]Percentages in the total number of concatemers sequenced in this experiment.
[2)]Percentages in the total number of extracted tags.

biology and genomic studies. Also, the survey of the overall gene expression pattern would provide the important clues and the deeper insight into the basic cellular biology of the more comprehensive understanding of the silk processing mechanisms in spinning insects. Identification of a complete set of genes expressed in a genome is one of the ultimate goals of transcriptome studies. Such information is essential for genome annotation and for further study of the function of each gene.

In the present study, with aiming for not only a novel gene discovery but also a transcripts pattern survey in posterior silk gland (PSG) of the silkworm, we report a psgSAGE tag profile of the by analyzing two hundred SAGE tag concatemer clones. An average size of concatemers was calculated by colony PCR amplification with Sp6 and T7 primers. The mean size of the concatemers was 17 tags per clone and the rage from 5 to 40 tags. After elimination of 58 unreliable concatemers, 10 bp of each SAGE tag sequence were generated from 142 reliable concatemers sequences using the SAGE 2000 software. Thus, the total number of reliable tag sequences in the psgSAGE library was 2,406, of which 682 different unique tags were identified. This corresponds to 28% of the total tags extracted (Table 1).

The annotation of 682 psgSAGE unique tags by the silkworm EST database revealed that 462 (68%) of the unique tags represented known genes, whereas 220 (32%) tags had no matches in SAGE and EST databases. Annotated the tag sequences data were classified according to the number of matches per gene with its putative identification and its corresponding gene numbers, respectively. Among these tags, 1,982 tags were detected twice or more accounted for 82% of the total tag population, whereas 445 tags were detected only once accounted for 18% of the total tag population. Table 2 shows the top 40 highly abundant tags (more than eight times) with gene identifications. Four percent (27 tags) of the unique tags were detected at least ten times each, which corresponds to a representation of more than 53% of the total tag population. Forty of these high-abundance tag sequences were referred to the silkworm psgEST database (Choi *et al*., 2007), and 28 of these tags were identified as known genes with characterized functions (Table 2). As expected, well-known silk protein genes such as fib L-chain, fib H-chain and P25 genes were expressed at highly level. Of these tags, the highly redundant tags (5'-GCCAGCCAGG-3' and 5'-TTAAACCTTT-3') that matched sequences in the silkworm EST database were those reflecting respectively fib L-chain gene (EF421144.1) and silk protein P25 gene (X024226.1) which accounted for 384 times (16% of the total tags) and 117 times (5% of the total tags) respectively. Also, housekeeping gene tags (5'-CAAAAGAAAC-3', 5'-CAGATCTGCT-3', 5'-GGCTTCGGTC-3') such as ribosomal protein encoding genes were among the most highly expressed genes (Table 2). Other highly redundant tag that matched sequence was that reflecting the fib H-chain gene (EU346954.1), which accounted for 18 times. This corresponds to approximately 0.8% of the total tag population analyzed. However, among the 40 highly abundance unique tags, 12 tags were identified as unknown or novel transcripts tags. Interestingly, there were many functionally unknown genes among these highly expressed genes. The high copy level of expression of these genes suggests that they have an important functional role in the mechanism of silk protein synthesis of the silkworm.

The total abundance and information of psgSAGE unique tags were presented in this report. We presumed that the psgSAGE tag database represents a quantitative transcript aspect related in silk protein biosynthesis mechanism. In previous study (Choi *et al*., 2007), we reported a psgEST database with 2,867 ESTs which was derived from just the same mRNA sample with that of the psgSAGE library. We compared the relative abundance of SAGE tags to that of the EST database. In the psgEST report, 544 unique ESTs were generated after the assembly of 2,867 ESTs. Among the 544 unigenes, the most abundantly expressed genes were fib L-chain gene (19 %), elongation factor 1-alpha gene (14%), and silk fibroin P25 gene (5.5%). However, in the psgSAGE tag redundancy data, the most abundant genes were fib L-chain gene, silk fibroin P25, and fib H-chain gene. These correspond to respectively 16.0%, 4.9%, and 0.75% of total SAGE tag sequences analyzed. The silk protein of silkworm is consisted of essential three elements such as fib H-chain, fib L-chain, and P25. A strict combination of these elements is required for the efficient secretion of fibrous protein from the PSG cells, which has a regular

