Article

Transcriptomic Analysis Reveals Patterns of Expression of Stage-Specific Genes in Early *Apis cerana* Embryos

Runlang Su [1,2,3,†], Yuhui Chen [2,3,†], Rui Zhu [2,3], Guiling Ding [2,3], Kun Dong [1], Mao Feng [2,3,*] and Jiaxing Huang [1,2,3,*]

[1] College of Animal Science and Technology, Yunnan Agricultural University, Kunming 650201, China; surunlang@gmail.com (R.S.)
[2] State Key Laboratory of Resource Insects, Chinese Academy of Agricultural Sciences, Beijing 100093, China; chenyuhui0902@163.com (Y.C.); 82101225470@caas.cn (R.Z.)
[3] Key Laboratory for Insect-Pollinator Biology of the Ministry of Agriculture and Rural Affairs, Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing 100093, China
*   Correspondence: fengmao@caas.cn (M.F.); huangjiaxing@caas.cn (J.H.)
†   These authors contributed equally to this work.

Abstract: Background/Objectives: *Apis cerana* development is described as comprising four stages: embryo, larva, pupa, and adult. There are significant differences between workers and drones in terms of physiological functions and social roles, and the formation of the organ primordia occurs during the embryonic stage. Therefore, the objective of this study is to investigate the differential expression of and alternative splicing of genes in worker and drone embryos and to explain their unique developmental patterns. Methods: Long-read sequencing (PacBio Iso-Seq) and short-read sequencing (Illumina RNA-Seq) were used to investigate worker and drone embryo gene expression differences in *A. cerana* across five developmental points (12, 24, 36, 48, and 60 h). Results: The study identified 59,254 common isoforms, with 5744 and 5106 isoforms specific to worker and drone embryos, respectively. Additionally, a new transcript of the *csd* gene was identified. The number of differentially expressed genes (3391) and differential splicing events (470 genes) peaked at the 24-h embryonic stage. Differential splicing events of *csd*, *dsx*, and *Y-y* were observed in the worker and drone embryos. Conclusions: The gene expression results indicated that the 24-h embryonic point is a critical period for the expression of genes related to developmental and behavioral differences between workers and drones. The findings provide a theoretical basis for future research on the developmental differences between workers and drones.

Keywords: honeybees; *A. cerana*; worker; drone; embryonic transcriptomes

1. Introduction

The life cycle of honeybees' development includes four stages: egg, larva, pupa, and adult [1,2]. The embryo is the first and most critical stage, in which the basic organs of the honeybees are formed. Gene expression during this stage significantly influences the morphology, appearance, and behavior of adult honeybees, which is tightly controlled by environmental and intracellular signals [3]. The fertilized diploid embryos of honeybees generally develop into queens and workers, while unfertilized embryos develop into haploid males [4]. Although there are no phenotypic differences in the *A. cerana* Fabricius, 1793, embryos and throughout the first four larval stages [5,6], the developmental duration differs markedly for queens (16 d), workers (20 d), and drones (23 d) under normal conditions.

These differences in the developmental processes are responsible for their differences in physiological functions and social roles [7] and specific adaptations to environmental conditions. For instance, drone embryos and larvae have more precise temperature regulation compared with workers, and drone pupae are more tolerant to temperature variability [8]. Therefore, understanding the transcriptomic differences between worker and drone embryos could provide a better understanding of the molecular mechanisms underlying their developmental disparities. This could provide valuable insights for optimizing breeding programs by selecting desirable traits and developing targeted conservation strategies to enhance colony resilience and adaptability.

Significant advances have been made in understanding the molecular mechanisms underlying honeybee embryo development, including the visualization of the key stages of embryogenesis and the development of techniques for genetic and molecular manipulation [9–11]. The differentially expressed genes between workers and drones (*Apis mellifera* Linnaeus, 1758) during embryo development are sex-specific [4]. The doublesex gene plays a key role in sex determination and reproductive development and influences physiological and behavioral differences between workers and drones [12]. Studies on transcription factors and regulatory networks have identified gene expression patterns influencing sex determination and cell fate and have provided new insights into the relationship between honeybee ontogeny and social behavior [13]. However, the understanding of the molecular characteristics of workers and drones during development is still limited.

Alternative splicing generates different messenger RNA (mRNA) splice isoforms from precursor mRNAs. Alternative splicing is ubiquitous in eukaryotic organisms and significantly enhances the diversity of transcriptomes and proteomes by generating multiple mRNA isoforms from a single gene. Thus, it is a critical post-transcriptional regulatory mechanism in eukaryotes and a key contributor to protein diversity, playing an essential role in phenotypic variation in honeybees [14,15]. Differences in sex-related phenotypes are regulated by selective splicing. For example, male and female somatic cells in *Drosophila* are determined by the selective splicing of the *dsx* gene [16]. In honeybees, genetic mapping and chain analysis have revealed that sex is determined by the heterozygosity of a single gene, the complementary sex determiner (*csd*) gene. Heterozygous individuals develop into females, while hemizygous or homozygous individuals develop into males or diploid males, respectively [17]. Sex-specific splicing was confirmed in the adult stage of honeybees by examining the expression of *dsxF1* and *dsxF2* in female samples and *dsxM* in male samples [18], and selective splicing differences are common between the sexes [19,20].

The previous assembly annotations of the *A. cerana* genome were based on short-read-length sequencing (RNA-Seq), which may have led to underannotated transcripts [3]. Some genes, particularly the different splicing isoforms of sex-determination-related genes, may have been overlooked, and long-read-length sequencing could be used to obtain full-length complementary DNA (cDNA sequences), which could provide more accurate and diverse transcripts [21–23]. PacBio Iso-Seq and Illumina RNA sequencing are used for biological transcriptome research [11,24]. Illumina RNA sequencing technology has been used to study the transcriptomic elements and dynamics of embryonic development in the honeybee genome [3]. Iso-Seq produces longer reads, offering more complete information and identifying complex splicing events and structural variations [25]. Iso-Seq technology does not require transcript assembly, which reduces the likelihood of errors, which can occur during the assembly of RNA-Seq short-read fragments. Many extensive transcriptome studies have combined Iso-Seq and RNA-Seq, such as in actinopterygian fish (e.g., *Gasterosteus aculeatus* Linnaeus, 1758) [26], mammals (e.g., goats) [27], and invertebrates (e.g., *Sphingonotus tsinlingensis* Zheng, Tu & Liang, 1963) [28]. The integration of these two

sequencing technologies could provide more comprehensive information on the honeybee transcriptome and reveal differences between drones and workers in the honeybee society.

To understand the growth and development patterns of drone and worker embryos and the expression characteristics during five developmental time points (12, 24, 36, 48, and 60 h), long-read (PacBio Iso-Seq) and short-read sequencing (Illumina RNA-Seq) were integrated to investigate the genes and gene expression patterns. Quantitative analysis was conducted on the embryo transcriptome using Illumina RNA-Seq to verify the alternative splicing events of sex- and phenotypic-related genes and changes in gene expression at the five developmental time points. The findings will enhance the understanding of *A. cerana* sexual differentiation during embryonic development and provide basic data on the transcriptome landscape of honeybee embryogenesis. Additionally, they will provide valuable insights into broader biological processes regulated by alternative splicing.

