

## Evidence for Multiple Functional Copies of the Male Sex-Determining Locus, *Sry*, in African Murine Rodents

Barbara L. Lundrigan, Priscilla K. Tucker

Museum of Zoology and Department of Biology, University of Michigan, Ann Arbor, MI 48109-1079, USA

Received: 4 October 1996 / Accepted: 17 January 1997

**Abstract.** Southern hybridization data suggest that the male sex-determining locus, *Sry*, is often duplicated in rodents. Here we explore DNA sequence evolution of orthologous and paralogous copies of *Sry* isolated from six species of African murines. PCR amplification followed by direct sequencing revealed from two to four copies of *Sry* per species. All copies include a long open reading frame, with a stop codon that coincides closely with the stop codon of the house mouse, *Mus musculus*, a species known to have a single copy of *Sry*. A phylogenetic analysis suggests that there are at least seven paralogous copies of *Sry* in this group of rodents. Putative orthologues are identical; sequence divergence among putative paralogues ranges from 1 to 8% (excluding the CAG repeat), with much lower levels of divergence in the high-mobility group (HMG-box) region than in the C-terminal region. A high proportion of nucleotide substitutions in both regions result in amino acid replacement. The long open reading frame, conserved HMG-box, and pattern of evolution of the putative paralogues suggest that they are functional.

**Key words:** *Sry* — HMG-box — Gene duplication — Sex determination — *Arvicanthis* — *Lemniscomys* — *Pelomys* — *Rhabdomys*

### Introduction

In mammals, male sex determination is defined by the initiation of testis development. There is now consider-

able evidence that *Sry* (sex-determining region *Y* chromosome) is the testis-determining gene (Gubbay et al. 1990; Sinclair et al. 1990). With rare exception (see, e.g., Bianchi et al. 1993), *Sry* is male-specific in all mammal species examined thus far, a sample that includes representatives from three metatherian and seven eutherian orders (Sinclair et al. 1990; Foster et al. 1992; Pallsboll et al. 1992; Bianchi and Bianchi 1993). It is expressed in the genital ridge at a time consistent with its having a role in testis determination (Gubbay et al. 1990; Koopman et al. 1991b), and a 14.5-kb mouse genomic DNA fragment containing *Sry* can induce testis development when injected into (chromosomally female) XX mouse embryos soon after fertilization (Koopman et al. 1991a).

*Sry* consists of a single exon with a conserved amino acid DNA-binding domain (high-mobility group or HMG-box) of approximately 79 base pairs (bp) surrounded by N-terminal and C-terminal regions of variable length and composition. The *Sry* protein binds to specific target DNA sequences, inducing a sharp bend (Nasrin et al. 1991; Ferrari et al. 1992) that is thought to influence the transcription of (still unidentified) genes involved in the differentiation of Sertoli cells. Binding activity resides in the HMG-box and mutations in this region have been associated with sex reversal in humans (reviewed in Hawkins 1993).

The importance of the N-terminal and C-terminal regions to the function of *Sry* is not known. These non-HMG-box regions differ markedly between species and can be aligned only among closely related taxa. Comparative studies within primates (Whitfield et al. 1993) and rodents (Tucker and Lundrigan 1993) suggest a rapid

rate of amino-acid sequence evolution in the nonbox regions, indicating that they are either under directional selection or are not functionally constrained.

A single copy of *Sry* occurs in humans and the house mouse, *Mus musculus*, but Southern hybridization studies of both New World sigmodontine rodents (Bianchi et al. 1993) and Old World murine rodents (Nagamine 1994) suggest that *Sry* sometimes occurs in two or more copies. Here we explore DNA sequence evolution of orthologous and paralogous copies of *Sry* isolated from six species of African murine rodents.

## Materials and Methods

**DNA Samples and Sample Preparation.** The Arvicanthis division is an African radiation of herbivorous rodents (family Muridae, subfamily Murinae) that includes 14 genera united primarily on the basis of tooth morphology (Misonne 1969). The phylogenetic relationships among the species in this group are not well understood. We sampled two males and two females from each of six species from the Arvicanthis division: *Arvicanthis nairobae*, *Lemniscomys barbarus*, *L. massaicus*, *L. rosalia*, *Pelomys fallax*, and *Rhabdomys pumilio*. Tissues were obtained from field-collected specimens kindly donated by W. N. Verheyen. Genomic DNA was isolated from frozen tissue following the method of Jenkins et al. (1982).

