

complex class I antigens^{14,15} and the T-cell epitopes of the circumsporozoite protein of the malaria parasite *P. falciparum*¹⁶, for example, have both undergone directional selection.

It is not clear how the high number of amino-acid substitutions might have been subject to advantageous selection in primate *SRY* genes. Should a form of genetic conflict prove to be involved, it may not be intergenomic, but intragenomic or even intrachromosomal. It is clear, however, that if *SRY* is periodically subject to strong selective forces then it will be a major

driving force in Y-chromosome evolution, causing rapid substitution of Y-chromosome variants in strong linkage disequilibrium with the *SRY* variant undergoing substitution. The occurrence of such sweeps may contribute to the apparent absence of polymorphism on the Y chromosome^{17,18}. Moreover, given that reduced or inappropriate *SRY* activity can cause sex reversal, the rapid evolution of *SRY*, producing populations with different *SRY* sequences, could be a significant cause of reproductive isolation. □

Received 14 April; accepted 21 May 1993.

1. Sinclair, A. H. et al. *Nature* **346**, 240–244 (1990).
2. Su, H. & Lau, Y.-F. C. *Am. J. Hum. Genet.* **52**, 24–38 (1993).
3. Gubbay, J. et al. *Nature* **346**, 245–250 (1990).
4. Ner, S. S. *Curr. Biol.* **2**, 208–210 (1992).
5. Harley, V. R. et al. *Science* **255**, 453–456 (1992).
6. Ferrari, S. et al. *EMBO J.* **11**, 4497–4506 (1992).
7. Berta, P. et al. *Nature* **348**, 448–450 (1990).
8. Jager, R. J., Anvret, M., Hall, K. & Scherer, G. *Nature* **348**, 452–454 (1990).
9. Gubbay, J. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7953–7957 (1992).
10. Li, W. H., Wu, C. I. & Luo, C. C. *Molec. Biol. Evol.* **2**, 150–174 (1985).
11. Koopman, P., Gubbay, J., Collignon, J. & Lovell-Badge, R. *Nature* **342**, 940–942 (1989).
12. Palmer, M. S., Berta, P., Sinclair, A. H., Pym, B. & Goodfellow, P. N. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1681–1685 (1990).
13. Hill, R. E. & Hastie, N. D. *Nature* **326**, 96–99 (1987).
14. Hughes, A. L. & Nei, M. *Nature* **335**, 167–170 (1988).

15. Hughes, A. L., Ota, T. & Nei, M. *Molec. Biol. Evol.* **7**, 515–524 (1990).
16. Hughes, A. L. *Genetics* **127**, 345–353 (1991).
17. Malaspina, P. et al. *Ann. hum. Genet.* **54**, 297–305 (1990).
18. Ellis, N. et al. *Nature* **344**, 663–665 (1990).
19. Foster, J. W. et al. *Nature* **359**, 531–533 (1992).
20. GCG Program Manual for the GCG Package (575 Science Drive, Madison, Wisconsin 53711, USA, 1991).
21. Johns, M. B. & Paulus-Thomas, J. E. *Analyt. Biochem.* **180**, 276–278 (1989).
22. Felsenstein, J. *Cladistics* **5**, 164–166 (1989).

ACKNOWLEDGEMENTS. We thank J. A. Barrett, W. A. Amos, J. R. Hawkins, C.-I. Wu, A. Hughes and our colleagues for various comments, advice and computer programs, and the providers of primate cell lines and DNA samples. L.S.W. holds a Bursarship from the Imperial Cancer Research Fund. P.N.G. is supported by the Wellcome Trust and R.L.B. is supported by the UK MRC.

