pyNotation Documentation

Delestre Clément

August 28, 2014

Contents

1 About pyNotation

pyNotation is a software dealing with GenBank files, in order to analyse a genome and/or compare 2 genomes (or genes). It can :

- Analyze a genome file and compute several biological parameters like genome size, % of GC, number of genes etc.
- Looking for a gene in other genome file. If similar gene are found, pyNotation can draw the genetical neighborhoud.

1.1 Dependencies

pyNotation use BioPython, so this library should be properly installed on your computer if you want to use pyNotation or modify its source code.

pyNotation also uses SignalP & HMMER. The software can still work without these tools but it's strongly recommended to install them before using pyNotation.

2 How it works

Here we describe how to use pyNotation. Please note this software doesn't have GUI[1] so we assume you are able to deal with command lines. To display a simple & quick help use -h or - -help option.

2.1 Analyze a genome

Specify the file you want to analyze with "- -analyze" or "-a" option. If you want to analyze several files, write several file names separated by a space.

By default, pyNotation will display results to the STDOUT but it's possible to save results in a CSV file. In order to do this, you have to specify a prefix with the "-o" option. Two types of analysis are done by pyNotation, so if you use "-o" option, two files will be created their name will follow this format :

- prefix_DNAseq.csv : containing information about the DNA sequence (% of GC for example)
- prefix_genesFeatures.csv : containing information about the genes (number of RNA genes for example)

[1] **G**raphical **U**ser **I**nterface

2.2 Looking for a gene

`pyNotation` can find several genes in a file and then compare these genes with other files. To specify which gene you want select, use the "-s" option and write words that should be contain in gene name. For example if you use "-s a" `pyNotation` will select all genes containing the "A" letter (case insensitive).

After finding one or several genes, `pyNotation` will look for similar genes in other files. So when you use the "-s" option you have to specify at least 2 files separated by a space. For example the command : "pyNotatationAnalyzer.py -s abc File1 File2 File3" will search genes containing "abc" in File1 and compare them with genes present in File2 and File3.

If a gene is found, `pyNotation` will display results on the STDOUT and calculate the similarity between the found gene and the gene of interest using `Needle`. If `Needle` is not installed, the similarity will be 0%. Similar genes can be saved in a file using the "-o" option.

Visualization

If a similar gene is found, the genetical neibourhood can be display using the "- -pdf" option and specify a prefix. The file will be at `PDF` format and its name will be prefix_NameOfGene.pdf. You can also specify how many genes should be draw upstream and downstream the interest gene with the "- -interval" option.

2.3 Metadata

By default, `pyNotation` will display `metadata` on the STDOUT (date of starting an analysis for example). If don't you want to display them, use "- -quiet" or "-q" option.