Volume 17 Number 23 1989 Nucleic Acids Research

The XLR sequence family: dispersion on the X and Y chromosomes of a large set of closely related sequences, most of which are pseudogenes

Henri-Jean Garchon[1,2], Elwyn Loh[1], William Y.Ho[1], Laurence Amar[3], Philip Avner[3] and Mark M.Davis[1]

[1]Department of Microbiology and Immunology, Stanford University School of Medicine and The Howard Hughes Medical Institute, Stanford, California, USA, [2]INSERM U 25, Hôpital Necker, Paris and [3]Unité d'Immunogénétique Humaine. Institut Pasteur, Paris, France

Received February 13, 1989; Revised October 16, 1989; Accepted October 27, 1989

ABSTRACT

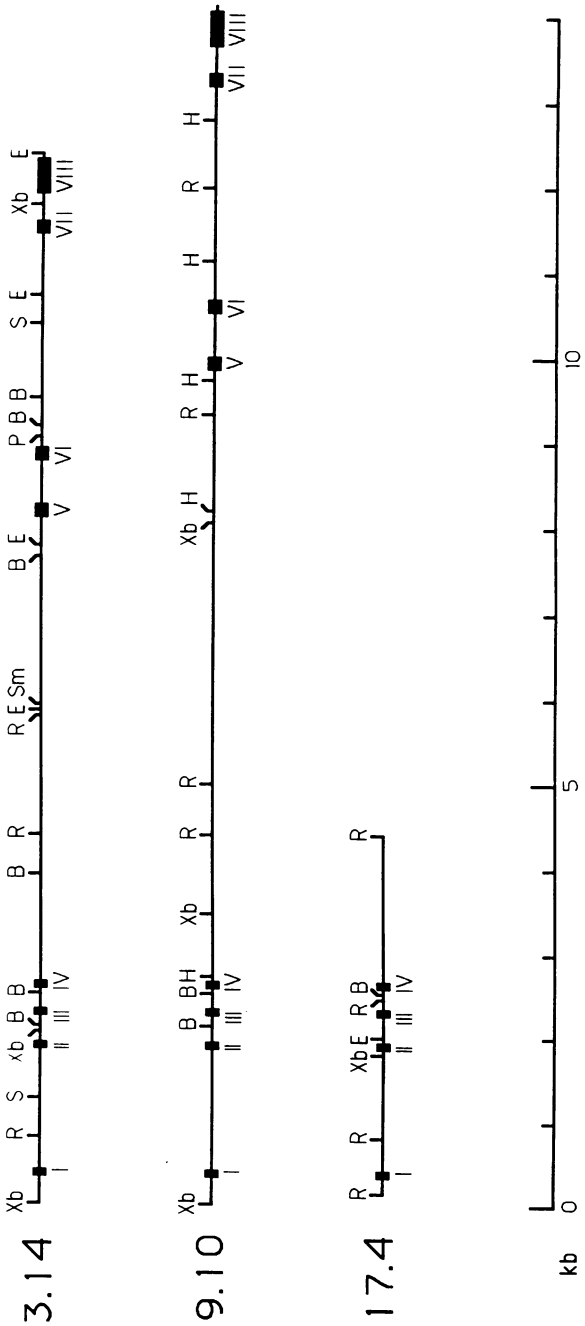
The XLR sequence family encodes RNA transcripts specific to late-stage T and B cells and their neoplasms. Only one apparently functional mRNA has been identified thus far and this encodes a novel 25 kDa nuclear protein. In this report, we find that the XLR gene family is composed of 50−75 copies per haploid genome which localize to at least two different portions of the mouse X chromosome. Neither of these locations are near the *xid* mutation that earlier work had correlated with XLR. In addition, some members of this family are also on the Y chromosome. Another surprising finding is that while the fourteen genomic clones examined to date have the same exon-intron structure and are closely related with respect to sequence conservation (90%), all appear (in most cases by multiple criteria) to be non-functional, raising the possibility that all but one of the members of this large semi-dispersed family are pseudogenes.

INTRODUCTION

The murine XLR (X-linked, Lymphocyte-Regulated) sequence family was first identified using subtractive cDNA probing and cloning methodologies and localized to the X-chromosome (1). Transcripts encoded by this family are predominantly found in lymphoid cells. In transformed cell lines representing different stages of B lymphocyte differentiation, XLR expression is restricted to the most mature stages of differentiation, coincident with immunoglobulin secretion, J chain synthesis and disappearance of the 14G8/BLA-2 membrane marker (2). Expression also seems to be stage-specific in the T-cell lineage (3). In B cells, the stage of differentiation at which XLR first appears coincides with the most pronounced effects of the *xid* mutation, an X-linked immune deficiency which seems to interfere with B-cell maturation (4−6). Comparisons of DNA from *xid* congenic mice and their parental strains with an XLR probe revealed several polymorphisms showing linkage between at least one of the XLR sequences and the *xid* locus (1).

Although this family is composed of many members, only one potentially functional transcript, pM1, has been isolated thus far from a series of cDNA libraries derived from thymus or spleen cells or two different B cell lines (3). pM1 belongs to the class of rare messengers (.002% in a plasma cell cDNA library) and encodes an open reading frame that predicts the existence of a 208 amino acid protein with an unmodified molecular weight of 25 kDa. This protein has been shown to be the *in vivo* product of the gene since antibodies raised against the predicted pM1 gene product expressed in bacteria recognize a nuclear protein in murine lymphoid cell lines (7). This protein could well play a important regulatory role in lymphocyte differentiation.

While the antibodies recognizing the XLR protein enable one to study the expression of XLR, little is known about the structure of the XLR sequence complex. Of particular


3.14
Xb R S Xb B B B R R E Sm B E P B B S E Xb E
I II III IV V VI VII VIII
9.10
Xb B BH Xb R R R H H R H
I II III IV V VI VII VIII
17.4
R R XbE R B R
I II III IV
kb 0 5 10

interest are the following questions : Is there more than one functional XLR gene ? What is the nature of the linkage between the XLR family, the functional pM1 gene and the *xid* mutation? It is also important to know the exon-intron structure of a typical XLR gene to see how it compares with others as well as aiding in the study of its regulation and its expression. In the present paper, we describe the general structure of members of this family with particular reference to sequence similarities, exon-intron organization and chromosomal localization. These data indicate that XLR is in a number of respects unlike any previously described gene family.

MATERIAL AND METHODS

The cosmid library was made by Mbo I partials of liver DNA from the CBA/Ca strain of mice by standard methods (8) using the Bates and Swift vector (9). The liver phage λJ1 library (generously provided by Dr Y. Chien and described in ref. 10) is also derived from Mbo I partials, but using liver DNA from the B10.A strain of mice. DNA probes were labeled using the random hexamer priming technique (11). Screening of libraries, subcloning of DNA fragments, Southern blot hybridization were done using classical procedures (8). Nucleotide sequencing was performed by the dideoxynucleotide chain termination method (12).

Exons I, IV and V of cosmid clones (Figures 3 and 4) were sequenced directly with double-stranded cosmid DNA (13) using conserved oligonucleotide primers for exon I (5'-ACGGCCAATCAGCAAGGGGCTTG-3') for exon IV (5'-TGTTTGTAGG-ACAAAACTGAGC-3'), and for exon V (5'-CCTTGGCTAGTCACATAAAG-3'). For each cosmid clone, sequences of reverse strands were derived from PCR-amplified (14) 160 bp fragments surrounding each exon using the following additional primers : for exon 1, 5'-CTTCAACAGGTCACTGAC-3', for exon IV, 5'-ATATAAATTATGTTAATAC-3' and for exon V, 5'-AAATGATACAATTGTGAC-3'. Amplified DNA fragments were then separated from primers and nucleotides by spermine precipitation (15) and sequenced using modified T7 DNA polymerase (16).

