Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 7953–7957, September 1992
Genetics

Inverted repeat structure of the *Sry* locus in mice

(testis determination/*Tdy*^{m1}/deletion/Y chromosome)

JOHN GUBBAY*, NIGEL VIVIAN*, ANDROULLA ECONOMOU*, DAVID JACKSON†, PETER GOODFELLOW†, AND ROBIN LOVELL-BADGE*‡

*Laboratory of Eukaryotic Molecular Genetics, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom; and †Human Molecular Genetics Laboratory, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, United Kingdom

Communicated by Mary F. Lyon, April 20, 1992 (received for review February 12, 1992)

ABSTRACT **The testis-determining gene *Sry* is located on the short arm of the mouse Y chromosome in a region known to have undergone duplications and rearrangements in comparison with the equivalent portion of the human Y chromosome. Detailed analysis of the *Sry* genomic locus reveals a further difference in that the mouse *Sry* open reading frame lies within 2.8 kilobases of unique sequence at the center of a large inverted repeat. This repeat, which is found in both *Mus musculus musculus* and *Mus musculus domesticus* Y chromosomes, is not present at the human *SRY* locus. Recombination involving the repeat region may have led to an 11-kilobase deletion, precisely excising *Sry* in a line of XY female mice.**

There is now abundant evidence that *Sry* (*SRY* in humans) is the testis-determining gene on the mammalian Y chromosome (1, 2). The gene has many properties consistent with being a switch gene responsible for initiating the male pathway. For example, it encodes a DNA-binding protein (3, 4) and is expressed for a brief period just prior to overt testis differentiation (5). Conclusive proof that it is the only Y-linked gene required for testis determination, and that it is equivalent to the classically defined locus *Tdy* (*TDF* in humans), has come from the demonstration that chromosomally female mice transgenic for a 14.5-kilobase (kb) *Sry* genomic DNA fragment can show normal male development (6). This experiment also revealed that all the sequences necessary for *Sry* expression are present within this 14.5-kb region.

Mutations are invaluable in the analysis of gene structure and function. Human XY females have been found who possess point mutations in the *SRY* open reading frame, proving that a functional *SRY* is necessary for testis determination (7, 8) and indicating regions of the protein important for recognition of target DNA sequences (3, 4). However, there are only a few cases where XY females have been reported in the mouse. In one set of cases, specific Y chromosome variants from *Mus musculus domesticus* subspecies fail to interact correctly with autosomal loci provided by the C57BL/6 strain, and this leads to high frequencies of XY females and hermaphrodites (9, 10). The Y variants are presumed to have an altered *Sry* gene compared with that of *Mus musculus musculus*, but this could be in any aspect of structure or regulation (11). The best case of a mutation directly involving *Sry* is that provided by the *Tdy*^{m1} mutation present in a line of XY females (12). This arose in an experiment designed to obtain insertion of a retroviral vector into *Tdy*; however, no retroviral sequences could be found associated with the mutation. Instead, initial findings demonstrated that at least part of the *Sry* open reading frame had been deleted (2).

Toward understanding the *Sry* genomic organization as well as the regulation of its transcription unit, we have compared this genomic region among normal, variant, and mutant mouse strains. We have analyzed approximately 30 kb surrounding *Sry* in *M. musculus musculus* and *M. musculus domesticus* and determined the extent of the deletion in *Tdy*^{m1} XY females. This analysis includes the complete DNA sequence of the 14.5-kb fragment that must contain the entire *Sry* gene,§ and it reveals a number of features about the locus that allow us to speculate on the mechanism of the deletion.

MATERIALS AND METHODS

Mice. For simplicity, the Y chromosome carrying the *Tdy*^{m1} mutation is referred to as Y⁻ in the text. It is of 129 strain origin, and therefore of *M. musculus musculus* type. It is maintained by breeding XXY⁻ females as described previously (13). These mice and normal 129 strain animals were from the colony at the National Institute for Medical Research. Animals carrying the *Mus musculus poschiavinus* Y chromosome were obtained from stocks maintained by Paul Burgoyne at the Medical Research Council Mammalian Development Unit, London.

Southern Blotting. DNA samples were electrophoresed on 0.7% agarose gels in 0.09 M Tris borate, pH 8.3/2 mM EDTA (TBE) and transferred onto Hybond-N filters (Amersham) in 0.5 M NaOH/1.5 M NaCl. Probes were labeled with [^{32}P]dCTP by using a multiprime kit (Amersham), added to the filter in 3× SSC/0.1% sodium pyrophosphate/5× Denhardt's solution/0.1% SDS/10% dextran sulfate containing denatured herring testis DNA at 50 μg/ml and hybridized for 16 hr at 65°C (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0; 1× Denhardt's solution = 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone). Filters were washed in 1× SSC/0.1% SDS at 65°C for 1 hr and then exposed to x-ray film for 1–3 days.