**Polymerase Chain Reaction (PCR).** Various combinations of forward (5' to 3': AGATCTTGATTTTTAGTGTTTC, TAGAGAGCATGGAGGGCCAT, TCCTACACAGAGAGAAATACC, GTCTGCTACAGTGGGACATAA, CTCCAAGGCCAGCAAAGCTT, ACCAGCAGAGCAAAGGACCAT) and reverse (5' to 3': TGTA-CAGTTT (T/C) GTTGAGGC, CACCAGTGATGTCAGCTGT-TAGTA, GTATCTAAACCACTAGGAATCTC, TGTCTGATGT-GATGGTGTGG, GAGGACAGGTGTGCAGCTCTA, AAGCTTTGCTGGGCCTTGGAG, ATGGTCCTTTGCTTTGC-TGGT, CAAGGAGGAGTATTGGCATAGGTA) primers were used to PCR amplify overlapping segments of *Sry* beginning 76 bp 5' to the open reading frame (ORF) and extending slightly beyond the putative stop codon of each sequence, for a total of from 658 to 685 bp. The first three forward primers and the first three reverse primers were designed from the published *Mus musculus Sry* sequence (Gubbay et al. 1992) and are highly conserved across Old World rodent species. The other primers were designed to match *Sry* of the species being investigated here, and once multiple copies were discovered, to target particular copies of the gene. The amplification protocol included 30 cycles, with denaturation at 95°C for 1 min, annealing at from 50 to 65°C for 1 min, and elongation at 72°C for 1 min 15 s. Because the DNAs were degraded, we were unable to explore the number of multiple copies of *Sry* using Southern hybridization.

To verify that amplified sequences were located on the Y chromosome, amplification experiments included both male and female DNA from each species under investigation. Amplification of the target sequences in males only was taken as evidence that the sequences were Y-specific. PCR products were visualized on a 2% NuSieve gel. If multiple bands were observed, they were cut from the gel and gene cleaned using the GeneClean II kit (Bio 101). Samples for which only a single band was observed were Centricron cleaned using Centricron 100 tubes (Amicon).

**Sequencing and Sequence Alignment.** DNA was sequenced using the Sanger dideoxy-sequencing technique (Sanger et al. 1977) and the Cycle Sequencing System (Gibco BRL). Because PCR products sometimes contained more than a single copy of *Sry*, each PCR product was

sequenced several times using several different sequencing primers. Many of our sequences are incomplete, particularly in the 5' region where the copies are very similar, and it was thus difficult to target and definitively identify particular copies of the gene. Sequences were aligned by eye with the published *Mus musculus Sry* sequence (Gubbay et al. 1992). The 3' end of the *Sry* sequences includes an imperfect CAG repeat of variable length, which could not be unambiguously aligned; we excluded that region from the phylogenetic analysis.

**Phylogenetic Analysis.** Aligned sequences were subjected to a parsimony analysis using PAUP (version 3.1.1, Swofford 1993). Characters were unordered (Fitch optimization) and uniformly weighted. We performed a single analysis, using the branch-and-bound option, with the *Mus musculus* sequence as an outgroup.

A decay index (Kallersjo et al. 1992) was calculated for each clade. The decay index of a given clade is the number of additional steps required to generate a nonminimal-length tree in which that clade has been dissolved. MacClade (ver. 3.01, Maddison and Maddison 1992) was used to trace character changes and document support for individual branches.

## Results

### Identification of Multiple Copies of *Sry*

We identified from two to four copies of *Sry* per species, for a total of 17 sequences. All sequences were male-specific, i.e., no amplification resulted when female DNA was used as a template. An alignment of the variable sites, excluding the CAG repeat region, is presented in Fig. 1.

### Parsimony Analysis

To distinguish orthologous copies of *Sry* (those resulting from speciation) from paralogous copies (those resulting from gene duplication), the aligned sequences in Fig. 1 were subjected to a parsimony analysis. This analysis resulted in 48 most parsimonious trees with a length of 40 and a consistency index of 0.975. A strict consensus is shown in Fig. 2.

