Insect science

Insect Science (2017) 00, 1–13, DOI 10.1111/1744-7917.12443

ORIGINAL ARTICLE

Identification and comparison of long non-coding RNAs in the silk gland between domestic and wild silkworms

Qiu-Zhong Zhou[1], Shou-Min Fang[2], Qiang Zhang[1], Quan-You Yu[1] and Ze Zhang[1]

[1]*Laboratory of Evolutionary and Functional Genomics, School of Life Sciences, Chongqing University, Chongqing, China and* [2]*College of Life Science, China West Normal University, Nanchong, Sichuan China*

Abstract Under long-term artificial selection, the domestic silkworm (*Bombyx mori*) has increased its silk yield tremendously in comparison with its wild progenitor, *Bombyx mandarina*. However, the molecular mechanism of silk yield increase is still unknown. Comparative analysis of long non-coding RNAs (lncRNAs) may provide some insights into understanding this phenotypic variation. In this study, using RNA sequencing technology data of silk gland in domestic and wild silkworms, we identified 599 lncRNAs in the silk gland of the silkworm. Compared with protein-coding genes, the silk gland lncRNA genes tend to have fewer exon numbers, shorter transcript length and lower GC-content. Moreover, we found that three lncRNA genes are significantly and differentially expressed between domestic and wild silkworms. The potential targets of two differentially expressed lncRNAs (DELs) (*dw4sg_0040* and *dw4sg_0483*) and the expression-correlated genes with the two DELs are mainly enriched in the related processes of silk protein translation. This implies that these DELs may affect the phenotypic variation in silk yield between the domestic and wild silkworms through the post-transcriptional regulation of silk protein.

Key words domestication; long non-coding RNAs; RNA-seq; silkworm; silk gland; transcriptome

Introduction

Long non-coding RNAs (lncRNAs) are a kind of non-coding RNAs (ncRNAs) with at least 200 nucleotides (nt) and they are transcribed by RNA polymerase II (RNAPII) (Mercer *et al.*, 2009; Kornienko *et al.*, 2013; Cao, 2014). As the development of high throughput RNA sequencing technology (RNA-seq), lncRNAs have been identified across eukaryotes including *Drosophila melanogaster*, *Nilaparvata lugens*, *Plutella xylostella* and *Bombyx mori* (Young *et al.*, 2012; Etebari *et al.*, 2015; Legeai & Derrien, 2015; Xiao *et al.*, 2015; Chen *et al.*, 2016; Wu *et al.*, 2016). Previous studies have revealed that lncRNAs play important regulation roles in various biological processes (Mercer *et al.*, 2009; Wang & Chang, 2011; Cao, 2014). For instance, the *Air* lncRNA that is mapped to the imprinted domain of *Igf2r* in the mouse implicates in the imprinting of the protein-coding gene (Zhu *et al.*, 2013). In *Drosophila*, the *rox1/2* lncRNA takes part in the dosage compensation of X-chromosome (Meller *et al.*, 1997). Moreover, yellow-achaete intergenic RNA (*yar*) could affect the sleep behavior of *Drosophila* (Soshnev *et al.*, 2011). More and more evidence supports the view that lncRNAs play important roles in many fundamental processes of life (Mercer *et al.*, 2009; Stanojcic *et al.*, 2011; Kapusta *et al.*, 2013; Morris & Mattick, 2014; Jenkins *et al.*, 2015; Mattick & Rinn, 2015).

Domestication accelerates the phenotypic variation between domestic organisms and their wild progenitors. Previous studies have shown that lncRNAs can also take part in this process (Zhou *et al.*, 2014; Li *et al.*, 2015a). For example, *linc-sscg2561* may contribute to phenotypic variation in emotional behavior, including aggressive behavior

Correspondence: Ze Zhang, School of Life Sciences, Chongqing University, Chongqing 401331, China. Tel: +86 23 65678449; email: zezhang@cqu.edu.cn

and reduced fear of humans during domestication of the pig (Zhou *et al.*, 2014). The domestic silkworm, *B. mori*, is the only completely domesticated insect. Compared with its wild progenitor, *B. mandarina*, many phenotypic traits of the domestic silkworm have changed greatly, for example, silk yield, flight and body color. Indeed, silk yield of the domestic silkworm is about 3–5 times that of the wild silkworm (Normile, 2009; Takasu *et al.*, 2010; Fang *et al.*, 2015). However, little is known about the genetic mechanisms of increased silk yield in the domestic silkworm (Ma *et al.*, 2011). Our previous study suggested that the higher efficient biosynthesis and secretion of silk proteins may contribute to elevated silk yield in the domestic silkworm (Fang *et al.*, 2015). At present, most of the studies on silk yield focus on the protein-coding genes and microRNAs (miRNAs) (Li *et al.*, 2014b; Fang *et al.*, 2015). As shown earlier, lncRNAs are deemed to be an essential regulator in almost all biological processes. Moreover, a recent study reported that a family of antisense lncRNAs (SINEUPs) are able to increase translation of partially overlapping protein-coding messenger RNAs (mRNAs) (Zucchelli *et al.*, 2016). Therefore, it is valuable to know whether lncRNAs relate to silk yield variation between domestic and wild silkworms.

Herein, we systematically identified the lncRNAs expressed in the silk gland from two domestic silkworm strains and two geographical *B. mandarina* samples based on RNA-seq data. Then, comparisons of the domestic and wild silkworm transcriptomes revealed differentially expressed lncRNAs (DELs). We also surveyed the spatial and temporal expression patterns of these DELs and annotated the potential target genes and the expression-correlated genes of DELs based on Gene Ontology (GO) and the database KEGG (Kyoto Encyclopedia of Genes and Genomes). Our results not only expanded the repertoire of lncRNAs in the silkworm but also provided some insights into the molecular mechanism of silk yield variation between the domestic and wild silkworms.

Materials and methods

Data resources

The two domestic silkworm strains used were Chunhua (D_CH) and Chunyu (D_CY). The two wild silkworm samples (W_AKBH and W_AKSQ) were obtained from the Baihe and Shiquan counties of Ankang City, in Shaanxi Province, respectively (Fang *et al.*, 2015). The silk glands were dissected on day 3 of 5th instar larvae. The median silk gland (MSG) and posterior silk gland (PSG) were pooled and used for sequencing by the Illumina Hiseq 2000 (National Center for Biotechnology Information [NCBI] BioProject Accession: PRJNA262539). The genome sequence of the silkworm was downloaded from the Silkworm Genome Database (Xia *et al.*, 2004; Duan *et al.*, 2010). All non-redundant proteins and full-length complementary DNAs (cDNAs) of the silkworm were obtained from NCBI database (https://www.ncbi.nlm.nih.gov/) (Suetsugu *et al.*, 2013). The lncRNAs of a further eight species were downloaded from NONCODE (http://www.bioinfo.org/noncode/) (Xie *et al.*, 2014).

Transcriptome assembly and lncRNAs identification

A pipeline for prediction of lncRNAs in the silk gland integrates multiple methods of lncRNAs identification (Fig. 1) (Sun *et al.*, 2012; Young *et al.*, 2012; Ilott & Ponting, 2013; Chen *et al.*, 2014; Li *et al.*, 2014d; Zhou *et al.*, 2014). The pair-end clean reads were mapped to the reference genome with TopHat2 v2.0.9 (Nam & Bartel, 2012; Trapnell *et al.*, 2012). Unmapped reads were trimmed 10 bp on both sides and then remapped to the reference genome (Sun *et al.*, 2012). All mapped reads were used to reconstruct the transcriptome of silk gland with Cufflinks v2.1.1 (Trapnell *et al.*, 2012).

