

A Cluster of Genes at the Terminal Region of Chromosome 1 Contributes to Silk Yield in *Bombyx Mori*.

Yue Luan

Southwest university

Chunlin Li

Southwest University

Yanyu Liu

Southwest university

Bili Zhang

Southwest University

Rui Gao

Southwest University

Weidong Zuo

Southwest University

Hai Hu

Southwest University

Xiaoling Tong

Southwest university

Cheng Lu

Southwest University

Fangyin Dai (✉ fydai@swu.edu.cn)

State Key Laboratory of Silkworm Genome Biology, Southwest University <https://orcid.org/0000-0002-0215-2177>

Research

Keywords: Silkworm, silk yield, Chr1, gene cluster, co-regulation

Posted Date: November 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1031814/v1>

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

[Read Full License](#)

Abstract

Background:

The silk yield is the most economically important trait in sericulture. In the domestication of the silkworm, traits related to silk yield are the main selective targets. In previous studies, linkage and association analyses were performed, which narrowed the region of the locus controlling the cocoon shell weight (CSW), an important index of silk yield, to the interval of 0–134 kb on the 1st chromosome.

Results

In this study, selection pressure analysis showed that most of the loci in this interval were co-selected during silkworm domestication. Association analysis conducted with 95 strains found that there is a 110 kb region that was significantly associated with CSW. An expression profile correlation analysis of eight genes in the region showed that their expression patterns were very similar. Sequence variation analysis detected mutations in the exon regions of BMgn002066 and BMgn002073, leading to premature termination of translation. The sequence variation was significantly correlated with CSW. In addition, another two genes, BMgn002067 and BMgn002071, had significantly different expression levels in the posterior silk glands of 11 high-yield strains and 11 low-yield strains. Therefore, we speculate that the abovementioned four genes and the gene BmAbl1 detected in previous studies may co-regulate CSW.

Conclusions

This study is the first to identify the gene clusters that affect CSW by forward genetics, thereby providing new evidence for gene clusters that regulate quantitative traits.

Background

As a natural fiber, silkworm silk has a high industrial value, and it currently makes a huge contribution to the economic growth of many developing countries. The breeding of high-yield strains is conducive to increasing the income of sericulture farmers and reducing the costs involved in sericulture. A long-term breeding program has produced a large number of improved silkworm strains, and their silk yield has been significantly improved. In addition, due to its biological properties such as air and moisture permeability, non-toxicity, non-irritation to the skin, and strong human compatibility, silk has significant future commercial value. At present, silk has applications in the fields of beauty care and medical biomaterials [1–4]. Therefore, silk yield has always been one of the most significant aspects of the sericulture industry. However, researchers have found that it is difficult to achieve further yield breakthroughs using traditional breeding methods. Therefore, molecular breeding methods are needed to produce more high-yield strains.

Similar to the yield traits of most other species, silk yield carries the genetic signature of quantitative traits and is regulated by multiple genes [5]. Researchers have used omics analyses such as transcriptome and proteome [6–10] and forward genetics to explore the genes that regulate silk yield. However, because the silkworm has complex genetic characteristics such as female complete linkage and other factors that could affect yield-related indicators, it is difficult to formulate suitable mapping materials and methods in forward genetic studies, making it difficult to research the genetic basis of silk yield. Therefore, over the past 100 years, only a few major genes have been detected, including *BmGlcNase1* and *BmAbl1* [11, 12]. Since the inheritance of silk yield traits is dominated by additive effects [13] and the few identified genes have limited effects on molecular breeding, it is necessary to identify additional silk yield controlling genes.

In the previous work, our research group used linkage analysis to circumvent the effects of gender in order to locate a quantitative trait loci (QTL) at 0–134k of the 1st chromosome (Chr1) and found that *BmAbl1* (*BMgn002070*) was significantly related to the CSW in this region [11]. In the course of this research, we preliminarily found that the mapping region seemed to have been selected as a whole unit. In the classical genetic studies on the economically important traits of silkworm, Chr1 has long been regarded as the chromosome where the major effect loci are located. Many mapping studies indicated that Chr1 explained the largest percentage of the phenotypic variance [13–18] where the additive effect was dominant, implying that more genes on Chr1 contributed to CSW [13]. Therefore, it is important to decipher this region in detail to identify more genes controlling silk yield.

In this study, we conducted an overall examination of the 0–134 kb region at the terminal portion of Chr1 and performed expression and sequence analyses of genes in the region in order to reveal the significance of the terminal portion of Chr1 in silk yield.

Methods

1. Insects

Low-yield strain IS-Dazao, high-yield strain 872B, and 93 other low-yield and high-yield strains were all from the Silkworm Gene Bank of Southwest University, China. The larvae were fed with fresh mulberry leaves under a 12-h/12-h light/dark photoperiod at 25°C.

2. Expression analysis

The materials used in the temporal expression profiles were IS-Dazao and 872B posterior silk glands of 5th instar larvae. The glands were dissected in physiological saline at the same timepoint of each day of the 5th instar and frozen in liquid nitrogen. The posterior silk glands used for multi-strain association analysis were taken from 22 strains on day 3 of the 5th instar. Tissues were lysed using TRIzol (Invitrogen, Carlsbad, CA), and total RNA was extracted according to the manufacturer's instructions, followed by reverse transcription of the RNA using the PrimeScript RT reagent Kit with gDNA Eraser kit to obtain cDNA for each tissue for subsequent quantitative analysis.

We performed RT-qPCR using the Bio-rad CFX96 sequence detection system with an iTaqSYBR Green (Bio-rad) according to the manufacturer's instructions. The *B. mori* ribosomal protein L3 (rpL3, GenBank Accession Number AY769270.1) was used as the reference, and the primers used in this process are listed in the supporting Table 1. Each sample had three or more biological replicates and technical replicates, and the expression levels of the genes were analyzed using Student's *t* test.

3. Association analysis

Association analysis involved silkworm strains with different yield traits, and the CSWs were weighed. The indels were searched in the initial mapping region. Specific primers were designed based on the flanking sequences and sent to Huada Gene Synthesis. Primers were then amplified in the parental genomic samples to detect amplification specificity and polymorphisms. Polymorphic markers were developed for selective genotyping. This is a method of genetic mapping based on linkage disequilibrium. We classified the above developed indel markers in the 95 strains and conducted association analysis based on their phenotypic data. The genotype of the strain was classified, and the significance was judged by the degree of association between the band type and CSW. The statistical results were analyzed by one-way ANOVA (** = $P < 0.01$). The primer sequences are listed in supporting Table 1.

4. Sequence cloning

PCR products were recovered using the Glue Recovery Kit (OMEGA) according to manufacturer's instructions. After recovery, the product fragment was inserted into a PMD19-T vector (Takara) by solution I and ligated at 16°C for 12 h. The ligation product was transferred into *Escherichia coli* competent cells. Sequences were obtained by cloning and picking positive colonies for sequencing.

