



Advanced Molecular Detection

Southeast Region Bioinformatics

Outline



Updates and Reminders



Agenda



What is PHoeNix?



Dependencies & Install



Testing Install & Running PHoeNix



Output File Structure & Pipeline Summary



Questions

Updates and Reminders

- Please email us on bphl-sebioinformatics@flhealth.gov for any queries or requests.
- One of the group members will respond to your query.

Agenda

April 24 – PHoeNix #1 (Installation and Dependencies)

May 1 – PHoeNix #2 (Demo & Troubleshooting)

May 15 – Open OnDemand

May 29 – AMRFinder+ and Pha4ge's hAMRonization pipeline

Future Trainings

- ONT & FL's Flisochar pipeline
- StaPH-B Toolkit Programs/Pipelines
- GISAID flagged SARS-CoV-2
- Git (git clone, etc.)
- Generating R figures
- Outbreak/cluster training
- SRA human scrubber tool
- ...and more

What is PHoeNix?

- **PHoeNix** – Portable Healthcare Nextgen Informatics pipeline
- Developed by the Division of Healthcare Quality Promotion (DHQP) at the CDC for pathogens commonly encountered in healthcare settings (including but not limited to e.g., *Acinetobacter baumannii*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus*)
- Pipeline is built using **Nextflow** which runs tasks across multiple compute infrastructures in a portable manner
- Uses **Docker/Singularity** containers making installation trivial and highly reproducible results
- Pipeline provides a standardized approach for identifying and characterizing healthcare-associated bacterial pathogens, specifically for public health partners
- Nextflow DSL2 implementation of this pipeline uses one container per process which makes it easier to maintain and update software dependencies

PHoeNlx Pipeline Insights

- PHoeNlx uses Illumina paired-end reads and was designed for use with pathogens causing healthcare-associated bacterial infections. This comprehensive pipeline performs:
 1. Quality control
 2. Checks for contamination
 3. Confirms taxa ID
 4. Performs sequence typing
 5. Assembles reads into scaffolds
 6. Detects antibiotic resistance & hypervirulence genes
 7. Searches for plasmid markers
- PHoeNlx generates several files that are compatible with downstream analytic tools, such as those used for phylogenetic tree building
- This pipeline is available to run on Terra, Nextflow tower, CLI and is also incorporated into the StaPH-B toolkit

Dependencies & Install

- Install Nextflow with conda-forge channel and a bioconda channel using mamba
- Configure Nextflow for command-line interface (CLI) users
- Install container software
- Transfer Kraken2 database files to a new directory using WinSCP
- Edit configuration files to run on HPG

Install Nextflow

Install Nextflow environment using mamba:

1. Activate conda/mamba

```
$ module load conda
```

2. Install Nextflow

```
$ mamba create -yp /blue/<bphl-state>/<user>/conda_envs/nextflow -c conda-forge -c bioconda nextflow=21.10.6
```

3. Activate the mamba environment

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

4. Then, run PHoeNix from this environment

Configuring Nextflow for CLI users

Configuration is done in the form of a config file & is needed so that Nextflow knows how to fetch the required software. Options include:

- Pipeline comes with config files called **docker** & **singularity** which instruct the pipeline to use the named tool for software management. For example, **-profile test,singularity**
- Check nf-core/configs to see if a custom config file already exists for your institute, so you can use **-profile <institute>** in your command which enables either docker or singularity to set the execution for your local compute environment
- Use **--singularity_pull_docker_container** if you have issues downloading Singularity images. This will pull and convert the docker image instead.
- Alternatively, you can use **nf-core download** to download images first, before running the pipeline.
 - Set **NXF_SINGULARITY_CACHEDIR** or **singularity.cachedir** for Nextflow to store & reuse the images from a central location for future pipelines

Configuration set up

To add NXF_SINGULARITY_CACHEDIR to your bash profile run the following:

- Open your ~/.bash_profile

```
$ emacs ~/.bash_profile
```

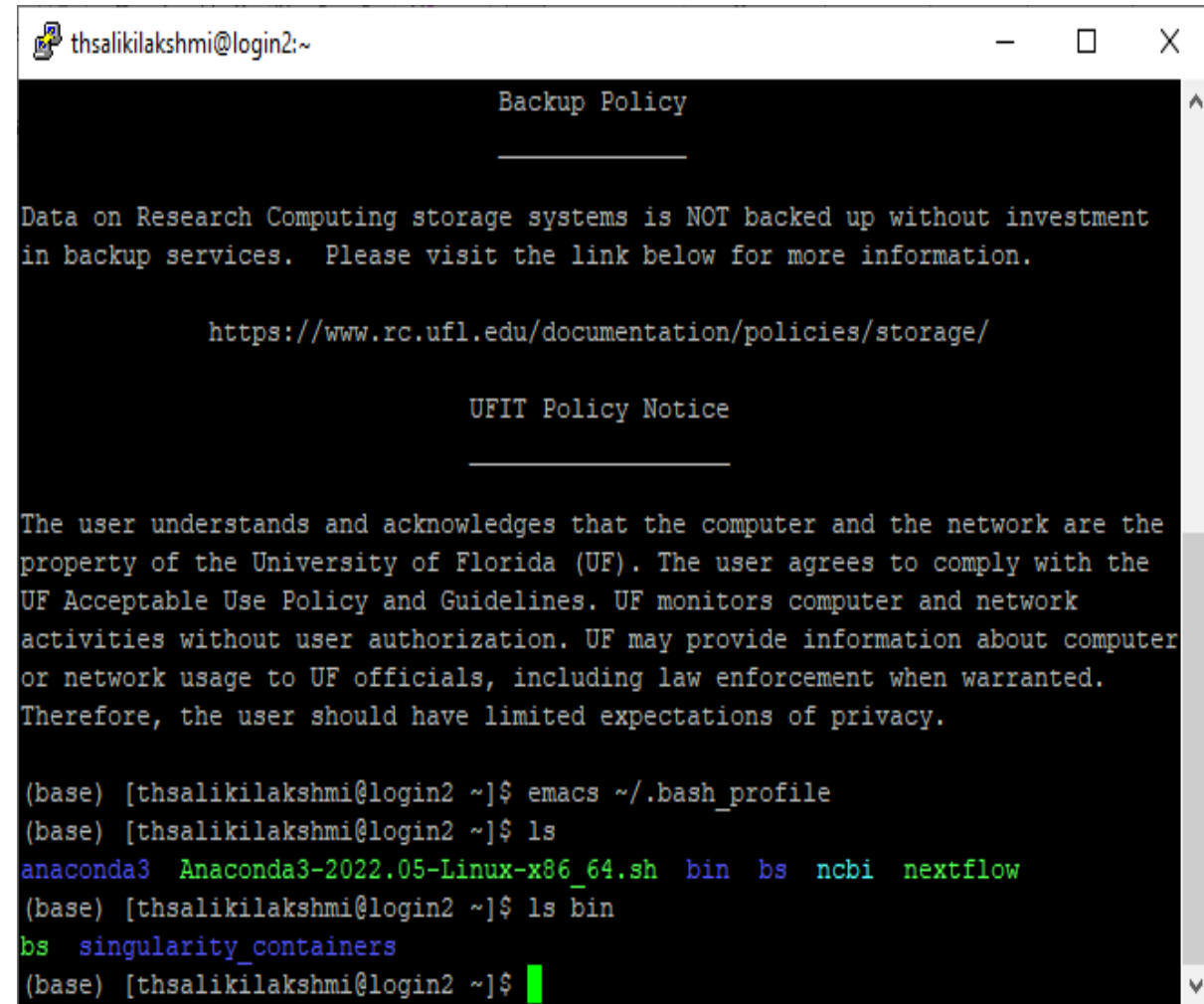
- Add these lines within ~/.bash_profile

```
$ export NXF_SINGULARITY_CACHEDIR=$PATH/Singularity_containers
export /path/to/singularity_containers
```

