African Journal of Biotechnology Vol. 11(49), pp. 10994-11012, 19 June, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.4121
ISSN 1684–5315 ©2012 Academic Journals

Full Length Research Paper

Long SAGE analysis of genes differentially expressed in the midgut and silk gland between the sexes of the silkworm *Bombyx mori*

Liping Gan[1,2], Ying Wang[1], Jian Xi[1], Yanshan Niu[1,] Hongyou Qin[3], Yanghu Sima[1] and Shiqing Xu[1]*

[1]National Engineering Laboratory for Modern Silk, Department of Applied Biology, Medical College, Soochow University, Suzhou, 215123, China.
[2]Biology Department, Chongqing Three Gorges University, Chongqing, 404000, China.
[3]National Engineering Center for Biochips at Shanghai, 151 Libing Road, Pudong, Shanghai 201203, China.

Accepted 21 May, 2012

There are great differences in silk production efficiency and quality between the male and female domestic silkworm (*Bombyx mori*). Many genes act together but are differentially expressed between the sexes during silk biosynthesis. Two long serial analyses of gene expression (SAGE) libraries were constructed from the midguts and silk glands of both males and females of a sex-limited strain using 5th instar larvae, yielding in total 96,713 and 98,126 SAGE tags, respectively. Among these tags, 202 were analyzed ($p < 0.05$ and at least a 2.0-fold change between sexes). Overall, 69 genes were then annotated in detail and 15 tags were annotated with expressed sequence tags (ESTs) based only on the NCBI, SilkDB and long-SAGE libraries. Of these genes, only three could be ascribed to sexual disparity as described by microarray-based expression resources of day three of 5th instar in Dazao *B. mori* microarray database (BmMDB). The other 66 genes were considered the SAGE-extracted genes that are differentially expressed between the sexes of the whole 5th instar larvae. Among the 66 genes and 15 tags, genes (and gene families) sex-specific storage-protein 1 gene (*Sp*1), low molecular mass 30 kDa lipoprotein 19G1 gene *(Lp-c*19), serine proteinase family [serine protease precursor gene *(Spp)* and chymotrypsin-like serine protease gene *(Ctlp*)], Serpins (*Spi*1 *and Spi*2) and *Ser*1, and the tag 1161 (annotated EST No. BY926524), which are involved in protein digestion in the midgut, synthesis of silk and inhibition of protein disintegration in silk gland, were verified by the remarkable disparities in gene expression between the sexes. The established SAGE library would contribute to the further identification of genes related to sexual disparity in silk protein production efficiency.

Key words: *Bombyx mori,* sexes, silk protein production efficiency, long-SAGE, differentially expressed genes.

INTRODUCTION

The domestic silkworm (*Bombyx mori* L.) is an important economic insect species and also an ideal organism for research (Mita et al., 2004; Sakudoh et al., 2010). Based on the fact that *B. mori* possesses more sex-specific proteins than *Drosophila*, it is an advantageous model for sex-specific transcription (Fujii and Shimada, 2007). In *B. mori*, investigation of common regulatory mechanisms of males are superior to females in integrated economic traits such as silk protein synthesis efficiency and silk quality (Zha et al., 2009), however, the molecular mechanisms involved in the determination of these properties are still not well understood (Li et al., 2009). Both the technology of female egg balanced-lethal system for male silkworm breeding and chromosomal trans-location mutants are widely used to improve the cocoon

*Corresponding author: E-mail: szsqxu@suda.edu.cn. Tel: +86 512 65880185. Fax: +86 512 65880185.

production efficiency, even if there are weak male larvae and female eggs drain. Therefore, attention to genes differentially expressed between the *B. mori* sexes have biological and silkworm-cocoon production significance.

The 5th instar is a significant stage for rapid development of *B. mori* larvae. During this stage, most of the larval developmental processes are completed and a mass of silk proteins are initiated to be rapidly synthesized (Xia et al., 2007). The silk gland and midgut are the main sites where pigment is transferred to the cocoon from the digestive tract; they are also the target organs of multi-gene regulation of the protein synthesis process for both sexes (Jin et al., 2004).

Thus, studies of the silk gland and midgut in this stage will help elucidate the regulatory mechanism of the complex synthesis process of silk proteins and help explain the disparity in protein synthesis efficiency between males and females. Serial analysis of gene expression (SAGE) is the most versatile method to study genome-wide gene expression. The frequency of each tag in the SAGE library provides an accurate estimate of the abundance of its corresponding mRNA (Poole et al., 2008; Bala et al., 2005).

Application of SAGE technology has led to the construction of SAGE libraries at different developmental stages and libraries that compare differentially genes following various treatments of *B. mori* both types have yielded many valuable findings (Huang et al., 2005; Funaguma et al., 2007; Zhang et al., 2007). A yellow cocoon sex-limited strain of Ysh and a W-chromosome mutation of carotenoid protein gene (*Cbp*) were selected to establish the long-SAGE libraries of the silk gland and midguts of the whole 5th instar larva stage for males and females, respectively. Disparity expression genes in silk gland and midgut that were differentially characterized between the sexes were screened in accordance with both the microarray expression profile on *B. mori* microarray database (BmMDB) and the unigenes notes and were analyzed for further research. The value of extracting and researching genes related to the differential regulation of protein synthesis between males and females was demonstrated by the sexual disparity of spatio-expression profiles and their relation to protein synthesis.

MATERIALS AND METHODS

Organisms

Ysh, a yellow cocoon color sex-limited strain of the silkworm *B. mori*, was provided by the Development and Genetic Molecular laboratory of the Medical College of Soochow University (Suzhou，China). Silkworm larvae were reared with fresh mulberry leaf at 25°C, a relative humidity of 70 to 80% and under a 12 h light/12 h dark cycle. Under these conditions, the two sexes of larvae were able to be accurately distinguished during the 4th and 5th instar stages by abdominal color. Most of the larvae reached the wandering stage in the late photophase on day seven at the 5th instar stage and spun their cocoons three days later at the pupa stage.

RNA isolation

All ribonucleic acid (RNAs) were prepared from two males and two females for each tissue. Overall, 10 tissues or organs, including epidermis, head, anterior silk gland (AMG), median silk gland (MSG), posterior silk gland (MSG), hemocyte, midgut (MG), malpighian tubule (MT), ovary and fat body (FB), were sampled on day three of the 5th instar larva (5L-3). Meanwhile, to find the exact time points that correspond to different expressions between the sexes, 13 time points from silk gland and midgut were taken from day 1 of the 4th instar larva (4L-1), which is the stage yellow color and white color larval can be initially discerned, today two of the wandering stage (S-2). All samples were stored immediately in liquid nitrogen. Total RNA was isolated using TRIzol reagent (Life Technologies) and purified using an Oligotex® Messenger ribonucleic acid (mRNA) Midi kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. All sample concentrations were estimated by spectrophotometer.

SAGE library construction

The SAGE procedure was performed on the mRNA from the silk glands and midguts of silkworms. The SAGE library of each gender contained seven silk gland samples and seven midgut samples from same-sex larvae of the Ysh strain, from day one of the 5th instar larvae to day one of the wandering stage, with each sample taken once a day. The sample of every library was mixed equally yielding 2 µg per original sample. The sample quality of two mixed libraries was confirmed by spectrophotometer and again by electrophoresis analyses.

The two SAGE libraries were constructed using the I-SAGE™ long kit (Invitrogen) with the method described previously (Huang et al., 2005). Purified mRNA was bound to oligo (dT) magnetic beads. Superscript™ II reverse transcriptase was used to synthesize double-stranded complementary deoxyribonucleic acid (cDNA) from the template mRNA-bead complexes. The cDNAs were digested with the anchoring enzyme *Nla*III, which recognizes a four-base sequence (CATG). The ligation products were digested with the tagging enzyme *Mme1*, which recognizes a sequence in the linkers and cleaves 17 bases downstream of this sequence. The concatemers with 454-SAGE adapters were size fractionated and those of 300 to 800 bp were isolated and linked with M270 magnetic beads. Single-stranded deoxyribonucleic acid (DNA) was isolated using a Qiagen kit and sequenced using a Roche 454 GS-FLX DNA Sequencer (Roche, Switzerland).

Gene identification (GLGI)

The conditions for identification of significant differences between the two libraries were set at a p-value < 0.05 and at least a twofold change. According to these criteria, different expression tags were selected and used to compare different expression levels. Some genes with no gene annotation were selected to generate longer cDNA fragments from SAGE tags for gene identification (GLGI) (Chen et al., 2002) according to the SAGE protocol. A sufficient number of templates for GLGI analysis were generated using low-cycle polymerase chain reaction (PCR). We used the SAGE primers 5′-GGATTTGCTGGTGCAGTACA-3′ as the sense primer and 5′-ACTATCTAGAGCGGCCGCTT-3′, which was located at the 3′ end of all cDNAs generated from the anchored oligo (dT) primers, as the antisense primer. The sense primer used for GLGI amplification contained the 17-base SAGE tag sequence, giving a total of 21 bases for each primer (5'- CAT GXX XXX XXX XXXX XXX XX-3'). The PCR-amplified products were purified and then cloned into the pMD18-T vector (Takara). The products were

Table 1. Tags distribution in the two SAGE libraries.

Abundance class (tag copy)	Number of tags detected in the SAGE		Number of unique SAGE tags	
	Female	Male	Female	Male
0			13049	10855
1	12391	13935	12391	13935
2 - 5	12509	14215	4490	5111
6 - 10	7398	7505	979	999
11 - 20	7941	8086	552	568
21 - 100	20252	19925	488	491
>100	36222	34460	150	140
Total	96713	98126	32099	32099

sequenced and provided full-length cDNA corresponding to the SAGE tags (Gan et al., 2010).

SAGE data analysis

SAGE 2002 was used to quantify the abundance of each tag. The two long-SAGE libraries were deposited in the SAGE map database on GenBank (Accession numbers soon to be available). Based on the NCBI EST database, we constructed a reference SAGE-tag database that represented the known silkworm ESTs. The EST sequences used to identify SAGE tags have a poly (A) tail (minimum of eight A nucleotides) (Lash et al., 2000) and the last CATG cleavage site in the sequence. All 17-base SAGE tags 3′-adjacent to the 3′-most NlaIII site (CATG) extracted from the reference sequences were used to build the reference SAGE database. The reference sequences allowed matching of experimental SAGE tags. The software we developed was used to match experimental SAGE tags to the reference SAGE database, which allowed the identification of genes that potentially corresponded to each SAGE tag. If the sequence had no corresponding gene as determined by BLAST analysis of the silkworm database of NCBI but could be confirmed in the WGS database of silkworm it was termed a ‘novel sequence’.

