Available online at www.sciencedirect.com

ScienceDirect

GENOMICS

Genomics 90 (2007) 416–420

www.elsevier.com/locate/ygeno

Short Communication

Identification of an active ID-like group of SINEs in the mouse

David H. Kass *, Nicole Jamison

Department of Biology, Eastern Michigan University, Ypsilanti, MI 48197, USA

Received 27 April 2007; accepted 11 May 2007
Available online 14 June 2007

Abstract

The mouse genome consists of five known families of SINEs: B1, B2, B4/RSINE, ID, and MIR. Using RT-PCR we identified a germ-line transcript that demonstrates 92.7% sequence identity to ID (excluding primer sequence), yet a BLAST search identified numerous matches of 100% sequence identity. We analyzed four of these elements for their presence in orthologous genes in strains and subspecies of *Mus musculus* as well as other species of *Mus* using a PCR-based assay. All four analyzed elements were identified either only in *M. musculus* or exclusively in both *M. musculus* and *M. domesticus,* indicative of recent integrations. In conjunction with the identification of transcripts, we present an active ID-like group of elements that is not derived from the proposed BC1 master gene of ID elements. A BLAST of the rat genome indicated that these elements were not in the rat. Therefore, this family of SINEs has recently evolved, and since it has thus far been observed mainly in *M. musculus,* we refer to this family as MMIDL.

Keywords: SINE; Retrotransposon; Mouse genome; ID elements

Short interspersed DNA elements (SINEs) represent a nonautonomous group of retrotransposons of approximately 75–400 nucleotides [1]. These consist of internal A- and B-box RNA polymerase III promoters, 3′-A tails, and flanking direct repeats that are generated as a result of the integration event [1]. Mice are known to contain four distinct SINE families referred to as B1, B2, B4/RSINE, ID, and MIR, consisting of 564 K, 348 K, 391 K, 79 K, and 115 K copies, respectively [2]. Of these SINEs, recent integration events have been observed in the mouse for B1 [3–5] and B2 elements [6]. No recent ID integrations have been identified to date in the mouse, although an example has been found in the rat [7], the genome of which contains an order of magnitude more ID elements than in the mouse [8].

Only a small number of SINEs are retrotranspositionally competent [9] and only a single SINE master gene has been identified. This is the BC1 RNA gene, which serves as the master gene for ID elements as supported by the expression of this gene in germ-line cells (although mainly in nerve cells), as well as its corresponding match to the consensus sequence of genomic ID elements among various rodents [10]. We performed RT-PCR to identify B2 transcripts in the germ-line of mouse and cloned and analyzed the sequences. We identified two clones that were not B2, but seemed related to ID (92.7% sequence identity excluding the primers), although not matching the BC1 gene. Upon performing a BLAST search, we found numerous 100% hits in the mouse genome, but when strictly searching the rat genome, the best hit was 96% and the variants indicated that this sequence was a rat ID element. Therefore, we further investigated the newly identified mouse element to determine its status as a member of a possible previously unidentified SINE family.

Results

Using RT-PCR from RNA isolated from mouse ovary incorporating a B2 primer, three clones were isolated. Using a BLAST search, one RT-PCR clone (MOR2-A) matched two chromosome 7 clones (Accession Nos. AC107862.10 and AC 114612.13), whereas the other two isolated clones (MOR2-B and MOR2-C) were identical to each other, including a G at the same position in the A tail. These two elements were clearly not B2, but more closely related to ID elements as demonstrated by

* Corresponding author. Fax: +1 734 487 9235.
E-mail address: dkass@emich.edu (D.H. Kass).

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.05.007

92.7% sequence identity (excluding the primer sequence) (Fig. 1). The generation of two identical cDNA clones of only three clones analyzed, yet with varied sequence identity to the B2 primer (Fig. 2), suggests the isolation of true transcripts, which appear to represent an uncharacterized family of SINEs that we will refer to as MMIDL (*Mus musculus* ID-like) and which was investigated further (see below). Two B2 RT-PCR clones were obtained from the mouse testes, one (MTR2-A) was similar to ID but identical to MOR2-B, whereas the other clone (MTR2-B) demonstrated 100% sequence identity (excluding the primer) to the 3′UTR of a mouse cerebellum cDNA for calcium/calmodulin-dependent protein kinase 2 (Accession No. AK160 407) as well as a genomic sequence (AC115728.11). Since this clone demonstrated little sequence identity to any other sequence in the database (about 77% was the next best match), it was apparent that it was truly the protein kinase gene that was isolated.

