

Chromatin dynamics of a large-sized genome provides insights into polyphenism and X0 dosage compensation of locusts

Received: 4 June 2024

Accepted: 4 August 2025

Published online: 1 September 2025

 Check for updates

Qing Liu^{1,7}, Feng Jiang^{1,2,7}, Ran Li^{1,2,7}, Shanlin Liu^{1,3,7}, Wenjing Fang^{1,2,7}, Xiao Li¹, Ting Hu^{1,2}, Lianyun Feng^{1,2}, Jie Zhang¹, Zhikang Liu¹, Jing He¹, Wei Guo¹, Xianhui Wang¹, Zongyi Sun³, Jingjing Li³, Yunan Gao³, Jiacheng Yi³, Qiye Li^{2,4}, Xiaoxiao Wang⁵, Liya Wei⁵ & Le Kang^{1,5,6} ✉

Locusts are characterized by a large genome size, polyphenism and an X0 sex determination system. Here we generated chromosome-level genomes for both desert and migratory locusts, as well as a comprehensive chromatin map for the latter. We found that genome enlargement is associated with an increased number of enhancers in expanded intronic regions. To explore the function of expanded enhancers, we identified a distal enhancer that contributes to behavioral differences between solitary and gregarious locusts. In the X0 sex system, H4K16ac enrichment and H4K20me1 depletion maintain balanced X-linked expression in male soma. Notably, the distance-dependent H4K16ac enrichment diminishes gradually in intergenic regions, revealing a special dosage compensation mechanism in large genomes. Furthermore, the distance-dependent H4K16ac results in a lag in the evolution of dosage compensation for the X-linked genes translocated recently from autosomes. Therefore, expanded intronic and intergenic regions shape a distinct chromatin regulation landscape in large genomes.

The migratory locust (*Locusta migratoria*), with a 6.9-Gb genome, is a typical representative of large-sized-genome insect species and is characterized by marked density-dependent polyphenism and an X0 sex determination system¹. Polyphenism or phenotypic plasticity, which is the ability of organisms with the same genotypes or genome to change their phenotype in response to a changing environment, has an obvious epigenetic characteristic. As a unique biology feature of locusts, polyphenism facilitates the rapid behavioral changes between solitary and gregarious phases to adapt to changes in population density¹. Behavioral changes lead to aggregation of locusts, which can ultimately result in the formation of destructive swarms. Therefore, behavioral dissimilarity between the two phases of locusts is the most important phase-related difference because it is the main cause of locust plague

outbreaks. The gene *Henna* encodes the most critical enzyme in dopamine biosynthesis, which determines behavioral changes in locusts^{1–3}. Behavioral plasticity is notably regulated at the transcriptional level, which is influenced by chromatin regulation. Compared to species with small genomes, such as the fruit fly, the gene size of *Henna* in the migratory locust has increased 32-fold, primarily because of the expansion in intron size. Therefore, *Henna* represents a unique case in the migratory locust for investigating the role of chromatin in the regulation of insect polyphenism, particularly with respect to expanded intronic regions.

Sexual differentiation is associated closely with chromatin regulation⁴. Locusts are typical representatives of the X0 sex determination system, which is characterized by male heterogamety. Although unpaired homologous chromosomes can disrupt the equilibrium of

¹State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing, China.

²College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. ³GrandOmics Biosciences, Wuhan, China. ⁴BGI Research, Wuhan, China. ⁵College of Life Science, Hebei University, Baoding, China. ⁶Institute of Cell and Gene Technology, Shenzhen University of Advanced Technology, Shenzhen, China. ⁷These authors contributed equally: Qing Liu, Feng Jiang, Ran Li, Shanlin Liu, Wenjing Fang. ✉e-mail: lkang@ioz.ac.cn

gene expression, differences in X chromosome number between the two sexes are mitigated by dosage compensation⁵. Dosage compensation strategies involving histone modifications are variable among different sex determination models⁶. However, the dosage compensation mechanism of an XO sex determination system, such as that of locusts, that adheres strictly to obligate sexuality following Mendelian inheritance, remains unresolved. In principle, translocated genes between sex chromosomes and autosomes are embedded more frequently in intergenic regions in large genomes than in small genomes because of more spacious context. The spread of H4K16ac, catalyzed by male-specific lethal (MSL) complex, can promote dosage compensation to the flanking intergenic that lack MSL complex binding⁷. Therefore, the expansive intergenic regions in large genomes provide an ideal system for exploring the effects of dosage compensation and its associated spreading efficiency in gene translocation.

Here we generated a chromosome-level genome assembly of the migratory locust for comprehensive annotation of chromatin structure. Taking polyphenism as an example, we compared chromatin differences to explore the roles of *cis*-regulatory elements in the behavioral changes of locusts. Furthermore, we determined allele-aware X-linked expression in XX females and investigated the role of histone modification in the loss and maintenance of balanced X-linked expression. We further generated a chromosome-level genome assembly of the desert locust and revealed the role of histone modification in X-linked gene translocation in locusts.

Results

Chromosome-level assembly of the migratory locust

We conducted genome sequencing on an adult female locust with a heterozygosity of 1.58%, using a combination of long-read sequencing, optical mapping and chromosome conformation capture based on proximity ligation (Figs. 1a,b, Supplementary Figs. 1 and 2, Supplementary Table 1 and Supplementary Note 1). In female locusts, the third longest scaffold, identified as the X chromosome based on Hoechst staining of metaphase chromosomes, showed equal read coverage to autosomes and about twice that in male locusts (Figs. 1c,d). The high contiguity of the LMv3.1 assembly greatly improved the completeness of the transposable elements (TEs) (Supplementary Fig. 3). Fluorescent in situ hybridization (FISH) of a pericentromeric satellite DNA sequence (Fig. 1e and Supplementary Note 1) showed fluorescent bands on one of the chromosome ends (Fig. 1f), demonstrating the anticipated chromosome end regions. We predicted 18,127 protein-coding genes using RNA expression and homology data (Supplementary Fig. 4). The LMv3.1 assembly presented higher mapping rates of full-length transcripts than the LMv2.4 assembly, indicating improved coverage of RNA transcript structures (Supplementary Figs. 5 and 6). The improvement in genome contiguity facilitates further investigation of the epigenomic landscape in locusts.

Increased number of intronic enhancers in large-sized genes

To determine the chromatin structure, we performed cleavage under targets and tagmentation (CUT&Tag) sequencing for a panel of 13 histone modifications, assay for transposase-accessible chromatin using sequencing (ATAC-seq), transcription start site (TSS) sequencing (TSS-seq) and strand-specific RNA sequencing (RNA-seq) in brains. The hierarchical clustering of read coverage clearly reflected the non-crossreactivity of the primary antibodies, the consistency of biological replicates and the differences among sequencing methods (Fig. 2a). We demonstrated that the abundance of histone modifications is associated with gene expression levels (Supplementary Note 2 and Supplementary Fig. 7). Approximately 40% of the genomic regions were classified into the 11 ChromHMM-defined chromatin states, which are associated with distinct regulatory elements (Fig. 2b, Supplementary Note 3 and Supplementary Figs. 8–11). As exemplified in Supplementary Fig. 12, chromatin states allow for inference of gene expression.

The intergenic region contributes the largest portion of chromatin states, due to its dominance in genomic composition (Supplementary Fig. 13). The permissive and repressed chromatin states have a considerable portion of TEs, indicating that TEs contribute to host regulatory innovation and are subject to epigenetic impact (Fig. 2c, Supplementary Note 4 and Supplementary Figs. 14 and 15).

The intron length in locusts exceeds the values observed in other insects (Supplementary Figs. 5 and 16). The chromatin states E5 (enhancer) and E7 (bivalent enhancer) occupy a broader region in the large introns than in the short introns, indicating a higher number of enhancers in large introns (Fig. 2d). The changes in intron length are correlated positively with the changes in enhancer number in the locust compared to the fruit fly and the honey bee, suggesting that the expansion of intron size is associated with an increase in enhancer number (Fig. 2e). The changes in gene length do not have a significant impact on the gene expression changes in the migratory locust compared to the fruit fly (Fig. 2f). Furthermore, expression values were distributed evenly among short-sized and large-sized genes in locusts (Supplementary Fig. 17), demonstrating that gene structure enlargement in locusts does not significantly impact gene expression. We also found that the RNA expression of the large-sized gene associated with an intronic enhancer is significantly higher than that of the large-sized gene lacking intronic enhancers (Fig. 2g), although it remains unclear whether these enhancers serve the role of enhancing transcription levels or adding extra control over gene expression. Compared with the short-sized genes, the large-sized genes show greater depletion of H3K4me2, H3K4me3 and H3K9ac, as well as greater deposition of H3K27me2 and H3K27me3 (Fig. 2h), which are catalyzed primarily by polycomb repressive complexes⁸. These results suggest that polycomb repressive complexes facilitate the spreading of H3K27me3, increase chromatin compaction and tether long-distance chromatin loops that promote contact between distal enhancer in the large-sized genes⁹. Taken together, the increase in enhancer numbers in large-sized genes, coupled with genome size expansion, may ensure equalized gene expression across genes of varying lengths in locusts.

Regulation of the intronic enhancer in behavioral plasticity

Because polyphenism is clearly modulated epigenetically¹⁰, we tried to use the behavioral changes of locusts as an example of epigenetic regulation in large genomes. Principal component analysis revealed the distinguishable spatial distribution of histone modifications linked to regulatory, repressive and transcription marks between the two locust phases (Fig. 3a). In the brain, we identified 1,069 genes that were upregulated and 485 genes that were downregulated in gregarious locusts. Gene ontology analysis suggests that the differentially expressed genes are involved in regulatory mechanisms underpinning behavioral regulation, signal transduction, synaptic plasticity and immune modulation (Supplementary Fig. 18). The two most significantly positively correlated histone modifications—H3K4me3 and H3K27ac—at the TSS region emphasize their role in promoter regulation (Fig. 3b). Conversely, the two most significantly negatively correlated histone modifications—H3K4me1 and H3K27me3—were situated 10 kb upstream of the TSS region, implying their involvement in a distal regulation as enhancers. Compared to the other three genomic categories, including promoters, introns and intergenic regions, H3K4me1 in the 10 kb region upstream of TSS showed the least overlap with H3K27ac (Supplementary Fig. 19), indicating primed enhancers lacking transcriptional enhancement¹¹. We identified the binding sites of transcription regulators associated with differential H3K4me3/H3K27ac at promoters and differential H3K4me1/H3K27me3 situated 10 kb upstream of the TSS region (Fig. 3c). We found higher frequencies of five transcription regulators located in these two regions, including *cg* (which establishes long-range chromatin contacts at promoters¹²), *Trl* (a transcription activator associated with nucleosome remodeling and long-distance enhancer–promoter communication¹³), *l(3)neo38*

