

Gen2Epi GUI Quick Tutorial

Locations and Name of the Test DATA Used in this tutorial:

Main Data directory: “/home/gen2epi/Desktop/Test_DATA”

Fastq files: “/home/gen2epi/Desktop/Test_DATA/WHO_Data”

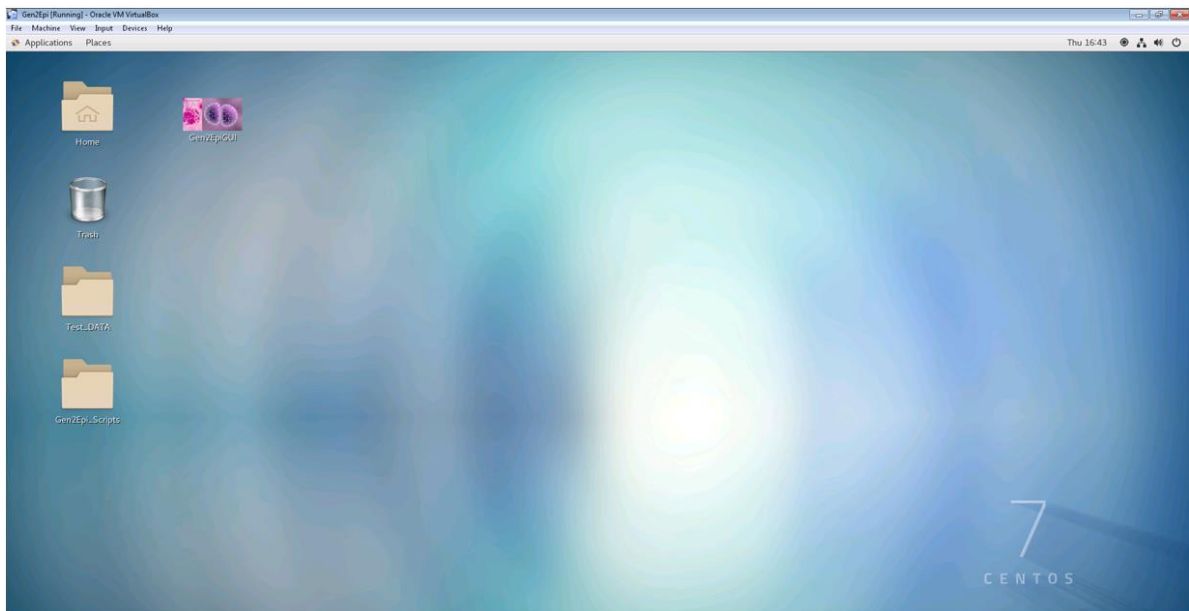
Genome files:

“/home/gen2epi/Desktop/Test_DATA/WHO_Full_Reference_genome/Chromosome”

Annotation files:

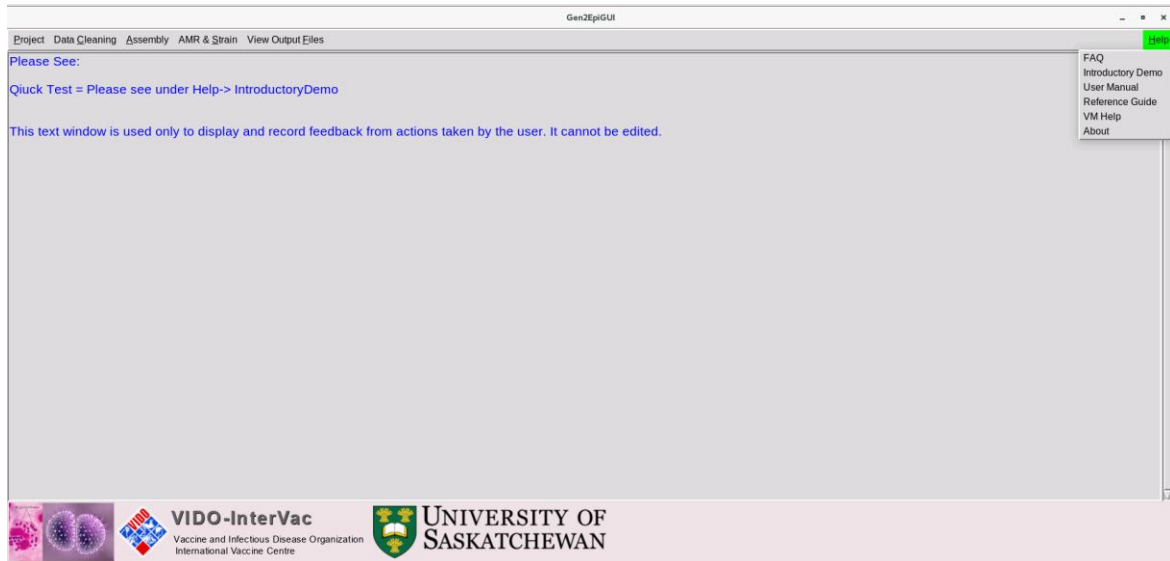
“/home/gen2epi/Desktop/Test_DATA/WHO_Genome_Annotation/Chromosome”

- 1) Click on the “Gen2EpiGUI” icon on the Desktop to start the program.

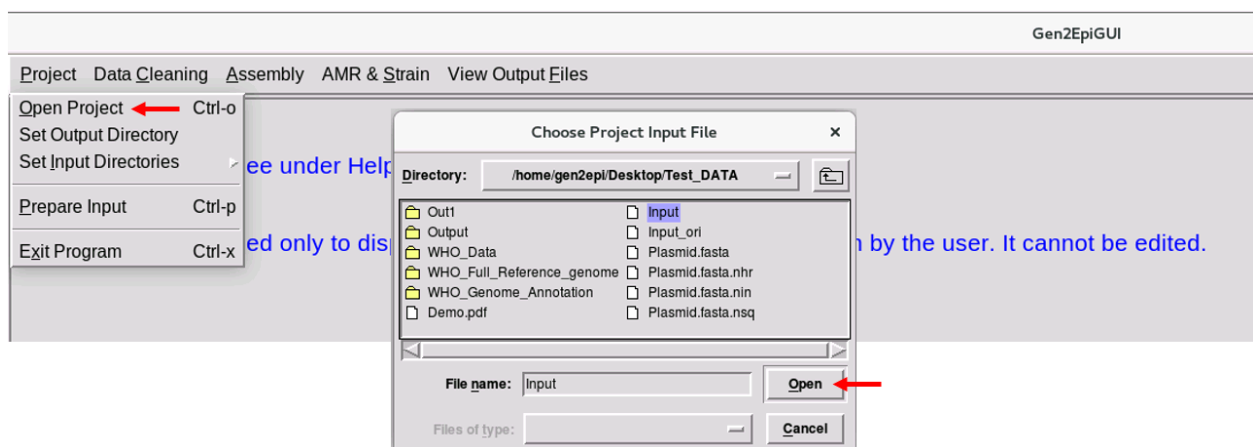


- 2) To use the program please follow the instruction as given in “Introductory Demo” under Help menu on the top right-hand side. For detailed usage please see the description below:
 - a. **Introductory demo**: A quick tutorial on “how to use Gen2Epi GUI on test dataset”.
 - b. User Manual: A step by step guide on “how can users analyze their own data using Gen2Epi GUI”.
 - c. Reference Guide: This document describes the complete command line usage of Gen2Epi v0.1.

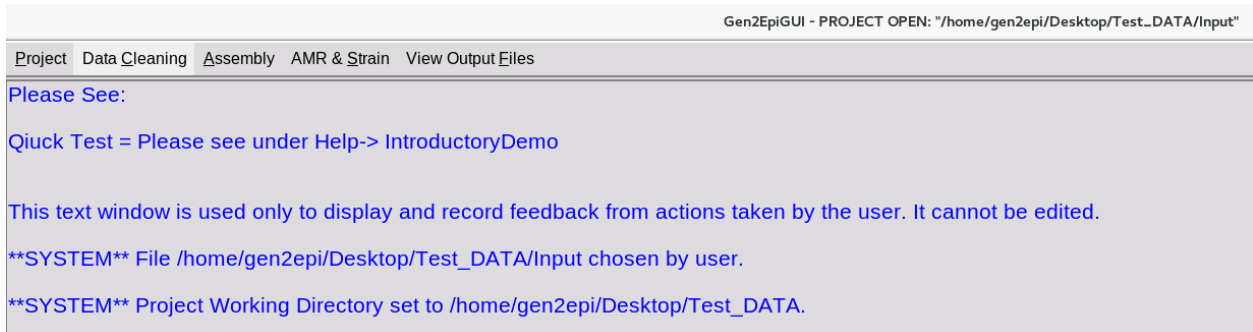
Please Note: For beginners “Introductory Demo” is highly recommended.



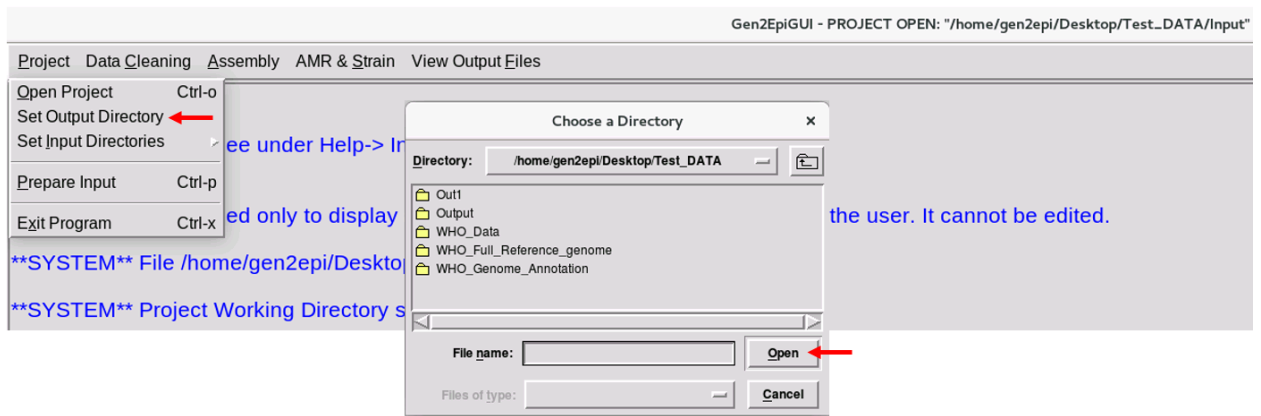
- 3) Load the input text (tab separated text file describes the name of all fastq files) file by clicking on the “Open Project” and browse (and open) the file as shown in the picture below. “Input” file is present in the Test_DATA folder at Desktop.



