Human and other mammalian genomes contain transposons of the *mariner* family

Corinne Auge-Gouillou[a], Yves Bigot[a], Nicolas Pollet[b], Marie-Helene Hamelin[a], Michele Meunier-Rotival[b], Georges Periquet[a,*]

[a]*Institut de Recherche sur la Biologie de l'Insecte, Faculté des Sciences, Parc Grandmont, 37200 Tours, France*
[b]*INSERM U347 affiliée au CNRS, 80, Rue du Général Leclerc, 94276 Le Kremlin-Bicêtre Cedex, France*

Received 31 May 1995

Abstract Internal fragments of the putative transposase gene of *mariner*-like elements (MLEs) were amplified from human, mouse, rat, chinese hamster, sheep and bovine genomic DNAs by polymerase chain reaction (PCR). The sequences identified in human, ovine and bovine genomes correspond to ancient degenerate transposons. Screening mammalian sequence libraries identified a truncated element in the human ABL gene and the sequence of its 5′-ITR was determined. This ITR sequences were used in PCR experiments with DNA from six mammalian species and detected full-sized and deleted MLEs. The presence of MLE in mammalian genomes demonstrates that they are ubiquitous mobile elements found from fungi to man. This observation strongly raises the possibility that MLE could constitute tools for the modification of eucaryotic genomes.

Key words: Evolution; Mammalia; Interspersed repeated DNA; Transposase

1. Introduction

In eukaryotic genomes, transposable elements moving via a DNA intermediate are called transposons [1–7]. Many of those described have been studied in invertebrate and plant species [8]. They possess inverted terminal repeats (ITRs of 8–40 bp) and encode a tranposase. Two sub-groups of transposons have been defined on the basis of similarities between the sequences encoding transposases. The HAT sub-group comprises *hobo*, *Ac* and Tam3 elements. The second sub-group comprises shorter transposons (1200–1500 bp) and includes several bacterial IS, Tc1, Tec1 and 2, *Bari*, TBE1, and *mariner*-like elements (MLE), which have a highly conserved 'D,D(35)E' motif [9] in their transposases. MLE are a family of about 1250 bp transposons with 20 to 40 bp ITRs encoding transposases which contain two highly conserved amino acid motifs (WVPHEL and YSPDLAP) separated by about 150 amino acids. MLEs were detected in about 20% of 400 insect species tested for these two motifs by polymerase chain reaction (PCR) using degenerate primers [1]. We previously used PCR, sequencing and the stepping-stone hybridization method to evidence MLEs in all the 40 hymenopteran species tested ([2] and personnal data). MLEs have also been detected in Nematodes [3], Crustaceae [4] and Fungi [5]. A *mariner*-like sequence has been obtained from *Fusarium oxysporum* [5] and its characteristics indicate that *mariner* and Tc1 transposon families are closely related. Recently, Tc1-like elements have been described in a primitive fish, the hagfish *Eptatretus stouti* [6], and in several species of salmonid fishes [7]. Tc1 and *mariner* elements may correspond to a 'transposon superfamily' of closely related elements, ubiquitous among Fungi, Invertebrates and Vertebrates. It has not been previously shown whether or not such transposons are present in mammals including man. Were they indeed present, they would be potentially valuable tools in transgenic studies [10]? In this paper, we report the presence of MLEs in the genome of human (*Homo sapiens*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), chinese hamster (*Mesocricetus auratus*), sheep (*Ovis aries*) and cattle (*Bos taurus*).

*Corresponding author. Fax: (33) 47 36 69 66.

Abbreviations: MLE, *mariner*-like element; ITR, inverted terminal repeats.

2. Materials and methods

2.1. DNA manipulations

All DNA purifications, syntheses of labelled DNA probes and agarose gel electrophoresis analyses were carried out as described [11,12]. Single and sandwich Southern blots were performed and hybridized with probes at the same specific activity, in 0.1% SDS, 0.5 M Na_2HPO_4/NaH_2PO_4 buffer, pH 7 at 65°C. Stringency of washing conditions at 65°C was either high in 0.5 × SSC (1 × SSC = 0.15 M NaCl/0.015 M trisodium citrate) or low in 2 × SSC. Exposure times are defined as short (4 h) or long (24 h).