Rapid evolution of the sex determining locus in Old World mice and rats

P. K. Tucker & B. L. Lundrigan

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

THE Y chromosome-linked sex determining locus (*Sry*) responsible for testis determination in mammals^{1–5} contains a DNA-binding motif (HMG box) that is conserved across species of marsupial and placental mammals (infraclasses Metatheria and Eutheria, respectively)^{1,4,6,7}. But little to no sequence similarity is observed in flanking sequences between these two infraclasses, or among orders within each infraclass. We investigated the rate and pattern of evolution for the coding sequence of *Sry* in Old World mice and rats (subfamily Murinae). We found typical rates of synonymous substitution throughout the gene, but high rates of non-synonymous substitution, especially in the C-terminal (non-HMG box) region, when compared to other genes⁸. This region is also characterized by a frame-shift mutation and variation in a trinucleotide repeat motif. These data suggest that the non-box region is either functionally unconstrained or has undergone species-specific adaptive divergence.

The mouse *Sry* transcript, like human *SRY*⁹, probably consists of a single exon with an open reading frame consisting of a central HMG (high mobility group) DNA-binding domain flanked by N-terminal and C-terminal regions (A. Hacker and R. Lovell-Badge, manuscript in preparation). We have compared a portion of the *Sry* transcript (86 base pairs (bp) of the 5'-untranslated region, the N-terminal region, the HMG box and the C-terminal region) in seven species of Old World mice and rats belonging to five genera within the subfamily Murinae, family Muridae (Fig. 1). Sequence alignment was possible for all regions except a portion of the C-terminal region (amino-acid positions 144–366) which, because of a highly variable trinucleotide repeat, could not be unambiguously aligned. The C-terminal region is also characterized by a frame-shift mutation at amino-acid position 138. A phylogenetic analysis of our aligned sequence (B.L.L. and P.K.T., manuscript in preparation) indicates that the frame-shift mutation results from the insertion

TABLE 1 Per cent sequence divergence between *Mus musculus* and six species of Old World mice and rats for aligned regions of the *Sry* locus calculated by the Jukes and Cantor method²²

Species	Total 515 bp	5' untranslated 86 bp	N-terminus/ HMG box 243 bp	C-terminus 186 bp
Ms	1.7	1.2	0.4	3.8
Mp	5.8	3.5	2.5	11.3
Mh	7.0	8.1	3.3	11.3
Ha	8.2	4.7	4.5	14.5
Sl	10.3	12.8	4.9	16.1
Re	11.7	12.8	6.2	18.3

Aligned regions include a portion of the 5' untranslated region, the N-terminal region and HMG box (amino-acid positions 1–81), and a portion of the C-terminal region (82–143). Species are as described in Fig. 1.

of a thymine in the ancestral lineage giving rise to *Mastomys hildebrandtii* and *Hylomyscus alleni*.

The trinucleotide repeat varies in length among species, from 60 bp in *Hylomyscus alleni* to 666 bp in *Mus musculus*. The predominant repeat motif is CAG in each species. But in each species this motif is interrupted at irregular intervals by single base pair changes and, in *Mus musculus*, by a TTC CAT motif. The frame-shift mutation, coupled with variation in both base pair composition and length of the trinucleotide repeat, produce extensive interspecific variation in amino-acid composition in this region of the gene.

Sequence divergence (in per cent) between *Mus musculus* and each of the other species is calculated for the various regions of the aligned sequence (Table 1). In every comparison, the N terminus/HMG box is the most conserved region of the gene and the C terminus is the least conserved. The pattern of synonymous and non-synonymous substitutions for these same comparisons is shown in Table 2. The ratio of non-synonymous to synonymous substitutions is greater in the C terminus than in the N terminus/HMG for all comparisons except that between *Mus musculus* and *Mastomys hildebrandtii*. This exception is probably due to the small number of total substitutions in the N-terminus/HMG region between these two species.

Our comparison between *Mus musculus* and *Rattus exulans* *Sry* sequence was examined in light of comparative data for the same genera from 28 other gene sequences⁸. The rates of

