

Article

Not peer-reviewed version

Mechanisms Behind the Formation of Ellipsoid Eggs in Silk-Worms (*Bombyx mori*): Insights from Transcriptomic Profiling

[Yanrong Chen](#)^{*}, Xinkai Wang, Manyun Xu, Xinyu Shen, [Pengchao Guo](#)^{*}, [Xiao-Hui Bai](#)

Posted Date: 2 September 2025

doi: 10.20944/preprints202509.0034.v1

Keywords: *Bombyx mori*; the ellipsoid egg; egg shape mutation; transcriptome sequencing



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Mechanisms Behind the Formation of Ellipsoid Eggs in Silk-Worms (*Bombyx mori*): Insights from Transcriptomic Profiling

Yanrong Chen ^{1,2,*}, Xinkai Wang ², Manyun Xu ¹, Xinyu Shen ¹, Pengchao Guo ^{3,4,*} and Xiao-Hui Bai ¹

¹ College of Life and Environment Science, Huangshan University, Huangshan Anhui 245041, China

² Jiangsu University of Science and Technology, Zhenjiang 212100, China

³ Integrative Science Center of Germplasm Creation in Western China (CHONGQING) Science City, Biological Science Research Center, Southwest University, Chongqing, 400716, China

⁴ State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, China

* Correspondence: 107025@hsu.edu.cn; guopc@swu.edu.cn

Abstract

Background/Objectives: The elongated egg is a morphological mutant of silkworm (*Bombyx mori*) eggs, yet the biochemical processes and molecular mechanisms underlying this trait remain unclear. **Methods:** In this study, we performed transcriptome sequencing on the ovaries of female pupae from the Nistari silkworm strain (comparing normal and elongated eggs) during the first three days post-pupation using high-throughput sequencing. **Results:** A total of 153.56 Gb of filtered data was obtained, identifying 23,366 genes and 35,798 mRNAs. Comparative analysis across three control groups revealed 374 differentially expressed genes (DEGs), with 131 upregulated and 243 downregulated genes. GO and KEGG enrichment analyses indicated that these DEGs were primarily associated with protein hydrolysis, DNA metabolic processes, and euchromatin/heterochromatin organization. Trend expression analysis revealed that transcriptional activity in elongated eggs was significantly higher than in normal eggs, particularly on day 3 of the pupal stage. **Conclusions:** WGCNA analysis classified gene expression patterns into twelve modules, with two modules showing specificity. Thirteen hub genes were identified, which are functionally linked to translation initiation, protein density regulation, post-translational modification, and protein turnover. These findings provide foundational insights into the molecular mechanisms driving the formation of elongated egg in silkworms.

Keywords: *Bombyx mori*; the ellipsoid egg; egg shape mutation; transcriptome sequencing

1. Introduction

The domestic silkworm (*Bombyx mori*) is an economically significant lepidopteran insect, playing a vital role in China's agricultural economy [1]. The hatching rate is a key indicator of silkworm egg quality, influenced by several factors such as egg quality, the season of egg production, nutritional content, and the duration of cold storage [2]. Among these, the shape index and weight of the silkworm eggs are particularly impactful on the hatching rate [2]. Therefore, studying the effect of the shape index on the hatching rate and understanding its molecular mechanisms are of great importance for improving the hatching rate, as well as for the selection and preservation of high-quality silkworm eggs.

The shape of the silkworm egg is determined by the eggshell, which is secreted by the follicular cells [3]. During ovarian development in the female moth, the follicular cells secrete and deposit a thick layer of yolk membrane, followed by the programmed secretion of chitin, which is sequentially deposited outside the oocyte [4]. This process leads to the coalescence and solidification of the chitin,

ultimately forming the eggshell [4]. The normal silkworm egg is typically short-oval in shape, slightly flattened, with a pointed anterior edge at the micropyle end [5]. However, there are various oval mutations in nature, one of which is the elongated egg [6]. The elongated egg is characterized by an extended oval shape, with a polygonal cell structure forming a reticulated pattern on the central surface. The long axis of the egg aligns with the egg's long axis [6,7].

The elongated egg trait exhibits mimetic matrilineal inheritance, with a one-generation delay in phenotypic dominant-recessive segregation compared to normal inheritance [6,7]. This trait is controlled by a recessive mutant gene, *elp*, located at locus 16.1 on linkage group 18 [8]. However, this gene has yet to be finely localized and cloned. Since embryos of the elongated eggs are capable of developing, hatching, and surviving normally, it is hypothesized that the *elp* gene likely elongates the shape of the follicular cells but does not affect their secretory function [8]. So far, the molecular mechanism underlying the formation of elongated eggs remains unclear.

In this study, we used normal silkworm eggs from the Nistari breed and its elongated egg mutant as research subjects. Ovarian tissues from the first three days after pupation were subjected to high-throughput transcriptome sequencing to analyze differences in transcript levels between normal and elongated eggs during days 1 to 3 of the pupal stage. By identifying differentially expressed genes during this period, we aimed to uncover the key genes associated with the formation of the elongated egg mutation. This research will provide insights into the molecular mechanisms underlying the development of elongated eggs.

2. Materials and Methods

2.1. Animals and Rearing Methods

The normal egg strain of the Nistari variety and the elongated egg strain of silkworms were selected as experimental materials. Both strains were provided by the Key Laboratory of Sericulture, Research Institute of the Chinese Academy of Agricultural Sciences, and Jiangsu University of Science and Technology. The eggs were incubated at 25°C until hatching. After hatching, the silkworms were reared on fresh mulberry leaves at 25°C and 70-85% humidity. At the pupal stage, ovarian tissues were dissected and collected for analysis. Three replicates were taken for each day across a three-day period, with transcriptome sequencing conducted on each of the collected samples. Simultaneously, normal ovarian tissues extracted from the pupal stage during the first three days were labeled as C1, C2, and C3, while elongated ovarian tissues from the same period were labeled as E1, E2, and E3, respectively.

2.2. Extraction of Total RNA and qRT-PCR Analysis

Ovaries were collected from female pupae on days 1 to 3 of pupation, with three biological replicates for each day. Total RNA was extracted using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The RNA quality ($OD_{260\text{ nm}}/OD_{280\text{ nm}}$) and concentration (ng/ μL) were assessed using a Thermo Scientific NanoDrop™ 1000 spectrophotometer (Thermo Scientific, Germany). Qualified samples were selected for the construction of sequencing libraries and subsequent Illumina transcriptome high-throughput sequencing.

To validate the reliability of the RNA-Seq data, 20 differentially expressed genes (DEGs) with distinct expression patterns were selected for qRT-PCR verification. Silkworm β -actin 3 and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were used as internal reference controls. Primers were designed using Oligo 7 software, and each sample included three biological replicates and three technical replicates. The relative expression levels of target genes were calculated using the $2^{-\Delta\Delta CT}$ method, based on the Cq values of the target and reference genes.

2.3. Transcriptome Sequencing and Assembly

Transcriptome sequencing of normal and elongated eggs from ovaries during the first three days of the pupal stage was conducted by Nanjing Jisi Huiyuan Biotechnology Co., Ltd. in Nanjing, China. Sequencing was performed with three biological replicates, and 18 libraries were constructed. After sequencing, the raw reads were subjected to quality control (QC) using the NGS QC Toolkit (v2.3.3) [9], and clean reads were obtained after filtering. The genome database of the University of Tokyo (<http://kaikocdna.dna.affrc.go.jp/>) was used as the reference genome sequence. Genome alignment and transcript reconstruction were performed using HISAT2 [10] and StringTie [11].

2.4. Differential Expression Gene Screening

Gene expression abundance in each sample was calculated using Cuffdiff [12] and StringTie [11], based on sequence similarity comparison. The expression levels of transcripts or genes were measured using the FPKM method. Differential expression analysis was performed using the Negative Binomial Distribution Test (NB) in the DESeq software package [13]. The NB test assessed the significance of differences in read counts, with the base mean value used to estimate gene expression. Differentially expressed genes were selected using the following criteria: $|\log_2(\text{Fold Change})| \geq 1$ and $\text{FDR} < 0.05$. To correct for multiple testing, the Benjamini-Hochberg method was applied to adjust the p-values, and FDR was used as the key index for identifying differentially expressed genes.

