J Mol Evol (1987) 25:201–206

Journal of
Molecular Evolution


# Multiple Forms of Male-Specific Simple Repetitive Sequences in the Genus *Mus*

Timothy H. K. Platt and Michael J. Dewey

Department of Biology, University of South Carolina, Columbia, South Carolina 29208, USA

**Summary.** Previous reports indicate that in laboratory strains of mice, males are distinct from females in possession of repetitive DNA, notably devoid of Eco RI and Hae III sites and rich in the simple tetranucleotides GATA/GACA. We report here that such sequences originated in an ancestor common to laboratory mice, *Mus hortulanus, M. spretus,* and possibly also *M. cookii.* Interestingly, other male-specific satellite sequences were detected in *M. caroli, M. cookii, M. saxicola,* and *M. minutoides.* This novel satellite is also likely to be composed of simple repetitious sequences, but does not contain GATA and GACA. Thus, the Y chromosome appears to contain a disproportionately large amount of simple repetitious DNA. An attractive explanation for these results is that long tandem arrays of simple repeated sequences are generated at high frequency throughout the genome and that they are retained for a longer time on the Y chromosome due to the absence of homologous pairing at meiosis.

**Key words:** GATA/GACA — Y chromosome — Male-specific — Simple repetitive sequences

## Introduction

The vertebrate genome consists of a combination of unique copy and repeated sequences. One type of repeat that has received growing attention in the past few years is the simple quadruplet repeat, particularly of GATA and GACA. The reason is their association with the heterogametic sex of snakes and mice. Singh et al. (1980) described a satellite DNA specific to females of the banded krait. Designated Bkm, this DNA hybridizes to the W chromosome of snakes and the Y chromosome in mice (Singh et al. 1981; Epplen et al. 1982). Sequence analysis has demonstrated that major components of Bkm are long arrays of the tetranucleotides GATA and GACA (Epplen et al. 1983a; Singh et al. 1984). Clearly these sequences are not exclusively located on the Y chromosome in mice (Kiel-Metzger and Erickson 1984), yet the most conspicuously male-specific forms of these simple repeats appear to be tightly associated with the sex-determining part of the Y chromosome (Singh and Jones 1982; Schafer et al. 1986). This was demonstrated in sex-reversed XX-karyotype males (*Sxr*), which result from translocation of a small part of the Y chromosome onto the X chromosome. Predictably this had led to the speculation that the GATA/GACA repeats play a role in the function of the part of the Y chromosome responsible for sex determination (Epplen et al. 1983b).

The form that these *Sxr*-associated GATA/GACA repeats take is high-molecular-weight (>5-kbp) fragments lacking restriction sites for either Hae III or Alu I. Both of these enzymes recognize four-base sequences. In genomic DNA, sequences of this size are expected to be highly represented, occurring in random DNA sequences once about every 0.3 kbp. Thus, it is likely that the high-molecular-weight Alu I and Hae III male-specific fragments are composed of very long stretches of GATA/GACA and possibly other simple repeats.

The study reported here is an examination of the evolution of GATA- and GACA-rich sequences in

*Offprint requests to:* M.J. Dewey

the genus *Mus* and, more generally, in the order Rodentia. Evidence is presented indicating that the generation of the high-molecular-weight male-specific forms of these tetranucleotides occurred recently in the evolution of *Mus* species. Furthermore, we report finding in the genus *Mus* other male-specific simple repetitive sequences, which lack GATA and GACA.

## Materials and Methods

*Animals.* The following strains and species were used in this study: C57BL/10Bg was provided by S.C. Maxson (University of Connecticut, Storrs); *Mus spretus, M. hortulanus, M. caroli, M. cookii, M. saxicola,* and *M. minutoides* were provided by M. Potter (National Institutes of Health) and F. Berger (University of South Carolina). Male and female *Peromyscus leucopus* were provided by W.D. Dawson (University of South Carolina). *Mus brevirostris* (Peru) was purchased from the Jackson Laboratories.