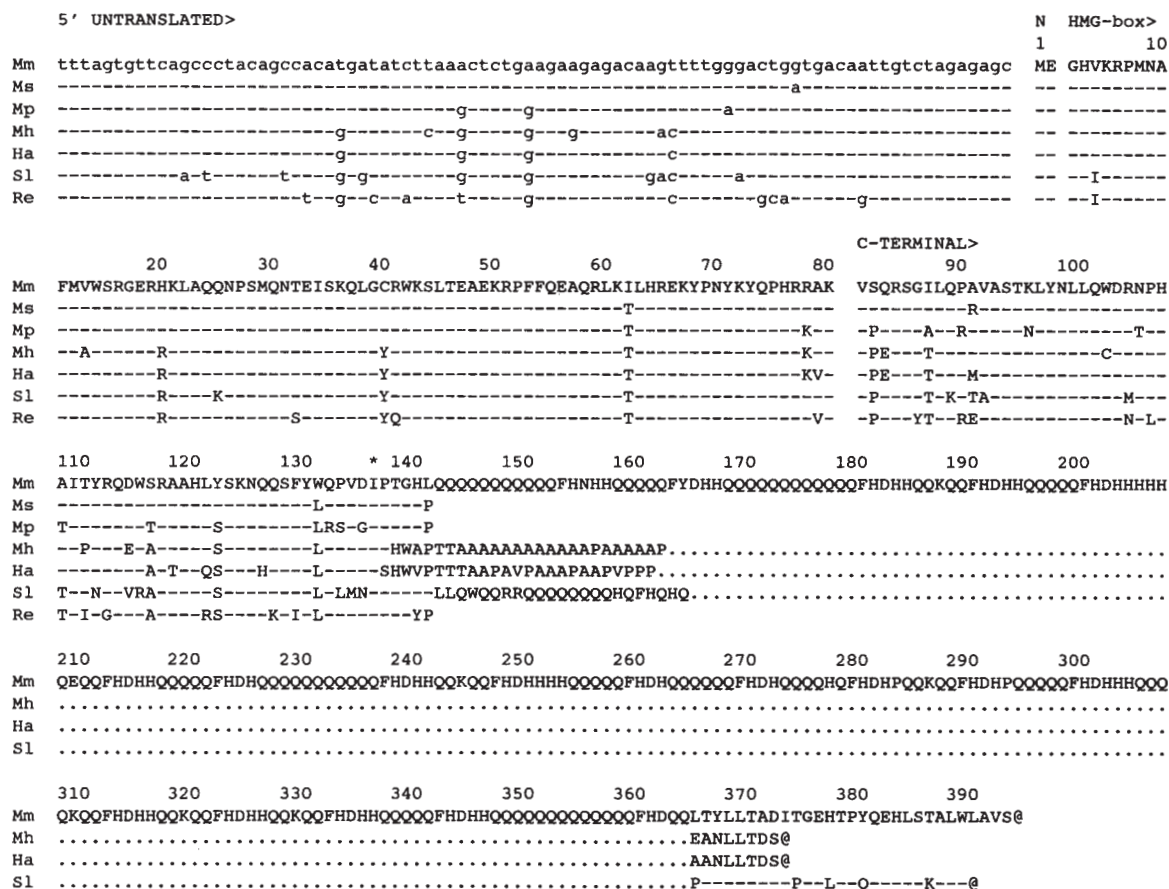


FIG. 1 Nucleotide and amino-acid sequences (single-letter code) for Sry from seven species of Old World mice and rats (subfamily Murinae, family Muridae). Aligned sequences include 86 bp of the 5' untranslated region, the N-terminal region (amino-acid positions 1–2), the HMG box (3–81) and a portion of the C-terminal region (82–143). A highly variable trinucleotide repeat motif in the C-terminal region (144–366) could not be unambiguously aligned at either the nucleotide or amino-acid level. The remainder of the C-terminal region (367–395) could not be aligned at the amino-acid level between Mm and Mh/Ha due to a single base pair insertion (*) at position 138 causing a frame-shift mutation. Dashes indicate sequence identity in aligned regions. Dots are gaps in the sequence. @, Stop codon; Mm, *Mus musculus*; Ms, *Mus spretus* (Spain); Mp, *Mus pahari* (Thailand); Mh, *Mastomys hildebrandtii* (Kenya); Ha, *Hylomyscus alleni* (Gabon); Sl, *Stochomys longicaudatus* (Gabon); Re, *Rattus exulans* (Philippines). Data for amino-acid positions 144–395

