Chromosoma (Berl) (1986) 93:502–510

CHROMOSOMA

The organization of the evolutionarily conserved GATA/GACA repeats in the mouse genome

Renate Schäfer, Sher Ali, and Jörg T. Epplen [1]

Junior Research Unit, Max-Planck-Institute for Immunobiology, D-7800 Freiburg, Federal Republic of Germany;
[1] addressee for reprint requests

Abstract. Simple repeated GATA and GACA sequences which were originally isolated from sex-specific snake satellite DNA have been found subsequently in all eukaryotes studied. The organization of these sequences within the mouse genome was investigated here by using synthetic oligonucleotide probes as a novel tool in comparison with conventional hybridization probes. Southern blot hybridization showed sex-specific patterns with both the $(GATA)_4$ and $(GACA)_4$ oligonucleotide probes, as previously described with conventional probes. The quantitative analysis of two mouse DNA phage libraries and of 25 isolated GATA-positive phage clones revealed intensive interspersion of GATA sequences with GACA, and with other repetitive and single-copy sequences. Ubiquitous interspersion and homogeneous genomic distribution of GATA and GACA sequences were confirmed by hybridization in situ of the oligonucleotide probes to metaphase chromosomes. The lengths of the GATA and GACA stretches were found to vary considerably in the individual phage clones. DNA inserts from 20 phages were assigned to autosomes and sex chromosomes and three genomic fragments were found to be confined to the Y chromosome. The organization of GATA and GACA sequences is discussed in the context of their evolutionary potential and possible conservation mechanisms.

Introduction

Although reiterated sequences appear ubiquitously in eukaryotes, the biological role of repetitive DNA is still enigmatic (Epplen and Ohno 1984). Eukaryotic genomes contain sequences of various degrees of repetition and repetitive DNA represents a substantial fraction of the genome (e.g. one-third in the mouse). A fraction of repetitive DNA is simple sequence DNA which is composed of short segments extensively reiterated in tandem fashion. Usually these sequences show some divergence often giving rise to long range periodicities. A priori, simple sequences have been regarded exclusively as extremely large uninterrupted repeating units. The finding that simple repeats may also be intimately interspersed with single copy and other repetitive DNA over the whole chromosomal complement requires their further analysis.

Only a few simple sequences have been characterized by sequence analysis. GATA and GACA simple repeats are but one example of the variety of this class of repetitive DNA. Investigation of GATA *s*imple *q*uadruplet *r*epeats (*sqr*) has shown sex-specific arrangements of these sequences in mammals (Epplen et al. 1982; Singh et al. 1984) although they had originally been isolated from sex-specific satellites of colubrid snakes. This is all the more unexpected since different chromosomal sex-determination mechanisms have been demonstrated in the snake and in mammals. The ZZ/ZW chromosomal mechanism with the heterogametic female operates in the reptilian species whereas in the XX/XY scheme of mammals the male is the heterogametic sex. In the mouse the sqr are interspersed over the whole chromosomal complement yet a substantial part of the GATA sqr is concentrated in the pericentric region of the Y chromosome (Singh and Jones 1982; Epplen et al. 1983b). There is further circumstantial evidence that GATA sqr are linked to genes associated with sex determination: *Sxr* is a sex-limited condition which causes XX mice to appear as phenotypic males. During meiosis of XY^{Sxr} male mice, the sex-reversal factor (Sxr) linked to GATA sqr is transferred from the distal Y^{Sxr} chromosome to a paternal X chromosome producing XX^{Sxr} male progeny. The question thus arises if GATA sqr themselves or closely linked sequences code for the Sxr factor.

The goal of the present report was therefore to describe the genomic organization of sqr with respect to length, distribution, DNA sequence interspersion and chromosomal localization. Furthermore an initial characterization of adjacent sequences was attempted. In the accompanying paper (Schäfer et al. 1986) studies on the expression of sqr are reported. Apart from any speculative present day functions, GATA sqr have proved to be effective tools for isolating Y-chromosomal sequences in the mouse.

Materials and methods

Library screening. Two genomic phage libraries from C57/Black 6 (B6) mice were kindly provided by R. Bruce Wallace. One represented Eco RI-digested B6 DNA cloned into Charon 4A, the other partially Sau 3AI-digested B6 DNA cloned into the Bam HI site of Charon 28. Library screening was done according to Benton and Davis (1977).

DNA preparation, gel electrophoresis and hybridization. Genomic DNA from mouse liver was prepared essentially according to Blin and Stafford (1976) and phage DNA according to Maniatis et al. (1982). DNA was digested with

restriction enzymes (BRL) according to the suppliers recommendations and fractionated on 0.5% horizontal agarose gels. The gels were either dried by a modified procedure as outlined by Tsao et al. (1983) for hybridization to oligonucleotides or the DNA was transferred onto nitrocellulose (Schleicher and Schüll) by Southern blotting (Southern 1975) for hybridization to nick-translated probes. Conventional probes were nick translated using [α-^{32}P]dATP (New England Nuclear) according to Rigby et al. (1977). Oligonucleotides, i.e. the 16mers $(GATA)_4$ and $(GACA)_4$, were labelled by [γ-^{32}P] ATP in a kinase reaction as described by Schöld et al. (1984). Filters of plaque lifts of the libraries were hybridized with conventional probes in 3 × SSC (68° C; 1 × SSC = 0.15 M NaCl, 0.015 M Na-citrate) and washed with 2 × SSC (68° C); Southern blots were hybridized in 6 × SSC (68° C) and washed with 2 × SSC (68° C).

