Mol. Cells *30*, 201-207, September 30, 2010
DOI/10.1007/s10059-010-0108-x

Molecules and Cells

Alu-Derived Old World Monkeys Exonization Event and Experimental Validation of the *LEPR* Gene

Jae-Won Huh[1,4], Young-Hyun Kim[1,2,4], Dae-Soo Kim[1,4], Sang-Je Park[1,3], Sang-Rae Lee[1], Sang-Hyun Kim[1], Ekyune Kim[1], Sun-Uk Kim[1], Myeong-Su Kim[1], Heui-Soo Kim[3], and Kyu-Tae Chang[1,*]

The leptin receptor (*LEPR*) is a crucial regulatory protein that interacts with Leptin. In our analysis of *LEPR*, novel *Alu*Jb-derived alternative transcripts were identified in the genome of the rhesus monkey. In order to investigate the occurrence of *Alu*Jb-derived alternative transcripts and the mechanism underlying exonization events, we conducted analyses using a number of primate genomic DNAs and adipose RNAs of tissue and primary cells derived from the crab-eating monkey. Our results demonstrate that the *Alu*Jb element has been integrated into our common ancestor genome prior to the divergence of simians and prosimians. The lineage-specific exonization event of the *LEPR* gene in chimpanzees, orangutans, and Old World monkeys appear to have been accomplished via transition mutations of the 5′ splicing site (second position of C to T). However, in New World monkeys and prosimians, the *Alu*Jb-related *LEPR* transcript should be silenced by the additional transversion mutation (fourth position of T to G). The *Alu*Jb-related transcript of human *LEPR* should also be silenced by a mutation of the 5′ splicing site (first position of G to A) and the insertion of one nucleotide sequence (minus fourth position of A). Our data suggests that lineage-specific exonization events should be determined by the combination event of the formation of splicing sites and protection against site-specific mutation pressures. These evolutionary mechanisms could be major sources for primate diversification.

INTRODUCTION

During mammalian evolution, a variety of transposable elements (TEs) occurred as the consequences of lineage-specific integration and numerous amplification events in different genomes, including those of humans, chimpanzees, rhesus monkeys, and mice (Costas, 2003; Han et al., 2007; International Human Genome Sequencing Consortium, 2001; The Chimpanzee Sequencing and Analysis Consortium, 2005). Approximately 50% of genome components were found to be composed of different TEs. Among them, a few retroelements have been identified as lineage-specific elements in humans (SVA, L1HS, LTR5_HS), chimpanzees (PtERV1 and PtERV2), and rhesus monkeys (MacERV and *Alu*YRb) (International Human Genome Sequencing Consortium, 2001; The Chimpanzee Sequencing and Analysis Consortium, 2005; Han et al., 2007). Thus, an investigation of lineage-specific integration events in relation with TEs may prove intriguing. However, the lineage-specific regions comprised of TEs rarely provide a direct link with functional components (genes or gene-related regions). Until now, only a small number of TE elements have been identified as functional elements on the basis of coding sequences, enhancers, and promoters (Huh et al., 2006; 2008; Piriyapongsa et al., 2007; Sverdlov, 1998; van de Lagemaat et al., 2003).

Alu elements are a primate-specific family of short interspersed elements (SINEs), which evolved from 7SL RNA ~65 million years ago; the human genome harbors > 1.1 million copies (11%) of these elements (Batzer and Deininger, 2002). Although a number of disease-related *Alu* elements mediated by novel integration and the following truncation events of functional genes have been previously identified, the results of a recent investigation showed that in humans, *Alu* elements could operate as potential functional elements by providing alternative exons (Sela et al., 2007). The internal sequences of *Alu* elements harbor the potential splicing donor "GT" and the acceptor site "AG", which could be recognized by human spliceosomes (Ast, 2004; Lev-Maor et al., 2003). Thus, *Alu* sequences scattered throughout the human genome might be transcribed by the so-called "exonization events". Furthermore, alternatively spliced *Alu*-derived exons are major sources for the enhancement of protein diversification in humans via the creation of novel protein isoforms (Kreahling and Graveley, 2004; Li et al., 2001; Sorek et al., 2002).

The leptin receptor gene (*LEPR*) interacts with an adipocyte-secreted hormone (leptin) that carries out weight regulatory and energy balance functions. The mammalian *LEPR* gene has been reported to harbor a range of splice variants derived from the 3′ terminal exons associated with the cytoplasmic domains of *LEPR* genes (Fruhbeck, 2006). Additionally, the human LEPR 219.1 isoform is derived via the integration event of the

[1]National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology, Ochang 363-883, Korea, [2]Functional Genomics, University of Science and Technology, Daejeon 305-333, Korea, [3]Division of Biological Sciences, College of Natural Sciences, Pusan National University, Busan 609-735, Korea, [4]These authors contributed equally to this work.
*Correspondence: changkt@kribb.re.kr

Received January 19, 2010; revised May 20, 2010; accepted May 27, 2010; published online August 23, 2010

Keywords: alternative transcript, AluJb, exonization event, leptin receptor, transposable elements

SVA element (its main components, SINE, VNTR, and *Alu*) and its expression activity was identified in the fetal liver and CD34+ hematopoietic stem cells on the RNA level (Damert et al., 2004).