The assembled transcripts were compared to protein-coding genes using the program cuffcompare (Sun *et al.*, 2012; Trapnell *et al.*, 2012). The transcripts in which categories corresponded to “I”, “j”, “o”, “u” or “x” were retained and then the strand unknown transcripts were deleted (Sun *et al.*, 2012). Putative lncRNAs were defined as the length of transcript ≥ 200 bp and open reading frames (ORFs) ≤ 100 amino acids (Li *et al.*, 2014d; Liao *et al.*, 2014; Xiao *et al.*, 2015). The protein-coding potential of each transcript was estimated using Coding Potential Calculator (CPC) software (Kong *et al.*, 2007). The candidate transcripts were searched against Pfam databases (http://pfam.xfam.org) with e-values of 0.001 and non-significant transcripts were retained (Xiao *et al.*, 2015). Transcripts which had alignment length ≥ 100 amino acids and identity ≥ 80% with non-redundancy (nr) were removed (Pauli *et al.*, 2012; Liao *et al.*, 2014). Due to the defect of RNA-seq and assembly, not all transcripts are full-length transcripts (Trapnell *et al.*, 2012). Probably, an incomplete transcript may be a part of full-length protein-coding cDNAs. So the next step, if one flank of the transcript is mapped to the flank (at least 15 bp) of full-length cDNA, this transcript will be removed (Suetsugu *et al.*, 2013). The lncRNAs were considered as high-quality lncRNAs when the coverage of lncRNAs was ≥ 80% and the mapped reads of lncRNAs were ≥ 6 at least in one silkworm sample (Zhou *et al.*, 2014).

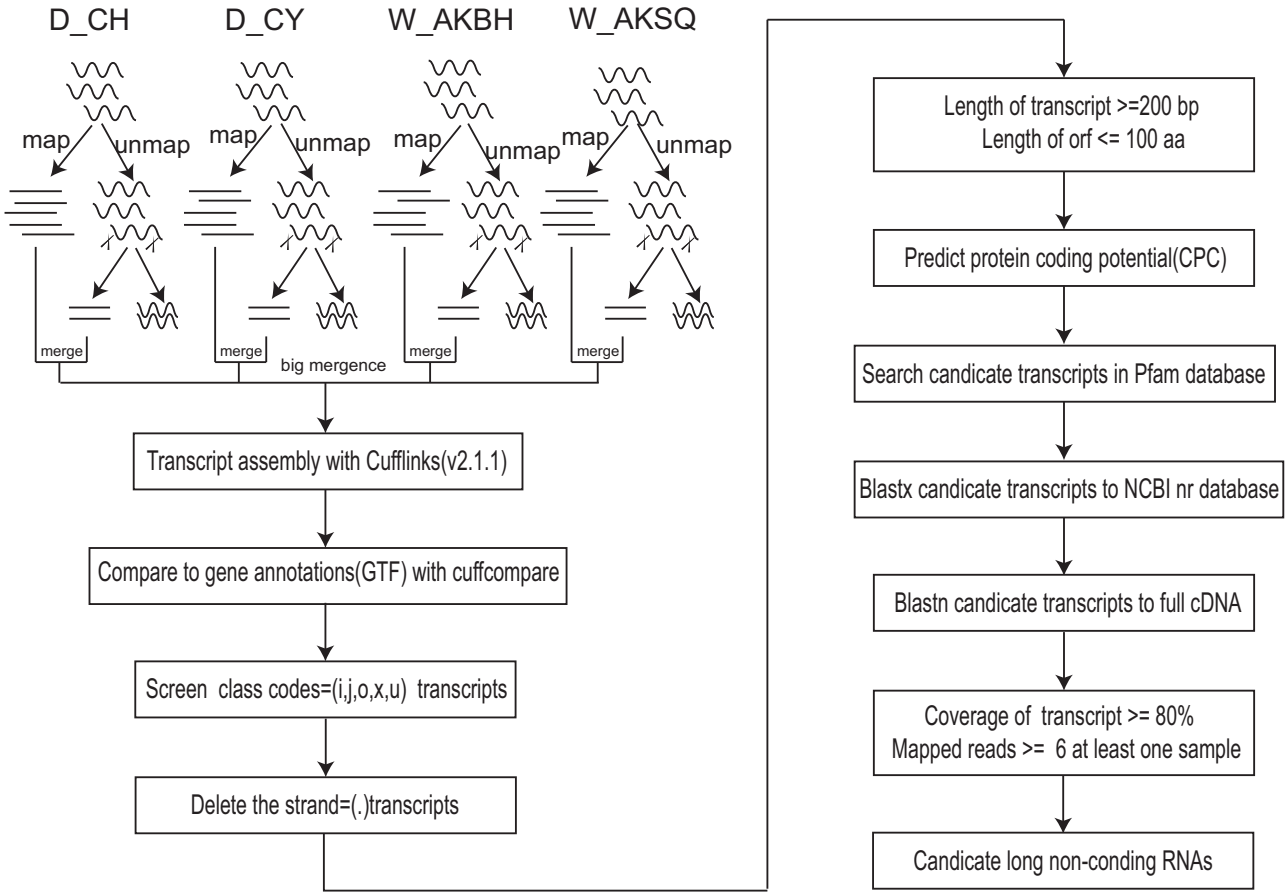

Fig. 1 Pipeline for the identification of silk gland long non-coding RNAs (lncRNAs). The top wavy lines represent RNA sequencing technology (RNA-seq) clean reads. Left straight lines represent the reads mapped to silkworm genome. Right wavy lines represent the reads unmapped to the silkworm genome. The 10 bp of the first unmapped reads were trimmed on each side and then the trimmed reads were remapped to the reference genome sequence. All the mapped reads (the straight lines) were assembled into the transcripts. The identification of silk gland lncRNAs was based on the transcriptome assembly.

Expression level of identified lncRNAs

We used HTSeq to count the uniquely mapped reads of lncRNAs in each silkworm sample (Anders *et al.*, 2014). DESeq, a package of R, was used to identify the differentially expressed lncRNAs between the domestic and wild silkworms (Anders & Huber, 2010). The genes with *P*-adjusted value ≤ 0.05 and absolute value of $\log_2$ fold-change ≥ 2 were considered as significantly differentially expressed (Chen *et al.*, 2014; Fang *et al.*, 2015).

Verification of differentially expressed lncRNAs by RT-PCR and qRT-PCR

Sequence-specific primers of DELs were designed using the online tool "primer-blast" with specificity checking parameter *B. mori* (taxid:7091) (Table S1) (Ye *et al.*, 2012). The reverse-transcription polymerase chain reaction (RT-PCR) of *dw4sg_0040* was performed under the following conditions: 94 °C for 4 min, 28 cycles of 94 °C for 30 s, 56 °C for 30 s, 72 °C for 50 s and a final extension at 72 °C for 10 min. The conditions for *dw4sg_0178* were: 94°C for 4 min, 29 cycles of 94 °C for 30 s, 56 °C for 30 s, 72 °C for 50 s and a final extension at 72 °C for 10 min. The conditions for *dw4sg_0483* were: 94 °C for 4 min, 28 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 50 s and a final extension at 72 °C for 10 min. Quantitative RT-PCR (qRT-PCR) was performed using CFX96^{TM} Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA) with SYBR Green qRT-PCR Mix (Bio-Rad). Expression patterns of DELs in nine tissues and 26 developmental stages of silkworm were surveyed by RT-PCR.

lncRNAs interacting in gene analysis and lncRNAs overlapping transposable elements

LncTar was used to find potential target genes of DELs in the 14 943 genes with option D -0.1 (Li *et al.*, 2015b). The potential target genes of DELs were annotated by BLASTx nr, interproscan, KOBAS and KAAS (Moriya *et al.*, 2007; Xie *et al.*, 2011; Jones *et al.*, 2014). Moreover, we sequenced silk glands of seven development time points from day 0 of the 5th instar to wandering stage in both domestic and wild silkworms (Zhou *et al.*, 2016, unpublished data). The expression-correlated genes of DELs were identified and the corresponding networks were constructed by using these unpublished RNA-seq data of silk glands. BLASTn was used to search the lncRNAs against DNA and RNA transposable elements (TEs) of silkworm with e-values $< e^{-5}$, coverage of lncRNA $\geq$ 80%, and identity $\geq$ 95% (Altschul *et al.*, 1997; Kapusta *et al.*, 2013).

