Available online at www.sciencedirect.com
ScienceDirect
Gene 407 (2008) 176–185
GENE
www.elsevier.com/locate/gene

Newly discovered young CORE-SINEs in marsupial genomes ☆

Maruo Munemasa [a], Masato Nikaido [a], Hidenori Nishihara [a], Stephen Donnellan [b], Christopher C. Austin [c], Norihiro Okada [a,*]

[a] *Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama, Japan*
[b] *Evolutionary Biology Unit, South Australian Museum, North Terrace, SA 5000, Australia*
[c] *Museum of Natural Science, Louisiana State University, LA 70803, USA*

Received 3 July 2007; received in revised form 2 October 2007; accepted 4 October 2007
Available online 12 October 2007
Received by Takashi Gojobori

Abstract

Although recent mammalian genome projects have uncovered a large part of genomic component of various groups, several repetitive sequences still remain to be characterized and classified for particular groups. The *s*hort *in*terspersed repetitive *e*lements (SINEs) distributed among marsupial genomes are one example. We have identified and characterized two new SINEs from marsupial genomes that belong to the CORE-SINE family, characterized by a highly conserved "CORE" domain. PCR and genomic dot blot analyses revealed that the distribution of each SINE shows distinct patterns among the marsupial genomes, implying different timing of their retroposition during the evolution of marsupials. The members of Mar3 (*Marsupialia* 3) SINE are distributed throughout the genomes of all marsupials, whereas the Mac1 (*Macropodoidea* 1) SINE is distributed specifically in the genomes of kangaroos. Sequence alignment of the Mar3 SINEs revealed that they can be further divided into four subgroups, each of which has diagnostic nucleotides. The insertion patterns of each SINE at particular genomic loci, together with the distribution patterns of each SINE, suggest that the Mar3 SINEs have intensively amplified after the radiation of diprotodontians, whereas the Mac1 SINE has amplified only slightly after the divergence of hypsiprimnodons from other macropods. By compiling the information of CORE-SINEs characterized to date, we propose a comprehensive picture of how SINE evolution occurred in the genomes of marsupials.

Keywords: CORE-SINE; Marsupials; Phylogeny; Evolution

1. Introduction

Mammalian genomes harbor a large amount of retroposons that propagate their copies in the host genome via an RNA intermediate generated from a "copy and paste" mechanism called retroposition (Rogers, 1985; Weiner et al., 1986; Brosius, 1991; Okada, 1991a,b). *S*hort *in*terspersed repetitive *e*lements (SINEs) belong to a class of retroposons that account for more than ten percent of nuclear DNA. The role of SINEs in the host genome still remains to be clarified; however, recent studies, including those from our laboratory, have found that some SINE-derived non-coding sequences are highly conserved (Nishihara et al., 2006a). This implies that these SINEs might have acquired some functionality during the evolution (Nishihara et al., 2006a; Bejerano et al., 2006; Mikkelsen et al., 2007). It may therefore be useful to characterize and categorize the genomic components of various mammals with respect to SINEs. Recent comprehensive genome sequencing projects have allowed us to investigate particular animals on the whole-genome level (e.g. Margulies et al., 2005), providing a very powerful tool for revealing a complete picture of SINE evolution. Indeed, owing to the completion of the human genome project, the contribution of SINEs to the human genome has been clarified in detail — the Alu fraction covers

Abbreviations: SINE, short interspersed repetitive element; LINE, long interspersed repetitive element; PCR, polymerase chain reaction; mya, million years ago.

☆ The nucleotide sequences reported in this paper have been submitted to GenBank and have been assigned accession numbers AB326393 to AB326416.

* Corresponding author. Department of Biological Sciences, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259-B21, Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan. Tel.: +81 45 924 5742; fax: +81 45 924 5835.
E-mail address: nokada@bio.titech.ac.jp (N. Okada).

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.10.008

more than 13% and *l*ong *in*terspersed repetitive *e*lements (LINEs) comprise more than 20% of the whole genome (International Human Genome Sequencing Consortium, 2001). Furthermore, recent genome project on short-tailed opossum (*Monodelphis domestica*) have revealed that SINEs cover more than 10% and LINEs comprise more than 29% of its genome (Gentles et al., 2007).

More than 30 SINE families have been characterized based on their structure. Usually SINEs are composed of a 5′ terminal tRNA-or 7SL RNA-related region containing a pol III promoter and a partner LINE-related 3′ tail. Furthermore, several SINE families are grouped into a superfamily based on the presence of a central conserved domain. To date, three superfamilies have been characterized as V-SINEs (Ogiwara et al., 2002), Due-SINEs (Nishihara et al., 2006a) and CORE-SINEs (Gilbert and Labuda, 1999), which exist in vertebrate and invertebrate genomes. Among these three superfamilies, CORE-SINEs are considered to be a rather young group and some intact CORE-SINEs are thought to possess retropositional activity in mammalian (especially non-eutherian) genomes (Gilbert and Labuda, 2000). The CORE element, which is the central conserved domain of CORE-SINEs, was initially reported as mammalian interspersed repeats (MIRs) and is widely distributed among mammalian genomes (Jurka et al., 1995; Smit and Riggs, 1995). Later, this MIR was divided into two families, Ther1 (MIR in RepBase Reports) and Ther2 (MIR3), which are distributed among the genomes of Theria (extant "Theria" consists of all mammals except for platypus and echidnas) (Gilbert and Labuda, 2000). MIRs are the most prevalent repeat in the human genome next to Alu, in that Ther1 shares 2.2% and Ther2 shares 0.3% of the draft human genome sequence (International Human Genome Sequencing Consortium, 2001). The Ther1 and Ther2 are highly divergent, and seem to have lost their retropositional activity before the split of monotremes, marsupials and eutherians, which occurred more than 110 mya (million years ago). Although the CORE-SINEs lack retropositional activity in the genomes of eutherians, they are still active in non-eutherian genomes. Gilbert and Labuda (2000) reported the presence of three additional CORE-SINE families (Mon1, Mar1 and Opo1). The members of these families are