2.2. DNA amplifications

Amplifications of internal fragment of MLE were performed on 100 ng of genomic DNA from each species using primers [1] MAR124F 5′-TGGGTNCCNAYGARYT-3′ and MAR276R 5′-GGNGCNARRTCNGGNSWRT-3′ (Y = C or T/R = A or G/S = C or G/W = A or T/N = A, C, G or T], in 100 μl of 10 mM Tris-HCl, pH 9/1.5 mM $MgCl_2$/50 mM KCl/0.1% Triton X-100/150 μM each dATP, dCTP, dGTP, and dTTP/0.8 mM of each oligonucleotide and 1 unit of Taq Polymerase (Appligene). A programmable thermal controller (Perkin-Elmer) was used for 40 cycles, 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min. At the end of the 40th cycle, an extension was allowed to proceed at 72°C for 5 min. The other amplification experiments were similarely performed, except that 0.1 μM rather than 0.8 μM of each oligonucleotide (P1: 5′-TTGATATTCGAGTACATTC-3′, P2: 3′-CTTTAAATAAAGTGTTATGT-5′, P3: 3′-CGAGTACATTCTTTAAATAA-5′) was used.

2.3. Cloning and sequencing

PCR products were separated on 1% low melting point agarose (Sigma type IX) gels, purified by melting at 55°C, and cloned into M13mp18 vector, dT-tailed at *Sma*I site of the cloning multisite. Cloning and phage preparations were performed by standard methods [11]. The inserts were sequenced on one strand with Sequenase v2.0 (USB), using denaturing gels either of polyacrylamide in buffer gradients or Hydrolink long ranger gels (AT-Biochem) as described [2]. Alignments of the mammalian MLEs sequences are performed from published data [1] and using CITI2 programs [13].

