



OPEN ACCESS

EDITED BY

Muniyandi Nagarajan,
Central University of Kerala, India

REVIEWED BY

Luc Swevers,
National Centre of Scientific Research
Demokritos, Greece
Kang He,
Zhejiang University, China

*CORRESPONDENCE

Yongzhen Huang,
✉ hyzsci@nwfufu.edu.cn
Shang Yang,
✉ yangshang416235551@163.com

[†]These authors have contributed equally to this work

RECEIVED 21 February 2025

ACCEPTED 28 May 2025

PUBLISHED 10 June 2025

CITATION

Song X, Lyu Y, Zhang Y, Ding W, Huang Y and Yang S (2025) Genome-wide association study of cocoon-producing traits in four Henan silkworm strains revealed new candidate genes. *Front. Genet.* 16:1580806. doi: 10.3389/fgene.2025.1580806

COPYRIGHT

© 2025 Song, Lyu, Zhang, Ding, Huang and Yang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome-wide association study of cocoon-producing traits in four Henan silkworm strains revealed new candidate genes

Xingya Song^{1,2†}, Yingxian Lyu^{3†}, Yuqiao Zhang², Wenjia Ding², Yongzhen Huang^{2*} and Shang Yang^{1*}

¹Henan Academy of Sericulture Science, Zhengzhou, Henan, China, ²College of Animal Science and Technology, Northwest Agriculture and Forestry University, Yangling, Shaanxi, China, ³Henan Agricultural Society, Zhengzhou, Henan, China

Background: The silkworm (*Bombyx mori*), has been farmed in China for over 5,000 years, and holds significant economic value. Recent genomic advances have deepened our understanding of silk production mechanisms, enabling the development of improved silkworm breeds. Further research is needed to identify effective molecular markers linked to silk production traits, as this will enhance genetic improvement efforts.

Results: In the genome-wide association analysis study, we identified 40 significant single nucleotide polymorphisms (SNPs) and 28 candidate genes which were related to cocoon shell weight, cocoon width and whole cocoon weight in four silkworm strains in Henan. Through linkage imbalance block analysis, we found that the *KWMTBOMO02490* (ATPase inhibitor-like protein) and *KWMTBOMO12678* (OTU domain-containing protein 7B) were strongly selected. By using online databases, we found that the *KWMTBOMO02490* gene was highly expressed in the posterior silk glands and may play an important role in the process of silk production.

Conclusion: The candidate genes in this study may affect the production and health traits of silkworms, and may provide an important reference for silkworm breeding.

KEYWORDS

genome-wide association study, cocoon-producing traits, Henan silkworm strains, candidate genes, *KWMTBOMO02490*

1 Background

The silkworm (*Bombyx mori*) is one of the insects with high economic value. Its main product, silk, played an important role in the development of the “Silk Road”, the first large-scale commercial exchange between the East and the West in the world history. According to historical records, China’s sericulture has an ancient origin of 5,000 years. Silkworms likely originated in China as trimoulting lines and diversified into local strains through independent spreads along the Silk Road (Xiang et al., 2018). China’s long history of sericulture established its industry in an important position in the world.

Silk is the most important product of silkworms, and it is very important to unravel the genetic mechanism of silk-production. In recent years, China has accelerated the research on silkworm genomics, which is now at the forefront of the world. In 2003, two teams

independently performed whole genome sequencing (WGS) of the domesticated silkworm genome and published the draft sequence for the silkworm genome (Xia et al., 2004; Mita et al., 2004). In 2009, Southwest University conducted whole genome resequencing of 29 strains of domesticated silkworms and 11 strains of wild silkworms, constructed a high-precision genetic variation map of the nuclear genome of silkworms, and found that the domestic silkworm had successfully differentiated from the wild silkworm after 5,000 years by phylogenetic analysis (Xia et al., 2009). Genome-based mapping studies have identified many important genes associated with mutations in body color (Futahashi et al., 2008), cocoon color (Sakudoh et al., 2007), development (Lu Z. et al., 2019), disease resistance (Yang et al., 2017), and other traits. A study on the domestication from wild silkworms to domestic silkworm has for the first time applied genome-wide association analysis (GWAS) to the mapping of important traits (cocoon color, blood color, disease resistance et al.) in the chromosomes of *B. mori*, confirming the feasibility and scientific validity of the methodology (Xiang et al., 2018). In 2022, the genome analysis of large-scale silkworm germplasm resources (“Thousand Silkworm Genome”) has been completed resulting in 87 the drawing of a super pan-genome map of silkworms (Tong et al., 2022). It provides a large amount of genomic information for the identification of important silkworm genes and the selection and breeding of silkworms. For silk production traits, we need to find more effective molecular markers to provide a scientific basis for genetic improvement of silkworms.

In this study, phenotypic statistics and whole-genome sequencing were performed on four silkworm strains in Henan Province with the aim of screening candidate genes for silkworm cocoon traits to help understand the genetic basis of silkworm production.

2 Materials and methods

2.1 Samples and phenotype collection

A total of 120 silkworms of the four strains were provided by the Henan Academy of Sericulture, and all of them were maintained under identical husbandry and management conditions. The four strains are Cuihua (CH, $n = 30$), Chunhui (CHH, $n = 30$), Chunyun A (CHY, $n = 30$), and Huayu A (HY, $n = 30$). Thirty individuals were selected from each strain, and the traits including whole cocoon weight (WCW), cocoon shell weight (CSW), cocoon length (CL), and cocoon width (CW) were measured. We used electronic analytical balances to measure WCW and CSW, and vernier calipers to measure CL and CW. Genomic DNA of each individual was extracted by a standard phenol-chloroform extraction method.

