Sequence of the Mouse Y Chromosome

By

Jessica E. Alföldi

BSc.(Honours) Biochemistry
Queen's University at Kingston, 2001

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2008

Signature of author ______________________________
Department of Biology
January 14, 2008

Certified by: ______________________________
David C. Page
Professor of Biology
Director, Whitehead Institute
Investigator, Howard Hughes Medical Institute
Thesis Supervisor

Accepted by: ______________________________
Stephen P. Bell
Chair, Biology Graduate Student Committee

Sequence of the Mouse Y Chromosome

by

Jessica E. Alföldi

Submitted to the Department of Biology on January 15, 2008 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

ABSTRACT

The mouse Y chromosome has been studied for over 50 years, from the early sex determination and immunological phenotypes attributed to it in the 1950s, to the several mouse Y permatogenic phenotypes and the sex ratio distortion phenotype that are still being studied today. However, even though the draft sequence of the mouse genome was published in 2000, only 750 kb of the 95 Mb mouse Y chromosome was sequenced at that time. To fill that void, we are sequencing the male-specific portion of the mouse Y, and present the analysis of 72 Mb of the mouse Y chromosome here.

We found that the X-degenerate portion of the mouse Y chromosome, the portion that descends from the autosomal progenitors of the mammalian X and Y chromosomes, is highly degenerate, even more so than the human Y's X-degenerate sequence. But the ampliconic portion of the mouse Y has expanded to an incredible degree, taking up 95% of the chromosome.

Almost all of the ampliconic portion of the mouse Y chromosome is made up of a single 92 Mb segmental duplication that we have named the Huge Repeat array. The Huge Repeat array is a collection of 150-200 515 kb repeat units. The repeat units are internally repetitive, gene-containing, and have a repeat unit to repeat unit similarity ranging from 99% to 99.999%. This array has no homology to the human Y chromosome. The functions of the Huge Repeat are still unclear, but we do know that it is required for proper sperm head morphology and an equal sex ratio in offspring.

Due to the high sequence similarity and the high copy number of the Huge Repeat array, we had to invent new sequencing strategies for the sequencing of the mouse Y chromosome. We used a BAC sequencing approach where BACs were selected for sequencing in iterative rounds of sample sequencing. The mouse Y sequencing effort is still in progress.

Thesis Supervisor: David C. Page
Title: Professor of Biology
Director, Whitehead Institute
Investigator, Howard Hughes Medical Institute

Acknowledgements

I'd like to start by thanking the people who have directly contributed to this work. Helen Skaletsky and Steve Rozen have taught me everything I know about genomics, and spent a lot of their precious time teaching it to me. Helen, who did a lot of the programming and assembly and thinking for this project, is an amazing person and a great scientist and a good friend of mine and I'm glad I've had the chance to know her. This work, like so many others in the Page lab, would have been impossible to do without her. I'd also like to thank Steve for putting a lot of thought into the mouse Y project as well, and Tatyana Pyntikova for her work towards the BAC selection part of this project.

I'd also like to thank the many hundreds of people at the Washington University at St. Louis Genome Sequencing Center – our collaborators who did all the sequencing for this project and who have been great partners in every way. There are way too many people for me to thank but I'd especially like to mention Tina Graves and Pat Minx and Rick Wilson for all of their invaluable work on the mouse Y chromosome. And of course, I would like to thank my advisor, David Page, who has been a great mentor and teacher and advisor in general.

Next I would like to thank my thesis defense committee – both Dave Bartel and Chris Burge who have been on my committee for years and have given me great advice, and also Amy Keating who agreed to help me out even though my work is a bit out of her field. And finally I'd really like to thank Rick Wilson, who flew out to Boston just for my thesis defense, and didn't torment me too much as part of my defense committee.

Of course, I also want to give thanks for being part of the Page lab – who collectively have managed to make grad school as fun as humanly possible, especially when I just needed a kindly ear to complain to. I wanted to especially mention Julian Lange, who has always kept me well supplied with chocolate and advice, and is one of the reasons I joined the Page lab in the first place. And I also need to thank Jana Hersch, the best baymate anyone could ever have, and one of my best friends. If the only thing that came out of our doctorates was your increased understanding of sarcasm, then it would still have been well worthwhile. Jennifer Hughes, Jake Mueller and Julian (again) also require thanks for helping me with this thesis (and actually reading the whole thing!) I'd also like to thank the rest of my friends, both old friends in Canada, and my new friends in Boston and everywhere else: For as much as a modicum of sanity is required to finish grad school, you have made all this possible.

And I'd like to thank my family, all of whom have been incredibly supportive and proud of me even when they have no idea what I am talking about. I'd like to especially thank Tom Alföldi, my Dad, who brought me back the academic calendar and course listings of MIT when I was sixteen years old, and who eventually let me convince him that biology is a real science.

And finally, I'd like to thank Daniel Doty, who is the best boyfriend and husband-to-be in the whole world, and who I could not possibly thank enough.

Table of Contents

CHAPTER 1

Introduction

The sex chromosomes of all mammals including the mouse are the X and Y chromosomes. Female mammals have two copies of the X chromosome, while male mammals have one X and one Y. The X and the Y chromosomes were a pair of autosomes in the common ancestor of all mammals 300 million years ago, and today have diverged enough that large majority of the chromosomes cannot recombine during meiosis. A small part of the X and Y chromosomes do recombine during meiosis; this region is known as the pseudo-autosomal region (PAR).

In the introduction to my thesis, I will describe the history of the study of the mouse Y chromosome. I will be focusing on the discovery of Y-linked genes, Y-linked phenotypes, and the attribution of genes to those phenotypes. I will also discuss other major features of the mouse Y chromosome such as the centromere and the PAR. This chapter should provide an overview of the state of knowledge of the mouse Y chromosome before we sequenced it.

In the beginning

The study of sex chromosomes has been primarily viewed through the lenses of sex determination and germ cell biology, and so it may come as a surprise that the first reported phenotype of the mouse Y chromosome was immunological, and had nothing at all to do with sexual differentiation or testis development. In 1955, Eichwald and Silmser were developing skin graft techniques by practicing on mice. They unexpectedly found that female mice would reject skin grafts from their brothers, but not vice-versa [1]. At this point, they did not understand all of the implications of their finding, but soon it was established that male mice produce male-specific antigens from loci on the Y chromosome [2]. The functions of these proteins (named H-Y proteins), or the location of the genes encoding them (named *Hya*) were unknown, and this was the beginning of a 40 year search for their identity.

The association of sex determination with the mouse Y chromosome was soon to follow. In the 1950s, the sex determination system of mammals was still unknown. It was already established that fruit flies (*Drosophila melanogaster*) determined sex by the ratio of X chromosomes to autosomes; the presence or absence of the Y was unimportant. Then in 1959, Welshons and Russell discovered XO mice – mice that had the full complement of autosomes, but only one X chromosome and no Y chromosome. These XO mice were phenotypically female and fertile, leading the authors to conclude that the mouse Y chromosome, unlike the *Drosophila* Y chromosome, must be sex-determining [3].

The search for *Tdy* and *Hya*

The next step in the search for the mouse sex-determining factor (which would soon be named *Tdy* - *Testis determining factor, Y-linked*) was the discovery of the *Sxr*^{a} phenotype. By 1971 it was well established that mice that carried the Y chromosome became male, and those that did not became female. Then Cattanach and colleagues discovered XO and XX mice that developed testes. They searched for a translocation from the Y chromosome to an autosome, came up empty-handed and attributed the sex reversal phenotype to an autosomal dominant mutation. What they had actually found was the *Sxr*^{a} translocation – a translocation that would be the primary focus of mouse Y chromosome research for the next 20 years [4].

The searches for the *Tdy* and *Hya* loci were joined in 1981, when Simpson and colleagues showed that *Sxr*^{a} mice (that were still posited to have an autosomal dominant mutation) also produced the H-Y antigens [5]. Earlier unpublished evidence of this linkage had already led others to propose that the two loci were one and the same [6]. This would not be disproved for several more years.

In 1982, *Sxr*^{a} was finally shown to be a translocation from the Y chromosome, and not an autosomal dominant mutation. Evans and colleagues provided cytological evidence that *Sxr*^{a} mice were created throught several steps (see Figure 1, a-c): First, one end of the Y chromosome was translocated to the other end and attached to the pseudo-autosomal region. In the second step, the *Sxr*^{a} chromosomal segment gets transferred from the Y to the X chromosome through crossing-over in the PAR during meiosis. The next generation of mice carry an X chromosome with the Y-derived *Sxr*^{a}. Thus, both XX*Sxr*^{a} and X*Sxr*^{a}O mice will have testes even without having a visible Y chromosome. The *Tdy* and *Hya* loci were now determined to be located at the non-PAR end of the mouse Y chromosome [7]. Ashley and colleagues found that the event leading to duplication of the *Sxr*^{a} region is due to a recombination between telomeric sequences. This was the first time that a telomere was found to participate in a chromosomal rearrangement [8].

The discovery of a *Sxr*^{b}, a variant of *Sxr*^{a} served to finally determine that *Tdy* and *Hya* are separate loci. McLaren and colleagues found a single mouse among their XX*Sxr*^{a} sex-reversed mice that did not produce the H-Y antigens. They did not know the genetic constitution of the mouse at the time, but we know now that it contained *Sxr*^{b}, a derived variant of *Sxr*^{a} attached to one of its X chromosomes (see Figure 1d) [9]. Upon closer examination of the XX*Sxr*^{b} mice, Burgoyne and colleagues discovered that XX*Sxr*^{b} mice were not only lacking the H-Y antigens, they also exhibited an earlier spermatogenic

Figure 1 Several partial deletions and transpositions of the mouse Y chromosome. Chromosomes depicted are not to scale. Horizontal bars denote the relative placement of genes. Horizontal rectangles denote the relative placement and size of the *Rbmy* gene family. Diagonal lines are used to indicate the location of the *Ssty* gene family. Y chromosome originating sequence is shown in blue, X chromosome originating sequence is shown in pink and PAR sequence is shown in purple. The location of the mouse Y centromere is unknown.

a A wild-type mouse Y chromosome

b The YSxr^a chromosome, where the Sxr^a region has been duplicated and transposed to the other end of the Y.

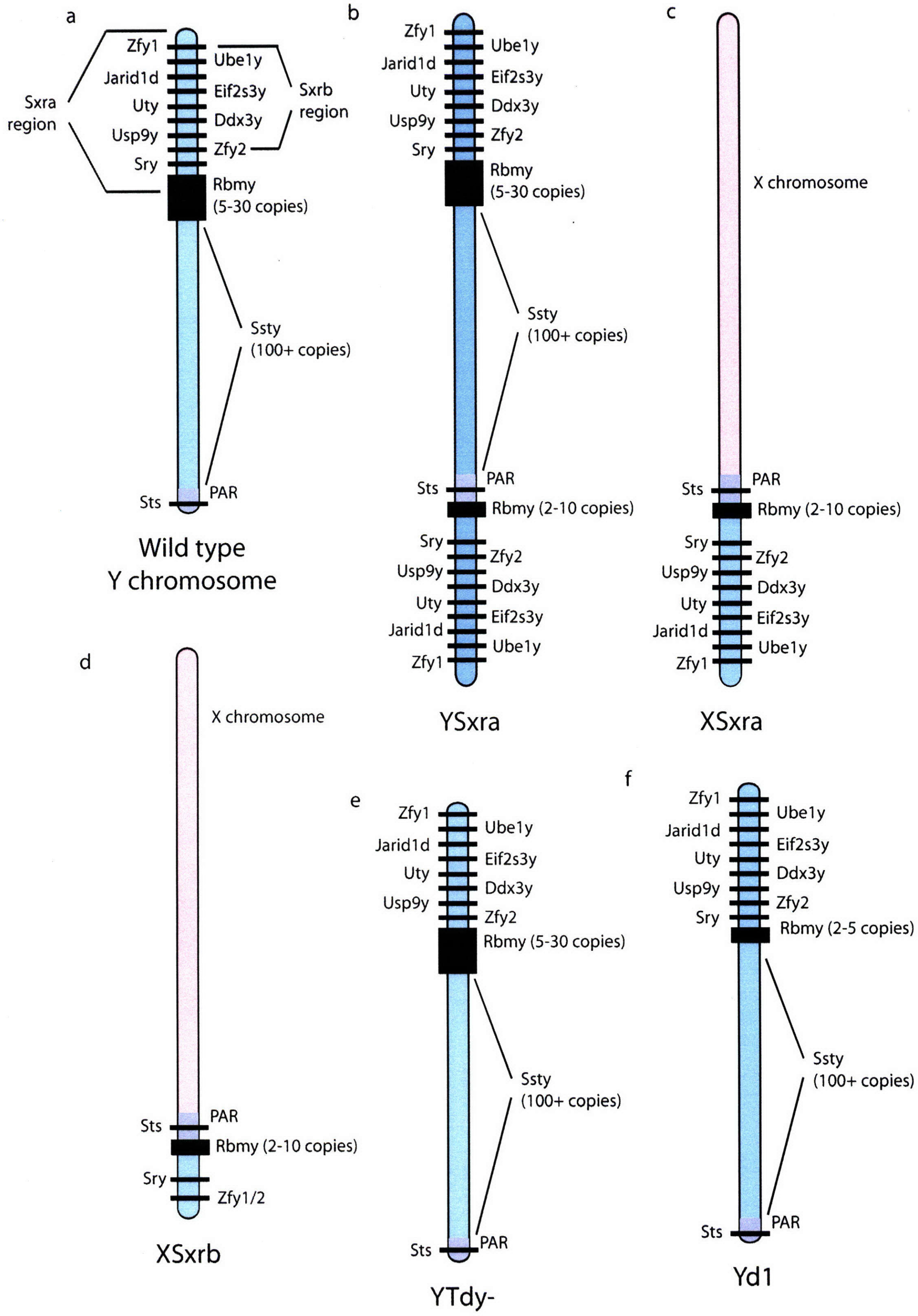
c The XSxr^a chromosome, where the Sxr^a region has been transferred to the X chromosome from the YSxr^a chromosome via the pseudoautosomal region.

d The XSxr^b chromosome, where an interstitial deletion has occurred in the Sxr^a portion of an XSxr^a chromosome

e The YTdy- chromosome, where a 14 kb segment including the *Sry* gene has been deleted from the Y chromosome.

f The Yd1 chromosome, where an interstitial deletion has removed most of the *Rbmy* family.

Figure 1


a
Sxra
region
Zfy1
Jarid1d
Uty
Usp9y
Sry
Ube1y
Eif2s3y
Ddx3y
Zfy2
Sxrb
region
Rbmy
(5-30 copies)
Ssty
(100+ copies)
PAR
Sts
Wild type
Y chromosome
b
Zfy1
Jarid1d
Uty
Usp9y
Sry
Ube1y
Eif2s3y
Ddx3y
Zfy2
Rbmy
(5-30 copies)
Ssty
(100+ copies)
Sts
PAR
Rbmy (2-10 copies)
Sry
Usp9y
Uty
Jarid1d
Zfy1
Zfy2
Ddx3y
Eif2s3y
Ube1y
YSxra
c
X chromosome
Sts
PAR
Rbmy (2-10 copies)
Sry
Usp9y
Uty
Jarid1d
Zfy1
Zfy2
Ddx3y
Eif2s3y
Ube1y
XSxra
d
X chromosome
Sts
PAR
Rbmy (2-10 copies)
Sry
Zfy1/2
XSxrb
e
Zfy1
Jarid1d
Uty
Usp9y
Ube1y
Eif2s3y
Ddx3y
Zfy2
Rbmy (5-30 copies)
Ssty
(100+ copies)
Sts
PAR
YTdy-
f
Zfy1
Jarid1d
Uty
Usp9y
Sry
Ube1y
Eif2s3y
Ddx3y
Zfy2
Rbmy (2-5 copies)
Ssty
(100+ copies)
Sts
PAR
Yd1

failure than XX*Sxr*^{a} mice. They could then ascribe a spermatogenic function to a portion of *Sxr*^{a} not present in *Sxr*^{b}.[10] This function was later named *Spy* (*Spermatogonial proliferation factor, Y-linked*).

In 1988, McLaren and colleagues found that a mouse of their *Sxr*^{b} line that they expected to have a X*Sxr*^{b}X genotype, unexpectedly produced the H-Y antigens, and no longer exhibited the Southern blot patterns charcteristic of *Sxr*^{b}. They proposed that this occurred by a recombination in the mouse's X*Sxr*^{b}Y father, where the *Sxr*^{b} moiety crossed-over with the complete Y chromosome, leading to an X*Sxr*^{a} chromosome descending to the anomalous mouse in question. They also noted that this theory required that the *Sxr*^{a} portion of the normal Y chromosome be on a short arm of the mouse Y [11]. This theory shortly became dogma in the mouse Y community, even though the evidence for the existence of a mouse Y short arm is far from clear.

Concurrently, Roberts and colleagues discovered the *Sxr*^{b} variant as well, and also promulgated the theory that the *Sxr*^{a} region must be found on a separate short arm. They also demonstrated that the *Sxr*^{b} segment was derived from *Sxr*^{a} by an interstitial deletion. Therefore, an XX*Sxr*^{a} mouse contains a segment of Y chromosomal DNA that an XX*Sxr*^{b} mouse does not. This Y chromosomal segment presumably contains both the *Hya* and *Spy* loci [12]. The *Sxr*^{b}-producing interstitial deletion was later shown to be mediated by a recombination between two paralogous genes in the *Sxr*^{a} region, *Zfy1* and *Zfy2* (*Zinc finger protein, Y-linked*) [13].

Sry, the sex-determining gene

After a prolonged and competitive search, the *Tdy* sex-determining locus was identified as the newly-discovered gene *Sry* (S*ex-determining region, Y-linked*). The *SRY* gene was first discovered in humans[14], but was first shown to be sex-determining in mice. Gubbay and colleagues infected mouse ES cells with a retrovirus to screen for phenotypically female XY mice [15]. Strangely enough, it was later found that this original *Sry* mutation could not have been generated by retroviral infection, and must have have occurred naturally. (see Figure 1e)[16] Soon thereafter, Koopman and colleagues showed that XX mice containing a genomic *Sry* transgene were phenotypically male (although sterile) [17]. *Sry* was later shown to be sex-determining in humans and other mammals [18].

In 1992, Gubbay and colleagues examined the genomic sequence surrounding the mouse *Sry* locus and found that *Sry*'s ORF was partly in the arm of an inverted repeat, and partly in the spacer between the two arms. This is true for the *Mus musculus Sry* locus, and for other closely related mice, but not in humans [19]. This led to another novel finding: the differential expression of *Sry* transcripts in embryos and adults. The embryonic sex-determining *Sry* transcript begins in the spacer of the inverted repeat, and continues into one arm. However, in adult mice, *Sry* transcription begins in one arm of the inverted repeat, continues through the unique spacer and ends in the other arm. Capel and colleagues found that this transcript is then spliced into a circle. It is still unknown whether this adult circular *Sry* transcript has any function [20].

The continued search for *Hya*

In the next several years, many scientists tried to map the *Sxr^a* region in order to discover the identity of *Hya*. McLaren and colleagues found many partial deletions of the *Sxr^a* region of the mouse Y chromosome, created by ectopic crossing-over between a *Sxr^a* region attached to a mouse's X chromosome, and the normal *Sxr^a* region near the centromere of the Y chromosome. They used these interstitial deletions to create the first maps of the mouse Y chromosome, ordering the very few known loci [21]. Later, Capel and colleagues, using the same approach to map the *Sxr*^{a} region, discovered an unusual phenotype from one of their partial *Sxr*^{a} deletions. Mice containing the Yd1 mutation were phenotypically female even though their *Sry* locus was intact (see Figure 1f). The Yd1 deletion was in fact, over 100 kb from the *Sry* locus. Capel and colleagues assumed this was due to a long range position effect, but the exact cause of this sex-reversal is still unknown [22]. Mitchell and colleagues took a different approach, creating cosmid and phage libraries containing small mouse genomic DNA inserts to map the *Sxr*^{a} region [23].

