Biochimie 93 (2011) 1928–1934

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Review

Exonization of transposed elements: A challenge and opportunity for evolution

Jürgen Schmitz*, Jürgen Brosius

Institute of Experimental Pathology (ZMBE), University of Münster, Münster, Germany

ARTICLE INFO

Article history:
Received 29 March 2011
Accepted 11 July 2011
Available online 21 July 2011

Keywords:
Retroposon exonization
Exonization phylogeny
Splicing
A-to-I editing

ABSTRACT

Protein-coding genes are composed of exons and introns flanked by untranslated regions. Before the mRNA of a gene can be translated into protein, the splicing machinery removes all the intronic regions and joins the protein-coding exons together. Exonization is a process, whereby genes acquire new exons from non-protein-coding, primarily intronic, DNA sequences. Genomic insertions or point mutations within DNA sequences often generate alternative splice sites, causing the splicing system to include new sequences as exons or to elongate existing exons. Because the alternative splice sites are not as efficient as the originals the new variants usually constitute a minor fraction of mature mRNAs. While the prevailing original splice variant maintains functionality, the additional sequence, free from selection pressure, evolves a new function or eventually vanishes. If the new splice variant is advantageous, selection might operate to optimize the new splice sites and consequently increase the proportion of the alternative splice variant. In some instances, the original splice variant is completely replaced by constitutive splicing of the new form. Because of the fortuitous presence of internal splice site-like structures within their sequences, portions of transposed elements frequently serve as modules of exonization. Their recruitment requires a long and versatile optimization process involving multiple changes over a time span of millions, even hundreds of millions, of years. Comparisons of corresponding genes and mRNAs in phylogenetically related species enables one to chronologically reconstruct such changes, from ancient ancestors to living species, in a stepwise manner. We will review this process using three different exemplary cases: (1) the evolution of a constitutively spliced mammalian-wide repeat (MIR), (2) the evolution of an alternative exon 1 from an alternative 5′-extended primary transcript containing an *Alu* element, and (3) a rare case of the stepwise exoniztion of an *Alu* element-derived sequence mediated by A-to-I RNA editing.

1. Introduction

An important process by which species evolve new properties is gene duplication and subsequent variation of the duplicate [1]. Duplicates are generally not under selection pressure and are, thereby, free to be reshaped and to gain variant or new features that might be advantageous for the individual, population, species, and lineage. This “trial-and-error” process has a chance to be successful as long as the original, functional form of the gene is maintained. A second way to fortuitously evolve novelties and versatility from a restricted set of genes is alternative splicing, a process by which a gene’s string of exons are reconnected in multiple ways at the RNA stage, enabling a single gene to encode multiple protein variants [2–4]. New alternative splice variants have the chance to evolve and be “tested” for variant functions while the original splice variants ensure functionality. New protein-coding exons are frequently acquired when new splice sites are formed in intronic or other untranslated regions. Remarkably, less than one year after the intron/exon structure of eukaryotic genes was revealed, Walter Gilbert predicted, in a short note entitled “Why genes in pieces?” [5], that the intron-exon structure of genes favors the recruitment of novel exons out of intronic space; a process that was soon experimentally verified [6]. This process was later termed exonization [7], and is an excellent example of exaptation at the genome level [8].

The present review focuses on the generation of new splice sites and recruitment of protein-coding sequences from randomly inserted retroposed elements, a special class of transposable/transposed units that arise via reverse transcription of RNA intermediates and subsequent insertion, for example, into the introns or UTRs of genes [9]. Today, the discernible parts of such retroposons account for more than 40% of the human genome [10]. Autonomous

* Corresponding author. Institute of Experimental Pathology (ZMBE), University of Münster, Von-Esmarch-Str. 56, 48149 Münster, Germany. Tel.: +49 251 8352133; fax: +49 251 8352134.
E-mail address: jueschm@uni-muenster.de (J. Schmitz).

0300-9084/$ – see front matter © 2011 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biochi.2011.07.014

retroposons, such as Long INterspersed Elements (LINEs) and Long Terminal Repeat (LTR) retrotransposons, encode their own enzymatic machinery to reverse transcribe RNA and insert the copy DNA (cDNA) back into the genome [11]. Fortuitously, in addition to retroposing their own RNA, this machinery can also retropose any (predominantly polyadenylated) RNA [12,13], thereby replicating themselves as well as other transcribed genes.

Transcriptionally active, non-autonomous Short INterspersed Elements (SINEs) do not have their own enzymatic machinery, but are reverse transcribed and inserted by the retropositional machinery of LINE elements [14]. Interestingly, although certain SINEs, such as the primate specific *Alu* elements, are trans-retroposed and inserted with the LINE machinery, contrary to the LINEs, they tend to accumulate in gene rich regions of the genome [15]. Because SINEs have no endogenous protein-coding capacity, usually only short sequence stretches, free of stop codons or other sequences that might interrupt the host gene open reading frame (ORF), contribute to new protein-coding sequences [16].

Biocomputational analyses of the human genome predicted that many primate-specific *Alu* SINEs might be part of protein-coding ORFs in mRNAs [9,17–19]. Gil Ast's group then demonstrated this in experiments whose results also provided mechanistic insights into the process of exonization [20–24]. Our laboratory conducted phylogenetic analyses of *Alu* elements in primates to study the various steps and time frames of novel exon exaptation; most of the observed *Alu* "exonizations" did not remain fixed in subsequent evolutionary relatives [25–27]. However, we did observe older persisting exonizations of Mammalian-wide Interspersed Repeats (MIRs) in mammals [28].

