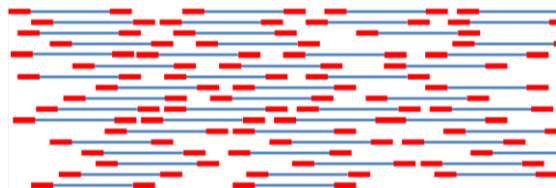


Metagenomic reads



Assembly

Contigs



Contigs with $\geq 95\%$ coverage with clean reads

Clean contigs



Dereplication with cd-hit-est

Non-redundant clean contigs

VIBRANT

Viral contigs

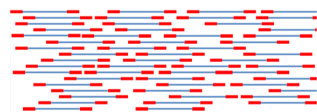
Predict rDNA

Map reads to potential cellular contaminant sequences: Bacteria, Archaea, fungi and protozoa

$\geq 5\text{‰}$ rDNA

$\geq 2\%$ mapped to cellular DNA

Datasets discarded



Clean reads not mapping to cellular sequences

Used to clean up contigs