*DNA Preparation, Electrophoresis, Blotting, and Hybridization Procedures.* Hepatic DNA was isolated by a modified version of the method of Blin and Stafford (1976) from animals that were fasted overnight prior to extraction. DNA preparations were digested overnight with double digests using Eco RI and Hae III obtained from Bethesda Research Laboratories. Digests were carried out using buffer and temperature conditions recommended by the manufacturer, with a three-fold excess of enzyme in 5 mM spermidine. Digests were electrophoresed on 0.8% agarose gels using TAE buffer (40 mM Tris, 20 mM $CH_3COONa$, 1 mM $Na_2EDTA$, pH adjusted to 8.0 with glacial acetic acid) at 40 V. The amount of DNA per lane ranged from 4 μg to about 10 μg. Gel-separated fragments were Southern transferred to nylon membranes (Amersham International) using the following protocol. Gels were depurinated in two 10-min washes in 0.24 N HCl, denatured in two 20-min washes in 1.5 M NaCl, 0.5 M NaOH, and neutralized in two 45-min washes in 1.5 M Tris, 3 M NaCl (pH 7.2). Transfer was effected using 12× SSPE (1.8 M NaCl, 120 mM $NaH_2PO_4$, 12 mM $Na_4EDTA$, pH adjusted to 7.4 with NaOH). Transferred DNA was covalently bound to the nylon membranes with UV irradiation using 5-min exposures on a UV transilluminator (Ultra Violet Products, Inc.).

Hybridizations were carried out as follows. Each filter was prehybridized for a minimum of 6 h in 15 ml of a solution containing 1% SDS, 0.08% PVP, 0.08% Ficoll, 27 μg/ml BSA, and 1 M NaCl, at a hybridization temperature determined by the base composition and size of the oligonucleotide probe (see below). During prehybridizations, oligonucleotides were 5′ end labeled in a T4 polynucleotide kinase reaction to a specific activity of $10^9$ cpm/μg DNA (Maxam and Gilbert 1980), and unincorporated nucleotides removed using a Sephadex G-25 fine (Pharmacia, Inc.) column with TE buffer (10 mM Tris, 1 mM $Na_4EDTA$, pH 8.0). For each filter, 250 ng of labeled nucleotide was added to 200 μl of 26 μg/μl *Torula* yeast RNA (Sigma Chemical Co.), heat denatured at 100°C for 10 min, and added to the prehybridization solution. Reactions were then carried out overnight. Except where stated, hybridization temperatures were determined according to the criteria of Wallace et al. (1981), at 45°C for prehybridization and hybridization of filters probed with $(GATA)_5$ and 55°C for prehybridization and hybridization of filters probed with $(GACA)_5$. This represents high stringency. Posthybridization processing consisted of one 30-min wash in 5× SSPE at room temperature followed by one 30-min wash in 5× SSPE at the hybridization temperature.

The oligonucleotides were synthesized with an Applied Biosystems 380B DNA Synthesizer and purified on a Beckman HPLC with an Altex C18 ODS column (Oligonucleotide Synthesis Facility, Department of Biology, University of South Carolina). The results reported here were replicated with more than one mouse per species. Also, parallel results were obtained with Alu I-generated digests.

## Results

### *Male-Specific Hybridization with $(GATA)_5$ and $(GACA)_5$*

Southern blots of Eco RI/Hae III-digested male and female DNA samples from laboratory mice and a variety of other rodents were hybridized to either $(GATA)_5$ or $(GACA)_5$. In all species examined, DNA from either sex hybridizes extensively with both probes (Figs. 1–3). Most of this hybridization common to males and females is to DNA fragments smaller than 3 kbp. However, as described previously (Shafer et al. 1986) and shown in Figs. 1a and 2a, male DNA from laboratory mouse strains characteristically contains high-molecular-weight (>20-kbp) species of GATA/GACA-homologous fragments (Figs. 1a and 2a). We have observed this pattern of male-specific hybridization to $(GATA)_5$ or $(GACA)_5$ in a number of laboratory strains and another commensal species, *M. brevirostris* (data not shown). Described below are the results obtained with the other rodent species roughly in the order of their increasing evolutionary distance from laboratory mice. In Fig. 5 is outlined the phylogeny of these species along with a summary of the results presented here.

### *Mus spretus* and *M. hortulanus*

Both of these species exhibit high-molecular-weight male-specific fragments complementary to either $(GATA)_5$ or $(GACA)_5$ (Fig. 1b and c). The hybridization to *M. hortulanus* male-specific bands is markedly reduced in comparison to that for *M. spretus.* Furthermore, in *M. spretus,* one of the two major male-specific GACA-containing fragments is clearly distinct from GATA-containing ones in any of these species, including *M. spretus* itself.

