A comparison of co-expression networks in silk gland reveals the causes of silk yield increase during silkworm domestication

Qiu-Zhong Zhou

Chongqing University

Ping Fu

Chongqing University

Shu-Shang Li

Chongqing University

Chang-Jiang Zhang

Chongqing University

Quan-You Yu

Chongqing University

Chuan-Zhen Qiu

Chongqing University

Hong-Bo Zhang

Chongqing University

Ze Zhang (✉ zezhang@cqu.edu.cn)

Chongqing University

Research article

Keywords: silkworm, domestication, silk gland, silk yield, transcriptome, co-expression

Posted Date: August 13th, 2019

DOI: <https://doi.org/10.21203/rs.2.12747/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Frontiers in Genetics on March 27th, 2020.
See the published version at <https://doi.org/10.3389/fgene.2020.00225>.

Abstract

Background Long-term domestication and breeding selection have led to that silk yield of the domestic silkworm (*Bombyx mori*) have increased several times higher than that of its wild ancestor (*B. mandarina*). However, little is known about the molecular mechanisms of silk yield increase during silkworm domestication. **Results** Based on dynamic patterns of functional divergence in silk gland between the domestic and wild silkworms, we found that at early and intermediate stages of silk gland development, the up-regulated genes of the domestic silkworm mainly include DNA integration, nucleic acid binding and transporter activity, which is related to division and growth of cell. This has led to that posterior silk gland (PSG) of the domestic silkworm has significantly more cells (“factories” of fibroin protein synthesis) than that of wild silkworm. At the late stage of silk gland development, the up-regulated genes in the domestic silkworm are enriched in protein processing and ribosome pathways, suggesting that protein synthesis efficiency is greatly improved during silkworm domestication. The synthetic capacity of fibroin proteins of the domestic silkworm has increased, however, the synthetic capacity of sericin proteins has decreased relative to wild silkworm. This reflects that the domestic and wild silkworms have been under different selection pressures. Importantly, we found that the network co-expressed with silk-coding genes for the domestic silkworm is much larger than that for wild silkworm. Furthermore, much more genes co-expressed with silk-coding genes in the domestic silkworm have been subject to artificial selection than those in wild silkworm. **Conclusion** Our results revealed that increase in silk yield during silkworm domestication is involved in improvement of a biological system which includes not only expansion of “factories” (cells of PSG) of protein synthesis but also high expression of silk-coding genes as well as silk production related genes such as biological energy, transport, and ribosome pathway genes. In addition, a combination of comparative multi-omics and dynamic network biological method used in this study provides a methodology reference for investigation of molecular mechanisms of formation of a complex trait for other domestic animals.

Background

The domestic silkworm (*Bombyx mori*) was domesticated from Chinese wild silkworm (*B. mandarina*) about 5000 years ago [1–3]. As an economical insect, silk produced by the domestic silkworm is an important material for not only textiles and industrial application but also biomaterials and cosmetics [4]. Sericulture is still one of income sources for farmers in some developing countries, especially China and India. Silk, which is composed of fibroin and sericin proteins, is synthesized in the silk gland, a silk-producing organ of the silkworm [5]. Because of the economic importance, the composition and genetic basis of silk in the domestic silkworm have been extensively studied up to now.

Silk fibroin consists of the fibroin heavy chain (*Fib-H*), fibroin light chain (*Fib-L*) and the 25-kD polypeptide proteins (*P25*) with a molar ratio of 6:6:1 [6], which is synthesized in the posterior silk gland (PSG), one of three specialized compartments of the gland. Sericin synthesized in middle silk gland (MSG) is mainly composed of a variety of glue proteins such as *sericin1*, *sericin2*, and *sericin3*. During the past few decades, the genes encoding silk fibroin (*Fib-H*, *Fib-L*, and *P25*) and sericin (*Sericin1*,

Sericin2, and *Sericin3*) have been identified and cloned [7–11]. The genetic loci underlying the variation of silk yield among the strains of the domestic silkworm have been mapped although the candidate genes responsible for silk yield remain to be identified [12–15]. Later, the comparative analysis of transcriptomes among the domestic silkworm strains with different silk yields was used to identify the genes associated with silk yield [16, 17]. It was found that the differential expression genes between the strains with different silk yields were mainly associated with the processing and biosynthesis of proteins [16], and silk gland development or protein synthesis [17]. Although these studies have provided some insight into the genetic basis of silk production, no gene that regulates silk yield in silkworm has been functionally verified. This implies that the molecular mechanisms underlying the silk production may be much more complex than thought before.

Compared with its wild ancestor, long-term artificial breeding and selection have led to the domestic silkworm having very different phenotypes. Silk yield of domestic silkworm is 3–5 times of silk yield higher than that of wild silkworm [18]. Thus, understanding the molecular mechanisms of difference in silk yield between the domestic and wild silkworms is important for not only improvement of complex traits but also evolutionary biology. A transcriptome comparison of silk gland of the day 3 of fifth-instar larvae between the domestic and wild silkworms revealed sixteen up-regulated genes in the domestic silkworm, which were related to secretion of proteins, tissue development and metabolism [18]. In addition, a shotgun proteomics approach with label-free quantification analysis was used for comparing proteomics of PSGs between the domestic and wild silkworms, and 50 differentially expressed proteins were identified [19]. Our recent study also revealed that there is a big difference in the abundance of sericin proteins in cocoon between the domestic and wild silkworms [20]. However, these comparisons of the transcriptomes or proteomes were only based on one stage or two stages of silk gland development or cocoon [18–20]. The dynamics of divergence of silk protein synthesis and regulation in silk gland between domestic and wild silkworms remains to be investigated.

