

Chromatin dynamics of a large-sized genome provides insights into polyphenism and XO dosage compensation of locusts

In the format provided by the
authors and unedited

Tables of contents

Supplementary Note 1: Chromosome-level assembly of the migratory locust.....	2
Supplementary Note 2: Combinational chromatin patterns of genes.....	3
Supplementary Note 3: Chromatin functional organization of the locust genome	4
Supplementary Note 4: TE contribution to chromatin states.....	5
Supplementary Note 5: SXL-independent initiation of dosage compensation.....	5
Supplementary Note 6: Gene configurations of the X chromosome of insects.....	6
Supplementary Note 7: Relocation of female-related genes.....	7
Supplementary Note 8: Selection of histone modifications	7
Supplementary Figures.....	9
Supplementary Tables	43
References	44

Supplementary Note 1: Chromosome-level assembly of the migratory locust

We conducted genome sequencing on an adult female locust (Supplementary Figure 1), which has a genomic heterozygosity of 1.58% (Supplementary Figure 2). A combination of three sequencing technologies including long-read sequencing (778.27 Gb of PacBio Sequel data and 713.02 Gb Illumina HiSeq X Ten data), optical mapping (3405.06 Gbp of Bionano optical map data, Direct Label Enzyme labeling), and chromosome conformation capture technique based on proximity ligation (1649.75 Gb of Hi-C data) was applied to construct the chromosome-level genome assembly of the migratory locust (Figure 1a). The chromosome-level genome (LMv3.1) includes 12 large chromosomes, covering 91.56% of the total assembled sequences (6.95 Gb), with a contig N50 of 2.23 Mb and a maximum chromosome length of 1.12 Gb (Figure 1b, Supplementary Table 1). The LMv3.1 assembly covers 96.9% of complete conserved orthologs in Arthropoda gene groups of Benchmarking Universal Single-Copy Orthologs (BUSCO). In X0/XX sex determination, females have an equal number of copies of the X chromosome compared to other autosomes but twice that of males. Therefore, we performed single-individual DNA-seq sequencing of both male and female locusts to determine the X chromosome using a chromosome quotient method. In females, the third longest scaffold (977 Mb), corresponding to the X chromosome, exhibited equal read coverage compared to the other 11 autosomes, as well as approximately twice the read coverage of the X chromosome observed in males (Figure 1c). The length of the X chromosome identified using the chromosome quotient method correlated well with the length determined by Hoechst staining of metaphase chromosomes (Figure 1d). Compared to the previous genome assembly LMv2.4 (564,428 scaffolds with a contig N50 of 9.3 Kb), the LMv3.1 assembly exhibits a 240-fold improvement in contig N50 and a 2,255-fold improvement in scaffold N50. Approximately 79.34% and 95.28% of the properly mapped Illumina reads aligned to the LMv2.4 and LMv3.1 assemblies, respectively. Transposable elements (TEs) accounted for approximately 70.34% of the locust genome, in which non-LTR retrotransposons (29.92%) are the most prevalent types. The high contiguity of the LMv3.1 assembly greatly improved the completeness of the TEs (Supplementary Figure 3). Fluorescent in situ hybridization of a satellite DNA sequence located in pericentromeric regions (Figure 1e) showed fluorescent bands on one of the chromosome ends (Figure 1f), demonstrating the anticipated chromosome end regions. Thus, a significant quality improvement has been achieved in the completeness and

accuracy of the LMv3.1 assembly.

To predict protein-coding genes, we further generated 113.01 Gb of short-read RNA-seq and 112.03 Gb of PacBio ISO-seq cDNA data obtained from 11 different tissues and organs of the locust. We predicted 18,127 protein-coding genes by utilizing Illumina RNA-seq and PacBio ISO-seq data, along with homologous proteins from eight insect species (Supplementary Figure 4). The geneset covers 91.4% of complete conserved orthologs in Arthropoda gene groups of BUSCO. The average gene length in the LMv3.1 assembly is approximately two-fold longer than that in the LMv2.4 assembly (Supplementary Figure 5). The LMv3.1 assembly presented the higher mapping rates of full-length ISO-seq transcripts than the LMv2.4 assembly, indicating improved coverage of RNA transcript structures (Supplementary Figure 6). The locust genome exhibits a unique phenomenon associated with large genomes, namely distant transcription initiation, due to intron size expansion in the mRNA leader, which spans from upstream of the start codon to transcription initiation sites of protein-coding genes¹. These distant promoters, located far upstream of the start codon, display high transcription initiation activity and are regulated by epigenetic control. Hence, the improvement in genome contiguity facilitates further investigation of the epigenomic landscape in locusts.

The satellite DNA sequences, which are commonly found in the (peri)centromeric region, are located proximal to one of the chromosome ends (Figure 1e). In the pericentromeric regions, a high copy number of two satellite DNA sequences, LmCentrol83 and LmCentrol88, are present with monomer lengths of 183 bp and 188 bp, respectively. Fluorescent in situ hybridization of LmCentrol88 showed FISH bands on one of the chromosome ends (Figure 1f), demonstrating the anticipated chromosome end regions.

Supplementary Note 2: Combinational chromatin patterns of genes

To determine the chromatin structure, we performed the assays for Cleavage Under Targets and Tagmentation sequencing (CUT&Tag) of histone modification, transposase-accessible chromatin sequencing (ATAC-seq), Oligo-capping sequencing (Transcription start site sequencing, TSS-seq), and strand-specific RNA sequencing (stranded RNA-seq) in brain tissues of gregarious locusts. We performed CUT&Tag sequencing for a panel of 13 histone modifications, including histone H3 lysine 9 acetylation (H3K9ac), H3K14ac, H3K18ac, H3K27ac, histone H3 lysine 27 monomethylation (H3K27me1), H3K27me2, H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K9me3 and H4K20me3. We visualized chromatin signal distribution along the gene

body and flanking regions for two categories of gene expression levels: low and high (Supplementary Figures 7). We found the strength of chromatin signals is linked to gene expression levels. At TSSs, H3K4me3, H3K4me2, and H3K9ac are the most abundant chromatin marks, while H3K36me3, H3K27me2, and H3K27me3 are the least abundant. Permissive chromatin marks such as H3K4me3, H3K27ac, and H3K9ac exhibit stronger marking at the TSSs of highly expressed genes compared to low-expressed genes. We detected the reduced levels of H3K27me2 and H3K27me3 at the TSSs of highly expressed genes in contrast to low-expressed genes. Thus, the signals of the repressive histone marks at the TSSs are not artifacts of accessible site signals in CUT&Tag experiments. Additionally, we observed a significant rise in H3K9me3, H3K27me2, and H4K20me3 downstream of transcription termination sites in low-expressed genes, indicating their involvement in gene silencing spread. Although the instructive or consequential nature of chromatin marks remains a matter of debate, we demonstrated that the abundance of histone modification is associated with gene expression levels in locusts.

Supplementary Note 3: Chromatin functional organization of the locust genome

The chromatin states E3 and E4 are characterized by a high frequency of H3K4me3 and a high depletion of H3K36me3, indicating their promoter-associated roles in permissive chromatin states. The major association of the promoter-associated chromatin states with the Oligo-capping TSSs indicated a good concordance between the chromatin-based and the TSS-based methods in promoter identification (Supplementary Figure 8). However, E7 (bivalent enhancer that is occupied by both permissive and repressive chromatin marks) were found dominantly in the intergenic region. The chromatin state E1 and E12 is predominantly marked by H3K36me3 and H3K27me1 deposition, respectively. They show remarkable depletion in the intergenic region, in consistent with its transcription-associated role in gene-body demarcation. The three chromatin states (E8, E10 and E11), which are occupied by repressive chromatin marks, are linked with establishment and maintenance of heterochromatin. The chromatin state E8 characterized by H3K27me3 deposition is mainly located within intergenic region, supporting its role in establishment of facultative (reversible) heterochromatin. H3K9me3, H3K27me3 and H4K20me3 are the three distinct chromatin marks associated with transcriptional repression. In all the large and medium chromosomes, we found the highly enriched H3K9me3 concentrations at one of the chromosome ends, while H3K27me3 and H4K20me3 signals are discontinuous and distributed

throughout the chromosome (Supplementary Figures 9-11). Centromeric and surrounding pericentromeric regions, which are clustered into constitutive heterochromatin foci enriched for repressive mark H3K9me3, are composed of large stretches of satellite DNA (Figure 1e). Therefore, chromatin states reflect functional organization of the locust genome.

Supplementary Note 4: TE contribution to chromatin states

Because TEs are dominantly comprised of the locust genome, we explored the TE contribution to the chromatin states. The permissive and repressed chromatin states have a considerable portion of TEs (Supplementary Figure 14), indicating that TEs contribute to host regulatory innovation and are subject to be epigenetic impacted. TEs comprise 34.84%, 27.82%, 45.64% and 53.11% of the chromatin states E3, E4, E5 and E7, respectively, indicating a more TE enrollment in enhancers (E5 and E7) than in promoters (E3 and E4) in the locust. Compared to active regulatory chromatin states, TEs are more overrepresented in the heterochromatin-associated states (76.39% in E10, 72.59% in E11 and 71.78% in E12). Notably, non-LTR retrotransposons exhibited a significantly higher level of involvement in the establishment of heterochromatin-associated states compared to other regulatory states ($P = 0.016$, Wilcoxon rank sum exact test; Supplementary Figure 14). Moreover, the composition of TE subfamilies within chromatin states does not correspond proportionally to their representation in the locust genome (Supplementary Figure 15).

Supplementary Note 5: SXL-independent initiation of dosage compensation

In the fruit fly, the female-specific expression of Sex-lethal (SXL) prevents *msl-2* mRNA from being translated into MSL2 protein through posttranscriptional mechanisms, thereby preventing the assembly of the Male-Specific Lethal (MSL) complex in females ². In contrast, the absence of SXL expression in males allows MSL2 protein expression and subsequent MSL complex assembly, leading to H4K16ac enrichment on the X chromosome to achieve dosage compensation. We compared the RNA expression levels of *sxl* and the core MSL complex subunits (MSL1, MSL2, MSL3, MOF and MLE), and found no significant differences in the expression levels of these genes between females and males (Supplementary Figure 23). In the fruit fly, the proteins MSL1, MSL2, and MSL3, but not MOF and MLF, show sex-specific expression ³. Therefore, we further compared the protein expression of these genes using 4D-DIA (four-dimensional data-independent acquisition) quantitative proteomic analysis (Supplementary Figure 24). We failed to detect the

protein expression of *msl-2* in both sexes, possibly due to its low expression level. However, we found the protein expressions of *sxl*, *msl-1* and *msl-3* in both females and males, differing from the female-specific SXL-MSL axis in the fruit fly ^{3,4}. *Msl-3*, shows high affinity to H4K20me1, suggesting an involvement of H4K20me1 in twofold upregulation of the male X chromosome in the fruit fly ². In fact, the role of H4K20me1 in gene regulation is still controversial due to its precise role in both gene activation and repression in varying contexts. H4K20me1 is enriched in the inactive X chromosome of female humans and in both X chromosomes of worm hermaphrodites (*Caenorhabditis elegans*), pointing to its specific role in X-linked gene expression silencing ⁵. Anyhow, the dosage-compensated X chromosome in locusts is not regulated by SXL and is devoid of H4K20me1 signals, indicating an SXL-independent initiation and a distinct repressive role of H4K20me1 in dosage compensation in X0 sex-determining insects.

Supplementary Note 6: Gene configurations of the X chromosome of insects

Gene configurations of the X chromosome among different insect orders persist over long periods ⁶⁻⁸. To examine the X-linked gene conservation of locusts with other insects, we identified 10,177 1-to-1 orthologs between the locust and other six insect species. Assessment of chromosome conservation by the orthologs of Muller elements in fruit fly revealed large-scale gene movement among locust chromosomes, suggesting the extensive genomic rearrangement and reshuffling of the X chromosome between the fruit fly and the locust (Supplementary Figure 25). To investigate whether sexual-benefit selection played a role in preserving X-linked genes on the locust X chromosome, we compared the ortholog repertoire of X chromosomes in insects from Orthoptera, Hemiptera, Coleoptera, and Diptera. Apart from the fruit fly, the locust and other five insect species, possessed a substantial number of X-linked genes (ranging from 338 to 2,489) and X-linked orthologs (ranging from 208 to 1,051). However, only 10 X-linked orthologs were shared among these insects (Supplementary Figure 26). Furthermore, we did not detect any large syntenic region of Muller elements among the X chromosomes of these seven insect species. We excluded the fruit fly from this analysis because its ancestral X chromosome, Muller element F, reverted to an autosome ⁹. Among the rest six insect species, only 19 X-linked orthologs were shared (Supplementary Figure 27). Recursively removing other insect species also revealed insufficient conservation of X-linked orthologs, as indicated by shared ortholog repertoires (26 in

Orthoptera|Hemiptera|Coleoptera and 98 in Orthoptera|Hemiptera). The scarcity of shared X-linked orthologs across four insect orders suggested the current gene configuration of the locust X chromosome is a result of extensive inter-chromosomal gene exchanges, rather than selection pressures to preserve a substantial number of sex-benefit genes on the locust X chromosome. Consequently, our results ruled out the long-term conservation of locust X-linked genes over the course of approximately 421 million years in insect evolution¹⁰. However, the previous studies have shown that the gene configurations of the X chromosome across different insect orders persist over long periods⁶⁻⁸. Our results are different from previous studies, possibly due to the inclusion of longer evolutionary timeframes in the present study.

To further investigate the formation of X-linked genes of the migratory locust within Orthoptera, we generated a chromosome-level genome of the desert locust (*Schistocerca gregaria*, SG, Caelifera, Acrididae) using Nanopore long-read sequencing (890.01 Gb) and Illumina data (745.60 Gb) with Hi-C scaffolding (739.23 Gb). The desert locust genome, with a contig N50 of 7.74 Mb and a scaffold N50 of 8,695 Mb, has 12 large chromosomes, covering 91.41% of the total genome assembly.