- \$ /path/ is the full path of the folder where you want to store. (**Singularity_Containers** can be named whatever you want)

```
$ source ~/.bash_profile
```

- This command allows Nextflow to see the new path



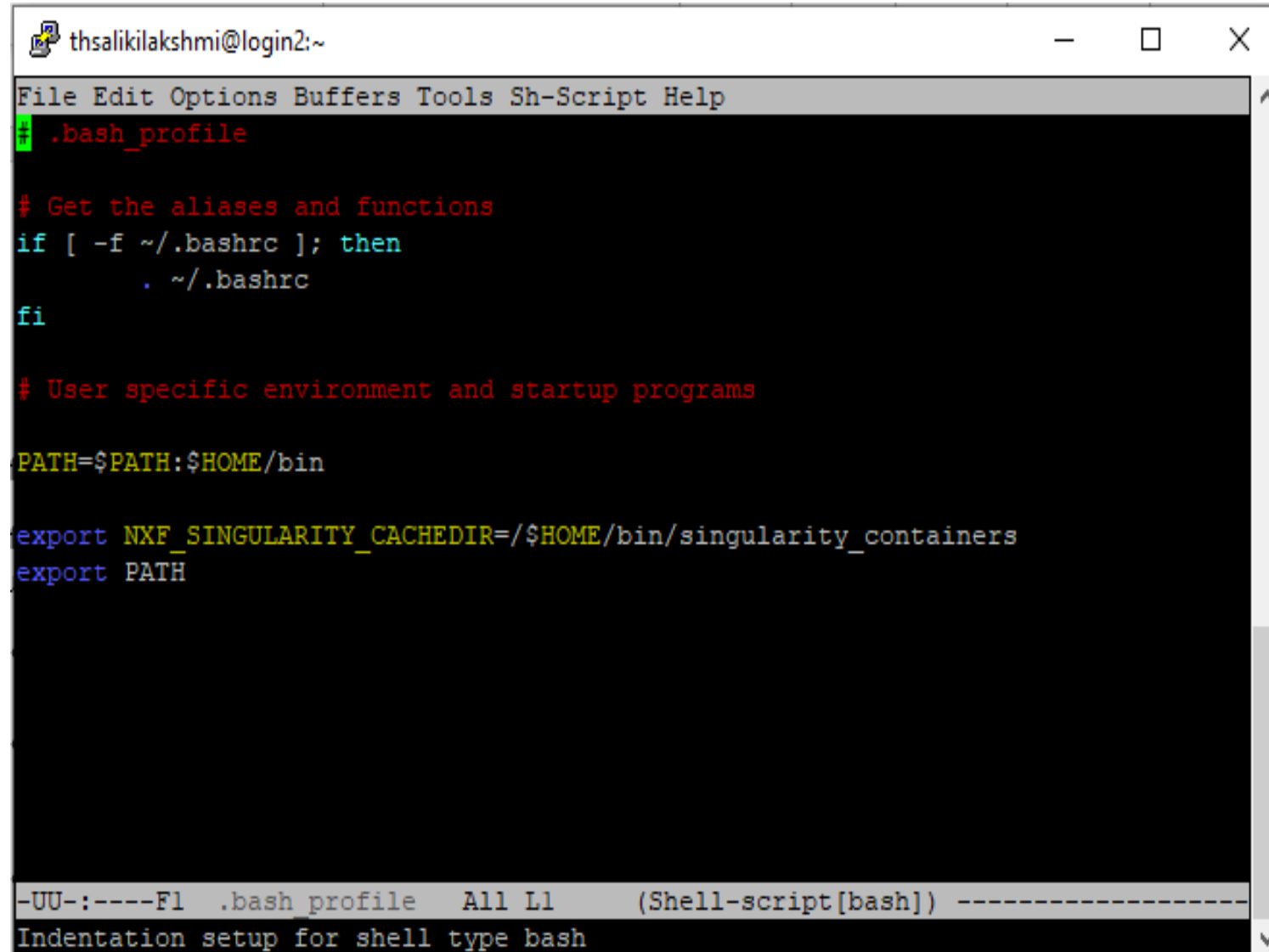
```
thsalikilakshmi@login2:~
Backup Policy
-----
Data on Research Computing storage systems is NOT backed up without investment
in backup services. Please visit the link below for more information.

https://www.rc.ufl.edu/documentation/policies/storage/

UFIT Policy Notice
-----
The user understands and acknowledges that the computer and the network are the
property of the University of Florida (UF). The user agrees to comply with the
UF Acceptable Use Policy and Guidelines. UF monitors computer and network
activities without user authorization. UF may provide information about computer
or network usage to UF officials, including law enforcement when warranted.
Therefore, the user should have limited expectations of privacy.

(base) [thsalikilakshmi@login2 ~]$ emacs ~/.bash_profile
(base) [thsalikilakshmi@login2 ~]$ ls
anaconda3  Anaconda3-2022.05-Linux-x86_64.sh  bin  bs  ncbi  nextflow
(base) [thsalikilakshmi@login2 ~]$ ls bin
bs  singularity_containers
(base) [thsalikilakshmi@login2 ~]$
```