2. Materials and Methods

2.1. Sample Collection

A. cerana was collected from the Yunnan Cangyuan Experimental Station of the Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Cangyuan County, Yunnan Province. Five *A. cerana* colonies were randomly selected, and the mated sister queens were confined to egg-free frames for 6 h before being placed back into the hives. From each colony, 20–40 developing embryos were collected at 12, 24, 36, 48, and 60 h after egg laying. The samples were treated with lysate and stored at $-80\ ^{\circ}$C. A total of 30 samples (each with three biological replicates) were collected for Illumina RNA-Seq and PacBio Iso-Seq.

2.2. Total RNA Extraction and PacBio Iso-Seq and Illumina RNA-Seq

The total RNA was extracted from embryo samples stored in PRL lysate using the Flash Pure Total RNA Micro Kit (Cat, R516-PRL, Beijing, China) and following the manufacturer's protocol. Then, it was qualified and quantified as follows: (1) the RNA purity and concentration were examined using a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) and (2) the RNA integrity and quantity were measured using the Agilent 2100/4200 system (Agilent Technologies, Santa Clara, CA, USA).

An RNA library for RNA-Seq was prepared as follows. The mRNA was purified from the total RNA using Poly A and then fragmented into 300–350 bp fragments. Then, the first-strand cDNA was reverse-transcribed using fragmented mRNA and deoxynucleotide triphosphates (dNTPs; dATP, dTTP, dCTP, and dGTP), and second-strand cDNA synthesis was performed. The remaining double-strand cDNA overhangs were converted into blunt ends via exonuclease/polymerase activities. After the adenylation of the 3′ ends of the DNA fragments, the sequencing adaptors were ligated to the cDNA, and the library fragments were purified. The template was enriched by polymerase chain reaction (PCR), and the PCR products were purified to obtain the final library. The sequencing libraries were generated using the VAHTS mRNA-seq v2 Library Prep Kit for Illumina (Vazyme, Nanjing, China) following the manufacturer's recommendations, and index codes were added to identify the sequences. After library construction, the concentration of the library was measured using the Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The size distribution of the library was detected by agarose gel electrophoresis. After library preparation and pooling of the different samples, the samples were subjected to Illumina sequencing.

A total amount of 2 μg of RNA was used for sample preparation. PacBio single-molecule real-time (SMRT) bell library preparation was performed using the SMRTbell Express Template Prep Kit 2.0 (PN 101-853-100, Pacific Biosciences, Menlo Park, CA, USA)

following the manufacturer's instructions (Procedure & Checklist-Iso-Seq Express Template Preparation for Sequel and Sequel II Systems). In brief, (i) the first strand of cDNA was synthesized by reverse transcription after adding primer with Oligo dT paired with Poly (A). (ii) Then, template switching was performed by adding the other primer paired with TS oligo. (iii) Subsequently, full-length cDNA was obtained by cDNA amplification with barcoded cDNA PCR primers. (iv) The amplified cDNA was purified with ProNex Beads and quantified by a Qubit 2.0 Fluorometer (Life Technologies). (v) Next, the SMRTbell library was prepared using Kit 2.0. (vi) Then, single-strand overhangs were removed, and DNA damage reparation, end-repair, A-tailing, and adapter ligation were conducted. After library purification using 0.45X AMPure PB beads and 1–10 Kb size selection, the library was prepared for sequencing on the Sequel II system (Pacific Biosciences).

2.3. Long-Read RNA Alignment and Isoform Identification

The worker and drone embryo samples produced an average of 1.24 million original sub-read segments (15.3 GB in total). Circular consensus sequences (CCSs) were created using ISOseq3 (v4.0.0; https://github.com/ylipacbio/IsoSeq3, accessed on 3 July 2024), and the barcode and primer sequences were removed by lima (v2.7.1; https://github.com/PacificBiosciences/barcoding, accessed on 3 July 2024) to generate FLNC sequences. These were compared to the reference genome (GCA_029169275-1) using pbmm2 (v1.13.1), and the transcript was generated and incorporated into isoforms by TAMA (v1.1.2) [29]. Then, SQANTI3 (5.1.1) was used to compare and classify the isoforms, filter false positive transcripts, and generate descriptive charts [30,31]. To maximize transcriptional identification, the total RNA was extracted from the worker and drone embryos of *A. cerana* at five different developmental time points (12, 24, 36, 48, and 60 h). The RNA samples of the worker and drone embryos at each time point were combined to construct two independent libraries for Iso-Seq. Each sample generated an average of 1,242,655 CCSs from the original segments, which were then filtered and clustered into full-length transcripts, generating an average of 703,781 high-quality consensus segments per sample (Supplemental Table S1).

2.4. Mapping and Quality Control of the Illumina RNA-Seq Data

Initially, the raw data (raw reads) in FASTQ format were processed using fastp (0.23.2). Clean reads were obtained by removing (1) reads containing adapters, (2) reads with more than 3 bases, and (3) reads with more than 20% bases with a Qphred ≤ 5 (Ref). Then, the Q20, Q30, and GC content of the clean data were calculated. Paired-end clean reads were then aligned to the reference genome of *A. cerana* (GCA_029169275.1) using Hisat2 (v2.2.1) [32]. Samtools (v1.6) was used to compare and sort the aligned SAM files into BAM files [33]. FeatureCount (v2.0.1) was used to count the number of reads mapped to each gene [34]. Finally, the Trinity (v2.15.1) software was used to perform differential expression analysis using the DESeq2 algorithm [35].

2.5. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Analysis

Emapper.py (2.1.10) was used to annotate the longest protein sequence in the reference of *A. cerana* genome using eggNOG database annotation, and the annotated file was converted into the org.db format for the gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis using the clusterProfiler package in R [36].

2.6. Quantitative Real-Time Polymerase Chain Reaction

We employed quantitative real-time PCR (qRT-PCR) technology to quantify the expression levels of alternatively spliced variants of the target gene. To eliminate genomic

DNA contamination and ensure RNA purity, a 5 × gDNA Clean Reaction Mix (Hunan Accurate Biotechnology Co., Ltd., Changsha, China, Catalog No. AG11728) was used. This kit effectively removes genomic DNA from RNA samples, preventing DNA contamination in subsequent reverse transcription and real-time PCRs and ensuring the accuracy and reliability of the samples. For the reaction system, 2.0 μL of 5 × gDNA Clean Reaction Mix was used, and the reaction was conducted at 42 °C for 2 min. The RNA template amount was less than 1000 ng, and RNase-Free water (Beijing Solarbio Science & Technology Co., Ltd., Beijing, China) was added to bring the total volume to 10 μL. Subsequently, the extracted RNA was reverse-transcribed into cDNA using a reverse transcription kit (Hunan Accurate Biotechnology Co., Ltd., Catalog No. AG11728). The reverse reaction was performed at 37 °C for 15 min, which was followed by a heat treatment at 85 °C for 5 s.

For the qRT-PCR amplification, a qRT-PCR kit (catalog number: AG11739; Hunan Accurate Biotechnology Co., Ltd.) was used for qRT-PCR. The qRT-PCR system included 2 × SG Green qRT-PCR Mix, forward and reverse primers (each 0.2 μL), and 1 μL of cDNA. The reaction conditions included an initial denaturation at 95 °C for 2 min, followed by 40 cycles of denaturation at 95 °C for 5 s, and an extension at 60 °C for 30 s.

Data analysis was performed using the $2^{-\Delta\Delta CT}$ method, with internal reference genes (Actin) employed to correct for differences in RNA concentration and reverse transcription efficiency between the samples. Subsequently, *t*-tests were conducted to evaluate the differences in the relative expression levels of the target gene among the different experimental treatment groups, and the relative expression levels for each group were calculated (Supplemental Table S7).

