Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes

Y.Q. Shirleen Soh,[1,2,7] Jessica Alföldi,[1,2,7,8] Tatyana Pyntikova,[1] Laura G. Brown,[1,3] Tina Graves,[4] Patrick J. Minx,[4] Robert S. Fulton,[4] Colin Kremitzki,[4] Natalia Koutseva,[1] Jacob L. Mueller,[1,9] Steve Rozen,[1,10] Jennifer F. Hughes,[1] Elaine Owens,[5] James E. Womack,[5] William J. Murphy,[5] Qing Cao,[6] Pieter de Jong,[6] Wesley C. Warren,[4] Richard K. Wilson,[4] Helen Skaletsky,[1,3] and David C. Page[1,2,3,*]

[1]Whitehead Institute, 9 Cambridge Center, Cambridge, MA 02142, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[3]Howard Hughes Medical Institute, Whitehead Institute, Cambridge, MA 02142, USA
[4]The Genome Institute, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
[5]College of Veterinary Medicine and Biomedical Sciences, 4458 Texas A&M University, College Station, TX 77843, USA
[6]BACPAC Resources, Children's Hospital Oakland, 747 52nd Street, Oakland, CA 94609, USA
[7]Co-first author
[8]Present address: Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[9]Present address: Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109, USA
[10]Present address: Duke-NUS Graduate Medical School, Singapore 169857, Singapore
*Correspondence: dcpage@wi.mit.edu
http://dx.doi.org/10.1016/j.cell.2014.09.052

SUMMARY

We sequenced the MSY (*male-specific region of the Y* chromosome) of the C57BL/6J strain of the laboratory mouse *Mus musculus*. In contrast to theories that Y chromosomes are heterochromatic and gene poor, the mouse MSY is 99.9% euchromatic and contains about 700 protein-coding genes. Only 2% of the MSY derives from the ancestral autosomes that gave rise to the mammalian sex chromosomes. Instead, all but 45 of the MSY's genes belong to three acquired, massively amplified gene families that have no homologs on primate MSYs but do have acquired, amplified homologs on the mouse X chromosome. The complete mouse MSY sequence brings to light dramatic forces in sex chromosome evolution: lineage-specific convergent acquisition and amplification of X-Y gene families, possibly fueled by antagonism between acquired X-Y homologs. The mouse MSY sequence presents opportunities for experimental studies of a sex-specific chromosome in its entirety, in a genetically tractable model organism.

INTRODUCTION

The mammalian sex chromosomes arose from an ordinary pair of autosomes (Lahn and Page, 1999; Ohno, 1967). Following the emergence of the sex-determining locus on the Y chromosome, a series of inversions on the Y chromosome suppressed X-Y crossing over. Suppression of X-Y crossing over liberated the X and Y chromosomes to radically differentiate. They remain identical only in the pseudoautosomal region, where X-Y crossing over still occurs.

Reconstruction of mammalian sex chromosome evolution has identified two major developments and highlighted their implications for human health and disease. First, the Y chromosome lost most of its ancestral genes, whereas the X chromosome retained them (Hughes et al., 2005, 2010, 2012; Ross et al., 2005; Skaletsky et al., 2003). Surviving ancestral Y genes are enriched for widely expressed, dosage-sensitive regulators of gene expression that may play crucial roles in Turner syndrome and sexual dimorphism in disease (Bellott et al., 2014). Second, both X and Y chromosomes have acquired and amplified genes independently of each other. These gene families are often expressed specifically in the testis, suggestive of functions in male gametogenesis (Hughes et al., 2010; Mueller et al., 2008, 2013; Murphy et al., 2006; Paria et al., 2011; Ross et al., 2005; Skaletsky et al., 2003; Warburton et al., 2004). The ampliconic structure of these amplified gene families makes them prone to recurrent rearrangements that can cause spermatogenic failure, sex reversal and Turner syndrome (Kuroda-Kawaguchi et al., 2001; Lange et al., 2009; Reijo et al., 1995; Repping et al., 2002, 2003).

Study of mammalian sex chromosomes and their relevance to human biology is hindered by the lack of high-quality reference sequences of X and Y chromosomes across a broad range of mammals. Our understanding of mammalian sex chromosomes and their evolution is presently supported by a limited set of high-quality single-haplotype sequences: complete X chromosome sequences from human and mouse, complete MSY sequences from human, rhesus, and chimpanzee, and the ancestrally derived sequences of the MSY in marmoset, mouse, rat, bull, and opossum (Bellott et al., 2014; Church et al., 2009; Hughes et al., 2005, 2010, 2012; Mueller et al., 2013; Ross et al., 2005;

Skaletsky et al., 2003). Other analyses rely on cDNA sequences, physical maps, and partial genomic sequences (Li et al., 2013; Paria et al., 2011). None of the fully sequenced Y chromosomes is from a genetically tractable model organism.

The mouse could potentially serve as such a model, offering rich opportunities to study mammalian Y chromosome biology. We thus embarked on this sequencing effort, whose results we now report, in 2002. Prior to our efforts, only limited sequences from the presumptive short arm were available to facilitate genetic studies of the mouse Y chromosome. The mouse MSY, which does not undergo crossing over with a homolog, is impervious to conventional genetic mapping methods, requiring instead approaches such as deletion mapping. When this study began, most identified genes, including the sex-determining gene *Sry*, mapped to the short arm (Burgoyne, 1998; Mazeyrat et al., 1998). The long arm was comparatively impenetrable due to its highly repetitive nature (Eicher et al., 1989; Nishioka and Lamothe, 1986; Phillips et al., 1982). Its repetitive sequences were known to produce at least one testis-specific transcript and were thought to play roles in spermatogenesis and fertility (Bishop and Hatat, 1987; Burgoyne et al., 1992; Conway et al., 1994; Moriwaki et al., 1988; Prado et al., 1992; Styrna et al., 1991; Touré et al., 2004a).

We recognized that mapping and sequencing the mouse MSY, given the repetitive long-arm sequences, would require a methodology with an extraordinary level of accuracy and precision. Our laboratory previously developed and used single-haplotype iterative mapping and sequencing (SHIMS) to assemble ampliconic sequences, long stretches of duplicated sequences that share high nucleotide identity, which are a common feature of primate sex chromosomes (Hughes et al., 2010, 2012; Mueller et al., 2013; Skaletsky et al., 2003). Based on our previous success, we employed the same approach with the mouse, obtaining high-quality, almost-complete sequence of the short arm, centromere, and long arm. These sequences, made immediately available in GenBank, have already enabled design and interpretation of numerous studies (Cocquet et al., 2009, 2010, 2012; Ellis et al., 2005, 2007, 2011; Ferguson et al., 2009; Pertile et al., 2009; Reynard et al., 2009; Touré et al., 2005; Wang et al., 2013a, 2013b).

Here, we report the sequence of the mouse MSY and its implications for mammalian sex chromosome evolution and biology. Like previously sequenced primate MSYs, the mouse MSY has lost most of its ancestral genes and acquired and amplified other genes during the past 200–300 million years. These processes of gene decay and acquisition are especially prominent in the mouse compared to the primate MSYs. The mouse MSY retains only nine of 639 ancestral genes, far fewer than in primates. Remarkably, the overwhelming majority of the mouse MSY consists instead of newly acquired, massively amplified, rodent-specific sequence, which contains three major Y gene families. These Y gene families have X homologs that are products of convergent acquisition and amplification. The work reported here sheds light on the paradoxical combination of divergence and convergence between the X and Y chromosomes in mouse sex chromosome evolution and offers a genetically tractable model for experimentation in Y chromosome biology.

RESULTS

Sequencing, Mapping, and Assembly of the Mouse Y Chromosome

The mouse MSY presented a technical challenge insurmountable by typical sequencing strategies. Prior evidence suggested that the mouse MSY contains highly repetitive sequences. We confirmed this by analyzing fingerprint contigs from C57BL/6J XY and XX bacterial artificial chromosome (BAC) libraries (RPCI-24 and RPCI-23, British Columbia Cancer Agency and Michael Smith Genome Science Center): male-specific fingerprint contigs displayed clone depths as great as 162-fold, whereas expected depth for single-copy Y sequence was 5-fold. Despite this unprecedented clone depth, we surmised from fingerprint analysis that individual repeat units were sufficiently large and complex to be amenable to BAC-based sequencing (Extended Experimental Procedures).

We sequenced the C57BL/6J mouse MSY using SHIMS (Extended Experimental Procedures). We started by identifying 170 mouse MSY BACs containing unique base-pair substitutions termed Sequence Family Variants (SFVs) (Saxena et al., 2000) (Data S1 available online). These BACs were used to seed 121 contigs, which were expanded by iteratively selecting and sequencing BAC clones with SFV matches to existing contig ends. To ensure accurate SFV identification and BAC assembly, we aimed to sequence BACs with large overlaps of about 50–60 kb. We obtained a tiling path of 88.8 Mb in 19 contigs, of which 60.0 Mb was sequenced in at least two independent BACs (Table S1). This redundancy enabled a high degree of accuracy, an error rate of about one in 143,000 nucleotides (Extended Experimental Procedures), which allowed us to resolve large, almost perfect repeats. We were able to resolve and assemble the massively ampliconic mouse MSY sequences only by SHIMS; whole-genome shotgun methods would have produced a collapsed assembly, not representative of actual sequences. SHIMS, as demonstrated here, will enable sequencing and assembly of complex ampliconic genomic structures, variants of which are often associated with human disease (Bailey and Eichler, 2006; Freeman et al., 2006).

To confirm and validate the sequence map, and order and orient the 19 sequence contigs, we constructed a radiation hybrid (RH) map spanning the entire length of the chromosome. We generated a panel of 93 high-resolution RH clones, which we genotyped for 215 markers designed to distinguish sequence differences between mouse MSY amplicons (Data S1). RH mapping independently confirmed our sequence assembly in each of the 19 sequence contigs, and enabled us to order and orient the contigs (Table S2).

