Sequence analysis in *Bos taurus* reveals pervasiveness of X–Y arms races in mammalian lineages


1Whitehead Institute, Cambridge, Massachusetts 02142, USA; 2Howard Hughes Medical Institute, Whitehead Institute, Cambridge, Massachusetts 02142, USA; 3College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas 77843, USA; 4Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA; 5The McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri 63108, USA; 6Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

Studies of Y Chromosome evolution have focused primarily on gene decay, a consequence of suppression of crossing-over with the X Chromosome. Here, we provide evidence that suppression of X–Y crossing-over unleashed a second dynamic: selfish X–Y arms races that reshaped the sex chromosomes in mammals as different as cattle, mice, and men. Using super-resolution sequencing, we explore the Y Chromosome of *Bos taurus* (bull) and find it to be dominated by massive, lineage-specific amplification of testis-expressed gene families, making it the most gene-dense Y Chromosome sequenced to date. As in mice, an X-linked homolog of a bull Y-amplified gene has become testis-specific and amplified. This evolutionary convergence implies that lineage-specific X–Y coevolution through gene amplification, and the selfish forces underlying this phenomenon, were dominantly powerful among diverse mammalian lineages. Together with Y gene decay, X–Y arms races molded mammalian sex chromosomes and influenced the course of mammalian evolution.

[Supplemental material is available for this article.]

Mammalian sex chromosomes evolved from an ordinary pair of autosomes (Ohno 1967; Lahn and Page 1999; Bellott et al. 2010). Differentiation of the X and Y Chromosomes began 200–300 million years ago, propelled by gradual suppression of X–Y crossing-over, most likely resulting from inversions on the Y Chromosome (Lahn and Page 1999; Ross et al. 2005). Although the X Chromosome continues to participate in crossing-over with its homolog during female meiosis, the male-specific region of the Y Chromosome (MSY) lacks a partner for crossing-over during male meiosis. Theoretical and empirical studies of Y Chromosome evolution have largely focused on one specific outcome of crossover suppression—genetic decay (for review, see Bachtrog 2013). Gene loss is a nearly universal feature of MSY evolution and has been reported in species as diverse as plants (papaya, white campion, and heartwing sorrel), flies, fish, frogs, lizards, and mammals (Liu et al. 2004; Chibalina and Filatov 2011; Kaiser et al. 2011; Miura et al. 2012; Zhou et al. 2012; Bellott et al. 2014; Hough et al. 2014; Lisachov et al. 2019; Peichel et al. 2020).

Complete sequencing of four mammalian Y Chromosomes—human, chimpanzee (*Pan troglodytes*), rhesus macaque (*Macaca mulatta*), and mouse (*Mus musculus*)—revealed that MSY decay has been counterbalanced to varying degrees by sequence acquisition and amplification (Skaletsky et al. 2003; Hughes et al. 2010, 2012; Soh et al. 2014). The super-resolution sequencing method used for these MSY sequences is a requirement for elucidating the complex structure of ampliconic sequences, which are euchromatic repeats that display >99% identity over >10 kb. More limited sequencing in other species, ranging from *Drosophila miranda* to mammals, suggests that MSY amplification may have been widespread (Skinner et al. 2015; Tomaszkiewicz et al. 2016; Brashear et al. 2018; Janečka et al. 2018; Bachtrog et al. 2019).

The mouse MSY stands in stark contrast to the three primate MSYs because it has been overtaken by ampliconic sequence. Nearly 98% of the mouse MSY is ampliconic, compared to 45%, 57%, and 5% in human, chimpanzee, and rhesus, respectively. The amplified mouse sequence bears no homology with any primate MSY sequence and contains three gene families that are expressed exclusively in the male germline. These MSY amplified genes have independently acquired and amplified germline-

Present addresses: 7Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA; 8Institute for Genomic Medicine, Nationwide Children’s Hospital, Columbus, OH 43205, USA; 9College of Food and Agriculture, United Arab Emirates University, Dubai, United Arab Emirates

Corresponding author: jhughes@wi.mit.edu

Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.269902.120.
specific gene families on the mouse X Chromosome. It has been hypothesized that mouse X–Y coamplification may be a manifestation of sex-linked meiotic drive, in which the X and Y Chromosomes compete against each other for transmission to the next generation (Ellis et al. 2011; Cocquet et al. 2012; Soh et al. 2014).

Sex-linked meiotic drive has been reported in dozens of insect species, but in mammals, this phenomenon has only been observed in a few rodent species (Jaenike 2001). In natural mouse populations, autosomes are known to engage in meiotic drive, with the t-haplotypes of Chromosome 17 harboring a strong segregation distorer (Silver 1985). It is not known if insects and rodents are particularly susceptible to the emergence of meiotic drive systems or if these species are simply studied more thoroughly in this context.

