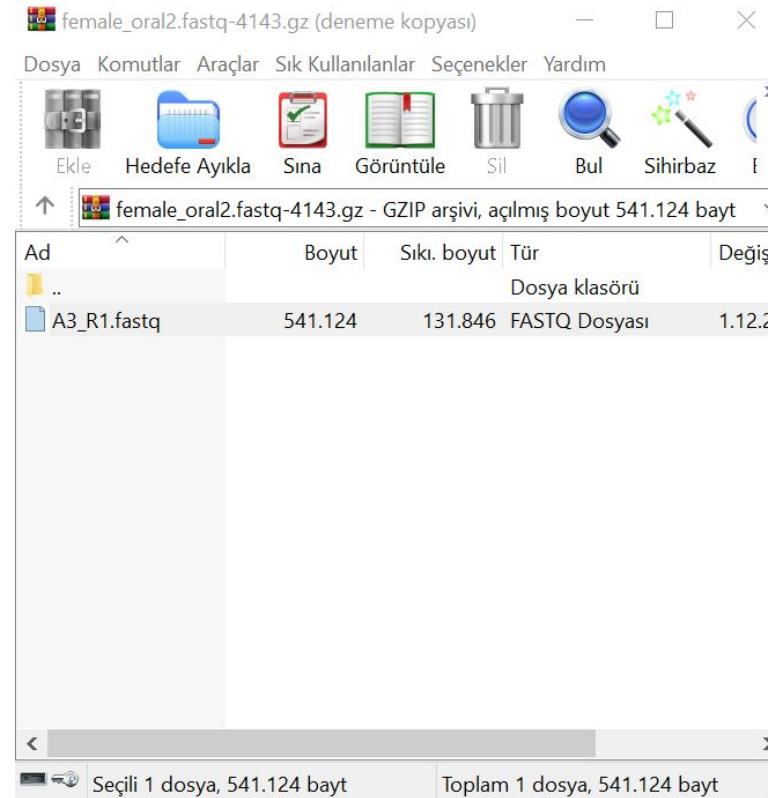


- Complete the Tutorial on Quality Control on the Galaxy platform:
<https://training.galaxyproject.org/training-material/topics/sequence-analysis/tutorials/quality-control/tutorial.html>
 - Share a screenshot for every step you take.
 - Tip: You can use arrows or other shapes to help you explain what you've done. You can use Google Slides or MS Powerpoint to help you with this.
 - Share screenshots of your FastQC results. Explain what each graph tells.
 - Make your Galaxy history visible and add a link to your Galaxy history at the end of the assignment.

downloading the data:



uploading the dowloaded data:

Download from web or upload from disk

Regular Composite Collection Rule-based

Please wait...1 out of 1 remaining.

Name	Size	Type	Genome	Settings	Status
 A3_R1.fastq	528.4 KB	fastq	unspecified (?)		Adding to history...

Type (set all):  Genome (set all):

 Choose local files  Choose remote files  Paste/Fetch data Start Pause Reset Close

renaming the dataset

Edit Dataset Attributes

Attributes

Convert

Datatypes

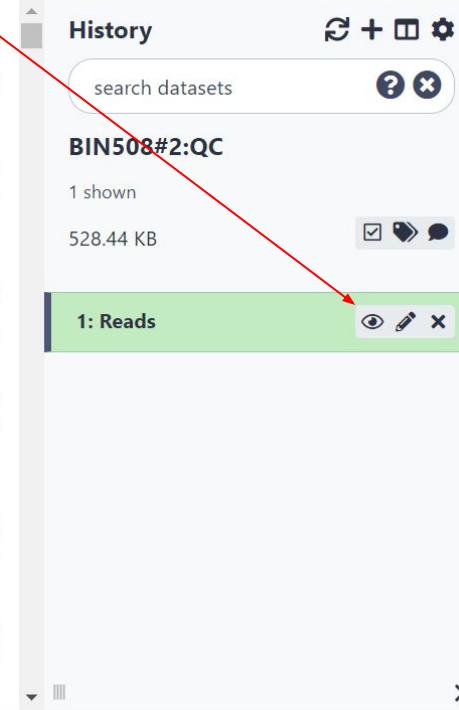
Permissions

Name

Reads

Info

inspecting the data by clicking the eye icon



Using FASTQ

Tools



search tools



Upload Data

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

Trimmomatic flexible read trimming
tool for Illumina NGS data

FASTQE visualize fastq files with
emoji's 🎉😎

PRINSEQ to process quality of
sequences

Draw quality score boxplot



FastQ data



1: Reads



Bin scores



No

Score types to show



Select/Unselect all

* Mean |

Maximum read length

500

Enable long reads up to this many bp long.

execution process:



Executed **FASTQE** and successfully added 1 job to the queue.

The tool uses this input:

- **1: Reads**

It produces this output:

- **2: FASTQE on data 1**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History



search datasets



BIN508#2:QC

2 shown

528.44 KB



2: FASTQE on data 1



1: Reads



dataset information:

Dataset Information

Number	2
Name	FASTQE on data 1
Created	Saturday Nov 6th 11:46:09 2021 UTC
Filesize	1.6 KB
Dbkey	?
Format	html
File contents	contents
History Content API ID	bbd44e69cb8906b5780a7a16891ce072
History API ID	2e47d762018d7067
UUID	84467c9b-7561-4335-89c1-7254df185821

Tool Parameters

Input Parameter	Value
FastQ data	1 Reads
Bin scores	False
Score types to show	Mean
Maximum read length	500

History

search datasets

BIN508#2:QC

2 shown

530.04 KB

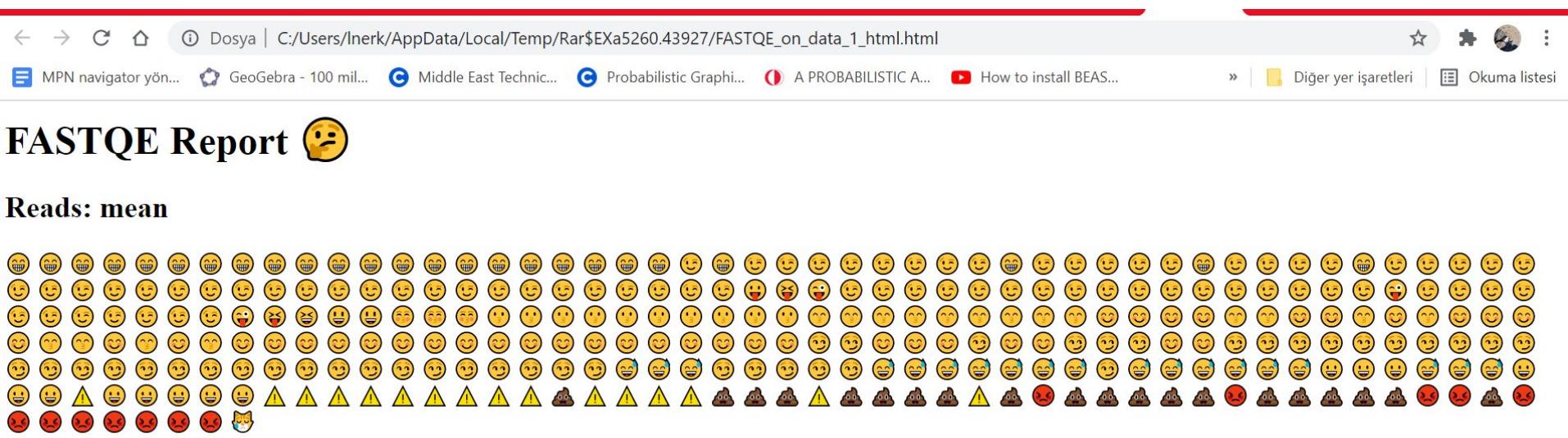
2: FASTQE on data 1

1.6 KB
format: html, database: ?

Download

1: Reads

html file of the FASTQE evaluation:





FastQC assesment:

Upload Data

Draw quality score boxplot

MultiQC aggregate results from bioinformatics analyses into a single report

FastQC Read Quality reports

FASTQ Summary Statistics by column

Compute quality statistics

Draw nucleotides distribution chart

SAM/BAM

BED

execution:



Executed **FastQC** and successfully added 1 job to the queue.

The tool uses this input:

- **1: Reads**

It produces 2 outputs:

- **3: FastQC on data 1: Webpage**
- **4: FastQC on data 1: RawData**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

txt output:

```
Galaxy4-[FastQC_on_data_1_RawData] - Not Defteri
Dosya Düzen Biçim Görünüm Yardım
##FastQC 0.11.8
>>Basic Statistics      pass
#Measure    Value
Filename   Reads
File type  Conventional base calls
Encoding   Sanger / Illumina 1.9
Total Sequences 812
Sequences flagged as poor quality      0
Sequence length 296
%GC      44
>>END_MODULE
>>Per base sequence quality      fail
#Base  Mean  Median  Lower Quartile  Upper Quartile  10th Percentile  90th Percentile
1     37.716748768472904  38.0  38.0  38.0  38.0  38.0
2     37.8423645320197  38.0  38.0  38.0  38.0  38.0
3     37.639162561576356  38.0  38.0  38.0  37.0  38.0
4     37.69458128078818  38.0  38.0  38.0  38.0  38.0
5     37.6884236453202  38.0  38.0  38.0  38.0  38.0
6     37.77093596059113  38.0  38.0  38.0  38.0  38.0
7     37.84605911330049  38.0  38.0  38.0  38.0  38.0
8     37.73891625615764  38.0  38.0  38.0  38.0  38.0
9     37.706896551724135  38.0  38.0  38.0  38.0  38.0
10-14 37.6243842364532  38.0  38.0  38.0  37.8  38.0
15-19 37.72364532019704  38.0  38.0  38.0  38.0  38.0
20-24 37.505665024630545  38.0  38.0  38.0  37.2  38.0
25-29 37.416009852216746  38.0  38.0  38.0  37.2  38.0
30-34 37.464532019704436  38.0  38.0  38.0  37.0  38.0
35-39 37.37389162561576  38.0  38.0  38.0  37.0  38.0
40-44 37.41674876847291  38.0  38.0  38.0  37.0  38.0
45-49 37.217487684729065  38.0  38.0  38.0  37.0  38.0
50-54 37.22463054187192  38.0  38.0  38.0  36.4  38.0
55-59 37.23251231527094  38.0  38.0  38.0  36.8  38.0
60-64 37.19088669950739  38.0  38.0  38.0  36.4  38.0
65-69 37.02857142857143  38.0  38.0  38.0  35.8  38.0
70-74 37.02857142857143  38.0  38.0  38.0  35.8  38.0
75-79 37.02857142857143  38.0  38.0  38.0  35.8  38.0
80-84 37.02857142857143  38.0  38.0  38.0  35.8  38.0
85-89 37.02857142857143  38.0  38.0  38.0  35.8  38.0
90-94 37.02857142857143  38.0  38.0  38.0  35.8  38.0
95-99 37.02857142857143  38.0  38.0  38.0  35.8  38.0
```

html output:

Sat 6 Nov 2021
Reads

FastQC Report

Summary

- ✓ Basic Statistics
- ✗ Per base sequence quality
- ✗ Per tile sequence quality
- ! Per sequence quality scores
- ✗ Per base sequence content
- ✗ Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences

Produced by FastQC (version 0.11.8)

Basic Statistics

Measure	Value
Filename	Reads
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	812
Sequences flagged as poor quality	0
Sequence length	296
%GC	44

Galaxy4-[FastQC_o...txt] FASTQE_on_data_1.zip female_oral2.fastq....gz

Tümünü göster

per-base sequence quality

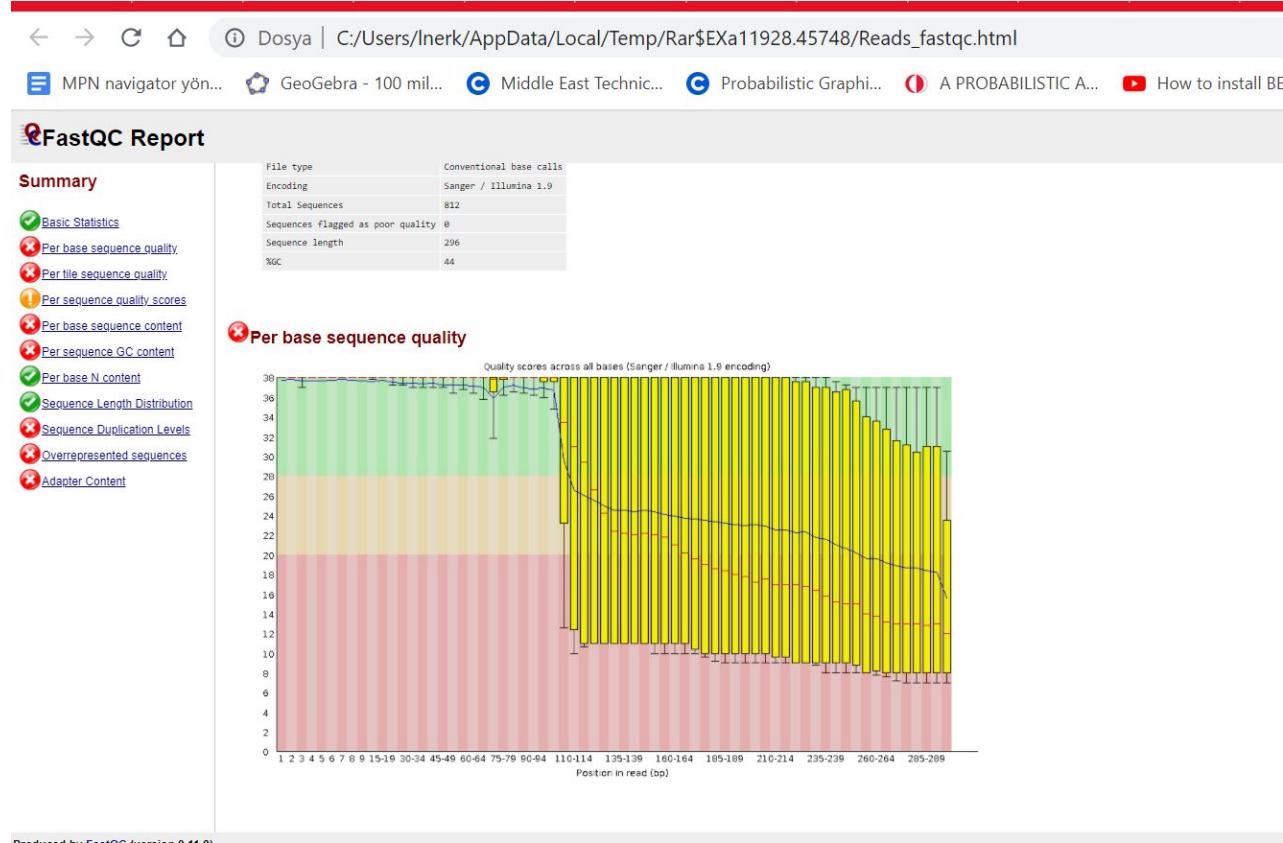
this graph represents
the range of the quality
as phred scores.
y-axis==quality scores→
higher scores == better

in this graph:

very good==green

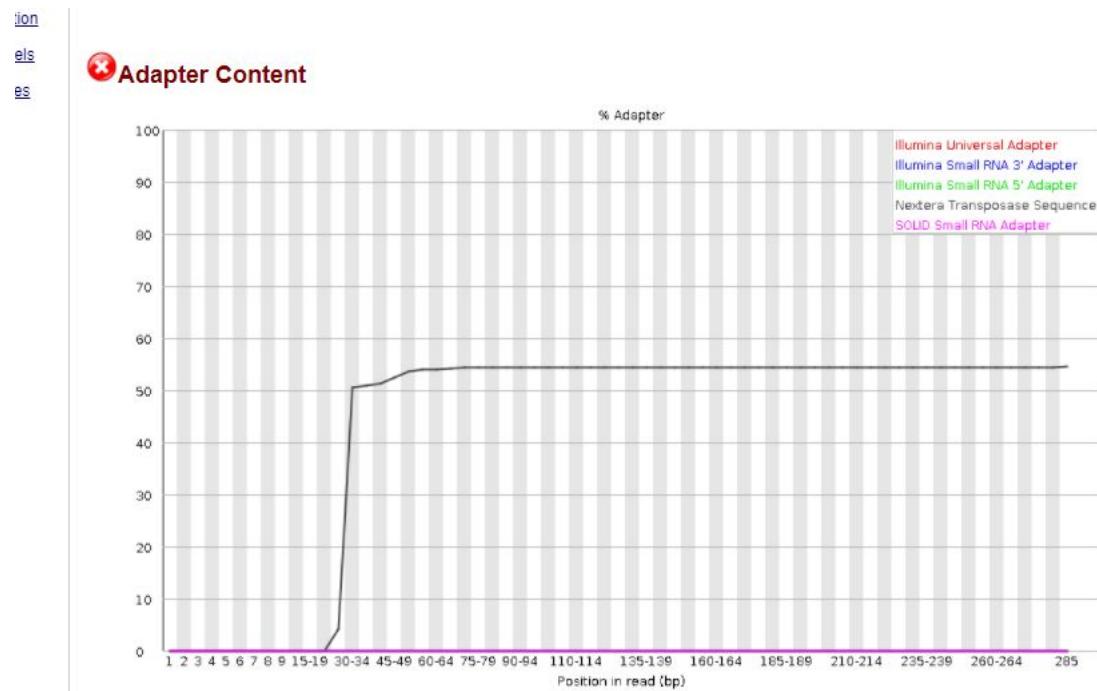
reasonable=blue

poor=red



adapter content:

cumulative (%) of reads in adapter seqs in each position.



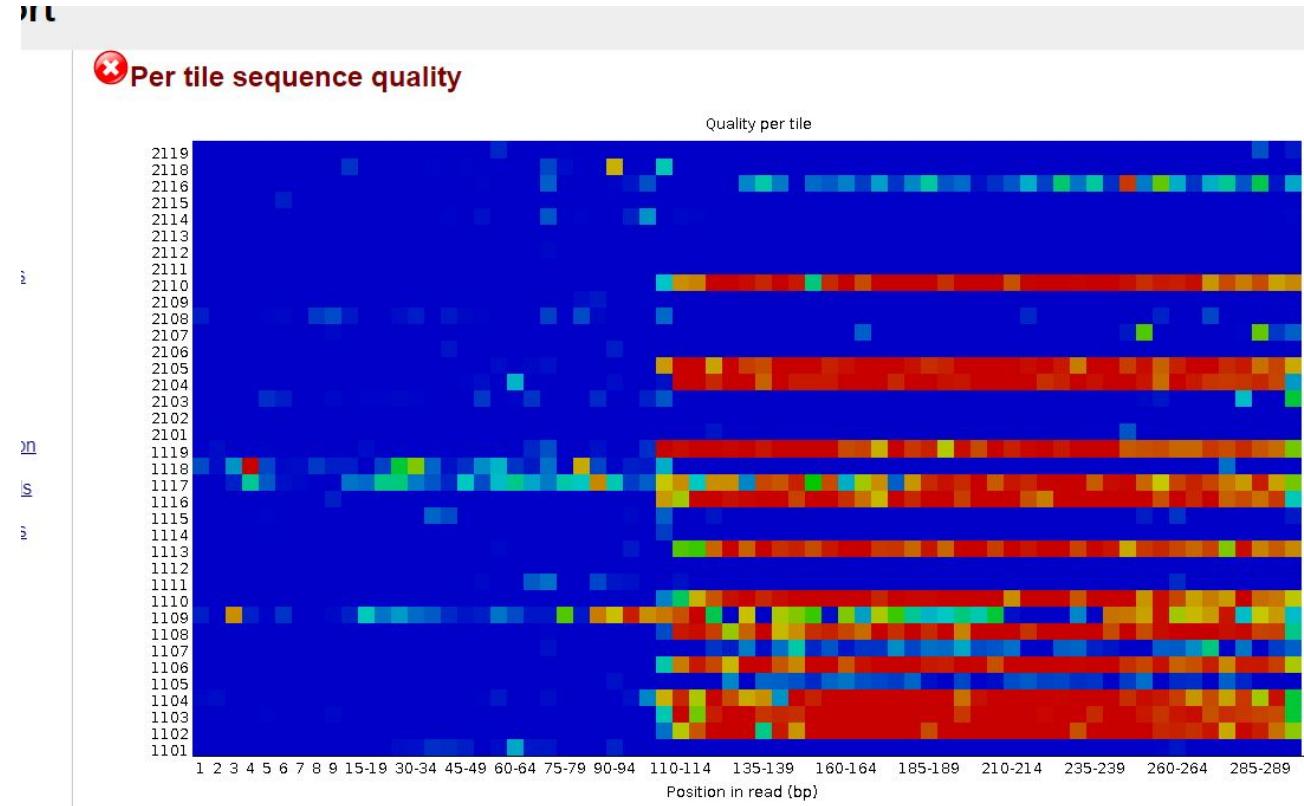
per tile sequence quality:

quality scores across all bases.

plot represent the deviation from average quality.

