Virology 256, 22–35 (1999)

Characterization of Human MMTV-like (HML) Elements Similar to a Sequence That Was Highly Expressed in a Human Breast Cancer: Further Definition of the HML-6 Group

Hong Yin,*[,1] Patrik Medstrand,* Anna Kristofferson,* Ursula Dietrich,† Pierre Åman,‡ and Jonas Blomberg*

*Section of Virology, Department of Medical Sciences, Uppsala University, Dag Hammarskjölds väg 17, 751 85 Uppsala, Sweden; †Georg-Speyer-Haus, Paul-Erlich-Strasse 42-44, D-60596 Frankfurt, Germany; and ‡Department of Clinical Genetics, University Hospital, S-221 85 Lund, Sweden

Received October 29, 1998; returned to author for revision December 10, 1998; accepted December 24, 1998

Previously, we found a retroviral sequence, HML-6.2*BC1,* to be expressed at high levels in a multifocal ductal breast cancer from a 41-year-old woman who also developed ovarian carcinoma. The sequence of a human genomic clone (HML-6.28) selected by high-stringency hybridization with HML-6.2*BC1* is reported here. It was 99% identical to HML-6.2*BC1* and gave the same restriction fragments as total DNA. HML-6.28 is a 4.7-kb provirus with a 5′LTR, truncated in RT. Data from two similar genomic clones and sequences found in GenBank are also reported. Overlaps between them gave a rather complete picture of the HML-6.2*BC1*-like human endogenous retroviral elements. Work with somatic cell hybrids and FISH localized HML-6.28 to chromosome 6, band p21, close to the MHC region. The causal role of HML-6.28 in breast cancer remains unclear. Nevertheless, the ca. 20 Myr old HML-6 sequences enabled the definition of common and unique features of type A, B, and D (ABD) retroviruses. In Gag, HML-6 has no intervening sequences between matrix and capsid proteins, unlike extant exogenous ABD viruses, possibly an ancestral feature. Alignment of the dUTPase showed it to be present in all ABD viruses, but gave a phylogenetic tree different from trees made from other ABD genes, indicating a distinct phylogeny of dUTPase. A conserved 24-mer sequence in the amino terminus of some ABD envelope genes suggested a conserved function.

Key Words: human MMTV-like sequence; HERV; chromosomal localization; evolutionary relationship; breast cancer.

INTRODUCTION

Human endogenous retroviruses (HERVs) are estimated to constitute at least 0.6% of the human genome and are distributed over all human chromosomes (for a review see Wilkinson *et al.,* 1994). These retroviruses were probably inserted into the germ line from relatively recently to several hundred million years ago by infection with exogenous retroviruses. Mutation and selection against viral functions presumably made many of these human endogenous retroviruses replication-defective. However, endogenous retroviruses can have profound effects on the organism where they reside. A selective Vβ-specific deletion of T-cells in mice is caused by superantigen encoded by endogenous or exogenous mouse mammary tumor virus (MMTV) sequences (see, e.g., Niimi *et al.,* 1995). In humans, several examples of biologic function of HERVs are known (Ting *et al.,* 1992; Schulte *et al.,* 1996), but much remains to be elucidated. Many HERVs are transcriptionally active (reviewed in Wilkinson *et al.,* 1994). HERVs are also under investigation as potential pathogens in cancer (see, e.g., Sauter *et al.,* 1995) and in autoimmune disorders (Bengtsson *et al.,* 1996; Conrad *et al.,* 1997; Perron *et al.,* 1997).

One family of HERVs, HERV-K, was characterized by its sequence homology with the B type mouse mammary tumor virus (MMTV) and Syrian hamster intracisternal A type particles (IAP) (Ono, 1986). Several MMTV-related transcriptional products have been observed in humans (Franklin *et al.,* 1988; Medstrand and Blomberg, 1993; Yin *et al.,* 1997). In a previous study, we identified six groups of HERVs related to MMTV (HML) by sequencing human genomic clones from human lymphocytes (Medstrand and Blomberg, 1993). Recently, the differential expression of HML sequences in human normal peripheral blood mononuclear cells (Andersson *et al.,* 1996) and in human breast cancer (Yin *et al.,* 1997) has been reported.

It is logical to investigate whether HML sequences are causally involved in human breast cancer. Exogenous and endogenous MMTV is a major etiologic factor in breast cancer in mice. There are both environmental and genetic contributions to the risk of breast cancer in humans (Willett and Stampfer, 1989). Theoretically, variants of HML sequences could account for some of these risks. However, we addressed this question in a previous publication (Yin *et al.,* 1997) where the degree of HML expression in 60 breast cancers was compared to non-malignant control tissue. We detected only one patient with a clearly abberrant expression pattern. In this woman, a high expression of the HML-6 group (Medstrand and Blomberg, 1993) was found by RT-PCR and dot hybridization under group-specific conditions. A probe containing this sequence (HML-6.2*BC1*) gave a

[1] To whom correspondence and reprint requests should be addressed. Fax: 46-18-515615. E-mail: Hong.Yin@infektion.uu.se.

0042-6822/99 $30.00

single restriction fragment under high-stringency Southern hybridization conditions, from most human normal and malignant breast tissues, human breast cancer cell lines, and human normal peripheral blood mononuclear cells. There were, however, signs of polymorphism, because all individuals did not yield the same band pattern (Yin *et al.,* 1997). We now report the finding of a human endogenous retroviral element, HERV-K (HML-6.28), identified from a human genomic library by using clone HML-6.2*BC1* as probe. The sequence is intermediate to MMTV, HERV-K10, and the intracisternal type A particle (IAP) sequences of rodents. This sequence, and some newly identified related sequences reported here, extends the knowledge of the HML-6 sequence group beyond what we previously reported (Medstrand *et al.,* 1997) and of ABD retroviruses in general. A comparison of the sequences reported here and other ABD sequences shows that (i) LTRs contained motifs discernible in other ABD LTRs but no ORFs similar to those of MMTV; (ii) HML-6 Gag consists mainly of matrix, capsid, and nucleocapsid proteins, suggesting that the additional unique Gag proteins of other ABD viruses are late acquisitions; (iii) dUTPase-like sequences were detected in all ABD sequences but gave phylogenetic trees different from those of other HML-6 genes, suggesting that ABD dUTPase probably had an evolutionary history different from that of other ABD genes; (iv) there exists a conserved amino-terminal stretch in many ABD envelope proteins, a region that normally is highly variable. HML-6.28 was localized to chromosome 6, band p21, near or within the MHC region.

RESULTS

The HML-6.28 and HML-6.29 genomic clones

The *pol* sequence of HML-6.3 was used as probe to screen 8×10^5 plaques of a human genomic library under hybridization conditions specific for the HML-6 group (Medstrand *et al.,* 1997). Thirty-nine clones containing HML-6 *pol* were identified and amplified using the ABDPOL/ABDPOR primers. The HML-6.29 clone (see below) was taken at this stage. Amplification products were hybridized with the HML-6.2*BC1* probe under high wash stringency. Only one phage clone, HML-6.28, gave high-stringency hybridization. To identify whether this clone represented a full-length HML-6 genome, LTR, *gag,* and *env* probes (see Materials and Methods) were used in Southern blot with digested DNAs of clone 6.28. The LTR and *gag,* but not the *env,* probe hybridized. The *pol* probe, clone HML-6.2*BC1,* detected the same 3-kb *Hin*dIII and 2-kb *Eco*RI fragments as previously detected with the same probe in genomic Southern blots (Yin *et al.,* 1997). A restriction map of clone 6.28 later revealed the same fragments. Thus, Southern blot analysis of HML-6.28 indicated the presence of a truncated provirus with one LTR, a complete *gag,* and a partial *pol,* but no *env.* Sequencing revealed that the HML-6.28 element is 4668 bp long. It contains a 5′ LTR (862 bp), *gag* (1899 bp), and a partial *pol* (1907 bp). An *Alu* sequence is inserted into the untranslated leader region in the beginning of *gag.*

The general structure of HML-6.29 has been described previously (Medstrand *et al.,* 1997). It is less deleted and contains a long *env* region. It was sequenced to facilitate interpretation. It is 7370 bp long. It has both 5′ LTR and 3′LTR and seemingly complete *pol* and *env* genes. The NC domains of *gag* and *pro* are, however, deleted.

The nucleic acid sequence of HML-6.28 in the region amplified by the ABDPOL/ABDPOR primers is 99% identical to the same stretch of the HML-6.2*BC1* clone (Fig. 1). *Hin*dIII and *Eco*RI yield the same HML-6.2*BC1* hybridizing fragments from HML-6.2*BC1* as from total human DNA. Restriction site analysis of the HML-6.28 sequence revealed that *Hin*dIII cleaved at positions 615 and 3715 and that *Eco*RI cleaved at position 1305. These fragments were not present in other genomic HML-6 clones (data not shown). These findings, and the sequence data discussed below and shown in Fig. 1, show that HML-6.28 is the likely origin of HML-6.2*BC1.*

The restriction map of the additional clone HML-6.29 has been determined previously (Medstrand *et al.,* 1997). The sequence analysis revealed high similarities to the other HML-6 elements. The degrees of similarity are given in Table 1A and are diagrammatically shown in Fig. 2. HML-6.29 is especially interesting because of its long *env* sequence. It is 66% similar to HML-6.17 in the TM domain. HML-6 *env* was only incompletely known before (Medstrand *et al.,* 1997). Env contains interrupted ORFs that may have encoded surface (SU) and transmembrane (TM) glycoproteins. SU is also disrupted by insertion of a 657-bp sequence similar to one found on chromosome X (Z82211). It contains an *Alu* sequence (310 bp) between the SU and the TM domains. Both HML-6.28- and HML-6.29-related sequences contain full-length RT domains that are disrupted by multiple termination codons and frameshifts. The coding regions of Gag and the TM domain of Env in HML-6.29 contain one stop codon each. Thus, functional proteins are not expected to be produced.

