

SeekViralReads results for sample “*sample*”

Adapters removal by cutadapt

command line:

```
cutadapt -n 5 -a CAGCGGACGCCTATGTGATG -g CATCACATAGGCGTCCGCTG -A CAGCGGACGCCTATGTGATG -G CATCACATAGGCGTCCGCTG
```

variable	value
Total read pairs processsed	1,000
R1 with adapters	44 (4.4%)
R2 with adapters	33 (3.3%)
R1 total length before	249,799 bp
R1 total length after	247,625 bp
R2 total length before	244,066 bp
R2 total length after	242,511 bp
Total length of filtered sequences (R1+R2)	490,136 bp (99.2%)
Reads pairs written	1,000 (100.0%)

Quality trimming by cutadapt

command line:

```
cutadapt -a CAGCGGACGCCTATGTGATG -g CATCACATAGGCGTCCGCTG -A CAGCGGACGCCTATGTGATG -G CATCACATAGGCGTCCGCTG
```

variable	value
Total reads pairs processsed	1,000
R1 with adapters	38 (3.8%)
R2 with adapters	43 (4.3%)
R1 total length before	247,625 bp
R1 total length after	224,515 bp
R2 total length before	242,511 bp

variable	value
R2 total length after	182,841 bp
Total length of quality trimmed sequences	407,356 bp (83.1%)
Reads pairs written	940 (94.0%)

Mapping on contaminant sequences

```
bwa mem -t 15 Data/Bacterial16S_and_Aedes_vexans_genome.fasta cleaned_sample_R1.fastq cleaned_sample_R2.fastq
```

variable	value
Reads input	940
Properly Mapped	391
Properly Mapped %	41.60%
Mapped with mate to a different chrom	24
Reads pair written	525

Merging overlapping reads with flash

variable	value
Reads input	525
Combined pairs	175
Percent combined	33.33%
Uncombined pairs	350
Number of output sequences	875

Dereplicate dataset with vsearch

variable	value
Input file name	sample_reads_potential_viral.fasta
Number of sequences	875

variable	value
Total length of sequences (nt)	211772
Number of unique sequences	724