3. Results

3.1. Comparative Analysis of Sex-Specific Splicing and Gene Expression

The selective splicing difference analysis of the sex genes in the honeybee embryos identified specific shearing events in both embryos at five time points (Figure 1a). To improve the accuracy of gene annotation using Iso-Seq, isoforms with expression levels below 15 were filtered out. Of the isoforms expressed in embryos of both sexes (total 59,254), 5744 (7%) were worker-specific isoforms, while 5106 (6%) were drone-specific isoforms (Supplemental Figure S1). In comparison with the reference annotation, the 24-h stage exhibited the largest disparity in the number of new genes and splice events, especially for workers (Figure 1b). Compared with the reference gene, the stage-specific expression of the *csd* gene was identified at different developmental stages, and a novel isoform was identified (G109027.21). After filtering out the low expression values in the RNA-Seq data, the expression differences were analyzed (Figure 1c,d). The results showed that the isoform *G10927.1* exhibited the highest expression level in the 36-h embryos, while it was significantly suppressed in the drone embryos across all developmental stages. Conversely, the isoform *G10927.39* reached its peak expression in the drone embryos at the 48-h stage (Figure 1c, Supplemental Table S4). Notably, the overall expression of the novel isoform *G109027.21* was higher in the drones than in the worker embryos.

3.2. Analysis of Differentially Expressed Genes

Through RNA-Seq, 30 libraries generated 2,984,509,350 raw reads, with an average of 99,483,645 reads per sample. After quality control, a total of 2,970,242,921 clean reads were obtained, with an average of 99,008,097 reads per sample. The average quality scores for the clean reads, Q20 and Q30, were 98.87% and 96.38%, respectively, indicating high data quality (Supplemental Table S2). The sliding window density analysis showed that more than 94% of the transcripts were uniquely mapped to the reference genome (Supplemental Table S3). Among the five time points, the most significant gene expression

differences between the sexes were observed at 24 h, with 3391 differentially expressed genes identified (Figure 2a). After filtering out the low-expressed genes and adjusting the *p*-values (padjust > 0.05 and |logFC| > 1), the volcanic maps of the embryonic development were drawn to show the number of up-regulated, down-regulated, and non-significant genes at each stage. The 10 genes with the most significant *p*-values in each of the five time points were ranked, and a large number of new genes or new isoforms of differential expression were identified. In the drone and worker embryos, the 24-h stage exhibited the highest number of differentially expressed genes, reaching 3391, while the fifth stage had the lowest number of significant differentially expressed genes, at 207 (Figure 2b).
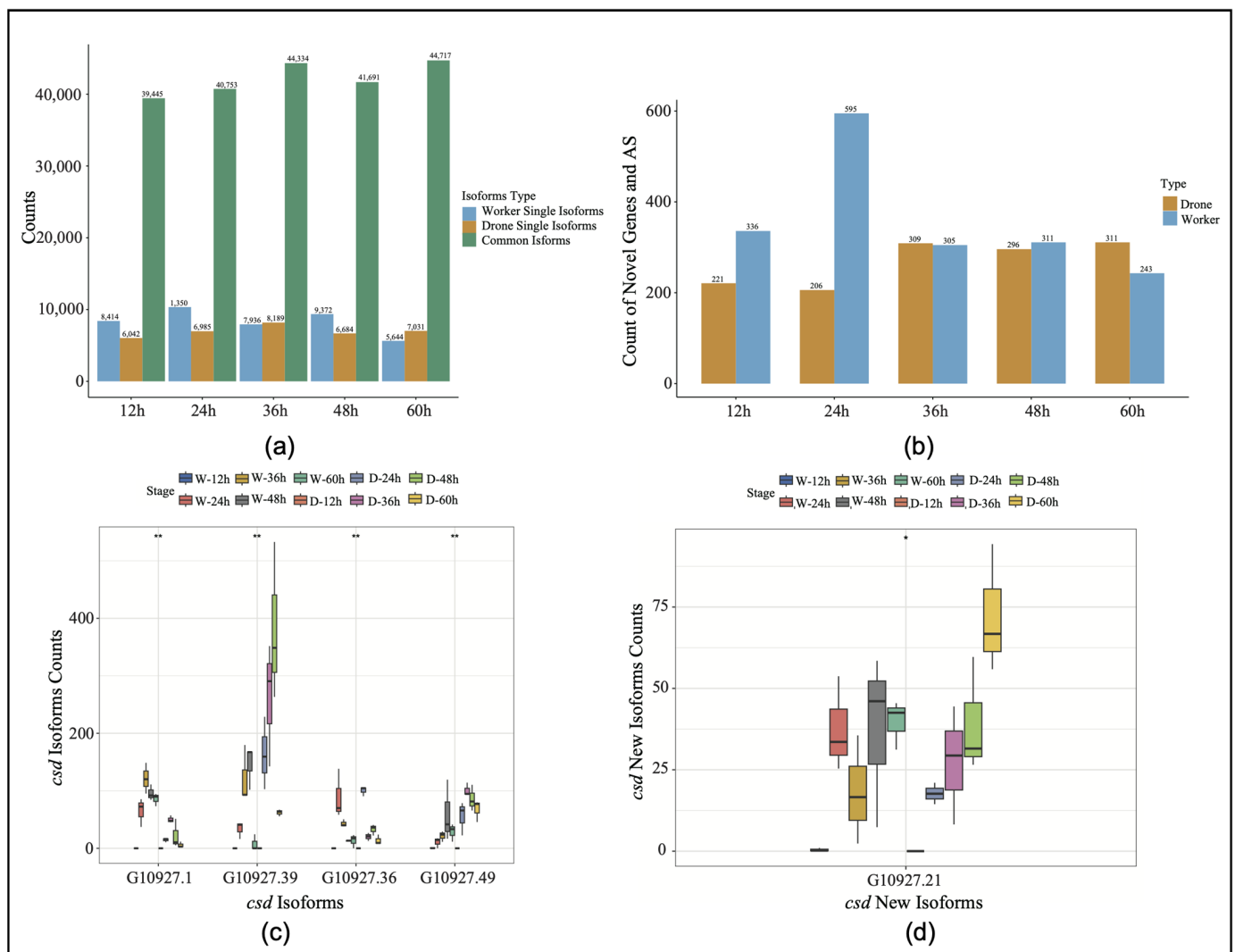
Figure 1. (**a**) Expression levels of specific isoforms and novel genes across five time points. W, worker; D, drone. (**b**) Discovery of novel genes and new splicing events in drones and workers. AS, alternative splicing. (**c**) The expression levels of the different isoforms of the *csd* gene (* *p* < 0.05, ** *p* < 0.01). (**d**) The expression levels of the various novel *csd* isoforms (* *p* < 0.05, ** *p* < 0.01).

