



Article Dynamic Evolution of Repetitive Elements and Chromatin States in Apis mellifera Subspecies

Nick Panyushev ^{1,2,†}, Max Selitskiy ^{1,†}, Vasilina Melnichenko ³, Egor Lebedev ¹, Larisa Okorokova ² and Leonid Adonin ^{1,4,*}

- ¹ Institute of Environmental and Agricultural Biology (X-BIO), Tyumen State University, 625003 Tyumen, Russia; panyushev@nextmail.ru (N.P.); m.v.selickij@utmn.ru (M.S.)
- ² Bioinformatics Institute, 197342 St. Petersburg, Russia; larisa.okorokova@etu.univ-cotedazur.fr
- ³ International Scientific and Research Institute of Bioengineering, ITMO University, 197101 St. Petersburg, Russia; vdmelnichenko@edu.hse.ru
- ⁴ Institute of Biomedical Chemistry, Group of Mechanisms for Nanosystems Targeted Delivery, 119121 Moscow, Russia
- * Correspondence: l.s.adonin@utmn.ru
- ⁺ These authors contributed equally to this work.

Abstract: In this study, we elucidate the contribution of repetitive DNA sequences to the establishment of social structures in honeybees (*Apis mellifera*). Despite recent advancements in understanding the molecular mechanisms underlying the formation of honeybee castes, primarily associated with Notch signaling, the comprehensive identification of specific genomic cis-regulatory sequences remains elusive. Our objective is to characterize the repetitive landscape within the genomes of two honeybee subspecies, namely *A. m. mellifera* and *A. m. ligustica*. An observed recent burst of repeats in *A. m. mellifera* highlights a notable distinction between the two subspecies. After that, we transitioned to identifying differentially expressed DNA elements that may function as cis-regulatory elements. Nevertheless, the expression of these sequences showed minimal disparity in the transcriptome during caste differentiation, a pivotal process in honeybee eusocial organization. Despite this, chromatin segmentation, facilitated by ATAC-seq, ChIP-seq, and RNA-seq data, revealed a distinct chromatin state associated with repeats. Lastly, an analysis of sequence divergence among elements indicates successive changes in repeat states, correlating with their respective time of origin. Collectively, these findings propose a potential role of repeats in acquiring novel regulatory functions.

Keywords: Hymenoptera; honeybee *Apis mellifera*; repetitive DNA; transposable elements (transposons; TE); noncoding RNA (ncRNA); chromatin landscape

1. Introduction

Honeybees play a vital role in preserving biodiversity and ensuring the food security of our planet [1]. They serve as essential pollinators for numerous crop species, facilitating their reproduction and yield [2]. Beyond their agricultural significance, honeybees also contribute to the maintenance of ecosystems and the balance of natural communities [3,4].

A honey bee colony consists of a queen, workers (females), and drones (males). The drones and the queen perform a reproductive function. The tasks of workers, associated with the upkeep of the colony, partly vary with their age [5]. Nurse workers build and clean the comb, store incoming pollen, protect the hive, and feed the newly hatched bees in their first weeks of life. Forager workers gather pollen, nectar, and water from their environment to provide for the colony. The caste development of social insects represents a significant evolutionary transition, playing a central role in their ecological success. Eusociality, which is defined as the division of labor, a shared care of offspring, and the presence of a sterile worker caste, has evolved repeatedly, especially among Hymenoptera. Each eusocial lineage is unique, with different social traits and its own evolutionary history [6,7]. A



Citation: Panyushev, N.; Selitskiy, M.; Melnichenko, V.; Lebedev, E.; Okorokova, L.; Adonin, L. Dynamic Evolution of Repetitive Elements and Chromatin States in *Apis mellifera* Subspecies. *Genes* **2024**, *15*, 89. https://doi.org/10.3390/ genes15010089

Academic Editor: David Lando

Received: 15 December 2023 Revised: 7 January 2024 Accepted: 10 January 2024 Published: 11 January 2024



Copyright: © 2024 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/). significant body of research data has accumulated on species with derived eusocial behavior, such as honeybees [8–14], but the genomic mechanisms regulating social behavior remain poorly understood.

Repetitive elements (REs), including transposable elements (TEs), are one of the main components of the eukaryotic genome. In insects, genome size and TE content can vary greatly [15]. Even among closely related *Drosophila* species, the TE content varies from 40% (in *Drosophila ananassae*) to 10% (in *Drosophila miranda* and *Drosophila simulans*) [16]. The highest TE content was found in the large genome of the migratory locust (*Locusta migratoria*, Orthoptera), exceeding 60% [17]. Previously seen as useless components of genomes, transposons, or "jumping genes", are now recognized to have a significant influence on the evolution of the host genome's structure [18,19]. Transposons have the ability to move and insert themselves into genes or regulatory sequences, disrupting coding sequences and gene regulation, resulting in chromosomal rearrangements [20]. Despite the potential negative impact on gene regulation, evidence suggests that TEs can also drive genomic innovations that offer advantages to the host [21–25].

REs are major constituents of long noncoding RNAs (lncRNAs, over 200 nucleotides) [26]. IncRNAs are involved in the epigenetic regulation of gene expression, acting as a regulatory link between chromatin-modifying complexes and the DNA sequence [27–29], maintaining loop interactions between promoters and enhancers [30], thereby acting as organizers of the three-dimensional architecture of the genome [31,32]. LncRNAs have been described to be involved in biological processes associated with honey bee behavior, such as dance behavior [33], sensory perceptions of odor, and olfactory receptor activity [34]. The rate of evolution of noncoding sequences has also been shown to correlate with major social transitions in bees [35].

Here, we relied on genomic and transcriptomic data available in public databases to assess the transcriptional activity of repetitive elements in the bee brain. We also analyzed the "historical" origin of repetitive elements and their domestication processes to assess the influence of repetitive elements on the regulation of eusocial behavior in the honey bee.

2. Materials and Methods

2.1. RE Identification

The RepeatModeler program (v2.0.3) was utilized to identify de novo repeat elements in the reference honey bee genome for *A. m. mellifera*, GCF_003254395.2_Amel_HAv3.1 (Amel_HAv3.1), and for *A. m. ligustica* (GCA_019321825.1). Following that, the Repeat-Classifier script and the Dfam3.6 database were employed to classify the consensus sequences. Subsequently, the obtained library of classified repetitive element (RE) families was used to mask the genomes of *A. m. mellifera* and *A. m. ligustica* using RepeatMasker with the -lib and -a options. To generate tab files with the percentage of divergence and Kimura plots, the resulting alignment output file was parsed through the parseRM script, https://github.com/4ureliek/Parsing-RepeatMasker-Outputs (accessed on 1 May 2023) (Figures S1–S3).

2.2. RNA-seq Analysis

Raw bulk RNA-seq reads ranging from SRR6727829 to SRR6727860 [36] were obtained from the SRA database and assessed for quality using FastQC (version 0.12.0). A pseudoalignment of the reads was performed using kallisto [37], utilizing an index generated from the reference transcriptome and consensus sequences of REs derived from RepeatModeler. A subsequent analysis was carried out in Rstudio using the DESeq2 (version 1.42.0) [38] R package.

For single-cell RNA-seq analysis, the raw reads were downloaded from the SRA database (SRP338028) [39] using the fasterq-dump tool from the SRA-toolkit, pseudo-aligned using the Kallisto bustools (kb counts), and processed using the Seurat R package [40].