Here, we present the complete MSY sequence of the bull (Bos taurus, Hereford breed). This species represents a third major branch of the mammalian tree (Fig. 1A) and is of great agricultural, anthropological, and economic importance. The bull and mouse lineages diverged from each other nearly 100 million years ago, and bull and mouse differ profoundly in life histories and reproductive strategies. However, their MSYs show structural convergence, which has important implications for our understanding of mammalian MSY evolution and biology.

Results

Mapping, sequencing, and assembly of the bull MSY

We produced and assembled a nearly complete sequence of the bull MSY (Fig. 1B; Supplemental Fig. S1; Supplemental File S1), accurate to one nucleotide per megabase. To achieve this high level of completeness and accuracy, we used the super-resolution single-haplotype iterative mapping and sequencing (SHIMS) methodology (Bellott et al. 2018), which we previously used to sequence sex chromosomes in human, chimpanzee, rhesus macaque, mouse, and chicken (Skaletsky et al. 2003; Bellott et al. 2010, 2017; Hughes et al. 2010, 2012; Mueller et al. 2013; Soh et al. 2014). We assembled a total of 40 Mb of sequence from a tiling path of 349 bull bacterial artificial chromosome (BAC) clones, all derived from a single male donor (L1 Domino, Bos taurus, Hereford breed) (Supplemental Table S1). The assembly contains 15 gaps, 12 of which are associated with heterochromatin or tandem arrays. We used fluorescence in situ hybridization (FISH) to visualize these tandem arrays, to order and orient sequence contigs within the assembly, and to estimate gap sizes within single-copy regions (Supplemental Figs. S2, S3). Where possible, we also confirmed linkage between contigs using radiation

Figure 1. Structure and ampliconic sequence content of bull, mouse, and primate Y Chromosomes. (A) Phylogenetic relationships among five species with SHIMS-assembled MSY sequences. Branch lengths are drawn to scale to indicate species divergence times. Nonhuman photos from Unsplash .com. (B–D) Triangular dot plots, to scale, of DNA sequence identities within the euchromatic MSYs of bull (B), mouse (C), and primates (human, chimpanzee, and rhesus) (D). Each dot represents 100% intrachromosomal identity within a 200-bp window. Direct repeats appear as horizontal lines, inverted repeats as vertical lines, and palindromes as vertical lines that nearly intersect the baseline. Below plots, schematic representations of chromosomes are shown. Sequence classes are color coded as indicated. (Other) Single-copy male-specific sequences that are not homologous to the X Chromosome. (E) Bull and mouse MSYs are represented as spiral diagrams drawn to scale. Major repeat units in bull and mouse long-arm amplicons are color coded as indicated. Repeat units between species are unrelated.
that spans 44 Mb (Supplemental File S1); all our subsequent analyses are based on this sequence assembly.

Comparison of five sequenced MSYs reveals convergent evolution in bull and mouse

Our analysis of the mouse MSY revealed that it followed an evolutionary path distinct from that of the three sequenced primate MSYs: human, chimpanzee, and rhesus macaque. The primate MSYs are by far the smallest chromosomes in their respective genomes; their euchromatin sequence content ranges from 12 Mb in rhesus to 26 Mb in chimpanzee. In contrast, the euchromatic sequence content of the mouse MSY totals nearly 90 Mb (Soh et al. 2014). The size disparity between the mouse and primate MSYs is attributed solely to massive amplification of lineage-specific sequences in mouse, because ancestral X-homologous regions of the MSY, presumed to be present in the mammalian ancestor, are markedly reduced in mouse. With only four complete MSY sequences in hand, it was impossible to know if this rampant MSY sequence expansion is a peculiarity of the mouse lineage.