Library screening filters as well as dried gels and Southern blots were hybridized with the 16mers $(GATA)_4$ at 35° C and $(GACA)_4$ at 43° C in 5 × SSPE (1 × SSPE = 180 mM NaCl, 10 mM $(Na_{1.5})PO_4$, 1 mM Na_2 EDTA, pH 8.0), 0.1% SDS, and washed with 6 × SSC on ice followed by a 1 min wash in 6 × SSC at the respective hybridization temperatures.

Determination of the length of GATA and GACA sqr stretches. For sqr length determination the 16mers $(GATA)_4$ and $(GACA)_4$ were labelled as described above and purified by 8% polyacrylamide gel electrophoresis under denaturating conditions. Bands containing the labelled 16mers were cut, eluted in TE buffer (10 mM Tris, pH 8.0; 1 mM EDTA) overnight at 4° C and desalted on a Sephadex G25 column. About 10 ng of cloned phage DNA plus 2×10^5 cpm of purified ^{32}P-labelled oligonucleotide were denatured at 95° C for 5 min and then rapidly cooled on ice. Reannealing of the oligonucleotides to phage DNA took place during the incubation for ligation. The ligation of the labelled oligonucleotides was done in a total volume of 5 µl with 0.6 U of T4 DNA ligase (BRL) for 3 h at 15° C in 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 10 mM dithiothreitol. The reaction mixture was then loaded on an 8% gradient polyacrylamide gel under denaturing conditions (Biggin et al. 1983), the gel was dried and exposed overnight to XAR-Film (Kodak). S_1 nuclease treatment of the annealed and ligated oligomer/template complex under various conditions neither altered nor improved the sqr length estimate. If two or more sqr stretches were present in an individual cloned phage DNA, then obviously only the longest sqr stretch could be determined.

Oligonucleotide hybridization in situ of the $(GATA)_4$ and $(GACA)_4$ probes to mouse metaphase chromosomes. Chromosome spreads were prepared from mouse bone marrow cells according to standard methods. Slides were treated with RNase A (100 µg/ml, Boehringer) for 1 h at 37° C, rinsed in 2 × SSC and dried in an ethanol series. After DNA denaturation (70% formamide, 2 × SSC at 70° C for 2 min), hybridization was carried out in situ under coverslips in 5 × SSPE, 0.5% SDS, 50 µg/ml *Escherichia coli* DNA and 2.5×10^6 cpm/ml probe for 3 h [$(GATA)_4$, 35° C; $(GACA)_4$, 43° C]. Slides were washed 3 times for at least 15 min each in 6 × SSC on ice and 1 min at the respective hybridization temperatures and subsequently dried. Slides were coated with NTB3 emulsion (Kodak) and exposed for 4 to 8 days at −70° C. Individual metaphases were photographed after G-banding according to standard procedures (Nesbitt and Francke 1973). As a control for the specificity of hybridization of the $(GATA)_4$ and $(GACA)_4$ probes, the length-matched oligonucleotide 5′-AATTCCG-TATCGATGC was used. This sequence is not present in the mouse genome. Hybridization, washing and exposure were carried out as outlined above except that the hybridization temperature was 41° C according to the differing base composition.

Results

GATA/GACA repeats in the mouse genome

Using uncharacterized nick-translated probes (Singh et al. 1981) or conventional probes containing sqr (Epplen et al. 1982), sex-specific hybridization patterns have been demonstrated with Southern blots of mouse DNA. $(GATA)_4$ and $(GACA)_4$ oligonucleotide probes produce more distinct sex-specific signals (Fig. 1): XY and XX^{Sxr} male Alu I- or Hae III-digested DNA probed with $(GATA)_4$ shows very similar patterns which differ from those of XX female DNA. In contrast to females, males show several bands hybridizing in the high molecular weight range (>4 kilobase (kb) in the Alu I digestion; >10 kb in the Hae III digestion). The $(GACA)_4$ probe gives similar results although sex-specific hybridization is less pronounced in Hae III- as compared to Alu I-digested DNA.

Table 1. Screening of recombinant genomic phage clones from the mouse[a]

Probe	Number of cross-hybridizing phages per haploid genome B6 library		Isolated sqr-positive phages ($n=25$)
	Sau 3A I	Eco RI	
Nick-translated $(GATA)_n$ probe	200–400[b]	185	4
$(GATA)_4$	7,200	9,100	4
Nick-translated $(GACA)_n$ probe	245	220–420[b]	0
$(GACA)_4$	7,900	8,600	0
Nick-translated $(GATA_n + (GACA)_n$ probes	21	N.T.[c]	0
$(GATA)_4 + (GACA)_4$	900	1,100	0
Nick-translated $(GATA)_n$ probe + nick-translated B6 repetitive DNA	>30,000	>30,000	18
Nick-translated $(GATA)_n + (GACA)_n$ nick-translated + B6 repetitive DNA	>30,000	>30,000	3

[a] 200,000 genomic phage clones were screened, which corresponds to one haploid mouse genome

[b] The lower estimate was obtained by counting only those recombinant phages that annealed strongly to the probe. The higher estimate includes phage that annealed weakly to the probe

