Y chromosome-linked genes implicated in spermatogenesis in cattle

Wan-Sheng Liu and Ti-Cheng Chang*

Department of Animal Science, The Center for Reproductive Biology and Health (CRBH), College of Agricultural Sciences, The Pennsylvania State University, University Park, PA 16802, USA; *Current address: Department of Plant Pathology, University of California Davis, Davis 95616, USA

Summary

The mammalian sex chromosomes evolved from an ordinary pair of autosomes during evolution. Unlike the X chromosome that is highly conserved, the Y chromosome is poorly conserved among mammalian lineages. Several special features set the Y chromosome apart from the rest of genome: male-limited transmission, absence of recombination, abundance of Y-specific repetitive sequences, degeneration of Y-linked genes during evolution, acquisition of autosomal genes, and accumulation and functional cluster of "testis genes" for maleness and reproduction. Since the degeneration process is lineage-dependent, different lineages retain different subsets of genes from the ancestral proto-Y chromosome, resulting in a diverse and lineage-specific Y chromosome gene content. During bovine evolution, a lineage-specific ‘autosome-to-Y’ transposition event resulted in three bovid-specific Y chromosome gene families, PRAMEY, ZNF280BY and ZNF280AY. Together, the male-specific region (MSY) of the bovine Y chromosome (BTAY) contains ~ 1200 protein coding genes that can be classified into 12 single copy and 16 multiple copy protein families. The copy number (CN) of these Y-linked gene families varies from 13 for PRAMEY to 236 for ZNF280BY, with significant differences between the taurine and indicine Y lineages. In addition, 367 non-coding RNA families (ncRNAs) were also identified on BTAY. Transcriptome analysis revealed that 95% of the BTAY genes/ncRNAs are expressed predominantly in testis and may be involved in spermatogenesis and male fertility. Though the functional role for the majority of the Y-linked genes needs to be determined, the preliminary data on PRAMEY clearly indicated a role in spermiogenesis. Furthermore, copy number variations (CNVs) of PRAMEY, ZNF280BY, TSPY and HSFY were found to be associated with testis size, sperm quality and fertility in dairy bulls. The authors discuss several challenges that influence male fertility selection associated with the bovine Y chromosome.

Corresponding author E-mail: wul12@psu.edu
Introduction to the mammalian Y chromosome

Evolution of the mammalian sex chromosomes and the degeneration of the Y chromosome

Sex chromosomes in mammals, birds and reptiles originated independently from different pairs of ordinary autosomes (Ohno 1967, Bull 1983, Quinn et al. 2007). The therian sex chromosomes are proposed to have evolved ~ 166 million years ago (MYA), while the human and bovine sex chromosomes are believed to have diverged ~ 80 MYA from the eutherian ancestral sex chromosomes (Veyrunes et al. 2008). The evolution of the therian X and Y chromosomes started from the acquisition of a gene, SRY (sex-determining region Y), which endowed one of the autosomes with a role in sex-determination. A stepwise suppression of recombination between the X and the Y chromosome driven by SRY led to evolutionary strata corresponding to individual suppression events and resulted in the current significantly different gene numbers on the X and the Y (Lahn & Page 1999a, Graves 2006). For example, there are ~ 1,100 protein coding genes on the human X (Ross et al. 2005), whereas there are < 200 genes on the human Y (Skaletsky et al. 2003). This dramatic reduction in the number of functional genes on the Y is caused by a process termed Y chromosome degeneration, in which the Y chromosome loses most of its original genes over evolutionary time (Charlesworth 1991, Rice 1996, Bachtrog et al. 2011). Various evolutionary models have been proposed to explain the degeneration of the Y sequences (Charlesworth & Charlesworth 2000, Bachtrog 2013). These include the Muller’s ratchet (Muller 1918, Charlesworth 1978, Haigh 1978), genetic hitch-hiking (Rice 1987), Hill-Robertson effect (Hill & Robertson 1966, Charlesworth & Charlesworth 2000), and Ruby in the rubbish (Peck 1994, Bachtrog 2013). A common feature of these models is that the efficacy of natural selection is reduced, and the effective population size is decreased as a result of selective events in the non-recombining Y chromosome (Charlesworth & Charlesworth 2000, Bachtrog 2013). The degeneration process is lineage-dependent, i.e. different lineages retain different subsets of genes from the ancestral Y chromosome (also known as proto-Y), resulting in a diverse and lineage-specific Y chromosome gene content (Graves 2006, Pearks Wilkerson et al. 2008, Yang et al. 2011). Therefore, unlike the X chromosome that is highly conserved (Ohno 1967), the Y chromosome is poorly conserved among mammalian lineages (Graves 1998, Ross et al. 2005, Wilson & Makova 2009, Liu 2010).

