



Article

A Global Survey of the Full-Length Transcriptome of *Apis mellifera* by Single-Molecule Long-Read Sequencing

Shuang-Yan Zheng^{1,2,†}, Lu-Xia Pan^{3,4,†}, Fu-Ping Cheng^{3,4}, Meng-Jie Jin^{3,4} and Zi-Long Wang^{3,4,*}

¹ College of Animal Science and Technology, Jiangxi Agricultural University, Nanchang 330045, China

² Sino-German Joint Research Institute, Nanchang University, Nanchang 330047, China

³ Honeybee Research Institute, Jiangxi Agricultural University, Nanchang 330045, China

⁴ Jiangxi Province Key Laboratory of Honeybee Biology and Beekeeping, Jiangxi Agricultural University, Nanchang 330045, China

* Correspondence: zilongwangnc@jxau.edu.cn

† These authors contributed equally to this work.

Abstract: As important pollinators, honey bees play a crucial role in both maintaining the ecological balance and providing products for humans. Although several versions of the western honey bee genome have already been published, its transcriptome information still needs to be refined. In this study, PacBio single-molecule sequencing technology was used to sequence the full-length transcriptome of mixed samples from many developmental time points and tissues of *A. mellifera* queens, workers and drones. A total of 116,535 transcripts corresponding to 30,045 genes were obtained. Of these, 92,477 transcripts were annotated. Compared to the annotated genes and transcripts on the reference genome, 18,915 gene loci and 96,176 transcripts were newly identified. From these transcripts, 136,554 alternative splicing (AS) events, 23,376 alternative polyadenylation (APA) sites and 21,813 lncRNAs were detected. In addition, based on the full-length transcripts, we identified many differentially expressed transcripts (DETs) between queen, worker and drone. Our results provide a complete set of reference transcripts for *A. mellifera* that dramatically expand our understanding of the complexity and diversity of the honey bee transcriptome.

Keywords: PacBio sequencing; alternative splicing; single-molecule sequencing; alternative polyadenylation site; differentially expressed transcript



Citation: Zheng, S.-Y.; Pan, L.-X.; Cheng, F.-P.; Jin, M.-J.; Wang, Z.-L. A Global Survey of the Full-Length Transcriptome of *Apis mellifera* by Single-Molecule Long-Read Sequencing. *Int. J. Mol. Sci.* **2023**, *24*, 5827. <https://doi.org/10.3390/ijms24065827>

Academic Editor: Sarath Chandra Janga

Received: 19 January 2023

Revised: 8 March 2023

Accepted: 12 March 2023

Published: 18 March 2023



Copyright: © 2023 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) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Honey bees are important pollinating insects, playing a crucial role in promoting biodiversity and maintaining ecological balance. They are characterized by high reproductive performance, short generation cycles, easiness of artificial breeding and division of labor. *A. mellifera* is the world's most widely bred honey bee species with the excellent ability to collect large nectar sources and gums [1]. The first version of the *A. mellifera* complete genome sequence was published in 2006 [2]; since then, several versions of genome sequences have been released on Genbank. So far, the genomic sequences [2,3], transcriptome [4,5] and genetic linkage map [6–8] of *A. mellifera* have been studied extensively.

Transcriptomes provide valuable gene transcription information, including alternative splicing (AS) events, long non-coding RNAs (lncRNAs) and alternative polyadenylation (APA) sites. This is especially useful for species lacking genome sequence information. Alternative splicing of pre-mRNAs is an effective means of regulating gene expression and increasing proteome diversity in higher eukaryotes [9,10]. Alternative splicing is coordinately controlled by multiple transcriptional and post-transcriptional regulatory mechanisms, and is involved in many biological processes including growth, development, signal transduction and response to stress [11]. The abundance of alternative splicing events reaches >95–100% in human genes [12], 63% in mouse genes [13] and 20% to 60% in plant intron genes [14–17]. Long non-coding RNAs are abundant in organisms and are important

regulators of gene expression [18]. Alternative polyadenylation events are a ubiquitous form of gene expression regulation in eukaryotic cells and they play an important role in many biological processes, including development, proliferation, differentiation and neuronal activation [19].

High-throughput transcriptome sequencing technology has been widely used to obtain transcript sequences and to analyze gene expression differences [20,21]. However, the second-generation sequencing technology requires the assembly of short reads (100–150 bp) that lead to low accuracy in the subsequent gene structure analysis, such as alternative splicing and gene fusion, because of the incomplete and inaccurate assembly process. In addition, the sequences of complex regions, such as highly repetitive and high GC content sequences, cannot be assembled. A third-generation sequencing platform can obtain high-quality full-length transcripts directly without PCR amplification and assembly during the sequencing and data analysis. Based on these full-length transcripts, we can precisely identify AS events, APA sites and fusion genes from the transcriptome of an organism.

At present, there are two representative third-generation sequencing (TGS) technologies: the single-molecule real-time (SMRT) sequencing technology from Pacific Biosciences (PacBio) [22] and Oxford Nanopore Technology (ONT) [23]. The SMRT sequencing method uses zero-mode waveguide (ZMW) technology that enables direct observation of the synthesis of DNA strands by a single DNA polymerase molecule. On the other hand, Oxford nanopore sequencing can directly identify a single nucleotide when a single DNA or RNA molecule passes through the nanopore (nanometer size) by electrophoresis. Both technologies are similar in terms of reading length, error rate and throughput and both can produce long reads that exceed most transcripts. They are thus widely used for genome assembly [24,25] and the identification of full-length transcripts [26,27].

The full landscape of the western honey bee transcriptome is not well understood. In this study, we attempt to build a catalog of all the full-length transcripts of *A. mellifera* and explore the complexity of the transcriptome using TGS technologies. To this end, we sequenced the full-length transcriptome of three mixed samples of queen and drone and worker of *A. mellifera* using the SMRT sequencing technique. The results of this study will improve honey bee genome sequence analysis and provide a complete set of reference transcriptomes for gene function study in *A. mellifera*.

2. Results

2.1. Sequencing Results and Data Assembly

Sequencing using the PacBio sequencing platform resulted in 1,100,885, 1,042,133 and 1,126,347 polymerase reads for the queen, worker and drone samples, respectively (Table 1, Figure S1). From these polymerase reads, 43,491,193, 40,055,924 and 40,663,357 subreads as well as 795,183, 735,061 and 792,515 circular consensus sequences (CCSs) were obtained. The final numbers of full-length non-chimeric (FLNC) reads were 556,221, 514,591 and 546,907, with an average length of 1504, 1511 and 1651 bp (Figure S2). The full-length transcripts accounted for 83.10%, 80.04% and 83.58% of the total clean reads in the three samples, respectively. After eliminating redundant FLNC sequences from the same transcript, 116,535 non-redundant FLNC sequences from 30,045 genes were obtained for subsequent analysis (Table S1).

Table 1. Summary of the PacBio sequencing data.

Sample	Total Bases (Gbp)	Number of Polymerase Reads	Number of Subreads	Number of CCSs	Number of FLNCs	Average Length of FLNCs (bp)
Queen	64.80	1,100,885	43,491,193	795,183	556,221	1504
Worker	59.74	1,042,133	40,055,924	735,061	514,591	1511
Drone	65.49	1,126,347	40,663,357	792,515	546,907	1651

2.2. Functional Annotation of Transcripts

The 116,535 transcripts were compared with the 12,328 known annotated genes in the *A. mellifera* reference genome (Genbank accession No.: GCA_003254395.2). A total of 11,130 gene loci and 20,359 transcripts overlapped with the genome annotation; 18,915 gene loci and 96,176 transcripts were newly identified in this study. These new transcripts accounted for 82.53% of the total transcripts (Figure 1A,B). Of the new transcripts, 76,018 were new transcripts from known genes and 20,158 were from new genes. After blast comparison of the 116,535 transcripts with the NR, GO, KO, KOG and Swiss-Prot databases, 92,477 transcripts were annotated (Tables S2 and S3). For the 18,915 new genes, 2967 genes were annotated.

