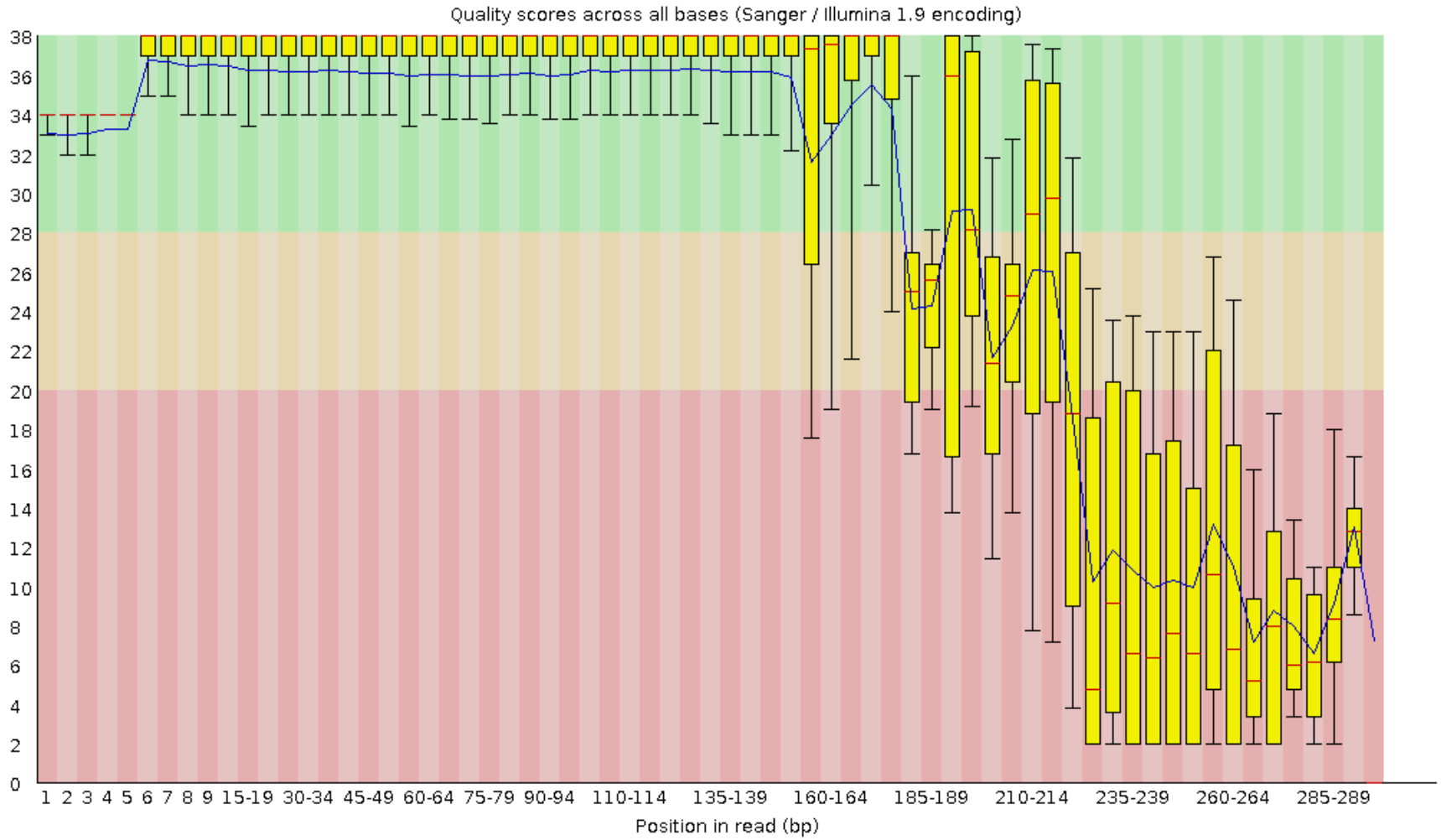
NGS sample QC

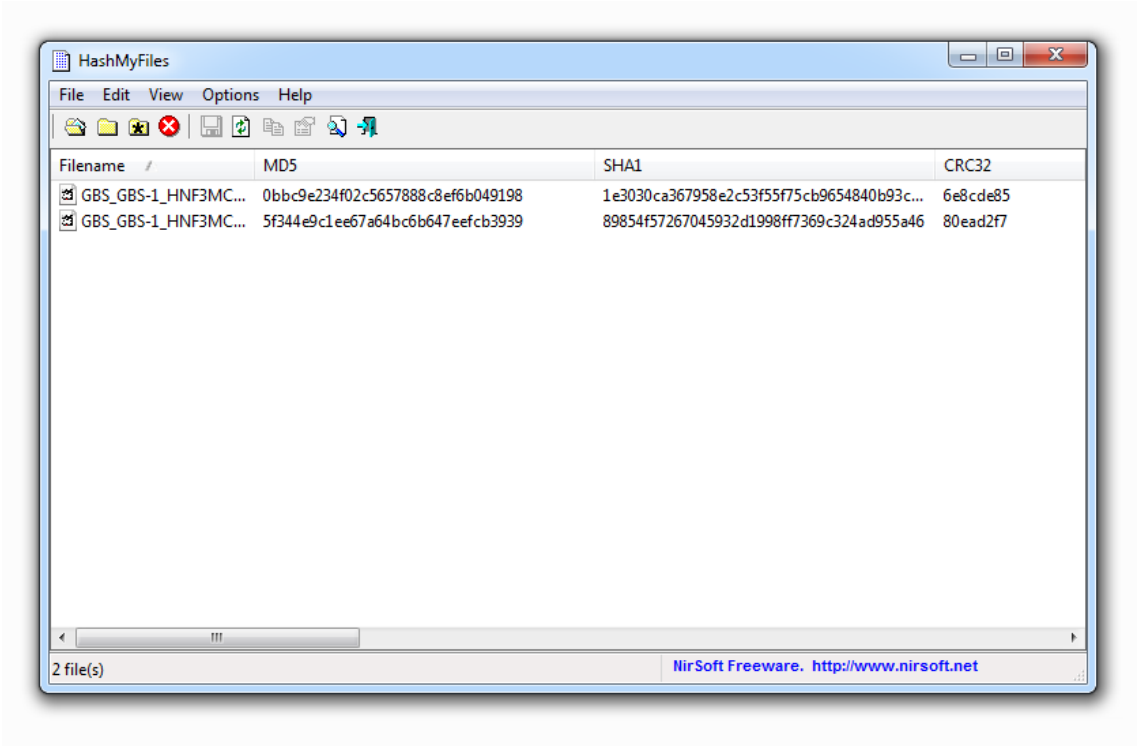
Per base sequence quality



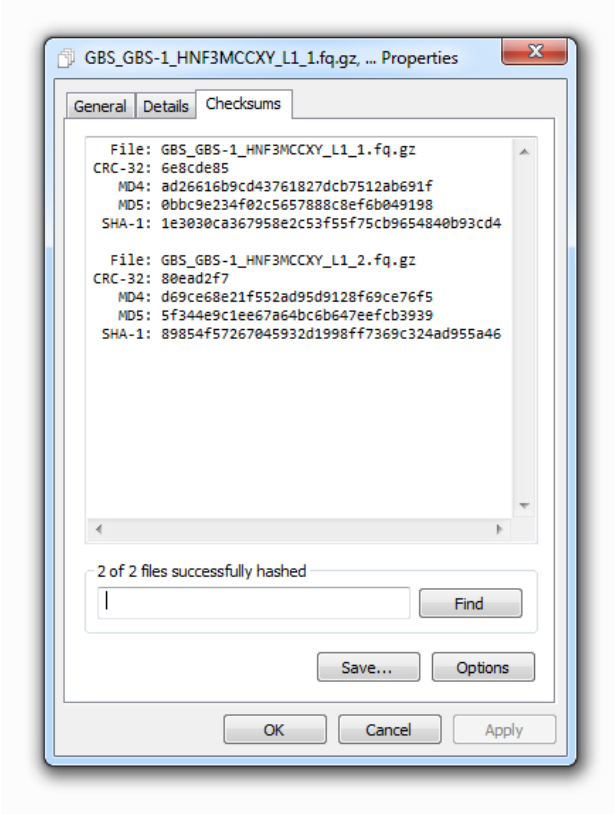
ID337_M11_TKD180701028-3_HNF5NCCXY_L4_1.fq.gz	21/07/2018 2:43 PM	GZ File	4,334,625 KB
ID337_M11_TKD180701028-3_HNF5NCCXY_L4_2.fq.gz	21/07/2018 2:43 PM	GZ File	4,970,835 KB
MD5.txt	24/07/2018 7:11 PM	TXT File	1 KB

