

Length variation of CAG repeats in *Sry* across populations of *Mus domesticus*

K.E. Miller, B.L. Lundrigan, P.K. Tucker

The Museum of Zoology, Room 3036, 1109 Geddes, and Department of Biology, University of Michigan, Ann Arbor, Michigan 48109-1079, USA

Received: 7 September 1994 / Accepted: 16 November 1994

Sry is the Y Chromosome (Chr)-linked, sex-determining locus responsible for testis determination in mammals (Gubbay et al. 1990; Koopman et al. 1990, 1991; Sinclair et al. 1990). The gene is thought to contain a single exon consisting of a central DNA-binding motif (HMG box) 79 amino acids in length, and flanking sequences. The DNA sequence of the HMG domain is conserved across species of marsupial and placental mammals (Foster et al. 1992; Gubbay et al. 1990; Sinclair et al. 1990; Whitfield et al. 1993). However, flanking regions are not conserved even at lower taxonomic levels (Tucker and Lundrigan 1993; Whitfield et al. 1993). In Old World mice and rats (subfamily Murinae), for example, the carboxy non-box region contains both perfect and imperfect intervals of the trinucleotide repeat, CAG, and varies in length from 92 amino acids in the African rat (*Hylomyscus alleni*) to 313 amino acids in the house mouse (*Mus musculus*). Significant length variation is evident even when comparing *Mus musculus* with its sibling species, *Mus domesticus*. The carboxy non-box region in the latter species is considerably shorter, from 153 to 155 amino acids (Coward et al. 1994).

Experiments investigating sex determination in mice have involved the transfer of the *Mus domesticus* Y Chr onto the genetic background of C57BL/6J, a standard inbred laboratory mouse strain consisting of a mixture of the genomes of *Mus musculus* and *Mus domesticus*; the Y Chr of C57BL/6J is of Asian *Mus musculus* origin (Nagamine et al. 1992; Tucker et al. 1992). The transfer of different *Mus domesticus* Y Chrs onto the C57BL/6J background can result in either sex reversal, delayed development of the testicular cord, or normal development (Biddle and Nishioka 1988; Biddle et al. 1991; Eicher et al. 1982; Eicher 1994; Nagamine et al. 1987; Washburn and Eicher 1983, 1989; Washburn et al. 1990).

Recent studies have shown that *Sry* from *Mus musculus* and *Mus domesticus* differs in three respects: (1) there are seven single base pair differences, including a C-to-T transitional mutation at nucleotide position 9906 (numbers for nucleotide positions throughout the paper refer to the base pair positions of *Sry* in GenBank entry X67204) that results in the replacement of the amino acid glutamine in *Mus musculus* with a stop codon in *Mus domesticus*; (2) *Mus musculus* has one additional CAC repeat (8808–8810); and (3) there are differences in the number of CAG repeats in the intervals 8733–8765 and 8811–8846 (Coward et al. 1994). However, only the number of CAG repeats in the interval 8811–8846 has been found to vary within *Mus domesticus* (Coward et al. 1994). This variation is correlated with the occurrence of delayed testis development and sex reversal when *Mus domesticus* *Sry* alleles are placed on a C57BL/6J genetic background, i.e., C57BL/6J and the normal functioning *Mus domesticus* alleles have 12 CAG repeats in the interval 8811–8846, a *Mus domesticus* variant (AKR/J), which when crossed with C57BL/6J results in

males with delayed testicular formation [for details see Eicher 1994], has 13 repeats in this region, and another *Mus domesticus* variant (Y^{pos}), which when crossed with C57BL/6J results in male sex reversal, has 11 CAG repeats. These data suggest that the number of CAG repeats in the interval 8811–8846, perhaps in conjunction with fixed allelic differences in *Sry* between *Mus musculus* and *Mus domesticus*, affects the normal functioning of *Sry*, at least on a hybrid *Mus musculus* × *Mus domesticus* genetic background.

Unstable trinucleotide repeats within or near the coding sequence of several human disease genes (Aslanidis et al. 1992; Brook et al. 1992; Buxton et al. 1992; Fu et al. 1991, 1992; Harley et al. 1992; Koide et al. 1994; Kremer et al. 1991; Mahadevan et al. 1992; Martin 1993; Morell 1993; Nagafuchi et al. 1994; The Huntington's Disease Collaborative Research Group 1993; Verkerk et al. 1991), and one mouse gene (Ghosh et al. 1993), have been implicated in the deleterious functioning of those genes. *Sry* may be yet another gene whose function is affected by an unstable trinucleotide repeat sequence.

