Frequently Asked Questions (FAQs)

1) Can I use the shared folder between host and VM as the output folder in Gen2Epi GUI?

   Answer: Yes, You can. Please follow instructions described in "**ReferenceGuide.pdf**" under Help menu in Gen2Epi GUI.

2) I have set a shared folder between VM and my host machine. However, I am not able to edit any file in this folder within VM image?

   Answer: Because VirtualBox VM image do not have writing permission for the shared folder. Please see the post reporting bug related to writing permissions in shared directory https://forums.virtualbox.org/viewtopic.php?f=7&t=10535

3) If you have pre-assembled genomes- how to further analyze them using Gen2Epi? (How to import pre-assembled genomes in Gen2Epi GUI)?

   Answer: Please use following steps:

   1. Project->Set Output Directory -> "Enter the name of the directory" and click on confirm.
   2. Project->Set Input Directories->Fastq and select the folder that has all your fastq files.
   3. Project->Prepare Input-> Enter the number e.g 5 (this number will set the length of your output directory name).
   4. Make sure to copy the "Chrom_AssemblyTrimmedReads" folder that contains *de novo* assemblies for each samples into the output directory set in step 1. If you are confuse about the "Chrom_AssemblyTrimmedReads" folder then you can run Gen2Epi on Test_DATA and see the arrangement of files in this folder.
   5. Project-> Open project -> select the "Prepare_Input.txt" file that you just prepared in step 3 and click on open.
   6. Go to Project-> Set Input Directories->Genome File and now select the folder that has genome files in it.
   7. Go to Project-> Set Input Directories->Annotation and select the folder that has annotation files in it.
   8. Now go to Assembly and click on Preassembled Genome tab. This step will start the scaffolding on your preassembled genomes
   9. Once completed, you can access the results in "Chr_Scaffolds_Sameref" under your output directory set in step 3.

4) How to increase the Gen2Epi Virtual Machine image hard disk size?

   Answer: Please follow step 6b described in "**ReferenceGuide.pdf**".

5) I have Whole Genome Sequencing (WGS) reads from bacteria other than *Neisseria gonorrhoeae*. Can I use Gen2Epi to assemble it?

   Answer: Yes, you can. However, you will only be able to use Gen2Epi until the *de novo* assembly step.

6) Is it necessary to perform the scaffolding?

   Answer: No, you do not have to perform scaffolding if you are interested only in the de novo assembly. However, AMR and strain typing functions currently using scaffolds.

7) I am getting error in one of the strain typing database update.

   Answer: We would suggest you to first complete the strain typing of your datasets using the existing databases and then redo this after updating the databases. This way you will still be able to complete your analysis even if the database update fail.

8) Why can't I create a "new folder" as an output folder while "browsing the path" during output directory setting?

   Answer: In order to create a new output folder in Gen2Epi simply follow these steps:
   1. Click on Project->Set Output Directory -> Browse the path where you want to create the output folder and click on open
   2. Type the name of the output file and click on confirm.

9) Defining the length of the filename?

   Answer: This is basically the number of characters that you would like to use in your output directory name. Fore example, let's say you have a fastq files with name "WHO-F_S2_R1.fastq.gz" and the name that you would like to assign to your output folder is WHO-F. WHO-F has 5 character in them so the length of the filename is 5. Instead of 5 if you will use 8 then the name of your output folder would be WHO-F_S2 and so on. Furthermore, this is the same number that you need to define during "Prepare Input" step in Gen2Epi.

10) Why I need to set up the input directories for "fastq, genome, and annotation"- I already set up input directory for "fastq" files in step 2b. Why do I need to do this again?

    Answer: In the current version of Gen2Epi, it is compulsory to set all the input files at the beginning of the analysis.

11) Quality check and Trimming: I did a quality check. For trimming, can I adjust the no. of input files? e.g., I did quality check in 4 samples and I need to perform trimming only in two of these samples then what are the steps for that? Can I customize it?

Answer: Yes, you can customize it. Please follow the below given steps:

1. Make a copy of your original "Input.txt or Prpeare_Input.txt" files so that you have a backup copy in case you lost it.
2. Now edit the "Input.txt" file by removing the entries that you do not want to analyze further.
3. Upload the file again in Gen2Epi via Project->Open Project and proceed further with the trimming step.

12) Execute java time error during running data cleaning on single sample?

Answer: If you are trying to check the quality of a single WGS read dataset then Gen2Epi will through out this error. This is because the "MultiQC" program used in the pipeline don not have enough file to merge. Nothing to worry about – the pipeline still complete the accurate analysis.

13) What is the minimum computation power with which I can perform each step in the GUI version of Gen2Epi?

Answer: The minimum requirement to perform each step in the GUI version of Gen2Epi is 2GB RAM, 2 Processor and 28 GB storage space.

14) I do not want to trim my raw fastq files, Can I generate the assembly using raw reads?

Answer: Yes you can. To perform d*e novo* assembly from raw reads, first follow the steps 3-4 given in "Gen2Epi-GUI-User-Manual" and then directly jump onto step 6.

15) What is the purpose of "Multiple Genomes" and "Single Genomes" tab under Scaffolding-> Chromosome submenu?

Answer: "Assembly->Scaffolding->Chromosome->Multiple Genomes" – Follow this step if you have multiple reference genomes for multiple fastq files (For example, data given in Gen2Epi Test_DATA).

"Assembly->Scaffolding->Chromosome->Single Genomes" – Follow this step if you have single reference genome for multiple fastq files.