A BLAST search of the entire GenBank database was performed utilizing the MOR2-B cDNA clone (excluding primer sequence). Ninety matches of 100% sequence identity were identified (all mouse sequences), including mRNA-derived sequences and overlapping BACs. However, the numerous 100% hits are indicative of a young group of SINEs potentially derived from the transcripts obtained from the ovary and the testis. We isolated some of the top hits, to obtain full-length elements, since the B2 primer was used for the RT-PCR and hence the 5′ sequence was unknown. By identifying the flanking direct repeats, full-length elements were defined. We used the entire element to repeat the BLAST search screening the mouse genome assembly (Build 36) and identified 66 hits of 100% sequence identity and 10 hits with only a single variant for the 76-nucleotide sequence. Overall there were 122 hits of at least 93% sequence identity. When only rat sequences were screened, the top hit was 96%, but the nucleotide variants were diagnostic for ID elements. This provides evidence that the MMIDL SINE family is not found in the rat and represents a small, relatively young group of potentially active SINEs. An alignment of four MMIDL elements isolated from a BLAST search along with the top 2 rat hits and the BC1-based ID elements is shown in Fig. 3.

We utilized a PCR assay to assess the presence and absence of the MMIDL in orthologous loci of various strains, subspecies of *Mus musculus,* and related species of *Mus.* Four sequences (non-mRNA and on different chromosomes to avoid getting the same sequence from overlapping BACs) in which primers containing unique sequences could be obtained were analyzed. The elements MMIDL2, MMIDL3, MMIDL5, and MMIDL6 with the accession numbers and the loci names that they were identified in are provided in Table 1. Additionally, the presence and absence of the elements for the four loci among the various analyzed mice are provided in Table 2. A sample gel is shown in Fig. 4. For MMIDL2, clearly a recent integration has occurred, possibly strictly in the C57 strain of mouse as this was the only one identified, although since it is inbred our data cannot conclude with certainty that it is not found in other *M. musculus.* MMIDL3 was identified in some strains of *M. musculus,* but not others, and was found in *M. domesticus,* although not in *M. hortulanus* or *M. caroli,* indicative of an integration event within the past 0.5 million years. Similar results were obtained for MMIDL5, although the one difference was that this element is present in *M. musculus castaneus.* MMIDL6 was identified only within three strains of *M. musculus,* but found to be heterozygous in the ICR Swiss mouse, which, being outbred, would not be unusual and would be indicative of a polymorphism in *M. musculus.* In general, of the four elements analyzed, all demonstrated recent integrations within *Mus* as none was found in *M. spretus, M. hortulanus,* or *M. caroli,* supporting this as an active mouse SINE family that is ID-like but not derived from the BC1 master gene and therefore derived from a different master gene.

```
B2F primer:    ccgaccgtaggggctggtgagatggctcag
B2:                           .....................
ID (BC1):                     ....t...g..tt.a......
MMIDL:                        ....t...g..tt.a......
B1:                             .c.g..cat.g....c..c
```

Fig. 2. Alignment of 5′ ends of mouse SINEs in relation to the B2 forward primer used for RT-PCR primer. Note that sequences were added to the B2F primer at the 5′ for the purpose of cloning into alternative constructs (for use in a separate study). MMIDL represents the new family of SINEs identified in this study. Dots indicate nucleotide sequences identical to the B2F primer.

