Gen2Epi GUI Quick Tutorial

"Filename", 'foldername', menu and submeu items, **Open files** and **confirm**

Locations and name of the test data used in this tutorial:

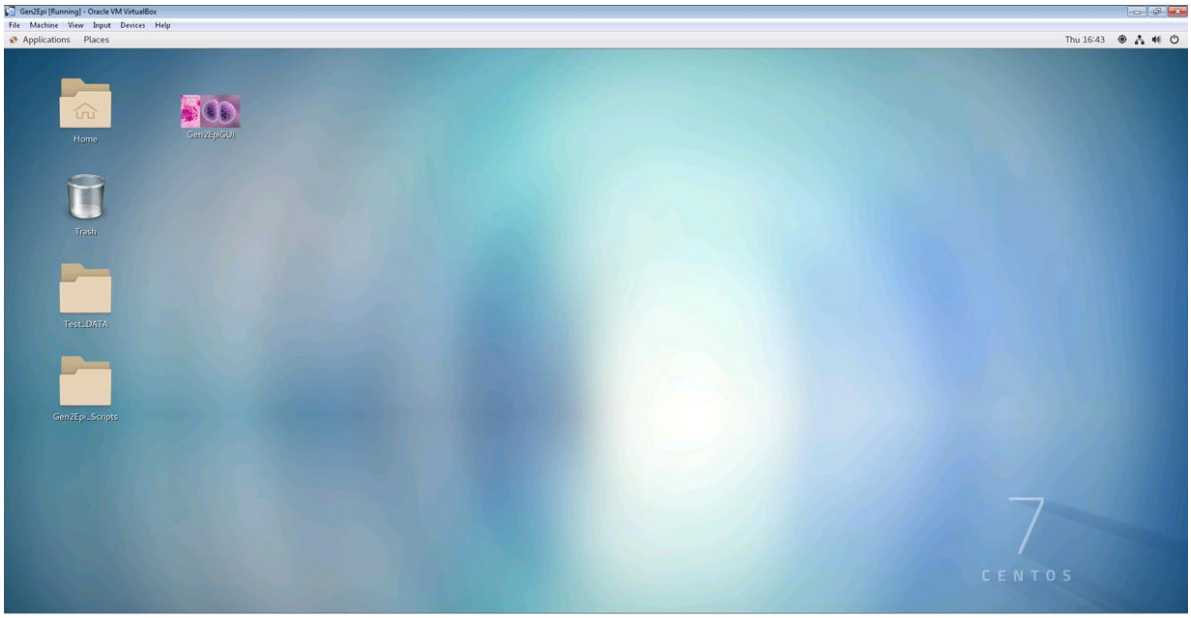
Test data directory: '*/home/gen2epi/Desktop/Test_DATA*'

FASTQ files: '*/home/gen2epi/Desktop/Test_DATA/WHO_Data*'

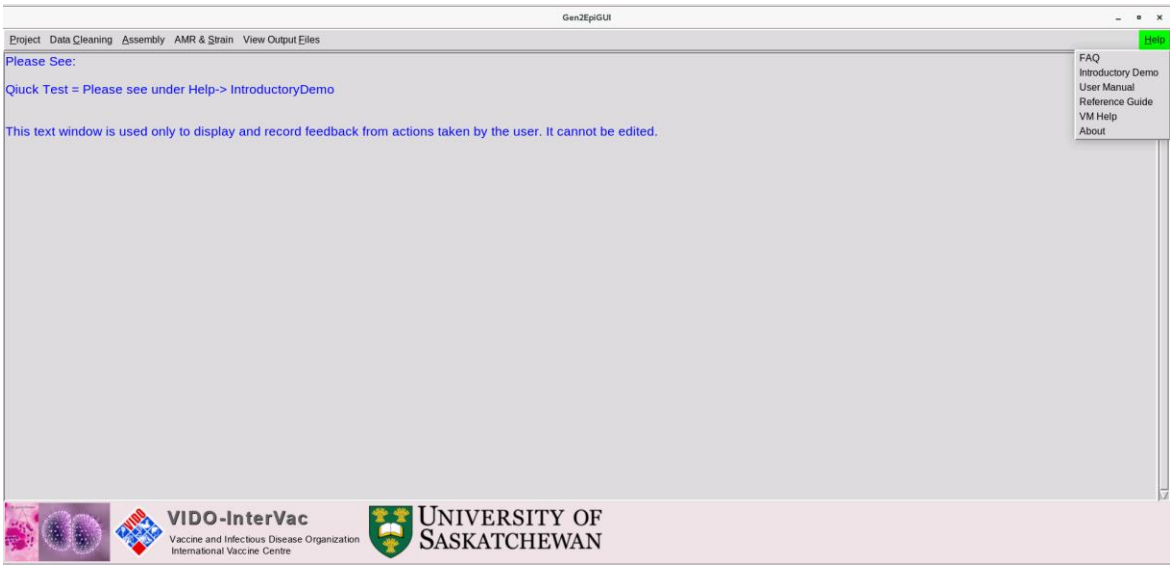
Genome files: '*/home/gen2epi/Desktop/Test_DATA/WHO_Full_Reference_genome/Chromosome*'

Annotation files: '*/home/gen2epi/Desktop/Test_DATA/WHO_Genome_Annotation/Chromosome*'

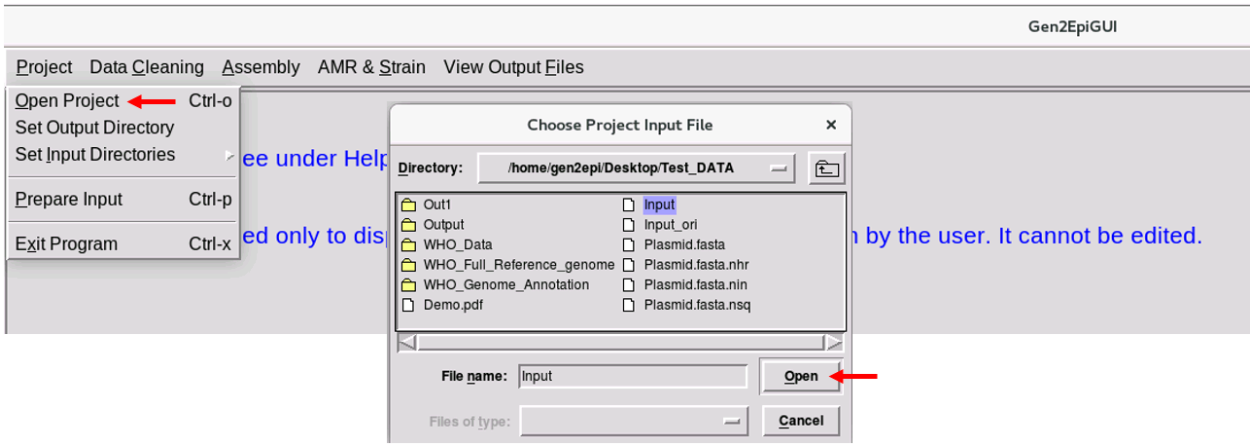
1) Click on the `Gen2EpiGUI` icon on the Desktop to start the program.



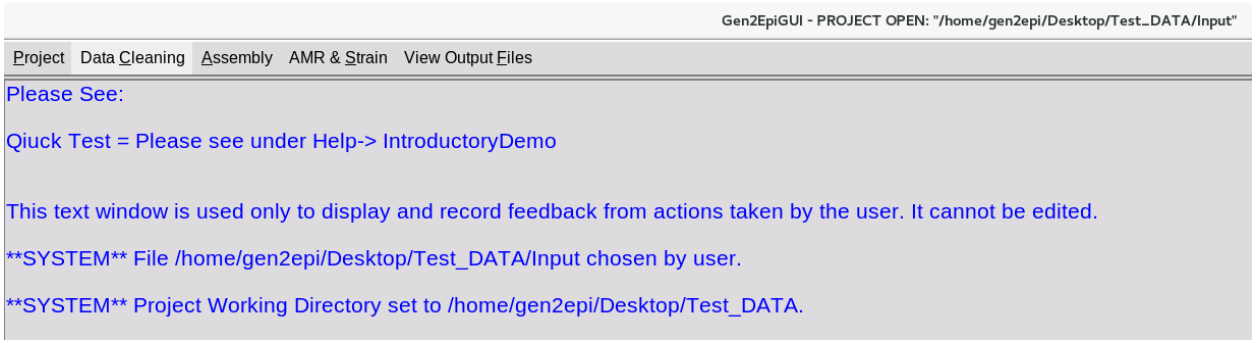
2) To use the program please follow the instruction as given in Introductory Demo under Help menu on the top right-hand side.



