1. Installation Guide

Installation through Installer: PanRV installer script is available (https://sourceforge.net/projects/panrv2./Installer.sh). In order to run Installer as an executable file, permissions should be set by changing the properties (right click on file -> properties -> permissions -> execute -> check "Allow executing file as program"). In order to run installer, enter command "./Installer.sh" in terminal. It will run all the commands and thus each component for PanRV will be downloaded and installed automatically. If at any step installer fails, open a new terminal and install that step separately, and then continue. Otherwise, close the terminal and pursue commands from the readme file.

Installation through Readme: PanRV Installation readme file (https://sourceforge.net/projects/panrv2/files/Readme/download) comprises of all commands in a step-wise manner. User may install PanRV by following each command.

Note: During installation, system may require an Ubuntu password and paths. In case of any difficulty during installation and path setting consult the Trouble Shooting section**Post Installation Steps:**

- Following Installation, navigate to PanRV/Configs directory and make changes in files (config.properties & log4j.properties) by replacing "USER" with your system's user name.
- Recheck all paths added in .Bashrc.
- Check permissions of following directories
    1. /home/USER/PanRV
    2. /usr/local/psort
    3. /gpsr

Availability and Implementation: PanRV, its installer and Readme files are freely accessible at https://sourceforge.net/projects/panrv2/ .

2. User Guide

PanRV provides an easy to use graphical user interface; each feature is presented with guidelines in this section. User can run PanRV pipeline by accessing PanRV directory and typing the following command in terminal:

java –jar PanRV.jar

In case of any error, please consult section Trouble Shooting.

2.1 Main Window

Once the command is entered, a main window will be displayed. Main window is divided into two panels: left panel provides a list of modules and right window to display the selected module. Above these panels, a *job name* field is provided. Setting up a job name is necessary because the results will be saved with the same name in form of a directory within the results folder. Re-entering the same name will overwrite the previous results. Hence, for every new analysis a unique job name must be provided (See Figure 1 for main window display and Figure 2 for setting a job name).

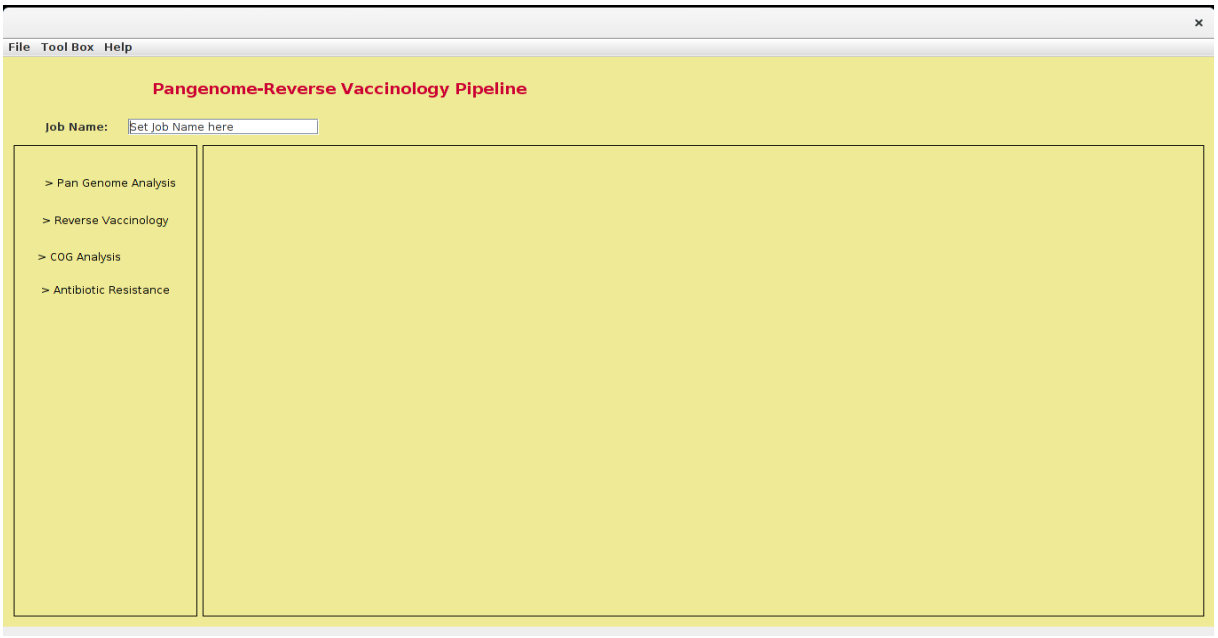

Figure 1: Main window of PanRV. It consists of a job name field to set a directory name in results, along with a list of modules on the left and a window to display the selected module on the right.

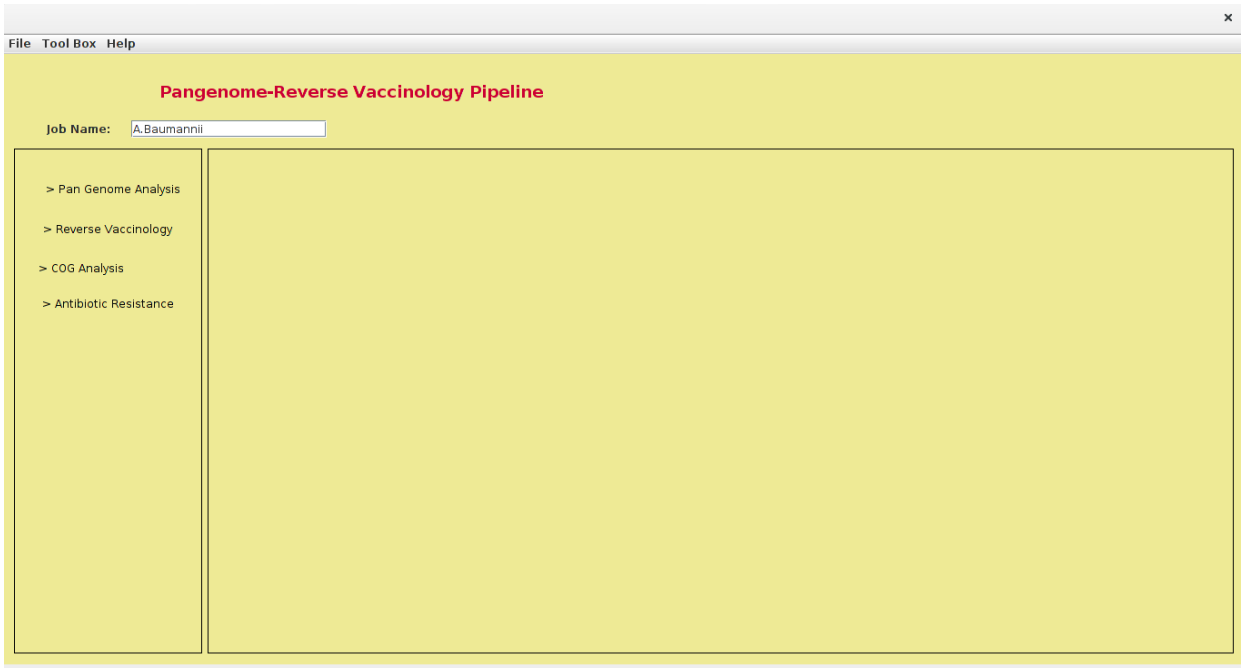

Figure 2: Setting up a Job name. Before starting any analysis, a unique job name must be provided as displayed above.

2.2 Pangenome Analysis Window

First module of PanRV is pangenome analysis, which can be selected by clicking the “Pan Genome Analysis” tab on the left panel list. It will highlight the tab in red and will display the module components in the window on the right, as shown in Figure 3.

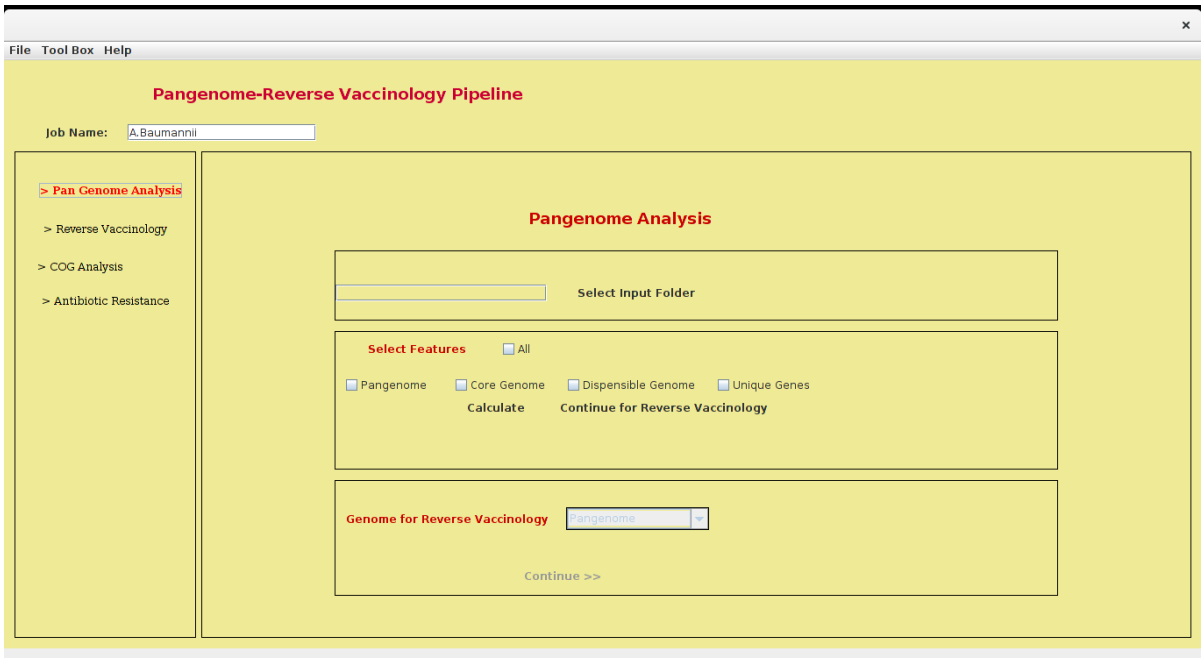

Figure 3: Pangenome analysis module: This module is displayed when “Pan Genome analysis” tab is selected from the left panel. User can select different features according to their study.