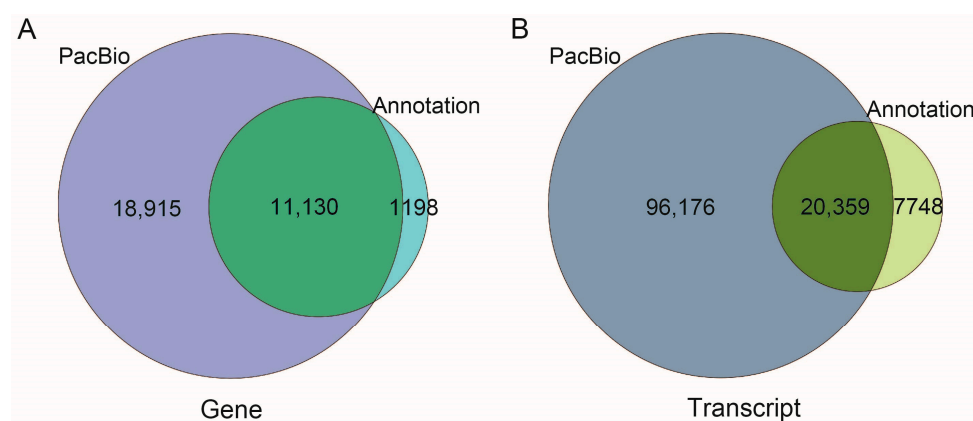


Figure 1. Comparison of transcripts (A) and genes (B) generated by the PacBio sequencing and *A. mellifera* genome annotation.

2.3. Unique Transcripts from Queen, Worker and Drone Data

There were 19,256, 18,644 and 20,186 transcripts unique to the queen, worker and drone datasets with an average length of 1828 bp, 1875 bp and 1830 bp, respectively (Figure 2A, Table S4). GO analysis of these unique transcripts showed that queen-unique transcripts were primarily enriched in 249 GO terms ($p < 0.05$, Table S5), worker-unique transcripts in 376 GO terms ($p < 0.05$), and drone-unique transcripts in 337 GO terms ($p < 0.05$). The largest GO terms in queen, worker and drone were “catalytic activity”, “binding” and “ion binding”, respectively. KEGG analysis showed that queen-unique transcripts were significantly enriched in three pathways ($q < 0.05$, Figure 2B), drone-unique transcripts in 18 pathways ($q < 0.05$, Figure 2C), and worker-unique transcripts in 16 pathways ($q < 0.05$, Figure 2D). Of these, many pathways were related to carbohydrate metabolism. Four pathways related to honey bee caste differentiation were significantly enriched in the worker, including the “Longevity regulating pathway–worm”, “Longevity regulating pathway–multiple species”, “MAPK signaling pathway–fly” and the “MAPK signaling pathway”.

2.4. Alternative Splicing

Gene transcript analysis revealed that 20,481 (68.17%) genes contained one isoform, 1663 (5.54%) contained two isoforms, and 3173 (10.56%) contained >10 isoforms (Figure 3A). The *A. mellifera* *Mob3* gene was found to contain six isoforms, though it was only annotated as one in the reference genome (Figure 3B).

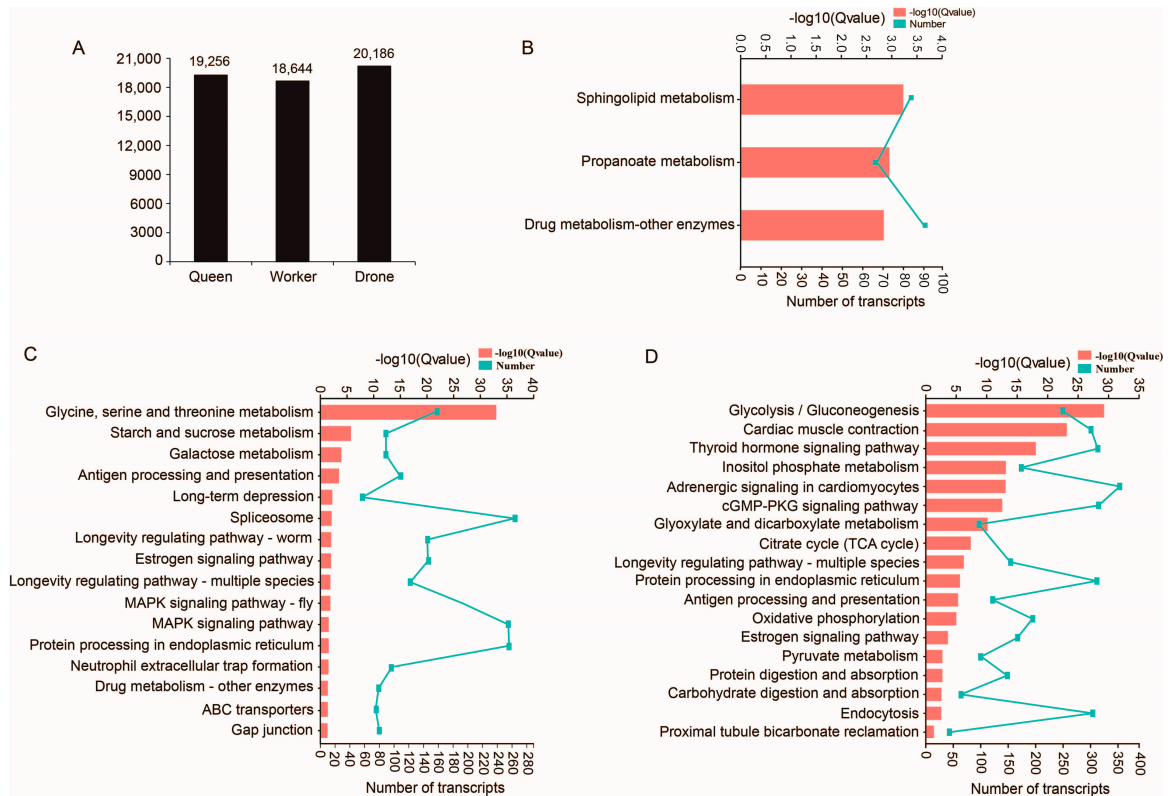


Figure 2. Unique transcripts isolated from queen, worker and drone datasets. (A) Numbers of unique transcripts in queen, worker and drone datasets. (B–D) show the significantly enriched KEGG pathways for queen-, worker- and drone-unique transcripts, respectively.

