Informatics studies on the sex chromosome evolution in genus *Tokudaia*, focusing on the X chromosome inactivation mechanism

X 染色体不活性化に焦点をあてたトゲネズミ属における性染色体進化の研究

A DISSERTATION

Submitted to the Graduate School of Life Science,

Hokkaido University

In partial fulfilment of the requirements for the degree

DOCTOR OF LIFE SCIENCE

Matiz-Ceron Luisa

September 2024

Contents

General Introduction

Sex determination is a process that occurs in individuals of a species with characteristics that allow them to be identified as males, females, or, in some cases, hermaphrodites. In animals, sex can be determined by environmental cues (such as temperature), or genetically; these systems are called environmental sex determination (ESD, especially temperature-dependent sex determination is called TSD) and genetic sex determination (GSD), respectively (Hake & O'Connor, 2008; Moore & Roberts, 2013). In the case of TSD, the temperature at which eggs are incubated determines the sex of the offspring. Higher or lower temperatures during incubation can result in different sexes. GSD is determined when the characteristics are present in sex chromosomes that direct gonad development towards distinct but reproductively compatible outcomes (Moore & Roberts, 2013). Inside genetic determination, individuals with two different sex chromosomes are of heterogametic sex, and they are thus able to produce two types of gametes. When both are the same, it is called homogametic (Brashear et al., 2021).

There are multiple examples of heterogametic-determined sex in animals, like the ZW and XY systems. ZW determination is seen in birds, reptiles, and some fish. Females have one Z and one W chromosome (ZW), while males have two Z chromosomes (ZZ). XY system is found in mammals and other groups. Males have one X and one Y chromosome (XY), while females have two X chromosomes (XX) (Hake & O'Connor, 2008). Since the XY system is the determined process in humans, it has been widely studied.

Despite originating from autosomes, the X and Y chromosomes have undergone differentiation through arrested recombination along the Y chromosome. This leads to reduced natural selection pressure and subsequent gene decay, as well as the accumulation of repeated sequences as transposable elements (TE) and tandem repeats (Cortez et al., 2014; Kejnovsky et al., 2009). The reduced selection and gene decay associated with X and Y chromosomes have resulted from the lack of recombination during meiosis, causing them to become heterochromatic frequently. As a result, heterochromatic sex chromosomes often contain high proportions of repeated sequences. Especially in non-recombining regions of the Y chromosome, repeated DNA sequences are accumulated, representing a dominant and early force forming the Y chromosome, probably before genes degenerate (Kejnovsky et al., 2009). The evolution and emergence of sex chromosomes have been closely linked to the dynamics of these repetitive elements and TE (Dechaud et al., 2019).

The sex determination process behind the XY mainly occurs due to unique genes in the Y chromosome. The mammalian Y-chromosomal testis-determining gene (*SRY*) induces male sex determination (Kashimada & Koopman, 2010). The presence of the *SRY* gene initiates a cascade of genetic and hormonal events leading to male differentiation, primarily by inducing the expression of SRY-box 9 (*SOX9*), another critical gene in the male development pathway. *SOX9* transcript is the master regulator of Sertoli cell differentiation and the male type of gonad development (Kozhukhar', 2012). In standard conditions, the absence of *SRY* (as in XX embryos), the default pathway, leads to the development of ovaries and female characteristics.

There are other processes and genes specific to X and Y chromosomes. These mechanisms are key components in the sex determination and dosage compensation in mammals. One example is the process of addressing the gene dosage imbalance on the X chromosome between females and males. In this case, mammalian cells have evolved a mechanism called X chromosome inactivation (XCI), wherein one of the X chromosomes in females is transcriptionally silenced (Loda et al., 2022; Payer & Lee, 2008; Wutz & Gribnau, 2007). This process ensures equitable X-linked gene expression levels in somatic cells (Lyon, 1971). Also, XCI involves silencing one of the two X chromosomes in each cell of a female's body. This process is not only necessary to realize the balance between females and males, but the modes of X chromosome regulation, especially mosaicism and skewing of XCI, influence the severity of diseases caused by X-linked mutations (Deng et al., 2014a). In humans, age-related X reactivation could further influence the diversity of phenotypes observed in diseases. For example, cancer and autoimmune diseases as neurological disorders have been known. That is prevalent in women, to the assumed deleterious effect of uncovering mutations (Spatz et al., 2004; Z. Sun et al., 2022).

To achieve the XCI, a complex interplay of multiple long non-coding RNAs (lncRNAs) and proteins occurs at a pivotal locus known as the X inactivation center (Xic). In mice and humans, the Xic encompasses an approximately 660-kb region (Chureau et al., 2002). Within the Xic, critical sequences govern the counting, selection, and initiation of silencing of the future inactive X (Xi) while maintaining gene expression from the active X (Xa) (Augui et al., 2011). Several genes reside in

the Xic, including the lncRNAs *Tsx*, *Xist*, *Tsix*, *Jpx*, and *Ftx*, and protein-coding genes *Cdx4*, *Chic1*, *Slc16a2*, and *Rlim*. There is a complex relation and interaction between these genes. The cornerstone of Xic regulation is the lncRNA X-inactive specific transcript, *Xist* (Hierholzer et al., 2022; J. Lee et al., 1999; Loda et al., 2022). *Xist* depends on a series of repeats to perform the early and late steps of XCI. *Xist* is the central transcript in a regulatory network orchestrated by two topologically associating domains (TADs) on the X chromosome, TAD-E, and TAD-D, each playing a vital role in governing the XCI process and playing a positive and negative effect respectively (van Bemmel et al., 2019).

A few exceptional mammalian species possess unique sex chromosomes that could mean differences in the conserved process, like the Y-bearing females, Y-loss between, among others (Saunders & Veyrunes, 2021). The genus *Tokudaia* (Muridae, Rodentia) stands out as a prime example of Y-loss (Fig. 1). This genus includes three species, each endemic to a single island in southernmost Japan, and is faced with significant conservation challenges (Yamada et al., 2010). The Tokunoshima spiny rat (*Tokudaia tokunoshimensis*, TTO, 2*n*=45, XO/XO) and Amami spiny rat (*Tokudaia osimensis*, TOS, 2*n*=25, XO/XO) lack one X chromosome and the Y chromosome. The *Sry* gene, not only determining for therian mammals, including marsupials and placental mammals, was also completely lost (Honda et al., 1977; Honda et al., 1978; Murata et al., 2010; Terao et al., 2022). Both males and females share a single X chromosome, and their genome sequences are nearly identical, with only a male-specific 34 kb duplication upstream of *Sox9* on chromosome 3 in TOS (Terao et al.,

2022), resulting in equal gene dosage of the X chromosome in the XO/XO species. By contrast, the Okinawa spiny rat (*Tokudaia muenninki*, TMU, 2n=44) exhibits the typical XX/XY sex chromosome system found in most mammals (Tsuchiya et al., 1989). However, the sex chromosomes have fused with a pair of autosomes, forming recently acquired regions called as the neo-X and neo-Y on the X and Y chromosomes, respectively. These neo-X and neo-Y regions display partial genetic differentiation, indicating recombination suppression between them (Murata et al., 2012).

Based on the above information, I hypothesized structural and expression changes in the sex chromosomes of species with unusual determination systems, like the loss of one X chromosome and the Y chromosome, even in regions usually conserved in Eutherian, like the Xic. To verify this hypothesis, I used the three species of genus *Tokudaia*, the Okinawa spiny rat (TMU; *Tokudaia muenninki*, XX/XY), the Tokunoshima spiny rat (TTO; *Tokudaia tokunoshimensis*, XO/XO), and the Amami spiny rat (TOS; *Tokudaia osimensis*, XO/XO). Using a bioinformatics approach, their reciprocal genome and RNA-Seq were used to analyze gene structure, order, expression, and repeats location of Xic in the X chromosome.

In Chapter I, I compared the gene location and order in Xic of the genus *Tokudaia* to test whether Xic was different between the XO/XO species and XX/XY species. I annotated the Xic region and the genes inside by using homology and RNA-Seq-based prediction. Intra and inter-genic regions were evaluated.

In Chapter II, I analysed the detailed structure and expression of defined nine genes (*Cdx4, Chic1, Tsx, Xist, Tsix, Jpx, Ftx, Rlim*, and *Slc16a2*, gene name is shown in Table 1) in Xic of *Tokudaia* compared with mice. Exons, introns, and coding sequence (CDS) were determined. For the key factor, *Xist*, repeats that indicate function were identified. To explore whether the Xic genes operate in the XO/XO species an expression analysis was performed.

In Chapter III, to understand whether TEs are the potential elements of mutation in Xic, I performed a repeat analysis on the X chromosome. I identified types of TEs in the X chromosome and regions with high accumulation of repeat sequences. The TEs located in Xic were identified. Rat and mouse chromosomes were compared to the genus *Tokudaia*.

Based on the results obtained through these experiments, I discussed the potential contribution of the loss of one X chromosome and the Y chromosome, the consequences of the degradation of the XCI system, and possible sources of these changes.

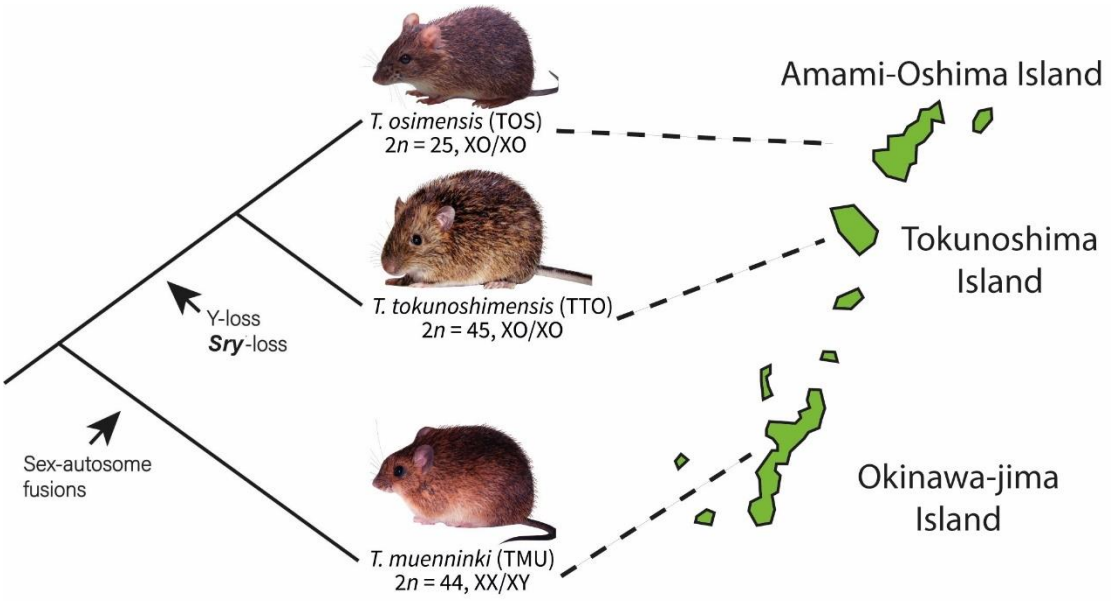

Fig. 1 Evolution of sex chromosomes and geographical distribution of the genus *Tokudaia*

A pair of autosomes fused to the X and Y chromosomes in the lineage of *Tokudaia muenninki*, resulting this species acquired neo-X and neo-Y. The Y-loss and *Sry*-loss caused by Y-to-X translocation occurred in common ancestral population of *Tokudaia osimensis* and *Tokudaia tokunoshimensis*. In the two species, a single X chromosome is shared with male and female, namely dosage compensation is need not between sexes.

Chapter I

Comparison of gene location and order in the X inactivation center

1.1 Abstract

XCI is an essential mechanism for dosage compensation between females and males in mammals. Mammalian cells have evolved the mechanism of XCI to address the gene dosage imbalance on the X chromosome between females and males. In females, XCI is controlled by a complex, conserved locus termed the X inactivation center, in which the lncRNA *Xist* is the key regulator. However, little is known about the Xic in species with unusual sex chromosomes. The genus *Tokudaia* includes three rodent species endemic to Japan. *Tokudaia osimensis* (TOS) and *Tokudaia tokunoshimensis* (TTO) lost the Y chromosome (XO/XO), while *Tokudaia muenninki* (TMU) acquired a neo-X region by fusion of the X chromosome and an autosome (XX/XY). I compared the gene location and structure in the Xic among *Tokudaia* species and mouse to predict order Xic loci. Compared to mouse, the Xic gene order and location were conserved in *Tokudaia* species. However, remarkable structure changes were observed in lncRNA genes, *Xist* and *Tsix*, in the XO/XO species. Additionally, three deletions and one inversion were confirmed in the intergenic region between *Jpx* and *Ftx* in the XO/XO species. I hypothesized that the significant structure change in *Xist, Tsix* and the intergenic region of *Jpx-Ftx* resulted in the inability to perform the X chromosome inactivation.

1.2 Introduction

Mammalian cells have evolved the mechanism of XCI to address the gene dosage imbalance on the X chromosome between females and males. In this process, one of the X chromosomes in every cell of females is transcriptionally silenced early in female development, leading to mosaic expression of either the maternal (Xm) or paternal (Xp) X chromosome (Loda et al., 2022; Panning, 2008; Payer & Lee, 2008; Wutz & Gribnau, 2007). This process ensures equitable expression levels in the X-linked gene of somatic cells (Lyon, 1971). The XCI mechanism is governed by the X chromosome-to-autosome ratio (X:A), and one of the X chromosomes is inactivated to achieve a balanced ratio (Monkhorst et al., 2009). This process is orchestrated primarily by the *Xist* gene (Hierholzer et al., 2022; J. Lee et al., 1999; Loda et al., 2022; Wutz & Gribnau, 2007). A complex interplay of *Xist*, multiple lncRNAs and proteins occurs on the pivotal Xic locus. In mice and humans, the Xic encompasses an approximately 660-kb region and located in Xq13 (Chureau et al., 2002). Within the Xic, critical sequences govern the counting, selection, and initiation of silencing of the future Xi, while maintaining gene expression from the Xa (Augui et al., 2011). Several genes reside in the Xic, including the lncRNAs genes *Tsx*, *Tsix*, *Jpx*, and *Ftx*, and protein-coding genes *Cdx4*, *Chic1*, *Slc16a2*, and *Rlim*.

In the early stages of embryonic development, one of the two X chromosomes in each female cell is randomly selected for inactivation, creating the Xi. The one that remains active is the Xa. The *Xist* gene on the chosen chromosome is expressed, remains within the nucleus, and spreads across the length of the X chromosome from

which it is transcribed (Pollex & Heard, 2019). This coating of the chromosome is essential for silencing gene expression. The recruits of various proteins and other factors modify the chromatin structure, such as histone deacetylases and methyltransferases (Panning, 2008). These modifications result in the chromatin condensation, forming a heterochromatic state in the Xi. The heterochromatin formation caused the silencing of most genes on the Xi. Some genes may escape inactivation and continue to be expressed. These escaping genes contribute to the functional differences between males and females (Berletch et al., 2011). This state is maintained throughout the life cycle and passed on during cell division (Loda et al., 2022).

XCI regulatory network is orchestrated by two TADs on the X chromosome, TAD-E, and TAD-D, each playing a vital role in governing the XCI process. TAD-E is positioned at the chromosome end, while TAD-D is located closer to the centromere (van Bemmel et al., 2019). TAD-E exerts a positive effect, along with *Ftx*, *Jpx*, *Slc16a2*, and *Rlim*, on *Xist* production during early development (Chureau et al., 2010; Tian et al., 2010). By contrast, TAD-D, which includes *Chic1*, *Cdx4*, *Tsx* and *Tsix* participates in negative regulation (Augui et al., 2011). In this way, the regulatory landscape of *Xist*, and presumably the Xic, probably spans at least 800 kb comprising these two TADs. The mechanism of insulation between the two TADs remains unclear.

The mechanism of XCI is typically conserved among mammals. However, the genus *Tokudaia* might be an exception because of its distinctive sex chromosome

system. Our lab previously sequenced, localized and assessed *Xist* RNA expression in two species with distinct sex chromosome systems, TOS (XO/XO) and TMU (XX/XY) (Zushi et al., 2017). In TOS, loss-of-function mutations was observed in the *Xist* gene. Notably, RT-PCR and northern blotting analyses confirmed the absence of *Xist* RNA expression in both male and female TOS cells. By contrast, TMU exhibited female-specific *Xist* RNA expression, with localization predominantly to the long arm of the original X chromosome and partial localization to the neo-X region of Xi (Kudo et al., 2023). Although the neo-X region in TMU does not exhibit heterochromatinization and does not contribute to Barr body formation, it forms a slightly condensed structure. Other genes essential for XCI within the genus *Tokudaia* have not been characterized.

In this chapter, I first compared the X chromosome sequences of the *Tokudaia* genus with those of the mouse to identify the Xic loci. Using the mouse as a reference, a homology-based approach was complemented by RNA-Seq results for the genes present in these loci. I established identity, gene order, and phylogenetic relationships between TMU, TTO, TOS, and mouse.

1.3 Materials and Methods

The high-quality genome assemblies and gene annotations of the genomes of three *Tokudaia* species were previously reported (Okuno et al., 2023), however, these annotations do not include ncRNA genes. Therefore, I predicted here the structure of nine genes in Xic. I firstly predicted gene structure for TMU using the following methods: 1) an RNA-Seq-based method that predicts gene structures based on transcriptome sequencing results, 2) a homology method that predicts genes structures on the basis of gene sequences of related species, mouse. The two prediction methods are described below.

1) RNA-Seq-based method

Gene prediction was performed using both mapping and de novo methods using RNA-Seq data (TMU, DRR059296-302; TTO, DRR496898, DRR496899; and TOS, DRR426688, DRR426689) and genome sequence data (Okuno et al., 2023)(TMU, PRJDB16411; TTO, PRJDB16412; and TOS, PRJDB16413). In the mapping method, HISAT2 v2.2.1 (Kim et al., 2015) was used to map RNA-Seq reads to genome sequences. The mapped reads were assembled into gtf files using StringTie v2.2.1 (Shumate et al., 2022). In de novo method, RNA-Seq reads were used to assemble into transcripts by Trinity v2.9.1 (Grabherr et al., 2011). Redundant sequences were removed from the assembled contigs using CD-Hit v4.8.1 (W. Li & Godzik, 2006)

and the contigs were aligned to the genome sequence using GMAP v2023.10.10 (Wu & Watanabe, 2005).

2) Homology-based method

The mouse sequence of each gene (*Cdx4*, *Chic1*, *Tsx*, *Tsix*, *Xist*, *Jpx*, *Ftx*, *Slc16a2* and *Rlim*) was used as the query of GMAP v2023.10.10 with cross-species parameter to annotate the TMU genome. Accession numbers of mouse genes were shown in Table 1. The gene structure results predicted in TMU were used for prediction of TTO and TOS genes according to the same methodology.

The prediction results of the two methods were manually curated. For the gene predictions of TTO and TOS, the predicted genes of TMU were aligned to the TTO and TOS genomes by GMAP, partially corrected based on the RNA-Seq mapping results.

Comparison of Xic between mouse and *Tokudaia* species

To generate the synteny plot of Xic loci, the lylaplot (https://github.com/mokuno3430/emu/tree/main/lyla_plot) tool was used. To run this tool, sequence alignment and comparison were performed by BLAST (v2.9.0) (Altschul et al., 1990). The mouse Xic genome sequence of mm39 X chromosome

(from 5 kb upstream *Cdx4* to 5 kb downstream *Rlim*, 102,360,005-103,029,892) was used for comparison with *Tokudaia* Xic.

Dot-plot alignment for pairwise comparison of Xic were generated using the nucmer command in MUMmer (V.3.23)(Delcher, 2002). All output files generated were transformed into matching coordinates by using the mummerplot standard command. The matching coordinate files were plotted using gnuplot (V.5.9)(Vaught, 1996).

1.4 Results

The gene content and locations in the Xic are largely conserved in *Tokudaia* species I compared the gene content and order in the Xic between *Tokudaia* species and mouse. In mouse, *Cdx4*, *Chic1*, *Tsx*, *Tsix*, *Xist*, *Jpx*, *Ftx*, *Rlim*, and *Slc16a2* were present in the 660-kb Xic (102,365,005–103,024,892 bp from *Cdx4* to *Rlim*) on the X chromosome. All these genes were annotated to the X chromosome in all *Tokudaia* species, and the gene order was conserved (Fig. 2). The estimated lengths of the Xic were 636, 582, and 595 kb in TMU, TTO, and TOS, respectively. Sequence identities between species are shown in Table 2. The sequence identity of *Tokudaia* species and mice was >86%; TTO and TOS showed the highest sequence identity of 99.12%.