Results

Silk gland transcriptome assembly

First, the short reads from RNA-seq data of silk glands of four different silkworms were aligned to the silkworm reference genome sequence (Xia *et al.*, 2004). In total, more than 163 million reads from two domestic silkworm strains (D_CH and D_CY) and two wild silkworm samples (W_AKBH and W_AKSQ) were successfully mapped to the reference genome sequence by TopHat2 v2.0.9 (Table 1). The mapped ratios for the silkworms D_CH, D_CY, W_AKBH, and W_AKSQ were 71.08%, 77.73%, 67.99% and 74.09%, respectively. All the mapped reads were assembled into 36 255 transcript loci and 50 630 transcripts by running the program Cufflinks v2.1.1. Then, we compared initial assemblies with the annotations of protein-coding genes. Among the initial assemblies, 12 697 transcripts completely matched to introns of known protein-coding genes (classcode " = ") and 18 793 transcripts were divided into unknown or intergenic transcripts (class "u").

Identification and characterization of silk gland lncRNAs

After excluding shorter nucleotide sequences and smaller ORFs (transcripts $<$ 200 nt and ORF $<$ 100 aa), 2015 transcripts were retained to the next pipeline; 537 of 2015 transcripts had protein-coding scores more than zero and were discarded by the software CPC (Kong *et al.*, 2007). Moreover, 187 transcripts were found to have similar domains when searching Pfam, and 684 transcripts hit proteins in the nr database of NCBI by BLASTx. These transcripts were discarded. Then the remained sequences were checked one by one. To identify high-quality sets of silk gland lncRNAs, we removed one transcript with low coverage of mapped reads and two transcripts with lower numbers of mapped reads. In the end, we identified 599 candidate lncRNAs in the silk gland (Table 2 and Table S2), and 409 of 599 candidates are novel compared with recently reported silkworm lncRNAs (Wu *et al.*, 2016; Zhou *et al.*, 2016). About 74.83% of the silk gland lncRNAs identified in this study correspond to unknown or intergenic transcripts (Table 2).

Previous studies have reported that TEs might have taken part in the origin of lncRNAs (Kapusta *et al.*, 2013; Kapusta & Feschotte, 2014). Among the silk gland lncRNAs, 23 lncRNAs overlapped TE-related regions (e-value $< e^{-5}$, coverage $\geq$ 80% and identity $\geq$ 95%) (Table S3 and Fig. S1). Twelve of 23 lncRNAs overlapped DNA transposons and the remaining 11 overlapped RNA transposons.

We estimated the exon numbers, transcript lengths and GC-contents of the silk gland lncRNAs and compared these estimates with the corresponding values of the silkworm protein-coding genes and other species protein-coding genes (Fig. S2). The silk gland lncRNAs have fewer exons than protein-coding genes in the silkworm. This is consistent with previous observations in mouse and other tested species (Sun *et al.*, 2012; Wu *et al.*, 2016; Zhou *et al.*, 2016) (Fig. S2A). In addition, lncRNAs in all species excluding mouse have shorter transcript lengths than the protein-coding genes (Fig. S2B). For the GC-content, compared with the silkworm protein-coding genes, the silk gland lncRNAs have a lower GC-content (47.80% vs. 37.79%), while the value is similar to the level of the silkworm genome (38.80%) (Fig. S2C). In other species, we also observed a similar pattern, indicating that this feature is common for the lncRNAs in different species (Fig. S2C). The lower GC-content in lncRNAs implies that GC-content may have an influence on the protein-coding potential of transcripts (Fickett, 1982). In general, the silk gland lncRNAs have fewer exons, shorter transcript length and lower GC-content than protein-coding genes.

Differentially expressed lncRNAs and experimental verification

The silk yield of the domestic silkworm is much higher than that of wild silkworm because of long-term artificial and breeding selection (Normile, 2009; Fang *et al.*, 2015). Like the differentially expressed genes (DEGs), the DELs

Table 1 Number of RNA sequencing technology reads aligned to the reference genome.

Replicate	Clean reads	First mapped	Second mapped	Merged mapped	Unmapped
D_CH	54 702 284	33 882 924 (61.94%)	4 999 457 (9.14%)	38 882 381 (71.08%)	15 819 903 (28.92%)
D_CY	57 496 084	40 079 037 (69.71%)	4 615 100 (8.03%)	44 694 137 (77.73%)	12 801 947 (22.27%)
W_AKBH	51 956 962	297 91 621 (57.34%)	5 535 309 (10.65%)	35 326 930 (67.99%)	16 630 032 (32.01%)
W_AKSQ	60 689 850	39 116 996 (64.45%)	5 849 869 (9.64%)	44 966 865 (74.09%)	15 722 985 (25.91%)
Total	224 845 180	142 870 578 (63.54%)	20 999 735 (9.34%)	163 870 313 (72.88%)	60 974 867 (27.12%)

Table 2 Category summary of all identified silk gland long non-coding RNAs.

Class code	Transcript no.	Percentage	Description
u	448	74.79%	Unknown, intergenic transcript
x	45	7.51%	Exonic overlap with reference on the opposite strand
i	27	4.51%	A transfrag falling entirely within a reference intron
j	29	4.84%	At least one splice junction is shared with a reference transcript
o	50	8.35%	Generic exonic overlap with a reference transcript
Total	599	100%	Total

between the domestic and wild silkworms may be related to the phenotypic variation of silk yield. In this study, we identified three DELs between the domestic and wild silkworms (Fig. 2A and Table S4). Among these DELs, the expression of one lncRNA (*dw4sg_0483*) was up-regulated in the silk gland of the domestic silkworm, while two lncRNAs (*dw4sg_0040* and *dw4sg_0178*) were up-regulated in wild silkworm.

RT-PCR and qRT-PCR experiments were performed to verify the DELs (Table S1). Both the experiments obtained similar results and confirmed the RNA-seq data (Fig. 2). Expression levels of *dw4sg_0040* and *dw4sg_0178* in the two wild silkworm samples are significantly higher than that in the two domestic silkworm strains ($P =$ 3.243e-5 and $P =$ 5.376e-5), respectively. In contrast, the *dw4sg_0483* was significantly highly expressed in the domestic silkworm compared with the wild silkworm.

Expression patterns of DELs

We investigated the expression patterns of DELs in different tissues and at different developmental stages using RT-PCR. Three surveyed lncRNAs have transcription signals (Fig. 3). The lncRNA *dw4sg_0040* was obviously expressed in silk gland and testis of the domestic silkworm and highly expressed in the 5th instar larvae. The *dw4sg_0483* was highly expressed in all tested tissues and at different developmental stages. The *dw4sg_0178* showed low expression levels in all tested larval tissues but it was highly expressed at pupal and adult stages. The three DELs were expressed not only in silk gland but also in other tissues, indicating their diverse functions. The 5th instar larvae is a vital stage for the development of silk gland (Blaes *et al.*, 1980; Li *et al.*, 2014b). Thus, it will be valuable to know in the future the function of lncRNA *dw4sg_0040* highly expressed in the 5th instar larvae.