There is only one homoplasious character on this tree, a nucleotide site with a character state that is shared by the outgroup and the lineage labeled copy 7. The large number of most parsimonious trees and the low decay index for most clades reflect the large amount of missing data. Missing data also tend to inflate the consistency index, since for missing entries, PAUP assigns the character state that would be most parsimonious given the tree topology.

The consensus tree suggests that there are at least seven paralogous copies of *Sry* in this group of rodents (copy 1–7 in Fig. 2). Each is diagnosed by three or more unique and unreversed synapomorphies. For copies 1, 2, and 6, some of these diagnostic characters are present in only a single terminal taxon and have been assumed for the other members of the clade, taxa for which the relevant sites are missing. However, even when those characters are not counted as support, copies 1, 2, and 6 are



ter in the parsimony analysis. All of the species examined had at least one copy of *Sry* with this indel, and one copy without it, suggesting that at least one duplication event preceded the speciation event giving rise to this group of rodents.

#### *Pattern of Evolution of Putative Paralogues*

The phylogenetic analysis suggests considerable variation in the rate of nucleotide evolution for different copies of the gene. This cannot be the result of missing data, since there is at least one complete sequence within each terminal clade. An examination of the relative proportion of nonsynonymous to synonymous nucleotide substitutions across the tree revealed a high proportion of nonsynonymous substitutions in both the HMG-box (50%) and C-terminal regions (70%). This pattern was not restricted to particular lineages, as might be expected if some copies are free to diverge while others are functionally constrained.

An alignment of the ORFs from the seven putative paralogues of *Sry* is shown in Fig. 3. The HMG-box is conserved among these copies, with pairwise sequence divergence never greater than 3%. In contrast, the C-terminal region is evolving relatively rapidly. Pairwise sequence divergence reaches 14% (excluding the repeat) and insertions and deletions result in frameshift mutations that preclude alignment of amino acids beyond the middle of the region. Because of the frameshift mutations, there are two alternative stop codons among the seven copies—one in copies 1, 2, and 6 and another in copies 3, 4, 5, and 7. These are located 39 bp and 64 bp 5' to the putative stop codon of the *Mus musculus Sry* sequence, respectively.

The repeat region which makes up over half of the ORF of *Mus musculus Sry* (666 bp) is greatly truncated in all seven of the *Sry* copies identified here (range 45 to 75 bp). A shortened repeat region, with respect to the *Mus musculus* sequence, has been reported in other rodent genera (Tucker and Lundrigan 1993).

#### **Discussion**

The male sex-determining locus, *Sry*, has duplicated several times in this group of rodents. The duplicate copies are Y-specific, but it is not known whether they are tandemly arranged or dispersed on the chromosome. Duplicate copies of *Sry* have been detected in other rodent species, including both New World (Bianchi et al. 1993) and Old World (Nagamine 1994) murids. The high incidence of duplicate copies among the few rodent species that have been examined and the occurrence of duplicate copies among rodent species that are not closely related suggest that duplicate copies of *Sry* are common in rodents.

It is not possible to estimate the age of these gene duplications, as little is known about the phylogenetic

history of the rodents examined in this study. However, it is notable that putative orthologues are identical to one another. A comparative study of the evolution of putative orthologues of *Sry* in the genus *Mus* (excluding the repeat region) found up to 6% interspecific sequence divergence (Lundrigan and Tucker 1994). The repeat region has been surveyed in only one species, *Mus domesticus*, where it exhibited variation among populations (Miller et al. 1995). This suggests that the speciation event that gave rise to the rodent species examined in our study was relatively recent.

In contrast to the lack of variation among orthologues, paralogues have diverged considerably. The most likely explanation for this pattern is that the speciation event that gave rise to the group of rodents examined here was preceded by the gene duplications that gave rise to the seven paralogous copies of *Sry* isolated from these species. The fact that for each species there is at least one *Sry* copy with, and one without, the distinctive 12-bp indel is consistent with this hypothesis. However, we cannot eliminate the possibility that one or more of the gene duplications occurred after the origin of the species group and that paralogous copies have diverged more rapidly than orthologous copies. Gene duplication is usually followed by a period of accelerated amino-acid substitution, presumably the result of relaxed constraints on redundant copies or selection for their functional divergence (Li 1985; Ohta 1988, 1994).