The search for *Hya* was continued in earnest in 1994 when King and colleagues created partial deletions of the *Sxr*^{a} region using an X*Sxr*^{a}O cell line. Each deletion line was then tested for the presence of H-Y antigens. Not only did this create an extensive map of the region, but also determined that there were at least two, and as many as five loci responsible for the production of the H-Y male specific antigens [24]. Then, in 1995, Scott and colleagues showed that the already known *Sxr*^{b} region gene *Jarid1d* (*Jumonji, AT-rich, interactive domain, 1d*) encoded one of the H-Y antigens. Its X homolog, *Jarid1c*, did not encode the same epitope [25]. The second and final H-Y antigen was identified by Greenfield and colleagues by screening for sex-specific transcripts in mouse embryos. They identified a new gene, *Uty* (*Ubiquitously transcribed tetratricopeptide repeat gene, Y-linked*), and showed that it also encoded a H-Y antigen. They then demonstrated that *Uty* and *Jarid1d* together completely accounted for the *Hya* phenotype, explaining the original mouse Y chromosome immunological phenotype [26].

Mouse Y chromosome genes

Some of the first mouse Y chromosome genes were discovered in the late 1980s as a consequence of the search for *Tdy*, the testis determining factor. I have already detailed the discovery of mouse *Sry*, but chronologically, the *Zfy* family was discovered first. In 1987, Page and colleagues had published the cloning of human *ZFY*, a gene they mistakenly hypothesized was *Tdy* [27]. As a result, there was intense competition to find the mouse homolog of this putative testis determining factor. In 1989, four papers were published describing the mouse *Zfy* family [28-31]. Collectively, they showed that the mouse Y chromosome contains two *Zfy* family members (*Zfy1* and *Zfy2*), published the cDNA sequence for both genes, showed that the genes were expressed in the mouse testis, and found polymorphisms in the genes between different subspecies of *Mus musculus*. Mardon and colleagues also theorized on whether one or both of the *Zfy* genes could be *Tdy*, but this would soon prove to be a moot point, as *Sry* was shown functionally to be *Tdy* by 1990.

After Sry, the next mouse Y chromosome gene to be discovered was *Ube1y* (*Ubiquitin-activating enzyme E1, Y-linked*). *Ube1y* is not found on the human Y chromosome, but does have an X-linked homolog, *Ube1x,* in both mice and humans. Mitchell and colleagues found that *Ube1y* was expressed exclusively in the testis. They also used a Southern 'zooblot' to show that homologs of *Ube1y* could be found in horse, pig, and rabbit [32]. Much later, in 2000, Levy and colleagues showed that even though there was only one functional copy of *Ube1y* in the mouse, the mouse Y chromosome also contained six *Ube1y* pseudo-genes [33]. In 1994, the previously mentioned *Jarid1d* was discovered on the mouse Y chromosome. Agulnik and colleagues showed that *Jarid1d* was located in the region deleted in *Sxr*^{b} mutants (Δ*Sxr*^{b}) and that it was expressed ubiquitously in male adult and embryonic tissues. They also used a zooblot to show that *Jarid1d* has homologs in many mammals, even as distant as marsupials [34].

As I described earlier, Capel and colleagues discovered sex-reversed mice with an intact Sry locus. They named this deletion Yd1. In 1995, Laval and colleagues followed up on this work and recreated Yd1 in crosses between different Mus musculus subspecies. They then discovered the presence of *Rbmy* (*RNA binding motif, Y-linked*), the mouse homolog of an already known human Y chromosome gene. They showed that this gene is present in several copies on the mouse Y, and that many of those copies are deleted in the Yd1 mutant [35]. Shortly thereafter, Elliott and colleagues analysed the expression of the *Rbmy* family and found them expressed exclusively in the germ cells of the testis, both in the embryo and the adult [36].

By 1998, Mahadevaiah and colleagues had analysed the Yd1 mutant mice further and found that they exhibited a new phenotype. XYd1 mice are phenotypically female, due to the unknown sex-reversing effect. But if an *Sry* transgene is added to these mice, they become phenotypically male, but they exhibit a defect in spermiogenesis, the post-meiotic stage of sperm development. In other words, they produce sperm with abnormal heads – a phenotype distinct from that of the previously described *Spy* locus [37]. Assuming that this abnormal head phenotype was due to the absence of *Rbmy*, Szot and colleagues added back a *Rbmy* transgene with a spermatid-specific promoter. Unfortunately, there was no rescue of the phenotype. They then discovered that they had been mistaken in the expression of *Rbmy* – it was not expressed in spermatids at all [38].

Also in 1998, Mazeyrat and colleagues discovered two new mouse Y genes in the Δ*Sxr*^{b} interval: *Ddx3y* (*DEAD box polypeptide 3, Y-linked*) and *Eif2s3y* (*Eukaryotic* *initiation factor 2, subunit 3, Y-linked*). To do this they first created a 750 kb contig made of BACs and cosmids spanning the entire Δ*Sxr*^{b} region from *Zfy1* to *Zfy2*. Then they exon trapped that region to pull up any new genes. This was in the hope of identifying candidates for the *Spy* spermatogonial phenotype [39]. The search for new Y genes was sometimes conducted in paralled in both mice and humans as shown when Brown and colleagues discovered *Usp9y* (*Ubiquitin specific peptidase 9, Y-linked*), another gene found in the Δ*Sxr*^{b} interval, in both species [40].

Some Y-linked genes present on the human Y chromosome are not present on the mouse Y chromosome. For example, Vogel and colleagues searched for a mouse homolog of the human *TSPY* gene, but only identified a pseudogene with no intact open reading frame. As *TSPY* homologs had already been found in bulls, they hypothesized that the gene had lost its function during rodent evolution [41]. Boettger-Tong and colleagues also tried to exon-trap a contig they had made covering the Δ*Sxr*^{b} interval, but the only new transcripts they found were pseudogene homologs of the *RHOA* family that they named *RhoAy1*, *2*, and *3* [42].

The gene responsible for the *Spy* spermatogonial proliferation phenotype was finally identified in 2001 when two different groups used the same add-back strategy to tease out its identity. Both groups took X*Sxr*^{b}O mice and added back individual genes from the Δ*Sxr*^{b} interval to see which gene could rescue the phenotype. Agulnik and colleagues only succeeded in showing that Jarid1d did not rescue the *Spy* phenotype [43]. However, Mazeyrat and colleagues discovered that *Eif2s3y* rescued the *Spy* phenotype. X*Sxr*^{b}O + *Eif2s3y* mice were not entirely normal though; while their spermatogonia did proliferate normally, they were still blocked before the second meiotic division. This was expected as it was the same phenotype found in X*Sxr*^{a}O mice [44]. This also marked the end of new gene discovery on the supposed 'short arm' of the mouse Y chromosome until we began our sequencing efforts.

The mystery of the mouse Y centromere

All published work on the mouse Y chromosome since 1988 has assumed that the mouse Y, unlike all the other mouse chromosomes, has a short arm, and that the *Sxr*^{a} segment of the chromosome is found on that short arm. There is not much data supporting this theory, and there is also not much data to refute it.

The cytogeneticists of the 1960s wanted to find a way to distinguish the mouse Y chromosome from chromosome 19, since they appeared to be about the same size under the microscope. Ford contended that the Y chromosome could be distinguished from chromosome 19 by the visible short arm of the mouse Y. However, he then qualified this statement by clarifying that he could not determine whether the Y chromosome really had a short arm, or just had a large centromeric region [45]. Others could not distinguish the Y chromosome from chromosome 19 by this criterion [46]. Unfortunately, this less than ringing endorsement was often used later to assert the existence of a mouse Y short arm [47].

The placement of the mouse Y centromere was largely forgotten until 1988, when two different groups studying the *Sxr*^{a} translocation came to the same conclusion. They believed that the only way that the *Sxr*^{a} region could have been transposed to the mouse Y PAR is if it was on a different arm of the mouse Y chromosome. Since most people cannot cytogenetically distinguish a mouse Y short arm, both groups assumed that this short arm must be less than 5 Mb in size. They then cited back to the cytogenetic research of the 1960s that I previously mentioned [11, 12]. One of the groups, Roberts and colleagues, went further and hybridized *Sxr*^{a} probes to the mouse Y chromosome and stated that they could assign these probes definitively to a short arm. Still, it is difficult for a reader using this data to determine whether the *Sxr*^{a} region is on a short arm, or merely adjacent to the centromere of a telocentric chromosome [12]. However, we do know that the mouse Y centromere is at the opposite end of the chromosome from its pseudoautosomal region. Using novel staining techniques, Schendl stained all the mouse centromeres including the Y centromere and showed its approximate location. Unfortunately, he could shed no light on the question of the existence of the mouse Y short arm [48].

Not only is the location of the mouse Y centromere uncertain, but the sequence of this centromere is also unknown. In 1970, Pardue and Gall hybridized mouse satellite DNA to mouse cells and noted that the satellite DNA stained the centromeric regions of almost all the chromosomes. That is, satellite DNA hybridized to every single mouse chromosome except for the Y chromosome [49]. Subsequent studies showed that both the major and the minor satellite sequences, found in all other mouse centromeres, are completely absent from the mouse Y chromosome [50, 51]. The sequence of the mouse Y centromere is still not known, and this has not helped in the search for its specific location.

The mouse's pseudoautosomal region

PARs are different from the rest of the Y chromosome in that they can regularly recombine with the X chromosome. In fact, the X and Y chromosome PARs are as similar as two autosomes. The mouse Y chromosome contains a single PAR, unlike the two PARs of the human Y. The mouse PAR was first described by Sachs in 1955 when he depicted the relative placement of the X chromosome and the Y chromosome at meiosis. He stated that during meiosis, the two chromosomes paired end-to-end, but he did not think it was likely that they formed true chiasmata and crossed-over [52]. This was disputed by Ohno and colleagues in 1959, who described in fine detail the interactions between the X and the Y chromosomes at meiosis, and showed that crossing-over did occur [53].

The first understanding of the mouse PAR sequence came from Nagamine and colleagues's study of XY*Sxr*^{a} mice. They found that the gene *Sts* (*Steroid sulfatase*) was tightly linked with the *Sxr*^{a} region, demonstrating that *Sts* was located on the mouse pseudoautosomal region. The human homolog, *STS*, is located just on the X-specific portion of the human X chromosome and not in the human PAR, thereby showing that the two human and mouse pseudoautosomal regions could not be completely homologous [54]. In 1996, Salido and colleagues cloned and sequenced *Sts* in the mouse and showed that it physically mapped to the mouse PAR [55].

In 1997, Palmer and colleagues found another locus mapping to the mouse PAR – the gene *Mid1* (*Midline 1*). *Mid1* spans the pseudoautosomal boundary (PAB) on only the X chromosome. This results in a complete copy of *Mid1* on the X chromosome, but only the 3' end of the gene on the Y chromosome. Therefore, we cannot really call it a pseudoautosomal gene, but its existence did help to define the pseudoautosomal boundary [56].

Several years later, Perry and colleagues built a Yeast Artificial Chromosome (YAC) map of the whole distal end of the mouse X chromosome in order to define the

PAB. They also determined the size of the mouse PAR by integrating a restriction enzyme site into the X chromosome near the PAB. They then used pulse field gel electrophoresis to show that the PAR was 700 kb long [57].

The *Yaa* locus

In 1980, a rather unexpected phenotype was attributed to the mouse Y chromosome – an accelerator of autoimmunity. Male BXSB strain mice were known to be particularly susceptible to a lupus-like autoimmune disease. Eisenberg and Dixon castrated the mice at a young age and showed that since the male mice were still prone to lupus, this must be attributed to a genetic effect from the Y chromosome, and not to their hormones [58]. Several years later, Izui and colleagues found that in placing the BXSB Y chromosome in different mouse backgrounds that the BXSB Y could accelerate autoimmune disease in mice that already possessed other susceptibility loci, but had no effect on most lab strains, such as C57BL/6. It was now hard to deny that a gene on the Y chromosome was responsible for this effect – and the locus was named *Yaa* (*Y chromosome-linked autoimmune acceleration*) [59]. Narrowing the region containing *Yaa* was not an easy task as the Y chromosome was always inherited as a whole and could only be subdivided by natural deletions or translocations [60].

The mystery was finally solved in 2006, when Subramanian and colleagues showed that the *Yaa* locus was not, in fact, a normal part of the Y. In BXSB mice, the most distal part of the X chromosome had been translocated onto the Y chromosome, creating an extra large PAR (see Figure 2a-b). They first suspected that this was the case when they found higher expression levels of a cluster of distal X genes in mice with the *Yaa* phenotype by microarray. They then confirmed the translocation by FISH. It is

Figure 2 Several more partial deletions and transpositions of the mouse Y chromosome. Chromosomes depicted are not to scale. Horizontal bars denote the relative placement of genes. Horizontal rectangles denote the relative placement and size of the *Rbmy* gene family. Diagonal lines are used to indicate the location of the *Ssty* gene family. Y chromosome originating sequence is shown in blue, X chromosome originating sequence is shown in pink and PAR sequence is shown in purple. The location of the mouse Y centromere is unknown.

a A wild-type mouse Y chromosome

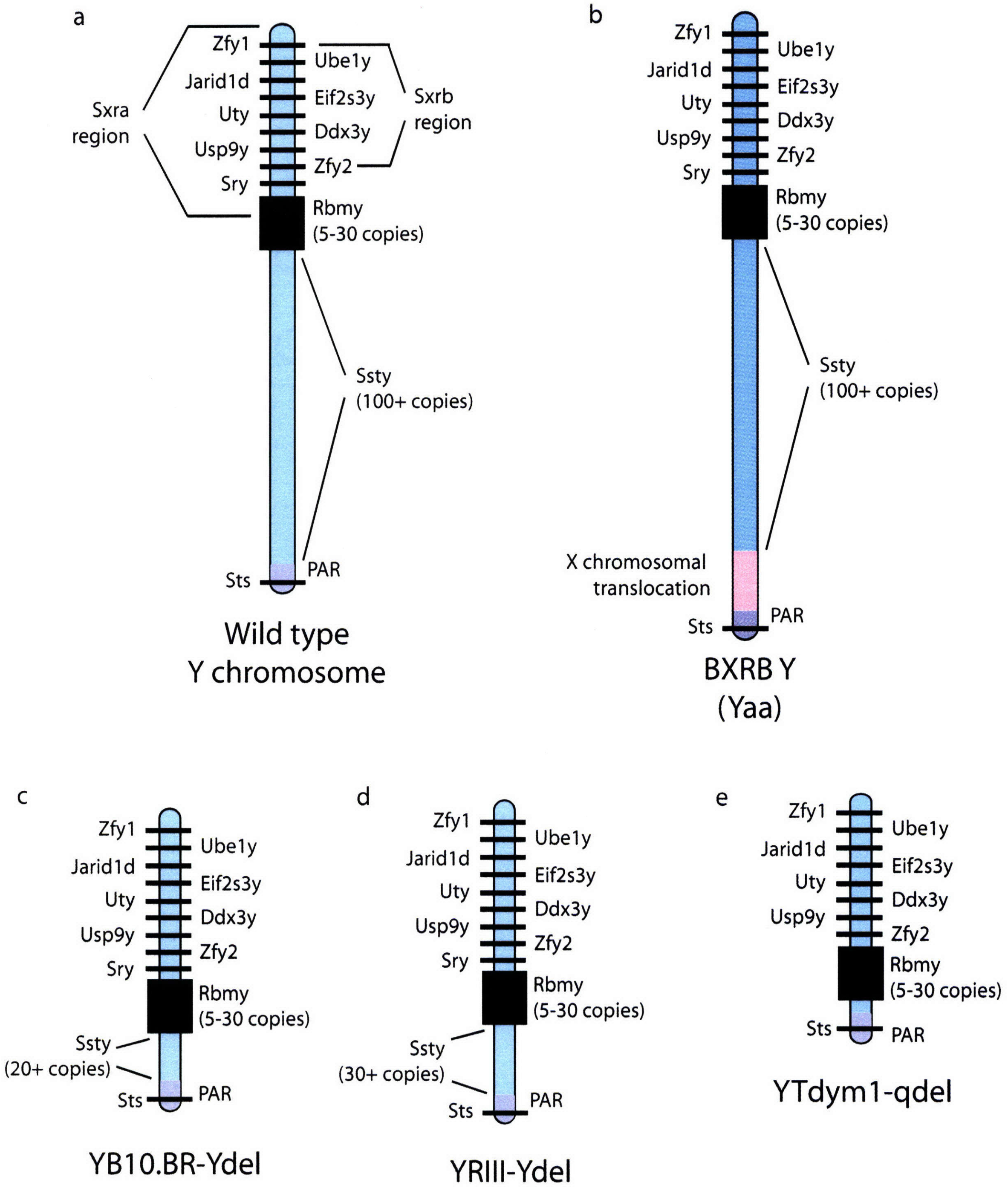
b The BXRB strain Y chromosome, a chromosome conferring the *Yaa* phenotype. Part of the distal X chromosome has been transposed onto this Y chromosome near the PAR.

c The YB10.BR-Ydel chromosome, a chromosome one third the size of a normal Y chromosome, but still containing the entire *Sxr*^{a} region.

d The YRIII-Ydel chromosome, a chromosome one quarter the size of a normal Y chromosome, but still containing the entire *Sxr*^{a} region.

e The YTdym1-qdel chromosome, a chromosome originating from the YTdy- chromosome and one tenth the size of a normal Y chromosome. It still retains almost all of the *Sxr*^{a} region, except for 14 kb surrounding *Sry*.

Figure 2


a
Zfy1
Jarid1d
Uty
Usp9y
Sry
Sxra
region
Ube1y
Eif2s3y
Ddx3y
Zfy2
Sxrb
region
Rbmy
(5-30 copies)
Ssty
(100+ copies)
Sts
PAR
Wild type
Y chromosome
b
Zfy1
Jarid1d
Uty
Usp9y
Sry
Ube1y
Eif2s3y
Ddx3y
Zfy2
Rbmy
(5-30 copies)
Ssty
(100+ copies)
X chromosomal
translocation
Sts
PAR
BXRB Y
(Yaa)
c
Zfy1
Jarid1d
Uty
Usp9y
Sry
Ube1y
Eif2s3y
Ddx3y
Zfy2
Rbmy
(5-30 copies)
Ssty
(20+ copies)
Sts
PAR
YB10.BR-Ydel
d
Zfy1
Jarid1d
Uty
Usp9y
Sry
Ube1y
Eif2s3y
Ddx3y
Zfy2
Rbmy
(5-30 copies)
Ssty
(30+ copies)
Sts
PAR
YRIII-Ydel
e
Zfy1
Jarid1d
Uty
Usp9y
Ube1y
Eif2s3y
Ddx3y
Zfy2
Rbmy
(5-30 copies)
Sts
PAR
YTdym1-qdel

presumably a higher dosage of genes in this X chromosomal region that leads to a *Yaa* phenotype [61].

Mouse Y repetitive sequences

In 1982, Phillips and colleagues started a longstanding trend by reporting highly repetitive sequences found on the Y chromosomes of many mouse species. In this first case, they identified retroviral sequences that were found in up to one hundred copies on the mouse Y chromosome. These repeats were among the first few sequences assigned to the mouse Y chromosome. From the quantity of retroviral sequences they found, they assumed that up to 3% of the chromosome could be composed of retroviral DNA. They found similar quantities of these Y-specific sequences in other members of the *Mus* genus. This was the first time anyone reported numerous retroviral sequences on the same chromosome [62].

Bishop and colleagues used a different multi-copy Y sequence to probe ten different inbred mouse lines and determine whether their Y chromosomes originated from the *Mus musculus musculus* subspecies (found in eastern Europe and much of Asia) or from the *Mus musculus domesticus* subspecies (found in western Europe). Surprisingly, they discovered that nine out of those ten strains carried a *Mus musculus musculus* Y chromosome. They hypothesized that this imbalance occurred through the whims of the mouse fanciers that originated most of today's inbred mouse strains [63]. Bishop and colleagues used a random Y probe (Y353/B) for this study, but this probe would prove to be very important in the future study of the mouse Y chromosome. Tucker and colleagues expanded on Bishop and colleagues's work in 1992, using several probes to examine the origins of 39 inbred mouse strains. They found that 31 of the 39 strains possessed a *Mus musculus musculus* Y chromosome [64]. Tucker and colleagues also used a 305 bp multi-copy Y probe to construct a phylogenetic tree of the *Mus* genus [65].