[c] N.T., not tested

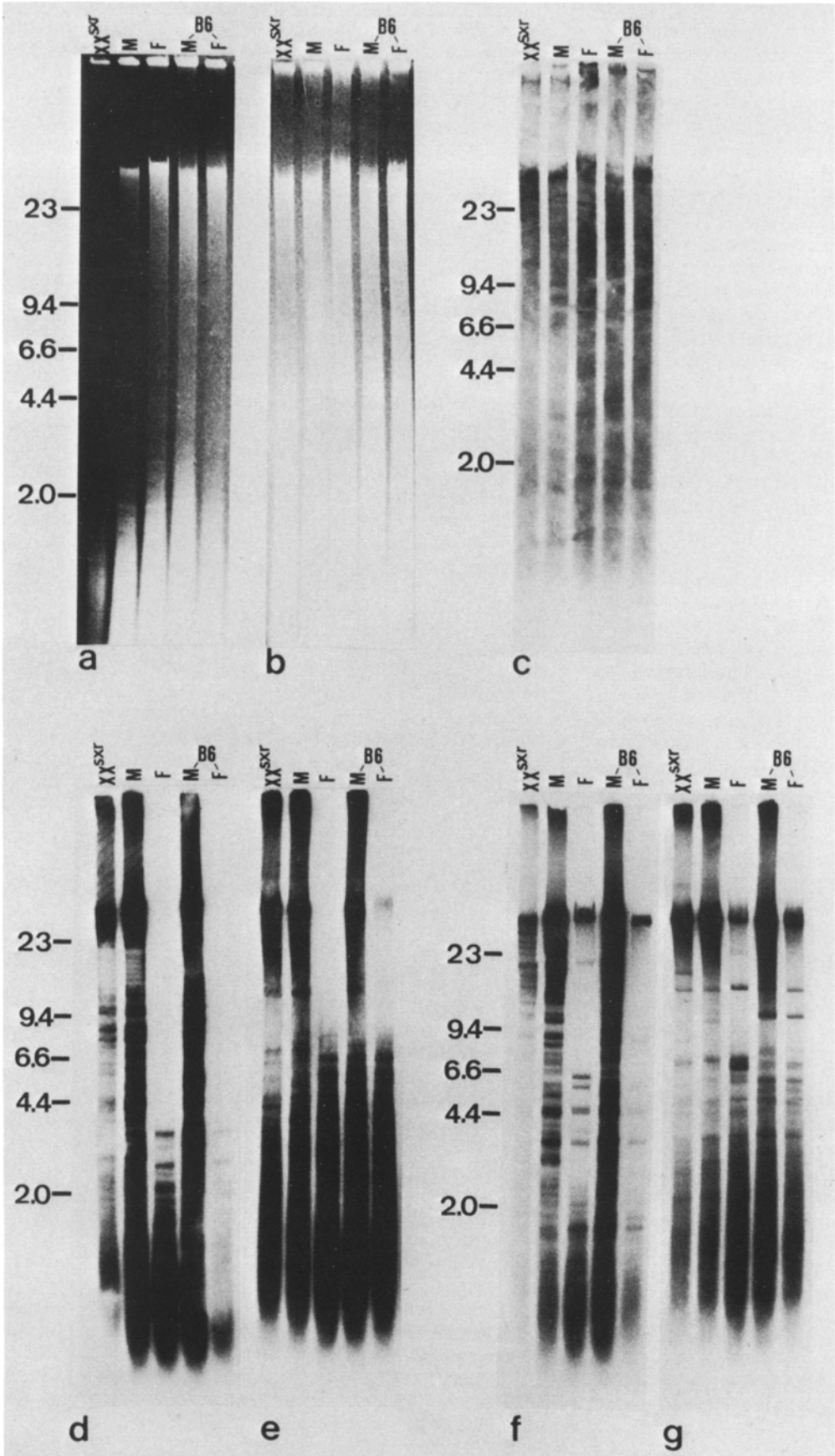

Fig. 1. Dried gel hybridization with restriction enzyme digested mouse DNA using different probes. For ethidium bromide staining (*a, b*) and all hybridizations (*c–g*) the same gel was used. DNA from left to right: XX^{Sxr}, male (M) and female (F) litter mates; male (M) and female (F) B6. *a, d, f,* Alu I digestion; *b, c, e, g* Hae III digestion. Probes: *c,* total B6 DNA; *d, e,* $(GATA)_4$; *f, g,* $(GACA)_4$. Molecular weight markers are given in kilobases on the left. For methodological details see Materials and methods

Table 2. Chromosomal localization of DNA inserts from sqr-positive phages

	A	X	Y	R	A+Y	X+Y	Y+R	A+X+Y	X+Y+R	A+X+Y+R
Number of phages ($n=20$)	6	0	3	1	3	1	2	1	2	1

R repeated autosomal DNA (refers to 10 copies and more present per genome); *A* autosomal; *Y* Y-chromosomal; *X* X-chromosomal