Structural changes in intergenic sequence in the XO/XO species

To visualize detailed Xic structure differences between species, a dot plot alignment of all loci in mouse and each *Tokudaia* species was generated (Fig. 3). The dot plot illustrates the high conservation of genomic structure between mouse and TMU in the Xic (Fig. 3a). By contrast, deletions and inversions were identified in TTO and TOS (Fig. 3b, c). I focused on the two regions indicated by gray color in Fig. 3b and c, and the enlarged images were shown in Fig. 3d-g. An insertion (6.9 kb) was observed in TOS *Tsx* (Fig. 3f). Deletions were confirmed in *Xist* and *Tsix* (Fig. 3d and f). Additionally, three deletions and one inversion were confirmed in the intergenic region between *Jpx* and *Ftx* (Fig. 3e, g). The size of each deletion was 10.2 kb, 1.8 kb,

and 10.9 kb, and the size of inversion was 11.4 kb. The location and size of deletions and inversion were mostly identical in TTO and TOS. None of these changes were present in TMU.

1.5 Discussion

I observed that the gene content and order in the Xic are conserved mainly between the genus *Tokudaia* and mice, however, with partial structural variation in TTO and TOS. Within *Tokudaia* species, TTO and TOS showed the highest sequence identity. A previous molecular phylogenetic analysis revealed that TMU (XX/XY) diverged first among the three species, and TOS and TTO (XO/XO) are phylogenetically closely related (Murata et al., 2010). Therefore, the patterns of sequence identity likely reflect the phylogenetic relationships within the genus.

Transcaucasian mole vole, *Ellobius lutescens* (2*n*=17), also has been known as the XO/XO species. The closely related species, *Ellobius talpinus*, has a 54, XX karyotype in both females and males and the XCI mechanism is most likely still intact (Just et al., 2007; Mulugeta et al., 2016). The previous report showed that the X remained monovalent without chiasma formation in female and male meiosis in *E. lutescens*, XO/XO (Matveevsky et al., 2017). The genome sequencing of Xic in *E. lutescens* indicated that all genes known to control XCI are still present, and the whole region showed >90% conservation between *E. lutescens* and *E. talpinus* (Mulugeta et al., 2016). This comparison, with my findings reveals a notable conservation of the Xic across both the genus *Tokudaia* and the genus *Ellobius* species despite their unique sex chromosome compositions. This conservation underscores the evolutionary stability of the Xic and XCI mechanisms across diverse mammalian lineages.

I confirmed deletions and one inversion in *Jpx–Ftx* in the XO/XO species, both in size and sequence. Mutations in this region likely originated in a common ancestor, raising the possibility that the loss of functional XCI occurred in the ancestral lineage of the XO/XO species, likely due to the loss of one X and the Y. This intergenic region of *Jpx–Ftx*, that within TAD-E, upregulates *Xist*. Notably, the loss of the broader *Jpx–Rlim* region reduces XCI (Barakat et al., 2011), emphasizing the significance of this region in mediating chromatin interactions in the Xic. The presence of TADs, including TAD-D and TAD-E, makes the interactions between genes and intergenic regions essential for initiating and maintaining XCI (van Bemmel et al., 2019). Successful XCI will create regions of heterochromatin and euchromatin, the latter harboring genes that escape XCI in the Xi (Fang et al., 2019; Loda et al., 2022). Evidence shows that *Ftx* serves as an intermediate in the cascade of events leading from *Rlim* to *Xist* activation (Chureau et al., 2010).

I hypothesized that both such structural changes would be more likely to occur because of having a single X chromosome and self-synapsis. In XO mice, an XO pachytene oocyte in which an unsynapsed univalent X is silenced shows pachytene arrest. Alternatively, an XO pachytene oocyte with a self-synapsed X chromosome remains active (Burgoyne et al., 2009). Self-synapsis potentially facilitates chromosomal rearrangements. Since it is impossible to obtain tissues from the endangered spiny rat for observations of synapsis, sequence analyses of whole X chromosomes are an essential approach to confirm my hypothesis.

1.6 Tables

Table 1 Accession number of NCBI RefSeq mouse genes.

Symbol	Name	Accession number
Cdx4	Caudal type homeobox 4	NM_007674
Chic1	Cysteine-rich hydrophobic domain 1	NM_009767
Tsx	Testis specific X-linked	NM_009440
Xist	X inactive specific transcript	NR_001463
Tsix	X (inactive)-specific transcript, opposite strand	NR_002844
Jpx	Just proximal *Xist*	NR_015508
Ftx	Five Prime To *Xist*	NR_028381
Slc16a2	Solute Carrier Family 16 Member 2	NM_009197
Rlim	Ring Finger Protein, LIM Domain Interacting	NM_011276
Tbp	TATA-Box Binding Protein	NM_013684

Table 2 Nucleotide sequence identity of the X inactivation center (Xic) of mouse, *Tokudaia muenninki* (TMU), *Tokudaia osimensis* (TOS), and *Tokudaia tokunoshimensis* (TTO).

Species	**Identity (%)**			
(Size of Xic)	**Mouse (660 kb)**	**TMU (636 kb)**	**TTO (582 kb)**	**TOS (595 kb)**
Mouse	-	86.62	86.66	86.75
TMU	-	-	96.11	96.01
TTO	-	-	-	99.12

1.7 Figures

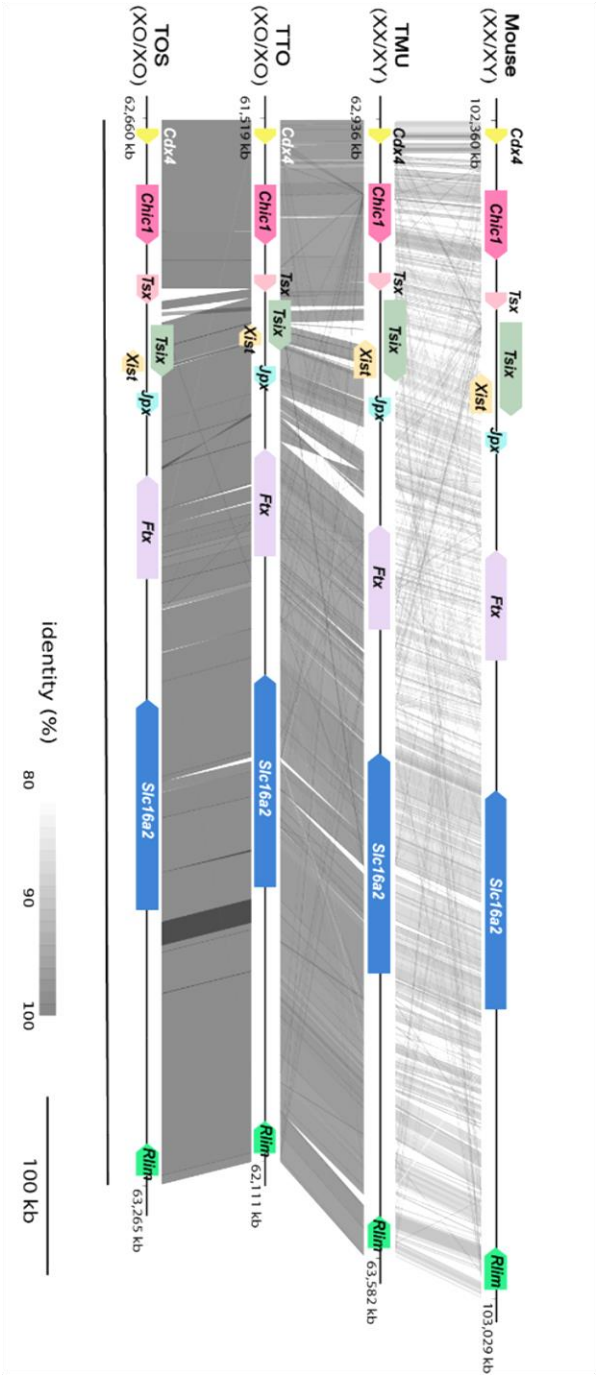

Fig. 2 Xic loci comparison

Gene location and sequence identity in the X inactivation center (Xic) of mouse, *Tokudaia muenninki* (TMU), *Tokudaia osimensis* (TOS), and *Tokudaia tokunoshimensis* (TTO). The lines on the left side mean phylogenetic relationship.

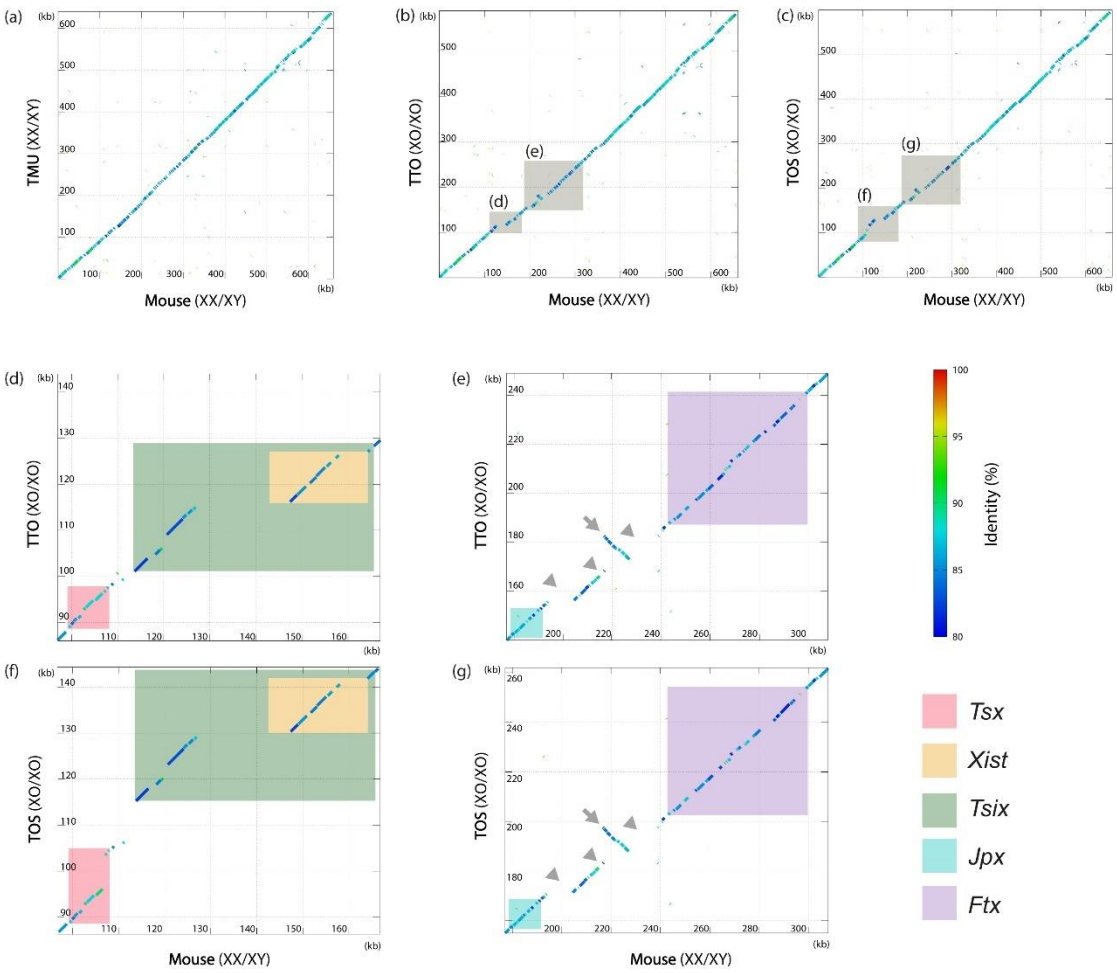

Fig. 3 Dot plot alignment of the X inactivation center (Xic)

Dot plot alignment between mouse and *Tokudaia muenninki* (TMU) (a), *Tokudaia tokunoshimensis* (TTO) (b), and *Tokudaia osimensis* (TOS)(c). Gray squares indicate the misaligned regions. (d, f) for *Tsx*-*Xist*/*Tsix*, and (e, g) for *Jpx*-*Ftx* are enlarged views of the regions indicated by gray squares in (b) and (c), respectively. Arrow heads and an arrow indicate deletions and an inversion in (e) and (g).

Chapter II

Structure and expression of genes in the X inactivation center

2.1 Abstract

To achieve XCI, a complex interplay of multiple lncRNAs and proteins occurs at a pivotal locus known as the Xic. Several genes reside in the Xic, including the lncRNAs *Tsx, Xist, Tsix, Jpx*, and *Ftx*, and protein-coding genes *Cdx4, Chic1, Slc16a2*, and *Rlim*. The cornerstone of XCI regulation is the lncRNA *Xist*. Mouse *Xist* has six conserved repeat sequences, A-, B-, C-, D-, E-, and F-repeats. The A-repeat had been identified at the most important one. The other repeats function as binding sites for chromatin regulatory complexes, loop maintenance, localization, and late stages of XCI. Previously my laboratory member sequenced the *Xist* gene and assessed *Xist* RNA expression and localization in two species with distinct sex chromosome systems, TOS (XO/XO) and TMU (XX/XY). In TOS, I observed loss-of-function mutations in the *Xist* gene. Other genes essential for XCI within the genus *Tokudaia* have not been characterized. I compared the gene structure and expression location in the nine genes (*Cdx4, Chic1, Tsx, Tsix, Xist, Jpx, Ftx, Slc16a2, Rlim*) characterized in the Xic loci among *Tokudaia* species. Compared to mouse, the Xic protein-coding genes were conserved in *Tokudaia* species. Some unique changes were found in genus *Tokudaia*, specifically in the XO/XO species. In *Xist*, important functional repeats, B-, C-, D-, and E-repeats, were partially or completely lost due to deletions in these species. RNA-seq data showed that female-specific expression patterns of *Xist* and *Tsix* were confirmed in TMU, however not in the XO/XO species. The RNA structure were calculated for every lncRNA.

2.2 Introduction

A complex relation between *Xist* and all the other genes in the Xic should be performed to produce a successful XCI. The Xic loci characterized the presence of nine genes, *Cdx4*, *Chic1*, *Tsx*, *Tsix/Xist*, *Jpx*, *Ftx*, *Slc16a2* and *Rlim,* in that order. In this group lncRNAs (*Xist*, *Tsix*, *Jpx*, and *Ftx*) and protein-coding genes (*Cdx4*, *Chic1*, *Rlim*, and *Slc16a2*) interplay from early stages to produce the XIC. The strategies utilized to achieve X-chromosome dosage compensation differ between different mammals (Deakin et al., 2009). In the mouse, the best studied eutherian mammal, XCI occurs in two waves during development. Both require the non-coding transcript *Xist*, which is present only in eutherians. First, shortly after fertilization, Xp undergoes inactivation. Inactivity of the Xp is maintained in extraembryonic tissues, but it is reactivated at the blastocyst stage in cells that give rise to the embryo, where random XCI (rXCI) occurs (Loda et al., 2022). In XX ES cells, two active X chromosomes stabilize the pluripotency gene expression network, and ectopic *Xist* expression leads to faster cell differentiation and downregulation of pluripotency factors (Schulz et al., 2014).

How the TAD is contributing to imprinted and random inactivation has been studied recently. Some mechanisms related to the positive regulation from TAD-E, along with *Ftx*, *Jpx*, *Slc16a2*, and *Rlim*, on *Xist* production during early development have been described (Chureau et al., 2010; Tian et al., 2010). *Ftx* located on the five

primes side to *Xist* has been described as a *cis* transcript to promote *Xist* transcriptional activation and establishment of XCI (Furlan et al., 2018). This gene transcript occupies distinct nucleoplasm loci, with multiple isoforms through a remarkable complexity of alternative promoter usage and splicing patterns (Chureau et al., 2010). *Jpx* presents just proximal to *Xist* and is probably the most well-study transcript after *Tsix*. It has been recognized for its role in the initiation, recruitment of chromatin modifiers, stabilization, and regulation of the spreading of *Xist* (Carmona et al., 2018; S. Sun et al., 2013; Tian et al., 2010). On the other hand, the protein-coding genes *Slc16a2* and *Rlim* ensure the rapid and stringent inhibition of gene expression in the future Xi (Barakat et al., 2014). A lack of 11-zinc finger protein or CCCTC-binding motifs occurring in a "tandem" (CTCF), suggests a euchromatin region between *Slc16a2* and *Rlim* (Berletch et al., 2015). Remarkedly, all the genes in this TAD have been reported as escaped genes in both or one of the X chromosomes on females (Berletch et al., 2015; Katsir & Linial, 2019).

On the other hand, in the negatively regulated TAD-D, which includes *Cdx4*, *Chic1*, *Tsx*, and *Tsix*, many relations have been established. In the case of *Cdx4*, it is highly syntenic in placental mammals. Just in the marsupial opossum, the conserved element lies on chromosome 2, while *Cdx4* and *Rsx* (the marsupial equivalent to *Xist*) lie on the X chromosome (Galupa et al., 2020). In the case of *Chic1*, the interactions with *Xite*/*Tsix* shape the structure of the opposite TAD-E by favoring conformations of the chromatin fiber wherein they mutually colocalize (Giorgetti et al., 2014). Contacts between *Chic1* and *Xite* are also observed. Each of these harbors a set of CTCF sites

involved in mediating the observed physical contacts within each locus; most CTCF motifs present the same orientation (Galupa et al., 2022). *Tsx* could produce protein in mouse testis (Cunningham et al., 1998); however, in meiotic germ cells, embryonic stem cells, and brain is non-coding and a long transcript (Anguera et al., 2011). *Tsix* which is the antisense transcription across *Xist*, is thought to be important in the regulation of *Xist* expression and in mechanisms of choice in XIC (Avner & Heard, 2001). *Tsix* lncRNA does not affect primary choice during XCI. This lncRNA prevents ectopic inactivation of the Xa during epiblast differentiation (Gayen et al., 2015). Inside the TAD-D, the *Tsx* and *Tsix* safeguard the future Xa from inactivation by blocking *Xist* coiling on its structure (Anguera et al., 2011; J. T. Lee, 2000). Meanwhile, *Chic1* and *Cdx4* contribute to the stabilization of the chromatin landscape, which is essential for the accurate silencing of the X chromosome (Hwang et al., 2015). A specific CTCF-binding site, located inside *Tsix*, was suggested to have a role (Kung et al., 2015; Navarro et al., 2006) as its deletion was reported to lead to *Xis*t misregulation and female-specific defect in mouse ES cells. Nevertheless, the deletion of the 58 kb spanning region is significant, as it abolishes isolation and produces aberrant gene expression in both TADs (Nora et al., 2012).

Finally, *Xist* lncRNA is indispensable for XCI. *Xist* initiates the process early during development by spreading in *cis* across the X chromosome from which it is transcribed, the future Xi. The *Xist* gene has been localized to a 450-kb in mouse X chromosome, expressed exclusively from the Xi (Herzing et al., 1997). During XCI, *Xist* RNA triggers gene silencing, recruits a group of chromatin modifying factors,

and drives a major structural reorganization of the X chromosome (Loda & Heard, 2019). Some parts of *Xist* are conserved at the sequence level, suggestive of functional elements. However, there could be variation inside the introns. Some of these conserved regions are highly repetitive and are referred to A-F repeats (Pontier & Gribnau, 2011). Mouse *Xist* has six conserved repeat sequences, A-, B-, C-, D-, E-, and F-repeats, and two promoters, P1 and P2 (C. J. Brown et al., 1992; Maenner et al., 2010; Navarro et al., 2006; Nesterova et al., 2001; Pugacheva et al., 2005; Raposo et al., 2021; Sheardown et al., 1997). These repeats are distributed across exon 1 (A–D and F) and exon 7 (E). The pivotal role of *Xist* in XCI is attributed to the presence of these highly conserved repeats, particularly the A-repeat (Raposo et al., 2021). *Xist* lacking the A-repeat is expressed, but unable to form the characteristic Xi-associated structure due to the lack of initial coiling (Schoeftner et al., 2006). The A-repeat includes 7 to 8 copies of a conserved sequence known as the AUCG tetraloop (Duszczyk et al., 2011). The other repeats function as binding sites for chromatin regulatory complexes, loop maintenance, localization, and late stages of XCI (Lee, 2011; Nesterova et al., 2001; Sarma et al., 2010). The key factor of Xic and the other corresponding genes had been mostly studied in mice and humans; however, how these conserved loci could be modified when there is a unique system of determination is still unknown. The genus *Tokudaia* can be a suitable candidate.

In this chapter, I evaluated in detail the Xic-identified genes (*Cdx4*, *Chic1*, *Tsx*, *Tsix*, *Xist*, *Jpx*, *Ftx*, *Slc16a2*, and *Rlim*) differences between TMU, TTO and TOS by comparing with mouse sequences. First, the introns and exons were determined in

every case by homology and expression mapping. The CDS was determined to *Cdx4*, *Chic1*, *Slc16a2*, and *Rlim*. Also, I examined the *Xist* repeats to identify the regions with a possible degradation in the XO/XO species. I annotated the detailed structure of genes present in Xic loci for TMU, TTO and TOS compared with mouse.

2.3 Materials and methods

Detailed identification of introns, exons and CDS

The predicted structures in the Chapter I were examined from *Cdx4* to *Rlim*. First, the identified gene structure for TMU sequence by using RNA-Seq-based and homology were extracted. That method predicts gene structures based on transcriptome sequencing results, and homology method based on gene sequences of related species, mouse.

