SAGE analysis of early oogenesis in the silkworm, *Bombyx mori*

**Shunsuke Funaguma ^a, Shin-ichi Hashimoto ^b, Yutaka Suzuki ^c, Naoko Omuro ^a,
Sumio Sugano ^c, Kazuei Mita ^d, Susumu Katsuma ^a, Toru Shimada ^{a*}**

* Corresponding author

Affiliation: ^a Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan.

^b Department of Molecular Preventive Medicine, School of Medicine The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

^c Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo 5-1-5, Kashiwanoha, Kashiwa, Chiba 277-8562, Japan.

^d National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8643, Japan.

Running title: Construction of *Bombyx* 3'Long SAGE libraries.

Key words: SAGE; ovary; differentiation; *Bombyx mori*

Corresponding author. Tel.: +81-3-5841-8124; fax: +81-3-5841-5089.

E-mail address: shimada@ss.ab.a.u-tokyo.ac.jp (T. Shimada)

Abstract

To identify genes involved in the differentiation of *Bombyx* cystoblast, we constructed two 3' long serial analysis of gene expression (Long SAGE) libraries from stage 1-3 or stage 2-3 egg chambers and compared their gene expression profiles. In both libraries, the most frequent tags were derived from the same novel transcript. The transcript does not have any open reading frame capable of encoding a protein with over 100 amino acids in length. RNA blot analysis revealed that this transcript is specifically and abundantly expressed in the *Bombyx* ovary, mainly the germ line cells in the ovarioles. These results suggest that *Bombyx* oogenesis may be regulated by a previously unidentified non-coding RNA. Comparison of the gene expression profiles between the stage 1-3 and stage 2-3 egg chamber libraries revealed that 272 tags were significantly more abundant in stage 1-3 egg chambers ($p < 0.05$ and at least two-fold change) than in library 2. Among the differentially expressed transcripts were the sequences that correspond to *ATP synthase subunit d* (3.1-fold enriched) and *ATP synthase coupling factor 6* (9.1-fold enriched), suggesting that they are involved in regulation of cell cycle of cystocytes.

55 **Introduction**

56 Oogenesis in insects consists of three steps, oocyte-nurse cell differentiation,
57 previtellogenesis and vitellogenesis. In *Bombyx mori*, there are 3 mitotic divisions of the
58 cystoblast with 3 incomplete cytokinesis which results in cytoplasmic bridges between
59 all daughter cystocytes. As is typical for Lepidoptera, all eight cystocytes initiate
60 meiosis together. One cystocyte later differentiates into the oocyte and continues along
61 the meiotic pathway, whereas the remaining cystocytes divert to an endomitotic
62 pathway and become nurse cells (Orikasa et al., 1993). Previtellogenesis involves the
63 growth of both oocytes and nurse cells before deposition, while during vitellogenesis,
64 the oocyte starts to accumulate yolk proteins and its growth becomes accelerated with
65 respect to the nurse cells (Büning, 1994).

66 In *Drosophila*, cystoblasts turn on the *bag-of-marbles* (*bam*) gene, which is required
67 for differentiation, and divide synchronously to generate a 16-cell germ cell cyst
68 interconnected by ring canals that become surrounded by somatic cells to form a new
69 follicle (Kai et al., 2004; McKearin and Spradling, 1990). However, homologs of *bam*
70 have not been isolated in any other organisms including *Bombyx mori*. To better
71 understand control of cystoblast differentiation in non-*Drosophila* species, we have
72 initiated studies aimed at identification of genes that control differentiation in *Bombyx*

73 *mori*.

74 Techniques based on complementary DNA subtraction or differential display can be
75 quite useful for comparing gene expression in two different cell types. However, these
76 methods provide only a partial picture of differential expression, providing with no
77 direct information about abundance of specific transcripts in the two different cell types.
78 The expressed sequence tag (EST) approach is a valuable tool for gene discovery, but
79 like RNA blotting, ribonuclease protection, and reverse transcriptase-polymerase chain
80 reaction analysis, EST evaluates only a limited number of genes at a time (Velculescu et
81 al., 1995).

82 Serial analysis of gene expression (SAGE), by contrast, is one of the more versatile
83 methods for genome-level studies of gene expression. The original SAGE technology
84 samples short sequence tags (14 bases long) from mRNA transcripts found in the
85 population of interest. Each tag contains sequence information that can identify, in basic
86 local alignment search tool analysis, the transcript from which an individual tag has
87 been derived (Velculescu et al., 1995). The frequency of each tag in the SAGE library is
88 an accurate estimate of the abundance of its corresponding mRNA transcript. Moreover,
89 SAGE is particularly well suited to the analysis of organisms whose genomes are not
90 completely sequenced, since it does not require a hybridization probe for each transcript

and provides quantitative information on gene expression that is not readily obtainable using other methods (Velculescu et al., 1995).

The original SAGE technique does, however, have a shortcoming in that some tags are short enough that they have multiple matches because of insufficiency of tag length.

Recently, Saha et al. developed the 3' long serial analysis of gene expression (Long SAGE) method, which generates 20 or 21 bp tags derived from 3' end of transcripts that can be rapidly analyzed and precisely matched to genomic sequence data and allows a detailed analysis of thousands of transcripts simultaneously (Saha et al., 2002). This method can be used not only to obtain quantitative information on the abundance of known transcripts but also to identify novel expressed genes. There were some reports on SAGE libraries in *Bombyx mori* (Huang et al., 2005a,b). Because their libraries and libraries constructed in this study were derived from different tissues and developmental stages, many kinds of SAGE tags were wished to be obtained. They will be useful for the silkworm genome project (Mita et al., 2004; Xia et al., 2004) by providing detailed information about transcriptome.

