Ancestral Y-linked genes were maintained by translocation to the X and Y chromosomes fused to an autosomal pair in the Okinawa spiny rat Tokudaia muenninki.
Ancestral Y-linked genes were maintained by translocation to the X and Y chromosomes fused to an autosomal pair in the Okinawa spiny rat *Tokudaia muenninki*

Chie Murata, Yoko Kuroki, Issei Imoto, Asato Kuroiwa

C. Murata
Department of Human Genetics, Institute of Biomedical Sciences, Tokushima University
Graduate School, 3-18-15 Kuramoto-cho, Tokushima 770-8503, Japan

Y. Kuroki
RIKEN, Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, Kanagawa 230-0045, Japan

Present address: Division of Pediatric Disease Genomics, Department of Genome Medicine, National Research Institute for Child Health and Development, 2-10-1, Okura, Setagaya-ku, Tokyo 157-8535, Japan

I. Imoto
Department of Human Genetics, Institute of Biomedical Sciences, Tokushima University
Graduate School, 3-18-15 Kuramoto-cho, Tokushima 770-8503, Japan

A. Kuroiwa
Laboratory of Animal Cytogenetics, Faculty of Science, Hokkaido University, Kita 10 Nishi 8, Kita-ku, Sapporo, Hokkaido 060-0810, Japan

Tel: +81-11-706-2752; Fax: +81-11-706-2619; E-mail: asatok@sci.hokudai.ac.jp

Correspondence: A. Kuroiwa

Keywords: sex chromosome, evolution, translocation, gene amplification, RNA-seq, BAC

A short running title: Y-linked gene translocations in the Okinawa spiny rat.
Introduction

The mammalian Y chromosome is male-specific and triggers the differentiation of the developing fetus into a male. The Y is one of the smallest chromosomes with the fewest genes, despite previous homology to the X chromosome (Ross et al. 2005). Most present therians (placentas and marsupials) maintain the sex chromosomes formed in the therian ancestor; a core set of Y-linked genes are shared among many eutherian species, while many multiple-copy genes are species- or lineage-specific (Paria et al. 2011; Bellott et al. 2014; Cortez et al. 2014). Models of Y gene loss dynamics revealed that Y chromosomal gene decay proceeds rapidly upon differentiation and then markedly levels off when small, but apparently essential, gene repertoires are defined (Hughes et al. 2012; Bellott et al. 2014; Cortez et al. 2014). In particular, the sex-determining gene SRY has persisted over long evolutionary time periods in most therians, with rare examples of SRY loss in the Tokudaia and Ellobius lineages (Just et al. 1995; Soullier et al. 1998; Vogel et al. 1998; Suto et al. 2001). Several long-lived ancestral Y-linked genes were also lost in at least one mammalian lineage during recent evolution in the clade (Bellott et al. 2014). However, other genes have been conserved by compensatory mechanisms, e.g., the rescue of Y-linked gene loss via gene transposition to other chromosomes (Hughes et al. 2015). This compensatory mechanism was first detected in Tokudaia species (Kuroiwa et al. 2010). The Amami and Tokunoshima spiny rats, Tokudaia osimensis and Tokudaia tokunoshimensis, respectively, have an XO/XO sex chromosome constitution; they lack the entire Y chromosome, except for a small region translocated to the X chromosome (Honda et al. 1977, 1978; Kuroiwa et al. 2010). The proto-Y region on the X chromosome contains at least four Y-linked genes, EIF2S3Y, KDM5D, TSPY, and ZFY, but functional RBMY and SRY have been absent from the genome since the secondary Y chromosome loss (Arakawa et al. 2002; Kuroiwa et al. 2010). Furthermore, KDM5D and EIF2S3Y are expressed in the gonads of both sexes in T. osimensis (Kuroiwa et al. 2010).

The Okinawa spiny rat Tokudaia muenninki has the XX/XY sex chromosome
constitution and multiple \textit{SRY} copies on the large Y chromosome fused to an autosome (Tsuchiya et al. 1989; Murata et al. 2010, 2012). It has an ancestral Y chromosome region and many \textit{SRY} sequences, but these \textit{SRY} loci may have not retained sex-determining functions (Murata et al. 2010; Kimura et al. 2014). \textit{SRY} proteins bind to the core region within a testis-specific enhancer of \textit{SOX9} (TESCO) with SF1 to upregulate \textit{SOX9} expression in the undifferentiated gonads of XY embryos of humans and mice (Sekido et al. 2008). TESCO of the genus \textit{Tokudaia} lost enhancer activity due to mutations in its SRY- and SF1-binding sites, and \textit{T. muenninki} \textit{SRY} does not activate the mutant TESCO or mouse TESCO (Kimura et al. 2014).

In our previous study, we found that two Y-linked genes, \textit{UTY} and \textit{DDX3Y}, are single-copy genes on the pericentromeric region of Yp, which is the ancestral Y chromosome region in \textit{T. muenninki} (Murata et al. 2012). By contrast, most other Y-linked genes are excessively duplicated, like the \textit{SRY}s. We investigated six apparently multi-copy Y-linked genes, \textit{ZFY}, \textit{UBAIY}, \textit{EIF2S3Y}, \textit{TSPY}, \textit{USP9Y}, and \textit{RBMY}, and two single-copy genes, \textit{UTY} and \textit{DDX3Y}, to examine whether the functions of multiple-copy genes were compensated by translocation or transposition. We performed RNA-seq using three tissue types to identify transcripts of ancestral Y-linked genes. We then conducted reverse transcription PCR (RT-PCR) and 3’ rapid amplification of cDNA ends (RACE) using six tissue types from an adult male to examine the functional conservation of these Y-linked genes based on expression patterns and the complete coding sequences, and to determine whether mRNA sequence variants existed for each gene. Furthermore, we performed fluorescent \textit{in situ} hybridization (FISH) using cDNA clones of Y-linked genes as probes to display their chromosomal localization. Using quantitative real-time PCR (qPCR) with genomic DNA, we estimated the copy number of each Y-linked gene in the male genome. The six genes were excessively duplicated on the Y chromosome, and several copies of \textit{ZFY} and \textit{UBAIY} were located on the X chromosome. We then performed next-generation sequencing of five bacterial artificial chromosome (BAC) clones including \textit{UTY} and/or \textit{DDX3Y} to reveal the genomic structure of the two genes as well as adjacent genes in the
ancestral Y euchromatic region. Our results show that functional Y-linked genes are compensated by translocations to other sex chromosomal regions and amplifications in *T. muenninki*.
**Materials and Methods**

**Sample preparation**

*T. muenninki* is endangered (The IUCN Red List of Threatened Species; http://www.iucnredlist.org/, 3/8/2016) and has been protected by the Japanese government as a natural treasure since 1972 (Yamada et al. 2010). With permission from the Agency for Cultural Affairs and the Ministry of the Environment in Japan, all spiny rats were released at their capture sites after a small piece was cut from the tip of their tail. Using the tail tissues, cell cultures and DNA extraction were performed as previously reported (Murata et al. 2010). Total RNA was extracted from the brain, liver, testis, kidney, lung, and spleen of one adult male *T. muenninki* who died of accidental causes in March 2013 using an RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer’s protocol. All experimental procedures conformed to the guidelines for animal experimentation of Tokushima University and all institutional and national guidelines for the care and use of laboratory animals were followed.