In this study, we first obtain time-series transcriptome data during whole silk gland development for both the domestic and wild silkworms. Then, we investigate dynamic patterns of functional divergence in silk gland between the domestic and wild silkworms. Our results provide some new insights into the molecular mechanisms of silk yield increase during silkworm domestication.

Results

Assembly of silk gland transcriptomes

For either male or female, the weights of larvae, whole cocoon, pupa, and cocoon shell in the domestic silkworm (Xiafang strain, XF) are all significantly higher than those in wild silkworm, respectively (Fig. S1). Especially, the cocoon shell weight (CSW) of the domestic silkworm is about ten-fold of CSW of wild silkworm. The silk gland is considered as the most important organ for silk protein synthesis [4]. To find if there is any transcriptional divergence during silk gland development between the domestic and wild silkworms, we carried out the RNA-seq at seven developmental stages of silk gland and comprehensively

characterized the gene expression dynamics in silk gland of the domestic and wild silkworms. After removing the low-quality reads, 559,843,084 and 545,393,238 clean reads are obtained from the domestic and wild silkworms, respectively (Table S1). The clean data are mapped to the silkworm genome and then assembled into the 29,691 gene loci using the tools Tophat2-Cufflinks [21]. And then we assessed the gene expression levels based on the uniquely mapped reads of gene loci. The gene expression patterns are similar for both silkworms during silk gland development (Fig. S2). More than 67.5% of genes have low expression levels ($0 < \text{FPKM} < 1$) in the domestic and wild silkworms (Table S2). Only 95~125 and 101~127 genes are considered as the super high expression (>1000 FPKM) for both silkworms, respectively. Specifically, both sericin and fibroin genes are super highly expressed genes in the domestic and wild silkworms, indicating that the function of silk gland is largely conserved between the domestic and wild silkworms.[18]

Divergence of transcriptional level during silk gland development between the domestic and wild silkworms

To compare the dynamic process of silk gland development, we identified differentially expressed genes (DEGs) at seven developmental stages between the domestic and wild silkworms (Fig. 1). Finally, 1,282, 1,149, 1,215, 1,291, 1,076, 1,139 and 1,264 genes are identified as DEGs between the domestic and wild silkworms in the time points of 0p, 1p, 2p, 3p, 4p, 5p and wandering stage (w), respectively (Table S3). The up-regulated genes of the domestic silkworm during silk gland development are prominently enrichment in the functions of DNA integration, nucleic acid binding and transporter activity (Fig. 2A). The genes related to transporter activity such as BGIBMGA004507, BGIBMGA004510, BGIBMGA012890, and XLOC_018800) are up-regulated in the domestic silkworm at least in five time points (Fig. S3A). Furthermore, the pathway of ECM-receptor interaction is strongly associated with the up-regulated genes which are mediated by transmembrane molecules such as integrins and proteoglycans (Fig. S3B). The integrins and proteoglycans such as BGIBMGA000915, BGIBMGA001498, and BGIBMGA002430 play important roles in the control of cellular activities and neurotransmitter release, which are higher expression in the domestic silkworm at most of the developmental stages of silk gland (Fig. S3B). This suggests that the cells of silk gland in the domestic silkworm perform more frequent exchange of biomolecule and signal. In the late stage, the up-regulated genes in the domestic silkworm display a strong enrichment for the genes related to protein processing in endoplasmic reticulum and ribosome (Fig. 2A and Fig. S4A). The down-regulated genes are assigned to oxidation-reduction, oxidoreductase activity and metabolic pathways (Fig. 2B and Fig. S4B), which are related to antioxidant systems and consistent with the previous report [18].

As a complementary method to further examine the functional shift between the domestic and wild silkworm, we identified the functional term which is more or less enrichment in all time points (Fig. 2C and Fig. S5). The translational initiation and ribosome pathway are more enrichment in the up-regulated gene at all time points (Fig. 2C). Most of the ribosome genes are highly expressed in the domestic silkworm especially at the late stages of silk gland development (Fig. S6). This indicates that protein synthesis is more active in the domestic silkworm than wild silkworm. More activity of protein synthesis

will need more biological energy [22]. Interestingly, the electron transport chain shows more enrichment in the up-regulated genes during the silk gland development. The mitochondrial genes such as *Mt_ATPase6*, *Mt_COI* and *Mt_COII* are all up-regulated expression in silk gland of the domestic silkworm (Fig. 2D).

Expression patterns of silk-coding genes during silk gland development in the domestic and wild silkworms

