CBW2023 Contextual Data Curation Scenarios

The following mock scenarios describe contextual data for SARS-CoV-2 sequences submitted during a public health emergency to the National COVID-19 Surveillance Database. The scenarios in this exercise are fictional and do not represent real data or programs at any public health organization in Canada, and are intended for training purposes only.

Scenario 1

The **BCCDC Public Health Laboratory** obtained a **nasopharyngeal swab** for **diagnostic testing (sample ID Bc-12345-ab)** on **March 1 2023** from a **symptomatic, 44 year old female** that had been **hospitalized** in the **ICU**. The individual had been exhibiting a **cough, fever, muscle weakness**, as well as other symptoms of **Acute Respiratory Distress Syndrome**. The individual recently **travelled** to the **United States** on holiday and returned on **Feb 19 2023**. The sample was flagged for sequencing as part of the lab's **International travel surveillance** program. The sample was sequenced on **March 7 2023** using an **Illumina MiSeq** instrument. The raw data was processed using **ncov-tools** (https://github.com/jts/ncov2019-artic-nf/blob/master/README.md) and dehosted using **BWA (version 0.7.17)**. The consensus sequence was generated using **iVar 2.3.1**. The sequence was uploaded to GISAID and assigned the accession number **EPI_ISL_436489**. Drs **Tejinder Singh, Fei Hu and Joe Blogs** helped to generate the sequence.

Note: GISAID "isolate" identifiers are generated by data providers in the following format
hCoV-19/COUNTRY/ISO regional code-Identifier/year
e.g. hCoV-19/CANADA/BC-provlab1234/2020
"hCoV-19" is always the same
The country is capitalized e.g. CANADA
ISO regional codes are 2 letter codes for the province or state e.g. BC, ON, QC, SK, AB
The identifier corresponds to the sample ID.
The "year" corresponds to the year of sample collection.

Scenario 2

Alberta Precision Labs (APL) were sent a swab from a local hospital lab (sample ID 192038475y) that had been acquired for diagnostic testing on Dec 21 2022. The sample was an oropharyngeal swab from a 25 year old (undeclared gender). The individual was considered fully vaccinated, having received 3 doses of the Pfizer Comirnaty vaccine. The swab was selected for sequencing as part of the public health lab's Vaccine escape surveillance program. The sample was sequenced on Jan 11 2023 using an Oxford Nanopore MinION. The sequence data was trimmed and processed using the Nextflow pipeline for running the ARTIC Network's field bioinformatics tools (https://github.com/phac-nml/ncov2019-artic-nf). Human reads were removed using the ncov-dehoster (https://github.com/phac-nml/ncov-dehoster and the consensus sequence was generated using iVar 2.3.1.

Scenario 3

A sample from an asymptomatic, 78 year old male was sequenced as part of an outbreak investigation in Ontario. Public Health Ontario received a nasal cavity swab (sample ID brc-333-666) on Feb 7 2023 for diagnostic testing. On Feb 14 2023 the sample was sent for sequencing as part of baseline surveillance. The sample was sequenced on an Oxford Nanopore GridION, and the data was processed using software described at https://github.com/phac-nml/ncov2019-artic-nf and dehosted using software described at https://github.com/phac-nml/ncov-dehoster. The consensus sequence was produced using iVar 2.3.1.

Scenario 4

The Roy Romanow Provincial Laboratory in Saskatchewan received a saline gargle sample for diagnostic testing (sample ID SC2-NS-12495857) on Sept 26 2022. The sample was from a 6 year old Female who had recently attended a family gathering with parents. The sample was sequenced on Oct 19 2022 as part of the province's baseline surveillance program on an Illumina MiSeq, and the raw data was processed and dehosted using the SIGNAL pipeline (https://github.com/jaleezyy/covid-19-signal) and the consensus sequence was generated using iVar 2.3.1.

Scenario 5

As part of its Respiratory Viruses in the Air surveillance program, the Laboratoire de santé publique du Québec (LSPQ) collected material extracted from the air filter of an air vent in a hospital. Two filters per room were pooled before sequencing to increase throughput. The sample was collected in May 2020 just after the start of the pandemic (sample ID fgh-34556-12xy). The sample was sequenced as part of a research project to understand viral spread via contact surfaces and airborne transmission. The sample was sequenced on September 23 2023 on an Illumina MiSeq. The raw data was processed and dehosted using the SIGNAL pipeline (https://github.com/jaleezyy/covid-19-signal) and the consensus sequence was generated using iVar 2.3.1.