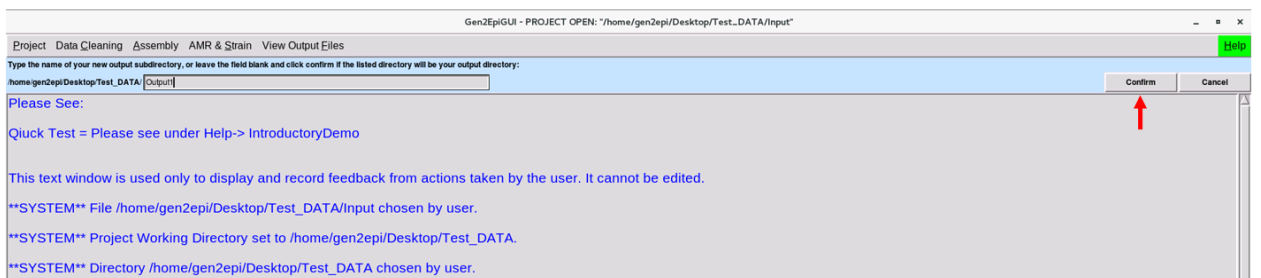
You will see following ****SYSTEM**** message in the window which shows that the “Input” file is successfully loaded.



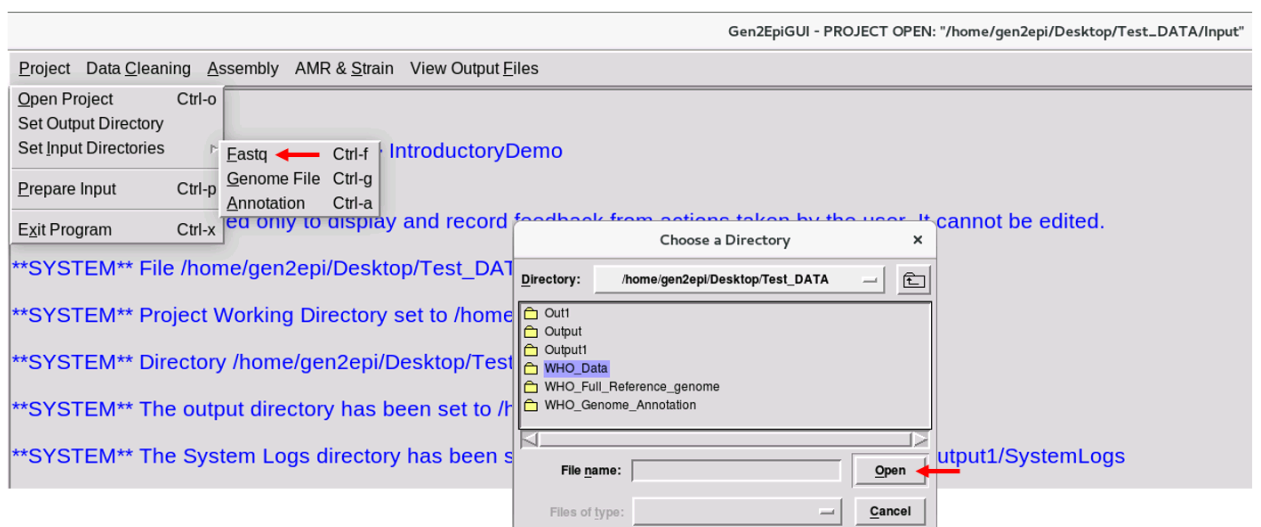
- 4) Set the output directory by clicking on “Set Output Directory” under “Project” menu. Browse the path (where you want to save the output files generated by Gen2Epi e.g. in this case path is “/home/gen2epi/Desktop/Test_DATA/”) and click open.



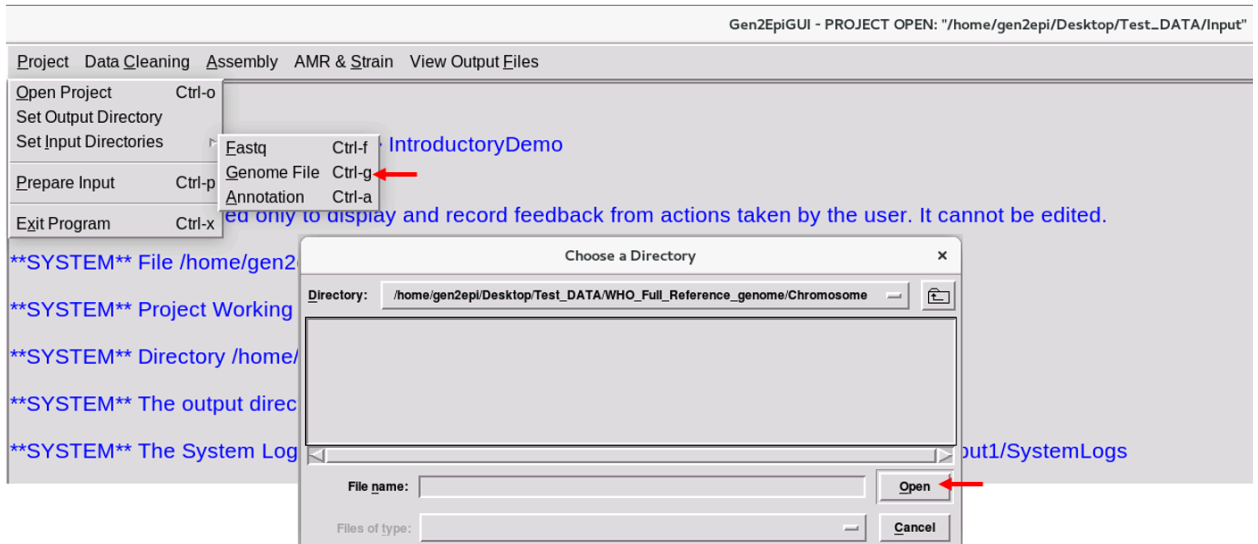
- 5) Write the name of your output directory (e.g. Output1) in the search box and click confirm.



- 6) Set Input directories for “Fastq” files by clicking on “Set Input Directories->Fastq”. Browse the folder “WHO_Data” and click “Open” as shown in the figure below:

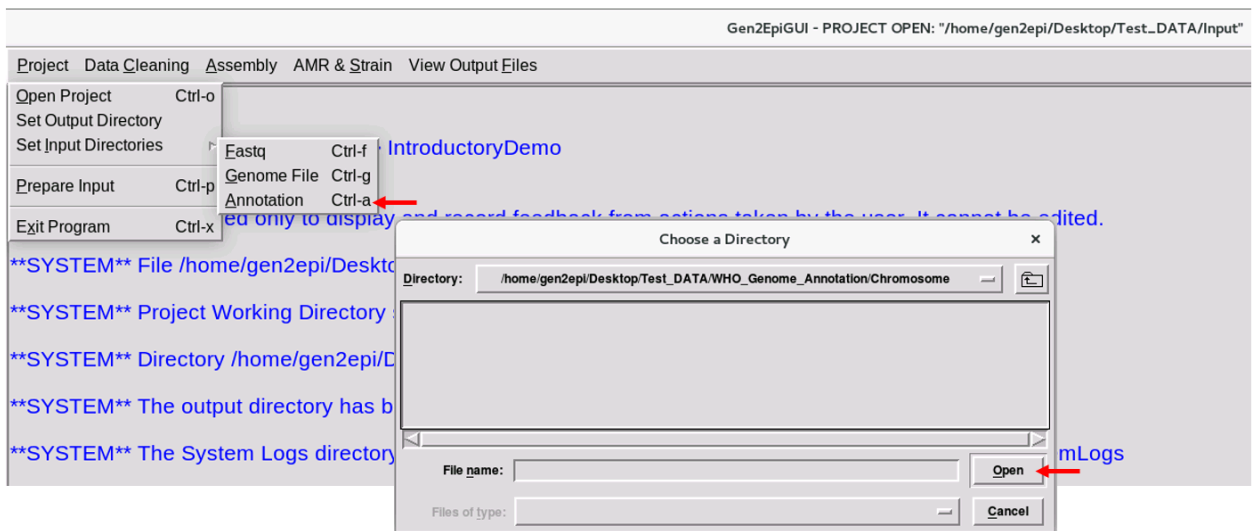


- 7) Set Input directories for “genome” files by clicking on “Set Input Directories”-> Genome File”. Select the folder “/home/gen2epi/Desktop/Test_DATA/WHO_Full_Reference_genome/Chromosome” and click “Open” as shown in the figure below:



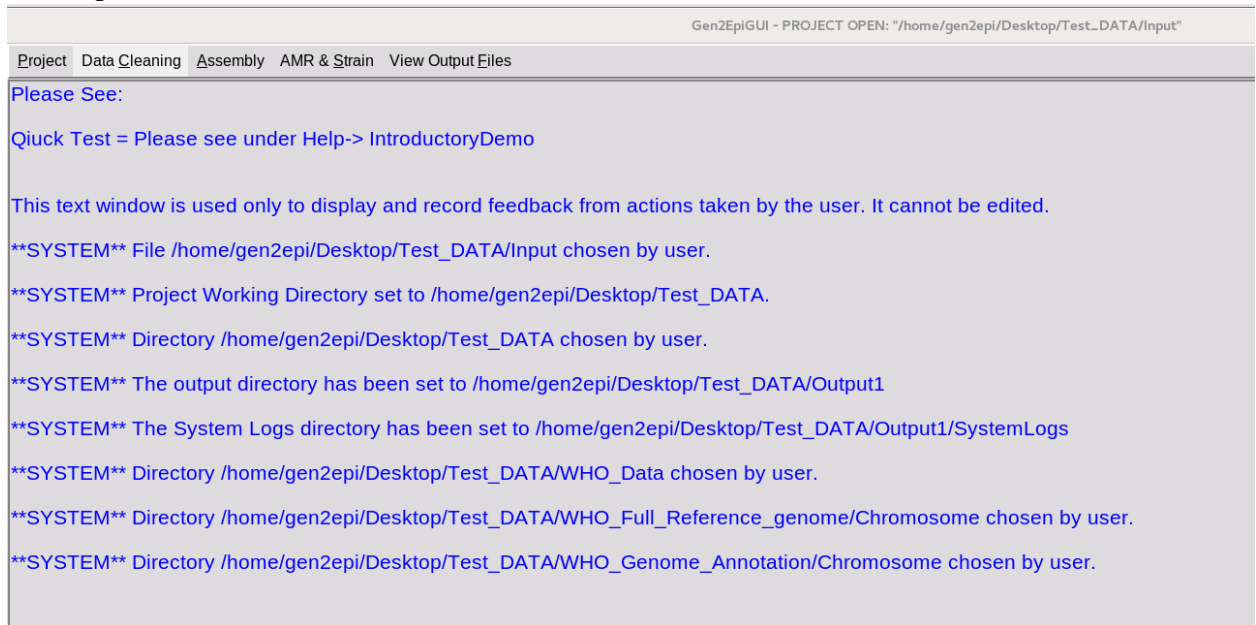
Please Note: You will not be able to see anything when you choose a directory as shown above but that’s all right.

- 8) Set Input directories for “annotation” files by clicking on “Set Input Directories ->Annotation”. Select the folder “/home/gen2epi/Desktop/Test_DATA/WHO_Genome_Annotation/Chromosome” and click on “Open” as shown in the figure below:

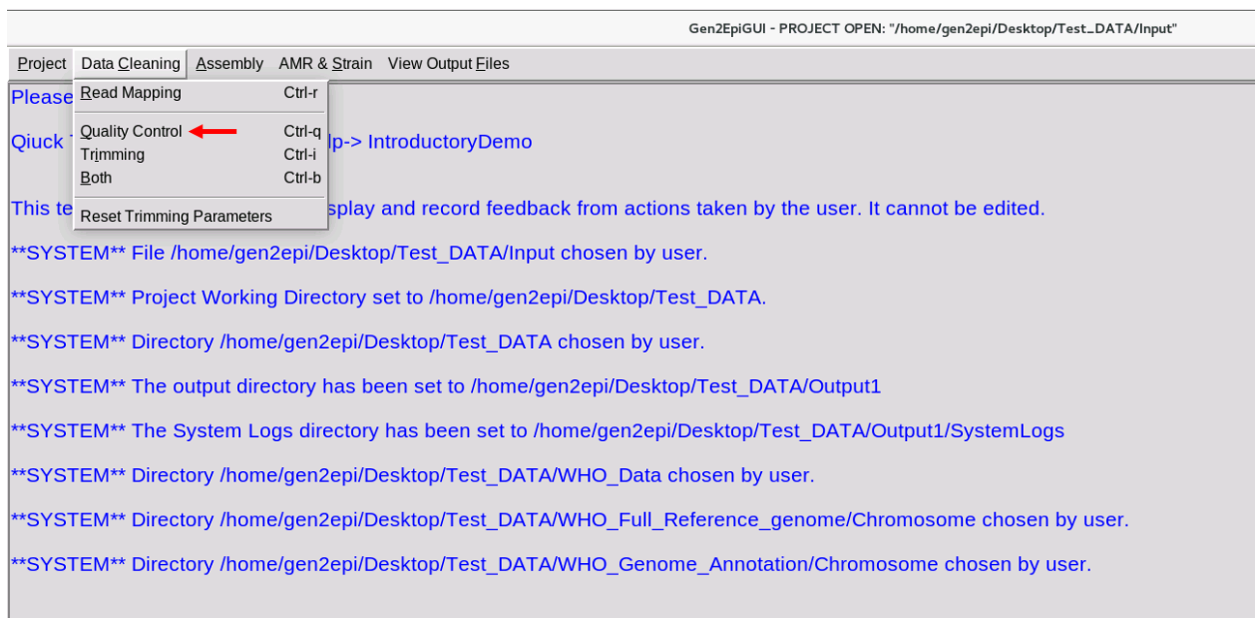


Again, you will not be able to see anything under your chosen directory as shown above but that’s all right.

- 9) Once all paths are set properly you will see following ****SYSTEM**** MESSAGES in Gen2Epi window



- 10) Now to clean your data: Under “Data Cleaning” menu – click on “Quality Control” submenu to check the quality of your raw fastq files.



Start of the analysis

```
Gen2EpiGUI - PROJECT OPEN: "/home/gen2epi/Desktop/Test_DATA/Input"

Project Data_Cleaning Assembly AMR & Strain View Output Files

**SYSTEM** Executing Quality_Control script, please wait...

perl /home/gen2epi/Desktop/Gen2Epi_Scripts/WGS_SIBP_P1.pl /home/gen2epi/Desktop/Test_DATA/Input /home/gen2epi/Desktop/Test_DATA/WHO_Data qualitycheck /home/gen2epi/Desktop/Test_DATA/Output1 2>&1

Started analysis of WHO-F_S2_L001_R1_001.fastq.gz
Approx 5% complete for WHO-F_S2_L001_R1_001.fastq.gz
Approx 10% complete for WHO-F_S2_L001_R1_001.fastq.gz
Approx 15% complete for WHO-F_S2_L001_R1_001.fastq.gz
Approx 20% complete for WHO-F_S2_L001_R1_001.fastq.gz
```

End of the analysis: On Successful completion, you will see following message

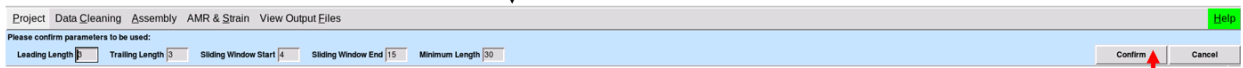
```
Analysis complete for WHO-L_S5_L001_R2_001.fastq.gz
[WARNING] multqc: MultiQC Version v1.7 now available!
[INFO] multqc: This is MultiQC v1.5
[INFO] multqc: Template : default
[INFO] multqc: Searching /home/gen2epi/Desktop/Test_DATA/Output1/QualityControl/WHO-F_S2_L001_R1_001_fastqc.zip/
[INFO] multqc: Searching /home/gen2epi/Desktop/Test_DATA/Output1/QualityControl/WHO-F_S2_L001_R2_001_fastqc.zip/
[INFO] multqc: Searching /home/gen2epi/Desktop/Test_DATA/Output1/QualityControl/WHO-G_S3_L001_R1_001_fastqc.zip/
[INFO] multqc: Searching /home/gen2epi/Desktop/Test_DATA/Output1/QualityControl/WHO-G_S3_L001_R2_001_fastqc.zip/
[INFO] multqc: Searching /home/gen2epi/Desktop/Test_DATA/Output1/QualityControl/WHO-K_S4_L001_R1_001_fastqc.zip/
[INFO] multqc: Searching /home/gen2epi/Desktop/Test_DATA/Output1/QualityControl/WHO-K_S4_L001_R2_001_fastqc.zip/
[INFO] multqc: Searching /home/gen2epi/Desktop/Test_DATA/Output1/QualityControl/WHO-L_S5_L001_R1_001_fastqc.zip/
[INFO] multqc: Searching /home/gen2epi/Desktop/Test_DATA/Output1/QualityControl/WHO-L_S5_L001_R2_001_fastqc.zip/
Searching 8 files.
[INFO] fastqc: Found 8 reports
[INFO] multqc: Compressing plot data
[INFO] multqc: Report : Desktop/Test_DATA/Output1/MultiQC-Raw/multqc_report.html
[INFO] multqc: Data : Desktop/Test_DATA/Output1/MultiQC-Raw/multqc_data
[INFO] multqc: MultiQC complete

**SYSTEM** Script completed (Quality_Control)...

**SYSTEM** Logged script output to /home/gen2epi/Desktop/Test_DATA/Output1/SystemLogs/Quality_Control.log
```

Note: Quality control results are present under “/home/gen2epi/Desktop/Test_DATA/Output1”. There will be two folders “MultiQC-Raw” and “QualityControl”. Please open the .html file under these folders to visualize the quality of each sample.

- 11) In order to trim the raw reads, users have to click on the “trimming” submenu that will further ask for the parameters confirmation. Users have the option to use either the default parameters or change them according to their requirements. After confirmation, the program will start trimming the reads.