In *Bombyx mori*, the process of oogenesis is divided into twelve detailed stages. During stage 1, eight cystocytes start to differentiate into oocytes or nurse cells as described above (Yamauchi and Yoshitake, 1984). A purpose of this experiment was to

identify genes that control cystoblast differentiation into oocytes or nurse cells in
Bombyx mori. But collecting intact stage 1 egg chambers is technically very difficult.
Thus, we constructed two 3' Long SAGE libraries from stage 1-3 egg chambers and
stage 2-3 egg chambers and compared the gene expression profiles at these stages to
find transcripts that were significantly more abundant in library 1 (stage 1-3).

Materials and Methods

Insects

The inbred silkworm strain p50T (Daizo) that is maintained at the University of Tokyo was used as described previously (Daimon et al., 2003).

Construction of the SAGE libraries

To isolate stage 2-3 egg chambers, we removed a germarium from stage 2-3 egg chambers under a light microscope. But we could not investigate the degree of cross contamination in the two selected pools of egg chambers because no gene was known to be expressed specifically in stage 1 or stage 4 egg chambers. Total RNA was isolated from egg chambers (stage 1-3 or stage 2-3) using the TRIzol reagent (Invitrogen) and mRNA was selected using a Micro-Fast Track 2.0 mRNA-isolation kit (Invitrogen) according to the protocol provided by the manufacturer. Long SAGE was performed with 2.5 µg mRNA as described previously (Hashimoto et al., 2003). SAGE2000 (ver.4.5) was used to quantify the abundance of each tag (Velculescu et al., 1995). The two ovary SAGE libraries were deposited in the SAGEmap database at NCBI (Accession numbers **GSM119465** and **GSM119467**).

SAGE tag annotation

We assembled the cDNA library sequences (published on the Web at <http://morus.ab.a.u-tokyo.ac.jp>.) by using HASTE (Paracel) and CAP4 (Paracel). The identity of mRNAs corresponding to the SAGE tags was determined via comparison with the assembled *Bombyx* EST database (Shimada et al., unpublished). Annotation of the EST database was performed as described below. The combined set of the consensus EST sequences was analyzed as a putative “unigene” set. Each unigene was compared against the fly protein database by using BLASTX. *Bombyx* unigene were automatically assigned a gene annotation corresponding to the highest statistical match found ($E \text{ value} < 1 \times 10^{-10}$) and that match was considered to be a putative fly homolog of the *Bombyx* gene. Gene Ontology (GO) classifications (Ashburner et al., 2000) of the corresponding fly homologs were obtained from Flybase (<http://flybase.bio.indiana.edu/>).

Statistical analysis

Statistical significance (p -value) in the evaluation is an expansion of Audic and Claverie’s method to manage more than two samples (Audic and Claverie, 1997). To estimate the extent of similarity between the two libraries we constructed and tested,

library 2 was normalized to the tag number of library 1 (122060). To avoid division by zero, a tag value of 1 was used for any tag assigned the value zero in library 2. The condition for significant differences between the two libraries was set at a *p*-value of < 0.05 and at least two-fold change.

cDNA cloning and DNA sequencing

Full-length cDNA of the most abundant transcript in both SAGE libraries was isolated by rapid amplification of cDNA ends (RACE) using GeneRacer Kit (Invitrogen) according to the manufacturer's instruction. Primers for 5'- and 3'-RACE are SAGETopGSP-F (5'-GCGCATATCCATAATGGGGGACT-3'), SAGETopGSP-R (5'-AGTGAAGTTACTCGCGACTCTGC-3'), SAGETopGSP-Fnest (5'-CCATCGCGTCATTCGTCGAACTT-3'), and SAGETopGSP-Rnest (5'-CTGCCAGCCATGCCTAAACTAAG-3'). These cloned DNAs were sequenced with the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit and the ABI PRISM 3100 Genetic Analyzer.

RNA blot analysis

Total RNA was extracted from indicated tissues using the TRIzol reagent (Invitrogen),

and 3 µg of total RNA were separated electrophoretically on 1.2% agarose gel containing 2.2M formamide in a MOPS buffer and transferred onto a Biodyne Plus nylon membrane (Pall). The membrane was hybridized at 68°C overnight in DIG Easy Hyb Granules (Roche). Next, the membrane was washed twice in 2 x SSC, 0.1% SDS at room temperature and then twice in 0.1 x SSC, 0.1% SDS at 68°C. Hybridization signals were detected using a LAS 1000 (Fujifilm) according to the manufacturer's instruction. A DIG-labeled probe that recognized the most abundantly expressed transcript in the *Bombyx* ovary was transcribed from a DNA template using a DIG RNA labeling kit (Roche). The DNA template for the RNA probe was amplified by PCR with primers SAGEtopF (5'-AGACAAACGCGCATATCCAT-3') and T7SAGEtopR (5'-TAATACGACTCACTATAGGGCTGACTGTTGCGAATGAAA-3').