The prediction results of the two methods were manually curated. For the gene predictions of TTO and TOS, the predicted genes of TMU were aligned to the TTO and TOS genomes by GMAP, partially corrected based on the RNA-Seq mapping results. The CDS for protein-coding *Cdx4*, *Chic1*, *Slc16a2*, and *Rlim* were predicted based on the previous annotation of *Tokudaia* (Okuno et al., 2023) available in https://figshare.com/articles/dataset/Annotation_data_for_three_Tokudaia_species/24105600/1?file=42292524. To compare the predicted sequences an alignment was produced for every gene by ClustalW (Li, 2003). The secondary structures of the identified lncRNA were calculated by using RNAFold (by ViennaRNA Package v.2.0)(Lorenz et al., 2011). Due to the restrictions of the software available the exon 1 and 7 of *Xist* were calculated individually.

Xist repeats identification and comparison

Xist repeats of *Tokudaia* species were predicted by comparing the sequences of all repeats in mouse (Raposo et al., 2021) to the *Xist* sequence of *Tokudaia* species by

using BLAT (v.36) (Kent, 2002). A customized command was used to improve the sensitivity (blat -noHead -minIdentity=70 -stepSize=1 -tileSize=9).

Gene expression analysis

The expression level of genes in Xic were confirmed. The RNA-Seq reads were mapped to each species genome using STAR (v.2.7.9)(Dobin et al., 2013). Whole genome annotation results were used in which the Xic region was replaced with nine curated transcripts. Then, the TPM was calculated by using RSEM (V.1.3.3) (Li & Dewey, 2011) with default parameters for paired end reads. A bar plot of the TPM per gene was perform using RStudio (V.2012.15.2) and the package ggplot2. *Tbp* was used as an internal control, and its accession number of mouse was shown in Table 1.

2.4 Results

Structural changes in lncRNA genes

Owing to its importance as the master regulator of XCI, I performed a comparative sequence analysis of *Xist*. I newly determined the *Xist* sequence of TTO and updated the results of a previous study that our group reported the *Xist* sequences of TMU and TOS (Zushi et al., 2017). All repeat regions were conserved in TMU, however, with a partial deletion in B- and C-repeat (Fig. 4, Table 3). Although A- and F-repeats were confirmed in TTO and TOS, the D-repeat was completely lost, and C- and E-repeats were partially lost due to deletions in exons 1 and 7. A partial deletion in B-repeat was confirmed in TOS but completely lost in TTO. *Tsix*, which is the antisense lncRNA to *Xist*, showed significant structural changes in the two XO/XO species (Fig. 5a). The sequence including mouse exon 3 was absent in all *Tokudaia* species, and the sequence of exon 2, additionally lost in the XO/XO species, TTO and TOS.

By comparison of the gene structure between mouse and *Tokudaia* species, *Tokudaia*-specific changes were found in three lncRNA genes, *Tsx*, *Jpx*, and *Ftx* (Fig. 5b-d). Although the sequence of mouse *Tsx* exon 6 was present in all *Tokudaia* genome sequences, it was not transcribed (Fig. 5b). In *Jpx*, the sequence which includes mouse exon 2 was absent in the genome sequence of all *Tokudaia* species (Fig. 5c). The size of exon 3 was longer than that of mouse. Therefore, the total length of exons in *Tokudaia* species is about 800-1,000 bp longer than that of the mouse (Table 4). *Ftx* had the same number of exons as mouse but with *Tokudaia*-specific

features in several exons (Fig. 5d). The sequence of mouse exons 3, 5, and 10 were absent in *Tokudaia* species genomes. The *Tokudaia* exons 3, 4, and 5 sequences showed the homologies with the mouse exons 4, 6, and 7, respectively. *Tokudaia* exons 6, 7, and 10 were the genus-specific sequence. The exon 11 was longer than that of the mouse. Therefore, the total length of exons in the *Tokudaia* species was about 1,300 bp longer than that of the mouse (Table 4). While these changes were observed in lncRNA genes, I confirmed conservation in protein-coding genes, *Cdx4*, *Chic1*, *Slc16a2*, and *Rlim* (Fig. 6 and Table 4).

Expression levels of Xic genes were affected by the loss of one X and the Y in the XO/XO species

Since I detected notable structural changes in the two genes, *Xist* and *Tsix*, in the XO/XO species, I estimated TPM values using RNA-Seq data to evaluate expression patterns (Fig. 7). The expression levels of *Tbp* as an internal control mainly were same between sexes in all species (Fig. 7a), supporting the validity of this TPM calculation. TMU *Xist* showed a high female-specific expression, while TTO and TOS showed no expression in either sex (Fig. 7b). Our group previously reported that the *Xist* lncRNA was not expressed in TOS (Zushi et al., 2017). Therefore, my result supported the previous data. I have newly confirmed that the *Xist* lncRNA was not expressed even in TTO. *Tsix* expression was quite a low level in TMU (Fig. 7c). The reason for the extremely low expression level of *Tsix* was thought to be that this lncRNA is mainly transcribed in undifferentiated cells such as ES cells (Mise et al., 1999). I used differentiated somatic cells (fibroblasts) from each *Tokudaia* species for

RNA-Seq analysis. Although the expression level was quite low, TMU showed the female-specific expression. By contrast, different expression patterns were confirmed in the XO/XO species (Fig. 7c). Neither the XO/XO species showed a female-specific expression pattern, suggesting that *Tsix* expression is not regulated in a sex-dependent manner observed in general XX/XY species.

The expression of other genes in Xic loci (*Chic1, Cdx4, Tsx, Jpx, Ftx, Slc16a2* and *Rlim*) was examined (Fig. 8). There were no sexual differences in all genes of all species. There was not significant difference of expression level between species in *Chic1, Cdx4, Tsx, Jpx, Ftx,* and *Rlim* (Fig. 8a-e, g). *Cdx4* was not expressed in any species. The expression pattern of *Chic1* and *Ftx* was similar (Fig. 8a, e). The patterns were same between the XO/XO species. The expression pattern of *Rlim* was similar between species and sex. *Slc16a2* showed the same expression between TMU and TTO. By contrast, TOS showed a higher expression than the other two species. The *Tsx* expression was lower than the other gene expression (under 1.0 TPM value). Finally, *Jpx* had the highest expression in female TMU. *Jpx* expression pattern was similar in TOS and TTO.

The secondary structures were estimated for *Tsx, Tsix, Xist* exon 1 and 7, *Jpx* and *Ftx* (Fig. 9, Table 5). The structures of TOS and TTO were similar in all lncRNAs. In the case of *Tsx,* the structure was conserved in *Tokudaia* species (Fig 9a). In mouse and TMU the structures were conserved for *Tsix*, *Jpx,* and *Xist* exon 1 and 7 (Fig 9b, c, e, f). The free minimum energy of the estimated the lncRNAs secondary structures in *Tokudaia* was close values in *Tsx, Tsix, Jpx* and *Ftx* (Table 5). Nevertheless, TMU

Ftx structure was different compare with mouse and *Tokudaia* species (Fig. 9d). The *Xist* exon 1 and 7 in TOS and TTO were less negative than mouse and TMU (Fig. 9e, f).

2.5 Discussion

In this chapter, I detailed the structure of the Xic loci, focusing on nine genes: *Cdx4*, *Chic1*, *Tsx*, *Xist*, *Tsix*, *Jpx*, *Ftx*, *Slc16a2*, and *Rlim*. The protein-coding genes were confirmed to be conserved, even in the XO/XO species. However, significant differences were observed in the lncRNAs. The expression patterns of key Xic transcripts were evaluated, revealing that *Xist* has underwent changes in the XO/XO species due to the loss of one X and Y chromosomes.

Protein-coding genes showed a consistent size and high identity between *Tokudaia* species and mouse, whereas partial variation was discovered in the regions within lncRNA genes. The variation in the three lncRNA genes, *Tsx*, *Ftx*, and *Jpx*, represent a distinct mark of the *Tokudaia* species, and these variations are not related to the sex chromosome constitution.

Only in the XO/XO species remarkable structure changes were observed in *Xist* and *Tsix*, which are a key regulator of XIC and an integral part of its downregulation, respectively. All functional repeats were completely conserved in TMU, suggesting that the *Xist* RNAs of TMU are functionally conserved. This finding supports the results of our previous study of XCI in TMU (Kudo et al., 2023; Zushi et al., 2017). By contrast, TTO and TOS exhibited a partial loss of exons 1 and 7, including C-, D-, and E-repeats, and B-repeat is lost in TTO. Notably, the C-repeat recruits the polycomb group complex (PcG). PcG works by maintaining the silent state of the Xi through the action of B- and C-repeats (Bousard et al., 2019). In the same way, the E-

repeat is essential for anchoring *Xist* RNA molecules to the Xi territory by facilitating *cis*-localization (Ridings-Figueroa et al., 2017). The function of the D-repeat is not well-characterized; the region is more heavily degenerated in rodents than in other mammals, suggesting that the mouse D-repeat is not functionally relevant (Nesterova et al., 2001; Raposo et al., 2021). Based on these previous findings, the deletions in *Xist* in TTO and TOS could affect the recognition and maintenance of the future Xi. The *Xist* lncRNAs are not expressed in the XO/XO species. However, my findings indicate that even if they are expressed, they cannot produce a successful and lasting XCI in the XO/XO species. The deletions of *Tsix* in TOS and TTO could also cause XCI regulation failure.

The *Xist* lncRNA was not expressed in the XO/XO species; similarly, female-specific expression of *Tsix* was not observed in them. Since individuals with XX chromosomes have not been confirmed in TOS and TTO, XX is thought to be lethal in the early stages of development due to failure of XCI. I therefore hypothesized that the significant structure changes in the intergenic region of *Jpx-Ftx* resulted in the inability to perform XCI and, as a result, a lack of *Xist* expression. The expression was not significant for *Chic1, Tsx, Ftx, Jpx,* and *Rlim*. Moreover, the expression pattern for protein-coding genes were conserved between sexes in the different species. There was an exception of high expression for *Slc16a2* in TOS. *Slc16a2* is also responsible of intercellular transportation of thyroid hormones and brain development in mammals (Bernal et al., 2015; Wilpert et al., 2020). Since this high expression was found, it could be interesting to see if there are any abnormal

development of function related with it expression in TOS individuals. This value should be confirmed in the future with other samples, since not particular difference were found in *Slc16a2* sequence for TOS. *Jpx* in female TMU had the highest expression, accordingly with the necessity of proceed with the XCI. *Jpx* acts in cis, as a mature RNA by regulating the initiation and maintain of *Xist* transcription, through the eviction of CTCF from Xist 5' region (Barakat et al., 2011; Rosspopoff et al., 2023; S. Sun et al., 2013).

The secondary structures of lncRNAs between mouse and TTO and TOS were conserved in each gene. The free minimum energy values were close too. Thermodynamic models are commonly employed in RNA secondary structure prediction, as they quantify the stability of an RNA structure by calculating its folding free energy change and then select the structure with the lowest free energy or the maximum expected accuracy as the most probable (Gong et al., 2024). Topological change and decrease in thermodynamical energy were reflected in the XO/XO species at the exon 1 and 7 of *Xist* when compared with TMU and mouse. However, similar nearest-neighbor loops were found for *Tsx* and *Jpx* structure in *Tokudaia* species, supporting the changes described in the genomic sequences as unique for *Tokudaia* genus. The secondary structure of the genes in Xic had not been studied previously in mouse. Accurate predictions of RNA secondary structures can help uncover the roles of functional non-coding RNAs (Sato et al., 2021). For that reason, a detailed study in the future of the structures founded in processes that highly involve lncRNAs, as XCI, could bring light into the mechanism. In the case of *Tokudaia* species, I concluded

that the changes of secondary strcuture of lncRNAs could be related with the deletions/insertions identified in the gene sequence, as with the loss of the necessity of XCI in the XO/XO species.

To examine whether the deletions and inversion identified came from TE, I would like to confirm the presence of TE in the X chromosome and especially in Xic. My study marks the first documentation of the Xic structure in the XO/XO species. My findings provide a basis for investigating XCI and sex chromosome evolution and contribute to a better understanding of the mechanism underlying XCI.

2.6 Tables

Table 3 Length *Xist* repeats of the mouse, *Tokudaia muenninki* (TMU), *Tokudaia osimensis* (TOS), and *Tokudaia tokunoshimensis* (TTO).

Species	Repeats (bp)					
	A	F	B	C	D	E
Mouse	365	35	200	1,615	2,700	1,125
TMU	352	36	140	1,079	2,764	1,193
TTO	323	36	0	694	0	913
TOS	324	36	100	978	0	913

Table 4 Length of genes in Xic. *Tokudaia muenninki* (TMU), *Tokudaia* tokunoshimensis (TTO), and *Tokudaia* osimensis (TOS).

Species	Genes (bp)								
	Cdx4	***Chic1***	***Tsx***	***Xist***	***Tsix***	***Jpx***	***Ftx***	***Slc16a2***	***Rlim***
Mouse	1,592	7,369	712	17,911	4,193	3,657	4,261	4,151	7,435
TMU	1,588	7,284	710	17,239	4,141	4,691	5,525	4,191	7,513
TTO	1,585	7,243	706	7,303	3,696	4,489	5,511	4,173	7,492
TOS	1,585	7,240	706	8,071	3,688	4,493	5,515	4,193	7,466

Table 5 Minimum free energy (kcal/mol) prediction of lncRNA in Xic. Mouse, *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS).

Species	lncRNAs					
	Tsx	***Tsix***	***Xist* exon 1**	***Xist* exon 7**	***Jpx***	***Ftx***
Mouse	-196.9	-1158.9	-2658.67	-2052.5	-1132.5	-1279.39
TMU	-153.8	-1166	-2471.5	-2011.2	-1363.8	-1568.7
TTO	-158.8	-969.5	-960.8	-734.1	-1300.1	-1566.5
TOS	-175.1	-969.9	-969.6	-917.3	-1300.8	-1554.7

2.7 Figures

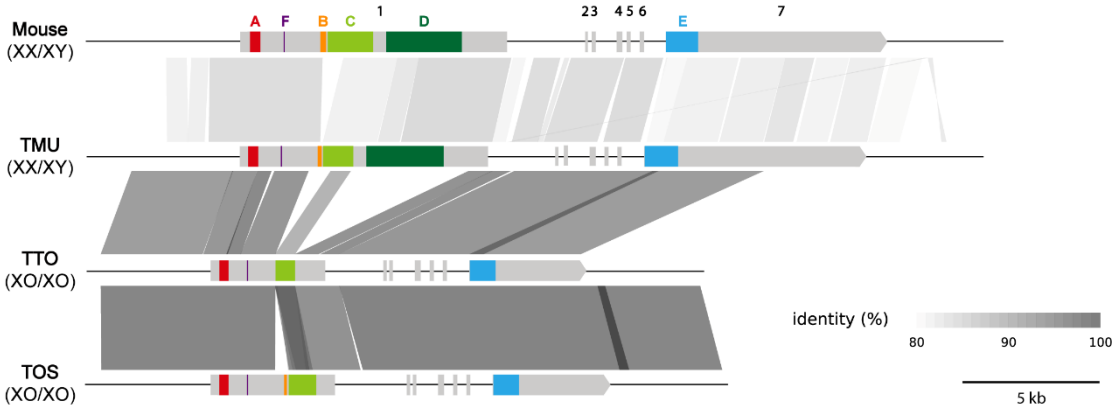

Fig. 4 Comparison of the gene structure of *Xist* in mouse, *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS).

Functional repeats of *Xist* are in highlighted in color, A (red), F (purple), B (orange), C (yellow green), D (green), and E (light blue). The number indicates exon number.

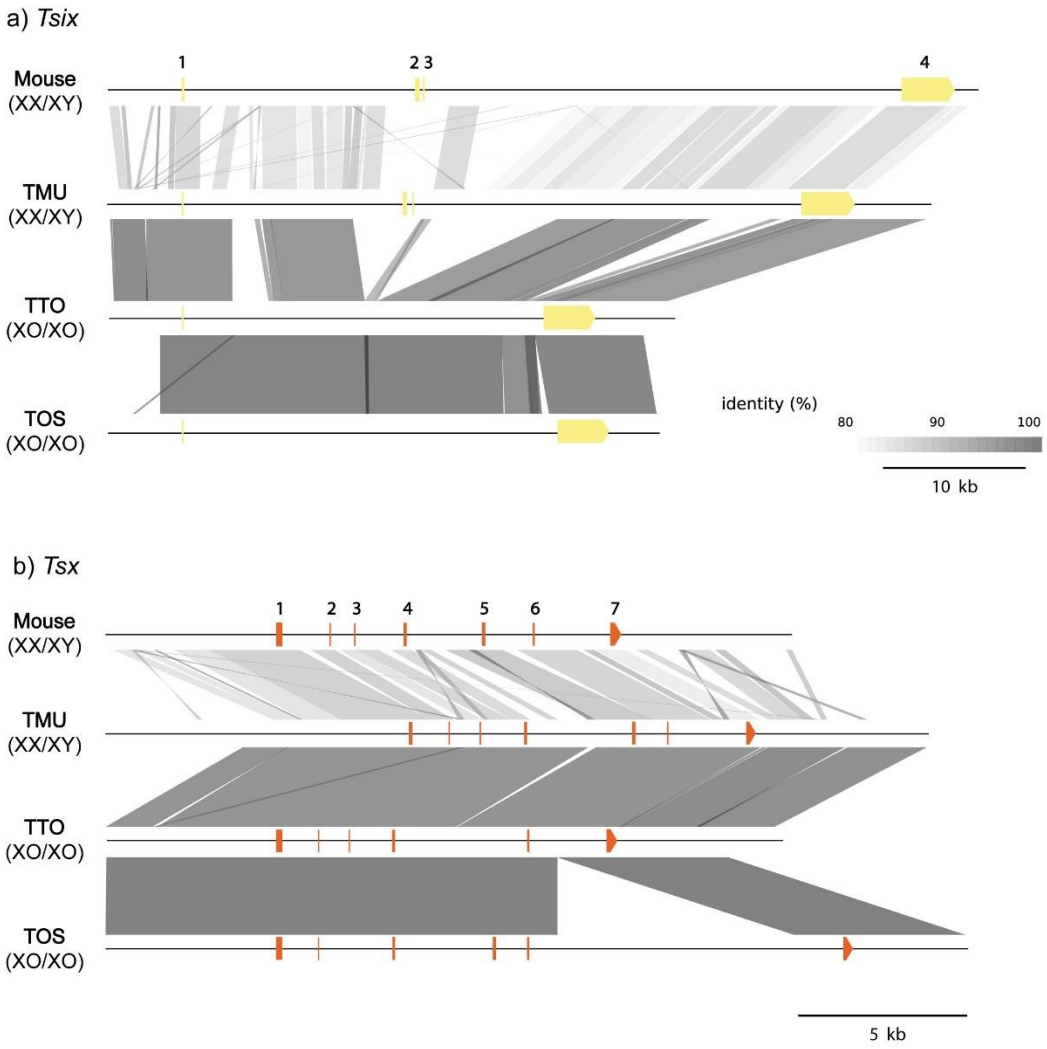

Fig. 5 Comparison of the structure of lncRNA genes.

Tsix (a), *Tsx* (b) in mouse, *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS). The number indicates exon number.