Reverse transcriptase (RT)-PCR

According to the manufacturer’s protocol (Invitrogen), the first-strand cDNA synthesis was conducted with prime of oligo-$(dT)_{18}$, and PCR was performed using the appropriate following gene (or EST)-specific primers. Expression levels were normalized to the expression of silkworm *Actin 3* genes. PCR conditions were as follows: 94°C for 3 min and 30 cycles of 94°C for 30 s, 54 to 57°C for 40 s, 72°C for 40 s and a last cycle of 72 for 10 min. The products were detected by agarose gel electrophoresis and analyzed with gel image system (GIS) software package (Tanon Company, China).

Annotation of gene function

To understand gene function, we performed functional annotation of the candidate genes systematically by using bioinformatics tools. First, we annotated with gene ontologies (GO) categories based on the three levels of biological process, molecular function, and cellular component. The candidate gene sequences were assigned to the nr database in GenBank and GO annotations (http://www.geneontology.org/) were assigned. Then, with Web Gene Ontology Annotation Plot (WEGO), the GO annotation results were plotted. The other annotation analysis was conducted using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov/home.jsp), which is an online tool that provided the gene symbols. Differentially expressed genes with significant enrichment were then compared.

RESULTS AND ANALYSIS

Construction results of long-SAGE

Total SAGE-tag copy numbers of 96,713 and 98,126 were available from female and male silkworm libraries, respectively. In total, 32,099 unique tags were detected, among which 21,244 and 19,050 tags were found in females and males, respectively. Additionally, the alternative number of tags was 13,049 in females and 10,855 in males, with a mutual number of 8195 in both sexes. Most of the tags were found with one or two copies, but there were a few tags with frequencies higher than 100 (Table 1). Nearly 8.9% of the 2,854 tags were matched to the public silkworm ESTs.

Analysis of highly expressed tags

Of the 100 highly expressed tags, 95 tags had gene annotations and five tags had only EST evidence. Of the annotated genes, more than 47 were ribosomal protein genes. There were also some genes with very high expression levels, such as the six cytochrome genes, five NADH oxidoreductase genes, five protease genes and two lipase genes. These genes play imperative roles during transportation and synthesis of silk (Tabunoki et al., 2004; Mahendran et al., 2006).The GO functional categories for these highly expressed genes are shown in Figure 1. In total, 88 genes had GO functional classifications. Nearly half of the genes were shown to be involved in cell and cell part, in cellular component and metabolic processes, and in biological processes. Other enriched GO function terms were intracellular, macromolecular complex, non-membrane-bounded organelle, ribonucleoprotein complex and structural molecule, and

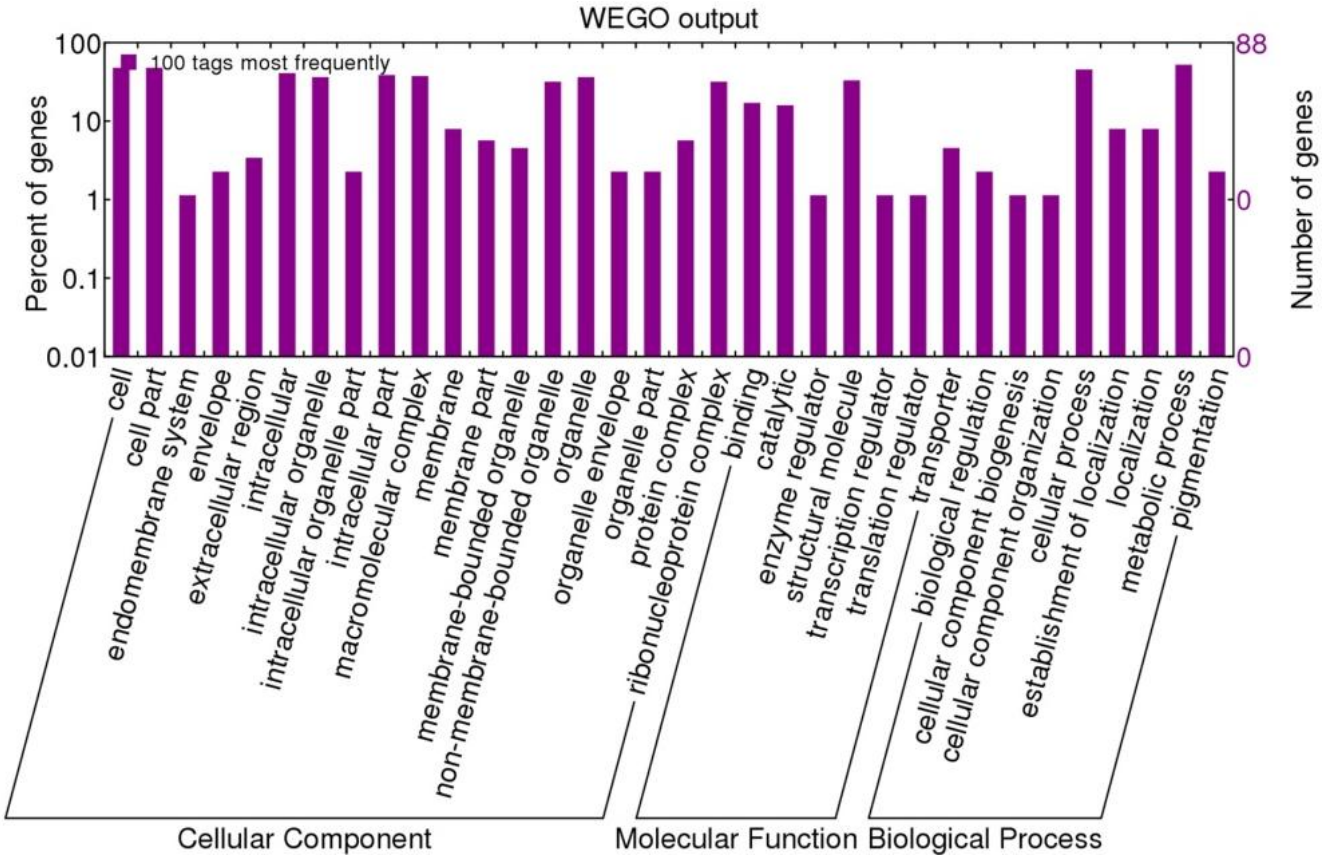

Figure 1. Gene ontology (GO) annotation of 100 highly expressed transcripts in SAGE libraries. The identified genes were classified into cellular component, molecular function and biological process by WEGO according to the GO terms. The number of genes is the number of times the GO term is used to annotate genes in the cluster. The left-hand shows its proportion in total genes with GO terms

cellular processes.

Annotation of differentially expressed tags

Based on the conditions of $P < 0.05$ and at least a twofold change between the sexes, 202 tags were selected from 32,099 unique tags in the two SAGEs as differentially expressed tags. Among these tags, 94 tags were up-regulated and 108 tags were down-regulated in females as compared to males. Overall, 76 tags were annotated to 69 genes in detail and 15 tags were annotated with ESTs based only on the NCBI, SilkDB and the long-SAGE libraries, while 111 tags (55%) were left unmatched (Supplementary Table 1). Table 2 lists the annotation results for 48 tags that were differentially expressed (fold ≥ 10) between both the SAGE libraries. Among them, 18 tags had gene annotations, two tags were gained through EST evidence, while 28 tags (58%) were unmatched SAGE tags. The results indicate that the constructed SAGE libraries were of great value in extracting novel genes related to protein synthesis.

Function analysis of the annotated genes

Based on the transcripts per million (TPM) value of each tag, the 76 tags above were clustered (Figure 2A) by Matlab R2009a software, in which up-regulation in females and males were clustered (parts (a) and (b), respectively). A similar expression pattern based on BmMDB is shown in Figure 2B, indicating corresponding gene probes of these tags from Supplementary Table 1. Those genes with similar expression patterns were clustered on the same platforms, showing tissue-specificity in two platforms. BmMDB (Xia et al., 2007) provided gene expression microarray-based data on day three of 5^{th} instar of Dazao (females and males separately). Application value of the SAGE libraries can be assessed by the ratio of genes that are differentially expressed between the sexes from SAGE and BmMDB. In total, 76 tags were annotated to 69 genes, of which 59 genes can be probed by 64 valid probes in BmMDB. The results show that 56 genes annotated with 63 tags (94.9%) had inconsistent expression patterns (Figure 2).

Only three genes, annotated with tags of sex-specific

Table 2. Annotation of 48 tags that were differentially expressed (Fold≥10) between both SAGE libraries.

Tag sequence	Female	Male	Fold of TPM	Annotation	Class
TGTGACTGTGTGATTGT	0	49	>49	Osiris 19	A
ATACTTTGAATGGGTTC	0	47	>47	hypothetical protein LOC692620	A
AAAAGTTGAACCGTCAT	0	34	>34	TPA: putative cuticle protein	A
ATCTGAGTTCAAACCGC	0	25	>25	No hits	C
GAACCGAGCGAGATAAA	0	24	>24	DY231379	B
AACGATTAATCAAATAA	0	22	>22	lipase-1	A
TGTATATCAAGTTAGAA	0	18	>18	TPA: putative cuticle protein	A
ATAATTGTAATTAACGT	0	17	>17	No hits	C
TTAATTTATTTTGTGAG	0	15	>15	Bombyx mori phosphate transport protein	A
TGCTTCTATGTTTATGA	0	14	>14	Osiris 9	A
CAACCAATATACACATT	0	13	>13	No hits	C
GAGGTTCGGGAGTCAGC	0	13	>13	cuticular protein tweedle motif 3	A
TTCAGGAAGTGCTTGGA	0	12	>12	No hits	C
ACTCCTGATGAGCTTGT	0	11	>11	No hits	C
TTAGGATCCTTGGAAAA	0	11	>11	non-coding RNA, ovarian small RNA-31474	A
AATCAATATGTGCTAAC	0	10	>10	No hits	C
AGAGGTGTAGCATAACT	0	10	>10	No hits	C
CTTAATTTGTACATACA	0	10	>10	No hits	C
GAGGTGGCGGATATGGT	0	10	>10	ribosome-associated protein P40	A
TCCTTAATTATATATGA	1	24	24	No hits	C
AATCGGAGCCCGACCTC	1	16	16	No hits	C
ATGGCGAGTGTGTTTTT	2	31	16	ribosomal protein P2	A
TTATATCTGCAAAAACG	1	15	15	No hits	C
CAAAAGAAACTTAAAAA	1	13	13	No hits	C
TGTCATCTATAGGTCTT	1	13	13	No hits	C
ATCTGAGTTCAAACCAG	1	12	12	No hits	C
GTAAATATTTTCAAGAT	1	12	12	No hits	C
AGTTTGAAGAAATAGAG	1	11	11	No hits	C
ATCTGAGTTCAAACCGA	3	31	10	No hits	C
TCCGTACATTTGTCGCT	2	20	20	No hits	C
AACCCTGTAGAGCATCT	1	10	10	No hits	C
CTATTCCTGTTCTACGA	1	10	10	No hits	C
TACCTATTATTTTGAGG	1	10	10	Lysophospholipase (LOC692950)	A
TGGGAATCTTTGAGGTT	1	10	10	No hits	C
CAAAGAATAGCTAAGCA	10	1	10	No hits	C
TAATCTTTCCTGTTCTT	10	1	10	Genomic DNA, chromosome 14	A
TTTCTCAACGGTCTACA	21	2	11	Beta-1,3-glucan recognition protein 4	A
CGTAATTGTTGCAGCAA	13	1	13	No hits	C
TCACTGGATCTGACGTT	13	1	13	No hits	C
TTGGCGTTTAAGAAATC	15	1	15	BB988365	B
ACAGTGTTGATATATCT	25	1	25	Proteasome 26S non-ATPase subunit 4	A
CCGAAGTTGAATTTAAT	32	1	32	Beta-tubulin (Tub2)	A
AATATAAAGTATTCAGA	15	0	>15	Histone H2B-like protein	A
TGTTTGGTAACAAATCA	13	0	>13	Chorion locus sequence encompassing gene	A
TTAGATTTGAATTAGAA	12	0	>12	No hits	C
GGTCTGGAATGTTCCGA	10	0	>10	No hits	C
TCGTATAGGCCGCAATT	10	0	>10	No hits	C
TTAATACATTATAATGT	10	0	>10	No hits	C

A, B and C indicate tags with annotation genes, EST evidence and no hits, respectively.