2.2 Sequencing, mapping and SNP calling

Genomic sequencing of these samples was performed by Huazhi Rice Bio-Tech Company (Changsha, China) using DNBSEQ-T7. Reads of 300 bp of qualified DNA samples were used to construct DNA libraries. Clean data were obtained by FASTP (Chen et al., 2018) after removing low quality reads, adaptors, and reads with multiple N bases.

Read quality was assessed through FastQC (v0.11.9). Subsequently, clean reads were mapped to the silkworm reference genome (Kawamoto et al., 2019) using Sentieon (Freed et al., 2017). QualiMap (v2.2.1) (Okonechnikov et al., 2016) was used to evaluate sequencing alignment data. Each sample was detected for SNPs by Sentieon (Freed et al., 2017), and then these samples are combined into a variant call format (VCF) file and filtered with parameters including “QD < 2.0, FS > 60.0, MQ < 40.0, SOR > 3.0, MQRankSum < -12.5, ReadPosRankSum < -8.0”. Then, we used ANNOVAR (Wang et al., 2010) to annotate these SNPs.

The quality of the resulting SNP genotype data was improved by the removal of SNPs with detection rate < 95%, and criteria such as minor allele frequency (MAF) < 0.05, and a Hardy-Weinberg Equilibrium (HWE) P -value < $1e-4$. Finally, we used PLINK (version 1.90) (Purcell et al., 2007) to filter the SNP data by “--indep-pairwise 500 50 0.5” parameters to remove the SNPs which have high linkage disequilibrium (LD) values. This procedure can reduce the amount of non-essential computation.

