

Introns: functions and origin



Very little is known about the genomic architecture even though there is widespread proliferation of the introns. Evidences from the recent advances emerged through the large-scale genomic sequencing projects and the functional analysis of the mRNA-processing events support the idea that the spliceosomal introns were not only present in the early eukaryote but also diverged into a minimum of two eukaryotic classes in the early stages of evolution (Lynch, 2002). The origin of introns is an issue and several researchers proposed different theories on the origin of introns. The present day debate goes on the important issues of evolution such as the introns-early versus introns-late, the mini gene hypothesis, the protein-splice site hypothesis, rates of intron gain, loss, and the ratio of parallel gain. Introns are not distributed uniformly, along with the excess of phase 0 introns. The introns early theory proposes that introns were present in the last universal common ancestor (LUCA) of prokaryotes and eukaryotes (Gilbert, 1978). To say further, it is postulated that the premier genetic elements encoded small domains, similar in length to typical modern exons, which recombined via non-coding intronic sequences present in some of these elements to facilitate protein evolution (Roy, 2003). During subsequent evolutionary history, the introns underwent diverged evolutionary courses in the different lineages: they were erased in prokaryote lineages, but maintained in eukaryotes as introns through the appearance of the spliceosome (Gilbert, 1978). The loss of introns in prokaryotes has been explained as “ genome streamlining” (Roy, 2003). According to the streamlining hypothesis, the replication efficiency is the cardinal object of the main pressure in the evolution of prokaryotes, and thus non-essential parts of the genomes would be eliminated. Introns would not exist under such intense negative selection.

The introns late theory supposes that spliceosomal introns arose in remotely antiquated eukaryotes from self-splicing introns. These group II introns were present in the mitochondrial organelles of endosymbionts, and invaded previously undivided genes and intron-less genomes, and the spliceosome evolved as a way to remove them (Cavalier-Smith, 1991). The argument for self-splicing introns giving rise to spliceosomal introns and the spliceosomes is based on functional and structural similarities between self-splicing group II introns and spliceosomal introns. In both types of introns, the 5' end becomes bound to an adenine near the 3' end, forming a lasso structure that is excised (Newman, 1997). Furthermore, the group II introns appear to be phylogenetically limited to eubacteria (Bonen & Vogel, 2001), and the organellar genomes of eukaryotes, such as mitochondrial genomes, are considered to share common ancestors with a variety of eubacteria (Gray, 1999). Genes in the organelles had been transferred to the nucleus in a large scale (Gray, Burger, & Lang, 1999), which may have resulted in group II intron-like elements invading into the eukaryote nucleus.

Even though the results of research cannot give a definitive answer, Nguyen, Yoshihama, and Kenmochi (2005) report that the evolution of the eukaryotes is not characterized by a general decrease in the intron density, as predicted in the intron early theory. In the intron evolution, during the period between the last common ancestor of the eukaryotes and the crown ancestor there is an either a decrease or increase in the intensity of introns. There is an intensive debate on the intron early and the intron late even though some lines of support were provided in favor of the intron late primarily as the decrease in density of introns was not observed. However, it was observed

that there is no conservation of the position of introns in the 25 cytoplasmic ribosomal proteins (CRP) genes of the archeal region of the mitochondrial genes of the bacterial origin, that are supposed to have diverged at the LUCA (Nguyen, Yoshihama, & Kenmochi, 2006). Many approaches are directed to infer that evolution of intron takes the patterns of the conservation of the positions of introns. The vital components to the translational machinery for the cellular life are the ribosomes and they are conserved to a high extent throughout the evolutionary process. This has created a possibility to compare the ribosomal proteins across the distantly diverged species, and the mitochondrial ribosomal proteins (MRP) are supposed to have evolved from bacteria, as there is a considerable homology between the MRP and the bacterial proteins, while the CRP have evolved independently. Genes of MRP are passed to the nuclear genome and this genome contains spliceosomal introns. In this situation, it is possible to determine the existence of the spliceosomal introns in the ancestors through comparisons of MRP and CRP intron-exon structures. Yoshihama, Nakao, Nguyen, and Kenmochi (2006) reports that no clarity of the conservation of the intron position has been observed between the genes of CRP and MRP and this indicates that in the last common ancestor of the MRP and CRP genes, the spliceosomal introns was absent. With the available results, Yoshihama, Nakao, Nguyen, and Kenmochi (2006) support the intron-late theory. Palmer and Langsdon (1991) report that the evidence accumulated is in favor of the restricted phylogenetic distribution of introns. They strongly support the view that the introns are inserted late in the eukaryotic evolution and there is no role for exon shuffling in the primordial assemblage of genes.