5. Selective pressure analysis

SNP calling was performed using resequencing data of seven wild silkworm strains, 42 local strains, and 87 improved strains for subsequent selection pressure analysis [19]. We calculated the population nucleotide diversity (π) of wild silkworms, local silkworms, and improved silkworms with a window size of 200 kb and a step size of 100 bp. The *Fst* values of the differentiation degree were calculated to determine the degree of selection for the gene.

Results

The terminal region of Chr1 was subject to domestication selection

In order to study whether the 0–134 kb region of Chr1 (Fig. 1a) was subject to selection during domestication, we selected wild strains, local strains, and improved strains to analyze the nucleotide diversity of this region. The results showed that the nucleotide diversity (π) of wild silkworm from 23.2k

of Chr1 was much higher than that of local and improved strains. The average π value of wild silkworms was 0.0090, while being only 0.0044 and 0.0041 in local and improved strains, respectively. The domestication process from wild strains to local strains reduced the π of wild silkworms by 51%, while the π value of the breeding process from local strains to improved strains was reduced by only 7% (Fig. 1b). This result suggests that the entire region was subjected to purifying selection, and the reduction of nucleotide diversity mainly occurred in the domestication stage. In order to further verify the above inferences, we used wild silkworm strains vs local strains and local strains vs improved strains as materials to simulate the domestication process and the breeding process and analyzed the degree of sequence divergence (*Fst*) in the candidate region. The results showed that consistent with the analysis of nucleotide diversity, the value of *Fst* in the range of 0–23.2k was not very different, and significant divergence began at the 23.2k point. The highest *Fst* of wild silkworm strains vs local strains was 0.793, with an average of 0.392; the highest *Fst* of local strains vs improved strains was 0.301, with an average of 0.063 (Fig. 1c). This result suggested that the loci were selected during the domestication process rather than during the breeding process. Since the reduction of nucleotide diversity may also be affected by factors such as genetic drift and bottleneck effects, we calculated Tajima's D in this region. The data showed that Tajima's D of local strains was always less than 0 in this region, suggesting that the decrease of nucleotide diversity in this region during domestication was indeed the result of purifying selection. Moreover, we also noticed that the value of Tajima's D of the wild population was always less than 0, implying that this region was also subject to purifying selection in wild silkworms. It is necessary to conduct an in-depth study on the specific reasons and the significance of this result (Fig. 1d). In summary, the analysis shows that the region of Chr1 from 23.2 kb to 134 kb was subjected to selection during domestication, and that selection in the region has maintained its integrity.

The sequence variation in this region is associated to the CSW

In sericulture, the CSW is the most important index of silk yield. Therefore, CSW was our first consideration for breeding target traits. Whether this locus is associated with CSW needed to be further explored through association analysis between polymorphic loci and CSW. In order to find the polymorphic loci, we sequenced the entire region in the low-yield strain IS-Dazao and the high-yield strain 872B and found 25 indels. We then genotyped these indels in 95 different silkworm strains with known CSW to detect the association between the variants and silk yield. Through association analysis, we found that 23 indels in the region showed a high level of association with CSW (Fig. 2a). By comparing the position information, 14 of the 25 indel markers were located in the gene region of interest. Eleven indels were located in the intron region; three were located in the exon region, and 11 were located in the intergenic region, including two indels located in the upstream 5K regulatory region (Fig. 2b). Using the indel of this region as a marker for linkage disequilibrium analysis, we found that the degree of linkage disequilibrium in this region was very high, with an average D of 0.86 (Fig. 2c), implying that the region as a whole has been co-selected. In summary, the region screened by the association analysis was similar to the results of the selection pressure analysis. The sequence variation in the interval from 25.2 kb to 134 kb was related to CSW, and the region is in a state of linkage disequilibrium (Fig. 2a).

The expression patterns of genes in the region are significantly similar

Selection pressure analysis, association analysis, and linkage disequilibrium analysis showed that the region may be co-selected for improving CSW. We therefore speculated that multiple genes in this region co-regulated CSW. To explore this, we first searched for genes in the region and found that the functions of these genes were mostly related to cell division and growth. Multiple genes working together are often accompanied by similar expression profiles. We determined the expression levels of genes in the region. The silk gland is the main organ secreting silk protein. The middle silk gland secretes sericin, and the posterior silk gland secretes silk fibroin. Silk fibroin is the main raw material used in the silk industry; so, the posterior silk gland was used as the material in this research. Because the 5th instar of the silkworm is the critical period for silk gland development and silk protein synthesis, we chose the 5th instar posterior silk gland as the experimental material to analyze the expression profiles of genes. The qPCR analysis results for eight genes in the region found that these genes had similar expression profiles. Their expression levels gradually increased with the development of silk glands, and their expression levels reached a peak on day 6 of the 5th instar, which was the most rapid period of silk gland development. We then calculated the correlation coefficients between gene expression profiles and found that the expression profiles of each gene pair had extremely high correlations, and the average Pearson's correlation coefficient (r) was always higher than 0.5 (Fig. 3b). The frequency of the correlation coefficient (r) between genes in the 0.8–1.0 interval was as high as 69.4% (Fig. 3c). The above results suggest that genes work together to promote silk gland development.

We matched the physical locations of these genes and the values of F_{st} in the region where the genes were located and found that these genes had a high degree of population differentiation between wild and local strains (supporting Table 2). The average values of F_{st} of *BMgn002068*, *BMgn002066*, *BMgn002067*, *BMgn002073*, and *BMgn002070* were all above 0.45 (supporting Table 2). We annotated the functions of these genes and found that many of them were related to energy metabolism and cell growth and division (supporting Table 2). According to the results of correlation analysis of the expression levels of multiple genes and the correlations of function annotation, there may be a regulatory mechanism involving multiple genes that co-regulates the CSW in the candidate region.

Table 1
Functional annotation of genes in the candidate region

Gene ID	Position (bp)	Fst (wild vs local)	Functional annotation
BMgn002068	92459-92686	0.6119	Unknown Function
BMgn002065	134200-134445	0.0402	Unknown Function
BMgn002066	126208-132433	0.6394	Adenylate and Guanylate cyclase
BMgn002067	100923-115702	0.4605	Alpha tubulin acetyltransferase 1
BMgn002073	32689-45286	0.602	SLC4-like anion exchanger
BMgn002072	4787-4996	0.0402	Unknown Function
BMgn002071	9533-15921	0.0424	Methenyltetrahydrofolate synthetase
BMgn002070	51384-77081	0.5	Abl tyrosine kinase

Genes Are Associated With Csw

There were 14 markers located in the gene region and three markers located in the exon region in the candidate region that were significantly associated with CSW (Fig. 2b). In order to explore whether an indel located in the exon region affected the function of the corresponding gene, we performed a sequence blast between high-yield and low-yield silkworm strains. The results showed that *BMgn002066* and *BMgn002073* had 1 bp and 11 bp deletions in the 6th and 2nd exon regions, respectively. These sequence variations led to premature termination of translation and breakage of the domain, which in turn caused significant changes in the protein structure (Fig. 4a, b). In order to explore whether the sequence variations were related to CSW, we divided 95 association analysis materials into two groups based on the sequence type, namely H haplotype and L haplotype, and performed an association analysis with CSW. There was a significant difference between H and L haplotypes CSW. The mutation in *BMgn002066* reduced CSW by 0.069 g, which was about 27% of the H haplotye CSW. The mutation in *BMgn002073* reduced CSW by 0.080 g, which was about 38% of the H haplotye CSW (Fig. 4a, b). The above results indicate that the two sequence variants that can cause changes in protein structure are significantly related to CSW and can explain yield trait variation to a large extent.

