

A Strand-Specific Transcriptomic Atlas of Corona-Discharge-Induced Diapause Termination in *Bombyx mori* Eggs

Yuanyuan Wen ^{1,†}, Linrong Wan ^{2,†} , Yuli Zhang ^{3,†} , Aichun Zhao ¹ , Wenfu Xiao ^{2,*} and Dingpei Long ^{1,*} 

¹ State Key Laboratory of Resource Insects, Institute of Sericulture and Systems Biology, Southwest University, Chongqing 400715, China; 18723669022@163.com (Y.W.); zhaoaichun@hotmail.com (A.Z.)

² Sericultural Research Institute, Sichuan Academy of Agricultural Sciences (SAAS), Nanchong 637000, China; wanlinrong1987@scsaas.cn

³ Sericultural Research Institute, Guangxi Zhuang Autonomous Region, Nanning 530007, China; zyl8324@126.com

* Correspondence: xiaowenfu@scsaas.cn (W.X.); dplong@swu.edu.cn (D.L.)

† These authors contributed equally to this work.

Abstract

Embryonic diapause is a reversible arrest of embryonic development that enables insects to survive adverse environmental conditions. In the silkworm (*Bombyx mori*), diapause is established during early egg development and can be efficiently terminated by corona discharge; however, the transcriptional responses associated with this physical stimulus remain insufficiently characterized. Here, we present a strand-specific RNA-seq dataset generated from diapause-destined silkworm eggs subjected to early or late corona discharge treatment. The dataset comprises 36 libraries from three experimental groups across eight developmental time points, with two biological replicates per condition. In total, 283.9 Gb of high-quality clean data were obtained, with average Q20 and Q30 values of 97.60% and 94.20%, respectively, and an average genome-mapping rate of 82.62%. The dataset includes raw reads, gene-expression matrices, lists of differentially expressed genes, novel transcript annotations, alternative splicing events, and SNP/InDel calls. This resource provides a time-resolved transcriptomic framework for investigating physical-stress-induced diapause termination and supports comparative studies of diapause regulation in silkworms and other insects.

Keywords: *Bombyx mori*; embryonic diapause; diapause termination; corona discharge; strand-specific RNA-seq; time-series transcriptome; data descriptor



Academic Editor: Rüdiger Pryss

Received: 1 June 2026

Revised: 26 June 2026

Accepted: 3 July 2026

Published: 13 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Summary

Diapause is a reversible embryonic developmental arrest that enables insects to synchronize development with seasonal and environmental conditions. Depending on the species, diapause may occur at the egg, larval, pupal, or adult stage. The silkworm (*Bombyx mori*) is both an economically important insect and a representative lepidopteran model, and its eggs enter diapause during early embryogenesis [1]. A transcriptomic dataset describing diapause termination in silkworm eggs is therefore valuable for studies of insect developmental regulation, seasonal adaptation, and sericultural production.

Corona discharge provides a non-chemical physical stimulus that can terminate diapause in silkworm eggs. Xian and Kobayashi first improved the corona discharge method to terminate silkworm egg diapause, and high hatching rates were obtained when diapause-destined eggs were treated at 20 h post oviposition [2]. However, the molecular events

linking this physical stimulus to diapause termination remain poorly understood. Recent work identifying CYCLE as a key regulator of diapause variation in *B. mori* highlights the importance of transcriptional and circadian regulatory pathways in seasonal developmental plasticity [3], but genome-wide, time-resolved data for corona-discharge-induced diapause termination remain limited.

In this study, strand-specific RNA sequencing was used to profile gene-expression dynamics in silkworm eggs after corona discharge treatment at different developmental stages. The resulting dataset captures early and late transcriptional responses, provides quality-controlled raw and processed transcriptomic resources, and supports downstream analyses of candidate genes, pathways, transcript-structure variation, and sequence variation associated with diapause termination.

2. Data Description

2.1. Data Overview

This dataset represents a strand-specific RNA-seq atlas of diapause termination in silkworm eggs. It includes 36 libraries from three treatment groups across eight developmental time points, with two biological replicates per sampled condition. Paired-end sequencing (2×150 bp) was performed on the Illumina HiSeq platform (Illumina, San Diego, CA, USA), generating 283.9 Gb of high-quality clean data. The data volume per sample ranged from 5.74 to 16.28 Gb, and the average Q20 and Q30 values were 97.60% and 94.20%, respectively.

Clean reads were mapped to the silkworm reference genome (Ensembl release 91), with mapping rates ranging from 79.50% to 84.03% and an average mapping rate of 82.62%. The dataset contains raw FASTQ files, gene-expression matrices, lists of differentially expressed genes (DEGs), novel transcript annotations in GTF format, alternative splicing events, and SNP/InDel calls. Raw sequencing data were deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA1470546.

2.2. Experimental Design and Sample Characteristics

To assess transcriptional responses associated with corona-discharge-induced diapause termination, three treatment groups were designed according to the developmental stage of silkworm eggs after oviposition. The experimental design enabled comparison between untreated diapause-destined eggs and eggs treated at early or late embryonic stages. All treatments were performed using standardized corona discharge parameters, and the experimental timeline is shown in Figure 1.

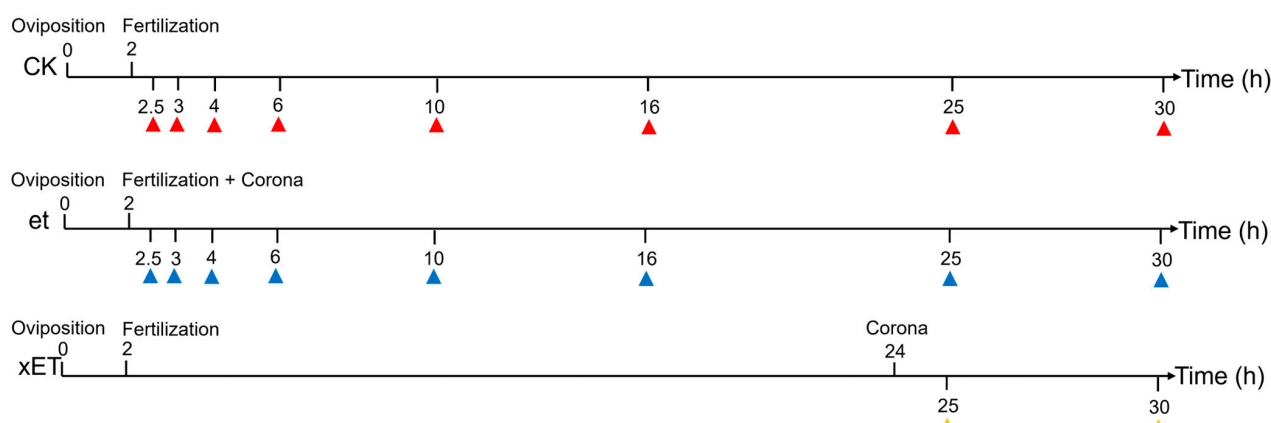


Figure 1. Experimental design and sampling time points for the control, early-treatment, and late-treatment groups. Triangles of different colors represent the sampling time points of different treatment groups.

The control group (CK) was not subjected to corona discharge. The early-treatment group (et) received corona discharge at 2 h post-oviposition (HPO), and samples were collected from 2.5 to 30 HPO. The late-treatment group (xET) received corona discharge at 24 HPO, and samples were collected at 25 and 30 HPO. Detailed sample information is provided in Table 1.