In our analysis, we focused on the identification and molecular characterization of retrotransposon-derived alternative splicing and exonization events on *LEPR* by comparing 14 primate species. Additionally, *Alu*-derived alternative transcripts were experimentally validated using crab-eating monkey adipose tissue and adipose primary cells, in an effort to trace the evolutionary mechanisms of exonization events.

MATERIALS AND METHODS

Computational screening and identification of repeat elements on *LEPR*

The RefSeq mRNA was BLAST (blastn) analyzed to identify the *LEPR* gene transcripts. RefSeq mRNA matching the *LEPR* gene was used to identify the splicing patterns, using the UCSC genome browser (http://genome.ucsc.edu/). Basically, the 98% similarity with RefSeq mRNA sequences only allowed spliced transcripts in specific lineages. The repeat elements integrated into the alternatively-spliced *LEPR* gene transcripts and the genomic sequences were detected using RepeatMasker software (http://www.repeatmasker.org/). Descriptions of specific repetitive elements were obtained from Repbase Update (Jurka, 2000). In order to reconstruct correctly the structure of the *LEPR* gene (individual exons), all identified sequences were compared with previously reported RefSeq mRNA sequences. Additionally, undesignated parts of identified sequences were used as query sequences in the UCSC genome browser (http://genome.ucsc.edu/). Finally, we could correctly reconstruct the structure (exon/intron boundary) of the *LEPR* gene.

Genomic DNA and RNA

Primate DNA (HU: *Homo Sapiens*, CH: *Pan troglodytes*, OR: *Pongo sp.*, RH: *Macaca mulatta*, CR: *Macaca fascicularis*, MA: *Mandrillus sphinx*, AF: *Chlorocebus aethiops*, CO: *Procolobus badius*, LA: *Trachypithecus sp.*, SP: *Ateles geoffroyi*, NI: *Aotus nigriceps*, TA: *Saguinus midas*, MAR: *Callithrix jacchus,* LE: *Lemur catta*) and mRNA samples from crab-eating monkey adipose tissue (one individual) and adipose tissue primary cells (two individuals) employed in this study were provided by the late Prof. Osamu Takenaka from the Primate Research Institute of Kyoto University of Japan and the National Primate Research Center (NPRC) of Korea. Pure mRNA was extracted with an RNA Mini kit mRNA isolation system (QIAGEN).

Experimental validation of *Alu*-derived exonization event via PCR, RT-PCR amplification, and sequencing procedure

In order to determine the integration lineage of the *Alu*Jb element in primates, a specific primer set was employed. The specific primer pair S1 (5′- TTT AGC CTG TTG TTT TAA TCT CCT -3′) and AS1 (5′- ACC ATC CAC CCT ATG TTT CA -3′) was designed for investigations of the *Alu*Jb element in different primate genomic DNAs. Different alternative transcripts of the *LEPR* gene were analyzed and validated via RT-PCR and a sequencing procedure. M-MLV reverse transcriptase with an annealing temperature of 42°C was utilized for the reverse transcription reaction with an RNase inhibitor (Promega). We also conducted control PCR amplification with pure mRNA samples that were not subjected to reverse transcription. No amplification of control PCR indicated that the prepared mRNA samples did not harbor genomic DNA. As a standard control, G3PDH was amplified by the primers G3PDH-S (5′- GAA ATC CCA TCA CCA TCT TCC AGG - 3′) and G3PDH-AS (5′- GAG CCC CAG CCT TCT CCA TG- 3′) from the human G3PDH gene [GenBank: AC068657]. The alternative transcript of the *LEPR* gene was amplified with the following primer pairs: 1S (5′- ATG GAA GGA GTG GGA AAA CC - 3′) and 1AS (5′ - TCC TCT TCT TTT GGA GCC TG - 3′) [GenBank: NM_001032819]. The primer locations of PCR and RT-PCR are indicated in Supplementary Fig. S1. PCR and RT-PCR were conducted under the following conditions: 5 min at 94°C followed by 35/40 cycles consisting of 1 min at 94°C, 1 min at the primer–specific annealing temperature, and 30 s at 72°C. A following final elongation step was conducted for 7 min at 72°C. The PCR products were then separated via electrophoresis on 1.5% agarose gel and visualized by ethidium bromide staining, purified with the Gel SV extraction kit (GeneAll), and cloned into pGEM-T-easy vector (Promega). The cloned DNA was isolated via alkaline lysis using the Exprep Plasmid SV plasmid isolation kit (GeneAll). The sequencing of 14 primate DNA samples and 5 alternative transcripts was conducted by the commercial sequencing company, Genotech.