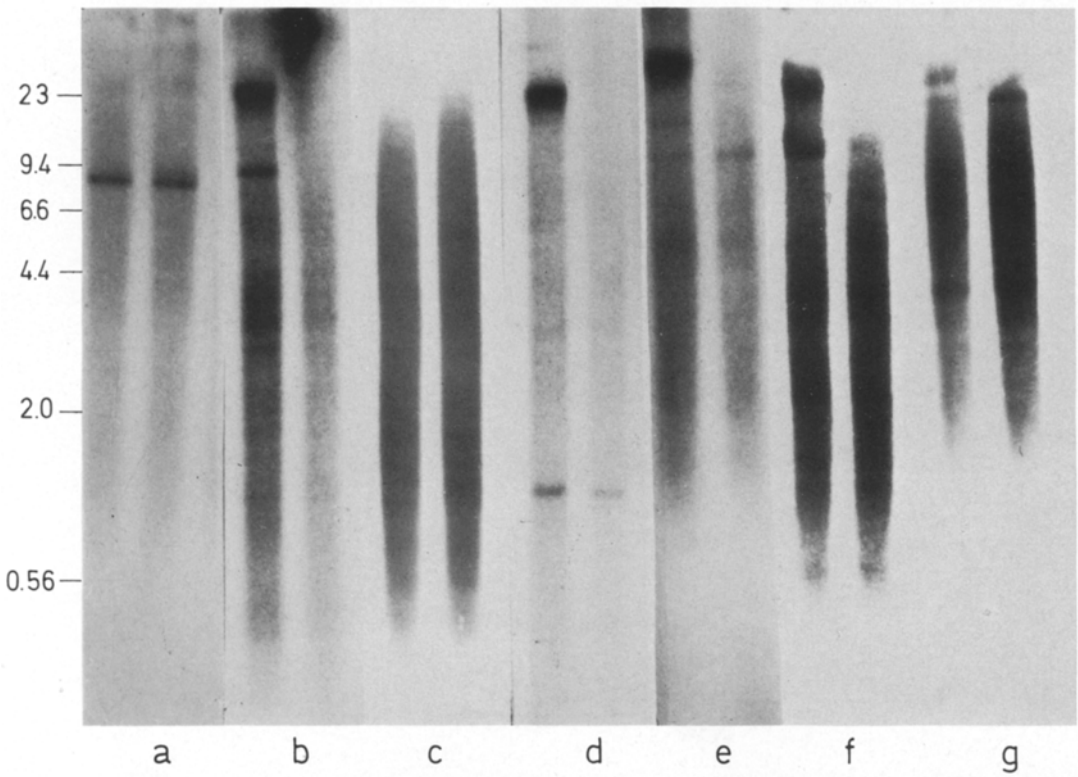

Fig. 2. Southern blot hybridization with restriction enzyme digested male (left lane of each lane pair) and female (right lane) B6 mouse DNA using individual phage DNAs as probes. Hybridization patterns produced by autosomal (*a*), Y-chromosomal (*b*), repetitive DNA (excluding GATA/GACA sqr) (*c*). Superimposition of autosomal and Y-chromosomal (*d*), X- and Y-chromosomal patterns (*e*); patterns characteristic for repetitive in addition to Y-chromosomal sequences (*f*), and complex pattern produced by repetitive sequences on autosomes and sex chromosomes (*g*). Molecular weight markers are given in kilobases on the left. For methodological details see Materials and methods

In order to determine the copy numbers of sqr clusters and subsequently to isolate male-specific DNA sequences, two genomic DNA libraries of male mice were screened. Besides the 16mers $(GATA)_4$ and $(GACA)_4$, two additional probes were used under moderate stringency conditions for genomic mouse DNA library screening. (a) A conventional "GATA" probe of some 750 basepairs (bp) containing an almost uninterrupted stretch of 37 GATA sqr. This clone was isolated from a *Drosophila melanogaster* genomic DNA library (Levinson and Epplen, unpublished) and has some 98% sequence identity with the sqr flanking region of the previously described CS 319 (Singh et al. 1984). The sequence of a second "GATA" probe used is given in the accompanying paper (Schäfer et al. 1986, see pmlc 4). (b) A conventional "GACA" probe of 670 bp contains approximately 20 exact copies of GACA in a stretch of "degenerating" GACA sqr. The full sequence is depicted in the accompanying paper (see pmlc 2).

The average insert length is more than 12 kb in the Sau 3A I library and more than 13 kb in the Eco RI library. From each 200,000 phage plaques were surveyed (Table 1) which corresponds to one haploid mouse genome (3.5×10^6 kb). In both libraries the conventional "GATA" and "GACA" probes revealed nearly identical numbers of cross-hybridizing phages. Among the GATA-positive plaques, 10% were additionally GACA-positive. Hybridization of the Eco RI library and the Sau 3A I library with the 16 mers yielded from each 30–50 times more positive clones per haploid genome (Table 1). Therefore, the majority of sqr are organized in short stretches which cannot be detected by conventional cloned probes. In each library about 12% of the clones hybridize to both 16mers. The numbers of double-positive phage clones with either oligonucleotides or conventional probes are thus significantly higher than would be expected, if the distributions of both types of sqr were completely independent of each other ($p < 0.01\%$ according to the Chi^2 test).

Of the conventional "GATA" probe positive phages of the Sau 3A I library 25 were isolated by plaque purification. This corresponds to more than 300 kb of the mouse genome. DNA was prepared and digested with several restriction enzymes before electrophoresis. Southern blots of

