Erika Maldonado-Rosado

03 February 2019

Sequence Assembly

BINF 6203

**Introduction**

In this lab we are assembling a tomato chloroplast, two E. coli fragments with different base pairs, and sets of PacBio CCS and CLR reads for E. coli. Our goal with this lab is to learn how to assembly genome fragments while using mamba our cloud base. We are also learning how to use Spades and Quast, programs to help assemble the genomes.

**Method**

For this lab you need access to the Dropbox link given to us by Dr. Gibas to get all the files we need, access to the Mamba where you will be able to use Spades, and Quast which we will be putting onto the Mamba server. To access Mamba one needs to use a VPN, for the purpose of this lab I used Cisco VPN which one can access using the student email and password. To get Quast and our file onto Mamba one must use the ‘sftp’ and ‘put’ command. Once done, one will have to get on the mamba to server to unzip and tar our zip files. Lastly, you need a text editor, for this lab I used Nano to create my script. In order to submit any request one must use qsub. For further instructions one can find them at University of North Carolina at Charlotte Mamba User notes.

**Results**

I numbered each script with an image below. The first three are the tomato chloroplast on the first I did it without the careful and only assembler, second I did both careful and only assembler, and the third I did only assembler. Careful is used to minimize number of mismatches in the final contigs. Only assembler is used to run assembly modules only. The difference between all of them was the number of contigs. The first one had 5 contigs, the second had 9, and the last one had 3.

Four and five are the E. coli genome and I did not notice a difference between the two. They both had the same number of contigs and the same number of base pairs (bp) per each of the results.

Six is the CLR and seven is the CCS both using the E. coli genomes. The CLR used ERR008613, while the CCS used ERR022075. The difference between CLR and CCS is that CLR is used for hybrid assemblies, there is no need to pre-correct this data. While CCS data needs to be pre-corrected and is usually single reads. CLR showed no difference between the E. Coli shown in four or five. CCS, however, shows immense difference with smaller amounts of contigs total and for bp.

**1) BC30**

#!/bin/bash

#===OPTIONS=======

#PBS -q mamba

#PBS -N spades

#PBS -l nodes=1:ppn=16

#PBS -l walltime=2:00:00

#PBS -l prologue=/users/emaldon1/prologue.sh,epilogue=/users/emalldon1/torque/epilogue.sh

#========END PBS OPTIONS=======

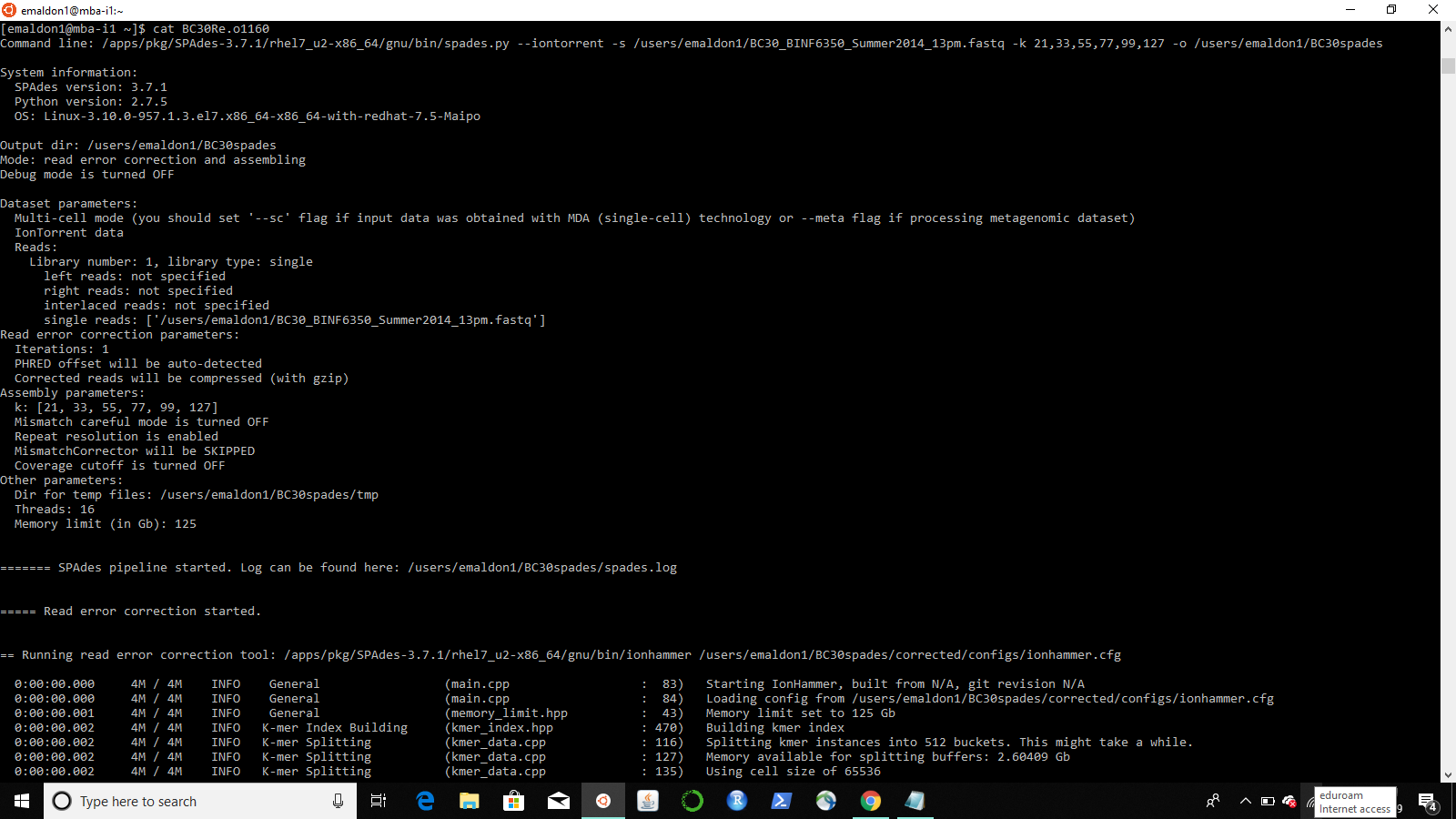
module load spades

spades.py --iontorrent -s BC30\_BINF6350\_Summer2014\_13pm.fastq -k 21,33,55,77,99,127 --careful -o BC30spadesl

/users/emaldon1/quast-5.0.2/quast.py /users/emaldon1/BC30spades/contigs.fasta -R /users/emaldon1/NC\_007898.fasta -G /users/emaldon1/NC\_007898.gff -o BC30quast