Soon thereafter, several papers were published describing a wide variety of repetitive mouse Y-specific sequences. These sequences were highly repetitive and were estimated to occur in as many as 200 copies on the mouse Y chromosome. In many cases, these Y-repetitive sequences were found to be conserved in the close relatives of *Mus* musculus, and less conserved in distant relatives within in the *Mus* genus [66]. However, sometimes, these sequences exhibited strange phylogenetic patterns. For instance, one sequence described by Nishioka was shown to be present in many copies in the closest relatives of *Mus musculus* and also more distant members of other subgenera, but were absent in other *Mus* species of an intermediate distance [67]. This led Nishioka and Dolan to conclude that these highly repetitive Y sequences must be evolutionary unstable, and are perhaps evolving at an unusually fast rate [68].

In 1989, Eicher and colleagues followed up on Phillips and colleagues in analyzing multi-copy mouse Y-specific retroviral sequences. Specifically, they analysed a 7 kb fragment that hybridized specifically to the *Mus musculus* Y chromosome approximately 500 times. They identified a new family of Y-specific retroviruses that they named MuRVY (Murine Repeated Virus, Y-linked). They found that one copy of MuRVY was contained in their 7 kb fragment and that this fragment was also found in many copies of other *Mus* Y chromosomes [69]. In that same year, Hutchison and Eicher cloned and sequenced MuRVY [70]. Fennelly and colleagues expanded on this work in 1996, when they showed that MuRVY occurred in a repeat unit alongside another retrovirus, IAPE-Y (Intracisternal A Particle – Y-linked), in a repeat unit of 25 kb. They believed that if there were 500 of these retroviral repeat units on the mouse Y chromosome, they could make up 20% of the entire chromosome. They confirmed that MuRVY could be found on many other *Mus* Y chromosomes, and showed that IAPE-Y was confined to the *Mus musculus* species [71].

Navin and colleagues identified six novel mouse Y repetitive sequences in 1996 through a technique known as RDA (representational difference analysis). Essentially, they selected for mouse Y specific sequences by subtracting a female mouse genome from a male mouse genome [72]. Bergstrom and colleagues expanded on this work the next year by slightly altering the RDA protocol, which resulted in the discovery of ten more Y-specific repetitive probes. They hoped to use these probes to aid in the construction of a contiguous clone-based map of the mouse Y chromosome [73].

Bergstrom and colleagues took a different approach in 1998 when they purified the Y chromosome by bivariate flow cytometry. They then used their purified mouse Y DNA to create Y-enriched small insert plasmid libraries. They confirmed the enrichment of Y DNA by identifying many known Y sequences in their libraries. They thoroughly characterized 103 clones, and produced 50 kb of mouse Y sequence [74]. Another important piece of data that emerged from the flow cytometry purification of the mouse Y was a more accurate size estimate. Previously, cytogeneticists had estimated the mouse Y chromosome's size at 50-60 Mb [75]. But bivariate flow cytometry estimated the mouse Y chromosome to be 94.7 Mb in length [74]. This paper marked the end of published attempts to clone random mouse Y sequences.

'Long arm' deletions and genes

As I stated before, the location of the mouse Y centromere is unknown. However, since any sequence not found in the *Sxr*^{a} region has conventionally been referred to as belonging to the mouse Y 'long arm', I will use that designation for the rest of this section.

Earlier, I mentioned the random Y chromosome probe Y353/B that Bishop and colleagues used to determine the origins of the Y chromosomes of many inbred mouse strains. In 1987, Bishop and Hatat began to understand the importance of this Y probe, when they discovered that it matched a large family of testis transcripts. When they performed FISH using a Y353/B probe, they found that it painted the whole of the mouse Y chromosome, except for the (minute) *Sxr*^{a} region of the Y. This newly-discovered long arm gene family, which was soon to be named *Ssty* (*Spermiogenesis-specific transcript, Y-linked*) was actually the first gene ever discovered on the mouse Y chromosome [76]. Prado and colleagues discovered two other cDNAs that were 84-97% identical to Y353/B and showed that they too were found in hundreds of copies on the mouse Y and were expressed in the testis. They also discovered that there are *Ssty* homologs in the female mouse genome [77].

In 1991, a new area of mouse Y chromosome research opened when Styrna and colleagues discovered the first mice with partial deletions of their mouse Y long arm – named B10.BR-Ydel (see Figure 2c). These mice had a significantly shorter Y chromosome, (later estimated to be 25% the size of a normal Y) and produced many more sperm with deformed heads. While wild type mice produced only 23% abnormal sperm, B10.BR-Ydel mice made 64% abnormal sperm. These mice were also generally less fertile and produced noticeably more female offspring than male [78]. Xian and colleagues examined this sperm defect further and showed that B10.BR-Ydel mice had lower fertility even when their sperm was used in *in vitro* fertilization [79].

As I discussed earlier, X*Sxr*^{a}O mice are phenotypically male and develop testes. Nevertheless X*Sxr*^{a}O mice are sterile as their germ cells cannot complete meiosis. In 1992, Burgoyne and colleagues showed that the meiotic block could be rescued by providing the X*Sxr*^{a} chromosome with a meiotic pairing partner. The rescued mouse's germ cells could then complete meiosis even if the meiotic pairing partner contained no mouse Y DNA. However, even when meiosis was completed, the mice were not fertile, as all the sperm they produced had grossly abnormal heads. They noted that the B10.BR-Ydel phenotype was similar to the one they had produced, but was less severe. This was expected since their X*Sxr*^{a} mice were missing much more of the Y chromosome than the B10.BR-Ydel mice. They then decided that there must be a spermiogenesis (post-meiotic sperm development) factor on the mouse Y long arm, and that the factor was probably in multiple copies, which would be consistent with varying degrees of the sperm head defect in different abnormal Y chromosomes. *Ssty*, being the only known multi-copy Y long arm-linked testis gene, was a promising candidate [80].

Conway and colleagues discovered another partial deletion of the mouse Y long arm in 1994, which they named RIII-Ydel (see Figure 2d). The RIII-Ydel Y chromosome was slightly larger than the B10.BR-Ydel chromosome (about 33% the size of a normal Y) and also resulted in abnormal sperm heads, decreased fertility, and more female offspring. They also showed that RIII-Ydel mice had many fewer genomic copies and transcripts of *Ssty* than the wild-type Y chromosome. Since *Ssty* was also shown here to be solely expressed in round spermatids (the stage just before the sperm head is formed), this supported *Ssty*'s possible role in the long arm deletion phenotypes [81].

After a long lull, the study of mouse Y long arm deletions resumed when Styrna and colleagues further examined their B10.BR-Ydel mutant in 2002. They found that the mutant mice contained degenerate testicular tubules and that even when sperm reached the uterus and oviducts of female mice, offspring were only sometimes produced [82]. Later, Grzmil and colleagues showed that the B10.BR-Ydel mice also had impaired sperm motility [83].

Then in 2004, Toure and colleagues discovered a third partial deletion of the mouse Y chromosome – the most severe yet. The new mutant was more difficult to analyze as the deletion had occurred in a mouse deleted for *Sry*. Therefore, in order to examine the phenotype of the new deletion, an *Sry* transgene had to be added to the mice. This new deletion was named YTdym1-qdel and the YTdym1-qdel chromosome was approximately 10% the size of a normal Y chromosome (Figure 2e). Toure and colleagues found that YTdym1-qdel mice were completely infertile, produced only sperm with severely abnormal heads, and had no detectable *Ssty* expression [84].

Toure and colleagues performed further analysis of the *Ssty* gene family, and determined that it could be subdivided into two subfamilies: *Ssty1* and *Ssty2*. They also produced an antibody to the SSTY1 protein. This protein was expressed from an unusual *Ssty1* locus, which contained an intron within its 5' region (all previously known *Ssty* copies were intronless). They showed that no SSTY1 protein was produced in X*Sxr*^{a} mice, and that RIII-Ydel mice had normal levels of the SSTY1 protein. This meant that *Ssty1* could not completely account for the long arm deletion phenotypes [85].

In 2005, Ellis and colleagues provided more insight into the workings of these long arm deletions when they used microarrays to examine differences in testis transcription between several mutants and wild type mice. They used mice with the following mutant chromosomes: RIII-Ydel, YTdym1-qdel, and X*Sxr*^{a}. They discovered that mice with long arm deletions showed higher transcription levels of many X linked genes, implying that a function of the mouse Y long arm is to repress those genes. Interestingly, the long arm deleted mice had higher levels of *Slx* (*Sycp3-like, X-linked*), the multi-copy X homolog of *Sly* (*Sycp3-like, Y-linked*). *Sly* is a multi-copy Y long arm gene that we had discovered in the course of sequencing the mouse Y chromosome. Ellis and colleagues proposed that the X and Y chromosomes were competing to be passed on to more offspring, in a process known as meiotic drive. In this situation, if the Y chromosome was partially deleted, the Y-linked genes could not suppress the X-linked genes as efficiently, and more female offspring would be produced [86]. This theory dovetails perfectly with the sex ratio distortion phenotype of mouse Y long arm deletions.

Ward and Burgoyne examined the sex ratio distortion phenotype further when they used intracytoplasmic sperm injection (ICSI) on sperm from mice with partial long arm deletions. In this procedure, sperm are injected directly into the egg, which should rescue any phenotype related to sperm motility or fertilization ability. They noted that when mice with YTdym1-qdel or RIII-Ydel chromosomes were injected into eggs, an equal amount of male and female offspring were produced. They also observed that both types of mutant mice produce equal numbers of X and Y sperm and that *in vitro* fertilization does not rescue the sex ratio distortion. They could then conclude that the increased number of female offspring from long arm deleted mice was due to a fertilization defect in the Y bearing sperm. Now, two functions could be attributed to the long arm of the mouse Y chromosome, a sperm head morphology function, and a meiotic drive function.

Summary

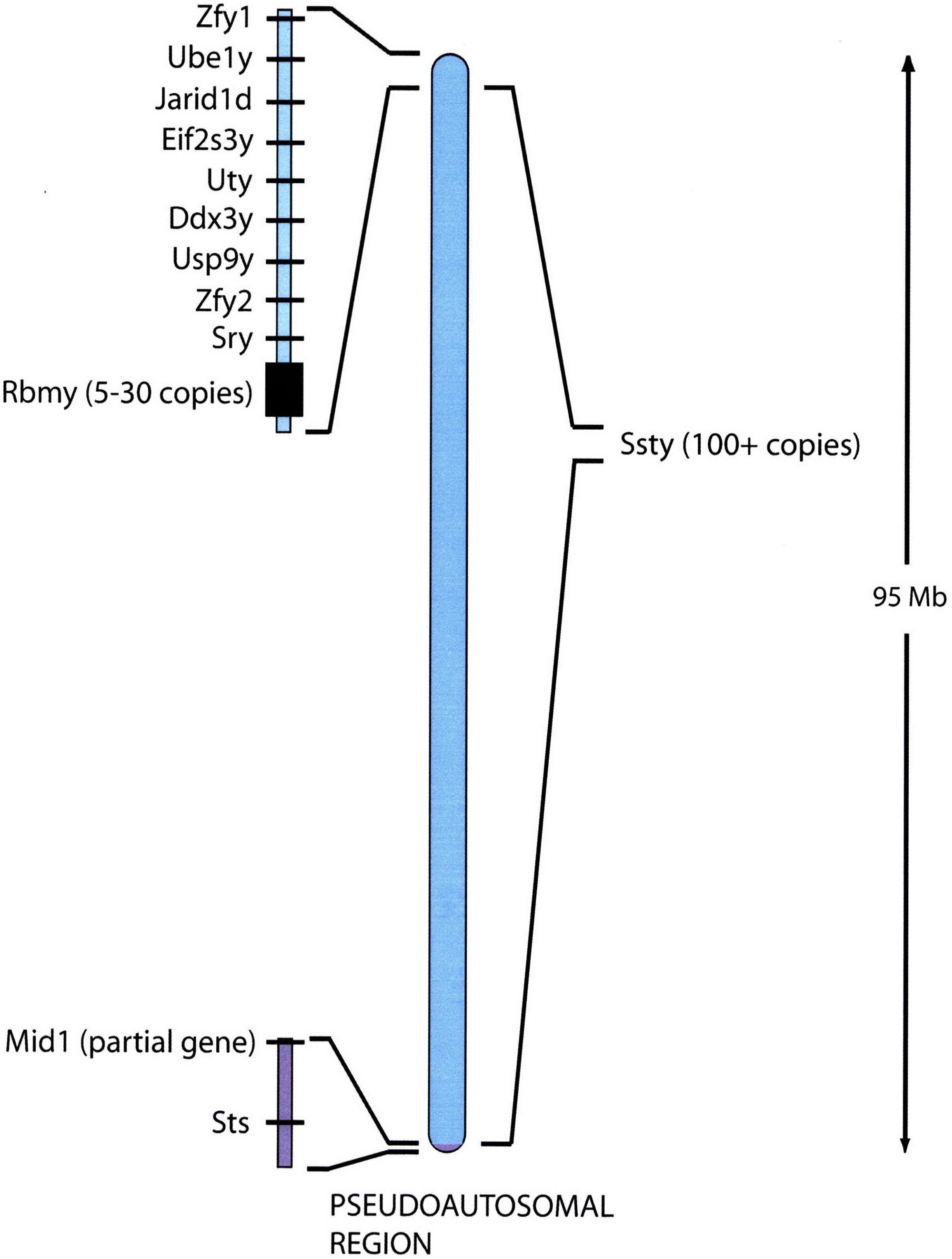
Before the sequencing of the mouse Y chromosome, knowledge about the genomic content of the chromosome was very limited (Figure 3). The mouse Y chromosome was predicted to be 95 Mb long, and to contain a short arm containing the *Sxr*^{a} region. All other mouse Y sequence was believed to be on the long arm. The *Sxr*^{a} region was well mapped, with BAC contigs extending over 2 Mb of sequence. Also, eight genes and two gene families were mapped to that region. The Δ*Sxr*^{b} region, a 750 kb segment of *Sxr*^{a}, had already been sequenced.

As for the rest of the chromosome, no serious sequencing attempts had been made, and there was an insignificant amount of sequence available. All sequences known to exist on the 'long arm' of the mouse Y chromosome were highly repetitive. This included a couple of retroviral sequences and the *Ssty* gene family. The pseudoautosomal boundary of the mouse Y had been discovered, and the gene *Mid1* was shown to straddle the border on the X chromosome side. Another gene, *Sts*, was known to be located in the mouse PAR. The sequence of the PAR was considered to be part of the mouse X chromosome sequencing project, but was difficult to clone into BACs or plasmids and was therefore still unsequenced.

Several functions of the mouse Y chromosome had been discovered and many had been attributed to specific genes. Researchers had found that *Sry* was the sex-determining gene, that *Eif2s3y* was required to complete male meiosis, and that *Jarid1d*

Figure 3 A depiction of the state of mouse Y chromosome knowledge as of 2002. The chromosome is shown to scale. The *Sxr*^{a} region is shown in an enlargement at the top of the figure. Relative gene locations were known but not their exact locations. The pseudoautosomal region is enlarged at the bottom of the figure. Genes are indicated by horizontal bars, including the partial gene, *Mid1*. The *Rbmy* gene family is designated by a black rectangle. The location of the *Ssty* gene family is also shown. The location of the mouse Y centromere was unknown.

Figure 3


Zfy1
Ube1y
Jarid1d
Eif2s3y
Uty
Ddx3y
Usp9y
Zfy2
Sry
Rbmy (5-30 copies)
Ssty (100+ copies)
95 Mb
Mid1 (partial gene)
Sts
PSEUDOAUTOSOMAL
REGION

and *Uty* produced male-specific antigens. It had also been determined that genes outside of the *Sxr*^{a} region were required for proper sperm head morphology and to raise the levels of male offspring to parity with female offspring. However the genes responsible for these last two functions were still unknown. It should also be noted that the only mouse Y chromosome genes with known functions were *Sry* and *Eif2s3y*.

There were many other mysteries surrounding the mouse Y chromosome. We did not know the location or even the nature of the mouse Y centromere. The 'long arm' of the mouse Y was especially puzzling – such a large piece of euchromatic sequence with this level of repetitiveness had never been seen before. Another puzzling mystery was that even with years of attempts, no one had managed to create a knockout of a mouse Y gene. As I mentioned earlier, the *Sry* 'knockout' was later found to be a naturally occurring deletion, and not the product of mutagenesis. In 2002, Rohozinski and colleagues published a paper declaring that they had successfully targeted *Eif2s3y* via homologous recombination in ES cells. However, a mouse was never produced [87]. The same year, Simpson and colleagues provided a strategy for the maintenance of future mouse Y gene knockouts. This would be necessary as some mouse Y genes are required for fertility. They proposed that mouse Y knockouts should be generated in a XY*Sxr*^{a} ES cell line that they created. The extra *Sxr*^{a} segment on the end of the Y's PAR could complement any gene knocked out from the normal *Sxr*^{a} region [88]. Still, no one has ever explained why knocking out a genes via homologous recombination has failed so consistently and specifically on the mouse Y chromosome.

Aims of my thesis

The focus of my graduate work has been the sequencing and analysis of the male-specific portion of the mouse Y chromosome. There are three major reasons why obtaining and analysing its sequence would be both important and interesting.

First, that the mouse Y chromosome should be sequenced to help provide a finished version of the mouse genome. When the draft sequence of the mouse genome was published in 2002, it did not include the Y chromosome. The Mouse Sequencing Consortium sequenced a female mouse [89]. This was not unexpected, as out of all the genomes sequenced to date, only the human and chimp Y chromosomes have been sequenced. Sequencing the homogametic sex (the sex with identical chromosomes, ie XX) is highly recommended in sequencing projects for two different reasons. If a male mouse had been sequenced, there would have been half as much sequence produced of the X and the Y chromosome as of the autosomes, resulting in X and Y assemblies of poor quality. Also, Y chromosomes are known to be difficult to sequence because of their repetitiveness.

The second reason why the sequence of the mouse Y chromosome would be useful comes from the sequencing of the human Y chromosome. When the human Y chromosome was sequenced, it revealed remarkable structures known as palindromes. Palindromes are large inverted repeats with a unique region, known as a spacer, separating the two repeated arms. The arms can have as high a sequence identity as 99.99% and often contain testis genes [90]. I hoped that the mouse Y chromosome would contain similar palindromes. This would not only demonstrate that palindromes are a common feature of mammalian Y chromosomes, but would also provide a more tractable model organism for the study of these palindromes.

Third, I hoped that in sequencing the mouse Y chromosome I would learn something from comparing the mouse Y sequence to the already complete human Y sequence and the nearly complete chimpanzee Y sequence. Humans and chimpanzees are very close relatives, separated by only six million years. Comparing those Y chromosomes could tell us quite a bit about recent primate Y evolution. But in order to know more about the general properties of mammalian Y chromosomes and to examine the earlier history of the mammalian sex chromosomes, we would need to sequence the Y chromosome of a more distant relative such as the mouse (separated from humans by 100 million years).

To accomplish these goals, I, along with many others, began the sequencing of the mouse Y chromosome. Chapter 2 of this thesis will describe the process of sequencing the mouse Y, and will also discuss our analysis of the mouse Y sequence obtained to date, about 75% of the chromosome. In Chapter 3, I will explain the impact of the sequencing of the mouse Y chromosome on the scientific community and I will propose further mouse Y experimentation and sequencing.

