

Structures of Endogenous Noncotropic Murine Leukemia Virus (MLV) Long Terminal Repeats in Wild Mice: Implication for Evolution of MLVs

KEIZO TOMONAGA AND JOHN M. COFFIN*

*Department of Molecular Biology and Microbiology, Tufts University,
School of Medicine, Boston, Massachusetts 02111*

Received 29 October 1998/Accepted 13 February 1999

To develop a better understanding of the interaction between retroviruses and their hosts, we have investigated the polymorphism in endogenous murine leukemia proviruses (MLVs). We used genomic libraries of wild mouse DNAs and PCR to analyze genetic variation in the proviruses found in wild mouse species, including *Mus musculus* (*M. m. castaneus*, *M. m. musculus*, *M. m. molossinus*, and *M. m. domesticus*), *Mus spretus*, and *Mus spicilegus*, as well as some inbred laboratory strains. In this analysis, we detected several unique forms of sequence organization in the U3 regions of the long terminal repeats of these proviruses. The distribution of the proviruses with unique U3 structures demonstrated that xenotropic MLV-related proviruses were present only in *M. musculus* subspecies, while polytropic MLV-related proviruses were found in both *M. musculus* and *M. spretus*. Furthermore, one unique provirus from *M. spicilegus* was found to be equidistant from ecotropic provirus and noncotropic provirus by phylogenetic analysis. This provirus, termed HEMV, was thus likely to be related to the common ancestor of these MLVs. Moreover, an ancestral type of polytropic MLV-related provirus was detected in *M. spretus* species. Despite their “ancestral” phylogenetic position, proviruses of these types are not widespread in mice, implying more-recent spread by infection rather than inheritance. These results imply that recent evolution of these proviruses involved alternating periods of replication as virus and residence in the germ line.

All inbred laboratory strains of mice contain numerous endogenous proviruses, of which those related to murine leukemia virus (type C MLVs) are the best-characterized group. Endogenous MLVs are divided into two major groups, ecotropic and noncotropic, classified by their potential host ranges (4). Ecotropic proviruses are present in one to five copies in some, but not all, common laboratory mouse strains (26, 30, 40). Noncotropic viruses are subdivided into three major groups, termed xenotropic, polytropic, and modified polytropic, and are present in about 20 copies each in the genome of inbred mice (21, 30, 50). Many of these proviruses have been chromosomally mapped in laboratory strains, and several have been molecularly cloned and sequenced (3, 7, 19, 20, 21, 23, 27, 28, 32, 37). These studies demonstrated that the noncotropic MLV proviruses are highly polymorphic in their insertion sites and exhibit limited genetic variation from one provirus to the next. Our previous studies have shown that the members of each group of noncotropic proviruses share a set of linked polymorphisms in *env* and the long terminal repeat (LTR) regions that distinguishes them from the members of other groups (12, 49). Most usefully, the polymorphisms allowed us to develop a set of oligonucleotide probes that unambiguously detect members of each noncotropic group in the mouse genome (18, 50).

The noncotropic proviruses are more widely distributed than the ecotropic proviruses and are also found in wild mouse species, especially *Mus musculus*, which are the progenitors of common inbred laboratory strains (7, 25, 26, 30, 54, 55, 63). The distribution of the noncotropic proviruses in taxonomi-

cally distinct wild mouse species indicates that these germ line sequences were acquired independently and have remained largely segregated in *Mus* (30, 55). Either interbreeding between different subspecies or cross-species infection could have contributed to this spread. Some of the noncotropic proviruses in wild mice have been cloned and analyzed (7, 55). Recently, we have demonstrated that, although in common laboratory strains the linkage of group-specific sequences of the proviruses is strict, proviruses that combine *env* and LTR sequences from different groups are commonly observed in *M. musculus* subspecies (55). Furthermore, we have found extensive genetic variation of noncotropic proviruses in the wild mice (55). These characteristics of the endogenous noncotropic MLV proviruses provide better understanding not only of the host-retrovirus interaction but also of coevolution of MLVs and their murine host. Furthermore, because MLVs have survived as both viruses and endogenous proviruses in their murine hosts, analysis of the polymorphism of the noncotropic MLV proviruses could give us a good way to understand adaptation of the MLVs to the hosts. In fact, the presence of endogenous proviruses that have undergone recombination with other endogenous viruses in wild mice implies that the endogenous proviruses have also adapted as viruses in their hosts (55).

We describe here a systematic investigation of polymorphism of noncotropic MLV proviruses in wild mice, including *M. musculus* subspecies (*M. m. castaneus*, *M. m. musculus*, *M. m. molossinus*, and *M. m. domesticus*), *Mus spretus*, and *Mus spicilegus* (formerly known as *Mus hortulanus*), as well as some inbred laboratory strains. Using genomic libraries of wild mouse DNA and specific oligonucleotide probes, we detected several unique U3 regions of noncotropic proviruses and determined their detailed distribution in the wild mice. Although most of the structures were found only in *M. musculus* subspe-

* Corresponding author. Mailing address: Department of Molecular Biology and Microbiology, Tufts University, School of Medicine, 136 Harrison Ave., Boston, MA 02111. Phone: (617) 636-6528. Fax: (617) 636-4086. E-mail: jcoffin_par@opal.tufts.edu.

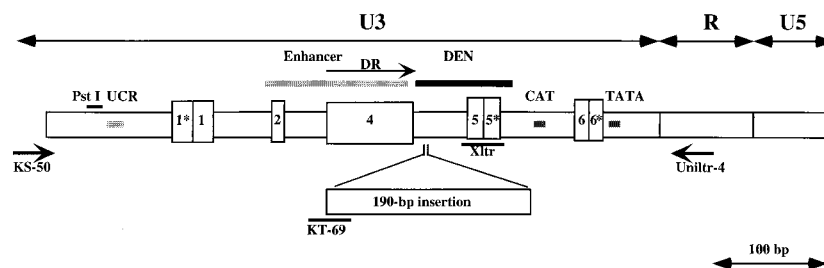


FIG. 1. Structure of the MLV LTR. The unique sequences in the U3 region are indicated. The locations of primers and probes used in this study are also shown. The numbering of the boxed regions corresponds to that in our previous study (55). UCR, upstream conserved region (17); DR, direct repeat.

cies, some were distributed in both *M. musculus* and other distinct species. Furthermore, we could detect possible ancestral forms of the noncotropic MLVs. This paper reports a possible evolutionary relationship between MLVs and wild mouse species.

MATERIALS AND METHODS

DNAs. In addition to four common inbred laboratory strains (AKR/J, HRS/J, C3H/HeJ, and C57BL/6J), DNAs from several species of *Mus* were used in this study. The DNAs of CAST/Ei, CASA/Rk (*M. m. castaneus*), CZECH II/Ei (*M. m. musculus*), MOLC/Rk, MOLF/Ei, MOLG/Dn (*M. m. molossinus*), WSB/Ei, ZALENDE/Ei (*M. m. domesticus*), SF/CamEi, PERA/Rk, PERC/Ei (*M. musculus*), SPRET/Ei (*M. spretus*), and PANCEVO/Ei (*M. spicilegus*) strains were obtained from the Mouse DNA Resource of The Jackson Laboratory, Bar Harbor, Maine. Samples of *M. spicilegus* (Halbturn), *Mus cervicolor*, *Mus caroli*, and *Mus cookii* DNA were kindly supplied by Christine A. Kozak of the National Institute of Allergy and Infectious Diseases, Bethesda, Maryland.

Preparation of genomic libraries. DNAs from the CZECH II/Ei (*M. m. musculus*) and MOLC/Rk (*M. m. molossinus*) strains were partially digested with the restriction enzyme *Bam*HI. The DNAs were then ligated into the *Bam*HI site of lambda EMBL3 vector (Stratagene), packaged, and subsequently amplified in XL1-blue MRA-P2 cells. The library was plated out and then screened by using a 5'-end ³²P-labeled xenotropic MLV-related or polytropic MLV-related (KT-69) oligonucleotide probe. The xenotropic virus-related probe (Xltr) is described elsewhere (18). Positive clones were selected and purified. Some of the unique clones were subcloned into the *Bam*HI site of the pBluescript SK(-) plasmid vector (Stratagene) and sequenced.

Dried gel hybridization and oligonucleotide probes. Hybridization in dried agarose gels (unblotting) was described previously (51). Briefly, genomic DNA digested with appropriate restriction enzymes was electrophoresed in a 0.8% agarose gel. After staining with ethidium bromide (EtBr), the DNA in the gel was denatured. After drying, the gel was hybridized for 16 h with a 5'-end ³²P-labeled oligonucleotide probe (0.5 × 10⁶ cpm/ml). The dried gel was then washed, briefly air dried, and exposed to X-ray film for 1 to 5 days by using an intensifying screen at -70°C. Sequences of oligonucleotide probes specific for each noncotropic provirus LTR region (Xltr, Pltr, and Mltr) and the hybridization temperatures are described elsewhere (18). Sequences of oligonucleotide probes specific for other types of the noncotropic proviruses (and the hybridization temperatures) are as follows: KT-51, 5'-TAC TAG GAC AAG GGC CAA ACA GG-3' (62°C); KT-53, 5'-GAC AAG GGC CAA GAA CCG ATG GTA C-3' (62°C); KT-55, 5'-GGA TAT CTG TGG TCG AGC ACC TGG-3' (60°C); KT-58, 5'-GGC TGA ATA GGT ATC GGT GGT-3' (58°C); KT-59/60, 5'-GGA AGT TCA GTT A(G/A)A GAT CAA GGC TG-3' (60°C); KT-61, 5'-GCC ATA AGC AAG CTA GCA ATA GTA AC-3' (60°C); KT-69, 5'-GAA CCA GCA ACA GAC ACA GAA G-3' (59°C); and KT-76, 5'-GCT GCC ATT TTG CAA GGC ATA G-3' (59°C).

Synthetic oligonucleotide primers and PCR analysis. To detect noncotropic MLV proviruses in mouse genomes by PCR, several amplification primers were generated in this study. The nucleotide sequences of the primers are as follows: Unltr-4, 5'-CGG GCG ACT CAG TCT ATC GG-3'; KS-50, 5'-CAG TAT CAC CAA CTC AAA TC-3'; Uniev-3, 5'-GGA TAC ACG CCG CTC ACG TA-3'; HE-8, 5'-CTA CAG AAC CGT AGA GGA CT-3'; and KA-59C, 5'-CAG CCT TGA TCT CTA ACT GAA CTT CC-3'. The locations of these sequences on the proviruses are shown in the relevant figures. PCRs were carried out in a total volume of 50 μl containing 0.5 μg of genomic DNA, 50 pmol each of sense and antisense primers, and 2.5 U of *Thermus aquaticus* DNA polymerase (*Taq* polymerase; Perkin-Elmer Cetus). The reaction mixtures for amplification were incubated at appropriate temperatures, and the cycle was repeated 30 times in a GeneAmp PCR system 2400 (Perkin-Elmer Cetus).

Cloning and sequencing analysis. The endogenous provirus fragments detected by PCR were purified from agarose gels and blunt ended by T4 DNA polymerase (New England Biolabs, Inc.). The fragments were then cloned into

the *Sma*I site of the pUC119 vector. DNA sequences were determined in accordance with the double-stranded dideoxy-chain termination method (45) by using the Sequenase version 2.0 kit (United States Biochemical Co.).