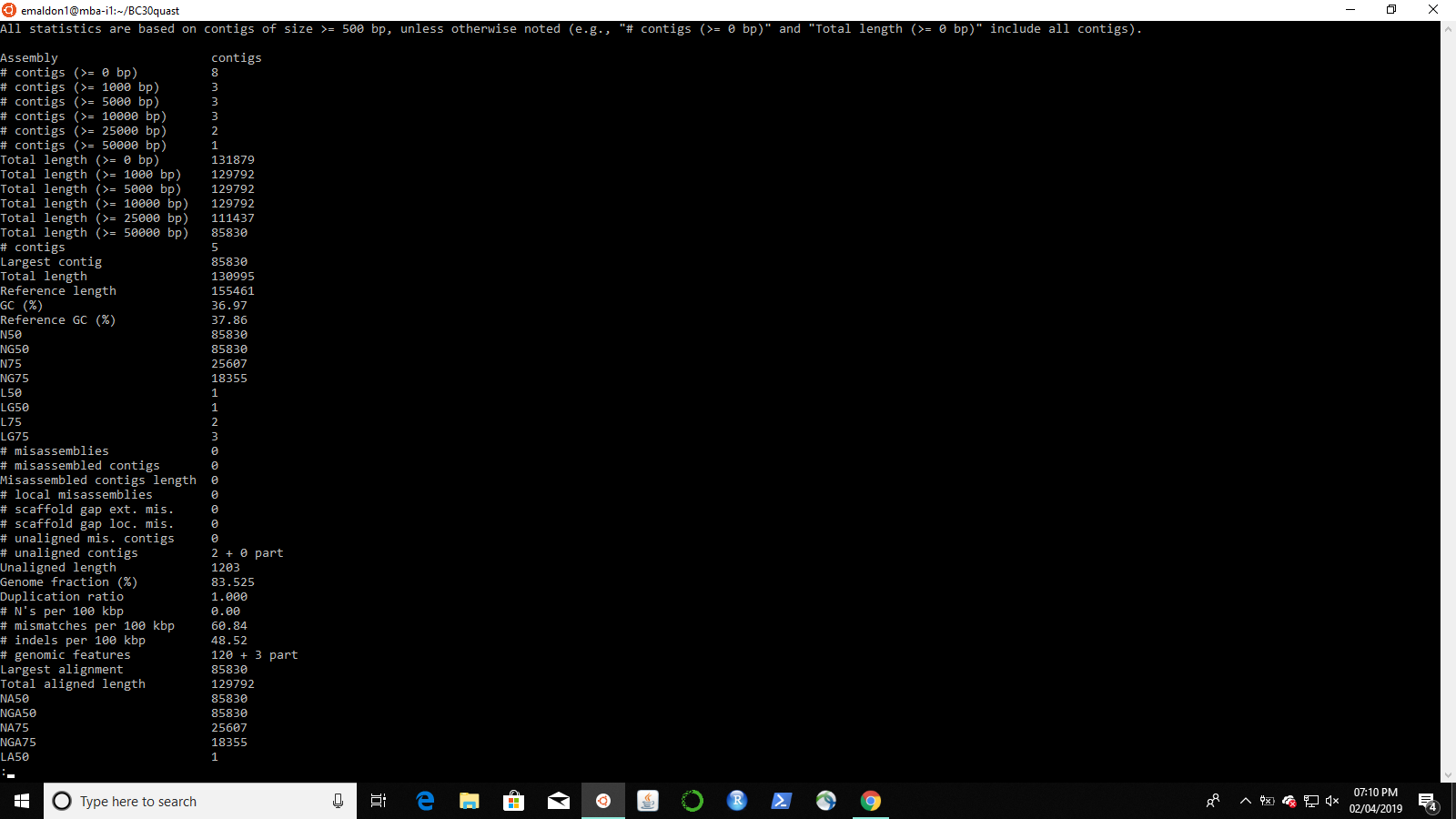
qsub -N "BC30Re" -q mamba spades.sh

cat BC30Re.o1160



cd BC30quast

less report.txt



**2) BC30 Assembly and Careful**

#!/bin/bash

#===OPTIONS=======

#PBS -q mamba

#PBS -N spades

#PBS -l nodes=1:ppn=16

#PBS -l walltime=2:00:00

#PBS -l prologue=/users/emaldon1/prologue.sh,epilogue=/users/emalldon1/torque/epilogue.sh

#========END PBS OPTIONS=======

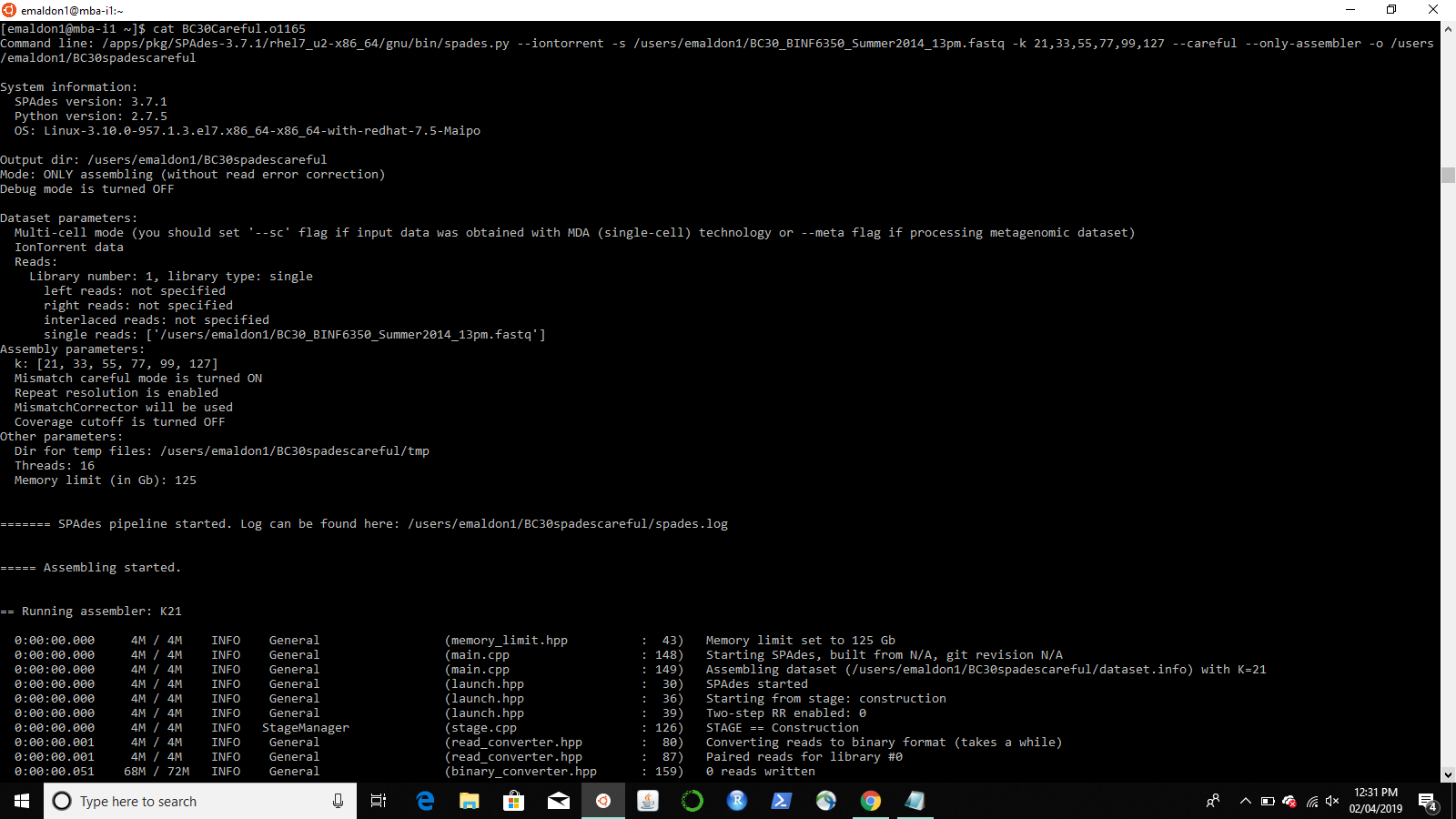
module load spades

spades.py --iontorrent -s BC30\_BINF6350\_Summer2014\_13pm.fastq -k 21,33,55,77,99,127 --careful --only-assembler -o BC30spadescareful

/users/emaldon1/quast-5.0.2/quast.py /users/emaldon1/BC30spadescareful/contigs.fasta -R /users/emaldon1/NC\_007898.fasta -G /users/emaldon1/NC\_007898.gff -o BC30carefulquast

