

Gain of a New Exon by a Lineage-Specific *Alu* Element-Integration Event in the *BCS1L* Gene during Primate Evolution

Sang-Je Park^{1,4}, Young-Hyun Kim^{1,2,4}, Sang-Rae Lee^{1,2,4}, Se-Hee Choe^{1,2}, Myung-Jin Kim¹, Sun-Uk Kim¹, Ji-Su Kim¹, Bo-Woong Sim¹, Bong-Seok Song¹, Kang-Jin Jeong¹, Yeung-Bae Jin¹, Youngeon Lee¹, Young-Ho Park¹, Young Il Park³, Jae-Won Huh^{1,2,*}, and Kyu-Tae Chang^{1,2,*}

BCS1L gene encodes mitochondrial protein and is a member of conserved AAA protein family. This gene is involved in the incorporation of Rieske FeS and Qcr10p into complex III of respiratory chain. In our previous study, *AluYRa2*-derived alternative transcript in rhesus monkey genome was identified. However, this transcript has not been reported in human genome. In present study, we conducted evolutionary analysis of *AluYRa2*-exonized transcript with various primate genomic DNAs and cDNAs from humans, rhesus monkeys, and crab-eating monkeys. Remarkably, our results show that *AluYRa2* element has only been integrated into genomes of *Macaca* species. This *Macaca* lineage-specific integration of *AluYRa2* element led to exonization event in the first intron region of *BCS1L* gene by producing a conserved 3' splice site. Intriguingly, in rhesus and crab-eating monkeys, more diverse transcript variants by alternative splicing (AS) events, including exon skipping and different 5' splice sites from humans, were identified. Alignment of amino acid sequences revealed that *AluYRa2*-exonized transcript has short N-terminal peptides. Therefore, AS events play a major role in the generation of various transcripts and proteins during primate evolution. In particular, lineage-specific integration of *Alu* elements and species-specific *Alu*-derived exonization events could be important sources of gene diversification in primates.

INTRODUCTION

Alternative splicing (AS) is an important molecular mechanism that regulates genes and increases genome diversity and complexity without increasing genome size in eukaryotes (Kim et al., 2014). In the case of humans, high-throughput sequencing data revealed that approximately 95% of multi-exon genes undergo AS (Pan et al., 2008). AS events are classified into four main types, including exon skipping, alternative 5' splice sites, alternative 3' splice sites, and intron retention, as well as other less frequent types, such as mutual exclusion, alternative promoter usage, and alternative polyadenylation (Ast, 2004; Keren et al., 2010). The exon skipping mechanism accounts for nearly 40% of AS events and is conserved in higher eukaryotes genomes. This mechanism could produce specific exon-skipped mRNA transcripts (Alekseyenko et al., 2007). Alternative 5' splice sites and alternative 3' splice sites account for nearly 18% and 8% of AS events, respectively. Additionally, these events are conserved in higher eukaryote genomes. Specifically, these events could occur when two or more splice sites are recognized at one end of an exon (Ast, 2004). Therefore, different sized exons are transcribed as different mRNA transcripts. Intron retention accounts for less than 5% of all AS events in vertebrate and invertebrate genomes (Park et al., 2012); in this case, intron sequences remain in the mature mRNA transcript without being spliced out.

A previous comparative analysis between human and mouse genomes indicated that AS is often associated with recent exon creation and/or loss (Lev-Maor et al., 2003). Three mechanisms, including exon shuffling, transition of a constitutive exon to an alternative exon, and exonization of intronic sequences, have been identified as causes of the production of alternatively spliced new exons (Corvelo and Eyras, 2008). In mammals, most newly exonized intronic sequences are generated by transposable elements (TEs) (Almeida et al., 2007; Amit et al., 2007; DeBarry et al., 2006; Huh et al., 2009; 2010; Piriyaopongsa et al., 2007). In particular, *Alu* elements are the most abundant type of TEs in the human genome and account for more than 10% of the human genome; consequently, they are a major source of new exon creation in human and non-human primate genomes (Huh et al., 2010; Lander et al., 2001; Lev-Maor et al., 2003; Lin et al., 2008). *Alu* elements are typically

¹National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology, Cheongju 363-883, Korea, ²University of Science & Technology, National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology, Cheongju 363-883, Korea, ³Graduate School Department of Digital Media, Ewha Womans University, Seoul 120-750, Korea, ⁴These authors contributed equally to this work.

*Correspondence: huhjw@kribb.re.kr (JWH); changkt@kribb.re.kr (KTC)

Received 7 May, 2015 ; revised 17 July, 2015 ; accepted 20 July, 2015 ; published online 4 November, 2015

Keywords: alternative splicing, *AluYRa2*, *BCS1L*, exonization, transposable element

300 nucleotides long and usually inserted with the antisense orientation by retrotransposition into the introns of primate genes (Keren et al., 2010; Lev-Maor et al., 2003). Antisense orientation of *Alu* elements contains potential 5' and 3' splice sites that could be recognized by spliceosomes, and 85% of exonizations occur within the right arm (Ast, 2004; Sela et al., 2010). In the human genome, more than 5% of the alternatively spliced new exons are derived from *Alu* elements (Sorek et al., 2002). Therefore, *Alu*-derived exonization events could be considered major sources of increased genome diversity in human and non-human primates by creation of novel protein isoforms.

The BC1 (ubiquinol-cytochrome c reductase) synthesis-like (*BCS1L*) gene is located on human chromosome 2q33. This gene encodes a homolog of the *Saccharomyces cerevisiae* *bcs1* protein (419-amino-acid), which is required for the assembly of the Rieske iron-sulfur subunit of complex III (the BC1 complex) of the mitochondrial respiratory chain (de Lonlay et al., 2001; Visapaa et al., 2002). It is a mitochondrial inner-membrane protein with a single transmembrane domain that belongs to the conserved mitochondrial protein AAA family, which consists of ATPases associated with various cellular activities (Fernandez-Vizarra et al., 2007; Hinson et al., 2007). The AAA families are involved in folding, unfolding, assembly, and degradation of other proteins (Hanson and Whiteheart, 2005; Snider and Houry, 2008; Tucker and Sallai, 2007). Previous studies have shown that mutations in the *BCS1L* gene are associated with various diseases such as GRACILE syndrome (which includes symptoms of intrauterine growth retardation, aminoaciduria, cholestasis, iron overload, lactic acidosis, and early death), Björnstads syndrome (sensorineural hearing loss and pili torti), mitochondrial encephalopathy, neonatal tubulopathy, encephalopathy, liver failure, and complex III deficiency (de Lonlay et al., 2001; Fernandez-Vizarra et al., 2007; Hinson et al., 2007; Visapaa et al., 2002).

In our analysis, we focused on the identification and molecular characterization of *Alu*-derived exonization events in the *BCS1L* gene in primates, specifically in the rhesus monkey (*Macaca mulatta*) and crab-eating monkey (*Macaca fascicularis*). Additionally, we validated that lineage-specific AS occurred and resulted in an integration event of an *Alu* element in the *BCS1L* gene during primate evolution.

MATERIALS AND METHODS

Ethics statement

Animal preparation and study design were conducted in accordance with the Guidelines of the Institutional Animal Care and Use Committee (KRIBB-AEC-15031) of the Korea Research Institute of Bioscience and Biotechnology (KRIBB).