Supplementary Note 7: Relocation of female-related genes

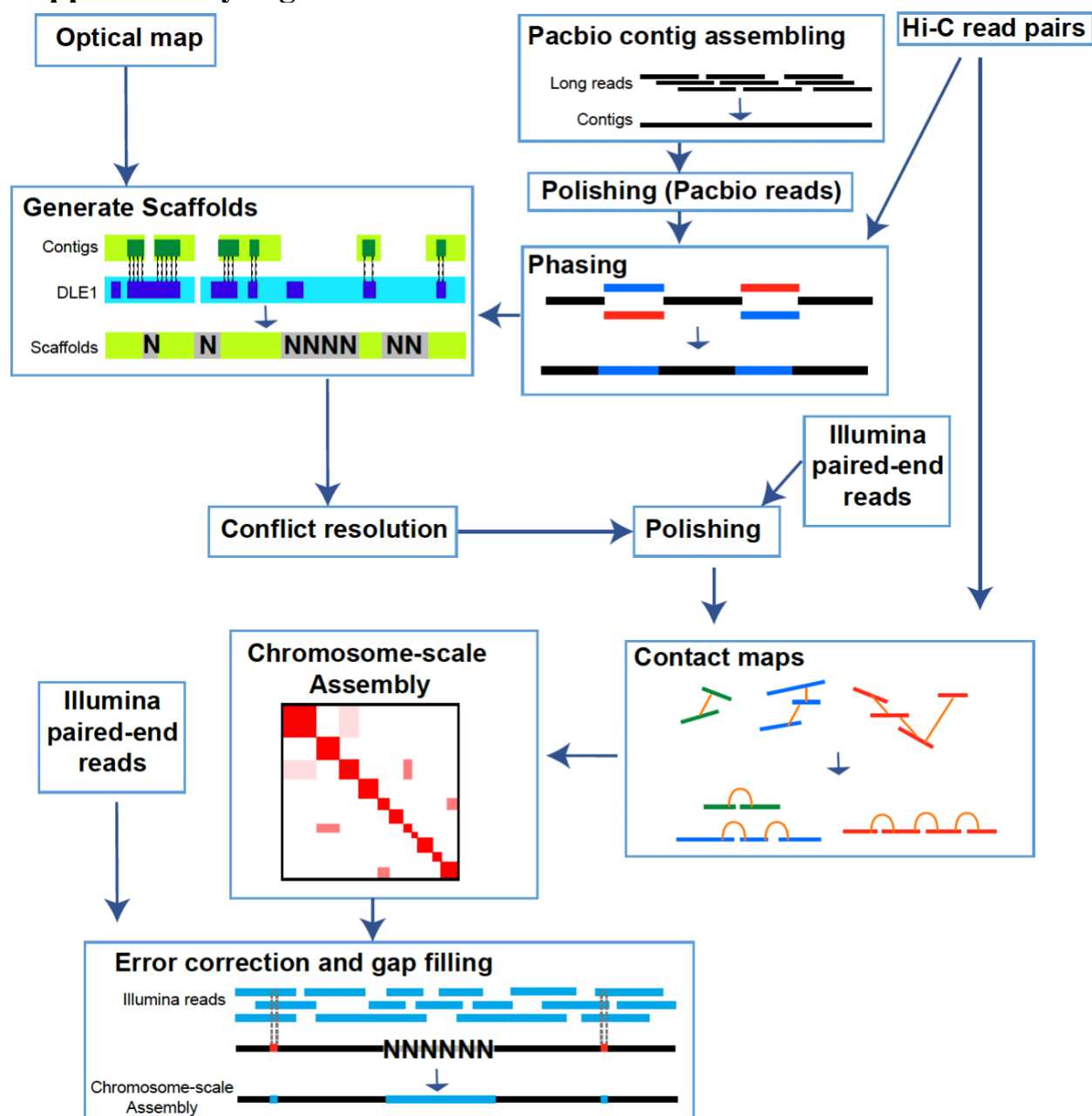
In the recent AtoX category, several genes were previously reported to be associated with functions that harm males or benefit females. For example, in the absence of testosterone, the gene encoding ornithine aminotransferase (*Oat*) shows higher expression in females compared to males, facilitating the efficient conversion of L-ornithine into L-glutamate¹¹. Additionally, *msps* promotes oocyte growth and cell fate maintenance by enhancing dynein-dependent nurse cell-to-oocyte transport during oogenesis¹². Both these two genes show similar RNA expression in the brains of females and males. However, compared to their expression levels in ovaries, these two genes exhibited significant depletion in RNA expression in testes. Therefore, sexually antagonistic selection promotes the relocation of female-related genes from autosomes to X chromosome in locusts. Consequently, these genes can be repressed by meiotic sex chromosome inactivation, facilitating to mitigate female antagonism and to increase male fitness during spermatogenesis.

Supplementary Note 8: Selection of histone modifications

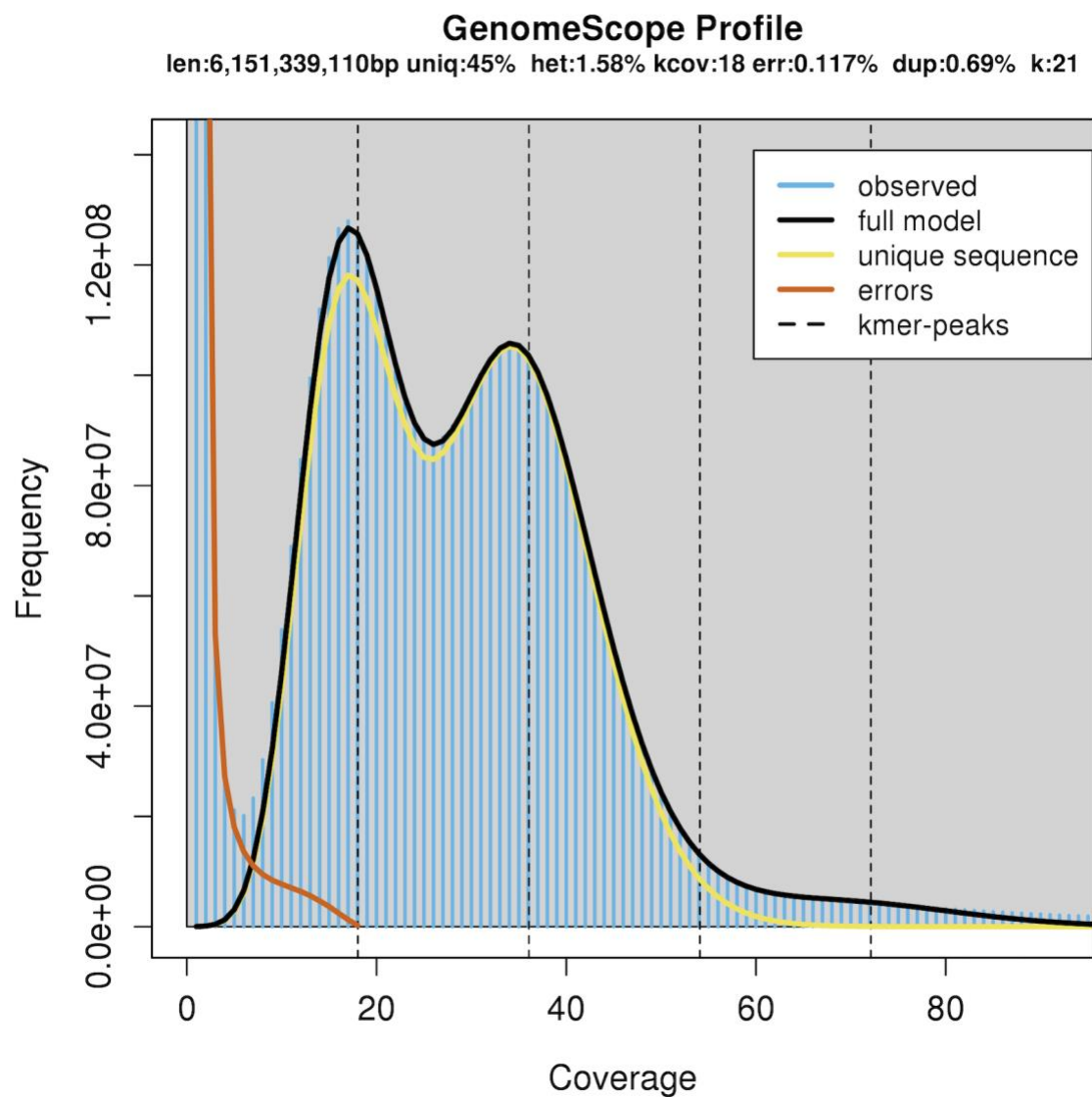
Histone modification, of which specific types mark distinct genomic regions and regulatory elements, is an essential component of chromatin structure^{13,14}. H3K9ac is associated with active gene promoters. Similarly, H3K14ac is found at active gene promoters and enhancers. H3K18ac

marks active gene promoters and tissue-specific enhancers, supporting tissue-specific gene expression. H3K27ac is a hallmark of active enhancers and promoters, distinguishing them from poised or inactive states. Conversely, H3K27me1 is predominantly found within the gene bodies of active genes, while H3K27me2 occupies intergenic or intragenic regions, helping to inactivate enhancers. H3K27me3, on the other hand, marks facultative heterochromatin and is enriched at promoters and within gene bodies of silent genes. H3K36me3 is deposited within gene bodies of actively transcribed genes and is associated with transcription elongation. For enhancer regions, H3K4me1 is characteristic of both primed and active enhancers, while H3K4me2 is found at active promoters and enhancers. H3K4me3 is a hallmark of active gene promoters, directly linked to transcriptional initiation. In contrast, H3K9me3 is a mark of constitutive heterochromatin, typically associated with transcriptionally inactive regions and the silencing of transposable elements (TEs) and their derived elements. H4K20me3, localizing primarily to telomeric regions, is involved in heterochromatin maintenance. Therefore, we used these 13 histone modifications to comprehensively decipher the chromatin structure in the brains of gregarious locusts. To compare histone modification profiles between the two phases of locusts, we performed the CUT&Tag assay for nine histone modifications, including H3K4me1, H3K4me3, and H3K27ac (which are strongly associated with active gene expression), as well as H3K9me3, H3K27me2, and H3K27me1 (which are strongly linked to the silencing of gene expression and TE-derived elements), and H3K36me3 (which is associated with transcription elongation) in the brains of solitary locusts. We did not include all 13 histone modifications in the phase comparison of locusts because some are unrelated to gene regulation or represent an intermediate phase between two specific types of histone modification. Different histone modifications are involved in various modes of dosage compensation. H4K20me1, H4K16ac, H3K27me3, and H3K9me3 are associated with dosage compensation in different animals. All four of these modifications have been reported to be associated with dosage compensation, even though they may serve functional roles in some species but not in others. We profiled the genomic distributions of these four histone modifications to determine whether these chromatin modifications are involved in dosage compensation in locusts.

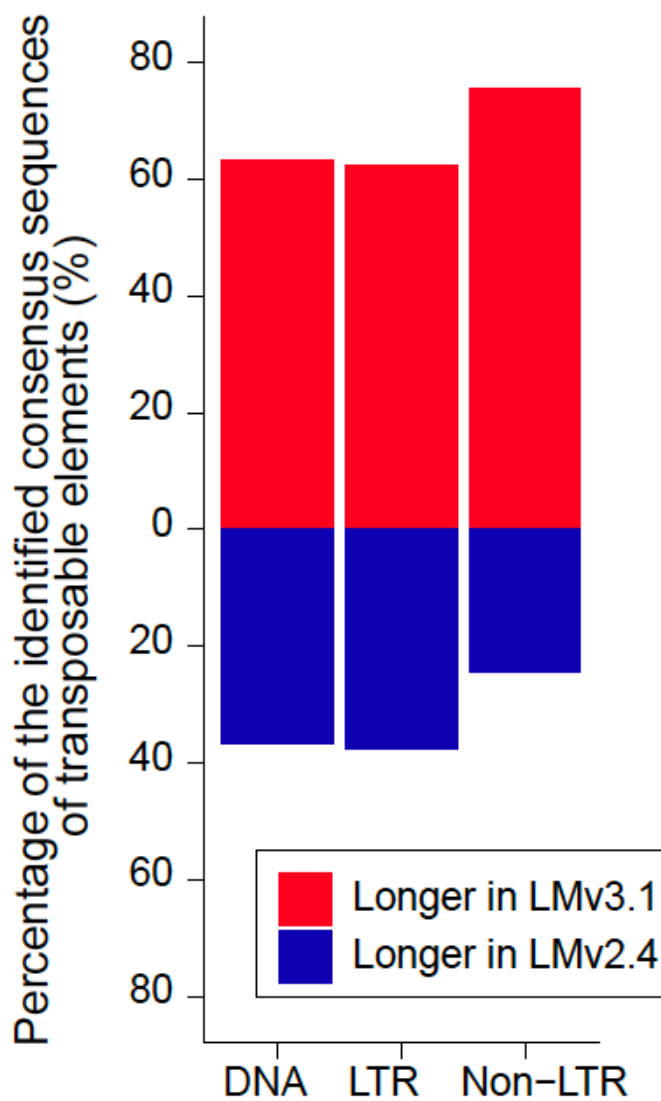
Supplementary Figures



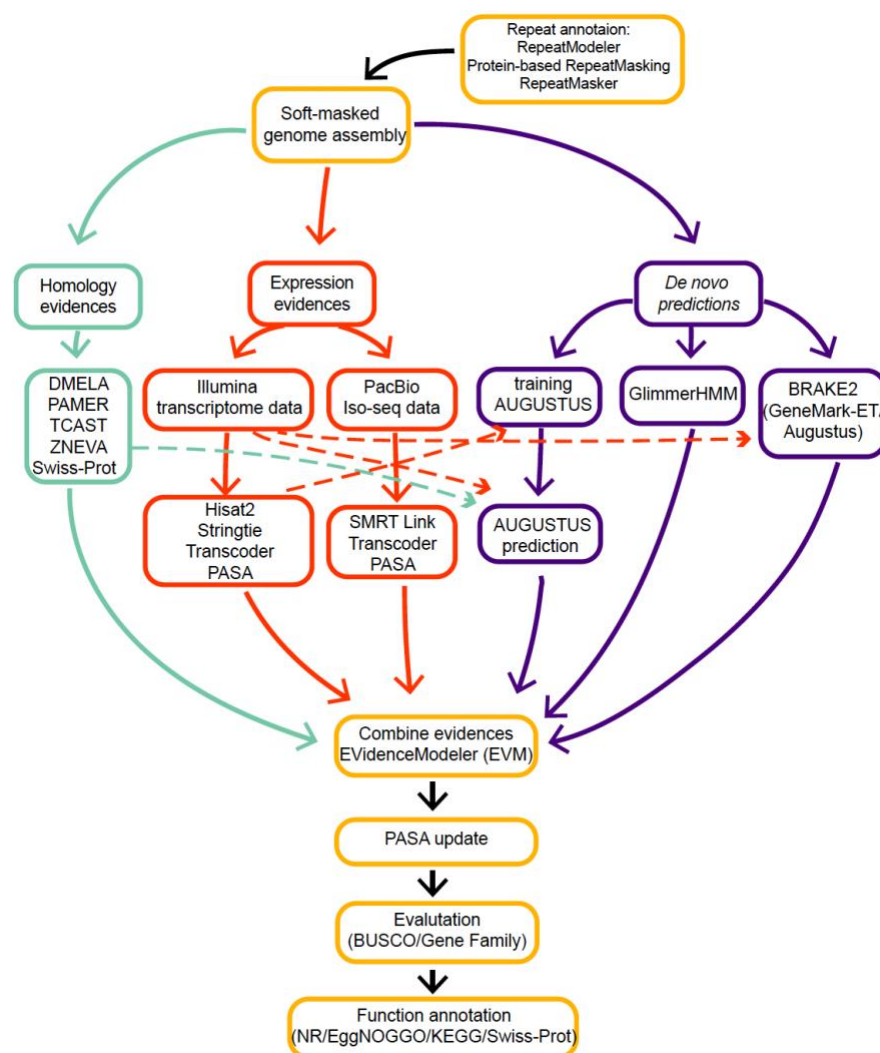
Supplementary Figure 1. Genome assembly of the migratory locust. The chromosome-scale genome assembly of the migratory locust was constructed in a stepwise manner using a combination of PacBio, Bionano, and Hi-C data.



Supplementary Figure 2. Estimation of genome size and heterozygosity rate for the migratory locust using Illumina data.