Alternative splicing analysis verified the expression levels of the different gene transcripts at the five time points of embryonic development, and the number of alternative splicing genes and events at the five time points of embryonic development significantly overlapped with the distribution of differential gene expression. At the 24-h stage of embryonic development, the number of alternative first exon events reached its peak at 215, accounting for 45% of all splicing events. Additionally, the highest total number of alternative splicing events occurred at this time point (470; involving 352 genes). In contrast,

the stage with the fewest alternative splicing events had only 97 events, corresponding to 78 genes (Figure 2c,d).
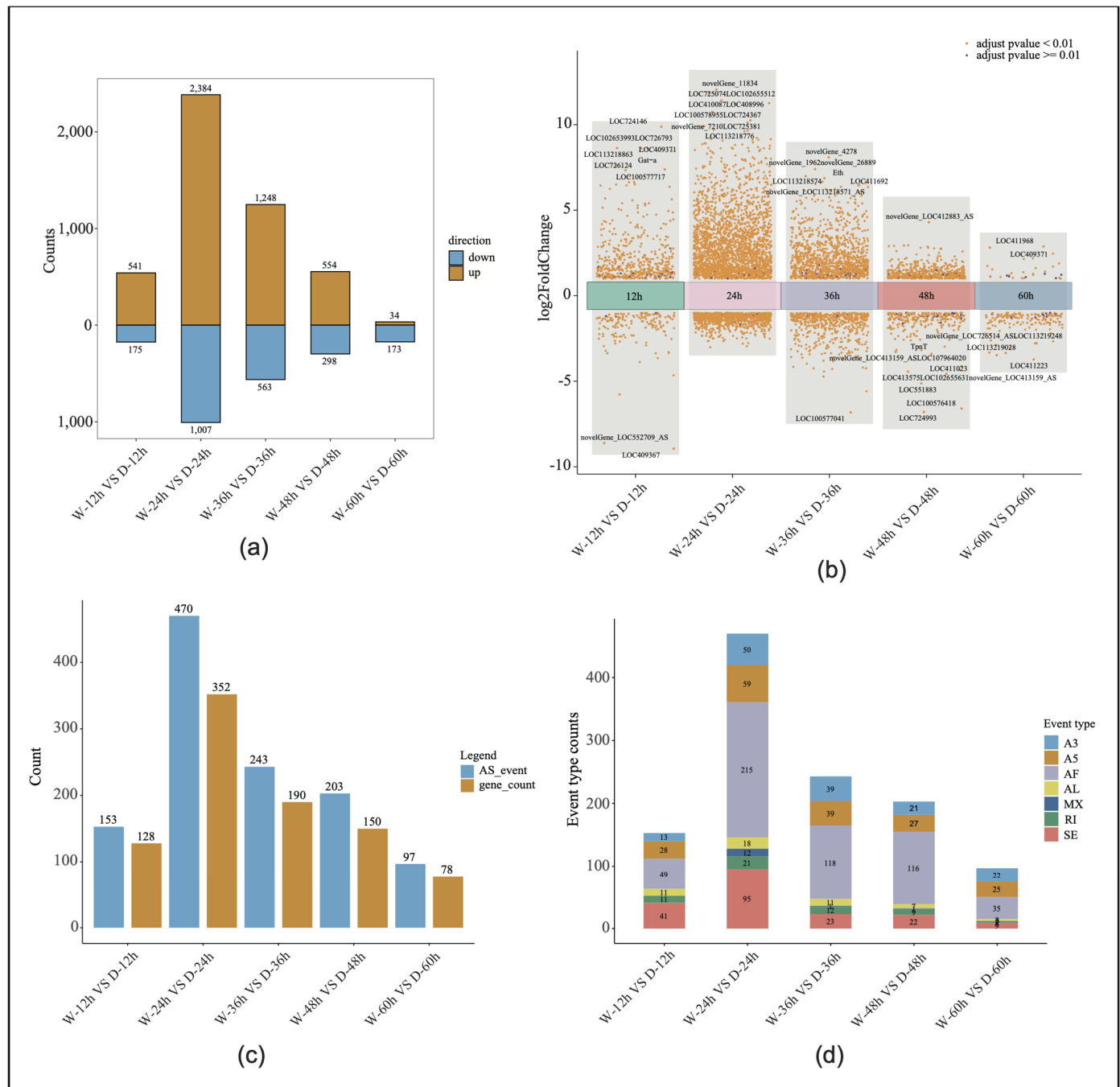

Figure 2. Differential gene expression analysis: (**a**) The number of significantly differentially expressed genes in both drone and worker embryos across five developmental time points. (**b**) Volcano plots depicting the differentially expressed genes in both drone and worker embryos at five developmental time points. W, worker; D, drone. (**c**) Variable splicing events and the corresponding gene count at five time points during worker and drone embryonic development. AS, alternative splicing. (**d**) Graph showing the number of seven types of alternative splicing events across five time points of worker and drone embryonic development.

3.3. Analysis of Gene Ontology Functional Enrichment

The results of the GO enrichment analysis (Supplemental Table S5) showed that the number of pathways involving significantly differentially expressed genes in the embryos

from the first to the fifth time points were 3, 108, 178, 363, and 54, respectively, with the fourth stage having the highest number of pathways. We selected 60 differentially expressed genes for the GO enrichment analysis (Figure 3a). The results indicated that genes related to sexual development (e.g., mating and male courtship behavior) showed significant differences during the third time point of embryonic development in both worker and drone embryos. Specifically, genes related to gonad development, learning and memory, eye development, muscle development, and neurological development also exhibited significant differential expression at the fourth time point.
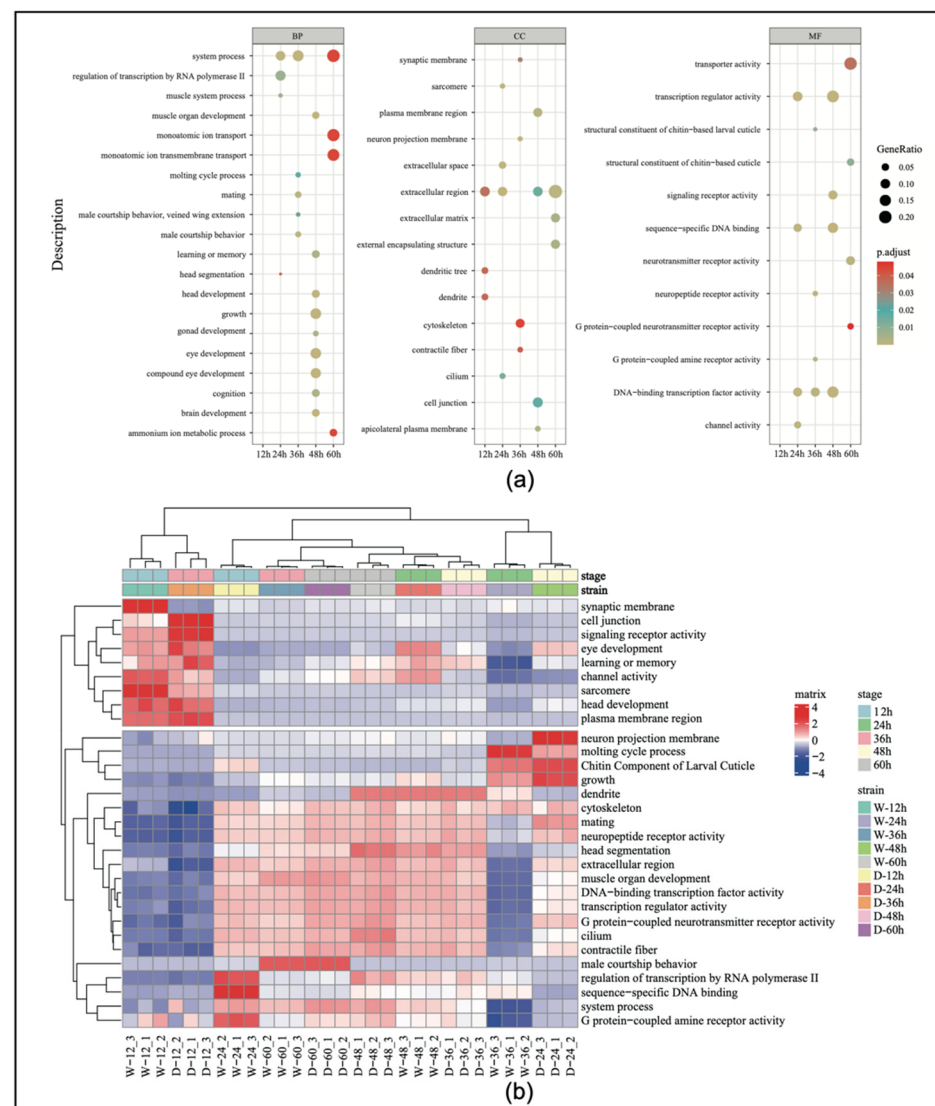
Figure 3. (**a**) Gene ontology (GO) enrichment categorization chart during embryonic development. (**b**) The heatmap of the top 30 genes by expression level in the GO enrichment analysis. BP, biological process; CC, cellular component; MF, molecular function.