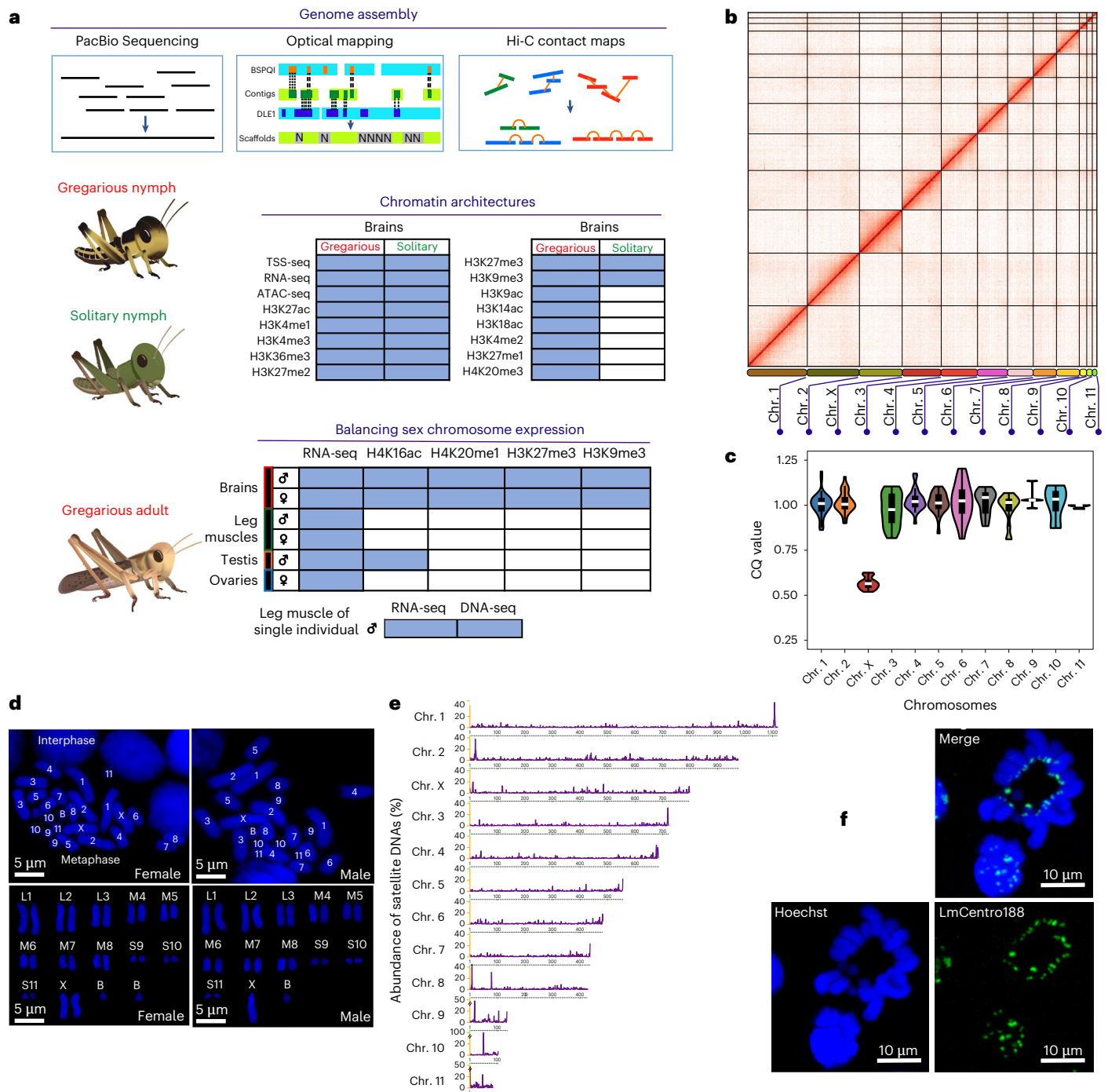
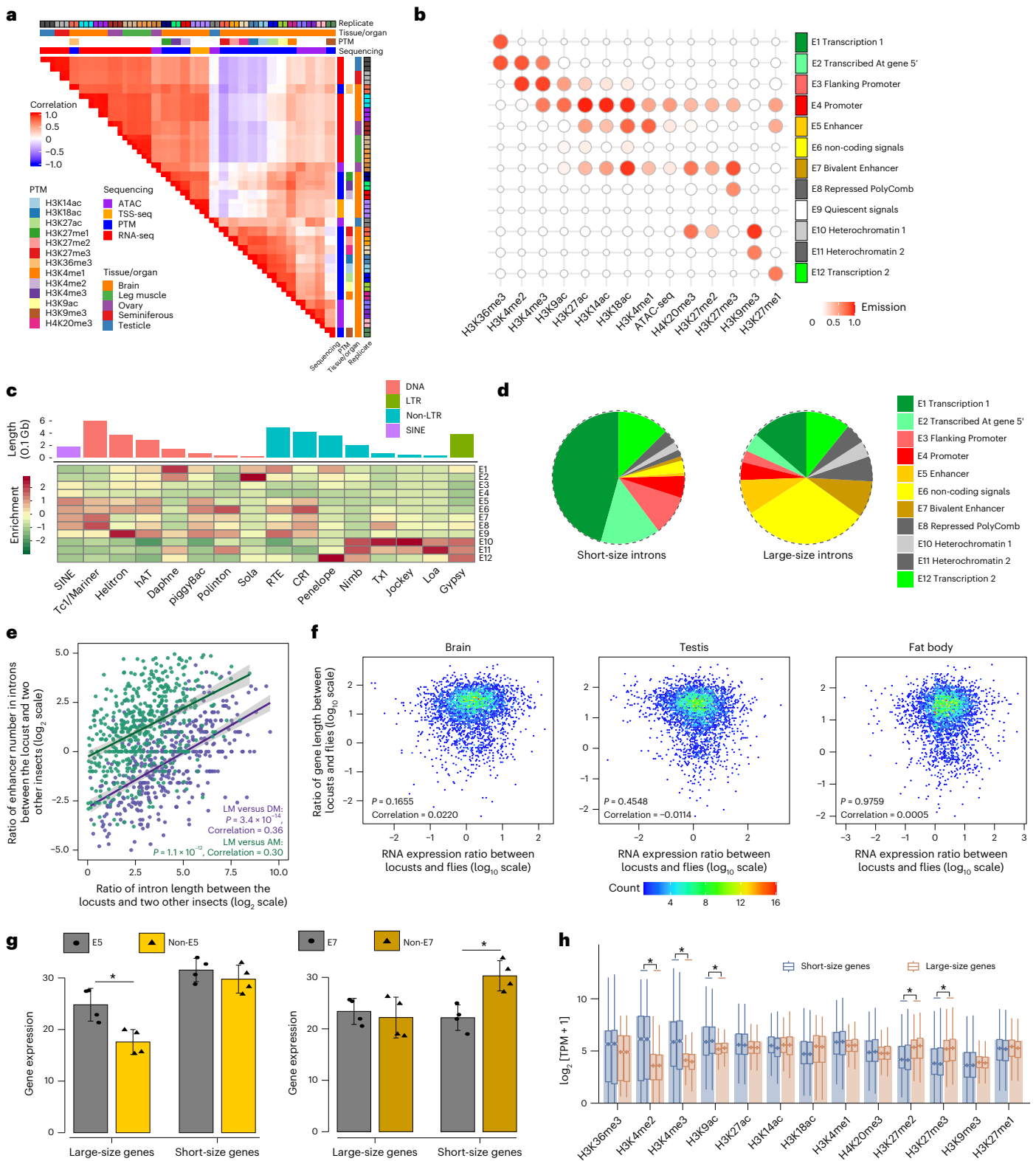


Fig. 1 | Chromosome-level assembly of the migratory locust. a, Summary of sequencing datasets in this study. **b**, Heatmap showing frequency of Hi-C contacts along the migratory locust genome assembly. **c**, Identification of the X chromosome using the chromosome quotient method. The X chromosome was identified by aligning the -30X DNA-seq data from both male and female locusts to the genome assembly. To calculate the read coverage, the chromosomes were split into 30-Mb bins. The bin numbers for chromosomes 1–11 and X are 40, 33, 27, 24, 20, 18, 16, 16, 5, 4, 3 and 27, respectively. Boxplots indicate the median (center line), the first and third quartiles (box limits) and whiskers extending to 1.5× the interquartile range. **d**, Characterization of locust

chromosomes. Embryos of the migratory locust were subjected to chromosome characterization using Hoechst 33342 nucleic acid stain; $n = 1$ independent biological replicates for demonstration purpose. **e**, The percentage distribution of satellite DNA sequences along locust chromosomes. Because centromeric satellite DNAs show a typical size greater than 100 bp, short tandem repeats with monomer lengths <100 bp were excluded from this analysis. The chromosomes were divided by partitioning into 1-Mb bins. **f**, FISH of the satellite DNA LmCentro188. The signal probe for LmCentro188 is conjugated with Alexa Fluor 488, and the chromosomes were labeled using Hoechst nucleic acid staining; $n = 2$ independent biological replicates.

(an essential regulator of nervous-system development¹⁴), *Sp1* (which controls gene expression within dopaminergic neurons¹⁵) and *Clamp* (a pioneer factor that increases chromatin accessibility at promoters¹⁶). Three of these five transcription regulators, including

Trl, *Sp1* and *Clamp* were differentially expressed between the solitary and gregarious locusts. Therefore, these three transcription regulators were probably involved in regulating differentially histone modification regions in brains of solitary and gregarious locusts.



Because the enlargement of genome size results in the presence of long introns¹⁷, we sought to explore whether the intronic *cis*-regulatory elements are involved in the regulation of behavioral plasticity of locusts. We used the gene *Henna* as an example because we have shown previously that it is the most critical gene in determining behavioral plasticity in dopamine biosynthesis pathway^{2,3}. The dominant TSS of *Henna* in brains, as identified by TSS-seq, was TSS1, whose chromatic pattern was correlated with differential RNA-seq expression between

solitary and gregarious locusts (Fig. 3d). In gregarious locusts, we observed upregulated deposition of ATAC, H3K4me3 and H3K27ac in the TSS1 region, as well as H3K36me in the gene-body region. Notably, we found an enhancer, EH1, which was decorated with high H3K4me1 and low H3K4me3 signals, in the 10.3-kb intronic region upstream of the first *Henna* coding exon. The significantly higher interaction frequency in quantitative chromosome conformation capture (3C) DNA confirmed the close three-dimensional physical proximity between

Fig. 2 | Difference in enhancer number between large-sized and short-sized genes. **a**, Pearson correlations among sequencing assays, tissues/organs and biological replicates based on the normalized signals in 10-kb window bins. **b**, Emission patterns of the 12 chromatin states. **c**, Enrichment profiles of chromatin states overlapped with different TE superfamilies. The 16 most abundant TE superfamilies were included in this analysis. **d**, Genomic region occupancy of chromatin states in short and long introns. Short-size introns correspond to the lower 25% of data, whereas large-size introns correspond to the upper 75% of data. **e**, Correlation between intron length changes and changes in enhancer number in the brains of the migratory locust (LM), the fruit fly (DM) and the honey bee (AM). A peak was considered an active enhancer if the abundance of both H3K4me1 and H3K27ac exceeded their first quartile (Q1). Error bands represent the 95% confidence interval around the fitted smoothing curve.

f, Correlation between gene length changes and expression changes in somatic and reproductive tissues. Transcripts per kilobase million (TPM) was used as the unit for gene expression quantification. **g**, RNA expression of genes varies with different sizes in the presence and absence of chromatin states E5 and E7. Error bars, s.d. Short-size genes correspond to the lower 25% of data, whereas large-size genes correspond to the upper 75% of data; $n = 4$ independent biological replicates. Data are shown as mean \pm s.e.m. $P = 0.0125$ for the E5 comparison and $P = 0.0058$ for the E7 comparison. $*P < 0.05$. **h**, Histone modification deposition on gene-body regions in short-size and large-size genes. Deposition abundance was determined using TPM + 1 on a \log_2 scale. $*False\ discovery\ rate < 0.01$. Boxplots indicate the median (center line), the first and third quartiles (box limits), and whiskers extending to $1.5 \times$ the interquartile range. PTM, post-translational modification.

TSS1 and EHI (Fig. 3e). The EHI-knockout mutants in gregarious locusts resulted in a significant reduction of *Henna* expression at mRNA and protein levels (Figs. 3f,g,h). Consequently, the EHI-knockout mutants of gregarious locusts showed a significant behavioral shift toward solitary behavioral traits in behavioral arena assays (Fig. 3i). Indeed, our previous studies have demonstrated that RNAi silencing of *Henna* expression induces a behavioral change from solitary to gregarious phases². Therefore, the knockout of EHI in gregarious locusts resulted in altered behavior—a phenotype copied by silencing *Henna* expression. Collectively, the enhancers located in long intron interact with promoters to regulate behavioral plasticity of locusts.

Chromatin changes of X chromosome during meiotic silencing

In the brain and leg muscles, we found a strong positive diagonal correlation between male and female gene expression on the X chromosome and autosomes (Fig. 4a), indicating a comparable expression between X chromosome and autosomes in the two sexes. Unlike autosomes, the X chromosome exhibited higher gene expression in ovaries compared to testes. So, the gene expression between the sexes is re-equilibrated in soma but not in gonads.