Sequence analysis and nucleotide sequence accession numbers

Multiple alignment (ClustalW) and manual correction of determined primate sequences were conducted using Bioedit software. The nucleotide sequences of the *LEPR* gene data reported in this paper have been deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases with the following accession numbers: [DDBJ: AB471813] from genomic DNA of *Homo sapiens*, [DDBJ: AB471814] from genomic DNA of *Pan troglodytes*, [DDBJ: AB471815] from genomic DNA of *Pongo sp.*, [DDBJ: AB471816] from genomic DNA of *Macaca mulatta*, [DDBJ: AB471817] from genomic DNA of *Macaca fascicularis*, [DDBJ: AB471818] from genomic DNA of *Mandrillus sphinx*, [DDBJ: AB471819] from genomic DNA of *Chlorocebus aethiops*, [DDBJ: AB471820] from genomic DNA of *Procolobus badius*, [DDBJ: AB471821] from genomic DNA of *Trachypithecus sp.*, [DDBJ: AB471822] from genomic DNA of *Ateles geoffroyi*, [DDBJ: AB471823] from genomic DNA of *Saguinus midas*, [DDBJ: AB471824] from genomic DNA of *Aotus nigriceps*, [DDBJ: AB471825] from genomic DNA of *Callithrix jacchus*, [DDBJ: AB471826] from genomic DNA of *Lemur catta*, [DDBJ: AB471890, AB471891] from mRNA samples of *Macaca fascicularis.*

RESULTS AND DISCUSSION

In silico analysis of the *LEPR* gene in humans, rhesus monkeys, mice, and rats

With the aid of computational analysis, alternative transcripts of the *LEPR* gene were identified in the genomes of humans (4 transcripts), rhesus monkeys (4 transcripts), mice (4 transcripts), and rats (3 transcripts) (Fig. 1). Interestingly, all major differences identified on the *LEPR* gene transcripts were detected within their 3′ terminal regions (Fig. 2). Additionally, the majority of these alternative transcripts occurred as the result of the exonization event of retrotransposons. We therefore focused our analysis into exonized retrotransposons on the *LEPR* gene using the RepeatMask program. In the tested species, different retrotransposons--including SVA in humans, *Alu*Jb in rhesus monkeys, B2_Mn1a, B3A, B1_Mur4, and MIRb in mice, and B3A in rats were identified (Fig. 2).

In order to investigate the integration status of exonized retrotransposons, genomic DNA sequences corresponding to the *LEPR* gene in related species (six species) were analyzed using bioinformatic tools. Among the retrotransposons identified

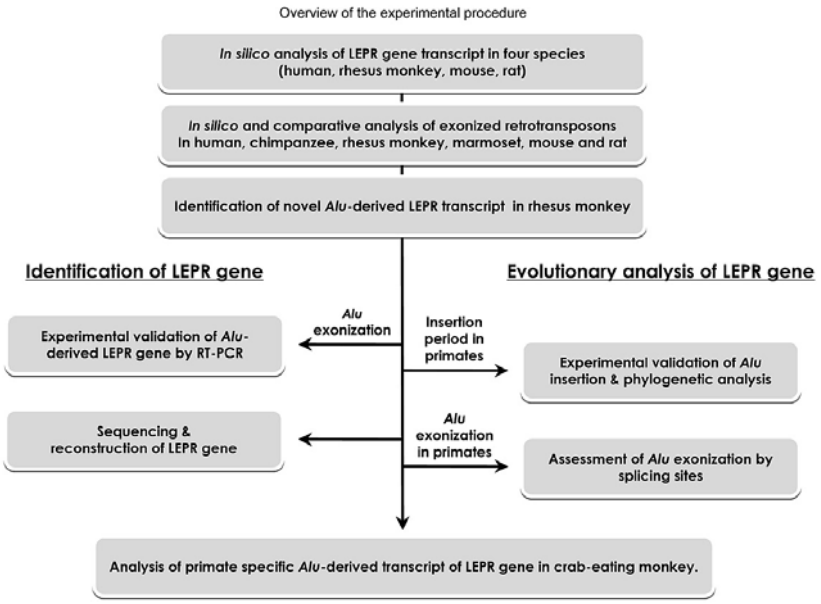


Fig. 1. Overview of experimental procedures. To better understand our experiment, we prepared a flowchart for a comprehensive overview of the experimental procedures.

