



Version 1

Feb 11, 2021

Cecret Workflow Protocol for SARS-CoV-2 Assembly and Lineage Classification V.1

In 1 collection

Erin L Young¹, TOAST CDC Workgroup²¹Utah Public Health Laboratory; ²Centers for Disease Control and Prevention
TOAST CDC Workgroup: toast@cdc.gov;

In Development

This protocol is published without a DOI.

Coronavirus Method Development Community

ycm6

SUBMIT TO PLOS ONE

ABSTRACT

This protocol provides instructions on how to install the staphb_toolkit and run the Cecret workflow. Cecret produces SARS-CoV-2 consensus sequences from single or paired-end fastq.gz or fastq reads and generates lineage classifications using Pangolin and Nextclade. This document applies to all WGS on the Illumina platform and downstream bioinformatics for CDC partner laboratories.

EXTERNAL LINK

https://github.com/StaPH-B/staphb_toolkit

EXTERNAL LINK

https://github.com/StaPH-B/staphb_toolkit

PROTOCOL CITATION

Erin L Young, TOAST CDC Workgroup 2021. Cecret Workflow Protocol for SARS-CoV-2 Assembly and Lineage Classification . **protocols.io**
<https://protocols.io/view/cecret-workflow-protocol-for-sars-cov-2-assembly-a-bsbhnaj6>

COLLECTIONS ⓘ

Menu

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

CREATED

Feb 10, 2021

LAST MODIFIED

Feb 11, 2021

PROTOCOL INTEGER ID

47177

PARENT PROTOCOLS

Part of collection

[Menu](#)

Load Dependencies

1 Load Software Dependencies

The staphb_toolkit requires either singularity or docker, Python 3.6 or greater, and Java version 8 or later. If you do not yet have these installed, do so at this time, and ensure they are in your PATH. Within the CDC high-performance computing cluster, these dependencies can be loaded into the environment using the following GNU module commands.

Loading dependencies for staphb_toolkit

```
module load Python/3.9.1  
module load java/jdk1.8.0_221  
module load nextflow/20.04.1  
module load singularity/3.5.3
```

Install staphb_toolkit

2 Install the staphb_toolkit which includes the Cecret assembly workflow

staphb_toolkit Install

```
git clone https://github.com/StaPH-B/staphb_toolkit.git  
cd staphb_toolkit/packaging/  
python3 setup.py install --user  
cd ../  
export PATH=$PATH:$(pwd)
```

Running Cecret Workflow

3 Run the Cecret workflow to perform sequence assembly and lineage classification

The input directory for Cecret contains a set of single or paired-end (default) fastq.gz or fastq reads from amplicon prepared libraries. By default, Cecret is configured to use the Arctic SARS-CoV-2 V3 primer set for primer trimming. However, this is customizable if another primer set is used. In this run we set the *--profile* argument to use Singularity containers but Cecret works with Docker containers as well (*--profile docker*).

Cecret Input and Output File Paths

```
#Input Sequencing Reads File Path:  
/Full_Path_to_Fastq_File_Directory/INPUT/SRR11953697  
  
#Output Directory:  
/Full_Path_to_Cecret_Output_Directory/SRR11953697_cecret_output
```

Cecret Workflow Command

```
staphb-wf cecret /Full_Path_to_Fastq_File_Directory/INPUT/SRR11953697 --output  
/Full_Path_to_Cecret_Output_Directory/SRR11953697_cecret_output --profile  
singularity
```

Access Output Files

- 4 Paths for the assembled consensus sequence and the summary report file which contain Pangolin and Nextclade lineage classifications

Cecret Consensus Sequence and Summary Report Paths

#Consensus Sequence Path:

/Full_Path_to_Cecret_Output_Directory/SRR11953697_cecret_output/consensus/SRR

#Summary Report Path:

/Full_Path_to_Cecret_Output_Directory/SRR11953697_cecret_output/summary.txt

Step 4 includes a Step case.

Pangolin Error