1. Eichwald, E.J. and C.R. Silmser, *Skin.* Transplant Bull, 1955. **2**: p. 148-9.
2. Eichwald, E.J., C.R. Silmser, and I. Weissman, *Sex-linked rejection of normal and neoplastic tissue. I. Distribution and specificity.* J Natl Cancer Inst, 1958. **20**(3): p. 563-75.
3. Welshons, W.J. and L.B. Russell, *The Y-Chromosome as the Bearer of Male Determining Factors in the Mouse.* Proc Natl Acad Sci U S A, 1959. **45**(4): p. 560-6.
4. Cattanach, B.M., C.E. Pollard, and S.G. Hawker, *Sex-reversed mice: XX and XO males.* Cytogenetics, 1971. **10**(5): p. 318-37.
5. Simpson, E., et al., *H-Y antigen in Sxr mice detected by H-2-restricted cytotoxic T cells.* Immunogenetics, 1981. **13**(4): p. 355-8.
6. Wachtel, S.S., et al., *Possible role for H--Y antigen in the primary determination of sex.* Nature, 1975. **257**(5523): p. 235-6.
7. Evans, E.P., M.D. Burtenshaw, and B.M. Cattanach, *Meitoic crossing-over between the X and Y chromosomes of male mice carrying the sex-reversing (Sxr) factor.* Nature, 1982. **300**(5891): p. 443-5.

8. Ashley, T., J. Lieman, and D.C. Ward, *Multicolor FISH with a telomere repeat and Sry sequences shows that Sxr (Sex reversal) in the mouse is a new type of chromosome rearrangement.* Cytogenet Cell Genet, 1995. **71**(3): p. 217-22.
9. McLaren, A., et al., *Male sexual differentiation in mice lacking H-Y antigen.* Nature, 1984. **312**(5994): p. 552-5.
10. Burgoyne, P.S., E.R. Levy, and A. McLaren, *Spermatogenic failure in male mice lacking H-Y antigen.* Nature, 1986. **320**(6058): p. 170-2.
11. McLaren, A., et al., *Location of the genes controlling H-Y antigen expression and testis determination on the mouse Y chromosome.* Proc Natl Acad Sci U S A, 1988. **85**(17): p. 6442-5.
12. Roberts, C., et al., *Molecular and cytogenetic evidence for the location of Tdy and Hya on the mouse Y chromosome short arm.* Proc Natl Acad Sci U S A, 1988. **85**(17): p. 6446-9.
13. Simpson, E.M. and D.C. Page, *An interstitial deletion in mouse Y chromosomal DNA created a transcribed Zfy fusion gene.* Genomics, 1991. **11**(3): p. 601-8.
14. Sinclair, A.H., et al., *A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif.* Nature, 1990. **346**(6281): p. 240-4.
15. Gubbay, J., et al., *A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes.* Nature, 1990. **346**(6281): p. 245-50.
16. Lovell-Badge, R. and E. Robertson, *XY female mice resulting from a heritable mutation in the primary testis-determining gene, Tdy.* Development, 1990. **109**(3): p. 635-46.
17. Koopman, P., et al., *Male development of chromosomally female mice transgenic for Sry.* Nature, 1991. **351**(6322): p. 117-21.
18. Berta, P., et al., *Genetic evidence equating SRY and the testis-determining factor.* Nature, 1990. **348**(6300): p. 448-50.
19. Gubbay, J., et al., *Inverted repeat structure of the Sry locus in mice.* Proc Natl Acad Sci U S A, 1992. **89**(17): p. 7953-7.
20. Capel, B., et al., *Circular transcripts of the testis-determining gene Sry in adult mouse testis.* Cell, 1993. **73**(5): p. 1019-30.
21. McLaren, A., et al., *Recombination between the X and Y chromosomes and the Sxr region of the mouse.* Genet Res, 1992. **60**(3): p. 175-84.
22. Capel, B., et al., *Deletion of Y chromosome sequences located outside the testis determining region can cause XY female sex reversal.* Nat Genet, 1993. **5**(3): p. 301-7.
23. Mitchell, M.J. and C.E. Bishop, *A structural analysis of the Sxr region of the mouse Y chromosome.* Genomics, 1992. **12**(1): p. 26-34.
24. King, T.R., et al., *Deletion mapping by immunoselection against the H-Y histocompatibility antigen further resolves the Sxra region of the mouse Y chromosome and reveals complexity of the Hya locus.* Genomics, 1994. **24**(1): p. 159-68.
25. Scott, D.M., et al., *Identification of a mouse male-specific transplantation antigen, H-Y.* Nature, 1995. **376**(6542): p. 695-8.

26. Greenfield, A., et al., *An H-YDb epitope is encoded by a novel mouse Y chromosome gene.* Nat Genet, 1996. **14**(4): p. 474-8.

27. Page, D.C., et al., *The sex-determining region of the human Y chromosome encodes a finger protein.* Cell, 1987. **51**(6): p. 1091-104.

28. Nagamine, C.M., et al., *Chromosome mapping and expression of a putative testis-determining gene in mouse.* Science, 1989. **243**(4887): p. 80-3.

29. Mardon, G., et al., *Duplication, deletion, and polymorphism in the sex-determining region of the mouse Y chromosome.* Science, 1989. **243**(4887): p. 78-80.

30. Mardon, G. and D.C. Page, *The sex-determining region of the mouse Y chromosome encodes a protein with a highly acidic domain and 13 zinc fingers.* Cell, 1989. **56**(5): p. 765-70.

31. Ashworth, A., S. Swift, and N. Affara, *Sequence of cDNA for murine Zfy-1, a candidate for Tdy.* Nucleic Acids Res, 1989. **17**(7): p. 2864.

32. Mitchell, M.J., et al., *Homology of a candidate spermatogenic gene from the mouse Y chromosome to the ubiquitin-activating enzyme E1.* Nature, 1991. **354**(6353): p. 483-6.

33. Levy, N., et al., *The ubiquitin-activating enzyme E1 homologous genes on the mouse Y chromosome (Ube1y) represent one functional gene and six partial pseudogenes.* Mamm Genome, 2000. **11**(2): p. 164-8.

34. Agulnik, A.I., et al., *A mouse Y chromosome gene encoded by a region essential for spermatogenesis and expression of male-specific minor histocompatibility antigens.* Hum Mol Genet, 1994. **3**(6): p. 873-8.

35. Laval, S.H., et al., *Y chromosome short arm-Sxr recombination in XSxr/Y males causes deletion of Rbm and XY female sex reversal.* Proc Natl Acad Sci U S A, 1995. **92**(22): p. 10403-7.

36. Elliott, D.J., et al., *An RBM homologue maps to the mouse Y chromosome and is expressed in germ cells.* Hum Mol Genet, 1996. **5**(7): p. 869-74.

37. Mahadevaiah, S.K., et al., *Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities.* Hum Mol Genet, 1998. **7**(4): p. 715-27.

38. Szot, M., et al., *Does Rbmy have a role in sperm development in mice?* Cytogenet Genome Res, 2003. **103**(3-4): p. 330-6.

39. Mazeyrat, S., et al., *The mouse Y chromosome interval necessary for spermatogonial proliferation is gene dense with syntenic homology to the human AZFa region.* Hum Mol Genet, 1998. **7**(11): p. 1713-24.

40. Brown, G.M., et al., *Characterisation of the coding sequence and fine mapping of the human DFFRY gene and comparative expression analysis and mapping to the Sxrb interval of the mouse Y chromosome of the Dffry gene.* Hum Mol Genet, 1998. **7**(1): p. 97-107.

41. Vogel, T., et al., *A murine TSPY.* Chromosome Res, 1998. **6**(1): p. 35-40.

42. Boettger-Tong, H.L., et al., *Transposition of RhoA to the murine Y chromosome.* Genomics, 1998. **49**(2): p. 180-7.

43. Agulnik, A.I., W.R. Harrison, and C.E. Bishop, *Smcy transgene does not rescue spermatogenesis in sex-reversed mice.* Mamm Genome, 2001. **12**(2): p. 112-6.

44. Mazeyrat, S., et al., *A Y-encoded subunit of the translation initiation factor Eif2 is essential for mouse spermatogenesis.* Nat Genet, 2001. **29**(1): p. 49-53.

45. Ford, C.E., *The murine Y chromosome as a marker.* Transplantation, 1965. **4**(3): p. 333-335.

46. Eichwald, E.J., N. Davidson, and M. Moore, *The murine Y chromosome as a marker.* Transplantation, 1965. **4**(2): p. 332-3.

47. Mitchell, M.J., *Spermatogenesis and the mouse Y chromosome: specialisation out of decay.* Results Probl Cell Differ, 2000. **28**: p. 233-70.

48. Schnedl, W., *End-to-end association of X and Y chromosomes in mouse meiosis.* Nat New Biol, 1972. **236**(62): p. 29-30.

49. Pardue, M.L. and J.G. Gall, *Chromosomal localization of mouse satellite DNA.* Science, 1970. **168**(937): p. 1356-8.

50. Broccoli, D., et al., *Isolation of a variant family of mouse minor satellite DNA that hybridizes preferentially to chromosome 4.* Genomics, 1991. **10**(1): p. 68-74.

51. Matsuda, Y. and V.M. Chapman, *In situ analysis of centromeric satellite DNA segregating in Mus species crosses.* Mamm Genome, 1991. **1**(2): p. 71-7.

52. Sachs, L., *The possibilities of crossing-over between the sex chromosomes of the house mouse.* Genetica, 1955. **27**(5-6): p. 309-22.

53. Ohno, S., W.D. Kaplan, and R. Kinosita, *On the end-to-end association of the X and Y chromosomes of Mus musculus.* Exp Cell Res, 1959. **18**: p. 282-90.

54. Nagamine, C.M., et al., *Linkage of the murine steroid sulfatase locus, Sts, to sex reversed, Sxr: a genetic and molecular analysis.* Nucleic Acids Res, 1987. **15**(22): p. 9227-38.

55. Salido, E.C., et al., *Cloning and expression of the mouse pseudoautosomal steroid sulphatase gene (Sts).* Nat Genet, 1996. **13**(1): p. 83-6.

56. Palmer, S., et al., *A gene spans the pseudoautosomal boundary in mice.* Proc Natl Acad Sci U S A, 1997. **94**(22): p. 12030-5.

57. Perry, J., et al., *A short pseudoautosomal region in laboratory mice.* Genome Res, 2001. **11**(11): p. 1826-32.

58. Eisenberg, R.A. and F.J. Dixon, *Effect of castration on male-determined acceleration of autoimmune disease in BXSB mice.* J Immunol, 1980. **125**(5): p. 1959-61.

59. Izui, S., et al., *The Y chromosome from autoimmune BXSB/MpJ mice induces a lupus-like syndrome in (NZW x C57BL/6)F1 male mice, but not in C57BL/6 male mice.* Eur J Immunol, 1988. **18**(6): p. 911-5.

60. Rowe, C., *The sexual male, part one: The flight of the spermatozoon*, in *Playboy.* 2007. p. 50-54, 130-132.

61. Subramanian, S., et al., *A Tlr7 translocation accelerates systemic autoimmunity in murine lupus.* Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9970-5.

62. Phillips, S.J., et al., *Male and female mouse DNAs can be discriminated using retroviral probes.* Nature, 1982. **297**(5863): p. 241-3.

63. Bishop, C.E., et al., *Most classical Mus musculus domesticus laboratory mouse strains carry a Mus musculus musculus Y chromosome.* Nature, 1985. **315**(6014): p. 70-2.

64. Tucker, P.K., et al., *Geographic origin of the Y chromosomes in "old" inbred strains of mice.* Mamm Genome, 1992. **3**(5): p. 254-61.

65. Tucker, P.K., B.K. Lee, and E.M. Eicher, *Y chromosome evolution in the subgenus Mus (genus Mus).* Genetics, 1989. **122**(1): p. 169-79.
66. Nishioka, Y., et al., *Molecular evolution of a Y-chromosomal repetitive sequence family in the genus Mus.* Mol Biol Evol, 1994. **11**(1): p. 146-53.
67. Nishioka, Y., *Evolutionary characterization of a Y chromosomal sequence conserved in the genus Mus.* Genet Res, 1988. **52**(2): p. 145-50.
68. Nishioka, Y., B.M. Dolan, and L. Zahed, *Molecular characterization of a mouse Y chromosomal repetitive sequence amplified in distantly related species in the genus Mus.* Genome, 1993. **36**(3): p. 588-93.
69. Eicher, E.M., et al., *A repeated segment on the mouse Y chromosome is composed of retroviral-related, Y-enriched and Y-specific sequences.* Genetics, 1989. **122**(1): p. 181-92.
70. Hutchison, K.W. and E.M. Eicher, *An amplified endogenous retroviral sequence on the murine Y chromosome related to murine leukemia viruses and viruslike 30S sequences.* J Virol, 1989. **63**(9): p. 4043-6.
71. Fennelly, J., et al., *Co-amplification to tail-to-tail copies of MuRVY and IAPE retroviral genomes on the Mus musculus Y chromosome.* Mamm Genome, 1996. **7**(1): p. 31-6.
72. Navin, A., et al., *Mouse Y-specific repeats isolated by whole chromosome representational difference analysis.* Genomics, 1996. **36**(2): p. 349-53.
73. Bergstrom, D.E., et al., *An expanded collection of mouse Y chromosome RDA clones.* Mamm Genome, 1997. **8**(7): p. 510-2.
74. Bergstrom, D.E., et al., *The mouse Y chromosome: enrichment, sizing, and cloning by bivariate flow cytometry.* Genomics, 1998. **48**(3): p. 304-13.
75. Bishop, C.E., *Mapping the mouse Y chromosome.* Annales d'Endocrinologie (Paris), 1993. **54**: p. 331-335.
76. Bishop, C.E. and D. Hatat, *Molecular cloning and sequence analysis of a mouse Y chromosome RNA transcript expressed in the testis.* Nucleic Acids Res, 1987. **15**(7): p. 2959-69.
77. Prado, V.F., et al., *Molecular characterization of a mouse Y chromosomal repetitive sequence that detects transcripts in the testis.* Cytogenet Cell Genet, 1992. **61**(2): p. 87-90.
78. Styrna, J., J. Klag, and K. Moriwaki, *Influence of partial deletion of the Y chromosome on mouse sperm phenotype.* J Reprod Fertil, 1991. **92**(1): p. 187-95.
79. Xian, M., et al., *Effect of a partial deletion of Y chromosome on in vitro fertilizing ability of mouse spermatozoa.* Biol Reprod, 1992. **47**(4): p. 549-53.
80. Burgoyne, P.S., et al., *Fertility in mice requires X-Y pairing and a Y-chromosomal "spermiogenesis" gene mapping to the long arm.* Cell, 1992. **71**(3): p. 391-8.
81. Conway, S.J., et al., *Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse Y chromosome.* Mamm Genome, 1994. **5**(4): p. 203-10.
82. Styrna, J., B. Bilinska, and H. Krzanowskaa, *The effect of a partial Y chromosome deletion in B10.BR-Ydel mice on testis morphology, sperm quality and efficiency of fertilization.* Reprod Fertil Dev, 2002. **14**(1-2): p. 101-8.
83. Grzmil, P., et al., *The influence of the deletion on the long arm of the Y chromosome on sperm motility in mice.* Theriogenology, 2007. **67**(4): p. 760-6.

84. Toure, A., et al., *A new deletion of the mouse Y chromosome long arm associated with the loss of Ssty expression, abnormal sperm development and sterility.* Genetics, 2004. **166**(2): p. 901-12.

85. Toure, A., et al., *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, 2004. **83**(1): p. 140-7.

86. Ellis, P.J., et al., *Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids.* Hum Mol Genet, 2005. **14**(18): p. 2705-15.

87. Rohozinski, J., et al., *Successful targeting of mouse Y chromosome genes using a site-directed insertion vector.* Genesis, 2002. **32**(1): p. 1-7.

88. Simpson, E.M., et al., *Novel Sxr(a) ES cell line offers hope for Y chromosome gene-targeted mice.* Genesis, 2002. **33**(2): p. 62-6.

89. Waterston, R.H., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.

90. Skaletsky, H., et al., *The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes.* Nature, 2003. **423**(6942): p. 825-37.

CHAPTER 2

The Mouse Y: The rapid expansion of a degenerating chromosome

Jessica Alföldi, Helen Skaletsky, Tina Graves, Patrick J. Minx, Tatyana Pyntikova, Steve Rozen, Richard K. Wilson, and David C. Page

Contributions of the authors:

Jessica Alföldi contributed to the design of the sequencing strategy, participated in BAC selection, performed other PCR based experiments, performed sequence analysis including Ks, gene and amplicon analysis, designed figures, and coordinated the sequencing effort.

Helen Skaletsky assembled sequenced BACs into the mouse Y chromosome build, contributed to the design of the sequencing strategy, participated in BAC selection, wrote computer programs in the aid of sequence analysis and gave advice and direction.

Tina Graves, Patrick J. Minx, Richard K. Wilson and many others at the Genome Sequencing Center at the Washington University School of Medicine sequenced and assembled BACs, participated in BAC selection and contributed to the design of the sequencing strategy.

Tatyana Pyntikova participated in BAC selection.

Steve Rozen contributed to the design of the sequencing strategy and gave advice and direction.

David C. Page gave advice and direction.

Abstract

Today, on the mouse Y chromosome, only a 3 Mb region remains of the original ~150 Mb autosome pair from which all marsupial and eutherian X and Y chromosomes are derived. However in sequencing the mouse Y, we have found that a large and remarkable segmental duplication, the Huge Repeat, has expanded to form the remaining 95% of the chromosome. Each Huge Repeat unit is 515 kb in length and is repeated 150-200 times to make up the 90 Mb array. The repeat units are 99% to 99.999% similar, euchromatic, internally repetitive, and have no homology to the human Y chromosome. The rest of the mouse Y chromosome is similar in organization to the human Y chromosome, but has degenerated much further, leaving fewer and more specialized genes. Conversely, the ampliconic regions of the mouse Y have expanded much further than the ampliconic regions of the primate Y chromosomes, or those of any other chromosome sequenced to date.

Introduction

The sex chromosomes of the mouse and almost all other mammals are the X chromosome and the Y chromosome, where females are XX and males are XY. Mammalian X and Y chromosomes are thought to have descended from an ordinary pair of autosomes in the common ancestor of all mammals 300 million years ago (mya)[1]. The initiating step is unclear, but it is possible that the chromosomes first began to diverge when the sex-determining gene *Sry* (sex determining region, Y-linked) differentiated from its homologous gene on the X chromosome *Sox3* (sex determining region, Y, box 3)[2-4]. At that point any descendant carrying a single copy of *Sry* would become male, while those who did not would become female. Over time, we believe that male benefit genes were added to the Y chromosome [5], and portions of the Y inverted. Now, the male beneficial genes in the inversions would always be inherited together with *Sry*, since those portions of the Y could no longer recombine with the X chromosome. This process continued until almost all of the sequence of the mammalian Y chromosomes was purely male-specific, and could not recombine with X chromosomes [6-9]. This part of the Y chromosome is known as the Male-Specific Y, or MSY. The remaining parts of Y chromosomes that do recombine with X chromosomes are called pseudoautosomal regions (PARs).

The human Y chromosome has been the most studied of all the Y chromosomes and has generally been regarded as a degenerated version of the X chromosome, full of repetitive elements and containing few genes [10-12]. Some proposed that this degeneration was destined to continue and eventually lead to the disappearance of the human Y chromosome altogether [13, 14]. Historically, the mouse Y chromosome has been viewed as even more degenerate than the human Y chromosome. The mouse MSY contained fewer known genes than the human MSY, and while the human MSY contains many ubiquitously expressed genes, the vast majority of known mouse MSY genes were more specialized and were expressed exclusively in the testis [15]. Furthermore, only a few Mb of the mouse MSY contained the vast majority of known genes on the chromosome [16]. The remaining 90 Mb of the mouse MSY were thought to be a repetitive, retroviral element-filled wasteland, their only gene content a single multi-copy gene family, *Ssty* [17, 18].

In 2001, the image of an entirely degenerate human Y chromosome was shaken when large ampliconic regions containing the segmental duplications known as palindromes were discovered in its sequence. These palindromes are large, inverted repeats, with a unique region known as a spacer separating the inverted repeat units (or arms). The palindrome arms are 99.9% to 99.99% identical to each other, and most contain testis genes [19]. It has been hypothesized that the ongoing gene conversion between these arms can serve to protect the genes within them from degeneration and disappearance through sequence homogenization [20].