Total RNA and genomic DNA samples

Total RNA from humans (*Homo sapiens* bone marrow, whole brain, fetal brain, fetal liver, heart, kidney, liver, lung, placenta, prostate, skeletal muscle, spleen, testis, thymus, trachea, uterus, colon, small intestine, spinal cord, and stomach) and rhesus monkeys (*Macaca mulatta* cerebrum, colon, liver, lung, kidney, pancreas, and stomach) were purchased from Clontech. An adult female (6 years of age) crab-eating monkey (*Macaca fascicularis*) that originated from Vietnam was imported from China with a Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) permit; she weighed 5 kg. The crab-eating monkey was provided by the National Primate Research Center (NPRC) of Republic of Korea. Five tissue samples from the cerebrum, heart, kidney, liver, and lung were obtained from the specific pathogen free (SPF) adult fe-

male crab-eating monkey following deep anesthesia using ketamine (5 mg/kg) by intramuscular injection, and perfusion with diethylpyrocarbonate (DEPC)-treated cold phosphate-buffered saline (PBS) was conducted via the common carotid artery with RNase inhibitors to inhibit blood contamination and promote the recovery of intact RNA molecules from the tissue samples. Total RNA was extracted from five tissue samples of the crab-eating monkey using the RNeasy Mini kit (Qiagen), according to the manufacturer's instructions. RNase-free DNase (Qiagen) was used to eliminate DNA contamination from the total RNA preparations. The RNA concentration and the absorbance ratio at 260 nm and 280 nm (A260/A280) were determined with a NanoDrop® ND-1000UV-Vis Spectrophotometer (NanoDrop Technologies).

We used a standard protocol to isolate genomic DNA from heparinized blood samples from the following species: (1) HU: humans (*Homo sapiens*); (2) hominoids: CH: chimpanzees (*Pan troglodytes*), BO: bonobos (*Pan paniscus*); (3) Old World monkeys (OWM): RH: rhesus monkeys (*Macaca mulatta*), JA: Japanese monkeys (*Macaca fuscata*), CR: crab-eating monkeys (*Macaca fascicularis*), PI: pig-tail monkeys (*Macaca nemestrina*), AF: African green monkeys (*Chlorocebus aethiops*), MA: mandrills (*Mandrillus sphinx*), CO: colobus monkeys (*Procolobus* sp.), LA: langurs (*Trachypithecus* sp.); (4) New World monkeys (NWM): MAR: marmosets (*Callithrix jacchus*), TA: tamarins (*Saguinus midas*), CA: capuchins (*Cebus apella*), SQ: squirrel monkeys (*Saimiri sciureus*), NI: night monkeys (*Aotus nigriceps*), SP: spider monkeys (*Ateles geoffroyi*); and (5) prosimians: RL: ring-tailed lemurs (*Lemur catta*).

RT-PCR and PCR amplification

BCS1L transcripts were analyzed by RT-PCR amplification. Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase with an annealing temperature of 42°C was used with an RNase inhibitor (Promega). We performed PCR amplification of pure mRNA samples without reverse transcription to demonstrate that the mRNA samples did not contain genomic DNA (data not shown). As a standard control, RPL32 was amplified from humans, rhesus monkeys, and crab-eating monkeys. The expression levels of the *BCS1L* gene of humans (NM_004328.4 and NM_001079866.1), rhesus monkeys, and crab-eating monkeys were calculated by RT-PCR experiments using specific primer pairs (Supplementary Table S1). RT-PCR experiments were carried out for 35 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s. Genomic DNA from various primates were PCR amplified using well-designed primer pairs on the highly conserved sequences in human and non-human primates (Supplementary Table S1). Genomic PCR experiments were carried out for 35 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 50 s.

Rapid amplification of cDNA ends (RACE) of the *BCS1L* gene

Following the protocols of the CapFishing™ Full-length cDNA Premix kit (Seegene), about 1–3 µg of total RNA from placental tissues of rhesus monkeys were reverse transcribed with the CapFishing™ adaptor for 5' RACE and the Oligo dT adaptor for 3' RACE. The first round of PCR was performed with either of the following primer combinations: 5'-RACE primer and 5'-target site primer (TSP; the target is the well-conserved region of *BCS1L*) or 3'-RACE primer and 3'-TSP (Supplementary Table S1). PCR products were size-fractionated by agarose gel electrophoresis, and the smear between 100–500 bp was purified with the Gel SV Extraction kit (GeneAll). Subsequently, nested PCR was performed under normal PCR conditions to

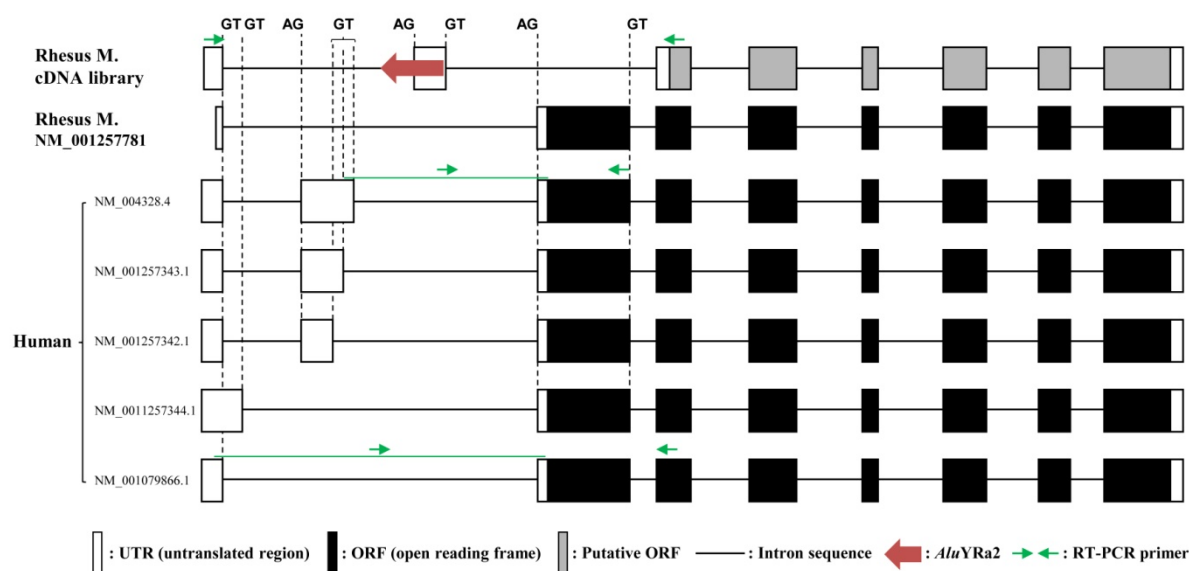


Fig. 1. Structural analysis of human and rhesus monkey *BCS1L* gene transcripts. The antisense-oriented *AluYRa2* element is located on the first intron of *BCS1L* in rhesus monkeys. Open, black, and gray boxes represent the untranslated region of the exons, protein-coding region, and putative protein-coding region, respectively. Vertical dashed lines indicate the 3' (GT) and 5' (AG) splice sites and the horizontal line indicates intron sequences. This figure is a structural illustration and is not to scale.

remove non-specific products with either of the following primer combinations: 5'-RACE primer and 5'-specific primer (SP) or 3'-RACE primer and 3'-SP. All of the primers used in RACE reactions are listed in Supplementary Table S1.