Genomic Library Construction and Screening. DNA from the appropriate mouse strain was partially digested with the restriction enzyme *Sau*3a and, following methods previously described (2), DNA in the size range 18–23 kb was obtained by gel purification. The ends of this DNA were partially filled in by incubating 50 μg for 15 min at room temperature in 50 mM NaCl/10 mM Tris·HCl, pH 7.5/1 mM dGTP/1 mM dATP containing 15 units of the Klenow fragment of DNA polymerase. Then 0.4 μg of this DNA was ligated to a λ FixII vector (Stratagene). The reaction products were packaged by using Gigapack gold packaging extract (Stratagene) according to the manufacturer's instructions. Phage were plated in 0.7% top agarose containing 10 mM $MgSO_4$ after adsorption to plating bacteria for 15 min at 37°C (ref. 14, p. 63). Plating bacteria were prepared by resuspending an overnight culture of strain DL652 in ice-cold 10 mM $MgSO_4$. Phage DNA was

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. X67204).

transferred to Hybond-N filters, denatured in 0.5 M NaOH/1.5 M NaCl, neutralized in 1 M ammonium acetate, and hybridized as described in *Southern Blotting*.

Sequencing. The dideoxy method of sequencing (15), using the T7 sequencing kit (Pharmacia), was performed according to the manufacturer's instructions. Sequence analysis was performed by using software designed by the Genetics Computer Group at the University of Wisconsin (16).

Polymerase Chain Reaction (PCR). For PCR analysis 0.1 μg of genomic DNA was added to a 50-μl reaction mix containing 50 mM Tris·HCl at pH 9.0, 15 mM ammonium sulfate, 7 mM $MgCl_2$, bovine serum albumin at 0.17 mg/ml, 0.05% Nonidet P-40, 3.75 mM each dATP/dCTP/dGTP/dTTP, 1 unit of *Thermus aquaticus* (*Taq*) DNA polymerase (Anglian), and between 250 ng and 1 μg of each primer. Amplification consisted of denaturation for 3 min at 94°C followed by 30 cycles of 96°C for 40 s, 65°C for 70 s, and 72°C for 100 s in a Hybaid thermal reactor model HBTR1. PCRs were carried out in duplicate on each sample: (*i*) with 1 μg of a single primer (5′-GTGTCTCAAAGCCTGCTCTTC-3′) designed to amplify a 522-base-pair (bp) fragment across the deleted region in genomic DNA from X~~Y~~ female mice; (*ii*) with 500 ng each of primers for the *Sry* box region (5′-GAGAGCATGGAGGGCCAT-3′ and 5′-CCACTCCTCTGTGACACT-3′) and 250 ng each of primers for *Zfy-1* (5′-GACTAGACATGTCTTAACATCTGTCC-3′ and 5′-CCTATTGCATGGACTGCAGCTTATG-3′) (6). For visualization an 8-μl aliquot was electrophoresed on a 2% agarose/TBE gel containing ethidium bromide at 0.1 μg/ml.

RESULTS

The *Sry* Locus Consists of an Inverted Duplication. The cloning of a 14.5-kb genomic fragment (L741) presumed to contain the entire *Sry* gene has been described previously (2). To further characterize the locus we used a nonrepetitive probe (p2.3/A; Fig. 1) from one end of L741 to isolate additional phage clones from a mouse genomic library. The *Eco*RI restriction maps of L741 and overlapping phage inserts L823 and L832 are shown in Fig. 1. The *Sry* box is a region of strong homology between human and mouse SRY proteins corresponding to an "HMG box" type of DNA-binding domain (2). This is encoded by a 237-bp region, indicated by the shaded box in Fig. 1, which is part of a longer open reading frame. The presence of identically sized restriction fragments and their symmetrical arrangement on either side of the *Sry* box indicated that the sequence surrounding the gene was an inverted repeat. Combining the phage mapping data, we have extended the restriction map to cover approximately 33 kb of the *Sry* locus (Fig. 1). This map suggests that the repeat has a minimum length of 15.5 kb. The entire sequence of L741 has now been obtained (EMBL data base accession no. X67204), confirming the repeated nature of this locus. A 2.8-kb region of unique sequence, including the *Sry* box, can be defined within the inverted repeat as shown in Fig. 1.