2.5. GO and KEGG Functional Enrichment

Gene Ontology (GO, <http://www.geneontology.org/>) is an international standard classification system for gene function, divided into three categories: molecular function, biological process, and cellular component. After screening for differentially expressed genes, GO enrichment analysis was conducted to investigate their distribution across these categories and to clarify the functional differences in the samples. The enrichment analysis is based on the hypergeometric distribution, which calculates the probability of observing a given number of differential genes within a specific GO term compared to the whole genome. A small p-value indicates significant enrichment of differential genes in that GO category.

KEGG pathway analysis was conducted on the differentially expressed genes using annotations from the Kyoto Encyclopedia of Genes and Genomes (KEGG, <https://www.genome.jp/kegg/>). The significance of gene enrichment in each pathway was assessed using the hypergeometric distribution test, where a small p-value indicates significant enrichment in that pathway. Both GO and KEGG enrichment analyses were performed and visualized using the OmicShare online platform [14] for data analysis.

2.6. Trend Analysis and Gene Co-Expression Network Analysis

Expression trend analysis of differentially expressed genes during the first three days of the pupal stage (days 1, 2, and 3) was performed using the STEM software [15]. The analysis used default parameters, with $\text{FDR} < 5\%$ and a fold change greater than 2 between any two time points. Gene co-expression network analysis was carried out using the WGCNA module in TBtools software [16], and the resulting gene regulatory networks were visualized with Cytoscape [17]. Module analysis was also conducted to explore the functional relationships between the genes.

3. Results

3.1. Quality Control and Sequence Splicing of Transcriptome Data

Using sequencing-while-synthesizing technology, each sample generated approximately 32-35 million raw reads, with a total read length of about 3 Gb, approximately seven times the size of the silkworm genome. Most of the bases achieved or exceeded the Q30 quality score. After removing low-quality reads (where more than 50% of the bases had a quality score of $Q \leq 5$), a total of 153.56

Gb of clean data was obtained. The base quality evaluation showed that 89.77% of the bases had a Q30 score, as detailed in Table 1.

Table 1. The statistics for the sequencing data estimates of all samples.

Sample	ReadSum	BaseSum	GC(%)	Q20(%)	CycleQ20(%)	Q30(%)
C1A	32744728	9823418400	46.62	96.86	100	91.81
C1B	28496796	8549038800	48.33	96.76	100	91.66
C1C	28193104	8457931200	48.40	97.03	100	92.24
C2A	26138394	7841518200	47.06	96.00	100	89.77
C2B	27039171	8111751300	50.69	96.86	100	91.90
C2C	25498980	7649694000	51.85	96.88	100	91.97
C3A	29240149	8772044700	48.42	96.89	100	91.90
C3B	29106631	8731989300	50.77	96.71	100	91.57
C3C	30092318	9027695400	48.04	97.03	100	92.20
E1A	27297421	8189226300	48.32	96.74	100	91.65
E1B	24407125	7322137500	48.03	96.91	100	91.99
E1C	31966938	9590081400	50.04	96.81	100	91.81
E2A	29731372	8919411600	48.78	96.93	100	92.00
E2B	32581082	9774324600	50.45	96.76	100	91.75
E2C	26406376	7921912800	48.00	96.94	100	92.04
E3A	26070073	7821021900	49.69	96.93	100	92.06
E3B	31237010	9371103000	49.37	97.02	100	92.26
E3C	24622649	7386794700	48.95	96.56	100	91.34

C1, C2, and C3: Samples from the 1st, 2nd, and 3rd days of the normal egg pupa stage. E1, E2, and E3: Samples from the 1st, 2nd, and 3rd days of the elongated egg pupa stage. A, B, C: Three replicate samples. ReadSum: The total number of paired-end reads in Clean Data. BaseSum: The total number of bases in Clean Data. GC content: The percentage of G and C bases in Clean Data relative to the total bases. Q30: The percentage of bases with Phred values greater than or equal to 30 in Clean Data relative to the total bases.

Reference genome alignment using HISAT2 revealed that about 81%-84% of the clean reads from each sample aligned with the silkworm genome. The coverage of all detected genes was approximately 80%, with 45%-60% of the reads mapping to known exonic regions, 11%-31% to intronic regions, and 7%-12% to intergenic regions. After merging our sequencing data with publicly available mRNA sequences, genome splicing and gene reconstruction were performed using StringTie. This analysis predicted a total of 23,366 genes and 35,798 mRNAs, significantly surpassing the 14,624 genes predicted by the Silkworm Genome Database (SilkDB) [18]. These results indicate that the sequencing data provides a more comprehensive representation of gene expression in silkworm tissues.

3.2. qPCR Assay and DEGs Screening

The expression levels of 20 genes were assessed using qRT-PCR according to the manufacturer's instructions. A strong correlation ($r = 0.939$, Figure 1) was observed between the RNA-Seq data and the qRT-PCR results. The expression profiles obtained from qRT-PCR closely matched those derived from RNA-Seq, validating the reliability of the RNA-Seq data in this study.

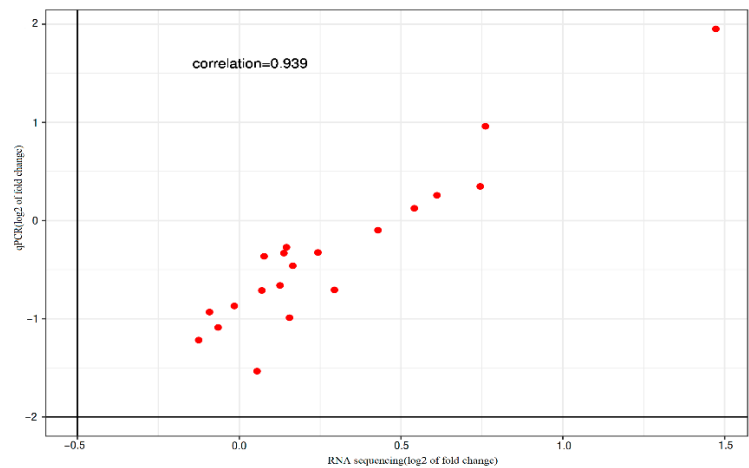


Figure 1. The Pearson correlation analysis of qRT-PCR results. RNA sequencing data for the 20 selected DEGs among the samples. Each point represents a value of fold change of expression level in C1-C3 groups comparing with that in E1-E3 groups. Fold-change values were log10 transformed.

The elongated eggs of the silkworm are a characteristic of the eggshell trait, secreted by the egg follicle cells. The three days before pupation are crucial for follicle development [19]. To investigate the gene expression patterns associated with this trait, a comparative analysis of normal and elongated egg samples was performed. The data were normalized, and differentially expressed genes (DEGs) were identified using $FDR \leq 0.05$ and \log_2 fold change > 1 as screening criteria. A total of 374 DEGs were detected between the two groups during the first three days of the pupal stage, including 131 up-regulated and 243 down-regulated genes. Specifically, the C1 vs E1 group showed 60 up-regulated and 35 down-regulated genes. The C2 vs E2 group showed 38 up-regulated and 22 down-regulated genes, while the C3 vs E3 group had 49 up-regulated and 192 down-regulated genes. The specific pathways associated with these genes are presented in Table 2. Differential expression was further analyzed using Venn and MA plots (Figures 2 and 3), which revealed changes in gene expression patterns between the two groups during the early pupal stage.

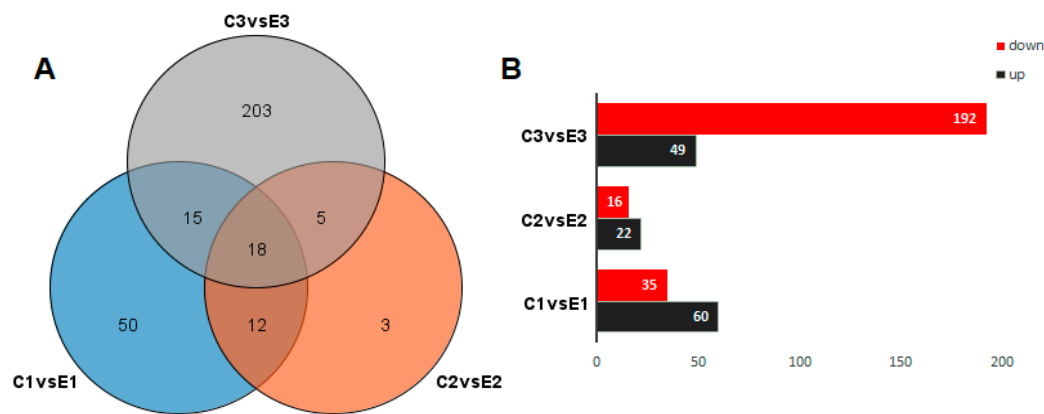


Figure 2. Venn diagrams of DEGs (A) and number of DEGs (B).

