Gen2Epi GUI Quick Tutorial

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

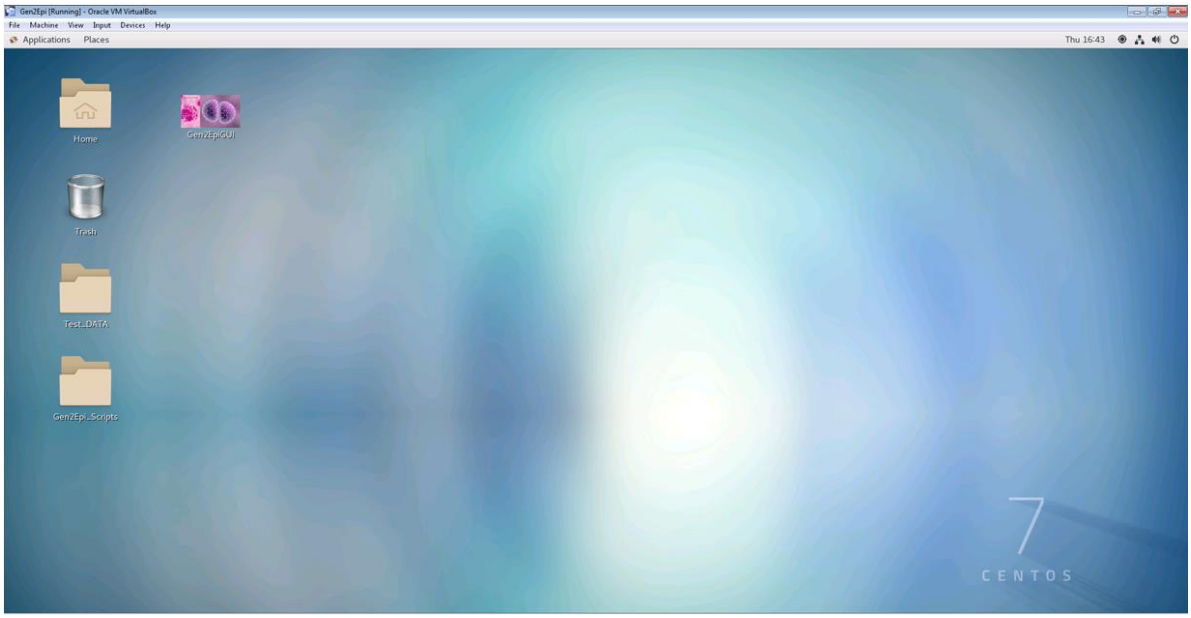
Main Data directory: "/home/gen2epi/Desktop/Test_DATA"

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"

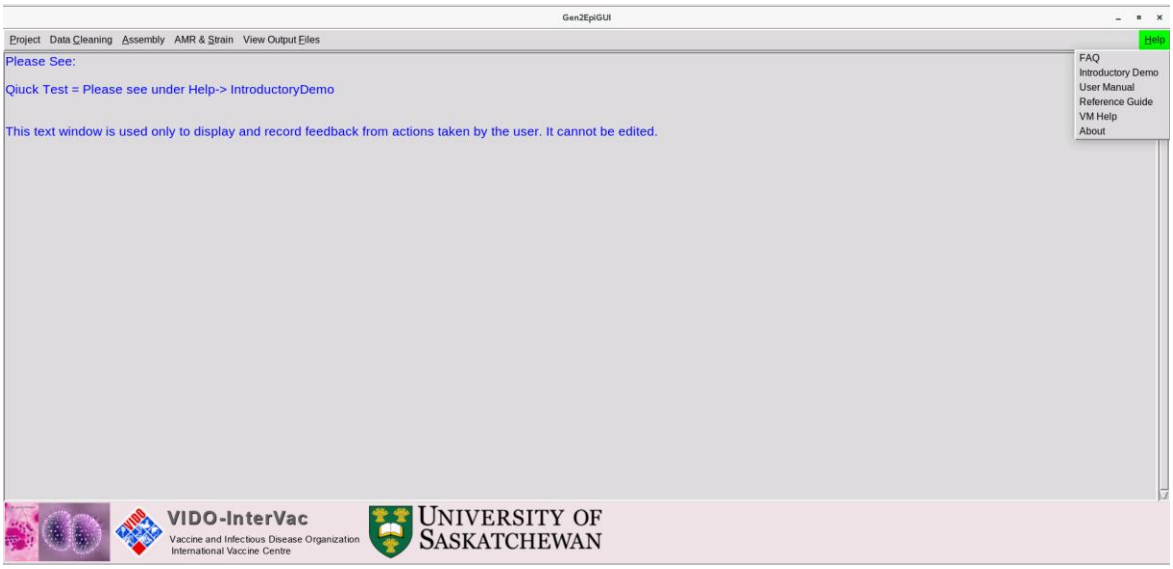
1) Click on the "Gen2EpiGUI" icon on the Desktop to start the program.



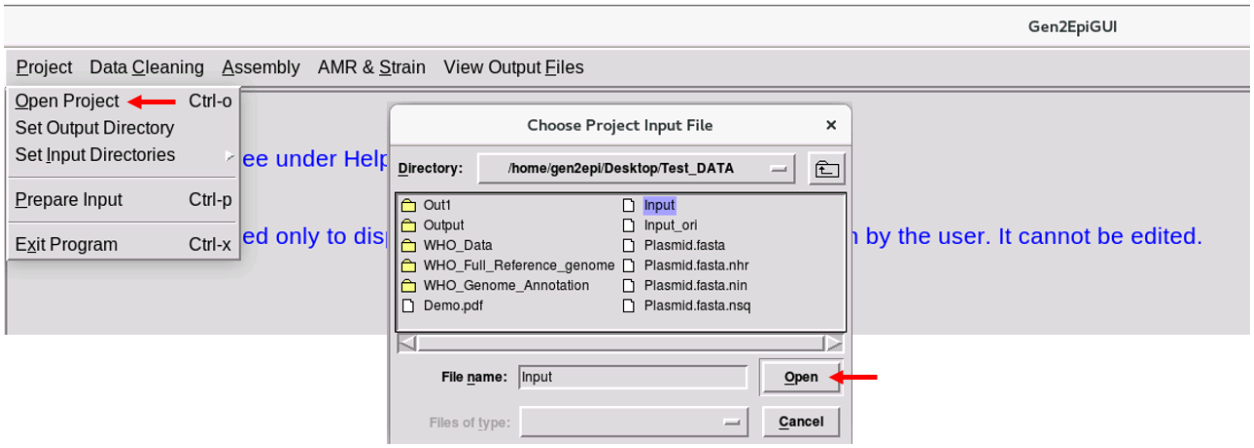
2) To use the program please follow the instruction as given in "Introductory Demo" under Help menu on the top right-hand side. For detailed usage please see the description below:

    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.

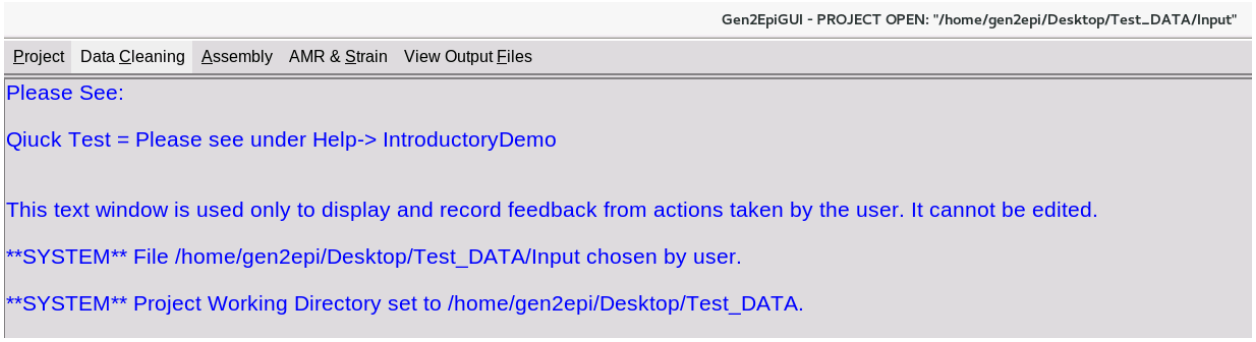
Please Note: For beginners "Introductory Demo" is highly recommended.