Chromosomal localization

The chromosomal localization of HML-6.28 sequence was first determined by Southern blot analysis of DNA from mouse–human and Chinese hamster–human hybrids carrying a single human chromosome. Mouse, Chinese hamster, and human lymphocyte DNA was used as controls. The DNA was *Hin*dIII digested, electrophoresed, transferred to nylon membranes, and hybridized with the radiolabeled HML-6.2*BC1* probe under high-stringency washing conditions. This probe detected a single strongly hybridizing 3-kb fragment in Southern blots of *Hin*dIII-digested human DNA under high-strin-

```
                                                                                                    60
HML-6.2BC1      GATCCAAAAATGTCAGGAAAATGGCAACTGCTACATGATTTGAGAGCTATTAAATGCACA
HML-6.28        ............................................................
Z84814          AT........A...C....GG...TG......G....C..G...------.T........

                                                                                                   120
HML-6.2BC1      GATTAAACCAGTGGGTGCATTACAGCAAGGTCTGTCATCCTCAGCAGCCATTCCAAGAGA
HML-6.28        ...........................................................-....
Z84814          A...G.....A.....C..............T..C..A.TC...TG.T......T.....

                                                                                                   180
HML-6.2BC1      TTGGACTCTTGTAGTAATAGGTCTTAAAGATTTTTTTTTTTTAATATACCATTACACAAAA
HML-6.28        ............................G.............................
Z84814          ....CT.............CA......G..-..G...C..C.C..............G.

                                                                                                   240
HML-6.2BC1      AGGATAAGCCTCAATTTGCCTTCTCTGTGCCTTCTATTAATCAAAGA-GAGCCTGTCTCT
HML-6.28        ...............................................A............
Z84814          ............G................................A.-..A.....T...

                    246
HML-6.2BC1      CGCTCT
HML-6.28        ......
Z84814          T...A.
```

Percentage of nucleotide identity

	HML-6.2*BC1*	HML-6.28	Z84814
HML-6.2*BC1*	–	99.2	82.4
HML-6.28		–	82.4
Z84814			–

FIG. 1. The comparison of the nucleotide sequence of the ABDPOL/ABDPOR *pol* region of HML-6.28 clone, HML-6.2*BC1* clone (Yin *et al.,* 1997), and ABDPOL/ABDPOR region on Z84814 (chromosome 6p21). Identity is shown by dots, and gaps are shown by dashes. Percentage nucleotide identity is shown at the bottom.

gency conditions, as reported previously (Yin *et al.,* 1997), suggesting that it is a single-copy gene in the human genome. This fragment was detected also in blots with DNA from the human–rodent hybrids. When membranes were washed at low stringency, numerous fragments were detected with both human and rodent DNA. At medium stringency, a single fragment was detected with human DNA, mouse DNA, and some human–mouse hybrids (data not shown). After high-stringency washing, a single band was obtained only from total human DNA

TABLE 1A

Percentage Identity of the HML-6.28 Sequences to Other Retroviruses at the Amino Acid Level

	HML-6.29	HML-6p	HML-6.17	HERVK10	MMTV	HERVKC(4)	IAPm	MPMV	JSRV
Gag									
MA	60.5	54.5		22.1	18.0	22.1	16.3	19.8	20.3
CA	64.8	66.5		27.0	23.5	—	29.3	27.9	32.1
NC	39.6	54.7		32.8	28.1	—	27.3	31.7	17.2
Pro									
DU	—	73.4		41.9	41.1	—	25.0	33.1	33.1
PR	—	63.3		34.8	28.3	—	36.0	31.7	30.2
Pol									
RT	70.9		74.1	51.7	48.7	45.1	48.8	52.5	50.2
RNaseH	70.7		78.0	34.1	34.1	17.1?	26.8	29.3	36.6
IN	76.8		76.8	39.4	41.6	—	37.9	38.5	37.6
Env									
SU	87.8		—	23.4	15.7	19.4	—	11.7	17.1
TM	67.3		66.2	22.5	21.6	11.7?	—	16.8	22.1
pSAG[a]	26.7		—	23.1	—	23.1	—	—	9.5

Note. A dash indicates absence of data; a question mark denotes uncertainty due to frameshifts.

[a] Conserved 24-aa stretch in the amino terminus of SU.

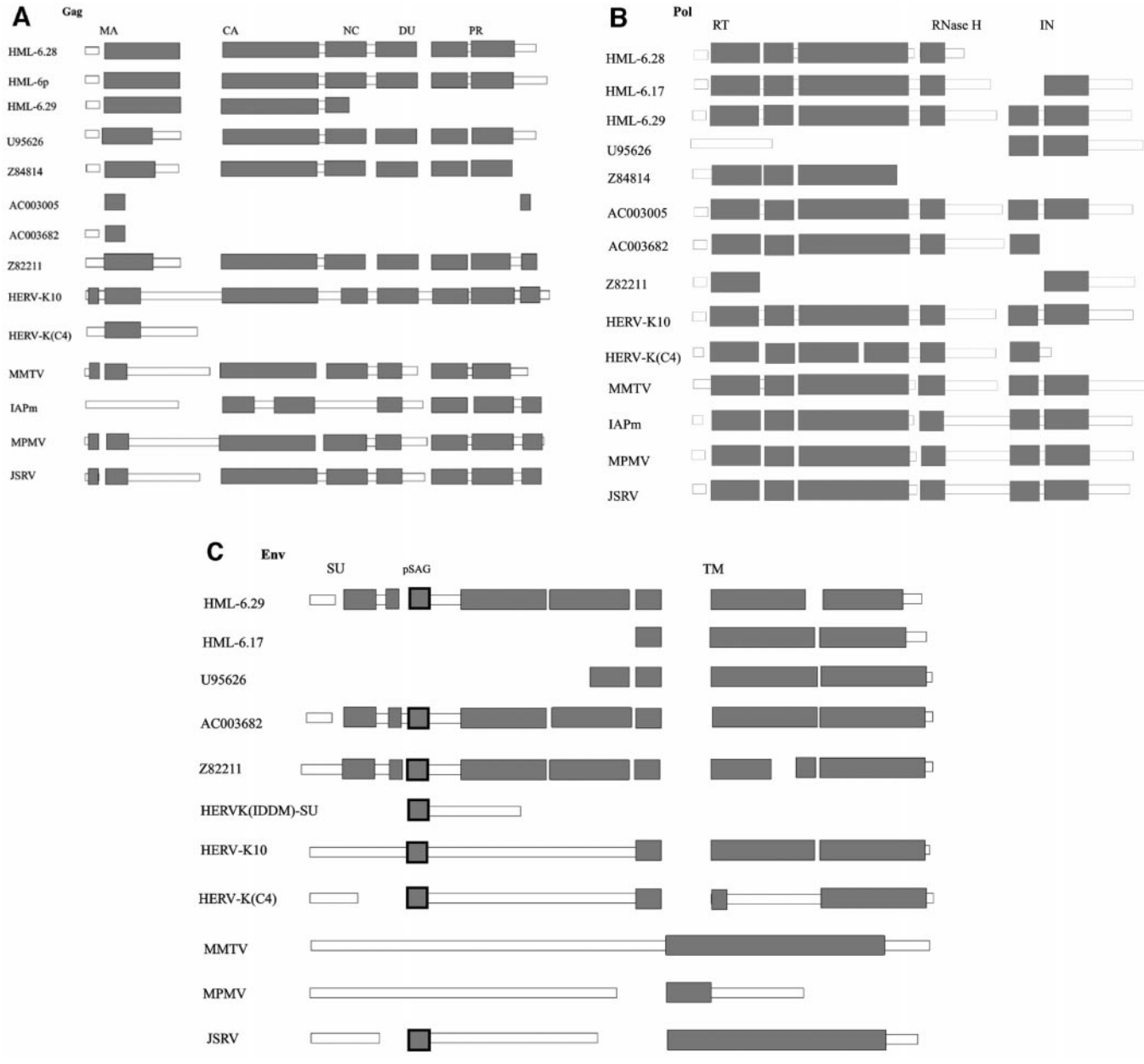

FIG. 2. Schematic representation of amino acid sequence similarity in (A) Gag, (B) Pol, and (C) Env regions between the HML-6 (HML-6.28, HML-6.29, HML-6p, and HML-6.17) group and HERV-K10 (Ono *et al.,* 1986), HERV-K (C4) (Dangel *et al.,* 1994), MMTV (Moore *et al.,* 1987), IAPm (Mietz *et al.,* 1987), MPMV (Sonigo *et al.,* 1986), JSRV (York *et al.,* 1992), and human chromosomal clones (for GenBank accession number, see Table 1). Shaded boxes show amino acid sequences that are at least 10% identical to the compound consensus. A frame indicates sequences with similarity to a conserved 24-amino-acid stretch of Env, which is contained in a putative superantigen (pSAG) encoded by a human endogenous retrovirus (Conrad *et al.,* 1997). Alignment and statistical scorings were done with MACAW software (National Center for Biotechnology Information, Bethesda, MD), implementing the algorithm of Karlin and Altschul (1990).

and DNA from hybrid GM10629A. GM10629A is a human–Chinese hamster ovary somatic cell hybrid retaining only human chromosome 6. The *Hin*dIII restriction fragment was about 3 kb. We conclude that, although sequences more or less related to the HML-6.2*BC1 pol* gene are widely dispersed in the human genome, there is an HML-6.2*BC1*-specific locus on chromosome 6. No cross-hybridization was seen with DNA from Chinese hamster and mouse (Fig. 3). To further localize the HML-6.2*BC1*-like sequence on chromosome 6, the clone HML-6.28 was used as a probe for FISH analysis. As expected from the Southern blot results, there was a hybridization signal only from chromosome 6 in FISH. The pattern of hybridization was identical between two different individuals. The signal was located at band p21–p22 (Fig. 4), showing that a sequence highly related to HML-6.2*BC1* is integrated at the short arm of chromosome 6.

