Birth and Evolution of Human Exons

Maria Dulcetti Vibranovski, *The University of Chicago, Chicago, Illinois, USA*
Noboru Jo Sakabe, *Institute of Chemistry, Universidade de São Paulo, São Paulo, Brazil*
Manyuan Long, *The University of Chicago, Chicago, Illinois, USA*

Advanced article

Article Contents

doi: 10.1002/9780470015902.a0020777

Exons can be independently gained or created in genes by exon shuffling, duplication and mutation of intronic sequences, especially repetitive elements as Alu. Such new exons are more prone to be alternatively spliced, included in transcripts in low frequencies and evolve faster than ancient exons. Alternative splicing is an important component of the evolution of new exons since it allows new gene parts to be slowly added to existing genes without compromising the original protein function. Eventually, new exons may be integrated to the original gene and evolve new functions.

Introduction

Eukaryotic genes are, in general, composed by an intercalated structure of exons and introns. After transcription of the gene into a pre-messenger ribonucleic acid (pre-mRNA), such intercalated structure undergoes a process named splicing by which introns are excised and exons are fused in one continuous molecule, the mature mRNA. **See also**: Alternative Splicing: Evolution

Although exons belong to a higher structure, the gene, because they have to be independently recruited during splicing, they are also somewhat stand-alone units. For this reason, an individual exon can be gained or lost by a gene, as well as evolve in a different rate than the remaining exons. Here we discuss how new exons appear in human genes and how they evolve. For practical purposes, new exons are considered to be those that can only be found in human, when comparing orthologous genes.

Gain of Pre-existing Exons

New exons can be gained by a gene as an entire, functional (coding and amenable to splicing) unit. For example, an existing, ancient exon can be duplicated by unequal crossover during meiosis, leading to the loss of an exon in one allele and gain in another (originating a new exon in this gene). This mechanism of intra-gene gain seems to be the source of a non-negligible fraction of new exons (at least 6% of all exons, Fedorov *et al*., 1998). **See also**: Insertion and Deletion of Exons during Human Gene Evolution

Exons can also be 'shuffled' in a genome, i.e. moved among distinct genes, resulting in inter-gene gain. Crossover between nonhomologous genes can promote the gain of an exon. The chances that this new, shuffled exon remains in the acceptor gene are associated with the preservation of the reading frame. The new exon should encode a protein segment that does not change the reading frame. **See also**: Insertion and Deletion of Exons during Human Gene Evolution; Exons: Shuffling

Although exon shuffling seems to have played an important role in the evolution of metazoan proteins, there are evidences that it did not play a major part in the gain of new human exons (Zhang and Chasin, 2006). **See also**: Exons and Protein Modules

Existing exons can also be gained by genes by retrotransposition: partial or complete mRNAs containing one or more exons retrotranscribed into complementary deoxyribonucleic acid (cDNA) sequences can be inserted in a genomic region through recombination. If the retrotransposition occurs within an intron, the result will be the insertion of an exon or set of exons into a different gene. In the latter case, as introns intervening the retrotranscribed exons are lost in retrotranscription, only one single, long new exon will be created. **See also**: Insertion and Deletion of Exons during Human Gene Evolution

The Birth of New Exons

Intronic sequences can also be a source of new exons. Mutations can lead to the 'birth' of exons by creating splice sites that allow intronic segments to be recognized by the splicing machinery.

Repetitive sequences located in introns seem to be an important 'cradle' for new exons, being detected in a large fraction of human exons and representing most new exons (Zhang and Chasin, 2006). One important family of repetitive elements in the human genome is the SINEs (short interspersed nuclear elements), of which the primate-specific Alu is an important member. Alu sequences are retrotransposons of about 280 nt, GC-rich and widespread in the human genome, constituting at least 10% of the sequence. These retrotransposons are similar to transcripts produced by RNA polymerase II, containing a poly A tail. **See also**: Short Interspersed Elements (SINEs)

These repetitive elements contain motifs that resemble splice sites. The antisense Alu consensus presents nine sites with at least four positions similar to the 5′ splice site consensus AG/GTRA. Potential 3′ splice sites also exist, three sites in the antisense Alu consensus differ from the YYYYYYYYYYNYAG/R consensus by no more than two nucleotides. The polypyrimidine tract of the 3′ splice site (poly Y) is formed by Ts in the reverse complement of the poly A tail (Makalowski *et al*., 1994). This way, just a few mutations in these motifs may be enough to create splice sites and turn an Alu segment into an exon (Lev-Maor *et al*., 2003; Sorek *et al*., 2004), mainly its antisense. The exonization of Alus, however, may require the presence of an exonic splicing enhancer (ESE) to promote splicing in addition to mutations creating splice sites (Lei *et al*., 2005) or be limited by the eventual presence of regulatory motifs that inhibit splicing. **See also**: Exonic Splicing Enhancers

