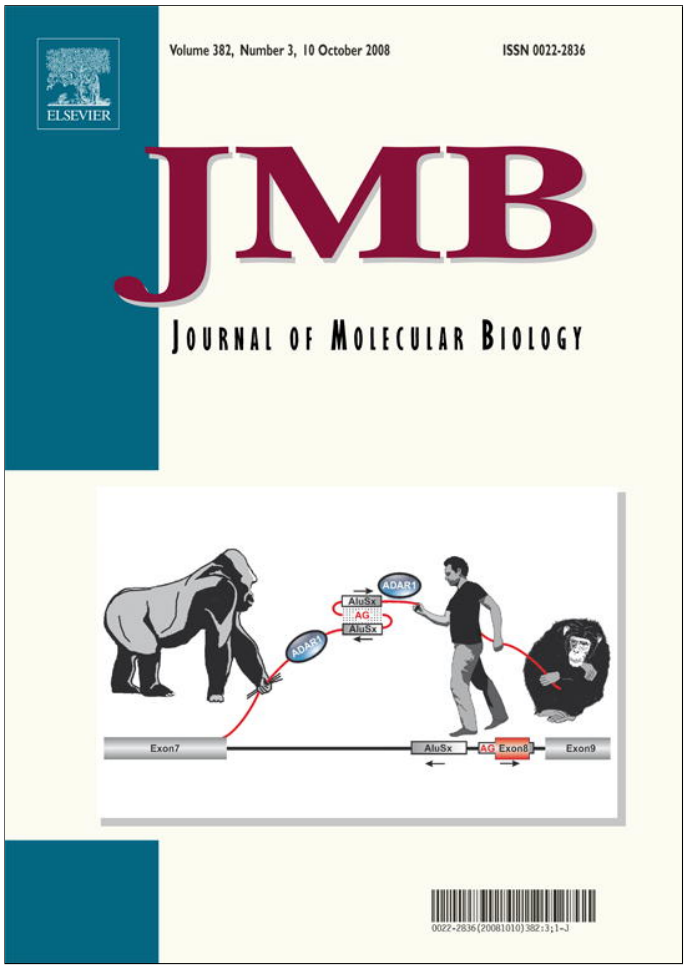

doi:10.1016/j.jmb.2008.07.014 J. Mol. Biol. (2008) 382, 601–609

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

# Beyond DNA: RNA Editing and Steps Toward *Alu* Exonization in Primates

**Maren Möller-Krull[1], Anja Zemann[1], Christian Roos[2], Jürgen Brosius[1] and Jürgen Schmitz[1]***

[1]*Institute of Experimental Pathology (ZMBE), University of Münster, Von-Esmarch-Str. 56, 48149 Münster, Germany*

[2]*Gene Bank of Primates, Primate Genetics, German Primate Centre, Kellnerweg 4, 37077 Göttingen, Germany*

*Received 15 February 2008; received in revised form 30 June 2008; accepted 8 July 2008*
Available online 16 July 2008

The exaptation of transposed elements into protein-coding domains by a process called exonization is one important evolutionary pathway for generating novel variant functions of gene products. Adenosine-to-inosine (A-to-I) modification is a recently discovered, RNA-editing-mediated mechanism that contributes to the exonization of previously unprocessed mRNA introns. In the human nuclear prelamin A recognition factor gene transcript, the alternatively spliced exon 8 results from an A-to-I editing-generated 3′ splice site located within an intronic *Alu* short interspersed element. Sequence comparisons of representatives of all primate infraorders revealed the critical evolutionary steps leading to this editing-mediated exonization. The source of exon 8 was seeded within the primary transcript about 58–40 million years ago by the head-to-head insertions of two primate-specific *Alu* short interspersed elements in the common ancestor of anthropoids. The latent protein-coding potential was realized 34–52 million years later in a common ancestor of gorilla, chimpanzee, and human as a result of numerous changes at the RNA and DNA level. Comparisons of 426 processed mRNA clones from various primate species with their genomic sequences identified seven different RNA-editing-mediated alternative splice variants. In total, 30 A-to-I editing sites were identified. The gorilla, chimpanzee, and human nuclear prelamin A recognition factor genes exemplify the versatile interplay of pre- and posttranscriptional modifications leading to novel genetic potential.



*Keywords:* nuclear prelamin A recognition factor (NARF); A-to-I RNA-editing; alternative splicing; *Alu* exonization

***Edited by J. Karn***

## Introduction

Evolution promotes new genetic function by at least two different ways: first, by triggering formation of new genes *via* various duplication events and, second, by including new exons in existing genes. New exons arise, for example, by exonization of intronic sequences, and often are composed, usually in part, of transposed elements. Such elements invade predominantly non-protein-coding regions and are either fortuitously exonized soon after integration or millions of years thereafter by acquisition of critical changes.[1,2] Exonized sequences, usually non-protein coding in their prior existence, contribute to protein-coding domains predominantly by alternative[1,3] and, far less frequently, by constitutive splicing.[4,5] In primates, predominantly *Alu* short interspersed elements are recruited as novel exons.[1,6] *Alu* retroposons represent a 7SL RNA-derived primate-specific repeat family.[7] They distributed during different temporal waves of primate evolution, leaving behind more than 1.1 million copies in the human genome.[5] The major waves of primate *Alu* insertion activity consist of the old *Alu*J subfamily [including *Alu*Jo and *Alu*Jb; about 81 million years ago (MYA)], the middle-aged *Alu*S subfamily (including *Alu*S, *Alu*Sx, *Alu*Sg, *Alu*Sq, *Alu*Sp, *Alu*Sc, and *Alu*Sb; about 48–19 MYA), and the

**Corresponding author.* E-mail address: jueschm@uni-muenster.de.

Present address: M. Möller-Krull, Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

Abbreviations used: A-to-I, adenosine-to-inosine; MYA, million years ago; ADAR, adenosine deaminase that acts on RNA; dsRNA, double-stranded RNA; NARF, nuclear prelamin A recognition factor; ORF, open reading frame; RT, reverse transcription; FLAM, fossil left Alu monomer; NMD, nonsense-mediated decay.