2.3 Population structure analysis

Since the silkworms involved in this study came from different populations, distinct subgroups may exist among these silkworms. We used principle component analysis (PCA) to explore the stratification of the populations by smartPCA (Patterson et al., 2006), and PopLDdecay (Zhang et al., 2019) was used to explore the LD degree.

2.4 GWAS analysis

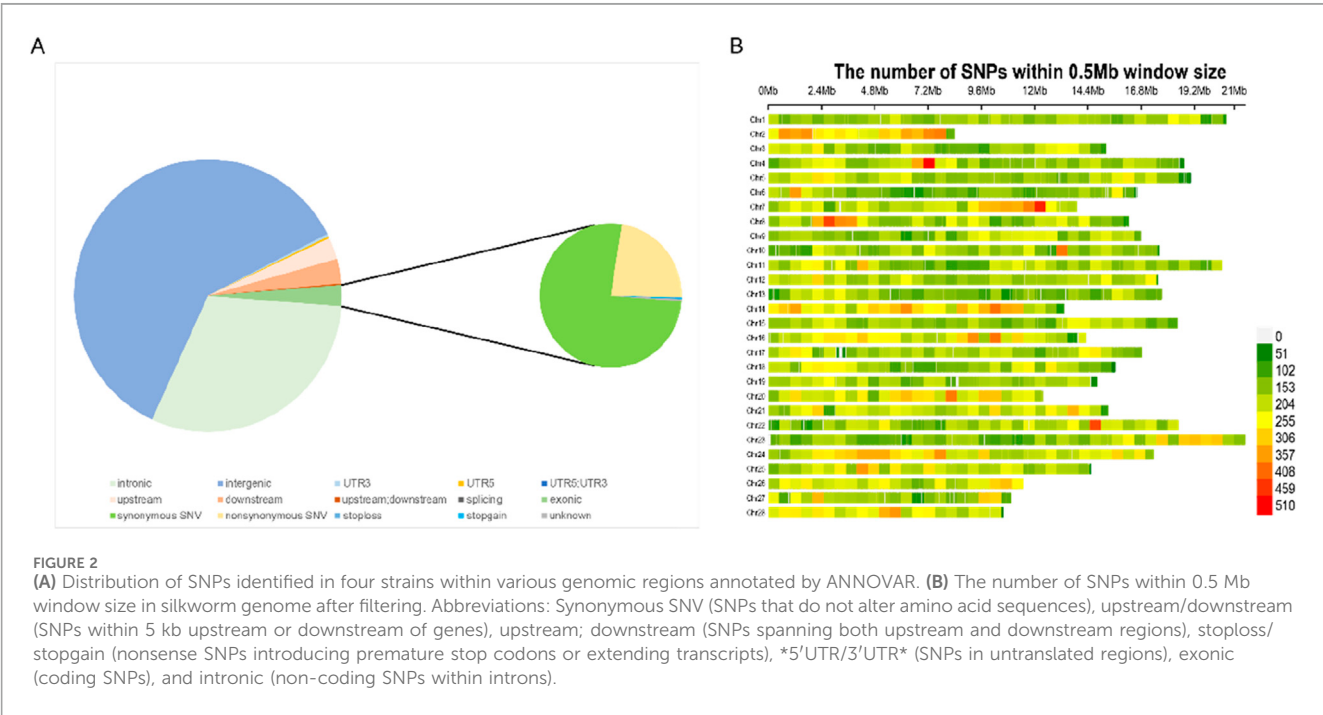
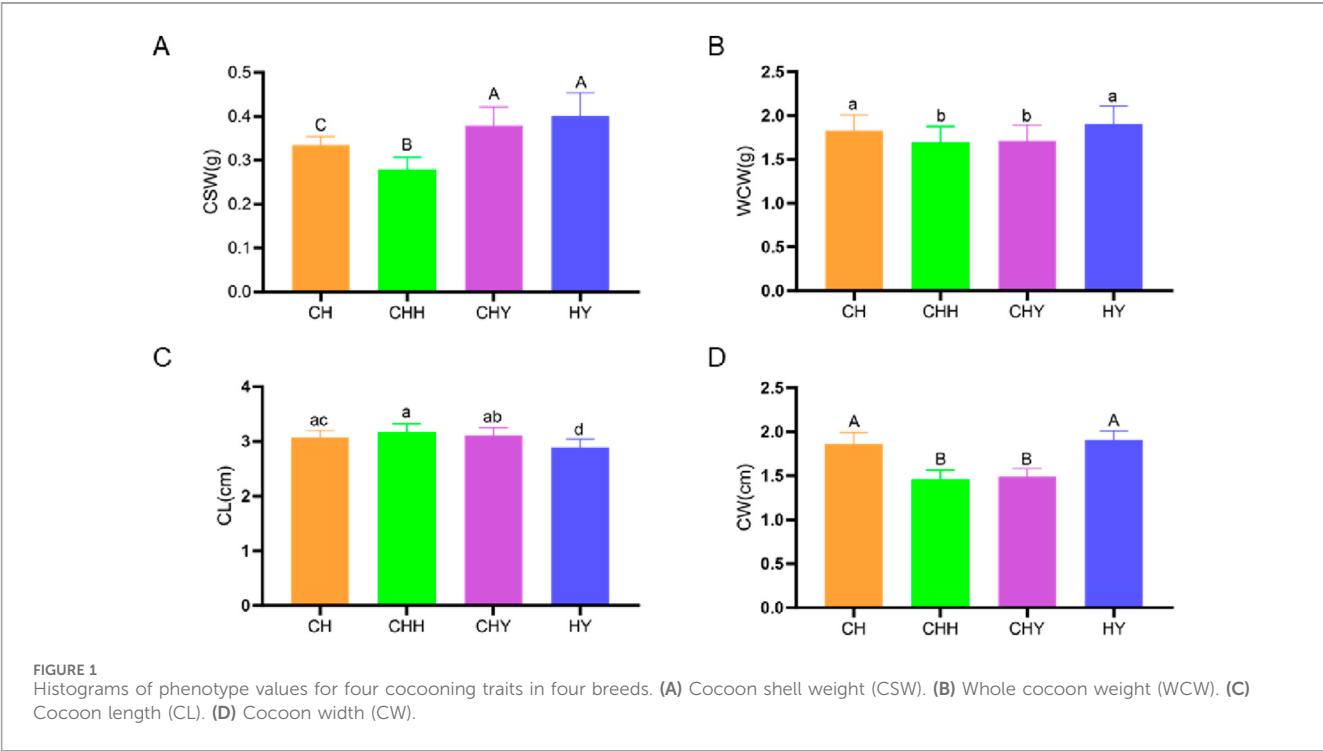
We used the Linear Mixed Model (LMM) model of GEMMA (v0.98.5) (Zhou and Stephens, 2012) software for GWAS analysis. If points in quantile-quantile (Q-Q) plots are off the diagonal (especially at the left end), population stratification or false positives in the model are suggested. λ corresponds to the genome inflation factor, with λ closer to 1 indicating a better model fit. Each point in the Manhattan plot was used to display SNPs, in which values in the vertical coordinate represent the strength of the association of the SNP with the phenotype. In order to reduce the effect of population stratification on GWAS, we used the first three principal components as covariates. The specific statistical model of GWAS is as follows:

$$y = SNP + COV + Kin + e$$

With y the phenotype value, Kin the kinship matrix, SNP the SNP marker effect, COV is the covariate, and e is the residual effect. The significance threshold (the red dotted lines in Manhattan plots) was calculated using the Bonferroni correction method ($0.05/175,704 = 2.845e-7$).

2.5 Gene annotation and LD blocks plotting

Genes within a 10 kb region centered on the significant SNPs were identified using BEDTools (Quinlan and Hall, 2010). The 10 kb



range was selected based on the gene annotation region in the previous study (5 kb each upstream and downstream) (Lu et al., 2023). We used LDBlockshow (Dong et al., 2021) to map LD blocks in the 10 kb upstream and downstream range of genes. R^2 represents the degree of association, and the closer the value is to 1, the redder the color in the LD blocks and the higher the degree of linkage disequilibrium.

3 Result

3.1 Phenotypic values and sequencing data

The phenotypic data of the 120 samples were screened for outliers, and the results showed that all the data were within a reasonable range (Figure 1). It can be seen that there are significant

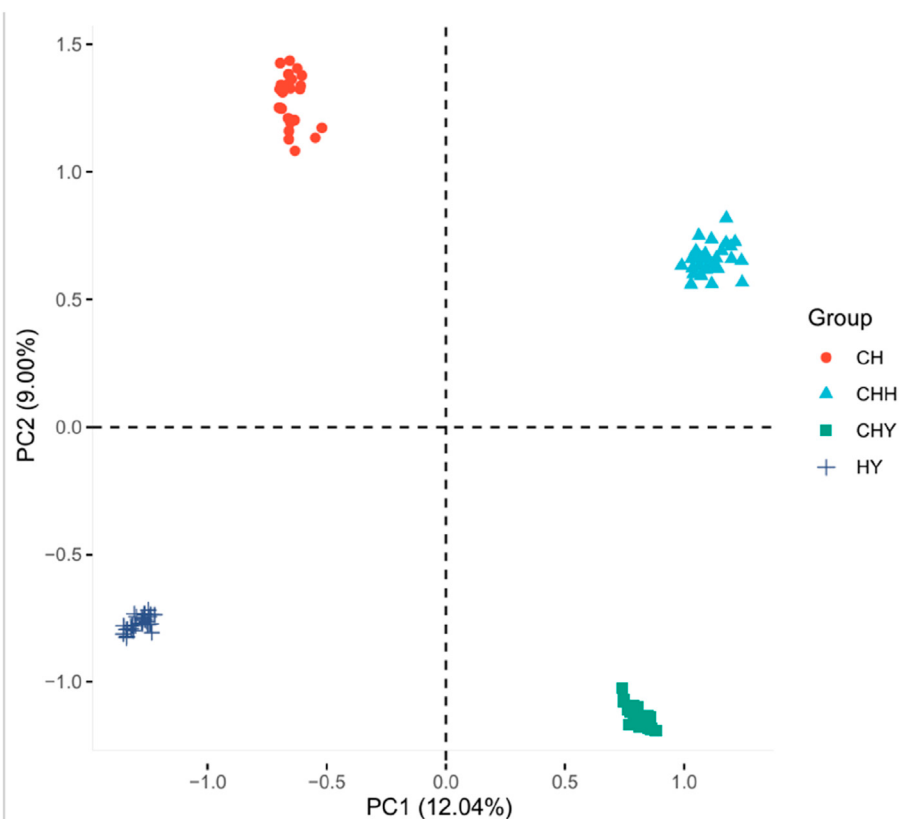


FIGURE 3
Principal component analysis (PCA) of four Henan silkworm strains. The first three principal components (PC1, PC2, PC3) explain 12.04%, 9.00%, and 7.37% of the total genetic variance, respectively. Each point represents an individual, colored by strain.

differences in the four traits among the four breeds of silkworm. The differences provide a basis for the conduction of GWAS.

A total of 772.51 Gb of data were obtained after sequencing. After the raw data were filtered out by FASTP (Chen et al., 2018), the clean data of all samples totaled 772.51 Gb, with an average of 6.43 Gb per sample, and the average of the cleanQ30 quality score was 93.98%. The sequencing depth of these samples ranged from 10.79× to 18.7×, with a mean depth of 12.47×. After the calling and filtering, we obtained 6,063,447 SNPs. The resulting SNPs were annotated and it was found that 2.49% of them were located in the exon region (Figure 2A). In the end, we obtained 175,704 SNPs after genotypic filtration and LD pruning. The final SNPs were evenly distributed in the silkworm genome (Figure 2B).

3.2 Population stratification analysis

PCA results based on genome-wide SNP data showed that population stratification might exist among these silkworms (Figure 3). The first PC explained 12.04% of the variation, and the second PC explained 9.00% of the variation. In addition, we found that the third PC also had a higher percentage, at 7.37%. Therefore, the first three PCs need to be

used as covariates in the GWAS model to explain the population stratification effect.

3.3 Significant SNPs and candidate genes

The LMM model was used to analyze the association between the whole genome of the silkworms and the four economic traits, and the relationship between population stratification and individuals was considered. The Manhattan plots and Q-Q plots of the GWAS results based on the LMM are shown in Figure 4. The inflation factors (λ) for the four traits were 1.011, 0.933, 1.001 and 0.975. Based on the results of Q-Q plots and inflation factor, some errors are expected in the model of cocoon width trait, while the model of CSW trait is more accurate than other traits. These results showed that the introduction of the first three principal components as covariates effectively eliminated the effect of population stratification.

The significance threshold obtained by Bonferroni correction is $2.845e-7$. Below this threshold, we detected three significant SNPs for the cocoon weight trait, two significant SNPs for the CSW trait, and 35 significant SNPs for the WCW trait (Table 1). By searching within 5 kb upstream and downstream of these significance SNPs, we obtained 28 related genes (Table 1).