Molecular cloning and sequencing procedures

PCR products were separated on a 1.2% agarose gel, purified with the Gel SV Extraction kit (GeneAll), and cloned into the pGEM-T-easy vector (Promega). The cloned DNA was isolated using the Plasmid DNA Mini-prep kit (GeneAll). Sequencing of primate DNA samples and alternative transcripts was performed by a commercial sequencing company (Macrogen).

Calculating the integration time of *AluYRa2*

To estimate the integration time of *AluYRa2* in *BCS1L* gene region (*BCS1L* *AluYRa2*), we calculate the time for most recent ancestor (TMRCA) and divergence time. TMRCA was estimated by NETWORK version 4.6.1.3 computer software (Fluxus Technology Ltd.). And also, to calculate the divergence time of *AluYRa2* elements, one hundred twenty intact macaque-specific *AluYRa2* elements were selected from the University of California Santa Cruz (UCSC) Genome Browser (<http://genome.ucsc.edu>). The phylogenetic tree of each *AluYRa2* was made by the neighbor-joining method of MEGA6.0 software (Tamura et al., 2013). The actual number of nucleotide substitutions was calculated by pairwise distance method of the Molecular Evolutionary Genetics Analysis (MEGA) software 6.0. Poly(A) tails in the sequences were excluded from the analysis.

RESULTS

Structure and comparative analysis of the *BCS1L* gene in humans and rhesus monkeys

In our previous study, we constructed and sequenced a full-length cDNA library as well as conducted a comparative gene analysis with humans to identify numerous full-length gene

transcripts using the placental tissue of rhesus monkeys (Kim et al., 2010). A rhesus monkey-specific AS event was identified on the *BCS1L* gene, and this event occurred by integration of the *AluYRa2* element (Fig. 1). To investigate the structural difference of the *BCS1L* gene between humans and rhesus monkeys, we performed comparative analysis with mRNA sequences of rhesus monkeys (*Alu*-exonized transcript and reference gene; NM_001257781) and humans (NM_004328.4, NM_001257343.1, NM_001257342.1, NM_001079866.1, and NM_001079866.1). The *Alu*-exonized transcript is composed of eight exons and transcribed into a 1169-base pair (bp) mRNA sequence with a 212 bp 5' untranslated region (UTR), 900-bp coding sequence (CDS), and 57 bp 3' UTR. In this transcript, the antisense-oriented *AluYRa2* element is inserted on the first intron and provides a canonical splicing donor site that produces a new *AluYRa2*-derived exon via exonization events. Additionally, 369 bp of an exon skipping event was also found. This event resulted in a short CDS compared with the reference sequences of rhesus monkeys and humans; however, 3' and 5' splice sites are conserved on both ends of the skipped exon. Alternatively, human *BCS1L* transcripts including NM_004328.4, NM_001257343.1, and NM_001257342.1 have one more exon (a second exon) than the other two transcripts. This additional exon was not found in rhesus monkeys (Fig. 1). Intriguingly, the rhesus monkey *BCS1L* gene also has a conserved 3' splice site (AG) and three different 5' splice sites (GT) in the orthologous sequences; however, only humans had additional exon sequences. We also confirmed that the human and rhesus monkey reference genes of *BCS1L* encoded 419 amino acids, and the *AluYRa2*-exonized transcript encoded 299 amino acids (Supplementary Fig. S1).

Evolutionary analysis of the *AluYRa2* insertion in the *BCS1L* gene

Comparative structure analysis revealed that the integrated *AluYRa2* element led to creation of a new exon in the *BCS1L*

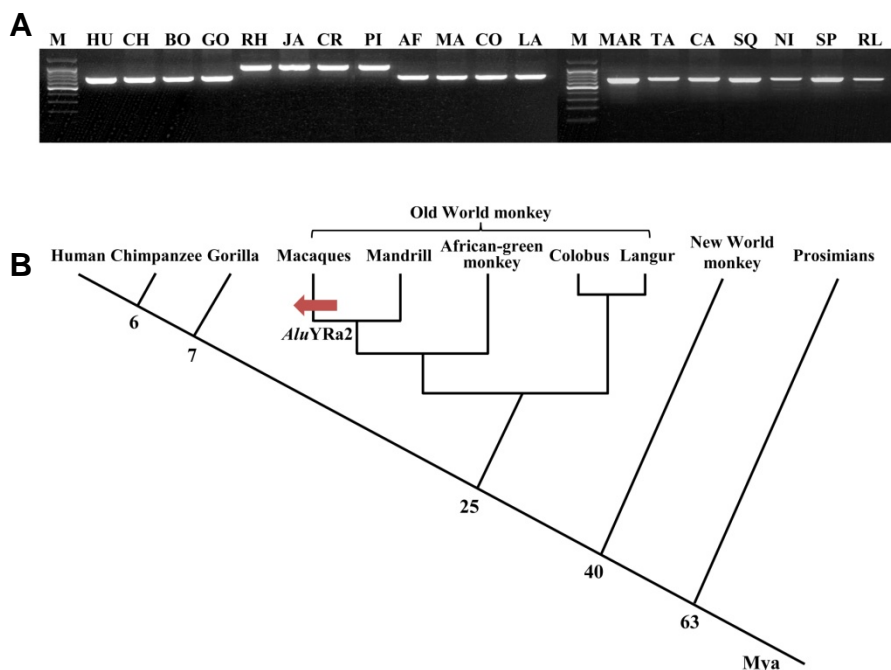


Fig. 2. Evolutionary investigation of the *AluYRa2* element during primate evolution. (A) PCR amplification of *AluYRa2* in various primates. Primate DNA samples were utilized for integration analyses of the *AluYRa2* elements in the *BCS1L* gene. M indicates the molecular size marker. Primate DNA samples are abbreviated as follows: (1) HU: human; (2) hominoids: CH: chimpanzee, BO: bonobo, GO: gorilla; (3) Old World monkeys: RH: rhesus monkey, JA: Japanese monkey, CR: crab-eating monkey, PI: pig-tail monkey, AF: African green monkey, MA: mandrill, CO: colobus, LA: langur; (4) New World monkeys: MAR: Marmoset, TA: tamarin, CA: capuchin, SQ: squirrel monkey, NI: night monkey, SP: spider monkey; and (5) prosimian: RL: ring-tailed lemur. (B) Schematic representation of the integration events of *AluYRa2* during primate evolution for the *BCS1L* gene. The integration event of the *AluYRa2* element is indicated as a red arrow. Mya, million years ago.

gene of rhesus monkeys; however, this element was not found in the human *BCS1L* gene. Therefore, to investigate when the *AluYRa2* element was incorporated into primate lineages, we performed genomic PCR amplification using locus-specific primer pairs that are designed for highly conserved regions in the genomic DNA of 19 primates, including hominoids (humans, chimpanzees, and bonobos), OWM (rhesus monkeys, Japanese monkeys, crab-eating monkeys, pig-tail monkeys, African-green monkeys, mandrills, colobus monkeys, and langurs), NWM (marmosets, tamarins, capuchins, squirrel monkeys, night monkeys, and spider monkeys), and prosimians (ring-tailed lemurs) (Fig. 2A). To evaluate if there were false-positive results in the amplified PCR products, we sequenced the products of 15 primate species (Fig. 3). The results demonstrated that integration of the *AluYRa2* element was only found only in *Macaca* species (rhesus monkeys, crab-eating monkeys, Japanese monkeys, and pig-tail monkeys) (Fig. 2B). Figure 3 depicts that a canonical 3' splice site is well conserved in the *AluYRa2* sequences, and a 5' splice site is also well conserved on the 3' flanking sequence of the *AluYRa2* element in Japanese monkeys, rhesus monkeys, crab-eating monkeys, and pig-tail monkeys.