Sequence alignments and phylogenetic analysis. Sequences were aligned by using the algorithm of Needleman and Wunsch (39) as implemented in the PILEUP program in the Genetics Computer Group (GCG) program (14). For phylogenetic analysis, genetic distances between pairs of the sequences were calculated by the DNADIST or PROTDIST program in PHYLIP version 3.5 (16), using Kimura's two-parameter model (29). A neighbor-joining tree was estimated by NEIGHBOR program and a bootstrap analysis (15) using 1,000 bootstrap replications to assess the support at each of the internal nodes of the neighbor-joining tree.

Nucleotide sequence accession numbers. Provirus sequences reported in this study have been deposited in GenBank under accession no. AF070719 to AF070732.

RESULTS

Our previous study demonstrated that noncotropic MLV proviruses show extensive genetic variation in wild mice (55). In particular, U3 regions of the proviruses are highly polymorphic, reflecting evolution of the MLVs in the mice. To investigate genetic variation of U3 regions in the endogenous noncotropic MLV proviruses in more detail, we generated genomic libraries from *M. m. musculus* and *M. m. molossinus* DNAs. To analyze unique U3 forms of the noncotropic proviruses, the libraries were screened by using xenotropic or other type-specific oligonucleotide probes, and individual fragments were selected and subcloned (see Materials and Methods). A number of clones containing either the 5' or the 3' LTR sequences of noncotropic MLV proviruses were selected for analysis.

In addition to genomic libraries, U3 regions of the proviruses were also cloned following PCR amplification. For the PCR, we designed an antisense primer, Unltr-4, in the R region of the LTR that annealed to all groups of noncotropic proviruses and a sense primer, KS-50, derived from a conserved region near the 5' end of the LTR. We used this primer pair for amplification of U3 regions from wild mouse DNAs. The exact sequences and locations of the primers are shown in Materials and Methods and Fig. 1, respectively. Amplified products were cloned into a plasmid vector and then sequenced. We have analyzed a total of 55 U3 clones from *M. musculus* and *M. spretus* species, designated by the prefix Mxv, Mcv, or SPR (Table 1). U3 sequences of unique clones were aligned with those of previous isolates of MLVs. Although U3 sequences of the noncotropic proviruses are highly conserved, reflecting their recent evolutionary relationship, the proviruses could be clearly classified by unique structural features resulting from duplication, deletion, or point mutation of the sequences.

The sequence regions in U3 useful for classification of the structures are indicated in Fig. 1. The numbers of the boxed regions correspond to those in our previous study (55). The 5' end is the most conserved site among known MLV U3 regions.

TABLE 1. Summary of noncotropic MLV U3 clones from wild mice

Species	Clone	Type	
<i>M. m. musculus</i>	Mxv2	X-II	
	Mxv3	X-II	
	Mxv4	P-III	
	Mxv7	X-II	
	Mxv8	X-II	
	Mxv9	X-I	
	Mxv10	X-II	
	Mxv11	X-IV	
	Mxv16	X-II	
	Mxv28	X-IV	
	Mxv36	P-IV	
	<i>M. m. molossinus</i>	Mcv3	X-III
		Mcv4	X-III ^a
		Mcv6	X-I/X-II, X-III ^b
Mcv7		X-III	
Mcv11		X-III	
Mcv15		X-III ^a	
Mcv16		X-III ^a	
Mcv17		X-I	
Mcv18		X-I	
Mcv20		X-III ^a	
<i>M. spretus</i>		SPR1	P-I
	SPR2	P-II	
	SPR5	P-IV	
	SPR6	P-I	
	SPR8	P-II	
	SPR9	P-I	
	SPR14	P-V	

^a Type X-III proviruses that have an additional region 4 enhancer sequence.

^b Possible mosaic provirus between different xenotropic types.

It includes a *Pst*I recognition site and the upstream conserved region (UCR) motif, CGCCAT (17). The approximately 100-bp sequence following region 1 contains a number of core-enhancer binding sequences that are well conserved among noncotropic proviruses. Some MLVs contain a direct repeat of the enhancer-rich region 4. The 93-bp sequence immediately 3' to region 4 includes several transcriptional factor binding sites and is called the downstream-of-the-enhancer (DEN) region (57). This region has been shown to play an important role in transcriptional activation of the LTR (10, 57). Characteristically, the polytropic MLV-related proviruses contain a unique 190-bp insertion in the 5' end of the DEN region. The U3 regions of MLVs also contain both CAT and TATA promoter-associated motifs in the 3' ends of the sequences.

In the following paragraphs, we describe some specific features of these regions, which are useful for classifying the various provirus types and understanding their evolutionary relationships.

Structure of endogenous xenotropic MLV-related provirus U3 regions. The xenotropic MLV-related proviruses could be divided into four subgroups, termed X-I to X-IV (Fig. 2). The type X-I proviruses include the previously identified proviruses *Bxv1*, NFS, and CWM, which encode infectious xenotropic virus (28, 35). *Bxv1* is also involved in the generation of oncogenic mink cell focus-forming (MCF) viruses in certain mice (23). The X-I proviruses differed from the others by the presence of only one copy of region 1 as well as the enhancer-rich region 4. This group was also distinguished from the others by a number of single-base differences, some of which affected enhancer motifs.

Type X-II and X-III xenotropic MLV-related proviruses were isolated from *M. m. musculus* and *M. m. molossinus*, respectively, as well as inbred laboratory strains. The structures of X-II and X-III were virtually identical to each other (Fig. 2). Both proviruses contained a 14-bp duplication of region 1 in addition to region 1 (1* and 1), and both types included common mutations in several regions, as well as a common insertion located 12 bp upstream of the CAT box. The X-III proviruses were distinguished by a 45-bp deletion in the middle of the enhancer region, resulting in the absence of CARG, LVb, and core-binding enhancer motifs. The deletion probably occurred by recombination between the 7-bp direct repeats, GGGCAA, found at both ends of the deleted sequence in the X-II proviruses. Furthermore, half of the type X-III clones we examined contained an additional direct repeat of region 4 (data not shown). In addition, one X-II sequence, *Xmv44* (52), contained an 11-bp deletion in the small region (region 5) repeat.

A few clones from the *M. m. musculus* library were classified as type X-IV. The U3 regions of these proviruses had the same structure as type A Poly/Xeno recombinant proviruses isolated from *M. m. musculus* and *M. m. domesticus* (55). Although the U3 structure of this type of provirus was similar to that of the X-I proviruses, several common mutations found in the X-I proviruses were not conserved in type X-IV proviruses. For example, the region 1 sequence was different from that of the X-I provirus (Fig. 2). Furthermore, the sequences of both enhancer and DEN regions clearly distinguished the X-IV proviruses from the X-I proviruses. Interestingly, this type of provirus was very similar to polytropic proviruses in the 3' end of the U3 region, including the CAT box.

Structure of endogenous polytropic MLV-related provirus U3 regions in wild mice. We next analyzed the U3 regions of polytropic MLV-related proviruses (Fig. 3). Since previous studies showed that *M. spretus* species also carry polytropic MLV-related proviruses (7, 30), we also isolated proviruses from *M. spretus* as well as *M. musculus* DNAs. U3 clones from *M. spretus* species were named with designations beginning with SPR. All proviruses contained a unique 190-bp insertion characteristic of this group (49). Despite the similarity of these proviruses, they could readily be subdivided into several types.

The structures of type P-I and P-II proviruses were identical to those of proviruses isolated from inbred laboratory strains and previously classified as polytropic (MX27) and modified polytropic (MX33), respectively (49). Both types were also found in *M. spretus*. The P-I and P-II proviruses shared a 14-bp duplication of region 1. P-I proviruses also differed by two deletions from other sequences. The lack of region 2 resulted in deletion of the CARG and LVb enhancer sequences. Furthermore, there was a 49-bp deletion in the variable 5' boundary region between the 190-bp insertion and original viral sequences where the P-II proviruses had an 11-bp deletion relative to the apparently complete P-III, -IV, and -V sequences. Both types lacked the region 5* duplication but contained a complete duplication of region 6 (Fig. 3).

Mxv4 was isolated from the *M. musculus* library and classified as a type P-III provirus, although its U3 region was virtually identical to that of the P-II proviruses, since it lacked some nucleotide changes commonly found in the P-II proviruses and it carried an "intact" 190-bp inserted sequence. Type P-IV and P-V proviruses were detected in both *M. musculus* and *M. spretus* DNAs. The 5' half of the P-IV U3 sequence had the same structure as type B Xeno/mPoly recombinant proviruses from *M. m. domesticus* (55). The sequences of the P-IV and P-V U3 regions were almost identical to each other but contained distinguishing deletions in the 5' portion. As with the P-I proviruses, the 23-bp deletion in the P-IV proviruses re-

