DNA Sequence Analysis of Sry Alleles (Subgenus Mus) Implicates Misregulation as the Cause of C57BL/6J-Y<sup>POS</sup> Sex Reversal and Defines the SRY Functional Unit

Kenneth H. Albrecht and Eva M. Eicher

The Jackson Laboratory, Bar Harbor, Maine 04609

Manuscript received January 8, 1997
Accepted for publication July 21, 1997

ABSTRACT

The Sry (sex determining region, Y chromosome) open reading frame from mice representing four species of the genus Mus was sequenced in an effort to understand the conditional dysfunction of some M. domesticus Sry alleles when present on the C57BL/6J inbred strain genetic background and to delimit the functionally important protein regions. Twenty-two Sry alleles were sequenced, most from wild-derived Y chromosomes, including 11 M. domesticus alleles, seven M. musculus alleles and two alleles each from the related species M. spicilegus and M. spretus. We found that the HMG domain (high mobility group DNA binding domain) and the unique regions are well conserved, while the glutamine repeat cluster (GRC) region is quite variable. No correlation was found between the predicted protein isoforms and the ability of a Sry allele to allow differentiation of ovarian tissue when on the C57BL/6J genetic background, strongly suggesting that the cause of this sex reversal is not the Sry protein itself, but rather the regulation of SRY expression. Furthermore, our interspecies sequence analysis provides compelling evidence that the M. musculus and M. domesticus SRY functional domain is contained in the first 143 amino acids, which includes the HMG domain and adjacent unique region (UR-2).

Although sex determination is an integral part of the development of most animals, sex determination mechanisms are not evolutionarily conserved and seem to be rapidly evolving. For example, a comparison of the sex determination process in three of the most studied experimental systems (Mus musculus/domesticus, Drosophila melanogaster and Caenorhabditis elegans) reveals little similarity (RYNER and SWAIN 1995). Mammalian male sex determination is initiated by the expression of the Y chromosomal Sry gene. Interspecific DNA sequence analysis among human, mouse, rat, rabbit, and dunnart indicates that only one region within the Sry protein (SRY) coding unit, the HMG box DNA binding motif, is well conserved (WHITFIELD et al. 1993). A comparison of partial Sry sequences among seven species of Old World mice and rats indicated that Sry is rapidly evolving in the regions outside the HMG box in these rodents (TUCKER and LUNDRIGAN 1993). While it is clear that the Sry sequence is rapidly evolving, it is not yet clear if Sry function also is rapidly evolving. For example, in laboratory mice Sry is probably expressed exclusively in the fetal XY gonad during a very restricted time period; whereas in humans and marsupials, Sry expression is widespread both spatially and temporally (KOOPMANN et al. 1990; CLEPET et al. 1993; HARRY et al. 1995). Furthermore, Sry exists in multiple copies in some Old World rodents but is single copy in some rodent species, including M. musculus and M. domesticus (NAGAMINE 1994).

We have sequenced the Sry open reading frame (ORF) from 22 different Y chromosomes representing four different species in the genus Mus (family Muridae) and have identified 15 structural alleles: seven from M. domesticus, five from M. musculus, two from M. spretus and one from M. spicilegus. We reasoned that this approach would address the biological question of why transfer of certain M. domesticus Y chromosomes onto the C57BL/6J (B6) inbred strain causes a disturbance of male sex determination that is manifested by the development of ovarian tissue in XY individuals. This phenomenon is termed C57BL/6J-Y<sup>POS</sup> (B6-Y<sup>POS</sup>) sex reversal because it was first described for a Y chromosome derived from the M. domesticus poschiavinus subspecies (EICHER et al. 1982). B6-Y<sup>POS</sup> sex reversal is genetically complex. First, both a Y chromosome component and autosomal components are involved (EICHER and WASHBURN 1983). The Y chromosome component is Sry and the autosomal components are comprised of B6 alleles of genes (referred to as tda, testis determining, autosomal) that interact inappropriately with certain M. domesticus Sry alleles (EICHER et al. 1995). Three of the tda genes recently were genetically mapped (EICHER et al. 1996). Second, not all M. domesticus Y chromosomes (i.e., Sry alleles) behave similarly to the M. d. poschiavinus Y chromosome when present on the B6 inbred strain genetic background (and conversely, not all inbred strains are similar to B6). Some M. domesticus Y chromosomes function as normal male determinants when on the B6 genetic background, i.e., there is no sex ratio distortion or appearance of ovarian.
tissue in XY individuals. Other \textit{M. domesticus} Y chromosomes function in a potentially intermediate manner in that their presence causes a delay in the completion of testis cord differentiation; however, no ovarian tissue is formed (Washburn and Eicher 1983; Washburn and Eicher 1989; Washburn et al. 1990; Palmer and Burgoyne 1991; Burgoyne and Palmer 1993). Thus, \textit{Sry} sequence information from different \textit{M. domesticus} alleles could be informative, especially if these \textit{Sry} alleles were tested on the B6 background. Our study included analysis of \textit{Sry} alleles in several B6-Y consomic strains where the Y chromosome was derived from \textit{M. domesticus} or \textit{M. musculus} male mice and transferred onto the B6 background. Many of the Y chromosomes were isolated from males of wild-derived strains or feral populations.

Based on sequence data from \textit{M. domesticus} \textit{Sry} alleles present in inbred strains, Coward and colleagues proposed that the number of glutamine residues (CAG trinucleotide repeats) present in one region (here referred to as the third glutamine repeat cluster or GRC-3, Figure 1) correlated with the presence or absence of sex reversal (Coward et al. 1994). Specifically, they hypothesized that 11 glutamine residues correlated with sex reversal, 12 glutamine residues allowed normal male sex determination, and 13 glutamine residues caused a delay in testis cord formation. While our DNA sequencing experiments were in progress, a second publication from this group suggested that the number of glutamine residues did not correlate with the presence or absence of sex reversal (Coward et al. 1994). Their studies concentrated on partial sequence analysis of the \textit{Sry} ORF. In contrast, we sequenced the entire ORF in each \textit{Sry} allele analyzed in an effort to implicate or vindicate the \textit{SRY} protein as causative in the B6-Y\textsuperscript{POS} sex reversal condition.