Identification of other HML-6.2*BC1*-like sequences

BLAST searches in GenBank for individual LTR, *gag,* and *pol* regions showed similarities to previously known

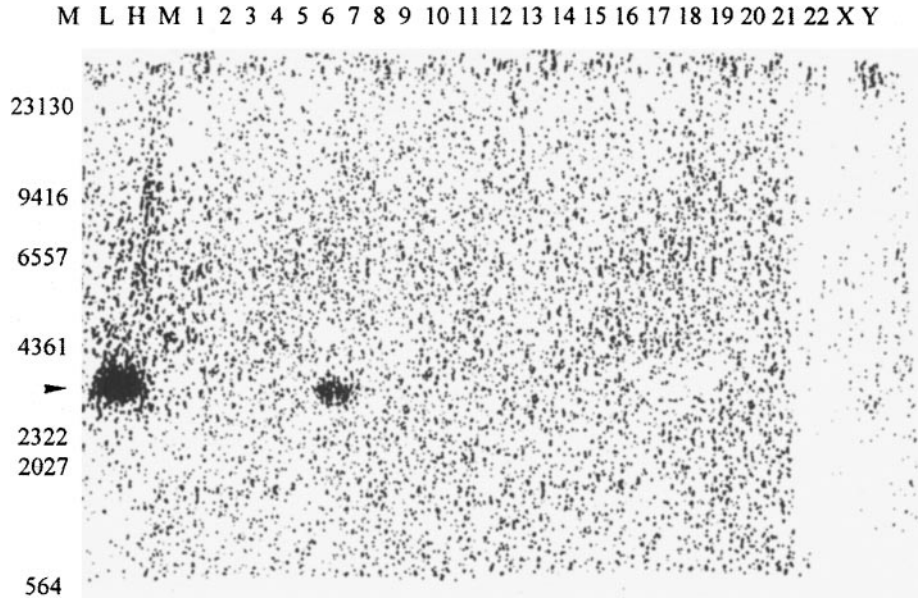

FIG. 3. Southern blot hybridization of a somatic cell hybrid mapping panel. *Hind*III-digested DNA from human–Chinese hamster and human–mouse hybrid cell lines and from controls was hybridized with the probe HML-6.2*BC1* clone under high-stringency washing conditions. Lane 1, M, molecular marker; lane 2, L, human lymphocyte DNA; lane 3, H, Chinese hamster DNA; lane 4, M, mouse DNA; lanes 1–Y, the numbers designating the lines refer to the human chromosomes retained in each hybrid. The arrowhead indicates a 3-kb hybridizing DNA fragment from human lymphocyte and chromosome 6.

type A, B, and D retroviruses and revealed previously undefined human genomic sequences (see above). HML-6.2*BC1,* HML-6.28, and HML-6.29 sequences were used. The result revealed homologous sequences on human chromosomes 3, 6, 19, and X (23–83% amino acid identity in Gag, Pol, and Env; 57–80% nucleotide identity

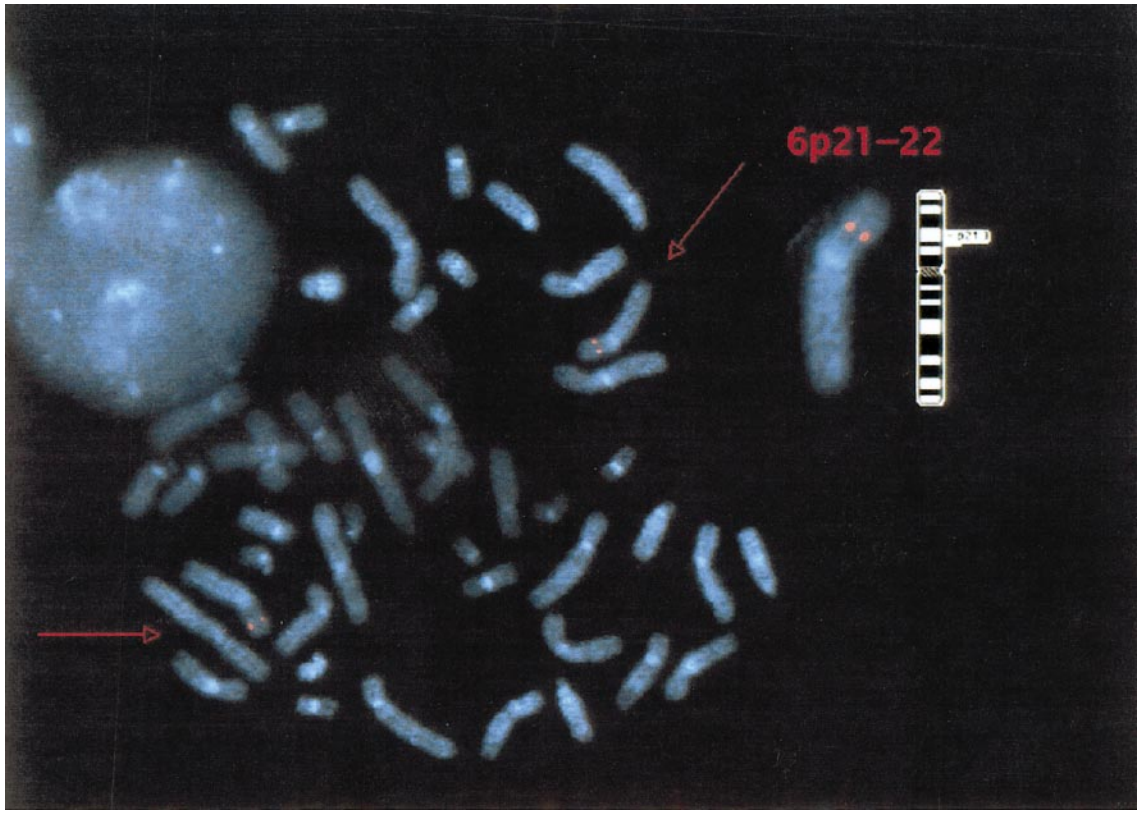

FIG. 4. FISH experiment. The human genomic clone HML-6.28 was used as a probe and hybridized to slides containing normal human metaphases. The signals were required to be localized at the same chromosome site in at least two metaphases in order to be scored as positive.

TABLE 1B

Distribution and Percentage Identity to HML-6.28 of Sequences Similar to It on Human Chromosomes Found in GenBank

Position	U95626[a] 115142–122217 3[b]	Z84814 55573–59666 6p21	AC003005 41256–44898 19q13.4	AC003682 1326–7215 19q13.4	Z82211 95540–101916 X
5′LTR[c]	56.9	59.2	80.4	69.8	57.0
3′LTR	+			+	+
Gag					
MA	58.5	62.8	33.7	56.1	58.1
CA	65.2	60.8	—	—	64.2
NC	53.1	53.8	—	—	51.6
Pro					
DU	72.1	67.2	—	—	69.7
PR	23.2	53.2	—	—	46.4
Pol					
RT	—	71.4	67.5	71.8	—
RNaseH	—	—	70.7	70.7	—
IN	75.8	—	73.2	—	67.7
Env					
SU	82.8	—	—	62.6	69.3
TM	67.3	—	—	61.9	66.8
pSAG	—	—	—	20.0	24.0
Gene features:	Contains ccr2, ccr5, and ccr6 genes	Contains HLA class II DRA, DRB3, and DRB9	Contains ZNF gene between D19S303 and ZNF303	Contains ZNF gene between D19S773 and ZNF303	Contains ESTs and STS genes

Note. +, presence; —, absence.

[a] Accession number and position of the sequence in GenBank.

[b] Human chromosome.

[c] Nucleotide for LTRs; amino acid for the rest of sequence.

in LTR) (Table 1B and Fig. 2). For example, U95626 (chromosome 3) contains two LTRs, a complete *gag,* a *pol* with RT and RNase H domains missing, and an *env* with SU missing. Z84814 (chromosome 6) contains a 5′LTR, a complete *gag,* and a *pol* with parts of RNase H and integrase domains deleted. Two sequences were found on chromosome 19. One, AC003005 (chromosome 19q13.4), is from clone F25491, and another, AC003682, was found in clones F18547, F11133, R27945, and R28830 of a human chromosome 19-specific cosmid library. They are found in the same chromosome region, but they are in different loci. The AC003005 element has a 5′LTR, a small part of *gag,* and a complete *pol.* AC003682 has two LTRs, a small part of *gag, pol* with an integrase deletion, and a complete *env.* Z82211 (chromosome X) has two LTRs, a complete *gag, pol* containing parts of RT and integrase, and a complete *env.* Although a sequence highly similar to HML-6.28 was found in sequence Z84814 that contains sequences from chromosomal location 6p21–p22, the same position as HML-6.28, it differed in several respects from HML-6.28. It was 83% similar to HM-6.2*BC1* in the ABDPOL/ABDPOR amplifiable stretch of *pol,* whereas the HML-6.28 sequence was 99% similar (Fig. 1). The chromosome 6 element had a sequence difference in *gag* and deletions in PR and *pol* relative to HML-6.28. The two sequences therefore must be different. The complete sequencing of the MHC region will probably eventually show their exact locations.