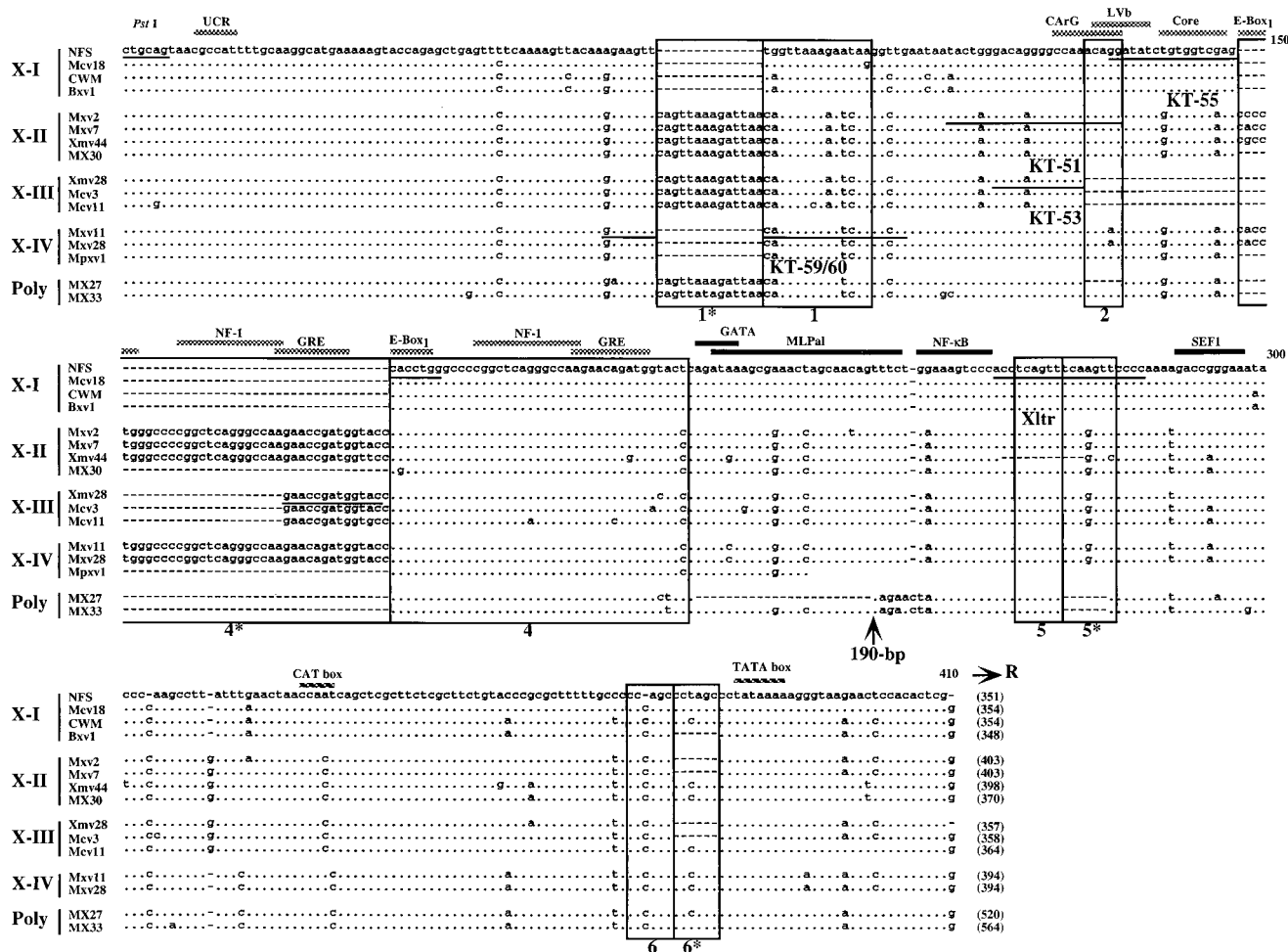


FIG. 2. U3 regions of xenotropic MLV-related proviruses. U3 sequences were cloned from *M. musculus* DNAs following amplification with the KS-50 and Unltr-4 primers. Sequences which were potentially reactive the Xltr probe and which lacked the 190-bp insert are aligned with those of several previously sequenced provirus isolates: NFS (NFS-Th-1) (28), CWM (CWM-S-5X) (35), *Bxv1* (23), *Xmv44* (52), MX30 (49), *Xmv28* (6), MX27, and MX33 (49). The sequence of NFS xenotropic provirus was used as a standard. Dots indicate nucleotide identity. Dashes indicate absence of a nucleotide. Direct repeats and unique sequences present in the proviruses are boxed. Potential enhancer sequence regions and transcriptional factor binding sites are indicated by the shaded and black bars, respectively. Locations of two promoter-associated motifs, the CAT and TATA boxes, are also indicated. The position of the 190-bp insertion in polytropic virus-related proviruses, MX27 and MX33, is shown by an arrow. The sequences of oligonucleotide probes used for unblotting analysis are underlined. The conserved *Pst*I recognition site is also shown. The size of the U3 region of each provirus is indicated at the end of the sequence in parentheses.

sulted in the absence of CARg and LVb enhancer motifs. Both types also contained intact 190-bp insertions in the sequences. Interestingly, unlike other polytropic virus-related proviruses, both proviruses lacked the 1* and 6* duplications but contained the region 5* duplication. Moreover, the CAT boxes of these types resembled those of type X-II and X-III proviruses rather than the other polytropic MLV-related proviruses.

Distribution of nonectropic MLV proviruses. To investigate in detail the distribution of nonectropic MLV proviruses in wild mice, we used structural features unique to each U3 type to design specific oligonucleotide probes. The locations of the probes are shown in Fig. 2 and 3, and their exact sequences are listed in Materials and Methods. Using these probes, we analyzed *Pvu*II-digested genomic DNAs from the wild and laboratory mouse strains listed in Table 2. The approximate number of fragments reactive with each probe is also indicated in Table 2. Because of the similarity of the U3 sequences of some proviruses, some probes detected more than one type. For example, a probe used for detection of X-IV proviruses, KT-59/60, could also hybridize to the P-IV and P-V proviruses

(Fig. 2 and 3). In this case, we determined the number of the X-IV provirus by subtracting the numbers of P-IV and P-V proviruses from those of KT-59/60 reactive fragments. As shown in Table 2, wild mice contained large and variable numbers of proviruses.

The X-I, X-II, and X-III proviruses showed similar distributions in the wild mouse species. These proviruses were found only in *M. musculus* subspecies (Fig. 4A to C and Table 2). Although the probe specific for the X-I provirus detected a few fragments in *M. m. domesticus* DNAs (Fig. 4A, lanes j, and Table 2), the dominant carriers of the X-I and X-II proviruses were *M. m. musculus* and *M. m. molossinus* (Fig. 4A and B, lanes f to i, and Table 2) with more than 50 fragments each in the latter DNA. *M. m. molossinus* DNAs also had a large number of X-III fragments, but only 12 fragments weakly reactive with the probe were found in *M. m. musculus* DNA (Fig. 4C, lane f, and Table 2). By contrast, X-IV-reactive proviruses were found in only two distinct species, *M. musculus* and *M. spicilegus* (Fig. 4D and Table 2). The X-IV probe (KT-59/60) also detected fragments in *M. spretus* DNA (Fig. 4D, lanes

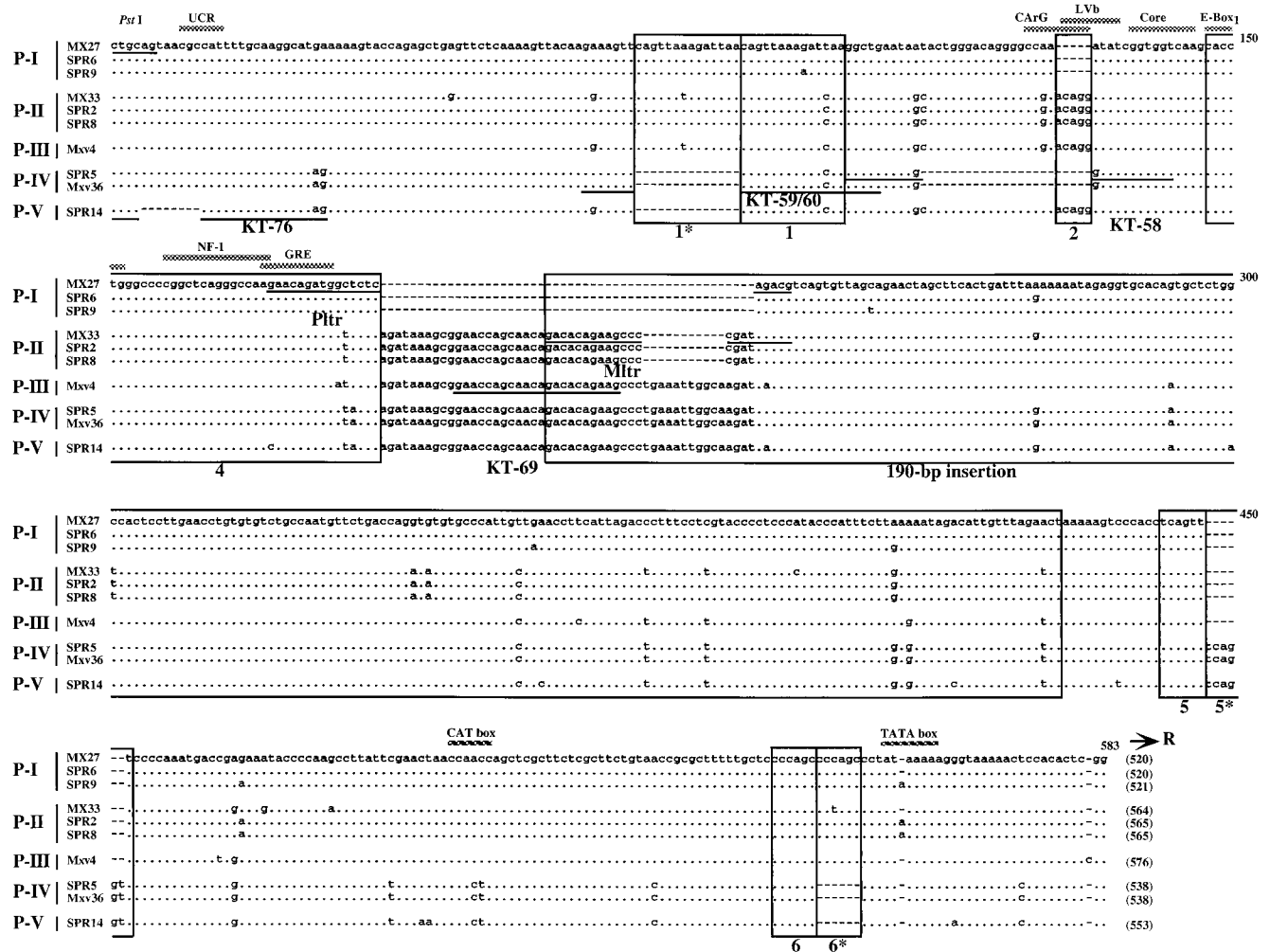


FIG. 3. Alignment of U3 sequences of polytropic MLV-related proviruses. Sequences of polytropic MLV-related U3 regions cloned from *M. musculus* and *M. spretus* DNAs as described in the previous legend are aligned with those of polytropic (MX27) and modified polytropic (MX33) proviruses (49). The MX27 sequence was used as a standard. Dots indicate nucleotide identity. Dashes indicate absence of a nucleotide. Direct repeats and unique sequences present in the proviruses are boxed (55). The 190-bp inserted region is also boxed. Potential enhancer sequence regions are indicated by the shaded bar. Locations of two promoter-associated motifs, the CAT and TATA boxes, are also indicated. The sequences of oligonucleotide probes are underlined. The conserved *Pst*I recognition site is also shown. The size of the U3 region of each provirus is indicated at the end of the sequence in parentheses.

o), but these bands belonged to either type P-IV or P-V proviruses (see below, Fig. 5C and D, lanes a to c). Comigration of the fragments found in the two different *M. spicilegus* strains (PANCEVO/Ei strain and Halbturn) (Fig. 4D, lanes p and q) implies that this endogenous provirus is widespread in this species. Furthermore, the two X-IV fragments found in *M. molossinus* species comigrated with those of two laboratory strains (Fig. 4D, lanes b, d, and g to i).

Almost all the fragments reactive with the xenotropic MLV-related provirus probes could be detected by the more generalized xenotropic MLV-related provirus LTR probe, Xltr (Fig. 4E) (55), because all types of the xenotropic MLV-related proviruses conserved region 5 (Fig. 2 and Table 2). The type-specific probes, however, allowed us to investigate the detailed distribution of each type of xenotropic MLV-related proviruses. For example, using the X-I probe we could see that the *Bxv1* provirus fragment comigrated among three laboratory strains and one *M. m. molossinus* strain (Fig. 4A, arrow). Furthermore, although we detected many xenotropic LTR-reactive fragments in both *M. m. musculus* and *M. m. molossinus*

subspecies with the Xltr probe (Fig. 4E and Table 2), X-II and X-III are the dominant types in *M. m. musculus* and *M. m. molossinus*, respectively (Fig. 4B and C and Table 2). Laboratory strains contained roughly equal numbers of the three types of proviruses, together accounting for most or all of the Xltr-reactive fragments. By contrast, fewer fragments altogether were detected by the type-specific probes in *M. m. castaneus*, *M. m. domesticus*, and *M. spretus* than by the Xltr probe (Table 2), indicating that these mice contain still another type(s) of xenotropic MLV-related provirus in the genomes. Note that the P-IV and P-V proviruses should also be detected by the Xltr probe (Fig. 3).