2.3. Genome Segmentation with ChromHMM

Datasets were downloaded from SRA [41]. To perform an initial quality control, we utilized FastQC and MultiQC (version 1.15). The reads from Chip-seq and ATAC-seq experiments were aligned to the HAv3.1 reference genome using STAR(version 2.7.11a) [42]. Afterwards, we employed ChromHMM [43] to binarize the aligned reads from Chip-seq (H3K27ac, H3K4me3, H3K4me1, and H3K27me3 histone modifications), ATAC-seq, and RNA-seq, with a 50-nucleotide genomic resolution.

Following the alignment and binarization steps, the combined data were utilized to train the model using the ChromHMM (version 1.25) LearnModel command. The input for training included 64 states (26 binary combinations), a bin size of 50, a file specifying the size of chromosomes, a file containing anchors and coordinates, and the model error change threshold.

Once the model was trained, the ChromHMM StatePruning command was executed to prune the model, resulting in 61 model files ranging from 2 to 63 states. The models were then compared to the original model using the ChromHMM CompareModels command. This comparison generated a graph that depicted the median and average loss values for each model in relation to the original model. Through an analysis of the obtained information, a model containing 23 states was chosen, as it covered 95% of the median losses observed in comparison to the base model.

Using the Python script with the pandas, numpy, and matplotlib libraries, a plot of emission and state transition probabilities was generated for the model, consisting of 23 states. Segmentation files for the three honey bee castes (worker, queen, and drone) were created using the ChromHMM MakeSegmentation command. From these segmentation files, a graph of the state transition probabilities, excluding transitions to itself, was constructed. Furthermore, based on this model, additional graphs were created. These included graphs depicting the starting site of transcription and the final site of transcription for each caste, as well as a graph displaying additional information for each state.

Based on the RNA-seq data, the coordinates of expressed and repressed exons and genes were determined. Additionally, the coordinates of known and unknown repeat classes were identified. The state in which these objects were located was determined using the ChromHMM OverlapEnrichment command (https://github.com/vasimel/Repeats_ in_honeybee_genome/tree/apis_mellifera_model (accessed on 10 November 2023)).

To find information about genes intersecting with state 13 and the REs closest to them, the samtools intersect and samtools closest programs were used, respectively. To display graphs, a python script and the ShinyGO 0.77 program (http://bioinformatics.sdstate.edu/go/ (accessed on 1 December 2023)) were used.

3. Results

3.1. Expansion of Previously Uncharacterized Repeat Families Discriminate the A. m. mellifera and A. m. ligustica Subspecies

The Kimura plots presented in Figure 1A illustrate the evolutionary landscape of repetitive sequences in the genomes of the western honey bee, *A. m. mellifera* (*A.m.m.*), and one of its subspecies, *A. mellifera ligustica* (*A.m.l.*).

The TE distribution pattern on the *A.m.m.* Kimura plot is characterized by two peaks located at 3–4% and 15–22% of substitutions from the transposable element (TE) consensus. The first peak is characterized by a high proportion of DNA transposon sequences, the highest in the whole diagram. The peak was preceded by a systematic increase in their copies in the previous stages. It is worth noting that, at this stage, which is relatively recent in evolutionary history, unclassified sequences contribute the least to TE diversity.

The second peak at 15–22% includes a great diversity of TEs, but the greatest contribution to this diversity is made by uncharacterized repeats. At this stage, there is an increase in DNA transposon diversity and the highest percentage of LTR elements, as well as the highest RC-Helitron copy number. The last two classes have almost no copies that have appeared in recent evolutionary history (1–2%).



Figure 1. Pattern of repetitive elements in the *A. m. mellifera* and *A. mellifera ligustica* genomes. (A) Interspersed repeat landscapes. TE classes are marked with colors: DNA transposons, green; LTR TE, red; rolling-circle TEs (helitrons), yellow; unknown repeats, blue and purple; (B) the classification of repetitive sequences in the genomes; and (C) the distribution of rnd-4_family-321 elements on chromosome maps.

The Kimura plot of *A.m.l.* has three peaks. The first peak at 4% includes copies of DNA transposons and unclassified repeats. The latter occupy less than half of the diversity at this stage. The second peak at 8% is characterized by a burst of diversity of uncharacterized sequences. The third peak at 15–16% is a distinctive feature of this genome compared to the *A.m.m.* genome, as it has the greatest amount of RE copies, but unfortunately, they cannot

be characterized at the present stage of research. It can be seen that one single element has expanded the most at this peak—it is rnd-4_family-321. As in the *A.m.m.* genome, there are two peaks in the increase in DNA transposon diversity, corresponding to substitution levels of 3–4% and 15%. The absence of new copies of RC and LTR is shown, and their maximum diversity is noted at the level of 17–18%.

Repetitive elements (repeatome) account for 10.83% and 13.15% of the *A.m.m.* and *A.m.l.* genome sequences, as shown in Figure 1B. Simple repeats, low complexity, and unclassified repeats occupy the highest proportion of the genome. TEs represent 2.37% of the *A.m.m.* genome sequence, while for *A.m.l.*, TEs account for 2.8%. Retroelements are poorly represented in the compared genomes, and their percentage does not exceed 0.2% of *A.m.m.* and 0.17% of the *A.m.l.* genome sequence. Representatives of class II are the most abundant among all TEs. The percentage of the Tc1-IS630-Pogo superfamily in the *A.m.l.* genome is 1.25%, which significantly exceeds this index in the other subspecies (0.95%—*A.m.m.*). At the same time, the percentage of another group of DNA transposons, helitrons, is comparable in the studied genomes, and equals 1.13% and 1.08% in *A.m.m.* and *A.m.l.*, respectively.

Figure 1C shows the expansion of rnd-4_family-321 in the centromere and telomere proximal regions of the chromosomes of both subspecies. This specific repetitive element is scientifically intriguing due to its considerable expansion in the genome of *A.m.l.* It is noteworthy that it comprises several copies of AvaI and AluI repeats, making up a longer and more intricate structure.

In both genomes, rnd-4_family-321 elements are mostly located in regions proximal to telomeres and centromeres. These elements grew in size mostly in the telomeres of *A.m.l.* chromosomes (LG1, LG4, LG5, LG7, LG9, LG11, LG13, LG15).

3.2. Integrating Multi-Omics Data via Markov Models for Genome-Wide Chromatin State Assignment

In our study, we employed Chip-seq data for four distinct histone marks (H3K27ac, H3K4me3, H3K4me1, and H3K27me3) along with RNA-seq and ATAC-seq data. The purpose of using these datasets was to train a ChromHMM model and subsequently refine it through a process called model pruning. By applying this approach, we successfully identified a total of 23 distinct chromatin states (Table S1). The features enriched in each chromatin state are similar for all castes (Figure 2A for queens and drones and Figure S5 for workers).

The distinct enrichment of H3K4me3 and H3K27ac, along with depletion of H3K4me1 and H3K27me3, is clearly evident in E1 and E3 (E before the state number is our internal designation) chromatin states, allowing them to be grouped together as TSS-associated states [41]. E1 and E3 are likely to represent two alternative states of TSS—inactive and active ones, respectively.