FileZilla The free FTP solution

ID337_M11_TKD180701028-3_HNF5NCCXY_L4_1.fq.gz	21/07/2018 2:43 PM	GZ File	4,334,625 KB
ID337_M11_TKD180701028-3_HNF5NCCXY_L4_2.fq.gz	21/07/2018 2:43 PM	GZ File	4,970,835 KB
MD5.txt	24/07/2018 7:11 PM	TXT File	1 KB



https://www.nirsoft.net/utils/hash_my_files.html



http://implbits.com/products/hashtab/

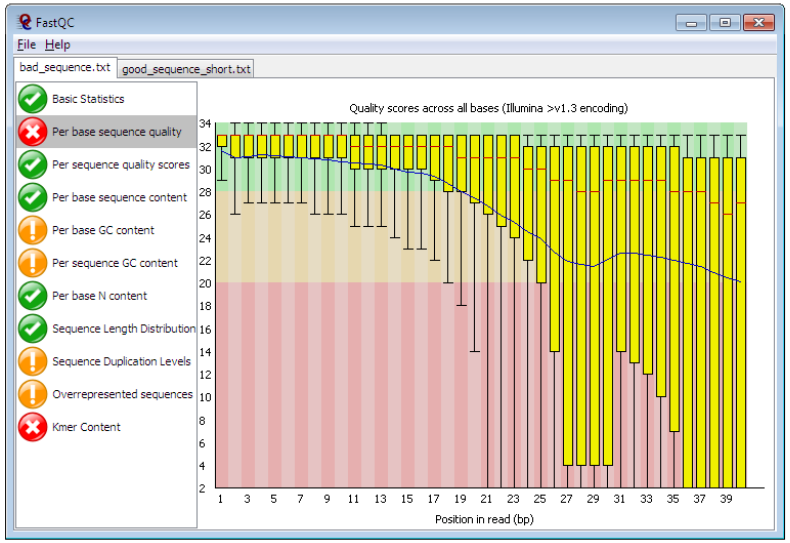
md5sum mysequencing_1.fq.gz

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

FastQC

Function	A quality control tool for high throughput sequence data.
Language	Java
Requirements	A suitable Java Runtime Environment The Picard BAM/SAM Libraries (included in download)
Code Maturity	Stable. Mature code, but feedback is appreciated.
Code Released	Yes, under GPL v3 or later.
Initial Contact	Simon Andrews

Download Now



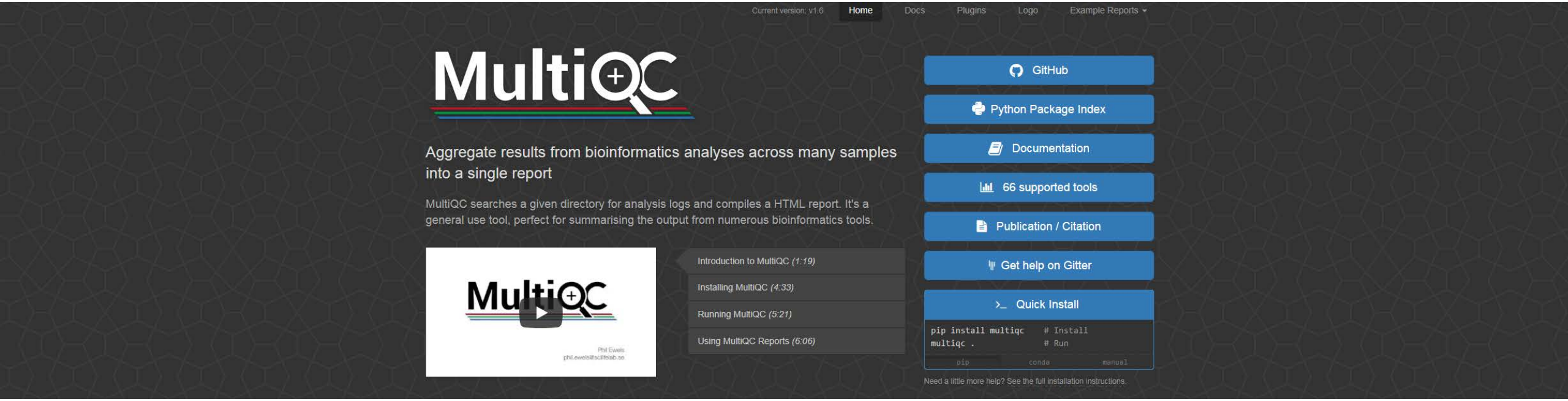
https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/

FastQ Screen

Function	FastQ Screen allows you to screen a library of sequences in FastQ format against a set of sequence databases so you can see if the composition of the library matches with what you expect.
Language	Perl
Requirements	Linux-based operating system Bowtie or Bowtie2 or BWA gzip (optional) SAMtools (optional) GD::Graph (optional) Bismark (bisulfite mapping only)
Code Maturity	Stable - has been working in production for some time
Code Released	Yes, under GPL v3 or later.
Documention	Online Documentation here
Initial Contact	Steven Wingett
	Download Now

https://sequencing.qcfail.com/

http://multiqc.info/



multiqc .