Real-time quantitative RT-PCR

Total RNA was subjected to reverse transcription by using avian myeloblastosis virus (AMV) reverse transcriptase and random primer (TaKaRa). Real-time quantitative PCR was carried out as previously described (Koike et al., 2003). Primers for PCR were as follows: ATPsynthasecouplingfactor6, 5'-AGCGGCAGCTCAGAAAGCTA-3' and 5'-CCCGCGCTCTTCTGTTTGTA-3'; ATPsynthasesubunitd, 5'-

CGGCACTTGTAGAATCGCAG-3' and 5'- TTTTGGTAGGATGCAATGTTGG-3';
the most frequently expressed RNA identified in stage 1-3 and 2-3 egg chambers, 5'-
TGGCAGAGTCGCGAGTAAGTT-3' and 5'- TTGCTCCCAGAACAAGAATTCA-3';
Actin3, 5'- GATGTGGATCTCGAAACAGGAGTAC-3' and 5'-
TGCTGTAGGGCATCATAACTTAAAGTC-3'. Relative expression levels were
calculated after correction of expression of Actin3.

Nucleotide sequence accession number

The nucleotide sequence reported in this paper has been submitted to the DDBJ,
EMBL, and NCBI Genbank data banks (Accession number **AB266084**).

Results

SAGE tag abundance in stage 1-3 egg chambers

A total of 122,060 tag sequences from the stage 1-3 egg chambers (library 1) allowed the identification of 40,311 different transcripts (Table 1). 1,773 unique tags hit single *Bombyx* EST sequences. 284 unique tags hit two or more *Bombyx* EST sequences (Table 1). Table 3-1 shows the top 50 transcripts in library 1. This library contains a lot of genes encoding ribosomal protein-encoding genes. However, the EST corresponding to the most frequently expressed tag in library 1 (expression frequency of 2.8%) did not share significant similarity any known gene (Table 3-1).

SAGE tag abundance in stage 2-3 egg chambers

A total of 37,073 tag sequences from the stage 2-3 egg chambers (library 2) allowed the identification of 16,740 different genes (Table 1). 1,123 unique tags corresponded to single *Bombyx* EST sequences. 209 unique tags corresponded to two or more *Bombyx* EST sequences (Table 1). Table 3-2 shows the top 50 transcripts in the stage 2-3 egg chamber library. As with library 1, this library contains a lot of genes that encode ribosomal proteins. The EST corresponding to most frequently expressed tag in library 2 (expression frequency of 3.0%) was the same transcript as that found in library 1

(Tables 3-1 & 3-2).

Isolation of the most frequently expressed RNA identified in stage 1-3 and 2-3 egg chambers.

To obtain a full-length cDNA corresponding to the most frequent expressed tag in both libraries, we performed 5'- and 3'-RACE. When 3'-RACE was performed, three different types of nucleotide sequence were obtained. When 5'-RACE was performed, a single type of the sequence was obtained (Fig. 1). The full-length cDNA isoforms we isolated had nucleotide sequence similarity to no genes. Moreover, none of the transcripts had an open reading frame that encoded more than 100 amino acids. The longest and shortest clones were 843bp and 709bp, respectively. An 816bp-long cDNA was also identified. By the BLASTn search against *Bombyx* genome database (Mita et al., 2004; Xia et al., 2004), all isoforms appeared to have three exons and two introns (Fig. 1). While first exon, first intron, second exon, and second intron were the same sequences among the three isoforms, the end sites of third exon varied. We speculated that these three variants might to be produced by the alternative polyadenylation mechanism (Fig. 1). We next analyzed expression of the transcript in various *Bombyx* larval or pupal tissues by RNA blotting. Detection of the transcript on the RNA blot

clearly shows that the transcript was specifically detectable in the ovary using an anti-sense probe. The molecular size of this transcript was in good agreement with the predicted size from RACE experiment (Fig. 2(a)). There was no SAGE tag corresponding to type C isoform (Fig. 1). Type A and type B isoforms (Fig. 1) are almost same in length (843bp and 816bp, respectively). Thus, we did not investigate the ratio between these two isoforms by performing RNA blot. No signals were detected using sense probe (data not shown). Furthermore, the transcript was detected in isolated mature eggs which consist of only germ line cells like in whole ovarioles (Fig. 2(b)). These results suggest that this transcript is a novel non-coding RNA that is specifically and abundantly expressed in the *Bombyx* ovary.

Comparison of stage 1-3 and 2-3 egg chamber gene expression profiles

Comparison of the expression patterns of the two libraries revealed that most genes (97.5% of library 1) were expressed at similar levels in the two libraries. However, we also identified 272 transcripts that were significantly more abundant in library 1 ($p < 0.05$ and at least two-fold change; Table 2). Fifty-four tags of 272 tags differently expressed between two SAGE libraries got hit to single *Bombyx* EST. We annotated these 54 tags by using the Flybase information to obtain Gene Ontology classification.

ESTs corresponded to 41 tags of these 54 tags had fly homologs. On the other hand, ESTs corresponded to 13 tags expect for these 41 tags did not have any homologous transcript in NCBI database. One of 272 tags were identified as *ATP synthase subunit d* (3.1-fold enriched) and *ATP synthase coupling factor 6* (9.1-fold enriched) (Table2). In addition, *Heat shock protein 68*, *stress-sensitive B*, *Ecdysone-inducible gene L2*, *CG11267*, *CG7430*, *Ribosomal protein L18A*, *Ribosomal protein L37a*, *Rm62* and *elongation factor 1 beta* were expressed more abundantly in library 1 (at least 5-fold change or greater).