In addition to a better understanding of the \textit{Sry} gene itself, we reasoned that defining the molecular nature of the dysfunction of some \textit{M. domesticus} \textit{Sry} alleles when present on a B6 genetic background could lead to a better understanding of the genetic position of the \textit{ida} genes in the sex determination pathway. The "functional polymorphism" in the \textit{M. domesticus} \textit{Sry} alleles that causes conditional sex reversal may be structural, regulatory, or both in nature. If the functional polymorphism is structural, then the \textit{ida} genes are likely to be \textit{Sry} responders, since \textit{Sry} is most likely a transcription factor that regulates downstream sex determination genes. If the functional polymorphism is regulatory (i.e., the timing, levels of expression, or stability of \textit{SRY} protein is influenced), then the position of the \textit{ida} genes in the sex determination genetic hierarchy is not obvious.

We also envisioned that an investigation of sequence divergence among \textit{Sry} alleles in the genus \textit{Mus} might identify conserved protein regions and thus delineate functionally important \textit{SRY} domains. These data would add information to how this important gene functions in mammalian sex determination. Biologically, our results indicate that there is no correlation between the different \textit{Sry} protein isoforms and the presence or absence of ovarian tissue in XY mice. We conclude that the cause of B6-Y\textsuperscript{POS} sex reversal lies in the regulation of \textit{Sry} and not in a protein structural polymorphism. The results also indicate that only the GRC region is highly variable within the \textit{Mus} species studied and that the HMG domain and the unique regions are well conserved. From these results, we propose that the GRC region is dispensable for normal \textit{Sry} function and that the \textit{M. musculus/domesticus} \textit{SRY} functional unit is contained in the HMG domain and perhaps the flanking unique regions.

\section*{MATERIALS AND METHODS}

\textbf{Mice:} C57BL/6JEi-Y consomic lines were established by standard backcrossing methods using wild-derived or inbred male mice and the C57BL/6JEi (B6) inbred strain subline. Inbred wild-derived strains were established by conventional inbreeding methods. Mouse species, strain designations, and the original collecting localities for the \textit{Y} chromosomes used in this study are listed in Table 1. In this report we refer to \textit{M. domesticus} and \textit{M. musculus} as distinct species, instead of subspecies or semispecies, based on data from hybrid zone investigations of these two taxa (Tucker et al. 1992b).

Sex reversal in a B6-Y consomic stock was initially recognized in preweanling mice by the presence of an intermediate anal-genital distance together with mammary gland associated yellow pigmented hairs. Sex reversal was confirmed by the presence of ovarian tissue in XY individuals. At least 250 offspring from multiple matings and litters were examined from each non-sex reversing Y chromosome consomic strain and in these cases, no evidence was found for sex reversal nor deviation from the expected 1:1 sex ratio (data not shown). For the two sex reversing B6-\textit{M. domesticus} consomic strains first described here (B6-Y\textsuperscript{DAM} and B6-Y\textsuperscript{YAM}) fetal gonads were analyzed at 14.5-15.5 days of fetal development as described in Eicher et al. (1989).

\textbf{PCR amplification of \textit{Sry}:} The \textit{Sry} ORF was PCR amplified for direct sequencing as follows: 100–150 ng of genomic DNA served as template for primers \textit{Sry}-8212 (5'-TTGTATTTTTAGTGTTCACCCCTACAGCGS') and \textit{Sry}-9791 (5'-AGCTGTTTGCTGTCTTTGTGCTAGCC-3') in a 100-uL reaction. Primers are designated by the 5' base using numbering in GenBank entry X67204 and the \textit{M. musculus} ORF is between \~8304 and 9491. PCR was by conventional techniques using \\textit{Taq} DNA polymerase (PERRIN-ELMER) and 1.5 mm MgCl\textsubscript{2} employing 35 cycles of 94° for 30 sec, 59° for 30 sec and 72° for 90 sec. A sample of each PCR was assayed for specificity on a 1% agarose gel (the expected PCR product size is \~1.6 kb) and the remainder purified for sequencing using either Wizard PCR Prep (PRIMEGA) or QIAGen spin columns (QIAGEN). Genomic DNA from at least two individual males from each strain/species served as PCR template and the reactions were combined before purification and sequencing to control for individual mouse variations or mutation. No sequence heterozygosity was detected in any of the templates analyzed.

\textbf{Sequencing and analysis of \textit{Sry} ORF:} The purified \textit{Sry} ORF PCR product was directly sequenced using \textit{Sry}-8212, \textit{Sry}-9791, \textit{Sry}-8653 (5'-GGAGTAGAGCAGCTGACAGCTGGTACTC-3') and \textit{Sry}-3475 (5'-CCACGTGCTAGACTGCCAACC-3') as prim-
RESULTS

The *M. musculus* Sry gene encodes a ~395-amino-acid protein that is here divided into five regions for descriptive purposes (Figure 1). The first region is comprised of the first two amino acids (Met-Glu) and is referred to as unique sequence region-1 (UR-1). This region was invariant among all the *Sry* alleles sequenced. The second region is the 80-amino-acid HMG domain. The third region is a unique 61-amino-acid unit (unique sequence region-2, UR-2) that precedes the fourth region, which consists of highly repetitive polyglutamine repeat clusters (GRC region). The fifth region is a unique 29-amino-acid C-terminal unit (unique sequence region-3, UR-3). In contrast, the *M. domesticus* Sry gene encodes a protein that is about half as large (~250 amino acids) because there is a “premature” TAG translation termination codon in the eighth GRC (GRC-8) (Coward et al. 1994). Thus, part of the GRC region and UR-3 are absent from the putative *M. domesticus* SRY protein. For convenience, the GRCs are numbered consecutively from 1 to 22 although no single *Sry* allele contains more than 21 GRCs. This arrangement results from the alignment presented here because gaps were introduced when no clear GRC alignment was evident (Figure 2).