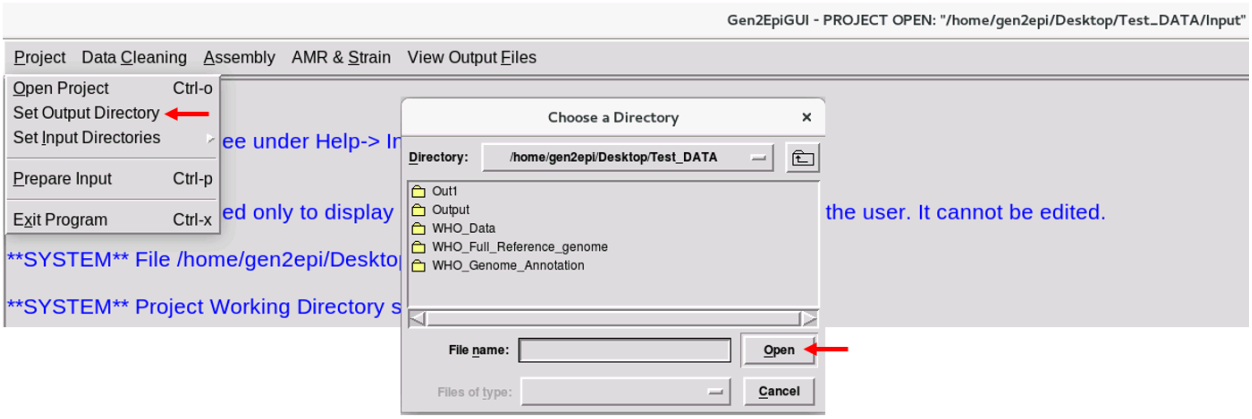
3) Load the text file (tab-separated text file listing the names of all FASTQ files), named "*Input*", by clicking on the Open Project under Project menu and browse (and **open**) the file as shown in the picture below. File "*Input*" is present in the "*Test_DATA"* folder at the Desktop folder.



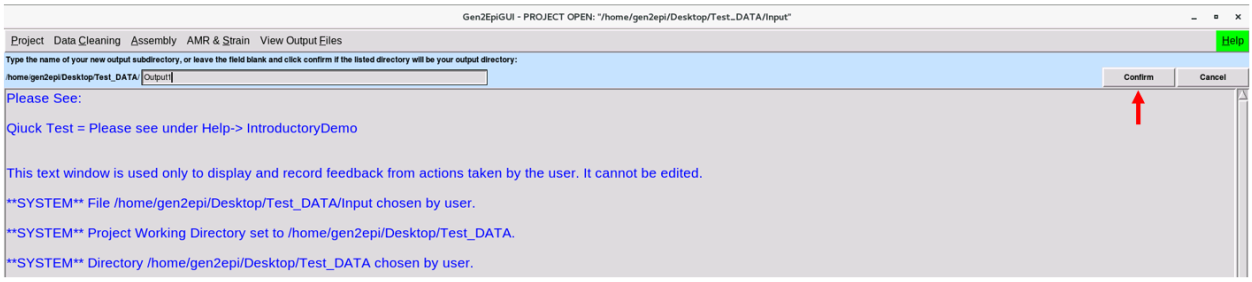
You will see the following **SYSTEM** message in the text window when you click on **open** which shows that the "*Input*" file is successfully loaded.



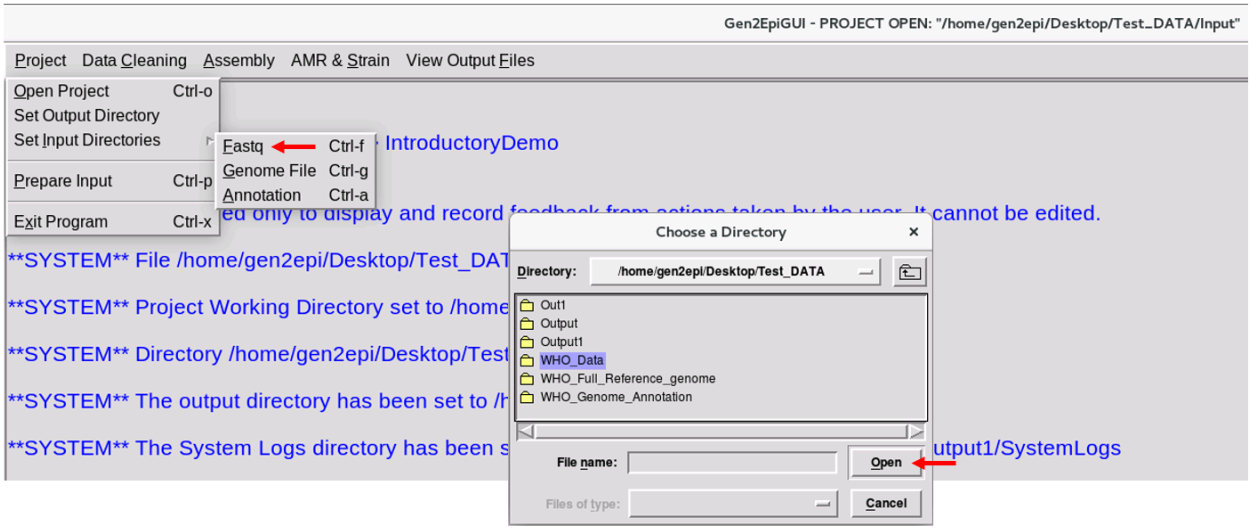
4) Set the output directory by clicking on **Set Output Directory** under the **Project** menu. Browse the path (navigate to where you want to save the output files generated by "Gen2EpiGUI" e.g. in this case the path is '*/home/gen2epi/Desktop/Test_DATA/*') and click **open**.



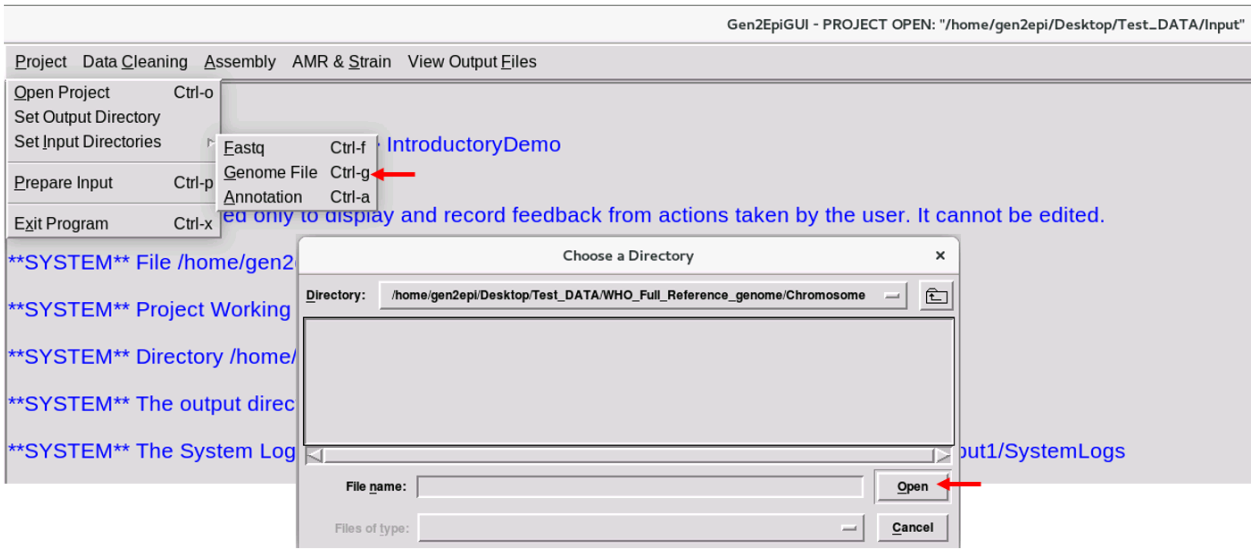
5) Write the name of your output directory (e.g. "*Output1"*) in the search box and click **confirm**.



6) Set input directories for FASTQ files by clicking on **Set Input Directories -> Fastq** under the **Project** menu. Browse the folder "*WHO_Data*" and click **open** as shown in the figure below:

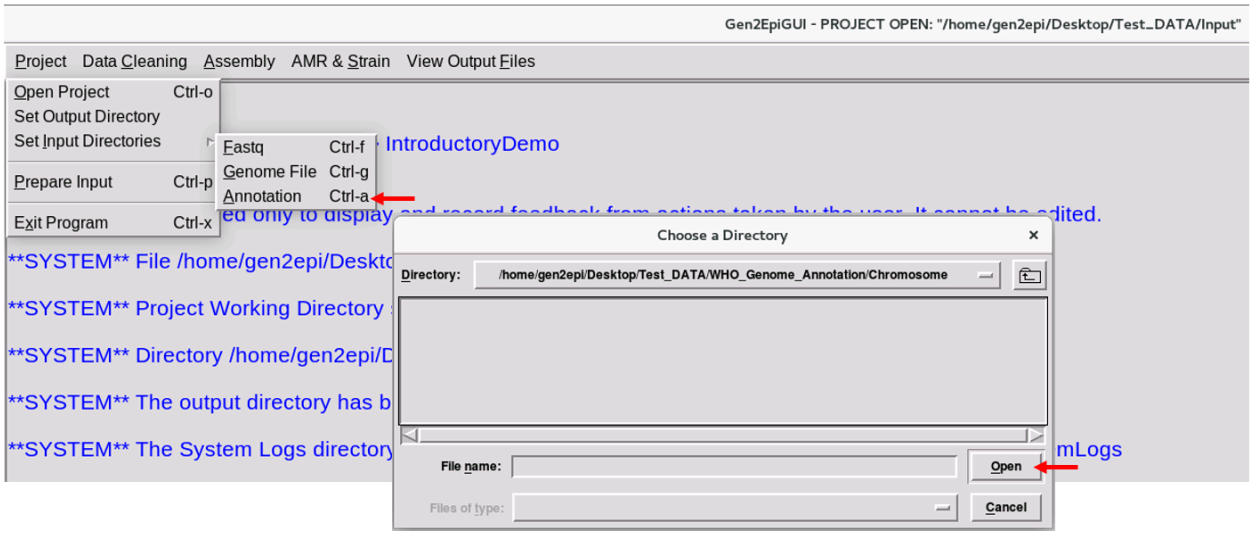


7) Set input directories for genome files by clicking on Set Input Directories -> Genome File. Select the folder (with genome files in it) by browsing the path '*/home/gen2epi/Desktop/Test_DATA/WHO_Full_Reference_genome/Chromosome*' and click **open** as shown in the figure below:



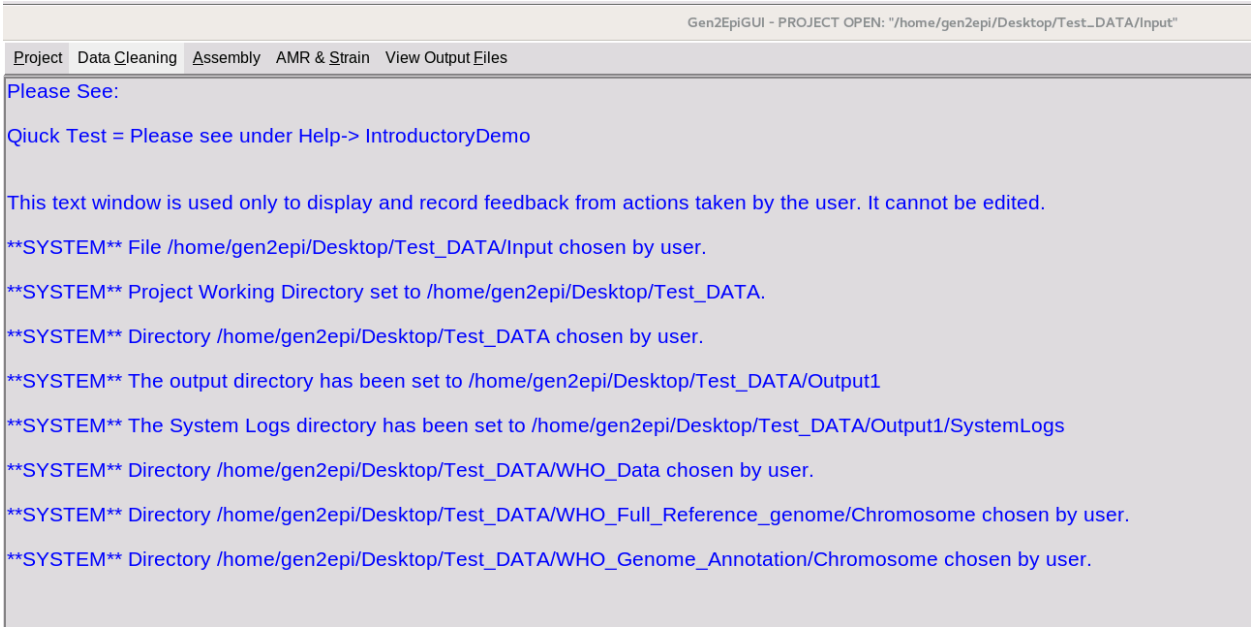
Please Note: You will not be able to see anything when you choose a directory as shown above but that's all right.

8) Set input directories for annotation files by clicking on Set Input Directories -> Annotation menu. Select the folder (with annotation files in it) by browsing the path '*/home/gen2epi/Desktop/Test_DATA/WHO_Genome_Annotation/Chromosome*' and click on **open** as shown in the figure below:

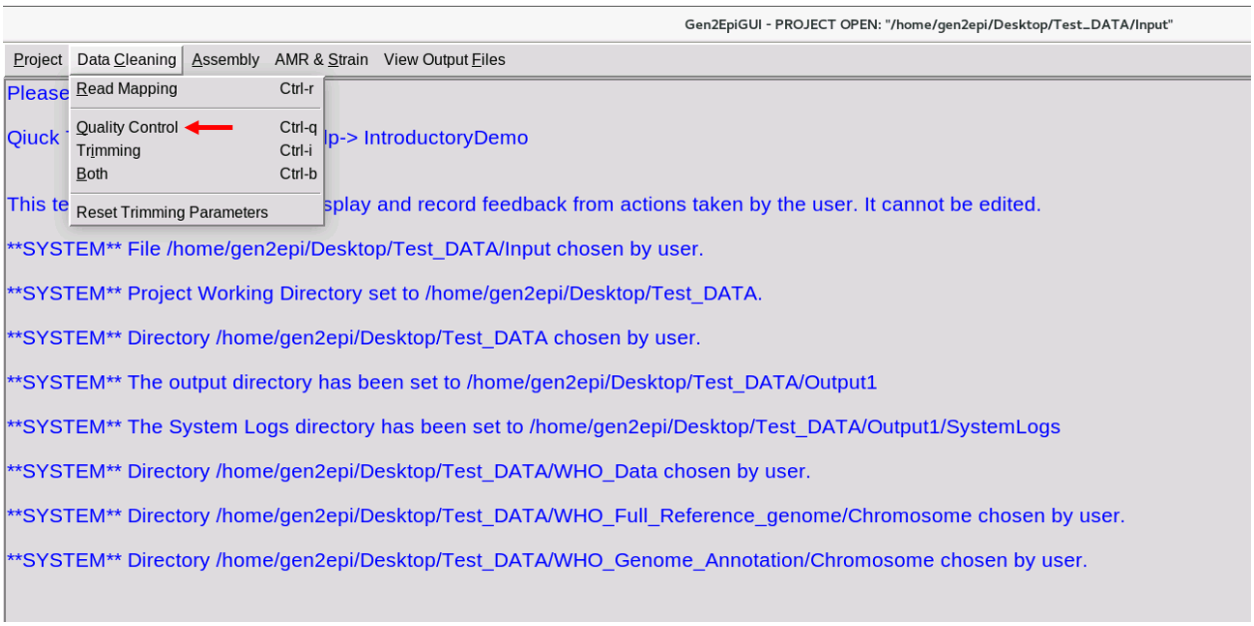


Again, you will not be able to see anything under your chosen directory as shown above but that's all right.

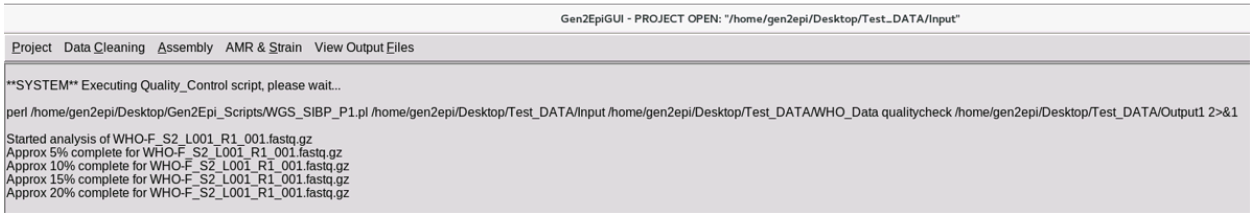
9) Once all paths are set properly you will see following **SYSTEM** MESSAGES in "Gen2EpiGUI" main text window.



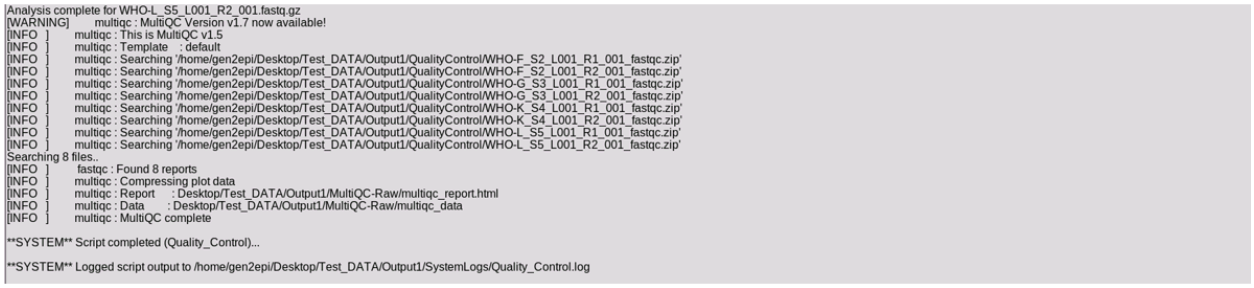
10) Now to clean your data: Under Data Cleaning menu – click on Quality Control submenu to check the quality of your raw fastq files.



