/\*Commands used for Quake

\*<http://www.cbcb.umd.edu/software/quake/manual.html>

\*/

/\*Tried multiple streaming commands and none worked.

\*The error was found out to be caused by the mapper file

\* not reading from standard input and reading from input file

\*/

/\*Noticed that I did not use the ‘-file’ parameter to point to the directory

\*for the mapper and reducer (as shown in the lecture videos).

\*This is because if the files were already on hdfs (which I did transfer over to hdfs),

\* then there is no need to specify the directory of the mapper/reducer files.

\*/

$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming.jar -input /user/staph\_data/ERR064898\_1.fastq -output count\_test.txt -mapper /user/staph\_data/count-qmers.cpp -reducer /user/staph\_data/reduce-qmers.cpp

/\*Since the above command failed, I tried to specify the -input and -output \*/

$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming.jar -input /user/staph\_data/ERR064898\_1.fastq -output count\_test.txt -file /user/staph\_data/count-qmers.cpp -mapper /user/staph\_data/count-qmers.cpp -file /user/staph\_data/reduce-qmers.cpp -reducer /user/staph\_data/reduce-qmers.cpp

$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming.jar -input /user/staph\_data/ERR064898\_1.fastq -output /user/staph\_data/count\_test.txt -file /home/biotools/Quake/src/count-qmers.cpp -mapper /home/biotools/Quake/src/count-qmers.cpp -file /home/biotools/Quake/src/reduce-qmers.cpp -reducer /home/biotools/Quake/src/reduce-qmers.cpp

/\* Tried command with file count\_qmers.cpp instead of count-qmers.cpp \*/

$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming.jar -input /user/staph\_data/ERR064898\_1.fastq -output /user/staph\_data/count\_test.txt -file /home/biotools/Quake/src/count-qmers.cpp -mapper /home/biotools/Quake/src/count-qmers.cpp -file /home/biotools/Quake/src/reduce-qmers.cpp -reducer /home/biotools/Quake/src/reduce-qmers.cpp

/\* None of the hadoop streaming command worked

\* because of the mapper file. If the mapper file was changed

\* to read from standard input instead of from the file, there is

\* a chance that this command might work and produce

\* the appropriate counts files. If the command works,

\* then the below steps will be the following:

\*/

/\*Run post-correction after getting counts.

\*This is done on command-line and outside of Hadoop

\*/

$ cov\_model.py [counts file]

/\*Correct the reads. The developer said that this step can

\*be ran in parallel inside Hadoop. However, the scripts for that

\*has not been written yet. A mapper and reducer will have to be

\*written for this. If ran outside of Hadoop, the command would be

\*/

$ correct -f [fastq list file] -k [k-mer size] -c [cutoff] -m [counts file]