ISSN: 2574-1241
DOI: 10.26717/BJSTR.2018.11.002172
Maoshan Chen. Biomed J Sci & Tech Res

Appendix 2: Detailed description of pipeline_config.txt.

Example:

```
[Para]
Para_inputdir=/home/lusf/software/program/example
Para_outputdir=/home/lusf/software/CPLSTool/program/example/qiime_result
[DB]
DB_reference=/home/lusf/software/CPLSTool/program/example/97_otus.fasta
```

DB_reference=/home/lusf/software/CPLSTool/program/example/97_otus.fasta

Description:

1. There are two main parts, *[Para]* and *[DB]*, which are the definition of variables in the corresponding programs for the pipeline set in the *BIN_module.txt file*. *[Para]* is representative of Parameters, and *[DB]* is representative of Database.

2. The parameters Para_inputdir, Para_outputdir and DB_reference are used in the *BIN_module.txt* file.

3. *[sample]* is another important parameter, which can be set if there are more than one sample, especially when the samples need to be analyzed one by one. The example is as the following:

```
[sample]$
D140214^IA3-1^I500bp^IPE250^I7.5Gb^Iraw data$
D140215^IA3-2^I500bp^IPE250^I7.5Gb^Iraw data$
D140218^IA6-1^I500bp^IPE250^I7.5Gb^Iraw data$
D140219^IA6-2^I500bp^IPE250^I7.5Gb^Iraw data$
D140222^IA14-1^I500bp^IPE250^I7.5Gb^Iraw data$
D140223^IA14-2^I500bp^IPE250^I7.5Gb^Iraw data$
```

(1) The first column are experiment numbers of the samples;

(2) The second column are the names of the samples,which will be used in the result;

(3) The third column is the insert length of experiment libraries;

(4) The fourth column is the sequencing length;

(5) The fifth column is the base number of the samples;

(6) The sixth column is quality control of the samples,such as raw data or clean data.

There are more useful descriptions after running the pipeline.

1. If a minor step is broken, the process of all the pipeline would not be affected. The status of every step can be checked by *show_process.py* program. The broken minor step can be re-analyzed after the accomplishment of the whole pipeline.

2. If the major step is broken, the process of all the pipeline would certainly be affected, the major step and subsequent steps need to be delivered again upon killing the running job.

3. There are five status of the tasks including *running, break, plan, hold and end*. The status of '*running*' means the task is under running, the status of '*break*' means the task is broken, '*plan*' means the task is about to run, '*end*' means the task is finished, and '*hold*' means the storage is not enough and more storage is needed.