Our previous study demonstrated that polytropic MLV-related proviruses are widely distributed in *M. musculus* subspecies and also suggested that the proviruses existed in the *Mus* germ line before subspeciation of *M. musculus* (55). To verify and extend this observation, we next analyzed the distribution of the polytropic MLV-related proviruses using the type-specific probes (Fig. 5). The P-I and P-II specific probes (Pltr and Mltr) detected specific fragments in *M. musculus*

TABLE 2. Distribution of nonectropic proviral sequences in wild mice^a

Species	Strain	No. of fragments for:									
		Xenotropic MLV-related provirus					Polytropic MLV-related provirus				
		X-I (KT-55)	X-II (KT-51)	X-III (KT-53)	X-IV (KT-59-60)	Xltr	P-I (Pltr)	P-II (Mltr)	P-III (KT-69 ^b)	P-IV (KT-58)	P-V (KT-76)
<i>M. musculus</i> (laboratory strains)	AKR/J	6	4	19	2	25	38	24	>1	0	0
	HRS/J	10	2	7	2	19	46	23	0	0	0
	C3H/HeJ	6	3	12	0	16	34	17	0	0	0
	C57BL/6J	19	13	20	2	36	46	26	>2	0	0
<i>M. musculus</i>	SF/CamEi	4	4	4	0	13	41	40	>1	0	0
	PERA/Rk	1	0	2	4	6	21	30	>1	0	0
	PERC/Ei	0	0	0	5	5	20	34	>1	2	2
<i>M. m. castaneus</i>	CAST/Ei	2	8	6	0	35	30	5	0	0	0
	CASA/Rk	6	9	6	0	33	37	11	0	0	0
<i>M. m. musculus</i>	CZECH II/Ei	31	>60	12	19	>100	23	1	>50	2	0
<i>M. m. molossinus</i>	MOLC/Rk	>50	>50	>100	2	>100	14	6	0	0	0
	MOLF/Ei	>50	>50	>100	2	>100	14	3	0	0	0
	MOLG/Dn	>50	>50	>100	2	>100	14	5	0	0	0
<i>M. m. domesticus</i>	WSB/Ei	2	0	0	2	9	39	35	>5	0	2
	ZALENDE/Ei	0	0	0	0	11	23	>50	>3	2	2
<i>M. spretus</i>	SPRET/Ei	0	0	0	0	5	15	2	0	2	2
<i>M. hortulanus</i>	PANCEVO/Ei	0	0	0	2	0	0	0	0	0	0
	Halbturn	0	0	0	2	0	0	0	0	0	0
<i>M. cervicolor</i>		0	0	0	0	0	0	0	0	0	0
<i>M. caroli</i>		0	0	0	0	0	0	0	0	0	0
<i>M. cookii</i>		0	0	0	0	0	0	0	0	0	0

^a Numbers of fragments after *Pvu*II digestion and hybridization of dried agarose gels with the probes are shown. Note that an intact provirus gives rise to two such fragments.

^b This probe is not strictly specific (see text).

and *M. spretus* DNAs but not in others (Fig. 5A and B). The *M. spretus* DNA contained 15 and 2 fragments specific for the P-I and P-II probes, respectively (Fig. 5A and B, lanes f, and Table 2).

The distribution of the P-IV and P-V proviruses was also examined by using specific probes. The P-IV provirus-specific probe, KT-58, detected two comigrating bands, most likely representing a single copy of this type of provirus, in the *M. musculus* and *M. spretus* species DNAs (Fig. 5C). A similar observation was made by using the P-V specific probe, KT-76 (Fig. 5D), with which fragments common to *M. spretus* and one strain of *M. musculus* were also observed. These observations suggested that these proviruses were integrated into the *Mus* germ line before separation between *M. musculus* and *M. spretus* species. Unfortunately, we could not generate a P-III provirus-specific probe because it lacks a sequence feature to distinguish it from the other polytropic MLV-related proviruses. However, hybridization with the probe, KT-69, that detected all types of polytropic MLV-related proviruses except for P-I provirus implied that the P-III provirus was found only in *M. musculus* subspecies and distributed predominantly in *M. m. musculus* subspecies (Table 2).

Sequence and phylogenetic analyses of *M. spicilegus* endogenous MLV provirus. Analyses using type-specific probes suggested the possibility that the X-IV-related provirus might represent the oldest type of nonectropic provirus among those examined, since it was the only one detected in the relatively distantly related *M. spicilegus*. Thus, we cloned an X-IV-reactive fragment from this species in two steps. First, we used a PCR primer that recognized a conserved MLV group *env* sequence, Unienv-3 (Fig. 6A), and an antisense primer, KA-59C, complementary to the X-IV-specific probe (Fig. 2 and 6A) to amplify and clone the whole *env* region. To determine whether this cloned fragment was the same X-IV-related provirus detected in the analysis presented Fig. 4D, we hybrid-

ized *Pvu*II-digested *M. spicilegus* DNAs with a specific U3 region probe, KT-61, designed from the cloned product. As shown in Fig. 6B, KT-61 detected two fragments identical to those seen with KT-59/60 (Fig. 4D), indicating that the cloned fragment was indeed derived from the X-IV-related provirus. This provirus was named *hortulanus* endogenous MLV (HEMV).

To infer the host range of HEMV, we determined the nucleotide sequence of part of its *env* gene. The host range of MLVs is specified by two variable regions (VRA and VRB) in the 5' portion of the surface (SU) protein (Fig. 6A) (1). The analysis indicated that the amino acid sequences of VRA and VRB of HEMV were quite different from those of other type C leukemia viruses (Fig. 6C), although other parts of the *Env* protein were highly conserved. Most strikingly, HEMV encoded a very short VRA sequence in the genome, implying a host range distinct from that of other known MLVs.

For the second step, we designed an HEMV-specific sense primer located in the *env* region, HE-8 (Fig. 6A), and performed PCR amplification with it and the Uniltr-4 antisense primer. The PCR product was purified and sequenced. By using a specific probe, the PCR product was confirmed to encode the HEMV U3 region (data not shown). The sequence of the HEMV U3 region is shown in Fig. 7. Although this sequence was distinct from the other proviruses, it was still closely related to both ecotropic and nonectropic MLV proviruses. A highly conserved 24-bp region, including the UCR motif, was present just 3' of the absent *Pst*I site (Fig. 7A). HEMV encoded only one region 1 and the sequence showed a 1-bp difference from that of the X-IV provirus. This difference probably accounts for the relatively weak hybridization signals seen with *M. spicilegus* DNAs by using the KT-59/60 probe (Fig. 4D). Interestingly, the enhancer regions, including region 4, were well conserved among all MLV proviruses (Fig. 7A). The core-binding and E-box₁ motifs of HEMV were found in ecotropic and nonectropic types, respectively. The DEN region of

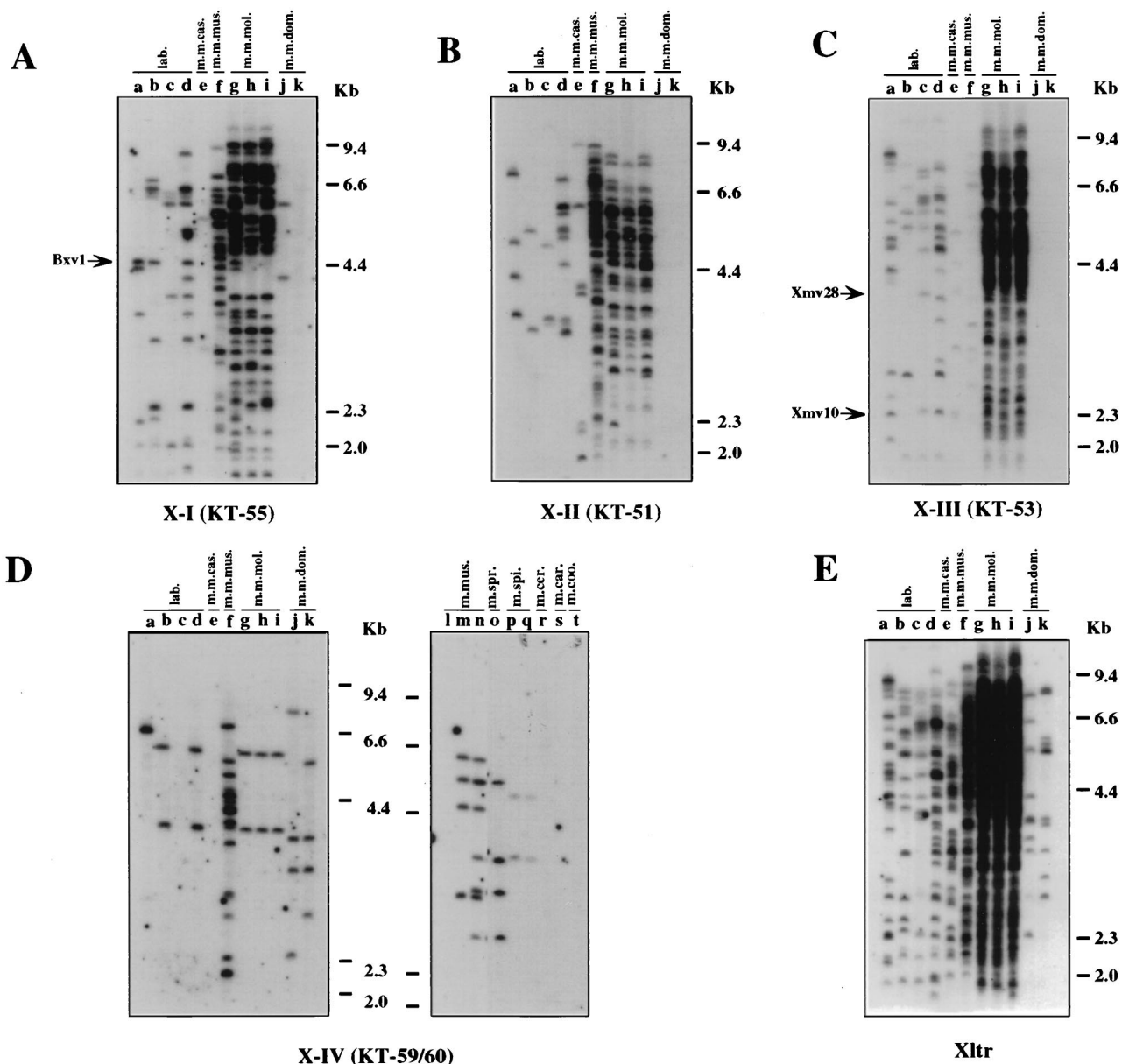


FIG. 4. Distribution of xenotropic MLV-related proviruses in wild mice. Analysis of *Pvu*II-digested mouse DNAs was performed by hybridization of dried agarose gels with an X-I-specific oligonucleotide probe (KT-55) (A), an X-II-specific oligonucleotide probe (KT-51) (B), an X-III-specific oligonucleotide probe (KT-53) (C), and the X-IV-reactive oligonucleotide probe (KT-59/60) (D). Lanes (a to d contain laboratory [lab.] strains): a, AKR/J; b, HRS/J; c, C3H/HeJ; d, C57BL/6J; e, CAST/Ei (*M. m. castaneus*) (m.m.cas.); f, CZECH II/Ei (*M. m. musculus*) (m.m.mus.); g, MOLC/Rk (*M. m. molossinus*); h, MOLF/Ei (*M. m. molossinus*); i, MOLG/Dn (*M. m. molossinus*) (m.m.mol.); j, WSB/Ei (*M. m. domesticus*); k, ZALENDE/Ei (*M. m. domesticus*) (m.m.dom.); l, SF/CamEi (*M. musculus*); m, PERA/Rk (*M. musculus*); n, PERC/Ei (*M. musculus*) (m.mus.); o, SPRET/Ei (*M. spretus*) (m.spr.); p, PANCEVO/Ei (*M. spicilegus*); q, Halbturn (*M. spicilegus*) (m.spi.); r, *M. cervicolor* (m.cer.); s, *M. caroli* (m.car.); t, *M. cookii* (m.coo.). Some known provirus loci, *Bxv1* (23), *Xmv10*, and *Xmv28* (19), are shown by arrows. The approximate positions of molecular markers are also shown.