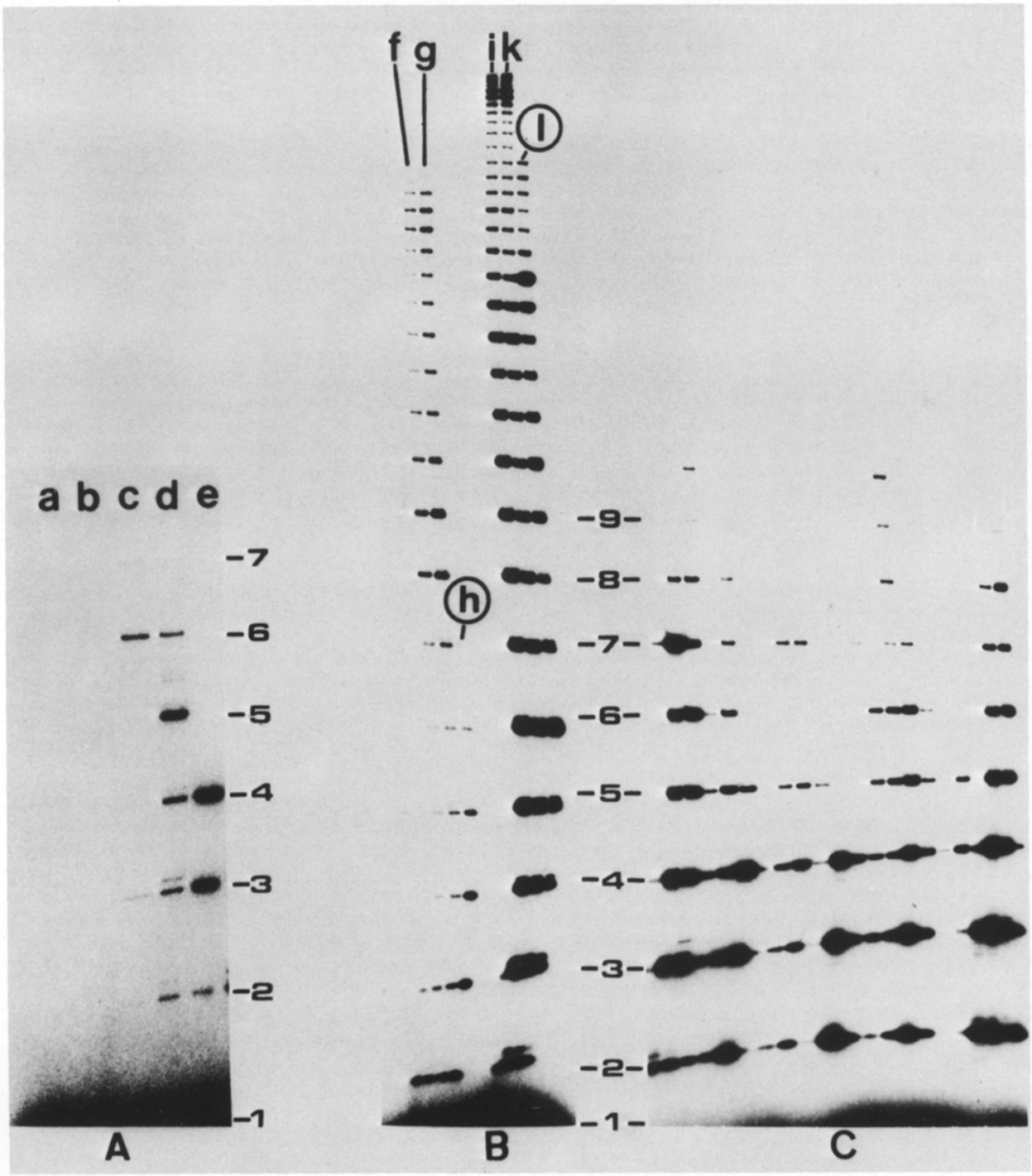

Fig. 3A–C. Length determination of GACA, GATA and GATA/GACA stretches of representative mouse genomic phage clones. A template/ligation assay was developed as described in Materials and methods. Ligation products of labelled oligonucleotides were analysed by polyacrylamide gradient gel electrophoresis under denaturing conditions. **A** Radiolabelled $(GATA)_4$ was used in this template-ligation experiment. Templates: *a* Charon 30; *b* pBR322; *c* pmlc4; *d* phage 4; *e* phage 26. **B** Both labelled oligonucleotides, $(GATA)_4$ and $(GACA)_4$, were used separately or in combination: in *f*, *g* and *h*, $(GACA)_4$; in *i*, *k*, and *l*, $(GACA)_4$ plus $(GATA)_4$. Templates: *f* and *i*, phage 25; *g* and *k*, phage 28; *h* and *l*, phage 31. **C** Radiolabelled $(GATA)_4$ was used in this template/ligation assay. Individual lanes represent different isolated phages. The numbers of ligated 16mers in the respective bands are indicated between (**A**) and (**B**) and between (**B**) and (**C**)

the gels were hybridized to the conventional "GATA" and "GACA" probes as well as to total B6 mouse DNA. Total mouse DNA was accompanied by a 100-fold excess of unlabelled GA^T_CA sqr as competitor in order to exclude sqr hybridization (since repeats present in 100 copies or more could be detected). The cumulative numerical result is given in Table 1: 21 of the 25 clones hybridize with total B6 DNA which demonstrates that GATA sqr stretches are randomly interspersed with other abundant sequences (with probably more than 1,000 copies per genome) on a DNA sequence

length of 12–13 kb. About 10% of the isolated phages hybridized with the "GACA" probe, which corresponds to the library screenings. Alu I digestion of the clones and their subsequent hybridization with $(GATA)_4$ yields hybridizing bands in the size range of 350 bp to 3,600 bp. This is within the range where both male and female Alu I-digested genomic mouse DNA shows hybridizing bands (Fig. 1). It is thus possible that clones containing longer sqr-positive fragments could not be cloned or that the sqr-positive Hae III/Alu I fragments had been shortened by the Sau 3A I digestion employed for the preparation of the library. Restriction mapping data revealed 3 of the 25 phages to be indistinguishable.

Autosomal or gonosomal localization of sqr-positive phage DNAs

Individual phage DNAs were digested with appropriate restriction enzymes. Wherever possible insert subfragments negative for conventional "GATA", "GACA" and highly repetitive sequences were isolated from low melting point agarose gels. Southern blots of male and female B6 mouse DNAs digested with the same restriction enzyme were challenged with the labelled insert subfragments. These subfragments often cross-hybridize with two or more identical or closely related sequences on autosomes and gonosomes. Thus, as shown in Table 2, nine principle patterns of hybridization were observed: Figure 2a relates to autosomal or pseudoautosomal localization of single-copy sequences. The bands observed in Figure 2b, c are best explained by Y-chromosome specific and repetitive DNA, respectively. Other patterns appear as mixes of the three single-locus patterns. Because of their extensive cross hybridization, only three of the phage inserts could be assigned exclusively to the Y chromosome.

