

## Raw data

795 million sequence reads



## Read trimming & adapter removal

609 million sequence reads retained (76.59%)

*trimmomatic*

min. length: 35 bp; leading 30; trailing 30; headcrop 6

*fastq-mcf*

l:50, q=20, qual\_mean=20

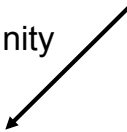


## PhiX174 & algal host sequence removal

123 million sequence reads retained (15.46%)

*BBSplit*

Viral community



## Taxonomic classification

263,878 read pairs retained  
(bacteria, archaea, viruses)

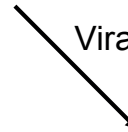
*CLARK*



## Virus taxonomy

41 families, 144 taxa  
49,557 read pairs

Viral gene expression



## De novo assembly & annotation

56,064 contig (N50 = 1,462bp)

4,856 contigs w/ viral genes

*rnaSPAdes & Prokka*



## Read mapping to assembly

(paired & unpaired reads)

*Kallisto-0.42.5*

paired reads: -b 10; single reads: -l 70 -s 18



## Differential gene expression

43 diff. exp. viral contigs

*metagenomeSeq*