Discussion

We present an as yet uncharacterized family of SINEs in the mouse genome that is very similar to ID elements and which we refer to as MMIDL. Although our transcriptional analysis is limited, MMIDL may be highly expressed in both the ovaries and the testes, as we used a B2 primer in our RT-

```
MTR2A:      TGGTAAAGTGCTTGCCTAGCAAGCCCAAGGCCCTGGGTTCGGTCCTCA
MOR2B:      ..................................................
MOR2C:      ..................................................
Mouse ID:   .....G..C..............G.........................
Rat ID:     .....G..C..............G.........................

MTR2A:      GCTCCGGAAAAAAGAAAAAAAAAAAAAAAAAAAA
MOR2B:      ...........................
MOR2C:      ...........................
Mouse ID:   ....T..
Rat ID:     ......
```

Fig. 1. Alignment of clones derived from RT-PCR and their correspondence with the most closely related known rodent SINE family (ID). Note that primer sequences were excluded from the comparison. Dots indicate nucleotide sequences identical to MTR2A.

```
MMIDL6:    GGGGTTGGGGATTTAGCTCAGTGGTAAAGTGCTTGCCTAGC
MMIDL5:    .........................................
MMIDL3:    .........................................
MMIDL2:    .........................................
RAT #1:    ........................G................
RAT #2:    ........................G................
MOUSE ID:  ........................G..C............
RAT ID:    ........................G..C............

MMIDL6:    AAGCCCAAGGCCCTGGGTTCGGTCCTCAGCTCCGGAAAAAA
MMIDL5:    ........................................G
MMIDL3:    .........................................
MMIDL2:    ............................A............
RAT #1:    ....G.........................A.......
RAT #2:    ....G.................C........A.......
MOUSE ID:  ....G.........................T..
RAT ID:    ....G.........................

MMIDL6:    GAAAAAAAAGAAAAAAAAAGAAAAAAAAAGAAAGAAAAAAA
MMIDL5:    .........A.G.......A.G.......A
MMIDL3:    A......G.A.........AG........A...A..
MMIDL2:    .........A.........A.....
RAT #1:    .....
RAT #2:    ......G..A..

MMIDL6:    ATAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGAAGAAAAAAA

MMIDL6:    AAAGAAAAAGAAAAAAGAAAAAAAAAAAAC
```

Fig. 3. Alignment of four MMIDL genomic sequences identified by a BLAST search queried with the RT-PCR products shown in Fig. 1, with the two top rat-specific BLAST search hits, and the mouse ID derived from the BC1 master gene. Dots indicated nucleotide sequence identical to MMIDL6.

PCR assay and yet MMIDL was the primary observed product. Our analysis of orthologous mouse loci demonstrates that MMIDL is an active family (or ID subfamily) of elements not derived from BC1, and its absence in the rat supports that this SINE family is of recent origin. Recent integrations of B1 elements in mouse have been studied in detail [3], in contrast to B2, although recent B2 integrations have been identified in mice [6] and mouse cell lines [11]. We, however, show that MMIDL is highly active and maybe currently to a much greater extent than B2 as a result of transcriptional levels. This would explain the lack of B2 RT-PCR products in contrast to MMIDL, which was particularly surprising considering a B2-based forward primer was used and that B2 elements have been demonstrated to be expressed at high levels in mouse germ-line tissue [12,13]. Additionally, BC1 RNA has been shown to be expressed in substantial levels in mouse male germ cells [14], yet although ID and MMIDL are identical in sequence in the primer region, none of the RT-PCR clones was ID, whereas two of three ovary clones and one of two testis clones were MMIDL. Additional studies are under way to analyze transcriptional levels of MMIDL and other SINEs to help validate this finding.

MMIDL, being a low-copy-number active SINE family, offers the unique opportunity of increased potential for identifying a master gene. BLAST searches of the mouse genome queried with MOR2B and MTR2A sequences, inclusive of the A-tail punctuated at the same position by a guanosine (but differing in A-tail length, likely the result of binding location of the oligo(dT)-based primer), generated 16 and 7 perfect matches, respectively. Utilizing sequences identified between the A-tail and the RNA Pol III termination signals of this narrowed group of loci can serve to develop 3′-end primers for RT-PCR

Table 1
Analyzed MMIDL loci, including primer sequences, accession numbers from which the elements were identified, and chromosomal location