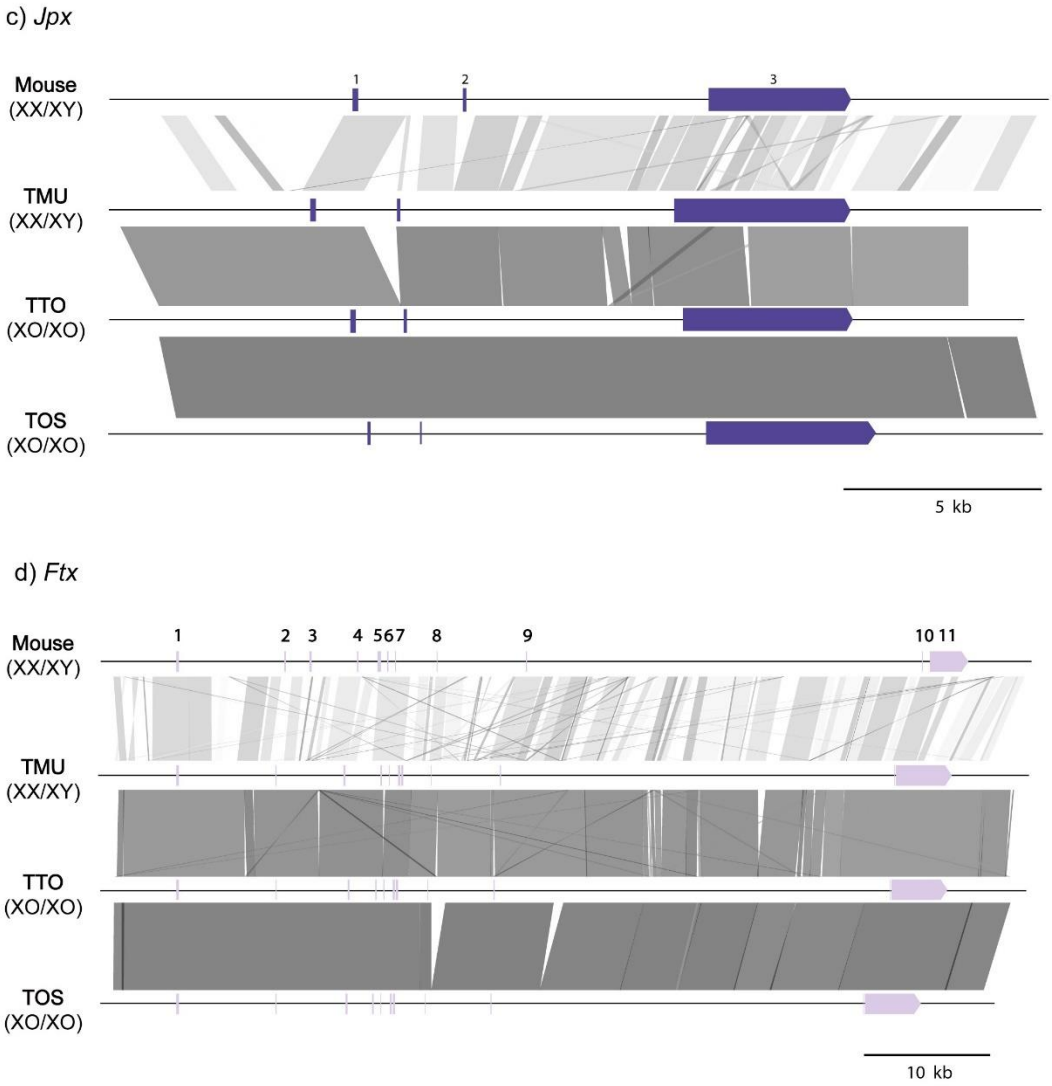

Fig. 5 (Continued) Comparison of the structure of lncRNA genes.

Jpx (c), and *Ftx* (d) in mouse, *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS). The number indicates the exon number.

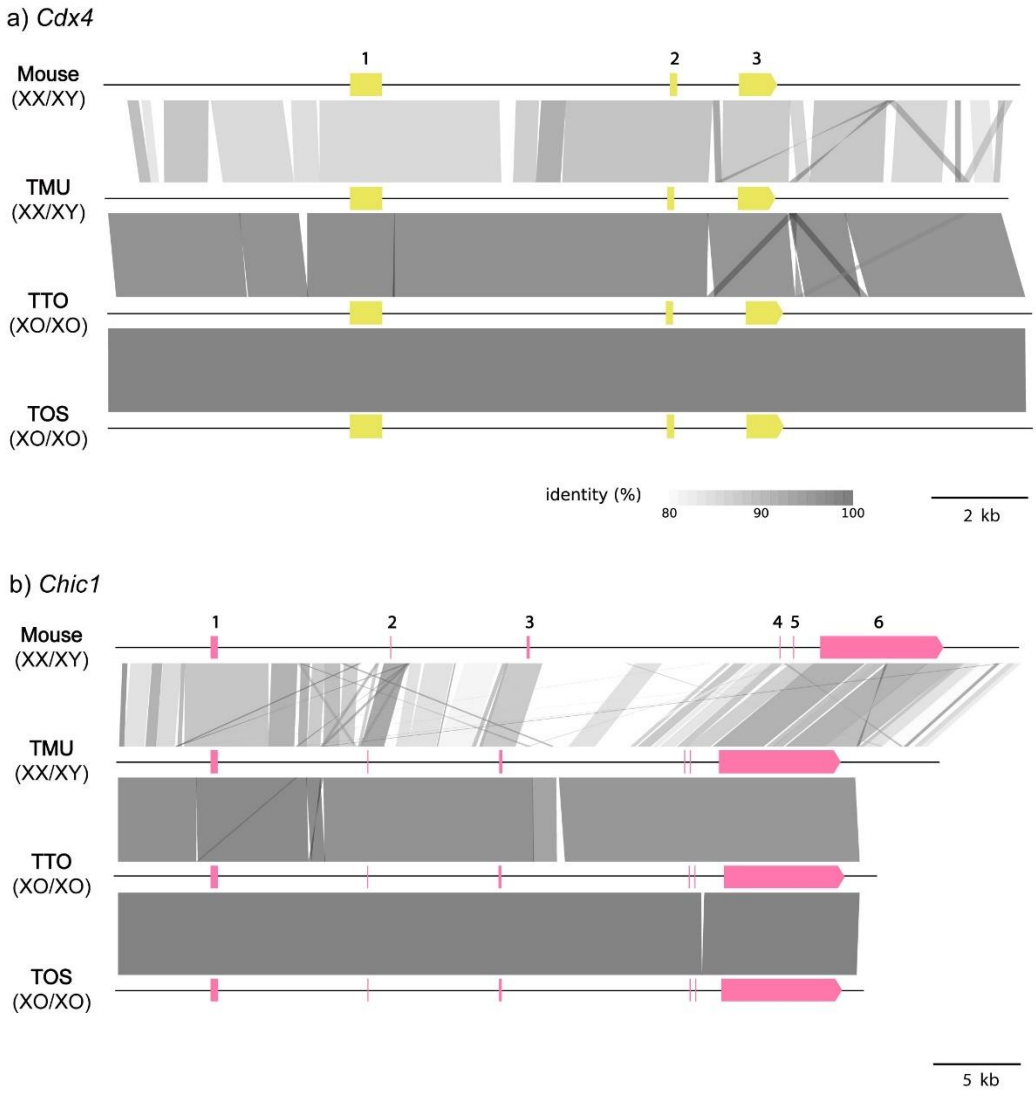

Fig. 6 Comparison of the structure of protein-coding genes.

Cdx4 (a), *Chic1* (b) in mouse, *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS). The number indicates exon number.

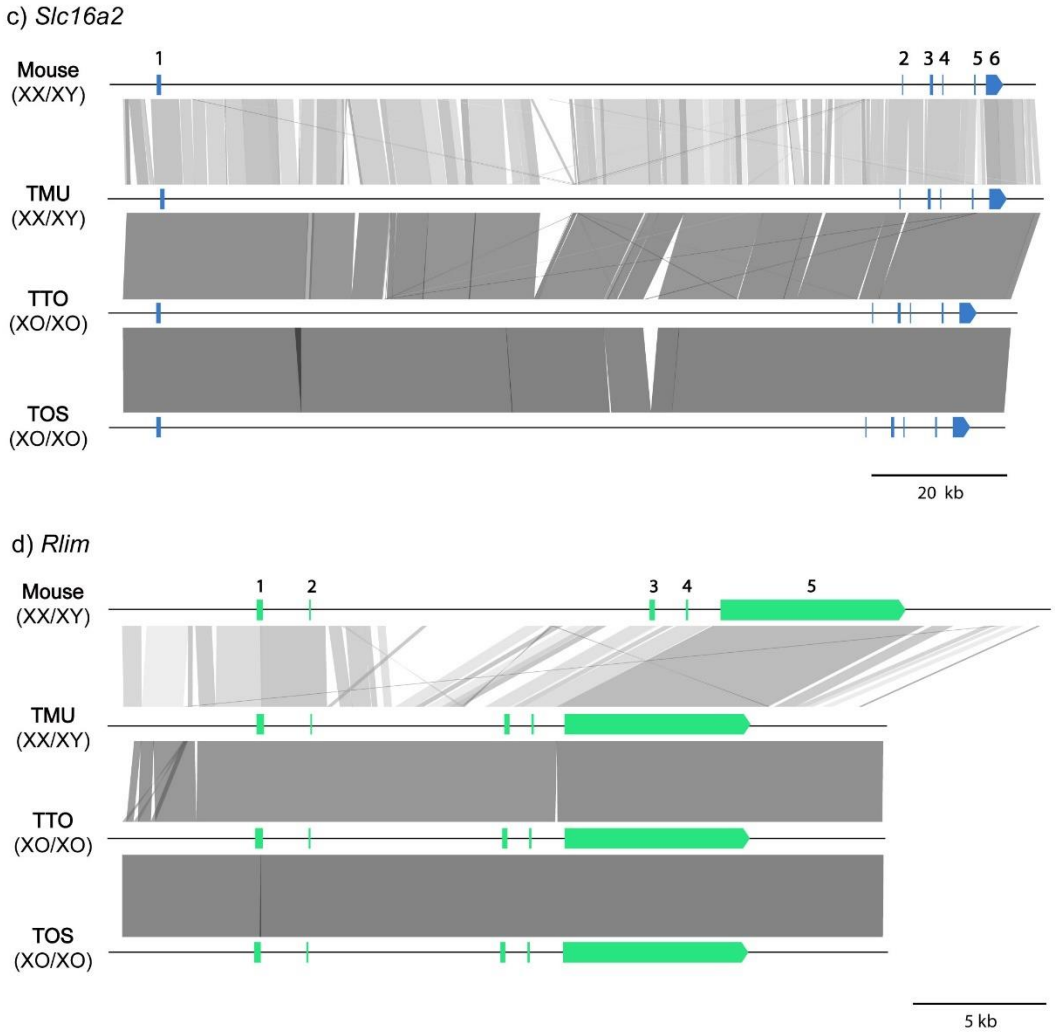

Fig. 6 (Continued) Comparison of the gene structure of protein-coding genes.

Slc16a2 (c), and *Rlim* (d) in mouse, *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS). The number indicates exon number.

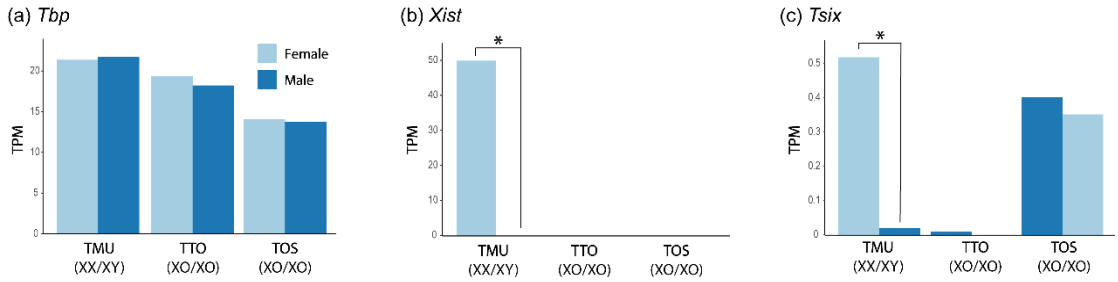

Fig. 7 Expression level of lncRNAs, *Xist* and *Tsix*

TPM value of *Tbp* (a) as internal control, *Xist* (b), and *Tsix* (c) in *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS). *$p < 0.05$.

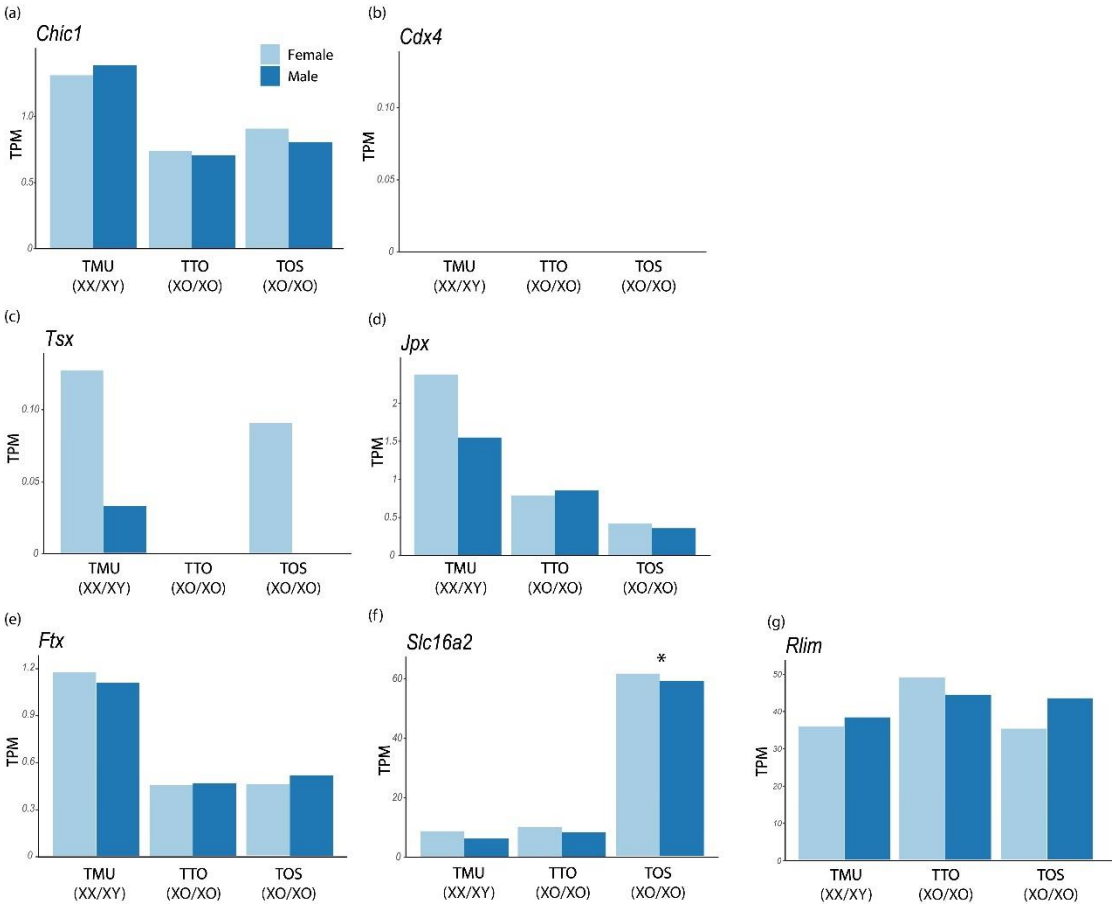

Fig. 8 Expression level of *Chic1, Cdx4, Tsx, Jpx, Ftx, Slc16a2* and *Rlim*

TPM value of *Chic1* (a), *Cdx4* (b), *Tsx* (c), *Jpx* (d), *Ftx* (e), *Slc16a2* (f) and *Rlim* (g) in *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS). *$p < 0.05$.

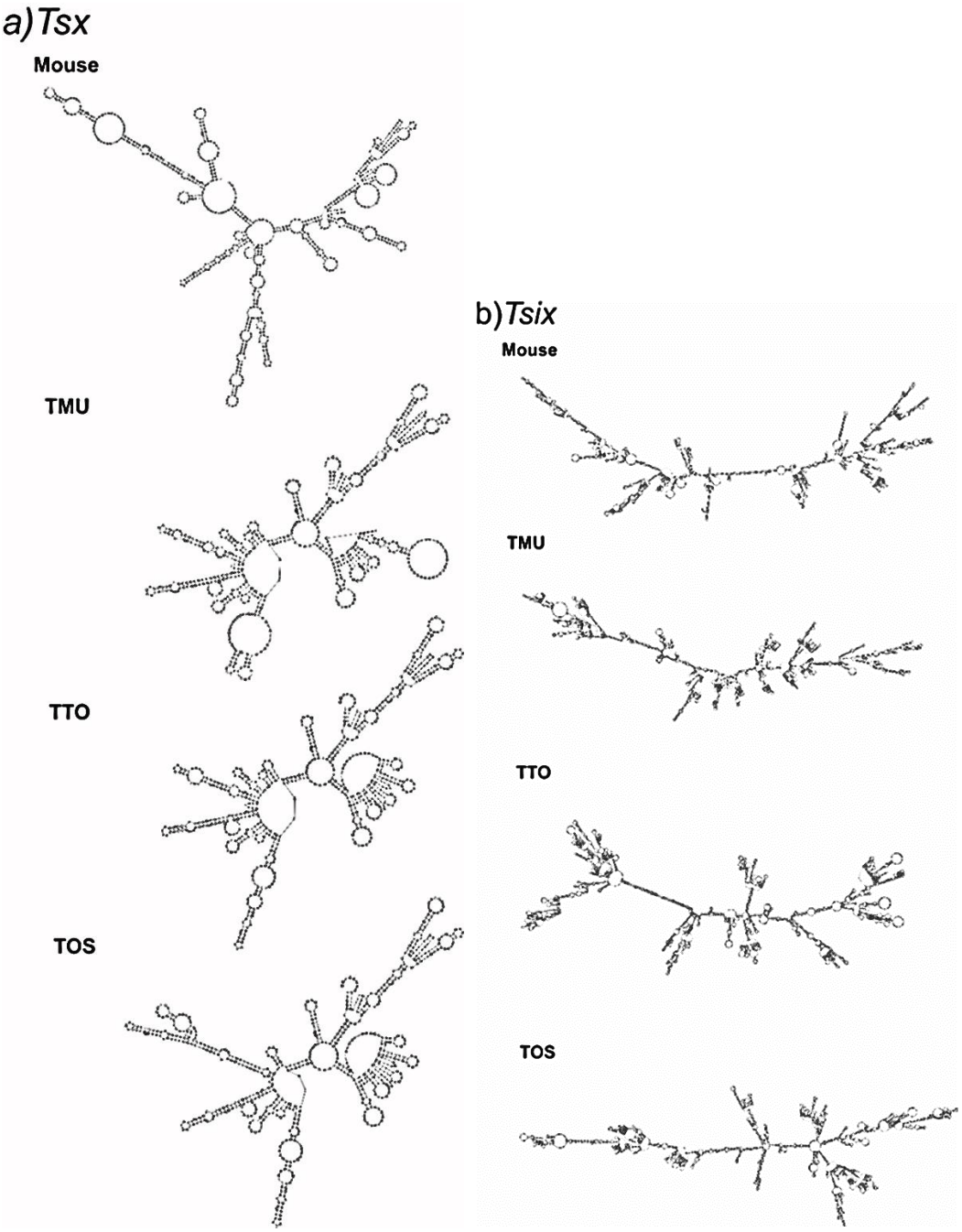

Fig. 9 Secondary structure of lncRNA of *Tsx* (a) *and Tsix* (b)

Mouse, *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS).

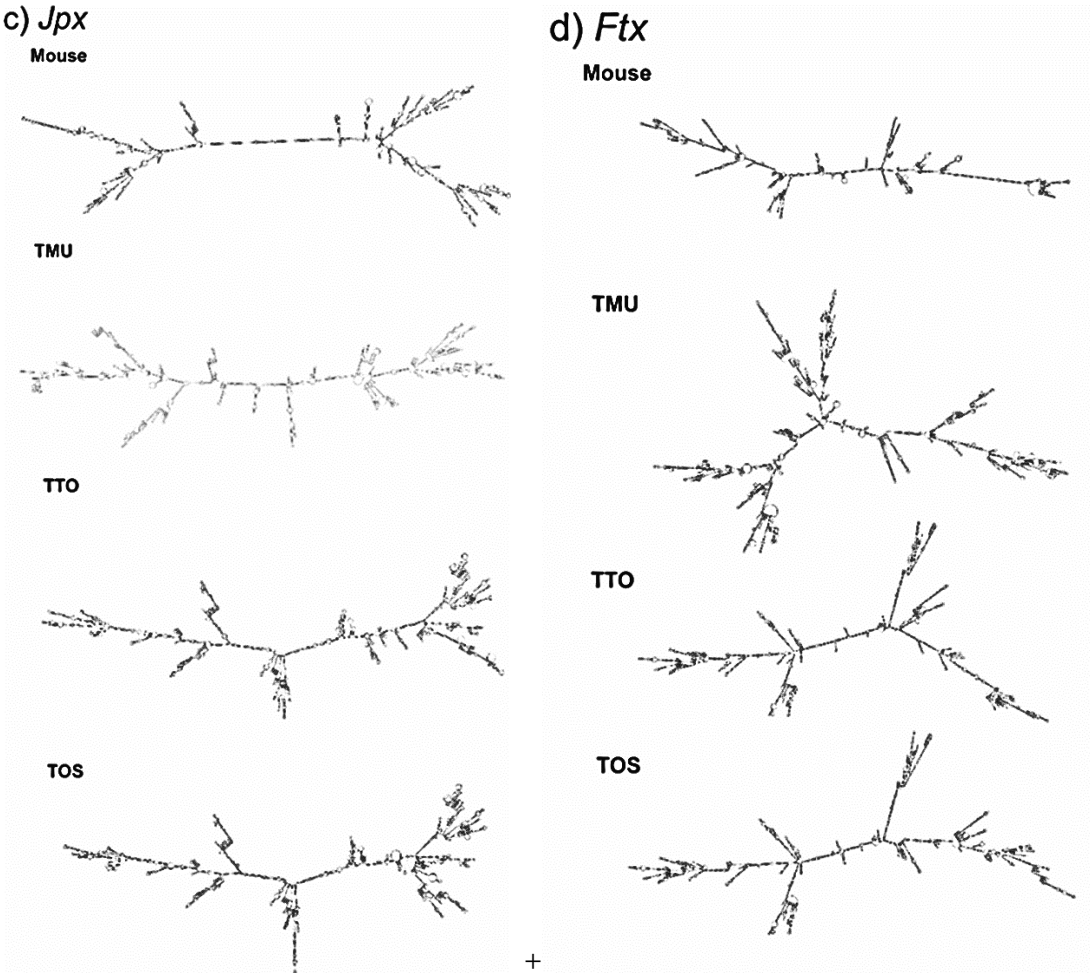

Fig. 9 (Continued) Secondary structure of lncRNA of *Jpx* (c) and *Ftx* (d)

Mouse, *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS).