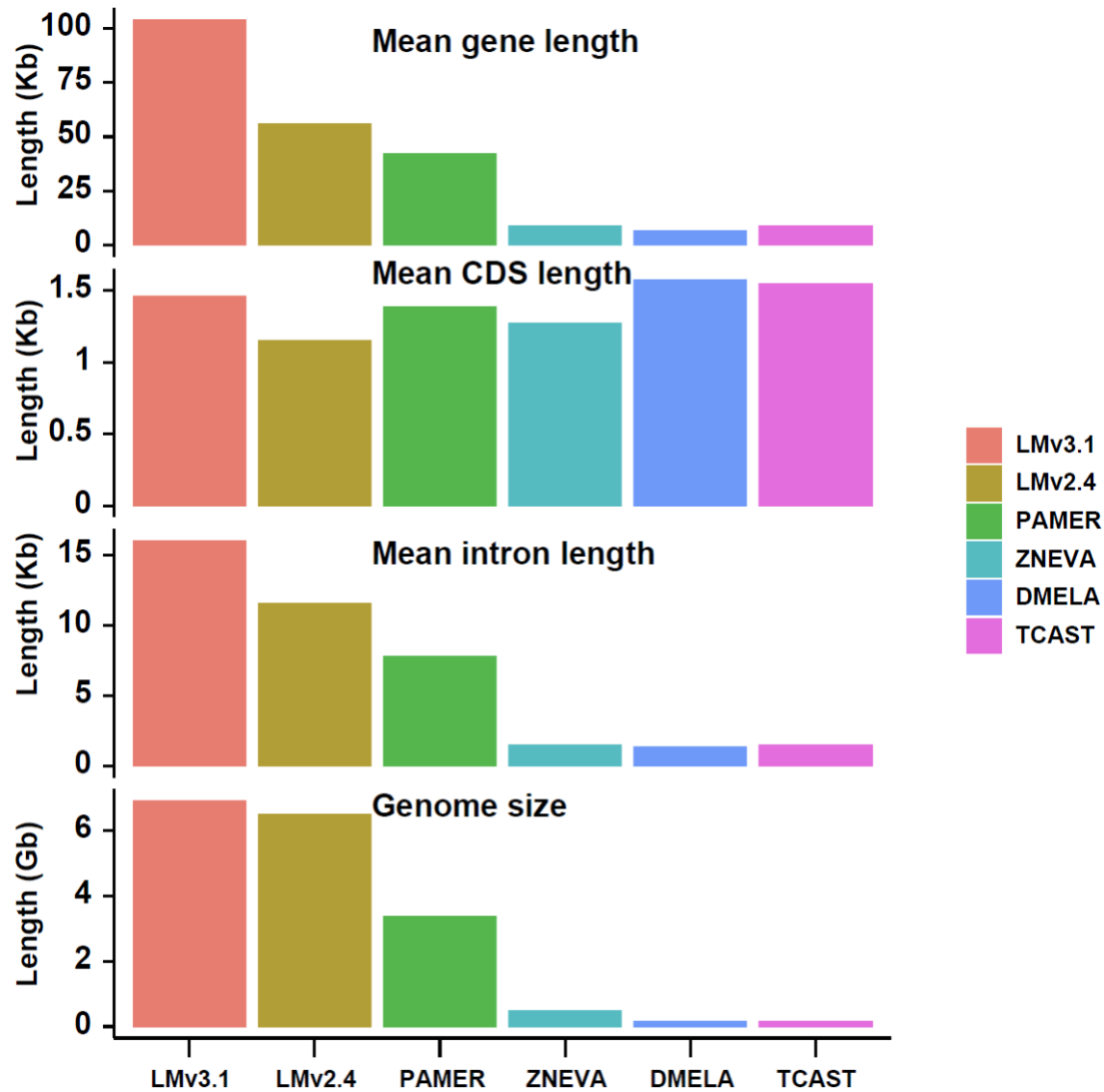


Supplementary Figure 3. Length comparison of consensus sequences of TE elements in the LMv3.1 and LMv2.4 assemblies.

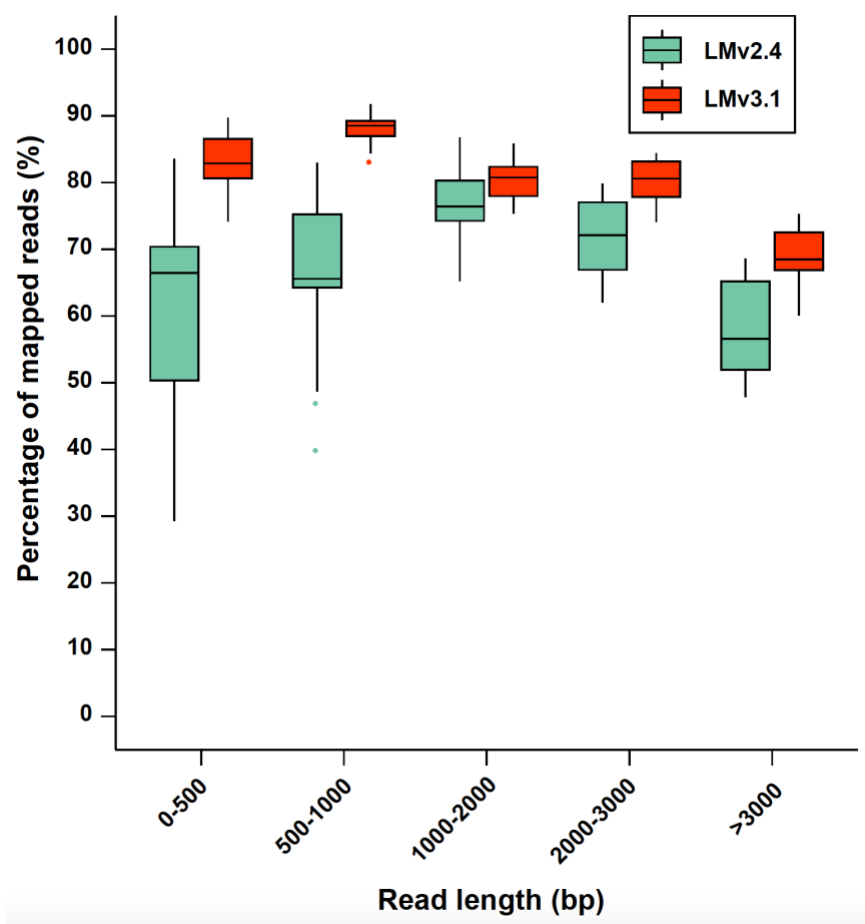


Supplementary Figure 4. Protein-coding gene prediction in the migratory locust genome using a combination of homology evidences, expression evidences and *De novo* predictions. For homology-based prediction, Exonerate was used to improve gene models derived from the protein sets of five insects, including *Drosophila melanogaster* (DMELA), *Periplaneta americana* (PAMER), *Tribolium castaneum* (TCAST) and *Zootermopsis nevadensis* (ZNEVA). For expression-based predication, the Illumina transcriptome data and PacBio ISO-seq data were used to predict protein-coding regions using Transcoder and to refine gene structures using PASA. For *de novo* prediction, Augustus, GlimmerHMM and BRAKER2 were employed on the repeat-masked genome sequences. Finally, EvidenceModeler was used to merge the three types of

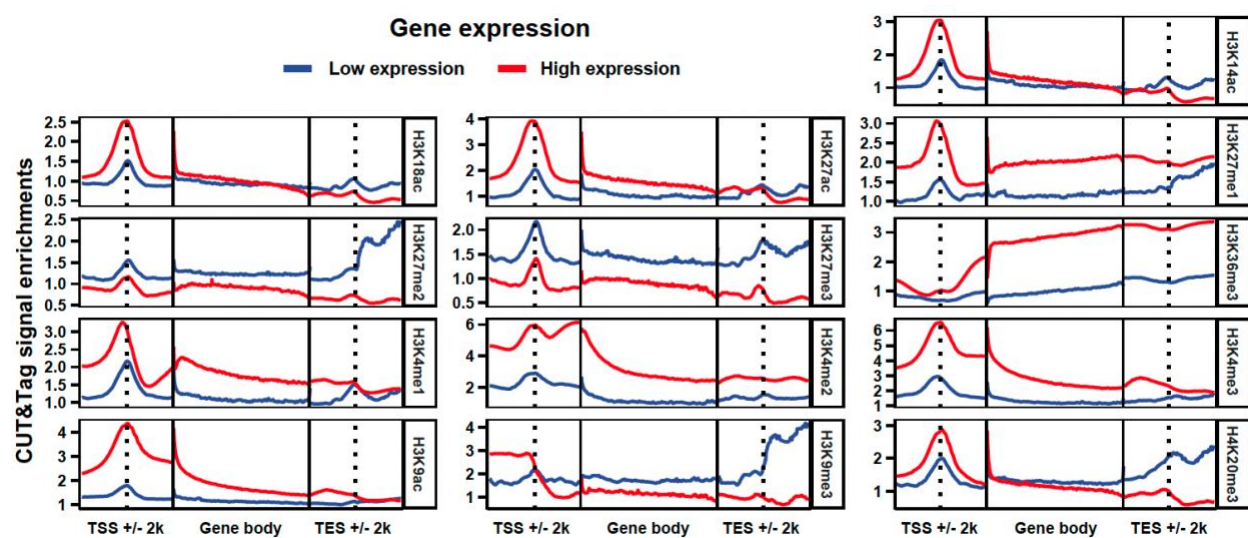
evidences into a comprehensive and non-redundant reference gene set, which was further refined by PASA.



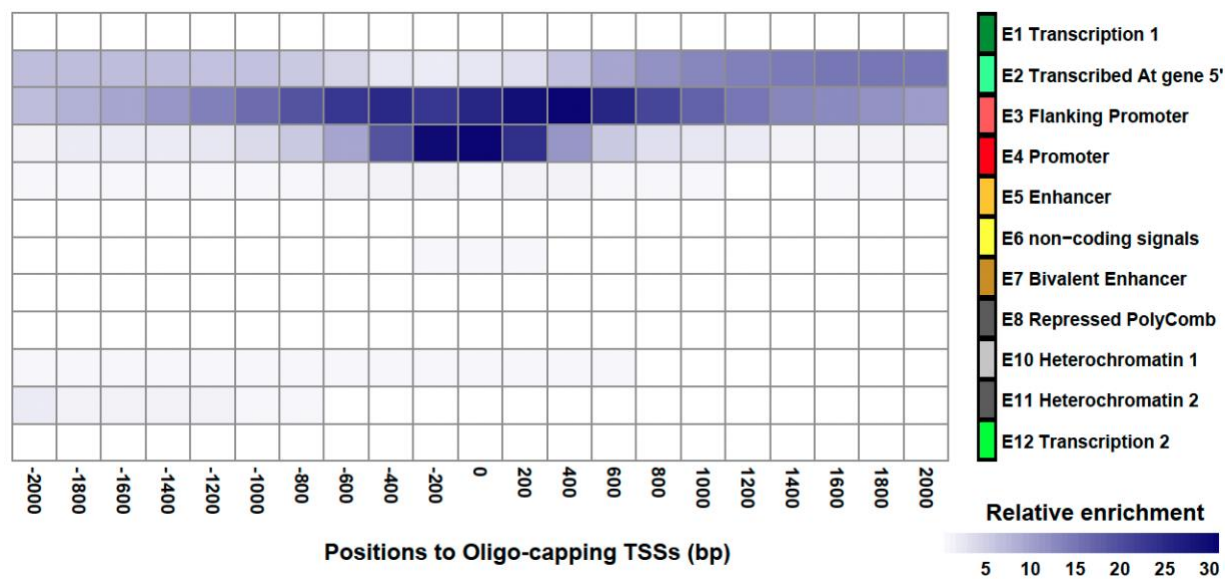
Supplementary Figure 5. Mean gene length, mean CDS length mean intron length and genome size in four insect species. *Drosophila melanogaster* (DMELA); *Periplaneta americana* (PAMER); *Tribolium castaneum* (TCAST); *Zootermopsis nevadensis* (ZNEVA). CDS, coding sequence.



Supplementary Figure 6. Mapping statistics for full-length transcripts using PacBio ISO-seq data. Box plots indicate the median (center line), the first and third quartiles (box limits), and whiskers extending to $1.5 \times$ the interquartile range.



Supplementary Figure 7. Plots of the mean CUT&Tag enrichment signals across genomic loci of the protein-coding genes. The protein-coding genes showing less than 10% and more than 90% quantile of gene expression (Transcripts per million, TPM) were excluded to remove the effects of outlier values. The remaining protein-coding genes were equally divided into two groups based on their TPM values. TSS, transcription start sites; TES, transcription end site.

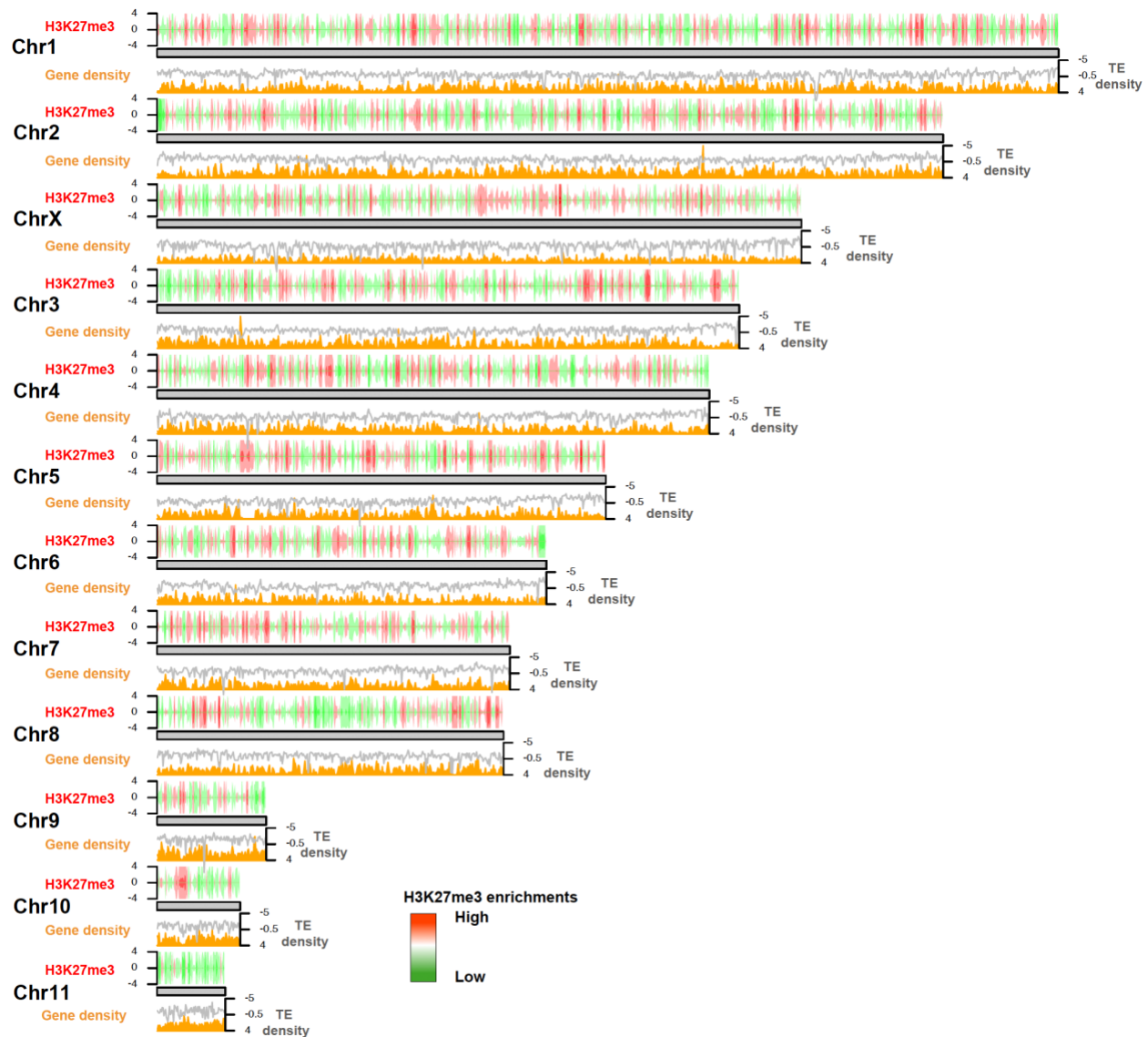


Supplementary Figure 8. Overlapping of the chromatin-based and the TSS-based methods in promoter identification.