We created a model Y chromosome assembly of 89.6 Mb, which comprises the 19 ordered and oriented contigs connected by estimated gaps (Data S1 and S2; Extended Experimental Procedures). We used this model assembly for all subsequent analyses. Based on the total sequence we obtained (88.8 Mb) and estimated gap sizes, we estimate our sequence to be 99.1% complete. This total sequence is within 6.2% of previous size estimates of the mouse Y chromosome as measured by flow cytometry (94.7 Mb) (Bergstrom et al., 1998).

The Mouse Y Chromosome Is Euchromatic and Massively Ampliconic

The mouse MSY dwarfs the human, chimp, and rhesus MSYs in both the absolute and relative amounts of euchromatic and ampliconic sequence: 89.5 Mb, or 99.9%, of the mouse MSY is euchromatic, and 87.7 Mb, or 98.0%, of this euchromatin is ampliconic (Figures 1A and 1B). The MSY's ampliconic sequences are distinct from what are typically referred to as repetitive sequences, which are often thought of as gene poor, high in interspersed repeats, or even heterochromatic. In contrast, the mouse MSY ampliconic sequence (and consequently the entire mouse MSY) is gene dense and is similar in interspersed-repeat content to the sequenced primate MSYs or to mouse autosomes (Figures 1C and 1D). Thus, the mouse MSY stands in contrast to theories that Y chromosomes should be degenerate, small, gene poor, and heterochromatic (Bachtrog, 2013; Charlesworth and Charlesworth, 2000; Graves, 2006).

A Complete Sequence of a Mammalian Centromere

We obtained the complete sequence of the mouse Y centromere (Figure S1). Consisting of 90 kb of satellite repeats, the centromere is the only heterochromatic sequence (defined as satellite sequence) that we identified in the entire mouse MSY. This centromeric sequence has been shown to associate with kinetochore-specific histone H3 CENPA (Pertile et al., 2009). It is located between 3.5 Mb of short-arm and 86.0 Mb of long-arm sequence, confirming that the mouse Y is the only acrocentric chromosome among all the other telocentric mouse chromosomes (Ford, 1966; McLaren et al., 1988; Roberts et al., 1988).

The Mouse MSY Retains Little Ancestral Sequence, but Newly Acquired Sequence Is Massively Amplified

The mouse MSY euchromatin contains two sequence classes with distinct origins: ancestral and acquired. Ancestral sequence, which originates from the autosomal ancestors of the mammalian sex chromosomes, occupies only 2.2%, or 2 Mb, of the total euchromatin and is located entirely within the short arm (Figures 1 and 2). Of this 2 Mb of ancestral sequence, 1.6 Mb is nonampliconic and contains seven single-copy genes and one duplicated gene (Bellott et al., 2014); the remaining 0.4 Mb is ampliconic and contains one ampliconic gene family (Table 1; Data S1). Relative to the primate MSYs, the mouse MSY retains fewer distinct ancestral genes, within a much smaller region. (Human, rhesus, and chimp retain 17, 18, and 13 ancestral genes, respectively.) Thus, it appears to have experienced greater degeneration than the primate MSYs and correspondingly has diverged more from its homolog, the mouse X chromosome.

The remaining 97.5%, or 87.4 Mb, of euchromatin consists of acquired sequence not originally on the ancestral autosomes, and not found on other sequenced mammalian MSYs. Almost all of this acquired sequence is ampliconic. One megabase of acquired ampliconic sequence is located on the short arm, interspersed among ancestral sequence (Figure 2); it includes one amplified gene pair, *H2al2y*, and one amplified testis-expressed transcript (Table S3; Extended Experimental Procedures). Taken in its entirety, the acquired sequence contains only two genes that exist in single copy, *Prssly* (*Protease, serine-like, Chr Y*) and *Teyorf1* (*Testis-expressed Y open reading frame 1*); these

genes map to the distal tip of the short arm, adjacent to telomeric sequence (Extended Experimental Procedures).

The remaining 86.4 Mb of acquired sequence is located on the long arm and is massively amplified (Figure 3A). The long-arm ampliconic sequence consists of a half-megabase unit amplified about 200 times. Each half-megabase unit comprises three core blocks defined by the boundaries of an internal duplication: internally unique sequences (depicted as red and blue) are bounded by the duplication (yellow) (Figure 3B). Amplification of the half-megabase unit results in >96% of the mouse MSY sharing at least 98% intrachromosomal identity (Figure S2). This amplified sequence makes up 3% of the haploid male mouse genome. Within the long-arm amplicons are two regions of extended and outstanding identity: one pair of 7 Mb direct repeats of 99.999% identity, and a second pair of 4.5 Mb tandem repeats of 99.995% identity (Figures 3A and S2). In comparison, the next largest pair of sequenced repeats from any organism is the P1 palindrome on the human MSY, which has a span of 1.45 Mb for each arm and arm-to-arm identity of 99.97% (Kuroda-Kawaguchi et al., 2001). The mouse MSY long-arm amplicons exist in two varieties: more regular and prevalent 515 kb tandem repeats (Figure 3C), and less regular, less prevalent 400 kb palindromes (Figure 3D). The two varieties occur in clusters along the length of the long arm (Figures 3A and 3E; Table S4).

Each long-arm ampliconic unit contains three protein-coding gene families, which are consequently massively amplified (Figures 3B–3D). The *Sly* and *Srsy* gene families were discovered through our sequencing efforts, and *Ssty* was previously described (Bishop and Hatat, 1987; Prado et al., 1992). Members of the *Ssty* family fall into two subfamilies, *Ssty1* and *Ssty2*. We find 126, 197, and 306 copies of *Sly*, *Srsy*, and *Ssty*, respectively, with intact open reading frames (ORFs) (Extended Experimental Procedures; Table 1; Data S1). Two of the three Y gene families have identifiable homology to autosomal genes whose products associate with chromosomes and have functions in meiotic chromosome synapsis and segregation. *Sly* is related to *Sycp3* (chromosome 10), a component of the synaptonemal complex (Moens and Spyropoulos, 1995), and *Ssty* is related to *Spin1* (chromosome 13), which has been shown to associate with the meiotic spindle in the mouse oocyte (Oh et al., 1997) (Figure S3). In addition to the three gene families, several non-coding transcripts also map to the long-arm amplicons (Table S3).

The 86.4 Mb of long-arm amplicons are interrupted in only 11 locations by a total of 0.76 Mb of sequence (Figure 3A; Data S1): nine copies of a 40 kb segment originating from chromosome 3, and two copies of a 200 kb segment that contains the acquired and amplified gene *Rbm31y* (Figure S3).

The Mouse MSY Long-Arm Amplicons Evolved at Least 3 Million Years Ago

The mouse MSY long-arm amplicons are present not only in C57BL/6J, but also in other *Mus musculus* strains and even other *Mus* species, demonstrating that the amplicons are at least 3 million years old. Nevertheless, the sequence structure and size are highly variable. We surveyed Y chromosomes from *M. musculus domesticus* (from AKR/J), *M. musculus castaneus* (CAST/EiJ), and *M. spretus* (SPRET/EiJ), which are estimated to have diverged from the C57BL/6 *Mus musculus musculus* Y
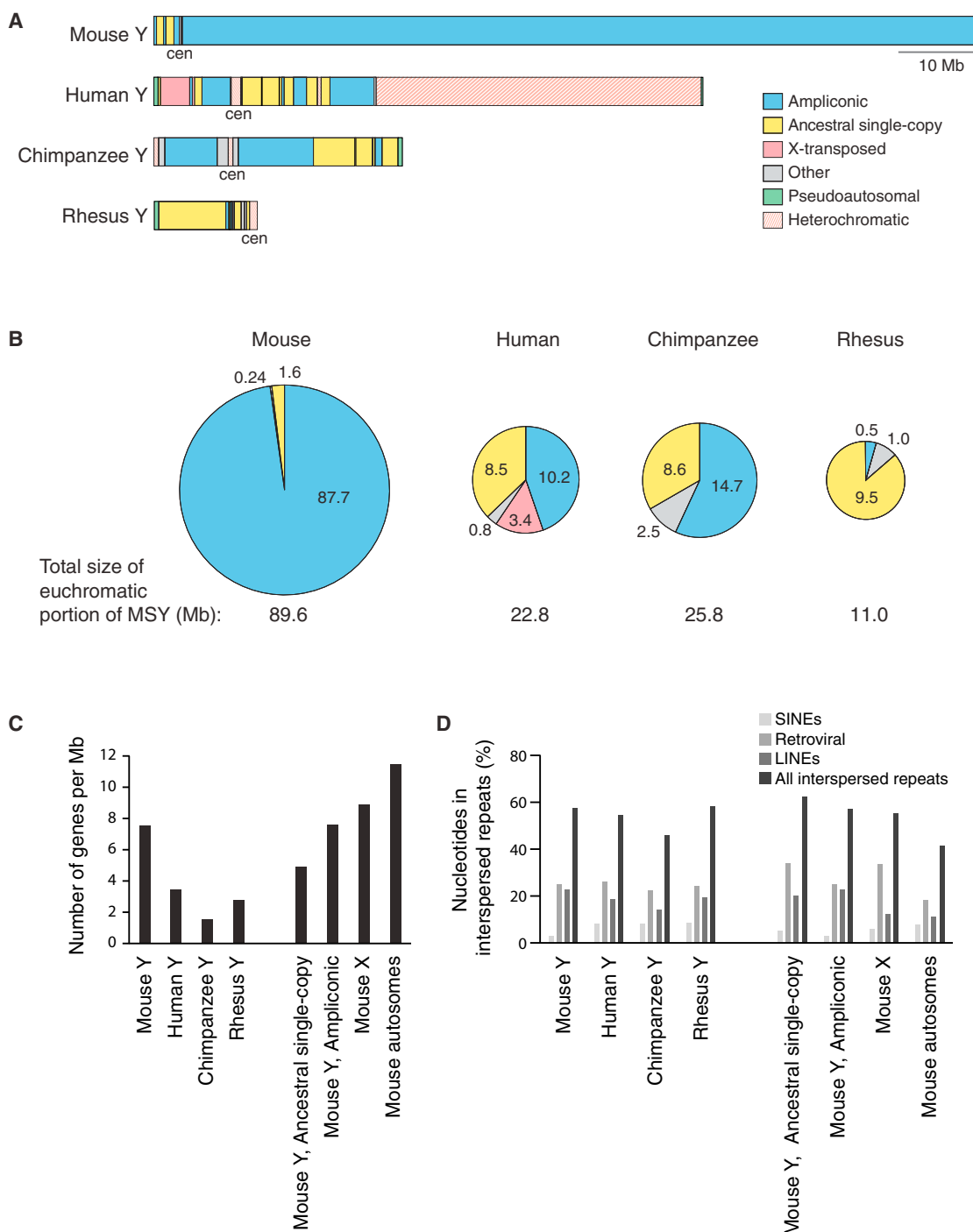
Figure 1. Comparison of Mouse, Human, Chimpanzee, and Rhesus Y Chromosomes

(A) Schematic representations of mouse, human, chimpanzee, and rhesus Y chromosomes, to scale. Ancestral single-copy corresponds to what was previously referred to as X-degenerate (Hughes et al., 2010, 2012; Skaletsky et al., 2003).