The bull MSY sequence suggests that massive MSY sequence amplification may be a broad characteristic of mammalian MSYs. The bull MSY is more similar in structure to the mouse MSY than to any of the primate MSYs (Fig. 1B–D). Single-copy X-homologous regions in bull and mouse account for only 4% and 2% of their respective MSY euchromatin contents, compared to 37%, 33%, and 86% in human, chimpanzee, and rhesus, respectively. The long arm of the bull MSY, which spans roughly 35 Mb, consists almost entirely of ampliconic sequence, with four basic repeat units (Fig. 1E; Supplemental Fig. S4). The mouse MSY long-arm ampliconic sequence is 86 Mb in length and contains three basic repeat units (Fig. 1E). There is no homology between the ampliconic sequences in bull and mouse, indicating that these amplifications were independent. The bull amplicons contain two distinct multicopy gene families: HSFY, which is found in primates and originates from the ancestral autosomal pair that gave rise to the mammalian X and Y (Hamilton et al. 2011), and ZNF280BY, which is lineage-specific and originates from an autosomal-to-Y transposition event (Yang et al. 2011). The bull MSY amplicons also contain about 400 inactivated pseudogenes (Supplemental Figs. S2, S5). In comparison, the mouse MSY amplicons contain three lineage-specific gene families that were acquired by the Y Chromosome during rodent evolution. In total, the mouse MSY has about 700 protein-coding genes and more than 1000 pseudogenes. The large number of pseudogenes found in bull and mouse is a consequence of the amplification of one or a few inactivated genes along with intact genes in the same repeat unit rather than multiple inactivation events. We conclude that the MSYs in the bull and mouse lineages experienced extensive but independent amplifications.

In all of the sequenced MSYs, ampliconic regions display extremely high inter-repeat nucleotide identities (up to 99.99%) because of ongoing gene conversion (as well as other forms of homologous recombination) between repeat units (Rozen et al. 2003). Both the bull and mouse MSYs have ample substrates for gene conversion, so we generated maps of intrachromosomal identities across these chromosomes to visualize the extent of such sequence homogenization in bull and mouse (Fig. 2). Overall, the long-arm amplicons in bull are more homogeneous than in mouse, as evidenced by a higher density of >99.9% matches in the intrachromosomal identity map (Fig. 2). Compared to human (as a representative primate), both bull and mouse MSY regions show high sequence identity spanning great distances (often tens of megabases), implying that gene conversion is not constrained by physical distance between substrates (Fig. 2).

Although gene conversion, which is homologous recombination without crossing-over, between MSY amplicons results in sequence homogenization, homologous recombination with crossing-over can create inversions, duplications, and deletions, with duplications and deletions in ampliconic regions giving rise to copy number variation (Lange et al. 2013). To examine copy number variation within bull MSY amplicons, we analyzed more than 200 publicly available bull whole-genome shotgun sequences from four different breeds (Daetwyler et al. 2014). Using read-depth mapping of raw sequencing data, we determined that copy number variability is indeed prevalent among bull MSY amplicons (Supplemental Fig. S6). Among Holstein cattle, for example, the copy number of the TSPY2-PRAME2 array varies nearly fivefold, and the copy number of the long-arm ampliconic sequence varies nearly threefold.

Evolutionary conservation of bull MSY ampliconic regions

The MSYs of bull and mouse differ radically in structure from the sequenced primate MSYs; one common feature that distinguishes these two disparate species from the three primates is domestication. Previous studies have shown that wild Mus species carry amplified sequences homologous to the long-arm amplicons found on the sequenced MSY (from Mus musculus musculus, strain C57BL/6) (Scavetta and Tautz 2010; Ellis et al. 2011; Soh et al. 2014), ruling out the possibility that amplification was a consequence of selective breeding. Similarly, we sought to determine if the bull MSY amplification predated the domestication of cattle. We used FISH analysis to examine six different ampliconic regions on the bull MSY (Fig. 3A), including the long-arm amplicon, across eight bovine species spanning 17 million years of evolution (Loftus et al. 1994; Bradley et al. 1996; Hernandez Fernández and Vrba 2005). The long-arm amplicon, as well as three other tandem arrays, are conserved and amplified on the MSY in all eight species (Fig. 3B). Although the arrangement and size of the amplicons differs between distantly related species, the presence of these sequences in diverse species of wild bovids confirms that these amplification events are not a consequence of selective breeding by humans, but instead occurred in a common ancestor of these species. Large-scale repeats, such as those that compose the bull MSY sequence, are prone to rearrangement, so the conservation of these evolutionarily volatile sequences across millions of years of evolution implies the action of purifying selection.

Gene content of bull MSY

A second feature that distinguishes the bull and mouse MSYs from those of primates is their high gene densities, a consequence of the massive amplification of genes within their ampliconic regions (Fig. 4A,B; Supplemental Fig. S5; Supplemental Table S3; Soh et al. 2014). In bull, the MSY gene density approaches the genome average of 10 per megabase (The Bovine Genome Sequencing and Analysis Consortium et al. 2009), counter to previous generalizations that Y Chromosomes are poor in genes (Mank 2012). Bull and mouse have fewer ancestral genes (13 and 9 unique genes,
respectively) than human and rhesus (17 and 18 genes, respectively) (Bellott et al. 2014). Chimpanzee contains only 13 intact ancestral genes because of lineage-specific gene losses (Hughes et al. 2005; Perry et al. 2007). The bull pseudoautosomal region (PAR) is much larger than the PARs of primates (Das et al. 2009), and three genes that are male-specific in human remain pseudoautosomal in cattle, accounting for some of the disparity in MSY ancestral gene content.