SSDI 0014-5793(95)00735-0

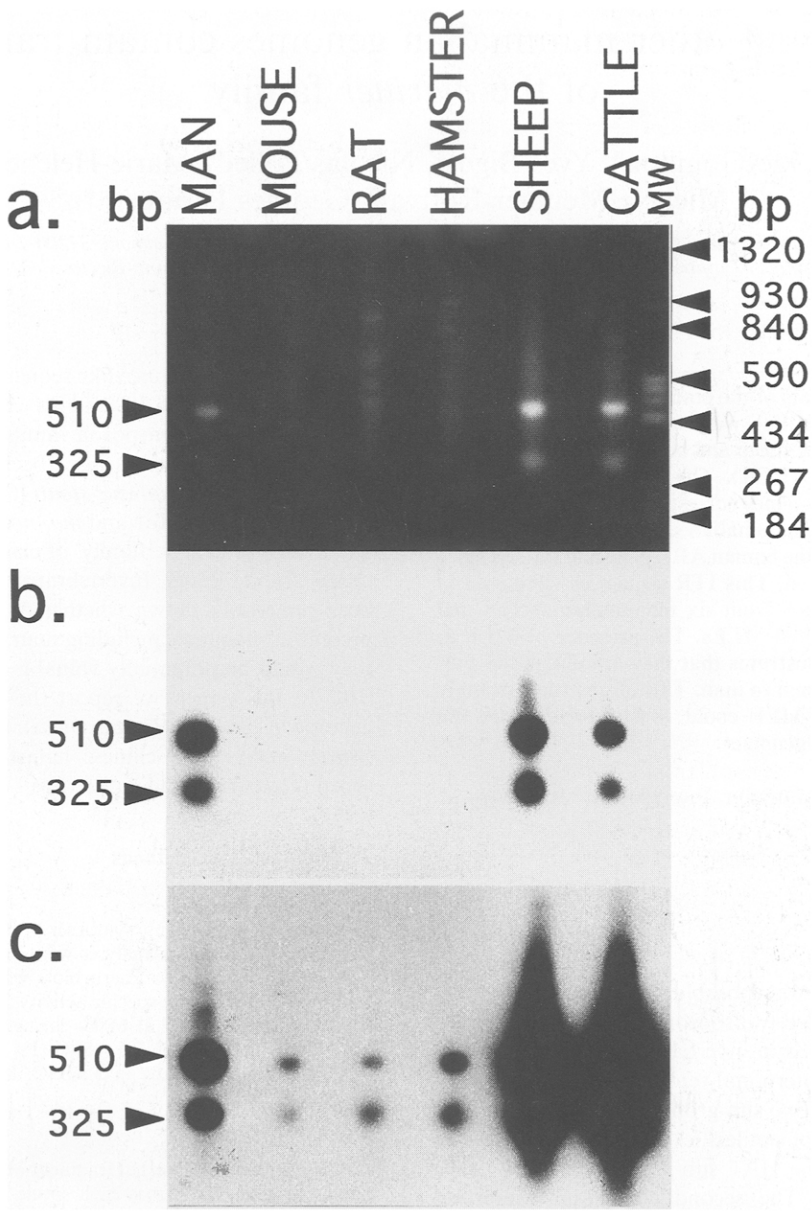


Fig. 1. Amplifications of part of the MLE transposase from human, mouse, rat, hamster, sheep and cattle (lanes 1 to 6 respectively) genomic DNA; (lane 7) DNA size markers indicated in bp on the right. (A) Separation of 10 ml PCR products on 1.5% agarose gel stained wih ethidium bromide (EtBr). The location and molecular weights of the two main PCR products are indicated on the left. (B) Autoradiogram of Southern blot of the gel in (A), hybridized with a mix of the five radiolabelled DNA fragments purified from human MLE clones (Hs2173, Hs2177, Hs2178, Hs2179 and H2217: see below) under high stringency conditions and short exposure. Similar results are obtained with probes from ovine or bovine MLE clones. (C) Autoradiogram of Southern blot of the gel in (A), hybridized with six radiolabelled DNA fragments of the bovine MLE clones (Bt2231, Bt2232,Bt2235, Bt2237, Bt2237 and Bt2239, see below) under low stringency conditions and prolonged exposure.

3. Results and discussion

A 450–500 bp internal fragment of MLE was first searched by PCR using degenerate primers designed from the two amino acid conserved regions (WVPHEL and YSPDLAP) of the known MLE transposases. Using genomic DNAs purified from lymphocytes by ultracentrifugation on ethidium bromide-cesium chloride gradient [11] or by chromatography on hydroxylapatite [12], an intense fragment of about 510 bp was obtained from human, ovine and bovine (Fig. 1A). The use of sperm, liver or kidney genomic DNAs gave similar results. PCR products of 510 bp were also obtained with the three rodent liver genomic DNAs but in smaller amounts and with other minor bands.