E2 and E22 also show enrichment in H3K27ac and H3K4me1, as well as a slight enrichment in the H3K4me3 label, which may be characteristic of intron regions [44]. According to the state transition graph (Figure 2B), E22 can transit to the E2 state along with losing RNA-seq features. Taking this into account, we can suggest an active exon feature for E22 and E2 as the intronic regions of those genes. E20 has a very similar pattern of features to E22, but shows no transitions to neither E22 nor E2, likely to be highly expressed housekeeping genes. The E19 state is enriched in all inherent histone marks, but has a low level of ATAC-seq. E19 shows a high association with expressed exons. E19 can only transit to the E6 or E20 state, making it also closely associated with gene expression. Potentially, the E19 state represents poised promoters, containing marks characteristic of both active and inactive chromatin. Both E21 and E14 are marked by high levels of H3K27ac and H3K4me1, while being depleted in H3K4me3. The H3K4me1 mark correlates with enhancers [45], while H3K27ac indicates their active state [46]. E21 also shows a high enrichment in RNA-seq, and is also associated with exons. On the contrary, the E14 state shows no association with expression and exon regions, but it is associated with chromatin accessibility. Thus, in

our opinion, E21 and E14 could be seen as enhancers. E9, in addition, shows a high level of H3K27me3, which characterizes this condition as a poised enhancer [45].



Figure 2. Features of chromatin states and genome occupancy in *A. m. mellifera* brains. (**A**) Features of the different chromatin states in queens/drones. Columns from left to right—occupied genome fraction (purple), features comprising states (in blue), and relative enrichment of respective genomic regions (in green); (**B**) the state transition graph; transition probabilities < 0.05 are not shown. The edge thickness indicates the probability of transition.

The E4 state and, to a lesser extent, the E5 state are characterized by the almost complete absence of histone marks, but by the enrichment of RNA-seq. Due to the transition graph, the E4 and E5 states can only transit to inactive states (E8, E15). Thus, they can be treated as states with residual background expression.

The E6 and E7 states are characterized by H3K27ac enrichment and low levels of H3K4me3, H3K4me1, and H3K27me3. E6 shows a strong association with transcribed genes. E7 shows no association with genes, but the E14 state can transit to E7 and revert back, hinting that it can be a potential cis-regulatory element in the switched-off state.

E8 shows a strong depletion in all features used and is present in the highest genome fraction (56.3%), being labeled as an extensive silent domain [44]. The E15 state of chromatin shows a similar label distribution, except for a slight enrichment in ATAC-seq.

The E10 state has a moderate enrichment in H3K4me1, with weak H3K4me3 and H3K27me3. Based on the transition graph, it can change only to E8, along with losing all marks. Based on this evidence, we can characterize it as the transition state. On the contrary, E11 has the strongest enrichment in H3K27me3, suggesting its inactivation, along with strong enrichments in H3K4me1 and H3K4me3. It also shows a strong affinity to repressed exon features. This indicates that these are regulatory elements of repressed genes.

The E10 and E11 states show a similar enrichment of the H3K4me3, H3K4me1, and H3K27me3 labels, and a low level for the rest of the labels. They do not show any association with exons/genes. They are likely to be transitory states of sorts.

The E12 state is distinguished by a high enrichment of the label with H3K27me3, which is characteristic of regions of Polycomb-mediated repression [44].

E16 has a strong association with TSS regions, and the strongest association with CpG islands. Also, this state is most tightly associated with repressed genes. Taken together, this state demarcates the repressed gene transcription start sites.

The region referred to as E17 exhibits a moderate enrichment in the histone mark H3K4me3, which is associated with active gene transcription. Meanwhile, it displays a depletion or decrease in other types of histone modifications.

Finally, E18 is characterized by a low level of H3K4me1, but together with a high level of H3K27ac, H3K4me3, and H3K27me3. Together with low but not absent ATAC-seq and RNA-seq signals, this can suggest this state to be the switched-off promoter. This assumption is strengthened by the transition graph, because it can transit to the inactive TSS-associated states (E1 and E16). Therefore, together with E23, E18 states represent the smallest fraction of the genome in workers (0.09% and 0.08%, respectively).

E23 is quite poorly represented, comprising no more than 0.09% of the total genome in the worker caste, representing the smallest fraction. This is likely to be the transition state for TSS regions. If they lose their RNA-seq signal, they will become repressed, similar to E16, or, alternatively, transit to the active condition, together with H3K27ac (like E3).

3.3. REs Are Significantly Enriched in Special Chromatin State (E13)

For E13 (Figure 3), state enrichment in all histone marks is inherent, but it has no transcription activity. This state is, however, the unique one, specifically enriched with repeats. This specific state can only transit to E15 and E8, believed to be non-active states.



Figure 3. Localization features of E13 chromatin state regions and GO analysis of their closest genes. (A) Venn diagrams indicating intersection between closest genes to E13 regions in different castes. (B) Lollipop diagram indicating the Gene Ontology molecular function enrichment of closest genes to E13.

3.4. REs Demonstrate Opposing Trends in E8 and E15 Chromatin State Proportions within 0–3 and 20-40 Kimura Distances

In the E8 state, characterized as having extensive silent domains, an increase in chromatin state proportions is observed as the Kimura substitution level increases (Figure 4 top). However, for higher Kimura levels (Figure 4 bottom and Figure S4), the trend changes



towards a decrease in proportion, and at the same time, an increase in the E15 state—close to E8, but more enriched in ATAC-seq.

Figure 4. Bar plots depicting chromatin states in young (Kimura 0–3) and middle-aged (Kimura 20–40) repeats. (**Top**) Young repeats; (**bottom**) middle-aged repeats. Stacked bar graphs illustrate the proportional genome occupancy of each chromatin state across various Kimura distance intervals.

3.5. Transcriptional Activity of Repetitive Elements during the Larval Stage in Queens and Workers

The most actively expressed repeats are unclassified ones: rnd-1_family-83 (base-Mean = 12.28) and rnd-6_family-3082 (baseMean = 11.93). Among the classified elements, DNA elements are the most actively expressed (rnd-5_family-1919#DNA/CMC-EnSpm (baseMean = 10)) as well as Mariner elements (rnd-1_family-1 (baseMean = 9.23), rnd-6_family-371 (baseMean = 7.69), and rnd-6_family-1748 (baseMean = 7.03)).

Compared to some genes that are expressed differentially in queens and workers, repetitive elements have much less difference in expression (max. log fold change is 1.4 for an unknown repeat, rnd-1_family-6). Some DNA elements, however, are differentially expressed in the brains of larvae: rnd-5_family-1919#DNA/CMC-EnSpm is downregulated in queens (log2 fold change = 0.9, *p*-value < 0.05); rnd-1_family-78#DNA/TcMar-Mariner is upregulated, with log2 fold change = 0.71 (*p*-value < 0.05), as well as rnd-1_family-11#DNA/TcMar-Mariner, with log2 fold change = 0.67 (Figure 5).

3.6. Expression of Eight Repeats Segregates Brain Cell Populations

A total of five main cell populations were identified (Figure 6A). To identify glia (cluster 15), the marker genes LOC410151 (*repo*) [47], *Tret1*, and *GlnS* [48] were used. Hemocytes (cluster 17) were identified using markers LOC411597 (*hml*) and LOC551684 (*fer2LCH*) [49]. Cells expressing the neuronal marker *elav*, the functions of which have been shown for *D*. *m*. [50] (LOC410689), were defined as neurons (all clusters except 15 and 17). These cells are divided into olfactory projection neurons (OPNs), optic lobe cells (OLCs), and Kenyon cells (KCs) (Figure 6B).