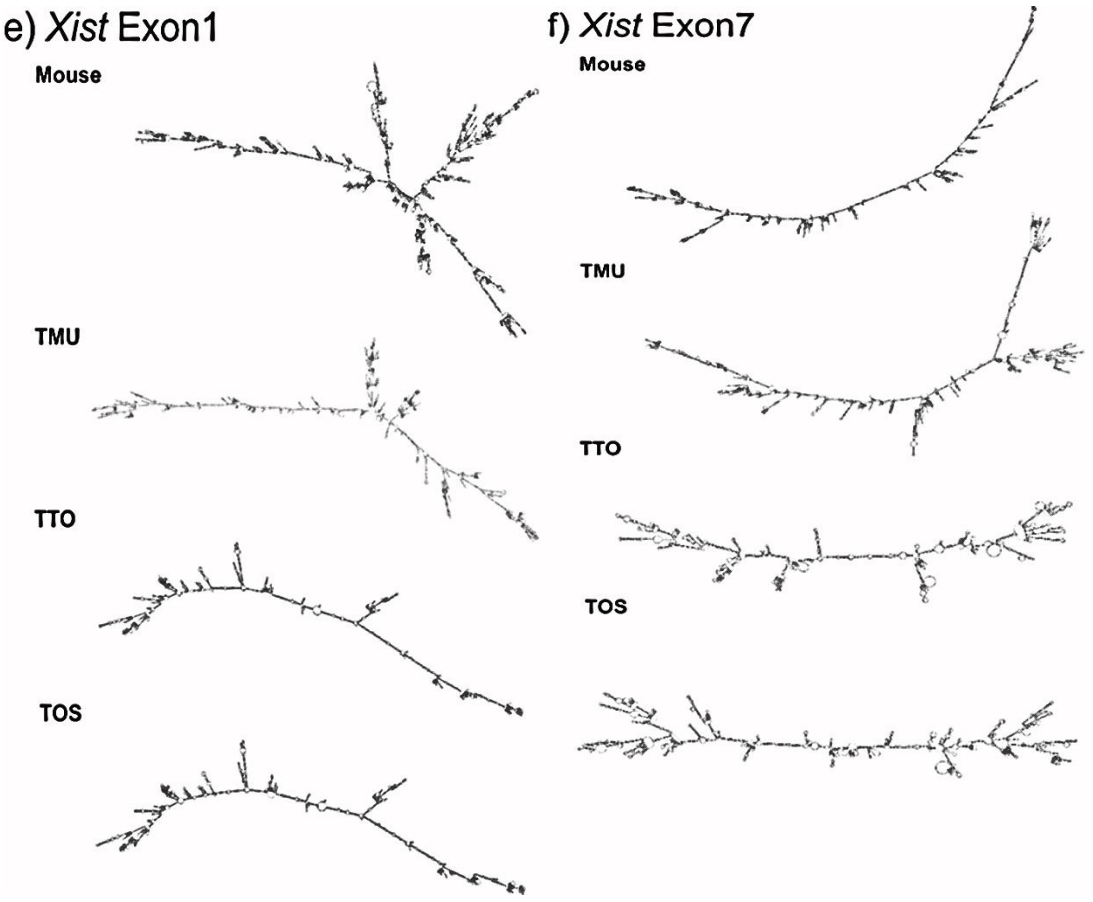

Fig. 9 (Continued) Secondary structure of lncRNA of *Xis*t exon 1 (e) and 7 (f),

Mouse, *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS).

Chapter III

A potential contribution of transposable elements in the X chromosome

3.1 Abstract

Transposable elements (TEs) are significant in genome evolution and species diversification, functioning by integrating and spreading within genomes through a "jumping" mechanism. TEs are classified into two major classes: Class I (retrotransposons) and Class II (DNA transposons). Each TE sequence can be categorized into a subfamily, superfamily, subclass, and class, sharing common genetic organization and monophyletic origin. The accumulation of TEs on sex chromosomes has driven their structural divergence from autosomes, influencing the unique evolutionary paths of the X and Y chromosomes. In the context of the X chromosome, TEs play a significant role in XCI. They help spread the silencing signal initiated by the *Xist* RNA, ensuring efficient inactivation of the entire X chromosome. Transposable elements are crucial in the evolution and divergence of sex chromosomes, contributing to genomic variability through mutations, gene disruptions, and alterations in gene regulation. I analyzed the distribution of TEs on the X chromosome of *Tokudaia* species and compared them with those of mice and rats. The research involved identifying the detailed distribution of TE classes, families, and subfamilies. It was found that Class I were prevalent on the X chromosome across the compared species. Unique TEs, such as the gypsy element, were identified specifically in *Tokudaia* species.

3.2 Introduction

TEs have emerged as significant players in the evolution of genomes and the diversification of species (Biémont & Vieira, 2006). These DNA sequences have the peculiarity of integrated -by "jumping"-into and spread within genomes. Generally, they are repetitive, with the possibility of recombining and inducing genome rearrangements (Dechaud et al., 2019). The estimated rate of TE insertions differs between organisms. Among individual fruit flies, 50–80% of mutations are due to such insertions. Meanwhile, in humans is around 0.1–1%, where the numerous copies of the long terminal repeats (LTRs) retrotransposons are less active and mostly fixed in position (Biémont & Vieira, 2006).

TEs can be divided into two major classes. Class I elements, also known as retrotransposons, mobilize through a "copy-and-paste" mechanism whereby an RNA intermediate is reverse-transcribed into a cDNA copy that is integrated elsewhere in the genome (Boeke et al., 1985; Bourque et al., 2018). Class I could be divided by the mechanism of chromosomal integration. The LTR retrotransposons correspond to sequences that used a cleavage and strand-transfer reaction catalyzed by an integrase, much like retroviruses (Brown et al., 1987). Meanwhile, non-LTR retrotransposons, long and short interspersed nuclear elements (LINEs and SINEs, respectively), are integrated by reverse transcription through a process referred to as target-primed reverse transcription (Luan et al., 1993). Class II, also known as DNA transposons,

are mobilized via a DNA intermediate directly through a “cut-and-paste” or by a “peel-and-paste” replicative mechanism involving a circular DNA intermediate (Bourque et al., 2018). Class I also has multiple subclasses, like Helitron or Maverick/Polintron. Each TE subclass is further divided into subgroups (or subfamilies) that are typically found across a wide range of organisms. They are sharing a common genetic organization and a monophyletic origin. In eukaryotes, Ty3/gypsy and Ty1/copia elements are two major subfamilies of LTR retrotransposons (Malik & Eickbush, 2001). Similarly, for DNA transposons Tc1/mariner, haT (hobo-Ac-Tam3), and MULEs (Mutator-like elements) are three superfamilies largely widespread (Feschotte & Pritham, 2007). Thus, in principle, every TE sequence in a genome can be affiliated with a subfamily, superfamily, subclass, and class (Bourque et al., 2018).

As their evolutionary success depends on their vertical transmission, transposable elements are intrinsically linked to reproduction. Sex chromosomes are often enriched in TEs. In organisms with sexual reproduction, transposable elements manifest their activity in germ cells or sexual chromosomes. For example, all TE types show a higher density on the Y chromosome compared to autosomes, except for SVA short retrotransposons (Dechaud et al., 2019). In the Y chromosome, the density is 30 times higher than the genome average for LTR elements and four times higher for Alu and L1 elements (Sellis et al., 2007). This accumulation might sometimes be the consequence of the impossibility for sex chromosomes to recombine and thus eliminate deleterious insertions.

The mammalian X chromosome is enriched in features favoring inactivation, part of the importance of dosage balance. Interspersed repeat elements, and in particular LINEs, have been suggested as the relevant enriching features to develop the regions of inactivation (Lyon, 2003). In the autosomes, LINEs are mainly concentrated in dark G-bands. However, in the mouse X chromosome, LINEs are distributed in both regions of positive and negative G-bands, demonstrated by fluorescence *in situ* hybridization (Boyle et al., 1990). The higher density of TEs on chromosomes might increase their probability of regulating some key genes and consequently impact sexual development (Dechaud et al., 2019). One process that relates inactivation with the X chromosome is the XCI, where *Xist* RNA coats and silences one of the two X chromosomes in female cells. Since there is poor knowledge about how XCI spreads across the chromosome, LINE-1 elements have been proposed to play a role in inactivation (Lyon, 2006). Chow et al (2010) proposed that genes become inactivated by the "move" in the *Xist* silenced region via a modification of the 3D conformation of the chromosome.

Conversely, LINE-poor regions are physically distant from *Xist* silenced regions. This hypothesis has been tested in the Amami spiny rat (TOS). Since no dosage compensation by X inactivation is required here, LINEs would not be required on this X chromosome region. However, a similar high concentration of LINEs on a single X chromosome of TOS was described compared to humans or mice (Scott et al., 2006). They are concluding that the accumulation of TEs on X chromosomes might be only a by-product of reduced recombination. However, the specific location of the LINEs

elements had not been determined nor the comparison with the other *Tokudaia* species (Scott et al., 2006).

Due to this information, I analyzed repeat sequences on the X chromosome in three *Tokudaia* species in Chapter III to examine the possibility that TEs are the potential elements of mutation in Xic. I identified the types of TEs in the X chromosome and regions with high accumulation of repeat sequences.

3.3 Materials and methods

Transposable elements identification

The genomes of *Tokudaia* previously used in Chapter I and II were run into RepeatModeler (V.2.0.4)(Flynn et al., 2020) and RepeatMasker (V. 4.1.5)(Smit et al., 2016). The genus *Tokudaia* X chromosome was compared with mice and rats with the respective id numbers NC_000086.8 and NC_086039.1. The X chromosome of TMU has two regions: the ancestral X chromosome and the neo-X, which originates from autosome. Therefore, I used here the genome information of the ancestral X chromosome. The size of X chromosome of each species used in this study is shown in Table 2.

RepeatModeler was used in the X chromosome of *Tokudaia*, moue and rat, adding the flag of -LTRStructur. The pan-genome annotations were established for every sequence. The elements successfully classified and those that remain as unclassified or unknown elements were separated and counted. The classified repetitions for every sequence were used as a library in Repeatmasker.

Repeatmasker was run with standard parameters. The output was summarized as tables and figures by using RStudio. Repeats in Xic were extracted and summarized. The repeats presented in Xic were compared with the other regions of the X chromosome.

Density calculation of transposable elements

Using Kernel density estimation (KDE), I estimated the probability density function (PDF) of a random variable based on observed data points. In this case, the positions of the repeats were based on the results of RepeatMasker. Location, class, and family were extracted for every point. Significance of the counts per species were calculated in each type of TE by using the Kruskal-Wallis test (since the distribution and number of samples were not parametric). A Gaussian distribution was used to calculate the distribution. This determines the shape, while the bandwidth parameter governs its width and thus the smoothness of the resulting estimate.

A kernel function involves placing at each data point and summing these kernels to form a smooth, continuous estimate of the PDF. The process involves evaluating the kernel function at numerous points across the variable's range and summing these contributions to compute the density estimate. This estimate represents the relative likelihood of observing different values of the variable. KDE is particularly useful for visualizing the underlying distribution of data, providing insights into its shape and characteristics without imposing assumptions about its parametric form (Kristan et al., 2011).

3.4 Results

LINEs are the most abundant repeats in the X chromosomes of *Tokudaia*

I performed a comparative analysis of TEs in the X chromosome of *Tokudaia*, focusing on the identification of TEs and their distribution on the X chromosome between TMU, TTO, TOS, mice, and rats. The ratio of annotated the TE sequence to each X chromosome is 46.82% (65,611,617 bp, TMU), 49.41% (69,684,013 bp, TTO), 48.28% (68,451,947 bp, TOS), 53.81% (91,194,617 bp, mice), and 49.3% (77,768,701 bp, rats).

Class I retrotransposons comprised 42.47-50.33% of the X chromosome in all species, indicating they are the highest bulk of the repeats (Fig. 10). Mostly consisted of sequences were L1 of LINEs (Table 6, Fig. 11). However, a small amount of L2 was identified in TMU and TOS (Table 6). Class I consisted of SINEs, LINEs, LTR family in all species. SINEs and LTR elements comprised about 3.73-5.57% and 5.82-7.46% of the X chromosome, except for mouse LTR, which was 9.88%.

The Class II DNA transposons, consisting of superfamily hAT and Tc, were the least represented (less than 1% in all species). Another element, Academ-1, was present in mouse and rat with a very few representations, however it was not identified in the *Tokudaia* species (Table 6, Fig. 12). Finally, the normalized number of counts between LINEs, SINEs, LTRs and DNA Transposon per species were not significant.

Unique TE subfamilies present in the X chromosome of *Tokudaia*

In the subfamilies of SINEs, Alu constituted the highest percentage in TMU (52.12%) and TTO (63.88%), meanwhile, it was B2 in TOS (43.57%) (Table 6, Fig. 13a). ID and MIR subfamilies were present in all *Tokudaia* species at a similar value (ID: 5.83-8.79, MIR: 0.49-0.92). tRNA and U type appeared in small quantity in the two the XO/XO species, TTO and TOS.

ERVL-MaLR showed the highest percentage in LTR subfamilies in all *Tokudaia* species (around 37%) (Table 6, Fig. 13b), followed by ERVK (around 36%), ERV1 and ERVL (around 9%). A small amount of Gypsy was also identified in all *Tokudaia* species.

Finally, in the subfamilies of DNA transposons, TcMar-Tigger and TcMar-Mariner showed similar percentages in all *Tokudaia* species (Table 6, Fig. 13c). The amount of haT-Charlie was almost same in TTO and TOS; however, it was 20% lower in TMU than in them. By contrast, haT-Tip100 was uniquely found in TMU.

A high LINEs density in the XO/XO species

The kernel density of the principal classes, LINEs, SINEs, LTRs and DNA transposons, were estimated in the *Tokudaia* species, mouse, and rat (Fig. 14). The estimation indicated that the class with the highest number of representations was

LINEs, followed by LTRs, SINEs, and DNA transposons, respectively, across all species. The density of LINEs was the higher in genus *Tokudaia*, specially for TOS around Xic. The density of DNA transposons was constant and the smallest in entire X chromosome distribution in all species. In SINEs, the density was also mostly constant and similar in all species with a peak sround the Xic for all the species. There was a similar pattern of distribution of LTRs in three *Tokudaia* species and rat. The significance was calculated by LINE, SINE, LTR, and DNA transposons in each species (based on the values of the Table 6). The p-value were higher than 0.05, indicating than the founded values were not significant.

3.5 Discussion

I observed that the TE content and density in the X chromosome are mostly conserved between *Tokudaia* species, mice, and rats, however, with partial structural variation. I identified 46.82-53.81% of repeats in the X chromosome, suggesting a total accumulation of repeats is the similar level between species. A previous cytogenetical analysis of our group reported that LINE-1 (L1) was present in all chromosomes of TOS (XO/XO) and accumulated preferentially on the X chromosome (Scott et al., 2006) with a similar value compared with other species. My study could provide new results about TMU and TOS LINEs, indicating a similar tendency of LINEs accumulation in three *Tokudaia* species. Principally composed by L1 in around 99% of the total LINEs, and L2 were less than 1% in TMU and TOS (0.07 and 0.04%, respectively). Both L1 and L2 play significant roles in genetic diversity and evolution by acting as mobile genetic elements that create new mutations. They influence gene regulation by altering expression patterns and affect genomic organization by impacting chromatin structure and accessibility (Roller et al., 2021). L2 is less active than L1, but still contribute significantly to genome dynamics, whereas L1 is more abundant and has a greater impact on genome evolution and structure due to their higher activity and more frequent insertions (Chow et al., 2010; Sellis et al., 2007).

For the other chromosomes, in mice and humans, the most prevalent type of TE are non-LTR retrotransposons (around 17–20% of the genome) and with the highest concentration in the X chromosome (Dechaud et al., 2019; Mangiavacchi et al., 2021). This tendency corresponded with my results of genus *Tokudaia*. In the case of

the DNA transposons, the presence of haT and Tc could be related to the structural conformation of X chromosome processes. In humans, hAT-Charlie family of DNA transposons has contributed extensively to CTCF sites involved in 3D genome folding (Choudhary et al., 2023). In the case of Tc1, it had been seen in freshwater bony fishes the relation with a longer and more active of expansion. It could be related to a more favorable environment for the proliferation of this type of transposons (Yuan et al., 2018). In that way, I can know about the environment of the species by looking into the proliferation of certain kinds of TEs.

The high accumulation of Alu sequences in the genus *Tokudaia*, especially in TMU and TTO could be related to the nature of Alu. Alu is a SINE class without the capacity to produce copies of themselves or integrate (also called nonautonomous) and double stranded DNA. For those activities, they rely on LINEs, called L1 (Pray, 2008). In contrast, B2 transposable elements, derived from tRNA genes, are specific to rodents like mice and rats. While B2 elements also play roles in genome dynamics and gene regulation, their impact is less pronounced due to their lower abundance and restricted distribution than Alu elements in primates (Horton et al., 2023). The presence of tRNA, the primer molecules that guide the reverse transcription of retrotransposons and U (unique) type sequences, could be related to a new process of post-transcriptional regulation in the XO/XO species after the loss of Y and one of X chromosomes (Martinez, 2017).

I also found a unique presence of gypsy in the *Tokudaia* species. Generally, there is an extremely low copy number in other mammalian genomes (Volff & Schartl,

2001). It had been seemed that insertion into gypsy sequences that are beneficial for the host would result in their disruption and loss of function. Individuals carrying such "knock-out" would undergo negative selection and may or may not disappear from the population (Srikulnath et al., 2022) Thus, analyzing Ty3/Gypsy interruptions offers a novel means of investigating selective pressures on drive sex chromosome differentiation.

The constant density of LINEs, LTRs, SINEs, and DNA transposons across the X chromosome could be related to the presence of conserved regions and the importance of the 3D structure in the X chromosome across rodents. However, it has seen a high presence of TEs in some regions. Many species are supporting the highly concordant presence of TEs in centromeres. However, direct involvement of TEs in defining centromere identity remains elusive (Klein & O'Neill, 2018). In humans, the density of LINEs is lowest in a region of the short arm where numerous genes escape inactivation, favoring the completeness (Bailey et al., 2000). Based on that, in our results, the decrease of LINEs in the XO/XO species could be related to the loss of the mechanism to produce the XCI.

Although differences in the concentration of TEs between mouse, rat, and genus *Tokudaia*, the representation of each class and subclass is conserved, however, some specific elements are prevalent in the genus *Tokudaia*, like gypsy (Table 6, Fig. 13). In this chapter, I found LINE elements as the most prevalent TEs type in the X chromosome, followed by LTRs, SINEs, and DNA transposons. Moreover, I found the increasing density of LINEs in the XO/XO species (Fig. 14). It is necessary to

verify the contribution of LINEs and their relation to the inactivation and the other TE elements in the genus *Tokudaia* by immunostaining inhibitory elements on the X chromosome, such as H3K27me3. To further validate the significance of the observed differences, additional samples should be included in future studies.

3.6 Tables

Table 6 Summary of transposable elements identified using a species-specific *de novo* library in the *Tokudaia* species.