Table 2. The 40 most frequently detected SAGE tags of posterior silk glands of the silkworm [1)]

Tag sequence (5'-3')	Count.	Abundance (%)	*Bombyx mori* Gene	Accession No.	No. of corresponding gene
GCCAGCCAGG	384	15.96	fibroin L-chain gene, exon 3	EF421144.1	1
TTAAACCTTT	117	4.86	silk protein P25 mRNA	X04226.1	3
CGGAGATTAT	81	3.37	No match [2)]	-	-
TATGTAATTT	78	3.24	mRNA for Seroin 2, mRNA	NM001046815.1	1
ATGGCGAGTG	63	2.62	ribosomal protein P2, mRNA	NM001043748.1	1
AGGTGATGCT	57	2.37	small nuclea ribonuclear protein F	NM001099806.1	3
CAAAAGAAAC	54	2.24	ubiquitin/ribosomal protein L40 fusion protein	NM001043907.1	2
CAGATCTGCT	53	2.20	ribosomal protein L37A (Rpl37a) mRNA	NM001043784.1	1
GGCTTCGGTC	49	2.04	mRNA for 60S acidic ribosomal protein P2	AM260708.1	2
TCCTTAATTA	45	1.87	No match	-	-
TCCGCATCCT	33	1.37	translationally controlled tumor protein (TCTP) mRNA	CF210320.1	2
TTGTTAATGT	32	1.33	No match	-	-
TCCCCGTACA	27	1.12	No match	-	-
TAATAATTGG	21	0.87	ATPase subunit 6 gene; cytochrome oxidase subunit III gene	DQ468099.1	1
GAATACTTGG	18	0.75	fibroin heavy chain Fib-h gene, exon 2 and partial cds	EU346954.1	2
TACGTCGAGA	18	0.75	mRNA for glyceraldehyde-3-phosphate dehydrogenase	AB262581.1	2
TCCTTCCTGA	18	0.75	kazal-type serine proteinase inhibitor 1 (Spi1), mRNA	NM001043579.1	1
TTTGGTTAAA	18	0.75	No match	-	-
AATCAATATG	15	0.62	ribosomal protein L37 (Rpl37), mRNA	NM001043782.1	1
ACTATACACT	15	0.62	No match	-	-
AGATTACAAT	15	0.62	ribosomal protein S15 (RpS15) mRNA	AY769330.1	1
CAACAGTACA	12	0.62	No match	-	-
ACATAGGATT	11	0.50	ribosomal protein S29 (Rps29), mRNA	NM001043814.1	1
TAATCGTTCG	11	0.50	No match	-	-
AATCGGAGCC	10	0.42	Xiafang mitochondrion	AY048187.1	1
GCGGTGATGG	10	0.42	ubiquitin-ribosomal protein S27a fusion (UERPS27A) mRNA	EF210321.1	1
AAGACGTTGT	10	0.42	No match	-	-
GACCTCTCGA	9	0.37	spatzle-1 (Spz), mRNA	NM001114594.1	3
AATGAAATCA	9	0.37	Xiafang mitochondrion	AY048187.1	1
CAATCAGAAC	9	0.37	18S ribosomal RNA gene	DQ347470.1	1
GAATCCTATA	9	0.37	hemocytin, transcriptional regulatory region	AB014342.1	1
GTTTGTACAT	9	0.37	vacuolar-type H+-ATPase subunit c (VATPc) mRNA	EU082222.1	1
GTTGTCACGG	9	0.37	18S ribosomal RNA gene	DQ347470.1	1
CAAAAGTACA	9	0.37	No match	-	-
TAATTTTTTT	8	0.33	silk protein P25 mRNA	X04226.1	2
TAATTAAGAT	8	0.33	juvenile hormone esterase (Jhe), mRNA	NM001043562.1	1
TACCATCTAC	8	0.38	ribosomal protein S28 (Rps28), mRNA	AY583363.1	1
TAGAATTTTG	8	0.38	DEAD box polypeptide 5 (LOC732932), mRNA	NM001044117.2	1
TCAGACAGTC	8	0.38	No match	-	-
TAATGATAGA	8	0.33	No match	-	-