To better understand the expression patterns and relationships of the genes across the different samples, the top 30 genes with the highest expression levels were extracted and expression cluster analysis was conducted. The results showed that genes associated with reproductive organ development began to increase in expression at 24 h of embryonic development (Figure 3b). Genes related to eye development showed significant changes at 24 h, and then, their level of expression gradually decreased. Genes associated with growth and development exhibited high specificity at 24 h, indicating that this period is a critical window for developmental regulation.

In this analysis, the clustering focused on five key developmental time points, and the selection of the top 30 genes was based on the highest expression levels and biological relevance across these stages. This approach captured dynamic expression changes and identified genes likely to play crucial roles in the embryonic development of both workers and drones.

3.4. Analysis of Sex-Specific Alternative Splicing

To further explore the gene differences between the sexes during embryonic development, we studied the retention intron involved in drone courtship behavior (*Y-y*), *dsx* gene, and *cds* gene in both worker and drone embryos at 24 h (Figure 4a–c). In addition, variable shear differences in brain development (*LOC725754*) and eye development (*LOC102654776*) were also identified (Supplemental Figure S3a,b). Additionally, quantitative PCR (qPCR) experiments were conducted to validate the alternative splicing of these genes. The results showed that the specific spliced forms of the *dsx*, *csd*, and *Y−y* genes were significantly more highly expressed in drones than in workers (Figure 5).
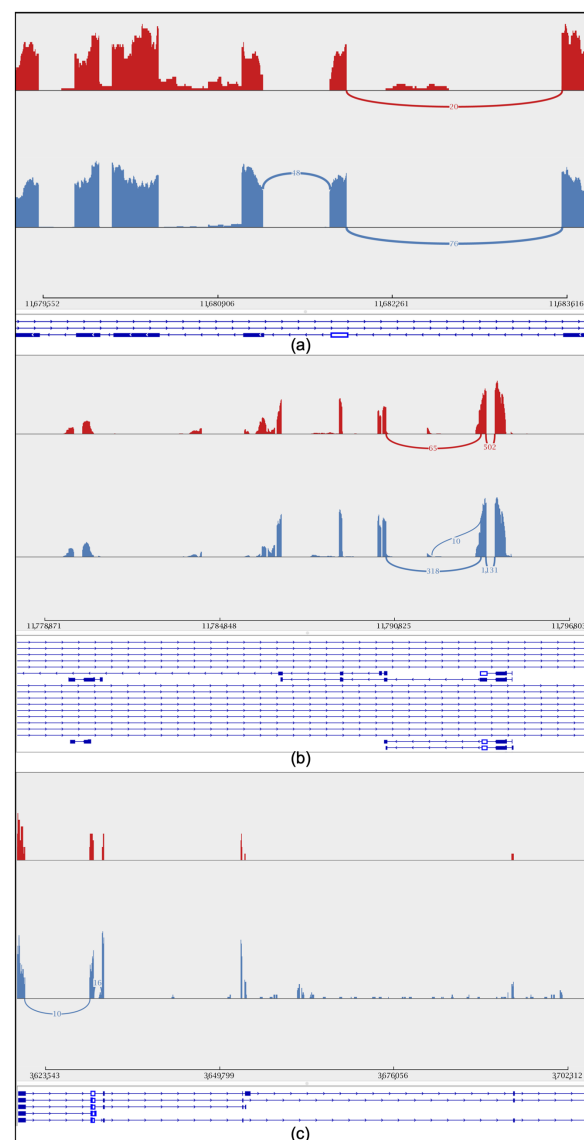

Figure 4. Sashimi plot showing the intron retention level of the sex gene. Peaks in red represent short-read coverage. For each isoform, blocks in blue represent exons, and lines between blocks represent introns. The red color represents workers, while the blue color represents drones. (**a**) Drone courtship behavior gene (*Y-y*). (**b**) *csd* gene. (**c**) *dsx* gene.
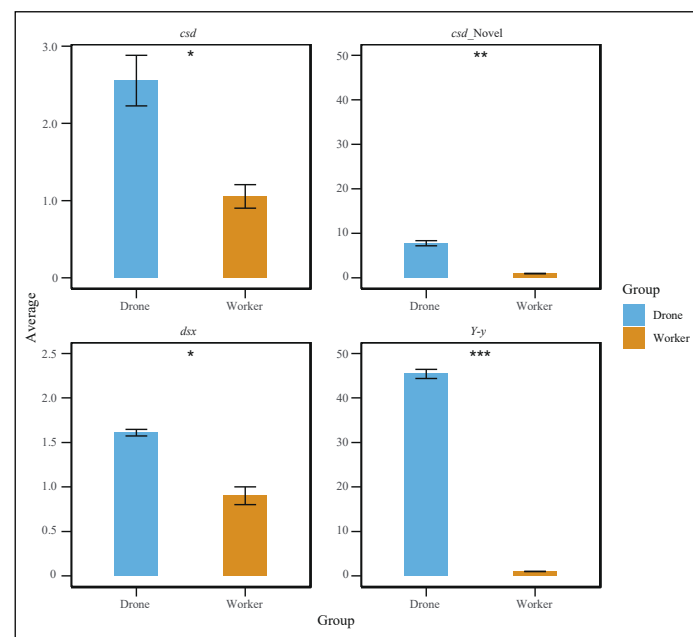
Figure 5. Quantitative polymerase chain reaction (qPCR) results showing that the specific spliced forms of the *dsx*, *csd*, and *Y−y* genes were significantly more highly expressed in drones than in workers (*t*-test, $p < 0.05$). (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

4. Discussion

Embryonic development is a crucial stage in the development of insects, with gene expression during the embryonic period playing an important role in ensuring normal development throughout their life cycle [4,37–39]. Iso-Seq technology enables the direct reading of full-length transcripts, avoiding the uncertainties associated with transcript assembly in RNA-Seq and significantly improving the accuracy of genome annotation [40]. This technology precisely identifies splicing variants and new splice isoforms and provides richer functional information [40]. Furthermore, the combination of Iso-Seq and RNA-Seq integrates the quantitative advantages of short-read data with the structural integrity of long-read data, allowing for the more accurate analysis of gene expression patterns and diversity. This mixed sequencing strategy has achieved higher-precision annotations in multiple species, including *Ptychobarbus chungtienensis* (Tsao, 1964) [41], *Ovis aries* Linnaeus, 1758 [42], *Gossypium purpurascens* Poir [43], *Procambarus clarkii* (Girard, 1852), and pigs [44].