There is often a certain degree of association between a trait and the expression level of the gene that regulates the trait. Therefore, we performed multi-strain expression level analysis on the seven genes detected in the selection pressure analysis (excluding the published gene *BMgn002070*) to further screen for co-regulatory genes. In the expression profile analysis, we found that the genes reached the peak of expression level the day before mounting the cocooning frame, i.e., day 6 of the 5th instar. Therefore, we selected 11 high-yield strains and 11 low-yield strains to determine the gene expression levels in the posterior silk glands on day 6 of the 5th instar. The results showed that the expression levels of genes

BMgn002067 and *BMgn002071* showed significant differences. This result indicates that these two genes may co-regulate silk gland development, thereby affecting the CSW.

Discussion

In this study, we found a region significantly associated with CSW through artificial selection analysis and association analysis. Through sequence and expression profile analyses, we found that genes within the region appeared to function together. The expression levels of *BMgn002067* and *BMgn002071* were significantly different between high-yield and low-yield strains, while the exon regions of *BMgn002066* and *BMgn002073* were mutated, leading to changes in the domain.

Among the genes with domain mutations, *BMgn002066* encodes adenylate and guanylate cyclase, an important signal transduction enzyme. After activation, it can not only activate the NO-SGC-cGMP signaling pathway but also inhibit the TGF- β signaling pathway [20-23]. cGMP is an important second messenger *in vivo* and can regulate downstream effector molecules such as protein kinase G. cGMP-dependent phosphodiesterase and cGMP-gated ion channels participate in a series of physiological or pathological reactions, including vasodilatation of blood vessels, inhibition of platelet aggregation, inhibition of cell proliferation, and other physiological regulation processes [21]. The inhibition of TGF- β signaling pathway can induce physiological effects on the inhibition of tissue fibrosis and cell proliferation. In the research on rectal and thyroid cancer, it has been found that adenylate and guanylate cyclase plays an important role in cancer proliferation [21]. As the silk gland undergoes cell division during the embryonic period, the number of cells is significantly related to the silk yield, and thus, we speculate that *BMgn002066* affects the development of the silk gland by affecting cell proliferation. *BMgn002073* encodes an SLC4-like anion exchanger. SLC4 gene products play an important role in CO₂ transport in red blood cells, in H⁽⁺⁾ and HCO₃⁽⁻⁾ absorption or secretion by various epithelial cells, and in regulating cell volume and intracellular pH [24, 25, 26]. Studies have shown that AE2 of the SLC4 family can regulate cell volume increase by Cl⁽⁻⁾ uptake [26]. In the silkworm, enlargement of silk gland cells occurs during the larval stage after the completion of silk gland cell division in the embryo. Combined with the function of *SLC4* to regulate the cell volume, we speculate that this enzyme may affect the silk gland development by regulating the cell volume.

The genes *BMgn002067* and *BMgn002071* encode a tubulin-N-acetyltransferase 1 (ATAT1) and 5-formyltetrahydrofolate cyclase (MTHFs), respectively, and there were significant differences in the expression level between high- and low-yield strains. *ATAT1* is used to catalyze the acetylation of tubulin and is clustered in the nucleus during the G1-G2 phase. At late mitosis, *ATAT1* co-locates with chromatids and spindles and eventually migrates to daughter nuclei, newly synthesized centrioles, and midbodies [27]. The specific distribution of *ATAT1* in the cell cycle suggests that *ATAT1* has multiple functions, including microtubule acetylation, RNA transcription activity, microtubule cut-off, and cytokinesis completion [28-33]. In the study of the function of *ATAT1* in cell division, researchers found that knocking out *ATAT1* in *Tetrahymena* slowed its growth rate [34]. In the silkworm, silk gland development is closely related to cell division. Investigations have shown that the number of silk gland cells in high-yield strains

is higher than that in low-yield strains. The number of silk gland cells does not change after the completion of silk gland cell division in the embryonic stage. ATAT1 functional research on the regulation of cell division indicates that this gene may regulate silk gland development by affecting embryonic mitosis. MTHFs are important enzymes in the pathway of folate metabolism, being involved in the metabolism of folate, purines, vitamins, and coenzyme factors. MTHFs regulate carbon flow through folate-dependent single-carbon metabolic networks [35, 36]. This regulatory network provides carbon for the biosynthesis of purines, thymine, and amino acids and influences DNA synthesis *in vivo*, thus influencing cell division and maturation [37, 38, 39]. MTHFs have been demonstrated to inhibit cell growth in human MCF-7 breast cancer cells; it has been confirmed in mice that MTHFs are an essential part of the purinosome and provide conditions for cancer cells to rapidly synthesize purine nucleotides [40, 41]. These studies of MTHFs show that they are closely related to cell division. In the silkworm, silk gland cell proliferation and growth directly affect silk gland development. Combined with the function of MTHFs in cell growth and division in other species, it can be inferred that genes that regulate silk yield are very likely to participate in energy metabolism and DNA synthesis. Therefore, it is possible that MTHFs can influence the energy supply of silk gland development, provide carbon for DNA synthesis, and affect the division and maturation of silk gland cells, thus affecting the CSW.

In this study, multiple genes were shown to coordinate and regulate the same trait within a certain region. Generally, there are two types of gene clustering distributions. The first type is a gene family in the form of a cluster in a region; this is the most common. The second type is where genes with similar functions are clustered together, although they are not necessarily in the same gene family [42]. Clustering distribution of genes has been studied in other species, for example, the conserved supergene loci affecting butterfly diversity [43], the supergene loci affecting butterfly mimicry [44], and a group of supergenes influencing the shape of male water birds' neck hair [45]. Studies have shown that gene clustering is an effective gene regulation mode in biological evolution. Although some clustered genes belong to different gene families, they always show similar expression profiles. This may be due to the initiation of the regulation of gene expression or the structural changes of chromosomes. There is more than one gene in the same open box that can be activated and transcribed at the same time, thereby greatly improving the efficiency of transcription. There have also been studies on silkworms through mixed pool sequencing analysis; these have found that most of the genes related to silk gland development in the silkworm are clustered in the genome and have similar expression profiles [46]. The selected genes *ATAT1*, *MTHFs*, *Adenylate and Guanylate cyclase*, and *SLC4-like anion exchanger* appear in clusters along with the silk yield regulating gene *BmAbf1* detected in previous studies, and the expression profiles are significantly similar. Based on this, we speculate that they may share the same enhancer to regulate the open reading frame, and the efficiency of multi-gene action can be greatly improved by regulating the expression of gene clusters at the same time. The simultaneous accumulation of gene effects can jointly regulate cell division, protein synthesis, and energy supply in order to regulate silk yield traits more efficiently. This finding implies that it is not only morphological traits but also a similar mode of regulation in quantitative traits that will provide a new reference for the study of quantitative traits.

