Int. J. Indust. Entomol. Vol. 1, No. 2, 2000, pp. 171~175

[Communication]

International Journal of
Industrial Entomology

Identification of Highly Transcribed Genes in Japanese Oak Silkworm, *Antheraea yamamai, U*sing PCR-Based cDNA Library

Jin Sung Lee, Ki Hwan Kim, Tae Won Goo, Eun Young Yun, Seok Woo Kang, Dong Sang Suh[1] and Jae Sam Hwang*

Department of Sericulture and Entomology, National Institute of Agricultural Science and Technology, R.D.A., Suwon 441-100, Korea. [1]Department of Genetic Engineering, Sungkyunkwan University, Suwon 440-746, Korea.

(Received 22 October 2000; Accepted 22 November 2000)

Determined sequences of 384 randomly selected clones in a PCR-based cDNA library of *Antheraea yamamai* could identify expressed sequence tags (ESTs) of highly expressed gene. One EST (fibroin) appeared 15 times, one EST (40S ribosomal protein S18) twelve times, one EST (ribosomal protein S24a) eleven times, ten times (ribosomal protein S8), nine times (60S ribosomal protein L10A), seven times (40S ribosomal protein S15A, S17, S17 and seroin), six times (ribosomal protein S8), five times (ribosomal protein S24, mariner transposase and P8 protein), four times (serpin 2), three times (heat shock protein 70 and poly A binding protein), and the remaining 6 ESTs twice (amylase, KIAA1006, elongation factor-1, transposon mag, translation initiation factor 4C, QM protein, transposase). Therefore, the 94 EST make it possible to identify 24 redundant clones that are candidates for highly expressed genes in posterior silk gland of this insect. The 24 redundant EST clones were identified in GenBank, but none of them was related to *A. yamamai*, suggesting that there are many unidentified genes which are highly expressed in the *A. yamamai* genome.

Key words : *Antheraea yamamai*, PCR-based cDNA library, ESTs, Posterior silk gland

Introduction

The japanese oak silkmoth, *A. yamamai,* has been regarded as a very important insect for the production of silk and its by-products for a long time in human life. Thus, the insect has been regarded as an inducstrial one for the production of the cloth and food in oriental nations. The silk produced by the insect can be utilized as various textile sources. This insect is a potential host for the expression of foreign genes encoding industrially, agriculturally and medically important proteins, because it modifies protein post translationally, and they are abundant enough to be utilized for cellular function (Nagaraju *et al.*, 1996).

Recent progress in the recombinant technologies of insect as *Drosophila melanosgaster*, Zebrafish, and *Bombyx mori* can make recombinant strains easier than ever before. These technologies include an efficient way of DNA transfer into the organism, development of various vector system as well as identification of inducible and strong promoters (fibroin promoter of *Bombyx mori,* etc), isolation of many genes of industrial and agricultural value (Nagaraju *et al.*, 1996; Steller and Pirrotta, 1985; Stuart *et al.*, 1988). However promoters walking strongly in *A. yamamai* still need to be isolated, since strong promoters are not found in *A. yamamai.*

In the absence of the information on highly expressed genes, analysis of the mRNA population is a basis for the identification of the promoter of highly transcribed genes of which will be isolated. Because of difficulty of direct analysis, the mRNA population can be analyzed by indirect methods. One of the indirect methods for analyzing mRNA population is sequencing hundreds or thousands of randomly selected clones in a PCR-based cDNA library since only one cDNA molecule corresponding to the 3' end of mRNA is synthesized, and every vector molecule has only one cDNA insert (Belyavsky *et al.*, 1989; Bertioli *et al.*, 1995; Carninci *et al.*, 1996; Kimmel and Berger, 1987; Suzuki; 1997). Therefore, sequencing of the randomly selected cDNA clones generate a detailed profile of

*To whom correspondence should be addressed.
Department of Sericulture and Entomology, National Institute of Agricultural Science and Technology, R.D.A., Suwon 441-100, Korea. Tel: +82-31-290-8432; Fax: +82-31-290-2176; E-mail: hwangjs@rda.go.kr

the expressed gene in a cell type or in a diease (Adaye *et al.*, 1997; Affara *et al.*, 1994; Frigerio *et al.*, 1995; Hwang *et al.*, 1995; Korneev *et al.*, 1994; Konishi *et al.*, 1994; Liew *et al.*, 1994).

