Biopipeline

Biopipeline's purpose is to find miRNA candidate sequences within a *fasta* file containing a genome (vegetal or animal).

Requisites
Biopipeline must be run in Linux, preferably Ubuntu 12.04 or later. The script was written in Perl.
The input is a *fasta* file containing a genome.
The output is a *fasta* file containing the miRNA cadidate sequences.

Dependencies
The following software must be previously installed:

- **Blast**
  $sudo apt-get install blast2

- **Vienna** (http://www.tbi.univie.ac.at/~ronny/RNA/index.html)
  Download from: http://www.tbi.univie.ac.at/RNA/index.html#download
  Or install via command line:
  $sudo apt-add-repository ppa:j-4/vienna-rna
  $sudo apt-get update
  $sudo apt-get install vienna-rna

- **EMBOSS**
  Open Package Manager, search for "emboss" and install the first occurency, or install via command line:
  $sudo apt-get install emboss

- **Bioperl**
  Install via Package Manager (search for "bioperl") or via command line:
  $sudo apt-get install bioperl

- **ForkManager**
  Install via Package Manager (search for "forkmanager") or via command line:
  $sudo apt-get install libparallel-forkmanager-perl

- **Only in Ubuntu 14.04 or later**, install the following, via command line:
  $sudo apt-get install libswitch-perl
  $sudo cpan App::cpanminus
  $sudo cpan Switch

- **Install R** via command line:
  $sudo apt-get update
  $sudo apt-get install r-base
  $sudo apt-get install r-base-dev

- **Random Forest**
  Open Package Manager, search for "randomforest" and install the first occurency (NOT i386 for 64 bits version), or install via command line:
  $ sudo apt-get update
  $ sudo apt-get install r-cran-randomforest

Installation Guide

Unpack Biopipeline archive into the desired directory. The following fiiles must be in the directory after installation:

- genome.fasta
- hairpin.fasta
- mainParallelVegetal.pl
- mainParallelAnimal.pl
- mature.fa
- matureVegetalAcron
- microRNAcheck_parallel.pl
- model.RData
- Rfam.fasta

User Guide

Before runninng Biopipeline for the first time, satisfy all dependencies in the Dependencies section above, and unpack Biopipeline as described in the Installation Guide.

If necessary, before running Biopipeline, convert the genome file to the *fasta* format:

```
$ fastq_to_fasta -v -n -Q 33 -i <inputFile.fq> -o <outputFile.fasta>
$ fastx_collapser -v -i <inputFile.fq> -o <outputFile.fasta>
$ ? fastq_quality_filter
```

Replace the file genome.fast in the Biopipeline directory with your genome file, renaming it to genome.fasta.

Type the following command line to run Biopipeline (choose the appropriate file to run according to the organism):

```
$ perl mainParallelVegetal.pl
```

or

```
$ perl mainParallelAnimal.pl
```

Technical and maintaing documentation

Please refer to the diagrams.pdf file.

Biopipeline article

de Souza Gomes, Matheus, et al. "Genome-wide identification of novel microRNAs and their target genes in the human parasite Schistosoma mansoni." *Genomics* 98.2 (2011): 96-111.

http://dx.doi.org/10.1016/j.ygeno.2011.05.007