(B) Sizes (in Mb) of euchromatic sequence classes in MSYs.

(C and D) Gene and interspersed repeat content of mouse MSY euchromatic sequence compared to human, chimpanzee, and rhesus Y chromosomes and breakdown by euchromatic sequence classes compared to mouse X chromosome and autosomes. (C) Gene densities (numbers per Mb) of coding genes. (D) Percentages of nucleotides contained in SINEs, retroviral, LINEs, and total interspersed repeats. Human, chimpanzee, and rhesus data are previously described (Hughes et al., 2010, 2012; Skaletsky et al., 2003).

See also Data S1 for mouse MSY sequence assembly and Data S2 for complete annotated sequence of mouse Y chromosome.

A

Ampliconic
Ancestral single-copy
Other
Heterochromatic

→long arm

cen

500 kb

B

Prssly
Ube1y
Kdm5d
Eif2s3y
H2al2y
Rbmy
array

Protein-coding +
genes −

Teyorf1
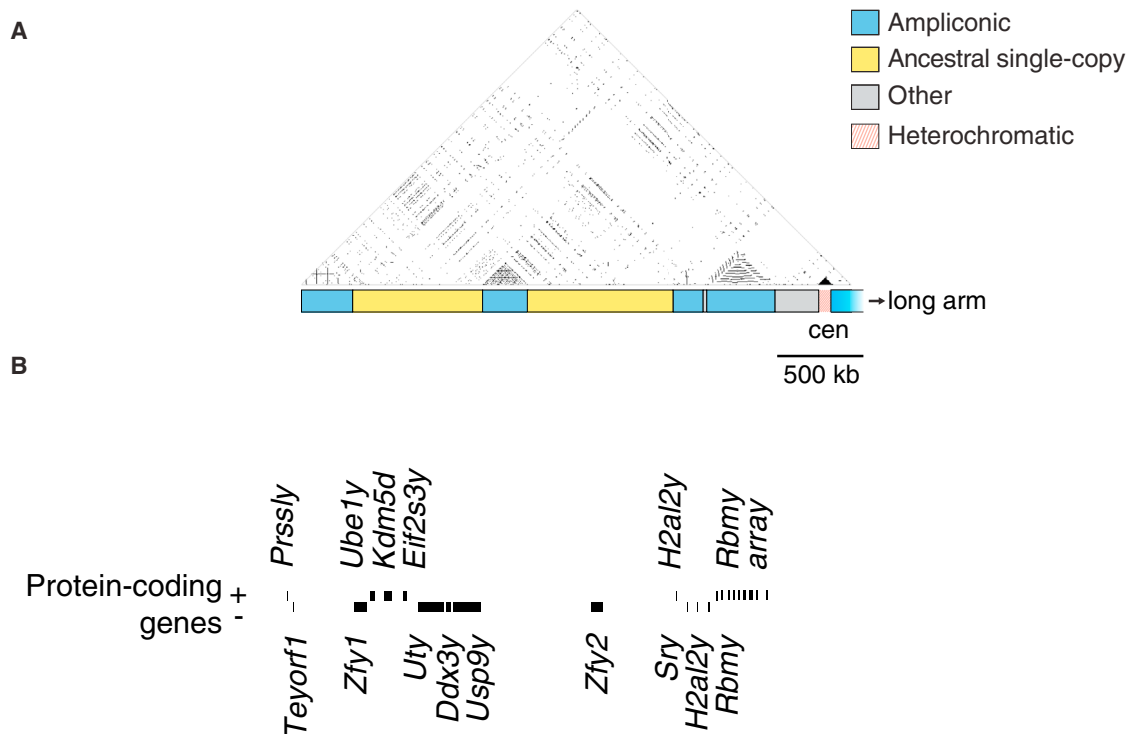Zfy1
Uty
Ddx3y
Usp9y
Zfy2
Sry
H2al2y
Rbmy

Figure 2. Gene Content and Structure of Mouse Y Short Arm
(A) Triangular dot plot of DNA sequence identities within the mouse Y short arm. Each dot represents 100% intrachromosomal identity within a 100 bp window. Direct repeats appear as horizontal lines, inverted repeats as vertical lines. The centromere (cen) is visible as a dense triangle representing highly identical heterochromatic sequence.
(B) Twelve families of protein-coding genes on the short arm.
See also Figure S1 for structure of the mouse Y centromere.

chromosome approximately 1, 1, and 3 million years ago, respectively (Silver, 1995). For each Y chromosome, we identified and sequenced three BACs containing sequence similar to the C57BL/6 long-arm amplicon (Table S5). Dot-plot analysis demonstrated that sequences from all three additional Y chromosomes align to the C57BL/6 long-arm amplicon, albeit with rearrangement (Figures 4A and S4A). In comparison, dot-plot analysis of a region of autosomal sequence shows no rearrangements between C57BL/6 and SPRET/EiJ (Figure S4B). From the sequenced BACs, we identified intact ORFs for *Sly* and *Ssty* in all three additional Y chromosomes. We identified intact ORFs for *Srsy* in AKR/J and CAST/EiJ Y chromosomes and found sequence alignment but no intact ORF within the three SPRET/EiJ Y chromosome BACs sequenced (Data S1). To determine whether these sequences are amplified in the three additional *Mus* Y chromosomes as in C57BL/6, we used fluorescence in situ hybridization to probe each Y chromosome with a BAC that contains long-arm amplic_onic sequence (Figure 4B). We observed that the long-arm amplicons are amplified to different degrees in the four Y chromosomes. We conclude that the ampliconic structure is at least 3 million years old, but rapidly evolving.

Convergently Acquired and Amplified Homologs on the Mouse X Chromosome
All acquired and amplified MSY genes have convergently acquired counterparts on the X chromosome (Table 1; Figure 5;

Data S1) (Mueller et al., 2008; Reynard et al., 2007), and, in all but one case, the X homolog is also amplified. All acquired and amplified X and Y genes are specific to the rodent or mouse lineage (Figure S3).

We considered the two ways by which X-Y gene families, in particular, the three most massively amplified X-Y gene families, could have been co-amplified: by recombination between the X and Y chromosomes or by recombination within each chromosome. Three pieces of evidence served to rule out X-Y recombination. First, the global physical distributions of the X-Y homologs make X-Y pairing and recombination unlikely: the three Y gene families are physically intermingled throughout the mouse Y long arm, whereas each X ampliconic gene family is found in distinct clusters on the X chromosome (Figure 5A). Second, local X-Y homology is limited: we find that for *Sly/Slx/Slxl1* and *Ssty/Sstx*, homology is limited to the genes themselves. For *Srsy/Srsx*, X-Y sequence similarity is limited to 30 kb of sequence surrounding the gene (Figure S5). Finally, each pair of X and Y homologs, and their surrounding sequences, display sequence divergence that is inconsistent with on-going interchromosomal recombination (Table S6; Figure 5B). We conclude that intrachromosomal recombination both within the X chromosome and within the Y chromosome is the likely mechanism of co-amplification of X and Y gene families.

This hypothesis is supported by abundant evidence of rearrangements within the mouse Y long-arm amplicons. Earlier in

Table 1. Genes and Gene Families on the Mouse Y Chromosome and Their X Homologs

Sequence Class	Y Gene	Copy Number	X-Homolog	Copy Number
Ancestral	*Zfy1, Zfy2*	2	*Zfx*	1
	Ube1y1	1	*Ube1x*	1
	Kdm5d	1	*Kdm5c*	1
	Eif2s3y	1	*Eif2s3x*	1
	Uty	1	*Utx*	1
	Ddx3y	1	*Ddx3x*	1
	Usp9y	1	*Usp9x*	1
	Sry	1	*Sox3*	1
	Rbmy	30[a]	*Rbmx*	1
Acquired	*H2al2y*	2	*H2al1*	14[b]
	Rbm31y	2	*Rbm31x*	1
	Sly	126	*Slx*	25[b]
			Slxl1	14[b]
	Ssty1	85	*Sstx*	11[b]
	Ssty2	221		
	Srsy	197	*Srsx*	14[b]
	Prssly	1[a]		
	Teyorf1	1[a]		

See Data S1 for sequences of individual members of amplified gene families. See Figure S3 for phylogenetic analyses of origins of acquired ampliconic genes. See Table S3 for accession numbers, and non-coding transcripts identified on the mouse Y chromosome.
[a]Estimate, see Extended Experimental Procedures.
[b]Estimates (Mueller et al., 2008).

this manuscript, we described 7 and 4.5 Mb segments of 99.999% identity and 99.995% identity, which most likely resulted from intrachromosomal duplications that either occurred recently (within the last few thousand years), or have been maintained by extremely efficient gene conversion. We found several long-arm amplicons that deviate from the canonical red-yellow-blue-yellow arrangement, probably resulting from recombination between ampliconic units (Figures 3E and S6). Finally, we observed a deletion of several megabases in the C57BL/6J strain when compared to the C57BL/6JTac strain, from which the RH panel was constructed (Extended Experimental Procedures). These two C57BL/6 strains have been reproductively isolated only since 1951. Several naturally occurring deletions of significant portions of the long arm have also been documented (Conway et al., 1994; Styrna et al., 1991; Touré et al., 2004a). Thus, the massively ampliconic sequences of the long arm likely act as extensive substrates for intrachromosomal recombination and facilitate rearrangements.