Various lines of evidence suggest that this is the genuine structure of the region rather than an artefact of the cloning procedure used. (*i*) A number of overlapping genomic clones (some of which are shown in Fig. 1) have been isolated from three different unamplified genomic phage libraries made from two different mouse strains (see below), and the structures of all of these agree with the proposed restriction map. (*ii*) There is agreement between the phage maps shown and fragment sizes seen in genomic Southern blots (data not shown and see ref. 2). (*iii*) The *Eco*RI site immediately to the right of the *Sry* box, being in the unique region, is the only asymmetric *Eco*RI site present (see Fig. 1). Thus a probe covering one end of the inverted repeat such as the 1.5-kb *Eco*RI fragment from L741 when used on a Southern blot of genomic DNA digested with *Eco*RI should recognize both a 1.5-kb fragment (itself) and a 3.5-kb fragment that contains the same sequence as part of the other copy of the repeat. We have described this result in an earlier report, although its significance was not understood at that time (2). (*iv*) Although the duplicated regions on either side of the *Sry* box are almost identical, there are at least 7 single-base differences. These differences have been verified by sequencing from both strands. If the duplication was introduced during cloning such mismatches would be unlikely. In addition, a number of blocks of repeats of the dinucleotide CA, also known as microsatellites, are present within the repeated region. Some of these microsatellites have been found to be polymorphic in length on either side of the duplication. (*v*) Finally, we have previously shown that L741, which contains the inverted duplication, is functionally active in that XX individuals transgenic for L741 can be sex reversed (6).

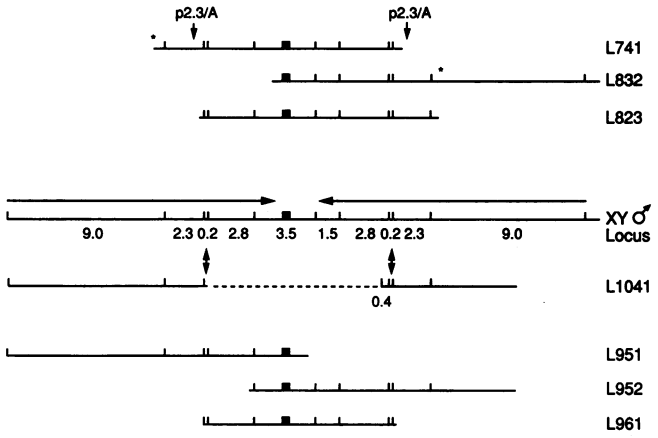

FIG. 1. Guide to *Sry* genomic clones and structure of the wild-type and X~~Y~~ female locus. *Eco*RI restriction maps of the following genomic phage inserts are shown: L741, L832, and L823, *Sry* locus from a 129 mouse (*M. musculus musculus* Y chromosome); L951, L952, and L961, *Sry* locus from an *M. musculus poschiavinus* mouse (*M. musculus domesticus* Y chromosome); and L1041, *Sry* locus from an X~~Y~~ female mouse carrying the *Tdy*^{m1} mutation (the broken line indicates deleted sequences). The *Sry* open reading frame is in the 5′-to-3′ orientation from left to right. The orientation of this clone with respect to the *Sry* box is not known. Restriction sites are indicated by vertical dashes. The position of the *Sry* box is indicated by a shaded box. The positions of probe p2.3/A are shown by arrows. XY ♂ Locus indicates the derived restriction map of the wild-type locus. Restriction fragment sizes are indicated in kb. The extent and end points of the duplicated sequences are indicated by horizontal arrows. The presence of a second 9.0-kb *Eco*RI fragment on the left side of the *Sry* box was inferred in the following way: A probe from the extreme left end of L741 is known to hybridize in two places (shown by asterisks), but recognizes only a single 9.0-kb *Eco*RI fragment on a genomic Southern blot (data not shown). The approximate positions of a single oligonucleotide primer used for PCR amplification of the deletion breakpoints are indicated by vertical arrows between the two loci.

The Inverted Duplication Predates the Divergence of *M. musculus musculus* and *M. musculus domesticus*. The *Sry* genomic locus was initially cloned from mouse strain 129, which carries a Y chromosome of *M. musculus musculus* origin (which we will call the *musculus* Y chromosome). To address the question of the evolutionary age of this inverted repeat, we have cloned the *Sry* genomic locus from a Y chromosome of *M. musculus domesticus* origin (the *domesticus* Y chromosome). The *M. musculus poschiavinus* substrain has been proposed to carry a late-acting allele of *Tdy* (10, 11) responsible for the high frequency with which XY females and hermaphrodites are seen when it is present on a C57BL/6 genetic background. This was chosen, therefore, as an example of a *domesticus* Y chromosome to also eventually define the molecular basis of this effect. An *M. musculus*

poschiavinus genomic library was screened with the *Sry* box. Fig. 1 shows the *Eco*RI restriction maps of three of the phage inserts obtained: L951, L952, and L961. The inverted duplication is clearly present on this Y chromosome, as shown by the symmetrical arrangement of fragment lengths around *Sry*. The *Eco*RI restriction pattern is identical to that seen in the *musculus* Y chromosome. These data suggest that the inverted repeat at the *Sry* locus arose before the divergence of the *musculus* and *domesticus* Y chromosomes. Given the similarity in the restriction maps of the *Sry* loci from the two Y chromosomes, it is clear that a much higher resolution of analysis will be required to correlate *Sry* structure with functional differences between alleles.