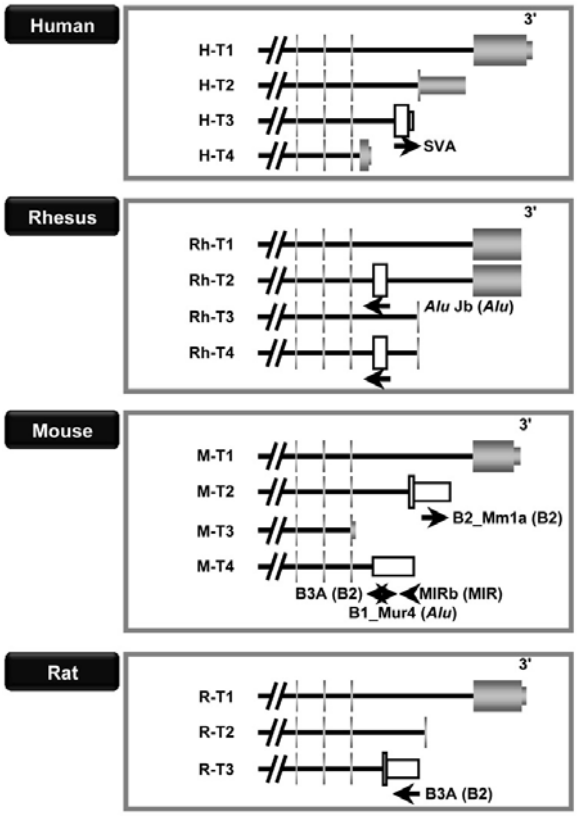


Fig. 2. Schematic representation of lineage-specific exonization events in the *LEPR* gene. Different retrotransposons are integrated into the *LEPR* gene in a lineage-specific manner. The boxed gene structures demonstrate the different 3′ terminal regions of the *LEPR* gene in humans, rhesus monkeys, mice, and rats. Open and closed boxes indicate the retrotransposon-derived exons and unrelated exons, respectively. Protein-coding regions and untranslated regions (UTRs) of the exons are illustrated as vertically thick and thin boxes, respectively. Horizontal arrows below the open boxes represent the indicated direction of the integrated retrotransposons.

in the *LEPR* gene, SVA was identified only in the human genome and B1_Mur4, B2_Mn1a, and MIRb were identified only in the mouse genome. That means SVA, B1_Mur4, B2_Mn1a, and MIRb are lineage-specific retrotransposons. However, *Alu*Jb (human, chimpanzee, rhesus monkey, and marmoset) and B3A (mouse and rat) elements are identified as common retrotransposons in different species. Intriguingly, although two common retrotransposons of *Alu*Jb and B3A were identified on the *LEPR* gene locus, the functional fates differed from each other.

The results of recent studies have shown that the majority of *Alu*-derived exonization events occurred via integration events of antisense orientation, via the provision of potential splicing sites (Kreahling and Graveley, 2004; Lev-Maor et al., 2003). Our analysis also demonstrated the antisense-oriented *Alu*Jb element in the 16th intron region of the human and rhesus monkey *LEPR* gene (Fig. 2). The rhesus monkey *Alu*Jb element was exonized as the 17th exon by providing a splicing donor and acceptor site. Additionally, exonized *Alu*Jb-derived sequences did not interrupt the coding sequences of *LEPR* (Van Vugt et al., 2006). This indicates that the exonization event of the *Alu*Jb element contributed to the occurrence of the *Alu*Jb exonized isoform, which is analogous to the case of human-specific LEPR 219.1 isoform occurring via an integration event of the SVA element (Damert et al., 2004). Until now, most *Alu* exonization events have been reported in humans, whereas these events have only been rarely reported in non-human primates (Krull et al., 2005; Lin et al., 2008; Singer et al., 2004). Thus, our analysis of exonization events occurring in rhesus monkeys might provide us with a clearer understanding of exonization events occurring during primate evolution. Additionally, in order to make it easier to understand this experiment, we prepared a flow chart describing our experimental procedures (Fig. 1).

Experimental validation of *Alu*Jb-related exonization event

Owing to limitations in the various primate RNA samples, previous *Alu* exonization events could not be validated by experimental approaches (Krull et al., 2005). Fortunately, however, we were able to obtain the crab-eating monkey adipose samples (one adipose tissue and two adipose tissue primary cells).

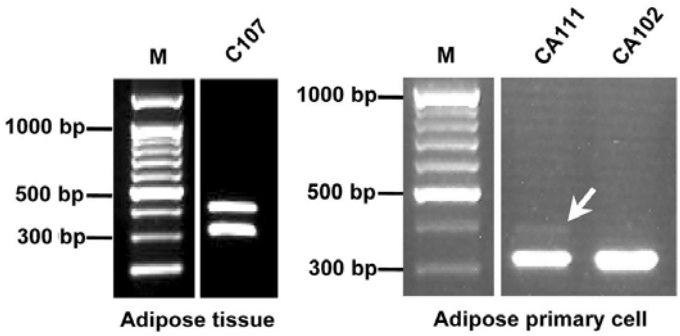


Fig. 3. RT-PCR amplification of the *LEPR* gene. The expressional analysis adipose tissue (one individual) and adipose tissue primary cells (two individuals) with the primer pair 1S and 1AS was conducted on crab-eating monkey samples. Predicted as 394 bp is the *Alu*Jb containing the alternative transcript of the *LEPR* gene (long form). Predicted as 301 bp is the original transcript (short form). Specific amplifications were confirmed via sequencing analysis. M indicates the molecular size marker.