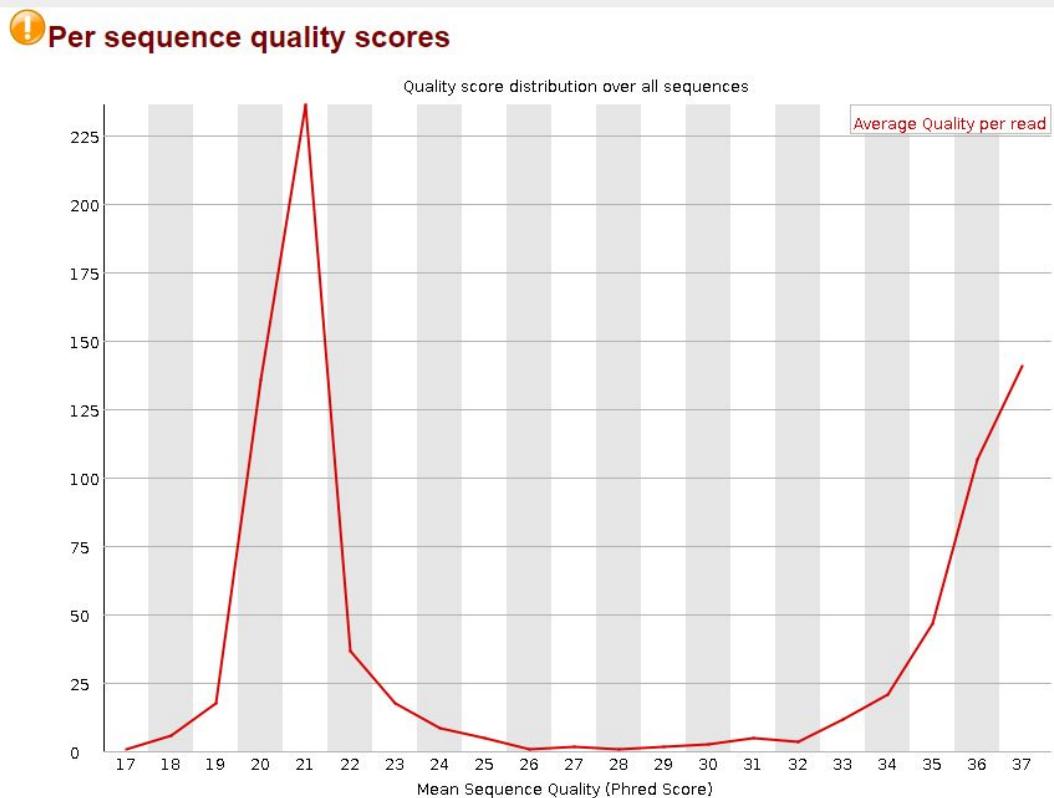
hotter colors=worst quality

blue colors=good quality



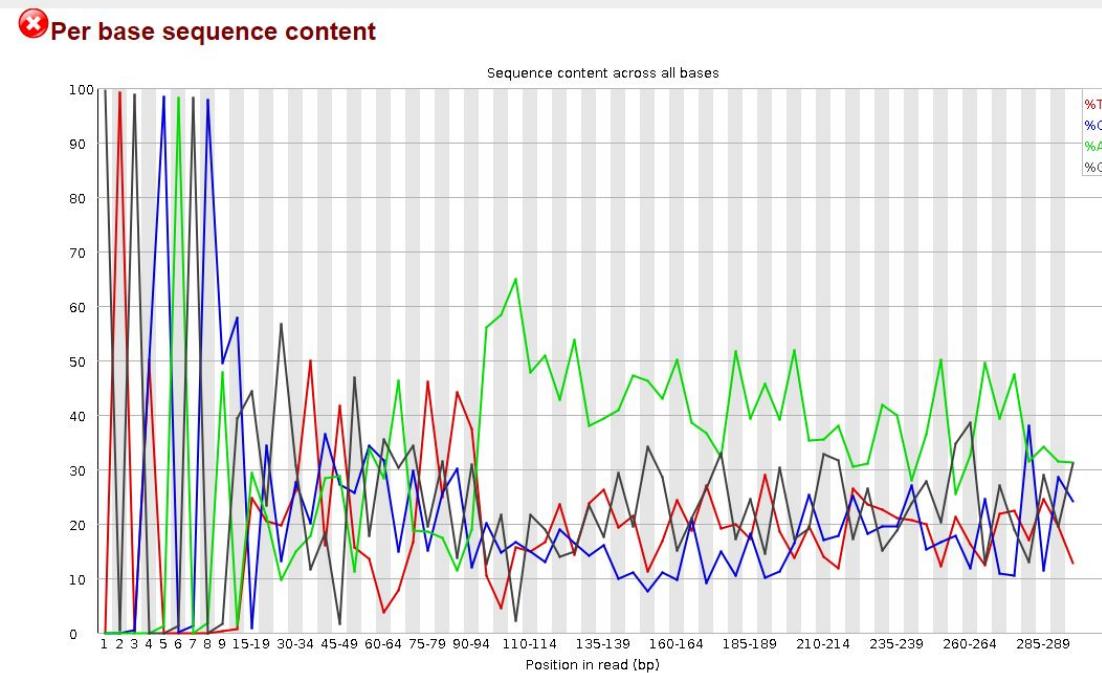
per sequence quality scores

plots the average quality score as distribution graph.



per base sequence content

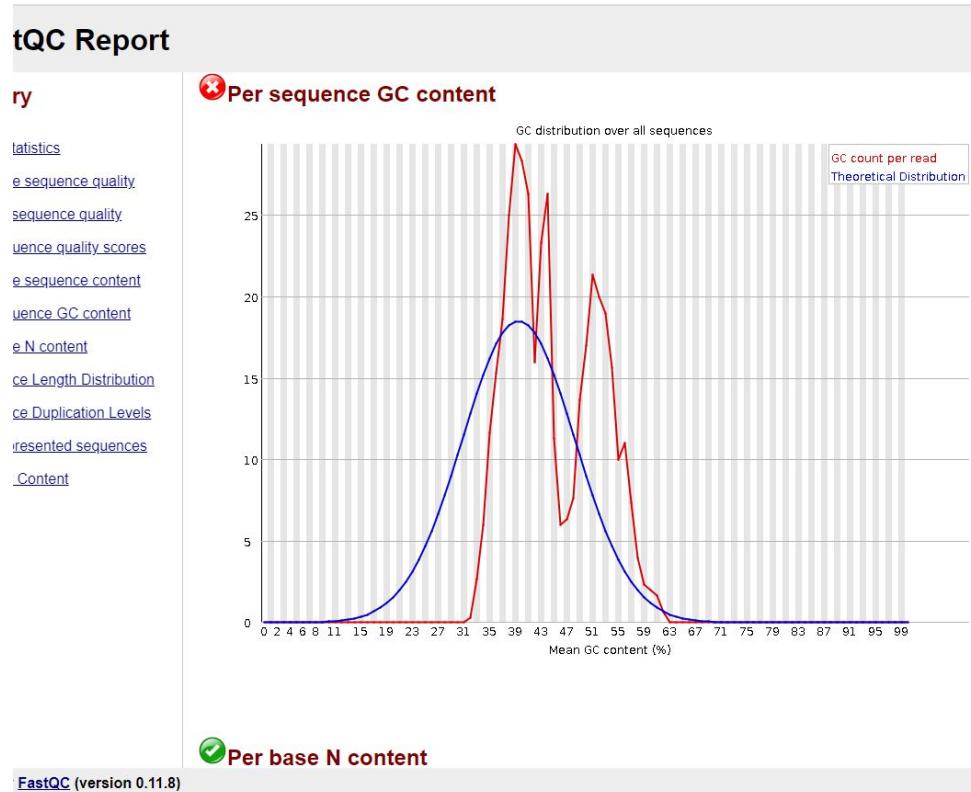
shows the % of average nucleotides in each position, in amplification data, the first positions are expected to be not really good, as in the graph.



Per sequence GC content

per sequence GC content

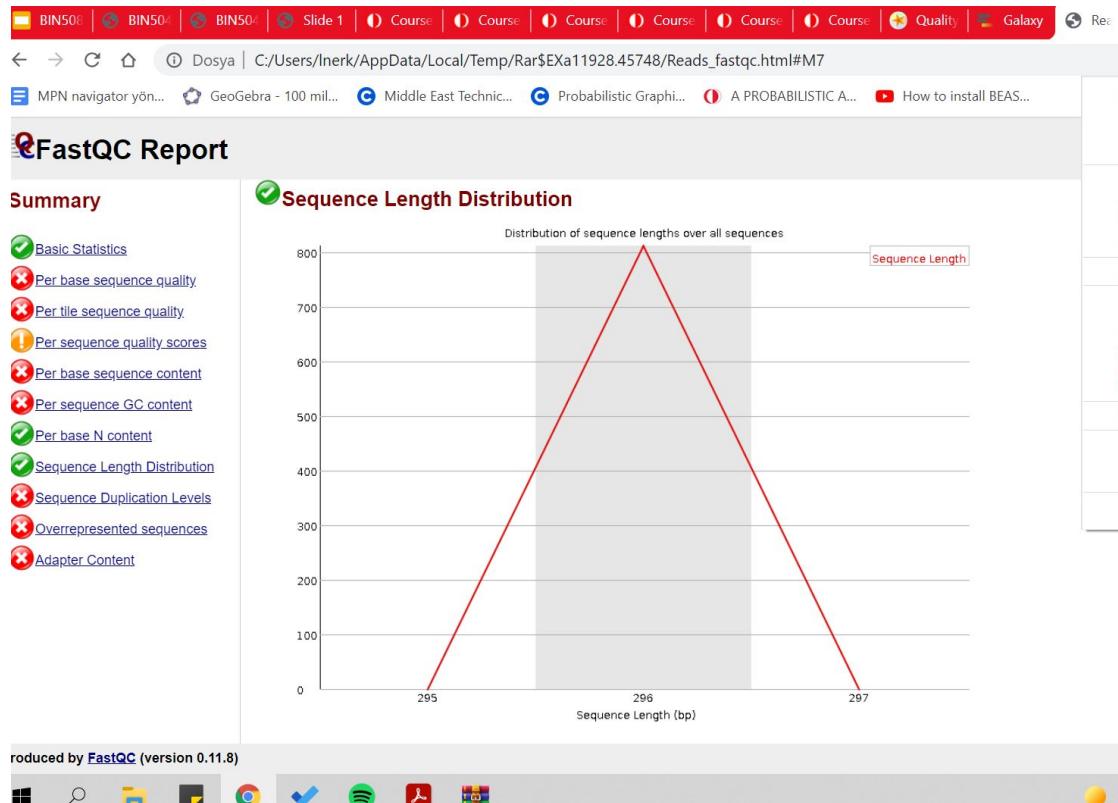
shows number reads vs % of bases,
shows mean GC content.



sequence length distribution

shows the distribution of the fragment sizes,

generally there is only one peak, but in some cases this can be uniform or more than one peak.