2. Why SINEs are especially predestinated to exonization

The diverse impacts of exonization on different species can be easily appreciated by comparing the only 12 detected transposable element-derived exons in the nematode *Caenorhabditis elegans* with the ~1800 retroposon-derived exons in humans [23]. This bias is probably influenced by a combination of low retropositional activity and selection toward short introns in invertebrates [29]. In contrast to those of invertebrate species, vertebrate and especially mammalian genomes "tolerate" large introns, which mitigate the potential negative impacts of additional transposed elements. Furthermore, the enhanced activity of the mammalian-specific LINE 1 elements and their non-specific selection of nearly any RNA for retroposition, increases the possibility of retroposon exonization in therian mammals [30]. However, the potential for exonization is not only a statistical matter of intron size and element numbers. Certain SINE elements, such as *Alu*s and MIRs, are predestined for exonization because they contribute (in antisense orientation) preexisting cryptic 3′ AG splice sites (conserved consensus sequences at the end of introns) and 5′ oligo-pyrimidine tracts (reversing the typical sense-oriented oligo(A) region(s) of SINE elements) [21,28,31], both of which are important [28] recognition signals for the splicing apparatus. Consequently, significantly more exonized *Alu*s and MIRs are located in the antisense orientation corresponding to the transcription of the host gene [28,29]. A 5′ splice site is then randomly activated downstream of the retroposon-specific 3′ splice site, either inside or outside of the retroposed element. In the latter case, part of the intron is included in the protein-coding sequence (e.g., Fig. 1, black cylinder).

3. Alternative vs. constitutive splicing of exonized sequences

In most instances, novel duplicated genes or splice variants are initially free from selection pressure as long as the original gene or mRNA remains functional and the duplicate or variant is not disadvantageous [32]. However, favorable or advantageous alternative splice variants can also replace the original splice variant. Although they initially feature relatively weak, cryptic, splice sites and low representation, eventually they may acquire stronger, canonical, splice sites and, even exclusive representation, without the continued need of the original splice product, and are thus constitutively spliced.

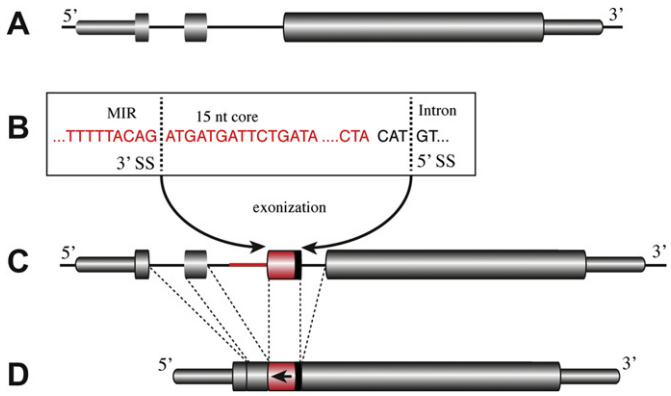

Fig. 1. Exonization of a portion of a more than 160-million-year-old MIR element (red) into the protein-coding region of the zinc finger protein *ZNF639* and its constitutive expression in all living mammals. Thick gray cylinders represent protein-coding exons, medium thick gray cylinders indicate the 5′ and 3′ untranslated regions (UTRs). Introns and intergenic regions are shown as black lines. The MIR-derived exon is shown as a thick red cylinder, the remainder of the intronic MIR element as a red line. The exonized part of the original intronic sequence (CAT) is shown as a narrow black cylinder. (A) Structure of the *ZNF639* gene prior to insertion of a MIR element into intron 2. (B) Nucleotide sequence of parts of the inserted MIR element for human (in red) and the exonized MIR and intronic portions, including the 3′- and 5′ splice sites (SS; dotted vertical lines). The 3′ splice site is adjacent to the 15-nt highly conserved core region of MIR elements and the 5′ splice site is part of the intron. (C) Gene structure including the exonized MIR cassette. (D) The mature, constitutively expressed *ZNF639* mRNA. Splicing is shown by dotted lines and the arrow indicates that the MIR element is in the antisense orientation.

We demonstrated the mammalian-specific constitutive expression of a new variant with an additional exon for the zinc finger protein 639 gene *ZNF639* (Fig. 1). More than 160 million years ago, a MIR element inserted in an antisense orientation into intron 5 of the ancestral *ZNF639* gene. Today, the exonized MIR sequence (132 MIR nts plus 3 nts of the adjacent intronic sequence) is constitutively included in the mRNA of all mammalian species in all tissues examined to date [28]. This implies that exonization might have occurred shortly after insertion or at least prior to the divergence of Mammalia ~160 million years ago. And, because all mammals constitutively express the version of this gene containing the MIR cassette, in all likelihood, all steps toward constitutive splicing happened before monotremes diverged from therians. Lin et al. [33] subsequently described fifteen potential constitutively expressed MIR cassettes, but only six cases of constitutive expression of *Alu* cassettes were found. However, a more extensive comparative analysis including non-mammalian species would be necessary to confirm these cases. It should be cautioned that there might be alternative splice variants that are active only at specific developmental stages or in special cell types that were not investigated. In Krull et al. [28], we propose that generally, long evolutionary times are necessary to establish constitutively expressed retroposon-containing cassettes. If all components for exonization and constitutive expression occur (and are fixed) before a speciation event, all following members of the lineage will express this exonization. If the time between speciation events is too short, some subsequent lineages will establish the remaining necessary steps later and others may loose the exonization [25]. By way of comparison, MIR

elements were present for up to 140 million years of mammalian evolution before an internal mammalian split occurred, while the *Alu* elements in primates, where many splits occurred over short time intervals, had only about half the available time (87 million years) to become fixed in genomes, and that may be why we do not see as many fixed *Alu* elements as we do MIRs [28].