**RNA-seq and data assembly**

The cDNA libraries from the brain, liver, and testis were constructed using the TruSeq RNA Sample Preparation Kit v2 (Illumina, San Diego, CA, USA). Three indexed libraries from each tissue were pooled and sequenced using one lane to obtain 101 nucleotide single-end reads with an Illumina HiSeq 2500 RNA-seq data were deposited in the DDBJ Sequence Read Archive under the accession number DRA004624. *De novo* assembly and reference mapping were performed after removing polyA/T tails and filtering low-quality reads. *De novo* assembly was carried out using Trinity (Grabherr et al. 2011; Haas et al. 2013) with the default settings. BLAST searches using the assembled sequences were performed with Blast+ ver. 2.2.30 (Camacho et al. 2009; ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/). The National Center for Biotechnology Information (NCBI) transcript reference sequences (refseq_rna.*.tar.gz) were downloaded via FTP from ftp://ftp.ncbi.nlm.nih.gov/blast/db/. The *de*
assembled sequences matched to Y-linked genes, and these X homologs in the BLAST search were extracted using the ‘grep’ command. To fill in gaps between the extracted contigs, RNA-seq reads were mapped to RefSeq mRNA sequences from the mouse (refMrna.fa.gz downloaded from http://hgdownload.cse.ucsc.edu/goldenPath/mm10/bigZips/) using BWA 0.5.9 with a 15% uniform base error rate (Li and Durbin 2009). The Integrative Genomics Viewer was used to visualize reads mapped to the reference genome (Robinson et al. 2011). Accession numbers of the eight X-linked gene sequences determined in this analysis are listed in Table S1.

RT-PCR and 3’ RACE

RT-PCR and 3’ RACE were performed using testis tissue to determine the complete coding sequences of ancestral Y-linked genes by filling in gaps in the de novo assembled sequences and to confirm the sequence variation for the transcripts of each Y-linked gene. RT-PCR using multiple tissues from a male was also conducted to determine the expression patterns of the Y-linked genes. The cDNA was synthesized using the SuperScript III First-Strand System for RT-PCR Kit (Thermo Fisher Scientific, MA, USA). The primers for each RT-PCR amplification were designed using the partial transcript sequences determined by de novo assembly and the ancestral Y genome sequence determined by BAC sequencing. ACTB was used as a positive control, and its primer pair was designed from the common sequence of mouse and rat. The primer sequences are shown in Table S2. Amplification was carried out in a 20 µl mixture containing 0.5 µl of cDNA, 0.3 µM each primer, 0.2 mM each of the four deoxynucleotide triphosphates (dNTPs), 0.25 U of PrimeSTAR GXL DNA Polymerase, and a one-fifth volume of 5× PrimeSTAR GXL Buffer (TaKaRa, Shiga, Japan). The RT-PCR conditions were as follows: 98°C for 2 min; 30 cycles at 98°C for 10 s, 55–60°C for 15 s, and 68°C for 30 s to 4 min; and a final extension at 72°C for 7 min. For 3’ RACE, cDNA synthesis was initiated at the poly(A) tail of mRNA using the adapter primer 5’-TGG AAG AAT TCG CGG CCG CAG TTT TTT TTT TTT T-3’ with the PrimeScript II 1st Strand cDNA Synthesis Kit (TaKaRa). The
unknown coding region of **USP9Y** was amplified using the Ex Taq HS System (TaKaRa) with the gene-specific primer 5′-CAC CAG CAC CAC AAG ATA GA-3′ and the universal amplification primer 5′-TGG AAG AAT TCG CGG CCG CAG-3′. PCR products were cloned using the TArget Clone Kit (ToYoBo, Osaka, Japan) after overhanging dA was added to the 3′ ends of the products using A-attachment mix (ToYoBo). Both the PCR products and cDNA clones were sequenced using an ABI 3500xL Genetic Analyzer. The identified Y gene and homologous X sequences were compared with orthologous sequences of the mouse (**Mus musculus**), rat (**Rattus norvegicus**), **T. osimensis**, and **Apodemus sylvaticus** using ClustalX 2.0 and MEGA5.2 (Larkin et al. 2007; Tamura et al. 2008). Furthermore, we performed phylogenetic analyses for each X/Y gene pair including **T. muenninki** and other rodent genes using MEGA5.2 to confirm whether X- and Y-linked genes formed separate clusters. Accession numbers of the Y-linked transcript sequences determined by RT-PCR and 3′ RACE with **de novo** assembled sequences are listed in Table S1. Accession numbers of the X and Y gene sequences in other rodents are also listed in Table S1.

**Chromosome preparation and FISH**

The preparation of R-banded chromosomes was performed as described by Matsuda et al. (1992), Matsuda and Chapman (1995), and Kobayashi et al. (2008). The cDNA clones of **ZFY** (2,263 bp), **UBAIY** (1,778 bp), **EIF2S3Y** (1,458 bp), **TSPY** (691 bp), **USP9Y-2** (3,616 bp), and **RBMY** (1,123 bp) were used as probes. The cDNA clones **ZFX** (2,816 bp) and **UBAI** (2,114 bp) from the brain were also used, and the primer pairs used to detect the two X-linked genes by RT-PCR are shown in Table S2. The FISH analysis using the cDNA clones was performed as described by Kuroiwa et al. (2010). The FISH signals were observed under an Olympus fluorescence microscope BX53, and the images were captured with a DP73 digital camera (Olympus, Tokyo, Japan).
qPCR

The copy number of each amplified ancestral Y-linked gene was quantified using gene-specific primers designed using Primer3 ver. 0.4.0 (http://bioinfo.ut.ee/primer3-0.4.0/). For each qPCR, *DDX3Y* and *UTY*, single-copy genes in the *T. muenninki* male genome, were used as an internal control for copy number. The primer pairs used for this analysis are shown in Table S2. To achieve maximum efficiency of the qPCR, the product size was limited to 90–109 bp.

qPCR was conducted using TaKaRa SYBR Premix Ex Taq (Perfect Real Time) in accordance with the manufacturer’s protocol, and amplification was detected using the Applied Biosystems 7500 Real-Time PCR System. The qPCR reactions were performed in triplicate in a 10 µl reaction volume in a 96-well plate under the following conditions: 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 34 s. Genomic DNA samples from five *T. muenninki* males were used in the analysis. The presence of a single melting temperature peak, which represented a single amplicon, was used to validate primer specificity, and the slope and $R^2$ values for serial dilutions of genomic DNA were used to confirm the efficiency of the reaction.

To detect potential contamination during plate preparation, nuclease-free water was included in each set of reactions as a negative control.

BAC clone sequencing, data assembly, and sequence comparison

BAC libraries of a single male *T. muenninki* were previously constructed (Murata et al. 2015). Five BAC clone sequences were determined, including *UTY* and/or *DDX3Y* genes (TMB1-066O12, 237P18, 265P11, 324H06, and 389K21), by multiplex sequencing on an Illumina HiSeq 1000 using a 100 bp paired-end protocol, as described in our previous study (Murata et al. 2015). The primers used to screen the BAC libraries were designed within an exon from the transcript sequences of the two genes determined by RT-PCR and sequencing. The primer pairs used for gene amplification were as follows: *UTY*, Primers no. 2 and 3 in Table S2; *DDX3Y*, 5′-GGA AGG ATT CAT TGA CTT TAG TA-3′ and 5′-TCT GTG A TC TGT CTC
CAT GA-3’. Amplification was carried out in a 10 µl mixture containing 1.0 µl of library DNA, 0.5 µM each primer, 0.2 mM each of the four dNTPs, 0.5 U of BIOTaq HS DNA Polymerase, and a one-tenth volume of 10× ImmoBuffer. PCR was performed under the following conditions: denaturation for 10 min at 95°C, followed by 45 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 55°C, and extension for 30 s at 72°C. BAC sequences were de novo assembled as previously reported (Murata et al. 2015). Gaps within de novo assembled sequences were bridged by PCR and sequencing. Finally 372,407 bp of combined sequence from five BAC clones was determined without gaps (LC066213).

