

MetaCluster

User manual v0.1 *beta*

Released on Jan 21st 2010

Outline

- 1 File Information
- 2 Runtime Environment
- 3 Input & Output
- 4 Command Format

File Information

The version 0.1 *beta* release package contains the C++ source code of “MetaCluster_0.1_beta” and one PDF version user manual: “MetaCluster_Manual.pdf”. You can also check our website to download simple examples: <http://www.cs.hku.hk/~alse/MetaCluster>

Runtime Environment

64-bit processor and 4G or more memory is very necessary for better performance or solving large dataset (over 10,000 DNA fragments and 100,000 is the maximum value). MetaCluster is developed and tested under Ubuntu 8.10 AMD 64 version and Debian 5.0 AMD 64 version and should be suitable for all Linux environment.

Input & Output

The input of MetaCluster should be a sequencing dataset contain several DNA fragments with FASTA format like:

```
>read0001
TACGCGGACTTCGATGCG ... GCTAGGCTGACG
>read0002
GCCTGCGAAGCGCGATTT ... AGCTGAGGCGTC
```

An estimated approximate cluster number say n , is another input. And according to this value, the MetaCluster will divide the input dataset to n taxonomic specific output files with FASTA format as the output under the same path of the binary file, like:

```
gourp0.fasta; group1.fasta ...
```

Command Format

The typical command format of MetaCluster is:

```
$ MetaCluster input_file [--classes n]
```

The default value of n is 2.

*For research and
academic use
All copyrights for
commercial application
reserved*