Features of the mammalian Y chromosome

The mammalian Y is typically the smallest chromosome of the genome, comprising < 3% of the haploid genome (Krausz & Degl‘Innocenti 2006). It is usually a metacentric or acrocentric chromosome and contains a short (Yp) and long arm (Yq). A small region (5% of the Y) located in the distal part of either Yp or Yq that mediates X and Y segregation is known as the pseudoautosomal region (PAR), where X and Y chromosomes pair and recombine during meiosis. The rest of the Y (95%) contains Y chromosome male-specific sequences (MSY) that do not recombine with the X during meiosis (Rice 1996). Several special features set the MSY apart from the rest of genome: absence of homologous recombination, male-limited transmission, abundance of Y-specific repetitive sequences with unique genomic structures (i.e. massive palindromes, or palindrome-like sequences), tendency of MSY genes to degenerate during evolution, acquisition of autosomal genes, and accumulation and functional cluster of “testis genes” for maleness and reproduction (Lahn & Page 1997, Tilford et al. 2001, Rozen et al. 2003, Gvozdev et al. 2005, Liu 2010). Investigating Y chromosomes is challenging as the absence of recombination between the X and Y makes classical linkage-mapping of MSY virtually impossible, and the complexity of the repetitive sequences makes sequencing
extremely difficult (Liu & Ponce de León 2007). This explains why the Y was excluded from most mammalian genome sequencing projects. Most of today’s knowledge regarding the mammalian Y chromosome is based on the three sequenced primate (human, chimpanzee and rhesus macaque) Y chromosomes (Skaletsky et al. 2003, Hughes et al. 2010, Hughes & Rozen 2012, Hughes et al. 2012) and the partially sequenced mouse (Alföldi 2008) and bovine Y chromosomes (Chang et al. 2013b).

**What we have learned from the primate, rodent and other mammalian Y chromosomes**

Genomic analyses have revealed that the primate Y chromosomes are composed of a heterogeneous mix of sequences. The human MSY, roughly 55 Mb, is divided into two regions: euchromatic and heterochromatic. The euchromatic portion of the MSY is roughly 23 Mb, while the heterochromatic region varies in length (polymorphism) among individuals, ranging from undetectable in some men to over half of the chromosome in some others (Krausz & Degl’Innocenti 2006). To date, the DNA sequence and gene annotation are available only for the euchromatic region of the human MSY (Skaletsky et al. 2003).

The euchromatic region of the human MSY contains three classes of sequences: X-transposed (3.4 Mb), X-degenerate (8.6 Mb) and ampliconic (10.2 Mb) (Skaletsky et al. 2003). The X-transposed sequence is unique to humans (not present in Chimpanzee and other primates), that transposed to the Y chromosome ~3-4 MYA (Page et al. 1984) and contains only two genes. The X-degenerate sequences are a deteriorated version of the ancestral Y chromosome. It contains 16 single-copy genes in humans with homologues on the X chromosome that mostly have housekeeping functions (Lahn & Page 1997, Skaletsky et al. 2003). Genes located in the X degenerate regions are completely conserved between human and rhesus macaque, whereas four of the 16 genes in this region were pseudogenized or lost in chimpanzee during evolution (Hughes & Rozen 2012).

The ampliconic sequences are highly repetitive with some massive palindromes. These palindromes exhibit intrachromosomal identities of 99.9% or greater (Skaletsky et al. 2003) and are believed to play an essential role in conserving Y gene functions across evolutionary time through a mechanism of the Y-to-Y gene conversion (Rozen et al. 2003). The human and chimpanzee MSY have 8 and 19 palindromes that span 5.5 Mb and 7.5 Mb, respectively, whereas the rhesus MSY has only three small palindromes, spanning ~ 0.4 Mb. Unlike in humans, most palindromes in chimpanzees exist in multiple copies so that each palindrome arm has multiple potential partners for both intra- and inter-palindrome gene conversion (Bachtrog 2013). The human ampliconic region contains nine distinct MSY-specific protein-coding gene families, with copy numbers ranging from two (VCY, XKRY, HSFY, PRY), three (BPY2), four (CDY, DAZ), six (RBMY) to ~ 35 (TSPY). All of the nine gene families are expressed predominantly or exclusively in testes (Skaletsky et al. 2003). Three (XKRY, HSFY, and PRY) out of nine gene families present in humans are pseudogenized or are simply absent in chimpanzees (Hughes et al. 2010). Thus, the gene repertoire of the chimpanzee MSY is much smaller and simpler than that of the human MSY. In addition, the human ampliconic region also contains 75 putative non-coding transcription units that also are expressed predominantly in testes (Skaletsky et al. 2003).

Genes in the human ampliconic region are derived through three converging processes: i) amplification of X-degenerate genes (for example, HSFY and VCY); ii) transposition and amplification of autosomal genes (DAZ); and iii) retroposition and amplification of autosomal genes (CDY) (Skaletsky et al. 2003). These processes are common across mammalian species and are considered as evidence for the Y chromosome’s ability to accumulate and to maintain
maleness and reproductive genes. In addition to the human DAZ and CDY, several lineage-specific Y chromosome gene families, including the mouse Ssty1 and Ssty2, the cat TETY1 and FLJ36031, and the bovine PRAMEY, ZNF280AY and ZNF280BY, have been reported (Lahn & Page 1999b, Skaletsky et al. 2003, Hughes et al. 2005, Murphy et al. 2006, Church et al. 2009, Chang et al. 2011, Yang et al. 2011). DAZ and CDY appear as main candidates for the human Azoospermia Factor (AZF) (Eberhart et al. 1996, Ruggiu et al. 1997, Slee et al. 1999, Houston & King 2000, Kleiman et al. 2001, Lahn et al. 2002, Dorus et al. 2003, Kleiman et al. 2003, Yen 2004, Kee et al. 2009). Although these two primate lineage-specific Y genes are autosomal in cattle and other non-primates, their functions in spermatogenesis and male fertility are highly conserved even in the non-primate autosomal orthologs (Dorus et al. 2003, Skaletsky et al. 2003, Liu et al. 2007, Wang et al. 2008, Kee et al. 2009). Similarly, the mouse autosomal ortholog (Pramel1) of the bovine Y-linked PRAMEY has also been proposed to play an essential role during spermatogenesis (Mistry et al. 2013). It is evident that the delineation of the MSY gene content in different species will reveal a set of core genes involved in spermatogenesis and male fertility which are fundamental for understanding the cause of infertility (Chang et al. 2013b).