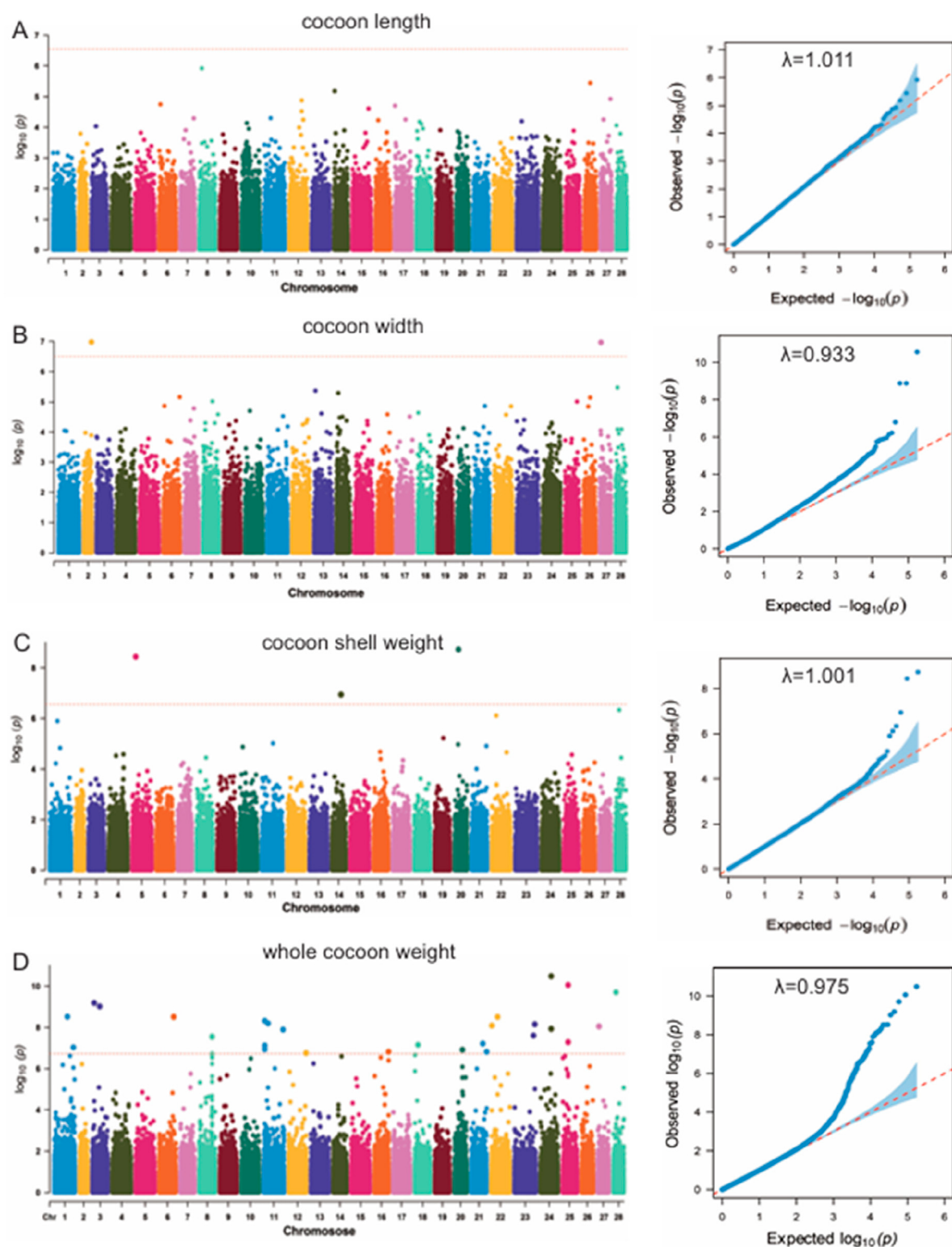


FIGURE 4

(A) Manhattan plot and Q-Q plot of cocoon length (CL) trait. (B) Manhattan plot and Q-Q plot of cocoon width (CW) trait. (C) Manhattan plot and Q-Q plot of cocoon shell weight (CSW) trait. (D) Manhattan plot and Q-Q plot of whole cocoon weight (WCW) trait.

In the LD block analysis, we found that there was the high linkage disequilibrium degree near the Chr5: 3377758 and Chr21: 11446386 (overlapping with the GWAS-significant SNPs for CSW

and WCW), as shown in Figures 5A,B. It indicates that these two regions were strongly selected in the process of silkworm breeding. These two LD blocks contained the candidate genes

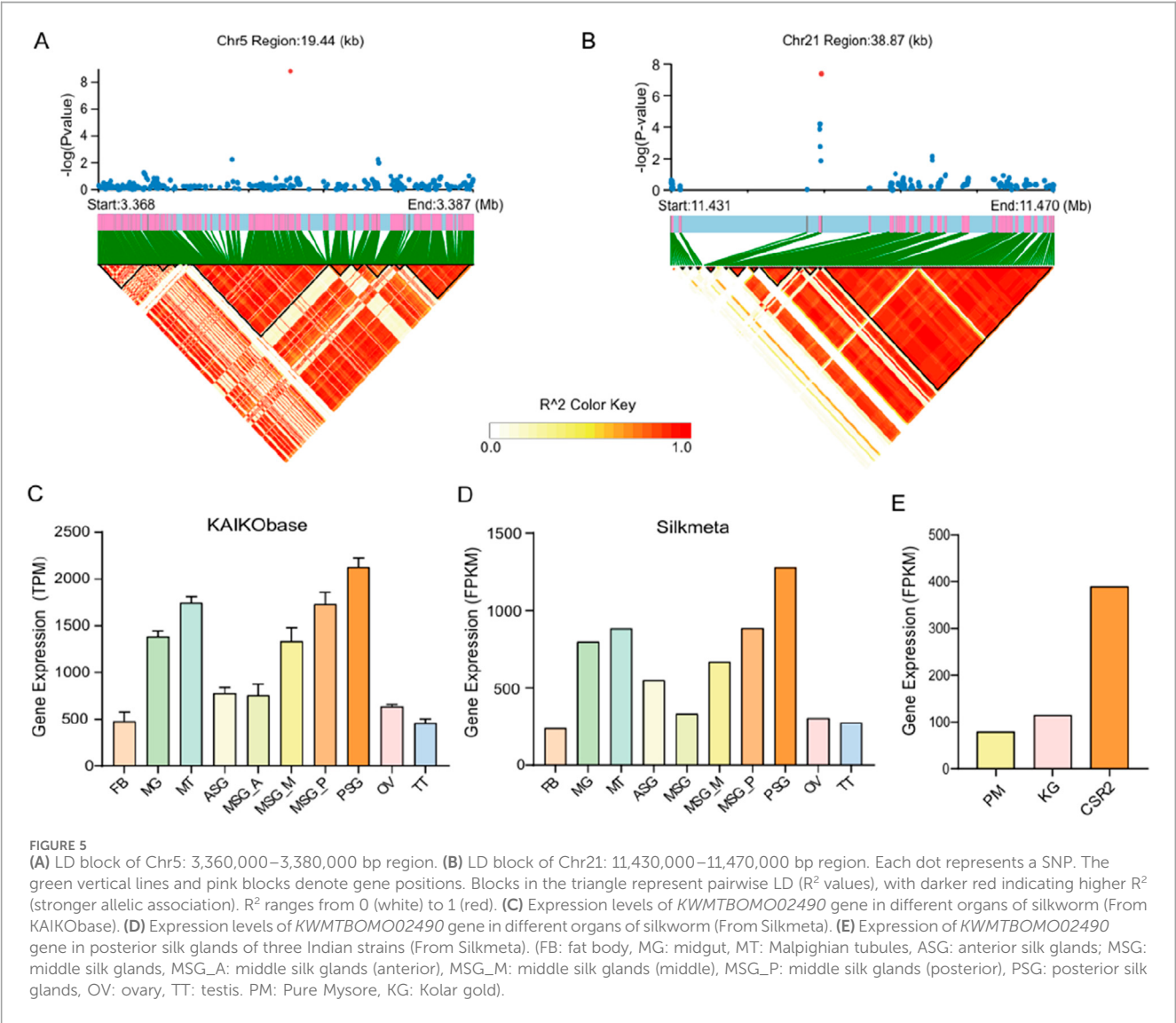
TABLE 1 Significant SNPs and their information from genome-wide association analysis.