Start of the analysis

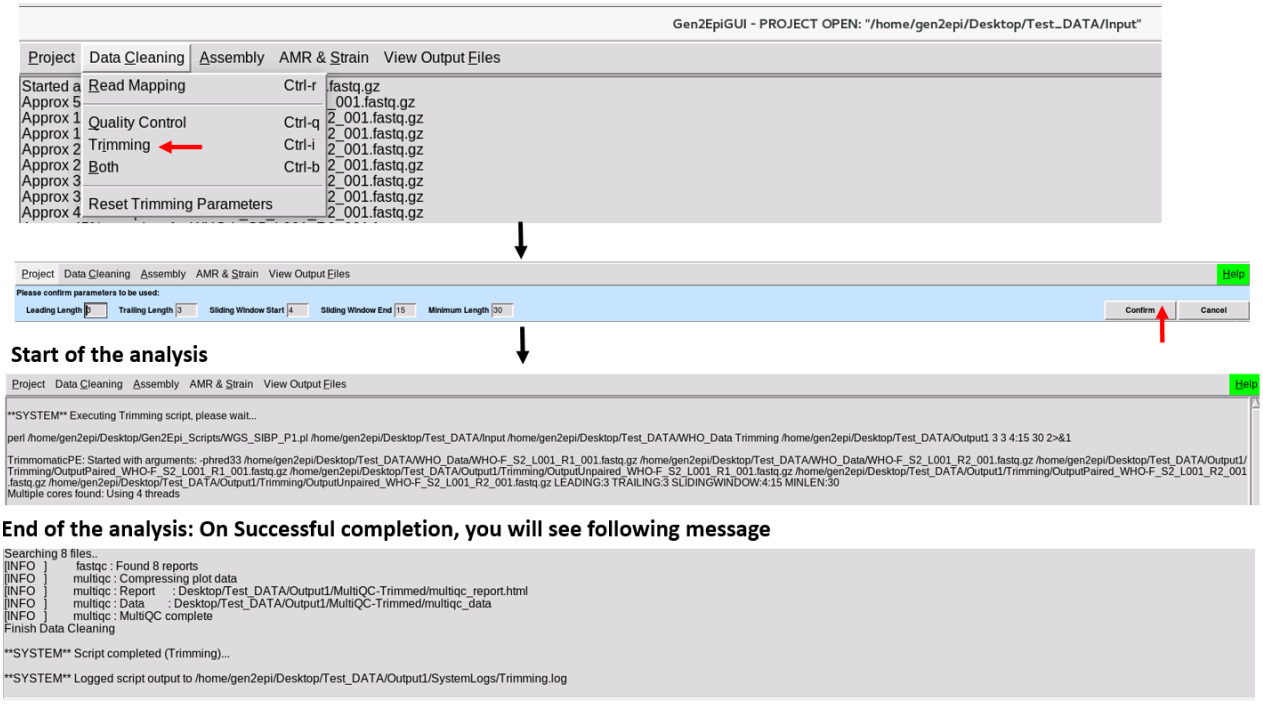


End of the analysis: On Successful completion, you will see following message



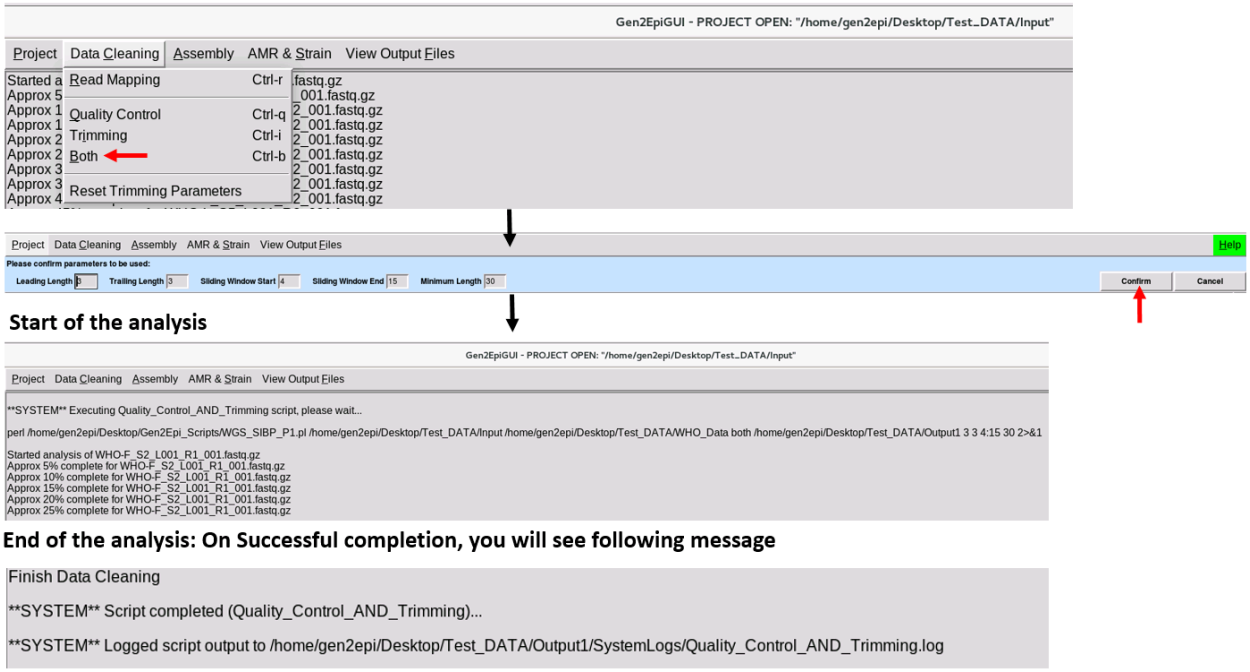
Note: Quality control results are present under '*/home/gen2epi/Desktop/Test_DATA/Output1*'. There will be two folders "*MultiQC-Raw*" and "*QualityControl*". Please open the .html files under these folders to visualize the quality of each sample.

11) In order to trim the raw reads, users have to click on the trimming submenu that will further ask for the confirmation of the parameters. 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.



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 files 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 the "*MultiQC-Raw*", "*quality control*", "*MultiQC-Trimmed*", "*Trimming*" and "*Trimmed_QC*" folders. Now go to "Gen2EpiGUI" and click on the both tab under Data Cleaning menu.

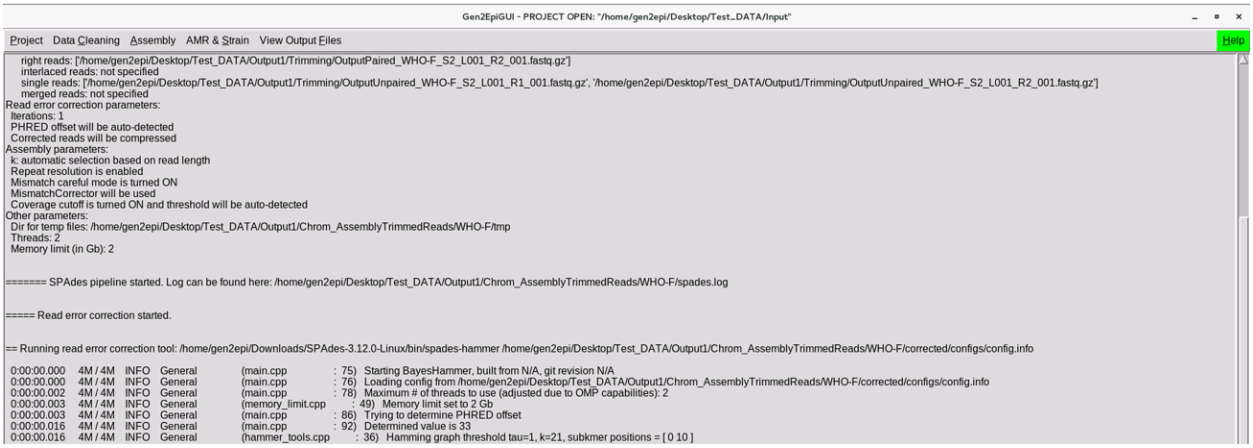


End of the analysis: On Successful completion, you will see following message

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** menu. 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.