Compared to the primate MSY, the mouse MSY is significantly larger, spanning ~ 95 Mb. The majority (95%) of the mouse MSY is euchromatic with approximately 150-200 large repeats. Each repeat unit is ~ 515 Kb in length and is internally repetitive (Alföldi 2008). Several mouse ampliconic-specific genes have been amplified up to as high as 100 copies and confirmed to be involved in spermatogenesis (Conway et al. 1994, Toure et al. 2004, Alföldi 2008, Riel et al. 2013).

The Bos taurus Y chromosome (BTAY)

The genomic structure of the bovine Y chromosome

BTAY is ~ 51 Mb in size and is the smallest chromosome in the genome (Liu & Ponce de León 2007). The PAR is ~ 6 Mb (Das et al. 2009), and the MSY is ~ 45 Mb. Cytogenetically, the size and morphology of the Y chromosome differ among bovid lineages (Di Meo et al. 2005). BTAY is submetacentric, while the zebu (Bos indicus, BIN) and river buffalo (Bubalus bubalis, BBU) Y chromosomes are acrocentric (Kieffer & Cartwright 1968). This morphological difference is the consequence of Y chromosomal rearrangements through either centromeric transposition or pericentric inversion as revealed by comparative fluorescent in situ hybridization (FISH) (Di Meo et al. 2005). By using Y-linked repetitive sequences as FISH painting probes, Di Meo et al. found that the Y chromosome in different bovid lineages has undergone genomic rearrangements and accumulated various classes of repetitive sequences during the bovid evolution (Di Meo et al. 2005).

The bovine Y is being sequenced (http://www.ncbi.nlm.nih.gov/bioproject/20275), and a draft sequence assembly of ~ 43.3 Mb is available (GenBank acc. no. CM001061.2). The bovine MSY (bMSY) comprises three major regions: X-degenerate (Xd), Y-transitional (Yt) and Y-ampliconic (Ya) regions. The Xd region, split into two sub-regions at either end of the MSY, spans 1.4 Mb (Xd1) in Yp and 1.1 Mb (Xd2) in Yq. The sequences in Xd share similarities from 70 - 95% with their X-linked counterparts, and all genes in Xd (Table 1) are present as single-copy (Chang 2012, Chang et al. 2013b).

The Yt region (~ 3.3 Mb) resides between Xd1 and Ya with a transitional feature between X-degenerate and ampliconic sequences, containing a block of intermingled repetitive and non-repetitive sequences.
### Table 1 Genes/Gene families of the bovine MSY compared with the human and mouse orthologs.

<table>
<thead>
<tr>
<th>Sequence types</th>
<th>Gene</th>
<th>Copy number (range)*</th>
<th>Tissue expression</th>
<th>Chromosomal locations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bovine paralogs§</td>
</tr>
<tr>
<td>X-degenerate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EIF1AY</td>
<td>1</td>
<td>Testis predominant</td>
<td>X (82%)</td>
</tr>
<tr>
<td></td>
<td>OFD1Y</td>
<td>1</td>
<td>Ubiquitous</td>
<td>X (88%)</td>
</tr>
<tr>
<td></td>
<td>USP9Y</td>
<td>1</td>
<td>Ubiquitous</td>
<td>X (89%)</td>
</tr>
<tr>
<td></td>
<td>UTY</td>
<td>1</td>
<td>Ubiquitous</td>
<td>X (84%)</td>
</tr>
<tr>
<td></td>
<td>DDX3Y</td>
<td>1</td>
<td>Testis predominant</td>
<td>X (87%)</td>
</tr>
<tr>
<td></td>
<td>ZFY</td>
<td>1</td>
<td>Ubiquitous</td>
<td>X (94%)</td>
</tr>
<tr>
<td></td>
<td>EIF2S3Y</td>
<td>1</td>
<td>Tissue-specific</td>
<td>X (87%)</td>
</tr>
<tr>
<td></td>
<td>SRY</td>
<td>1</td>
<td>Testis predominant</td>
<td>X (77%)</td>
</tr>
<tr>
<td></td>
<td>RBMY</td>
<td>1</td>
<td>Ubiquitous</td>
<td>X (73%)</td>
</tr>
<tr>
<td></td>
<td>ZRSR2Y</td>
<td>1</td>
<td>Ubiquitous</td>
<td>X (88%)</td>
</tr>
<tr>
<td></td>
<td>RPL23AY</td>
<td>1</td>
<td>Ubiquitous</td>
<td>Autosome</td>
</tr>
<tr>
<td>Y-transitional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UBE1Y</td>
<td>1</td>
<td>Ubiquitous</td>
<td>X (86%)</td>
</tr>
<tr>
<td></td>
<td>PRAMEY</td>
<td>10 (2-31)</td>
<td>Testis-specific</td>
<td>Autosome</td>
</tr>
<tr>
<td></td>
<td>TSPY</td>
<td>19</td>
<td>Testis-specific</td>
<td>X, autosome</td>
</tr>
<tr>
<td>Y-ampliconic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZNF280BY</td>
<td>230   (22-380)</td>
<td>Testis predominant</td>
<td>Autosome</td>
</tr>
<tr>
<td></td>
<td>ZNF280AY</td>
<td>79</td>
<td>Testis predominant</td>
<td>Autosome</td>
</tr>
<tr>
<td></td>
<td>HSFY</td>
<td>190 (21-308)</td>
<td>Testis predominant</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>TSPY(-M2)</td>
<td>157          (37-250)</td>
<td>Testis-specific</td>
<td>X, autosome</td>
</tr>
<tr>
<td>ncRNAs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ncRNA1-367</td>
<td>mc†</td>
<td>Testis predominant</td>
<td>-</td>
</tr>
</tbody>
</table>

* The copy number was estimated from the bovine Y chromosome draft sequence assembly (acc. no. CM001061). The range was obtained from Yue et al. 2013, 2014, and Hamilton et al. 2009, 2011, 2012.

