Galaxy for virologist training Exercise 4: Nanopore mapping 101

Title	Galaxy
Training dataset:	Nanopore MinION Sequencing of a Monkey Pox Virus (MPXV) from Spain 2022 oubreak. Data is publicly available at SRA with ID ERR10297654. Paper
Questions:	• How Nanopore reads are differently assembled from Illumina?
Objectives:	• Understand the concept of assembly • Learn how to interpret assembly quality control metrics
Estimated time:	40 min

1. Description

Nanopore techology is a third generation sequencing technique which allows to get longer sequences, but with reduced sequence quality. Different technologies have different formats, qualities, and specific known biases which make the analysis different among them. In this tutorial, we are going to see an example of how to assemble long reads from a Nanopore sequencing run.

2. Upload data to galaxy

Training dataset

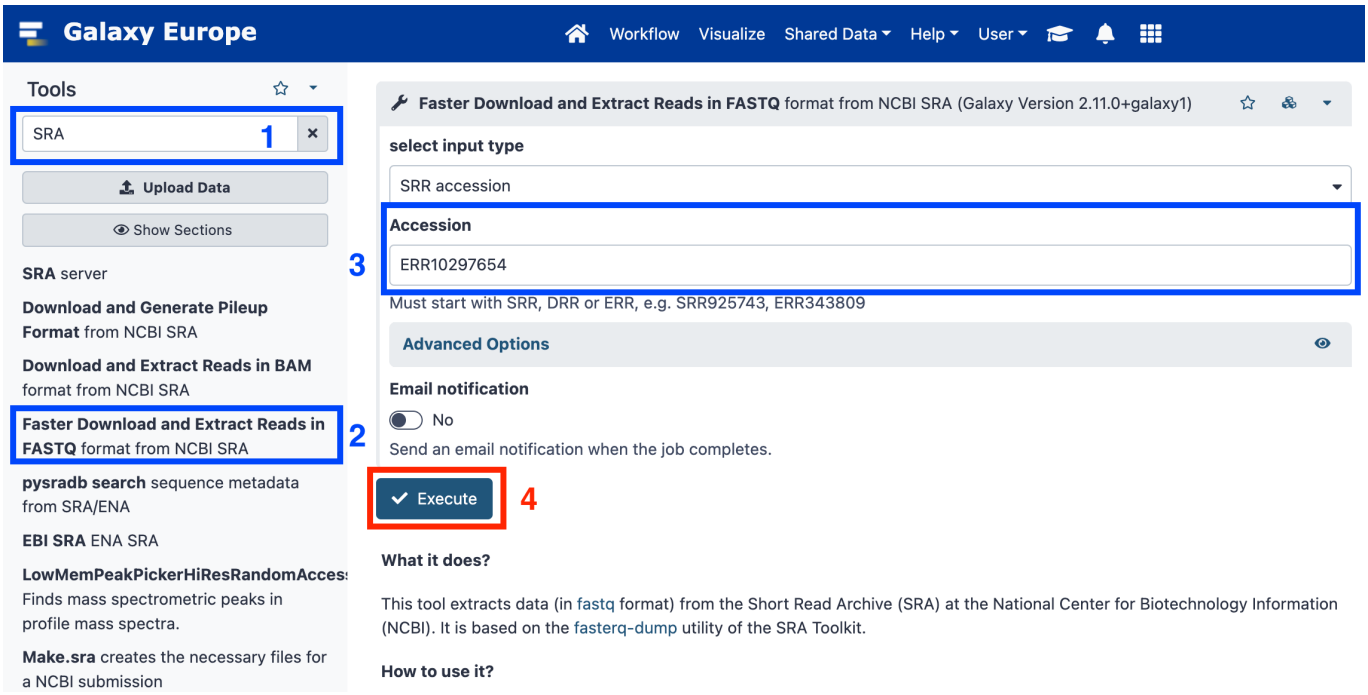
- [SRA ID: ERR10297654](https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=ERR10297654&display=metadata

Create new history

- Click the `+` icon at the top of the history panel and create a new history with the name `nanopore assembly 101 tutorial` as explained here