Table 1. Sample information.

Sample	Group	Description
CK2.5-1	CK2.5	Collection: 2.5 HPO
CK2.5-2	CK2.5	Collection: 2.5 HPO
CK3-1	CK3	Collection: 3 HPO
CK3-2	CK3	Collection: 3 HPO
CK4-1	CK4	Collection: 4 HPO
CK4-2	CK4	Collection: 4 HPO
CK6-1	CK6	Collection: 6 HPO
CK6-2	CK6	Collection: 6 HPO
CK10-1	CK10	Collection: 10 HPO
CK10-2	CK10	Collection: 10 HPO
CK16-1	CK16	Collection: 16 HPO
CK16-2	CK16	Collection: 16 HPO
CK25-1	CK25	Collection: 25 HPO
CK25-2	CK25	Collection: 25 HPO
CK30-1	CK30	Collection: 30 HPO
CK30-2	CK30	Collection: 30 HPO
et2.5-1	et2.5	Corona treatment: 2 HPO; Collection: 2.5 HPO
et2.5-2	et2.5	Corona treatment: 2 HPO; Collection: 2.5 HPO
et3-1	et3	Corona treatment: 2 HPO; Collection: 3 HPO
et3-2	et3	Corona treatment: 2 HPO; Collection: 3 HPO
et4-1	et4	Corona treatment: 2 HPO; Collection: 4 HPO
et4-2	et4	Corona treatment: 2 HPO; Collection: 4 HPO
et6-1	et6	Corona treatment: 2 HPO; Collection: 6 HPO
et6-2	et6	Corona treatment: 2 HPO; Collection: 6 HPO
et10-1	et10	Corona treatment: 2 HPO; Collection: 10 HPO
et10-2	et10	Corona treatment: 2 HPO; Collection: 10 HPO
et16-1	et16	Corona treatment: 2 HPO; Collection: 16 HPO
et16-2	et16	Corona treatment: 2 HPO; Collection: 16 HPO
et25-1	et25	Corona treatment: 2 HPO; Collection: 25 HPO
et25-2	et25	Corona treatment: 2 HPO; Collection: 25 HPO
et30-1	et30	Corona treatment: 2 HPO; Collection: 30 HPO
et30-2	et30	Corona treatment: 2 HPO; Collection: 30 HPO
xET25-1	xET25	Corona treatment: 24 HPO; Collection: 25 HPO
xET25-2	xET25	Corona treatment: 24 HPO; Collection: 25 HPO
xET30-1	xET30	Corona treatment: 24 HPO; Collection: 30 HPO
xET30-2	xET30	Corona treatment: 24 HPO; Collection: 30 HPO

2.3. Sequencing Data Quality Assessment

The basic metadata of each library is listed in Table 2. Sequencing quality was assessed for all 36 libraries (Tables S1 and S2; Figures S1–S3). The total raw data volume was approximately 540 GB (Table S1). After removal of adapter sequences and low-quality reads using Cutadapt v5.2, clean reads accounted for more than 99% of total reads, and clean data accounted for more than 98.5% of the raw data, yielding approximately 279 Gb of valid data (Table S2). Per-base quality plots showed that most bases exceeded Q20 (Figure S1). In sample CK3-2, minor quality fluctuations in the first six bases were consistent with random-primer-associated sequencing bias rather than sample degradation (Figure 2A,B). Base-composition plots showed no evident AT/GC bias

(Figures S2 and 2C,D), and average read-quality distributions showed sharp unimodal peaks centered at Q30–Q35 (Figures S3 and 2E,F). Strand specificity was further evaluated using the `infer_experiment.py` module in RSeQC. Across the 36 libraries, an average of 96.83% of assigned reads were consistent with the expected fr-firststrand orientation, whereas 1.72% were assigned to the opposite orientation and 1.45% were ambiguous or unassigned. The resulting expected-strand/opposite-strand ratio was approximately 56.3:1, confirming that the dUTP-based strand-specific library construction successfully preserved strand information. These results indicate high sequencing quality and support downstream transcriptomic analyses.