Figure 5. The MA plot visualizes expression differences of the genes and repeats in queen and worker larvae. The *x*-axis represents the average expression level, while the *y*-axis depicts the log-fold change, highlighted with orange—transcripts with differential expression.



Figure 6. Clustering of single-cell transcriptome of the honey bee brain. (**A**) The UMAP of single cells from the brain of a honey bee, divided into 5 main clusters: hemocytes; glial cells; olfactory projection neurons (OPNs); optic lobe cells (OLCs); and Kenyon cells (KCs). OLCs were subdivided into Tm5c, Lawf2, and PM neurons. Kenyon cells were subdivided into class I small KCs and class I large KCs. (**B**) Dot plot of predictive gene markers used for cluster annotation. The black frame shows the marker/combination of markers used to identify the cell population.

OPNs: The axons of the olfactory receptors of the honey bee project into the lobes of the antennae, which consist of glomeruli. The glomeruli are connected to each other by local interneurons, and from the glomeruli, projection neurons are directed to the center of the brain of a higher order, such as the mushroom bodies and the lateral protocerebrum [51]. Using a combination of markers (LOC413466 (*oaz*), LOC410657 (*acj6*), and LOC724282 (*opt*)), we identified two clusters of olfactory projection neurons (clusters 11 and 12) [49,52].

OLCs: The optic lobes of the honey bee are responsible for the transduction of light stimuli and conductivity to mushroom bodies, where they evoke a reaction [53]. *Hiscl1* is highly expressed in cluster 14 and is associated with the type of Tm5c neurons. We also identified a population of lamina wide-field neurons of the second type—Lawf2 (cluster 16; LOC552079 (*hth*), LOC100577751 (*lim1*)). Finally, using the marker LOC410658 (*lim3*), we identified the PM cell type of the optic lobe (clusters 4, 5, and 7) [39,49].

KCs: The mushroom bodies are the processing centers for sensory information and are also involved in learning and memory [54]. We found five mushroom-body cell clusters (clusters 3, 6, 9, 10, and 13; LOC408804 (*plc*) and LOC408372 (*mub*)) [39].

Mushroom-body Kenyon cells are divided into two classes based on their morphology, function, and localization [52]. In addition, the first class is subdivided into three subclasses (small, medium, and large), and the population of FoxP-expressing cells is described separately [55]. Also, class I small KCs (clusters 6, 10; *e*74) and class I large KCs (clusters 3, 8, 9; mblk-1, cAmKii) [56] were identified. Class 1 medium cells expressing *mKast* gene and FoxP population cells were not found due to a lack of the necessary marker genes.

3.7. Repetitive Elements

Eight REs have been identified as markers of cell populations (Figure 7; Table S2). Unknown-1/2 is expressed in clusters 6, 13 (KCs), and 7 (OLCs). Unknown-4/303 is expressed in 11, 12 (OPNs), and 16 (OLCs). Unknown-5/2258 is expressed in three, eight, and nine clusters, identified as class 1 large KCs. Copia-5/1071 is expressed in 11 (OPNs) and 17 (hemocytes). Unknown-6/719 is expressed in four (OLCs), 10 (KCs), and 17 (hemocytes). EnSpm-5/1919 and Unknown-1/33 show a similar expression pattern in 10 (KCs) and 17 (hemocytes). Hemocytes play a crucial role in the cell-mediated immunity of insects [57]. They are responsible for various functions, such as detecting infectious agents, the phagocytosis of small particles, and the encapsulation of larger foreign objects [58]. Mariner-1/1 is found in two clusters: 7 (OLCs) and 13 (KCs).



Figure 7. Dot plot of eight REs predicted as markers. Among the identified markers, the highest expression level (average expression 2.5) was observed for the Copia-5/1071 in cluster 11, which was identified as olfactory projection neurons. Additionally, the EnSpm-5/1919 showed high expression in cluster 17, characterized as hemocytes.

4. Discussion

4.1. Eusociality

Eusociality represents an intricate form of social conduct marked by a division of labor between reproductive and non-reproductive castes for achieving a highly structured and cooperative social organization in the natural world. This high level of social organization has independently emerged in numerous Arthropod taxa, including Decapoda (shrimps), Isoptera (termites), Aphididae (aphids), Thysanoptera (thrips), Coleoptera: Scolytidae (bark beetles), and Platypodidae (ambrosia beetles) [59–63], and particularly in Hymenoptera (sawflies, wasps, bees, and ants) [64–67]. The origin of eusociality is estimated to have occurred at least fifteen times within the aculeate Hymenoptera, without taking into account a sole origin amongst ants [7]. The development of eusociality represents a significant adaptive advantage for these species and has culminated in their global prevalence and diverse ecological roles today. The evolutionary origins of eusociality have attracted biologists' interest for many years, resulting in several theories that explore various evolutionary processes. The precise emergence of eusociality in Hymenoptera cannot be pinpointed in the fossil record, rendering the timeline a subject of ongoing analysis. However, according to phylogenetic studies and molecular clock analyses, it is estimated that eusociality emerged approximately 100 million years ago during the Cretaceous period [68].

4.2. Distribution of TEs

Transposon-mediated gene regulation is likely one of the mechanisms underlying the eusocial transition [69]. Although eusocial species of shrimps [59], cockroaches, and termites [70] have been found to have a higher number of transposons, bees show a reduced quantity in conjunction with smaller genome sizes. The decrease in the number of transposable elements observed in eusocial bees may be a consequence of their social life, and possibly linked to a balance between genomic diversity and integrity, driven by recombination and TE suppression [71,72].

Repetitive elements comprise a significant proportion of the genome in many metazoans. While repeatomes are usually similar in closely related organisms, their content and landscape may significantly vary in different representatives of one genus [73]. Moreover, differences in repetitive element distribution can be observed between closely related species [16], which is caused by the expansion and elimination of repetitive element copies. The expansion of the AluI element observed in the genome of *A. m. ligustica* suggests that REs are constantly evolving and may be the source of intraspecies variations.

The honey bee contains an unusually low amount of repetitive DNA. Many elements are fragmentary and incomplete, and although they may contain sequences of typical transposable element protein domains, they are often rendered inactive by stop codons and frameshift mutations [74]. Social complexity in bees has been shown to negatively correlate with the abundance and diversity of transposable elements (TEs) [16,18].

In multicellular organisms, a number of protective mechanisms guard against the uncontrolled proliferation of transposable elements. These defense systems, particularly piwi-interacting RNA (piRNA) [75,76] and zinc-finger genes [77,78], are notable among these mechanisms. They adapt to changing transposable elements, thereby applying selection pressure on the TEs [79,80]. Major biological shifts frequently involve the expansion of transcription factor families [81].

Repetitive elements have the potential to evolve into cis-regulatory elements [82]. In mammalian genomes, transposons are increasingly recognized as a significant source of diverse cis-regulatory sequences. For instance, in mice and humans, LTRs linked to several pluripotent transcription factors are commonly found to be enriched with enhancer-associated chromatin signatures such as H3K27ac and H3K4me1, indicating their role in the regulation of gene expression [83].

