

Origin of Spliceosomal Introns: Mainstream Hypotheses Comparison

Litianqing Yuan

*College of Science, Xi'an Jiaotong-Liverpool University, Suzhou, China
Litianqing.Yuan21@student.xjtlu.edu.cn*

Abstract. Spliceosomal introns widely exist in the RNA of eukaryotic cells. In recent years, spliceosomal introns are believed to be important for gene expression in eukaryotes. However, the specific functions, connections, and differential mechanisms of introns located at distinct hnRNA positions remain poorly understood. The research value and mysteries of spliceosomal intron catch people's attention. Among the many mysteries in this field, the origin and evolution route is one of the most significant and enlightening research directions. Due to the lack of direct evidence, however, the research progress about the spliceosomal introns has been slow. There are still many mysteries about the origin and evolution route of introns, so a definite conclusion is still impossible based on existing discoveries and experimental results, thus, summarizing the available information can be heuristic for researchers. This article will summarize existing research and unknown fields about three commonly believed origins of spliceosomal introns: introns-early, introns-late, and introns-first. A perfect conclusion is impractical based on current evidence, but a relatively reasonable hypothesis is still valuable. This article will review previous studies in order to organize the line of thought, to help further research.

Keywords: intron's Origin, intron's evolution route, Intron gain, models of introns' evolution route, introns-early and introns-late

1. Introduction

The spliceosomal intron is a name for the non-translated RNA sequence, which is always excised during the modification of hnRNA. The spliceosomal intron has been found for several decades, and some of its functions have been recognized, but it is still not completely understood, especially the details of its evolutionary path. For example, it remains an open question how spliceosomal intron complexity expanded in eukaryotic mRNA, given that prokaryotes barely contain such RNA fragments [1]. There are three different hypothesis which are widely accepted, though two of them are non-exclusive and they were integrated later [2, 3].

In one hypothesis called "Introns-early", the spliceosomal intron in eucaryon evolved from primordial introns in the prokaryotic ancestors. In the "introns-late" hypothesis, which is not exclusive to the hypothesis above, the spliceosomal intron appears continuously along the evolutionary path. However, some people also believe that the spliceosomal intron appeared earlier than the exon. It asserts that mRNA evolves from non-functional RNA in primitive prokaryotes, and

the spliceosomal intron is a kind of "relic" left in the mRNA in descendant eukaryotes [4]. Details of these hypotheses will be introduced below. In recent years, some old hypotheses have been adjusted as new evidence has been found, and new mysteries and controversies have appeared. Though there is still no definite conclusion about the origin of the spliceosomal intron, the summary is heuristic for future research. The three representative evolutionary hypotheses are systematically reviewed in the following sections, including their core standpoints, supporting evidence and unresolved defects.

2. Introns-early hypothesis

Introns-early, as its name suggests, holds that introns appeared in mRNA before the emergence of the DNA world. This hypothesis proposes that mRNA with introns could have appeared in primitive prokaryotes and functioned similarly to that in eukaryotic cells. Some evidences support this hypothesis. As is recognized by the mainstream academic community, intron widely exists in nearly all kinds of eucaryon cells. In comparisons of introns across different species, the positions of many introns were found to be similar, with a percentage over 10 times higher than expected by random distribution [5]. Based on this, introns-early hypothesis earn some supporters in early years. However, the first query to this hypothesis is about protokaryon: Nowadays people can hardly find intron in mRNA in the cells of protokaryon. If introns in mRNA originated from primitive prokaryotes, where are those prokaryotes that possess such introns?

An explanation was mentioned by the supporters of introns-early. As was discovered in modern times, the structure of spliceosome, in its simplest form, is a complex of several different kinds of RNAs and over 10 kinds of proteins [6]. Though it was probable that spliceosome in the cells of the origin protokaryon was not so complicated, several different kinds of proteins and RNAs are still necessary [6]. Thus, extra energy expenditure for intron was necessary for these protokaryon's cells to play a sufficient role. For prokaryotes, extra energy used for introns is costly, especially compared with eukaryotic cells, which obtain much more energy from their mitochondria. With the natural selection, only prokaryotes without spliceosomal introns could fit into the modern environment.