Recently, such analysis has been carried out in a few organisms and cell types (Krizman *et al.*, 1996; Lu *et al.*, 1988; Ryo *et al.*, 1998). In this paper, we constructed a PCR-based cDNA library from posterior silk gland (PSG) of *A. yamamai* and sequenced 384 clones to identify ESTs of highly expressed genes, which subsequently will be used as probes for isolation of cognate promoters using genomic DNA library or cosmid library

Materials and Methods

Insect and cloning vector

The posteior silk gland of *A. yamamai* was used for a raw material for PCR-baed cDNA library construction. *E. coli* XL1-Blue was used for host cell of pGem-T easy vector.

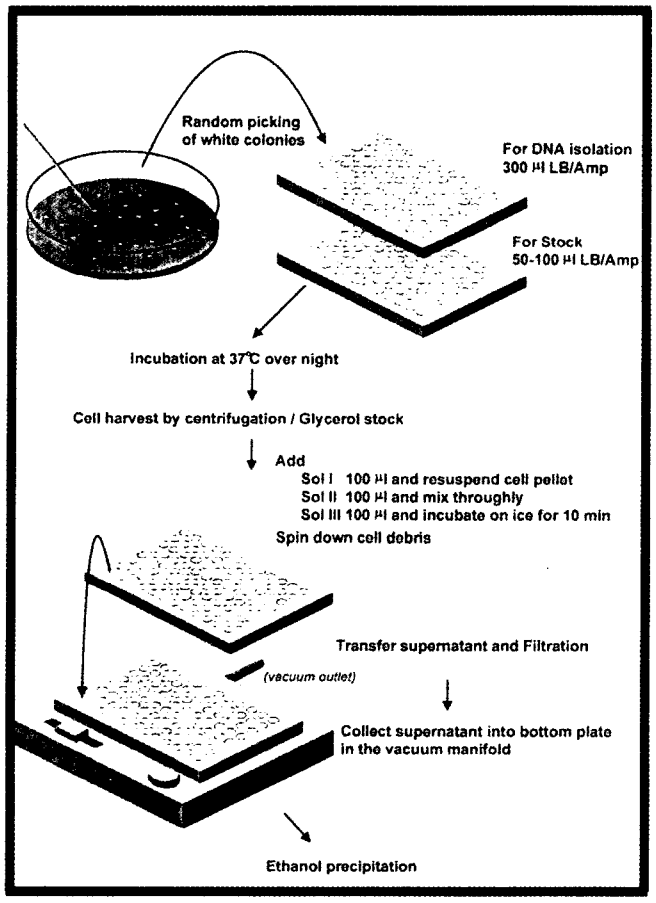

Fig. 1. Example for procedure of high-throughput plasmid DNA isolation for single-pass sequencing. Plasmid DNA containing supernatant was purified using 96-well formatted 0.22 m pore GV filtering unit.

E. coli strain was grown on an LB medium with addition of ampicillin at the concentration of 50 ug/ml.

Constuction of PCR-based cDNA library

Total RNA was obtained by grinding the PSG frozen in the presence of liquid nitrogen followed centrifugation on a $CsCl_2$ gradient of cell-free extract after being dissolved in guanidinethiocyanate (GTC) buffer. mRNA was isolated by affinity chromatography of isolated total RNA using oligo d(T)-cellulose column as proviously described (Sambrook et al., 1989). A PCR-based cDNA library using *Not*I and *Eco*RI adopter from PSG of *A. yamamai* was constructed from the isolated mRNA by SuperScript choice system for cDNA library (Gibco-BRL, USA). The recombinant plasmid harboring cDNA insert was prepared by high-throughput 96 well format (Fig. 1). Nucleotide sequence were determined in the direction of 5' to 3' of mRNA using ABI PRISM 377 DNA Sequencer (Applied Biosystems, USA) after polymerase chan reaction by Big-Dye ready reaction kit (Applied Biosystems, USA).

Sequence similarity analysis

EST sequences were identified by using a computer program, Lasergene (DNASTAR, USA). Similarities of the obtained sequences in the GenBank database was searched using BLAST network program in National Center for Biotechnology Information (Benson *et al.*, 1993).

