SeqSender

Advanced Molecular Detection
Southeast Region Bioinformatics

OVERVIEW

Purpose

- A CLI tool developed by the CDC to streamline and automate the submission of viral and pathogen WGS data to public databases such as GenBank, SRA, GISAID and BioSample.
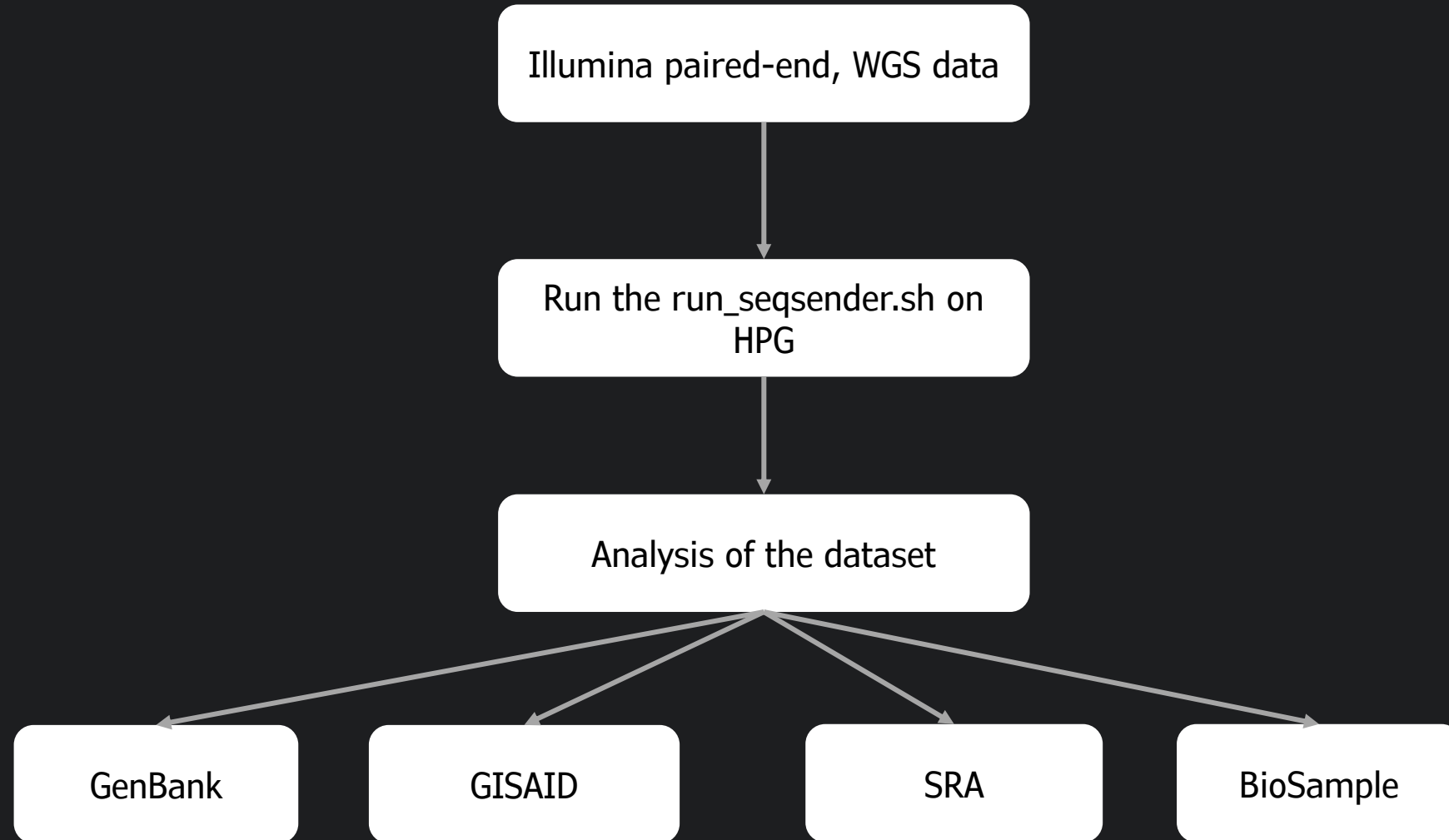
Usage

- Used by public health labs, researchers, and sequencing centers to generate standardized submission files and upload assembled genomes with metadata to repositories.