However, there are some evidences showing that this explanation has deficiencies. Recent analyses of gene structural evolution (Figure 1) provide counterevidence: introns are not merely lost over evolutionary timescales. From Figure 1, one can observe directly how intron densities changed along the evolutionary routes of different species. Though intron density of some species dropped during their evolution routes shown in Figure 1, the intron density of some other species kept growing. It is clear that only some introns can possibly be inherited from the last eukaryotic common ancestor (LECA). Generally speaking, the introns-early hypothesis does not fit new research results well.

Because the explanation mentioned above is somewhat far-fetched and incomplete, early supporters of introns-early tried to change their minds. Some tried to merge their hypothesis with Introns-late (explained later), while others tried to put forward a new hypothesis—introns-first—to explain the origin of mRNA and spliceosomal introns.

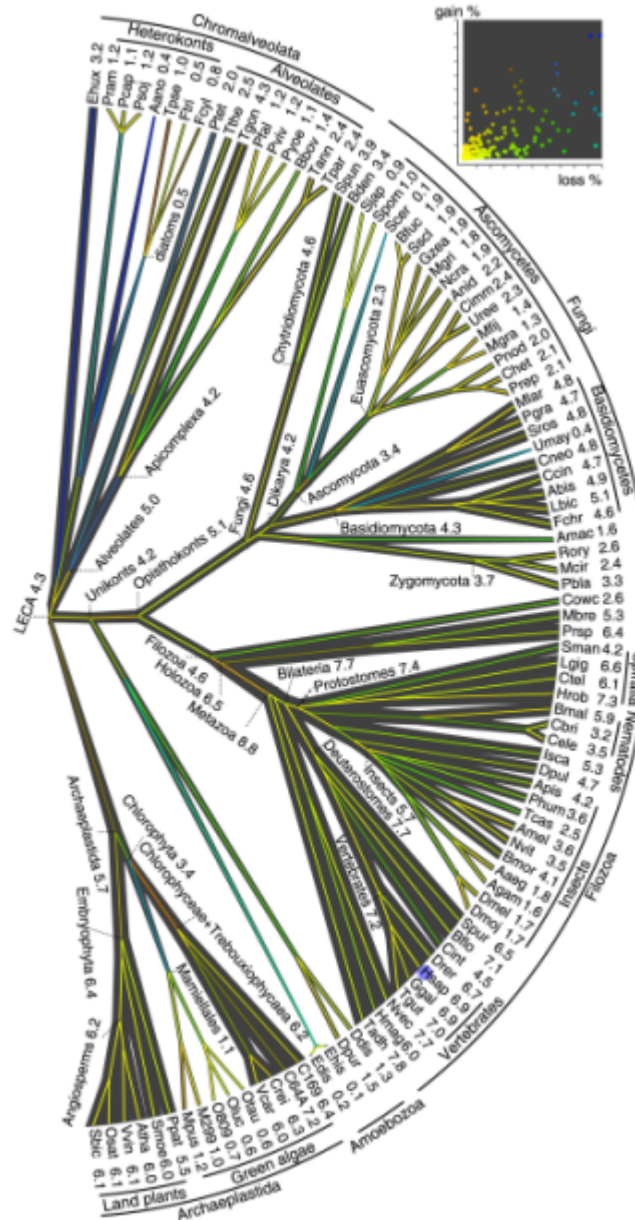


Figure 1. Reconstruction of intron gains and losses in the evolution of eukaryotes and intron density in ancestral eukaryote forms [7]

3. Intron-late hypothesis

Intron-late means that the appearance of intron in cells can be later than the appearance of DNA world on the evolutionary tree. Supporters of this hypothesis believe that introns emerge continuously within the evolutionary process of eukaryotic cells. Here are some evidence supporting this hypothesis: according to some studies, evidence of intron position diversity shows the high possibility that most introns have arrived at their current locations only recently [8]. At this stage of research, however, there are still some problems that can hardly be solved in this hypothesis. For example, the question of spliceosomal intron's origin has still not been affirmed. Also, some problems still need to be solved for intron-late hypothesis. Figure 1 also provides evidence that the loss of introns during the routes of evolution has advantages in a larger group of species [7], which

can not be explained well. This result indicates a new possibility: both intron-early and intron-late happened in the evolution routes of most species. These two hypotheses have been merged together as one new hypothesis later, which was called 'many introns early in eukaryotic evolution' [3]. Up to now, the appearance stage of spliceosomal intron has reached an initial consensus. However, the specific way of spliceosomal intron's formation is still not clear. More hypotheses would be proposed and verified.