Start of the analysis

```
Project Data_Cleaning Assembly AMR & Strain View Output Files

**SYSTEM** Executing Trimming script, please wait...

perl /home/gen2epi/Desktop/Gen2Epi_Scripts/WGS_SIBP_P1.pl /home/gen2epi/Desktop/Test_DATA/Input /home/gen2epi/Desktop/Test_DATA/WHO_Data Trimming /home/gen2epi/Desktop/Test_DATA/Output1 3 3 4 15 30 2>&1

TrimomaticPE: Started with arguments: -phred33 /home/gen2epi/Desktop/Test_DATA/WHO_Data/WHO-F_S2_L001_R1_001.fastq.gz /home/gen2epi/Desktop/Test_DATA/WHO_Data/WHO-F_S2_L001_R2_001.fastq.gz /home/gen2epi/Desktop/Test_DATA/Output1/Trimming/OutputPaired_WHO-F_S2_L001_R1_001.fastq.gz /home/gen2epi/Desktop/Test_DATA/Output1/Trimming/OutputPaired_WHO-F_S2_L001_R2_001.fastq.gz /home/gen2epi/Desktop/Test_DATA/Output1/Trimming/OutputUnpaired_WHO-F_S2_L001_R1_001.fastq.gz /home/gen2epi/Desktop/Test_DATA/Output1/Trimming/OutputUnpaired_WHO-F_S2_L001_R2_001.fastq.gz LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30
Multiple cores found. Using 4 threads
```

End of the analysis: On Successful completion, you will see following message

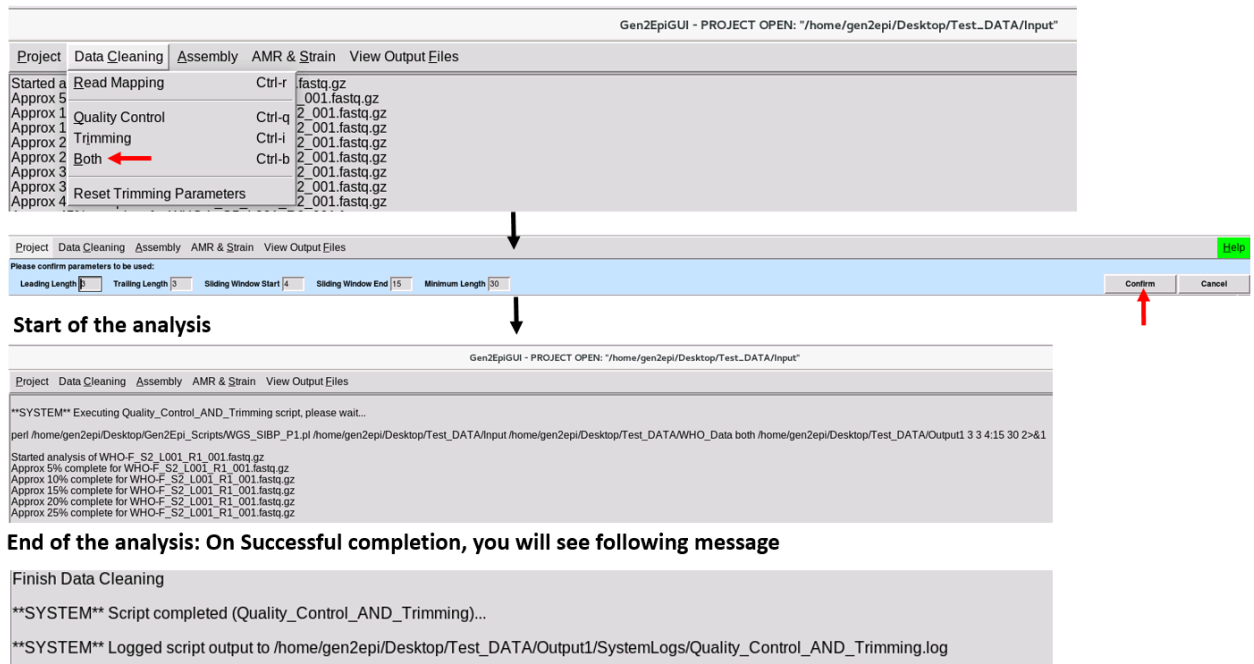
```
Searching 8 files.
[INFO] fastqc: Found 8 reports
[INFO] multqc: Compressing plot data
[INFO] multqc: Report : Desktop/Test_DATA/Output1/MultiQC-Trimmed/multqc_report.html
[INFO] multqc: Data : Desktop/Test_DATA/Output1/MultiQC-Trimmed/multqc_data
[INFO] multqc: MultiQC complete
Finish Data Cleaning

**SYSTEM** Script completed (Trimming)...

**SYSTEM** Logged script output to /home/gen2epi/Desktop/Test_DATA/Output1/SystemLogs/Trimming.log
```


Note: Trimming results are present under “/home/gen2epi/Desktop/Test_DATA/Output1”. There will be three folders “MultiQC-Trimmed”, “Trimming” and “Trimmed_QC”. Please open the .html file under these folders to visualize the quality of each sample.

- 12) It is also possible to run step 10 and 11 together as a single command. First open “/home/gen2epi/Desktop/Test_DATA/Output1” folder and delete “MultiQC-Raw”, “quality control”, “MultiQC-Trimmed”, “Trimming” and “Trimmed_QC” folders. Now go to Gen2Epi GUI and click on the “both” tab under “Data Cleaning” menu.



Note: - You will find all the results under “/home/gen2epi/Desktop/Test_DATA/Output1” in five different folders i.e. “MultiQC-Raw”, “quality control”, “MultiQC-Trimmed”, “Trimming” and “Trimmed_QC”

- 13) Now, perform the *de novo* assembly of trimmed reads by clicking on “De Novo” under Assembly tab. You will see the output as shown in the picture below.

Please Note: - This step may take a while depending on individual computer configurations. It could be possible that nothing happens when you try to press enter in the VM image, no need to worry VM will resume itself after completing the *de novo* assembly part. You can minimize the VM image window and come back later.

```

Gen2EpiGUI - PROJECT OPEN: "/home/gen2epi/Desktop/Test_DATA/Input"
Project Data Cleaning Assembly AMR & Strain View Output Files Help
right reads: [/home/gen2epi/Desktop/Test_DATA/Output1/Trimming/OutputPaired_WHO-F_S2_L001_R2_001.fastq.gz]
interlaced reads: not specified
single reads: [/home/gen2epi/Desktop/Test_DATA/Output1/Trimming/OutputUnpaired_WHO-F_S2_L001_R1_001.fastq.gz', '/home/gen2epi/Desktop/Test_DATA/Output1/Trimming/OutputUnpaired_WHO-F_S2_L001_R2_001.fastq.gz']
merged reads: not specified
Read error correction parameters:
Iterations: 1
PHRED offset will be auto-detected
Corrected reads will be compressed
Assembly parameters:
k: automatic selection based on read length
Repeat resolution is enabled
Mismatch careful mode is turned ON
MismatchCorrector will be used
Coverage cutoff is turned ON and threshold will be auto-detected
Other parameters:
Dir for temp files: /home/gen2epi/Desktop/Test_DATA/Output1/Chrom_AssemblyTrimmedReads/WHO-F/tmp
Threads: 2
Memory limit (in Gb): 2

===== SPAdes pipeline started. Log can be found here: /home/gen2epi/Desktop/Test_DATA/Output1/Chrom_AssemblyTrimmedReads/WHO-F/spades.log

===== Read error correction started.

== Running read error correction tool: /home/gen2epi/Downloads/SPAdes-3.12.0-Linux/bin/spades-hammer /home/gen2epi/Desktop/Test_DATA/Output1/Chrom_AssemblyTrimmedReads/WHO-F/corrected/configs/config.info
0:00:00.000 4M / 4M INFO General (main.cpp : 75) Starting BayesHammer, built from N/A, git revision N/A
0:00:00.000 4M / 4M INFO General (main.cpp : 76) Loading config from /home/gen2epi/Desktop/Test_DATA/Output1/Chrom_AssemblyTrimmedReads/WHO-F/corrected/configs/config.info
0:00:00.002 4M / 4M INFO General (main.cpp : 78) Maximum # of threads to use (adjusted due to OMP capabilities): 2
0:00:00.003 4M / 4M INFO General (memory_limit.cpp : 49) Memory limit set to 2 Gb
0:00:00.003 4M / 4M INFO General (main.cpp : 86) Trying to determine PHRED offset
0:00:00.016 4M / 4M INFO General (main.cpp : 92) Determined value is 33
0:00:00.016 4M / 4M INFO General (hammer_tools.cpp : 36) Hamming graph threshold tau=1, k=21, subkmer positions = [0 10]

```

14) Once the *de novo* assembly (step 13) complete. You will see the following message

```

===== Mismatch correction finished.

* Corrected reads are in /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/corrected/
* Assembled contigs are in /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/contigs.fasta
* Assembled scaffolds are in /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/scaffolds.fasta
* Assembly graph is in /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/assembly_graph.fastg
* Assembly graph in CFA format is in /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/assembly_graph_with_scaffolds.cfa
* Paths in the assembly graph corresponding to the contigs are in /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/contigs.paths
* Paths in the assembly graph corresponding to the scaffolds are in /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/scaffolds.paths

===== SPAdes pipeline finished.

SPAdes log can be found here: /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/spades.log

Thank you for using SPAdes!

**SYSTEM** Script completed (De_Novo)...

**SYSTEM** Logged script output to /home/gen2epi/Desktop/Test_DATA/Output1/SystemLogs/De_Novo.log

```

Note: Users will find the results under “/home/gen2epi/Desktop/Test_DATA/Output1” in four folder i.e. Chrom_AssemblyTrimmedReads (Assembled chromosome contigs in FASTA format generated from trimmed reads), ChromContigAssemblyTrimmedStat (Chromosome assembly statistics), Plasmid_AssemblyTrimmedReads (Assembled plasmid contigs in FASTA format generated from trimmed reads), and PlasmidContigAssemblyTrimmedStat (Plasmid assembly statistics)