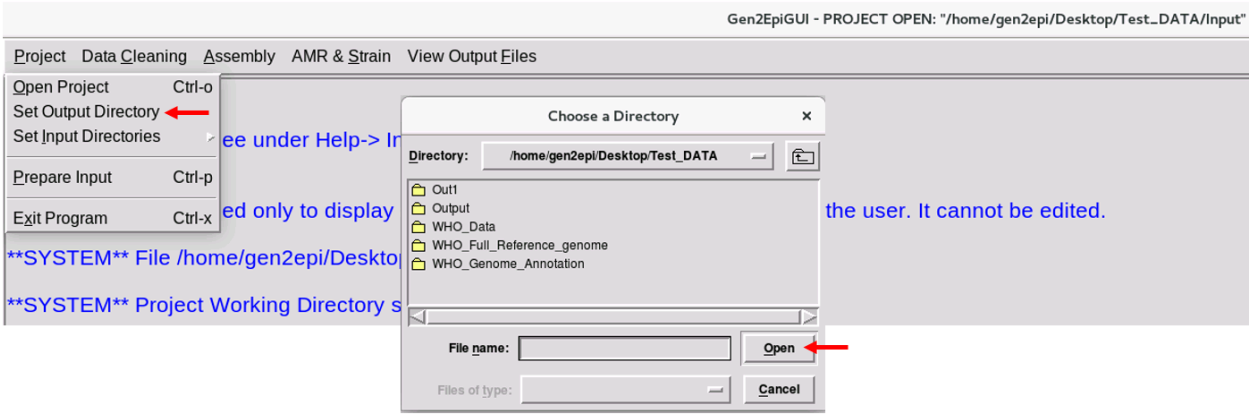
3) Load the input text (tab separated text file describes the name of all fastq files) file by clicking on the "Open Project" and browse (and open) the file as shown in the picture below. "Input" file is present in the Test_DATA folder at Desktop.



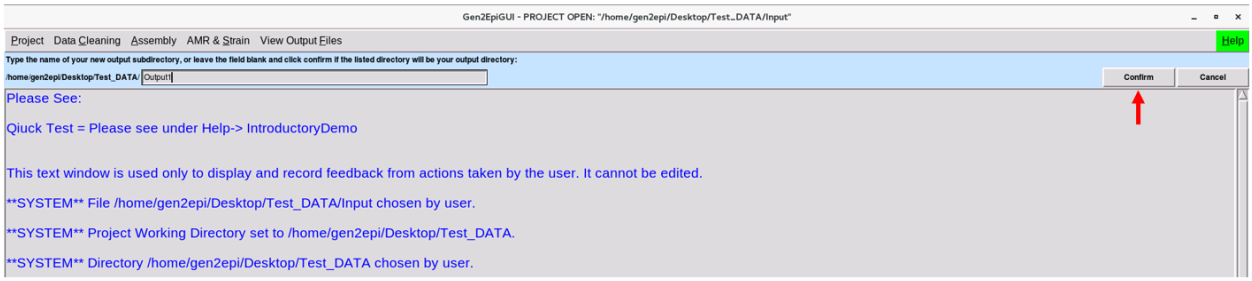
You will see following **SYSTEM** message in the window which shows that the "Input" file is successfully loaded.



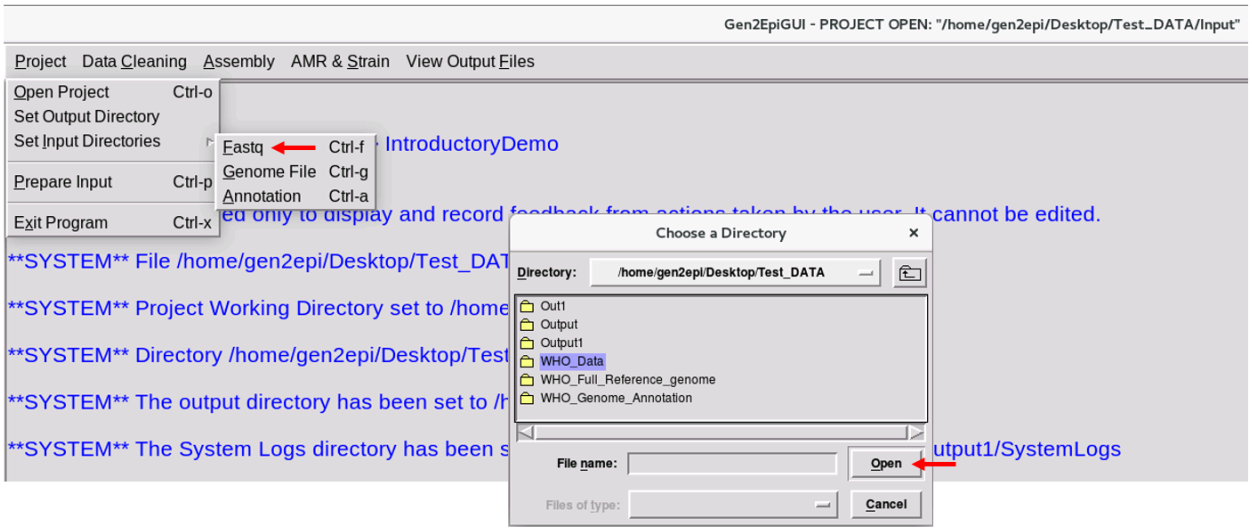
4) Set the output directory by clicking on “Set Output Directory” under “Project” menu. Browse the path (where you want to save the output files generated by Gen2Epi e.g. in this case path is “/home/gen2epi/Desktop/Test_DATA/”) and click open.



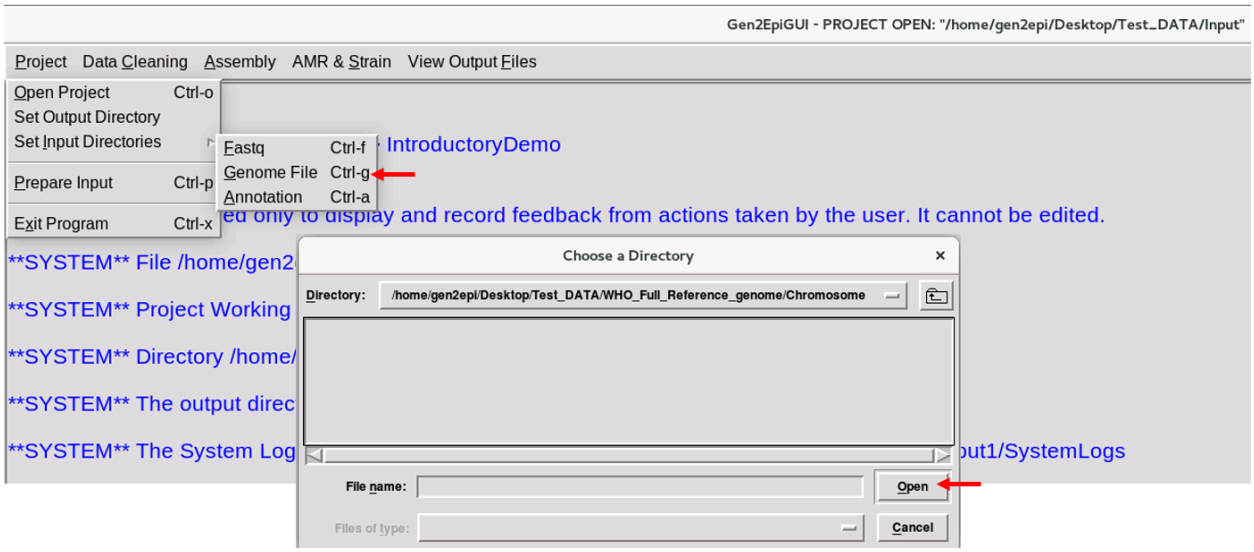
5) Write the name of your output directory (e.g. Output1) in the search box and click confirm.