User can set input folder path by the “select input folder” button (Figure 4). Input of this module is a directory containing protein gff files of all the strains that require a pangenome analysis. It allows user to select the features of pangenome which are required in the analysis including pangenome, core genome, dispensable genes or unique genes (Figure 5). “Calculate” button is used to run pangenome analysis (Figure 6). It displays its count (Figure 7). “Continue for Reverse Vaccinology” button can be selected for target identification analysis. Further analysis will require an input file that can be obtained from the results of pangenome analysis, which can be selected from a dropdown list (including pangenome, core genome, dispensable or unique genes). Analysis can then be continued using the “Continue” button (Figure 8 and 9).

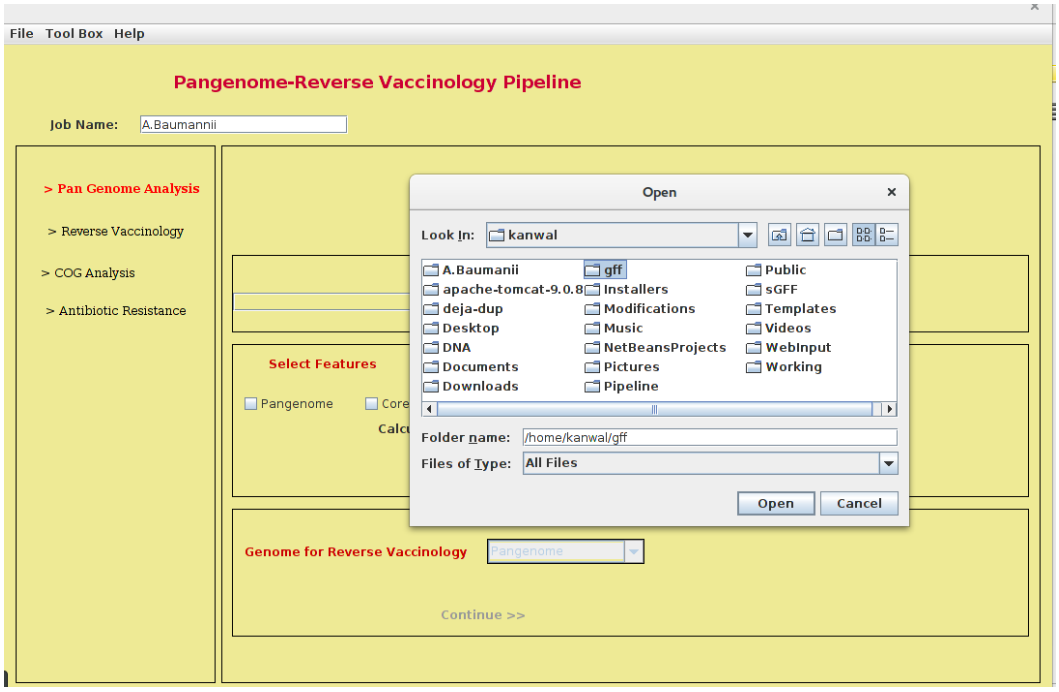

Figure 4: Folder selection: Complete sets of gff files (must obtained by Prokka1.2) of different microbial strains to analyze pangenome

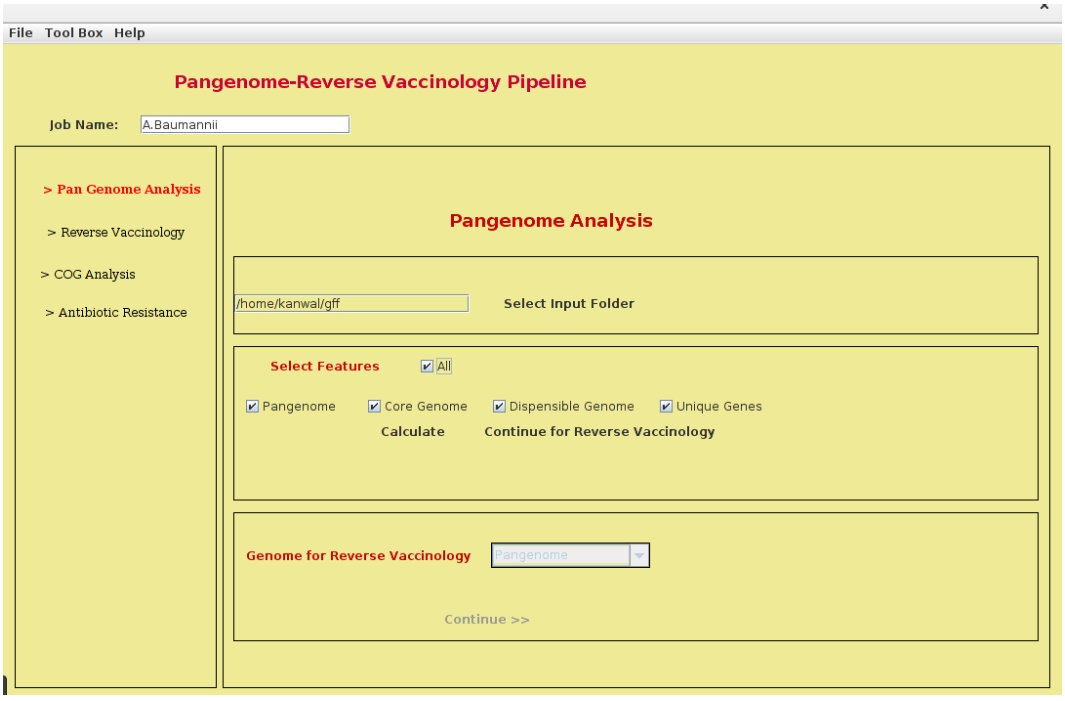

Figure 5: Feature selection: Users can select features which they deem fit for their analysis.

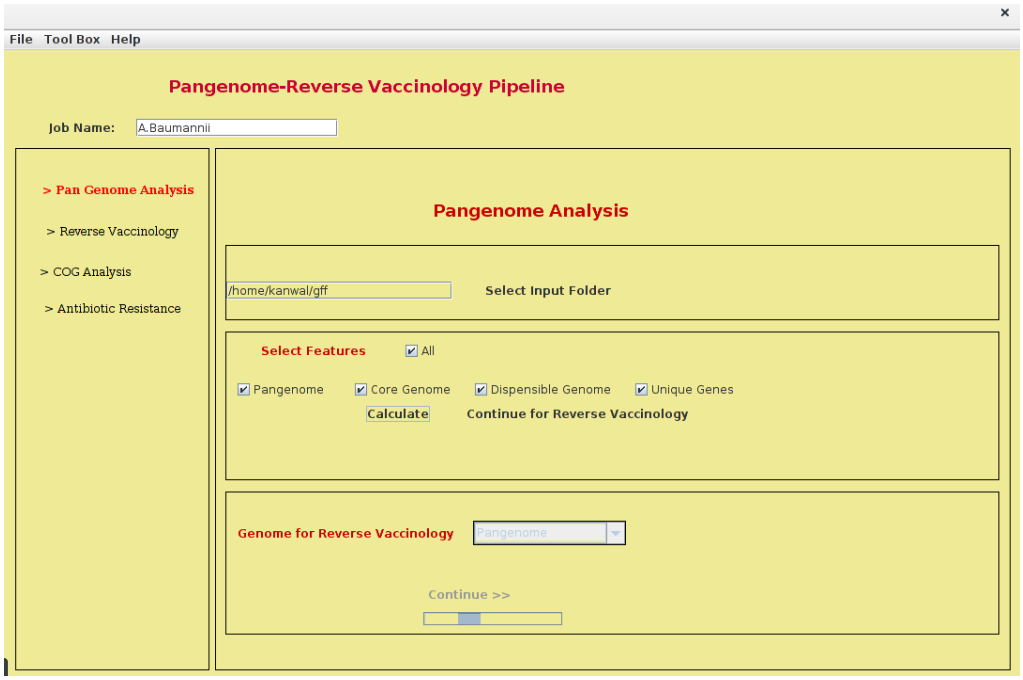

Figure 6: Pangenome analysis: Users can calculate only pangenome if they want to continue analysis for therapeutic target prediction.