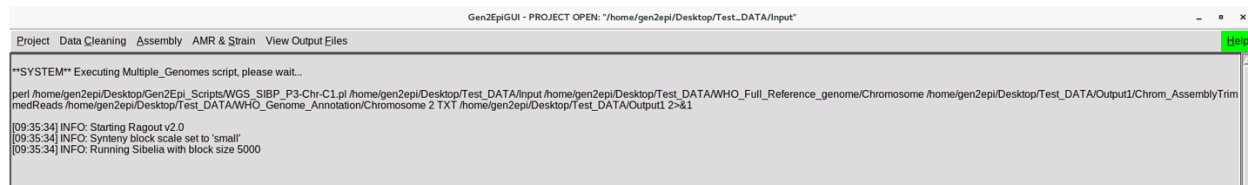
15) Perform the scaffolding of *de novo* assembly by clicking on “Assembly->Scaffolding->Chromosome->Multiple Genomes” as shown below:

```

Gen2EpiGUI - PROJECT OPEN: "/home/gen2epi/Desktop/Test_DATA/Input"
Project Data Cleaning Assembly AMR & Strain View Output Files
[bwa_index] Construct BWT De Novo Ctrl-d
[bwa_index] 0.01 seconds Scaffolding
[bwa_index] Update BWT Scaffolding Chromosome Multiple Genomes
[bwa_index] Pack forward Preassembled Genome Plasmid Ctrl-l Single Genomes
[bwa_index] Construct SA from BWT and Occ... 0.01 sec
[main] Version: 0.7.12-r1039
[main] CMD: /home/gen2epi/Downloads/SPAdes-3.12.0-Linux/bin/spades-bwa index -a is /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/misc/
[main] Real time: 0.054 sec; CPU: 0.029 sec
0:01:07.786 4M / 80M INFO DatasetProcessor (dataset_processor.cpp : 173) Running bwa mem .../home/gen2epi/Downloads/SPAdes-3.12.0-Linux/bin/spades-bwa mem -
Reads/WHO-L/misc/assembled_scaffolds.fasta /home/gen2epi/Desktop/Test_DATA/Output1/Trimming/OutputUnpaired_WHO-L_S5_L001_R2_001.fastq.gz > /home/gen2epi/Desktop/T
st6/lib2_VXtd9s/tmp.sam
[main] Version: 0.7.12-r1039
[main] CMD: /home/gen2epi/Downloads/SPAdes-3.12.0-Linux/bin/spades-bwa mem -v 1 -t 2 /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/mis
ng/OutputUnpaired_WHO-L_S5_L001_R2_001.fastq.gz
[main] Real time: 0.025 sec; CPU: 0.034 sec
0:01:07.812 4M / 80M INFO DatasetProcessor (dataset_processor.cpp : 224) Adding scaffolds from /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedRea

```


Start of the Scaffolding process



```
Gen2EpiGUI - PROJECT OPEN: "/home/gen2epi/Desktop/Test_DATA/Input"

Project Data Cleaning Assembly AMR & Strain View Output Files Help

**SYSTEM** Executing Multiple_Genomes script, please wait...

perl /home/gen2epi/Desktop/Gen2Epi_Scripts/WGS_SIBP_P3-Chr-C1.pl /home/gen2epi/Desktop/Test_DATA/Input/home/gen2epi/Desktop/Test_DATA/WHO_Full_Reference_genome/Chromosome /home/gen2epi/Desktop/Test_DATA/Output1/Chrom_AssemblyTrimmedReads /home/gen2epi/Desktop/Test_DATA/WHO_Genome_Annotation/Chromosome 2 TXT /home/gen2epi/Desktop/Test_DATA/Output1 2-&1

[09:35:34] INFO: Starting Ragout v2.0
[09:35:34] INFO: Synteny block scale set to 'small'
[09:35:34] INFO: Running Sibelia with block size 5000
```

Scaffolding process completion



```
Finished: 2019-06-03 09:41:13
Elapsed time: 0:00:07.934315
Total NOTICES: 5; WARNINGS: 5; non-fatal ERRORS: 0

Thank you for using QUAST!

Attaching package: &#x26;#x26;dplyr&#x26;#x26;

The following objects are masked from &#x26;#x26;package:stats&#x26;#x26;:
  filter, lag

The following objects are masked from &#x26;#x26;package:base&#x26;#x26;:
  intersect, setdiff, setequal, union

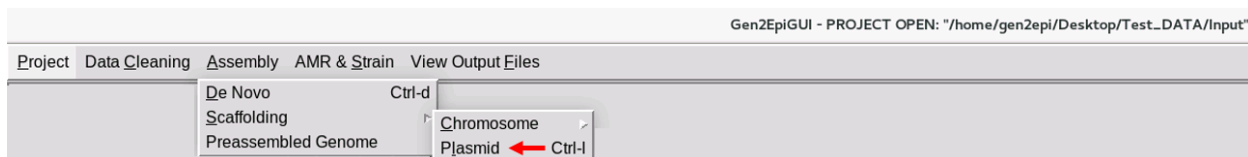
Joining, by = "samples"

**SYSTEM** Script completed (Multiple_Genomes)...

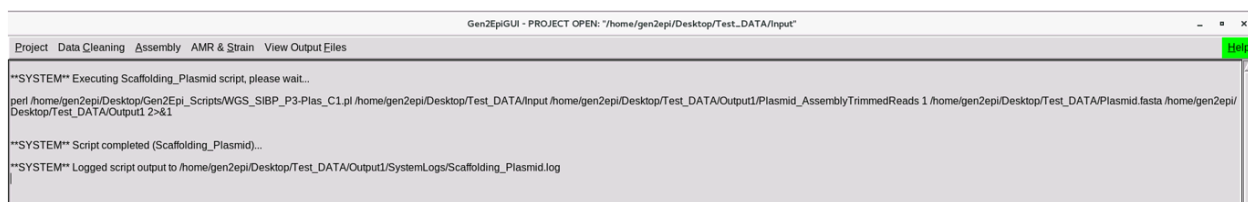
**SYSTEM** Logged script output to /home/gen2epi/Desktop/Test_DATA/Output1/SystemLogs/Multiple_Genomes.log
```

Note: Output files generated from this step will be stored in folder “Chr_Scaffolds” under “/home/gen2epi/Desktop/Test_DATA/Output1”.

- 16) To identify the different types of *Neisseria gonorrhoeae* plasmids using assembled plasmid contigs (in step 13), select “Assembly->Scaffolding->Plasmid”.



Once you click on “Plasmid”. You will see the following output



```
Gen2EpiGUI - PROJECT OPEN: "/home/gen2epi/Desktop/Test_DATA/Input"

Project Data Cleaning Assembly AMR & Strain View Output Files Help

**SYSTEM** Executing Scaffolding_Plasmid script, please wait...

perl /home/gen2epi/Desktop/Gen2Epi_Scripts/WGS_SIBP_P3-Plas_C1.pl /home/gen2epi/Desktop/Test_DATA/Input/home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads 1 /home/gen2epi/Desktop/Test_DATA/Plasmid.fasta /home/gen2epi/Desktop/Test_DATA/Output1 2-&1

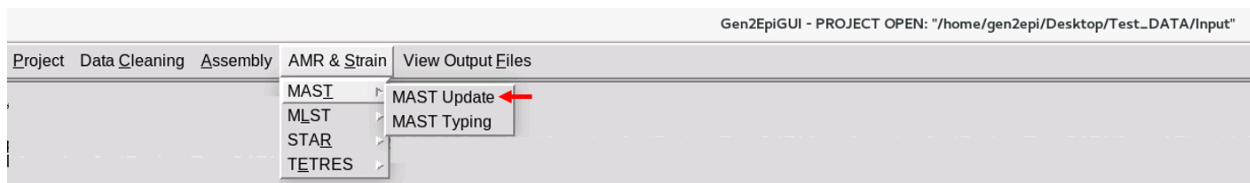
**SYSTEM** Script completed (Scaffolding_Plasmid)...

**SYSTEM** Logged script output to /home/gen2epi/Desktop/Test_DATA/Output1/SystemLogs/Scaffolding_Plasmid.log
```

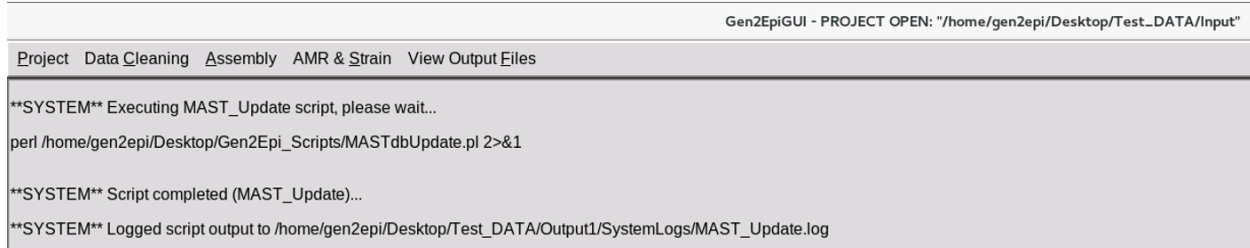
Note: Output generated from this step will be stored in folder “Plasmid_Identification” under “/home/gen2epi/Desktop/Test_DATA/Output1”.

17) NgMAST Typing

- a. Update the underlying MAST database by clicking MAST->MAST Update tab



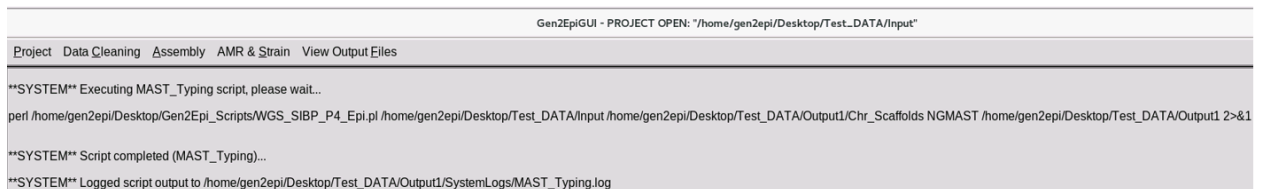
MAST database update



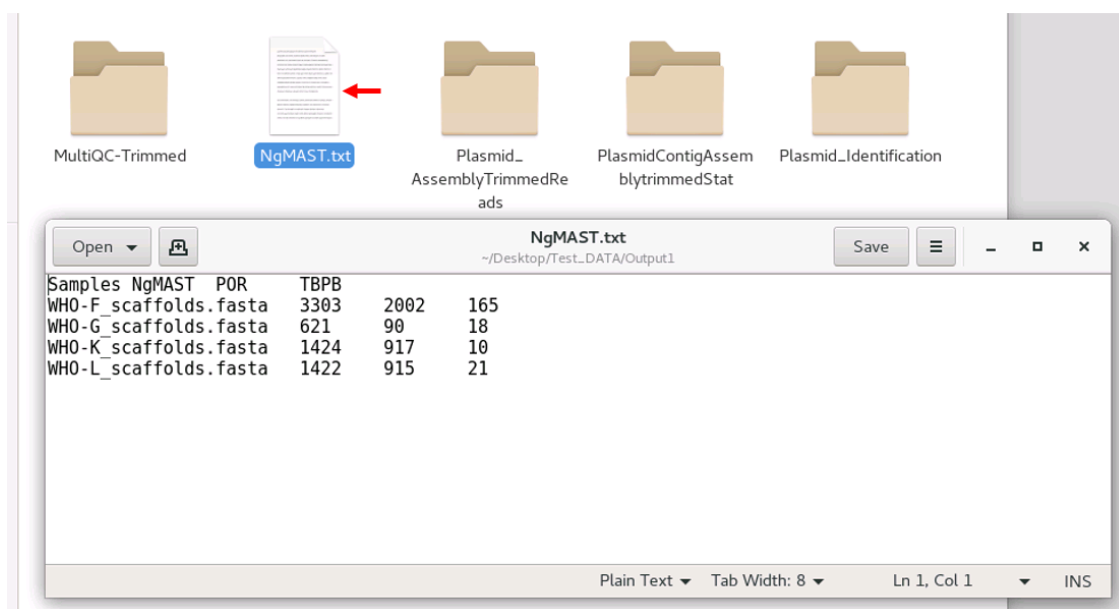
b. Perform the MAST typing by selecting MAST->MAST Typing