Functional annotation of DELs' potential targets and the expression-correlated genes

Previous studies suggested that lncRNAs could regulate the expression of target genes through chromatin remodeling, control of transcription initiation and post-transcriptional processing (Wang & Chang, 2011; Chujo *et al.*, 2016). In order to understand the potential functions of DELs, the putative target genes of the DELs were identified (Table S5). It was found that *dw4sg_0040*, *dw4sg_0178* and *dw4sg_0483* have 24, 32 and five potential target genes, respectively (Table S5). These target genes were used to perform homology search against the nr database in NCBI (Table 3). In total, 29 of 61 potential targets have homologous hits (Table 3). Interestingly,

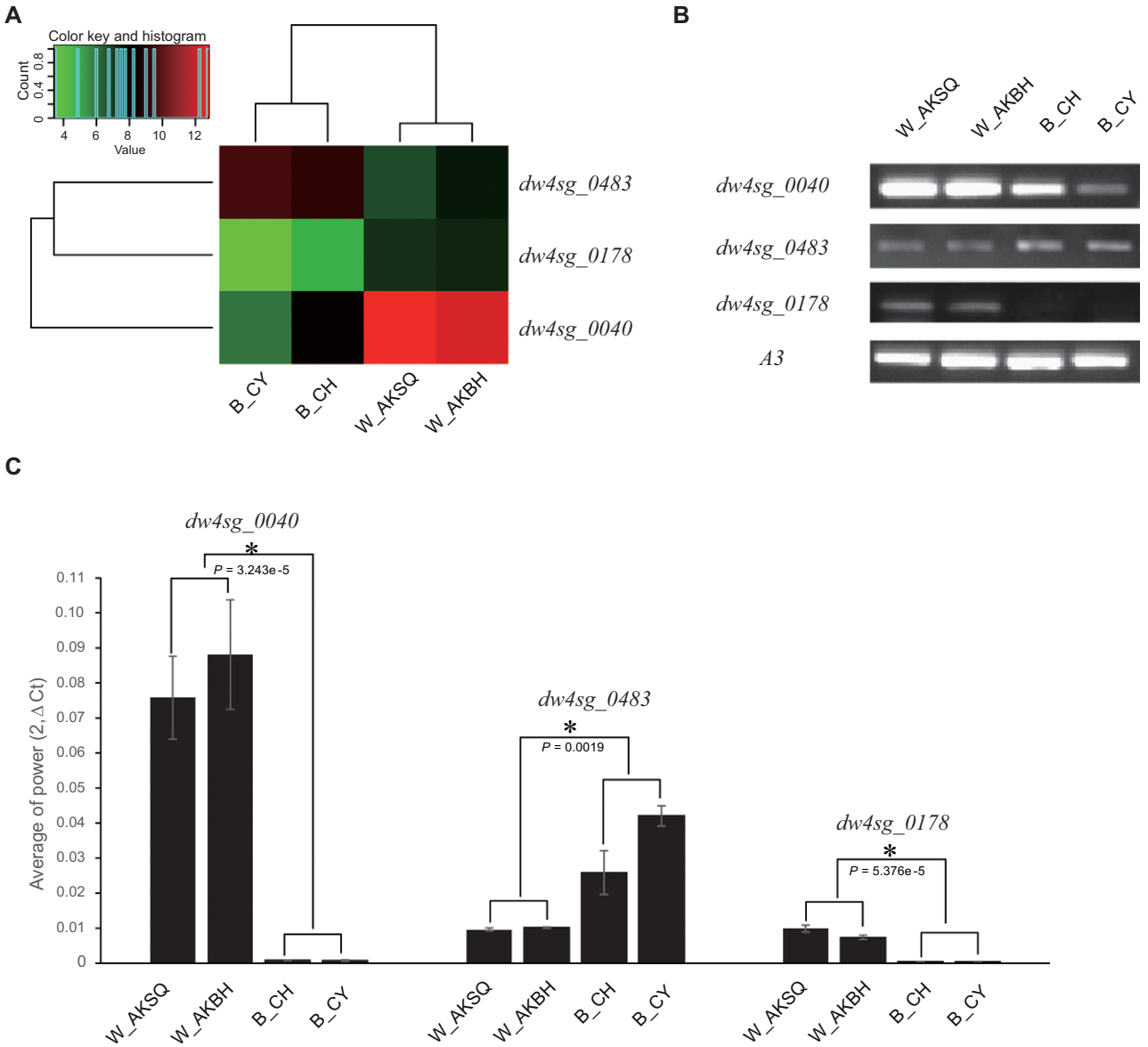

Fig. 2 Differentially expressed long non-coding RNAs (lncRNAs) between domestic and wild silkworms. (A) Differentially expressed lncRNAs of silk glands. Expression levels were measured by RNA sequencing technology (RNA-seq). lncRNAs were clustered by vertical clustering. The samples were clustered by horizontal clustering. (B) Reverse-transcription polymerase chain reaction (RT-PCR) validation of differentially expressed lncRNAs in four samples. The silkworm cytoplasmic actin 3 gene was used as an internal control. (C) Quantitative RT-PCR validation of differentially expressed lncRNAs in four samples.

one potential target (*BGIBMGA004605*) of *dw4sg_0040* is involved in the stress response and protein folding and it resides in the pathway of protein transport in endoplasmic reticulum of the silkworm (Table 3) (Sakano *et al.*, 2006; Moriya *et al.*, 2007).

Using our new transcriptome data of silk glands from day 0 of the 5th instar to wandering stage in domestic and wild silkworms, the expression-correlated networks of DELs were constructed (Fig. 4A and Fig. S3). The results indicated that the expression profiles of 164 genes are significantly correlated with that of *dw4sg_0040* in both domestic and wild silkworms (Pearson correlation coefficient, $r > 0.8$, $P<0.05$). Five genes showed expression profiles significantly correlated with *dw4sg_0040* only in the domestic silkworm, while 191 genes showed similar expression profiles only in the wild silkworm.

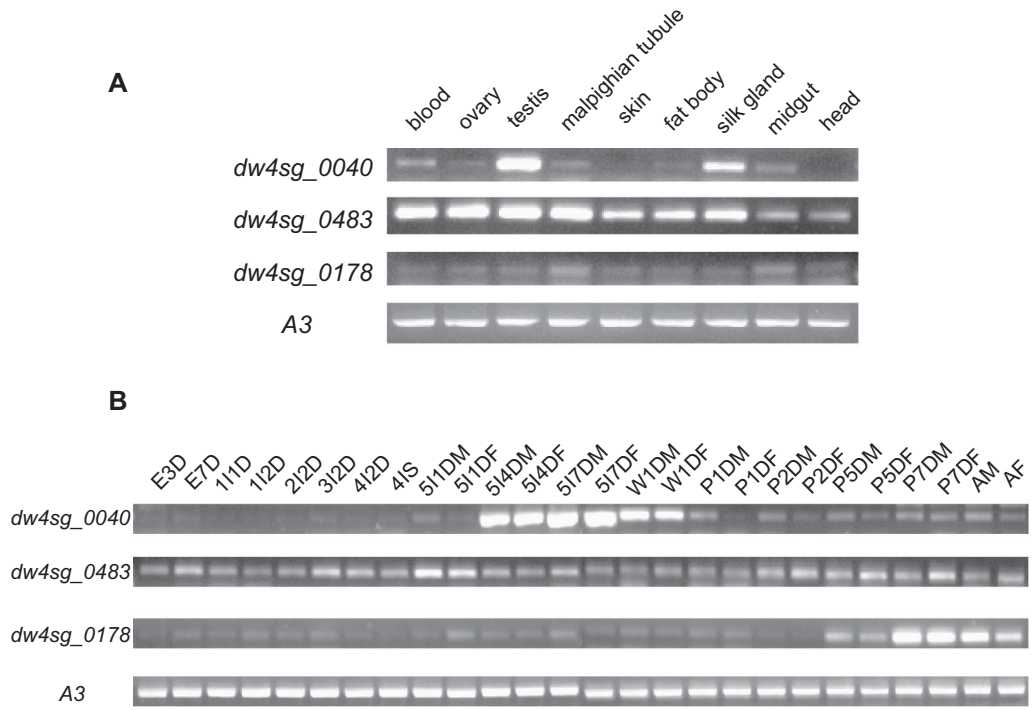

Fig. 3 Expression patterns of differentially expressed long non-coding RNAs (DELs) in nine tissues and at 26 developmental stages. *(*A) Expression patterns of DELs in the nine tissues of *Bombyx mori*. (B) Expression patterns of DELs at silkworm developmental stages. E, egg; I, instar; D, day; S, sleep; M, Male; F, Female; W, Wandering; P, pupa; A, adult, e.g. third day of egg (E3D), first day of the 1st instar (1I1D), sleep of the 4th instar (4IS), first day of the 5th instar in the male (5I1DM).