To capture genome rearrangements, a dot-plot analysis was performed for the T. muenninki Y and mouse Y regions. The mouse genome sequence (NC_000087) was obtained from NCBI. A nucleotide-based alignment was generated using NUCmer, which is a mode in MUMmer 3.0, for the large-scale genome comparisons with default settings (minimum length of a maximal exact match: 20, and minimum cluster length: 65; Kurtz et al. 2004). The exon-intron boundaries were defined for the combined sequence of five BAC clones by comparison to the transcript sequences of T. muenninki and orthologous genes in the mouse genome. To verify whether the genes located on the ancestral Y genome were transcribed into mRNAs, the predicted mRNA sequences were compared with the actual mRNA sequences of T. muenninki using ClustalX 2.0. Furthermore, the sequences surrounding SRY with masked repetitive elements were compared between T. muenninki, mouse, and rat using ClustalX 2.0. Repetitive elements were identified using the RepeatMasker Web Server (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker).
Identification of expressed Y-linked genes and their X homologs

Y-linked genes diverge more rapidly than other chromosomal genes among species. Therefore, the de novo assembly of RNA-seq data is a useful strategy to find Y-linked genes without a genome sequence, such as in non-laboratory animals. We performed RNA-seq using three tissue types to identify Y-linked genes in *T. muenninki*. We generated 115,946,604, 101,801,400, and 111,401,004 single-end reads from the testis, brain, and liver, respectively. A total of 91–92% of the sequenced reads passed the quality control check and were used for de novo Trinity assembly. After the de novo assembly of all available reads from the three tissues, we obtained 111,185 components with an N50 of 3,437 bp, defined as the maximum length such that 50% of the total assembled sequences were contained in contigs of at least this length. We extracted the partial transcript sequences that matched eight mouse and rat Y-linked genes (*ZFY*, *UBA1Y*, *EIF2S3Y*, *TSPY*, *UTY*, *DDX3Y*, *USP9Y*, and *RBMY*) by BLAST searches of the de novo assembled sequences. We also identified the complete coding sequences of six X-linked genes (*ZFX*, *UBA1X*, *KDM6A/UTX*, *DDX3X*, *USP9X*, and *RBMX*) from the assembled sequences, and we determined the sequences of two genes, *EIF2S3X* and *TSPYL2*, by filling in gaps using the read sequences mapped to the mouse RefSeq mRNA sequences. The similarities of the nucleotide and amino acid sequences were 91.8–97.6% and 93.4–99.4%, respectively, between X-linked genes of *T. muenninki* and mouse (Table S3).

Predicted amino acid sequences and expression patterns of Y-linked genes

Using RT-PCR and 3′ RACE with an adult testis tissue sample, we determined the complete coding sequences of seven Y-linked genes, *ZFY*, *EIF2S3Y*, *TSPY*, *UTY*, *DDX3Y*, *USP9Y*, and *RBMY*, but not *UBA1Y* (Fig. S1). The identification of X and Y genes was performed by phylogenetic analysis with each X/Y gene sequences of *T. muenninki* and other rodents, revealing that their Y-linked genes were grouped into the same cluster and formed separate
cluster to the X homologs (Fig. S2). We observed several nucleotide substitution sites within the transcript sequences of \textit{ZFY}, \textit{EIF2S3Y}, \textit{TSPY}, \textit{USP9Y}, and \textit{RBMY}, and no sequence variation within the \textit{UTY} and \textit{DDX3Y} transcripts (Fig. S1). The predicted amino acid sequences of the seven Y-linked genes exhibited 63.3--99.6\% similarity between \textit{T. muenninki} and other rodents (Table 1). Some of the \textit{ZFY} transcripts had 1 bp and 7 bp deletions at positions 369 and 494--496, respectively, in \textit{T. muenninki} (Fig. S1a). Therefore, the predicted amino acid sequences of the transcripts without either deletion were compared with the mouse and \textit{T. osimensis} orthologs (Table 1, Fig. S1b). We did not detect sequence variation in \textit{UBA1Y}, but we observed a partial coding sequence caused by an internal stop codon. The internal stop codon was generated by a frameshift 86 bp deletion, which resulted in the deletion of more than 230 amino acid residues (Fig. S1c, d). The predicted amino acid sequences of \textit{EIF2S3Y}, \textit{TSPY}, \textit{UTY}, and \textit{DDX3Y} were highly similar to the orthologs of other rodents without frameshift mutations (Table 1, Fig. S1e--l). In \textit{USP9Y}, we identified a difference between \textit{T. muenninki} and mouse orthologs at the last 22 amino acid residues due to a frameshift mutation (Fig. S1m--n). Based on a sequence comparison of \textit{RBMY} among \textit{T. muenninki}, mouse, and rat, we inferred that the frameshift mutation occurred in the mouse lineage, and sequence similarity was higher for the comparison between \textit{T. muenninki} and the rat than between \textit{T. muenninki} and the mouse (Table 1, Fig. S1o, p).

We examined the expression patterns of the eight Y-linked genes by RT-PCR using six tissues obtained from an adult male. We observed higher \textit{ZFY} expression in the testis and brain than in other tissues, and we detected \textit{UBA1Y} expression in all tissue types, except for the kidney. \textit{EIF2S3Y}, \textit{UTY}, and \textit{DDX3Y} showed ubiquitous expression, \textit{TSPY} and \textit{RBMY} were specifically expressed in the testis, and \textit{USP9Y} was expressed more highly in the testis than in other tissues in \textit{T. muenninki} (Fig. 1).

\textbf{Chromosomal localization of Y-linked genes}
We mapped the *T. muenninki* cDNA clones of six Y-linked genes, *ZFY, UBA1Y, EIF2S3Y, TSPY, USP9Y*, and *RBMY*, to chromosomes of a male by direct R-banding FISH. All six genes were located on the long arm of the Y chromosome as multiple-copy genes (Fig. 2). *ZFY* was also distributed on the distal region of Xq as well as Yq, and *ZFX* was located on the middle region of Xq (Fig. 2a–c, Fig. S3a, b). Furthermore, we observed FISH signals of *UBA1Y* along the heterochromatin block of the X chromosome as well as Yq (Fig. 2d–f), and we detected *UBA1* signals in the pericentromeric region of Xq (Fig. S3c, d). *EIF2S3Y* and *USP9Y* copies were mainly located on the pericentromeric and middle region of Yq, respectively (Fig. 2g–i). The other four genes were amplified on the pericentromeric and middle region of Yq (Fig. 2a–f, m–r). We also observed *USP9Y* on the pericentromeric region of Yp (Fig. 2j–l), which was the predicted ancestral Y euchromatic region including two single-copy genes *UTY* and *DDX3Y* (Murata et al. 2012).

Copy numbers of Y-linked genes

We performed qPCR using genomic DNA of five *T. muenninki* males to estimate the copy numbers of the six Y-linked genes (*ZFY, UBA1Y, EIF2S3Y, TSPY, USP9Y*, and *RBMY*) that were identified as multiple-copy genes in the FISH analysis (Fig. 2). We also examined the copy number of *SRY*, which was previously demonstrated to have multiple copies (Murata et al. 2010). We used *UTY* and *DDX3Y* as internal control genes with a single copy per male diploid genome. Setting the copy number value of *UTY* (or *DDX3Y*) to 1.0, the relative copy numbers for *DDX3Y* (or *UTY*), *ZFY, UBA1Y, EIF2S3Y, TSPY, USP9Y, RBMY*, and *SRY* were approximately 1, 63, 66, 10, 61, 34, 50, and 61, respectively (Fig. 3, Table 1).