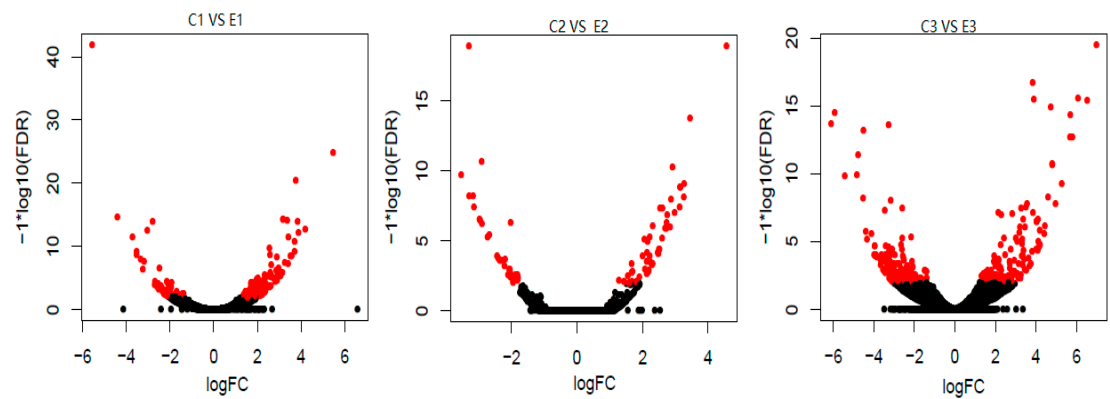


Figure 3. The Volcano plot of DEGs. The graph shows a for C1 vs E1 group, b for C2 vs E2 group, and c for C3 vs E3 group, with log10 (FDR) on the vertical axis and logFC on the horizontal axis, and black dots indicate genes with no significant differences, and red dots indicate genes with significant differences.

Table 2. The results of KEGG enrichment analysis for the C1 vs E1, C2 vs E2, and C3 vs E3 groups.

Group	Pathway	KEGG ID	Enrichment_Factor	Qvalue
C1 vs E1	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	ko00563	0.04	0.26096
	Cysteine and methionine metabolism	ko00270	0.08	0.28168
	Ribosome biogenesis in eukaryotes	ko03008	0.15	0.32021
	Spliceosome	ko03040	0.24	1
	Endocytosis	ko04144	0.24	1
	Protein processing in endoplasmic reticulum	ko04141	0.29	1
	Metabolic pathways	ko01100	0.78	1
	MAPK signaling pathway	Ko04013	1.90	1
	Hippo signaling pathway	Ko04391	2.60	1
C2 vs E2	Endocytosis	ko04144	0.04	0.04036
C3 vs E3	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	ko00563	0.04	0.26096
	MAPK signaling pathway	Ko04013	2.19	1
	Hippo signaling pathway	Ko04391	0.33	1
	Cysteine and methionine metabolism	ko00270	0.08	0.28168
	Ribosome biogenesis in eukaryotes	ko03008	0.15	0.32021
	Spliceosome	ko03040	0.24	1
	Endocytosis	ko04144	0.24	1
	Protein processing in endoplasmic reticulum	ko04141	0.29	1
	Metabolic pathways	ko01100	0.78	1

KEGG: Kyoto Encyclopedia of Genes and Genomes.

3.3. GO Enrichment Analysis of DEGs

In the C1 vs E1 group, DEGs were annotated using the whole transcriptome background and GO/KO annotations. GO enrichment analysis revealed that 138 DEGs were associated with biological processes, 52 with molecular functions, and 36 with cellular components. Among the biological process-related genes, most were involved in multicellular biological processes, protein hydrolysis, DNA metabolism, and GPI anchor biosynthesis. Cellular component-related genes were mainly involved in signal transduction, heterochromatin, euchromatin, integral components of the

B

membrane, and multilineage chromosome bands. For molecular functions, the majority of genes were related to RNA-directed DNA polymerase activity, peptidyl chain endonuclease activity, DNA binding, and hydrolase activity (Figure 4).

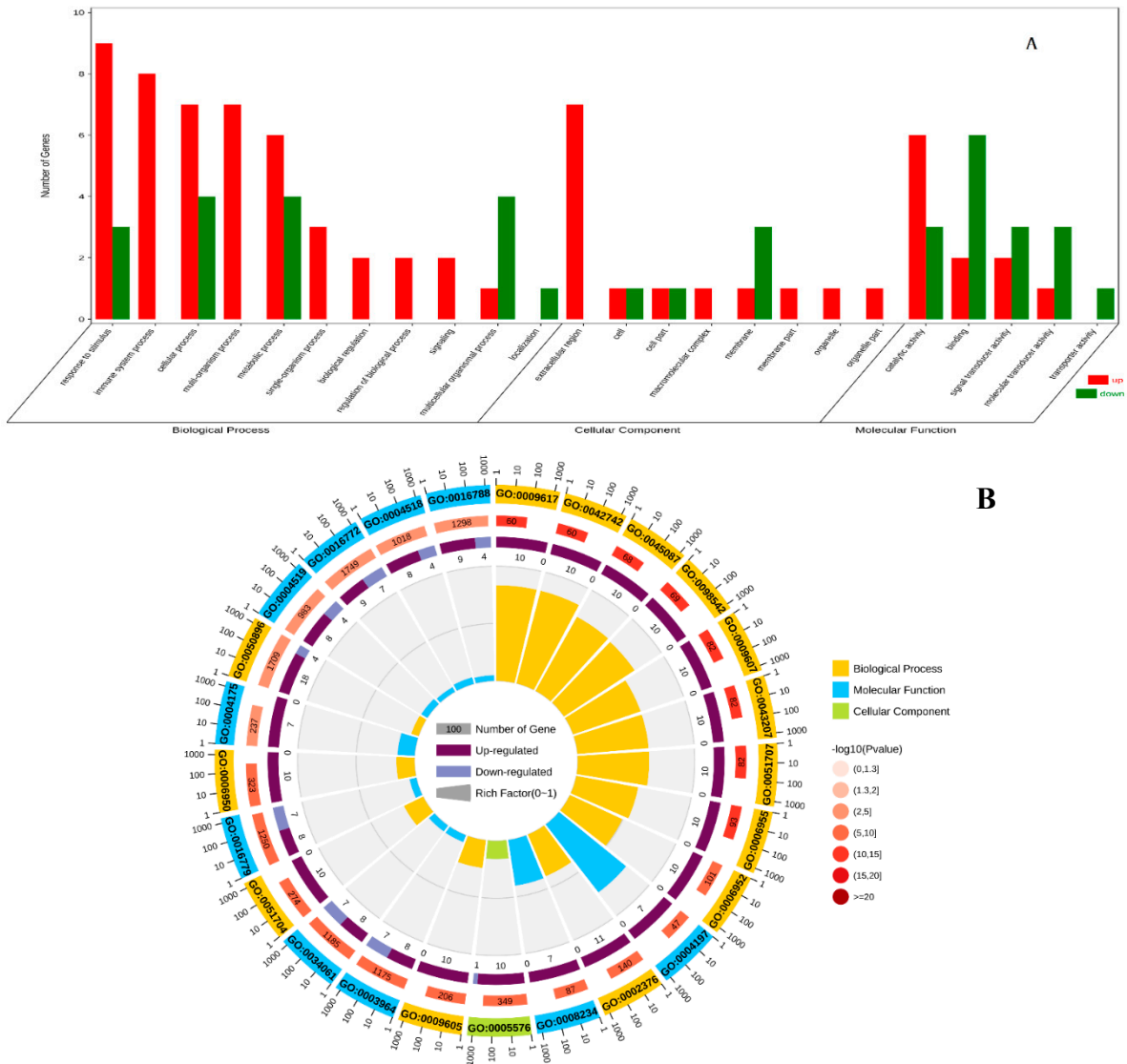


Figure 4. Gene ontology (GO) classification of DEGs in C1 vs E1 group. (A) GO classification analysis of DEGs. (B) GO enrichment analysis of DEGs. The abscissa of figure is the GO classification, the ordinate is the percentage of the number of genes on the left and the number of genes on the right.

