#########################################################################################

# virus\_detect.pl --file\_type [String] --reference [FILE] --coverage [Float]

# --host\_reference [FILE] --objective\_type [String]

# --max\_dist[INT] --max\_open [INT] --max\_extension [INT] --len\_seed [INT] --dist\_seed [INT] --thread\_num [INT]

# --strand\_specific --min\_overlap [INT] --max\_end\_clip [INT] --cpu\_num [INT] --mis\_penalty [INT] --gap\_cost [INT] --gap\_extension [INT]

# input\_file

#

# Basic options:

# --file\_type Format of input file (fastq or fasta) [fastq] warning if wrong format

# --reference The name of a fasta file containing all of the virus reference sequences [vrl\_plant]

# --coverage Minimum coverage of a reported reference by virus contigs [0.3]

# --host\_reference Name of a host reference file for subtraction [Null]

#

# BWA-related options (align sRNA to reference virus database or host sequences):

# --max\_dist Maximum edit distance [1]

# --max\_open Maximum number of gap opens [1]

# --max\_extension Maximum number of gap extensions [1]

# --len\_seed Take the first INT subsequence as seed [15]

# --dist\_seed Maximum edit distance in the seed [1]

#

# Megablast-related options (remove redundancy within virus contigs):

# --strand\_specific Only for sequences assembled from strand-specific RNA-seq [Not selected]

# --min\_overlap The minimum overlap length between two contigs to be combined [30]

# --max\_end\_clip The maximum length of end clips [6]

# --mis\_penalty Penalty for a nucleotide mismatch [-1]

# --gap\_cost Cost to open a gap [2]

# --gap\_extension Cost to extend a gap [1]

#

# Megablast-related options (align virus contigs to reference virus database for virus identification):

# --word\_size [11]

# --exp\_value [1e-5]

# --identity\_percen [80]

# --mis\_penalty\_b Penalty for a nucleotide mismatch [-1]

# --gap\_cost\_b Cost to open a gap [2]

# --gap\_extension\_b Cost to extend a gap [1]

#

# Result filter options(4):

# --hsp\_cover Coverage cutoff of a reported virus contig by reference virus sequences [0.75]

# --coverage\_cutoff Coverage cutoff of a reported virus reference sequences by assembled virus contigs [0.1]

# --depth\_cutoff Depth cutoff of a reported virus reference [5]

###########################################################################################

1. Align reads to reference virus sequence database

2. Align reads to host reference sequences

3. De novo assembly

4. Remove redundancies in virus contigs

5. Virus identification