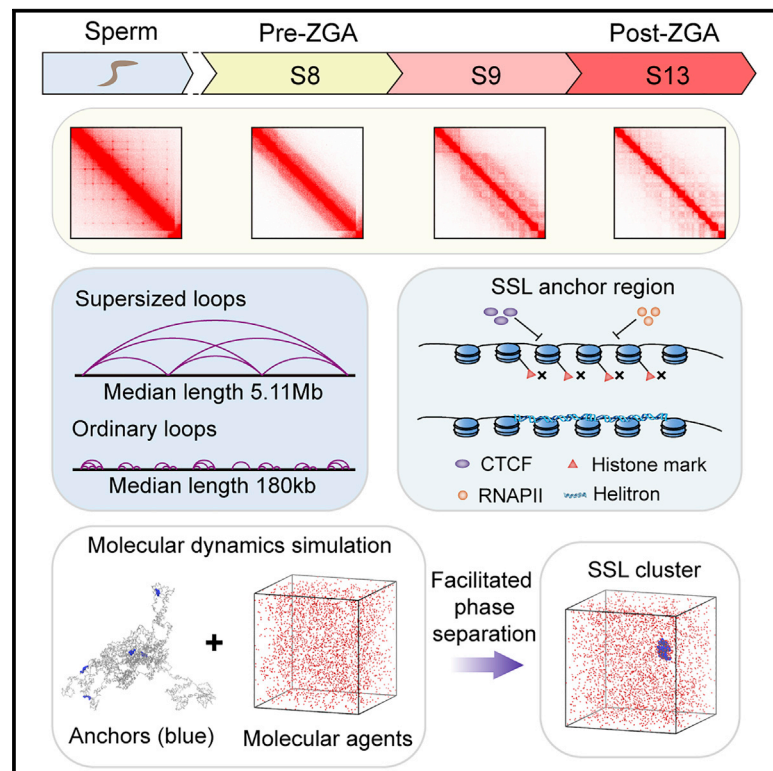**Report**

Evolutionarily distinct and sperm-specific supersized chromatin loops are marked by Helitron transposons in *Xenopus tropicalis*

Graphical abstract

Authors

Zhaoying Shi, Jinsheng Xu, Longjian Niu, ..., Yonglong Chen, Li Li, Chunhui Hou

Correspondence

huangkai@szbl.ac.cn (K.H.),
chenyl@sustech.edu.cn (Y.C.),
li.li@hzau.edu.cn (L.L.),
houchunhui@mail.kiz.ac.cn (C.H.)

In brief

Through systematic comparison and comprehensive analysis of chromosome conformation, Shi et al. show an evolutionarily distinct and sperm-specific genome structure of supersized loop clusters in *Xenopus tropicalis* that are demarcated by transposon Helitrons, suggesting the existence of a link between the control of Helitrons and genome architecture during spermatogenesis.

Highlights

- Multimegabase supersized loops (SSLs) form clusters in mature *X. tropicalis* sperm

- SSL anchors are inaccessible and devoid of CTCF, RNAPII, and modified histones

- SSL anchors are explicitly enriched with class II DNA transposon of Helitrons

- Molecular dynamics simulation indicates SSL clusters may form via phase separation

CellPress

Cell Reports

Report

Evolutionarily distinct and sperm-specific supersized chromatin loops are marked by Helitron transposons in *Xenopus tropicalis*

Zhaoying Shi,[1,10] Jinsheng Xu,[2,7,10] Longjian Niu,[3,5,6,10] Wei Shen,[2,7,10] Shuting Yan,[4,10] Yongjun Tan,[1,3] Xuebo Quan,[4] Edwin Cheung,[8,9] Kai Huang,[4,*] Yonglong Chen,[1,*] Li Li,[2,7,*] and Chunhui Hou[3,11,*]

[1]Department of Biology, Southern University of Science and Technology, Shenzhen 518055, China
[2]Department of Bioinformatics, Huazhong Agricultural University, Wuhan 430070, China
[3]China State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[4]Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518132, China
[5]School of Public Health and Emergency Management, Southern University of Science and Technology, Shenzhen 518055, China
[6]Shenzhen Key Laboratory of Cardiovascular Health and Precision Medicine, Southern University of Science and Technology, Shenzhen 518055, China
[7]Hubei Key Laboratory of Agricultural Bioinformatics, Huazhong Agricultural University, Wuhan 430070, China
[8]Cancer Centre, Faculty of Health Sciences, University of Macau, Taipa, Macau 999078, China
[9]Frontier Science Centre for Precision Oncology of Ministry of Education, University of Macau, Taipa, Macau 999078, China
[10]These authors contributed equally
[11]Lead contact
*Correspondence: huangkai@szbl.ac.cn (K.H.), chenyl@sustech.edu.cn (Y.C.), li.li@hzau.edu.cn (L.L.), houchunhui@mail.kiz.ac.cn (C.H.)
https://doi.org/10.1016/j.celrep.2023.112151

SUMMARY

Transposable elements (TEs) are abundant in metazoan genomes and have multifaceted effects on host fitness. However, the mechanisms underlying the functions of TEs are still not fully understood. Here, we combine Hi-C, ATAC-seq, and ChIP-seq assays to report the existence of multimegabase supersized loop (SSL) clusters in the *Xenopus tropicalis* sperm. We show that SSL anchors are inaccessible and devoid of the architectural protein CTCF, RNA polymerase II, and modified histones. Nearly all SSL anchors are marked by Helitrons, a class II DNA transposon. Molecular dynamics simulations indicate that SSL clusters are likely formed via a molecular agent-mediated chromatin condensation process. However, only slightly more SSL anchor-associated genes are expressed at late embryo development stages, suggesting that SSL anchors might only function in sperm. Our work shows an evolutionarily distinct and sperm-specific genome structure marked by a subset of Helitrons, whose establishment and function remain to be explored.

INTRODUCTION

High-order genome architectures form three-dimensional (3D) scaffolds that facilitate long-range control of many biological activities.[1–4] Generally, metazoan genomes fold into three types of distinct structures, including A/B compartments,[5] topologically associating domains (TADs),[6–9] and structural and functional chromatin loops.[10–16] In recent years, the establishment of these structures has been found to be associated with transposable elements (TEs) in animals and plants. For example, class I retrotransposons L1 and B1/Alu are associated with A/B compartment segregation in mammalian genomes,[17] while the human endogenous retrovirus subfamily H (HERV-H) retrotransposons demarcate TAD borders when actively transcribed in human pluripotent stem cells (hPSCs).[18] In addition, transposons also help form Knot structures, in which specific genomic regions of all five *Arabidopsis* chromosomes interact together in the somatic nucleus.[19]

Structural changes associated with TEs also have consequences at the gene expression level. For instance, the mobilization of TEs can trigger genome instability, leading to the deregulated expression of essential genes and deleterious effects on host fitness.[20] In contrast, TEs can also provide a reservoir of DNA motifs that can be utilized for the proper control of gene regulation[21,22] and the improvement of adaptability.[23–25] Accordingly, TE repression and utilization for regulatory purposes have been two intertwined themes throughout evolution. Although several class I TEs have been reported to be associated with 3D genome structure in mammals, whether TEs are involved in forming the 3D genome structure in gametes has not been reported.

In this study, we mapped the 3D genome architecture of the *Xenopus tropicalis* (*X. tropicalis*) sperm genome and unveiled supersized loops (SSLs) that are enriched with class II DNA transposon Helitrons.[26] Through molecular dynamics simulations, we proposed two different mechanisms underlying the formation of

SSLs. However, the biological function and importance of this structure remain to be explored.

RESULTS

Genome folding patterns are evolutionarily distinct in vertebrate sperm

During spermiogenesis, vertebrate genomes transform into a more condensed state.[27] To understand whether this process is evolutionarily conserved in vertebrates, we carried out a comparative analysis of the 3D genome structures in the mature sperm of several species, including humans,[28] monkeys,[29] mice,[29–31] and zebrafish.[32] Overall, our results show that the 3D genome structures of sperm are dramatically different among vertebrate animals (Figures S1A–S1J and S2A–S2D). Specifically, the human sperm genome lacks TADs (Figure S1A), whereas only a hinge-like structure indicative of TAD border location exists in zebrafish sperm (Figure S1G). Furthermore, compartments in zebrafish sperm are barely segregated (Figure S1H), while the genome of human sperm is well compartmentalized (Figure S1B). In contrast, both TADs and compartments are prominent 3D genome features of monkey and mouse sperm (Figures S1C–S1F and S2A–S2D). These structural differences strongly suggest a lack of conserved principles governing genome folding in vertebrate sperm.

The above observations prompted us to examine the genome architecture of sperm in another animal model, *X. tropicalis*, which occupies a unique position in the phylogenetic tree. Hi-C analysis revealed a striking absence of TADs and compartments in the *X. tropicalis* sperm genome (Figures S1I, S1J, and S2E). In contrast, human K562 cells, used as a Hi-C spike-in control, have TAD and compartment structures that are clearly present (Figures S2F and S2G). Together, observations from our work and others show that the sperm genomes of various species likely adopt different structures associated with species-specific factors.

Supersized loops form in the *X. tropicalis* sperm genome

When visually inspecting the contact heatmaps of the *X. tropicalis* sperm genome, we unexpectedly discovered dots indicative of clustered loops (Figures 1A and S3A–S3C; Table S1). We identified 144 of these loops covering 25 multimegabase genomic regions in the ten *X. tropicalis* chromosomes. The interaction frequency between sperm loop anchors remained the highest as the resolution increased from 25 kb to 5 kb, similar to the ordinary chromatin loops in brain cells (Figure 1B). Notably, the loops within these clusters span huge genomic distances with a median size of 5.11 Mb, approximately 28 times longer than regular chromatin loops in brain cells (Figure 1C; p < 0.001). We named these genomic structures SSLs to distinguish them from the regular-sized chromatin loops in somatic cells.