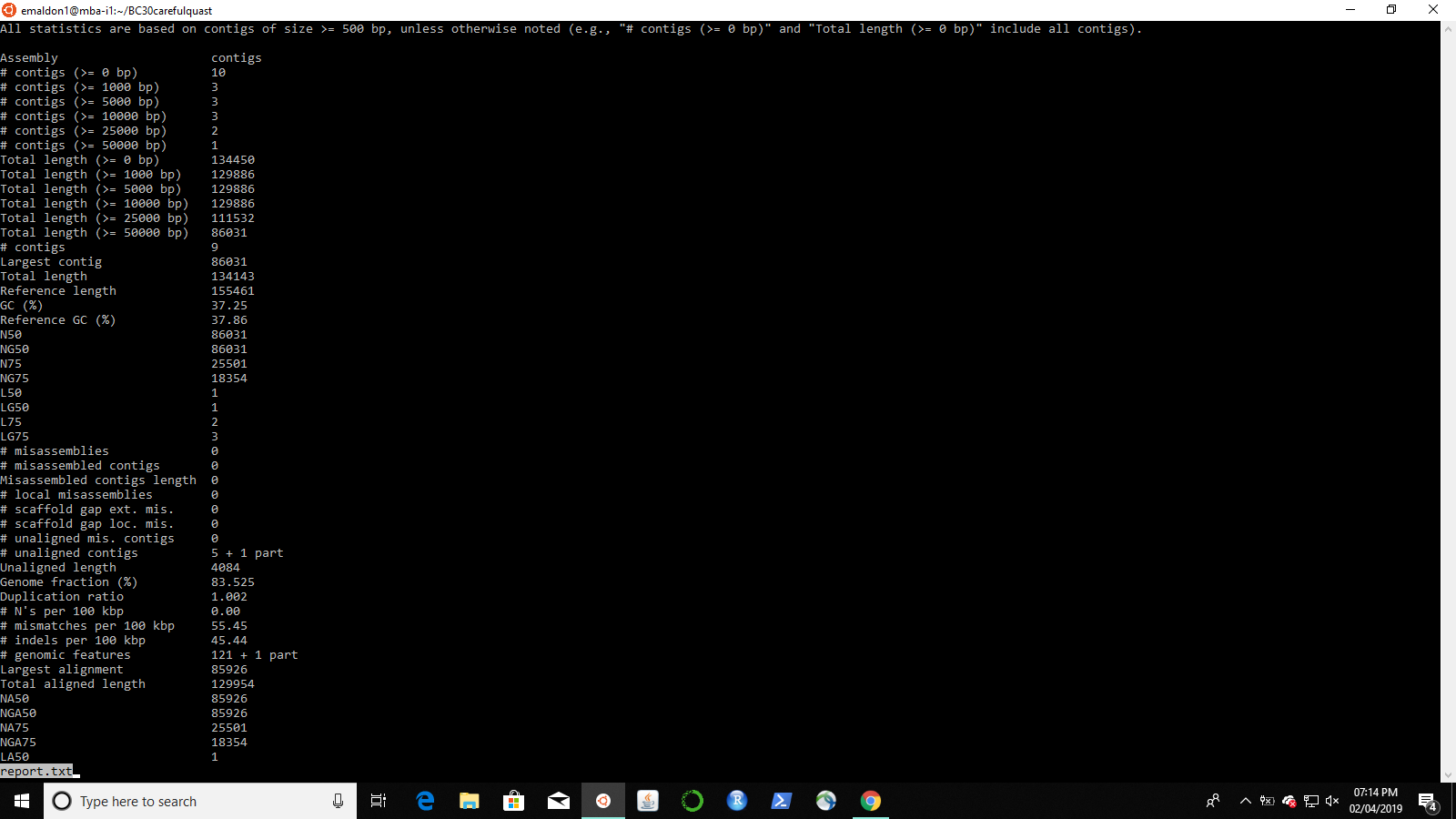
qsub -N "BC30Careful" -q mamba spadescareful.sh

cat BC30Careful.o1165



cd BC30Carefulquast

less report.txt



**3) BC30 Careful Only**

#!/bin/bash

#===OPTIONS=======

#PBS -q mamba

#PBS -N spades

#PBS -l nodes=1:ppn=16

#PBS -l walltime=2:00:00

#PBS -l prologue=/users/emaldon1/prologue.sh,epilogue=/users/emalldon1/torque/epilogue.sh

#========END PBS OPTIONS=======

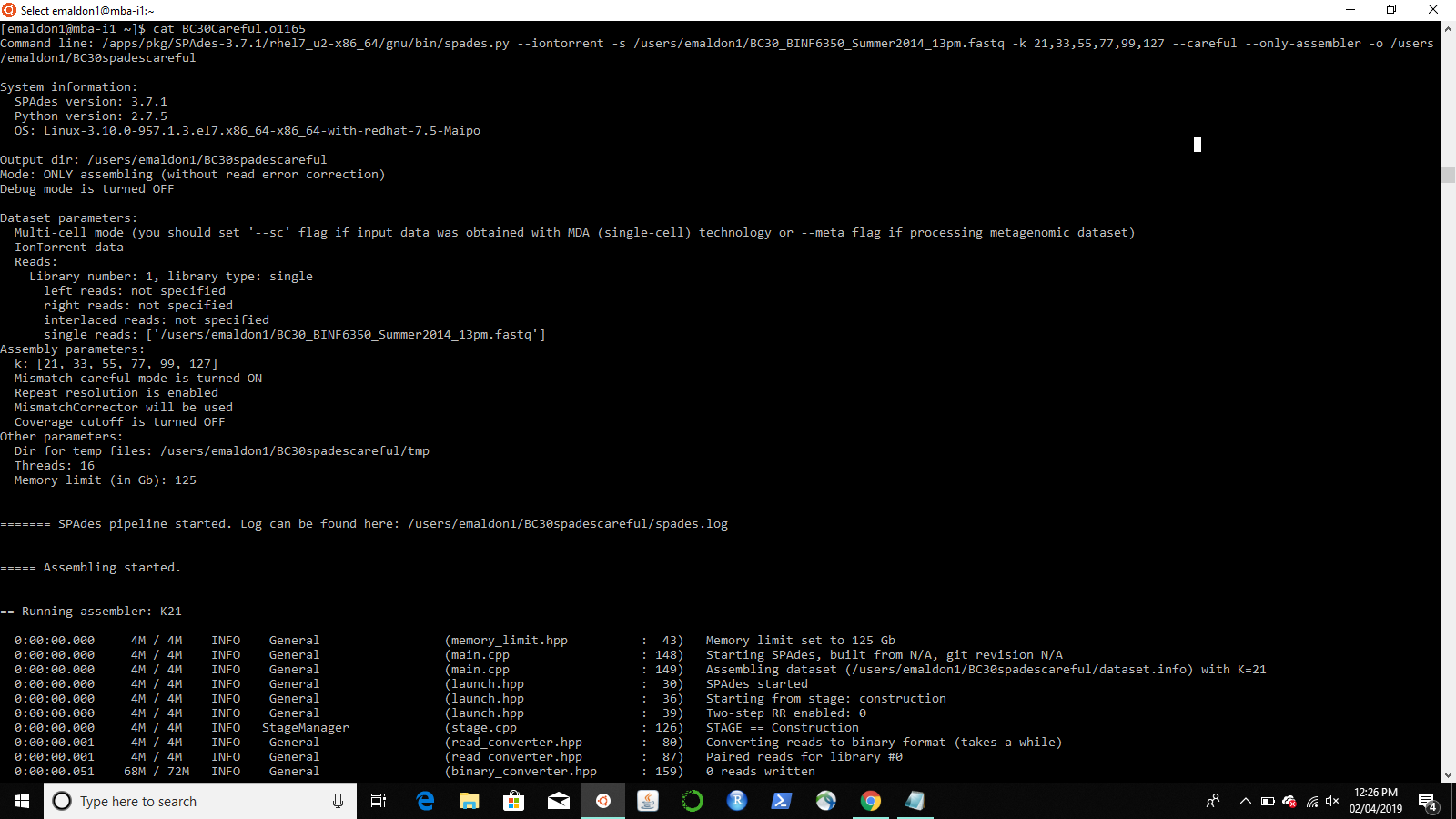
module load spades

spades.py --iontorrent -s BC30\_BINF6350\_Summer2014\_13pm.fastq -k 21,33,55,77,99,127 --careful -o BC30spadescarefulonly