young and still active *Alu*Y wave (including *Alu*Y, *Alu*Ya5, *Alu*Ya8, and *Alu*Yb8; starting about 6 MYA and continuing to the present).[8]

Recently, it was demonstrated that certain *Alu* exonizations might depend particularly on adenosine-to-inosine (A-to-I) RNA editing,[6] whereby an adenosine is substituted by an inosine that is then recognized as a guanosine.[9] A-to-I RNA editing can generate, delete, or alter the meaning of triplet codons as well as permit new splice sites.[10,11]

The editing reaction is catalyzed by members of the enzyme family known as adenosine deaminases that act on RNA (ADARs).[12] ADARs are double-stranded RNA (dsRNA)-specific enzymes that generally act on duplexes longer than 30 base pairs.[13,14] They are composed of a domain structure that includes variable numbers of dsRNA binding motifs. Two ADAR enzymes (ADAR1 and 2), varying in their number of dsRNA binding motifs and their editing specificities, have been identified in mammals.[15,16] ADAR2 preferentially edits at a single nucleotide or at few nucleotides, converting the arginine codon (AGA) to glycine (IGA = GGA), whereas ADAR1 edits numerous adenosines in RNA duplexes in a more promiscuous way.[17,18] However, the immediate structural surroundings of the edited site appears to be more important than the sequence motif itself.[6,18]

Human *Alu* sequences appear to be preferred targets for A-to-I editing, a phenomenon that may be explained by their high incidence (e.g., 90% of all known human genes contain *Alu* insertions that, in case they are in tandem inverse orientation, can form intermolecular dsRNA structures that are subject to ADAR editing).[13] Recently, Lev-Maor *et al.* described the A-to-I RNA editing and resultant alternative splicing of one such *Alu* element that led to the birth of exon 8 of the human nuclear prelamin A recognition factor (NARF; Fig. 1).[6] This posttranscriptional editing generates a 3′ splice site functionally equivalent to AG that facilitates exonization and eliminates a premature stop codon important for maintaining an open reading frame (ORF). The newly described splice variant is expressed in diverse human cell lines and tissues.[6]

Although the above study demonstrated a distinctly novel pathway to generate novel gene domains, the evolutionary steps leading to such an event and its primate wide distribution have yet to be delineated. In the present study, we successfully PCR-amplified the NARF gene between exons 7 and 9, including their embedded intron(s), in repre-

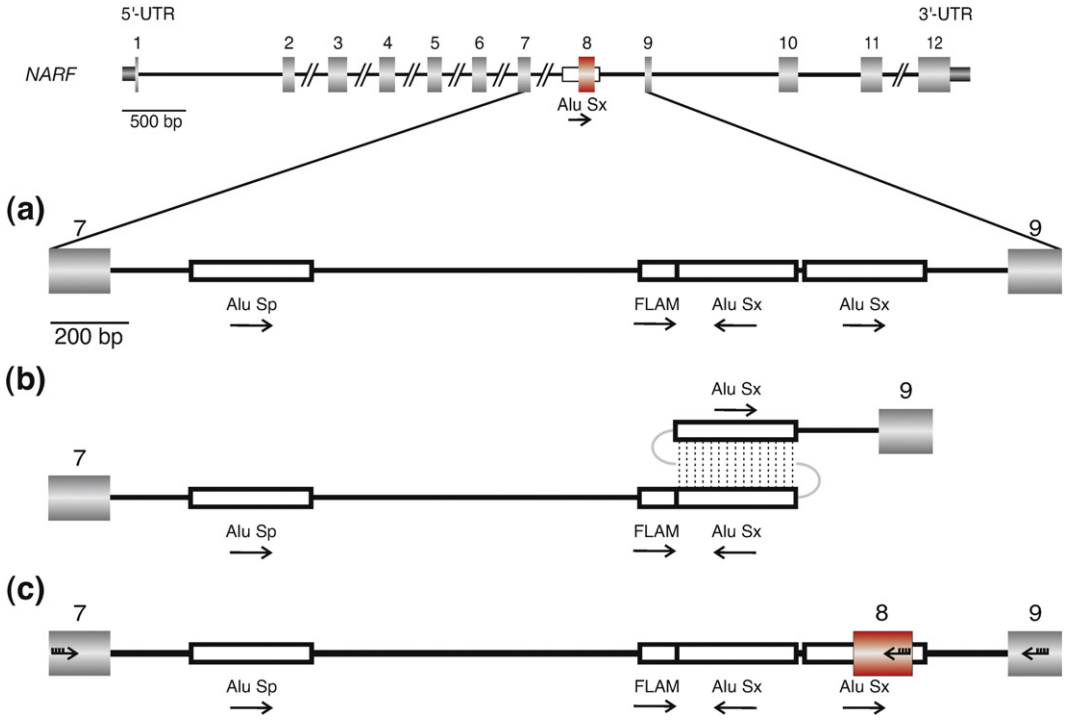


**Fig. 1.** Posttranscriptional RNA editing and gain of a protein-coding alternative exon. Structure of the complete NARF gene harboring the internal, exonized, sense-oriented *Alu*Sx element (top). Numbered boxes represent protein-coding exons (CDS). The 5′ and 3′ untranslated regions (UTRs) are shown as thick gray lines. Introns are indicated as black lines; double slashes denote gaps in larger introns. The orientations of transposed elements are indicated by arrows. White boxes represent the intronic parts of transposed elements, and the red area denotes the exonized protein-coding region. The framed area labeled with FLAM indicates an intronic FLAM element, split by the insertion of an *Alu*Sx element (a–c) (about 10 nt of the FLAM element are located between the two *Alu*Sx sequences). (a) Magnification of the region from exon 7 to exon 9 of the gorilla, chimpanzee, and human NARF gene. (b) Scheme of the double-stranded pre-mRNA region, an essential precondition for the ADAR-mediated RNA editing. The predicted dsRNA secondary structure is indicated by dotted lines. (c) Gene structure including alternative spliced exon 8 derived from part of the *Alu*Sx element (red). Arrows within exon boxes indicate the location and orientation of primers.

sentatives of all primate infraorders. Analyses of these products enabled us to outline specific details of the evolutionary history of the described splice variant of NARF, including the insertion of the targeted *Alu* element in the lineage leading to anthropoids. The critical RNA-editing steps leading to exonization of exon 8 were analyzed by reverse transcription (RT)-PCR and cDNA sequencing. We identified 10 alternative splice forms, 7 of which were mediated by A-to-I editing, and dozens of further modification sites. Our analyses enabled a retrospective delineation of the period and mode of exonization.