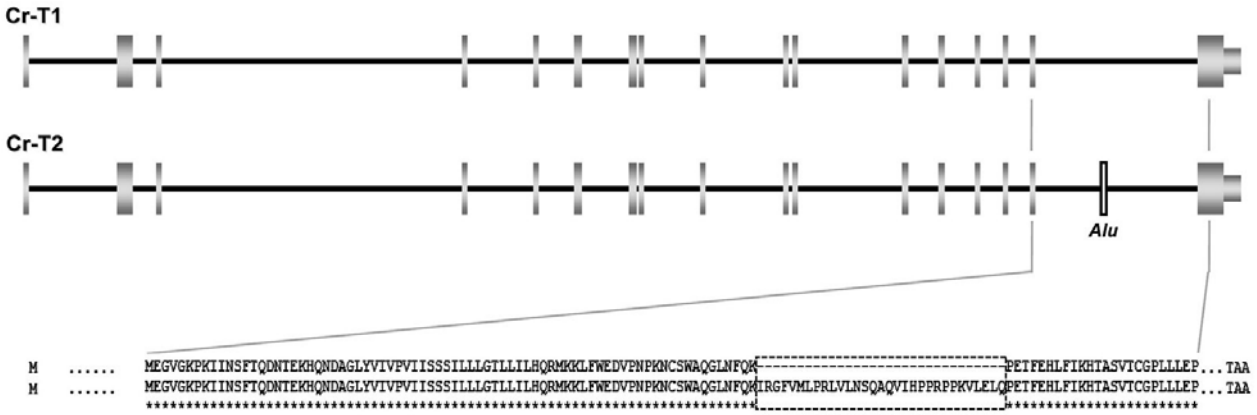


Fig. 4. Genomic structure and alternatively spliced transcripts of crab-eating monkey *LEPR* gene (Cr-T1, Cr-T2). Magnification focused on the *Alu*Jb-derived exon (exonization) and adjacent exons. The primer pairs 1S and 1AS were used for the amplification and sequencing procedures. Sequences were translated and analyzed. Dotted box indicated the *Alu*Jb-derived protein coding region. Protein-coding regions and untranslated regions (UTRs) of the exons are illustrated as vertically thick and thin boxes, respectively.

Although we were not able to obtain the rhesus monkey RNA samples, samples from the closely related crab-eating monkey are sufficient for the validation of the exonization event of *Alu*Jb-associated transcripts on *LEPR* via RT-PCR analysis and sequencing (Fig. 3, Supplementary Fig. S2). Among the three different samples, only the adipose tissue sample shows the marked expression activity of the *Alu*Jb-derived alternative transcript (Fig. 3). Although we were unable to identify the *Alu*Jb-derived alternative transcript in CA102 samples of adipose primary cells, and CA111 samples show low-level expression patterns. These phenomena could be explained by the possibility that the *Alu*JB-derived transcript had lost its transcription activity during the cell culture procedure. Additionally, epigenetic differences could also represent an effective solution for the different fates of *Alu*Jb-derived transcripts in different cell types. However, owing to limitations of the RNA samples, our questions could not be answered satisfactorily.

In C107, CA111, and CA102 samples, all amplified 394 bp and 301 bp fragments were also sequenced and analyzed for experimental confirmation (Supplementary Fig. S2). RT-PCR analysis of crab-eating monkey samples revealed the existence of the *Alu*Jb-derived alternative transcript of the *LEPR* gene. Additionally, sequence analysis of crab-eating monkey samples also showed clear open reading frames which could be identified in rhesus monkey *LEPR* gene (Fig. 4). Thus, the crab-eating monkey genome may derive its functional protein sequences via exonization events involving *Alu*Jb elements.

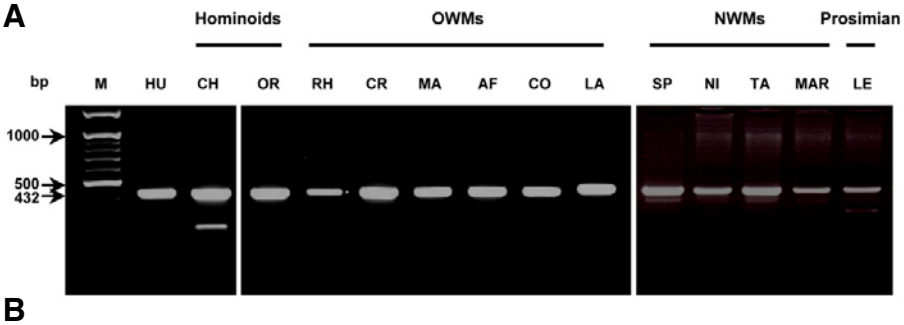
As a result, our experimental validation confirmed the existence and expression of the *Alu*Jb-derived alternative transcript of the *LEPR* gene in the crab-eating monkey.

Evolutionary analysis of *Alu*Jb insertion in the *LEPR* gene

In an effort to trace the possibility of exonization events of the *Alu*Jb element in the *LEPR* gene, locus-specific primer pairs (S1 and AS1) for the identification of orthologous *Alu*Jb sequences of 14 primate species were designed (Supplementary Fig. S1). These sequences were then compared with the consensus *Alu*Jb sequences retrieved from the Repbase updates (Jurka, 2000).

First, we evaluated the integration lineage of *Alu*Jb during primate radiation via a PCR amplification procedure with different primate DNA samples. The results demonstrated that all tested primate lineages harbored the same 432 bp fragment (Fig. 5A). For the thorough validation of the false positive results from the amplified 432 bp products, we performed a variety of sequencing procedures. These results demonstrate that *Alu*Jb was integrated into our common ancestor genome approximately 60 million of years ago (MYA), prior to the divergence of simians and prosimians.