The results of the domestication analysis in this study were very interesting. In the selection pressure analysis, we found that the location of the gene cluster received stronger selection during the domestication process than during the breeding process. This was indicated since from wild to local strains, CSW only increased from 0.06 g to about 0.15 g, while in the process of breeding, the CSW of improved strains reached up to 0.4 g [12]. The silk yield has been greatly improved during the breeding process, and the researchers therefore speculated that the breeding process may have entailed greater selection for silk yield-related genes. In addition, the results of a principal component analysis (PCA) of the silk gland transcriptome in a previous study showed that most of the differentially expressed genes had little difference in principal components between wild and local strains, while the difference between local and improved strains was greater [10]. The PCA results confirmed the above speculation. However, in this study, the loci of the gene cluster were more strongly selected during the domestication process. In response to the completely different results of this study, we found the same situation in a cotton yield study by consulting the literature. The genetic divergence degree of cotton from local strains to improved strains was only 0.04, while the *Fst* of wild and local strains was as high as 0.10. This suggests that the process of breeding has had little effect on the genetic diversity of cotton, while the process of domestication greatly reduced the genetic diversity of cotton [47]. From this result, we hypothesized that our result might be due to the fact that some important yield regulatory genes were selected and fixed in advance during the domestication process to ensure the base yield.

This study has identified the first gene cluster region mapped by the forward genetic method, providing evidence for the regulation of gene clusters in silk yield. However, the regulation mode of gene clusters in the region and the mechanism of how each gene regulates silk yield need to be studied further in future.

Conclusion

This study found that there is a cluster of genes at the terminal region of chromosome 1 that can jointly regulate the CSW of the silkworm. The genes within the cluster have similar expression profiles, and the region where the gene cluster is located is co-selected and is significantly related to CSW. Four genes related to CSW were identified through gene sequence variation and gene expression level analyses among multiple strains. These results are significant for our understanding of the genetic basis of silk yield, and the data offer a reference to direct the molecular breeding of silkworms.

Abbreviations

CSW
cocoon shell weight
QTL
quantitative trait loci
TATA1
tubulin-N-acetyltransferase 1
MTHFs

5-formyltetrahydrofolate cyclase.

Declarations

Authors' contributions: YL, CLL and FYD formulated and designed the experiments; YL, YYL, HH and RG acquired and analyzed the data; YL, CLL, BLZ and YYL performed the experiments; YL, CLL and Wdz constructed the figures; YL prepared the manuscript; CLL, YYL, XLT, CL and FYD revised and proofread the manuscript. All authors have read and approved the final manuscript.

Acknowledgments: We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 31830094, No. U20A2058 and No. 31902211) and the Municipal Graduate Student Research Innovation Project of Chongqing [CYB19113]

Availability of data and materials: The datasets supporting the conclusions of this study are included within the article and as additional files

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Competing interests: The authors have no competing interests to declare.

References

1. Alsbjorn B. Biologic wound covering skin burn treatment. *World J Surg* 1992; 16: 43–46.
2. Minoura N, Aiba S, Gotoh Y, Tsukada M, Imai Y. Attachment and growth of cultured fibroblast cells on silk protein matrices. *J Biomed Mater Res* 1995; 29: 1215–1221.
3. Liu H, Liu Y, Qian J, Yu J, Denga J. Fabrication and features of a Methylene Green-mediating sensor for hydrogen peroxide based on regenerated silk fibroin as immobilization matrix for peroxidase. *Talanta* 1996; 43: 111–8.
4. Tamada Y. Sulfation of silk fibroin by chlorosulfonic acid and the anticoagulant activity. *Biomaterials* 2004; 25: 377–383.
5. Shen DH. On the inheritance of some quantitative characters in *Bombyx mori* L. *Jour. College of Agri. Imp. Univ. of Tokyo* 1928; 10: 39–66.
6. Wang SH, You ZY, Ye LP, Che J, Qian KJ, Nanjo Y, Komatsu S, Zhong BX. Quantitative proteomic and transcriptomic analyses of molecular mechanisms associated with low silk production in silkworm *Bombyx mori*. *J Proteome Res* 2014; 13: 735–51.

7. Ma L, Ma Q, Li X, Cheng LL, Li K, Li S. Transcriptomic analysis of differentially expressed genes in the Ras1CA-overexpressed and wildtype posterior silk glands. *BMC Genomics* 2014; 15: 182.
8. Li J, Qin S, Yu HJ, Zhang J, Liu N, Yu Y, Hou CX, Li MW. Comparative transcriptome analysis reveals different silk yields of two silkworm strains. *PLoS One* 2018; 11(5): e0155329.
9. 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.
10. Luan Y, Zuo W, Li C, Gao R, Zhang H, Tong XL, Han MJ, Hu H, Lu C, Dai FY. Identification of genes that control silk yield by RNA sequencing analysis of silkworm (*Bombyx mori*) strains of variable silk yield. *Int J Mol Sci* 2018; 19(12): 3718.
11. Luan Y, Li CL, Zuo WD, Hu H, Gao R, Zhang H, Tong XL, Lu C, Dai FY. Gene mapping reveals the association between tyrosine protein kinase *Ab11* and the silk yield of *Bombyx mori*. *Anim Genet*. 2021; 52(3): 342–350.
12. Li CL, Tong XL, Zuo WD, Hu H, Xiong G, Han MJ, Gao R, Luan Y, Lu KP, Gai TT, et al. The beta-1, 4-N-acetylglucosaminidase 1 gene, selected by domestication and breeding, is involved in cocoon construction of *Bombyx mori*. *PLoS Genet* 2020; 16(7): e1008907.
13. Li CL, Zuo WD, Tong XL, Hu H, Qiao L, Song JB, Xiong G, Gao R, Dai FY, 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: 426–32.
14. Fuji H, Kawaguchi Y, Kihara H, Banno Y, Shimada T, Abe H, Entomol AR, Asakawa S, Shimizu N, Kawasaki S, et al. Genetic Stocks and Mutations of *Bombyx mori*. *Important Genetic Resources* 1998; pp. 54.
15. Nagaraja GM, Mahesh G, Satish V, Madhu M, Muthulakshmi M, Nagaraju J. Genetic mapping of Z chromosome and identification of W chromosome-specific markers in the silkworm, *Bombyx mori*. *Heredity* 2005; 95: 148–57.
16. Zhan S, Huang JH, Guo QH, Zhao YP, Li WH, Miao XX, Goldsmith MR, Li MW, Huang YP. 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.
17. Hou CX, Wang X, Li B, Xu A, Huang YP, Li MW. Mapping and Analysis of *Bombyx mori* Sex-linked QTLs Related to Cocoon and Silk Traits. *Science of Sericulture* 2013; 39: 35–9.
18. Darvasi A, and Soller M. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor Appl Genet* 2019; 85: 353–9.
19. Xiang H, Liu X, Li M, Zhu Y, Wang L, Cui Y, Liu L, Fang G, Qian H, Xu A, Wang W, Zhan S. The evolutionary road from wild moth to domestic silkworm. *Nature Ecology and Evolution* 2018; 2: 1268–1279.
20. Sharma RK, Duda T, Makino CL. Integrative Signaling Networks of Membrane Guanylate Cyclases: Biochemistry and Physiology. *Front Mol Neurosci* 2016; 9: 83.
21. Garbers DL. Guanylate cyclase receptor family. *Recent Prog Horm Res* 1990; 46: 85–96.