Mouse strains, interspecific cross construction and the other techniques used for the mapping of the XLR sequences are as described previously (17, 18). The deletion hybrid panel has been described in Avner et al. (19). Homology searches were done using computer programs of the University of Wisconsin genetics computer group (20).

RESULTS

An initial set of ten cosmid clones picked from a CBA/Ca liver library using a full-length XLR cDNA probe was analyzed by restriction mapping and cDNA subclone probing. The results (data not shown) indicated that, although closely related to each other, all the cosmids were different and non-overlapping. We chose to analyze two of them, 3.14 and 1.37, in greater detail as they appeared the most closely related to the pM1 cDNA. Sequencing data obtained in the course of their characterization, revealed that segments of these clones which would appear later to correspond to exons III and VIII had diverged markedly from

Figure 1. Restriction maps of cosmid clone 3.14 and λJ1 phage clones 9.10 and 17.4, showing conserved exon-intron organization; B: Bgl II; E: EcoRV; H: Hind III; P: Pst I; R: EcoRI; S: SacI ; Sm: SmaI; Xb: XbaI. Roman numerals refer to exons as defined in Fig. 2.

the pM1 sequence. Another mouse genomic library in phage λJ1 (10) was therefore screened with the pM1 cDNA probe and about 400 positive clones per million recombinant phages were obtained. Of these, 55 were rescreened using oligonucleotide probes matching the pM1 sequence and five clones were chosen for further analysis since they were the most closely related to pM1 on the basis of thermal stability of the hybridization. Restriction fragment analysis showed that three of the five clones were overlapping, leaving a total of three independent genomic clones, 9.10, 17.4 and 9.7, for detailed study. For reasons of clarity, the results concerning these five different genomic clones (9.10, 17.4, 9.7 from phage and 3.14 and 1.37 from cosmid) are presented together.

Exon-intron structure of XLR genes

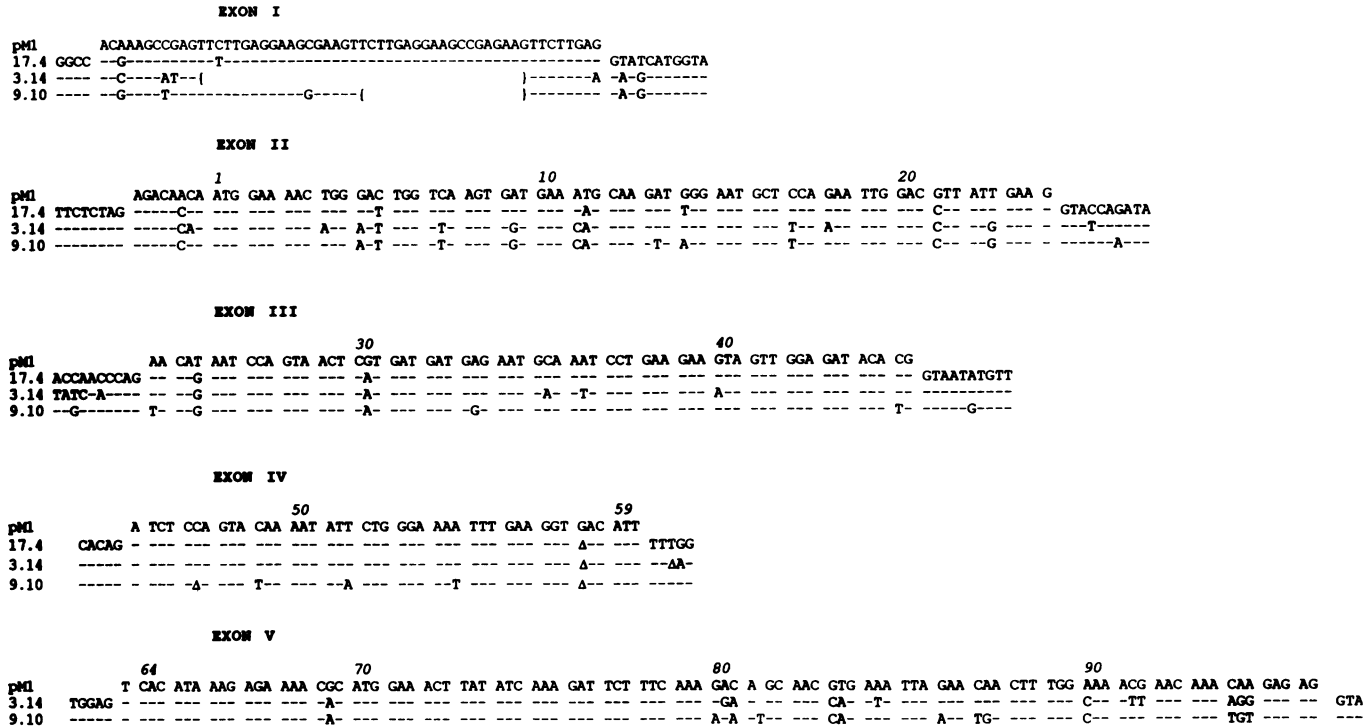
Restriction site mapping and sequencing of fragments hybridizing to the pM1 cDNA show that all five clones have a common exon-intron organization. Figure 1 shows detailed maps of 3.14, 9.10, and 17.4. Clones 3.14 and 9.10 contain a complete XLR gene, as defined by the full-length pM1 cDNA split into eight exons. The first four exons are clustered into 2.5 kb and are separated from the four last exons by a 5−7 kb intron. The first four exons are smaller (58, 78, 64 and 44 bp respectively) than the last four exons (100, 99, 100 and 396 bp). The sequences of the segments of the five clones hybridizing to the pM1 cDNA are shown in Figure 2. While 9.10 and 3.14 contain all eight exons, 17.4 has only the four first exons (as shown in Figure 1), 1.37 only the last two exons and 9.7 only exon VIII. The entire length of the pM1 cDNA has been matched to genomic sequences with the exception of eleven nucleotides located downstream of exon IV and encompassing codons corresponding to 60−63 of the pM1 open reading frame. In fact, the 3′ border of exon IV does not show canonical splice donor sequences (21) and has therefore been assigned solely on the basis of its sequence similarity to the cDNA. There is however no ambiguity at the 5′ end of exon V that shows a regular splice acceptor sequence. The splice donor that is attached to the 3′ end of exon I is also altered in two of the clones, 9.10 and 3.14. The third clone, 17.4, follows the canonical sequence and shows no gap at the junction between exons I and II. With these exceptions, splice signals match the consensus sequences.

Since the initiation codon is located right after the beginning of exon II, the first exon must be untranslated. The termination codon lies within the last exon.

XLR sequences are diverse but closely related

XLR sequences not only have a common exon-intron organization but also are clearly related at the nucleotide sequence level. Table I shows comparisons of the sequences of various subclones analyzed in the course of this study. These comparisons bear both on successive exons of the various XLR clones and on their flanking regions. Homology is equal to or greater than 89%. Similar levels of homology are obtained irrespective of whether exons or their flanking sequences are considered (data not shown). The highest level of homology observed is 98.2% between exons VII and VIII for the clones 9.10 and 1.37. This includes a 964 bp region of which only 237 bp is coding sequence. This result suggests that the genes have diverged only very recently .