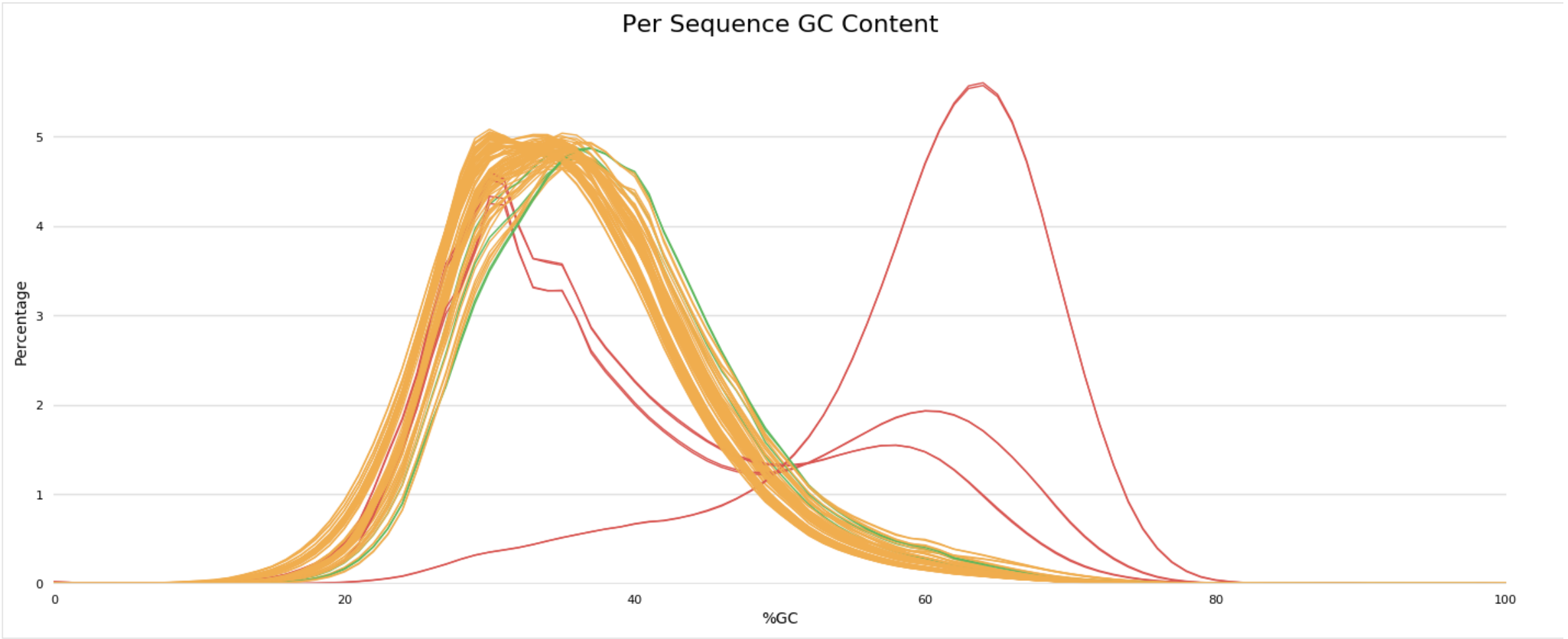
Per Sequence GC Content

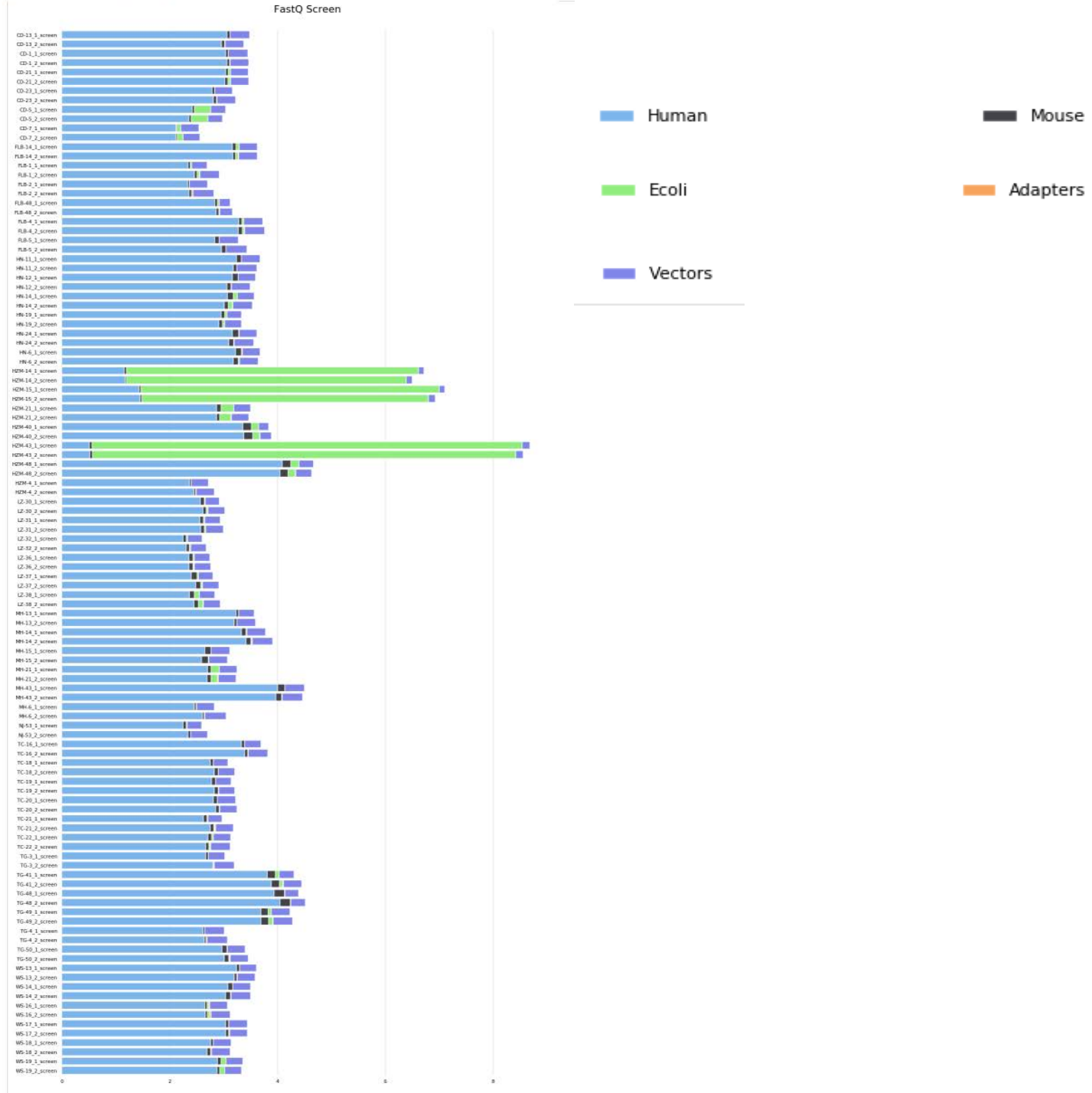
102

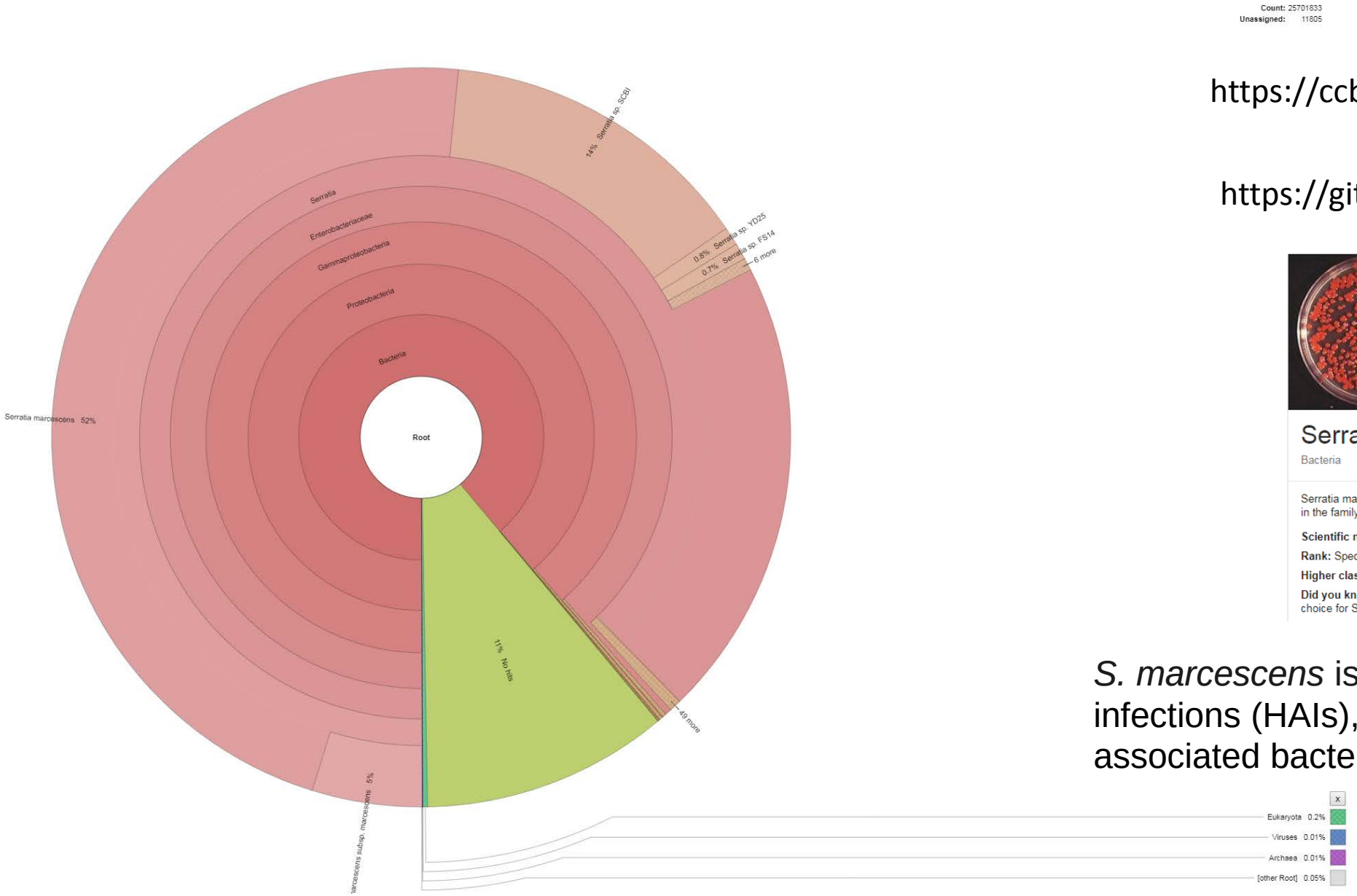
The average GC content of reads. Normal random library typically have a roughly normal distribution of GC content. See the FastQC help.

Flat image plot. Toolbox functions such as highlighting / hiding samples will not work (see the docs).

