Journal of Genetics (2019) 98:41
https://doi.org/10.1007/s12041-019-1091-4

RESEARCH ARTICLE

Identification and expression profiling of MSY genes of yak for bull fertility

PARTHA PRATIM DAS[1], G. KRISHNAN[2], JUWAR DOLEY[1], TAPAS KUMAR BISWAS[3], VIJAY PAUL[1], PRITHIVIRAJ CHAKRAVARTY[1], SITANGSHU MOHAN DEB[4] and PRANAB JYOTI DAS[1,5]*

[1]*Indian Council of Agricultural Research-National Research Centre on Yak, West Kameng 790 101, India*
[2]*Indian Council of Agricultural Research-National Institute of Animal Nutrition and Physiology, Bengaluru 560 030, India*
[3]*Indian Council of Agricultural Research-Indian Veterinary Research Institute Eastern Regional Station, 37, Belgachia Rd, LIG Housing, MIG Housing, Tala, Kolkata 700 037, India*
[4]*Indian Council of Agricultural Research-National Dairy Research Institute, Karnal 132 001, India*
[5]*Indian Council of Agricultural Research-National Research Centre on Pig, Rani, Guwahati 790 101, India*
*For correspondence. E-mail: drpranabjyotidas@gmail.com.

Received 1 November 2018; revised 27 December 2018; accepted 10 January 2019; published online 4 May 2019

Abstract. Yak (*Bos grunniens*) is a unique bovine species and considered as lifeline of highlanders. The male subfertility in yak is a matter of concern that causes huge economic loses. The spermatogenesis and male reproduction machinery are critically governed by Y-linked genes which tend to acquire necessary information in the course of evolution. The Y-linked fertility genes are present in multiple copies with testis-limited expression. To understand this novel complexity, 12 male-specific region of Y chromosome (MSY) genes have been studied in the yak. Targeted genes are amplified in male and female genomic DNA and confirmed the male-derived specificity. Moreover, testis and sperm-specific expressions of MSY genes are distinct among different tissues. The quantitative polymerase chain reaction results validate the expression pattern of these genes in various tissues with predominant expression in testis and sperm. The sequencing of resultant yak MSY genes gives significant result and shows similarity with cattle (*Bos indicus*), but few nucleotide mismatches define the proposition of infertile male in the F_1 hybrid of cattle and yak. The identified MSY genes can be used to establish male-specific characteristics and to differentiate male and female yak genotypically. Further, these genes may act as valuable resources to understand the capacity of spermatogenesis, embryogenesis, cellular growth, azoospermia and male subfertility in the yak.

Keywords. yak; Y-chromosome; male-specific region of Y chromosome gene; male-fertility; testis; spermatozoa.

Introduction

Yak (*Bos grunniens*) is a multipurpose bovidae of high altitude which can thrive in the extreme cold up to −40°C and yak husbandry is considered to be the lifeline of highlanders. Yak provides milk, meat, wool and fuel, and serves as a pack animal to the people living in high-altitude areas, mainly of the central Himalayan region (Krishnan *et al.* 2009). It possesses the same number of chromosomes ($2n = 60$) like cattle, however, attains sexual maturity late and low fertility rate as compared to cattle (Lan *et al.* 2014). Indeed, their ability to survive in the extreme environmental conditions makes yak socially and economically an important animal. Infertility in yak causes huge economic loses to the highlanders where other livestock husbandry/agriculture is unfeasible. Since yaks are maintained under a migratory system of rearing, at lower altitude yak is crossed with hill cattle, which is a common practice in yak husbandry to enhance productivity. As a result, the F_1 hybrid females are highly productive but the males are infertile (Medugorac *et al.* 2017). Therefore, the enhancement of reproductive efficiency and fertility of the yak which is primarily governed by male-specific region of Y chromosome (MSY) genes is the foremost important for

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s12041-019-1091-4) contains supplementary material, which is available to authorized users.

Table 1. Different genes undertook for the present study, their location and characteristics.

	Locus symbol	Gene name	Location (human)	Characteristics
1	*SRY*	Sex determining region Y	Yp11.3	Protein coding gene
2	*TSPY*	Testis-specific protein Y-linked 1	Yp11.2	Protein coding gene
3	*TSPY4*	Testis-specific protein Y-linked 4	Yp11.2	Protein coding gene
4	*TSPY6P*	Testis-specific protein Y-linked 6	Yp11.2	Pseudogene
5	*DDX3Y*	DEAD(Asp-Glu-Ala-Asp) Box Helicase 3, Y-linked	Yq11.21	Protein coding gene
6	*UTY*	Ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	Yq11.221	Protein coding gene
7	*HSFY1*	Heat shocked transcription factor, Y-linked 1	Yq11.222	Protein coding gene
8	*HSFY2*	Heat shocked transcription factor, Y-linked 2	Yq11.222	Protein coding gene
9	*AMELY*	Amelogenin, Y-linked	Yp11.2	Protein coding gene
10	*USP9Y*	Ubiquitin-specific peptidase 9, Y-linked	Yq11.21	Protein coding gene
11	*TXLNG2P*	Taxilin gamma 2	Yq11.222	Pseudogene
12	*FAM197Y1*	Family with sequence similarity 197, Y-linked, Member1	Yp11.2	Protein coding gene

genetic improvement. Further, there is a dearth of information on the facts of genes governing the yak fertility and its functioning of molecular mechanisms.

The defects on Y chromosome are responsible for 10–25% of male infertility in mammals (Fincham and Simmer 2007). The Y chromosome is the smallest acrocentric, heterochromatic sex chromosome in mammals which contains master-switch genes responsible for gender differentiation (Waters *et al.* 2007; Bachtrog 2013; Tobler *et al.* 2017). It encompasses the male-specific region (MSY) or non-recombining region of Y chromosome (NRY) as well as the pseudoautosomal region (PAR) (Das *et al.* 2009; Paria *et al.* 2011). Approximately 95% of Y chromosome is male-specific, which does not recombine during meiosis and only the 5% PAR facilitates recombination with X chromosome (Das *et al.* 2010, 2013a).

The whole genome sequence is characterized in cattle and recently in yak (Elsik *et al.* 2009; Medugorac *et al.* 2017). However, the information on the Y chromosome of yak is very limited and inadequate, might be due to highly repetitive sequences, palindromic region and a higher rate of mutation that usually hamper fruitful sequencing. The present study is designed to characterize important male-specific genes and to evaluate their expression pattern in different tissues of yak. This study was undertaken to characterize and analyse 12 very important MSY genes, i.e. *SRY*, *TSPY*, *TSPY4*, *TSPY6P*, *FAM197Y1*, *USP9Y*, *UTY*, *DDX3Y*, *AMELY*, *TXLNG2P*, *HSFY1* and *HSF2Y* in the yak with the available sequence information and location of Y chromosome in the mammalian genome (table 1).