4. From birth to maturity of retroposon-derived exons

Fig. 2 shows the estimated times of various SINE insertions and their subsequent exonizations in representative mammalian species. For both MIR and *Alu* elements, exonization occurred either shortly after insertion in the same ancestral branch, as shown for insertions number 5 (MIR in *ZNF639*; [28]), 6 (*Alu*Jb in *LEPR*; [34]), 11 (*Alu*Sp in*MTO1-3*; [25]), and 12 (*Alu*Sc in*PKP2b-4*; [25]) or as much as 150 million years later, as exemplified by the exonization of a MIR3 element (insertion 1) in the cholinergic receptor nicotinic alpha 1 gene (*CHRNA1*; [28]) in Great Apes. Frequently, a series of changes was required to generate all the necessary genomic conditions for successful splicing. Fig. 3 illustrates the reconstruction of all steps that occurred for the emergence of the novel *Alu*-derived exon 1 in the human tumor necrosis factor receptor gene type 2 (*p75TNFR*; [27]), including the insertion of the *Alu*Jo element into the 5′ UTR of the gene and the untimed acquisition of an alternative transcription start site upstream of the element. In addition, a point mutation generated a new ATG start codon derived from an ATA sequence within the *Alu*Jo element. This “future” alternative start codon is present in marmosets and all other anthropoids. In the lineage leading to Old World monkeys, a GC–GT mutation generated a functional 5′ splice site. Finally, in the same lineage, a random deletion of 7 nts generated an intact ORF and set the stage for expression of the alternative splice form *icp75TNFR* in all Old World monkeys including Apes. Xing and Lee [35] suggest that phylogenetic reconstruction of the changes leading to exonization offers a unique possibility to recognize and better understand the historical requirements for successful exonization.

5. Editing-associated alternative splicing

The repertoire of possible molecular mechanisms to arrive at the acquisition of new exons from retroposon cassettes can be quite versatile, and another such mechanism is, albeit rare, exemplified by the nuclear prelamin A recognition factor gene (*NARF*; Fig. 4). Lev-Maor et al. [20] aligned ESTs and cDNAs from GenBank (version 136) to sequences of the human genome and screened for exonized retroposons that were not flanked by canonical splice sites (at the genomic level) and recovered an alternative splice variant of the *NARF* gene, in which A-to-I editing in the pre-mRNA mediates alternative splicing [36], producing a new *Alu*Sx-derived protein-coding sequence. In the genome, the *NARF AluSx* exon 8 is flanked by an AA sequence that, at the pre-mRNA level, is edited to a 3′ AI (functions as AG) splice site. Editing is mediated by the 5′ adjacent, antisense-orientated *Alu*Sx element that forms a double-stranded pre-mRNA structure with the sense-oriented partially exonized *Alu*Sx, a precondition for activating the adenosine deaminase-mediated RNA editing. From a phylogenetic perspective, both *Alu*Sx elements inserted sometime between 81 and 43 million years ago after Tarsius split from the common ancestor of higher primates