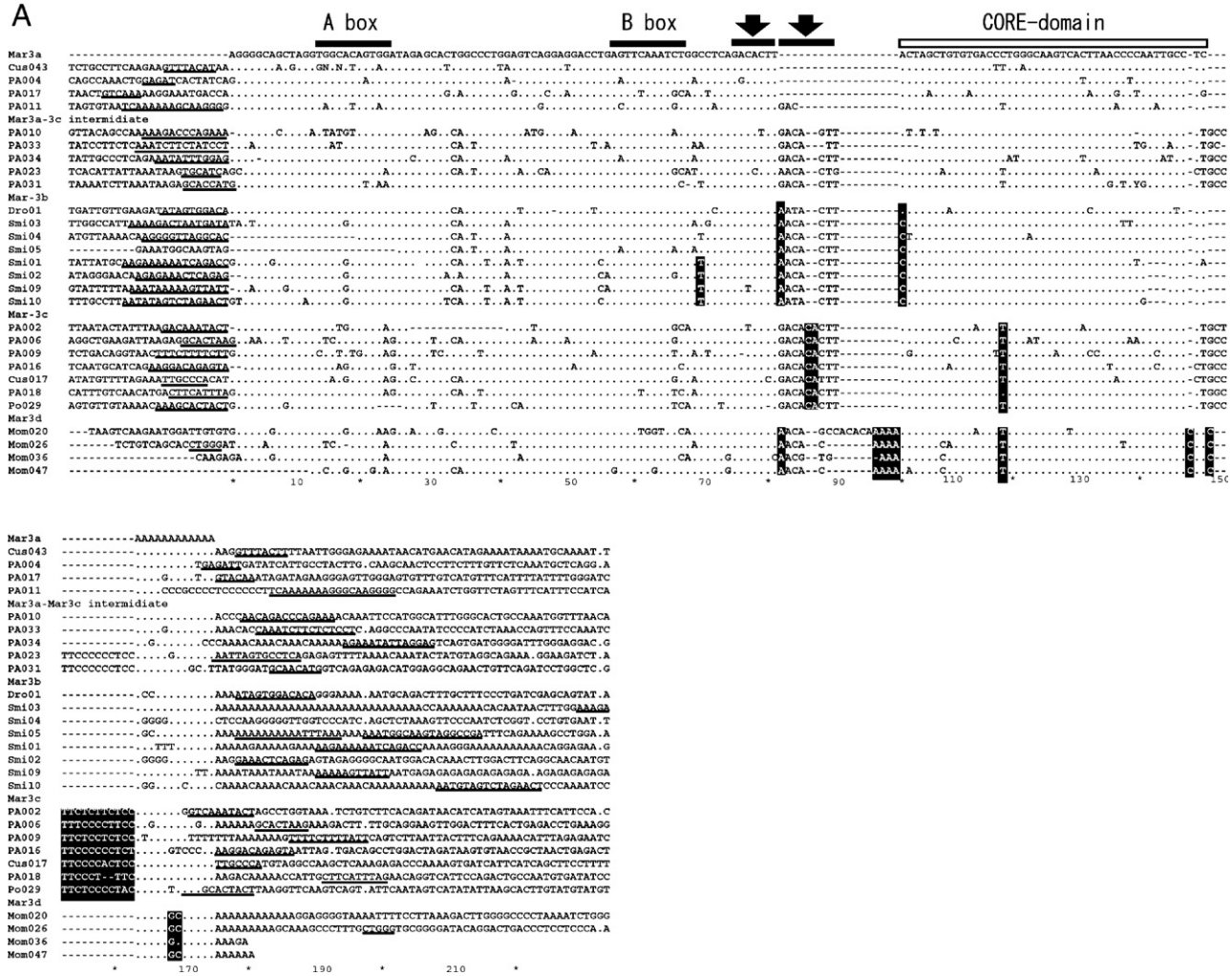

Fig. 1. Sequence alignments of the newly identified SINE subfamilies. (A) The Mar3 subfamily is subdivided into Mar3a, b, c, and d. (B) The Mac1 subfamily. The dots indicate nucleotides identical to the consensus sequence at the top. The A box and B box, which are typical for the tRNA region of each SINE are shown by thick bar. The diagnostic nucleotides for each subfamily are shaded in black. The insertions immediately upstream of all Mar3 CORE domains caused by the duplication of the 3′ end of the tRNA-related region are indicated by arrows. Underlined nucleotides indicate the target site duplications of each SINE loci.

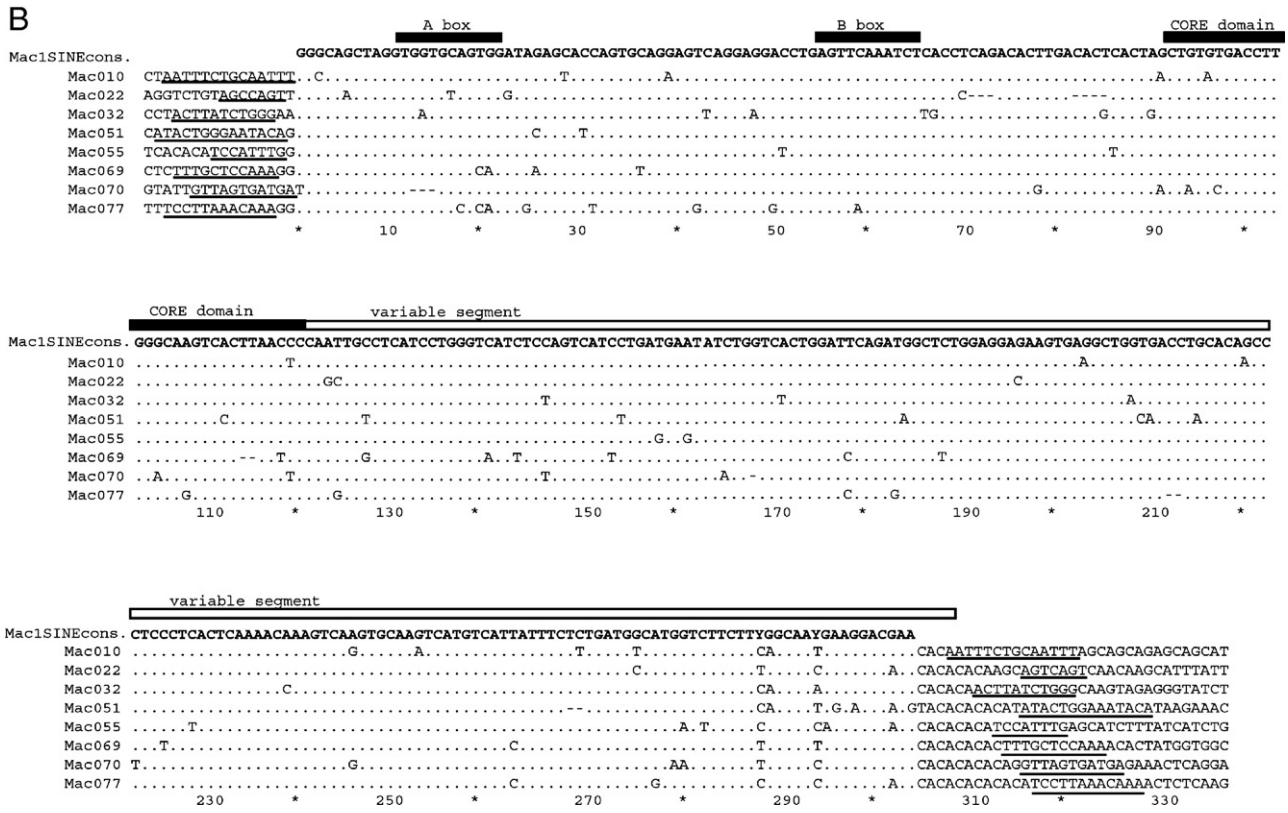

Fig. 1 (*continued*).

more recent than those of Ther1 and Ther2. Mon1 is specifically dispersed in the genomes of monotremes and further subdivided into three subfamilies (Mon1a, Mon1b and Mon1c). Mar1 is distributed in the genomes of all living marsupials. In contrast, Opo1 has been shown to be distributed in the genomes of American marsupial (Virginia opossum: *Didelphis virginiana*, and Gray Short-tailed Opossum: *M. domestica*), and absent in Australian marsupial genomes. The Opo-1 of *M. domestica* is deposited in RepBase as SINE_1 MD.