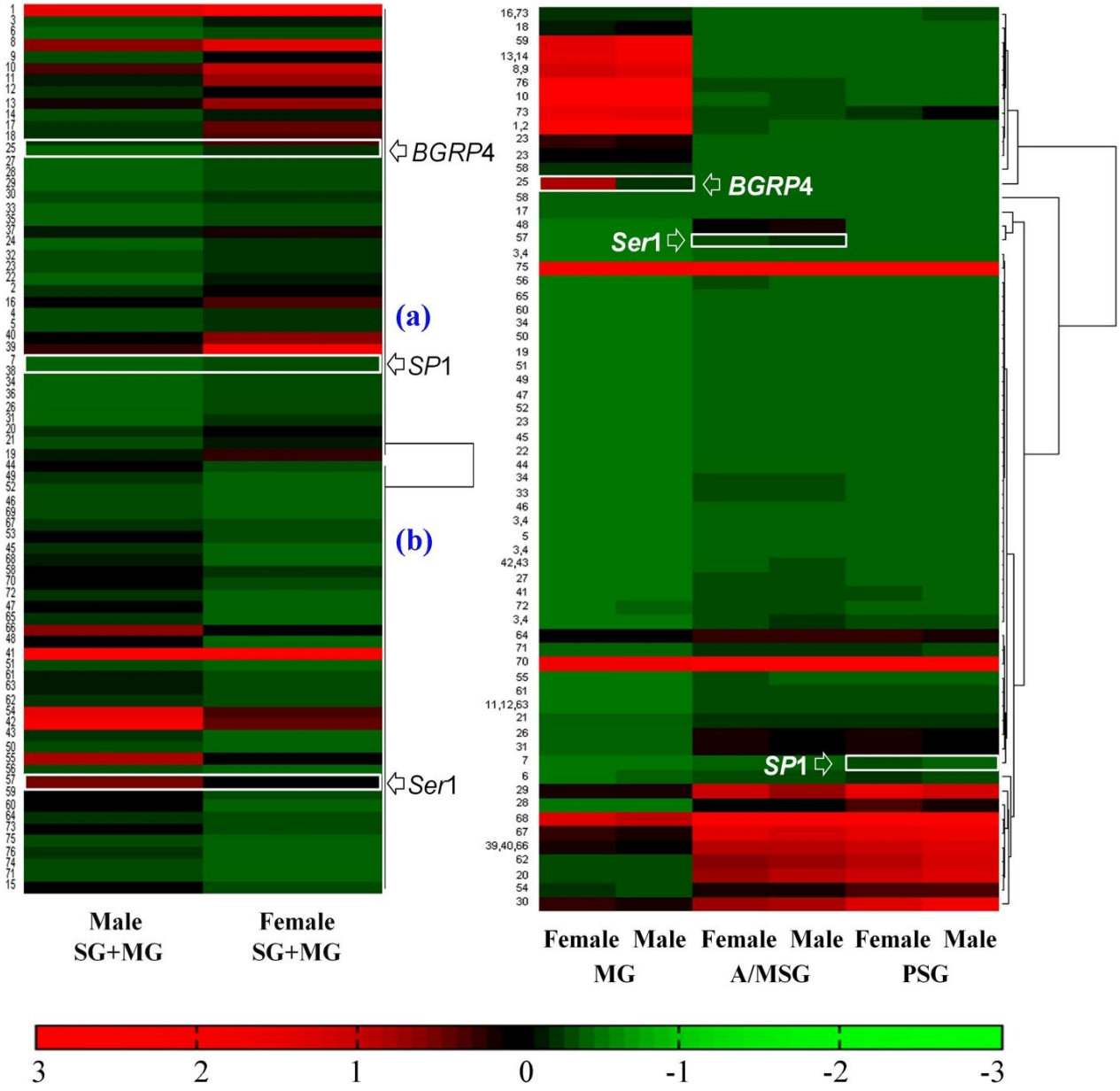

Figure 2. Expression profiles of 76 tags based on TPM of SAGE data (A) and 64 probes (corresponding to 59 genes) based on intensity value units on BmMDB (B). The series numbers in both A and B were mapped to the first line of Supplementary Table 1. The red, black and green colors indicate tags or probes that were expressed at high, low and zero levels, respectively. Tags and probes were aligned horizontally, and libraries are shown vertically. White rectangles indicate the similar expression tendency in both SAGE and BmMDB

storage-protein 1 (*SP*1) (Tag No. 7 in Supplementary Table 1) and *beta*-1, 3-glucan recognition protein 4 (*BGRP4*) (Tag No.25), were shown to be up-regulated in females, and *sericin* 1 (*Ser*1) (Tag No. 57) was up-regulated in males by both SAGE and silkDB analyses (Figure 2). The results suggest that more genes differentially expressed between the sexes could be extracted from the SAGE tags during the whole 5th instar larva stage than that from BmMDB on day three of the 5th instar. Furthermore, a functional classification of the 55 differentially expressed genes was performed from those gene symbols given by the DAVID database (Table 3). Function annotation results indicated that nine genes (*LOC692357*, *LOC732998*, *LOC732967*, *HEL*, *LOC692513*, *LOC692849*, *CHI*, *LOC692977*, and *LOC100134929*) annotated proteolysis, followed by two genes encoding lipase and two genes encoding superoxide metabolic process, suggesting that the

Table 3. Functional annotations of 55 differentially expressed genes.

Category term	Count	%	P Value	Genes	Fold Enrichment
Extracellular space	4	0.1129	0.0019	LP-C23, LP-C6, SP1, LP-C19	13.31
Cytoplasm	6	0.1693	0.0057	RPS3AE, RPSA, TCTP, LOC732998, SOD, EIF3I	4.80
Extracellular region part	4	0.1129	0.0060	LP-C23, LP-C6, SP1, LP-C19	9.21
Protein of unknown function DUF1676	3	0.0847	0.0107	OSI9, OSI18, OSI19	17.42
Lepidopteran low molecular weight lipoprotein	3	0.0847	0.0244	LP-C23, LP-C6, LP-C19	11.61
PIRSF005527: microvitellogenin	3	0.0847	0.0292	LP-C23, LP-C6, LP-C19	10.21
Peptidase activity	6	0.1693	0.0333	LOC692357, LOC100134929, LOC732998, LOC732967, LOC692977, HEL	3.11
Lipoprotein	3	0.0847	0.0379	LP-C23, LP-C6, LP-C19	9.11
Hydrolase	7	0.1975	0.0495	LOC692849, LOC692357, LOC732998, CHI, LOC732967, LOC692513, HEL	2.47
Lipase	2	0.0564	0.0553	LOC692849, LOC692513	34.84
Lipase, N-terminal	2	0.0564	0.0553	LOC692849, LOC692513	34.84
Superoxide metabolic process	2	0.0564	0.0611	SOD	30.42
Proteolysis	5	0.1411	0.0648	LOC692357, LOC732998, LOC732967, LOC692977, HEL	3.00
Superoxide dismutase activity	2	0.0564	0.0767	SOD	24.55
Oxidoreductase activity, acting on superoxide radicals as acceptor	2	0.0564	0.0767	SOD	24.55
Oxygen and reactive oxygen species metabolic process	2	0.0564	0.0806	SOD	22.82
Peptidase activity, acting on L-amino acid peptides	5	0.1411	0.0849	LOC692357, LOC100134929, LOC732998, LOC732967, HEL	2.83

Column 1, Annotation term; column 2, gene count belonging to the term; column 3, percentage of genes in sum genes taking part in analysis; column 4, gene list belong to the term; last column, fold value. Red rectangle indicate enriched category terms with P Value<0.05.

differential expression was a key issue in disparities observed with respect to metabolic regulation of alimentation and silk protein synthesis between the sexes of larvae.

In addition, the three genes *OSI9*, *OSI18* and *OSI19* encoded a protein of unknown function, DUF1676, indicating that the annotated genes from SAGE tags had further value in investigating the mechanism of the difference of silk protein synthesis efficiency between males and females. Interestingly, three genes encoding low-molecular-weight lipoprotein shared the same annotation of microvitellogenin, which had been previously reported to show expression differences between the gonad of male and female silkworms (Yano et al., 1994).

Verification on differentially expressed tags

Overall, five annotated genes and one EST evidence tag was further selected to verify the results of the SAGE libraries, and the results are as follows (Primers are listed in Supplementary Table 2).

Annotation genes of different expression between the sexes

In the SAGE results of the whole 5th instar larva silk gland and midgut, a 3.75-fold sexual disparity tag (ATGAACTGATCACACGA) was noted for the gene *Sp*1. The expression profile on BmMDB indicated that *Sp*1 was mainly translated in the fat body, gonad and so forth of the females, but the sex-specific expression in the silk gland was not that noticeable (histogram in Figure 3). The results of saturated RT-PCR demonstrated that no obvious disparity in the midgut was observed at mRNA transcription level, but the mRNA production was mostly detected in the silk gland of females rather than males (electropherogram in Figure 3). Hence, the SAGE results provided more accurate information of sex-related differentially expressed genes in the silk gland and midgut.