X-Y Acquired and Amplified Genes Are Expressed Specifically in the Male Germline

We next considered pressures that may have selected for co-acquisition and amplification of X-Y gene families. We assessed likely functions of X-Y gene families by measuring their expression across a panel of adult mouse tissues (ENCODE Project Consortium, 2012; Merkin et al., 2012). To determine if expression within the testis is in germ cells, we also examined wild-type (*Kit*+/*KitWv*) and germ-cell-deficient (*KitW*/*KitWv*) testes (Mueller et al., 2013). From this and previous analyses (Mueller et al., 2008; Reynard et al., 2007, 2009; Touré et al., 2004b), we conclude that *Sly*, *Ssty*, and *Srsy* and their X homologs, *Slx* and *Slxl1*, *Sstx*, and *Srsx*, are expressed predominantly in the male germline (Figure 6). By examining SFVs that distinguish individual members of each gene family, we found evidence of transcription of at least one-third of the members of the *Ssty* and *Sly* gene families (Extended Experimental Procedures). Male germline expression of these genes is consistent with observations that mice bearing deletions of the mouse Y long arm—and therefore possessing reduced numbers of *Sly*, *Ssty*, and *Srsy* genes—suffer sperm abnormalities and subfertility (Burgoyne et al., 1992; Conway et al., 1994; Moriwaki et al., 1988; Reynard et al., 2009; Styrna et al., 1991; Touré et al., 2004a). Other acquired Y genes and their X counterparts—*H2al2y/H2al2x*, *Rbm31y/Rbm31x*, *Prssly*, and *Teyorf1*—are also expressed predominantly in testicular germ cells, again supporting the idea that acquired genes on the sex chromosomes have functions in male gametogenesis (Bellott et al., 2010; Hughes et al., 2010, 2012; Lahn and Page, 1997; Skaletsky et al., 2003). As a control, we analyzed expression of the ancestral single-copy Y genes and their X homologs. Many of these ancestral genes are ubiquitously expressed, as previously noted, consistent with the idea that surviving ancestral genes are widely expressed, dosage-sensitive regulators of gene expression (Bellott et al., 2014; Lahn and Page, 1997).

DISCUSSION

We have assembled, using SHIMS, the sequence of the mouse MSY. Despite the shared evolutionary origin of placental mammalian Y chromosomes, the mouse MSY is spectacularly different from the human, rhesus, and chimpanzee MSYs. Only 2 Mb, or 2.2% of its sequence, has shared ancestry with the primate MSYs. Instead, the mouse MSY is dominated by a single family of acquired amplicons not found in the primate MSYs. These amplicons, each of which spans 500 kb and contains three gene families, comprise 86.4 Mb, or 96.5% of the chromosome.

How has the mouse MSY chromosome evolved to be so different from the primate MSYs? The same processes underlie the evolution of both mouse and primate MSYs; the difference lies in the extent to which they have played out. Mouse and primate MSYs alike have lost most genes from the ancestral autosomes, with the mouse retaining fewer genes than the primates. Both mouse and primate MSYs have acquired and amplified testis-specific genes. However, the relative and absolute scale of amplification in the mouse is unmatched among sequenced sex chromosomes.

Amplification of Acquired Sequence Is Consistent with Sex-Linked Meiotic Drive

We suggest that massive amplification of acquired sequence on the mouse MSY resulted from sex-linked meiotic drive. Sex-linked meiotic drive occurs when a driver arises on a sex chromosome, causing it to be transmitted to offspring more often than its counterpart. The resulting skew in sex ratio exerts strong countervailing selection for a suppressor, on the autosomes or
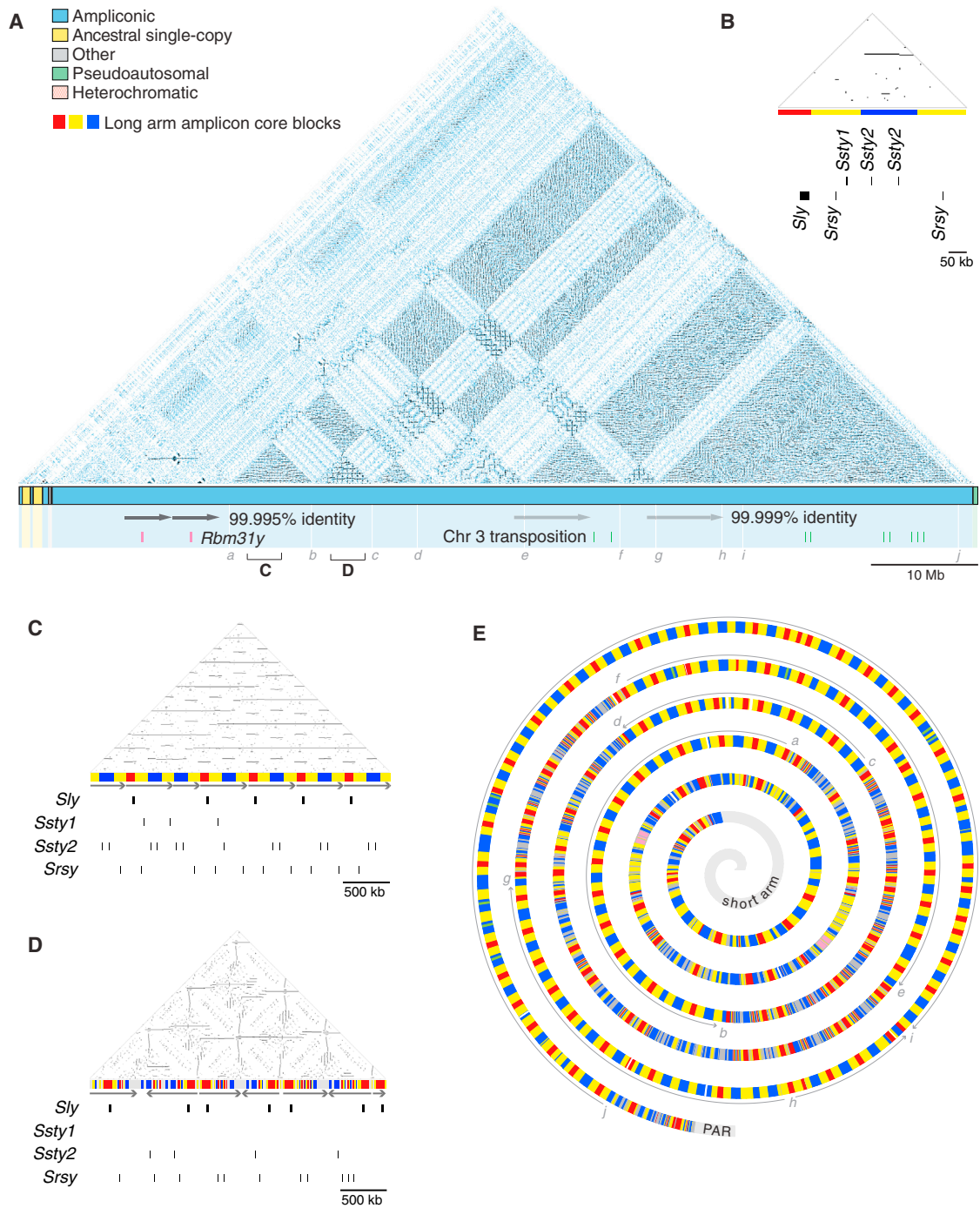
Figure 3. Gene Content and Structure of Mouse Y Long Arm

(A) Triangular dot plot of DNA sequence identities of the entire mouse Y chromosome. Each black or blue dot represents 100% intrachromosomal identity within a 500 or 200 bp window, respectively. Underneath the dot plot: gray arrows represent regions of extended and high identity; pink and green bars represent locations of *Rbm31y* and chromosome three transposition, respectively. Gray, italicized small letters label boundaries between clusters of different amplicon organization.

(B) Triangular dot plot of DNA sequence identity and gene content of a consensus tandem amplicon unit.

(C and D) Triangular dot plots of DNA sequence identity of two major amplicons type and organization: tandem (C) and palindromic (D).

For (B)–(D), each black dot represents 100% intrachromosomal identity within a 100 bp window. Underneath each dot plot is a representation of the substructure of each ampliconic unit, consisting of red, yellow, and blue core blocks. Directionality of the amplicons is indicated by gray arrows. Locations of protein coding genes within the amplicons are indicated by black bars.

(legend continued on next page)

the sex chromosome counterpart, to restore sex ratio balance. If both driver and suppressor are dosage sensitive, they would undergo iterated cycles of expansion, resulting in rapid co-amplification of both driver and suppressor (Jaenike, 2001; Partridge and Hurst, 1998). In *Drosophila melanogaster*, the X- and Y-linked multicopy genes *Stellate* and *Suppressor of Stellate* are hypothesized to be such a pair of meiotic driver and suppressor (Hurst, 1992, 1996; Palumbo et al., 1994). The mouse MSY's three acquired and massively amplified gene families and their X homologs are reminiscent of a meiotic driver and suppressor pair: in all three cases, both the X and Y genes are highly amplified, they are expressed specifically in testicular germ cells, and perturbation of gene family copy number results in sex ratio distortion. Mice that have fewer members of the Y gene families, due to partial deletions of the Y long arm, produce more female than male offspring (Conway et al., 1994; Moriwaki et al., 1988). Knock-down of *Sly* or *Slx*, one of the three X-Y gene pairs, also distorts sex ratio in favor of females or males, respectively (Cocquet et al., 2009, 2012).