Percentages | Counts

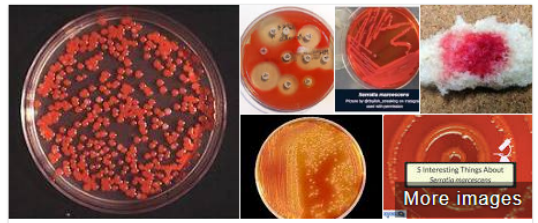



FastQ Screen
Human
Mouse
Ecoli
Adapters
Vectors



https://ccb.jhu.edu/software/kraken/

https://github.com/marbl/Krona/wiki



Serratia marcescens

Bacteria

Serratia marcescens is a species of rod-shaped gram-negative bacteria in the family Enterobacteriaceae. Wikipedia

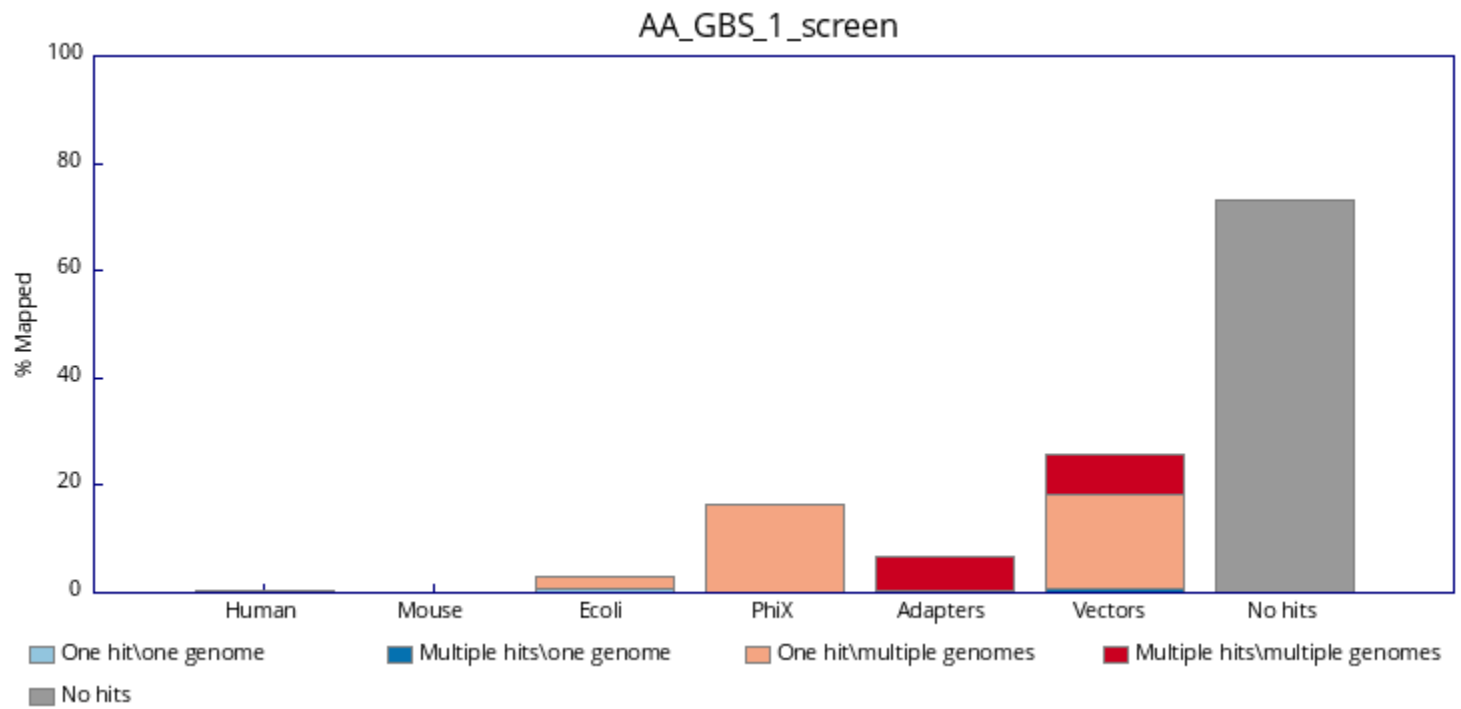
Scientific name: Serratia marcescens

Rank: Species

Higher classification: Serratia

Did you know: There is no antibiotic that is widely considered the drug of choice for S. marcescens. idstewardship.com

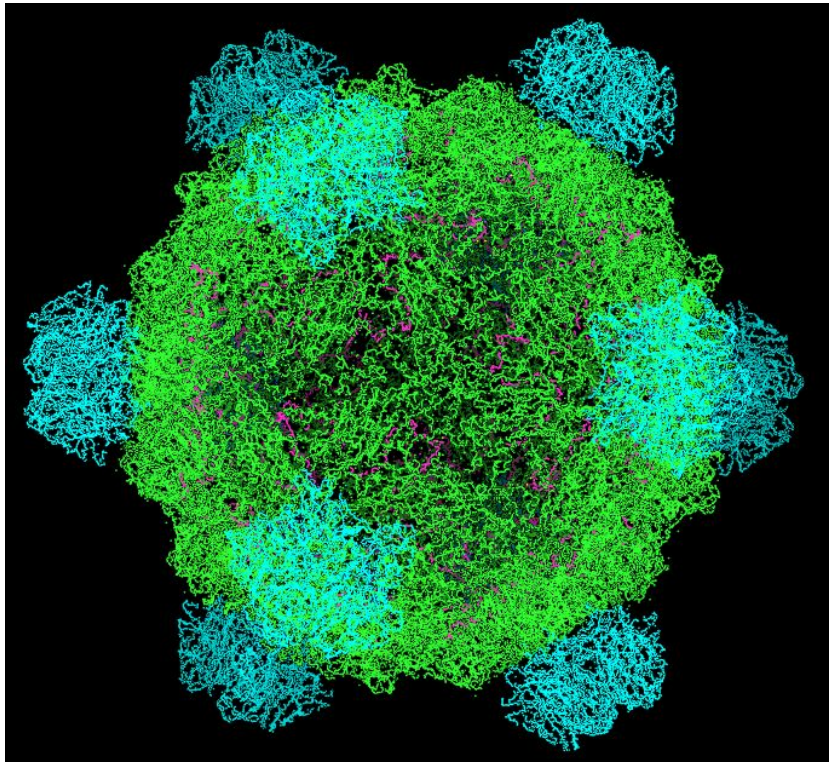
S. marcescens is involved in hospital-acquired infections (HAIs), particularly catheter-associated bacteremia


AA_GBS_1_screen
100
80
60
40
20
0
% Mapped
Human
Mouse
Ecoli
PhiX
Adapters
Vectors
No hits
One hit\one genome
Multiple hits\one genome
One hit\multiple genomes
Multiple hits\multiple genomes
No hits

What is PhiX??

