J Mol Evol (2000) 51:256–264
DOI: 10.1007/s002390010087

JOURNAL OF MOLECULAR EVOLUTION

Evolutionary History of B1 Retroposons in the Genus *Mus*

David H. Kass, Melissa E. Raynor,* Thomas M. Williams*

Department of Biology, 316 Mark Jefferson, Eastern Michigan University, Ypsilanti, MI 48197, USA

Received: 25 February 2000 / Accepted: 5 June 2000

Abstract. Short interspersed DNA elements (SINEs) amplify by retroposition either by (i) successive waves of amplification from one or a few evolving master genes or by (ii) the generation of new master genes that coexist with their progenitors. Individual, highly conserved, elements of the B1 SINE family were identified from the GenBank nucleotide database using various B1 subfamily consensus query sequences to determine their integration times into the mouse genome. A comparison of orthologous loci in various species of the genus *Mus* demonstrated that four subfamilies of B1 elements have been amplifying within the last 1–3 million years. Therefore, B1 sequences are generated by coexisting source genes. Additionally, three B1 subfamilies have been concurrently propagated during subspecies divergence and strain formation in *Mus*, indicating very recent activity of this retroposon family. The patterns of intra- and interspecies variations of orthologous loci demonstrate the usefulness of B1 integrations as a phylogenetic tool. A single inconsistency in the phylogenetic trends was depicted by the presence of a B1 insert in an orthologous locus exclusively in *M. musculus* and *M. pahari*. However, DNA sequence analysis revealed that these were independent integrations at the same genomic site. One highly conserved B1 element that integrated at least 4–6 million years ago suggests the possibility of occasional function for B1 integrations.

Key words: B1 retroposon — SINE subfamilies — *Mus* — B1 integration — Master gene — SINE evolution

**Present address:* Department of Molecular Biology, Parke-Davis, 2800 Plymouth Road, Ann Arbor, MI 48105, USA
Correspondence to: David Kass; e-mail: david.kass@emich.edu

Introduction

B1 sequences are a family of short interspersed DNA elements (SINEs) specific to the genomes of rodents, originating from the 7SL RNA gene (Rogers 1985). There are an estimated 80,000–100,000 B1 copies per haploid genome in *Mus* (Krayev et al. 1980; Rogers 1985). These elements, referred to as retroposons, are amplified in the genome by RNA-mediated transposition. The majority of B1 elements are pseudogenes, as only a few loci (termed master or source genes) are capable of serving as the source for retroposition (Deininger et al. 1992). B1 elements have demonstrated recent activity by yielding a polymorphism in an androgen-related mouse gene (King et al. 1986). This element was found integrated into this gene in the DBA strain of *Mus musculus*, but lacking in the C57BL/6 and BALB/c strains.

Distinct SINE families are concurrently active in rodent genomes. For instance, current activity of the tRNA-derived B2 repeats has been demonstrated by recent insertions present only in certain strains of mice (Kominami et al. 1983; Roy et al. 1998), as well as in a cell line (Oberbäumer 1992). No evidence exists for recent activity of the tRNA-derived ID retroposons in mice, although these elements have generated insertional polymorphisms in the rat (Schuler et al. 1983; Kim and Deininger 1996).

The rodent B1 family, based on a detailed sequence analysis (Quentin, 1989), has been grouped into six subfamilies distinguished by diagnostic variants (Fig. 1). The subfamily structure of B1, based on diagnostic positions and sequence divergence (pairwise similarities), implies successive waves of subfamily amplification during the course of rodent evolution (Fig. 1), the youngest subfamily represented by B1–B. This evolutionary

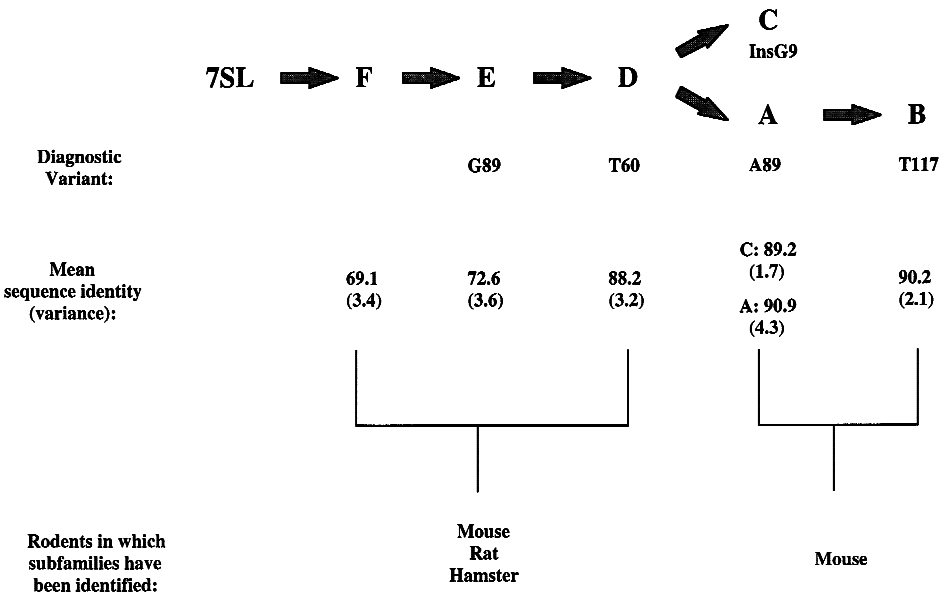

Fig. 1. Evolution of B1 subfamilies as determined by Quentin (1989), including diagnostic nucleotide changes, pairwise similarities, and rodents known to harbor these sequences.