In all previously sequenced MSYs (human, chimpanzee, rhesus, and mouse), we found that multicopy genes were expressed exclusively or predominantly in testes, but most single-copy ancestral genes were found to be expressed more broadly (Skaletsky et al. 2003; Bentley et al. 2008; Hughes et al. 2010, 2012). Y-linked testis-specific genes likely have specialized functions related to spermatogenesis, whereas the broadly expressed genes have more widespread functions related to gene regulation (Bellott et al. 2014). We analyzed the expression of bull MSY genes, as well as their homologs on the X Chromosome or autosomes, across eight somatic tissues and testis using previously published data sets (Merkin et al. 2012). All multicopy bull MSY genes, both ancestral and acquired, displayed largely testis-specific expression patterns (Fig. 4D; Bellott et al. 2014). Because of these patterns of gene expression, the bull and mouse MSYs as a whole can be viewed as more skewed toward testis-specific expression (Fig. 4C). Using a germ-cell-depleted mouse model, we were able to refine the expression pattern of the mouse MSY multicopy gene families further, finding that most are expressed exclusively in germ cells (Soh et al. 2014). Although a similar model is not available for bull, we analyzed previously published RNA-seq data sets generated from sorted germ cells (pachytene spermatocytes and round spermatids; purity >90% via StaPut gradient) (Lesch et al. 2016) and were able to detect transcription of bull MSY genes and their X Chromosome and autosome counterparts in these samples (Supplemental Fig. S7), providing evidence that these genes are transcribed in both pre- and post-meiotic male germ cells. In contrast, most single-copy ancestral genes of the bull MSY are expressed at appreciable levels in all tissues examined, as are their X-linked homologs (Fig. 4D, E).

X–Y coamplification in bull

The final common feature shared between the bull and mouse MSYs—the coamplification of X and Y gene families—has the greatest biological implications. The three acquired, amplified gene families on the mouse MSY long arm have nonallelic homologs on the X Chromosome that have been amplified in parallel, and all of these sex-linked gene families are expressed exclusively in the male germline (Soh et al. 2014). We speculated that this X–Y coevolution may be a manifestation of sex-linked meiotic drive, which distorts the balanced ratio of male and female offspring.

X–Y coamplification is also found in cattle. One of the amplified gene families on the bull MSY long arm is HSFY, which is present in at least 79 copies. HSFY is also multicopy in human and rhesus, but with only two or three copies, respectively. The X homolog of this gene, HSFX, is highly amplified in bull (Fig. 5A). Although the Bos taurus reference X Chromosome assembly contains just one copy of HSFX (The Bovine Genome Sequencing and Analysis Consortium et al. 2009), our analysis revealed higher than average BAC coverage within this region. We used SHIMS to sequence a total of eight X Chromosome clones (Supplemental Table S4) and found that HSFX is in fact a multicopy gene in Bos

Figure 2. Intrachromosomal similarities in bull, human, and mouse MSYs. Circos plots were generated to visualize intrachromosomal DNA similarities at three different percent identity cutoffs. Bull, mouse, and human MSYs are represented, to scale, as circle diagrams; black tick mark at top of circle denotes the artificial junction between the Yp and Yq termini; color-coding corresponds to sequence class as in Figure 1. For each MSY sequence, stepwise 50-kb segments were compared to the remaining unmasked sequence within the MSY. Lines within plots connect each 50-kb query to its top hit (only hits >10 kb in length are plotted). Line colors indicate minimum percent nucleotide identity of hits as shown.

Sequencing in Bos taurus reveals X–Y competition

www.genome.org
There are three distinct HSFX variants in the bull, and one variant has undergone further amplification, with at least nine highly similar copies (Fig. 5B). HSFX is also multicopy in human and rhesus: two variants are highly diverged from each other, and in human, each variant is duplicated (Fig. 5B). In bull, HSFY and HSFX copies display high levels of intra-family nucleotide identity, implying ongoing intrachromosomal gene conversion, but the Y- and X-linked families are highly diverged from each other (their predicted proteins show only 34% amino acid identity), indicating that there is no ongoing X-Y recombination at these loci (Fig. 5B). The amplification of HSFY and HSFX therefore appears to have occurred independently on the bovine Y and X Chromosomes, and phylogenetic analysis shows that the same is true in primates (Fig. 5B). The HSFY and HSFX gene families are expressed predominantly or exclusively in the testis in both human and bull, and are expressed in male germ cells in bull (Fig. 4; Supplemental Fig. S7), mirroring the expression pattern of the sex-linked amplified gene families in mouse. The bull provides a second example of rampant amplification on the MSY accompanied by large-scale X-linked amplification of homologous gene families, demonstrating that these events are linked and may be widespread among mammalian lineages.