Real-time quantitative RT-PCR

To confirm our SAGE data for egg chambers (stage 1-3 and stage 2-3), we performed real-time quantitative RT-PCR. Two differently expressed genes (*ATP synthase subunit d* and *ATP synthase coupling factor 6*) and an equally expressed gene (the most abundant transcript identified in stage 1-3 and 2-3 egg chambers) were selected and evaluated. The two differently expressed genes were expressed abundantly in stage 1-3 egg chamber (Table 4).

Discussion

In *Bombyx mori*, one of the set of eight sister cystocytes that derive from the cystoblast differentiates into the oocyte and the remainder become nurse cells (Orikasa et al. 1993). Here, we used the 3' Long SAGE approach to identify candidate genes that may be involved in cystoblast differentiation in *Bombyx mori*. To do this, we compared the gene expression profiles of a stage 1-3 and a stage 2-3 egg chamber library.

Interestingly, we found that most abundant transcript in the stage 1-3 and 2-3 egg chamber libraries lacked similarity to any known gene and does not appear to be a protein-coding gene. The transcript is abundantly and specifically expressed in the ovary (Fig. 2(a)). Furthermore, the transcript was detected in isolated mature eggs as well as in whole ovarioles (Fig. 2(b)). These results suggested that this transcript existed in germ line cells. Some non-coding RNAs are known to have biological functions in *Drosophila*. For example, the *RNA on the X* (*roX1* and *roX2*) transcripts function in localizing the chromatin remodeling activity of the male-specific lethal complex by binding specifically to male-specific lethal proteins (Kelly, 2004). Also, RNA produced by the *polar granule component* (*pgc*) gene, which is found in the pole plasm of *Drosophila* embryos (Nakamura et al., 1996), is known to mediate transcriptional repression, apparently as a result of transcription factor sequestration (Martinho et al.,

2004). Similarly, the most frequent expressed transcript identified in this work may have some important roles in oogenesis.

Fifty-four tags of 272 tags differently expressed between two SAGE libraries got hit to single *Bombyx* EST and could be annotated. On the other hand, 208 tags of them did not get hit to *Bombyx* ESTs. Thus, if we made cDNA libraries derived from *Bombyx* pupal ovaries, we could get more information about these 272 tags. Among the 272 transcripts that were more abundant in the stage 1-3 library (Table 2) were *Bombyx* homolog of *ATP synthase subunit d* (3.1-fold enriched) and *ATP synthase coupling factor 6* (9.1-fold enriched) (Table2). Recent studies revealed that the mutation, which they named *tenured*, mapped to a deletion in the gene encoding the cytochrome c oxidase subunit Va (CoVa) resulting in reducing Cyclin E abundance (but not expression of the cyclin E transcript) with reduced levels of ATP and block of cell cycle progression in *Drosophila melanogaster* (Mandal et al., 2005). Furthermore, endocycle is controlled by Cyclin E in *Drosophila* nurse cells (Lilly et al., 1996). All of cytochrome c oxidase subunit Va (CoVa), ATP synthase are involved in oxidative phosphorylation and production of ATP (Saraste et al., 1999). In *Bombyx* egg chamber, they may be involved in regulation of cell cycle of cystocytes by regulation of cyclin E abundance.

Recently, use of an RNA interference-mediated loss-of-function approach was

reported for *Bombyx* (Quan et al., 2002; Tabunoki et al., 2004; Isobe et al., 2004; Tanaka et al., 2005; Ohnishi et al., 2006). To clarify the role of specific genes in differentiation of *Bombyx* oogenesis, including the roles of and of the most abundant transcript detected via 3' Long SAGE in this work, we are currently planning to determine the effect on oogenesis of knock-down of these transcripts using this newly reported RNAi-dependent gene silencing approach.

References

- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25-29.
- Audic, S., and Claverie, J.M. 1997. The significance of digital gene expression profiles. Genome Res. 7, 986-995.
- Büning, J. 1994. In: The Insect Ovary. Ultrastructure, Previtellogenic Growth and Evolution. Chapman & Hall, London, Glasgow, Weinheim, New York, Tokyo, Melbourne, Madras, 400pp.
- Daimon, T., Hamada, K., Mita, K., Okano, K., Suzuki, M.G., Kobayashi, M., Shimada, T. 2003. A *Bombyx mori* gene, *BmChi-h*, encodes a protein homologous to bacterial and baculovirus chitinases. Insect Biochem.Mol.Biol. 33, 749-59.