Interestingly, most SSLs appear as clusters (138/144 SSLs in 19 clusters), except for 6 SSLs formed between single pairs of anchors (Table S1). We also examined the genome structures of sperm from other animals. However, none of them have structures that are similar to SSLs. We conducted Hi-C experiments on the mature sperm of *X. laevis* and did not observe SSL-like structures in the syntenic genomic regions that contain SSLs in the sperm of

X. tropicalis (Figure S4). These results suggest that SSLs, to the best of our knowledge, are established only in *X. tropicalis* sperm.

Next, we asked whether SSL structures are maintained after fertilization and in terminally differentiated somatic cells of *X. tropicalis* by examining chromatin contact maps of SSL-spanned genomic regions in the brain, liver, and embryos at different developmental stages. No structures similar to sperm SSLs were detected in these tissues or cells (Figures 1D–1I). Because the embryos we used for Hi-Cs are diploid, we determined that it may be necessary to exclude the maternally inherited genome from our analysis. Thus, we conducted Hi-C on hybrid embryos derived from male *X. tropicalis* and female *X. laevis* parents and examined the genome architecture of the paternal genome only. Again, we found no SSL structures for the *X. tropicalis* genome in embryos of late development stage 33 or the tail of tadpoles (Figures S3D and S3E). In sum, we did not identify any SSL or SSL-like structures in any of the embryos and somatic cells that we examined. These results indicate that SSLs are likely to be sperm specific in *X. tropicalis*. Thus, we uncover a 3D structure of SSLs that has yet to be reported.

SSL anchors are located in closed chromatin

Regular chromatin loops form between anchors accessible to CCCTC-binding factor (CTCF) and cohesins that mediate chromatin interactions. To determine whether SSL anchors are open for protein binding in the sperm genome, we conducted assay for transposase-accessible chromatin using sequencing (ATAC-seq) and chromatin immunoprecipitation sequencing (ChIP-seq) analyses (Figures 2A and S5A–S5E). Surprisingly, we found that SSL anchors are less accessible than flanking DNA sequences (Figure 2B), suggesting that these regions are in a relatively closed chromatin state. In contrast, many gene promoters are accessible to Tn5 digestion, even though transcription is silent in sperm (Figures 2A and S5A–S5E). Consistent with these observations, we found RNA polymerase II (RNAPII) enriched at promoters (Figures 2A and S5A–S5E) but not at SSL anchors (Figure 2C), confirming that SSL anchors are located in inactive genomic regions. Together, these results suggest that establishing SSLs may be associated with repressive effects.

We hypothesized that architectural proteins such as CTCF might mediate the maintenance of SSLs in mature sperms. However, contrary to our speculation, ChIP-seq analysis showed that CTCF is not enriched at SSL anchors (Figure 2D), even though CTCF binding occurs throughout the sperm genome (Figures 2A and S5A–S5D). Furthermore, a motif search showed no enrichment of CTCF binding sites at SSL anchor regions. Together, these results suggest that CTCF is unlikely to be the factor responsible for maintaining or establishing SSLs.

Repressive chromatin modifications, including H3K27 methylation and H3K9 methylation, are associated with the binding of Polycomb-group (PcG) proteins and heterochromatin protein 1 (HP1), respectively. These marks are reportedly involved in long-range chromatin interactions.[33,34] Therefore, to determine if these two chromatin types are involved in maintaining or establishing SSLs, we performed H3K27me3 and H3K9me2 ChIP-seq analysis (Figure 2E). Our results showed that neither H3K27me3 nor H3K9me2 was enriched at SSL anchors (Figures 2F and 2G). We examined several additional histone modifications and
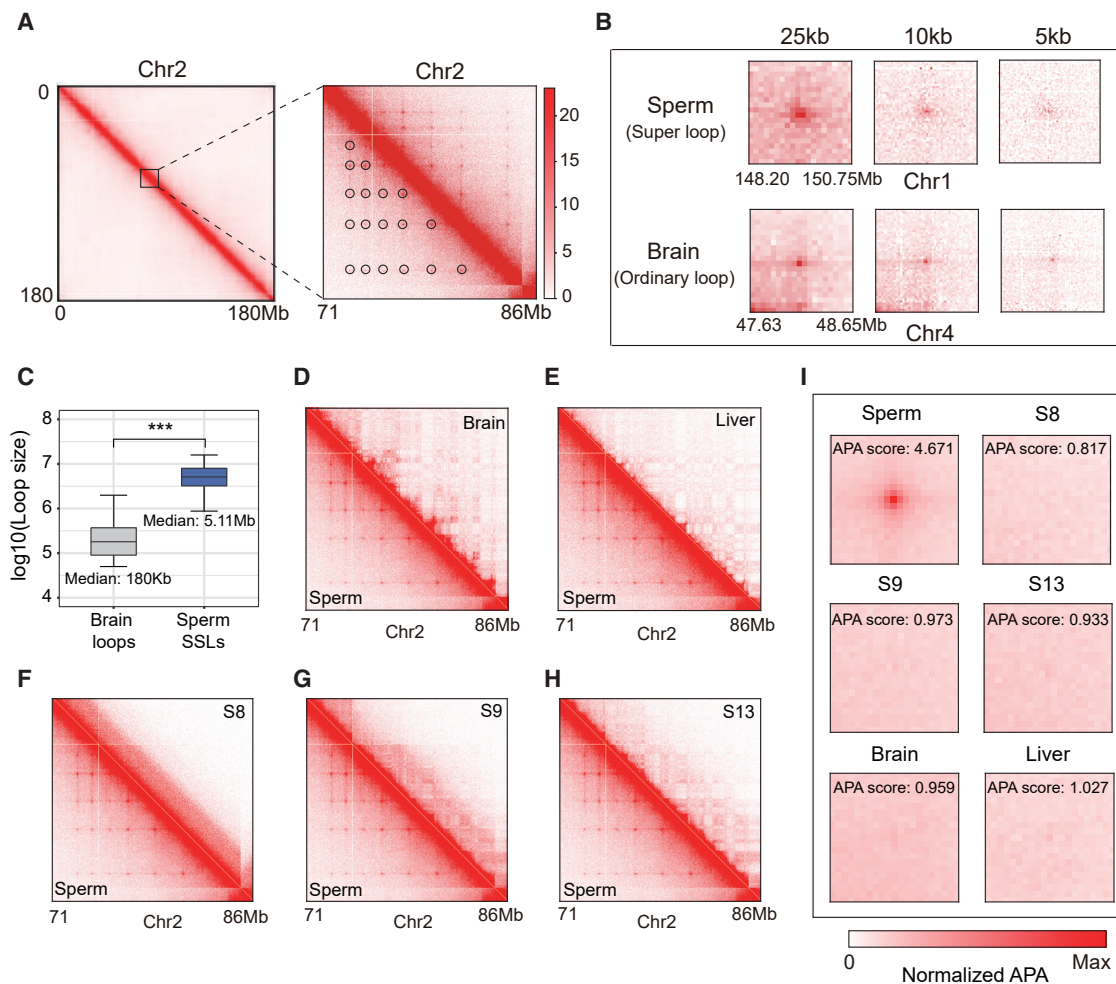
Figure 1. Supersized loops form in the *X. tropicalis* sperm genome

(A) The Hi-C contact matrix heatmap of chromosome 2 in *X. tropicalis* sperm (left). The inlet is marked as a black box, and a zoomed-in view is shown on the right (Chr2: 62–80 Mb at 25 kb resolution). SSLs are marked by black circles on the bottom-left corner.

(B) Heatmap visualization of an SSL in sperm (top) and a typical chromatin loop in the brain (bottom). The resolutions are 25, 10, and 5 kb, respectively.

(C) Boxplot representing the size of chromatin loops and SSLs. Boxes represent the 25th, 50th, and 75th percentiles, while whiskers show 1.5× the interquartile range. The median size is 180 kb for chromatin loops and 5.11 Mb for SSLs. ***p < 0.001 for size comparison between chromatin loops and SSLs (two-sided Wilcoxon rank-sum test).

(D–H) Example regions showing the disappearance of SSLs in embryos (developmental stages 8, 9, and 13) and somatic cells of the brain and liver.

(I) Aggregate peak analysis (APA) of SSLs confirms the disappearance of SSLs in embryos (developmental stages 8, 9, and 13) and somatic cells of the brain and liver.

Two or more biological replicates were carried out for Hi-C (n ≥ 2).

See also Figures S1–S4 and Table S1.

components of PRC1 and PRC2 complexes and found that they were also not enriched at SSL anchors (Figures 2E–2G, S5F, and S6). Instead, we observed active marks such as H3K4me3 and H3K27ac enriched at promoters (Figure S5G). These results imply that clustered SSLs are unlikely established or maintained through the aggregation of heterochromatin.