The construction of the first genes is associated with the exon shuffling, and the intron position is influenced by different dynamics of the phylogenetic distribution. The statistical analysis tests of de Souza, Long, Klein, Roy, Lin, and Gilbert (1998), support the intron/exon structure of the genes is a consequence of the assemblage of the first genes by the exon shuffling at the early stages of evolution. Some experiments have suggested that introns have gained during the evolution of eukaryotes at the proto-splice sites. The conserved sequence, which flanks the introns, may have been used as a site for the intron gain during the process of evolution (Dibb & Newman, 1989). The clear example of the insertion of the spliceosomal introns are observed in the U2 and U6 small nuclear genes of certain yeast species. The coding sequences that flank these introns, however, are random and the example cannot suggest the proto-splice site for the real insertion of introns. An excess of the symmetric exons existing in the modern or the ancient conserved genes, supports the key role of the exon shuffling in the early and late stages of evolution before the divergence of eukaryotes and prokaryotes (Long, de Souza, Rosenberg, & Gilbert, 1998). The intron-exon structures have an intricate history, and evolution occurred with different stages of introns and exons. The ancient introns prefer to be in the protein module boundaries, as they are involved in the exon shuffling. Modern introns are inserted at the proto-splice sites. Fedorov, Cao, Saxonov, de Souza, Roy, and Gilbert (2001) observed that the ancient and the modern introns are distributed in the ancient conserved regions and the non-ancient conserved samples of the genes. Excess of phase 0 intron positions in the boundary region of the modules of the ancient proteins that are common to the eukaryotes and prokaryotes support the idea that, the introns are used in the

<https://assignbuster.com/introns-functions-and-origin/>

construction of the ancient genes, by the exon shuffling of the modules at the early stages of evolution (Fedorov, Cao, Roy, & Gilbert, 2003). In conclusion, Introns early and introns late supporters now agree that both the spliceosome and spliceosomal introns originated long before the most recent common ancestor of living eukaryotes, and that little of the primigenial exon/intron boundary distribution is left due to rapid intron turnover (Wolf, Kondrashov, & Koonin, 2001). In addition, different lineages reveal disparate intron loss and gain patterns; intron sliding is a rare event and the majority disclose simply one base relocation (Rogozin, Lyons-Weiler, & Koonin, 2000; Sakharkar, Tan, & de Souza, 2001).

A new theory of intron origin was emerged in 1998 by Jeffares, Poole, and Penny, the “introns first” theory, which posited that introns and the spliceosome are remnants from the RNA world. A potential RNA must be so essential to metabolism that various systems rely on such RNA and its existence could not be substituted by proteins, and that this RNA should also be prevalent in different organisms and catalytic in order for being conceived as remnants from the RNA world. Among manifold RNA molecules, small nucleolar RNAs (snoRNAs) achieve the requirements. Furthermore, as snoRNAs are necessitated to guide chemical modifications of ribosomal RNAs (rRNAs), they must antedate the origin of genetically encoded proteins. Because snoRNA genes are often encoded by introns of proteins involved in response to stimulus (e. g., heat-shock proteins) and ribosome synthesis, at least these introns must forerun the protein-coding exons that surround them. It is then logical to postulate that a splicing mechanism must have been present as well. The origin of spliceosomes in the RNA world would be

consistent with the assembly of ribonucleoproteins (RNPs) due to its essentiality in the snoRNA maturation, which plays a crucial role in the processing of rRNAs, a vital element of the proto-ribosome, from spliced introns. To date, proponents of all hypotheses about the origin of introns lack sufficient evidence to refute the other theories, and the controversy continues.

Functions of introns

It is predicted that spliceosomal introns have no special or general functions, and their non-existence in the prokaryotes and massive losses in the eukaryotic lineages, suggests the fact that there is no essential function for introns (Roy & Irimia, 2008). Intron presence in eukaryotes has several evident shortcomings that include waste of time and energy during gene expression while transcribing long intronic fragments of pre-mRNA molecules, and possible errors in ordinary splicing because long introns contain abundant false splicing sites (pseudo exons). Hence, introns must be endowed with some functional gains to counteract these drawbacks. Several important functions of introns have been uncovered that counter the concept of introns as selfish, non-functional genomic elements. For example, Roy and Irimia (2008) reported that a very short sequence of introns plays a role in genomic stability, chromatin structure and also in promoting recombination. The evolution of the spliceosomal introns by the removal of the restrictions of the internal sequences in the genome might be free to drift with minimal constraints. In this situation, it would be reasonable to predict that any mutational change that is beneficial might have a positive selection value and can be retained. It can be speculated that, with the entry of introns, a