**Sry ORF polymorphisms:** *M. domesticus:* The ORF from 11 *M. domesticus* *Sry* alleles was sequenced and conceptually translated, and seven structural alleles were identified. These results are presented in Figure 2 and summarized in Table 1. All *M. domesticus* *Sry* alleles analyzed contain seven clusters of polyglutamine repeats preceded by a Phe-His-Asp-His (or similar) motif plus one polyglutamine repeat lacking this motif (GRC-1). The protein sequence of all 11 alleles is identical except for two sites of polymorphism: (1) GRC-1 contains nine glutamine residues in all alleles except *Sry*<sup>135</sup>, which contains 10, and (2) GRC-3 contains a variable number of glutamine residues. GRC-3 was proposed to play a role in B6-Y<sup>POS</sup> sex reversal (Coward et al. 1994). Among the alleles analyzed, the number of glutamine residues in GRC-3 varies from 11 to 15; *Sry*<sup>POS</sup> and *Sry*<sup>TR</sup> possess 11; *Sry*<sup>CD</sup>, *Sry*<sup>RB</sup> and *Sry*<sup>TR</sup> possess 12; *Sry*<sup>APP</sup>, *Sry*<sup>AP</sup> and *Sry*<sup>R</sup> possess 13; *Sry*<sup>DP</sup> and *Sry*<sup>R</sup> possess 14; and *Sry*<sup>RB</sup> possesses 15.

Nine of the 11 *M. domesticus* *Y* chromosomes were tested for their ability to cause sex reversal on the B6 genetic background (Table 1). Six produced XY sex reversed mice. It is clear from the data presented in Table 1 that the number of glutamine residues in GRC-3 does not correlate with sex reversal. For example, *Sry*<sup>APP</sup>, *Sry*<sup>AP</sup> and *Sry*<sup>RB</sup> have 13 glutamine residues in GRC-3 but only *Sry*<sup>APP</sup> causes sex reversal. Also, *Sry*<sup>TR</sup> and *Sry*<sup>RB</sup> contain 12 glutamine residues, but only *Sry*<sup>TR</sup> causes sex reversal. Furthermore, *M. domesticus* *Sry* alleles with 11, 12, 13, 14 or 15 glutamine residues in GRC-3 can cause sex reversal on a B6 genetic background.

*M. musculus:* The ORF from seven *M. musculus* *Sry* alleles was sequenced and conceptually translated, and five structural alleles were identified (Figure 2 and Table 1). The *M. musculus molossinus*-derived alleles, *Sry*<sup>135</sup> and *Sry*<sup>AP</sup>, were identical at the nucleotide level across the region sequenced, as were two of the European wild-derived alleles, *Sry*<sup>WRB</sup> and *Sry*<sup>DP</sup>. In contrast to *M. domesticus* *Sry* alleles, *M. musculus* *Sry* alleles con-
FIGURE 2.—Amino acid sequence comparison of the predicted SRY protein for the Sry alleles analyzed in this study. The 29 alleles are grouped into 15 protein structural variants as indicated: seven from *M. domesticus*, five from *M. musculus*, two from *M. spretus* and one from *M. spicigus*. Sry<sup>M. domesticus</sup> alleles that cause sex reversal on a B6 genetic background are indicated by an asterisk (*). The *M. domesticus* stop codon in GRW is also indicated by an asterisk (*) and an arrowhead. For comparison, the predicted *M. domesticus* protein sequence is continued past the stop codon. The HMG domain is indicated by a line above the protein sequence and the GRCs are numbered sequentially beginning with the Phe residue at the start of each cluster (except for GRG1 where this motif is absent). Amino acid identities are shaded, and gaps are indicated by dashes (-) with amino acid residue numbers indicated at the right.

Many of the remaining polyglutamine repeat clusters (GRC-4 to GRC-22) also are quite polymorphic, and the variability is contained in the glutamine and histidine containing regions. Most of the amino acid substitutions are the result of single nucleotide changes or repeat expansions/contractions. The GRC region is flanked by unique sequence regions with the HMG domain lying adjacent to UR-1 (Figure 1). There is one amino acid polymorphism in the HMG domain that is peculiar to the *M. m. molossinus* alleles, an Ile for Thr substitution. There also is one amino acid polymorphism in UR-3 that is specific to five of the seven *M. musculus* Y chromosomes were tested for sex reversal on the B6 genetic background (Table 1). Interestingly, none provided any evidence of sex reversal.

*M. spretus* and *M. spicigus*: To further delineate the conserved (and therefore potentially critical) SRY regions, we included Sry alleles from two populations each of *M. spicigus* and *M. spretus* in the analysis. (These Y chromosomes were not analyzed for sex reversal be-
cause interspecific male F₁ hybrids between these species and B6 are sterile.) The two *M. spicilegus* alleles were identical at the nucleotide level. The two *M. spreitus* alleles only differed at a single nucleotide that results in a Lys-Arg difference in GRC15. *M. spicilegus*, *M. spreitus*, and the progenitor of *M. musculus* and *M. domesticus* diverged from a common ancestor about three to five million years ago (FERRIS et al. 1983). *M. musculus* and *M. domesticus* subsequently diverged from each other one to two million years ago. The *M. spicilegus* and *M. spreitus* Sry alleles are more dissimilar to each other and to either *M. musculus* or *M. domesticus* than any *M. musculus* or *M. domesticus* allele is to any other member of its respective species (Figure 2). Among the four *Mus* species analyzed, the predicted proteins range from 230 to 234 amino acids (aa) for *M. domesticus*, 395–420 aa for *M. musculus*, 355 aa for *M. spreitus*, and 311 aa for *M. spicilegus*.