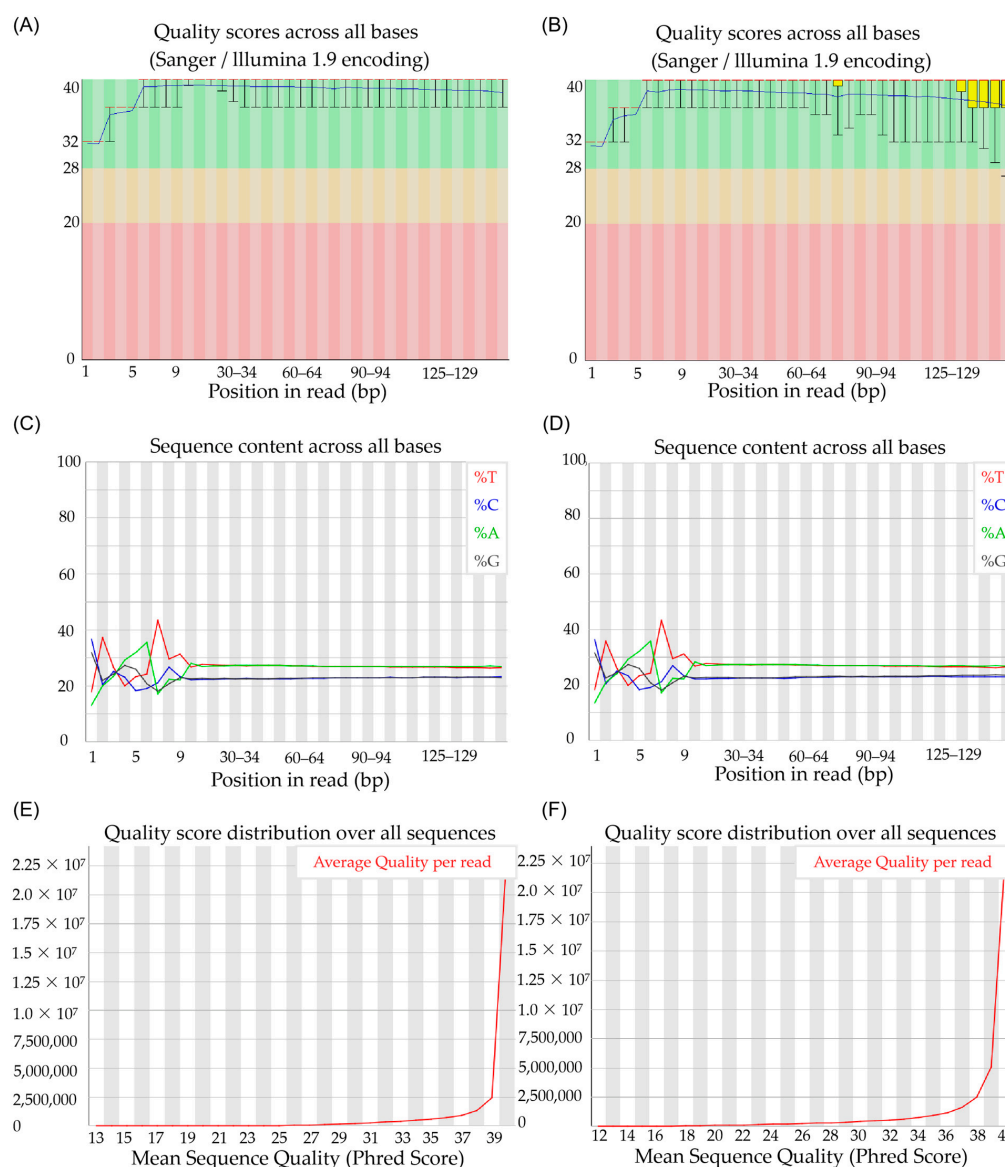


Figure 2. Sequencing-quality profiles of RNA-seq reads. (A) Per-base quality distribution of read 1 (R1) in sample CK3-2. The green zone indicates Q values ≥ 30 (high-quality bases), the beige zone corresponds to Q values of 20–28 (acceptable-quality bases), and the red zone represents Q values < 20 (low-quality bases). (B) Per-base quality distribution of read 2 (R2) in sample CK3-2. The green zone indicates Q values ≥ 30 (high-quality bases), the beige zone corresponds to Q values of 20–28 (acceptable-quality bases), and the red zone represents Q values < 20 (low-quality bases). (C) Base-composition distribution of R1 in sample CK2.5-1. (D) Base-composition distribution of R2 in sample CK2.5-1. (E) Average read-quality distribution of R1 in sample CK2.5-1. (F) Average read-quality distribution of R2 in sample CK2.5-1.

Table 2. Basic Information of the Libraries.

Sample	Library	Library Insert Size	Sequencing Platform	Sequencing Mode
CK10-1	LR8k8344	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK10-2	LR8k8345	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK16-1	LR8k8346	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK16-2	LR8k8347	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK2.5-1	LR8k8348	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK2.5-2	LR8k8349	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK25-1	LR8k8350	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK25-2	LR8k8351	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK30-1	LR8k8352	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK30-2	LR8k8353	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK3-1	LR8k8354	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK3-2	LR8k8355	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK4-1	LR8k8356	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK4-2	LR8k8357	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK6-1	LR8k8358	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
CK6-2	LR8k8359	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et10-1	LR8k8360	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et10-2	LR8k8361	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et16-1	LR8k8362	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et16-2	LR8k8363	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et2.5-1	LR8k8364	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et2.5-2	LR8k8365	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
xET25-1	LR8k8366	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
xET25-2	LR8k8367	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et25-1	LR8k8368	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et25-2	LR8k8369	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et30-1	LR8k8370	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et30-2	LR8k8371	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
xET30-1	LR8k8372	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
xET30-2	LR8k8373	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et3-1	LR8k8374	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et3-2	LR8k8375	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et4-1	LR8k8376	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et4-2	LR8k8377	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et6-1	LR8k8378	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp
et6-2	LR8k8379	380 bp	Illumina HiSeq	Paired-end, 2 × 150 bp

2.4. Mapping Result Statistics

Clean reads were aligned to the silkworm reference genome (Ensembl release 91) using TopHat2 v2.1.1 [4]. Mapping statistics are summarized in Table S3, and chromosomal read distributions are shown in Figure S4. The number of mapped reads per sample ranged from 31 M to 86 M, with an average mapping rate of 82.62%; sample CK2.5-1 showed the highest mapping rate (84.03%). In sample CK30-1, read coverage across chromosomes was broadly even, with no obvious chromosome-level mapping bias (Figure 3A).

Read distributions across genomic regions are summarized in Table S4 and Figure S5. On average, 74.08% of reads mapped to genic regions, 97.27% of genic reads mapped to exonic regions, and 25.92% of reads mapped to intergenic regions, which is consistent with eukaryotic RNA-seq data. The distribution patterns were highly similar across the 36 samples, indicating stable library construction and sequencing performance.

Gene-body coverage analysis (Figure S6) showed the expected pattern of relatively lower coverage near the 5' and 3' ends and higher coverage across internal gene regions.

For sample CK25-1, the coverage curves were highly consistent, with no evidence of severe RNA degradation or systematic positional bias (Figure 3B).

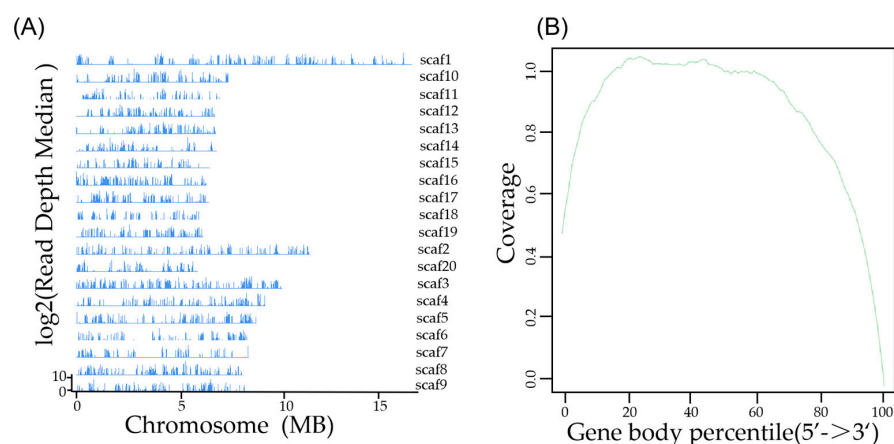


Figure 3. Genomic distribution of sequencing reads. (A) Chromosomal read distribution in sample CK30-1. (B) Gene-body sequencing-depth distribution in sample CK25-1.

2.5. Gene Expression Quantification

Gene expression levels were normalized as fragments per kilobase of transcript per million mapped reads (FPKM), and expression matrices for all samples are provided in Table S5. Across the 36 samples, 9380–10,680 expressed genes ($\text{FPKM} \geq 0.1$) were detected, with an average of 9818 genes per sample. The FPKM density plot (Figure 4A) and box plot (Figure 4B) showed comparable expression distributions among samples, indicating that the data were suitable for cross-sample comparison.