pattern is consistent with the evolution of *Alu* sequences in primates, in which the sequential amplification of subfamilies has been derived by a limited number of master genes (Deininger et al. 1992). Four of the six B1 subfamilies demonstrate equivalent pairwise similarities (Quentin 1989; Fig. 1), indicating either that subfamily evolution occurred in a brief period of time or that there has been concurrent evolution of B1 subfamilies (suggestive of multiple master loci). This study addresses the more recent activity of the B1 retroposon by analyzing the most conserved elements and estimating integration times by comparing orthologous loci of various species of *Mus.*

SINE integrations are highly stable and provide highly useful phylogenetic markers (Cook and Tristem 1997), since the shared insertion of the element within a specific DNA sequence would be identical by descent rather than by state. Analysis of these integrations has been utilized to determine phylogenetic relationships in salmon and trout (Murata et al. 1993, 1996) and cetartiodactyls (Nikaido et al. 1999). Therefore, this work also addresses the possible utility of isolating young B1 elements as a phylogenetic tool within the genus *Mus.*

Methods

Samples

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.* Liver tissue from the *M. musculus* strains ICR, BALB/c, and C57BL/6 were provided by Dr. Astrid Engel, Louisiana State University Medical Center.

DNA Isolation

DNA was extracted from the livers of the three strains of *M. musculus* using the Super Quik Gene kit (Analytical Genetic Testing Center, Inc., Denver, CO).

Identification of Loci Containing Young B1 Elements

Highly conserved B1 sequences were selected by a bias screening of the GenBank database by a BLAST search (Altschul et al. 1990) using segments of the subfamily consensus sequences (Quentin 1989).

PCR, Cloning, and DNA Sequencing

To analyze orthologous loci of different species of *Mus,* and subspecies and strains of *M. musculus,* primers were designed that flank the B1 sequences and thereby test for the presence or absence of B1 elements. GeneWorks (Oxford Molecular Group) was used to determine the effectiveness of the primers for amplification. The absence of the B1 element yields a fragment approximately 135 bp smaller than templates containing the B1 element. Divergence times of these mice were used to estimate times of integration of individual B1 elements.

PCR amplifications were performed in 20-μl volumes containing 1× Taq buffer (GIBCO), a 200 μ*M* concentration of dNTPs (Pharmacia), 3 m*M* $MgCl_2$, a 0.25 μ*M* concentration of each primer (Table 1), 1 U Taq DNA polymerase (GIBCO), and 50 ng DNA. Reactions were performed using an MJ thermal cycler at various annealing temperatures (NN_1/NN_2) for the different loci (see Table 1) under the following conditions: 94°C for 2 min, 1×; 94°C for 15 s, NN_1°C for 15 s, and 72°C for 30 s, 5×; 94°C for 15 s, NN_2°C for 15 s, and 72°C for 30 s, 25×; and 72°C for 5 min. Amplification products were analyzed by 1.5% agarose gel electrophoresis.

For verification of the correct locus, at least one product with an

Table 1. Conditions and primers used for amplifying B1-containing loci

B1 subfamily	Locus (accession No.)	Primers	Annealing conditions	Expected size (bp)[a]
B	MMTDGF1 (X93470)	5′-acgcgatcggtctttccagttc-3′ 5′-cgacttacctgaaatgtaaga-3′	53°C/49°C	680
B	MMILBP (U00938)	5′-ccagtgttacagaaggttgt-3′ 5′-ccacagagataatgcaagcata-3′	59°C/56°C	614
B	MMAFPGEN (X87098)	5′-tgagtttgggataggtaagat-3′ 5′-ttatgctacagtcggcctgttc-3′	60°C/57°C	560
B	MMWHNGENE (Y12488)	5′-gccctgcttctttgtgaact-3′ 5′-tgcctggacaatcttcagaac-3′	53°C/49°C	640
A	MMHC438N12 (AF049850)	5′-gctttctgatgaaaggcctg-3′ 5′-gctcatttcctctgacccaa-3′	61°C/59°C	370
A	AF030883 (AF030883)	5′-ttcggtgagttggaggctag-3′ 5′-ccaagtagctcccaacgaga-3′	60°C/57°C	595
C	MUSBAND31 (M16536)	5′-tggcctcgaactccctaagt-3′ 5′-atggctcctgtacgaccatg-3′	57°C/54°C	467
C	AC002121 (AC002121)	5′-aactgtaacgcaaggctgtg-3′ 5′-cccttcatcctgaagactcg-3′	53°C/49°C	525
D	MMU39442 (U39442)	5′-gccacgtattgctctggtcta-3′ 5′-cacatcacatccttgccact-3′	58°C/55°C	454
D	MMHC135G15 (AF050157)	5′-catggttgtccacactctcc-3′ 5′-acaagcctttcactggcttg-3′	61°C/59°C	485

[a] Expected size of the product containing the B1 insert. The subfamily represented by the B1 element is indicated.

insert and one without an insert (if applicable) were cloned and sequenced. Direct cloning of the PCR products was performed by TA-cloning into either the pGEMT vector (Promega) or the pCR2.1-TOPO vector (Invitrogen).