Strain typing

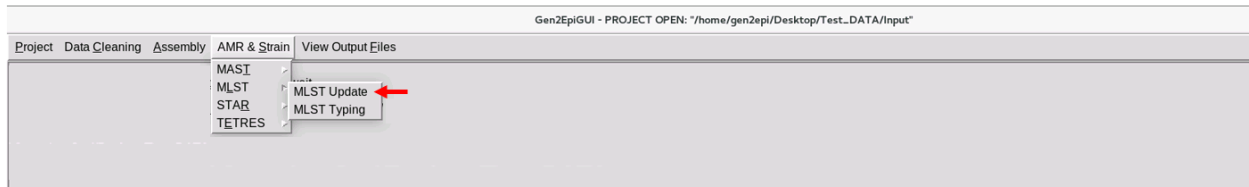


Note: MAST strain typing for each sample is present in file “NgMAST.txt” under “/home/gen2epi/Desktop/Test_DATA/Output1” as shown below.

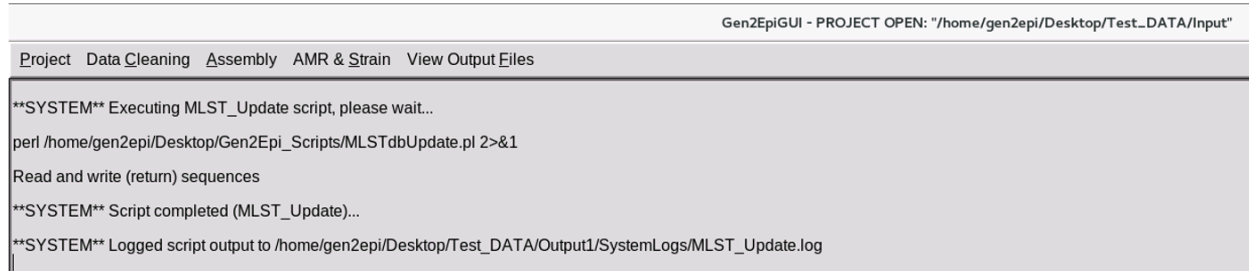


18) NgMLST Typing

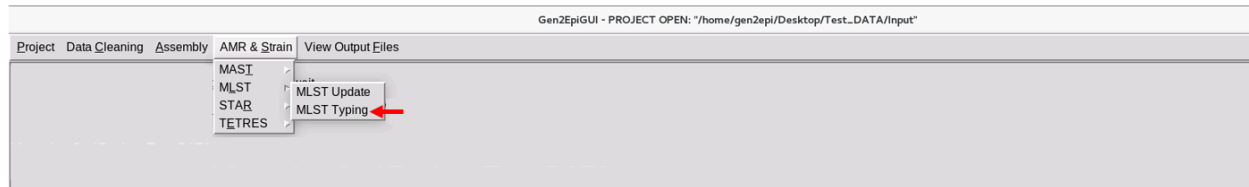
- a. Update the underlying MLST database by clicking MLST->MLST Update tab



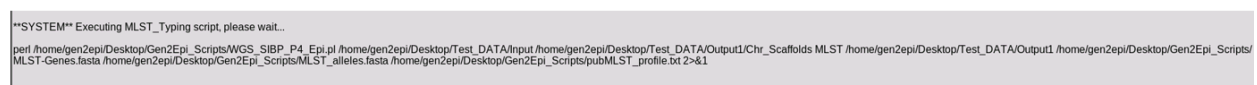
MLST database update



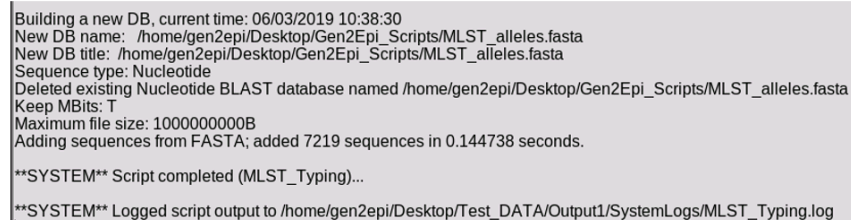
- b. Perform the MLST typing by selecting MLST->MLST Typing



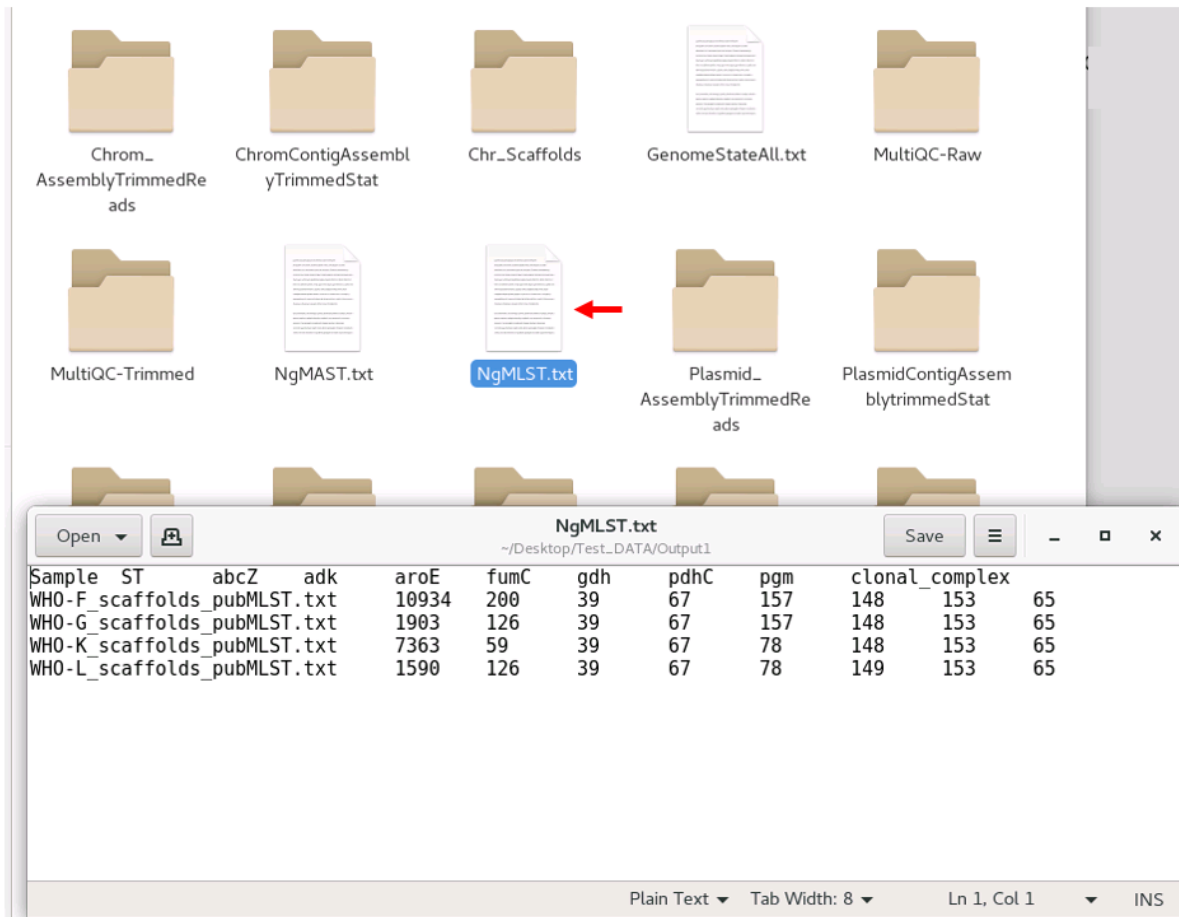
Start of the strain typing analysis



Completion of the strain typing analysis

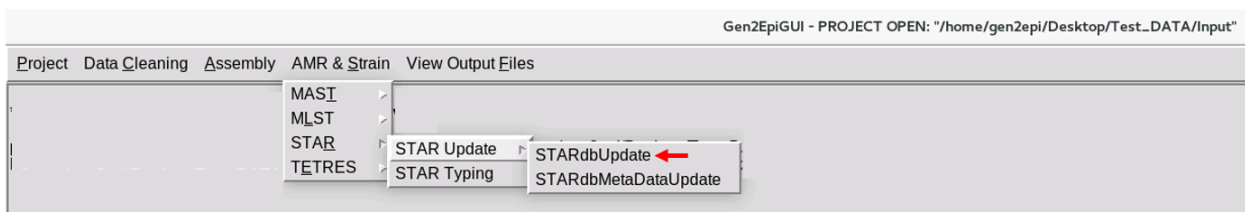


Note: MLST strain typing for each sample is present in file “NgMLST.txt” under “/home/gen2epi/Desktop/Test_DATA/Output1” as shown below.