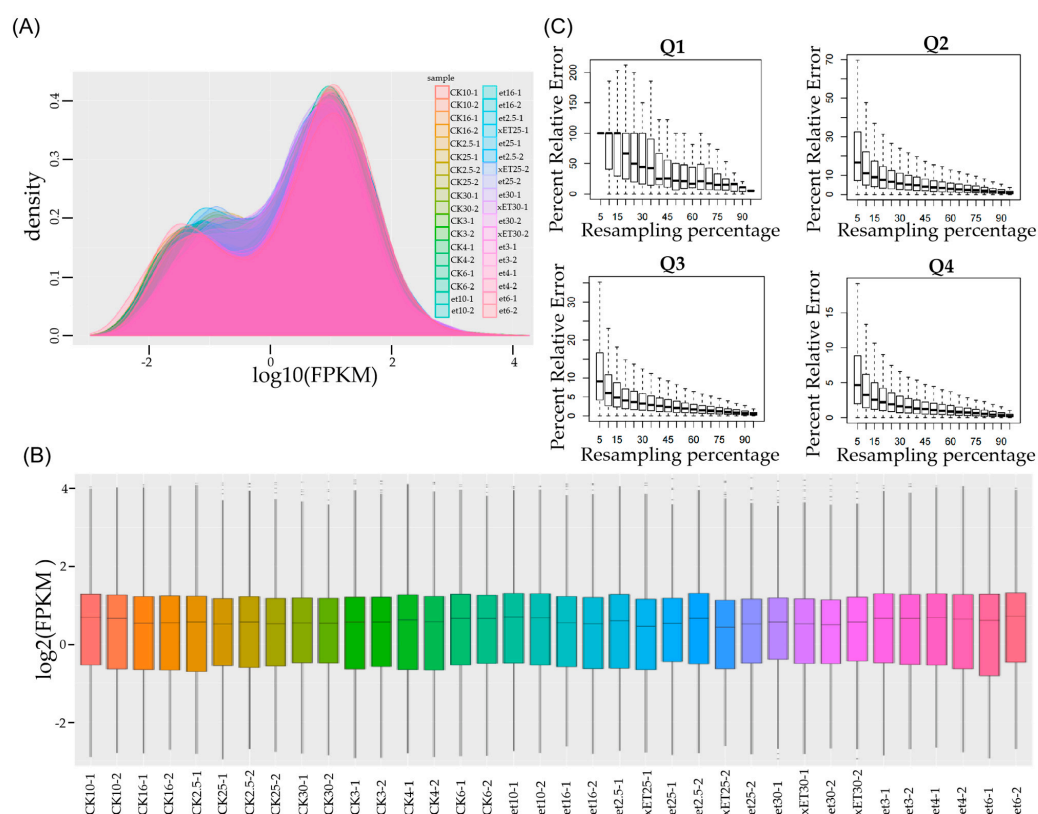


Figure 4. Gene-expression distribution and sequencing saturation. (A) FPKM density distribution. (B) FPKM box plot. (C) RPKM sequencing-saturation curve for sample CK10-1.

Sequencing saturation was evaluated using RSeQC v2.6.4 [5] (Figure S7). In sample CK10-1, lowly expressed genes reached saturation more slowly (Q1), with the relative error decreasing below 5% at a 95% resampling rate, whereas highly expressed genes reached saturation rapidly (Q4), with the relative error decreasing below 5% at a 5% resampling rate (Figure 4C). These results indicate that the sequencing depth was sufficient for robust expression quantification across a wide expression range.

Pearson correlation analysis showed that biological replicates within the same group were highly correlated, supporting the reproducibility of the experimental design (Figure 5). Lower correlations among samples from different developmental stages reflected marked temporal changes in gene-expression profiles during silkworm egg development.

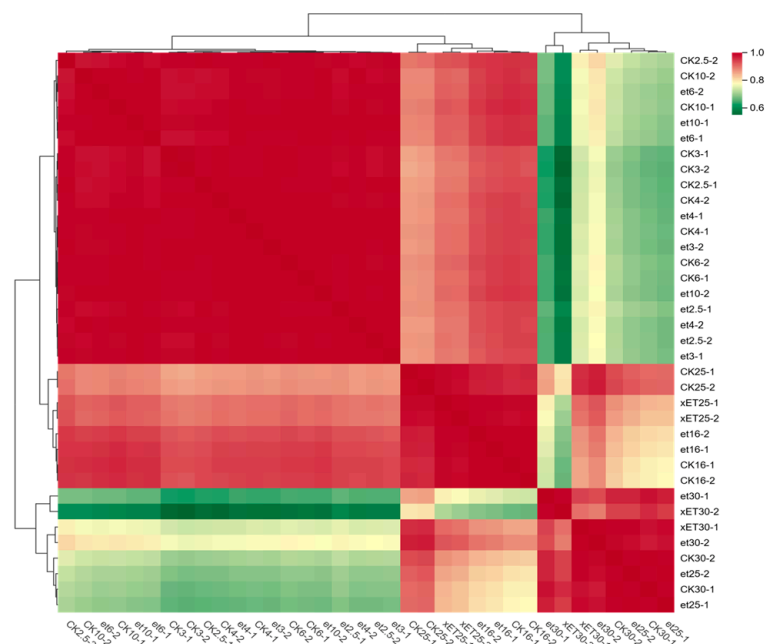


Figure 5. Pearson correlation analysis among RNA-seq samples.

2.6. Differential Expression Analysis

Differentially expressed genes were identified using $|\log_2\text{FoldChange}| > 1$ and adjusted p value (padj) < 0.05 as screening thresholds. DEG statistics for each pairwise comparison are provided in Tables S6 and S7. For example, the CK30 versus CK25 comparison identified 1856 DEGs, including 826 upregulated and 1030 downregulated genes. Volcano and MA plots for this comparison showed balanced distributions of differentially expressed and non-differentially expressed genes, supporting the reliability of normalization and differential analysis (Figure 6A,B). Additional volcano and MA plots are provided in Figures S8 and S9.

To explore candidate genes associated with corona-discharge-induced diapause termination, DEG overlaps were examined among CK30_vs_et30, CK30_vs_xET30, CK25_vs_et25, and CK25_vs_xET25. Eleven overlapping DEGs were identified across two sets of comparisons designated F1 and F2, where F1 represents the two comparisons CK25_vs_et25 and CK25_vs_xET25, and F2 represents CK30_vs_et30 and CK30_vs_xET30 (Figure 6C), suggesting a set of candidate transcriptional markers responsive to corona discharge. A DEG matrix was also constructed to compare stage-specific overlaps (Figure 6D). Venn diagrams, matrix plots, and DEG information tables for additional comparisons are provided in Figure S10 and Table S8. Transcriptional regulation sample grouping information is provided in Table S9.