DNA sequencing was performed using universal primers (M13R, T7, SP6) either manually, using Sequenase (USB) with separation by electrophoresis using a denaturing 5% Long Ranger (FMC) gel and autoradiography, or by the use of an ABI prism automated DNA sequencer.

Results

Identification of Younger B1 Elements Among Various Subfamilies

We chose 10 loci to analyze, representing constituents of the four youngest subfamilies (A, B, C, D) of B1 elements, three of which (A, B, C) are found in mice but not rats (Quentin 1989). These elements are among the most highly conserved in relation to their respective consensus sequences. We isolated four B subfamily loci (Table 2), as these represent the youngest subfamily based on the evolutionary scheme of Quentin (1989) and are therefore predicted to have a higher probability of being more recently integrated. The D-subfamily element in the locus MMHC135G15 was the first listed sequence of the BLAST search using differing subfamily consensus sequences. This was not an anticipated result since members of an older subfamily would typically demonstrate greater divergence.

Determination of Integration Times by PCR of Ortholgous Loci: B Subfamily Analysis

Each of the 10 loci was analyzed by using various species of *Mus* to determine the integration time of the B1 element (copies of raw data available by request). A representative locus (MMAFPGEN) is shown in Fig. 2. Divergence times during the course of evolution of *Mus* species and subspecies are based on Bonhomme and Guénet (1996) and Boursot et al. (1993) and represented in Fig. 3. An amplified fragment approximately 135 bp less than expected indicates the absence of the B1 element, which was verified by cloning and sequencing at least one absence variant. At least one presence variant for each locus was sequenced to confirm the correct locus and SINE integration. In cases where the B1–B integration appeared to be specific for *M. musculus* (Table 2), several subspecies and strains were subsequently analyzed (Table 3).

Among the four B1–B elements analyzed, two (MMTDGF1 and MMWHNGENE) were identified only in certain strains of *M. musculus* (Table 2 and 3), indicating that this subfamily has been active since the radiation of subspecies within the past 500,000 years (Fig. 3). Because several strains have these inserts, the insertion predates the development of the strains. The presence of the B1 within these two loci in the genomes of ICR, BALB/c, and ZALENDE/Ei (*M. m. domesticus*) suggests that these strains were derived from the *domesticus* subspecies. Recent retropositions of B1 elements may therefore provide a tool for assessing the origins of subspecies and strains of *Mus.*

M. pahari seemed to contain a B1 insert in the MMWHGENE locus. More surprisingly, the analysis of the *M. pahari* sequence for this B1-containing locus (Fig. 4) suggests a separate insertion into precisely the same site. The evidence supporting independent integrations includes (i) two distinct variants (note the nucleotides labeled with an asterisk in Fig. 4) within the flanking

Table 2. Analysis of integrations of conserved B1 elements, representing different B1 subfamilies, using various species of *Mus*[a]

No.	Locus	Insertion							Time of insertion (mya)	Variants
		Mm^1	Mm^2	Mmd	Ms	Mh	Mc	Mp		
	B subfamily									
1	MMTDGF1	+	+	+	−	−	−	0	<0.5	4 (3)
2	MMILBP	+	+	+	−	+	−	−	<1	3
3	MMAFPGEN	+	+	+	+	+	−	−	1–3	7 (6)
4	MMWHNGENE	+	+	+	−	−	−	+*	<0.5	4 (3)
	A subfamily									
5	MMHC438N12	+	+	+	+	+	−	−	1–3	5 (4)
6	AF030883	−	+	−	−	−	0	0	<0.5	4 (3)
	C subfamily									
7	MUSBAND31	+	+	+	+	+	+	0	≥3–6	7 (6)
8	AC002121	+	+	+	+	+	0	0	≥1–3	9 (8)
	D subfamily									
9	MMU39442	+	+	+	+	+	−	−	1–3	5 (4)
10	MMHC135G15	+	+	+	+	+	+	+	>4–6	3

[a] Loci correspond to GenBank designations. A plus sign indicates the presence of the B1 element within the particular locus; a minus sign absence of the B1; a zero indicates no amplification product. Species designations: Mm^1 and Mm^2, *Mus musculus musculus;* (BALB/c and ICR strains, respectively); Mmd, *Mus musculus domesticus;* Ms, *Mus spretus;* Mh, *Mus hortulanus;* Mc, *Mus caroli;* Mp, *Mus pahari.* Variants are in relation to subfamily consensus sequences described by Quentin (1989); the number in parentheses takes into account a common variant to the consensus. Times of insertion of individual B1 elements are based on their presence/absence in various *Mus* species. The asterisk represents an independent same-site insertion (see text).