Another way by which Alu-exons can be created, although not permanently in the genome, is by A-to-I RNA-editing. This 'dynamic' mode of creation, that involves an enzymatic process, generates exons in intronic regions of the mRNA by editing one nucleotide and creating a splice site 'on the fly' (Lev-Maor *et al*., 2007). The intron between exons 7 and 9 of the *NARF* gene contains an Alu element flanked by 3′-AA/GU-5′ nucleotides. A-to-I editing of mRNAs from this gene yields a functional 3′-AI/GU-5′ splice site pair, since the spliceosome recognizes inosine as guanosine. The level of inclusion of such exons in transcripts is therefore controlled by the level of RNA-editing and was found to vary among different tissues (Lev-Maor *et al*., 2007). **See also**: mRNA Editing

Not only Alu gives birth to new exons. Other repetitive elements as long interspersed nuclear elements (LINEs), long terminal repeat (LTR) and other DNA transposons have also been exonized, contributing to the creation of a considerable fraction of human exons (Zhang and Chasin, 2006).

Independently of the mechanism of exon gain or birth, retention of new exons as parts of the existing gene is constrained by several factors. Mainly, the presence of stop codons would truncate the protein or lead to degradation of the mRNA by nonsense mediated decay (Maquat, 2005) and the exons need to present minimum conditions to be spliced. Indeed, new exons are avoided in coding regions, being more frequently observed in untranslated regions (UTRs) than ancient exons, mainly in the 5′ ($>30\%$ of the recent exons, Zhang and Chasin, 2006). New exons in the 3′ UTR would probably trigger degradation of mRNAs by nonsense mediated decay.

Alternative Splicing and the Gain/Birth of New Exons

Despite the avoidance of the presence of new exons in the coding region of genes, there might be a mechanism by which coding exons can be allowed to appear. In humans and other vertebrates, most genes were shown to be spliced in more than one way, i.e. not necessarily all exons are incorporated in all transcribed mRNAs (alternative splicing). Although this phenomenon seems to have many physiological functions, it may also be an evolutionary mechanism to slowly incorporate new gene parts without affecting the original gene completely. In the case of exon creation, the most relevant type of alternative splicing would be the alternative inclusion of an entire exon (exon skipping). **See also**: Alternative Splicing: Evolution

An association between acquired exons and alternative splicing was noticed in both exons duplicated in tandem and Alu-derived exons. In the former case, duplicated exons were found to be included in a mutually exclusive fashion (both exons never occurred together in the same transcript), so as to avoid disruption of the original protein in at least 10% of all cases of alternative exon inclusion (Kondrashov and Koonin, 2001). Supposedly, the new duplicated exon can potentially evolve a new function, since the original one is maintained in other transcripts. In the latter case, Alu-containing exons were found to be more often alternatively included than other exons and only in a minority of transcripts. Also, most Alu-exons are alternatively spliced (Sorek *et al*., 2002).

The frequency at which the new exon is used is highly relevant. By diverting only a minor part of the transcribed pre-mRNAs to include the new exon, the cell can maintain the original gene function, while allowing a new protein product to be 'tested' (Modrek and Lee, 2003). Indeed, an analysis of exons included in the minority of transcripts showed that they evolve faster (Xing and Lee, 2005), i.e. they are under less negative selective pressure to maintain their function. Extending these observations, most new exons were found to be alternatively spliced and are included in the minority of transcripts and evolve at higher rates than older ones as measured by the density of SNPs and the number of nonsynonymous/synonymous changes (Zhang and Chasin, 2006).

The mechanistic reason why new exons are frequently subjected to alternative splicing and are included in the minority of transcripts is because they bear weak splice sites. As these splice sites appeared by mutation of existing intronic sequences, it might take some time for them to become strong splice sites, and then allow inclusion of the exon in the majority or in all transcripts.

Although according to this logic alternative splicing would be a primitive trait, it is possible that in some cases it is a derived state, since it can potentially increase the diversity of a proteome and contribute to the organism's plasticity and fitness.