Table1	Summary of transposable elements identified using a species-specific de novo library									
Species (Size of X, bp)		**TMU (140141768)**			**TTO (140990620)**			**TOS (141763114)**		
Type	**Subtype**	**Number**	**Length (bp)**	**%**	**Number**	**Length (bp)**	**%**	**Number**	**Length (bp)**	**%**
SINEs		54730	7807223	5.57	57048	7680007	5.45	50601	5287443	3.73
Sub-families	**Alu**	29906	-	2.90	38100	-	3.48	21014	-	1.48
	B2	19543	-	1.89	15173	-	1.38	23021	-	1.62
	B4	4038	-	0.39	2023	-	0.18	3215	-	0.22
	ID	3531	-	0.34	3480	-	0.31	4644	-	0.32
	MIR	547	-	0.02	547	-	0.05	383	-	0.22
	tRNA	47	-	0.004	177	-	0.01	459	-	0.03
	Other	34	-	0.003	146	-	0.01	97	-	0.006
LINEs		65126	41330662	29.49	82363	46159822	32.73	102868	47045842	33.18
	L1	65078	41327117	29.49	82363	46159822	32.73	102822	47042079	33.18
	L2	48	3545	0	0	0	0	46	3763	0
LTRs		37003	10377753	7.41	39394	10240643	7.26	39113	10570241	7.46
	Gypsy	26	16932	0.01	32	16509	0.01	45	35869	0.03
	Retroviral	36977	10360821	7.39	39310	10214015	7.24	39057	10532981	7.43
Sub-families	**ERVL-MaLR**	15950	-	2.79	17758	-	2.97	17703	-	3.04
	ERV1	4684	-	0.84	3926	-	0.65	5263	-	0.9
	ERVK	15346	-	2.76	15759	-	2.63	15974	-	2.74
	ERVL	5451	-	0.98	5709	-	0.95	4084	-	0.7
	ERV	-	-	-	56	-	0.009	90	-	0.01
	Unknown	-	-	-	-	-	-	12	-	0.002
Class I		**156859**	**59515638**	**42.47**	**178805**	**64080472**	**45.44**	**192582**	**62903526**	**44.37**
DNA transposons		3615	667424	0.48	3099	555739	0.39	3356	616423	0.43
Sub-families	**hAT**	2669	503100	0.36	2350	417027	0.3	2531	471839	0.33
	hAT-Charlie	2636	-	0.35	2350	-	0.29	2531	-	0.33
	hAT-Tip100	33	-	0.005	0	-	0	0	-	0
	Tc	946	164324	0.12	749	138712	0.1	825	144584	0.1
	TcMar-Mariner	685	-	0.03	599	-	0.02	599	-	0.03
	TcMar-Tigger	261	-	0.08	150	-	0.07	226	-	0.07
	Academ-1	0	-	0	0	-	0	0	-	0
Class II		**3615**	**667424**	**0.48**	**3099**	**555739**	**0.39**	**3356**	**616423**	**0.43**
Unclassified		7672	2017166	1.44	6059	1950957	1.38	8357	1910962	1.35
Total interspersed repeats		168146	62200228	44.38	187963	66587168	47.22	204295	65430911	46.15

Table 6 (Continued) Summary of transposable elements identified using a species-specific *de novo* library in mouse and rat.

Table1	Summary of transposable elements identified using a species-specific de novo library						
Species (Size of X, bp)		**Mouse (164206332)**			**Rat (157468448)**		
Type	**Subtype**	**Number**	**Length (bp)**	**%**	**Number**	**Length (bp)**	**%**
SINEs		60673	9312006	5.49	58632	6937434	4.4
Sub-families	**Alu**	43718	-	3.65	34754	-	3.11
	B2	18594	-	1.55	17431	-	1.56
	B4	650	-	0.05	601	-	0.05
	ID	2077	-	0.17	8674	-	0.77
	MIR	573	-	0.04	533	-	0.03
	tRNA	0	-	0	0	-	0
	Other	71	-	0.006	91	-	0.008
LINEs		101530	59254870	34.96	92104	55985212	35.49
	L1	101530	59254870	34.96	92104	55985212	35.49
	L2	0	0	0	0	0	0
LTRs		47194	16736484	9.88	37980	9184160	5.82
	Gypsy	0	0	0	0	0	0
	Retroviral	46957	16683197	9.84	37751	9148353	5.8
Sub-families	**ERVL-MaLR**	25243	-	4.69	20901	-	2.84
	ERV1	6771	-	1.25	6233	-	0.84738
	ERVK	18102	-	3.36	13211	-	1.79568
	ERVL	2390	-	0.44	1976	-	0.26854
	ERV	129	-	0.02	99	-	0.01
	Unknown	271	-	0.05	246	-	0.03
Class I		**209397**	**85303360**	**50.33**	**188716**	**72106806**	**45.71**
DNA transposons		3949	747677	0.44	3324	610130	0.39
Sub-families	**hAT**	3201	587843	0.35	2608	489754	0.31
	hAT-Charlie	3201	-	0.33	2608	-	0.31
	hAT-Tip100	0	-	0	0	-	0
	Tc	748	130419	0.08	688	118609	0.08
	TcMar-Mariner	571	-	0.02	500	-	0.05
	TcMar-Tigger	177	-	0.06	188	-	0.02
	Academ-1	304	-	0.03	28	-	0.003
Class II		**3949**	**747677**	**0.44**	**3324**	**610130**	**0.39**
Unclassified		5039	1429129	0.84	3894	734214	0.47
Total interspersed repeats		218385	87480166	51.62	195934	73451150	46.56

3.7 Figures

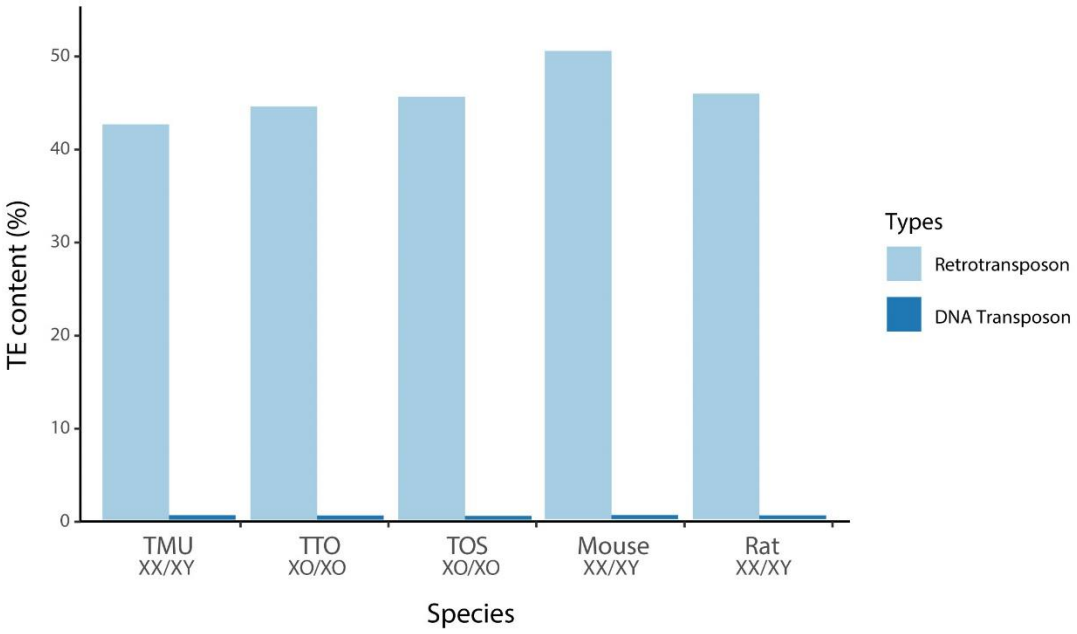

Fig. 10 Contents of Class I and II repeats in the X chromosome.

The ratio of Class I (retrotransposons, pale blue) and Class II (DNA transposons, blue) sequences in each X chromosome of *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), *Tokudaia osimensis* (TOS), mouse, and rat.

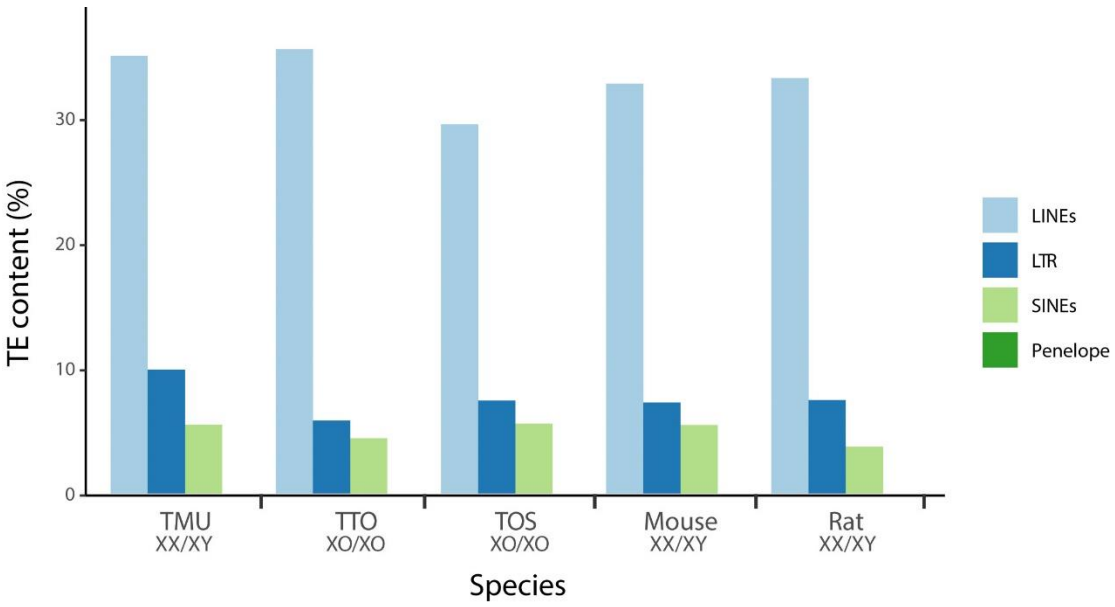

Fig. 11 Contents of Class I repeat in the X chromosome.

Contents of Class I (retrotransposons) were estimated by Repeatmasker in *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), *Tokudaia osimensis* (TOS), mouse, and rat.

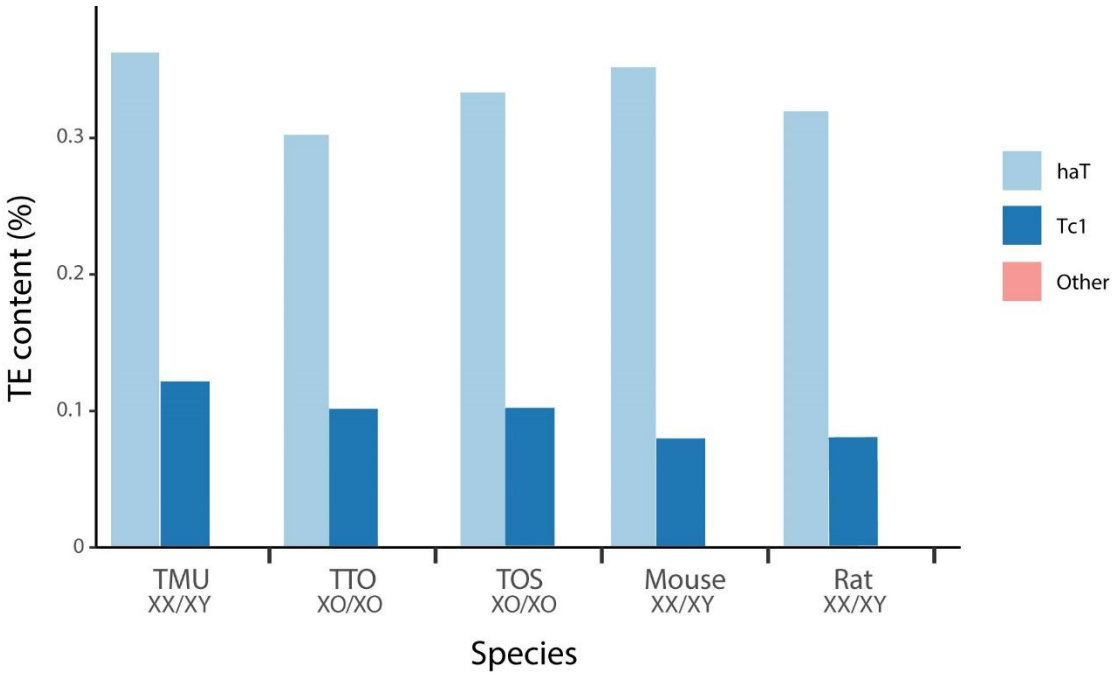

Fig. 12 Contents of Class II repeats in the X chromosomes.

Contents of subfamilies in Class II (DNA transposons) were estimated by Repeatmasker in *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), *Tokudaia osimensis* (TOS), mouse, and rat.

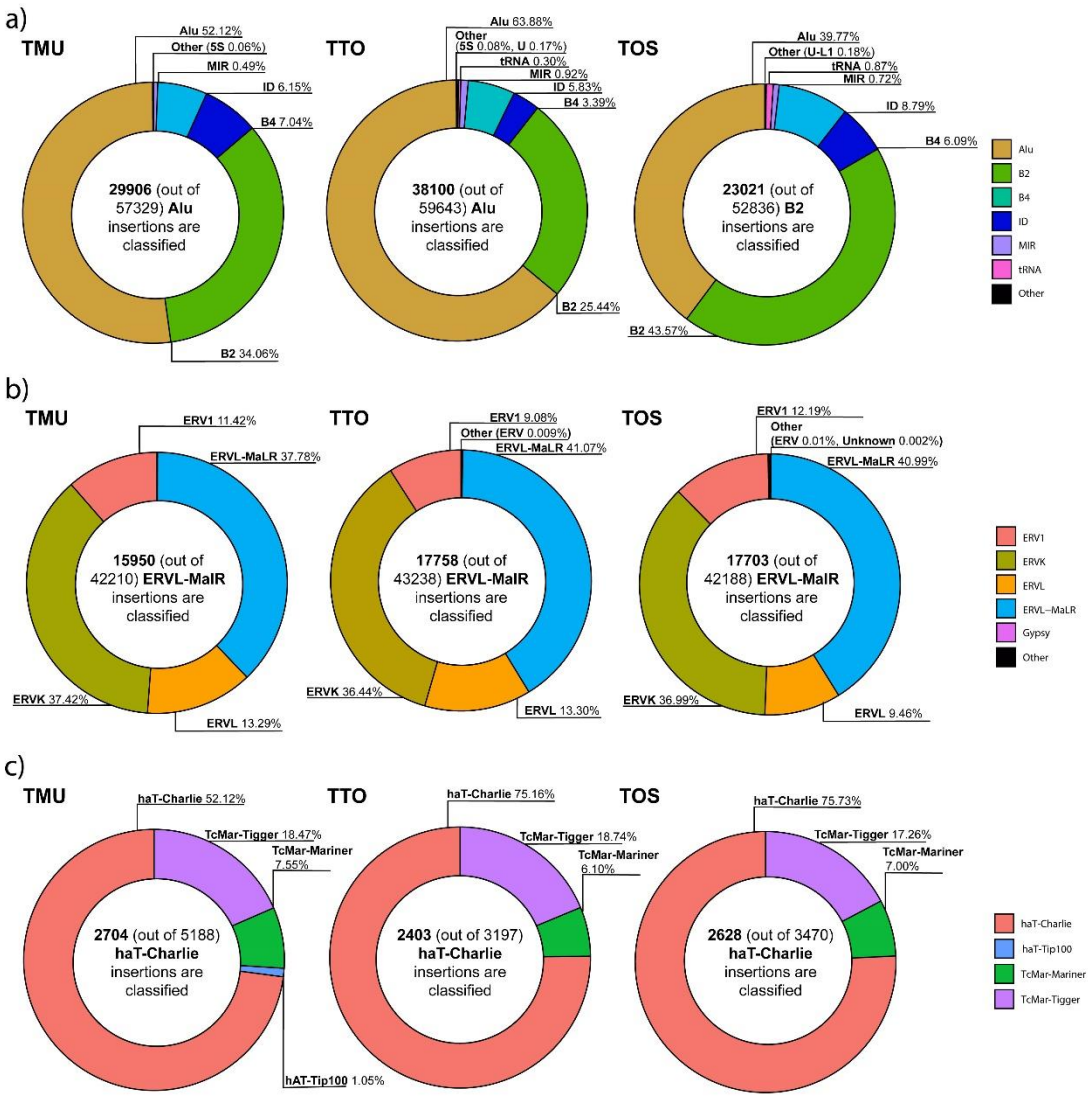

Fig. 13 Contents of subfamilies of Class I and II repeats in the X chromosome.

Contents of subfamilies in (a) SINEs, (b) LTRs, and (c) DNA transposons in *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), and *Tokudaia osimensis* (TOS).

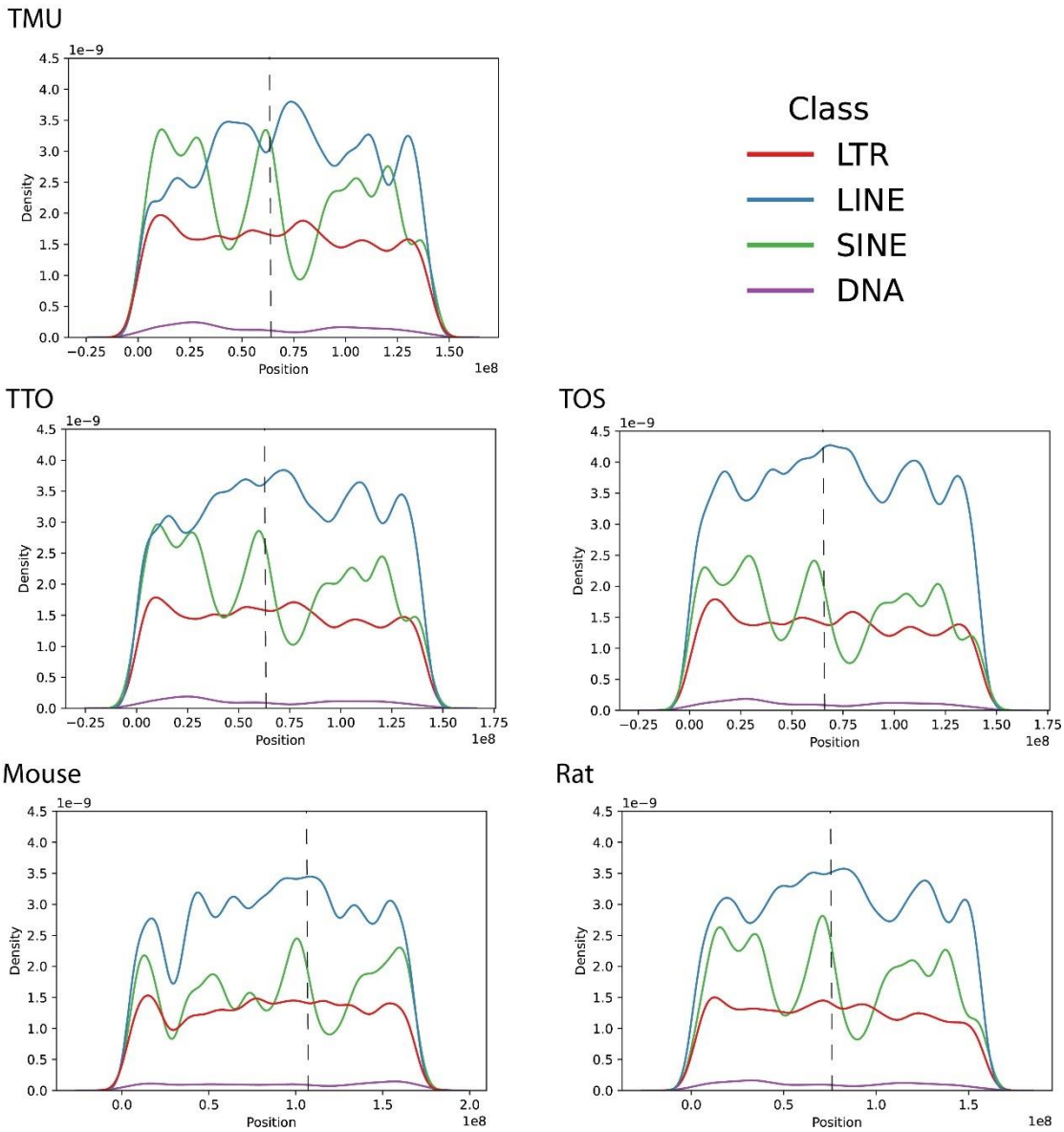

Fig. 14 Density estimation of repeats in the X chromosome.

Kernel density estimation (KDE) was performed in *Tokudaia muenninki* (TMU), *Tokudaia tokunoshimensis* (TTO), *Tokudaia osimensis* (TOS), mouse, and rat to show the repeats distribution based on the X chromosome location. Dotted line indicates the Xic position.