Similarity of HML-6.28 and HML-6.29 to other retroviruses

At the amino acid level, HML-6.28 and HML-6.29 elements revealed ABD-like *gag, pol,* and *env* genes, where a total of 10 domains exhibited significant protein similarity to the previously published HML-6 sequences (40 to 88% identity; see Figs. 1, 2, and 5–7, and Table 1). This was also seen in amino acid sequence alignments. Other similar sequences were, in approximate decreasing order of similarity, HERV-K10 (Ono *et al.,* 1986), HERV-K (C4) (Dangel *et al.,* 1994), murine IAP (Mietz *et al.,* 1987), MMTV (Moore *et al.,* 1987), Jaagsiekte retrovirus (JSRV) (York *et al.,* 1992), and Mason–Pfizer monkey retrovirus (MPMV) (Sonigo *et al.,* 1986) (Table 1A). Figure 2 shows a schematic representation of the amino acid similarities in Gag (Fig. 2A), Pol (Fig. 2B), and Env (Fig. 2C). The similarities to the new GenBank-derived human genomic sequences are shown in Table 1B. To further define the relationship of HML-6.28 and HML-6.29 to known retroviruses, phylogenetic trees were constructed, based on multiple sequence alignments of Gag, Pol, and Env amino acid sequences. Figure 5 shows the phylogenetic relationships of Gag and Pol of HML-6.28 and HML-6.29 to other retroviruses in unrooted neighbor

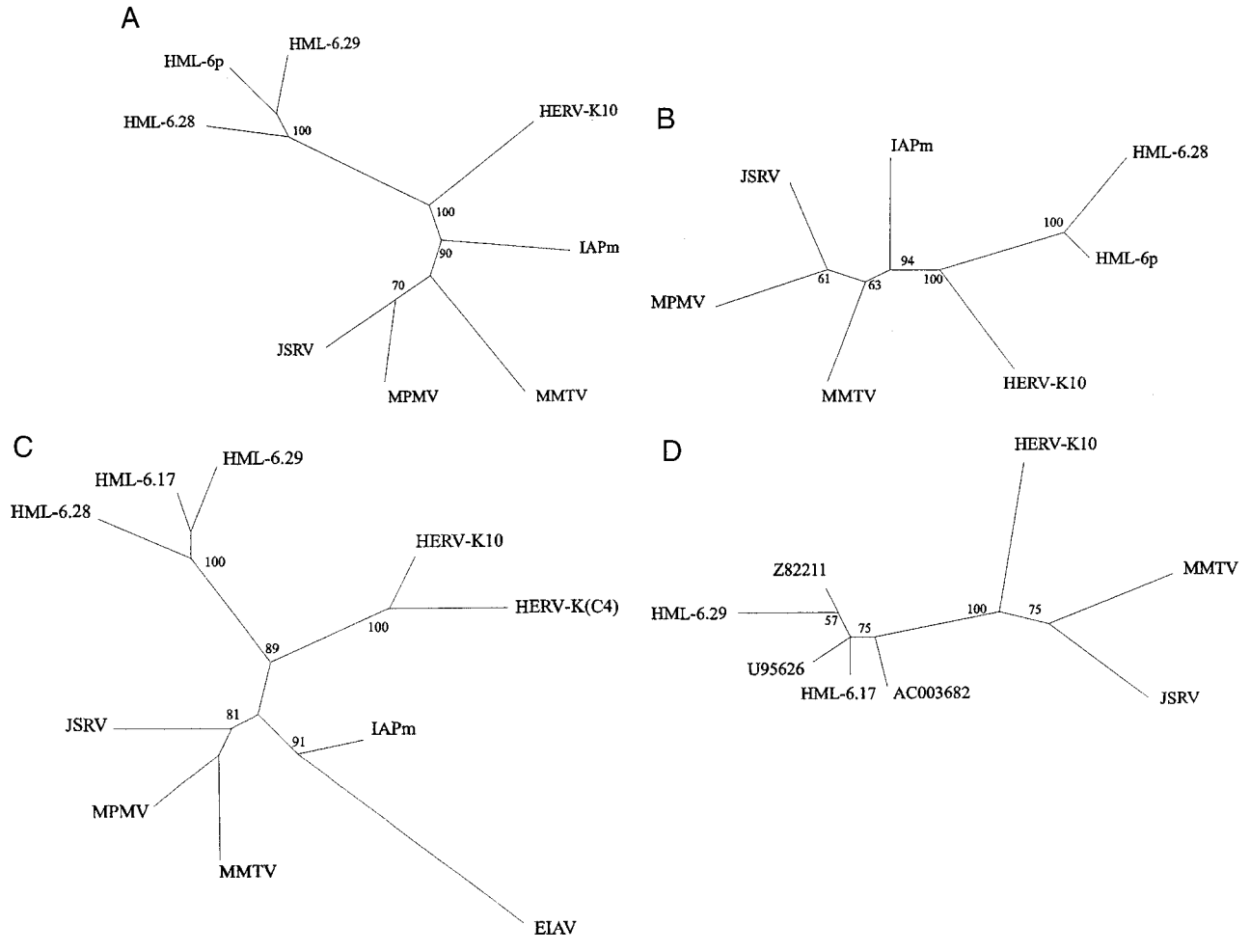

FIG. 5. Phylogenetic analysis of HML-6.28, using the neighbor-joining method. (A) Unrooted phylogenetic tree of CA amino acid sequence (230 aa) of HML-6.28 from the Gag region compared with HML-6.29, HML-6p, HERV-K10, IAPm, MMTV, MPMV, and JSRV. (B) Unrooted tree for PR amino acid sequences (139 aa). (C) Unrooted phylogenetic tree depicting the relationship of amino acid sequences from the RT region (263 aa) of HML-6.28, HML-6.29, HML-6.17, HERV-K10, HERV-K (C4), IAPm, MMTV, MPMV, JSRV, and EIAV. (D) Unrooted tree for TM amino acid sequences (224 aa) of HML-6.29, HML-6.17, HERV-K10, MMTV, JSRV, U95626, AC003682, and Z82211. The values at the branches indicate the percentage supporting that node after 100 bootstrap replicates.

joining trees. CA (Fig. 5A), PR (Fig. 5B), RT (Fig. 5C), and TM (Fig. 5D) based trees all had the same branch order and similar relative branch lengths. Known extant exogenous ABD retroviruses all branch together. HML-6 sequences branch together, distinct from HERV-K10. Specific features of HML-6.28, HML-6.29, and related sequences found in GenBank are detailed below.

LTRs

The LTRs complemented previous knowledge of HML-6 LTRs (Medstrand *et al.,* 1997). Polyadenylation signals (AATAAA) and a probable hormone-responsive element (HRE) (Scheidereit and Beato, 1984), a possible enhancer core identified in the HML-6.17 5′LTR (Medstrand *et al.,* 1997) and HERV-K10 LTRs (Ono *et al.,* 1986), and a presumed polyadenylation site, CA, were identified. A poly(A) signal and CA site were detected in the 5′LTR of HML-6.28 and 6.29 but not in the 3′LTR of HML-6.29. Both elements contain a likely primer binding sequence (PBS; Fig. 6). Fourteen and 12 of 17 bases of the probable PBS of the two elements, respectively, were complementary to the corresponding region of rat lysine UUU tRNA. This tRNA is used as primer by MMTV (Varmus, 1982). The putative PBS is identical to that found in HML-6.17 and is closely related to the PBSs of other HERV elements belonging to the HERV-K superfamily. Probable U3, R, and U5 regions were identifiable in both HML-6.28 and HML-6.29 LTRs. HML-6.29, U95626, AC003682, and Z82211 have both 5′LTR and 3′ LTRs. Like the earlier reported HML-6.17, which has a 13% 5′-3′LTR difference (Medstrand *et al.,* 1997), they differed by 15, 11, 14, and 14%, respectively. In none of these LTRs did we find an open reading frame similar to the MMTV superantigen.

Gag

The amino terminus of MA of HML-6.28 and HML-6.29 could be aligned with most other ABD MAs. However, the

