QC with fastqc

We'll use FastQC from Babraham Bioinformatics.

A guided example: 454

First we'll make a directory where to save the output of the QC:

```
mkdir ~/web/454_QC/
```

Remember that tilda (~) is the shortcut for your **home directory**. If your username is bsb99 then you'll create a /home/bsb99/web/454_QC directory. Easy?

Now we can run FASTQC:

```
/bsb/bin/fastqc -o ~/web/454_QC/ /bsb/denovo/datasets/454/SRP001673.fastq
```

Since we saved the output in the special directory “web”, we can access it from the internet (changing `bsb07` with your username). Visit the page:

1. http://my.seq.space/public/researcher/bsb07/454_QC/

Then click on the HTML file to view it.

Average quality

What do you think about the quality boxplots across the read length?

Adaptor

Check the “Adaptors” section: do you think it worked as expected on our dataset?

k-mer distribution

What about over represented k-mers?

Try yourself: Illumina MiSeq

Illumina MiSeq data is very common for bacterial genomes. Try performing the QC, saving it to a folder like `~/web/MiSeq_QC` so that you'll be able to actually see the output from the web. Being paired ends, you'll find two files, not just one:

1. /bsb/denovo/datasets/illumina/DRR075740_1.fastq
2. /bsb/denovo/datasets/illumina/DRR075740_2.fastq