4.3. Limitations of ChromHMM

Our study has several technical limitations. The alignment of the results from a biological experiment with sequencing reads of 150 bp does not provide 100% accuracy, as the repetitive elements in the genome of *A. mellifera* are lengthy, with an average length of 343.4 bp and a maximum length of 11,867 bp. Aligning genome pieces with 150 bp reads cannot cover the length of REs, leading to low alignment accuracy for features such as Chip-seq and RNA-seq regions with REs [84]. The genome resolution of 50 bp for the mean or median value also hinders achieving high accuracy. The limited number of available features (ATAC-seq, RNA-seq, and four modified histones) can obscure potentially active regions of the genome that may play a role in certain processes. For example, adding a new feature can decrease state 8 in our model and expand another state, or even introduce a new state, thereby describing a greater number of genome regions with non-silent states [43,85]. The above factors can significantly impact the final result of clustering and analysis. Improving alignment accuracy, increasing the length of sequenced

reads to cover repetitive elements of the genome (for instance, by using Oxford Nanopore), incorporating more features (a broader range of histone modifications, for example, the histone group, h1, h2a, h2b, h3, h4, and their modifications such as h3k9me3, as well as other features like RNA-seq, atac-seq, and wgbs), and exploring other clustering methods can enhance the accuracy and resolution of the methods [85,86].

Overall, the combination of proposed improvements could increase the accuracy and resolution of the methods, even if they do not fully overcome the problem of repetitive elements and allow more genomes to be clustered with greater accuracy.

4.4. REs Markers in scRNA-seq Data

Eight TE sequences were identified as markers in different populations of scRNA-seq data from the adult honey bee brain. The limited number of markers can be attributed to the low activity of transposable elements in adult tissues [87] and the relatively low abundance of TEs in the bee genome compared to other insects [14]. Studies have shown patterns of transposon expression in the adult *Drosophila* brain [88], but there is still insufficient data to identify expression patterns of mobile elements and their potential role in caste determination in bees.

5. Conclusions

Based on our extensive investigations, we have concluded that the repetitive elements present in the *A. m.* species are associated with the divergence observed among its subspecies. Despite the recent expansion of these repetitive elements, our analysis suggests that they have no discernible effect on caste formation at the transcriptomic level. However, our examination of the chromatin landscape has revealed the existence of a distinct state, specifically associated with these repetitive elements. This discovery, combined with the apparent process of "domestication", serves as the basis for uncovering compelling evidence for the direct involvement of genomic repeats in the evolution of social behavior. The emergence of a unique chromatin state associated with repetitive elements provides preliminary evidence for the potential involvement of repeat-derived sequences in the evolution of *A. mellifera*. The ongoing dynamics of domestication provide an avenue for this involvement in the immediate evolutionary future.

These findings shed new light on the complex mechanisms underlying the evolution of social behavior and provide valuable insights into the role of repetitive elements in shaping the genomic landscape of honey bees.

Supplementary Materials: The following supporting information can be downloaded at https://www. mdpi.com/article/10.3390/genes15010089/s1: Figure S1: Scatter plot of repeat coverage vs repeat copy number in genome using different modes of alignment. Every dot depicts a distinct repeat element. The x-axis depicts the repeat copy number and the y-axis depicts the median coverage of these elements. Blue colour dots—the unique alignment mode, ignoring all the multi-mapper hits. Orange dots represent values obtained with the multialignment mode, reporting all the hits for every element; Figure S2: Boxplot of repeat coverage comparing the alignments in multimapper mode and single mapper mode. To the right—the unique alignment mode, ignoring all the multi-mapper hits. The left box represents values obtained with the multi alignment mode, reporting all the hits for every element; Figure S3: Density plot of the lengths distribution of the repeats of different classes identified by RepeatModeler; Figure S4: Bar plots of chromatin states in repeats throughout the total range of Kimura distances Stacked bar plots indicate the occupied fraction of the genome by each chromatin state for every Kimura distance window value; Figure S5: Features of chromatin states and genome occupancy in brain of the workers of A.m.mellifera. Columns from left to right—Occupied genome fraction (purple), Features comprising states (in blue), and relative enrichment of respective genomic regions (in green); Table S1: ChromHMM chromatin states interpretation; Table S2: Repeats, which serve as cluster markers in sc-RNA-seq analysis. Average expression shows the mean counts by all cells in a given cluster. Percent expressed shows the portion of cells expressing the repeat.

Author Contributions: L.A., N.P. and L.O. formed the concept of this study; methodology and software, N.P., L.O., V.M., M.S. and E.L.; L.A., N.P. and L.O. supervised the preparation of the draft.

All authors contributed to the article and approved the submitted version. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Ministry of Science and Higher Education of the Russian Federation within the framework of the Federal Scientific and Technical Program for the Development of Genetic Technologies for 2019–2027 (agreement no.: 075-15-2021-1345, unique identifier: RF-193021X0012).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article and supplementary materials.

Acknowledgments: This research would not have been possible without the assistance of A. Drozdov and A. Lisitsa.

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

References

- Aizen, M.A.; Garibaldi, L.A.; Cunningham, S.A.; Klein, A.M. How Much Does Agriculture Depend on Pollinators? Lessons from Long-Term Trends in Crop Production. *Ann. Bot.* 2009, 103, 1579–1588. [CrossRef]
- Christmann, S. Do We Realize the Full Impact of Pollinator Loss on Other Ecosystem Services and the Challenges for Any Restoration in Terrestrial Areas? *Restor. Ecol.* 2019, 27, 720–725. [CrossRef]
- 3. Patel, V.; Pauli, N.; Biggs, E.; Barbour, L.; Boruff, B. Why Bees Are Critical for Achieving Sustainable Development. *Ambio* 2021, 50, 49–59. [CrossRef] [PubMed]
- 4. Dangles, O.; Casas, J. Ecosystem Services Provided by Insects for Achieving Sustainable Development Goals. *Ecosyst. Serv.* 2019, 35, 109–115. [CrossRef]
- 5. Kohno, H.; Kubo, T. Genetics in the Honey Bee: Achievements and Prospects toward the Functional Analysis of Molecular and Neural Mechanisms Underlying Social Behaviors. *Insects* **2019**, *10*, 348. [CrossRef]
- Wilson, E.O.; Hölldobler, B. Eusociality: Origin and Consequences. Proc. Natl. Acad. Sci. USA 2005, 102, 13367–13371. [CrossRef] [PubMed]
- 7. da Silva, J. Life History and the Transitions to Eusociality in the Hymenoptera. *Front. Ecol. Evol.* **2021**, 9. [CrossRef]
- 8. Ashby, R.; Forêt, S.; Searle, I.; Maleszka, R. MicroRNAs in Honey Bee Caste Determination. Sci. Rep. 2016, 6, 18794. [CrossRef]
- Alaux, C.; Sinha, S.; Hasadsri, L.; Hunt, G.J.; Guzmán-Novoa, E.; DeGrandi-Hoffman, G.; Uribe-Rubio, J.L.; Southey, B.R.; Rodriguez-Zas, S.; Robinson, G.E. Honey Bee Aggression Supports a Link between Gene Regulation and Behavioral Evolution. *Proc. Natl. Acad. Sci. USA* 2009, 106, 15400–15405. [CrossRef]
- Greenberg, J.K.; Xia, J.; Zhou, X.; Thatcher, S.R.; Gu, X.; Ament, S.A.; Newman, T.C.; Green, P.J.; Zhang, W.; Robinson, G.E.; et al. Behavioral Plasticity in Honey Bees Is Associated with Differences in Brain microRNA Transcriptome. *Genes Brain Behav.* 2012, 11, 660–670. [CrossRef]
- 11. Eyer, M.; Dainat, B.; Neumann, P.; Dietemann, V. Social Regulation of Ageing by Young Workers in the Honey Bee, *Apis mellifera*. *Exp. Gerontol.* **2017**, *87*, 84–91. [CrossRef] [PubMed]
- de Paula Junior, D.E.; de Oliveira, M.T.; Bruscadin, J.J.; Pinheiro, D.G.; Bomtorin, A.D.; Coelho Júnior, V.G.; Moda, L.M.R.; Simões, Z.L.P.; Barchuk, A.R. Caste-Specific Gene Expression Underlying the Differential Adult Brain Development in the Honeybee *Apis mellifera*. *Insect Mol. Biol.* 2021, 30, 42–56. [CrossRef] [PubMed]
- 13. Wang, M.; Xiao, Y.; Li, Y.; Wang, X.; Qi, S.; Wang, Y.; Zhao, L.; Wang, K.; Peng, W.; Luo, G.-Z.; et al. RNA m6A Modification Functions in Larval Development and Caste Differentiation in Honeybee (*Apis mellifera*). Cell Rep. **2021**, 34, 108580. [CrossRef]
- 14. 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]
- 15. Gregory, T.R.; Nicol, J.A.; Tamm, H.; Kullman, B.; Kullman, K.; Leitch, I.J.; Murray, B.G.; Kapraun, D.F.; Greilhuber, J.; Bennett, M.D. Eukaryotic Genome Size Databases. *Nucleic Acids Res.* **2007**, *35*, D332–D338. [CrossRef]
- Petersen, M.; Armisén, D.; Gibbs, R.A.; Hering, L.; Khila, A.; Mayer, G.; Richards, S.; Niehuis, O.; Misof, B. Diversity and Evolution of the Transposable Element Repertoire in Arthropods with Particular Reference to Insects. *BMC Evol. Biol.* 2019, 19, 11. [CrossRef] [PubMed]
- 17. Jiang, F.; Yang, M.; Guo, W.; Wang, X.; Kang, L. Large-Scale Transcriptome Analysis of Retroelements in the Migratory Locust, Locusta Migratoria. *PLoS ONE* 2012, 7, e40532. [CrossRef]
- Gilbert, C.; Peccoud, J.; Cordaux, R. Transposable Elements and the Evolution of Insects. *Annu. Rev. Entomol.* 2021, 66, 355–372.
 [CrossRef]
- Feschotte, C. Transposable Elements and the Evolution of Regulatory Networks. *Nat. Rev. Genet.* 2008, *9*, 397–405. [CrossRef]
 [PubMed]