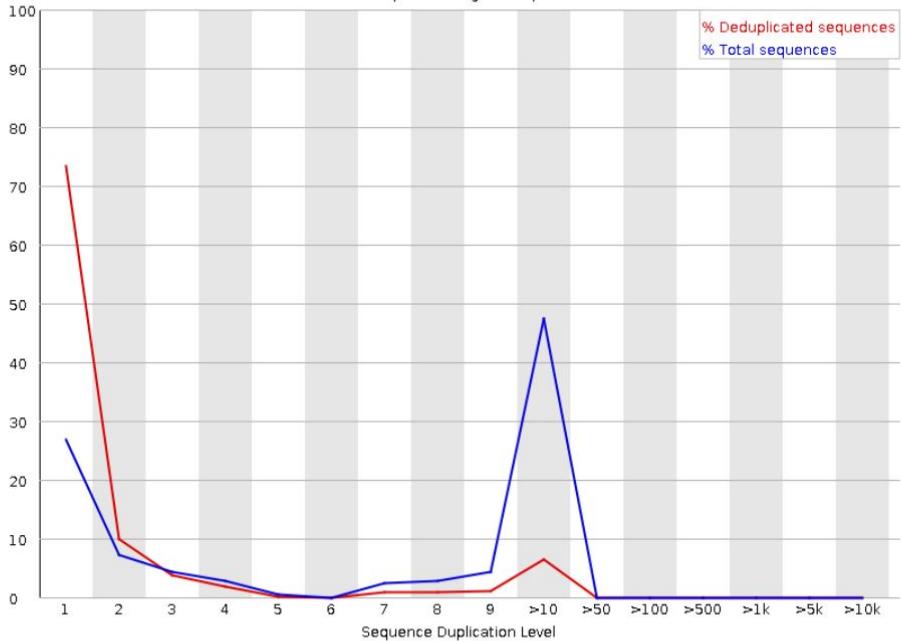


sequence duplication levels

shows the sequence duplication levels.

Sequence Duplication Levels

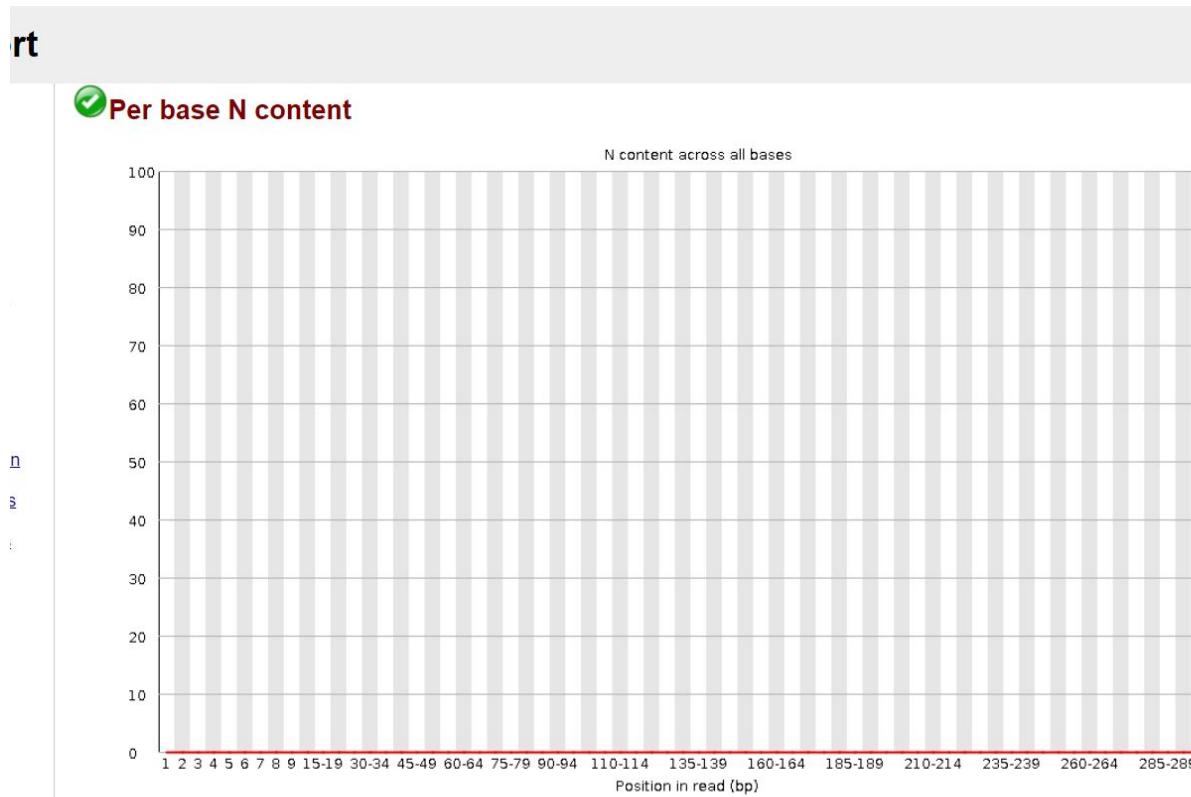
Percent of seqs remaining if deduplicated 36.82%



per base N content

shows the bases that have sufficient confidence,

N==conventional base call



Cutadapt: trimming

The screenshot shows the Galaxy web interface. At the top, there is a dark header bar with the word "Galaxy" in white. Below it, the main content area has a light gray background. In the top left corner of this area, the word "Tools" is displayed in bold black text. To the right of "Tools" are two small icons: a star and a three-line menu icon. Below "Tools", there is a search bar containing the text "cuta" with a clear button (an "X") to its right. Underneath the search bar is a large, light-gray rectangular button with a small upward arrow icon and the text "Upload Data". Further down is another light-gray button with a circular eye icon and the text "Show Sections". The main content area contains several tool descriptions. The first description is for "Cutadapt Remove adapter sequences from FASTQ/FASTA", which is highlighted with a red arrow pointing to it from the "Cutadapt: trimming" section above. The second description is for "Sequence Logo generator for fasta (eg Clustal alignments)". Below these descriptions is a light-gray horizontal bar labeled "WORKFLOWS" in bold black text. Underneath this bar, the text "All workflows" is visible.

Tools

cuta

Upload Data

Show Sections

Cutadapt Remove adapter sequences from FASTQ/FASTA

Sequence Logo generator for fasta (eg Clustal alignments)

WORKFLOWS

All workflows

 **Cutadapt** Remove adapter sequences from FASTQ/FASTA (Galaxy Version 3.4+galaxy0)

Single-end or Paired-end reads?

Single-end

FASTQ/A file



1: Reads



Should be of datatype "fastq.gz" or "fasta"

3' (End) Adapters

1: 3' (End) Adapters



Source

Enter custom sequence



Enter custom 3' adapter name (Optional if Multiple output is 'No')

Enter custom 3' adapter sequence

CTGTCTCTTATACACATCT

(-a)

Filter Options



Discard Trimmed Reads



Discard reads that contain the adapter instead of trimming them. Use the 'Minimum overlap length' option in order to avoid throwing away too many randomly matching reads! (--discard-trimmed)

Discard Untrimmed Reads



Discard reads that do not contain the adapter. (--discard_untrimmed)

Minimum length (R1)

20

Discard trimmed reads that are shorter than LENGTH. Reads that are too short even before adapter removal are also discarded. (--minimum-length)

Maximum length (R1)

Read Modification Options



Quality cutoff

20



Trim low-quality bases from 5' and/or 3' ends of each read before adapter removal. Applied to both reads if data is paired. If one value is given, only the 3' end is trimmed. If two comma-separated cutoffs are given, the 5' end is trimmed with the first cutoff, the 3' end with the second. (--quality-cutoff)

NextSeq trimming

0

Experimental option for quality trimming of NextSeq data. This is necessary because that machine cannot distinguish between G and reaching the end of the fragment (it encodes G as 'black'). This

execution



Executed **Cutadapt** and successfully added 1 job to the queue.

The tool uses this input:

- **1: Reads**

It produces 2 outputs:

- **6: Cutadapt on data 1: Read 1 Output**
- **7: Cutadapt on data 1: Report**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

inspecting txt file (report)

```
Galaxy7-[Cutadapt_on_data_1_Report] - Not Deterri
Dosya Düzen Biçimi Görünüm Yardım
This is cutadapt 3.4 with Python 3.9.2
Command line parameters: -j=6 -a CTGTCTTATACACATCT --output=out1.fq --error-rate=0.1 --times=1 --overlap=3 --action=trim --minimum-length=20 --quality-cuto
Processing reads on 6 cores in single-end mode ...
Finished in 0.11 s (137 µs/read; 0.44 M reads/minute).

==== Summary ===

Total reads processed: 812
Reads with adapters: 461 (56.8%)
Reads that were too short: 0 (0.0%)
Reads written (passing filters): 812 (100.0%)

Total basepairs processed: 240,352 bp
Quality-trimmed: 84,277 bp (35.1%)
Total written (filtered): 116,483 bp (48.5%)

==== Adapter 1 ===

Sequence: CTGTCTTATACACATCT; Type: regular 3'; Length: 19; Trimmed: 461 times

No. of allowed errors:
1-9 bp: 0; 10-19 bp: 1

Bases preceding removed adapters:
A: 0.0%
C: 0.0%
G: 100.0%
T: 0.0%
none/other: 0.0%
WARNING:
The adapter is preceded by "G" extremely often.
The provided adapter sequence could be incomplete at its 5' end.

Overview of removed sequences
```

FASTQE after trimming

The screenshot shows the Galaxy web interface with the FASTQE tool running. A red arrow points from the title "FASTQE after trimming" down to the "Score types to show" section of the tool's configuration panel. Another red arrow points from the "Maximum read length" input field down to its descriptive text below.