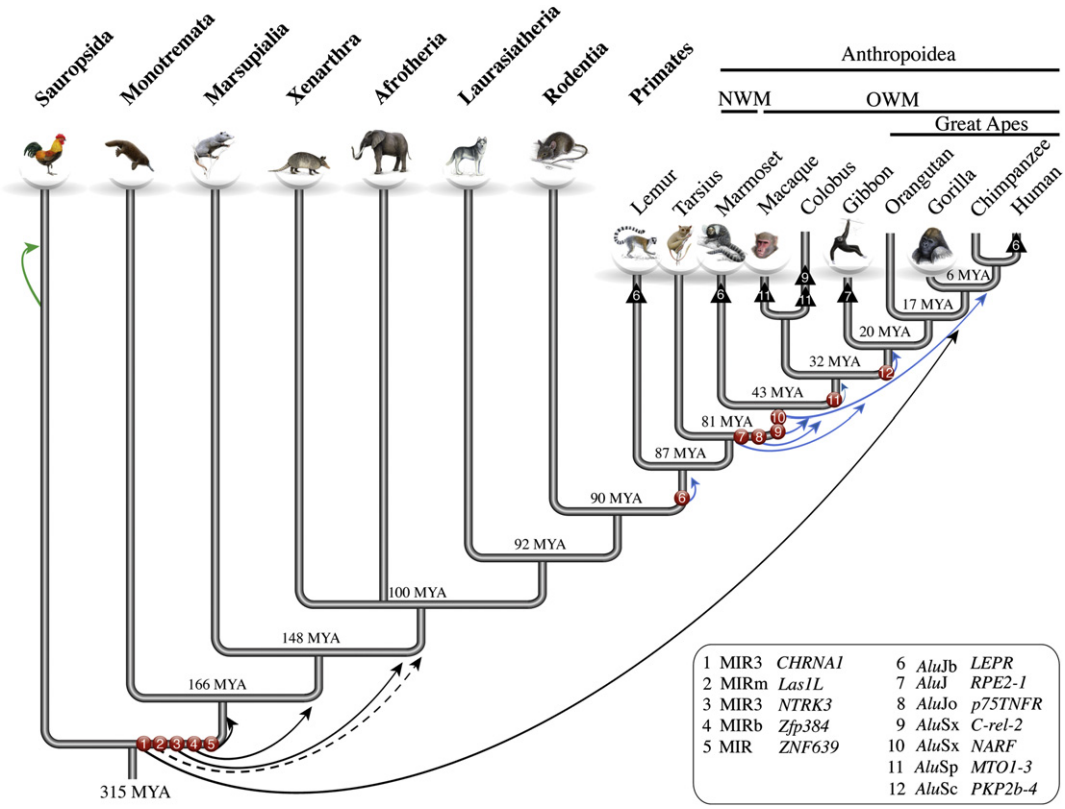

Fig. 2. Mammalian SINE insertions and exonizations. Various MIR and *Alu* retroposons inserted into the respective genes (listed in the box) at about the times shown by the red dots. The times of exonization after insertion are indicated by arrows (black for MIRs and blue for *Alus*). The dashed arrow indicates that the exonization might have happened earlier. Some of the *Alu* exonizations did not persist in all primate lineages (losses depicted by triangles). For the chicken *ZNF639* gene, a constitutively expressed exonized intronic sequence is located in the same intron as the mammalian exonized MIR element (green arrow). NWM = New World monkeys, OWM = Old World monkeys. The *Alu*Jb insertion (#6) was taken from Huh et al. [34]; all remaining insertions are from Krull et al. [25] and Krull et al. [28]. Dating was adopted from Warren et al. [47], Wible et al. [48], Liu et al. [49], and Perelman et al. [50]. For an early study on MIR exonization see Hughes [51].

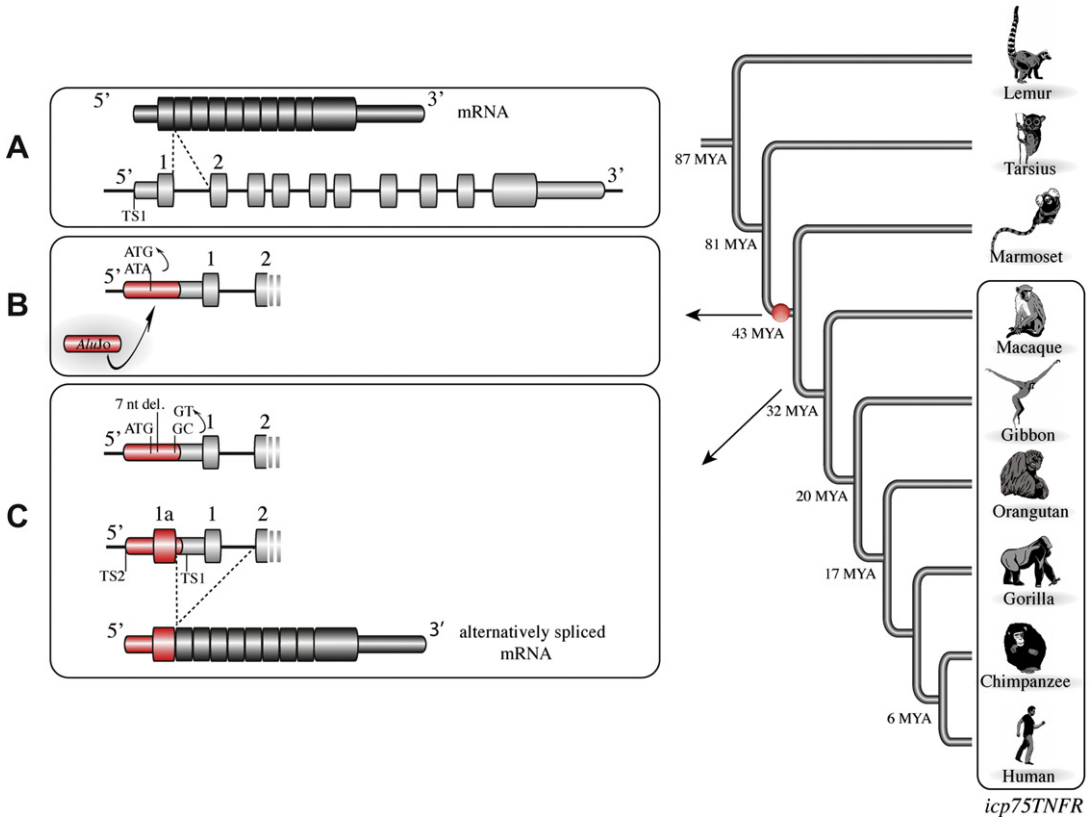

Fig. 3. Step-by-step changes required in the tumor necrosis factor receptor gene type 2 (*p75TNFR*) following the insertion of an *Alu*Jo to generate an alternatively spliced mRNA containing a novel exon 1. (A) *top;* Schematic of mRNA of the original primate *p75TNFR* containing 10 exons; *bottom* DNA schematic of the gene. (B) About 43 million years ago, in the common ancestor of higher primates (including New and Old World monkeys), an *Alu*Jo element (red cylinder/dot) inserted into the 5′ UTR of *p75TNFR*. Within the same ancestral lineage, an ATG start codon was generated from an ATA within the *Alu*Jo sequence. (C) After the divergence of New World monkeys, a 3′ splice site was generated by a GC to GT mutation in the common ancestor of Old World monkeys. In the same internal branch, a seven-nucleotide deletion within the inserted *Alu*Jo generated a new ORF, linking the new exon 1 in frame with the remainder of the original exons. This was the final step in acquiring an alternatively spliced exon 1 that is now conserved in all Old World monkey lineages, including Apes. Wide gray cylinders represent exonic sequences, narrow gray cylinders indicate the 5′ and 3′ untranslated regions (UTRs). Introns and intergenic regions are shown as black lines. Splicing is indicated by dotted lines. The original (A) and alternative (C) transcription start sites are denoted TS1 and TS2, respectively.