- 20. Bourque, G.; Burns, K.H.; Gehring, M.; Gorbunova, V.; Seluanov, A.; Hammell, M.; Imbeault, M.; Izsvák, Z.; Levin, H.L.; Macfarlan, T.S.; et al. Ten Things You Should Know about Transposable Elements. *Genome Biol.* **2018**, *19*, 199. [CrossRef]
- Carareto, C.M.A.; Hernandez, E.H.; Vieira, C. Genomic Regions Harboring Insecticide Resistance-Associated Cyp Genes Are Enriched by Transposable Element Fragments Carrying Putative Transcription Factor Binding Sites in Two Sibling Drosophila Species. *Gene* 2014, 537, 93–99. [CrossRef] [PubMed]
- 22. Wu, C.; Lu, J. Diversification of Transposable Elements in Arthropods and Its Impact on Genome Evolution. *Genes* **2019**, *10*, 338. [CrossRef] [PubMed]
- Ellison, C.E.; Bachtrog, D. Dosage Compensation via Transposable Element Mediated Rewiring of a Regulatory Network. *Science* 2013, 342, 846–850. [CrossRef] [PubMed]
- Pardue, M.-L.; Rashkova, S.; Casacuberta, E.; DeBaryshe, P.G.; George, J.A.; Traverse, K.L. Two Retrotransposons Maintain Telomeres in Drosophila. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* 2005, 13, 443–453. [CrossRef] [PubMed]
- Jangam, D.; Feschotte, C.; Betrán, E. Transposable Element Domestication as an Adaptation to Evolutionary Conflicts. *Trends Genet. TIG* 2017, 33, 817–831. [CrossRef] [PubMed]
- Johnson, R.; Guigó, R. The RIDL Hypothesis: Transposable Elements as Functional Domains of Long Noncoding RNAs. RNA 2014, 20, 959–976. [CrossRef] [PubMed]
- 27. Bonasio, R. Emerging Topics in Epigenetics: Ants, Brains, and Noncoding RNAs. Ann. N. Y. Acad. Sci. 2012, 1260, 14–23. [CrossRef]
- 28. Bonasio, R.; Shiekhattar, R. Regulation of Transcription by Long Noncoding RNAs. Annu. Rev. Genet. 2014, 48, 433–455. [CrossRef]
- 29. Rinn, J.L.; Chang, H.Y. Genome Regulation by Long Noncoding RNAs. Annu. Rev. Biochem. 2012, 81, 145–166. [CrossRef]
- 30. Hou, Y.; Zhang, R.; Sun, X. Enhancer LncRNAs Influence Chromatin Interactions in Different Ways. *Front. Genet.* **2019**, *10*, 936. [CrossRef] [PubMed]
- Amaral, P.P.; Leonardi, T.; Han, N.; Viré, E.; Gascoigne, D.K.; Arias-Carrasco, R.; Büscher, M.; Pandolfini, L.; Zhang, A.; Pluchino, S.; et al. Genomic Positional Conservation Identifies Topological Anchor Point RNAs Linked to Developmental Loci. *Genome Biol.* 2018, 19, 32. [CrossRef]
- 32. Engreitz, J.M.; Haines, J.E.; Perez, E.M.; Munson, G.; Chen, J.; Kane, M.; McDonel, P.E.; Guttman, M.; Lander, E.S. Local Regulation of Gene Expression by lncRNA Promoters, Transcription and Splicing. *Nature* **2016**, *539*, 452–455. [CrossRef]
- 33. Feng, W.; Huang, J.; Zhang, Z.; Nie, H.; Lin, Y.; Li, Z.; Su, S. Understanding of Waggle Dance in the Honey Bee (*Apis mellifera*) from the Perspective of Long Non-Coding RNA. *Insects* **2022**, *13*, 111. [CrossRef]
- Liu, F.; Shi, T.; Qi, L.; Su, X.; Wang, D.; Dong, J.; Huang, Z.Y. lncRNA Profile of *Apis mellifera* and Its Possible Role in Behavioural Transition from Nurses to Foragers. *BMC Genom.* 2019, 20, 393. [CrossRef] [PubMed]
- 35. Rubin, B.E.R.; Jones, B.M.; Hunt, B.G.; Kocher, S.D. Rate Variation in the Evolution of Non-Coding DNA Associated with Social Evolution in Bees. *Philos. Trans. R. Soc. B Biol. Sci.* **2019**, *374*, 20180247. [CrossRef] [PubMed]
- Wojciechowski, M.; Lowe, R.; Maleszka, J.; Conn, D.; Maleszka, R.; Hurd, P.J. Phenotypically Distinct Female Castes in Honey Bees Are Defined by Alternative Chromatin States during Larval Development. *Genome Res.* 2018, 28, 1532–1542. [CrossRef] [PubMed]
- 37. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-Optimal Probabilistic RNA-Seq Quantification. *Nat. Biotechnol.* **2016**, *34*, 525–527. [CrossRef]
- 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]
- Zhang, W.; Wang, L.; Zhao, Y.; Wang, Y.; Chen, C.; Hu, Y.; Zhu, Y.; Sun, H.; Cheng, Y.; Sun, Q.; et al. Single-Cell Transcriptomic Analysis of Honeybee Brains Identifies Vitellogenin as Caste Differentiation-Related Factor. *iScience* 2022, 25, 104643. [CrossRef] [PubMed]
- Satija, R.; Farrell, J.A.; Gennert, D.; Schier, A.F.; Regev, A. Spatial Reconstruction of Single-Cell Gene Expression Data. *Nat. Biotechnol.* 2015, 33, 495–502. [CrossRef]
- 41. Lowe, R.; Wojciechowski, M.; Ellis, N.; Hurd, P.J. Chromatin Accessibility-Based Characterisation of Brain Gene Regulatory Networks in Three Distinct Honey Bee Polyphenisms. *Nucleic Acids Res.* **2022**, *50*, 11550–11562. [CrossRef]
- 42. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 2013, 29, 15–21. [CrossRef] [PubMed]
- 43. Ernst, J.; Kellis, M. Chromatin-State Discovery and Genome Annotation with ChromHMM. *Nat. Protoc.* **2017**, *12*, 2478–2492. [CrossRef] [PubMed]
- Kharchenko, P.V.; Alekseyenko, A.A.; Schwartz, Y.B.; Minoda, A.; Riddle, N.C.; Ernst, J.; Sabo, P.J.; Larschan, E.; Gorchakov, A.A.; Gu, T.; et al. Comprehensive Analysis of the Chromatin Landscape in Drosophila Melanogaster. *Nature* 2011, 471, 480–485. [CrossRef]
- 45. Rada-Iglesias, A. Is H3K4me1 at Enhancers Correlative or Causative? Nat. Genet. 2018, 50, 4–5. [CrossRef]
- 46. Barral, A.; Déjardin, J. The Chromatin Signatures of Enhancers and Their Dynamic Regulation. *Nucleus* **2023**, *14*, 2160551. [CrossRef] [PubMed]
- Edwards, T.N.; Meinertzhagen, I.A. The Functional Organisation of Glia in the Adult Brain of Drosophila and Other Insects. *Prog. Neurobiol.* 2010, 90, 471–497. [CrossRef]