In the C2 vs E2 group, DEGs were annotated similarly using the whole transcriptome background and GO/KO annotations. GO enrichment analysis showed that 63 DEGs were assigned to biological processes, 18 to cellular components, and 45 to molecular functions. Among the biological process-related genes, most were involved in RNA-directed DNA polymerase activity, somatic muscle development, protein hydrolysis, and DNA integration. The cellular component-related genes were primarily involved in signal transduction, membrane integration, cytoplasm, euchromatin, and protein complex. Genes related to molecular functions were mainly associated with RNA-directed DNA polymerase activity, zinc-finger structure, DNA binding, and neuropeptide Y receptor activity (Figure 5).

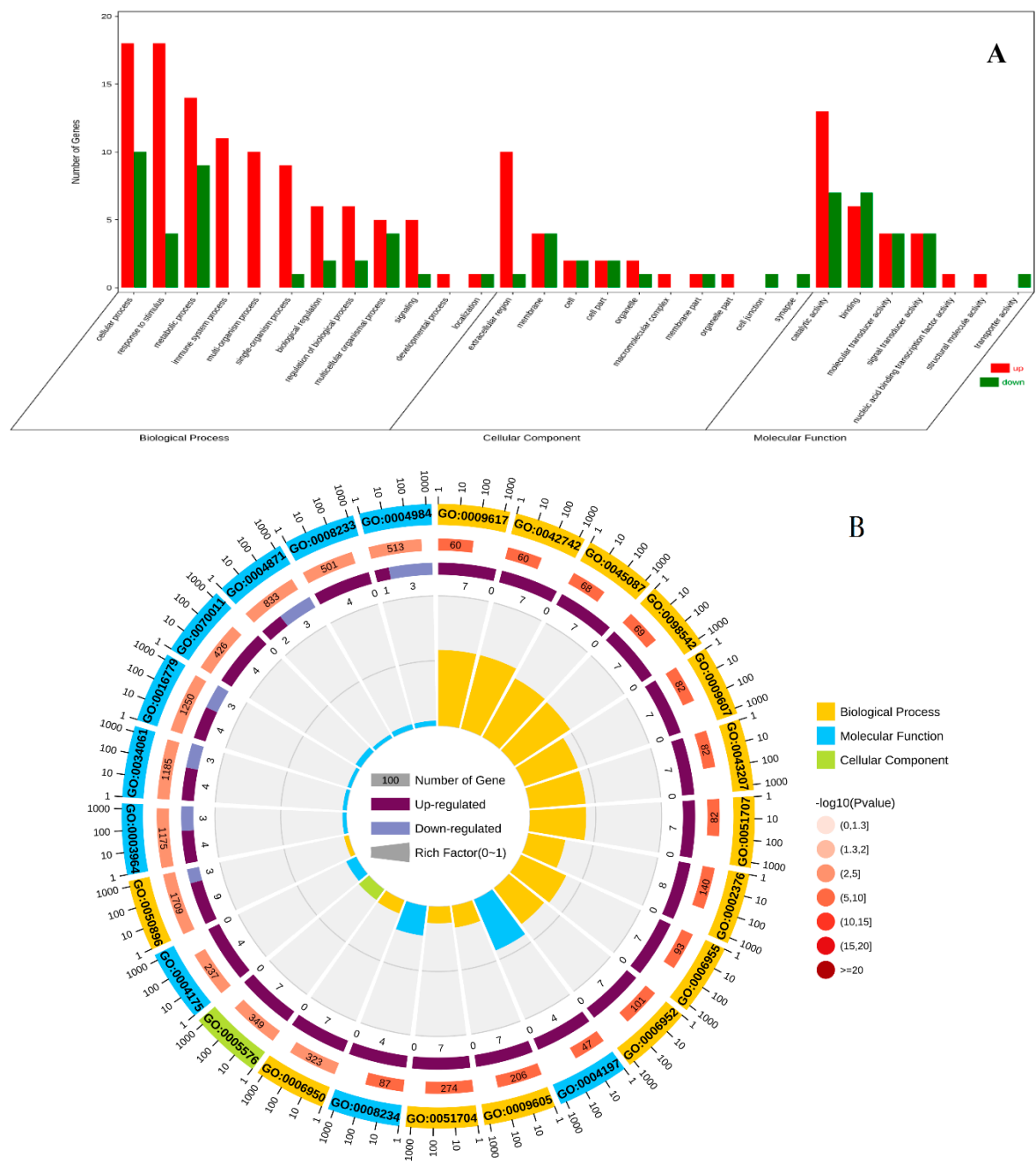


Figure 5. Gene ontology (GO) classification of DEGs in C2 vs E2 group. (A) GO classification analysis of DEGs. (B) GO enrichment analysis of DEGs. The abscissa of figure is the GO classification, the ordinate is the percentage of the number of genes on the left and the number of genes on the right.

For the C3 vs E3 group, DEGs were annotated in the same way, and GO enrichment analysis revealed that 311 DEGs were linked to biological processes, 143 to cellular components, and 143 to molecular functions. Biological process-related genes were predominantly involved in translation initiation, RNA phosphodiester bond hydrolysis, somatic muscle development, and nucleus development. Cellular component-related genes mainly contributed to signal transduction, heterochromatin, negative regulation of smooth signaling pathways, dorsal closure, and ion transport across membranes. Molecular function-related genes were primarily linked to monooxygenase activity, deconjugating enzyme activity, cysteine-type peptidyl endonuclease activity, and calcium-dependent phospholipid binding (Figure 6).

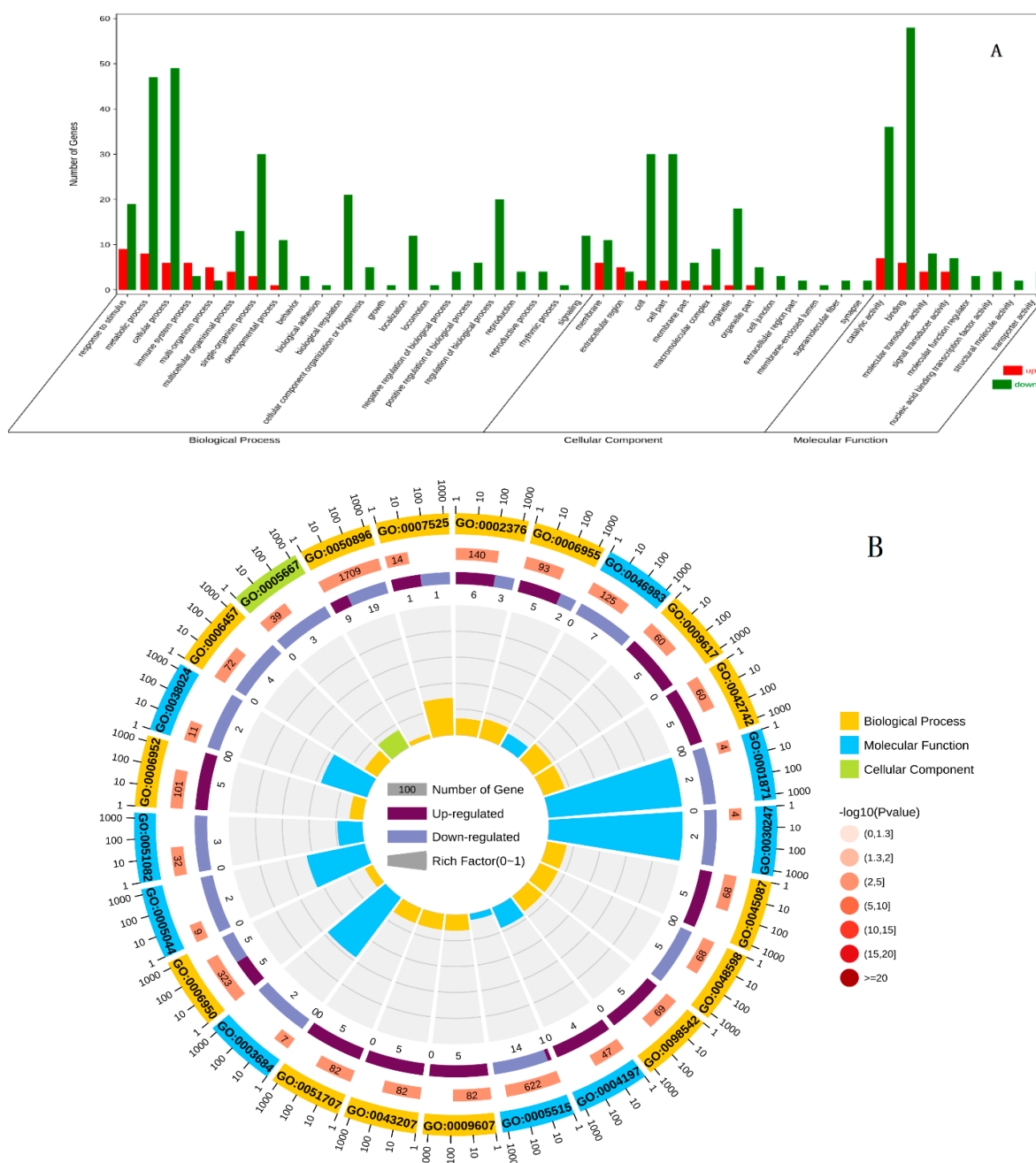


Figure 6. Gene ontology (GO) classification of DEGs in C2 vs E2 group. (A) GO classification analysis of DEGs. (B) GO enrichment analysis of DEGs.

