

## Appendix 1: Detailed description of BIN\_module.txt.

### Example:

```

[Job Start]
Name      pickotu
Memory    2G
Queue     sci.q
CPU        N
Major      T
Order      1
Thread     1
Qsub       False
Command    BIN/pick_open_reference_otus.py -i Para_inputdir/all.fa -r DB_reference -o
Para_outputdir -f
  
```

[Job End]

Description :

1. Every step is a module, which begins with [Job Start] and ends with [Job End]. They are highlighted in red. The required parameters for every module are fixed, they are at the beginning of every line and they are shown in orange.

2. The details of the parameters are as the following:

*Name*: The name of every module, it must be set.

*Memory*: The internal storage used in every module. It can be set to a large value if it is unknown. The default value of Memory is 3G. It's important to note that OUTPUTDIR is the parameter of output directory.

*Queue*: The queue which will be used to analyze the step, such as super.q or sci.q, users can set it according to the system configuration of the used computer cluster. The default value of Queue is sci.q.

*CPU*: The maximum parallel tasks of every module. If it is a integer, then each run is delivered with a fixed number; If it is N, then all input samples will be delivered; If it is N/2 or N/3, 1/2 or 1/3 of the input samples will be delivered. The default value of CPU is 3.

*Major*: If the value is T, it is a major step, which means that the next step must wait this step, the next step can be analyzed only when this step finished; if the value is F, the step is a minor step, which means that this step can be run with the next step. The default value of Major is T.

*Order*: The order of this step, such as first, second, third and so on. It is a required parameter.

*Thread*: Number of threads for every task. The default value of Thread is 1.

*Command*: The operation command lines for this module. If it is a multi-line command, users can keep writing until meet [Job End].

3. Notably, in the command line, variables are set (in blue), all except BIN are necessary to be defined in the *pipeline\_config.txt* file.