Genomic structure of ancestral Y genes

We screened five BAC clones including *UTY* and/or *DDX3Y*, which are single-copy genes on Yp, to identify the genomic structure of ancestral Y genes. We determined a 372.4 kb sequence
by next-generation sequencing using the BAC clones. Based on a comparison between *T. muenninki* and the mouse Y chromosome sequence, we identified a large homologous block without apparent inversions (Fig. 4). We obtained the sequences of *SRY, EIF2S3Y, TSPY, UTY, DDX3Y*, and *USP9Y* (Fig. 4a, b). We detected the reversed *SRY* sequence upstream of *EIF2S3Y* in *T. muenninki*, although *SRY* was located more than 1 Mb downstream of *USP9Y* in the mouse genome (Fig. 4c). Exons 1–4 and exon 12, including the complete 447 bp coding region, were absent from *EIF2S3Y*, although the upstream and downstream sequences of *EIF2S3Y* were found in the BAC clone sequence of *T. muenninki* (Fig. 4, Fig. S1e). We did not detect exons 1–28 in *USP9Y* because the terminal region of the BAC sequence was within intron 29 of *USP9Y* (Fig. 4a). Additionally, the ancestral *USP9Y* genomic sequence had an internal stop codon within exon 29 (Fig. S1m). The predicted transcript sequences inferred from the genomic sequence of three multiple-copy genes, *EIF2S3Y, TSPY*, and *USP9Y*, were different from the actual mRNA sequences (Table 1, Fig. S1e, g, m). For single-copy genes, the *UTY* and *DDX3Y* mRNA sequences were completely consistent with the predicted sequences from the genome, and the exon-intron boundaries for the two genes corresponded to those of mouse orthologs, except for a 3 bp slip of an accepter splicing site in exon 16 of *DDX3Y* (Table 1, Fig. S1i, k and Fig. S4).

**Function and structure of the duplicated *SRY***

The *SRY* copy located immediately upstream of *EIF2S3Y* (named *SRY25*) in the euchromatic region was compared with previously reported *SRY* sequences of *T. muenninki*, mouse, and rat in the coding region and the surrounding region, to reveal whether it functions as a sex-determining gene. *SRY25* matches the previously identified *SRY5*, which is a pseudogene with an internal stop codon (Fig. S1q). In the region upstream of *SRY/Sry*, an approximately 3.1 kb sequence in the rat showed more than 80% and 71% similarity with *T. muenninki SRY* sequences segmented by long interspersed element (LINE) insertions (Fig. 5a, Fig. S5a). We did
not detect conservation between mice and the other two species, except for an approximately 0.7
kb region just before the start codon (Fig. 5a, Fig. S5a). Downstream of SRY/Sry, we observed
more than 6 kb of three conserved segments separated by insertions of several retrotransposable
elements between mouse and T. muenninki, but the conservation was limited to an
approximately 1.3 kb region between rat and the other two species (Fig. 5b, Fig. S5b). In the
SRY surrounding region, sequence similarity was higher between T. muenninki and other rodents
than between mouse and rat.

We then identified the distal end of a SRY duplication unit to identify factors related to
SRY amplification in T. muenninki by comparing the genomic region surrounding SRY between
T. muenninki and other rodents. We found a long-terminal repeat (LTR) at the distal end of a
SRY duplication unit, which was adjacent to the region with high similarity to mouse upstream
of Eif2s3y (Fig. 5b).
Discussion

We determined the complete coding sequences of ZFY, EIF2S3Y, TSPY, UTY, DDX3Y, USP9Y, and RBMY, but not UBA1Y, by RNA-seq, RT-PCR, and 3′ RACE (Fig. S1). We observed higher sequence similarity between T. muenninki and mouse than between mouse and rat for the single-copy genes UTY and DDX3Y, but slightly lower similarity between T. muenninki and mouse/Apodemus sylvaticus (the Wood mouse) than between mouse/A. sylvaticus and rat for the multiple-copy genes EIF2S3Y, TSPY, USP9Y, and RBMY (Table 1). Based on molecular phylogenetic studies, the Tokudaia lineage (sister to Apodemus) is more closely related to Mus than Rattus (Rowe et al. 2008), suggesting that the rate of evolution of the latter four Y-linked genes accelerated in T. muenninki relative to other rodents owing to the lineage-specific gene amplification. In the mouse and rat, ZFY, UBA1Y, USP9Y, TSPY, and RBMY are predominantly expressed in testes, whereas EIF2S3Y, UTY, and DDX3Y are broadly expressed in male tissues (Bellott et al. 2014; GEO Profiles in National Center for Biotechnology Information). Similar expression patterns are observed in T. muenninki and other rodents for the Y-linked genes, except for UBA1Y (Fig. 1). Our results indicated that the T. muenninki genome retained the seven functional Y-linked genes. We also observed several nucleotide substitution sites within the transcript sequences of ZFY, EIF2S3Y, TSPY, USP9Y, and RBMY (Fig. S1), indicating that mRNAs of each gene are transcribed from multiple gene regions with different sequences. Additionally, some of the ZFY transcripts had deletions within the coding region in T. muenninki, suggesting that both functional and pseudogene sequences of ZFY are expressed in T. muenninki testes (Fig. S1a).

We observed excessive amplification of the ancestral Y-linked genes on the Y chromosome in T. muenninki (Fig. 2, Fig. 3). Based on a FISH analysis, multiple copies of six genes (ZFY, UBA1Y, EIF2S3Y, TSPY, USP9Y, and RBMY) were mainly localized to the long arm of the Y chromosome. Our previous study demonstrates that the long arm of the Y chromosomes is entirely heterochromatic (Murata et al. 2010). EIF2S3Y and USP9Y copies were located on the
pericentromeric and middle region of Yq, respectively (Fig. 2g–l), while the other four genes and SRY were on the pericentromeric and middle regions (Fig. 2a–f, m–r, Murata et al. 2010). Consistent with the FISH results, which showed that EIF2S3Y and USP9Y copies were not located on the middle and pericentromeric regions, respectively, we detected fewer copies of these two genes than the other multiple-copy genes located in both regions by qPCR (Fig. 3). These results indicated that the Y segment that includes the seven Y-linked genes translocated from Yp to Yq, and the segment was subsequently amplified in the middle and pericentromeric regions of Yq after the loss of EIF2S3Y and USP9Y, respectively. Y-linked gene amplification is common during mammalian evolution; for example, it has been reported for ZFY, RBMY, TSPY, and SRY (Bianchi et al. 1993; Chai et al. 1998; Verkaar et al. 2004; Turner et al. 2007). However, this is the first report of the massive amplification of these seven Y-linked genes in a mammalian species. We also found that the Y-linked genes were translocated to the X chromosome at least two times in T. muenninki. ZFY and UBA1Y were distributed on the distal region of Xq and the centromeric heterochromatin of the X chromosome, respectively (Fig. 2a–f). In the T. muenninki X chromosome, the long and short arms consist of the ancestral X chromosome and a fused autosome, respectively (Murata et al. 2012). Dobigny et al. (2004) suggested that constitutive heterochromatin between the X chromosome and the autosome fused to the X prevented the spread of facultative heterochromatinization of the inactive X to an autosome. Therefore, UBA1Y would be translocated to the centromeric heterochromatin of the X chromosome and would then be involved in this amplification related to the autosomal fusion in T. muenninki. The order and direction of the five Y-linked genes determined by BAC clone sequencing of T. muenninki revealed that the genome structure of the ancestral Y euchromatic region was highly conserved between T. muenninki and mouse, except for SRY (Fig. 4), suggesting that we identified the ancestral genomic sequences of three multiple-copy genes (EIF2S3Y, TSPY, and USP9Y) in the Y euchromatic region of T. muenninki. We observed the
loss of five exons in *EIF2S3Y* and the presence of an internal stop codon in *USP9Y* (Fig. 4, Fig. S1). The deletion in *EIF2S3Y* and a missense mutation in *USP9Y* might occur in each original sequence after duplication. The predicted transcript sequences inferred from the genomic sequence of *TSPY* as well as *EIF2S3Y* and *USP9Y* were different from the actual transcript sequences (Fig. S1), suggesting that the amplified copies of these genes were functional, unlike the original sequences. Thus, our results provide evidence for compensatory effects of the amplified gene copies.