HEMV was relatively close to that of ecotropic proviruses. Furthermore, HEMV had a complete CAT box, CCAAT, as was the case for the X-I and ecotropic proviruses. Moreover, unlike the U3 sequences of known MLV proviruses (22), HEMV lacked both region 6 and region 6* in the 3' end of U3. Interestingly, simian sarcoma virus (SSV) (13), gibbon ape leukemia virus (GaLV) (56), and feline leukemia virus (48) U3 sequences also lack these regions (Fig. 7B).

To assess the phylogenetic relationship between HEMV and the other proviruses, we generated unrooted neighbor-joining phylogenetic trees for the *env* and U3 sequences of the proviruses presented here (Fig. 8). In the *env* tree (Fig. 8A),

each group of proviruses was clearly separated from the others. Interestingly, HEMV was found to be equidistant from the ecotropic and noncrotropic proviruses as well as from the viruses of various species. The U3 tree also supports the phylogenetic position of HEMV and is also supported by bootstrap analysis (Fig. 8B and C). These observations support the possibility that HEMV might resemble a common ancestor of recent MLVs.

DISCUSSION

Genetic variation and distribution of noncrotropic MLV proviruses in wild mice. Endogenous noncrotropic MLV pro-

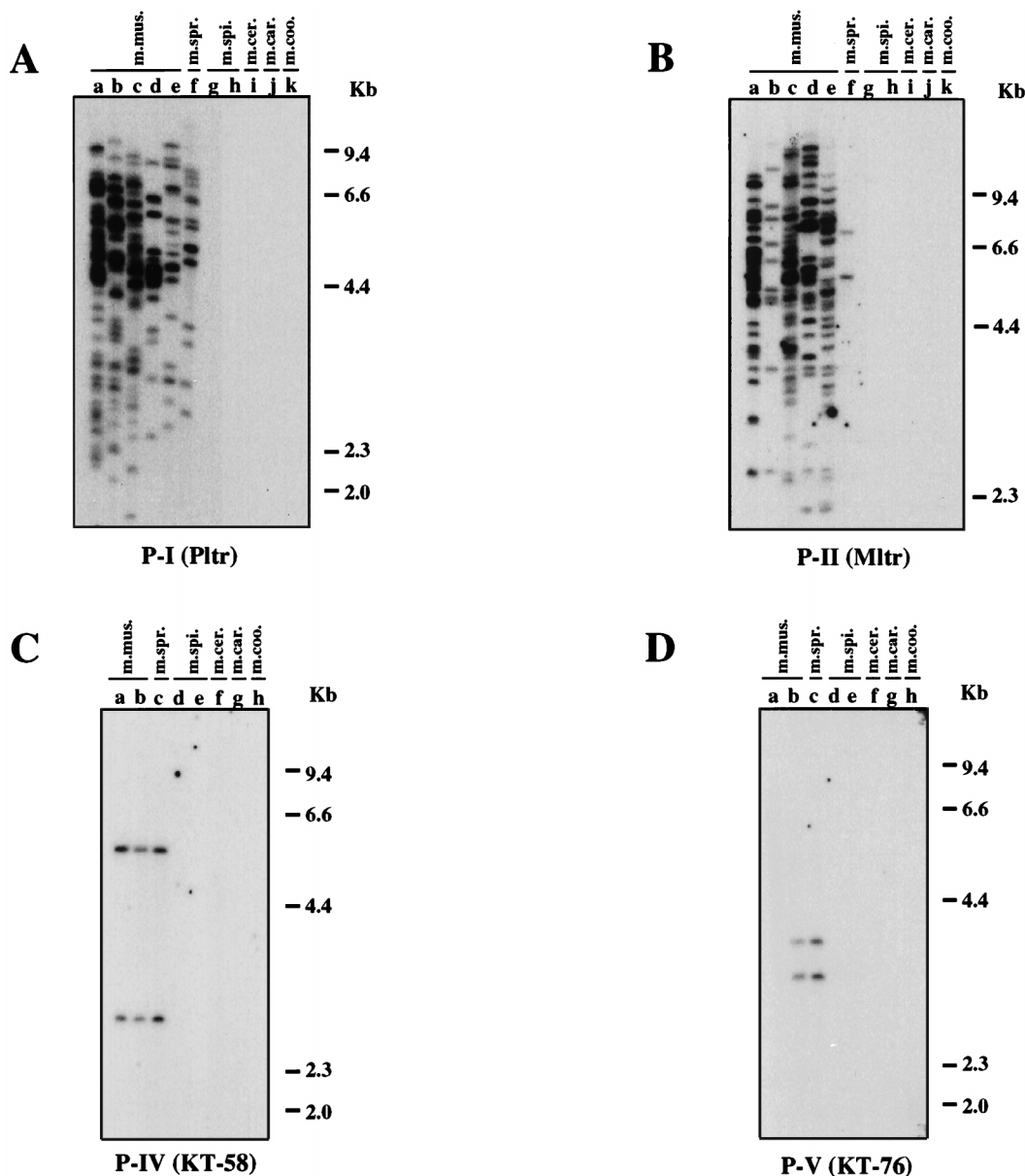


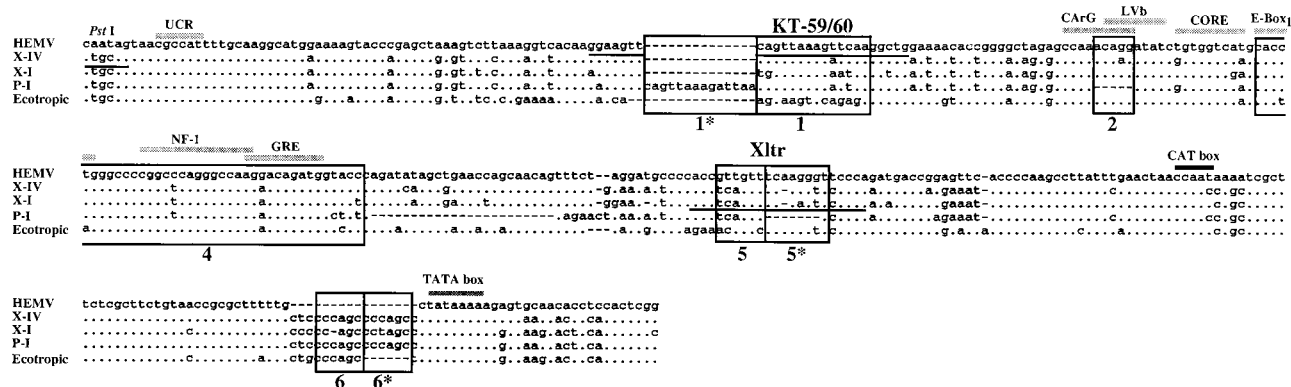
FIG. 5. Distribution of polytropic MLV-related proviruses in wild mice. Analysis of *Pvu*II-digested mouse DNAs was performed by hybridization of the DNA in dried agarose gels with the P-I-specific oligonucleotide probe (Pltr) (A), the P-II-specific oligonucleotide probe (Mltr) (B), the P-IV-specific oligonucleotide probe (KT-58) (C), and the P-V-specific oligonucleotide probe (KT-76) (D). (A and B) Lanes: a, C57BL/6J; b, CASA/Rk (*M. m. castaneus*); c, SF/CamEi (*M. musculus*); d, PERA/Rk (*M. musculus*); e, PERC/Ei (*M. musculus*) (m.mus.); f, SPRET/Ei (*M. spretus*) (m.spr.); g, PANCEVO/Ei (*M. spicilegus*); h, Halbturm (*M. spicilegus*) (m.spi.); i, *M. cervicolor* (m.cer.); j, *M. caroli* (m.car.); k, *M. cookii* (m.coo.). (C and D) a, CZECH II/Ei (*M. m. musculus*); b, ZALENDE/Ei (*M. m. domesticus*) (m.mus.); c, SPRET/Ei (*M. spretus*) (m.spr.); d, PANCEVO/Ei (*M. spicilegus*); e, Halbturm (*M. spicilegus*) (m.spi.); f, *M. cervicolor* (m.cer.); g, *M. caroli* (m.car.); h, *M. cookii* (m.coo.). The approximate positions of molecular markers are also shown.

viruses are stable genetic elements that were fixed recently in the *Mus* germ line and are therefore highly polymorphic. Endogenous ecotropic MLV proviruses are also polymorphic but are neither widely distributed nor greatly amplified in mice (4, 12), indicative of much more recent origin. Study of the non-ecotropic MLV proviruses should provide a tool to obtain a better understanding of the association between retroviruses and their host during their evolutionary history. Here, we analyzed the extended genetic variation of non-ecotropic endogenous MLV proviruses in wild mice including *M. musculus*, *M. spretus*, and *M. spicilegus*. We report here several new and distinct classes of non-ecotropic provirus U3 sequences and

their relationships to one another and to known endogenous proviruses.

The proviruses we studied showed a high degree of sequence similarity in U3, consistent with a recent evolutionary relationship. The 5' portion including the UCR motif, which is highly conserved among a large number of type C retroviruses (22), was also the most conserved among the wild-mouse proviruses. This motif has been reported to function as a negative regulatory element in several murine type C viruses (17). Other regions of U3 are less well conserved and allowed us to group the proviruses into a number of subtypes, using polymorphisms arising from duplication, deletion, insertion, or point mutation.

A



B

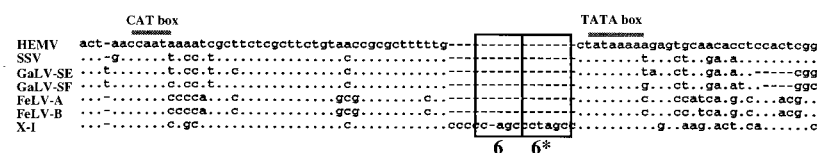


FIG. 7. Alignment of HEMV U3 sequences. (A) The U3 region of the HEMV provirus was obtained as described in the text and is aligned with the corresponding sequences from the X-IV (Mxv11), X-I (NFS-Th-1) (28), P-I (MX27) (49), and ecotropic (Akv) (58) proviruses. Dots indicate nucleotide identity. Dashes indicate absence of a nucleotide. The sequence of an oligonucleotide probe, KT-59/60, reactive with X-IV proviruses is underlined. Direct repeats in the enhancer regions found in ecotropic proviruses are not shown. The *Pst*I recognition site is also shown. (B) Lack of the region 6 direct repeat in HEMV. The 3' portion of the HEMV U3 sequence is aligned with the analogous regions of the several type C retroviruses. Abbreviations and strains of the viruses included in this alignment are as follows: GaLV-SE and GaLV-SF, SEATO and San Francisco isolates (56); FeLV-A and FeLV-B, feline leukemia virus, subgroup A (Glasgow-1) and subgroup B (Gardner-Armstein) (48); X-I, NFS-Th-1 (28).

to the murine AIDS (MAIDS) virus, were identified as proviruses containing the X-II-type U3 in their genomes. These two MAIDS-related viruses are endogenous in laboratory strains and have donated the *gag* sequences characteristic of MAIDS viruses to the related exogenous viruses by recombination (11, 31). The X-III proviruses have also been cloned before. Insertion of *Xmv28*, an X-III provirus, into the mouse genome is associated with the *rd* mutation (6).

