



Advanced Molecular Detection

Southeast Region Bioinformatics

Office Hours – Session 3
04/03/2023

Outline



ELC Updates/Reminders



Agenda



What is Nextflow?



Importance of Nextflow



Using Nextflow on HPG



Nextflow Training



Questions

ELC Updates/Reminders

- Program A, Activity 1f – AMD Regional Workforce Development Training Participant
 - Please apply for this to send participants to an in-person training workshop (hosted by ABiL) in Atlanta, GA in Fall 2023
 - \$3,000-\$10,000 available for each jurisdiction
- Program A, Activity 6a – AMD Platform Support Core Activities
 - Please apply and participate in at least two Communities of Practice
 - You do not need to be a subject matter expert in any of the CoPs. Participation in CoPs mainly allows supporting states to provide their needs and input for features for the new AMD Platform
 - \$25,000 available for each jurisdiction
 - Can ask for maintenance of equipment, procurement of supplies, staff time
 - ***CDC said that most jurisdictions did not apply. There is an award available for all ELC jurisdictions.**

Agenda

April 24 - PHoeNIx #1 (Installation & Dependencies)

May 1 - PHoeNIx #2 (Demo and Questions/Troubleshooting)

May 15 – AMRFinder+ and PHA4GE's hAMRonization pipeline

May 29 - ONT & FL's Flisochar pipeline

Future Trainings

- StaPH-B Toolkit Programs/Pipelines
- GISAID flagged SARS-CoV-2
- Git (git clone, etc.)
- Generating R figures
- ...and more

What is Nextflow?

- [Nextflow's documentation! — Nextflow 22.10.6 documentation](#)
- “Its fluent DSL simplifies the implementation and deployment of complex parallel and reactive workflows on clouds and clusters”
 - DSL – Domain Specific Language
- Enables scalable and reproducible scientific workflows using containers
 - **Singularity/Apptainer** or Docker
- Adaptable pipelines which are written in the most common scripting languages
- Can be used with conda environments
- There is also cloud support/utility
 - AWS, Google, Azure

Nextflow Processes and Channels

- A Nextflow script is made by joining multiple processes
- These processes can be written in any scripting language and are executed on a linux platform
 - Bash, Perl, Ruby, Python, etc.
- Processes can be run independently and are isolated from one another
 - Each process defines a channel as input and output
 - The string of processes (with inputs and outputs) make up the pipeline
- The pipeline runs as a channel, running the processes (can be simultaneous)

Importance of Nextflow

- We will be using Nextflow (as a dependency) for installing and using PHoeNix
- Also, all our [BPHL-Molecular](#) pipelines are being converted to Nextflow
- Most bioinformatics pipelines are now developed using Nextflow
- Nextflow is also the most common and will be the main format for the AMR Platform pipelines

Using Nextflow on HPG

- Load Nextflow module

```
$ module load nextflow
```

- Or, you can create a conda/mamba specific nextflow environment
(Note: we will be doing this for **PHoeNix #1**)

```
$ module load conda
```

```
mamba create -yp /blue/bphl-<state>/<user>/<conda_envs>/nextflow
```

```
mamba activate /blue/bphl-<state>/<user>/<conda_envs>/nextflow
```

```
mamba install -c bioconda nextflow=21.10.6
```


Using Nextflow on HPG

```
$ nextflow -h
```

```
Usage: nextflow [options] COMMAND [arg...]

Options:
  -C                Use the specified configuration file(s) overriding any defaults
  -D                Set JVM properties
  -bg              Execute nextflow in background
  -c, -config       Add the specified file to configuration set
  -config-ignore-includes Disable the parsing of config includes
  -d, -dockerize    Launch nextflow via Docker (experimental)
  -h               Print this help
  -log             Set nextflow log file path
  -q, -quiet       Do not print information messages
  -syslog          Send logs to syslog server (eg. localhost:514)
  -trace           Enable trace level logging for the specified package name - multiple packages can be provided separating them with a comma e.g. '-trace nextflow,io.segura'
  -v, -version     Print the program version

Commands:
  clean            Clean up project cache and work directories
  clone           Clone a project into a folder
  config          Print a project configuration
  console         Launch Nextflow interactive console
  drop            Delete the local copy of a project
  help            Print the usage help for a command
  info            Print project and system runtime information
  kuberun         Execute a workflow in a Kubernetes cluster (experimental)
  list            List all downloaded projects
  log             Print executions log and runtime info
  pull           Download or update a project
  run            Execute a pipeline project
  secrets         Manage pipeline secrets (preview)
  self-update     Update nextflow runtime to the latest available version
  view           View project script file(s)
```

Using Nextflow on HPG

```
$ nextflow -version
```

```
[mollymitchell@login5 ~]$ nextflow -version  
  
N E X T F L O W  
version 22.11.1-edge build 5834  
created 29-11-2022 15:16 UTC (10:16 EDT)  
cite doi:10.1038/nbt.3820  
http://nextflow.io
```



Basics of a Nextflow command

- What's included in a basic Nextflow command
 - Calling nextflow
 - Options
 - Command argument (run, pull, clone, clean, help, etc.)
 - Script file
 - Profile
 - Entry – if necessary

```
$ nextflow [options] COMMAND [arg...] script_file -profile <singularity/docker/custom> -entry [option]
```

Nextflow Training

- [Nextflow: implementing a simple pipeline | Microbiome binfies \(telatin.github.io\)](#)
 - Andrea Telatin's Nextflow Training
- [Nextflow Training Workshop | Seqera Labs](#)
- Use these links to learn more about writing and executing your own Nextflow workflows!

References

[Introduction to Nextflow](#) by Andrew Severin, PhD.

[Nextflow's documentation! — Nextflow 22.10.6 documentation](#)

Time for Questions & Feedback

- Questions?
 - Do you need help with anything?
 - Requests for separate trainings?
- Feedback
 - What would you like to see?



Advanced Molecular Detection Southeast Region Bioinformatics

Questions?

bphl16bioinformatics@flhealth.gov

Lakshmi Thsaliki, MS

Bioinformatician

Lakshmi.Thsaliki@flhealth.gov

Molly Mitchell

Bioinformatician

Molly.Mitchell@flhealth.gov

Sarah Schmedes, PhD

Lead Bioinformatician

Sarah.Schmedes@flhealth.gov