Supplementary Figure 9. Enrichment profiles of H3K9me3 in brain tissues of 4-intar gregarious locusts. The log₂-based CUT&Tag signals of H3K9me3 were normalized using a Z score

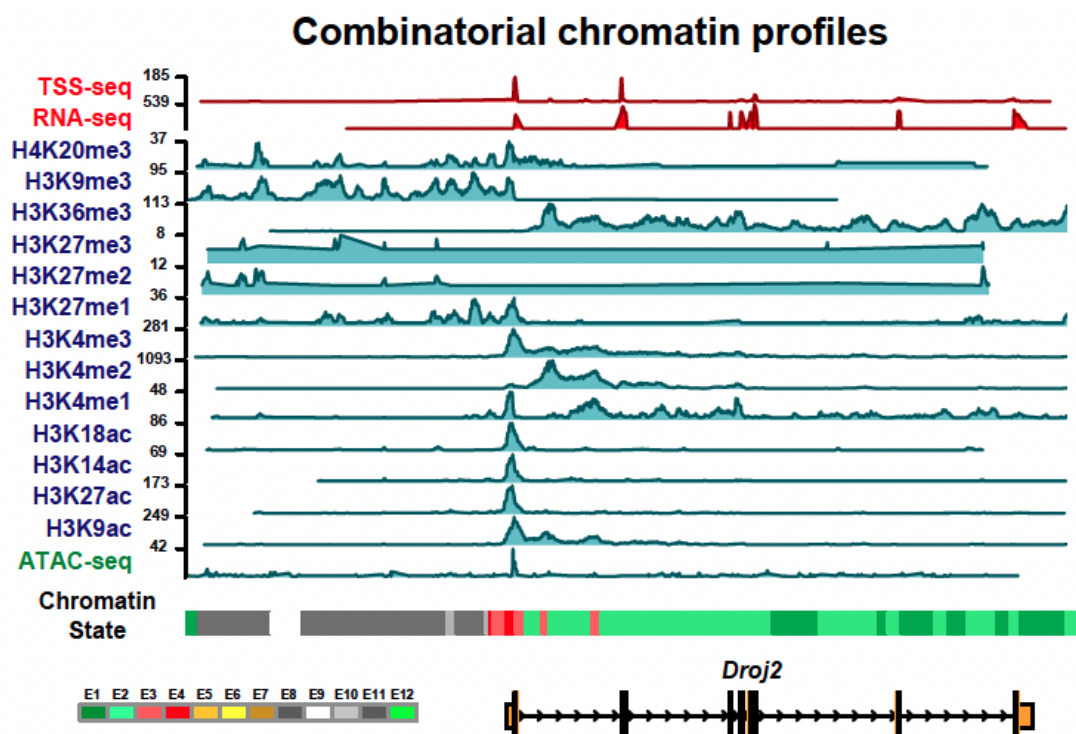
transformation. The chromosomes were divided by partitioning into 1-Mb bins. TE, transposable element.



Supplementary Figure 10. Enrichment profiles of H3K27me3 in brain tissues of 4-intar gregarious locusts. The log₂-based CUT&Tag signals of H3K27me3 were normalized using a Z score transformation. The chromosomes were divided by partitioning into 1-Mb bins. TE, transposable element.

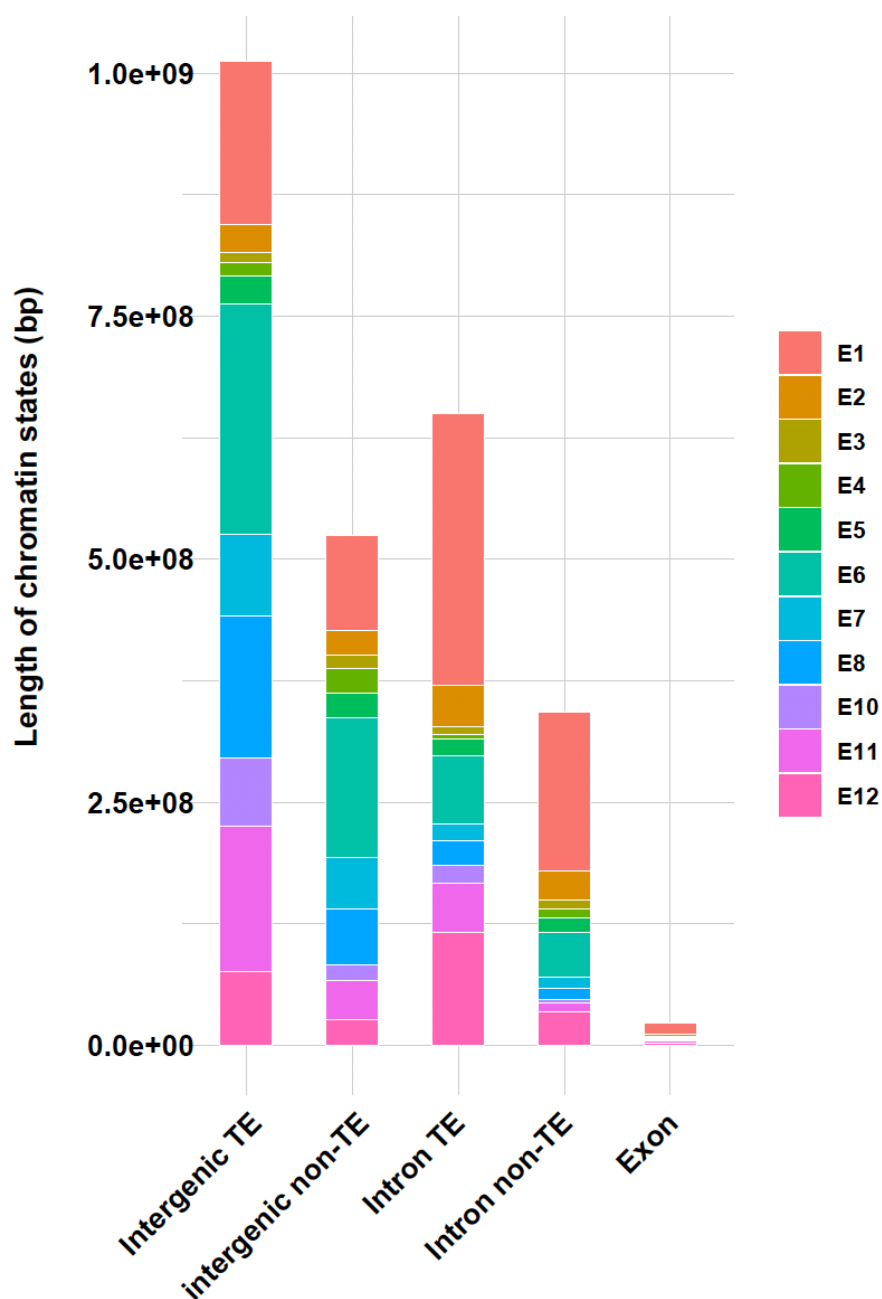


Supplementary Figure 11. Enrichment profiles of H4K20me3 in brain tissues of 4-intar gregarious locusts. The log₂-based CUT&Tag signals of H4K20me3 were normalized using a Z score transformation. The chromosomes were divided by partitioning into 1-Mb bins. TE, transposable element.



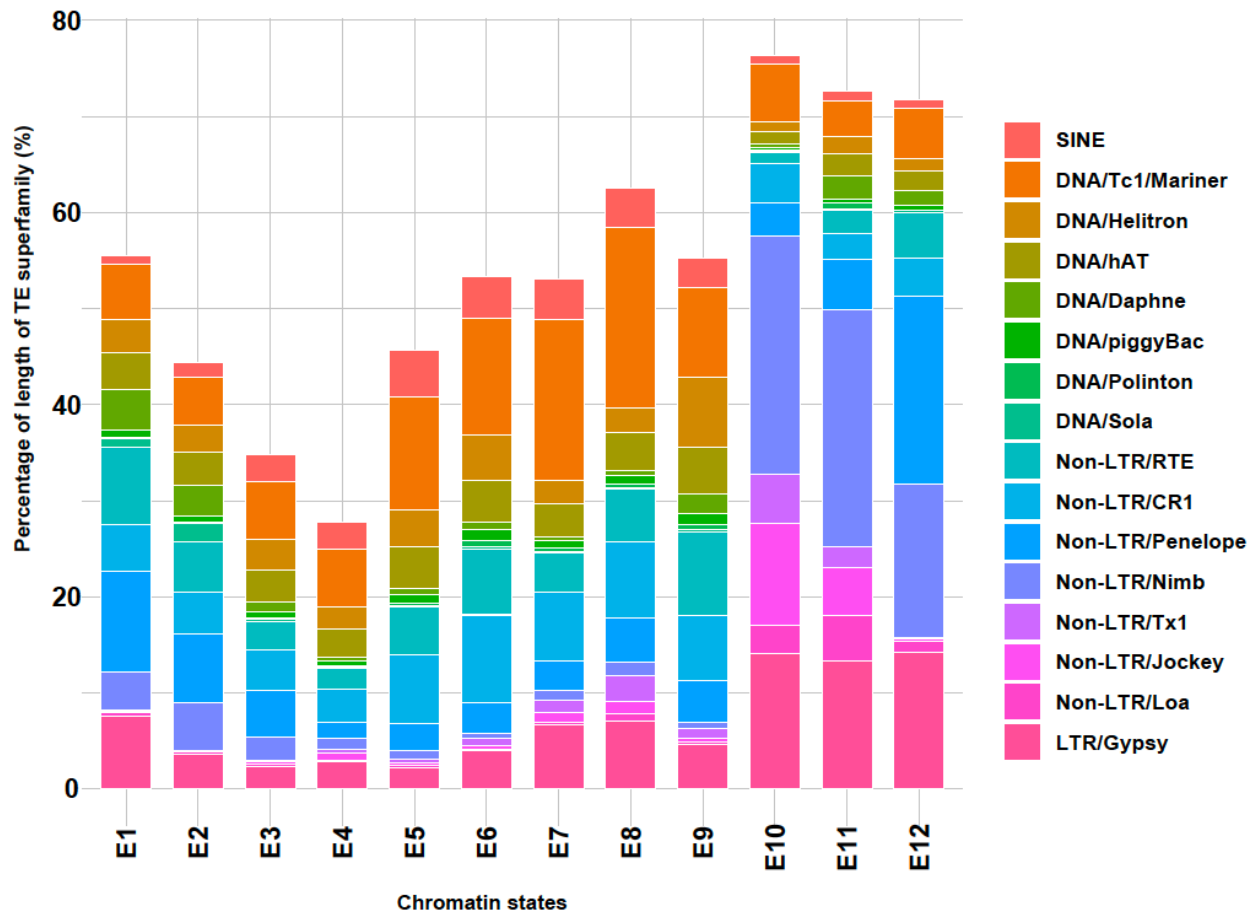
Supplementary Figure 12. Combinatorial chromatin signal profiles around the genic region of the protein-coding gene *Droj2*. The predominant read distribution of TSS-seq is found at the first and second exons due to the paired-end sequencing linking these two exons, whereas the read distribution of RNA-seq covers all exons. Consequently, the two transcriptomic tracks highlight the promoter and gene body regions of protein-coding genes. In the ATAC-seq track, the promoter region also shows a high level of read coverage, indicating that the chromatin is accessible to allow the binding of transcriptional initiation machinery. The promoter region exhibits high levels of enrichment for multiple histone modifications (such as H3K27ac, H3K14ac, and H3K4me3) associated with active transcription, and low levels of histone modifications (such as H3K27me2, H3K27me3, and H3K9me3) associated with repressed transcription. Additionally, H3K36me3, which is associated with transcription elongation, is found in the gene body region of transcriptionally active genes. Therefore, the promoter region and its downstream regions can be classified into four chromatin states: E4 (promoter), E3 (flanking promoter), E2 (transcribed at gene 5'), and E1 (transcription). The upstream region of the promoter is decorated with H3K9me3 and H4K20me3, classifying it into two heterochromatin-associated chromatin states (E10 and

E11). As a result, the gene transcription level of genomic loci can be inferred from these chromatin states.

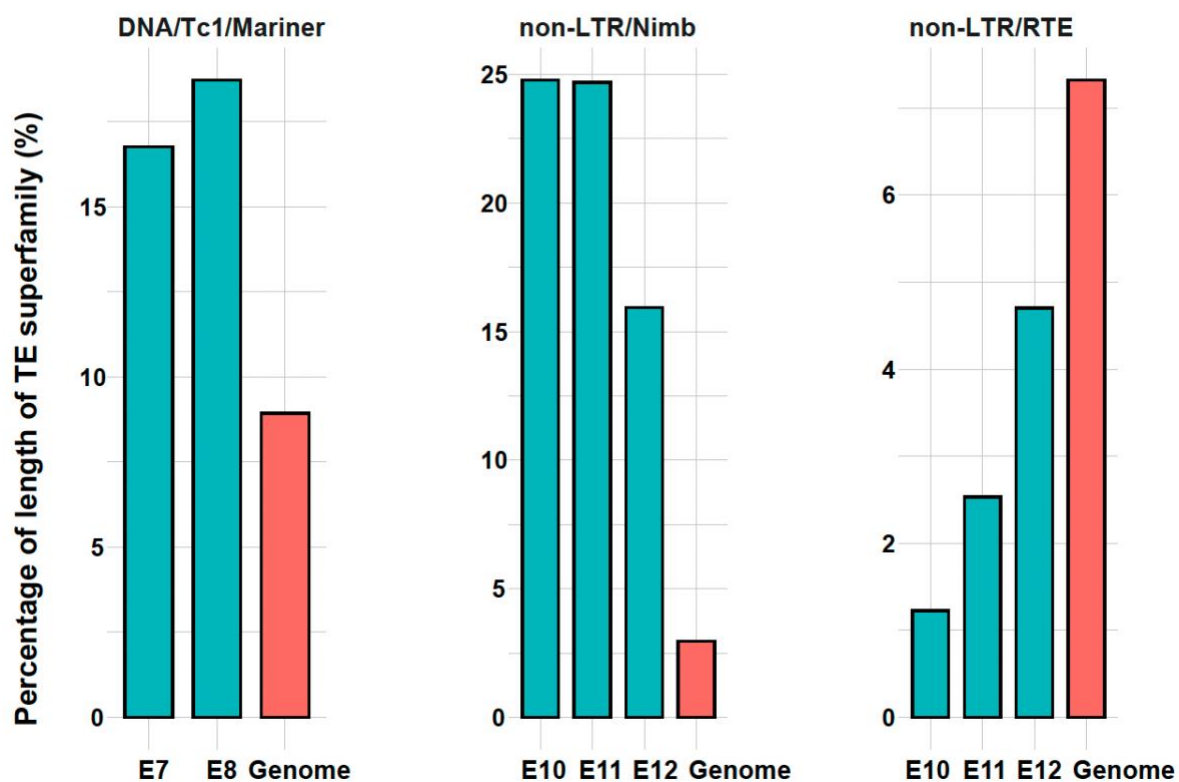


Supplementary Figure 13. Length of chromatin states in genomic regions. The length of the 12 chromatin states, which were defined by ChromHMM, were summed to calculate the total length

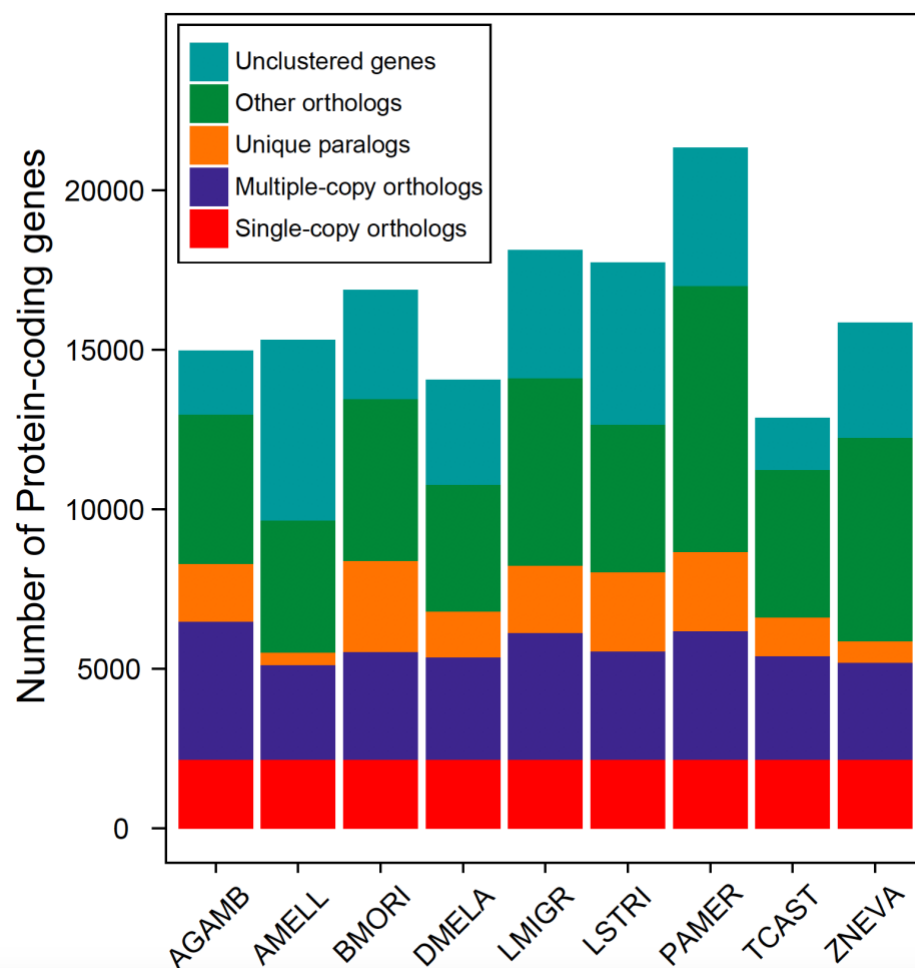
of chromatin states in genomic region. The chromatin state E9, which was associated with quiescent signals, wasn't shown in this figure.