Dependencies

- Conda

WORKFLOW

Application

Objective
Use a SARS-CoV-2 dataset and generate an output from the SeqSender pipeline

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/seqsender0715/

```
Name
    ..
    __pycache__
    config
    docs
    inst
    man
    output
    shiny
    test_data
    _pkgdown.yml
    argument_handler.py
    biosample_sra_handler.py
    DESCRIPTION
    docker-compose.yaml
    Dockerfile
    env.yaml
    file_handler.py
    genbank_handler.py
    gisaid_handler.py
    LICENSE
    mypy.ini
    ncbi_handler.py
    NEWS.md
    README.md
    README.Rmd
    run_seqsender.sh
    seqsender.7907554.err
    seqsender.7907554.out
    seqsender.py
    seqsender-kickoff
    settings.py
    setup.py
    tools.py
    upload_log.py
```

cd /blue/bphl-<state>/<user>/repos/bphl-molecular/

git clone https://github.com/CDCgov/seqsender

mkdir analysis/

cd analysis/

cp /blue/bphl-<state>/<user>/repos/bphl-molecular/ seqsender/

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

```
GNU nano 5.6.1                                          run_seqsender.sh
#!/usr/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=seqsender
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=20
#SBATCH --mem=20gb
#SBATCH --time=48:00:00
#SBATCH --output=seqsender.%j.out
#SBATCH --error=seqsender.%j.err


python seqsender.py submit \
    --organism COV \
    --submission_name run1 \
    --submission_dir output \
    --config_file test_data/COV/cov_config.yaml \
    --metadata_file test_data/COV/cov_genbank_metadata.csv \
    --fasta_file test_data/COV/cov_sequence.fasta \
    --genbank
```

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/seqsender0715/

Name
```
..
__pycache__
config
docs
inst
man
output
shiny
test_data
_pkgdown.yml
argument_handler.py
biosample_sra_handler.py
DESCRIPTION
docker-compose.yaml
Dockerfile
env.yaml
file_handler.py
genbank_handler.py
gisaid_handler.py
LICENSE
mypy.ini
ncbi_handler.py
NEWS.md
README.md
README.Rmd
run_seqsender.sh
seqsender.7907554.err
seqsender.7907554.out
seqsender.py
seqsender-kickoff
settings.py
setup.py
tools.py
upload_log.py
```

Submitting To NCBI

Enter your username and password in the config file

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

```
module load conda

conda env –f env.yaml

conda activate seqsender

sbatch run_seqsender.sh
```

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/seqsender0715/

Name
- ..
- __pycache__
- config
- docs
- inst
- man
- output
- shiny
- test_data
- _pkgdown.yml
- argument_handler.py
- biosample_sra_handler.py
- DESCRIPTION
- docker-compose.yaml
- Dockerfile
- env.yaml
- file_handler.py
- genbank_handler.py
- gisaid_handler.py
- LICENSE
- mypy.ini
- ncbi_handler.py
- NEWS.md
- README.md
- README.Rmd
- run_seqsender.sh
- seqsender.7907554.err
- seqsender.7907554.out
- seqsender.py
- seqsender-kickoff
- settings.py
- setup.py
- tools.py
- upload_log.py

Application Cont.

Application Cont.

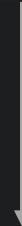/.../n.yengalareddy/repos/bphl-molecular/analysis/seqsender0715/output/run1/submission_files/

Name
..
GENBANK
submission_status_report.csv

gb-sample_name	genbank_status	genbank_accession	genbank_message
SARS-CoV-2/human/USA/GA_2741/2020			
SARS-CoV-2/human/USA/GA_3742/2020			

Application Cont.