FASTQE visualize fastq files with emoji's 🎉🥳 (Galaxy Version 0.2.6+galaxy2)

FastQ data

6: Cutadapt on data 1: Read 1 Output
5: Cutadapt on data 1: Read 1 Output
1: Reads

Bin scores

No

Score types to show

Select/Unselect all

Mean

Maximum read length

500

Enable long reads up to this many bp long.

History

search datasets

BIN508#2:QC

7 shown
5.48 MB

7: Cutadapt on data 1: R eport

81 lines
format: **txt**, database: ?

This is cutadapt 3.4 with Python 3.9.
Command line parameters: -j=6 -a CTG
Processing reads on 6 cores in singl
Finished in 0.11 s (137 µs/read; 0.4

execution



Executed **FASTQE** and successfully added 1 job to the queue.

The tool uses this input:

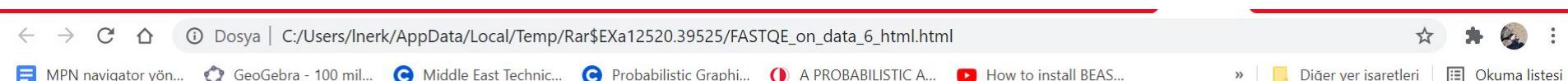
- **6: Cutadapt on data 1: Read 1 Output**

It produces this output:

- **8: FASTQE on data 6**

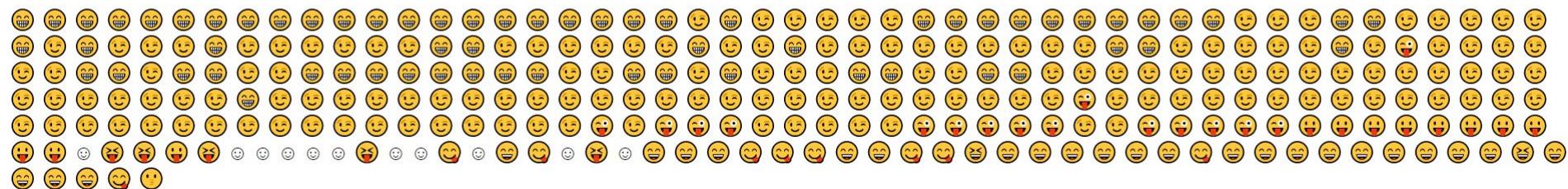
You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

inspecting html



FASTQE Report 😕

Cutadapt on data 1_ Read 1 Output: mean



Scratchbook

The screenshot shows a web browser window with the URL usegalaxy.org in the address bar. The main content is a Galaxy interface for a task titled "BIN508#2:QC".

Galaxy Header: Includes links for Home, Workflow, Visualize, Shared Data, Help, User, and a grid icon.

Left Panel: Titled "BIN508#2:QC". It has sections for "Search tools" (with a search bar and clear button), "Upload Data" (button), and "STATISTICS AND VISUALIZATION" (with "Statistics" and "Machine Learning" sub-sections).

Middle Panel: Shows a success message: "Executed **FASTQE** and successfully added 1 job to the queue." It details the input and output steps:

- Input: "6: Cutadapt on data 1: Read 1 Output"
- Output: "8: FASTQE on data 6"

A note at the bottom says: "You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed."

Right Panel: Titled "History". It shows "BIN508#2:QC" with "8 shown" results, a total size of "5.48 MB", and icons for "Format", "Run", "Database", and help.

comparing the outputs:

Screenshot of a web browser comparing two FASTQE reports side-by-side.

The browser window title is "usegalaxy.org".

The tabs shown in the header are: BIN50, BIN50, BIN50, Slide, Courses, Courses, Courses, Courses, Courses, Quality, Ge... (partially visible), Reads, FASTC, +, and several others.

The address bar shows the URL "usegalaxy.org".

The page content displays two FASTQE reports:

- Left Report (BIN50#2:QC: FASTQE on data 1):** Titled "FASTQE Report 😊". Sub-section "Reads: mean" is shown. The data is represented by a grid of smiley face emojis. A few warning icons (yellow triangles with exclamation marks) are scattered among the smileys.
- Right Report (BIN50#2:QC: FASTQE on data 6):** Titled "FASTQE Report 😕". Sub-section "Cutadapt on data 1_ Read 1 Output: mean" is shown. The data is represented by a grid of smiley face emojis. A few warning icons (yellow triangles with exclamation marks) are scattered among the smileys.

At the bottom of the right report, there is a status message: "2: FASTQE on data 1" followed by a file icon, edit icon, and delete icon.

FastQC after trimming

 **FastQC** Read Quality reports (Galaxy Version 0.72+galaxy1)

Short read data from your current history

6: Cutadapt on data 1: Read 1 Output

Contaminant list

No tabular dataset available.

tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA

Adapter list

No tabular dataset available.

list of adapters adapter sequences which will be explicitly searched against the library. tab delimited file with 2 columns: name and sequence. (--adapters)

Submodule and Limit specifying file

Nothing selected

a file that specifies which submodules are to be executed (default=all) and also specifies the thresholds for each submodules warning parameter

Disable grouping of bases for reads >50bp

No

Using this option will cause fastqc to crash and burn if you use it on really long reads, and your plots may end up a ridiculous size. You have been warned! (--nogroup)

Lower limit on the length of the sequence to be shown in the report

History

search datasets

BIN508#2:QC

8 shown

5.48 MB

format: **html**, database: ?

Picked up _JAVA_OPTIONS
Djava.io.tmpdir=/corral/...
-Xmx7g -Xms256m

HTML file

2: FASTQC on data 1

1.6 KB

format: **html**, database: ?

View details

1: Reads

execution

The screenshot shows a web-based bioinformatics tool interface. At the top, there is a navigation bar with several tabs: 'ebra - 100 mil...', 'Middle East Technic...', 'Probabilistic Graphi...', 'A PROBABILISTIC A...', 'How to install BEAS...', and 'Diğer yer işaret...'. Below the navigation bar is a dark header with icons for home, workflow, visualize, shared data, help, user, and a grid.

The main content area has a green background. It displays a success message: 'Executed **FastQC** and successfully added 1 job to the queue.' followed by a checkmark icon. Below this, it says 'The tool uses this input:' and lists one item: '• 6: Cutadapt on data 1: Read 1 Output'. It then says 'It produces 2 outputs:' and lists two items: '• 9: FastQC on data 6: Webpage' and '• 10: FastQC on data 6: RawData'. A note at the bottom states: 'You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.'

To the right of the main content area, there is a sidebar titled 'History' with a search bar. It shows a list of recent history items:

- BIN508#2:0
- 10 shown
- 5.48 MB
- File type
- 3: FastQC on data 6: Webpage
- 923.5 KB
- format: html, c
- Picked up _JAVA_OPTIONS
- Djava.io.tmpdir
- Xmx7g -Xms

At the bottom right, there are icons for file download, refresh, and other navigation functions, along with the text 'HTML file'.

inspecting the report

Sat 6 Nov 2021
Cutadapt on data 1_ Read 1 Output

FastQC Report

Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✗ Per tile sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ✗ Per sequence GC content
- ✓ Per base N content
- ! Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences

Produced by [FastQC](#) (version 0.11.8)

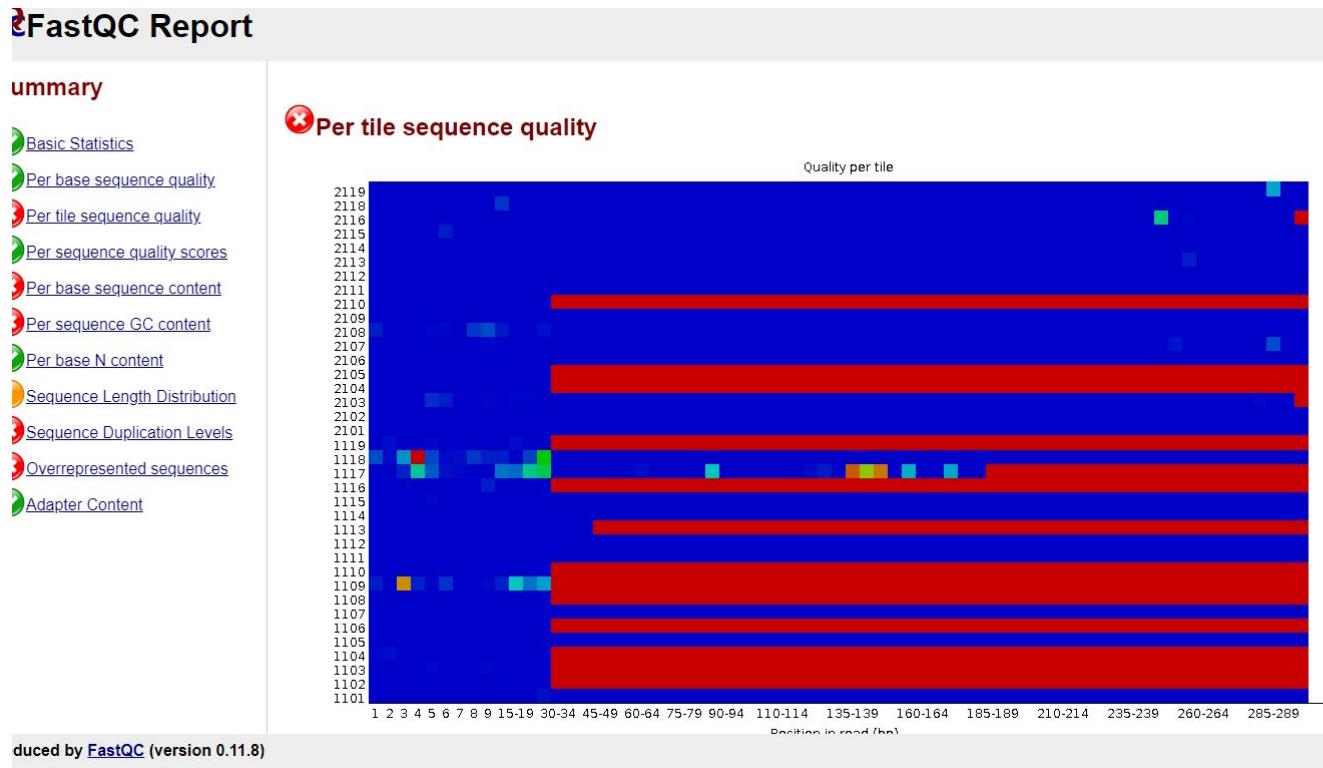
Basic Statistics

Measure	Value
Filename	Cutadapt on data 1_ Read 1 Output
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	812
Sequences flagged as poor quality	0
Sequence length	26-296
%GC	54