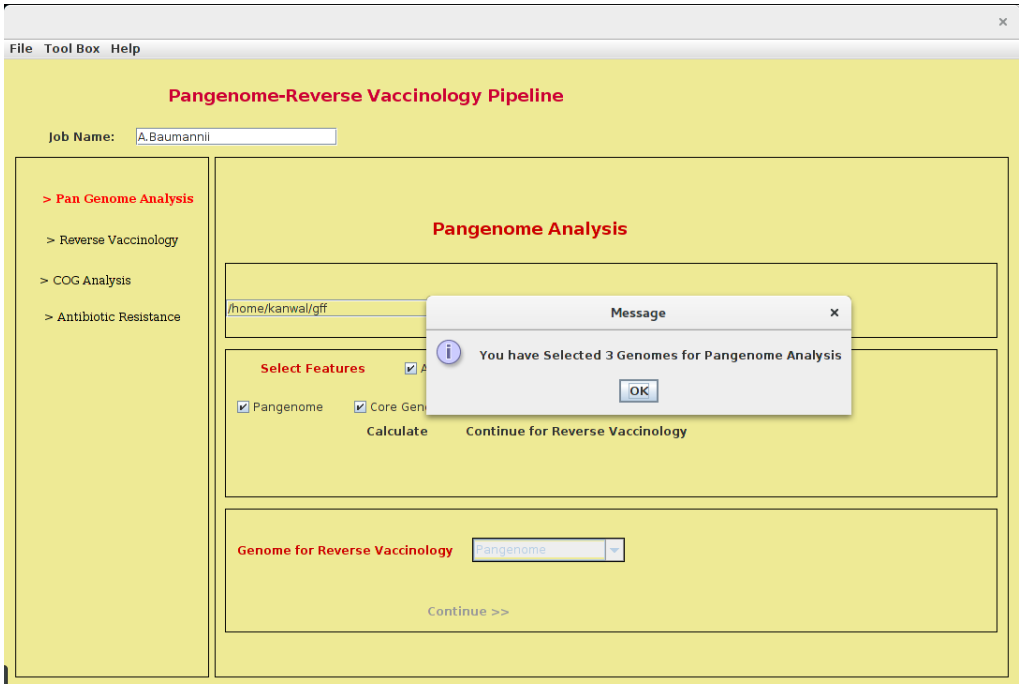

Figure 7: Before starting Pangenome analysis, system will show number of files.

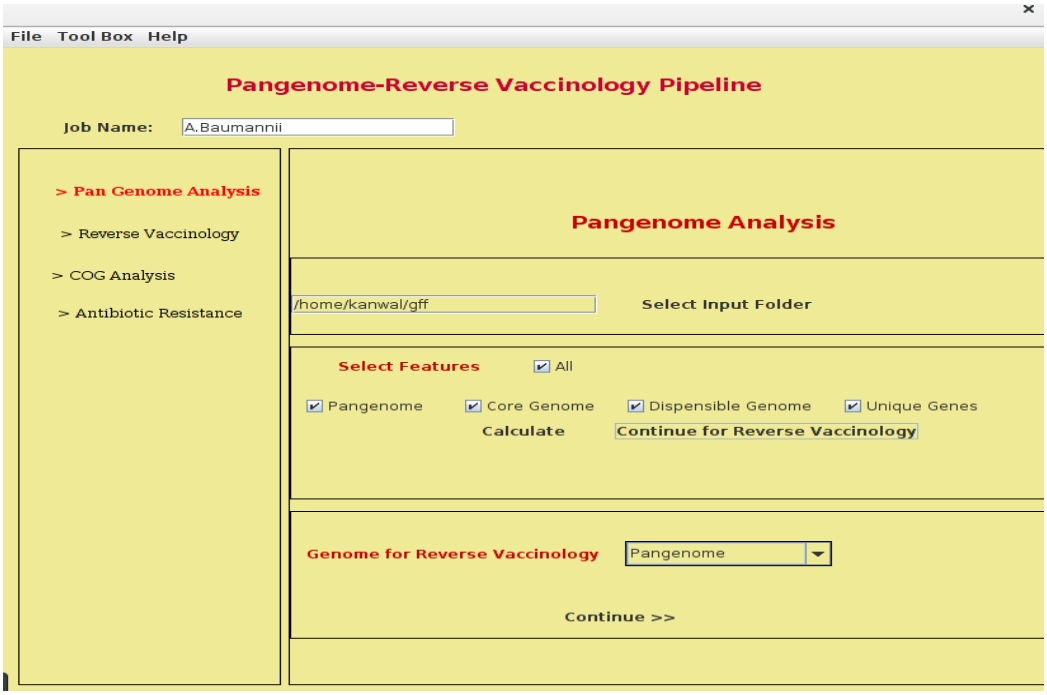

Figure 6: Continuation of pangenome analysis for target identification: This will continue with the output of pangenome analysis for therapeutic target prediction analysis

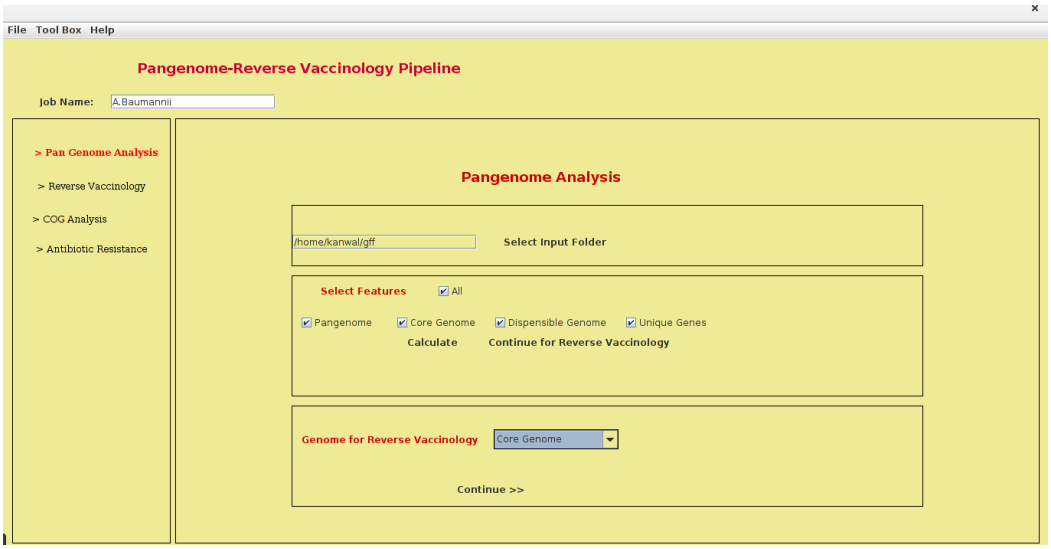

Figure 7: Selection of resultant FASTA file from pangenome analysis for vaccine target identification using reverse vaccinology module

2.3 Reverse Vaccinology Window

This module works alone as well as in continuation with pangenome analysis by employing the output of pangenome as an input for this module. It contains various filters and components to analyze input protein sequences separately. The filtered sequences are then selected as vaccine candidates.

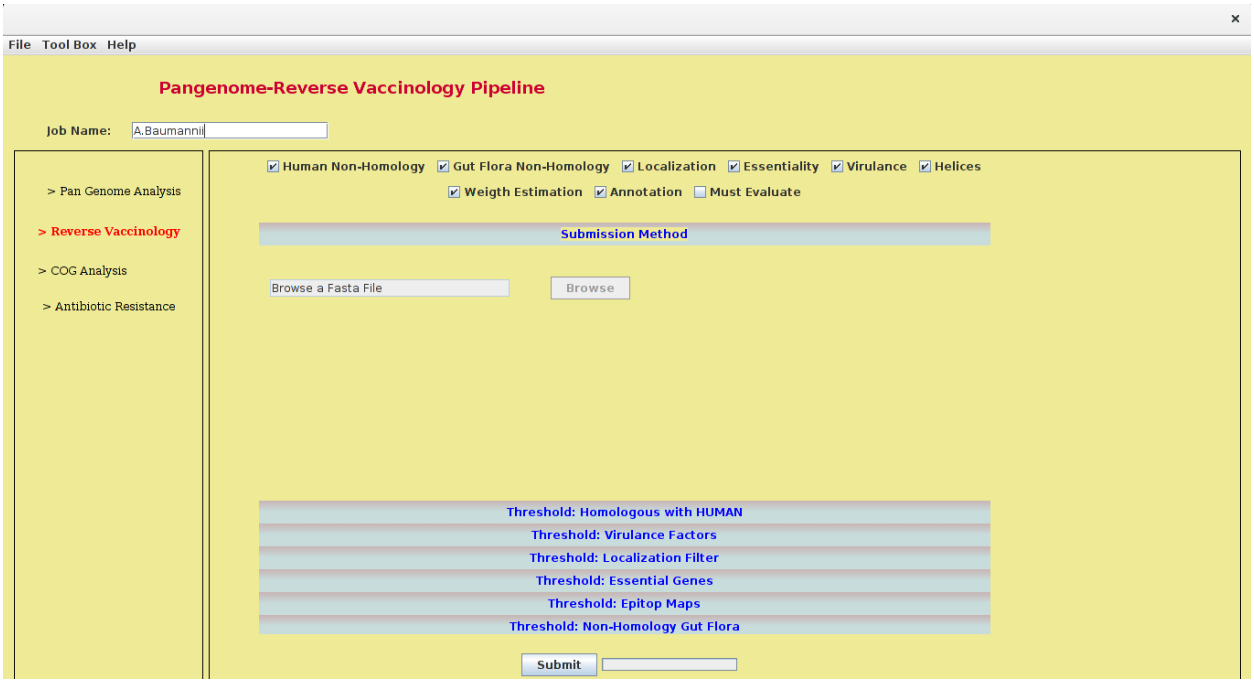

Figure 8: Reverse Vaccinology module: Contains additional various components; Users can add these components into analysis.

The components in the top panel of reverse vaccinology window are shown in the form of checkboxes (Figure: 10). User can select these components as desired.

Human Non-Homology: Selecting this check box will add the feature to the analysis. The threshold values can also be set under "Threshold: Homologous with HUMAN" tab. It is selected by default.

Gut Flora Non-Homology: Select this check box to add this feature to the analysis, its threshold values can also be set under "Threshold: Non-Homology Gut Flora" tab. It is selected by default.