Discussion

The bull MSY is the seventh sex chromosome to be completely sequenced using the SHIMS approach, the only proven strategy for producing super-resolution assemblies of ampliconic regions. Of the finished sex chromosomes, the bull MSY is second only to the mouse MSY in total ampliconic sequence content, and SHIMS was vital for the elucidation of the complex repeat architecture of the bull Y long arm and an accurate representation of its gene content. Because the majority of ampliconic repeats composing the bull Y long arm differ by less than one base pair per 1000 (Fig. 2), a whole-genome shotgun approach, in which read lengths are typically <500 bp in length, would collapse the majority of amplicons into a single unit. Thus, the ampliconic repeats of the bull MSY, which are of considerable biological interest, were only accessible through SHIMS.

Until now, the mouse MSY stood alone, contrasting starkly with the three sequenced primate MSYs in terms of overall size and structure, leaving open the question of whether the comparatively large and mostly ampliconic mouse MSY was unique among mammals. The MSY of bull, which represents a third major branch of the mammalian tree, is decidedly more similar to the MSY of mouse, demonstrating that massive lineage-specific MSY amplification of testis-specific gene families is not a peculiarity of the mouse lineage, but is widely distributed among mammals. More limited sequencing surveys of additional mammalian MSYs (including pig, gorilla, cat, and horse) show evidence for large-scale amplification of testis-specific gene families as well (Skinner et al. 2015; Tomaszkiewicz et al. 2016; Brashear et al. 2018; Janczka et al. 2018). Horse, which has multiple Y- and X-linked copies of a parasite-derived, testis-specific transcript, may provide another example of X-Y coamplification (Janczka et al. 2018). To understand the extent, mechanism, and evolutionary trajectory of MSY amplification, complete SHIMS-based assemblies of ampliconic regions from these and other mammals will be required.

The MSY amplification of testis-specific gene families in both bull and mouse occurred in parallel with amplifications of related testis-specific gene families on the X Chromosome (Fig. 5; Soh et al. 2014). It has been hypothesized that X-versus-Y
meiotic drive may induce an evolutionary arms race between the sex chromosomes in mouse (Ellis et al. 2011; Cocquet et al. 2012; Soh et al. 2014). In this scenario, the proportion of female and male offspring is influenced by a meiotic “driver” and “suppressor” located on the X and Y Chromosomes, respectively (Presgraves 2008; Meiklejohn and Tao 2010). This proposal is based on prior observations of such phenomena in other animals: sex-chromosome-associated meiotic drive has been

<table>
<thead>
<tr>
<th>Bull MSY gene</th>
<th>Copy number</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANEL1</td>
<td>1</td>
</tr>
<tr>
<td>DDX3Y</td>
<td>1</td>
</tr>
<tr>
<td>EIF1AY</td>
<td>1</td>
</tr>
<tr>
<td>EIF233Y</td>
<td>1</td>
</tr>
<tr>
<td>HSFY</td>
<td>79</td>
</tr>
<tr>
<td>OFD1Y</td>
<td>1</td>
</tr>
<tr>
<td>RBMY</td>
<td>1</td>
</tr>
<tr>
<td>SRY</td>
<td>1</td>
</tr>
<tr>
<td>TSPY1*</td>
<td>170</td>
</tr>
<tr>
<td>TSPY2</td>
<td>10</td>
</tr>
<tr>
<td>USP9Y</td>
<td>1</td>
</tr>
<tr>
<td>UTY</td>
<td>1</td>
</tr>
<tr>
<td>ZFY</td>
<td>1</td>
</tr>
<tr>
<td>ZRS2Y</td>
<td>1</td>
</tr>
<tr>
<td>PRAHYEY1*</td>
<td>20</td>
</tr>
<tr>
<td>PRAHYEY2</td>
<td>5</td>
</tr>
<tr>
<td>PRAHYEY3</td>
<td>1</td>
</tr>
<tr>
<td>PRSSLY</td>
<td>1</td>
</tr>
<tr>
<td>ZNF288BY</td>
<td>69</td>
</tr>
</tbody>
</table>

