**Metagenomics pipeline usage:**

1. **Install software and databases**

The software and databases list below must be correctly installed before using the pipeline. Software installation instruction were provided for download or look up for help.

|  |  |  |  |
| --- | --- | --- | --- |
| Software |  | Download path and install instruction | Install Database |
| Fastqc | yes | https://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc | NO |
| Prokka | yes | <https://github.com/tseemann/prokka> | NO |
| Multiqc | yes | https://github.com/ewels/MultiQC |  |
| Trimmomatic | yes | http://www.usadellab.org/cms/?page=trimmomatic | NO |
| Kneaddata |  | https://github.com/biobakery/kneaddata | YES |
| Kraken2 | yes | http://ccb.jhu.edu/software/kraken2/index.shtml?t=downloads | YES |
| megahit | yes | <https://github.com/voutcn/megahit> | YES |
| cd-hit | yes | http://weizhongli-lab.org/cd-hit/ | NO |
| Bracken | yes | https://ccb.jhu.edu/software/bracken/ | NO |
| Humann3 | Humann2 | https://huttenhower.sph.harvard.edu/humann/ | YES |
| Metaphlan3 | Yes | https://github.com/biobakery/MetaPhlAn/wiki/MetaPhlAn-3.0 | YES |
| Bowtie2 | yes | http://bowtie-bio.sourceforge.net/bowtie2/index.shtml | NO |
| Samtools | yes | https://github.com/samtools/samtools | NO |
|  |  |  |  |

1. **Run example:**

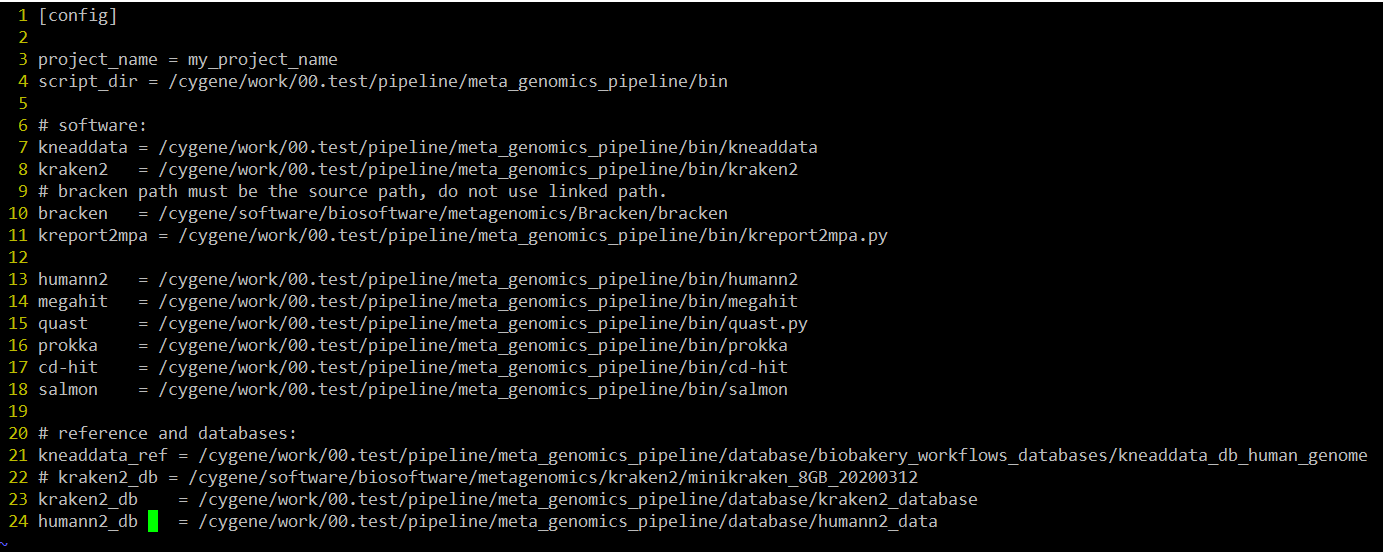
After finishing install software and databases, run this example to check the installation:

2.1 Go to example folder: `cd example`

configuring config.txt

`vi config.txt`

Modify these manually then save and quit edit.



* 1. configuring `data.list` file.

`vi data.list`

put your data full path correctly each line and give it a right name with a tab separate.

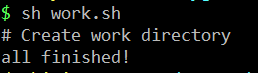


Note: this pipeline only suitable for SE fastq data.

2.3. Run it.

Do not change anything in work.sh file,

`sh work.sh`



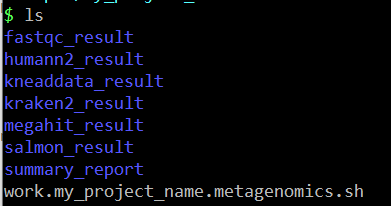
Then the project folder will be generated automatically. Go to the project file.

* 1. run the pipeline script.

`sh work.my\_project\_name.metagenomics.sh`

1. **Check the results.**

The result folder will be generated by the pipeline.



Checking each result in the corresponding location.

The html report will be generated in `summary\_report` folder, but not yet finished.