It is particularly striking that the HMG domain and three unique sequence regions are well conserved among these four *Mus* species. For example, the HMG domain contains one difference, the *M. m. molossinus* specific Ile for Thr substitution. UR-2 displays three differences: one is shared between *M. spicilegus* and *M. spreitus* (an Arg for Ala substitution at amino acid 92) and the other two are specific to the *M. musculus* alleles (a Leu for Trp substitution at amino acid 133) and a Pro for Leu substitution at amino acid 143. UR-3 displays a

**Figure 2.**—Continued
single difference that is shared: a Lys for Thr substitution at amino acid 304 in *M. spicilegus*, which is amino acid 348 in *M. spretus*. A second UR-3 difference is specific to *M. m. castaneus* and is detailed above.

The GRC region is highly divergent between *M. spicilegus* and *M. spretus* and when compared to *M. musculus* and *M. domesticus*. *M. spicilegus* contains 14 GRCs while *M. spretus* contains 18 GRCs. Furthermore, based on the alignment presented, *M. spicilegus* has two histidine and seven glutamine residues in GRC-3 while *M. spretus* does not contain GRC-3. In fact, many of the GRCs were difficult to align. Additionally, some of the amino acid differences represent multiple nucleotide changes within a given codon and therefore represent changes that are more divergent than the single nucleotide changes present in the *M. musculus* alleles.

**Sry 3’ UTR polymorphisms: M. domesticus:** The sequenced region for the *M. domesticus* Sry alleles encompassed ~800 bp of 3’ untranslated region (3’ UTR), including what are GRC-9 to GRC-22 in Figure 2. If this region were translated, it would encode 21 GRCs and would be more similar to the European wild-derived *M. musculus* alleles than to the *M. m. molossinus* alleles. Unexpectedly, there is very little polymorphism in the 3’ UTR as just three polymorphic sites were identified. There is an additional His residue (CAG) in GRC-12 of the 3’ UTR specific to Sry<sup>POS</sup> and Sry<sup>TR</sup>. One polymorphism is shared by Sry<sup>POS</sup>, Sry<sup>TR</sup> and Sry<sup>CD1</sup> and is a G for C substitution in GRC-21 that changes a His to a Glu residue. The third polymorphism is specific to Sry<sup>CD1</sup> and is a T for A substitution at position 9727 (Table 2). The *M. musculus* stop codon is conserved (in frame if the *M. domesticus* stop codon is ignored) in all *M. domesticus* alleles analyzed.

It should be noted that the sequence we obtained for Sry<sup>TR</sup> differs from previously published results and from what has been deposited in GenBank (accession number U03645). GenBank entry U03645 contains an additional CAC repeat at position 1622 (numbering according to COWARD et al. 1994) compared to that previously published by COWARD and the sequence we obtained. Additionally, our allele has two differences:

---

**TABLE 1**

*Description of Mus strains utilized including salient features of the Sry allele sequenced from each strain*

<table>
<thead>
<tr>
<th>Taxa</th>
<th>Strain or stock designation&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Collecting locality or inbred strain origin for Y chromosome&lt;sup&gt;b&lt;/sup&gt;</th>
<th>His/Glu number&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Repeat number&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Sex reversal&lt;sup&gt;f&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. domesticus</em></td>
<td>B6JE-Y&lt;sup&gt;POS&lt;/sup&gt;</td>
<td>Zalende, Val Poschiavo, Switzerland</td>
<td>1/11</td>
<td>8</td>
<td>Yes&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>B6JE-Y&lt;sup&gt;TR&lt;/sup&gt;</td>
<td>Tirano, Val Poschiavo, Italy</td>
<td>1/11</td>
<td>8</td>
<td>Yes&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>B6JE-Y&lt;sup&gt;SAM&lt;/sup&gt;</td>
<td>Kammerberg, Bavaria, Germany</td>
<td>1/12</td>
<td>8</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>B6JE-Y&lt;sup&gt;APP&lt;/sup&gt;</td>
<td>Appercha, Bavaria, Germany</td>
<td>1/13</td>
<td>8</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>B6JE-Y&lt;sup&gt;LP&lt;/sup&gt;</td>
<td>Lipari, Isola Eolie, Italy</td>
<td>1/14</td>
<td>8</td>
<td>Yes&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>B6JE-Y&lt;sup&gt;ORB&lt;/sup&gt;</td>
<td>Orobie, Bergamo, Italy</td>
<td>1/15</td>
<td>8</td>
<td>Yes&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>B6JE-Y&lt;sup&gt;UBB&lt;/sup&gt;</td>
<td>BUB/BNJ inbred strain</td>
<td>1/12</td>
<td>8</td>
<td>No&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>B6JE-Y&lt;sup&gt;ARR&lt;/sup&gt;</td>
<td>AKR/J inbred strain</td>
<td>1/13</td>
<td>8</td>
<td>No&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>B6JE-Y&lt;sup&gt;WSB&lt;/sup&gt;</td>
<td>Centreville, Maryland, U.S.A.</td>
<td>1/13</td>
<td>8</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>WMP/Ei</td>
<td>Monastir, Tunisia</td>
<td>1/14</td>
<td>8</td>
<td>NT</td>
</tr>
<tr>
<td></td>
<td>CD-I</td>
<td>CD-1 random bred strain</td>
<td>1/12</td>
<td>8</td>
<td>NT</td>
</tr>
<tr>
<td><em>M. musculus</em></td>
<td>B6JE-Y&lt;sup&gt;ACH&lt;/sup&gt;</td>
<td>Achering, Bavaria, Germany</td>
<td>3/21</td>
<td>21</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>B6JE-Y&lt;sup&gt;VRB&lt;/sup&gt;</td>
<td>Vejurnbro, Viborg, Denmark</td>
<td>4/17</td>
<td>21</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>B6JEi</td>
<td>C57BL/6J inbred strain subline</td>
<td>2/12</td>
<td>20</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>B6JEi&lt;sup&gt;129/J&lt;/sup&gt;</td>
<td>129/J inbred strain</td>
<td>2/12</td>
<td>20</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>B6JE-Y&lt;sup&gt;CZECH/El&lt;/sup&gt;</td>
<td>Bratislava, Slovakia, Czechoslovakia</td>
<td>3/17</td>
<td>21</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>SKIVE/Ei</td>
<td>Skive, Denmark</td>
<td>4/17</td>
<td>21</td>
<td>NT</td>
</tr>
<tr>
<td></td>
<td>CAST/Ei</td>
<td>Chonburi, Chonburi Prov., Thailand</td>
<td>4/12</td>
<td>20</td>
<td>NT</td>
</tr>
<tr>
<td><em>M. spicerus</em></td>
<td>PANCEVO/Ei</td>
<td>Pancevo, Serbia, Yugoslavia</td>
<td>2/7</td>
<td>14</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Halb</td>
<td>Halbturn, Burgenland Prov., Austria</td>
<td>2/7</td>
<td>14</td>
<td>NA</td>
</tr>
<tr>
<td><em>M. spretus</em></td>
<td>SPRET/Ei</td>
<td>Puerto Real, Cadiz Province, Spain</td>
<td>deleted</td>
<td>18</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Azrou</td>
<td>Azrou, Morocco</td>
<td>deleted</td>
<td>18</td>
<td>NA</td>
</tr>
</tbody>
</table>