## Results and Discussion

To compare the prerequisites for alternative splicing and editing at the DNA level, we cloned and sequenced amplicons of approximately 2.6 kb spanning the sequences between exons 7 and 9 of representatives from all primate infraorders. Then, we derived the mRNA sequences coding for the spliced forms by RT-PCR, cloning, and cDNA sequencing of 900 clones. Because the levels of ADAR-editing activities vary from tissue to tissue,[6] it was necessary to analyze the same tissue for all species examined. However, due to the difficulty of obtaining organ tissues, especially for great apes, we were restricted to skeletal muscle, which exhibits the lowest expression of the alternatively spliced exon 8 compared to other tissues,[6] and RT-PCR using primate-specific primers located in exons 7 and 9 failed to detect the alternatively spliced exon 8. Therefore, we also generated species-specific primers for sequences in exon 7 and at the 3′ terminus of exon 8 (Fig. 1; Supplementary Information). We gel-eluted the cDNAs of potential alternative splice products, cloned them, and sequenced a total of 900 cDNAs using a proofreading polymerase. A total of 426 cDNAs contained the expected exonized exon 8 while the other sequences constituted nonspecific amplifications. Only fully overlapping and 100% identical forward and reverse sequence clones were selected for further analysis to ensure error-free comparisons of cDNAs.

### Prosimian intron between NARF exons 7 and 9

Beginning with the prosimians, the first two primate branches preceding the Anthropoidea, the respective intron in strepsirrhines (the first prosimian primate branch represented by *Eulemur coronatus* and *Microcebus murinus*) contains no discernible repeat elements. In Tarsiiformes, the second prosimian divergence (represented by *Tarsius syrichta*), two *Alu*J elements (*Alu*Jb embedded in *Alu*Jo) were inserted in the same (sense) orientation (Fig. 2).

### Platyrrhini intron between NARF exons 7 and 9

A first prerequisite for RNA-editing-mediated exonization was acquired 58–40 MYA following an insertion of two oppositely oriented *Alu*Sx elements

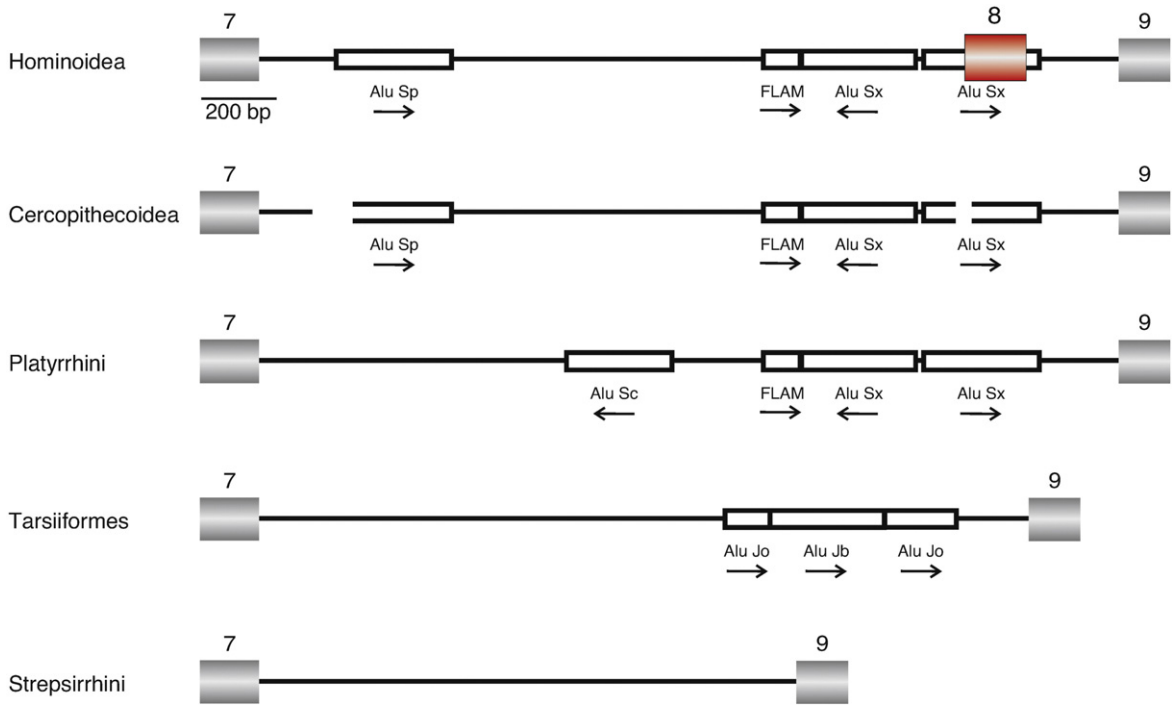


**Fig. 2.** Schematic structure of the NARF region spanning exons 7 to 9 in different primate infraorders. Symbols are analogous to those used in Fig. 1. Two specific deletions within *Alu* elements in Cercopithecoidea are depicted as gaps in the schematic structure. A specific *Alu*S insertion downstream of the antisense *Alu*Sx in *M. mulatta* and an insertion of a long terminal repeat (LTR) element downstream of exon 7 in *T. auratus* are not indicated in the figure. An exonized protein-coding *Alu* exon 8 is only detectable in gorilla, chimpanzee, and human.