In this study, we comprehensively analyzed gene expression differences in embryonic development between workers and drones using PacBio Iso-Seq and Illumina RNA-Seq to enhance the understanding of sex differentiation during embryonic development in *A. cerana*. Transcriptome analysis across five developmental time points (12, 24, 36, 48, and 60 h) revealed that the 24-h stage is critical for gene differential expression and alternative splicing events during the embryonic development of workers and drones. In particular, during embryonic development, novel isoforms of the *csd* gene (G10927.21) were identified. Through qPCR validation, the existence of these newly identified isoforms was confirmed (Figure 5, Supplemental Table S7). These findings provide a foundation for the functional study of the *A. cerana* genome [45].

Based on Iso-Seq, specific transcripts and novel genes peaked after 24 h of embryonic development, and then, their expression gradually decreased. This trend was consistent with changes in the number of differentially expressed genes and alternative splicing events observed with RNA-Seq. Additionally, the principal component analysis plot indicated that the RNA-Seq data had good reproducibility (Supplemental Figure S2), further supporting the reliability of the data. These findings reveal the complexity of gene activity after 24 h of

embryonic development in *A. cerana*. Iso-Seq transcriptome annotations showed that more than 7% of the variable transcripts were only present in drones or workers (Supplemental Figure S1). This suggests that, in addition to sex-biased gene expression, sex-specific isoforms may be a common mechanism for achieving sex-specific functions, which was similarly reported in three spine stickleback fish [26]. During this development stage, we identified unknown functional genes that may play a key role in sex differentiation in *A. cerana*.

The differences in the number of isoforms and the overlapping distribution of their corresponding genes across the five developmental time points further explained the role of the different isoforms in gene and functional expression. These isoforms may contribute to the formation of morphological and behavioral differences between adult drones and workers, and at the embryonic stage, genes related to these characteristics exhibited differential expression. The ten most significantly different novel genes at each developmental stage were screened, and 12 new isoforms were identified. Additionally, there were significant trends and quantitative differences in isoform expression between embryos of different sexes. For example, the two novel isoforms, *novelGene_11834* and *novelGene_7210*, exhibited significantly higher expression levels in worker embryos at 24 h compared with drones. Although the specific functions of these new genes remain unclear, their significant expression differences between worker and drone embryos suggest that they may play important roles in sex-specific development. The *novelGene_11834* gene was significantly more expressed in worker embryos at 24 h compared with drones, and *novelGene_7210* also showed significantly higher expression in workers than in drones. These findings suggest that these genes may be involved in the formation of the unique physiological and morphological characteristics of workers (Supplemental Figure S4, Table S6) [18]. A further analysis of the functional annotations and associated pathways of these novel genes may help with a better understanding of their roles in sex differentiation.

Differences in alternative splicing have been reported to significantly impact sex differentiation and development [46,47]. Moreover, numerous studies have shown that sex-specific isoforms play key roles in honeybees [18], *Drosophila* [16], *Artemia franciscana* Kellogg, 1906 [47], birds [20], and zebrafish [48]. In *Drosophila*, sex-specific differences in transcript diversity are also present during embryonic development [49]. Although the functions of these isoforms have not yet been clearly identified, their significant differences in sex-specific expression provide important clues for future functional studies. In this study, through detailed expression analysis, we identified the alternative splicing forms of key sex-determination genes, such as *csd* and *dsx*, at the 24-h developmental stage. This indicates that sex-specific alternative splicing events occur as early as the embryonic stage and may play a crucial role in initiating sex differentiation. The presence of distinct splice isoforms of key sex-determination genes further suggests that alternative splicing serves as a central regulatory mechanism driving sex differentiation in *A. cerana*. Based on these findings, further investigation into the functional roles of these sex-specific isoforms could provide new insights into the molecular mechanisms underlying sex determination and differentiation in *A. cerana*. Furthermore, this regulation may affect genes related to eye structure, potentially resulting in differences in compound eye morphology between drones and workers (Supplemental Figure S5). Understanding how these isoforms influence developmental pathways could enhance the understanding of honeybee development and provide insights into their developmental biology and genetics.

The significance of the yellow gene family in various biological processes has been well documented across multiple species. For example, mutations in the yellow gene family in fruit flies can lead to changes in pigment deposition in the wings, abdomen, and thorax, and can affect behavioral patterns through interactions with the sex determination pathway.

Similarly, when the *Yellow-y* gene is knocked out in silkworm larvae, the originally black larvae exhibit yellow or lighter body colors [50–53]. Additionally, studies have found that the yellow gene family plays a critical role in body color formation in the fall armyworm (*Spodoptera frugiperda* J.E. Smith, 1797) and the black soldier fly (*Hermetia illucens* (Linnaeus, 1758)) [53,54]. Notably, the yellow gene family affects eggshell formation in the red flour beetle (*Tribolium castaneum* (Herbst, 1797)), and these genes are essential during eggshell formation and early embryonic development [52].

In studies on Asian honeybees (*A. cerana*), the yellow gene family was associated with the differentiation of workers and queens [49]. These genes are expressed in the brains, venom glands, and other tissues of larvae, pupae, and adult honeybees [49], indicating that they play important roles in the development and behavior of honeybees. The *Yellow-y* gene in honeybees is believed to be closely related to sex differentiation, which is consistent with research findings in other species. In this study, we confirmed the expression of different isoforms of the yellow gene family in 24-h-old embryos of both sexes (Figure 5, Supplemental Table S7). This suggests that the *Yellow-y* gene may be involved in sex differentiation, wing development, and pigmentation differences during embryonic development.

However, the study has certain limitations, such as the inability to fully explore the functional mechanisms of the *Yellow-y* gene in sex differentiation and development due to the lack of experimental validation. Future research could focus on deeper molecular analyses and the application of gene knockdown or overexpression techniques to further elucidate the role of the yellow gene family in honeybee development and behavior.

5. Conclusions

In conclusion, to explore the molecular differences between the transcriptomes of worker and drone embryos in *A. cerana*, PacBio Iso-Seq and Illumina RNA-Seq were conducted, and the 24-h development time point was found to be the most significant stage of gene differential expression and alternative splicing events during the development of worker and drone embryos.

This study provides new data for genome annotation of *A. cerana* and reveals the important role of sex-specific transcripts in the sex differentiation of *A. cerana* at the embryo stage. These findings lay the foundation for future studies of honeybee transcriptomics and gene regulation and provide insights into the molecular mechanisms of honeybee sex differentiation.

Author Contributions: Conceptualization, R.S. and Y.C.; methodology, R.S.; software, R.S.; validation, R.S. and Y.C.; formal analysis, R.S.; investigation, R.S.; resources, J.H.; data curation, R.S.; writing—original draft preparation, R.S.; writing—review and editing, Y.C., G.D., K.D., M.F. and J.H.; visualization, R.S.; supervision, J.H.; project administration, J.H.; funding acquisition, J.H.; sample collection, R.Z. and R.S. All authors have read and agreed to the published version of the manuscript.