Analysis of the *Sry* Locus in Mice Carrying the Tdy^{m1} Mutation. Our previous results suggested that the Tdy^{m1} mutation was due to a deletion encompassing the *Sry* box, and one breakpoint of this deletion was already indicated (2). To fully define this deletion, we used probe p2.3/A (see Fig. 1), shown previously to be present in the Y chromosome carrying the mutation (referred to subsequently as Y⁻) to screen a phage genomic library of DNA from an XY⁻ female mouse. An *Eco*RI restriction map of the phage insert obtained, L1041, is shown in Fig. 1. This indicates that the entire deletion covers just under 11 kb of sequence including and adjacent to the *Sry* box. Upon digestion with *Eco*RI the only novel fragment produced from L1041 not seen in any phage inserts from wild-type DNA was a 400-bp fragment. The position of this is shown in Fig. 1. This fragment was sequenced, as it was presumed to contain the deletion breakpoints themselves. Fig. 2 shows the sequence of this fragment and one possible alignment of this sequence to the two arms of the inverted duplication at the *Sry* locus. The alignment shown in Fig. 2 represents the mutation as two deletions, one being only 239 bp, while the other spans from one arm of the inverted duplication to the other. This is shown as Scheme 1 in Fig. 3. However, as the sequences on either side of the *Sry* box are palindromic, an alternative interpretation is that a single deletion has occurred accompanied by a small inversion of 264 bp. This is shown as Scheme 2 in Fig. 3. To confirm that L1041 accurately represents the genome in XY⁻ female mice, we designed an assay to amplify a fragment at the breakpoint in DNA from XY⁻ females by using PCR. Because the deletion juxtaposes sequences that are palindromic, a single PCR primer should be sufficient to amplify a 522-bp fragment around the breakpoint (Fig. 1). We were able to amplify such a fragment from XY⁻ female DNA but not wild-type DNA (Fig. 4). The fragment obtained was of the appropriate size and its sequence was identical to that of the breakpoint fragment from L1041.

DISCUSSION

We have shown that the open reading frame encoding the mouse *Sry* DNA-binding domain is contained within 2.8 kb of unique sequence flanked by long inverted repeats of at least 15.5 kb. Inverted repeats are a known feature of eukaryotic genomes, but, as far as we know, *Sry* represents the first example where the gene itself is not duplicated but present in a unique interrepeat region (17–19). For example, this may be contrasted with α-amylase genes, where two members of the family are organized in the form of a large inverted repeat with a unique central core about 2 kb in length (20). Other inverted repeats are associated with very large gene families, such as the mouse major urinary proteins (21), with centromeric regions (22, 23), or with instances of gene amplification (24). It has been suggested (24) that the mechanisms of gene

```
          1                                                                  80
Left.     GAATTCCACACTGAGGCCTAATGCCTGCTGTTTCTTCTGTGTCTCAAAGCCTGCTCTTCCTACCTGAAATAGGATAGTCC
Del.      GAATTCCACACTGAGGCCTAATGCCTGCTGTTTCTTCTGTGTCTCAAAGCCTGCTCTTCC.....

          81                                                                 160
Left.     TTTTTTTCCCCCTTGGTACTTCTTTACCTTATTACCACACCAAGACTTAATACTCTTGCTTTTATTTCTCCTTTCAGTTA
Del.      ................................................................................

          161                                                                240
Left.     CCTGTTATGCCTTTCAGATTGTCACCCTCTATCTGAATTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
Del.      ................................................................................

          241                                                                320
Left.     GTGTGTGTGTGTGTGTGTGTATGTGTGTGTGTGTGTTTGTGTGTGTCTTTTTGTGTGTGATGTGTGATTGGCGATT
Del.      ..............................................................TGTGATGTGTGATTGGCGATT

          321                                                                400
Left.     GTTGTCTTTGCTTTGTACACACACTTTACTTTGCACACACACACACACACACACACACACACACA........GAGTG
Del.      GT...CTTTGCTTTGTACACACACTTTACTTTGCACACACACACACACACACACACACACACACACACAGAGTG

          401                                                                480
Left.     TTTATTGATATGTAAATCACTGGATTCAGTTTACAGAGATTCAGATGTCGTTTGGAATATGATAGCTGCTAGCAGTACAT
Del.      TTTATTGATATGTAAATCACTGGATTCAGTTTACAGAGATTCAGATGTCGTTTGGAATATGATAGCTGCTAGCAGTACAT

          481                                                                560
Left.     GTTGCTTGTTCAGCTTTGCTTCCTGATTCAGATGTGTGCCCACTGTTGCTTTAGGTTGCTTCAGGCCTGCCTAACCACAC
Del.      GTTGCTTGTTCAGCTTTGCTTCCTGATTCAGATGTGTGCCCACTGTTGCTTTAGGTTGCTTCAGGCCTGCCTAACCACAC

          561                                                                640
Left.     CACCTCCT→
Del.      CACCTCaaacacacacacacacacatacacacacacacacacacacacacacacacacacacacacacacacacacac
Right.    ←acacaaacacacacacacacacatacacacacacacacacacacacacacacacacacacacacacacacacacac

          641        654
Del.      acacacatgaattc
Right.    acacacatgaattc
```