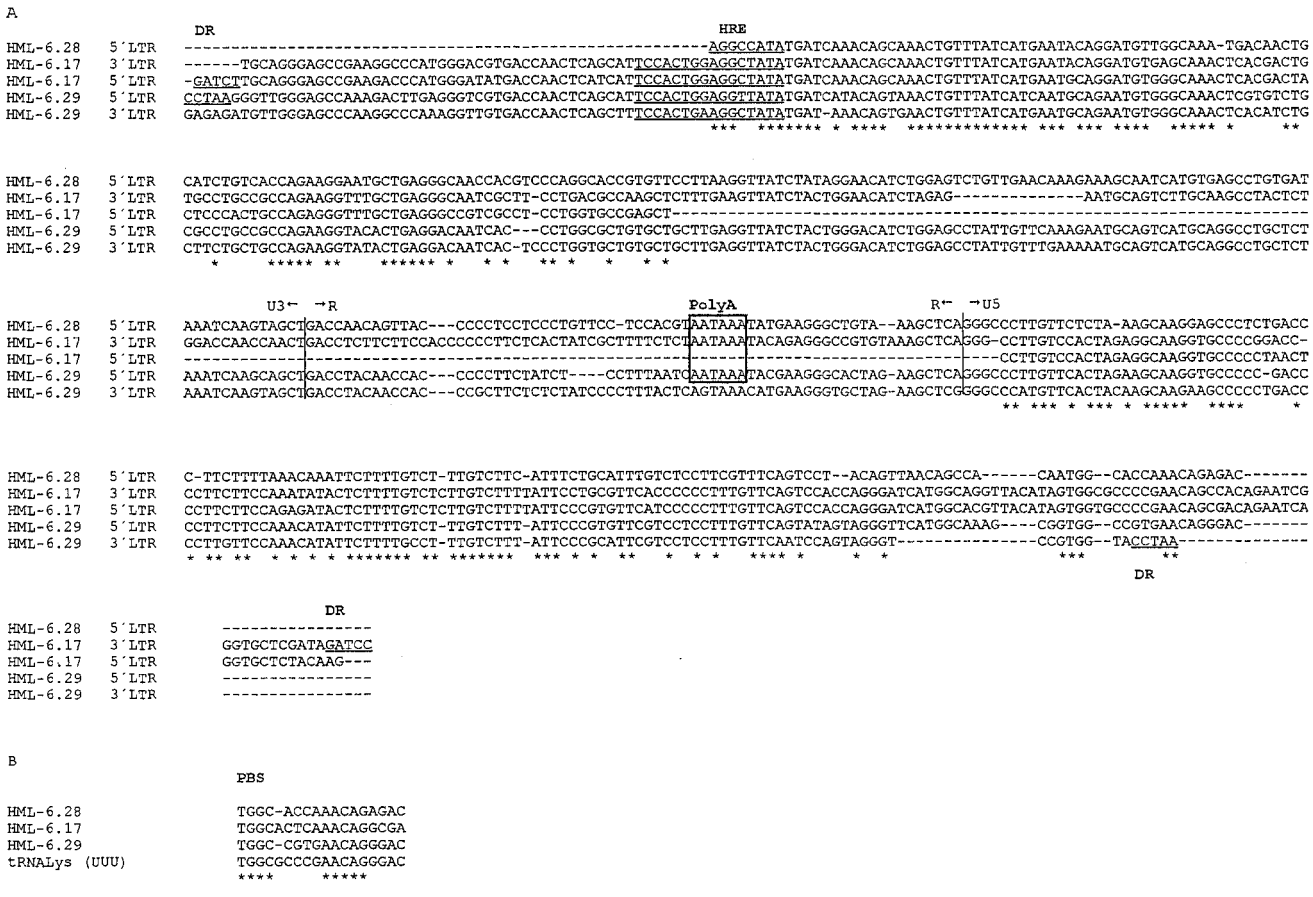

FIG. 6. (A) Multiple alignment of HML-6 LTRs. The absence of a nucleotide is denoted by a dash. Asterisks indicate base identities. Probable DR (direct repeats) and HRE (hormone responsive element) are underlined. A polyadenylation signal (AATAAA) is boxed. U3/R and U5/R boundaries are indicated according to Ono (1986). (B) Alignment of the PBS (Primer Binding Site) of HML-6.28, HML-6.17 (Medstrand *et al.,* 1997), and HML-6.29 5′LTRs. The sequence of PBS is compared to a rat lysine tRNA with UUU anticodons.

ensuing region was more diverse; HML-6 had no counterparts of the intervening Gag proteins MMTV p24 (Racevskis and Prakash, 1984; Moore *et al.,* 1987) and MPMV p12 and p10 (Bradac and Hunter, 1984; Sonigo *et al.,* 1986). Also IAPm, JSRV, MMTV, MPMV, HERV-K10, and HERV-K(C4) contained additional unique stretches between MA and CA (Fig. 2A). The C terminus of HML-6 MA contains a 20-amino-acid proline-rich stretch with high similarity to HTLV-I MA (Seiki *et al.,* 1983) with consensus motifs like "PTAP" and "PPPY" (cf. Wills *et al.,* 1994). The overall arrangement of HML-6 *gag* thus is simple, similar to that of HTLV (Seiki *et al.,* 1983). HML-6 CA was colinear with CA of other ABD viruses and was easily aligned (Table 1 and Fig. 7). It contained several conserved motifs including the major homology region (MHR) (Craven *et al.,* 1995). A consensus HML-6 MHR is (LVI)MQGSQEPYX3F(LV)X2LQEXV, where X denotes any amino acid, the numbers denote the number of intervening Xs, and a conserved position is shown either by a single amino acid letter or several amino acids within parentheses. Within the NC protein of HML-6.28, a first zinc finger domain with the consensus CYNCG(KQ)(IT)-GH(FL)KK(EN)C, followed by a second imperfect zinc finger domain, was identified. The putative protease domain of HML-6.28 shows 63.3% amino acid identity to the corresponding region of HML-6p (Medstrand *et al.,* 1997).

dUTPase

Within the HML-6.28 *gag* gene, a region with amino acid similarity to the putative dUTPase was defined by aligning. Its sequence is similar to that of HML-6p, HERV-K10, MMTV (Jacks *et al.,* 1987), MPMV, JSRV, and EIAV (equine infectious anemia virus; Yaniv *et al.,* 1986), by 73, 42, 42, 34, 34, and 30%, respectively. However, there also seems to exist a similar but more distantly related (25% identity) protein in intracisternal type A particles. The human dUTPase (McIntosh *et al.,* 1992; Strahler *et al.,* 1993) is 24% similar and *Escherichia coli* enzyme (Lundberg *et al.,* 1983) is 22% similar (Fig. 8A). MMTV, EIAV, and *E. coli* have five strongly conserved motifs (McGeoch, 1990). The sequence identity between the human and the *E. coli* enzyme (32%), and between the human and retroviral enzymes (MMTV and EIAV) (33%) has been reported (Björnberg, 1995). To study the phylogeny of this separately acquired retroviral enzyme, a similarity tree was constructed with the PROTDIST and NEIGHBOR methods (Fig. 8B). The DU-based tree was similar to the

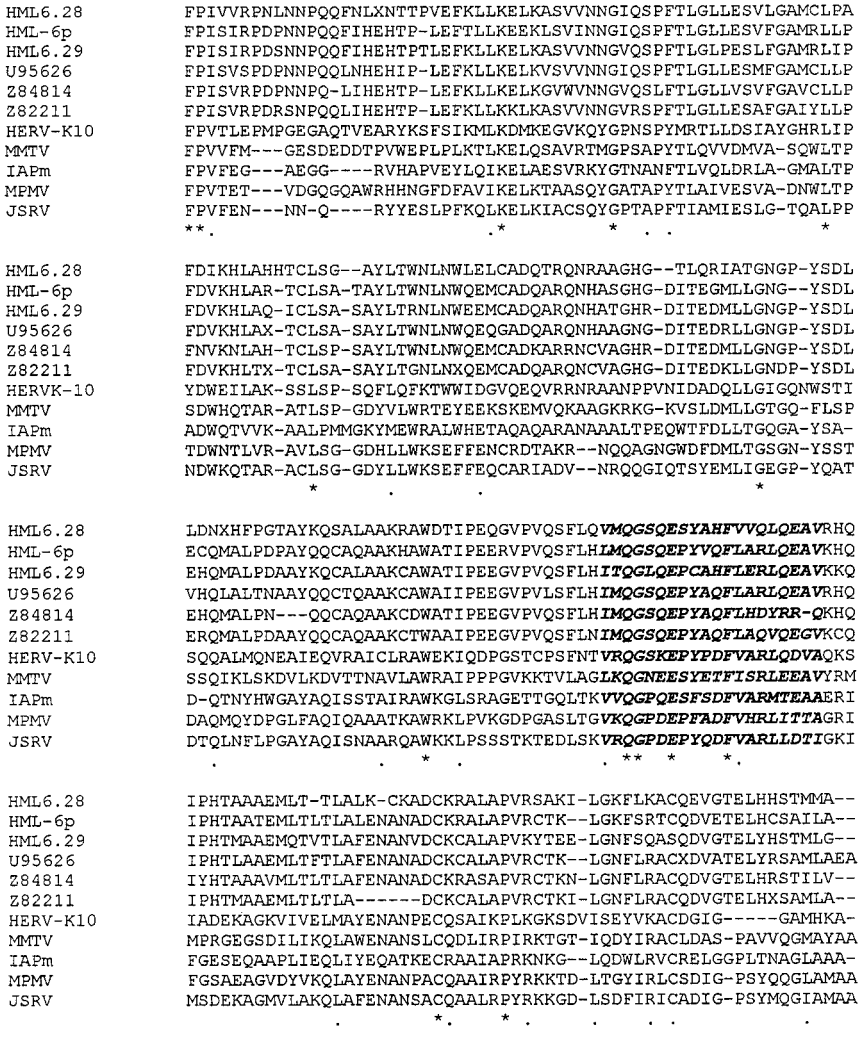

FIG. 7. Multiple alignment of known and deduced amino acid sequences of the CA domain of ABD retroviruses. The motif region is depicted as boldface italic. Asterisks indicate total identity, and dots indicate partical identity.

other trees (Fig. 5), but had one difference. The IAP DU-like sequence clustered distantly with JSRV, whereas its PR and RT sequences are the most HML-like of the nonhuman sequences. A non-ABD virus, EIAV (a lentivirus), also has a dUTPase but it is located in *pol.* It clustered with human and *E. coli* dUTPases.

Pol

Retroviral *pol* genes are generally the most conserved sequences among retroviruses (McClure *et al.,* 1988). HML-6.28 and HML-6.29 both share more than 70% amino acid identity with HML-6.17 and 41.7 and 41.5% identity with HERV-K10 and MMTV, respectively. As shown in Table 1, the RT region of HML-6.28 exhibits the highest identity to other ABD retroviruses. The tether domain that connects RT and RNase H is less conserved (60% identity to HML-6.17). HML-6.28 RNase H shows 70 and 78% identity to HML-6.29 and HML-6.17 and 34% identity to both HERV-K10 and MMTV, respectively. In addition to the universally conserved LPQG and YXDD motifs, HML-6 Pol contains the following conserved motifs, potentially useful for constructing broadly amplifying primers: WNLKI(GA)(PL)EKVQ, LN(DG)FQQLLGDI(HN)W, WQMDVTHI, KLSYVHVTIDT, (TI)DNGPAY, and HKTGIPYNPRGQGI.