§ The number in the parenthesis is the similarity between X and Y paralogs.

† mc: multiple copies on the Y.
The Ya region extends to a much greater degree compared to Xd and Yt and spans ~34.8 Mb (85% of the MSY). A large fraction of the ampliconic sequences (69%) are palindrome-like and share > 99% intra-chromosomal similarities. Dot plot analyses revealed that Ya was formed based on elaborate arrays of inverted repeats embedded with ampliconic gene families (Yang et al. 2011, Chang et al. 2013b), a genomic structure that is similar to the one identified in the primate and the rodent Y ampliconic regions (see above). Each inverted repeat represented a repeat unit with a relatively fixed size of ~ 420 Kb, which is different from the human and chimpanzee MSY-palindromes with variable sizes. In contrast, the bovine inverted repeats are imperfect palindromes, and a larger number of repeat units are present in bMSY. It is estimated that ~ 80 repeat units are present on the sequenced Y chromosome of a Hereford bull (L1 Domino 99375) (GenBank acc. no. CM001061.2) (Chang et al. 2013b). In general, the genomic structure of bMSY is more like the mouse MSY (Alföldi 2008) than the primate MSY (Skaletsky et al. 2003, Hughes et al. 2010, Hughes et al. 2012).

The gene content of the bovine Y chromosome

Recent transcriptome analysis of bMSY (GenBank acc. no. GAQ000000000) identified a total of 1,274 protein-coding genes/families and 367 additional non-coding RNA (ncRNA) families, making the bMSY gene density (~31.2 genes/Mb) the highest in the genome, in comparison with ~ 9.4 genes/Mb for the bovine X (BTAX) and ~10.2 genes/Mb for the entire genome (Chang et al. 2013b). The discovery of the higher gene density along with the high transcriptional activities observed from these Y chromosome genes (see below) challenges the widely accepted hypothesis that the MSY is gene poor and transcriptionally inert.

Protein-coding genes. The 1,274 MSY-linked genes belong to 28 gene families including 12 single- and 16 multicopy genes (Table 1). Sixteen of these gene families are bovid specific (Chang et al. 2013b). A total of ten genes (EIF1AY, AMELY, OFD1Y, USP9Y, UTY, DDX3Y, ZFY, EIF2S3Y, RBMY and SRY) located in the Xd region (Xd1 and Xd2) are all single-copy with X-counterparts. Two additional single-copy genes, UBE1Y and CYorf15, are located in the Yt region, both of which have X-counterparts. An array of TSPY genes and a bovid-specific gene family, PRAMEY, are also present in the Yt (Verkaar et al. 2004, Chang et al. 2011). The TSPY array comprises ~ 19 gene copies duplicated in tandem within a ~ 600 kb region while the PRAMEY gene family is composed of ten duplicates in the Yt region on the sequenced Hereford Y chromosome (Chang 2012). Notably, the amplification of PRAMEY was limited to the Yt region and did not extended to the Ya, indicating a unique genomic context within the Yt.

Four major ampliconic gene families map in Ya, including HSFY, TSPY, ZNF280BY and ZNF280AY. They were largely amplified on BTAY with an estimated copy number of 190, 136, 234, and 79, respectively (Chang et al. 2013b). Their copy numbers were increased by a factor of 80 corresponding to the predicted number of repeat units on BTAY. HSFY and TSPY were conserved in mammalian species (Bhowmick et al. 2007) and evolved as a consequence of degeneration of the proto-XY chromosomes. The amplification of the bovine TSPY and HSFY is greater and broader than the human TSPY (35 copies) and HSFY (2 copies) (Skaletsky et al. 2003). The ZNF280BY and ZNF280AY genes were instead bovid-specific. In addition, BTAY was dotted with 11 putative, bovid-specific coding genes, including EGLY and Bovid-specific Transcript, Y-linked (BTY) 1-10, with a copy number from 2 to 174 copies (Chang et al. 2013b) (Table 1).

ZNF280BY, ZNF280AY and PRAMEY formed a 60 kb gene block on BTAY, with a paralogous gene block located on BTA17 (Chang et al. 2011, Yang et al. 2011). A cross-species comparison indicated that the autosomal gene block is highly conserved not only in eutherians but also in non-placental vertebrates, including opossum, chicken, frog, and zebrafish. In contrast, the Y-linked
orthologs were only present in the bovid lineages (Yang et al. 2011, Chang 2012, Chang et al. 2013a, Chang et al. 2013b). Phylogenetic trees of these three gene families revealed a consistent topology that ZNF280BY, ZNF280AY and PRAMEY always form a tight cluster with the BTA17 paralog before clustering with other orthologs. These results suggested that these three Y-linked gene families were evolved from an autosome-to-Y transposition of the BTA17 gene block and amplified differentially thereafter (Chang et al. 2011, Yang et al. 2011, Chang et al. 2013b).