3.4. KEGG Pathway Enrichment Analysis of DEGs

KEGG enrichment analysis identified a total of 123 DEGs annotated to 53 KEGG pathways across all three sample pairs. After filtering for pathways with a p-value ≤ 0.05 , several pathways were found to be significantly enriched in DEGs. In the C1 vs E1 group, the pathways mismatch repair, fanconi anemia pathway, metabolic pathway, and GPI-anchor biosynthesis were significantly enriched. The C2 vs E2 group enriched only one KEGG pathway, namely, endocytosis. In the C3 vs E3 group, seven pathways were significantly enriched, including those related to cysteine and methionine metabolism, protein processing in the endoplasmic reticulum, spliceosome, endocytosis, ribosome biosynthesis, and GPI-anchor biosynthesis (Figure 7). These enriched pathways provide valuable insights into the specific processes, functions, and pathways associated with stunting in the silkworm *B. mori*.

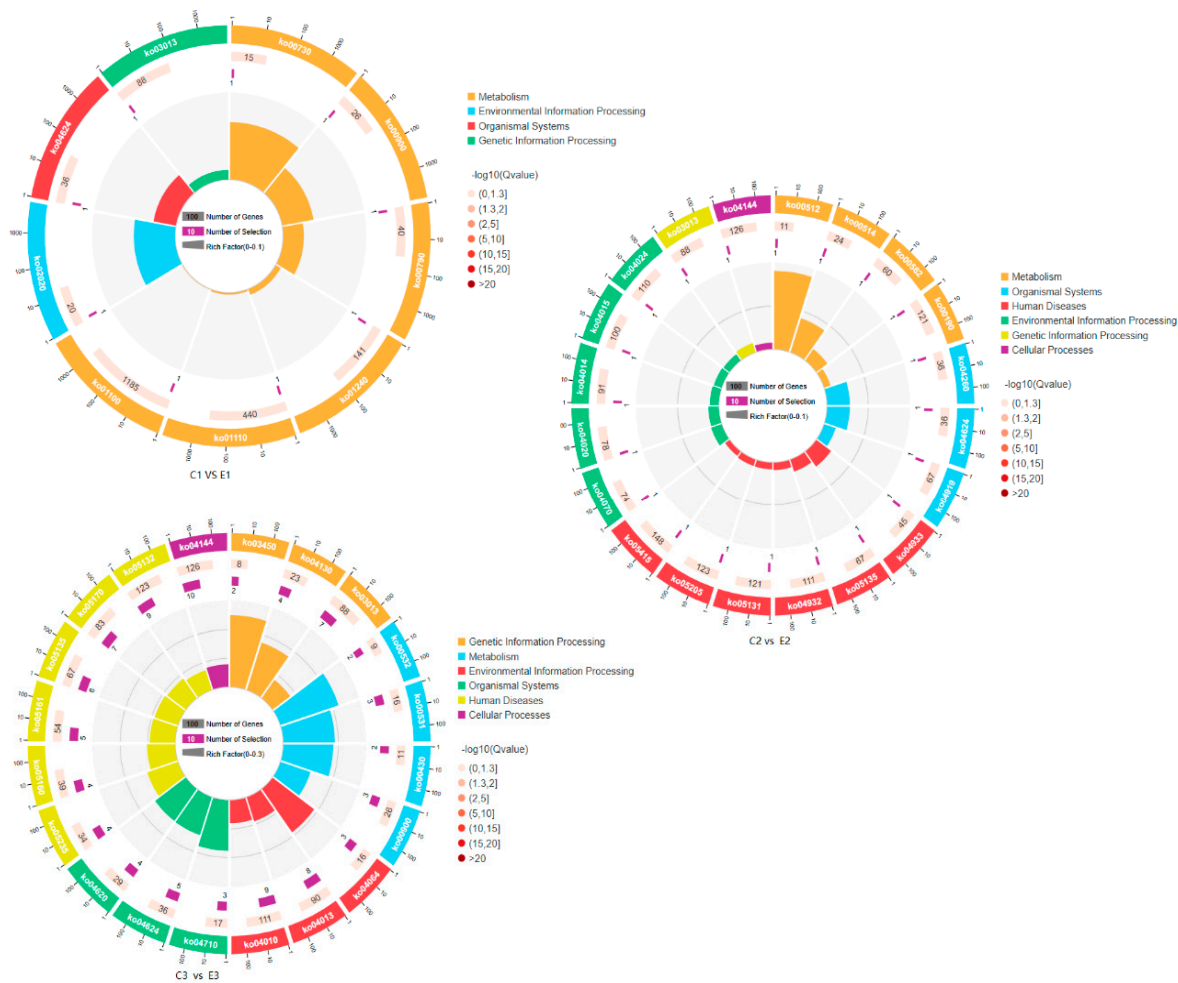
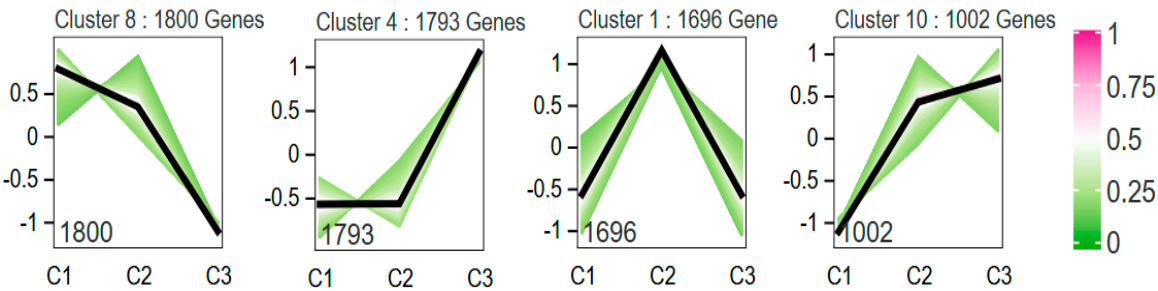


Figure 7. Kyoto Encyclopedia of Genes and Genomes (KEGG) categories for DEGs.

3.5. Trend Analysis

To investigate the expression patterns of differentially expressed genes (DEGs) in the normal egg group (C1, C2, C3) and the elongated egg group (E1, E2, E3), trend analysis was performed using the Short Time-series Expression Miner (STEM) software [20]. The 13,678 DEGs in the normal egg group were clustered into 10 expression patterns, with clusters 8, 4, 1, and 10 showing significantly enriched trends ($p < 0.05$). Specifically, cluster 8 contained 1,800 genes, cluster 4 contained 1,793 genes, cluster 1 contained 1,696 genes, and cluster 10 contained 1,002 genes (Figure 8). In the elongated egg group, 14,005 DEGs were clustered into 10 expression patterns. Profiles 4, 1, 10, and 5 exhibited significant enrichment trends ($p < 0.05$), with profile 4 containing 1,971 genes, profile 1 containing 1,838 genes, profile 10 containing 1,646 genes, and profile 5 containing 1,333 genes (Figure 8).