Env

We found a 24-amino-acid conserved motif in the amino terminus of the SU domain of Env of several ABD sequences, including HML-6.29 (Figs. 2C and 9). The features of the motif are (IVS)X2(IVL)X(WG)X6-(IV)Y(HV)(ND)(DQW)X2(LW)X(SPG)(RGK)(TPK). In this motif, the deduced HML-6.29 Env has 28% amino acid sequence identity to HERV-K(IDDM) (Conrad *et al.,* 1997) and 24% to HERV-K10 and HERV-K (C4). The Env-containing HML-6 sequences found in GenBank also contained the motif. The carboxy termini of SU were conserved within HMLs, but distinct from those of exogenous ABD viruses (Fig. 2C). As is evident from Fig. 5D, a TM-based phylogenetic tree was similar to CA-, PR-, and RT-based

A

	HML-6p	HERV-K10	HERV-K(C4)	MMTV	JSRV	IAPm	Human	EIAV	*E.coli*
HML-6.28	72.8	42.1	42.1	34.2	34.2	25.4	23.7	29.8	21.9
HML-6p	-	40.0	39.1	33.9	30.4	23.5	25.2	27.8	24.3
HERV-K10		-	40.5	41.4	42.2	29.3	22.4	25.0	22.4
MMTV			-	50.8	44.4	29.7	32.2	32.2	32.2
MPMV				-	60.7	30.5	29.7	29.7	29.7
JSRV					-	32.5	27.4	28.2	31.6
IAPm						-	23.7	16.9	16.9
Human							-	32.2	30.5
EIAV								-	25.4
E.coli									-

B

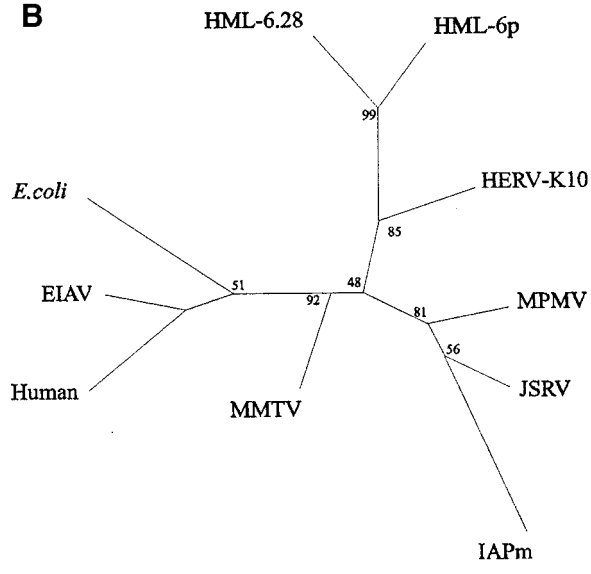

FIG. 8. (A) Percentage identity between the amino acid sequences of dUTPase domains of HML-6.28 and HML-6p, HERV-K10, MMTV (Jacks *et al.,* 1987), MPMV, JSRV, IAPm, EIAV (equine infectious anemia virus, Yaniv *et al.,* 1986), *E. coli,* and human dUTPase (McIntosh *et al.,* 1992). The analysis was performed on an alignment covering 120 amino acids. (B) Phylogenetic analysis of dUTPase from HML-6.28, HML-6p, HERV-K10, MMTV, MPMV, JSRV, IAPm, EIAV, *E. coli,* and human.

trees. HML-6 TM contained at least three imperfect heptad repeats in amino acids 47–69 (LHNRINTELQTEVAMLKSIVLWL), approximately conforming to the leucine zipper repeat pattern of MLV (LREVEKSISNLEKSLTSLSEVVL) (Ramsdale *et al.,* 1996; Delwart *et al.,* 1990). The other aligned ABD TMs (HERV-K10, MMTV, and JSRV) had a similar arrangement. The most conserved motif of HML-6 TM (its "immunosuppressive sequence"; see, e.g., Gottleib *et al.,* 1989) is LWLGEKAQSLQLQQQLHC.

DISCUSSION

Sequencing and genomic mapping of HERVs will yield a wealth of information on the evolution and biology of primate retroviruses. This knowledge will also profoundly influence the understanding of the functions of the human genome. HERVs can exert both *cis* effects by transposition to new loci, which causes gene disruption and gene activation, and *trans* active regulatory retroviral functions, which influence host genes at a distance (Varmus, 1982). Some of these elements have retained active enhancer and promoter functions (Ting *et al.,* 1995; Schulte *et al.,* 1996). A number of HERV LTRs have been found in the vicinity of the MHC genes, where they are suggested to mediate recombination and possibly to influence the expression of MHC genes (Svensson *et al.,* 1995; Dangel *et al.,* 1994; Kambhu *et al.,* 1990; Mayer *et al.,* 1993). A significant portion of the high variability and polymorphism in this region is caused by insertions or deletions involving retroviral sequences (Horton *et al.,* 1998).

HML-6.28 is located in or near this region. Single LTR similar to those of HML-6.28 and HML-6.29 were also found in PAC clone172K2 (GenBank Accession No. Z84814) close to the DRB3 gene of MHC (our unpublished observation; data not shown). In chimpanzees,

alignment of amino acids

```
HERVK(IDDM)      VTPVTWMDNPIEVYVNDSVWVPGP
HML-6.29         VRPILWSDALSEIYHDQGAWGSRT
HERV-K10         IRAVTWMDNPTEVYVNDSVWVPGP
HERV-K(C4)       IRPVTWLKPPVEVYVNNSVWIPKP
JSRV             IQSLGWDREIVPVYVNDTSLLGGK
AC003682         VRPVLGSNTPPEIYHDQGAWTPGP
Z82211           SMPILWSDTPPGIYHDWGEWAPGP
                          *
```

percentage of amino acid identity

	HML-6.29	HERV-K10	HERV-K(C4)	JSRV	AC003692	Z82211
HERVK(IDDM)	28	80	60	32	40	36
HML-6.29	-	24	24	8	56	48
HERV-K10		-	68	32	36	32
HERV-K(C4)			-	28	36	28
JSRV				-	8	12
AC003682					-	60
Z82211						-

FIG. 9. Percentage identity and alignment of the amino acid sequence of the conserved motif which occurs in the amino-terminus of SU of a putative superantigen, pSAG, between HML-6.29, AC003682, Z82211, HERVK (IDDM), HERV-K10, HERV-K (C4), and JSRV. Asterisks indicate total identity, and dots indicate partial identity.

expression of DRB6 seems to be driven by an HML-6 LTR (Mayer *et al.,* 1993). Although MHC encoded surface antigens are involved in tumor rejection, it is difficult to extrapolate an oncogenetic mechanism for HML-6.28 based on available evidence. Judging from the divergence of 5′ and 3′ LTRs of HML-6 elements with double LTRs (13–15%), and calculating with a 0.3% neutral mutational drift per million years (see, e.g., Futuyma, 1998), this sequence family may have been integrated 18–25 million years ago. This contrasts with the much smaller divergence of the HML-2 element HERV-K10, which has a 0.2% 5′-3′ LTR dissimilarity (Ono, 1986), which indicates a surprisingly recent integration around 0.3 million years ago. HML-6.28 and HML-6.29 had structural features of ABD proviral DNA with widespread similarities to MMTV, MPMV, and IAP, but have multiple stop codons in *gag, pol,* and *env.* The primer binding site is complementary to tRNAlys, as for the previously reported HML-6p (Medstrand *et al.,* 1997), justifying the names HERV-K (HML-6.28) and HERV-K (HML-6.29). For a discussion on this nomenclature, see Andersson *et al.,* (1999). All class II element PBSs identified so far are complementary to tRNAlys (Wilkinson *et al.,* 1993; Goodchild *et al.,* 1993; Haltmeier *et al.,* 1995; Medstrand *et al.,* 1997).

HML-6 Gag is shorter than that of other ABD viruses. ABD Gag sequences intervening between MA, CA, and NC are very diverse. They were probably acquired separately for each virus group, at different time points. The simplest explanation is that ancient ABD viruses had a simple MA-CA-NC Gag arrangement, like HTLV (Seiki *et al.,* 1983).

Using the new sequence information provided by the sequencing of HML-6.28, HML-6.29, and the HML-6 sequences identified in GenBank, we could also reconsider the phylogeny of the dUTPases of ABD retroviruses. We compared all dUTPases that have five conserved motifs in the amino acid sequence in the same order (McGeoch, 1990). The position of IAP was different in the dUTPase tree than in trees made from other genomic components and it was even more distant from the rest than *E. coli.* It may therefore be that IAP DU has had a phylogeny different from other retroviral dUTPases and the rest of the ABD genome. Although bootstrap values are high enough to support the interpretation, a more complete spectrum of sequences is necessary for firm conclusions. This interpretation differs somewhat from earlier published interpretations (Coffin, 1996), where IAP was claimed not to have a dUTPase. According to our data, dUTPase must have been acquired very early in retroviral evolution, but on several different occasions.

We found that the amino terminus of the deduced Env of the HML-6.29 element, as well as the HML-6 elements on chromosomes 3, 19, and X, contain a 24-amino-acid motif conserved between several endo- and exogenous ABD retroviruses. The most common function of amino-terminal halves of retroviral surface glycoproteins is re-

ceptor binding (for a review, see Coffin, 1996). The conservation in this otherwise highly variable Env portion indicates that some ABD viruses share a common mechanism mediated by this stretch. It may be of interest that a putative Vβ7-specific superantigen (pSAG) encoded by the amino terminus of *env* of an HML-2 group element, HERV-K (IDDM), was recently described (Conrad *et al.,* 1997). HERV-K(IDDM) is highly related to the prototypic sequence HERV-K10 and was implicated in the pathogenesis of human type I diabetes. The HML-6.29 version of the motif has 23% amino acid identity to the HERV-K (IDDM) counterpart.