**Figure 4.** Bull MSY gene content and expression analysis. (A) Tabulation of ancestral and acquired protein-coding genes in bull MSY. Only intact genes, not pseudogenes, are counted in these totals. We confirmed transcription of all single-copy genes and representative members of multigene gene families. Copy numbers for TSPY1 and PRAHYE1 are estimates because these gene families are located within tandem arrays, for which we have only partial sequence. Copy numbers were calculated by determining BAC coverage within each array, which gives an estimate of its total size. (B) Plot of protein-coding gene density (number of protein-coding genes per Mb) in the five sequenced MSYS: bull, mouse, human, chimpanzee, and rhesus. (C) Plot of percentage of MSY protein-coding genes with testis-specific expression in the five sequenced MSYS. (D,E) Expression analysis includes RNA-seq data sets previously generated from nine adult male **Bos taurus** (Holstein) tissues. Expression levels for MSY genes (D) and their X or autosome homologs (E) were estimated in transcript per million (TPM) units. TPM values are plotted on a log10 scale. Three biological replicates were analyzed for each tissue; means with standard errors are plotted.
documented at least 24 times in insects and at least three times in rodents (Jaenike 2001).

In addition to HSFXY, there is another example of X–Y coamplification in human (Table 1). There are multiple copies of VCY (two copies) and VCX (four copies) in both human and chimpanzee, but no other primates. Both gene families are expressed exclusively in the male germline: VCX encodes a highly acidic protein, whereas VCY encodes a highly basic protein, suggestive of antagonistic functions (Lahn and Page 2000). The extent of X–Y coamplification varies widely in these three examples. Roughly 96% of the mouse MSY euchromatic sequence arose as a consequence of this phenomenon, compared to <2% of the human MSY (Table 1). In bull, mouse, and human, the X Chromosome sequence has not been affected to as large a degree as the Y Chromosome (Table 1), perhaps because the requirement for recombination between X homologs during female meiosis serves to constrain runaway amplification. One study found intense positive selection and selective sweeps associated with numerous X Chromosome ampliconic regions in primates, leading the investigators to speculate that these amplicons are involved in meiotic drive (Nam et al. 2015). Y-linked antagonists of these drivers are not necessarily homologous loci, so this phenomenon may have shaped the evolution of both X and Y Chromosomes more deeply and extensively than is immediately apparent.

In mouse, this “arms-race” hypothesis is supported by knockdown studies in which disruption of the X-linked homologs of one coamplified gene family (Sylko) skew sex ratios toward males, and disruption of the Y-linked homologs (Sly) skew sex ratios toward females (Cocquet et al. 2009, 2012). Moreover, targeted deletion and duplication of the Sylko gene family skew sex ratios toward males and females, respectively (Kruger et al. 2019). Progress toward a mechanistic understanding of this X–Y competition has come from a recent study demonstrating that the proteins encoded by Sylko and Sly compete to interact with the product of another amplified gene family (Styx) at the promoters of thousands of genes (Moretti et al. 2020).

### Table 1. Extent of coamplification of testis-specific gene families on X and Y Chromosomes in bull, mouse, and human

<table>
<thead>
<tr>
<th></th>
<th>X or Y euchromatin occupied by X–Y coamplified sequence (%)</th>
<th>X–Y coamplified gene families</th>
<th>Total copy number of X–Y coamplified genes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
<td>X</td>
</tr>
<tr>
<td>Bull</td>
<td>2</td>
<td>82</td>
<td>HSFX</td>
</tr>
<tr>
<td>Mouse</td>
<td>8</td>
<td>96</td>
<td>Sstx, Slx, Srsx</td>
</tr>
<tr>
<td>Human</td>
<td>0.06</td>
<td>1.65</td>
<td>HSFX</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.33</td>
<td>VCX</td>
</tr>
</tbody>
</table>

![Figure 5. Coamplification of testis-specific gene families on X and Y Chromosomes.](image)
perturbations of sex ratios in cattle would be of great interest, both biologically and commercially.

The bull Y Chromosome sequence provides a second example of massive amplification driven by X–Y coevolution in mammals, substantiating a second major theme of mammalian Y Chromosome evolution, one that is distinct from the process of Y decay (Fig. 6). Both of these evolutionary processes (decay and amplification) stem from the same initial processes in sex chromosome evolution: large-scale inversions on the Y Chromosome suppressing crossing-over with the X Chromosome. The cascading effects of this X–Y divergence reach far beyond the sex chromosomes and have been influenced differently in the soma and germline. Massive loss of Y Chromosome genes forms the basis for the X-linked recessive model of inheritance, in which X-linked recessive alleles lack “sheltering” from wild-type homologs in males. A second consequence of Y decay that played out in the soma is X-Chromosome inactivation, which is the manifestation of a complex evolutionary process involving Y gene loss, consequent up-regulation of the X homolog in males and females, and silencing of one X-linked copy in females (Ohno 1967; Jegalian and Page 1998). In contrast, the germline fostered selfish evolutionary processes. Suppression of X–Y crossing-over essentially creates distinct X and Y linkage groups in the male germline, producing an attractive environment for factors involved in meiotic drive (Lyttle 1991; Jaenike 2001). The downstream consequences of sex-linked meiotic drive may include X–Y incompatibilities that manifest as hybrid male sterility, promoting speciation (Frank 1991; Hurst and Pomiankowski 1991), and the evolution of meiotic sex chromosome inactivation, which may have arisen to constrain X-versus-Y competition during spermatogenesis through transcriptional silencing of the sex chromosomes (Ellis et al. 2005; Meiklejohn and Tao 2010).