General discussion

XCI is an essential mechanism for dosage compensation between females and males in mammals. In females, XCI is controlled by a complex, conserved locus termed the Xic, in which the lncRNA *Xist* is the key regulator (Loda et al., 2022; Panning, 2008). However, little is known about the Xic in species with unusual sex chromosomes. The genus *Tokudaia* includes three rodent species (*T. muenninki*, *T. osimensis*, *T. tokunoshima*) endemic to Japan with a reported loss of one X and Y chromosome on the last two species (Honda et al., 1977, 1978; Murata et al., 2010). In this study, I used reciprocal genome information and RNA-Seq data of three *Tokudaia* species to analyze the Xic loci to understand the changes related to the unique sex chromosomes of these species. In Chapter I, I revealed that the Xic loci are still present in all species of the genus *Tokudaia*. The gene contents and order were conserved well between *Tokudaia* species and mouse. However, some changes in *Tsix* and *Xist*, and the intergenic regions of *Jpx-Ftx* were observed. Three deletions and one inversion were confirmed in the intergenic region between *Jpx* and *Ftx* in the XO/XO species, indicating that these mutations probably had occurred in the common ancestral population of these two species. In Chapter II, I compared the structure and expression of nine genes in the Xic between *Tokudaia* species and mouse. A comparison of gene structure showed that protein-coding genes were conserved well between *Tokudaia* species and mouse. By contrast, *Tokudai*a-specific structures were found in the lncRNA genes. Remarkable structural changes were observed in *Xist* and *Tsix* which were unique to the XO/XO species. Important functional repeats, B-, C-,

D-, and E-repeats were partially or completely lost due to deletions in these species. RNA-seq data showed that female-specific expression patterns of *Xist* and *Tsix* were confirmed in TMU however not in the XO/XO species, TTO and TOS. Based on these previous findings, the deletions in *Xist* in TTO and TOS could affect the recognition and maintenance of the future Xi. The *Xist* lncRNAs are not expressed in the XO/XO species, however, our findings indicate that even if they are expressed, they are incapable of producing a successful and lasting XCI in the XO/XO species due the deletions in *Xist/Tsix* and intergenic *Jpx-Ftx*. Finally, in chapter III I analyze the presence and distribution of repeats (as transposons) in X chromosome of genus *Tokudaia*, mouse and rat. LINEs type was the highest bulk in all the species. However, some unique TEs were found in genus *Tokudaia*, and a peak of SINEs was found close to Xic loci for all the species. In this study, I analyzed with bioinformatics tools the presence, structure, and gene order of the Xic loci in the genus *Tokudaia*, also the presence of TEs in the X chromosome.

The Xic locus is a critical region on the X chromosome responsible for initiating XCI in mammals, a process essential for dosage compensation between males and females (Augui et al., 2011; Deng et al., 2014b; Loda et al., 2022). The conservation of the Xic locus across different mammalian species underscores its vital role in development. This region contains key genes and regulatory elements, such as *Xist*, which are necessary to correct inactive one X chromosome in females (Raposo et al., 2021). The *Xist* gene coats the X chromosome from which it is transcribed, initiating a cascade of events leading to its silencing (Pontier & Gribnau, 2011). Without the

expression of *Xist*, XCI cannot occur, as this RNA is fundamental for recruiting the necessary chromatin-modifying complexes that establish and maintain the inactive state (Bousard et al., 2019; Raposo et al., 2021). The complex interactions between genes (*Cdx4, Chic1, Tsx, Tsix, Jpx, Ftx, Rlim,* and *Slc16a2*) within the TAD-E and TAD-D are integral to this process (van Bemmel et al., 2019). These interactions facilitate the proper regulation and expression of *Xist* and other associated genes, ensuring the coordinated silencing of the X chromosome. Thus, the genes within the Xic loci and the expression of *Xist* are indispensable for the successfully implementing XCI, highlighting their critical roles in maintaining dosage compensation and genomic stability in mammals.

In our findings within *Tokudaia* species, TTO and TOS showed the highest sequence identity in the Xic loci. A previous molecular phylogenetic analysis revealed that TMU (XX/XY) diverged first among the three species, and TOS and TTO (XO/XO) are phylogenetically closely related (Murata et al., 2010). Therefore, the patterns of sequence identity likely reflect the phylogenetic relationships within the genus. Protein-coding genes showed a consistent size and high identity between *Tokudaia* species and mouse, whereas partial variation was discovered in the regions within ncRNA genes. The conservation of the protein-coding genes is mediated due to a constant evolutionary pressure to maintain the sequence, due to its role in Xic or/and other process of the body. On the other hand, for long non-coding RNA, the space for variability and change been higher that protein-coding was reflected in my results. In the case of the XO/XO species, even if the *Xist* and *Tsix* lncRNAs are expressed, they

cannot produce a successful and lasting XCI. The variation in the three ncRNA genes, *Tsx*, *Ftx*, and *Jpx*, represent a distinct mark of the *Tokudaia* species and these variations are not related with sex chromosome constitution. It remains unclear how the early expression of *Xist* and other Xic loci genes in genus *Tokudaia* express. However, the results indicates that in female cells that have lost one X chromosome, XCI no longer occurs, resulting in loss of function of the *Xist* gene. Nucleotide mutations in genes that have lost their function will not be subject to natural selection. Like the lack of selection pressure in the intergenic region between the *Jpx*-*Ftx*, Unraveling the RNA-Seq data from initial cell stages or transgenic mice could help to understand how the Xic loci were influenced by the loss of the XCI in the XO/XO species.

In the case of when XCI becomes unnecessary due to changes in genetic or environmental factors, the structure and function of the Xic locus may evolve, leading to variations in its sequence or regulation. I hypothesized that such structural changes would be more likely to occur because of having a single X chromosome and self-synapsis. In XO mice, an XO pachytene oocyte in which an unsynapsed univalent X is silenced shows pachytene arrest. By contrast, an XO pachytene oocyte with a self-synapsed X chromosome remains active (Burgoyne et al., 2009). Self-synapsis potentially facilitates chromosomal rearrangements. Since it is not possible to obtain tissues from the endangered spiny rat for observations of synapsis, sequence analyses of whole X chromosomes are an important approach to confirm our hypothesis. Despite these potential changes, the overall structural conservation and precise order

of elements within the Xic locus are crucial for ensuring the accurate and efficient initiation and spread of XCI (Nesterova et al., 2001; van Bemmel et al., 2019). This structural integrity helps maintain proper gene dosage and prevents disorders associated with X chromosome misregulation, highlighting the importance of the Xic locus in mammalian development and evolution.

In the last part of this research, I described the distribution of TEs in the X chromosome of *Tokudaia* species and compared them to that of mouse and rat. First, the detailed distribution of classes, families, and subfamilies of TEs were identified. I showed that the LINEs were prevalent in the X chromosome of the compared species. Some unique TEs were identified in the *Tokudaia* species, such as a gypsy. The density of the TEs was also calculated. An increasing density in LINEs was identified in the XO/XO species around the Xic loci. Transposable elements, particularly LINEs, play a significant role in the evolution and divergence of sex chromosomes (Roller et al., 2021). These mobile genetic elements contribute to genomic variability by inserting themselves into new locations within the genome, which can lead to mutations, gene disruptions, and alterations in gene regulation (Bourque et al., 2018). Additionally, the accumulation of transposable elements, including LINEs, on the sex chromosomes has driven their structural divergence from autosomes, contributing to the unique evolutionary paths of the X and Y chromosomes. On the X chromosome, LINEs are particularly important as they facilitate the process of XCI. LINEs help to spread the silencing signal initiated by the *Xist* RNA across the X chromosome, ensuring that the entire chromosome is inactivated efficiently (Lyon, 2003). The

findings of unique transposable elements specific to *Tokudaia* could help us understand the mutational forces that have modified the conserved Xic in mammals. Furthermore, the decreased density of crucial LINE elements in the XO/XO species indicates a loss of the inactivation capacity of the X chromosome, shedding light on the evolutionary and functional dynamics of sex chromosome regulation. As a result, in species like TTO and TOS, where the XIC is no longer necessary, TEs may accumulate over time, accelerating mutations in the sequence. These mutations could occur because there is no selective pressure to maintain the functionality of these genes, given that their primary role (regulating XCI) is redundant in an XO/XO context. This creates new evolutionary landscapes in species that have lost the necessity of this process, as reported in this research for the genus *Tokudaia* and the Xic loci. It is still worthwhile to analyze other evolutionary adaptations and potential functional divergences of the X chromosome in *Tokudaia*, especially in the context of species-specific mechanisms for understanding mammalian diversity. However, since the difference count values of TE were not significant, more samples will be need it in the future to confirm the hypothesis about TE role in diversification and the role of LINE in *Xist* spread on X chromosome. In another hypothesis, the high p-value could be related with the recent specification of the species in *Tokudaia* genus. With more time, it could be possible to reflect a remarkable number of changes that illustrates the role of TE in the mutation and posterior change of the X chromosome.

Acknowledgements

My time at the Graduate School of Life Science, Hokkaido University was fast paced but incredibly fulfilling. Thanks to this experience, my attitude towards science and my abilities have changed significantly. I have grown both as a person and as a scientist.

First, I express my sincere appreciation to my supervisor, Professor Asato Kuroiwa. Prof. Kuroiwa always guided me with patience and care, helping to improve not only our research but also my awareness of critical details. Even in the most challenging moments of my research, she never let go of my hand. She is a role model for me, and I highly admire her work and character.

Second, I am also grateful to my research supervisors. Professor Atsushi Kimura for helpful suggestion, discussion, and practical guidance. Similarly, I would like to thank Assistant Professor Ikuya Yoshida for his valuable comments and revisions of my ideas and results. Additionally, I am grateful to Assistant Professor Shusei Mizushima for his feedback and for always boosting my morale to continue my research.

I would like to thank Professor Takehiko Itoh, the School of Life Science and Technology, Tokyo Institute of Technology and Lecturer Miki Okuno, the Department of Infectious Medicine at Kurume University. I could not have completed this research without their help and feedback. Their critical knowledge changed my perspective on the research and significantly improved the quality of our

investigation. They also provided high-quality genome and RNA-seq data for the genus *Tokudaia*.

I am also grateful to the members of the Kuroiwa laboratory for stimulating discussions and their supportive help in pursuing my thesis research.

I am thankful for the financial support provided by the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT). I never imagined fulfilling my dream of studying and living in Japan without your support. You have changed my life. I extend my gratitude to the Japanese people and government for supporting this program. I hope to contribute positively to the relationship between Japan and Colombia.

To my love, Yaka: Without your emotional support, I would never have reached this goal. Thank you for believing in me and making me feel special every day.

Thanks to my parents, Carlos and Luz Stella. No matter how crazy my ideas are, they always support me with love and patience. I will never tire of saying that I am who I am because of their efforts.

To my sister, Juliana: Thank you for supporting my dream to come to Japan and for taking care of our parents while I pursued my dream. You are the most important person in my life.

Coming to Hokkaido University was a great decision that helped me grow as a person, a scientist, and a global citizen. I am deeply grateful for all the wonderful people and opportunities that have come my way since making this decision.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anguera, M. C., Ma, W., Clift, D., Namekawa, S., Iii, R. J. K., & Lee, J. T. (2011). Tsx Produces a Long Noncoding RNA and Has General Functions in the Germline, Stem Cells, and Brain. *PLOS Genetics*, *7*(9), e1002248. https://doi.org/10.1371/journal.pgen.1002248

Augui, S., Nora, E. P., & Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nature Reviews Genetics*, *12*(6), Article 6. https://doi.org/10.1038/nrg2987

Avner, P., & Heard, E. (2001). X-chromosome inactivation: Counting, choice and initiation. *Nature Reviews Genetics*, *2*(1), 59-67. https://doi.org/10.1038/35047580

Barakat, T. S., Gunhanlar, N., Pardo, C. G., Achame, E. M., Ghazvini, M., Boers, R., Kenter, A., Rentmeester, E., Grootegoed, J. A., & Gribnau, J. (2011). RNF12 Activates Xist and Is Essential for X Chromosome Inactivation. *PLOS Genetics*, *7*(1), e1002001. https://doi.org/10.1371/journal.pgen.1002001

Berletch, J. B., Ma, W., Yang, F., Shendure, J., Noble, W. S., Disteche, C. M., & Deng, X. (2015). Escape from X Inactivation Varies in Mouse Tissues. *PLOS Genetics*, *11*(3), e1005079. https://doi.org/10.1371/journal.pgen.1005079

Berletch, J. B., Yang, F., Xu, J., Carrel, L., & Disteche, C. M. (2011). Genes that escape from X inactivation. *Human Genetics*, *130*(2), 237-245. https://doi.org/10.1007/s00439-011-1011-z

Bernal, J., Guadaño-Ferraz, A., & Morte, B. (2015). Thyroid hormone transporters—Functions and clinical implications. *Nature Reviews Endocrinology*, *11*(7), 406-417. https://doi.org/10.1038/nrendo.2015.66

Biémont, C., & Vieira, C. (2006). Junk DNA as an evolutionary force. *Nature*, *443*(7111), 521-524. https://doi.org/10.1038/443521a

Boeke, J. D., Garfinkel, D. J., Styles, C. A., & Fink, G. R. (1985). Ty elements transpose through an RNA intermediate. *Cell*, *40*(3), 491-500. https://doi.org/10.1016/0092-8674(85)90197-7

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology*, *19*(1), 199. https://doi.org/10.1186/s13059-018-1577-z

Bousard, A., Raposo, A. C., Żylicz, J. J., Picard, C., Pires, V. B., Qi, Y., Gil, C., Syx, L., Chang, H. Y., Heard, E., & da Rocha, S. T. (2019). The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO Reports*, *20*(10), e48019. https://doi.org/10.15252/embr.201948019

Boyle, A. L., Ballard, S. G., & Ward, D. C. (1990). Differential distribution of long and short interspersed element sequences in the mouse genome: Chromosome karyotyping by fluorescence in situ hybridization. *Proceedings of the National Academy of Sciences*, *87*(19), 7757-7761. https://doi.org/10.1073/pnas.87.19.7757

Brashear, W. A., Bredemeyer, K. R., & Murphy, W. J. (2021). Genomic architecture constrained placental mammal X Chromosome evolution. *Genome Research*, *31*(8), 1353-1365.

https://doi.org/10.1101/gr.275274.121

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J., & Willard, H. F. (1992). The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, *71*(3), 527-542. https://doi.org/10.1016/0092-8674(92)90520-M

Brown, P. O., Bowerman, B., Varmus, H. E., & Bishop, J. M. (1987). Correct integration of retroviral DNA in vitro. *Cell*, *49*(3), 347-356. https://doi.org/10.1016/0092-8674(87)90287-X

Burgoyne, P. S., Mahadevaiah, S. K., & Turner, J. M. A. (2009). The consequences of asynapsis for mammalian meiosis. *Nature Reviews Genetics*, *10*(3), Article 3. https://doi.org/10.1038/nrg2505

Choudhary, M. N. K., Quaid, K., Xing, X., Schmidt, H., & Wang, T. (2023). Widespread contribution of transposable elements to the rewiring of mammalian 3D genomes. *Nature Communications*, *14*(1), 634. https://doi.org/10.1038/s41467-023-36364-9

Chow, J. C., Ciaudo, C., Fazzari, M. J., Mise, N., Servant, N., Glass, J. L., Attreed, M., Avner, P., Wutz, A., Barillot, E., Greally, J. M., Voinnet, O., & Heard, E. (2010). LINE-1 Activity in Facultative Heterochromatin Formation during X Chromosome Inactivation. *Cell*, *141*(6), 956-969. https://doi.org/10.1016/j.cell.2010.04.042

Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P., & Rougeulle, C. (2010). Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Human Molecular Genetics*, *20*(4), 705-718. https://doi.org/10.1093/hmg/ddq516

Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P., & Duret, L. (2002). Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Research*, *12*(6), 894-908. https://doi.org/10.1101/gr.152902

Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., Grützner, F., & Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature*, *508*(7497), 488-493. https://doi.org/10.1038/nature13151

Cunningham, D. B., Segretain, D., Arnaud, D., Rogner, U. C., & Avner, P. (1998). The Mouse*Tsx*Gene Is Expressed in Sertoli Cells of the Adult Testis and Transiently in Premeiotic Germ Cells during Puberty. *Developmental Biology*, *204*(2), 345-360. https://doi.org/10.1006/dbio.1998.9004

Deakin, J. E., Chaumeil, J., Hore, T. A., & Marshall Graves, J. A. (2009). Unravelling the evolutionary origins of X chromosome inactivation in mammals: Insights from marsupials and monotremes. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *17*(5), 671-685. https://doi.org/10.1007/s10577-009-9058-6

Dechaud, C., Volff, J.-N., Schartl, M., & Naville, M. (2019). Sex and the TEs: Transposable elements in sexual development and function in animals. *Mobile DNA*, *10*(1), 42. https://doi.org/10.1186/s13100-019-0185-0

Delcher, A. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research*, *30*(11), 2478-2483. https://doi.org/10.1093/nar/30.11.2478

Deng, X., Berletch, J. B., Nguyen, D. K., & Disteche, C. M. (2014a). X chromosome regulation:

Diverse patterns in development, tissues and disease. *Nature Reviews Genetics*, *15*(6), Article 6. https://doi.org/10.1038/nrg3687

Deng, X., Berletch, J. B., Nguyen, D. K., & Disteche, C. M. (2014b). X chromosome regulation: Diverse patterns in development, tissues and disease. *Nature Reviews Genetics*, *15*(6), Article 6. https://doi.org/10.1038/nrg3687

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. https://doi.org/10.1093/bioinformatics/bts635

Duszczyk, M. M., Wutz, A., Rybin, V., & Sattler, M. (2011). The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. *RNA*, *17*(11), 1973-1982. https://doi.org/10.1261/rna.2747411

Fang, H., Disteche, C. M., & Berletch, J. B. (2019). X Inactivation and Escape: Epigenetic and Structural Features. *Frontiers in Cell and Developmental Biology*, *7*. https://doi.org/10.3389/fcell.2019.00219

Feschotte, C., & Pritham, E. J. (2007). DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics*, *41*(Volume 41, 2007), 331-368. https://doi.org/10.1146/annurev.genet.40.110405.090448

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451-9457. https://doi.org/10.1073/pnas.1921046117

Furlan, G., Gutierrez Hernandez, N., Huret, C., Galupa, R., van Bemmel, J. G., Romito, A., Heard,

E., Morey, C., & Rougeulle, C. (2018). The Ftx Noncoding Locus Controls X Chromosome Inactivation Independently of Its RNA Products. *Molecular Cell*, *70*(3), 462-472.e8. https://doi.org/10.1016/j.molcel.2018.03.024

Galupa, R., Nora, E. P., Worsley-Hunt, R., Picard, C., Gard, C., van Bemmel, J. G., Servant, N., Zhan, Y., El Marjou, F., Johanneau, C., Diabangouaya, P., Le Saux, A., Lameiras, S., Pipoli da Fonseca, J., Loos, F., Gribnau, J., Baulande, S., Ohler, U., Giorgetti, L., & Heard, E. (2020). A Conserved Noncoding Locus Regulates Random Monoallelic Xist Expression across a Topological Boundary. *Molecular Cell*, *77*(2), 352-367.e8. https://doi.org/10.1016/j.molcel.2019.10.030

Galupa, R., Picard, C., Servant, N., Nora, E. P., Zhan, Y., van Bemmel, J. G., El Marjou, F., Johanneau, C., Borensztein, M., Ancelin, K., Giorgetti, L., & Heard, E. (2022). Inversion of a topological domain leads to restricted changes in its gene expression and affects interdomain communication. *Development*, *149*(9), dev200568. https://doi.org/10.1242/dev.200568

Gayen, S., Maclary, E., Buttigieg, E., Hinten, M., & Kalantry, S. (2015). A Primary Role for the Tsix lncRNA in Maintaining Random X-Chromosome Inactivation. *Cell Reports*, *11*(8), 1251-1265. https://doi.org/10.1016/j.celrep.2015.04.039

Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., Tiana, G., & Heard, E. (2014). Predictive Polymer Modeling Reveals Coupled Fluctuations in Chromosome Conformation and Transcription. *Cell*, *157*(4), 950-963. https://doi.org/10.1016/j.cell.2014.03.025

Gong, T., Ju, F., & Bu, D. (2024). Accurate prediction of RNA secondary structure including

pseudoknots through solving minimum-cost flow with learned potentials. *Communications Biology*, *7*(1), 1-13. https://doi.org/10.1038/s42003-024-05952-w

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644-652. https://doi.org/10.1038/nbt.1883

Hake, L., & O'Connor, C. (2008). *Genetic mechanisms of sex determination* [Education]. Nature Education. http://www.nature.com/scitable/topicpage/genetic-mechanisms-of-sex-determination-314

Herzing, L. B., Romer, J. T., Horn, J. M., & Ashworth, A. (1997). Xist has properties of the X-chromosome inactivation centre. *Nature*, *386*(6622), 272-275. https://doi.org/10.1038/386272a0

Hierholzer, A., Chureau, C., Liverziani, A., Ruiz, N. B., Cattanach, B. M., Young, A. N., Kumar, M., Cerase, A., & Avner, P. (2022). A long noncoding RNA influences the choice of the X chromosome to be inactivated. *Proceedings of the National Academy of Sciences*, *119*(28), e2118182119. https://doi.org/10.1073/pnas.2118182119

Honda, T., Suzuki, H., & Itoh M. (1977). Itoh M. An unusual sex chromosome constitution found in Amami spinous country-rat, Tokudaia osimensis osimensis. Jpn J Genet, 52(3), 247-249. https://doi.org/10.1266/jjg.52.247

Honda, T., Suzuki, H., & Itoh M., & Hayashi, K. (1978). Karyotypical differences of the Aamami spinous country-rats, Tokudaia osimensis osimensis obtained from two neighbouring

islands. Jpn J Genet, 53(4), 297-299. https://doi.org/10.1266/jjg.53.297

Horton, I., Kelly, C. J., Dziulko, A., Simpson, D. M., & Chuong, E. B. (2023). Mouse B2 SINE elements function as IFN-inducible enhancers. *eLife*, *12*, e82617. https://doi.org/10.7554/eLife.82617

Hwang, J. Y., Choi, K.-H., Lee, D.-K., Kim, S.-H., Kim, E. B., Hyun, S.-H., & Lee, C.-K. (2015). Overexpression of OCT4A ortholog elevates endogenous XIST in porcine parthenogenic blastocysts. Journal of Reproduction and Development, advpub, 2015-2017. https://doi.org/10.1262/jrd.2015-017

Just, W., A. Baumstark, A. Süss, A. Graphodatsky, W. Rens, N. Schäfer, I. Bakloushinskaya, H. Hameister, W. Vogel. (2007). Ellobius lutescens: Sex determination and sex chromosome. Sexual Development : Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation, 1(4). https://doi.org/10.1159/000104771

Kashimada, K., & Koopman, P. (2010). Sry: The master switch in mammalian sex determination. *Development (Cambridge, England)*, *137*(23), 3921-3930. https://doi.org/10.1242/dev.048983

Katsir, K., & Linial, M. (2019). Human genes escaping X-inactivation revealed by single cell expression data. *BMC Genomics*, *20*(1), 201. https://doi.org/10.1186/s12864-019-5507-6

Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z., & Vyskot, B. (2009). *The role of repetitive DNA in structure and evolution of sex chromosomes in plants | Heredity*. https://www.nature.com/articles/hdy200917

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Research*, *12*(4), 656-664.

https://doi.org/10.1101/gr.229202

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357-360. https://doi.org/10.1038/nmeth.3317

Kozhukhar', V. G. (2012). [SRY and SOX9: The main genetic factors of mammalian sex determination]. *Tsitologiia*, *54*(5), 390-404.