In conclusion, the finding of high expression of an HML-6 element in a human breast cancer gave us an incentive to elucidate the diversity of related sequences and to understand their evolution. We hope that this information also will aid the understanding of the role of HERVs in human disease, in particular that of HMLs in breast cancer.

MATERIALS AND METHODS

Sources of DNA and library screening

A panel of monochromosomal somatic cell hybrids (hybrid mapping panel 2) was obtained from the Human Genetic Mutant Cell Repository (HGMCR, Coriell Institute, Camden, NJ). All cell lines were cultivated as described by the provider. DNA was isolated from somatic cell hybrids, human normal peripheral blood mononuclear cells, and the parental cell lines (GM05862, mouse cell line 3T6; and GM10658, Chinese hamster cell line RJK88) according to the method presented in Ausubel *et al.,* (1987). Dubois *et al.,* (1993) and Drwinga *et al.,* (1993) have characterized the human chromosome content of these hybrid cells.

We used a human genomic library that contained partially digested 15–20 kb *Sau*3A fragments cloned in the phage vector GEM-12. It was kindly provided by Dr. Dixie Mager. The library was first screened with an HML-6.3 *pol* probe at medium stringency, which detects the HML-6 group of retroviral sequences, as described (Medstrand *et al.,* 1997). The HML-6.29 clone originated from this first screening. The identified HML-6 clones were then rehybridized with the HML-6.2*BC1* clone and washed at high stringency in 0.1× SSPE, 0.1% SDS at 68°C. Only one hybridizing clone, HML-6.28, remained.

PCR

PCR using ABDPOL/ABDPOR primers was performed as described (Yin *et al.,* 1997). The PCR contained 1 U *Taq* DNA polymerase (Perkin–Elmer Cetus), 100 ng of each primer, and 5 μl of the supernatant of plaques. Amplification products of clones were then transferred to nylon membrane (Hybond-N$^+$, Amersham International, Amersham, England) and hybridized with the HML-6.2*BC1* clone as described below.

Southern blot

Several DNAs were analyzed: (i) 10 μg from somatic cells hybrids and human lymphocytes was digested with *Hin*dIII, (ii) the products that were amplified by PCR using ABDPOL/ABDPOR primers from human genomic clones, and (iii) about 10 μg of recombinant phage λ DNAs (Sambrook *et al.,* 1989). These were digested with restriction endonucleases (*Bam*HI, *Eco*RI, *Hin*dIII, and *Sac*I; Boehringer Mannheim), singly or in binary combinations, and electrophoretically separated on 0.7% Tris–acetate–EDTA buffered agarose gels. They were Southern transferred to nylon membranes (Hybond-N$^+$, Amersham International), according to the manufacturer's instructions. The membranes contained somatic hybrid DNA, PCR products, and *pol, gag, env,* and LTR containing fragments from the λ DNAs and were hybridized to probes consisting of the ABDPOL/ABDPOR amplified portion of HML-6.2*BC1 pol* as well as HML-6 *gag, env,* and LTR fragments (see below). The following standard hybridization solution (10 ml) was used for all prehybridizations and hybridizations: 6× SSC, 1× Denhardt's solution, 1% (w/v) SDS, and heat-denatured herring sperm DNA (0.1 mg/ml). After 4 h of prehybridization at 60°C, fresh hybridization solution and 2–5 × 10^5 cpm/ml labeled probe was added and further hybridized for 18 h. Washings were performed in 0.1× SSPE, 0.1% SDS at 68°C for HML-6.2*BC1,* 0.4× SSPE, 0.1% SDS at 65°C for *gag, env,* and LTR probes.

Probes

The HML-6.2*BC1* (298 bp) probe was amplified from a human breast cancer patient by RT-PCR using ABDPOL/ABDPOR primers (Yin *et al.,* 1997). The *gag* probe was a 1.8-kb PCR fragment between the PBS site and the NC region of *gag* of clone hml-6.17. The *env* probe was a 0.8-kb *Apa*I/*Eco*RI fragment from clone hml-6.17, containing only *env.* The LTR probe was a 0.8-kb *Eco*RI fragment of clone HML-6.17 (Medstrand *et al.,* 1997). All probes were labeled with [α-^{32}P]dCTP by a random priming protocol (Sambrook *et al.,* 1989).

Sequence analysis

Restriction fragments were subcloned into vector pT7. Sequencing was performed with an automatic ABI Prism 310 sequencer by the dye-labeled terminator method using AmpliTaq DNA polymerase FS (Perkin–Elmer). After sequencing in both directions was carried out, a translation in all three reading frames was done with PC-GENE (Intelligenetics, Campbell, CA). The most likely frame was chosen by comparison with other retroviral sequences mainly using the BLAST programs (Altschul *et al.,* 1990) and the GenBank databases. In most cases, frameshifts are evident by comparison with exogenous sequences or the less mutated HERV-K10 sequence. Less conserved stretches can, however, cause transla-

tion ambiguities. An example is HERV-K (C4) (Table 1A). Alignment of amino acid and nucleic acid sequences was then carried out using the Clustal W (1.5) multiple sequence alignment program (Thompson *et al.,* 1994). Columns containing gaps were edited out of alignments used for phylogenetic analysis. All alignments can be obtained from the corresponding author and at the URL http://www.kvir.uu.se. After alignment, intersequence distances were calculated with PROTDIST with a PAM similarity matrix, from the PHYLIP program package (J. Felsenstein, 1993, PHYLIP; Phylogeny Inference Package, Version 3.5p. Department of Genetics, University of Washington, Seattle). Similarity based trees were subsequently produced by the neighbor-joining algorithm (Saitou and Nei, 1987), as implemented in PHYLIP. Bootstrap values and drawing of trees were also made using PHYLIP. A program written by Dr. Patrik Medstrand produced the nucleotide similarity matrices. The sequence of clone HML-6.28 was prepared by (i) sequencing of overlapping subclones of restriction fragments and (ii) primer walking to bridge a few gaps not covered by the fragment sequences. All sequences were done in both forward and reverse directions. The sequences of HML-6.28 and HML-6.29 have been submitted to GenBank. Their accession numbers are AF069508 (HML-6.28), and AF079797 (HML-6.29).

Fluorescence in situ hybridisation (FISH)

FISH was carried out as described by Pinkel *et al.,* (1986). The probe (clone HML-6.28) was a clone derived from the human genomic library and detected by PCR and hybridization using HML-6.2*BC1* as probe under high stringency (see "Sources of DNA"). The probe was pooled and labeled with biotin-16–deoxyuridine triphosphate (Boehringer Mannheim) using random hexanucleotides (Amersham) and hybridized to slides containing metaphase chromosomes. The biotinylated probe was detected using one layer of avidin-fluorescein isothiocyanate and one layer of goat antiavidin-FITC, both at a concentration of 5 μg/ml. The slides were mounted in 1.0 M Tris–HCL:glycerol (1:9) containing 2% DABCO (Sigma), 0.02% thimerosal, and propidium iodide. The hybridization signals were analyzed in an Axioskop microscope (Zeiss) coupled to a Cyto Vision Ultra system (Applied Imaging) using a Charged Coupled Device (CCD) camera.

ACKNOWLEDGMENTS

We thank Per Alm and Jolanta Juraszczyk for technical assistance. The European Commission, project GENE-CT930019, and funds at the Medical Faculty of Lund and local funds at Uppsala, Sweden supported this work.

REFERENCES

Altschul, S. F., Gish, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.

Andersson, M. L., Medstrand, P., Yin, H., and Blomberg, J. (1996). Differential expression of human endogenous retroviral sequences similar to mouse mammary tumor virus in normal peripheral blood mononuclear cells. *AIDS Res. Hum. Retroviruses* **12,** 833–840.

Andersson, M. L., Lindeskog, M., Medstrand, P., Westley, B., and Blomberg, J. (1999). Diversity of human endogenous retrovirus class II sequences. *J. Gen. Virol.* **80,** 255–269.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1987). "Current Protocols in Molecular Biology." Wiley, New York.

Bengtsson, A., Blomberg, J., Nived, O., Pipkorn, R., Toth, L., and Sturfelt, G. (1996). Selective antibody reactivity with peptides from human endogenous retroviruses and nonviral poly (amino acids) in patients with systemic lupus erythematosus. *Arthritis Rheum.* **39,** 1654–1663.

Björnberg, O. (1995). "Viral dUTPases: Recombinant Expression, Purification, and Substrate Specificity." Ph.D. thesis in biochemistry, Lund University.

Bradac, J., and Hunter, E. (1984). Polypeptides of Mason–Pfizer monkey virus. I. Synthesis and processing of the *gag*-gene products. *Virology* **138,** 260–275.

Coffin, J. M. (1996). Retroviridae: The viruses and their replication. *In* "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, *et al.,* Eds.), 3rd ed., pp. 1767–1847. Lippincott-Raven, Philadelphia.

Conrad, B., Weissmahr, R. N., Böni, J., Arcari, R., Schupbach, J., and Mach, B. (1997). A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. *Cell* **90,** 303–313.

Craven, R. C., Leure-duPree, A. E., Weldon, R. A., Jr., and Wills, J. W. (1995). Genetic analysis of the major homology region of the Rous sarcoma virus Gag protein. *J. Virol.* **69,** 4213–4227.

Dangel, A. W., Mendoza, A. R., Baker, B. J., Daniel, C. M., Carroll, M. C., Wu, L.-C., and Yu, C. Y. (1994). The dichotomous size variation of human complement C4 genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* **40,** 425–436.

