# Galaxy for virologist training Exercise 2: Quality control and trimming

Despite the improvement of sequencing methods, there is no error-free technique. A correct measuring of the sequencing quality is essential for identifying problems in the sequencing, thus, this must be the first step in every sequencing analysis. Once the quality control is finished, it's important to remove those low quality reads, or short reads, for which a trimming step is mandatory. After the trimming step it is recommended to perform a new quality control step to be sure that trimming worked.

## 1. Illumina Quality control and trimming

| Title | Pre-processing |
|---|---|
| **Training dataset:** | PRJEB43037 - In August 2020, an outbreak of West Nile Virus affected 71 people with meningoencephalitis in Andalusia and 6 more cases in Extremadura (south-west of Spain), causing a total of eight deaths. The virus belonged to the lineage 1 and was relatively similar to previous outbreaks occurred in the Mediterranean region. Here, we present a detailed analysis of the outbreak, including an extensive phylogenetic study. This is one of the outbreak samples. |
| **Questions:** | • How do I check whether my Illumina data was correctly sequenced?<br>• How can I improve the quality of my data? |
| **Objectives**: | • Perform a quality control in raw Illumina reads<br>• Perform a quality trimming in raw Illumina reads<br>• Perform a quality control in trimmed Illumina reads |
| **Estimated time**: | 25 min |

### 1.1. Quality control

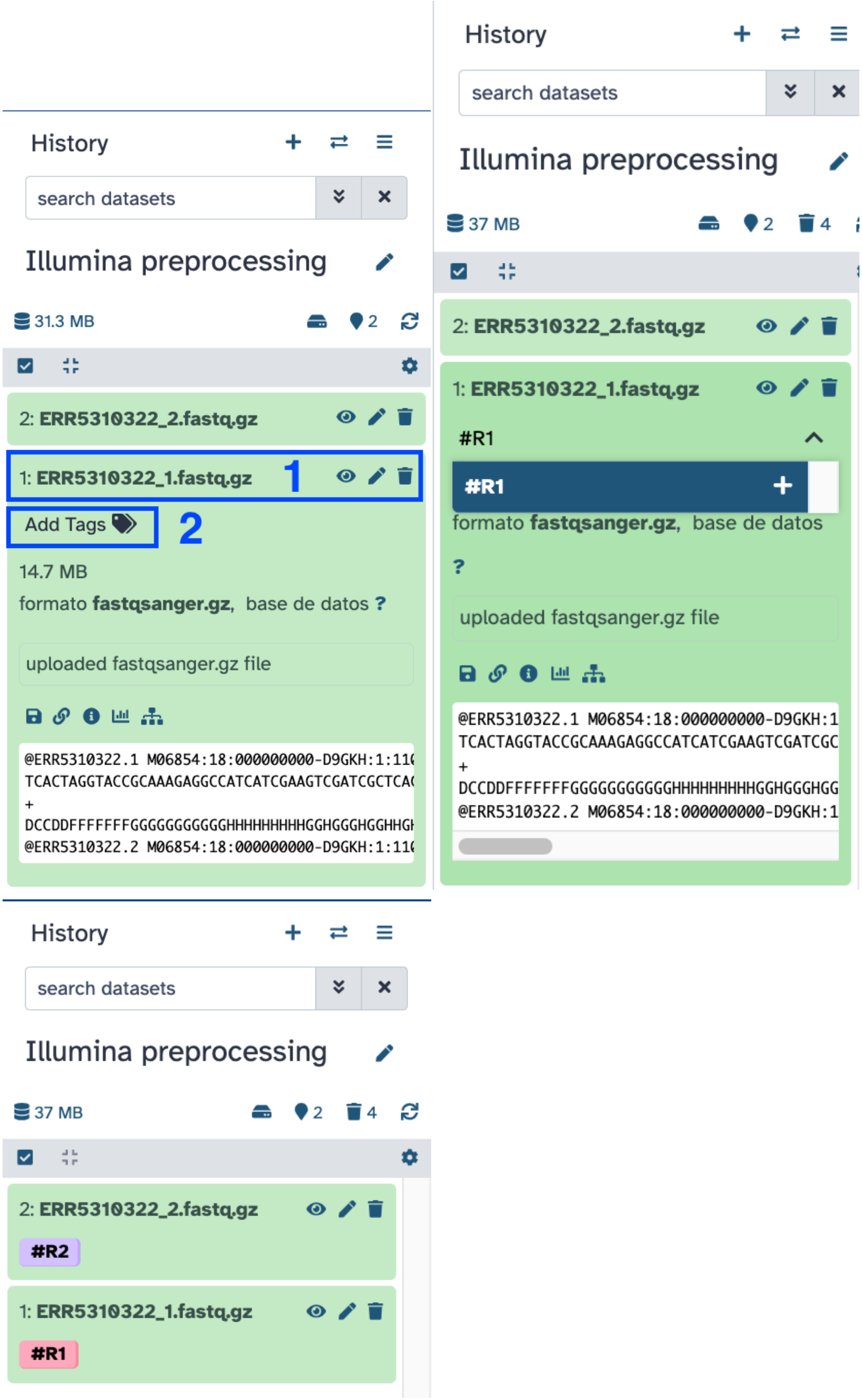
#### 1.1.1. Upload data

To run the quality control over the samples, follow these steps:

1. Create a new history, as we explained yesterday named **Illumina preprocessing**
2. Upload data as seen yesterday, copy and paste the following URLs:

```
ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR531/002/ERR5310322/ERR5310322_1.fastq.gz
ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR531/002/ERR5310322/ERR5310322_2.fastq.gz
```

3. Add some tags to the files. ***It is mandatory that the tag starts with # to be propagated to the processes***.



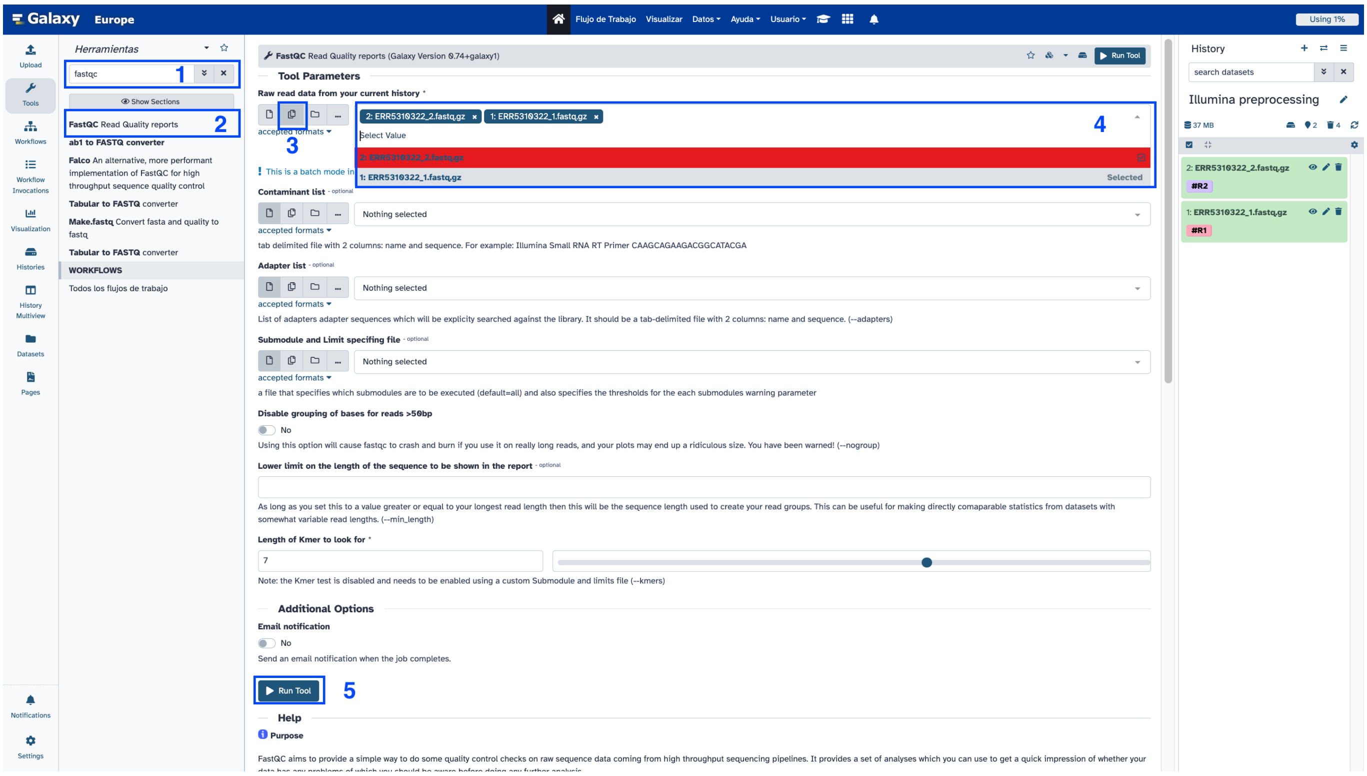
```
@ERR5310322.1 M06854:18:000000000-D9GKH:1:11
TCACTAGGTACCGCAAAGAGGCCATCATCGAAGTCGATCGCTCA
+
DCCDDFFFFFFFGGGGGGGGGGHHHHHHHHHGGHGGGHGGHHG
@ERR5310322.2 M06854:18:000000000-D9GKH:1:11
```

```
@ERR5310322.1 M06854:18:000000000-D9GKH:1
TCACTAGGTACCGCAAAGAGGCCATCATCGAAGTCGATCGC
+
DCCDDFFFFFFFGGGGGGGGGGHHHHHHHHHGGHGGGHGG
@ERR5310322.2 M06854:18:000000000-D9GKH:1
```

