The Metagenomic *in silico* Lab: Design, Analysis and Validation of Metagenomic Investigations

Martin Norling, Oskar E. Karlsson, Hadrien Gourlé, Erik Bongcam-Rudloff and Juliette Hayer
SLU Global Bioinformatics Center, Swedish University of Agricultural Sciences, Uppsala, Sweden
Contact: juliette.hayer@slu.se

The overall aim is to provide virologists within veterinary medicine with an easy to use tool for experimental design, simulation and analysis of viral metagenomics experiments.

Experimental design

Provide an estimate of needed sequencing depth

Ensure enough coverage to detect all viral species in a sample, even those of low abundance

Bioinformatic analysis

Data preprocessing: quality control

Assembly of the sequencing reads

Taxonomic assignment / Binning

Evaluation

Testing tools on datasets with known content:
→ Simulated datasets

Results analysis and comparison

Validation

Experimental design module

Run pipelines module

- Environmental sample
- Environmental sample with assembly

Upload host genome

Data cleaning parameters

Simulate datasets module

- Ion Proton
- Ion Torrent (200bp or 400bp)

- exponential

- Estimation of the sequencing depth using an adaptation of Stevens' theorem (1).

- Prediction of the confidence value for finding all genomes present in a sample.
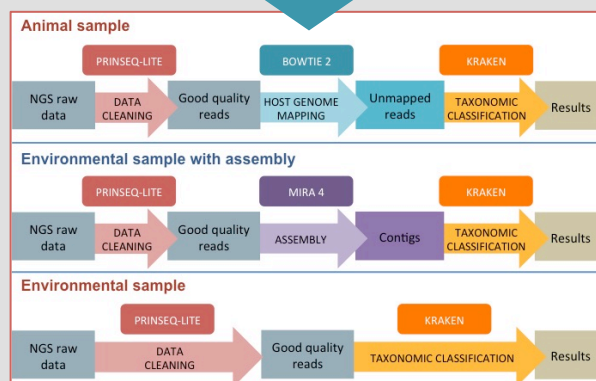
MetLab

Simulates datasets from:
- a selected sequencing technology
- a number of included species and their distribution.

Then downloads random viral genomes, generates the fastq file(s) and summarizes the taxonomic content.