INTRON EVOLUTION: SLIDING AND VARIABILITY OF LENGTH

I.V. Poverennaya[1]*, D.D. Gorev[2], T.V. Astakhova[3], M.A. Roytberg[2, 3]
[1] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*
[2] *Moscow Institute of Physics and Technology, Moscow, Russia*
[3] *Institute of Mathematical Problems of Biology RAS, Pushchino, Russia*
* *Corresponding author: ipoverennaya@gmail.com*

Key words: *introns, intron evolution, intron phase, sliding, exon-intron structure*

Motivation and Aim: Intron sequences evolve so fast that even their lengths seem to be highly unconserved. Thus, intron evolution is usually considered in terms of evolution of exon-intron structures (EIS). One of the rarest and intriguing evolutionary event is intron sliding that is a shifting of exon-intron boundaries over short distances. Such relocation could lead to the change of intron phase, i.e. the position of intron relative to the open reading frame that might affect the "golden ratio" of intron phase distribution. Here we analyze the exon-intron structure of four different eukaryotic genes in order to find out the preferable choice of intron phase during intron sliding and to study the correlation between lengths of orthologous introns.
Methods and Algorithms: Identification of orthologous introns requires exon-intron structure alignment of orthologous genes. To construct them we have applied our own protocol of manual EIS aligning based on multiple sequence alignment of genes products and exon lengths. The obtained alignments have been used as training and testing sets for multiple EIS alignment program developed by us.
Results: For each analyzed gene about 100 orthologues from different vertebrates (mammals, amphibians, fishes, birds, etc.) was obtained through the Annotation Pipeline NCBI database (http://www.ncbi.nlm.nih.gov/gene/) to make a dataset. Analysis of EIS alignments has revealed several cases of sliding for each dataset. In every case, the sliding caused an intron phase change; however, there seem to be no preference of novel or initial phase. Analysis of the intron lengths showed that despite high variability of intron length, some correlations could be observed especially in separate taxa. Moreover, if instead of intron length L we will consider a normalized length N = (L-A)/A, where A is an average length within a group of orthologous introns. E.g. for *ptprd* genes of birds (28 species were considered) the normalized value is in the interval (-0.15, 0.15) for 85.2% of introns what is significantly higher than the values for random set of lengths in accordance with the distribution of the lengths of the introns. Also, for the interval (-0.5; 0.5) the according proportion of introns is 96.8%.
Conclusion: Obtained results did not confirm our initial hypothesis that in the process of sliding introns prefer to change its phase to 0 more frequently. However, it is necessary to expand the analysis on a larger dataset of genes for making a final conclusion. Despite the wide range of orthologous intron lengths, some intron length conservation could still be observed and leads us to the question what intron length the ancient introns had.