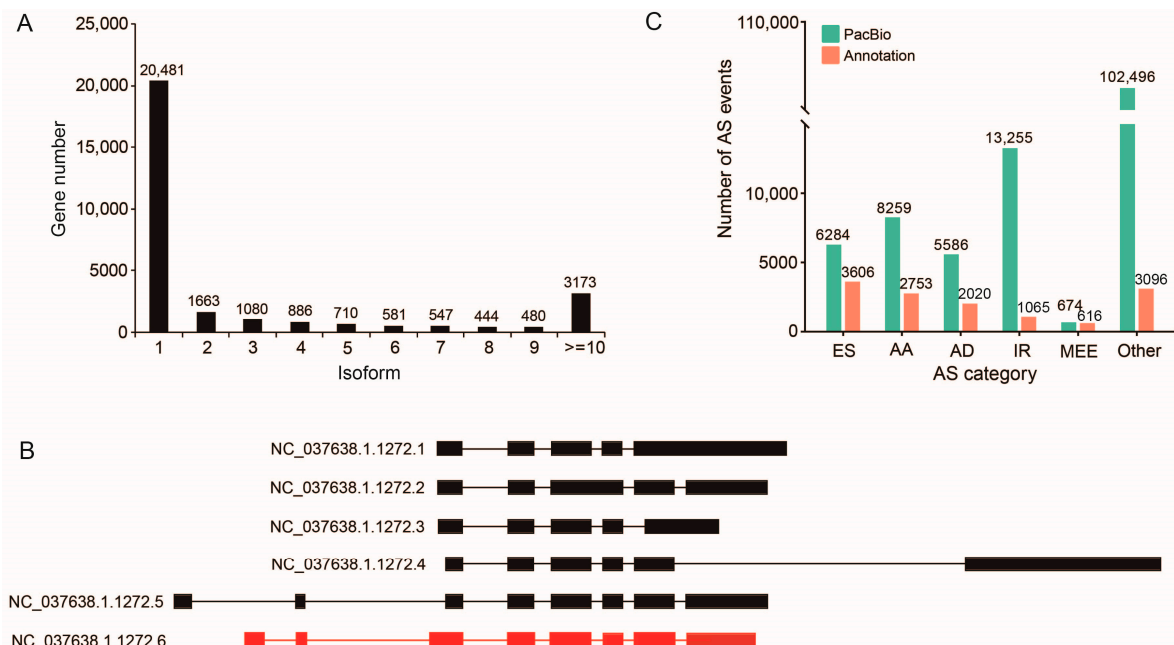


Figure 3. The alternative splicing events identified from PacBio honey bee transcriptome. (A) Distribution of splice isoforms of genes. (B) The exon/intron structure of the six isoforms of the *Mob3* gene. The filled boxes represent exons and lines represent introns. The isoform in red was obtained by genome annotation. (C) The comparison of each alternative splicing type between this PacBio data and the NCBI reference transcript set.

The alternative splicing events in the three mixed samples were analyzed, and 25,289, 62,103 and 42,352 alternative splicing events were detected in the queen, worker and drone datasets, respectively (Figure S3). By combining the data of the three RNA samples, a total of 136,554 AS events originating in 6435 genes were detected. Of the five major AS types, intron retention (IR) was the predominant type, originating in 4074 genes (Figure 3C). In contrast, only 13,156 AS events were detected in the NCBI reference transcript set of *A. mellifera* (Figure 3C); this is less than the AS events obtained in this study by an order of magnitude.

Five genes were randomly selected to validate the accuracy of AS events using RT-PCR. The results indicated that the size of each amplified fragment was consistent with what was expected (Figure 4).

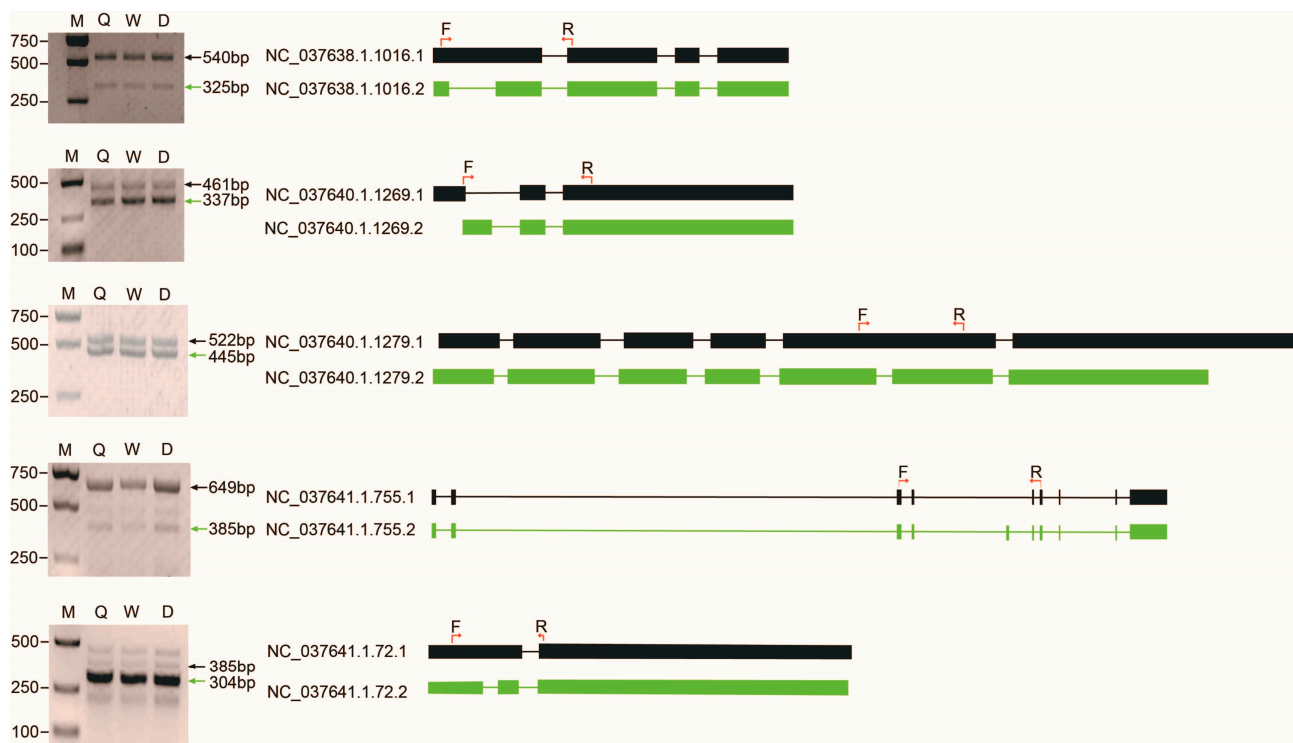


Figure 4. Verification of the alternative splicing events in five genes by RT-PCR and Sanger sequencing. The exon/intron structure of each isoform of each gene is shown in the right panel. The filled boxes represent exons and the lines represent introns. Different isoforms of each gene are displayed in green or black. The locations of the PCR primers are shown with red arrows on the first isoform of each gene. M: marker, Q: queen, W: worker, D: drone.

2.5. APA Sites

In the queen, worker and drone datasets, 14,491, 14,021 and 15,213 APA loci from 5364, 5262 and 5650 gene loci were identified, respectively (Figure S4). Based on the 116,535 PacBio transcripts, a total of 23,376 APA events at 5714 gene loci were identified, of which 5321 (93.12%) genes had 1–9 poly (A) sites and 393 (6.88%) genes had more than 10 poly (A) sites, with an average of 4.09 poly (A) sites per gene (Figure 5A). The nucleic acid composition 50 bp upstream and downstream of the poly (A) sites were enriched with uracil (U) and adenine (A) (Figure 5B), and a conserved element (AAAAAAAAAA) was over-presented in this region (Figure 5C).