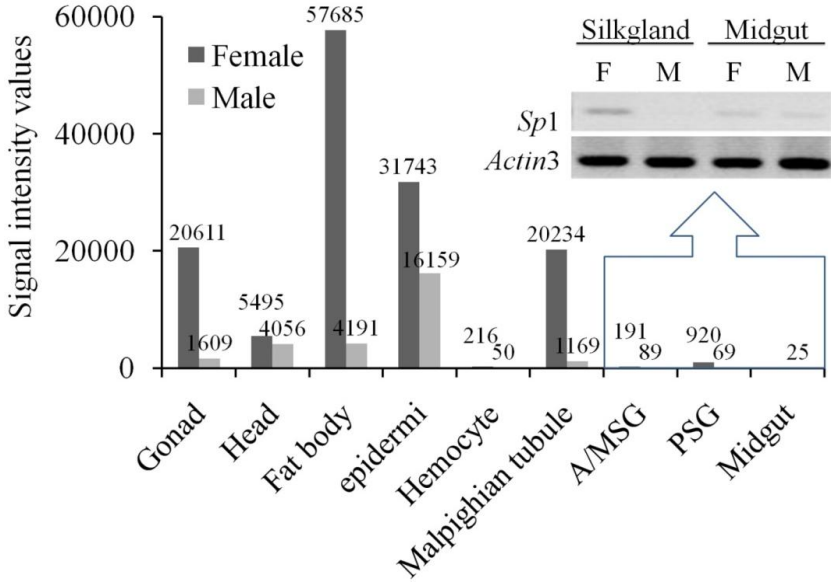

Figure 3. Tissue expression profiles of *Sp*1 gene between the sexes. The histogram shows the results measured by BmMDB in the Dazao strain using day three 5th instar larvae, and the electropherogram shows the RT-PCR results in the Ysh strain using whole 5th instar larvae.

The 30K is a type of protein that exists most prevalently in hemocytes of the 5th instar larvae; it is also both the main component of ovary protein and an important nutritional resource in embryonic development (Kerr, et al., 1969; Kim et al., 2003). In the current study, four tags (GCCTGGGGATACAATGG, GCCTGGGGATACAACGG, GTACATCGTCCCCTACT and AATAATGTAATAAACGA) corresponding to the three genes of low-molecular-weight 30-kDa lipoprotein family were differentially expressed with folds of 3.0, 3.2, 4.7 and 4.3 between the two libraries. One of the three genes, low molecular mass 30 kDa lipoprotein 19G1 gene (*Lp-c*19), was examined for spatio-temporal expression during 5th instar larvae.

The RT-PCR result indicated that the gene was expressed in several tissues, higher in midgut and epidermis (Figures 4A and B) in both sex on day three of 5th instar larvae. Therefore, the midgut was selected to present the developmental expression profile. The results indicate that the expression began from day one of the 5th instar larva stage to day one of the wandering stage in female, but it was not consistently expressed during this stage in male (Figures 4C and D). Gene expression results were accorded to the difference in SAGE tags frequency (Supplementary Table 1) and that this kind of protein is coincided to the fact that a large amount of this protein exists in females.

Serine protease genes

Serine protease (SP) is a key enzyme in protein digestion and absorption. Overall, 143 putative genes encoding SPs had been identified, including two representative genes of serine protease precursor gene (*Spp*) and chymotrypsin-like serine protease gene (*Ctlp*) (Zhao, et al., 2010). In the current study, the *Spp* gene (Tag ATTTTTTTGTAAATAAA) and the *Ctlp* gene (Tag AATTTCTTCAATCAACA and Tag CCAGAAGTA-AATGTAAG) were highly expressed in the female library (2.3-fold and 2.1 to 4.4-fold, respectively). Investigations in both sex on day 3 of 5th instar larvae indicated that the two genes were only expressed overwhelmingly in the midgut (Figures 5A and B). The expression profile in midgut of *Spp* gene indicated that it was highly expressed in females than in males in a later period of 5th instar and wandering stage (from 5L-5 to S-2), presuming that the female consumed more protein nutrition in the ovum making (Figures 5C and D).

Kazal-type serine proteinase inhibitor

Kazal-type serine proteinase inhibitor (serpins) is a kind of enzyme that participates in inhibiting the activity of serine proteinase. Overall, 34 putative genes encoding serpins had been identified (Zou et al., 2009), including two representative genes of Kazal-type serine proteinase inhibitor 1 (*Spi*1) (Kurioka et al., 1999) and serine proteinase inhibitor 2 (*Spi2*). In the SAGE libraries, *Spi*1 (Tag TCCTTAATTATATATGG) and *Spi*2 (Tag TGAATTTAGGGTTGCCA) were expressed higher in males than in females (2.3- to 5.3-fold and 3.8-fold). The

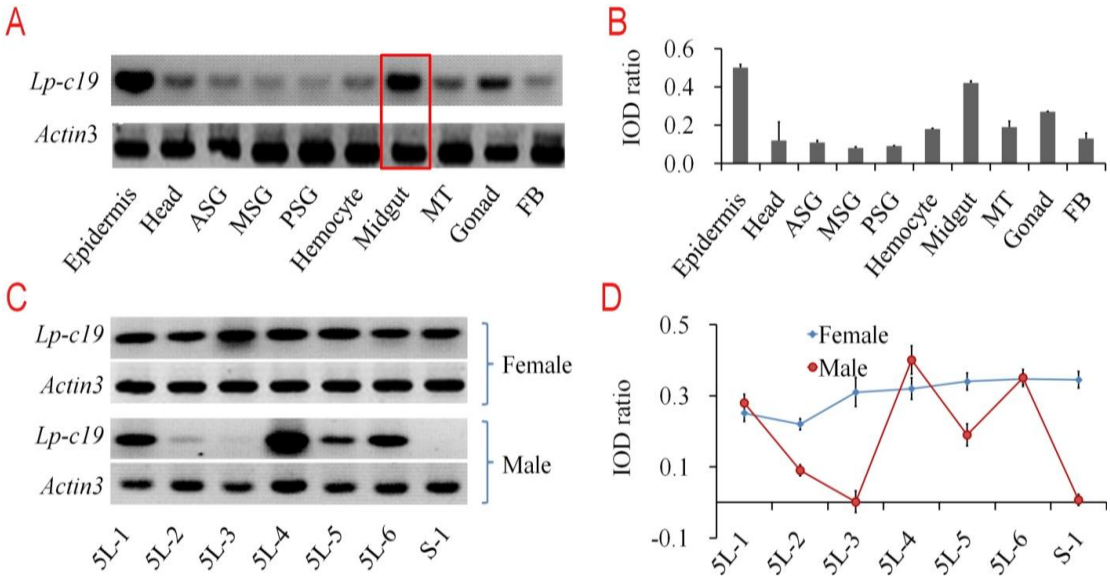

Figure 4. Spatio-temporal expression profile of *Lp-c19* gene was measured by RT-PCR between the sexes in larvae of the Ysh strain. Tissues expression profile of electropherogram (A) and diagram based on integrated optical density (IOD) from electropherogram (B) of *Lp-c*19 for in both sex on day 3 of 5th instar larvae. Stage expression profile of electropherogram (C) and diagram based on integrated optical density (IOD) from electropherogram (D) of *Lp-c*19. *Actin* 3 was used as a control.

results of RT-PCR indicated that *Spi*1 was highly expressed in the midgut and silk gland, while *Spi*2 was highly expressed in the MSG, epidermis and head in both sex on day three of 5th instar larvae (Figures 6A and B).

Further investigation studied the expression change in the silk gland and the development stages. Between the sexes, the *Spi*1 gene was only expressed from day three of the 5th instar larvae to the spinning time in females but was expressed longer in males and at a higher expression level (Figures 6C and D); these findings imply that the inhibition activity of protein degradation in the silk gland was stronger in males than females during the middle and later period of the 5th instar larvae.

Sericin 1

Sericin genes, specifically expressed in the MSG of silkworm larvae, are a family of genes that include five identified genes (*Ser1* to *Ser5*). *Ser*1 is only expressed in MSG cells with transcriptions of a 2.8 kb mRNA and a type of mRNA precursor; three of the precursor's cleaved products would conduct the translation of sericin proteins S1, S2, and S3. The *Ser*1 gene could be annotated by tag (ATGGAAGCGTAACATCC) in the SAGE libraries, in which the copy number in males was twice that of females. The RT-PCR results indicated that the expression level of *Ser*1 gene in the silk gland of males was remarkably higher than in females (electropherogram in Figure 7). The expression profile on BmMDB showed that *Ser*1 gene was highly expressed only in the mixed samples of ASG and MSG and that the mRNA level was higher in males (histogram in Figure 7). The results above demonstrated that the synthesis efficiency of the silk protein in males was higher than that of females.

Novel tag

A SAGE tag of GCTTCCGCCGTGCCTGC (Tag.1161) was annotated to EST BY926524 of the silkworm database at NCBI. Copy numbers of the tag in the female and male SAGE libraries were 3 and 18, respectively. Expression abundance of the EST in 11 tissues of Ysh strain was investigated by RT-PCR. The results show that this EST was only expressed in midgut on day three of 5th instar larvae (Figures 8A and B), and more highly expressed in males from day two to five of the 5th instar, during which period there was no expression in females (Figures 8C and D). It is indicated that present SAGE tags from midgut and silk gland hold the potential value for excavating the differentially expressed genes between the sexes.

DISCUSSION

Contribution of the established SAGE libraries to the extraction of genes related to the sexual disparity in silk protein production efficiency

The purpose of raising silkworms is to produce a cocoon,

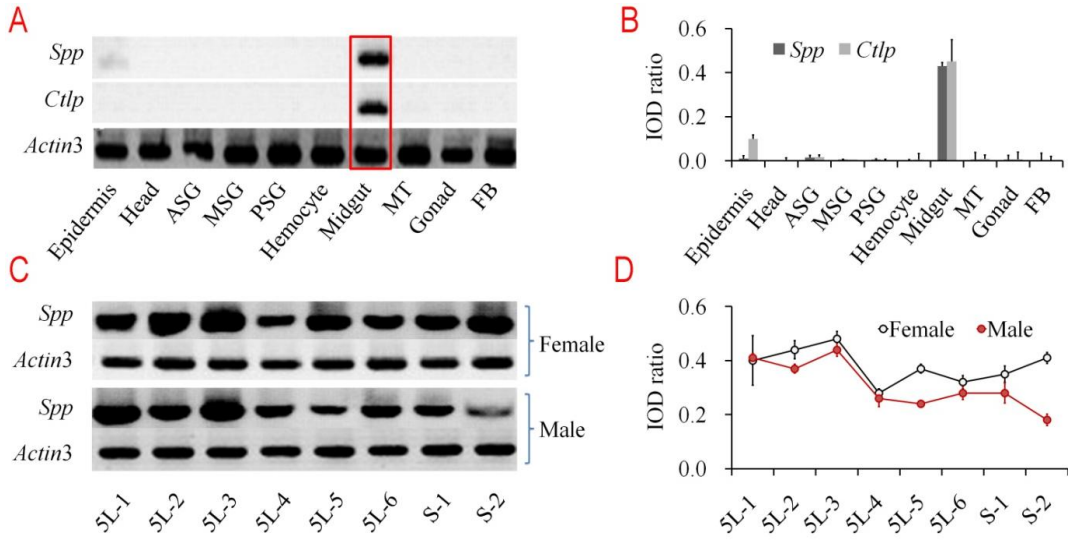

Figure 5. The spatio-temporal expression profiles of *Spp* and *Ctlp* measured by RT-PCR between the sexes in larvae of the Ysh strain. Tissue expression profile of electropherogram (A) and diagram based on integrated optical density (IOD) from electropherogram (B) of *Spp* and *Ctlp* in both sex on day three of 5th instar larvae. Stage expression profile of electropherogram (C) and diagram based on integrated optical density (IOD) from electropherogram (D) of *Spp*. *Actin* 3 was used as a control.