Materials and methods

Sample collection

Blood was collected from yak in a yak-breeding tract of Arunachal Pradesh of India in 0.5 M sterile EDTA. The tissue samples were collected from different organs of yak (both male and female) immediately after slaughter. Utmost care was taken to avoid RNase contamination, and tissues were sliced into small pieces which were dipped into RNAlater (Invitrogen, Carlsbad, USA) in the ratio of 1:3 volumes and stored at −80°C until use. Semen samples were also collected from healthy yak bulls by the artificial vagina method using the Missouri model (Das *et al.* 2013b). Further, the semen samples were processed and purified and stored at −80°C until further use.

DNA and RNA isolation

The genomic DNA was isolated from all the blood samples of both male and female yaks by the conventional phenol–chloroform method with a slight modification. The RNA was isolated from eight major tissues of yaks, namely liver, kidney, heart, spleen, lung, muscle, ovary and testis using the Trizol reagent (Das *et al.* 2010). The quality of DNA and RNA was checked using a spectrophotometer and agarose gel electrophoresis. The RNA from sperm was isolated using a published protocol with the help of a 27 gauge needle and Trizol (Das *et al.* 2010). Finally, cDNA was synthesized from RNA of different tissues in 20 μL reaction volume using 100 μM random primers and 200 U RevertAid reverse transcriptase (Invitrogen).

Amplification of yak MSY gene by polymerase chain reaction (PCR)

Primers of all the genes in the present study were designed using Primer3 software (details of the primer list are given in table 2). Amplification of the genes was carried out by PCR in a 25 μL of reaction volume containing 1× Dream *Taq* buffer, 2.5 U DNA polymerase (Invitrogen, Carlsbad, CA, USA), 0.2 mM dNTPs (Invitrogen), 25 pmol primers and 100 ng DNA. The reactions were carried out as follows: first denatured the DNA at 94°C for 5 min, repeat

Table 2. List of gene names and primer sequences with annealing temperature and product size.

Gene symbol	Forward primer	Reverse primer	T_m (°C)	Product size (bp) (genomic DNA)	Product size (bp) (RNA)
AMELY[a]	CTATGAACCCGTTGGTGGAT	ATTGAGTGTTGGCCAGGAAC	60	167	167
DDX3Y[a]	GTGAAGGTTTTGGCAGTCGT	TGTGCGGAGGACAGTTATTG	58	299	225
FAM197Y1[a]	CCTGCAGAAGACATCATGGA	CTTCAGCTCCTGGGAGGAC	58	78	78
HSFY1[a]	CAGACTTTGTCTGAGGATCTTGTG	TCTTTCACCTCATCTGTTTCAGAG	58	80	80
HSFY2[a]	CTTTGAGGACCTCTGTGATGG	AACGGAGGTTTCTGAACCAG	60	190	190
SRY[a]	GTCCAGCTGTGGTACAGCAA	AGCTGCTTGCTGATGTCTGA	58	232	232
TSPY[a]	GCACCTTCCAAGTTGTGAGC	CACCTCCTCCACGATGTCTT	58	254	254
TSPY4[a]	GGGGGCAGAAAAGTCTAGGT	GTGGAATGAGATGCGTGGTA	58	91	91
TSPY6P[a]	TACCACGCATCTCATTCCAC	CACCTCAGCAATCCAGTCAG	58	147	147
TXLNG2P[a]	TCAACTCCAGAGGAGAAGCTG	ATGGGCCTCACTCCCTTAAT	58	77	77
USP9Y[a]	TTGGGGTGAGCCTGTTAATC	AAGGAACCTCCCAGGACTTG	58	135	135
UTY[a]	CAGCAACCACATTCTGCTGT	CTTTCTGGCACCTCAAAACC	58	120	120
AR[b]	AGCAGCAACAGGAGACCAGT	TGCTTAAGCCTGGGAAAGTG	58	273	273
GAPDH[c]	CAAGGTCATCCATGACCACTTTG	GTCCACCACCCTGTTGCTGTAG	58	496	496
PRM2[c]	GGTCTACGGGAGGACTCACA	CCTCCTCCTCCTCATCCTTC	58	349	169
PTPRC[c]	CCACGGGTATTCAGCAAGTT	TTGATCCTGCATCTCCGTTT	58	228	136
TUBB2B[d]	TGTCCCTCGTGCTATCTTGGT	CACATCCAGGACCGAGTCAA	58	180	180
ACTB[d]	CGGCATCGAGGACAGGAT	CATCGTACTCCTGCTTGCTGAT	58	169	169
TBP[d]	CAGAGAGCTCCGGGATCGT	CACCATCTTCCCAGAACTGAATAT	58	194	194
GAPDH[d]	CCTGCCCGTTCGACAGATA	GGCGACGATGTCCACTTTG	58	150	150

The primer used in different experiments: [a]PCR, RT-PCR, sequencing and qRT-PCR; [b]PCR; [c]RT-PCR and [d]qRT-PCR.

a cycle of 94°C for 30 s, 58/60°C for 30 s, 72°C for 1 min for 32 cycles and a final extension of 5 min at 72°C in a thermal cycler.

Sequencing and phylogenetic analysis of MSY genes

Standard double-stranded sequencing reactions were performed using 50 ng of purified PCR product, 4 pmol of primer and a BigDye Terminator ready reaction kit (Applied Biosystems, Foster City, USA). Cycle sequencing was carried out in a GeneAmp9600 thermal cycler (Perkin Elmer, Waltham, USA) employing 30 cycles at 96°C for 10 s, 49−50°C for 5 s and 60°C for 4 min. Extended products were purified by alcohol precipitation followed by dissolving in Hi-Di formamide and analysing in an ABI3700 automated DNA Analyzer (Applied Biosystems).

The sequences of yak MSY genes were submitted to the NCBI databank and accession numbers were assigned subsequently. Phylogenetic and molecular evolutionary analyses were performed using MEGA v6 (Tamura *et al.* 2013). Consensus sequence and gap weight were analysed by using MegAlign of DNAstar (Clustal 1994). Distance metrics between the species were calculated with GeneBee methods and algorithm (Castresana 2000; Edgar 2004; Dereeper *et al.* 2008). Mismatch analysis values were estimated using 100% bootstrap values. These values were used to reconstruct the neighbour-joining tree by comparing different male-specific genes of yak and other species (Chevenet *et al.* 2006; Guindon *et al.* 2010).