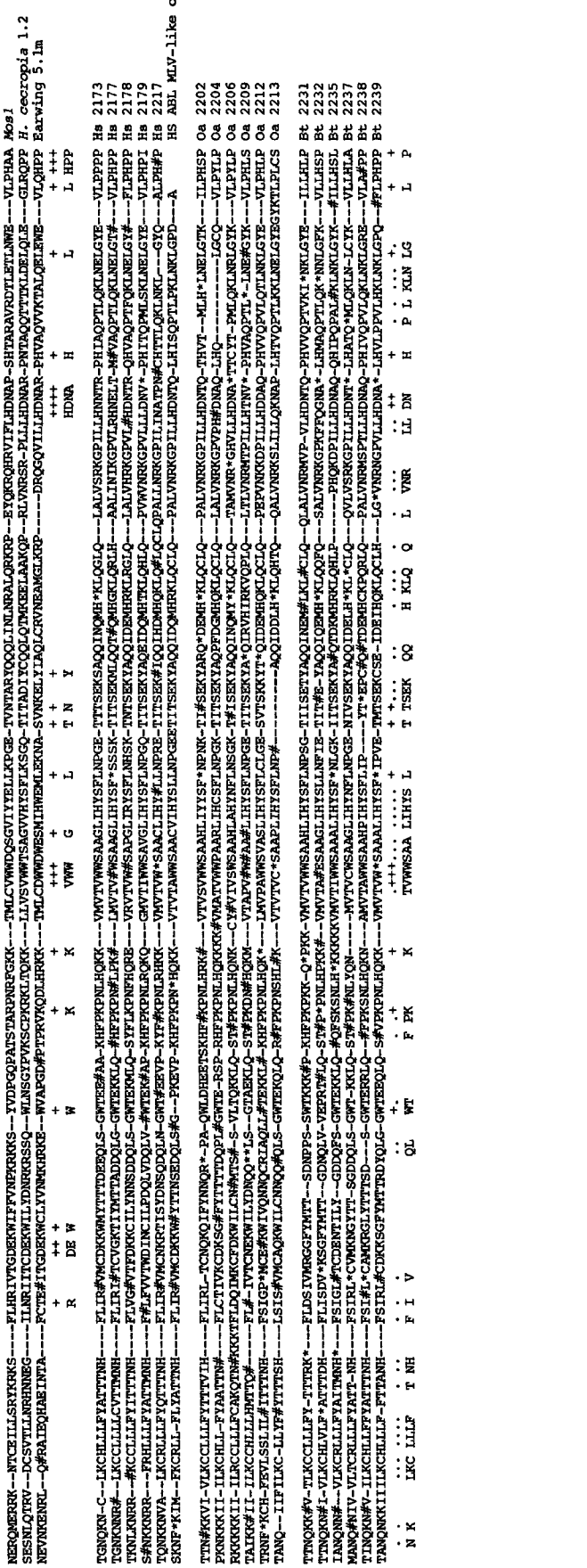
We verified that the human, ovine and bovine PCR 510 bp fragments correspond to part of a MLE tranposase-encoding region by cloning and sequencing them. At least five clones from each species were sequenced. The best match for each of these sequences in Genbank (release 85] was *Hyalophora cecropia*, *Ephestia cautella* (Lepidoptera) or *Forficula auriculata* (Dermaptera) MLEs sequences as assessed by FASTA or BLAST [13]. Human, ovine and bovine genomic DNAs were

digested and hybridized on Southern blot with MLEs probes of the same species (data not shown). We thus confirmed that these species contain many MLEs, with a very minimal estimate of one hundred elements per genome.

Another abundant PCR product observed in human, ovine and bovine DNAs was a fragment of about 325 bp (Fig. 1A). This fragment hybridized with the mammalian MLEs clones (Fig. 1B and C), indicating that it corresponds to a deleted 510 bp fragment. This suggests that these mammalian genomes contain different MLEs forms. None of the PCR products from rodents hybridized to human or sheep MLEs probes, even at low washing stringency and after long film exposure. However hybridization with the bovine MLEs probe and low washing stringency allow the detection of the 510 bp and 325 bp rodent fragments (Fig. 1C). Therefore, MLEs are also present in rodent genomes and are more similar to bovine than to human and ovine MLEs.

To align the putative amino-acid sequences of cloned fragments with those of insect MLE transposases [1], many small deletions and insertions resulting from frameshift and stop codons were required (Fig. 2). None of the sequences obtained showed an uninterrupted long open reading frame. Intraspecific and interspecific nucleotide similarities were similar (78% ± 5 and 74% ± 5) as were amino-acid sequences (54% ± 6 and 46% ± 6). Thus the MLEs found in human, ovine and bovine genomes have strongly diverged and many may be inactive elements. We were unable to identify a putative origin of these sequences because of their substantial sequence divergence. Consequently, in phyletical trees, it is unclear whether the cladograms and the nodes join groups of related sequences, or separate classes of randomly associated sequences of different genomic origins with high rates of sequence divergence. The similarities between *Drosophila mauritiana* and *H. cecropia* MLEs were 48% and 31% for DNA and amino-acid sequences respectively [14]. Between *H. cecropia* on one hand and human, ovine and bovine MLEs on the other, the average similarity were 63% ± 3 and 36% ± 4 and with *D. mauritiana*, they were 57% ± 4 and 28% ± 3. This indicates that these similarities are due to ancient horizontal transfer(s) of MLE between insect and mammalian species and that *H. cecropia* MLEs are the most similar to those in the human, ovine and bovine genomes.