SSL anchors are enriched explicitly with Helitron transposons

Since the potential role of proteins and histone marks known to function in chromatin loop formation had been ruled out, we next

wondered if DNA sequences may provide some clues about how SSLs are formed. For this purpose, we searched for DNA sequences with different genetic features across SSL-spanned genomic regions. First, we screened DNA sequences around the anchor regions but failed to identify consensus sequences of any known transcription factors in animals. Next, we found that gene density, in general, was slightly lower at SSL anchors than at random controls (Figure S7A). Finally, by analyzing the density of different transposons, we identified that Helitrons, the only group of rolling-circle transposons in class II TEs, are enriched explicitly at SSL anchors (Figures 3A, 3B, and S7B).
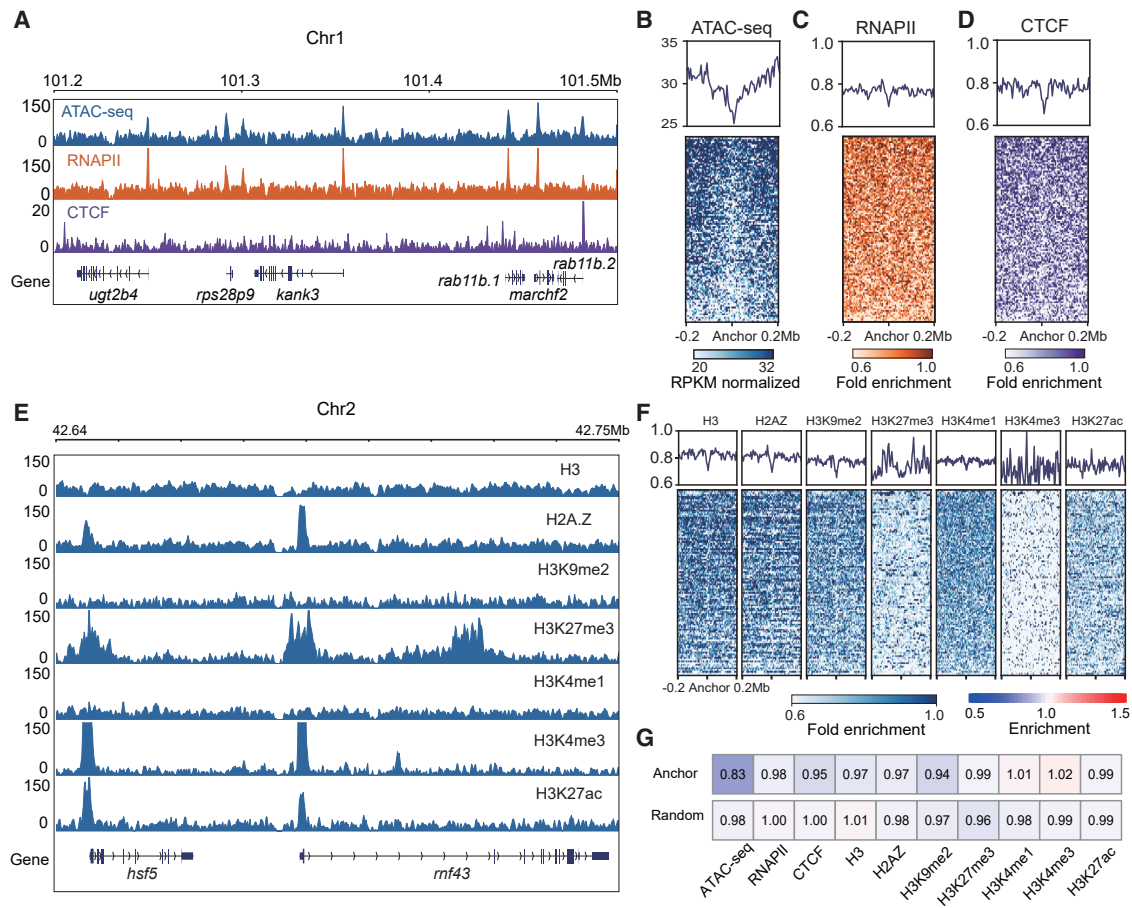
Figure 2. SSL anchors are located in closed chromatin
(A) An example region showing chromatin accessibility, RNAPII, and CTCF occupancy in *X. tropicalis* sperm.
(B–D) The density of ATAC-seq and ChIP-seq signals for RNAPII and CTCF around SSL anchors (±0.2 Mb).
(E) An example region showing the enrichment of histone modifications in *X. tropicalis* sperm.
(F) The density of histone modification signals is shown around SSL anchors (±0.2 Mb).
(G) Comparison of the relative enrichment of histone variants and modifications signals at SSL anchors and random control loci in the genome.
At least two biological replicates were carried out for ATAC-seq and ChIP-seq (n = 2).
See also Figures S5 and S6.

However, at this time, we cannot determine whether Helitron transposition mediates SSL formation.

Because SSLs exist only in *X. tropicalis* sperm but not in the sperm of other species, including *X. laevis*, we asked whether Helitrons in humans, monkeys, mice, or zebrafish genomes are different from those in *X. tropicalis*. A comparison of the different vertebrate genomes shows that the total number of Helitrons in mammals is much lower (Figure 3C). Moreover, although there are thirteen times more Helitrons in zebrafish (Figure 3C), the largest Helitrons are in *X. tropicalis* (Figure 3D). The lower number of Helitrons in mammalian genomes might pose a less immediate source of threat to genome integrity. It is also possible that a Helitron repression/deletion mechanism has long been established in mammalian cells. Nevertheless, even in *X. tropicalis*, many genomic loci harboring Helitrons do not form SSLs (Figures 3E and S7B–S7D), suggesting that unknown variances exist in different Helitrons.

Langevin dynamics simulations suggest that SSLs may form through molecular agent-facilitated condensation
Tracing the dynamic change in chromosome conformation through spermiogenesis is technically challenging. Therefore, we took an alternative approach by conducting coarse-grained Langevin dynamics simulations on the SSL formation process. We first constructed a polymer of 5,000 beads to represent part of the chromatin. SSL anchors are represented by five short segments (ten beads) separated by 800 beads. We then analyzed the number of clusters formed by the anchors as a function of time. In each simulation trajectory, the number ranged from one (all loop anchors cluster together simultaneously) to five (no contact between any pair of loop anchors). We ran ten independent simulations to mimic loop formation from different cells and calculated the average number of clusters against time. The statistical uncertainty reflects the variance between the stochastic clustering processes. Since each anchor
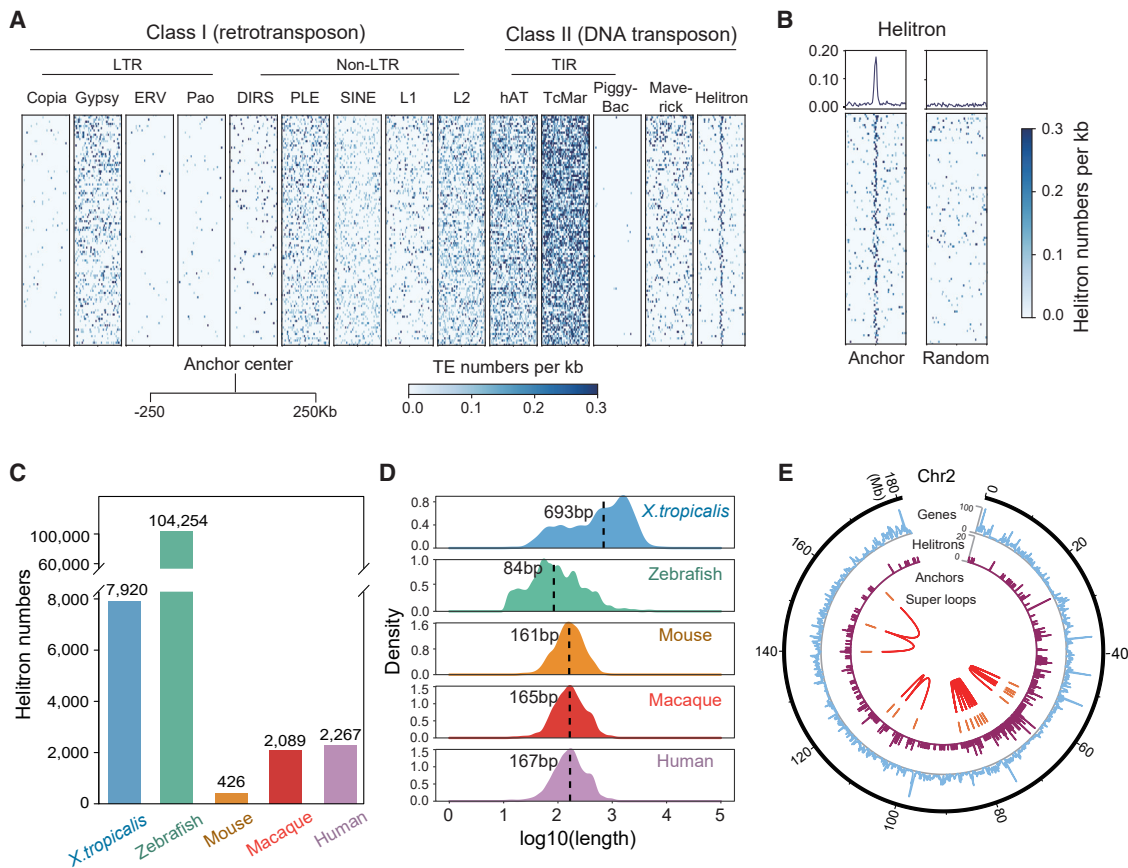
Figure 3. SSL anchors are enriched explicitly with Helitron transposons

(A) The density of class I retrotransposons and class II DNA transposons around SSL anchors are calculated and displayed (number of TEs per kb) (±250 kb).

(B) Enrichment of Helitrons at SSL anchors with randomly shuffled genomic regions as controls (number of TEs per kb) (±250 kb).

(C) Number of Helitrons in different species.

(D) Size distribution of Helitrons in different species. The dashed lines show the median lengths of Helitrons in each species.

(E) Circos plot shows anchors of clustered SSLs on chromosome 2 enriched with Helitrons.

See also Figure S7.

in our simulation has multiple interacting sites, we assigned weak interactions between them. Our results revealed that multivalent weak homotypic attraction between the anchors is insufficient to establish SSLs. As shown in Figure 4A, the average number of clusters is approximately five throughout the simulation, meaning anchors remain well separated (Figures 4B and 4C). In a control simulation system, we demonstrate that reducing the contour distance between the anchors allows them to cluster up (Figure S8A). From a polymer physics point of view, long polymer segments between anchors pose considerable entropic barriers, preventing them from being close to each other. It is worth noting that strong multivalent interactions between loop anchors can also drive the formation of SSLs over a great genomic distance without the aid of agents (Figures S8B–S8G). However, SSLs formed through strong and direct interactions between loop anchors could be too stable to be reversible. Nevertheless, we do not exclude this possibility at this stage.