### 1.1.2. Run FastQC

1. Search for the **fastqc** tool
2. Select **FastQC Read Quality reports** and set the following parameters:
3. Select multiple file data set in Raw read data from your current history
4. Select the two datasets

5. Then go down and select **Run tool**



fastqc

**FastQC** Read Quality reports

**ab1 to FASTQ converter**

**Falco** An alternative, more performant implementation of FastQC for high throughput sequence quality control

**Tabular to FASTQ** converter

**Make.fastq** Convert fasta and quality to fastq

**Tabular to FASTQ** converter

Todos los flujos de trabajo

**FastQC** Read Quality reports (Galaxy Version 0.74+galaxy1)

**Tool Parameters**

**Raw read data from your current history** *

2: ERR5310322_2.fastq.gz × 1: ERR5310322_1.fastq.gz ×

Select Value

2: ERR5310322_2.fastq.gz

1: ERR5310322_1.fastq.gz Selected

accepted formats

! This is a batch mode in

**Contaminant list** - optional

Nothing selected

accepted formats

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

**Adapter list** - optional

Nothing selected

accepted formats

List of adapters adapter sequences which will be explicity searched against the library. It should be a tab-delimited file with 2 columns: name and sequence. (--adapters)

**Submodule and Limit specifing file** - optional

Nothing selected

accepted formats

a file that specifies which submodules are to be executed (default=all) and also specifies the thresholds for the 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** - optional

As long as you set this to a value greater or equal to your longest read length then this will be the sequence length used to create your read groups. This can be useful for making directly comaparable statistics from datasets with somewhat variable read lengths. (--min_length)

**Length of Kmer to look for** *

7

Note: the Kmer test is disabled and needs to be enabled using a custom Submodule and limits file (--kmers)

**Additional Options**

**Email notification**

No

Send an email notification when the job completes.

**Help**

**Purpose**

FastQC aims to provide a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines. It provides a set of analyses which you can use to get a quick impression of whether your

Illumina preprocessing

2: ERR5310322_2.fastq.gz

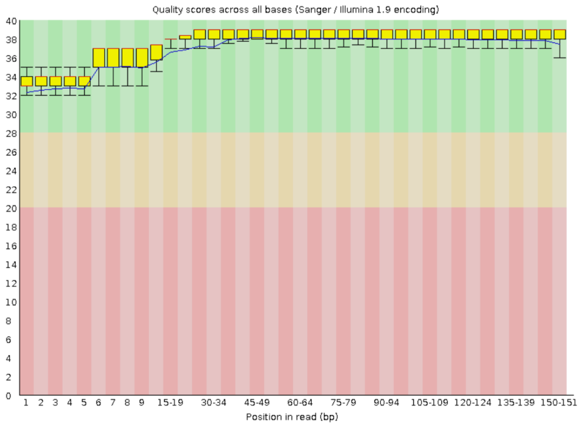
1: ERR5310322_1.fastq.gz

To see the results we are going to open the jobs with **Web page** in their name for both data 1 and data 2.

**Basic Statistics**

| Measure | Value |
|---|---|
| Filename | ERR5310322_1_fastq_gz.gz |
| File type | Conventional base calls |
| Encoding | Sanger / Illumina 1.9 |
| Total Sequences | 265989 |
| Sequences flagged as poor quality | 0 |
| Sequence length | 35-151 |
| %GC | 51 |

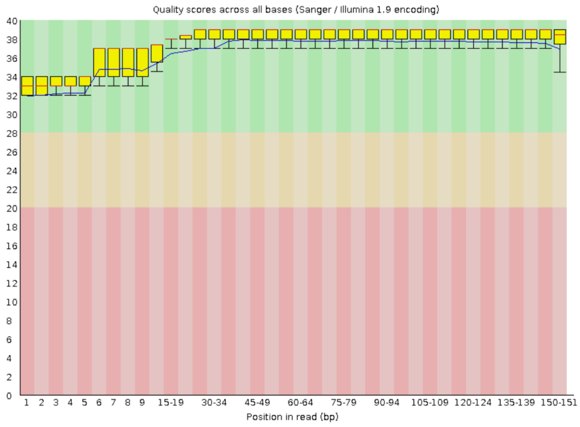
**Per base sequence quality**



**Basic Statistics**

| Measure | Value |
|---|---|
| Filename | ERR5310322_2_fastq_gz.gz |
| File type | Conventional base calls |
| Encoding | Sanger / Illumina 1.9 |
| Total Sequences | 265989 |
| Sequences flagged as poor quality | 0 |
| Sequence length | 35-151 |
| %GC | 51 |

**Per base sequence quality**



Here, you can see the number of reads in each file, the maximum and minimum length of all reads in the sample, and the quality plots for both R1 and R2. They look quite good, but we are going to run trimming over the samples.

▶ How many reads do the samples have?

## First question

▶ How do I check whether my Illumina data was correctly sequenced?

## 1.2. Trimming

Once we have performed the quality control, we have to perform the quality and read length trimming:

### 1.2.1. Run Fastp

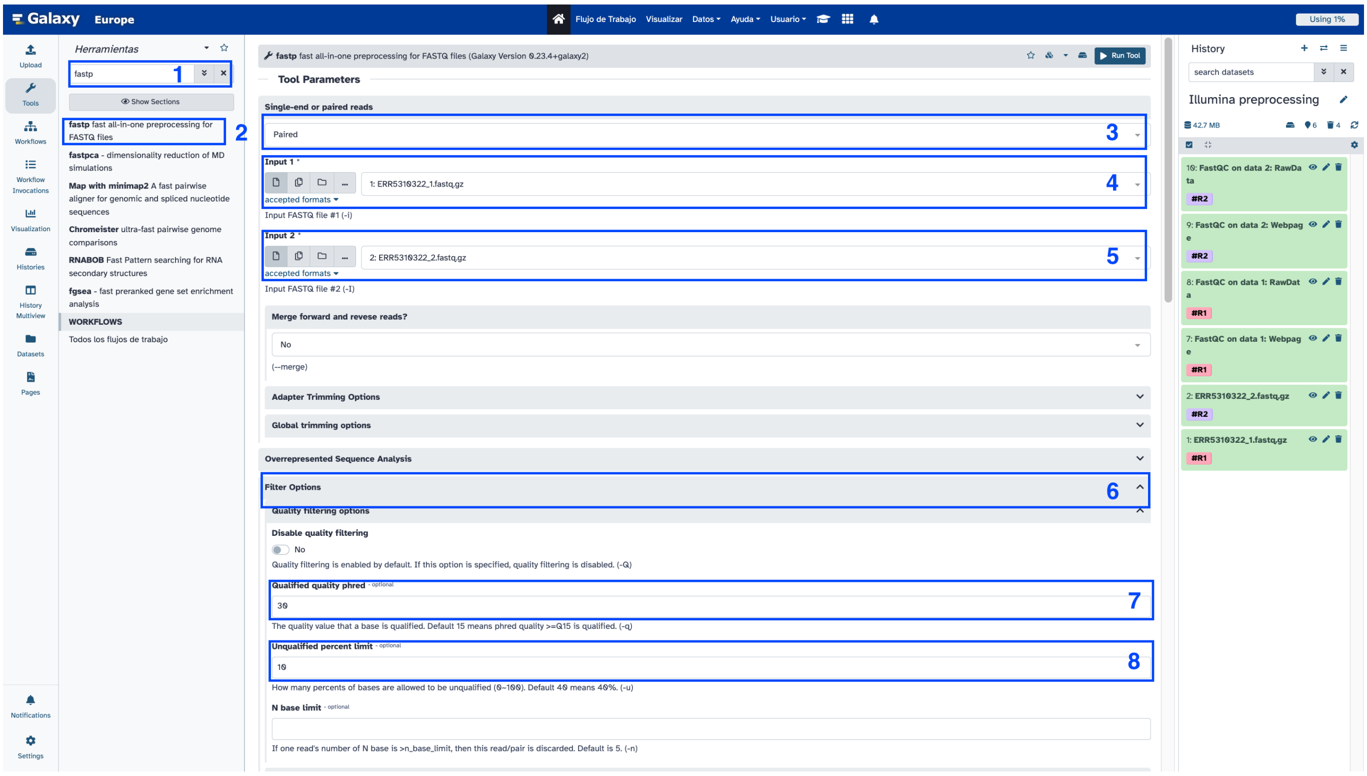
1.Search for **fastp** in the tools

2.Then select **fastp - fast all-in-one preprocessing for FASTQ files**

-Select custom parameters:

```
3.Single-end or paired reads > Paired

    4.Input 1 > Browse datasets (right folder icon) > Select
ERR5310322_1.fastq.gz

    5.Input 2 > Browse datasets > Select ERR5310322_2.fastq.gz

6.Display Filter Options

    -Quality Filtering options

        7.Qualified Quality Phred = 30

        8.Unqualified percent limit = 10

    -Length Filtering Options

        9.Length required = 50

10.Read modification options

    11.PoliX tail trimming > Enable polyX tail trimming

    -Per read cutting by quality options

        12.Cut by quality in front (5') > Yes

        13.Cut by quality in tail (3') > Yes

        14.Cutting mean quality = 30
```