The 164 expression-correlated genes shared by domestic and wild silkworms were enriched in oxidation-reduction process, heme binding, response to oxidative stress and oxidoreductase activity in GO (Fig. 4B) as well as in fructose and mannose metabolism, longevity regulating and tryptophan metabolism in KEGG (Fig. 4C). It was found that 12 expression-correlated genes with *dw4sg_0178* were shared by domestic and wild silkworms, 55 genes only in the domestic silkworm and one gene only in the wild silkworm. These expression-correlated genes are not enriched in any GO term or KEGG pathway. However, 356 expression-correlated genes with *dw4sg_0483* shared by domestic and wild silkworms are enriched in transfer RNA (tRNA) processing and ribosomal RNA (rRNA) processing in GO as well as in ribosome biogenesis in eukaryotes and aminoacyl−tRNA biosynthesis in KEGG (Fig. 4B, 4C).

Discussion

The silk gland lncRNAs

Although lncRNAs are widely identified in some insect species with the application of RNA-seq technology (Etebari *et al.*, 2015; Jayakodi *et al.*, 2015; Jenkins *et al.*, 2015; Legeai & Derrien, 2015; Xiao *et al.*, 2015; Chen *et al.*, 2016; Wu *et al.*, 2016), the information about lncRNAs in the silkworm is relatively rare. A previous study suggested that nine silk gland-enriched ncRNAs may be involved in the development of silk gland and fibroin synthesis (Li *et al.*, 2014a). However, lengths of these silk gland-enriched ncRNAs are less than 200 bp and they do not belong to lncRNAs. Recently, Wu *et al.* (2016) identintlyfied 11 810 lncRNAs in the domestic silkworms using the RNA-seq data from 21 tissues. Furthermore, they found that eight modules were related to the silk gland. However, the information of the silk gland lncRNAs is obscure. It is well known that the expressions of lncRNAs are often specific to a tissue or a particular developmental stage (Clark & Blackshaw, 2014; Li *et al.*, 2014c; Paralkar *et al.*, 2014; Yang *et al.*, 2014). In this study, therefore, we focused on the identification of lncRNAs in the silk gland as well as the comparison of lncRNAs in silk gland between the domestic and wild silkworms. Using RNA-seq data from silk glands of two domestic silkworm strains and two wild silkworm samples, we identified 599 silk glandlncRNAs (Table 2 and Table S2) and 409 of 599 candidates are novel compared

Table 3 Functional annotation of potential target genes.

lncRNA_id	Target gene_id	Best BLAST hit in nr database; putative function	KEGG pathway	E-value
dw4sg_0040	*BGIBMGA000039*	Antibacterial peptide enbocin 2 precursor; immune response	No hit	6.50E-19
dw4sg_0040	*BGIBMGA000730*	SPARC-related modular calcium-binding protein 1; limb development	No hit	2.01E-17
dw4sg_0040/ dw4sg_0178	*BGIBMGA004590*	Calsyntenin-1; molecular transducer	No hit	1.72E-08
dw4sg_0040	*BGIBMGA004605*	Heat shock protein hsp20.8; stress response and correct folding protein	Protein processing in endoplasmic reticulum, longevity regulating pathway	2.45E-13
dw4sg_0040	*BGIBMGA005109*	Doublesex isoform X2	No hit	4.60E-29
dw4sg_0040	*BGIBMGA005299*	Hypothetical protein 33	No hit	1.24E-34
dw4sg_0040	*BGIBMGA006868*	Uncharacterized protein LOC101742095	No hit	5.03E-14
dw4sg_0040	*BGIBMGA007101*	Uncharacterized protein LOC105842473	No hit	1.32E-20
dw4sg_0040	*BGIBMGA009458*	GDP-Man:Man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase-like; transferase activity	Metabolic pathways, N-glycan biosynthesis	1.75E-26
dw4sg_0040	*BGIBMGA011613*	Uncharacterized protein LOC101743046	No hit	7.08E-27
dw4sg_0040	*BGIBMGA011940*	Hypothetical protein KGM_15136	Ubiquitin-mediated proteolysis	7.93E-11
dw4sg_0040	*BGIBMGA012497*	Uncharacterized protein LOC105841306	No hit	1.83E-29
dw4sg_0040	*BGIBMGA013615*	Xylulose kinase; phosphorylates D-xylulose to produce D-xylulose 5-phosphate	Metabolic pathways, pentose and glucuronate interconversions	1.61E-20
dw4sg_0178	*BGIBMGA000041*	Uncharacterized protein LOC101744079	No hit	2.17E-11
dw4sg_0178	*BGIBMGA000282*	Cuticular protein RR-2 motif 82 precursor	No hit	1.68E-34
dw4sg_0178	*BGIBMGA003798*	Nephrin-like	No hit	4.28E-28
dw4sg_0178	*BGIBMGA004129*	Uncharacterized protein LOC105842239	No hit	4.55E-23
dw4sg_0178	*BGIBMGA004512*	Uncharacterized protein LOC101747207	No hit	1.03E-08
dw4sg_0178	*BGIBMGA005383*	DNA-C5-methyltransferase-1; epigenetic modification	Metabolic pathways, cysteine and methionine	7.96E-22
dw4sg_0178	*BGIBMGA008055*	Katanin p60 ATPase-containing subunit A-like 1	No hit	5.08E-26
dw4sg_0178	*BGIBMGA010108*	CDC42 small effector 2-like protein	No hit	5.77E-18
dw4sg_0178	*BGIBMGA011243*	Acidic fibroblast growth factor intracellular-binding protein	No hit	1.51E-33

(to be continued)

Table 3 Continued.

lncRNA_id	Target gene_id	Best BLAST hit in nr database; putative function	KEGG pathway	E-value
dw4sg_0178	*BGIBMGA011339*	Syntaxin-binding protein 5 isoform X2; molecular transducer	No hit	7.61E-24
dw4sg_0178	*BGIBMGA012433*	Neuroendocrine convertase 1	No hit	9.08E-19
dw4sg_0178	*BGIBMGA013249*	Myotubularin-related protein 3; dephosphorylates phosphatidylinositol 3-phosphate	Metabolic pathways, inositol phosphate metabolism, phosphatidylinositol signaling system	1.46E-24
dw4sg_0178	*BGIBMGA014075*	Uncharacterized protein	Spliceosome	8.11E-24
dw4sg_0483	*BGIBMGA005700*	Epsin-1	Endocytosis	6.15E-165
dw4sg_0483	*BGIBMGA008276*	Anaphase-promoting complex subunit 1-like	Ubiquitin-mediated proteolysis	2.92E-76
dw4sg_0483	*BGIBMGA012444*	Importin-11; nuclear transport receptor	No hit	0

lncRNA, long non-coding RNA; BLAST, Basic Local Alignment Search Tool; KEGG, Kyoto Encyclopedia of Genes and Genomes.

with recently reported silkworm lncRNAs (Wu *et al.*, 2016; Zhou *et al.*, 2016). Nevertheless, the silk gland lncRNAs showed several different genetic features from the protein-coding genes such as fewer exon numbers, shorter transcript length and lower GC-content, consistent with the previous observations (Wu *et al.*, 2016; Zhou *et al.*, 2016). In addition, these results are also consistent with the previously reported lncRNAs in other species, suggesting that these characteristics are common for lncRNAs in most species (Young *et al.*, 2012; Li *et al.*, 2014d; Xie *et al.*, 2014; Wu *et al.*, 2016).

Comparison of silk gland lncRNAs between domestic and wild silkworms

After domestication, silk yield of the domestic silkworm is much higher than that in the wild silkworm (Fang *et al.*, 2015). Silk yield is an important economic trait and a domestication trait. However, little is known about the molecular mechanism of the silk yield variation between domestic and wild silkworms (Tang *et al.*, 2007; Li *et al.*, 2014a; Li *et al.*, 2014b; Wu *et al.*, 2016). The silk gland is a specialized organ to synthesize and secrete silk proteins. Silk fibroin, a major component of silk protein, is composed of a complex ([fibroin heavy chain (*Fib-H*) - fibroin light chain (*Fib-L*)]$_6$ - *P25*) (Inoue *et al.*, 2000; Zhou *et al.*, 2001; Tang *et al.*, 2007). Our previous study showed that there is no difference in expression levels of the genes encoding silk proteins between domestic and wild silkworms (Fang *et al.*, 2015). Nevertheless, the expressions of the genes related to protein secretion were up-regulated in the domestic silkworm (Fang *et al.*, 2015). A previous study on pigs indicated that lncRNAs might have taken part in the formation of the domestication trait (Zhou *et al.*, 2014; Li *et al.*, 2015a). Thus, comparison of lncRNAs in the silk gland between the domestic and wild silkworms would provide some insights into the molecular mechanisms of silk yield improvement in the domestic silkworm.