Pairwise synonymous substitution rate analyses applied for the three most amplified coding genes HSFY, TSPY, and ZNF280BY indicated that the bovine ampliconic region underwent at least two major evolutionary expansions (Chang et al. 2013b). The earlier expansion occurred in the Miocene, ~ 14–20 MYA, during which the earth’s climate experienced dramatic changes and the Bovinae diverged from Antilopinae (Hassanin & Douzery 2003). The later expansion occurred in the Pliocene, within 5 MYA, during which increase of the species richness and abundance of Bovidae was observed (Bobe & Eck 2001). Therefore, the amplification of the bovine ampliconic gene families may have contributed to the diversification and speciation of the Bovidae (Chang et al. 2013b).

Non-coding RNAs. A total of 367 BTAY-linked ncRNA families were identified by a testis RNA-seq approach. The biological roles and biogenic mechanisms underlying these ncRNAs remain unknown. However, expression analyses of 21 ncRNAs have consistently shown a predominant expression in testis, suggesting their involvement in spermatogenesis. Furthermore, motif analysis indicated that 11 ncRNA families comprised motifs associated with transposable elements, which have implicated roles in genomic duplications (Johnson et al. 2006, Chang et al. 2013b). Therefore, the expression of these ncRNAs may be essential in the expansion of the ampliconic regions and the duplications of the ampliconic genes.

The expression of BTAY-linked genes during testis development and spermatogenesis

Transcriptome of bMSY

RNA-seq on bovine testis of three different stages, 20 days (20d), 8 months (8m) and 2 years (2y), indicated that the transcriptional activity of bMSY was vigorous with ~ 95% of the bMSY genes expressed (Chang et al. 2013b). A hierarchical clustering of the differentially expressed genes across the three different stages of testes, including 108 bMSY genes/families, revealed five major expression patterns. The majority (61%) of the differentially expressed bMSY genes were clustered into a main pattern with an increased expression from 20d towards 2y. Gene ontology enrichment analyses indicated that these genes were associated with reproduction and spermatogenesis. The second pattern displayed the highest expression in 8m testis and covered 36% of the bMSY genes which were associated with ion transport and cell growth. The remaining 3% of bMSY genes with the other three expression patterns were associated with diverse biological process. These results recapitulate the fundamental roles of the bMSY genes during spermatogenesis and male reproduction (Chang et al. 2013b).

The expression of the antisense RNA of Y-linked genes

In situ hybridization (ISH) analyses indicated that both the sense and antisense RNAs of ZNF280BY were expressed in adult testis (Yang et al. 2011). The sense RNA is detected broadly in different types of the seminiferous tubule cells. The antisense RNA was instead detected specifically in spermatids. ISH of PRAMEY displayed a pattern in contrast to ZNF280BY. The sense RNA of
PRAMEY is detected only in spermatids, whereas the antisense RNA is detected across different cells of seminiferous tubules with a predominant signal detected in spermatids (Chang et al. 2011). Another BTAY associated gene, DDX3Y and an autosomal gene, CDYL (the ortholog of the human CDY) have also been demonstrated with antisense RNA expression in spermatocytes and spermatids (Wang et al. 2008, Liu et al. 2009). These results suggested that the antisense RNAs may provide an important layer of regulation during spermiogenesis.

Protein localization and function of the Y-linked genes

Besides predominant expression of the BTAY-linked genes in testes, knowledge of the molecular function and cellular (subcellular) localization of the Y-gene encoded proteins is lacking. A recent work revealed that the bovine PRAMEY protein is expressed in different steps of spermatids and in acrosome and flagellum of spermatozoa (Zhao 2013). Immunogold electron microscopy further revealed that PRAMEY was firstly localized in the acrosomal granule of round spermatids, migrated with the content of acrosomal granule during spermiogenesis, and was finally present in the acrosomal matrix of mature spermatozoa, suggesting that PRAMEY is involved in acrosome biogenesis (Zhao 2013). Similar to the bovine PRAMEY, the mouse ortholog (Pramel1) (Mistry et al. 2013) and the human ortholog (PRAME) (Zhao et al. 2013) also were found to be expressed in acrosome. Further analysis indicated that PRAMEY interacts with the protein phosphatase 1 gamma isoform 2 (PP1γ2) (Zhao et al. 2013). The latter is a key component for regulation of spermatozoa motility and male fertility (Mishra et al. 2003, Fardilha et al. 2011).

The bovine HSFY protein was detected in spermatogenic cells and Sertoli cells. Western blot analysis revealed HSFY in the testicular tissues with normal spermatogenesis, maturation arrest, and Sertoli cell-only syndrome, but the amount of the protein in the samples of the maturation arrest and Sertoli cell-only syndrome was different. The expression of HSFY was low or absent in spermatogenic cells of maturation arrest specimens, and the ratio of HSFY expressed in Sertoli cells was different in the samples with maturation arrest and with Sertoli cell-only syndrome (Sato et al. 2006). Therefore, HSFY may play a role in the regulation of spermatogenic cell differentiation. However, a recent study on infertile men with the HSFY deletion discovered that HSFY has only a minor contribution to male fertility (Kichine et al. 2012).

Y chromosome variations and male fertility in cattle

In addition to the bovid Y chromosome structure variations described above, there are at least three more types of sequence variations. These include single nucleotide polymorphism (SNP), insertion/deletion (indel), and copy number variation (CNV), which provide an opportunity to develop Y-linked genetic markers for male reproduction evaluation. Due to the repetitive sequences and the multicopy of Y locus, chromosome-wide SNP and indel analysis and their association with male fertility are still lacking in cattle. In contrast, considerable progress has been made in CNVs of Y-linked genes thanks to the qPCR (quantitative real-time PCR) approach (Hamilton et al. 2011, Yue et al. 2013).