In an effort to characterize the stability of *Sry* CAG repeat intervals 8733–8765 and 8811–8846 [the two longest intervals of the perfect repeat CAG in *Sry*] in natural populations of *Mus domesticus*, we surveyed sequence variation in the following wild caught animals: 13 *Mus domesticus* from a single barn population, 12 *Mus domesticus* from three different outbuildings within several kilometers of each other, and an additional five *Mus domesticus* from geographically disjunct populations. The number of *Mus domesticus* sampled, their diploid chromosome number, and exact collecting locality are given in Table 1.

Polymerase Chain Reaction (PCR) amplifications of regions of

Table 1. Number of CAG repeats in the intervals 8733–8765 and 8811–8846 of the carboxy-terminal portion of *Sry* in populations of *Mus domesticus*.

Locality	Diploid number	Sample size	Number of CAG repeats	
			First interval	Second interval
Italy: Molise; 6.1 km W by road of Bonefro, single barn	2N = 22	13	9	13
Italy: Lazio; 11.4 km WNW by road of Cassino Train Station 3, out-buildings	2N = 40	12	9	13
Egypt: Beheira Govt.; Wadi el Natrun	2N = 40	1	10	14
Morocco: Tafilalt Oasis; Erfoud	2N = 40	1	11	20
Switzerland: Grisons Canton; Zalende, Poschiavo	2N = 26	1	9	11
United States: Maryland; Queen Anne's Co., Centreville	2N = 40	1	9	13

Sry containing the trinucleotide repeat were accomplished with the following primer pairs: forward 5'-GGCCATGTCAAGCGC-CCCATG-3') and reverse (5'-CACCAGTGATGTCAGCTGT-TAGTA-3'), and forward (5'-TCCTACACAGAGAAAATACC-3') and reverse (5'-GCTGTTTGTCTTTGTGTAGC-3'). To ensure PCR amplification of a Y-specific band, male and female specimens of *Mus domesticus* were included in each PCR experiment. The Sanger dideoxy method of sequencing (Sanger et al. 1977) with the Gibco BRL dsDNA Cycle Sequencing System was performed according to manufacturer's instructions. The forward sequencing primer (5'-TCCTACACAGAGAAAATACC-3') was used to sequence PCR products.

No variation was found in the nucleotide sequence of the 13 individuals trapped in a single barn near Bonefro, Italy (Table 1). All of these individuals possessed 9 CAG repeats in the first interval (8733–8765), and 13 CAG repeats in the second interval (8811–8846). Additionally, we found no variation among the 12 males collected from three different outbuildings separated by a few kilometers in Cassino, Italy. Like the population from Bonefro, all individuals had 9 CAG repeats in the first interval and 13 CAG repeats in the second interval (Table 1).

In contrast to the within population samples, *Mus domesticus* from geographically disjunct populations differed in number of CAG repeats in both the first and second intervals (Table 1). Variation in the first interval ranged from 9 to 11 repeats, with the European and North American populations having 9 repeats, and the Egyptian and Moroccan populations having 10 and 11 repeats, respectively. Variation in the second interval ranged from 11 to 20 repeats, with European and North American populations having 11 or 13 repeats and the Egyptian and Moroccan populations having 14 and 20 respectively.

Comparison of these two longest intervals of CAG repeat in *Sry* suggests that these repeats are stable within local populations, but vary across geographically disjunct populations. Our earlier studies indicate that these two intervals may be the only polymorphic regions of *Sry* in *Mus domesticus* (Lundrigan and Tucker 1994). Variation in repeat number, in contrast to lack of variation in the portions of the gene composed of unique sequence, probably reflects the more rapid rate of mutation of repeat regions relative to regions of unique sequence. Trinucleotide repeat expansion (or reduction in rare cases) is thought to be the result of dynamic mutations (Richards and Sutherland 1994). The process of mutation is "dynamic" in the sense that the number of repeats initially present determines the probability that additional changes will occur.