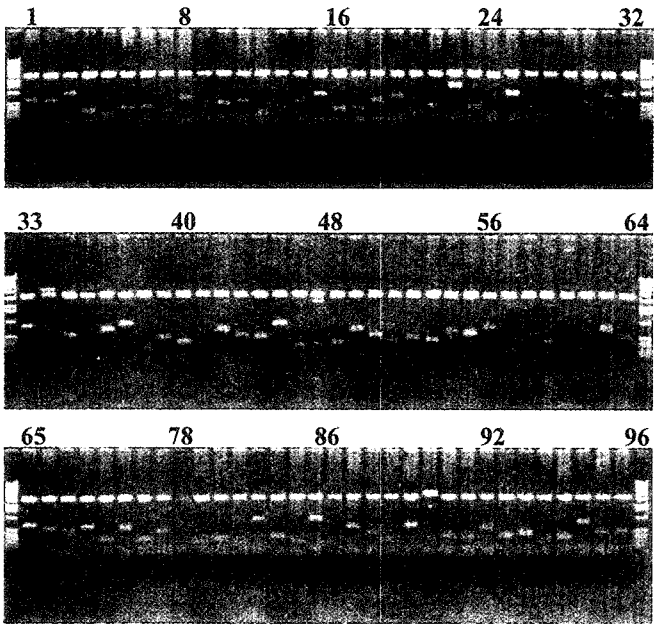

Fig. 2. Size distribution and cloning efficiency of 96 cDNA inserts. Cloned cDNA insert in pGem-T easy vector was digested with *Eco*RI restriction endonuclease and separated onto 1.2%/0.5 TBE agarose gel. Both side lanes represent 1 kb DNA ladder.

Results and Discussion

cDNA library construction and sequencing

The PCR-based cDNA library was constructed from posterior silk gland of *A. yamamai*. The quality of this library was determined by analyzing colony forming units (cfu), insert DNA size and ratio. The primary cfu is an important factor, when massive EST analysis is to be carried out. The primary cfu of cDNA library was 4.11×10^5, which is sufficient to include low abundant mRNA species. The insert ratio was measured by analyzing an insert from 96 randomly selected cDNA clone digested with *Eco*RI of pGem-T easy vector. All the clones that were examined harbored inserts with an average size of 0.5 to 2.0 kb (Fig. 2). Therefore, this cDNA library was estimated to be adequate for EST analysis. In total, nucleotide sequences of 384 randomly selected cDNA clones were determined. Among 384 clones, 24 clones containing insert shorter than 50 bp were excluded in the further analysis. In total, nucleotide sequences of 360 randomly selected cDNA clones were determined. Among the 360 clones, 21 having an insert shorter than 200 bp were excluded from sim-

Table 1. Redundancy of highly expressed gene from *Antheraea yamamai* posterior silk gland

Redundancy[1]	Identification	DB[2]	Species	E-value
15	Fibroin	P	*A.yamamai*	3e-91
12	40S ribosomal protein S18	P	*D. melanogaster*	8e-64
11	Ribosomal protein S24	P	*H. sapiens*	7e-15
10	Ribosomal protein S8	P	*A. pellifera*	4e-21
09	60S ribosomal protein L10A	P	*M. musculus*	4e-37
07	40S ribosomal protein S15A	P	*D. melanogaster*	8e-46
	40S ribosomal protein S17	P	*D. melanogaster*	2e-51
	Seroin	P	*G. mellonella*	0.009
	Ribosomal protein S27A	P	*M. secta*	8e-66
	40S ribosomal protein S18	P	*D. melanogaster*	2e-37
06	Ribosomal protein S8	P	*A. mellifera*	4e-45
05	Ribosomal protein S24	P	*M. musculus*	1e-44
	Mariner transposase	P	*A. mylitta*	1e-16
04	Serpin 2	P	*M. secta*	1e-20
03	Poly A binding protein	P	*D. melanogaster*	3e-46
	Heak shock protein 70	P	*T. ni*	7e-66
	P8 protein	P	*M. musculus*	3e-07
02	α-amylase	P	*B. licheniformis*	3e-92
	KIAA1006	P	*H. sapiens*	6e-15
	Elongation factor-1	P	*B. mori*	2e-28
	Transposon MAG	P	*B. mori*	3e-09
	Translation initiation factor 4C	P	*A. gambiae*	7e-40
	QM protein	P	*B. mandarina*	6e-77
	Transposase	P	*B. mori*	3e-34
	γ-aminobutyric acid	P	*H. sapiens*	0.003
	Polycomb protein	P	*D. melanogaster*	0.004
	Ceropin D	P	*H. cecropia*	3e-34
	cAMP regulated phosphoprotein	P	*H. sapiens*	1e-06
	Copper/Zinc SOD	P	*C. elegans*	0.006
	Glycosylceramidase	P	*R. rattus*	0.004
	Gag-like protein	P	*B. mori*	8e-43

[1]The redundancy of the EST among sequenced 360 clones
[2]DB indicates BLASTX similarity search of NCBI network service.

ilarity analysis, since such a short sequence may not show significant similarity when analyzed in the GenBank by chance.