For anchors with long contour distances, we wondered if there are any external driving forces underlying their clustering. We postulated that there are molecular agents that might facilitate

the establishment of SSLs. We added agent particles into our simulation and assigned cohesive agent-anchor interactions to test this idea. We designed two agents: a weakly self-attractive agent that does not condense (Figure 4D) and a strongly self-attractive agent that can phase separate (Figure 4E). Our simulation showed that the weak agent caused anchors to cluster gradually (Figure 4F). Although all anchors may form a single cluster, as in one simulation example (Figure 4G), the average number of clusters reaches two as the systems evolve, indicating different subsets of loops formed in different cells. Bright spots in the contact matrix indicate anchor-anchor contacts (Figures 4G–4I).

When we examined the effect of the strong agent, we observed that the clustering process was remarkably accelerated (Figure 4J), with even more intense contact signals between anchors under this condition (Figure 4K). We show a typical simulation snapshot in Figure 4L, with the zoomed-in panel revealing the condensed agent surrounding the cluster. The effect on phase separation from the strong agent is evident (Figure 4M) compared with the homogeneous distribution of the weak agent (Figure 4I). To summarize, our designed simulations
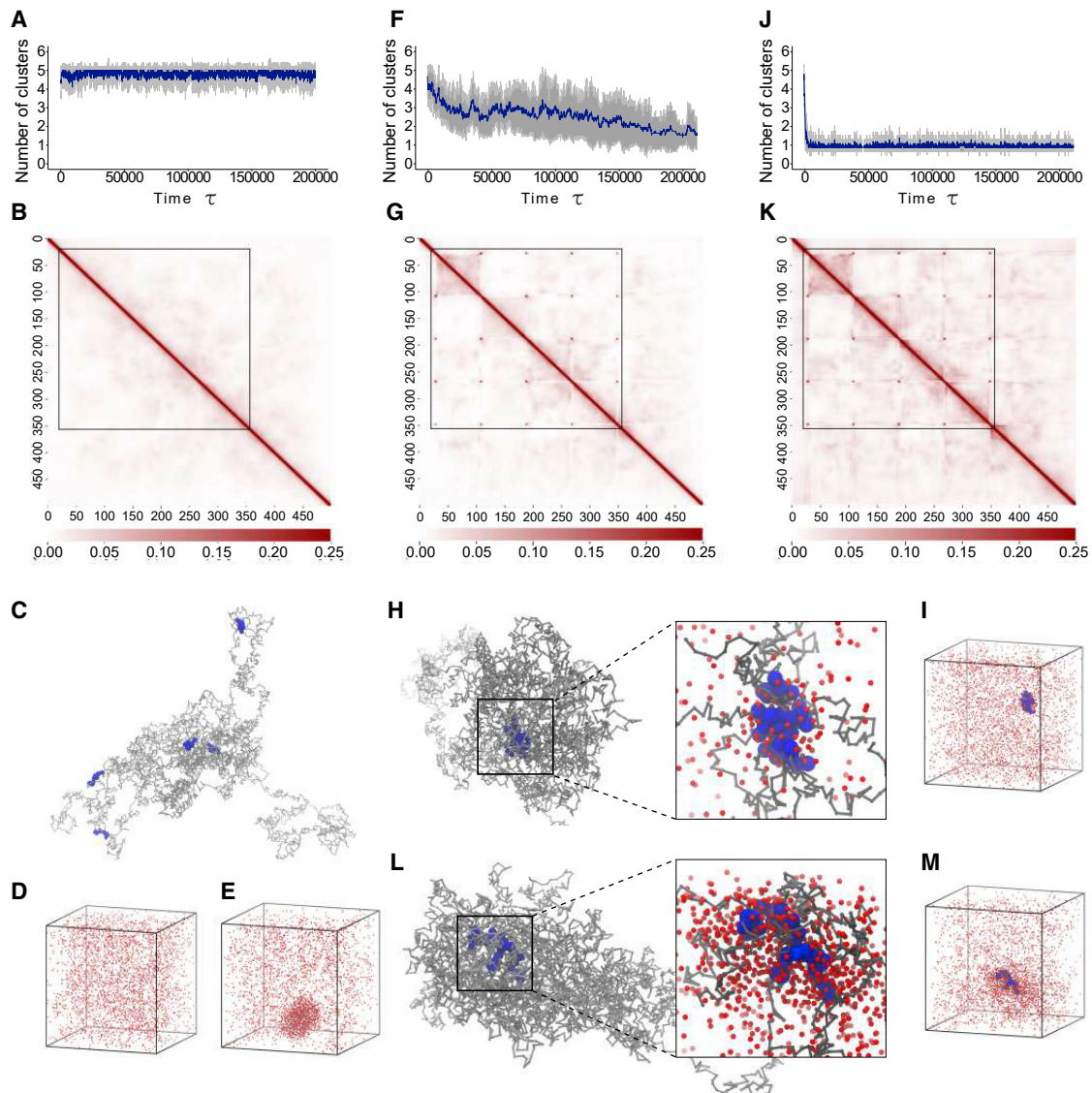
Figure 4. Molecular dynamics simulations suggest that SSL may form through molecular agent-facilitated condensation

(A) The number of clusters as a function of time in the absence of agents. The gray area shows the uncertainty.

(B) The average contact matrix of the simulation without agents. The color bar corresponds to the contact probability of 0–1.

(C) A typical simulation snapshot shows that anchors (blue) cannot cluster without agents.

(D) Weak agents do not phase separate on their own in a simulation box.

(E) Strong agents phase separate on their own.

(F) The average number of clusters as a function of time in the presence of a weak agent.

(G) The average contact matrix of the simulation in the presence of a weak agent.

(H) A typical simulation snapshot shows that anchors (blue) cluster in the presence of weak agents (red, only shown in the zoomed-in panel).

(I) A weak agent does not phase separate, as it facilitates the clustering of the anchors.

(J) The average number of clusters as a function of time in the presence of strong agents.

(K) The average contact matrix of the simulations in the presence of a strong agent.

(L) A typical simulation snapshot shows that anchors (blue) cluster in the presence of strong agents (red, only shown in the zoomed-in panel).

(M) A strong agent phase separates as it facilitates the clustering of anchors.

See also Figure S8.

show that weak and strong agents can promote the clustering of anchors. At this point, we cannot determine whether the SSL pattern is due to an entire cluster of all anchors or a superposition of different subsets of loops.

SSL anchors transit into TAD borders in embryos and differentiated cells

Although monitoring the SSL formation process is technically challenging, we can trace what these structures transit during
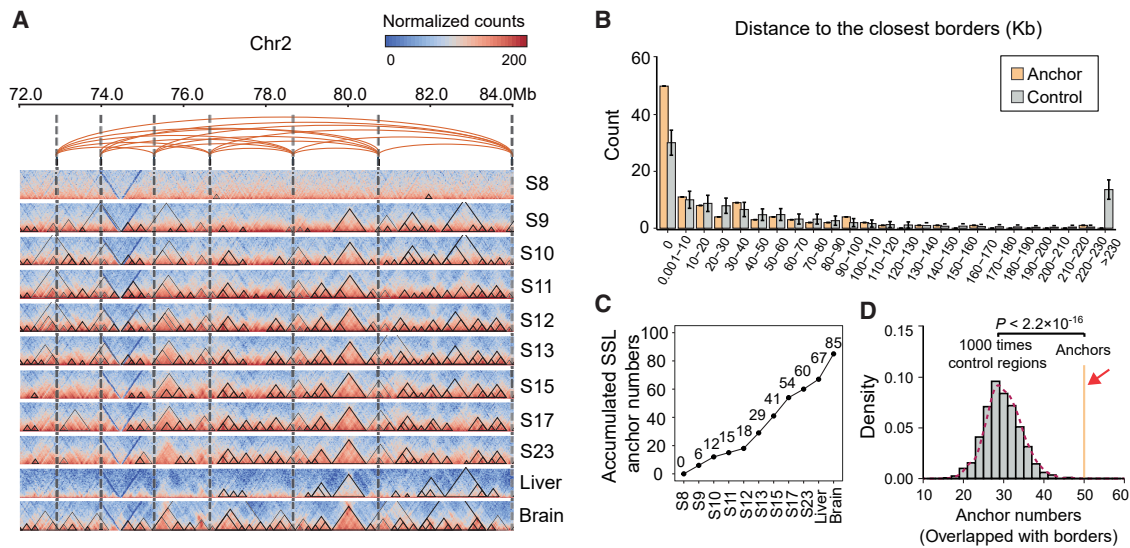
Figure 5. SSL anchors transit into TAD borders in embryos and differentiated cells

(A) An example of a region that shows the transition of SSL anchors into TAD borders during development and in brain and liver cells. Black triangles depict TAD domains. Interactions between SSLs are shown as orange arcs, and the locations of SSL anchors are shown as vertical dashed lines.

(B) Distance distribution of SSL anchors to the closest TAD borders. The shuffled non-SSL anchor regions containing Helitrons are repeated 1,000 times for comparison and shown as the mean ± SD.

(C) Accumulation of SSL anchor transit into TAD borders during embryogenesis and differentiation.

(D) The number of SSL anchors overlapping with TAD borders shows that SSL anchors tend to transit into TAD borders more than the non-SSL anchor regions containing Helitrons. p < 2.2 × 10^{-16} from Student's t test.

See also Figure S9.

embryogenesis. To this end, we examined the chromatin structure of SSL anchors in embryos from multiple developmental stages using our previously published Hi-C datasets.[35] Interestingly, SSL anchors tended to be at or near TAD borders (Figure 5A). We then combined all the TAD borders in embryos and somatic cells and calculated the average distance of SSL anchors to the nearest TAD borders. As shown in Figure 5B, approximately 50% of SSL anchors are located at TAD borders, much higher than non-SSL anchor Helitron sites. During embryo development, up to 60% of SSL anchors are located at or immediately adjacent to TAD borders. An additional 25% are at or adjacent to TAD borders in the liver and brain (Figure 5C). Furthermore, a 1,000-time permutation confirmed that the localization of SSL anchors at TAD borders is statistically significant compared with a non-SSL anchor Helitron control (Figure 5D; p < 2.2 × 10^{-16}).

