Corrections

BIOCHEMISTRY
Correction for “Purified vitamin K epoxide reductase alone is sufficient for conversion of vitamin K epoxide to vitamin K and vitamin K to vitamin KH2,” by Pei-Hsuan Chu, Teng-Yi Huang, Jason Williams, and D. W. Stafford, which appeared in issue 51, December 19, 2006, of Proc Natl Acad Sci USA (103:19308–19313; first published December 12, 2006; 10.1073/pnas.0609401103).

The authors note that Fig. 5 and its corresponding legend appeared incorrectly. The corrected figure and its corrected legend appear below.

Also, the authors note that on page 19312, right column, 3rd full paragraph, line 5 “The reaction was carried out for 1 hr.” should instead appear as “The reaction was carried out for 20 min.”

Fig. 5. Conversion of vitamin K to vitamin KH2 by VKOR. The reaction was performed by using purified VKOR that had been dialyzed in the presence of THP. VKOR activity is represented as turnover number per second. Bar 1, DTT with elution buffer as background control; Bar 2, purified VKOR after dialysis against buffer A with 4 mM THP. Data are represented as mean ± SD (n = 3).

www.pnas.org/cgi/doi/10.1073/pnas.1401722111

EVOLUTION

The authors note that the author name Anne M. Bronikowski should instead appear as Anne M. Bronikowski. The corrected author line appears below. The online version has been corrected.


www.pnas.org/cgi/doi/10.1073/pnas.1324133111
**GENETICS**


The authors wish to note, “We have recently updated the data associated with our GenBank depositions to include age, tissue, and developmental stage of the bovine testis RNA-seq data. We have also deposited new data for information discussed in the Supporting Information of our article. We apologize for not providing this information at the time of publication. The updated accession numbers are as follows:

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Sample</th>
<th>Developmental Stage</th>
<th>Tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRX357350</td>
<td>Bos taurus 789_20D</td>
<td>postnatal-20 d</td>
<td>Testis</td>
</tr>
<tr>
<td>SRX357349</td>
<td>Bos taurus 789_20D</td>
<td>postnatal-20 d</td>
<td>Testis</td>
</tr>
<tr>
<td>SRX357348</td>
<td>Bos taurus 74_8M</td>
<td>puberty-8 mo</td>
<td>Testis</td>
</tr>
<tr>
<td>SRX357347</td>
<td>Bos taurus 74_8M</td>
<td>puberty-8 mo</td>
<td>Testis</td>
</tr>
<tr>
<td>SRX357346</td>
<td>Bos taurus 645_2Y</td>
<td>maturity-2 y</td>
<td>Testis</td>
</tr>
<tr>
<td>SRX357345</td>
<td>Bos taurus 645_2Y</td>
<td>maturity-2 y</td>
<td>Testis</td>
</tr>
<tr>
<td>SRX388838</td>
<td>Direct cDNA selection of the bovine Y chromosome</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“In addition, the project ‘Transcriptome analysis of the bovine Y chromosome,’ together with the bovine testis cDNA selection reads and assembled transcripts/ncRNAs (> 200 bp), were submitted to the Transcriptome Shotgun Assembly (TSA) database, www.ncbi.nlm.nih.gov/genbank/tsa (Bioproject accession no. PRJNA230872; reads accession no. SRX388838).

“The assembled contigs (> 200 bp) have been deposited at DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank under the accession GAQ00000000. The version described in this paper is the first version, GAQ001000000.”

**IMMUNOLOGY**


The authors note that James A. Thomas should be included as a cocorresponding author. Correspondence can be addressed to him at james.thomas@bcm.edu.

Also, the authors note that they omitted references to articles by Juliana et al. and Fernandes-Alnemri et al. The complete references appear below.


www.pnas.org/cgi/doi/10.1073/pnas.1401725111
Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development

Ti-Cheng Chang,a,b Yang Yang,a,† Ernest F. Retzel,b and Wan-Sheng Liu,a,b,c

*Department of Animal Science, Center for Reproductive Biology and Health, College of Agricultural Sciences and bIntegrative Biosciences Graduate Program Option in Bioinformatics and Genomics, The Pennsylvania State University, University Park, PA 16802; and cNational Center for Genome Resources, Santa Fe, NM 87505

Edited* by Harris A. Lewin, University of California, Davis, CA, and approved June 17, 2013 (received for review December 5, 2012)