Determination of the lengths of the longest GATA/GACA stretches in sqr-positive phage DNAs

As described in Materials and methods, a simple method for estimating the lengths of the GATA/GACA stretches was established. Several DNAs were used in control experiments: pmlc 2 is a cDNA clone containing only GACA repeats (the sequence is given in the accompanying paper). When used as a template for annealing and ligation of $(GATA)_4$ subunits, pmlc 2 does not give any bands in addition to the signal of the unreacted $(GATA)_4$ monomer (Fig. 3). The same is true for pBR322 or Charon 30 used as control templates. If $(GACA)_4$ is used in the pmlc 2 template/ligation experiment, three bands occur; the first and the second band (32mer) are more pronounced than the third one. The signal of the 32mer can be expected considering the uninterrupted array of six GACA sqr separated from another GACA sqr by GA. The third weaker band of the 48mer can be explained by the two GACA complementary tetramers exactly 12 bases away from the longer cluster. Thus two GACA complementary sqr seem to be sufficient for annealing and ligation of an additional 16mer.

Clone pmlc 4 (sequence shown in Schäfer et al. 1986) contains two uninterrupted arrays of GATA of 13 and 14 sqr separated by three bases and flanked at the 5′ end by GATAGA and by GATT at the 3′ end. Accordingly, bands 3,6 and 7 (12,24 and 28 GATA sqr, respectively) are prevalent (Fig. 3). In the GATA-positive phage clones the number of ligated $(GATA)_4$ is in the range of 5 to more than 30 16mers (20 to more than 120 sqr), 7 (28 sqr) being most frequent (Table 3). The three GACA-positive clones contain stretches of 6 and twice 19 $(GACA)_4$ (Table 4). In the "double-positive" clones the use of both oligonucleotides together in the template/ligation experiment leads to bands of higher molecular weights than using each 16mer alone (Table 4). This finding demonstrates that both sqr types are interspersed with each other and agrees with hybridization of both 16mers to the same individual band of Alu I-digested phage DNA (data not shown).

Table 3. Distribution of sqr-stretch lengths in 'GATA'-positive phages

Maximal number of ligated $(GA^T_CA)_4$ (n)	No. of phages containing	
	$n \times (GATA)_4$ or	$n \times (GACA)_4$
5	1	
6	1	1
7	5	
10	1	
11	1	
12	2	
13	2	
15	2	
16	2	
19	1	2
21	1	
22	1	
23	1	
24–26	3	
30–35	1	

Table 4. Length of GATA, GACA, and GATA + GACA stretches from sqr-positive phages

Clone	Maximal number of ligated $(GATA)_4$	Maximal number of ligated $(GACA)_4$	Maximal number of ligated $(GATA)_4 + (GACA)_4$
25	23	19	30–35
28	12	19	30–35
31	16	6	21
pmlc 2	–	3	3

Oligonucleotide probes demonstrate random distribution of sqr on mouse chromosomes by in situ hybridization

Conventional nick-translated probes have proved that sqr are preferentially detectable by in situ hybridization in the pericentric region of the mouse Y chromosome (Epplen et al. 1982; Singh et al. 1982) and secondarily on mouse chromosome 17 (Kiel-Metzger and Erickson 1984). Due to the inherent limitations of conventional probes, only relatively long stretches of sqr can be demonstrated. Therefore we attempted to introduce a hybridization method employing oligonucleotide probes in situ. The sensitivity of the method is certainly limited because by standard 5′ kinase labelling only *one* radioative reporter group can be attached per probe molecule. However, since sqr are present up to