(Anthropoidea). During this same time, the 5′ GT splice site was generated after the mutation of a GC dinucleotide into a canonical GT splice site. Nearly 40 million years later in the common ancestor of gorilla, chimpanzee and human, the RNA-editing process produced the canonical 5′ AI dimer that functions like an AG splice site (see above). The corresponding exonized RNA sequence still featured a UAG stop codon, which is converted – also by A-to-I editing – to a UIG triplet that functions like a UGG codon for the amino acid tryptophane. Finally, a second TGA stop codon was converted at the DNA-level into a CGA codon for arginine in the common ancestor of chimpanzee and man yielding an ORF for the alternative exon 8 and subsequent exons [26].

Evolution is a random and undirected process. As much as new exons generated from retroposed genes and elements can enrich the transcriptome and functionality, others can cause serious disadvantages [31,37]; for a recent review see Vorechovsky [38]. However, the majority of newly exonized elements probably do not leave any traces in their genomes, providing no advantageous function, they are eventually lost again. Interestingly, A-to-I editing not only has the potential to support exonization but also to prevent (aberrant) exonization as recently documented by Sakurai et al. [39].

6. Evolutionary time and retroposon exonization

Usually, inheritable acquisition of a retroposon exonization requires long periods of time, starting with the insertion of a retroposon into the germ line. Once inserted, the elements have to be fixed in the population, and species, a period that can exceed several million years [40]. Initially, the additional element rarely provides an advantage for the individual, and therefore its distribution within the population is determined mainly by random drift and is less effected by selection. Consequently, fixation time depends on the effective population size, whereby small populations can fix the changes over shorter times [41]. At least in some evolutionary lineages, most exonizations disappear by accumulating random mutations (e.g., that remove splice sites, introduce stop codons, or disturb the order of the downstream reading frame). This was shown for lineage-specific *Alu* exonizations in primates (triangles in Fig. 2; [25]). Only if the exonizations provide advantageous functions to their hosts or are expressed constitutively in important genes (example in Fig. 1; [28]) are they “protected” by purifying selection and persist. This is to say that evolutionary time is a definitive factor determining the importance and evolvability of new exons. In most instances, the 87 million years of primate evolution were not sufficient for many newly exonized *Alu* exons to persist in primates [25]. The situation is somehow different for very old elements, such as MIRs, that had plenty of evolutionary time within ancestral mammalian branches (e.g., MIRs existed for ~140 million years in the first mammalian ancestor before the mammalian radiation) to evolve and establish novelties. All MIR exonizations described (in phylogenetic species order) have persisted thus far [28,34]. Presumably, there were also numerous transient exonizations of MIR elements in the past.