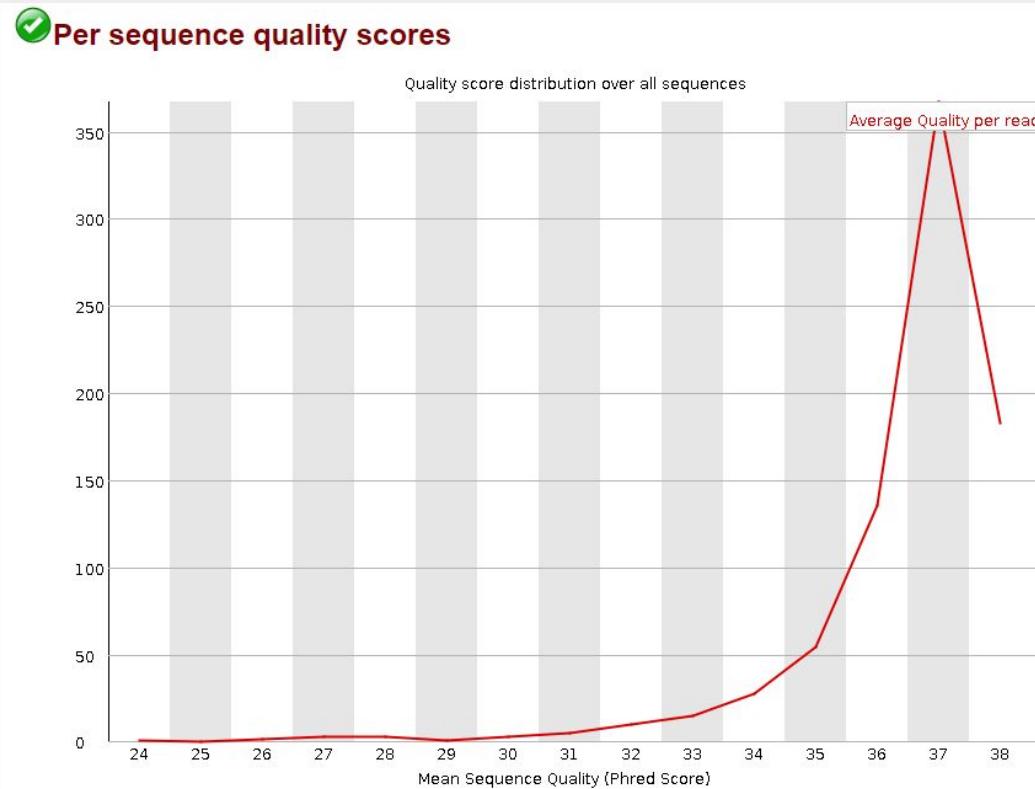
FastQC_on_data_6_.zip FASTQE_on_data_6.zip Galaxy7-[Cutadapt...].txt Galaxy5-[Cutada...].fastq