To estimate the integration time of *BCS1L*_AluYRa2, we performed the NETWORK analysis (Bandelt et al., 1999). The age of *BCS1L*_AluYRa2 of rhesus monkey are estimated to be in the range of 4.57-5.49 myrs. And also, we additionally calculate the integration time of *BCS1L*_AluYRa2(t_1) using the divergence time from other *AluYRa2*. From the phylogenetic tree analysis, most similar *AluYRa2* (ms*AluYRa2*) was selected among the one hundred twenty intact *AluYRa2* sequences (Fig. S2). And then, we calculated pairwise distance between *BCS1L*_AluYRa2 and consensus *AluYRa2* ($d_{AluYRa2} = 0.042$), and *BCS1L*_AluYRa2 and ms*AluYRa2* ($d_{BCS1L_AluYRa2} = 0.033$). Previous study indicated that *AluYRa2* was integrated 8.6 million years ago (t) in the macaque ge-

nome (Han et al., 2007a). According to calculation of divergence time ($t_1 = (d_{BCS1L_AluYRa2}/d_{AluYRa2})t$), *AluYRa2* was integrated in the *BCS1L* gene region about 6.76 myrs (Chou et al., 2002).

Expression pattern of the *BCS1L* transcript and experimental validation of the *AluYRa2*-derived exonization event

In order to understand the expression features of the *BCS1L* gene and validate the *AluYRa2*-derived exonization event on the *BCS1L* gene, we performed RT-PCR experiments and sequenced the products using transcript-specific primer pairs in 20 different human tissues, seven different rhesus monkey tissues, and five different crab-eating monkey tissues (Figs. 3 and Supplementary Fig. S3). We could not perform RT-PCR in the Japanese and pig-tail monkeys because we did not have RNA samples of these two species. In the human tissues, to detect the two representative transcripts (NM_004328.4 and NM_001079866.1) of the human *BCS1L* gene, each transcript-specific primer pair was designed based on the exon junctions using the Primer3 program (Supplementary Table S1). The results showed that the two transcripts of the *BCS1L* gene have ubiquitous expression patterns in all of the tissues. Moreover, the *BCS1L* gene showed a ubiquitous expression pattern in the rhesus monkeys. On the other hand, different expression patterns and different sized product bands were found in each tissue sample of the crab-eating monkeys (Fig. 4). The sequencing data indicated that 10 total transcript variants were identified in the rhesus monkeys (V1-V3) and crab-eating monkeys (V4-V10) (Fig. 4 and Supplementary Fig. S3). Overall, based on comparative structural analysis, most of transcript variants showed different structures as the result of exonization, exon skipping, and different 5' splice sites. V2-V7 transcripts showed different lengths of the first exon, and V3, V6, V7, and V10 had a 25-bp short exon (E4). In particular, only the V1, V8,

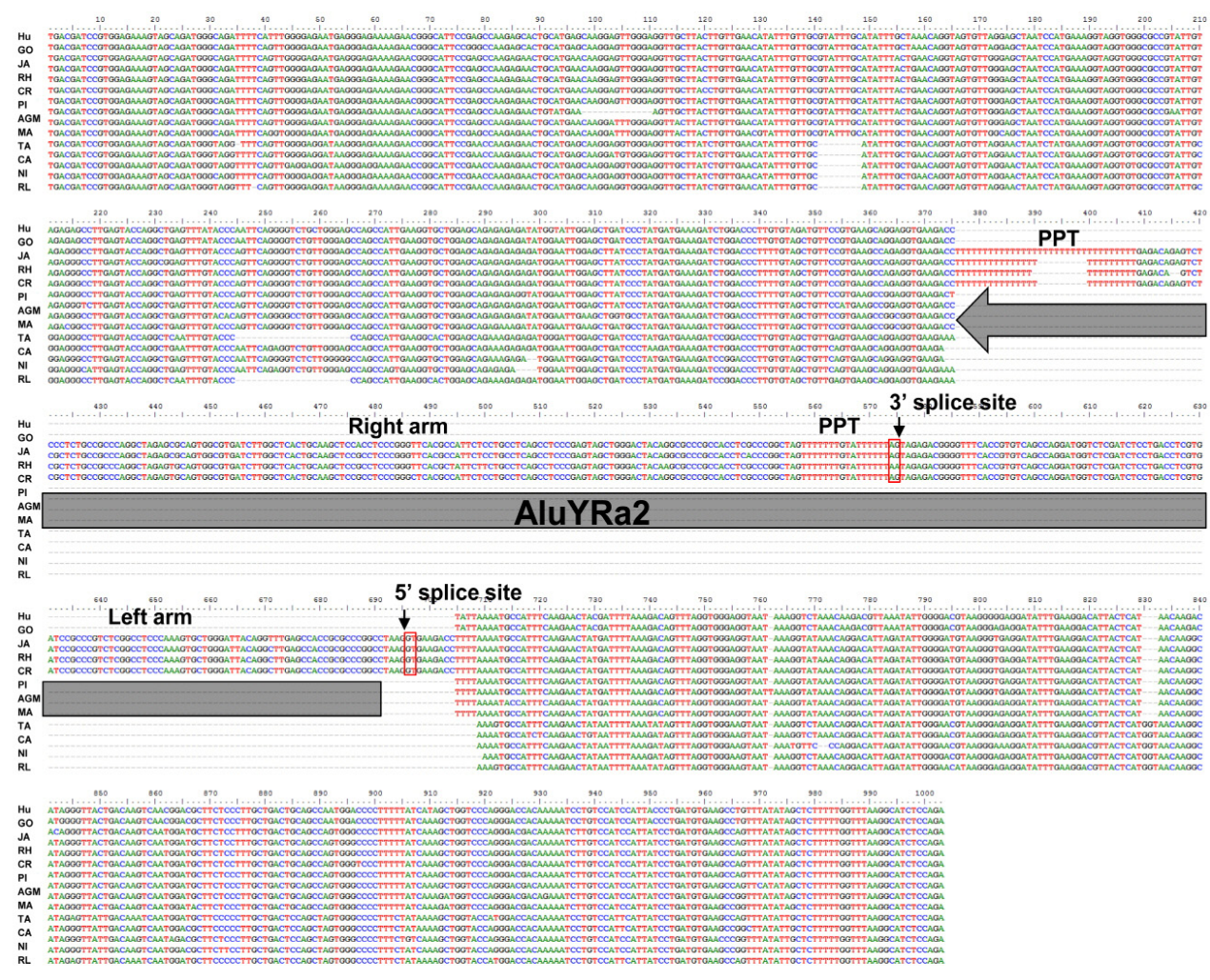


Fig. 3. Multiple nucleotide sequence alignment of the integrated *AluYRa2* region among different species. The gray arrow indicates the *AluYRa2* element. The species analyzed were: (1) HU: human; (2) hominoids: CH: chimpanzee, BO: bonobo, GO: gorilla; (3) Old World monkeys: RH: rhesus monkey, JA: Japanese monkey, CR: crab-eating monkey, PI: pig-tail monkey, AF: African green monkey, MA: mandrill, CO: colobus, LA: langur; and (4) New World monkeys: MAR: Marmoset, TA: tamarin, CA: capuchin, SQ: squirrel monkey, NI: night monkey, SP: spider monkey; and (5) prosimian: RL: ring-tailed lemur. PPT, polypyrimidine tract.