### *Mus caroli*

Little or no male-specific hybridization to either probe was evident. The hybridization common to the sexes, however, differed markedly between the two probes (Fig. 2b). $(GATA)_5$ hybridized to a range of sizes in a broad smear of reactivity. No particular repeat family was sufficiently represented to allow its appearance as a single band. In contrast, a major fraction of $(GACA)_5$ reactivity was localized to two <1-kbp diffuse families (seen in the lowermost region of the gel in Fig. 2b).

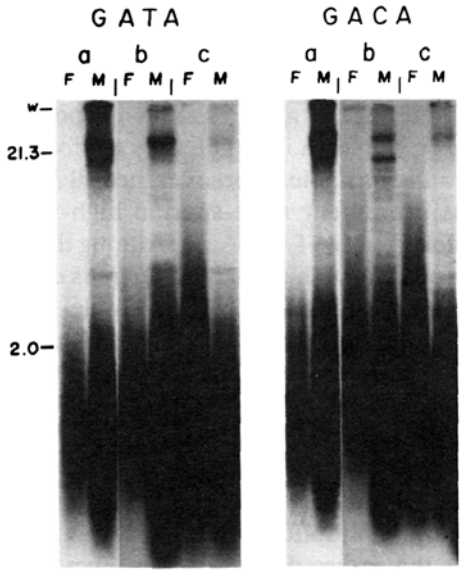


**Fig. 1.** Hybridization of end-labeled $(GATA)_5$ or $(GACA)_5$ with female (F) or male (M) DNA digested with a combination of Eco RI and Hae III and subjected to agarose gel electrophoresis. In the margin are indicated in kilobase pairs the positions of molecular-weight markers; w indicates the well position. a, C57BL/10Bg; b, *M. spretus*; c, *M. hortulanus*. The amount of C57BL/10Bg DNA was considerably greater than that of the other two species, *M. spretus* and *M. hortulanus*. However, the gel contained equal amounts of DNA from these latter species

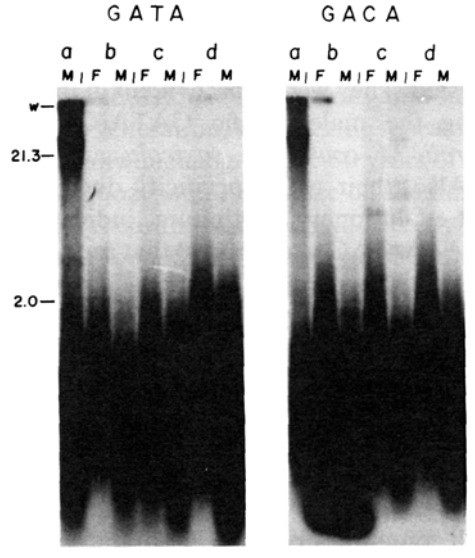


**Fig. 2.** Hybridization of end-labeled $(GATA)_5$ or $(GACA)_5$ with female (F) or male (M) DNA digested with a combination of Eco RI and Hae III. In the margin are indicated in kilobase pairs the positions of molecular-weight markers; w indicates the well location. a, C57BL/10Bg; b, *M. caroli*; c, *M. cookii*; d, *M. minutoides*

*Mus cookii*

No male-specific hybridization was apparent with $(GATA)_5$. However, blots probed with $(GACA)_5$ displayed several male-specific bands in the 3–20-kbp region of the gel. The intensity of these bands

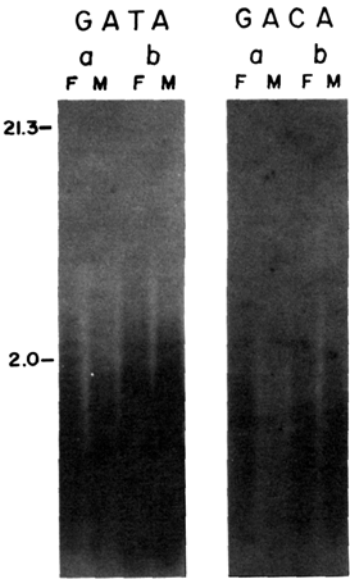


**Fig. 3.** Hybridization of end-labeled $(GATA)_5$ or $(GACA)_5$ with female (F) or male (M) DNA digested with a combination of Eco RI and Hae III. Along the side are indicated in kilobase pairs the positions of molecular-weight markers. a, *Rattus rattus*; b, *P. leucopus*