/users/emaldon1/quast-5.0.2/quast.py /users/emaldon1/BC30spadescarefulonly/contigs.fasta -R /users/emaldon1/NC\_007898.fasta -G /users/emaldon1/NC\_007898.gff -o BC30carefulonlyquast

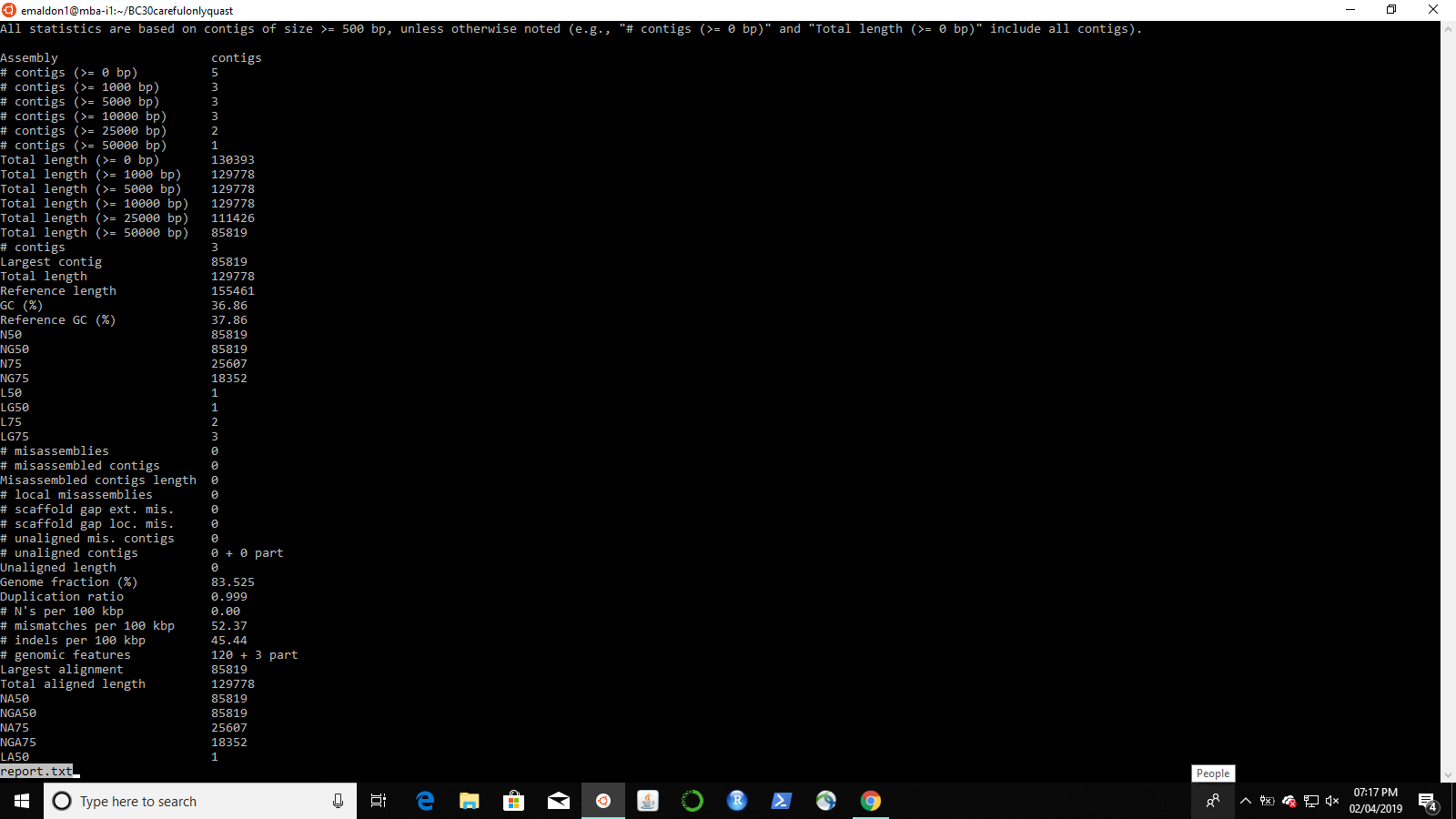
qsub -N "BC30Carefulonly" -q mamba spadescarefulonly.sh

cat BC30Carefulonly.o1179



cd BC30Carefulonlyquast

less report.txt



**4) E. coli 008613**

#!/bin/bash

#===OPTIONS=======

#PBS -q mamba

#PBS -N spades

#PBS -l nodes=1:ppn=16

#PBS -l walltime=2:00:00

#PBS -l prologue=/users/emaldon1/prologue.sh,epilogue=/users/emalldon1/torque/epilogue.sh