Localization: Selection of this checkbox will add protein sub-cellular localization feature into the analysis, which is selected by default. The threshold values can be set under "Threshold: Localization" tab. By default, the selected locations include extracellular, periplasmic, outermembrane and the unknowns. Unknown locations are the ones which can't be predicted by

this filter. This feature has been included so that the unknown locations can be manually analyzed by other means. User can select locations by choice. Organism type and gram stain values must be set using these options.

Essentiality: Selecting this feature will add it to the analysis. The threshold values can be set under "Threshold: Essential Genes" tab, which are already selected by default.

Virulence: Selection of this check box will add virulence factors identification for analysis, the threshold values can be set under "Threshold: Virulence Factors" tab. This tab is selected by default. User can also select databases i.e. MvirDB, VFDB or both.

Helices: To calculate transmembrane helices, select this checkbox. It will add the proteins in final result that have less than one helix. It is selected by default.

Weight Estimation: Select this check box to add molecular weight estimation feature in analysis. This feature is selected by default.

Annotation: Selection of this check box indicates that the annotation of protein sequence with UniProt data base is required.

Selection of all these features is necessary for accurate analysis of sequences for vaccine target identification.

Must Evaluate:

By selecting this checkbox sequences will be analyzed using each filter. It will give complete analysis of all sequences. Unselecting this box will stop further analysis of protein in case it fails to pass a filter. Only the proteins that sequentially pass the previous filter will be promoted to the next filter. The only difference between selection and non-selection is that former is slower as compared to latter, because it gives complete results of each one filter.

After the selection of these criterion, click on the "Submit" button to start analysis. If only reverse vaccinology is selected it will start analysis (Figure 10) and if it is continuation of pangenome, then pangenome analysis will be followed by reverse vaccinology steps. The pangenome module starts at first (Figure 11) and when it ends, (Figure 12) reverse vaccinology module will start processing. A progress bar indicates progress (Figure 13). When both processes end, the"Submit"button is changed to "Result" button (Figure 14). The result button depicts that both processes have been completed.

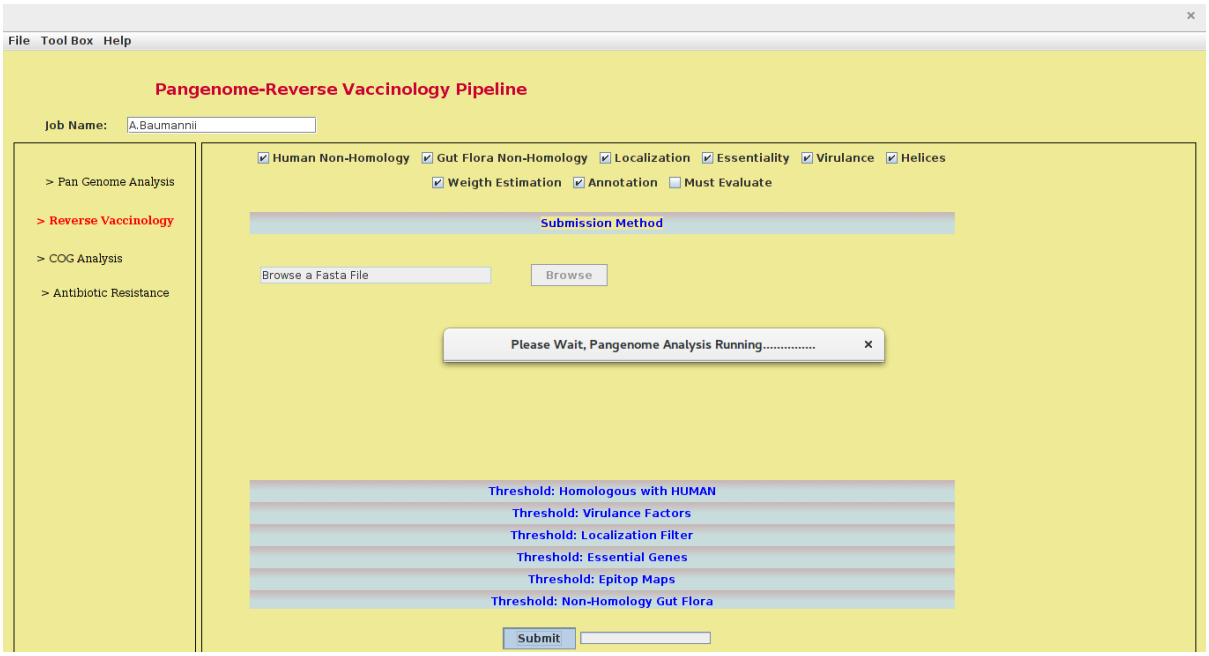

Figure 9: Target identification followed by Pangenome analysis

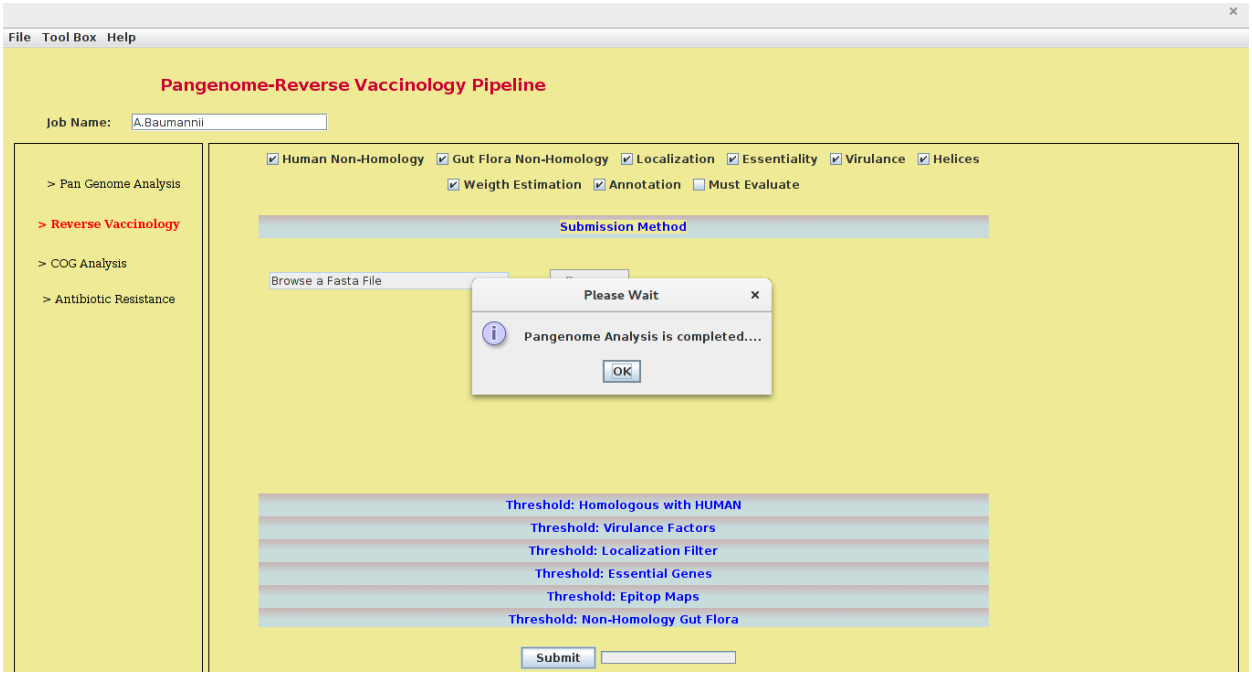

Figure 10: Pangenome analysis completion

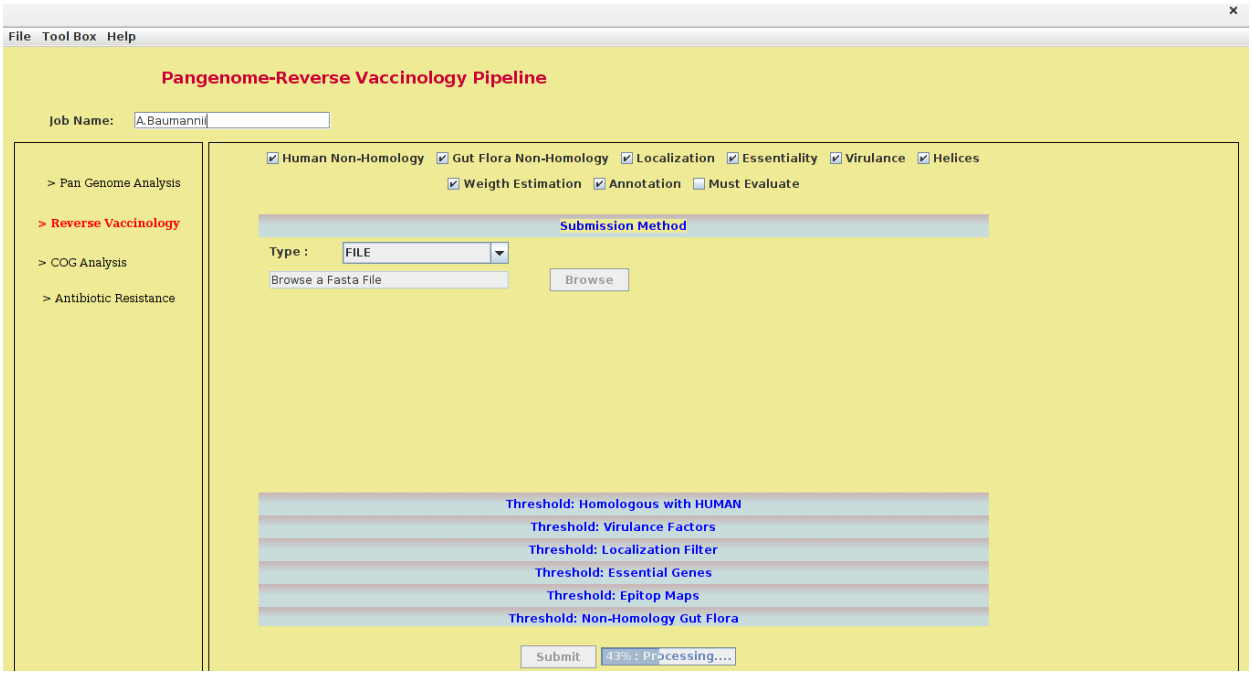

Figure 11:Reverse Vaccinology Analysisprocessing (progress bar indication)