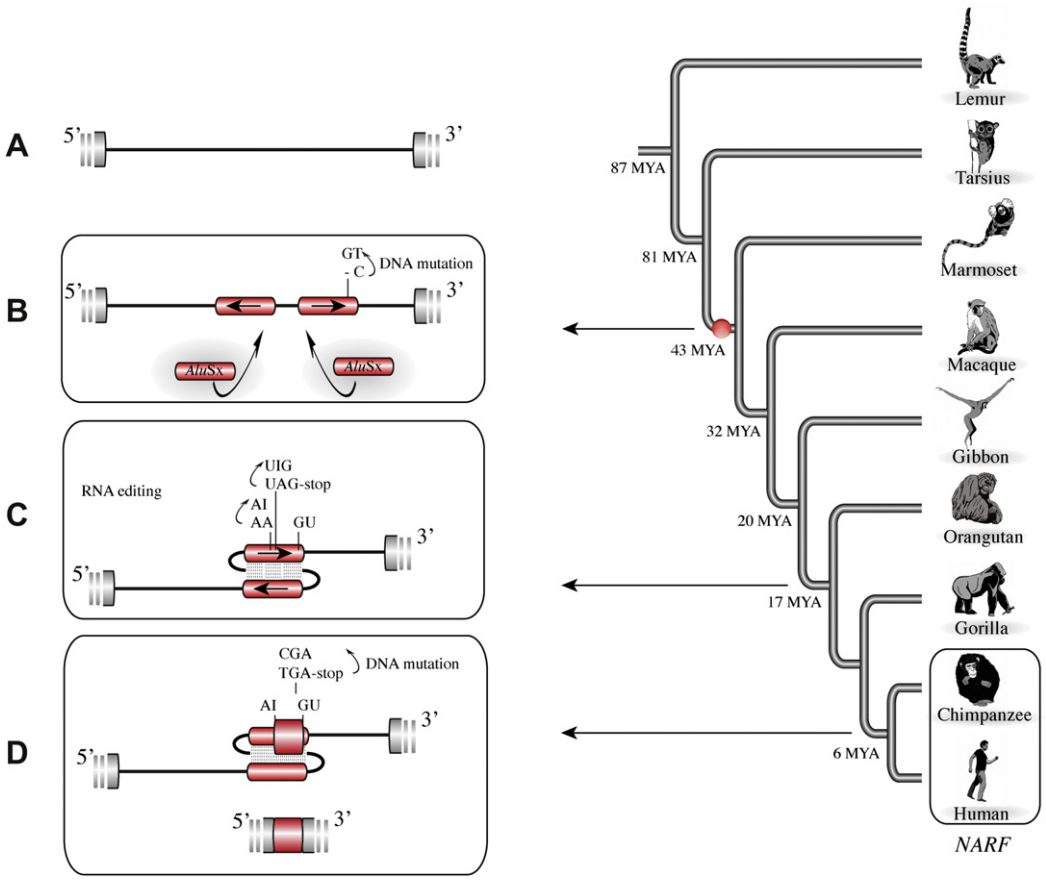

Fig. 4. Schematic showing the successive steps involved in the exonization of an *Alu* element in the nuclear prelamin A recognition factor gene (*NARF*) involving A-to-I RNA editing in primate species. (A) Schematic of DNA of the original primate *NARF* gene showing the region of exon 7, intron 7, and exon 8. (B) In the common ancestor of anthropoids (including New and Old World monkeys) two *Alu*Sx elements integrated head-to-head into intron 7 of *NARF*(red cylinders). In the same ancestor, a 3′ GT splice site appeared in the distal *Alu*Sx element. (C) About 26 million years later, in the common ancestor of gorilla, chimpanzee, and human, a 3′ AI (functions as AG) splice site arose and a UAG stop codon was eliminated, both by A-to-I editing of the pre-mRNA; one of the prerequisites for such editing is duplex RNA formation between the inverted *Alu*Sx elements. (D) Finally, in the common ancestor of chimpanzees and humans, a preexisting TGA stop codon mutated to CGA. With this change, an alternative, potentially functional, splice variant was generated. Wide gray cylinders represent exonic sequences, short gray cylinders indicate the flanking 5′ and 3′ sequences, and introns are shown as black lines. The shaded area indicates the region of double-stranded RNA contributed by the two SINEs and A-to-I editing. The spliced mRNA is shown at the bottom.

It is interesting that the 3' SS of exonized MIRs abuts the so-called core sequence (Fig. 1; 15 nt core [42]), a region significantly more conserved than the remainder of CORE-SINEs [43]. Apart from the possibility that this conservation is due to unknown, perhaps epigenetic mechanisms, one could speculate on a function for this region - or perhaps a function in the not so distant past.

The low number of more recent exonizations observed in rodents, e.g., in comparison to primates, could be due - apart from a reduced suitability of rodent-specific SINEs - to their high rates of sequence evolution and rapid speciation. Although the activity of retroposition in human and mouse is similar, more than 1,800 cases of exonization have been detected in human compared to only about 500 in mouse ([29] and personal observations). However, it must be taken into account that the transcriptome coverage and annotation of the human genome is much more complete compared to mouse.

While the present review focused on only two groups of SINE exonizations, it should be mentioned that other retroposed elements could also be exonized. While exonized sequences from well-known retroposed elements are easily being recognized, anonymous exonized sequences are possibly much more frequent but their detection is more difficult and requires comparative screening for presence or absence in multiple genomes. For example, only by chance we found the exonization of 132 nts and 129 nts of intronic sequences in ostrich and chicken respectively ([28]; Fig. 2), that demonstrate the inclusion of anonymous intronic sequences into the protein-coding part of a gene.

Among others, in mammals the basic splice signals are relative simple and easy to be generated randomly, an efficient way to develop new functional modules [5,44,45].

7. Conclusions

Retroposons are well known to be co-opted into protein-coding functions in the process of exaptation. Especially in mammals, there are large varieties and quantities of different retroposon elements and processes that lead to new exons, and internal cryptic splice sites and branch points can lead to splicing variations.

Usually long evolutionary times are necessary to acquire and establish sufficient genomic conditions for exonization. Therefore elements such as MIRs with a long mammalian history (and long internodes to secure fixation) are more frequently found persistently exonized in all branching lineages, whereas for younger elements such as *Alu*s in primates, it is not unusual for the gain and loss of exonized elements to vary from lineage to lineage. As a caveat, it should be noted, however, that retroposed elements may be "predestinated" candidates for exonization, but they were also studied earlier and more frequently than other protein-coding

cassettes, as they can be easily detected in genomic screenings of ESTs and cDNAs via repeat masking. Other acquired protein-coding cassettes derived from anonymous intronic sequences are not directly recognizable by such screening methods, and can only be found by comparatively analyzing genomes of different species. Their contribution to the evolution of protein-coding sequences will be better understood with the increasing progression of comparative genome sequencing projects.

Except for the constitutively expressed exonizations, the functions of alternative splice variants containing exonized moieties as protein cassettes remains hypothetical and only their unchanged persistence over long evolutionary periods implies possible function [46]. Future intensive protein sequencing and structural as well as functional studies will provide more significant insights into the contributions of such alternative variants.

Acknowledgments

We are grateful to Marsha Bundman for editorial assistance, and thank Jón Baldur Hlíðberg for providing the mammalian paintings of Fig. 2 and Silke and Andy Scheffel for preparing the primate silhouettes of Figs. 3 and 4. This work was supported by the Deutsche Forschungsgemeinschaft (SCHM1469) and by the Nationales Genomforschungsnetz (0313358A).