Examination of the sequences (Figure 2) reveals that they differ from each other and from the pM1 sequence mostly by point mutations although some deletions have occurred as in exons I, VI, and VIII. Deletions in exon I are of particular interest and are discussed in more detail later on. The mutational variations appear to be scattered throughout the sequences and are equally frequent in both the coding and untranslated regions. Not only are several positions mutated in all the clones shown in Figure 2 but some of the positions


EXON I
pM1 ACAAAGCCGAGTTCTTGAGGAAGCGAAGTTCTTGAGGAAGCCGAGAAGTTCTTGAG
17.4 GGCC --G----------T---------- GTATCATGGTA
3.14 ---- --C----AT--( )-------A -A-G-------
9.10 ---- --G----T----------G-----( )-------- -A-G-------
EXON II
1 10 20
pM1 AGACAACA ATG GAA AAC TGG GAC TGG TCA AGT GAT GAA ATG CAA GAT GGG AAT GCT CCA GAA TTG GAC GTT ATT GAA G
17.4 TTCTCTAG -----C-- --- --- --- --- --T --- --- --- --- --- -A- --- --- T-- --- --- --- --- --- --- C-- --- --- - GTACCAGATA
3.14 --------- -----CA- --- --- --- A-- A-T --- -T- --- -G- --- CA- --- --- --- --- --- T-- A-- --- --- C-- --G --- - ---T------
9.10 --------- -----C-- --- --- --- --- A-T --- -T- --- -G- --- CA- --- -T- A-- --- --- T-- --- --- --- C-- --G --- - ------A---
EXON III
30 40
pM1 AA CAT AAT CCA GTA ACT CGT GAT GAT GAG AAT GCA AAT CCT GAA GAA GTA GTT GGA GAT ACA CG
17.4 ACCAACCCAG -- --G --- --- --- --- -A- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -- GTAATATGTT
3.14 TATC-A----- -- --G --- --- --- --- -A- --- --- --- --- -A- -T- --- --- --- A-- --- --- --- --- --- -- ----------
9.10 --G------- T- --G --- --- --- --- -A- --- --- -G- --- --- --- --- --- --- --- --- --- --- --- --- T- -----G----
EXON IV
50 59
pM1 A TCT CCA GTA CAA AAT ATT CTG GGA AAA TTT GAA GGT GAC ATT
17.4 CACAG - --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- Δ-- --- TTTGG
3.14 ----- - --- --- --- --- --- --- --- --- --- --- --- --- --- --- Δ-- --- --ΔA-
9.10 ----- - --- -Δ- --- T-- --- --A --- --- --T --- --- --- Δ-- --- -----
EXON V
64 70 80 90
pM1 T CAC ATA AAG AGA AAA CGC ATG GAA ACT TAT ATC AAA GAT TCT TTC AAA GAC A GC AAC GTG AAA TTA GAA CAA CTT TGG AAA ACG AAC AAA CAA GAG AG
3.14 TGGAG - --- --- --- --- --- -A- --- --- --- --- --- --- --- --- --- --- -GA - -- --- CA- -T- --- --- --- --- --- C-- -TT --- --- AGG --- -- GTA
9.10 ----- - --- --- --- --- --- -A- --- --- --- --- --- --- --- --- --- --- A-A -T-- --- CA- --- --- A-- TG- --- --- C-- --- --- --- TGT --- -- ---

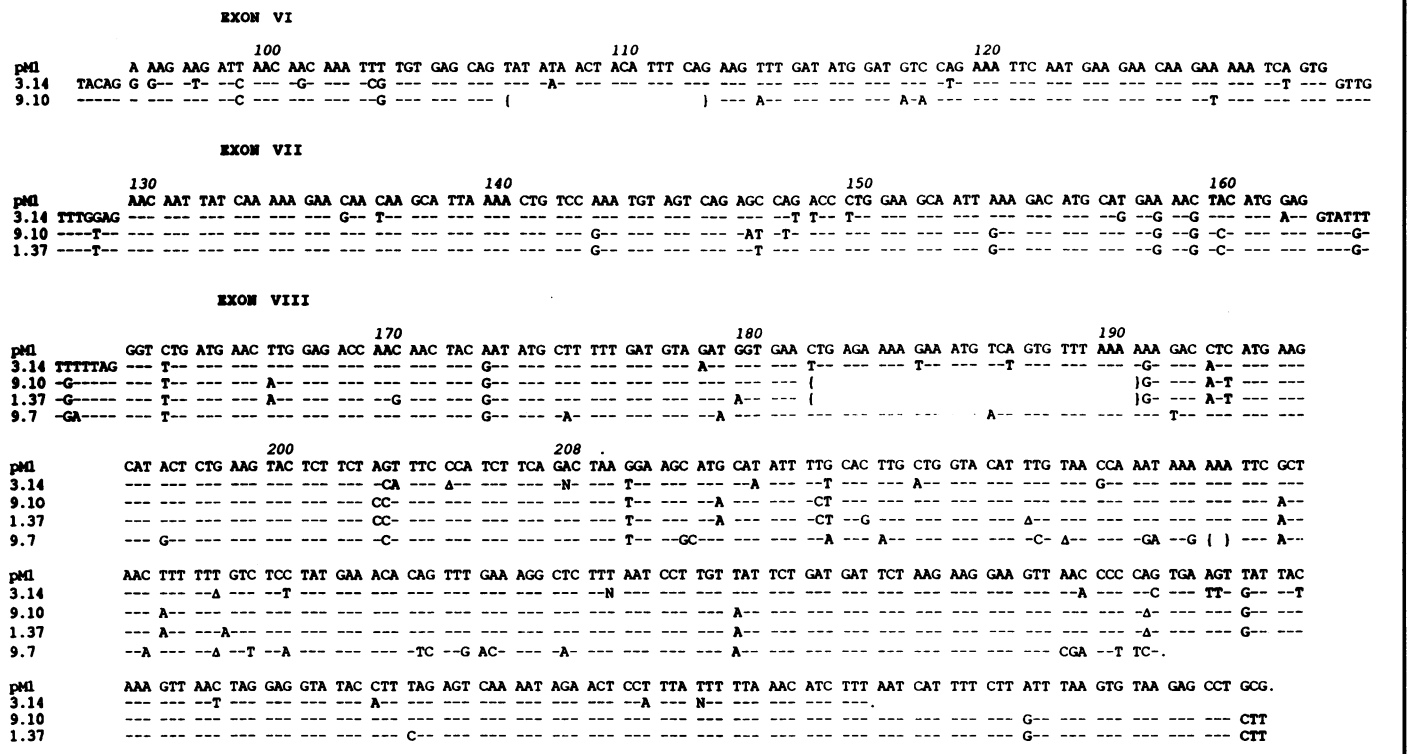

EXON VI
EXON VII
EXON VIII
pM1
3.14
9.10
1.37
9.7

Table I – Sequence comparisons of various segments of XLR genes

Origin of subclones	Exons	Total length of exons (bp)	Length of regions compared (bp)	Percentage of homology (%)
3.14–17.4	I	58	669	90.3
9.10–3.14	II,III	141	897	91
17.4–9.10	II, III	141	874	91
17.4–3.14	II,III,IV	199	1006	89.5
9.10–3.14	V,VI	199	1248	89
9.10–3.14	VII	99	723	89.1
1.37–9.10	VII, VIII	495 (237*)	964	98.2

* length of coding region.

undergo an identical change from the pM1 consensus sequence in all the clones examined. It is clear that some of these sequence changes may alter the functional capabilities of the genes. The A to C mutation at position −3 upstream of the AUG codon in exon II leaves, for example, the sequence at odds with the consensus sequence described by M. Kozak (22) and this could decrease translation efficiency. Similarly, the deletion of a G in codon 58 of exon IV introduces a shift in the pM1 reading frame soon after initiation. Among other noteworthy modifications is the CA dinucleotide substitution at codon 83. This substitution is actually surrounded by several other alterations including a single insertion which represents the only insertion seen within the pM1 coding region.