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:

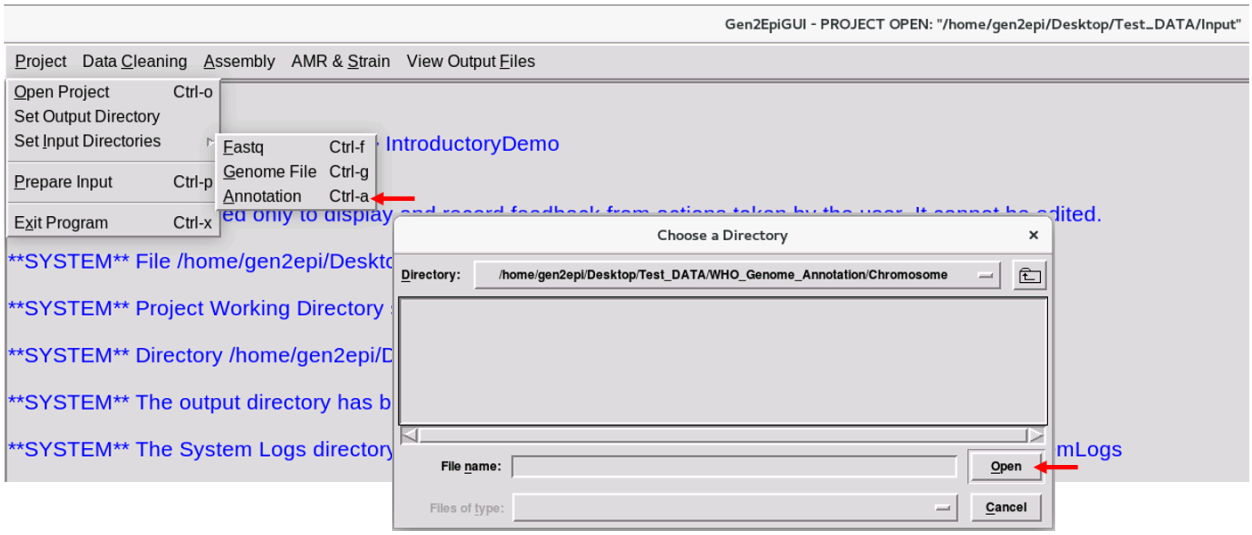


7) Set Input directories for “genome” files by clicking on “Set Input Directories”-> Genome File”. Select the folder “/home/gen2epi/Desktop/Test_DATA/WHO_Full_Reference_genome/Chromosome” and click “Open” as shown in the figure below:



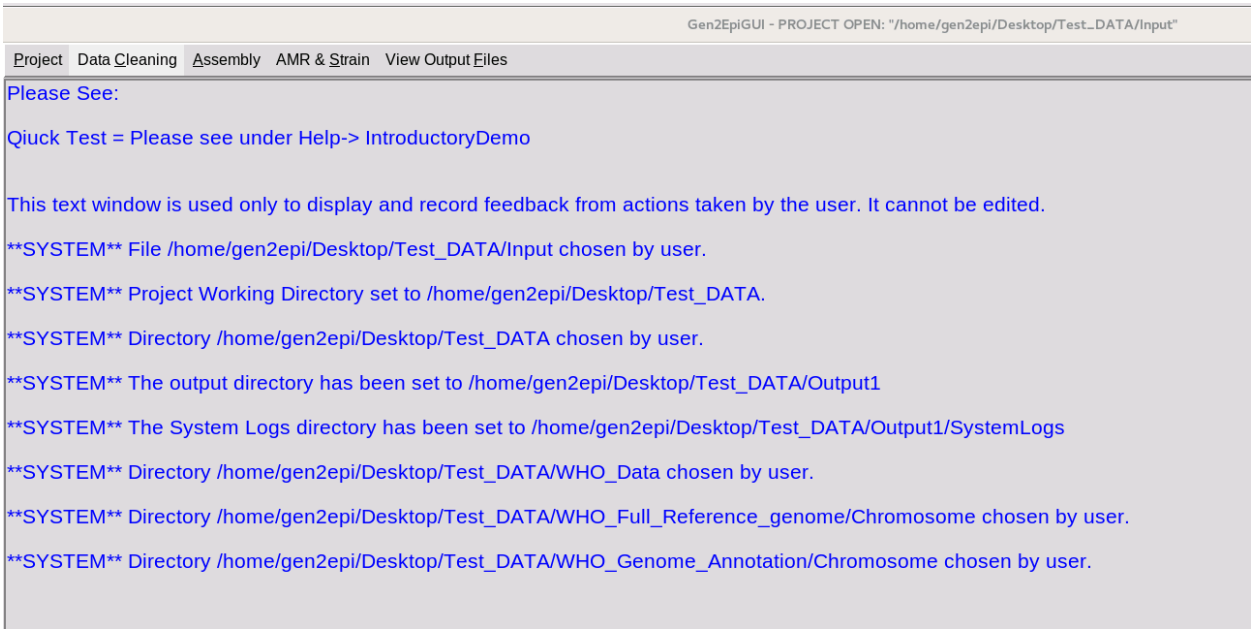
Please Note: You will not be able to see anything when you choose a directory as shown above but that’s all right.

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:

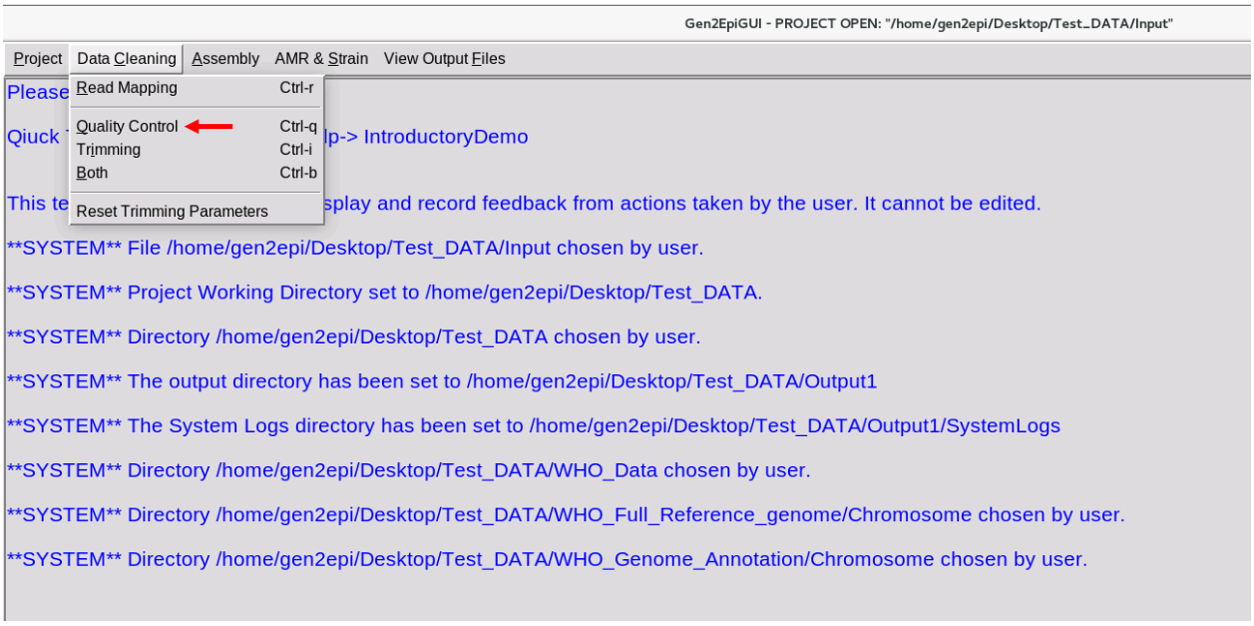


Again, you will not be able to see anything under your chosen directory as shown above but that’s all right.

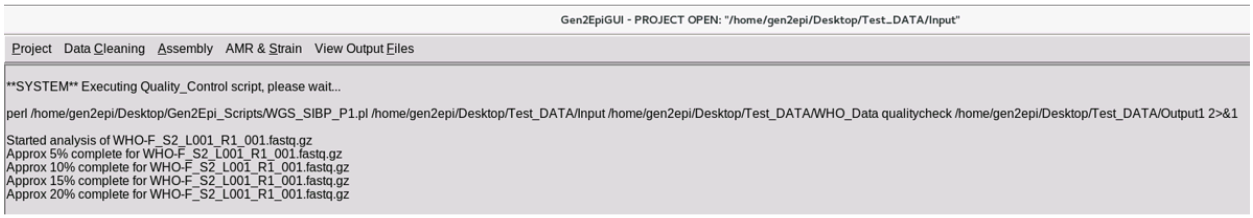
9) Once all paths are set properly you will see following **SYSTEM** MESSAGES in Gen2Epi window



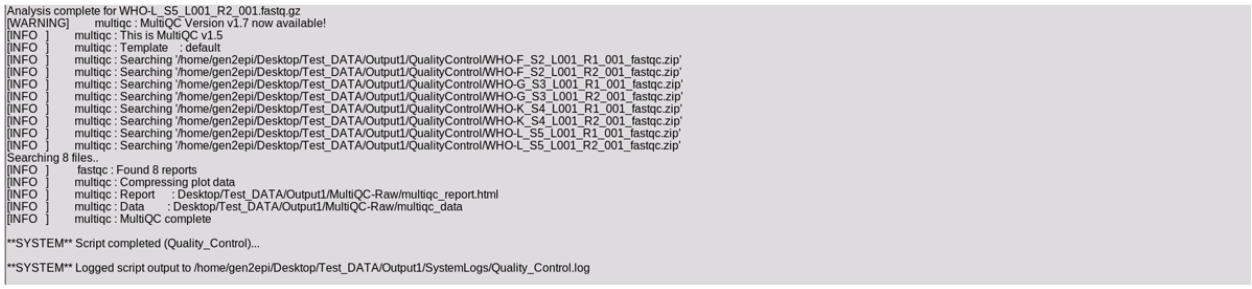
10) Now to clean your data: Under "Data Cleaning" menu – click on "Quality Control" submenu to check the quality of your raw fastq files.