3.1. Intron sliding

Intron sliding is a phenomenon in which the splicing site between intron and exon moves for a short distance (less than 60 bases). Though the facticity of intron sliding remains controversial for a long time, research confirmed it by statistical analysis [9]. By moving the splicing site on mRNA, the length of intron can be changed. Some researchers connected this phenomenon with gain and loss of intron: perhaps the appearance of intron was due to the translation of splicing site. Also, further research on closely located introns among 12 *Drosophila* genomes confirmed that intron sliding is a phenomenon during the formation process of some kinds of introns [10]. However, how introns take their shape is still disputable under this hypothesis: intron sliding may not be the only way for introns to be formed.

3.2. Proto-splice sites

Drawing on some hypotheses and evidence from some studies, some researchers believe that parallel acquisition can also be a source of introns. One prevailing theory states that conserved flanking coding sequences on both sides of introns represent residual recognition motifs termed proto-splice sites within mRNA transcripts [11, 12]. Studies that support this hypothesis show their evidence. Cryptic splicing signals are a kind of sequence which are similar to a spectral signal but do not participate in the splicing of mRNA. They look just like 'disabled spectral signal', and according to some studies they are always at the same positions, compared with any spectral signal in the direct homologous gene of another species (Figure 2). The Figure 2 shows the similarity of position between intron spectral signal in some species and proto-splice sites in others. It is worth noting that this evidence can also fit the intron-late hypothesis. However, results of some other studies are contrary to that through the finding of exceptions: recently gained introns in animal genes of alpha-amylase [13]. These researches with disharmonious results may indicate that the origin of spliceosomal intron may be complex and multi-pathway.

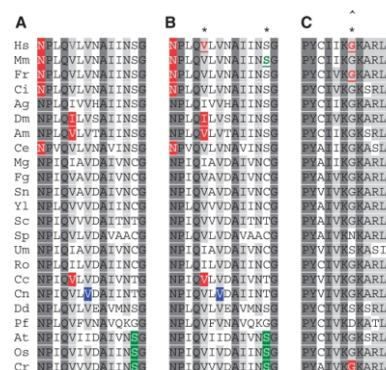


Figure 2. Comparison of intron and css positions for a small part of the ribosomal protein gene database [12]

4. The compromise evolutionary model

Because the different evidence supporting their own hypotheses could not be explained by any single way, some researchers tried to compromise them. Through their work, a complete and credible history of spliceosomal introns was built. At first, the initial spliceosomal introns were already on the RNA chain of the LECA. The precursors of introns may be inherited from the primordial genetic pool and evolve with the initial cells together through a complex way (figure 3). Then after the appearance of LECA, different parts of spliceosomal introns were inherited while others were lost in different species, due to natural selection (figure 1). Also, new spliceosomal introns entered the mRNA of eukaryotes, perhaps through intron sliding, and also may have come from the mitochondrial endosymbiont (figure 4).

Based on this compromise hypothesis, the evidences from the supporters of both introns-early and introns-late can be explained well. Whether conservation of intron position or the finding of newly acquired introns, current evidence can only show part of the situation of introns. Thus compromise them together may be the comprehensive model, which fits the truth better. However, this compromise hypothesis still might not be accepted by everyone. Some researchers choose to support another hypothesis, introns-first, which was mentioned above.