References

[1] S. Ohno, Evolution by Gene Duplication. Springer-Verlag, New York, 1970.
[2] G. Ast, How did alternative splicing evolve? Nat. Rev. Genet. 5 (2004) 773–782.
[3] B. Modrek, C.J. Lee, Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss, Nat. Genet. 34 (2003) 177–180.
[4] R. Sorek, R. Shamir, G. Ast, How prevalent is functional alternative splicing in the human genome? Trends Genet. TIG 20 (2004) 68–71.
[5] W. Gilbert, Why genes in pieces? Nature 271 (1978) 501.
[6] R. Treisman, N.J. Proudfoot, M. Shander, T. Maniatis, A single-base change at a splice site in a beta 0-thalassemic gene causes abnormal RNA Splicing, Cell 29 (1982) 903–911.
[7] J. Parma, D. Christophe, V. Pohl, G. Vassart, Structural organization of the 5′ region of the thyroglobulin gene. Evidence for intron loss and "exonization" during evolution, J. Mol. Biol. 196 (1987) 769–779.
[8] J. Brosius, S.J. Gould, On “genomenclature”: acomprehensive (and respectful) taxonomy for pseudogenes and other “junk DNA”, Proc. Natl. Acad. Sci. USA 89 (1992) 10706–10710.
[9] A. Nekrutenko, W.H. Li, Transposable elements are found in a large number of human protein-coding genes, Trends Genet. 17 (2001) 619–621.
[10] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. Fitzhugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. Levine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R.A. Gibbs, D.M. Muzny, S.E. Scherer, J.B. Bouck, E.J. Sodergren, K.C. Worley, C.M. Rives, J.H. Gorrell, M.L. Metzker, S.L. Naylor, R.S. Kucherlapati, D.L. Nelson, G.M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D.R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H.M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R.W. Davis, N.A. Federspiel, A.P. Abola, M.J. Proctor, R.M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D.R. Cox, M.V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G.A. Evans, M. Athanasiou, R. Schultz, B.A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W.R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J.A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D.G. Brown, C.B. Burge, L. Cerutti, H.C. Chen, D. Church, M. Clamp, R.R. Copley, T. Doerks, S.R. Eddy, E.E. Eichler, T.S. Furey, J. Galagan, J.G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L.S. Johnson, T.A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W.J. Kent, P. Kitts, E.V. Koonin, I. Korf, D. Kulp, D. Lancet, T.M. Lowe, A. Mclysaght, T. Mikkelsen, J.V. Moran, N. Mulder, V.J. Pollara, C.P. Ponting, G. Schuler, J. Schultz, G. Slater, A.F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y.I. Wolf, K.H. Wolfe, S.P. Yang, R.F. Yeh, F. Collins, M.S. Guyer, J. Peterson, A. Felsenfeld, K.A. Wetterstrand, A. Patrinos, M.J. Morgan, P. De Jong, J.J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y.J. Chen, Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.
[11] J.L. Goodier, E.M. Ostertag, K. Du, H.H. Kazazian Jr., A novel active L1 retrotransposon subfamily in the mouse, Genome Res. 11 (2001) 1677–1685.
[12] J. Brosius, RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements, Gene 238 (1999) 115–134.
[13] K. Ohshima, N. Okada, SINEs And LINEs: symbionts of eukaryotic genomes with a common tail, Cytogenet. Genome Res. 110 (2005) 475–490.
[14] P.L. Deininger, M.A. Batzer, Mammalian retroelements, Genome Res. 12 (2002) 1455–1465.
[15] H.H. Kazazian Jr., Mobile elements: drivers of genome evolution, Science 303 (2004) 1626–1632.
[16] L. Lin, S. Shen, A. Tye, J.J. Cai, P. Jiang, B.L. Davidson, Y. Xing, Diverse splicing patterns of exonized Alu elements in human tissues, PLoS Genet. 4 (2008) e1000225.
[17] W. Makalowski, G.A. Mitchell, D. Labuda, Alu sequences in the coding regions of mRNA: a source of protein variability, Trends Genet. 10 (1994) 188–193.
[18] R. Sorek, G. Ast, D. Graur, Alu-containing exons are alternatively spliced, Genome Res. 12 (2002) 1060–1067.
[19] M. Wu, L. Li, Z. Sun, Transposable element fragments in protein-coding regions and their contributions to human functional proteins, Gene 401 (2007) 165–171.
[20] G. Lev-Maor, R. Sorek, E.Y. Levanon, N. Paz, E. Eisenberg, G. Ast, RNA-editing-mediated exon evolution, Genome Biol. 8 (2007) R29.
[21] G. Lev-Maor, R. Sorek, N. Shomron, G. Ast, The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons, Science 300 (2003) 1288–1291.
[22] O. Ram, S. Schwartz, G. Ast, Multifactorial interplay controls the splicing profile of Alu-derived exons, Mol. Cell Biol. 28 (2008) 3513–3525.
[23] N. Sela, E. Kim, G. Ast, The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates, Genome Biol. 11 (2010) R59.
[24] R. Sorek, G. Lev-Maor, M. Reznik, T. Dagan, F. Belinky, D. Graur, G. Ast, Minimal conditions for exonization of intronic sequences: 5′ splice site formation in Alu exons, Mol. Cell 14 (2004) 221–231.
[25] M. Krull, J. Brosius, J. Schmitz, Alu-SINE exonization: en route to protein-coding function, Mol. Biol. Evol. 22 (2005) 1702–1711.
[26] M. Möller-Krull, A. Zemann, C. Roos, J. Brosius, J. Schmitz, Beyond DNA: RNA editing and steps toward Alu exonization in primates, J. Mol. Biol. 382 (2008) 601–609.
[27] S.S. Singer, D.N. Mannel, T. Hehlgans, J. Brosius, J. Schmitz, From "junk" to gene: curriculum vitae of a primate receptor isoform gene, J. Mol. Biol. 341 (2004) 883–886.
[28] M. Krull, M. Petrusma, W. Makalowski, J. Brosius, J. Schmitz, Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs), Genome Res. 17 (2007) 1139–1145.
[29] N. Sela, B. Mersch, N. Gal-Mark, G. Lev-Maor, A. Hotz-Wagenblatt, G. Ast, Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu’s unique role in shaping the human transcriptome, Genome Biol. 8 (2007) R127.
[30] R. Baertsch, M. Diekhans, W.J. Kent, D. Haussler, J. Brosius, Retrocopy contributions to the evolution of the human genome, BMC Genomics 9 (2008) 466.
[31] G.A. Mitchell, D. Labuda, G. Fontaine, J.M. Saudubray, J.P. Bonnefont, S. Lyonnet, L.C. Brody, G. Steel, C. Obie, D. Valle, Splice-mediated insertion of an Alu sequence inactivates ornithine delta-aminotransferase: a role for Alu elements in human mutation, Proc. Natl. Acad. Sci. USA 88 (1991) 815–819.
[32] M. Gerstein, D. Zheng, The real life of pseudogenes, Sci. Am. 295 (2006) 48–55.
[33] L. Lin, P. Jiang, S. Shen, S. Sato, B.L. Davidson, Y. Xing, Large-scale analysis of exonized mammalian-wide interspersed repeats in primate genomes, Hum. Mol. Genet. 18 (2009) 2204–2214.
[34] J.W. Huh, Y.H. Kim, D.S. Kim, S.J. Park, S.R. Lee, S.H. Kim, E. Kim, S.U. Kim, M.S. Kim, H.S. Kim, K.T. Chang, Alu-derived Old World monkeys exonization event and experimental validation of the LEPR gene, Mol. Cells 30 (2010) 201–207.
[35] Y. Xing, C. Lee, Alternative splicing and RNAselection pressure - evolutionary consequences for eukaryotic genomes, Nat. Rev. Genet. 7 (2006) 499–509.
[36] K. Nishikura, Functions and regulation of RNA editing by adar deaminases, Annu. Rev. Biochem. 79 (2010) 321–349.
[37] M.R. Wallace, L.B. Andersen, A.M. Saulino, P.E. Gregory, T.W. Glover, F.S. Collins, A de novo Alu insertion results in neurofibromatosis type 1, Nature 353 (1991) 864–866.
[38] I. Vorechovsky, Transposable elements in disease-associated cryptic exons, Hum. Genet. 127 (2010) 135–154.
[39] M. Sakurai, T. Yano, H. Kawabata, H. Ueda, T. Suzuki, Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome, Nat. Chem. Biol. 6 (2010) 733–740.

