

Viral genome analysis from wastewater sequencing data

Type of project: Bachelor End Project, Master End Project, Literature Review

Student background: CS, Nanobiology, LST

Daily supervisor: Jasmijn Baaijens, Pattern Recognition and Bioinformatics
(j.a.baaijens@tudelft.nl)

Background

Domestic wastewater is a rich source of information which can provide insight into the health of communities in an anonymized fashion. For example, wastewater monitoring of SARS-CoV-2 RNA has proven an effective and efficient technique to approximate COVID-19 case rates in the population. Indeed, COVID-19 patients excrete fragments of the viral genome through feces or urine, which ends up in domestic wastewater. These genome fragments can be extracted and sequenced: for each fragment the exact RNA sequence is determined. These short sequences, also called *reads*, can then be used to analyze the viral genome: which mutations do we see compared to previously observed sequences? How much viral diversity is present in the sample? Wastewater sequencing can also be used to track the prevalence of specific variants over time without having to analyze many patient samples: the population of an entire geographic area is represented in a single wastewater sample.

In this project you will expand and improve on existing techniques for SARS-CoV-2 analysis through wastewater sequencing, and investigate applications for other viruses.

Project ideas

Note that these are suggestions, the scope and focus of the project will be adapted to suit your expertise and interests, as well as your program requirements.

- Characterize the diversity between SARS-CoV-2 lineages and investigate how to select an “optimal” reference set
- Design and implement a framework to analyze SARS-CoV-2 diversity in wastewater at a desired granularity: lineage, WHO variants, or pre-defined classes of lineages.
- Explore application on other viruses (presence in wastewater, availability of reference sequences, genomic diversity) and evaluate performance