are missing for three species (Ms, Mp and Re). PCR amplifications of this region in Ms and Mp indicate that it is roughly equal in length to that for Mm. The repeat region in Re is considerably shorter. Combinations of forward (5'AGATCTTGATTTTATGTGTC3', 5'GGCCATGTCAGGCCCATG3', 5'TCCTACACAGAGAAAATACC3') and reverse (5'TGCAGCTCTACTCCAGTCTTG3', 5'CACCAGTGATGCTGTTAGTA3', 5'GCTGTTGCTGCTTTGTGCTAGC3') primers were used in amplification experiments²⁴. Sequences were aligned to the sequence from mouse strain 129²⁵, which has an Asian *Mus musculus* Y chromosome²⁶, using the SS2 algorithm²⁷, available in the Eugene package of sequence analysis programs. Alignments were subsequently modified by visual inspection. 5' untranslated, N-terminus/HMG box, and C-terminal regions correspond to nucleotide positions 8,217–8,303, 8,304–8,546 and 8,547–9,488, respectively, from the 14.5-kb genomic fragment (L741) containing the Sry gene isolated from mouse strain 129²⁵.

TABLE 2 Synonymous (ds) ± s.e. and non-synonymous (dn) ± s.e. substitutions per 100 sites and their ratio (dn/ds) for aligned regions of the Sry locus

Species	ds	Total		N terminus/HMG			C terminus		
		dn	dn/ds	ds	dn	dn/ds	ds	dn	dn/ds
Mm	3.3 (1.87)	1.5 (0.66)	0.45	0.0	0.5 (0.52)	—	6.9 (3.86)	2.8 (1.38)	0.41
Ms	11.8 (3.33)	4.8 (1.17)	0.41	8.3 (3.98)	1.0 (0.73)	0.12	15.4 (5.37)	10.0 (2.52)	0.64
Mp	16.7 (3.88)	6.5 (1.35)	0.39	6.2 (3.44)	2.6 (1.15)	0.41	28.5 (6.82)	11.6 (2.69)	0.40
Mh	22.7 (4.35)	7.5 (1.44)	0.33	12.4 (4.74)	2.7 (1.17)	0.21	33.3 (7.20)	13.8 (2.88)	0.41
Ha	20.9 (4.21)	6.8 (1.38)	0.33	14.3 (4.99)	2.6 (1.16)	0.18	28.4 (6.79)	12.3 (2.76)	0.43
Sl	22.1 (4.31)	8.6 (1.53)	0.39	15.4 (5.18)	3.9 (1.40)	0.25	29.3 (6.86)	14.8 (2.98)	0.50
Re									

Comparisons are between *Mus musculus* and six species of Old World mice and rats. The number of synonymous and non-synonymous sites compared for each analysis are as follows. Total, 91.0 to 93.5 synonymous and 332.50 to 334.75 non-synonymous; N terminus/HMG, 48.0 to 49.0 synonymous and 188.58 to 192.0 non-synonymous; C terminus, 42.58 to 45.33 synonymous and 140.67 to 143.42 non-synonymous. Calculations followed the method of Nei and Gojobori²³ using the program NAG provided by A. Hughes. Species are as described in Fig. 1; N-terminus/HMG and C-terminal regions of the gene are as described in Table 1. Comparisons between Mm and Ms, Mm and Mp, and Mm and Mh for the N terminus/HMG are based on a small number of substitutions and may not be statistically meaningful.

synonymous substitution for the N-terminus/HMG and C-terminal regions fall well within the range of synonymous substitutions for these 28 genes (median = 21.6, $r = 9.4-60.2$). In contrast, the rates of non-synonymous substitution for the N-terminus/HMG region and the combined N-terminus/HMG and C-terminal regions are among the highest of the 28 genes (median = 1.3, $r = 0-11.6$). Indeed, the C-terminal region alone shows a higher rate of non-synonymous substitution than any of these 28 genes.