We performed comparative analysis of lncRNAs in silk glands between domestic and wild silkworms and identified three DELs (*dw4sg_0040*, *dw4sg_0483* and *dw4sg_0178*). Interestingly, two lncRNAs (*dw4sg_0040* and *dw4sg_0178*) were down-regulated in the domestic silkworm. RT-PCR and qRT-PCR confirmed these results (Fig. 2). To further understand the functions of these DELs, we predicted their potential interacting genes based on the minimum free energy joint structure and expression correlation analyses. The potential targets of DELs did not show differential expressions in silk glands between domestic and wild silkworms (Fang *et al.*, 2015). However, eight of 32 DEGs identified in a previous study are also expression-correlated genes with DELs in this study. Five of the eight genes are up-regulated in wild silkworm and related to immune response and one potential target (*BGIBMGA000039*) of *dw4sg_0040* is also related to antibacterial peptide (Table 3). Furthermore, the functional annotation showed that the expression-correlated genes with *dw4sg_0040* shared by domestic and wild

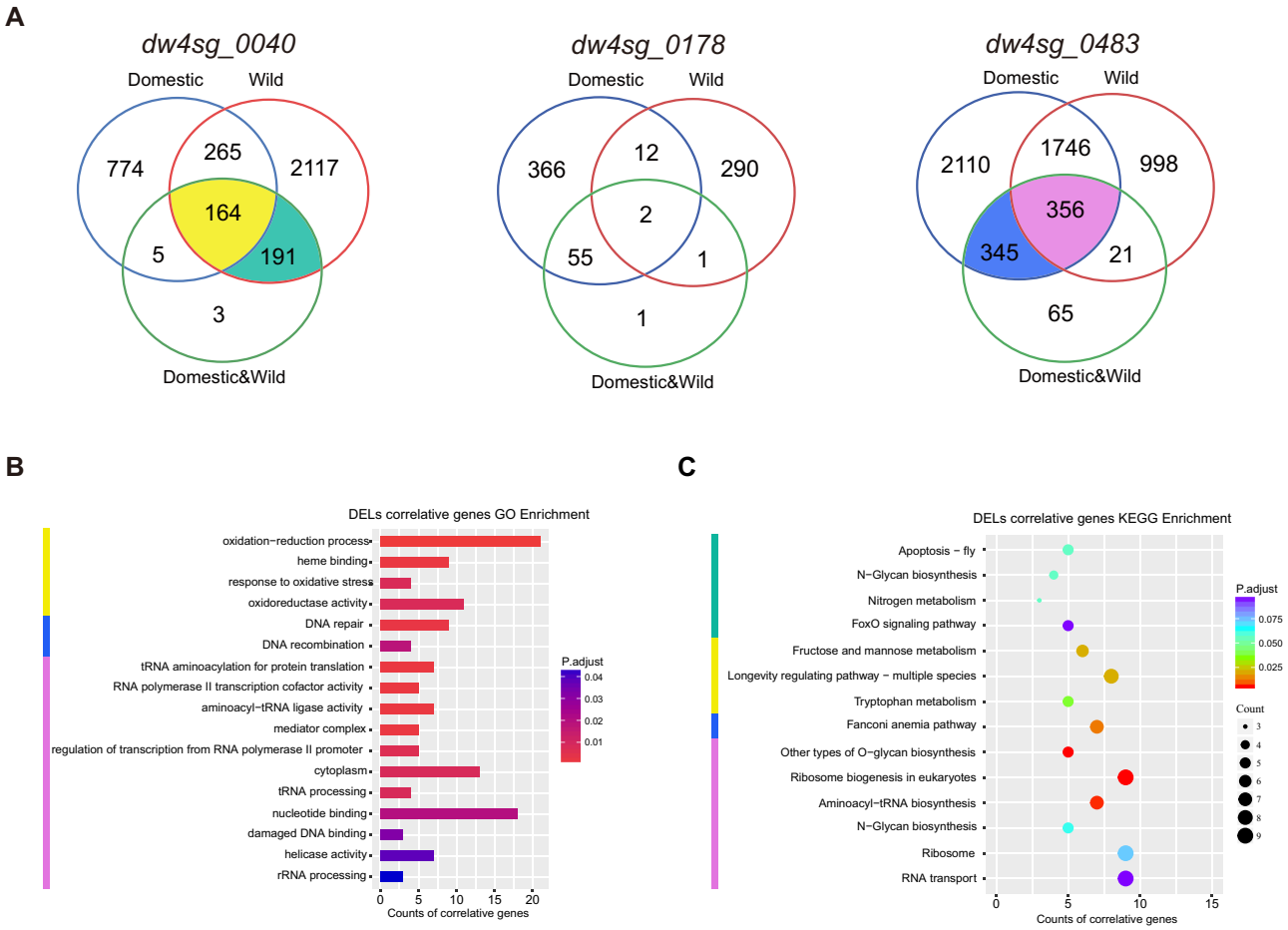

Fig. 4 Functional enrichment of the expression correlated genes of differentially expressed long non-coding RNAs (DELs). (A) Venn diagrams of the expression-correlated genes in the domestic silkworm, wild silkworm and domestic and wild silkworms. Blue circle, domestic; red circle, wild; green circle, domestic and wild. Yellow indicates that the expression-correlated genes with *dw4sg_0040* were shared by domestic and wild silkworms. Turquoise indicates the expression-correlated genes with *dw4sg_0040* only in the wild silkworm. Royal blue indicates the expression-correlated genes with *dw4sg_0483* only in the domestic silkworm. Magenta indicates the expression-correlated genes with *dw4sg_0483* shared by the domestic and wild silkworms. (B) The Gene Ontology (GO) enrichment of the expression-correlated genes of DELs. (C) Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment of the expression correlated-genes of DELs. The meanings of color bars are the same as those in (A).

silkworms are enriched in the oxidation-reduction process and oxidative stress response (Fig. 4B). These results suggested that *dw4sg_0040* may have an impact on the process to resist adverse circumstances in the silkworm, especially in the wild silkworm. Another three genes are up-regulated in the domestic silkworm and are involved in the function of protein export. Moreover, *dw4sg_0483* is up-regulated in the domestic silkworm. The expression-correlated genes with *dw4sg_0483* shared by domestic and wild silkworms are enriched in post-transcriptional regulation such as tRNA processing, rRNA processing, ribosome biogenesis in eukaryotes and aminoacyl−tRNA biosynthesis in GO and KEGG (Fig. 4B, 4C). These processes are related to the translation process of silk protein. Thus, this implies that *dw4sg_0483* may contribute to the phenotypic variation of silk yield between domestic and wild silkworms through post-transcriptional regulation of silk protein.

Our previous study showed that up-regulated genes in the domestic silkworm are involved in secretion of proteins (Fang *et al.*, 2015). Moreover, two potential targets reside in the pathway of ubiquitin-mediated proteolysis on protein degradation (Table 3). The transcription levels of silk-related genes might not be direct reasons for the phenotypic variation. *dw4sg_0040* and *dw4sg_0483* may affect the phenotypic variation in silk yield between domestic and wild silkworms in the post-transcriptional process, including silk protein translation,

protein transport and protein degradation. Further experiments are needed in the future to confirm these speculations.