Second, 14 amplified primate sequences were carefully sequenced and aligned (Supplementary Fig. S3). Our detailed sequence analysis showed that the 3′ splice site of newly acquired exons by *Alu*Jb were identical to the *Alu*Jb consensus sequences (Fig. 5B). However, different locus-specific mutation events were identified within the boundary region of the 5′ splice site. Basically, all 14 tested primate species evidence



B

	3'SS					5'SS									
	−2	−1	1 2 3			−4	−3	−2	−1	1	2	3	4	5	6
Alu Jb	A	G	AGA	···	TTA	—	C	A	G	G	C	G	T	G	A
Human*	A	G	ATA	···	TTA	A	C	A	G	A	T	G	T	G	A
Chimpanzee	A	G	ATA	···	TTA	—	C	A	G	G	T	G	T	G	A
Orangutan	A	G	ATA	···	TTA	—	C	A	G	G	T	G	T	G	A
Rhesus	A	G	ATA	···	TTA	—	C	A	G	G	T	G	T	G	A
Crab-eating	A	G	ATA	···	TTA	—	C	A	G	G	T	G	T	G	A
Mandrill	A	G	ATA	···	TTA	—	C	A	G	G	T	G	T	G	A
African green	A	G	ATA	···	TTA	—	C	A	G	G	T	G	T	G	A
Colobus	A	G	GTA	···	TTA	—	C	A	G	G	T	G	T	G	A
Langur	A	G	GTA	···	TTA	—	C	A	G	G	T	G	T	G	A
Spider*	A	G	GTA	···	TTA	—	C	A	G	G	T	G	G	G	A
Night*	A	G	GTA	···	TTA	—	C	A	G	G	T	G	G	G	A
Tamarine*	A	G	ATA	···	TTA	—	C	A	G	G	T	G	G	G	A
Marmoset*	A	G	GTA	···	TTA	—	C	A	G	G	T	G	G	G	A
Lemur*	A	G	GTA	···	TTA	—	C	A	G	G	T	G	G	G	A

Fig. 5. PCR amplification of *Alu*Jb and summary of boundary sequences of potential 3′ and 5′ alternative splice sites in various primates. (A) Primate DNA samples were utilized for integration analyses of the *Alu*Jb elements in the *LEPR* gene. M indicates the molecular size marker. Primate DNA samples are abbreviated as follows: (1) HU: human (*Homo sapiens*); (2) hominoids: CH: chimpanzee (*Pan troglodytes*), OR: orangutan (*Pongo pygmaeus*); (3) Old World monkey: RH: rhesus monkey (*Macaca mulatta*), CR: Crab-eating mankey (*Macaca fascicularis*), MA: Mandrill (*Mandrillus sphinx*), AF: African green monkey (*Cercopithecus aethiops*), CO: Colobus (*Procolobus sp*), LA: Langur (*Trachypithecus sp*), (4) SP: Spider monkey (*Ateles geoffroyi*), NI: Night monkey (*Aotus nigriceps*), TA: Tamarin (*Saguinus midas*), MAR: Marmoset (*Callithrix jacchus*), (5) LE: Lemur (*Lemur catta*). (B) Summary of boundary sequences of potential 3′ and 5′ alternative splice sites. Consensus sequences of *Alu*Jb and corresponding primate sequences are indicated. Canonical splicing sites of "AG" and "GT" are designated by gray coloring. Red and blue colors indicate the "G" to "A" transition mutation and 1nt-insertion mutation of "A" in the adjacent region of the 5′ splice site, respectively. In humans, the bold line indicates the frame-shift mutation by 1nt-insertion. The bold line in the crab-eating monkey indicates the uninterrupted frame for translation. Green colors represent the "C" to "T" transition mutation in the 5′ splice site. Yellow colors represent the "T" to "G" transversion mutation in the adjacent region of the 5′ splice site. An asterisked species might not harbor the *Alu*Jb-derived exons.

common transition mutations (second position of C to T). Therefore, all primate species could acquire the canonical 5′ splice site of "GT". However, the critical 1nt-insertion event of "A", which could result in a frame shift and non-sense mutation in the minus fourth position was identified in the human genome (Fig. 5B). Additionally, a transition mutation in the 5′ splice site (first position of G to A) was also identified. These human-specific mutations could result in the failure of the recognition of spliceosomes for the appropriate removal of intron sequences during the splicing event; additionally, frame-shift and non-sense mutations could induce the production of truncated (non-functional) proteins (Ast, 2004).

In New World monkeys and prosimian lineages, canonical splicing signals were identified in the 3′ and 5′ splice sites (Fig. 5B). However, they evidenced "T" to "G" transversion mutations located in the fourth position of the 5′ splicing site. Previous *in vitro* experimental analysis of *Alu*-derived exonization events showed that the third and fourth positions of the 5′ splicing site are critically involved in the control of the exon inclusion level (Ram et al., 2008; Sorek et al., 2004). Additionally, the preferred sequences of the fourth position for the creation of *Alu*-derived alternative transcripts are "T" or "C" (Ram et al., 2008; Sorek et al., 2004). Although we were unable to validate experimentally the existence or presence of *Alu*Jb-related transcripts on New World monkeys and prosimian lineages, the *Alu*Jb elements on the *LEPR* gene may not be exonized.