Delwart, E. L., Mosialos, G., and Gimore, T. (1990). Retroviral envelope glycoproteins contain a "leucine zipper"-like repeat. *AIDS Res. Hum. Retroviruses* **6,** 703–706.

Drwinga, H. L., Toji, L. H., Kim, C. H., Greene, A. E., and Mulivor, R. A. (1993). NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2. *Genomics* **16,** 311–314.

Dubois, B. L., and Naylor, S. L. (1993). Characterization of NIGMS human/rodent somatic cell hybrid mapping panel 2 by PCR. *Genomics* **16,** 315–319.

Franklin, G. C., Chretien, S. C., Hanson, I. M., Rochefort, H., May, F. E. B., and Westley, B. R. (1988). Expression of human sequences related to those of mouse mammary tumor virus. *J. Virol.* **62,** 1203–1210.

Futuyma, D. J. (1998). The neutral theory of molecular evolution. *In* "Evolutionary Biology" (D. J. Futuyma, Ed.), 3rd ed., pp. 320–327. Sunderland, MA.

Goodchild, N. L., Wilkinson, D. A., and Mager, D. L. (1993). Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. *Virology* **196,** 778–788.

Gottlieb, R. A., Lennarz, W. J., Knowles, R. D., Cianciolo, G. J., Dinarello, C. A., Lachman, L. B., and Kleinerman, E. S. (1989). Synthetic peptide corresponding to a conserved domain of the retroviral protein p15E blocks IL-1-mediated signal transduction. *J. Immunol.* **15,** 4321–4328.

Haltmeier, M., Seifarth, W., Blusch, J., Erfle, V., Hehlmann, R., and Leib-Mösch, C. (1995). Identification of S71-related human endogenous retroviral sequences with full-length *pol* genes. *Virology* **209,** 550–560.

Horton, R., Niblett, D., Milne, S., Palmer, S., Tubby, B., Trowsdale, J., and Beck, S. (1998). Large-scale sequence comparison reveal unusually high levels of variation in the *HLA-DQB*1 locus in the class II region of the human MHC. *J. Mol. Biol.* **282,** 71–79.

Jacks, T., Townsley, K., Varmus, H. E., and Majors, J. (1987). Two efficient ribosomal frameshifting events are required for synthesis of mouse

mammary tumor virus *gag*-related polyproteins. *Proc. Natl. Acad. Sci. USA* **84,** 4298–4302.

Kambhu, S., Falldorf, P., and Lee, J. S. (1990). Endogenous retroviral long terminal repeats within the HLA-DQ locus. *Proc. Natl. Acad. Sci. USA* **87,** 4927–4931.

Karlin, S., and Altschul, S. F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA* **87,** 2264–2268.

Lundberg, L. G., Thoresson, H. O., Karlström, O. H., and Nyman, P. O. (1983). Nucleotide sequence of the structure gene for dUTPase of *Escherichia coli* K-12. *EMBO J.* **2,** 967–972.

Mayer, W. E., O'Huigin, C., and Klein, J. (1993). Resolution of the HLA-DRB6 puzzle: A case of grafting a de novo-generated exon on an existing gene. *Proc. Natl. Acad. Sci. USA* **90,** 10720–10724.

McClure, M. A., Johnson, M. S., Feng, D. F., and Doolittle, R. F. (1988). Sequence comparisons of retroviral proteins: Relative rates of change and general phylogeny. *Proc. Natl. Acad. Sci. USA* **85,** 2469–2473.

McGeoch, D. J. (1990). Protein sequence comparisons show that the "pseudoproteases" encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Res.* **18,** 4105–4110.

McIntosh, E. M., Ager, D. D., Gadsden, M. H., and Haynes, R. H. (1992). Human dUTPase pyrophosphatase: cDNA sequence and potential biological importance of the enzyme. *Proc. Natl. Acad. Sci. USA* **89,** 8020–8024.

Medstrand, P., and Blomberg, J. (1993). Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: Differential transcription in normal human tissues. *J. Virol.* **67,** 6778–6787.

Medstrand, P., Mager, D. L., Yin, H., Dietrich, U., and Blomberg, J. (1997). Structure and genomic organization of a novel human endogenous retrovirus family: HERV-K (HML-6). *J. Gen. Virol.* **78,** 1731–1744.

Mietz, J. A., Grossman, Z., Lueders, K. K., and Kuff, E. L. (1987). Nucleotide sequence of a complete mouse intracisternal A-particle genome: Relationship to known aspects of particle assembly and function. *J. Virol.* **61,** 3020–3029.

Moore, R., Dixon, M., Smith, R., Peters, G., and Dickson, C. (1987). Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: Two frameshift suppression events are required for translation of *gag* and *pol*. *J. Virol.* **61,** 480–490.

Niimi, N., Wajjwalku, W., Ando, Y., Nakamura, N., Ueda, M., and Yoshikai, Y. (1995). A novel V beta 2-specific endogenous mouse mammary tumor virus which is capable of producing a milk-borne exogenous virus. *J. Virol.* **69,** 7269–7273.

Ono, M. (1986). Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J. Virol.* **58,** 937–944.

Ono, M., Yasunaga, T., Miyata, T., and Ushikubo, H. (1986). Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *J. Virol.* **60,** 589–598.

Perron, H., Garson, J. A., Bedin, F., Beseme, F., Paranhos-Baccala, G., Komurian-Pradel, F., Mallet, F., Tuke, P. W., Votsset, C., Blond, J. L., Lalande, B., Seigneurin, J. M., and Mandrand, B. (1997). Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. *Proc. Natl. Acad. Sci. USA* **94,** 7583–7588.

Pinkel, D., Gray, J. W., Trask, B., Van Den Engh, G., Fuscoe, J., and Van Dekken, H. (1986). *Cold Spring Harb Symp. Quant. Biol.* **1,** 151–157.

Racevsks, J., and Prakash, O. (1983). Proteins encoded by the long terminal repeat region of mouse mammary tumor virus: Identification by hybrid-selected translation. *J. Virol.* **51,** 604–610.

Ramsdale, E. E., Kingsman, S. M., and Kingsman, A. (1996). The "putative" leucine zipper region of murine leukemia virus transmembrane protein (P15e) is essential for viral infectivity. *Virology* **220,** 100–108.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sauter, M., Schommer, S., Kremmer, E., Remberger, K., Dolken, G., Lemm, I., Buck, M., Best, B., Neumann-Haefelin, D., and Mueller-Lantzsch, N. (1995). Human endogenous retrovirus K10: Expression of Gag protein and detection of antibodies in patients with seminomas. *J. Virol.* **69,** 414–421.

Scheidereit, C., and Beato, M. (1984). Contacts between hormone receptor and DNA double helix within a glucocorticoid regulatory element of mouse mammary tumor virus. *Proc. Natl. Acad. Sci. USA* **82,** 3029–3033.

Schulte, A. M., Lai, S., Kurtz, A., Czubayko, F., Riegel, A. T., and Wellstein, A. (1996). Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. *Proc. Natl. Acad. Sci. USA* **93,** 14759–14764.

Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983). Human adult T-cell leukemia virus: Complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc. Natl. Acad. Sci. USA* **80,** 3618–3622.

Sonigo, P., Barker, C., Hunter, E., and Wain-Hobson, S. (1986). Nucleotide sequence of Mason–Pfizer monkey virus: An immunosuppressive D-type retrovirus. *Cell* **45,** 375–385.

Strahler, J. R., Zhu, X. X., Hora, N., Wang, Y. K., Andrews, P. C., Roseman, N. A., Neel, J. V., Turka, L., and Hanash, S. M. (1993). Maturation stage and proliferation-dependent expression of dUTPase in human T cells. *Proc. Natl. Acad. Sci. USA* **90,** 4991–4995.

Svensson, A. C., Setterblad, N., Sigurdardottir, S., Rask, L., and Andersson, G. (1995). Primate DRB genes from the DR3 and DR8 haplotypes contain ERV9 LTR elements at identical positions. *Immunogenetics* **41,** 72–82.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **11,** 4673–4680.

Ting, C. N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C., and Meisler, M. H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* **6,** 1457–1465.

Varmus, H. E. (1982). Form and function of retroviral proviruses. *Science* **216,** 812–820.

Wilkinson, D. A., Mager, D. L., and Leong, J. C. (1994). Endogenous human retroviruses. *In* "The Retroviridae" (J. Levy, Ed.), Vol. 3, pp. 465–535. Plenum, New York.

Willett, W., and Stampfer, M. (1989). Dietary fat intake and breast cancer risk. *J. Natl. Cancer Inst.* **81,** 1422–1424.

Wills, J. W., Cameron, C. E., Wilson, C. B., Yan, X., Nennett, R. P., and Leis, J. (1994). An assembly domain of the Rous sarcoma virus Gag protein required late in budding. *J. Virol.* **68,** 6605–6618.

Yaniv, A., Dahlberg, J., Gazit, A., Sherman, L., Chiu, I. M., Tronick, S. R., and Aaronson, S. A. (1986). Molecular cloning and physical characterization of integrated equine infectious anemia virus: Molecular and immunologic evidence of its close relationship to ovine and caprine lentiviruses. *Virology* **154,** 1–8.

Yin, H., Medstrand, P., Andersson, M. L., Borg, Å., Olsson, H., and Blomberg, J. (1997). Transcription of human endogenous retroviral sequences related to mouse mammary tumor virus in human breast cancer and placenta: Similar pattern in most malignant and nonmalignant breast tissues. *AIDS Res. Hum. Retroviruses* **13,** 507–516.

York, D. F., Vigne, R., Verwoerd, D. W., and Querat, G. (1992). Nucleotide sequence of the jaagsiekte retrovirus, an exogenous and endogenous type D and B retrovirus of sheep and goats. *J. Virol.* **66,** 4930–4939.