Phi X 174
ΦX174

First DNA based genome to be sequenced
in 1977 by Fred Sanger

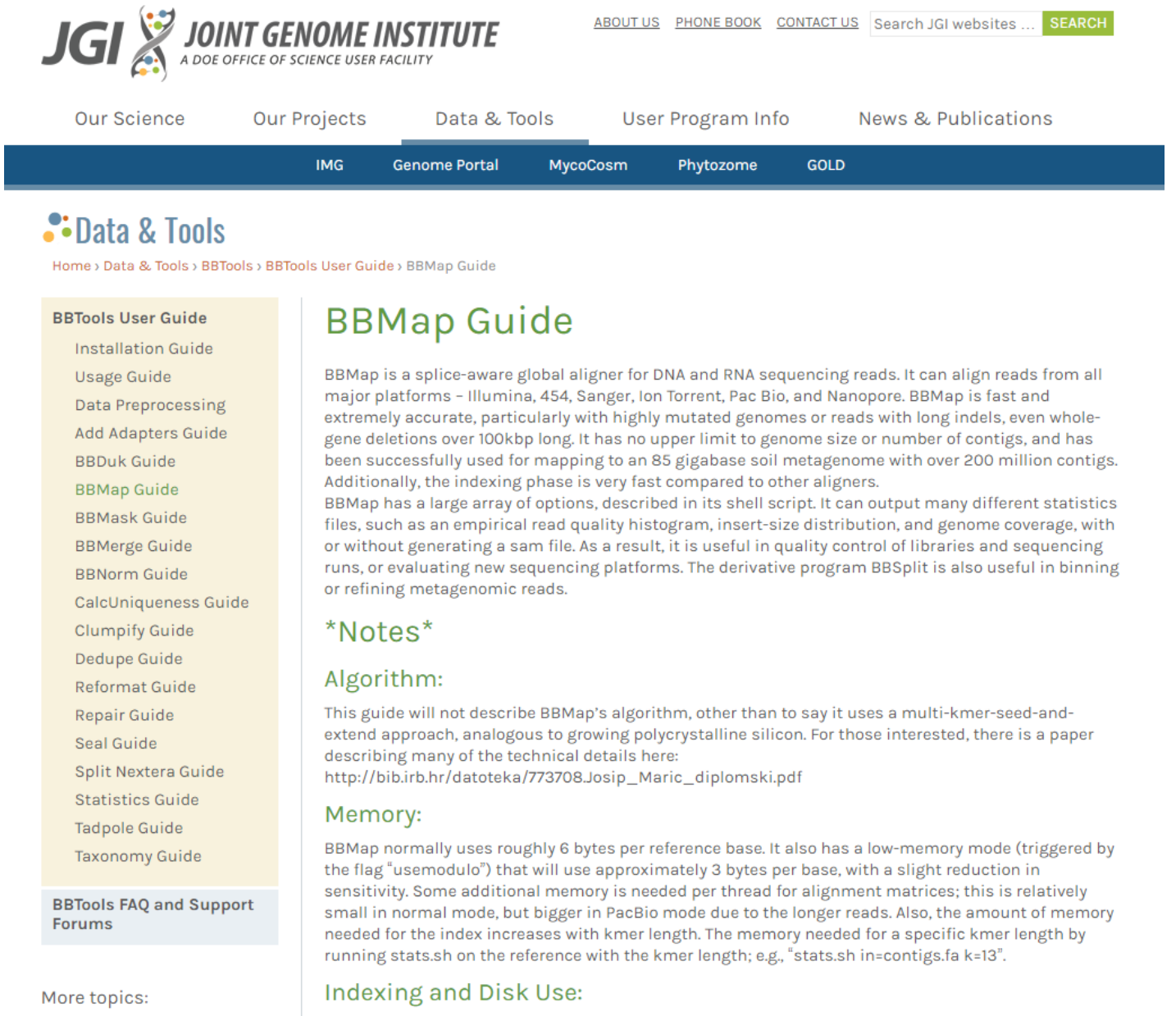


Control for illumina sequencers

Used for base quality calibration

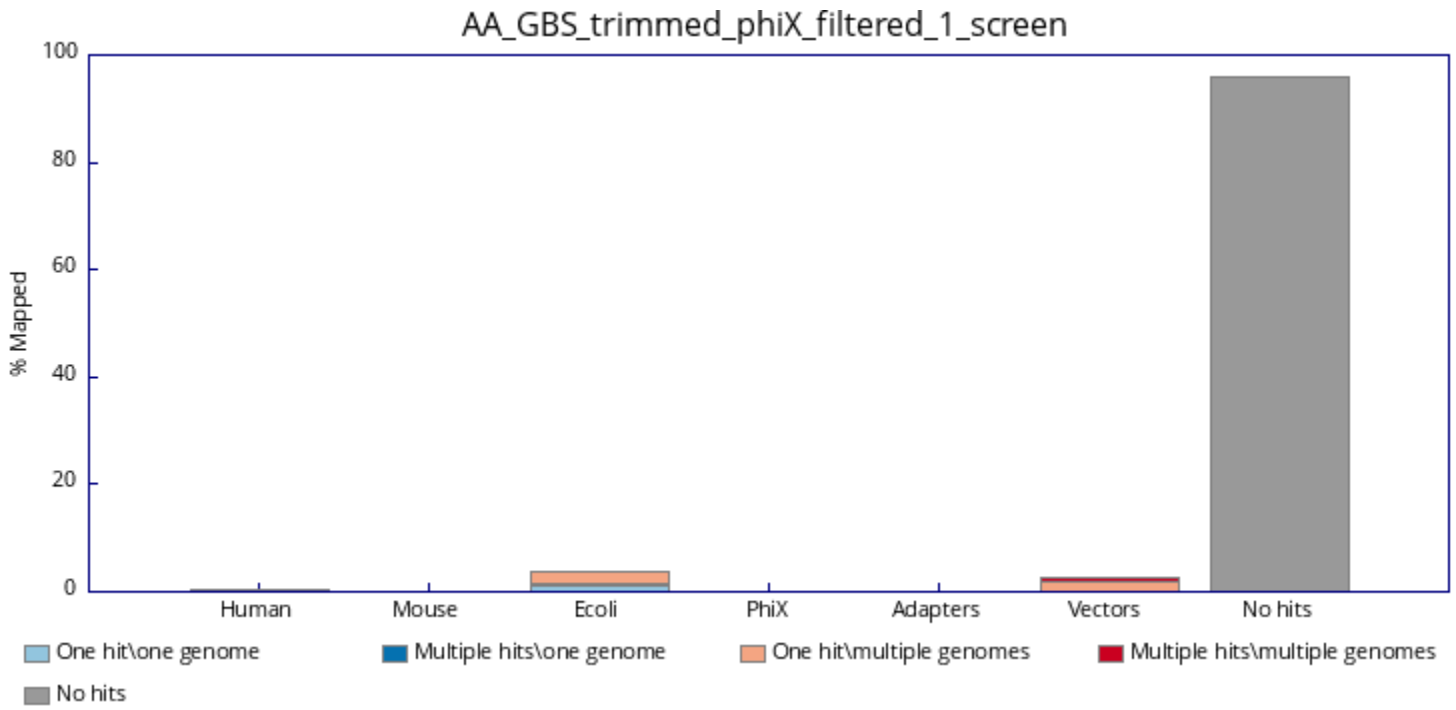
Sometimes added as spike-in for low diversity libraries

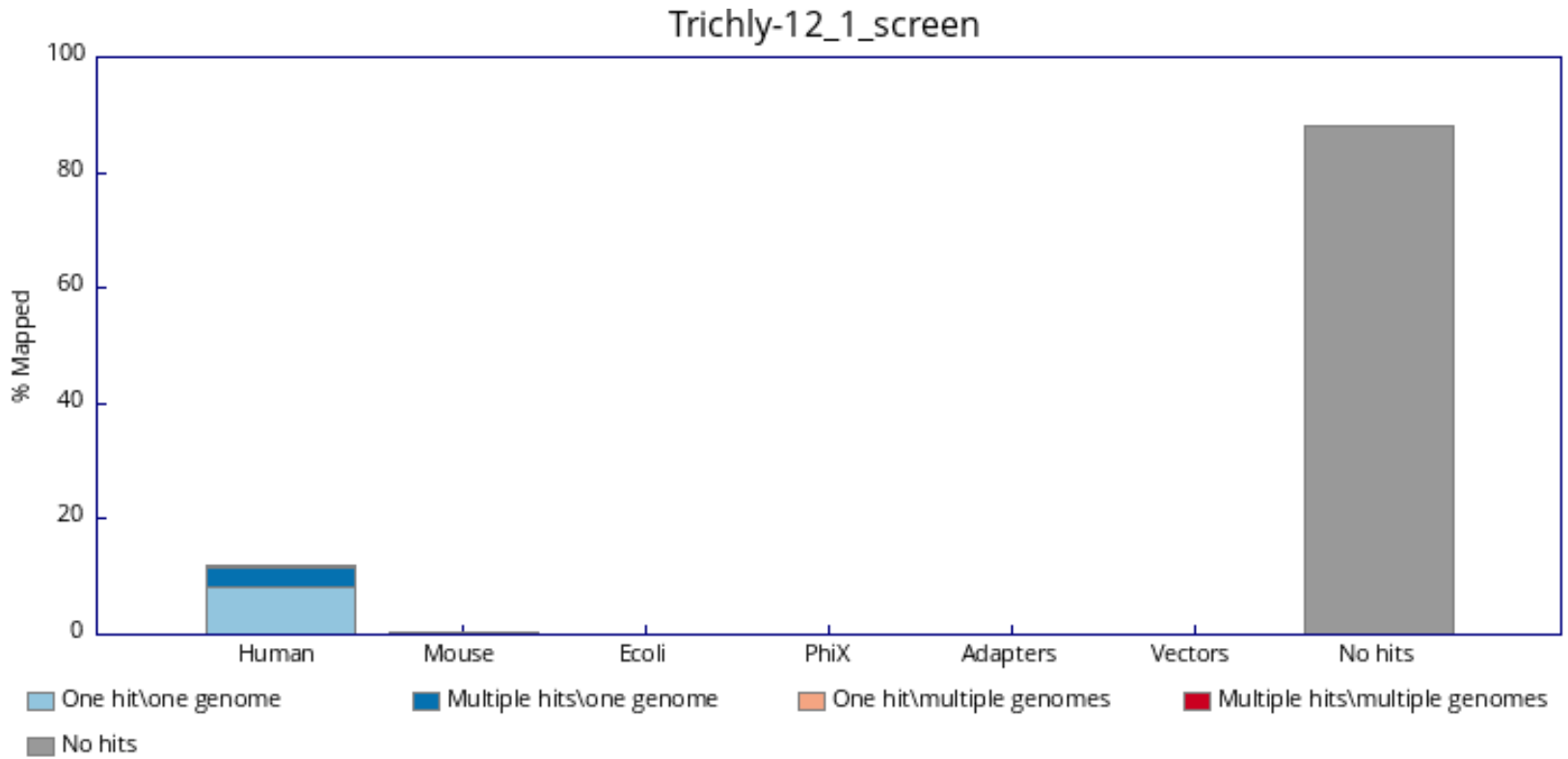
https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/



bbduk.sh in1=AA_GBS_1.fastq.gz out1=AA_GBS_trimmed_1.fq.gz minlen=40 ktrim=r k=23 mink=11 hdist=1 tbo tpe ref=/home/wfl/programs/bbmap/resources/adapters.fa maxns=0 qtrim=r trimq=10

bbduk.sh in1=AA_GBS_trimmed_1.fq.gz out1=AA_GBS_trimmed_phiX_filtered_1.fq.gz ref=/home/wfl/programs/bbmap/resources/phix174_ill.ref.fa.gz,/home/wfl/programs/bbmap/resources/sequencing_artifacts.fa.gz k=31




