***Idotea baltica* RNA-Seq Preprocessing**

Reptile 1.1

**INSTALLING**

Change directory to USERAPPL and create new directory called *reptile*

cd $USERAPPL

mkdir reptile

Go to the *reptile* directory and download the installation package

cd reptile

wget http://aluru-sun.ece.iastate.edu/lib/exe/fetch.php?media=source:reptile-v1.1.zip

Unzip the package

unzip fetch.php?media=source:reptile-v1.1.zip

Go to the *src* directory

cd reptile-v1.1

cd src

Compile reptile, and then copy to the *reptile-v1.1* directory

make all

cp reptile-v1.1 ..

Go to the *seq-analy* dir and compile; then copy to the *reptile-v1.1* dir

make all

cp seq-analy /homeappl/home/ketaya/appl\_taito/reptile/reptile-v1.1

Go to *reptile\_merger* and compile, and then copy to the *reptile-v1.1* dir

make all

cp reptile\_merger /homeappl/home/ketaya/appl\_taito/reptile/reptile-v1.1

Tell the shell the location of the executable

export PATH=${PATH}:${USERAPPL}/reptile/reptile-v1.1

**NOTE!** This command must be run every time logging into Taito

**PREPROCESSING**

Files must be unzipped

gunzip out\_fw\_paired.fq.gz

gunzip out\_rev\_paired.fq.gz

\*\*\* ~20min/file, ~13GB to ~74GB

Rename files from ‘.fq’ to ‘.fastq’

mv out\_fw\_paired.fq out\_fw.paired.fastq

mv out\_rev\_paired.fq out\_rev\_paired.fastq

Run the preprocessing Perl script, flag=2(Keep reads intact)

perl $USERAPPL/reptile/reptile-v1.1/utils/fastq-converter-v2.0.pl \

timmomatic\_unzipped . 2

\*\*\* ~25min/file, NOTE! renames reads to counting number

**PARAMETER TUNING**

Make a copy of the example seq-analy configuration file

cp $USERAPPL/reptile/reptile-v1.1/utils/seq\_analy/config-example \

seq-analy-config

Update the input and output file names in the configuration file

InFaFile: out\_fw\_paired.fa

IQFile: out\_fw\_paired.q

OKmerHistFile: out\_fw\_paired.kmerhist

QHistFile: out\_fw\_paired.qualhist

vi seq-analy-config

Create batch jobs file and run

sbatch reptile\_seq-analy.sh

\*\*\* ~4hrs (8 CPUs per task, 8GB memory) ~4hrs (16 cores, 16GB memory)

Open the quality histogram file

less out\_fw\_paired.qualhist

Update *seq-analy-config* file according to the README file

QThreshold: 70/70

Qlb: 66/66 (for the Reptile configuration file)

MaxBadQPerKmer: 8/8

KmerLen: 24/24

OKmerHistFile: out\_fw\_paired2.kmerhist

QHistFile: out\_fw\_paired2.qualhist

vi seq-analy-config

Re-run ‘seq-analy’

sbatch reptile\_seq-analy.sh

\*\*\* ~4.5hrs (16 cores, 16GB memory) ~ 4.5hrs (8 cores, 8GB memory)

Make a copy of the example Reptile configuration file

cp $USERAPPL/reptile/reptile-v1.1/src/config-example reptile-config

Open the k-mer histogram file

less out\_fw\_paired2.kmerhist

Update the Reptile configuration file according to the README file

InFaFile: out\_fw\_paired.fa

IQFile: out\_fw\_paired.q

OErrFile: out\_fw\_paired.errors

T\_expGoodCnt: 390/310

T\_card: 2/2

KmerLen: 12/12

QThreshold: 70/70

Qlb: 66/66

MaxBadQPerKmer: 8/8

Step: 12/12

vi reptile-config

**RUNNING**

Create batch jobs file and run

sbatch reptile.sh

\*\*\* ~12.5hrs (8 cores, 16GB memory, c=2), ~12.5hrs (8cores, 32GB memory, c=1)

Create batch jobs file and run to generate corrected sequences

sbatch reptile\_merger.sh

# output: out\_fw\_corrected.fa

\*\*\* ~ 4hrs (8 cores, 16GB memory)

**REAPEAT FOR OTHER PAIR**