15.Finally, click on **Run tool**



fastp

**fastp** fast all-in-one preprocessing for FASTQ files

**fastpca** - dimensionality reduction of MD simulations

**Map with minimap2** A fast pairwise aligner for genomic and spliced nucleotide sequences

**Chromeister** ultra-fast pairwise genome comparisons

**RNABOB** Fast Pattern searching for RNA secondary structures

**fgsea** - fast preranked gene set enrichment analysis

WORKFLOWS

Todos los flujos de trabajo

**fastp** fast all-in-one preprocessing for FASTQ files (Galaxy Version 0.23.4+galaxy2)

Tool Parameters

Single-end or paired reads: Paired

Input 1 *: 1: ERR5310322_1.fastq.gz
accepted formats
Input FASTQ file #1 (-i)

Input 2 *: 2: ERR5310322_2.fastq.gz
accepted formats
Input FASTQ file #2 (-I)

Merge forward and reverse reads?: No
(--merge)

Adapter Trimming Options

Global trimming options

Overrepresented Sequence Analysis

Filter Options

Quality filtering options

Disable quality filtering: No
Quality filtering is enabled by default. If this option is specified, quality filtering is disabled. (-Q)

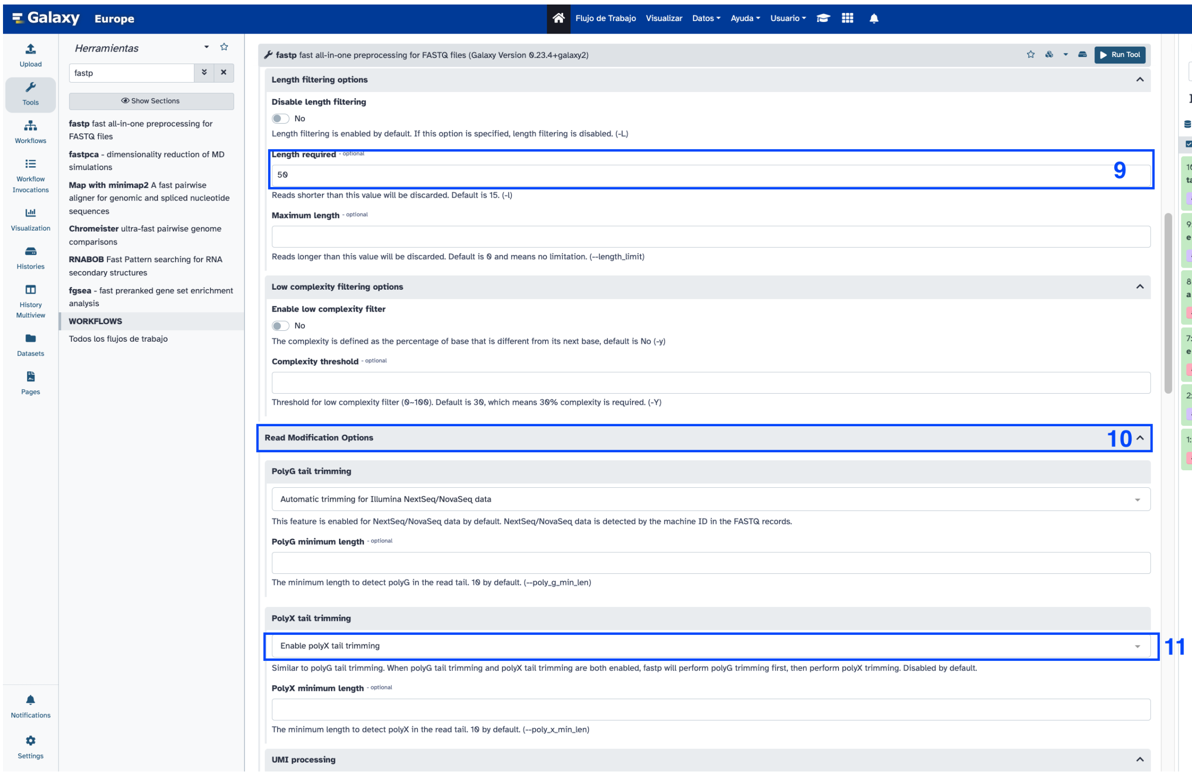
Qualified quality phred - optional: 30
The quality value that a base is qualified. Default 15 means phred quality >=Q15 is qualified. (-q)

Unqualified percent limit - optional: 10
How many percents of bases are allowed to be unqualified (0~100). Default 40 means 40%. (-u)

N base limit - optional
If one read's number of N base is >n_base_limit, then this read/pair is discarded. Default is 5. (-n)

Illumina preprocessing
42.7 MB 6 4

10: FastQC on data 2: RawData #R2
9: FastQC on data 2: Webpage #R2
8: FastQC on data 1: RawData #R1
7: FastQC on data 1: Webpage #R1
2: ERR5310322_2.fastq.gz #R2
1: ERR5310322_1.fastq.gz #R1



**fastp** fast all-in-one preprocessing for FASTQ files (Galaxy Version 0.23.4+galaxy2)

Length filtering options

Disable length filtering: No
Length filtering is enabled by default. If this option is specified, length filtering is disabled. (-L)

Length required - optional: 50
Reads shorter than this value will be discarded. Default is 15. (-l)

Maximum length - optional
Reads longer than this value will be discarded. Default is 0 and means no limitation. (--length_limit)

Low complexity filtering options

Enable low complexity filter: No
The complexity is defined as the percentage of base that is different from its next base, default is No (-y)

Complexity threshold - optional
Threshold for low complexity filter (0~100). Default is 30, which means 30% complexity is required. (-Y)

Read Modification Options

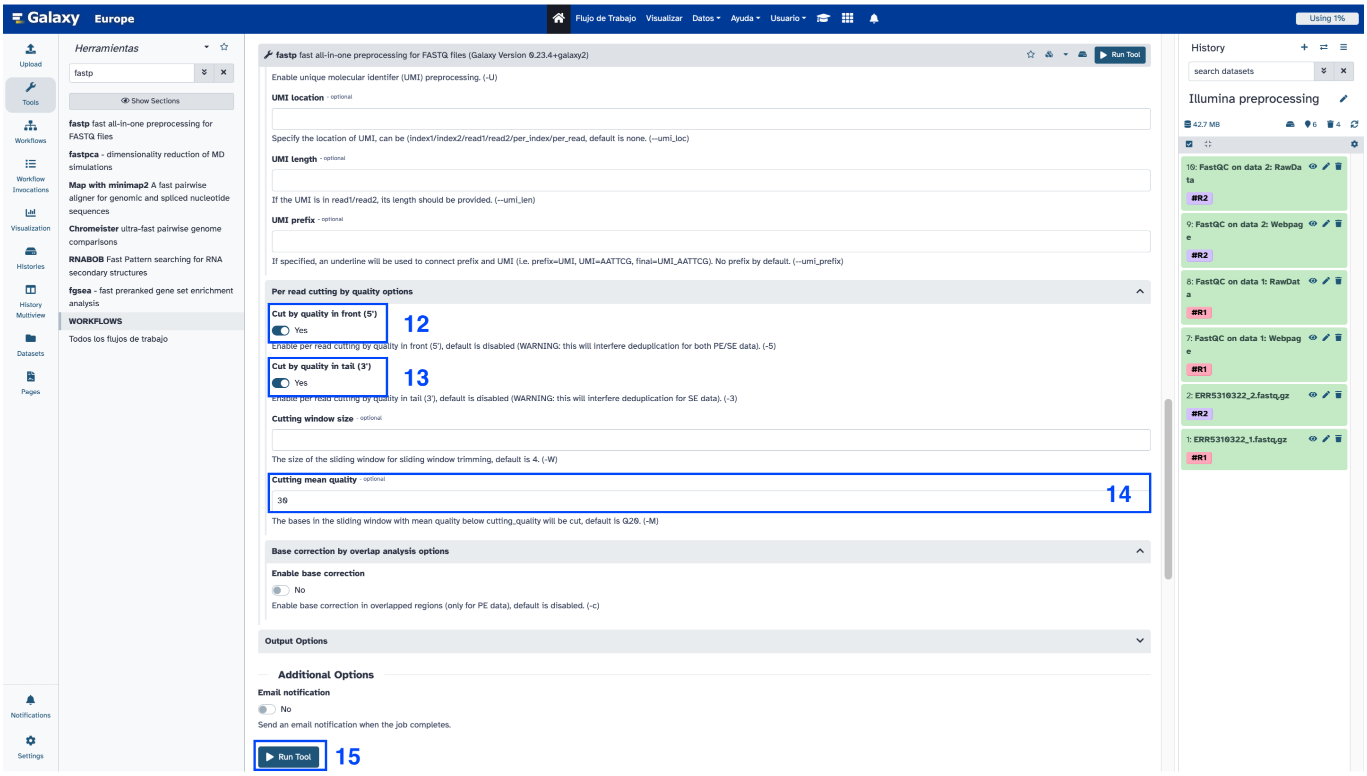
PolyG tail trimming: Automatic trimming for Illumina NextSeq/NovaSeq data
This feature is enabled for NextSeq/NovaSeq data by default. NextSeq/NovaSeq data is detected by the machine ID in the FASTQ records.