References

1. Winston, M. *The Biology of the Honey Bee*; Harvard University Press: Cambridge, MA, USA, 1991.

2. Fleig, R.; Sander, K. Embryogenesis of the honeybees Apis mellifera L. (Hymenoptera: Apidae): An SEM study. *Int. J. Insect Morphol. Embryol.* **1986**, *15*, 449–462. [CrossRef]

3. Hu, X.; Ke, L.; Wang, Z.; Zeng, Z. Dynamic transcriptome landscape of Asian domestic honeybees (*Apis cerana*) embryonic development revealed by high-quality RNA sequencing. *BMC Dev. Biol.* **2018**, *18*, 11. [CrossRef] [PubMed]

4. Wang, M.; Chen, D.; Zheng, H.; Zhao, L.; Xue, X.; Yu, F.; Cheng, C.; Niu, Q.; Wang, S.; Zhang, Y.; et al. Sex-specific development in haplodiploid honeybees is controlled by the female-embryo-specific activation of thousands of intronic lncRNAs. *Front. Cell Dev. Biol.* **2021**, *9*, 690167. [CrossRef] [PubMed]

5. Koetz, A.H. Ecology, behaviour and control of *Apis cerana* with a focus on relevance to the Australian incursion. *Insects* **2013**, *4*, 558–592. [CrossRef] [PubMed]

6. Bertholf, L.M. The moults of the honeybees. *J. Econ. Entomol.* **1925**, *18*, 380–384. [CrossRef]

7. Hrassnigg, N.; Crailsheim, K. Differences in drone and worker physiology in honeybees (Apis mellifera). *Apidologie* **2005**, *36*, 255–277. [CrossRef]

8. Li, Z.; Huang, Z.Y.; Sharma, D.B.; Xue, Y.; Wang, Z.; Ren, B. Drone and worker brood microclimates are regulated differentially in honey bees, *Apis mellifera. PLoS ONE* **2016**, *11*, e0148740.

9. Osborne, P.W.; Dearden, P.K. Expression of Pax group III genes in the honeybees (Apis mellifera). *Dev. Genes Evol.* **2005**, *215*, 499–508. [CrossRef] [PubMed]

10. Dearden, P.K.; Duncan, E.J.; Wilson, M.J. Immunohistochemistry on Honeybee (Apis mellifera) Embryos. *Cold Spring Harb. Protoc.* **2009**, *2009*, pdb.prot5227. [CrossRef]

11. Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **2009**, *323*, 133–138. [CrossRef]

12. Baral, S.; Arumugam, G.; Deshmukh, R.; Kunte, K. Genetic architecture and sex-specific selection govern modular, male-biased evolution of *doublesex*. *Sci. Adv.* **2019**, *5*, eaau3753. [CrossRef]

13. Khamis, A.M.; Hamilton, A.R.; Medvedeva, Y.A.; Alam, T.; Alam, I.; Essack, M.; Umylny, B.; Jankovic, B.R.; Naeger, N.L.; Suzuki, M.; et al. Insights into the Transcriptional Architecture of Behavioral Plasticity in the Honey Bee Apis mellifera. *Sci. Rep.* **2015**, *5*, 11136. [CrossRef] [PubMed]

14. Zhang, H.; Chen, W.; Zhu, D.; Zhang, B.; Xu, Q.; Shi, C.; He, H.; Dai, X.; Li, Y.; He, W.; et al. Population-level exploration of alternative splicing and its unique role in controlling agronomic traits of rice. *Plant Cell* **2024**, *36*, 4372–4387. [CrossRef] [PubMed]

15. Calixto, C.P.; Guo, W.; James, A.B.; Tzioutziou, N.A.; Entizne, J.C.; Panter, P.E.; Knight, H.; Nimmo, H.G.; Zhang, R.; Brown, J.W. Rapid and Dynamic Alternative Splicing Impacts the Arabidopsis Cold Response Transcriptome. *Plant Cell* **2018**, *30*, 1424–1444. [CrossRef] [PubMed]

16. Burtis, K.C.; Baker, B.S. Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* **1989**, *56*, 997–1010. [CrossRef]

17. Beye, M.; Hasselmann, M.; Fondrk, M.K.; Page, R.E., Jr.; Omholt, S.W. The Gene csd Is the Primary Signal for Sexual Development in the Honeybee and Encodes an SR-Type Protein. *Cell* **2003**, *114*, 419–429. [CrossRef]

18. Cho, S.; Huang, Z.Y.; Zhang, J. Sex-specific splicing of the honeybees *doublesex* gene reveals 300 million years of evolution at the bottom of the insect sex-determination pathway. *Genetics* **2007**, *177*, 1733–1741. [CrossRef] [PubMed]

19. Gibilisco, L.; Zhou, Q.; Mahajan, S.; Bachtrog, D. Alternative Splicing within and between Drosophila Species, Sexes, Tissues, and Developmental Stages. *PLOS Genet.* **2016**, *12*, e1006464. [CrossRef] [PubMed]

20. Rogers, T.F.; Palmer, D.H.; E Wright, A. Sex-Specific Selection Drives the Evolution of Alternative Splicing in Birds. *Mol. Biol. Evol.* **2020**, *38*, 519–530. [CrossRef] [PubMed]

21. An, D.; Cao, H.X.; Li, C.; Humbeck, K.; Wang, W. Isoform sequencing and state-of-art applications for unravelling complexity of plant transcriptomes. *Genes* **2018**, *9*, 43. [CrossRef]

22. Wang, J.; Dean, D.C.; Hornicek, F.J.; Shi, H.; Duan, Z. RNA sequencing (RNA-Seq) and its application in ovarian cancer. *Gynecol. Oncol.* **2019**, *152*, 194–201. [CrossRef] [PubMed]

23. Wijeratne, S.; Gonzalez, M.E.H.; Roach, K.; Miller, K.E.; Schieffer, K.M.; Fitch, J.R.; Leonard, J.; White, P.; Kelly, B.J.; Cottrell, C.E.; et al. Full-length isoform concatenation sequencing to resolve cancer transcriptome complexity. *BMC Genom.* **2024**, *25*, 122. [CrossRef]

24. Wang, B.; Kumar, V.; Olson, A.; Ware, D. Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing. *Front. Genet.* **2019**, *10*, 384. [CrossRef] [PubMed]

25. Xia, Y.; Jin, Z.; Zhang, C.; Ouyang, L.; Dong, Y.; Li, J.; Guo, L.; Jing, B.; Shi, Y.; Miao, S.; et al. TAGET: A toolkit for analyzing full-length transcripts from long-read sequencing. *Nat. Commun.* **2023**, *14*, 5935. [CrossRef]

26. Naftaly, A.S.; Pau, S.; White, M.A. Long-read RNA sequencing reveals widespread sex-specific alternative splicing in threespine stickleback fish. *Genome Res.* **2021**, *31*, 1486–1497. [CrossRef] [PubMed]

27. Zhang, H.; Liang, Y.; Chen, S.; Xuan, Z.; Jiang, Y.; Li, R.; Cao, Y. Full-length transcriptome sequencing reveals extreme incomplete annotation of the goat genome. *Anim. Genet.* **2023**, *54*, 421–424. [CrossRef]

28. Zhao, L.; Wang, H.; Li, P.; Sun, K.; Guan, D.-L.; Xu, S.-Q. Genome Size Estimation and Full-Length Transcriptome of Sphingonotus tsinlingensis: Genetic Background of a Drought-Adapted Grasshopper. *Front. Genet.* **2021**, *12*, 678625. [CrossRef] [PubMed]