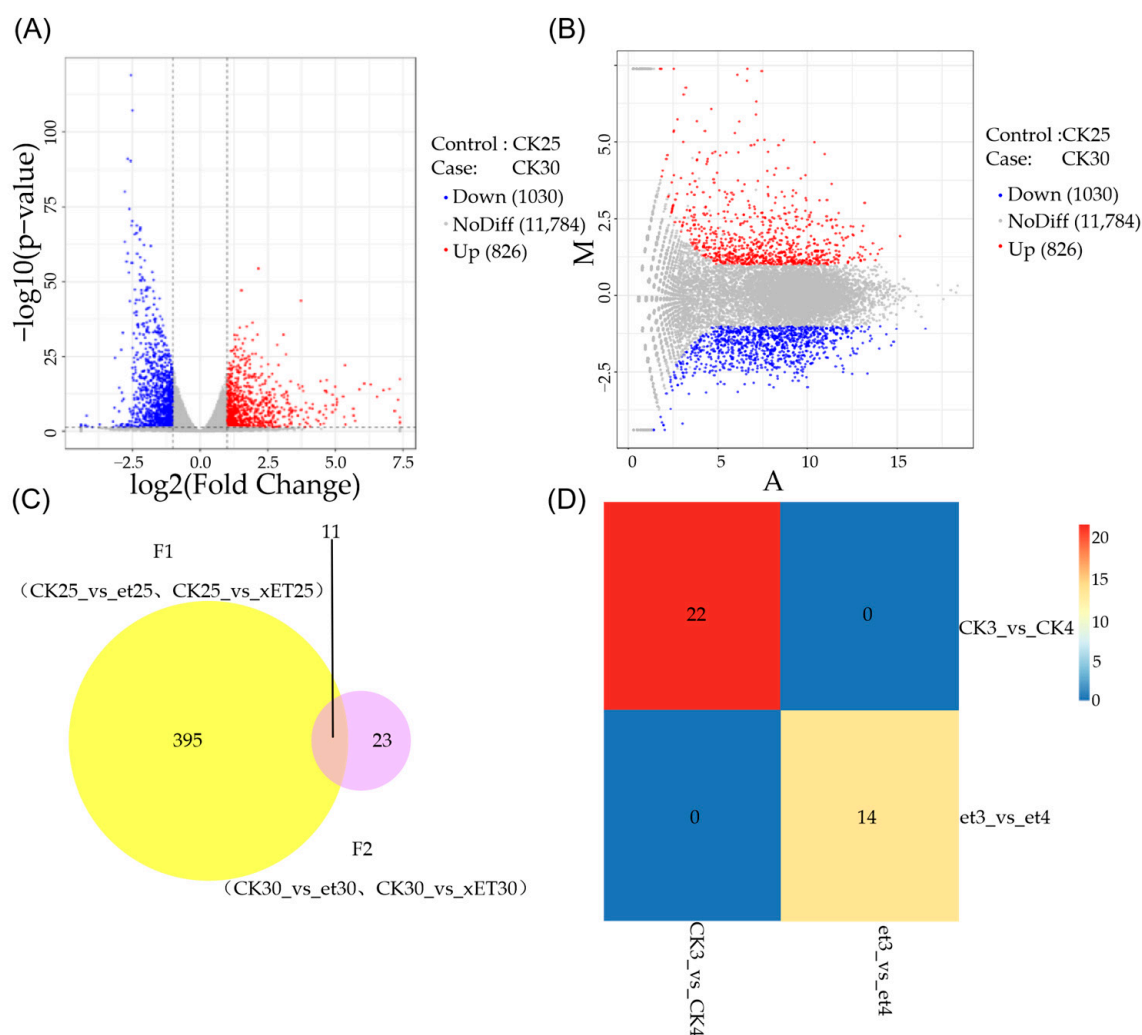


Figure 6. Differential expression analysis. **(A)** Volcano plot of DEGs in the CK30 versus CK25 comparison. **(B)** MA plot for CK30 versus CK25. **(C)** Venn diagram showing overlapping DEGs between F1(CK25_vs_et25 and CK25_vs_xET25) and F2 (CK30_vs_et30 and CK30_vs_xET30). **(D)** Heatmap showing expression clustering of DEGs from the two comparisons, CK3_vs_CK4 and et3_vs_et4.

Two-way clustering separated samples into two broad developmental phases, corresponding to early (2.5–10 HPO) and late (16–30 HPO) egg development (Figure 7). All CK samples clearly split into early and late subgroups along this developmental trajectory, while late corona-treated samples (et25 and et30) clustered closely with the late CK group. This pattern suggests that early corona discharge treatment may accelerate, or partially compress, embryonic developmental progression, which requires further verification in future research. DEGs were further grouped into 28 expression clusters (Figure S11). In Cluster 26, for example, BGIBMGA007397 (Chrb) was progressively upregulated during the late stage in the CK group but was expressed at lower levels in the et group at matched time points, indicating that corona treatment altered the temporal expression pattern of the selected diapause-related gene (Table S10).

2.7. Functional Enrichment Analysis of DEGs

Gene Ontology (GO) enrichment analysis of differentially expressed genes was conducted via a hypergeometric test with the whole genome as the background (Table S11). GO directed acyclic graphs (DAGs) were used to visualize the hierarchical relationships among enriched terms in the biological process, cellular component, and molecular function categories (Figure S12).

Furthermore, KEGG pathway enrichment analysis was performed to identify significantly enriched metabolic and signaling pathways. KO pathway maps were applied to clarify the distribution and expression variation characteristics of differentially expressed genes within core pathways (Table S12, Figure S13), and relevant pathway URLs were summarized in Table S12.

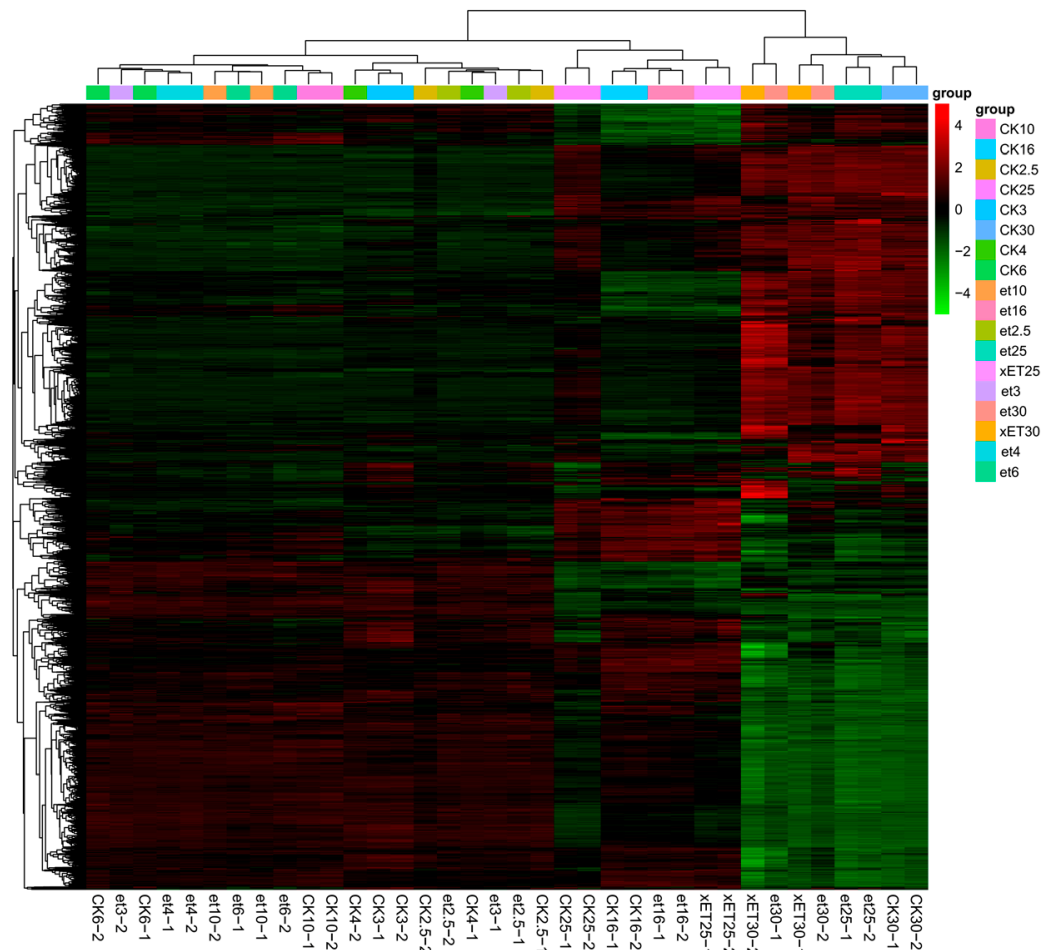


Figure 7. Two-way clustering analysis of DEGs across developmental stages and treatment groups.