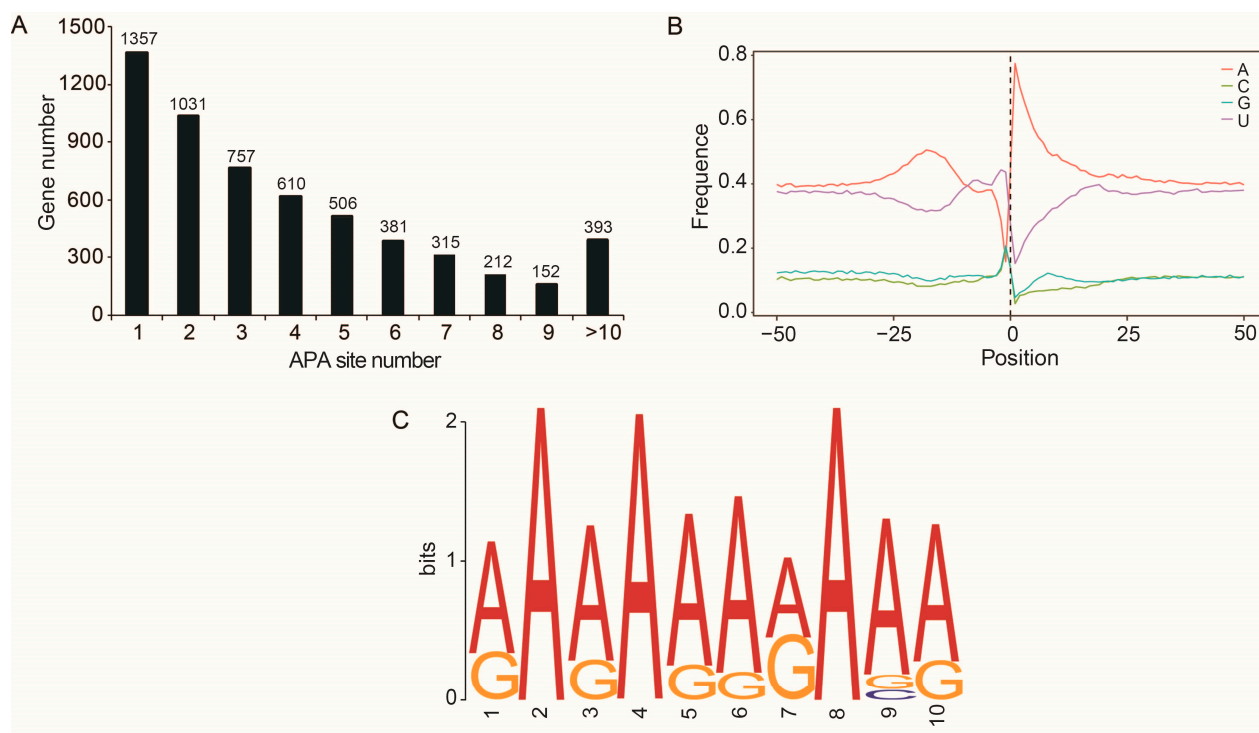


Figure 5. Alternative polyadenylation sites identified from the PacBio honey bee transcriptome. (A) Distribution of the number of APA sites per gene. (B) Nucleotide composition 50 bp upstream and downstream of the poly (A) sites. (C) A conserved element near the poly (A) sites predicted by MEME analysis.

2.6. Long Non-Coding RNAs

A total of 21,813 lncRNAs were identified from the 116,535 transcripts and, of these, 1168 lncRNAs overlapped with known lncRNAs from genome annotation (Figure 6A). A total of 20,645 lncRNAs were novel lncRNAs predicted by CPC/CNCI/CPAT/Pfam analysis (Figure 5B). The average length of all the lncRNAs was significantly shorter than that of the mRNAs (Figure 6C). All the lncRNAs were classified into four types according to their positional relationship in the *A. mellifera* genome. Finally, 9758 (44.73%), 6262 (28.71%), 3248 (14.89%) and 2545 (11.67%) lncRNAs were classified as located in the intronic region, sense, antisense or intergenic region, respectively, where intronic region lncRNA was the dominant type (Figure 6D).

2.7. DETs and DEGs between the Three Honey Bee Castes

Based on full-length transcripts obtained from this study, we used the RNA-seq data from 2-day and 4-day larvae of the three honey bee castes published previously [28] to explore whether expression differences between the castes are more evident at the transcript level.

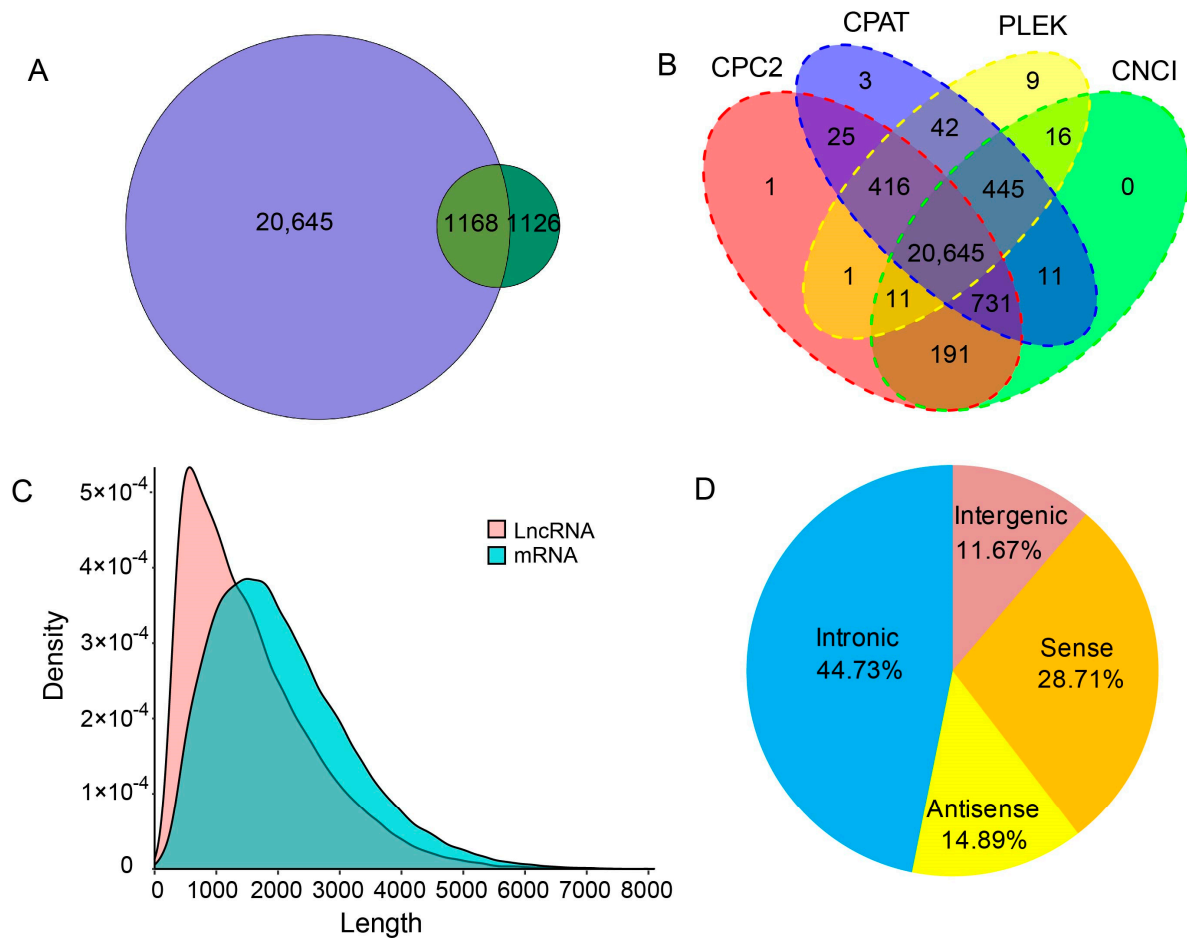


Figure 6. LncRNAs identified from the PacBio honey bee transcriptome. **(A)** Comparison of the lncRNAs identified in this study with those from genome annotation. **(B)** The Venn diagram of the number of lncRNAs predicted by CPC2, CPAT, PLEK and CNCI. **(C)** The length density distribution of lncRNAs and mRNAs. **(D)** Proportions of the four kinds of lncRNAs.