presumably in the common ancestor of all Anthropoidea. The intron between exons 7 and 9 in the platyrrhine species (represented by *Callithrix jacchus*, *Saimiri sciureus*, and *Lagothrix lagotricha*) contains the *Alu*Sx duplex together with a fossil left Alu monomer (FLAM) and a platyrrhine, lineage-specific *Alu*Sc element (Fig. 2). The potential precursor 3′ splice site AA dinucleotide that eventually undergoes A-to-I RNA editing and contributes to the generation of *Alu* exon 8 is already present (Fig. 3, positions 9–10). Likewise, the 5′ GT splice site was generated after insertion and is already present in platyrrhines. The *Alu*Sx consensus sequence is −C at this position (Fig. 3, positions 161–162). However, effective exonization in terms of protein-coding function is still not likely because a frameshift (immediately upstream of the 5′ splice site) in the hypothetically exonized domain of the sense-oriented *Alu*Sx sequence would lead to premature stop codons in exon 9. In addition, an in-frame TGA stop codon within the same element is present in *C. jacchus* and *S. sciureus* (Fig. 3, positions 114–116) and an additional deletion of one nucleotide was detected in *L. lagotricha*. The RT-PCRs in *C. jacchus* revealed only an amplicon lacking the *Alu* exon 8.

## Cercopithecoidea intron between NARF exons 7 and 9

In addition to the inserted short interspersed element cluster of FLAM and the two head-to-head *Alu*Sx elements, the Cercopithecoidea (represented

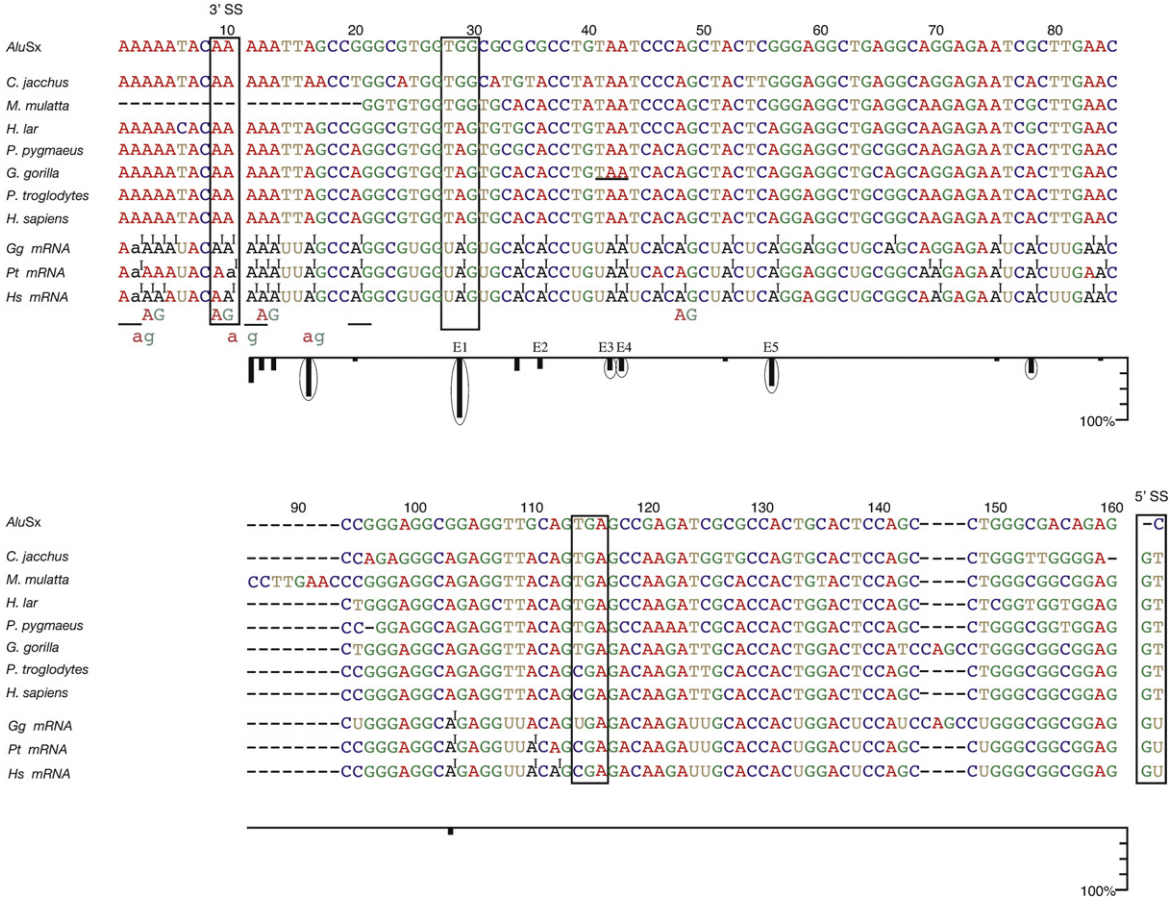


**Fig. 3.** Comparison of genomic DNA and processed mRNA sequences of the NARF*Alu* exonized region of primates. The top sequence represents the *Alu*Sx consensus sequence. Representative genomic DNA sequences are shown for Platyrrhini (*C. jacchus*), Cercopithecoidea (*M. mulatta*), and all Hominoidea (*H. lar*, *P. pygmaeus*, *G. gorilla*, *P. troglodytes*, and *H. sapiens*). The three bottom sequences show the processed mRNAs of gorilla (Gg), chimpanzee (Pt), and human (Hs). RNA editing sites experimentally demonstrated in gorilla, chimpanzee, and human are indicated by black nucleotides labeled with an I, and those described by Lev-Maor *et al.* are indicated as E1–E5.[6] Effective (in-frame) splice sites are labeled with AG for chimpanzee and/or human and ag for gorilla below the mRNA; those leading to truncated forms (out-of-frame for human and chimpanzee) are underlined. The lower case letter a indicates A-to-I editing that generates a functional AG splice site but is not directly detectable. The first boxed area shows the RNA-editing-generated, major alternative 3′ splice site (3′ SS: AA→AG). The second boxed area shows the RNA-edited converted stop codon (TAG→TGG). The third box indicates a substitution at the DNA level removing a premature stop codon (TGA→CGA) in human and chimpanzee. The last boxed area represents the 5′ splice site (5′ SS). This splice site was generated by a genomic change of C (corresponding to the *Alu*Sx consensus sequence) to a GT dinucleotide. The frequencies of gorilla–chimpanzee–human common editing sites, following the predominant splice site (first box), within the 426 analyzed gorilla, chimpanzee, and human cDNAs are given as bars in percentage below the sequences. Encircled frequencies represent previously published high-efficiency *Alu* sequence editing positions[19] (for numerical values, see Supplementary Information).