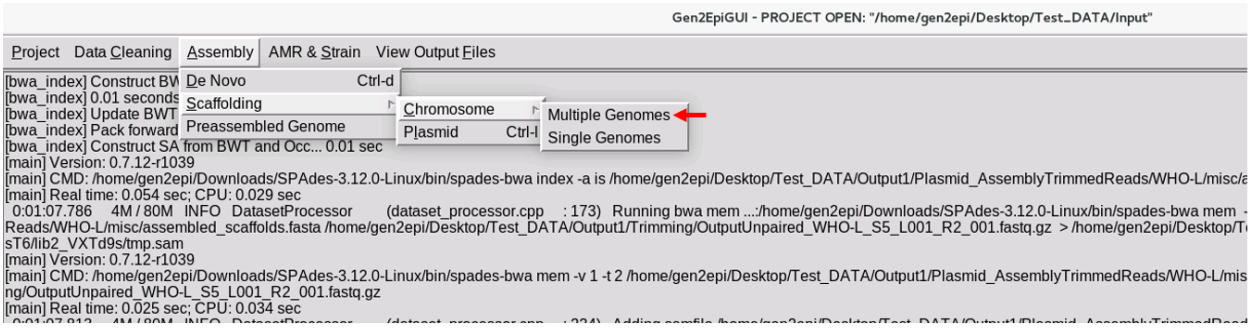


14) Once the *de novo* assembly (step 13) is completed. You will see the following message

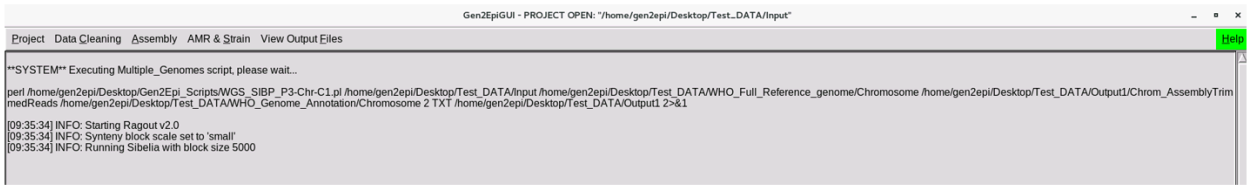
```
===== Mismatch correction finished.

* 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 GFA format is in /home/gen2epi/Desktop/Test_DATA/Output1/Plasmid_AssemblyTrimmedReads/WHO-L/assembly_graph_with_scaffolds.gfa
* 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: You 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:



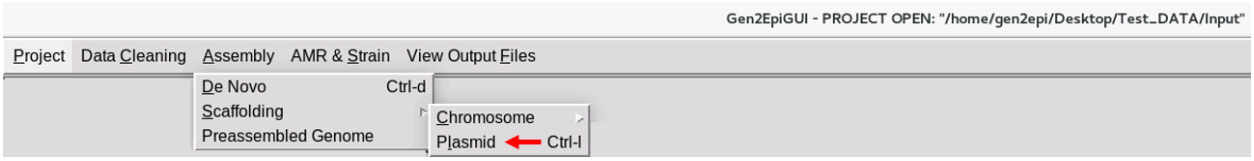
Start of the Scaffolding process



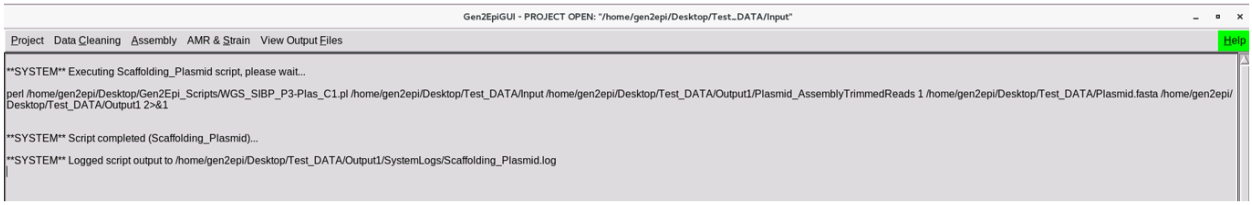
Scaffolding process completion

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

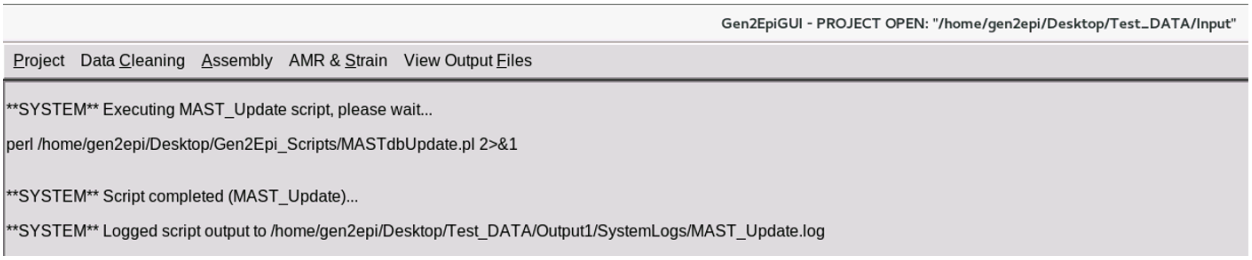


Note: Output generated from this step will be stored in folder "*Plasmid_Identification*" under '*/home/gen2epi/Desktop/Test_DATA/Output1*'.

17) NgMAST Typing

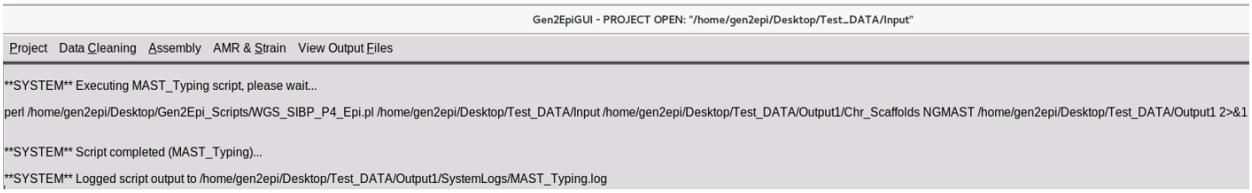
a. Update the underlying MAST database by clicking on AMR & Strain -> MAST -> MAST Update submenu.

MAST database update

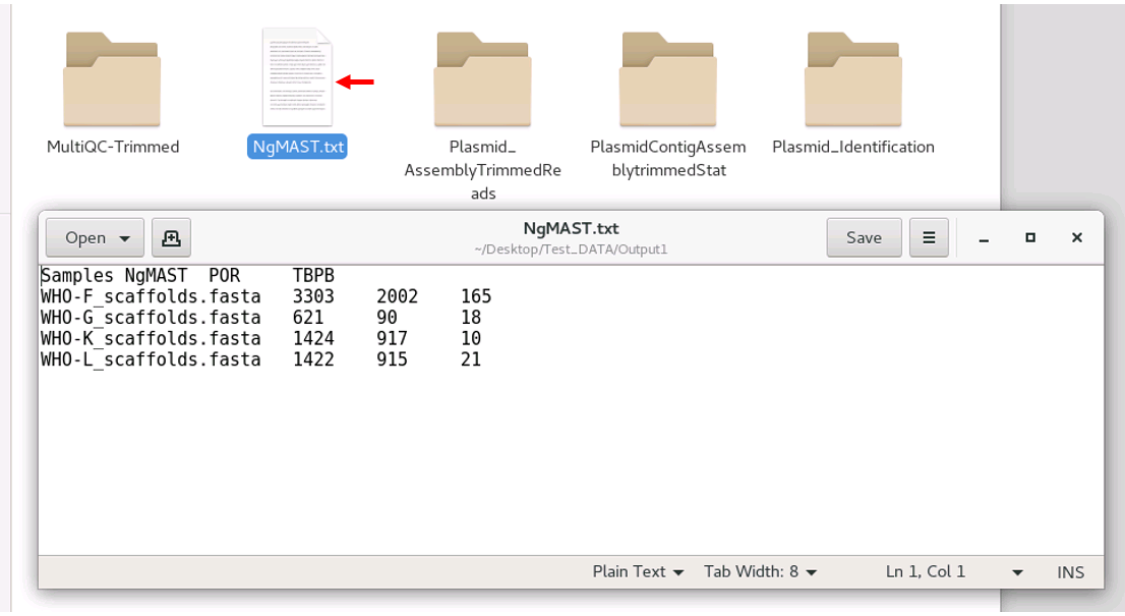


b. Perform the MAST typing by selecting AMR & Strain -> MAST -> MAST Typing submenu

Strain typing

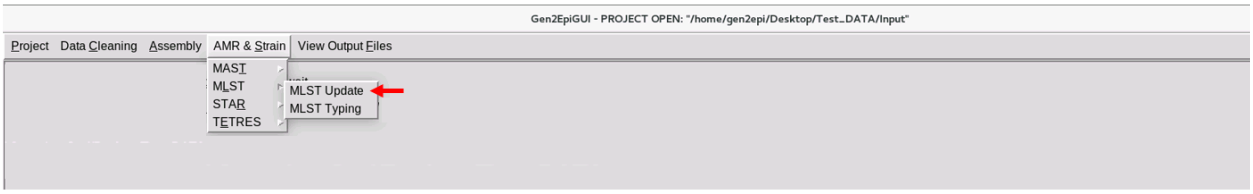


Note: MAST strain typing for each sample is present in file "*NgMAST.txt*" under '*/home/gen2epi/Desktop/Test_DATA/Output1'* as shown below.

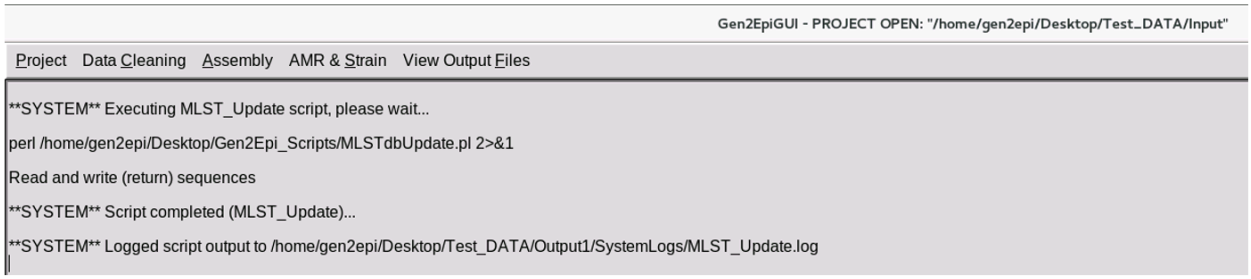


18) NgMLST Typing

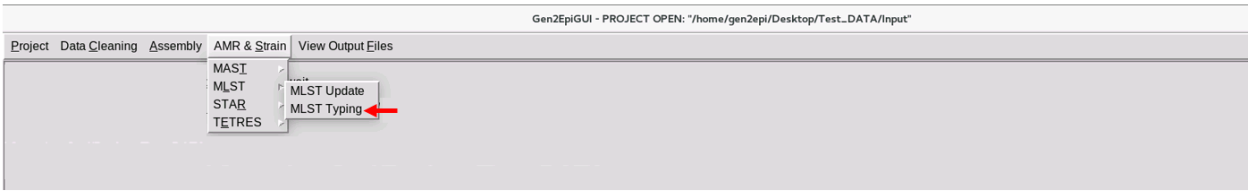
a. Update the underlying MLST database by clicking on AMR & Strain -> MLST -> MLST Update submenu.



MLST database update



b. Perform the MLST typing by selecting AMR & Strain -> MLST -> MLST Typing submenu.



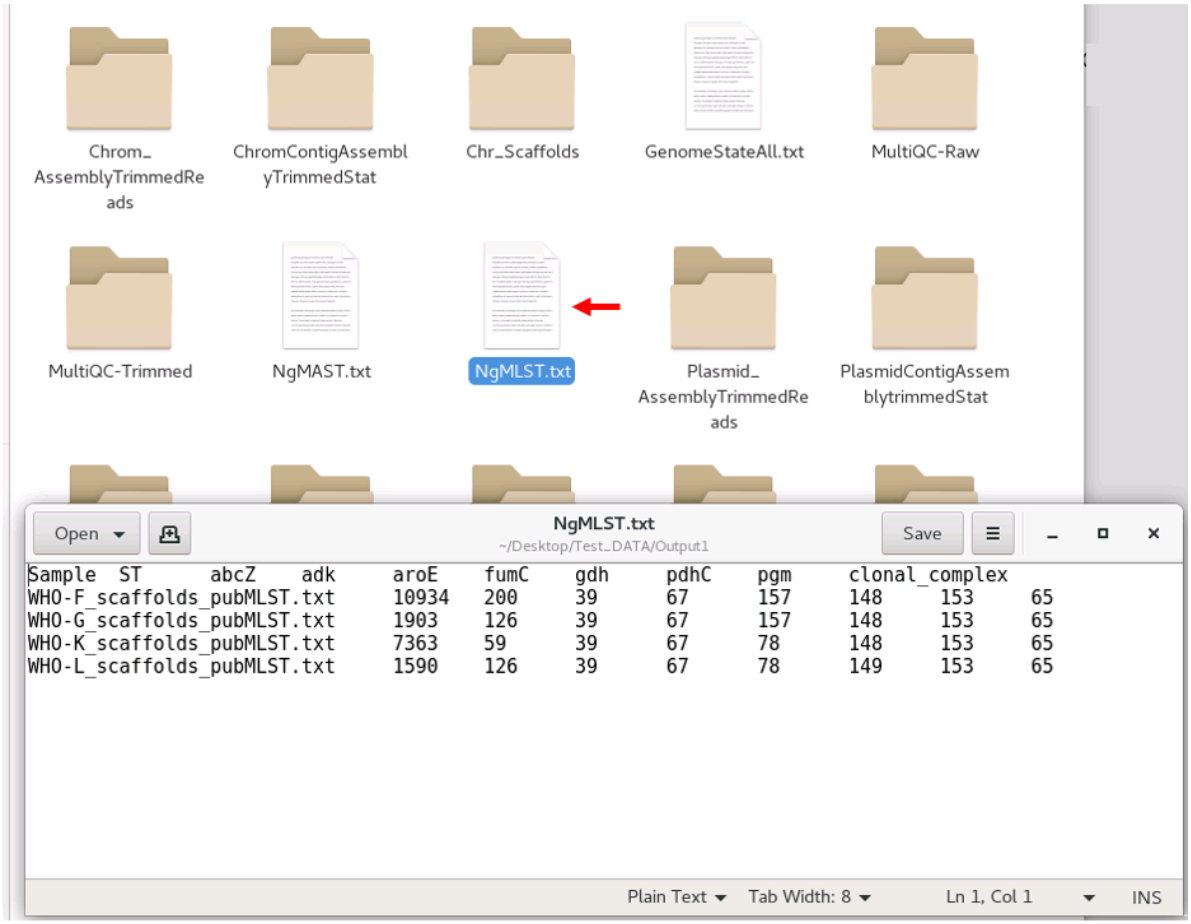
Start of the strain typing analysis