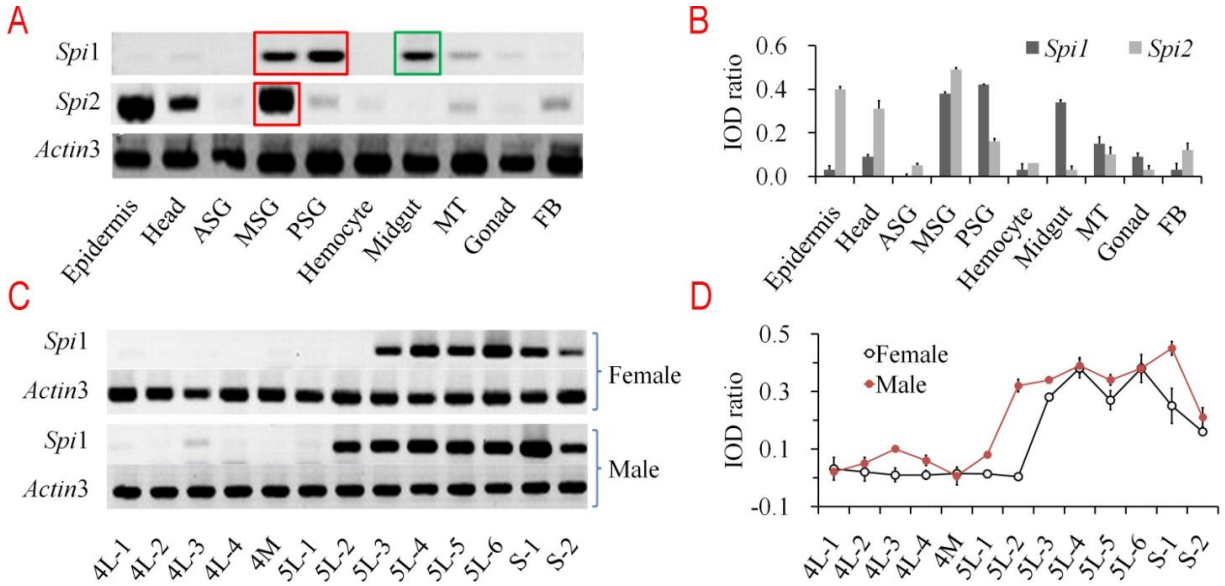

Figure 6. The spatio-temporal expression profiles of *Spi*1 and *Spi*2 measured by RT-PCR between the sexes in larvae of Ysh strain. Tissues expression profile of electropherogram (A) and diagram based on integrated optical density (IOD) from electropherogram (B) of *Spi1* and *Spi*2 in both sex on day three of 5th instar larvae . Stage expression profile of electropherogram (C) and diagram based on integrated optical density (IOD) from electropherogram (D) of *Spi1*. *Actin* 3 was used as a control.

specifically a cocoon shell. However, there are striking differences between the feeding efficiencies of the two larva sexes. The 5th instar of *Bombyx* is a transition period for the metamorphosis from larva to pupa, and some 90% of silk proteins were biosynthesized within this period. Based on the fact that it is difficult to discern the sex of normal larvae in sericulture, heredity strains linked to the W chromosome, such as sex-limited egg color, sex-limited stripe, and sex-limited cocoon color, were obtained by creating and applying chromosomal translocation mutants.

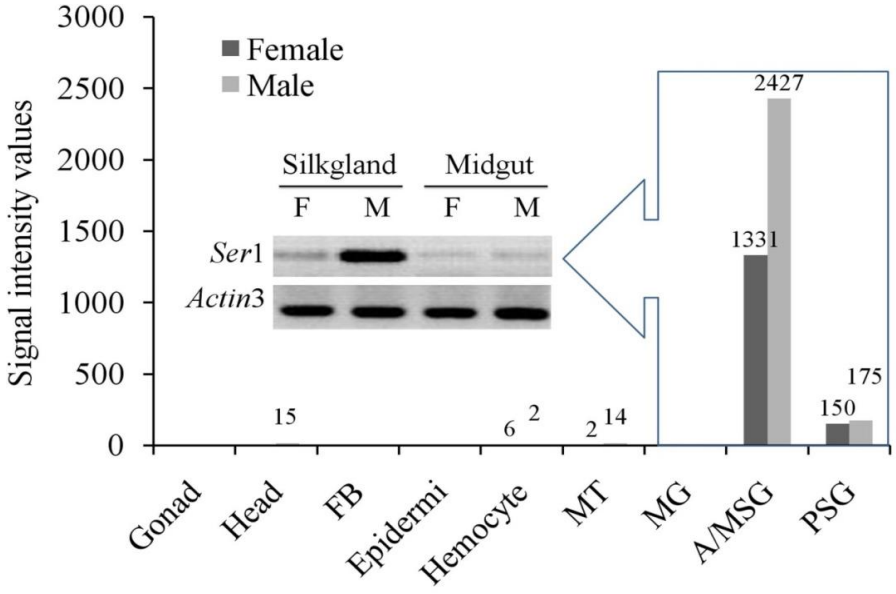

Figure 7. Tissue expression profile of the *Ser*1 gene between the sexes. The histogram shows the results measured by BmMDB in the Dazao strain for day three 5th instar larvae, and the electropherogram shows RT-PCR results in the Ysh strain for the whole 5th instar larvae .

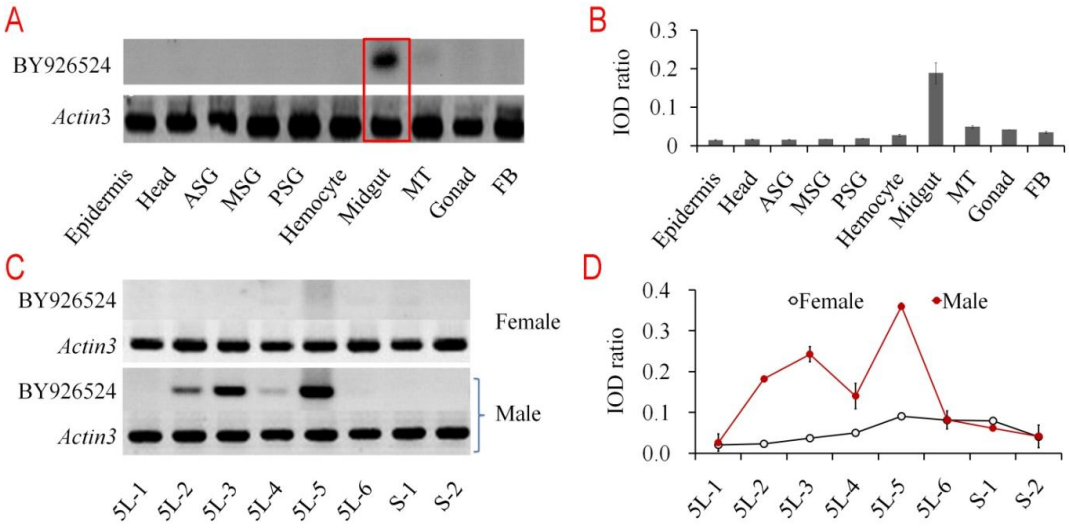

Figure 8. Spatio-temporal expression profile of the novel tag annotated EST BY926524 was measured by RT-PCR between the sexes in larvae of the Ysh strain. Tissues expression profile of electropherogram (A) and diagram based on integrated optical density (IOD) from electropherogram (B) of EST BY926524 for day three 5th instar larvae. Stage expression profile of electropherogram (C) and diagram based on integrated optical density (IOD) from electropherogram (D) of EST BY926524. *Actin* 3 was used as a control.

These strains are convenient for sericulture and provide good materials for investigating the disparity in feeding efficiency between the sexes.

Sex-dependent visible mutations of *B. mori* have been identified and the differences between the sexes of proteins translated in the middle silk gland have also been investigated (Jin et al., 2004). However, the genes involved in differential expression of yellow cocoon color sex-limited varieties have not yet been identified nor have these genes been analyzed in the published data from corresponding silkworm tissues. Therefore, it is necessary to establish an expression library for excavating and

identifying silk synthesis related genes that are differentially expressed between the sexes of silkworm.

Although the public database BmMDB provided the gene expression profile of 10 tissues on day three of 5th instar of the Dazao strain, it does not cover the differentially expressed genes during all the silk protein synthesis stages (Xia et al., 2004). Therefore, the database BmMDB cannot meet the requirements of excavation and analysis of these genes. Thus, studies of the silk gland and midgut in 5th instar will help elucidate the complex regulatory mechanism of the silk protein synthesis process and help explain the disparity in synthesis efficiency between males and females. In the SAGE libraries, more than 70 high-abundance genes involving a silk protease and a special protein in the midgut or silk gland were observed. Verification of these annotated genes and tags with EST evidence showed that their expression tendencies in both sexes were similar to those in corresponding SAGE libraries.

Low-abundance tags with additional utility value in the SAGE libraries

Previously, SAGE analysis showed that higher expression tags can be matched to known gene sequences, while the low expression tags do not match to any of the silkworm genome sequence, indicating that these low-expression tags corresponded to unknown or poorly characterized transcripts with specific tissue distribution (Virlon et al., 1999).

Thus, we did not pay enough attention to these low-expression tags and may have missed some important information. Several sex-determination homologous genes have previously been reported to show expression differences between male and female silkworms (Wang et al., 2009, Suzuki et al., 2003). Of these genes, groucho (*gro*) and transcript variant F (*tra2*) were expressed at higher levels in the ovary while sex-lethal (*sxl*) and intersex (*ix*) were expressed at higher levels in the testis (Schütt and Nöthiger, 2000, Zha et al., 2009).