[1)]A total of 2,406 SAGE tag sequences were identified, representing 682 unique transcripts. The table listed the top 40 tag sequences with unambiguous matches or no matches in EST database of the PSG library.
[2)]Tags without match in the PSG database.

molar ratio of 6H : 6L : 1P25 (Inoue *et al*., 2000; Mori *et al*., 1995). As shown in the psgSAGE tag frequency data, a molar ratio of fib H-chain : L-chain : P25 was not concord with the theoretical molar ratios. The molar ratio of

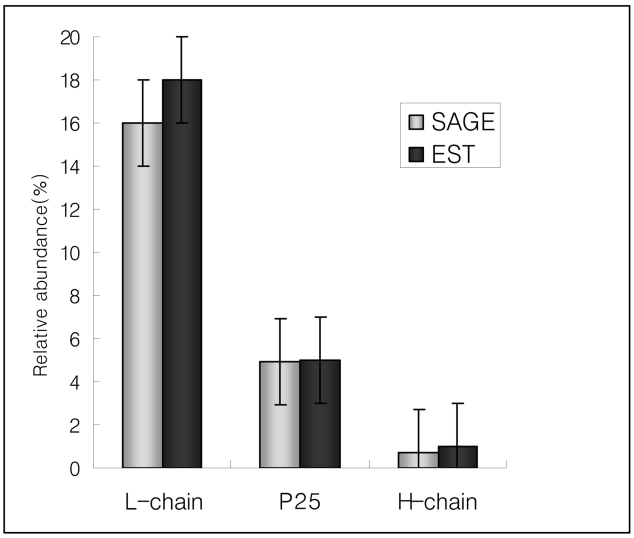

Fig. 1. Comparison of transcript abundance. Bars represent the percent of relative abundance (%) of the fibroin L- and H-chain and P25 as determined by EST analysis (dark bars, courtesy of Choi *et al.*, 2007) and SAGE tag analysis (light bars). Error bars represent the standard deviation.

fib H-chain, L-chain, and P25 was approximately 1 : 21 : 7. Fig. 1 shows the aspect of a relative molar ratio of the fib H-chain, L-chain, and P25 between the EST and SAGE methods. In the psgEST database, fib H-chain, L-chain and P25 gene were represented 30, 546, and 158 times, respectively, Therefore, the relative molar ratio of the fib H-chain, L-chain, and P25 in the psgEST was approximately 1 : 18 : 5, which was almost the same molar ratio with that of the psgSAGE.

It is true that any approaches do not reflect the accurate statistics of mRNA abundance in its transcripts, but the SAGE method is a simple and suitable method for providing quantitative information on gene expression that is not readily obtainable using other methods (Sun *et al.*, 2004). However, the SAGE method has many limitations for transcript detection, which are multiple matching and false matching problems of SAGE tags (Huang *et al.*, 2005; Sun *et al.*, 2004). In this study, these problems were also happened, nine tag sequences have double or triple matches in the 40 most redundant SAGE tags (Table 2). In addition, the third high expressed tag, "5'-CGGAGATTAT-3" does not match at the silkworm EST database, which corresponds to 3.4% of the total unique tags analyzed. Therefore, we presumed that its corresponding gene is involved in some mechanism of the silkworm biology. Further genetic biological analysis of this gene will give rise to more detailed identification and gene function in the silkworm.