Tümünü göster

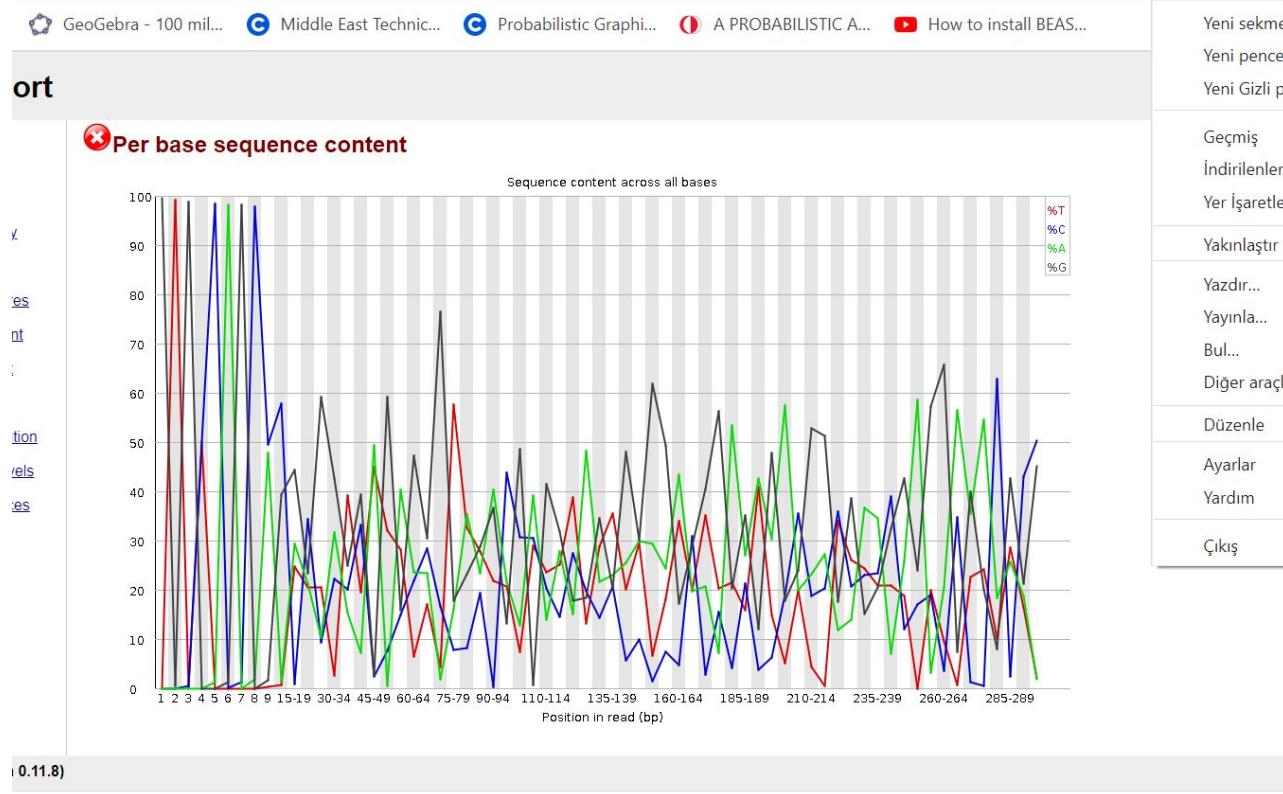
quality per tile



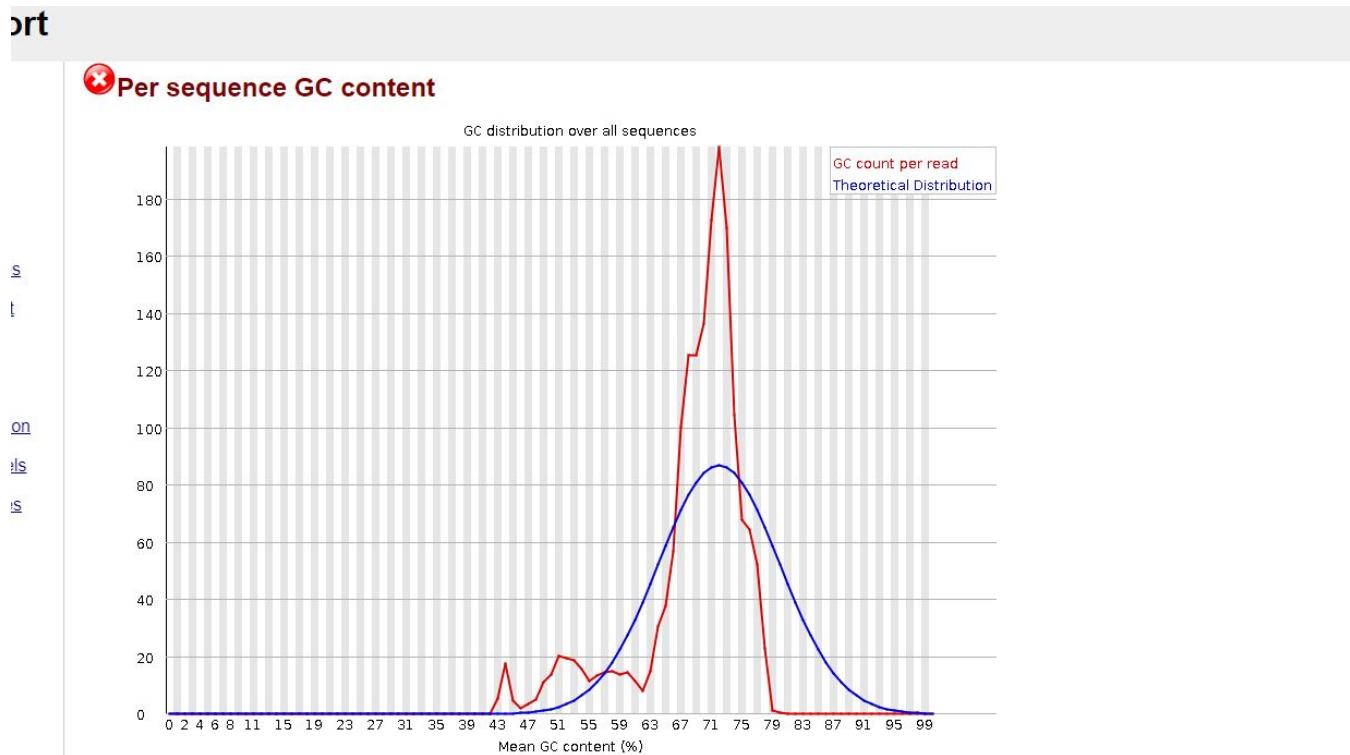
per sequence quality scores



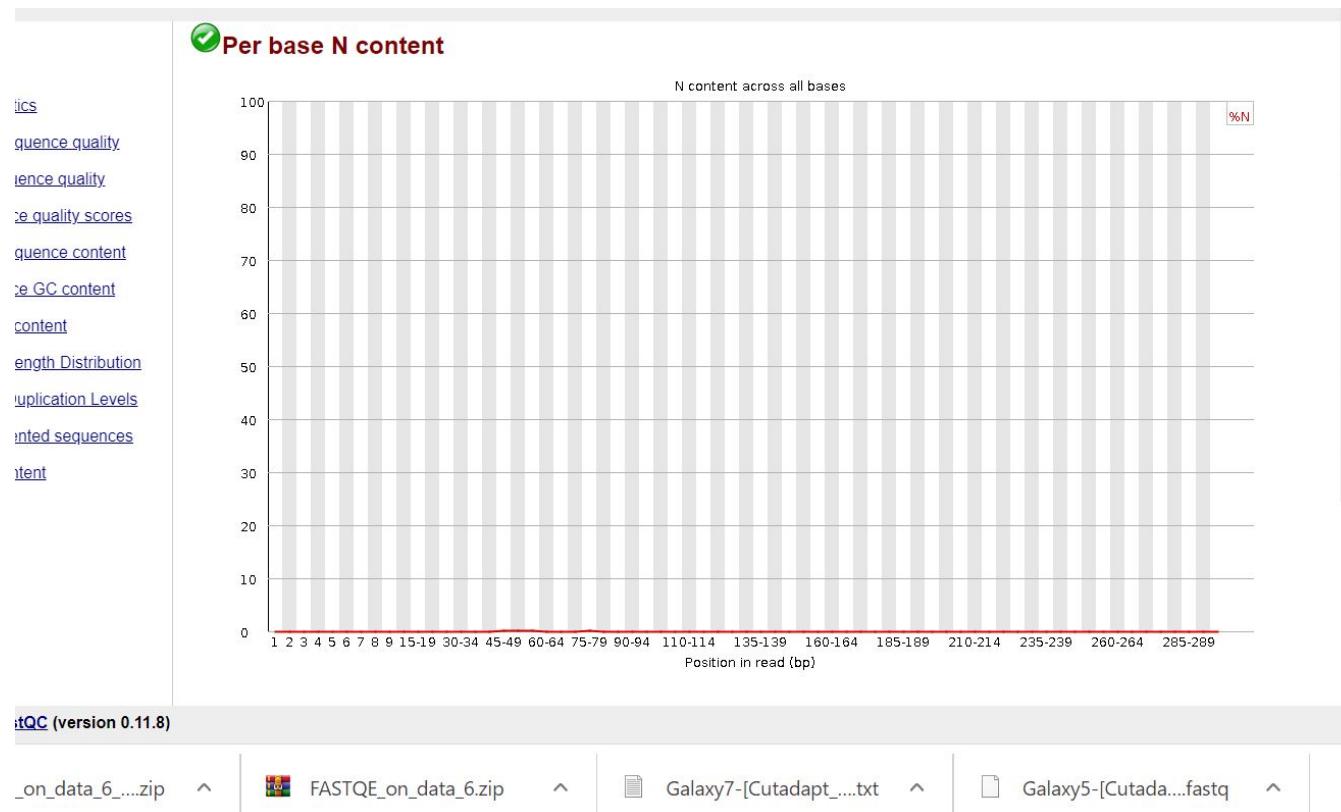
per base sequence content



per sequence GC content



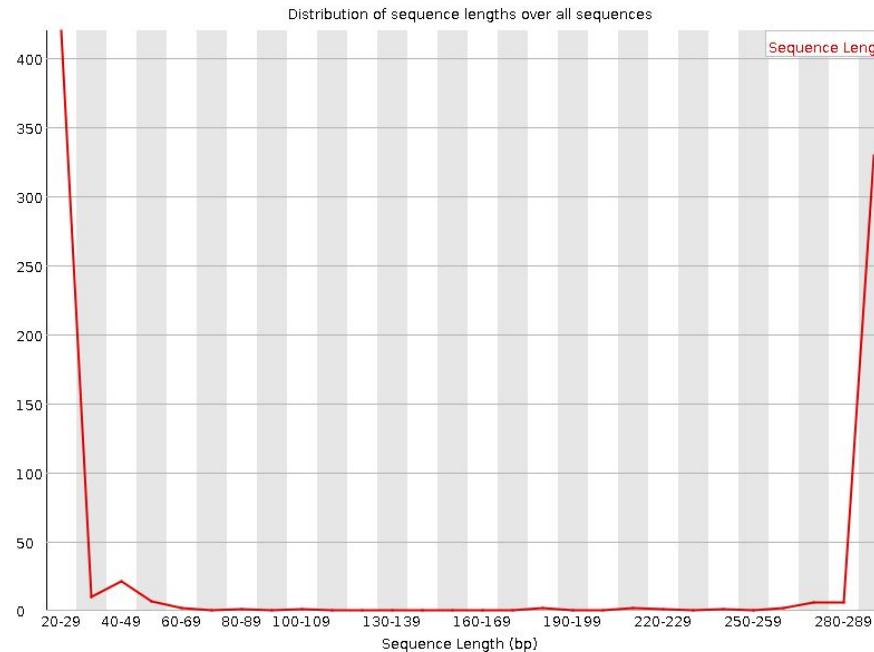
per base N content



sequence length distribution

t

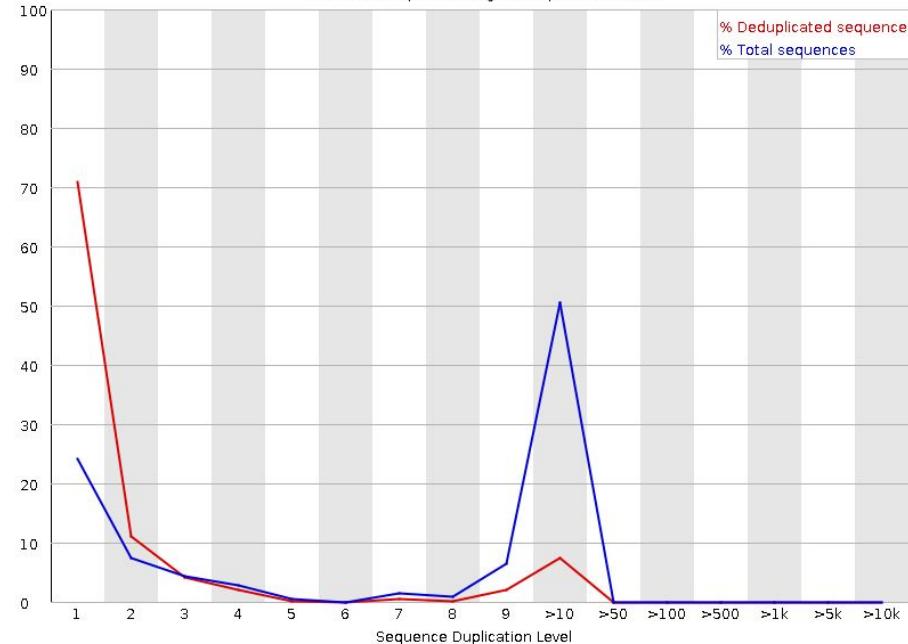
⚠ Sequence Length Distribution



sequence duplication levels

Sequence Duplication Levels

Percent of seqs remaining if deduplicated 34.11%



importing new data

The screenshot shows a software interface for managing biological data. At the top, there is a navigation bar with links for Home, Workflow, Visualize, Shared Data, Help, User, and a profile icon. Below the navigation bar, a title bar reads "Download from web or upload from disk". Underneath this, there is a tab bar with four options: Regular (selected), Composite, Collection, and Rule-based.

The main area displays a table with columns: Name, Size, Type, Genome, Settings, and Status. A single row is present in the table:

Name	Size	Type	Genome	Settings	Status
New File	155 b	Auto-det...	unspecified (?)		100%

Below the table, there is a green-highlighted input field containing two URLs:

```
https://zenodo.org/record/61771/files/GSM461178_untreat_paired_subset_1.fastq  
https://zenodo.org/record/61771/files/GSM461178_untreat_paired_subset_2.fastq
```

At the bottom of the dialog, there are several buttons and dropdown menus:

- Type (set all): Auto-detect
- Genome (set all): unspecified (?)
- Buttons: Choose local files, Choose remote files, Paste/Fetch data, Start, Pause, Reset, Close

execution FastQC

 Workflow Visualize Shared Data ▾ Help ▾ User ▾   



Executed **FastQC** and successfully added undefined jobs to the queue.

The tool uses 2 inputs:

- 11: **GSM461178_untreat_paired_subset_1.fastq**
- 12: **GSM461178_untreat_paired_subset_2.fastq**

It produces 4 outputs:

- 13: **FastQC on data 11: Webpage**
- 14: **FastQC on data 11: RawData**
- 15: **FastQC on data 12: Webpage**
- 16: **FastQC on data 12: RawData**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

MultiQC

The screenshot shows the MultiQC software interface for configuring a new analysis. The main window has a dark header bar with tabs for 'Workflow', 'Visualize', 'Shared Data', 'Help', and 'User'. On the left, there's a sidebar with a green button labeled 'New' and a red 'X' button. The main content area is titled '1: Results'.

Which tool was used to generate logs?
FastQC

Software name

FastQC output
1: FastQC output

Type of FastQC output?
Raw data

FastQC output

- 16: FastQC on data 12: RawData
- 15: FastQC on data 12: Webpage
- 14: FastQC on data 11: RawData
- 13: FastQC on data 11: Webpage
- 12: GSM461178_untreat_paired_subset_2.fastq
- 11: GSM461178_untreat_paired_subset_1.fastq

Buttons:
+ Insert FastQC output
+ Insert Results

Report title

History
BIN508
16 shown
53.59 MB

On the right side, there's a vertical 'History' panel with a search bar at the top. It lists several entries:

- 16: FastQC RawData
- 15: FastQC Webpage
- 14: FastQC RawData
- 13: FastQC Webpage
- 12: GSM_untreat_paired_
- 11: GSM_untreat_paired_
- 10: FastQC

execution



Executed **MultiQC** and successfully added 1 job to the queue.