#========END PBS OPTIONS==========

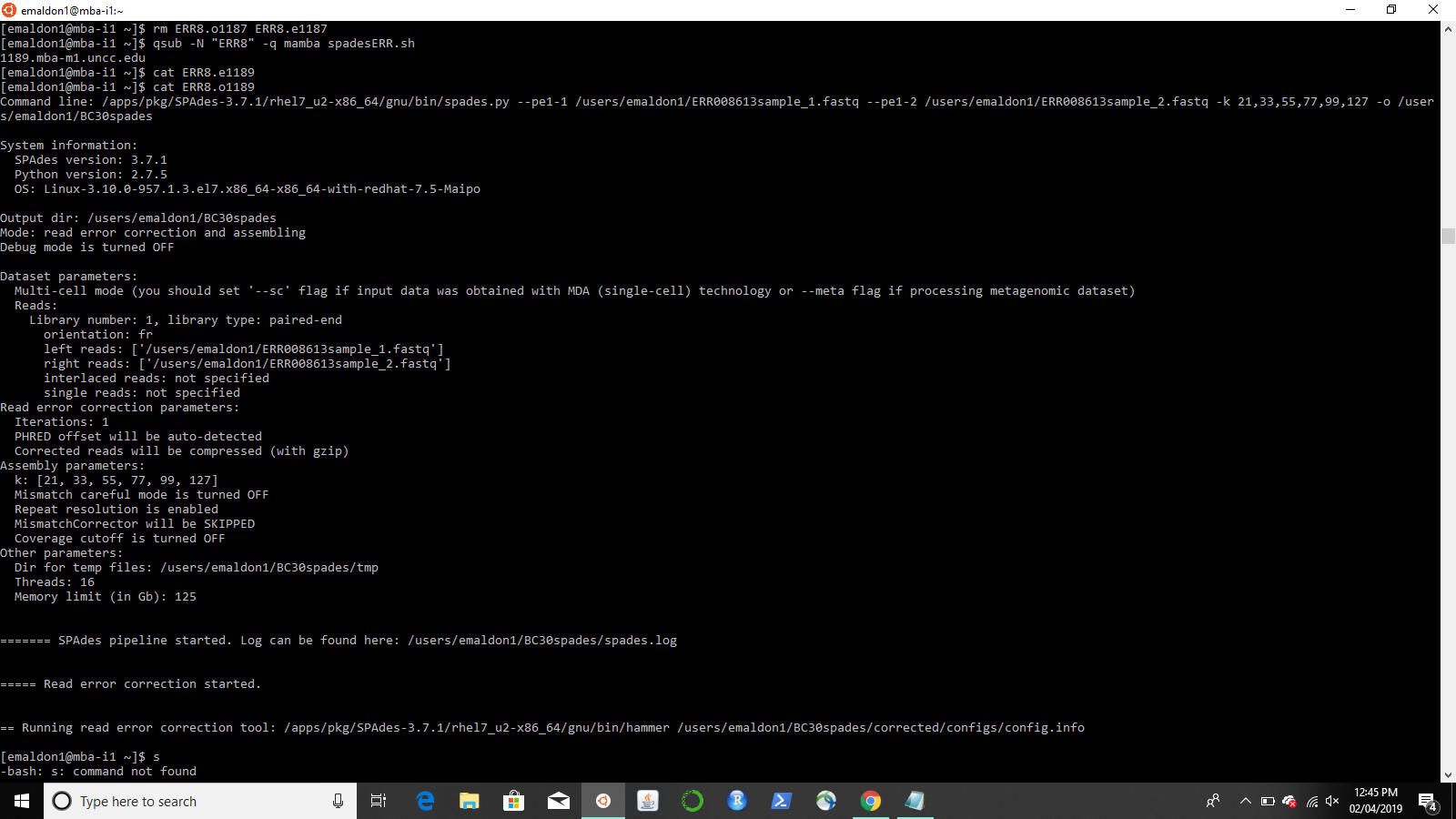
module load spades

spades.py --pe1-1 ERR008613sample\_1.fastq --pe1-2 ERR008613sample\_2.fastq -k 21,33,55,77,99,127 -o BC30spades

/users/emaldon1/quast-5.0.2/quast.py /users/emaldon1/BC30spades/contigs.fasta -R NC\_007898.fasta -G NC\_007898.gff -o ERR8quast

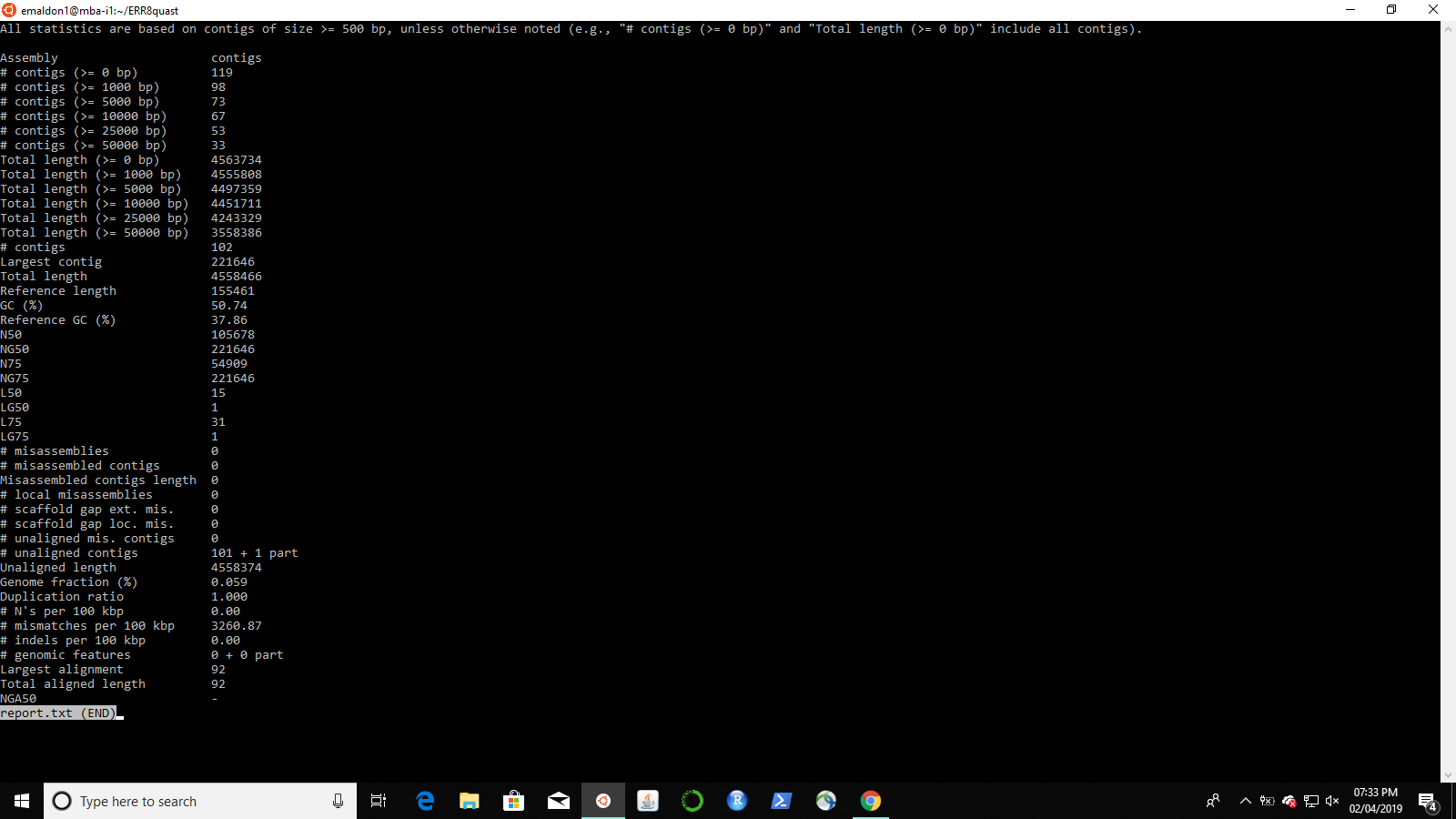
qsub -N "ERR8" -q mamba spadesERR.sh

cat ERR8.o1189



cd ERR8quast

less report.txt



**5) E. coli 022075**

#!/bin/bash

#===OPTIONS=======

#PBS -q mamba

#PBS -N spades

#PBS -l nodes=1:ppn=16

#PBS -l walltime=2:00:00

#PBS -l prologue=/users/emaldon1/prologue.sh,epilogue=/users/emalldon1/torque/epilogue.sh