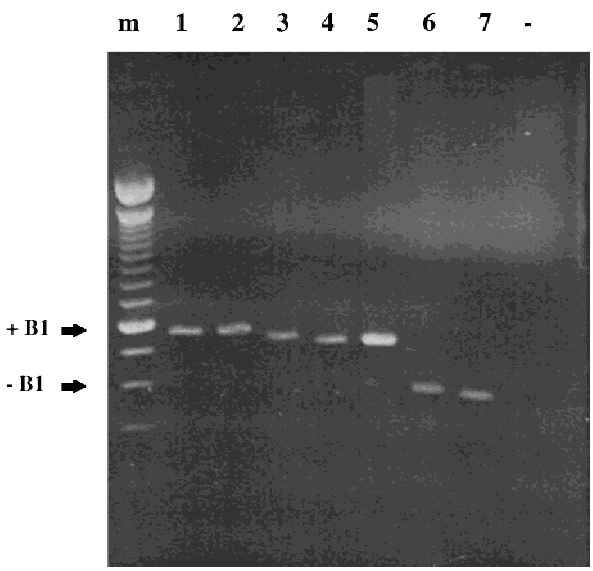

Fig. 2. Analysis of the MMAFPGEN locus for the presence or absence of a B1–B integration. Lane m, 100-bp ladder (Gibco BRL); lane 1, *M. m. musculus* (BALB/c); lane 2, *M. m. musculus* (ICR); lane 3, *M. m. domesticus;* lane 4, *M. spretus;* lane 5, *M. hortulanus;* lane 6, *M. caroli;* lane 7, *M. pahari;* lane −, negative control (no template DNA).

direct repeats between *M. pahari* and *M. musculus* both precede and succeed the B1 integrations in their respective species (one variant, however, is an insertion/deletion of a nucleotide and either may have mutated postdating the integration or may indicate that the independent integrations occurred at positions that differed by a single nucleotide); (ii) *M. hortulanus*, which lacks the insert, contains the sequence identical to the direct repeat of *M. musculus* (Fig. 4), representing the ancestral form of this locus between these two species; (iii) the B1 insert in *M. pahari* appears to be a B1–C/B1–B intermediate rather than B1–B; and (iv) the sequences of the A-rich tails are considerably more variant between the species than are the A-rich flanking direct repeats, plus there are no shared variants that punctuate the polyA tail. In contrast, a comparison of the MMTDGF1 locus between that obtained for *M. m. domesticus* and the published *M. musculus* (strain 129/SV) sequence shows features representing a single integration, such as identical flanking direct repeats and identical punctuations (two cytosines) within the A-tail.

Although most similar to the consensus sequence, the B1–B element in the MMILBP locus was present in all *M. musculus* strains analyzed in this study, as well as in *M. hortulanus* (Table 2). However, it was absent in the orthologous locus of *M. spretus*. The evolution of these three species of *Mus* is unclear (Prager et al. 1996) and usually shown as a trichotomous split (Fig. 3), although this result may suggest that *M. spretus* is a sister taxon. Identification of additional B1 markers may potentially resolve this uncertainty. The apparently oldest B1–B analyzed (MMAFPGEN) is found in *M. hortulanus*, *M. spretus*, and *M. musculus* and has integrated within 1–3 mya (Fig. 3). As predicted, the B1–B subfamily is very young, with half of the analyzed B1–B-containing loci demonstrating variation within *M. musculus* subspecies.

A Subfamily Analysis

Two B1–A-containing loci were analyzed for integration times. The B1–A located in the AF030883 locus was

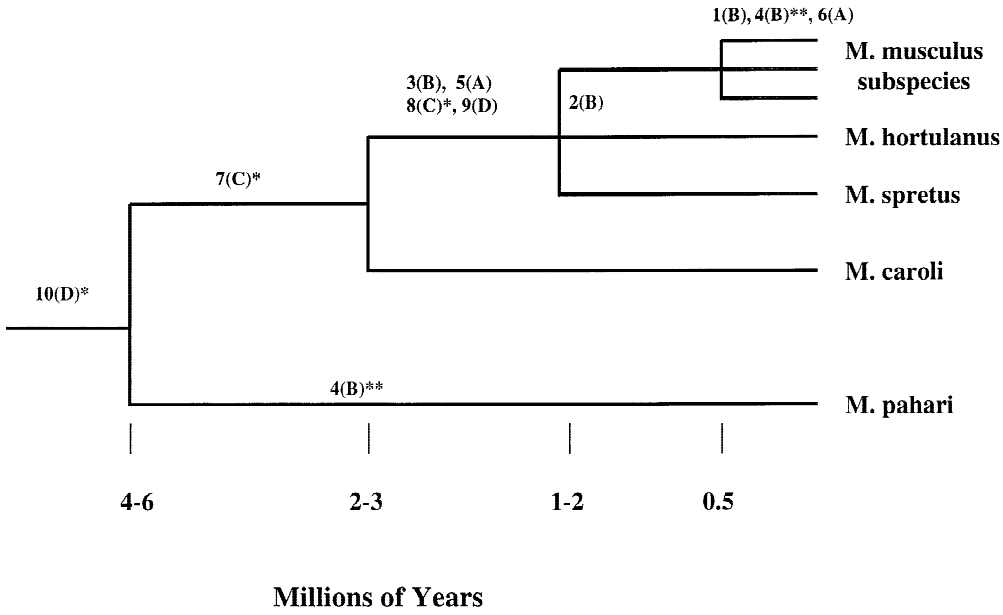

Fig. 3. Phylogenetic relationship between *Mus* species and B1 integrations. Evolutionary relationships and divergence times are based on Boursot et al. (1993) and Bonhomme and Guénet (1996). Numbers correspond to loci in Table 2. *Letters in parentheses* refer to the B1 subfamily. *Single asterisks* indicate that the time of integration may be greater for these loci due to the lack of identification of amplification products of more divergent species. *Double asterisks* indicate independent, same-site, integrations.