The analysis of the chimpanzee Y chromosome has provided further insight into the nature of Y chromosome degeneration and expansion. Hughes *et al.* found that even though the chimpanzee Y chromosome had lost 4 genes since its split from the human Y, the human Y had not lost a single gene since the split 6 million years ago [21]. Also, the chimpanzee Y's ampliconic regions were much larger and more complex than the human Y's, demonstrating that expansionary forces can coexist on Y chromosomes alongside the more well-known degenerative forces. The sequencing of the chimpanzee MSY also showed that the Y chromosome was by far the fastest evolving chromosome in the hominid genomes [22].

We suspected that the same would be true of the mouse Y chromosome. Many people had sought to find genes and other sequences on the mouse MSY homologous to those on the human MSY. Very little was found – and what was found was confined to a 3 Mb region out of a 95 Mb chromosome [23]. Since genes homologous to human MSY genes could only be found on 3 Mb of mouse Y chromosome sequence, it looked as if the mouse Y chromosome was degenerating much faster than its primate relatives. However, we also had clues that much of the mysterious remaining 92 Mb of mouse MSY sequence was incredibly repetitive, which suggested to us that the Y chromosome expansionary forces were also especially vigorous on the mouse MSY.

Results

The Huge Repeat array

Structurally, the mouse MSY is very unusual. 95% of its 95 Mb is a euchromatic, gene-containing segmental duplication that we have named the Huge Repeat (HR). This Huge Repeat array is not homologous to any sequence on the human Y. The remaining 3 Mb of non-HR sequence is located almost entirely at the end of the mouse Y chromosome opposite to the PAR and contains several smaller segmental duplications. In contrast, the segmental duplications of the human Y chromosome compose only 25% of its euchromatin [8]. (Figure 1)

Structure of the Huge Repeat array

The Huge Repeat unit is 515 kb in length and is internally repetitive. A single Huge Repeat unit is composed of two black amplicons, a blue amplicon and a red amplicon. We estimate that there are 150-200 HR units on the mouse MSY, composing an estimated 92 Mb of the 95 Mb mouse MSY. (Figure 2) Due to the numerous copies of the Huge Repeat unit found on the mouse MSY, and to the internal repetition in the HR unit, we propose that the Huge Repeat array has undergone significant intrachromosomal recombination. Indeed, many mouse MSY sequences appear to be derived from events of homologous recombination between Huge Repeat units – such as a recombination between two black amplicons, leading to an interstitial deletion of a red and a black amplicon. (Figure 3) There are also many sequences that are so highly rearranged such that it is not possible to reconstruct a path from the canonical Huge Repeat unit to their

Figure 1 Schematic representation of the mouse and human Y chromosomes, to scale. The precise location of the mouse Y centromere is unknown. One of the possible locations is shown.

Figure 1

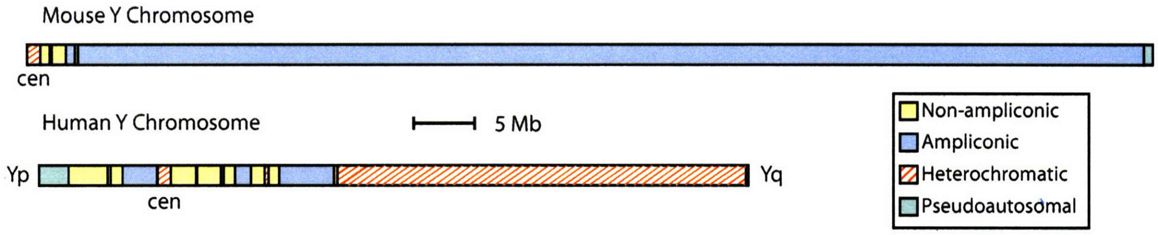

Mouse Y Chromosome
cen
Human Y Chromosome
5 Mb
Yp
Yq
cen
Non-ampliconic
Ampliconic
Heterochromatic
Pseudoautosomal

Figure 2 Diagram of the mouse Y chromosome with expanded views of two selected regions. **a** Enlargement of three adjacent mouse Y contigs located distally from the pseudoautosomal region. Genes denoted by black arrows are to scale; genes denoted by arrowheads are smaller than the scale indicates. Amplicons are represented as coloured triangles. The blue segments of triangles found in the rightmost rectangle signify the presence of Huge Repeat sequence similar to that represented by blue triangles in part c below. **b** The mouse Y chromosome as shown in figure 1. **c** Enlargement of an unanchored mouse Y contig. The red, blue, and black triangles represent the repeat units that comprise the Huge Repeat array. Beige rectangles indicate insertions of non-Huge Repeat sequence. Genes are shown below. Arrows denote genes drawn to scale; arrowheads denote genes that are smaller than the scale indicates.

Figure 2

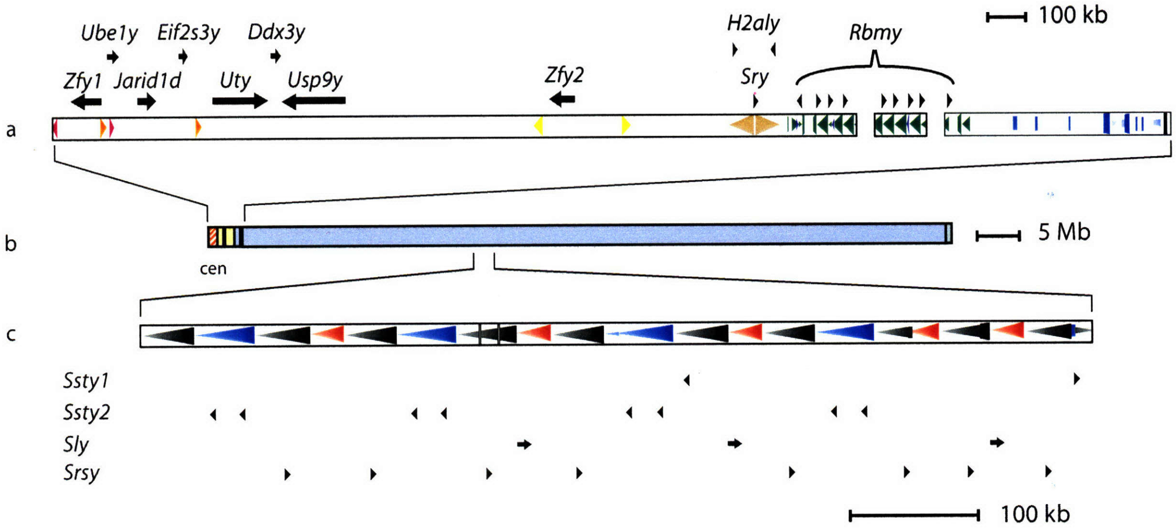

100 kb
Ube1y
Eif2s3y
Ddx3y
H2aly
Rbmy
Zfy1
Jarid1d
Uty
Usp9y
Zfy2
Sry
a
b
5 Mb
cen
c
Ssty1
Ssty2
Sly
Srsy
100 kb

Figure 3 Sequence based map of a sample Huge Repeat contig.

This is a detailed map of the region shown in Figure 2c. All sequence mapped here is ampliconic and therefore is shown with a blue background.

a, Protein-coding genes. Previously reported genes and novel, experimentally verified transcription units for which cDNA sequencing suggests protein-coding potential. Plus (+) strand above, minus (-) strand below.

b, Scale, in Mb.

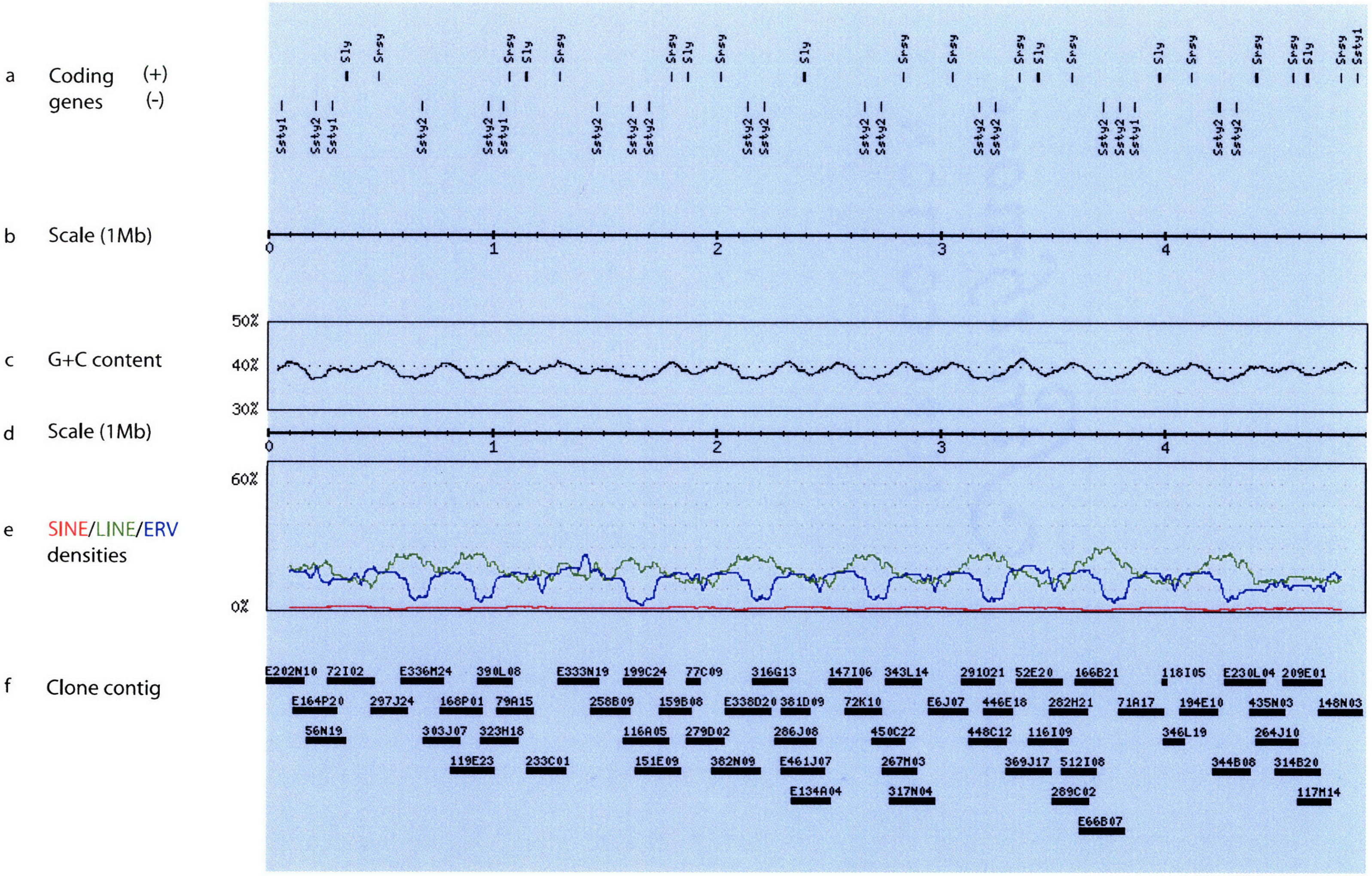
c, G+C content (%) calculated in 100-kb sliding window with 1-kb steps.

d, Scale, in Mb.

e, SINE, LINE, and endogenous retroviral (ERV) repeat content, expressed as percentage of nucleotides, calculated in a 200-kb sliding window with 1-kb steps.

f, 58 BAC clones that have been completely or partially sequenced. Each bar represents the size and position of one BAC clone, labelled by its BAC library identifier. All BACs whose label begins with the letter E are from the CHORI-36 BAC library; all other BAC clones are from the RPCI-24 library.

Figure 3


a Coding genes (+) (-)
Sly
Srsy
Srsy
Sly
Srsy
Srsy
Sly
Srsy
Sly
Srsy
Srsy
Srsy
Sly
Srsy
Sly
Srsy
Srsy
Srsy
Sly
Srsy
Ssty1
Ssty1
Ssty2
Ssty1
Ssty2
Ssty2
Ssty1
Ssty2
Ssty2
Ssty2
Ssty2
Ssty2
Ssty2
Ssty2
Ssty2
Ssty2
Ssty2
Ssty2
Ssty1
Ssty2
Ssty2
b Scale (1Mb)
0
1
2
3
4
c G+C content
50%
40%
30%
d Scale (1Mb)
0
1
2
3
4
e SINE/LINE/ERV densities
60%
0%
f Clone contig
E202N10
72I02
E336M24
390L08
E333N19
199C24
77C09
316G13
147I06
343L14
291021
52E20
166B21
118I05
E230L04
209E01
E164P20
297J24
168P01
79A15
258B09
159B08
E338D20
381D09
72K10
E6J07
446E18
282H21
71A17
194E10
435N03
148N03
56N19
303J07
323H18
116A05
279D02
286J08
450C22
448C12
116I09
346L19
264J10
119E23
233C01
151E09
382N09
E461J07
267M03
369J17
512I08
344B08
314B20
E134A04
317N04
289C02
117M14
E66B07

current structure. However, a plurality of the Huge Repeat sequence (18% of the mouse MSY) is in completely intact repeat units. (Figure 4)

High similarity of repeat units

The sequence identity between any two of the 150-200 HR units ranges from 99% to 99.999%. Therefore, there can be as few as five nucleotides different between two Huge Repeat units. Largely due to this extreme degree of sequence identity between Huge Repeat units, there is a remarkably high level of intrachromosomal sequene similarity on the mouse MSY. 50% of the sequence of the chromosome has a sequence identity of 99.9% or higher to another locus on the mouse MSY. Also, 89% of sequences on the mouse MSY have a sequence identity of 99% or higher to another part of the chromosome. In contrast, only 25% of the human MSY has an intrachromosomal sequence similarity as high as 99%. (Figure 5)

Genes of the Huge Repeat array

We have found that each complete HR unit can contain up to seven gene copies from three different gene families: *Sly* (Sycp3-like, Y-linked), *Srsy* (Serine-rich, secreted, Y-linked) and *Ssty* (Spermiogenesis specific transcript, Y-linked). The *Ssty* gene family can be further sub-divided into the *Ssty1* and *Ssty2* families[24], where *Ssty1* is always found within the black amplicon and *Ssty2* is always found within the blue amplicon. These three testis-specific gene families (see Figure 6) all have multi-copy mouse X homologs, but have no orthologs on the human Y. There is considerable variation between members of each gene family, but since there are often several loci with identical coding sequences, it is impossible to determine whether any specific locus is transcribed. We therefore define a *Sly*, *Srsy*, or *Ssty* locus to be a gene by whether it

Figure 4 Electronic fractionation of mouse MSY sequence by the maximum length a segment can align with a canonical Huge Repeat unit without major insertions, deletions or inversions. This figure shows percentages of mouse MSY sequence is found in complete or partial HR units. The rightmost bar contains all sequence whose longest match to the canonical Huge Repeat unit is an entire HR unit (515 kb). The coloured triangles are diagrams of Huge Repeat sequence (as shown in figure 2) of actual mouse Y contigs that represent Huge Repeat unit segments of 120 kb, 150 kb, and 515 kb respectively.

Figure 4

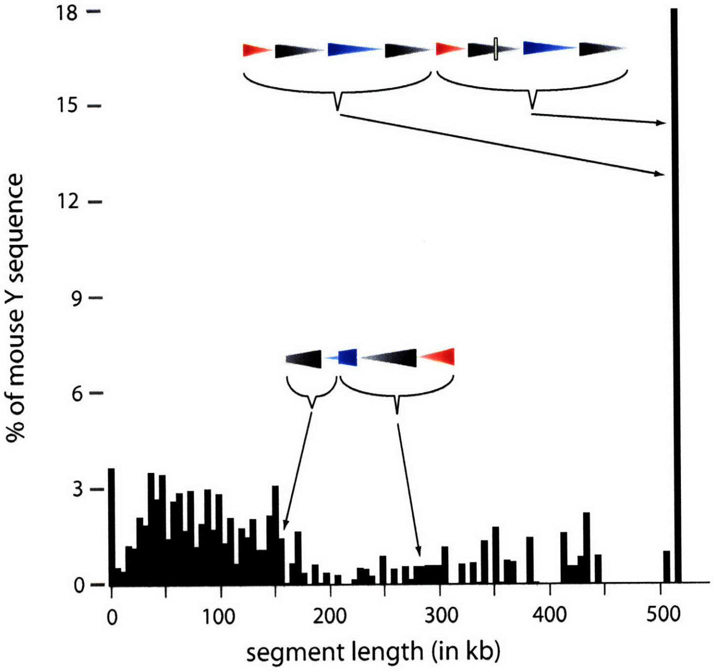

18
15
12
9
6
3
0
% of mouse Y sequence
0
100
200
300
400
500
segment length (in kb)

Figure 5 The mouse Y chromosome exhibits significantly higher levels of intrachromosomal similarity than the human Y chromosome. **a**, Electronic fractionation of the mouse MSY by percent identity to other mouse MSY sequences, plotted on a logarithmic scale. Values <70% are not shown. **b**, Electronic fractionation of the human MSY by percent identity to other human MSY sequences, plotted on a logarithmic scale. Values <70% are not shown. Adapted from Figure 9 of Skaletsky et al.

Figure 5

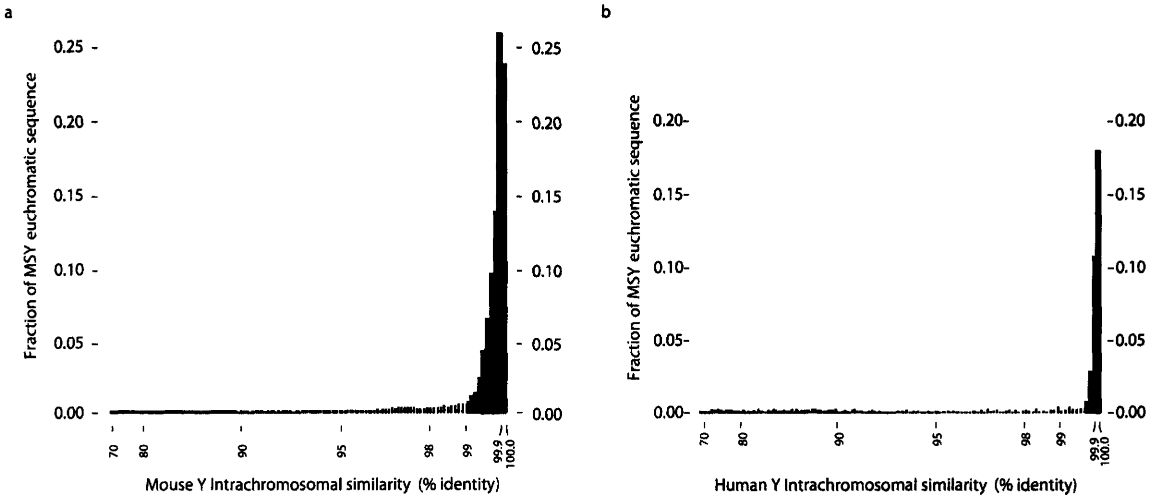
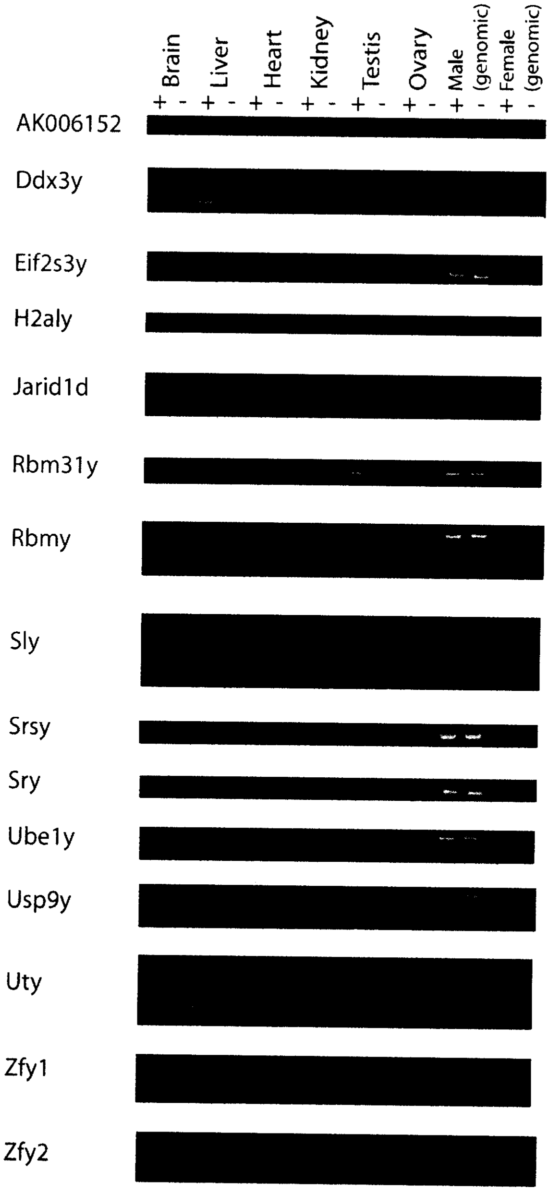

a
Fraction of MSY euchromatic sequence
0.25
0.20
0.15
0.10
0.05
0.00
70
80
90
95
98
99
99.9
100.0
Mouse Y Intrachromosomal similarity (% identity)
b
Fraction of MSY euchromatic sequence
0.20
0.15
0.10
0.05
0.00
70
80
90
95
98
99
99.9
100.0
Human Y Intrachromosomal similarity (% identity)

Figure 6

RT-PCR analysis of all known mouse MSY genes. Intron-spanning primers were used for all intron-containing genes. Two types of negative controls were used in this study. For intron-containing genes, genomic DNA controls were used (male genomic). Also, RT-controls were used for all genes (-). Primer sequences are listed in Supplementary File 1 at http://jura.wi.mit.edu/page/Jessica_Alfoldi/index.html.