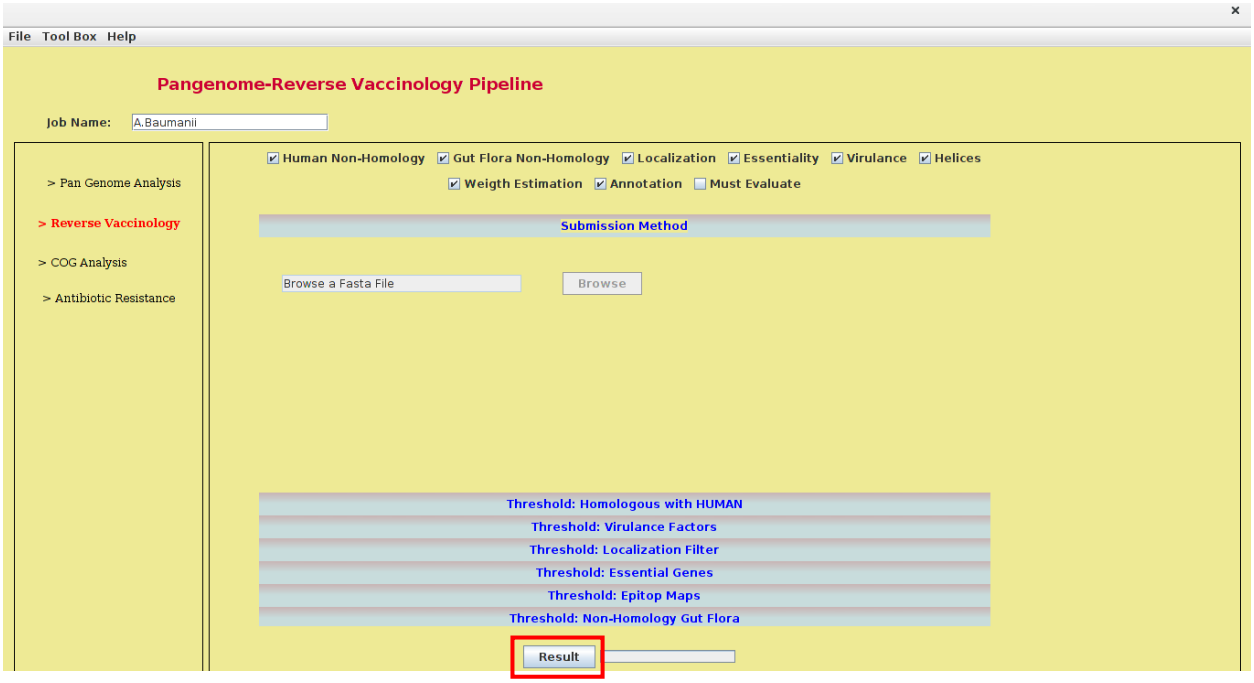

Figure 12: Reverse Vaccinology analysis completion

2.4 COG Analysis

COG analysis can be done by selecting its tab in left modules list (Figure 15). Upload protein FASTA file using “Upload file” button. By default, threshold values are set; user can change these values and click on “submit” button tostart the analysis. An ongoing process is shown in the form of progress bar. When analysis is completed progress bar stops and disappears. Results will be saved in COG directory in Results.

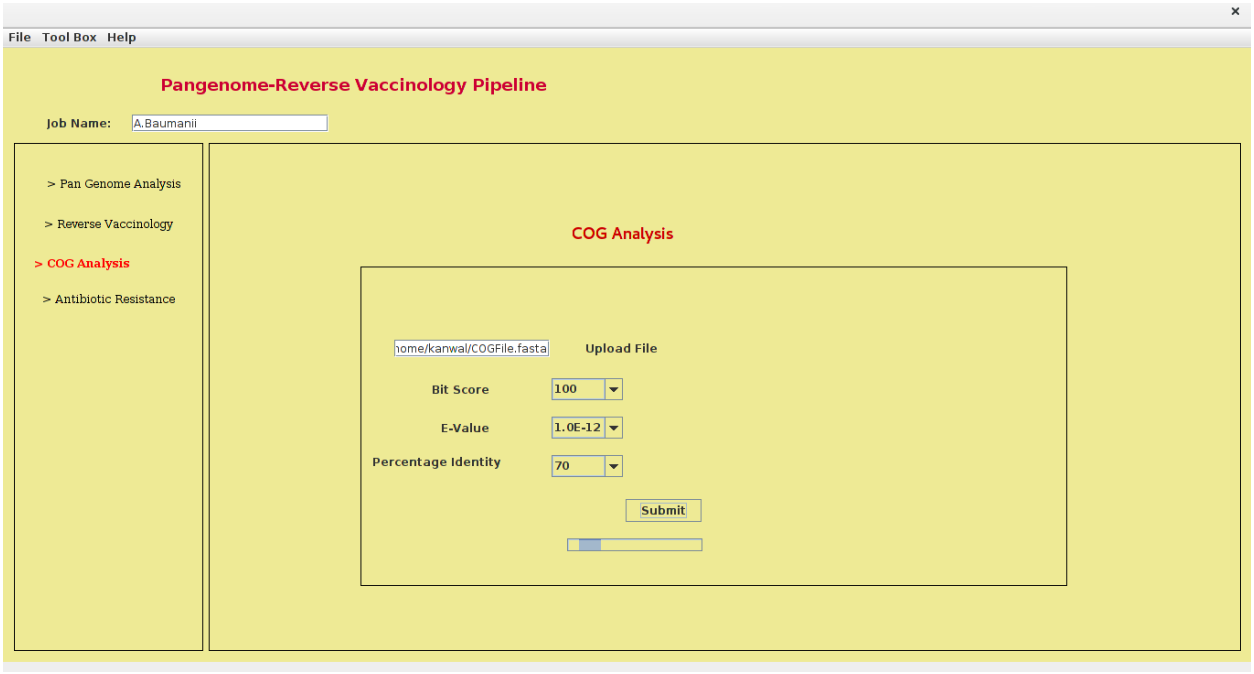

Figure 13: COG analysismodule

2.5 Antibiotic Resistance Analysis

Antibiotic resistance analysis can be done by selecting its tab in left modules list (Figure 16). Upload protein FASTA file using “Upload file” button. By default, threshold values are set; user can change these values and click on the “submit” button in order to start the analysis. The running process is shown in the form of a progress bar. When analysis is completed progress bar will stop and disappear. Results will be saved in CARD directory in results.

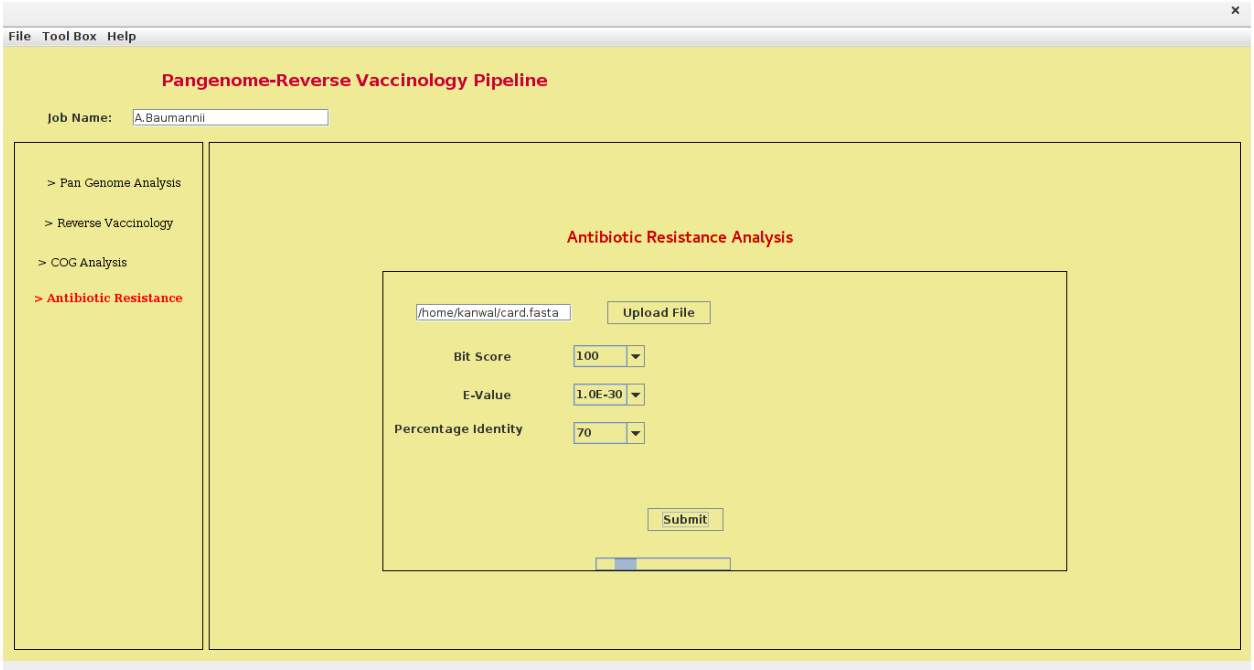

Figure 14: Antibiotic Resistance Analysis

3. Results

Final results of each module will be saved in Results within a directory labeled by your job name with their respective names including Pangenome, Vaccine Targets, COG and CARD (Figure 17).