When comparing queen larvae vs. worker larvae, worker larvae vs. drone larvae and queen larvae vs. drone larvae, at the 2-day larval stage, there were 1006, 3836 and 5109 DETs related to 778, 2707 and 3341 genes as well as 341, 715 and 1461 DEGs (Figure 7A,B; Table S6). At the 4-day larval stage, there were 4160, 4130 and 4381 DETs related to 2660, 2448 and 2562 genes as well as 1644, 1643 and 1905 DEGs (Figure 7A,B; Table S6). Many of the DETs were related to genes involved in caste differentiation or sex determination of honey bees, such as *Igf1*, *Fem* and *Amdsx* (Table S6). The numbers of overlapping genes between the DETGs and DEGs occupied 17.99–56.56% and 68.11–76.06% of the DETGs and DEGs, respectively (Figure 7C).

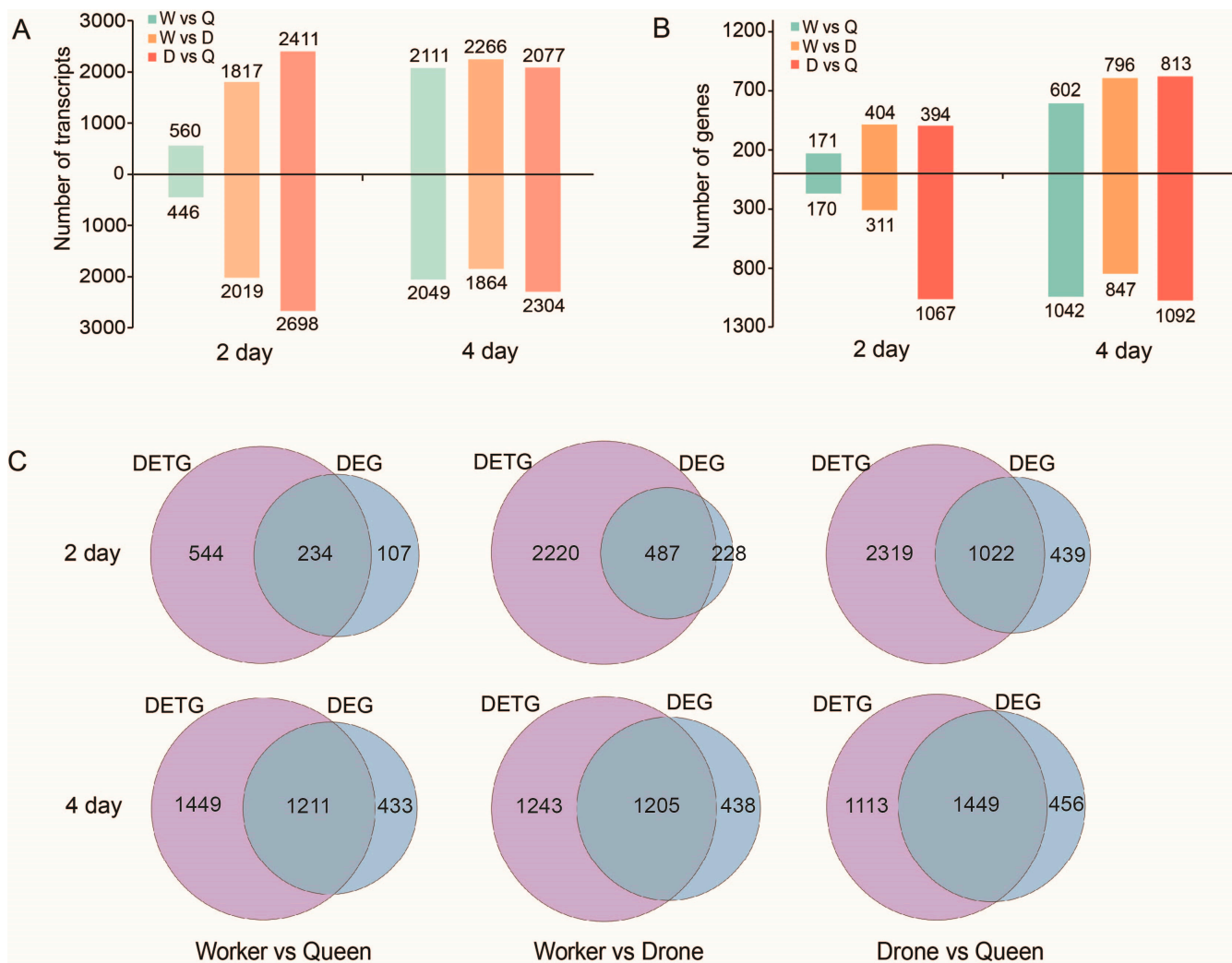


Figure 7. The DETs and DEGs between castes at larva stage. The DETs (A) and DEGs (B) among queen, worker and drone at 2-day and 4-day larva stages. (C) The overlapped genes between DETGs and DEGs.

3. Discussion

Although the genome sequence of the western honey bee *A. mellifera*, a very important Hymenoptera model organism, has been published [2], its transcriptome features and mRNA structure have not been analyzed in depth. With the improvement of sequencing technologies, SMRT provides the possibility to profile the *A. mellifera* transcriptome. Compared with second-generation sequencing technologies, the TGS technologies, represented by PacBio’s SMRT sequencing, can directly sequence the cDNA to obtain full-length transcripts. In this study, 116,535 transcripts were obtained by PacBio sequencing of the honey bee transcriptome; this is 4.15 times the number of transcripts recorded in NCBI annotations. Also, many novel or previously unidentified protein-coding genes were identified. These transcript datasets represent the largest and most complete set of full-length transcripts of the western honey bee, containing a broad diversity of transcript isoforms that are well-suited for studying alternative splicing, gene fusion, etc.

Alternative splicing is an important transcriptional process in eukaryotes and one of the methods that produce different transcriptional isoforms that can significantly expand the diversity of gene function [9,10]. In this study, we detected 136,554 AS events from 6435 genes, occupying 21.42% of genes identified in this study; this is similar to the ratio reported in *Drosophila* [29] and higher than that in *Bombyx mori* [30] and *Plutella*

xylostella [31]. These results suggest that AS is a widely used mRNA post-transcriptional process mechanism in honey bees.

Selection of different polyadenylation sites in pre-mRNAs will result in isoforms with different lengths of coding or 3' untranslated regions, further increasing transcriptome diversity and thus finely regulating mRNA stability, localization and translation efficiency. About half of the genes in *Drosophila* [32], worms [33] and zebrafish [34] have more than two APA sites. Recent studies using high-throughput sequencing have shown that APA regulates gene expression through multiple mechanisms in plants and animals [35–37]. In the present study, we identified a large number of APA loci in honey bees; moreover, genes with more than two APA loci account for 76.25% of the total number of APA loci-related genes in the honey bee transcriptome. Our findings suggest that APA is one of the important processes for promoting transcriptome diversity of honey bees.

In this study, we identified a large number of lncRNAs, of which 94.65% were newly identified lncRNAs. In mammals, lncRNAs can account for up to 4–9% of total RNA, whereas protein-coding RNAs account for only 1% of total RNA [38]. This suggests that the prevalence of lncRNAs from both coding and noncoding genes is higher in honey bees than previously thought. Interestingly, lncRNAs from intron regions were the largest type, while in other insects lncRNAs from intergenic regions are the most abundant type [39–42]. This suggests that introns in honey bee genes may play an important role in gene transcription regulation through lncRNAs. Recent studies have shown that lncRNAs are numerous and are involved in many important biological processes, including X chromosome silencing [43,44], chromosome structure and genome rearrangements [45–47], imprinting [48,49], transcriptional activation and suppression [50,51]. The large number of lncRNAs identified in this study may also be extensively involved in various biological processes in honey bees.