DISCUSSION

Spermatogenesis involves homologous crossover and recombination, control of TEs, and nearly complete repackaging of the genome to generate sperm whose primary function is to transmit genetic materials required for reproduction fitness. However, how the genome is regulated through spermatogenesis remains to be investigated. We show that genome folding in vertebrate sperm is drastically diverse through evolution.

Unexpectedly, we discovered many clusters of multimegabase SSLs in the *X. tropicalis* sperm genome. SSL anchor sequences are devoid of proteins and histone modifications

implicated in long-range interactions. The superlong distances between SSL anchors cause a nearly unsurmountable energy barrier that prevents the interaction between anchors. Although loop extrusion can create chromatin loops in an energy-driven manner, previous simulations have shown that typical loops formed by this mechanism are under 240 kb.[36] Our molecular dynamics simulation findings indicate that molecular agent-mediated condensation or phase separation may facilitate the formation of SSLs. However, the molecular agent that associates with SSL anchors and mediates condensation or phase separation during spermiogenesis is still being determined. Nevertheless, the discovery of SSLs reminds us that more unknown 3D genome structures might exist in other cell types or at specific developmental stage. Thus, our study highlights the importance of mapping genome folding patterns in more cell types and other organisms to uncover 3D genome structures that are still unknown.

Another surprising discovery from our study is the enrichment of transposon Helitrons at SSL anchors. Helitros were first identified by *in silico* prediction[26] and have been reported in the eukaryotic genomes of fungi,[37] plants, and animals.[38–41] Helitrons transpose through a rolling-circle replication process that requires spatial proximity between the donor and the target sites.[42,43] The transposition of TEs is potentially harmful to genome integrity because TE insertions may disrupt genes and regulatory elements, which can cause many types of human diseases.[44] Thus, the maintenance of genomic integrity is critically important in eukaryotic cells, especially in reproductive cells such as sperm and oocytes. Additional Helitron knockout

experiments and Helitron transposition activity detection assays are required to reveal the relationship between the establishment of SSLs and the transposition (or repression of transposition) of Helitrons. It is also possible that newly replicated Helitron DNA, enzymes involved in transposition, and other factors might act as molecular agents to facilitate SSL formation through condensation.

Although we noticed slightly more genes expressed at relatively late embryo developmental stages (Figure S9), we cannot conclude whether SSL anchors play a role in gene expression regulation during embryogenesis. Nevertheless, our results suggest that a genome can adopt an unexpected structure such as SSL to deal with a possible deleterious transposition of Helitrons during spermatogenesis in *X. tropicalis*.

In summary, we unveiled an evolutionarily distinct and sperm-specific 3D genome structure of SSLs whose anchors are enriched with transposon Helitrons. This study adds an additional type of genome structure to the current list of 3D genome folding patterns. In addition, our work provides insight into the intriguing relationship between TEs and the 3D genome structure.

Limitations of the study

Although we discover that Helitron-demarcated clusters of supersized chromatin loops can be established in the mature sperm of *X. tropicalis*, we note the lack of experimental evidence confirming the requirement of Helitrons. If Helitrons are later proven critical for establishing SSL clusters, then why can only a small portion of them establish such structures? What other factors affect the selective establishment of SSLs at Helitron-containing genomic loci? More importantly, what are the functions of SSL clusters? Fully addressing these questions will expand our understanding of the complex principles of 3D genome folding, Helitron transposition control, and the biological functions of Helitrons during animal development.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
  - Frogs
- METHOD DETAILS
  - Sperm isolation
  - Cross-fertilization
  - Embryo fixation and crosslinking
  - Human K562, mouse ESC, and frog sperm crosslinking
  - RNA isolation and sequencing
  - Hi-C library preparation
  - ChIP-seq library preparation
  - Hi-C data processing

- Spike-in Hi-C
- Hybrid embryo Hi-C
- Calculation of contact heatmap resolution
- A/B compartment analysis
- Reproducibility score
- Contact probability $P(s)$ plots
- Insulation score analysis
- TAD calling
- Super loop calling
- Aggregate peak analysis
- ChIP-seq analysis
- ATAC-seq analysis
- Hybrid embryo RNA-seq analysis
- Motif search
- Molecular dynamics simulation
- QUANTIFICATION AND STATISTICAL ANALYSIS

REFERENCES

1. Oudelaar, A.M., and Higgs, D.R. (2021). The relationship between genome structure and function. Nat. Rev. Genet. *22*, 154–168. https://doi.org/10.1038/s41576-020-00303-x.

2. Misteli, T. (2020). The self-organizing genome: principles of genome architecture and function. Cell *183*, 28–45. https://doi.org/10.1016/j.cell.2020.09.014.

3. Zheng, H., and Xie, W. (2019). The role of 3D genome organization in development and cell differentiation. Nat. Rev. Mol. Cell Biol. *20*, 535–550. https://doi.org/10.1038/s41580-019-0132-4.

4. Rowley, M.J., and Corces, V.G. (2016). The three-dimensional genome: principles and roles of long-distance interactions. Curr. Opin. Cell Biol. *40*, 8–14. https://doi.org/10.1016/j.ceb.2016.01.009.

5. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293. https://doi.org/10.1126/science.1181369.

6. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472. https://doi.org/10.1016/j.cell.2012.01.010.

7. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385. https://doi.org/10.1038/nature11049.

8. Hou, C., Li, L., Qin, Z.S., and Corces, V.G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol. Cell *48*, 471–484. https://doi.org/10.1016/j.molcel.2012.08.031.

9. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380. https://doi.org/10.1038/nature11082.

10. Jung, I., Schmitt, A., Diao, Y., Lee, A.J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., et al. (2019). A compendium of promoter-centered long-range chromatin interactions in the human genome. Nat. Genet. *51*, 1442–1449. https://doi.org/10.1038/s41588-019-0494-8.

11. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004.

12. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell *162*, 900–910. https://doi.org/10.1016/j.cell.2015.07.038.

13. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.

14. Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295. https://doi.org/10.1016/j.cell.2013.04.053.

15. Hou, C., Zhao, H., Tanimoto, K., and Dean, A. (2008). CTCF-dependent enhancer-blocking by alternative chromatin loop formation. Proc. Natl. Acad. Sci. USA *105*, 20398–20403. https://doi.org/10.1073/pnas.0808506106.

16. Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311. https://doi.org/10.1126/science.1067799.

17. Lu, J.Y., Chang, L., Li, T., Wang, T., Yin, Y., Zhan, G., Han, X., Zhang, K., Tao, Y., Percharde, M., et al. (2021). Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. Cell Res. *31*, 613–630. https://doi.org/10.1038/s41422-020-00466-6.

18. Zhang, Y., Li, T., Preissl, S., Amaral, M.L., Grinstein, J.D., Farah, E.N., Destici, E., Qiu, Y., Hu, R., Lee, A.Y., et al. (2019). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat. Genet. *51*, 1380–1388. https://doi.org/10.1038/s41588-019-0479-7.

19. Grob, S., Schmid, M.W., and Grossniklaus, U. (2014). Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Mol. Cell *55*, 678–693. https://doi.org/10.1016/j.molcel.2014.07.009.

20. Ågren, J.A., and Clark, A.G. (2018). Selfish genetic elements. PLoS Genet. *14*, e1007700. https://doi.org/10.1371/journal.pgen.1007700.

21. Judd, J., Sanderson, H., and Feschotte, C. (2021). Evolution of mouse circadian enhancers from transposable elements. Genome Biol. *22*, 193. https://doi.org/10.1186/s13059-021-02409-9.

22. Ellison, C.E., and Bachtrog, D. (2015). Non-allelic gene conversion enables rapid evolutionary change at multiple regulatory sites encoded by transposable elements. Elife *4*, e05899. https://doi.org/10.7554/eLife.05899.

23. Senft, A.D., and Macfarlan, T.S. (2021). Transposable elements shape the evolution of mammalian development. Nat. Rev. Genet. *22*, 691–711. https://doi.org/10.1038/s41576-021-00385-1.

24. Jangam, D., Feschotte, C., and Betrán, E. (2017). Transposable element domestication as an adaptation to evolutionary conflicts. Trends Genet. *33*, 817–831. https://doi.org/10.1016/j.tig.2017.07.011.

25. Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. *18*, 71–86. https://doi.org/10.1038/nrg.2016.139.

26. Kapitonov, V.V., and Jurka, J. (2001). Rolling-circle transposons in eukaryotes. Proc. Natl. Acad. Sci. USA *98*, 8714–8719. https://doi.org/10.1073/pnas.151269298.

27. Rathke, C., Baarends, W.M., Awe, S., and Renkawitz-Pohl, R. (2014). Chromatin dynamics during spermiogenesis. Biochim. Biophys. Acta *1839*, 155–168. https://doi.org/10.1016/j.bbagrm.2013.08.004.

28. Chen, X., Ke, Y., Wu, K., Zhao, H., Sun, Y., Gao, L., Liu, Z., Zhang, J., Tao, W., Hou, Z., et al. (2019). Key role for CTCF in establishing chromatin structure in human embryos. Nature *576*, 306–310. https://doi.org/10.1038/s41586-019-1812-0.

29. Wang, Y., Wang, H., Zhang, Y., Du, Z., Si, W., Fan, S., Qin, D., Wang, M., Duan, Y., Li, L., et al. (2019). Reprogramming of meiotic chromatin architecture during spermatogenesis. Mol. Cell *73*, 547–561.e6. https://doi.org/10.1016/j.molcel.2018.11.019.

30. Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., Yao, X., Li, F., Zhu, W., Gao, L., et al. (2017). 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. Cell *170*, 367–381.e20. https://doi.org/10.1016/j.cell.2017.06.029.

31. Jung, Y.H., Sauria, M.E.G., Lyu, X., Cheema, M.S., Ausio, J., Taylor, J., and Corces, V.G. (2017). Chromatin states in mouse sperm correlate with embryonic and adult regulatory landscapes. Cell Rep. *18*, 1366–1382. https://doi.org/10.1016/j.celrep.2017.01.034.