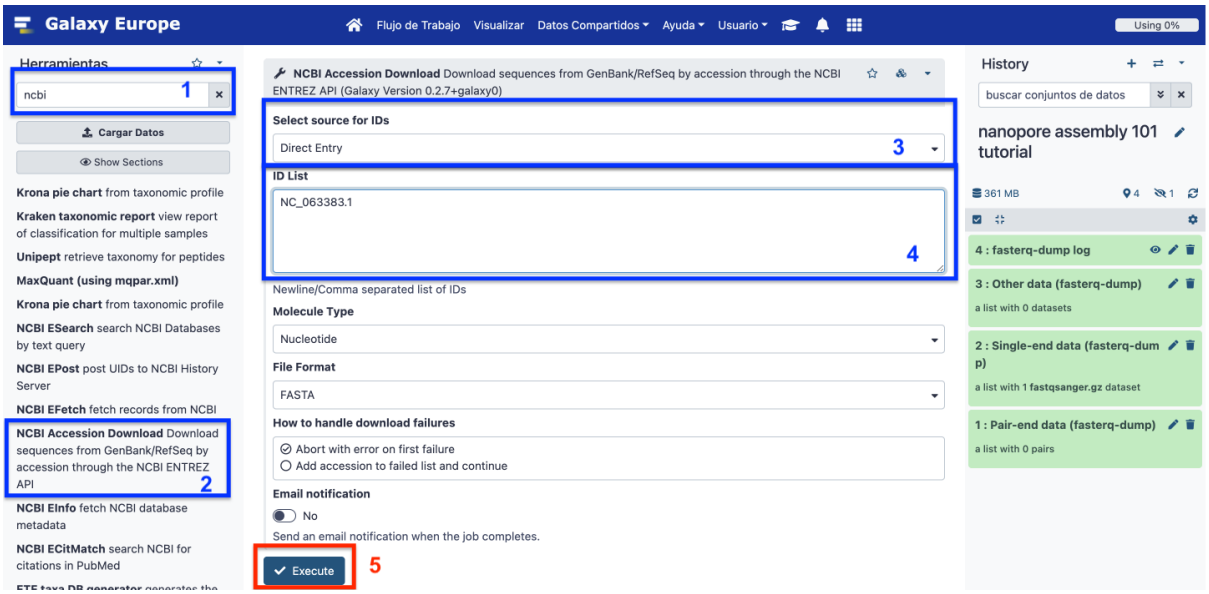
Upload data

1. Look for `SRA` in the tool search bar and select `Faster Download and Extract Reads in FASTQ format from NCBI SRA`
2. Accession = `ERR10297654`
3. Execute