On the other hand, screening mammalian sequences data in Genbank revealed that the intron 1b of the human ABL gene (Ac. No. U07562; chromosomal location 9q34; Fig. 3A) contains part of a MLE. It is truncated at the 3′ end and showed a translation start codon for a MLE transposase. Independent searches for conserved internal motifs of the *H. cecropia* MLE inverted terminal repeats in the region up to the start codon identified a putative ITR of this human MLE (Fig. 3B). We



→

Fig. 2. Alignment of MLE transposase deduced amino acid sequences – *Mos1* (*Drosophila mauritiana*), *H. cecropia* 1.2 and Earwing 5.1m (*Forficula auriculata*) (1) – five human clones (Hs2173 to H2217), the sequence of the first intron of the human ABL gene (ABL MLV-like oncogene), six ovine clones (Oa2202 to Oa2213) and six bovine clones (Bt2231 to Bt2239). The sequences have been deposited in the EMBL Nucleotide Sequence Database under Accession numbers X84285 to X84301). The symbol (#) indicates an introduced frameshift to maintain an aligned reading frame, and (*) indicates a stop codon in the aligned reading frame. Gaps (−) were introduced to maximize alignments.

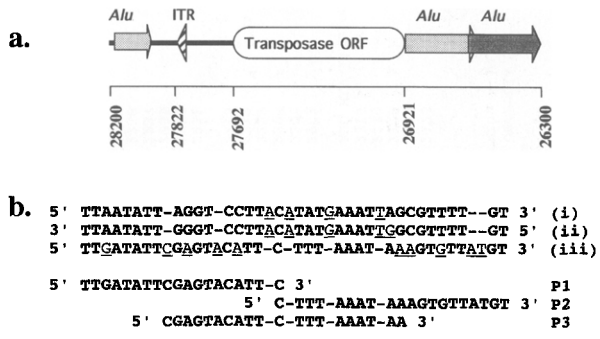


Fig. 3. (A) A schematic representation of the 3′-truncated MLE found in the first intron of human ABL gene (Genbank Ac. No. U07562) between nucleotide positions 26300 and 28200. The putative transposase ORF is on the complementary strand of the U07562 sequence which is 84.5% similar to Hs2173. Grey arrows point to the polydA tail of *Alu* repeats located in this region. The hatched arrow corresponds to the 5′-ITR of the MLE. (B) Sequence alignments of the 5′ (i) and 3′ (ii) ITRs of *H. cecrocopia* MLE with the 5′-ITR (iii) of the MLE found in the first intron of human ABL gene. P1, P2, P3: synthetic primers derived from the MLE ITR located in the human ABL gene.

designed three overlapping primers (P1, P2 and P3; Fig. 3B) from this ITR sequence, which were used to test MLEs in human, mouse, rat, hamster, sheep and cattle genomes by PCR. To allow for the probable degeneration of the sequences of these ITRs, the DNAs were tested with six primer mixes (each primer alone and three primer pairs). Each genomic DNA gave PCR products with 3 or 4 of the primer mixes and contained several amplified fragments (Fig. 4). These PCR products were hybridized with probes corresponding to mixes of the five human, six ovine or six bovine MLE clones. Southern blots washed at high stringency revealed that MLEs had indeed been amplified from human, ovine and bovine DNA (Fig. 4). However, some of the amplified fragments did not hybridize with the probe. They were presumably either spurious products or very substantially altered MLEs. Only the use of the bovine MLE probe at low washing stringency revealed MLEs in rodent genomes. These results were thus consistent with those obtained by Southern blotting with PCR products of the first amplifications using the degenerate primers (Fig. 1B and 1C). They confirmed that MLEs are present in the genomes of the six mammalian species analyzed: human, mouse, rat, hamster, sheep and cattle.