V9, and V10 transcripts had the *AluYRa2*-derived exon; among them, the V1 and V9 transcripts showed the same gene structure as the transcript from the rhesus monkey cDNA library, and the V8 transcript had one more exon (E3) between second and third exon.

To demonstrate the transcriptional start site (TSS) and polyadenylation sequence of the *BCS1L* gene in rhesus monkeys, we conducted bidirectional RACE analyses in the placental tissue of rhesus monkeys (Fig. 6). As a result, we obtained seven different transcripts (T1-T7), and all of these transcripts showed different TSSs except for T2 and T6. However, the result of 3' RACE analysis only showed the same sequences. Additionally, AATAAA was detected as the polyadenylation signal motif and polyadenylation tail at the 3' end of the mRNA sequence (Supplementary Fig. S4). Specifically, only T1 and T2 had a 369 bp-long second exon (E3), and other transcripts showed exon skipping of this exon. Although the seven transcript variants were identified in the placental tissue of rhesus monkeys, the *AluYRa2*-exonized transcript was not detected by 5' RACE.

DISCUSSION

More than 40% of the human genome is comprised of TEs. According to previous studies, TEs can influence transcription and biological function of genes via various mechanisms (Gogvadze and Buzdin, 2009; Han et al., 2007b; Park et al., 2012). For these reasons, TEs play an important role in evolutionary processes and genome diversification. Moreover, about 90% of TEs are retroelements in the human genome (Bannert and Kurth, 2004). Of the retroelements, about 8% are human endogenous retrovirus (HERV), 20% are long interspersed elements (LINEs), and 13% are short interspersed elements (SINEs) (Schmitz and Brosius, 2011). *Alu* elements are primate-specific SINEs, and the human genome has more than 1.1 million copies of these elements (Batzer and Deininger, 2002). This large number of *Alu* elements can influence functional genes that may cause several genetic disorders when *Alu*-mediated recombination occurs (Deininger and Batzer, 1999). Alternatively, these elements can lead to AS events and

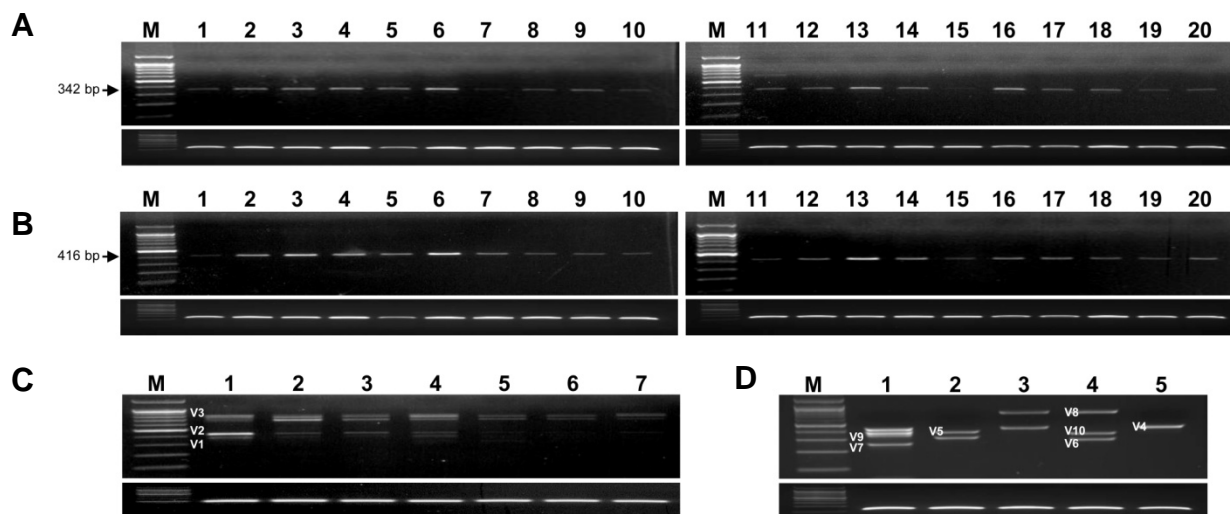


Fig. 4. RT-PCR amplification of the *BCS1L* gene. (A) NM_004328.4 transcript (342 bp) of a human, (B) NM_001079866.1 transcript (416 bp) of a human, (C) rhesus monkey, and (D) crab-eating monkey. RPL32 (80 bp) indicates the positive control. M indicates the size marker. Human cDNA samples are as follows: 1: bone marrow; 2: brain (whole); 3: fetal brain; 4: colon; 5: small intestine; 6: heart; 7: kidney; 8: liver; 9: fetal liver; 10: lung; 11: placenta; 12: prostate; 13: skeletal-muscle; 14: spinal cord; 15: spleen; 16: stomach; 17: testis; 18: thymus; 19: trachea; 20: uterus. Rhesus monkey cDNA samples are as follows: 1: brain; 2: colon; 3: liver; 4: lung; 5: kidney; 6: pancreas; 7: stomach. Crab-eating monkey cDNA samples are as follows: 1: brain; 2: heart; 3: kidney; 4: liver; 5: lung. V1-V10 transcripts were validated by sequencing the RT-PCR products.

exonization by provision of new splice sites in the intragenic regions. Furthermore, intergenic or intragenic *Alu* elements could provide polyadenylation sites, promoters, enhancers, and silencers for a number of adjacent functional genes (Lee et al., 2009). Therefore, although *Alu* elements are harmful elements for host genome survival, they can be important sources that enhance genome diversity. Additionally, they could lead to the creation of novel transcript variants and protein isoforms in primate genomes (Li et al., 2001; Sorek et al., 2002).

In our previous study, we identified an *AluYRa2*-exonized *BCS1L* gene in the placental cDNA library of the rhesus monkey (Kim et al., 2010). This *AluYRa2*-derived transcript was not found in the human genome, because integration of the *AluYRa2* element did not occur in the human genome (Fig. 1). During primate evolution, about 110,000 OWM-specific *Alu* elements existed in the OWM genomes, and a total of 14 different OWM lineage-specific *AluY* subfamilies such as *AluYRa*, *AluYRb*, *AluYRc*, and *AluYRd* were identified (Han et al., 2007a). Among them, insertion of *AluYRa2* elements was found in Cercopithecinae species (rhesus monkeys, pig-tail monkeys, Japanese monkeys, olive baboons, and green monkeys) and 11 other subfamilies (including YRa3, YRa4, YRb2-4, YRc1-2, and YRd1-4) are restricted to *Macaca* species. However, our results showed that only *Macaca* species, including Japanese monkeys, rhesus monkeys, crab-eating monkeys, and pig-tail monkeys, have an *AluYRa2* element insertion; this insertion was not present in other species such as hominoids, the other OWM (mandrills and African-green monkeys), NWM, and prosimians (Fig. 2). These results were not consistent with those of a previous study (Han et al., 2007a). Therefore, to resolve this incongruence, integration events of the *AluYRa2* element and identification analysis in different loci should be investigated in a large number of *Mandrillus* and *Chlorocebus*

populations.

