

Project 2 Part 1: Analysis of metagenomic sequences
Xiaojun Gao

Part 1:

Run the following command for kraken on out data against provided data base:

```
kraken --db /opt/ccb/data/Metagenomics/minikraken_20171101_8GB_dustmasked  
/opt/ccb/data/Metagenomics/MiSeq_metareads_50K.fq > kraken_output
```

Part 2:

Run the following command to generate a kraken report for result from kraken-report --db /opt/ccb/data/Metagenomics/

```
minikraken_20171101_8GB_dustmasked kraken_output > kraken_report
```

Next, write a script to produce a summary of the 2 lists of counts for genus and species. Run the script on kraken_report and produce our own kraken_lab_report:

```
g++ -o myProgram kraken-report-summary.cpp  
./myProgram kraken_report kraken_lab_report
```

The files submitted related to Kraken are the following:

kraken_report

kraken_lab_report

kraken-report-summary.cpp (not required, the script for generating kraken_lab_report)

This pdf document

Part 3:

For the first species, the identifier of the read is

"SRR1613382.30803.1". The initial Kraken classification is "Tai Forest ebolavirus". The species I identified with megablast is "Zaire ebolavirus". The BLAST-E value is 1e-68.

For the second species, the identifier of the read is

"SRR1613382.10923.1". The initial Kraken classification is

"Bradyrhizobium sp. ORS 285". The species I identified with megablast is "Bradyrhizobium sp. PSBB068". The BLAST-E value is 2e-70.

The file submitted for megablast is this pdf document.