The male-specific region of the mammalian Y chromosome (MSY) contains clusters of genes essential for male reproduction. The highly repetitive and degenerative nature of the Y chromosome impedes genomic and transcriptomic characterization. Although the Y chromosome sequence is available for the human, chimpanzee, and macaque, little is known about the annotation and transcriptome of nonprimate MSY. Here, we investigated the transcriptome of the MSY in cattle by direct testis cDNA selection and RNA-seq approaches. The bovine MSY differs radically from the primate Y chromosomes with respect to its structure, gene content, and density. Among the 28 protein-coding genes/families identified on the bovine MSY (12 single- and 16 multicopy genes), 16 are bovid specific. The 1,274 genes identified in this study made the bovine MSY gene density the highest in the genome; in comparison, primate MSYs have only 31–78 genes. Our results, along with the highly transcriptional activities observed from these Y-chromosome genes and 375 additional noncoding RNAs, challenge the widely accepted hypothesis that the MSY is gene poor and transcriptionally inert. The bovine MSY genes are predominantly expressed and are differentially regulated during the testicular development. Synonymous substitution rate analyses of the multicopy MSY genes indicated that two major periods of expansion occurred during the Miocene and Pliocene, contributing to the adaptive radiation of bovids. The massive amplification and vigorous transcription suggest that the MSY serves as a genomic niche regulating male reproduction during bovid expansion.

gene amplification | evolution | transposition | expression

The genomics of the mammalian Y chromosome is poorly characterized compared with that of the X chromosome and autosomes because of the difficulties imposed by the abundance of repetitive sequences and the prevalent notion that the Y chromosome is degenerate with poor gene content and limited transcriptional potential (1). During evolution, the Y chromosome underwent progressive degeneration as a consequence of stepwise cessation of recombination that prevented exchange of sexual differentiation genes between the Y and X chromosomes. As a result, the majority (95%) of the present-day Y chromosome is a male-specific region that does not recombine with the X chromosome during meiosis, and only a small portion (5%) of the Y chromosome (the pseudautosomal region, PAR) maintains recombination with the X chromosome (1, 2). The male-specific region of the Y chromosome (MSY) contains two major regions, the X-degenerate region that comprises single-copy genes, representing evolutionary relics of the proto-sex chromosomes, and the ampliconic region that contains largely amplified gene families derived from either X-degenerate genes or autosomal genes acquired via transposition or retroposition (3). The biological significance and causative mechanism of the MSY gene amplification still are not fully understood.

The Y chromosome has evolved independently and accumulated lineage-specific genes (3–5). A majority of the Y-chromosome genes are expressed predominantly in testis, suggesting that they play a role in spermatogenesis and male fertility. Because the orthologs of the MSY genes, regardless of their chromosomal locations, also were implicated in spermatogenesis (6, 7), 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. To date, the Y chromosome has been sequenced only in the primates, with 27, 18, and 23 protein-coding gene families identified in the human, chimpanzee, and macaque MSY, respectively (3, 8, 9). Unique features identified in the primate MSY are palindrome, which provide a basis for Y-to-Y gene conversion (3, 8) that homogenizes the palindromic sequences and leads to structural polymorphisms, such as deletions. However, information about the Y chromosome in the other mammalian lineages is still lacking. A broader characterization of the Y-chromosome–linked genes is critical to our understanding of how the Y chromosome has shaped mammalian evolution.

Because of its unique phylogenetic position relative to the other mammals, Bos taurus, as a representative of ruminants, is important in studying the genetics of complex traits and evolution. Although the bovine genome has been sequenced (10), knowledge of the bovine Y chromosome (BTAY) is limited, because the genome sequence was from a cow. The bovine Y chromosome is the smallest chromosome in the genome, representing ~1.7% of the haploid genome (11). The size and morphology of the Y chromosome differ among the bovid lineages (12). BTAY is submetacentric, whereas the Y chromosome in the zebu (Bos indicus) and river buffalo (Bubalus bubalis) is acrocentric (13). Further cytogenetic characterization indicated that the bovid Y chromosome underwent rearrangements caused by centromeric transposition or pericentric inversion (12). Only eight genes on BTAY—DDX3Y, OFD1Y, PRAMEY, SRY, TSPY, HSFY, ZNF280AY, and ZNF280BY—have been characterized previously (14–22).

The clarification of the gene content and genomic features of the bovine MSY (bMSY) is essential for better understanding of the mammalian MSY gene dynamics. Here, we applied direct testis cDNA selection and RNA-seq approaches and found that the bMSY is rich in protein-coding genes and noncoding RNAs (ncRNA), contrary to the traditional perspective. The bMSY genes were expressed predominantly in testes and were up-c

—Author contributions: Y.Y. and W.-S.L. designed research; T.-C.C., Y.Y., and E.F.R. performed research; E.F.R. contributed new reagents/analytic tools; T.-C.C. analyzed data; and T.-C.C. wrote the paper.

The authors declare no conflict of interest.

This Direct Submission article had a prearranged editor.

Data deposition: The sequences reported in this paper have been deposited in the GenBank Database (accession nos. JX899360–JX899373 and KC662371).

†Present address: Department of Pathology and Laboratory Medicine, School of Medicine, University of North Carolina-Chapel Hill; Chapel Hill, NC 27599.

‡Deceased November 8, 2012.

§To whom correspondence should be addressed. E-mail: wul12@psu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221104110/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1221104110
regulated during testes development, suggesting their involvement in spermatogenesis. These genes were greatly amplified during two major periods of expansion during the Miocene and Pliocene, coincident with the adaptive radiation of bovids.