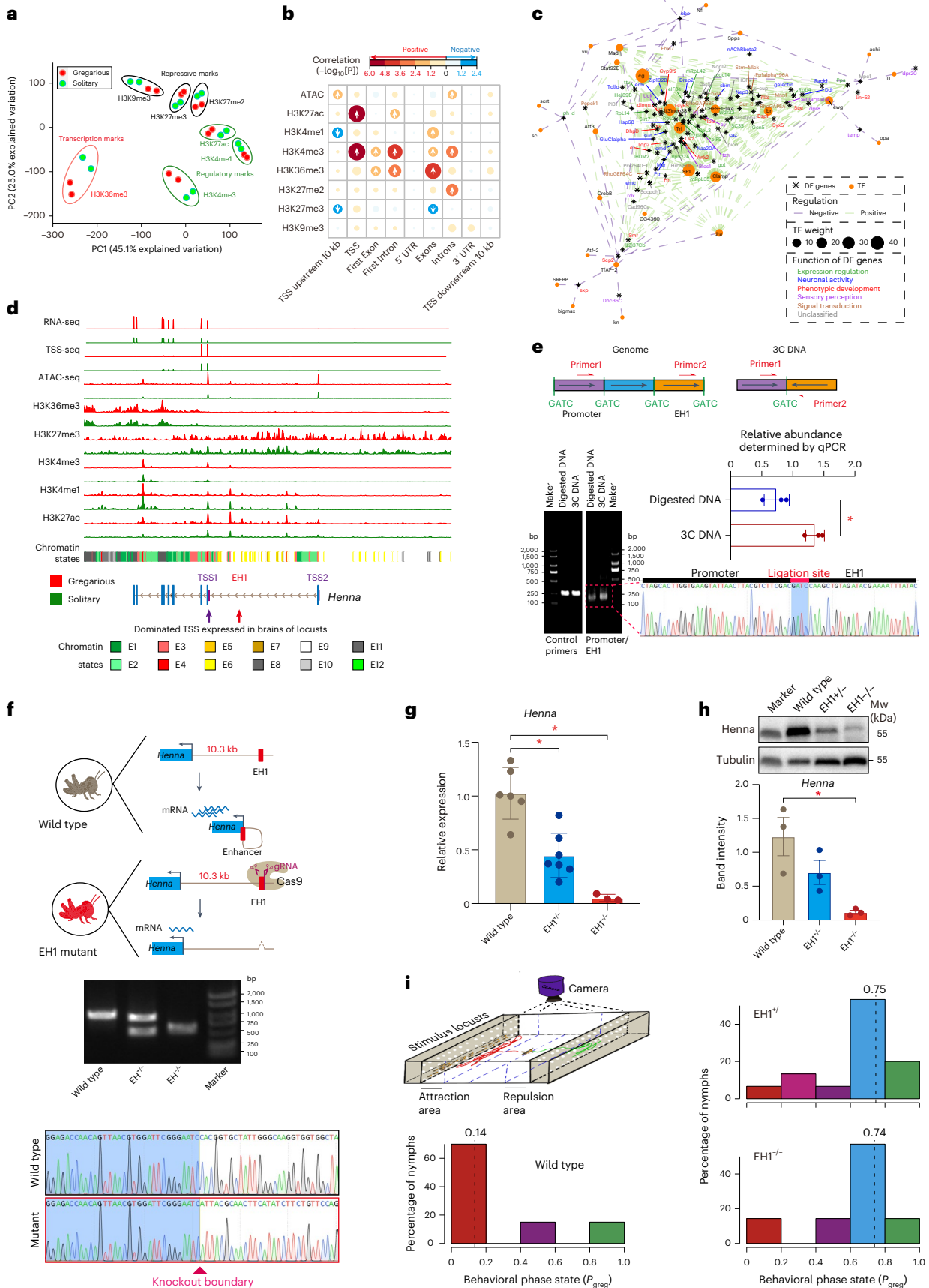
The X/A ratio in testes held near 0.5—a key indicator of meiotic sex chromosome inactivation (Fig. 4b). The X chromosome showed increased Hi-C signals/interaction frequencies, as evidenced by the presence of parallel lines along the main diagonal in the interaction map (Fig. 4c). The increased interaction frequencies are unique to the X chromosome and not observed in the other autosomes (Supplementary Fig. 20). The log–log plot of interaction frequencies against distance showed that the X chromosome deviated gradually from the autosomes as distance increased (Fig. 4d). Compared to autosomes, the slower interaction frequency decay in the X chromosome indicates the relatively higher interaction frequencies at long-range distances, from 10 Mb to 60 Mb. Consequently, the inactivated X chromosome is more compacted in higher-order chromatin structures, consistent with the observation that active chromatin regions tend to interact¹⁸.

We quantified the pairwise chromosomal interactions within (intra-chromosomal *cis* contacts) and between (interchromosomal *trans* contacts) chromosomes to measure spatial distance between the X chromosome and autosomes. As expected, the intrachromosomal *cis* contacts are more prevalent than the interchromosomal *trans* contacts (Supplementary Fig. 21). Furthermore, the interchromosomal *trans* contacts are less frequent, specifically on the X chromosome (Fig. 4e), indicating a more distal spatial position of the X chromosome compared to autosomes within chromosomal territories. The DNA FISH assay of a single-copy X-linked gene (*Pex5*) reveals that the X chromosome shows greater condensation than autosomes (Fig. 4f). These results suggest that the inactivated X chromosome is farther from the transcription-associated hubs at nuclear speckles, which are connected closely among activated autosomes¹⁹.

The global reduction of H4K16ac signals of X-linked genes in testes is in accordance with the loss of expression balance between male and female locusts, indicating chromatin suppression of gene expression on the X chromosome in male locusts (Fig. 4g). The early germ cells displayed an enrichment of H4K16ac (Fig. 4h), whereas the spermatocytes, undergoing meiotic division, exhibited a depletion of H4K16ac, corresponding to the global reduction of H4K16ac signals in X-linked genes in the testes²⁰. The binary classification of A (active transcriptional state) and B (inactive transcriptional state) compartments showed that different H4K16ac distribution was correlated strongly with transcription activity (Fig. 4i and Supplementary Fig. 22). The number of high expressed genes in compartment A in the X chromosome (0.572 Mb^{-1}) was lower than that in autosomes (0.915 Mb^{-1}). These results show that the loss of expression balance of the X chromosome in testes might be correlated with interchromosomal spatial distance, compartment organization and H4K16ac depletion. Therefore, the relatively remote spatial position of the gigantic X chromosome, decorated with distinct chromatin structures, results in the formation of segregated chromosome territories compared to autosomes during meiotic silencing.

Fig. 3 | Chromatin differentiation of locust phase changes. **a**, Principal component analysis of seven histone modifications. **b**, Pearson correlation analysis between chromatin signals and the phase-related gene expression of gregarious and solitary locusts. **c**, Association of transcription factors (TFs) with differential histone modifications near promoters. Differentially expressed (DE) genes, which were positively and negatively correlated with the differential H3K4me3/H3K27ac at promoters and the differential H3K4me1/H3K27me3 situated 10 kb upstream of the promoters, respectively, were included in this analysis. TF weight is calculated based on the frequencies of nonredundant TF binding sites. Positive and negative regulation refer to the relationships between histone modification abundance and gene expression, respectively. **d**, Combinatorial chromatin signals of *Henna*. The y axis was scaled separately for different sequencing methods. **e**, Quantitative 3C assays to assess the spatial interaction between EHI and TSS1 in gregarious locusts. Control DNA represents the fragments amplified by two internal primers between two enzyme sites. Digested DNA represents random ligated genomic

DNA. The ligation site represents the site where EHI is ligated to TSS1; $n = 3$ independent biological replicates. Data are shown as mean \pm s.e.m. $P = 0.0131$; $*P < 0.05$. **f**, Knockout EHI by the CRISPR–Cas9 system in gregarious locusts. EHI^{+/+} and EHI^{-/-} indicate heterozygous and homozygous mutants, respectively; $n = 1$ independent biological replicates for demonstration purpose. **g**, qPCR quantification of *Henna* mRNA expression in wild type, heterozygous mutants and homozygous mutants. At least $n = 3$ independent biological replicates. Data are shown as mean \pm s.e.m. P values for comparisons of wild type with EHI^{+/+} and EHI^{-/-} are 0.0004 and 0.0000, respectively. $*P < 0.05$. **h**, Western blot of *Henna* protein expression; $n = 3$ independent biological replicates. Data are shown as mean \pm s.e.m.; $P = 0.017$, $*P < 0.05$. **i**, Arena behavioral assays for quantifying phase-related behavior. Independent biological replicates; $n = 15$ (EHI^{-/-}), $n = 7$ (EHI^{+/+}) and $n = 20$ (wild type). A P_{greg} (probabilistic metric of gregariousness) close to 0 indicates typical gregarious behavior; $P = 0.0014$ for EHI^{+/+} and $P = 0.0414$ for EHI^{-/-}. UTR, untranslated region.



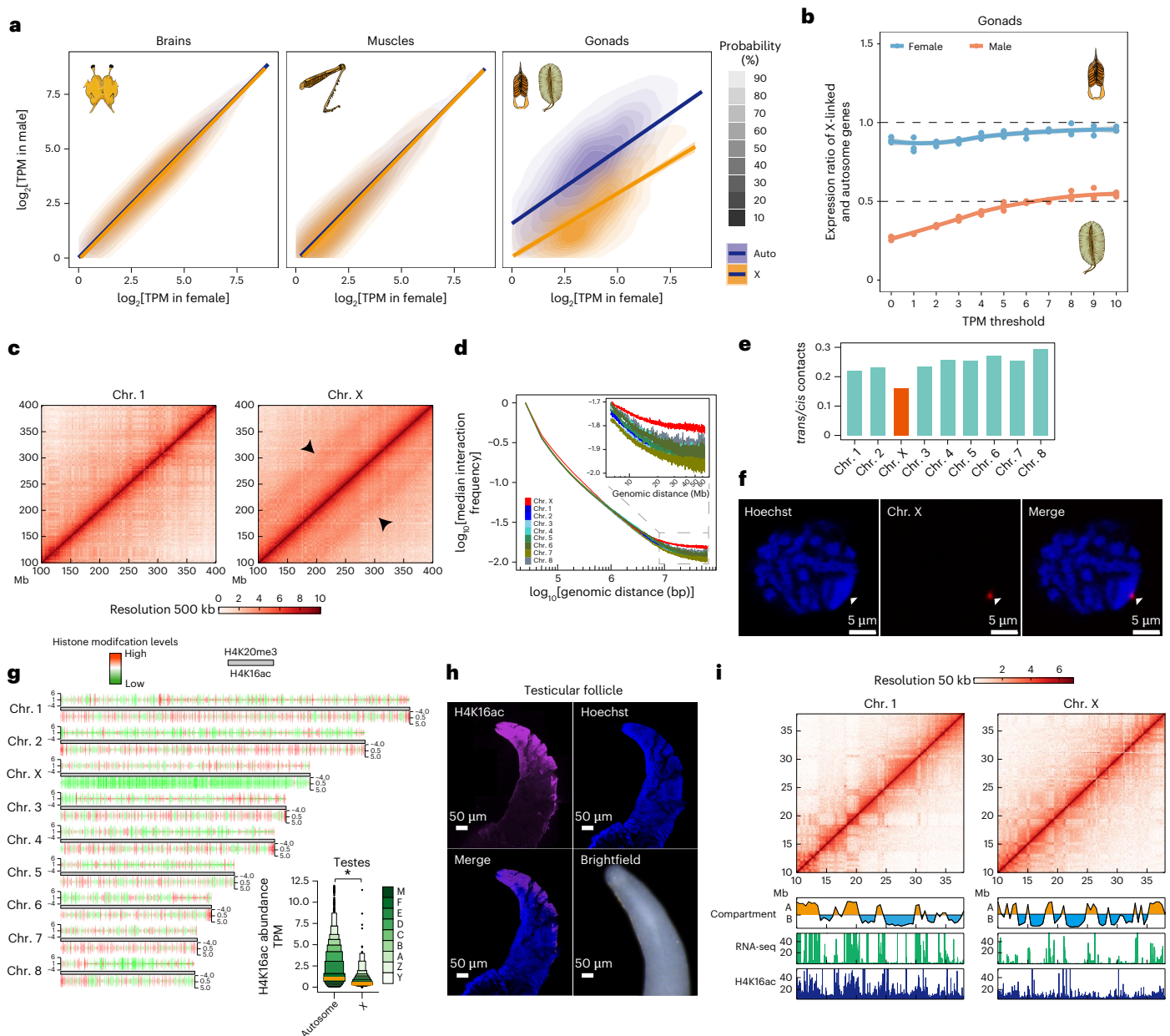


Fig. 4 | Global interaction and chromatin change of the X chromosome during meiotic silencing. a, Correlation of gene expression in brains, leg muscles and gonads. Error bands, 95% confidence interval of the linear model fit. P values $< 2.2 \times 10^{-16}$, Pearson correlation test. **b**, Expression ratios between X-linked and autosomal genes in gonads under different minimum TPM cutoffs. Error bands, 95% confidence interval around the fitted smoothing curve. **c**, Pearson correlations of Hi-C interaction frequency matrices of chr. 1 and chr. X at 500-kb resolution. The intensity of each pixel represents the normalized frequency of interaction between a pair of genomic loci. Arrowheads indicate parallel lines along the main diagonal of the interaction map. **d**, A log-log plot of interaction frequency decay with distance in testes. We assumed independence between the X chromosome and the autosomes. Normalized Hi-C data binned at 25-kb scale were converted to interaction frequencies. The x axis shows the \log_{10} genomic distance, and the y axis shows the median \log_{10} Hi-C interaction

frequency for any given genomic distance. **e**, Relative enrichments of *trans* contacts by calculating the ratio of the number of *trans*-over *cis*-mapping read pairs for each chromosome. **f**, Condensation of the X chromosome in a meiotic cell of spermatocytes in testes. A series of 24 DNA probes (the single-copy X-linked gene *Pex5*) was selected randomly. Nuclei were stained with Hoechst 33342; $n = 2$ independent biological replicates. The position of the X chromosome is indicated by arrowheads. **g**, Read abundances of H4K16ac and H4K20me3 are shown in 1-Mb bin units. $P = 0.0000$, $*P < 0.05$. **h**, The distribution and abundance of H4K16ac in testicular follicles of adult locusts. Nuclei were stained with Hoechst 33342 staining; $n = 2$ independent biological replicates. **i**, Association of Hi-C interaction frequency with gene expression and H4K16ac signals in chr. 1 and chr. X. The A and B compartments, determined by the first principal component (PC1), reflected the transcription and H4K16ac state on a linear genomic scale.