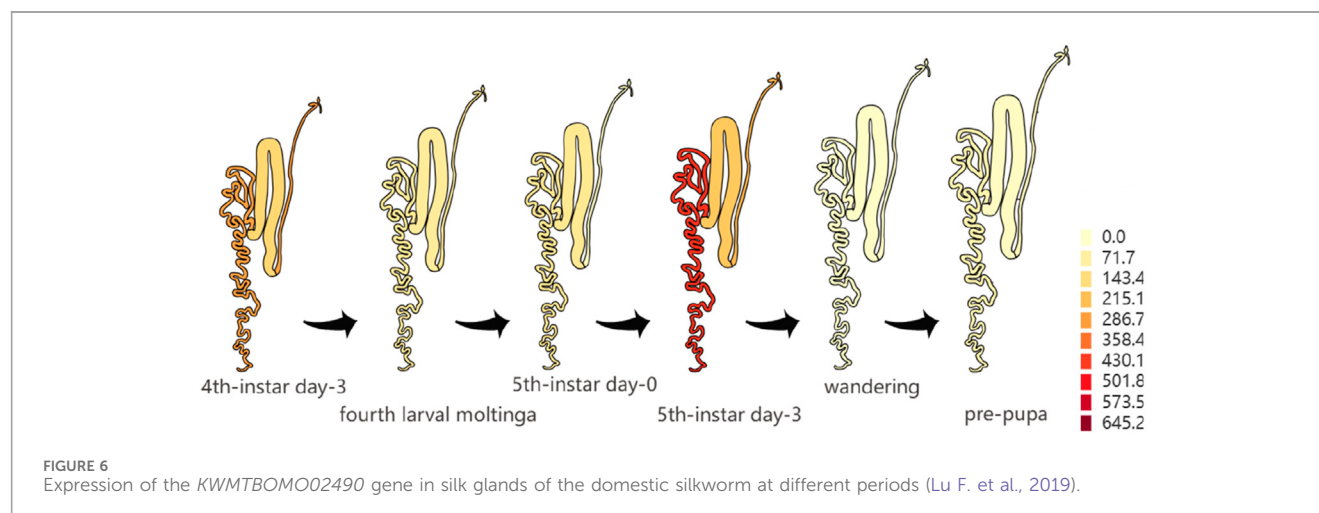
Trait	Chr	Position	P-value	Gene ID	Description
CW	2	7653741	1.04E-07	KWMTBOMO00968	no hit
				KWMTBOMO00969	uncharacterized LOC114250257 isoform X1
	27	1126676	1.07E-07		
CSW	5	3377758	2.86E-09	KWMTBOMO02490	ATPase inhibitor-like protein
				KWMTBOMO02491	uncharacterized LOC114241584
	14	8763775	1.18E-07	KWMTBOMO08425	synaptotagmin-10-like isoform X2
	20	3882656	1.48E-09	KWMTBOMO11938	microvitellogenin-like
				KWMTBOMO11939	LOC101744209
				KWMTBOMO11940	microvitellogenin-like
WCW	1	13219760	3.03E-09	KWMTBOMO00047	gamma-soluble NSF attachment protein
	1	15848686	2.39E-07	KWMTBOMO00539	GRB10-interacting GYF protein 2
	1	19161216	9.22E-08	KWMTBOMO00617	protein angel
	3	1854481	6.53E-10	KWMTBOMO01088	fatty-acid amide hydrolase 2-like
	3	7250590	9.59E-10	KWMTBOMO01276	inhibitor of growth protein 3
	6	14325958	3.08E-09	KWMTBOMO03645	LOC114358757
	8	12957991	2.78E-07	KWMTBOMO03581	uncharacterized LOC114253097
	8	12979488	2.89E-08		
	8	12987280	2.30E-07		
	8	12996278	2.74E-08		
	8	13072500	1.85E-07	KWMTBOMO04710	ubiquitin carboxyl-terminal hydrolase 5
	11	1153628	1.15E-07		
	11	1162600	4.78E-09		
	11	1166658	7.50E-08		
	11	4472852	6.28E-09		
	11	19350288	1.27E-08		
	12	17277581	1.75E-07	KWMTBOMO07549	WD repeat-containing protein 26
				KWMTBOMO07550	trafficking protein particle complex subunit 4
	14	8327852	2.53E-07		
	16	14233874	1.51E-07		
	18	365866	2.18E-07	KWMTBOMO10679	hypothetical protein B5V51_11010; partial
	18	3367456	6.95E-08	KWMTBOMO10803	uncharacterized LOC110385585
				KWMTBOMO10804	uncharacterized LOC114252976
	20	8101424	1.21E-07	KWMTBOMO12129	Retrovirus-related Pol polyprotein from transposon TNT 1–94
	21	11446386	6.05E-08	KWMTBOMO12678	OTU domain-containing protein 7B
	21	15275468	1.52E-07		
	22	721868	8.13E-09	KWMTBOMO12824	uncharacterized LOC105842570
	22	6354809	3.10E-09	KWMTBOMO12987	histidine-rich glycoprotein-like

(Continued on following page)

TABLE 1 (Continued) Significant SNPs and their information from genome-wide association analysis.