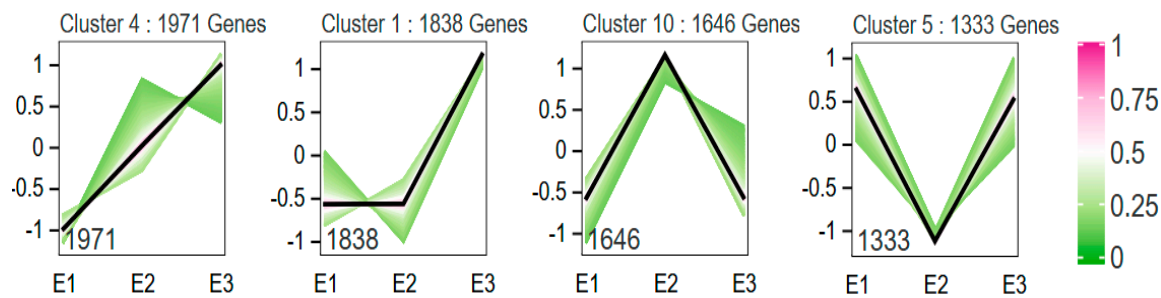


Figure 8. DEGs expression profiles of the wild type eggs and the elongate eggs. C1, C2, and C3 denote days 1, 2, and 3 of the normal egg pupal stage, respectively; E1, E2, and E3 denote days 1, 2, and 3 of the pupal stage of the elongate egg, respectively.

For each expression pattern, GO functional enrichment analysis was performed. In the normal egg group, genes in profiles 4 and 10 contained the majority of GO function entries, while genes in profile 1 were mainly enriched in cellular components and molecular functions. In the elongated egg group, profiles 4 and 1 contained the most GO function entries, while genes in profile 5 were enriched in cellular components and molecular functions. Additionally, genes in profile 8 were concentrated in the biological process and molecular function categories. In terms of expression trends, the first three days after pupation showed higher transcriptional activity in the elongated eggs compared to the normal eggs, particularly on the third day of pupation.

3.6. Weighted Gene Co-Expression Network Analysis (WGCNA)

To further investigate the relationship between gene expression in the ovaries during the first three days of the pupal stage and the formation of elongated eggs, we employed Weighted Gene Co-expression Network Analysis (WGCNA) [21] to examine the expression patterns of 23,366 genes filtered from 18 samples. This analysis identified 12 distinct gene co-expression modules based on their expression patterns (Figure 9A). Among them, two modules—yellow and pink—exhibited specific gene expression characteristics. The yellow module showed a negative correlation with the number of days until pupation (correlation coefficient = -0.89, $p < 0.01$), while the pink module showed a strong positive correlation (correlation coefficient = 0.59, $p < 0.01$). Additionally, the correlations between Module Membership (MM) and Gene Significance (GS) for these two modules were 0.6 and 0.91, respectively, indicating robust associations (Figure 9B and 9C).

To better understand the gene co-expression relationships, we used Cytoscape software [17] to visualize the co-expression network of differentially expressed genes with a weight greater than 0.2 in the yellow and pink modules. In the yellow module, four hub genes were identified: BGIBMGA005781, BGIBMGA002542, BGIBMGA000462, and BGIBMGA013932 (Figure 10). In contrast, the pink module contained nine core genes: BGIBMGA000562, BGIBMGA004157, BGIBMGA002058, BGIBMGA011121, BGIBMGA011120, BGIBMGA012351, BGIBMGA002576, BGIBMGA001201, and BGIBMGA006755 (Figure 11).

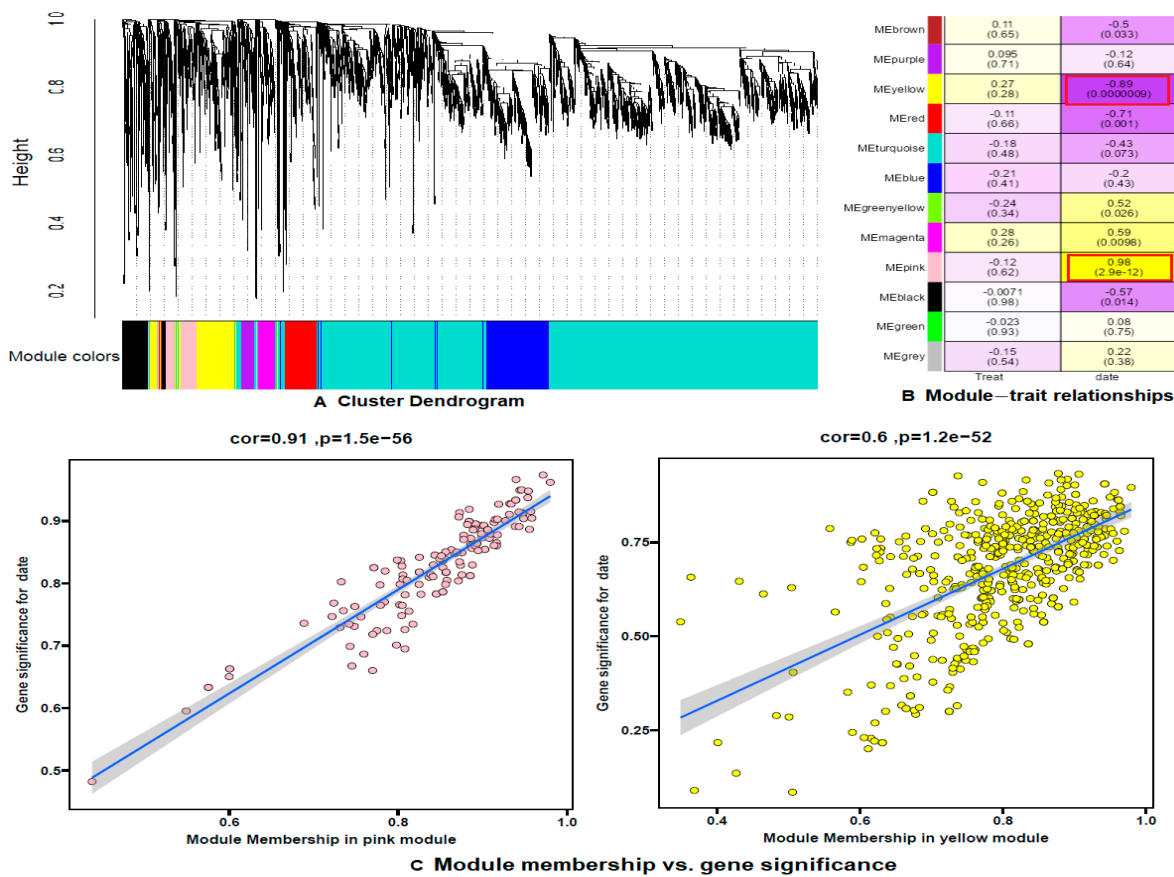


Figure 9. Identification of co-expression gene modules and the association with traits. A Clustering dendrogram of DEGs; B The heatmap of correlation between ME and traits of egg shape, the day of the pupation. C Scatter plot of ME in the pink and yellow modules.

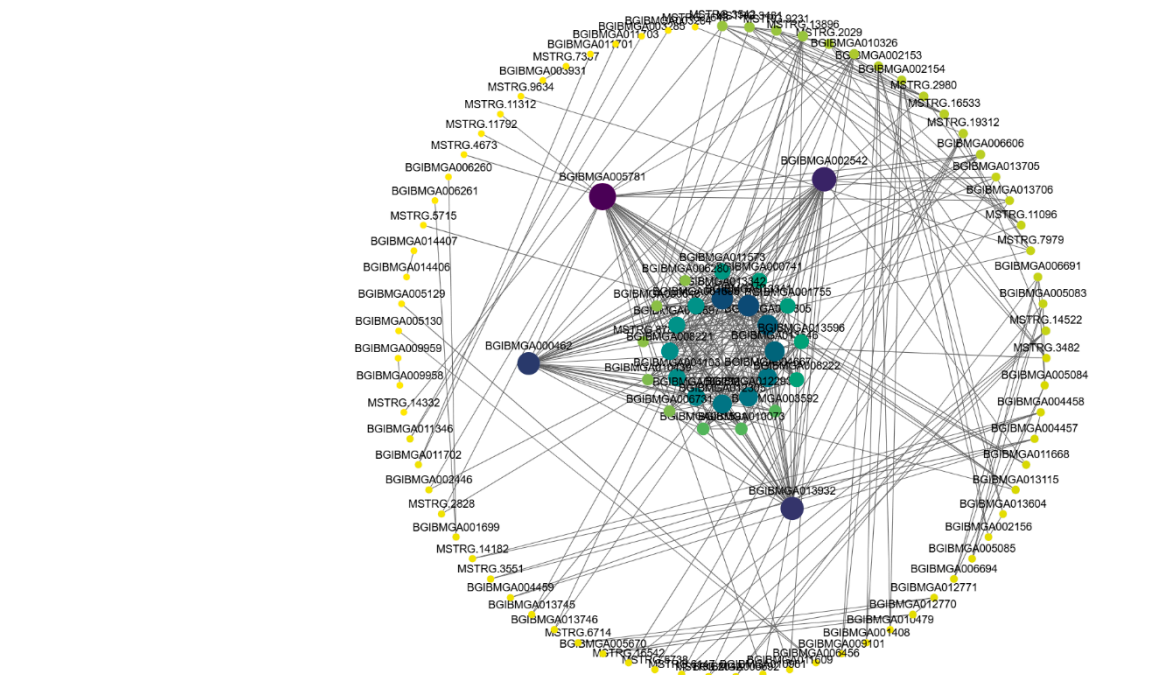


Figure 10. The visualization of the pink module. The hub genes in the module were bold with red and grey.