We showed that the *SRY* surrounding region was highly conserved between *T. muenninki* and other rodents, although *SRY25* was a nonfunctional pseudogene with the internal stop codon (Fig. S1q). There is a possibility that *SRY* could be duplicated on the Y chromosome to compensate its function like as other Y-linked genes. However, it is suggested that *SRY* has not retained sex-determining functions (Murata et al. 2010; Kimura et al. 2014).

The 5’ and 3’ flanking region of *SRY25* showed high similarity to mouse and rat (Fig. 5), suggesting that the gene was duplicated by not transposition but translocation. The 5’ flanking region of *SRY25* was segmented by LINE insertion (Fig. 5a yellow colored), indicating that this LINE insertion occurred after *SRY* duplication. In contrast, a LTR were found at the distal end of a conserved *SRY* unit (Fig 5b, pink colored). Furthermore, the downstream of this LTR showed high homology to the 5’ flanking region of mouse *Eif2S3y* (Fig. 5b, black colored). These observations indicated that the LTR was boundary for *SRY* duplication unit and the duplication occurred via these repetitive sequences. Several genetic mechanisms can determine the numbers of repeating units in tandem arrays, including gene amplification by processes such as replication slippage, rolling circle amplification, and unequal exchange (Charlesworth et al. 1994). Because replication slippage and rolling circle amplification involve only a short stretch of DNA at a time, these processes probably play only minor roles in the amplification of very long repeat units (Charlesworth et al. 1994). By contrast, the expansion of repeat units may be strongly associated with unequal exchange; this pattern has been inferred based on a
sequence analysis of satellites consisting of higher-order repeat units (Cabot et al. 1993; Warburton et al. 1993). Furthermore, extensive accumulation of retrotransposable elements has been observed in the Y chromosomes of Mus species (Eicher et al. 1989). Actually, unequal sister chromatid exchange mediated by retrotransposable elements could have been involved in SRY amplification along the heterochromatic block of the X chromosome in Cabrera’s vole, Microtus cabrerae (Fernández et al. 2002; Marchal et al. 2008). Gene amplification by unequal exchange mediated by retrotransposable elements has also been described for human olfactory receptor genes (Glusman et al. 2000), the primate γ-globin gene (Fitch et al. 1991), and mouse zinc-finger protein genes (Chen et al. 2002). Therefore, the amplification process for the Y-linked genes of T. muenninki can be explained by the same mechanism mediated by retrotransposable elements. In support of this hypothesis, we did not detect an intronless sequence generated by retrotransposition in the six multiple Y-linked genes (ZFy, UBA1Y, EIF2S3Y, TSPY, USP9Y, and RBMY) by PCR using the genomic DNA of T. muenninki (data not shown).

Our results revealed that the seven Y-linked genes were translocated to the X and Y chromosomes fused to an autosomal pair and were subsequently amplified in T. muenninki. Additionally, we showed that the amplified copies of the Y-linked genes are functional, unlike the original genes in the ancestral Y euchromatic region. The rescue of Y-linked genes via transposition to other chromosomes is widespread among mammals (Hughes et al. 2015). In the common ancestor of the two XO/XO species, T. osimensis and T. tokunoshimensis, the Y chromosome was lost after a small euchromatic region containing essential Y-linked genes was translocated to the X chromosome and a few Y-linked gene transposed to autosomes (Kuroiwa et al. 2010; Fig. 6). In the lineage of the T. muenninki, the Y chromosome was likely rescued from the risk of loss by sex-autosome fusions (Murata et al. 2012; Fig. 6). Nonetheless, the Y-linked genes were specifically translocated to the other region of the Y chromosome in the T. muenninki lineage, suggesting that the translocation was necessary to rescue Y-linked genes.
from progressive degeneration in the ancestral Y region. As the Y chromosome diverged from
the X chromosome, the correction of replication errors by homologous recombination was not
possible, leading to the accumulation of deleterious mutations in Y-linked genes (Bachtrog
2013). Therefore, it is hypothesized that becoming multiple copied genes is one strategy to
survive on the non-recombining Y (Bellott et al. 2014). Also, gene duplication coupled with
gene conversion between duplicates can potentially maintain gene function by counteracting
evolutionary Y decay (Rozen et al. 2003; Charlesworth 2003; Noordam and Repping 2006). A
combination of analytical and computer simulation methods also showed that gene conversion
effectively maintains the functionality of Y-linked duplicates that have already become fixed,
although it has little impact on the probability that duplicates become fixed within a population
(Connallon and Clark 2010). Therefore, the functions of *T. muenninki* Y-linked genes may have
been maintained not only by translocation, but also by amplification on the sex chromosomes
fused to an autosomal pair.
Figure Legends

Figure 1. Expression patterns of eight Y-linked genes in six tissue types from an adult male of *Tokudaia muenninki*.

*ACTB* was used as a positive control. N.C.: negative control. Marker: AccuRuler 1 kb DNA RTU Ladder (Maestrogen).

Figure 2. FISH mapping of six Y-linked genes on *T. muenninki* male chromosomes.

All six genes, *ZFY* (a–c), *UBA1Y* (d–f), *EIF2S3Y* (g–i), *USP9Y* (j–l), *TSPY* (m–o), and *RBMY* (p–r), were mapped along the long arm of the Y chromosome. The hybridization signals are indicated by arrowheads. Propidium iodide-stained R- (a, b, d, e, g, h, j, k, m, n, p, q) and Hoechst G-banding patterns (c, f, i, l, o, r) are shown. The scale bars represent 10 µm. (s) A summary of the Y-linked gene locations on the sex chromosomes. *: Murata et al. 2012; **: Murata et al. 2010.

Figure 3. Quantification of the copy numbers of ancestral Y-linked genes in *T. muenninki*.

The copy number of *UTY* (a) or *DDX3Y* (b) was set to 1.0, and the relative values for the other eight genes were estimated in five males by qPCR. The graph shows the means ± SD.

Figure 4. Comparison of the Y chromosome region of mouse (x-axis) to the ancestral Y chromosome region of *T. muenninki* (y-axis).

(a) Dot-plot alignments of a 391 kb region of mouse (NC_000087: 960,000–1,350,695) to a 372.4 kb region of *T. muenninki* (LC066213: 1–372,407). (b) Dot-plot alignments of an 8.8 kb region of mouse (NC_000087: 2,655,600–2,664,400) to a 15.9 kb region of *T. muenninki* (LC066213: 14,001–29,875). These alignments were generated using a mummerplot script and gnuplot. Aligned segments are represented as dots or lines in the NUCmer alignment. The positions of the genes on the mouse Y chromosome are shown on the dot-plot alignments. (c)
Comparison of Y gene order between the mouse and *T. muenninki*. *SRY* was located approximately 1.2 Mb downstream of *USP9Y* in the mouse genome.

**Figure 5. Sequence conservation of the regions upstream (a) and downstream (b) of SRY among *T. muenninki*, mouse, and rat.**

Percentages between bars indicate sequence similarity between the segments. Species-specific regions, except for retrotransposable elements, are colored white. TMU: *T. muenninki*.