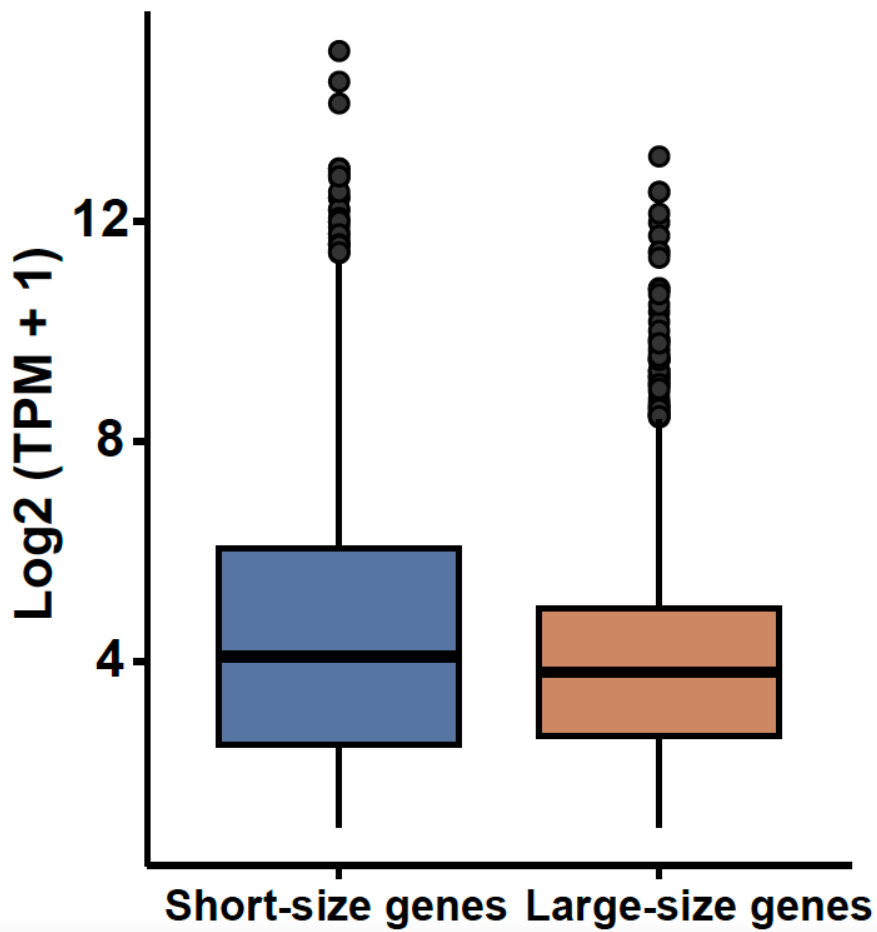
Supplementary Figure 14. Accumulated percentages of length of transposable element (TE) superfamily in chromatin states. The chromatin state E4, which represents promoter-associated state, shows the greatest depletion of transposable elements among the twelve chromatin states.



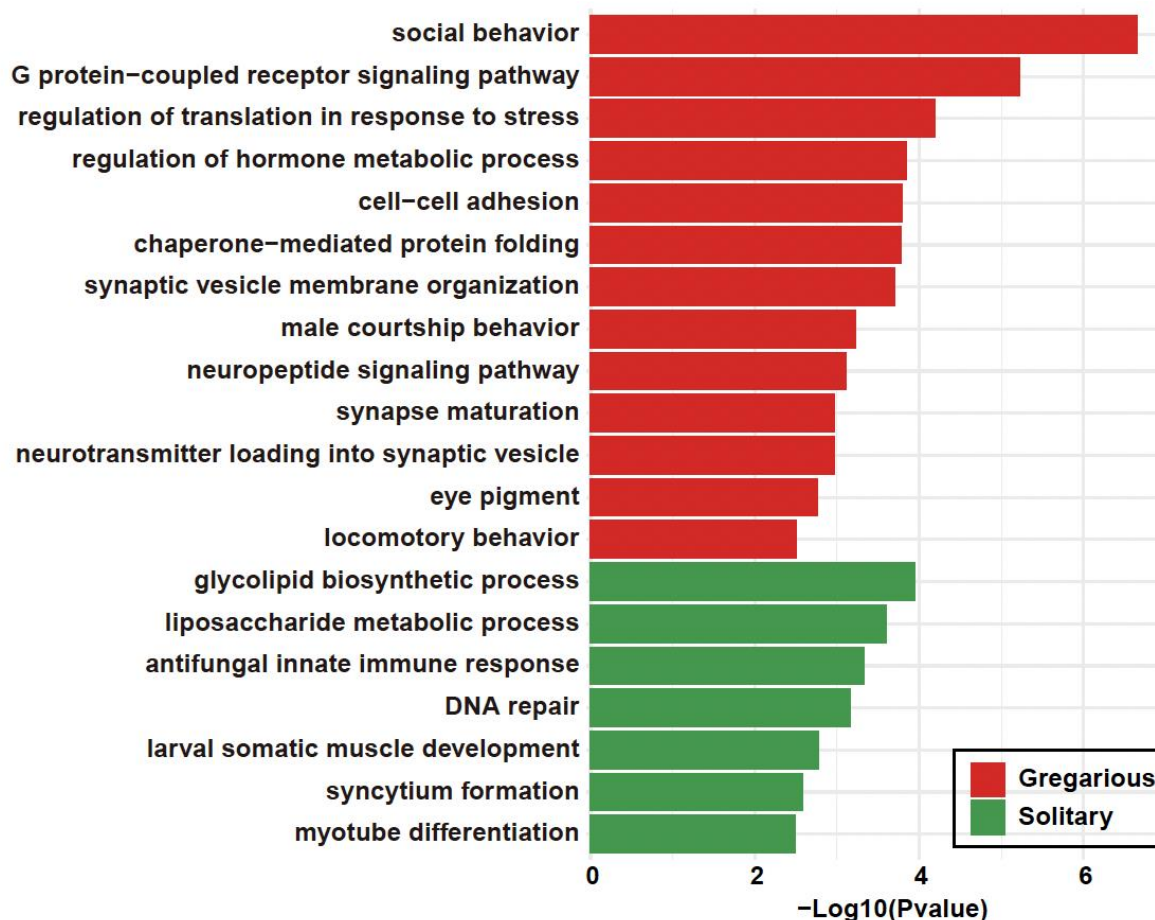
Supplementary Figure 15. Percentages of length of DNA/Tc1/Mariner, non-LTR/Nimb and non-LTR/RTE superfamilies in genomic regions.



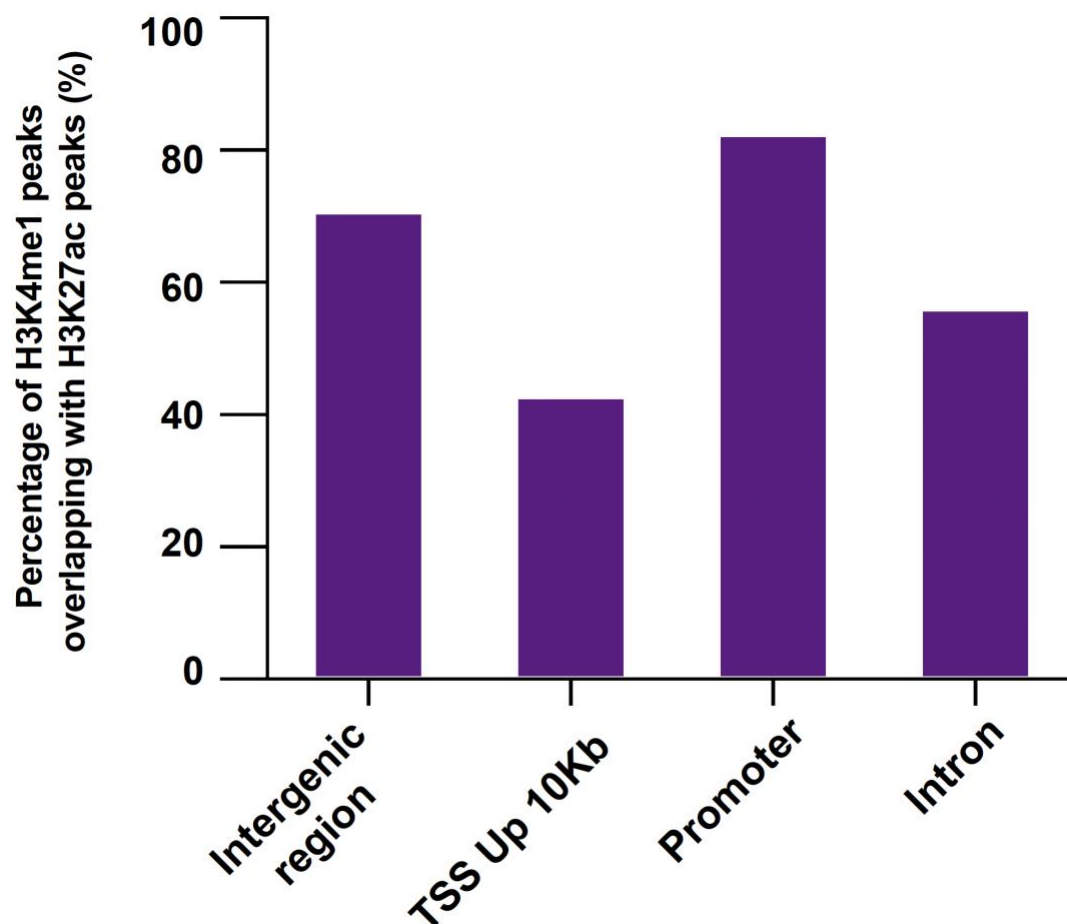
Supplementary Figure 16. Orthology relationships of protein-coding genes in the multiple insects. *Anopheles gambiae*, AGAMB; *Apis mellifera*, AMELL; *Bombyx mori*, BMORI; *Drosophila melanogaster* (DMELA); *Locusta migratoria*, LMIGR; *Laodelphax striatellus*, LSTRI; *Periplaneta americana* (PAMER); *Tribolium castaneum* (TCAST) and *Zootermopsis nevadensis* (ZNEVA).



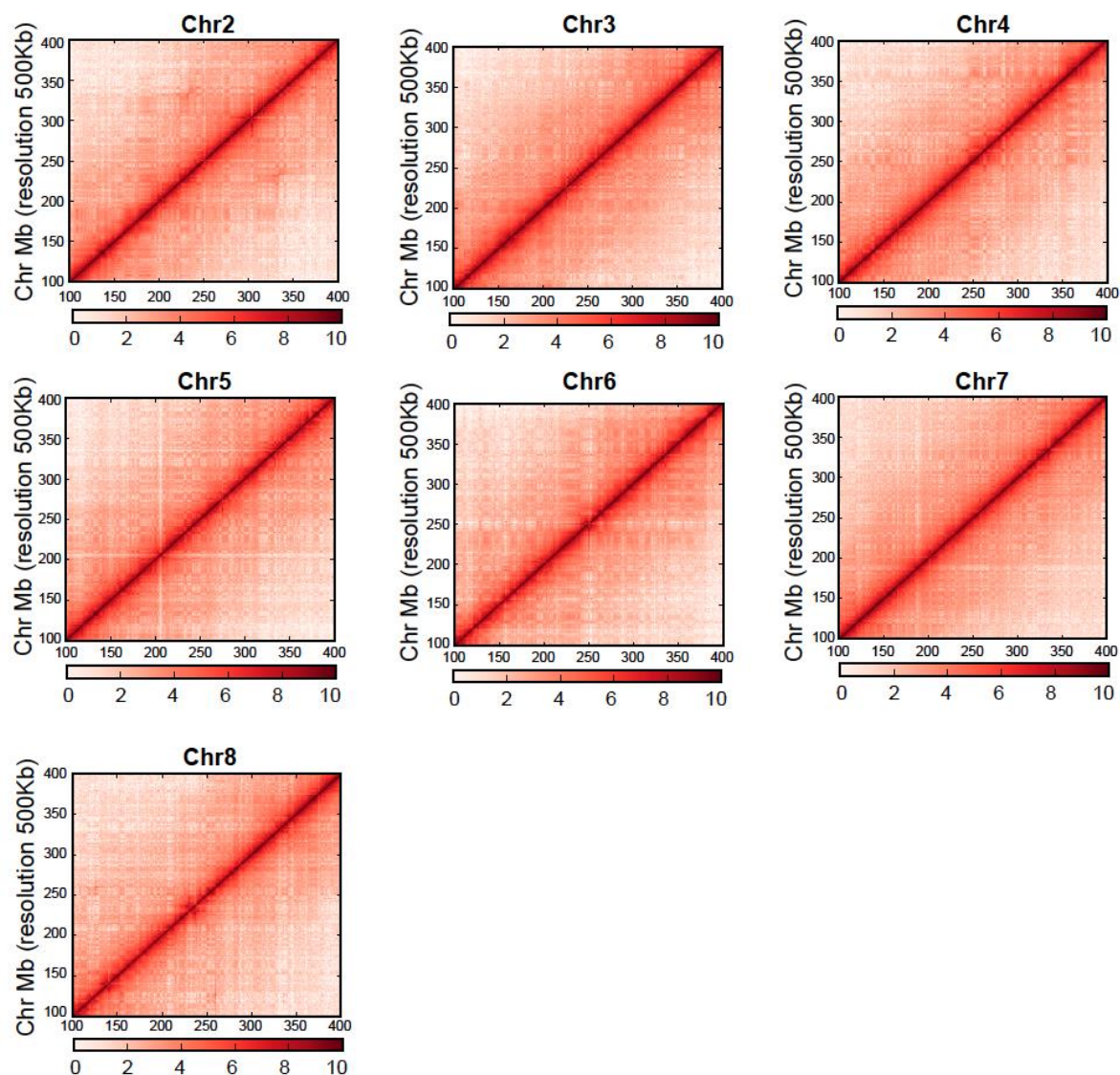
Supplementary Figure 17. Expression values (TPM+1 in log2 scale) in short-size and long-size genes in the brains of locusts. Short-size genes correspond to the lower 25% of the data, while long-size genes correspond to the upper 75% of the data. Box plots indicate the median (center line), the first and third quartiles (box limits), and whiskers extending to $1.5\times$ the interquartile range.



