

VIRIDIC v1.1 stand-alone – user manual

Developer: Cristina Moraru, email liliana.cristina.moraru@uni-oldenburg.de

Pre-requisites

- Install on your computer/server the Singularity v. 3.5.2 software (<https://sylabs.io/>)
- Download the viridic-singularity from viridic.icbm.de
- Remove from archive the folder (it contains the starting script and the singularity file)

To run VIRIDIC stand-alone

- Go to the folder with the viridic singularity
- Type `./viridic.bash projdir=FOLDER_name in=input_file`
- Note the syntax above: parameter name followed by "=" and then parameter value. For example: `res=similarity`. When forming the command, DON'T use the double quotes for the around the parameters.

To see the help file

- Go to the folder with the viridic singularity
- Type `./viridic.bash help`

To see the version of VIRIDIC

- Go to the folder with the viridic singularity
- Type `./viridic.bash version`

Parameters for VIRIDIC stand-alone

Mandatory

projdir the folder where VIRIDIC will save its output

in input .fasta file for VIRIDIC, containing all phage genomes

Optional

ram_max sets maximum ammount of RAM which can be used per CPU during the distance calculations and genome clustering. To be set if you encounter an error involving "future_globals_max_size_default".

Default: 2000 Mb

res which type of results should VIRIDIC return?
Default: "similarity" (intergenomic similarities default)
Others: "distance" (intergenomic distances)

steps which VIRIDIC steps do you want to run?
Default: "ALL" (intergenomic similarity calculation, clustering and heatmap)
Others:

“sim_clust” (only the calculation of intergenomic similarities and the clustering)

“clust_heatmap” (for previously run projects, it runs only the clustering and the heatmap)

“heatmap” (for previously run projects, it runs only the heatmap)

bln parameters for aligning with BLASTN. The parameters have to be surrounded by single quotes (the double quotes should be ignored when giving the command).

Default: “-word_size 7 -reward 2 -penalty -3 -gapopen 5 -gapextend 2”

Others:

“-word_size 20 -reward 1 -penalty -2”

“-word_size 11 -reward 2 -penalty -3 -gapopen 5 -gapextend 2”

“-word_size 28 -reward 1 -penalty -2”

ncor number of cores for BLASTN

clust which clustering method should VIRIDIC use?

Default: “complete”

Others: “ward.D”, “ward.D2”, “single”, “complete”, “average”, “mcquitty” (see hclust

R package)

thsp threshold for clustering at species level. It should be given as similarity.

Default: “95”

Range: 50-100 (below 50 the clustering can be wrong)

thgen threshold for clustering at genus level. It should be given as similarity.

Default: “70”

Range: 50-100 (below 50 the clustering can be wrong)

Optional for visualization of the heatmap

sim_cols Color fill for similarities/distances

Default: PuBuGn

Others: “Blues”, “BuGn”, “BuPu”, “GnBu”, “Greens”, “Greys”, “Oranges”, “OrRd”, “PuBu”, “PuRd”, “Purples”, “RdPu”, “Reds”, “YlGn”, “YlGnBu”, “YlOrBr”, “YlOrRd”

cols_Align Color fill for aligned genome fractions

Default: “peachpuff”

Others: “steelblue1”, “slategray2”, “skyblue1”, “lightsteelblue”, “thistle1”, “wheat1”, “moccasin”, “sandybrown”, “khaki1”, “antiquewhite”, “plum2”, “palegreen”, “seagreen1”

cols_Frac Color fill for genome length ratios

Default: “black”

Others: “none”, “blue”, “darkblue”, “cadetblue”, “darkgreen”, “chartreuse4”, “chartreuse”, “blueviolet”, “darkmagenta”, “coral4”, “firebrick4”

col_border_sim Border color for intergenomic similarities/distances

Default: “gray80”

Others: "none", "white", "gray98", "gray95", "gray90", "gray80", "gray70", "gray60", "gray50", "gray40", "gray30", "gray20", "gray10", "black"

col_border_frac Border color for fractions
Default: "gray80"
Others: "white", "gray98", "gray95", "gray90", "gray80", "gray70", "gray60", "gray50", "gray40", "gray30", "gray20", "gray10", "black"

show_sim show similarity values
Default: "TRUE"
Others: "FALSE"

show_sqLenFrac show values for the genome length ratios
Default: "TRUE"
Others: "FALSE"

show_qAligFrac show values for the aligned fraction genome1
Default: "TRUE"
Others: "FALSE"

show_sAligFrac show values for the aligned fraction genome2
Default: "TRUE"
Others: "FALSE"

font_sim font size for the similarity/distance values
Default: "8"
Others: only integers, min 1

font_sqLenFrac font size for the genome length ratios
Default: "4"
Others: only integers, min 1

font_qAligFrac font size for the aligned fraction genome1
Default: "4"
Others: only integers, min 1

font_sAligFrac font size for the aligned fraction genome2
Default: "4"
Others: only integers, min 1

font_row font size for the row names
Default: "12"
Others: only integers, min 1

font_col font size for the column names
Default: "12"
Others: only integers, min 1

annot_height Height of the top annotation displaying genome lengths

	Default: "10" Others: only integers, min 1, max 100
annot_font	font size for the annotation title Default: "100" Others: only integers, min 1, max 300
annot_rot	rotation of the annotation title Default: "270" Others: "0", "90" and "180"
lgd_width	width of the legends Default: "40" Others: only integers, min 1, max 100
lgd_font	font size for the legend titles Default: "3" Others: only integers, min 1, max 100
lgd_lab_font	font size for the legend labels Default: "3" Others: only integers, min 1, max 100
lgd_pos	position of the legend title Default: "topleft" Others: "topcenter", "leftcenter", "lefttop"
sim_for_frac	threshold to display intergenomic similarity/distances values Default: "0" if res="similarity", 100 if res="distance" Others: integers only, max 100
sim_for_frac	threshold to display the values for the genome length fraction and the aligned fraction genome 1 and genome 2 Default: "0" if res="similarity", 100 if res="distance" Others: integers only, max 100
sim_for_sim	threshold to display intergenomic similarity/distances values Default: "0" if res="similarity", 100 if res="distance" Others: integers only, max 100
lgd_height	height of the legends Default: "3" Others: only integers, min 1, max 100

Outputs of VIRIDIC stand-alone

All the files produced by VIRIDIC will be saved in the user defined project directory ("projdir" option). In here, the main outputs are found in subfolder 04_VIRIDIC_out:

- Heatmap.PDF (the heatmap – the main output)
- clusters.csv (a tab separated file with the genomes clustered at species and genus level, as defined by the “thsp” and “thgen” parameters)
- sim_MA_genCol.csv (a tab separated file containing the intergenomic distances between all genome pairs)

Other intermediary files are available in RDS format (storing single R objects), which can directly be implemented into R pipelines, if desired.