In the present study, based on genomic library screening, we found two CORE-SINEs that are closely related to Mar1. Exploration of the genome database for these SINEs, together with PCR and genomic dot blot analyses revealed that they are newly recognized CORE-SINEs, Mar3 and Mac1, which are distinct, but closely related to the members of Mar1. Furthermore, detailed sequence alignment and nucleotide diversity estimation of the CORE-SINEs characterized to date have enabled us to provide a nearly complete picture of the evolution of these two new SINEs. Descriptions of repetitive sequences such as Mar3 and Mac1 may be very useful in understanding the genetic components of non-eutherian mammals and for advancing comparative genome analysis between broad mammalian taxa.

2. Materials and methods

2.1. DNA samples

The DNA samples used in the study were from the following genomes: Red-necked wallaby (*Macropus rufogriseus*), swamp wallaby (*Wallabia bicolor*), western hare wallaby (*Lagorchestes hirsutus*), northern Nailtail wallaby (*Onychogalea unguifera*), quokka (*Setonix brachyurus*), dusky pademelon (*Thylogale brunii*), little rock wallaby (*Peradorcus concinna*), yellow-footed rock wallaby (*Petrogale xanthopus*), Goodfellow's tree kangaroo (*Dendrolagus goodfellowi*), greater forest wallaby (*Dorcopsis hageni*), brush-tailed rat kangaroo (*Bettongia pencillata*), long-nosed potoroo (*Potorous tridactylus*), musky rat kangaroo (*Hypsiprimnodon moschatus*), spotted cuscus (*Phalanger maclatus*), Herbert River ringtail possum (*Pseudochirulus herbertensis*), sugar glider (*Petaurus breviceps*), mountain pygmy possum (*Burramys parvus*), feathertail glider (*Acrobates pygmaeus*), koala (*Phascolarctos cinereus*), fat-tailed antechinus (*Antechinus agilis*), tiger quoll (*Dasyurus maclatus*), brindled bandicoot (*Isoodon macrourus*), Virginia opossum (*D. virginiana*), domestic pig (*Sus domesticus*), horse (*Equus caballus*), house mouse (*Mus musculus*), and chicken (*Gallus gallus*). Total genomic DNAs were isolated by phenol and chloroform extraction and ethanol precipitation (Blin and Stafford, 1976) and stored at 4 °C.

2.2. Screening of genomic libraries to identify new SINEs

We used the *P. breviceps*, *P. maclatus*, and *M. rufogriseus* genomes as representatives of marsupials for screening CORE-SINEs. Total genomic DNAs were first digested with *Hind*III restriction endonuclease. Digested DNA fragments were fractionated by ultracentrifugation through sucrose density gradients (10–40% w/v), and the optimal size fractions (about 2 kbp) were ascertained by agarose-gel electrophoresis.

Genomic libraries were constructed by ligation of the DNA fragments into the plasmid vector pUC18. DNA probes used for screening the genomic libraries were prepared by PCR with the primer set Mar1For: 5′-AGCTAGGTGGCGCAGTGGA-3′ and Mar1Rev: 5′-AGTCGGACACGACTGAAACG-3′ using the genome of the *M. rufogriseus* as a template. The PCR products were labeled internally with [α-^{32}P]dCTP by the primer extension method. The labeled DNA probes were purified using QIAquick PCR Purification Kit (QIAGEN) and used for subsequent screening. Southern hybridization was performed at 50 °C overnight in a solution of 6× SSC, 1% SDS, 2× Denhardt's solution, and 100 mg/ml herring sperm DNA and washed at 50 °C for 50 min in a solution of 2× SSC and 1% SDS (Sambrook, Fritsch, and Maniatis 1989). The positively hybridized clones that appeared to contain CORE-SINE sequence were isolated and purified. Purified plasmids were sequenced using the BigDye terminator cycle sequencing kit (Applied Biosystems) in the forward direction using the primer M4: 5′-GTTTTCCCAGTCACGAC-3′ and in the reverse direction using the primer RV: 5′-CAGGAAACAGCTATGAC-3′ using an automated sequencer (ABI3100, Applied Biosystems).

2.3. GenBank searches

The CORE-SINE sequences identified by the genomic library screen above were applied to the FASTA program with default parameters in order to obtain additional new CORE-SINE sequences. These CORE-SINE sequences were complied and aligned using Clustal X (Thompson et al., 1997) and a consensus sequence was deduced from the alignment. This resulted in the establishment of the Mar1, Mar3 and Mac1 consensus SINE sequences. We also collected 50 sequences of Opo1 SINE, whose 3′ tail was not clarified in the previous study by Gilbert and Labuda (2000). The alignment of these 50 Opo1 SINE sequences enabled us to clarify the 3′ tail.

2.4. Dot blot analysis

Genomic DNA of *G. gallus*, *M. musculus*, *E. caballus*, *S. domesticus*, *D. virginiana*, *D. maclatus*, *I. macrourus*, *A. agilis*, *P. cinereus*, *P. breviceps*, *B. parvus*, *A. pygmaeus*, *P. maclatus*, *H. moschatus*, *P. tridactylus* and *M. rufogriseus* were spotted onto a hybridization transfer membrane (GeneScreen Plus, Perkin Elmer) using a dot-blot apparatus (model DP-96,