by *Macaca mulatta*, *Colobus guereza*, and *Trachypithecus auratus*) contain a sense-oriented *Alu*Sp about 250 nt downstream of exon 7. An additional macaque-specific *Alu*S element was detected upstream of the antisense-oriented *Alu*Sx, and *T. auratus* features a specific insertion of 383 nt of a long terminal repeat (LTR) element. Due to a 45-nt deletion in the sense-oriented *Alu*Sx element, including the first 10 nt of the eventual *Alu* exon 8 and the precursor 3′ splice site AA dinucleotide (Fig. 3), effective exonization of the sense *Alu*Sx is unlikely in Cercopithecoidea. Accordingly, the RT-PCRs in *M. mulatta* revealed only an amplicon lacking *Alu* exon 8.

### Hominoidea intron(s) between NARF exons 7 and 9

The Hominoidea analyzed in this study do not contain any additional transposed elements or other structural elements that might serve as a substrate for A-to-I editing. Once more, editing-mediated exonization with protein-coding potential is unlikely in *Hylobates lar* and *Pongo pygmaeus*. The in-frame TGA stop codon already present in the platyrrhines *C. jacchus* and *S. sciureus* persists in *H. lar*, *P. pygmaeus*, and *Gorilla gorilla* genomic DNA (Fig. 3, positions 114–116). Gorilla, in addition, features a 4-nt insertion, leading to a frameshift (Fig. 3, positions 144–147). The RT-PCRs for *H. lar* and *P. pygmaeus* revealed only the splice variant without the alternative exon 8.

Interestingly, the RT-PCR for *G. gorilla* revealed a distinct amplification signal (Supplementary Information, Fig. SI1). Cloning and sequencing of 200 clones revealed 144 with exon 8 sequences. From a total of nine different splice variants, we detected three (51 clones) with an ORF that includes the additional gorilla-specific 4-nt insertion; in two of these variants (2 clones each), the in-frame stop codon was edited (TAA → TGG; Fig. 3, positions 41–43), leading to a potentially functional insertion of exon 8 with an intact ORF. The gorilla alternative exon 8 ORF is shifted in-phase compared to exon 8 in human,[6] but the 4-nt insertion restores the original exon 9 ORF.

The RT-PCRs for *Pan troglodytes* and *Homo sapiens* revealed strong amplification signals, indicating the presence of the alternative exon 8. We sequenced 200 cDNAs for each and verified 130 and 152 clones, respectively. At the DNA level, the original in-frame TGA stop codon (Fig. 3, positions 114–116) is changed to CGA, and at the RNA level, the hominoid-specific UAG stop codon at positions 28–30 (Fig. 3) is efficiently edited to UGG (in 96% and 91% of clones, respectively). Thus, function with respect to protein-coding capacity of the inserted *Alu*Sx element was possible, after A-to-I RNA editing of the predominant 3′ splice site at positions 9–10 (Fig. 3; see also Lev-Maor *et al.*[6]).

At the genomic level, all investigated human groups (Japanese, Chinese, Koreans, African Americans, and Caucasians) exhibit identical DNA sequences.

### 3′ splice sites for exon 8

In total, we reverse transcribed and sequenced 426 verified mRNAs from gorilla, chimpanzee, and human. Ten different 3′ splice sites in the region of the sense-oriented *Alu*Sx element were detected (Fig. 3; Supplementary Information). Seven of these originate from A-to-I editing of AA to AG splice sites (gorilla, 70 clones; chimpanzee, 48 clones; human, 66 clones); the remaining three are already present at the genomic level. Correspondingly, 4 gorilla, 39 chimpanzee, and 49 human clones contain an intact ORF. This suggests that A-to-I editing is similarly efficient in the three great apes, but in terms of providing potential alternative protein-coding capacity, there is a large difference in gorilla, compared to chimpanzee and human. This can be explained by the fact that the prevailing splice variant starting at position 11 (Fig. 3; Supplementary Information) in chimpanzee (40 sequences) and human (42 sequences) is interrupted in gorilla (33 sequences) by a stop codon at positions 114–116 (Fig. 3; Supplementary Information) and the out-of-frame 4-nt insertion (see above). Interestingly, the highly edited stop codon UAG at positions 28–30 (Fig. 3) facilitates an ORF in chimpanzee and human but is not important for the alternative ORF in gorilla, even though it is frequently edited (24/33 clones) in this species. This suggests that editing may not necessarily be connected to selection for functionality.

### Editing sites for exon 8

Lev-Maor *et al.* described six RNA editing sites (the edited splice site and sites E1–E5),[6] two of which are essential for inclusion of *Alu* exon 8 in the human NARF gene (Fig. 3). The first facilitates the alternative splicing of NARF *Alu* exon 8 by producing a new 3′ splice site (Fig. 3, position 10), and the second is the editing of an in-frame UAG stop codon resulting in a UGG tryptophan codon (Fig. 3, position 29, E1).

Analyzing the sequence data of our RT-PCR clones revealed 24 additional editing sites in the *Alu* exon 8 region (Fig. 3), leading to further alternative splice sites, changes in amino acids, silent changes (preserving the sequence of amino acids), and out-of-frame splice variants. In a favorable dsRNA environment, provided in this case by the duplex structure of the two *Alu*Sx elements, ADAR1 editing of many adenosines in the immediate neighborhood of one another is not unusual.[17,18] The editing frequencies of all 30 A-to-I editing sites found in the 426 analyzed RNAs are calculated and displayed in Fig. 3 and Supplementary Information. Many of the editing sites are modified in only one clone or in a few clones, suggesting that these sites are without important functional consequences for the alternatively spliced NARF gene.