In fact, there is some evidence of a transition in the composition of splice sites from organisms that do not present alternative splicing like *Saccharomyces cerevisiae* and *Schyzosacharomyces pombe* to mice and human (which present abundant alternative splicing, Ast, 2004). Intronic nucleotides of the splice sites of *S. cerevisiae* are highly conserved, allowing accurate splicing every time. These nucleotides in the distantly related *S. pombe*, however, are more variable, though still considerably conserved. Splice

sites of mice and human, however, present important differences to the former organisms. There is an increased conservation of nucleotides in the exonic portion and a decreased conservation of nucleotides in the intronic portion. Apparently, vertebrate splice sites are less 'rigid', they are spread along more nucleotides, but are more variable, thus allowing alternative splicing. However, it is not yet possible to discard that the evolutionary direction was the loss of alternatively splicing in the fast growing *S. pombe* and *S. cerevisiae*, since diversity can also be a risk.

Regulatory Elements in Exons are Under Selective Pressure

Once new exons have been incorporated in the host genes, their evolution rate is gradually slowed down, due to selective pressure to maintain the new function. Interestingly, protein coding exons seem to be under strong selection at the nucleotide level, even at the expense of the amino acid composition (Parmley *et al*., 2007). This is because splicing of exons involves short degenerate sequences named splicing regulatory elements, often located in exons, and in higher density near splice borders, that help in the recruitment of the splicing machinery. Such pressure to conserve regulatory elements seems to be even more pronounced in exons included in the minority of transcripts. It is possible that these poorly spliced sequences depend on regulatory sequences to be spliced in a greater extent and are therefore under stronger selection. As alternatively spliced exons grow older, the selective pressure on the RNA is increased even more (Xing and Lee, 2005). **See also**: Exonic Splicing Enhancers

As a consequence of increased selective pressure in nucleotide composition, since shorter exons are more densely packed with regulatory elements, they seem to evolve more slowly than longer ones (Parmley *et al*., 2007).

A summary of the events related to the birth of a new exon from Alu elements described earlier can be found in **Figure 1**.