Given that more activity of protein synthesis needs more abundance transcripts of silk-coding genes, from 0p to w stages, we found that the expression levels of fibroin genes in the domestic silkworm are higher than those of wild silkworms at almost seven stages, which are confirmed by the qPCR (Fig. 3A, 3B, 3C and Fig. S7). This indicates that the silk gland generates more transcripts of fibroin to provide the RNA template for the high-efficient proteins synthesis in the domestic silkworm. However, the expression patterns of sericin genes are different from fibroin genes (Fig. 3D, 3E and 3F). The expression levels of *Sericin1* and *Sericin3* genes increased as developmental stage of the silk gland, especially reached to very high levels at the late stage of silk gland development in wild silkworm. In contrast, the expression level of *Sericin2* gene decreased as developmental stage of the silk gland, and the level in the domestic silkworm is lower than that in wild silkworm. These results imply that the silk gland of the domestic silkworm produces more fibroins but less sericins relative to wild silkworm. To further confirm these results, we inspected the microstructural morphology of cross-section of silk fibers by scanning electron microscopy. The microstructural morphology of silk fibers of the domestic silkworm exhibits larger fibroins than wild silkworm in the outer, middle and inner layers of cocoon silk (Fig. 3G, 3H, 3J and 3K). However, in the inner layer of cocoon silk fiber, which are produced at the late stage, the fibroins are surrounded by more sericins in wild silkworm (Fig. 3I and 3L). This is consistent with the expression patterns of sericins during the silk gland development. Moreover, our previous study also found that the cocoon of the domestic silkworm contained more silk fibroins but less sericins than wild silkworm cocoon based on the comparative proteomics approach [20]. These results indicated that the synthetic capacity of fibroin proteins has increased but the synthetic capacity of sericin proteins has decreased during silkworm domestication.

Co-expression network analysis of the silk-coding genes in the domestic and wild silkworms

We have identified 3,453 DEGs at least at one stage and found that fibroin genes have higher expression levels in the domestic silkworm than those in wild silkworm. To look at which DEGs are closely co-expressed with silk-coding genes, we performed co-expression analysis. A total of 32 consistent modules are detected in the domestic and wild silkworms. We found that the genes in the lightsteelblue module are enriched in the functions including protein folding, biological energy, ribosome, and RNA transport (Fig. 4A and 4B). These functions are overlapped with the functional divergence in the silk gland between the domestic and wild silkworms (Fig. 2). Interestingly, fibroin genes (*Fib-H*, *Fib-L*, and *P25*) and sericin genes (*Sericin1* and *Sericin2*) are also found in the lightsteelblue module. In the lightsteelblue module of the domestic silkworm, we detected 400 DEGs co-expressed with silk-encoding genes, such as ribosome (BGIBMGA008335 and BGIBMGA006919), RNA transport (BGIBMGA005438 and BGIBMGA001699) and

oxidative phosphorylation (BGIBMGA007211) related genes, which were up-regulated in the domestic silkworm (Table S3 and Table S4). For wild silkworm, we detected 258 DEGs co-expressed with silk-coding genes. Some of them (BGIBMGA013791 and BGIBMGA000029) are up-regulated in wild silkworm, which is related to ribosome and RNA transport. After removing the genes that show no difference in expression level between the domestic and wild silkworms, co-expression networks of silk-coding genes and DEGs are constructed for the domestic and wild silkworms, respectively (Fig. 4C and 4D).

For the domestic silkworm, 142 DEGs are co-expressed with fibroin gene *Fib-H*, in which 53 genes are only co-expressed with *Fib-H* (Fig. 4C). 119 DEGs are co-expressed with *Sericin1* in the domestic silkworm and most of them are common co-expressed with *Fib-H* (Fig. 4C). *Fib-L*, *P25*, and *Sericin2* are co-expressed with 29, 50 and 28 DEGs, respectively. Most of DEGs co-expressed with *Fib-L*, *P25* and *Sericin1* are shared by *Fib-H* and *Sericin1*. Moreover, the DEGs in the network are associated with the functions including the ATP-binding (e.g. BGIBMGA009554 and BGIBMGA008085), transport (e.g. BGIBMGA009170 and BGIBMGA001462) and ribosome (e.g. BGIBMGA006919 and BGIBMGA000074) (Fig. 4C). Especially, two protein-coding genes (BGIBMGA000074 and BGIBMGA013169) related to ribosome are co-expressed with *Fib-H*. Strikingly, the topology of the network for wild silkworm is quite different from that for the domestic silkworm (Fig. 4 C vs D) and the network for wild silkworm is much smaller than that for the domestic silkworm. Only 19 DEGs are co-expressed with *Fib-H* and there is no ribosome gene that connects to the *Fib-H* in wild silkworm (Fig. 4D). Wild silkworm *Fib-L* and *P25* also have fewer connections compared with the domestic silkworm counterparts (Fig. 4D). In addition, some co-expressed genes shared by silk-coding genes are involved in the ATP-binding, transport, and ribosome in wild silkworm, but the number is smaller than that in the domestic silkworm (Fig. 4C and 4D). These results indicate that divergence in biological energy, transport, and ribosome pathways might have led to the difference in silk yield between the domestic and wild silkworms.

To further survey the influence of domestication, the whole-genome resequenced data of the domestic and wild silkworm populations are used to identify genomic regions under artificial selection. After removing the genes without selection signature from the co-expression network, two small co-expression networks of silk-coding genes are constructed in the domestic and wild silkworms, respectively (Fig. 5A and 5B). Interestingly, three genes with selection signature (BGIBMGA000074, BGIBMGA012537, and BGIBMGA006919) related to ribosome biogenesis are found in the network of the domestic silkworm (Fig. 5A). BGIBMGA000074 is especially co-expressed with *Fib-H*. The nucleotide polymorphisms and population differentiation of BGIBMGA000074 are shown in Fig. 5C. BGIBMGA006919 is shared by *Fib-L* and *Sericin1*. BGIBMGA012537 is especially co-expressed with the *Sericin1*. However, the ribosome genes were not detected in the small network of wild silkworm (Fig. 5B). These results suggest that ribosome genes may be subject to artificial selection during silkworm domestication. Besides the ribosome genes, we also found that BGIBMGA001462 related to transport function is co-expressed with *Fib-H* in the domestic silkworm (Fig. 5A), which is specifically expressed in the silk gland (Fig. 5E). *Fib-H* is the largest protein molecule in the fibroin and is the most important gene for silk production.^[11] Strikingly, we found that *Fib-H* in the network of the domestic silkworm has much more connections than *Fib-H* of wild silkworm and *Sericin1* also shows a similar pattern (Fig. 5A, 5B and 5D). This indicates that artificial

selection during silkworm domestication might have directly acted not only on the silk-coding genes but also on silk production related genes such as biological energy, transport, and ribosome pathway genes.