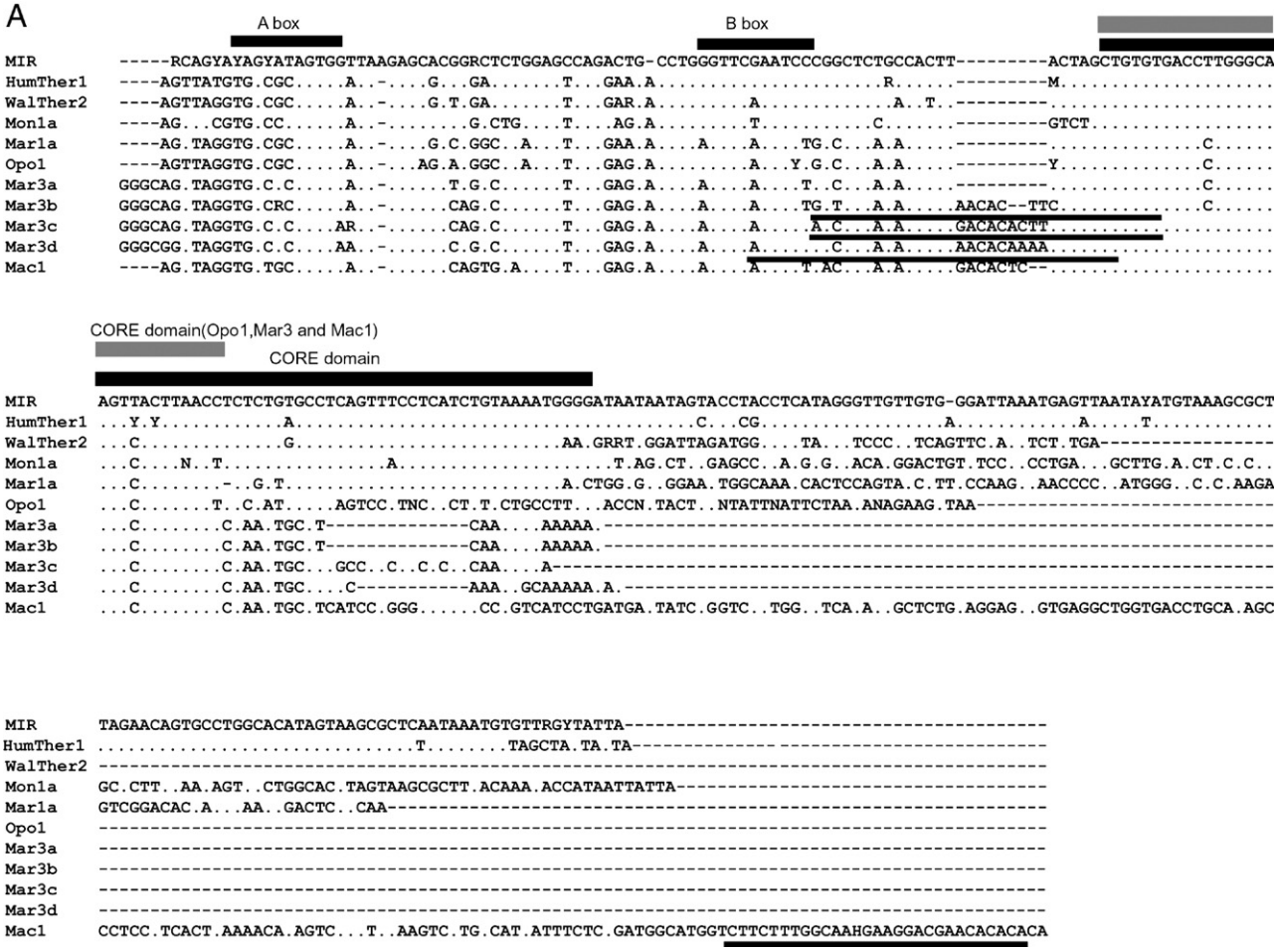

Fig. 2. Comparison of mammalian CORE-SINEs. (A) Alignment of the consensus sequences of mammalian CORE-SINEs. The thick underlines indicate the annealing position of the oligonucleotide probes used in the genomic dot blot hybridization analysis. The shorter CORE domain, characteristic of the newly identified SINEs and Opo1, is indicated by the gray bar. (B) Schematic representation comparing the structure of the CORE-SINEs characterized to date. The patterns within the boxes indicate distinct diagnostic elements for each SINE.

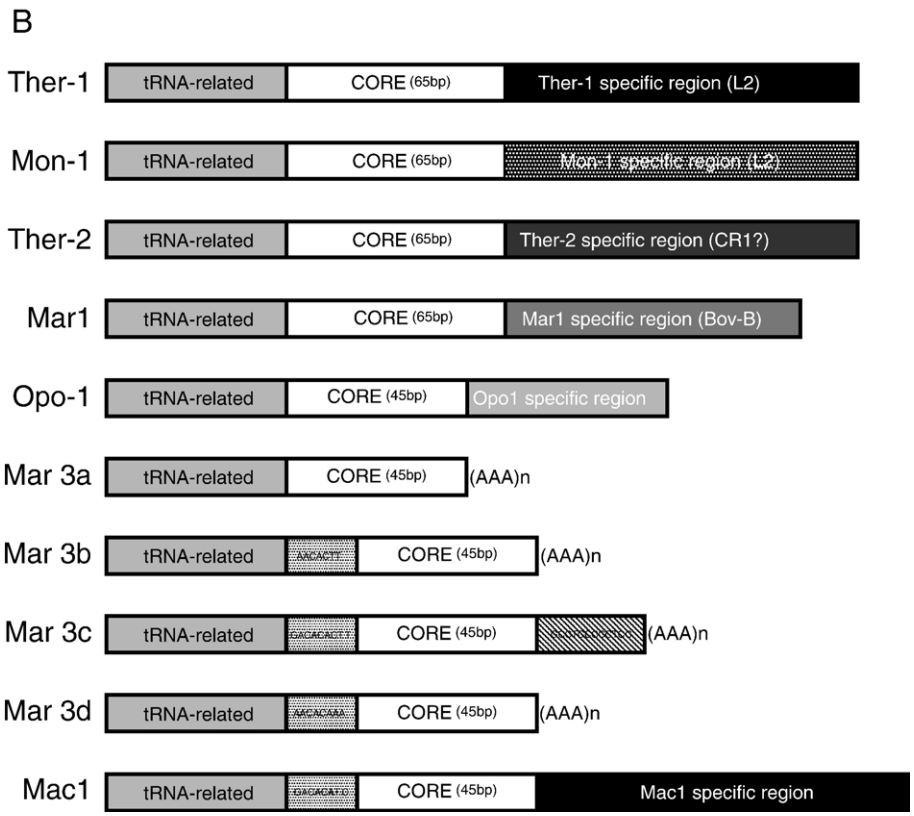

Fig. 2 (*continued*).

Advantech). The following oligonucleotide probes were designed to anneal specifically to consensus CORE-SINE sequences: Mar3b: 5′-GGTCTCAGACACTTAACACTTCCTAGC TGTGT-3′, Mar3c: 5′-AGCCTCAGACACTTGACA-CACTTACTAGCTGTGT-3′, Mar3d: 5′-AAATCCCGCCT-CAGACACTTAACACAAAAACTAGCT-3′, Mac1: 5′-CCC CAATTGCCTCATCCTGGGTCATCTCCAGTCATCC-3′ (Fig. 2A). The oligonucleotide probes were end-labeled with [γ-^{32}P] ATP. Labeled probes were purified using the QIAquick Nucleotide Removal Kit (QIAGEN). The plasmid DNAs containing consensus Mar3 and Mac1 sequence were used as positive controls. Southern hybridization was performed at 55 °C overnight, and washing was performed two times at 60 °C for 50 minutes.