361 Hashimoto, S., Morohoshi, K., Suzuki, T., Matsushima, K. 2003.
 362 Lipopolysaccharide-inducible gene expression profile in human monocytes. Scand. J.
 363 Infect Dis. 35, 619-627.
 364
 365 Huang, J., Miao, X., Jin, W., Couble, P., Mita, K., Zhang, Y., Liu, W., Zhuang, L., Shen,
 366 Y., Keime, C., Gandrillon, O., Brouilly, P., Briolay, J., Zhao, G., Huang, Y. 2005a.
 367 Serial analysis of gene expression in the silkworm, *Bombyx mori*. Genomics 86,
 368 233-241.
 369
 370 Huang, J., Miao, X., Jin, W., Couble, P., Zhang, Y., Liu, W., Shen, Y., Zhao, G., Huang,
 371 Y. 2005b. Radiation-induced changes in gene expression in the silkworm revealed by
 372 serial analysis of gene expression (SAGE). Insect Mol. Biol. 14, 665-674.
 373
 374 Isobe, R., Kojima, K., Matsuyama, T., Quan, G.X., Kanda, T., Tamura, T., Sahara, K.,
 375 Asano, S.I., Bando, H. 2004. Use of RNAi technology to confer enhanced resistance to
 376 BmNPV on transgenic silkworms. Arch Virol. 149, 1931-1940.
 377
 378 Kai, T., Williams, D., and Spradling, A.C. 2005. The expression profile of purified

379 *Drosophila* germline stem cells. Dev. Biol. 283, 486-502.

380

381 Kelly, R.L. 2004. Path to equality strewn with *roX*. Dev. Biol. 269, 18-25.

382

383 Koike, Y., Mita, K., Suzuki, M.G., Maeda, S., Abe, H., Osoegawa, K., deJong, P.J.,

384 Shimada, T. 2003. Genomic sequence of 320 kb containing a *kettin* orthologue on the Z

385 chromosome in *Bombyx mori*. Mol. Gen. Genomics 269, 137-149.

386

387 Lilly, M.A., Spradling, A.C. 1996. The *Drosophila* endocycle is controlled by Cyclin E

388 and lacks a checkpoint ensuring S-phase completion. Genes Dev. 10, 2514-2526.

389

390 Mandal, S., Guptan, P., Owusu-Ansah, E., Banerjee, U. 2005. Mitochondrial regulation

391 of cell cycle progression during development as revealed by the tenured mutation in

392 *Drosophila*. Dev. Cell 9, 843-854.

393

394 Martinho, R.G., Kunwar, P.S., Casanova, J., Lehmann, R. 2004. A noncoding RNA is

395 required for the repression of RNAPIII-dependent transcription in primordial germ

396 cells. Curr. Biol. 14, 159-165.

397

398 McKearin, D.M., and Spradling A.C. 1990. *bag-of-marbles*: a *Drosophila* gene required
399 to initiate both male and female gametogenesis. *Genes Dev.* 12, 2242-2251.

400

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

406

407 Nakamura, A., Amikura, R., Mukai, M., Kobayashi, S., and Lasko, P.F. 1996.
408 Requirement for a noncoding RNA in *Drosophila* polar granules for germ cell
409 establishment. *Science* 274, 2075-2079.

410

411 Ohnishi, A., Hull, J.J., and Matsumoto, S. 2006. Targeted disruption of genes in the
412 *Bombyx mori* sex pheromone biosynthetic pathway. *Proc. Natl. Acad. Sci. USA* 103,
413 4398-4403.

414

415 Orikasa, C., Yamauchi, H., Nagasawa, H., Suzuki, A., and Nagata, M. 1993. Induction
 416 of oocyte-nurse cell differentiation in the ovary by the brain during the initial stage of
 417 oogenesis in the silkworm, *Bombyx mori* (Lepidoptera: Bombycidae). Appl. Entomol.
 418 Zool. 28, 303-311.
 419
 420 Quan, G.X., Kanda, T., Tamura, T. 2002. Induction of the white egg 3 mutant
 421 phenotype by injection of the double-stranded RNA of the silkworm white gene. Insect
 422 Mol. Biol. 11, 217-222.
 423
 424 Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler,
 425 K.W., and Velculescu V.E. 2002. Using the transcriptome to annotate the genome. Nat.
 426 Biotechnol. 20, 508-512.
 427
 428 Saraste, M. 1999. Oxidative phosphorylation at the fin de siecle. Science 283,
 429 1488-1493.
 430
 431 Tabunoki, H., Higurashi, S., Ninagi, O., Fujii, H., Banno, Y., Nozaki, M., Kitajima, M.,
 432 Miura, N., Atsumi, S., Tsuchida, K., Maekawa, H., Sato, R. 2004. A carotenoid-binding

433 protein (CBP) plays a crucial role in cocoon pigmentation of silkworm (*Bombyx mori*)
 434 larvae. FEBS Letters. 567, 175-178.
 435
 436 Tanaka, H., Yamamoto, M., Moriyama, Y., Yamao, M., Furukawa, S., Sagisaka, A.,
 437 Nakazawa, H., Mori, H., Yamakawa, M. 2005. A novel Rel protein and shortened
 438 isoform that differentially regulate antibacterial peptide genes in the silkworm *Bombyx*
 439 *mori*. Biochim. Biophys. Acta. 1730, 10-21.
 440
 441 Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W. 1995. Serial Analysis of
 442 Gene Expression. Science. 270, 484-487.
 443
 444 Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai,
 445 C., Pan, G., Xu, J., Liu, C., Lin, Y., Qian, J., Hou, Y., Wu, Z., Li, G., Pan, M., Li, C.,
 446 Shen, Y., Lan, X., Yuan, L., Li, T., Xu, H., Yang, G., Wan, Y., Zhu, Y., Yu, M., Shen,
 447 W., Wu, D., Xiang, Z., Yu, J., Wang, J., Li, R., Shi, J., Li, H., Li, G., Su, J., Wang, X.,
 448 Li, G., Zhang, Z., Wu, Q., Li, J., Zhang, Q., Wei, N., Xu, J., Sun, H., Dong, L., Liu, D.,
 449 Zhao S., Zhao, X., Meng, Q., Lan, F., Huang, X., Li, Y., Fang, L., Li, C., Li, D., Sun, Y.,
 450 Zhang, Z., Yang, Z., Huang, Y., Xi, Y., Qi, Q., He, D., Huang, H., Zhang, X., Wang, Z.,