Dosage compensation of the X chromosome in soma

In brains and leg muscles, X-linked gene expression in female and male locusts was equal to that of the autosomal genes, indicating complete dosage compensation in male locusts (Fig. 5a). X chromosome expression in female locusts can be achieved through biallelic expression of both alleles or by random inactivation of one allele with

hyperactivation of the other in animals²¹. The transcriptome analysis of heterozygosity in a female individual revealed biallelic expression of X-linked genes in female locusts (Fig. 5b). To eliminate potential influences from A-to-I RNA editing, the results, even after removing adenine sites, robustly supported the biallelic expression of X-linked genes in female soma.

We then profiled the genomic distributions of H4K20me1, H4K16ac, H3K27me3 and H3K9me3 to determine whether these chromatin modifications are involved with dosage compensation in locusts. H4K16ac was more prevalent in the genic region compared to H4K20me1, H3K27me3 and H3K9me3 (Fig. 5c). In autosomes, the equal genic and intergenic signals between female and male locusts for these four chromatin marks corresponded to the balanced autosomal gene expression in brains (Fig. 5d). However, within the genic region of the X chromosome, only the density of female/male enrichment ratios of H4K16ac signals centered around 1, consistent with its role in complete dosage compensation. All four histone modifications showed similar levels between autosomal and X-linked genes across genic regions in females (Fig. 5e). To account for the difference in chromosome numbers between females and males, we estimated male data (the abundance of histone modification in male brains) theoretically assuming no dosage changes in females. As expected, the H4K16ac levels were comparable between the autosomal and X-linked genes in males, aligning with the well-recognized role of H4K16ac in establishing dosage compensation in heterogametic organisms²². In fact, the enrichment levels of H4K16ac on X-linked genes in males were higher than those theoretically estimated for male data. Different from the widespread distribution of H4K16ac across the entire gene body, H4K20me1 localized mainly to the 5' end of the gene body. However, the H4K20me1 level of X-linked genes in males was more depleted than that of the theoretically estimated male data. In fruit flies, the absence of Sex-lethal (SXL) expression in males permits MSL complex assembly, leading to H4K16ac enrichment on the X chromosome for dosage compensation²³. Despite the close association between H4K20me1 and MSL complex, we found no significant differences in RNA and protein expression of SXL and MSL subunits between females and males, indicating an SXL-independent initiation of dosage compensation in locusts (Supplementary Figs. 23 and 24 and Supplementary Note 5). Overall, the depletion of H4K20me1 and the enrichment of H4K16ac suggest that they are connected functionally with maintaining balanced expression of the X chromosome in the somatic organs of male locusts.

Different from H4K20me1, H3K27me3 and H3K9me3, levels of H4K16ac decreased with the increase of the distance to the nearest gene in the intergenic region on the X chromosome (Fig. 5f), indicating a distance-dependent dosage effect of H4K16ac and partial dosage compensation. Different from the fruit fly²⁴, we did not detect a dominant immunofluorescence signal of H4K16ac across the entire male X genome (Fig. 5g). To further confirm the uniqueness of distance-dependent H4K16ac enrichment in the large genome of locusts, we analyzed H4K16ac profiling data from the fruit fly²⁵. In contrast to the locust genome, we observed complete dosage compensation in both genic and intergenic regions of the compact genome of the fruit fly (Fig. 5h). As expected, the distance-dependent H4K16ac enrichment was not observed on the X chromosome of the fruit fly (Fig. 5i). Therefore, the dosage effects of H4K16ac, initiated from genic region, faded gradually during subsequent spreading into intergenic regions, which constrain the spreading of dosage effects in locusts.

Dosage compensation on translocated X-linked genes

By examining the conservation of X-linked genes across insects at order level, we found that the gene configuration of the locust X chromosome is the result of extensive interchromosomal gene exchanges through drastic genome rearrangement (Supplementary Note 6 and Supplementary Figs. 25–27). 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). We determined the pairwise chromosome associations of the two locusts, the pygmy grasshopper (*Zhengitettix transpicula* (ZT), Caelifera, Tetrigidae) and the pygmy mole cricket (*Xya riparia* (XR), Caelifera, Tridactyloidea), which belongs to a basal taxon of Caelifera. The chromosomes of the migratory locust are identical to those of the desert locust but exhibit one-to-several or one-to-zero correspondence with those of the pygmy grasshopper and the pygmy mole cricket, demonstrating that genome size expansion accompanies large-scale chromosome rearrangement during the early diversification of Caelifera species (Fig. 6a and Supplementary Figs. 28–30). The X-linked genes of the migratory locust have largely persisted on the X chromosome since the emergence of Caelifera (Fig. 6b), although it cannot be concluded whether the X chromosome of the pygmy mole cricket represents the ancestral X chromosome of Caelifera or has undergone a chromosome fusion event between an autosome and a sex chromosome. To investigate the chromosomal rearrangement of X-linked gene of the migratory locust, the one-to-one orthologs among the four species were divided into the five gene categories: ancient autosome (LMAxSGaXRaZTa), ancient X (LMxSGxXRxZTx), ancient AtoX (translocated from Autosomes to X chromosome before the divergence of two locust species, LMxSGxXRaZTa), ancient XtoA (LMAxSGaXRxZTx) and recent AtoX (LMxSGaXRaZTa). Compared with the ancient AtoX and ancient X categories, the lower ratios of female/male in H4K20me1 signals showed insufficient depletion of H4K20me1 in the recent AtoX category. Furthermore, the recent AtoX category showed a higher ratio of female/male in H4K16ac signals than the ancient AtoX and ancient X categories (Fig. 6c). This suggests an incomplete recovery of dosage compensation of recently translocated X-linked genes in the migratory locust.

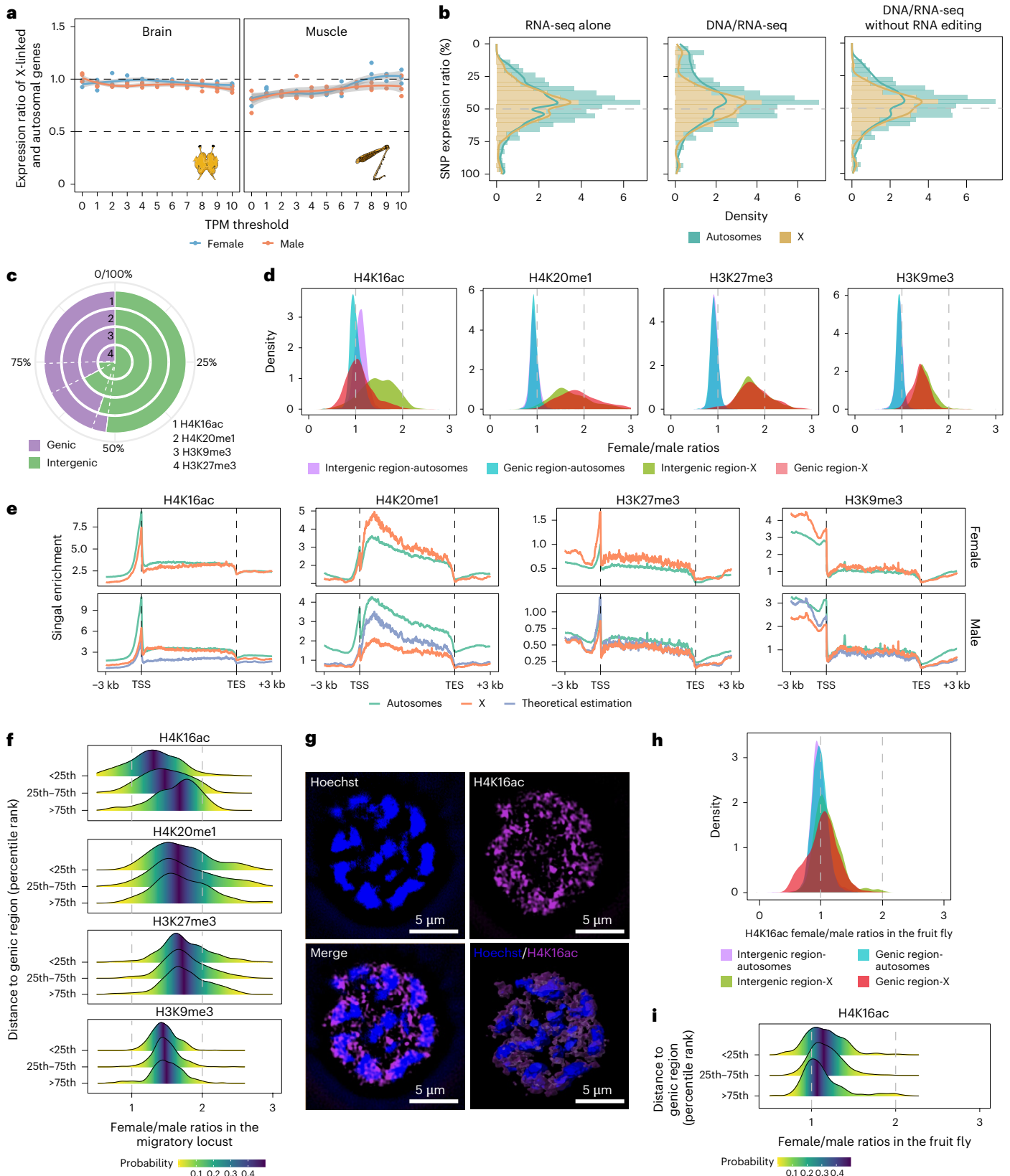
Compared to the ancient X and the two autosomal categories, the ancient AtoX category, but not the recent AtoX category, had significantly higher dN/dS ratios (Fig. 6d), suggesting that adaptive evolution is more likely to have affected the ancient AtoX category. Accordingly, the ancient AtoX category displayed a broader range and higher levels of histone modification than the recent AtoX category (Fig. 6e). This implies that X-linked translocations have the potential to accumulate sequence divergence or to maintain mutations that cannot be effectively eliminated by recombination, ultimately contributing to gene function innovation over time. In the recent AtoX category, several genes, including *Oat* and *msps*, were reported previously to be associated with functions that harm males or benefit females^{26,27}. Therefore, sexually antagonistic selection promotes the relocation of female-related genes from autosomes

Fig. 5 | Balanced expression of the X chromosome in soma. **a**, Expression ratios between X-linked and autosomal genes in brains and muscles under different minimum TPM cutoffs for gene inclusion. Error bands, 95% confidence interval around the fitted smoothing curve. **b**, Allelic X chromosome expression in a female. To ensure accurate identification of heterozygous alleles, the biallelic heterozygosity sites were determined using DNA-seq for the same individual locust that was subjected to RNA-seq. **c**, Genomic prevalence of the histone modifications H4K20me1, H4K16ac, H3K27me3 and H3K9me3. The genomic prevalence was determined by the average portion of the reads located in the genic and intergenic regions. **d**, Density distribution of the ratios (female/male) of histone modification abundances in 200-kb bins between the two sexes in brains. **e**, Enrichment profiles of histone modification of H4K20me1, H4K16ac, H3K27me3 and H3K9me3 across genic region. The theoretical estimation

in males was inferred assuming no dosage changes in females. **f**, Density distribution of the ratios (female/male) of histone modification abundances between the two sexes in different genomic regions in brains of the migratory locust. The different genomic regions were classified based on the distance to the nearest genes. **g**, Immunofluorescence staining of H4K16ac in brains. Nuclei were stained with Hoechst 33342 staining; $n = 2$ independent biological replicates. **h**, Density distribution of the ratios (female/male) of histone modification abundances between the two sexes in the fruit fly. The average chromatin immunoprecipitation sequencing signals in 2-kb bins were compared between female and male locusts. **i**, Density distribution of the ratios (female/male) of histone modification abundances between the two sexes in different genomic regions in 2–4 h embryos of the fruit fly. The different genomic regions were classified based on the distance to the nearest genes.

to X chromosome in locusts (Supplementary Note 7). Consequently, these genes can be repressed by meiotic sex chromosome inactivation, facilitating to mitigate female antagonism and to increase male fitness during spermatogenesis. Therefore, the translocated X-linked genes in the migratory locust are likely to be convergently recruited to the X chromosome and subject to relocation of dosage

compensation due to the need to suppress their functional roles favoring females during spermatogenesis. Taken together, the recent AtoX category showed distinct differences in histone modification, suggesting a lag in the relocation of dosage compensation for the X-linked genes that were recently translocated autosomes in the migratory locust.