Table 3. Analysis of B1-containing loci that are variant within *Mus musculus*[a]

	Strain							
Locus	BALB	C57	ICR	CZECH	CASA	MOLF	SKIVE	ZALEN
1(B)								
MMTDGF1	+	−	+	−	−	−	−	+
4(B)								
MMWHNGENE	+	+	+	−	−	−	−	+
6(A)								
AF030883	−	+	+	−	+	−	−	−

[a] A plus sign refers to the presence of the B1 and a minus sign indicates the absence. Subspecies and strain designations: BALB, BALB/c *M. musculus* ssp.; C57, C57BL/6 *M. m.* ssp.; ICR, ICR *M. m.* ssp., CZECH, CZECH II/Ei *M. m. musculus;* CASA, CASA/Rk *M. m. castaneus;* MOLF, MOLF/Ei *M. m. molossinus;* SKIVE, SKIVE/Ei *M. m. domesticus;* ZALEN, ZALENDE/Ei *M. m. domesticus.*

found only in some strains of *M. musculus* (Tables 2 and 3), indicating a recent integration (Fig. 3). The C57BL/6, ICR, and CASA/Rk strains contain the insert. The B1–A within MMHC438N12 apparently integrated prior to the *spretus, hortulanus, musculus* split (Fig. 3). Like the B1–B elements, B1–A has been active for about 1–3 million years and demonstrates recent activity as well. Although the B subfamily master gene was apparently derived from the A subfamily master gene (Quentin 1989), both have been concurrently active.

C Subfamily Analysis

The most conserved B1–C elements demonstrated more variation from the consensus than did the B1–A and B1–B elements to their resepective consensus sequences. The B1–C element within MUSBAND31 was observed in *M. caroli,* as well as *M. hortulanus, M. spretus,* and *M. musculus* (Table 2), indicating that its integration occurred at least between 3–6 mya (Fig. 3). This time frame could not be narrowed since no amplification product was obtained for *M. pahari.* The AC002121 B1–C insert was identified in *M. musculus, M. hortulanus,* and *M. spretus,* although no product was obtained for *M. caroli* or *M. pahari* (Table 2), possibly the result of an acquired mutation in the primer region. Although only two B1–C loci were analyzed, King et al. (1986) identified a polymorphic B1 element in the RP2 gene, which was present in two DBA strains and absent in the BALB/cJ and C57BL/6J strains of *M. musculus.* Application of subfamily diagnostic nucleotides places this element in the B1–C subfamily. Therefore, although apparently older in