Trait	Chr	Position	P-value	Gene ID	Description
	23	19027987	2.46E-08		
	23	20344373	7.03E-09	KWMTBOMO14200	fatty acyl-CoA reductase wat-like
	24	10672189	1.17E-08		
	24	10672312	3.26E-11		
	25	2533920	2.46E-07		
	25	5634141	5.14E-08	KWMTBOMO15132	zinc finger protein 420 isoform X2
				KWMTBOMO15133	uncharacterized LOC101738733
	25	5641640	8.77E-11		
	27	1361478	8.87E-09		
	28	2794335	1.98E-10		





KWMTBOMO02490 (ATPase inhibitor-like protein) and *KWMTBOMO12678* (OTU domain-containing protein 7B).

4 Discussion

Analyzing the genetic factors of cocoon production is critical for advancing silkworm breeding. GWAS is an important method to identify the genes that regulate economic traits, which can provide new ideas for studying the genetic background of economic traits in silkworms. By GWAS, Tong et al. identified a cluster of tandemly arranged genes coding for sugar transporter proteins (Str) in silkworms, which are synergistically involved in the uptake of flavonoids in silkworms through ‘dose-sharing’ and thus determine the formation of green cocoons (Lu et al., 2023). *KWMTBOMO02490* gene, one of the genes related to the CSW of silkworms in this study, is located in a genome region with a high degree of linkage disequilibrium. It indicated that this gene has been selected with high intensity in the breeding process of domestic silkworms. Notably, we searched three online databases [Silkmeta (Lu et al., 2024), KAIKObase (Shimomura et al., 2009) and SilkDB3.0 (Lu F. et al., 2019)] and found that *KWMTBOMO02490*, has high expression in the posterior silk glands which are the organs that produce fibroin protein (Figures 5C,D). Otherwise, we used transcriptome data of the silk glands from three Indian strains of silkworm in Silkmeta database and found that the expression level of this gene was very different in these three strains. CSR2 (a strain with high CSW) has the highest expression level among these three strains (Figure 5E). Moreover, we found that this gene was also highly expressed in the posterior silk glands of silkworms on the third day of the fifth instar (Figure 6) (Lu F. et al., 2019). These studies suggested that this gene is likely to play an important role in the synthesis of silk. *KWMTBOMO02490* encodes an ATPase inhibitor-like protein, and has ATPase inhibitor activity (GO: 0042030). It is involved in the negative regulation of ATPase activity (GO: 0032780), and locates mainly in mitochondria (GO: 0005739), indicating functions closely related to energy-consuming activities within the cell. The process of fibroin protein synthesis in the posterior silk gland requires a large amount of energy, which comes from ATP hydrolysis. In summary, we speculate that ATPase inhibitor-like proteins affect the synthesis of silk fibroin in the posterior silk glands of silkworms by regulating the

activity of ATPase, thereby affecting silk yield. For the WCW-related genes, we found that the region of the *KWMTBOMO12678* gene also showed a high degree of linkage disequilibrium. This gene encodes OTU domain-containing protein 7B (OTUD7B), which is a deubiquitylase belonging to the A20 subgroup of ovarian tumor (OTU) protein superfamily (Tang et al., 2021). A previous study has shown that the OTUD7B deficiency results in B-cell hyper-responsiveness to antigens, lymphoid follicular hyperplasia in the intestinal mucosa, and elevated host-defence ability against an intestinal bacterial pathogen, *Citrobacter rodentium* (Hu et al., 2013).

5 Conclusion

In this study, GWAS was used to study the cocoon production traits of four silkworm varieties in Henan Province, and 40 significant SNPs with 28 related candidate genes were identified. These genes are closely related to traits such as CSW, CW, and full cocoon weight. The results suggest that these candidate genes may affect the production and health traits of silkworms, and in particular, *KWMTBOMO02490* gene is considered to be an excellent candidate genes affecting cocoon shell traits due to its potential role in the posterior silk glands. These findings not only enhance our understanding of the genetic basis of cocoon production traits, but also provide a new perspective on molecular marker-assisted selection of silkworm varieties.

Although the identified candidate genes may be related to cocoon production traits, their specific biological functions and mechanisms of action need to be further experimentally verified. Future research will focus on the functional validation and application of these candidate genes, with a view to breeding new silkworm varieties with higher cocoon yields.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because There is no ethical approval for invertebrates in this institution, and it is being improved.