Ethics statement

Procurement of blood and tissue samples was performed in accordance with the approval of the Institute Animal Ethics Committee of Indian Council of Agricultural Research-National Research Centre on Yak, Dirang, India. The approved animal use protocol number is 4(17)/NRCY/IAEC-02 dated 01.08.13.

Results

DNA and RNA isolation

A good quality DNA was isolated from whole blood following the recommended protocol, and subsequently, quality was checked using a spectrophotometer and 0.8% agarose gel. RNA was also isolated from eight different tissues and sperm of yak using a published protocol with a slight modification and quality was checked (Das *et al.* 2010). The purity of RNA was further checked by reverse transcriptase (RT)-PCR using the intron-spanning *PTPRC* (*CD45*) gene for liver, kidney, heart, spleen, lung, muscle, ovary and testis (figure 1) and intron-spanning *PRM2* and *PTPRC* (*CD45*) genes for sperm and testis (figure 2, a&b).

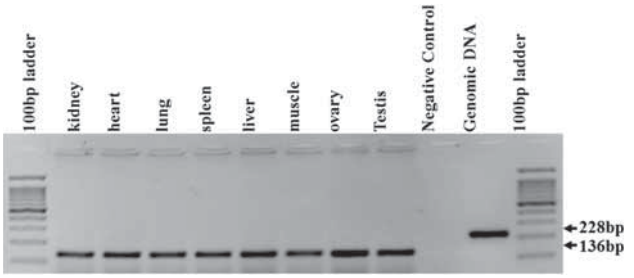

Figure 1. RT-PCR validation of RNA, isolated from different tissues of yak using the intron-spanning *PTPRC* (*CD45*) gene run on 2% agarose gel. No genomic band in the tissue panel indicates samples are free of DNA contamination.

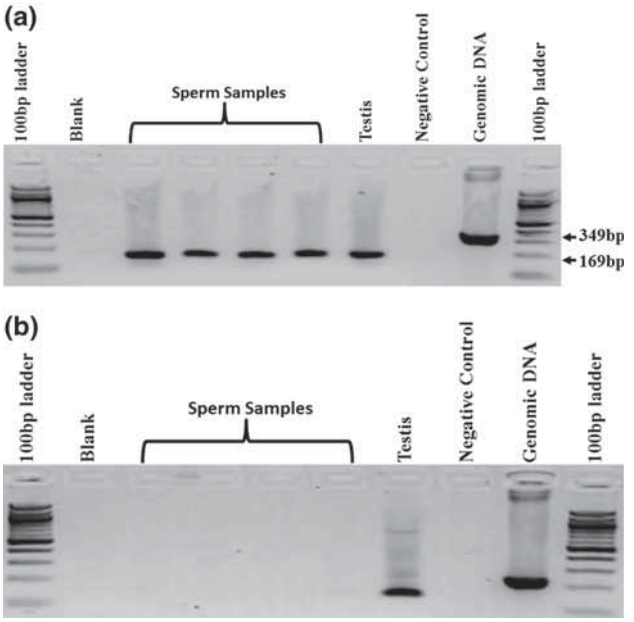

Figure 2. (a) RT-PCR validation of RNA, isolated from yak sperm and testis using the intron-spanning *PRM2* genes run on 2% agarose gel. No genomic band of sperm and testis indicates samples are free of DNA contamination. (b) RT-PCR validation of RNA, isolated from yak sperm and testis using the intron-spanning *PTPRC* (*CD45*) genes run on 2% agarose gel. No RNA and DNA bands in sperm indicate samples are free of DNA and RNA contamination.

Amplification and sequencing of yak MSY gene

All the 12 genes were amplified in male genomic DNA with no (*TSPY6P*, *FAM197Y1*, *HSFY1* and *HSF2Y*) or very faint (*SRY*, *TSPY*, *TSPY4*, *USP9Y*, *UTY*, *DDX3Y* and *AMELY*) bands in the female DNA of yak (figure 3). This result revealed that all these genes are very specific to the male genome which can be used to differentiate male and female DNA of yak. On the other hand, positive control *AR* gene, which is present in the X chromosome amplified in both male and female genomic DNA of yak (figure 3).

All the 12 genes were sequenced successfully by Sanger's dideoxy method and able to retrieve the expected sequences. The sequences were submitted to the NCBI databank and accession numbers were assigned subsequently (*USP9Y*–KR051078, *SRY*–KR263869, *DDX3Y*–KR422623, *UTY*–KR611085, *TSPY6P*–KT354500, *HSFY2*–KT354501, *AMELY*–KT354502, *TSPY*–MG254720, *TSPY4*–MG254721, *FAM19Y1*–MG270064, *TXLNG2P*–MG270065 and *HSFY1*–MG270066).

Phylogenetic and sequencing data analysis

Phylogenetic analysis has been carried out for all yak MSY genes with the available sequence of other eutherian mammals and the orthologous of these genes was confirmed in other mammals as well (table 3).

Phylogenetic analysis has shown 100% sequence similarity of yak *SRY*, *USP9Y*, *HSF2Y*, *UTY* and *AMELY* genes with cattle (figures 1, a&b; 3, a&b; 4, a&b; 5, a&b; 6, a&b; 5b and tables 1, 3, 4, 5, 6 in electronic supplementary material at http://www.ias.ac.in/jgenet/). However, a few interesting mismatches in yak and cattle were detected in the *DDX3Y* and *TSPY6P* genes with a sequence identity of 99.3 and 98.6%, respectively (figure 2b and 7b and tables 2 and 7 in electronic supplementary material). The sequenced yak DDX3Y gene showed mismatch with cattle at the positions C.106A>G and C.182G>A (figure 2, a&c in electronic supplementary material). Further, the yak *TSPY6P* gene showed mismatch with cattle at the positions A.48C>T and A.57A>T (figure 7, a&c in electronic supplementary material).