- 48. Allen, A.M.; Neville, M.C.; Birtles, S.; Croset, V.; Treiber, C.D.; Waddell, S.; Goodwin, S.F. A Single-Cell Transcriptomic Atlas of the Adult Drosophila Ventral Nerve Cord. *eLife* 2020, *9*, e54074. [CrossRef]
- 49. Davie, K.; Janssens, J.; Koldere, D.; De Waegeneer, M.; Pech, U.; Kreft, Ł.; Aibar, S.; Makhzami, S.; Christiaens, V.; Bravo González-Blas, C.; et al. A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **2018**, *174*, 982–998.e20. [CrossRef]
- 50. Robinow, S.; White, K. Characterization and Spatial Distribution of the ELAV Protein during Drosophila Melanogaster Development. J. Neurobiol. 1991, 22, 443–461. [CrossRef]
- 51. Galizia, C.G.; Menzel, R. Odour Perception in Honeybees: Coding Information in Glomerular Patterns. *Curr. Opin. Neurobiol.* **2000**, *10*, 504–510. [CrossRef]
- 52. Groh, C.; Rössler, W. Analysis of Synaptic Microcircuits in the Mushroom Bodies of the Honeybee. *Insects* 2020, 11, 43. [CrossRef] [PubMed]
- 53. Roat, T.C.; da Cruz-Landim, C. Mitosis and Cell Death in the Optic Lobes of Workers, Queens and Drones of the Honey Bee (*Apis mellifera*) during Metamorphosis. *J. Biosci.* 2010, 35, 415–425. [CrossRef]
- Caron, S.; Abbott, L.F. Neuroscience: Intelligence in the Honeybee Mushroom Body. Curr. Biol. CB 2017, 27, R220–R223. [CrossRef] [PubMed]
- Suenami, S.; Oya, S.; Kohno, H.; Kubo, T. Kenyon Cell Subtypes/Populations in the Honeybee Mushroom Bodies: Possible Function Based on Their Gene Expression Profiles, Differentiation, Possible Evolution, and Application of Genome Editing. *Front. Psychol.* 2018, 9. [CrossRef] [PubMed]
- 56. Kaneko, K.; Suenami, S.; Kubo, T. Gene Expression Profiles and Neural Activities of Kenyon Cell Subtypes in the Honeybee Brain: Identification of Novel "middle-Type" Kenyon Cells. *Zool. Lett.* **2016**, *2*, 14. [CrossRef] [PubMed]
- 57. Eleftherianos, I.; Xu, M.; Yadi, H.; Ffrench-Constant, R.H.; Reynolds, S.E. Plasmatocyte-Spreading Peptide (PSP) Plays a Central Role in Insect Cellular Immune Defenses against Bacterial Infection. *J. Exp. Biol.* **2009**, *212*, 1840–1848. [CrossRef]
- Negri, P.; Maggi, M.; Ramirez, L.; Szawarski, N.; De Feudis, L.; Lamattina, L.; Eguaras, M. Cellular Immunity in *Apis mellifera*: Studying Hemocytes Brings Light about Bees Skills to Confront Threats. *Apidologie* 2016, 47, 379–388. [CrossRef]
- 59. Chak, S.T.C.; Harris, S.E.; Hultgren, K.M.; Jeffery, N.W.; Rubenstein, D.R. Eusociality in Snapping Shrimps Is Associated with Larger Genomes and an Accumulation of Transposable Elements. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2025051118. [CrossRef]
- 60. Thorne, B.L.; Breisch, N.L.; Muscedere, M.L. Evolution of Eusociality and the Soldier Caste in Termites: Influence of Intraspecific Competition and Accelerated Inheritance. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12808–12813. [CrossRef]
- Lawson, S.P.; Legan, A.W.; Graham, C.; Abbot, P. Comparative Phenotyping across a Social Transition in Aphids. *Anim. Behav.* 2014, 96, 117–125. [CrossRef]
- 62. Chapman, T.W.; Crespi, B.J.; Kranz, B.D.; Schwarz, M.P. High Relatedness and Inbreeding at the Origin of Eusociality in Gall-Inducing Thrips. *Proc. Natl. Acad. Sci. USA* 2000, *97*, 1648–1650. [CrossRef]
- 63. Biedermann, P.H.W.; Taborsky, M. Larval Helpers and Age Polyethism in Ambrosia Beetles. *Proc. Natl. Acad. Sci. USA* 2011, 108, 17064–17069. [CrossRef] [PubMed]
- 64. Oeyen, J.P.; Baa-Puyoulet, P.; Benoit, J.B.; Beukeboom, L.W.; Bornberg-Bauer, E.; Buttstedt, A.; Calevro, F.; Cash, E.I.; Chao, H.; Charles, H.; et al. Sawfly Genomes Reveal Evolutionary Acquisitions That Fostered the Mega-Radiation of Parasitoid and Eusocial Hymenoptera. *Genome Biol. Evol.* **2020**, *12*, 1099–1188. [CrossRef]
- 65. Torres, V.O.; Montagna, T.S.; Raizer, J.; Antonialli-Junior, W.F. Division of Labor in Colonies of the Eusocial Wasp, Mischocyttarus Consimilis. J. Insect Sci. 2012, 12, 21. [CrossRef]
- 66. Cardinal, S.; Danforth, B.N. The Antiquity and Evolutionary History of Social Behavior in Bees. *PLoS ONE* **2011**, *6*, e21086. [CrossRef]
- 67. Favreau, E.; Martínez-Ruiz, C.; Rodrigues Santiago, L.; Hammond, R.L.; Wurm, Y. Genes and Genomic Processes Underpinning the Social Lives of Ants. *Curr. Opin. Insect Sci.* 2018, 25, 83–90. [CrossRef]
- 68. Opachaloemphan, C.; Yan, H.; Leibholz, A.; Desplan, C.; Reinberg, D. Recent Advances in Behavioral (Epi)Genetics in Eusocial Insects. *Annu. Rev. Genet.* **2018**, *52*, 489–510. [CrossRef]
- 69. Berger, J.; Legendre, F.; Zelosko, K.-M.; Harrison, M.C.; Grandcolas, P.; Bornberg-Bauer, E.; Fouks, B. Eusocial Transition in Blattodea: Transposable Elements and Shifts of Gene Expression. *Genes* **2022**, *13*, 1948. [CrossRef]
- Korb, J.; Poulsen, M.; Hu, H.; Li, C.; Boomsma, J.J.; Zhang, G.; Liebig, J. A Genomic Comparison of Two Termites with Different Social Complexity. *Front. Genet.* 2015, 6, 9. [CrossRef]
- 71. Kapheim, K.M.; Pan, H.; Li, C.; Salzberg, S.L.; Puiu, D.; Magoc, T.; Robertson, H.M.; Hudson, M.E.; Venkat, A.; Fischman, B.J.; et al. Genomic Signatures of Evolutionary Transitions from Solitary to Group Living. *Science* **2015**, *348*, 1139–1143. [CrossRef]
- 72. Fouks, B.; Brand, P.; Nguyen, H.N.; Herman, J.; Camara, F.; Ence, D.; Hagen, D.E.; Hoff, K.J.; Nachweide, S.; Romoth, L.; et al. The Genomic Basis of Evolutionary Differentiation among Honey Bees. *Genome Res.* **2021**, *31*, 1203–1215. [CrossRef] [PubMed]
- Schartl, M.; Kneitz, S.; Volkoff, H.; Adolfi, M.; Schmidt, C.; Fischer, P.; Minx, P.; Tomlinson, C.; Meyer, A.; Warren, W.C. The Piranha Genome Provides Molecular Insight Associated to Its Unique Feeding Behavior. *Genome Biol. Evol.* 2019, *11*, 2099–2106. [CrossRef] [PubMed]
- 74. Elsik, C.G.; Worley, K.C.; Bennett, A.K.; Beye, M.; Camara, F.; Childers, C.P.; de Graaf, D.C.; Debyser, G.; Deng, J.; Devreese, B.; et al. Finding the Missing Honey Bee Genes: Lessons Learned from a Genome Upgrade. *BMC Genom.* **2014**, *15*, 86. [CrossRef] [PubMed]