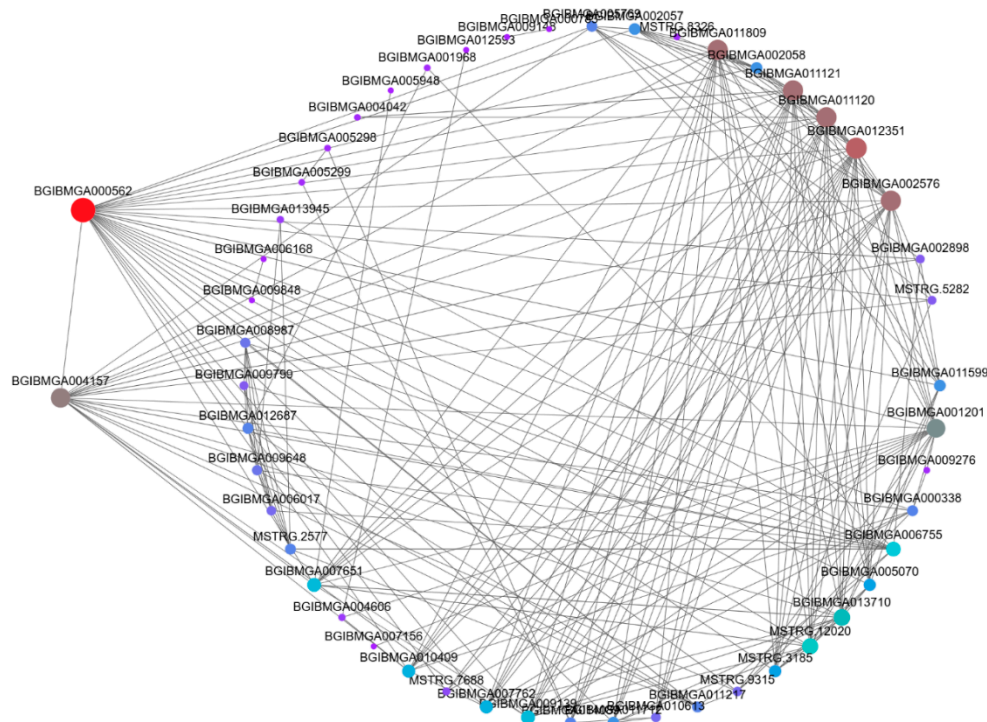


Figure 11. The visualization of the yellow module. The hub genes in the module were bold with lividity.

The core gene BGIBMGA005781 encodes a heat shock protein-like protein, which may play an important role in the formation of lipid bodies in silkworms [22]. BGIBMGA002542 encodes α -tubulin, a key component of the cytoskeleton [23]. BGIBMGA000462 encodes 27-kDa hemolymph proteins with an unknown function DUF1397 [24,25]. However, the specific function of this protein has not yet been reported. BGIBMGA013932 encodes a Serpin7 protein, which is involved in inhibiting the degradation of silk proteins and the generation of oxidative radicals, thereby maintaining homeostasis during silk formation [26]. The protein encoded by BGIBMGA000562 is associated with the sex-linked chocolate recessive color mutation in silkworm larvae [27], but its specific function remains unknown. Chen et al. reported that the expression level of the BGIBMGA004157 gene is more significant in oocytes during the steroid biosynthesis process, suggesting its involvement in the initiation of diapause [28]. BGIBMGA002058 is a gene linked to the Z sex chromosome [29]. BGIBMGA011120 encodes a protein involved in plasmalogen biosynthesis [28], while BGIBMGA001201 encodes muscular protein 20, which contains a putative actin-binding surface [30]. The functions of the remaining genes, BGIBMGA011121, BGIBMGA012351, BGIBMGA002576, and BGIBMGA006755, have not been reported.

4. Discussion

The hatching rate of silkworm eggs is a crucial indicator for evaluating the quality of silkworm breeding stock, directly affecting the yield, quality of cocoons, and the economic benefits for farmers [31]. Several factors influence the hatching rate, including genetic, physiological, pathological, and environmental factors [32]. Among these, the quality of the silkworm eggs, such as egg shape and egg weight, significantly impacts the hatching rate [33]. An excessively large egg shape index or a relatively long egg shape can hinder fertilization, leading to a decrease in the fertilization rate. Conversely, an excessively small egg shape index may cause difficulties in embryo development, leading to mortality in the later stages, thus reducing the hatching rate [33]. Therefore, investigating the molecular mechanisms underlying the formation of elongated silkworm eggs can provide valuable insights into the genetic and molecular basis of egg shape regulation, offering a theoretical foundation for optimizing silkworm breeding and enhancing the hatching rate.

In previous studies, we identified a mutant silkworm from the Nistari variety, which exhibited oval-shaped eggs that elongated into a more cylindrical form [33]. The ratio of the long axis to the short axis of these mutant eggs was 1.7, significantly higher than the 1.26 ratio of normal eggs, and the hatching rate was 59.49%, much lower than the 86.43% observed in normal eggs [34]. However, the molecular mechanism underlying the formation of elongated eggs remains unclear. In this study, we used normal Nistari eggs as the control group and performed transcriptome sequencing on the ovaries of both elongated and normal eggs from the first to the third day of the pupal stage. We then identified and analyzed differentially expressed genes (DEGs) between the two groups through GO and KEGG enrichment analysis. The results indicated that the DEGs were primarily involved in DNA replication, protein synthesis and degradation, mRNA splicing, snRNA regulatory complex activity, and the composition of the cytoskeleton and chromosomes. From a cellular perspective, the formation of elongated eggs involves the stretching of normal follicles into an elongated shape, a process controlled by the *elp* gene [8]. This process likely requires cytoskeletal rearrangement and energy conversion, which may, in turn, influence the transcriptional regulation of related genes.

Through WGCNA analysis, two specific co-expression modules were identified, revealing 13 hub genes that play roles in processes such as steroid biosynthesis, phosphatidylcholine synthesis, and cytoskeleton composition. Notably, the core hub gene BGIBMGA000562 encodes a protein related to the sex-linked chocolate recessive spot mutation in silkworm larvae [27], although its specific function remains to be further investigated. The core gene BGIBMGA002058 is associated with the Z chromosome [29], which is of particular interest, as Chen et al. reported that the target gene responsible for the formation of giant egg mutants in the Yun7 silkworm variety is also located on the Z chromosome [35]. This suggests that BGIBMGA002058 may play a role in the formation of elongated egg mutants in the Nistari silkworm variety, a mechanism that differs significantly from the *elp* and *sp* genes in terms of chromosomal location [36]. Additionally, the functions of the remaining four core genes (BGIBMGA011121, BGIBMGA012351, BGIBMGA002576, and BGIBMGA006755) have yet to be reported. Further research into the functions of these genes may help uncover the molecular mechanisms behind the formation of elongated eggs.

Author Contributions: Y.C. and P. G. designed the experiments; Y.C., X.W., M.X., X.S., P.G., and X.H.B. performed the experiments; Y.C., X.W., P.G., and X.H.B. analyzed data; Y.C., P.G., and X.H.B. wrote the manuscript.

Funding: This work was supported by the Talent Launching Fund of Huangshan University(2019xkjq010), the open project of the State Key Laboratory of Silkworm Genome Biology(SKLSGB-ORP202106), the Master's degree program of Huangshan University (hsxyssd007, 2022xzx005), the College Student's Innovation and Entrepreneurship Training Program (202310375022, 202410375035, 202410375042, S202410375082), and the First-class Discipline at Huangshan University (ylxk202101).

Data Availability Statement: Data are available from the corresponding author upon a reasonable request.

Conflicts of Interest: The authors declare that they have no competing financial interests.