```
**SYSTEM** Executing MLST_Typing script, please wait...

perl /home/gen2epi/Desktop/Gen2Epi_Scripts/WGS_SIBP_P4_Epi.pl /home/gen2epi/Desktop/Test_DATA/Input /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds MLST /home/gen2epi/Desktop/Test_DATA/Output1 /home/gen2epi/Desktop/Gen2Epi_Scripts/
MLST-Genes.fasta /home/gen2epi/Desktop/Gen2Epi_Scripts/MLST_alleles.fasta /home/gen2epi/Desktop/Gen2Epi_Scripts/pubMLST_profile.txt 2>&1
```

Completion of the strain typing analysis

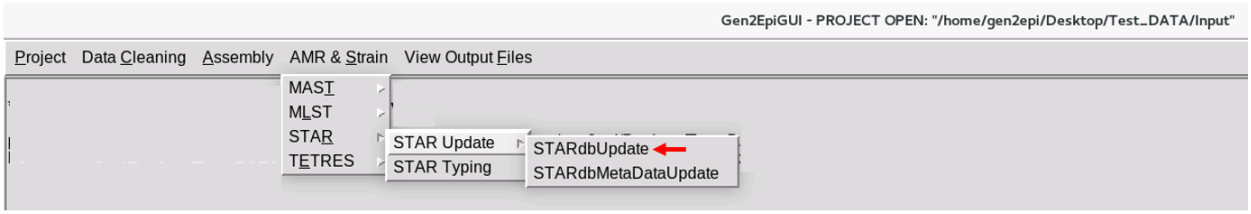
```
Building a new DB, current time: 06/03/2019 10:38:30
New DB name:   /home/gen2epi/Desktop/Gen2Epi_Scripts/MLST_alleles.fasta
New DB title:  /home/gen2epi/Desktop/Gen2Epi_Scripts/MLST_alleles.fasta
Sequence type: Nucleotide
Deleted existing Nucleotide BLAST database named /home/gen2epi/Desktop/Gen2Epi_Scripts/MLST_alleles.fasta
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 7219 sequences in 0.144738 seconds.

**SYSTEM** Script completed (MLST_Typing)...

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

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

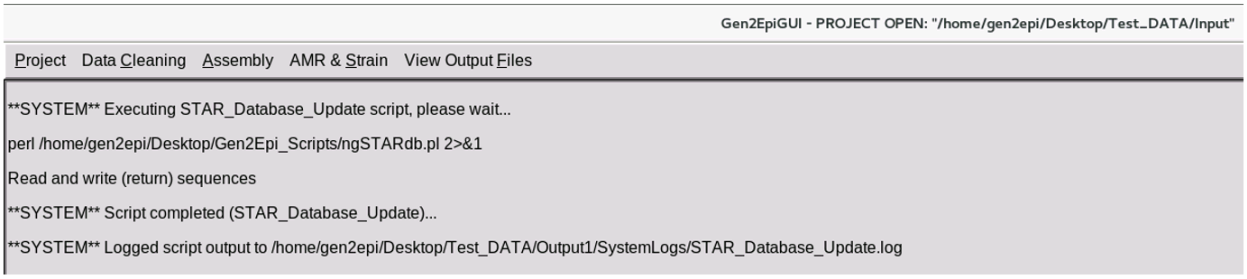


19) NgSTAR

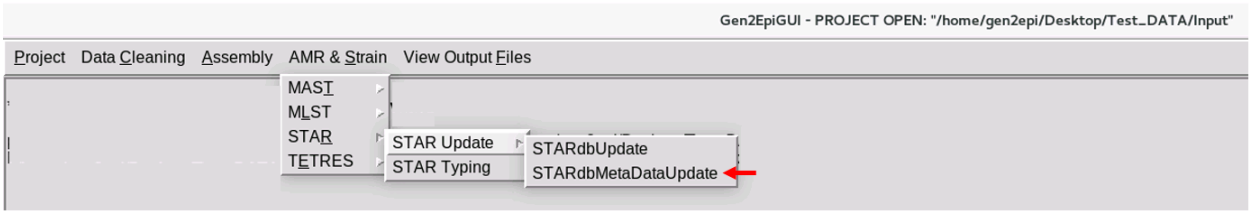
a. Update the underlying STAR database by clicking on AMR & Strain -> STAR -> STAR Update -> STARdbUpdate submenu.