Although we presently lack sufficient information to reconstruct the evolutionary history of the X-Y acquired and amplified gene families with certainty, we speculate that in mouse, one or more meiotic drivers were initially acquired by the X chromosome, and then countered by one or more suppressors acquired by the Y chromosome. We observe that the X chromosome has many more acquired genes than the Y chromosome (Mueller et al., 2013), whereas all Y-acquired and amplified genes have X homologs, consistent with their having been acquired in response to a subset of X-acquired genes. Subsequently, intrachromosomal recombination within acquired ampliconic regions would facilitate rapid evolution and expansion of both drivers and suppressors. We note that the amplification of X-acquired genes is restricted to discrete, localized clusters, whereas Y-acquired genes are intermingled and massively amplified, likely reflecting different constraints with respect to crossing over: amplification on the X chromosome is constrained by the need to maintain crossing over between X homologs in the female, whereas amplification on the MSY has no such limitations.

Strong selective pressure to acquire and amplify Y-linked suppressors of meiotic drive may account not only for the massively amplified Y-acquired genes, but also for the much-decayed Y ancestral genes, thereby shaping the character of the entire mouse MSY. Specifically, we speculate that ancestral gene decay may be a by-product of strong positive selection for meiotic drive suppressors on the mouse MSY. Because the MSY does not cross over with a homolog, it is inherited and selected as a unit. Strong selection for beneficial mutations in acquired, amplified MSY genes could propel deleterious mutations in ancestral MSY genes to fixation in a population (Charlesworth and Charlesworth, 2000; Rice, 1987). Indeed, our laboratory previously drew attention to such a correlation between enhanced gene acquisition/amplification and increased ancestral gene

decay within the primates: the chimpanzee MSY, which has acquired twice as many palindromes as the human MSY, has sustained more inactivating mutations in ancestral genes than human (Hughes et al., 2005, 2010). Thus, in both mouse and chimpanzee, single-copy, ancestral MSY genes may have been casualties of selective forces directed at the ampliconic, acquired genes.

X-Y interchromosomal conflict, and its consequent impact on gene acquisition and amplification on sex chromosomes, may be widespread in mammals. In both human and mouse—the only two species with high-quality reference sequences for both sex chromosomes—the X and Y have co-acquired and amplified genes. As in mouse, the human gene families *VCX* and *VCY* are also testis specific and may also be involved in meiotic drive (Lahn and Page, 2000). These X-Y gene families are lineage specific, consistent with rapid evolution due to meiotic drive: *VCX* is detected only in simian primates, and *Sly*, *Ssty*, and *Srsy* only in murid rodents. The scale of amplification on the mouse MSY is much greater than that of the sequenced primate MSYs. However, the mouse may not be exceptional: the cat and horse MSYs also appear to have highly amplified gene families (Murphy et al., 2006; Paria et al., 2011). To determine the true phylogenetic range of lineage-specific acquisition and amplification of X-Y genes, SHIMS assemblies must first be constructed for the X and Y chromosomes across more mammals.

Genetic conflict between the X-Y acquired ampliconic genes may contribute to hybrid sterility and consequent reproductive isolation. It was previously proposed (Frank, 1991; Hurst and Pomiankowski, 1991), and recently demonstrated in *Drosophila* (Phadnis and Orr, 2009; Tao et al., 2001), that segregation distortion can cause male hybrid sterility. In *Mus*, hybrid sterility loci map at or near ampliconic regions on the X chromosome, including *Slx* and *Slxl1* (Elliott et al., 2001, 2004; Good et al., 2008; Mueller et al., 2013), and they may result from incompatibilities between X and Y chromosomes (Campbell et al., 2012). Further study of the rapidly co-evolving, lineage-specific X-Y ampliconic gene families will be needed to test the proposed link between X-Y genetic conflict and hybrid sterility in mammals.

The Mouse Y Chromosome as an Experimentally Tractable System

The mouse MSY sequence provides unprecedented opportunities for experimentation. Its two major sequence classes—acquired and amplified, versus ancestral, single-copy—require distinct experimental approaches.

The three gene families that are intermingled and massively amplified on the mouse Y long arm cannot be deciphered by the previous major experimental approach, which utilizes massive, naturally occurring deletions of the long arm: all three intermingled gene families would be deleted, which would preclude parsing the roles of individual gene families. It would also be challenging to use a traditional genetic targeting approach

---

(E) Detailed long-arm amplicon substructure. Green and pink correspond to chromosome 3 transposition and *Rbm31y*, respectively. Gray corresponds to other sequence, including the short arm and PAR. Gray, italicized small letters label boundaries between different organizations and directions of amplicons, and correspond to labels in Figure 3A. Gray arrows indicate directionality of amplicons.

See also Figure S2 for details on 7 and 4 Mb regions of high identity, and overall intrachromosomal similarity. See Figure S6 for evidence of rearrangements within long-arm amplicons.

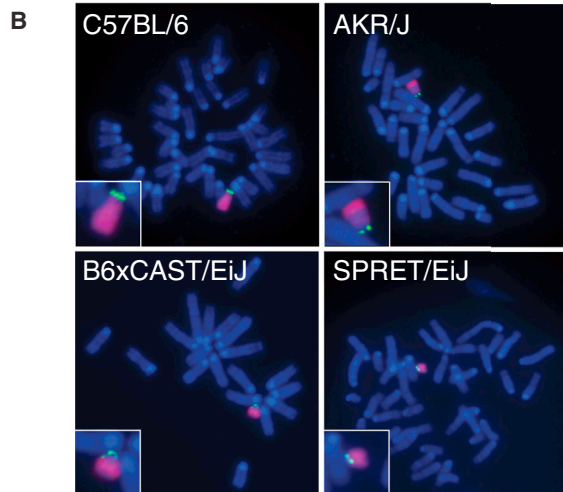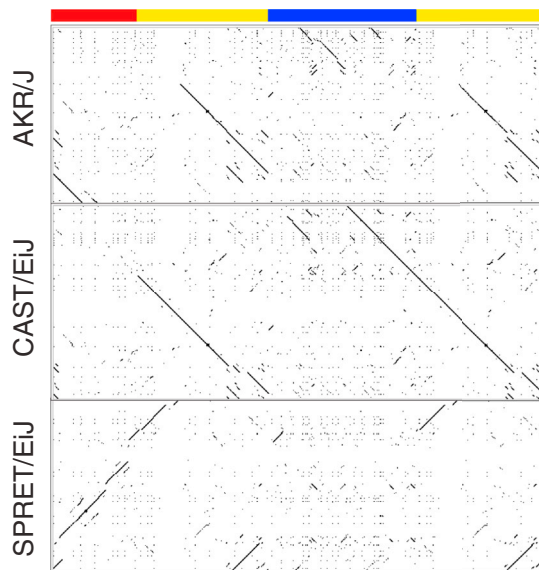**A** C57BL/6J long arm amplicon consensus unit

Figure 4. Comparison of Long-Arm Ampliconic Sequence in Related *Mus* Species

(A) Dot plots of DNA sequence identity between the C57BL/6J Y chromosome long-arm consensus amplicon and BACs from AKR/J (CH33-213P12), CAST/EiJ (CH26-73N1), and SPRET/EiJ (CH35-73N05). Each dot represents 100% identity within a 25 bp window.

(B) DNA FISH on cell spreads containing the Y chromosome of C57BL/6Tac (*Mus musculus musculus*), AKR/J (*Mus musculus domesticus*), CAST/EiJ (*Mus musculus castaneus*), and SPRET/EiJ (*Mus spretus*). Chromosomes were labeled with a single probe from C57BL/6J from the short-arm sequence (green) and a single probe deriving from each respective strain from the long-arm sequence (red). Short-arm C57BL/6J probe, RP24-084F20; long-arm C57BL/6J probe, RP24-088I20; long-arm AKR/J probe, CH33-204D11; long-arm CAST/EiJ probe, CH26-073N11; long-arm SPRET/EiJ probe, CH35-062A13. See also Figure S4 for additional dot plots of long-arm ampliconic sequences from related *Mus* species.

to ablate all members of any one of the gene families. At present, the most feasible approach is to target members of individual gene families by knockdown. Knowledge of the mouse MSY sequence and structure that emerged early during this project led investigators to adopt knockdown strategies to target *Sly* and its X homologs, *Slx* and *Slxl1* (Cocquet et al., 2009, 2010, 2012). These approaches have recapitulated sperm defects and sex ratio distortion observed with MSY long-arm deletions and enabled further investigations of gene function. The other massively amplified X-Y gene families remain to be similarly investigated, and the sequence presented here can guide knockdown strategies.

The prominent amplification on the mouse MSY long arm presents an opportunity for biochemical studies of chromatin. Unbiased profiling of chromosomal proteins, by purifying proteins associated with a specific DNA locus, has been hindered by the inability to purify sufficient amounts of a target region. One successful strategy took advantage of telomeric sequences, which are abundant relative to most DNA loci (Déjardin and Kingston, 2009). The mouse long-arm amplicons, comprising 3% of the mouse haploid genome, constitute a significantly enriched substrate for such unbiased chromatin protein-profiling strategies.

The mouse MSY sequence also provides a foundation for genetic manipulation of the ancestral, single-copy genes. Until recently, efforts at targeted mutagenesis of Y genes met with poor success. As such, studies have relied on Y chromosome translocations, deletions, and transgenesis, which are not optimal for dissecting individual gene functions (Burgoyne, 1998; Mazeyrat et al., 2001; Vernet et al., 2011, 2012; Yamauchi et al., 2014). The sequence we present, in combination with new genetic targeting methods such as TALENs and CRISPR, has already enabled genetic targeting of Y single-copy genes in mice (Wang et al., 2013a, 2013b). The research community can now explore the diverse biology of the male-specific chromosome in the premier mammalian genetic model, armed with a comprehensive, high-quality reference sequence.