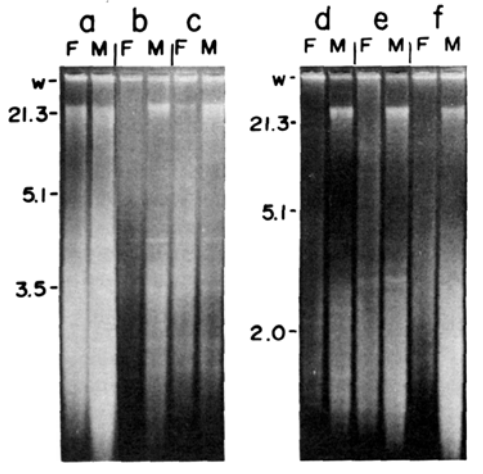


**Fig. 4.** Ethidium bromide-stained DNA fragments of female (F) and (M) genomic DNA digested with a combination of Eco RI and Hae III and electrophoresed through agarose. a, C57BL/10Bg; b, *M. spretus*; c, *M. hortulanus*; d, *M. caroli*; e, *M. cookii*; f, *M. minutoides*

was much reduced in comparison to the male-specific hybridization observed with C57BL/10Bg.

*Mus saxicola, M. minutoides, Rattus rattus,* and *P. leucopus*

No high-molecular-weight hybridization was evident with either probe in DNA samples from either sex.

*Other Male-Specific Satellites*

During the course of these studies it became apparent that in ethidium bromide-stained gels of Eco

RI/Hae III-digested genomic DNA, several *Mus* species possess high-molecular-weight male-specific repeated sequences (Fig. 4). From their species distribution, these repeat families are clearly distinct from the GATA/GACA-containing families. Described below are results with the different rodent species listed roughly in the order of their increasing evolutionary distance from laboratory mice.

Laboratory mice (C57BL/10 Bg)

DNA samples from both sexes gave prominent staining in the high-molecular-weight (>20-kbp) region of the gel.

*Mus spretus*

Marked staining localized to a diffuse region >20 kbp was evident in male samples only.

*Mus hortulanus*

The high-molecular-weight ethidium bromide band was evident in both sexes, though much stronger in male DNA.

*Mus caroli, M. cookii, M. saxicola,* and *M. minutoides*

A prominent region of ethidium bromide staining was apparent in male DNA from all these species. Such a band was much reduced or absent from female DNA.

*Rattus rattus* and *P. leucopus*

Genomic DNA from neither sex of these species displayed detectable repeat families in the higher-molecular-weight regions of the agarose gels (not shown).

## Discussion

Sites for restriction enzymes that recognize sequences of only four bases are relatively highly represented and interspersed in genomic DNA. Thus, digestion with enzymes such as Hae III or Alu I results in the generation of numerous fragments, most of which are less than 3–4 kbp. Larger fragments from such digests are likely to possess sequence characteristics differing markedly from that of the genome as a whole. This certainly is the case in male laboratory mice, for the high-molecular-weight Eco RI/Hae III digestion products that hybridize to $(GATA)_5$ or $(GACA)_5$. Thus, from the results presented here and those of others (Epplen et al. 1982; Singh et al. 1984; Schafer et al. 1986), the Y chromosome of laboratory mice is distinguished by containing long stretches rich in GATA, GACA, and possibly other simple repeated sequences.

In this study, various *Mus* and other rodent species were examined for the presence of male-specific GATA and GACA sequences. The results indicate that although these repeats are present in all species and in both sexes, the form in which they are found is highly plastic and variable. Moreover, not all *Mus* species possess prominent male-specific high-molecular-weight forms of GATA/GACA. Among those species that do possess such sequences, there is considerable quantitative variation. Also, the male-specific profiles with GATA and GACA vary independently of each other both within and among the *Mus* species. Based on these observations, the generation and subsequent existence of these male-specific GATA/GACA polymers is likely to have been the result of multiple modifying events. Included among these is one (or more) amplification event(s), as well as other events that cause diminution or even loss of these male-specific fragments. The relative contribution and timing of each is not easy to assess. Yet, from the species distribution of the male-specific GATA/GACA fragments, their generation must have occurred in an ancestor common to at least the laboratory strains, *M. spretus,* and *M. hortulanus,* and possibly including *M. caroli* and *M. saxicola* (Fig. 5).