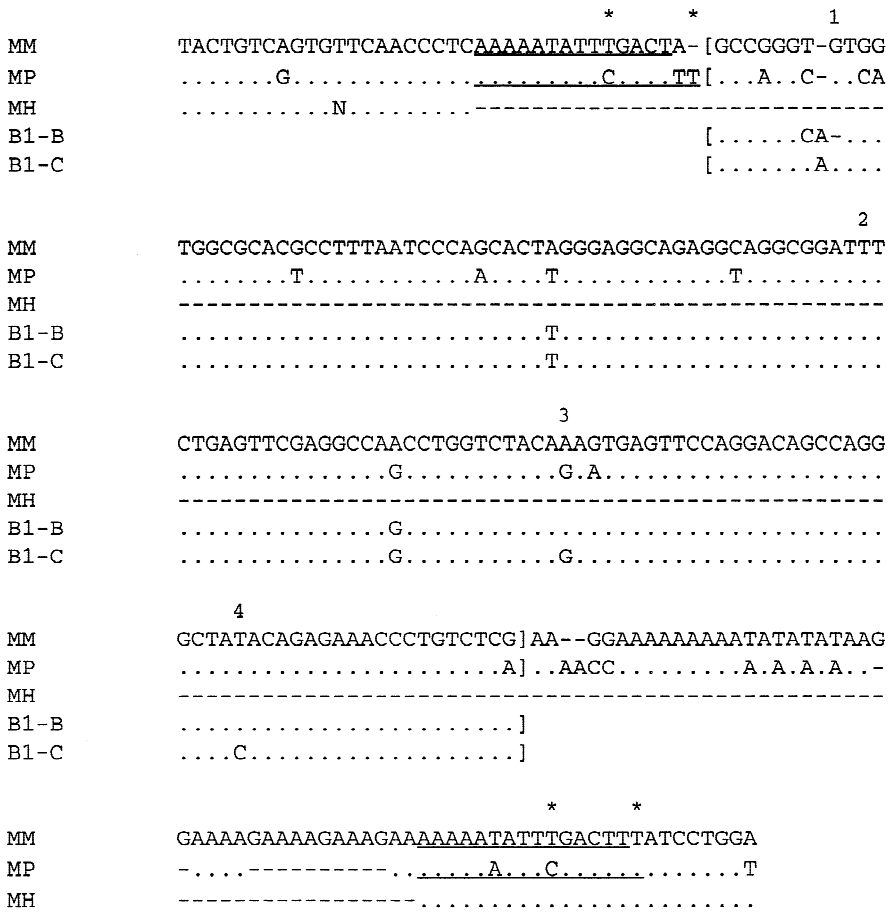

Fig. 4. Sequences of orthologous B1–B-containing loci (4B in Table 2). MM refers to *Mus musculus*, MP refers to *M. pahari*, MH refers to *M. hortulanus* (which lacks the B1 insert), and B1–B and B1–C are consensus sequences for their respective B1 subfamilies. The B1 sequences are *bracketed. Dots* indicate sequence identity to MM. *Dashes* are used for maximal alignment. *Underlined sequences* represent flanking direct repeats. *Asterisks* refer to variants that both precede and succeed the B1 integration. The *numbers* indicate the four diagnostic nucleotide positions that distinguish the B1 subfamilies (Fig. 1).

origin than B1–A and B1–B, it has demonstrated recent activity.

D Subfamily Analysis

Although not restricted to the mouse, the B1–D sequences demonstrate a pairwise similarity approaching that of the mouse exclusive elements (Fig. 1). Two highly conserved B1–D elements were analyzed. The B1–D integration into MMU39442 is absent in *M. caroli* and *M. pahari* but present in the other analyzed species (Table 2) and therefore inserted between 1 and 3 mya (Fig. 3). This indicates somewhat recent activity of this B1 subfamily. The B1–D integration into the MMHC135G15 locus was the most conserved sequence (along with MMILBP) in relation to its respective subfamily consensus sequence (Table 2), suggesting a recent insertion. Additionally, there was no variation between the *M. m. domesticus* B1–D sequence (obtained in this study) within MMHC135G15 and the GenBank sequence. However, because this insert was found in all analyzed species of *Mus*, the integration must have occurred at least 4–6 mya. The relative lack of random mutations of this B1–D element suggests the possibility of function.

Discussion

Concurrent Generation of B1 Elements Arising from Differing Master Genes

Ten B1-containing orthologous loci, representing four subfamilies (based on diagnostic nucleotides (Quentin 1989)), were analyzed for insertion times. The results are summarized in Fig. 3. All four subfamilies were concurrently active between 1 and 3 mya as depicted by their presence in loci in *M. spretus*, *M. hortulanus*, and *M. musculus* but absence in *M. caroli* and *M. pahari*. According to Quentin (1989), B1–C and B1–A were both derived from B1–D, while B1–A yielded B1–B (Fig. 1). These results are in agreement with Quentin, as the four combined B1–C and B1–D elements analyzed were generally older than the four B1–B and two B1–A elements. However, the integration time of a B1–C element (MUSBAND31) is older than a B1–D element (MMU39442), indicating that the subfamily evolution was not the result of a change in the master gene but was, rather, the result of the formation of a new master gene. In addition, three master genes have been active during *M. musculus* subspecies formation, as evidenced by the two B1–B and one B1–A presence/absence variants

within *M. musculus* and the previously identified B1–C variant within the RP2 gene (King et al. 1986). Therefore, although successive waves of amplification may occur based on the successive generation of master genes, the original master gene apparently maintains its activity. At least three B1 subfamilies (master genes) have been active during "recent" times, therefore continuing to have an impact on the genome of this species.

Utilization of B1 Integrations as a Phylogenetic Tool

SINE integrations have been proposed to be a highly valuable tool in phylogenetics (Cook and Tristem 1997; Murata et al. 1993, 1996; Nikaido et al. 1999). The power of this type of analysis comes from use of data from multiple loci. Additionally, this type of analysis is simple to employ by PCR and agarose gel electrophoresis and can provide as much information as detailed sequence analyses and Southern blotting as depicted by the use of *Alu* integrations in human population studies to support the Out-of-Africa hypothesis (Batzer et al. 1994). The limiting factor in the use of SINEs involves the identification of individual elements that provide useful markers. By using a biased search of the genetic database, we have demonstrated rapid identification of B1 elements useful in the analysis of *Mus* phylogeny. All but possibly 2 elements postdate a particular point (node) within the evolution of this genus, and 3 of the 10 analyzed elements demonstrate variation within *M. musculus* species. Therefore, we predict that employing a biased search of the genome would generate numerous additional markers, particularly using the A, B, and C subfamilies.