[40] J. Schmitz, C. Roos, H. Zischler, Primate phylogeny: molecular evidence from retroposons, Cytogenet. Genome Res. 108 (2005) 26–37.
[41] M. Kimura, The length of time required for as electively neutral mutant to reach fixation through random frequency drift in a finite population, Genet. Res. 15 (1970) 131–133.
[42] A.F. Smit, A.D. Riggs, MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation, Nucleic Acids Res. 23 (1995) 98–102.
[43] N. Gilbert, D. Labuda, CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs, Proc. Natl. Acad. Sci. USA 96 (1999) 2869–2874.
[44] J. Brosius, The fragmented gene, Ann. N. Y. Acad. Sci. 1178 (2009) 186–193.
[45] R. Sorek, The birth of new exons: mechanisms and evolutionary consequences, RNA 13 (2007) 1603–1608.
[46] V. Gotea, W. Makalowski, Do transposable elements really contribute to proteomes? Trends Genet. 22 (2006) 260–267.
[47] W.C. Warren, L.W. Hillier, J.A. Marshall Graves, E. Birney, C.P. Ponting, F. Grutzner, K. Belov, W. Miller, L. Clarke, A.T. Chinwalla, S.P. Yang, A. Heger, D.P. Locke, P. Miethke, P.D. Waters, F. Veyrunes, L. Fulton, B. Fulton, T. Graves, J. Wallis, X.S. Puente, C. Lopez-Otin, G.R. Ordonez, E.E. Eichler, L. Chen, Z. Cheng, J.E. Deakin, A. Alsop, K. Thompson, P. Kirby, A.T. Papenfuss, M.J. Wakefield, T. Olender, D. Lancet, G.A. Huttley, A.F. Smit, A. Pask, P. Temple-Smith, M.A. Batzer, J.A. Walker, M.K. Konkel, R.S. Harris, C.M. Whittington, E.S. Wong, N.J. Gemmell, E. Buschiazzo, I.M. Vargas Jentzsch, A. Merkel, J. Schmitz, A. Zemann, G. Churakov, J.O. Kriegs, J. Brosius, E.P. Murchison, R. Sachidanandam, C. Smith, G.J. Hannon, E. Tsend-Ayush, D. McMillan, R. Attenborough, W. Rens, M. Ferguson-Smith, C.M. Lefevre, J.A. Sharp, K.R. Nicholas, D.A. Ray, M. Kube, R. Reinhardt, T.H. Pringle, J. Taylor, R.C. Jones, B. Nixon, J.L. Dacheux, H. Niwa, Y. Sekita, X. Huang, A. Stark, P. Kheradpour, M. Kellis, P. Flicek, Y. Chen, C. Webber, R. Hardison, J. Nelson, K. Hallsworth-Pepin, K. Delehaunty, C. Markovic, P. Minx, Y. Feng, C. Kremitzki, M. Mitreva, J. Glasscock, T. Wylie, P. Wohldmann, P. Thiru, M.N. Nhan, C.S. Pohl, S.M. Smith, S. Hou, M. Nefedov, P.J. de Jong, M.B. Renfree, E.R. Mardis, R.K. Wilson, Genome analysis of the platypus reveals unique signatures of evolution, Nature 453 (2008) 175–183.
[48] J.R. Wible, G.W. Rougier, M.J. Novacek, R.J. Asher, Cretaceous eutherians and Laurasian origin for placental mammals near the K/T boundary, Nature 447 (2007) 1003–1006.
[49] G.E. Liu, L.K. Matukumalli, T.S. Sonstegard, L.L. Shade, C.P. Van Tassell, Genomic divergences among cattle, dog and human estimated from large-scale alignments of genomic sequences, BMC Genomics 7 (2006) 140.
[50] P. Perelman, W.E. Johnson, C. Roos, H.N. Seuanez, J.E. Horvath, M.A.M. Moreira, B. Kessing, J. Pontius, M. Roelke, Y. Rumpler, M. Paula, M.P.C. Schneider, S.J. O'Brien, J. Pecon-Slattery, A molecular phylogeny of living primates, PLoS Genet. 7 (2011) e1001342.
[51] D.C. Hughes, MIRs as agents of mammalian gene evolution, Trends Genet. 16 (2000) 60–62.