EXPERIMENTAL PROCEDURES

BAC Selection and Sequencing/Single-Haplotype Iterative Mapping and Sequencing

We sequenced 746 BACs from the RPCI-24 C57BL6/J library and 232 BACS from the CHORI-36 C57BL6/J library (Table S1). BAC selection and sequencing occurred in two phases. In the first phase, we used eight sequence-tagged site (STSs) to identify 121 nonoverlapping BACs containing unique SFVs (Data S1). In the second phase, we used these 121 BACs to seed contigs, which we expanded by iteratively selecting and sequencing partially overlapping BAC clones with matching SFV patterns. See Extended Experimental Procedures for details.

We also identified and sequenced three clones each, representative of the long-arm amplicons, from CHORI-33 (AKR/J), CHORI-26 (CAST/EiJ), and CHORI-35 (SPRET/EiJ) (Table S5).

Radiation Hybrid Mapping

215 STS markers were tested on a 25,000 rad panel consisting of 93 hybrid clones (Table S2; Data S1). An RH map was constructed using RHMAPPER (Slonim et al., 1997).

Interspersed Repeats

Interspersed repeats were electronically identified with RepeatMasker (Smit et al., 1996).
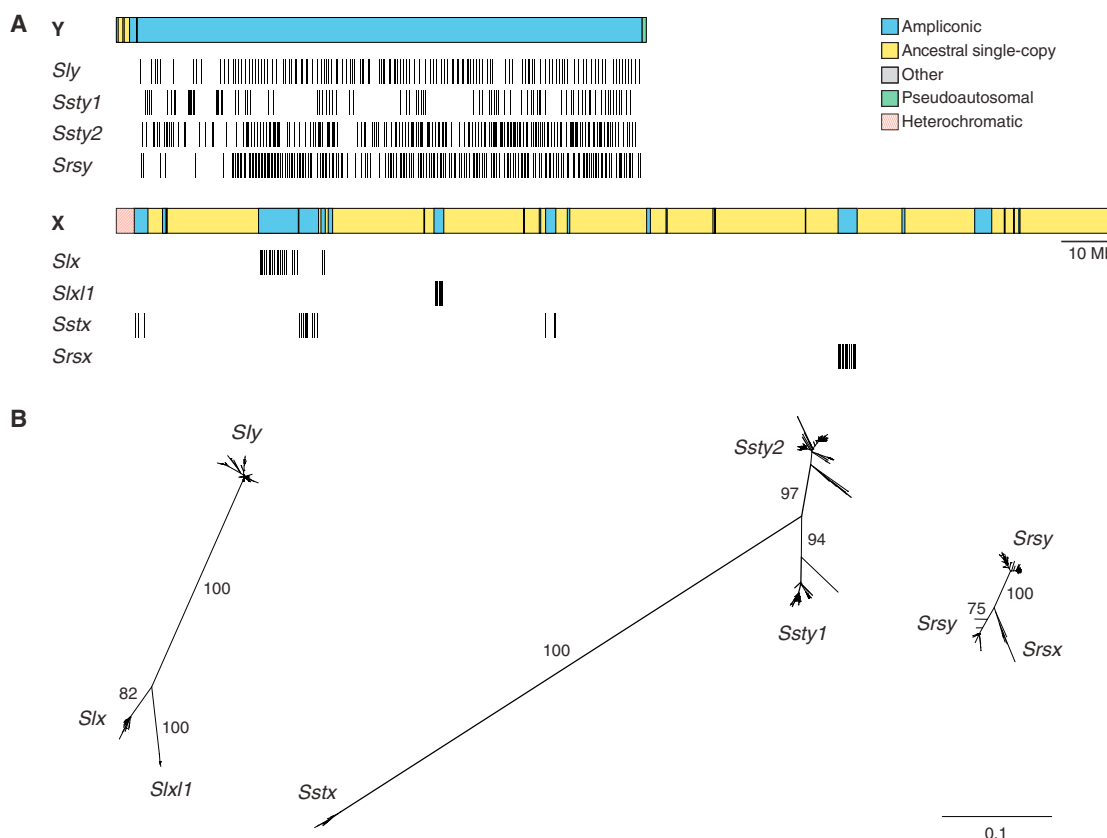
Figure 5. Comparison of X and Y Ampliconic Genes

(A) Location of ampliconic genes on the mouse Y long arm, and their homologs on the mouse X chromosome, to scale. (Note that the location of the *Sstx* cluster at 30 Mb is likely misplaced in this genome assembly: based on BAC-end sequence matches, it should be located together with the *Sstx* cluster at the proximal tip.)
(B) Phylogenetic analysis of mouse Y long-arm ampliconic genes and their X homologs. All trees are drawn on same scale. Unit length represents expected substitutions per site. Branch labels indicate bootstrap confidence values. See also Figure S5 for dot-plot analyses of X and Y ampliconic regions. See Table S6 for nucleotide identity between X and Y amplified gene families.

Dot Plots

Triangular dot plots (representing intrachromosomal sequence similarity) and square dot plots (representing interchromosomal sequence similarity) were generated by a custom Perl script available at http://pagelab.wi.mit.edu/material-request.html.

Identification of Genes and Transcription Units

We identified genes and transcripts as previously described (Skaletsky et al., 2003). See Extended Experimental Procedures for details.

Chromosomal Fluorescence In Situ Hybridization /Fluorescence In Situ Hybridization Analyses

Fluorescence in situ hybridization (FISH) assays were performed on mouse embryonic fibroblasts using probes for the C57BL6/J Mus *musculus musculus* Y chromosome. Mouse embryonic fibroblasts were derived from embryos of C57BL6/NTac, *Mus musculus domesticus* (AKR/J), *Mus musculus castaneus* (C57BL6/NTac × CAST/EiJ), and *Mus spretus* (SPRET/EiJ). Metaphase FISH analyses were performed as previously described (Saxena et al., 1996).

Gene Expression Analyses

We measured expression across mouse tissues using previously published mRNA sequencing data: *Kit^W/Kit^Wv* and *Kit^+/Kit^W* testis, SRA060831; ovary, SRX135150; all other tissue, SRP016501 (ENCODE Project Consortium, 2012; Merkin et al., 2012; Mueller et al., 2008). Normalized expression for each gene was calculated as in Bellott et al. (2014). For multicopy or ampliconic gene families, we counted the number of reads that aligned to any member of the gene family. See Extended Experimental Procedures for details.

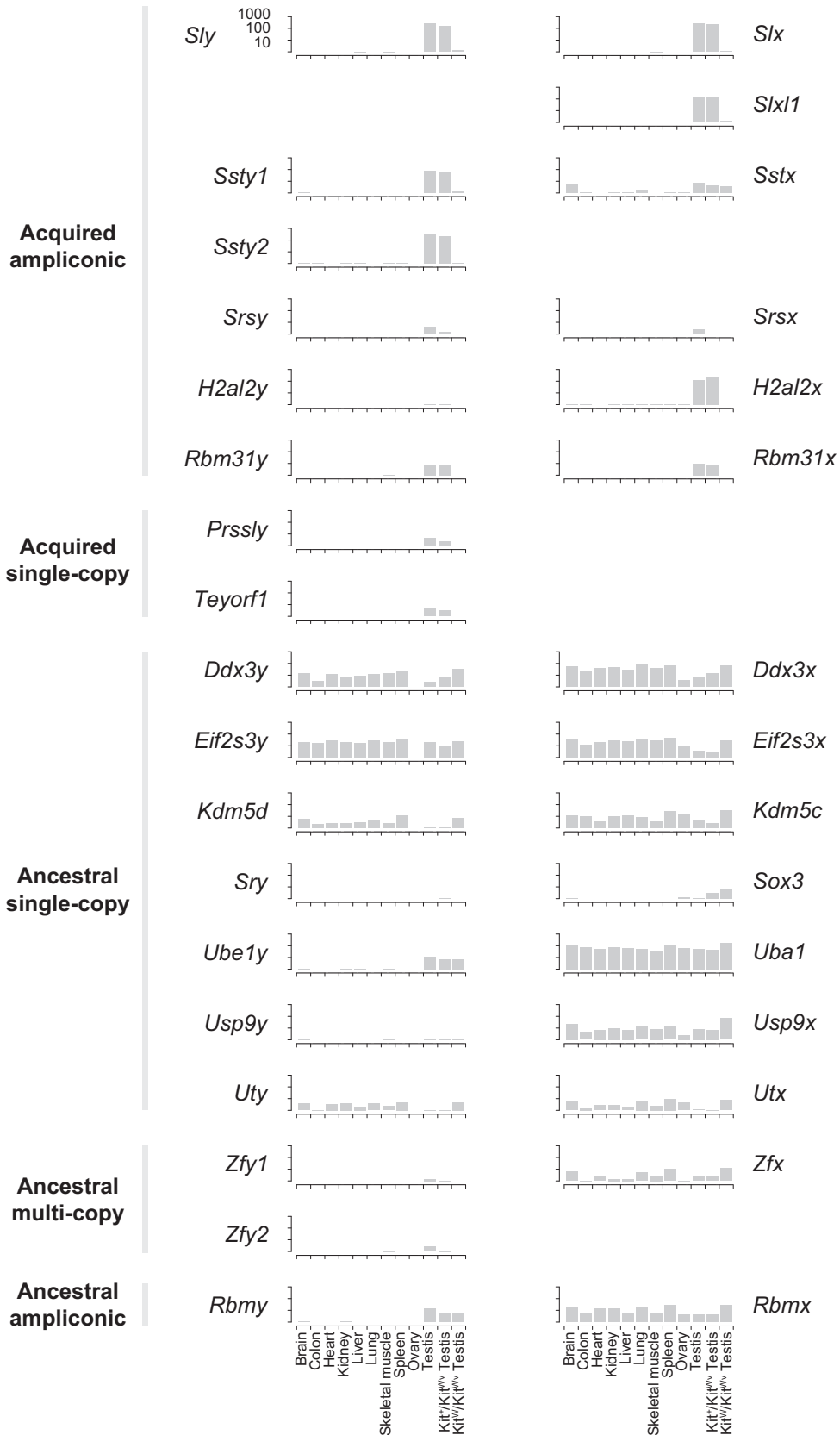**Multiple Alignments and Phylogenetic Analyses**

To align mouse MSY repeat units, we used Fast Statistical Alignment (Bradley et al., 2009). For phylogenetic analyses of X-Y ampliconic gene families, nucleotide sequences for members of each gene family were aligned using Clustalw (Larkin et al., 2007), and phylogenetic trees were generated by DNAML in the Phylip package (Felsenstein, 1989).