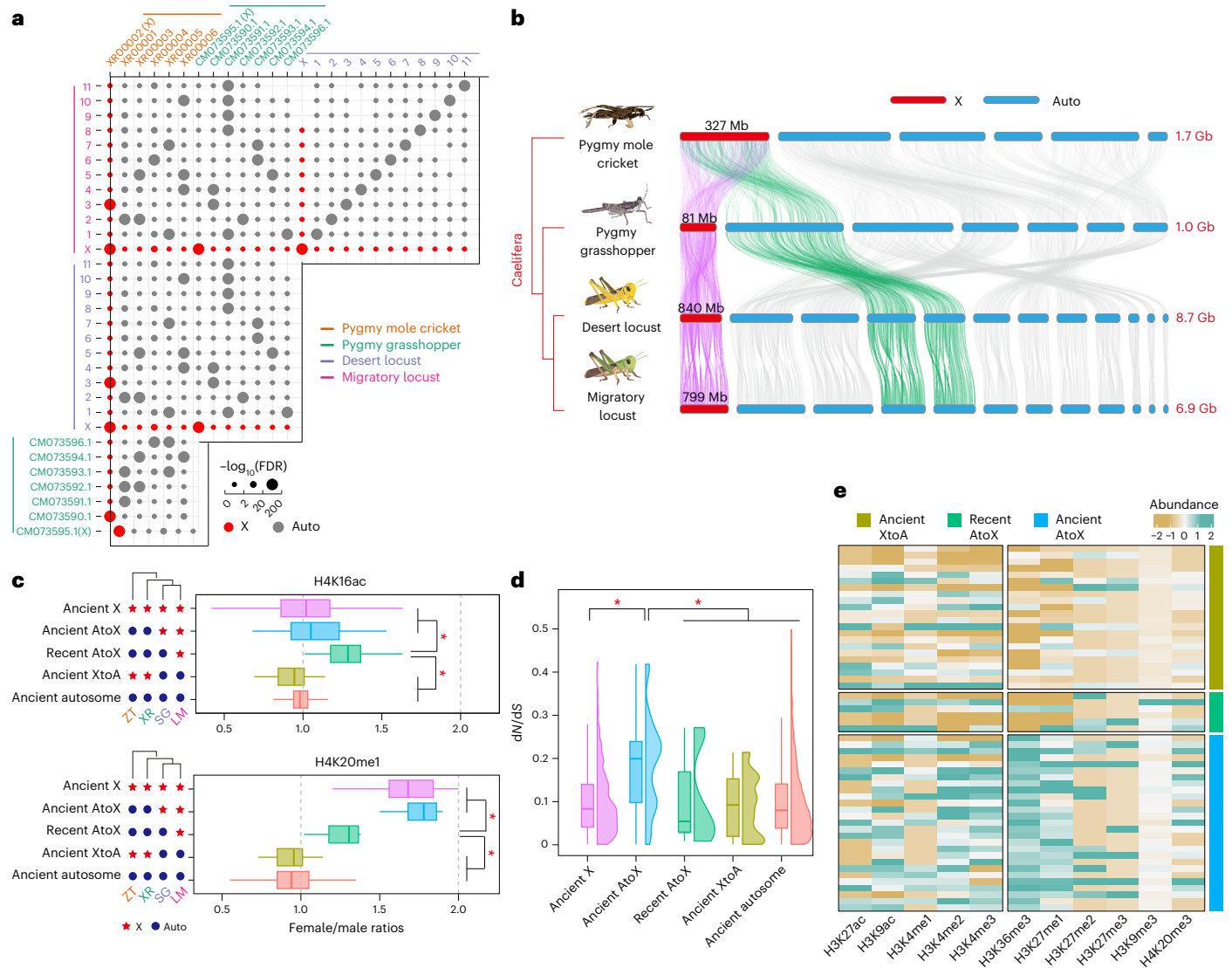


Fig. 6 | Dosage compensation turnover on X-linked gene derived from autosomes. a, Pairwise dotplots showing the significance of chromosome–chromosome associations among the migratory locust, the desert locust, the pygmy grasshopper and the pygmy mole cricket. **b**, One-to-one ortholog linking among the chromosomes of four orthopteran species. Vertical lines connect orthologs across the four species. Only connections between chromosome pairs with significantly associated associations are shown. **c**, Ratios (female/male) of genic H4K16ac and H4K20me1 signals between the two sexes. Reads per kilobase million values were used to determine the histone modification signals. The one-to-one orthologs of the migratory locust were divided into the six gene categories: ancient autosome (LMaSGaXRaZTa, $n = 4,661$), ancient X (LMxSGxXRaZTx, $n = 314$), ancient AtoX (LMxSGxXRaZTa, $n = 27$), ancient XtoA (LMaSGaXRaZTx, $n = 22$) and recent AtoX (LMxSGaXRaZTa, $n = 6$); n , number of one-to-one orthologs. The recent XtoA category was not included due to the

absence of any ortholog. Boxplots indicate the median (center line), the first and third quartiles (box limits), and whiskers extending to 1.5x the interquartile range. The P values for comparisons of recent AtoX with ancient X, ancient AtoX, ancient XtoA and ancient autosomes were (0.0151, 0.0011), (0.0403, 0.0017), (0.0023, 0.0017) and (0.0017, 0.0011) for H4K16ac and H4K20me1, respectively; $*P < 0.05$. **d**, dN/dS comparison in the five gene categories. dN/dS , the ratio of nonsynonymous substitutions per nonsynonymous site (dN) to synonymous substitutions per synonymous site (dS). Boxplots indicate the median (center line), the first and third quartiles (box limits), and whiskers extending to 1.5x the interquartile range. The P values for comparisons of ancient AtoX with ancient X, recent AtoX, ancient XtoA and ancient autosomes were 2×10^{-6} , 0.0359, 0.0002 and 2×10^{-6} , respectively; $*P < 0.05$. **e**, Heatmap showing abundance of histone modification signals for the recent AtoX, ancient XtoA and ancient AtoX categories.

Discussion

In this study, an epigenomic map, based on a high-quality genome, allowed us to accurately identify chromatin modifications of spacious noncoding regions of locusts in an unprecedented manner due to improved assembly completeness in intronic and intergenic regions. The increased number of enhancers in expanded intronic regions and the distance-dependent H4K16ac dosage compensation in expanded intergenic regions reveal distinct epigenomic characteristics in the dark matter of large-sized genomes.

We observed that the phase-related histone modification differences decorating promoters and enhancers were correlated with differential transcription between solitary and gregarious phases, underscoring the significant influences of chromatin regulation on determining alternative developmental programming in locusts^{2,28}. Indeed, the involvement of H3K4me3 and H3K27ac on promoters has also been confirmed previously in caste determination of honey bees and carpenter ants, suggesting that these histone modifications play a conserved role in the phenotypic plasticity of a wide range of

insect species^{29,30}. We found that the *cis*-regulatory element EHI functions specifically as an enhancer to regulate the behavioral plasticity of locusts. We also identified an EHI-like sequence in the homologous intron of *Henna* in the desert locust (Supplementary Fig. 31), but not in the two *Caeliferan* insects or in the three insects with small-sized genomes. Since *Henna* has been reported to be involved in the regulation of phenotypic plasticity in the honey bee³¹, we further checked its epigenomic data but did not find a positionally homologous enhancer of *Henna*³². These results indicate that EHI may be present only in locust genomes and may contribute solely to the specialized regulatory function of behavioral plasticity in locusts. Therefore, these findings reveal that, accompanying genome size expansion, an increasing number of enhancers have been recruited in large-sized genes in locusts.

Different from the fruit fly, where the male X chromosome exhibits predominant immunostaining intensity of H4K16ac, we did not observe this phenomenon in male locusts³³. Despite H4K16ac not exhibiting preferential enrichment on the male X chromosome in mosquitoes, a common characteristic of dosage compensation mechanisms is the elevated H4K16ac levels on dosage-compensated sex chromosomes²⁴. The intergenic regions occupy a broader area in the locust X chromosome than in the fruit fly X chromosome. TEs, which make up 71.1% of the locust X chromosome, may play a role in assembling and organizing constitutive heterochromatin in intergenic regions by providing docking sites for protein complexes associated with repressive histone modifications or through the presence of hotspots for TE insertions^{34,35}. Therefore, we found a mosaic pattern of immunostaining intensity, similar to that on autosomes, rather than a prevailing intensity along the X chromosome. These results are also supported by the distance-based decoration of H4K16ac outside genic regions and the lack of H4K16ac enrichment in intergenic regions in locusts. Therefore, the distance-dependent H4K16ac enrichments in intergenic regions demonstrate peculiarly large genome effects on dosage compensation.

Our results ruled out the long-term conservation of X-linked genes due to extensive interchromosomal gene exchanges through drastic genome rearrangement (Supplementary Note 6 and Supplementary Figs. 25–27). A recent study published another chromosome-level genome assembly of the migratory locust³⁶. However, this present study presents different conclusions from the latter study regarding the conservation of X-linked genes. The discrepancy is because of the use of different criteria to define conservation levels. The numbers of X-linked genes vary greatly between the migratory locust (1,985 and 12 pairs of chromosomes) and the fruit fly (80 and four pairs of chromosomes). They identified only 35 X-linked homologs, which may provide limited evidence for concluding that the X chromosome has maintained long-term gene content conservation between these two insect species. Indeed, our results show that only ten X-linked orthologs are present on the sex chromosomes of all the four insect orders (Supplementary Figs. 26 and 27). These findings suggest that, throughout insect evolution, there has been little evidence of strong selective pressure to preserve a substantial proportion of X-linked genes on the X chromosome. The incomplete recovery of H4K16ac-mediated dosage compensation and insufficient depletion of H4K20me1, as well as restricted histone modification decoration in the recent AtoX category indicate a lagging relocation of dosage compensation on recently translocated X-linked genes. This contrasts with the re-established compensation and abundant histone decoration in the ancient AtoX category, indicating that the dosage compensation effects of flanking pre-existing X-linked genes cannot spread efficiently to the newly translocated X-linked genes. The spreading of H4K16ac, catalyzed by the MSL complex, to the flanking intergenic and genic regions lacking MSL complex binding is important for dosage compensation²². Since the MSL complex loads at chromatin entry sites, mainly in introns and nearby 3-kb genic regions⁷, the limited efficiency of H4K16ac spread may require a greater number of chromatin entry sites in the locust genome compared to smaller genomes. Consequently, the newly translocated

X-linked genes in the locust genome may not be able to quickly establish recognition by the MSL complex. Therefore, distance-dependent dosage compensation in intergenic region results in a lagging relocation of dosage compensation to X-linked genes recently translocated from autosomes in large genomes.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-025-02330-y>.