22. Robison GA, Butcher RW, Sutherland EW. Adenyl cyclase as an adrenergic receptor. *Ann N Y Acad Sci* 1967; 139(3): 703–23.
23. Toda T, Uno I, Ishikawa T, Powers S, Kataoka T, Broek D, Cameron S, Broach J, Matsumoto K, Wigler M. In Yeast, RAS Proteins Are Controlling Elements of Adenylate Cyclase. *Cell* 1985; 40(1): 27-36.
24. Piermarini PM, Grogan LF, Lau K, Wang L, Beyenbach KW. A SLC4-like anion exchanger from renal tubules of the mosquito (*Aedes aegypti*): evidence for a novel role of stellate cells in diuretic fluid secretion. *Am J Physiol Regul Integr Comp Physiol* 2010; 298(3): R642-60.
25. Linser PJ, Oviedo MN, Hirata T, Seron TH, Smith KE, Piermarini PM, Romero MF. Slc4-like anion transporters of the larval mosquito alimentary canal. *J Insect Physiol*. 2012; 58(4): 551–62.
26. Romero MF, Chen AP, Parker MD, Boron WF. The SLC4 family of HCO₃ transporters. *Mol Aspects Med* 2013; 34(2-3): 159–82.
27. Yoko NM, Takashi N., Yoshimi N, Hideyuki T, Kenjiro A, Yoshiko K, Toshio M, Haruo H. Dynamic localization of α -tubulin acetyltransferase ATAT1 through the cell cycle in human fibroblastic KD cells. *Med Mol Morphol* 2018; 51(4): 217–226.
28. Aguilar A, Becker L, Tedeschi T, Heller S, Iomini C, Nachury MV. Alpha-tubulin K40 acetylation is required for contact inhibition of proliferation and cell-substrate adhesion. *Mol Biol Cell* 2014; 12: 1854–1866.
29. Kalebic N, Martinez C, Perlas E, Hublitz P, Bilbao-Cortes D, Fiedorczuk K, Andolfo A, Heppenstall PA. Tubulin acetyltransferase α TAT1 destabilizes microtubules independently of its acetylation activity. *Mol Cell Biol* 2013; 33: 1114–1123.
30. Nakakura T, Asano-Hoshino A, Suzuki T, Arisawa K, Tanaka H, Sekino Y, Kiuchi, Y., Kawai K, Hagiwara H. The elongation of primary cilia via the acetylation of alpha-tubulin by the treatment with lithium chloride in human fibroblast KD cells. *Med Mol Morphol* 2015; 48: 44–53.
31. Nakakura T, Suzuki T, Nemoto T, Tanaka H, Asano-Hoshino A, Arisawa K, Nishijima Y, Kiuchi Y, Hagiwara H. Intracellular localization of alpha-tubulin acetyltransferase ATAT1 in rat ciliated cells. *Med Mol Morphol* 2016; 49: 133–143.
32. Piperno G, LeDizet M, Chang XJ. (1987) Microtubules containing acetylated alpha-tubulin in mammalian cells in culture. *J Cell Biol* 104, 289–302.
33. Hubbert C, Guardiola A, Shao R, Kawaguchi Y, Ito A, Nixon A, Yoshida M, Wang XF, Yao TP. HDAC6 is a microtubule-associated deacetylase. *Nature* 2012; 417: 455–458.
34. Akella JS, Wloga D, Kim J, Starostina NJ, Lyonsabbott S, Morrissette NS, Dougan ST, Kipreos ET, Gaertig J. MEC-17 is an α -tubulin acetyltransferase. *Nature* 2010; 467: 218–222.
35. Appling DR. Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. *FASEB J* 1991; 5: 2645–51.
36. Fox JT, Stover PJ. Folate-mediated one-carbon metabolism. *Vitam Horm* 2008; 79: 1–44.
37. Farber S, Diamond LK, Mercer RD, Sylvester RF, Wolff VA. Temporary remission in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl glutamic acid (aminopterin). *N Engl J*

- Med 1948; 238: 787–93.
38. Chattopadhyay S, Moran RG, Goldman DI. Pemetrexed: biochemical and cellular pharmacology, mechanisms, and clinical applications. *Mol Cancer Ther* 2007; 6: 404–17.
 39. Carreras CW, Santi DV. The catalytic mechanism and structure of thymidylate synthase. *Annu Rev Biochem* 1995; 64: 721–62.
 40. Wu D, Li Y, Song GJ, Cheng CY, Zhang RG, Joachimiak A, Shaw N, Liu ZJ. Structural Basis for the Inhibition of Human 5,10-Methenyltetrahydrofolate Synthetase by N10-Substituted Folate Analogues. *Cancer Res* 2009; 69(18): 7294–301.
 41. Field SM, Anderson DD, Stover PJ. Mthfs is an essential gene in mice and a component of the purinosome. *Front Genet* 2011; 2: 36.
 42. Boutanaev AM, Kalmykova AI, Shevelyov YY, Nurminsky DI. Large clusters of co-expressed genes in the *Drosophila* genome. *Nature* 2002; 420: 666–669.
 43. Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter SW, Ferguson L, et al. (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*. 477, 203–206.
 44. Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, Baxter S, Abanto M, Bermingham E, Humphray SJ, Rogers J, et al. A Conserved Supergene Locus Controls Colour Pattern Diversity in *Heliconius* Butterflies. *PLoS Biol* 2006; 4(10): e303.
 45. Küpper C, Stocks M, Risse JE, Remedios ND, Farrell LL, McRae SB, Morgan TC, Karlionova N, Pinchuk P, Verkuil YI, et al. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat Genet* 2016; 48(1): 79–83.
 46. Li CL, Tong XL, Zuo WD, Luan Y, Gao R, Han MJ, Xiong G, Gai TT, Hu H, Dai FY, et al. QTL analysis of cocoon shell weight identifies BmRPL18 associated with silk protein synthesis in silkworm by pooling sequencing. *Sci Rep* 2017; 7(1), 17985.
 47. Fang L, Wang Q, Hu Y, Jia YH, Chen JD, Liu BL, Zhang, ZY, Guan XY, Chen SQ, Zhou BL, et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat Genet* 2017; 49(7): 1089–1098.