Author contributions

XS: Data curation, Visualization, Writing – original draft. YL: Data curation, Funding acquisition, Resources, Writing – review and editing. YZ: Methodology, Software, Visualization, Writing – review and editing. WD: Data curation, Software, Visualization, Writing – review and editing. YH: Data curation, Methodology, Resources, Supervision, Writing – original draft, Writing – review and editing. SY: Data curation, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported

by the Autonomous innovation project of Henan Academy of Agricultural Sciences (2023ZC117, 2024ZC128).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17), i884–i890. doi:10.1093/bioinformatics/bty560
- Dong, S. S., He, W. M., Ji, J. J., Zhang, C., Guo, Y., and Yang, T. L. (2021). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings Bioinforma.* 22 (4), bbaa227. doi:10.1093/bib/bbaa227
- Freed, D., Aldana, R., Weber, J. A., and Edwards, J. S. (2017). The Sentieon Genomics Tools - a fast and accurate solution to variant calling from next-generation sequence data. *bioRxiv*, 115717. doi:10.1101/115717
- Putahashi, R., Sato, J., Meng, Y., Okamoto, S., Daimon, T., Yamamoto, K., et al. (2008). Yellow and ebony are the responsible genes for the larval color mutants of the silkworm *Bombyx mori*. *Genetics* 180 (4), 1995–2005. doi:10.1534/genetics.108.096388
- Hu, H., Brittain, G. C., Chang, J.-H., Puebla-Osorio, N., Jin, J., Zal, A., et al. (2013). OTUD7B controls non-canonical NF- κ B activation through deubiquitination of TRAF3. *Nature* 494 (7437), 371–374. doi:10.1038/nature11831
- Kawamoto, M., Jouraku, A., Toyoda, A., Yokoi, K., Minakuchi, Y., Katsuma, S., et al. (2019). High-quality genome assembly of the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* 107, 53–62. doi:10.1016/j.ibmb.2019.02.002
- Lu, F., Wei, Z., Luo, Y., Guo, H., Zhang, G., Xia, Q., et al. (2019b). SilkDB 3.0: visualizing and exploring multiple levels of data for silkworm. *Nucleic acids Res.* 48 (D1), D749–D755. doi:10.1093/nar/gkz919
- Lu, K., Pan, Y., Shen, J., Yang, L., Zhan, C., Liang, S., et al. (2024). SilkMeta: a comprehensive platform for sharing and exploiting pan-genomic and multi-omic silkworm data. *Nucleic acids Res.* 52 (D1), D1024–D1032. doi:10.1093/nar/gkad956
- Lu, Y., Luo, J., An, E., Lu, B., Wei, Y., Chen, X., et al. (2023). Deciphering the genetic basis of silkworm cocoon colors provides new insights into biological coloration and phenotypic Diversification. *Mol. Biol. Evol.* 40 (2), msad017. doi:10.1093/molbev/msad017
- Lu, Z., Meng, Z., Wen, M., Kang, X., Zhang, Y., Liu, Q., et al. (2019a). Overexpression of BmFoxO inhibited larval growth and promoted glucose synthesis and lipolysis in silkworm. *Mol. Genet. genomics MGG* 294 (6), 1375–1383. doi:10.1007/s00438-019-01550-2
- Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., et al. (2004). The genome sequence of silkworm, *Bombyx mori*. *DNA Res. Int. J. rapid Publ. Rep. genes genomes* 11 (1), 27–35. doi:10.1093/dnares/11.1.27
- Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinforma. Oxf. Engl.* 32 (2), 292–294. doi:10.1093/bioinformatics/btv566
- Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2 (12), e190. doi:10.1371/journal.pgen.0020190
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* 26 (6), 841–842. doi:10.1093/bioinformatics/btq033
- Sakudoh, T., Sezutsu, H., Nakashima, T., Kobayashi, I., Fujimoto, H., Uchino, K., et al. (2007). Carotenoid silk coloration is controlled by a carotenoid-binding protein, a product of the Yellow blood gene. *Proc. Natl. Acad. Sci. U. S. A.* 104 (21), 8941–8946. doi:10.1073/pnas.0702860104
- Shimomura, M., Minami, H., Suetsugu, Y., Ohyanagi, H., Satoh, C., Antonio, B., et al. (2009). KAIKObase: an integrated silkworm genome database and data mining tool. *BMC genomics* 10, 486. doi:10.1186/1471-2164-10-486
- Tang, J., Wu, Z., Tian, Z., Chen, W., and Wu, G. (2021). OTUD7B stabilizes estrogen receptor α and promotes breast cancer cell proliferation. *Cell Death and Dis.* 12 (6), 534. doi:10.1038/s41419-021-03785-7
- Tong, X., Han, M.-J., Lu, K., Tai, S., Liang, S., Liu, Y., et al. (2022). High-resolution silkworm pan-genome provides genetic insights into artificial selection and ecological adaptation. *Nat. Commun.* 13 (1), 5619. doi:10.1038/s41467-022-33366-x
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603
- Xia, Q., Guo, Y., Zhang, Z., Li, D., Xuan, Z., Li, Z., et al. (2009). Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Sci. (New York, NY)* 326 (5951), 433–436. doi:10.1126/science.1176620
- Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., et al. (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Sci. (New York, NY)* 306 (5703), 1937–1940. doi:10.1126/science.1102210
- Xiang, H., Liu, X., Li, M., Zhu, Y., Wang, L., Cui, Y., et al. (2018). The evolutionary road from wild moth to domestic silkworm. *Nat. Ecol. and Evol.* 2 (8), 1268–1279. doi:10.1038/s41559-018-0593-4
- Yang, J. G., Liu, T. H., Dong, X. L., Wu, Y. F., Zhang, Q., Zhou, L., et al. (2017). *In vivo* RNA interference of BmNHR96 enhances the resistance of transgenic silkworm to BmNPV. *Biochem. biophysical Res. Commun.* 493 (1), 332–339. doi:10.1016/j.bbrc.2017.09.022
- Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinforma. Oxf. Engl.* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44 (7), 821–824. doi:10.1038/ng.2310