451 Li, W., Cao, Y., Yu, Y., Yu, H., Li, J., Ye, J., Chen, H., Zhou, Y., Liu, B., Wang, J., Ye,
452 J., Ji, H., Li, S., Ni, P., Zhang, J., Zhang, Y., Zheng, H., Mao, B., Wang, W., Ye, C., Li,
453 S., Wang, J., Wong, G.K., Yang, H.; Biology Analysis Group. 2004. A draft sequence
454 for the genome of the domesticated silkworm (*Bombyx mori*). Science 306, 1937-1940.

455

456 Yamauchi, H., and Yoshitake, N. 1984. Developmental stages of ovarian follicles of the
457 silkworm, *Bombyx mori* L. J. Morphol. 179, 21-31.

458

459

460

461

462

463

464

465

466

467

468

469

470 **Acknowledgement**

471 We are grateful to Hiroyuki Izu and Kensuke Suzuki for silkworm EST assembly. This
472 work was supported by the NIAS/MAFF Insect Technology Program (to T.S.),
473 Grants-in-Aid for Scientific Research from MEXT/JSPS (No. 17018007 and 16208006
474 to T.S.), JSPS Research Fellowships for Young Scientists (16•11612 to S.F.) and the
475 Professional Program for Agricultural Bioinformatics, JST, Japan.

476

477

478

479

480

481

482

483

484

485

486

487

488 **Figure legends**

489 **Table 1**

490 Experimental matching of SAGE tags to the *Bombyx* EST database.

491

492 **Table 2**

493 Comparison of expression profiles of the stage1-3 and 2-3 egg chamber libraries.

494 Column one, gene ontology (GO) terms according to Flybase; column two, gene names

495 according to Flybase; column three, tag sequence; column four, the ratio of the tag

496 number in library 1 (stage 1-3 egg chambers) to the tag number in library 2 (stage 2-3

497 egg chambers). The tag number in library 2 was normalized to that of library 1 (i.e.

498 122,060).

499

500 **Table 3-1**

501 The 50 3'-SAGE tags most frequently isolated from the stage 1-3 egg chamber library

502 (library 1). Column one, the tag sequence; column two, gene names according to

503 Flybase. Column three, tag abundance among 122,060 tags. 'EST' indicates that this tag

504 can be found in the *Bombyx* EST database but function is unknown. 'Multi hit' indicates

that the tag corresponds to two or more EST sequences. ‘No hit’ indicates that the tag does not correspond to any EST sequence.

Table 3-2

The 50 3’-SAGE tags most frequently isolated from stage 2-3 egg chambers (library 2). Column one, tag sequence; column two, gene name (according to Flybase). Column three shows the tag abundance among 37,073 tags. ‘EST’ indicates that this tag is found in the *Bombyx* EST database but function is unknown. ‘Multi hit’ indicates that the tag corresponds to two or more unassembled EST sequences. ‘No hit’ indicates that the tag does not correspond to any known gene.

Table 4.

Relative amount of some transcripts in stage 1-3 egg chambers relative to stage 2-3 egg chambers. a. An average value of triplicate or tetraplicate; b. the standard deviation of three or four separate determinations.

Fig. 1.

Schematic representation of the genomic structure of the most frequently expressed

transcript identified in stage 1-3 and 2-3 SAGE libraries. Three isoforms were identified, each with different polyadenylation sites. Open boxes, exons.

Fig. 2.

(a) RNA blot analysis of the most frequently expressed transcript. 3 µg of total RNA was loaded on each lane. An rRNA band stained with ethidium bromide is shown as a loading control. Left, RNA size markers. Lane 1, midgut of fifth instar larvae; lane 2, silk gland of fifth instar larvae; lane 3, female pheromone gland; lane 4, fat body four days after pupation; lane 5, ovary four days after pupation; lane 6, testis four days after pupation.

(b) RNA blot analysis of the most frequently expressed transcript. 3 µg of total RNA was loaded on each lane. An rRNA band stained with SYBR Gold (Molecular Probes) is shown as a loading control. Left, RNA size markers. Lane 1, whole ovariole four days after pupation; lane 2, isolated mature egg of female moth.