PolyG minimum length - optional
The minimum length to detect polyG in the read tail. 10 by default. (--poly_g_min_len)

PolyX tail trimming: Enable polyX tail trimming
Similar to polyG tail trimming. When polyG tail trimming and polyX tail trimming are both enabled, fastp will perform polyG trimming first, then perform polyX trimming. Disabled by default.

PolyX minimum length - optional
The minimum length to detect polyX in the read tail. 10 by default. (--poly_x_min_len)

UMI processing


Galaxy Europe
Flujo de Trabajo
Visualizar
Datos
Ayuda
Usuario
Using 1%
Upload
Tools
Workflows
Workflow Invocations
Visualization
Histories
History Multiview
Datasets
Pages
Notifications
Settings
Herramientas
fastp
Show Sections
fastp fast all-in-one preprocessing for FASTQ files
fastpca - dimensionality reduction of MD simulations
Map with minimap2 A fast pairwise aligner for genomic and spliced nucleotide sequences
Chromeister ultra-fast pairwise genome comparisons
RNABOB Fast Pattern searching for RNA secondary structures
fgsea - fast preranked gene set enrichment analysis
WORKFLOWS
Todos los flujos de trabajo
History
search datasets
Illumina preprocessing
42.7 MB
6
4
10: FastQC on data 2: RawData
#R2
9: FastQC on data 2: Webpage
#R2
8: FastQC on data 1: RawData
#R1
7: FastQC on data 1: Webpage
#R1
2: ERR5310322_2.fastq.gz
#R2
1: ERR5310322_1.fastq.gz
#R1
12
13
14
15

**fastp** fast all-in-one preprocessing for FASTQ files (Galaxy Version 0.23.4+galaxy2) ▶ Run Tool

Enable unique molecular identifer (UMI) preprocessing. (-U)

**UMI location** - optional

Specify the location of UMI, can be (index1/index2/read1/read2/per_index/per_read, default is none. (--umi_loc)

**UMI length** - optional

If the UMI is in read1/read2, its length should be provided. (--umi_len)

**UMI prefix** - optional

If specified, an underline will be used to connect prefix and UMI (i.e. prefix=UMI, UMI=AATTCG, final=UMI_AATTCG). No prefix by default. (--umi_prefix)

**Per read cutting by quality options**

**Cut by quality in front (5')**
Yes

Enable per read cutting by quality in front (5'), default is disabled (WARNING: this will interfere deduplication for both PE/SE data). (-5)

**Cut by quality in tail (3')**
Yes

Enable per read cutting by quality in tail (3'), default is disabled (WARNING: this will interfere deduplication for SE data). (-3)

**Cutting window size** - optional

The size of the sliding window for sliding window trimming, default is 4. (-W)

**Cutting mean quality** - optional
30

The bases in the sliding window with mean quality below cutting_quality will be cut, default is Q20. (-M)

**Base correction by overlap analysis options**

**Enable base correction**
No

Enable base correction in overlapped regions (only for PE data), default is disabled. (-c)

**Output Options**

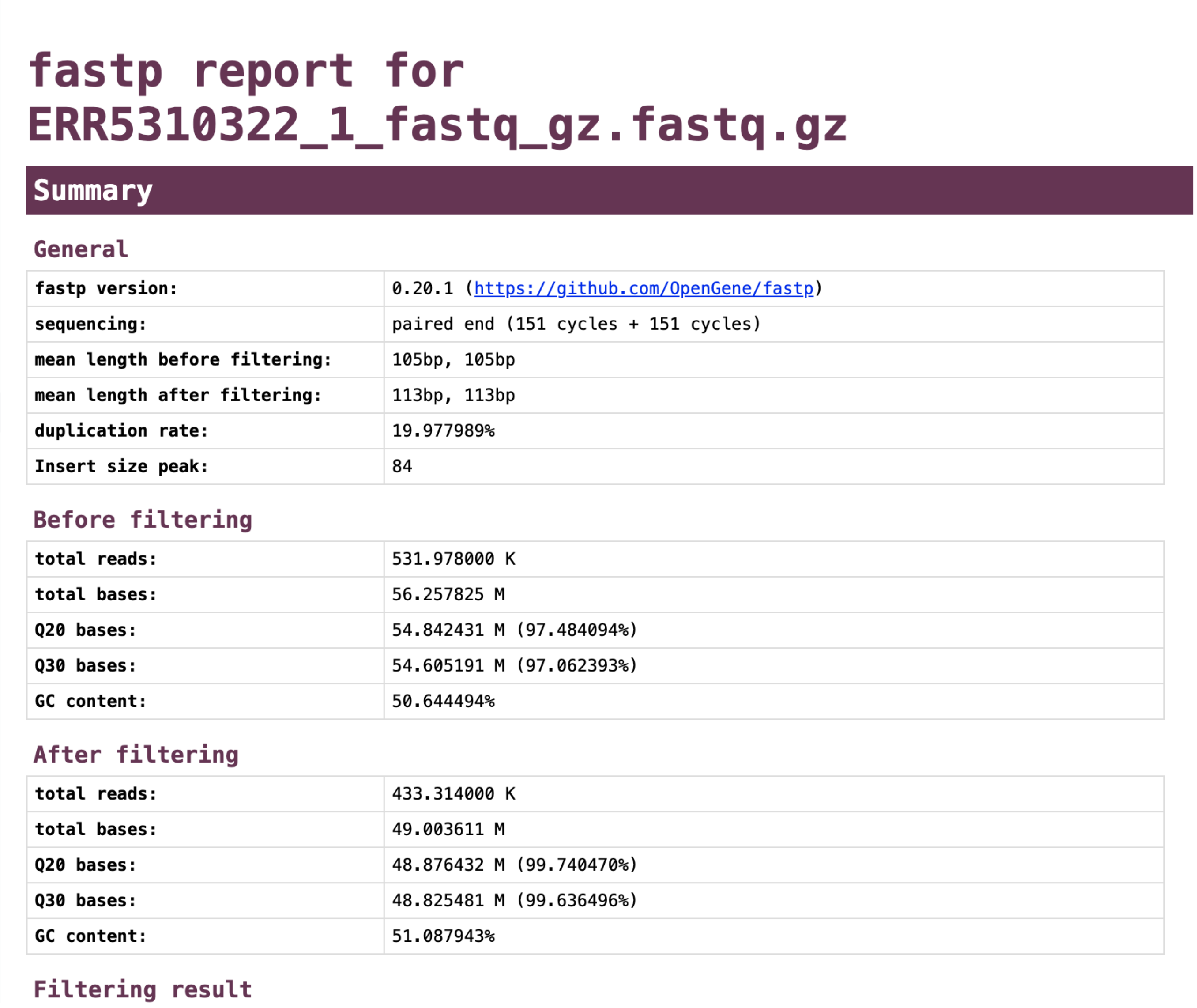
**Additional Options**

**Email notification**
No

Send an email notification when the job completes.

▶ Run Tool

To see the trimming stats, have a look at the **fastp on data 2 and data 1: HTML report** file. You should see something like that.



# fastp report for ERR5310322_1_fastq_gz.fastq.gz

## Summary

### General

| | |
|---|---|
| fastp version: | 0.20.1 (https://github.com/OpenGene/fastp) |
| sequencing: | paired end (151 cycles + 151 cycles) |
| mean length before filtering: | 105bp, 105bp |
| mean length after filtering: | 113bp, 113bp |
| duplication rate: | 19.977989% |
| Insert size peak: | 84 |

### Before filtering

| | |
|---|---|
| total reads: | 531.978000 K |
| total bases: | 56.257825 M |
| Q20 bases: | 54.842431 M (97.484094%) |
| Q30 bases: | 54.605191 M (97.062393%) |
| GC content: | 50.644494% |

### After filtering

| | |
|---|---|
| total reads: | 433.314000 K |
| total bases: | 49.003611 M |
| Q20 bases: | 48.876432 M (99.740470%) |
| Q30 bases: | 48.825481 M (99.636496%) |
| GC content: | 51.087943% |

### Filtering result

▶ How many reads have we lost?