In addition, based on the transcripts obtained in this study, we identified a large number of DETs among the three honey bee castes using RNA-seq data. The number of DET-related genes is 2.7 times the number of DEGs reported in a previous study [28]. Moreover, the overlapping genes between the DETGs and DEGs just occupy a middle proportion of the DETGs. It suggests that many genes showed differential expression between the honey bee castes at some splicing isoforms and not at the whole gene level. These findings suggest that phenotypic differences between the three honey bee castes are more likely to be regulated at the transcript isoform level, rather than at the gene level.

In conclusion, this study greatly expanded the data set of *A. mellifera* gene transcripts, providing a basis for discovering new genes and transcripts and for conducting gene function studies in *A. mellifera*.

4. Materials and Methods

4.1. Sample Collection

The experimental colonies of *A. mellifera* were raised by the Honeybee Research Institute, Jiangxi Agricultural University, Nanchang, Jiangxi province, China (28.46 °N, 115.49 °E). For the worker samples, a fertilized queen from a healthy colony was controlled on an empty comb to lay eggs for 6 h. Then, ten developmental time points and eight organs/tissues of newly emerged workers were sampled. The same queen and colony were used for collecting queen samples. Similarly, the queen was allowed to lay eggs on an empty worker comb for 6 h; after 72 h, 1-day-old larvae in the worker cells were moved to queen cells to cultivate queens. Ten developmental time points and eight organs/tissues of newly emerged queens were sampled. For the drone samples, an unfertilized queen was allowed to lay eggs on an empty drone cell comb for 6 h. Then, ten developmental time points and eight organs/tissues of newly emerged drones were sampled. The developmental time points and organs/tissues sampled are listed in Table S7. All the samples were frozen and stored in liquid nitrogen for further use.

4.2. Library Construction and Sequencing

The total RNA of each sample was extracted separately using the TRIzol™ Reagent (Invitrogen, Carlsbad, CA, USA). The quality of the RNA samples was measured using a Nanodrop 2000 (Thermo Fisher Scientific, Wilmington, DE, USA) and the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). The RNA samples from the same caste were equally mixed into one sample for library construction. The full-length cDNAs were synthesized using the SMARTer™ PCR cDNA Synthesis Kit (Clontech, Palo Alto, CA, USA) using oligodT as primers and were amplified by PCR. The PCR products were purified using PB magnetic beads to remove cDNA fragments less than 1 kb in length. Subsequently, the full-length cDNAs were end-repaired and connected with SMRT adapters, then purified with PB magnetic beads again to obtain the sequencing libraries. Finally, the libraries were sequenced using PacBioSeque II (Pacific Biosciences of California, Inc., Menlo Park, CA, USA).

4.3. Analysis of the Raw Data

The raw sequencing data were preprocessed using the SMRTlink v10.1 (Pacific Biosciences of California, Inc., Menlo Park, CA, USA) software, and the full-length transcripts were obtained by Iso-Seq analysis process. The single molecule sequencing polymerase reads were split into subreads, where subreads with lengths less than 50 bp or greater than 15,000 bp were filtered; the remaining subreads were considered clean data. The subreads obtained from the same polymerase reads formed circular consensus sequences (CCS) after self-correction of errors. The CCS sequences were classified by detecting concatemer (chimeric sequence) and 5' and 3' sequencing primers, to find full-length non-concatemer (FLNC) sequences. The FLNC sequences were used for subsequent analysis.

4.4. Identification of Genes and Transcripts

The corrected FLNC sequences were aligned to the *A. mellifera* reference genome version. Amel_HAv3.1 (Available online: ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/003/254/395/GCA_003254395.2_Amel_HAv3.1/ (accessed on 21 June 2022).) using GMAP v2021-05-27 software (Department of Bioinformatics Genentech, Inc., South San Francisco, CA, USA) [52] to obtain their precise positions on the genome. According to the alignment position of each FLNC sequence, the genes (loci) and transcripts (isoforms) were identified. If two FLNC sequences mapped in the same direction and had more than 20% overlap and at least one exon had more than 20% overlap, they were considered to be transcripts from the same gene. Redundant and low-reliability transcripts from the same gene were removed, and the remaining sequences were considered the final non-redundant transcript datasets.

The obtained gene loci were compared with the annotated loci of the *A. mellifera* reference genome version Amel_HAv3.1 through the GMAP v2021-05-27 software (Department of Bioinformatics Genentech, Inc., South San Francisco, CA, USA), and the sequenced genes were referred to as new genes if they met any of the following criteria: (1) no overlap or less than 20% overlap with the annotated genes; (2) more than 20% overlap with the annotated genes, but with the opposite orientation. The transcripts obtained by PacBio sequencing were compared with the annotated genes of *A. mellifera* reference genome, and the transcript was considered as a new one if one or more new splicing sites existed in the transcript, or if the transcript and the annotated gene were not both single exons. Functional annotation of new genes and transcripts was performed by searching against the NR, GO, KEGG, COG/KOG and Swiss-Prot databases using Diamond v2.0.7 software (University of Tübingen, Tübingen, Germany) [53].

4.5. Identification of Alternative Splicing Events, lncRNAs and APA Sites

Alternative splicing events, including exon skipping (ES), intron retention (IR), alternative donor site (AD), alternative acceptor site (AA) and mutually exclusive exon (MEE), were classified and counted by Astalavista v3.2 software (Johns Hopkins University School of Medicine, Baltimore, MD, USA) [54].

To identify lncRNAs, the transcript sequences of new genes and known genes were compared with NR, KOG, KO and Swiss-Prot databases to filter out possible coding sequences. The coding potential of the remaining sequences was further evaluated with the CPAT v1.2.4 (Mayo Clinic College of Medicine, Rochester, MN, USA) [55], CNCI v2.0 (Chinese Academy of Sciences, Beijing, China) [56], CPC2 vbeta (Peking University, Beijing, China) [57] and PLEK v1.2 (Xidian University, Xi'an, Shanxi, China) [58] softwares. All sequences with coding potential were then filtered and the remaining sequences were considered the final noncoding RNA datasets. The APA sites of each gene were detected using Tapis v1 (Colorado State University, Fort Collins, Colorado, USA) [59] software with default parameters and conserved elements near the poly (A) sites were predicted using MEME v5.3.3 software (University of Queensland, Brisbane, Queensland, Australia) [60].

4.6. Identification of DETs and DEGs

The RNA-seq data of *A. mellifera* queens, workers and drones from the 2-day and 4-day larvae stages were downloaded from NCBI; all transcripts obtained from this study, as well as partial transcripts from NCBI, were used as reference sequences to screen for differentially expressed transcripts. The clean reads obtained from RNA-seq were mapped to the non-redundant full-length transcripts using Bowtie2 v2.3.5 (Johns Hopkins University, Baltimore, Maryland, USA) [61] and the expression level of each transcript in each sample was calculated. DESeq2 v1.22.2 (Harvard School of Public Health, Boston, MA, USA) [62] was used for differential expression significance analysis and the $|\log_2(\text{fold change})| > 1$, FDR < 0.05 was used as the screening criterion. To identify DEGs, reads from all transcripts of the same gene were used to count the expression level of each gene and the screening criterion of DEGs was the same as DETs.

4.7. RT-PCR

Transcripts from five randomly selected genes were chosen for experimental validation using RT-PCR. Total RNAs were extracted from newly emerged queen, worker and drone and were reverse transcribed to cDNA. Primers were designed to amplify all the detected transcripts at the same time based on full-length sequences using *Primer Premier 5.0* (Table S8). PCR conditions were as follows: pre-denaturation at 95 °C for 5 min; 30 amplification cycles of denaturation at 95 °C for 30 s, 58.5 °C for 45 s, 72 °C for 45 s and final elongation at 72 °C for 10 min. The PCR amplification products were detected on 1% agarose gel and followed by Sanger sequencing.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24065827/s1>.