Expression profiling of yak MSY genes

The *PRM2* gene is expressed at a higher level only during spermatogenesis which can be used as a positive marker for spermatozoa, and *PTPRC*, a cell surface protein-encoding gene served as a positive marker for all the body tissues except for spermatozoa. The introns spanning primers of *PRM2* and *PTPRC* give different product sizes for DNA and RNA, which facilitated detection of genomic DNA contamination of RNA as well as other cell RNA contamination in sperm (figures 1 and 2, a&b). The expression profiles of all 12 MSY genes were examined in cDNA of different tissue panels (figure 4) and sperm (figure 5). The RT-PCR result confirmed that yak MSY genes were expressed predominantly in testis and sperm (figures 4 and 5) and a few were expressed ubiquitously. The *SRY* gene does not express in any of the tissues except in sperm. Another male-specific gene *AMELY*, not expressed in any of the tissues including sperm which is known to express only in the enamel of teeth (Salido *et al.* 1992; Fincham and Simmer 2007). The MSY genes *TSPY*, *TSPY4* and *TSPY6P* were uniquely expressed in testis and sperm whereas *TXLNG2P* and *USP9Y* were expressed in testis, sperm as well as in spleen. However, *DDX3Y*, *FAM197Y1* and *UTY* were expressed in all the major tissues except in

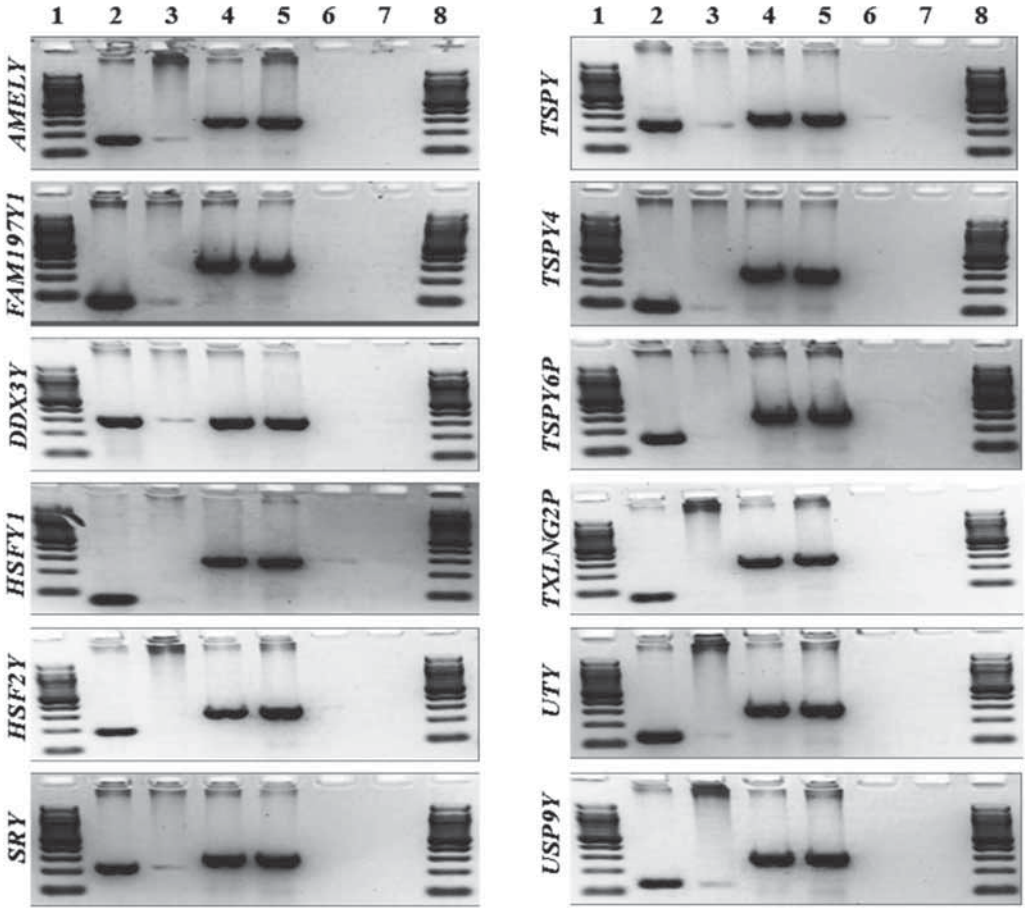

Figure 3. Electrophoretic analysis of MSY genes amplified in male genomic DNA and not in female genomic DNA (lanes 1 and 8, 100-bp ladder; lane 2, male genomic DNA; lane 3, female genomic DNA; lane 4, positive control (*AR*) with male genomic DNA; lane 5, positive control (*AR*) with female genomic DNA; lanes 6 and 7, negative control).

Table 3. Orthologues of different genes present in different species.

Gene symbol	Yak	Cattle	Human	Horse	Dog	Pig
SRY	+	+	+	+	+	+
TSPY	+	+	+	+	+	+
TSPY4	NR	+	+	NR	+	+
TSPY6P	NR	+	+	NR	+	NR
FAM197Y1	NR	NR	+	NR	NR	NR
USP9Y	NR	+	+	+	+	+
UTY	NR	+	+	+	+	+
DDX3Y	+	+	+	+	+	+
AMELY	NR	+	+	+	+	+
TXLNG2P	NR	NR	+	NR	NR	NR
HSFY1	NR	+	+	+	+	+
HSFY2	NR	+	+	+	+	NR

NR, not reported.

muscle and ovary. The *HSFY1* gene does not express in any of the tissues but the *HSF2Y* gene expressed predominantly in testis and sperm. Further, *DDX3Y* has shown a different band size in different tissues (figure 4). Of these 12 genes, 10 were expressed predominantly in testis and sperm (figures 4 and 5) with varying intensity.

Quantification of yak MSY gene by qRT-PCR

qRT-PCR was used to validate qualitative RT-PCR results as well as to examine MSY genes copy number variation in different tissues using yak MSY genes. The yak MSY genes *SRY*, *TSPY*, *TSPY4*, *TSPY6P*, *FAM197Y1*,