new round of molecular evolution can be initiated based on RNA rather than protein. Because of this process, alternative splicing increases the scope of gene complexity, which allows the production of a set of proteins related with different properties from a single gene. Although small, the sequences involved in alternative splicing are the bearers of the many of the signatures of information, which includes high sequence complexity, non-random base distribution, and intriguing patterns of conservation. Regulated splicing can serve some biological functions, which better explains the retention of the occasion for introns in the reduced eukaryotic genes. The last common ancestor of the extant eukaryotes has a complex spliceosomes and a large number of introns, and those introns might be having degenerate sequences. (Roy & Irimia, 2008). Irimia, Rukov, Penny, and Roy (2007) state that the ancient genes of the modern organisms have high levels of alternative splicing and these genes might be having many introns in their common ancestors, which is an important requirement for alternative splicing. Interpretation based on the experimental results of Irimia, Rukov, Penny, and Roy (2007) reveal that alternative splicing might have appeared before the rise of the multicellular organisms, and the alternative splicing might be having an important role in the biological functions of the ancient unicellular organisms. A previous study has identified that the secondary structure of introns coordinates splice site pairing, prevents exon skipping and enforces the inclusion of the internal exons in yeast. The elements of these types in vertebrate genes might assist in the splicing of the very large introns and in the evolution of alternative splicing (Howe & Ares, 1997). Currently, there are five alternative splicing modes that are commonly recognized, including

5'-splice sites, alternative 3'-splice sites, exon skipping, mutually exclusive exons, and retained introns (Black, 2003).

Introns have regulatory sequences and the transcription of the ribosomal protein genes is regulated extensively in a coordinated way; however, it has been reported that introns have no direct involvement with the regulation of the transcription of the highly expressed group of ribosomal protein genes (Zhang, Vingron, & Ropcke, 2008). Liu, He, Amasino, and Chen (2004), on the basis of their research in the plant species *Arabidopsis* report that the transposable elements inserted in the intron has a role in evolution and gene expression. In a recent research it was found that intronic sequences are linked to the putative regulatory mechanism in order to modulate the properties of the membranes and ion channel gradients of the hippocampal neurons (Bell et al., 2008; Tsigos & Rigoutsos, 2008). The second intron of the human nestin gene has been reported to bear an evolutionarily conserved region conducting gene expression to central nervous system (CNS) progenitor cells and to early neural crest cells (Lothian & Lendahl, 1997). Likewise, in order to express the human apolipoprotein B gene in liver, the second intron of this gene is crucial (Brooks et al., 1994). Cenik, Derti, Mellor, Berriz, and Roth (2010) also reveal that introns within the 5'UTR of human genes enhance the expression of some genes in a length dependent manner. Transcriptional regulatory functions are performed by some intron sequence motifs. Positive regulatory sequences or the enhancers along with the negative regulatory sequences or repressors have been reported in first introns of considerable number of human genes. The introns play a regulatory role in the processing of the primary transcript by

the modulation of the mRNA splicing or they influence the stability of the mRNA through the interactions of RNA-RNA, RNA-DNA, or the RNA-Protein (Cooper, 1999).

Introns contain several types of non-coding, but functional, RNA sequences. The snoRNAs and microRNAs (miRNAs) play a key role in a range of processes that includes biogenesis of the ribosome and gene regulation. SnoRNAs are located inside introns, and are produced during post-splicing processing of intronic RNA (Hüttenhofer, Brosius, & Bachellerie, 2002). The snoRNAs may also modify the other RNA targets that include smaller nuclear RNAs (snRNAs) of the spliceosome, and they also play a role in the regulation in the alternative splicing of mRNA (Hoepfner, Simon, White, Jeffares, & Poole, 2009), and in guiding the process of pseudouridylation and methylation in pre-rRNA by complementary pairing of their guide sequences with rRNAs (Maden & Hughes, 1997). MiRNA is frequently found inside introns (Bartel, 2004). They are short RNA molecules and are capable of “silencing” genes by the binding of complementary sequences in the 3'UTR of one or more target mRNAs. In vertebrates, 40-70% of miRNAs seem to be expressed from introns of both coding and non-coding transcripts. Intronic miRNAs are less common in worms and flies, 15% and 39%, respectively, in protein coding genes (Griffiths-Jones, Saini, van Dongen, & Enright, 2008). Hoepfner, Simon, White, Jeffares, and Poole (2009) report that the intronic snoRNAs and miRNAs are more likely and significantly stable than the intergeneric mRNAs. A minority of the snoRNAs and miRNAs remain in the same location throughout the process of evolution and this may confer an advantage. A transposon inserted into the intron becomes an intronic

miRNA, by taking the advantage of protein synthesis machinery for the processing and maturation. These miRNAs play an important role in the diverse biological systems of various organisms (Ying, 2008). However, there may be many more undiscovered functions and functional elements in spliceosomal introns.