Locus	Accession No.	Chr.	Forward primer	Reverse primer	Size +IDL	Size −IDL
MMIDL-2	AC202775.6	8	5′-GATTGGGATGAGTGCTGGCTC-3′	5′-GCCCTGGACACACAAAACTGC-3′	231	120
MMIDL-3	AC153624.20	6	5′-AATTCCTAGTGAGGGGACATTATGG-3′	5′-TTCTGCTTCCCAAGCCCTGAC-3′	413	290
MMIDL-5	AC188461.3	7	5′-TCCAAGAAGCCAGGGAACTGTC-3′	5′-TCACCTCTGGGATCAGCAAACC-3′	383	263
MMIDL-6	AC099576.12	3	5′-AAAGGGATAGGAATGGTGTGGG-3′	5′-TTCAGGTTGTTGCTCAGTTGACAG-3′	441	207

with a 5′-end MMIDL primer. Alternatively the C-RACE technique [15] may be incorporated, as it has previously been used to analyze SINE-containing transcribed loci.

The identification of this novel SINE family demonstrates how dynamic the mammalian genome is and that new SINE families can continue to be derived. MMIDL may have been derived from an ID element that incorporated mutations but integrated in a fortuitous location in the genome that allowed for high levels of germ-line transcription and the opportunity to "hijack" LINE machinery for further amplification. This is also the first example of an ID-related element demonstrated to jump recently within the mouse genome, although recent B1 and B2 integrations have been characterized. BC1-based ID elements have been found to jump recently only in rat, plus they have been highly active in the rat since the split from a common ancestor of the mouse [7]. Additionally, the diagnostic positions for the MMIDL (27-A, 30-T, 45-C), with only position 30 being a mutation-susceptible CpG site, are not found in any of the more recent rat ID subfamilies described by Kim and Deininger [7]. There was an additional variant for MMIDL-2 (72-A), which, based on a BLAST search, is unique for this element and not representative of another subfamily of IDLs. As with other SINEs, recent integrations being identical by descent (i.e., a shared integration is extremely unlikely to be the result of homoplasy [16]) may allow its use as a phylogenetic tool depending on the point in *Mus* evolutionary history at which these elements emerged, as well as potentially to assess the origins of various strains and population patterns of *M. musculus*, further contributing to the data generated using B1 SINEs [3,5].

Materials and methods

DNA from the following *Mus* species and strains were purchased from The Jackson Laboratory: *M. musculus musculus* (CZECH II/Ei), *M. m. domesticus* (ZALENDE/Ei), *M. m. domesticus* (SKIVE/Ei), *M. m. molossinus* (MOLF/Ei), *M. m. castaneus* (CASA/Rk), *M. hortulanus* (PANCEVO/Ei), *M. spretus* (SPRET/Ei), *M. caroli,* and *M. pahari.* DNA from the *M. musculus* strains ICR, BALB/c, and C57/BL6 were kindly provided by Dr. Astrid Engel, at Tulane University Medical Center. Mouse ovary and testes tissues were obtained from Rockland Immunochemicals (Gilbertsville, PA, USA).

RNA was extracted from mouse ovary and testes utilizing the SV Total RNA Isolation System kit from Promega (Milwaukee, WI, USA) according to the manufacturer's protocols. RT-PCR was performed using the Promega Reverse Transcription System. cDNA was synthesized from 2 μl isolated RNA in a 20-μl volume with 5 mM $MgCl_2$, Reverse Transcription 1× Buffer (Promega), 1 mM dNTP, recombinant RNasin ribonuclease inhibitor, AMV reverse transcriptase, and 0.025 μg oligo(dT)-based Adaptor primer (5′-GCCTTCGAATTCAG-GTTTTTTTTTTTTTT-3′). The reaction was incubated at 42°C for 15 min, 95°C for 5 min, and 4°C for 5 min. This was followed by PCR using 10 μl cDNA in a 50-μl volume containing 0.18 mM dNTP, 1.875 mM $MgCl_2$, Reverse Transcription 0.98× Buffer, 0.01 μg adaptor-based primer (5′-GCCTTCGAA-TTCAGGTT-3′), 1 mM B2-based primer (5′-CCGACCGTAGGGGCTGGT-GAGATGGCTCAG-3′) (note that additional 5′-end sequences were added simply for cloning into alternative constructs), and 2 U *Taq* polymerase (Promega). The reactions were carried out in an MJ Research thermocycler under the following conditions: 94°C for 2 min; 94°C for 20 s, 42°C for 20 s, 72°C for 20 s for 32 cycles; 72°C for 5 min.

