

Virus detection using own host organism genome

The session course_VirusDetect_raspberry contains short RNA-seq from raspberry (*Rubus occidentalis*). As Chipster doesn't have the genome information for raspberry, we need to obtain it and use the tool **VirusDetect with own host genome**.

1. Select the file **raspberryReads.fasta** and check the number and length distribution of the reads using the tool **Quality control / Read quality statistics with PRINSEQ** so that in parameters you set **Input file format = FASTA**.
-How many reads are there and how long are they? How is the base quality?
2. Select **raspberryReads.fasta** and run **VirusDetect** with default parameters.
-How many reads aligned to virus reference database? How many contigs were assembled and how many viruses detected?
-Extract the html files to see which viruses were detected. Is the coverage good?
3. The raspberry genome sequence (*Rubus_occidentalis_v1.0.a1.scaffolds.fasta.gz*) was obtained from the Washington State University using the tool **Utilities / Download file from a URL directly to server**. Check what the file contains using the tool **Sequence file summary** (type the tool name in the search box to locate it).
-How many sequences (genomic contigs) does the file contain?
4. Select **raspberryReads.fasta** and **Rubus_occidentalis_v1.0.a1.scaffolds.fasta.gz** and run the tool **VirusDetect with own host genome** (check that the files are correctly assigned). This will take a while, because VirusDetect has to build the BWA index of the raspberry genome for the alignment.
-When the results arrive, can you locate the BWA index in the output files?
-How many reads aligned to the host genome?
-Did the number of de novo contigs increase when the host genome was included in the analysis? Were any host-derived contigs detected?
-Extract the html files and check if the same viruses were detected as before.
5. Practice using the ready-made host genome indexes: Select **raspberryReads.fasta** and **raspberryReads_hostgenome_bwa_index.tar** and run the tool **VirusDetect with own host genome** (check that the files are correctly assigned).