Of the 30 posttranscriptional modifications identified, 6 represent "high-efficiency" *Alu* editing sites[19] (encircled in Fig. 3). Editing of the in-frame stop codon that is essential to facilitate an intact ORF

was detected, for example, in 93% of the clones with the E1 editing site, confirming previous results.[6] By chemical inactivating nonsense-mediated decay (NMD), a mechanism that is known to remove mRNAs with premature internal stop codons, Lev-Maor *et al.* showed that the lack of UAG stop-codon-containing mRNAs was not due to NMD-mediated degradation.[6] Furthermore, considering the high number of in-frame stop codons within the sequenced clones (e.g., 131 for human), NMD seems to be not greatly relevant for the alternative splice products of NARF. Curiously, the editing frequency decreases towards the 3′ end of the exon. This is possibly due to declining complementarity of the inversely oriented *Alu*Sx sequences, which include one insertion and one deletion (indels; Supplementary Information, Fig. SI2), reducing the efficiency of the ADAR1 enzyme.

Interestingly, the identical great ape sequences around the editing sites (Fig. 3) as well as the base complementarities between the two inversely oriented *Alu*Sx elements (Supplementary Information, Fig. SI3) fail to explain why gorilla, chimpanzee, and human RNAs are edited, while apparently the corresponding sequences in orangutan and other anthropoids remain unspliced and unedited. Despite using species- and anthropoid-specific primers for RT-PCR and amplification of orangutan, gibbon, and marmoset from skeletal muscle, as well as from some other tissues that more highly express the alternative exon 8 in human (orangutan: liver, marmoset: brain; data not shown), we failed to generate any specific PCR products supporting the inclusion of exon 8 in these species. Elution, cloning, and sequencing of material from the expected size range of the electrophoretic gels also revealed no traces of exon 8.

One important regulatory mechanism at the exon inclusion level is the altering of exonic splicing enhancers and silencers. Lev-Maor *et al.* defined 12 different important splicing regulatory regions for the human NARF alternative exon 8.[6] The sequences of all 12 regions were compared in all investigated species and found to be identical in great apes, and therefore, they cannot account for the differences in alternative splicing observed in these species. Moreover, their relevance to splicing does not seem to be

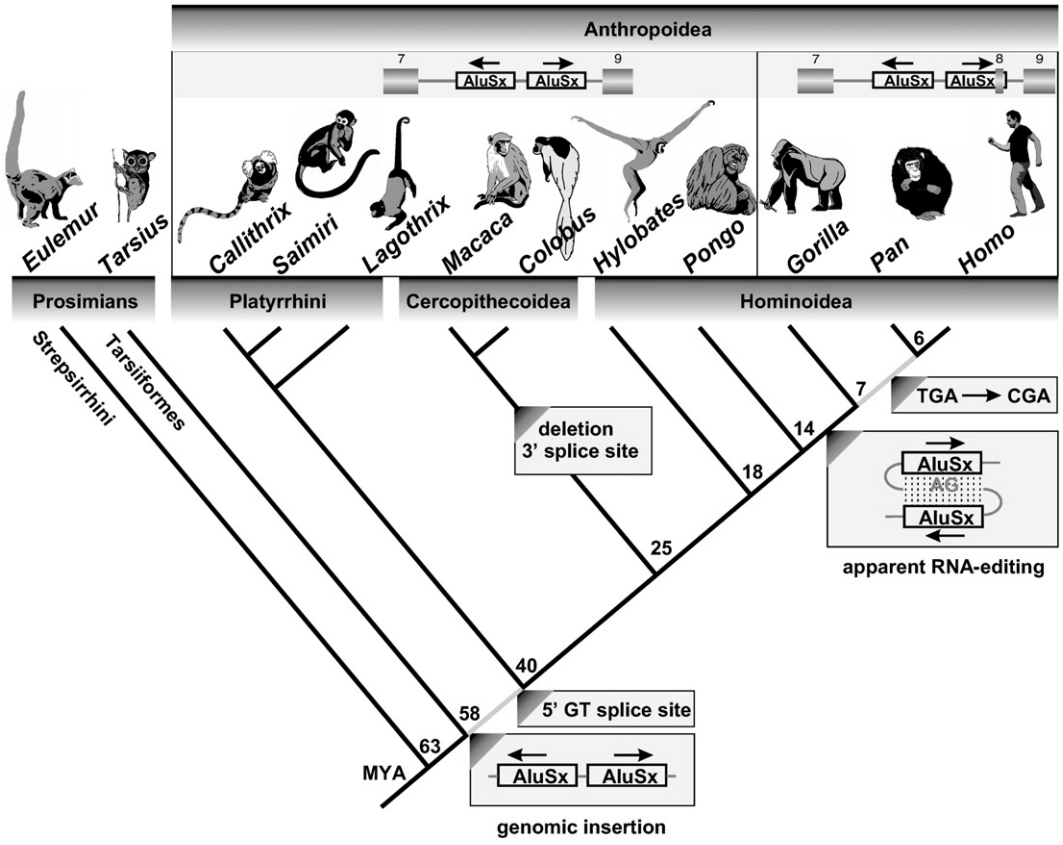


**Fig. 4.** Evolutionary scenario for the birth of NARF *Alu* exon 8. Phylogenetic tree representing all primate infraorders with approximate dating for each divergence point.[20] About 58–40 MYA in the common ancestor of higher primates (Anthropoidea), two genomic changes set the stage for the evolution of NARF *Alu* exon 8: the insertion of head-to-head-oriented *Alu*Sx elements, including the eventual AA editing site, and the generation of a potential 5′ GT splice site. About 50–35 million years later, in a common ancestor of human, chimpanzee, and gorilla, three further important changes appeared: ADAR1-mediated RNA editing within the intronic dsRNA structure of the two *Alu*Sx sequences facilitated the editing of an AA to a 3′ AG splice site and the editing of a premature stop codon (UAG→UGG), and genomic modification of a TGA to CGA converted a premature stop into an arginine codon. A schematic structure of the genomic situation for editing critical components is shown in the upper part. The newly exapted NARF *Alu* exon 8 is shown for gorilla, chimpanzee, and human.

essentially dependent on editing of the suggested regulatory sequences.[6] Furthermore, sequencing also showed that the 50-nt region upstream of the 3′ splice site and the 26-nt region downstream of the 5′ splice site, also known to be important for splicing, are identical in both orangutan and human. It remains to be elucidated why, despite the presence of all known necessary requisites, inclusion of exon 8 is apparently not occurring in orangutan.