Table 1

	library 1		library 2	
	Number of tags	Number of unique tags	Number of tags	Number of unique tags
All tags	122060	40311	37073	16740
Single hit to the <i>B.mori</i> EST databases (proportion of SAGE tags matched)	28452 (23.3%)	1773 (4.4%)	9772 (26.4%)	1123 (6.7%)
Multi hit to the <i>B.mori</i> EST databases	14729 (12.1%)	284 (0.7%)	6037 (16.3%)	209 (1.2%)
No hit to the <i>B.mori</i> EST databases	78879 (64.6%)	38254 (94.9%)	21264 (57.3%)	15408 (92.0%)

Table 2

GO BP	Description	Long tag	Fold
autophagic cell death	eRF1	CCACCAACCAGGAAA	3.9
cation transport	stress-sensitive B	CCGCGGGCCGTGTCG	5.5
cell adhesion	Ecdysone-inducible gene L2	TATACTTAATATGGA	5.5
cell proliferation	14-3-3zeta	TTTTACCAATAAAC	7.5
cell surface receptor linked signal transduction	CG10602	ATGCACGTGTGTGCT	3.6
cell-cell signaling	Imaginal disc growth factor 2	TATCCGAAATTGTAA	3.6
cytoskeleton organization and biogenesis	Act79B	TACCCCGGTATCGCC	3.8
'de novo' protein folding	CG11267	TGATTCTAATTCTGT	8.8
defense response	Heat shock protein 68	TTGAAATTAAATAAA	13.4
	Heat shock protein cognate 3	GAGATTGATAACATC	2.4
electron transport	CG7430	TGCCGAAGCGCTACG	5.2
	ATPase coupling factor 6	ACAGCATTCCCATCT	9.1
establishment of cell polarity	abstrakt	CTGGGTGAGCCCGAT	3.0
glycolysis	Gapdh2	TACGTCGAGATTATC	2.2
iron ion homeostasis	Ferritin 1 heavy chain homologue	GCGCGTCAAATAGAA	2.6
mitochondrial electron transport, ubiquinol to cytochrome c	Rieske iron-sulfur protein	GACGAAGGCTTGCTA	3.9
phospholipid metabolism	Aps	TTATGCTTGGACTTC	3.0
protein biosynthesis	Ribosomal protein L7A	CTTAGCTCTAACAAA	2.7
	Ribosomal protein P2	TGTTTACTCTGCTCT	3.9
	Ribosomal protein L19	TAGAGTGAGTGGTAT	2.1
	Ribosomal protein L18A	TTGTGTACACCTGAA	8.2
	Ribosomal protein L37a	ACGAAGCTACAGCCT	5.8
	Ribosomal protein L14	ACGTTCTTACAACGT	2.7
	Ribosomal protein L31	AAGTCGTTTGTGTAA	4.6
	Elongation factor 1 α 8D	AAAGATCTGCGTGTG	3.9
	Ribosomal protein S15	AGATTACAATAAAGT	2.8
	Ribosomal protein S15A	AAGAAGCCAGAAGAA	2.4
protein transport	CG8860	TGTGTTAGAACATCC	2.4
proteolysis	Rpn1	GCCTCTGTTCAATAA	3.6
	Rpn11	CTTGACCTAAACTTG	3.0
proton transport	ATP synthase, subunit d	TTATATCATTTGAAA	3.1
	ATP synthase-beta	TGAATTGATGAGAAA	2.6
receptor mediated endocytosis	Dynamin associated protein 160	GTGCTGCAGGAGAGG	2.3
RNA interference	Rm62	TAGAATTTTATTAAT	7.3

Table 2 (continued)

GO BP	Description	Long tag	Fold
translational elongation	Elongation factor 1 beta	TAATATTTATAAGTG	5.3
translational initiation	Eukaryotic initiation factor 3 p40 subunit	GAACGTGTTGATGAA	2.2
tRNA metabolism	CG3045	TTCAACGCTACACGA	2.6

Table 3-1

Long tag	Description	Percentage
GCTGGCAGAGTCGCGA	EST	2.8%
CAGATCTGCTGTCAGG	Multi hit	1.0%
TCCGCATCCTGGCAGG	Multi hit	0.7%
ATCTGAGTTCAAACCG	Multi hit	0.6%
GGCTTCGGTCTCTTTG	Multi hit	0.6%
CGCGCGAAGCGTTTAT	No hit	0.6%
GGCAGGTGATCTACAC	Ribosomal protein S25	0.6%
GTGGTCTTGTGCACGG	Multi hit	0.4%
GCGGTGATGGAGGACC	Ribosomal protein S27A	0.4%
TCACGTCCACAGGCAG	Ribosomal protein S26	0.4%
TTGCCGTGGGACCAGC	Multi hit	0.4%
ATGGCGAGTGTGTTTC	Ribosomal protein P1	0.4%
TCTTGTCTCTGAGCTC	CG15697	0.4%
AGGTGATGCTAAACAT	Multi hit	0.4%
GCAACTCTGGCAGTGT	Multi hit	0.3%
CTTTCGTCGCTGGTAT	No hit	0.3%
GCCCCTCGTTTTGAAA	No hit	0.3%
GATGTAAAGTGCCCTG	CG10423	0.3%
GCATCTACCCTCTGCG	Ribosomal protein S3A	0.3%
ACGCTTACGAACTGCG	Ribosomal protein L22	0.3%
GACTCCGCTGAAGGCG	Multi hit	0.3%
CCTTTTGCACGGTACG	Ribosomal protein L14	0.3%
GACCTCTCGACGCTTG	Qm	0.3%
TCGTCTCCCTGCCACA	CG3203	0.3%
CCTGTACAGAAACAAA	Multi hit	0.3%
GATCTTGTAAGACAG	Multi hit	0.3%
GACTGTCCTGTACAGG	Multi hit	0.3%
CTTAAATGAGCGTCTG	No hit	0.3%
GTGTCTCTTCGAAGAA	Multi hit	0.3%
CCAACCAAGATCCTGC	Ribosomal protein S20	0.3%
GTTACATATCGCGTCA	lethal (2) essential for life	0.3%
TCAAGCACTCCAAGTG	Multi hit	0.3%
CCCGCTGAGTTTCTTG	No hit	0.3%
GGCTTTGGTCTCTTCG	Ribosomal protein P0	0.3%
ACATCATCGAAGTAGA	Ribosomal protein S17	0.3%
GTGTCAACCAGTTGAA	Ribosomal protein L15	0.3%