EST similarity analysis

The single-pass sequencing of randomly selected cDNA clones containing insert DNA larger than 200 bp were performed to generate the ESTs derived from posterior silk gland of *A. yammamai*. The average length of the EST was 476 nucleotides. Each of the EST sequences were translated into three reading frames. And the primary sequences (amino acids) were analyzed using BLASTX or BLASTN public program (www.ncbi.nlm.nih.gov/BLAST). By single-pass sequencing, 349 ESTs were generated from this cDNA library. Out of 349 ESTs, 277 clones showed similarities to the GenBank registed genes, and 72 ESTs had not been identitied to the previously identified genes in the public database. Although these clones were isolated randomly, some clones had similar inserts due to amplification of original cDNA library and existence of abundant gene.

In case when more than one EST sequence was matched with the same gene in GenBank, a clone showing the lowest BLASTX e-value was considered as a non-redundant EST. As a result, 349 cDNA sequences yielded 151 different ESTs. The redundancy analysis reveled that 24 ESTs were considered as redundant clones. A summary of these results is presented in Table 1. The ESTs were matched genes originated from various organisms, such as bacteria, plant, insect and human. The 16 ESTs showed similarities to the proteins identified in insects, 4 ESTs to the proteins identified in mammal and, 4 ESTs in other eukaryotes.

Among 24 redundant EST clones, one EST (fibroin) showed up 15 times, one EST (40S ribosomal protein S18) twelve times, one EST (rbosomal protein S24a) eleven times, ten times (ribosomal protein S8), nine times (60S ribosomal protein L10A), seven times (40S ribosomal protein S15A, S17, S17 and seroin), six times (ribosomal protein S8), five times (ribosomal protein S24, mariner transposase and P8 protein), four times (serpin 2), three times (heat shock protein 70 and poly A binding protein), and the remaining 6 ESTs twice (amylase, KIAA 1006, elongation factor-1, transposon mag, translation initiation factor 4C, QM protein, tansposase). Therefore, the 349 EST make it possible to identify 24 redundant clones that are candidates for highly expressed genes in posterior silk gland of *A. yamamai*. If more ESTs are analyzed, it will be more accurate to identify the genes most highly expressed.

The sequence homology search showed that the ESTs involving ribosomal protein-related genes and silk synthesis-related genes were the most abundantly observed ones with the frequency of 26% (91/349) and 6.3% (22/349), respectively. This suggests that these genes are highly expressed in the posterior silk gland of silkmoths such as *B. mori, B. mandarina,* and *A. pernyi.* Although no seroin was discovered from major silkworms such as *B. mori*, and *B. mandarina* (Goldsmith and Kafatos, 1984). From this EST analysis, novel silk-related gene and seroin, which was first identified and cloned in *Galleria mellonella* was identified (Zurovec *et al.*, 1998). Otherwise, we suggest that *A. yamamai* has the silk-related genes of another type as well as H-H type fibroin (Kusuda *et al.*, 1986). Unexpectively, no EST was matched to *A. yamamai.* This result suggested that there are many of unidentified genes, the promoter of which acts strongly in *A. yamamai.*

We are now analyzing the relative expression levels between genes having the redundant ESTs. At present, we are trying to isolate the cognate prompters with two-dimensional cosmid library using redundant ESTs as molecular probes.

Acknowledgments

This study was supported by grants (G7 project) from the Ministry of Science and Technology of Korea.

References

Adjaye, J., R. Daniels, V. Bolton and M. Monk (1997) cDNA libraries from single human preimplantation. *Genomics* **46**, 337-344

Affara, N. A., E. Bentley, P. Davey, A. Pelmear and M. H. Johns (1994) The identification of novel gene sequences of the human adult testis. *Genome* **22**, 205-210.

Belyavsky, A., T. Vinogradova and K. Rajewsky (1989) PCR-based cDNA library construction: general cDNA libraries at the level of a few cells. *Nucleic Acids Res.* **17**, 2919-2932.

Bertioli, D., J. Schlicher, M. J. Adams, P. R. Burrows, H. H. Steinbiss and J. F. Anthoniw (1995) An analysis of differential display shows a strong bias towards high copy number mRNAs. *Nucleic Acids Res.* **23**, 4520-4523.