Figures

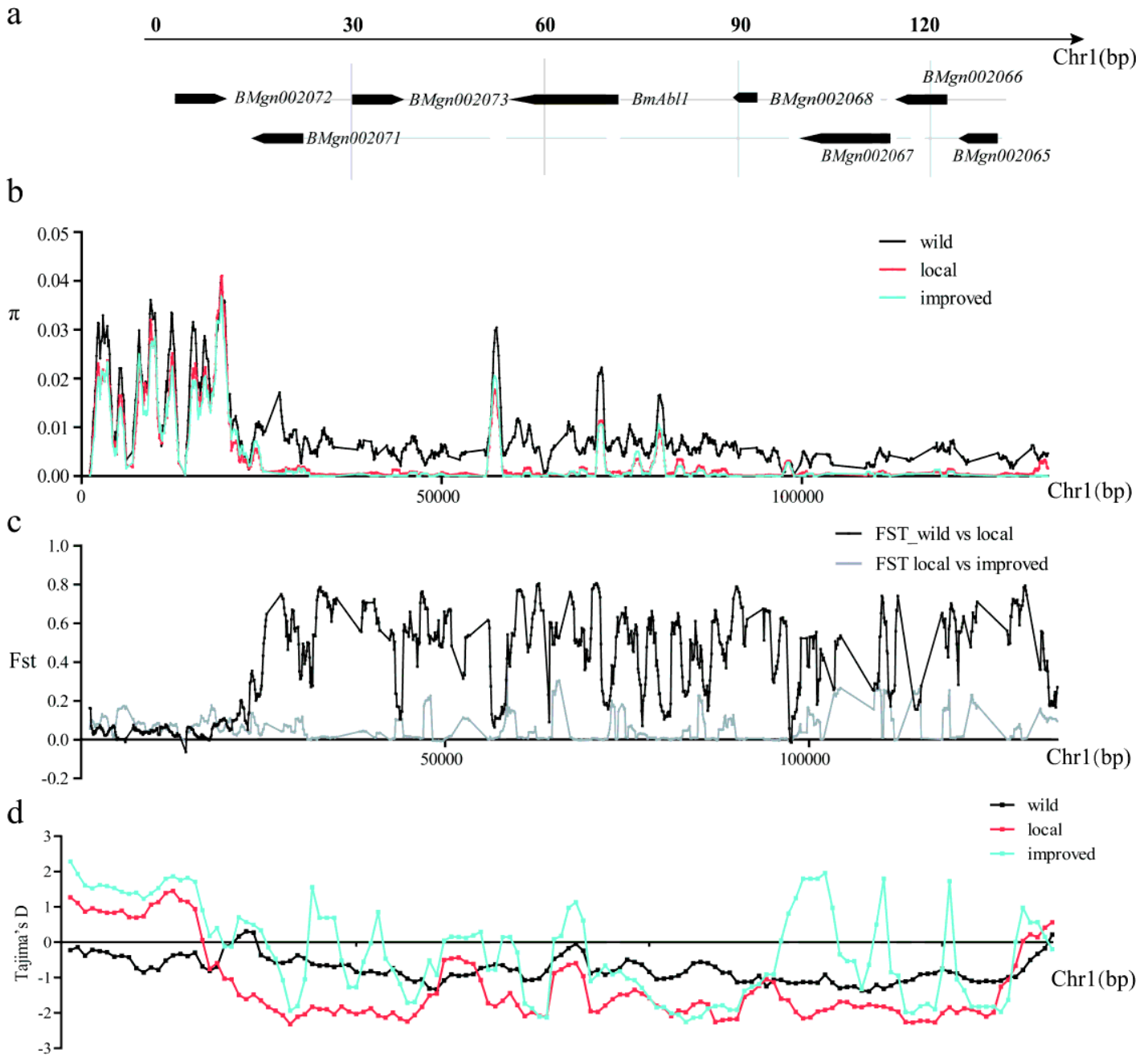


Figure 1

Selective sweeping of Chr1 0–134 kb. a. The black arrow marks the position of the terminal portion of Chr1, and the black pentagon represents the positions and numbers of genes contained in the region. b. Nucleotide diversity (π) of wild silkworms, local strains, and improved strains. The black line represents wild silkworms; the red line represents local strains, and the green line represents improved strains. The x-axis represents the physical location of Chr1. c. The F_{st} values of wild vs local strains and local vs improved strains. The black line represents the F_{st} between wild silkworms and local strains; the gray line represents the F_{st} between local strains and improved strains, and the x-axis represents the physical location of Chr1. d. Tajima's D analysis of the candidate region. The black line represents the wild strain population; the red line represents the local strain population, and the green line represents the improved strain population. The gray background color marks the purifying selection region.