2.8. Transcript Structure Analysis and Genome Variation

Transcript-structure analysis identified 146,769 novel transcripts corresponding to 14,406 gene loci. Class codes “j” and “u” accounted for 32.13% and 52.59% of novel transcripts, respectively. Known and novel transcript statistics are shown in Figure 8A, and detailed novel transcript annotations are provided in GTF format (Table S13). In addition, 821 5' UTRs and 1141 3' UTRs were extended or optimized, involving 1950 genes (Table S14). Twelve categories of alternative splicing events were identified, with per-sample counts summarized in Table S15 and Figure 8B.

Variant calling identified 173,260 SNP loci and 11,737 InDel loci. Homozygous and heterozygous SNP counts per sample are shown in Figure 8C, and the transition/transversion ratio was 1.77 (Figure 8D). Homozygous and heterozygous InDel counts are shown in Figure 8E. Detailed SNP and InDel information, including genomic position and nucleotide changes, is provided in Tables S16 and S17.

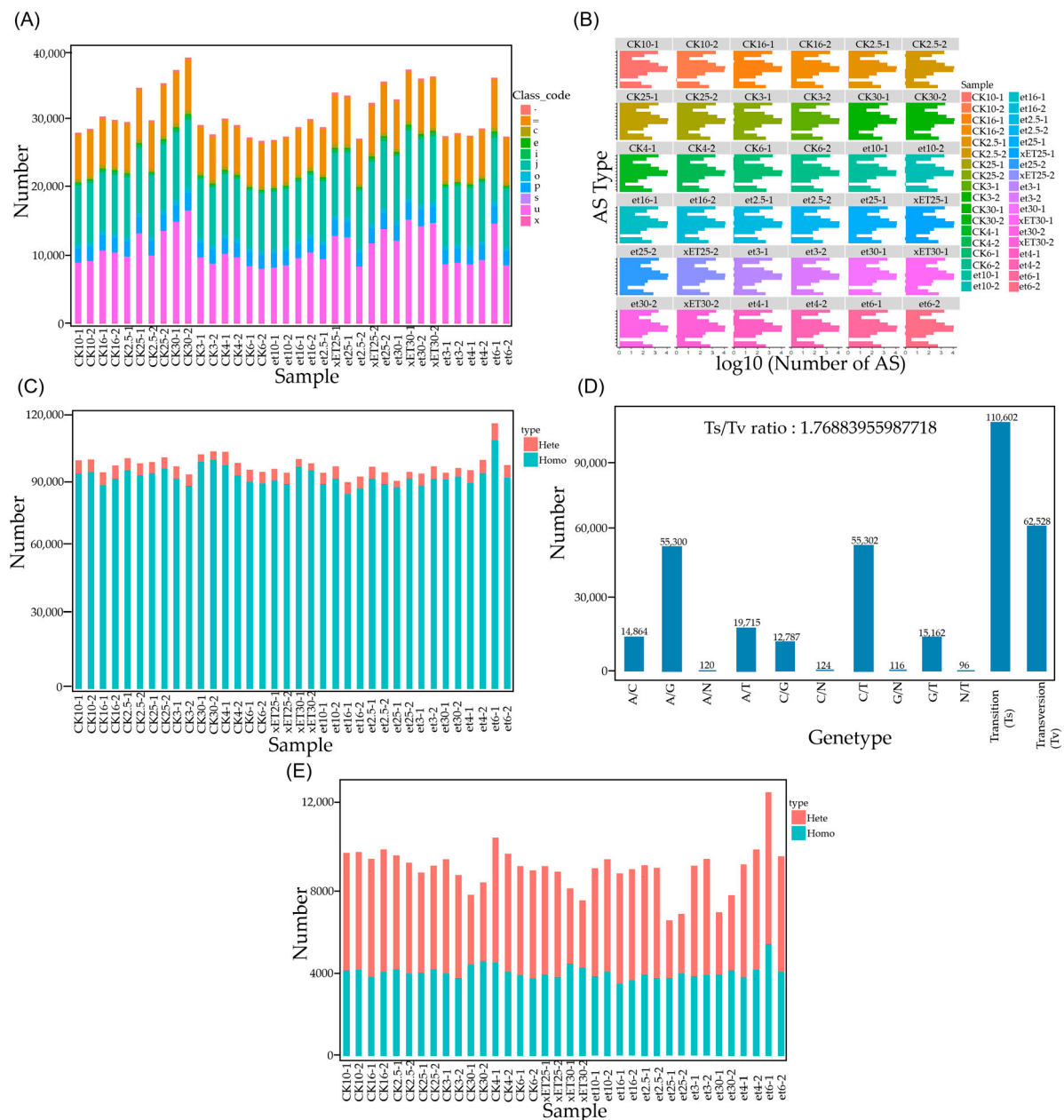


Figure 8. Summary of transcript features and sequence-variation statistics. **(A)** Known and novel transcript statistics. **(B)** Statistics of alternative splicing (AS) events. The horizontal axis shows the log10-transformed number of AS events, and each panel represents one sample. The y-axis shows different AS event types, listed from top to bottom as follows: AE, alternative exon ends; IR, intron retention; MIR, multi-intron retention; MSKIP, multi-exon skipping; SKIP, skipped exon; TSS, alternative transcript start site; TTS, alternative transcript termination site; XAE, approximate alternative exon ends; XIR, approximate intron retention; XMIR, approximate multi-intron retention; XMSKIP, approximate multi-exon skipping; XSKIP, approximate skipped exon. Different colors represent different sample groups. **(C)** SNP statistics. **(D)** Transition/transversion (Ti/Tv) distribution. **(E)** InDel statistics.

3. Methods

3.1. Experimental Materials and Sample Preparation

The diapause strain Dazao was provided by the Gene Bank of the State Key Laboratory of Resource Insects High-Efficiency Cultivation and Utilization, Southwest University. Silkworms were reared at 25 ± 1 °C and 75% relative humidity under a 12 h light/12 h dark

photoperiod. Eggs were assigned to three groups: CK (untreated control), et (corona discharge at 2 HPO), and xET (corona discharge at 24 HPO). Corona treatment was performed at 12 kV for 1 min with an 8 mm electrode distance. The 2 h post-oviposition (2 HPO) treatment was selected because this time point corresponds to early zygotic development, where male and female pronuclei converge and fuse to form the zygotic nucleus under standard incubation conditions [6]. Previous work showed that corona treatment within 2–4 h after oviposition efficiently terminates diapause [7]. The 24 HPO treatment was selected to represent a later embryonic stage around germ-band development. The sampling scheme is shown in Figure 1. Collected eggs were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction.

Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. RNA integrity was examined by 1% agarose gel electrophoresis, and RNA quantity and integrity were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Only RNA samples with RNA integrity number (RIN) values ≥ 7.0 were used for library construction.

3.2. Library Construction and Sequencing

Poly(A)+ mRNA was enriched using oligo(dT) magnetic beads and fragmented to 200–300 bp. Strand-specific libraries were constructed using a dUTP-based second-strand synthesis strategy, in which dUTP replaced dTTP during second-strand cDNA synthesis. The library insert size was approximately 380 bp, and the final library fragments were 300–400 bp. Qualified libraries were sequenced on the Illumina HiSeq platform using paired-end 2×150 bp sequencing.