**Results**

**Bovine MSY Is Gene Rich.** To determine the gene content and transcriptomic activity of the Y chromosome, we performed BTAY transcriptome analyses through direct testis cDNA selection combined with testis RNA deep sequencing with replicates (n = 2) at three developmental stages: at postnatal day 20 (20d), at puberty (8 mo, 8m), and at maturity (2 y, 2y). The physical coverage of paired-end reads from the selected cDNA (6.7 million) was 32.7 Mb, covering 80.9% of the bMSY in the BTAY draft assembly (GenBank accession no. CM001061, Project ID: 20275) (Fig. 1). Alignment of the three testis RNA-seq pair-end reads (60.1 million) against the BTAY draft assembly produced a similar coverage (82.0%), confirming the reliability of our direct testis cDNA selection experiment (17). Together, these results suggest a large-scale transcriptional activity in BTAY.

The Y-chromosome–specific reads were assembled de novo into contigs and clustered as transcript units (TUs) (Table S1). After annotation, we identified 18 known genes/families, of which 12 were single-copy genes in the X-degenerate region and six were multicopy gene families in the ampliconic region (Fig. 1 and Table S2). Comparative analysis indicated that ZRSR2Y and RPL23AY are specific to cattle, UBE1Y and ELY233Y are found among nonprimates, and the rest are conserved in all mammals studied to date (Table S2) (23, 24). The BTAY draft assembly did not contain the RBMY gene. To position RBMY, we performed radiation hybrid (RH) mapping and found that RBMY is located at the distal Yq (Fig. 1).

We further examined the origin of the two bovid-specific genes, ZRSR2Y and RPL23AY. The orthologs of ZRSR2Y are present on the eutherian X chromosome, suggesting that ZRSR2Y was lost from the nonnvoid lineages during evolution. RPL23AY has an intronless X-chromosome–linked paralog (RPL23AX) with 90.7% sequence similarity and a highly similar autosomal paralog (99.3%) on BTA19 (NM_001045958). Because this autosomal paralog is located in a conserved syntenic block in eutherians and contains an exon–intron structure, the sex-linked RPL23AY/X may have been derived via recombination of the autosomal paralog.

The BTAY ampliconic gene families have two major evolutionary origins: X-degenerate genes, including TSPY, HSFY, and EGLY, of which TSPY and HSFY are conserved in some other mammalian species (25), and autosome-to-Y transposed genes, including ZNF280AY, ZNF280BY, and PRAMEY, which are bovid specific (Table S2) (17, 18). Except for PRAMEY and EGLY, these gene families have been amplified extensively on BTAY with a copy number ranging from 79 (ZNF280AY) to 234 (ZNF280BY) based on the sequenced Y chromosome (GenBank accession no. CM001061) (Table 1). We have demonstrated that the formation of the bMSY ampliconic region is based on a repeat unit containing ZNF280BY, ZNF280AY, HSFY, and TSPY families (17). The present analysis indicated that the copy numbers of the ampliconic gene families appear to increase by a factor of 80, suggesting that bMSY consists of ~80 repeat units (~420 kb per unit) and that each unit contains one to three copies of the corresponding gene families (17). Bovine TSPY and HSFY, with 136 and 192 copies, respectively, have been amplified to a greater extent than human TSPY (35 copies) and HSFY (2 copies) (3). The bovid-specific Y-chromosome gene families were derived from an autosome-to-Y chromosome transposition of a gene block on BTA17 followed by differential amplification on BTAY (17, 18). In contrast to the widespread amplification of ZNF280BY and ZNF280AY, PRAMEY was amplified to ~10 copies in a narrow region (~5 Mb on BTAY) where an array of 19 TSPY genes is present. This region appears to have a transitional feature between the X-degenerate and the ampliconic regions (Fig. 1). These results suggest that the bMSY is fundamentally different from the human MSY (hMSY) (3) in the genomic organization and composition.

An intronless gene, EGLY, was identified on the bMSY with three copies and encodes a protein with a retroviral envelope protein motif associated with transposable elements (TEs). EGLY has intronless homologs on the X chromosome (EGLX) and two autosomes (BTA14 and BTA24) with which it shares 90% and 99% sequence similarity, respectively. Because orthologous sequences of EGLY were identified in the other mammalian X chromosomes (tBLASTn), EGLY may be a surviving relic of the sex chromosome evolution. The emergence of the autosomal paralogs may result from the transposition of EGLY.