Discussion

After having undergone the long-term artificial selection of domestication, the larvae weight of the domestic silkworm is about quadruple compared with wild silkworm no matter in the male or female (Fig. S1A). However, silk yield (cocoon shell weight) of the domestic silkworm is nine times more than that of wild silkworm (Fig. S1D). By correlation analysis, we found that the silk yield is largely dependent on the larvae weight in wild silkworm (Fig. S8A and 8B). Nevertheless, after corrected for larvae weight difference, the domestic silkworm also produces about two-fold silk yield in comparison with wild silkworm (Fig. S1E). This suggests that high efficiency of silk protein synthesis and/or other mechanisms has led to an increase in silk yield in the domestic silkworm.

Indeed, our analysis on time-series transcriptomes revealed some dynamic patterns of functional divergence during development of silk gland between the domestic and wild silkworms. At early and intermediate stages of silk gland development, the up-regulated genes of the domestic silkworm mainly include DNA integration, nucleic acid binding, and transporter activity (Fig. 2A). This is related to division and growth of cell [23–25]. In fact, silk gland of the domestic silkworm is larger than that of wild silkworm (Fig. 1). The posterior silk gland cell is considered as the “factory” for synthesis of fibroins.[5] While the DAPI was used to stain cells of posterior gland and the number of the cells was counted under fluorescence microscope (Fig. S9), we found that the number of cells of silk gland in the domestic silkworm is significantly larger than that in wild silkworm (491 ± 24 vs. 338 ± 20 , $P = 8.34E-11$, t-test). That is, the domestic silkworm has much more “factories” than wild silkworm to produce more silk fibroins after domestication and breeding improvement. Apart from more “factories” of silk fibroin production, our results also revealed that in the late stage of silk gland development, the up-regulated genes in the domestic silkworm relative to wild silkworm are enriched in protein processing in endoplasmic reticulum and ribosome pathways (Fig. 2A and Fig. S4A), indicating that protein synthesis is more active in the domestic silkworm than wild silkworm. These results are consistent with previous studies [18, 19, 26].

In this study, divergence in the expression pattern of silk-coding genes and in the composition of silk proteins between the domestic and wild silkworms was investigated. The results suggested that the synthetic capacity of fibroin proteins has increased but the synthetic capacity of sericin proteins has decreased during silkworm domestication (Fig. 3 and Fig. S7). This may be the results of different selection pressures acted on different components of cocoon proteins. High yield in silk fibroin is always a target of long-term and strongly artificial and breeding selection in the domestic silkworm because of its economic value [5]. However, the genes encoding both sericin and fibroin proteins in wild silkworm are under natural selection because of the importance of these proteins in protecting the cocoon in wild conditions [20].

To systematically understand the molecular mechanisms of silk yield increase during domestication, we constructed co-expression networks of silk-coding genes in the domestic and wild silkworms, respectively. Strikingly, we found that the network co-expressed with silk-coding genes for the domestic silkworm is much larger than that for wild silkworm (Fig. 4). Furthermore, much more genes co-expressed with silk-coding genes in the domestic silkworm have been subject to artificial selection than those in wild silkworm (Fig. 5A and 5B). This implies that the increase of silk yield during silkworm domestication and improvement is involved in many genes rather than a few genes. A recent resequencing study suggested that the nitrogen metabolism pathway is the most significantly enriched in the set of domestication-associated genes [27]. In the involved pathway, three genes participate in the synthesis of alanine, aspartate, and glutamate, providing important substances for the synthesis of silk protein [27]. Our results do not include the same genes as the previous study. This may be due to the fact that our results are based on dynamic patterns of transcriptomes in silk gland while the previous study is based on the static selection signature in the genomes. Importantly, our results show that artificial selection during silkworm domestication might have directly acted not only on the silk-coding genes but also on silk production related genes such as biological energy, transport, and ribosome pathway genes.

Many economically important traits of domestic animals, such as milk yield of dairy cows and egg production of chicken as well as silk yield of the silkworm, exhibit continuous distribution in hybrid populations and are called quantitative traits. These traits are generally thought to be controlled by multiple genes or loci. Therefore, at the beginning, an “infinitesimal” model was proposed to formalize the polygene background, which assumes genetic variation of quantitative traits is caused by a very large number of genes with very small and equal allelic effects [28]. Later, the distribution of allelic effects was found to be more nearly exponential, a few loci having large effects while a large number of loci showing very smaller effects [29]. Most importantly, the effects of genes affecting quantitative traits vary by different genetic and sexual backgrounds and external environment. The heritability of quantitative traits, the proportion of the phenotypic variance that is caused by genetic factors, typically ranges from ~5% to 50% [30]. Even though, evolution and improvement of complex traits are involved in many genetic factors including genotype by genotype interaction (epistasis), genotype by sex interaction, genotype by environment, and pleiotropy. Thus, quantitative traits are also known as complex traits. Our results suggested that increase in silk yield during silkworm domestication has been involved in the improvement of a biological system which includes not only expansion of “factories” of protein synthesis but also high expression of silk-coding genes as well as silk production related genes such as biological energy, transport, and ribosome pathway genes. This is due to the property of genetic architecture of complex traits.