32. Wike, C.L., Guo, Y., Tan, M., Nakamura, R., Shaw, D.K., Díaz, N., Whittaker-Tademy, A.F., Durand, N.C., Aiden, E.L., Vaquerizas, J.M., et al. (2021). Chromatin architecture transitions from zebrafish sperm through early embryogenesis. Genome Res. *31*, 981–994. https://doi.org/10.1101/gr.269860.120.

33. Zenk, F., Zhan, Y., Kos, P., Löser, E., Atinbayeva, N., Schächtle, M., Tiana, G., Giorgetti, L., and Iovino, N. (2021). HP1 drives de novo 3D genome

reorganization in early Drosophila embryos. Nature *593*, 289–293. https://doi.org/10.1038/s41586-021-03460-z.

34. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G.L., Chang, J.M., and Cavalli, G. (2018). Polycomb-dependent chromatin looping contributes to gene silencing during Drosophila development. Mol. Cell *71*, 73–88.e5. https://doi.org/10.1016/j.molcel.2018.05.032.

35. Niu, L., Shen, W., Shi, Z., Tan, Y., He, N., Wan, J., Sun, J., Zhang, Y., Huang, Y., Wang, W., et al. (2021). Three-dimensional folding dynamics of the Xenopus tropicalis genome. Nat. Genet. *53*, 1075–1087. https://doi.org/10.1038/s41588-021-00878-z.

36. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of chromosomal domains by loop extrusion. Cell Rep. *15*, 2038–2049. https://doi.org/10.1016/j.celrep.2016.04.085.

37. Cultrone, A., Domínguez, Y.R., Drevet, C., Scazzocchio, C., and Fernández-Martín, R. (2007). The tightly regulated promoter of the xanA gene of Aspergillus nidulans is included in a helitron. Mol. Microbiol. *63*, 1577–1587. https://doi.org/10.1111/j.1365-2958.2007.05609.x.

38. Grabundzija, I., Messing, S.A., Thomas, J., Cosby, R.L., Bilic, I., Miskey, C., Gogol-Döring, A., Kapitonov, V., Diem, T., Dalda, A., et al. (2016). A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nat. Commun. *7*, 10716. https://doi.org/10.1038/ncomms10716.

39. Xiong, W., He, L., Lai, J., Dooner, H.K., and Du, C. (2014). HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. Proc. Natl. Acad. Sci. USA *111*, 10263–10268. https://doi.org/10.1073/pnas.1410068111.

40. Yang, L., and Bennetzen, J.L. (2009). Structure-based discovery and description of plant and animal Helitrons. Proc. Natl. Acad. Sci. USA *106*, 12832–12837. https://doi.org/10.1073/pnas.0905563106.

41. Yang, L., and Bennetzen, J.L. (2009). Distribution, diversity, evolution, and survival of Helitrons in the maize genome. Proc. Natl. Acad. Sci. USA *106*, 19922–19927. https://doi.org/10.1073/pnas.0908008106.

42. Grabundzija, I., Hickman, A.B., and Dyda, F. (2018). Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. Nat. Commun. *9*, 1278. https://doi.org/10.1038/s41467-018-03688-w.

43. Kapitonov, V.V., and Jurka, J. (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet. *23*, 521–529. https://doi.org/10.1016/j.tig.2007.08.004.

44. Goodier, J.L., and Kazazian, H.H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell *135*, 23–35. https://doi.org/10.1016/j.cell.2008.09.022.

45. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595. https://doi.org/10.1093/bioinformatics/btp698.

46. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. *3*, 95–98. https://doi.org/10.1016/j.cels.2016.07.002.

47. Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke, T., Backofen, R., Ramírez, F., and Grüning, B.A. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids Res. *46*, W11–W16. https://doi.org/10.1093/nar/gky504.

48. Yu, W., He, B., and Tan, K. (2017). Identifying topologically associating domains and subdomains by Gaussian Mixture model and Proportion test. Nat. Commun. *8*, 535. https://doi.org/10.1038/s41467-017-00478-8.

49. Shin, H., Shi, Y., Dai, C., Tjong, H., Gong, K., Alber, F., and Zhou, X.J. (2016). TopDom: an efficient and deterministic method for identifying topological domains in genomes. Nucleic Acids Res. *44*, e70. https://doi.org/10.1093/nar/gkv1505.

50. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

51. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

52. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

53. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

54. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

55. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

56. Gibeaux, R., Acker, R., Kitaoka, M., Georgiou, G., van Kruijsbergen, I., Ford, B., Marcotte, E.M., Nomura, D.K., Kwon, T., Veenstra, G.J.C., and Heald, R. (2018). Paternal chromosome loss and metabolic crisis contribute to hybrid inviability in Xenopus. Nature *553*, 337–341. https://doi.org/10.1038/nature25188.

57. Akkers, R.C., van Heeringen, S.J., Jacobi, U.G., Janssen-Megens, E.M., Françoijs, K.J., Stunnenberg, H.G., and Veenstra, G.J.C. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. Dev. Cell *17*, 425–434. https://doi.org/10.1016/j.devcel.2009.08.005.

58. Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature *523*, 240–244. https://doi.org/10.1038/nature14450.

59. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208. https://doi.org/10.1093/nar/gkp335.

60. Thompson, A.P., Aktulga, H.M., Berger, R., Bolintineanu, D.S., Brown, W.M., Crozier, P.S., in 't Veld, P.J., Kohlmeyer, A., Moore, S.G., Nguyen, T.D., et al. (2022). LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. Comput. Phys. Commun. *271*, 108171.

61. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33-38–27-38.

62. Stukowski, A. (2009). Visualization and analysis of atomistic simulation data with OVITo the Open Visualization Tool. Model. Simul. Mat. Sci. Eng. *18*, 015012.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-POLR2B	ABclonal	cat#A5928; RRID:AB_2766666
Rabbit anti-CTCF	Active Motif	cat#61311; RRID:AB_2614975
Rabbit anti-H2AK119ub1	Cell Signaling Technology	cat#8240
Rabbit anti-RING1B	Cell Signaling Technology	cat#5694
Rabbit anti-EZH2	Cell Signaling Technology	cat#5246
Rabbit anti-SUZ12	Cell Signaling Technology	cat#3737
Rabbit anti-histone H3	Proteintech	cat#17168-1-AP; RRID:AB_2716755
Rabbit anti-histone H2A.Z	Active Motif	cat#39943; RRID:AB_2793401
Rabbit anti-histone H3K4me1	Active Motif	cat#39297; RRID:AB_2615075
Rabbit anti-histone H3K4me3	Active Motif	cat#39159; RRID:AB_2615077
Rabbit anti-histone H3K27me3	Active Motif	cat#39155; RRID:AB_2561020
Rabbit anti-histone H3K27ac	Active Motif	cat#39133
Rabbit anti-H3K9me2	ABclonal	cat#A2359; RRID:AB_2764319
Chemicals, peptides, and recombinant proteins		
Biotin-14-dCTP	LifeTechnologies	cat#19518018
DpnII	New England Biolabs	cat#R0543L
DNA PolymeraseI Klenow Fragment	New England Biolabs	cat#M0210L
T4 DNA Ligase	New England Biolabs	cat#M0202L
Buffer 3.1	New England Biolabs	cat#B7203
Dynabeads MyOne Streptavidin C1	LifeTechnologies	cat#65002
Proteinase K	Transgen	cat#GE201-01
RNase A	Transgen	cat#GE101-01
TransStart® FastPfu Fly DNA Polymerase	Transgen	cat#AP231-13
Critical commercial assays		
NEBNext Ultra RNA Library Prep Kit for Illumina	New England Biolabs	cat#E7530L
TruePrep® DNA Library Prep Kit V2	Vazyme	cat#TD501
VAHTS Universal DNA Library Prep Kit	Vazyme	cat#ND607-03
TransZol Up Plus RNA Kit	Transgen	cat#ER501-01
Deposited data		
X.tropicalis sperm Hi-C, 6 replicates	Niu et al.[35]	PRJNA606649
X.tropicalis sperm Hi-C, 2 replicates	This study	GEO:GSE197877 GSA: CRA006259
K562 Spike-in Hi-C	This study	GEO:GSE197877 GSA: CRA006259
X. laevis sperm Hi-C (K562 spike-in)	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm H2AK119ub1 ChIP-seq (mESC spike-in)	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm Ring1B ChIP-seq (mESC spike-in)	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm EZH2 ChIP-seq (mESC spike-in)	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm SUZ12 ChIP-seq (mESC spike-in)	This study	GEO:GSE197877 GSA: CRA006259