References

- Wang, X. & Kang, L. Molecular mechanisms of phase change in locusts. *Annu Rev. Entomol.* **59**, 225–244 (2014).
- Ma, Z., Guo, W., Guo, X., Wang, X. & Kang, L. Modulation of behavioral phase changes of the migratory locust by the catecholamine metabolic pathway. *Proc. Natl Acad. Sci. USA* **108**, 3882–3887 (2011).
- Yang, M. et al. MicroRNA-133 inhibits behavioral aggregation by controlling dopamine synthesis in locusts. *PLoS Genet.* **10**, e1004206 (2014).
- Gagnidze, K., Weil, Z. M. & Pfaff, D. W. Histone modifications proposed to regulate sexual differentiation of brain and behavior. *Bioessays* **32**, 932–939 (2010).
- Meyer, B. J. The X chromosome in *C. elegans* sex determination and dosage compensation. *Curr. Opin. Genet. Dev.* **74**, 101912 (2022).
- Vielle, A. et al. H4K20me1 contributes to downregulation of X-linked genes for *C. elegans* dosage compensation. *PLoS Genet.* **8**, e1002933 (2012).
- Straub, T., Grimaud, C., Gilfillan, G. D., Mitterweger, A. & Becker, P. B. The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. *PLoS Genet.* **4**, e1000302 (2008).
- Wiles, E. T. & Selker, E. U. H3K27 methylation: a promiscuous repressive chromatin mark. *Curr. Opin. Genet. Dev.* **43**, 31–37 (2017).
- Kraft, K. et al. Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. *Proc. Natl Acad. Sci. USA* **119**, e2201883119 (2022).
- Ernst, U. R. et al. Epigenetics and locust life phase transitions. *J. Exp. Biol.* **218**, 88–99 (2015).
- Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
- Vorobyeva, N. E., Krasnov, A. N., Erokhin, M., Chetverina, D. & Mazina, M. Su(Hw) interacts with Combgap to establish long-range chromatin contacts. *Epigenetics Chromatin* **17**, 17 (2024).
- Chetverina, D., Erokhin, M. & Schedl, P. GAGA factor: a multifunctional pioneering chromatin protein. *Cell. Mol. Life Sci.* **78**, 4125–4141 (2021).
- Reddington, J. P. et al. Lineage-resolved enhancer and promoter usage during a time course of embryogenesis. *Dev. Cell* **55**, 648–664 (2020).
- Kvetnansky, R., Sabban, E. L. & Palkovits, M. Catecholaminergic systems in stress: structural and molecular genetic approaches. *Physiol. Rev.* **89**, 535–606 (2009).
- Duan, J. et al. CLAMP and Zelda function together to promote *Drosophila* zygotic genome activation. *eLife* **10**, e69937 (2021).
- Liu, Q., Jiang, F., Zhang, J., Li, X. & Kang, L. Transcription initiation of distant core promoters in a large-sized genome of an insect. *BMC Biol.* **19**, 62 (2021).

18. Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
19. Monahan, K., Horta, A. & Lomvardas, S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* **565**, 448–453 (2019).
20. Arafat, E. A., El-Samad, L. M., Moussian, B. & Hassan, M. A. Insights into spermatogenesis in the migratory locust, *Locusta migratoria* (Linnaeus, 1758)(Orthoptera: Acrididae), following histological and ultrastructural features of the testis. *Micron* **172**, 103502 (2023).
21. Disteché, C. M. Dosage compensation of the sex chromosomes. *Annu Rev. Genet.* **46**, 537–560 (2012).
22. Gelbart, M. E., Larschan, E., Peng, S., Park, P. J. & Kuroda, M. I. *Drosophila* MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nat. Struct. Mol. Biol.* **16**, 825–832 (2009).
23. Samata, M. & Akhtar, A. Dosage compensation of the X chromosome: a complex epigenetic assignment involving chromatin regulators and long noncoding RNAs. *Annu. Rev. Biochem.* **87**, 323–350 (2018).
24. Keller Valsecchi, C. I., Marois, E., Basilicata, M. F., Georgiev, P. & Akhtar, A. *Life Sci. Alliance* **4**, e202000996 (2021).
25. Rieder, L. E., Jordan, W. T. 3rd & Larschan, E. N. Targeting of the dosage-compensated male X-chromosome during early *Drosophila* development. *Cell Rep.* **29**, 4268–4275 (2019).
26. Levillain, O., Diaz, J. J., Blanchard, O. & Dechaud, H. 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 (2005).
27. Lu, W., Lakonishok, M. & Gelfand, V. I. The dynamic duo of microtubule polymerase mini spindles/XMAP215 and cytoplasmic dynein is essential for maintaining *Drosophila* oocyte fate. *Proc. Natl Acad. Sci. USA* **120**, e2303376120 (2023).
28. Yang, P., Hou, L., Wang, X. & Kang, L. Core transcriptional signatures of phase change in the migratory locust. *Protein Cell* **10**, 883–901 (2019).
29. Simola, D. F. et al. A chromatin link to caste identity in the carpenter ant *Camponotus floridanus*. *Genome Res.* **23**, 486–496 (2013).
30. Wojciechowski, M. et al. Phenotypically distinct female castes in honey bees are defined by alternative chromatin states during larval development. *Genome Res.* **28**, 1532–1542 (2018).
31. Zayed, A., Naeger, N. L., Rodriguez-Zas, S. L. & Robinson, G. E. Common and novel transcriptional routes to behavioral maturation in worker and male honey bees. *Genes Brain Behav.* **11**, 253–261 (2012).
32. 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.* **50**, 11550–11562 (2022).
33. Turner, B. M., Birley, A. J. & Lavender, J. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. *Cell* **69**, 375–384 (1992).
34. Marsano, R. M. & Dimitri, P. Constitutive heterochromatin in eukaryotic genomes: a mine of transposable elements. *Cells* **11**, 761 (2022).
35. Lawson, H. A., Liang, Y. & Wang, T. Transposable elements in mammalian chromatin organization. *Nat. Rev. Genet.* **24**, 712–723 (2023).
36. Li, X., Mank, J. E. & Ban, L. The grasshopper genome reveals long-term gene content conservation of the X Chromosome and temporal variation in X Chromosome evolution. *Genome Res.* **34**, 997–1007 (2024).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Methods

Ethics declarations

No ethical approval was required for the use of locusts in this study.

Insects

The locust colonies were reared under a 14 h/10 h light/dark photo regimen at 30 °C and were fed fresh wheat seedlings and wheat bran. An adult female individual was dissected to remove intestinal tracts, which were frozen immediately in liquid nitrogen and stored at –80 °C until DNA extraction.

Chromosome staining and estimation of chromosome size

To obtain embryos for chromosome staining analysis, locust eggs were dissected from egg pods to isolate embryos. The embryos were immersed in 0.04% colchicine for 90 min followed immediately by hypotonic treatment for 30 min, and were then immersed in 3:1 ethanol–acetic acid for 90 min. Afterward, the embryos were immersed in 80% acetic acid for 2 min and dispersed and suspended gently for 3 min. The cell suspension was aspirated into a micropipette and dropped onto a clean slide at 60 °C to dry the solution. The slide was stained with Hoechst 33342 (Thermo Fisher Scientific) for 5 min and visualized using a confocal microscopy system (Zeiss LSM 710). We measured the chromosome bivalents from one female embryonic cell at metaphase of mitosis. All the bivalents were well separated. The pixels of each chromosome were measured by the Image J software and arranged in order of decreasing value.

Genomic library construction and DNA sequencing

The DNA samples used for genomic sequencing were obtained using a sodium dodecyl sulfate method. Initially, the tissues were ground carefully with liquid nitrogen, and the cell nucleus was cleaned subsequently using a modified homogenization buffer. Following this step, the residual material was washed with 1× PBS buffer (Invitrogen) and the cell nucleus was lysed using a total lysis buffer. Subsequently, RNA and protein were eliminated using RNase A (Qiagen) and protease K (Qiagen), respectively. Finally, the DNA was eluted using an elution buffer (Qiagen) after phenol/chloroform extraction. For the migratory locust, two libraries with insert sizes of 20 kb and 40 kb were constructed separately using SMRTbell Template Preparation 1.0 using BluePippin size-selection (Sage Science), and were subsequently sequenced using on a PacBio Sequel instrument with the Sequel Sequencing Kit v.2.1. To obtain the optical mapping data, genomic DNA was labeled using the Bionano Prep DNA Labeling Kit-DLS, and loaded into a nanochannel array on a Saphyr instrument (Bionano Genomics). For the desert locust, DNA extraction was constructed for Nanopore libraries and then sequenced using a PromethION platform. The sequencing libraries were constructed using the ligation sequencing kit (SQK-LSK109) and loaded onto FLO-PRO002 flow cells. The genomic DNA was also used to construct a paired-end library using Nextera XT DNA Library Prep Kit (Illumina). This short-read library was sequenced on an Illumina HiSeq X Ten platform.

DNA labeling for optical mapping

Direct labeling was performed using the Direct Labeling and Staining Kit (Bionano Genomics, catalogue number 80005) according to the manufacturer's manual. In brief, high-molecular-weight DNA was incubated at 37 °C with DLE labeling mix. Following proteinase K digestion and cleanup of the unincorporated DL-Green label, the labeled DNA was prestained, homogenized and quantified using on a Qubit Fluorometer (Thermo Fisher). The labeled sample was loaded into a nanochannel array of flow cells, imaged and digitized in a Saphyr instrument (Bionano Genomics) according to the manufacturer's instructions.

Genome assembly

GenomeScope v.1.0.0 was used to conduct a K-mer frequency survey of genome composition based on 21-mer frequency distribution to

estimate genome size and heterozygosity. For the migratory locust, The raw PacBio reads were corrected and trimmed using Canu v.1.7³⁷, and the contig-level assembly were performed using FALCON v.2.0.4 followed by the FALCON-Unzip v.0.4.0 module³⁸. The remaining alternative haplotypes from the initial primary contigs were removed using Purge Haplotigs v.1.0.0, and haplotype switch errors were eliminated using the FALCON-Phase v.0.1.0³⁹. All PacBio reads were aligned to the resulting assemblies using BLASR v.5.3.3 and then the consensus sequences were called using the Arrow program in SMRT link v.7.0. The PacBio-corrected assemblies were further improved by two rounds of polishing with Pilon v.1.23 using paired-end Illumina data⁴⁰. Data visualization, de novo assembly and hybrid scaffolding of the optical map data with a minimum size of 150 kb were performed using the Bionano Access v.1.2.1 and Bionano Solve v.3.2.1 software. Gaps in the hybrid-scaffolded assembly were closed using PBjelly from PBSuite v.15.8.24 using PacBio and Nanopore read datasets⁴¹. We subsequently reperformed error correction procedures to polish the sequences in the gap regions. For the desert locust, de novo genome assemblies were generated by NextDenovo v.2.4.0 and NextPolish v.1.3.1 using long-read Nanopore sequencing.

Scaffolding of a chromosome-level genome assembly

For Hi-C experiments, tissues were cross-linked with 1% fresh formaldehyde and then quenched with 0.2 M glycine for 5 min. The resulting cells were incubated with 0.1% SDS then 1% Triton X-100. Genomic DNA was digested with the *DpnII* restriction enzyme (NEB) and incubated at 37 °C overnight. The cohesive ends were filled in with biotin-14-dATP and 5 U μl^{-1} Klenow fragment of DNA polymerase I (NEB), followed by a 90 min incubation at 37 °C with gentle rotation. The resulting DNA fragments were then ligated by adding a mix of T4 DNA ligase (Thermo Fisher Scientific), 10% Triton X-100 and bovine serum albumin (NEB). The cross-linking was reversed with proteinase K and the biotinylated fragments were pulled down using Dynabeads MyOne Streptavidin C1 beads (Thermo Fisher Scientific). The Hi-C library for Illumina sequencing was prepared using the NEBNext Ultra II DNA library Prep Kit for Illumina (NEB). Trimmed Hi-C reads were aligned to the polished genome assembly using bwa v.0.7.17 and filtered for Hi-C-specific artifacts using the HiC-Pro v.2.11.1 pipeline^{42,43}. The Juicer v.1.5.6 and 3D-DNA v.180922 pipelines were used to group, order and orientate scaffolds and to generate a chromosome-level genome assembly. The potential misjoins, translocations and inversions within the boundaries of each chromosome were corrected by manually introducing breakpoints at the mis-assembly sites based on visualization with Juicebox v.1.5.3⁴⁴.

Sex chromosome identification

Because both male and female locusts have the same complement of autosomes, the read coverage of autosomes is expected to be roughly equal in the male and female data in XO/XX sex determination. The sex chromosome was identified by using a chromosome quotient method that relies on the simple assumption that female locusts have a twofold increase of read coverage compared to male locusts⁴⁵. The DNA-seq reads from an individual of both a male and female locust were aligned separately to the genome assembly using the Bowtie2 v.2.4.1 program⁴⁶. The resulting bam files were filtered to remove low-quality alignments using the SAMtools v.1.9 program⁴⁷. Contigs derived from the X chromosome were also identified in the contig-level assembly to verify the mis-scaffolding links in Hi-C scaffolding.

Analysis of transposable elements

Transposable elements and simple repeats were predicted de novo and classified using the RepeatModeler v.2.01 package with RMBlast v.2.9.0-p2 engine⁴⁸. Other prerequisite modules including RECON v.1.08, RepeatScout v.1.0.6, TRF v.4.0.9, Dfam v.2.0 and Repbase-derived libraries were installed with RepeatModeler v.2.01. A combination of the consensus sequences of RepeatModeler output and the locust

transposable elements deposited in Repbase were used as input for RepeatMasker v.4.11 to determine the genomic coordinates of annotated transposable elements and generate a masked version of the locust genome assembly. SINE retrotransposons were classified into non-LTR retrotransposons based on the classification system of Repbase. The Kimura distances between genome copies and TE consensus from the RepeatModeler library were determined using RepeatLandscape after genome masking with RepeatMasker. Tandem repeats were identified using Tandem Repeats Finder v.4.09 and were processed by pyTanFinder (<https://github.com/Kirovez/pyTanFinder>).