Indeed, sometimes, it is possible that a certain effect will be obtained by affecting one node of a complex system. For example, previous studies overexpressed *Ras1^{CA}* and *York1^{CA}* in the posterior silk gland in the silkworm strain with a low level of silk yield, respectively, and observed the corresponding increase of silk yield [22, 31]. However, the researchers did not observe overlapping improved silk yield in the hybrid progeny of *Ras^{CA}*-overexpressing silkworms and *York1^{CA}*-overexpressing silkworms [31]. This suggests

an unknown and complex relationship between *Ras1* and *Yoekie* in affecting silk yield. Unexpectedly, when *Ras^{CA}*-overexpressing silkworms were hybridized with silkworm strain with high level of silk yield, their hybrid progeny did not show increase in silk yield (Gong CL, personal communication). Obviously, the genetic background may affect the role of *Ras^{CA}*-overexpression. Therefore, high silk yield in elite varieties in current sericulture is a result of long-term accumulation of mutations beneficial for silk-producing pathways as well as of continuous optimization of the related pathways and silk-producing organ. To understand the genetic basis and molecular mechanisms of improvement of a complex trait such as silk yield in the domestic silkworm, a combination of comparative multi-omics and dynamic network biological method is useful. In this sense, our study provides a methodology reference for investigation of molecular mechanisms of formation of a complex trait in other domestic animals.

Methods

Phenotypic investigation of silk in the domestic and wild silkworms

Wild silkworms are collected in Chongqing City, China. The domestic silkworms (Xiafang strain) are obtained from the sericultural research institute of Nanchong City. The method of rearing wild silkworm is the same as previously described [18]. First, we surveyed and compared the weight of cocoon silk between the domestic and wild silkworms, each sample was composed of 16 cocoons. Second, the cell number of posterior silk gland was counted for the domestic and wild silkworms by the method of nucleus staining with 4',6-diamidino-2-phenylindole (DAPI), respectively [23]. Moreover, we also used the scanning electron microscopy to compare the microstructural morphology of silk fibers in the cross-section between the domestic and wild silkworms [32, 33]. The cocoon was divided into three layers including inner, middle and outer [34]. The bundled silk fibers of each layer were put through a small tube and cut into short section, and then sputter-coated with palladium for 2 min [33]. The small tubes fixing onto the platform are used for the cross-sectional view of the silk fibers by the SU3500 (Hitachi, Japan).

Sample preparation, rna extraction, and rna sequencing

The fifth instar duration of the domestic silkworm is longer than that of wild silkworm. Thus, the silk gland of the domestic silkworm was dissected at the fourth molt (D0p), the 28h of the fifth instar (D1p), the 56h of the fifth instar (D2p), the 84h of the fifth instar (D3p), the 112h of the fifth instar (D4p), the 140h of the fifth instar (D5p) and the wandering stage (Dw). The silk gland of wild silkworm was dissected at the fourth molt (W0p), the 24h of the fifth instar (W1p), the 48h of the fifth instar (W2p), the 72h of the fifth instar (W3p), the 96h of the fifth instar (W4p), the 120h of the fifth instar (W5p) and the wandering stage (Ww) (Fig. 1). All the silk gland samples were frozen in liquid nitrogen for later investigation. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. 1.5 µg of total RNA was used to produce barcoded RNA sequencing libraries using the NEBNext® Ultra™ RNA Library Prep Kit (NEB, USA). Libraries were pooled in five different pools based on barcode compatibility, and each pool was sequenced on the Illumina platform.

Quality control and transcriptome assembly

The low-quality reads containing the adapter or >10% of N bases or > 50% of bases which quality scores ≤ 5 were filtered from the raw data. The retained reads are considered as clean reads and mapped to the reference genome of silkworm from the Ensembl database release 31 (*Bombyx mori*: GCA_000151625) with TopHat2 v2.1.1, respectively [35]. The transcripts of silk gland were assembled in each sample using Cufflinks v2.2.1 with the option -N 3—read-gap-length 3—read-edit-dist 3 -G [35, 36].

Quantification of gene expression and identification of differentially expressed genes

FPKM (Fragments per Kilobase of transcript Per Million mapped reads) is used to estimate the level of gene expression. To ensure the precision of expression, only the uniquely mapped reads were used to quantify the gene expression. The read counts of gene were calculated in the 25 developmental samples using the software HTseq v0.6.0 [37]. The read counts were standardized by the R package (edgeR) with the method of TMM (the weighted trimmed mean of M-values) and then transformed to the FPKM based on the gene length and sequencing depth [38, 39]. For the biological replicates, the mean of FPKM is chosen as the transcriptional signal of a gene. RNA-Seq of tissues was performed with the same pipeline [40].

The differential expression analysis was conducted at the same developmental stage between the domestic and wild silkworms using the edgeR [38]. The P -value was adjusted by the method of BH (Benjamini & Hochberg) [41]. The adjusted P -value < 0.05 and $|\log_2 \text{fold-change}| > 1$ are considered as significantly differential expression for the replicated sample. The criteria of adjusted P -value < 0.01 and $|\log_2 \text{fold-change}| > 1$ are for the non-replicated sample.