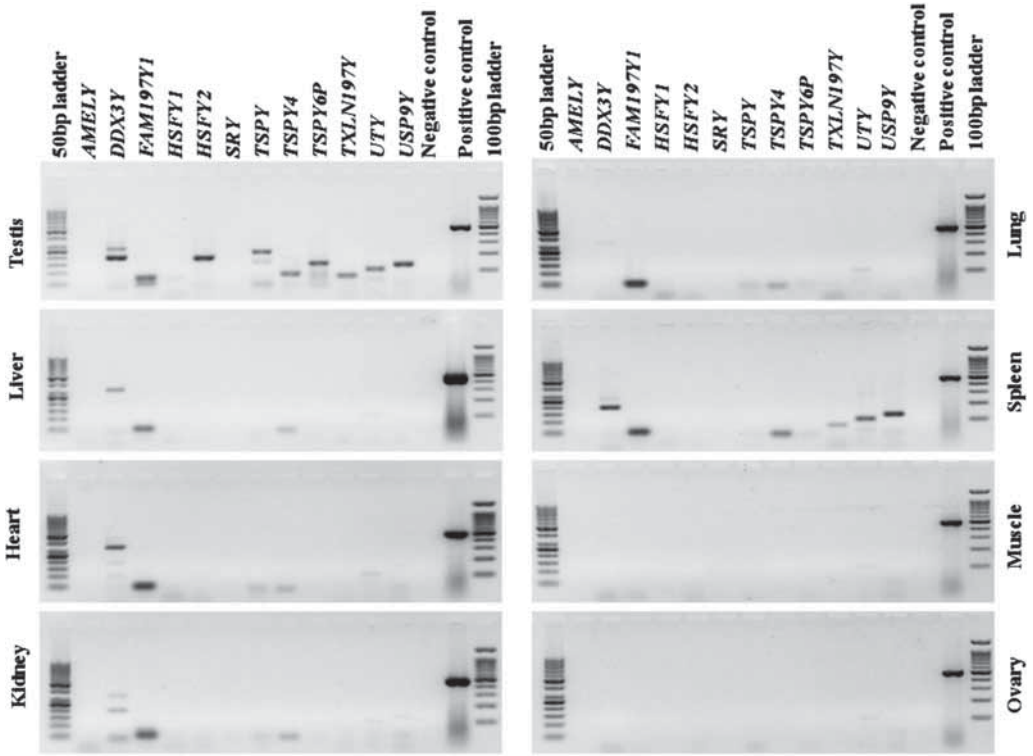

Figure 4. Electrophoretic analysis of RT-PCR amplification of 12 MSY genes in eight major tissues (organ) showing different expression patterns with *GAPDH* as a positive control.

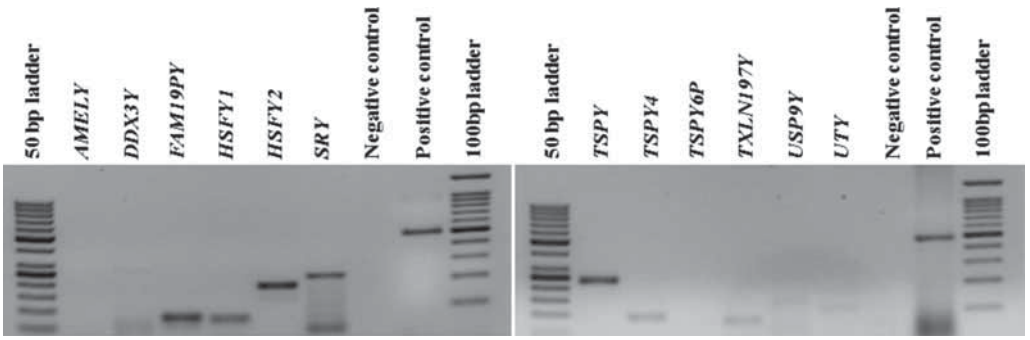

Figure 5. Electrophoretic analysis of RT-PCR amplification of 12 MSY genes in yak sperm.

USP9Y, *UTY*, *DDX3Y*, *TXLNG2P* and *HSF2Y* were constitutively expressed in a range of tissues, namely liver, lung, kidney, heart, spleen, skeletal muscle, testis and sperm (figures 6 and 7). However, *AMLEY* and *HSFY1* genes were not expressed in any of the tissue (figures 6, a&d and 7). Based on the gene expression profiling carried out in the present study, the four MSY genes *DDX3Y*, *FAM197Y1*, *TSPY* and *TSPY4* were constitutively expressed in a broad range of tissue types (figures 6, b, c, g, h and 7), including the testis and sperm. Four MSY genes *HSF2Y*, *TXLNG2P*, *UTY* and *USP9Y* were predominantly expressed in testis along with sperm and spleen (figures 6, e, j, k, l and 7). On the other hand, *TSPY6P* and *SRY* genes were uniquely expressed in testis and sperm, respectively (figures 6, f&i and 7). A composite fold change of all these genes along with four housekeeping genes in different tissue panels is depicted in figure 7.

Discussion

The Y chromosome is identified as the sex-determining chromosome dated back to 1905 during a study of *Tenebrio molitor* which contains genes for normal sperm production in human and other mammals. The Y chromosome has accumulated genes responsible for male reproduction and fertility which evolved from ancestral autosomal chromosome about 300 million years ago