Kristan, M., Leonardis, A., & Skočaj, D. (2011). Multivariate online kernel density estimation with Gaussian kernels. *Pattern Recognition*, *44*(10), 2630-2642. https://doi.org/10.1016/j.patcog.2011.03.019

Kudo, R., Yoshida, I., Matiz Ceron, L., Mizushima, S., Kuroki, Y., Jogahara, T., & Kuroiwa, A. (2023). The neo-X does not form a Barr body but shows a slightly condensed structure in the Okinawa spiny rat (Tokudaia muenninki). *Cytogenetic and Genome Research*. https://doi.org/10.1159/000531275

Kung, J. T., Kesner, B., An, J. Y., Ahn, J. Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B. C., Pinter, S. F., Erwin, J. A., & Lee, J. T. (2015). Locus-Specific Targeting to the X Chromosome Revealed by the RNA Interactome of CTCF. *Molecular Cell*, *57*(2), 361-375. https://doi.org/10.1016/j.molcel.2014.12.006

Lee, J., Davidow, L. S., & Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genetics*, *21*(4), Article 4. https://doi.org/10.1038/7734

Lee, J. T. (2000). Disruption of Imprinted X Inactivation by Parent-of-Origin Effects at Tsix. *Cell*, *103*(1), 17-27. https://doi.org/10.1016/S0092-8674(00)00101-X

Lee, J. T. (2011). Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nature Reviews Molecular Cell Biology*, *12*(12), 815-826.

https://doi.org/10.1038/nrm3231

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. https://doi.org/10.1186/1471-2105-12-323

Li, K.-B. (2003). ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics*, *19*(12), 1585-1586. https://doi.org/10.1093/bioinformatics/btg192

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)*, *22*(13), 1658-1659. https://doi.org/10.1093/bioinformatics/btl158

Loda, A., Collombet, S., & Heard, E. (2022). Gene regulation in time and space during X-chromosome inactivation. *Nature Reviews. Molecular Cell Biology*, *23*(4), 231-249. https://doi.org/10.1038/s41580-021-00438-7

Loda, A., & Heard, E. (2019). Xist RNA in action: Past, present, and future. *PLOS Genetics*, *15*(9), e1008333. https://doi.org/10.1371/journal.pgen.1008333

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *6*(1), 26. https://doi.org/10.1186/1748-7188-6-26

Luan, D. D., Korman, M. H., Jakubczak, J. L., & Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell*, *72*(4), 595-605. https://doi.org/10.1016/0092-8674(93)90078-5

Lyon, M. F. (1971). Possible Mechanisms of X Chromosome Inactivation. *Nature New Biology*,

232(34), Article 34. https://doi.org/10.1038/newbio232229a0

Lyon, M. F. (2003). The Lyon and the LINE hypothesis. *Seminars in Cell & Developmental Biology*, *14*(6), 313-318. https://doi.org/10.1016/j.semcdb.2003.09.015

Lyon, M. F. (2006). Do LINEs Have a Role in X-Chromosome Inactivation? *BioMed Research International*, *2006*, e59746. https://doi.org/10.1155/JBB/2006/59746

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianférani, S., Dorsselaer, A. V., Clerc, P., Avner, P., Visvikis, A., & Branlant, C. (2010). 2-D Structure of the A Region of Xist RNA and Its Implication for PRC2 Association. *PLOS Biology*, *8*(1), e1000276. https://doi.org/10.1371/journal.pbio.1000276

Malik, H. S., & Eickbush, T. H. (2001). Phylogenetic Analysis of Ribonuclease H Domains Suggests a Late, Chimeric Origin of LTR Retrotransposable Elements and Retroviruses. *Genome Research*, *11*(7), 1187-1197. https://doi.org/10.1101/gr.185101

Mangiavacchi, A., Liu, P., Della Valle, F., & Orlando, V. (2021). New insights into the functional role of retrotransposon dynamics in mammalian somatic cells. *Cellular and Molecular Life Sciences*, *78*(13), 5245-5256. https://doi.org/10.1007/s00018-021-03851-5

Martinez, G. (2017). tRNAs as primers and inhibitors of retrotransposons. *Mobile Genetic Elements*, *7*(5), 1. https://doi.org/10.1080/2159256X.2017.1393490

Matveevsky, S., Kolomiets, O., Bogdanov, A., Hakhverdyan, M., & Bakloushinskaya, I. (2017). Chromosomal Evolution in Mole Voles Ellobius (Cricetidae, Rodentia): Bizarre Sex Chromosomes, Variable Autosomes and Meiosis. *Genes*, *8*(11), E306. https://doi.org/10.3390/genes8110306

Mise, N., Goto, Y., Nakajima, N., & Takagi, N. (1999). Molecular Cloning of Antisense

Transcripts of the Mouse*Xist*Gene. *Biochemical and Biophysical Research Communications*, *258*(3), 537-541. https://doi.org/10.1006/bbrc.1999.0681

Monkhorst, K., Hoon, B. de, Jonkers, I., Achame, E. M., Monkhorst, W., Hoogerbrugge, J., Rentmeester, E., Westerhoff, H. V., Grosveld, F., Grootegoed, J. A., & Gribnau, J. (2009). The Probability to Initiate X Chromosome Inactivation Is Determined by the X to Autosomal Ratio and X Chromosome Specific Allelic Properties. *PLOS ONE*, *4*(5), e5616. https://doi.org/10.1371/journal.pone.0005616

Moore, E. C., & Roberts, R. B. (2013). Polygenic sex determination. *Current Biology*, *23*(12), R510-R512. https://doi.org/10.1016/j.cub.2013.04.004

Mulugeta, E., Wassenaar, E., Sleddens-Linkels, E., van IJcken, W. F. J., Heard, E., Grootegoed, J. A., Just, W., Gribnau, J., & Baarends, W. M. (2016). Genomes of Ellobius species provide insight into the evolutionary dynamics of mammalian sex chromosomes. *Genome Research*, *26*(9), 1202-1210. https://doi.org/10.1101/gr.201665.115

Murata, C., Yamada, F., Kawauchi, N., Matsuda, Y., & Kuroiwa, A. (2010). Multiple copies of SRY on the large Y chromosome of the Okinawa spiny rat, Tokudaia muenninki. *Chromosome Research*, *18*(6), 623-634. https://doi.org/10.1007/s10577-010-9142-y

Murata, C., Yamada, F., Kawauchi, N., Matsuda, Y., & Kuroiwa, A. (2012). The Y chromosome of the Okinawa spiny rat, Tokudaia muenninki, was rescued through fusion with an autosome. *Chromosome Research*, *20*(1), 111-125. https://doi.org/10.1007/s10577-011-9268-6

Navarro, P., Page, D. R., Avner, P., & Rougeulle, C. (2006). Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program.

Genes & Development, *20*(20), 2787-2792. https://doi.org/10.1101/gad.389006

Nesterova, T. B., Slobodyanyuk, S. Y., Elisaphenko, E. A., Shevchenko, A. I., Johnston, C., Pavlova, M. E., Rogozin, I. B., Kolesnikov, N. N., Brockdorff, N., & Zakian, S. M. (2001). Characterization of the Genomic Xist Locus in Rodents Reveals Conservation of Overall Gene Structure and Tandem Repeats but Rapid Evolution of Unique Sequence. *Genome Research*, *11*(5), 833-849. https://doi.org/10.1101/gr.174901

Okuno, M., Mochimaru, Y., Matsuoka, K., Yamabe, T., Matiz-Ceron, L., Jogahara, T., Toyoda, A., Kuroiwa, A., & Itoh, T. (2023). Chromosomal-level assembly of Tokudaia osimensis, Tokudaia tokunoshimensis, and Tokudaia muenninki genomes. *Scientific Data*, *10*(1), 927. https://doi.org/10.1038/s41597-023-02845-1

Panning, B. (2008). X-chromosome inactivation: The molecular basis of silencing. *Journal of Biology*, *7*(8), 30. https://doi.org/10.1186/jbiol95

Payer, B., & Lee, J. T. (2008). X chromosome dosage compensation: How mammals keep the balance. *Annual Review of Genetics*, *42*, 733-772. https://doi.org/10.1146/annurev.genet.42.110807.091711

Pollex, T., & Heard, E. (2019). Nuclear positioning and pairing of X-chromosome inactivation centers are not primary determinants during initiation of random X-inactivation. *Nature Genetics*, *51*(2), 285-295. https://doi.org/10.1038/s41588-018-0305-7

Pontier, D. B., & Gribnau, J. (2011). Xist regulation and function eXplored. *Human Genetics*, *130*(2), 223-236. https://doi.org/10.1007/s00439-011-1008-7

Pray, L. (2008). *Functions and Utility of Alu Jumping Genes | Learn Science at Scitable*. http://www.nature.com/scitable/topicpage/functions-and-utility-of-alu-jumping-genes-561

Pugacheva, E. M., Tiwari, V. K., Abdullaev, Z., Vostrov, A. A., Flanagan, P. T., Quitschke, W. W., Loukinov, D. I., Ohlsson, R., & Lobanenkov, V. V. (2005). Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. *Human Molecular Genetics*, *14*(7), 953-965. https://doi.org/10.1093/hmg/ddi089

Raposo, A. C., Casanova, M., Gendrel, A.-V., & da Rocha, S. T. (2021). The tandem repeat modules of Xist lncRNA: a swiss army knife for the control of X-chromosome inactivation. *Biochemical Society Transactions*, *49*(6), 2549-2560. https://doi.org/10.1042/BST20210253

Ridings-Figueroa, R., Stewart, E. R., Nesterova, T. B., Coker, H., Pintacuda, G., Godwin, J., Wilson, R., Haslam, A., Lilley, F., Ruigrok, R., Bageghni, S. A., Albadrani, G., Mansfield, W., Roulson, J.-A., Brockdorff, N., Ainscough, J. F. X., & Coverley, D. (2017). The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes & Development*, *31*(9), 876-888. https://doi.org/10.1101/gad.295907.117

Roller, M., Stamper, E., Villar, D., Izuogu, O., Martin, F., Redmond, A. M., Ramachanderan, R., Harewood, L., Odom, D. T., & Flicek, P. (2021). LINE retrotransposons characterize mammalian tissue-specific and evolutionarily dynamic regulatory regions. *Genome Biology*, *22*(1), 62. https://doi.org/10.1186/s13059-021-02260-y

Rosspopoff, O., Cazottes, E., Huret, C., Loda, A., Collier, A. J., Casanova, M., Rugg-Gunn, P. J., Heard, E., Ouimette, J.-F., & Rougeulle, C. (2023). Species-specific regulation of *XIST* by the *JPX/FTX* orthologs. *Nucleic Acids Research*, *51*(5), 2177-2194.

https://doi.org/10.1093/nar/gkad029

Sarma, K., Levasseur, P., Aristarkhov, A., & Lee, J. T. (2010). Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. *Proceedings of the National Academy of Sciences*, *107*(51), 22196-22201. https://doi.org/10.1073/pnas.1009785107

Sato, K., Akiyama, M., & Sakakibara, Y. (2021). RNA secondary structure prediction using deep learning with thermodynamic integration. *Nature Communications*, *12*(1), 941. https://doi.org/10.1038/s41467-021-21194-4

Saunders, P. A., & Veyrunes, F. (2021). Unusual Mammalian Sex Determination Systems: A Cabinet of Curiosities. *Genes*, *12*(11), 1770. https://doi.org/10.3390/genes12111770

Schoeftner, S., Sengupta, A. K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., & Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *The EMBO Journal*, *25*(13), 3110-3122. https://doi.org/10.1038/sj.emboj.7601187

Schulz, E. G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., & Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell*, *14*(2), 203-216. https://doi.org/10.1016/j.stem.2013.11.022

Scott, L. A., Kuroiwa, A., Matsuda, Y., & Wichman, H. A. (2006). X accumulation of LINE-1 retrotransposons in Tokudaia osimensis, a spiny rat with the karyotype XO. *Cytogenetic and Genome Research*, *112*(3-4), 261-269. https://doi.org/10.1159/000089880

Sellis, D., Provata, A., & Almirantis, Y. (2007). Alu and LINE1 Distributions in the Human

Chromosomes: Evidence of Global Genomic Organization Expressed in the Form of Power Laws. *Molecular Biology and Evolution*, *24*(11), 2385-2399. https://doi.org/10.1093/molbev/msm181

Sheardown, S. A., Duthie, S. M., Johnston, C. M., Newall, A. E. T., Formstone, E. J., Arkell, R. M., Nesterova, T. B., Alghisi, G.-C., Rastan, S., & Brockdorff, N. (1997). Stabilization of Xist RNA Mediates Initiation of X Chromosome Inactivation. *Cell*, *91*(1), 99-107. https://doi.org/10.1016/S0092-8674(01)80012-X

Shumate, A., Wong, B., Pertea, G., & Pertea, M. (2022). Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLoS Computational Biology*, *18*(6), e1009730. https://doi.org/10.1371/journal.pcbi.1009730

Smit, A., Hubley, R., & Green, O. (2016). *Repeatmasker* [Software]. Nucleic Acids Res. http://repeatmasker.org.

Spatz, A., Borg, C., & Feunteun, J. (2004). X-Chromosome Genetics and Human Cancer. *Nature Reviews Cancer*, *4*(8), 617-629. https://doi.org/10.1038/nrc1413

Sun, S., Del Rosario, B. C., Szanto, A., Ogawa, Y., Jeon, Y., & Lee, J. T. (2013). Jpx RNA Activates *Xist* by Evicting CTCF. *Cell*, *153*(7), 1537-1551. https://doi.org/10.1016/j.cell.2013.05.028

Sun, Z., Fan, J., & Wang, Y. (2022). X-Chromosome Inactivation and Related Diseases. *Genetics Research*, *2022*, 1391807. https://doi.org/10.1155/2022/1391807

Terao, M., Ogawa, Y., Takada, S., Kajitani, R., Okuno, M., Mochimaru, Y., Matsuoka, K., Itoh, T., Toyoda, A., Kono, T., Jogahara, T., Mizushima, S., & Kuroiwa, A. (2022). Turnover of mammal sex chromosomes in the *Sry*-deficient Amami spiny rat is due to male-specific

upregulation of *Sox9*. *Proceedings of the National Academy of Sciences*, *119*(49), e2211574119. https://doi.org/10.1073/pnas.2211574119

Tian, D., Sun, S., & Lee, J. T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X-chromosome inactivation. *Cell*, *143*(3), 390-403. https://doi.org/10.1016/j.cell.2010.09.049

Tsuchiya, K. (1989). Taxonomic study of Tokudaia (Rodentia: Muridae) : I. Genetic differentiation. *Memoirs of the National Science Museum, Tokyo*, *22*, 227-234.

van Bemmel, J. G., Galupa, R., Gard, C., Servant, N., Picard, C., Davies, J., Szempruch, A. J., Zhan, Y., Żylicz, J. J., Nora, E. P., Lameiras, S., de Wit, E., Gentien, D., Baulande, S., Giorgetti, L., Guttman, M., Hughes, J. R., Higgs, D. R., Gribnau, J., & Heard, E. (2019). The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of Tsix and Xist. *Nature genetics*, *51*(6), 1024-1034. https://doi.org/10.1038/s41588-019-0412-0

Vaught, A. (1996). Graphing with Gnuplot and Xmgr: Two graphing packages available under Linux. *Linux Journal*, *1996*(28es), 7-es.

Wilpert, N.-M., Krueger, M., Opitz, R., Sebinger, D., Paisdzior, S., Mages, B., Schulz, A., Spranger, J., Wirth, E. K., Stachelscheid, H., Mergenthaler, P., Vajkoczy, P., Krude, H., Kühnen, P., Bechmann, I., & Biebermann, H. (2020). Spatiotemporal Changes of Cerebral Monocarboxylate Transporter 8 Expression. *Thyroid: Official Journal of the American Thyroid Association*, *30*(9), 1366-1383. https://doi.org/10.1089/thy.2019.0544

Wu, T. D., & Watanabe, C. K. (2005). GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics (Oxford, England)*, *21*(9), 1859-1875.

https://doi.org/10.1093/bioinformatics/bti310

Wutz, A., & Gribnau, J. (2007). X inactivation Xplained. *Current Opinion in Genetics & Development*, *17*(5), 387-393. https://doi.org/10.1016/j.gde.2007.08.001

Yamada, F., Kawauchi, N., Nakata, K., Abe, S., Kotaka, N., Takashima, A., Murata, C., & Kuroiwa, A. (2010). Rediscovery After Thirty Years Since the Last Capture of the Critically Endangered Okinawa Spiny Rat Tokudaia muenninki in the Northern Part of Okinawa Island. *Mammal Study*, *35*(4), 243-255. https://doi.org/10.3106/041.035.0404

Yuan, Z., Liu, S., Zhou, T., Tian, C., Bao, L., Dunham, R., & Liu, Z. (2018). Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genomics*, *19*(1), 141. https://doi.org/10.1186/s12864-018-4516-1

Zushi, H., Murata, C., Mizushima, S., Nishida, C., & Kuroiwa, A. (2017). Unique XCI evolution in Tokudaia: Initial XCI of the neo-X chromosome in Tokudaia muenninki and function loss of XIST in Tokudaia osimensis. *Chromosoma*, *126*(6), 741-751. https://doi.org/10.1007/s00412-017-0639-4

Main Publication

Matiz-Ceron, L., Okuno, M., Itoh, T., Yoshida, I., Mizushima, S., Toyoda, A., Jogahara, T., Kuroiwa, A. (2024) Loss of one X and the Y chromosome changes the configuration of the X inactivation center in the genus *Tokudaia*. *Cytogenetics and Genome Research* doi: 10.1159/000539294.

List of other publications

Okuno, M., Mochimaru, Y., Matsuoka, K., Yamabe, T., **Matiz-Ceron, L.**, Jogahara, T., Toyoda, A., Kuroiwa, A., Itoh, T. (2023) Chromosomal-level assembly of *Tokudaia osimensis*, *Tokudaia tokunoshimensis*, and *Tokudaia muenninki* genomes. *Scientific Data*, 10(1), 927. doi: 10.1038/s41597-023-02845-1

Kudo, R., Yoshida, I., **Matiz Ceron, L.**, Mizushima, S., Kuroki, Y., Jogahara, T., Kuroiwa, A. (2022) The Neo-X does not form a Barr body but shows a slightly condensed structure in the Okinawa spiny rat (*Tokudaia muenninki*). *Cytogenetic and Genome Research*, 162(11-12), 632–643. doi: 10.1159/000531275.

Oral presentation

Matiz-Ceron, L., Okuno, M., Itoh, T., Yoshida, I., Mizushima, S., Toyoda, A., Jogahara, T., Kuroiwa, A. (2023). Loss of the Y chromosome changes the configuration of the X inactivation center in the genus *Tokudaia*. Asia-Pacific Chromosome Colloquium. Tekirdag, Turkiye. September 18, 2023.

Poster presentations

Matiz-Ceron, L., Mochimaru, Y., Matsuoka, K., Itoh, T., Mizushima, S., Kuroiwa, A. Loss of Y chromosome changes the configuration of X inactivation center in genus *Tokudaia*. The 45th Annual Meeting of the Molecular Biology Society of Japan, Chiba, Japan. December 1, 22.

Matiz-Ceron, L., Itoh, T., Mizushima, S., Kuroiwa, A. Loss of Y chromosome changes the configuration of X inactivation center in genus *Tokudaia*. 73th Annual Conference of the Society of Chromosome Research. Online. November 16, 2022.