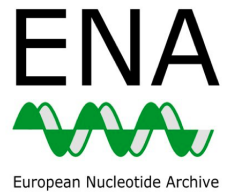


Jul 09, 2020

SARS-CoV2 EBI assembly submission protocol

DOI

<https://dx.doi.org/10.17504/protocols.io.bhwqj7dw>



Nabil-Fareed Alikhan¹, Emma Griffiths², Ruth Timme³, Duncan MacCannell⁴

¹Quadram Institute Bioscience; ²University of British Columbia; ³US Food and Drug Administration;

⁴Centers for Disease Control and Prevention

Coronavirus Method De...

PHA4GE



Nabil-Fareed Alikhan

University of Oxford

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.bhwqj7dw>

Protocol Citation: Nabil-Fareed Alikhan, Emma Griffiths, Ruth Timme, Duncan MacCannell 2020. SARS-CoV2 EBI assembly submission protocol . **protocols.io** <https://dx.doi.org/10.17504/protocols.io.bhwqj7dw>



License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: June 25, 2020

Last Modified: November 11, 2021

Protocol Integer ID: 38576

Keywords: metadata, INSDC, ERC000033, ENA, EBI, SARS-CoV2, COVID-19, cov2 ebi assembly submission protocol, biosamples submit sar, pha4ge guidance for sar, submit sar, included protocol, established bioproject, pha4ge guidance, existing bioproject, new bioproject, first data submission, cov2, new webin laboratory submission account, protocol, bioproject, metadata, european nucleotide archive, ena, ebi, raw sequence data, sar, biosample,

Disclaimer

Please note that this protocol is public domain, which supersedes the CC-BY license default used by protocols.io.

Abstract

PURPOSE:

This protocol covers the steps for submitting a SARS-CoV-2 assembly to ENA

For new submitters, there's quite a bit of groundwork that needs to be established before a laboratory can start its first data submission. We recommend that one person in the laboratory take a few days to get everything set up in advance of when you expect to do your first data submission.

Two protocols cover the PHA4GE guidance for SARS-CoV-2 submission to ENA (Raw sequence data, metadata, and assemblies)

Complete in order (1 then 2):

1. SARS-CoV-2 EBI submission protocol: ENA, BioSample, and BioProject

- Step-by-step instructions for establishing a new Webin laboratory submission account and for creating and linking a new BioProject to an existing umbrella effort.
- Submit SARS-CoV-2 raw data to ENA (European Nucleotide Archive) and metadata.

2. SARS-CoV-2 EBI assembly submission protocol (included protocol)

Required: established BioProject and BioSamples

- Submit SARS-CoV-2 assemblies to ENA linking to existing BioProject, BioSamples, and raw data.

- 1 The Webin-CLI program is described as the only way to upload assembled sequences. This includes consensus sequences of SARS-COV2. Generally, you should follow the guidance here: <https://ena-docs.readthedocs.io/en/latest/submit/general-guide/webin-cli.html>

The process, briefly, is as follows:

1. Download the Webin-cli tool. It is a Java program, so you will need the Java runtime environment installed as well.
2. Create manifest files, one for each assembly,
3. Submit the manifest file, and the associated assembly data via the Webin tool.

To begin SARSCoV2 consensus sequences will need to be submitted as a chromosome assembly, see <https://ena-docs.readthedocs.io/en/latest/submit/assembly/genome.html#chromosome-assembly>

Note

Assemblies can only be submitted using **Webin-CLI**, using-context genome. During the process, you must define metadata in the **manifest file(s)**. Please specify 'COVID-19 outbreak' as the 'ASSEMBLY_TYPE'.

- 2 The assembly submission system is set up to point to an existing sample record. You must have already created your Project/Study and registered your samples to proceed with the assembly submission.

For each record you will need 3 different files:

1. A manifest file that details the sample the assembly should be associated, and some other metadata.
2. The assembly sequence itself, this should be a FASTA file compressed with gz.
3. The chromosome file which details the order of the sequences in the FASTA file. This again, should be compressed with gz.

The manifest file could look something like this:



```
STUDY    ERP123456
SAMPLE   ERS123456
RUN_REF  ERR123456
FASTA    SARSCOV_Seq_Example.fasta.gz
NAME     SARSCOV_Seq_Example
ASSEMBLY_TYPE COVID-19 outbreak
PROGRAM   ARTIC-ivar
PLATFORM   Illumina
COVERAGE   1000
CHROMOSOME_LIST Illumina_NORW-EA35E.chrom.gz
```

Note here, that the STUDY, SAMPLE and RUN/EXPERIMENT must be specified, which means you should have already created these records and you should have the accession numbers to populate these fields.

See the Webin documentation for more information <https://ena-docs.readthedocs.io/en/latest/submit/general-guide/webin-cli.html>

3 Creating the supporting files

The supporting files include the consensus sequence and the chromosome file. Both must be compressed (gz format) to be submitted by the Webin-Cli program.

The chromosome file which details the order of the sequences in the FASTA file. Since the your COVID19 consensus sequence should be single contiguous sequence, the file is very simple, with the name of the sequence (the FASTA header), being the first (and only sequence) [1].

```
SARSCOV_Seq_Example 1      Chromosome
```

The FASTA sequence is a standard FASTA format. The header should match the name given in the chromosome file.

```
>SARSCOV_Seq_Example
ATAGTCACATAGCAATCTTTATCACATAGCAATCTTTATCACATAGCAA
TCTTTATCACATAGCAATCTTTA...
```



You need one chromosome file, for each FASTA file.

See the Webin documentation for more information <https://ena-docs.readthedocs.io/en/latest/submit/general-guide/webin-cli.html>

4 **Submitting a consensus sequence**

With all the supporting files ready you can submit them with the webin tool, example:

```
java -jar webin-cli-3.0.0.jar -context genome --  
manifest manifest_file.txt -inputDir myData -  
outputDir mydata_reports -submit -userName Webin-12345 -  
passwordFile mypassword
```