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
X.tropicalis sperm ATAC-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm RPB2 ChIP-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm CTCF ChIP-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm SMC1 ChIP-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm H3 ChIP-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm H2A.Z ChIP-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm H3K4me1 ChIP-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm H3K4me3 ChIP-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm H3K9me2 ChIP-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm H3K27ac ChIP-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis sperm H3K27me3 ChIP-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis Brain Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis liver Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis stage 8 embryo Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis stage 9 embryo Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis stage 10 embryo Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis stage 11 embryo Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis stage 12 embryo Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis stage 13 embryo Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis stage 15 embryo Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis stage 17 embryo Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis stage 23 embryo Hi-C	Niu et al.[35]	PRJNA606649
X.tropicalis (male) x *X.laevis (female)* stage 33 hybrid embryos Hi-C	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* tail of hybrid embryos Hi-C	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 3 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 5 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 7 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 8 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 9 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
X.tropicalis (male) x *X.laevis (female)* stage 10 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 11 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 12 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 13 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 16 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 24 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 34 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
X.tropicalis (male) x *X.laevis (female)* stage 41 hybrid embryos RNA-seq	This study	GEO:GSE197877 GSA: CRA006259
Human sperm Hi-C	Chen et al.[28]	GSA:CRA000852
Monkey (*Macaca mulatta*) sperm Hi-C	Wang et al.[29]	GEO: GSE109344
Mouse sperm Hi-C (a)	Jung et al.[31]	GEO: GSE79230
Mouse sperm Hi-C (b)	Ke et al.[30]	GSA: CRA000108
Zebrafish (*Danio rerio*)	Wike et al.[32]	GEO: GSE152744
Experimental models: Cell lines		
Human K562 cells	ATCC	CCL-243
Experimental models: Organisms/strains		
Xenopus tropicalis	Nasco (Fort Atkinson, WI, USA)	N/A
Xenopus laevis	Nasco (Fort Atkinson, WI, USA)	N/A
Hybrid embryos of male Xenopus tropicalis and female Xenopus laevis	This study	N/A
Software and algorithms		
BWA (0.7.17)	Li et al.[45]	https://github.com/lh3/bwa
Distiller (v0.3.3)	N/A	https://github.com/open2c/distiller-nf
Pairtools (v0.3.0)	N/A	https://github.com/open2c/pairtools
Juicer/juicebox	Durand et al.[46]	https://github.com/aidenlab/juicer
HiCExplorer (v3.6)	Wolff et al.[47]	https://github.com/deeptools/HiCExplorer
GenomeDISCO	N/A	https://github.com/kundajelab/genomedisco
rGMAP	Yu et al.[48]	http://tanlab4generegulation.org/rGMAP_1.1.tar.gz.
TopDom	Shin et al.[49]	http://zhoulab.usc.edu/TopDom/
Trimmomatic (v0.38)	Bolger et al.[50]	http://www.usadellab.org/cms/index.php?page=trimmomatic
Samtools (v1.7)	Danecek et al.[51]	https://github.com/samtools/samtools
PicardTools (v2.25.0)	N/A	http://broadinstitute.github.io/picard/
Deeptools (v3.5.0)	Ramírez et al.[52]	http://deeptools.ie-freiburg.mpg.de/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MACS2 (v2.2.7.1)	Zhang et al.[53]	https://github.com/taoliu/MACS/
STAR (v2.7.9a)	Dobin et al.[54]	https://github.com/alexdobin/STAR/releases
RSEM (v1.2.28)	Li et al.[55]	http://deweylab.biostat.wisc.edu/rsem
R	R Core Team	https://www.r-project.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Chunhui Hou (houchunhui@mail.kiz.ac.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All raw sequencing data generated in this study have been deposited under accession numbers GSE197877 in the NCBI GEO repository and CRA006259 in the National Genomics Data Center of China. The Hi-C sequencing data used in this study are publicly available, and their accession numbers are listed in the key resources table.
- The custom codes used in this study are available at https://github.com/hzjsxu/SSCL and https://doi.org/10.5281/zenodo.7580981.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

Human erythroleukemic K562 cells were grown in RPMI 1640 (Corning, 10-040-CV) with 10% fetal bovine serum and 1% penicillin/streptomycin (Sigma, P0781) to a density between 500,000 and 1 million cells per mL at 37°C. Mouse ESCs were cultured on feeder-free dishes coated with 0.2% gelatin (Millipore, Cat# 901771) in N2B27$^{2i/lif}$ medium, which included N2B27 medium supplemented with 1 μM PD035901 (LC Laboratories, Cat# P-9688), 3 μM Chir99021 (LC Laboratories, Cat# C-6556), and 1x10^3 units/mL hLIF (human Leukemia Inhibitory Factor, Millipore, Cat# ESG1107).

Frogs

All animal procedures were conducted following international standards and were approved by the Animal Care and Use Committee of the Southern University of Science and Technology (Approval No: SUSTech-JY2021024Z). Wild-type male *X. tropicalis* (one to two years old) and *Xenopus laevis* (*X. laevis*) (two to three years old) of both sexes were used for sperm isolation or cross-fertilization.

METHOD DETAILS

Sperm isolation

We injected one male *X. tropicalis* (one to two years old), or *Xenopus laevis* (*X. laevis*) (two to three years old), with 150 U, or 300 U, of human chorionic gonadotropin (HCG), respectively. After 2 hours, two *Xenopus* testes were collected and homogenized in 1 ml 1x MBS (80 mM NaCl, 10 mM HEPES, 2.4 mM NaHCO$_3$, 1 mM KCl, 0.82 mM MgSO$_4$, 0.33 mM Ca(NO$_3$)$_2$, 0.41 mM CaCl$_2$, pH 7.4) and filtered with a 40 μm strainer to remove tissues and fat. Next, the filtered cell solution was brought to 14 ml with 1x MBS and centrifuged for two rounds at 180xg. Finally, the supernatant was centrifuged at 1500xg, and pure sperm were harvested for experiments.

Cross-fertilization

We carried out ovulation and cross-fertilization[56] with some modifications. Briefly, female *X. laevis* (two to three years old) were injected with 500 U HCG. After 12 hours, male *X. tropicalis* (one to two years old) were injected with 150 U HCG. After 2 hours, *X. laevis* females were squeezed gently to deposit eggs onto a 9 cm clean Petri dish. Two *X. tropicalis* testes were then collected in 0.5 ml 1xMBS and homogenized using razor blades. Any liquid in the Petri dishes was removed, and eggs were fertilized with 0.1 ml of sperm premixed with 0.9 ml of water. Eggs were swirled as much as possible to ensure that they were all separated and then incubated for 10 min. Jelly coats were removed with a 2% cysteine solution (NaOH was used to adjust the pH to 7.8-8.0). After extensive washing (at least four times) with 0.1xMBS, embryos were incubated at 23°C and collected at the desired stages for RNA-seq and Hi-C.

Embryo fixation and crosslinking

Hybrid embryos (NF developmental stages 3, 5, 7, 8, 9, 10, 11, 12, 13, 16, 24, 34, and 41; sex unknown, mixed) were collected and fixed for 40 min in 1.5% formaldehyde. Fixation was stopped after a 10 min incubation in 0.125 M glycine/0.1 × MBS, followed by three washes with 0.1 × MBS. Fixed embryos were frozen at −80°C in 1.5 ml microcentrifuge tubes (200 embryos per tube).

Human K562, mouse ESC, and frog sperm crosslinking

Ten million K562, or mouse embryonic stem cells, or frog sperm cells were centrifuged and resuspended in 10 ml fresh RPMI medium and 1xMBS, respectively. Cells were fixed with 0.28 ml of 37% formaldehyde for 10 minutes at room temperature (RT). The reaction was then stopped by adding 0.56 ml 2.5 M glycines. The cell suspension was incubated for 5 minutes at RT, followed by 15 minutes on ice. The suspension of cells was centrifuged at 1000-1500xg for 5 minutes at 4°C. The cell pellet was gently washed twice with 20 ml of ice-cold 1× PBS. At this stage, the fixed cells can be stored at −80°C for up to one year.

RNA isolation and sequencing

Total RNA was extracted from hybrid embryos (NF developmental stages 3, 5, 7, 8, 9, 10, 11, 12, 13, 16, 24, 34, and 41; sex unknown, mixed) using the *TransZol* Up Plus RNA Kit (Transgen, ER501-01). According to the manufacturer's instructions, cDNA libraries were prepared with an NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, E7530 L). The libraries were sequenced on an Illumina HiSeq 2500 platform, generating 150-base paired-end reads.

Hi-C library preparation

Hi-C libraries were optimized according to a previous protocol.[13] Briefly, 1-2 million sperm cells, or 50-200 *X. laevis* (♀) x *X. tropicalis* (♂) embryos, were crosslinked with 1% formaldehyde for 10 min using vacuum infiltration. We added human K562 cells as a spike-in control. Isolated nuclei were digested with 80 U of DpnII (NEB, R0543 L) at 37°C for 4 hrs, and restriction fragment overhangs were marked with biotin-labeled nucleotides. After labeling, chromatin fragments in proximity were ligated with 4000 U T4 DNA ligase for 6 hrs at 16°C. Next, chromatin was reverse crosslinked, purified, and precipitated with ethanol. Biotinylated DNA was sheared into 250-500 bp fragments and pulled down with MyOne Streptavidin T1 beads (Life Technologies, 65602). Immobilized DNA fragments were end-repaired, A-tailed, and ligated with adaptors. Fragments were amplified with the Q5 master mix (NEB, M0492 L). Hi-C libraries were sequenced on the Illumina HiSeq X10 platform (paired-end sequencing 2× 150 bp).

ChIP-seq library preparation

We performed ChIP assays[57] on sperm and spike-in control cells. Briefly, ∼5 million sperm cells were crosslinked with 1% formaldehyde for 10 min. Chromatin was sheared to an average size of 150 bp with a sonicator (Bioruptor Pico, Diagenode). Sonicated chromatin fragments were immunoprecipitated with 3-5 μg of antibody. Chromatin-bound antibodies were recovered with 30 μl Protein A/G magnetic beads (Millipore 16-663). After reverse crosslinking, the ChIPed DNA was recovered using the MinElute Reaction Cleanup Kit (Qiagen 28206) and amplified with the VAHTS® Universal DNA Library Prep Kit for Illumina V3 (Vazyme ND607). Amplified ChIP libraries were sequenced on the Illumina HiSeq X10 platform. Human K562 and mouse embryonic stem cells were used as spike-in controls.

Hi-C data processing

Hi-C raw sequencing reads were aligned to the reference genome using the distiller-nf mapping pipeline with default parameters (https://github.com/mirnylab/distiller-nf). In brief, raw fastq reads were mapped to reference genomes (hg19 for humans, rheMac2 for rhesus macaques, mm10 for mice, xenTro10 for *X. tropicalis*, and danRer11 for zebrafish) using bwa mem.[45] Aligned reads were then classified and deduplicated using pairtools (https://github.com/open2c/pairtools) such that the uniquely mapped reads were retained while duplicated pairs were removed. The retained reads were referred to as valid pairs. Next, the valid pair files were transformed to.hic files using pre-command from juicer tools, and read pairs with a mapping quality score < 1 were filtered out.[46] Finally, all Hi-C contact matrices were normalized using the Knight-Ruiz (KR) method for further analysis. The.hic files were also converted into.cool files using the hic2cool convert command from HiCExplorer.[47]

Spike-in Hi-C

The *X. tropicalis* and human reads were separated from the mixture using the xenTro10 and hg19 reference genomes. The mixed *X. tropicalis* and human reads were first mapped to the xenTro10 and hg19 reference genomes using bwa mem, respectively. The separated data were processed following the "Hi-C data processing" section.

