Lab 4

Jeffrey Miller

2022 Feb 24

render('lab-4-2023.md', output_format = "pdf_document")

Review Last Week Objectives:

Change working directories

```
cd gen711/shell_data/untrimmed_fastq
```

Do word find in files with grep

```
grep 'GNATNACCACTTCC' SRR098026.fastq
```

Use a single command to navigate multiple steps in your directory structure, including moving backwards (one level up).

```
cd ../
```

Employ navigational shortcuts to move around your file system. Having touble changing directories in terminal to your lab folder? Try:

```
cd ~
```

Exc. 3a "3 ways to change directories to HOME from untrimmed_fastq"
1. cd ~
2. cd ../../
3. cd /home/unhAW/jtmiller

Interconvert between absolute and relative paths.

```
ls gen711/shell_data/
realpath gen711
```

Work with hidden directories and hidden files and perform operations on files in directories outside your working directory

```
pwd
ls -a gen711/shell_data/
ls -a gen711/shell_data/.hidden
```

3b. How many programs in /bin 1. Start with the letter c 2. Start with the letter a 3. Start with the letter o

```
ls /bin/c*
```

Pipe is for making the output of one command go into another.

```
ls /bin/c* | wc -l
grep 'GNATNACCACTTCC' gen711/shell_data/untrimmed_fastq/SRR098026.fastq | wc -l
```

The '>' command is directing the output to a file

```
grep 'GNATNACCACTTCC' gen711/shell_data/untrimmed_fastq/SRR098026.fastq > the readifound.match
grep -B1 -A2 NNNNNNNNNN gen711/shell_data/untrimmed_fastq/*.fastq > bad_reads.fastq
```

To open a file, we used head, tail, and catas well as a program called 'less'

```
cat bad_reads.fastq
less bad_reads.fastq
```

New material

Make a new directory for new project files, change directories to there, and then use a command called curl to download a fastq from the internet:

```
cd ~
mkdir -p ~/dc_workshop/data/untrimmed_fastq/
cd ~/dc_workshop/data/untrimmed_fastq
curl -O ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR258/003/SRR2584863/SRR2584863_2.fastq.gz
```

The data comes in a compressed format, which is why there is a .gz at the end of the file names. This makes it faster to transfer, and allows it to take up less space on our computer. Let's unzip one of the files so that we can look at the fastq format.

```
gunzip SRR2584863_1.fastq.gz
ls
head -n 4 SRR2584863_1.fastq
```

Quality encoding: !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJ | | | | | Quality score: 01……..11……..21……..31……..41

Exercise: What is the last read in the SRR2584863_1.fastq file? How confident are you in this read?

Lets copy the rest of the files from tmp on RON cp /tmp/*fastq.gz .

Q:How big are the files? (Hint: Look at the options for the ls command to see how to show file sizes.) hint, it involves 'ls'

Lets activate the environment. A conda environment is ...

```
conda activate genomics
```

At this point, lets validate that all the relevant tools are installed. Test with

```
fastqc -h
```

FastQC can accept multiple file names as input, and on both zipped and unzipped files, so we can use the *.fastq* wildcard to run FastQC on all of the FASTQ files in this directory

```
fastqc *.fastq*
```

After that has finished, see the files and size:

```
ls -lh
```

Lets organize these files a bit. Sep. the fastqc files from the data by making a directory and moving the fastq files into it.

```
mkdir -p ~/dc_workshop/results/fastqc_untrimmed_reads
mv *.zip ~/dc_workshop/results/fastqc_untrimmed_reads/
mv *.html ~/dc_workshop/results/fastqc_untrimmed_reads/
```

lets head into the fastqc results folder and take a look into one of the summary files:

```
cd ~/dc_workshop/results/fastqc_untrimmed_reads/
```

The output files are zipped up. Lets see if we can unzip them with all in one command.

```
unzip *.zip
```

Then, we will use less to look a the summary

```
less SRR2584863_1_fastqc/summary.txt
```

Lastly, lets try to download one of those html files to our computer with sftp

```
realpath SRR2589044_1_fastqc.html
sftp jtmiller@ron.sr.unh.edu
get
```

Q: Which samples failed at least one of FastQC's quality tests? What test(s) did those samples fail?

```
cat */summary.txt > ~/dc_workshop/docs/fastqc_summaries.txt
```