<sup>a</sup>Strain or stock designation denotes whether the Y chromosome, and hence Sry allele sequenced, was from an inbred strain (e.g., WMP/Ei) or a Y chromosome congenic strain (e.g., B6JEi-Y<sup>POS</sup>). The Halbturn and Azrou stocks were at F<sub>7</sub> and F<sub>10</sub>, respectively.

<sup>b</sup>Localities listed represent the original collecting location of wild-caught male mice that were the founders of the strains listed in a chromosome from conventional inbred strains also are included here and so designated.

<sup>c</sup>His/Glu number is the number of histidine and glutamine residues present in GRC-3.

<sup>d</sup>Repeat number is the total number of potentially translated GRC repeats.

<sup>f</sup>Sex reversal denotes whether sex reversal, as defined in MATERIALS AND METHODS, was observed in a given strain. NT, strains that were not tested; NA, strains that could not be tested because of hybrid sterility, not applicable.

<sup>i</sup>Sex reversal previously reported in EICHER et al. (1982).

<sup>j</sup>Sex reversal previously reported in NAGAMINE et al. (1987) and confirmed here.

<sup>k</sup>Sex reversal previously reported in BIDDLE et al. (1991) and confirmed here.
TABLE 2

<table>
<thead>
<tr>
<th>Nucleotide position</th>
<th>Canonical nucleotide</th>
<th>Nucleotide variation (allele)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9522</td>
<td>C</td>
<td>A (M. spicilegus)</td>
</tr>
<tr>
<td>9526</td>
<td>C</td>
<td>T (M. spicilegus)</td>
</tr>
<tr>
<td>9581</td>
<td>G</td>
<td>A (M. spicilegus)</td>
</tr>
<tr>
<td>9581</td>
<td>C</td>
<td>A (Sry(^{1990}) and Sry(^{1993}))</td>
</tr>
<tr>
<td>9582</td>
<td>T</td>
<td>G (M. spicilegus)</td>
</tr>
<tr>
<td>9610–9611</td>
<td>CC</td>
<td>TT (M. spretus)</td>
</tr>
<tr>
<td>9622</td>
<td>C</td>
<td>T (M. musculus)</td>
</tr>
<tr>
<td>9648–9655</td>
<td>TGTAAGGA</td>
<td>Deletion (M. spretus)</td>
</tr>
<tr>
<td>9727</td>
<td>A</td>
<td>T (M. m. castaneus)</td>
</tr>
</tbody>
</table>

Numbering according to position in GenBank entry X67204 from the 129 inbred strain.

One at position 2032, where our allele does not have a T, and one at position 2086, where our allele has an additional T. While it is possible that these differences are peculiar to the individual mice used for sequencing, the Sry\(^{1990}\) sequence we obtained (for these three positions) is shared by all M. domesticus alleles we have analyzed, thus making it likely that the sequence we present here is correct.

M. musculus: The sequenced region for the M. musculus Sry alleles encompassed ~295 bp of 3' UTR because the M. domesticus ORF is substantially longer than the M. domesticus ORF and, therefore, the 3' UTR begins further downstream in the sequenced region. A single polymorphism was observed in this region and is an A for C substitution at position 9581 shared by the Sry\(^{1990}\) and Sry\(^{1993}\) alleles (Table 2).

M. spretus and M. spicilegus: The sequenced region contains 279 bp of 3' UTR for M. spretus and 294 bp for M. spicilegus. The 3' UTR is more divergent in the M. spretus and M. spicilegus alleles than in the M. domesticus or M. musculus alleles, when compared to the Sry\(^{1990}\) allele. There are nine regions of variation including one 8-bp deletion (Table 2). Four differences are single nucleotide substitutions unique to the M. spicilegus allele: A for C at position 9522, T for C at position 9526, A for G at position 9531 and G for T at position 9582. One difference is a 2-bp change unique to M. spretus: TT for CC at positions 9610–9611. One difference is unique to M. musculus: T for C at position 9622, and one difference is unique to M. m. castaneus: T for A at position 9727. Additionally, the M. spretus allele contains an 8-bp deletion for the sequence TGTAAGGA at positions 9648–9655.

Sry 5' UTR polymorphisms: Approximately 55 bp of sequence immediately upstream of the putative initiation codon was obtained for each Sry allele. No substitutions within or between species were detected in this interval. The significance of this finding is unclear given the short length of the interval and the absence of significant DNA motifs.