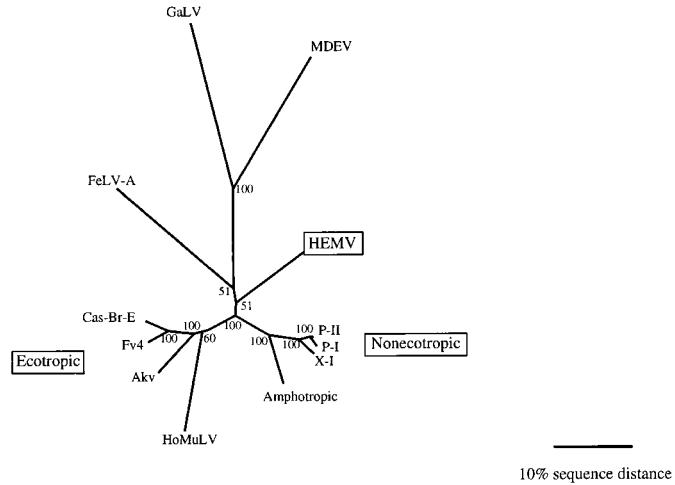
The distribution of the proviruses also has important implications for the evolutionary relationship of the xenotropic MLVs. First, most of the xenotropic relatives were found only in *M. musculus* subspecies (Fig. 4 and Table 2). In particular, *M. m. musculus* and *M. m. molossinus* were the predominant carriers of the X-I, X-II, and X-III proviruses, indicating that they entered the mouse germ line after separation between *M. musculus* and *M. spretus*, probably around the time of subspeciation between *M. m. musculus* and *M. m. domesticus*. Despite their presence, the dissimilarities of the distributions and numbers of the proviruses among the subspecies imply that a variety of amplification and germ line integration events occurred after separation of the subspecies. Furthermore, the type-specific probes revealed introgression of some proviruses involved in pathogenicity or mutation into certain laboratory mice. The inbred strains could have inherited the *Bxv1* MCF-related provirus from the *M. m. molossinus* subspecies (Fig. 4A). The MAIDS-related proviruses could have been generated after subspeciation of *M. musculus* and then acquired in laboratory strains from either *M. m. musculus* or *M. m. molossinus* (Fig. 4B). Two mutation-associated proviruses, *Xmv10* and *Xmv28* (6, 61), were probably inherited from the *M. m. molossinus* subspecies (Fig. 4C). Further study of the distribu-

tion among mice of provirus integrated at specific sites will be required to determine the relationship between integration and speciation events.

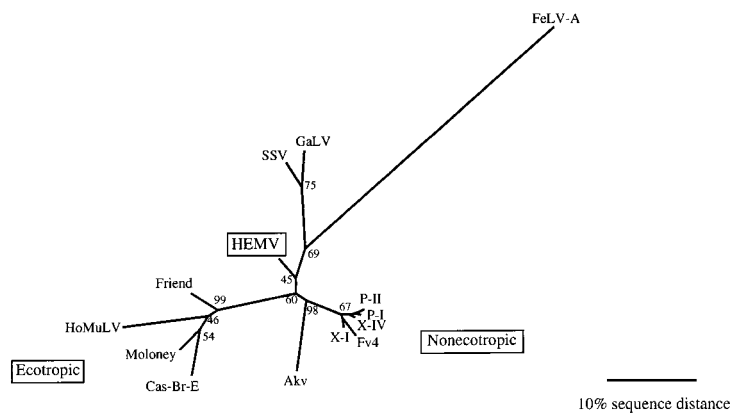
Among xenotropic MLV-related proviruses, the X-IV type showed unique genetic features and a unique distribution pattern in the mice. Despite a xenotropic MLV-related U3 structure, the X-IV proviruses shared some common features with polytropic MLV-related proviruses, including identical region 1 and CAT box sequences. Furthermore, some X-IV proviruses and Poly/Xeno recombinant proviruses from *M. m. musculus* and *M. m. domesticus* contain polytropic virus-type *env* sequences (55). Sequence analysis of a clone from our genomic library, Mxv11, indicated that the SU region of the provirus was more similar to that of polytropic viruses than to that of xenotropic viruses, although it contained a large deletion (data not shown). These observations imply that this type of provirus could be related to a recent common ancestor of the nonecotropic proviruses. It could be necessary to investigate larger numbers of samples from *M. spretus* or the related species to verify this hypothesis.

Polytropic MLV-related proviruses showed less diversity than xenotropic MLV-related proviruses in the wild mice (Fig. 3). Sequence analysis demonstrated that proviruses analogous to polytropic (type P-I) and modified polytropic (P-II) proviruses in laboratory strains (49) were also present and widely distributed in *M. spretus*. This result was also confirmed by unblotting (Fig. 5 and Table 2). Among the polytropic MLV-related proviruses, the predominant proviruses were of the P-I and P-II types, while P-IV and P-V were not widespread, although some proviruses yielded comigrating bands from DNAs from two distinct species, *M. musculus* and *M. spretus*

A. env



B. U3



C. U3

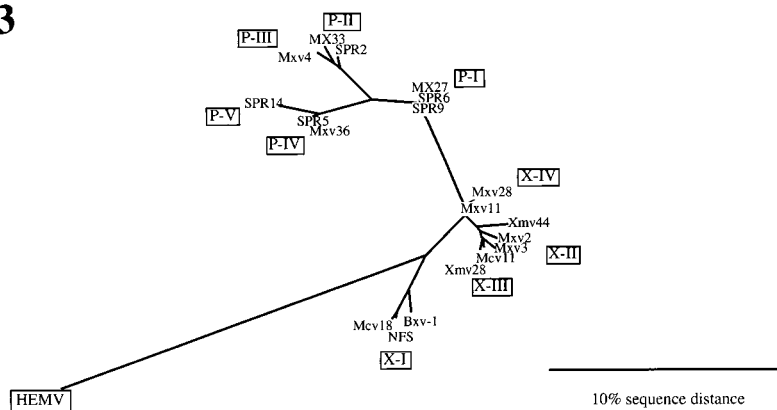


FIG. 8. Phylogenetic analysis of the nonectotropic U3 proviruses. Unrooted phylogenetic trees for the *env* (A) and U3 (B and C) regions of HEMV and several type C leukemia viruses were estimated by neighbor joining. Branch lengths are drawn to scale. To illustrate consistency all bootstrap values obtained with 1,000 replications of bootstrap sampling are shown. Viruses used for this analysis are as follows: ecotropic viruses, Akv (58), Moloney (47), Friend (FE29 strain) (43), Fv4 (M33884), Cas-Br-E (P08360), and HoMuLV (60); nonectotropic viruses, P-I (MX27) (49), P-II (MX33) (49), and X-I (NFS-Th-1) (28); amphotropic viruses, (4070A) (42), MDEV (62), GaLV (SEATO) (56), and feline leukemia virus subgroup A (FeLV-A) (Glasgow-1) (48).

(Fig. 5C and D). This result implies that polytropic MLV-related proviruses were integrated into the *Mus* germ line before separation between these species, but that only the P-I and P-II proviruses were able to spread within the *Mus* ge-

nomes. Furthermore, the sequence analysis revealed that both the P-I and P-II proviruses contain an internal deletion in the 5' portion of the 190-bp unique insertion (Fig. 3), while the P-III, P-IV and P-V proviruses contained intact forms of the

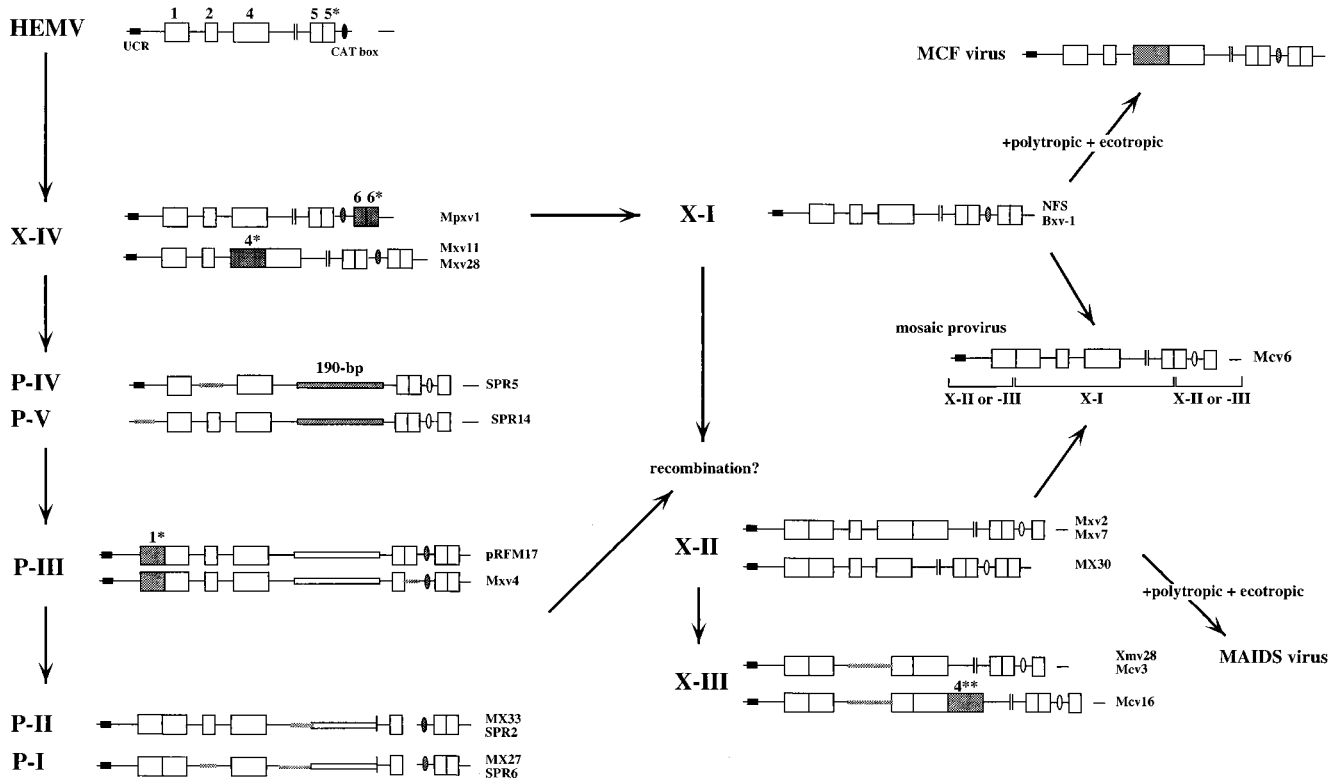


FIG. 9. A possible evolutionary scheme relating nonectropic MLV U3 regions. Schematic representations of the U3 structures are shown. Locations of the UCR and CAT box are also shown. Corresponding clones are shown on the right of the structures. Clone pRFM17 was described previously (8). Note that this scheme is parsimonious with respect to insertions and deletions. Discrepancies relative to the tree shown in Fig. 8 probably stem from these events having occurred multiple times.

inserted sequence in their genomes. This observation suggests that the P-I and P-II proviruses were derived from other polytropic MLV-related proviruses with an intact 190-bp insertion. It is likely that P-II was generated from the P-III provirus by simple deletion in the inserted sequence (see below).

