Brief description

Introduction to NCBI Pathogen Detection and Google Cloud Based Resources for Pathogen NGS Analysis

Course Objective:
Cloud computing allows users to bring the compute to the data. There is an increasing push to make data FAIR (findable, accessible, interoperable, and repeatable) by depositing data in cloud environments. NIH is starting to take advantage of cloud-based resources, and NCBI/NLM are starting to make data, including next generation sequencing (NGS) data from SRA available in the cloud along with a number of tools.

The NCBI Pathogen Detection team will host a very brief beginner-level workshop introducing some of these concepts. We will show some of the following: existing cloud-based datasets that NCBI/NLM/NIH are making available and how to access them, cloud-based compute tools, and command-line tools that can be used for NGS analysis. Participants will use published data from a hospital metagenomics study to learn how to access and work with these tools and datasets.

Participants will gain practical experience Using Google Cloud Platform (GCP) with:

Accessing NCBI SRA data and metadata in the cloud
Accessing NCBI tools with cloud-capable computing such as ElasticBLAST
Utilizing other NCBI genomic analysis tools of interest to analyze bacterial pathogens, including NCBI's AMRFinderPlus
Using the Pathogen Detection interface to search and identify interesting sequences to be used for further analyses

Target Audience:
Biologists and bioinformaticians working in the area of pathogen genomics, who are familiar with command line tools, and looking to learn more about NCBI cloud-based tools and datasets, and the Pathogen Detection project.

Prerequisites for attendance:
Familiarity with Linux environment and UNIX commands and command-line interfaces.

If interested:

Contact Bill Klimke @ NCBI by Nov 20th, 2020.

klimke@ncbi.nlm.nih.gov