In summary, we assumed that the total number of psgSAGE tags analyzed is not enough to represent all of the genes expressed in the PSG of silkworm. However, the psgSAGE data provide a brief overview of a quantitative aspect of silk processing mechanism related gene in the silkworm. The information reported in this study should be useful for researchers in the field to study on the mechanism of silk processing in insects.

Acknowledgement

This study was supported by the research grant of Rural Development Administration research (RDA) of Korea.

References

Choi, K. H., T. W. Goo, E. Y. Yun, J. S. Hwang and S. W. Kang (2007) Gene expression profile of the posterior silk glands of the silkworm, *Bombyx mori* L. *Korean J. Genetics* **29**, 193-201.

David, G. H. (2003) Genomics pure and applied entomology. *Annu. Rev. Entomol.* **48**, 235-260.

Funaguma, S., S. I. Hashimoto, Y. Suzuki, N. Omuro, S. Sugano, K. Mita, S. Katsuma and T. Shimada (2007) SAGE analysis of early oogenesis in the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* **37**, 147-154.

Goldsmith M. R., T. Shimada and H. Abe (2005) The genetics and genomics of the silkworm, *Bombyx mori*. *Annu. Rev. Entomol.* **50**, 71-100.

Huang, J., X. Miao, W. Jin, P. Couble, K. Mita, Y. Zhang, W. Liu, L. Zhuang, Y. Shen, C. Keime, O. Gandrillon, P. Brouilly, J. Briolay, G. Zhao and Y. Huang (2005) Serial analysis of gene expression in the silkworm, *Bombyx mori*. *Genomics* **86**, 233-241.

Imamura, M., J. Naki, S. Inoue, X. Q. Guo, T. Kanda and T. Tamura (2003) Targeted gene expression using the GAL4/UAS system in the silkworm *Bombyx mori*. *Genetics* **165**, 1329-1340.

Inoue, S., K. Tanaka, F. Arisaka, 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.

Jin, H. J. and L. K. David (2003) Mechanism of silk processing in insects and spiders. *Nature* **424**, 1057-1061.

Keith, A. B., D. Li, S. M. Jeffrey and C. M. Aldaz (2003) Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics* **19**, 1477-1483.

Marguiles, E. H. and J. W. Innis (2000) eSAGE: Managing and analyzing data generated with serial analysis of gene expression (SAGE). *Bioinformatics* **14**, 650-651.

Mori, K., K. Tanaka, Y. Kikuch, M. Waga, S. Waga and S. Mizuno (1995) Production of a chimeric fibroin light0chain

polypeptide in a fibroin secretion-deficient naked pupa mutant of the silkworm *Bombyx mori*. *J. Mol. Biol*. **251**, 217-228.

Sun, M., G. Zhou, S. Lee, J. Chen, R. Z. Shi and S. M. Wang (2004) SAGE is far more sensitive than EST for detecting low-abundance transcripts. *BCM genomics* **5**, 1-4.

Tanaka, K., N. Kajiyama, K. Ishikura, S. Waga, A. Kikuch, K. Ohtomo, T. Takai and S. Mizuno (1999) determination of the site of disulfide linkage between heavy and light chains of silk fibroin produced by *Bombyx mori*. *Biochem. Biophys. Acta*. **1432**, 92-103.

Tomita, M., H. Munetsuna, T. Sato, T. Adachi, R. Hino, M. Hayashi, K. Shimizu, N. Nakamura, T. Tamura and K. Yoshizato (2003) Transgenic silkworms produce recombinant human type III procollagen in cocoons. *Nature Biotech*. **21**, 52-56.

Velculescu, V. E., L. Zhang, B. Vogelstein and K. W. Kinzler (1995) Serial analysis of gene expression. *Science* **270**, 484-487.

Zhang, Y., J. Huang, S. Jia, W. Liu, M. Li, S. Wang, X. Miao, H. Xiao and Y. Huang (2007) SAGE tag based cDNA microarray analysis during larval during larval to pupal development and isolation of novel cDNA in *Bombyx mori*. *Genomics* **90**, 372-379.