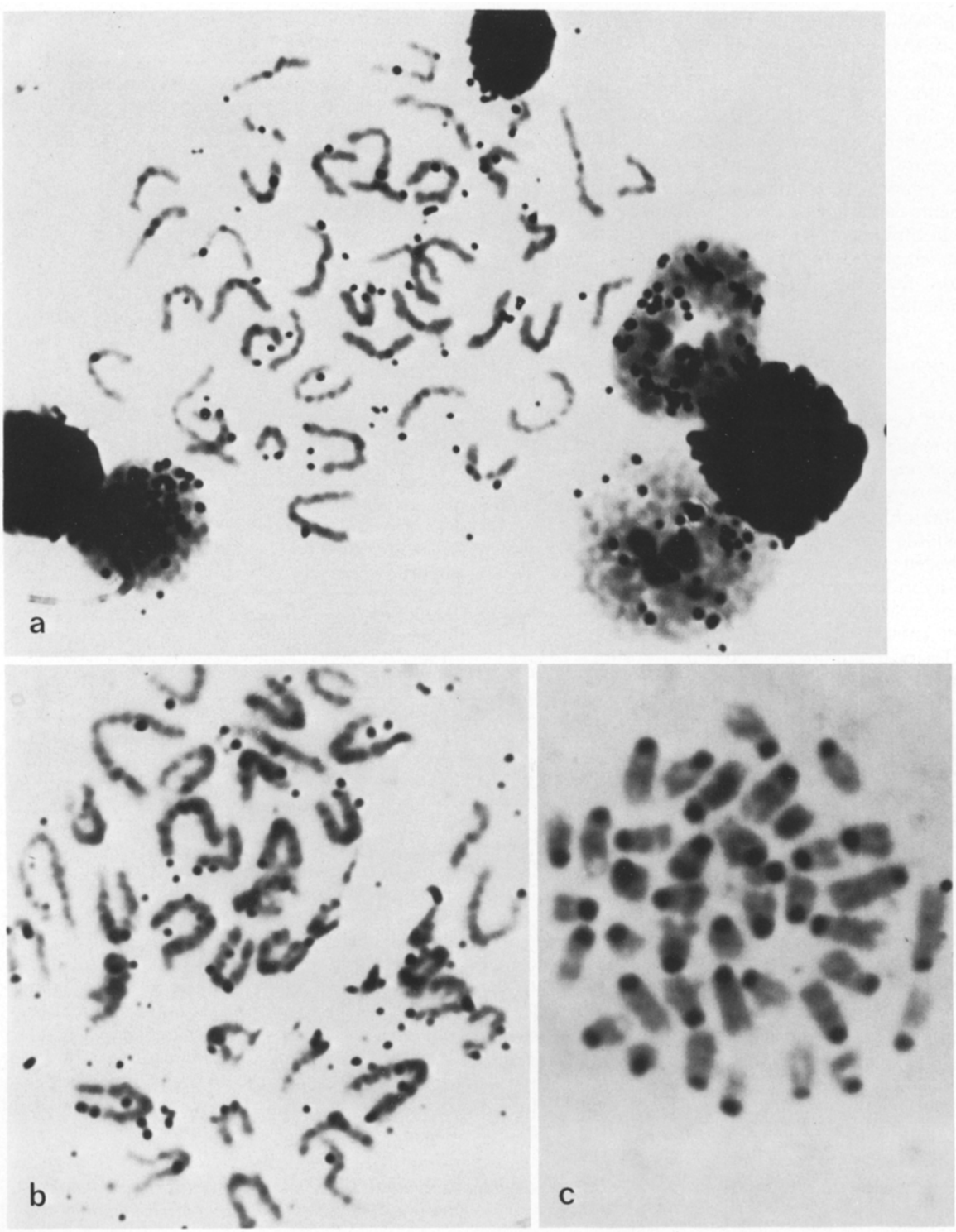

Fig. 4a–c. Oligonucleotide hybridization in situ to mouse metaphase chromosomes. Standard mouse chromosome preparations were challenged by the radiolabelled oligonucleotide probes indicated below. After exposure to the film emulsion, the standard G-banding technique was used to identify the chromosomes. Note the homogeneous distribution of silver grains with both GATA and GACA sqr as well as the absence of signal with the unrelated oligonucleotide. Probes and temperature of hybridization: **a** 5′-$(GATA)_4$; 35° C, **b** 5′-$(GACA)_4$; 43° C, **c** 5′-AATTCCGTATCGATGC; 41° C

10^4 times per haploid genome (see Table 1), specific hybridization signals should, nevertheless, be obtained in due exposure time.

As can be seen in Figure 4, the $(GATA)_4$ and $(GACA)_4$ oligonucleotide probes indeed show specific labelling over the mouse chromosomes (and interphase nuclei). A survey of several hundred metaphases revealed random distribution of silver grains on the chromosomes. Moreover no increased signal density was observed in the pericentric region of the Y chromosome. This finding stands in clear

contrast to earlier reports where the conventional cloned or uncloned probe had been used. Hence, it seems possible that *short* autosomal sqr far outnumber long Y chromosomal sqr thus preventing sex-specific hybridization patterns. This explanation is in agreement with the finding that there is no quantitative difference detectable by the hybridization of sqr probes to male and female mouse DNA (Ali and Epplen, unpublished). Sex-specific hybridization patterns are obvious only after digestion of mouse DNA with certain restriction enzymes and gel electrophoretic separation of the resulting DNA fragments (see Fig. 1).

Discussion

Simple quadruplet GATA/GACA tandem repeats occur in every eukaryotic species investigated thus far by hybridization analysis. These sqr have been identified at the DNA sequence level in *Drosophila* (Singh et al. 1984; Levinson and Epplen, unpublished), *Elaphe radiata* (Epplen et al. 1982), mouse (Epplen et al. 1983b; Singh and Jones 1984) and rat (Alonso et al. 1983). In *Elaphe radiata,* both types of sqr are found in the DNA of a sex-specific satellite. GACA is also adjacent to GATA in a mouse cDNA clone (pmc 14) isolated from a thymoma library (Epplen et al. 1983a). The results of the library screenings here also show intensive interspersion of both sqr to a significantly higher extent than expected from random distribution. This conserved linkage can be explained because both sqr could easily have evolved from each other by an initial single transition mutation and subsequent concerted amplification events.

In the isolated phage clones the sqr stretches show a considerable length variability in the range of about 20 to more than 120 sqr, 28 sqr being prevalent and 60 sqr being average in the phage panel. Similar diversity has been reported to be a general feature of certain simple tandemly repeated sequence arrays in eukaryotes (Jeffreys et al. 1985): They may lead to arrays of short repeats which are highly polymorphic within, for example, a human population. This diversity might have been produced by unequal crossing over or gene conversion like events or by slipped strand mispairing. The latter mechanism is thought to result in hot spots for mutation in the presence of repeated sequences (Jones and Kafatos 1982). Theoretically the abundant short sqr that have been observed in our library screenings might be precursors of longer stretches or conversely might have evolved from the latter. Longer stretches of sqr might have been formed preferentially in evolution in certain chromosomal areas, e.g. in a region with reduced crossing over rates on the Y chromosome. Indeed, there is a tendency towards longer sqr stretches in phages containing mouse Y-chromosomal fragments compared to the other phages. This is reflected in our gel hybridization data with the $(GA^T_CA)_4$ oligonucleotide probes which reveal sex-specific organization of the sqr.