Gene duplications have occurred frequently in the evolution of the mammalian genome. The location of *Sry* on the Y chromosome may make it particularly susceptible to duplication events. Unlike other chromosomes in the genome, the Y chromosome recombines only over a restricted region. Much of the chromosome, including the sex-determining region, is thus clonally inherited from father to son. This unique mode of inheritance has resulted in an unusually rapid rate of duplication, rearrangement, addition, and deletion (reviewed in Marshall Graves 1995). In *Mus musculus*, *Sry* lies within 2.8 kb of unique sequence at the center of a large inverted repeat (Gubbay et al. 1992), a possible relic of multiple copies of *Sry* in the *Mus* lineage. Two closely linked genes, zinc finger protein, *Zfy* (Page et al. 1987), and a gene of unknown function, *Ube1Y* (Mitchell et al. 1991), have also multiplied on the *Mus musculus* Y chromosome. *Zfy* is present in two copies in *Mus musculus* (Mardon et al. 1989; Nagamine et al. 1989) and has amplified up to 26 times on the Y chromosome in other rodents (Bianchi et al. 1992; Lau et al. 1992; Nagamine 1994). The two *Mus musculus Zfy* copies are expressed in the testis and do not appear to be functionally redundant (Koopman et al. 1989).

We do not know whether the *Sry* copies identified in this study represent functional genes or pseudogenes; the presence of transcripts has not been explored. However, the available evidence suggests that the copies are func-



different regions of the *Sry* copies, a pattern that would not be expected of pseudogenes.

But why should selection conserve redundant copies of *Sry*? Either there is some advantage to maintaining functionally equivalent, multiple versions of the gene, or these apparently redundant copies have taken on new functions. Inbred mice with two copies of *Sry* are fertile (Lyon et al. 1981) as are XYY humans (Gorlin 1977). However, there is no obvious selective advantage conferred on individuals that have these duplications. If the copies have acquired new functions, what are they? Clearly, an investigation of gene expression will be needed to answer these questions.

*Acknowledgments.* We wish to thank W.N. Verheyen for providing the tissue samples used in this study. This research was supported by National Science Foundation grant DEB-9209550 (to P.K.T.).