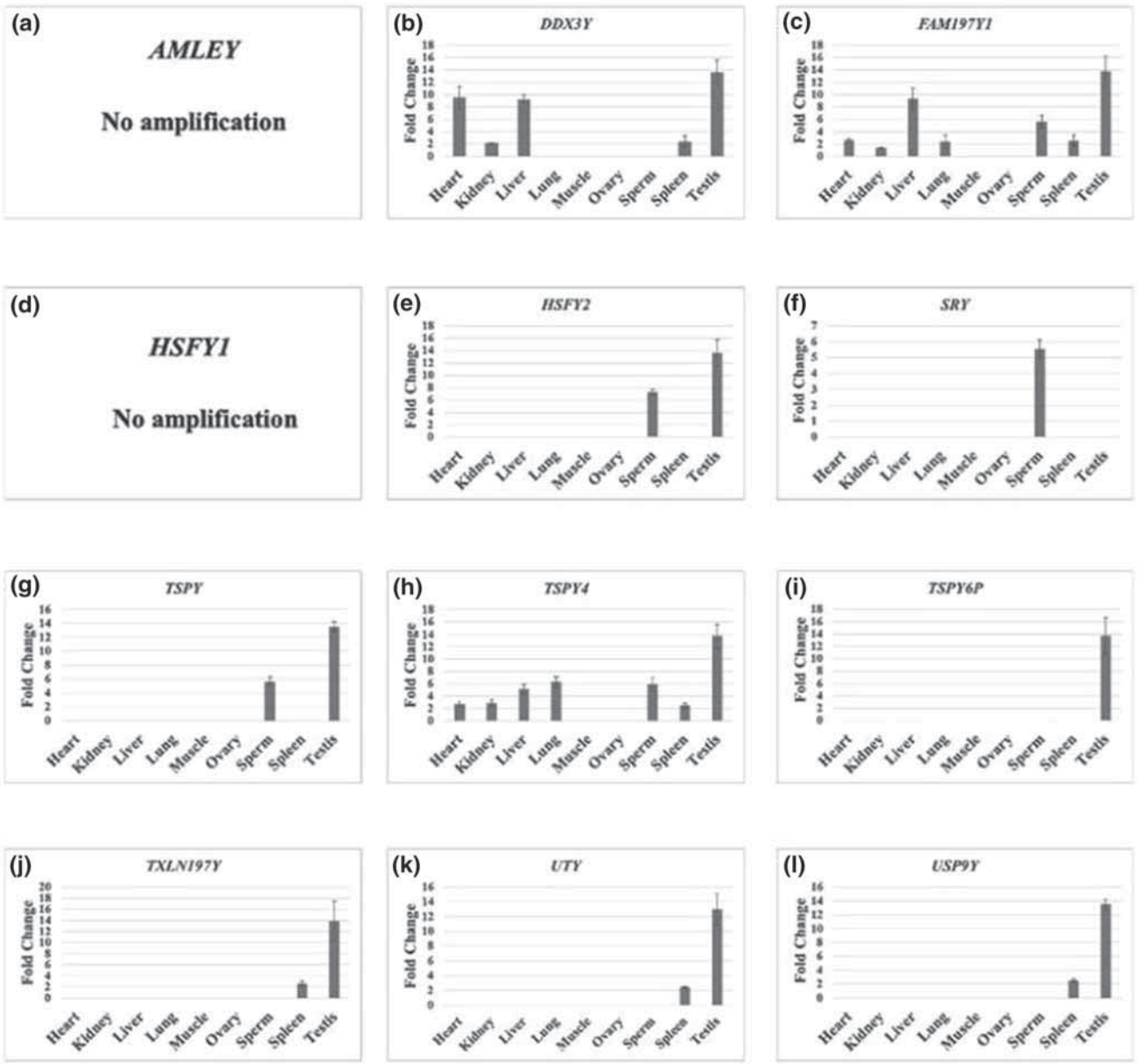

Figure 6. Quantitative real-time PCR expression analysis of MSY genes in different tissues of yak. (a) No expression was observed for *AMELY* gene in any the tissue. (b) *DDX3Y* predominantly expressed in testis, heart and liver and moderately expressed in kidney and spleen. (c) *FAM197Y1* predominantly expressed in testis and liver, and moderately expressed in all tissues except in muscle and ovary. (d) No expression observed for the *HSFY1* gene in any of the tissues. (e) *HSFY2* gene expressed only in testis and sperm. (f) *SRY* gene is exclusively expressed in sperm. (g) *TSPY* gene expressed only in testis and sperm. (h) *TSPY4* gene expressed in most of the tissues, predominantly in testis but not expressed in muscle and ovary. (i) *TSPY6P* gene exclusively expressed in testis. (j) *TXLN197Y* gene is predominantly expressed in testis and scanty expression in spleen. (k) *UTY* gene is predominantly expressed in testis and little expression in spleen. (l) *USP9Y* gene is predominantly expressed in testis and diminutive expression in spleen.

(Lahn and Page 1999; Bachtrog 2013). Our results clearly revealed that this gene could be an excellent sex-specific marker in the yak. Some of the MSY genes show faint bands on female genomic DNA, which can also be well explained by genetic analyses of the eutherian sex-chromosome evolutionary process. Genomic and genetic analyses have revealed that the eutherian sex chromosomes are composed of the heterogeneous mix of sequences with different evolutionary trajectories like PARs, heterochromatin, X-transposed region, X-degenerate region and ampliconic regions. Most of the Y-specific genes characterized in this study have come under the categories of X-degenerate region and ampliconic region of the Y chromosome. The X-degenerative region is a deteriorated version of the ancestral autosome that formed the Y chromosome. The X-degenerate region of the Y

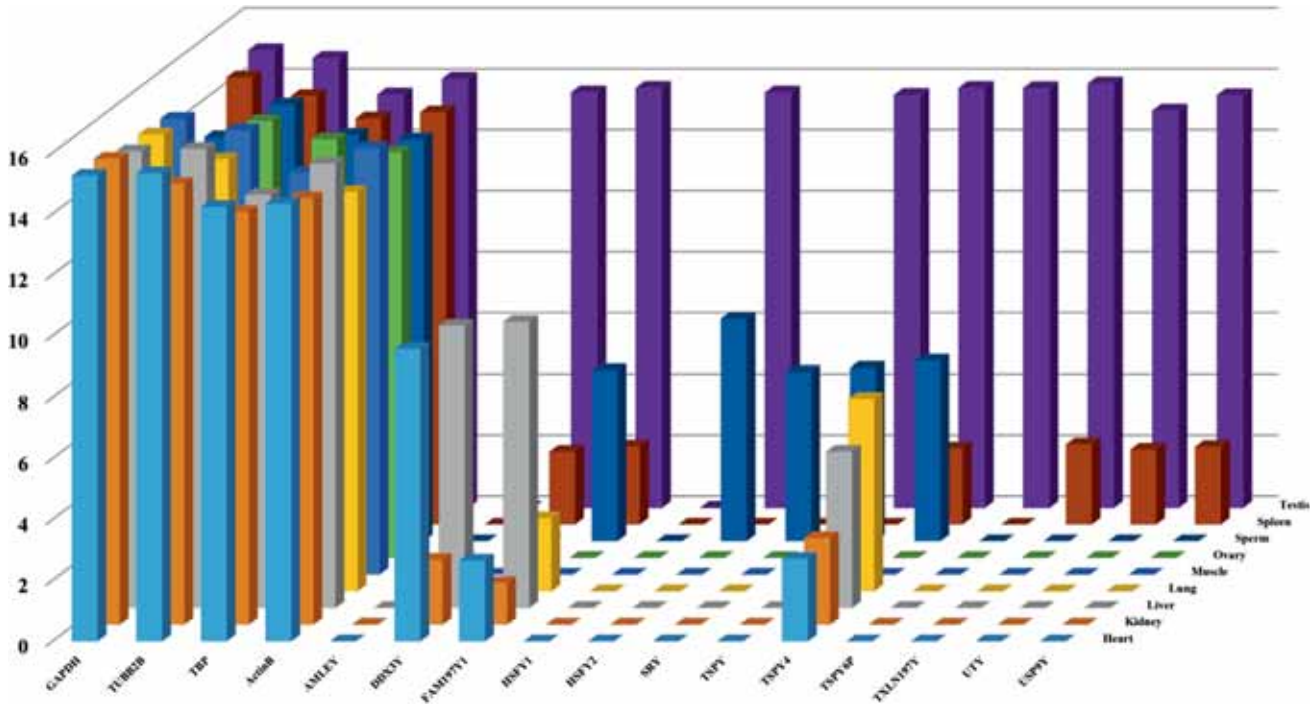

Figure 7. Quantitative real-time PCR expression analysis of 12 MSY and four housekeeping genes in different tissues of yak. The *x*-axis represents different genes tested in the present study, the *y*-axis represents fold change and the *Z*-axis represents different tissues tested for expression analysis.

chromosome displays between 60 and 96% sequence identity with their X-linked homologues, and seems to be surviving remnants of ancient autosome from which the X and Y chromosomes coevolved.