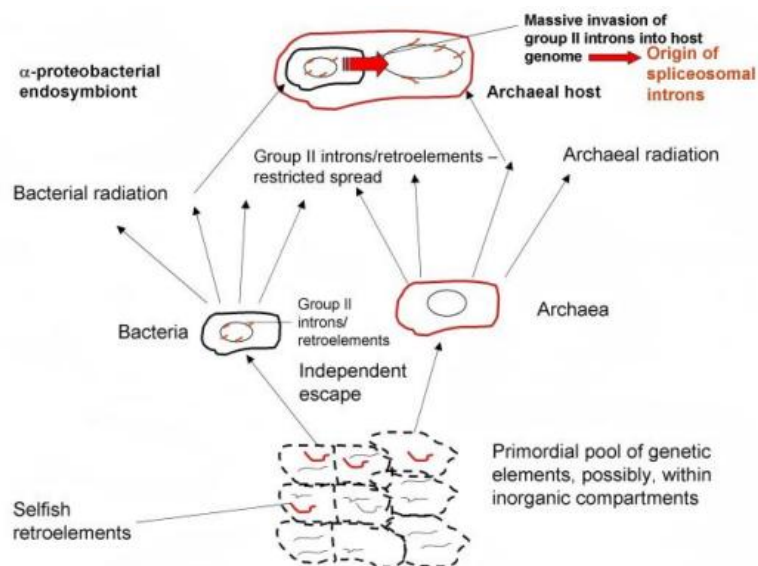


Figure 3. A brief early history of spliceosomal introns [3]

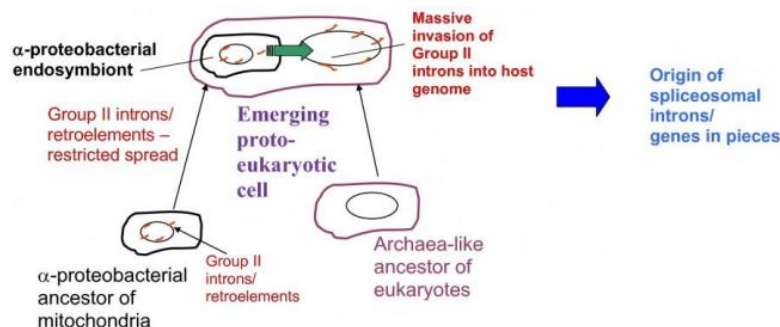


Figure 4. A hypothetical scenario of early history of spliceosomal introns [14]

5. Introns-first hypothesis

Introns-first hypothesis denies the origin model of intron mentioned before. They put forward a new idea that introns appeared even earlier than exons—in other words, exons emerged from non-coding regions of RNA genes in prokaryotes, which they consider as predecessors of mRNA [14]. In their model, introns emerged in the RNA world, which existed earlier than protein-coding synthesis. During the RNA world, the introns, which they consider as predecessors of mRNA, may have been used as functional RNA. These RNA fragments evolved during species evolution, and new functions developed with the emergence of exons, which finally became mRNA.

For the introns-first hypothesis, RNA-Continuity was necessary. They should try to prove that at least part of intron was directly from the RNA world. Though their function may have changed a lot, they need to have a function at nearly all steps of intron evolution. However, although the model has been built, a continuous series of functional intermediates has not been successfully established. In general, introns-first is not the best choice at present.

6. Conclusion

This article mainly introduces the hypotheses of spliceosomal introns' origin. Though there are still some disputes, the compromise solution is accepted by more people. By reviewing the historical development of theories, researchers have already found a relatively well-developed hypothesis and model to explain the origin of spliceosomal introns. The compromise model can explain the results of most studies. In comparison, the introns-first hypothesis is more far-fetched, with less evidence and more problems, and is not accepted by the mainstream academic community. However, all theories are still at the hypothesis stage: due to the lack of direct evidence, even the compromise solution of intron-early and intron-late is not the final answer.

Through this article, the process of theoretical development was organized, and the future research direction became clearer. However, there are still so many questions waiting for researchers. As shown in Figure 1, although intron density can decrease in eukaryote evolution, almost all the eukaryotes did not lose their 'last spliceosomal introns'. Compared with prokaryotes, in which spliceosomal introns are hardly found, this phenomenon still awaits a more reasonable explanation. In addition, the specific function of introns in eukaryotes is waiting for clarification. Many mysteries remain unsolved in evolutionary science, and many hypotheses await verification. As one of the least understood topics in biology, most of the conclusions obtained about spliceosomal introns remain speculative and uncertain. Generally speaking, there would be a long way for further study.