3. Results

3.1. Newly isolated CORE-SINEs in Australian marsupials

The CORE-SINEs isolated through screening of *P. breviceps*, *P. maclatus*, and *M. rufogriseus* genomic libraries consisted of Mar1, Ther1 and uncharacterized SINEs. These uncharacterized SINEs were further divided into two subgroups. The first subgroup consists of SINEs with a 5′ tRNA-derived region, a CORE-domain, and a 3′ poly A tail (Fig. 1A). The second subgroup is distinguished by a unique 3′ tail region containing a long $(CA)_n$ terminal repeat (Fig. 1B), which is specific to this subgroup (Fig. 1B). Although typical CORE-SINEs possess a highly conserved 65 bp CORE domain, the COREs of the newly isolated SINEs are 20 bp shorter in length. This shorter CORE is also detected in Opo1, implying the same ancestry of these new SINEs and Opo1 (Fig. 2A, B). The new SINEs were classified as either Mar3 family (*Marsupialia* 3) or Mac1 family (*Macropodoidea* 1) SINES to indicate the distribution of the SINE families in the genomes of all marsupials or of macropods (the members of the superfamily *Macropodoidea*), respectively. Recently, “MAR2_MD” SINE has been deposited in RepBase, but this sequence is just a short fragment of Bov-B LINE (RTE-1) and has no relationship with CORE-SINEs. To avoid confusion, we do not use the name “Mar2” but use “Mar3” for the newly characterized CORE-SINEs.

3.2. Characterization of the Mar3 SINE

Fig. 1 shows the sequence alignment of Mar3 and Mac1. The length of Mar3 varies from 116 to 144 bp without the poly A tail. Variable regions are located at both ends of the CORE domain. Most of the members of Mar3 have a diagnostic 7 bp insertion immediately upstream of the CORE domain (Fig. 2B). This diagnostic insertion is derived from duplication of the 3′ end of the tRNA-related region (Fig. 1A, indicated by arrows). The members of Mar3 are further subdivided into four subfamilies. We named these SINE subfamilies Mar3a, Mar3b, Mar3c and Mar3d (Figs. 1 and 2). Mar3a is the only subfamily lacking the diagnostic insertion upstream of the CORE domain, implying that this group is the most primitive Mar3 SINE. The Mar3c subfamily also has a 5′-GACACACTT-

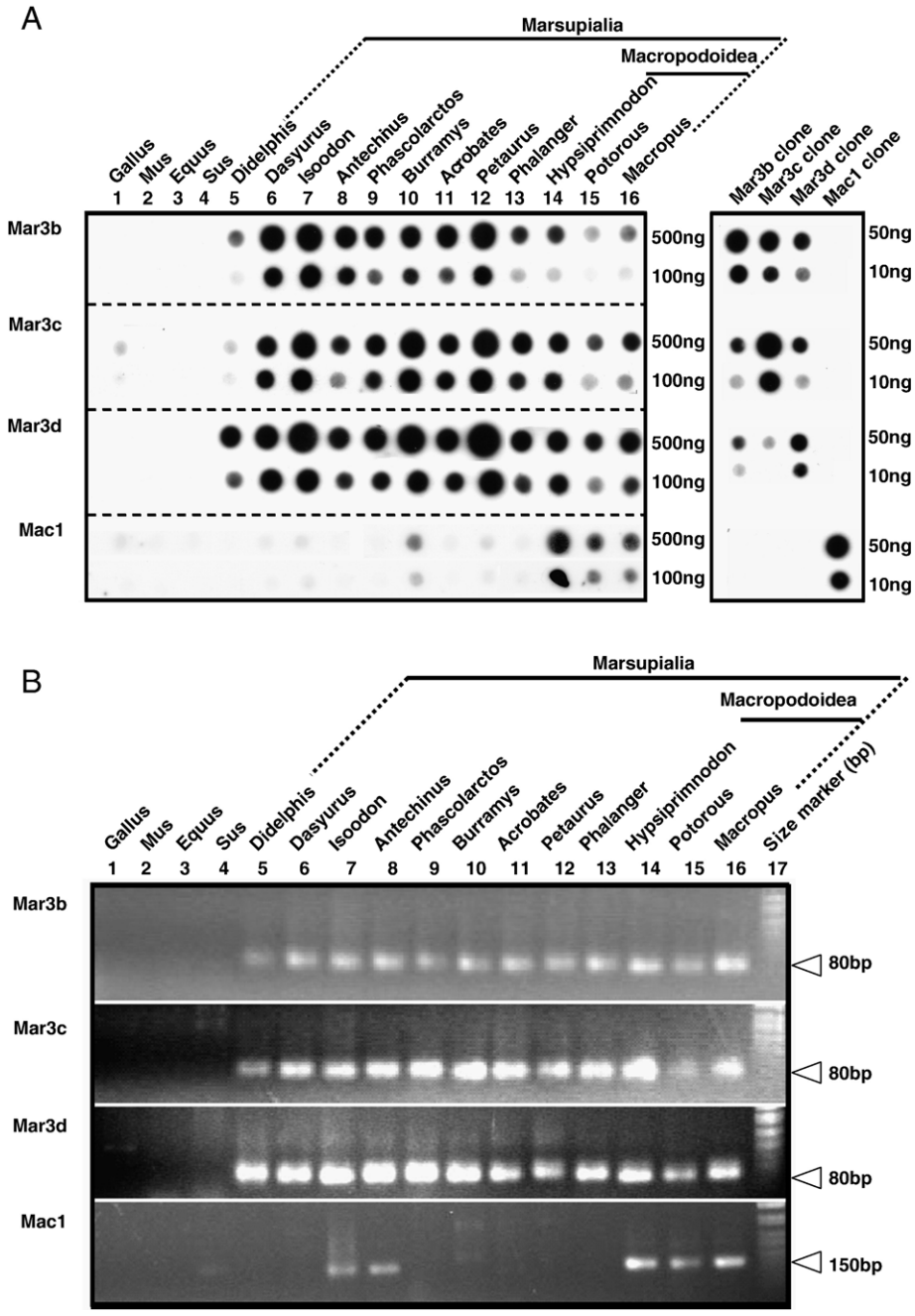

Fig. 3. The distribution of Mar3 and Mac1 among the genomes of various vertebrate taxa. (A) Dot blot analysis on marsupial genomic DNA (left panel) and control plasmid DNA (right panel). The radiolabeled probe used for each sample is indicated at left. The amount of input DNA is indicated to the right of each panel. (B) PCR analysis of marsupial genomic DNA to confirm dot blot data. PCR products from reactions containing 10 ng genomic DNA and 5 pmol of the PCR primers indicated at left (see Supp. Table 1) were subjected to 3% agarose gel electrophoresis and stained with ethidium bromide. DNA size markers (bp) are indicated at right.