The variability in the two repeat intervals among populations of *Mus domesticus* also suggests that testis determination in *Mus domesticus* is not dependent upon the strict conservation of these intervals. Although evidence linking variability in trinucleotide repeat number and disease genes is extensive, it must be noted that trinucleotide repeats are very common. They are found in at least 50 human genes and within every animal studied (Morell 1993). The occurrence of nucleotide repeats and their maintenance through time suggest that they serve some functional role. King (1994) hypothesized that trinucleotide repeats may serve a regulatory role, and that variation in the number of repeats within the normal functioning range of the gene may provide a mechanism for generating phenotypic variation within a population. Unstable repeat sequences may be an especially significant mechanism for generating genotypic variation in a Y-specific gene, such as *Sry*, that does not undergo recombination and is clonally inherited through paternal lineages.

Although variation in CAG repeat length is correlated with delayed testis development and sex reversal, it is not known whether this polymorphism is responsible for the abnormal functioning of certain *Mus domesticus Sry* alleles on the C57BL/6J genetic background. A causative association would imply that the addition or deletion of even a single CAG can result in the abnormal

functioning of *Sry* (Coward et al. 1994). If this is indeed the case, it would be a unique example of a trinucleotide repeat region in which alteration of repeat number by deletion, as well as expansion, disrupts the normal functioning of the gene. Alternatively, the correlation between variation in polyglutamine repeats and sex reversal may be coincidental, resulting from the fact that this repeat region is quite variable in nature.

Acknowledgments. We gratefully acknowledge M. Potter, National Cancer Institute (NCI contract NO1-CB-71085); E. Eicher, The Jackson Laboratory; M. Nachman, Cornell University; and S. Goodman, Field Museum of Natural History, for providing frozen organs or DNA from wild house mouse populations. We thank E. Eicher for reviewing an earlier draft of this paper. This publication was supported by National Science Foundation grant DEB-9209950 (to P.K. Tucker).

References

- Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevan, M., Tsilifidis, C., Chen, C., Alleman, J., Wormskamp, N.G.M., Vooijs, M., Buxton, J., Johnson, K., Smeets, H.J.M., Lennon, G.G., Carrano, A.V., Korneluk, R.G., Wieringa, B., de Jong, P.J. (1992). Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature* 355, 548–551.
- Biddle, F.G., Nishioka, Y. (1988). Assays of testis development in the mouse distinguish three classes of domesticus-type Y chromosome. *Genome* 30, 870–878.
- Biddle, F.G., Eales, B., Nishioka, Y. (1991). A DNA polymorphism from five inbred strains of the mouse identifies a functional class of domesticus-type Y chromosome that produces the same phenotypic distribution of gonadal hermaphrodites. *Genome* 34, 96–104.
- Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J., Hudson, T., Sohn, R., Zemelman, B., Snell, R.G., Rundle, S.A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P.S., Shaw, D.J., Housman, D.E. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 68, 799–808.
- Buxton, J., Shelbourne, P., Davies, J., Jones, C., Van Tongeren, T., Aslanidis, C., de Jong, P., Jansen, B., Anvret, M., Riley, B., Williamson, R., Johnson, K. (1992). Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* 355, 547–548.
- Coward, P., Nagai, K., Chen, D., Thomas, H.D., Nagamine, C.M., Lau, Y.-F.C. (1994). Polymorphism of a CAG trinucleotide repeat within *Sry* correlates with B6.Y^{Dom} sex reversal. *Nature Genet.* 6, 245–250.
- Davies, K. (1993). Triplet repeats on the rise. *Nature* 365, 88.
- Eicher, E.M. (1994). Sex and trinucleotide repeats. *Nature Genet.* 6, 221–223.
- Eicher, E.M., Washburn, L.L., Whitney III, J.B., Morrow, K.E. (1982). *Mus poschiavimus* Y chromosome in the C57BL/6J murine genome causes sex reversal. *Science* 217, 535–537.
- Foster, J.W., Brennan, F.E., Hampikan, G.K., Goodfellow, P.N., Lovell-Badge, R., Selwood, L., Renfree, M.B., Cooper, D.W., Graves, J.A.M. (1992). Evolution of sex determination and the Y chromosome: SRY-related sequences in marsupials. *Nature* 359, 531–533.
- Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J.M.H., Holden, J.J.A., Fenwick, Jr., R.G., Warren, S.T., Oostra, B.A., Nelson, D.L., Caskey, C.T. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* 67, 1047–1058.
- Fu, Y.-H., Pizzuti, A., Fenwick, Jr., R.G., King, J., Rajnarayan, S., Dunne, P.W., Dubel, J., Nasser, G.A., Ashizawa, T., de Jong, P., Wieringa, B., Korneluk, R., Perryman, M.B., Epstein, H.F., Caskey, C.T. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 255, 1256–1258.
- Ghosh, S., Palmer, S.M., Rodrigues, N.R., Cordell, H.J., Heame, C.M., Cornell, R.J., Prins, J.-B., McShane, P., Lathrop, G.M., Peterson, L.B., Wicker, L.S., Todd, J.A. (1993). Polygenic control of autoimmune diabetes in non-obese diabetic mice. *Nature Genet.* 4, 404–409.
- 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.

