JOURNAL OF VIROLOGY, Sept. 1989, p. 4043–4046
0022-538X/89/094043-04$02.00/0

Vol. 63, No. 9

An Amplified Endogenous Retroviral Sequence on the Murine Y Chromosome Related to Murine Leukemia Viruses and Viruslike 30S Sequences†

KEITH W. HUTCHISON[1]* AND EVA M. EICHER[2]

Department of Biochemistry, University of Maine, Orono, Maine 04469-0131,[1] and The Jackson Laboratory, Bar Harbor, Maine 04609[2]

Received 8 February 1989/Accepted 19 May 1989

A highly repeated sequence on the murine Y chromosome was cloned and characterized. The DNA sequence of the viral long terminal repeats (LTRs) showed that the 5′ and 3′ LTRs were ~90% homologous. The LTRs are generally unrelated to any previously reported viral LTR but are somewhat similar to the viruslike 30S sequences.

Mice carry in their genomes large complements of retroviral and retrovirus-related sequences that are inherited in a Mendelian fashion. These include but are not limited to the intracisternal A particles (1,000 to 2,000 copies per genome) (12), the mouse mammary tumor viruses (~20 copies per genome) (1, 3, 13), the murine leukemia viruses (MuLV; 30 to 50 copies per genome) (6, 20), and the viruslike 30S (VL30S) sequences (100 to 200 copies per genome) (11). New families of retrovirus-related sequences are continually being discovered (10, 18).

We previously identified in inbred strains of laboratory mice a retrovirus-related sequence (16) that is part of a large repeat unit present on the Y chromosome in approximately 500 copies (1a). Hybridization data from different species of *Mus* suggest that integration of the retrovirus into the Y chromosome occurred prior to the divergence of a closely related group in the subgenus *Mus*, including *Mus spretus*, *M. abbotti*, *M. hortulanus*, *M. musculus*, and *M. domesticus* (1a). Subsequent amplification of the virus and its flanking cellular sequences occurred in the Y chromosome of the male progenitor to *M. musculus* and *M. domesticus* (1a, 16).

This paper reports the further analysis of this highly repeated retrovirus-related sequence, designated MuRVY (murine repeated virus on the Y chromosome). We use the term MuRVY to refer specifically to those highly repeated proviral sequences on the Y chromosome resulting from a single integration event. Although autosomal sequences can be detected by hybridization of MuRVY probes to the DNA of female mice (1a), we do not know the structure of these autosomal sequences, and they are referred to as MuRVY-related sequences.

For this study, MuRVY sequences were subcloned from two genomic clones, λY6 and λY45. The two genomic clones were identified by probing a C57BL/10 genomic library with the MuLV M720. Their isolation and the characterization of the genomic clones are described elsewhere (1a). Clone λY6 contained a complete copy of a MuRVY sequence, whereas clone λY45 contained only the 3′ end of the provirus. A restriction enzyme map constructed for the λY6 copy of MuRVY is shown in Fig. 1. The total length of the provirus was 8.8 kilobases, with the 5′-to-3′ orientation based on the analysis of DNA sequence data.

The 5′ and 3′ long terminal repeats (LTRs) of λY6 were subcloned into M13 and sequenced by dideoxy chain termination as previously described (7). For portions of the LTR, the sequence was obtained from a nested series of BAL 31 deletions under conditions described by the supplier (IBI). The 5′ and 3′ LTRs were 627 and 628 base pairs (bp) long, respectively (Fig. 2). There were 49 base changes between the two LTRs, with 11 bases either inserted or deleted. Most of these changes occurred in the regions of the LTR that would approximate the R and U5 regions (Fig. 2). Overall, the LTRs were approximately 90% homologous. Potential TATA and CCAAT boxes were identified at positions 418 and 372, respectively (Fig. 2). A potential poly(A) addition signal with the sequence ATTAAAG was identified at position 537. The MuRVY LTRs ended with a 4-bp inverted repeat, TGAA----TTCA (Fig. 3A). The inverted repeats, primer-binding sites, and polypurine tract of the MuRVY LTRs were similar to those of other murine LTRs, particularly those of the VL30S sequences (Fig. 3B). The primer-binding sequence of MuRVY resembled the $tRNA^{Gly}$ primer-binding sites of VL30S sequences (Fig. 3B) (2, 14). There were three positions at which the MuRVY tRNA binding-site sequence differed from those of the VL30S sequences (Fig. 3B). We do not know whether this difference represents evolutionary changes in MuRVY or an uncharacterized tRNA species. Finally, there was a 4-bp direct repeat of cellular DNA (GTAG) immediately flanking the MuRVY LTRs (Fig. 2).

