

Biopipeline

O propósito do Biopipeline é encontrar sequências candidatas a miRNA em um arquivo *fasta* contendo o genoma de um organismo, que pode ser planta ou animal.

Requisitos

O Biopipeline deve rodar em Linux, preferencialmente Ubuntu 12.04 ou superior. A linguagem de programação utilizada é o Perl.

A entrada é sempre um arquivo *fasta* contendo um genoma.

A saída é um arquivo *fasta* contendo sequências candidatas a miRNA.

Dependências

Os seguintes *softwares* devem estar instalados previamente:

- **Blast**

`$sudo apt-get install blast2`

- **Vienna** (<http://www.tbi.univie.ac.at/~ronny/RNA/index.html>)

Baixe o instalador em: <http://www.tbi.univie.ac.at/RNA/index.html#download>

Também é possível instalar via linha de comando:

`$sudo apt-add-repository ppa:j-4/vienna-rna`

`$sudo apt-get update`

`$sudo apt-get install vienna-rna`

- **EMBOSS**

Abra o Gerenciador de Pacotes, procure por “emboss” e instale a primeira ocorrência, ou instale via linha de comando:

`$sudo apt-get install emboss`

- **Bioperl**

Instale pelo Gerenciador de Pacotes (procure por “bioperl”) ou via linha de comando:

`$sudo apt-get install bioperl`

- **ForkManager**

Instale pelo Gerenciador de Pacotes (procure por “forkmanager”) ou via linha de comando:

`$sudo apt-get install libparallel-forkmanager-perl`

- **Apenas no Ubuntu 14.04 ou superior**, instale também, usando linha de comando:

`$sudo apt-get install libswitch-perl`

`$sudo cpan App::cpanminus`

`$sudo cpan Switch`

- **R**

`$sudo apt-get update`

`$sudo apt-get install r-base`

`$sudo apt-get install r-base-dev`

- **Random Forest**

Abra o Gerenciador de Pacotes, procure por “randomforest” e instale a primeira ocorrência (não selecione a opção i386 para 64 bits), ou use a linha de comando:

`$ sudo apt-get update`

`$ sudo apt-get install r-cran-randomforest`

Guia de instalação

Descompacte o pacote do Biopipeline no diretório escolhido. Esse diretório deve conter os seguintes arquivos após a instalação:

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

Guia de utilização

Antes da primeira utilização, satisfaça todas as dependências constantes da seção Dependências acima, e descompacte o pacote do Biopipeline conforme a seção Guia de Instalação.

Se necessário, antes de executar o Biopipeline, converta o arquivo contendo o genoma para o formato fasta:

```
$ 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
```

Substitua, no diretório de instalação do Biopipeline, o arquivo genome.fasta pelo seu arquivo contendo o genoma a ser analisado, mantendo o nome do arquivo como genome.fasta.

Utilize a seguinte linha de comando para executar o Biopipeline (escolha o arquivo apropriado de acordo com o organismo):

```
$ perl mainParallelVegetal.pl
ou
$ perl mainParallelAnimal.pl
```

Documentação técnica e de manutenção

Por favor, consulte o arquivo diagrams.pdf.

Artigo do Biopipeline

DE SOUZA GOMES, Matheus et al. Genome-wide identification of novel microRNAs and their target genes in the human parasite *Schistosoma mansoni*. Genomics, v. 98, n. 2, p. 96-111, 2011.

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