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

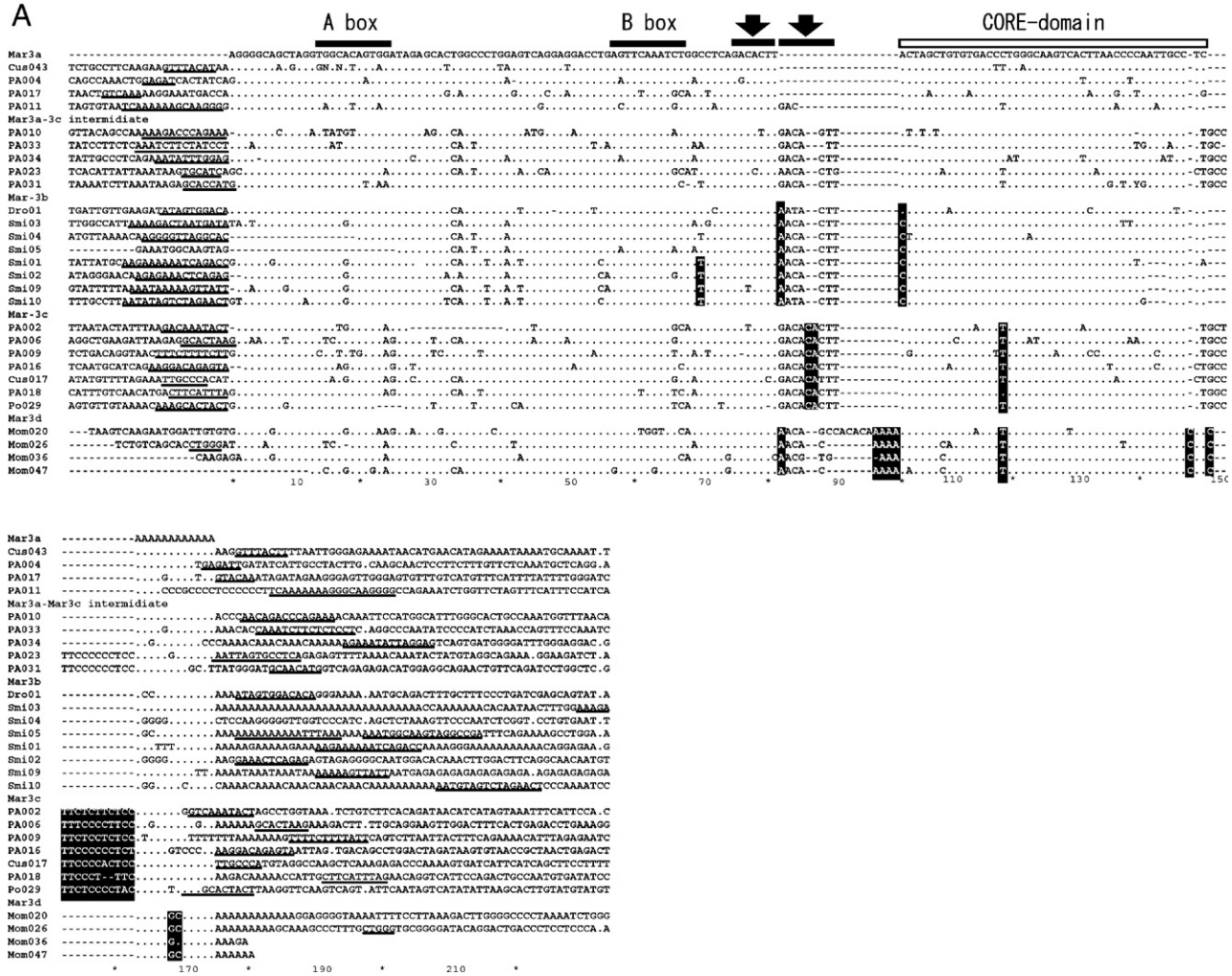

Fig. 1. Sequence alignments of the newly identified SINE subfamilies. (A) The Mar3 subfamily is subdivided into Mar3a, b, c, and d. (B) The Mac1 subfamily. The dots indicate nucleotides identical to the consensus sequence at the top. The A box and B box, which are typical for the tRNA region of each 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.