Although tags corresponding to sex-determining cascade genes were searched in our SAGE libraries, their copy numbers had no notable differences between the sexes because these expression tags are gonad-specific and neither silk gland nor midgut-specific (Supplementary Table 3). In the present study, 79.3% of the unique tags (copy number less than 5) are of low frequency in the two SAGE libraries. Additionally, most of these tags were unmatched SAGE tags, and more than 74.5% of them were alternative tags in both sexes. Tag GCGAACG-CTTGGCGCAT was annotated to a carotenoid-binding protein gene (*cbp*) (Tabunoki et al., 2004; Sakudoh et al., 2007), a key protein gene involved in cocoon coloration that was translocated onto the W chromosome in the Ysh strain; however, the tag copy numbers were 1 and 0 in the female and male libraries, respectively.

This tag belonged to low-abundance tags and is not well characterized. Most of the unmatched SAGE tags are novel tags derived from novel transcripts. These novel transcripts may originate from the alternatively spliced transcripts (Mironov et al., 1999; Pauws et al., 2001); they may also belong to the non-coding transcripts that have multiple regulatory functions (Erdmann et al., 2001).Therefore, a suitable method for the future study of these tags is necessary. SAGE analysis elucidated a novel collection of mRNAs, exclusively genes expressed differentially between the sexes in the silk gland and midgut cells during the 5th instar.

Conclusion

The results of the present study indicate that the constructed SAGE libraries are of great value to excavate genes related to protein synthesis, and they are conducive to understand the mechanisms of sexual disparity in silk protein production efficiency. We can speculate that, if true, more genes annotated by those low-abundance tags in our SAGE libraries may intervene in silk protein synthesis efficiency and silk quality between the sexes of *Bombyx.*

ACKNOWLEDGEMENTS

This work was supported by the National High-tech Research and Development (R&D) Program of China (863 Program) (Grant No. 2011AA100306), (Suzhou University) National Engineering Laboratory for Modern Silk Fund, National Natural Science Foundation of China (Grant No. 31172264), Major Applied Research Program of Soochow University (Project No. Q3034850), the University Innovation Team Construction Project of Chongqing Education Committee (201040) and the Scientific Research and Innovation Team of Chongqing Three Gorges University.

Abbreviations:

A/MSG, Anterior/median silk gland; **BmMDB,** *Bombyx mori* microarray database; **GO,** gene ontology; **MG,** midgut; **PSG,** posterior silk gland; **SAGE,** serial analysis of gene expression; **TPM,** transcripts per million; ***Sp*1,** sex-specific storage-protein 1 gene; ***Lp-c*19,** low molecular mass 30 kDa lipoprotein 19G1 gene; ***Spp,*** serine protease precursor gene; ***Ctlp*,** chymotrypsin-like serine protease gene; ***Spi*,** kazal-type serine proteinase inhibitors gene; ***Ser*1,** sericin1 gene.

REFERENCES

Bala P, Georgantas RW 3rd, Sudhir D, Suresh M, Shanker K, Vrushabendra BM, Civin CI, Pandey A (2005). TAGmapper, a

web-based tool for mapping SAGE tags. Gene, 364: 123-129.
Chen J, Lee S, Zhou G, Wang SM (2002). High-throughput GLGI procedure for converting a large number of serial analysis of gene expression tag sequences into 3′complementary DNAs. Genes Chromosomes Cancer, 33: 252-261.
Erdmann VA, Barciszewska MZ, Hochberg A, Groot N, Barciszewski J (2001). The non-coding RNAs as riboregulators. Nucleic Acids Res., 29(1): 189-93.
Fujii T, Shimada T (2007). Sex determination in the silkworm, *Bombyx mori*: a female determinant on the W chromosome and the sex-determining gene cascade. Semin. Cell Dev. Biol., 18: 379-388.
Funaguma S, Hashimoto S, Suzuki Y, Omuro N, Sugano S, Mita K, Katsuma S, Shimada T (2007). SAGE analysis of early oogenesis in the silkworm, *Bombyx mori*. Insect. Biochem. Mol. Biol., 37: 147-154.
Gan LP, Xi J, Xu L, Sima YH , Xu SQ (2010). Improved GLGI Method to Identify Tags from LongSAGE of Silkworm, *Bombyx mori*. China Biotechnol., 30: 68-74. (in Chinese).
Huang J, Miao X, Jin W, Couble P, Mita K, Zhang Y, Liu W, Zhuang L, Shen Y, Keime C, Gandrillon O, Brouilly P, Briolay J, Zhao G ,Huang Y (2005). Serial analysis of gene expression in the silkworm, *Bombyx mori*. Genomics, 86: 233-241.
Jin YX, Chen YY, Xu MK, JIANG YH (2004). Studies on middle silk gland proteins of cocoon colour sex-limited silkworm, *Bombyx mori* L. using two-dimensional polyacrylamide gel electrophoresis J. Biosci., 29: 45-49.
Kerr MS (1969). The hemolyrnph proteins of the blue crab, *Callinectes sapidus*. A lipoprotein serologically identical to oocyte lipovitellin. Dev. Biol., 20: l-17.
Kim EJ, Park JH, Park TH (2003). Inhibition of apoptosis by recombinant 30K protein originating from silkworm hemolymph. Biochem. Biophys. Res. Commun., 308: 523-528.
Kurioka A, Yamazaki M, Hirano H (1999).Primary structure and possible functions of a trypsin inhibitor of Bombyx mori. Eur. J. Biochem., 259: 120-126.
LASH AE, Tolstoshev CM, Wagner L (2000). SAGE map: a public gene expression resource. Genome Res., 10(7): 1051-1060.
Li Y, Chen K, Yao Q, Li J, Wang Y, Liu H, Zhang C, Huang G (2009). The effect of calorie restriction on growth and development in silkworm, *Bombyx mori*. Arch. Insect. Biochem. Physiol., 71: 159-72.
Mahendran B, Ghosh SK, Kundu SC (2006). Molecular phylogeny of silk-producing insects based on 16S ribosomal RNA and cytochrome oxidase subunit I genes J. Genet., 85: 31-38.
Mironov AA, Fickett JW, Gelfand MS, Kerr MS (1999). Frequent alternative splicing of human genes. Genome Res., 9(12): 1288-1293.
Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-Okuda K, Yamamoto K, Ajimura M, Ravikumar G, Shimomura M, Nagamura Y, Shin-I T, Abe H, Shimada T, Morishita S, Sasaki T (2004). The genome sequence of silkworm, *Bombyx mori*. DNA Res., 11: 27-35.
Pauws E, van Kampen AH, van de Graaf SA, de Vijlder JJ, Ris-Stalpers C (2001). Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. Nucleic Acids Res., 29(8): 1690-1694.
Poole RL, Barker GL, Werner K, Biggi GF, Coghill J, Gibbings JG, Berry S, Dunwell JM, Edwards KJ (2008). Analysis of wheat SAGE tags reveals evidence for widespread antisense transcription. BMC Genomics, 9: 475.
Sakudoh T, Iizuka T, Narukawa J, Sezutsu H, Kobayashi I, Kuwazaki S, Banno Y, Kitamura A, Sugiyama H, Takada N, Fujimoto H, Kadono-Okuda K, Mita K, Tamura TYamamoto K, Tsuchida K (2010). A CD36-related transmembrane protein is coordinated with an intracellular lipid-binding protein in selective carotenoid transport for cocoon coloration. J. Biol. Chem., 285: 7739-7751.
Sakudoh T, Sezutsu H, Nakashima T, Kobayashi I, Fujimoto H, Uchino K, Banno Y, Iwano H, Maekawa H, Tamura T, Kataoka H, Tsuchida K (2007). Carotenoid silk coloration is controlled by a carotenoid-binding protein, a product of the Yellow blood gene. Proc. Natl. Acad. Sci. USA., 104: 8941-8946.
Schütt C, Nöthiger R (2000). Structure, function and evolution of sex-determining systems in Dipteran insects. Development, 127: 667-677.
Suzuki MG, Funaguma S, Kanda T, Tamura T, Shimada T (2003). Analysis of the biological functions of a doublesex homologue in *Bombyx mori*. Dev. Genes. Evol., 213: 345-354.
Tabu`noki H, Higurashi S, Ninagi O, Fujii H, Banno Y, Nozaki M, Kitajima M, Miura N, Atsumi S, Tsuchida K, Maekawa H, Sato R (2004). A carotenoid-binding protein, CBP plays a crucial role in cocoon pigmentation of silkworm, *Bombyx mori*. larvae. FEBS Lett., 567: 75-178.
Virlon B, Cheval L, Buhler JM, Billon E, Doucet A, Elalouf JM (1999).Serial microanalysis of renal transcriptomes. Proc. Natl. Acad. Sci. USA., 96: 15286-15291.
Wang Z, Zha X, He N, Xiang Z, Xia Q (2009).Molecular cloning and expression analysis of Bmrbp1, the *Bombyx mori* homologue of the Drosophila gene rbp1. Mol. Biol. Rep., 37: 2525-2531.
Xia Q, Cheng D, Duan J, Wang G, Cheng T, Zha X, Liu C, Zhao P, Dai F, Zhang Z, He N, Zhang L, Xiang Z (2007). Microarray-based gene expression profiles in multiple tissues of the domesticated silkworm. *Bombyx mori*. Genome Biol., 8.R162.
Yano K, Sakurai MT, Izumi S, Tomino S (1994). Vitellogenin gene of the silkworm, *Bombyx mori*, structure and sex-dependent expression. FEBS Lett., 356: 207-211.
Zha X, Xia Q, Duan J, Wang C, He N,Xiang Z (2009). Dosage analysis of Z chromosome genes using microarray in silkworm, *Bombyx mori*. Insect. Biochem. Mol. Biol., 39: 315-321.
Zhang Y, Huang J, Jia S, Liu W, Li M, Wang S, Miao X, Xiao H, Huang Y (2007). SAGE tag based cDNA microarray analysis during larval to pupal development and isolation of novel cDNAs in *Bombyx mori*. Genomics, 90: 372-379.
Zhao P, Wang GH, Dong ZM, Duan J, Xu PZ, Cheng TC, Xiang ZH, Xia QY (2010). Genome-wide identification and expression analysis of serine proteases and homologs in the silkworm *Bombyx mori*. BMC, Genomics, 11: 405.
Zou Z, Picheng Z, Weng H, Mita K, Jiang H (2009). A comparative analysis of serpin genes in the silkworm genome. Genomics, 93: 367-375.

Supplementary Table 1. Differently expression tags from two SAGE libraries. Part A: 76 tags annotated to 69 genes in detail.