According to our results, careful comparative analyses of potential splicing sites and adjacent sequences strongly indicated the possibility of lineage-specific exonization events involving *Alu*Jb on the *LEPR* gene in Old World monkey, chimpanzee, and orangutan lineages.

Evolutionary perspective of exonization event in *LEPR* gene

The leptin receptor gene (*LEPR*) is the primary element in the regulation of body weight, and controls the hypothalamus (Friedman and Halaas, 1998). *LEPR* affects diverse signaling pathways in relation to T-cells and pancreatic beta-cells (Load et al., 1998). Among the various functional regions of *LEPR*, the 3′ cytoplasmic tail is the most important. This is because the phosphotyrosines of *LEPR* in their 3′ cytoplasmic tail are crucial for the interaction between JAK2 and STAT5 (Fruhbeck, 2006). However, a variety of lineage-specific retrotransposons were integrated into the 3′ terminal regions of *LEPR* in humans, rhesus monkeys, rats, and mice (Fig. 2). Thus, the protein diversity acquired by the exonization event of retrotransposons could exert an advantageous or harmful effect on various signaling pathways in their specific lineages. *LEPR* shows intriguing evidence of enhanced 3′ terminal diversity by lineage-specific integration events involving different retrotransposons in humans, rhesus monkeys, mice, and rats. The accumulation of integrated lineage-specific retrotransposons provided additional protein sequences for the diversity of the functional signaling pathway in association with *Leptin,* via an alternative splicing mechanism. Although we were unable to conduct any functional analysis of the alternatively spliced *LEPR* gene by

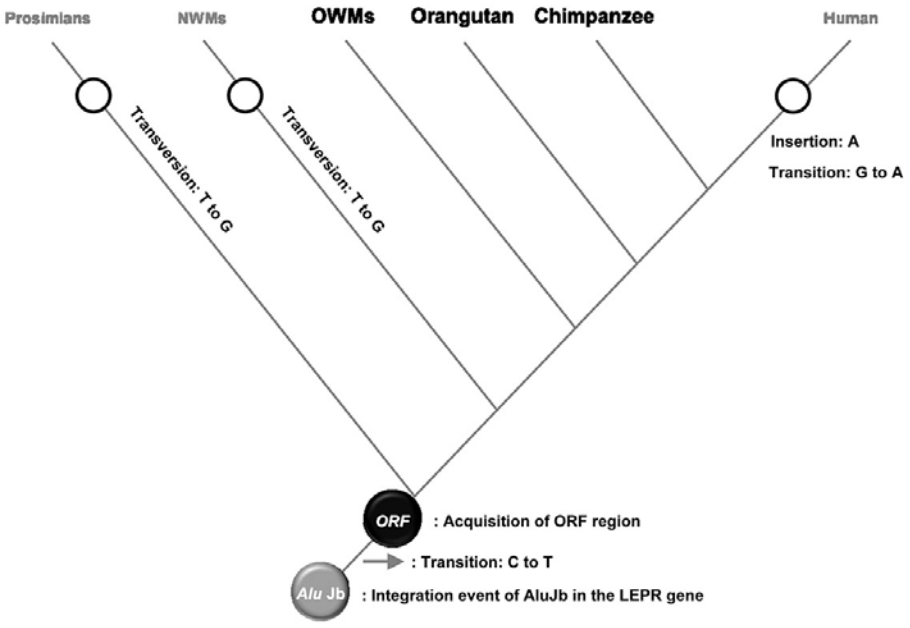


Fig. 6. Evolutionary pathway of lineage specific *Alu*Jb-derived exonization events during the primate evolution. The *Alu*Jb integration event is shown as a gray circle in the lineage of the common ancestor genome of primates. The closed circle indicates the acquisition of an open reading frame of *Alu*Jb-derived sequences created via an exonization event. Open circles indicate the mutation events causing the interruption of the exonization event of the *Alu*Jb element, including the "T" to "G" transversion (prosimians and NWMs), the "A" insertion (human), and the "G" to "A" transition (human).

assessing the exonization events, those retrotransposon-derived regions may be one of the most important regions that interact with specific proteins for species-specific characteristics.

In our analysis, we carefully traced the evolutionary scenario and validated the fate of the *Alu* element on the *LEPR* gene in different lineages of primates. In the case of the *LEPR* gene, exonization events might have occurred abruptly via transition mutations following the integration event of the *Alu*Jb element in our common ancestor genome prior to the divergence of simians and prosimians (Fig. 6). And then, lineage specific mutations events were occurred in different lineages. The results of our analysis confirmed the diverse routes to protein-coding functions of *Alu* exonization events, as indicated by Krull et al. (2005). Our experimental validation could facilitate the necessary functional approaches to *Alu*-exonization-associated protein isoforms.

All in all, retrotransposons may prove to be valuable factors for lineage-specific gene diversification and their different evolutionary fates (selection or not) could generate lineage-specific proteins as the result of different mutation pressures.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENT

This study was supported by a research grant from the Ministry of Education, Science, and Technology (KBM4311022, KGC 5401011), Republic of Korea.