Table 2
Results of PCR-based assay for detection of presence (+) or absence (−) of the MMIDL element

Locus	Balb/c	ICR	C57	Mmc	Mmm	Md	Ms	Mh	Mc	Mp
MMIDL-2	−	−	+	−	−	−	0	0	0	0
MMIDL-3	+	+	+	−	−	+	0	−	−	0
MMIDL-5	+	+	+	+	−	+	−	−	−	0
MMIDL-6	+	+/−	+	−	−	−	−	−	−	0

A (0) indicates no PCR product was observed. Analyzed mouse strains and species include Balb/c, ICR Swiss (outbred), C57/BL6, *Mus musculus castaneus* (Mmc), *M. musculus molossinus* (Mmm), *M. domesticus* (Md), *M. spretus* (Ms), *M. hortulanus* (Mh), *M. caroli* (Mc), and *M. pahari* (Mp).

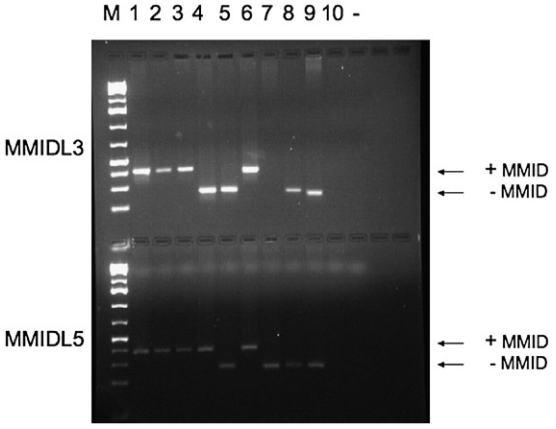

Fig. 4. Examples of analyses of presence/absence variations of MMIDL elements among strains and subspecies of *M. musculus*, and related *Mus* species. PCR amplicons were separated on 2% agarose gels. Lane M, 1 kb plus ladder (Fermentas); 1, Balb/c; 2, C57/BL6; 3, ICR 4; *M.m. castaneus*; 5, *M.m. molossinus*; 6, *M. domesticus*; 7, *M. spretus*; 8, *M. hortulanus*; 9, *M. caroli*; 10, *M. pahari*; -, negative control (no DNA).

A BLAST search using the RT-PCR MMIDL sequence was performed, and top hits were selected to investigate recent integration events. To analyze orthologous loci of different species of *Mus* and subspecies and strains of *M. musculus,* primers were designed that flank the MMIDL sequences and thereby test for presence or absence of B1 elements. The MacVector (Oxford Molecular Group) DNA sequence software program was used to determine the effectiveness of the primers for amplification. The absence of the MMIDL element yields a fragment approximately 120 bp smaller than templates containing the MMIDL element (including the A-tail and target-site duplication). The MMIDL6 presence form, however, was 234 bp larger than the absence form, as the A-tail was an enormous 119 nucleotides, punctuated with 11 guanines and 1 thymine.

PCR amplifications were performed in 20-μl volumes containing 1× GoTaq buffer (Promega), 200 μM dNTPs (Fermentas), 1.5 mM $MgCl_2$, 0.25 μM each primer (Table 1), 1 U GoTaq DNA polymerase (Promega), and 50 ng DNA. Reactions were performed using an MJ Research thermocycler with various annealing temperatures under the following conditions: 94°C for 2 min, 1×; 94°C for 15 s, 53°C for 15 s (except for MMIDL2, for which a 56°C annealing temperature was used), 72°C for 30 s (except for MMIDL2, 15 s), 30×; and 72°C for 5 min. Amplification products were analyzed by 2% agarose gel electrophoresis with 1× TAE buffer, stained with ethidium bromide, and analyzed on a UV gel documentation system (Ultra-Lum, Claremont, CA, USA).