Figure 6


+ - Brain
+ - Liver
+ - Heart
+ - Kidney
+ - Testis
+ - Ovary
+ - Male (genomic)
+ - Female (genomic)
AK006152
Ddx3y
Eif2s3y
H2aly
Jarid1d
Rbm31y
Rbmy
Sly
Srsy
Sry
Ube1y
Usp9y
Uty
Zfy1
Zfy2

contains an intact Open Reading Frame (ORF). By this standard, we estimate that there are 113 copies of *Sly*, 163 copies of *Srsy*, and 311 copies of *Ssty* on the mouse Y chromosome. (Figure 2) All Huge Repeat gene intact ORF sequences can be found as Supplementary Files 2, 3, and 4 at http://jura.wi.mit.edu/page/Jessica_Alfoldi/index.html.

A Huge Repeat palindrome

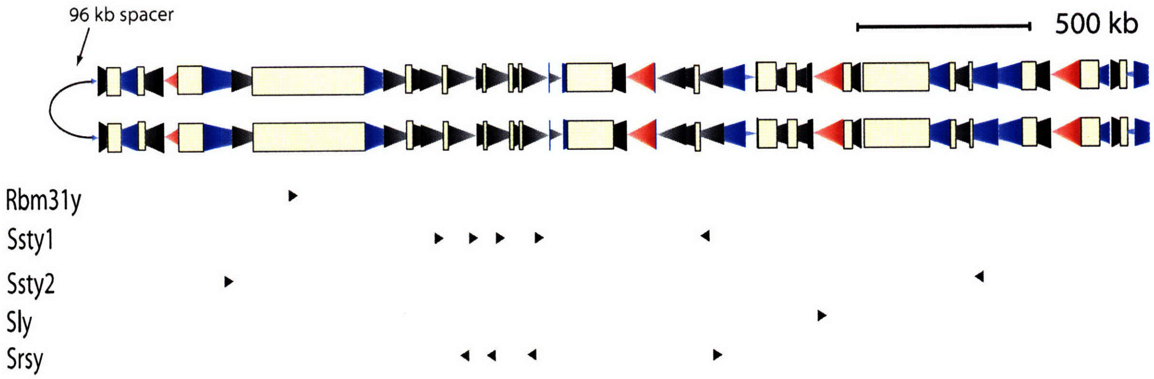
One Huge Repeat rearrangement stands out among the rest – a remarkable palindrome composed of rearranged HR sequence. Each arm of this Huge Repeat palindrome is 3.5 Mb in length. The unique spacer separating the two homologous arms is only 96 kb, yielding a total length of 7 Mb for the entire palindrome. Each arm not only contains a wide variety of rearranged Huge Repeat units and many Huge Repeat genes, but also a 200 kb non-Huge Repeat sequence containing a novel testis gene, *Rbm31y* (RNA binding motif, 31, Y-linked). The two arms of this HR palindrome exhibit 99.999% sequence identity, the highest we have found anywhere on the chromosome. (Figure 7) In contrast, the largest palindrome on the human Y chromosome, the P1 palindrome, has a total length of only 2.9 Mb and an arm-to-arm identity of 99.97%. We currently have the sequence of this Huge Repeat palindrome in eight contigs, but we are confident from the sequence we do have, and the high degree of similarity of the contigs, that they do form one large palindrome.

The rest of the mouse Y chromosome

Whereas 90 Mb of the mouse MSY is composed of Huge Repeat sequence, the remaining 3 Mb is not. This sequence can be divided into two parts. The first part consists of the many short (<200 kb) unique sequences and interspersed repeats scattered

Figure 7 Diagram of a 7 Mb mouse Y palindrome largely composed of Huge Repeat sequence. The coloured triangles represent Huge Repeat seqence, and the beige rectangles represent non-Huge Repeat sequence as in figure 2. The non-Huge Repeat unique spacer between the two near-identical arms is indicated by a curved line. The gene *Rbm31y* is smaller than the scale indicates.

Figure 7



among the Huge Repeat array; and the second is 3 Mb of sequence at the opposite end of the chromosome from the PAR. Before we started sequencing the mouse MSY, seven genes and two gene families had been mapped to these distal 3 Mb, and several BACs (Bacteria Artificial Chromosomes) covering 750 kb of this region had been sequenced. [16]

In sequencing this 3 Mb of sequence, we found that it is very similar in structure to a condensed human MSY. It contains several low-copy segmental duplications, often arranged in palindromes. It also contains seven genes and three small gene families, some of which are located in the arms or spacers of palindromes and many of which are testis-specific (Figure 6). All of those genes have homologs on the mouse and human X chromosomes. Seven of them also have homologs on the human Y chromosome. This includes the sex-determining gene *Sry* (sex determining region, Y-linked), which straddles the spacer and one arm of a 150 kb palindrome. Another gene, H2aly (H2A-like, Y-linked) is found in each of the arms of that palindrome. There is one highly repetitive region (denoted as green triangles in Figure 2), where each 40 kb repeat unit contains a copy of the gene *Rbmy* (RNA-binding motif, Y-linked). (Figure 2) The size of these repeat units is not amenable to our sequencing method, and so the structure of this region, and the exact number of repeat units found there, have not been determined. However, we estimate from our sequence and other data that the region is 1 Mb in length, comprising 25 repeat units containing up to 25 Rbmy intact ORFs. The gene density of the region near *Zfy1* (Zinc finger protein 1, Y-linked) is roughly equal to the genome average (one gene/ 50 kb), but the region near *Zfy2* (Zinc finger protein 1, Y-linked) is very low[25]. (Figure 8).

Figure 8 Sequence based map of the non Huge Repeat portion of the mouse Y chromosome located distally from the pseudoautosomal region.

This is a detailed map of the region shown in Figure 2a. Shown via background colouring are the positions of three classes of euchromatic MSY sequences: X-degenerate (yellow), ampliconic (blue), and other (white). There are two gaps in the sequence which can be seen in part g.

a, Major features. All amplicons with a repeat unit size of at least 10 kb. Arrows indicate amplicons that are drawn to scale; arrowheads indicate amplicons that are not drawn to scale due to their small size.

b, Protein-coding genes. Previously reported genes and novel, experimentally verified transcription units for which cDNA sequencing suggests protein-coding potential. Plus (+) strand above, minus (-) strand below.

c, Scale, in Mb.

d, Pseudogenes. Previously reported and novel sequences with homology to known genes but no protein-coding capacity.

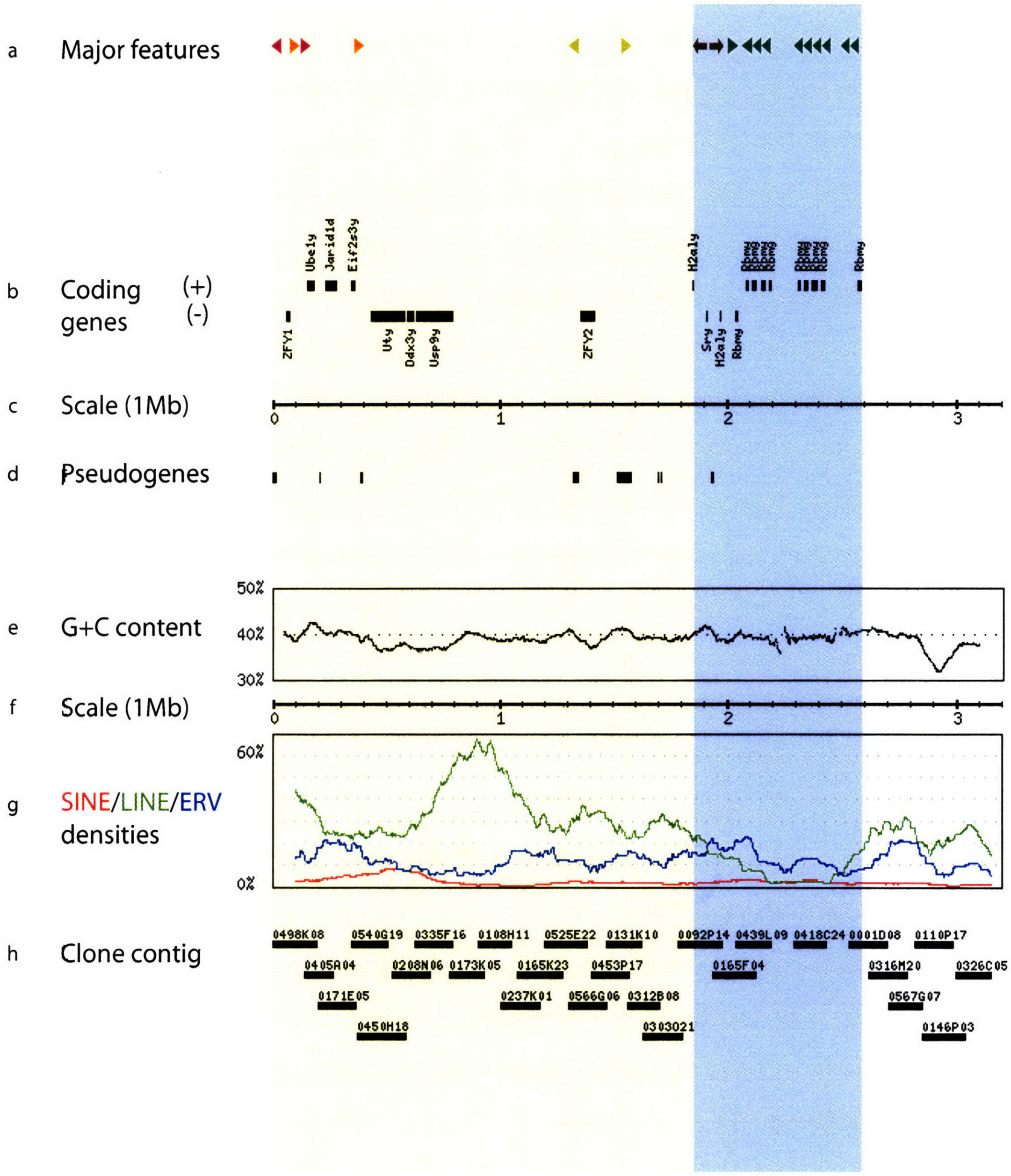
e, G+C content (%) calculated in 100-kb sliding window with 1-kb steps.

f, Scale, in Mb.

g, SINE, LINE, and endogenous retroviral (ERV) repeat content, expressed as percentage of nucleotides, calculated in a 200-kb sliding window with 1-kb steps.

h, 27 BAC clones that have been completely or partially sequenced. Each bar represents the size and position of one BAC clone, labelled by its BAC library identifier. All BACs shown come from the RPCI-24 BAC library.

Figure 8


a Major features
b Coding genes (+) (-)
c Scale (1Mb)
d Pseudogenes
e G+C content
f Scale (1Mb)
g SINE/LINE/ERV densities
h Clone contig
Ube1y
Jarid1d
Eif2s3y
H2a1y
Rbmy
Rbmy
Rbmy
Rbmy
Rbmy
Rbmy
Rbmy
Rbmy
Rbmy
ZFY1
Uty
Ddx3y
Usp9y
ZFY2
Sry
H2a1y
Rbmy
0
1
2
3
50%
40%
30%
60%
0%
0498K08
0540G19
0335F16
0108H11
0525E22
0131K10
0092P14
0439L09
0418C24
0001D08
0110P17
0405A04
0208N06
0173K05
0165K23
0453P17
0165F04
0316M20
0326C05
0171E05
0237K01
0566G06
0312B08
0567G07
0450H18
0303021
0146P03

The mouse Y chromosome's centromere

Unusually, the location of the mouse Y centromere is unknown. All other mouse chromosomes are known to be telocentric, but historically, the mouse Y chromosome was declared to be acrocentric. The 3 Mb of non-Huge Repeat sequence opposite to the mouse Y PAR was assumed to be on a mouse Y short arm [26]. There is no visible short arm in karyotypes of the mouse genome, but a small (< 5 Mb) short arm containing that sequence could still be present. Alternatively, the mouse Y chromosome could have only a single arm, and in such a case that 3 Mb of sequence would be adjacent to the centromere (as shown in Figure 1). The existence of a short arm would also make certain translocations of parts of the mouse Y chromosome easier to explain [27]. The difficulty arises from the lack of any DNA probes for the mouse Y centromere, as the satellite sequences prevalent in all the other mouse centromeres are absent from the Y [28]. We do know the approximate location of the mouse Y centromere, as others have used Giemsa staining to show that the centromere is at the opposite end of the chromosome from the pseudoautosomal region [29]. We can not narrow down the location of the centromere any further. It could be located at the mouse Y telomere opposite from the PAR, leaving the mouse Y chromosome with just one arm. It could be located in any of the gaps in our sequence build of that 3 Mb region, as shown in Figure 2a. It could even be anywhere near that end of the chromosome, in the Huge Repeat array or not, as we cannot identify the centromere by its sequence.

Sequencing the mouse MSY

Traditional sequencing methods

Whole genome shotgun is the standard method that has been used in the sequencing of all the genome projects of the past ten years. In this method, an entire genome is cut into small pieces (~1 kb) and cloned into plasmids. Each of those plasmid inserts is then sequenced individually, and computer programs are used to stitch all the small sequences into whole chromosomal sequences. Whole genome shotgun is the most cost- and time-efficient method of sequencing genomes, but since it fails in assembling together repetitive sequences it would be completely useless in the sequencing of the mouse MSY. The problem lies in the high similarity between Huge Repeat units, and between other segmental duplications. With only 1 kb sequences to work with, it would be impossible to determine which sequence belongs to which repeat unit, and the small sequences could not be assembled into a whole chromosome. Therefore, we decided to sequence the mouse MSY using a BAC by BAC approach.

Using BACs to sequence segmental duplications

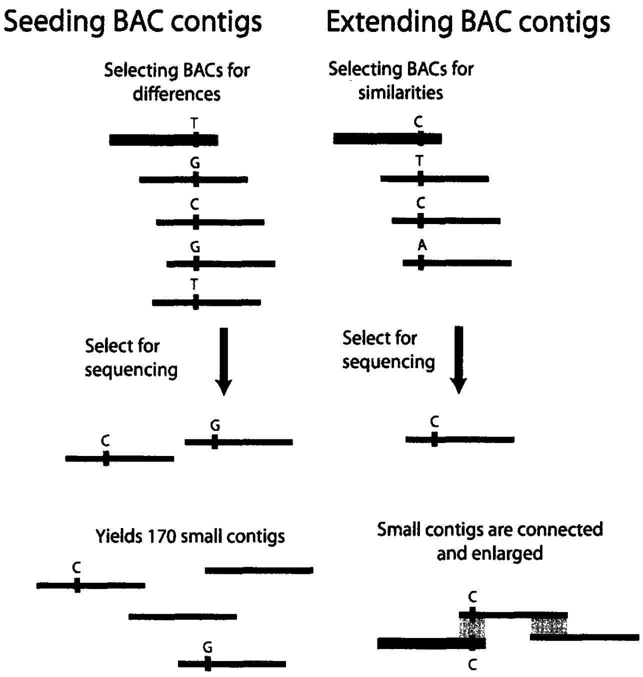
Before the advent of whole genome shotgun, the earliest sequenced genomes were sequenced using large insert clones. In this method, the genome is cut into many pieces (~150 kb) and inserted into BACs or cosmids. The insert in each clone is then cut into small pieces (~3 kb) and cloned into plasmids. The sequence of each clone is assembled individually before all the BAC or cosmid clones are assembled together into a chromosome. Repetitive sequence can be assembled using this method as long as the repeat units are larger than the size of the clone. This is why a variation on the traditional BAC by BAC approach was used to sequence the human [8, 19] and chimpanzee [21, 22] MSYs, as well as the 3 Mb of non-Huge Repeat mouse MSY sequence shown in Figure 2. This region and these chromosomes contained both single-copy sequence and segmental duplications consisting of 2-13 repeat units. To sequence these regions, BAC clones were chosen to be sequenced using known markers, and then individual copies of segmental duplications were teased apart by the presence of sequence family variants (SFVs) – individual nucleotide differences between repeat units. The Huge Repeat array was sequenced BAC by BAC, but it required very different BAC clone selection techniques, as there were no known markers along its 92 Mb of sequence, and there were hundreds of copies of repeat units to tease apart instead of just a few.

The sequencing of the Huge Repeat array

And so we first determined which BAC clones were from the MSY by using BAC fingerprints and sequenced BAC ends to eliminate any clones that we could assign to the already sequenced female mouse genome. We then selected Huge Repeat BAC clones for sequencing in two major phases. (Figure 9) In the first phase, we created our own markers along the mouse MSY by taking advantage of the individual base pair substitutions between Huge Repeat units that we call Sequence Family Variants (SFVs). We sequenced an equivalent part of the Huge Repeat unit in every HR BAC clone and looked for unique SFV patterns. We selected those clones that exhibited a unique pattern of SFVs. This resulted in 170 small sequence contigs made up of the sequence of a single BAC clone. In the second phase, our goal was to expand each of these 170 small contigs by finding partially overlapping BAC clones. To accomplish this, we looked for identical SFV patterns between our already sequenced BACs and all of its potential neighbours. We then selected any perfectly matching BAC clones for sequencing. This process of picking neighbours for already sequenced clones was repeated in several cycles to expand contigs until they joined together.

Figure 9 Process of selecting mouse Y Bacterial Artificial Chromosomes (BACs) for sequencing. Horizontal bars represent BAC clones; thick horizontal bars represent already sequenced BAC clones. Short vertical lines within bars denote single nucleotide differences between highly similar sequences [sequence family variants (SFVs)]. Gray shading indicates true overlaps between two BAC clones.

Figure 9


Seeding BAC contigs
Extending BAC contigs
Selecting BACs for differences
T
G
C
G
T
Selecting BACs for similarities
C
T
C
A
Select for sequencing
C
G
Select for sequencing
C
Yields 170 small contigs
C
G
Small contigs are connected and enlarged
C
C

Our mouse MSY sequence

In the course of sequencing the mouse MSY, we acquired 72 Mb of sequence, constituting roughly three quarters of the chromosome. This sequence has been assembled into 79 contigs, with a median length of 540 kb and a maximum of 4.9 Mb. The large majority of these contigs are unanchored, with the exception of the non-HR contigs shown in Figure 2 and the pseudoautosomal boundary.