CNVs of PRAMEY

The copy number (CN) of PRAMEY varied from 2 to 31 with a median CN (MCN) of 13 among 460 bulls tested. Significant variations were observed among 15 cattle breeds. Holstein bulls had
the lowest MCN of 12, whereas Limousin bulls possessed the highest MCN of 26. Furthermore, bulls in the taurine (BTA) lineage had a significantly lower MCN (13) than those bulls in the indicine (BIN) lineage (20). Association analysis revealed that the PRAMEY CN was correlated negatively with scrotal circumference (SC), relative scrotal circumference (RLSC), percentage of normal spermatozoa (PNS), and non-return rate (NRR), but had no significant association with post thaw motility (PTM), incubated motility (IM), percentage of intact acrosome (PIA), sire conception rate (SCR) and relative breeding efficiency (RBE) in Holstein bulls (Yue et al. 2013).

**CNVs of ZNF280BY**

The MCN of ZNF280BY was 236, ranging from 28 to 380 among 460 bulls (of 15 breeds). The taurine bulls had a significantly lower MCN (231) of ZNF280BY than indicine bulls (284). Association analysis in Holstein bulls (n = 140) revealed that the CNVs of ZNF280BY were correlated negatively with SC and RLSC, while positively with SCR (Yue et al. 2014).

**CNVs of HSFY**

A previous study indicated that the CN of HSFY was around 70 on BTAY with no variation among 24 Holstein bulls (Hamilton et al. 2011). A recent report found that the MCN of HSFY among 460 bulls was 197, varying from 21 to 308. Bulls in the BTA-lineage had a significantly higher MCN (202) than bulls in the BIN-lineage (178) (Yue et al. 2014). The CN discrepancy between the two studies was due to the difference in the gene PCR primer design, in which Hamilton et al.’s primers targeted only the highly conserved portion of the HSFY loci, whereas Yue et al.’s primers amplified all HSFY loci on the Y. Similar to ZNF280BY, the CNVs of HSFY were negatively associated with RLSC and SC (Yue et al. 2014).

**CNVs of TSPY**

The bovine TSPY is one of the most investigated genes on BTAY. The CN of TSPY varied from 37 to 200 according to different reports (Vogel et al. 1997, Verkaar et al. 2004, Hamilton et al. 2012, Chang et al. 2013b) and the MCN was 94 among 14 cattle breeds investigated (Hamilton et al. 2009). The CNVs of TSPY was positively correlated with adjusted NRR and negatively correlated with the TSPY (mRNA) expression level in the testis, but no correlation with the semen parameters was measured (Hamilton et al. 2012). In humans, the CN of TSPY was different among Y lineages and was associated with spermatogenesis (Giachini et al. 2009, Nickkholgh et al. 2010). The TSPY CN was proposed as a potential risk factor for male infertility (Vodicka et al. 2007).

The lower the CN of Y-linked genes, the larger the testicular size

It is interesting to see that the CNVs of PRAMEY, ZNF280BY and HSFY are negatively correlated to SC and RLSC, indicating that bulls with a low CN of Y-linked genes tend to have a larger testicular size. In a remarkable case, a Holstein bull who had the lowest CN in PRAMEY (CN = 2), HSFY (21) and ZNF280BY (28) gene families was found to have the largest RLSC among all animals tested (Yue et al. 2013, Yue et al. 2014). It is likely that this particular Y chromosome
has a large deletion in the Ya region, which supports the discovery that lesser CN is associated with larger testis in Holsteins.

**CNVs of Y-linked genes among paternal lineages**

In today’s Holstein population, the effective population size of males is very small (see discussion below), leading to a few dominant paternal lineages in North America Holsteins. A recent paternal pedigree analysis of 192 Holstein AI bulls found all but one were descendants of only 3 patrilineal founders (HOUSA1427381, HOUSA1441440 and HOUSA1491007) that were born in the 1960s. The founder effect on the CNVs of Y-linked genes and male reproductive traits was examined. The results indicated that the founders had no significant effects on the reproductive traits, and the interactions between the founders and the CNVs of the Y-linked genes also had no significant effects (Yue et al. 2013, Yue et al. 2014).

**What are the challenges to male fertility selection associated with the Y chromosome in cattle?**

**Significant reduction of the effective population size in cattle**

A very unique feature of the Y chromosome is that it is always transmitted from father to son. Male-limited transmission means that the Y chromosome is an ideal part of the genome to carry genes that increase male fitness. In addition, male-beneficial mutations on the Y chromosome are always transmitted and accumulated through this mechanism (Rice 1996, Bachtrog 2013). One of the consequences of male-limited transmission is the reduction of the effective population size (Wilson Sayres et al. 2014). In a natural population, because of its paternal transmission, the effective population size of Y chromosomes relative to X chromosomes and autosomes is reduced. When the female-to-male breeding sex ratio is 1, Y-linked genes are only 1/4 as many as autosomes and 1/3 as many as X chromosomes (Tucker & Lundrigan 1995). If the female-to-male breeding sex ratio is greater than 1, as are the cases in most farm animal breeding programs, then the effective population size of Y chromosomes relative to X and autosomes is further reduced.