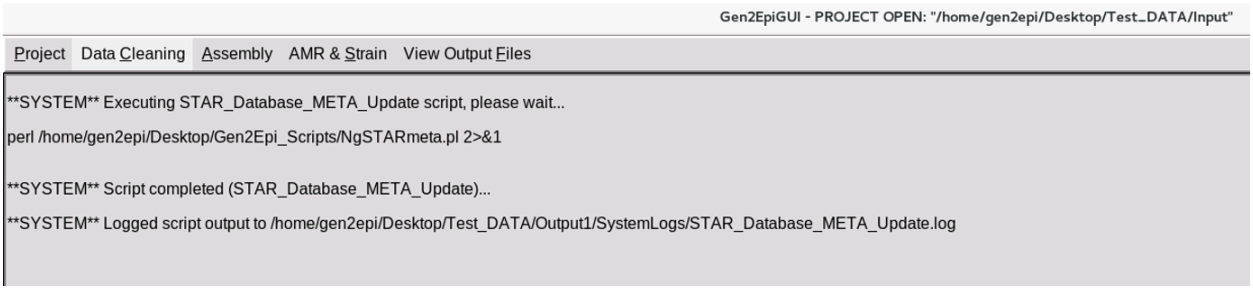
STAR database update



b. Update the underlying STAR database metadata by clicking on AMR & Strain -> STAR -> STAR Update -> STARdbMetaDataUpdate submenu.

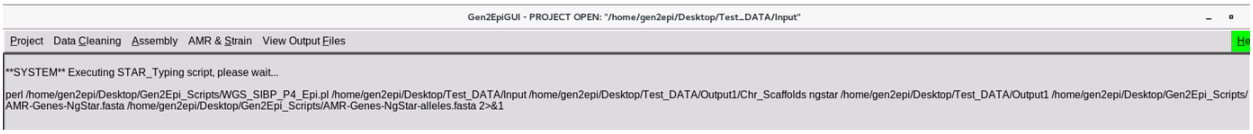


STAR database metadata update

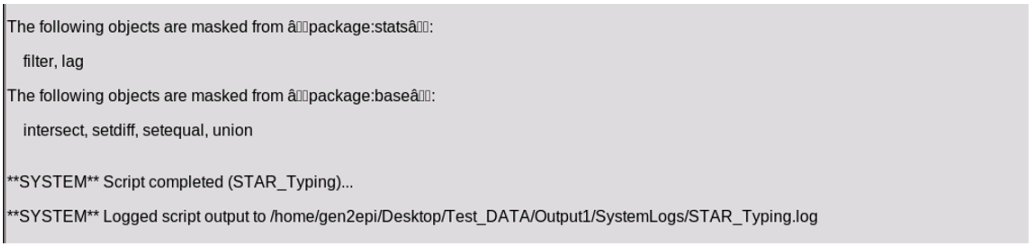


c. Perform the STAR typing by selecting AMR & Strain -> STAR -> STAR Typing submenu.

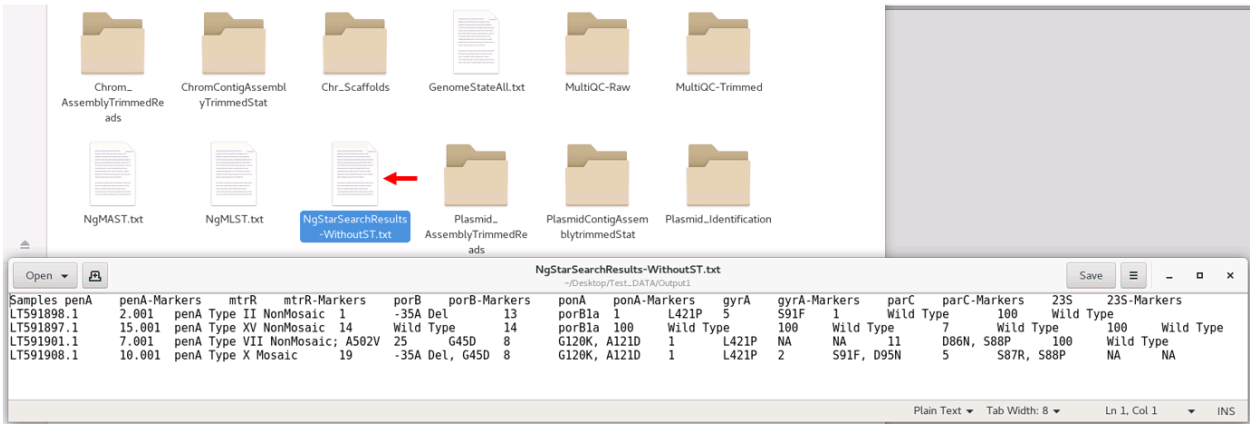
Start of the STAR typing process



Completion of the STAR typing process

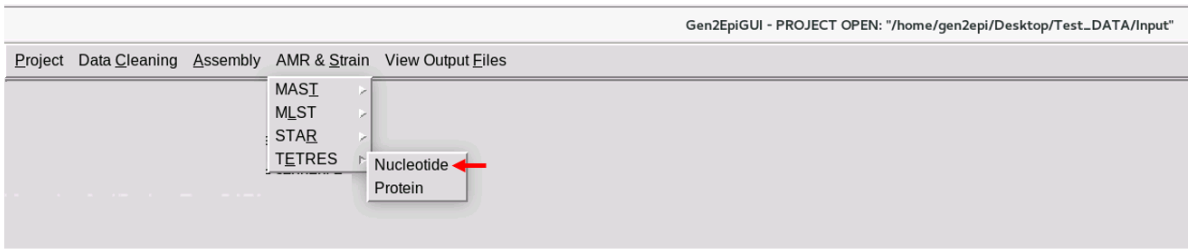


Note: NgSTAR output for each sample is present in file "*NgStarSearchResults-WithoutST.txt*" under '*/home/gen2epi/Desktop/Test_DATA/Output1*' as shown below.



20) Tetracycline Resistance

1) Click on TETRES -> Nucleotide submenu under AMR & Strain menu to extract the rpsJ nucleotide sequences. These fasta sequences can be visualized under any multiple sequence alignment program.



