

TrieFinder

Gabriel Renaud

Max Planck Institute for Evolutionary Anthropology

gabriel.reno [at sign] gmail.com

August 23, 2012

Contents

1	Description	1
2	Installation	1
2.1	Building the binaries	1
2.2	Setting up the databases	2
3	Running the software	2

1 Description

Triefinder is a software to build an index of all the tags of a certain restriction enzyme site in Refseq, Unigene and Genomic databases and query a fastq of sequenced tags to assign them to known tags in the index.

2 Installation

2.1 Building the binaries

Make sure the main script and worker scripts are executable:

```
chmod u+x mainScript.pl
chmod u+x src/mergeAnnotation.pl
chmod u+x src/readFiles.pl
```

Build the C++ binaries:

```
cd src/ make
```

2.2 Setting up the databases

Download the databases:

- Download the RefSeq for Zebrafish. for example we used

```
wget ftp://ftp.ncbi.nih.gov/refseq/D_rerio/mRNA_Prot/zebrafish.rna.fna.gz
```

- Download the Unigene for Zebrafish for example we used

```
wget ftp://ftp.ncbi.nih.gov/repository/UniGene/Danio_rerio/Dr.seq.all.gz
```

- Download the Genome for Zebrafish
- Concatenate all the genomic sequences for the Zebrafish chromosomes (genomic database) into a single file.

3 Running the software

Type:

```
./mainScript.pl
```

for a list of the available options.