The locus referred to as 4B appears to be the single exception for integrations following phylogenetic trends. Based on the DNA sequence of locus 4B in *M. pahari*, this anomaly is apparently the result of same-site independent integrations. However, the impact of independent integrations warrants only minimal consideration when utilizing B1 integrations as a phylogenetic tool, since there was no evidence for parallel integrations in the other analyzed B1-containing loci. As an example, a contrast of sequences of the AC002121 locus between *M. hortulanus* (this study) and *M. musculus* (GenBank), species that diverged about 1–2 mya, reveals that (i) the direct repeats flanking the inserts are identical between the species, (ii) the inserted B1 element is in the same SINE subfamily, and (iii) the species share the same two variants that punctuate the A-tail. In each representative presence variant that was sequenced, there was no indication of independent integrations in relation to the published sequence. Additional criteria we observe in some cases for distinguishing a single-shared integration are provided when flanking direct repeats are not identical, but the orthologous repeats between the two mouse genomes are the same. The rarity of independent same-site integration is also supported by analyses of recent *Alu* integrations in humans, which indicate that either these SINEs are absent in orthologous loci of other primates or, in a few cases, integration of a "young" *Alu* element predates the divergence of humans and chimpanzees. In either case, there is a correspondence with primate phylogeny. In a rare case, a gene conversion event altered the integrated element to represent a different *Alu* subfamily (Kass et al. 1995), although it was the *Alu* sequence, not the integration event, that deviated from phylogenetic trends. Only a single recognized example exists of independent *Alu* integrations within the vicinity of each other in an orthologous locus between the chimpanzee and the human (Arcot et al. 1998) yielding similarly sized PCR fragments. However, unlike locus 4B, these were not within the same site. Nikaido et al. (1999) also identified proximal integrations of retroposons. However, these independent insertional events were observed in genomes of organisms distinguished at more distant taxonomic levels, which significantly increases the chance of identifying this type of occurrence.

The divergence of *M. hortulanus*, *M. spretus*, and *M. musculus* has been controversial (Prager et al. 1996), hence usually written as a trichotomous split. Although the data we present are limited, the finding of a shared variant between *M. hortulanus* and *M. musculus* is suggestive of a clade, with *M. spretus* the sister group. This evidence supports the detailed mitochondrial DNA sequence analysis by Prager et al. (1996, and references therein). Another key advantage of using SINE integrations as markers is that the insertions are stable, and therefore each variant undeniably represents a synapomorphy; i.e., these types of variants are identical by descent rather than by state.

The origin of strains and geographic races of *Mus* has been difficult to assess (Santos et al. 1993; Prager et al. 1996; Boissinot and Boursot 1997). Rapid identification and ease of use make B1 integration markers a promising new tool for this purpose. The limited data of this study do not allow for definitive conclusions. However, identical patterns between CZECH II/Ei and MOLF/Ei are consistent with the close phylogenetic positioning of these strains by RFLP analysis of genomic DNA (Santos et al. 1993). The C57BL strain was shown to contain the Y chromosome of Asian origin (Tucker et al. 1992), which is the area inhabited by *M. m. castaneus* (Bonhomme and Guenet 1996), although *M. m. musculus* inhabits the more northern regions with areas of intermixing. Additionally, Santos et al. (1993) found the *M. m. castaneus* strain CAST/Ei [which has the same origins as CASA/Ei (Festing 1996)] to be more related to C57BL/6J than either is to MOLF/Ei or CZECH II, although less related to each other than CZECH II and MOLF/Ei are. Our data are consistent with these findings, as two of three variants are shared between C57 and CASA in relation to all three variants shared between MOLF and

CZECH. This might suggest that C57 contains predominantly the *M. m. castaneus* genome. BALB/c has the identical pattern to ZALENDE/Ei, suggesting that BALB/c is derived from *M. m. domesticus.*

Genetic Impact of Retroposons

Retroposon integrations can alter the genetic makeup of an organism by insertional mutagenesis (Wallace et al. 1991) or by providing sites for unequal homologous recombination (Lehrman et al. 1987). Some DNA regions appear to be hot spots for retroposon integrations, as indicated by the many *Alu* elements within the low-density lipoprotein receptor gene (Lehrman et al. 1987). Burton et al. (1991) identified an L1 retrotransposition within the same genomic site of *M. musculus* and *M. pahari,* but absent in *M. caroli,* and could not resolve the possibilities of a precise deletion, an ancestral polymorphism, or independent insertions. The rationale arguing against the latter scenario, in addition to being extremely unlikely, was the lack of insertional specificity previously observed for L1 elements. A profoundly striking result in this study was the identification of independent B1 integrations into the same genomic site (locus 4B). This was supported by the conserved variants of the flanking direct repeats, highly distinctive A-rich tails (but not the A-rich flanking sequences), the absence of shared variants that punctuate the A tail, the sequence of *M. hortulanus* (which lacks the insert), and the subfamily affiliations of the inserts. Therefore, the combined findings of Burton et al. (1991) and this study support the existence of highly specific genomic sites susceptible to integrations of RNA-mediated transposable elements.

Occasionally, a retroposon integration may yield a new and/or improved function, a process referred to as exaptation (Brosius 1991). The BC200 RNA gene, a primate gene primarily expressed in the brain, provides an example of an *Alu* integration that was exapted (Brosius 1991). Additionally, a B2 integration yielded a new polyadenylation signal in a major histocompatibility gene (Kress et al. 1984). The relative conservation (to the consensus sequence) of the B1–D element in the mouse MMHC135G15 locus may suggest functional significance. Alternatively, this conservation may have been the consequence of a gene conversion event, which would further support this mechanism as a secondary mode of SINE evolution (Kass et al. 1995).

