BIO-312: Practical on Environmental DNA Analysis

Session 3: Environmental DNA from Freshwater Sampling

By Katherine Delevaux and Prof. Sebastian Waszak
Laboratory of Computational Neuro-Oncology, Swiss Institute of Experimental Cancer Research, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Switzerland
https://www.epfl.ch/labs/upwaszak/

Background

For this session, we will shift gears to taxonomic analysis of freshwater samples of the river that flows through Cambridge, UK. The sequencing data was published as part of https://elifesciences.org/articles/61504. PDF is also available on Moodle.

To summarize, the authors collected 1 liter of water from many sites along the river in April, June, and August 2018. From each sample, DNA was extracted, amplified, and then sequenced with a Nanopore MinION device. The data was made publicly available through the European Nucleotide Archive.

Part 0: Connect to czid.org and join the *Freshwater Project*

In this part, we will look at taxonomic classification using czid.org. It is an open-source cloud-based metagenomics pipeline.

Open up czid.org and make an account using your EPFL email address. Join the Freshwater project by giving your account email address to Katherine and she'll add you.

Part 1: Taxonomic Classification Exploration on czid.org

We have downloaded the sequencing data and uploaded the samples to czid.org for taxonomic classification.

If you didn't get the chance last week, take a few minutes to explore the site for what information is available. Note that some of the samples failed the pipeline. Unfortunately, czid.org does not provide information for debugging why they failed.

In the Total Reads column, you can get a sense of how many reads there were per sample.

Are you surprised by the significant variability of reads per sample?
How is the quality of the reads based on the Passed QC column?

Click on a sample, in the top right *Download > View Pipeline Visualization*. Click on some of the pipeline steps to see a more in-depth description.

Czid.org is geared towards pathogen detection. At the top right you can filter for only known pathogens.

How many known pathogens are present in the sample? Do you recognize any of the species?

Part 2: Data Exploration and Taxonomic Analysis on Noto

Let’s switch over to Noto https://noto.epfl.ch, to analyze the czid.org outputs.

Download **Freshwater_Analysis.ipynb** and the data file **all_samples_data.tsv** from the Moodle and upload them to a new folder in Noto. You will follow along the **Freshwater_Analysis.ipynb** for the analysis.

We downloaded the classified reads from czid.org and then used the *R* package taxize to connect to NCBI to add genus and family names based on the tax_id.