- Harley, H.G., Brook, J.D., Rundle, S.A., Crow, S., Reardon, W., Buckler, A.J., Harper, P.S., Housman, D.E., Shaw, D.J. (1992). Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* 355, 545–546.
- King, D.G. (1994). Triple repeat DNA as a highly mutable regulatory mechanism. *Science* 263, 595–596.
- Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., Saito, M., Tomoda, A., Miike, T., Naito, H., Ikuta, F., Tsuji, S. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nature Genet.* 6, 9–13.
- Koopman, P., Münsterberg, A., Chapel, B., Vivian, N., Lovell-Badge, R. (1990). Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* 348, 450–452.
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for *Sry*. *Nature* 351, 117–121.
- Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R., Richards, R.I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)*n*. *Science* 252, 1711–1714.
- Lundrigan, B.L., Tucker, P.K. (1994). Tracing paternal ancestry in mice using the Y-linked, sex-determining locus, *Sry*. *Mol. Biol. Evol.* 11, 483–492.
- Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J., O'Hoy, K., LeBlond, S., Earle-Macdonald, J., de Jong, P., Wieringa, B., Korneluk, R.G. (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 255, 1253–1255.
- Martin, J.B. (1993). Molecular genetics of neurological diseases. *Science* 262, 674–676.
- Morell, V. (1993). The puzzle of the triple repeats. *Science* 260, 1422–1423.
- Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I., Murayama, N., Tanaka, Y., Kikishima, H., Umino, K., Kurosawa, H., Furukawa, T., Nihei, K., Inoue, T., Sano, A., Komure, O., Takahashi, M., Yoshizawa, T., Kanazawa, I., Yamada, M. (1994). Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nature Genet.* 6, 14–18.
- Nagamine, C.M., Taketo, T., Koo, G.C. (1987). Studies on the genetics of tda-1 XY sex reversal in the mouse. *Differentiation* 33, 223–231.
- Nagamine, C.M., Nishioka, Y., Moriwaki, K., Boursot, P., Bonhomme, F., Lau, Y.F.C. (1992). The musculus-type chromosome of the laboratory mouse is of Asian origin. *Mamm. Genome* 3, 84–91.
- Richards, R.I., Sutherland, G.R. (1994). Simple repeat DNA is not replicated simply. *Nature Genet.* 6, 114–116.
- Sanger, T., Nicklen, S., Coulson, A.R. (1977). DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A., Lovell-Badge, R., Goodfellow, P. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346, 240–244.
- The Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971–983.
- Tucker, P.K., Lee, B.K., Lundrigan, B.L., Eicher, E.M. (1992). Geographic origin of the Y chromosome in "old" inbred strains of mice. *Mamm. Genome* 3, 254–261.
- Tucker, P.K., Lundrigan, B.L. (1993). Rapid evolution of the sex determining locus in Old World mice and rats. *Nature* 364, 715–717.
- Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., Fu, Y.-H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F., Eussen, B.E., van Ommen, G.-J.B., Blonden, L.A.J., Riggins, G.J., Chastain, J.L., Kunst, C.B., Galjaard, H., Caskey, C.T., Nelson, D.L., Oostra, B.A., Warren, S.T. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65, 905–914.
- Washburn, L.L., Eicher, E.M. (1983). Sex reversal in XY mice caused by dominant mutation on chromosome 17. *Nature* 303, 338–340.
- Washburn, L.L., Eicher, E.M. (1989). Normal testis determination in the mouse depends on genetic interaction of a locus on chromosome 17 and the Y chromosome. *Genetics* 123, 173–179.
- Washburn, L.L., Lee, B.K., Eicher, E.M. (1990). Inheritance of T-associated sex reversal in mice. *Genet. Res.* 56, 185–191.
- Whitfield, L.S., Lovell-Badge, R., Goodfellow, P.N. (1993). Rapid sequence divergence of the mammalian sex-determining gene *SRY*. *Nature* 364, 713–715.