## Conclusions

Four major events distributed over more than 50 million years of primate evolution gave rise to the birth of the novel exon 8 in the NARF gene (Fig. 4). This chain of events started with the insertion of two head-to-head-oriented *Alu*Sx elements, already harboring one of the potential essential A-to-I editing sites, in the intron between exons 7 and 9 in the common ancestor of anthropoid primates. During the same period, a genomic C-to-GT change near the 3′ end of the sense-oriented *Alu*Sx element occurred, facilitating a functional 5′ splice site some 45 million years later. In the next 34 million years of primate evolution, no significant changes occurred in the respective intron, and the two *Alu*Sx elements, essential for duplex mRNA formation, did not diverge enough from each other to preclude the essential duplex RNA formation (19% sequence divergence in humans for the human exonized region; Supplementary Information, Fig. SI2).

In the pre-mRNA, the double-strand structure of the oppositely oriented *Alu*Sx elements facilitated the ADAR machinery in generating, at some point (perhaps even concomitantly with the second insertion), novel 3′ splice sites, supported by splice products in human, chimpanzee, and gorilla, albeit at very low potentially protein-inclusion levels in the latter great ape. In the common ancestor of human and chimpanzee, additional critical changes lead to more efficient inclusion of the potentially functional novel NARF *Alu* exon 8. At the genomic level, a TGA stop codon within the sense-oriented *Alu*Sx element mutated to CGA (arginine). At the RNA level, an additional premature stop codon UAG is being edited to a UGG (tryptophane) to yield the new protein-coding exon. With these changes, the novel, potentially functional, efficiently alternatively spliced NARF *Alu* exon 8 was born, persisting over the last 6 million years of chimpanzee and human evolution and developing numerous additional splice variants and editing sites. The efficiency of the acquisition of *Alu* exon 8 in the common ancestor of human and chimpanzee underlines the close relationship of the two species, indisputably demonstrated for the first time by Salem *et al.*[21] However, the gorilla genome also evolved a collection of alternative splice variants, most of them also mediated by A-to-I editing of exon 8; nevertheless, potential inclusion of these forms in protein-coding domains maintaining an intact ORF is limited to very few variants. The exaptation of NARF *Alu* exon 8 displays the complex interplay of genomic and post-transcriptional changes accumulated over a period of more than 50 million years, eventually leading to novel protein-coding material. Selection pressure is evident in the nearly 100% effective editing of the *Alu* exon 8 internal genomic stop codon in chimpanzee and human. On the other hand, a high level of variable alternative splice and editing sites continues to maintain a state of posttranscriptional plasticity—compared to 100% sequence conservation at the genomic level—to chance upon potential novel functions for these alternative splice variants in the hominin lineage.

## Materials and Methods

### DNA extraction

Following standard protocols,[22] genomic DNA was isolated from tissue samples of the Hominoidea: *G. gorilla*, *P. pygmaeus*, and *H. lar*; the Cercopithecoidea (Old World monkeys): *C. guereza* and *T. auratus*; the Platyrrhini (New World monkeys): *C. jacchus*, *L. lagotricha*, and *S. sciureus*; the Tarsiiformes: *T. syrichta*; and the Strepsirrhini: *E. coronatus* and *M. murinus*. Genomic DNA from Hominoidea blood of *H. sapiens* (four different ethnic groups: Japanese, Chinese, Koreans, and African Americans) and *Pan paniscus* was isolated using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

### PCR amplification of genomic DNA

Information about the NARF gene was available from the UCSC Genome Browser† for *H. sapiens* (Caucasian), *P. troglodytes*, and *M. mulatta*. Based on the information in these three primates, the conserved primers NARF-F, NARF-F1, and NARF-R were designed for genomic PCR amplification (Supplementary Information). Primers NARF-F1 and NARF-R were used for PCR amplification of *H. sapiens* (Japanese, Chinese, Koreans, and African Americans), *G. gorilla*, and *C. guereza*. Amplification in all other species was performed using primers NARF-F and NARF-R. All PCR reactions were performed using the Thermoprime Plus DNA Polymerase (ABgene, Hamburg, Germany) under standard conditions.

Purification from 1% agarose gels was followed by ligation of the PCR products into the pDrive Cloning Vector (Qiagen) and electroporation into TOP10 cells (Invitrogen, Groningen, The Netherlands). Plasmids containing the PCR products were isolated using the High Pure Plasmid Isolation Kit (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. Sequencing was performed with the Ampli Taq FS Big Dye Terminator Kit (PE Biosystems, Foster City, CA) using standard M13FSP/RSP primers (Supplementary Information).

### RNA extraction

Total RNA was extracted from frozen muscle of *H. sapiens*, *P. troglodytes*, *G. gorilla*, *P. pygmaeus*, *H. lar*, *M.*

† http://genome.ucsc.edu/cgi-bin/hgBlat

*mulatta*, and *C. jacchus* using the TRIzol method (Invitrogen). In addition, brain material was used for *M. mulatta* and *C. jacchus* and liver for *P. pygmaeus*. In human, it has been shown that although the levels of tissue-specific expression are different, alternatively spliced NARF *Alu* exon 8 is readily detectable in various tissues,[6] including muscle.