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

Funding: This work was supported by the National Natural Science Foundation (No. 32160134, 31402147) and the Natural Science Foundation of Jiangxi province (No. 2018ACB21028), China.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The subreads obtained by PacBio sequencing has been submitted to the Sequence Read Archive (SRA) database and are available from NCBI under the BioProject number PRJNA941379.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cheng, S.L. Honey bee germplasm resources. In *The Apicultural Science in China*, 1st ed.; Liu, B.H., Ed.; Chinese Agricultural Press: Beijing, China, 2001; pp. 16–24.
2. Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **2006**, *443*, 931–949. [[CrossRef](#)] [[PubMed](#)]
3. Wallberg, A.; Bunikis, I.; Pettersson, O.V.; Mosbech, M.-B.; Childers, A.K.; Evans, J.D.; Mikheyev, A.S.; Robertson, H.M.; Robinson, G.E.; Webster, M.T. A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC Genom.* **2019**, *20*, 275. [[CrossRef](#)] [[PubMed](#)]
4. Yokoi, K.; Wakamiya, T.; Bono, H. Meta-Analysis of the Public RNA-Seq Data of the Western Honeybee *Apis mellifera* to Construct Reference Transcriptome Data. *Insects* **2022**, *13*, 931. [[CrossRef](#)] [[PubMed](#)]
5. Bresnahan, S.T.; Döke, M.A.; Giray, T.; Grozinger, C.M. Tissue-specific transcriptional patterns underlie seasonal phenotypes in honey bees (*Apis mellifera*). *Mol. Ecol.* **2022**, *31*, 174–184. [[CrossRef](#)]
6. Hunt, G.J.; Page, R.E., Jr. Linkage map of the honey bee, *Apis mellifera*, based on RAPD markers. *Genetics* **1995**, *139*, 1371–1382. [[CrossRef](#)] [[PubMed](#)]
7. Solignac, M.; Vautrin, D.; Baudry, E.; Mougél, F.; Loiseau, A.; Cornuet, J.-M. A Microsatellite-Based Linkage Map of the Honeybee, *Apis mellifera* L. *Genetics* **2004**, *167*, 253–262. [[CrossRef](#)]
8. Solignac, M.; Mougél, F.; Vautrin, D.; Monnerot, M.; Cornuet, J.-M. A third-generation microsatellite-based linkage map of the honey bee, *Apis mellifera*, and its comparison with the sequence-based physical map. *Genome Biol.* **2007**, *8*, R66. [[CrossRef](#)]
9. Keren, H.; Lev-Maor, G.; Ast, G. Alternative splicing and evolution: Diversification, exon definition and function. *Nat. Rev. Genet.* **2010**, *11*, 345–355. [[CrossRef](#)]
10. Nilsen, T.W.; Graveley, B.R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **2010**, *463*, 457–463. [[CrossRef](#)]
11. Staiger, D.; Brown, J.W. Alternative Splicing at the Intersection of Biological Timing, Development, and Stress Responses. *Plant Cell* **2013**, *25*, 3640–3656. [[CrossRef](#)]
12. Pan, Q.; Shai, O.; Lee, L.J.; Frey, B.J.; Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **2008**, *40*, 1413–1415. [[CrossRef](#)] [[PubMed](#)]
13. Merkin, J.; Russell, C.; Chen, P.; Burge, C.B. Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. *Science* **2012**, *338*, 1593–1599. [[CrossRef](#)] [[PubMed](#)]
14. Wang, B.-B.; Brendel, V. Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 7175–7180. [[CrossRef](#)] [[PubMed](#)]
15. Barbazuk, W.B.; Fu, Y.; McGinnis, K.M. Genome-wide analyses of alternative splicing in plants: Opportunities and challenges. *Genome Res.* **2008**, *18*, 1381–1392. [[CrossRef](#)]
16. Filichkin, S.A.; Priest, H.D.; Givan, S.A.; Shen, R.; Bryant, D.W.; Fox, S.E.; Wong, W.-K.; Mockler, T.C. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* **2010**, *20*, 45–58. [[CrossRef](#)]
17. Marquez, Y.; Brown, J.W.; Simpson, C.; Barta, A.; Kalyna, M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* **2012**, *22*, 1184–1195. [[CrossRef](#)]
18. Gil, N.; Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat. Rev. Genet.* **2020**, *21*, 102–117. [[CrossRef](#)]
19. Elkon, R.; Ugalde, A.P.; Agami, R. Alternative cleavage and polyadenylation: Extent, regulation and function. *Nat. Rev. Genet.* **2013**, *14*, 496–506. [[CrossRef](#)]
20. Reuter, J.A.; Spacek, D.V.; Snyder, M.P. High-Throughput Sequencing Technologies. *Mol. Cell* **2015**, *58*, 586–597. [[CrossRef](#)]
21. Hrdlickova, R.; Toloue, M.; Tian, B. RNA -Seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA* **2017**, *8*, e1364. [[CrossRef](#)] [[PubMed](#)]
22. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [[CrossRef](#)] [[PubMed](#)]
23. Ambardar, S.; Gupta, R.; Trakroo, D.; Lal, R.; Vakhlu, J. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian J. Microbiol.* **2016**, *56*, 394–404. [[CrossRef](#)]
24. Zimin, A.V.; Stevens, K.A.; Crepeau, M.; Puiu, D.; Wegrzyn, J.; Yorke, J.A.; Langley, C.H.; Neale, D.B.; Salzberg, S.L. An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. *Gigascience* **2017**, *6*, 1–4. [[CrossRef](#)] [[PubMed](#)]
25. Wang, Y.-X.; Chen, H.-F.; Yin, Z.-Y.; Chen, W.-L.; Lu, L.-T. The genetic adaptations of *Toxoptera aurantii* facilitated its rapid multiple plant hosts dispersal and invasion. *Genomics* **2022**, *114*, 110472. [[CrossRef](#)] [[PubMed](#)]
26. Zhang, H.; Xu, H.; Liu, H.; Pan, X.; Xu, M.; Zhang, G.; He, M. PacBio single molecule long-read sequencing provides insight into the complexity and diversity of the *Pinctada fucata martensii* transcriptome. *BMC Genom.* **2020**, *21*, 481. [[CrossRef](#)] [[PubMed](#)]
27. Byrne, A.; Beaudin, A.E.; Olsen, H.E.; Jain, M.; Cole, C.; Palmer, T.; DuBois, R.M.; Forsberg, E.C.; Akeson, M.; Vollmers, C. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat. Commun.* **2017**, *8*, 16027. [[CrossRef](#)]
28. He, X.; Jiang, W.; Zhou, M.; Barron, A.B.; Zeng, Z. A comparison of honeybee (*Apis mellifera*) queen, worker and drone larvae by RNA-Seq. *Insect Sci.* **2019**, *26*, 499–509. [[CrossRef](#)]
29. 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](#)]