#### 1.2.2. Other trimming tools: Trimmomatic

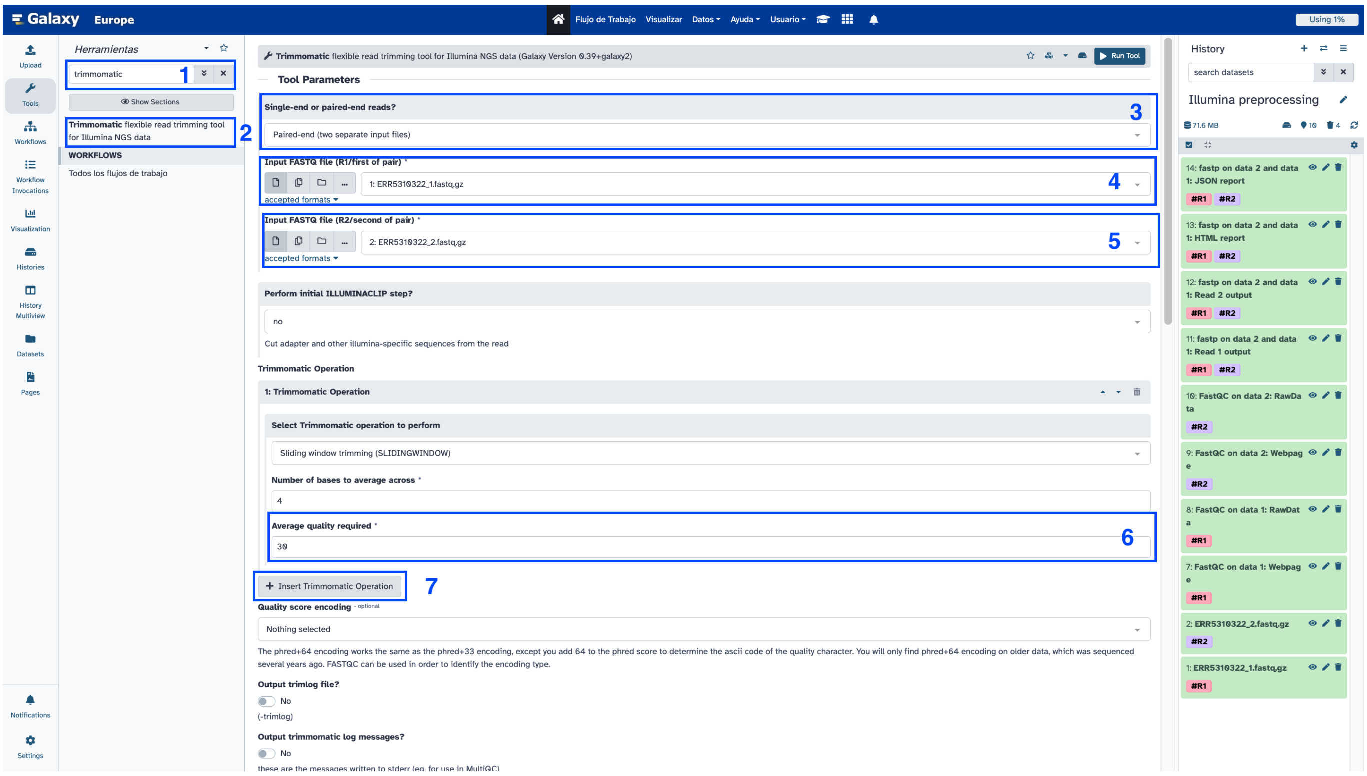
1.Search for **trimmomatic** in the tools

2.Select **Trimmomatic flexible read trimming tool for Illumina NGS data**

-Select custom parameters:

```
3.Single-end or paired-end reads? = Paired-end (two separated files)

4.Input FASTQ file (R1/first of pair) = ERR5310322_1.fastq.gz

5.Input FASTQ file (R2/second of pair) = ERR5310322_2.fastq.gz

6.Average quality required = 30

7.Insert Trimmomatic Operation:

    8.Select Trimmomatic operation to perform: **MINLEN**
    9.Minimum length of reads to be kept = 50
```

10.Select **Run tool**



trimmomatic

**Trimmomatic** flexible read trimming tool for Illumina NGS data

**Trimmomatic** flexible read trimming tool for Illumina NGS data (Galaxy Version 0.39+galaxy2)

**Tool Parameters**

**Single-end or paired-end reads?**
Paired-end (two separate input files)

**Input FASTQ file (R1/first of pair)** *
1: ERR5310322_1.fastq.gz
accepted formats

**Input FASTQ file (R2/second of pair)** *
2: ERR5310322_2.fastq.gz
accepted formats

**Perform initial ILLUMINACLIP step?**
no
Cut adapter and other illumina-specific sequences from the read

**Trimmomatic Operation**

**1: Trimmomatic Operation**

**Select Trimmomatic operation to perform**
Sliding window trimming (SLIDINGWINDOW)

**Number of bases to average across** *
4

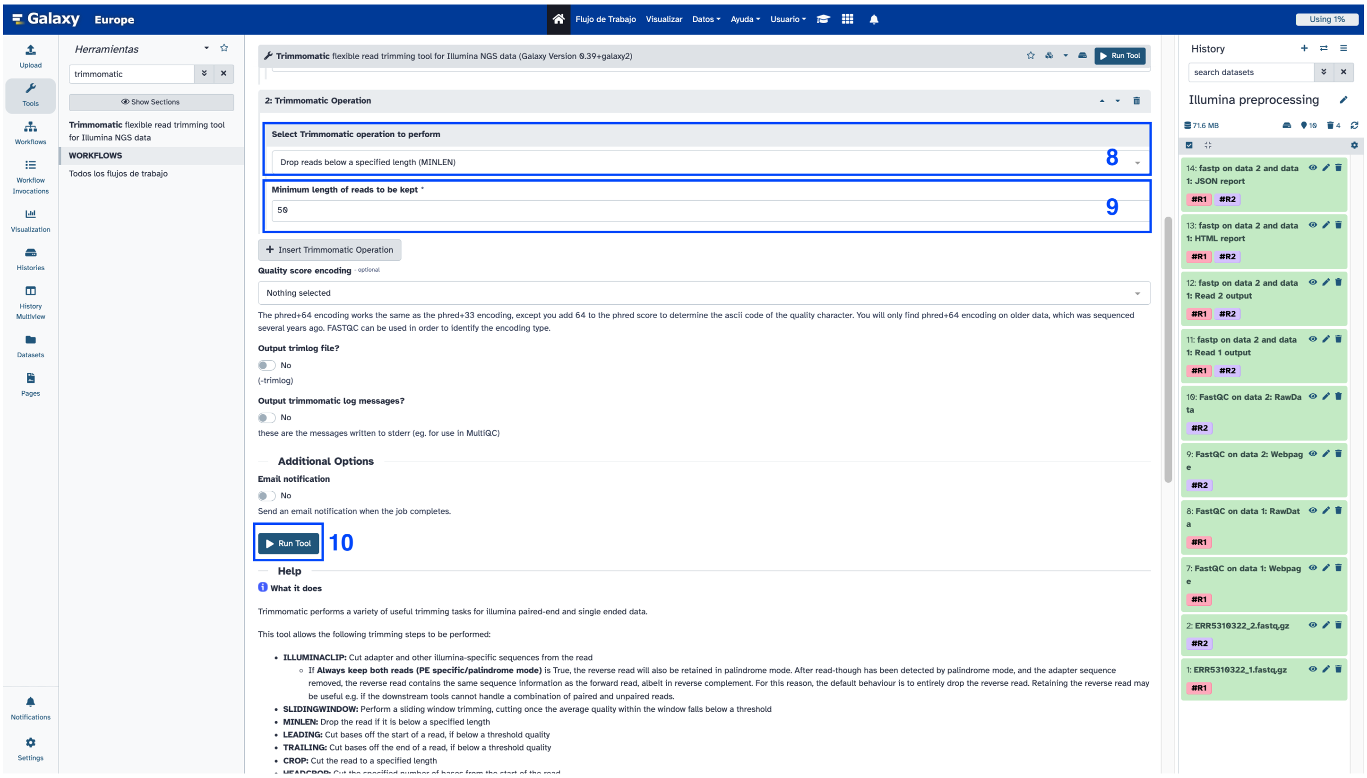
**Average quality required** *
30

Insert Trimmomatic Operation

**Quality score encoding** - optional
Nothing selected
The phred+64 encoding works the same as the phred+33 encoding, except you add 64 to the phred score to determine the ascii code of the quality character. You will only find phred+64 encoding on older data, which was sequenced several years ago. FASTQC can be used in order to identify the encoding type.

**Output trimlog file?**
No
(-trimlog)

**Output trimmomatic log messages?**
No
these are the messages written to stderr (eg. for use in MultiQC)



trimmomatic

**Trimmomatic** flexible read trimming tool for Illumina NGS data

**Trimmomatic** flexible read trimming tool for Illumina NGS data (Galaxy Version 0.39+galaxy2)

**2: Trimmomatic Operation**

**Select Trimmomatic operation to perform**
Drop reads below a specified length (MINLEN)

**Minimum length of reads to be kept** *
50

Insert Trimmomatic Operation

**Quality score encoding** - optional
Nothing selected
The phred+64 encoding works the same as the phred+33 encoding, except you add 64 to the phred score to determine the ascii code of the quality character. You will only find phred+64 encoding on older data, which was sequenced several years ago. FASTQC can be used in order to identify the encoding type.

**Output trimlog file?**
No
(-trimlog)

**Output trimmomatic log messages?**
No
these are the messages written to stderr (eg. for use in MultiQC)

**Additional Options**

**Email notification**
No
Send an email notification when the job completes.

Run Tool

**Help**

**What it does**

Trimmomatic performs a variety of useful trimming tasks for illumina paired-end and single ended data.

This tool allows the following trimming steps to be performed:

- **ILLUMINACLIP:** Cut adapter and other illumina-specific sequences from the read
    - If **Always keep both reads (PE specific/palindrome mode)** is True, the reverse read will also be retained in palindrome mode. After read-though has been detected by palindrome mode, and the adapter sequence removed, the reverse read contains the same sequence information as the forward read, albeit in reverse complement. For this reason, the default behaviour is to entirely drop the reverse read. Retaining the reverse read may be useful e.g. if the downstream tools cannot handle a combination of paired and unpaired reads.
- **SLIDINGWINDOW:** Perform a sliding window trimming, cutting once the average quality within the window falls below a threshold
- **MINLEN:** Drop the read if it is below a specified length
- **LEADING:** Cut bases off the start of a read, if below a threshold quality
- **TRAILING:** Cut bases off the end of a read, if below a threshold quality
- **CROP:** Cut the read to a specified length
- **HEADCROP:** Cut the specified number of bases from the start of the read

Trimmomatic does not perform statistics over trimmed reads, so we need to perform FastQC again over the Trimmomatic results.