emacs ~/.bash_profile



```
thsalikilakshmi@login2:~
File Edit Options Buffers Tools Sh-Script Help
~/.bash_profile

# Get the aliases and functions
if [ -f ~/.bashrc ]; then
    . ~/.bashrc
fi

# User specific environment and startup programs

PATH=$PATH:$HOME/bin

export NXF_SINGULARITY_CACHEDIR=$HOME/bin/singularity_containers
export PATH

-UU-:----F1 .bash_profile All L1 (Shell-script[bash]) ----
Indentation setup for shell type bash
```

Download Database Files for Kraken2

- Email HAISeq@cdc.gov, with the subject line "**krakenDB invite request**" to request access to the sharefile link & provide the email address to send invite to
- Then, download the **hash.k2d**, **opts.k2d**, and **taxo.k2d** files needed for the Kraken2 subworkflow of PHoeNIx from the CDC sharefile link
- Other kraken2 databases cannot be used – there is a specific **ktax_map.k2** file needed for the pipeline
- Download via WinSCP
- Once downloaded, the folder containing these files should be passed to PHoeNIx via by **--kraken2db**

We've shared the database files via HPG public-share

Git Clone PHoeNix

Install the latest version via cloning PHoeNix GitHub repo into a folder of your choosing

```
$ cd $/path/to/local/phoenix/repo/  
$ git clone https://github.com/CDCgov/phoenix
```

Testing Install

- Test that the pipeline is installed and configured correctly:

```
$ nextflow run phoenix/main.nf --profile test,singularity --entry  
PHOENIX --kraken2db $PATH_TO_DB
```