Because MuRVY is present in at least 500 copies on the Y chromosome of laboratory mice (1a), we questioned whether the DNA sequences from λY6 were representative of the total population of MuRVY. This question was addressed in two ways. First, a partial sequence was obtained for the 3′ LTR of clone λY45 (Fig. 2). The two 3′ LTRs differed from each other in 7 of 331 bp. This is about one-fifth the degree of dissimilarity found between the 5′ and 3′ LTRs of clone λY6. Because there are differences between the 3′ LTRs of clones λY6 and λY45, we conclude that the two clones represent different copies of MuRVY. However, the overall homology (98%) of the two 3′ LTRs supports the interpretation that the differences between the 5′ and 3′ LTRs of clone λY6 are representative of most of the MuRVY population.

* Corresponding author.
† Maine Agricultural Experiment Station publication no. 1375.

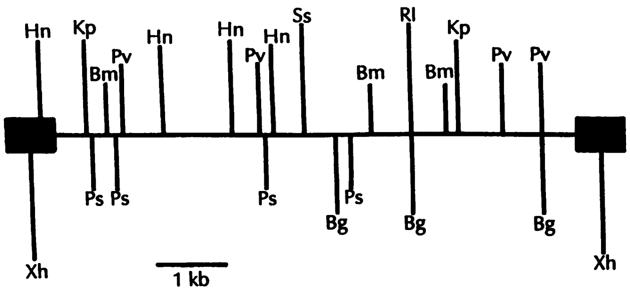

FIG. 1. Restriction enzyme map of the λY6 copy of MuRVY. MuRVY is drawn 5′ to 3′ with respect to the potential coding direction of the viral genome. Restriction enzyme site abbreviations: Bm, *Bam*HI; Bg, *Bgl*II; Hn, *Hin*dIII; Kp, *Kpn*I; Ps, *Pst*I; Pv, *Pvu*II; R, *Eco*RI; Ss, *Sst*I; Xh, *Xho*I. There were no sites for the enzymes *Xba*I, *Sal*I, and *Sma*I.

Second, the 5′ LTR of λY6 contains a *Hin*dIII site which was not present in the 3′ LTR of either λY6 or λY45 (Fig. 4B). To determine whether the divergence between the 5′ and 3′ LTR sequences was present in additional copies of MuRVY, we hybridized an LTR probe to genomic blots of *Bgl*II-digested and *Bgl*II-*Hin*dIII double-digested DNA from C57BL/6J male mice (Fig. 4). As shown in Fig. 4A, the major proportion of hybridization was to the fragments predicted from the results shown in Fig. 4B. A few additional bands were seen in the *Bgl*II and *Bgl*II-*Hin*dIII double digest, suggesting that some variation does occur in the MuRVY population. Nevertheless, these data generally suggest that most of the copies of MuRVY have the *Hin*dIII site in the 5′ LTR but not in the 3′ LTR. This further supports the position that the sequence data for the 5′ and 3′ LTRs of λY6 are representative of the predominant fraction of copies of MuRVY.