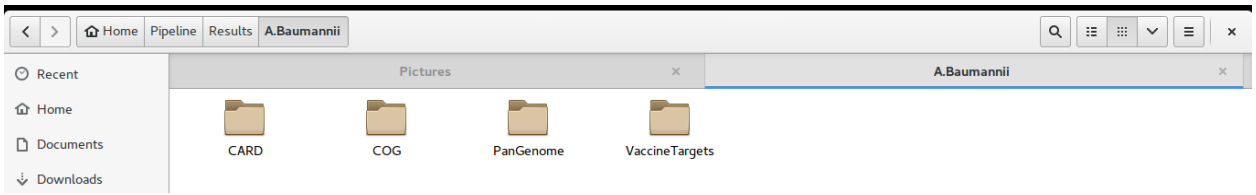

Figure 15: Final results saved in results directory

3.1 Pangenome Result Directory:

Pangenome directory as shown in figure 17 is shown in detail in Figure 18 where it displaysall input FASTA files and Result Directory. Result directory contains final results and intermediate files. User can find output file in respective folders at the end of list of files (Core genes, dispensable genes, unique genes and pangenome) shown in Figure 19.

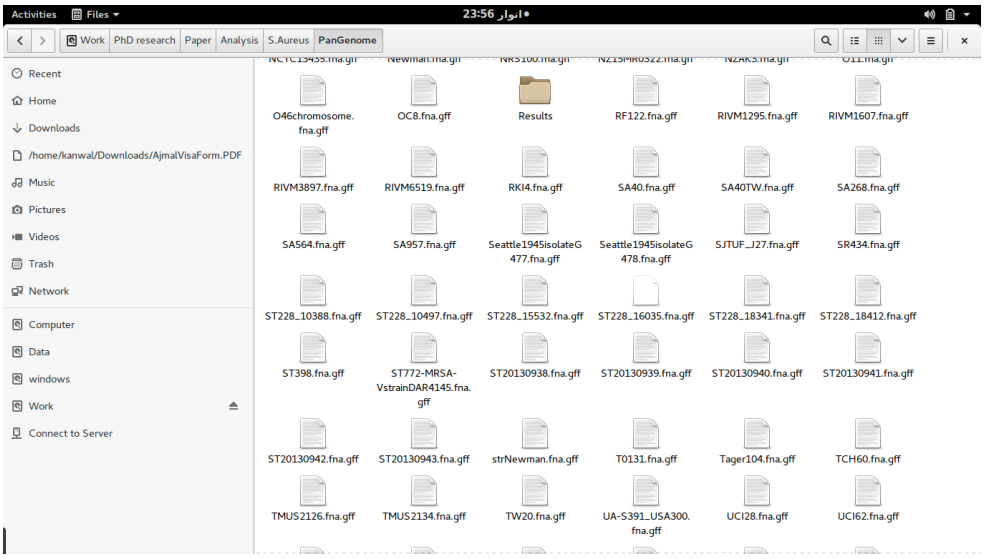

Figure 16: Pangenome Folder: Contain all gff files and result folder after analysis completion

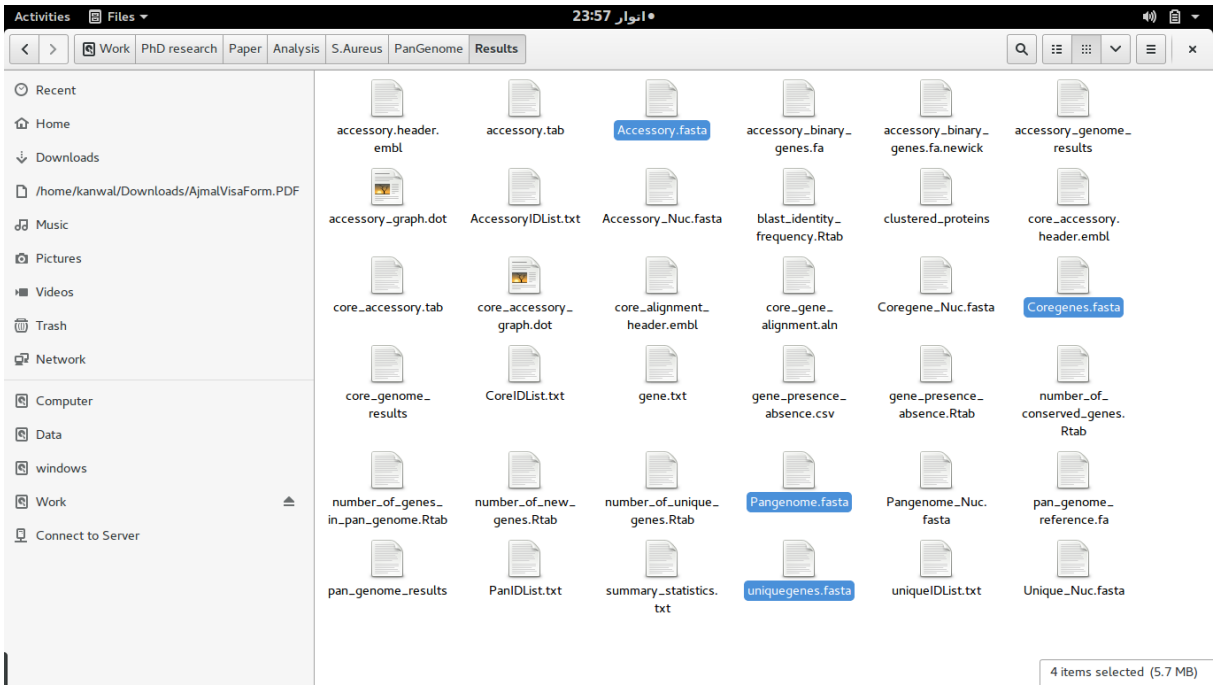

Figure 17: Pangenome Results folder. Contains various files including categories: pan,core, dispensable and uniquegenes in FASTA file and other Roary generated files.

3.2 Reverse Vaccinology Results:

Reverse vaccinology results are saved in the “Vaccine Target” directory (Figure 20). Results are shown in various formats (Json, xml, html, pdf). Results of each filter are also shown in separate files.

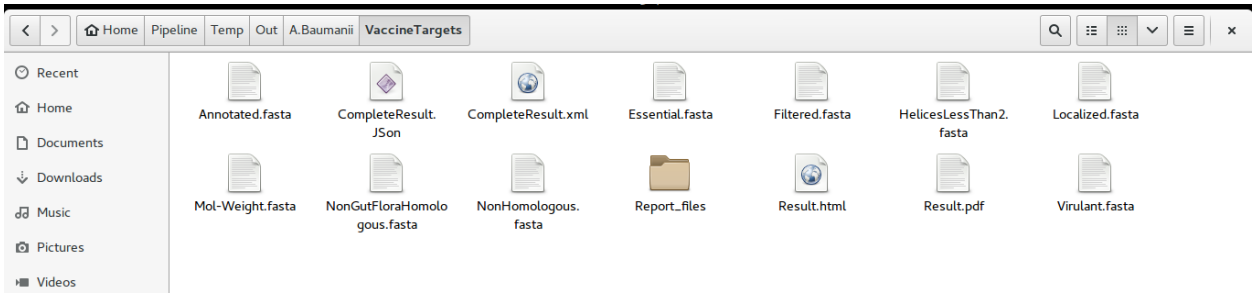

Figure 18: Vaccine Target folder. Contains results of reverse vaccinology. Resultsare shown in different formats and results of each filter are also shown in separate files.

3.3 COG Result:

COG Results are saved in “COG” directory (Figure 21). This directory contains an input.fasta file (input file with tagged protein ID). Tagged protein ID is used to map results in COGResults.csv. It is desirableto open COGResult.csv in gedit in Ubuntu and notepad++ in windows to avoid format differences.

COGResult.csv displaysresults in various columns (Figure 22). The file contains following columns:

1. Tagged ID in input.fasta, 2. NCBI protein GI matched with input sequences, 3. COG ID, 4. COG symbol, 5. COG description, 6. matched genome name from NCBI, 7. Protein domain ID matched with input sequence

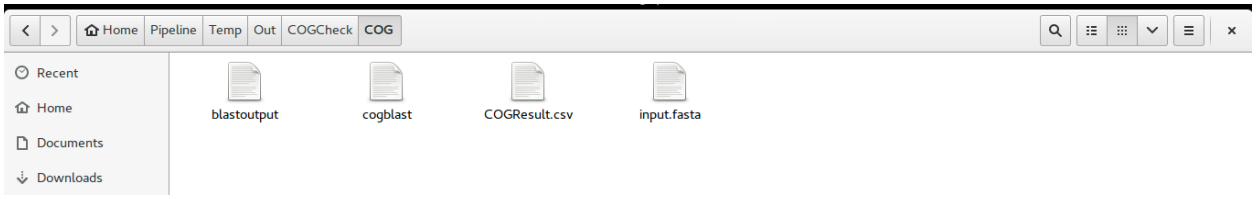

Figure 19: COG analysis results folder: this folder contains an input.fasta and COGResult.csv. Input file is tagged with a No. which is used in final results.