3′ insertion immediately downstream of the CORE domain. Several intermediates between Mar3a and Mar3c were obtained throughout the screening (Fig. 1A), which suggests that Mar3c was derived from Mar3a. Because we also obtained such intermediates when comparing Mar3a with Mar3b and Mar3d, the latter two members also might be derived from Mar3a. All members of these subfamilies have 3′ poly A tails that are 4 to 30 bp in length. The presence of the poly A tail in the members of the Mar3 family implies that, like Alu in the human genome, they use L1 proteins for retroposition. After completion of our work, Gentles et al. (2007) reported the presence of a SINE family named SINE2_MD in the genome of short-tailed opossum. This SINE2_MD share obvious sequence homology with Mar3 SINE, implying the common ancestry among them.

3.3. Characterization of the Mac1 SINE

The consensus sequence of Mac1 is 310 bp in length. Mac1 is composed of a 5′ tRNA-related region, a shorter CORE domain with a diagnostic upstream insertion and a 3′ tail region

that is unique to Mac1. This SINE family, the sequences of which can be found in GenBank, was also found in the genomes of Tammar wallaby (*Macropus eugenii*) and swamp wallaby (*W. bicolor*). The 3′ tail of Mac1 is about 250 bp and ends with $(CA)_n$ repeats. The sequence divergence among the members of Mac1 is very small, showing that this SINE family seems to be very young.

3.4. The distribution of new SINEs

We investigated the distribution pattern of Mar3 and Mac1 among vertebrates using oligonucleotide probes specific for Mar3b, Mar3c, Mar3d and Mac1. Fig. 3A shows the dot blot pattern generated by Southern hybridization using each probe against the genomic DNA of various vertebrates. It was very difficult to design probes to completely distinguish between each Mar3 subfamily because the nucleotide differences between these subfamilies are very small. Indeed, although we detected cross hybridization of particular probes to several negative control plasmids, highly dense signals were clearly detected in the case of the correct combination of probes and plasmids. The dot blot patterns showed the existence of the Mar3b, Mar3c and Mar3d subfamilies in all marsupial genomes. However, because of the possibility of false positives due to dot blot cross-hybridization, PCR analysis was used to confirm this conclusion (Fig. 3B). We sequenced the PCR products for each marsupial and confirmed the presence of the Mar3b, Mar3c and Mar3d subfamily SINEs among their genomes. Although, a very weak signal was detected in the mouse genome with the Mar3c probe (Fig. 3A), it was likely an artifact because no Mar3c sequence was found in the mouse genome database and no Mar3c-specific PCR product was generated from the mouse genome (Fig. 3B). Although dot blot signals for the Mar3 subfamilies' probes were weak in the *D. virginiana*, PCR products were successfully obtained. This data suggests that Mar3 is present in the genome of the *D. virginiana*, but the copy number is low and/or the sequence divergence is large, which results in less effectiveness of the probe.

Mac1 showed only a limited distribution within the genomes of the superfamily *Macropodoidea*, which was confirmed by the PCR analysis (Figs. 3A, B). Although weak signal was obtained in *B. parvus*, no obvious PCR band was obtained. On the contrary, faint PCR products were detected in the genomes of *I. macrourus* and *A. agilis*, whereas dot blot signals were not observed. We cloned and sequenced the PCR products of these two species, and found they were non-specific bands, which might be due to the miss annealing of the PCR primers. Therefore, we can conclude that Mac1 is distributed specifically in the genomes of the superfamily *Macropodoidea*.

3.5. Characterization of particular SINE loci

Supplementary Figures 1A and 1B show the insertion patterns of two Mar3 loci (PA031 and PA023) and one Mac1 locus (MS010). Insertion of Mar3 at the PA023 locus was detected in both *P. breviceps* (*Petauridae*) and *P. herbertensis* (*Pseudocheiridae*), implying a close relationship between them. At the locus PA031, the Mar3 insertion was detected in all members of *Macropodoidea*, confirming the monophyly of this superfamily. At many of the loci represented by MS010 (MS010, MS022, MS032, MS045 and MS055), the insertion of Mac1 was detected in *P. tridactylus*, *B. pencillata*, *D. hageni*, *D. goodfellowi*, *P. xanthopus*, *P. concinna*, *T. brunii*, *S. brachyurus*, *O. unguifera*, *L. hirsutus*, *W. bicolor*, and *M. rufogriseus*, but not in *H. moschatus*. These data indicate that *Hypsiprimnodon* is the most basal genus among the kangaroos used in the present analysis. This idea is consistent with the result of Burk et al. (1998).

4. Discussion

4.1. The relationships of CORE-SINE families

We discovered two novel CORE-SINE families and clarified the distribution patterns of these families among marsupials. Dot blot and PCR analyses indicated that members of the Mar3 family are present in all marsupial genomes. CORE-SINE families typically possess a 65 bp CORE domain, which is highly conserved in Ther1, Ther2, Mon1 and Mar1. However, the CORE domain of these newly identified SINE families is 20 bp shorter than the canonical CORE domain (Fig. 1). The shorter CORE domain is commonly observed in several CORE-SINEs, including the Opo1 family. The presence of CORE-SINEs possessing this shorter CORE region implies that 20 bp of the 3′ terminal region of the CORE domain lacked importance for retroposition during the evolution of marsupials.

The Mar3 family consists of at least four subfamilies. Based on the sequence divergence, the oldest and most primitive subfamily is Mar3a. This subfamily has no diagnostic insertion immediately upstream of the CORE domain, whereas Mar3b, Mar3c, and Mar3d share the insertion element. Therefore, we speculated that these three subfamilies are derived from Mar3a. The presence of several intermediate sequences between Mar3a and Mar3c enabled us to describe the transition from Mar3a to Mar3c (Fig. 1A). Mar3b and Mar3d might also be derived from Mar3a; namely, one of the intermediate sequences between Mar3a and Mar3c could be the origin of these two subfamilies (Fig. 4). The insertion in the upstream region of the CORE domain, which is diagnostic of Mar3b, c and d, is also observed in Mac1 (Fig. 2A). Furthermore, Mac1 has the short CORE domain, like all members of the Mar3 family. These two lines of evidence suggest that Mac1 originated from one of the members of Mar3 and proliferated in the genome of the common ancestor of the superfamily *Macropodoidea* (Fig. 4).