No.	Tag	F (SG+MG)	M (SG+MG)	Accession number	Description	Gene symbol	Probe	Way
Genes up-regulated in female								
1	AGGTACCCATTCCATTT	959	204	NM_001046694.1	Triacylglycerol lipase	LOC692849	sw18759	RL
2	CTTTACAAGTTTGGCAG	58	15	NM_001046694.1	Triacylglycerol lipase	LOC692849	sw18759	GLGI
3	GCCTGGGGATACAATGG	33	7	Q00802.1	Low molecular mass 30 kda lipoprotein 19G1	Lp-c19	sw12676 sw17422 sw22699 sw22776	RL
4	GCCTGGGGATACAACGG	21	7	Q00802.1	Low molecular mass 30 kda lipoprotein 19G1	Lp-c19	sw12676 sw17422 sw22699 sw22776	RL
5	GTACATCGTCCCCTACT	25	8	P09334.1	Low molecular 30 kda lipoprotein PBMHP-6	Lp-c6	sw06038	RL
6	AATAATGTAATAAACGA	13	3	P09338.1	Low molecular 30 kda lipoprotein PBMHPC-23	Lp-c23	sw19103	RL
7	ATGAACTGATCACACGA	15	4	P09179.1	Sex-specific storage-protein 1	SP 1	sw12337	RL
8	AATTTCTTCAATCAACA	226	106	NM_001046965.1	Chymotrypsin-like serine protease	LOC732967	sw03707	RL
9	CCAGAAGTAAATGTAAG	48	11	NM_001046965.1	Chymotrypsin-like serine protease	LOC732967	sw03707	GLGI
10	ATTTTTTTGTAAATAAA	183	80	NM_001043361.1	Serine protease precursor	LOC692357	sw22515	RL
11	CGATTGCAATAAACCAA	139	27	NM_001044073.1	Eukaryotic translation initiation factor 5A	LOC693075	sw21925	GLGI
12	GACGTACCTCACGTGAA	49	18	DQ201182.1	Eukaryotic translation initiation factor 5A	LOC693075	sw21925	RL
13	CACTTGCCGCTTTCGCC	137	58	FJ556992.1	Putative membrane protein		sw08572	RL
14	CTAGAAAACCTATGGTA	34	9	ACL99854.1	Putative membrane protein		sw08572	RL
15	TTTAATTTAGACTTTTG	16	34	AB499730.1	Hypothetical protein			RL
16	TACATAGTCGATGGCAA	100	45	NM_001044109.2	Fatty acid binding protein	LOC732863	sw22673	RL
17	GGATTCGTTGAAACCAC	102	22	NM_001111333.1	Bmbr-C gene for Broad-Complex isoform Z2	Br	sw21738	GLGI
18	CGATTGCAATAAACCAC	93	23	U07847.1	Truncated alpha-amylase (amy) gene		sw21252	GLGI
19	TTCCTATTAGATGATTT	86	33	NM_001044015.1	Chitinase protein [Hyphantria cunea]	Chi	sw19054	RL
20	TAATGAAGAATATTTGT	50	17	FJ602785.1	TGF-beta-inducible nuclear protein 1	Tinp1	sw16922	RL
21	TTGGACCCGCTCTAGTT	36	10	ABD36116.1	Electron-transfer-flavoprotein beta polypeptide	LOC692811	sw01078	RL
22	CCGAAGTTGAATTTAAT	32	1	NP_001036964.1	Beta-tubulin (Tub2)	Tub2	sw04294	GLGI
23	CAGTCCACCCACAACAG	28	10	NM_001135883.1	Hatching enzyme-like protein	Hel	sw22496 sw18856 sw18724	RL
24	TTCTGGTTCTTCGGCAG	26	3	AP009039.1	Genomic DNA, chromosome 6			RL
25	TTTCTCAACGGTCTACA	21	2	NM_001166142.1	Beta-1,3-glucan recognition protein 4	BGRP4	sw20413	RL
26	CATCGGTAACAATAGTG	15	4	ABF51430.1	Proteasome subunit alpha type 6-A	LOC732998	sw09073	RL
27	AATATAAAGTATTCAGA	15	0	NM_001160196.1	Histone H2B-like protein	H2b-l	sw13053	RACE
28	CATTCTGTACAAAATTA	13	2	NM_001046687.1	Sr protein	LOC692842	sw03860	RL
29	AAAAGAACAATGGCACA	13	2	NM_001046815.1	Bombyx mori signal peptidase 18 kda subunit	LOC692977	sw17051	RL
30	TCACCGAAAAACTTTTG	19	6	ABF51367.1	H+ transporting ATP synthase gamma subunit	LOC732965	sw13289	RL
31	ACAGTGTTGATATATCT	25	1	NM_001098340.1	Proteasome 26S non-atpase subunit 4	LOC778515	sw13233	RL,GLGI
32	TTTTTGATGTCCGTTGA	27	9	ABF51406.1	Vacuolar ATP synthase subunit F	LOC732987		RL
33	GTTTTTTGAATACTAAT	13	3	AAR97568.1	Cu/Zn superoxide dismutase	SOD	sw00530	RL
34	ATGCAACATCGGGTTTT	7	0	ABK30932.2	NADPH oxidoreductase	LOC100037421	sw13877 sw02589	RL
35	TGTTTGGTAACAAATCA	13	0	X15557.1	Chorion locus sequence encompassing gene			GLGI
36	TAATCTTTCCTGTTCTT	10	1	AP009012.1	Genomic DNA, chromosome 14			RL
37	CTGAATGTTATCGTATT	69	33	AP009038.1	Genomic DNA, chromosome 12			RL
38	TTGGAACCTTTCATTTA	7	0	AP009008.1	Genomic DNA, chromosome 4			RL
39	GCCAGCCAGGGCGACTA	356	67	NM_001043790.1	Ribosomal protein s3Ae	Rps3ae	sw06340	GLGI
40	GCATCTACCCTCTGCGC	130	50	NM_001043790.1	Ribosomal protein S3Ae	Rps3ae	sw06340	RL

Supplementary Table 1. Contd.

No.	Tag	F (SG+MG)	M (SG+MG)	Accession number	Description	Gene symbol	Probe		Way
Genes down-regulated in female									
41	GCCAGCCAGGGCGATTA	362	848	NM_001113274.1	Carboxypeptidase inhibitor	LOC100134929	sw08920		GLGI
42	TCCTTAATTATATATGG	106	244	NM_001043579.1	Kazal-type serine proteinase inhibitor 1	Spi1	sw12002		RL
43	TCCTTAATTATATATAG	3	16	AAL83944.1	Kazal-type serine proteinase inhibitor 1	Spi1	sw12002		RL
44	TGAATTTAGGGTTGCCA	11	42	NM_001043582.1	silk proteinase inhibitor 2	Spi2	sw12507		RL
45	TGCTTCTATGTTTATGA	0	14	NM_001135888.1	Osiris 9	Osi9	sw05625		RL
46	TGGAGCACCACGTAGAC	0	8	NM_001135889.1	Osiris 18	Osi18	sw03878		RL
47	TGTGACTGTGTGATTGT	0	49	NM_001135890.1	Osiris 19	Osi19	sw11469		RL
48	AAAAGTTGAACCGTCAT	0	34	BR000473.1	TPA: putative cuticle protein		sw00693		RL,GLGI
49	TGTATATCAAGTTAGAA	0	18	FAA00626.1	TPA: putative cuticle protein	CPR123	sw11845		RL
50	TGATCGCATCTTAGAGT	0	7	NM_001173318.1	Cuticular protein glycine-rich 14	CPG14	sw07050		RL
51	GAGGTTCGGGAGTCAGC	0	13	NM_001043486.1	Cuticular protein tweedle motif 3	CPT3	sw12541		RL
52	CTTGTTATTAATACGTA	0	8	NM_001173311.1	Cuticular protein glycine-rich 24	CPG24	sw18518		RL
53	TAACAATAACAATCTTA	8	35	GQ338156.1	Putative fatbody protein 3rev-g1 mrna				RL
54	TTTGGTTAAAATTATAT	99	206	NM_001098353.1	Poly A binding protein	Pabp	sw15810		RL
55	ACTACGAATTATTGAAA	50	124	AAF73760.1	AF149768_4 NADH dehydrogenase 4L	ND4L	sw19438		RL
56	GCGCTTTCCGAAATTTT	0	7	NM_001114990.1	Glutathione S-transferase unclassified 1	GSTu1	sw12179		RL
57	ATGGAAGCGTAACATCC	46	94	AB112020.1	Sericin 1	Ser1	sw14946		RL
58	CACGCAAATTCGGACCC	21	42	NM_001043933.1	Muscle LIM protein, transcript variant 2	LOC692826	sw21951	sw21952	RL
59	TTAATGTTCTTCCAATA	8	47	NM_001142753.1	Hypothetical protein LOC100216500	LOC100216500	sw16045		RL
60	ATACTTTGAATGGGTTC	0	47	NM_001043602.1	Hypothetical protein LOC692620	LOC692620	sw14913		RL
61	TAGGAATAAATTATTAT	8	30	NM_001046851.1	NADH dehydrogenase (ubiquinone) Fe-S protein 8	Ndufs8	sw00924		RL
62	TTGTACTTGGTGGTGGT	8	22	NM_001046968.1	Eukaryotic translation initiation factor 3 subunit I	eIF3i	sw14403		RL
63	GACGTACCCCACGTGAA	7	28	DQ202521.1	Eukaryotic translation initiation factor 5A mRNA		sw21925		RL
64	TTTTCTATGTTGTTGAA	7	21	DQ311164.1	enOyl-CoA hydratase precursor 3 mRNA		sw19440		RL
65	AAACTAACAACTGAGCT	6	21	NM_001043365.1	Trachealess	trh	sw15312		RL
66	GACCTCACAACTGATAA	45	105	NM_001043790.1	Ribosomal protein s3Ae (Rps3ae)	Rps3ae	sw06340		RL,GLGI
67	AATTTGCGAGTGCGAAT	7	22	NM_001099607.1	Ribosomal protein S24	RpS24	sw06149		RL
68	ATGGCGAGTGTGTTTTT	2	31	NM_001043748.1	Ribosomal protein P2	Rpp2	sw18239		GLGI
69	GAGGTGGCGGATATGGT	0	10	BAB78527.1	Ribosome-associated protein P40	RpSA			RL
70	AGAGAGAGTGTTTTCGG	13	52	NM_001043895.1	Ribosome-associated membrane protein 4	LOC692773	sw11021		RL
71	TACCTATTATTTTGAGG	1	10	NM_001046790.1	Lysophospholipase (LOC692950)	LOC692950	sw06684		RL
72	TTAATTTATTTTGTGAG	0	15	NM_001047017.1	Bombyx mori phosphate transport protein	LOC733022	sw12463		RL
73	GTTGCGTTGCCCAAGTT	15	41	NM_001044109.2	Fatty acid binding protein	LOC732863	sw22549	sw22673	RL
74	TTAGGATCCTTGGAAAA	0	11	AB417664.1	Non-coding RNA, ovarian small RNA-31474				RL
75	TCCGCATCCTGGCAGAA	0	9	Q75VN3.1	Translationally-controlled tumor protein homolog	Tctp	sw22898		RL
76	AACGATTAATCAAATAA	0	22	NM_001043501.1	Lipase-1	LOC692513	sw14430		GLGI

Supplementary Table 1 (contd). Part B: 15 tags annotated with ESTs.