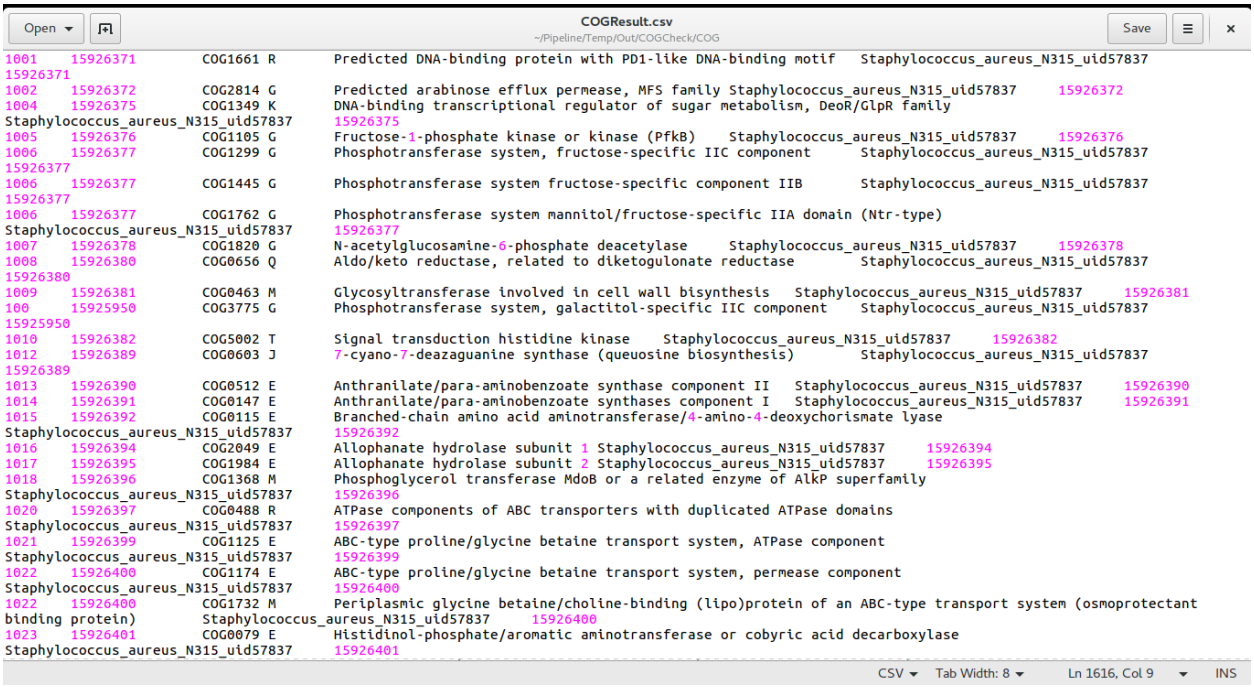

Figure 20: COGResult.csv. File contains columns 1. Tagged ID in input.fasta; 2. NCBI protein GI matched with input sequences; 3. COG ID; 4. COG symbol; 5. COG description; 6. matched genome name from NCBI; 7. Protein domain ID matched with input sequence.

3.4 Antibiotic Resistance Analysis Results:

Antibiotic resistance analysis results are saved in "CARD" directory (Figure 23). This directory contains input.fasta (input file with tagged protein ID). Tagged protein ID is used to map in CARDResult.csv. It is desirable to open CARD Result.csv in gedit in ubuntu and notepad++ in windows to avoid format differences.

COGResult.csv contains results in various columns (Figure 24). File contains columns 1. Tagged ID in input.fasta 2. NCBI protein GI matched with input sequences 3. ARO (Antibiotic Resistance Ontology) ID 4. ARO symbol 5. ARO description

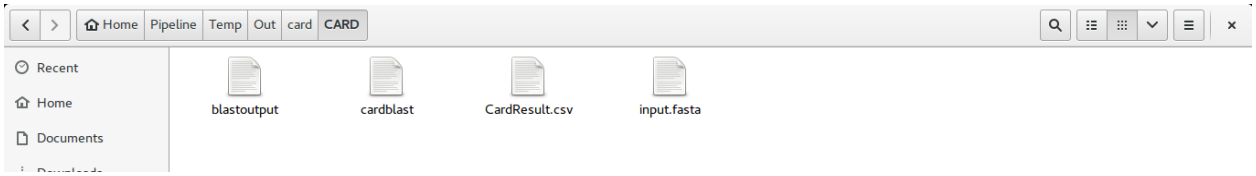

Figure 21: Antibiotic Resistance analysis results folder: This folder consists of aninput.fasta and CARDResult.csv. Input file is tagged with a No. which is used in final results.

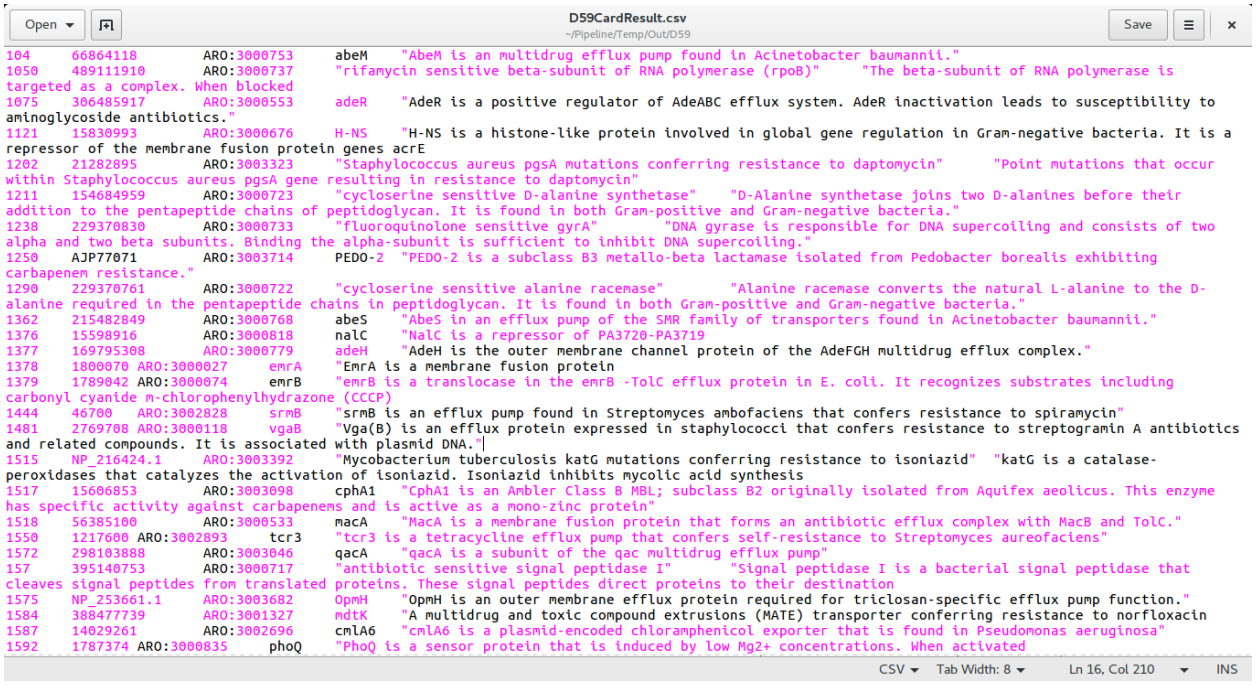

Figure 22: CARDResult.csv. File contains the following columns: 1. Tagged ID in input.fasta; 2. NCBI protein GI matched with input sequences; 3. ARO ID; 4. ARO symbol; 5. ARO description.

4. Trouble shooting

4.1 Blastall path

System will display a path where it has found blastall (Figure: 25). User will need to checkif the entered path is correct. Otherwise, find the path by entering the following command in another terminal:*"whichblastall"*. Enter correct path here.

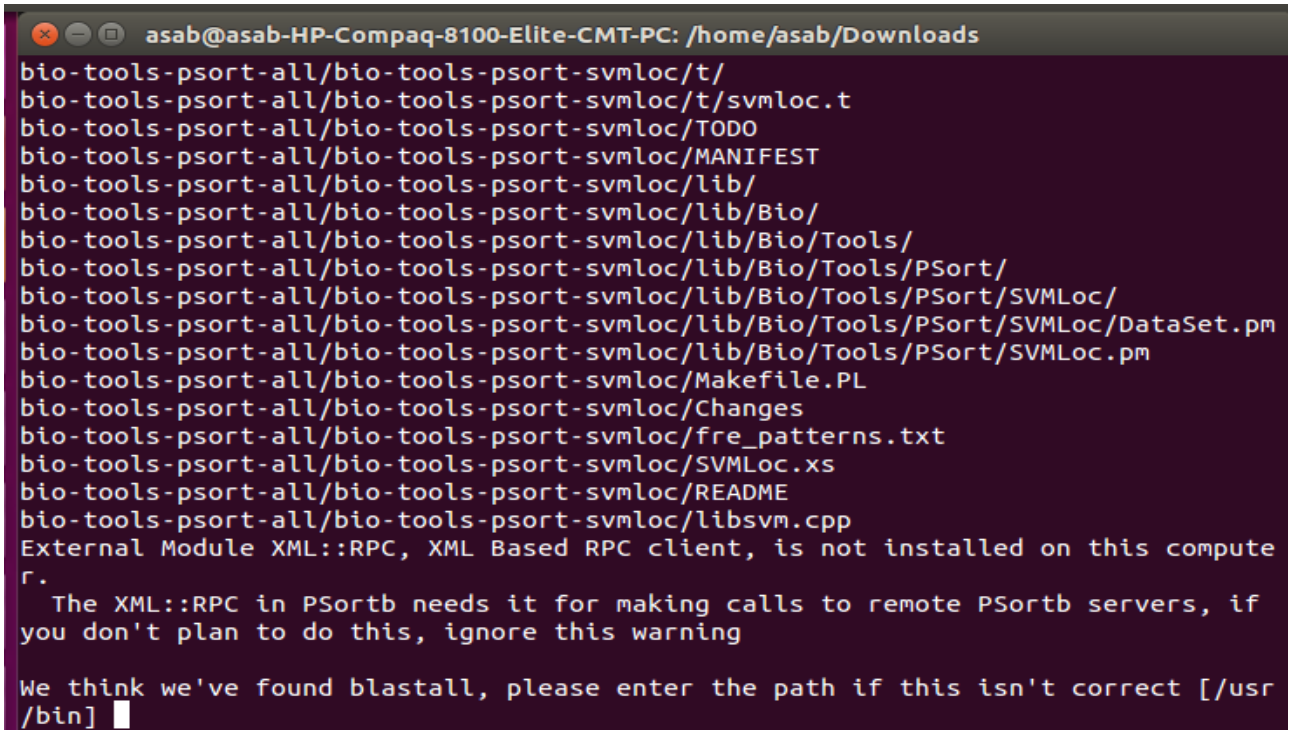

Figure 23: During installation of PSortb, it will require blastall path. First check if the system shows the correct path, otherwise type correct blastall path

4.2 Pfscan not found

We couldn't find pfscan issue can be resolved by entering correct path where pftools are saved. Mostly it is found in /usr/local/bin. First check its availability as shown in figure 27 then enter that path.