▶ Try to do it on your own.

#### Second question

▶ How can I improve the quality of my data?

- This hands-on history URL: https://usegalaxy.eu/u/svarona/h/illumina-preprocessing

## 2. Nanopore Quality control and trimming

| Title | Galaxy |
|---|---|
| **Training dataset:** | The data we are going to manage corresponds to Nanopore amplicon sequencing data using ARTIC network primers por SARS-CoV-2 genome. From the Fast5 files generated by the ONT software, we are going to select the pass reads, so they are already filtered by quality. |
| **Questions:** | • How do I know if my Nanopore data was correctly sequenced? |
| **Objectives**: | • Perform a quality control in raw Illumina reads<br>• Perform a quality trimming in raw Nanopore reads<br>• Perform a quality control in trimmed Nanopore reads |
| **Estimated time**: | 15 min |

### 2.1. Quality control

To run the quality control over the samples, follow these steps:

1. Create a new history has explained yesterday named **Nanopore quality**
2. Upload data as seen yesterday, copy and paste the following URLs:

```
https://raw.githubusercontent.com/nf-core/test-datasets/viralrecon/nanopore/minion/fastq_pass/barcode01/FAO93606_pass_barcode01_7650855b_0.fastq
https://raw.githubusercontent.com/nf-core/test-datasets/viralrecon/nanopore/minion/fastq_pass/barcode01/FAO93606_pass_barcode01_7650855b_1.fastq
https://raw.githubusercontent.com/nf-core/test-datasets/viralrecon/nanopore/minion/fastq_pass/barcode01/FAO93606_pass_barcode01_7650855b_2.fastq
```

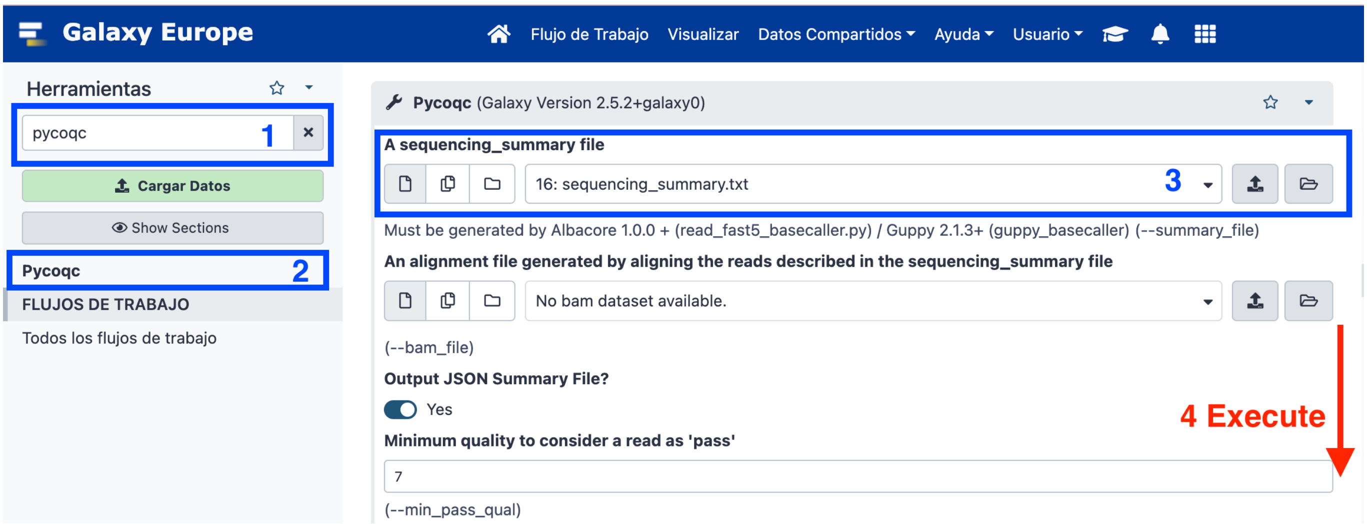
#### 2.1.1. PycoQC

To use PycoQC we need to use the `sequencing_summary.txt` provided by de Nanopore sequencing machine.

Upload data as seen yesterday, copy and paste the following URL:

```
https://raw.githubusercontent.com/nf-core/test-
datasets/viralrecon/nanopore/minion/sequencing_summary.txt
```

1. Search for the **Pycoqc** tool
2. Select **Pycoqc quality control for Nanopore sequencing data**
3. In *A sequencing_summary file*: Select the `sequencing_summary.txt` we just uploaded
4. Select **Run tool**



pycoqc

Pycoqc

**Pycoqc** (Galaxy Version 2.5.2+galaxy0)

**A sequencing_summary file**

16: sequencing_summary.txt

Must be generated by Albacore 1.0.0 + (read_fast5_basecaller.py) / Guppy 2.1.3+ (guppy_basecaller) (--summary_file)

**An alignment file generated by aligning the reads described in the sequencing_summary file**

No bam dataset available.

(--bam_file)

**Output JSON Summary File?**

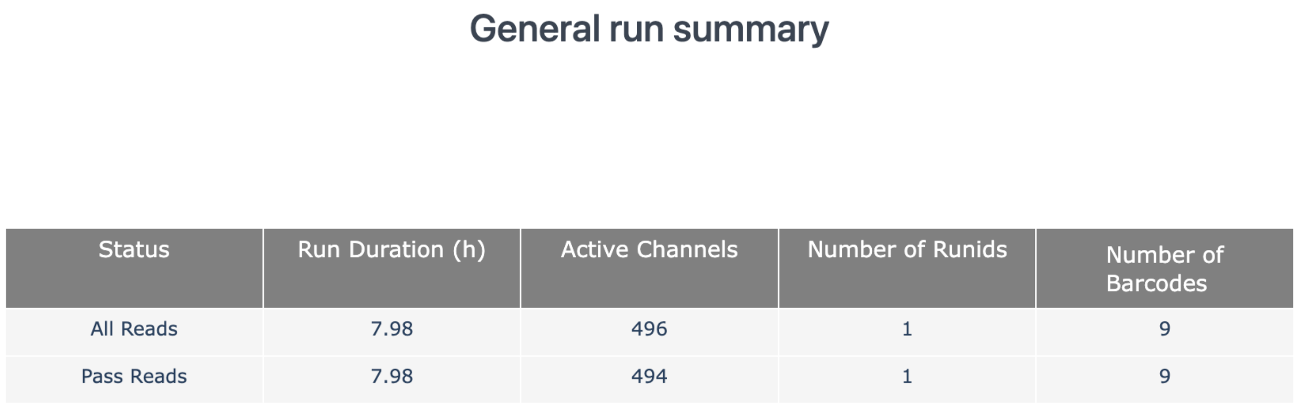
Yes

**Minimum quality to consider a read as 'pass'**

7

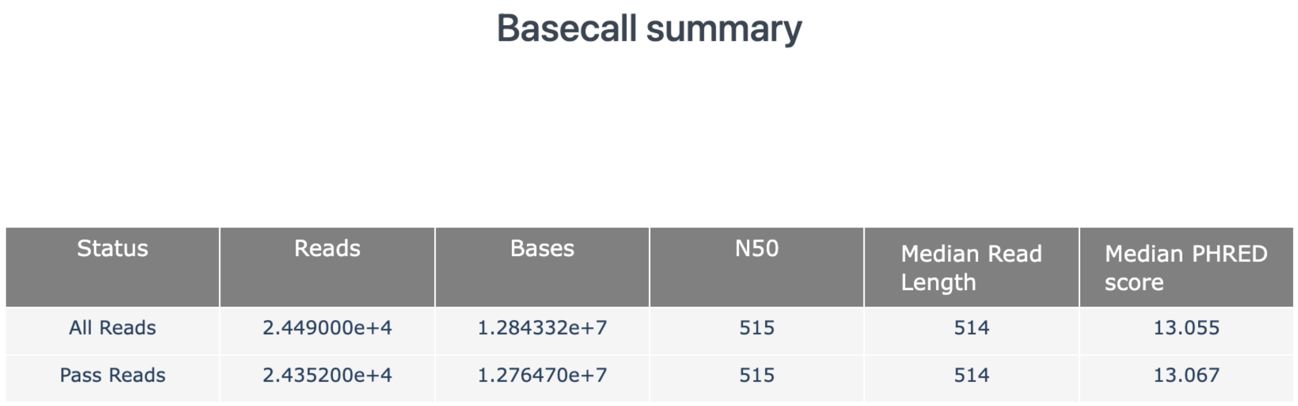
(--min_pass_qual)

Then inspect the resulting PycoQC HTML Report:



## General run summary

| Status | Run Duration (h) | Active Channels | Number of Runids | Number of Barcodes |
|---|---|---|---|---|
| All Reads | 7.98 | 496 | 1 | 9 |
| Pass Reads | 7.98 | 494 | 1 | 9 |