In the case of the dairy cattle, particularly in Holsteins and Jerseys, the effective population size \((Ne)\) (for all males and females) has been extremely reduced in the past 50 years as a consequence of artificial selection and the application of the artificial insemination (AI) technology. For example, the \(Ne\) of Holsteins in Northern America is estimated to be only about 100 (based on 2011 data from the Council on Dairy Cattle Breeding, using equation \(Ne = \frac{1}{2} \Delta F, \Delta F = 0.0053\), with an assumption that the generation interval is \(~4-5\) years). There are approximately 1200 active Holstein AI sires (with daughters) and \(~7\) million Holstein cows in US. The ratio of the Y chromosome to the X chromosome is as low as \(1:11,668\). Since many of these AI bulls are genetically related, the effective population size of males (or Y chromosomes) is even more reduced. We have recently carried out a paternal pedigree analysis with data from the Holstein Association databases in US and Canada. The pedigrees dated back to 1880s when the Holsteins were initially imported to U.S. from the Netherlands. To our surprise, all of the current Holstein bulls (with \(\geq 50\) daughters) in North America are descendants of two ancestors (886 H H B, Hulleman, 3/27/1881; 711 H H B, Neptune H, 3/23/1880), implying that only two independent Y chromosomes are present in today’s Holsteins. So, the actual ratio of the Y chromosome to the X chromosome is 1 to 4.7 million (X P Yue, C Dechow & W-S
Liu 2014, unpublished observations). Theoretically, the extremely small population size of Y chromosomes will reduce the variation at Y-linked loci (relative to X-linked and autosomal loci), resulting from stochastic bottleneck effects (Nei et al. 1975, Tucker & Lundrigan 1995). Practically, the effect of this extremely small Ne of Y chromosomes on the Holstein breeding program, especially on male reproduction traits, is unknown and has to be addressed.

Lack of molecular study on male reproductive efficiency

Reproductive efficiency is one of the most important economic traits in cattle production. Over the past half-century, reproductive efficiency of dairy and beef cattle has significantly declined both in cows (Lucy 2001, Inskeep & Dailey 2005, Royal et al. 2008) and bulls (Coulter 1980, Coulter & Kozub 1980, Courot & Colas 1986, DeJarnette et al. 2004, Kastelic 2013). Fertilization rate is heavily influenced by male factors. In the AI industry, a variation of up to 25% of NRR is evident within a population of bulls that meet the normal commercially acceptable standards (Linford et al. 1976, Saacke et al. 1988). Furthermore, embryonic death is influenced by male factors associated with lowered semen quality as measured by sperm viability and % normal sperm morphology (Bearden et al. 1956, Sullivan & Elliott 1968, Mattei 1984, Courot & Colas 1986, Saacke et al. 1988, Setchell et al. 1988, Chenoweth 2007). In the beef industry, it was estimated back to 1980 that as many as 18-30% of beef bulls used in natural service were reproductively deficient (Coulter 1980, Coulter & Kozub 1980). More than 30 years later, the problem not only remains, but even the percentage of the deficient bulls remains roughly the same (Kastelic 2013). Therefore, low fertility (or sub- or in-fertility) is a significant problem in cattle.

To address this problem, cattle breeders and geneticists need to understand the mechanisms that regulate spermatogenesis and male fertility. Unfortunately, bull fertility/infertility has not been studied at the molecular level because of the lack of molecular genetic markers and diagnostic tools. This has prevented the identification of high fertility, subfertile or infertile sires at an early age. Quite often, even if an infertile bull is identified, he is usually eliminated from the breeding program immediately without further investigation. The majority of the previous studies on bull fertility focused on the routine examination of semen quality, including sperm concentration and structural and morphological analyses (Amann & DeJarnette 2012) as well as testicular size and temperature (Kastelic 2013). The widely used field measurement, NRR, to estimate the bull fertility was based on the cows’ pregnancy status and is highly prone to bias and unexplained variability (reviewed in Amann & DeJarnette 2012). A new dairy bull fertility estimation system termed SCR proposed by the USDA Animal Improvement Programs Laboratory (AIPL) in 2008, is a modified version of NRR and is based on a large, nation-wide database (Norman et al. 2008, Amann & DeJarnette 2012). Nevertheless, both NRR and SCR are based on field data collected from breeders, not based on the sire’s genotype. Furthermore, because the heritability of NRR and SCR is very low, usually smaller than 0.05 (Kuhn & Hutchison 2008, Zhang 2008), these measurements do not accurately reflect actual male fertility.

Although genomic selection is becoming a routine practice in dairy breeding programs based on genotypic data from the bovine 50K and 778K SNP chips, very limited progress has been achieved in improving bull fertility because limited information on genes associated with bull fertility is available (Huang et al. 2010, Burns et al. 2011, Penagaricano et al. 2012). Several approaches, including candidate gene (or pathway) (Khatib et al. 2010, Ogorevc et al. 2011), seminal proteomics (Killian et al. 1993), sperm transcriptome (Feugang et al. 2010), and genome-wide associate analysis (GWAS) (Feugang et al. 2009, Huang et al. 2010, Penagaricano et al. 2012) have been applied to identify potential gene/DNA markers. These resulted in a small
number (<100) of candidate SNPs for male fertility. Given the fact that over 1000 known genes are involved in spermatogenesis (Hermo et al. 2010), the identification of a small number of associated SNPs to bull fertility is unexpected. Two possible factors may contribute to the result: a) the significantly reduced effective population size of dairy bulls that may lead to a decrease in genetic variation of fertility traits (described above); and b) the design of experiments and efficiency of the approach used in the associated analysis. An example for the latter factor is from a recent GWAS analysis that used the BovineSNP50 Bead Chip and SCR data from 1755 sires and identified only 8 (non-Y chromosome) male fertility-associated SNPs (Penagaricano et al. 2012). The outcome of this study was not surprising because that SCR ranking does not accurately reflect male fertility as we discussed above. So, critical thinking and well-designed GWAS experiments are required for genomic evaluation of bull fertility (Amann & DeJarnette 2012).