Hundred of BTAY-specific transcripts were identified, including 46 single-copy and 339 multicopy TU families. Ten of the multicopy TUs, named Bovid-specific Transcript, Y-linked (BfTY) 1–10, have coding potential based on their splicing signals, ORF

**Fig. 1.** Transcriptional landscape of the bovine MSY. (A) A gene map of bMSY. X-degenerate single-copy genes or transcripts (black) are clustered at either end of the bMSY, whereas multicopy genes/transcripts (red and orange) are present in the majority of the bMSY. RBMY was missing from the draft assembly. The relative position of RBMY (dashed line) was determined based on a RH-mapping analysis. (Note: The copy numbers of each gene or TU family are detailed in Table S1.) (B) The alignment of deep sequencing reads from the direct cDNA selection. About 80.9% of bMSY genes are transcriptionally active. The scale on the top of the ideogram is based on the BTAY draft assembly. The gray bars represent gaps in the assembly. (C) The predicted loci of six protein-coding, multicopy gene families. The PRAMEY gene family and the TSPY array (red) were amplified in a confined region (~5 Mb) near the boundary between the X-degenerate and ampliconic regions, representing a transitional region with a peculiar genomic context.
Table 1. Copy numbers of the protein-coding genes in the bovine MSY

<table>
<thead>
<tr>
<th>MSY region</th>
<th>Gene</th>
<th>Peptide length</th>
<th>Copy no.</th>
<th>No. of copies with coding potential*</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-d</td>
<td>EIF1AY</td>
<td>143</td>
<td>1</td>
<td>1</td>
<td>NM_001145757</td>
</tr>
<tr>
<td></td>
<td>OFD1Y</td>
<td>875</td>
<td>1</td>
<td>1</td>
<td>JN193532</td>
</tr>
<tr>
<td></td>
<td>USP9Y</td>
<td>2552</td>
<td>1</td>
<td>1</td>
<td>NM_001145509</td>
</tr>
<tr>
<td></td>
<td>ZRS2Y</td>
<td>130</td>
<td>1</td>
<td>1</td>
<td>JX899373</td>
</tr>
<tr>
<td>UTY</td>
<td>141</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>XM_002700446</td>
</tr>
<tr>
<td></td>
<td>DDX3Y</td>
<td>660</td>
<td>1</td>
<td>1</td>
<td>NM_001172595</td>
</tr>
<tr>
<td></td>
<td>RPL23AY</td>
<td>158</td>
<td>1</td>
<td>1</td>
<td>JX899372</td>
</tr>
<tr>
<td></td>
<td>ZFY</td>
<td>801</td>
<td>1</td>
<td>1</td>
<td>NM_177491</td>
</tr>
<tr>
<td></td>
<td>EIF23YY</td>
<td>473</td>
<td>1</td>
<td>1</td>
<td>XM_002700462</td>
</tr>
<tr>
<td></td>
<td>UBE1Y</td>
<td>158</td>
<td>1</td>
<td>1</td>
<td>KC662371</td>
</tr>
<tr>
<td>Y-a</td>
<td>HSFY</td>
<td>94–417</td>
<td>192</td>
<td>190</td>
<td>NM_001077006</td>
</tr>
<tr>
<td></td>
<td>TSPY</td>
<td>96–355</td>
<td>136</td>
<td>134</td>
<td>NM_001244608</td>
</tr>
<tr>
<td></td>
<td>PRAMEY</td>
<td>511–512</td>
<td>10</td>
<td>5</td>
<td>GU144301</td>
</tr>
<tr>
<td></td>
<td>ZNF280AY</td>
<td>101–263</td>
<td>79</td>
<td>79</td>
<td>HQ014564</td>
</tr>
<tr>
<td></td>
<td>ZNF280BY</td>
<td>114–543</td>
<td>234</td>
<td>234</td>
<td>NM_001078120</td>
</tr>
<tr>
<td></td>
<td>EGLY</td>
<td>122</td>
<td>3</td>
<td>3</td>
<td>JX899360</td>
</tr>
<tr>
<td></td>
<td>BTY1Y</td>
<td>105</td>
<td>4</td>
<td>1</td>
<td>JX899361</td>
</tr>
<tr>
<td></td>
<td>BTY2Y</td>
<td>103</td>
<td>2</td>
<td>1</td>
<td>JX899362</td>
</tr>
<tr>
<td></td>
<td>BTY3Y</td>
<td>120</td>
<td>78</td>
<td>76</td>
<td>JX899363</td>
</tr>
<tr>
<td></td>
<td>BTY4Y</td>
<td>148</td>
<td>83</td>
<td>60</td>
<td>JX899364</td>
</tr>
<tr>
<td></td>
<td>BTY5Y</td>
<td>110</td>
<td>87</td>
<td>43</td>
<td>JX899365</td>
</tr>
<tr>
<td></td>
<td>BTY6Y</td>
<td>135</td>
<td>96</td>
<td>92</td>
<td>JX899366</td>
</tr>
<tr>
<td></td>
<td>BTY7Y</td>
<td>152</td>
<td>174</td>
<td>82</td>
<td>JX899367</td>
</tr>
<tr>
<td></td>
<td>BTY8Y</td>
<td>95</td>
<td>146</td>
<td>65</td>
<td>JX899368</td>
</tr>
<tr>
<td></td>
<td>BTY9Y</td>
<td>91</td>
<td>98</td>
<td>92</td>
<td>JX899369</td>
</tr>
<tr>
<td></td>
<td>BTY10Y</td>
<td>95</td>
<td>117</td>
<td>105</td>
<td>JX899370</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,547, 1,274</td>
</tr>
</tbody>
</table>

X-d, X-degenerate region; Y-a, Y-ampliﬁc region.