3.3. Data Preprocessing and Mapping

Raw reads were processed with Cutadapt v5.2 to remove adapters and low-quality sequences. A reference index was built with Bowtie2 [8], and clean reads were aligned to the silkworm reference genome (Ensembl release 91) using TopHat2 v2.1.1 [4]. The alignment parameters allowed up to two mismatches, enabled microexon search, and specified the strand-specific library type as fr-firststrand. Strand specificity was assessed using the infer_experiment.py module in RSeQC v2.6.4 with the aligned BAM files and the silkworm reference gene annotation. Reads were classified according to whether their mapping orientation was consistent with the expected fr-firststrand pattern, assigned to the opposite orientation, or ambiguous/unassigned.

3.4. Gene Expression and Differential Analysis

Gene-level read counts were calculated using HTSeq v2.1.2 [9] in union mode, and expression levels were normalized as FPKM. Transcript assembly and gene/transcript quantification were performed using the TopHat-Cufflinks pipeline [10]. Sequencing saturation was evaluated with RSeQC v2.6.4 [5]. Pearson correlation coefficients and principal component analysis (PCA) were used to assess sample relationships. Differential expression analysis was performed using DESeq2 v1.18 [11], with DEGs defined as genes showing $|\log_2\text{FoldChange}| > 1$ and adjusted p -value (padj) < 0.05 . The Benjamini–Hochberg method was used to correct for multiple testing and to calculate adjusted p -values. Based on the sequencing saturation results shown in Figure 4C, the $|\log_2\text{FoldChange}| > 1$ cutoff was used to retain moderate early transcriptional changes in diapause-related genes while reducing background noise. Volcano plots, MA plots, and heatmaps were generated in R using ggplot2 v4.0.3 [12] and pheatmap v1.0.13 [13].

3.5. Functional Enrichment and Transcript Structure Variation Analysis

GO and KEGG enrichment analyses of DEGs were performed using the hypergeometric test against the whole-genome background. Gene functional classification and orthologous annotation were supported by COG [14] and eggNOG [15]. GO analysis was used to identify enriched biological process, cellular component, and molecular function terms [16,17]. KEGG analysis was used to identify enriched metabolic and signaling pathways [18–21].

3.5.1. Novel Transcript Identification

Transcripts were assembled for each sample using StringTie [22], and assembled transcripts were compared with the silkworm reference annotation (Ensembl release 91) using gffcompare [23]. Novel transcripts were classified by class code, with major categories defined in Table 3.

Table 3. Classification of Class Codes.

Class Code	Definition
j	Novel transcript sharing at least one splice site with the reference.
e	Single-exon transcript overlapping exon and ≥ 10 bp intron of reference; putative pre-mRNA.
i	Transcript fully contained within a reference intron.
o	Transcript mostly covering the exons of the reference.
p	Putative nascent transcript within 2000 bp of the reference.
r	Repeat-associated transcript ($\geq 50\%$ in repeat regions).
u	Novel intergenic transcript.
x	Antisense transcript overlapping reference exons.
s	Antisense transcript with intron overlaps; possible alignment artifact.
.	Transcript with multiple classifications.

3.5.2. UTR Region Extension

Candidate UTR extensions were defined as read-supported regions located upstream or downstream of annotated coding sequences. Novel 5' UTR and 3' UTR regions were identified by comparison with existing reference annotations to refine gene-structure models.

3.5.3. Alternative Splicing Analysis

Alternative splicing events were classified and quantified using ASprofile v1.0.4 [24] based on StringTie-assembled gene models. Twelve event types were identified, as summarized in Table 4.

Table 4. Twelve Types of Alternative Splicing Events.

Event Code	Description
AE	Alternative exon ends (5', 3', or both)
IR	Intron retention (labeled as IR_ON/IR_OFF)
MIR	Multi-intron retention (labeled as MIR_ON/MIR_OFF)
MSKIP	Multi-exon skipping (labeled as MSKIP_ON/MSKIP_OFF)
SKIP	Skipped exon; SKIP_ON and SKIP_OFF occur as a pair
TSS	Alternative transcript start site

Table 4. *Cont.*

Event Code	Description
TTS	Alternative transcript termination site
XAE	Approximate alternative exon ends
XIR	Approximate intron retention (labeled as XIR_ON/XIR_OFF)
XMIR	Approximate multi-intron retention (labeled as XMIR_ON/XMIR_OFF)
XMSKIP	Approximate multi-exon skipping (labeled as XMSKIP_ON/XMSKIP_OFF)
XSKIP	Approximate skipped exon (labeled as XSKIP_ON/XSKIP_OFF)

3.5.4. SNP and InDel Detection

SNP and InDel loci were detected using VarScan v2.3.8 [25]. The filtering criteria were as follows: base quality > Q20, locus coverage > 8 reads, variant-supporting reads > 2, and p -value < 0.01.

3.6. Data Analysis Software

The main software tools and key parameters used in this study are summarized in Table 5.

Table 5. Introduction to Bioinformatics Software (all accessed on 2 July 2026).

Software	Function	Parameters
Cutadapt v5.2 (http://cutadapt.readthedocs.io/en/stable/)	Data filtering	Minimum 10 bp overlap (AGATCGGAAG), 20% base error rate allowed
FastQC v0.11.7 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc)	Quality control	Default parameters
TopHat2 v2.1.1 (http://tophat.cbcb.umd.edu/)	Read alignment	Microexon-search enabled; library-type = fr-firststrand
RSeQC v2.6.4 (http://rseqc.sourceforge.net)	RPKM saturation analysis and strand-specificity assessment	Default parameters for saturation analysis; infer_experiment.py for strand-specificity evaluation
HTSeq v2.1.2 (https://pypi.org/project/HTSeq/)	Expression quantification	Union mode
DESeq2 v1.18 (https://www.bioconductor.org/packages/release/bioc/html/DESeq2.html)	Differential expression analysis	$ \log_2\text{FoldChange} > 1$ and adjusted p -value (padj) < 0.05; Benjamini–Hochberg multiple testing correction
ggplot2 v4.0.3 (https://cloud.r-project.org/web/packages/ggplot2/index.html)	Volcano plot, MA plot	Default parameters
pheatmap v1.0.13 (https://cran.r-project.org/web/packages/pheatmap/index.html)	Cluster analysis	Default parameters
topGO v2.8.0 (http://www.bioconductor.org/packages/release/bioc/html/RamiGO.html)	GO DAG enrichment	Default parameters

Table 5. Cont.

Software	Function	Parameters
StringTie v1.3.4 (http://ccb.jhu.edu/software/stringtie/)	Transcript assembly and lncRNA quantification	Default parameters
ASprofile v1.0.4 (http://ccb.jhu.edu/software/ASprofile/)	Alternative splicing event statistics	Default parameters
VarScan v2.3.8 (http://varscan.sourceforge.net/)	SNP and InDel detection	--p-value 0.01 --min-avg-qual 20

4. User Notes

This dataset provides a high-quality time-series transcriptomic resource for silkworm eggs subjected to untreated control, early corona discharge, and late corona discharge conditions. It can be used to compare temporal gene-expression dynamics, identify candidate genes associated with diapause termination, and examine transcript-structure and sequence-variation features during early embryonic development.