To determine the extent of these modifications in other members of the XLR family, we sequenced the regions corresponding to exons IV and V in other cosmid clones as described in Material and Methods. Figure 3 shows the result of this analysis. The sequences are aligned according to their relatedness. All of the genomic clones in the case of exon IV and 7 out of 8 of the clones in the case of exon V could thus be amplified and sequenced. This confirms the high level of conservation of the flanking regions. It also suggests that these additional XLR genes have a common exon-intron organization with the XLR sequences already described. All the exon IV sequences examined showed the loss of a G in codon 58 with the concomittant introduction of a frameshift. Similarly, all of the clones show a CA dinucleotide substitution at codon 83 of the pM1 reading frame. Remarkably, several positions in exon V are altered either in all the clones (codons 80 and 94) or in most of them (codons 87 and 90). Codon 81 has frequent mutations by insertion as this was observed in two thirds of the clones examined.

Most XLR sequences are non-functional

As mentioned earlier, a shift in the pM1 reading frame occurs close to the initiation site in all clones analyzed. This frameshift also introduces a stop codon into exon V of all the sequences examined so far : TAA or TAG at codon 85 of the pM1 reading frame for clones 2.8, 2.31 and 3.14 ; TGA at codon 87 for clones 2.38A or 3.11. The 3.12 and 9.10 clones moreover contain stop codons in all three reading frames of exon V. Such termination may well occur even earlier in many of these genes: at codon 45 in 9.10 or at codon 49 for 3.15, 1.36B and 2.31. Taken overall, it is clear that all the XLR clones presented here are probably pseudogenes.

Figure 2. Sequences of XLR exons compared to the pM1 sequence. The numbers above the pM1 sequence refer to the amino acids of the pM1 reading frame. Single base pair deletions are indicated by Δ, larger deletions by brackets.

EXON IV

pM1	A	TCT	CCA	GTA	CAA	AAT	ATT	CTG	GGA	AAA	TTT	GAA	GGT	GAC	ATT	
		50													*59*	
9.10	CACAG	-	---	-A-	---	T--	---	--A	---	---	--T	---	---	---	Δ--	---	TTTGG
17.4	-----	-	---	---	---	---	---	---	---	---	---	---	---	---	Δ--	---	-----
3.14	-----	-	---	---	---	---	---	---	---	---	---	---	---	---	Δ--	---	--ΔA-
2.8	-----	-	-A-	---	---	---	---	---	---	---	---	---	---	---	Δ--	---	--ΔA-
1.36A	-----	-	-A-	---	---	---	---	---	---	---	---	---	---	--A	Δ--	---	--ΔA-
2.38A	-----	-	---	---	---	---	---	--A	---	---	--T	---	---	---	Δ--	---	--ΔA-
3.11	-----	-	---	---	---	---	---	--A	---	---	--T	---	---	---	Δ--	---	-----
3.12	-----	-	---	---	---	---	---	--A	---	---	--T	---	---	---	Δ--	---	-----
3.15	-----	-	---	---	---	T--	---	--A	---	---	--T	---	---	---	Δ--	---	-----
1.36B	-----	-	---	---	---	T--	---	--A	---	---	--T	---	---	---	Δ--	---	-----
2.31	-----	-	---	---	---	T--	---	--A	---	---	--T	---	---	---	Δ--	---	--Δ--

EXON V

pM1	CGC	ATG	GAA	ACT	TAT	ATC	AAA	GAT	TCT	TTC	AAA	GAC	A		GC	AAC	GTG	AAA	TTA	GAA	CAA	CTT	TGG	AAA	ACG	AAC	AAA	CAA	GAG	AG	
	70											*80*												*90*							
3.14	-A-	---	---	---	---	---	---	---	---	---	---	-GA	-		--	---	CA-	-T-	---	---	---	---	---	---	-TT	---	---	AGG	---	--	GTAG
2.8	-A-	---	---	---	---	---	---	---	---	---	---	-GA	-		--	---	CA-	--G	---	---	---	---	---	C--	-T-	---	---	AGG	---	--	-----
1.36B	-A-	---	---	---	---	---	---	---	---	---	---	A-A	T		--	---	CA-	---	---	---	TG-	---	---	C--	---	---	---	TGG	---	--	-----
2.38A	-A-	---	---	---	---	---	---	---	---	---	---	A-A	-A		--	---	CA-	---	---	---	TG-	G--	---	C--	---	---	---	TGG	---	--	-----
3.11	-A-	---	---	---	---	---	---	---	---	---	---	A-A	-A		--	---	CA-	---	---	---	TG-	---	---	C--	--A	---	---	TGG	---	--	-----
3.12	-A-	---	---	---	---	---	---	---	---	---	---	A-A	-T		--	---	CA-	---	---	A--	TG-	---	---	C--	---	---	---	TGT	---	--	-----
9.10	-A-	---	---	---	---	---	---	---	---	---	---	A-A	-T		--	---	CA-	---	---	A--	TG-	---	---	C--	---	---	---	TGT	---	--	-----
2.31	-A-	---	---	---	---	---	---	---	---	---	---	A-A	-AAT	--	---	CA-	---	---	---	TG-	---	---	---	C--	---	---	---	TGG	---	--	-----
3.15	-A-	---	---	---	---	---	-C-	---	---	---	---	A-A	-AAT	--	---	CA-	---	---	---	TG-	---	---	---	C--	---	---	---	TGG	---	--	-----

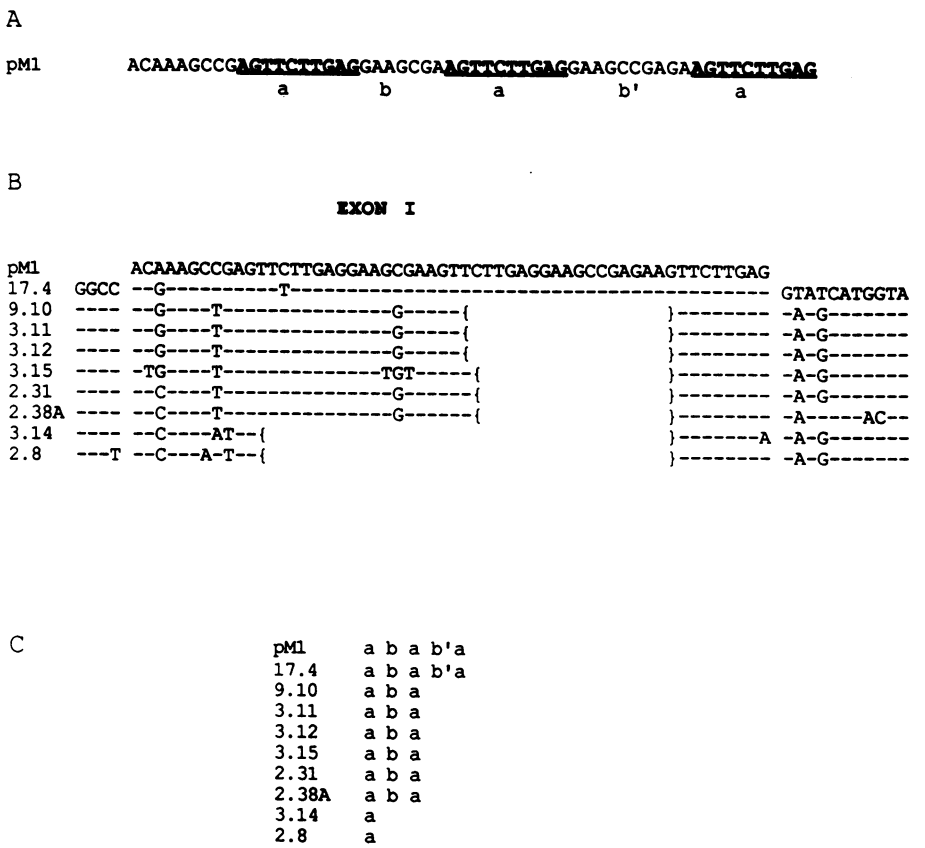


Figure 4. Repeated motifs in exon I are subject to deletions by homologous recombination. A. The 5′ untranslated region of the pM1 transcript shows a decanucleotide motif *a* repeated three times. B. Sequences of genomic clones in their segments homologous to exon I. C. Motif structure of these clones showing the deletion of the decamer repeat.