Abbreviations

The following abbreviations are used in this manuscript:

DEGs	Differentially expressed genes
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
QC	Quality control
NB	Negative binomial distribution test
GO	Gene Ontology
KEGG	the Kyoto Encyclopedia of Genes and Genomes
WGCNA	Weighted gene co-expression network analysis
SilkDB	the Silkworm Genome Database
STEM	the Short Time-series Expression Miner

MMModule Membership

GSGene significance

References

1. 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. Mol. Sci.* **2018**, *19*, 3718.

2. Sun, J.; Zheng, X.; Ouyang, G.; Qian, H.; Chen, A. Ebony plays an important role in egg hatching and 30k protein expression of silkworm (*Bombyx mori*). *Arch. Insect. Biochem. Physiol.* **2023**, *113*, e22014.

3. Kawasaki, H.; Sato, H.; Suzuki, M. Structural Proteins in the silkworm egg-shells. *Insect Biochem.* **1971**, *1*(2), 130-148.

4. Telfer, W. H. Egg formation in lepidoptera. *J Insect Sci.* **2009**, *9*, 1-21.

5. Nageswara Rao, S.; Muthulakshmi, M.; Kanginakudru, S.; Nagaraju, J. Phylogenetic relationships of three new microsporidian isolates from the silkworm, *Bombyx mori*. *J. Invertebr. Pathol.* **2004**, *86*(3), 87–95.

6. Lu, C.; Xiang, Z. Genetic studies of ellipsoid egg 2 in silkworm, *Bombyx mori*. *Acta Sericol. Sinica* **1991**, *17*, 137-140.

7. Kyeth, W. S. Studies on the inheritance of cocoon color and egg shape in *Bombyx mori*. The doctoral thesis. Fukuoka: Kyushu University, **1943**: 182.

8. Liu, X. F.; Ma, X.; Hou, C. X.; Li, B.; Li, M. W. Molecular mapping of test mapping strain for 18th linkage group recessive genes elp, ch-2 and mln in silkworm (*Bombyx mori*). *Hereditas (Beijing)* **2013**, *35*(3), 373–378.

9. Patel, R. K.; Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* **2012**, *7*(2), e30619.

10. Pertea, M.; Kim, D.; Pertea, G. M.; Leek, J. T.; Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **2016**, *11*(9), 1650–1667.

11. Pertea, M.; Pertea, G. M.; Antonescu, C. M.; Chang, T. C.; Mendell, J. T.; Salzberg, S. L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*(3), 290–295.

12. Trapnell, C.; Hendrickson, D. G.; Sauvageau, M.; Goff, L.; Rinn, J. L.; Pachter, L.; Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **2013**, *31*(1), 46–53.

13. Love, M. I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550.

14. Mu, H.; Chen, J.; Huang, W.; Huang, G.; Deng, M.; Hong, S.; Ai, P.; Gao, C.; Zhou, H. OmicShare tools: A zero-code interactive online platform for biological data analysis and visualization. *iMeta* **2024**, *3*, e228.

15. Ernst, J.; Bar-Joseph, Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinforma.* **2006**, *7*, 191.

16. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202.

17. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504.

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

19. Chen, Z. W. Identification, function and expression regulation of chorion genes in silkworm, *Bombyx mori*. The doctoral thesis. Chongqing: Southwest University, **2017**: 18-20.

20. Ernst, J.; Bar-Joseph, Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **2006**, *7*, 191.

21. Maertens, A.; Tran, V.; Kleensang, A.; Hartung, T. Weighted gene correlation network analysis (WGCNA) reveals novel transcription factors associated with bisphenol a dose-response. *Front. Genet.* **2018**, *9*, 508.

22. Ogata, N.; Yokoyama, T.; Iwabuchi, K. Transcriptome responses of insect fat body cells to tissue culture environment. *PLoS One* **2012**, *7*(4), e34940.

23. Wang, S. H.; You, Z. Y.; Ye, L. P.; Che, J.; Qian, Q.; Nanjo, Y.; Komatsu, S.; Zhong, B. X. Quantitative proteomic and transcriptomic analyses of molecular mechanisms associated with low silk production in silkworm *Bombyx mori*. *J. Proteome Res.* **2014**, 13(2), 735–751.
24. Hou, Y.; Zhang, Y.; Gong, J.; Tian, S.; Li, J.; Dong, Z.; Guo, C.; Peng, L.; Zhao, P.; Xia, Q. Comparative proteomics analysis of silkworm hemolymph during the stages of metamorphosis via liquid chromatography and mass spectrometry. *Proteomics* **2016**, 16(9), 1421–1431.
25. Kajiwar, H.; Imamaki, A.; Nakamura, M.; Mita, K.; Xia, Q.; Ishizaka, M. Proteome analysis of silkworm 1. Fat body. *J. Electrophoresis* **2009**, 53(2), 19–26.
26. Yi, Q.; Zhao, P.; Wang, X.; Zou, Y.; Zhong, X.; Wang, C.; Xiang, Z.; Xia, Q. Y. Shotgun proteomic analysis of the *Bombyx mori* anterior silk gland: An insight into the biosynthetic fiber spinning process. *Proteomics* **2013**, 13(17), 2657–2663.
27. Liu, C.; Yamamoto, K.; Cheng, T. C.; Kadono-Okuda, K.; Narukawa, J.; Liu, S. P.; Han, Y.; Futahashi, R.; Kidokoro, K.; Noda, H.; Kobayashi, I.; Tamura, T.; Ohnuma, A.; Banno, Y.; Dai, F. Y.; Xiang, Z. H.; Goldsmith, M. R.; Mita, K.; Xia, Q. Y. Repression of tyrosine hydroxylase is responsible for the sex-linked chocolate mutation of the silkworm, *Bombyx mori*. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107(29), 12980–12985.
28. Chen, Y. R.; Jiang, T.; Zhu, J.; Xie, Y. C.; Tan, Z. C.; Chen, Y. H.; Tang, S. M.; Hao, B. F.; Wang, S. P.; Huang, J. S.; Shen, X. J. Transcriptome sequencing reveals potential mechanisms of diapause preparation in bivoltine silkworm *Bombyx mori* (Lepidoptera: Bombycidae). *Comp. Biochem. Physiol. Part D, Genomics Proteomics* **2017**, 24, 68–78.
29. Kawamoto, M.; Koga, H.; Kiuchi, T.; Shoji, K.; Sugano, S.; Shimada, T.; Suzuki, Y.; Katsuma, S. Sexually biased transcripts at early embryonic stages of the silkworm depend on the sex chromosome constitution. *Gene* **2015**, 560(1), 50–56.
30. Zhong, X. W.; Zhao, P.; Zou, Y.; Nie, H. Y.; Yi, Q. Y.; Xia, Q. Y.; Xiang, Z. H. Proteomic analysis of the immune response of the silkworm infected by *Escherichia coli* and *Bacillus bombysepticus*. *Insect Sci.* **2012**, 19(5), 559–569.
31. Rahmathulla, V. K. Management of climatic factors for successful Silkworm (*Bombyx mori* L.) crop and higher silk production: A Review. *Psyche: A J. Entomol.* **2012**, 2012, 121234.
32. Gao, H. Q.; Tang, Y.; Gu, H. Y.; Yang, B. Influence factors on hatching rate of silkworm eggs and its solutions. *Jiangsu Sericulture* **2019**, 41(Z1), 15–17.
33. Wang, X.Z.; Xu, J.; Wang, S.S.; Pu, Y.X.; Shen, X.J.; Tang, S.M. Effects of egg shape index of silkworm (*Bombyx mori*) on hatching rate and related gene expression analysis. *J. Southern Agriculture* **2020**, 51(12), 3109–3115.
34. Wang, X. The genetic analysis and the transcriptome analysis of the formation mechanism in the new mutant ellipsoid egg of *Bombyx mori*. The master's thesis. Zhenjiang: Jiangsu University of Science and Technology, **2020**: 25.
35. Chen, A.; Liao, P.; Li, Q.; Zhao, Q.; Gao, M.; Wang, P.; Liu, Z.; Meng, G.; Dong, Z.; Liu, M. phytanoyl-CoA dioxygenase domain-containing protein 1 plays an important role in egg shell formation of silkworm (*Bombyx mori*). *PLoS One* **2021**, 16(12), e0261918.
36. Kawaguchi, Y.; Kusakabe, T.; Lee, J. M.; Koga, K. Characteristics of egg trait mutations in the silkworm, *Bombyx mori*. *J. Insect Biotech. Sericol.* **2009**, 78(3), 113–126.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.