Our genomic sequence analysis demonstrated integration of an antisense sequence analysis demonstrated integration of an antisense-oriented *AluYRa2* element in the first intron region of the rhesus monkey *BCS1L* gene (Supplementary Fig. S3). The *Alu* element consists of two similar monomers, the left and right arms, and this monomer is joined by an A-rich linker and followed by a poly(A) tail (Gal-Mark et al., 2008). The majority of *Alu*-derived exonization events are identified by antisense orientation of the *Alu* element, and the right arm is the main component where exonization events occur (Kreahling and Graveley, 2004; Lev-Maor et al., 2003; Sorek et al., 2002). However, exonization events can sometimes occur on the left arm of the *Alu* element, because the right and left arms have high sequence similarity (Gal-Mark et al., 2008). Our results revealed that the exonization event occurred in the left arm of *AluYRa2* and the 3' flanking sequence in rhesus monkeys and crab-eating monkeys (Fig. 4 and Supplementary Fig. S3). This element could provide recognition sequences for spliceosomes, such as 3' and 5' splice sites. Additionally, Japanese and pig-tail monkeys had conserved 3' and 5' splice sites, their boundary sequence, and polypyrimidine tract (PPT). Therefore, we predict that an *AluYRa2*-derived exonization event occurred in Japanese and pig-tail monkeys. However, this phenomenon needs to be confirmed by RT-PCR experiments.

RT-PCR experiments showed that the *BCS1L* gene was ubiquitously expressed in all tissues in humans and rhesus monkeys. However, different expression patterns and different sized bands were found in the five crab-eating monkey tissues (Fig. 4). These products were analyzed by cloning and sequencing procedures (Supplementary Fig. S3). As a result, we confirmed that not only an *AluYRa2*-derived exonization event but also AS events such as exon skipping and alternative 3' splice sites in rhesus and crab-eating monkeys. Moreover, most

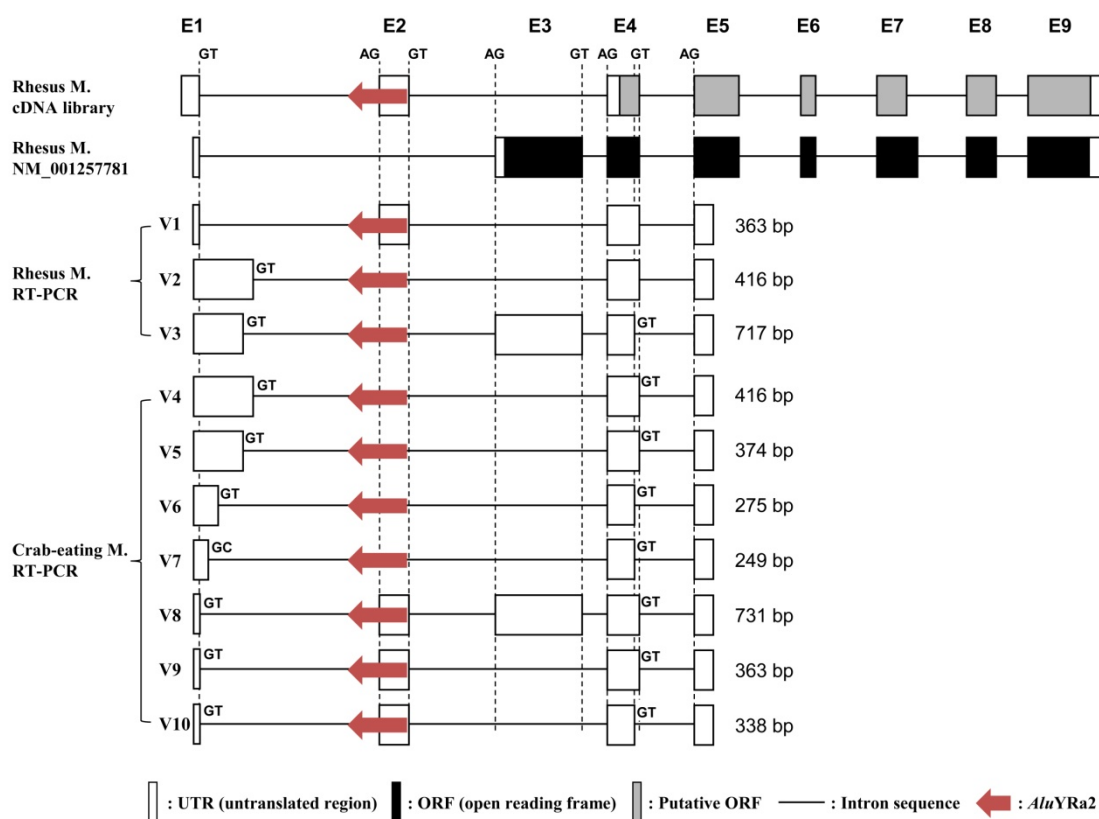


Fig. 5. Structural analysis of rhesus monkey and crab-eating monkey *BCS1L* gene transcripts by RT-PCR experiments and sequencing of the products. A total of 10 transcripts were identified from various tissues of these two species. E1-E9 indicate the consensus exon of the *BCS1L* gene. The antisense-oriented *AluYRa2* element is located on the first intron of *BCS1L* in rhesus monkeys. Open, black, and gray boxes represent the untranslated region of the exons, protein-coding region, and putative protein-coding region, respectively. Vertical dashed lines indicate the 3' (GT) and 5' (AG) splice sites and the horizontal line indicates intron sequences. This figure is a structural illustration and is not to scale.

of these events occurred in the 5' region (E1-E4) of the *BCS1L* gene (Fig. 5). In particular, the crab-eating monkey showed more abundant transcript variants as the result of AS events compared with humans and rhesus monkeys. Moreover, tissue-specific AS events were found in all five tissues of the crab-eating monkey. For example, the V4 transcript was found in the brain, kidney, and lung tissues, the V7 transcript in brain tissue, and the *AluYRa2*-derived transcript was found in the V8, V9, and V10 transcripts was also detected in all of the tissue samples (Fig. 4). On the contrary, tissue-specific AS events were not detected in the seven rhesus monkey tissues. This phenomenon of discordance between rhesus and crab-eating monkeys could be explained by different accumulation of mutated sequences in the cis elements or intronic regions after divergence from a common ancestor. In particular, the cis-sequence changes and variation of intronic sequence could lead to species-specific splicing patterns in the *MAGOH* gene (Lin et al., 2010). The genomic DNA sequence of the *BCS1L* gene region (E1-E4 region, except for the *AluYRa2* sequence) showed high sequence similarity between human and rhesus monkeys (94.5% identity), human and crab-eating monkeys (94.5% identity), and rhesus and crab-eating monkeys (99.4% identity) (data not shown). Therefore, we believe that this minor difference could lead to species-specific splicing patterns in the *BCS1L* gene. However,

further studies are needed to demonstrate the validity of this species-specific AS events.