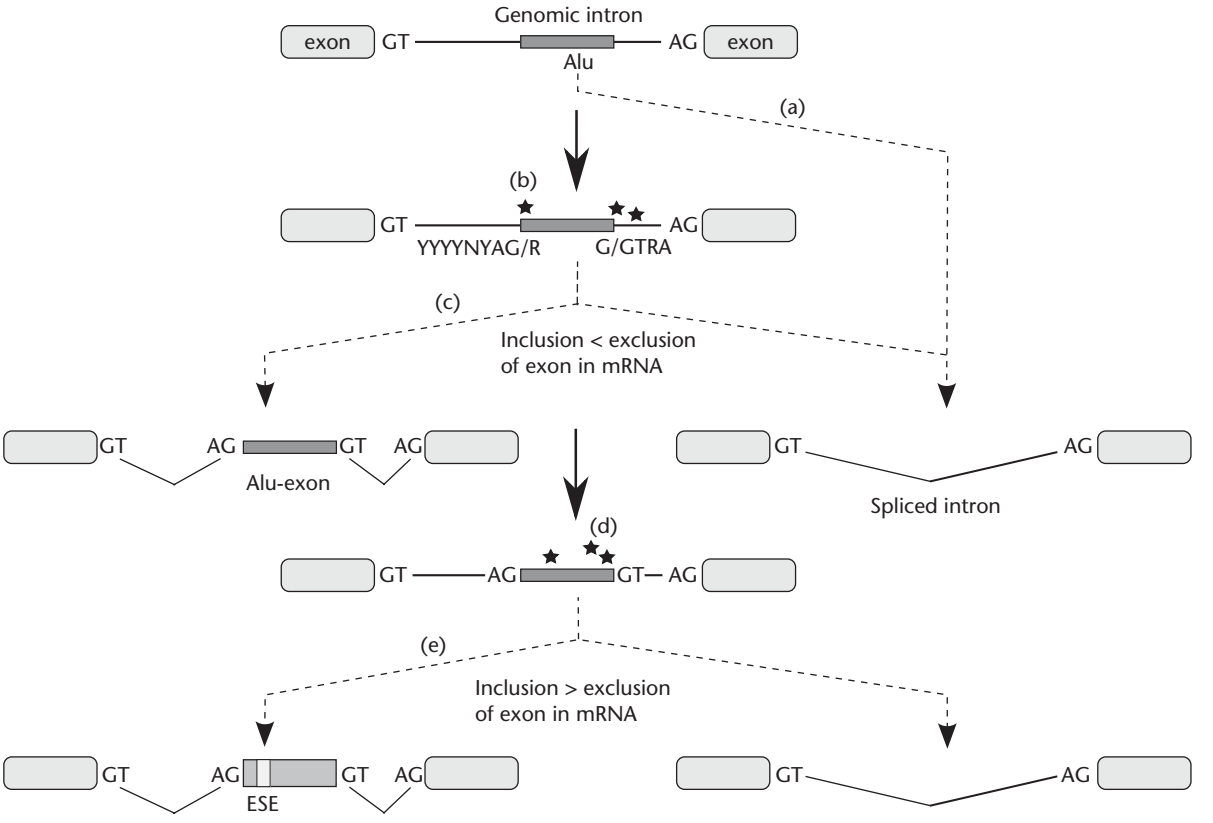

Figure 1 A general pathway for the birth of a new exon from intronic sequences. Solid lines indicate 'evolutionary pathways' and dashed lines 'splicing pathways'. Stars represent mutations in the DNA. (a) Original (ancient) splicing pathway. (b) An intronic segment (mainly repetitive sequences as Alu) undergoes mutations (stars) that create weak splice sites (and perhaps required ESEs). (c) The segment may be incorporated in a few mRNA transcripts (alternatively spliced), since it bears weak splicing signals. (d) The newly born exon is free to evolve, because it does not compromise the original gene function, thus the observed higher evolutionary rates. (e) Eventually, the new exon might undergo mutations (stars) that create proper splicing regulatory elements (ESEs) or improve splice sites, increasing its inclusion frequency (it might even become constitutively spliced). The high evolutionary rate might also allow the exon to evolve a new function. Owing to selective pressure on both the RNA and DNA levels, evolutionary rate tends to slow down. New exons can remain alternative, providing diversity to the genome.

References

Ast G (2004) How did alternative splicing evolve? *Nature Review Genetics* **5**: 773–782.

Fedorov A, Fedorova L, Starshenko V, Filatov V and Grigor'ev E (1998) Influence of exon duplication on intron and exon phase distribution. *Journal of Molecular Evolution* **46**: 263–271.

Kondrashov FA and Koonin EV (2001) Origin of alternative splicing by tandem exon duplication. *Human Molecular Genetics* **10**: 2661–2669.

Lei H, Day INM and Vorechovský I (2005) Exonization of AluYa5 in the human ACE gene requires mutations in both 3′ and 5′ splice sites and is facilitated by a conserved splicing enhancer. *Nucleic Acids Research* **33**: 3897–3906.

Lev-Maor G, Sorek R, Levanon EY *et al*. (2007) RNA-editing-mediated exon evolution. *Genome Biology* **8**(2): R29.

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.

Makalowski W, Mitchell GA and Labuda D (1994) Alu sequences in the coding regions of mRNA: a source of protein variability. *Trends in Genetics* **10**: 188–193.

Maquat LE (2005) Nonsense-mediated mRNA decay in mammals. *Journal of Cell Science* **118**: 1773–1776.

Modrek B and Lee CJ (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nature Genetics* **34**: 177–180.

Parmley JL, Urrutia AO, Potrzebowski L, Kaessmann H and Hurst LD (2007) Splicing and the evolution of proteins in mammals. *PLoS Biology* **5**: e14.

Sorek R, Ast G and Graur D (2002) Alu-containing exons are alternatively spliced. *Genome Research* **12**: 1060–1067.

Sorek R, Lev-Maor G, Reznik M *et al*. (2004) Minimal conditions for exonization of intronic sequences: 5′ splice site formation in Alu exons. *Molecular Cell* **14**: 221–231.

Xing Y and Lee C (2005) Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. *Proceedings of the National Academy of Sciences of the USA* **102**: 13526–13531.

Zhang XH and Chasin LA (2006) Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proceedings of the National Academy of Sciences of the USA* **103**: 13427–13432.

Further Reading

Chamary JV, Parmley JL and Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Review Genetics* **7**: 98–108.

Xing Y and Lee C (2006) Alternative splicing and RNA selection pressure – evolutionary consequences for eukaryotic genomes. *Nature Review Genetics* **7**: 499–509.