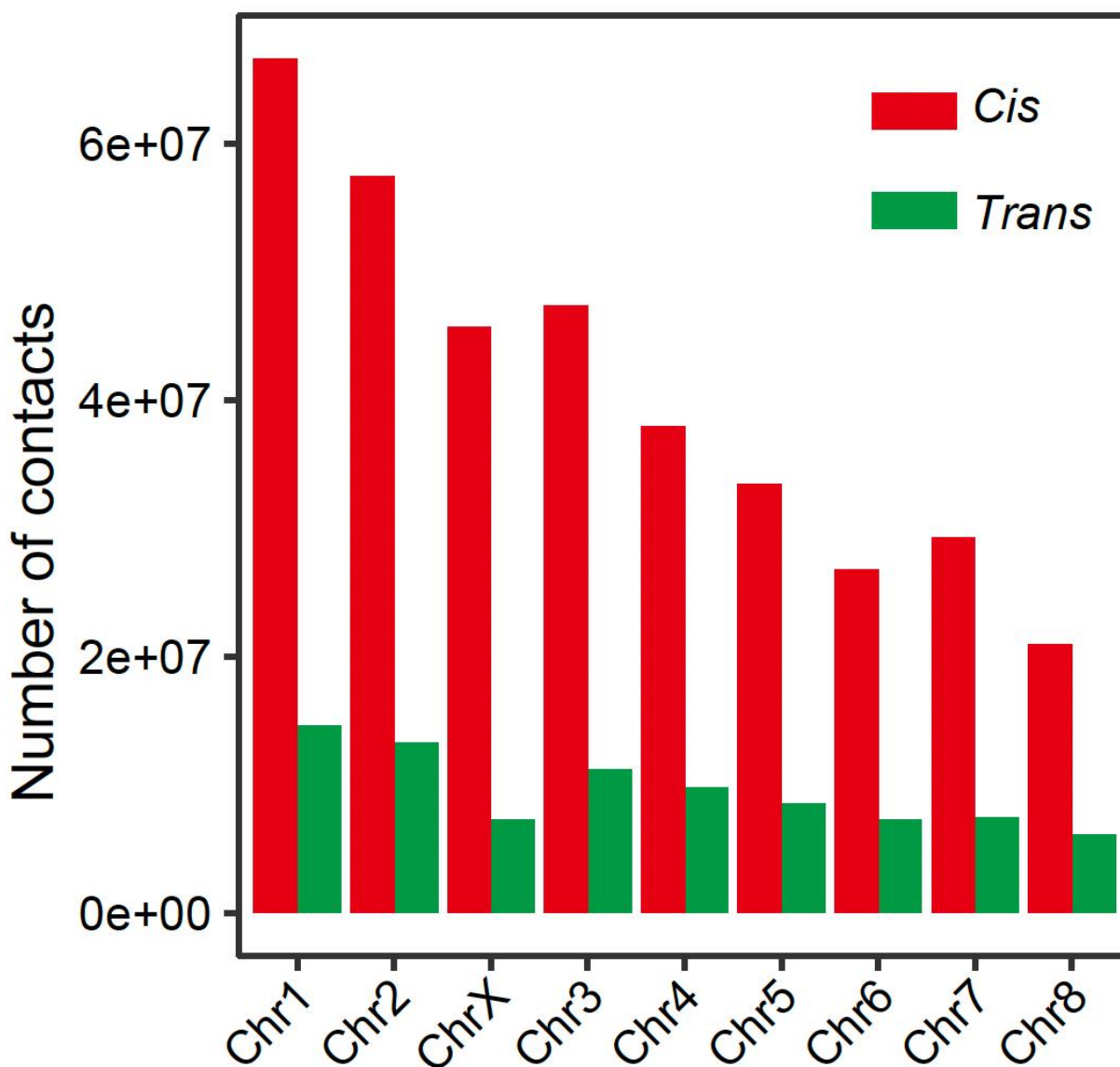
Supplementary Figure 18. Gene ontology enrichment of differentially expressed genes in the brains of solitary and gregarious locusts. The differentially expressed genes were determined using the DESeq2 package in R. Biological process enrichment was determined using the clusterProfiler and GO.db packages in R. Solitary indicates significant up-regulation of genes in solitary locusts; Gregarious indicates significant up-regulation of genes in gregarious locusts.



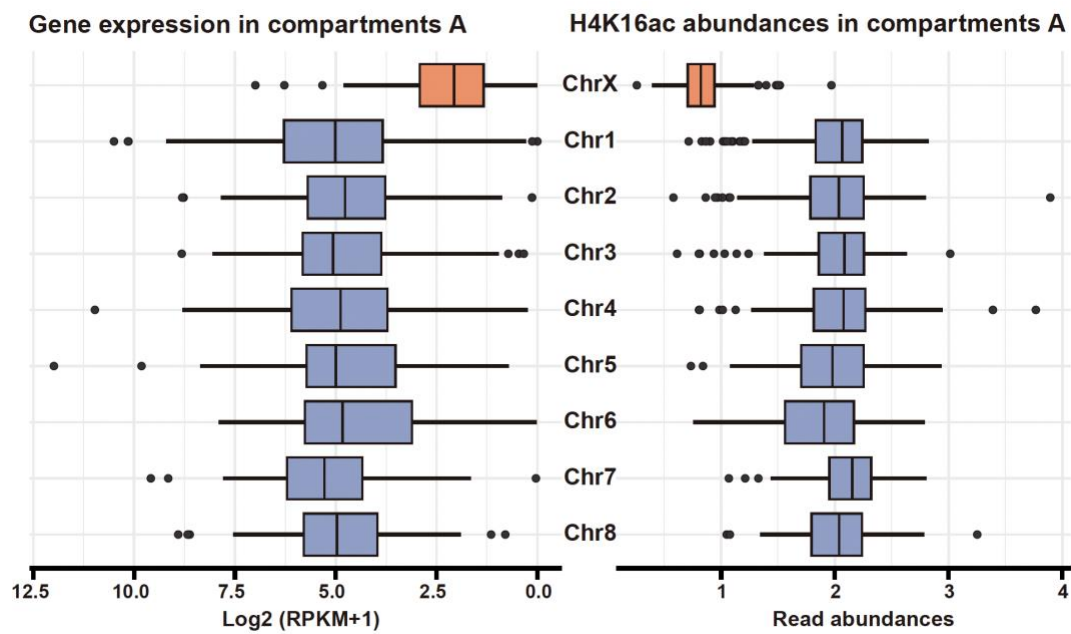
Supplementary Figure 19. The percentages of H3K4me1 peaks that overlap with H3K27ac peaks. The overlaps were determined using the intersect function in BEDTools.



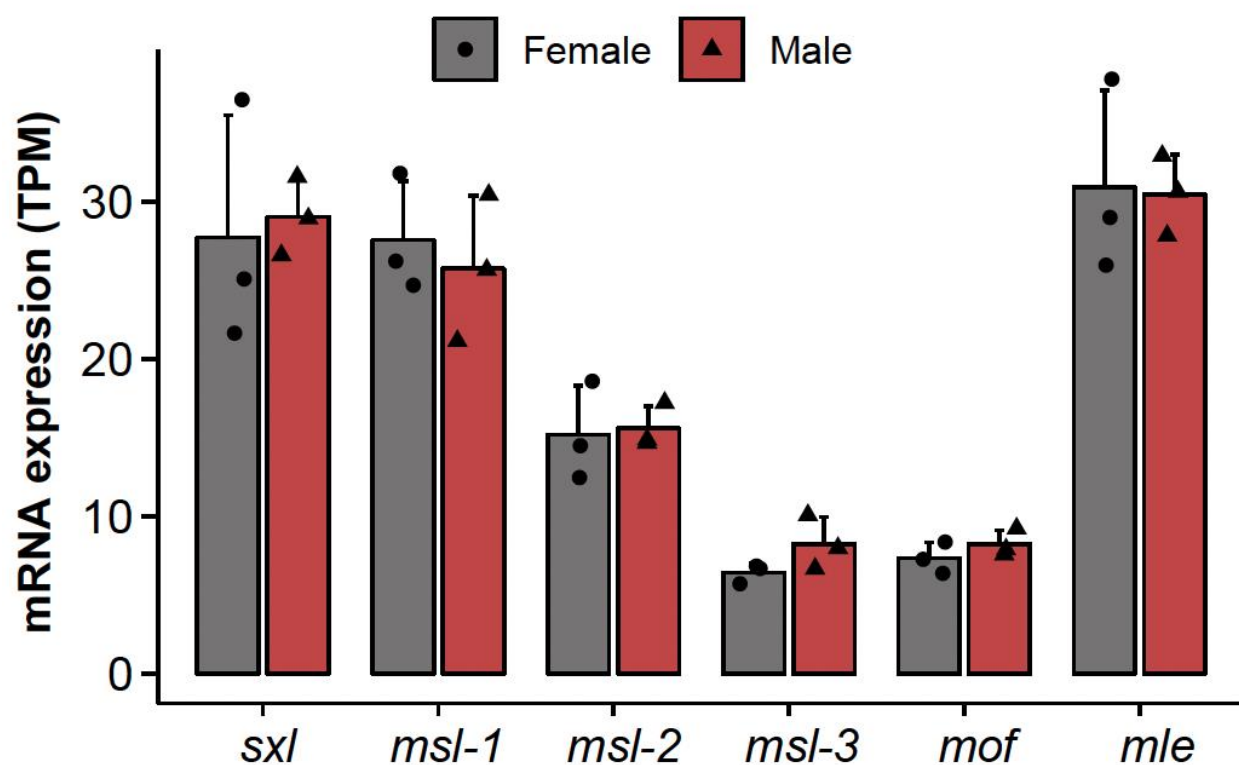
Supplementary Figure 20. Pearson correlations of Hi-C interaction frequency matrices of the chromosomes 2 to 8 at 500 Kb resolution. The intensity of each pixel represents the normalized frequency of interaction between a pair of genomic loci.



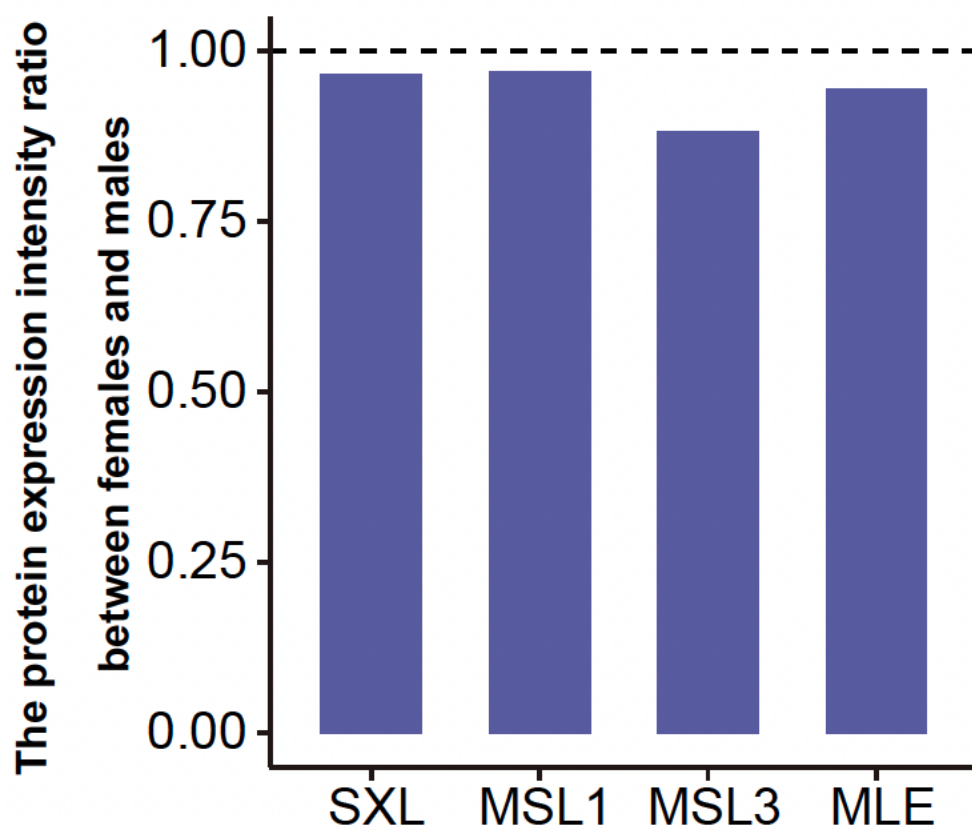
Supplementary Figure 21. Interaction contacts in different chromosomes. All pairwise chromosomal interactions within (intra-chromosomal *cis* contacts) and between (inter-chromosomal *trans* contacts) chromosomes were measured to determine spatial distance between the X chromosome and autosomes.



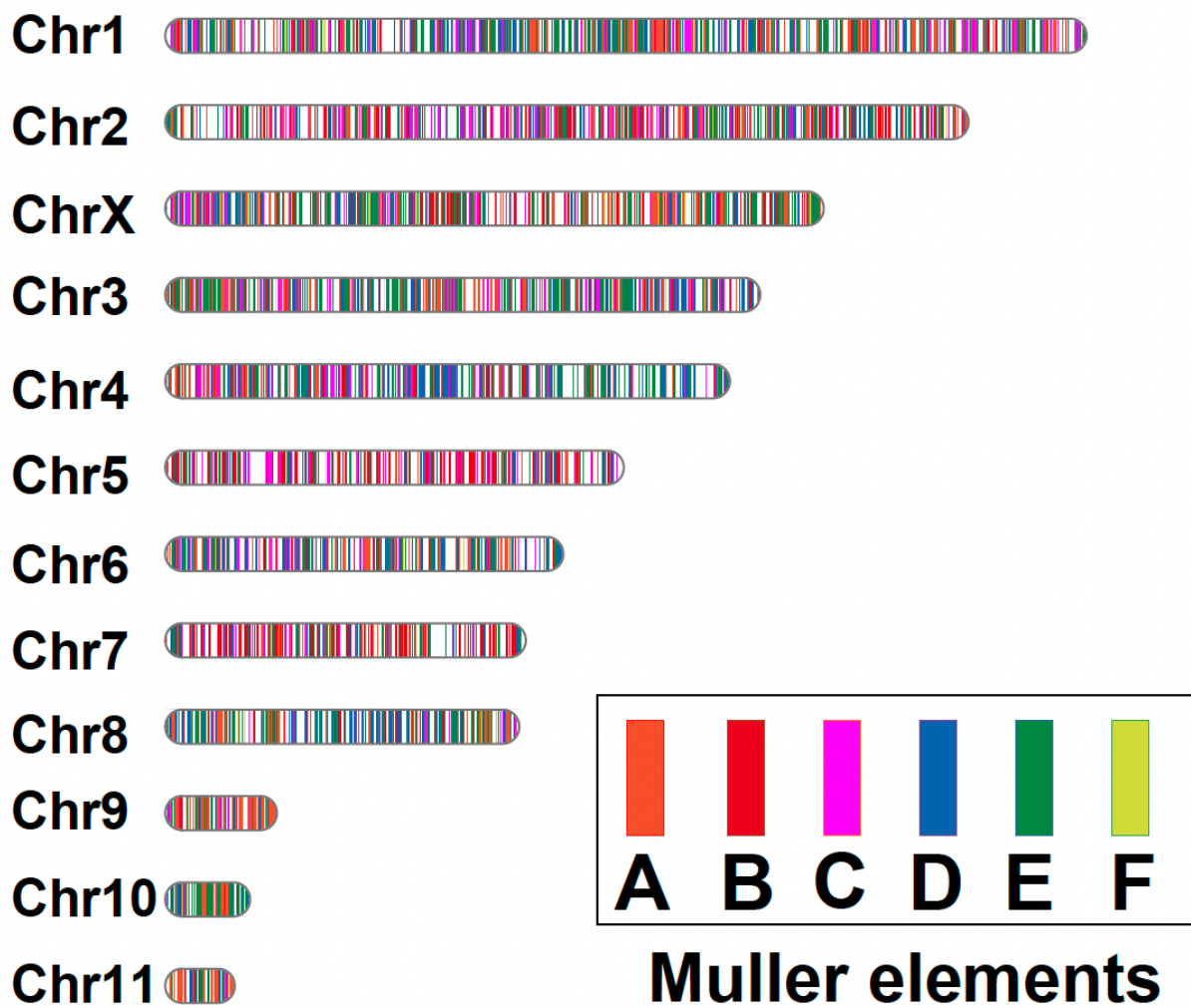
Supplementary Figure 22. Boxplot of gene expression and H4K16ac abundances in compartments A of the locust chromosomes. Compartments were delineated at a 0.5 Mb resolution using principal component analysis conducted with the HiTC package in R. The compartment with a higher gene density was defined as compartment A, and the other as compartment B. Box plots indicate the median (center line), the first and third quartiles (box limits), and whiskers extending to $1.5\times$ the interquartile range.