Four particular MLE-hybridizing fragments were amplified from nearly all the mammalian genomes tested. Their sizes were about 1300 bp, 980 bp, 850 bp and 650 bp. The 1300 bp fragments probably correspond to full-sized elements whereas 980 bp, 850 bp and 650 bp are presumably deleted forms. Forms with large deletions, having only 850 bp, 560 bp and 350 bp of the MLE have previously been described in hymenopteran genomes [2]. A 850 bp form is thus found in hymenopteran and mammalian species. As suggested for the *KP* element, and *Th* elements in the P [14] and *hobo* [15] transposon systems of *D. melanogaster*, this 850 bp deleted element might be involved in the regulation of the transposition activity of MLEs.

Our results and those previously published [1–7] demonstrate that elements of the Tc1-*mariner*-like superfamily are widespread in fungi, invertebrates and vertebrates species. As in vertebrate, they have been detected in fishes and in mammalians it is likely suggested that they would also be present in the genome of other vertebrate groups; thus, MLEs are the first DNA mobile ubiquitous elements described to date.

Their distribution among various and distant species could

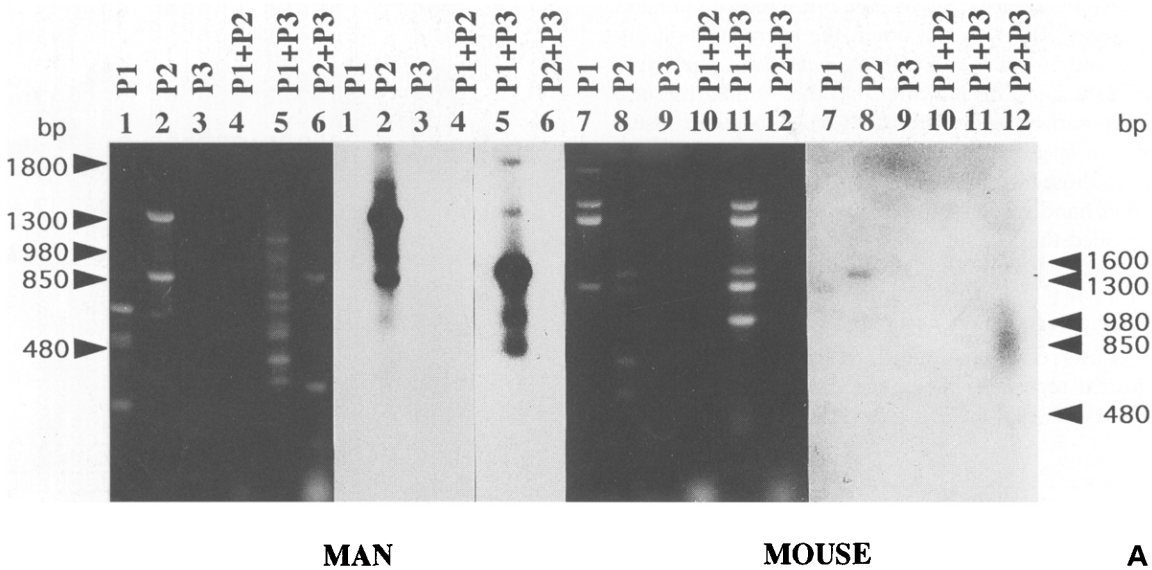


Fig. 4. PCR products from human (lanes 1 to 6), mouse (7 to 12), rat (13 to 18), hamster (19 to 24), ovine (25 to 30) and cattle (31 to 36) genomic DNA, analysed on 1.5% agarose gels stained with EtBr (left panel) and their respective Southern blots (right panel). Molecular weights on the left correspond to lanes 1 to 6, 13 to 18 and 25 to 30. Molecular weights on the right are for the remaining lanes. PCR products were obtained using six different primer mixes: lanes 1, 7, 13, 19, 25, 31 (P1); lanes 2, 8, 14, 20, 26, 32 (P2); lanes 3, 9, 15, 21, 27, 33 (P3); lanes 4, 10, 16, 22, 28, 34 (P1+P2); lanes 5, 11, 17, 23, 29, 35 (P1+P3); lanes 6, 12, 18, 24, 30, 36 (P2+P3). Southern blot hybridizations were performed with human (1 to 6), ovine (25 to 30) and bovine (7 to 24, and 31 to 36) MLE probes. Hybridizing fragments were observed with human, ovine and bovine PCR products at high stringency (1 to 6, 25 to 36). The bovine MLE probe hybridized at low stringency to rodent PCR products (7 to 24). PCR conditions are the same as in Fig. 1 (16), except that 0.1 μM rather than 0.8 μM of each oligonucleotide was used.

