

INTRON LENGTHS AND PHASES: REGULARITIES AND DATABASE

Astakhova T.V.¹, Tsitovich I.I.^{2,3}, Yakovlev V.V.^{1,3}, Roytberg M.A.^{1,3,4*}

¹*Institute of Mathematical Problems of Biology, RAS, Moscow Region, Russia;*

²*Institute of Information Transmission Problems, RAS, Moscow, Russia;*

³*Higher School of Economics, Moscow, Russia;*

⁴*Moscow Institute of Physics and Technology, Moscow Region, Russia*

e-mail: mroytberg@lpm.org.ru

** Corresponding author*

Key words: exon, intron, intron phase, intron length

Motivation and Aim: Understanding of regularities of exon-intron structure is important for study of mechanisms and evolution of exon-intron structure. Relationships between phase and length of introns are not sufficiently studied. We extend our previous results [1] to genomes of Plantae.

Methods and Algorithms: Phase of intron is a remainder of total length of preceding exons divided by three. In [1] we have considered 17 animal genomes of various classes (insects, fishes, amphibians, reptiles, birds, mammals) and demonstrated for all the genomes the following effect. It is known that introns are unevenly distributed over the phases, the phase ratio, i.e. proportion between number of introns in the phases 0, 1 and 2, is about 5:3:2. We have shown that the ratio changes, if we consider only T-long introns, i.e. introns, with length greater than a threshold T. With increasing of the threshold T part P1(T) of introns in phase 1 increases, while the part P0(T) of introns in phase 0 decreases and at a certain value we obtain P1(T)=P0(T). In this work we extend the study to 4 plant genomes, *A.thaliana*, *O.sativa*, *V.vinifera* and *P. trichocarpa*.

Results: For all considered plant genomes the baseline ratio is about 6:2:2, i.e. fraction of phase 0 introns in plant genomes is greater than it in animal genomes. For *V. vinifera* and *P. trichocarpa* the ratio P0(T):P1(T):P2(T) unlike all other considered genomes does not depend on T. However for *O. sativa* and *A. thaliana* the dependence exists. E.g. for *O. sativa* with T=1100 the phase ratio is 53:23:24 while for all introns we have ratio 58:20:22. Z-score ZP0(1100) of the decrement of P0(1100) compared to P0(0) is -5.4. Number of introns having length at least 1100 is 2915 that is 5.6 % of all introns. For *A. thaliana* ZP0(T) has two local extrema. For T=1200 we have ZP0(1200) = -5.6, number of long introns 132 (0.2 % of all introns) and phase ratio 34:27:29. For T=280 we have ZP0(280) = -4.0, number of long introns 6015 (9.2% of all introns) and phase ratio 55:21:23. Some other regularities related to intron lengths and phases were also revealed. To provide the study we have created the database available at http://server2.lpm.org.ru/static/introns_db/; the scheme of the database can be found at the same site.

Conclusion. The dependence between lengths and phases in plant genomes significantly differs from those for animals. That may reflect difference in evolution of the taxa.

References:

1. T.V. Astakhova, I.I. Tsitovich, V.V. Yacovlev, M.A. Roytberg. Dependence between lengths and phases of introns.// *MCCMB 2013: Proceedings*. 2013.