Benson, D., D. J. Lipman and J. Ostell (1993) GenBank. *Nucleic Acids Res.* **16**, 7583-7600.

Carninci, P., C. Kvam, A. Kitamura, T. Ohsumi, Y. Okazaki, M. Itoh, M. Kamiya, K. Shibata, N. Sasaki, M. Izawa, M. Muramatsu, Y. Hayashizaki and C. Scheneider (1996) High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* **37**, 327-336.

Frigerio, J. M., P. Benthezene, P. Garrido, E. Oritz, S. Barthellemy, S. Vasseur, B. Sastre, I. Sleznieff, J. C. Dagorn and

J. L. Iovanna (1995) Analysis of 2166 clones from a human coloretal cDNA library by partial sequencing. *Hum. Mol. Genet.* **4**, 37-43.

Goldsmith. M. G. and F. C. Kafatos (1984) Deveolpmentally regulated genes in silkmoths. *Anu. Rev. Genet.* **18**, 443-487.

Hwang, D. M., W. S. Hwang and C. C. Liew (1994) Single pass sequencing of a unidirectional human cDNA library to discover novel genes of the cardiovascular system. *J. Mol. Cell. Cardiol.* **26**, 1329-1333.

Kimmel, R. and S. L. Berger (1987) Preparation of cDNA and the generation of cDNA libraries: Overview. *Methods in Enzymol.* **52**, 307-316.

Korneev, S., S. Blackshaw and J. A. Davis (1994) cDNA libraries from a few neural cells. *Prog. Neurobiol.* **42**, 229-246.

Konishi, K., Y. Morishima, E. Ueda, Y. Kibe, K. Nonomura, K. Yamanish and H. Yasuno (1994) Cataloguing of the genes in human keratinocytes: Analysis of 607 randomly selected cDNA sequences. *Biochem. Biophs. Res. Comm.* **202**, 976-983.

Krizman, D. B., R. F. Chuaqui, P. S. Meltzer, J. M. Trent, P. H. Duray, W. M. Linehan, L. A. Liotta and M. R. Emmert-Buck (1996) Construction of a representative cDNA library from prostatic intraepithelial neoplasia. *Cancer Res.* **56**, 5380-5383.

Kusuda, J., Y. Tazzima, Y. Onimura, O. Ninaki and Y. Suzuki (1986) The sequence around the 5' end of the fibroin gene from the wild silkworm, *Bombyx mandarina*, and comparison with that of the domesticated species, *B. mori*. *Mol. Gen. Genet.* **203**, 359-364.

Liew, C., M. Hwang, Y. W. Fung, C. Laurenssen, E. Cukerman, S. Tsui and C. Y. Lee (1994) A catalogue of genes in the cardiovascular system as identified by express sequence taqs (ESTs). *Proc. Natl. Acad. Sci. USA* **91**, 10645-10649.

Lu. X. and D. Werner (1988) Construction and qualtity of cDNA libraries prepared from cytoplasmic RNA and not enriched in poly(A)$^+$ RNA. *Gene* **71**, 157-164.

Nagaraju, J. and T. Kanda (1996) Attempt at transgenesis of the silkworm, *Bombyx mori* by egg-injection of foreign DNA. *Appl. Entomol. Zool.* **31**, 587-596.

Ryo, A., N. Kodoh, T. Wakasaki, A. Hada, N. Yamamoto and M. Yamamoto (1998) A method for analyzing the qualitative and quantitative aspects of gene expression: a transcriptional profile revealed for HeLa cells. *Nucleic Acids Res.* **26**, 2586-2592.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989) Molecular cloning: a laboratory manual, 2nd ed.,Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Steller, H. and V. Pirrotta (1985) Fate of DNA insjected into early *Drosophila* embryos. *Deveop. Biol.* **109**, 54-62.

Stuart, G. W., J. V. McMurray and M. Westerfield (1988) Replication, integration and stable genr-line transmission of foreign sequences injected into early zebrafish embryos. *Development* **103**, 403-412.

Suzuki, Y., K. Yoshitomo-Nakagawa, K. Maruyama and A. Suyama (1997) Construction and characterization of a full length-enriched and a 5'-end enriched cDNA library. *Gene* **200**, 149-156.

Zurovec, M., C. Yang, D. Kodrick and F. Schnal (1998) Identification of a novel type of silk protein and regulation of its expression. *J. Biol. Chem.* **273(25)**, 15423-15428.