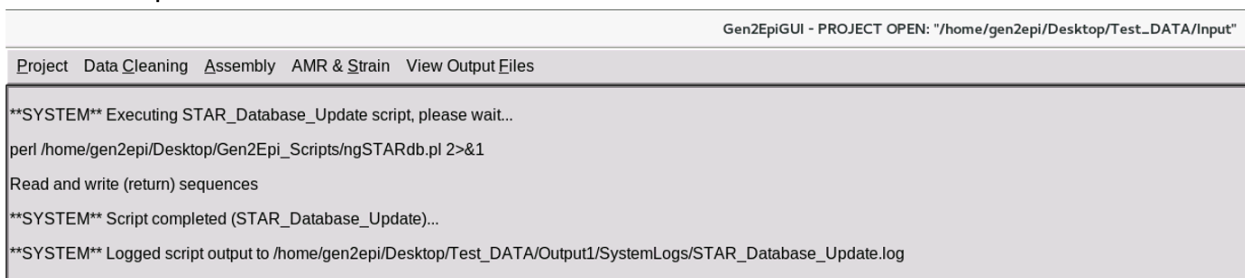


19) NgSTAR

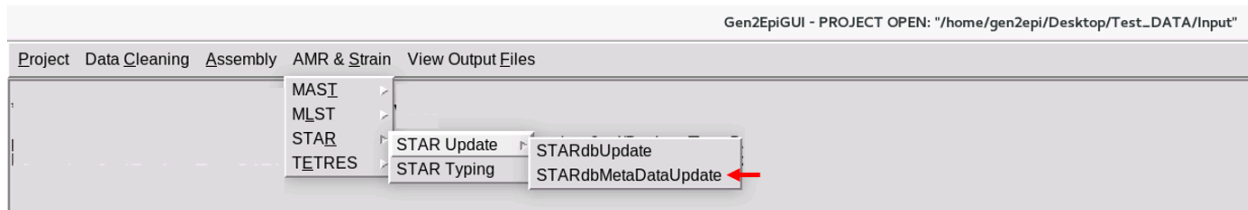
- Update the underlying STAR database by clicking STAR->STAR Update->STARdbUpdate tab



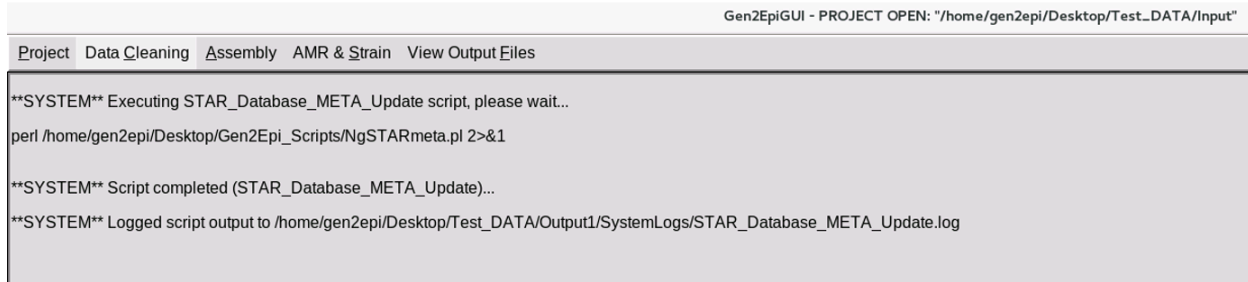
STAR database update



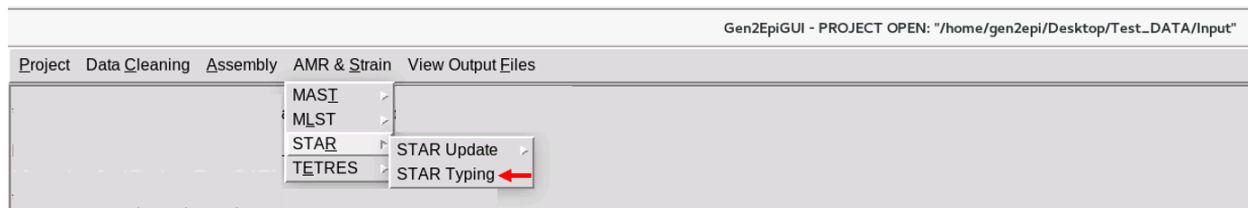
- b. Update the underlying STAR database metadata by clicking STAR->STAR Update->STARdbMetaDataUpdate tab



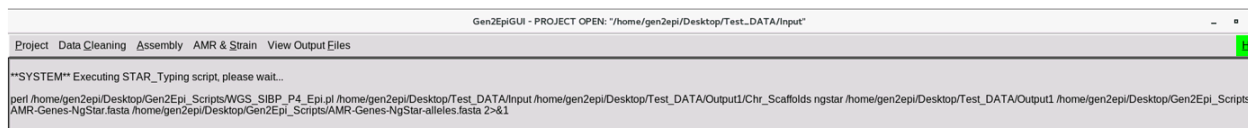
STAR database metadata update



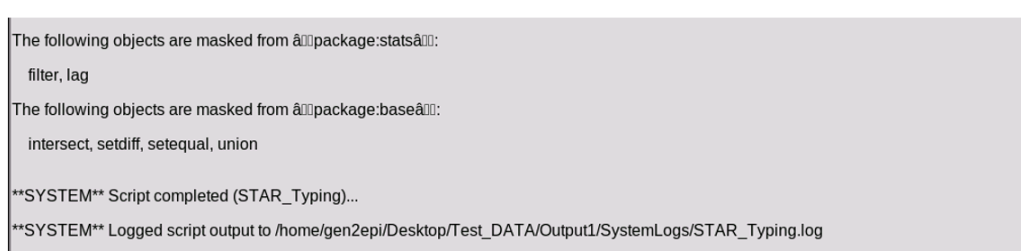
- c. Perform the STAR typing by selecting STAR->STAR Typing



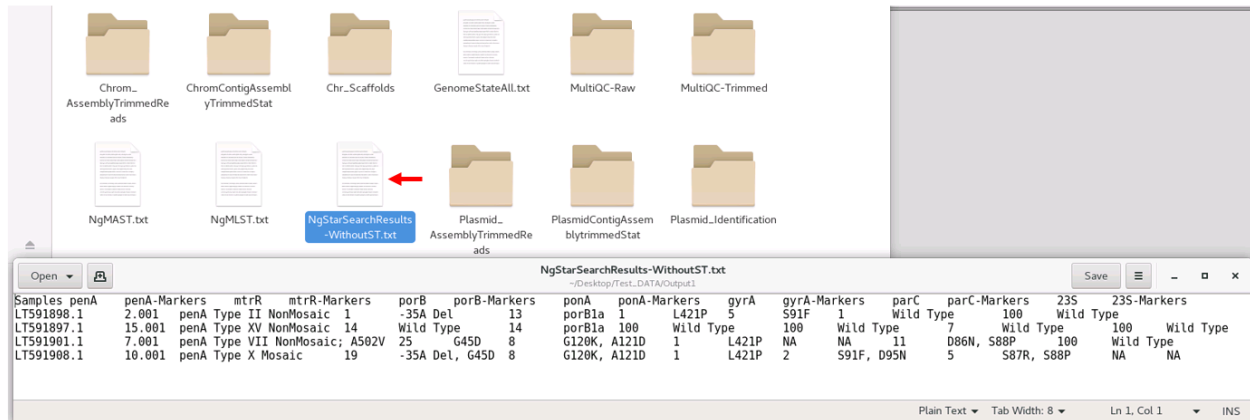
Start of the STAR typing process



Completion of the STAR typing process

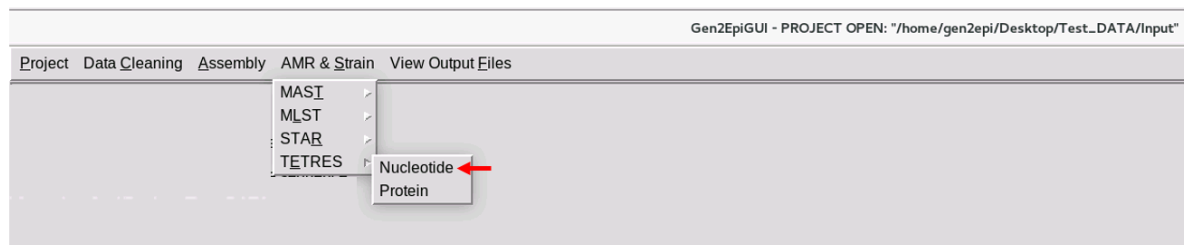


Note: NgSTAR output for each sample is present in file “NgStarSearchResults-WithoutST.txt” under “/home/gen2epi/Desktop/Test_DATA/Output1” as shown below.



20) Tetracycline Resistance

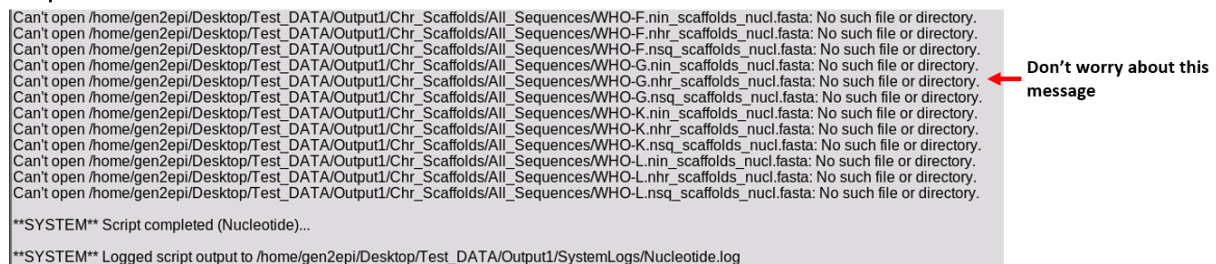
- 1) Click on “Nucleotide” under TETRES Tab to extract the rpsJ nucleotide sequences. These fasta sequences can be visualized under any multiple sequence alignment program.