#========END PBS OPTIONS==========

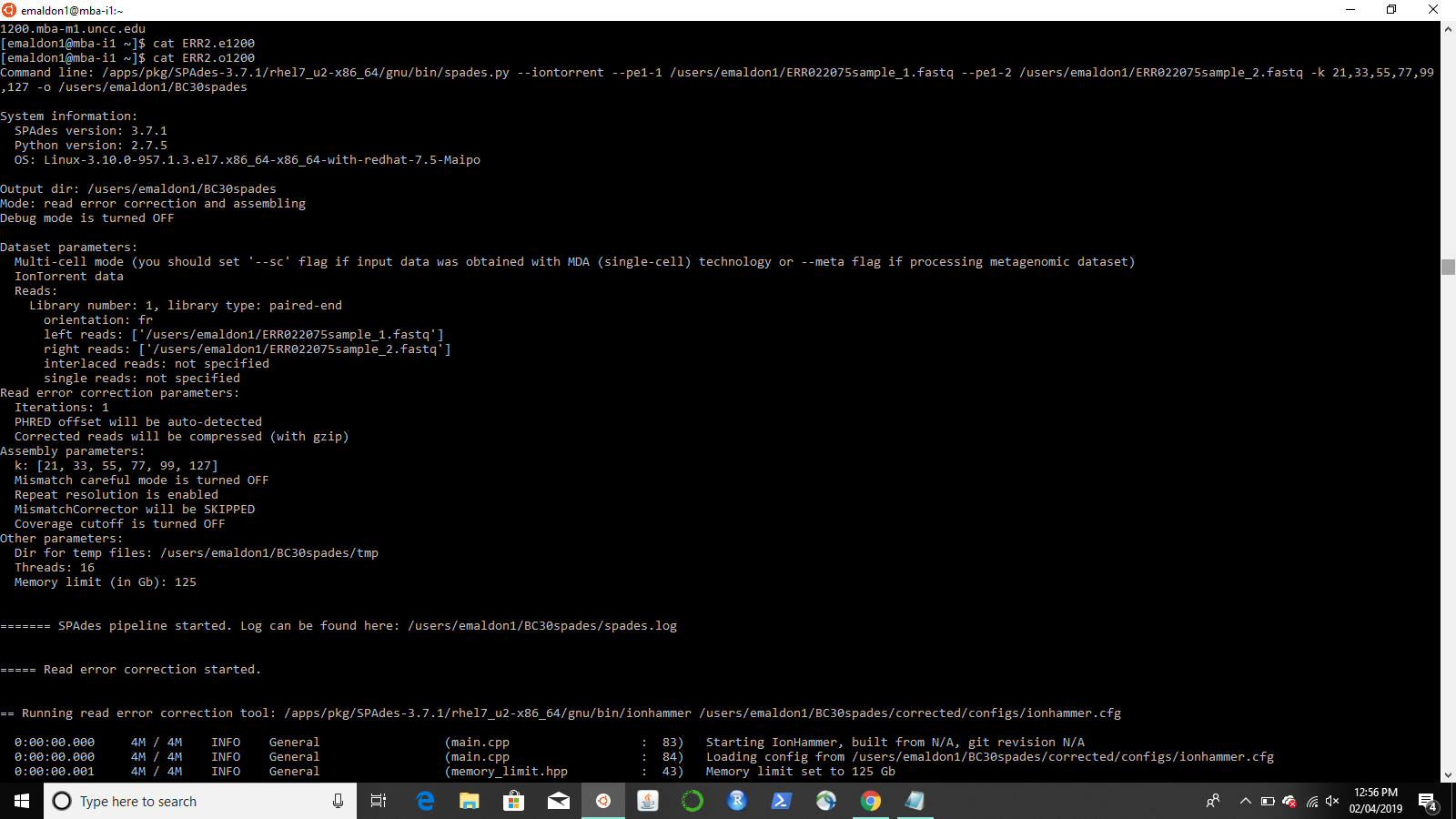
module load spades

spades.py --iontorrent --pe1-1 ERR022075sample\_1.fastq --pe1-2 ERR022075sample\_2.fastq -k 21,33,55,77,99,127 -o BC30spades

/users/emaldon1/quast-5.0.2/quast.py /users/emaldon1/BC30spades/contigs.fasta -R NC\_007898.fasta -G NC\_000913.gff -o ERR2quast

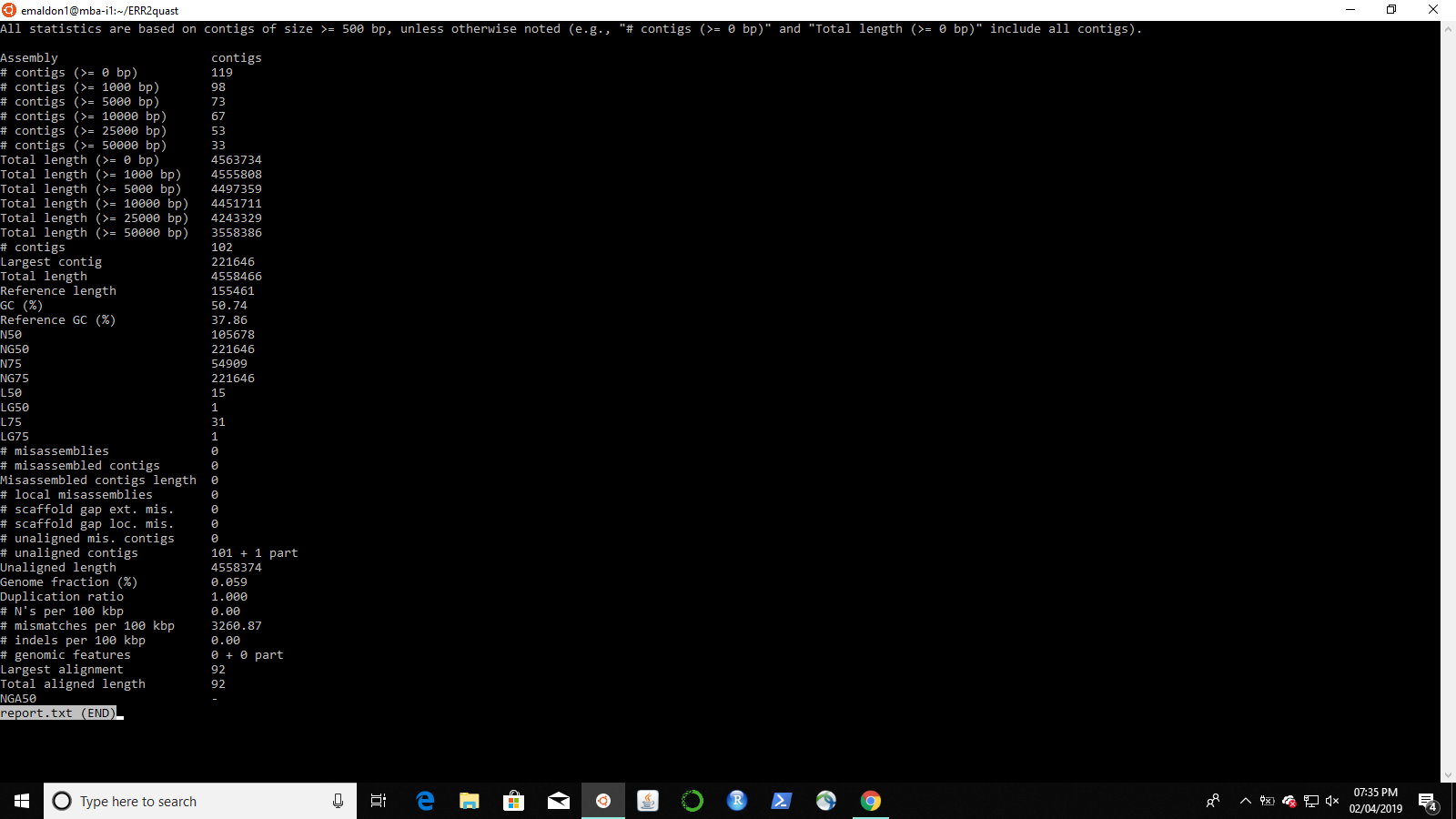
qsub -N "ERR2sp" -q mamba spadesERR22.sh

cat ERR2.e1200



cd ERR2quast/

less report.txt



**6) CLR E. coli**

#!/bin/bash

#===OPTIONS=======

#PBS -q mamba

#PBS -N spades

#PBS -l nodes=1:ppn=16

#PBS -l walltime=2:00:00

#PBS -l prologue=/users/emaldon1/prologue.sh,epilogue=/users/emalldon1/torque/epilogue.sh

