

Bioinformatics investigation of non-coding RNAs and transposable elements in plants

Daniel Longhi Fernandes Pedro¹, Nicolas Gil de Souza Aoki¹, Alan Péricles Rodrigues Lorenzetti², Douglas Silva Domingues¹, Alexandre R. Paschoal¹

1 UTFPR - PPGBIOINFO

2 USP

Abstract

Non-coding RNAs (ncRNAs) are transcripts that do not encode proteins. There are several classes of ncRNAs, which the most studied are microRNAs (miRNAs). Transposable Elements (TEs) are the major genomic component in eukaryotic genomes. They can comprise more than 45% of human and animal genomes, and in plants, they comprise up to 90% of the genome. Our research group recently developed the PlanTE-MIR DB, the first public database that studies the relationship between miRNA and TEs in plants. In this repository, users can search, extract and analyze these overlapping features in 10 plant species. Now, we intend to evaluate TEs relationship with all ncRNA classes, generating a new version of PlanTE-MIR DB. New bioinformatics analyses will use public genomic data available at Ensembl Plants portal and results will be accessible on a user-friendly website. Three steps cover the workflow of this investigation: a) Curate and intersect ncRNAs and repetitive DNA features from existing Ensembl annotation; b) Perform de novo TE prediction in plant genomes and intersect ncRNA annotation in order to find new potential overlaps; and c) Compare newly discovered TEs against public ncRNA databases. From 44 genomes available at Ensembl Plants, 25 species have ncRNA annotation. In 24 we found overlap with TEs. The species with most overlapped regions was *Zea mays* with 3,105 hits and the species with less hits was *Sorghum bicolor*, with one hit. Finally, we intend to develop a new method to identify plant TEs using deep learning techniques. These computational analyses will provide to the scientific community a friendly way to work with this knowledge.

Funding: UTFPR