Start of the analysis

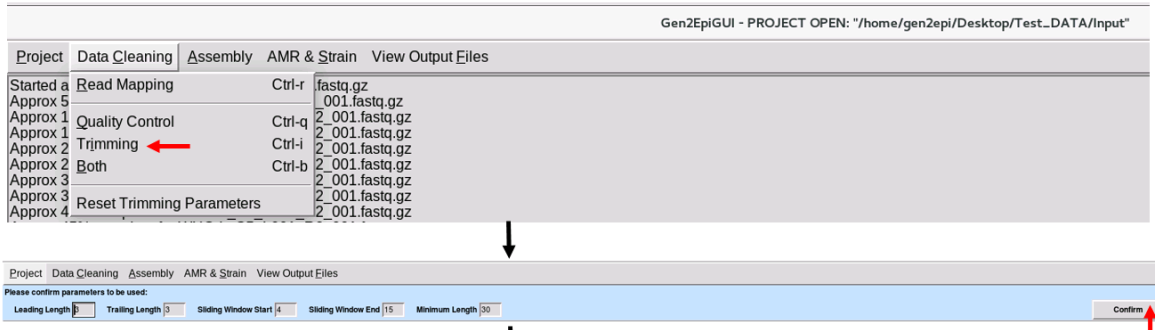


End of the analysis: On Successful completion, you will see following message

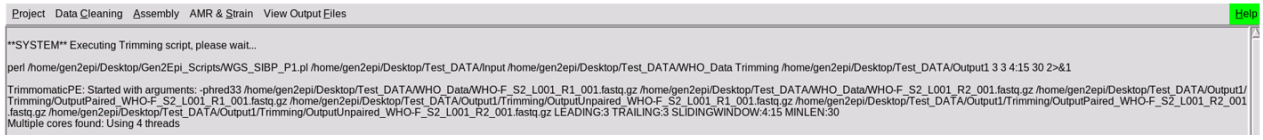


Note: Quality control results are present under "/home/gen2epi/Desktop/Test_DATA/Output1". There will be two folders "MultiQC-Raw" and "QualityControl". Please open the .html file under these folders to visualize the quality of each sample.

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.



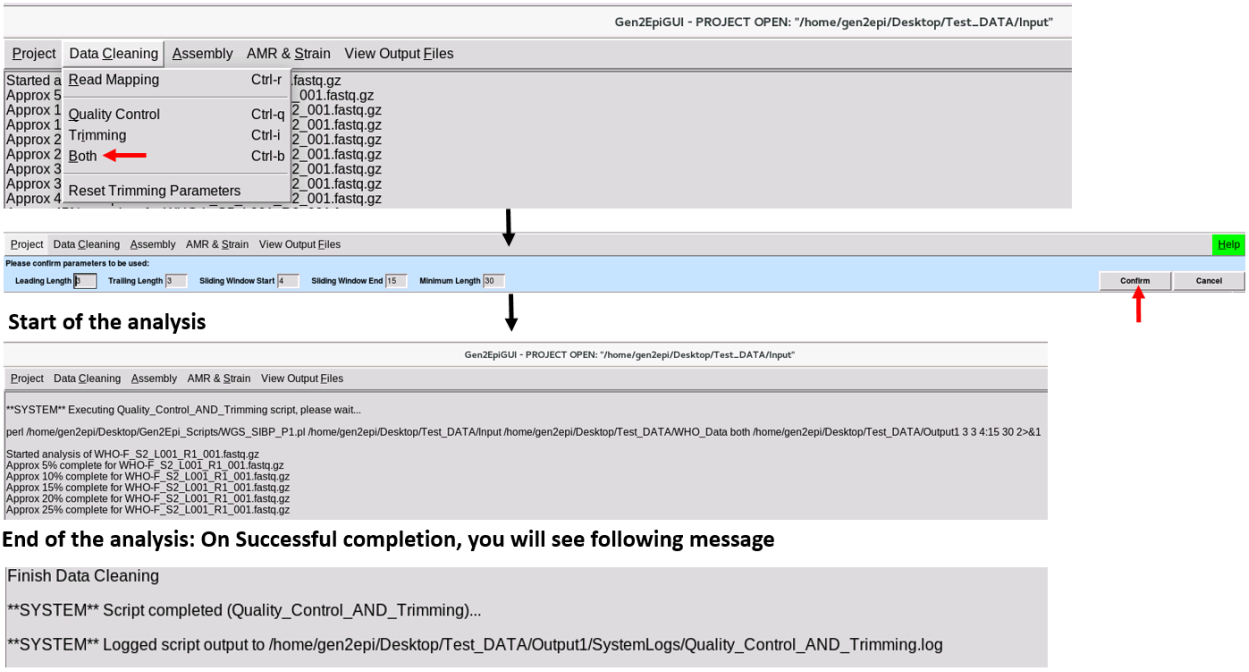
Start of the analysis



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

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.

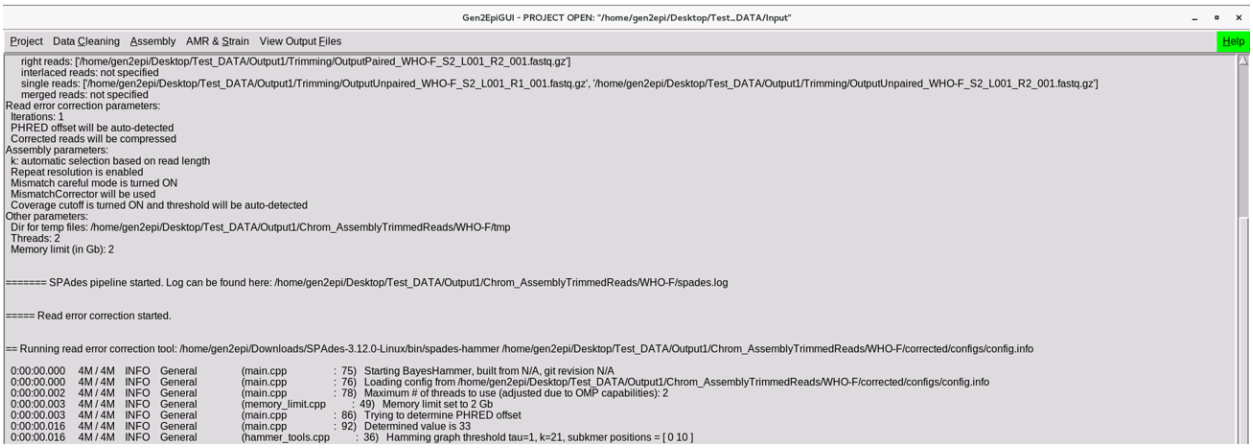


End of the analysis: On Successful completion, you will see following message

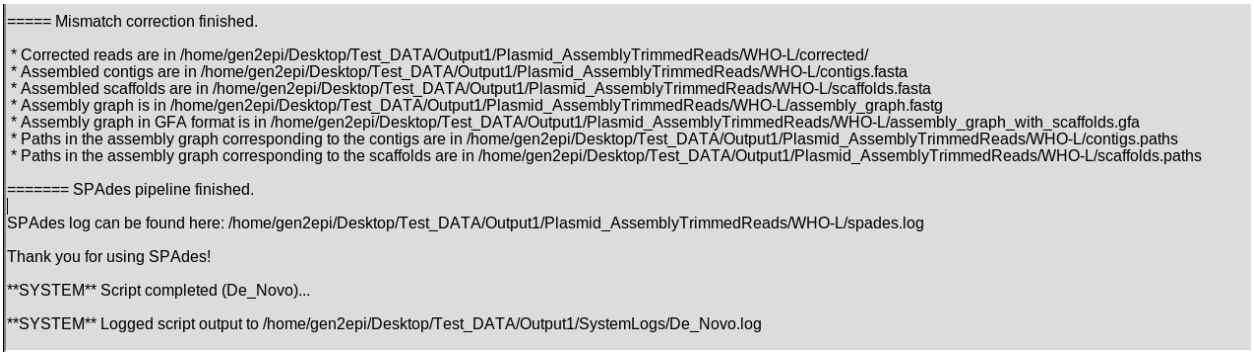
Note: - You will find all the results under "/home/gen2epi/Desktop/Test_DATA/Output1" in five different folders i.e. "MultiQC-Raw", "quality control", "MultiQC-Trimmed", "Trimming" and "Trimmed_QC"

13) Now, perform the *de novo* assembly of trimmed reads by clicking on "*De Novo*" under Assembly 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.