Trichly-12_1_screen
100
80
60
40
20
0
% Mapped
Human
Mouse
Ecoli
PhiX
Adapters
Vectors
No hits
One hit\one genome
Multiple hits\one genome
One hit\multiple genomes
Multiple hits\multiple genomes
No hits

Remove Human – from Brian Bushnell

http://seqanswers.com/forums/archive/index.php/t-42552.html

https://drive.google.com/file/d/0B3llHR93L14wd0pSSnFULUlhcUk/edit

hg19_main_mask_ribo_animal_allplant_allfungus.fa.gz

```
bbmap.sh minid=0.95 maxindel=3 bwr=0.16 bw=12 quickmatch fast minhits=2
path=/path/to/hg19masked/ qtrim=rl trimq=10 untrim -Xmx23g in=reads.fq outu=clean.fq
outm=human.fq
```

Activities Firefox Web Browser Wed 12:00

MultiQC Report - Mozilla Firefox

MultiQC Report | MultiQC Report | Krona - Root | MultiQC Report | Stacks: Stacks Manual | MultiQC Report

file:///media/wfl/6TB_RAID/SowThistle/STI6_600bp/multiqc_report.html

Search

MultiQC v1.0.dev0

- General Stats
- FastQ Screen
- FastQC
  - Sequence Quality Histograms
  - Per Sequence Quality Scores
  - Per Base Sequence Content
  - Per Sequence GC Content
  - Per Base N Content
  - Sequence Length Distribution
  - Sequence Duplication Levels
  - Overrepresented sequences
  - Adapter Content

Welcome! Not sure where to start? Watch a tutorial video (6:06) don't show again

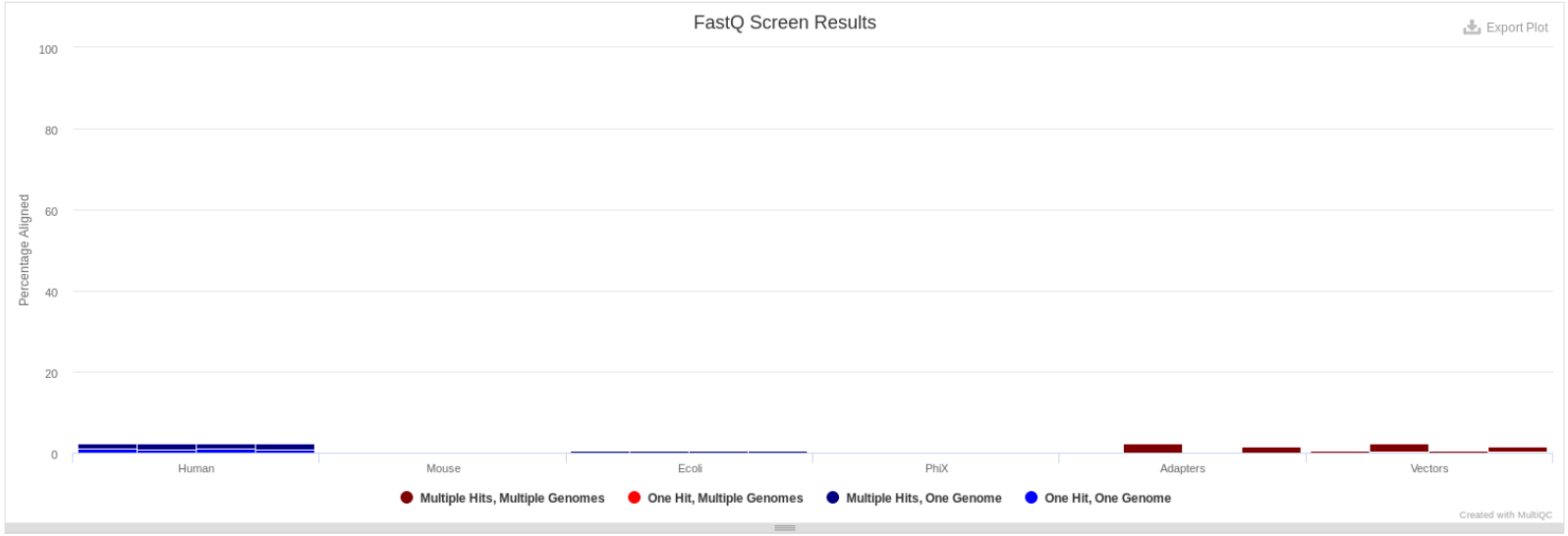
General Statistics

Copy table | Configure Columns | Plot | Showing 4/4 rows and 3/5 columns.

Sample Name	% Dups	% GC	M Seqs
STI6_600bp_L7_1	30.6%	37%	118.7
STI6_600bp_L7_2	27.3%	37%	118.7
STI6_600bp_L8_1	27.2%	37%	103.0
STI6_600bp_L8_2	27.2%	37%	103.0