Cleavage under targets and tagmentation

The selection of histone modifications is introduced in Supplementary Note 8. The commercial antibodies include H3K9ac (Abcam, catalogue number ab32129), H3K14ac (Abcam, catalogue number ab52946), H3K18ac (Abcam, catalogue number ab177870), H3K27ac (Abcam, catalogue number ab4729), H3K27me1 (Active Motif, catalogue number 61015), H3K27me2 (Abcam, catalogue number ab24684), H3K27me3 (Cell Signaling Technology, catalogue number 9733S), H3K36me3 (Active Motif, catalogue number 61022), H3K4me1 (Abcam, catalogue number ab8895), H3K4me2 (Abcam, catalogue number ab32356), H3K4me3 (Abcam, catalogue number ab8580), H3K9me3 (Abcam, catalogue number ab8898), H4K20me3 (Abcam, catalogue number ab9053), H4K16ac (Abcam, catalogue number ab109463) and H4K20me1 (Abcam, catalogue number ab9051). The locust tissues were isolated and immobilized to concanavalin A-coated beads (Novoprotein). The beads-bound cells were incubated with primary antibodies by rotating overnight. The resulting cells were incubated with secondary antibodies at room temperature for 1 h and were then washed three times with DIG-wash buffer to remove unbound antibodies. Cells were resuspended in DIG-300 buffer (300 mM NaCl) with pA/G-Tn5 adapter complex for 1 h at room temperature by rotating, followed by tagmentation at 37 °C for 1 h. Subsequent washes were also performed under stringent salt conditions (300 mM NaCl) to suppress the affinity of Tn5 for open chromatin⁴⁹. Barcoded I5 and I7 primers were used for PCR amplification to generate the sequencing libraries. The final sequencing libraries were sequenced on an Illumina NovaSeq 6000 system. Paired-end reads were aligned to LMv.3.1 using Bowtie2 v.2.4.1. The multiBigwigSummary and plotCorrelation functions from deepTools v.2.27.1 were used to assess replicate reproducibility. The computeMatrix and plotHeatmap functions from deepTools v.2.27.1 were used to generate the heatmap. The CUT&Tag peaks were called by MACS2 v.2.2.7.1 and SEACR v.1.3 for narrow and broad peaks, respectively^{50,51}. The fragment size plots of the CUT&Tag experiments showed periodical peaks corresponding to the mononucleosomes, dinucleosomes and trinucleosomes (Supplementary Fig. 32). The signal intensities of stranded RNA-seq correlated positively with H3K36me3 (Fig. 2a), which is consistent with their demarcation role in gene-body deposition. The negative correlation of the repressed histone modification marks (H3K27me3, H3K27me2 and H4K20me3) with gene-body transcription suggests these repressed histone modification marks are more likely to silence expression in the nongenic region. The pairwise-correlated permissive histone modification marks (H3K18ac, H3K14ac, H3K27ac and H3K9ac) exhibited a positive correlation with promoter-associated transcription (H3K4me3, H3K4me2 and TSS-seq). For the histone modification datasets, we obtained a total of 683,539 peaks with an average number of 52,579 (ranging from 13,555 for H3K4me2 to 99,091 for H3K27me2; Supplementary Fig. 33).

Transcription start sites

Total RNAs were isolated using TRIzol reagent (Invitrogen). Genomic DNA was removed using TURBO DNase (Invitrogen) and poly-A RNA was enriched twice using a Dynabeads Oligo (dT) 25 kit (Thermo Fisher Scientific). Poly-A RNA was treated with calf intestinal alkaline phosphatase at 37 °C for 1 h and then treated with tobacco acid

pyrophosphatase at 37 °C for 1 h. The sequencing libraries were constructed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) following the manufacturer's instructions and sequenced on an Illumina NovaSeq 6000 system. The trimmed reads were aligned to LMv.3.1 using HISAT2 v.2.2.1 without allowing soft clipping alignments. The resulting BAM files were used in the identification of oligo-capping transcription start sites using the CAGER v.1.24.0 package in R v.3.5.1⁵².

ATAC sequencing

Locust tissues were digested using trypsin/EDTA and resuspended in lysis buffer to permeabilize membranes. The nuclear pellets were resuspended in the tagmentation mix (Vazyme) and incubated at 37 °C for 30 min. DNA Clean Beads (Vazyme) was then added to terminate the tagmentation reaction and to purify DNA fragments. Barcoded I5 and I7 primers were used for PCR amplification to generate sequencing libraries. The resulting sequencing libraries were purified with DNA Clean Beads (Vazyme) and subjected to sequencing on an Illumina NovaSeq 6000 system. ATAC-seq data analyses were based on the ENCODE pipeline. Trimmed reads were aligned to LMv.3.1 using Bowtie2 v.2.4.1. The aligned reads +4 base pairs (bp) for positive strands and -5 bp for negative strands were shifted to adjust the 9-bp duplication created by Tn5 transposase. The ATAC-seq peaks were called on nucleosome-free reads by MACS2 v.2.2.7.1 using the narrow parameter⁵⁰. The consensus peaks were retained for further analyses. ATAC-seq experiments generated an enrichment of fragment size around 100 and 200 bp, indicating nucleosome-free and mononucleosome-bound fragments (Supplementary Fig. 34). ATAC-seq signals were enriched at TSSs that were captured by Oligo-capping sequencing.

Annotation of chromatin states

We used ChromHMM v.1.23 to analyze the combinatorial patterns of 13 histone modifications⁵³. The chromatin state annotation definitions are adopted from those used for the human genome⁵⁴. Briefly, the promoter state is associated primarily with H3K4me3 and H3K27ac. The flanking promoter state shows high levels of H3K4me3 but lacks H3K27ac, reflecting the broader distribution of H3K4me3 compared to H3K27ac. The transcribed at gene 5' end state is associated with H3K4me3 and H3K36me3. The two transcription states are associated specifically with H3K36me3 and H3K27me1, respectively. The enhancer state is linked with H3K27ac and H3K4me1, and the bivalent enhancer state is characterized by the presence of enhancer-associated marks (H3K18ac, H3K4ac and H3K27ac) together with repressive marks H3K27me3. The repressed polycomb state is defined by its exclusive association with H3K27me3, since H3K27me3 is catalyzed primarily by polycomb repressive complexes. The difference between the two H3K9me3-associated heterochromatin states is the presence of H4K20me3.

Fluorescent in situ hybridization

FISH was carried out using the FISH Kit (BersinBio) with small modifications. Chromosome preparation was incubated in the digestive working solution. After three washes in distilled water, the slides were dehydrated in an ethanol series then air-dried. DNA probes and slides were denatured and hybridized overnight. After hybridization, slides were washed under low stringency conditions (2× SSC, 0.1% NP-40/2× SSC and 2× SSC).

Immunofluorescence

Testes were fixed in 4% paraformaldehyde (PFA) at 4 °C overnight, and seminiferous tubules were then isolated. Brains were collected in PBS and then digested in collagenase/dispase solution (2 mg ml⁻¹, Roche) for 10–15 min at 28 °C. Incubation with primary antibodies against H4K16ac (Abcam, catalogue number ab109463) was performed at 4 °C in a moist chamber overnight. The secondary antibody, Alexa Fluor 546 goat anti-rabbit IgG (Thermo Fisher, catalogue number A-11035), was

diluted in immunofluorescence staining secondary antibody dilution buffer. Hoechst was used as nuclear marker.

Genomic fragment knockout using the CRISPR–Cas9 system

We masked the repetitive region to remove segments with the potential to generate off-target (guide RNAs (gRNAs)). Subsequently, we designed gRNAs based on nonrepetitive regions flanking EHI. The cutting efficiency of the gRNAs was determined using Illumina amplicon sequencing. gRNAs were transcribed using a template assembled with the GeneArt Precision gRNA Synthesis Kit (Thermo Fisher). Homology donors were produced by incorporating two homology arms located in the flanking region of EHI. The mixture of gRNAs, homology donors and TrueCut Cas9 Protein v.2 (Thermo Fisher) was microinjected into the eggs laid within 2 h. Microinjected eggs were placed in disposable sterile Petri dishes. Knockout of the 379-bp genomic fragment encompassing EHI was verified by Sanger sequencing.

Behavioral assays

Gregarious locusts were reared in large cages (40 cm × 40 cm × 40 cm) at a density of approximately 600 locusts per cage. Solitary locusts were cultured individually in white metal boxes (10 cm × 10 cm × 25 cm) supplied with charcoal-filtered compressed air. Development into the solitary or gregarious phase depends on rearing conditions. The behavioral assay for the third day of fourth-instar locust nymphs was conducted in a rectangular Perspex arena with opaque walls and a clear top. A locust was released into the arena and monitored for 5 min. Individual behavioral data were recorded automatically and analyzed using the EthoVision video tracking system. The three behavioral variables were as follows: total distance moved (TDM) and total duration of movement (TDMV), representing motor activity levels, and attraction index (AI), defined as the total duration in the stimulus area minus the total duration in the opposite area, representing attraction or repulsion to the stimulus group. The behavioral phase status of each locust was assessed by applying a single probabilistic metric of gregariousness, $P_{\text{greg}} = \frac{P_{\text{greg}}}{1 + P_{\text{greg}}}$, where $\eta = -2.11 + 0.005 \times \text{AI} + 0.012 \times \text{TDM} + 0.015 \times \text{TDMV}$. e represents Euler's number, approximately 2.718. A P_{greg} value of 1 indicates fully gregarious behavior, whereas a value of 0 indicates fully solitary behavior.

Quantitative chromosome conformation capture assays

Cells from locust brains were collected to prepare single-cell suspensions in 1 ml PBS. The resulting cells were incubated with 2% formaldehyde for cross-linking and were quenched with a final concentration of 0.125 mM glycine. The cells were not treated with cross-linking as a digested DNA control. Cells were lysed with 500 μl 1× cold lysis buffer (10 mM Tris–HCl pH 8.0, 10 mM NaCl, 0.2% NP-40). Each sample was digested overnight with *DpnII* enzyme. Ligation reaction was performed at 16 °C for –22 h in 1.5-ml tubes containing 10× T4 ligase buffer, 125 μl water, 4 μl 40 U per microliter of T4 ligase. The crosslinks were reversed with Proteinase K (Invitrogen) at 65 °C overnight. DNA were then purified using phenol–chloroform extraction and quantified by Qubit dsDNA HS Assay (Invitrogen). The primer sequences are provided in Supplementary Table 2.

Quantitative reverse transcription PCR

Total RNA was extracted using TRIzol reagent (Invitrogen). Quantitative reverse transcription PCR was performed using FastStart DNA Master SYBR Green I (Roche) with a LightCycler 480 instrument (Roche). Relative expression levels were calculated using the $2^{-\Delta\Delta C_t}$ method.

Statistics

For statistical analysis, a two-sided Student's *t*-test was used to evaluate differences in (1) mRNA and protein expression of *Henna*, (2) spatial interaction between EHI and TSS1 and (3) read abundances of H4K16ac between autosomes and the X chromosome. A two-sided

Wilcoxon rank-sum exact test was used to evaluate differences in (1) histone modification deposition on gene-body regions in short-size and large-size genes, (2) female-to-male ratios of genic H4K16ac and H4K20me1 signals (with Benjamini–Hochberg adjusted *P* values) and (3) arena behavioral assays. A two-sided Pearson's product-moment correlation test was used to assess the statistical significance of correlations in (1) intron length ratios between the locust and the two other insects, (2) RNA expression ratios between the locust and the fruit fly and (3) gene expression between females and males. A one-sided Student's *t*-test was used to evaluate differences in dN/dS values for translocated genes, with Benjamini–Hochberg adjusted *P* values. The significance between chromosome–chromosome associations in Caeliferan species was assessed with the two-sided Fisher's exact test against a null model of random ortholog permutation with a Bonferroni correction.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequencing data and genome assemblies have been deposited in the Genome Sequence Archive at the National Genomics Data Center of the China National Center for Bioinformation under the BioProject accession number [PRJCA043246](https://doi.org/10.6084/m9.figshare.27889818). The genome annotations are available via figshare at <https://doi.org/10.6084/m9.figshare.27889818> (ref. 55). The transcriptome data of the brain, testis and fat body of the fruit fly were retrieved from the NCBI database under accession number [PRJNA75285](https://doi.org/10.6084/m9.figshare.27889818). The H3K27ac and H3K4me1 profiling data from fruit fly and honey bee brains were retrieved from the NCBI GEO database under accession numbers [GSE210427](https://doi.org/10.6084/m9.figshare.27889818) and [GSE206992](https://doi.org/10.6084/m9.figshare.27889818), respectively. The H4K16ac profiling data from female and male fruit flies were retrieved from the NCBI database under GEO accession number [GSE133637](https://doi.org/10.6084/m9.figshare.27889818). Source data are provided with this paper.