The studies on yak MSY genes have revealed the significance of male viability in mice, cattle and human. Its role in sex determination, hormone regulation, testicular development, spermatogenesis and male fertility is well established in mice (Touré *et al.* 2004; Ellis and Affara 2006; Grzmil *et al.* 2007). The testis-specific expression of 10 genes in our study indicates its role in male reproduction. In addition, elucidation of sperm-specific expression of Y genes opens a new dimension for studying male infertility noninvasively. Moreover, the dispersed expressions of the Y-linked genes in different tissues of yak simultaneously corroborate their association with other functions apart from reproduction and spermatogenesis. The expression of *SRY* gene in yak sperm is the first such report not only on yak but also on other important domestic animals. Earlier studies have suggested that *SRY* is expressed only in human sperm predominantly (Modi *et al.* 2005). Very recently, Chun-Mei *et al.* (2018) studied whether the *SRY* gene and its mRNA transcript are present in the X and Y chromosome-bearing sperm of bulls or not. They established that the *SRY* gene and its mRNA transcript are present in the Y chromosome-bearing sperm of bulls. By using RT-PCR, they detected a single *SRY* transcript of 142 bp in the Y sperm of bull and confirmed this with fluorescence *in situ* hybridization analysis that *SRY* transcripts are present at the edges of the sperm heads of Y chromosome-bearing sperm but are absent from the X chromosome-bearing sperm. This is utmost important that for reproduction of male offspring the haploid y-bearing sperm must contain *SRY* genes but it not necessarily required for overall reproduction and fertility.

The *DDX3Y* gene functions are linked with alteration of RNA secondary structure such as translation initiation, nuclear and mitochondrial splicing and biogenesis of RNA degradation located at the azoospermia factor interval of the Y chromosome. The deletion of the *DDX3Y* region of Y chromosome may lead to infertility or subfertility in human and may cause Sertoli cell-only syndrome (Liu *et al.* 2009). Although alteration of *DDXY* gene may hamper in male reproduction but not necessarily be always causal gene for male infertility. Expression analysis of *DDX3Y* revealed that this gene is highly expressed in yak testis and sperm along with diminutive expression in liver, spleen, kidney and heart (figures 4, 5, 6b and 7). The expression pattern had shown two different product sizes in different tissues (255 bp testis and sperm, and ~310 bp in liver, kidney and heart) under the same reaction conditions which reflects the phenomenon of alternative splicing to carry out tissue-specific functions. The phylogenetic analysis and alignment report of yak *DDX3Y* gene showed sequence mismatches with cattle at the positions C.106A>G and

C.182G>A. These mismatches overturn the earlier belief of having a similar Y gene structure in yak and cattle.

The amelogenin Y (*AMELY*) proteins involved in amelogenesis found in the Y chromosome are essential for the development of enamel. This gene is very important from the evolutionary point of view of sex chromosome but not necessarily related to causal infertility of male. A copy of this gene is also present in X (*AMELX*). The differences between the Y and X chromosome versions of the amelogenin gene may enable to be used in the sex determination of unknown samples. Unexpressed *AMELY* in any of the tested yak tissue indicates the common evolution of this gene in mammals. Phylogenetic analysis specified that yak and cattle *AMELY* evolved from the same lineage (figure 3, a&b in electronic supplementary material). The phylogenetic and sequence analysis of yak *AMELY* gene exhibits 100% nucleotide identity with cattle and 80% with human (table 3 in electronic supplementary material).

The heat shock transcription factor, Y-linked 2 (*HSFY2*) gene is an encoding member of the heat shock factor family of transcriptional activator for heat shock protein. Phylogenetic analysis indicates that yak and cattle *HSFY2* evolved from the same lineage (figure 4, a&b in electronic supplementary material). Although the function of this gene is still elusive in mammals except for human, the expression of this gene exclusively in testis and sperm indicates the functionality of this gene that may play a greater role in spermatogenesis and male reproduction in the animal. The heat shock transcription factor, Y-linked 1 (*HSFY1*) is another protein-coding gene characterized in the present study which has an important paralogue of *HSFX2*. The *HSFY1* gene is specifically amplified from the male yak genomic DNA (figure 3) but not expressed in any of the tissue in the present study. The gene taxilin gamma 2, pseudogene (*TXLNG2P*) is also characterized in the present study and specifically amplified in the male genomic DNA (figure 3). The *TXLNG2P* is a pseudogene affiliated to the long noncoding RNA (lncRNAs) class. The constitutive expression of this gene in testis reveals the role of lncRNAs in an epigenetic modification in the early stages of life. In the present study, the expression of *TXLNG2P* is documented for the first time irrespective of any nonhuman mammal.

Overall, identification and characterization of gene content and genomic features of the yak MSY is essential for a better understanding of the ruminant MSY gene dynamics in male reproduction as well as the evolution of sex chromosome. Reproductive isolation of yak is essentially associated with the increased diversification which may be essential for their adoption. We presume that the massive amplification along with vigorous transcription makes the yak MSY a unique genomic niche to regulate male reproduction during the expansion of ruminant. In the present study, three approaches, namely identification of male-specific gene by PCR, qualitative expression profiling by RT-PCR, quantitative expression profiling by real-time PCR and sequencing; and phylogenetic analysis were carried out to understand yak MSY genes. Perceptibly, this study can be used as a tool for PCR-based detection of sex in yak which may facilitate in the forensic study, embryo sexing etc. The identification of yak MSY genes and their expression profiling provides a foundation for genetic manipulation of the ancestral, single-copy genes. Further, the present study will be on optimal for dissecting individual gene functions, Y chromosome translocations, deletions and transgenesis in the yak and related ruminant species. The basic information on yak MSY genes certainly helps exploring the diverse biology of the male-specific chromosome in the mammalian genetic model, armed with a comprehensive, high-quality reference sequence. The identification and differential expression of all other genes in yak uniquely established their importance in male specificity as well as other tissue functions. This study also substantiates that the occurrence of infertile F_1 male yak–cattle hybrid is due to recombination failure in male meiosis resulting in divergence of Y chromosome gene structure in both the species, while the structural conservation of X chromosome results in the development of a fertile female hybrid.

