Mirtrons: computational feature analysis and miRNA comparison

Tamires Priscila da Costa[1], Douglas Silva Domingues[2,3], Alexandre Rossi Paschoal[1,2]

[1] *Departamento de Computação – DACOM, Universidade Tecnológica Federal do Paraná, Cornélio Procópio,* [2] *Programa de Pós-Graduação em Bioinformática - PPGBIOINFO, Universidade Tecnológica Federal do Paraná, Cornélio Procópio and* [3] *Departamento de Botânica, Instituto de Biociências, Universidade Estadual Paulista, Rio Claro*

The advances in genome sequencing technologies, as well as in bioinformatics approaches allow numerous scientific analyses using publicly available data. Among these data are the microRNAs (miRNAs). They correspond to one class of non-coding RNAs that acts as a post-transcriptional regulator of mRNA level in cells. Recently, a novel microRNA class, named mirtrons, was identified in several model organisms and they have different biogenesis when compared to canonical miRNAs. Mirtrons are processed by splicing step instead of the cleavage by a specific enzyme. There are more than 140 databases specific for noncoding research, mostly devoted to miRNA data. However, the biological computational view of mirtrons is not sufficiently addressed. In this study, we did a computational features analysis on mirtron public data and compared against canonical miRNAs. Our interest is to identify mirtron-specific properties that could be applied in future *in silico* applications. Mirtron data was extracted from literature, using public curated data of *A. thaliana*, rice, human, mouse, fly and worm genomes. We analyzed five features based on sequence and structure aspects: GC content, length, sequence nucleotides distribution (up to K-mer <4), free energy and conservation. These analyses were performed using in-house PERL scripts, R studio and BLAST program for alignment. We have found differences in the free energy distribution, mainly in plant mirtrons. Specifically for conservation analysis we focused on plant and fly genomes. Our results suggest conservation in flies genomes while we did not found any significant results in plants. Overall, this work provides a feature overview on mirtrons literature data. We believe that this report will contribute to future research in computational approaches to understand mirtron distribution and characteristics.