#========END PBS OPTIONS==========

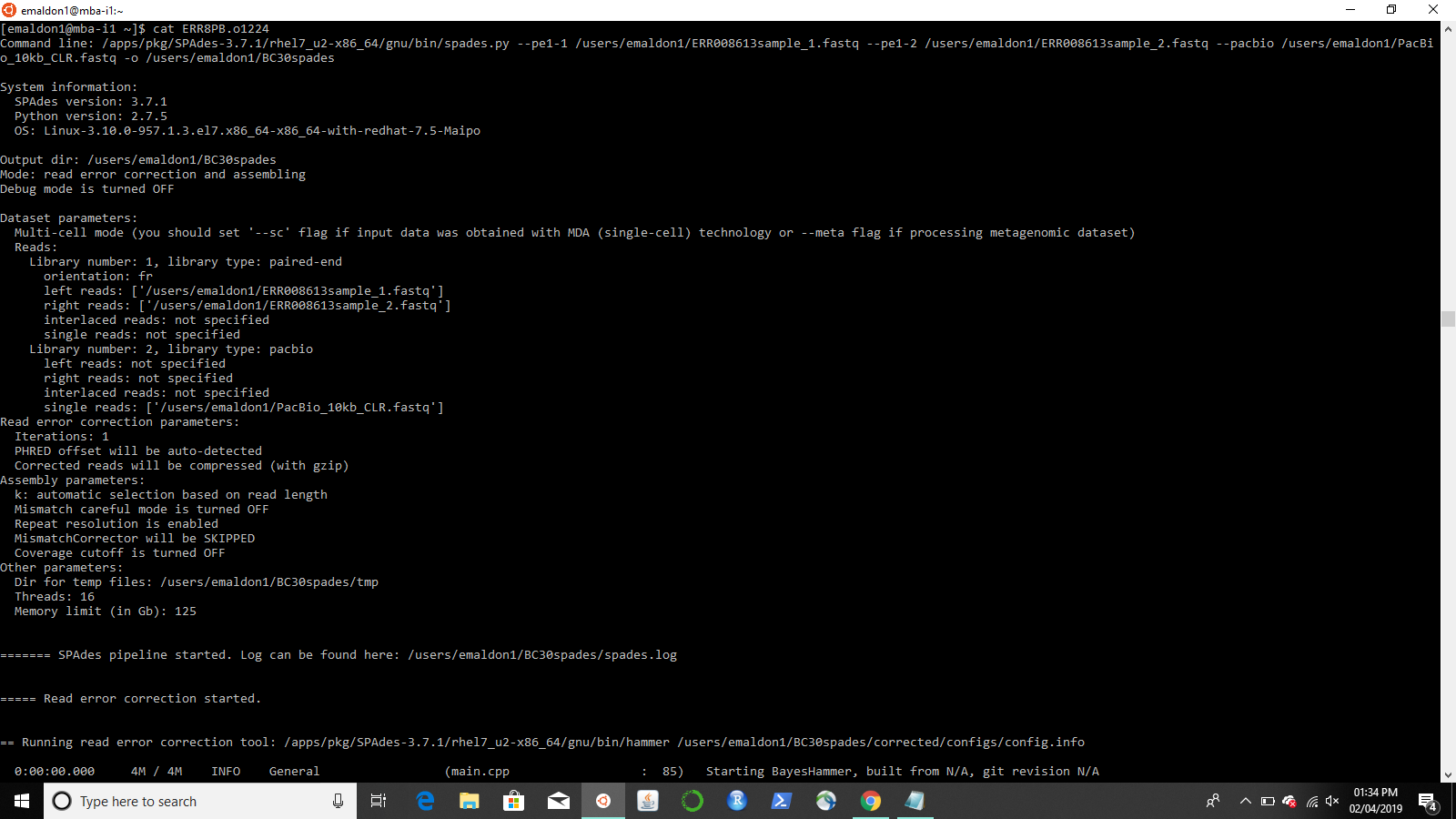
module load spades

spades.py --pe1-1 ERR008613sample\_1.fastq --pe1-2 ERR008613sample\_2.fastq --pacbio PacBio\_10kb\_CLR.fastq -o ERR8spades

/users/emaldon1/quast-5.0.2/quast.py -R NC\_007898.fasta -G NC\_000913.gff -o ERR8PBquast

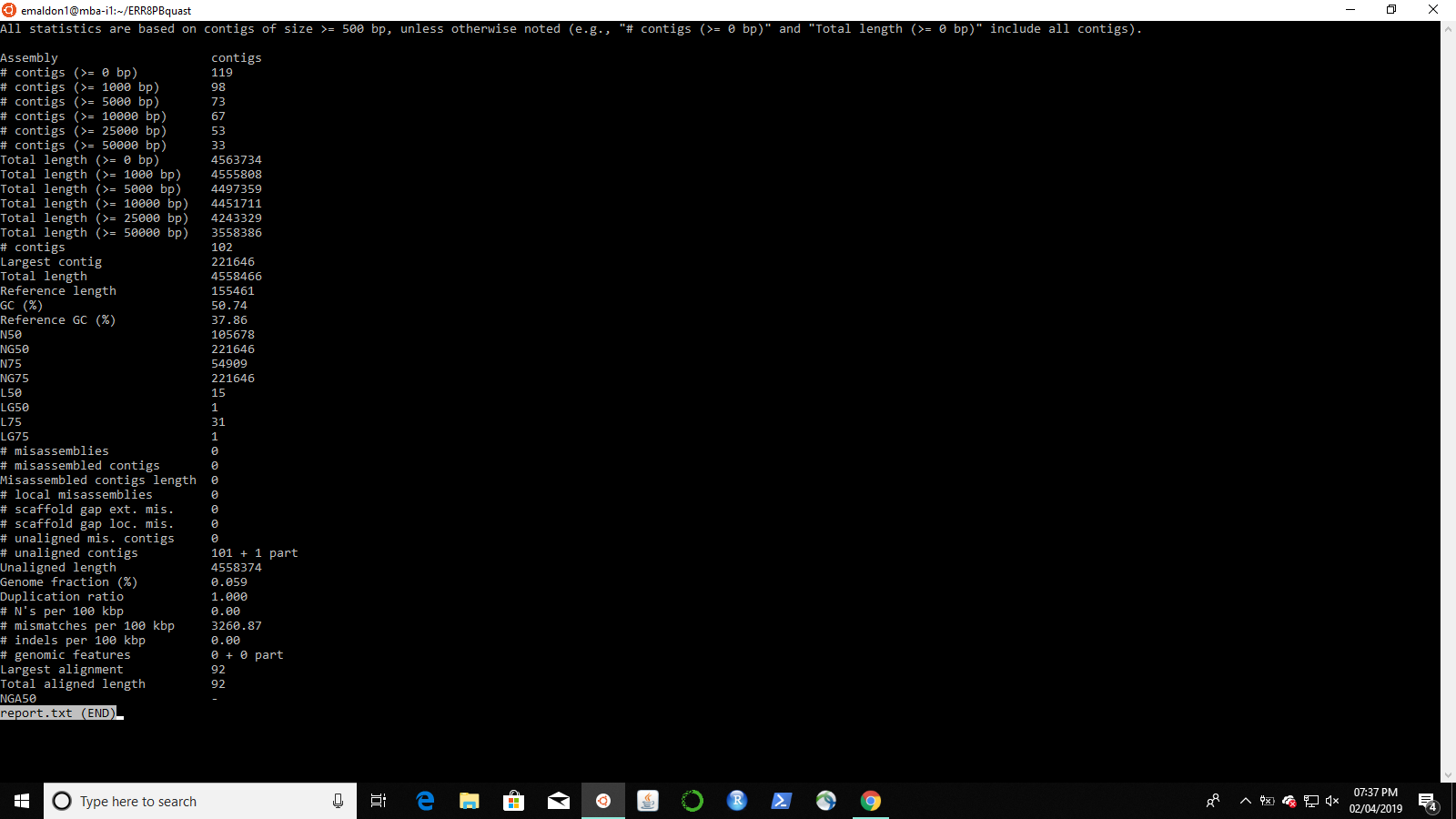
qsub -N "ERR8PB" -q mamba spadesPacBio.sh

cat ERR8PB.e1224



cd ERR8PBquast

less report.txt



**7) CCS E. coli**

#!/bin/bash

#===OPTIONS=======

#PBS -q mamba

#PBS -N spades

#PBS -l nodes=1:ppn=16

#PBS -l walltime=2:00:00

#PBS -l prologue=/users/emaldon1/prologue.sh,epilogue=/users/emalldon1/torque/epilogue.sh