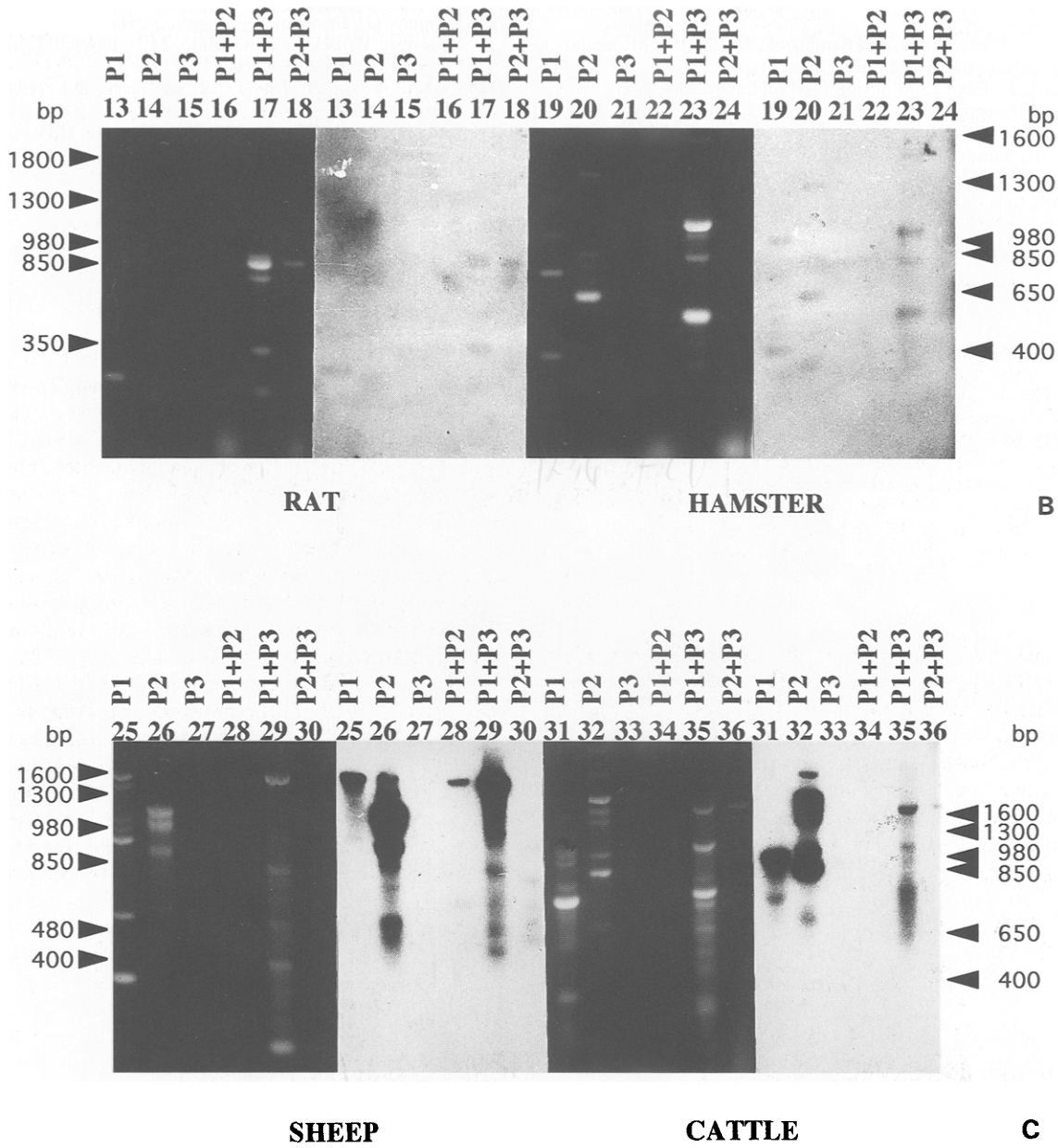


Fig. 4. (*continued*).

result from old horizontal transferts between the insects and the vertebrates. Such horizontal transferts have been recently evoked for Tc1-like elements between invertebrate (nematodes) and vertebrate (fishes) [10]. The presence of active MLEs in *D. mauritiana* [16] and *H. cecropia* [17] genomes suggests that MLEs could still be transfered between genomes.

Mammalian appear to carry numerous MLEs. They may thus be useful as probes and tags for genetic mapping. Moreover, the molecular features of MLEs among vertebrate and invertebrate species therefore offer a new perspective to use insect MLEs as xenovectors for transgenosis applications in vertebrate systems, as it was efficiently done in invertebrate systems [10,17].

Acknowledgements: This work was supported by grants from the C.N.R.S. (URA1298), M.E.S.R. and GREG 94.

References

[1] Robertson, H.M. (1993) Nature 362, 241–245.
[2] Bigot, Y., Hamelin, M.H., Capy, P. and Periquet, G. (1994) Proc. Natl. Acad. Sci. USA 91, 3408–3412.
[3] Sedensky, M.M., Hudson, S.J., Everson, B. and Morgan, P.G. (1994) Nucleic Acids. Res. 9, 1719–1723.
[4] Capy, P., Langin, T., Bigot, Y., Brunet, F., Daboussi, M.J., Periquet, G., David, J.R. and Hartl, D.L. (1994) Genetica 93, 161–170.
