

Virus Identification
(Genomad, VS2, CT3)

- fasta files



Preprocessing

Dereplication

- input: fasta
- output: directory
- + CheckV and BLAST

Decontamination

- input: fasta
- output: clean fasta
- + BLAST

Adjust seq

- input: fasta
- output: adjusted fasta
- + Biopython

Recombinaiton

- input: fasta
- output: filtered fasta
- + RDP, 3Seq, GENECONV, MaxChi, Chimaera, Bootscan, Siscan

output_clusters/.

|— clusters.tsv

|— cluster_sequences.fa

|— ani_results.tsv

|— blast_results.tsv

|— clustering.log

File with the first column as the cluster representative

Clustered FASTA file

Pairwise ANI (average identity) results

BLAST (n/p) results

Module log file

results/

|— clean_sequences.fasta

|— clean_sequences_stats.txt

|— clean_sequences_blast.tsv

|— clean_sequences_decontamination.log

Decontaminated sequences

Decontamination statistics

BLAST results (if keep-temp is used)

Detailed log file

output/.

|— recombination.log

|— aligned_sequences.3s.log

|— aligned_sequences.3s.pvalHist

|— aligned_sequences.3s.longRec

|— aligned_sequences.3s.rec

|— output.csv

log file

3Seq output

3Seq output

3Seq output

3Seq output