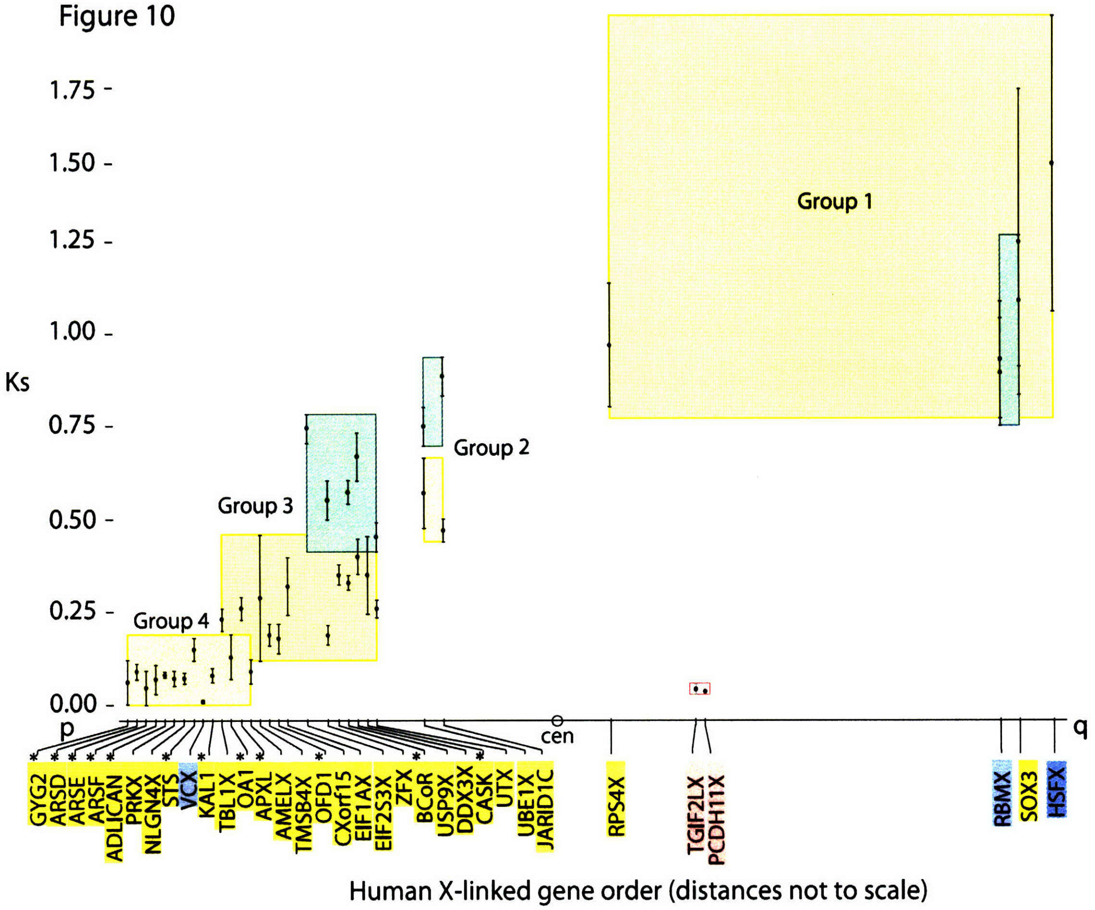
Mouse sex chromosome evolution

Ks value (a measure of synonomous nucleotide divergence) of X-Y gene pairs has often been used to estimate the length of time since the specialization of each gene pair [7, 30]. Lahn and Page suggested that the differentiation of the X and Y paralogs could only have occurred once they were no longer recombining with each other and therefore, that the Ks values are linearly dependent on time since gene pair divergence. They estimated the differentiation ages of human X-Y pairs from their Ks values and then grouped those genes into four groups or strata. They also found a correlation between those Ks values and the position of the X gene on the human X chromosome.

We also grouped mouse X-Y gene pairs into strata and we found that the older mouse strata coincide with the older human strata (groups 1-3). As mouse X chromosomes are highly rearranged with respect to most other known X chromosomes, we did not expect a similar correlation between mouse Ks value and mouse X gene position [31]. However, we did find a correlation between the mouse Ks values and the human X chromosome gene order. (Figure 10) We placed the mouse genes into groups using both the mouse and the human Ks values and the chromosomal location of the

Figure 10 Ks of both mouse and human X-Y gene pairs is correlated with Human X-linked gene order. Plot of K_s versus Human X-linked gene order for 32 X-Y gene pairs. Colour highlighting of X-linked gene names indicates whether Y homologues are X-degenerate (yellow), ampliconic (blue), or X-transposed (pink). Human genes are shown as black circles; mouse genes are shown as green circles. Within the plot, four yellow rectangles denote four previously defined evolutionary strata of human genes, three green rectangles denote three equivalent evolutionary strata of mouse genes; a small pink rectangle highlights two X-transposed genes. Standard errors for K_s values are shown. This figure is adapted from Figure 7 of Skaletsky et al.

Figure 10


1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Ks
Group 1
Group 2
Group 3
Group 4
p
cen
q
*GYG2
*ARSD
*ARSE
*ARSF
*ADLICAN
PRKX
NLGN4X
*STS
VCX
*KAL1
TBL1X
*OA1
*APXL
AMELX
TMSB4X
*OFD1
CXorf15
EIF1AX
EIF2S3X
ZFX
*BCoR
USP9X
DDX3X
*CASK
UTX
UBE1X
JARID1C
RPS4X
TGIF2LX
PCDH11X
RBMX
SOX3
HSFX
Human X-linked gene order (distances not to scale)

genes' homologues in *Monodelphis domestica*, a marsupial. *Monodelphis* diverged from the common human and mouse lineages over 100 mya, and at the time of that divergence, we know that the group 1 and 2 genes had become part of the MSY, but that all of the later group genes would have been in a PAR at that time. And so the *Monodelphis* X chromosome should contain the genes from human and mouse groups 1 and 2, but should not contain any genes from the later groups. We can then predict that genes in group 2 should be found on the X chromosome of *Monodelphis,* but that group 3 genes would be found on *Monodelphis* autosomes. For example, as the *Monodelphis* Eif2s3x is located on *Monodelphis* chromosome 4 and not on the X chromosome, we could assign the *Eif2s3x/Eif2s3y* gene pair to group 3.

Mouse Ks values are significantly higher than human Ks values for groups 2 and 3, as the rodent lineage is known to have a higher substitution rate[25]. In contrast, the Ks values for group 1 of both species are similar, since Ks is only a reliable measure of time since gene pair separation up to an approximate value of 1. (Table 2)

Discussion

The mouse MSY can be divided into two parts. The first, a 3 Mb segment of sequence located opposite to the pseudoautosomal region, is similar in structure to a miniaturized version of the human MSY. It contains many genes with homologues on the mouse X chromosome, as well as several low-copy segmental duplications that contain testis genes. Since this region contains much less sequence and fewer genes than its counterparts, the entire human and chimpanzee chromosomes, we assume that the

Table 2 - Analysis of sequence divergence in X-Y gene pairs

Gene pair	Ks (+/-s.d.)	Ka (+/-s.d.)	Ks/Ka	DNA divergence (coding region, %)	Protein divergence (%)
Rbm31y/Rbm31x	0.11 (+/-0.02)	0.075 (+/-0.012)	1	7.30	13.43
Uty/Utx	0.46 (+/-0.04)	0.12 (+/-0.01)	4	14.55	16.47
Zfy1/Zfx	0.56 (+/-0.06)	0.17 (+/-0.02)	3	20.24	29.18
Zfy2/Zfx	0.56 (+/-0.06)	0.16 (+/-0.02)	3	20.16	28.05
Usp9y/Usp9x	0.58 (+/-0.03)	0.090 (+/-0.006)	6	16.32	15.81
Ddx3y/Ddx3x	0.68 (+/-0.07)	0.049 (+/-0.008)	14	15.01	8.58
Ube1y/Ube1x	0.75 (+/-0.06)	0.10 (+/-0.01)	7	19.47	16.32
Eif2s3y/Eif2s3x	0.76 (+/-0.09)	0.011 (+/-0.004)	71	14.02	0.02
Jarid1d/Jarid1c	0.90 (+/-0.06)	0.13 (+/-0.01)	7	22.63	20.97
Rbmy/Rbmx	0.91 (+/-0.15)	0.39 (+/-0.04)	2	31.61	19.70
Sry/Sox3	1.11 (+/-0.18)	0.87 (+/-0.09)	1	47.88	70.29

Gene pair	Sequence compared (bp)	Genbank accession numbers		Stratum	Monodelphis chromosomal location	Male-specific age (millions of years)
Rbm31y/Rbm31x	1698	AK017055	Rbm31x?	4*	none	20
Uty/Utx	4281	NM_009484	NM_009483	3	4	100
Zfy1/Zfx	2406	AK076618	NM_001044386	3	4	100
Zfy2/Zfx	2406	AK030048	NM_001044386	3	4	100
Usp9y/Usp9x	7686	NM_148943	NM_009481	3	4	100
Ddx3y/Ddx3x	1992	NM_012008	NM_010028	3	4	100
Ube1y/Ube1x	3180	NM_011667	NM_009457	2	X	150
Eif2s3y/Eif2s3x	1419	NM_012011	NM_012010	3	4	100
Jarid1d/Jarid1c	4707	NM_011419	NM_013668	2	X	150
Rbmy/Rbmx	1218	NM_011253	NM_011252	1	X	280
Sry/Sox3	1203	NM_011564	NM_009237	1	X	280

*ORF sequences used for this analysis can be found in Supplementary File 5 at http://jura.wi.mit.edu/page/Jessica_Alfoldi/index.html.

degenerative forces acting on Y chromosomes have proceeded much further on the mouse Y than on the primate Y chromosomes.

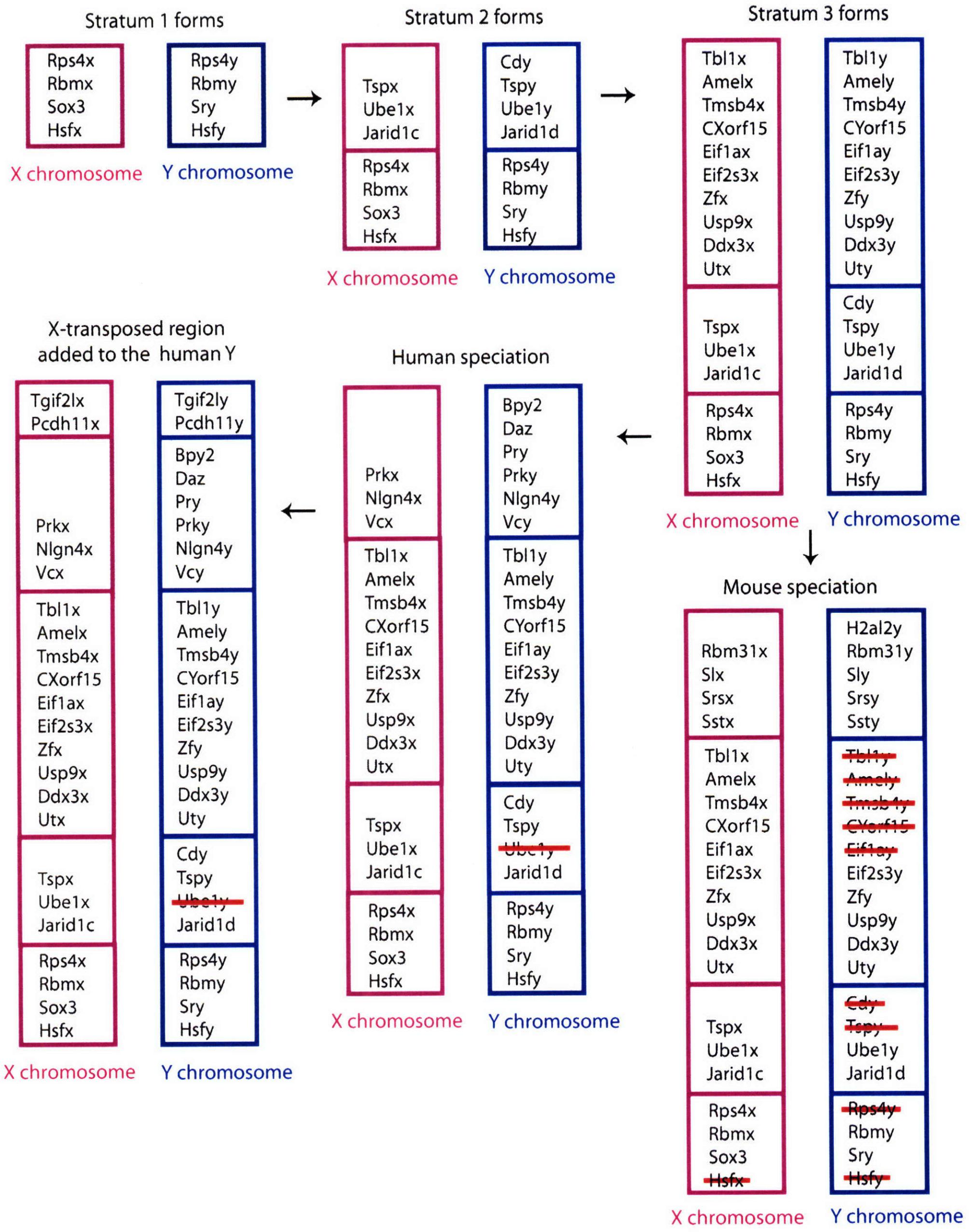
The second part of the mouse MSY is the Huge Repeat array, a giant gene-containing segmental duplication covering the other 92 Mb of the chromosome. It is the largest segmental duplication found to date, and has no homology to the human Y chromosome. We would like to discuss both the common and divergent evolutions of the mouse and human Y chromosomes, and the potential function of the most remarkable feature of the mouse Y chromosome, the Huge Repeat array.

Evolution of the mouse and human sex chromosomes

Using the sequence of the mouse and human X and Y chromosomes, along with other genomic data, we have reconstructed an outline of mouse and human sex chromosome evolution (Figure 11). In the common ancestor of mice and humans, 300 million years ago (mya), the ancestors of the X and Y chromosomes were an ordinary pair of autosomes, with alleles of the same genes on each chromosome. The first differentiation of the X and the Y occurred 280 mya, when *Sry* diverged from its X homologue, *Sox3*. Since the presence of Sry makes an organism male, the Y chromosome was now sex-determining. Around the same time, an inversion occurred on the Y chromosome, leading to the cessation of recombination in that region between the two chromosomes. This region contained Sry as well as the other group 1 genes. The rest of the chromosome still recombined with the X chromosome and was a large pseudoautosomal region. The next step occurred 150 mya, when another part of the PAR stopped recombining with the X chromosome and added on to the MSY. This new MSY

Figure 11 Schematic of the evolution of the human and mouse sex chromosomes through the acquistion and loss of MSY genes and their X homologues. Arrows indicate the flow of time and the split of the human and mouse lineages.

Figure 11


Stratum 1 forms
Rps4x
Rbmx
Sox3
Hsfx
Rps4y
Rbmy
Sry
Hsfy
X chromosome
Y chromosome
Stratum 2 forms
Tspx
Ube1x
Jarid1c
Rps4x
Rbmx
Sox3
Hsfx
Cdy
Tspy
Ube1y
Jarid1d
Rps4y
Rbmy
Sry
Hsfy
X chromosome
Y chromosome
Stratum 3 forms
Tbl1x
Amelx
Tmsb4x
CXorf15
Eif1ax
Eif2s3x
Zfx
Usp9x
Ddx3x
Utx
Tspx
Ube1x
Jarid1c
Rps4x
Rbmx
Sox3
Hsfx
Tbl1y
Amely
Tmsb4y
CYorf15
Eif1ay
Eif2s3y
Zfy
Usp9y
Ddx3y
Uty
Cdy
Tspy
Ube1y
Jarid1d
Rps4y
Rbmy
Sry
Hsfy
X chromosome
Y chromosome
Human speciation
Prkx
Nlgn4x
Vcx
Tbl1x
Amelx
Tmsb4x
CXorf15
Eif1ax
Eif2s3x
Zfx
Usp9x
Ddx3x
Utx
Tspx
Ube1x
Jarid1c
Rps4x
Rbmx
Sox3
Hsfx
Bpy2
Daz
Pry
Prky
Nlgn4y
Vcy
Tbl1y
Amely
Tmsb4y
CYorf15
Eif1ay
Eif2s3y
Zfy
Usp9y
Ddx3y
Uty
Cdy
Tspy
~~Ube1y~~
Jarid1d
Rps4y
Rbmy
Sry
Hsfy
X chromosome
Y chromosome
X-transposed region
added to the human Y
Tgif2lx
Pcdh11x
Prkx
Nlgn4x
Vcx
Tbl1x
Amelx
Tmsb4x
CXorf15
Eif1ax
Eif2s3x
Zfx
Usp9x
Ddx3x
Utx
Tspx
Ube1x
Jarid1c
Rps4x
Rbmx
Sox3
Hsfx
Tgif2ly
Pcdh11y
Bpy2
Daz
Pry
Prky
Nlgn4y
Vcy
Tbl1y
Amely
Tmsb4y
CYorf15
Eif1ay
Eif2s3y
Zfy
Usp9y
Ddx3y
Uty
Cdy
Tspy
~~Ube1y~~
Jarid1d
Rps4y
Rbmy
Sry
Hsfy
X chromosome
Y chromosome
Mouse speciation
Rbm31x
Slx
Srsx
Sstx
Tbl1x
Amelx
Tmsb4x
CXorf15
Eif1ax
Eif2s3x
Zfx
Usp9x
Ddx3x
Utx
Tspx
Ube1x
Jarid1c
Rps4x
Rbmx
Sox3
~~Hsfx~~
H2al2y
Rbm31y
Sly
Srsy
Ssty
~~Tbl1y~~
~~Amely~~
~~Tmsb4y~~
~~CYorf15~~
~~Eif1ay~~
Eif2s3y
Zfy
Usp9y
Ddx3y
Uty
~~Cdy~~
~~Tspy~~
Ube1y
Jarid1d
~~Rps4y~~
Rbmy
Sry
~~Hsfy~~
X chromosome
Y chromosome

region contained the four group 2 paired genes. Also around this time, the gene *Cdy* (Chromodomain protein, Y-linked) was transposed onto the MSY. The third sex chromosome differentiation event occurred 100 mya, adding the nine group 3 genes to the MSY.

The human and mouse lineages diverged 75 mya [25], placing their sex chromosomes on different paths. In the mouse, four more genes (mouse group 4) were transferred from the PAR to the MSY in the past 75 million years. Also, nine genes were lost from the mouse Y chromosome, including Hsfy (Heat shock transcription factor, Y-linked); it is probable that Hsfx (Heat shock transcription factor, X-linked), its X homolog was lost from the X chromosome during this time as we can not find any homologous sequence in the current build (Build 37.1) of the mouse X chromosome.

We can distinguish two distinct phases of human sex chromosome evolution since the separation of the human and mouse lineages. For the first step, many more genes (human group 4) were added to the MSY when recombination between the X and Y ceased for another segment of the Y chromosome. In the meantime, all but three of those Y genes degenerated into pseudogenes. Also, two older genes were lost from the MSY and three genes were added by transposition. The second step occurred 3-4 mya, when a segment of the human X chromosome was transposed over to the human Y, creating the human Y's X-transposed region. The X-transposed region contains two genes and does not recombine with the X chromosome [8].