Hybrid embryo Hi-C

The hybrid embryo Hi-C datasets were mapped to the combined *X. tropicalis* v10.0 and *X. laevis* v9.2 reference genomes using bwa mem with parameters '-SP -B 8'. Subsequent analyses were the same as in the "Hi-C data processing" section.

Calculation of contact heatmap resolution

We calculated the Hi-C data resolution[13] defining the contact heatmap resolution as the smallest bin size, with 80% of bins having at least 1,000 contact reads. The following script with minor modifications was used to calculate the contact heatmap resolution: https://github.com/aidenlab/juicer/blob/master/misc/calculate_map_resolution.sh.

A/B compartment analysis

We assigned A/B compartments using HiCExplorer.[5] By hicTransform, 500 kb corrected matrices were used to create the observed/expected matrix and calculate the Pearson correlation matrix. Next, principal component analysis was performed using hicPCA on the Pearson correlation matrix, and PC1 values were used to designate the A/B compartment status. Gene density was also considered such that eigenvalues with higher or lower gene densities were assigned to compartment A or B, respectively.

Reproducibility score

The reproducibility score was calculated with GenomeDISCO (https://github.com/kundajelab/genomedisco), which compares contact maps of 3D genome structures. It uses random walks on the contact map graph for smoothing before comparing the contact maps, resulting in a concordance score that can be used to evaluate the reproducibility of biological replicates. Here, the GenomeDISCO reproducibility score was calculated at 100 kb resolution for biological replicates.

Contact probability *P*(*s*) plots

Contact probability (*P*), a function of genomic separation (*s*), was calculated based on cis reads from the valid pairs using the python script pairsqc.py (https://github.com/4dn-dcic/pairsqc). In brief, logarithmically spaced bins along chromosomes with an increasing factor of $10^{0.1}$ were used. Then, for each bin, the contact probability was computed as number_of_reads/number_of_possible_reads/bin_size. Number_of_reads is the number of interactions at corresponding distances. Number_of_possible_reads was computed as the sum of L_chr – s_mid – 1 over the whole genome, where L_chr is the length of a chromosome. S_mid is the midpoint of the bin at the $\log_{10}$ scale (for example, bin $10^{6.1}$–$10^{6.0}$ has a midpoint of $10^{6.05}$). Bin_size was computed as max distance – min distance (for example, for bin $10^{6.0}$–$10^{6.1}$, the bin_size is $10^{6.1}$–$10^{6.0}$).

Insulation score analysis

We calculated the insulation score using a custom python script[58] using a 25 kb bin size and a 500 kb sliding window. Finally, the insulation score was normalized relative to all insulation scores across each chromosome by calculating the log2 ratio of each bin's insulation score versus the mean of all insulation scores. The valleys or minima along the normalized insulation score indicate the loci of reduced Hi-C interactions and often represent TAD boundaries.

TAD calling

We identified TADs using multiple software. In brief, domains were first annotated using arrowhead,[46] rGMAP,[48] and TopDom[49] with default parameters at 10 kb resolution. A metric defined as the "diamond score" was then used to measure the domains' strength and filter out domains. Domains were filtered out if they had a diamond score <0.6 and a domain size <100 kb. Next, domains detected by all three methods (arrowhead, rGMAP, and TopDom) were merged, and boundaries were set to the bin with the lowest insulation score. Finally, domains located in low contact density areas were also excluded.

Super loop calling

Supersized loops were called with KR-normalized contact matrices at 25 kb resolution using HiCCUPs[46] and filtered with a false discovery rate at 0.1 (calling parameters: -k KR -r 25,000 -f 0.1 -d 25000 -l 5 -p 2--ignore-sparsity). To minimize false positives, we performed manual filtering by visually inspecting heatmaps for each loop. In the end, 144 super loops were annotated in the *X. tropicalis* sperm Hi-C contact map.

Aggregate peak analysis

To measure the enrichment of annotated super loops in the contact matrix, aggregate peak analysis (APA) was performed using juicer tools.[46] In brief, submatrices around super loops were taken from the KR-normalized matrix and then summed. The APA matrix was plotted as a heatmap. The APA score was calculated as the ratio of the central pixels to the mean of the pixels in the lower-left corner. An APA score that is significantly above 1 indicates enrichment.

ChIP-seq analysis

Low-quality raw ChIP reads and adapter sequences were removed by Trimmomatic v0.38.[50] Clean reads were then aligned to the *X. tropicalis* v10.0 reference genome using bwa with default parameters. Low mapping quality (MAPQ <20) and PCR-duplicated reads were removed by SAMtools[51] and Picard (https://broadinstitute.github.io/picard/), respectively. Read coverage of genomic regions for filtered BAM files was computed to assess the reproducibility of replicates with a 10 kb bin size by multiBamSummary bins command in deeptools v3.5.[52] The Pearson correlation coefficients (PCCs) between two replicates were calculated based on the read coverage. If the PCC was larger than 0.85, the replicate BAM files were merged for further peak calling analysis. Peaks

were identified against a corresponding input control using MACS2[53] with the options "macs2 callpeak -f BAM -g 1.4e9 -B -q 0.01 –SPMR –nomodel –shift 0". Fold enrichment (ChIP/Input) was generated from bedGraph files of treat_pileup and control_lambda using MACS2 bdgcmp with the 'FE' (fold enrichment) method. ChIP-seq signal tracks for visualization were generated by bamCoverage with the options "–binSize 10 –extendReads –minMappingQuality 20 –ignoreDuplicates –normalizeUsing RPKM".

ATAC-seq analysis

Raw fastq reads were trimmed and aligned to the *X. tropicalis* v10.0 genome using bwa. Low mapping quality reads and PCR duplicates were removed by SAMtools and Picard. Replicate BAM files were merged after confirming a high Pearson correlation coefficient between any two replicates. Peaks were identified using MACS2 with the options "macs2 callpeak -f BAM -g 1.4e9 -B -q 0.01 –SPMR –nomodel –shift 75 –extsize 150 –call-summits". Fold enrichment was generated from bedGraph files of treat_pileup and control_lambda using MACS2 bdgcmp with the 'FE' (fold enrichment) method. ATAC-seq signal tracks for visualization were generated by bamCoverage with the option "–normalizeUsing RPKM".

Hybrid embryo RNA-seq analysis

Raw fastq reads and adapter sequences were removed by Trimmomatic v0.38. Clean reads were mapped to the combined *X. tropicalis* v10.0 and *X. laevis* v9.2 reference genome using STAR v2.7.9a.[54] To prevent misalignment to other species, we applied the following stringent criteria: ignoring a read mapped to more than one target (–outSAMmultNmax 1 – outSAMmapqUnique 10) and allowing no mismatch (–outFilterMismatchNmax 0). Gene expression TPM was calculated by RSEM v1.2.28.[55]

Motif search

We extracted a 25 kb sequence centered around each SSL anchor and carried out a de novo motif search using MEME (version 5.5.0)[59] with default parameters.

Molecular dynamics simulation

Coarse-grained Langevin dynamics simulations were conducted to model chromatin cluster formation. Polymer chains of 5000 monomers were initialized as random walks. All the monomers in the chain have mass $m=1$ and radius $\sigma=1$. The SSL anchors were modeled by ten successive beads separated by 800 units. Five SSL anchors were defined in the chain to investigate their clustering behavior. The polymer chain was held by harmonic bonds with $k=50$ and $r_0 = \sigma$. The Lennard–Jones potential with $\sigma=1$ was used for nonbonded interactions. To investigate the clustering of SSLs, we added molecular agents to the system as a multivalent interaction provider to crosslink distant anchors. Although multiple molecular species could be involved in the interaction, we introduced only one kind of agent for simplification. The agents were coarse-grained as monomers with relatively strong nonbonded interactions with anchors. For the polymer system in a cubic box of 40σ, 3200 agents with a density of 0.05 were added. We tested different levels of agent interaction strengths to study the effect of molecular agents on anchor clustering. Example systems are shown below.

LJ interaction parameters			
	Strong agent interaction	Weak agent interaction	No agent
Agent - agent	1.2ϵ	0.8ϵ	NA
Agent - anchor	1.4ϵ	1.4ϵ	NA
Anchor - anchor	0.6ϵ	0.6ϵ	0.2ϵ −1.2ϵ
Rest - rest	0.25ϵ	0.25ϵ	0.25ϵ
Anchor - rest	0.2ϵ	0.2ϵ	0.2ϵ
Agent - anchor	0.6ϵ	0.6ϵ	NA

All Langevin dynamics simulations were performed by LAMMPS[60] with a time step of 0.01 τ. The start and end temperatures were $T=1$, and the damping parameter was set to 100. Simulations were carried out for 20,000,000 steps. A contact cutoff of 5σ was used to analyze the time-dependent cluster counting and contact maps. Each analysis was based on ten independent simulations with different initial polymer conformations to mimic different cells. The first 2 million steps were treated as equilibration and not used for contact map calculation. Simulation screenshots were rendered in VMD[61] and Ovito.[62] In addition, polymer chains with anchor segments separated by 30 beads were investigated as a reference system. This system mimics small chromatin loops that can form easily under the same anchor-anchor-specific interaction strength.

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless stated otherwise within the text, statistical analysis was performed using R software. All data were assessed for Gaussian distribution using the Shapiro-Wilk normality test to decide whether to use parametric or non-parametric statistical measures. In this study, two-sided Wilcoxon rank-sum test, Student's t-test, and Chi-square test were used, p values of <0.05 were considered the cutoff for statistical significance and specific statistical details can be found in the figure legends. Error bars for all data represent the mean ± SEM.