REFERENCES

Ast, G. (2004). How did alternative splicing evolve? Nat. Rev. Genet. *5*, 773-782.

Batzer, M.A., and Deininger, P.L. (2002). Alu repeats and human genomic diversity. Nat. Rev. Genet. *3*, 370-379.

Costas, J. (2003). Molecular characterization of the recent intragenomic spread of the murine endogenous retrovirus MuERV-L. J. Mol. Evol. *56*, 181-186.

Damert, A., Lower, J., and Lower, R. (2004). Leptin receptor isoform 219.1: an example of protein evolution by LINE-1-mediated human-specific retrotransposition of a coding SVA element. Mol. Biol. Evol. *21*, 647-651.

Friedman, J.M., and Halaas, J.L. (1998). Leptin and the regulation of body weight in mammals. Nature *395*, 763-770.

Fruhbeck, G. (2006). Intracellular signalling pathways activated by leptin. Biochem. J. *393*, 7-20.

Han, K., Konkel, M.K., Xing, J., Wang, H., Lee, J., Meyer, T.J., Huang, C.T., Sandifer, E., Hebert, K., Barnes, E.W., et al. (2007). Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. Science *316*, 238-240.

Huh, J.W., Kim, T.H., Yi, J.M., Park, E.S., Kim, W.Y., Sin, H.S., Kim, D.S., Min, D.S., Kim, S.S., Kim, C.B., et al. (2006). Molecular evolution of the periphilin gene in relation to human endogenous retrovirus m element. J. Mol. Evol. *62*, 730-737.

Huh, J.W., Ha, H.S., Kim, D.S., and Kim, H.S. (2008). Placenta-restricted expression of LTR-derived NOS3. Placenta *29*, 602-608.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. *16*, 418-420.

Kreahling, J., and Graveley, B.R. (2004). The origins and implications of alternative splicing. Trends Genet. *20*, 1-4.

Krull, M., Brosius, J., and Schmitz, J. (2005). *Alu*-SINE exonization: en route to protein-coding function. Mol. Biol. Evol. *22*, 1702-1711.

Lev-Maor, G., Sorek, R., and Ast, G. (2003). The birth of an alternatively spliced exon: 3′ splice-site selection in *Alu* exons. Science *300*, 1288-1291.

Li, W.H., Gu, Z., Wang, H., and Nekrutenko, A. (2001). Evolutionary analyses of the human genome. Nature *409*, 847-849.

Lin, L., Shen, S., Tye, A., Cai, J.J., Jiang, P., Davidson, B.L., and Xing, Y. (2008). Diverse splicing patterns of exonized *Alu* elements in human tissues. PLoS Genet. *17*, e1000225.

Lord, G.M., Matarese, G., Howard, J.K., Baker, R.J., Bloom, S.R., and Lechler, R.I. (1998). Leptin modulates the T-cell immune response and reverses starvation-induced immunosuppression. Nature *394*, 897-901.

Piriyapongsa, J., Polavarapu, N., Borodovsky, M., and McDonald, J. (2007). Exonization of the LTR transposable elements in human genome. BMC Genomics *8*, 291.

Ram, O., Schwartz, S., and Ast, G. (2008). Multifactorial interplay controls the splicing profile of Alu-derived exons. Mol. Cell. Biol. *28*, 3513-3525.

Sela, N., Mersch, B., Gal-Mark, N., Lev-Maor, G., Hots-Wagenblatt, A., and Ast, G. (2007). Comparative analysis of transposed element insertion within human and mouse genes reveals *Alu*'s unique role in shaping the human transcriptome. Genome Biol.

8, R127.
Singer, S.S., Männel, D.N., Hehlgans, T., Brosius, J., and Schmitz, J. (2004). From "junk" to gene: curriculum vitae of a primate receptor isoform gene. J. Mol. Biol. *20*, 883-886.
Sorek, R., Ast, G., and Graur, D. (2002). *Alu*-containing exons are alternatively spliced. Genome Res. *12*, 1060-1067.
Sorek, R., Lev-Maor, G., Reznik, M., Dagan, T., Belinky, F., Graur, D., and Ast, G. (2004). Minimal conditions for exonization of intronic sequences: 5′ splice site formation in alu exons. Mol. Cell *14*, 221-231.
Sverdlov, E.D. (1998). Perpetually mobile footprints of ancient infections in human genome. FEBS Lett. *428*, 1-6.
The Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequencing of the chimpanzee genome and comparison with the human genome. Nature *437*, 69-87.
van de Lagemaat, L.N., Landry, J.R., Mager, D.L., and Medstrand, P. (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet. *19*, 530-536.
Van Vugt, D.A., Lujan, M.E., Froats, M., Krzemien, A., Couceyro, P.R., and Reid, R.L. (2006). Effect of fasting on cocaine-amphetamine-regulated transcript, neuropeptide Y, and leptin receptor expression in the non-human primate hypothalamus. Neuroendocrinology *84*, 83-93.
Xiao, R., Park, K., Oh, Y., Kim, J., and Park, C. (2008). Structural characterization of the genome of BERV gamma4, the most abundant endogenous retrovirus family in cattle. Mol. Cells *26*, 404-408.