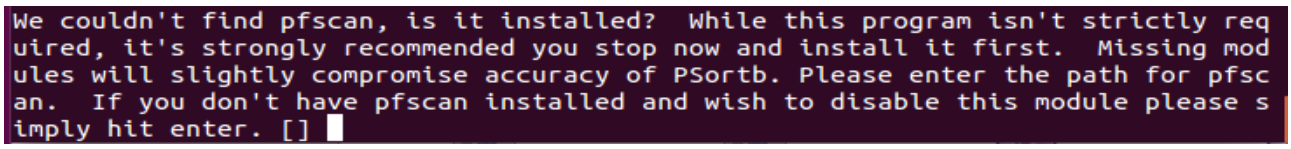

Figure 24: Pfsan path not found: give the correct path to where pftools are saved for correct installation. Mostly it is found in /usr/local/bin. First check its availability as shown in the next figure and then enter its path.

Figure 25: Checking availability of pftools and its path

4.3 lhmmer library not found

"No library found for –lhmmer" warning (Figure 28) can be addressed by finding this library in /usr/local/lib or /usr/local/lib64 if found there enter that path here. If system finds the library there, it can display a warningof not finding thedynamic linker path (Figure 29),continue by pressing enter. This warning will be shown for each library as displayed in figure 30. Do the same for all of them.

```
Warning (mostly harmless): No library found for -lhmmer

hmmer was not found in the dynamic linker path, is there somewhere else we shoul
d check?  [/usr/local/lib]
```

Figure 26:Library paths not found: usually these libraries are installed in /usr/local/lib64

```
We found hmmer in /usr/local/lib64, that's the good news.
The bad news is hmmer is likely not in your dynamic linker path, you must do one
 of two things:
- have your system administrator add /usr/local/lib64 to '/etc/ld.so.conf' (RECO
MMENDED!)
- add /usr/local/lib64 to the 'LD_LIBRARY_PATH' environment variable before each
 execution

Please consult your system administrator or email us for further assistance.

Do you want to continue anyways? [Y/n] [Y]
```

Figure 27: Not found in dynamic linker path. If found in path you entered it is not necessary to add it in dynamic linker path. Youwill only need to continue by pressing enter.

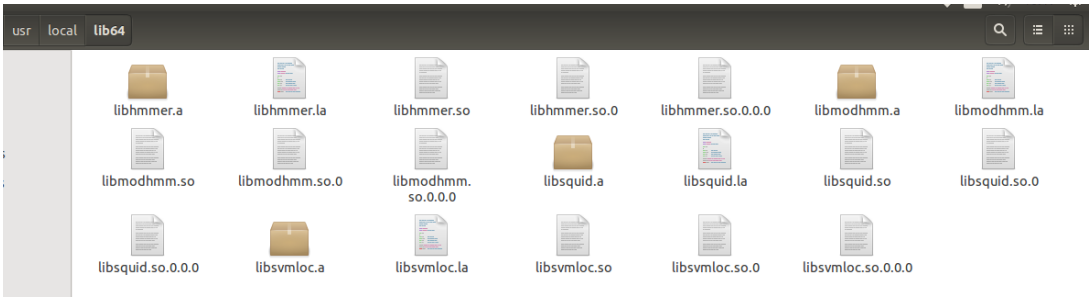

Figure 28: All libraries required for PSortb installation

4.4 PSortb installation Path:

System will ask where you want to install PSORTb. By default, it is found in */usr/local/psortb.* Enter the path to install psortb.(Figure 31)

```
Where do you plan to install the psort configuration files? [/usr/local/psortb]
```

Figure 29: Where to install PSORTb. Enter by default path that is /usr/local/psortb

4.5 JAVA incompatibility while running PanRV

After all installations run java –jar PanRV.jar by going to PanRV folder where jar is located. If an error like “unsupported major.minor version” occurs (Figure 32), it is due to more than one JAVA version installation. This issue can be resolved by uninstalling the older version throughsoftware center (Figure 33).

```
asab@asab-HP-Compaq-8100-Elite-CMT-PC:/home/asab$ cd PanRV/
asab@asab-HP-Compaq-8100-Elite-CMT-PC:/home/asab/PanRV$ java -jar PanRV.jar
Exception in thread "main" java.lang.UnsupportedClassVersionError: Presenter/Main : Unsupported major.minor version 52.0
        at java.lang.ClassLoader.defineClass1(Native Method)
        at java.lang.ClassLoader.defineClass(ClassLoader.java:803)
        at java.security.SecureClassLoader.defineClass(SecureClassLoader.java:142)
        at java.net.URLClassLoader.defineClass(URLClassLoader.java:449)
        at java.net.URLClassLoader.access$100(URLClassLoader.java:71)
        at java.net.URLClassLoader$1.run(URLClassLoader.java:361)
        at java.net.URLClassLoader$1.run(URLClassLoader.java:355)
        at java.security.AccessController.doPrivileged(Native Method)
        at java.net.URLClassLoader.findClass(URLClassLoader.java:354)
        at java.lang.ClassLoader.loadClass(ClassLoader.java:425)
        at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:308)
        at java.lang.ClassLoader.loadClass(ClassLoader.java:358)
        at sun.launcher.LauncherHelper.checkAndLoadMain(LauncherHelper.java:482)
asab@asab-HP-Compaq-8100-Elite-CMT-PC:/home/asab/PanRV$ which java
/usr/bin/java
asab@asab-HP-Compaq-8100-Elite-CMT-PC:/home/asab/PanRV$ java version
Error: Could not find or load main class version
asab@asab-HP-Compaq-8100-Elite-CMT-PC:/home/asab/PanRV$ java -version
java version "1.7.0_101"
OpenJDK Runtime Environment (IcedTea 2.6.6) (7u101-2.6.6-0ubuntu0.15.10.1)
OpenJDK 64-Bit Server VM (build 24.95-b01, mixed mode)
asab@asab-HP-Compaq-8100-Elite-CMT-PC:/home/asab/PanRV$
```

Figure 30: Unsupported major.minor version. This error means more than one versions of JAVA are working simultaneously. Check Java version. If it is less than 1.8 it is needed to delete older version

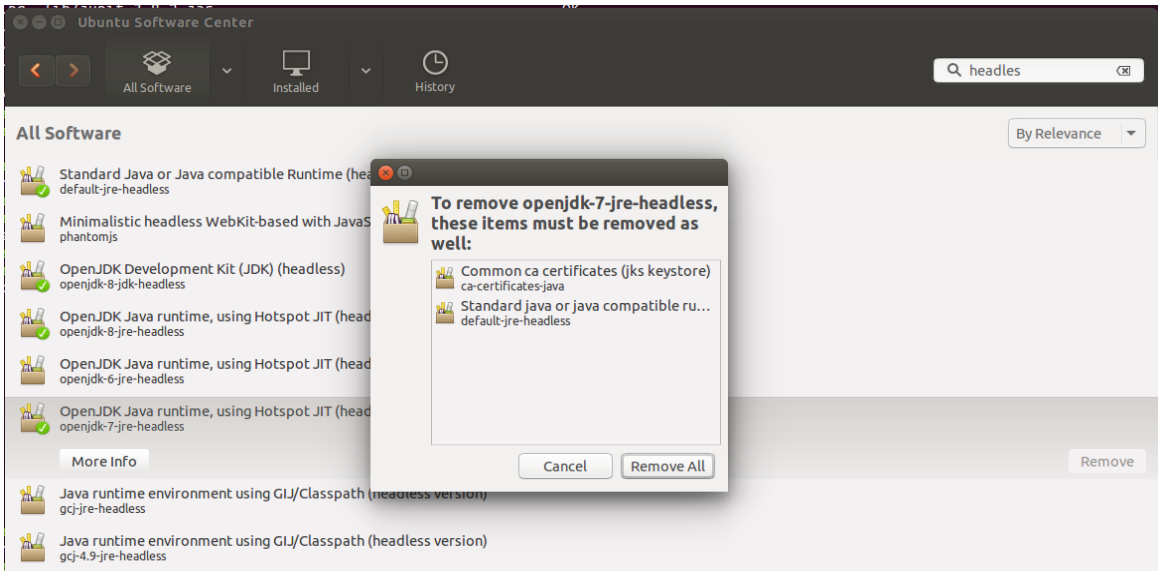

Figure 31: Java old version deletion: Go to software center; enter java, it displays different java libraries list. Remove openjdk-7-jre-headless, older JDK and JRE.Recheck Java version if updated and then start PanRV.