Table 3-1 (continued)

Long tag	Description	Percentage
CAAAAGAACTTAAAA	Multi hit	0.3%
GGTACACTCAAGAAGG	Multi hit	0.3%
TCCGTACATTTGTCGC	No hit	0.3%
CGGAAAATACTACAGA	Ribosomal protein L30	0.2%
GAGTACATCCACAGGA	Ribosomal protein L19	0.2%
ATGGCCGCACCATCCG	Ribosomal protein S4	0.2%
GGCGACGGTAACCACT	Multi hit	0.2%
GGTACGAAATGGAAGG	CG8415	0.2%
ACCAACAGATTGTTGG	CG3751	0.2%
CCGCGTGAAATCAAAG	Multi hit	0.2%
GAGACAACTCGCGGCT	No hit	0.2%
TGTAATGTTTAATTTT	No hit	0.2%
TCGAATCCTAGGCAGT	Ribosomal protein S18	0.2%
CCAGAATTTTGATCCG	No hit	0.2%

Table 3-2

Long tag	Description	Percentage
GCTGGCAGAGTCGCGA	EST	3.0%
ATCTGAGTTCAAACCG	Multi hit	2.8%
CAGATCTGCTGTCAGG	Multi hit	0.9%
GCATCTACCCTCTGCG	Ribosomal protein S3A	0.8%
GATGTTAAGTGCCCTG	CG10423	0.8%
TCCGCATCCTGGCAGG	Multi hit	0.7%
CCCGCTGAGTTTCTTG	No hit	0.7%
GGCTTCGGTCTCTTTG	Multi hit	0.7%
GACTGTCCTGTACAGG	Multi hit	0.6%
GGTACGAAATGGAAGG	CG8415	0.6%
TCTTGTCTCTGAGCTC	CG15697	0.6%
CTTAAATGAGCGTCTG	No hit	0.5%
GTGGTCTTGTGCACGG	Multi hit	0.5%
GACTCCGCTGAAGGCG	Multi hit	0.5%
GCAACTCTGGCAGTGT	Multi hit	0.5%
CTTTCGTCGCTGGTAT	No hit	0.5%
GGCAGGTGATCTACAC	Ribosomal protein S25	0.5%
TCGTCTCCCTGCCACA	CG3203	0.5%
CCTGTACAGAAACAAA	Multi hit	0.5%
CGCGCGAAGCGTTTAT	No hit	0.4%
ATGGCCGCACCATCCG	Ribosomal protein S4	0.4%
GTGTCAACCAGTTGAA	Ribosomal protein L15	0.4%
GACCTCTCGACGCTTG	Qm	0.4%
ACCAACAGATTGTTGG	CG3751	0.4%
TCAAGCACTCCAAGTG	Multi hit	0.4%
GTGTCTCTTCGAAGAA	Multi hit	0.4%
GGCTTTGGTCTCTTCG	Ribosomal protein P0	0.4%
CCAACCAAGATCCTGC	Ribosomal protein S20	0.3%
GGTACACTCAAGAAGG	Multi hit	0.3%
TCCGTACATTTGTCGC	No hit	0.3%
ATGGCGAGTGTGTTTC	Ribosomal protein P1	0.3%
TTGCCGTGGGACCAGC	Multi hit	0.3%
CCTTTTGCACGGTACG	Ribosomal protein L14	0.3%
TCACGTCCACAGGCAG	Ribosomal protein S26	0.3%
TGGCTGGGGTTCGCGTG	No hit	0.3%
ACATCATCGAAGTAGA	Ribosomal protein S17	0.3%

Table 3-2 (continued)

Long tag	Description	Percentage
GCCCCCTCGTTTTGAAA	No hit	0.3%
GCGGTGATGGAGGACC	Ribosomal protein S27A	0.3%
GACAAGTACCATCCTG	Multi hit	0.3%
AGGTGATGCTAAACAT	Multi hit	0.3%
CTTCAGTACAGCTAGT	oho23B	0.3%
GTCTACTGATGGACAA	No hit	0.3%
AAGCCATCCGTGCGGA	Multi hit	0.3%
TTGCTTGCTCAAGGTA	No hit	0.3%
CCAGAATTTTGATCCG	Multi hit	0.3%
CGGAAAATACTACAGA	Ribosomal protein L30	0.3%
ACGCTTACGAACTGCG	Ribosomal protein L22	0.2%
GAAGACACCATCAAAA	Multi hit	0.2%
GAGTACATCCACAGGA	Ribosomal protein L19	0.2%
AAAAAGCAAGGCCAG	Multi hit	0.2%

Table 4

stage	Mean ^a ±SD ^b		
	ATPase coupling factor 6	ATP synthase, subunit d	The most abundant RNA in stage 1-3 and stage 2-3 egg chambers
Stage 1-3	1.6±0.15	1.3±0.15	0.96±0.39
Stage 2-3	1±0.12	1±0.14	1±0.15

Figure 1

Figure 2

(a)

(b)