Supplementary Figure 23. mRNA expression of *sxl* and MSL-related genes in locust brains of females and males. The mRNA expression was determined by RNA-seq. TPM, Transcripts Per Million. $n = 3$ independent biological replicates. The data are shown as mean + SEM.

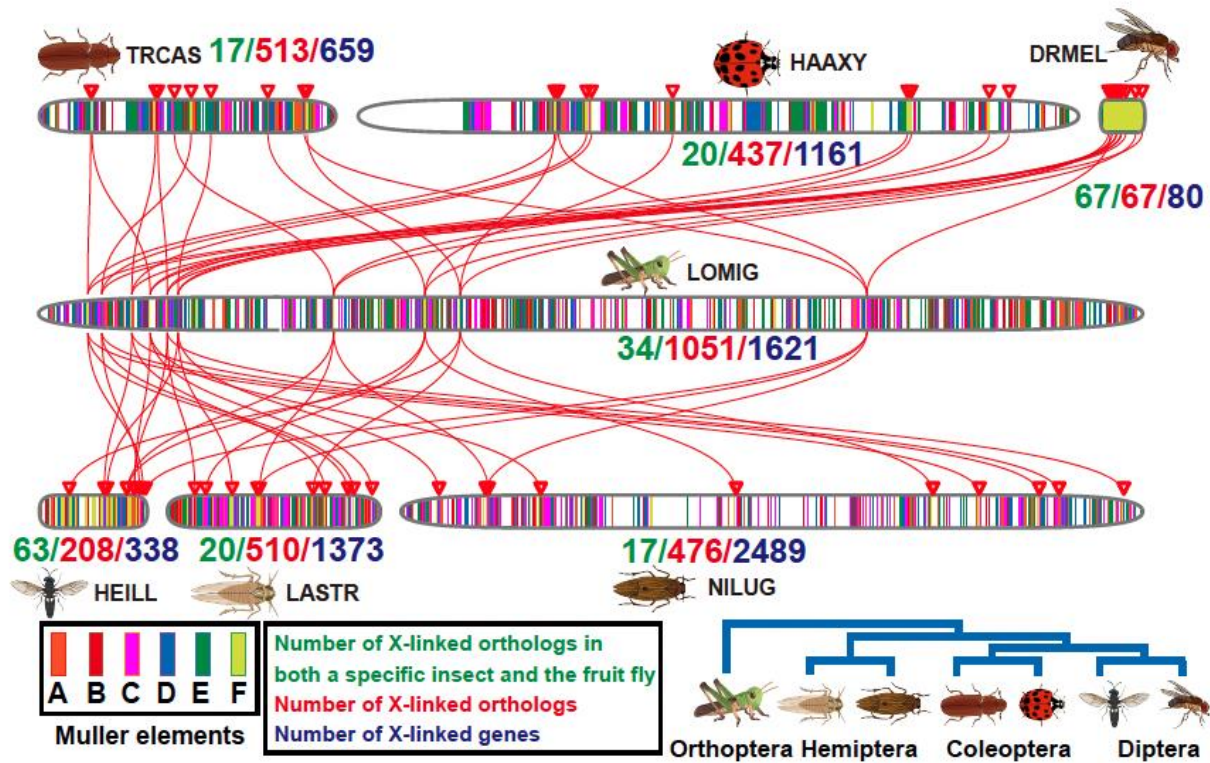


Supplementary Figure 24. The protein expression ratios of *sxl* and MSL-related genes in female and male locusts. The protein expression ratios of these genes were compared using 4D-DIA (four-dimensional data-independent acquisition) quantitative proteomic analysis.

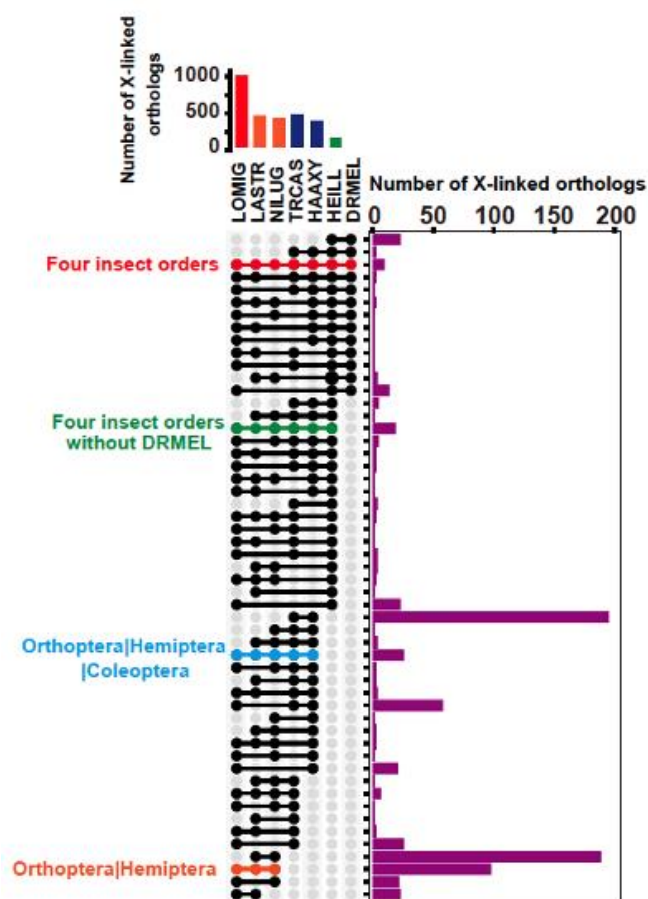


Supplementary Figure 25. Assessment of chromosome conservation of the migratory locust using the orthologs of Muller elements of *Drosophila melanogaster*.

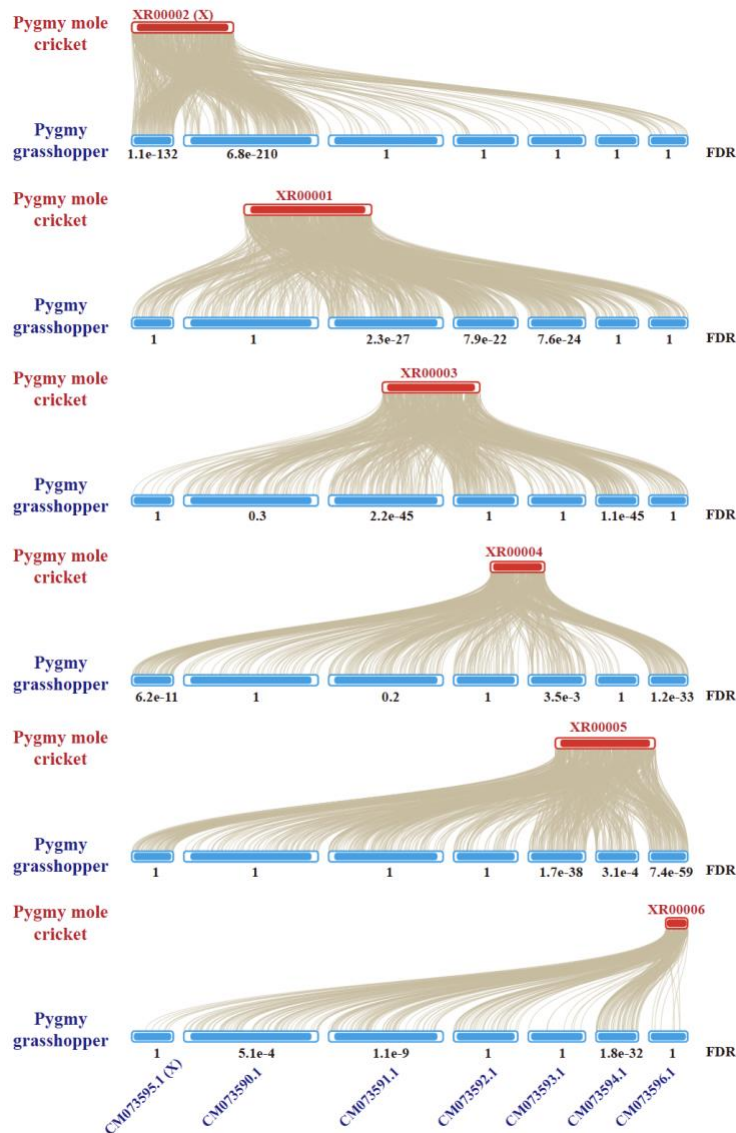
The X chromosome in insects from four different orders



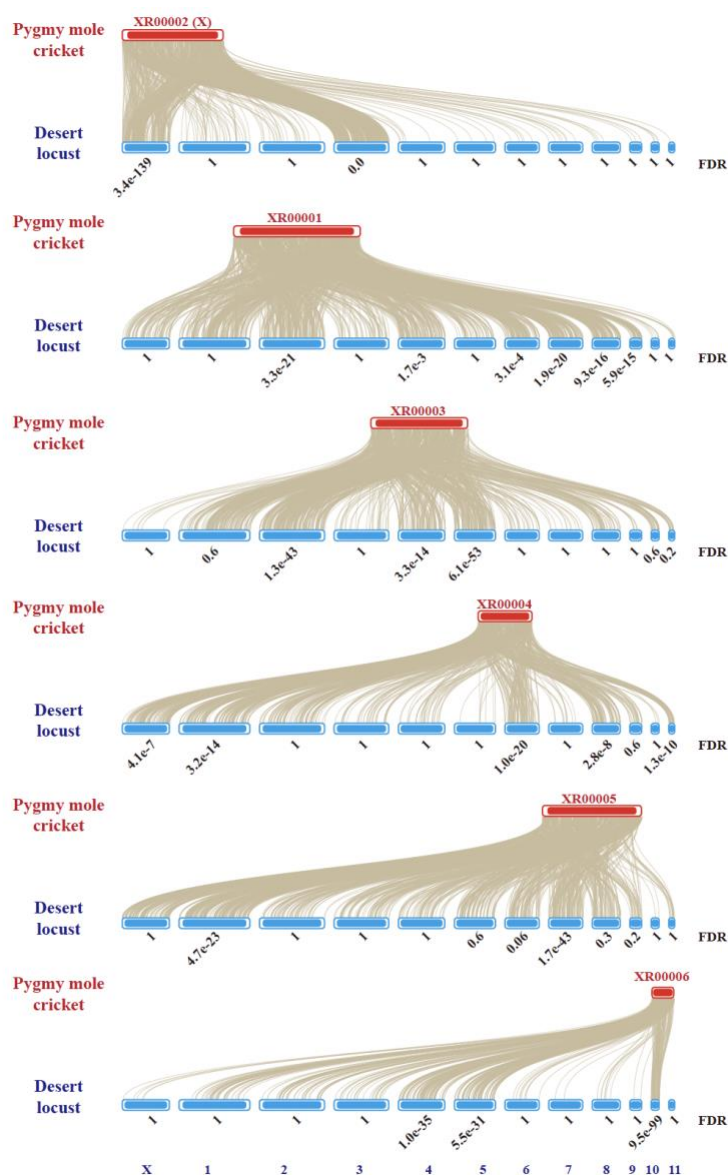
Supplementary Figure 26. X-linked orthologs between the migratory locust and the other six insect species from Hemiptera, Coleoptera, and Diptera. TRCAS, *Tribolium castaneum*; HAAXY, *Harmonia axyridis*; DRMEL, *Drosophila melanogaster*; HEILL, *Hermetia illucens*; LASTR, *Laodelphax striatellus*; NILUG, *Nilaparvata lugens*; LOMIG, *Locusta migratoria*.



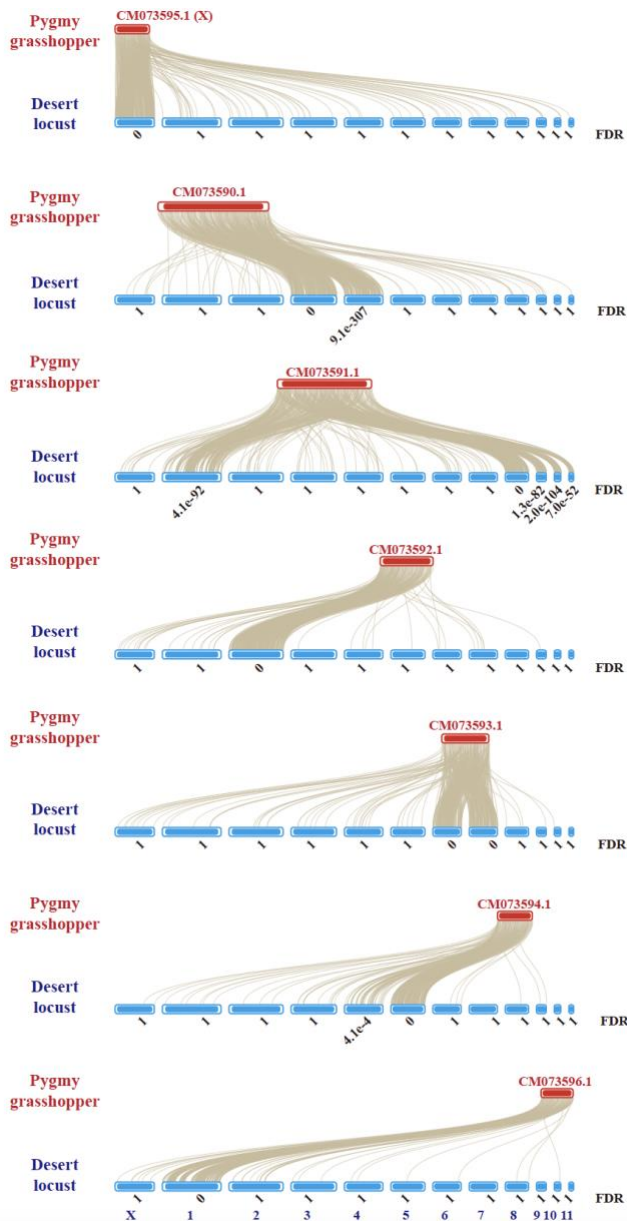
Supplementary Figure 27. Number of X-linked orthologs between Orthoptera and the other three insect orders. Only the 1-to-1 orthologs were included in this analysis. The lines between two chromosomes represent the 1-to-1 ortholog relationships.



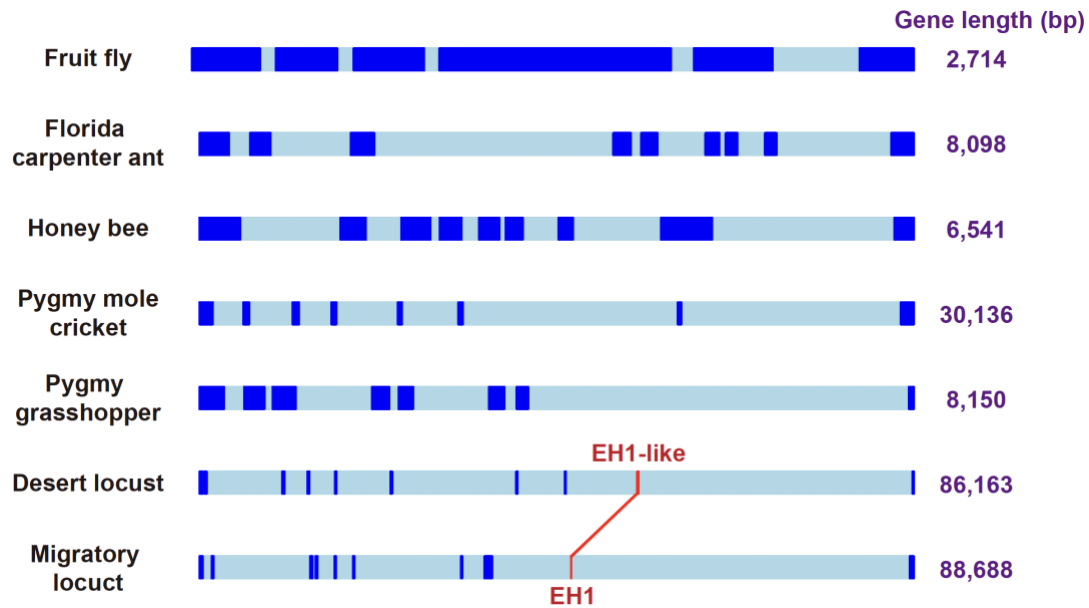
Supplementary Figure 28. The X-linked orthologs between the pygmy grasshopper and the pygmy mole cricket. The method for determining the significance of chromosome-chromosome associations between these insects was adapted from a previous study ¹⁵. The significance of chromosome-chromosome associations between these insects was evaluated using Fisher's exact test, compared against a null model based on random gene permutation. For each species pair, we applied a Bonferroni correction to account for multiple chromosome-chromosome comparisons. Adjusted P values are shown at the bottom of each chromosome.