- 75. Song, J.L.; Stoeckius, M.; Maaskola, J.; Friedländer, M.; Stepicheva, N.; Juliano, C.; Lebedeva, S.; Thompson, W.; Rajewsky, N.; Wessel, G.M. Select microRNAs Are Essential for Early Development in the Sea Urchin. *Dev. Biol.* 2012, 362, 104–113. [CrossRef] [PubMed]
- 76. Santos, D.; Feng, M.; Kolliopoulou, A.; Taning, C.N.T.; Sun, J.; Swevers, L. What Are the Functional Roles of Piwi Proteins and piRNAs in Insects? *Insects* 2023, 14, 187. [CrossRef]
- 77. Lukic, S.; Nicolas, J.-C.; Levine, A.J. The Diversity of Zinc-Finger Genes on Human Chromosome 19 Provides an Evolutionary Mechanism for Defense against Inherited Endogenous Retroviruses. *Cell Death Differ.* **2014**, *21*, 381–387. [CrossRef]
- 78. Baumgartner, L.; Handler, D.; Platzer, S.W.; Yu, C.; Duchek, P.; Brennecke, J. The Drosophila ZAD Zinc Finger Protein Kipferl Guides Rhino to piRNA Clusters. *eLife* **2022**, *11*, e80067. [CrossRef]
- 79. Catlin, N.S.; Josephs, E.B. The Important Contribution of Transposable Elements to Phenotypic Variation and Evolution. *Curr. Opin. Plant Biol.* **2022**, *65*, 102140. [CrossRef]
- Wells, J.N.; Chang, N.-C.; McCormick, J.; Coleman, C.; Ramos, N.; Jin, B.; Feschotte, C. Transposable Elements Drive the Evolution of Metazoan Zinc Finger Genes. *Genome Res.* 2023, 33, 1325–1339. [CrossRef]
- Harrison, M.C.; Jongepier, E.; Robertson, H.M.; Arning, N.; Bitard-Feildel, T.; Chao, H.; Childers, C.P.; Dinh, H.; Doddapaneni, H.; Dugan, S.; et al. Hemimetabolous Genomes Reveal Molecular Basis of Termite Eusociality. *Nat. Ecol. Evol.* 2018, 2, 557–566. [CrossRef]
- 82. Gebrie, A. Transposable Elements as Essential Elements in the Control of Gene Expression. *Mob. DNA* 2023, *14*, 9. [CrossRef] [PubMed]
- 83. Sundaram, V.; Wysocka, J. Transposable Elements as a Potent Source of Diverse Cis-Regulatory Sequences in Mammalian Genomes. *Philos. Trans. R. Soc. B Biol. Sci.* 2020, 375, 20190347. [CrossRef] [PubMed]
- 84. Alser, M.; Rotman, J.; Deshpande, D.; Taraszka, K.; Shi, H.; Baykal, P.I.; Yang, H.T.; Xue, V.; Knyazev, S.; Singer, B.D.; et al. Technology Dictates Algorithms: Recent Developments in Read Alignment. *Genome Biol.* **2021**, *22*, 249. [CrossRef] [PubMed]
- 85. van der Velde, A.; Fan, K.; Tsuji, J.; Moore, J.E.; Purcaro, M.J.; Pratt, H.E.; Weng, Z. Annotation of Chromatin States in 66 Complete Mouse Epigenomes during Development. *Commun. Biol.* **2021**, *4*, 239. [CrossRef]
- 86. Brown, E.J.; Bachtrog, D. The Chromatin Landscape of Drosophila: Comparisons between Species, Sexes, and Chromosomes. *Genome Res.* **2014**, 24, 1125–1137. [CrossRef]
- 87. He, J.; Babarinde, I.A.; Sun, L.; Xu, S.; Chen, R.; Shi, J.; Wei, Y.; Li, Y.; Ma, G.; Zhuang, Q.; et al. Identifying Transposable Element Expression Dynamics and Heterogeneity during Development at the Single-Cell Level with a Processing Pipeline scTE. *Nat. Commun.* **2021**, *12*, 1456. [CrossRef]
- 88. Treiber, C.D.; Waddell, S. Transposon Expression in the Drosophila Brain Is Driven by Neighboring Genes and Diversifies the Neural Transcriptome. *Genome Res.* 2020, *30*, 1559–1569. [CrossRef]

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.