- This command will run the pipeline on preloaded data
- If all goes well, the output will look like the screen on the right
- As seen from the image, pipeline takes **~21 mins**
- Note: some steps aren't run in this pipeline (i.e., SPADES_WF stats), this is normal.

```
thsalikilakshmi@login5:/blue/bphl-florida/thsalikilakshmi/training
[bb/5232b5] process > PHOENIX:PHOENIX_EXTERNAL:BRMAP_REFORMAT (Test_Sample) [100%] 1 of 1 â
[8f/74bbc8] process > PHOENIX:PHOENIX_EXTERNAL:MLST (Test_Sample) [100%] 1 of 1 â
[a3/a0512c] process > PHOENIX:PHOENIX_EXTERNAL:GAMMA_HV (Test_Sample) [100%] 1 of 1 â
[98/2ad711] process > PHOENIX:PHOENIX_EXTERNAL:GAMMA_AR (Test_Sample) [100%] 1 of 1 â
[d9/c70240] process > PHOENIX:PHOENIX_EXTERNAL:GAMMA_PF (Test_Sample) [100%] 1 of 1 â
[75/4ed988] process > PHOENIX:PHOENIX_EXTERNAL:QUAST (Test_Sample) [100%] 1 of 1 â
[e2/a6de1b] process > PHOENIX:PHOENIX_EXTERNAL:KRAKEN2_WTASMBLD:KRAKEN2_WTASMBLD (Test_Sample) [100%] 1 of 1 â
[90/999cf1] process > PHOENIX:PHOENIX_EXTERNAL:KRAKEN2_WTASMBLD:KRAKENTOOLS_MAKEKREPORT (Test_Sample) [100%] 1 of 1 â
[1d/55250a] process > PHOENIX:PHOENIX_EXTERNAL:KRAKEN2_WTASMBLD:KREPORT2KRONA_WTASMBLD (Test_Sample) [100%] 1 of 1 â
[fd/e217c8] process > PHOENIX:PHOENIX_EXTERNAL:KRAKEN2_WTASMBLD:KRAKEN2_BH_WTASMBLD (Test_Sample) [100%] 1 of 1 â
[0d/ba51cf] process > PHOENIX:PHOENIX_EXTERNAL:KRAKEN2_WTASMBLD:KRONA_KTIMPORTTEXT_WTASMBLD (Test_Sample) [100%] 1 of 1 â
[df/9befae] process > PHOENIX:PHOENIX_EXTERNAL:MASH_DIST (Test_Sample) [100%] 1 of 1 â
[22/7f9290] process > PHOENIX:PHOENIX_EXTERNAL:DETERMINE_TOP_TAXA (Test_Sample) [100%] 1 of 1 â
[3a/c37688] process > PHOENIX:PHOENIX_EXTERNAL:FASTANI (Test_Sample) [100%] 1 of 1 â
[cd/11b1d6] process > PHOENIX:PHOENIX_EXTERNAL:FORMAT_ANI (Test_Sample) [100%] 1 of 1 â
[2d/841f2a] process > PHOENIX:PHOENIX_EXTERNAL:DETERMINE_TAXA_ID (Test_Sample) [100%] 1 of 1 â
[77/da8448] process > PHOENIX:PHOENIX_EXTERNAL:PROKKA (Test_Sample) [100%] 1 of 1 â
[21/8bb531] process > PHOENIX:PHOENIX_EXTERNAL:AMRFINDERPLUS_UPDATE (update) [100%] 1 of 1 â
[83/2777d1] process > PHOENIX:PHOENIX_EXTERNAL:GET_TAXA_FOR_AMRFINDER (Test_Sample) [100%] 1 of 1 â
[7b/f699a1] process > PHOENIX:PHOENIX_EXTERNAL:AMRFINDERPLUS_RUN (Test_Sample) [100%] 1 of 1 â
[6e/alc3f4] process > PHOENIX:PHOENIX_EXTERNAL:CALCULATE_ASSEMBLY_RATIO (Test_Sample) [100%] 1 of 1 â
[f3/5854fa] process > PHOENIX:PHOENIX_EXTERNAL:GENERATE_PIPELINE_STATS_WF:GENERATE_PIPELINE_STATS (Test_Sample) [100%] 1 of 1 â
[13/dd23cc] process > PHOENIX:PHOENIX_EXTERNAL:CREATE_SUMMARY_LINE (Test_Sample) [100%] 1 of 1 â
[e4/42dd47] process > PHOENIX:PHOENIX_EXTERNAL:FETCH_FAILED_SUMMARIES [100%] 1 of 1 â
[c9/011ce8] process > PHOENIX:PHOENIX_EXTERNAL:GATHER_SUMMARY_LINES (1) [100%] 1 of 1 â
[f5/b965e6] process > PHOENIX:PHOENIX_EXTERNAL:CUSTOM_DUMPSOFTWAREVERSIONS (1) [100%] 1 of 1 â
[70/3f52ae] process > PHOENIX:PHOENIX_EXTERNAL:MULTIQC [100%] 1 of 1 â
~[cdcgov/phoenix] Pipeline completed successfully~

***Reminder: If your lab has received funds (ELC, SHARP, etc.) to sequence isolates under the HAI/AR component of the AR Lab Network, please upload relevant sequence files to CDC's NCBI HAI-Seq Umbrella Project (ID S31911) and update any relevant alerts records with the HAI WGS ID and SRR ID, within 7-10 business days from sequencing completion.***

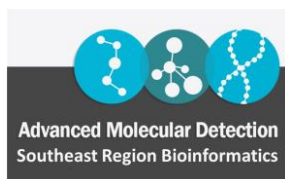
WARN: To render the execution DAG in the required format it is required to install Graphviz -- See http://www.graphviz.org for more info.
Completed at: 31-Jan-2023 08:24:08
Duration : 27m 48s
CPU hours : 0.8
Succeeded : 41

[/blue/bphl-florida/thsalikilakshmi/training/conda_envs/nextflow] [thsalikilakshmi@c0704a-s2 training]$ nextflow run cdcgov/phoenix -r vl.0.0 --profile test,singularity --entry PHOENIX --kraken2db /blue/bphl-florida/thsalikilakshmi/training/phoenix_testing/
```