29. Kuo, R.I.; Cheng, Y.; Zhang, R.; Brown, J.W.S.; Smith, J.; Archibald, A.L.; Burt, D.W. Illuminating the dark side of the human transcriptome with long read transcript sequencing. *BMC Genom.* **2020**, *21*, 751. [CrossRef]

30. Tardaguila, M.; de la Fuente, L.; Marti, C.; Pereira, C.; Pardo-Palacios, F.J.; del Risco, H.; Ferrell, M.; Mellado, M.; Macchietto, M.; Verheggen, K.; et al. SQANTI: Extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res.* **2018**, *28*, 396–411. [CrossRef] [PubMed]

31. Pardo-Palacios, F.J.; Arzalluz-Luque, A.; Kondratova, L.; Salguero, P.; Mestre-Tomás, J.; Amorín, R.; Estevan-Morió, E.; Liu, T.; Nanni, A.; McIntyre, L.; et al. SQANTI3: Curation of long-read transcriptomes for accurate identification of known and novel isoforms. *Nat. Methods* **2024**, *21*, 793–797. [CrossRef]

32. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef] [PubMed]

33. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]

34. Liao, Y.; Smyth, G.K.; Shi, W. feature Counts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef]

35. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.D.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef]

36. Cantalapiedra, C.P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [CrossRef] [PubMed]

37. Fujinaga, D.; Gu, J.; Kawahara, H.; Ogihara, M.H.; Kojima, I.; Takeshima, M.; Kataoka, H. Twenty-hydroxyecdysone produced by dephosphorylation and ecdysteroidogenesis regulates early embryonic development in the silkmoth, Bombyx mori. *Insect Biochem. Mol. Biol.* **2020**, *127*, 103491. [CrossRef] [PubMed]

38. Yang, Q.; De Schutter, K.; Chen, P.; Van Damme, E.J.M.; Smagghe, G. RNAi of the N-glycosylation-related genes confirms their importance in insect development and α-1,6-fucosyltransferase plays a role in the ecdysis event for the hemimetabolous pest insect Nilaparvata lugens. *Insect Sci.* **2022**, *29*, 91–99. [CrossRef]

39. Karaiskos, N.; Wahle, P.; Alles, J.; Boltengagen, A.; Ayoub, S.; Kipar, C.; Kocks, C.; Rajewsky, N.; Zinzen, R.P. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **2017**, *358*, 194–199. [CrossRef] [PubMed]

40. Zhao, Y.; Chen, Z.; Hu, M.; Liu, H.; Zhao, H.; Huang, Y.; Jiang, M.; Li, S.; Li, G.; Zhu, C.; et al. Integrating Iso-seq and RNA-seq data for the reannotation of the greater amberjack genome. *Sci. Data* **2024**, *11*, 675. [CrossRef] [PubMed]

41. Gao, Z.; Chong, Y.; Lu, Y.; Ma, S.; Wang, Z.; Hong, J.; Wu, J.; Li, M.; Xi, D.; Deng, W. Transcriptomic annotation of the Chungtien schizothoracin (Ptychobarbus chungtienensis) using Iso-seq and RNA-seq data. *Sci. Data* **2024**, *11*, 629. [CrossRef] [PubMed]

42. Wu, Z.; Zhuang, L.; Yan, M.; Ren, W.; Raza, S.H.A.; Al Abdulmonem, W.; Alghsham, R.S.; Alharbi, Y.M.; Wu, S. Characterization and analysis of the muscle transcriptome in black Tibetan sheep (Ovis aries) by hybrid sequencing of PacBio Iso-seq. *Small Rumin. Res.* **2023**, *227*, 107093. [CrossRef]

43. Rehman, A.; Tian, C.; He, S.; Li, H.; Lu, S.; Du, X.; Peng, Z. Transcriptome dynamics of Gossypium purpurascens in response to abiotic stresses by Iso-seq and RNA-seq data. *Sci. Data* **2024**, *11*, 477. [CrossRef]

44. Beiki, H.; Liu, H.; Huang, J.; Manchanda, N.; Nonneman, D.; Smith, T.P.L.; Reecy, J.M.; Tuggle, C.K. Improved annotation of the domestic pig genome through integration of Iso-Seq and RNA-seq data. *BMC Genom.* **2019**, *20*, 344. [CrossRef]

45. Paolillo, G.; De Iorio, M.G.; Filipe, J.F.S.; Riva, F.; Stella, A.; Gandini, G.; Pagnacco, G.; Lazzari, B.; Minozzi, G. Analysis of complementary sex-determiner (csd) allele diversity in different honeybees subspecies from Italy based on NGS data. *Genes* **2022**, *13*, 991. [CrossRef] [PubMed]

46. Gómez-Redondo, I.; Planells, B.; Navarrete, P.; Gutiérrez-Adán, A. Role of alternative splicing in sex determination in vertebrates. *Sex. Dev.* **2021**, *15*, 381–391. [CrossRef]

47. Viet, D.N.; Christiaens, O.; De Vos, S.; Smagghe, G.; Bossier, P. The Sex-Specific Splicing of Doublesex in Brine Shrimp *Artemia franciscana*. *Genes* **2022**, *13*, 1997. [CrossRef]

48. Lin, X.; Liu, F.; Meng, K.; Liu, H.; Zhao, Y.; Chen, Y.; Hu, W.; Luo, D. Comprehensive Transcriptome Analysis Reveals Sex-Specific Alternative Splicing Events in Zebrafish Gonads. *Life* **2022**, *12*, 1441. [CrossRef]

49. Ray, M.; Conard, A.M.; Urban, J.; Mahableshwarkar, P.; Aguilera, J.; Huang, A.; Vaidyanathan, S.; Larschan, E. Sex-specific splicing occurs genome-wide during early Drosophila embryogenesis. *ELife* **2023**, *12*, e87865. [CrossRef] [PubMed]

50. Ferguson, L.C.; Green, J.; Surridge, A.; Jiggins, C.D. Evolution of the Insect Yellow Gene Family. *Mol. Biol. Evol.* **2010**, *28*, 257–272. [CrossRef] [PubMed]

51. Cao, Y.; Zhan, S.; Zhang, Y.; Tan, A.; Huang, Y. Identification of yellow gene family in Agrotis ipsilon and functional analysis of Aiyellow-y by CRISPR/Cas9. *Insect Biochem. Mol. Biol.* **2018**, *94*, 1–9. [CrossRef]

52. Noh, M.Y.; Kramer, K.J.; Muthukrishnan, S.; Arakane, Y. Ovariole-specific Yellow-g and Yellow-g2 proteins are required for fecundity and egg chorion rigidity in the red flour beetle, Tribolium castaneum. *Insect Biochem. Mol. Biol.* **2023**, *159*, 103984. [CrossRef] [PubMed]

53. Dong, Y.; Xu, X.; Qian, L.; Kou, Z.; Andongma, A.A.; Zhou, L.; Huang, Y.; Wang, Y. Genome-wide dentification of yellow gene family in Hermetia illucens and functional analysis of yellow-y by CRISPR/Cas9. *Insect Sci.* **2024**. [CrossRef]

54. Han, W.; Tang, F.; Zhong, Y.; Zhang, J.; Liu, Z. Identification of yellow gene family and functional analysis of Spodoptera frugiperda yellow-y by CRISPR/Cas9. *Pestic. Biochem. Physiol.* **2021**, *178*, 104937. [CrossRef]