Acknowledgments

We thank Drs Wei Sun, Min-Jin Han and Ping Fu for their insightful comments and all other members of Zhang's lab for their laboratory assistance. This work was supported by the National Natural Science Foundation of China (No. 31272363 to ZZ) and the Chongqing Graduate Student Research Innovation Project (CYB14041 to QZZ).

Disclosure

The authors declare no conflict of interest.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389–3402.

Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biology*, 11, R106.

Anders, S., Pyl, P.T. and Huber, W. (2014) HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31, 166–169.

Blaes, N., Couble, P. and Prudhomme, J.C. (1980) The programming of silk-gland development in *Bombyx mori*. I. Effects of experimental starvation on growth, silk production, and autolysis during the fifth larval instar studied by electron microscopy. *Cell and Tissue Research*, 213, 311–324.

Cao, J. (2014) The functional role of long non-coding RNAs and epigenetics. *Biological Procedures Online*, 16, 11.

Chen, B., Zhang, Y., Zhang, X., Jia, S.L., Chen, S. and Kang, L. (2016) Genome-wide identification and developmental expression profiling of long noncoding RNAs during *Drosophila metamorphosis*. *Scientific Reports*, 6, 23330.

Chen, J.H., Quan, M.Y. and Zhang, D.Q. (2014) Genome-wide identification of novel long non-coding RNAs in *Populus tomentosa* tension wood, opposite wood and normal wood xylem by RNA-seq. *Planta*, 241, 125–143.

Chujo, T., Yamazaki, T. and Hirose, T. (2016) Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochimica et Biophysica Acta*, 1859, 139–146.

Clark, B.S. and Blackshaw, S. (2014) Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease. *Frontiers in Genetics*, 5, 164.

Duan, J., Li, R.Q, Cheng, D.J., Fan, W., Zha, X.F., Cheng, T.C., Wu, Y.Q., Wang, J., Mita, K., Xiang, Z.H. and Xia, Q.Y. (2010) SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology. *Nucleic Acids Research*, 38, D453–D456.

Etebari, K., Furlong, M.J. and Asgari, S. (2015) Genome wide discovery of long intergenic non-coding RNAs in diamondback moth (*Plutella xylostella*) and their expression in insecticide resistant strains. *Scientific Reports*, 5, 14642.

Fang, S.M., Hu, B.L., Zhou, Q.Z., Yu, Q.Y. and Zhang, Z. (2015) Comparative analysis of the silk gland transcriptomes between the domestic and wild silkworms. *BMC Genomics*, 16, 60.

Fickett, J.W. (1982) Recognition of protein coding regions in DNA-sequences. *Nucleic Acids Research*, 10, 5303–5318.

Ilott, N.E. and Ponting, C.P. (2013) Predicting long non-coding RNAs using RNA sequencing. *Methods*, 63, 50–59.

Inoue, S., Tanaka, K., Arisaka, F., Kimura, S., Ohtomo, K. and Mizuno, S. (2000) Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and *P25*, with a 6:6:1 molar ratio. *The Journal of Biological Chemistry*, 275, 40517–40528.

Jayakodi, M., Jung, J.W., Park, D., Ahn, Y.J., Lee, S.C., Shin, S.Y., Shin, C., Yang, T.J. and Kwon, H.W. (2015) Genome-wide characterization of long intergenic non-coding RNAs (lincRNAs) provides new insight into viral diseases in honey bees *Apis cerana* and *Apis mellifera*. *BMC Genomics*, 16, 680.

Jenkins, A.M., Waterhouse, R.M. and Muskavitch, M.A. (2015) Long non-coding RNA discovery across the genus *anopheles* reveals conserved secondary structures within and beyond the Gambiae complex. *BMC Genomics*, 16, 337.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W.Z., Mcanulla, C., Mcwilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R. and Hunter, S. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236–1240.

Kapusta, A. and Feschotte, C. (2014) Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends in Genetics*, 30, 439–452.

Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M. and Feschotte, C. (2013) Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genetics*, 9, e1003470.

Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L. and Gao, G. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, 35, W345–W349.

Kornienko, A.E., Guenzl, P.M., Barlow, D.P. and Pauler, F.M. (2013) Gene regulation by the act of long non-coding RNA transcription. *BMC Biology*, 11, 59.

Legeai, F. and Derrien, T. (2015) Identification of long non-coding RNAs in insects genomes. *Current Opinion in Insect Science*, 7, 37–44.

Li, A., Zhang, J., Zhou, Z., Wang, L., Sun, X. and Liu, Y. (2015a) Genome-scale identification of miRNA-mRNA and miRNA-lncRNA interactions in domestic animals. *Animal Genetics*, 46, 716–719.

Li, D.D., Liu, Z.C., Huang, L., Jiang, Q.L., Zhang, K., Qiao, H.L., Jiao, Z.J., Yao, L.G., Liu, R.Y. and Kan, Y.C. (2014a) The expression analysis of silk gland-enriched intermediate-size non-coding RNAs in silkworm *Bombyx mori*. *Insect Science*, 21, 429–438.

Li, J.S., Cai, Y.M., Ye, L.P., Wang, S.H., Che, J.Q., You, Z.Y., Yu, J. and Zhong, B.X. (2014b) MicroRNA expression profiling of the fifth-instar posterior silk gland of *Bombyx mori*. *BMC Genomics*, 15, 410.

Li, J., Chen, Z., Tian, L., Zhou, C., He, M.Y., Gao, Y., Wang, S., Zhou, F., Shi, S., Feng, X. *et al*. (2014c) LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut*, 63, 1700–1710.

Li, J., Ma, W., Zeng, P., Wang, J., Geng, B., Yang, J. and Cui, Q. (2015b) LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Briefings in Bioinformatics*, 16, 806–812.

Li, L., Eichten, S.R., Shimizu, R., Petsch, K., Yeh, C.T., Wu, W., Chettoor, A.M., Givan, S.A., Cole, R.A. and Fowler, J. *et al*. (2014d) Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biology*, 15, R40.

Liao, Q., Shen, J., Liu, J.F., Sun, X., Zhao, G.G., Chang, Y.Z., Xu, L.T., Li, X.R., Zhao, Y. and Zheng, H.Q. *et al*. (2014) Genome-wide identification and functional annotation of *Plasmodium falciparum* long noncoding RNAs from RNA-seq data. *Parasitology Research*, 113, 1269–1281.

Ma, L., Xu, H., Zhu, J., Ma, S., Liu, Y., Jiang, R.J., Xia, Q. and Li, S. (2011) *Ras1*^{CA} overexpression in the posterior silk gland improves silk yield. *Cell Research*, 21, 934–943.

Mattick, J.S. and Rinn, J.L. (2015) Discovery and annotation of long noncoding RNAs. *Nature Structural & Molecular Biology*, 22, 5–7.

Meller, V.H., Wu, K.H., Roman, G., Kuroda, M.I. and Davis, R.L. (1997) *roX1* RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell*, 88, 445–457.

Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long noncoding RNAs: insights into functions. *Nature Reviews Genetics*, 10, 155–159.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C. and Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, 35, W182–W185.

Morris, K.V. and Mattick, J.S. (2014) The rise of regulatory RNA. *Nature Reviews Genetics*, 15, 423–437.

Nam, J.W. and Bartel, D.P. (2012) Long noncoding RNAs in *C. elegans*. *Genome Research*, 22, 2529–2540.

Normile, D. (2009) Sequencing 40 silkworm genomes unravels history of cultivation. *Science*, 325, 1058–1059.

Paralkar, V.R., Mishra, T., Luan, J., Yao, Y., Kossenkov, A.V., Anderson, S.M., Dunagin, M., Pimkin, M., Gore, M., Sun, D., Konuthula, N., Raj, A., An, X.L., Mohandas, N., Bodine, D.M., Hardison, R.C., and Weiss, M.J. (2014) Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. *Blood*, 123, 1927–1937.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A. and Schier, A.F. (2012) Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Research*, 22, 577–591.

Sakano, D., Li, B., Xia, Q., Yamamoto, K., Banno, Y., Fujii, H. and Aso, Y. (2006) Genes encoding small heat shock proteins of the silkworm, *Bombyx mori*. *Bioscience, Biotechnology, and Biochemistry*, 70, 2443–2450.