*Estimated from the bovine Y chromosome draft sequence assembly (GenBank accession no. CM001061).

length, and motif analysis (Table 1). The remaining TUs were classified as ncRNAs. Similar to other bMSY gene families, the BTYs were ampliﬁed to 2–174 copies. These BTY proteins share similarity with putative proteins with unknown function in remote metazoans.

Collectively, 1,274 protein-coding genes were identiﬁed and distributed over 40.8 Mb of the bMSY, with a gene density of ~31.2 genes/Mb; this density is signiﬁcantly higher than the ~9.4 genes/Mb for BTAX and ~10.2 genes/Mb for the entire genome [University of Maryland (UMD) 3.1 assembly].

Transcriptomic Activity Is Vigorous in the Bovine MSY. The RNA-seq reads from the three different developmental stages (20d, 8m, and 2y) were aligned to the entire bovine genome sequence assembly (UMD 3.1 assembly) to determine the expression level of all transcripts in the testes. More than 95% of the transcripts on the bMSY were transcriptionally active, with ZNF280BY being the most highly expressed gene family across the three stages. The majority of the bMSY transcriptome, including 13/28 protein-coding genes and 220/375 TUs, were up-regulated during testis development. A total of 38, 106, and 17 MSY genes/TUs were signiﬁcantly differentially expressed in the 20d vs. 8m, 20d vs. 2y, and 8m vs. 2y comparisons, respectively (P < 0.05) (Fig. 2A and Fig. S1). To examine the functional association of the TUs further, we performed a hierarchical clustering of all of the signiﬁcantly diﬀerentially expressed genes, including 108 bMSY genes/TUs, 426 X-chromosome-linked genes, and 7,026 autosomal genes. This analysis revealed 5 major patterns of gene expression (Fig. 2B). Pattern I comprises genes/TUs that are signiﬁcantly up-regulated from 20 d to 8 m and 2 y. This group includes 31% of the diﬀerentially expressed bMSY genes/TUs and is enriched with Gene Ontology (GO) terms of reproduction and spermatogenesis, indicating their involvement in sexual maturation and spermatogenesis (Table S3). Pattern II genes/TUs, including 36% of the bMSY genes/TUs, are highly expressed in 8m testis and are enriched with GO terms of ion transport and cell growth. Pattern III-V genes/TUs, containing the remaining 3% of bMSY genes/TUs, are enriched with GO terms associated with diverse biological process, such as the mRNA metabolism and signaling pathway. These results indicate that the diﬀerentially expressed genes/TUs are associated mainly with spermatogenesis and male reproduction.

To validate our RNA-seq data, we examined temporal and spatial expression patterns of 56 genes/TUs by RT-PCR in five developmental stages of the bovine testis and 11 other adult tissues. Among the genes tested, 49 were conﬁrmed to have a male-speciﬁc expression pattern, and 42/49 were expressed in testis (Table S4); the remaining seven were not detected in testis, perhaps because of a low expression level. The seven TUs without a male-speciﬁc expression pattern were ampliﬁed with multiple bands from both male and female genomic DNA and re-run for further study. Furthermore, 55% (33/60) of the differentially expressed genes/TUs were up-regulated with age, supporting our RNA-seq analysis. The spatial expression analyses revealed that the bMSY genes were exclusively or predominantly expressed in the bovine testes (Table S4), reinforcing their roles in spermatogenesis.

Bovine MSY Gene Families Underwent Two Major Expansions. An evolutionary model is proposed in Fig. 3 to delineate the evolution of the X-degenerate and Y-ampliﬁcic genomic on BTAY. The extensive expansion of the MSY genes is distinctive in BTAY. To understand further how the Y-chromosome gene families were expanded, we performed pairwise synonymous substitution rate (Ks) analyses for HSFY, TSPY, and ZNF280BY, the three most ampliﬁed protein-coding gene families in bMSY (Table 1). By plotting the frequency of the gene duplication against the duplication time, we found a pattern following a bi-modal distribution for all of the three families, suggesting that at least two major expansions of the BTAY ampliﬁcic region occurred during evolution (Fig. 4). An earlier expansion occurred ~14–20 Mya during the Miocene (26), and a later expansion occurred within 5 Mya during the Pliocene. We found that the ampliﬁcation of the HSFY began ~18.1–20.0 Mya, much later than the ampliﬁcation of TSPY and ZNF280BY (~35 Mya), and coincided with the divergence between cattle and sheep (~19.2 Mya) (27, 28). These results suggest that the ampliﬁcation of the bMSY gene families may be involved in the diversiﬁcation of the Bovidae. Although gene conversion may lead to sequence homogenization, the congruent Ks distributions from three large ampliﬁed gene families indicate that gene duplication is a major mechanism accounting for the distinct Ks clusters.