DISCUSSION

B6-Ypos sex reversal is not explained by protein isoform variation: The primary goal of this comparative sequence analysis was to gain insight into the molecular nature of Sry dysfunction in B6-Ypos sex reversal by attempting to correlate a particular SRY protein isoform with the ability to produce ovarian tissue in XY mice. Although GRG-3 variability was the only coding region polymorphism found among 10 of the 11 M. domesticus alleles analyzed, we found that the number of glutamines in GRG-3 did not correlate with the presence or absence of sex reversal. A second polymorphic site was unique to Sry\(^{1990}\), which contained 10, instead of nine, glutamine residues in GRG-1. Although Sry\(^{1990}\) causes sex reversal when present on the B6 background, some alleles with nine glutamines in GRG-1 also do so. In fact, there was no correlation between any particular SRY protein isoform and sex reversal. These results strongly suggest that the cause of B6-Ypos sex reversal is not solely the structure of the Sry protein, but rather the regulation of SRY expression. (It is formally possible that a combination of protein structure and regulation causes B6-Ypos sex reversal.)

This conclusion is in contrast to that reached by Coward et al. (1994); however, the discrepancy is likely to be attributable to the different samples examined. The Sry alleles analyzed here were primarily from wild-derived Y chromosomes, whereas those studied by Coward et al. were from inbred strains. Because many inbred strains share a Y chromosome of common recent origin, the actual number of different Sry alleles represented in inbred strains is few (Tucker et al. 1992a).

Not all M. domesticus Sry alleles cause sex reversal when present on the B6 genetic background. Some, such as Sry\(^{1990}\), cause a transient delay in testis cord differentiation (Washburn and Eicher 1988; Washburn and Eicher 1989; Washburn et al. 1990; Palmer and Burgoyne 1991; Burgoyne and Palmer 1993). Further study will be necessary to determine if there is any correlation between delayed testis cord differentiation and SRY protein structure. However, existing data do allow us to draw a few conclusions. All M. domesticus Sry alleles, for which complete sequence data is available, contain a premature (relative to other Mus species) stop codon. The stop codon per se cannot be responsible for sex reversal because not all M. domesticus alleles cause sex reversal. Likewise, the premature stop codon is not likely to be responsible for delaying fetal testis cord differentiation because not all M. domesticus Sry alleles cause delay (Coward et al. 1994). It is possible, however, that other protein isoform differences (GRG-5 glutamine number variability) contribute to or cause this delay. We have suggested that delayed testis cord differentiation alone does not result in sex reversal, but may be the first step toward sex reversal (Eicher 1994; Eicher et al. 1995). If this hypothesis is true, then
additional Sry sequence differences must contribute to "push" the delay into complete sex reversal. The data now suggest that these sequence differences lie outside the ORF and therefore are regulatory. It is possible that both the delay in testis cord differentiation and sex reversal are caused by regulatory sequence differences and are different degrees of the same phenomenon.

We envision the following genetic scenario to explain the phenomenon of B6-Y\textsuperscript{POS} sex reversal. First, this sex reversal is genetically complex and arises from an aberrant interaction between certain \textit{M. domesticus} Sry alleles and autosomal genes of the B6 genome (\textit{tda} genes). Generally, \textit{M. domesticus} Sry alleles fall into three phenotypic groups when present on a B6 genetic background and can (1) allow ovarian tissue development, (2) cause a brief delay in testis cord differentiation, or (3) function as normal (wild-type) male determinants. Since C57BL/6J is an inbred strain, the autosomal component of the various B6-Y \textit{M. domesticus} consomic strains is identical (barring new mutation and negligible genetic drift), and therefore the phenotypic variability between different B6-Y \textit{M. domesticus} consomic strains must be due to differences in SRY expression. We previously proposed that differences in SRY expression cause the phenotypes manifest in B6-Y \textit{M. domesticus} consomic strains (Eicher 1994). It may be possible to delineate the regulatory sequence differences responsible for the misexpression of SRY by further sequence analysis of various Sry\textsuperscript{M. domesticus} alleles.

SRY misregulation as the cause of B6-Y\textsuperscript{POS} sex reversal is further supported by the \textit{M. musculus} ORF sequence analysis. Four \textit{M. musculus} Sry alleles (other than Sry\textsuperscript{B6}) were tested for sex reversal when present on the B6 genetic background (Table 1). Sry\textsuperscript{ACH}, Sry\textsuperscript{IB}, Sry\textsuperscript{IL} and Sry\textsuperscript{CZECH} have quite different GRC-2 amino acid compositions (three histidine and 21 glutamine residues, four histidine and 17 glutamine residues, two histidine and 12 glutamine residues and three histidine and 17 glutamine residues, respectively, see Figure 2). Coward et al. proposed that 12 glutamines in GRC-2 was the normal state and that deviations from this number affected SRY function. Strikingly, given the degree of polymorphism displayed, none of the four \textit{M. musculus} alleles caused sex reversal. However, these \textit{M. musculus} Sry alleles have not been examined for delayed testis cord differentiation.

The conclusion that B6-Y\textsuperscript{POS} sex reversal is due to a difference in SRY expression and not SRY structure does not allow us to suggest where the \textit{tda} genes function within the mammalian sex determination genetic hierarchy (i.e., if the \textit{tda} genes regulate Sry, are regulated by Sry, or act in parallel with Sry). For example, if the cause of B6-Y\textsuperscript{POS} sex reversal was attributable to SRY structure, one would predict that the \textit{tda} genes were acted upon by SRY, and hence downstream in the sex determination pathway.

The lack of SRY protein sequence variability is discordant with the range of variability in biological function of the parental \textit{Y} chromosomes on the B6 background. The idea that many phenotypic differences result from changes in the developmental expression of genes (regulation) rather than changes in protein structure (amino acid divergence) is not novel. Thirty years ago, it was proposed that "two organisms may be phenotypically more (or less) different than they are on the basis of the amino acid sequence of their polypeptide chains" (Zuckerkandl and Pauling 1965). In general, it has been variously hypothesized that much of evolution is caused by changes in gene expression rather than changes in protein structure (Wang et al. 1996).