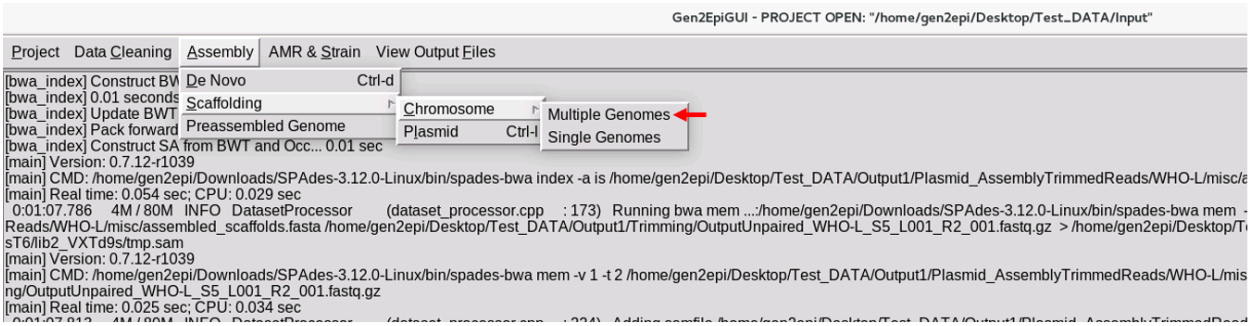


14) Once the *de novo* assembly (step 13) complete. You will see the following message

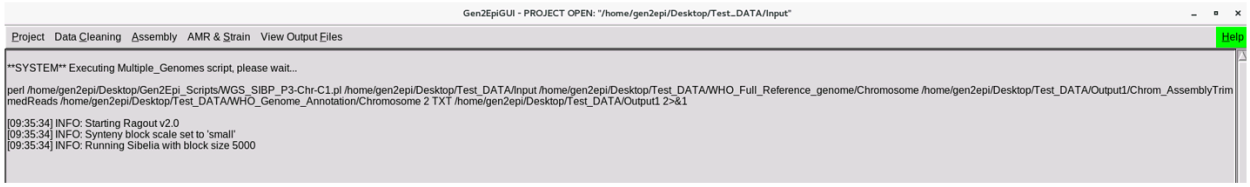


Note: Users will find the results under "/home/gen2epi/Desktop/Test_DATA/Output1" in four folder i.e. Chrom_AssemblyTrimmedReads (Assembled chromosome contigs in FASTA format generated from trimmed reads), ChromContigAssemblyTrimmedStat (Chromosome assembly statistics), Plasmid_AssemblyTrimmedReads (Assembled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



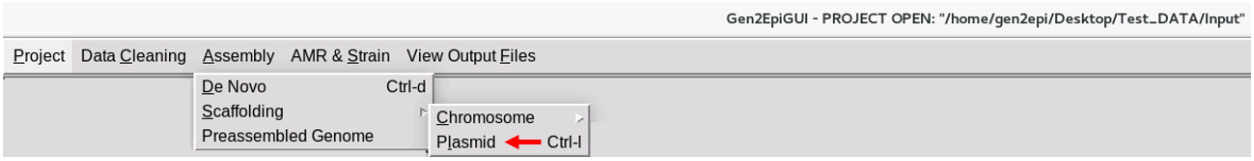
Start of the Scaffolding process



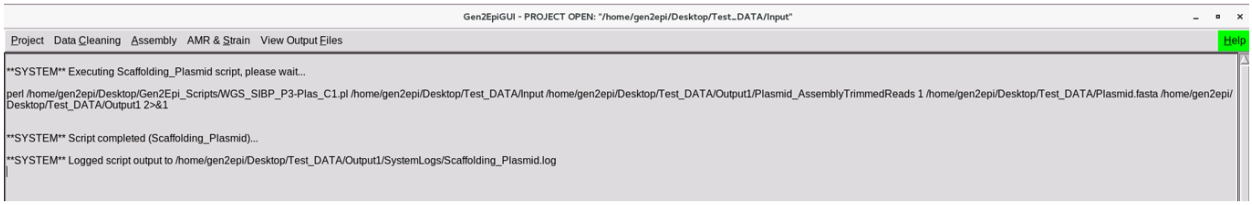
Scaffolding process completion

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

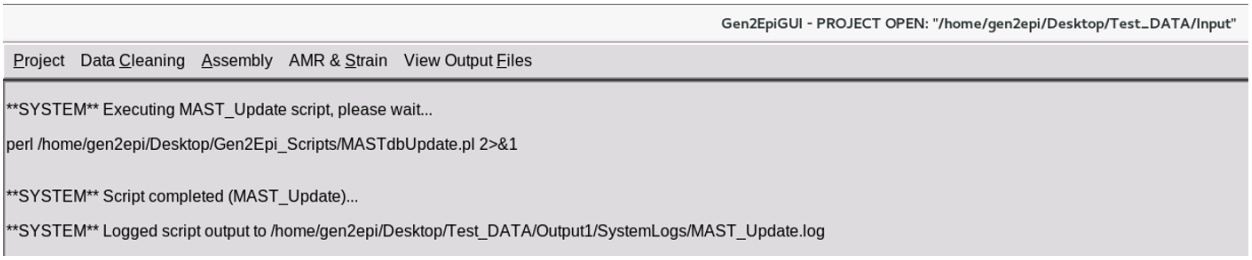


Note: Output generated from this step will be stored in folder "Plasmid_Identification" under "/home/gen2epi/Desktop/Test_DATA/Output1".

17) NgMAST Typing

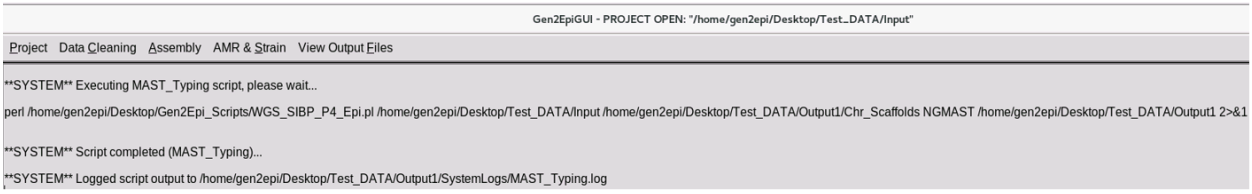
a. Update the underlying MAST database by clicking MAST->MAST Update tab