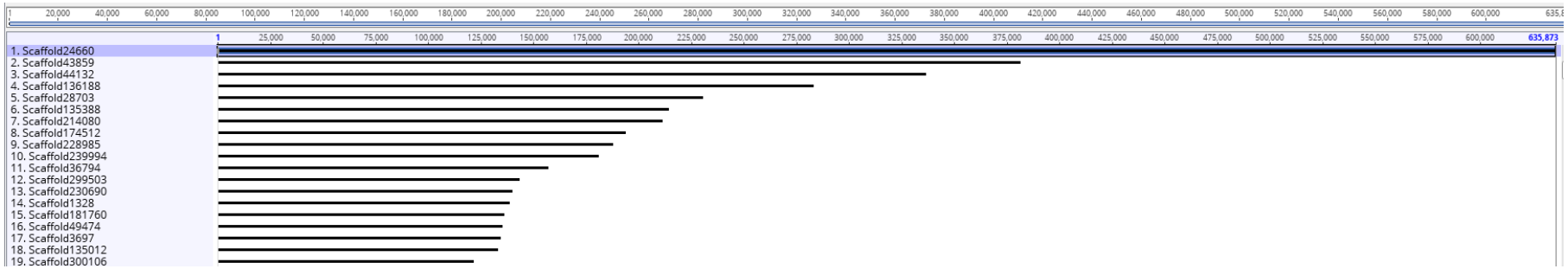
FastQ Screen

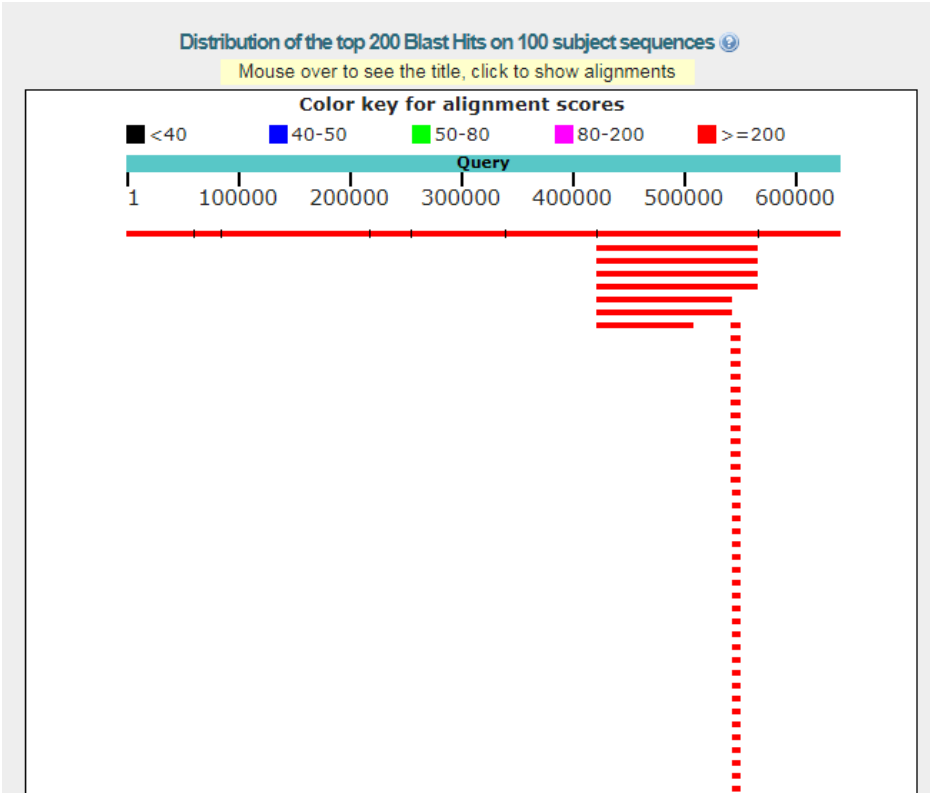
FastQ Screen allows you to screen a library of sequences in FastQ format against a set of sequence databases so you can see if the composition of the library matches with what you expect.

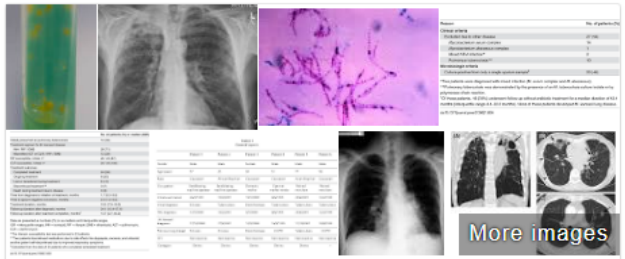


FastQC

FastQC is a quality control tool for high throughput sequence data, written by Simon Andrews at the Babraham Institute in Cambridge.







Mycobacterium kansasii

Bacteria

Mycobacterium kansasii is a bacterium in the Mycobacterium family. The genus includes species known to cause serious diseases in mammals, including tuberculosis and leprosy, but this species is generally not dangerous to healthy people. Wikipedia

Scientific name: Mycobacterium kansasii

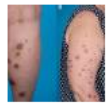
Phylum: Actinobacteria

Higher classification: Mycobacterium

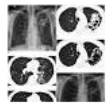
Order: Actinomycetales

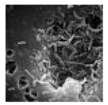
Rank: Species

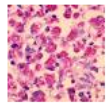
People also search for


Mycobact... chelonae


Bacteria

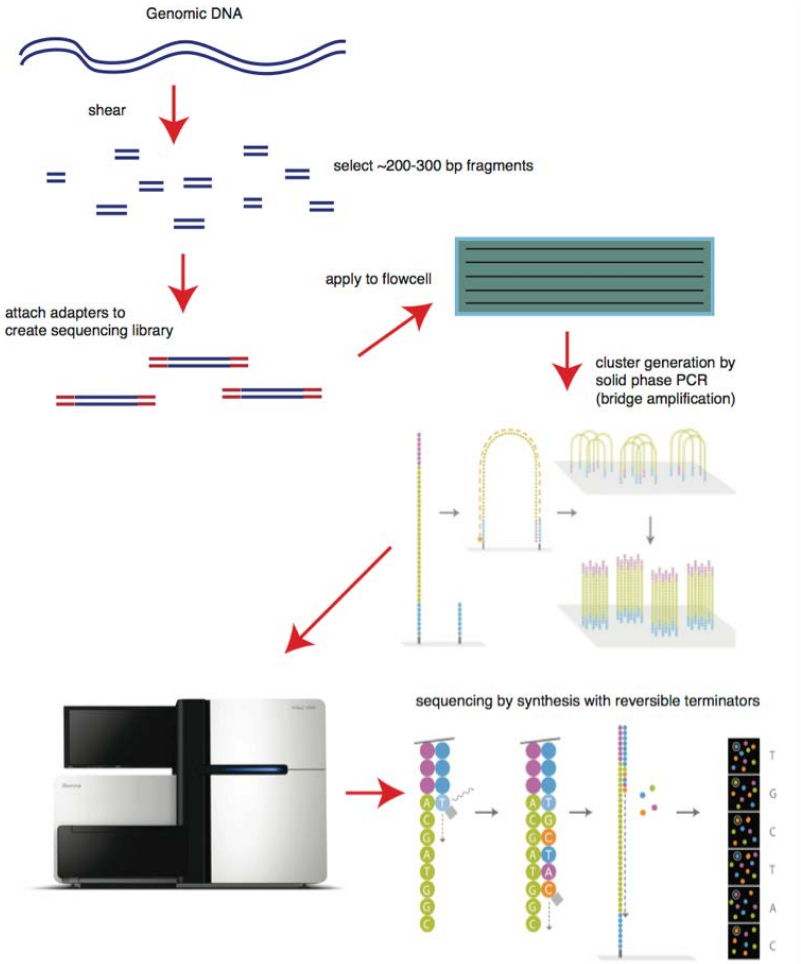

Mycobact... szulgai


Mycobact... malmoense


Mycobact... avium avium

Feedback

Adapter trimming



Adapter read-through

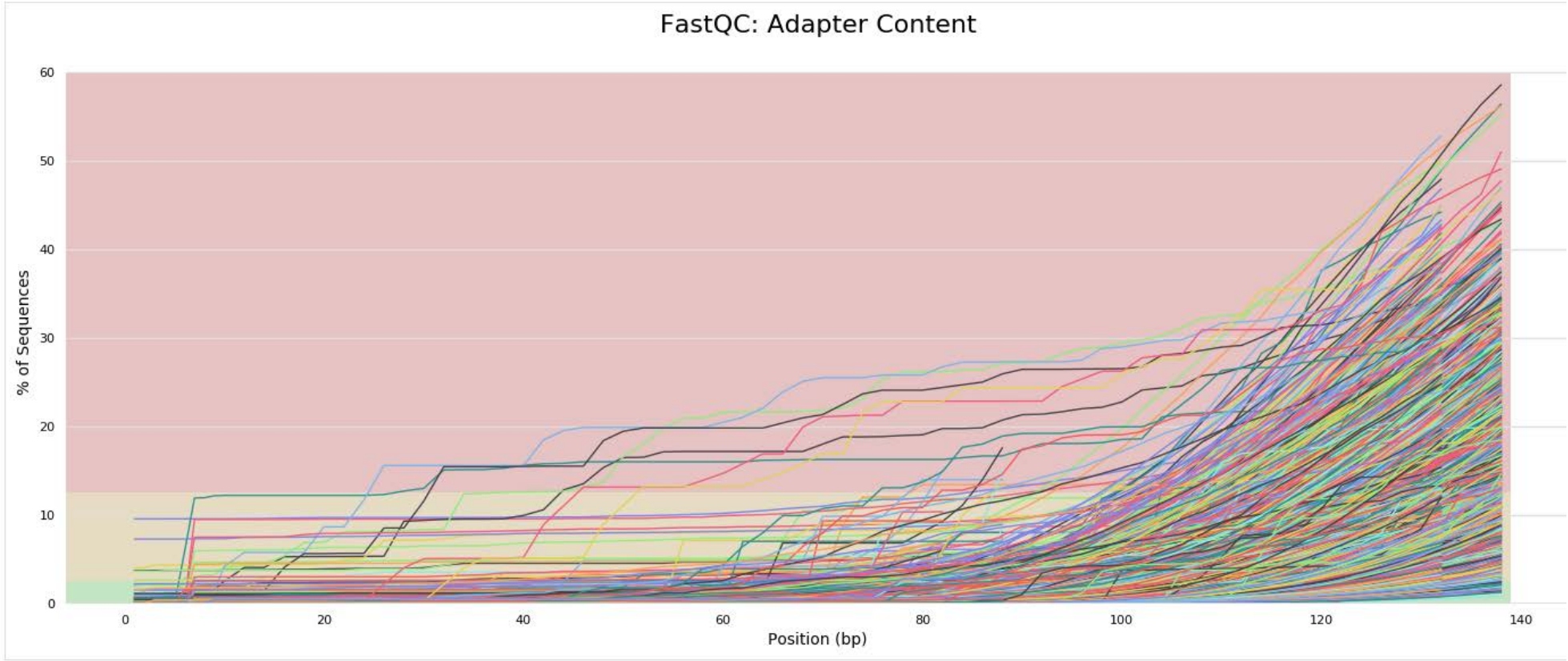
Before trimming:

Adapter Content

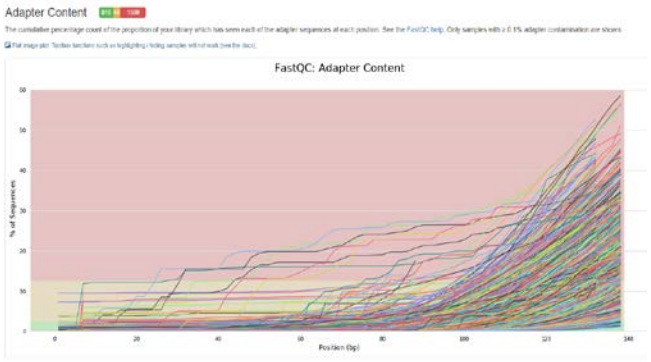
818 40 1500

The cumulative percentage count of the proportion of your library which has seen each of the adapter sequences at each position. See the FastQC help. Only samples with ≥ 0.1% adapter contamination are shown.