Acknowledgments. D.H.K. was supported through an Eastern Michigan University Spring/Summer Research Award and a University Research in Excellence Fund. T.M.W. was supported by a Departmental Hellwig Apprenticeship. We also thank Stephen Durkee for technical assistance and Dr. John Doucet for comments on the manuscript.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arcot SS, Adamson AW, Risch GW, LaFleur J, Robichaux MB, Lamerdin JE, Carrano AV, Batzer MA (1998) High-resolution cartography of recently integrated human chromosome 19-specific Alu fossils. J Mol Biol 281:843–856

Batzer MA, Stoneking M, Alegria HM, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL (1994) African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 91:12288–12292

Boissinot S, Boursot P (1997) Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the house mouse: Selection on the Y chromosome? Genetics 146:1019–1034

Bonhomme F, Guénet J-L (1996) The laboratory mouse and its wild relatives. In: Lyon MF, Rastan S, Brown SDM (eds) Genetic variants and strains of the laboratory mouse, Vol. 2. Oxford University Press, Oxford, pp 1577–1596

Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F (1993) The evolution of house mice. Annu Rev Ecol Syst 24:119–152

Brosius J (1991) Retroposons—Seeds of evolution. Science 251:753

Burton FH, Loeb DD, Edgell MH, Hutchison CA III (1991). L1 gene conversion or same-site transposition. Mol Biol Evol 8:609–619

Cook JM, Tristem M (1997) 'SINEs of the times'—Transposable elements as clade markers for their hosts. Trends Ecol Evol 12:295–297

Deininger PL, Batzer MA, Hutchison CA III, Edgell MH (1992) Master genes in mammalian repetitive DNA amplification. Trends Genet 8:307–311

Festing MFW (1996) Origins and characteristics of inbred strains of mice. In: Lyon M, Sohaila R, Brown SDM (eds) Genetic variants and strains of the laboratory mouse, Vol. 2. Oxford University Press, Oxford, pp 1537–1576

Kass DH, Batzer MA, Deininger PL (1995) Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. Mol Cell Biol 15:19–25

Kim J, Deininger PL (1996) Recent amplification of rat ID sequences. J Mol Biol 261:322–327

King D, Snider LD, Lingrel JB (1986) Polymorphism in an androgen-regulated mouse gene is the result of the insertion of a B1 repetitive element into the transcription unit. Mol Cell Biol 6:209–217

Kominami R, Muramatsu M, Moriwaki K (1983) A mouse type 2 Alu sequence (M2) is mobile in the genome. Nature 301:87–89

Krayev AS, Kramerov DA, Skryabin KG, Ryskov AP, Bayev AA, Georgiev GP (1980) The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. Nucleic Acids Res 8:1201–1215

Kress M, Barra Y, Seidman JG, Khoury G, Jay G (1984) Functional insertion of an Alu type 2 (B2 SINE) repetitive sequence in murine class I genes. Science 226:974–977

Lehrman MA, Goldstein JL, Russell DW, Brown MS (1987) Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. Cell 48:827–835

Murata S, Takasaki N, Saitoh M, Okada N (1993) Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc Natl Acad Sci USA 90:6995–6999

Murata S, Takasaki N, Saitoh M, Tachida H, Okada N (1996) Details of retropositional genome dynamics that provide a rationale for a generic division: The distinct branching of all the pacific salmon and trout (Oncorhynchus) from the Atlantic salmon and trout (Salmo). Genetics 142:915–926

Nikaido M, Rooney AP, Okada N (1999) Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. Proc Natl Acad Sci USA 96:10261–10266.

Oberbäumer I (1992) Retroposons do jump: A B2 element recently integrated in an 18S rDNA gene. Nucleic Acids Res 20:671–677

Prager EM, Tichy H, Sage RD (1996) Mitochondrial DNA sequence variation in the eastern house mouse, Mus musculus: Comparison with other house mice and report of a 75-bp tandem repeat. Genetics 143:427–446

Quentin Y (1989) Successive waves of fixation of B1 variants in rodent lineage history. J Mol Evol 28:299–305

Rogers JH (1985) The origin and evolution of retroposons. Int Rev Cytol 93:187–279

Roy AM, Gong C, Kass DH, Deininger PL (1998) Recent B2 element insertions in the mouse genome. DNA Sequence 8:343–348

Santos J, Cole Y, Pellicer A (1993) Phylogenetic relationships among laboratory and wild-origin Mus musculus strains on the basis of genomic DNA RFLPs. Mammal Genome 4:485–492

Schuler LA, Weber JL, Gorski J (1983) Polymorphism near the rat prolactin gene caused by insertion of an Alu-like element. Nature 305:159–160

Tucker PK, Lee BK, Lundrigan BL, Eicher EM (1992) Geographic origin of the Y chromosomes in “old” inbred strains of mice. Mammal Genome 3:254–261

Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS (1991) A de novo Alu insertion results in neurofibromatosis type 1. Nature 353:864–866