Comparative analysis of co-expression network

The weighted gene co-expression network analysis was carried out with the R package WGCNA v1.51 [42]. Since the lowly expressed or non-varying genes usually represent the 'noise' for the co-expression network analysis, the genes with FPKM < 1 more than 10 samples were filtered [38, 42, 43]. And the non-varying genes were also removed using the "goodSamplesGenes" function of WGCNA with default parameter [42, 43]. Gene expression data are used to generate the similarity matrix of Pearson correlations between gene and gene ($s_{ij} = \text{cor}(x_i)$) across the samples in the domestic and wild silkworms, respectively [42]. The adjacency matrices were created using the adjacency function with a weighted soft threshold ($a_{ij} = s_{ij}^\beta$, $\beta = 16$). The weighted soft threshold (β) was estimated by the criterion of approximate scale-free topology [44]. To weaken the effect of "noise" connections, the adjacency matrices were transformed into the topological overlap matrixes (TOM) using the "TOMsimilarity" function of WGCNA [42]. 1-TOM was used to perform average linkage hierarchical clustering and for the module detection [45, 46]. Fold change of weight value > 1.5 is considered as differential co-expression between the domestic and wild silkworms [47].

Functional analysis

TransDecoder is used to predict the protein of gene and only the longest one is retained [48]. All the genes are annotated for protein function performing by the InterProScan (v60) [49]. The results of InterProScan were transformed into gene ontology (GO) annotations. GO enrichment is carried out for differentially expressed genes and module genes with Goseq [50]. Significantly enriched analysis of KEGG is performed using KOBAS 3.0 based on the hypergeometric test and Benjamini-Hochberg correction [51].

Population genetic analysis

The 30X sequencing reads of whole-genome from 18 samples (10 for domestic silkworm and 8 for wild silkworm) were mapped to the reference genome using the software BWA-MEM (v0.7.7) [52]. The processes including sorting, duplicate marking, local realignment and base quality recalibration are then conducted with the alignment file (bam format) [53]. SNP and INDEL are discovered by the tools including GATK, SAMtools v1.4, and freebayes v1.0.2, respectively [53–55]. The common SNP of GATK, SAMtools, and freebayes are considered as high-quality SNP. Nucleotide polymorphism (π) and fixation index (F_{st}) of population differentiation are calculated by VCFtools [56]. Genes with selected signal are detected using the method describing in the references [57, 58].

Abbreviations

CSW: cocoon shell weight; DAPI: 4',6-diamidino–2-phenylindole; DEG: differentially expressed genes; Fib-H: fibroin heavy chain; Fib-L: fibroin light chain; MSG: middle silk gland; PSG: posterior silk gland; XF: Xiafang strain.

Declarations

Ethics approval and consent to participate

Experiments were conducted in accordance with the protocol approved by the Institutional Animal Care and Use Committee of the Chongqing University (permit number CBE-A201607020).

Consent for publication

Not applicable

Availability of data and material

We will release the RNA-seq raw data after the paper is published.

Competing interests

The authors declare no competing financial interests.

Funding

This work was supported by the National Natural Science Foundation of China (31772524) and the National High Technology Research and Development Program of China (2013AA102507–2).

Author contributions

ZZ and QYY participated in the research design. QCZ, HBZ, SSL and CJZ reared and collected the experiment materials. QZZ, FP, SSL and CJZ analyzed the data and performed the experiments. QZZ and FP wrote the manuscript. ZZ and QYY supervised and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Many thanks to Dr. Min-Jin Han and Dr. Wei Sun for his insightful comments in this study. We would like to thank Dr. Chun-Dong Zhang for his assistance in the experiment of cell nucleus staining. We thank Zhen-Tian Yan for help in the scanning electron microscope (SEM). We also thank the other members of Zhang's lab for their assistance in this study.

Reference

- 1.Astaurov BL, Rovinskaya IS. Chromosome complex of Ussuri geographical race of *Bombyx mandarina* M. with special reference to the problem of the origin of the domesticated silkworm, *Bombyx mori*. Cytology.1959;1:327–332.
- 2.Shimada T, Kurimoto Y, Kobayashi M. Phylogenetic relationship of silkmoths inferred from sequence data of the arylphorin gene. Mol Phylogenet Evol.1995;4(3):223–234.
- 3.Sun W, Yu H, Shen Y, Banno Y, Xiang Z, Zhang Z. Phylogeny and evolutionary history of the silkworm. Sci China Life Sci.2012;55(6):483–496.
- 4.Goldsmith MR, Shimada T, Abe H. The genetics and genomics of the silkworm, *Bombyx mori*. Annu Rev Entomol.2005;50:71–100.
- 5.Xia Q, Li S, Feng Q. Advances in silkworm studies accelerated by the genome sequencing of *Bombyx mori*. Annu Rev Entomol.2014;59:513–536.