Ethics Statement

All experiments involving mice conformed to ethical principles and guidelines approved by the Committee on Animal Care at the Massachusetts Institute of Technology (Institutional Animal Care and Use Committee no. 0711-075-14).

ACCESSION NUMBERS

GenBank accession numbers for BAC sequences are listed in Tables S1 and S5. GenBank accession numbers for previously unreported genes are *H2al2y*, EU315689, NM_001113395; *Srsy*, EU052291; *Prssly*, KJ780361, XM_006544527; *Teyorf1*, KJ780362; *Rbm31x*, EU315690; and noncoding transcript, KC170991 (see also Table S3).

(legend on next page)

REFERENCES

Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. 14, 113–124.

Bailey, J.A., and Eichler, E.E. (2006). Primate segmental duplications: crucibles of evolution, diversity and disease. Nat. Rev. Genet. 7, 552–564.

Bellott, D.W., Skaletsky, H., Pyntikova, T., Mardis, E.R., Graves, T., Kremitzki, C., Brown, L.G., Rozen, S., Warren, W.C., Wilson, R.K., and Page, D.C. (2010). Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature 466, 612–616.

Bellott, D.W., Hughes, J.F., Skaletsky, H., Brown, L.G., Pyntikova, T., Cho, T.-J., Koutseva, N., Zaghlul, S., Graves, T., Rock, S., et al. (2014). Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature 508, 494–499.

Bergstrom, D.E., Grieco, D.A., Sonti, M.M., Fawcett, J.J., Bell-Prince, C., Cram, L.S., Narayanswami, S., and Simpson, E.M. (1998). The mouse Y chromosome: enrichment, sizing, and cloning by bivariate flow cytometry. Genomics 48, 304–313.

Bishop, C.E., and Hatat, D. (1987). Molecular cloning and sequence analysis of a mouse Y chromosome RNA transcript expressed in the testis. Nucleic Acids Res. 15, 2959–2969.

Bradley, R.K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I., and Pachter, L. (2009). Fast statistical alignment. PLoS Comput. Biol. 5, e1000392.

Burgoyne, P.S. (1998). The role of Y-encoded genes in mammalian spermatogenesis. Semin. Cell Dev. Biol. 9, 423–432.

Burgoyne, P.S., Mahadevaiah, S.K., Sutcliffe, M.J., and Palmer, S.J. (1992). Fertility in mice requires X-Y pairing and a Y-chromosomal "spermiogenesis" gene mapping to the long arm. Cell 71, 391–398.

Campbell, P., Good, J.M., Dean, M.D., Tucker, P.K., and Nachman, M.W. (2012). The contribution of the Y chromosome to hybrid male sterility in house mice. Genetics 191, 1271–1281.

Charlesworth, B., and Charlesworth, D. (2000). The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 355, 1563–1572.

Church, D.M., Goodstadt, L., Hillier, L.W., Zody, M.C., Goldstein, S., She, X., Bult, C.J., Agarwala, R., Cherry, J.L., DiCuccio, M., et al.; Mouse Genome Sequencing Consortium (2009). Lineage-specific biology revealed by a finished genome assembly of the mouse. PLoS Biol. 7, e1000112.

Cocquet, J., Ellis, P.J.I., Yamauchi, Y., Mahadevaiah, S.K., Affara, N.A., Ward, M.A., and Burgoyne, P.S. (2009). The multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis. PLoS Biol. 7, e1000244.

Cocquet, J., Ellis, P.J.I., Yamauchi, Y., Riel, J.M., Karacs, T.P.S., Rattigan, A., Ojarikre, O.A., Affara, N.A., Ward, M.A., and Burgoyne, P.S. (2010). Deficiency in the multicopy Sycp3-like X-linked genes Slx and Slxl1 causes major defects in spermatid differentiation. Mol. Biol. Cell 21, 3497–3505.

Cocquet, J., Ellis, P.J.I., Mahadevaiah, S.K., Affara, N.A., Vaiman, D., and Burgoyne, P.S. (2012). A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. PLoS Genet. 8, e1002900.

Conway, S.J., Mahadevaiah, S.K., Darling, S.M., Capel, B., Rattigan, A.M., and Burgoyne, P.S. (1994). Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse Y chromosome. Mamm. Genome 5, 203–210.

Déjardin, J., and Kingston, R.E. (2009). Purification of proteins associated with specific genomic loci. Cell 136, 175–186.

Eicher, E.M., Hutchison, K.W., Phillips, S.J., Tucker, P.K., and Lee, B.K. (1989). A repeated segment on the mouse Y chromosome is composed of retroviral-related, Y-enriched and Y-specific sequences. Genetics 122, 181–192.

Elliott, R.W., Miller, D.R., Pearsall, R.S., Hohman, C., Zhang, Y., Poslinski, D., Tabaczynski, D.A., and Chapman, V.M. (2001). Genetic analysis of testis weight and fertility in an interspecies hybrid congenic strain for Chromosome X. Mamm. Genome 12, 45–51.

Elliott, R.W., Poslinski, D., Tabaczynski, D., Hohman, C., and Pazik, J. (2004). Loci affecting male fertility in hybrids between Mus macedonicus and C57BL/6. Mamm. Genome 15, 704–710.

Ellis, P.J.I., Clemente, E.J., Ball, P., Touré, A., Ferguson, L., Turner, J.M.A., Loveland, K.L., Affara, N.A., and Burgoyne, P.S. (2005). Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids. Hum. Mol. Genet. 14, 2705–2715.

Ellis, P.J.I., Ferguson, L., Clemente, E.J., and Affara, N.A. (2007). Bidirectional transcription of a novel chimeric gene mapping to mouse chromosome Yq. BMC Evol. Biol. 7, 171.

Ellis, P.J.I., Bacon, J., and Affara, N.A. (2011). Association of Sly with sex-linked gene amplification during mouse evolution: a side effect of genomic conflict in spermatids? Hum. Mol. Genet. 20, 3010–3021.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166.

Ferguson, L., Ellis, P.J.I., and Affara, N.A. (2009). Two novel mouse genes mapped to chromosome Yp are expressed specifically in spermatids. Mamm. Genome 20, 193–206.

Ford, C.E. (1966). The murine Y chromosome as a marker. Transplantation 4, 333–335.

Figure 6. Expression of Mouse Y Chromosome Genes and Their X Homologs
Expression of Y genes and their X homologs in various adult tissues, as measured by RNA sequencing. We also measured expression in germ-cell-deficient (Kit^W/Kit^Wv) and wild-type control (Kit^+/Kit^Wv) testes. Expression is measured by the total number of reads aligning to each gene, normalized by the length of the gene (or the average length of the gene for a multicopy or ampliconic gene family), and the total number of reads mapped to the transcriptome.

Frank, S. (1991). Divergence of meiotic drive–suppression systems as an explanation for sex-biased hybrid sterility and inviability. Evolution *45*, 262–267.

Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Altshuler, D.M., Aburatani, H., Jones, K.W., Tyler-Smith, C., Hurles, M.E., et al. (2006). Copy number variation: new insights in genome diversity. Genome Res. *16*, 949–961.

Good, J.M., Dean, M.D., and Nachman, M.W. (2008). A complex genetic basis to X-linked hybrid male sterility between two species of house mice. Genetics *179*, 2213–2228.

Graves, J.A.M. (2006). Sex chromosome specialization and degeneration in mammals. Cell *124*, 901–914.

Hughes, J.F., Skaletsky, H., Pyntikova, T., Minx, P.J., Graves, T., Rozen, S., Wilson, R.K., and Page, D.C. (2005). Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature *437*, 100–103.

Hughes, J.F., Skaletsky, H., Pyntikova, T., Graves, T.A., van Daalen, S.K.M., Minx, P.J., Fulton, R.S., McGrath, S.D., Locke, D.P., Friedman, C., et al. (2010). Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. Nature *463*, 536–539.

Hughes, J.F., Skaletsky, H., Brown, L.G., Pyntikova, T., Graves, T., Fulton, R.S., Dugan, S., Ding, Y., Buhay, C.J., Kremitzki, C., et al. (2012). Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Nature *483*, 82–86.

Hurst, L.D. (1992). Is *Stellate* a relict meiotic driver? Genetics *130*, 229–230.

Hurst, L.D. (1996). Further evidence consistent with *Stellate's* involvement in meiotic drive. Genetics *142*, 641–643.

Hurst, L.D., and Pomiankowski, A. (1991). Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. Genetics *128*, 841–858.

Jaenike, J. (2001). Sex chromosome meiotic drive. Annu. Rev. Ecol. Syst. *32*, 25–49.

Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S., and Page, D.C. (2001). The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat. Genet. *29*, 279–286.

Lahn, B.T., and Page, D.C. (1997). Functional coherence of the human Y chromosome. Science *278*, 675–680.

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

Lahn, B.T., and Page, D.C. (2000). A human sex-chromosomal gene family expressed in male germ cells and encoding variably charged proteins. Hum. Mol. Genet. *9*, 311–319.

Lange, J., Skaletsky, H., van Daalen, S.K.M., Embry, S.L., Korver, C.M., Brown, L.G., Oates, R.D., Silber, S., Repping, S., and Page, D.C. (2009). Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. Cell *138*, 855–869.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Li, G., Davis, B.W., Raudsepp, T., Pearks Wilkerson, A.J., Mason, V.C., Ferguson-Smith, M., O'Brien, P.C., Waters, P.D., and Murphy, W.J. (2013). Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. Genome Res. *23*, 1486–1495.