Use of the Y-linked repetitive markers in the bovine high density (HD) SNP chip

Among the 778K SNPs on the BovineHD beadchip assay (Illumina, San Diego, CA), 1,421 are Y-linked SNPs (McDaneld et al. 2012) that cover the entire Y chromosome draft sequence assembly (GenBank acc. no. CM001061.1). Our preliminary analysis of these so-called Y-linked SNPs demonstrated that the majority of them are not typical SNPs by definition; instead, they are nucleotide variations among repeat units (see BTAY structure above) (or among different loci in a gene family) within an individual Y chromosome (in this case, the Y chromosome of the Hereford bull, L1 Domino 99375). Our experience with the Illumina Genome Studio software (Illumina, San Diego, CA) indicated that the software is not suitable to analyze the Y-linked multicopy nucleotide variations.

There is only one report to date that utilized the Y-linked markers on the HD SNP chip for reproduction analysis in beef cattle (McDaneld et al. 2012). Results from this study were debatable as the authors reported that a large portion (550/1224) of these Y-linked SNPs was significantly associated with decreased reproductive efficiency in females. Furthermore, 21 to 29% of females in the low reproductive category (from various beef cattle herds across the U.S.) were positive for the Y-linked markers (McDaneld et al. 2012). Though this observation needs to be validated by independent studies, others who used the same BovineHD beadchip had encountered problems with artifacts of genotyping the Y-linked SNPs in Holstein females (Dr. Jerry Taylor, personal communication). We tend to believe that the atypical Y-linked SNPs (multicopy and hemizygous) and the limitation of the (Illumia) Genome Studio prevent a meaningful use of these Y-linked markers (on the BovineHD chip) from association studies on male reproductive efficiency in cattle.

Conclusions

Male-limited transmission makes the Y chromosome an ideal site to carry genes that increase male fitness, but, at the same time reduce the effective population size. As a result, each Y-lineage accumulates a cluster of testis expressed genes for maleness and reproduction. In cattle, the Y chromosome, the smallest in the genome, contains ~ 1200 protein coding genes and 367 ncRNA families, which make the Y chromosome the highest gene density in the genome. These genes are located in the palindrome-like sequences of the Y-specific repeat units. Approximately 80 repeat units are present in the bovine Y draft sequence assembly. The number of repeat unit varies from individual to individual, providing a genomic mechanism
for the CNV of BTAY genes. Three bovid-specific Y chromosome gene families were identified, which together with the rest of BTAY genes and ncRNAs are expressed predominantly in testis. CNVs of these Y-linked gene families have been associated with testis size and male fertility. However, the functional role of these genes in spermatogenesis and fertilization is largely unknown. Although working on Y chromosome related genes and markers is challenging, the rationale and significance of studying the Y chromosome to improve male reproductive efficiency is evident. It is particularly important in the dairy and beef breeding program where artificial insemination and genomic selection have been applied. How to maintain the current Y chromosome lineages and to prevent them from further reduction in Holsteins and other cattle breeds is an urgent issue. Future research should focus on the identification of variations in the Y chromosome genes/sequences (DNA level), transcription (RNA level) and translation (protein level) and their associations with semen quality and male fertility. It is the authors’ belief that a well-designed Y-linked gene chip would be valuable for bull fertility selection at an early age.

Acknowledgements

We thank Chad Dechow and Dale Olver (Department of Animal Science, The Pennsylvania State University, University Park) for valuable discussion and review of the manuscript. This project was supported by a grant (no. 2010-65205-20362) from the National Institute of Food and Agriculture (NIFA), USDA to WSL.

References


Das PJ, Chowdhary BP & Raussepp T 2009 Characterization of the bovine pseudoautosomal region and comparison with sheep, goat, and other mammalian pseudoautosomal regions. Cytogenetic and Genome Research 126 139-147.


Hamilton CK, Verduzco-Gomez AR, Favetta LA, Blondin P & King WA 2012 Testis-specific protein Y-encoded copy number is correlated to its expression and the field fertility of Canadian Holstein bulls. Sexual Development 6 231-239.

Hassanin A & Douzery EJ 2003 Molecular and morphological phylogenies of ruminantia and the alternative position of the moschidae. Systematic Biology 52 206-228.

Hermo L, Pelletier RM, Cyr DG & Smith CE 2010 Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. Part 5: intercellular junctions and contacts between germ cells and Sertoli cells and their regulatory interactions, testicular cholesterol, and genes/proteins associated with more than one germ cell generation. Microscopy Research and Technique 73 409-494.


chromosomes are remarkably divergent in structure and gene content. Nature 463 536-539.
Lahn BT & Page DC 1999a Four evolutionary strata on the human X chromosome. Science 286 964-967.
arr-scr1.htm.


Toure A, Grigoriev V, Mahadevaiah SK, Rattingan A, Ojarikre OA & Burgoyne PS 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.


Yue XP, Dechow C, Chang TC, Dejarnette JM, Marshall CE, Lei CZ & Liu WS 2014 Copy number variations of the extensively amplified Y-linked genes, HSFY and ZNF280BY, in cattle and their association with male reproductive traits in Holstein bulls. BMC Genomics 15113.