6. Inoue S, Tanaka K, Arisaka F, Kimura S, Ohtomo K, Mizuno S. Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6:6:1 molar ratio. *J Biol Chem*.2000;275(51):40517–40528.
7. Gamo T. Genetic variants of the *Bombyx mori* silkworm encoding sericin proteins of different lengths. *Biochem Genet*.1982;20(1–2):165–177.
8. Chevillard M, Couble P, Prudhomme JC. Complete nucleotide sequence of the gene encoding the *Bombyx mori* silk protein P25 and predicted amino acid sequence of the protein. *Nucleic Acids Res*.1986;14(15):6341–6342.
9. Michaille JJ, Couble P, Prudhomme JC, Garel A. A single gene produces multiple sericin messenger RNAs in the silk gland of *Bombyx mori*. *Biochimie*.1986;68(10–11):1165–1173.
10. Yamaguchi K, Kikuchi Y, Takagi T, Kikuchi A, Oyama F, Shimura K, Mizuno S. Primary structure of the silk fibroin light chain determined by cDNA sequencing and peptide analysis. *J Mol Biol*.1989;210(1):127–139.
11. Zhou CZ, Confalonieri F, Medina N, Zivanovic Y, Esnault C, Yang T, Jacquet M, Janin J, Duguet M, Perasso R, et al. Fine organization of *Bombyx mori* fibroin heavy chain gene. *Nucleic Acids Res*.2000;28(12):2413–2419.
12. Li B, Wang X, Hou C, Xu A, Li M. Genetic analysis of quantitative trait loci for cocoon and silk production quantity in *Bombyx mori* (Lepidoptera: Bombycidae). *Eur J Entomol*.2013;110(2):205–213.
13. Zhan S, Huang J, Guo Q, Zhao Y, Li W, Miao X, Goldsmith MR, Li M, Huang Y. An integrated genetic linkage map for silkworms with three parental combinations and its application to the mapping of single genes and QTL. *BMC Genomics*.2009;10:389.
14. Li C, Zuo W, Tong X, Hu H, Qiao L, Song J, Xiong G, Gao R, Dai F, Lu C. A composite method for mapping quantitative trait loci without interference of female achiasmatic and gender effects in silkworm, *Bombyx mori*. *Anim Genet*.2015;46(4):426–432.
15. Li C, Tong X, Zuo W, Luan Y, Gao R, Han M, Xiong G, Gai T, Hu H, Dai F, et al. QTL analysis of cocoon shell weight identifies BmRPL18 associated with silk protein synthesis in silkworm by pooling sequencing. *Scientific Reports*.2017;7(1):17985.
16. Li J, Qin S, Yu H, Zhang J, Liu N, Yu Y, Hou C, Li M. Comparative Transcriptome Analysis Reveals Different Silk Yields of Two Silkworm Strains. *PLoS One*.2016;11(5):e0155329.
17. Luan Y, Zuo W, Li C, Gao R, Zhang H, Tong X, Han M, Hu H, Lu C, Dai F. Identification of Genes that Control Silk Yield by RNA Sequencing Analysis of Silkworm (*Bombyx mori*) Strains of Variable Silk Yield. *Int J of Mol Sci*.2018;19(12):3718.

- 18.Fang SM, Hu BL, Zhou QZ, Yu QY, Zhang Z. Comparative analysis of the silk gland transcriptomes between the domestic and wild silkworms. BMC Genomics.2015;16:60.
- 19.Li JY, Cai F, Ye XG, Liang JS, Li JK, Wu MY, Zhao D, Jiang ZD, You ZY, Zhong BX. Comparative Proteomic Analysis of Posterior Silk Glands of Wild and Domesticated Silkworms Reveals Functional Evolution during Domestication. J of Proteome Res.2017;16(7):2495–2507.
- 20.Dai ZJ, Sun W, Zhang Z. Comparative analysis of iTRAQ-based proteomes for cocoons between the domestic silkworm (*Bombyx mori*) and wild silkworm (*Bombyx mandarina*).. J. Proteomics.2019;192:366–373.
- 21.Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols.2012;7(3):562–578.
- 22.Ma L, Xu H, Zhu J, Ma S, Liu Y, Jiang RJ, Xia Q, Li S. *Ras1^{CA}* overexpression in the posterior silk gland improves silk yield. Cell Res.2011;21(6):934–943.
- 23.Zhang CD, Li FF, Chen XY, Huang MH, Zhang J, Cui H, Pan MH, Lu C. DNA replication events during larval silk gland development in the silkworm, *Bombyx mori*. J Insect Physiol.2012;58(7):974–978.
- 24.Oh H, Irvine KD. In vivo analysis of Yorkie phosphorylation sites. Oncogene.2009;28(17):1916–1927.
- 25.Oh H, Irvine KD. In vivo regulation of Yorkie phosphorylation and localization. Development.2008;135(6):1081–1088.
- 26.Zhou QZ, Fang SM, Zhang Q, Yu QY, Zhang Z. Identification and comparison of long non-coding RNAs in the silk gland between domestic and wild silkworms. Insect Sci.2018;25(4):604–616.
- 27.Xiang H, Liu X, Li M, Zhu Y, Wang L, Cui Y, Liu L, Fang G, Qian H, Xu A, et al. The evolutionary road from wild moth to domestic silkworm. Nature Ecology & Evolution.2018(8):1268–1279.
- 28.Mackay TF. The genetic architecture of quantitative traits. Annu Rev Genet.2001;35:303–339.
- 29.Lynch M, Walsh B: Genetics and Analysis of Quantitative Traits. Oxford: Oxford University Press; 1998.
- 30.Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. Nature Reviews Genetics.2019;20(3):135–156.
- 31.Zhang P, Liu S, Song HS, Zhang G, Jia Q, Li S. *Yorkie^{CA}* overexpression in the posterior silk gland improves silk yield in *Bombyx mori*. J Insect Physiol.2017;100:93–99.
- 32.Chen F, Porter D, Vollrath F. Structure and physical properties of silkworm cocoons. J R Soc Interface.2012;9(74):2299–2308.