Integration of *Alu* elements in the intragenic region could lead to the production of different amino acid sequences by exonization events in associated genes (Park et al., 2012). Our six-frame translation sequence analysis results confirmed that the *AluYRa2*-exonized transcript encoded 299 amino acids, and this sequence was shorter than in human, mouse, and rhesus monkey reference gene sequences in the N-terminal region (Supplementary Fig. S1). These short amino acids were derived from AS events, including *AluYRa2*-derived exonization and E3 exon skipping in the 5' region of *BCS1L* (Fig. 5). Alternatively, the amino acid sequence of the C-terminal region was highly conserved in the *AluYRa2*-exonized *BCS1L* transcript. *BCS1L* has two functional domains including a sorting sequence of the N-terminal region (1-89 amino acids) and AAA-family ATPase (220-399 amino acids) (Supplementary Fig. S1) (Hinson et al., 2007). This sorting sequence consists of an intermembrane space (IMS), transmembrane (TM) stretch, and N-linker region. Additionally, they are related to translocation and insertion into the mitochondrial inner membrane (Wagener and Neupert, 2012). Therefore, the newly identified *AluYRa2*-exonized transcript likely does not have an important role in mitochondria because it is not inserted into the mitochondria. To demonstrate this phenomenon, further functional studies are

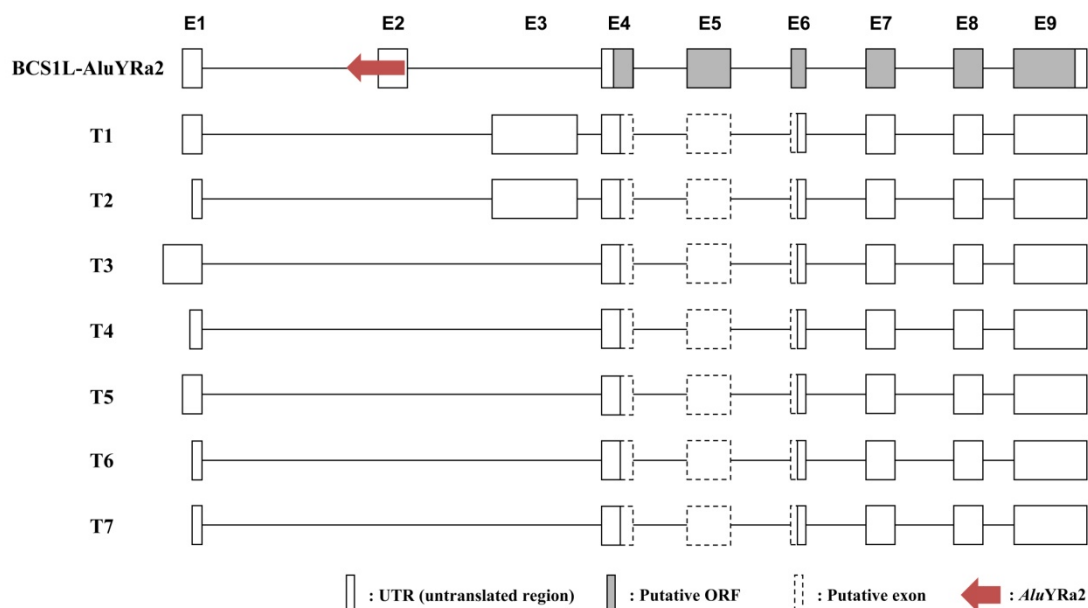


Fig. 6. Structural analysis of *BCS1L* transcript variants by bidirectional RACE experiments and sequencing of the products. A total of seven (T1-T7) different transcripts were identified from placental tissues of rhesus monkeys. The antisense-oriented *AluYRa2* element is located on the first intron of *BCS1L* in rhesus monkeys. Open, gray, and dashed boxes represent the untranslated region of the exons, putative protein-coding region, and putative exon, respectively. A horizontal line indicates the intron sequences. This figure is a structural illustration and is not to scale.

needed using various transcript variants of the *BCS1L* gene in rhesus and crab-eating monkeys. However, one definite result is the presence of novel transcripts by an *AluYRa2*-derived exonization event and other AS events on the *BCS1L* gene, and they produce different protein isoforms in these two primate species. As a result, these AS events could be a novel source of genome diversity.

Rhesus and crab-eating monkeys are the most valuable non-human primate animal model species for biomedical research, particularly in the fields of microbiology, neuroscience, and biochemistry (Gibbs et al., 2007; Lee et al., 2014; Park et al., 2013; 2015). The genus *Macaca* diverged with humans from a common ancestor between 25 and 31 million years ago (Kumar and Hedges, 2011). Thus, these two species have biological and behavioral similarities as well as a close genetic relationship (Carlsson et al., 2004; Huh et al., 2012). This evolutionary relationship could make these species more appropriate experimental animal models than other animals including rodents, rabbits, and dogs. Previous studies demonstrated that mutations of the *BCS1L* gene could lead to several diseases, such as GRACILE syndrome, Björnstads syndrome, mitochondrial encephalopathy, neonatal tubulopathy, encephalopathy, liver failure, and complex III deficiency. Therefore, our identification of full-length transcripts and transcript variants as well as assessment of expression pattern in rhesus and crab-eating monkeys could provide basic and useful information for investigation of various *BCS1L* gene-related diseases using non-human primate animal models.

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

ACKNOWLEDGMENTS

This research was supported by a grant from the Korea Research Institute of Bioscience and Biotechnology (KRIBB) Research Initiative Program (KGM4241541).

REFERENCES

- Alekseyenko, A.V., Kim, N., and Lee, C.J. (2007). Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA* 13, 661-670.
- Almeida, L.M., Silva, I.T., Silva, W.A., Jr., Castro, J.P., Riggs, P.K., Carareto, C.M., and Amaral, M.E. (2007). The contribution of transposable elements to *Bos taurus* gene structure. *Gene* 390, 180-189.
- Amit, M., Sela, N., Keren, H., Melamed, Z., Muler, I., Shomron, N., Izraeli, S., and Ast, G. (2007). Biased exonization of transposed elements in duplicated genes: a lesson from the TIF-IA gene. *BMC Mol. Biol.* 8, 109.
- Ast, G. (2004). How did alternative splicing evolve? *Nat. Rev. Genet.* 5, 773-782.
- Bandelt, H.J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37-48.
- Bannert, N., and Kurth, R. (2004). Retroelements and the human genome: new perspectives on an old relation. *Proc. Natl. Acad. Sci. USA* 101 Suppl 2, 14572-14579.
- Batzer, M.A., and Deininger, P.L. (2002). *Alu* repeats and human genomic diversity. *Nat. Rev. Genet.* 3, 370-379.
- Carlsson, H.E., Schapiro, S.J., Farah, I., and Hau, J. (2004). Use of primates in research: a global overview. *Am. J. Primatol.* 63, 225-237.
- Chou, H.H., Hayakawa, T., Diaz, S., Krings, M., Indriati, E., Leakey, M., Paabo, S., Satta, Y., Takahata, N., and Varki, A. (2002). Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl. Acad. Sci. USA* 99, 11736-11741.