Users should consider several factors when reusing the dataset. First, the dataset contains two biological replicates per condition, which supports comparative exploration but may limit statistical power for subtle expression changes. Some analyses are well supported by two biological replicates, whereas others should be interpreted with caution. Global-level analyses, including overall gene expression dynamics, functional enrichment of differentially expressed gene sets, and identification of core diapause-associated gene candidates, are adequately supported under this replicate scheme. In contrast, single-gene-level differential expression results, quantitative detection of weakly expressed transcripts, and minor transcriptional variations carry elevated false-negative risks due to limited sample replication. Researchers utilizing this resource for follow-up experimental validation are advised to verify key target genes via qPCR or expand biological replicates where feasible. Second, treatment timing and embryonic developmental stage are closely linked in this system; therefore, developmental-stage effects should be considered when interpreting treatment-responsive genes. Third, the distinction between et and xET groups should be maintained carefully because they represent early and late corona discharge treatments, respectively.

The dataset was processed using a standardized bioinformatics workflow that included read filtering, genome alignment, expression quantification, differential expression analysis, enrichment analysis, transcript assembly, alternative splicing analysis, and variant calling. Reuse of the dataset should include careful inspection of metadata, sample group definitions, and processing parameters, particularly for time-series and treatment-comparison analyses.

The resource is suitable for mining candidate diapause-related genes, reconstructing transcriptional trajectories during diapause termination, comparing early versus late responses to physical stimulation, and integrating transcriptomic data with future functional validation experiments in *Bombyx mori* and related lepidopteran species.

Supplementary Materials: The supporting information can be downloaded at: <https://zenodo.org/records/20473686>, accessed on 2 July 2026.

Author Contributions: Conceptualization, Y.W., L.W. and D.L.; methodology, Y.W., L.W. and Y.Z.; software, Y.W. and L.W.; validation, Y.W., L.W., W.X. and D.L.; formal analysis, Y.W. and L.W.; investigation, Y.W., L.W. and Y.Z.; resources, W.X. and D.L.; data curation, Y.W., L.W. and Y.Z.;

writing—original draft preparation, Y.W., L.W. and Y.Z.; writing—review and editing, Y.W. and D.L.; visualization, Y.W. and L.W.; supervision, A.Z., W.X. and D.L.; project administration, D.L.; funding acquisition, A.Z., W.X. and D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (32360860), the Fundamental Research Funds for the Central Universities (SWU-KF25005 and SWU-XJPY202309), the Sichuan Aquatic (silkworm) livestock and poultry breeding project (1+9KJGG004), the Natural Science Foundation of Chongqing (cstc2021jcyj-cxtt0005), the Venture and Innovation Support Program for Chongqing Overseas Returnees (cx2024003), and the Earmarked Fund for CARS (CARS-18-ZJ0201).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw FASTQ sequencing reads generated in this work have been deposited in the NCBI Sequence Read Archive (SRA) under accession PRJNA1470546 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1470546>, accessed on 2 July 2026). The dataset is released under the Creative Zero (CC0) license. Processed transcriptomic data, including the full FPKM expression matrix, DEG lists, associated analytical files, and all supplementary figures and tables, are openly available via Zenodo at <https://zenodo.org/records/20473686>. The full FPKM matrix can be found in Table S5, while DEG lists and relevant supporting data are available in Tables S6–S17 within the archived Supplementary Materials package.

Acknowledgments: We thank Shanghai Personalbio Technology Co., Ltd., Shanghai, China for professional transcriptome sequencing and bioinformatics assistance. We also thank members of our laboratory for technical assistance and helpful discussions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Noguchi, H.; Hayakawa, Y. Dopamine is a key factor for the induction of egg diapause of the silkworm, *Bombyx mori*. *Eur. J. Biochem.* **2001**, *268*, 774–780. [[CrossRef](#)] [[PubMed](#)]
2. Xian, X.; Kobayashi, M. Change of electric potential in the silkworm egg during the embryogenesis induced by artificial hatching. *J. Sericultural Sci. Jpn.* **1997**, *66*, 485–487. [[CrossRef](#)]
3. Zheng, S.; Wang, Y.; Li, G.; Qin, S.; Dong, Z.; Yang, X.; Xu, X.; Fang, G.; Li, M.; Zhan, S. Functional polymorphism of CYCLE underlies the diapause variation in moths. *Science* **2025**, *388*, eado2129. [[CrossRef](#)] [[PubMed](#)]
4. Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S.L. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **2013**, *14*, R36. [[CrossRef](#)] [[PubMed](#)]
5. Wang, L.; Wang, S.; Li, W. RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* **2012**, *28*, 2184–2185. [[CrossRef](#)]
6. Nakao, H. Early embryonic development of *Bombyx*. *Dev. Genes Evol.* **2021**, *231*, 95–107. [[CrossRef](#)] [[PubMed](#)]
7. Zhang, Y.L.; Huang, Y.; Wang, P.Y.; Li, Q.; Bi, L.H.; Zhao, A.C.; Xiang, Z.H.; Long, D.P. Very Early Corona Treatment-Mediated Artificial Incubation of Silkworm Eggs and Germline Transformation of Diapause Silkworm Strains. *Front. Bioeng. Biotechnol.* **2022**, *10*, 843543. [[CrossRef](#)] [[PubMed](#)]
8. Langdon, W.B. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Min.* **2015**, *8*, 1. [[CrossRef](#)] [[PubMed](#)]
9. Anders, S.; Pyl, P.T.; Huber, W. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **2015**, *31*, 166–169. [[CrossRef](#)] [[PubMed](#)]
10. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **2012**, *7*, 562–578. [[CrossRef](#)] [[PubMed](#)]
11. 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. [[CrossRef](#)] [[PubMed](#)]

12. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed.; Springer International Publishing: Cham, Switzerland, 2016; ISBN 978-3-319-24277-4. [\[CrossRef\]](#)
13. Kolde, R. *pheatmap: Pretty Heatmaps*, R Package Version 1.0.13; CRAN: Windhoek, Namibia, 2025. Available online: <https://cran.r-project.org/package=pheatmap> (accessed on 25 June 2026).
14. Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E.V.; Krylov, D.M.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; et al. The COG database: An updated version includes eukaryotes. *BMC Bioinform.* **2003**, *4*, 41. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Powell, S.; Szklarczyk, D.; Trachana, K.; Roth, A.; Kuhn, M.; Muller, J.; Arnold, R.; Rattei, T.; Letunic, I.; Doerks, T.; et al. eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* **2012**, *40*, D284–D289. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Aleksander, S.A.; Balhoff, J.; Carbon, S.; Cherry, J.M.; Drabkin, H.J.; Ebert, D.; Feuermann, M.; Gaudet, P.; Harris, N.L.; Hill, D.P.; et al. The Gene Ontology knowledgebase in 2023. *Genetics* **2023**, *224*, iyad031. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **2004**, *32*, D277–D280. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **2010**, *38*, D355–D360. [\[CrossRef\]](#) [\[PubMed\]](#)
22. 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*, 290–295. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Pertea, G.; Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Res* **2020**, *9*, 304. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Florea, L.; Song, L.; Salzberg, S.L. Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000Res* **2013**, *2*, 188. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Koboldt, D.C.; Chen, K.; Wylie, T.; Larson, D.E.; McLellan, M.D.; Mardis, E.R.; Weinstock, G.M.; Wilson, R.K.; Ding, L. VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **2009**, *25*, 2283–2285. [\[CrossRef\]](#) [\[PubMed\]](#)

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.