30. Shao, W.; Zhao, Q.-Y.; Wang, X.-Y.; Xu, X.-Y.; Tang, Q.; Li, M.; Li, X.; Xu, Y.-Z. Alternative splicing and *trans*-splicing events revealed by analysis of the *Bombyx mori* transcriptome. *Rna* **2012**, *18*, 1395–1407. [[CrossRef](#)]
31. Zhao, Q.; Zhong, W.; He, W.; Li, Y.; Li, Y.; Li, T.; Vasseur, L.; You, M. Genome-wide profiling of the alternative splicing provides insights into development in *Plutella xylostella*. *BMC Genom.* **2019**, *20*, 463. [[CrossRef](#)]
32. Smibert, P.; Miura, P.; Westholm, J.O.; Shenker, S.; May, G.; Duff, M.O.; Zhang, D.; Eads, B.D.; Carlson, J.; Brown, J.B.; et al. Global Patterns of Tissue-Specific Alternative Polyadenylation in *Drosophila*. *Cell Rep.* **2012**, *1*, 277–289. [[CrossRef](#)] [[PubMed](#)]
33. Jan, C.H.; Friedman, R.C.; Ruby, J.G.; Bartel, D.P. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature* **2011**, *469*, 97–101. [[CrossRef](#)] [[PubMed](#)]
34. Ulitsky, I.; Shkumatava, A.; Jan, C.H.; Subtelny, A.O.; Koppstein, D.; Bell, G.W.; Sive, H.; Bartel, D.P. Extensive alternative polyadenylation during zebrafish development. *Genome Res.* **2012**, *22*, 2054–2066. [[CrossRef](#)] [[PubMed](#)]
35. Shen, Y.; Venu, R.; Nobuta, K.; Wu, X.; Notibala, V.; Demirci, C.; Meyers, B.C.; Wang, G.-L.; Ji, G.; Li, Q.Q. Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing. *Genome Res.* **2011**, *21*, 1478–1486. [[CrossRef](#)] [[PubMed](#)]
36. Wu, X.; Liu, M.; Downie, B.; Liang, C.; Ji, G.; Li, Q.Q.; Hunt, A.G. Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 12533–12538. [[CrossRef](#)]
37. Pereira-Castro, I.; Moreira, A. On the function and relevance of alternative 3'-UTRs in gene expression regulation. *Wiley Interdiscip. Rev. RNA* **2021**, *12*, e1653. [[CrossRef](#)]
38. Ponting, C.P.; Oliver, P.L.; Reik, W. Evolution and Functions of Long Noncoding RNAs. *Cell* **2009**, *136*, 629–641. [[CrossRef](#)]
39. Wu, Y.; Cheng, T.; Liu, C.; Liu, D.; Zhang, Q.; Long, R.; Zhao, P.; Xia, Q. Systematic Identification and Characterization of Long Non-Coding RNAs in the Silkworm, *Bombyx mori*. *PLoS ONE* **2016**, *11*, e0147147. [[CrossRef](#)]
40. Li, W.-J.; Song, Y.-J.; Han, H.-L.; Xu, H.-Q.; Wei, D.; Smagghe, G.; Wang, J.-J. Genome-wide analysis of long non-coding RNAs in adult tissues of the melon fly, *Zeugodacus cucurbitae* (Coquillett). *BMC Genom.* **2020**, *21*, 600. [[CrossRef](#)]
41. Azlan, A.; Obeidat, S.M.; Das, K.T.; Yunus, M.A.; Azzam, G. Genome-wide identification of *Aedes albopictus* long noncoding RNAs and their association with dengue and Zika virus infection. *PLoS Negl. Trop. Dis.* **2021**, *15*, e0008351. [[CrossRef](#)]
42. Meng, L.; Yuan, G.; Chen, M.; Dou, W.; Jing, T.; Zheng, L.; Peng, M.; Bai, W.; Wang, J. Genome-wide identification of long non-coding RNAs (lncRNAs) associated with malathion resistance in *Bactrocera dorsalis*. *Pest Manag. Sci.* **2021**, *77*, 2292–2301. [[CrossRef](#)] [[PubMed](#)]
43. Leeb, M.; Steffen, P.A.; Wutz, A. X chromosome inactivation sparked by non-coding RNAs. *RNA Biol.* **2009**, *6*, 94–99. [[CrossRef](#)]
44. Hierholzer, A.; Chureau, C.; Liverziani, A.; Ruiz, N.B.; Cattanaach, B.M.; Young, A.N.; Kumar, M.; Cerase, A.; Avner, P. A long noncoding RNA influences the choice of the X chromosome to be inactivated. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2118182119. [[CrossRef](#)] [[PubMed](#)]
45. Rinn, J.L.; Kertesz, M.; Wang, J.K.; Squazzo, S.L.; Xu, X.; Bruggmann, S.A.; Goodnough, L.H.; Helms, J.A.; Farnham, P.J.; Segal, E.; et al. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* **2007**, *129*, 1311–1323. [[CrossRef](#)]
46. Engreitz, J.M.; Pandya-Jones, A.; McDonel, P.; Shishkin, A.; Sirokman, K.; Surka, C.; Kadri, S.; Xing, J.; Goren, A.; Lander, E.S.; et al. The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science* **2013**, *341*, 1237973. [[CrossRef](#)]
47. Loewer, S.; Cabili, M.N.; Guttman, M.; Loh, Y.-H.; Thomas, K.; Park, I.H.; Garber, M.; Curran, M.; Onder, T.; Agarwal, S.; et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* **2010**, *42*, 1113–1117. [[CrossRef](#)] [[PubMed](#)]
48. Sleutels, F.; Zwart, R.; Barlow, D.P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **2002**, *415*, 810–813. [[CrossRef](#)]
49. Zhang, H.; Zeitz, M.J.; Wang, H.; Niu, B.; Ge, S.; Li, W.; Cui, J.; Wang, G.; Qian, G.; Higgins, M.J.; et al. Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the *Kcnq1* locus. *J. Cell Biol.* **2014**, *204*, 61–75. [[CrossRef](#)]
50. Attaway, M.; Chwat-Edelstein, T.; Vuong, B.Q. Regulatory Non-Coding RNAs Modulate Transcriptional Activation During B Cell Development. *Front. Genet.* **2021**, *12*, 678084. [[CrossRef](#)]
51. Li, X.L.; Subramanian, M.; Jones, M.F.; Chaudhary, R.; Singh, D.K.; Zong, X.; Gryder, B.; Sindri, S.; Mo, M.; Schetter, A.; et al. Long Noncoding RNA PURPL Suppresses Basal p53 Levels and Promotes Tumorigenicity in Colorectal Cancer. *Cell Rep.* **2017**, *20*, 2408–2423. [[CrossRef](#)]
52. Wu, T.D.; Watanabe, C.K. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **2005**, *21*, 1859–1875. [[CrossRef](#)] [[PubMed](#)]
53. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [[CrossRef](#)] [[PubMed](#)]
54. Sammeth, M.; Foissac, S.; Guigó, R. A general definition and nomenclature for alternative splicing events. *PLoS Comput. Biol.* **2008**, *4*, e1000147. [[CrossRef](#)] [[PubMed](#)]
55. Wang, L.; Park, H.J.; Dasari, S.; Wang, S.; Kocher, J.-P.; Li, W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **2013**, *41*, e74. [[CrossRef](#)]
56. Sun, L.; Luo, H.; Bu, D.; Zhao, G.; Yu, K.; Zhang, C.; Liu, Y.; Chen, R.; Zhao, Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **2013**, *41*, e166. [[CrossRef](#)]

57. Kong, L.; Zhang, Y.; Ye, Z.-Q.; Liu, X.-Q.; Zhao, S.-Q.; Wei, L.; Gao, G. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **2007**, *35*, W345–W349. [[CrossRef](#)]
58. Li, A.; Zhang, J.; Zhou, Z. PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinform.* **2014**, *15*, 311. [[CrossRef](#)]
59. Abdel-Ghany, S.E.; Hamilton, M.; Jacobi, J.L.; Ngam, P.; Devitt, N.; Schilkey, F.; Ben-Hur, A.; Reddy, A.S.N. A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* **2016**, *7*, 11706. [[CrossRef](#)]
60. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37* (Suppl. S2), w202–w208. [[CrossRef](#)]
61. Langmead, B. Aligning Short Sequencing Reads with Bowtie. *Curr. Protoc. Bioinform.* **2010**, *32*, 11.7.1–11.7.14. [[CrossRef](#)]
62. 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](#)]

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.