Discussion

In this study we comprehensively examined the transcriptional activity on a mammalian Y chromosome using RNA-seq. We demonstrated that bMSY is rich in genes/transcripts that are transcriptionally dynamic and vigorous during testis development. These ﬁndings directly contradict the traditional view that the Y is largely heterochromatic with a paucity of genes and transcriptional activity (1). The bMSY evidently has a much higher gene density (31.2 genes/Mb) than the bovine, human, and mouse genomes (10.2, 12.6, and 12.8 genes/Mb, respectively). Although the bMSY and hMSY comprise similar numbers of protein families (28 and 27, respectively), the bMSY harbors eightfold more gene loci than the hMSY because of the distinct MSY genomic structure. The hMSY contains eight palindromes with two to six loci of coding genes in each palindrome, whereas
the bMSY contains a total of ~80 palindromic-like repeat units, each containing 4–12 loci of coding genes (17). These results suggest that the genomic amplification of MSY is an essential mechanism shaping the evolution of the Y chromosome.

The structural differences between the bMSY and hMSY observed in this study support the addition and attrition theory of mammalian Y chromosome evolution (2). The independent decay of the Y chromosome has resulted in gene loss in a lineage-dependent manner (Table S2). For the X-degenerate genes, ZRSR2Y is present only in the BTAY and has been degenerated in the nonbovid lineages during evolution. The OFD1Y gene is an active gene in bovids but is pseudogenized in humans (29). The EIF2S3Y and UBE1Y genes are conserved in most mammals, except for primates (3). Compared with the X-degenerate genes, the ampiconic gene content is more diversified among lineages. Except for the HSFY and TSPY gene families that were differentiated from the X-chromosome-linked counterparts before the eutherian radiation and are amplified in the Y chromosome of most mammalian species (25), the majority of the ampiconic genes originated either from a lineage-degenerate differentiation or from lineage-specific, autosome-to-Y transposition (Fig. 3), resulting in a diversified gene content in different lineages. As detailed in Table S2, lineage-specific MSY gene families are present in all species examined to date (e.g., DAZ and CDY in humans (30, 31), Soy1 and Hsy in mice (32); TET1Y1 and FLJ35031 in cats (5); and ZNF280BY, ZNF280AY, PRA1MEY, EGLY, and BTY1-10 in cattle (17, 18).

Compared with the primate MSY in which 3–19 palindromes of variable size (17.4 kb–2.9 Mb) are located (3, 8, 9), the bMSY is distinctive by having ~80 repeat units of similar size (420 kb) in this particularly sequenced Y chromosome. This finding is supported by previous FISH studies with probes containing the three most amplified bMSY gene families, HSFY, TSPY, and ZNF280BY (also known as 5’OY11.1) (14, 20, 21, 33). FISH signals were observed in the proximal region of the short arm (Yp) and the long arm (Yq) of the bovid Y chromosome, where the ampiconic region is located (Fig. 1). Strong TSPY signals were observed in the proximal region of Yp (14, 22), revealing the location of the TSPY array (Fig. 1) and supporting the draft sequence assembly and a previous report about the different clusters of the TSPY subfamilies (21). It is important to point out that the number of repeat units on the BTAY varies among individuals and among populations, as evidenced by previous studies on the copy number variations of TSPY, ranging from 45 to ~200 (21, 22).

Although lineage-specific noncoding transcripts have been reported on MSY of a few species, such as the 28 ncRNAs in the human (3) and at least three ncRNAs in the horse (4), the identification of 375 ncRNAs in the bMSY is surprising. The expression of 21 ncRNAs was validated (Tables S4 and S5) and consistently showed a predominant expression in testis. However, the biogenic mechanisms and function of these ncRNAs remain unknown. Previous studies have suggested that the ncRNAs in testis are involved mainly in gene regulation at the post-transcriptional and translational levels during spermatogenesis (34, 35). The activities of several autosomal ncRNAs have been associated with the initiation and maintenance of meiotic sex chromosome inactivation, an important gene-silencing process that is established at midpachytene stage and persists to the postmeiotic phase of spermatogenesis (36). In addition, recently it has been proposed that extensive interactions between ncRNAs and protein-coding genes control gametogenesis during meiotic development in yeast (37). Because spermatogenesis is a tightly regulated process of cell proliferation and differentiation, we anticipate that the ncRNAs in the bMSY may confer a layer of delicate gene regulation during spermatogenesis.