MAST database update

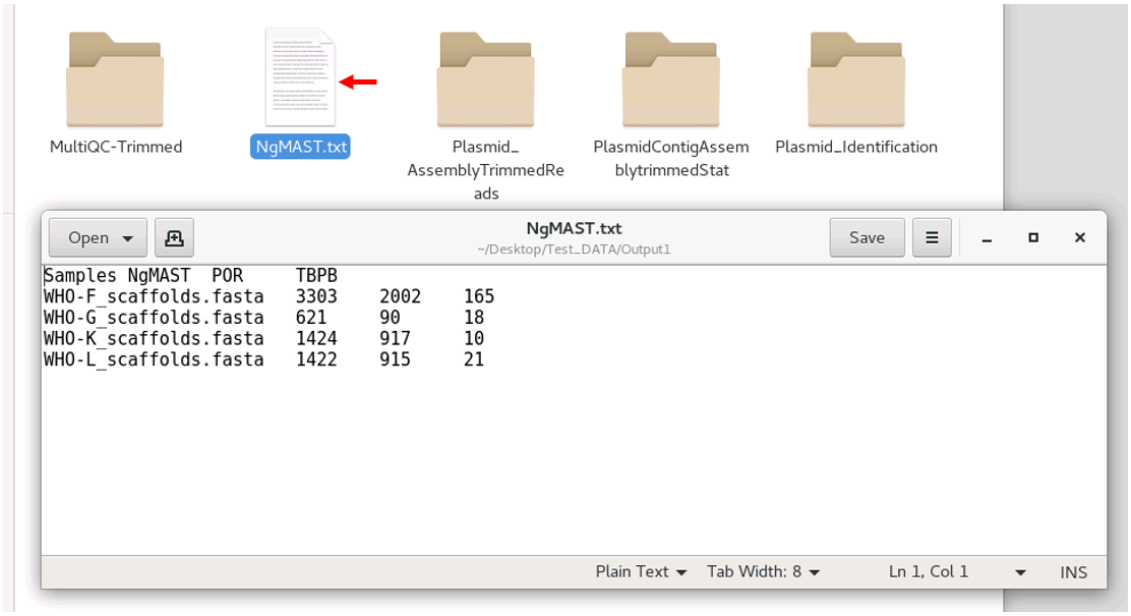


b. Perform the MAST typing by selecting MAST->MAST Typing

Strain typing

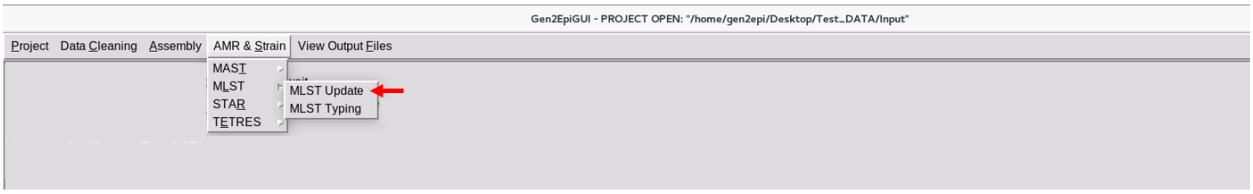


Note: MAST strain typing for each sample is present in file "NgMAST.txt" under "/home/gen2epi/Desktop/Test_DATA/Output1" as shown below.

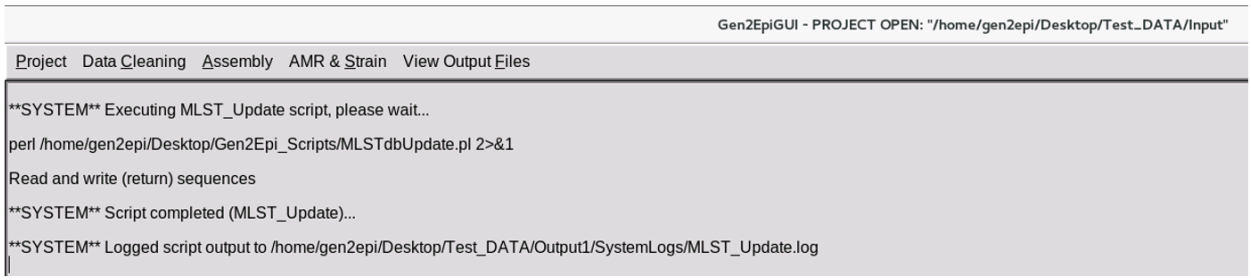


18) NgMLST Typing

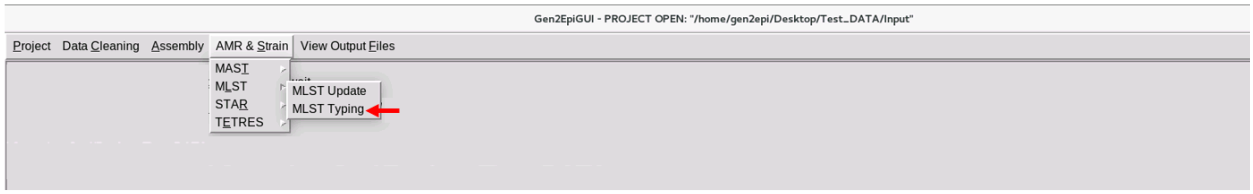
a. Update the underlying MLST database by clicking MLST->MLST Update tab



MLST database update

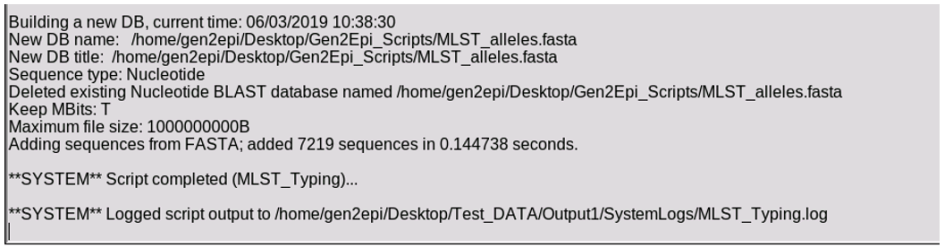


b. Perform the MLST typing by selecting MLST->MLST Typing

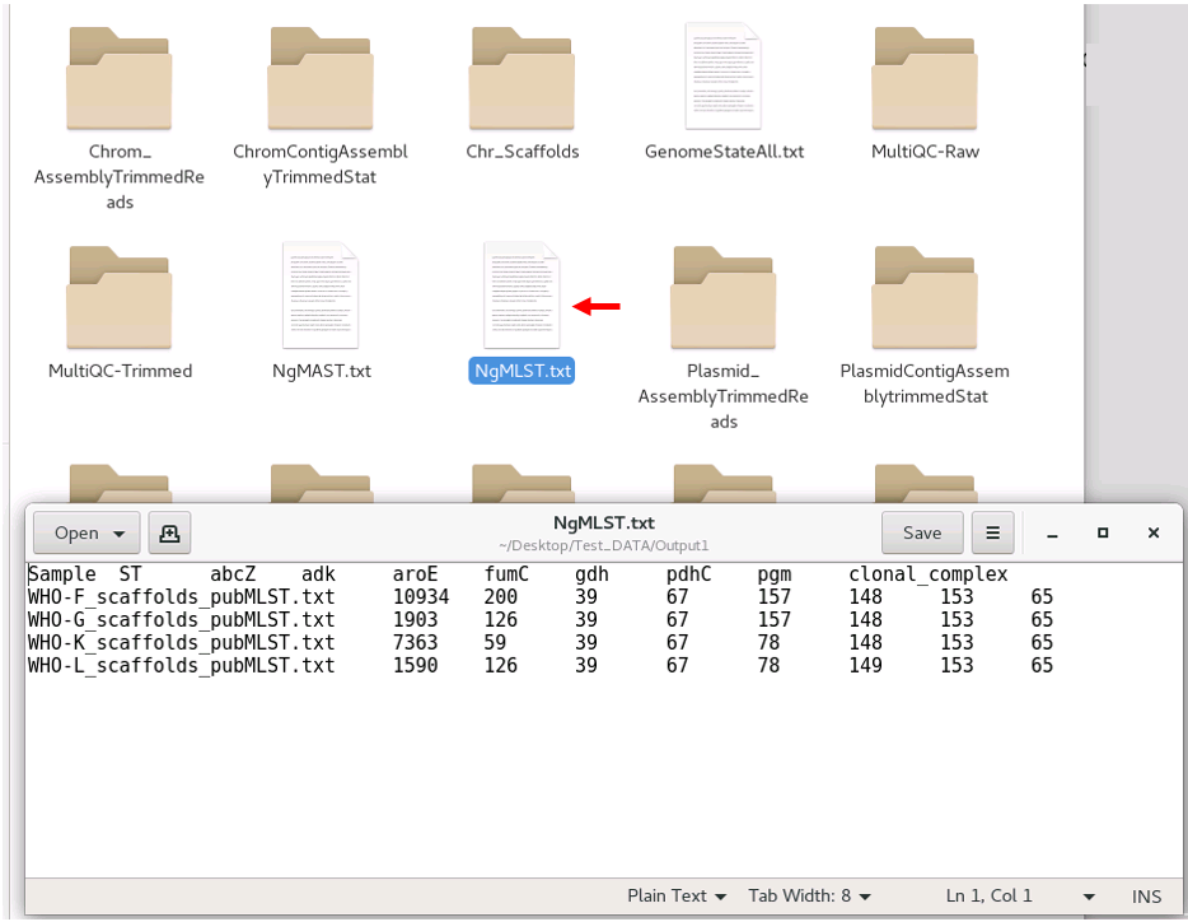


Start of the strain typing analysis

Completion of the strain typing analysis

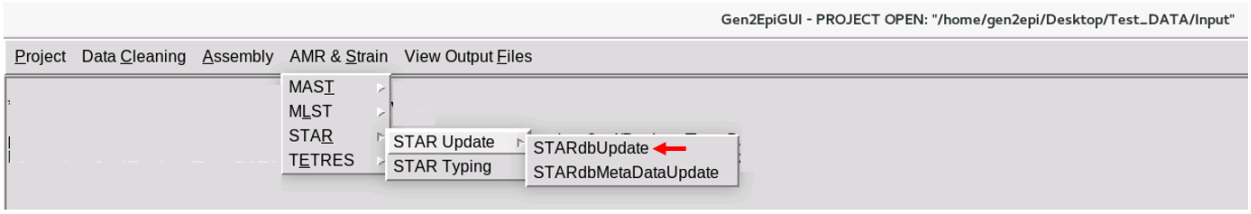


Note: MLST strain typing for each sample is present in file "NgMLST.txt" under "/home/gen2epi/Desktop/Test_DATA/Output1" as shown below.