Decanucleotide repeats in exon I are subject to deletions and duplications

The untranslated region of the pM1 cDNA, and of the corresponding genomic sequences for exon I, indicates a particularly fluid series of expansions and/or contractions within this portion of the gene. As shown in Figure 4A, the untranslated region of pM1 contains a decanucleotide, AGTTCTTGAG, that is perfectly repeated three times. These core motifs are intermingled with two equally short and closely related sequences, b and b′, giving an overall structure *abab'a*. Examination of the exon I sequences available initially (Figure

Figure 3. Sequences of exons IV and V. Sequences of available cosmid clones were determined for the regions related to exons IV and V as described in Material and Methods. Blank spaces in codon 81 are due to insertions in the corresponding genomic sequences.

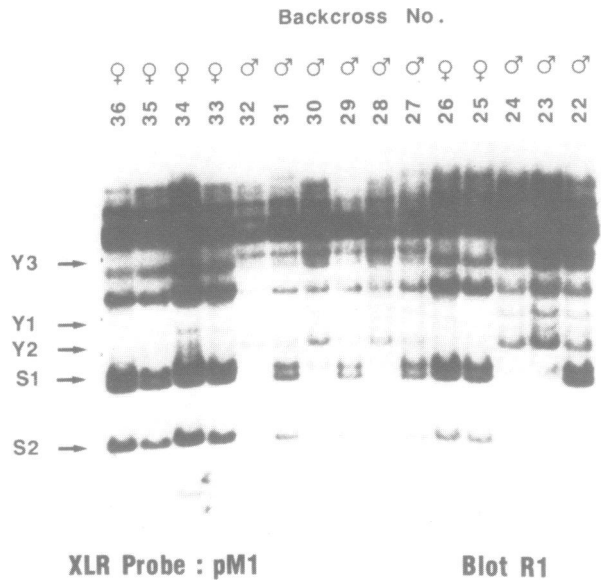


Figure 5. Interspecific backcross. Numbers refers to particular backcross progeny. DNA was cut with EcoRI and probed with pM1. Bands S1 and S2 are *M. spretus*-derived X chromosome bands. Y1, Y2 and Y3 are male-specific bands.

2) revealed deletions in two clones resulting in the loss of either one or two of the core repeats and the intervening oligomers, leaving a sequence having the structure, *aba* (clone 9.10) or *a* only (clone 3.14). Exon I sequences from additional XLR clones were obtained in a manner similar to that described previously for exon IV and V (Figure 4B). Although point mutations are also present, the sequences clearly fall into one of two categories (Figure 4C), the first being homologous to clone 9.10 (clones 2.31, 2.38A, 3.11, 3.12 and 3.15, *aba* structure) the second to 3.14 (clone 2.8, *a* structure). The only clone so far identified that is not deleted in exon I is 17.4. These deletions are most simply explained by an unequal exchange of DNA between sister chromatids occurring by homologous recombination.

At least two of the XLR sequences are located some 16–21 cM apart on the X chromosome

The large number of XLR sequences also raises the question of their organization on the X chromosome, i.e. whether they are clustered or dispersed on the X chromosome. Two different approaches were taken to this problem. The first one is based on pedigree analysis of an interspecific mouse cross (17, 18). Since *Mus spretus* mice are evolutionarily distant enough from laboratory mice belonging to either the *Mus musculus domesticus* or to the *Mus musculus musculus* sub-species to allow easy detection of RFLPs, the order and the position of successive loci on a given chromosome can then be determined in backcross progeny both by calculating recombination frequencies and by 'pedigree analysis'. This approach has been applied successfully to the mapping of a series of other X-chromosome loci (18, 23).

Figure 5 shows a Southern blot analysis of EcoRI-digested backcross progeny DNAs with the pM1 probe. Two *M. spretus*-derived polymorphic EcoRI fragments can be identified, S1 and S2. The S1 fragment is located distal to the M2C marker (DXPas7)

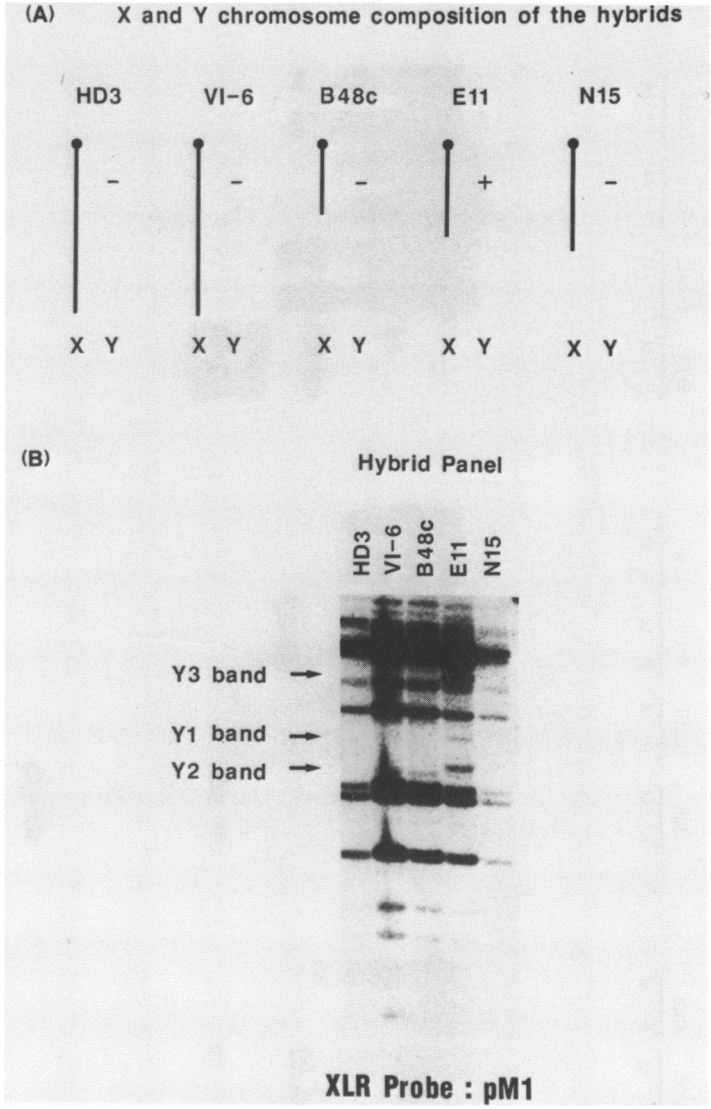
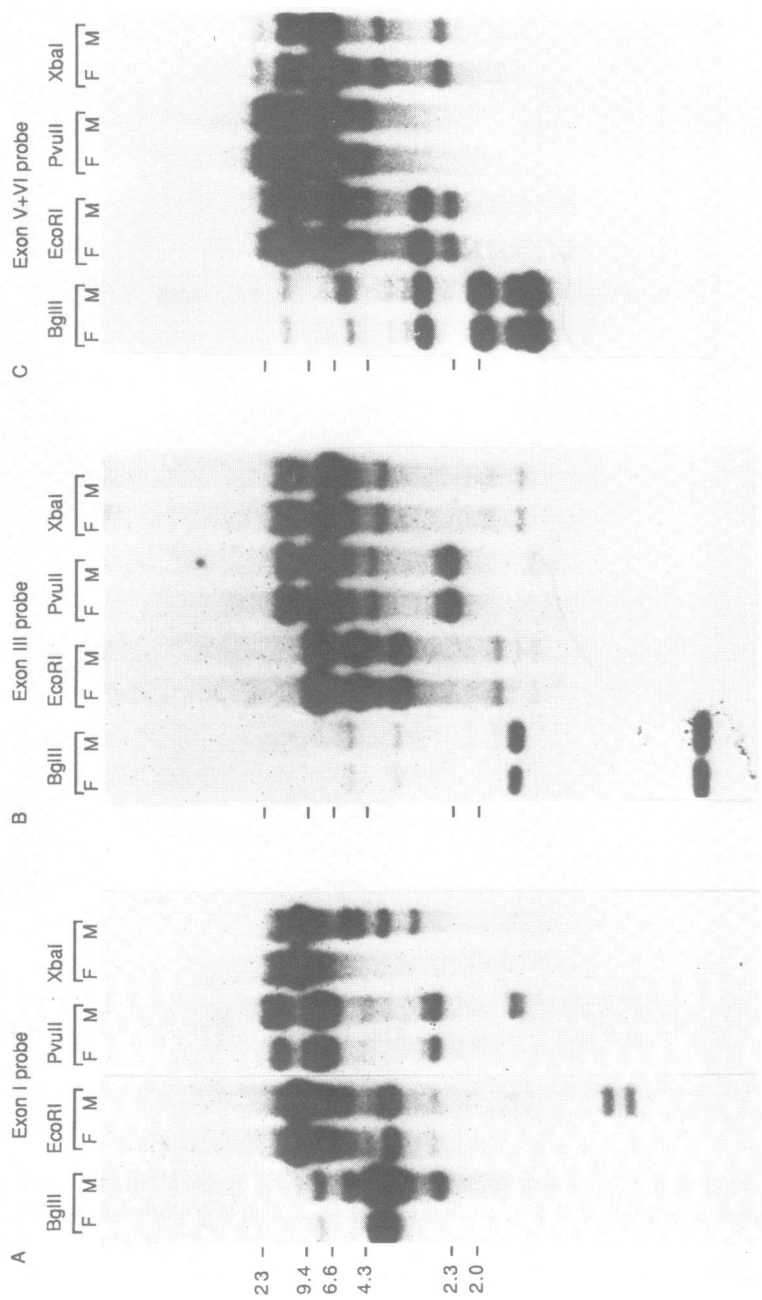