**Figure 6. Evolution of the X and Y chromosomes and genes in genus *Tokudaia*.**

The lineage of *T. muenninki* was firstly diverged in genus *Tokudaia* (Murata et al. 2010). Sex-autosome fusion (Murata et al. 2012), Y-to-Y and Y-to-X gene translocations occurred in the lineage. In the common ancestor of *T. osimensis* and *T. tokunoshimensis*, a small euchromatic region of Y chromosome translocated to a distal end of X chromosome, and several Y genes transposed to autosome (Kuroiwa et al. 2010).
Supplementary Figure Legends

Figure S1. Alignment of the nucleotide and amino acid sequences of nine Y-linked genes between *T. muenninki* and other rodents.

The coding region from the mRNA and/or genomic sequences of *ZF*Y (a, b), *UBA1*Y (c, d), *EIF2S*3Y (e, f), *TSPY* (g, h), *UT*Y (i, j), *DDX3*Y (k, l), *USP9*Y (m, n), *RBMY* (o, p), and *SRY* (q) in *T. muenninki* were compared with orthologous sequences in mouse, rat, *T. osimensis*, and/or *A. sylvaticus*. Identical residues are indicated by dots. TMU: *T. muenninki*. The deletion and the internal stop codon are colored in gray (a) and pink (m), respectively. Asterisks show the stop codon sites in the amino acid sequences (b, d, f, h, j, i, l, n, p). Single nucleotide and amino acid polymorphisms within *T. muenninki* transcripts of each gene are shown in red letters (e–h, m–p). Different nucleotide sites of genomic sequences from the mRNAs of *T. muenninki* are shown in blue letters (e, g, m, o). (q) The start codons and the stop codons are colored in blue. *SRY*25 is the genomic sequence observed upstream of *EIF2S*3Y in this study. *SRY*1–24 sequences of *T. muenninki* are reported by Murata et al. 2010, of which only three (*SRY*1–3) have the conserved coding sequences. The repeat sequence was partially skipped in the mouse (600 bp), and the skipped site is shown by double slashes.

Figure S2. Neighbor-joining trees constructed with each X/Y gene sequence of *T. muenninki* and other rodents.

Number above and below branches refer to bootstrap percentages. The values less than 70% are not shown. TMU: *T. muenninki*; MMU: *M. musculus*; RNO: *R. norvegicus*; ASY: *A. sylvaticus*

Figure S3. Chromosomal location of two X-linked genes in *T. muenninki*.

A fluorescent in situ hybridization (FISH) analysis showed that *ZFX* (a, b) and *UBA1* (c, d) are distributed on the middle and pericentromeric regions of Xq, respectively. The hybridization signals are indicated by arrowheads. R- (a, c) and G-banding (b, d) patterns are shown. The
scale bars represent 10 µm.

**Figure S4. Nucleotide alignment of the accepter splicing site within intron 15 of DDX3Y between T. muenninki and mouse.**

The intronic sites are colored in gray for each species. Arrows show the boundary between an exon and intron. Asterisks show the identical sites between two sequences.

**Figure S5. Nucleotide sequence alignment in the conserved upstream (a) and downstream (b) regions of SRY among T. muenninki, mouse, and rat.**

Identical residues are indicated by dots. The skipped site is shown by double slashes. TMU: T. muenninki.
Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors thank F. Yamada, T. Jogahara, C. Koshimoto, T. Mitani, and K. Nakata for their efforts in capturing the animals and the Fujii Memorial Institute of Medical Sciences, Tokushima University, for support with fluorescence microscopy. This work was supported by JSPS KAKENHI Grant Numbers 23870023 and 26870410.

Authors’ Contributions

CM conceived of and designed the study, performed molecular and cytogenetic experiments and all data analysis, and drafted the manuscript. YK conducted RNA-seq. II participated in total RNA and BAC DNA extraction and commented on the manuscript. AK conceived of and designed the study, and participated in drafting the manuscript. All authors read and approved of the final manuscript.
References


organization of the \textit{RBMY} genes on the human Y chromosome: transposition and amplification of an ancestral autosomal \textit{hnRNPG} gene. Genomics 49:283-289


mouse Y chromosome is composed of retroviral-related, Y-enriched and Y-specific sequences. Genetics 122:181-92


Honda T, Suzuki H, Itoh M (1977) An unusual sex chromosome constitution found in the

country-rats, *Tokudaia osimensis osimensis* obtained from two neighbouring islands. Jpn.
J. Genet. 53:297-299

Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, Dugan S, Ding Y,
followed rapid gene loss on human and rhesus Y chromosomes. Nature 483:82-86

chromosome-to-autosome transposition events counter Y-chromosome gene loss in
mammals. Genome Biol 16:104.

Sry in species of the vole *Ellobius*. Nat Genet 11:117-118

Kimura R, Murata C, Kuroki Y, Kuroiwa A (2014) Mutations in the testis-specific enhancer of
*SOX9* in the *SRY* independent sex-determining mechanism in the genus *Tokudaia*. PLoS
One 9:e108779