FIG. 2. Sequences at the deletion breakpoints in XY⁻ females. The sequence of the 0.4-kb *Eco*RI fragment from L1041 (Del.), which contains the XY⁻ female deletion breakpoints, is compared with sequences from either arm of the duplication. Sequences from the left arm of the duplication—i.e., 5′ of the *Sry* open reading frame (Left.)—are shown in uppercase, while those from the right arm—i.e., 3′ to the *Sry* open reading frame (Right.)—are shown in lowercase. The deletion sequence is shown in uppercase where it corresponds to the left arm and lowercase when it corresponds to the right arm. The arrangement shown corresponds to scheme 1 (Fig. 3)—i.e., one short deletion (shown by dots) and a major deletion at position 568. The horizontal arrows at this point represent a continuation of sequence from the left and right sides of the locus in wild-type DNA that is deleted in the XY⁻ female. The deletion sequence corresponds to sequences at the wild-type locus except at two points: (*i*) a TGT sequence is missing from the deletion sequence at position 323, and (*ii*) the length of one of the CA repeat motifs differs in the deletion sequence from that of the wild-type sequence. This particular motif is polymorphic in length on either side of *Sry* in wild-type mice, and thus it may represent a polymorphism that already existed in the Y chromosome that gave rise to the XY⁻ female line. Alternatively, additional CA residues may have been added during the deletion process.

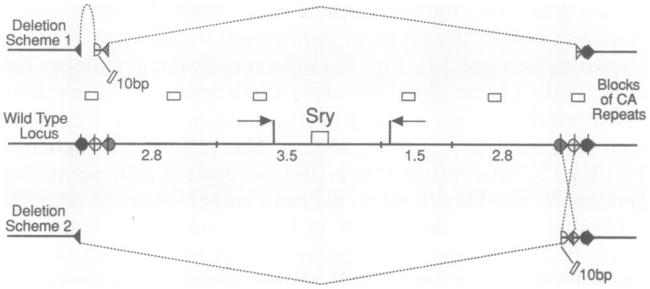

FIG. 3. Two possible schemes for the deletion of *Sry* in X$\cancel{Y}$ female mice. The wild-type locus is shown with restriction sites for *Eco*RI indicated by vertical dashes and fragment sizes in kb. The end points of the duplicated regions are marked by horizontal arrows above the locus. Rectangular blocks represent the approximate position of blocks of CA·GT repeats. Three points along the sequence, marked on both arms of the repeat, are illustrated by triangle/semicircle symbols. DNA breakage points have occurred at the centers of these symbols. As these symbols have polarity, the presence of either a triangle or a semicircle in the deleted locus (Schemes 1 and 2) indicates which side of a particular breakpoint sequences originate from. In both schemes there are three breakpoints. Two of these breakpoints are at almost identical positions on either side of *Sry* but differ by 10 bp. A 10-bp block of sequence is therefore shown being lost to compensate for this slight difference. Broken lines indicate the position of deletion and inversion events.

amplification and stable gene duplication during evolution may share common features, such as repeated initiation of DNA synthesis from a particular point on the chromosome followed by multiple recombination events.

In cases where a gene is duplicated there are obvious reasons why it may be a selective advantage to maintain the structure and homology between the genes. With *Sry*, it is the duplicated nongenic arms that show almost perfect conservation at the sequence level. There are at least four possible explanations for this phenomenon. (*i*) The duplication may have been a recent event, allowing little time for divergence.