Code availability

The processing commands for sequencing data analysis and Figs. 1–6 are available via figshare at <https://doi.org/10.6084/m9.figshare.27889818> (ref. 55).

References

- Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
- Chin, C. S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
- Kronenberg, Z. N. et al. Extended haplotype-phasing of long-read de novo genome assemblies using Hi-C. *Nat. Commun.* **12**, 1935 (2021).
- Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
- English, A. C. et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE* **7**, e47768 (2012).
- Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
- Hall, A. B. et al. A male-determining factor in the mosquito *Aedes aegypti*. *Science* **348**, 1268–1270 (2015).

46. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
47. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
48. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
49. Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat. Protoc.* **15**, 3264–3283 (2020).
50. Zhang, Y. et al. Model-based analysis of ChIP–Seq (MACS). *Genome Biol.* **9**, R137 (2008).
51. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by sparse enrichment analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).
52. Haberle, V., Forrest, A. R., Hayashizaki, Y., Carninci, P. & Lenhard, B. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res.* **43**, e51 (2015).
53. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
54. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
55. Jiang, F. Chromosome-level genome assemblies of the migratory locust and the desert locust. *figshare* <https://doi.org/10.6084/m9.figshare.27889818> (2025).

Acknowledgements

This work was supported by the National Key R&D Program of China (2022YFD1400500 to F.J.), the International Partnership Program of Chinese Academy of Sciences (151542KYSB20200016 to F.J. and W.G.), the National Natural Science Foundation of China (32088102 to L.K., 32300397 to Q. Liu, and 32270523 to J.H.), the Young Elite Scientists

Sponsorship Program by CAST (2024QNRC001 to Q. Liu) and the Initiative Scientific Research Program of Institute of Zoology, Chinese Academy of Sciences (2024IOZ0202 to F.J.). This work was also supported by the Agriculture Science and Technology Major Project (to F.J.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Q. Liu, F.J. and L.K. designed and supervised the study. Q. Liu, F.J. and S.L. performed data analysis. R.L., Q. Liu and W.F. performed experiments. X.L., T.H., L.F., J.Z., Z.L., J.H., W.G., Xianhui Wang, Xiaoxiao Wang and L.W. contributed to the experiments. Xianhui Wang, Z.S., J.L., Y.G., J.Y. and Q. Li contributed to the genome assembly and annotation. F.J., Q. Liu and L.K. wrote the paper with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-025-02330-y>.

Correspondence and requests for materials should be addressed to Le Kang.

Peer review information *Nature Genetics* thanks Steven Van Belleghem and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Chromosome staining was visualized using on a Zeiss LSM 710 confocal microscopy. The genome sequencing was conducted on a PacBio Sequel instrument and a Nanopore PromethION platform and an Illumina HiSeq X Ten System. DNA labeling for optical mapping was imaged and digitized in a Saphyr instrument (Bionano Genomics). All the libraries of CUT&Tag, ATAC-seq and oligo-capping experiments were sequenced on an Illumina NovaSeq 6000 System. Quantitative reverse transcription (qRT-PCR) was performed with a LightCycler 480 instrument (Roche). DNA FISH images were obtained using a Zeiss LSM 710 confocal microscope. Immunofluorescence images were obtained using an Olympus IXplore SpinSR microscope or a Nikon N-SIM S super resolution microscope.

Data analysis GenomeScope v1.0.0 was used to conduct a K-mer frequency survey of genome composition based on 21-mer frequency distribution to estimate genome size and heterozygosity. For the migratory locust, The raw PacBio reads were corrected and trimmed using Canu v1.7, and the contig-level assembly were performed using FALCON v2.0.4 followed by the FALCON-Unzip v0.4.0 module. The remaining alternative haplotypes from the initial primary contigs were removed using Purge Haplotigs v1.0.0, and the haplotypes switch errors were eliminated using the FALCON-Phase v0.1.0. All PacBio reads were aligned to the resulting assemblies using BLASR v5.3.3 and then the consensus sequences were called using the Arrow program in SMRT link v7.0. The PacBio-corrected assemblies were further improved by two rounds of polishing with Pilon v1.23 using paired-end Illumina data. Data visualization, De novo assembly, and hybrid scaffolding of the optical map data with a minimum size of 150 Kb were performed using the Bionano Access v1.2.1 and Bionano Solve v3.2.1 software. Gaps in the hybrid-scaffolded assembly were closed using PBJelly from PBSuite v15.8.24 using PacBio and Nanopore read datasets. For the desert locust, the de novo genome assemblies were generated by NextDenovo v2.4.0 and NextPolish v1.3.1 using long-read Nanopore sequencing. Trimmed Hi-C reads were aligned to the polished genome assembly using bwa v0.7.17 and filtered for Hi-C-specific artifacts using the HiC-Pro v2.11.1 pipeline. The Juicer v1.5.6 and 3D-DNA v180922 pipelines were used to grouping, ordering and orientation of scaffolds and to generate a chromosome-level genome assembly. The potential misjoins, translocations, and inversions within the boundaries of each chromosome were corrected by manually introducing breakpoints at the mis-assembly sites based on visualization with Juicebox v1.5.3. The sex chromosome

was identified by using a chromosome quotient method which relies on the simple assumption that female locusts have a two-fold increase of read coverage compared to male locusts. The DNA-seq reads from an individual of both the male and female locusts were aligned separately to the genome assembly using the Bowtie2 v2.4.1 program. The resulting bam files were filtered to remove low-quality alignments using the SAMtools v1.9 program. Transposable elements and simple repeats were predicted de novo and classified using the RepeatModeler v2.01 package with RMBlast v2.9.0-p2 engine 49. Other prerequisite modules including RECON v1.08, RepeatScout v1.0.6, TRF v4.0.9, Dfam v2.0 and Repbase-derived libraries were installed with the RepeatModeler v2.01 package. A combination of the consensus sequences of RepeatModeler output and the locust transposable elements deposited in Repbase were used as input for RepeatMasker v4.11 to determine the genomic coordinates of annotated transposable elements and generate a masked version of the locust genome assembly. Paired-end reads were aligned to LMv3.1 using Bowtie2 v2.4.1. The multiBigwigSummary and plotCorrelation functions from deepTools v2.27.1 were used to assess replicate reproducibility. The computeMatrix and plotHeatmap functions from deepTools v2.27.1 were used to generate the heatmap. The CUT&Tag peaks were called by MACS2 v2.2.7.1 and SEACR v1.3 for narrow and broad peaks, respectively. The trimmed transcriptomic reads were aligned to LMv3.1 using HISAT2 v2.2.1 without allowing soft clipping alignments. The resulting BAM files were used in the identification of oligo-capping transcription start sites using the CAGER version 1.24.0 package in R v3.5.1. The ATAC-seq data analyses were performed based on the ENCODE pipeline. The trimmed reads were aligned to LMv3.1 using Bowtie2 v2.4.1. The ATAC-seq peaks were called on nucleosome-free reads by MACS2 v2.2.7.1 using the narrow peak mode. ChromHMM v1.23 was used to analyze the combinatorial patterns of 13 histone modifications. The ortholog pairs between the migratory locust and the fruit fly were determined using OrthoFinder v2.5.5 inference. The read abundances plot and the letter value plots were generated using the karyoploteR v1.22.0 and Ivplot v0.2.1 packages, respectively.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The sequencing data and genome assemblies were deposited in the Genome Sequence Archive (GSA) at the National Genomics Data Center of the China National Center for Bioinformation, under accession number PRJCA043246. The genome annotations were deposited in Figshare under the DOI accession link: <https://dx.doi.org/10.6084/m9.figshare.27889818>. The transcriptome data of the brain, testis, and fat body of the fruit fly were retrieved from the NCBI database under accession number PRJNA75285. The H3K27ac and H3K4me1 profiling data from fruit fly and honey bee brains were retrieved from the NCBI GEO database under accession numbers GSE210427 and GSE206992, respectively. The H4K16ac profiling data from female and male fruit flies were retrieved from the NCBI database under GEO accession number GSE133637. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	<input type="text" value="N/A"/>
Reporting on race, ethnicity, or other socially relevant groupings	<input type="text" value="N/A"/>
Population characteristics	<input type="text" value="N/A"/>
Recruitment	<input type="text" value="N/A"/>
Ethics oversight	<input type="text" value="N/A"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample size	filed. The sample size, which followed previously published literature on locust experiments (e.g., Guo, Nature, 2020), was chosen empirically to ensure sufficient statistical power.
Data exclusions	No data was excluded
Replication	The CUT&Tag and ATAC experiments were performed with two independent biological replicates, in line with accepted practices in the epigenomics field. The RNA-seq, 3C, mRNA qPCR, and Western blot experiments were performed with at least three independent biological replicates. The behavioral assays were performed with at least seven independent biological replicates. The FISH and immunofluorescence experiments were performed with two independent biological replicates. All attempts to repeat the experiments were successful.
Randomization	The samples were allocated randomly into experimental groups.
Blinding	The investigators were blinded to group allocation during data collection and data analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	<p>H3K9ac, Abcam, ab32129, 1:100 H3K14ac, Abcam, ab52946, 1:100 H3K18ac, Abcam, ab177870, 1:100 H3K27ac, Abcam, ab4729, 1:100 H3K27me1, Active Motif, 61015, 1:100 H3K27me2, Abcam, ab24684, 1:100 H3K27me3, Cell Signaling Technology, 9733S, 1:100 H3K36me3, Active Motif, 61022, 1:100 H3K4me1, Abcam, ab8895, 1:100 H3K4me2, Abcam, ab32356, 1:100 H3K4me3, Abcam, ab8580, 1:100 H3K9me3, Abcam, ab8898, 1:100 H4K20me3, Abcam, ab9053, 1:100 H4K16ac, Abcam, ab109463, 1:100 H4K20me1, Abcam, ab9051, 1:100 Alexa Fluor 546 goat anti-rabbit IgG, Thermo Fisher, A-11035, 1:1000 Henna, 1:3000, The henna antibody is retrieved from the previous study (PMID: 24586212).</p>
Validation	<p>Histone modification is highly conserved in animals. The validation information for the histone modification antibodies is provided on the website of manufacturers. The henna antibody was validated in the previous study (PMID: 24586212). H3K9ac, https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-acetyl-k9-antibody-y28-chip-grade-ab32129 H3K14ac, https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-acetyl-k14-antibody-ep964y-chip-grade-ab52946 H3K18ac, https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-acetyl-k18-antibody-epr16595-chip-grade-ab177870 H3K27ac, https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-acetyl-k27-antibody-chip-grade-ab4729 H3K27me1, https://www.activemotif.com/catalog/details/61015 H3K27me2, https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-di-methyl-k27-antibody-chip-grade-ab24684, H3K27me3, https://www.cellsignal.com/products/primary-antibodies/tri-methyl-histone-h3-lys27-c36b11-rabbit-mab/9733 H3K36me3, https://www.activemotif.com/catalog/details/61021/histone-h3-trimethyl-lys36-antibody-clone-mab-clone-mabi-0333 H3K4me1, https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-mono-methyl-k4-antibody-chip-grade-ab8895 H3K4me2, https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-di-methyl-k4-antibody-y47-chip-grade-ab32356 H3K4me3, https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-tri-methyl-k4-antibody-chip-grade-ab8580 H3K9me3, https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-tri-methyl-k9-antibody-chip-grade-ab8898 H4K20me3, https://www.abcam.com/en-us/products/primary-antibodies/histone-h4-tri-methyl-k20-antibody-chip-grade-ab9053 H4K16ac, https://www.abcam.com/en-us/products/primary-antibodies/histone-h4-acetyl-k16-antibody-epr1004-ab109463 H4K20me1, https://www.abcam.com/en-us/products/primary-antibodies/histone-h4-mono-methyl-k20-antibody-chip-grade-ab9051</p>

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	The locust colonies, maintained at the Institute of Zoology, Chinese Academy of Sciences, were reared under a 14:10 light/dark photoperiod at 30°C and fed fresh wheat seedlings and wheat bran. Fourth-instar nymphs with a 1:1 sex ratio were used for behavioral plasticity experiments. Genome sequencing was conducted on an adult female migratory locust and an adult male desert locust. For dosage compensation experiments, both adult female and male locusts were used.
Wild animals	No wild animals were used in the study.
Reporting on sex	N/A
Field-collected samples	No field collected samples were used in the study.
Ethics oversight	This study does not require ethical approval.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A