Mazeyrat, S., Saut, N., Sargent, C.A., Grimmond, S., Longepied, G., Ehrmann, I.E., Ellis, P.S., Greenfield, A., Affara, N.A., and Mitchell, M.J. (1998). The mouse Y chromosome interval necessary for spermatogonial proliferation is gene dense with syntenic homology to the human *AZFa* region. Hum. Mol. Genet. *7*, 1713–1724.

Mazeyrat, S., Saut, N., Grigoriev, V., Mahadevaiah, S.K., Ojarikre, O.A., Rattigan A, Bishop, C., Eicher, E.M., Mitchell, M.J., and Burgoyne, P.S. (2001). A Y-encoded subunit of the translation initiation factor *Eif2* is essential for mouse spermatogenesis. Nat. Genet. *29*, 49–53.

McLaren, A., Simpson, E., Epplen, J.T., Studer, R., Koopman, P., Evans, E.P., and Burgoyne, P.S. (1988). Location of the genes controlling H-Y antigen expression and testis determination on the mouse Y chromosome. Proc. Natl. Acad. Sci. USA *85*, 6442–6445.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science *338*, 1593–1599.

Moens, P.B., and Spyropoulos, B. (1995). Immunocytology of chiasmata and chromosomal disjunction at mouse meiosis. Chromosoma *104*, 175–182.

Moriwaki, K., Suh, D., and Styrna, J. (1988). Genetic factors affecting sperm morphology in the mouse. Mouse Newsl. *82*, 138.

Mueller, J.L., Mahadevaiah, S.K., Park, P.J., Warburton, P.E., Page, D.C., and Turner, J.M.A. (2008). The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. Nat. Genet. *40*, 794–799.

Mueller, J.L., Skaletsky, H., Brown, L.G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren, W.C., Wilson, R.K., and Page, D.C. (2013). Independent specialization of the human and mouse X chromosomes for the male germ line. Nat. Genet. *45*, 1083–1087.

Murphy, W.J., Pearks Wilkerson, A.J., Raudsepp, T., Agarwala, R., Schäffer, A.A., Stanyon, R., and Chowdhary, B.P. (2006). Novel gene acquisition on carnivore Y chromosomes. PLoS Genet. *2*, e43.

Nishioka, Y., and Lamothe, E. (1986). Isolation and characterization of a mouse Y chromosomal repetitive sequence. Genetics *113*, 417–432.

Oh, B., Hwang, S.Y., Solter, D., and Knowles, B.B. (1997). *Spindlin*, a major maternal transcript expressed in the mouse during the transition from oocyte to embryo. Development *124*, 493–503.

Ohno, S. (1967). Sex chromosomes and sex-linked genes. Monographs on endocrinology, *Vol. 1* (Berlin, New York: Springer).

Palumbo, G., Bonaccorsi, S., Robbins, L.G., and Pimpinelli, S. (1994). Genetic analysis of *Stellate* elements of Drosophila melanogaster. Genetics *138*, 1181–1197.

Paria, N., Raudsepp, T., Pearks Wilkerson, A.J., O'Brien, P.C.M., Ferguson-Smith, M.A., Love, C.C., Arnold, C., Rakestraw, P., Murphy, W.J., and Chowdhary, B.P. (2011). A gene catalogue of the euchromatic male-specific region of the horse Y chromosome: comparison with human and other mammals. PLoS ONE *6*, e21374.

Partridge, L., and Hurst, L.D. (1998). Sex and conflict. Science *281*, 2003–2008.

Pertile, M.D., Graham, A.N., Choo, K.H.A., and Kalitsis, P. (2009). Rapid evolution of mouse Y centromere repeat DNA belies recent sequence stability. Genome Res. *19*, 2202–2213.

Phadnis, N., and Orr, H.A. (2009). A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science *323*, 376–379.

Phillips, S.J., Birkenmeier, E.H., Callahan, R., and Eicher, E.M. (1982). Male and female mouse DNAs can be discriminated using retroviral probes. Nature *297*, 241–243.

Prado, V.F., Lee, C.H., Zahed, L., Vekemans, M., and Nishioka, Y. (1992). Molecular characterization of a mouse Y chromosomal repetitive sequence that detects transcripts in the testis. Cytogenet. Cell Genet. *61*, 87–90.

Reijo, R., Lee, T.Y., Salo, P., Alagappan, R., Brown, L.G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O., et al. (1995). Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat. Genet. *10*, 383–393.

Repping, S., Skaletsky, H., Lange, J., Silber, S., Van Der Veen, F., Oates, R.D., Page, D.C., and Rozen, S. (2002). Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am. J. Hum. Genet. *71*, 906–922.

Repping, S., Skaletsky, H., Brown, L., van Daalen, S.K.M., Korver, C.M., Pyntikova, T., Kuroda-Kawaguchi, T., de Vries, J.W., Oates, R.D., Silber, S., et al. (2003). Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat. Genet. *35*, 247–251.

Reynard, L.N., Turner, J.M.A., Cocquet, J., Mahadevaiah, S.K., Touré, A., Höög, C., and Burgoyne, P.S. (2007). Expression analysis of the mouse multi-copy X-linked gene *Xlr-related, meiosis-regulated* (*Xmr*), reveals that *Xmr* encodes a spermatid-expressed cytoplasmic protein, SLX/XMR. Biol. Reprod. *77*, 329–335.

Reynard, L.N., Cocquet, J., and Burgoyne, P.S. (2009). The multi-copy mouse gene *Sycp3-like Y-linked* (*Sly*) encodes an abundant spermatid protein that interacts with a histone acetyltransferase and an acrosomal protein. Biol. Reprod. *81*, 250–257.

Rice, W.R. (1987). Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics *116*, 161–167.

Roberts, C., Weith, A., Passage, E., Michot, J.L., Mattei, M.G., and Bishop, C.E. (1988). Molecular and cytogenetic evidence for the location of *Tdy* and *Hya* on the mouse Y chromosome short arm. Proc. Natl. Acad. Sci. USA *85*, 6446–6449.

Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P., et al. (2005). The DNA sequence of the human X chromosome. Nature *434*, 325–337.

Saxena, R., Brown, L.G., Hawkins, T., Alagappan, R.K., Skaletsky, H., Reeve, M.P., Reijo, R., Rozen, S., Dinulos, M.B., Disteche, C.M., and Page, D.C. (1996). The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nat. Genet. *14*, 292–299.

Saxena, R., de Vries, J.W., Repping, S., Alagappan, R.K., Skaletsky, H., Brown, L.G., Ma, P., Chen, E., Hoovers, J.M.N., and Page, D.C. (2000). Four *DAZ* genes in two clusters found in the AZFc region of the human Y chromosome. Genomics *67*, 256–267.

Silver, L.M. (1995). Mouse Genetics: Concepts and Applications (New York: Oxford University Press).

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825–837.

Slonim, D., Kruglyak, L., Stein, L., and Lander, E. (1997). Building human genome maps with radiation hybrids. J. Comput. Biol. *4*, 487–504.

Smit, A., Hubley, R., and Green, P. (1996). RepeatMasker Open-3.0.

Styrna, J., Klag, J., and Moriwaki, K. (1991). Influence of partial deletion of the Y chromosome on mouse sperm phenotype. J. Reprod. Fertil. *92*, 187–195.

Tao, Y., Hartl, D.L., and Laurie, C.C. (2001). Sex-ratio segregation distortion associated with reproductive isolation in Drosophila. Proc. Natl. Acad. Sci. USA *98*, 13183–13188.

Touré, A., Szot, M., Mahadevaiah, S.K., Rattigan, A., Ojarikre, O.A., and Burgoyne, P.S. (2004a). A new deletion of the mouse Y chromosome long arm associated with the loss of *Ssty* expression, abnormal sperm development and sterility. Genetics *166*, 901–912.

Touré, A., Grigoriev, V., Mahadevaiah, S.K., Rattigan, A., Ojarikre, O.A., and Burgoyne, P.S. (2004b). 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.

Touré, A., Clemente, E.J., Ellis, P., Mahadevaiah, S.K., Ojarikre, O.A., Ball, P.A.F., Reynard, L., Loveland, K.L., Burgoyne, P.S., and Affara, N.A. (2005). Identification of novel Y chromosome encoded transcripts by testis transcriptome analysis of mice with deletions of the Y chromosome long arm. Genome Biol. *6*, R102.

Vernet, N., Mahadevaiah, S.K., Ojarikre, O.A., Longepied, G., Prosser, H.M., Bradley, A., Mitchell, M.J., and Burgoyne, P.S. (2011). The Y-encoded gene *Zfy2* acts to remove cells with unpaired chromosomes at the first meiotic metaphase in male mice. Curr. Biol. *21*, 787–793.

Vernet, N., Mahadevaiah, S.K., Ellis, P.J.I., de Rooij, D.G., and Burgoyne, P.S. (2012). Spermatid development in XO male mice with varying Y chromosome short-arm gene content: evidence for a Y gene controlling the initiation of sperm morphogenesis. Reproduction *144*, 433–445.

Wang, H., Hu, Y.-C., Markoulaki, S., Welstead, G.G., Cheng, A.W., Shivalila, C.S., Pyntikova, T., Dadon, D.B., Voytas, D.F., Bogdanove, A.J., et al. (2013a). TALEN-mediated editing of the mouse Y chromosome. Nat. Biotechnol. *31*, 530–532.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013b). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

Warburton, P.E., Giordano, J., Cheung, F., Gelfand, Y., and Benson, G. (2004). Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. Genome Res. *14* (10A), 1861–1869.

Yamauchi, Y., Riel, J.M., Stoytcheva, Z., and Ward, M.A. (2014). Two Y genes can replace the entire Y chromosome for assisted reproduction in the mouse. Science *343*, 69–72.