The pattern of substitution for *Sry* relative to other genes was assessed by examining the ratio of non-synonymous to synonymous substitution. The dn/ds for the N-terminus/HMG box, C-terminal and combined regions are among the highest of all gene comparisons (median = 0.07, $r = 0.01-0.58$). Taken together these data suggest that, relative to other genes, *Sry*, and in particular the C-terminal region, is undergoing rapid divergence in amino-acid sequence.

This rapid sequence divergence could be explained in one of two ways: either the *Sry* gene (in particular the C-terminal region) is not functionally constrained, or it is undergoing species-specific adaptive divergence by a process of positive darwinian selection. Circumstantial evidence against the hypothesis of relaxed selection in the C-terminal region includes the following observations. First, although both the human and mouse *SR Y* proteins bind the same consensus DNA sequence *in vitro*¹⁰, XX mice transgenic for a 25 kilobase (kb) genomic fragment containing the human *SR Y* gene, express the gene, but are not sex-reversed^{5,10}. One possible explanation for this result is that the highly variable region outside the DNA-binding domain is important for the functioning of *Sry*. Second, the trinucleotide repeat in the C-terminal region of *Sry* may affect expression of the gene. Several human genetic disorders result from the presence of trinucleotide repeats either within the coding portion of a gene^{11,12} or in the 5' or 3' untranslated regions¹³⁻²¹. In every case, variation in the length of the repeat is thought to affect the normal expression of the gene or the structure of the protein product. Likewise, variation in the repeat may affect structure and function of *Sry* in a species-specific way. Ultimately, the evolutionary significance of rapid amino-acid sequence divergence in *Sry* may be best understood using a transgenic approach to assess how different naturally occurring variants of the gene, or portions thereof, from one species, function in a second closely related species. □

Received 24 May; accepted 22 June 1993.

1. Sinclair, A. H. *et al. Nature* **346**, 240-244 (1990).
2. Berta, P. *et al. Nature* **348**, 448-450 (1990).
3. Jager, R. J., Anvret, M., Hall, K. & Scherer, G. *Nature* **346**, 452-454 (1990).
4. Gubbay, J. *et al. Nature* **346**, 245-250 (1990).
5. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. *Nature* **351**, 117-121 (1991).
6. Foster, J. W. *et al. Nature* **359**, 531-533 (1992).
7. Whitfield, L. S., Lovell-Badge, R. & Goodfellow, P. N. *Nature* **364**, 713-715 (1993).
8. O'Huigin, C. & Li, W.-H. *J. molec. Evol.* **35**, 377-384 (1992).
9. Su, H. & Lau, Y.-F. *C. Am. J. hum. Genet.* **52**, 24-38 (1993).
10. Lovell-Badge, R. *Phil. Trans. R. Soc. B* **339**, 159-164 (1993).
11. Laspada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fishbeck, H. *Nature* **352**, 77-79 (1991).
12. The Huntington's Disease Collaborative Research Group *Cell* **72**, 971-983 (1993).
13. Fu, Y.-H. *et al. Cell* **67**, 1047-1058 (1991).
14. Kremer, E. J. *et al. Science* **252**, 1711-1714 (1991).
15. Verkerk, A. J. M. H. *et al. Cell* **65**, 905-914 (1991).
16. Aslanidis, C. *et al. Nature* **355**, 548-551 (1992).
17. Brook, J. D. *et al. Cell* **66**, 799-808 (1992).
18. Buxton, J. *et al. Nature* **355**, 547-548 (1992).
19. Fu, Y.-H. *et al. Science* **255**, 1256-1259 (1992).
20. Harley, H. G. *et al. Nature* **355**, 545-546 (1992).
21. Mahadevan, M. *et al. Science* **255**, 1253-1255 (1992).
22. Jukes, T. H. & Cantor, C. R. in *Mammalian Protein Metabolism* Vol. 3 (ed. Monro, H. N.) 21-120 (Academic, New York, 1969).
23. Nei, M. & Gojobori, T. *Molec. Biol. Evol.* **3**, 418-426 (1986).
24. Tucker, P. K. & Lundrigan, B. L. *Meth. Enzym.* **224**, 517-525 (1993).
25. Gubbay, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 7953-7957 (1992).
26. Tucker, P. K., Lee, B. K., Lundrigan, B. L. & Eicher, E. M. *Mammal. Genome* **3**, 254-261 (1992).
27. Altschul, S. F. & Ericson, B. W. *Bull. math. Biol.* **48**, 603-616 (1986).