References

- [1] Mikina, W., Hałakuc, P., & Milanowski, R. (2024). Transposon-derived introns as an element shaping the structure of eukaryotic genomes. *Mobile DNA*, 15(1), 15. <https://doi.org/10.1186/s13100-024-00325-w>
- [2] Collins, L., & Penny, D. (2005). Complex spliceosomal organization ancestral to extant eukaryotes. *Molecular Biology and Evolution*, 22(4), 1053–1066. <https://doi.org/10.1093/molbev/msi091>
- [3] Koonin, E. V. (2006). The origin of introns and their role in eukaryogenesis: A compromise solution to the introns-early versus introns-late debate? *Biology Direct*, 1, 22. <https://doi.org/10.1186/1745-6150-1-22>
- [4] Penny, D., Hoepfner, M. P., Poole, A. M., & Jeffares, D. C. (2009). An overview of the introns-first theory. *Journal of Molecular Evolution*, 69(5), 527–540. <https://doi.org/10.1007/s00239-009-9279-5>
- [5] Fedorov, A., Merican, A. F., & Gilbert, W. (2002). Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proceedings of the National Academy of Sciences of the United States of America*, 99(25), 16128–16133. <https://doi.org/10.1073/pnas.242624899>

- [6] Lamolle, G., & Musto, H. (2018). Why prokaryotes genomes lack genes with introns processed by spliceosomes? *Journal of Molecular Evolution*, 86(9), 611–612. <https://doi.org/10.1007/s00239-018-9874-4>
- [7] Csuros, M., Rogozin, I. B., & Koonin, E. V. (2011). A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS Computational Biology*, 7(9), e1002150. <https://doi.org/10.1371/journal.pcbi.1002150>
- [8] Stoltzfus, A., Logsdon, J. M., Jr, Palmer, J. D., & Doolittle, W. F. (1997). Intron "sliding" and the diversity of intron positions. *Proceedings of the National Academy of Sciences of the United States of America*, 94(20), 10739–10744. <https://doi.org/10.1073/pnas.94.20.10739>
- [9] Cohen, N. E., Shen, R., & Carmel, L. (2012). The role of reverse transcriptase in intron gain and loss mechanisms. *Molecular Biology and Evolution*, 29(1), 179–186. <https://doi.org/10.1093/molbev/msr192>
- [10] Lehmann, J., Eisenhardt, C., Stadler, P. F., & Krauss, V. (2010). Some novel intron positions in conserved Drosophila genes are caused by intron sliding or tandem duplication. *BMC Evolutionary Biology*, 10, 156. <https://doi.org/10.1186/1471-2148-10-156>
- [11] Dibb, N. J., & Newman, A. J. (1989). Evidence that introns arose at proto-splice sites. *The EMBO Journal*, 8(7), 2015–2021. <https://doi.org/10.1002/j.1460-2075.1989.tb03609.x>
- [12] Kapustin, Y., Chan, E., Sarkar, R., Wong, F., Vorechovsky, I., Winston, R. M., Tatusova, T., & Dibb, N. J. (2011). Cryptic splice sites and split genes. *Nucleic Acids Research*, 39(14), 5837–5844. <https://doi.org/10.1093/nar/gkr203>
- [13] Da Lage, J. L., Maczkowiak, F., & Cariou, M. L. (2011). Phylogenetic distribution of intron positions in alpha-amylase genes of bilateria suggests numerous gains and losses. *PLoS One*, 6(5), e19673. <https://doi.org/10.1371/journal.pone.0019673>
- [14] Rogozin, I. B., Carmel, L., Csuros, M., & Koonin, E. V. (2012). Origin and evolution of spliceosomal introns. *Biology Direct*, 7, 11. <https://doi.org/10.1186/1745-6150-7-11>