Direct cloning of the PCR products and RT-PCR products was performed by TA cloning into the pGEMTEasy vector (Promega) according to the manufacturer's protocols and transforming competent JM109 *Escherichia coli* cells. Using blue/white colony screening, white colonies were selected and plasmid DNA was isolated using the Promega Wizard Plus SV Plasmid DNA Isolation Kit.

For verification of the correct locus, at least one product with an insert and one without an insert were cloned and sequenced. These samples and RT-PCR

clones were submitted to Functional Biosciences (Madison, WI, USA) for DNA sequencing. BLAST searches of the flanking sequences were performed to verify the products were obtained from orthologous loci.

Acknowledgments

This work was supported by NIH AREA Award GM062828 to D.H.K. and an Undergraduate Honors Thesis Award to N.J. We thank Dr. Joo Kim, Louisiana State University, for advice on database analysis.

References

[1] J.H. Rogers, The origin and evolution of retrotransposons, Int. Rev. Cytol. 93 (1985) 187–279.
[2] R.H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J.F. Abril, P. Agarwal, et al., Initial sequencing and comparative analysis of the mouse genome, Nature 420 (2002) 520–562.
[3] D.H. Kass, M.E. Raynor, T.M. Williams, Evolutionary history of B1 retroposons in the genus Mus, J. Mol. Evol. 51 (2000) 256–264.
[4] N. Gilbert, J.M. Bomar, M. Burmeister, J.V. Moran, Characterization of a mutagenic B1 retrotransposon insertion in the jittery mouse, Hum. Mutat. 24 (2004) 9–13.
[5] P. Munclinger, P. Boursot, B. Dod, B1 insertions as easy markers for mouse population studies, Mamm. Genome 14 (2003) 359–366.
[6] A.M. Roy, C. Gong, D.H. Kass, P.L. Deininger, Recent B2 element insertions in the mouse genome, DNA Seq. 8 (1998) 343–348.
[7] J. Kim, P.L. Deininger, Recent amplification of rat ID sequences, J. Mol. Biol. 261 (1996) 322–327.
[8] D.H. Kass, J. Kim, P.L. Deininger, Sporadic amplification of ID elements in rodents, J. Mol. Evol. 42 (1996) 7–14.
[9] P.L. Deininger, M.A. Batzer, C.A. Hutchison III, M.H. Edgell, Master genes in mammalian repetitive DNA amplification, Trends Genet. 8 (1992) 307–311.
[10] J. Kim, J.A. Martignetti, M.R. Shen, J. Brosius, P. Deininger, Rodent BC1 RNA gene as a master gene for ID element amplification, Proc. Natl. Acad. Sci. USA 91 (1994) 3607–3611.
[11] I. Oberbaumer, Retroposons do jump: a B2 element recently integrated in an 18S rDNA gene, Nucleic Acids Res. 20 (1992) 671–677.
[12] A.P. Ryskov, P.L. Ivanov, O.N. Tokarskaya, D.A. Kramerov, M.S. Grigoryan, G.P. Georgiev, Major transcripts containing B1 and B2 repetitive sequences in cytoplasmic poly(A)+ RNA from mouse tissues, FEBS Lett. 182 (1985) 73–76.
[13] R. Bachvarova, Small B2 RNAs in mouse oocytes, embryos, and somatic tissues, Dev. Biol. 130 (1988) 513–523.
[14] I.A. Muslimov, Y. Lin, M. Heller, J. Brosius, Z. Zakeri, H. Tiedge, A small RNA in testis and brain: implications for male germ cell development, J. Cell Sci. 115 (2002) 1243–1250.
[15] J. Kim, D.H. Kass, P.L. Deininger, Transcription and processing of the rodent ID repeat family in germline and somatic cells, Nucleic Acids Res. 23 (1995) 2245–2251.
[16] D.A. Ray, J. Xing, A.H. Salem, M.A. Batzer, SINEs of a nearly perfect character, Syst. Biol. 55 (2006) 928–935.