Methods

BAC selection and sequencing

The single-haplotype iterative mapping and sequencing (SHIMS) strategy (Kuroda-Kawaguchi et al. 2001) was used to assemble a path of sequenced clones selected from the CHORI-240 BAC library (https://bacpacresources.org/) and a custom BAC library (BTDAEX) constructed by Amplicon Express. Both libraries are derived from a single male donor (L1 Domino, Bos taurus Hereford breed). Fingerprint contigs and end sequences had been generated previously for the CHORI-240 library (Snelling et al. 2007), so we relied primarily upon this library. The BTDAEX library was used to fill gaps. BAC tiling paths were selected for sequencing as previously described (Hughes et al. 2012). The error rate in the finished sequence was estimated by counting mismatches in alignments between overlapping clones.

Assessing copy number variability

Illumina sequence data generated from 234 bull genomes from four different Bos taurus breeds (Daetwyler et al. 2014) were obtained from the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra; project accession number SRP039339). We also obtained male Bos taurus indicus whole-genome Illumina sequence data from NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) to use as an outgroup (project accession numbers PRJNA277147, PRJNA324270, and PRJNA324822). Bowtie was used to map reads (Langmead et al. 2009), allowing up to two mismatches and up to a 650-bp insert size for paired-end reads; data from each animal was analyzed separately. First, reads were mapped to our entire bull MSY assembly, and only mapped reads were analyzed further. Next, we used these reads to map to the female Bos taurus genome (bosTaur4) and removed mapped reads. We then mapped the remaining reads to five different MSY ampliconic regions (TSPY1 array, TSPY2-PRAME2 array, PRAME1 array, RBMY array, and long-arm amplicon) and to 1-Mb single-copy regions from both the MSY short and long arms. We also mapped these reads to HSFX and to a single-copy region on the X Chromosome as a control. Bowtie parameters for this step were adjusted to allow up to 500 alignments per read and to prioritize the best alignments. We calculated the size of each ampliconic region relative to the 1-Mb single-copy short-arm region based on comparative read depths. The statistical significance of differences between means was determined using Welch’s t-test.

Fluorescence in situ hybridization (FISH)

FISH on male cell lines derived from various cattle breeds was performed as previously described (Raudsepp and Chowdhary 2008) and to a single-copy region on the X Chromosome as a control. Bowtie parameters for this step were adjusted to allow up to 500 alignments per read and to prioritize the best alignments. We calculated the size of each ampliconic region relative to the 1-Mb single-copy short-arm region based on comparative read depths. The statistical significance of differences between means was determined using Welch’s t-test.

Radiation hybrid mapping

Twenty-five STS markers were tested on a 25,000-rad panel consisting of 132 hybrid clones. A cell line derived from the same donor animal (L1 Domino) as the BAC library was used to construct the panel. Data analyses were performed using RHMAPPER 1.22 (Slonim et al. 1997).

Figure 6. Possible consequences of X–Y recombination suppression: two major, parallel themes of Y Chromosome evolution. Summary of major events and processes in the evolution of the mammalian X and Y Chromosomes from a pair of autosomes. The initial step was the Y Chromosome’s acquisition of the testis-determining gene, which was followed by suppression of recombination between the X and Y Chromosomes in most of their length. Lack of X–Y recombination resulted in a cascade of effects apparent within the sex chromosomes and genome-wide. The soma (bottom) and germline (top) influenced independent aspects of sex chromosome evolution.
Dot plot and phylogenetic analyses

Triangular and square dot plot analyses were performed using custom Perl codes (Supplemental Code). Circos plots were generated as previously described (Krzewinski et al. 2009). For each MSY sequence, stepwise 50-kb segments were used as BLAST queries against the remainder of the unmasked MSY sequence, and non-gapped percent identities were calculated for top-hit alignments. For bull and human, tandem array-associated gaps (three in bull and two in human) are represented by model sequence. For phylogenetic analyses, alignments were generated using PRANK (Löytynoja 2014), and trees were generated using PhyML with default parameters (Guindon et al. 2009). HSFY/X sequences used in this analysis can be found in Supplemental File S3.

