2nd Bioinformatics Meeting
October 13 and 14, 2016, Universidad de Talca

SCIFORUM
BMEICB-02

A bioinformatic approach to search for active transposases in genomes.

Braulio Valdebenito and Gonzalo Riadi *

Centro de Bioinformática y Simulación Molecular, Facultad de Ingeniería, Universidad de Talca, 2 Norte 685, Casilla 721, Talca, Chile.

* Author to whom correspondence should be addressed; E-Mail: griadi@utalca.cl; Tel.: +5671 2201685; Fax: +5671 2201685.

Abstract: Eukaryotic transposons are DNA sequences able to move inside a genome. They are characterized by a sequence that encodes a transposase protein of ~300 aminoacids and flanking it, short terminal inverted repeats of ~30bp. Active DNA transposons are very difficult to predict computationally because: 1. Due to their activity, there are many copies, or paralogous, of the transposons of a family in a genome; 2. Due to mutation, a high diversity of sequences has resulted, and as consequence; 3. Many transposons are incomplete or mutated enough to render the element inactive. In order to circumvent these issues, we generated Hidden Markov Models (HMMs) for 12 families of eukaryotic transposases because HMMs are an appropriate technique for searching evolutionary divergent sequences.

In animals, during their development, transposons activity is regulated by piRNAs. This regulation occurs via Watson-Crick base pairing between the piRNA and the transposase transcript. In order to test the ability of our models to predict active transposases, we used as reference the mapping of known piRNAs sequences of an organism on its own genome, and compared it to our transposase predictions, and to those made by RepeatMasker, the current gold standard software for prediction of mobile elements. We found that, while RepeatMasker has a higher absolute number of predictions, its sensitivity and selectivity as classifier of active transposases is lower than our HMMs for all tested organisms. Although, there is a lot of room for improvement, these results are a step towards the improvement of the accuracy of prediction of active transposases.

Conflicts of Interest

The authors declare no conflict of interest.