## Basecall summary

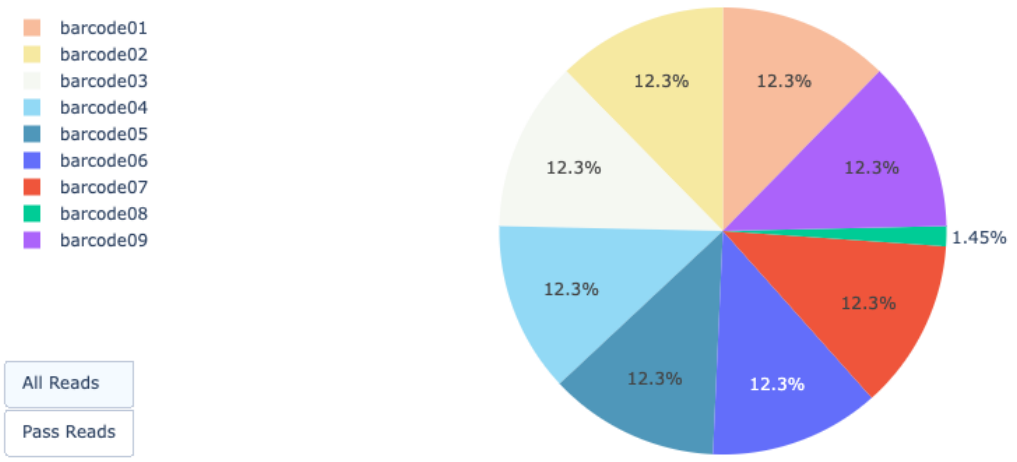
| Status | Reads | Bases | N50 | Median Read Length | Median PHRED score |
|---|---|---|---|---|---|
| All Reads | 2.449000e+4 | 1.284332e+7 | 515 | 514 | 13.055 |
| Pass Reads | 2.435200e+4 | 1.276470e+7 | 515 | 514 | 13.067 |

**Question**

▶ How many reads do the samples have?

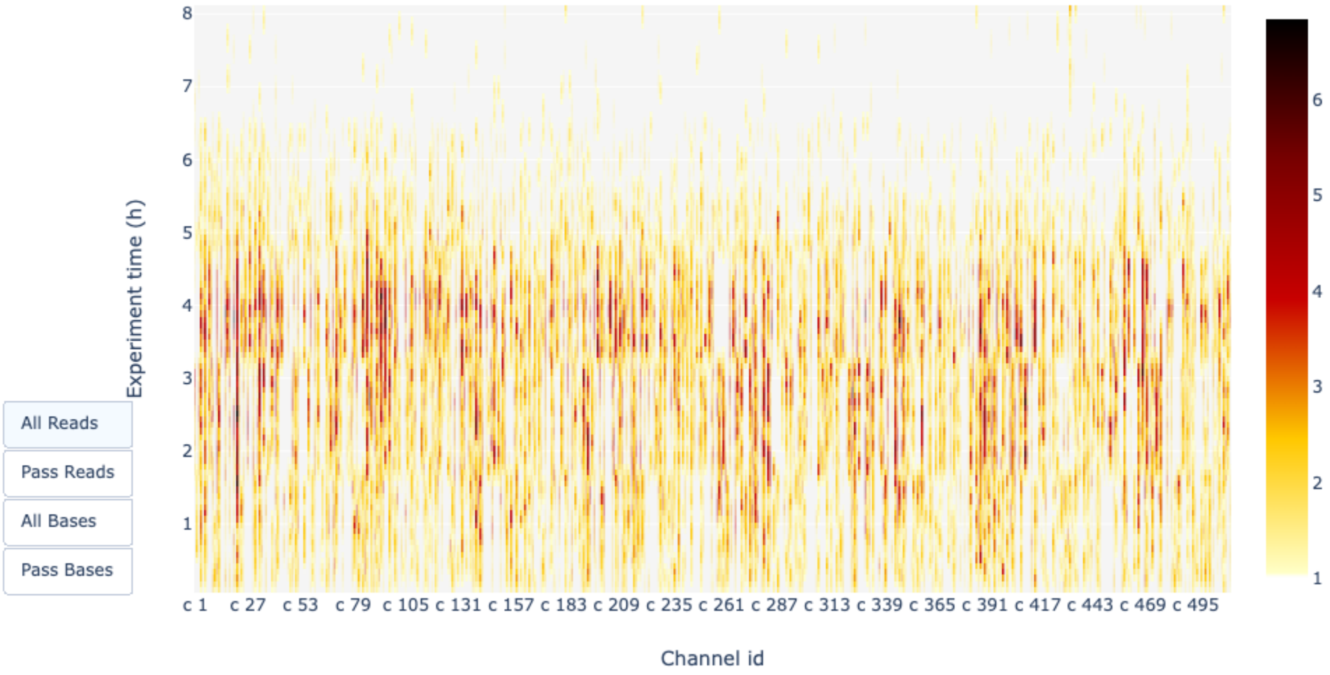
▶ Do you understand all the plots?

**Number of reads per barcode**:



This plot shows the number of reads per barcode, which means de number of reads per sample to be demultiplexed. In a goog experiment, all the barcodes should have the same number of reads. In this training we only used reads from barcode01 sample but we can see that barcode08 couldn't be correctly sequenced.

**Channel activity over time:**



It gives an overview of available pores, pore usage during the experiment, inactive pores and shows if the loading of the flow cell is good (almost all pores are used). In this case, the vast majority of channels/pores are inactive (white) after the 6h of experiment, so the run should have been dinished at that time. You would hope for a plot that it is dark near the X-axis, and with higher Y-values (increasing time) doesn't get too light/white. Depending if you chose "Reads" or "Bases" on the left the colour indicates either number of bases or reads per time interval.

▶ How do I check whether my Nanopore data was correctly sequenced?

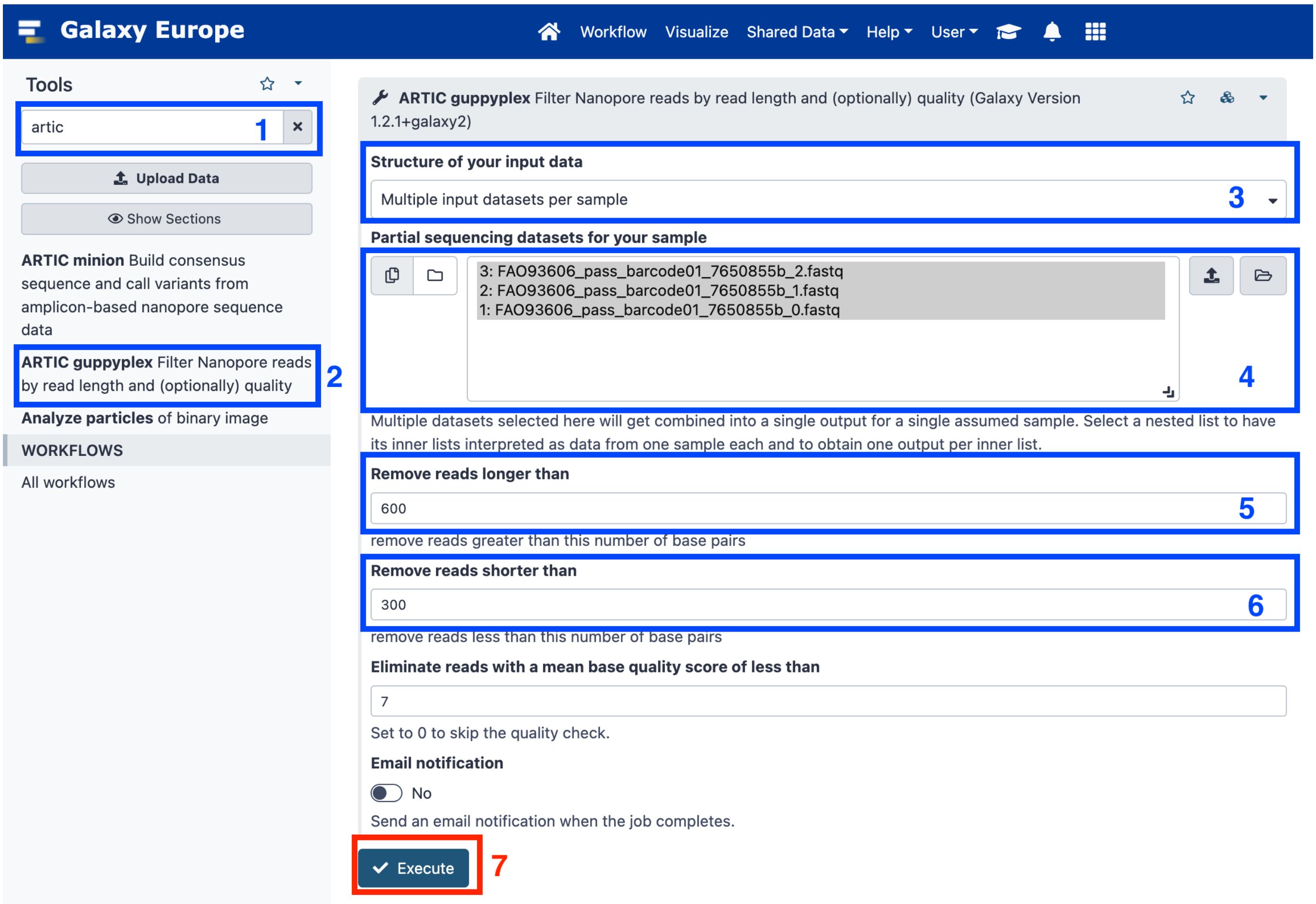
## 2.2. Trimming

When Nanopore reads are being sequenced, the MinKnown software splits Fast5 reads into quality pass and quality fail. As we will select only Fast5 pass reads, we won't need to perform a quality trimming, so even if we see that the reads have a bad Phred score, we know that the ONT software considered the reads as "good quality".