### RT-PCR

RT-PCR was carried out on all extracted RNAs (see above). Reverse primers (NARF-R* for gorilla, chimpanzee, and human; Supplementary Information) were used with 200 U SuperScript™II RNaseH-Reverse Transcriptase (Invitrogen) to synthesize cDNA from approximately 4 μg RNA (60 min at 45 °C). After RNaseH treatment, the cDNA was used for PCR amplification with Phusion High-Fidelity DNA Polymerase (proofreading activity) (Finnzymes, Espoo, Finland). For gorilla, we used primers NARF-HsPtGg-F and NARF-R3; for chimpanzee and human, we used primers NARF-HsPtGg-F and NARF-HsPt-R2. Further, used primers and their genomic locations are listed in Supplementary Information. After gel elution, A overhangs for TA cloning were produced with Thermoprime Plus DNA Polymerase (ABgene) at 70 °C for 20 min. PCR products were purified with the Plasmid Isolation Kit (Roche), cloned in the pDrive Cloning Vector (Qiagen), and sequenced under the same conditions described for genomic DNA. For full overlap sequencing, M13 forward (FSP) and reverse (RSP) primers were used (Supplementary Information) to sequence 200 clones for each species. In spite of the lack of any distinct signal after electrophoretic separation of gibbon and marmoset and an unspecific signal in orangutan PCR products, we cut out the areas of the gels containing the expected size range and derived 50, 50, and 200 sequences, respectively.

### Sequence analyses

Sequences were manually aligned to available sequences of *H. sapiens*, *P. troglodytes*, and *M. mulatta* obtained from the UCSC Genome Browser†. The RepeatMasker Server‡ was used to detect and classify inserted transposed elements. FASTA alignments for DNA are available upon request from the corresponding author.

### GenBank accession numbers

EF601655, *H. sapiens* RNA 1; EF601656, *H. sapiens* RNA 2; EF601662, *P. paniscus* DNA; EU232758, *P. troglodytes* RNA1; EU232759, *P. troglodytes* RNA2; EF601663, *G. gorilla* DNA; EF601657, *G. gorilla* RNA; EF601664, *P. pygmaeus* DNA; EF601658, *P. pygmaeus* RNA; EF601665, *H. lar* DNA; EF601659, *H. lar* RNA; EF601666, *C. guereza* DNA; EF601667, *T. auratus* DNA; EF601660, *M. mulatta* RNA; EF601668, *C. jacchus* DNA; EF601661, *C. jacchus* RNA; EF601669, *L. lagotricha* DNA; EF601670, *S. sciureus* DNA; EF601671, *T. syrichta* DNA; EF601672, *E. coronatus* DNA; EF601673, *M. murinus* DNA. Aligned cDNA sequences are presented in FASTA format as Supplementary Information.

‡Smit and Green, RepeatMasker at http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker

## Acknowledgements

We thank Michael Haberl, Bernd Karger, Andreas Ochs, Sandra Silinski, Svante Pääbo, and Peter Wohlsein for providing us with tissue samples and are grateful to Marsha Bundman for editorial assistance. We thank Silke and Andy Scheffel for providing the primate silhouettes of Fig. 4. We appreciate the valuable comments of the two anonymous referees. This work was supported by the Nationales Genomforschungsnetz (0313358A to J.B. and J.S.) and the Deutsche Forschungsgemeinschaft (SCHM1469 to J.S. and J.B.).

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.07.014

## References

1. Sorek, R., Ast, G. & Graur, D. (2002). Alu-containing exons are alternatively spliced. *Genome Res.* **12**, 1060–1067.
2. Krull, M., Brosius, J. & Schmitz, J. (2005). Alu-SINE exonization: en route to protein-coding function. *Mol. Biol. Evol.* **22**, 1702–1711.
3. Nekrutenko, A. & Li, W. H. (2001). Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **17**, 619–621.
4. Krull, M., Petrusma, M., Makalowski, W., Brosius, J. & Schmitz, J. (2007). Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). *Genome Res.* **17**, 1139–1145.
5. Sela, N., Mersch, B., Gal-Mark, N., Lev-Maor, G., Hotz-Wagenblatt, A. & Ast, G. (2007). Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol.* **8**, R127.
6. Lev-Maor, G., Sorek, R., Levanon, E. Y., Paz, N., Eisenberg, E. & Ast, G. (2007). RNA-editing-mediated exon evolution. *Genome Biol.* **8**, R29.
7. Kriegs, J. O., Churakov, G., Jurka, J., Brosius, J. & Schmitz, J. (2007). Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet.* **23**, 158–161.
8. Kapitonov, V. & Jurka, J. (1996). The age of Alu subfamilies. *J. Mol. Evol.* **42**, 59–65.
9. Basilio, C., Wahba, A. J., Lengyel, P., Speyer, J. F. & Ochoa, S. (1962). Synthetic polynucleotides and the amino acid code. *Proc. Natl Acad. Sci. USA*, **48**, 613–616.
10. Bass, B. L. (2001). *RNA-Editing: Frontiers in Molecular Biology.* Oxford University Press, Oxford.
11. Gott, J. M. & Emeson, R. B. (2000). Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* **34**, 499–531.
12. Bass, B. L., Nishikura, K., Keller, W., Seeburg, P. H., Emeson, R. B., O'Connell, M. A. *et al.* (1997). A standardized nomenclature for adenosine deaminases that act on RNA. *RNA*, **3**, 947–949.
13. DeCerbo, J. & Carmichael, G. G. (2005). SINEs point to abundant editing in the human genome. *Genome Biol.* **6**, 216.

14. Fierro-Monti, I. & Mathews, M. B. (2000). Proteins binding to duplexed RNA: one motif, multiple functions. *Trends Biochem. Sci.* **25**, 241–246.
15. Kim, U., Wang, Y., Sanford, T., Zeng, Y. & Nishikura, K. (1994). Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proc. Natl Acad. Sci. USA*, **91**, 11457–11461.
16. Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P. H. & Hguchi, M. (1996). A mammalian RNA editing enzyme. *Nature*, **379**, 460–464.
17. Bass, B. L. (2002). RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846.
18. Källman, A. M., Sahlin, M. & Ohman, M. (2003). ADAR2 A→I editing: site selectivity and editing efficiency are separate events. *Nucleic Acids Res.* **31**, 4874–4881.
19. Athanasiadis, A., Rich, A. & Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**, e391.
20. Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J. *et al.* (1998). Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* **9**, 585–598.
21. Salem, A. H., Ray, D. A., Xing, J., Callinan, P. A., Myers, J. S., Hedges, D. J. *et al.* (2003). Alu elements and hominid phylogenetics. *Proc. Natl Acad. Sci. USA*, **100**, 12787–12791.
22. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.