A global search of the GenBank DNA Sequence Database (Release 55.0) to identify sequences related to the MuRVY LTRs was made by using the Lipman-Pearson DNA Search and Alignment facility of the IBI/Pustell DNA Sequence Analysis programs. The degree of similarity of the MuRVY LTR sequence to other sequences was low. The greatest similarity detected was to the sequences for the VL30S LTRs and to an endogenous retrovirus associated with the mouse major histocompatibility complex *TL* locus (15). The alignment of the MuRVY 5′ LTR to the VL30S LTRs was strongest in the regions flanking the putative TATA box (Fig. 2). For clarity, only a portion of the VL30S sequence is shown. Other regions showed little relationship between the MuRVY LTR sequence and the VL30S LTR sequence, although a two-dimensional matrix comparison at low stringencies showed some similarity in the 3′ ends of the LTRs (K. W. Hutchison, unpublished data). Similar results were obtained when the MuRVY LTR sequence was compared with other published VL30S sequences (6, 8, 9, 14, 17). The

```
                             10        20        30        40        50        60        70        80        90       100       110       120       130
MuRVY 5' LTR   GTAG TGAAGAGCAAAAACTTATAATTTTGGGTCTTATAGCATTGTGGTATGTAAGTTCAAGAAATACAGGGCTGAAGAATAGACCAAGGCGGGTCCAGGCACAGCCAAACCCACCCTGTAGCTAAGAATGTCCA
MuRVY 3' LTR        ....C........................................................T...................C.....................T..T......................G

         140       150       160       170       180       190       200       210       220       230       240       250       260       270       280
GCCATAAAAGATAAAAAAGAGAATGCAGGAACTAGCCACTGCTATCAGGGCAGACCAAATGGCCATCTGGCAGGATGGCACCTCCCTATAGATCCCCATAAATCATACATTGACCAGGTGTCCTGCTGACCTCGAGAAGTCCCAAGCCCT
.............T...............................A......G........................................-........................................................
                                                                                         MuRVY 3' LTR*                              ...................

         290       300       310       320       330       340       350       360       370       380       390       400       410       420       430
TGTGCCTCCCTACACCTTAGACATGCACATATTACCCTGTTAATTTGTAATCATCCCGCCA-CATTTAAATAAGCCTATCATGTGTAATGGCCCAAATTCTTTTGGCTCCCCTGTTTTTGGTTTGTG-GTTTTTGCCTTTATATGCCTGT-
....T..TT..C..TT....-..G.T.............C................T...GC..............A..T............G.................................-...............A......-
....T..TT..C..TT....-..G.T.............C................T...GC..............A..T............A.................................-...............A......-
                                                                              VL30 LTR   .AC...CC..G...C.C.C.CCCCCCT..A......T...A....C......G
                                                                                              320       330       340       350       360

         440       450       460       470       480           490       500       510       520       530       540       550       560       570
AAAAAGTTTGGGCCGGGGTTGA-ACTTCCCTGCCCCCTGCGCAGTGCATGAGATTCGACCCC-AG-ATTGCTGGCTTGGGCGCTCCTTTTTTGTTTGTGTGTTTGTGTGAGATTAAAGC-TTGTATAGTTACAGCGAGTTCC-GCCTCAGT
.............T........-...C...............G.........T....GT..TC..C...-........A...AA.....G................AC.CC.......C.......A......T......TG.T......
.............T........-...C.........T...............T....GT..TC..C...-...A....A...AA.....G................A..CC.......C.......A......T......T-.T......
.....T.....CT..A......TT..C.T..A.A..A.A.TAG.........T.........-..-.GCT....TC.AA.T.---....C...-C...-.AC..-.A.T.A..CTTGC.C.C-..C.-..---TT....
    370       380       390       400       410       420         430       440         450         460        470        480           490

    580       590       600       610       620
GTCCTACTGGGTGCGCACCT-CCCTGAAGAGGGACTCACTGAAGGGGTCTTTCA
................G.T.G...-...A........C.C-.C.......... GTAG
.............G...
```

FIG. 2. Alignment of MuRVY 5′ (clone λY6) and 3′ (clones λY6 and λY45) LTRs. The 5′ LTR sequence is given in the upper line. Only the differences in the 3′ LTR sequences are shown. MuRVY 3′ LTR*, Sequence for the λY45 clone; VL30 LTR, a portion of the sequence for a VL30S LTR (9). Base-pair position is given below the VL30S sequence. Sequences are hyphenated to maintain maximal alignment. Position numbers are given for the 5′ LTR. GTAG at the beginning of the 5′ LTR and at the end of the 3′ LTR represents direct repeats of the cellular DNA. Potential TATA and CCAAT boxes are underlined.