Running PHoeNix

The help command will display required and optional options

```
nextflow run /blue/<bphl-state>/<user>/phoenix/phoenix help
```



```
-----  
PHOENIX  
cdcgov/phoenix v1.0.0  
-----  
Typical pipeline command:  
  
nextflow run cdcgov/phoenix -r $version -entry PHOENIX -profile <singularity,docker,test,custom> --input samplesheet.csv --kraken2db $PATH_TO_DB  
  
Required Options  
--input [string] Path to comma-separated file containing information about the samples in the experiment.  
--kraken2db [string] FULL PATH for where kraken2_db files are, which includes hash.k2d, taxo.k2d, opts.k2d and ktax_map.k2 files.  
[default: ${baseDir}/assets/databases/]  
  
Optional options  
--outdir [string] Path to the output directory where the results will be saved. MUST BE A FULL PATH. [default: ${PWD}/results]  
--email [string] Email address for completion summary.  
--multiqc_title [string] MultiQC report title. Printed as page header, used for filename if not otherwise specified.  
--busco_db_path [string] Path for where BUSCO files are if you want to run in offline mode. Use only with -entry CDC_PHOENIX  
[default: Not to run in offline mode.]  
--publish_dir_mode [string] null [default: copy]  
--singularity_pull_docker_container [string] null [default: false]  
  
Max job request options  
--max_cpus [integer] Maximum number of CPUs that can be requested for any single job. [default: 16]  
--max_memory [string] Maximum amount of memory that can be requested for any single job. [default: 128.GB]  
--max_time [string] Maximum amount of time that can be requested for any single job. [default: 240.h]
```

Inputs

- PHoeNix runs only on Illumina paired-end reads
- Multiple samples can be analyzed using a samplesheet.csv file

```
nextflow run cdcgov/phoenix -profile singularity -entry PHOENIX --input samplesheet.csv --kraken2db $PATH_TO_DB
```

Sample Sheet Input:

```
--input '[path to sample sheet file]'
```

- Sample sheet to be created with the information about the samples to be analyzed before running the pipeline
- Use **--input** parameter to specify the location