Soshnev, A.A., Ishimoto, H., Mcallister, B.F., Li, X., Wehling, M.D., Kitamoto, T. and Geyer, P.K. (2011) A conserved long noncoding RNA affects sleep behavior in *Drosophila*. *Genetics*, 189, 455–468.

Stanojcic, S., Gimenez, S., Permal, E., Cousserans, F., Quesneville, H., Fournier, P. and D'alencon, E. (2011) Correlation of LNCR rasiRNAs expression with heterochromatin formation during development of the holocentric insect *Spodoptera frugiperda*. *PLoS ONE*, 6, e24746.

Suetsugu, Y., Futahashi, R., Kanamori, H., Kadono-Okuda, K., Sasanuma, S., Narukawa, J., Ajimura, M., Jouraku, A., Namiki, N., Shimomura, M., Sezutsu, H., Osanai-Futahashi, M., Suzuki, M.G., Daimon, T., Shinoda, T., Taniai, K., Asaoka, K., Niwa, R., Kawaoka, S., Katsuma, S., Tamura, T., Noda, H., Kasahara, M., Sugano, S., Suzuki, Y., Fujiwara, H., Kataoka, H., Arunkumar, K.P., Tomar, A., Nagaraju, J., Goldsmith, M.R., Feng, Q.L., Xia, Q.Y., Yamamoto, K., Shimada, T. and Mita, K. (2013) Large scale full-length cDNA sequencing reveals a unique genomic landscape in a lepidopteran model insect, *Bombyx mori*. *G3-Genes Genomes Genetics*, 3, 1481–1492.

Sun, L., Zhang, Z., Bailey, T.L., Perkins, A.C., Tallack, M.R., Xu, Z. and Liu, H. (2012) Prediction of novel long non-coding RNAs based on RNA-Seq data of mouse *Klf1* knockout study. *BMC Bioinformatics*, 13, 331.

Takasu, Y., Hata, T., Uchino, K. and Zhang, Q. (2010) Identification of *Ser2* proteins as major sericin components in the non-cocoon silk of *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, 40, 339–344.

Tang, J., Li, W.F., Zhang, X. and Zhou, C.Z. (2007) The gene expression profile of *Bombyx mori* silkgland. *Gene*, 396, 369–372.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L.

(2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7, 562–578.

Wang, K.C. and Chang, H.Y. (2011) Molecular mechanisms of long noncoding RNAs. *Molecular Cell*, 43, 904–914.

Wu, Y.Q., Cheng, T.C., Liu, C., Liu, D.L., Zhang, Q., Long, R.W., Zhao, P. and Xia, Q.Y. (2016) Systematic identification and characterization of long non-coding RNAs in the silkworm, *Bombyx mori. PLoS ONE*, 11, e0147147.

Xia, Q.Y., Zhou, Z.Y., Lu, C., Cheng, D.J., Dai, F.Y., Li, B., Zhao, P., Zha, X.F., Cheng, T.C., Chai, C.L., Pan, G.Q., Xu, J.S., Liu, C., Lin, Y., Qian, J.F., Hou, Y., Wu, Z.L., Li, G.R., Pan, M.H., Li, C.F., Shen, Y.H., Lan, X.Q., Yuan, L.W., Li, T., Xu, H.F., Yang, G.W., Wan, Y.J., Zhu, Y., Yu, M.D., Shen, W.D., Wu, D.Y., Xiang, Z.H., Yu, J., Wang, J., Li, R.Q., Shi, J.P., Li, H., Li, G.Y., Su, J.N., Wang, X.L., Li, G.Q., Zhang, Z.J., Wu, Q.F., Li, J., Zhang, Q.P., Wei, N., Xu, J.Z., Sun, H.B., Dong, L., Liu, D.Y., Zhao, S.L., Zhao, X.L., Meng, Q.S., Lan, F.D., Huang, X.G., Li, Y.Z., Fang, L., Li, C.F., Li, D.W., Sun, Y.Q., Zhang, Z.P., Yang, Z., Huang, Y.Q., Xi, Y., Qi, Q.H., He, D.D., Huang, H.Y., Zhang, X.W., Wang, Z.Q., Li, W.J., Cao, Y.Z., Yu, Y.P., Yu, H., Li, J.H., Ye, J.H., Chen, H., Zhou, Y., Liu, B., Wang, J., Ye, J., Ji, H., Li, S.T., Ni, P.X., Zhang, J.G., Zhang, Y., Zheng, H.K., Mao, B.Y., Wang, W., Ye, C., Li, S.G., Wang, J., Wong, G.K.S. and Yang, H.M. (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*, 306, 1937–1940.

Xiao, H.M., Yuan, Z.T., Guo, D.H., Hou, B.F., Yin, C.L., Zhang, W.Q. and Li, F. (2015) Genome-wide identification of long noncoding RNA genes and their potential association with fecundity and virulence in rice brown planthopper, *Nilaparvata lugens*. *BMC Genomics*, 16, 749.

Xie, C., Mao, X.Z., Huang, J.J., Ding, Y., Wu, J.M., Dong, S., Kong, L., Gao, G., Li, C.Y. and Wei, L.P. (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, 39, W316–W322.

Xie, C.Y., Yuan, J., Li, H., Li, M., Zhao, G.G., Bu, D.C., Zhu, W.M., Wu, W., Chen, R.S. and Zhao, Y. (2014) NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Research*, 42, D98–D103.

Yang, X.F., Gao, L., Guo, X.L., Shi, X.H., Wu, H., Song, F. and Wang, B.B. (2014) A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases. *PLoS ONE*, 9, e87797.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T.L. (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134.

Young, R.S., Marques, A.C., Tibbit, C., Haerty, W., Bassett, A.R., Liu, J.L. and Ponting, C.P. (2012) Identification and properties of 1,119 candidate lincRNA loci in the *Drosophila melanogaster* genome. *Genome Biology and Evolution*, 4, 427–442.

Zhou, C.Z., Confalonieri, F., Jacquet, M., Perasso, R., Li, Z.G. and Janin, J. (2001) Silk fibroin: structural implications of a remarkable amino acid sequence. *Proteins*, 44, 119–122.

Zhou, Q.Z., Zhang, B., Yu, Q.Y. and Zhang, Z. (2016) BmncRNAdb: a comprehensive database of non-coding RNAs in the silkworm, *Bombyx mori*. *BMC Bioinformatics*, 17, 370.

Zhou, Z.Y., Li, A.M., Adeola, A.C., Liu, Y.H., Irwin, D.M., Xie, H.B. and Zhang, Y.P. (2014) Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs. *Genome Biology and Evolution*, 6, 1387–1392.

Zhu, J.J., Fu, H.J., Wu, Y.G. and Zheng, X.F. (2013) Function of lncRNAs and approaches to lncRNA-protein interactions. *Science China Life Sciences*, 56, 876–885.

Zucchelli, S., Patrucco, L., Persichetti, F., Gustincich, S. and Cotella, D. (2016) Engineering translation in mammalian cell factories to increase protein yield: The unexpected use of long non-coding SINEUP RNAs. *Computational and Structural Biotechnology Journal*, 14, 404–410.

Manuscript received September 8, 2016
Final version received December 3, 2016
Accepted December 29, 2016

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1 An example of high similarity between silk gland lncRNAs and transposable elements.

Fig. S2 Characteristics of the silk gland lncRNAs in the silkworm.

Fig. S3 Expression regulation networks of the genes correlated with DELs.

Table S1 The primer sequences of *dw4sg0040, dw4sg0178, dw4sg0483* and *A3* used for RT-PCR and qRT-PCR.

Table S2 The GTF annotations of all identified silk gland lncRNAs in the silkworm.

Table S3 The silk gland lncRNAs with high sequence similarity to transposable elements.

Table S4 Detailed expression levels of differentially expressed lncRNAs between the domestic and wild silkworms.

Table S5 Potential target genes of differentially expressed lncRNAs.