Gene expression analyses

To generate the histograms shown in Figure 4, D and E, we reanalyzed a total of 27 publicly available data sets (Merkin et al. 2012). These comprised three data sets each from the following Bos taurus tissues: brain, colon, heart, kidney, liver, lung, skeletal muscle, spleen, and testes (SRA accession numbers SRX196344–SRX196370). After our initial mapping, we noted that one of the skeletal muscle samples (accession number SRX196368) indicated strong expression of several ampliconic gene families—HSFY, TSPY, and ZNF280B—which are usually testis-specific, even though the overall expression pattern of SRX196368 clusters with that of other Bos taurus muscle samples (38). We suspected that this data set might be mislabeled or contaminated, so we reanalyzed 11 additional male Bos taurus skeletal muscle RNA-seq data sets (SRA accession numbers SRX317172–SRX317179, SRX317191–SRX317193). We found no transcription of the HSFY, TSPY, and ZNF280B gene families in any of the 11 other muscle data sets. We noted that the abundance of testis transcripts in SRX196368 was only ~6% of their abundance in testis samples, so we concluded that the skeletal muscle data set corresponding to SRA accession number SRX196368 was contaminated and removed from our gene expression analyses. For analysis in Supplemental Figure S7, RNA-seq was performed on samples derived from Bos taurus (Hereford breed) whole testis and isolated pachytene spermatocytes and round spermatids. Purified germ cell fractions were obtained as previously described (Bellvé 1993). Total RNA was isolated from samples using the RNeasy kit (Qigen), and sequencing libraries were made using the TruSeq RNA kit (Illumina). Using the Illumina HiSeq 2500 platform, 100-bp paired-end reads were obtained for testis and 40-bp paired-end reads were obtained for purified germ cells. For all data sets, RNA-seq reads were mapped to the Bos taurus transcriptome (Bos_taurus.ARS-UCD1.2; MSY genes manually added) using ballistic with sequence-based bias correction (Bray et al. 2016). For multicopy X and Y gene families, we summed the number of reads that mapped to any single member of the gene family. Three biological replicates were analyzed for each tissue, with expression levels estimated in transcripts per million (TPM). Plots with means and standard errors were generated using Prism 8.

Sequence annotation

Interspersed repeats were electronically identified using RepeatMasker (http://www.repeatmasker.org). Protein-coding genes were identified as previously described (Skalts et al. 2003). Active genes were distinguished from pseudogenes by (1) evidence for transcriptional activity by RT–PCR or RNA-seq, (2) intact splice sites (if multivicous), and (3) full-length open reading frames (compared to other members of Y-linked multicopy gene family, Y-homologs in other species, or X/autosomal homologs in Bos taurus). Loci with confirmed transcription but without significant ORFs were considered noncoding transcripts.

Data access

All BAC sequences generated in this study have been submitted to the NCBI Nucleotide Database (https://www.ncbi.nlm.nih.gov/nucleotide/), and accession numbers are listed in Supplemental Tables S1 and S4. RNA-seq data from Bos taurus adult testis generated in this study have been submitted to the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession number SRX358238. The bull Y assembly has been submitted to the NCBI Nucleotide Database under accession number CM001061.2. Custom Perl scripts used to generate dot plots are available as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank L. Teitz for data analysis and A. Larrauneta for assistance with RH mapping. We thank J.L. Mueller, D.G. de Rooij, T. Endo, K.A. Romer, and Y.Q.S. Soh for comments on the manuscript. This work was supported by the National Institutes of Health (RO1HG007852) and Howard Hughes Medical Institute.


References

Hughes et al.


Received August 6, 2020; accepted in revised form October 28, 2020.
Sequence analysis in *Bos taurus* reveals pervasiveness of X–Y arms races in mammalian lineages

Jennifer F. Hughes, Helen Skaletsky, Tatyana Pyntikova, et al.

*Genome Res.* 2020 30: 1716-1726 originally published online November 18, 2020

Access the most recent version at doi:10.1101/gr.269902.120

---

**Supplemental Material**

http://genome.cshlp.org/content/suppl/2020/11/16/gr.269902.120.DC1

**References**

This article cites 68 articles, 11 of which can be accessed free at:

http://genome.cshlp.org/content/30/12/1716.full.html#ref-list-1

**Creative Commons License**

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

**Email Alerting Service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here.](http://genome.cshlp.org/subscriptions)

---

To subscribe to *Genome Research* go to:

http://genome.cshlp.org/subscriptions

© 2020 Hughes et al.; Published by Cold Spring Harbor Laboratory Press