- Corvelo, A., and Eyraes, E. (2008). Exon creation and establishment in human genes. *Genome Biol.* 9, R141.
- de Lonlay, P., Valnot, I., Barrientos, A., Gorbatyuk, M., Tzagoloff, A., Taanman, J.W., Benayoun, E., Chretien, D., Kadhom, N., Lombes, A., et al. (2001). A mutant mitochondrial respiratory chain assembly protein causes complex III deficiency in patients with tubulopathy, encephalopathy and liver failure. *Nat. Genet.* 29, 57-60.
- DeBarry, J.D., Ganko, E.W., McCarthy, E.M., and McDonald, J.F. (2006). The contribution of LTR retrotransposon sequences to gene evolution in *Mus musculus*. *Mol. Biol. Evol.* 23, 479-481.
- Deininger, P.L., and Batzer, M.A. (1999) Alu repeats and human disease. *Mol. Genet. Metab.* 67, 183-193.
- Fernandez-Vizarrá, E., Bugiani, M., Goffrini, P., Carrara, F., Farina, L., Procopio, E., Donati, A., Uziel, G., Ferrero, I., and Zeviani, M. (2007). Impaired complex III assembly associated with BCS1L gene mutations in isolated mitochondrial encephalopathy. *Hum. Mol. Genet.* 16, 1241-1252.
- Gal-Mark, N., Schwartz, S., and Ast, G. (2008). Alternative splicing of Alu exons—two arms are better than one. *Nucleic Acids Res.* 36, 2012-2023.
- Gibbs, R.A., Rogers, J., Katze, M.G., Bumgarner, R., Weinstock, G.M., Mardis, E.R., Remington, K.A., Strausberg, R.L., Venter, J.C., Wilson, R.K., et al. (2007). Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316, 222-234.
- Gogvadze, E., and Buzdin, A. (2009). Retroelements and their impact on genome evolution and functioning. *Cell Mol. Life Sci.* 66, 3727-3742.
- 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. (2007a). Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. *Science* 316, 238-240.
- Han, K., Lee, J., Meyer, T.J., Wang, J., Sen, S.K., Srikanta, D., Liang, P., and Batzer, M.A. (2007b). Alu recombination-mediated structural deletions in the chimpanzee genome. *PLoS Genet.* 3, 1939-1949.
- Hanson, P.I., and Whiteheart, S.W. (2005). AAA+ proteins: have engine, will work. *Nat. Rev. Mol. Cell Biol.* 6, 519-529.
- Hinson, J.T., Fantin, V.R., Schonberger, J., Breivik, N., Siem, G., McDonough, B., Sharma, P., Keogh, I., Godinho, R., Santos, F., et al. (2007). Missense mutations in the BCS1L gene as a cause of the Bjornstad syndrome. *N Engl. J. Med.* 356, 809-819.
- Huh, J.W., Kim, Y.H., Lee, S.R., Kim, H., Kim, D.S., Kim, H.S., Kang, H.S., and Chang, K.T. (2009). Gain of new exons and promoters by lineage-specific transposable elements-integration and conservation event on CHRM3 gene. *Mol. Cells* 28, 111-117.
- Huh, J.W., Kim, Y.H., Lee, S.R., Kim, D.S., Park, S.J., Kim, H., Kim, J.S., Song, B.S., Kim, H.S., and Chang, K.T. (2010). Four different ways of alternative transcripts generation mechanism in ADRA1A gene. *Genes Genet. Syst.* 85, 65-73.
- Huh, J.W., Kim, Y.H., Park, S.J., Kim, D.S., Lee, S.R., Kim, K.M., Jeong, K.J., Kim, J.S., Song, B.S., Sim, B.W., et al. (2012). Large-scale transcriptome sequencing and gene analyses in the crab-eating macaque (*Macaca fascicularis*) for biomedical research. *BMC Genomics* 13, 163.
- Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* 11, 345-355.
- Kim, D.S., Huh, J.W., Kim, Y.H., Park, S.J., Lee, S.R., and Chang, K.T. (2010). Full-length cDNA sequences from Rhesus monkey placenta tissue: analysis and utility for comparative mapping. *BMC Genomics* 11, 427.
- Kim, T., Kim, J.O., Oh, J.G., Hong, S.E., and Kim do, H. (2014). Pressure-overload cardiac hypertrophy is associated with distinct alternative splicing due to altered expression of splicing factors. *Mol. Cells* 37, 81-87.
- Kreahling, J., and Graveley, B.R. (2004). The origins and implications of alternative splicing. *Trends Genet.* 20, 1-4.
- Kumar, S., and Hedges, S.B. (2011). TimeTree2: species divergence times on the iPhone. *Bioinformatics* 27, 2023-2024.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.
- Lee, J.R., Huh, J.W., Kim, D.S., Ha, H.S., Ahn, K., Kim, Y.J., Chang, K.T., and Kim, H.S. (2009). Lineage specific evolutionary events on SFTP gene: Alu recombination-mediated deletion (ARMD), exonization, and alternative splicing events. *Gene* 435, 29-35.
- Lee, Y., Kim, Y.H., Park, S.J., Huh, J.W., Kim, S.H., Kim, S.U., Kim, J.S., Jeong, K.J., Lee, K.M., Hong, Y., et al. (2014). Insulin/IGF signaling-related gene expression in the brain of a sporadic Alzheimer's disease monkey model induced by intracerebroventricular injection of streptozotocin. *J. Alzheimers Dis.* 38, 251-267.
- Lev-Maor, G., Sorek, R., Shomron, N., 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., Jiang, P., Sato, S., Davidson, B.L., and Xing, Y. (2010). Evolution of alternative splicing in primate brain transcriptomes. *Hum. Mol. Genet.* 19, 2958-2973.
- 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.* 4, e1000225.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413-1415.
- Park, S.J., Huh, J.W., Kim, Y.H., Kim, H.S., and Chang, K.T. (2012). Intron Retention and TE Exonization Events in ZRANB2. *Comp. Funct. Genomics* 2012, 170208.
- Park, S.J., Kim, Y.H., Huh, J.W., Lee, S.R., Kim, S.H., Kim, S.U., Kim, J.S., Jeong, K.J., Kim, K.M., Kim, H.S., et al. (2013). Selection of new appropriate reference genes for RT-qPCR analysis via transcriptome sequencing of cynomolgus monkeys (*Macaca fascicularis*). *PLoS One* 8, e60758.
- Park, S.J., Kim, Y.H., Nam, G.H., Choe, S.H., Lee, S.R., Kim, S.U., Kim, J.S., Sim, B.W., Song, B.S., Jeong, K.J., et al. (2015). Quantitative expression analysis of APP pathway and tau phosphorylation-related genes in the ICV STZ-induced non-human primate model of sporadic Alzheimer's disease. *Int. J. Mol. Sci.* 16, 2386-2402.
- Piriyapongsa, J., Polavarapu, N., Borodovsky, M., and McDonald, J. (2007). Exonization of the LTR transposable elements in human genome. *BMC Genomics* 8, 291.
- Schmitz, J., and Brosius, J. (2011). Exonization of transposed elements: A challenge and opportunity for evolution. *Biochimie* 93, 1928-1934.
- Sela, N., Mersch, B., Hotz-Wagenblatt, A., and Ast, G. (2010). Characteristics of transposable element exonization within human and mouse. *PLoS One* 5, e10907.
- Snider, J., and Houry, W.A. (2008). AAA+ proteins: diversity in function, similarity in structure. *Biochem. Soc. Trans.* 36, 72-77.
- Sorek, R., Ast, G., and Graur, D. (2002). Alu-containing exons are alternatively spliced. *Genome Res.* 12, 1060-1067.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725-2729.
- Tucker, P.A., and Sallai, L. (2007). The AAA+ superfamily—a myriad of motions. *Curr. Opin. Struct. Biol.* 17, 641-652.
- Visapaa, I., Fellman, V., Vesa, J., Dasvarma, A., Hutton, J.L., Kumar, V., Payne, G.S., Makarow, M., Van Coster, R., Taylor, R.W., et al. (2002). GRACILE syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. *Am. J. Hum. Genet.* 71, 863-876.
- Wagener, N., and Neupert, W. (2012). Bcs1, a AAA protein of the mitochondria with a role in the biogenesis of the respiratory chain. *J. Struct. Biol.* 179, 121-125.