/.../n.yengalareddy/repos/bphl-molecular/analysis/seqsender0715/output/run1/submission_files/

Name

📁 ..
📁 GENBANK
📄 submission_status_report.csv

/.../n.yengalareddy/repos/bphl-molecular/analysis/seqsender0715/output/run1/submission_files/GENBANK/

Name

📁 ..
📄 authorset.sbt
📄 comment.cmt
📁 run1.zip
📄 sequence.fsa
📄 source.src
📄 submission.xml
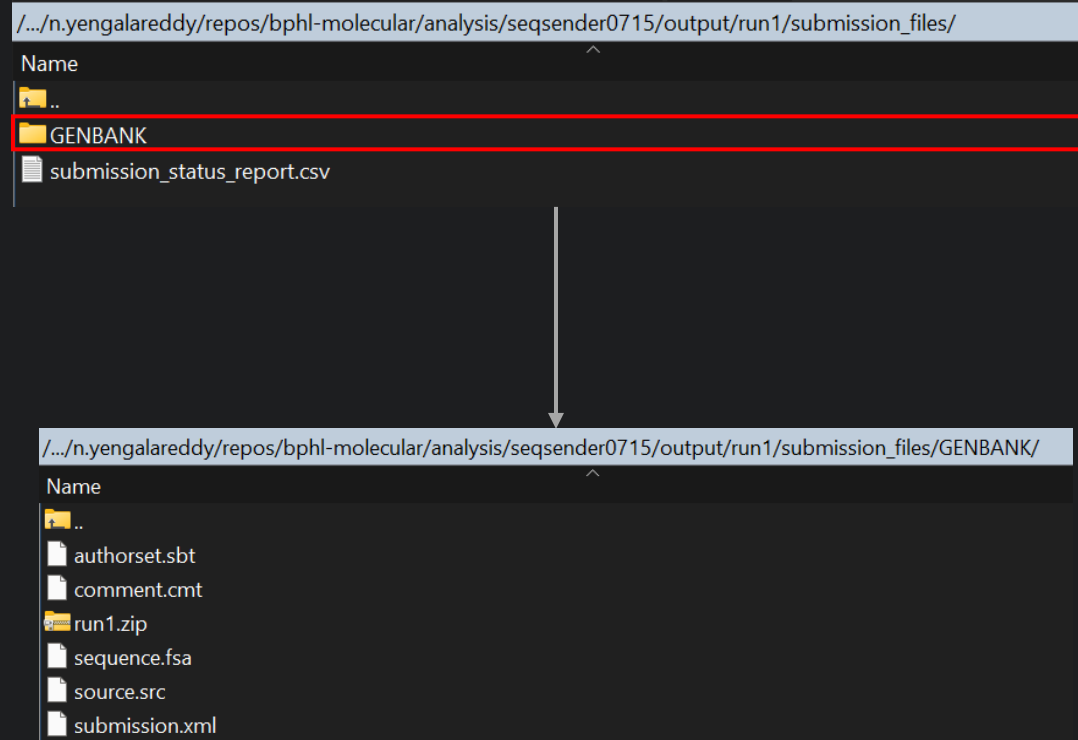
Conclusion

Fundamentals of SeqSender

Installation and setup of SeqSender in HPG

Successfully executed job query for SeqSender

Generated output files

Advanced Molecular Detection Southeast Region Bioinformatics

Questions?

bphl-sebioinformatics@flhealth.gov

Molly Mitchell, PhD
Bioinformatics Supervisor
Molly.Mitchell@flhealth.gov

Nikhil Reddy, MS
Bioinformatician
Nikhil.Yengala@flhealth.gov

Sam Bernhoft, MPH
Bioinformatician
Samantha.bernhoft@flhealth.gov