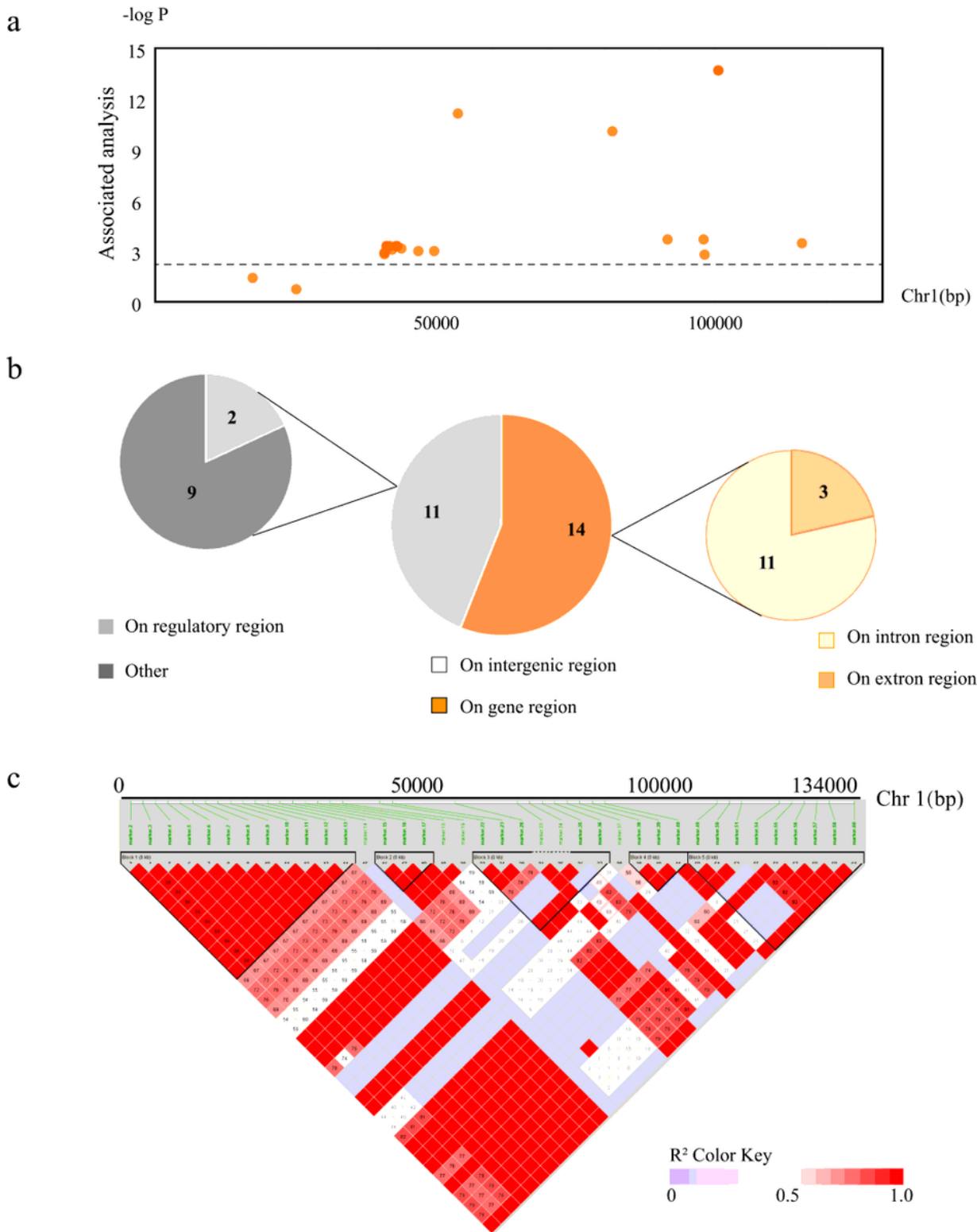


Figure 2

Analysis of association and linkage disequilibrium of Chr1 0–134k a. The x-axis represents the physical location of Chr1; a dot represents an indel mark, and a black box represents the degree of linkage between the marks. The y-axis shows the degree of association with the CSW. The threshold value calculation method is $p_value/number\ of\ markers = 0.05/25 = 0.002$ and is marked with a black dotted line. b. The percentage of markers located in the region. Gray represents the markers located in the

intergenic region, and orange represents the markers located in the gene region. The pie chart on the right is a detailed view of the markers of the gene region. The dark orange segment represents the proportion of markers located in the exon region among the markers in the gene region, and the light orange represents the proportion of markers located in the intron region. The pie chart on the left is a detailed view of the intergenic region in which the light gray part represents the proportion of markers located in the regulatory region to the number of markers in the intergenic region. c. Linkage disequilibrium heat map with 38 indel data as markers. The red box represents the tight linkage between the markers.