ACKNOWLEDGEMENTS. We thank M. Potter (National Cancer Institute), E. Eicher (Jackson Laboratory), R. Sage (University of Missouri), B. Patterson and L. Heaney (Field Museum of Natural History), R. Baker (The Museum, Texas Tech University) for providing frozen organ tissue from wild mice; R. Lovell-Badge for sharing unpublished data; and B. Taylor, K. Miller, M. O'Keefe and S. Jansa for technical assistance. This publication was supported by a NSF grant to P.K.T.

Isolation of a Miller-Dieker lissencephaly gene containing G protein β -subunit-like repeats

Orly Reiner*, Romeo Carrozzo*, Ying Shen*, Manfred Wehnert*, Fabrizia Faustinella*, William B. Dobyns†, C. Thomas Caskey*‡§ & David H. Ledbetter*

* Institute for Molecular Genetics, and † Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA

‡ Division of Pediatric Neurology, Departments of Neurology and Pediatrics, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA

LISSENCEPHALY (agyria-pachygyria) is a human brain malformation manifested by a smooth cerebral surface and abnormal neuronal migration^{1,2}. Identification of the gene(s) involved in this disorder would facilitate molecular dissection of normal events in brain development³. Type 1 lissencephaly occurs either as an isolated abnormality or in association with dysmorphic facial appearance in patients with Miller-Dieker syndrome^{4,5}. About 15% of patients with isolated lissencephaly and more than 90% of patients with Miller-Dieker syndrome have microdeletions in a critical 350-kilobase region in chromosome 17p13.3 (ref. 6). These deletions are hemizygous, so haplo-insufficiency for a gene in this interval is implicated. Here we report the cloning of a gene (*LIS-1*, lissencephaly-1) in 17p13.3 that is deleted in Miller-Dieker patients. Non-overlapping deletions involving either the 5' or 3' end of the gene were found in two patients, identifying *LIS-1* as the disease gene. The deduced amino-acid sequence shows significant homology to β -subunits of heterotrimeric G proteins, suggesting that it could possibly be involved in a signal transduction pathway crucial for cerebral development.

During a study to identify genes containing β -transducin-like repeats⁷, degenerate polymerase chain reaction (PCR) primers were designed for conserved amino-acid sequences⁸ and used to amplify DNA from a human fetal brain complementary DNA library. The product of this amplification, when used to screen the same library, identified three cDNAs. One corresponding to the known protein 12.3 (ref. 9) and the other two were new cDNAs. The sequence of these two cDNAs was very similar, but PCR analysis of a panel of monochromosomal hybrids mapped one clone to chromosome 2 and the other to chromosome 17 (data not shown). Multiple isolates of cDNAs from chromosomes 2 and 17 were obtained from brain and kidney libraries, but here we study the gene from chromosome 17.

Several cDNA clones mapping to chromosome 17 were isolated and sequenced (Fig. 1). The presence of at least two 5' sequences is indicative of alternative splicing. Regional mapping to a small panel of somatic cell hybrids containing different segments of chromosome 17 (ref. 10) localized the cDNAs to distal 17p (p12-pter). The cDNAs were next analysed in somatic cell hybrids containing chromosomes 17 isolated from patients with Miller-Dieker lissencephaly syndrome (MDS)⁶. Two cDNA clones, representing 3' (clone 6-1) and 5' (clone 8-1) portions of the gene, were hybridized to a panel of somatic cell hybrids containing a single chromosome 17 from several MDS patients. Representative results are shown in Fig. 2a, b. Hybrid KCB4 contains the deleted chromosome from patient MDS-19 (ref. 11), one of four MDS patients whose breakpoints define the centromeric boundary of the critical region¹². The Southern blot results indicate that there is a partial deletion involving only

§ To whom correspondence should be addressed.