repositioning in the X chromosome of XO/XO mammals, Ryukyu spiny rat. Chromosome


Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and


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 35:243-255
<table>
<thead>
<tr>
<th>Gene</th>
<th>Kidney</th>
<th>Lung</th>
<th>Spleen</th>
<th>Brain</th>
<th>Liver</th>
<th>Testis</th>
<th>N.C.</th>
<th>Marker (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZFY</td>
<td><img src="image1.png" alt="Image of ZFY" /></td>
<td><img src="image2.png" alt="Image of ZFY" /></td>
<td><img src="image3.png" alt="Image of ZFY" /></td>
<td><img src="image4.png" alt="Image of ZFY" /></td>
<td><img src="image5.png" alt="Image of ZFY" /></td>
<td><img src="image6.png" alt="Image of ZFY" /></td>
<td><img src="image7.png" alt="Image of ZFY" /></td>
<td>3,000</td>
</tr>
<tr>
<td>UBA1Y</td>
<td><img src="image8.png" alt="Image of UBA1Y" /></td>
<td><img src="image9.png" alt="Image of UBA1Y" /></td>
<td><img src="image10.png" alt="Image of UBA1Y" /></td>
<td><img src="image11.png" alt="Image of UBA1Y" /></td>
<td><img src="image12.png" alt="Image of UBA1Y" /></td>
<td><img src="image13.png" alt="Image of UBA1Y" /></td>
<td><img src="image14.png" alt="Image of UBA1Y" /></td>
<td>3,000, 1,500</td>
</tr>
<tr>
<td>EIF2S3Y</td>
<td><img src="image15.png" alt="Image of EIF2S3Y" /></td>
<td><img src="image16.png" alt="Image of EIF2S3Y" /></td>
<td><img src="image17.png" alt="Image of EIF2S3Y" /></td>
<td><img src="image18.png" alt="Image of EIF2S3Y" /></td>
<td><img src="image19.png" alt="Image of EIF2S3Y" /></td>
<td><img src="image20.png" alt="Image of EIF2S3Y" /></td>
<td><img src="image21.png" alt="Image of EIF2S3Y" /></td>
<td>1,500, 1,000</td>
</tr>
<tr>
<td>TSPY</td>
<td><img src="image22.png" alt="Image of TSPY" /></td>
<td><img src="image23.png" alt="Image of TSPY" /></td>
<td><img src="image24.png" alt="Image of TSPY" /></td>
<td><img src="image25.png" alt="Image of TSPY" /></td>
<td><img src="image26.png" alt="Image of TSPY" /></td>
<td><img src="image27.png" alt="Image of TSPY" /></td>
<td><img src="image28.png" alt="Image of TSPY" /></td>
<td>750, 500</td>
</tr>
<tr>
<td>UTY</td>
<td><img src="image29.png" alt="Image of UTY" /></td>
<td><img src="image30.png" alt="Image of UTY" /></td>
<td><img src="image31.png" alt="Image of UTY" /></td>
<td><img src="image32.png" alt="Image of UTY" /></td>
<td><img src="image33.png" alt="Image of UTY" /></td>
<td><img src="image34.png" alt="Image of UTY" /></td>
<td><img src="image35.png" alt="Image of UTY" /></td>
<td>3,000, 1,500</td>
</tr>
<tr>
<td>DDX3Y</td>
<td><img src="image36.png" alt="Image of DDX3Y" /></td>
<td><img src="image37.png" alt="Image of DDX3Y" /></td>
<td><img src="image38.png" alt="Image of DDX3Y" /></td>
<td><img src="image39.png" alt="Image of DDX3Y" /></td>
<td><img src="image40.png" alt="Image of DDX3Y" /></td>
<td><img src="image41.png" alt="Image of DDX3Y" /></td>
<td><img src="image42.png" alt="Image of DDX3Y" /></td>
<td>3,000, 1,500</td>
</tr>
<tr>
<td>USP9Y</td>
<td><img src="image43.png" alt="Image of USP9Y" /></td>
<td><img src="image44.png" alt="Image of USP9Y" /></td>
<td><img src="image45.png" alt="Image of USP9Y" /></td>
<td><img src="image46.png" alt="Image of USP9Y" /></td>
<td><img src="image47.png" alt="Image of USP9Y" /></td>
<td><img src="image48.png" alt="Image of USP9Y" /></td>
<td><img src="image49.png" alt="Image of USP9Y" /></td>
<td>3,000</td>
</tr>
<tr>
<td>RBMY</td>
<td><img src="image50.png" alt="Image of RBMY" /></td>
<td><img src="image51.png" alt="Image of RBMY" /></td>
<td><img src="image52.png" alt="Image of RBMY" /></td>
<td><img src="image53.png" alt="Image of RBMY" /></td>
<td><img src="image54.png" alt="Image of RBMY" /></td>
<td><img src="image55.png" alt="Image of RBMY" /></td>
<td><img src="image56.png" alt="Image of RBMY" /></td>
<td>1,000</td>
</tr>
<tr>
<td>ACTB</td>
<td><img src="image57.png" alt="Image of ACTB" /></td>
<td><img src="image58.png" alt="Image of ACTB" /></td>
<td><img src="image59.png" alt="Image of ACTB" /></td>
<td><img src="image60.png" alt="Image of ACTB" /></td>
<td><img src="image61.png" alt="Image of ACTB" /></td>
<td><img src="image62.png" alt="Image of ACTB" /></td>
<td><img src="image63.png" alt="Image of ACTB" /></td>
<td>750</td>
</tr>
</tbody>
</table>
(bp)

mouse Sry

TMU SRY25

rat SRY

(bp)

mouse Sry

TMU SRY25

rat SRY

: LTR  : LINE  : SINE  : unknown repeat  : homologous sequences between species

: the region with high homology to mouse Eif2s3y upstream
Y-to-X euchromatic region translocation

Y-to-autosome gene transposition

Y-loss SRY-loss

Sex-autosome fusion

Y-to-Y Y-to-X gene translocations

Amami spiny rat
*T. osimensis*
2n=25, XO/XO

Tokunoshima spiny rat
*T. tokunoshimensis*
2n=45, XO/XO

Okinawa spiny rat
*T. muenninki*
2n=44, XX/XY

Transposed Y gene

Proto-Y (Translocated Y region)

Autosome X

Neo-X

Neo-Y

Ancestral-Y

Heterochromatin
<table>
<thead>
<tr>
<th>USPY-2mRNA</th>
<th>USPY-1mRNA</th>
<th>USPY-1mRNA</th>
<th>USPY-1mRNA</th>
<th>USPY-1mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>USPY_mRNA</td>
<td>USPY_genome</td>
<td>USPY_genome</td>
<td>USPY_genome</td>
<td>USPY_genome</td>
</tr>
<tr>
<td>USPY_mouse</td>
<td>USPY_tat</td>
<td>USPY_tat</td>
<td>USPY_tat</td>
<td>USPY_tat</td>
</tr>
<tr>
<td>USPY_tat</td>
<td>USPY_mRNA</td>
<td>USPY_mRNA</td>
<td>USPY_mRNA</td>
<td>USPY_mRNA</td>
</tr>
<tr>
<td>USPY_tst</td>
<td>USPY_sst</td>
<td>USPY_sst</td>
<td>USPY_sst</td>
<td>USPY_sst</td>
</tr>
<tr>
<td>USPY_qm</td>
<td>USPY_qm</td>
<td>USPY_qm</td>
<td>USPY_qm</td>
<td>USPY_qm</td>
</tr>
<tr>
<td>USPY_sst</td>
<td>USPY_sst</td>
<td>USPY_qm</td>
<td>USPY_qm</td>
<td>USPY_qm</td>
</tr>
</tbody>
</table>

**Note:**
- The above table represents the nucleotide sequences for the USPY-2mRNA and USPY-1mRNA constructs.
- The sequences are aligned to show the positions of the different components (e.g., USPY_tst, USPY_mRNA).
- The sequences are marked with specific symbols to denote internal codons and stop codons.
- The overall structure highlights the integration of different segments within the USPY constructs.