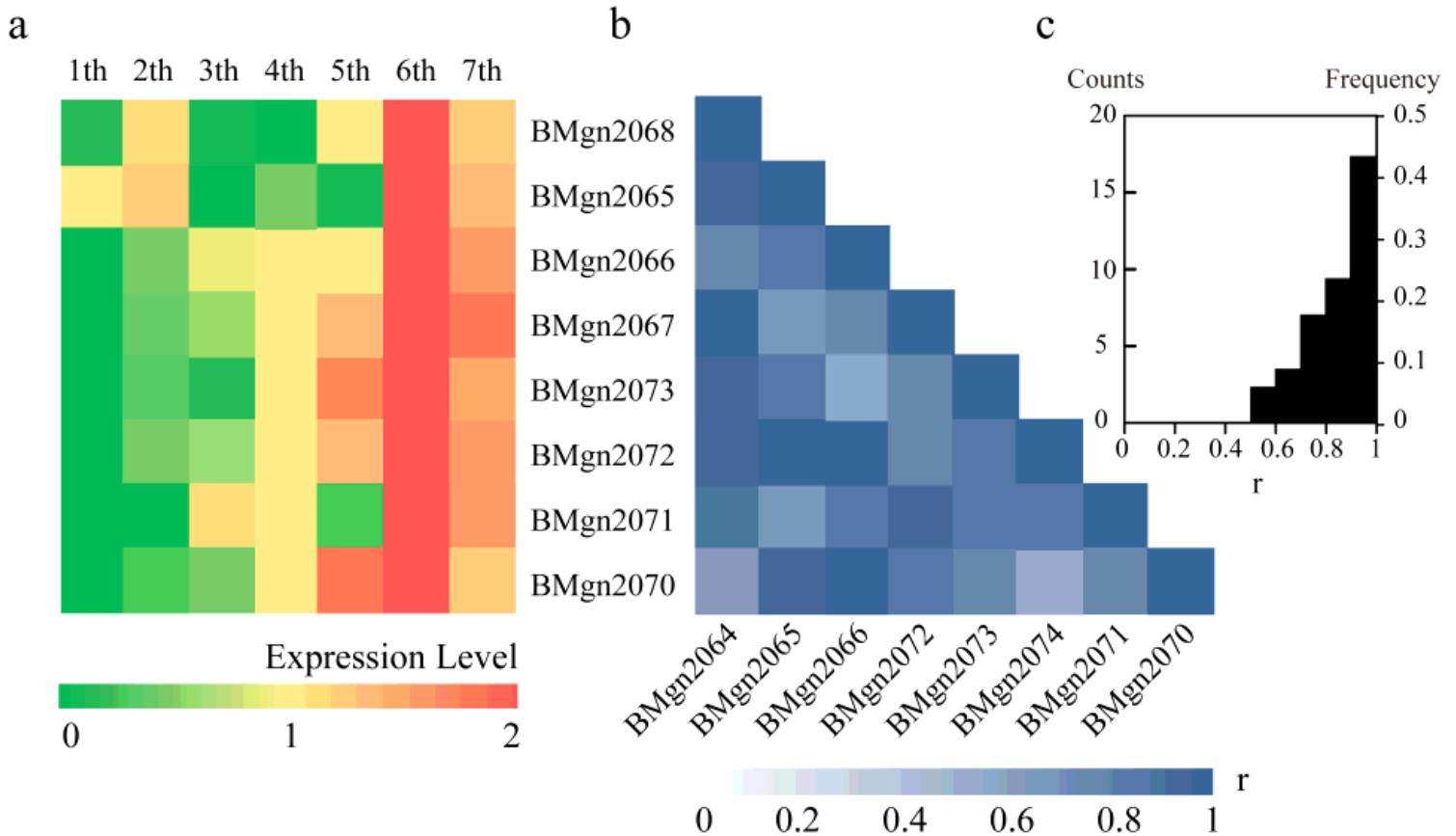


Figure 3

Expression profiles of eight genes in the candidate region. a Heat map of expression profiles of eight genes in the candidate region. Each row represents a gene, and the gene number is marked on the right side of the heat map. Each column represents a time point, from day 1 of the 5th instar to day 7 of the 5th instar, marked at the top of the heat map. The heat map colors represent the expression levels of the genes. Red represents a high expression level, and green represents a low expression level. The expression color scale is placed on the lower side of the heat map as a reference for the expression level. b Correlation heat map of the expression level of eight genes. Both the abscissa and ordinate are eight genes in the region. x-axis: Gene ID BMgn002068, BMgn002065, BMgn002066, BMgn002067, BMgn002071, BMgn002072, BMgn002073, and BMgn002070, from left to right; y-axis: Gene ID from top to bottom: BMgn002068, BMgn002065, BMgn002066, BMgn002067, BMgn002071, BMgn002072, BMgn002073, and BMgn002070. Each small square represents the Pearson's correlation coefficient (r) value of the expression profiles of a pair of genes. The darker the color, the stronger the correlation. The r

values were used to generate the color scale, and these are listed on the lower side of the heatmap as a numerical reference. c Frequency histogram of the Pearson's correlation coefficient (r) of gene expression levels. The x-axis represents the r value, and the y-axis on the left represents the frequency represented by a histogram.

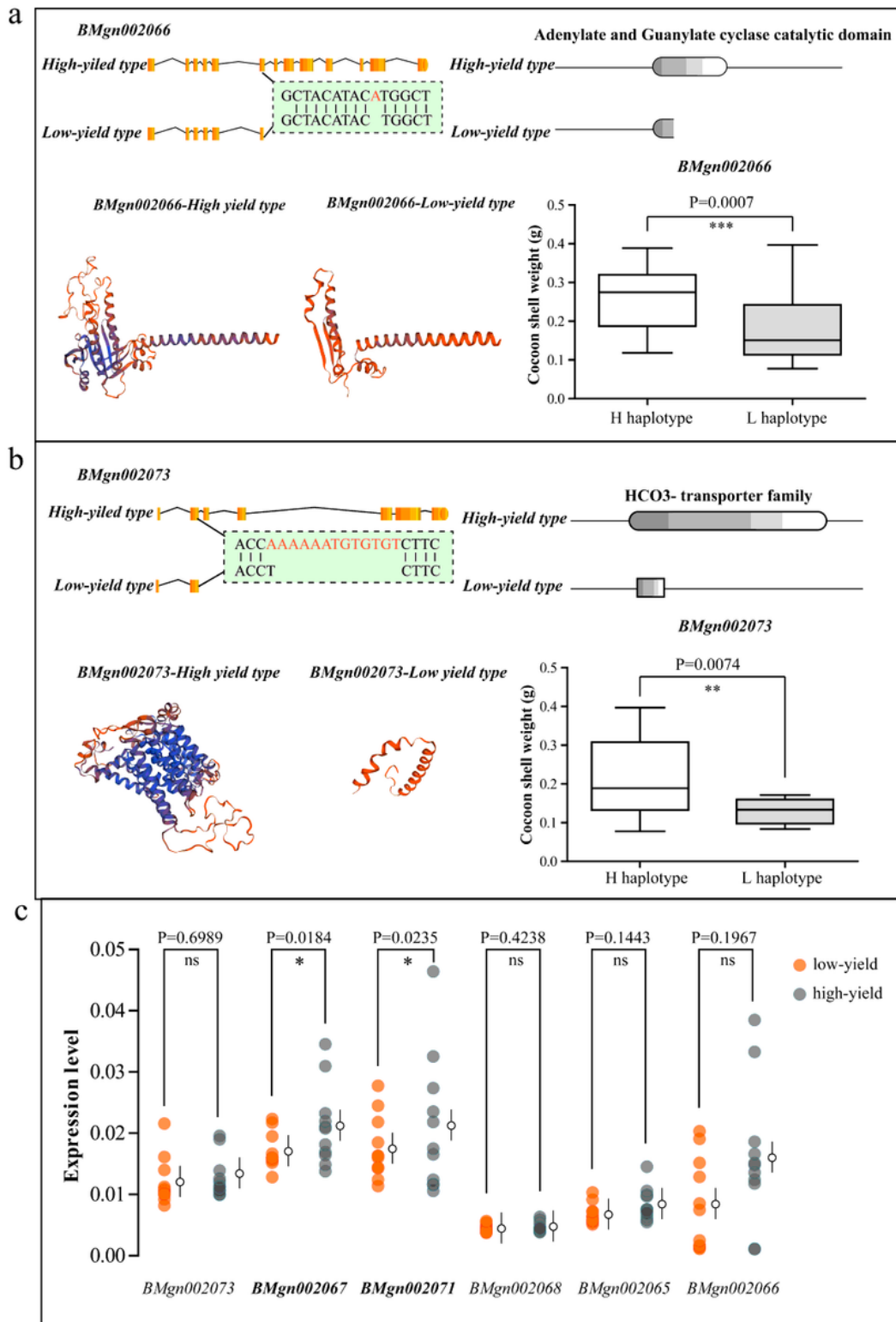


Figure 4

Genes are associated with CSW a. Sequence variation, domain variation, protein tertiary structure variation, and genotype association analysis of BMgn002066. The orange boxes in the upper left figure represent exons, and the black lines represent introns. The gray bars in the upper right figure represent the domains. The lower left picture shows the changes in the protein tertiary structures of high-yield and low-yield strains. The lower right figure shows the association analysis between sequence variation and CSW. The white box represents the strains with the H haplotype sequence; the gray box represents the strains with the L haplotype sequence; the y-axis represents CSW, and the significance is marked on the top of the box plot. b. Sequence variation, domain variation, protein tertiary structure variation, and genotype association analysis of BMgn002073. c. The expression levels of genes among multiple strains. The data are divided into six groups by gene number, and the gene number is indicated in the lower part of the figure. Each group consists of two columns: high-yield strains and low-yield strains. Each green dot represents a high-yield strain, and each red dot represents a low-yield strain. The y-axis is the expression level of the gene in the posterior silk gland of each strain. Circles represent the average expression levels. The significance degree of the difference in expression level between high-yield strains and low-yield strains is marked above the gene; '*' means significant, and 'ns' means not significant.

Supplementary Files

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

- [supportingTable1.xlsx](#)
- [supportingTable2.xlsx](#)