#========END PBS OPTIONS==========

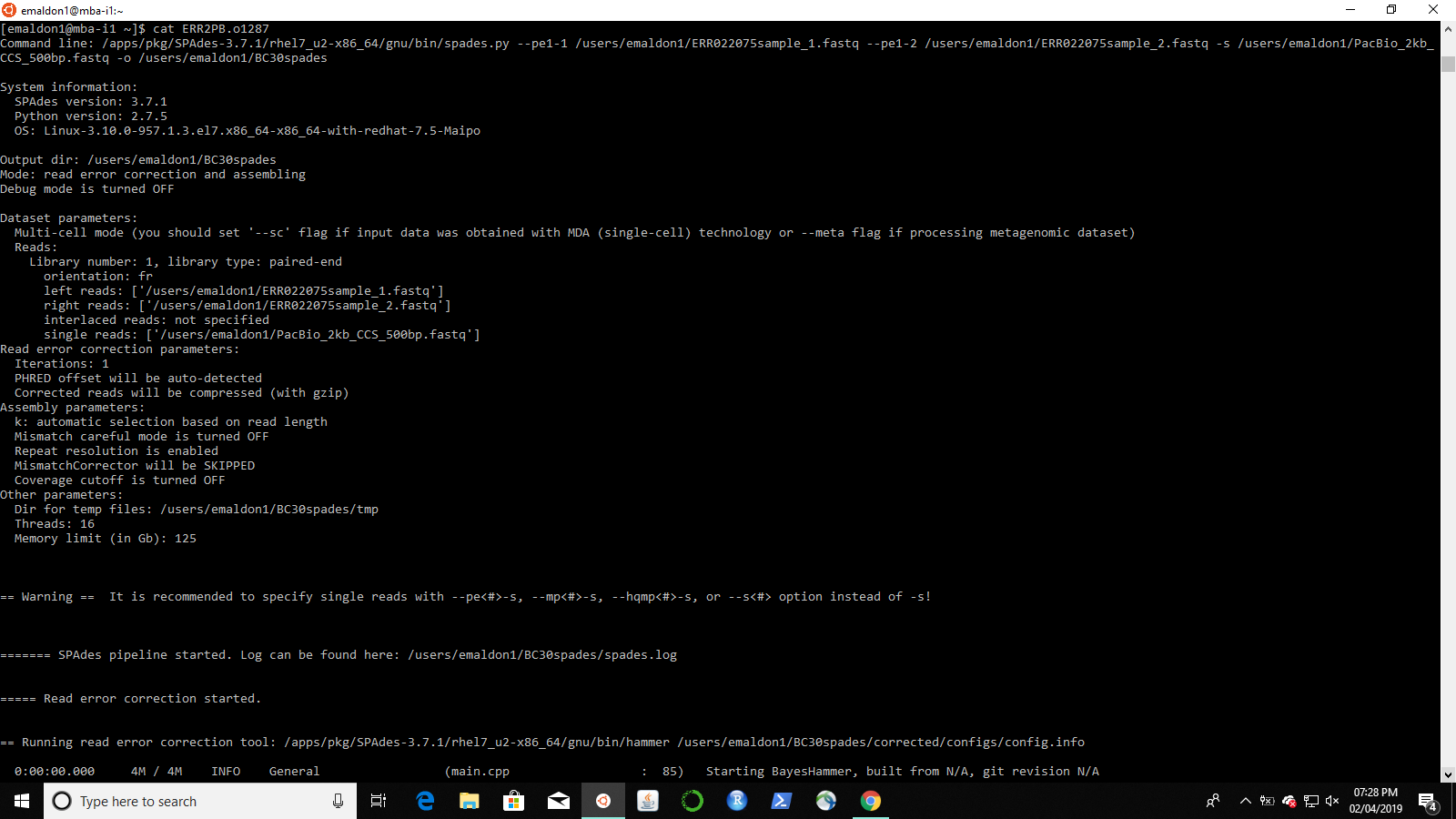
module load spades

spades.py --pe1-1 ERR022075sample\_1.fastq --pe1-2 ERR022075sample\_2.fastq -s PacBio\_2kb\_CCS\_500bp.fastq -o BC30spades

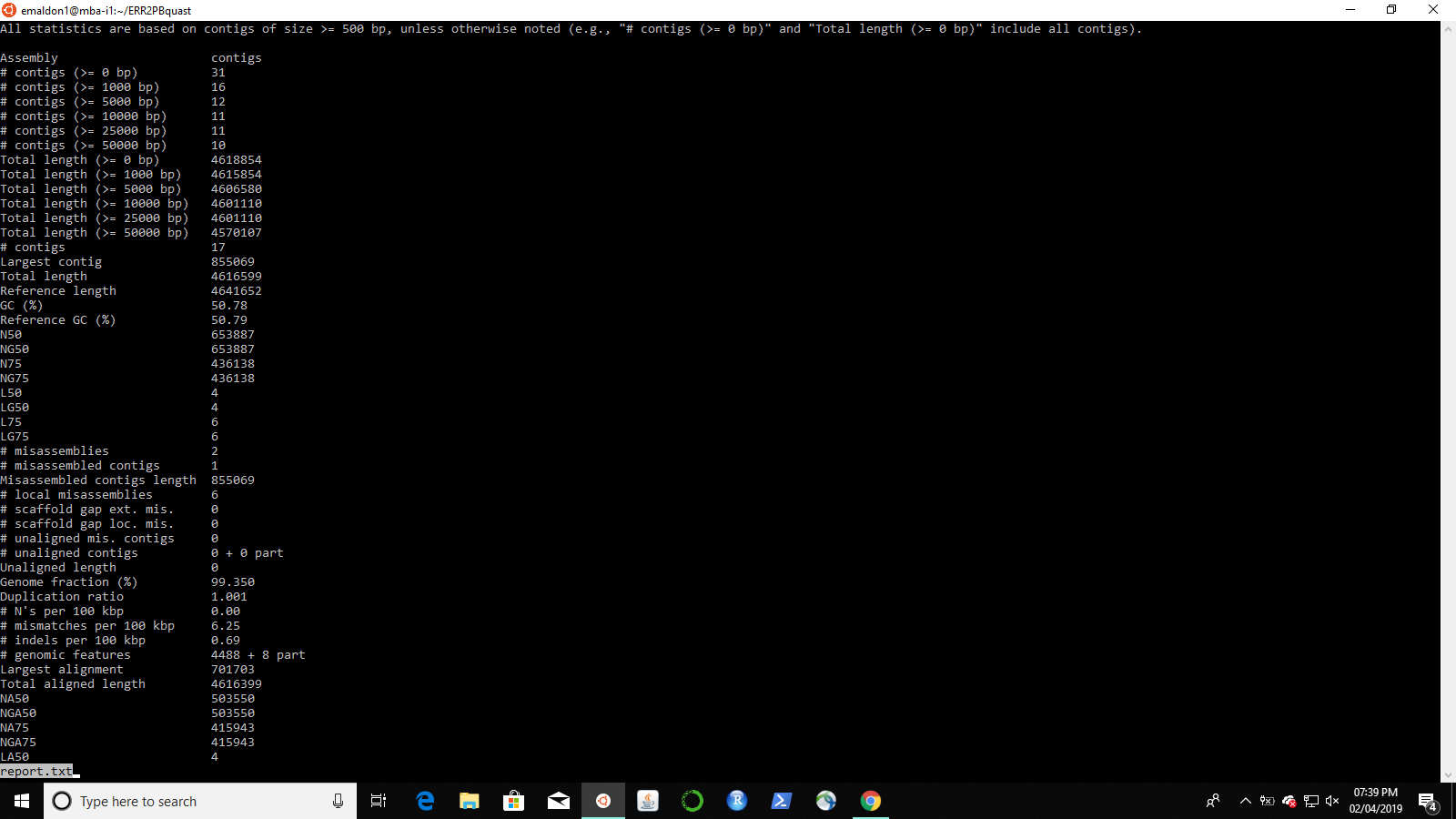
/users/emaldon1/quast-5.0.2/quast.py /users/emaldon1/BC30spades/contigs.fasta -R NC\_000913.fna -G NC\_000913.gff -o ERR2PBquast

qsub -N "ERR2PB" -q mamba spadesPacBiERR2.sh

cat ERR2PB.o1287

****

cd ERR2PBquast

less report.txt****