Figure 6. Somatic cell hybrid analysis. A. X and Y chromosome composition of the hybrids. B. Hybridization patterns of EcoRI-digested DNA with the XLR cDNA probe pM1.

and proximal to the Raf locus, a chromosomal span of some 8 cM, thus defining what we propose to call the xlr-l locus. The second fragment, S2, defining the xlr-2 locus is located more distally on the X-chromosome, between the Hprt and the cf-9 (coagulation factor IX) loci. This region comprises about 4 cM. The recombination distance between the xlr-1 and the xlr-2 loci is some 17 cM. Additional segregating sequences, Y1, Y2


A
Exon I probe
BglII
EcoRI
PvuII
XbaI
F M
B
Exon III probe
C
Exon V+VI probe
2.3
9.4
6.6
4.3
2.3
2.0

and Y3 are seen only in male progeny, suggesting the presence of XLR sequences on the Y chromosome (see below).

X-chromosome XLR sequences are located proximal to the T16H translocation breakpoint.

While the progeny analysis allowed us to define two XLR loci on the X chromosome, it did not allow a quantitative idea of the overall distribution of XLR sequences on the X chromosome. This complementary information was obtained by the use of a panel of somatic cell hybrids (19) deleted for various portions of the X-chromosome. These cell lines have been isolated by fusing splenocytes from mice bearing different reciprocal X-autosome translocations with the CH (chinese hamster) cell line. Figure 6 shows an analysis of such hybrids with the pM1 cDNA probe. As previously reported (1), there is no cross-hybridization with hamster DNA (not shown). Comparison of the T16H-derived B48c hybrid cell line containing the proximal part of the mouse X-chromosome (breakpoint between the DMD and MDXS120 loci, some 29−32 cM distal to the centromere) with the mouse HD3 line carrying at least one copy of the entire X, the X only hybrid VI-6 and the more distally broken N15 and E11 hybrid cell lines suggests that all detectable X-linked EcoRI sequences lie within the proximal part of the X present in the B48c hybrid cell line. The great majority, if not all the X-chromosome located XLR sequences, must therefore be proximal to the MDXS120 locus.

Some XLR sequences are linked to the Y chromosome

Data from the backcross progeny analysis suggesting the existence of Y-linked sequences are reinforced by results of the study of somatic cell hybrids (Figure 6). Extra bands not present on X-chromosome carrying hybrids such as VI-6 and B48c are present on hybrid E11 which is known to contain both part of the mouse X-chromosome as well as Y-chromosome sequences. To identify more clearly these Y-linked sequences, DNA from male and female CBA/Ca mice were probed comparatively with genomic XLR subclones (Figure 7). Male-specific polymorphic bands are most evident with the exon I probe (a 1.1 kb fragment surrounding exon I of phage 9.10). Up to five male-specific bands are visible in Pvu II and Xba I digests. Similarly, a 5.5 kb band is visualized in male DNA after Bgl II digestion and hybridization with an exon V+VI probe (a 1.4kb Pst I-Bgl II subclone of cosmid 3.14). Fainter bands (a 2.2 kb Bgl II fragment and a 1.7 kb Pvu II fragment) are also evidenced with the small (a 400 bp Bgl II fragment from phage 9.10) exon III probe. Of note is the low level of polymorphism of the Bgl II sites that surround exon III since most XLR genes show a 0.4 kb or a 1.8kb band.

There are 50−75 XLR sequences per haploid genome

The number of positive clones identified by screening a genomic library suggested the existence of 50 to 100 XLR sequences per genome, a number larger than previously estimated by merely counting bands on a Southern blot (~20, ref. 1). To confirm this estimate, EcoRI-digested genomic DNA was electrophoresed and blotted along with various amounts of two of the genomic probes used above, i.e. mapping either to exon I or to exon V+VI. Each filter was hybridized with the corresponding probe (Figures 8A and 8B). Autoradiographs were analyzed using a scanning densitometer. Areas under the peaks of hybridization of the genomic probes were plotted against the amount of loaded DNA

Figure 7. Some XLR sequences are linked to the Y chromosome. DNA from male (M) and female (F) CBA/Ca mice was cut with various enzymes and was hybridized at 68°C in aqueous solution with indicated genomic probes (see text). Filters were washed at 55°C in 0.2×SSPE, 0.1%SDS and were autoradiographed overnight.