![Fig. 2. Expression of the bovine MSY genes/transcripts in testes. (A) Volcano plot of the differential gene expression between 20d and 2y testes. A total of 106 BTAY genes or TUs were found to be differentially expressed. The x-axis represents log twofold change; the y-axis represents −log10 (P < 20). The genes, including HSFY and ZNF280BY, with a −log10 P value >20 are excluded from the plot. The red dots represent Y-chromosome–linked genes or TUs. Blue dots represent autosomal and X-chromosome–linked genes. The dashed orange line indicates P = 0.05; dots above the line have P < 0.05. The genes or TUs with a fold-change <2 are shown in gray. (B) Clustering and heatmap of differentially expressed Y-chromosome–linked genes or TUs. A hierarchical clustering of the differentially expressed genes revealed five major patterns (I–V). The heatmap of differentially expressed genes based on clustering is shown, in which each column represents the age of tests (20d, 8m, or 2y), and each row represents a gene or TU. Data were normalized to z-scores between −1 and 1 for each gene or TU. Red and blue indicate an increase or decrease, respectively, in gene expression compared with the universal mean (white) for each gene or TU. The differentially expressed bMSY genes or TUs were clustered mainly in pattern I, which is up-regulated in 8m/2y testes relative to 20d testes. Pattern II genes or TUs have the highest expression; pattern III has the lowest expression in 8m tests. In contrast to pattern I, the genes or TUs in patterns IV and V were down-regulated. GO enrichment analyses for each pattern are detailed in Table S3.](https://www.pnas.org/doi/10.1073/pnas.1221104110)

![Fig. 3. The evolutionary model of the bovine Y chromosome. The bovine Y chromosome evolved from a pair of ancestral autosomes. Genes on the BTAY originated through two major evolutionary paths. First, the X-degenerate genes evolved through gradual differentiation of the Y chromosome from the X chromosome, as explained by the addition and attrition theory (2). Second, selection forces facilitated the acquisition of autosomal, male-beneficial genes by the BTAY, in which the transposable elements have acted for the movement and amplification of the bMSY genes. The amplification occurred in both X-degenerate genes (EGLY, TSPY, and HSFY) and genes acquired from autosomes (ZNF280AY, ZNF280BY, and PRAMEY). Two subpaths diverged as a result of different strengths of gene amplification, in which TSPY, HSFY, ZNF280AY, and ZNF280BY were amplified extensively. The extensive amplification of the bMSY genes in the Pliocene and Miocene is associated with the adaptive diversification of Bovidae (see text and Fig. 4).](https://www.pnas.org/doi/10.1073/pnas.1221104110)
Despite the diverged gene origins and contents, the amplified genes in the MSY, whether conserved or lineage specific, are predominantly or exclusively expressed in testis and are implicated in male fertility (4, 5, 38, 39). Our results revealed that the majority of the bMSY genes and ncRNAs are up-regulated during testis development and are enriched with GO terms in spermatogenesis, reinforcing their functional involvement in male reproduction. Of particular interest are the EGLY gene family and 11 ncRNAs that all contain motifs associated with TEIs (Table S1). Because TEIs are involved in genomic duplications (40) and biogenesis of novel protein-coding genes and small RNAs (41, 42), these genes and transcripts may be essential for the expansion of the palindromic-like repeat units and the amplification of the multicopy genes on the bMSY. These data not only support the notion that the MSY accumulates genes crucial for male reproduction but also offer an explanation for the “selfish Y chromosome” theory (43, 44). Further exploration of the Y-chromosome gene content in the other mammalian lineages will disclose a broader diversity of lineage-specific Y-chromosome genes and a set of core genes involved in spermatogenesis and male fertility.

The massive amplification of the Y-chromosome gene, Sly, was reported in the mouse (45, 46). The biological role of the amplified genes was proposed to be either resisting Y-chromosome degeneration (47) or maintaining stable protein expression to achieve reproductive success (46). However, the mechanism behind the expansion of these Y-chromosome genes and related ampliconic sequences in the MSY is largely unknown. The gene expansion pattern in the bMSY was similar to that in the mouse MSY but not the hMSY. In the present study, we examined the most amplified three bMSY gene families and found two important periods of expansion. During these periods, major diversification of bovid species occurred. The first stage occurred during the Miocene (~14–20 Mya), when the climate became cooler and drier, and grasslands replaced forests as the dominant habitat in many regions, leading to the selection and divergence of large herbivores, including the ruminants (26). This timing is in agreement with a previous report that estimated that the Bovinae diverged from Antilopinae in the early Miocene (~22.6 Mya (26). It implies that the early expansion of the bovine ampliconic region may be involved in the adaptation and speciation of the bovid lineages. The later expansion, beginning 5 Mya in the Pliocene (~2–5 Mya), may have initiated another round of adaptive processes during which Bovidae started to dominate mammalian fauna. An increase in species richness in Bovidae was observed ~2.8 Mya, followed by changes in abundance between 2.8 and 2.0 Mya (48). Therefore, the recent expansion of the bovine ampliconic region is potentially associated with the increased diversification of Bovidae and may be essential for their reproductive isolation. We believe that the massive amplification along with vigorous transcription make the bMSY a unique genomic niche to regulate male reproduction during the bovid expansion.