4.2. The distribution of CORE-SINE families

Fig. 5 illustrates the distribution of each CORE-SINE in Marsupials. Mar1 is a marsupial-specific SINE (Gilbert and Labuda, 2000) present in both American and Australian marsupials. By using the database search, we found several orthologous loci, where Mar1 SINE is commonly present in *D. virginiana* and *M. rufogriseus* (data not shown), which indicates that this SINE was amplified before the ancestral split of

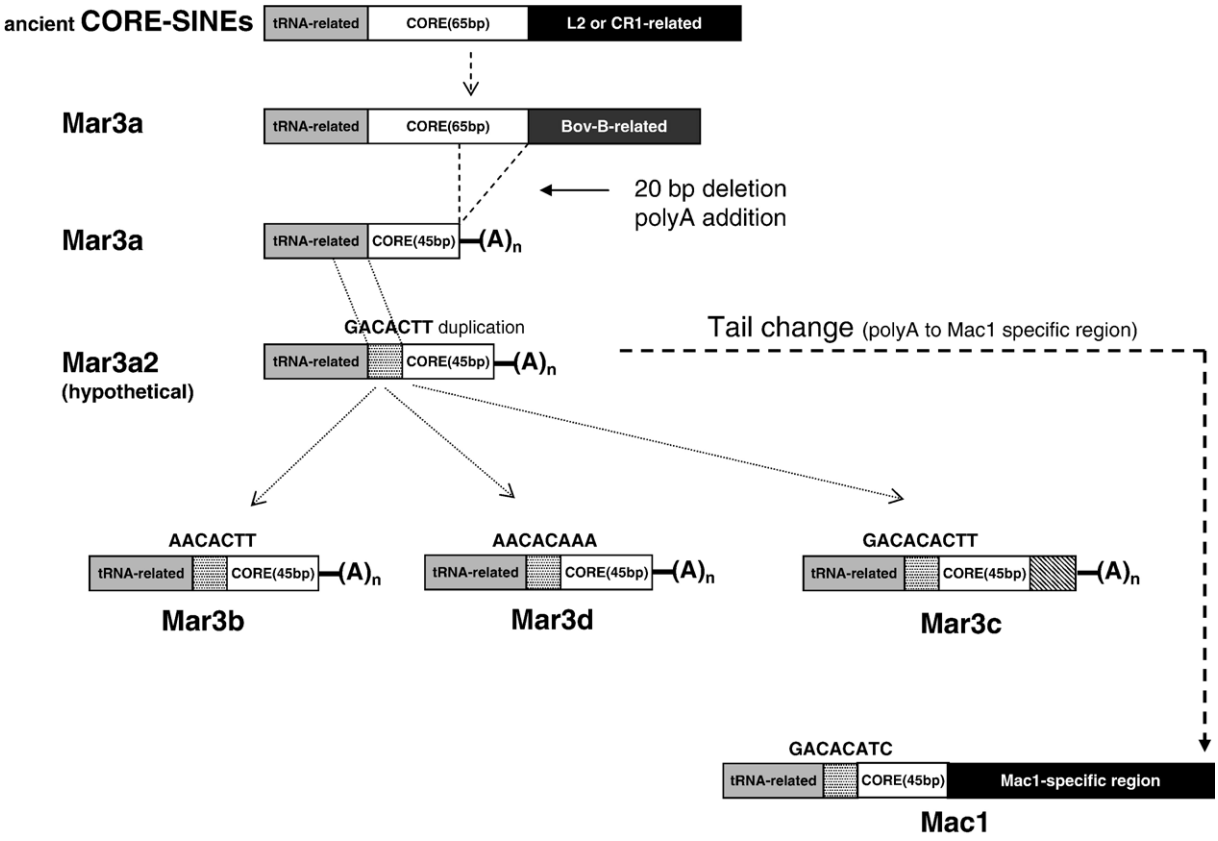

Fig. 4. Illustration of the evolutionary relationships among CORE-SINEs.

American and Australian marsupials and suggests a relatively old age for this SINE group. According to the previous study by Gilbert and Labuda (2000), Opo1 shows a distribution specific to American marsupials. Our results indicate that all members of Mar3 are distributed among marsupials. However, Southern blot analysis suggests that Mar3 occurs at a lower frequency in the American marsupial (represented by *D. virginiana*) genome than in Australian marsupials. Furthermore, several insertions of Mar3c are specific to particular lineages of Australian marsupials. In addition, average nucleotide diversities of Mar3c, and Mar3d are smaller than that of Mar1 (Suppl. Table 2). These three lines of evidence suggest that the proliferation of Mar3 is more recent than Mar1.

Mac1 is present in the genomes of all members of *Macropodoidea* (Figs. 3A, B). However our SINE insertion analyses of each particular locus revealed that Mac1 was inserted in all members of *Macropodoidea* (macropods) except *H. moschatus* (Suppl. Figs. 1B and 2). Therefore, majority of the members of Mac1 were likely amplified soon after the divergence of *H. moschatus* from other macropods, which is estimated to have taken place approximately 25 mya based on the mitochondrial DNA sequences analysis (Burk and Springer, 2000).

4.3. Phylogenetic impact of SINE insertions

The phylogenic relationships of marsupials are still enigmatic at the order and family level, particularly among Australian marsupials (Osborne et al., 2002; Nilsson et al., 2004). To resolve these problems, diagnostic SINE insertions provide useful information (Nikaido et al., 1999; Shedlock and Okada, 2000; Nishihara et al., 2005; Nishihara et al., 2006b). In the present study, two phylogenetically informative insertions of Mar3 were detected. At the locus PA023, Mar3 is common to *Petauridae* and *Pseudocheiridae*, but is not found in other diprotodontian marsupials, implying that this SINE was inserted

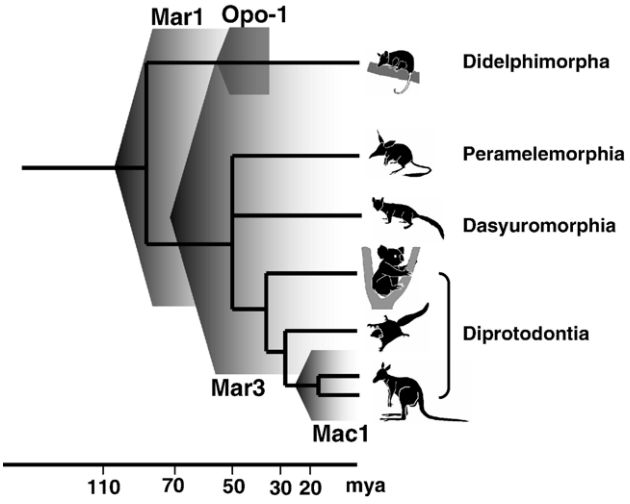

Fig. 5. The summary of CORE-SINE evolution. The presence of active SINEs is indicated by shaded regions. The intensity of shading is relative to the speculated retropositional activity of each SINE inferred from the distribution and the sequence divergence calculations of these SINEs. The time scale is shown below.