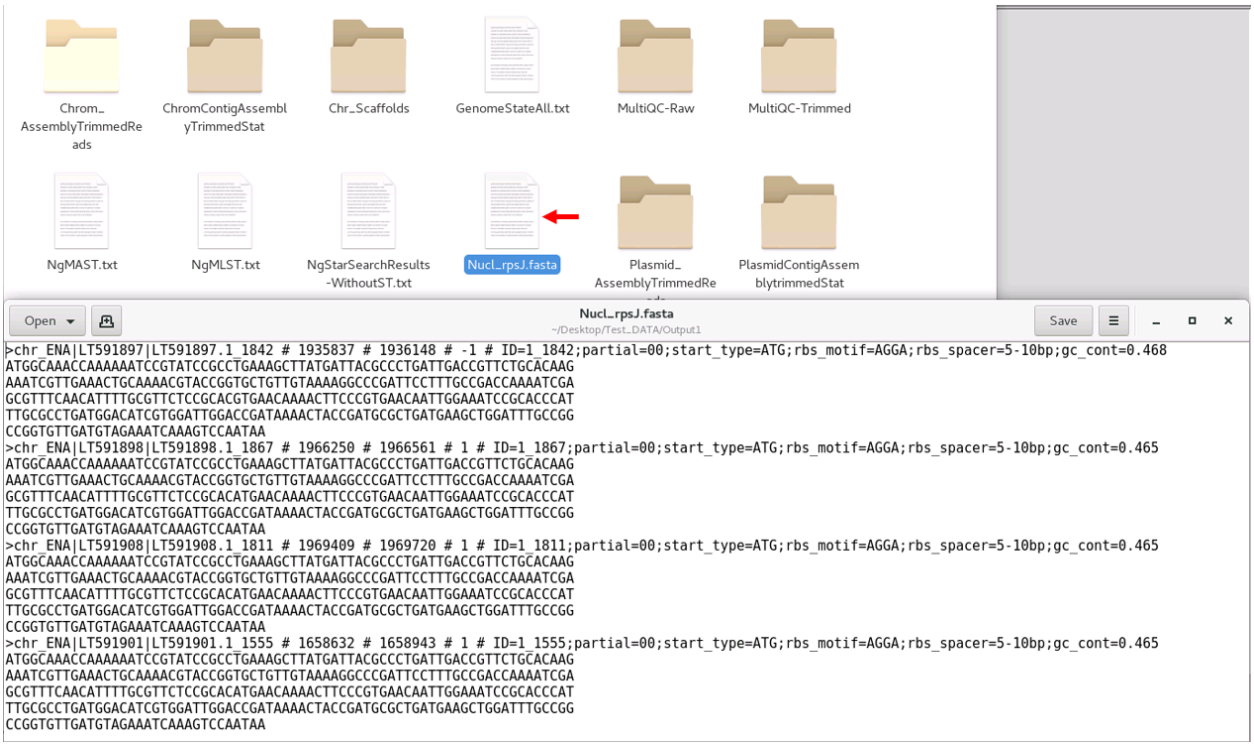
Starting

Completion

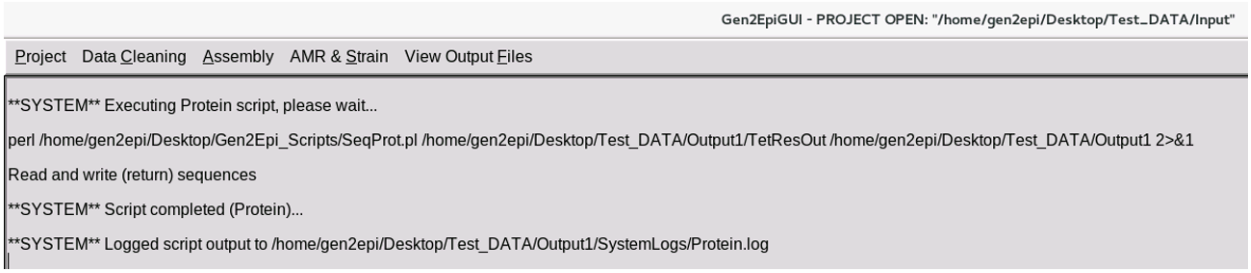
```
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-F.nin_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-F.nhr_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-F.nsq_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-G.nin_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-G.nhr_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-G.nsq_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-K.nin_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-K.nhr_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-K.nsq_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-L.nin_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-L.nhr_scaffolds_nucl.fasta: No such file or directory.
Can't open /home/gen2epi/Desktop/Test_DATA/Output1/Chr_Scaffolds/All_Sequences/WHO-L.nsq_scaffolds_nucl.fasta: No such file or directory.

**SYSTEM** Script completed (Nucleotide)...

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

Don't worry about this message

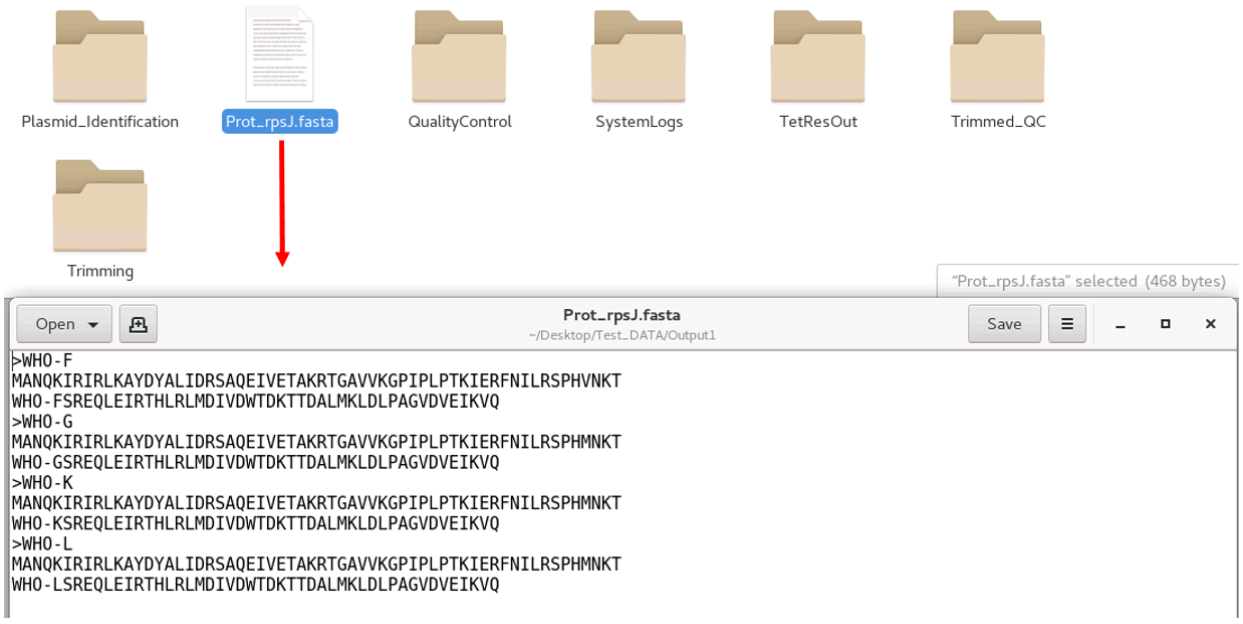
<u>Note</u>: You can see the rpsJ nucleotide sequence for each sample in file "*<u>Nucl_rpsJ.fasta</u>*" under '*/home/gen2epi/Desktop/Test_DATA/Output1*' as shown below.


Chrom_AssemblyTrimmedReads
ChromContigAssemblyTrimmedStat
Chr_Scaffolds
GenomeStateAll.txt
MultiQC-Raw
MultiQC-Trimmed
NgMAST.txt
NgMLST.txt
NgStarSearchResults-WithoutST.txt
Nucl_rpsJ.fasta
Plasmid_AssemblyTrimmedRe
PlasmidContigAssemblytrimmedStat
Open
Nucl_rpsJ.fasta
~/Desktop/Test_DATA/Output1
Save
>chr_ENA|LT591897|LT591897.1_1842 # 1935837 # 1936148 # -1 # ID=1_1842;partial=00;start_type=ATG;rbs_motif=AGGA;rbs_spacer=5-10bp;gc_cont=0.468
ATGGCAAACCAAAAAATCCGTATCCGCCTGAAAGCTTATGATTACGCCCTGATTGACCGTTCTGCACAAG
AAATCGTTGAAACTGCAAAACGTACCGGTGCTGTTGTAAAAGGCCCGATTCCTTTGCCGACCAAAATCGA
GCGTTTCAACATTTTGCGTTCTCCGCACGTGAACAAAACTTCCCGTGAACAATTGGAAATCCGCACCCAT
TTGCGCCTGATGGACATCGTGGATTGGACCGATAAAACTACCGATGCGCTGATGAAGCTGGATTTGCCGG
CCGGTGTTGATGTAGAAATCAAAGTCCAATAA
>chr_ENA|LT591898|LT591898.1_1867 # 1966250 # 1966561 # 1 # ID=1_1867;partial=00;start_type=ATG;rbs_motif=AGGA;rbs_spacer=5-10bp;gc_cont=0.465
ATGGCAAACCAAAAAATCCGTATCCGCCTGAAAGCTTATGATTACGCCCTGATTGACCGTTCTGCACAAG
AAATCGTTGAAACTGCAAAACGTACCGGTGCTGTTGTAAAAGGCCCGATTCCTTTGCCGACCAAAATCGA
GCGTTTCAACATTTTGCGTTCTCCGCACATGAACAAAACTTCCCGTGAACAATTGGAAATCCGCACCCAT
TTGCGCCTGATGGACATCGTGGATTGGACCGATAAAACTACCGATGCGCTGATGAAGCTGGATTTGCCGG
CCGGTGTTGATGTAGAAATCAAAGTCCAATAA
>chr_ENA|LT591908|LT591908.1_1811 # 1969409 # 1969720 # 1 # ID=1_1811;partial=00;start_type=ATG;rbs_motif=AGGA;rbs_spacer=5-10bp;gc_cont=0.465
ATGGCAAACCAAAAAATCCGTATCCGCCTGAAAGCTTATGATTACGCCCTGATTGACCGTTCTGCACAAG
AAATCGTTGAAACTGCAAAACGTACCGGTGCTGTTGTAAAAGGCCCGATTCCTTTGCCGACCAAAATCGA
GCGTTTCAACATTTTGCGTTCTCCGCACATGAACAAAACTTCCCGTGAACAATTGGAAATCCGCACCCAT
TTGCGCCTGATGGACATCGTGGATTGGACCGATAAAACTACCGATGCGCTGATGAAGCTGGATTTGCCGG
CCGGTGTTGATGTAGAAATCAAAGTCCAATAA
>chr_ENA|LT591901|LT591901.1_1555 # 1658632 # 1658943 # 1 # ID=1_1555;partial=00;start_type=ATG;rbs_motif=AGGA;rbs_spacer=5-10bp;gc_cont=0.465
ATGGCAAACCAAAAAATCCGTATCCGCCTGAAAGCTTATGATTACGCCCTGATTGACCGTTCTGCACAAG
AAATCGTTGAAACTGCAAAACGTACCGGTGCTGTTGTAAAAGGCCCGATTCCTTTGCCGACCAAAATCGA
GCGTTTCAACATTTTGCGTTCTCCGCACATGAACAAAACTTCCCGTGAACAATTGGAAATCCGCACCCAT
TTGCGCCTGATGGACATCGTGGATTGGACCGATAAAACTACCGATGCGCTGATGAAGCTGGATTTGCCGG
CCGGTGTTGATGTAGAAATCAAAGTCCAATAA
```
```

2) Click on TETRES -> protein submenu under AMR & Strain menu to extract the rpsJ protein alignment. These fasta sequences can be visualized under any multiple sequence alignment program.



Note: You can see the rpsJ nucleotide sequence for each sample in the 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.