Flat image plot. Toolbox functions such as highlighting / hiding samples will not work (see the docs).

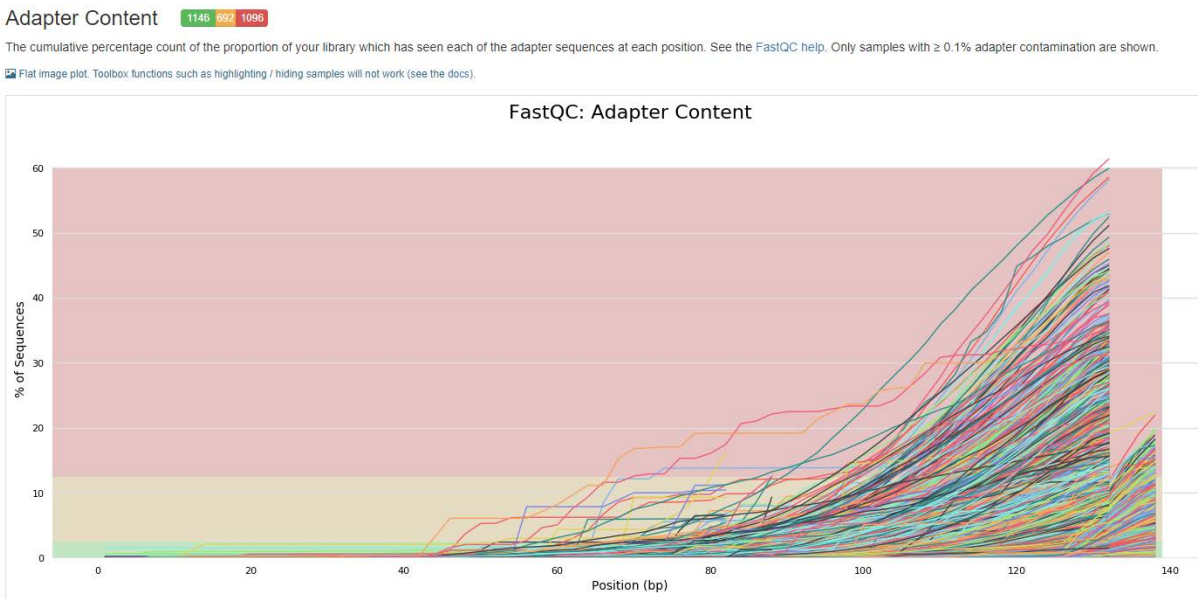


Before trimming:

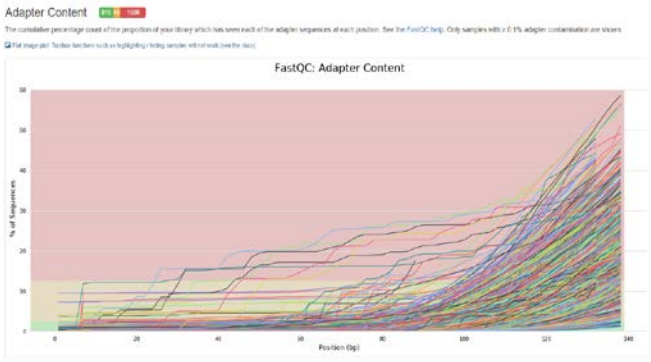


java -jar /mnt/d/EucMel/programs/Trimmomatic-0.36/trimmomatic-0.36.jar PE -threads 8 -phred33
/mnt/d/EucMel/DATA/AllDATA/Allosyncarpia_ternata_11530_R1.fastq
/mnt/d/EucMel/DATA/AllDATA/Allosyncarpia_ternata_11530_R2.fastq
/mnt/d/EucMel/DATA/trimmed/Allosyncarpia_ternata_11530_forward_paired.fq
/mnt/d/EucMel/DATA/trimmed/Allosyncarpia_ternata_11530_forward_unpaired.fq
/mnt/d/EucMel/DATA/trimmed/Allosyncarpia_ternata_11530_reverse_paired.fq
/mnt/d/EucMel/DATA/trimmed/Allosyncarpia_ternata_11530_reverse_unpaired.fq
ILLUMINACLIP:/mnt/d/EucMel/programs/Trimmomatic-0.36/adapters/TruSeq3-PE-2.fa:2:40:15 LEADING:3
TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50

Trimmomatic:

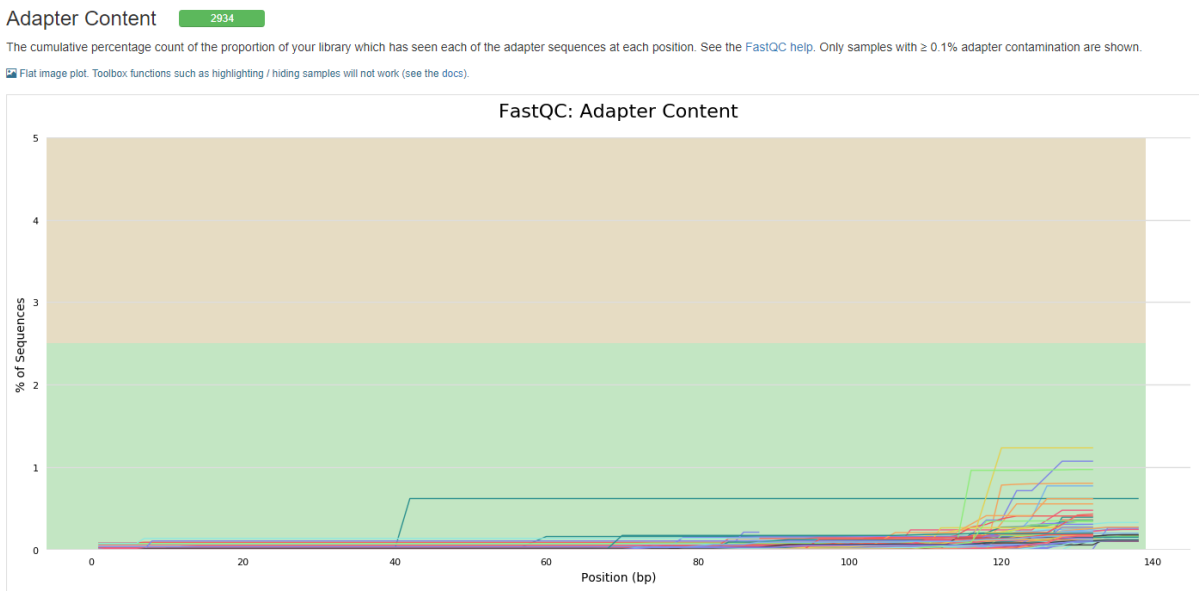


Before trimming:




```
bbduk.sh in1=/mnt/d/EucMel/DATA/AllDATA/".$samplefilename."_R1.fastq
in2=/mnt/d/EucMel/DATA/AllDATA/".$samplefilename."_R2.fastq
out1=/mnt/d/EucMel/DATA/bbtrimmed/".$samplefilename."_trimmed_1.fq
out2=/mnt/d/EucMel/DATA/bbtrimmed/".$samplefilename."_trimmed_2.fq
minlen=80 ktrim=r k=23 mink=8 hdist=1 tbo tpe
ref=/mnt/c/bbmap/resources/adapters.fa maxns=0 qtrim=r trimq=10
```

Bbduk:




QUALITY CONTROL
PASSED
INSPECTION