Comparison of satellite DNAs of different species usually reveals clear distinctions between the sequence of one species and that of another. This fact reflects the high rate of change in satellite sequences, which therefore must be faster than the rate of speciation. The evolutionary conservation of significant stretches of sqr sets them apart from other satellite DNAs and could be due to two mechanisms (which are not mutually exclusive): (a) The speed of sequence homogenization or de novo generation of sqr could be consistently far greater than the speed of degeneration. (b) Sqr may be maintained because of natural selective forces. The latter mechanism implies that GATA/GACA sqr should not be categorized as junk DNA (Epplen and Ohno 1984) but may exert some as yet, unknown function. In this context, these sqr sequences are particularly interesting due to the linkage of GATA sqr to the heterogametic sex over long evolutionary distances ($>3\times10^8$ years) and independent of male or female heterogameity. It is tempting to assume that sqr were originally the means of separating an autosomal chromosome pair, allowing it to develop into heteromorphic sex chromosomes.

The persistent sex-specific arrangement of sqr has provoked questions about fortuitous linkage to sex-determining genes or a more direct causal relationship to the formation of the gonads in the heterogametic sex. Linkage of sqr to the murine Sxr factor is confirmed beyond doubt by gel hybridization (Fig. 1) since specific oligonucleotide probes (without any flanking sequences) were employed. Aside from their preferential Y-chromosomal location in the mouse, long GATA sqr have been found to be strongly represented also on chromosome 17 of the mouse and, even there, have been invoked to play a role in sex-determination (Kiel-Metzger and Erickson 1984). Furthermore, in contrast to a recent report (Kiel-Metzger et al. 1985), we have been able to demonstrate clearly the presence of GATA sqr by oligonucleotide probing also in *human* Y chromosomal cosmid clones (Arnemann et al. 1986). Though not being conclusive, this finding at least does not contradict the possibility that sqr remain linked in evolution to a Y-chromosomal sex-determining factor in mammals.

Clonable sqr are always interspersed with other repetitive DNAs or single-copy sequences and therefore do not exhibit the previously expected, extremely long stretches of tandem repetition typically located, for example, in the centromeric heterochromatin (Novak 1984). In addition, for the first time, by virtue of in situ hybridization with oligonucleotides, ubiquitous genomic interspersion of simple repeats was demonstrated. Unlike other tandemly repeated DNA, these sqr are transcribed and appear in poly(A)$^+$ RNA preparations (see Schäfer et al. 1986). Several isolated cDNA clones are endowed with closely related inverted repeats flanking the sqr regions. Sequence characteristics of the inverted repeats are shared with eukaryotic transposons which almost always end in 5′ TG...CA3′ (Chen and Barker 1984). Thus a transposon-like activity might be envisioned as an elegant means of spreading sqr throughout the mouse genome and explain their complex organization.

Acknowledgements. We are grateful to R. Bruce Wallace for the libraries and reading of an early version of the manuscript which was kindly typed by Mrs. R. Schneider. SA was the recipient of a fellowship form the Alexander von Humboldt-foundation and JTE thanks the Max-Planck-Gesellschaft for continued support.

References

Alonso A, Kühn B, Fischer J (1983) A unusual accumulation of repetitive sequences in the rat genome. Gene 26:303–306

Arnemann J, Jakubiczka S, Schmidtke J, Schäfer R, Epplen JT (in press) Clustered BKM DNA sequences on the human Y chromosome. Hum Genet

Benton WD, Davis RW (1977) Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196:180–182

Biggin MD, Gibson TJ, Hong GF (1983) Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc Natl Acad Sci USA 80:3963–3965

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

Chen HR, Barker WC (1984) Nucleotide sequences of the retroviral long terminal repeats and their adjacent regions. Nucleic Acids Res 12:1767–1778

Epplen JT, Ohno S (1984) On structural and functional aspects of single-copy and interspersed repetitive DNA. In: Sparkes RS, de la Cruz FF (eds) Research perspectives in cytogenetics. University Park Press Baltimore pp 17–34

Epplen JT, McCarrey JR, Sutou S, Ohno S (1982) Base sequence of a cloned snake W-chromosome DNA 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 repetitive DNA sequences: Do "sex-specific" satellite components serve any sequence dependent function? Differentiation 23 (Suppl):60–63

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–73

Jones CW, Kafatos FC (1982) Accepted mutations in a gene family: evolutionary diversification of duplicated DNA. J Mol Evol 19:87–103

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

Kiel-Metzger K, Warren G, Wilson GN, Erickson RP (1985) Evidence that the human Y chromosome does not contain clustered DNA sequences (Bkm) associated with heterogametic sex determination in other vertebrates. N Engl J Med 313:242–245

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Cold Spring Harbor, NY, pp 387–389

Nesbitt MN, Francke U (1973) A system of nomenclature for band patterns of mouse chromosomes. Chromosoma 41:145–158

Novak U (1984) Structure and properties of a highly repetitive DNA sequence in sheep. Nucleic Acids Res 12:2343–2350

Rigby PWJ, Dieckmann M, Rhodes C, Berg P (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J Mol Biol 113:237–251

Schäfer R, Böltz E, Becker A, Bartels F, Epplen JT (1986) On the expression of the evolutionarily conserved GATA/GACA repeats in mouse tissues. Chromosoma 93:496–501

Schöld M, Colombero A, Reyes AA, Wallace RB (1984) Laboratory methods. Oligonucleotide – directed mutagenesis using plasmid DNA templates and two primers. DNA 3:469–477

Singh L, Jones KW (1982) Sex reversal in the mouse (Mus musculus) is caused by a recurrent nonreciprocal crossover involving the X and an aberrant Y chromosome. Cell 28:205–216

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

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

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Tsao SGS, Brunk CF, Pearlman RE (1983) Hybridization of nucleic acids directly in agarose gels. Anal Biochem 131:365–372

Received May 13, 1985 / in revised form February 6, 1986
Accepted by W. Beermann