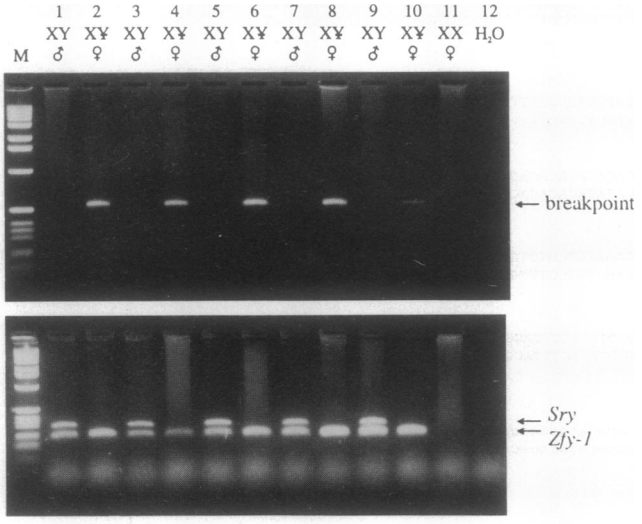

FIG. 4. PCR analysis of the deletion in X$\cancel{Y}$ female mice. Analysis was performed on 11 different genomic DNA samples (from 5 XY ♂ mice, 5 X$\cancel{Y}$ ♀ mice, and 1 XX ♀ mouse). (*Upper*) A 522-bp fragment (breakpoint) was amplified only in X$\cancel{Y}$ female samples when a single PCR primer adjacent to the deletion breakpoints was used. The strict correlation between amplification of the breakpoint fragment and the $\cancel{Y}$ chromosome indicates that plasmid DNA contamination is not present. (*Lower*) The same samples were tested for PCR amplification of the *Sry* box (*Sry*), for which a band is present only in normal males, and *Zfy-1*, for which a band is present in all samples except XX ♀. Marker bands (M) are 1353, 1078, 872, 603, 310, 281, 271, 234, 194, 118, and 72 bp.

Although we have shown that both a *musculus* and a *domesticus* Y chromosome carry this duplication, these chromosomes diverged within the last 1–2 million years (see ref. 25). Studies of more distantly related *Mus* species and of other species will be necessary to determine the evolutionary history of this event. (*ii*) Gene conversion events may be responsible for maintaining conservation between the two arms of the repeat. (*iii*) Sequence conservation can also result from a functional role; for instance, *Sry* regulatory sequences, such as enhancers, which may need to be present in duplicate, could reside within the repeated area. However, these are generally small motifs, which would not explain the large area of conservation. Studies on the regulation of *Sry* should shed light on this matter. (*iv*) Sequences in the duplication (or the duplication itself) may be required to maintain a specific chromatin configuration necessary for the correct regulation of *Sry* expression.

Human *SRY* is not surrounded by an inverted duplication (1, 26). Perhaps this is not surprising, as there are dramatic differences between the organization of the human and mouse Y chromosomes. Human *SRY* is located only 5 kb away from the pseudoautosomal boundary (27), while mouse *Sry* is located at the opposite end of the chromosome from the pseudoautosomal region (28). It is known that the short arm of the mouse Y chromosome (where *Sry* resides) has undergone a number of duplications and rearrangements. For instance, mouse *Zfy* genes are present in two copies as a large tandem repeat, while only one *Zfy* gene is found in most other species (29–31). No homology has been found outside the *Sry* box region with human *SRY*, and the organization of the transcription unit appears to be different (unpublished data).

The *Sry* deletion mutant described here lends further weight to the argument that *Sry* is normally necessary for testis determination. A comparison of the sequence from the 14.5-kb *Sry* genomic clone with 25 kb of sequence at the human *SRY* locus failed to reveal the presence of any other conserved genes (D.J. and A. Sinclair, unpublished results). As this comparison includes the entire portion deleted in the X$\cancel{Y}$ females, *Sry* must be the only gene affected (barring cis effects on any closely linked gene).

Lovell-Badge and Robertson (12) proposed that the mutation in the X$\cancel{Y}$ females must have occurred after the retroviral infection protocol and may therefore have been an entirely unconnected event. Retroviral excision events can give rise to deletions; however, a single long terminal repeat (LTR) is usually left behind as excision occurs by recombination between LTRs (32). As all the sequences at the breakpoint of this deletion are from the endogenous locus, it seems unlikely that insertion and subsequent excision of a retroviral or retroviral-like element was its cause.

