NCBI SRA Filter output of Fetch data rRNA contamination initial search database search for analysis chimera analysis Necessary files: Search parameters/queries: Filtering conditions: Analysis steps: · Virome studies only • Number of reads • SILVA SSU + LSU rRNA 1) BWA align each RNA-seg FASTQ on (>= 100000 & <= 15000000) database PRINSEQ-filtered the SILVA rRNA ref FASTA • Related to Homo sapiens or human organism Average read size/length for <=1% ambiguous N's 2) BWA align each RNA-seq FASTQ on NO RNA-seg strategy only (>=100 & <=500) (SILVA rRNA ref FASTA) the corresponding virus ref FASTA Single-end layout only Context-based search of • RNA-seg data per sample 3) Detect rRNA-virus chimeric reads Illumina or Ion torrent virus-related samples (RNA-seg FASTQ) as found in common between the platform Exclusion of DNA or Studied viral genome per alignments of step 1 & 2 overrepresented viruses in sample (virus ref FASTA) 4) Calculate rRNA contamination & the examined samples rRNA-virus chimera statistics rRNA-virus YES chimeras present?

Implementation of the compared methods

Compared methods:

- RAW method i.e. the raw unprocessed reads
- BWA method i.e. the unmapped reads after BWA alignment on the SILVA rRNA ref FASTA
- SORTMERNA method i.e. the unmapped reads after SortMeRNA alignment on the SILVA rRNA ref FASTA
- ZWA2 method i.e. the clean reads after ZWA2 mapping and trimming based on the SILVA rRNA ref FASTA

Analysis steps:

- 1) Implement the compared methods separately on each RNA-seq FASTQ
- 2) Measure running performance of each method

Virus mapping statistics / analysis

Analysis steps:

- 1) BWA align the treated RNA-seq FASTQ by each compared method on the virus ref FASTA
- Calculate virus mapping / alignment statistics by samtools

De novo assembly statistics / analysis

Analysis steps:

- 1) MEGAHIT *de novo* assembly on the RNA-seq FASTQ files treated by each compared method
- Construction of a local BLASTn database based on the corresponding virus ref FASTA
- 3) BLASTn of the generated MEGAHIT contigs on the local viral database
- 4) Calculate de novo assembly statistics based on BLASTn output report & virus-specific assembly metrics