In conclusion, the present study revealed that yak MSY encompasses rich transcripts which are transcriptionally dynamic, vigorous during spermatogenesis and testicular functions. It also directly showed that the Y-chromosome of yak has structural similarity with cattle and other eutherian mammals with broad transcriptional activity. Another important aspect of this study has depicted sequence mismatches between the yak MSY and cattle MSY which supports the attrition theory of recombination failure of male meiosis resulting in an infertile male F_1 hybrid of yak and cattle. Further, it supports the notion of independent decay of the yak Y chromosome, which has resulted in sequence mismatches with the cattle lineage. This characterization of MSY genes in yak may facilitate to address the male fertility-related issues and sex-chromosome-linked abnormalities in the near future. The results revealed that the majority of the yak MSY genes are predominantly expressed in testis and sperm indicating their necessity in spermatogenesis to support the physiological functions of male reproduction. In future, sequencing of whole Y-chromosome and annotation of its gene content will reveal a broader diversity of lineage-specific Y-chromosome with its involvement in spermatogenesis and male fertility.

Acknowledgements

The authors are thankful to the Director, ICAR-NRCY for granting permission to avail the facilities to carry out this research project. The study was supported by ICAR-NRC on Yak Institutional project grant 2013–2015.

References

Bachtrog D. 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113.

Castresana J. 2000 Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552.

Chevenet F., Brun C., Bañuls A.-L., Jacq B. and Christen R. 2006 Treedyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinf.* **7**, 439.

Chun-Mei H., Rong C., Tao L., Xiao-Li C., Yong-Fu Z., Meng-Ting M. *et al.* 2018 The bovine sex-determining region Y (*Sry*) gene and its mRNA transcript are present in Y sperm but not X sperm of bulls. *Anim. Biol.* **68**, 321–332.

Das P., Chowdhary B. and Raudsepp T. 2009 Characterization of the bovine pseudoautosomal region and comparison with sheep, goat, and other mammalian pseudoautosomal regions. *Cytogenet. Genome Res.* **126**, 139–147.

Das P. J., Paria N., Gustafson-Seabury A., Vishnoi M., Chaki S. P., Love C. C. *et al.* 2010 Total RNA isolation from stallion sperm and testis biopsies. *Theriogenology* **74**, 1099–1106. e2.

Das P., Mishra D. K., Ghosh S., Avila F., Johnson G., Chowdhary B. *et al.* 2013a Comparative organization and gene expression profiles of the porcine pseudoautosomal region. *Cytogenet. Genome Res.* **141**, 26–36.

Das P. J., McCarthy F., Vishnoi M., Paria N., Gresham C., Li G. *et al.* 2013b Stallion sperm transcriptome comprises functionally coherent coding and regulatory RNAs as revealed by microarray analysis and RNA-seq. *PLoS One* **8**, e56535.

Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F. *et al.* 2008 Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469.

Edgar R. C. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Ellis P. J. and Affara N. A. 2006 Spermatogenesis and sex chromosome gene content: an evolutionary perspective. *Hum. Fertil.* **9**, 1–7.

Elsik C. G., Tellam R. L. and Worley K. C. 2009 The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522–528.

Fincham A. G. and Simmer J. P. 2007 Ciba Foundation Symposium 205-Dental Enamel: Dental Enamel: Ciba Foundation Symposium 205, p. 118–118.

Grzmil P., Gołas A., Müller C. and Styrna J. 2007 The influence of the deletion on the long arm of the Y chromosome on sperm motility in mice. *Theriogenology* **67**, 760–766.

Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W. and Gascuel O. 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321.

Krishnan G., Ramesha K., Sarkar M., Chakravarty P., Kataktalware M. and Saravanan B. 2009 Modified temperature humidity index for yaks. *Indian J. Anim. Sci.* **79**, 788–790.

Lahn B. T. and Page D. C. 1999 Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967.

Lan D., Xiong X., Wei Y., Xu T., Zhong J., Zhi X. *et al.* 2014 RNA-Seq analysis of yak ovary: improving yak gene structure information and mining reproduction-related genes. *Sci. China Life Sci.* **57**, 925–935.

Liu W.-S., Wang A., Yang Y., Chang T.-C., Landrito E. and Yasue H. 2009 Molecular characterization of the *DDX3Y* gene and its homologs in cattle. *Cytogenet. Genome Res.* **126**, 318–328.

Medugorac I., Graf A., Grohs C., Rothammer S., Zagdsuren Y., Gladyr E. *et al.* 2017 Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat. Genet.* **49**, 470.

Modi D., Shah C., Sachdeva G., Gadkar S., Bhartiya D. and Puri C. 2005 Ontogeny and cellular localization of SRY transcripts in the human testes and its detection in spermatozoa. *Reproduction* **130**, 603–613.

Paria N., Raudsepp T., Wilkerson A. J. P., O'Brien P. C., Ferguson-Smith M. A., Love C. C. *et al.* 2011 A gene catalogue of the euchromatic male-specific region of the horse Y chromosome: comparison with human and other mammals. *PLoS One* **6**, e21374.

Salido E. C., Yen P., Koprivnikar K., Yu L. and Shapiro L. 1992 The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes. *Am. J. Hum. Genet.* **50**, 303.

Tamura K., Stecher G., Peterson D., Filipski A. and Kumar S. 2013 MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729.

Tobler R., Nolte V. and Schlötterer C. 2017 High rate of translocation-based gene birth on the drosophila Y chromosome. *Proc. Natl. Acad. Sci. USA* **114**, 11721–11726.

Touré A., Grigoriev V., Mahadevaiah S. K., Rattigan Á., Ojarikre O. A. and Burgoyne P. S. 2004 A protein encoded by a member of the multicopy Ssty gene family located on the long arm of the mouse Y chromosome is expressed during sperm development. *Genomics* **83**, 140–147.

Waters P. D., Wallis M. C. and Graves J. A. M. 2007 Mammalian sex-origin and evolution of the Y chromosome and SRY. In *Seminars in cell and developmental biology*, volume 18, no. 3, pp. 389–400. Elsevier.

Corresponding editor: H. A. RANGANATH