Clearly the mutation itself was a complex event, as shown in Fig. 3, involving either two individual deletions or one deletion accompanied by a short inversion. Either scheme would involve three points of DNA breakage, two of these being at almost identical points (±10 bp) in each arm of the inverted duplication. Rather than this being a completely random event, the unusual properties of the locus may have contributed to the mutation. The role of direct repeats in the generation of deletions is well documented. Such deletions occur by unequal sister chromatid exchange, unequal crossing-over, or replication slippage (33). In humans, these mechanisms have been proposed to explain deletions giving rise to diseases such as the thalassemias (34) and the generation of new length alleles at tandem repetitive hypervariable loci (35). The *Sry* deletion has occurred near blocks of CA repeats present on either side of the *Sry* box. Such repeats are thought to be randomly distributed about the genome but are not uncommon near genes (36). The proximity of the deletion breakpoints to microsatellites suggests that they may have acted as tandem repetitive elements to promote the deletion

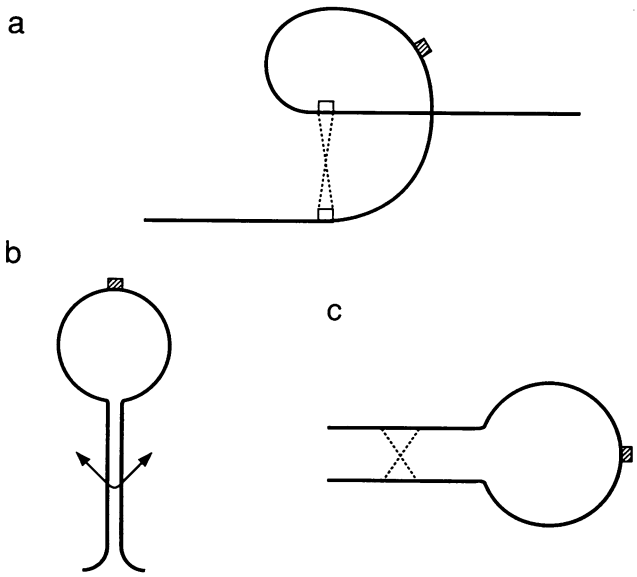

FIG. 5. Schematic representation of recombination events that may occur at the *Sry* locus. The hatched box indicates the position of the *Sry* box. Blocks of CA·GT repeats are indicated by open boxes. Broken lines indicate the position of double-stranded recombination events. (*a*) Tandemly repeated elements can promote deletions by intramolecular recombination. (*b*) DNA sequences (possibly single-stranded) can be juxtaposed due to their palindromic nature, and subsequent strand breakage and religation can result in the generation of a deletion (indicated by a double-headed arrow). (*c*) Intramolecular recombination promoted by palindromic sequences can lead to "flip-flop" inversion of intervening sequences.

by one of the mechanisms mentioned above (see Fig. 5). An alternative possibility is that the palindromic nature of the locus may have played a part in juxtaposing the deletion breakpoints by forming a stem–loop structure. It has been proposed that short palindromic sequences may have played such a role in a 5-kb deletion in the human low density lipoprotein receptor (37). This model invokes the formation of a stem–loop structure involving single-stranded DNA during replication. Nicking and subsequent religation within the stem leads to deletion of the looped-out unique sequences (see Fig. 5).

One other expected consequence of the inverted duplication at this locus may be a capacity to invert the unique sequences at the center of the duplication, including *Sry*. This could occur by homologous pairing between sequences in the inverted repeat followed by intramolecular recombination (see Fig. 5). Flip-flop inversion of specific DNA sequences flanked by inverted repeats is known to occur in bacteria, where it can affect expression of adjacent genes—for example, in flagellar phase variation (38, 39). It is also thought to be the mechanism for an inversion event at the *Drosophila melanogaster* α-Amylase locus, resulting in a null allele of this gene (40). Due to the extreme sequence conservation of the *Sry* inverted repeat, inversion events not accompanied by additional mutations may have no affect on *Sry* function. Once a DNA sequence outside the limits of the duplication has been isolated it should be possible to test for changes in the orientation of *Sry* with respect to this new marker.

In conclusion, we have shown that mouse *Sry* is at the center of a large inverted repeat, present on both the *musculus* and *domesticus* Y chromosomes. In addition, the structure of the mouse locus may have promoted a deletion of *Sry* that resulted in the generation of XY female mice.

We are indebted to Anne-Marie Frischauf for help with constructing the XY female genomic library and Paul Burgoyne for providing mice carrying the *M. musculus poschiavinus* Y chromosome. We also thank the members of our two groups—in particular, Blanche Capel, Peter Koopman, and Andrew Sinclair—who have contributed to this work in terms of ideas, unpublished data, and critical reading of the manuscript.

1. Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. W., Frischauf, A.-M., Lovell-Badge, R. & Goodfellow, P. N. (1990) *Nature (London)* **346,** 240–244.
2. Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Münsterberg, A., Vivian, N., Goodfellow, P. & Lovell-Badge, R. (1990) *Nature (London)* **346,** 245–250.
3. Harley, V. R., Jackson, D., Hextall, P., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell-Badge, R. & Goodfellow, P. N. (1992) *Science* **255,** 453–456.
4. Nasrin, N., Buggs, C., Kong, X. F., Carnazza, J., Goebl, M. & Alexander-Bridges, M. (1991) *Nature (London)* **354,** 317–320.
5. Koopman, P., Münsterberg, A., Capel, B., Vivian, N. & Lovell-Badge, R. (1990) *Nature (London)* **348,** 450–452.
6. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. (1991) *Nature (London)* **351,** 117–121.
7. Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B. L., Goodfellow, P. N. & Fellous, M. (1990) *Nature (London)* **348,** 248–250.
8. Jäger, R. J., Anvret, M., Hall, K. & Scherer, G. (1990) *Nature (London)* **348,** 452–454.
9. Eicher, E. M. (1988) *Philos. Trans. R. Soc. London B* **322,** 109–118.
10. Eicher, E. M. & Washburn, L. L. (1986) *Rev. Genet.* **20,** 327–360.
11. Palmer, S. J. & Burgoyne, P. S. (1991) *Development* **113,** 709–714.
12. Lovell-Badge, R. H. & Robertson, E. (1990) *Development* **109,** 635–646.
13. Gubbay, J., Koopman, P., Collignon, J., Burgoyne, P. & Lovell-Badge, R. (1990) *Development* **109,** 647–653.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
15. Sanger, F. S., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
16. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
17. Kant, J. A., Fornace, A. J., Saxe, D., Simon, M. I., McBride, O. W. & Crabtree, G. R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2344–2348.
18. Russnak, R. H. & Candido, E. P. M. (1985) *Mol. Cell. Biol.* **5,** 1268–1278.
19. Wang, S. W., Robins, A. J., d'Andrea, R. & Wells, J. R. E. (1985) *Nucleic Acids Res.* **13,** 1369–1387.
20. Groot, P. C., Mager, W. H., Henriquez, N. V., Pronk, J. C., Arwert, F., Planta, R. J., Eriksson, A. W. & Frants, R. R. (1990) *Genomics* **8,** 97–105.
21. Bishop, O., Selman, G. G., Hickman, J., Black, L., Saunders, R. D. P. & Clark, A. J. (1985) *Mol. Cell. Biol.* **5,** 1591–1600.
22. Polizzi, C. & Clarke, L. (1991) *J. Cell Biol.* **112,** 191–201.
23. Rattner, J. B. (1991) *Bioessays* **13,** 51–56.
24. Ford, M. & Fried, M. (1986) *Cell* **45,** 425–430.
25. Tucker, P. K., Lee, B. K. & Eicher, E. M. (1989) *Genetics* **122,** 169–179.
26. Ellis, N. & Goodfellow, P. N. (1989) *Trends Genet.* **5,** 406–410.
27. Palmer, M. S., Sinclair, A. H., Berta, P., Ellis, N. A., Goodfellow, P. N., Abbas, N. E. & Fellous, M. (1989) *Nature (London)* **342,** 937–939.
28. McLaren, A., Simpson, E., Epplen, J. T., Studer, R., Koopman, P., Evans, E. P. & Burgoyne, P. S. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6442–6445.
29. Mardon, G., Mosher, R., Disteche, C. M., Nishioka, Y., McLaren, A. & Page, D. C. (1989) *Science* **243,** 78–80.
30. Nagamine, C. M., Chan, K., Kozak, C. A. & Lau, Y.-F. (1989) *Science* **243,** 80–83.
31. Simpson, E. M. & Page, D. C. (1991) *Genomics* **11,** 609–620.
32. Mager, D. L. & Goodchild, N. L. (1989) *Am. J. Hum. Genet.* **45,** 848–854.
33. Crow, J. F. & Dove, W. F. (1988) *Genetics* **120,** 1–6.
34. Henthorn, P. S., Smithies, O. & Mager, D. L. (1990) *Genomics* **6,** 226–237.
35. Jeffreys, A. J., Royle, N. J., Wilson, V. & Wong, Z. (1988) *Nature (London)* **332,** 278–281.
36. Jongeneel, C. V., Acha-Orbea, H. & Blankenstein, T. (1990) *J. Exp. Med.* **171,** 2141–2146.
37. Lehrman, M. A., Russel, D. W., Goldstein, J. L. & Brown, M. S. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3679–3683.
38. Johnson, R. C., Bruist, M. F. & Simon, M. I. (1986) *Cell* **46,** 531–539.
39. Simon, M. & Herskowitz, I., eds. (1985) *Genome Rearrangements* (Liss, New York).
40. Schwartz, P. E. & Doane, W. W. (1989) *Biochem. Genet.* **27,** 31–46.