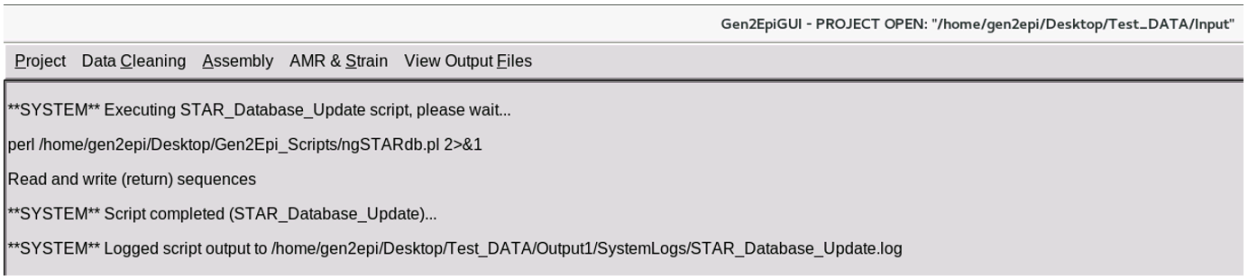


19) NgSTAR

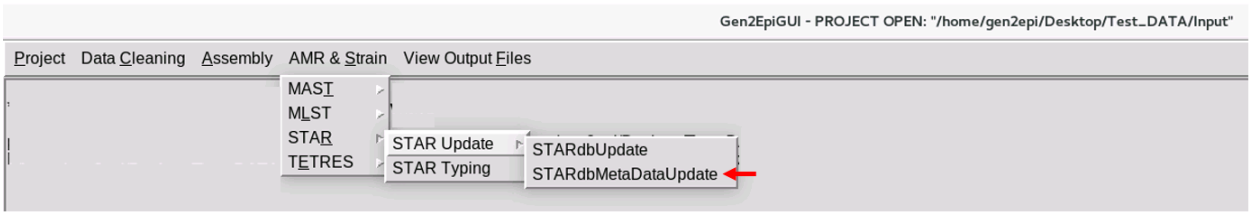
a. Update the underlying STAR database by clicking STAR->STAR Update->STARdbUpdate tab



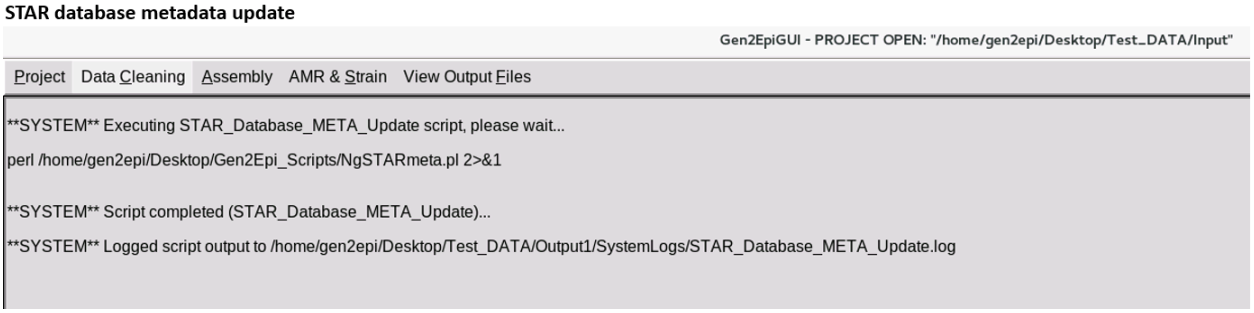
STAR database update



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

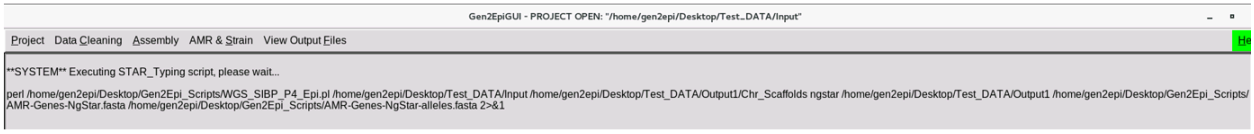


STAR database metadata update

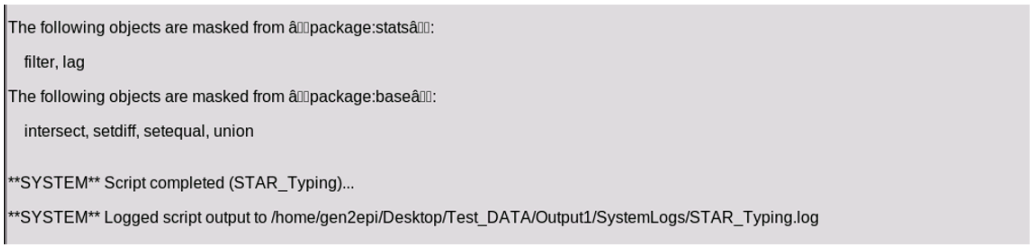


c. Perform the STAR typing by selecting STAR->STAR Typing

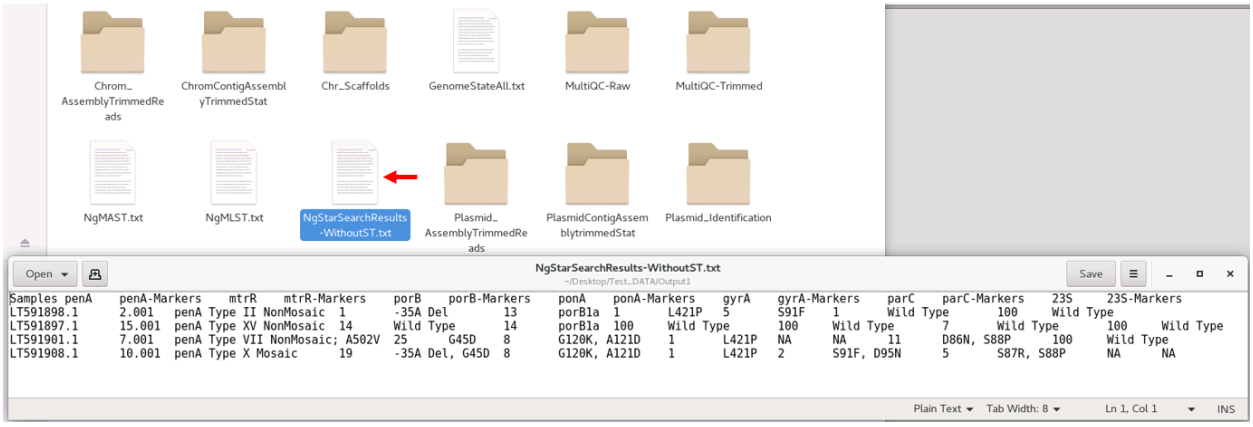
Start of the STAR typing process



Completion of the STAR typing process

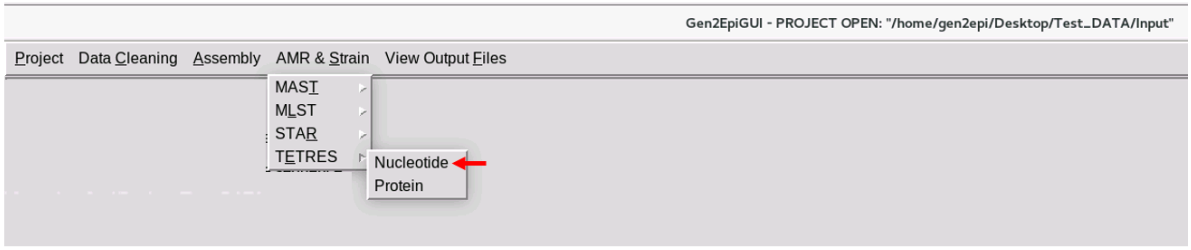


Note: NgSTAR output for each sample is present in file "NgStarSearchResults-WithoutST.txt" under "/home/gen2epi/Desktop/Test_DATA/Output1" as shown below.