Unexpectedly, we found from ethidium bromide-stained gels that other high-molecular-weight male-specific repeated sequences are in a number of the *Mus* species tested. They were most evident in those species lacking the male-specific GATA/GACA forms (*M. caroli, M. cookii, M. saxicola,* and *M. minutoides*). Also, their male specificity does not follow a strict evolutionary continuum, indicating (as with the GATA/GACA sequences) that multiple independent events are involved in their generation and modification. We are currently investigating the sequence structure of these newly discovered male-specific DNAs. Their lack of Alu I and Hae III sites and high molecular weight (>20 kbp) would predict a composition of simple repeated sequences, but distinct from GATA/GACA.

It is curious how the Y chromosome in these animals, which contains only about 1% of the entire genome, appears to favor the possession of high-molecular-weight simple repeats. There is yet no known function for the simple repeats, and it has been suggested that they are randomly generated as a consequence of anomalous replication from a pool of low-molecular-weight forms (e.g., see Tautz and Renz 1984). Thus, their exaggerated representation on the Y chromosome may simply be a consequence of its permanent state of haploidy. Long stretches of simple repeated sequences may be randomly generated on all chromosomes at a high frequency and, owing to mismatch at meiosis, lost from the autosomes and X chromosome at an equally high rate.

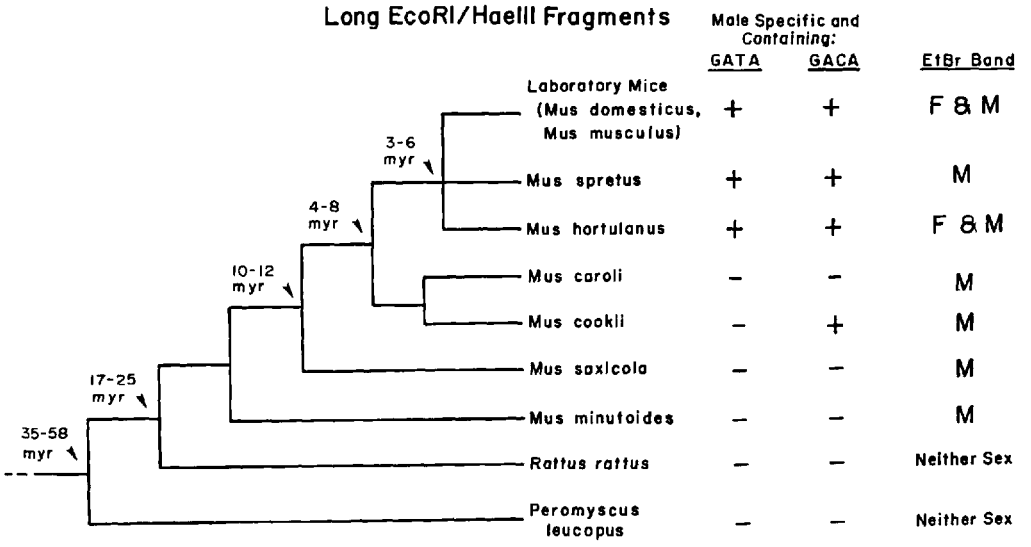


**Fig. 5.** Phylogenetic relationships of rodent species used in this study and presence of different male-specific high-molecular-weight simple repetitive sequences in each. The laboratory strain most used in this study was C57BL/10Bg, although other inbred strains and *M. brevirostrus* (Peru) were examined and found to give the same results. Many laboratory strains (including C57) are thought to consist of *M. domesticus* autosomes and mitochondrial DNA (Selander et al. 1969; Yonekawa et al. 1980) with an *M. musculus* Y chromosome (Bishop et al. 1985). This phylogeny was derived from comparative studies of protein polymorphism (Bonhomme et al. 1984), as well as divergence of total genomic (Callahan and Todaro 1978; Brownell 1983; Martin et al. 1985) and mitochondrial (Ferris et al. 1983) DNA sequences. The data are taken from results shown in Figs. 1–4. The results for *M. saxicola* are not shown. The ethidium bromide-stained gels of *Rattus* and *Peromyscus* also are not shown

Rapid loss of tandemly repeated sequences by such a mechanism has been nicely demonstrated for the polymorphic PR1 in mice (Kominami et al. 1985). Heterozygosity for this tandem repeat results in its loss or reduction at frequencies approaching 40% in a single generation, whereas the repeat has remained essentially unchanged over 30 years of maintenance as a homozygote. Residence on the nonpairing part of the Y chromosome would thereby protect a repeated sequence from this type of elimination. Other, perhaps slower, means such as unequal crossover of sister chromatids would be responsible for the decay and elimination of unneeded sequences on the Y.