One of the interesting findings of this study was obtained from analysis of the provirus from *M. spicilegus*, HEMV. Although originally detected with the type X-IV probe, the U3 region of this provirus does not belong with either the xenotropic or the polytropic group. HEMV junction fragments did comigrate in two different strains of *M. spicilegus*, indicating that it is indeed endogenous in these mice. In a previous study, a distantly related pathogenic ecotropic virus (Hortulanus MLV or HoMLV) was isolated from *M. spicilegus* species (59, 60). It is believed, however, that HoMLV is exogenous and was acquired from Asian wild mice (59, 60). The sequence and phylogenetic analyses of HEMV clearly demonstrate its close relationship to MLVs, but it is equidistant from ecotropic and nonectropic provirus groups (Fig. 7B and 8). These results suggest that HEMV is likely to be related to the common ancestor of MLVs. Despite the ancestral phylogenetic position of HEMV, however, the provirus is not widespread in wild mice (Fig. 6B). This result implies a more recent spread by infection of the provirus rather than inheritance. Furthermore, the phylogenetic position of HEMV was also found to be equidistant from simian leukemia viruses (GaLV and SSV) and MLVs. Indeed, the sequence analysis also revealed that the 3' end of its U3 region is more closely related to the simian viruses than to the murine proviruses (Fig. 7B). It is now believed that GaLV and SSV originated subsequent to a trans-species infection of primates by xenotropic viruses of murine

origin derived from *M. caroli*, *M. cervicolor*, or other related mouse species (2, 34). Although sequence analysis of the xenotropic viruses from *M. caroli* and *M. cervicolor* has not been carried out and the HEMV sequence was not detected in these species (Fig. 6B), there is a possibility that HEMV is the origin of the GaLV/SSV group in primates. Moreover, an endogenous virus (MDEV) that shows a unique host range has recently been isolated from *M. dunni* (5). The sequences of the vrA and vrB regions of both MDEV (62) and HEMV are consistent with a host range different from that of all other known viruses of this genus (Fig. 6C). However, they are also quite different from each other and occupy distinct phylogenetic positions (Fig. 8A).

Possible evolutionary relationships among MLVs in wild mice. The unique structural features and detailed distribution of the nonectropic provirus U3 regions raise several interesting possibilities regarding their evolutionary relationships. In the scheme proposed in Fig. 9, HEMV could represent the progenitor of the LTRs found in other nonectropic proviruses. The HEMV U3 sequence could have evolved into the nonectropic-type U3 and given rise to proviruses like the X-IV type following several mutations, including insertion of regions 6 and 6* and a number of duplication and recombination events. The X-IV sequences could generate P-IV and P-V proviruses by insertion of the 190-bp sequence and deletion of region 6*. The polytropic virus-like *env* sequence of the X-IV proviruses supports this evolutionary step. The P-IV and P-V proviruses seem to be ancestral forms of the polytropic MLV-related proviruses. The P-III sequence could have been derived from P-IV and P-V following duplication of region 1, and it could have given rise to P-I and P-II proviruses by deletions

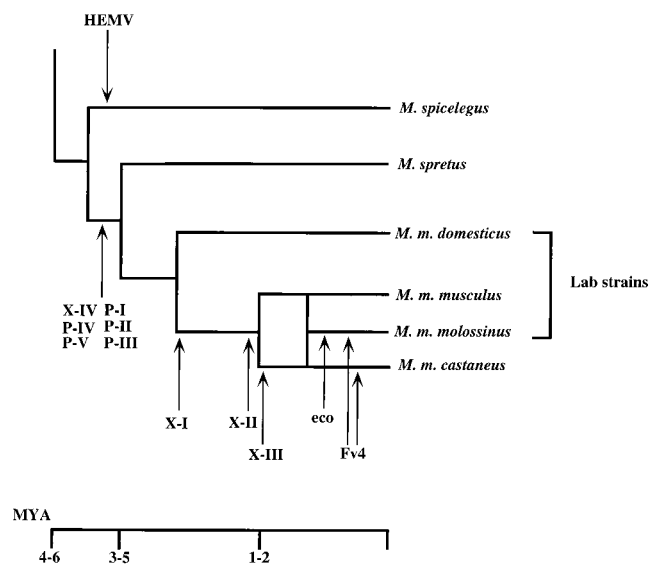


FIG. 10. Coevolution of MLVs and wild mice. A schematic phylogenetic relationship of wild-mouse species is shown. Possible periods of provirus integration are represented by arrows. The bar under the tree indicates the approximate time scale. MYA, million years ago.

including the 5' side of the 190-bp insertion. The apparent presence of type P-IV and P-V proviruses in both *M. musculus* and *M. spretus* suggests that the separation of these viruses occurred before the separation of *M. musculus* and *M. spretus* about 5 million years ago. Type X-I proviruses could have been derived from type X-IV proviruses by several mutations. The X-II and X-III proviruses could have been generated by duplication of region 1 from the X-I proviruses or recombination with polytropic MLV-related proviruses. These events could have occurred after the separation of *M. musculus* and *M. spretus*.

In order to accomplish these evolutionary steps, it would be necessary for the proviruses to have extended periods of replication as viruses and involve numerous mutational and recombination events. Both the nature of the mutational events (duplications, insertions, and deletions, particularly involving enhancer regions) and their rapid accumulation over short evolutionary periods are inconsistent with mutations occurring during residence in the mouse germ line. These events are consistent with those commonly found in the generation of the recombinant MCF and MAIDS viruses derived in part from the X-I and X-III proviruses, respectively. Furthermore, a mosaic provirus, which has the type X-I enhancer-DEN region on the X-II- or X-III-based sequence, was found in a strain of *M. m. molossinus*, Mcv6 (Fig. 9), directly implicating recombination during virus replication.

It is important to keep in mind that the scheme shown in Fig. 9 is the simplest, involving the smallest number of insertion, duplication, and deletion events. It shares many features with the point mutation-based tree shown in Fig. 8, but the two patterns are not perfectly concordant, particularly in the polytropic lineage. This discordance implies that some events may have occurred more than once during the evolution of these viruses. Given the high frequency of recombination and duplication events observed during infection of a single animal (53), it is likely that some of the duplication and deletion events have occurred multiple times during the evolution of these viruses.

Figure 10 shows a schematic representation of the coevolution of MLV proviruses and the wild-mouse species we included in this study. The observation that the ancestral-type

proviruses are not widespread in mice implies that these proviruses only recently expanded into these mice by infection rather than inheritance. Thus, rather than serving as either a static repository for ancient viruses or a reservoir from which exogenous viruses appear from time to time, endogenous proviruses should be viewed as a snapshot of a complex dynamic process involving long periods of replication as viruses, during which recombination and mutation events accumulate, and relatively recent germ line insertion. These proviruses could provide new genetic markers for the strain identification or evolutionary study of both MLVs and their murine host. The approach taken in this study will serve as the basis for further examination of the evolution and involvement of retroviral sequences in tumorigenicity in the murine host.

ACKNOWLEDGMENTS

We are grateful to Jonathan Stoye and Christopher Tipper for helpful comments, to Mary Bostic-Fitzgerald for preparing the manuscript, and to Christine Kozak for the generous gift of wild-mouse DNAs.

This work was supported by National Cancer Institute award R35CA44385 to J.M.C. and a Leukemia Society of America Special Fellowship to K.T. J.M.C. was a research professor of the American Cancer Society.

REFERENCES

- Battini, J.-L., J. M. Heard, and O. Danos. 1992. Receptor choice determinants in the envelope glycoprotein of amphotropic, xenotropic, and polytropic murine leukemia viruses. *J. Virol.* **66**:1468-1475.
- Benveniste, R. E., R. Callahan, C. J. Sherr, V. Chapman, and G. J. Todaro. 1977. Two distinct endogenous type C viruses isolated from the Asian rodent *Mus cervicolor*: conservation of viro gene sequences in related rodent species. *J. Virol.* **21**:849-862.
- Blatt, C., K. Mileham, M. Haas, M. N. Nesbitt, M. E. Harper, and M. I. Simon. 1983. Chromosomal mapping of the mink cell focus-inducing and xenotropic *env* gene family in the mouse. *Proc. Natl. Acad. Sci. USA* **80**: 6298-6302.
- Boeke, J. D., and J. P. Stoye. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, p. 343-435. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Bonham, L., G. Wolgamot, and A. D. Miller. 1997. Molecular cloning of *Mus dunni* endogenous virus: an unusual retrovirus in a new murine viral interference group with a wide host range. *J. Virol.* **71**:4663-4670.
- Bowes, C., T. Li, W. N. Frankel, M. Danciger, J. M. Coffin, M. L. Applebury, and D. B. Farber. 1993. Localization of a retroviral element within the *rd* gene coding for the *b* subunit of cGMP-phosphodiesterase. *Proc. Natl. Acad. Sci. USA* **90**:2955-2959.
- Ch'ang, L.-Y., W. K. Yang, F. E. Myer, C. K. Koh, and L. R. Boone. 1989. Specific sequence deletions in two classes of murine leukemia virus-related proviruses in the mouse genome. *Virology* **168**:245-255.
- Ch'ang, L.-Y., W. K. Yang, F. E. Myer, and D.-M. Yang. 1989. Negative regulatory element associated with potentially functional promoter and enhancer elements in the long terminal repeats of endogenous murine leukemia virus-related proviral sequences. *J. Virol.* **63**:2746-2757.
- Chattopadhyay, S. K., H. C. I. Morse, M. Makino, S. K. Ruscetti, and J. W. Hartley. 1989. Defective virus is associated with induction of murine retrovirus-induced immunodeficiency syndrome. *Proc. Natl. Acad. Sci. USA* **86**: 3862-3866.
- Chen, H., and F. K. Yoshimura. 1994. Identification of a region of a murine leukemia virus long terminal repeat with novel transcriptional regulatory activities. *J. Virol.* **68**:3308-3316.
- Cho, B. C., J. J. Shaughnessy, D. A. Largaespada, H. G. Bedigian, A. M. Buchberg, N. A. Jenkins, and N. G. Copeland. 1995. Frequent disruption of the *Nf1* gene by a novel murine AIDS virus-related provirus in BXH-2 murine myeloid lymphomas. *J. Virol.* **69**:7138-7146.
- Coffin, J. M., J. P. Stoye, and W. N. Frankel. 1989. Genetics of endogenous murine leukemia viruses. *Ann. N. Y. Acad. Sci.* **567**:39-49.
- Devare, S. G., E. P. Reddy, J. D. Law, K. C. Robbins, and S. A. Aaronson. 1983. Nucleotide sequence of the simian sarcoma virus genome: demonstration that its acquired cellular sequences encode the transforming gene product p28sis. *Proc. Natl. Acad. Sci. USA* **80**:731-735.
- Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387-395.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783-791.