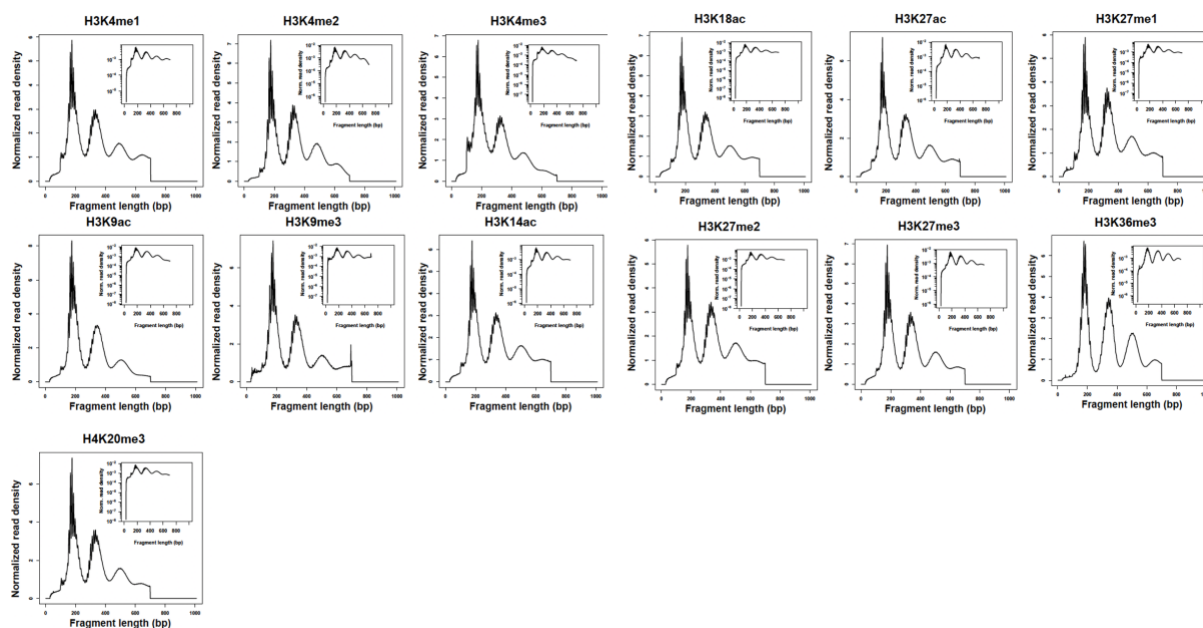
Supplementary Figure 29. The X-linked orthologs between the pygmy mole cricket and the desert locust. The method for determining the significance of chromosome-chromosome associations between these insects was adapted from a previous study ¹⁵. The significance of chromosome-chromosome associations between these insects was evaluated using Fisher's exact test, compared against a null model based on random gene permutation. For each species pair, we applied a Bonferroni correction to account for multiple chromosome-chromosome comparisons. Adjusted P values are shown at the bottom of each chromosome.



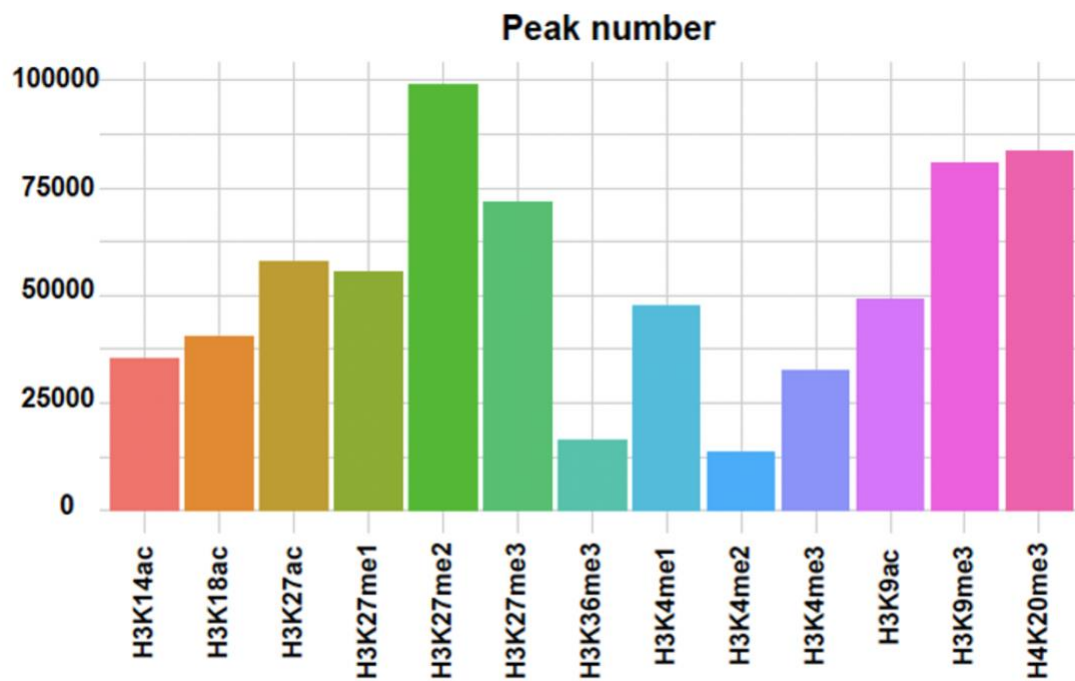
Supplementary Figure 30. The X-linked orthologs between the pygmy grasshopper and the desert locust. The method for determining the significance of chromosome-chromosome associations between these insects was adapted from a previous study ¹⁵. The significance of chromosome-chromosome associations between these insects was evaluated using Fisher's exact test, compared against a null model based on random gene permutation. For each species pair, we applied a Bonferroni correction to account for multiple chromosome-chromosome comparisons. Adjusted P values are shown at the bottom of each chromosome.



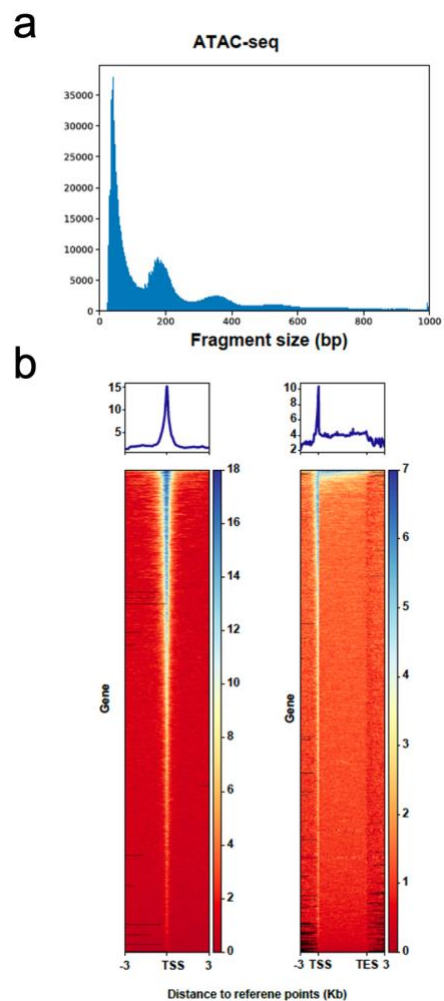
Supplementary Figure 31. The gene structures of *Henna* in the seven insect species. Dark blue and light blue rectangles indicate exons and introns, respectively. The EH1-like sequence in the desert locust shows 68% nucleotide similarity to EH1 in the migratory locust.



Supplementary Figure 32. Fragment size distribution of CUT&Tag experiments of histone modifications. The sequencing reads derived from large fragments (>700 bp) were excluded in these analyses.



Supplementary Figure 33. Peak numbers of the 13 histone modifications in the locust genome.



Supplementary Figure 34. Fragment size distribution (A) and TSS enrichments (B) of ATAC-seq experiments.

Supplementary Tables

Supplementary Table 1. Summary of assembly statistics for the migratory locust genomes (LMv3.1 and LMv2.4) and the desert locust genome (SGv3).

	LMv3.1	LMv2.4	SGv3
Contig N50 (bp)	2,232,395	9,292	7,740,000
Contig length>5kb (bp)	6,947,099,932	4,226,930,634	8,695,866,202
Scaffold N50 (bp)	722,372,665	324,817	825,498,460
Scaffold L50	4	3,615	5
Scaffold length>5kb (bp)	7,133,095,908	6,502,763,785	8,701,661,913
Scaffold longest (bp)	1,118,529,874	8,179,469	1,253,778,933
GC (%)	41.13	40.67	42.37

Supplementary Table 2. Sequences of the primers, probes and gRNAs used.

Primer	Sequence
Control primer1	CATGATAACCGACGCGCAAC
Control primer2	ACACACAGCGAATGAACCCA
3C primer1	GGCCTAGCACTTGGTGAAGTATTA
3C primer2	CGTACAAGAAGGCCAATGAAAGTT
Henna qPCR primer F	CGGGGAAGGGTAAAAAC
Henna qPCR primer R	CTTGGCTGAGTCGAGGC
CRISPR gRNA	GAAGTTGCGTAATGCCGCAA
X_FISHprobe-1	AAAAGCAACCATGCTTCTGC
X_FISHprobe-2	AACAAGTGGATCCTCACAGC
X_FISHprobe-3	ACTCTTGCAATAATTCTGCA
X_FISHprobe-4	AGGATAGTTCTCATTTGTCA
X_FISHprobe-5	ATAATGGCTGCAGGATCCTG
X_FISHprobe-6	CAAATCACTACCTTCACCAC
X_FISHprobe-7	CATTGAGTGCCTGTGAATTC
X_FISHprobe-8	CATTTTGTGCTGCATCAACC
X_FISHprobe-9	CCATTGTTTCAGTCATTTTGT
X_FISHprobe-10	CTCAAGGATGTAACATGCGT
X_FISHprobe-11	CTGGGTTGTCCTCATCAAAT
X_FISHprobe-12	CTTCTCATCCTGTTGACAAG
X_FISHprobe-13	GAAATGATGATCTCAGGCCA
X_FISHprobe-14	GAATGGCTCCAATAAGGTCT

X_FISHprobe-15	GCAAGAGCTTTTAAAGCCTT
X_FISHprobe-16	TAAAGCCTGAGCATTACAC
X_FISHprobe-17	TAGGATTTGCAAACAGCCAT
X_FISHprobe-18	TCACTGTTTTCTTCCTTTGG
X_FISHprobe-19	TCTGCAACTAGGTCTCTGTA
X_FISHprobe-20	TCTGTTTCTTCAGATTGTGT
X_FISHprobe-21	TGCTGCTAATCTCTCCTTTC
X_FISHprobe-22	TTGATGTAGCCTCTACACTG
X_FISHprobe-23	TTTGAAAGGACGTCTGCCAT
X_FISHprobe-24	TTTAAATGTTTCCATCAGGC

References

1. Liu, Q., Jiang, F., Zhang, J., Li, X., and Kang, L. (2021). Transcription initiation of distant core promoters in a large-sized genome of an insect. *BMC Biol* 19, 62. 10.1186/s12915-021-01004-5.
2. Samata, M., and Akhtar, A. (2018). Dosage Compensation of the X Chromosome: A Complex Epigenetic Assignment Involving Chromatin Regulators and Long Noncoding RNAs. *Annual review of biochemistry* 87, 323-350. 10.1146/annurev-biochem-062917-011816.
3. Keller Valsecchi, C.I., Marois, E., Basilicata, M.F., Georgiev, P., and Akhtar, A. (2021). Distinct mechanisms mediate X chromosome dosage compensation in *Anopheles* and *Drosophila*. *Life Sci Alliance* 4. 10.26508/lsa.202000996.
4. Palmer, M.J., Richman, R., Richter, L., and Kuroda, M.I. (1994). Sex-specific regulation of the male-specific lethal-1 dosage compensation gene in *Drosophila*. *Genes Dev* 8, 698-706. 10.1101/gad.8.6.698.
5. Vielle, A., Lang, J., Dong, Y., Ercan, S., Kotwaliwale, C., Rechtsteiner, A., Appert, A., Chen, Q.B., Dose, A., Egelhofer, T., et al. (2012). H4K20me1 contributes to downregulation of X-linked genes for *C. elegans* dosage compensation. *PLoS Genet* 8, e1002933. 10.1371/journal.pgen.1002933.
6. Meisel, R.P., Delclos, P.J., and Wexler, J.R. (2019). The X chromosome of the German cockroach, *Blattella germanica*, is homologous to a fly X chromosome despite 400 million years divergence. *BMC Biol* 17, 100. 10.1186/s12915-019-0721-x.
7. Li, Y., Park, H., Smith, T.E., and Moran, N.A. (2019). Gene Family Evolution in the Pea Aphid Based on Chromosome-Level Genome Assembly. *Mol Biol Evol* 36, 2143-2156. 10.1093/molbev/msz138.
8. Mathers, T.C., Wouters, R.H.M., Mugford, S.T., Swarbreck, D., van Oosterhout, C., and Hogenhout, S.A. (2021). Chromosome-Scale Genome Assemblies of Aphids Reveal Extensively Rearranged Autosomes and Long-Term Conservation of the X Chromosome. *Mol Biol Evol* 38, 856-875. 10.1093/molbev/msaa246.
9. Landeen, E.L., and Presgraves, D.C. (2013). Evolution: from autosomes to sex chromosomes--and back. *Curr Biol* 23, R848-850. 10.1016/j.cub.2013.08.021.
10. Montagna, M., Tong, K.J., Magoga, G., Strada, L., Tintori, A., Ho, S.Y.W., and Lo, N. (2019). Recalibration of the insect evolutionary time scale using Monte San Giorgio fossils suggests survival of key lineages through the End-Permian Extinction. *Proc Biol Sci* 286, 20191854. 10.1098/rspb.2019.1854.
11. Levillain, O., Diaz, J.J., Blanchard, O., and Dechaud, H. (2005). Testosterone down-regulates ornithine aminotransferase gene and up-regulates arginase II and ornithine decarboxylase genes for polyamines synthesis in the murine kidney. *Endocrinology* 146, 950-959. 10.1210/en.2004-1199.
12. Lu, W., Lakonishok, M., and Gelfand, V.I. (2023). The dynamic duo of microtubule polymerase Mini spindles/XMAP215 and cytoplasmic dynein is essential for maintaining *Drosophila* oocyte fate. *Proc Natl Acad Sci U S A* 120, e2303376120. 10.1073/pnas.2303376120.
13. Millan-Zambrano, G., Burton, A., Bannister, A.J., and Schneider, R. (2022). Histone post-translational modifications - cause and consequence of genome function. *Nat Rev Genet* 23, 563-580. 10.1038/s41576-022-00468-7.
14. Jamge, B., and Berger, F. (2022). Diversification of chromatin organization in eukaryotes. *Current opinion in cell biology* 74, 1-6. 10.1016/j.ceb.2021.12.002.

15. Simakov, O., Bredeson, J., Berkoff, K., Marletaz, F., Mitros, T., Schultz, D.T., O'Connell, B.L., Dear, P., Martinez, D.E., Steele, R.E., et al. (2022). Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci Adv* 8, eabi5884. 10.1126/sciadv.abi5884.