*Acknowledgments.* The authors acknowledge the contributions to this work of F. Berger in the form of many helpful and critical discussions. We also wish to thank B. Branham for help in the preparation of this manuscript. This work was supported by Public Health Service grant HD 17523 from the National Institute of Child Health and Human Development.

## References

Bishop CE, Bousot P, Baron B, Bonhomme F, Hatat D (1985) Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus musculus* Y chromosome. Nature (Lond) 315:70–72

Blin N, Stafford DW (1976) A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res 3:2303–2308

Bonhomme F, Catalan J, Britton-Davidian J, Chapman VM, Moriwaki K, Nevo E, Thaler L (1984) Biochemical diversity and evolution in the genus *Mus*. Biochem Genet 22:275–303

Brownell E (1983) DNA/DNA hybridization studies of muroid rodents. Evolution 37:1034–1051

Callahan R, Todaro GJ (1978) Four major endogenous retrovirus classes each genetically transmitted in various species of *Mus*. In: Morse HC (ed) Origins of inbred mice. Academic Press, New York, pp 689–713

Epplen JT, McCarrey JR, Sutow S, Ohno S (1982) Base sequence of a cloned snake W chromosome fragment and identification of a male specific putative mRNA in the mouse. Proc Natl Acad Sci USA 79:3798–3802

Epplen JT, Cellini A, Romero S, Ohno S (1983a) An attempt to approach the molecular mechanisms of primary sex determination: W- and Y-chromosomal conserved simple repetitive DNA sequences and their differential expression in mRNA. J Exp Zool 228:305–312

Epplen JT, Cellini A, Shorte M, Ohno S (1983b) On evolutionarily conserved simple repetitious DNA sequences: do "sex specific" satellite components serve any sequence dependent function? Differentiation 23(Suppl):60–63

Ferris SD, Sage RD, Prager EM, Ritte U, Wilson AC (1983) Mitochondrial DNA evolution in mice. Genetics 105:681–721

Kiel-Metzger K, Erickson RP (1984) Regional localization of sex-specific Bkm-related sequences on proximal chromosome 17 of mice. Nature (Lond) 310:579–581

Kominami R, Sudo K, Yoshikura H, Suzuki H, Moriwaki K, Hilgers J, Muramatsu M (1985) A polymorphic repetitive-sequence RR1 family. Evidence for meiotic instability. J Mol Biol 183:301–309

Martin SL, Voliva CF, Hardies SC, Edgell MH, Hutchinson CA III (1985) Tempo and mode of concerted evolution in the L1 repeat family of mice. Mol Biol Evol 2:127–140

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65: 499–560

Schafer R, Ali S, Epplen JT (1986) The organization of the evolutionarily conserved GATA/GACA repeats in the mouse genome. Chromosoma 93:502–510

Selander RK, Hunt WG, Yang SY (1969) Protein polymorphism and genetic heterozygosity in two European subspecies of the house mouse. Evolution 23:379–390

Singh L, Jones KW (1982) Sex reversal in the mouse (*Mus musculus*) is caused by a recurrent non-reciprocal crossover involving the X and an aberrant Y chromosome. Cell 27:205–216

Singh L, Purdom IF, Jones KW (1980) Sex chromosome-associated satellite DNA. Evolution and conservation. Chromosoma 79:137–157

Singh, L, Purdom IF, Jones KW (1981) Conserved sex-chromosome associated nucleotide sequences in eukaryotes. Cold Spring Harbor Symp Quant Biol 45:805–813

Singh L, Phillips C, Jones KW (1984) The conserved nucleotide sequences which define Sxr in the mouse are transcribed. Cell 36:111–120

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12:4127–4138

Wallace RB, Johnson MJ, Hirose T, Miyake T, Kawashima EH, Itakura K (1981) The use of synthetic oligonucleotides as hybridization probes II: hybridization of oligonucleotides of mixed sequence to rabbit $\beta$ globin DNA. Nucleic Acids Res 9:879–893

Yonekawa H, Moriwaki K, Gotsh O, Watanabe J, Hayashi J-C, Miyashita N, Petras ML, Tagashira Y (1980) Relationship between laboratory mice and the subspecies *Mus musculus domesticus* based on restriction endonuclease cleavage patterns of mitochondrial DNA. Jpn J Genet 55:289–296

Received September 30, 1986/Revised January 12, 1987