A

```
MuRVY    TGAAGAGCAAAAA ---------LTR--------- AGGGGTCTTTCA
VL30     TGAAGAATAGAAA ---------LTR--------- GAGGGTCTTTCA
AKR623   TGAAAGACCCC  ---------LTR---------  GGGGTCTTTCA
```

B

```
MuRVY    TTCA TGTGGTGCATTGGCTGGGAGA ---virus--- AAAGAAGTGGGTAA TGAA
VL30S    TTCA TTTGGTGCATTGGCCGGGAAA ---virus--- GAAGAAGTGGGGAA TGAA
```

FIG. 3. Comparison of the MuRVY LTR with other retroviral LTRs. (A) Alignment of the terminal inverted repeat of the MuRVY LTR with a VL30S sequence LTR (5) and an ecotropic MuLV LTR (4). Underlined bases show the actual inverted repeat. Vertical bars indicate identical bases in the MuRVY LTR and either the VL30S LTR or the ecotropic MuLV LTR. (B) Alignment of the primer-binding site and polypurine tract of MuRVY with the corresponding sequence from a VL30S sequence (9). Ends of the MuRVY proviral LTR are indicated by the solid box.

MuRVY LTRs were not significantly related to ecotropic or xenotropic MuLV LTRs, solo LTRs (22), Etn elements (18), or Gln elements (10). Neither was there any signficant homology to a recently reported retroviral provirus whose presence confers androgen-sensitive regulation to the mouse sex-limited protein gene (19).

The MuRVY family of proviruses was originally identified by their homology to the type C retrovirus M720 (16). They are highly repeated sequences representing a significant portion of the murine Y chromosome (1a, 16). We have used DNA sequencing of the putative LTRs to further support the supposition that MuRVY is a retrovirus or of retroviral origin. The MuRVY LTRs have the structural characteristics of other retroviral LTRs in that they terminate in an inverted repeat, they have a putative tRNA primer-binding site at the internal junction of the 5′ LTR, and they have a polypurine tract at the internal junction of the 3′ LTR (Fig. 3). The MuRVY provirus is also flanked by a 4-bp direct repeat of cellular DNA (Fig. 2), suggesting that MuRVY integrated into the Y chromosome by a normal retroviral mechanism, generating a short duplication of the cellular DNA. Although a potential promoter sequence has been assigned within the MuRVY LTR (Fig. 2), we do not know whether it is functional. A search of the GenBank database showed that there is little similarity between the MuRVY LTR and other mammalian retroviral LTRs. The LTRs showing the greatest similarity to the MuRVY LTR are those of the VL30S sequences. This similarity is not restricted to the LTRs but is also seen in the use of $tRNA^{Gly}$ as the primer for minus-strand DNA synthesis (Fig. 3B).

An approximate measure of when MuRVY integration and amplification occurred can be made from the differences in DNA sequences between the 5′ and 3′ LTRs of λY6 and between the 3′ LTRs of λY6 and λY45. After integration, the 3′ and 5′ LTRs become isolated genetic units, free to evolve independently. After amplification, individual copies of the proviral sequences are similarly free to evolve. Thus, differ-