[5]. Langin, T., Capy, P. and Daboussi, M.J. (1995) Genetica 93, 49–59.
[6] Heierhorst, J., Lederis, K. and Richter, D. (1992) Proc. Natl. Acad. Sci. USA 89, 6798–9802.
[7] Goodier, J.L. and Davidson, W.S. (1994) J. Mol. Biol. 241, 26–34.
[8] Berg, D.E. and Howe, M.M. eds. (1989) in: Mobile DNA (Berg, D.E. and Howe, M.H., Eds.) American Society for Microbiology, Washington, DC.
[9] Doak, T.G., Doerder, F.P., Jahn, C.L. and Herrick, G. (1994) Proc. Natl. Acad. Sci. USA 91, 942–946.

[10] Lidholm, D.A., Lohe, A.R. and Hartl, D.L. (1993) Genetics 134, 859–868.
[11] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning: A Laboratory Manual (Maniatis, T., Fritsch, E.F. and Sambrook, J., Eds.) Cold Spring Harbor Laboratory, NY.
[12] Bigot, Y., Hamelin, M.H. and Periquet, G. (1991) J. Evol. Biol. 4, 541–556.
[13] Dessen, P., Fondrat, C., Valencien, C. and Mugnier, C. (1990) Cabios 6, 355–356.
[14] Rio, D.C. (1991) Trends Genet. 5, 282–287.
[15] Periquet, G., Lemeunier, F., Bigot, Y., Hamelin, M.H., Bazin, C., Ladevèze, V., Eeken, J., Galindo, M.I., Pascual, L. and Boussy, I. (1994) Genetica 93, 79–90.
[16] Capy, P., Koga, A., David, J.R. and Hartl, D.L. (1992) Genetics 130, 499–506.
[17] Lidholm, D.A., Gumunson, G.H and Boman, H.G. (1991) J. Biol. Chem. 266, 11518–11521.