**SRY structure, function, and evolution**: Our second goal was to gain insight into the structure, function, and evolution of the Sry gene in the subgenus \textit{Mus}. Comparative sequence analysis is based on the premise that highly conserved regions represent domains that are evolutionarily constrained and therefore functionally important and that nonconserved regions are either not evolutionarily constrained and thus represent functionally less important domains, or they represent domains that are subject to positive Darwinian selection. The ORF of 22 Sry alleles was sequenced, including alleles from \textit{M. domesticus}, \textit{M. musculus}, \textit{M. spicilegus} and \textit{M. spretus}. Among these four species, the HMG domain and unique sequence regions are well conserved, even among the most distantly related species, while the GRC region is quite variable. This result suggests that the SRY functional domain is contained, maximally, in the first 143 amino acids, which includes UR-1, the HMG domain and UR-2. Despite apparently strong conservation, UR-3 must be dispensable for SRY function because it is predicted to be absent in all \textit{M. domesticus} Sry proteins (which all function normally within their genome of origin and many function normally on a B6 background). It is possible that the GRC region is subject to directional evolution and that the high degree of polymorphism merely reflects this process. Furthermore, it is possible that the apparent conservation of UR-2 reflects the fairly close relatedness of the species analyzed.

It has been suggested that because only the HMG box is conserved between genera, it represents the SRY functional domain (Whitfield et al. 1993). Our sequence analysis supports this view but suggests that the region of conservation in the subgenus \textit{Mus} may extend to 143 amino acids, which includes UR-2. Among the Sry alleles from the four species analyzed, there are only four amino acid differences in the first 143 amino acids (Figure 2). Three of the differences are unique to the \textit{M. musculus} lineage and are essentially conserved (Ile for Thr at amino acid 63, Trp for Leu at amino acid 133 and Leu for Pro at amino acid 143). The Ile for Thr substitution is within helix III of the HMG domain but is at a site that is not conserved among HMG domain proteins. An Ile at this position is unique to the
M. m. molossinus lineage. The fourth amino acid difference is a nonconservative Arg for Ala change and differentiates the M. musculus/M. domesticus (Arg) from M. spicilegus/M. spretus (Ala).

Although most human SRY sex reversing mutations are located within the HMG box, recently a mutation in M. m. molossinus located within the HMG box, recently a mutation differentiates the 3' to the HMG box was identified in two sisters with 46, XY complete gonadal dysgenesis (Kwok et al. 1996). This mutation changes a Leu residue to a premature stop codon and would eliminate the amino acids. This finding suggests that the region 3' to this premature stop codon is necessary for human SRY function. Perhaps non-HMG box regions are important for species-specific SRY functions.

DNA constructs containing either the entire M. m. molossinus Sry coding unit or only the M. m. molossinus GRC region alone are capable of transcriptional transactivation in vitro (Dubin and Ostrer 1994). On the other hand, the complete M. domesticus coding unit and complete human SRY coding unit are incapable of transcriptional transactivation in similar assays and were actually specific transcriptional repressors. However, a M. m. molossinus Sry construct coding for the first 137 amino acids behaved nearly identically to the complete human gene and the complete M. domesticus gene. We suggest that if the mouse SRY functional domain is contained in the first 143 amino acids, both M. musculus and M. domesticus Sry are identical and are, in fact, functionally similar to human SRY. This hypothesis argues that Sry/SRY is incapable of transcriptional transactivation and that Sry/SRY is primarily a sequence-specific DNA binding protein that bends DNA and physically acts as a transcriptional switch. This idea is compatible with the theory that HMG domain proteins are architectural elements (Grosschedl et al. 1994). A further extension of this thought is that Sry/SRY probably requires other context dependent transcription co-factors to execute its functions (be it activation or repression). The tda genes (and others) are candidates for these transcription cofactors.

We found little within-species polymorphism among the 11 M. domesticus Sry alleles analyzed. There were only two polymorphic sites and both involved differences in the number of glutamine residues within the GRC region. The first polymorphic site was specific to the SrySP allele and was discussed above. Additionally, Miller and colleagues (1995) showed that the number of glutamine residues in GRC-I varied from nine to 11 in the M. domesticus Sry alleles they sampled. The second polymorphic site in the M. domesticus Sry alleles is within GRC-3 where the number of glutamine residues varied between 11 and 15. Miller and colleagues (1995) showed that the glutamine residue number in GRC-3 varied from 13 to 20 in a different group of wild-derived M. domesticus Sry alleles.

Over all, the ORFs of the M. musculus Sry alleles displayed greater within species variation than the M. domestica s alleles. The number of GRCs was variable (either 20 or 21), as were most individual GRCs. Taken together, these data suggest that variation in the number of glutamine repeats is common and therefore any correlation with sex reversal is coincidental. The finding of greater within species variation among M. musculus Sry alleles compared to M. domesticus alleles is in contrast to data obtained from various Y chromosomes using the YB10 repeat as a probe (Tucker et al. 1989). This discrepancy may be attributable to a sampling phenomenon, or perhaps reflects differential mutation or fixation rates for different regions or types of loci on the Y chromosome.

The GRC region is highly divergent among the four Mus species examined. In this study, we identified Sry alleles with eight (M. domesticus), 14 (M. spicilegus), 18 (M. spretus), and 20 or 21 (M. musculus) GRCs. Also, almost every individual GRC displays differences relative to the other species. Even if one considers only the first seven GRCs, which represent the minimal GRC region of M. domesticus, there is still a high degree of divergence. These differences include deletions of certain GRCs in some species. For example, GRC-3 is not present in M. spretus and GRC-8 is not present in M. spicilegus. The GRC region is clearly not well conserved, suggesting that it is not functionally constrained.