Load reference file from NCBI

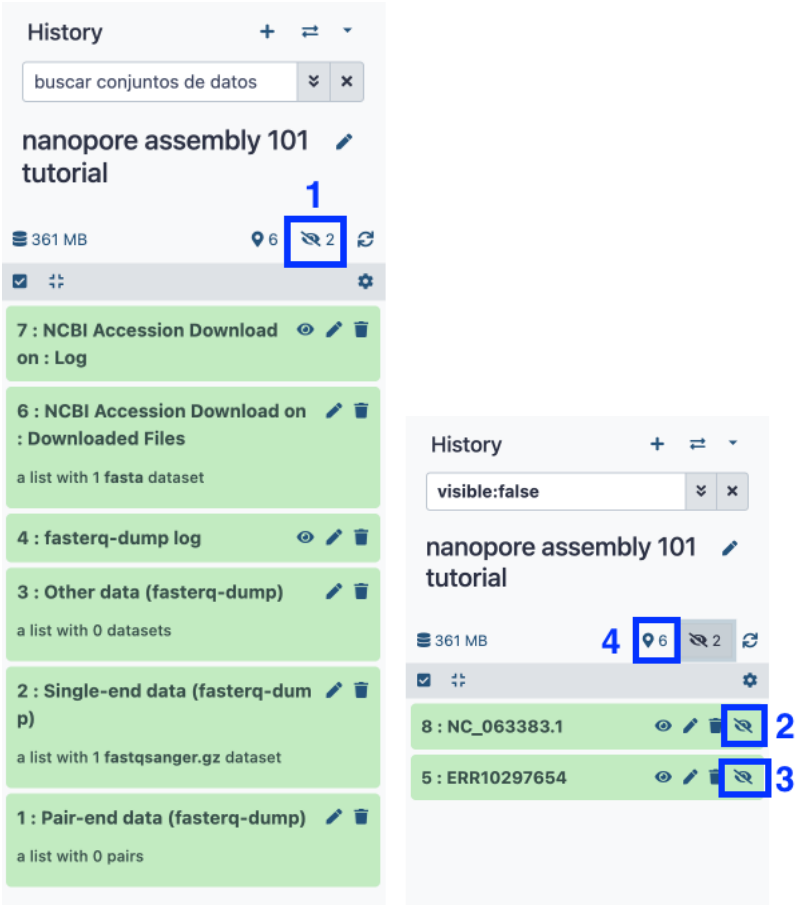
1. Search `NCBI` using the search toolbox and select `NCBI Accession Download Download sequences from GenBank/RefSeq by accession through the NCBI ENTREZ API`
2. Select source for IDs > Direct entry
3. ID List = NC_063383.1
4. Execute



Unhide data

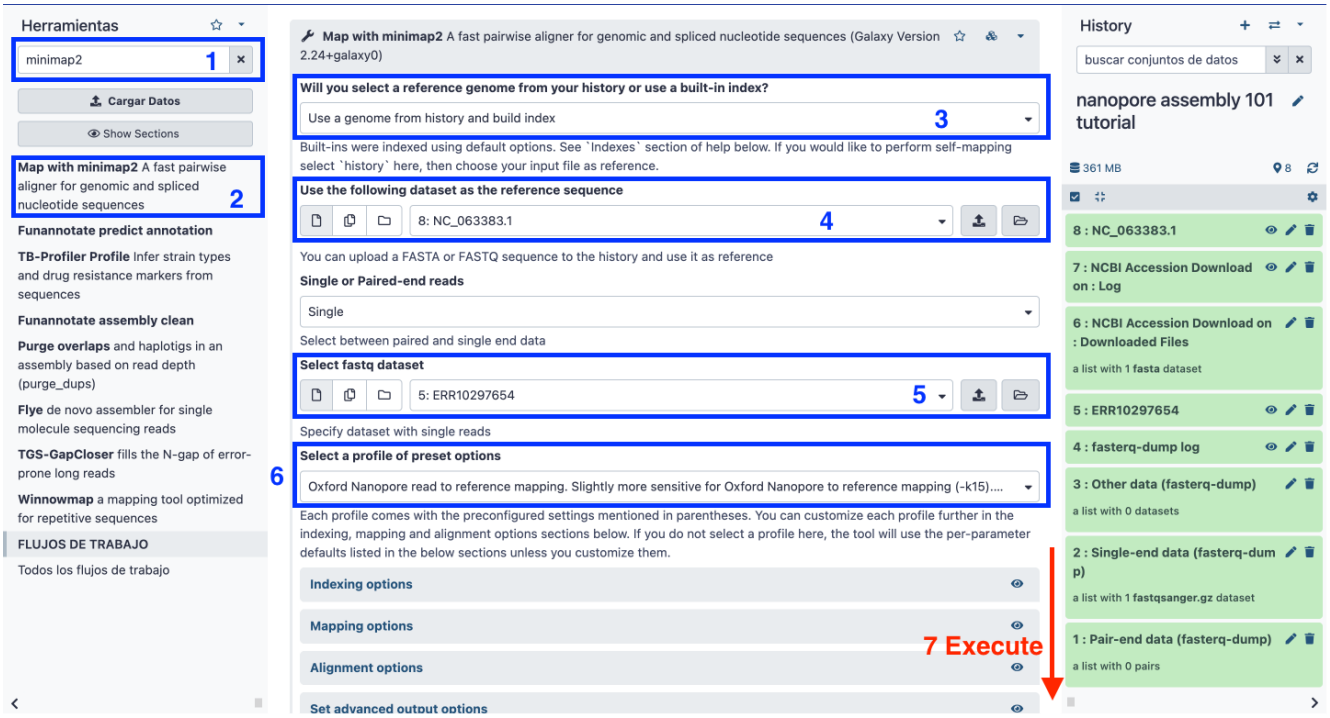
Using SRA and NCBI API downloads data as hidden so we are going to unhidde this data as follows:

1. Click on the strikethrough eye (Show hidden)
2. Select the strikethrough for ERR10297654 and NC_063383.1 datas.
3. Then select the location icon (show active)



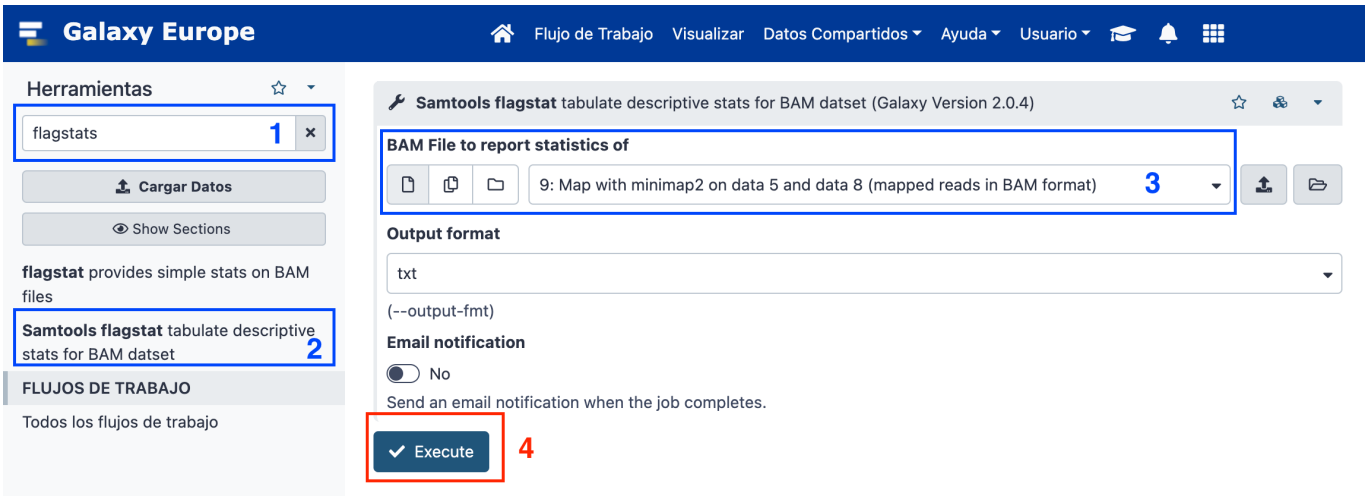
Mapping with Minimap2

1. Search `minimap2` using the search toolbox and select `Map with minimap2 A fast pairwise aligner for genomic and spliced nucleotide sequences`
2. Will you select a reference genome from your history or use a built-in index?: Use a genome from history and built-in index
    - Select NC_063383.1
3. Select fastq dataset: ERR10297654
4. Select a profile of preset options > Oxford Nanopore Read to reference mapping (map-ont)
5. Click execute and wait.



Mapping stats with samtools

1. Search `flagstatst` using the search toolbox and select `Samtools flagstat tabulate descriptive stats for BAM datset`
2. BAM File to report statistics of > Select Minimap2 bam output
3. Click execute and wait.
4. Click in the 👁 and see the bam stats.



▶ Which is the mapping rate?

▶ How many reads do we have in our dataset?

> This training history is available at: https://usegalaxy.eu/u/s.varona/h/nanopore-assembly-101-tutorial

> Note: Nanopore data is known to have more error than short sequencing reads. This is why assembly post-processing is strongly recommended, usually using combined sequencing aproximation with both Nanopore and Illumina reads.