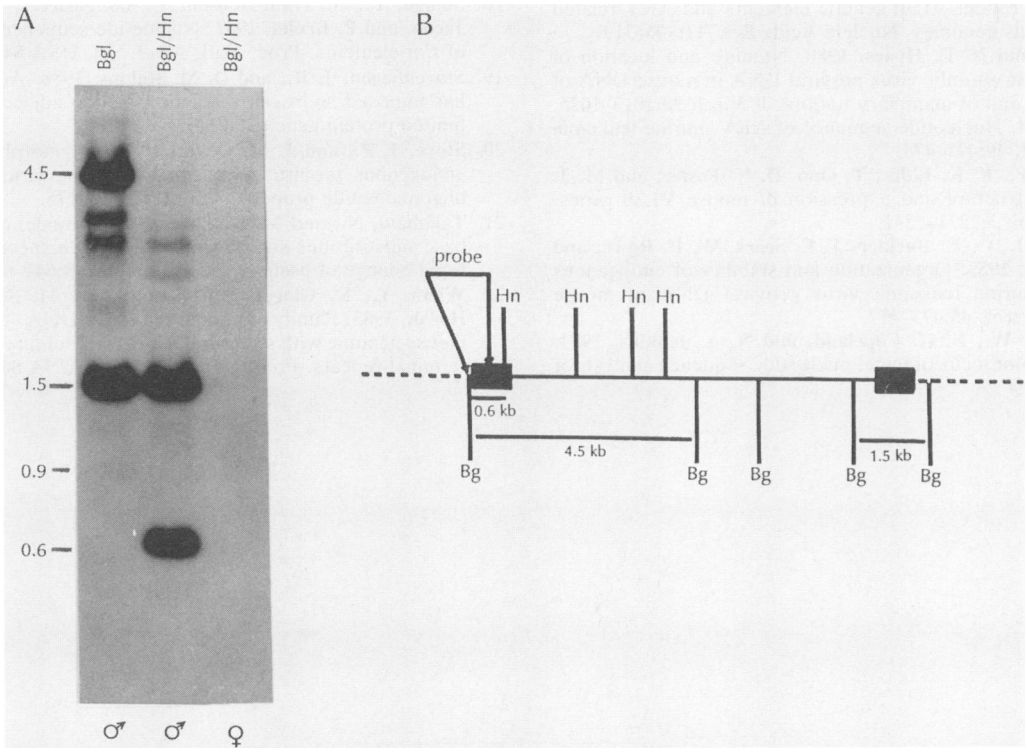

FIG. 4. Probing of DNA from a C57BL/10J male mouse with a MuRVY LTR probe to show that the copy of MuRVY contained in clone λY6 is representative of most copies of MuRVY. (A) DNA from a C57BL/10J male mouse (10 μg) was digested with *Bgl*II or *Bgl*II and *Hin*dIII, electrophoresed, blotted, and hybridized with a MuRVY LTR probe. DNA from a female C57BL/10J mouse was also digested with *Bgl*II and *Hin*dIII. The sizes of the major fragments are given in kilobases, as is the location of the potential band at 900 bp. (B) Schematic representation of the restriction enzyme sites and location of the probe for the panel A experiment. Predicted fragments are indicated. kb, Kilobases; Hn, *Hin*dIII; Bg, *Bgl*II.

ences between the LTR sequences are a measure of the evolutionary time involved. The sequence divergence between the 5′ and 3′ LTRs and between the two 3′ LTRs suggests that the virus integrated into the murine Y chromosome 5 to 10 million years ago, with a subsequent amplification event 1 to 2 million years ago (based on the data of Takahata and Kimura [21]). Although these values are clearly approximations and do not take into consideration such factors as corrections for dissimilarity due to gene conversion or unequal rates of evolution, the sequence analysis suggests that the integration and subsequent amplification of the provirus occurred at significantly different times and, therefore, that the two events were independent.

We acknowledge the expert technical assistance of Barbara K. Lee, Marie Ivey, and Yolanda Nieuwkerk. We are indebted to Edward H. Birkenmeier, Hendrick G. Bedigian, Neal G. Copeland, and Nancy A. Jenkins for their critical comments during the preparation of the manuscript.

This research was supported by Public Health Service grant GM20919 from the National Institutes of Health and by funds from the University of Maine and the Maine Agricultural Experiment Station.