Function of the Huge Repeat array

What forces led to the expansion and maintenance of the 92 Mb Huge Repeat array on the mouse Y chromosome? Our only data addressing this question come from the study of several mouse lines containing different partial deletions of the Y chromosome. We know now that these deletions ablate anywhere from 2/3 to 9/10 of the Huge Repeat array. In these mouse lines, increasing sperm head deformities are correlated with decreasing copies of the Huge Repeat [32-35]. Even though fertility is a powerful tool for selection, we feel that it is unlikely that the sperm head development function of one or more Huge Repeat genes is the driving force behind this array. It is highly unusual for any species to have hundreds of copies of any gene, the one major exception being the hundreds of copies of ribosomal DNA in many species. We find it unlikely that mice require as high a level of transcription of sperm head development genes as they do of ribosome genes, especially since these genes are not present, and therefore not required at all, in most other mammals.

Our favoured hypothesis for the force behind the expansion and continued maintenance of the Huge Repeat array is meiotic drive, or segregation distortion. Segregation distortion is the phenomenon where one allele is preferentially transmitted into offspring. In this case, we propose, as others have [36], that the mouse X and Y chromosomes are competing with each other to be preferentially transmitted to the next generation. In support of this, others have found that mice with partial deletions of what we know now to be the Huge Repeat array produce more female offspring [37]. Also, microarrays have shown that mice with fewer Huge Repeat copies show upregulation of several X chromosome genes [36]. This implies that one of the functions of the Huge Repeat array is to repress certain X chromosome genes, and to encourage the birth of more Y-bearing offspring. It is plausible that such competition between the mouse X and Y chromosomes could drive the expansion of the elements used in that competition, such as the Huge Repeat array. Certainly, this shows that there can be strong expansionary forces on Y chromosomes, along with the better-known degenerative forces.

Methods

BAC selection and sequencing

All BAC clones selected for sequencing are from the male C57BL/6 *Mus musculus* libraries RPCI-24 and CHORI-36. BACs were chosen for sequencing in several different ways.

1) We used a fingerprint contig database composed of fingerprints of the BAC clones in the male RPCI-24 and the female RPCI-23 BAC libraries. We decided that any fingerprint contigs containing more than 5 BAC clones and composed entirely of RPCI-24 BACs were likely to be from the MSY. PCR STSs were created from these presumed male BACs and tested in both male and female genomic DNA to confirm their male-specificity and contiguity. We could then choose tiling paths for any single- or low-copy fingerprint contigs.

2) 100 BAC clones were selected at random for sequencing from a single massive fingerprint contig, that we later knew to contain the Huge Repeat array

3) All Huge Repeat BAC clones were identified from the RPCI-24 and CHORI-36 ligraries through filter hybridizations using a variety of Huge Repeat overgo probes.

i) Huge Repeat BAC clones were chosen for sequencing for their unique SFV patterns, as described in Results.

ii) Huge Repeat BAC clones were chosen for sequencing for having identical SFV patterns to a target BAC, as described in Results.

Interspersed repeats

We used RepeatMasker (http://repeatmasker.genome.washington.edu) to identify interspersed repeats electronically.

Gene prediction

Potential genes were first identified using GenScan[38]. These potential genes were then selected for confirmation by RT-PCR if they had a significant BLAST match to a mouse EST.

RT-PCR analysis

RNA was extracted from C57BL/6 adult mouse tissues (all from male mice, except for ovary) using the TRIzol reagent (Invitrogen). cDNAs were produced from those RNAs using SuperScript II Reverse Transcriptase (Invitrogen), and contaminating RNA was removed using TURBO DNA-free (Ambion). All PCR primers used and product sizes are listed in Supplementary Table 2

Ks and Ka calculations

We used ClustalX [39] to align X and Y Open Reading Frames (Supplementary Table 3) and then manually adjusted the insertions/deletions. We calculated Ks and Ka using the diverge function of the Wisconsin Package (Version 10.3, Genetics Computer Group)

Electronic fractionation by length of Huge Repeat unit

Dotplots were created between every mouse Y contig and the canonical Huge Repeat unit using a custom Perl script available on request. Using these dotplots, we determined the longest contiguous match to the canonical Huge Repeat unit for every 5 kb interval. The sequence of the canonical Huge Repeat unit can be found in Supplementary File 6 at http://jura.wi.mit.edu/page/Jessica_Alfoldi/index.html.

Electronic fractionation by intrachromosomal similarity

We used BLAST (http://blast.wustl.edu) to find, for each 5 kb segment, the highest similarity to the rest of the mouse Y sequence.

1. Ohno, S., *Sex Chromosomes and Sex-linked Genes*. 1967, Berlin: Springer.
2. Stevanovic, M., et al., *SOX3 is an X-linked gene related to SRY.* Hum Mol Genet, 1993. **2**(12): p. 2013-8.
3. Foster, J.W. and J.A. Graves, *An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene.* Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1927-31.
4. Lahn, B.T., N.M. Pearson, and K. Jegalian, *The human Y chromosome, in the light of evolution.* Nat Rev Genet, 2001. **2**(3): p. 207-16.
5. Lahn, B.T. and D.C. Page, *Functional coherence of the human Y chromosome.* Science, 1997. **278**(5338): p. 675-80.
6. Spencer, J.A., et al., *Genes on the short arm of the human X chromosome are not shared with the marsupial X.* Genomics, 1991. **11**(2): p. 339-45.
7. Lahn, B.T. and D.C. Page, *Four evolutionary strata on the human X chromosome.* Science, 1999. **286**(5441): p. 964-7.
8. Skaletsky, H., et al., *The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes.* Nature, 2003. **423**(6942): p. 825-37.
9. Ross, M.T., et al., *The DNA sequence of the human X chromosome.* Nature, 2005. **434**(7031): p. 325-37.
10. Charlesworth, B., *Model for evolution of Y chromosomes and dosage compensation.* Proc Natl Acad Sci U S A, 1978. **75**(11): p. 5618-22.
11. Muller, H.J., *The Relation of Recombination to Mutational Advance.* Mutat Res, 1964. **106**: p. 2-9.

12. Rice, W.R., *Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome.* Genetics, 1987. **116**(1): p. 161-7.

13. Aitken, R.J. and J.A. Marshall Graves, *The future of sex.* Nature, 2002. **415**(6875): p. 963.

14. Graves, J.A., *The degenerate Y chromosome - can conversion save it?* Reprod Fertil Dev, 2004. **16**(5): p. 527-34.

15. Mitchell, M.J., *Spermatogenesis and the mouse Y chromosome: specialisation out of decay.* Results Probl Cell Differ, 2000. **28**: p. 233-70.

16. Mazeyrat, S., et al., *The mouse Y chromosome interval necessary for spermatogonial proliferation is gene dense with syntenic homology to the human AZFa region.* Hum Mol Genet, 1998. **7**(11): p. 1713-24.

17. Bishop, C.E. and D. Hatat, *Molecular cloning and sequence analysis of a mouse Y chromosome RNA transcript expressed in the testis.* Nucleic Acids Res, 1987. **15**(7): p. 2959-69.

18. Eicher, E.M., et al., *A repeated segment on the mouse Y chromosome is composed of retroviral-related, Y-enriched and Y-specific sequences.* Genetics, 1989. **122**(1): p. 181-92.

19. Kuroda-Kawaguchi, T., et al., *The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men.* Nat Genet, 2001. **29**(3): p. 279-86.

20. Rozen, S., et al., *Abundant gene conversion between arms of palindromes in human and ape Y chromosomes.* Nature, 2003. **423**(6942): p. 873-6.

21. Hughes, J.F., et al., *Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee.* Nature, 2005. **437**(7055): p. 100-3.

22. Hughes, J.F., et al., *The fastest evolving chromosome in the hominids.* 2007.

23. Bergstrom, D.E., et al., *The mouse Y chromosome: enrichment, sizing, and cloning by bivariate flow cytometry.* Genomics, 1998. **48**(3): p. 304-13.

24. Toure, A., et al., *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, 2004. **83**(1): p. 140-7.

25. Waterston, R.H., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.

26. Roberts, C., et al., *Molecular and cytogenetic evidence for the location of Tdy and Hya on the mouse Y chromosome short arm.* Proc Natl Acad Sci U S A, 1988. **85**(17): p. 6446-9.

27. McLaren, A., et al., *Location of the genes controlling H-Y antigen expression and testis determination on the mouse Y chromosome.* Proc Natl Acad Sci U S A, 1988. **85**(17): p. 6442-5.

28. Pardue, M.L. and J.G. Gall, *Chromosomal localization of mouse satellite DNA.* Science, 1970. **168**(937): p. 1356-8.

29. Schnedl, W., *End-to-end association of X and Y chromosomes in mouse meiosis.* Nat New Biol, 1972. **236**(62): p. 29-30.

30. Sandstedt, S.A. and P.K. Tucker, *Evolutionary strata on the mouse X chromosome correspond to strata on the human X chromosome.* Genome Res, 2004. **14**(2): p. 267-72.

31. Amar, L.C., et al., *Conservation and reorganization of loci on the mammalian X chromosome: a molecular framework for the identification of homologous subchromosomal regions in man and mouse.* Genomics, 1988. **2**(3): p. 220-30.
32. Toure, A., et al., *A new deletion of the mouse Y chromosome long arm associated with the loss of Ssty expression, abnormal sperm development and sterility.* Genetics, 2004. **166**(2): p. 901-12.
33. Grzmil, P., et al., *The influence of the deletion on the long arm of the Y chromosome on sperm motility in mice.* Theriogenology, 2007. **67**(4): p. 760-6.
34. Ward, M.A. and P.S. Burgoyne, *The effects of deletions of the mouse Y chromosome long arm on sperm function--intracytoplasmic sperm injection (ICSI)-based analysis.* Biol Reprod, 2006. **74**(4): p. 652-8.
35. Styrna, J., J. Klag, and K. Moriwaki, *Influence of partial deletion of the Y chromosome on mouse sperm phenotype.* J Reprod Fertil, 1991. **92**(1): p. 187-95.
36. Ellis, P.J., et al., *Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids.* Hum Mol Genet, 2005. **14**(18): p. 2705-15.
37. Conway, S.J., et al., *Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse Y chromosome.* Mamm Genome, 1994. **5**(4): p. 203-10.
38. Burge, C. and S. Karlin, *Prediction of complete gene structures in human genomic DNA.* J Mol Biol, 1997. **268**(1): p. 78-94.
39. Thompson, J.D., et al., *The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.* Nucleic Acids Res, 1997. **25**: p. 4876-4882.

CHAPTER 3

Conclusion

Our sequencing of the mouse Y chromosome has added a great deal to the sum total knowledge about the chromosome. However, there is still much more to be done, both in sequencing efforts and in further experimentation, before we can truly understand the forces that have shaped the mouse Y, the functions of its genes and the purpose of its amplicons, and even more to be done before we understand sex chromosome evolution as a whole.

Mouse Y sequencing

In the preceding chapter, I presented the analysis of 75% of the mouse Y chromosome, comprising 72 Mb of sequence. This sequence was assembled into 79 contigs with a median length of 540 kb. That build of the mouse Y chromosome is obviously not the complete sequence of the mouse Y chromosome. And so the question remains, what would we learn from a more complete mouse Y sequence? It is possible that there are unique pockets of non-Huge Repeat sequence hidden inside the Huge Repeat array. These would most easily be found by a more complete sequencing of the chromosome. But that aside, the question we really should be asking is: What would we learn from a complete copy of the Huge Repeat array?

It is possible that we would find a different subfamily of Huge Repeat sequences, but I think it is unlikely that we would have missed it thus far. Most likely, the profile of Huge Repeat sequence at a repeat unit level would remain essentially the same. The most important set of information we would gain from further sequencing of the mouse Y chromosome would be the superstructure of the Huge Repeat array. We could find out if giant Huge Repeat palindromes such as the *Rbm31y*-containing palindrome are unusual or the norm on the mouse Y, and we could determine if there are even more unusual structures lurking in this array. And if the entire chromosome was sequenced we would be able to determine the relative placement of different types of Huge Repeat sequence across the chromosome. Are highly similar Huge Repeat sequences located more closely together on the chromosome or further apart? We should also keep in mind that since the Huge Repeat's high copy number and high sequence identity leaves it very vulnerable to recombination, the superstructure of one mouse's Huge Repeat array could be very different from another mouse's Huge Repeat, even within the same sub-species.

This further sequencing could be complemented by FISH experiments to localize any non-Huge Repeat sequence found on the chromosome. For instance, we could find the approximate location of the *Rbm31y*-containing palindrome and also definitively show how many copies of that gene are on the mouse Y chromosome.

In fact, we are continuing to sequence the mouse Y chromosome, and hope to continue until we exhaust all the resources of the two existing male C57BL/6 BAC libraries. Table 1 shows both the current state of mouse Y sequencing and what results we predict from each further round of BAC selection as described in the previous chapter. This is our best estimate of the future results of our sequencing efforts, but it does depend on several assumptions that should be discussed here.

In reality, BAC selection does not take place in discrete rounds, where one set of contig extensions is selected, and then all the chosen BACs are sequenced before the next set of extending BACs is selected. The process is more continuous, as we actually sequence BACs as they are chosen, and choose new BACs as soon as there is enough sequence data to select a new contig extension. Also, our model assumes that our BAC

Round	# of contigs	Largest contig (Mb)	Total non-redundant sequence (Mb)	Total BACs sequenced
Current	**67**	**4.8**	**75.7**	**814**
+1	33	7.5	88.0	992
+2	18	12.9	93.0	1068
+3	7	25.7	95.0	1104
+4	2	68.6	95.5	1118

Table 1 – Current and predicted future statistics of the mouse Y sequencing effort

libraries contain all the BACs required to cover the entire euchromatic sequence of the mouse Y chromosome, which is likely to be false. First of all, even though our two BAC libraries are statistically supposed to together cover any given point on the Y chromosome ten times, because of the random way that the BACs were produced from genomic DNA, it is quite possible that there are isolated points that have no coverage at all. Second, if any mouse Y sequences are lethal to bacteria, those genes will not be present in any BACs in our BAC libraries. Third, it is possible that there are portions of the Huge Repeat array that are more than 99.999% identical, making it impossible to assemble the entire array with confidence. All that aside, we are optimistic that our finished mouse Y sequence will have less than 5 gaps, yielding less than 6 contigs.

Thus far this discussion has been focused on the efforts required to sequence the Huge Repeat array. However, we should not forget another feature of the mouse Y chromosome that has still not been completely sequenced: the *Rbmy* array. Currently, we have sequenced both borders of the *Rbmy* array, as well as a sample *Rbmy* array BAC in the middle. Due to the short length of the *Rbmy* repeat unit (~40 kb), obtaining that sequence required extraordinary measures. Obtaining the complete sequence of the array would require an entirely different sequencing strategy, perhaps based on cosmids. But if the entire *Rbmy* array was sequenced, we would finally know the number of *Rbmy* genes on a wild type mouse Y chromosome and their arrangement. Also, as one of the breakpoints of the *Sxr*^{a} transposition is within the *Rbmy* array, we could use the sequence to try to understand how this transposition occurred. We will not have a complete picture of the structure of the mouse Y chromosome until we understand the structures of both the *Rbmy* array and the Huge Repeat array.

Mouse Y experimentation

I have previously mentioned the mystery of the identity and the location of the mouse Y centromere, and I believe that it is an entirely solvable mystery. Even though the sequence of the mouse Y centromere is unknown, we do know that it binds to some of the common centromeric proteins, such as aurora B kinase (Goldberg and Allis, personal communication). Even though it would be technically difficult, it should be possible to combine the hybridization of aurora B kinase antibodies with FISH of *Sxr*^{a} probes in mouse cells to show whether or not the *Sxr*^{a} region is on a short arm. Also, chromatin immunoprecipitation of metaphase mouse cells with aurora B kinase should pull down the centromeres of all the mouse chromosomes, including the Y. Hopefully, these centromeres could then be at least partially sequenced.

Even though it would be very interesting to find out why mouse Y gene knockouts are so difficult, I believe that the best route to take in examining mouse Y gene function would be to turn to RNAi knockdowns. This would be very useful for all of the single- and few-copy genes, but would be absolutely essential to study the Huge Repeat genes, which are found in over a hundred copies. This would be the only way to tease apart the functions of each HR gene, instead of examining the result of removing many copies of all the Huge Repeat genes, as in the partial long arm deletions. Hopefully, this would finally establish which gene families are responsible for the sperm head morphology and sex ratio distortion phenotypes. RNAi would be especially useful in the study of the HR genes, not just because knocking out a hundred gene copies in an ES cell is a ridiculous idea, but because the small inhibiting RNAs could be tailored to knockdown subpopulations of *Ssty*, for instance. This would help establish how many and which gene family members are functional.

Sex chromosome evolution

The sequence of the mouse Y chromosome tells us a great deal about the chromosome itself and the mouse genome as a whole, but it is also a great resource in helping us to understand sex chromosome evolution, and specifically mammalian sex chromosome evolution. To date, the only sequenced Y chromosomes come from humans, mice, and chimpanzees. Human and chimpanzee are only separated by six mya, and so the mouse Y chromosome represents just the second really independent example we have of mammalian sex chromosome evolution.

And what we have learned is that mammalian sex chromosomes are much more different than anyone had previously imagined. The story of mammalian Y chromosome evolution has always been described in a unified way, with all the mammalian Ys descended from a common autosomal ancestor and affected by the same degenerative forces. It was therefore expected that all mammalian Y chromosomes would be somewhat alike in both sequence and structure. Mammalian Y chromosomes were expected to be gene-poor and to have extensive homology between their gene sets (reviewed in [1]).

I found that these expectations are only true for the mouse Y chromosome if you exclude the Huge Repeat array. The 3 Mb non-HR segment of the mouse MSY has very few genes, and those genes are mostly homologous to human Y genes. But it is impossible to exclude the Huge Repeat when talking about the evolution of the mouse Y chromosome. Amplification has obviously played a large role on both the mouse and human Y chromosomes, and has given them very distinct and widely divergent structures. Not only that, but their sequences are also incredibly divergent: The Huge Repeat array is the largest non-syntenic block between the mouse and human genomes.

In my opinion, the Huge Repeat array was not an inevitability for the mouse. Sperm head morphology could have been handled in a much simpler way, with single-copy genes, as it is done in other mammalian species. And if the meiotic drive hypothesis is correct, then the whole Huge Repeat was started by a fluke – one of the mouse sex chromosomes acquiring the ability to transmit itself to more than 50% of sperm. From there, the Huge Repeat inevitably snowballed, as the X and the Y chromosomes fought each other for reproductive supremacy. It didn't have to happen that way. It is likely that if the mouse sex chromosome evolution experiment was re-run a dozen times, it would result in a dozen radically different Y chromosomes – each shaped by random chance to some degree.

What does that mean for all the other mammalian sex chromosomes that are yet to be sequenced? I predict that all of them have amplicons of some kind, ranging from two-copy palindromes to hundred-copy amplicons. These amplicons will likely contain testis genes. But other than that, I do not think that the sequence or structure of any specific mammalian Y chromosome can be predicted. We cannot know which genes will have degenerated, which remain as single-copy genes, and which will be amplified a hundred-fold (other than the likely continued existence of *Sry*). I also believe that X chromosomes are not as pure and inviolate as Ohno's Law would have us believe. The sex chromosomes shape each other, and it cannot be coincidence that the Huge Repeat genes have multi-copy paralogs on the X chromosome. I think that parallel amplification is likely to be seen again in other species. And this can only be tested by sequencing many other Y chromosomes, particularly mammalian Y chromosomes. For example, the rat is only separated by 12-14 mya from the mouse, and I predict that the rat Y chromosome will not contain the Huge Repeat array – and will therefore have no homology with 95% of the mouse Y chromosome. Finding out what it has been doing for the past 12 million years will be truly fascinating.

Having the sequence of the mouse Y chromosome will benefit many different groups of scientists. It enables further experimentation on the mouse Y chromosome itself, such as RNAi knockdowns of new and already known genes and gene families. It allows genomicists to study the entire mouse genome, instead of just the female genome. And for those who study Y chromosomes, it is a new point of data – only the third Y chromosome ever sequenced and the only sequenced Y chromosome from an organism more than 6 million years distant from *Homo sapiens*, giving us a very new perspective on mammalian sex chromosome evolution.

1. Graves, J.A., *The origin and function of the mammalian Y chromosome and Y-borne genes--an evolving understanding.* Bioessays, 1995. **17**(4): p. 311-20.