miRNAfe: a tool for feature extraction in pre-miRNA prediction

Cristian Yones[1], Georgina Stegmayer[1], Laura Kamenetzky[2], Diego Milone[1]
[1]Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH-UNL, CONICET, Ciudad Universitaria UNL, (3000) Santa Fe, Argentina.
[2]Instituto de Investigaciones en Microbiología y Parasitología Médica (UBA), CONICET, Paraguay 2155, piso 13 (1121), Buenos Aires, Argentina.

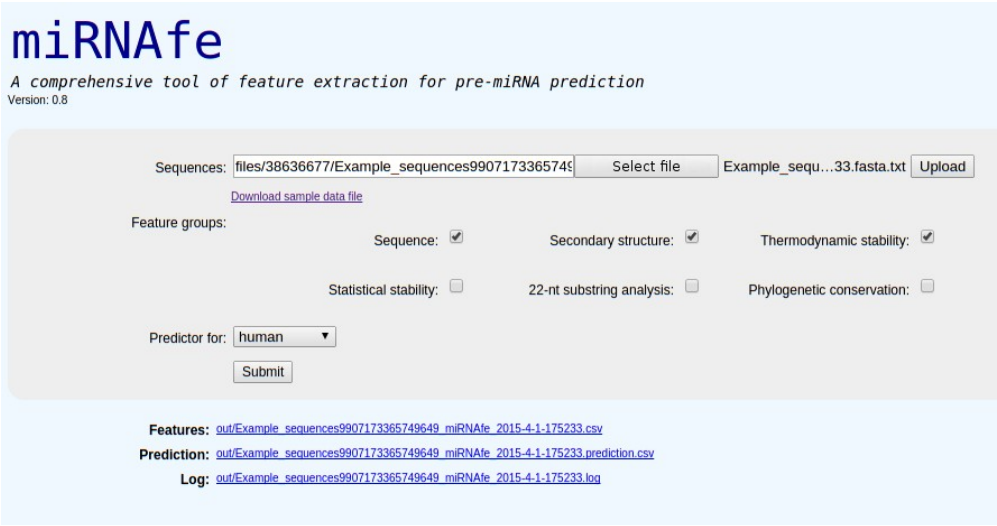
Background

MicroRNAs (miRNA) are a group of short (~22 nucleotides) non-coding RNA which can play important roles in gene regulation by targeting mRNAs for cleavage or translational repression [1]. Precursors of miRNA (pre-miRNA) are characterized by their hairpin structure. However, a large amount of similar sequences can be folded into this kind of structure in many genomes.
In order to predict miRNAs, the first step is to extract features from sequences of a complete genome to train a binary classifier. After that, the classifier is used to predict which sequences from a genome are more likely to be a pre-miRNA. The feature extraction step is very important for this whole process, in order to achieve high rates of true positives predictions. The amount of features that have been proposed in the last years is very large and there are many different tools that partially achieve this task. They are coded in different programming languages and have different access modes (web, command line, etc.). Besides, several tools are proprietary software and the source code is not even available.

Results

We have developed the miRNAfe tool that implements all existing state-of-the-art feature extraction processes used for miRNA prediction nowadays. The tool implements up to 80 features proposed in literature over the past 10 years [2]. The features are divided into six pre-defined groups according to the kind of information that must be extracted from the sequence: sequence, secondary structure, thermodynamic features, statistical stability, 22-nt substrings analysis and phylogenetic conservation.
MiRNAfe is composed by a set of Matlab functions and can be installed in the user machine as any other toolbox. Some remarkable capabilities of the toolbox are batch analysis, parallel processing, and report generation. In order to validate all feature extraction scripts, several sequences were analyzed with miRNAfe and with software of the original authors, and the outputs have been compared. In all tests performed, the results were always consistent to those of the original papers.
We have developed a simplified and easy to use web interface that can be accessed at http://fich.unl.edu.ar/sinc/web-demo/mirnafe/. Also, besides the main function of the tool is not prediction, a support vector-machine (SVM) is provided since it is the most frequently used method in pre-miRNA prediction techniques [2]. The Figure 1 shows the web interface after a feature extraction process and prediction on the example file provided with human sequences.

Figure 1



Web interface of miRNAfe: the user must load a fasta file with the sequences to be analyzed, select which group of features wants to extract and, if a prediction wants to be made, choose a pre-trained classifier.

Reference

1. M. Rosenzvit, M. Cucher, L. Kamenetzky, N. Macchiaroli, L. Prada, and F. Camicia: **MicroRNAs in Endoparasites**. *Nova Science Publishers 2013*.

2. Kleftogiannis, D., Korfiati, A., Theofilatos, K., Likothanassis, S., Tsakalidis, A., & Mavroudi, S.: **Where we stand, where we are moving: Surveying computational techniques for identifying miRNA genes and uncovering their regulatory role**. *Journal of biomedical informatics 2013*, ***46*(3)**:563-573.