LITERATURE CITED

1. **Cohen, J. C., and H. E. Varmus.** 1979. Endogenous mammary tumor virus DNA varies among wild mice and segregates during inbreeding. Nature (London) **278**:418–423.
1a. **Eicher, E. M., K. W. Hutchison, S. J. Phillips, P. K. Tucker, and B. K. Lee.** 1989. A repeated segment on the mouse Y chromosome composed of retroviral-related, Y-enriched and Y-specific sequences. Genetics **122**:181–192.
2. **Giri, G. P., C. P. Hodgson, P. K. Elder, M. G. Courtney, and M. J. Getz.** 1983. Discrete regions of sequence homology between cloned rodent VL30 genetic elements and AKV-related MuLV provirus genomes. Nucleic Acids Res. **11**:306–319.
3. **Groner, B., and N. E. Hynes.** 1980. Number and location of mouse mammary tumor virus proviral DNA in mouse DNA of normal tissue and of mammary tumors. J. Virol. **33**:1013–1025.
4. **Herr, W.** 1984. Nucleotide sequence of AKV murine leukemia virus. J. Virol. **49**:471–478.
5. **Hodgson, C. P., P. K. Elder, T. Ono, D. N. Foster, and M. J. Getz.** 1983. Structure and expression of mouse VL30 genes. Mol. Cell. Biol. **3**:2221–2231.
6. **Hoggan, M. D., C. E. Buckler, J. F. Sears, W. P. Rowe, and M. A. Martin.** 1983. Organization and stability of endogenous xenotropic murine leukemia virus proviral DNA in mouse genomes. J. Virol. **45**:473–477.
7. **Hutchison, K. W., N. G. Copeland, and N. A. Jenkins.** 1984. Dilute-coat-color locus of mice: nucleotide sequence analysis of the d^{+2J} and d^{+Ha} revertant alleles. Mol. Cell. Biol. **4**:2899–2904.
8. **Itin, A., and E. Keshet.** 1983. Nucleotide sequence analysis of the long terminal repeat of murine viruslike DNA (VL30) and its adjacent sequences: resemblance to retrovirus proviruses. J. Virol. **47**:656–659.
9. **Itin, A., and E. Keshet.** 1986. Diverse long terminal repeats are associated with murine retroviruslike (VL30) elements. Mol. Cell. Biol. **6**:1276–1282.
10. **Itin, A., and E. Keshet.** 1986. A novel retroviruslike family in mouse DNA. J. Virol. **59**:301–307.
11. **Keshet, E., and A. Itin.** 1982. Patterns of genomic distribution and sequence heterogeneity of a murine "retrovirus-like" multigene family. J. Virol. **43**:50–58.
12. **Lueders, K. K., and E. L. Kuff.** 1977. Sequences associated with intracisternal A particles are reiterated in the mouse genome. Cell **12**:963–972.
13. **MacInnes, J. I., V. L. Morris, W. F. Flintoff, and C. A. Kozak.** 1984. Characterization and chromosomal location of endogenous mouse mammary tumor virus loci in GR, NFS, and DBA mice. Virology **132**:12–25.
14. **Norton, J. D., J. Connor, and R. J. Avery.** 1984. Unusual long terminal repeat sequence of a retrovirus transmissible mouse (VL30) genetic element: identification of functional domains. Nucleic Acids Res. **12**:3445–3460.
15. **Pampeno, C. L., and D. Meruelo.** 1986. Isolation of a retroviruslike sequence from the *TL* locus of the C57BL/10 murine major histocompatibility complex. J. Virol. **58**:296–306.
16. **Phillips, S. J., E. H. Birkenmeier, R. Callahan, and E. M. Eicher.** 1982. Male and female mouse DNAs can be discriminated using retroviral probes. Nature (London) **297**:241–243.
17. **Rotman, G., A. Itin, and E. Keshet.** 1984. 'Solo' large terminal repeats (LTR) of an endogenous retrovirus-like gene family (VL30) in the mouse genome. Nucleic Acids Res. **12**:2273–2282.
18. **Sonigo, P., S. Wain-Hobson, L. Bougueleret, P. Tiollais, F. Jacob, and P. Brulet.** 1987. Nucleotide sequence and evolution of *Etn* elements. Proc. Natl. Acad. Sci. USA **84**:3768–3771.
19. **Stavenhagen, J. B., and D. M. Robins.** 1988. Ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. Cell **55**:247–254.
20. **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.
21. **Takahata, N., and M. Kimura.** 1981. A model of evolutionary base substitutions and its applications with special reference to rapid change of pseudogenes. Genetics **98**:641–657.
22. **Wirth, T., K. Gloggler, T. Baumruker, M. Schmidt, and I. Horak.** 1983. Family of middle repetitive DNA sequences in the mouse genome with structural features of solitary retroviral long terminal repeats. Proc. Natl. Acad. Sci. USA **80**:3327–3330.