Prepare the list of virulence genes data of each species retrieved from VFDB for initial processing with Perl script

Filter sequence identity and completeness based on filter thresholds

Assign gene names to the filtered list by associating subject ID with VFDB

Structure the correlation between strains and virulence genes into data matrix

Generate an Excel file from the data matrix

R script reads the excel file into data frame

Manipulating data frame by assigning 1 and 0 for presence and absence of virulence genes

Removing row of virulence genes which are absent in all the strains

Convert manipulated data frame into data matrix

Feed the data matrix into pheatmap for visualization