Then we will only be performing a read length trimming. As we are using amplicon sequencing data, we won't be expecting reads smaller than 400 nucleotides, nor higher than 600, which would obviously correspond to chimeric reads.

### 2.2.1. Artic

1. Search for **artic** tool
2. Select **ARTIC guppyplex Filter Nanopore reads by read length and (optionally) quality**
3. Structure of your input data: Multiple input datasets per sample
4. While pressing the *Ctrl* key, select the three samples
5. Remove reads longer than = 600
6. Remove reads shorter than = 300
7. Do not filter on quality score (speeds up processing) = Yes (we had already select pass reads)



artic

**ARTIC minion** Build consensus sequence and call variants from amplicon-based nanopore sequence data

**ARTIC guppyplex** Filter Nanopore reads by read length and (optionally) quality

**Analyze particles** of binary image

**ARTIC guppyplex** Filter Nanopore reads by read length and (optionally) quality (Galaxy Version 1.2.1+galaxy2)

**Structure of your input data**
Multiple input datasets per sample

**Partial sequencing datasets for your sample**
3: FAO93606_pass_barcode01_7650855b_2.fastq
2: FAO93606_pass_barcode01_7650855b_1.fastq
1: FAO93606_pass_barcode01_7650855b_0.fastq

Multiple datasets selected here will get combined into a single output for a single assumed sample. Select a nested list to have its inner lists interpreted as data from one sample each and to obtain one output per inner list.

**Remove reads longer than**
600
remove reads greater than this number of base pairs

**Remove reads shorter than**
300
remove reads less than this number of base pairs

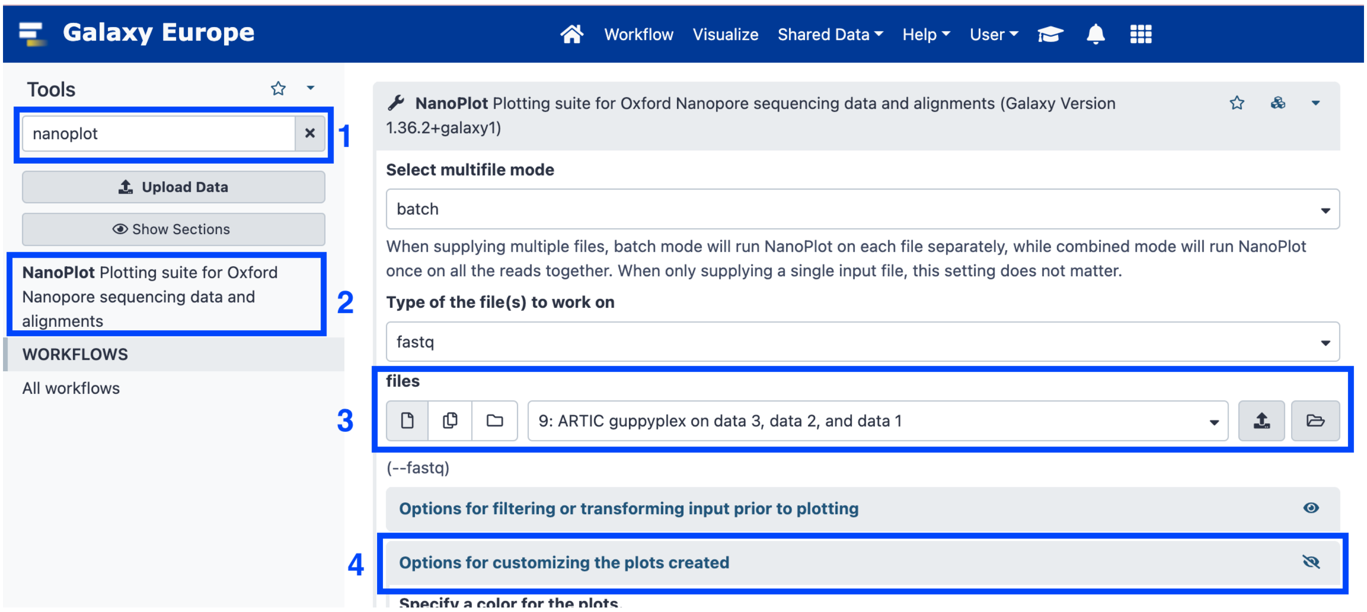
**Eliminate reads with a mean base quality score of less than**
7
Set to 0 to skip the quality check.

**Email notification**
No
Send an email notification when the job completes.

### 2.2.2. Nanoplot

Now we are going to run NanoPlot on filtered data:

1. Search for the **Nanoplot** tool and select **NanoPlot Plotting suite for Oxford Nanopore sequencing data and alignments**
2. Run the tool as follows:
   - In the *files* part, select ARTIC output file.
   - Display **Options for customizing the plots created**:
     - **Specify the bivariate format of the plots** > *Select all*
     - **Show the N50 mark in the read length histogram** > *Yes*
   - Select **Execute**



nanoplot

**NanoPlot** Plotting suite for Oxford Nanopore sequencing data and alignments

**NanoPlot** Plotting suite for Oxford Nanopore sequencing data and alignments (Galaxy Version 1.36.2+galaxy1)

**Select multifile mode**

batch

When supplying multiple files, batch mode will run NanoPlot on each file separately, while combined mode will run NanoPlot once on all the reads together. When only supplying a single input file, this setting does not matter.

**Type of the file(s) to work on**

fastq

**files**

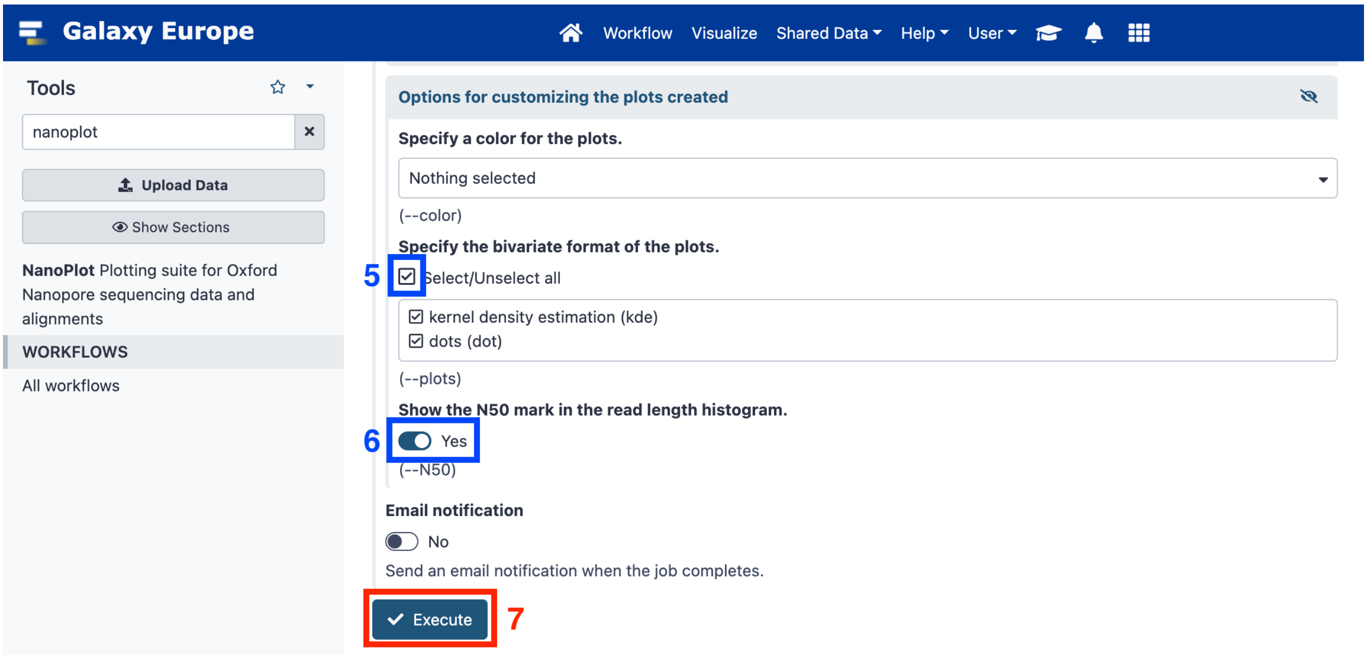
9: ARTIC guppyplex on data 3, data 2, and data 1

(--fastq)

**Options for filtering or transforming input prior to plotting**

**Options for customizing the plots created**

**Specify a color for the plots**



nanoplot

**NanoPlot** Plotting suite for Oxford Nanopore sequencing data and alignments

**Options for customizing the plots created**

**Specify a color for the plots.**

Nothing selected

(--color)

**Specify the bivariate format of the plots.**

☑ Select/Unselect all

☑ kernel density estimation (kde)
☑ dots (dot)

(--plots)

**Show the N50 mark in the read length histogram.**

● Yes

(--N50)

**Email notification**

○ No

Send an email notification when the job completes.

### Questions

- ▶ Did our data length and quality improve?
- ▶ How many reads did we lost during trimming step?

- This hands-on history URL: https://usegalaxy.eu/u/svarona/h/nanopore-quality

> ***NOTE:*** We can't use nanofilt because it is not installed in Galaxy