in a common ancestor of these two families (Suppl. Figs. 1, 2). That is, PA023 suggests the monophyly of *Petauridae* and *Pseudocheiridae*, which are grouped in the superfamily *Petauroidea* based on morphological classification (Aplin and Archer, 1987). The PA031 locus supports the monophyly of *Potoroidae* and *Macropodidae*, which are grouped in the superfamily *Macropodoidea* based on morphological study (Gray, 1821) (Supp. Figs. 1, 2). Our molecular data also support the monophyletic relationships of the superfamilies *Petauroidea* and *Macropodoidea* established by nuclear and mitochondrial sequence data (Amrine-Madsen et al., 2003; Burk et al., 1998; Burk and Springer, 2000; Osborne et al., 2002). The consistency of the phylogenetic trees constructed by SINE insertions, morphological classifications or several DNA sequence comparison analyses clearly show the reliability of SINE insertions as phylogenetic markers. In the case of all Mac1 loci isolated in the present study, SINE insertions were detected for all members of *Macropodoidea* except for *H. moschatus*. This SINE data suggests that *Potoroidae* are non-monophyletic, which has also been suggested by other molecular data (Burk et al., 1998). The combination of our data with that of Burk et al. (1998) suggests a need for taxonomic revision for *H. moschatus*. Thus, these newly characterized Mar3 and Mac1 SINEs might be very important for further understanding of marsupial phylogeny, particularly regarding diprotodontians. Especially, Mar3 SINEs will provide reliable information for the inter-familial phylogeny, and Mac1 SINE for intra-*Macropodoidea* phylogeny.

In this study, we identified and characterized two new CORE-SINEs in marsupial genomes and showed that they are relatively young SINEs compared with Ther1, Ther2 and Mar1, which have previously been characterized (Gilbert and Labuda, 1999, 2000). The distribution, sequence divergence and structure of each CORE-SINE revealed a nearly complete picture of the evolution of these SINEs, suggesting their promise as a phylogenetic tool. Our SINE data will provide the foundation for further understanding of the genomes and evolution of marsupials, whose complete genome sequences are now rapidly accumulating in the database.

Acknowledgments

We thank Drs. Ken Aplin, Yoko Satta and the Kanazawa zoo for various marsupial samples. This work was supported by a Grant-in-Aid to N. O. from the Ministry of Education, Science, Sports and Culture of Japan.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2007.10.008.

References

Amrine-Madsen, H., Scally, M., Westerman, M., Stanhope, M.J., Krajewski, C., Springer, M.S., 2003. Nuclear gene sequences provide evidence for the monophyly of australidelphian marsupials. Mol. Phylogenet. Evol. 28, 186–196.

Aplin, K.P., Archer, M., 1987. Recent advances in marsupial systematics with a new syncretic classification. In: Archer, M. (Ed.), Possums and Opossums. Studies in Evolution. Surrey Beatty and Sons, New South Wales, pp. 15–72.

Bejerano, G., et al., 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441, 87–90.

Blin, N., Stafford, D.W., 1976. A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res. 3, 2303–2308.

Brosius, J., 1991. Retroposons-seeds of evolution. Science 251, 753.

Burk, A., Springer, M.S., 2000. Intergeneric relationships among Macropodoidea (Metatheria:Diprotodontia) and the chronicle of kangaroo evolution. J. Mol. Evol. 7, 213–237.

Burk, A., Westerman, M., Springer, M.S., 1998. The phylogenetic position of the musky rat–kangaroo and the evolution of bipedal hopping in kangaroos (Macropodidae: Diprotodontia). Syst. Biol. 47, 457–474.

Gentles, A.J., et al., 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. Genome Res. 17, 992–1004.

Gilbert, N., Labuda, D., 1999. CORE-SINEs: Eukaryotic short interspersed retroposing elements with common sequence motifs. Proc. Natl. Acad. Sci. U. S. A. 96, 2869–2874.

Gilbert, N., Labuda, D., 2000. Evolutionary inventions and continuity of CORE-SINEs in mammals. J. Mol. Biol. 298, 365–377.

Gray, J.E., 1821. On the natural arrangement of vertebose animals. Lond. Med. Redeposit. 15, 296–310.

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–920.

Jurka, J., Zietkiewicz, E., Labuda, D., 1995. Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. Nucleic Acids Res. 23, 170–175.

Karolchik, D., et al., 2003. The UCSC Genome Browser Database. Nucleic Acids Res. 31, 51–54.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Biol. 16, 111–120.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Margulies, E., et al., 2005. Comparative sequencing provides insights about the structure and conservation of marsupial and monotreme genomes. Proc. Natl. Acad. Sci. U. S. A. 102, 3354–3359.

Mikkelsen, T.S., et al., 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. Nature 447, 167–177.

Munemasa, M., Nikaido, M., Donnellan, S., Austin, C.C., Okada, N., Hasegawa, M., 2006. Phylogenetic analysis of diprotodontian marsupials based on complete mitochondrial genomes. Genes Genet. Syst. 81, 181–191.

Nikaido, M., Rooney, A.P., 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. U. S. A. 96, 10261–10266.

Nilsson, M.A., Arnason, U., Spencer, P.B.S., Janke, A., 2004. Marsupial relationships and a timeline for marsupial radiation in South Gondwana. Gene 340, 189–196.

Nishihara, H., Satta, Y., Nikaido, M., Thewissen, J.G.M., Stanhope, M.J., Okada, N., 2005. A retroposon analysis of Afrotherian phylogeny. Mol. Biol. Evol. 22, 1823–1833.

Nishihara, H., Smit, A.F.A., Okada, N., 2006a. Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res. 16, 864–874.

Nishihara, H., Hasegawa, M., Okada, N., 2006b. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. Proc. Natl. Acad. Sci. U. S. A. 103, 9929–9934.

Ogiwara, I., Miya, M., Ohsima, K., Okada, N., 2002. V-SINE: A new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. Genome Res. 12, 316–324.

Okada, N., 1991a. SINEs. Curr. Opin. Genet. Dev. 1, 498–504.

Okada, N., 1991b. SINEs: short interspersed repeated elements of the eukaryotic genome. Trends Ecol. Evol. 6, 358–361.

Osborne, M.J., Christidis, L., Norman, J.A., 2002. Molecular phylogenetics of the Diprotodontia (kangaroos, wombats, koala, possums, and allies). Mol. Phylogenet. Evol. 25, 219–228.

Rogers, J.H., 1985. The origin and evolution of retroposons. Int. Rev. Cyt. 93, 187–279.

Shedlock, A.M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. BioEssays 22, 148–160.

Smit, A.F.A., Riggs, A.D., 1995. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. Nucleic Acids Res. 23, 98–102.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Weiner, A.M., Deininger, P.L., Efstratiadis, A., 1986. Nonviral retroposons: Genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. Allied Res. India 55, 631–661.