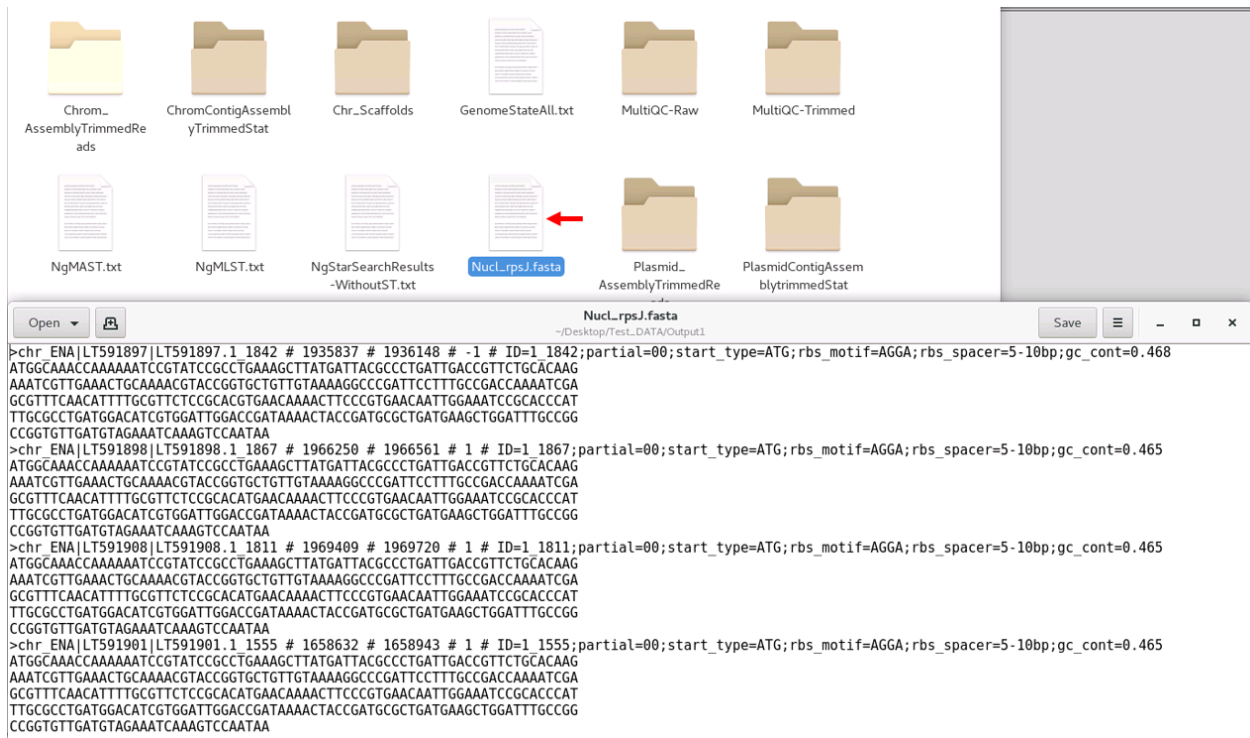
Starting



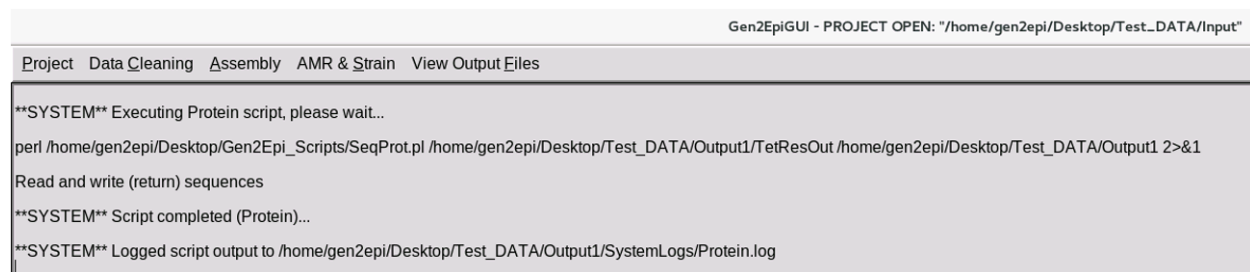
Completion



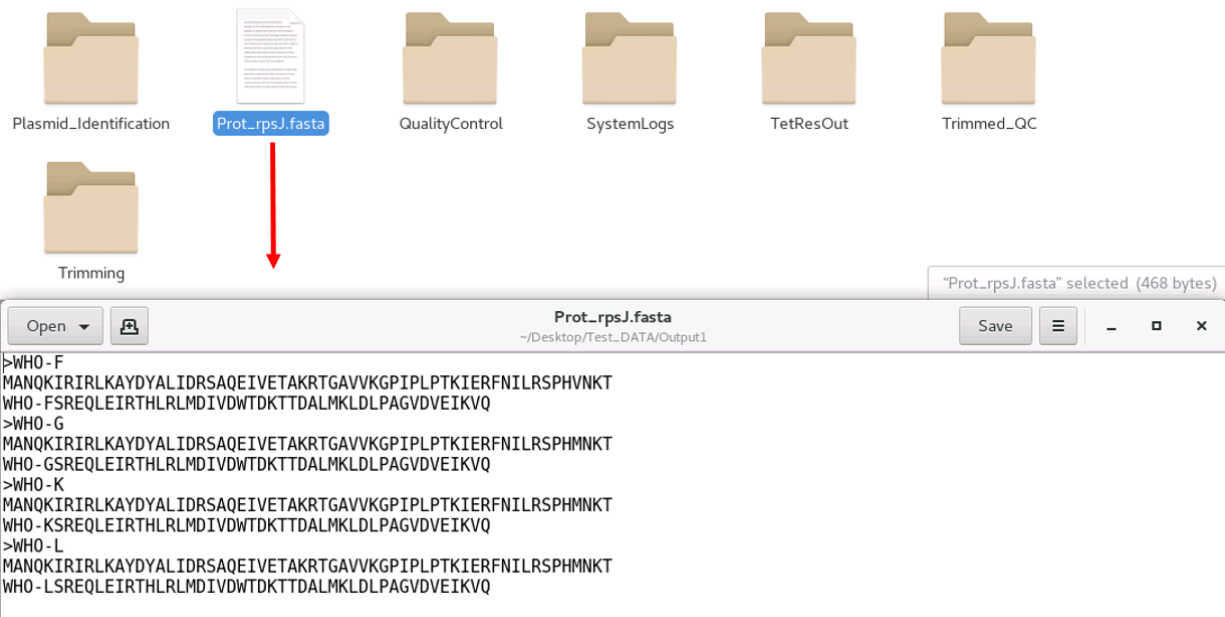
Note: Users can see the rpsJ nucleotide sequence for each sample in file “Nucl_rpsJ.fasta” under “/home/gen2epi/Desktop/Test_DATA/Output1” as shown below.



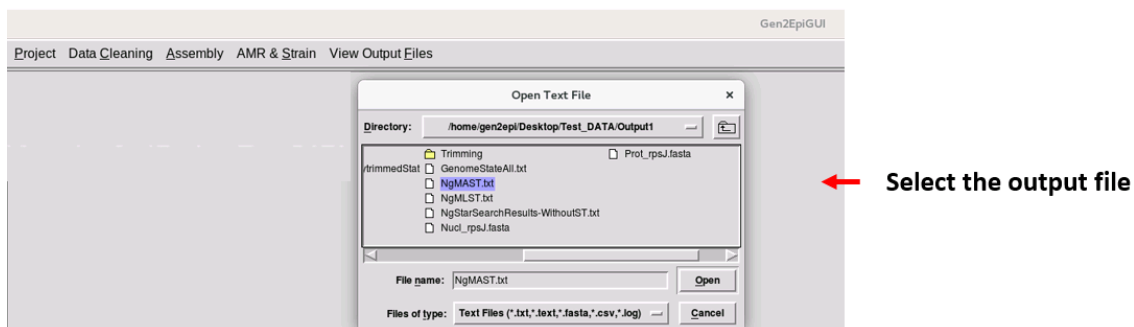
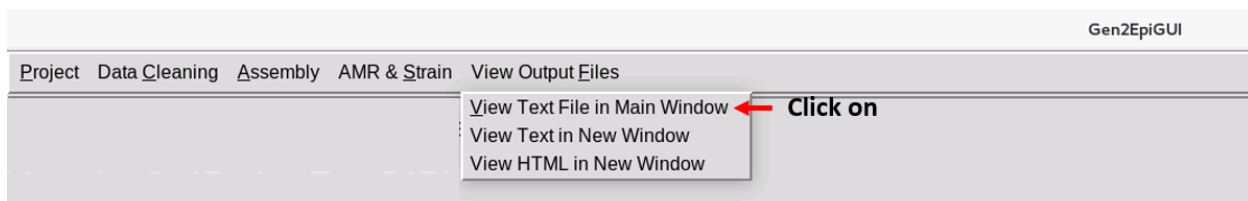
- Click on “Protein” under TETRES Tab to extract the rpsJ protein alignment. These fasta sequences can be visualized under any multiple sequence alignment program.



Note: Users can see the rpsJ nucleotide sequence for each sample in file “Prot_rpsJ.fasta” under “/home/gen2epi/Desktop/Test_DATA/Output1” as shown below.



21) View Output Files: - Finally, the text and HTML outputs generated from the above steps are accessible under “View Output File” tab.



Gen2EpiGUI

Project	Data Cleaning	Assembly	AMR & Strain	View Output Files
Samples	NgMAST POR	TBPB		
WHO-F_scaffolds.fasta	3303	2002	165	
WHO-G_scaffolds.fasta	621	90	18	
WHO-K_scaffolds.fasta	1424	917	10	
WHO-L_scaffolds.fasta	1422	915	21	

Visualize the results