- 33.Guan J, Zhu W, Liu B, Yang K, Vollrath F, Xu J. Comparing the microstructure and mechanical properties of *Bombyx mori* and *Antheraea pernyi* cocoon composites. *Acta Biomater.*2017;47:60–70.
- 34.Zhang Y-Y, Wu Z-Y, Zhou Y, Zhao F, Li J. An analysis on morphological structure of cocoon filaments from different layer of mulberry silkworm and eri-silkworm cocoons. *Science of Sericulture.*2013;39(6):1126–1130.
- 35.Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology.*2013;14(4):R36.
- 36.Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol.*2014;32(5):462–464.
- 37.Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics.*2014;31:166–169.
- 38.Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.*2010;26(1):139–140.
- 39.Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology.*2010;11(3):R25.
- 40.Wu Y, Cheng T, Liu C, Liu D, Zhang Q, Long R, Zhao P, Xia Q. Systematic Identification and Characterization of Long Non-Coding RNAs in the Silkworm, *Bombyx mori*. *PLoS One.*2016;11(1):e0147147.
- 41.Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical And Powerful Approach To Multiple Testing. *J Roy Stat Soc B Met.*1995;57(1):289–300.
- 42.Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.*2008;9:559.
- 43.Miller JA, Horvath S, Geschwind DH. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc Natl Acad Sci U S A.*2010;107(28):12698–12703.
- 44.Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.*2005;4:Article17.
- 45.Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science.*2002;297(5586):1551–1555.
- 46.Yang L, Li Y, Wei Z, Chang X. Coexpression network analysis identifies transcriptional modules associated with genomic alterations in neuroblastoma. *Biochimica et Biophysica*

Acta.2018;1864(6):2341–2348.

47.Lu X, Li QT, Xiong Q, Li W, Bi YD, Lai YC, Liu XL, Man WQ, Zhang WK, Ma B, et al. The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *The Plant Journal*.2016;86(6):530–544.

48.Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*.2011;29(7):644–652.

49.Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztanyi Z, El-Gebali S, Fraser M, et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res*.2017;45(D1):D190-D199.

50.Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*.2010;11(2):R14.

51.Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res*.2011;39:W316–322.

52.Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*.2010;26(5):589–595.

53.McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*.2010;20(9):1297–1303.

54.Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*.2009;25(16):2078–2079.

55.Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. <https://arxiv.org/abs/1207.3907> (2012). Accessed 15 Oct 2019.

56.Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics*.2011;27(15):2156–2158.

57.Qiu Q, Wang L, Wang K, Yang Y, Ma T, Wang Z, Zhang X, Ni Z, Hou F, Long R, et al. Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nature Communications*.2015;6:10283.

58.Wang MS, Zhang RW, Su LY, Li Y, Peng MS, Liu HQ, Zeng L, Irwin DM, Du JL, Yao YG, et al. Positive selection rather than relaxation of functional constraint drives the evolution of vision during chicken domestication. *Cell Res*.2016;26(5):556–573.

Figures

Figure 1

Morphology variation of the fifth instar larvae and silk gland between the domestic and wild silkworms. W: wild silkworm, D: domestic silkworm (Xiafang strain), p: time point, w: the wandering stage of silkworm (the last time point of the fifth instar), 0p: start of the fifth instar.

Figure 2

Functional divergences in the silk gland development between the domestic and wild silkworms. (A) Gene ontology enrichment of up-regulated genes in the domestic silkworm. (B) Gene ontology enrichment of down-regulated genes in the domestic silkworm. (C) The GO and KEGG term are more enrichment in the up-regulated genes than the down-regulated genes in all time points. (D) Expressed profile of all mitochondrial genes in the domestic and wild silkworms.

Figure 3

Expression pattern of silk-coding gene and micro-composition of silk between the domestic and wild silkworms (A-F) expression profile for *Fib-H*, *Fib-L*, *P25*, *Sericin1*, *Sericin2*, and *Sericin3*. Y-axis, the value of FPKM. X-axis, the differential stage. (G-I) Cross section of the outer, middle and inner layer silk of the domestic silkworm, respectively. (J-L) Cross section of the outer, middle and inner layer silk of the wild silkworm, respectively.

Figure 4

Functional enrichment of lightsteelblue module and differential co-expression network of silk-coding genes. (A) Gene ontology enrichment of lightsteelblue module. (B) KEGG enrichment of lightsteelblue module. (C) The co-expression network of silk-coding genes in the domestic silkworm. (D) The co-expression network of silk-coding genes in wild silkworm.

Figure 5

The influence of domestication for the co-expression network of silk-coding genes. (A-B) Domestic and differential co-expression network of silk-coding genes in the domestic and wild silkworms, respectively. (C) Nucleotide polymorphism (π) and Fixation index (F_{st}) of BGIBMGA000074. Red rectangle, Exon. (D) Degree of five silk-coding genes in the domestic and differential co-expression network of the domestic and wild silkworm. (F) Tissue expression of BGIBMGA006919, BGIBMGA012537, BGIBMGA000074, and BGIBMGA001462.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS5.pdf](#)
- [FigS7.pdf](#)
- [FigS2.pdf](#)
- [FigS4.pdf](#)
- [FigS1.pdf](#)
- [TableS4.xls](#)
- [TableS1.doc](#)
- [TableS2.doc](#)
- [FigS9.pdf](#)
- [FigS8.pdf](#)
- [TableS3.xls](#)
- [FigS3.pdf](#)
- [FigS6.pdf](#)