

Hands-on

Reads preprocessing

NGS Server

Putty 163.25.92.42 port:22

Username: student01~student030

Password: student01~student030

FASTQC

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

teacher@localhost:~\$

\$ fastqc -h

\$ mkdir -p www/test1

\$ fastqc -o www/test1 -f fastq -t 1 test1.fq

\$ mkdir -p www/test2

\$ fastqc -o www/test2 -f fastq -t 1 test2.fq

<http://163.25.92.42/~teacher/test1/>

<http://163.25.92.42/~teacher/test2/>

#**teacher** should be replaced by **student01~student030**

FASTX-Toolkit

http://hannonlab.cshl.edu/fastx_toolkit/

FASTQ-TO-FASTA

```
$ fastq_to_fasta -i test1.fq -o test1.fasta -v
```

```
$ fastq_to_fasta -i test2.fq -o test2.fasta -v
```

FASTX-Toolkit

FASTA/Q Clipper

```
$ fastx_clipper -a TCGTATGCCGTCTTCTGCTTG  
-l 15 -c -v -i test1.fasta  
-o test1_clean.fasta
```

Clipping Adapter: TCGTATGCCGTCTTCTGCTTG

Min. Length: 15

Non-Clipped reads - discarded.

Input: 998992 reads.

Output: 989581 reads.

discarded 6845 too-short reads.

discarded 1348 adapter-only reads.

discarded 1218 non-clipped reads.

discarded 0 N reads.

FASTX-Toolkit

FASTX Collapser

```
$ fastx_collapser -i test1_clean.fasta  
-o test1_clean_c.fasta -v
```

Input: 989581 sequences (representing 989581 reads)

Output: 84652 sequences (representing 989581 reads)

>1-70795

GGTAAAATGGCTGAGTGAAGCATTG

>2-52559

GGTAAAATGGCTGAGTGAAGCATT

>3-45500

TGAGGTAGTAGGTTGTATAGTT

FASTX-Toolkit

FASTA Formatter

```
$ fasta_formatter -i test1_clean_c.fasta -o  
test1.tag -t
```

```
1-70795  GGTAATGGCTGAGTGAAGCATTG  
2-52559  GGTAATGGCTGAGTGAAGCATT  
3-45500  TGAGGTAGTAGGTTGTATAGTT
```

```
fasta_formatter -i test1_clean_c.fasta -t | cut -d-  
-f 2 > test1.txt
```

```
70795  GGTAATGGCTGAGTGAAGCATTG  
52559  GGTAATGGCTGAGTGAAGCATT  
45500  TGAGGTAGTAGGTTGTATAGTT
```