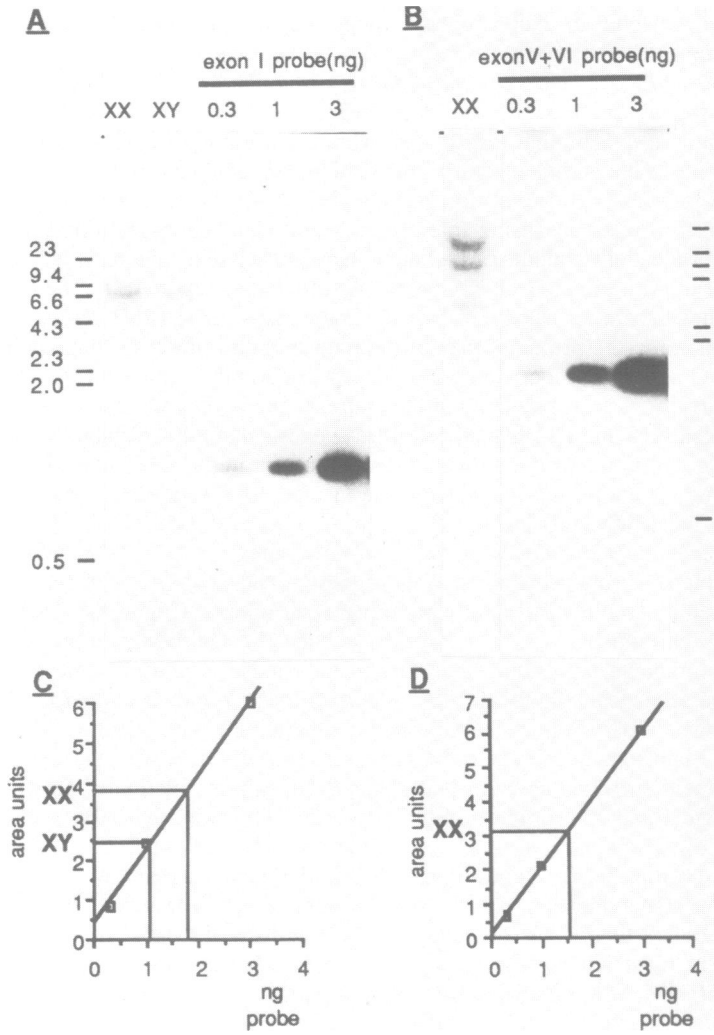


Figure 8. Estimation of the number of XLR sequences. A and B. 20 μg female (XX) or 10 mg male (XY) genomic DNAs were digested with EcoRI, electrophoresed and blotted along with increasing amounts of genomic probes matching either exon I (a 1.1 kb insert in Bluescript vector)(A) or exon V+VI (a 1.4 kb insert in pGEM3)(B). Each blot was hybridized with the corresponding test probe. Washes were at 55°C in 0.2×SSPE, 0.1%SDS. Filters were exposed overnight to Kodak XAR5 film without intensifying screens in order to avoid the saturation of the films. C and D. Autoradiographs were analysed with a scanning densitometer (XL, LKB) and peak areas were plotted against the corresponding amount of DNA for each test probe (C, exon I, and D, exon V+VI). Linear regression curves were fit allowing the number of XLR segments related to each probe to be calculated (see text). Data in C correspond to the autoradiograph shown in A. Autoradigraph shown in B was exposed twice longer than that actually analyzed in D.

(Figures 8C and D) allowing linear regression curves to be fit ($y = 0.440 + 1.860 * x$ for exon I and $y = 0.131 + 1.979 * x$ for V+VI). The hybridization signals given by genomic XX DNAs (20 μg) were summed and corresponding amounts of test probes

were calculated. Assuming a haploid genome size of 3×10^9 bp, copy numbers of 68 and 54 XLR sequences per haploid genome in CBA/Ca female mice mapping to exon I and exon V+VI respectively were thus determined. Similarly, probing of male CBA/Ca DNA (10 μg, XY, figure 8A and C) led to an estimate of 150 copies per total genome, suggesting that a similar number of exon I-like sequences are located on each sex chromosome.

DISCUSSION

The present paper reports the isolation and characterization of XLR sequences as well as data about their chromosomal location. Analysis of total genomic DNA with two different genomic probes led to an estimate of 50−75 XLR sequences per X chromosome, thus confirming the evaluation yielded by screening of a genomic library. Preliminary experiments were suggestive of a high level of similarity among XLR sequences: thus, hybridization of genomic blots at low stringency (58°C in aqueous solution for 48 hours, wash at 58°C in 2×SSPE, 0.1%SDS) with three different genomic probes mapping to either exon I or exon III or exon V+VI, failed to reveal additional bands (not shown) when compared to the blots shown on figure 7. In addition, an initial survey using restriction mapping and probing of XLR genomic clones suggested strongly that some of the fragments visualized by Southern blot hybridization have undergone recent amplification. Detailed analysis of five of these genomic clones complemented by sequencing data of particular segments from a larger panel of XLR clones has revealed that they all share a common exon-intron structure and that they are also closely related at the nucleotide sequence level ($\geq$ 89%). This conservation includes exon-flanking regions as well the exons themselves. Sequence comparisons show that all the XLR genomic sequences so far isolated differ from the pM1 cDNA sequence. Some of these modifications such as a CA dinucleotide substitution in exon V or the point deletion in exon IV are unusual. The latter alteration not only shifts the pM1 reading frame but also introduces stop codons nearby. In fact, multiple stop codons are present in two—including the pM1 reading frame—or even all three of the reading frames, casting serious doubt as to whether any of these fourteen sequences are functional genes. Additional arguments bearing on the alterations in the splice donor sequences flanking exons I and IV and on the substitution of the consensus purine by a pyrimidine at position −3 prior to the initiation codon, also largely support the view that all the XLR clones so far studied represent pseudogenes.

Our mapping studies have revealed that XLR sequences localize to at least two well-separated regions of the proximal part of the X chromosome as well as to the Y chromosome. A more quantitative evaluation of the distribution of the X-linked sequences using a hybrid cell panel showed that the great majority, if not all, lie proximal to the MDXS120 and *Tabby* loci. Thus, the bulk of the XLR genes are not closely linked to the *xid* mutation since this defect has been mapped distal to *Tabby* on the X chromosome, between the Pgk-1 and the α-galactosidase loci (24). The structure of the XLR family consisting as it does of some largely amplified segments makes it difficult to rule out absolutely the existence of XLR sequences located further away from the centromere, in a location compatible with the mapping of the *xid* mutation.

The more centromerically located xlr-1 locus lies close to at least two other loci defined by probes detecting X-Y cross-reacting sequences. One of them, the mouse cDNA pCR5/B (C.E. Bishop, unpublished) also localizes to between the M2C and A-raf loci with which for the moment no recombinants have been detected. The second locus detected by an X-Y cross-reacting probe, 80Y/B, has been defined by *in situ* and hybrid deletion studies

rather than by recombination analysis (25). A complicated family of repeat sequences spanning at least 1 Mbp detected by anonymous probe DXSmh141 has also recently been localized to this region (26). We conclude therefore the xlr-1 locus falls within a region known to harbour not only an extensive repeat family but also other X-Y homologous sequences, some of which have not apparently been extensively conserved during evolution and are absent from man.

The above mentioned probe 8OY/B is one of a family of related anonymous probes isolated from a Y-chromosome library which detect sequences on both the X and the Y chromosomes and which, unlike XLR sequences, are predominantly localized or amplified on the Y chromosome rather than the X. Although these X- and Y-defined sequences do not appear to be transcriptionally active, other genes on the mouse Y are known to be active and these include the ZfY gene thought to correspond to the testis-determining factor (27) and the steroid sulfatase gene (28). If the Y-linked XLR sequences are functional, a possible correlation with the Yaa factor which contributes to autoimmune disease in the BXSB mouse strain (29) could be worth exploring.