The tool uses 2 inputs:

- **14: FastQC on data 11: RawData**
- **16: FastQC on data 12: RawData**

It produces this output:

- **18: MultiQC on data 16 and data 14: Webpage**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

html file of the MultiQC

Dosya | C:/Users/Inerk/AppData/Local/Temp/Kar\$Exa18020.18009/MultiQC_on_data_16_and_data_14__Webpage_html.html

MPN navigator yön... GeoGebra - 100 mil... Middle East Technic... Probabilistic Graphi... A PROBABILISTIC A... How to install BEAS... » Diğer yer işaretleri Okuma listesi

MultiQC

v1.9

General Stats

FastQC

Sequence Counts

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

Status Checks

MultiQC

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2021-11-06, 08:16 based on data in: /corral4/main/jobs/038/884/38884820/working/multiqc_WDir

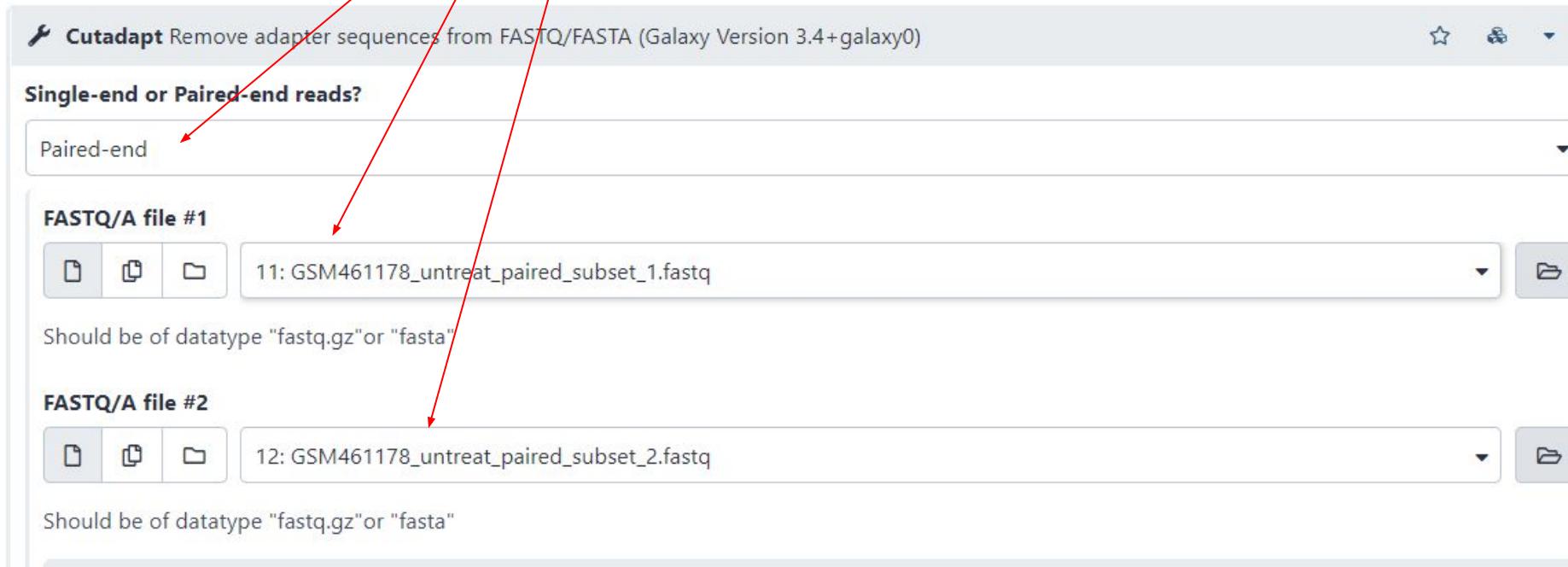
Welcome! Not sure where to start? Watch a tutorial video (6:06) don't show again X

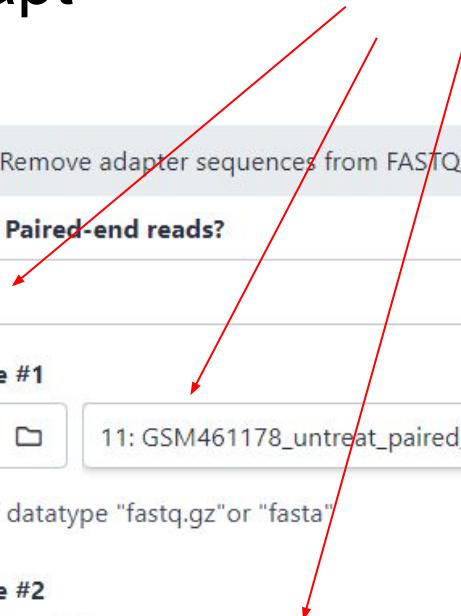
General Statistics

Showing 2/2 rows and 3/5 columns.

Sample Name	% Dups	% GC	M Seqs
GSM461178_untreat_paired_subset_1_fastq	8.8%	53%	0.1
GSM461178_untreat_paired_subset_2_fastq	7.3%	53%	0.1

cutadapt

A screenshot of the Cutadapt tool interface in Galaxy. The title bar says "Cutadapt Remove adapter sequences from FASTQ/FASTA (Galaxy Version 3.4+galaxy0)".
Single-end or Paired-end reads?
Paired-end
FASTQ/A file #1
11: GSM461178_untreat_paired_subset_1.fastq
Should be of datatype "fastq.gz" or "fasta"
FASTQ/A file #2
12: GSM461178_untreat_paired_subset_2.fastq
Should be of datatype "fastq.gz" or "fasta"



- A red arrow points from the top-left towards the "Paired-end" radio button.
- Two red arrows point downwards from the top-left towards the "FASTQ/A file #1" input field and the "FASTQ/A file #2" input field respectively.

Filter Options

Discard Trimmed Reads



Discard reads that contain the adapter instead of trimming them. Use the 'Minimum overlap length' option in order to avoid throwing away too many randomly matching reads! (--discard-trimmed)

Discard Untrimmed Reads



Discard reads that do not contain the adapter. (--discard_untrimmed)

Minimum length (R1)

20

Discard trimmed reads that are shorter than LENGTH. Reads that are too short even before adapter removal are also discarded. (--minimum-length)

Maximum length (R1)

Discard reads that did not pass CASAVA filtering (header has :Y). (--discard-cassava)

Read Modification Options



Quality cutoff

20

Trim low-quality bases from 5' and/or 3' ends of each read before adapter removal. Applied to both reads if data is paired. If one value is given, only the 3' end is trimmed. If two comma-separated cutoffs are given, the 5' end is trimmed with the first cutoff, the 3' end with the second. (--quality-cutoff)

NextSeq trimming

0

Experimental option for quality trimming of NextSeq data. This is necessary because that machine cannot distinguish between G and reaching the end of the fragment (it encodes G as 'black'). This option works like regular quality trimming (where one would use -q 20 instead), except that the qualities of G bases are ignored. (--nextseq-trim)

Trim Ns



No

This option allows to unconditionally remove bases from the beginning or end of each read. If the given length is positive, the bases are removed from the beginning of each read. If it is negative, the bases are removed from the end. (--cut)

Change negative quality values to zero



(--zero-cap)

Outputs selector

Select/Unselect all

- Report: Cutadapt's per-adapter statistics. You can use this file with MultiQC.
- Info file: write information about each read and its adapter matches.
- Rest of read: when the adapter matches in the middle of a read, write the rest (after the adapter).
- Wildcard file: when the adapter has wildcard bases (Ns) write adapter bases matching wildcard positions.
- Too short reads: write reads that are too short according to minimum length specified (default: discard reads).
- Too long reads: write reads that are too long (according to maximum length specified)
- Untrimmed reads: write reads that do not contain the adapter to a separate file, instead of writing them to the regular output file (default: output same file as trimmed)
- Multiple output: create a separate file for each adapter trimmed (default: all trimmed reads are in a single file)

Job Resource Parameters

execution

[Workflow](#)[Visualize](#)[Shared Data](#) ▾[Help](#) ▾[User](#) ▾

Executed **Cutadapt** and successfully added 1 job to the queue.

The tool uses 2 inputs:

- 11: **GSM461178_untreat_paired_subset_1.fastq**
- 12: **GSM461178_untreat_paired_subset_2.fastq**

It produces 3 outputs:

- 22: **Cutadapt on data 12 and data 11: Read 1 Output**
- 23: **Cutadapt on data 12 and data 11: Read 2 Output**
- 24: **Cutadapt on data 12 and data 11: Report**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

inspecting generated txt file:

Galaxy24-[Cutadapt_on_data_12_and_data_11_Report] - Not Defteri

Dosya Düzen Biçim Görünüm Yardım

This is cutadapt 3.4 with Python 3.9.2

Command line parameters: -j=6 --output=out1.fq --paired-output=out2.fq --error-rate=0.1 --times=1 --overlap=3 --action=trim --minimum-length=20 --pair-filter

Processing reads on 6 cores in paired-end mode ...

Finished in 0.90 s (9 µs/read; 6.69 M reads/minute).

==== Summary ===

Total read pairs processed: 100,000

 Read 1 with adapter: 0 (0.0%)

 Read 2 with adapter: 0 (0.0%)

Pairs that were too short: 1,376 (1.4%)

Pairs written (passing filters): 98,624 (98.6%)

Total basepairs processed: 7,400,000 bp

 Read 1: 3,700,000 bp

 Read 2: 3,700,000 bp

Quality-trimmed: 182,802 bp (2.5%)

 Read 1: 44,164 bp

 Read 2: 138,638 bp

Total written (filtered): 7,159,132 bp (96.7%)

 Read 1: 3,616,660 bp

 Read 2: 3,542,472 bp

making history visible

https://usegalaxy.org/u/leman_nur_nehri/h/bin5082qc

The screenshot shows a user interface for sharing a Galaxy history. At the top, there's a dark header bar with various navigation links like 'Home', 'Workflows', 'Datasets', 'Shared Data', 'Help', 'User', and 'Logout'. Below this is a white main content area with a dark sidebar on the left.

Share or Publish History 'BIN508#2:QC'

Two toggle buttons are shown:

- Make History accessible
- Make History publicly available in Published Histories

A message states: "This History is currently **accessible via link and published**. Anyone can view and import this History by visiting the following URL:"

Below this, there are two icons: a pencil for editing and a link for sharing, followed by the URL: [url: https://usegalaxy.org/u/leman_nur_nehri/h/bin5082qc](https://usegalaxy.org/u/leman_nur_nehri/h/bin5082qc)

At the bottom right, there's a button labeled "Share History with Individual Users ▾".