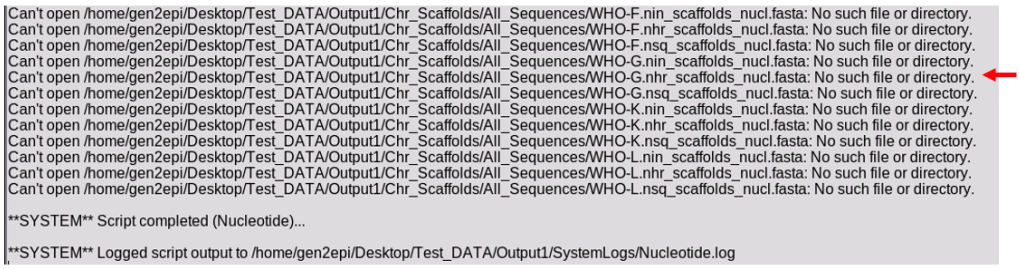
20) Tetracycline Resistance

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.



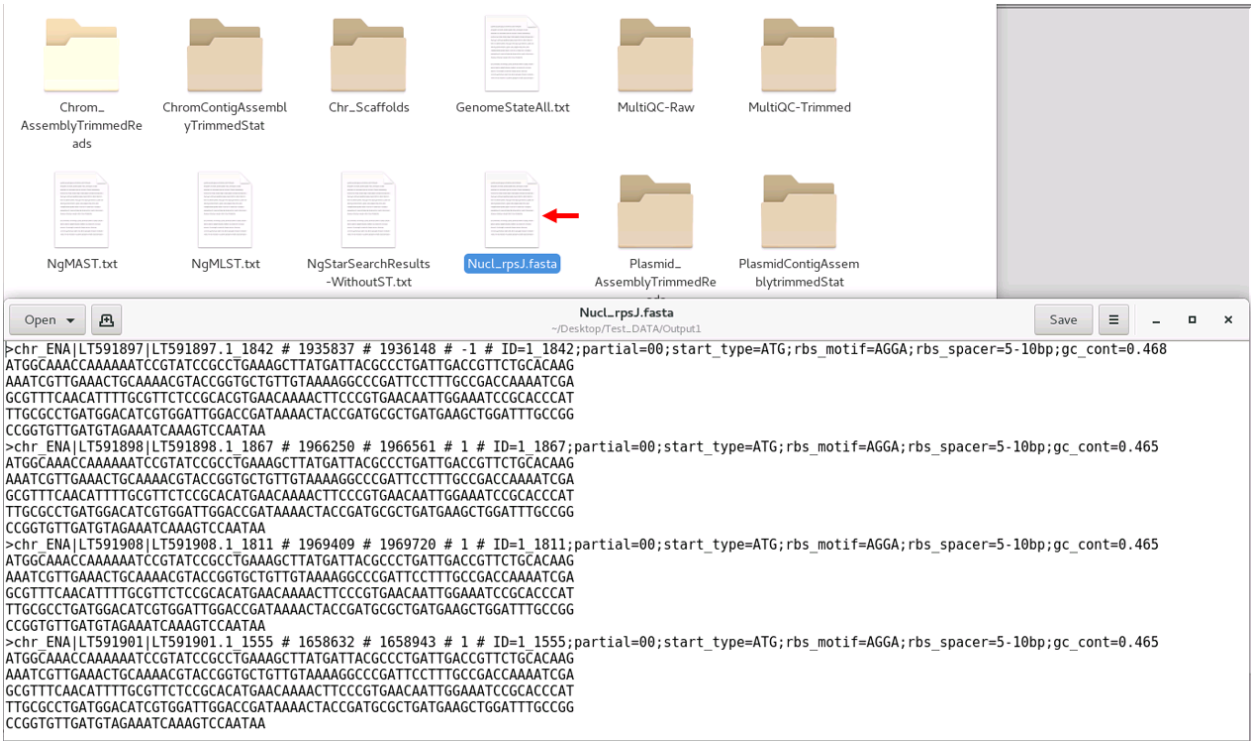
Starting

Completion

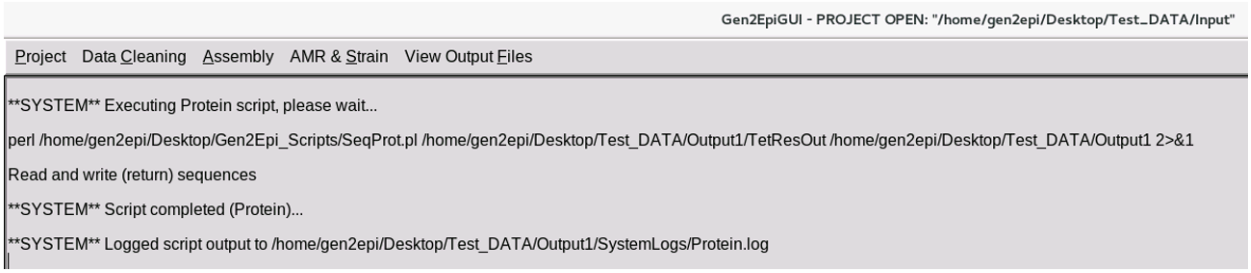


Don't worry about this message

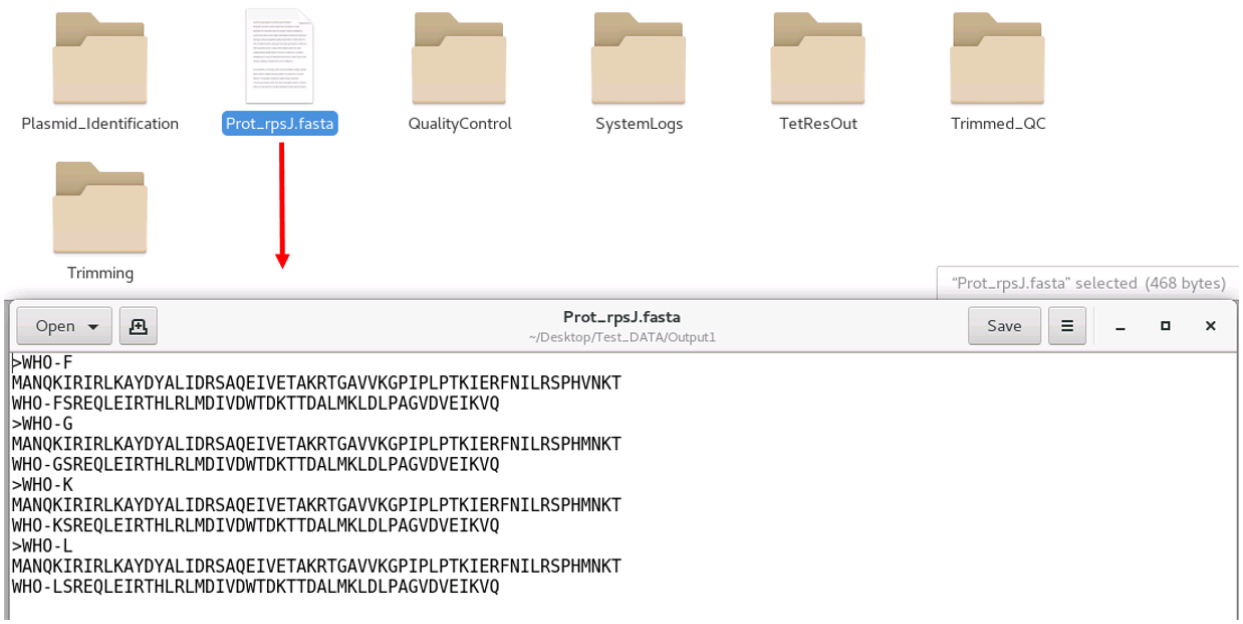
Note: Users can see the rpsJ nucleotide sequence for each sample in file "Nucl_rpsJ.fasta" under "/home/gen2epi/Desktop/Test_DATA/Output1" as shown below.



2) 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.