The conservation of exon-intron structure in XLR genes among which a minority, perhaps only one, is functional makes XLR unique among large multigene families. It appears to represent the first example of a multigene family containing a large proportion of pseudogenes which are not associated with retroposition. In multigenic families such as the β-actin family (one functional gene, 19 or more pseudogenes)(30), the phosphoglycerokinase (two functional genes including one retroposon, 7 pseudogenes)(31), argininosuccinate-synthetase (one functional gene, 14 pseudogenes)(32), and, most strikingly, phosphoglyceraldehyde dehydrogenase in the rat and mouse which show one functional gene and $\geq$ 200 pseudogenes (33), the pseudogenes are all retroposed. Such retroposed pseudogenes exhibit very specific features (34) including (i) loss of intervening sequences (ii) lack of conservation of flanking regions (iii) presence of flanking direct repeats preceded by AT-rich regions (iv) absence of linkage between the retroposed copies of the functional gene. Although it remains possible that some of these XLR sequences result from the retroposition of nuclear mRNA precursors (such as rat preproinsulin gene, ref.35), the conservation of intervening sequences and the restriction of XLR genes to sex chromosomes indicates rather that XLR behaves more like classical multigene families. There are examples of such gene families, class I-Qa (36), α-interferon (37), 28S ribosomal RNA genes (38), 5S rRNA genes (39) where pseudogenes occur frequently but they never account for more than 50% of the sequences. The status of XLR is therefore clearly unusual. The amplification of the clustered members may have occurred through unequal crossover following homologous recombination, as suggested by analysis of the exon I sequences. The origin of the dispersed members of the family including the Y chromosome, however, must involve another mechanism of gene amplification.

ACKNOWLEDGMENT

We wish to thank Dr Louis Gastinel for oligonucleotide synthesis and Ms Kathy Redman for preparation of the manuscript. The expert technical assistance of Laure Eloy and Monique DoutÄ is gratefully acknowledged. E.Y. Loh is a Hilda Irene Duggan Investigator of the Arthritis Foundation and M.M.Davis is a Scholar of the PEW Foundation. This research was supported by the National Institutes of Health, the Howard Hughes Medical Institute and the INSERM.

REFERENCES

1. Cohen, D.I., Hedrick, S.M., Nielsen, E.A., D'Eustachio, P., Ruddle, F., Steinberg, A.D., Paul, W.E., and Davis, M.M. (1985). *Nature* **314**, 369–372.
2. Cohen, D.I., Steinberg, A.D., Paul, W.E., and Davis, M.M. (1985). *Nature* **314**, 372–374.
3. Siegel, J.N., Turner, C.A., Klinman, D.M., Wilkinson, M., Steinberg, A.D., MacLeod, C.L., Paul, W.E., Davis, M.M., and Cohen, D.I. (1987). *J. Exp. Med.* **166**, 1702–1715.
4. Scher, I. (1982). *Adv. Immunol.* **33**, 1–71.
5. Kung, J.T., Sharrow, S.O., Ahmed, A., Habbersett, R., Scher, I., and Paul, W.E. (1982). *J. Immunol.* **128**, 2049–2056.
6. Hardy, R., Hayakawa, K., Parks, D., Herzenberg., L.A., and Herzenberg., L.A. (1984). *J.Exp. Med.* **159**, 1169–1188.
7. Garchon, H-J, and Davis, M.M. (1989). *J. Cell. Biol.* **108**, 779–787.
8. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). *Molecular cloning: a laboratory manual.* Cold Spring Harbor, New York.
9. Bates, P.F., and Swift, R.A. (1983). *Gene* **26**, 137–146.
10. Chien, Y., Becker, D.M., Lindsten, T., Okamura, M., Cohen, D.I., and Davis, M.M. (1984). *Nature* **312**, 31–35.
11. Feinberg, A.P., and Vogelstein, B. (1983). *Anal. Biochem.* **132**, 6–13.
12. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). *Proc. Natl. Acad. Sci. USA.* **74**, 5463–5467.
13. Chen, E.Y., and Seeburg, P.H. (1985). *DNA* **4**, 165–170.
14. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988). *Science* **239**, 487–491.
15. Hoopes, B.C., and McClure, W.R. (1981). *Nucl. Acids Res.* **9**, 5493–5504.
16. Tabor, S., and Richardson, C.C. (1987). *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
17. Amar, L.C., Arnaud, D., Cambrou, J., Guenet, J.-L., and Avner, P. (1985). *EMBO J.* **4**, 3695–3700.
18. Avner, P., Amar, L., Dandolo, L., and Guénet, J.-L. (1988). *Trends in Genetics* 4, 18–23.
19. Avner, P.A., Arnaud, D., Amar, L., Cambrou, J., Winking, H., and Russell, L.B. (1987). *Proc. Natl. Acad. Sci. USA* **84**, 5330–5334.
20. Devereux, J., Haeberli, P., and Smithies, O. (1984). *Nucl. Acids Res.* **12**, 387–395.
21. Mount, S.M. (1982). *Nucl. Acids Res.* **10**, 459–472.
22. Kozak, M. (1984). *Nucl. Acids Res.* **12**, 857–872.
23. Brockdorff, N., Fisher, E.M.C., Cavanna, J.S., Lyon, M.F., and Brown, S.D.M. (1987). *EMBO J.* **6**, 3291–3297.
24. Berning, A.K., Eicher, E.M., Paul, W.E., and Scher, I. (1980). *J. Immunol.* **124**, 1875–1877.
25. Avner, P., Bishop, C.E., Amar, L., Cambrou, J., Hatat, D., Arnaud, D., and Mattei, M.-G. (1987). *Development* **101**, 107–116.
26. Brown, S.D.M., Brockdorff, N., Cavanna, J.S., Fisher, E.M.C., Greenfield, A.G., Lyon, M.F., and Nasier, J. (1988). *Curr. Top. Microbiol. Immunol.* **137**, 3–11.
27. Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M.C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A., and Brown, L.G. (1987). *Cell* **51**, 1091–1104.
28. Nagamine, C.M., Michot, J.-L., Roberts, C., Guénet, J.-L., and Bishop, C.E. (1987). *Nucl. Acids Res.* **15**, 9227–9238.
29. Murphy, E.D., and Roths, J.B. (1979). *Arthritis Rheum.* **22**, 1188–1194.
30. Sun-Yu Ng, Gunning, P., Eddy, R., Ponte, P., Leavitt, J., Shows, T., and Kedes, L. (1985). *Mol. Cell. Biol.* **5**, 2720–2732.
31. Potten, H., Jendraschak, E., Hauck, S., Amar, L.C., Avner, P., and Müllhofer, G. (1988). *In press.*
32. Freytag, S.O., Bock, H.-G.O., Beaudet, A.L., and O'Brien, W.E. (1984). *J. Biol. Chem.* **259**, 3160–3166.
33. Yun Tso, J., Sun, X.-H., Kao, T.-h., Reece, K.S., and Wu, R. (1985). *Nucl. Acids Res.* **13**, 2485–2502.
34. Weiner, A.M., Deininger, P.L., and Efstratiadis, A. (1986). *Ann. Rev. Biochem.* **55**, 631–661.
35. Soares, M.B., Schon, E., Henderson, A., Karathanasis, S.K., Cate, R., Zeitlin, S., Chirgwin, J., and Efstratiadis, A. (1985). *Mol. Cell. Biol.* **5**, 2090–2103.
36. Weiss, E.H. (1987). *Immunol. Res.* **6**, 179–191.
37. Henco, K., Brosius, J., Fujisawa, A., Fujisawa, J.-I., Haynes, J.R., Hochstadt, J., Kovacic, T., Pasek,

M., Schamböck, M., Schmid, J., Todokoro, K., Wälchli, M., Nagata, S., and Weissmann, C. (1985). *J. Mol. Biol.* **185**, 227–260.
38. Roiha, H., Miller, J.R., Woods, L.C., and Glover, D.M. (1981). *Nature* **290**, 749–753.
39. Miller, J.R., Cartwright, E.M., Browlee, G.G., Fedoroff, N.V., and Brown, D.D. (1978). *Cell* **13**, 717–725.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.