Tag	F (SG+MG)	M (SG+MG)	EST
GAACCGAGCGAGATAAA	0	24	DY231379
TTTCATTATAAAATGGA	0	7	DC550967
GCTTCCGCCGTGCCTGC	3	18	BY926524
TGCACTTATTATTATAT	2	12	AV405687
TTTTTAATTGAAAGTGA	5	20	DN237319
AAATAAAAACTCTATTC	8	29	BY916534
ATTGACATTGCAAAGAT	5	17	DC551580
GAGAGCGTTTGACTGAA	13	42	CK515399
AAGTTGAAAGGACACCA	14	31	DY230799
GCCTTCAAGTGTATGGA	17	35	BP177687
TTAATAAATGATTACTT	52	26	DC544935
TGGCTGATAACAGCCGG	149	61	DC532049
CGGGAAAGTTATAGGAA	15	4	BJ984951
CTTATGAATAATGAAAT	26	4	CK513283
TTGGCGTTTAAGAAATC	15	1	BB988365
GAAACAGCACGCACGCT	2	18	BB991666

Supplementary Table 1 (contd). Part C: 111 tags (55%) left unmatched.

Tag	F (SG+MG)	M (SG+MG)	Annotation
ATCTGAGTTCAAACCGC	0	25	No hits found!
ATAATTGTAATTAACGT	0	17	No hits found!
CAACCAATATACACATT	0	13	No hits found!
TTCAGGAAGTGCTTGGA	0	12	No hits found!
ACTCCTGATGAGCTTGT	0	11	No hits found!
AATCAATATGTGCTAAC	0	10	No hits found!
AGAGGTGTAGCATAACT	0	10	No hits found!
CTTAATTTGTACATACA	0	10	No hits found!
AGCACCACGAACATCAC	0	8	No hits found!
TAAAAAAAATGGAAATA	0	8	No hits found!
TCAACAAATTAGATATT	0	8	No hits found!
TTACTGTTACCCCAGGT	0	8	No hits found!
TTCGCAAGAATCCATAT	0	8	No hits found!
ATCCGAGTTCAAACCGG	0	7	No hits found!
ATTGCTGATAAAGGAGC	0	7	No hits found!
CAAATATACTTAAATAG	0	7	No hits found!
CACTTACATAACAAAAA	0	7	No hits found!
CGGCGGAAGTGCCGCAC	0	7	No hits found!
GAATACAAAGATGGCGG	0	7	No hits found!
GAATCCTATAGAAGGGC	0	7	No hits found!
TTTAGATGAGGAAACAT	0	7	No hits found!
TCCTTAATTATATATGA	1	24	No hits found!
AATCGGAGCCCGACCTC	1	16	No hits found!
TTATATCTGCAAAAACG	1	15	No hits found!
CAAAAGAAACTTAAAAA	1	13	No hits found!
TGTCATCTATAGGTCTT	1	13	No hits found!
ATCTGAGTTCAAACCAG	1	12	No hits found!
GTAAATATTTTCAAGAT	1	12	No hits found!
AGTTTGAAGAAATAGAG	1	11	No hits found!

Supplementary Table 1 (contd). Part C: Contd.

AACCCTGTAGAGCATCT	1	10	No hits found!
CTATTCCTGTTCTACGA	1	10	No hits found!
TGGGAATCTTTGAGGTT	1	10	No hits found!
TCCGTACATTTGTCGCT	2	20	No hits found!
GCCAGCCAGGGCGATTG	2	19	No hits found!
TGCAATGGTGACTCTGG	2	19	No hits found!
AAGACAGTTCCAAGAGT	2	17	No hits found!
CAAAAGAAACTTAAATT	2	13	No hits found!
ATCTGAGTTCAAACCGA	3	31	No hits found!
AATTCAGTTTTATAAAT	3	23	No hits found!
GAAATATAATTTACAAC	3	23	No hits found!
CATTTTATCTATTGATG	3	19	No hits found!
TACAGTTGTTTAATAAA	3	19	No hits found!
ATCTGAGTTCAAACCGT	3	18	No hits found!
AATGTTTGTAGAAAACT	3	14	No hits found!
TCCGCATCCTGGCAGGG	3	14	No hits found!
TGTGTTAGAACATCCCT	4	25	No hits found!
ATATGGACAAAACACAC	4	22	No hits found!
ATCGCGTCGTGACTGGT	4	19	No hits found!
AGATTACAATAAAGTAG	4	16	No hits found!
CAGATCTGCTGTCAGGG	5	20	No hits found!
AACAATTCCTTCATTAA	6	20	No hits found!
AGGTACCCATTCCATTC	7	0	No hits found!
CTGCAAGCAAACACCGC	7	0	No hits found!
TATATTCCGTTCAGTTT	7	0	No hits found!
TATTAGTTGTAGTCATT	7	0	No hits found!
TGTGAGAGTCACAATAC	7	0	No hits found!
TTAATTCCCTATGTTCA	7	0	No hits found!
AACAATTCCTTCGTTAG	8	0	No hits found!
AGGCACCCATTCCATTT	8	0	No hits found!
CGATTGCAATAAAAAAA	8	0	No hits found!
GTCACATCCAAAAAGTT	8	0	No hits found!
GCCAACTGCTACAGAAA	9	0	No hits found!
TGTTATTTTAAGAATTT	9	0	No hits found!
CTTTTTATAAAATATAA	9	30	No hits found!
TGAGTTATTTTTAGTTT	9	30	No hits found!
GGTCTGGAATGTTCCGA	10	0	No hits found!
TCGTATAGGCCGCAATT	10	0	No hits found!
TTAATACATTATAATGT	10	0	No hits found!
CAAAGAATAGCTAAGCA	10	1	No hits found!
TTAGATTTGAATTAGAA	12	0	No hits found!
TTTGTTTATCAGAAATT	12	2	No hits found!
TGCAGACAAATAAATAT	13	33	No hits found!
TAAAAAGCTTAAAAGAT	13	3	No hits found!
TTCTGGAATATAATGAA	13	2	No hits found!
CGTAATTGTTGCAGCAA	13	1	No hits found!
TCACTGGATCTGACGTT	13	1	No hits found!
AACAATTCCTTCATTAC	14	33	No hits found!
ATTTTTTGTAAATAAAC	14	3	No hits found!
GTGTTAAACAAATTTTT	14	3	No hits found!
TAAATAAAAGAAGAAAA	14	3	No hits found!
AAGTGTAATAAATAAAT	14	2	No hits found!

Supplementary Table 1 (contd). Part C: Contd.

CAAATACGCGAAGTTTT	15	4	No hits found!
TTGTTGTAATTTATAGA	16	34	No hits found!
AGCTCCAAGATTGAAAC	16	4	No hits found!
ATTTAACTCATTTTATG	17	2	No hits found!
GTATCATCTAGCGATGT	18	2	No hits found!
TTCCTCAACGGTCTACA	18	2	No hits found!
GACCATTACATACAATA	19	55	No hits found!
GACGCTGCGAATTCTAG	19	48	No hits found!
AACACTCGCACGTATTG	20	44	No hits found!
ATGACACGACTAAAAAA	20	7	No hits found!
CCGGTATTTACTCGGAA	20	3	No hits found!
GACGAAGGCTTGCTAGT	21	7	No hits found!
CAGGCCTAGATTATACC	22	7	No hits found!
TATTTTCTCATATTATT	22	7	No hits found!
CTCTGTGCATTTAATTA	23	9	No hits found!
TTCAGTATTCCTACAGT	23	9	No hits found!
TACAGTTCTTTAATAAA	30	13	No hits found!
TAATAATTTGAATTTTA	31	10	No hits found!
TAAATCCGTGTCGTATC	35	8	No hits found!
GGCTACCAGTTCTGGCT	36	15	No hits found!
TGTGATTATGAATCTAA	40	19	No hits found!
GTATTCTTGATTAAAGT	47	22	No hits found!
CAACTGAGTCATAAATC	48	23	No hits found!
ACGATCACGACCAACAC	51	24	No hits found!
CTATATGATAGGAGTGT	66	29	No hits found!
GATGATTATACCTTCGA	68	29	No hits found!
TTTACTGTAAAACTATT	71	195	No hits found!
TAAAATCTATATAAATT	91	15	No hits found!
TGGGGATCAGGTGCCAA	98	49	No hits found!

Supplementary Table 2. RT-PCR primers used in Figures 3 to 8.

Genes	Primer
Sex-specific storage-protein 1 (Sp1)	ACCGACGACGCATACGGCATAA (S) CACCGCTGGGAAGACGGATTTT (A)
Low molecular mass 30 kDa lipoprotein 19G1 gene (Lp-c19)	GAAGAGCGAAGTCATCACAAAT (S) TGAACTCAACTGGGAAACAATC (A)
Serine protease precursor (Spp)	GCACCAGGGTCACCACCTCCAA (S) CCCAGGCCCAAGTACCAGCAAA (A)
Chymotrypsin-like serine protease (Ctlp)	CCGTTCAAGGCGGCAGTGTATC (S) CAAACCGGCGCAGAATTTGTTA (A)
Kazal-type serine proteinase inhibitor 1 (Spi1)	TTCACTAACGCTAATCCTG (S) TTCAACGCATTTGTCCT (A)
Silk proteinase inhibitor 2 (Spi2)	GTCACCACCTCCAATGTCC (S) GAGGCTCACTTGGCGTTT (A)

Supplementary Table 2. Contd.

Sericin 1 (Ser-1)	CGAGGAAGTTCTGGTGA (S) ACTACGGTGGTTGGAGC (A)
BY926524	ACTTCCCCGTTCCTCCTT (S) GTTTGATTTCGCTTTCCT (A)
Actin 3	CTGCGTCTGGACTTGGC (S) CGAGGGAGCTGCTGGAT (A)

Supplementary Table 3. Tags corresponding to sex-determining cascade genes in males and females of Ysh.

Tag	Copy number (F)	Copy number (M)	Gene	Accession ID
TACGTTAAGTTGCGCCC	0	1	*Sex-lethal* (*Sxl*)	NM_001173383.1
CAAATGCCTGTACCAGG	0	1	Intersex (*ix*)	NM_001043640.1
TTTATATAGATTAGTGT	0	2	Groucho (*gro*)	NM_001134889.1
GGAAAACCTACAATAAG	0	1	transcript variant F (*tra2*)	NM_001126233.1
GCAGATGTCACCTCAAT	9	13	Doublesex*(Dsx)*	NM_001043406.1