### Additional Notes:
- The USPY constructs are designed to study specific biological processes or for therapeutic applications.
- The sequences indicate the presence of regulatory elements and coding regions.
- The alignment helps in understanding the organization and potential functionalities of these constructs.
SRY25  CTCCTAAGGCTGTAGAGGCCTCTAGCA- ---TACGCA TAGCAGAGC ACGCTAGAGA ACGCTAGAGA TAGCCACTCT CATACACTG
SRY25  ......... ......... ......... .........  ------  -------  -------  -------  -------  -------  -------  -------
SRY6  ......... ......... ......... .........  ------  -------  -------  -------  -------  -------  -------  -------
SRY15  ......... .........  ..T......  ------  -------  -------  -------  -------  -------  -------  -------
SRY20  ......... .........  ..T......  ..C......  ------  -------  -------  -------  -------  -------  -------
SRY9  ......... ......... ......... .........  ------  -------  -------  -------  -------  -------  -------
SRY12  ......... ......... ......... .........  ------  -------  -------  -------  -------  -------  -------
SRY13  ......... ......... ......... .........  ------  -------  -------  -------  -------  -------  -------
SRY23  ......... ......... ......... .........  ------  -------  -------  -------  -------  -------  -------
SRY24  ......... ......... ......... .........  ------  -------  -------  -------  -------  -------  -------
SRY11  .........  A  ...........  ...........  ------  -------  -------  -------  -------  -------  -------
SRY18  .........  A  ...........  ...........  ------  -------  -------  -------  -------  -------  -------
SRY17  .........  A  ...........  ...........  ------  -------  -------  -------  -------  -------  -------
SRY14  .........  .........  .........  .........  ------  -------  -------  -------  -------  -------  -------
SRY14  .........  .........  .........  .........  ------  -------  -------  -------  -------  -------  -------
SRY7  .........  .........  .........  .........  ------  -------  -------  -------  -------  -------  -------
SRY1  .........  .........  .........  .........  ------  -------  -------  -------  -------  -------  -------
SRY2  .........  .........  .........  .........  ------  -------  -------  -------  -------  -------  -------
SRY3  .........  .........  .........  .........  ------  -------  -------  -------  -------  -------  -------
SRY8  .........  .........  .........  .........  ------  -------  -------  -------  -------  -------  -------
SRY10  .........  .........  .........  .........  ------  -------  -------  -------  -------  -------  -------
Sry-mouse  G..T......  A......  ...........  ------  -------  -------  -------  -------  -------  -------
SRY25  GCTCTGA---
SRY5  .........---
SRY6  .........---
SRY15  .........---
SRY20  .........---
SRY9  .........---
SRY12  .........---
SRY3  .........---
SRY24  .........---
SRY11  .........---
SRY16  .........---
SRY17  .........---
SRY21  .........---
SRY14  .........---
SRY19  .........---
SRY18  .........---
SRY22  .........---
SRY1  .........---
SRY2  .........---
SRY3  .........---
SRY6  .........---
SRY10  .........---
Sry-mouse  .......T.T---
Sry-rat  .TCTCTCC
<table>
<thead>
<tr>
<th>Genes</th>
<th>Accession No.</th>
<th>T. muenninki</th>
<th>M. musculus</th>
<th>R. norvegicus</th>
<th>T. osimensis</th>
<th>A. sylvaticus</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZFY</td>
<td>LC066211</td>
<td>NM_009570</td>
<td>-</td>
<td>E14275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UBA1Y</td>
<td>LC066205</td>
<td>NM_011667</td>
<td>XM_008773685</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIF2S3Y</td>
<td>LC066199</td>
<td>NM_012011</td>
<td>NM_001167666</td>
<td>AB441819</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSPY</td>
<td>LC066203</td>
<td>NR_027507</td>
<td>NM_022923</td>
<td>-</td>
<td>AF205112</td>
<td></td>
</tr>
<tr>
<td>UTY</td>
<td>LC066209</td>
<td>NM_009484</td>
<td>XM_008773696</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDX3Y</td>
<td>LC066197</td>
<td>NM_012008</td>
<td>NM_001167665</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP9Y</td>
<td>LC066207</td>
<td>NM_148943</td>
<td>XM_008773701</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBMY</td>
<td>LC066201</td>
<td>NM_001166384</td>
<td>GATN01000006</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZFX</td>
<td>LC066212</td>
<td>NM_001044386</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UBA1</td>
<td>LC066206</td>
<td>NM_001136085</td>
<td>NM_001014080</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIF2S3X</td>
<td>LC066200</td>
<td>NM_012010</td>
<td>NM_001100542</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSPYL2</td>
<td>LC066204</td>
<td>NM_029836</td>
<td>NM_001191618</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KDM6A/UTX</td>
<td>LC066210</td>
<td>NM_009483</td>
<td>XM_008773047</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDX3X</td>
<td>LC066198</td>
<td>NM_010028</td>
<td>NM_001108246</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP9X</td>
<td>LC066208</td>
<td>NM_009481</td>
<td>NM_001135893</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBMX</td>
<td>LC066202</td>
<td>NM_011252</td>
<td>NM_001025663</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table. S2 Primer pairs for RT-PCR and qPCR.

<table>
<thead>
<tr>
<th>Genes</th>
<th>No.</th>
<th>Primers for RT-PCR</th>
<th>Primers for qPCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZFY</td>
<td>*</td>
<td>1 GTCAGGAGCTGACTTAAGAAGA AAGTCATCACTGATCCTGACGACAGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 GCTTCAAAATCACAATCCCTA CCAGGAGCATCCCAATAGAAGA</td>
<td></td>
</tr>
<tr>
<td>UBA1Y</td>
<td>**</td>
<td>1 GGAGGAAGACAAACTGACAGA TGGAGAGACACTGACAGCTGAAGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 AGGTCTTTAAGACAGCAGA CAAGGAGCATCCCAATAGAAGA</td>
<td></td>
</tr>
<tr>
<td>EIF2S3Y</td>
<td>**</td>
<td>1 CTTTCGAGACGACGATGCA TCTGACACCTGACAGCTGAAGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 CCAGGATCCAGGAAATGGA GCTGACACCTGACAGCTGAAGA</td>
<td></td>
</tr>
<tr>
<td>TSPY</td>
<td>**</td>
<td>1 GGGAAGAGACAGAACACAGA GGAGGAAGACACTGACAGCTGAAGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 ACTCTTTACTCTGGAATGGA GGGTTTAAATCCTGGAAGAGA</td>
<td></td>
</tr>
<tr>
<td>UTY</td>
<td></td>
<td>1 ACAGGTGTACGTCCGTAAGA ATGGAGACACTGACAGCTGAAGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 CAGAAGCTGCTGCCGTAAAGC TAAGGTCACTGACAGCTGAAGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 TACCAGTCCACTCAGTGATGAT —</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 TGGCGTCAACAGCAGGAAT —</td>
<td></td>
</tr>
<tr>
<td>DDX3Y</td>
<td>*</td>
<td>1 CAAAACGCTGTCAGTACAGA AGGAGACACTGACAGCTGAAGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 ACCAAACGAGAAGAGAGGTG TTAAAGTGTCGAGACTGACAGCTGAAGA</td>
<td></td>
</tr>
<tr>
<td>USP9Y</td>
<td></td>
<td>1 AGGAGTACTTATCTGTTGCTAAGT TGTACGCTGAGCCTAAGGACTGAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 AGCATAAGTGAAGACGCTTACCTCCAGCAGATCCGACGACGATCCGAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 CCAGATGGTGATGGAAGAAT —</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 TCTTCACTGCTTTTCCTTAGA —</td>
<td></td>
</tr>
<tr>
<td>RBMY</td>
<td>**</td>
<td>1 GACAGAAGAAAGCTGAGAACATG TGTACGCTGAGCCTAAGGACTGAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 CTTTATGACTCTGCTCAGA ACGTCCCAAGACTCCTGGAAGAGA</td>
<td></td>
</tr>
<tr>
<td>SRY</td>
<td></td>
<td>— AACACATGCCAATCAGGAGA —</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>— TGTCCACAGGGTGAAAATTAA —</td>
<td></td>
</tr>
<tr>
<td>ZFX</td>
<td>**</td>
<td>1 GCGTATAACTTGCCCAAGC —</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 GAGTACCCATGTAAGGAGGAGA —</td>
<td></td>
</tr>
<tr>
<td>UBA1X</td>
<td>**</td>
<td>1 GAGGCAAGAGGTGACGACAGA —</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 ACTCCTTCTCTGCTGACAGA —</td>
<td></td>
</tr>
<tr>
<td>ACTB</td>
<td>*</td>
<td>1 GCCATCTGACCCCTGAGTA —</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 TCTCAGCTGTTGCTGAGA —</td>
<td></td>
</tr>
</tbody>
</table>

*The primer pairs used for RT-PCR using multiple tissues to reveal the expression profiles.
**The primer pairs used for FISH analysis.
Table. S3 Sequence homology of *T. muenninki* X-linked genes with the orthologues of mouse

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Homology of nucleotide with mouse (%)</th>
<th>Homology of amino acid with mouse (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZFX</td>
<td>94.6</td>
<td>95.4</td>
</tr>
<tr>
<td>UBA1</td>
<td>96.2</td>
<td>97.6</td>
</tr>
<tr>
<td>EIF2S3X</td>
<td>96.2</td>
<td>99.4</td>
</tr>
<tr>
<td>TSPLY2</td>
<td>94.9</td>
<td>93.4</td>
</tr>
<tr>
<td>KDM6A/UTX</td>
<td>97.6</td>
<td>96.2</td>
</tr>
<tr>
<td>DDX3X</td>
<td>97.2</td>
<td>99.4</td>
</tr>
<tr>
<td>USP9X</td>
<td>97.6</td>
<td>99.4</td>
</tr>
<tr>
<td>RBMX</td>
<td>91.8</td>
<td>96.9</td>
</tr>
</tbody>
</table>