The GRC repeat region is essentially a group of CAG/CAC trinucleotide repeat units and is the source of most of the sequence variation within Sry. Simple repetitive DNA tracts are unstable in all organisms examined to date and instability of trinucleotide repeats is a common hallmark for several human inherited diseases (Monckton and Caskey 1995). Many mechanisms have been proposed to explain repeat instability, including DNA polymerase slippage and unequal recombination (McMurray 1995). Unequal meiotic recombination is not likely to play a role in the instability of the Sry GRC region because this region of the mouse Y chromosome is outside of the shared X-Y homology and recombination region. However, aberrant sister chromatid exchange might explain the apparent inversion of GRC-14 and GRC-15 in the M. spretus alleles or may account for deletions/duplications of other GRCs.

The 3' UTRs (partially analyzed) of the four species are well conserved. The M. domesticus 3' UTR includes what would be GRC-9 to GRC-22 if the M. domesticus stop codon is ignored (Figure 2). It is somewhat surprising that this region is very well conserved within the M. domesticus species group (just two polymorphic sites in ~420 bp) when, by contrast, the corresponding GRC region is very polymorphic within the M. musculus species group. It is not immediately obvious why this region is highly conserved in M. domesticus, where it is not part of the protein coding region, and is not highly conserved in M. musculus, where it is part of the protein. However, the difference between M. musculus and M. domesticus may be primarily a result of one lineage being
older than the other and the older lineage being more polymorphic. These data suggest that the *M. domesticus* stop codon is an evolutionarily recent addition and that the additional 3' UTR segment has not yet diverged. The ~770 bp of *M. domesticus* 3' UTR that was sequenced contains just one additional polymorphism (within the *M. domesticus* species group), and the *M. musculus* stop codon is potentially conserved in frame. (We would point out that there is no direct in vivo evidence that the first *M. domesticus* stop codon is used for translational termination.) A further sampling of *M. domesticus* Sry alleles might uncover additional polymorphism similar to that seen in *M. musculus* Sry alleles.

The sequenced region of the 3' UTR downstream from the *M. musculus* stop codon (bp 9482–9770) also is fairly well conserved among the four species. *M. spicilegus* and *M. spreatus* are more divergent than *M. musculus* and *M. domesticus* when compared to each other or to the other two species. This result is consistent with their evolutionary histories. The major difference is an 8-bp deletion in *M. spreatus*. Since the sequenced region contains only a small proportion (~300/3465 bp) of the very large Sry 3' UTR, it is difficult to suggest structure/function relationships from the data. The only identified feature that is contained in the 3' UTR sequenced are four ATTTA putative mRNA instability elements (Hacker et al. 1995). The sequenced region includes the first three RNA instability elements that are conserved in all 22 alleles except *M. m. castaneous*. The first element is modified from ATTTA to ATITTAA in Sry CAST and this may not represent a significant variation for this type of element. The conservation of these four sites argues that they are a functionally important component of the Sry 3' UTR.

The Mus species examined here can be broadly classified by three elements of their Sry ORF sequences: (1) the presence of either an Ile or Thr residue at position 65 of the HMG domain, (2) the number of glutamine repeat clusters (GRCs), and (3) the presence or absence of a stop codon in GRC-8 (Table 1). It is important to add the caveat that we analyzed only two alleles each from *M. spicilegus* and *M. spreatus* and thus these alleles may not be representative of each respective species. By these criteria, *M. musculus* can be further subdivided into three types: *M. m. musculus* from Europe, *M. m. musculus* from Japan and *M. m. castaneous*.

Since the presence of a stop codon in GRC-8 is unique to *M. domesticus*, we propose that the stop codon present in the *M. domesticus* GRC-8 be hereafter referred to more properly as the “*M. domesticus* stop codon” instead of the “premature stop codon.” We cannot unequivocally discern if the addition of a stop codon to GRC-8 in the common ancestor of *M. m. musculus* and *M. domesticus* was a contributing factor to or a result of this species divergence, but the biological data outlined above argues that the presence of a *M. domesticus* stop codon per se does not cause either sex reversal or delayed testis cord differentiation when *M. domesticus* Y chromosomes are introgressed into a B6 genetic background. It is important to point out that the mutation of a CAG codon to a TAG stop codon represents a CG to TA transition, which is a frequent source of variation in many genes.

The data presented suggest a number of interesting and testable hypotheses. Since the cause of SRY misfunction cannot be attributed to protein structure, SRY's regulation must be aberrant. Further sequencing of the 5' and 3' regulatory regions may uncover the responsible sequence elements. Along this line, a transgene containing a *M. domesticus* protein (from a sex reversing Sry allele) driven by *M. musculus* regulatory regions would be predicted to behave like a *M. musculus* Sry allele. Also, if the CAG repeat region and UR-3 are dispensable for normal SRY function, then a transgene containing only the first 143 amino acids (UR-1, the HMG box, and UR-2) should function similar to a transgene containing the complete SRY protein. These are just a few of the experiments suggested by the above data.

We thank Linda Washburn and Leonia Gagnon for establishing and maintaining the Y chromosome consomic mouse strains, and Richard Sage for supplying some of the progenitor wild-caught males, including those carrying the *Yach*, *Yapr*, and *Ykom* chromosomes. We thank Douglas McMinnimy for his invaluable expert DNA sequence analysis. We also thank LuAnne Peters, Priscilla Tucker, Elizabeth Simpson and Linda Washburn for critically reviewing previous versions of this manuscript. This work was supported by National Institutes of Health (NIH) grant GM-20919 to E.M.E. and Core grant CA34196 to The Jackson Laboratory. K.H.A. was additionally supported by NIH fellowship GM-16726 and training grant HD-07065.

**LITERATURE CITED**

Biddle, F. G., B. A. Eales and Y. Nishikoa, 1991 A DNA polymorphism from five inbred strains of the mouse identifies a functional class of domestic-type Y chromosome that produces the same phenotypic distribution of gonadal hermaphrodites. Genomics 34: 96–104.


Carlisle, C., H. Winkler, D. Weichenhan and C. M. Nagamine, 1996 Absence of correlation between *Sry* polymorphisms and XY sex reversal caused by the *M. m. domesticus* Y chromosome. Genomics 38: 32–45.


Communicating editor: R. E. GANSCHOW