## References

- Bianchi NO, Bianchi MS (1993) Male specific *Sry* patterns in marsupials. *J Mamm* 74:531–534
- Bianchi NO, Bianchi MS, Pamilio P, Vidal-Rioja L, de la Chapelle A (1992) Evolution of zinc finger-Y and zinc finger-X in Oryzomyne-Akodontine rodents (Cricetidae). *J Mol Evol* 34:54–61
- Bianchi NO, Bianchi MS, Bailliet G, de la Chapelle A (1993) Characterization and sequencing of the sex determining region Y gene (*Sry*) in *Akodon* (Cricetidae) species with sex reversed females. *Chromosoma* 102:389–395
- Ferrari S, Harley VR, Pontiggia A, Goodfellow P, Lovell-Badge R, Bianchi ME (1992) *SRY*, like HMG1, recognizes sharp angles in DNA. *EMBO J* 11:4497–4506
- Foster JW, Brennan FE, Hampikian GK, Goodfellow PN, Sinclair AH, Lovell-Badge R, Selwood L, Renfree MB, Cooper DW, Marshall Graves JA (1992) Evolution of sex determination and the Y chromosome: *SRY*-related sequences in marsupials. *Nature* 359:531–533
- Gorlin RJ (1977) Classical chromosome disorders. In: Yunis JJ (ed) *New chromosomal syndromes*. Academic Press, New York, pp 60–117
- Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Munsterberg A, Vivian N, Goodfellow P, Lovell-Badge R (1990) 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* 346:245–250
- Gubbay J, Vivian N, Economou A, Jackson D, Goodfellow P, Lovell-Badge R (1992) Inverted repeat structure of the *Sry* locus in mice. *Proc Natl Acad Sci USA* 89:7953–7957
- Hawkins JR (1993) Mutational analysis of *SRY* in XY females. *Hum Mutat* 2:347–350
- Jenkins NA, Copeland NG, Taylor BA, Lee BK (1982) Organization, distribution and stability of endogenous ecotropic murine leukemia virus DNA sequence in chromosomes of *Mus musculus*. *J Virol* 43:26–36
- Kallersjo M, Farris JS, Kluge AG, Bult C (1992) Skewness and permutation. *Cladistics* 8:275–287
- Koopman P, Gubbay J, Collignon J, Lovell-Badge R (1989) *Zfy* gene expression patterns are not compatible with a primary role in mouse sex determination. *Nature* 342:940–942
- Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R (1991a) Male development of chromosomally female mice transgenic for *Sry*. *Nature* 351:117–121
- Koopman P, Munsterberg A, Capel B, Vivian N, Lovell-Badge R (1991b) Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* 348:450–452
- Lau Y, Yang-Feng TL, Elder B, Fredga K, Wiberg UH (1992) Unusual distribution of *Zfy* and *Zfx* sequences on the sex chromosomes of the wood lemming, a species exhibiting XY sex reversal. *Cytogenet Cell Genet* 60:48–54
- Li W-H (1985) Accelerated evolution following gene duplication and its implication for the neutralist-selectionist controversy. In: Ohta T, Aoki K (eds) *Population genetics and molecular evolution*. Japan Scientific Societies Press, Tokyo, pp 333–352
- Lundrigan BL, Tucker PK (1994) Tracing paternal ancestry in mice, using the Y-linked, sex-determining locus, *Sry*. *Mol Biol Evol* 11:483–492
- Lyon MF, Cattanach BM, Charlton HM (1981) Genes affecting sex differentiation in mammals. In: Austin CR, Edwards RG (eds) *Mechanisms of sex differentiation in animals and man*. Academic Press, New York, pp 329–386
- Maddison WP, Maddison DR (1992) *MacClade*. Analysis of phylogeny and character evolution, version 3. Sinauer, Sunderland, MA
- Mardon G, Mosher R, Disteché CM, Nishioka Y, McLaren A, Page DC (1989) Duplication, deletion, and polymorphism in the sex-determining region of the mouse Y chromosome. *Science* 243:78–80
- Marshall Graves JA (1995) The origin and function of the mammalian Y chromosome and Y-borne genes—an evolving understanding. *Bioessays* 17:311–320
- Miller KE, Lundrigan BL, Tucker PK (1995) Length variation of CAG repeats in *Sry* across populations of *Mus domesticus*. *Mamm Genome* 6:206–208
- Misonne X (1969) African and Indo-Australian Muridae, evolutionary trends. Tervuren Belgium Musee Royal De L’Afrique Centrale Annales. *Sci Zool* 172:1–219
- Mitchell MJ, Woods DR, Tucker P, Opp JS, Bishop CE (1991) Homology of a candidate spermatogenic gene from the mouse Y chromosome to the ubiquitin-activating enzyme E1. *Nature* 354:483–486
- Nagamine CM (1994) The testis-determining gene, *SRY*, exists in multiple copies in Old World rodents. *Genet Res Camb* 64:151–159
- Nagamine CM, Chan K, Kozak CA, Lau Y (1989) Chromosome mapping and expression of the putative testis-determining gene in mouse. *Science* 243:80–83
- Nasrin N, Buggs C, Kong XF, Carnazza J, Goebel M, Alexander-Bridges M (1991) DNA-binding properties of the product of the testis-determining gene and related protein. *Nature* 354:317–320
- Ohta T (1988) Further simulation studies on evolution by gene duplication. *Evolution* 42:375–386
- Ohta T (1994) Further examples of evolution by gene duplication revealed through DNA sequence comparisons. *Genetics* 138:1331–1337
- Page DC, Mosher R, Simpson EM, Fisher EMC, Mardon G, Pollack J, McGillivray B, de la Chapelle A, Brown LG (1987) The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* 51:1091–1104
- Pallsboll PJ, Vader A, Bakke I, El-Gewely MR (1992) Determination of gender in cetaceans by the polymerase chain reaction. *Can J Zool* 70:2166–2170
- Sanger FS, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467
- Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf A, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346:240–244
- Swofford DL (1993) *Phylogenetic analysis using parsimony (PAUP)*, version 3.1.1. University of Illinois, Champaign, IL
- Tucker PK, Lundrigan BL (1993) Rapid evolution of the sex determining locus in Old World mice and rats. *Nature* 364:715–717
- Tucker PK, Lee BK, Lundrigan BL, Eicher EM (1992) Geographic origin of the Y chromosome in “old” inbred strains of mice. *Mamm Genome* 3:254–261
- Whitfield LS, Lovell-Badge R, Goodfellow PN (1993) Rapid sequence evolution of the mammalian sex-determining gene *SRY*. *Nature* 364:713–715