**Materials and Methods**

**Collection of Bovine Tests and Other Tissue Samples and RNA Extraction.** All procedures were approved by the Institutional Animal Care and Use Committee at Penn State University. Bovine testes were collected from a local farm (in State College, PA) during bull castration at the age of 4 d, 20 d, 3 mo, and 8 mo. Mature testes and 10 other tissues (liver, kidney, spleen, cerebellum, adrenal gland, longissimus muscle, lymph node, semitendinosus, spinal cord, and lung) were collected from a 2-y-old bull. Ovarian tissue from a mature cow was collected from a slaughter house (State College, PA). Total RNAs were extracted from these samples as described (17).

**Direct Tests cDNA Selection and Sequencing.** Direct tests cDNA selection is detailed in Yang et al. (17). The BTAY DNA was isolated by microdissection (49). The DNA fragments were amplified by PCR and labeled with biotin-16-dUTP by nick translation (Roche). The selected cDNAs were PCR-amplified using the biotin-labeled primer (17). Selection efficiency was assessed by quantitative PCR with a Y-chromosome–linked gene, DDX3Y, as a positive control and β-actin and CYCL as negative controls. The selected cDNAs were sequenced at the Pennsylvania State University Genomics Core Facility using an ABI-3730XL DNA analyzer (17) and at the National Center for Genome Resources using an Illumina GAIIx (Illumina). Library construction and sequence methods were described previously (17). A total of 6,710,574 high-quality paired-end reads of 2 × 36 bp from direct cDNA selection were generated and were assembled de novo into 1,438 contigs by the AbySS program (50).

**RT-PCR.** Total RNAs from five stages (4d, 20d, 3m, 8m, and 2y) of testes, an adult ovary, and 10 somatic tissues were treated with DNase I (Ambion) and were reverse transcribed using the SuperScript III First-Strand Synthesis System (Invitrogen). RT-PCR was performed in a volume of 20 μL containing 10 ng cDNA, 200 μM dNTPs, 1.5 mM MgCl₂, 2.5 μM of each primer, and 1 unit Taq DNA polymerase (Biolime). The PCR conditions were 94 °C for 7 min followed by 35 cycles each of 95 °C for 40 s, 60–65 °C for 40 s, and 72 °C for 30 s, with a final extension at 72 °C for 7 min. PCR products were resolved on 1.5% agarose gels with ethidium bromide in 1× Tris-acetate-EDTA buffer.

**Sequence Analysis.** The assembled contigs were first blasted against different databases using BLAST (53), resulting in two sets of transcripts. One set comprised the transcripts matched to orthologous genes in other species or paralogous genes in bovids, which were annotated based on the information obtained from the orthologs or paralogs. The other set comprised boid-specific transcripts. The contigs were clustered into TUs using CD-HIT (parameters: –c 0.8 –g 1 –r 1 –as 0.8 –n 4) (54). The longest sequence of each TU was used as a seed sequence for predicting genes or paralogs on BTAY by aligning them against the BTAY draft assembly (GenBank accession no. CM001061) using Splign based on a compartmentalization and a refined alignment algorithm (55). The motif analyses were conducted using Interpro (56). The ORFs of the TUs were predicted using Getorf (57), Repeat elements were identified and filtered by RepeatMasker (58). The read coverage and genomic locations of TUs were visualized by Integrative Genomics Viewer (59) with customized tracks. Differential expression of genes was examined using an Illumina GAIIx platform.

**Fig. 4.** Age distribution of the paralogs in the bovine TSPY, HSFY, and ZNF280B gene families. Pair-wise Ks values for all loci of the bovine TSPY, HSFY, and ZNF280B gene families were calculated, and the gene duplication times were estimated. The gene duplication frequencies (y-axis) were plotted against the age of the gene duplication (x-axis). The plot displays a bi-modal distribution with peaks falling into two major time periods: 0–5 and 14–20 Mya (highlighted in gray), corresponding to the Pliocene and Miocene epochs, respectively. The red dashed line denotes the divergence of sheep and cattle, and around 19.2 Mya.
p1 and p2 are the paralogous gene pairs, and o1 and o2 are orthologous gene pairs between human (TSPY, HSFY, ZNF280B) and bovine (HSFY, HSFY, ZNF280BY).

**ACKNOWLEDGMENTS.** We thank Dr. Abel Ponce de León for providing the microdissected BTAY DNA fragments; Dr. John A. Crow for assistance in contig assembly of the deep sequencing reads; the Human Genome Sequencing Center at Baylor College of Medicine and Dr. David Page’s group at the Microdissected Chromosome Technology for the BTAY draft assembly; and Dr. Cooudavali Shashikant for comments on the manuscript. This work was supported by US Department of Agriculture-National Institute of Food and Agriculture Grant 2010-65205-20362 and by start-up funds from the Pennsylvania State University (to W.-S.L.).