16. Felsenstein, J. 1995. PHYLIP: phylogeny inference package, version 3.5. University of Washington, Seattle.
17. Flanagan, J. R., A. M. Krieg, E. E. Max, and A. S. Khan. 1989. Negative control region at the 5' end of murine leukemia virus long terminal repeats. *Mol. Cell. Biol.* **9**:739–746.
18. Frankel, W. N., and J. M. Coffin. 1994. Endogenous nonectropic proviruses mapped with oligonucleotide probes from the long terminal repeat region. *Mamm. Genome* **5**:275–281.
19. Frankel, W. N., J. P. Stoye, B. A. Taylor, and J. M. Coffin. 1989. Genetic analysis of endogenous xenotropic murine leukemia viruses: association with two common mouse mutations and the viral restriction locus *Fv-1*. *J. Virol.* **63**:1763–1774.
20. Frankel, W. N., J. P. Stoye, B. A. Taylor, and J. M. Coffin. 1989. Genetic identification of endogenous polytropic proviruses by using recombinant inbred mice. *J. Virol.* **63**:3810–3821.
21. Frankel, W. N., J. P. Stoye, B. A. Taylor, and J. M. Coffin. 1990. A linkage map of endogenous murine leukemia proviruses. *Genetics* **124**:221–236.
22. Golemis, E. A., N. A. Speck, and N. Hopkins. 1990. Alignment of U3 region sequences of mammalian type C viruses: identification of highly conserved motifs and implications for enhancer design. *J. Virol.* **64**:534–542.
23. Hoggan, M. D., R. R. O'Neill, and C. A. Kozak. 1986. Nonectropic murine leukemia viruses in BALB/c and NFS/N mice: characterization of the BALB/c *Bxv-1* provirus and the single NFS endogenous xenotrope. *J. Virol.* **60**:980–986.
24. Holland, C. A., C. Y. Thomas, S. K. Chattopadhyay, C. Koehne, and P. V. O'Donnell. 1989. Influence of enhancer sequences on thymotropism and leukemogenicity of mink cell focus-forming viruses. *J. Virol.* **63**:1284–1292.
25. Inaguma, Y., N. Miyashita, K. Moriwaki, W. C. Huai, J. Mei-Lei, H. Zinqiao, and H. Ikeda. 1991. Acquisition of two endogenous ecotropic murine leukemia viruses in distinct Asian wild mouse populations. *J. Virol.* **65**:1796–1802.
26. Jenkins, N. A., N. G. Copeland, B. A. Taylor, and B. K. Lee. 1982. Organization, distribution, and stability of endogenous ecotropic murine leukemia virus DNA sequences in chromosomes of *Mus musculus*. *J. Virol.* **43**:26–36.
27. Khan, A. S. 1984. Nucleotide sequence analysis establishes the role of endogenous murine leukemia virus DNA segments in formation of recombinant mink cell focus-forming murine leukemia viruses. *J. Virol.* **50**:864–871.
28. Khan, A. S., and M. A. Martin. 1983. Endogenous murine leukemia proviral long terminal repeats contain a unique 190-base-pair insert. *Proc. Natl. Acad. Sci. USA* **80**:2699–2703.
29. Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111–120.
30. Kozak, C. A., and R. R. O'Neill. 1987. Diverse wild mouse origins of xenotropic, mink-cell focus-forming, and two types of ecotropic proviral genes. *J. Virol.* **61**:3082–3088.
31. Kubo, Y., Y. Nakagawa, K. Kakimi, H. Matsui, K. Higo, L. Wang, H. Kobayashi, T. Hirama, and A. Ishimoto. 1994. Molecular cloning and characterization of a murine AIDS virus-related endogenous transcript expressed in C57BL/6 mice. *J. Gen. Virol.* **78**:881–888.
32. Lamont, C., P. Culp, R. L. Talbott, T. R. Phillips, R. J. Trauger, W. N. Frankel, M. C. Wilson, J. M. Coffin, and J. H. Elder. 1991. Characterization of endogenous and recombinant proviral elements of a highly tumorigenic AKR cell line. *J. Virol.* **65**:4619–4628.
33. Lenz, J., D. Celander, R. L. Crowther, R. Patarca, D. W. Perkins, and W. A. Haseltine. 1984. Determination of the leukemogenicity of a murine retrovirus by sequences within the long terminal repeat. *Nature* **308**:467–470.
34. Lieber, M. M., C. J. Sherr, G. J. Todaro, R. E. Benveniste, R. Callahan, and H. G. Coon. 1975. Isolation from the Asian mouse *Mus caroli* of an endogenous type C virus related to infectious primate type C viruses. *Proc. Natl. Acad. Sci. USA* **72**:2315–2319.
35. Massey, A. C., M. A. Coppola, and C. Y. Thomas. 1990. Origin of pathogenic determinants of recombinant murine leukemia viruses: analysis of *Bxv-1*-related xenotropic viruses from CWD mice. *J. Virol.* **64**:5491–5499.
36. Massey, A. C., S. S. Lawrenz, D. J. Innes, and C. Y. Thomas. 1994. Origins of enhancer sequences of recombinant murine leukemia viruses from spontaneous B- and T-cell lymphomas of CWD mice. *J. Virol.* **68**:3773–3783.
37. Meruelo, D., A. Rossomando, M. Offer, J. Buxbaum, and A. Pellicer. 1983. Association of endogenous viral loci with genes encoding murine histocompatibility and lymphocyte differentiation antigens. *Proc. Natl. Acad. Sci. USA* **80**:5032–5036.
38. Morrison, H. L., B. Soni, and J. Lenz. 1995. Long terminal repeat enhancer core sequences in proviruses adjacent to *c-myc* in T-cell lymphomas induced by a murine retrovirus. *J. Virol.* **69**:446–455.
39. Needleman, S. B., and C. D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**:443–453.
40. O'Brien, S. J., J. L. Moore, M. A. Martin, and J. E. Womack. 1982. Evidence for the horizontal acquisition of murine AKR virogenes by recent horizontal infection of the germ line. *J. Exp. Med.* **155**:1120–1123.
41. O'Neill, R. R., C. E. Buckler, T. S. Theodore, M. A. Martin, and R. Repaske. 1985. Envelope and long terminal repeat sequences of a cloned infectious NZB xenotropic murine leukemia virus. *J. Virol.* **53**:100–106.
42. Ott, D., R. Friedrich, and A. Rein. 1990. Sequence analysis of amphotropic and 10A1 murine leukemia viruses: close relationship to mink cell focus-inducing viruses. *J. Virol.* **64**:757–766.
43. Perryman, S., J. Nishio, and B. Chesebro. 1991. Complete nucleotide sequence of Friend murine leukemia virus, strain FB29. *Nucleic Acids Res.* **19**:6950.
44. Rosen, C. A., W. A. Haseltine, J. Lenz, R. Ruprecht, and M. W. Cloyd. 1985. Tissue selectivity of murine leukemia virus infection is determined by long terminal repeat sequences. *J. Virol.* **55**:862–866.
45. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
46. Shigemoto, K., S. Kubo, Y. Itoh, G. Tate, S. Handa, and N. Maruyama. 1992. Expression and structure of serum gp70 as an acute phase protein in NZB mice. *Mol. Immunol.* **29**:573–582.
47. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukaemia virus. *Nature* **293**:543–548.
48. Stewart, M. A., M. Warnock, A. Wheeler, N. Wilkie, J. I. Mullins, D. E. Onions, and J. C. Neil. 1986. Nucleotide sequences of a feline leukemia virus subgroup A envelope gene and long terminal repeat and evidence for the recombinational origin of subgroup B viruses. *J. Virol.* **58**:825–834.
49. Stoye, J. P., and J. M. Coffin. 1987. The four classes of endogenous murine leukemia virus: structural relationships and potential for recombination. *J. Virol.* **61**:2659–2669.
50. Stoye, J. P., and J. M. Coffin. 1988. Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes. *J. Virol.* **62**:168–175.
51. Stoye, J. P., W. N. Frankel, and J. M. Coffin. 1991. DNA hybridization in dried gels with fragmented probes: an improvement over blotting techniques. *Technique* **3**:123–128.
52. Stoye, J. P., N. Kaushik, S. Jeremiah, and S. Best. 1995. Genetic map of the region surrounding the retrovirus restriction locus, *Fv1*, on mouse chromosome 4. *Mamm. Genome* **6**:31–36.
53. Stoye, J. P., C. Moroni, and J. Coffin. 1991. Virological events leading to spontaneous AKR thymomas. *J. Virol.* **65**:1273–1285.
54. Taylor, B. A. 1978. Recombinant inbred strains: use in gene mapping, p. 423–438. *In* H. C. Morse III (ed.), *Origins of inbred mice*. Academic Press, New York, N.Y.
55. Tomonaga, K., and J. M. Coffin. 1998. Structure and distribution of endogenous nonectropic murine leukemia viruses in wild mice. *J. Virol.* **72**:8289–8300.
56. Trainor, C. D., M. L. Scott, S. F. Josephs, K. E. Fry, and M. J. Reitz. 1984. Nucleotide sequence of the large terminal repeat of two different strains of gibbon ape leukemia virus. *Virology* **137**:201–205.
57. Tupper, J. C., H. Chen, E. F. Hays, G. C. Bristol, and F. K. Yoshimura. 1992. Contributions to transcriptional activity and to viral leukemogenicity made by sequences within and downstream of the MCF13 murine leukemia virus enhancer. *J. Virol.* **66**:7080–7088.
58. Van Beveren, C., E. Rands, S. K. Chattopadhyay, D. R. Lowy, and I. M. Verma. 1982. Long terminal repeat of murine retroviral DNAs: sequence analysis, host-proviral junctions, and preintegration site. *J. Virol.* **41**:542–556.
59. Voytek, P., and C. Kozak. 1988. HoMuLV: a novel pathogenic ecotropic virus isolated from the European mouse, *Mus hortulanus*. *Virology* **165**:469–475.
60. Voytek, P., and C. A. Kozak. 1989. Nucleotide sequence and mode of transmission of the wild mouse ecotropic virus, HoMuLV. *Virology* **173**:58–67.
61. Winkes, B. M., M. M. Ollmann, and G. S. Barsh. 1994. Association of Xmv-10 and the non-agouti (a) mutation explained by close linkage instead of causality. *Mamm. Genome* **5**:3–10.
62. Wolgamot, G., L. Bonham, and A. D. Miller. 1998. Sequence analysis of *Mus dunni* endogenous virus reveals a hybrid VL30/gibbon ape leukemia virus-like structure and a distinct envelope. *J. Virol.* **72**:7459–7466.
63. Yonekawa, H., O. Gotoh, Y. Tagashira, Y. Matsushima, L. I. Shi, W. S. Cho, N. Miyashita, and K. Moriwaki. 1986. A hybrid origin of Japanese mice "Mus musculus molossinus". *Curr. Top. Microbiol. Immunol.* **127**:62–67.
64. Yoshimura, F. K., B. Davison, and K. Chaffin. 1985. Murine leukemia virus long terminal repeat sequences can enhance gene activity in a cell-type-specific manner. *Mol. Cell. Biol.* **5**:2832–2835.