Sample Sheet Input

- Sample sheet must be a comma-separated file (.csv) with only **1** column and a header row
- **DO NOT HAVE ANY SPACES IN THIS FILE**
- Make sure the paths are full paths and not relative
- Automated sample sheet also can be created
 - This will be shown in a few slides
- Final sample sheet file consisting of paired-end data looks something like the image on the right

```
sample,fastq_1,fastq_2
JBI22000743,/blue/bphl-
florida/thalikilakshmi/data/HAI/20220829_jax_220823_PLN_WLK_MS/fastqs/JBI22000743_1.fastq.gz,/blue/b
phl-florida/thalikilakshmi/data/HAI/20220829_jax_220823_PLN_WLK_MS/fastqs/JBI22000743_2.fastq.gz
JBI22000770,/blue/bphl-
florida/thalikilakshmi/data/HAI/20220829_jax_220823_PLN_WLK_MS/fastqs/JBI22000770_1.fastq.gz,/blue/b
phl-florida/thalikilakshmi/data/HAI/20220829_jax_220823_PLN_WLK_MS/fastqs/JBI22000770_2.fastq.gz
JBI22000771,/blue/bphl-
florida/thalikilakshmi/data/HAI/20220829_jax_220823_PLN_WLK_MS/fastqs/JBI22000771_1.fastq.gz,/blue/b
phl-florida/thalikilakshmi/data/HAI/20220829_jax_220823_PLN_WLK_MS/fastqs/JBI22000771_2.fastq.gz
```

Sample Sheet Description

Column	Description
sample	Custom sample name. This entry will be identical for multiple sequencing libraries/runs from the same sample. Spaces in sample names are automatically converted to underscores (_)
fastq_1	Full path to <i>fastq</i> file for Illumina short reads 1. File has to be gzipped and have the extension ".fastq.gz" or ".fq.gz"
fastq_2	Full path to <i>fastq</i> file for Illumina short reads 2. File has to be gzipped and have the extension ".fastq.gz" or ".fq.gz"

Automated Sample Sheet Creation

- The sample sheet can be created automatically using the below command from a directory of *fastq* files

```
phoenix/bin/create_samplesheet.sh <directory of fastq files> > <output_directory>/samplesheet.csv
```

- The script will search 1 directory deep & attempt to determine sample id names and pairing/multilane information to create a sample sheet automatically
- **Please review the sample sheet for accuracy before using it in the pipeline**
- samplesheet.csv can be changed to your preferred name

Outputs

ANI

- Developed for fast alignment-free computation of whole-genome Average Nucleotide Identity (ANI)
- Avoids expensive sequence alignments and uses Mashmap as its MinHash based sequence mapping engine to compute the orthologous mappings and alignment identity estimates

AMRFinder

- AMRFinder and the accompanying database identify acquired antimicrobial resistance genes in bacterial protein and/or assembled nucleotide sequences as well as known resistance-associated point mutations for several taxa

Assembly

- SPAdes – St. Petersburg genome assembler – is an assembly toolkit containing various assembly pipelines
- SPAdes scaffold files are used for downstream analysis.

BUSCO (only run with --entry CDC_PHOENIX)

- BUSCO output is based on evolutionarily-informed expectations of gene content of near-universal single-copy orthologs, thus the BUSCO metric is complementary to technical metrics like N50

Outputs

fastp

- A tool designed to provide fast all-in-one preprocessing (trimming) of *fastq* files. This tool was developed in C++ with multithreading supported to afford high performance

FastQC

- Provides general quality metrics about your sequenced reads
- Provides information about the quality score distribution across your reads, per base sequence content (%A/T/G/C), adapter contamination, and overrepresented sequences

GAMMA

- GAMMA (Gene Allele Mutation Microbial Assessment) is a command line tool that finds gene matches in microbial genomic data using protein coding (rather than nucleotide) identity then translates and annotates the match by providing the type (i.e., mutant, truncation, etc.) and a translated description (i.e., Y190S mutant, truncation at residue 110, etc.)

Kraken2

- A taxonomic classifier using exact k-mer matches to achieve high accuracy and fast classification speeds
- This classifier matches each k-mer within a query sequence to the lowest common ancestor (LCA) of all genomes containing the given k-mer

Outputs

MLST

- Scans assembly files against traditional PubMLST typing schemes

QUAST

- Evaluates the quality of genome assemblies

removedAdapters

- BBDUK was developed to combine most common data-quality-related trimming, filtering, and masking operations into a single high-performance tool

SRST2 (only run with --entry CDC_PHOENIX)

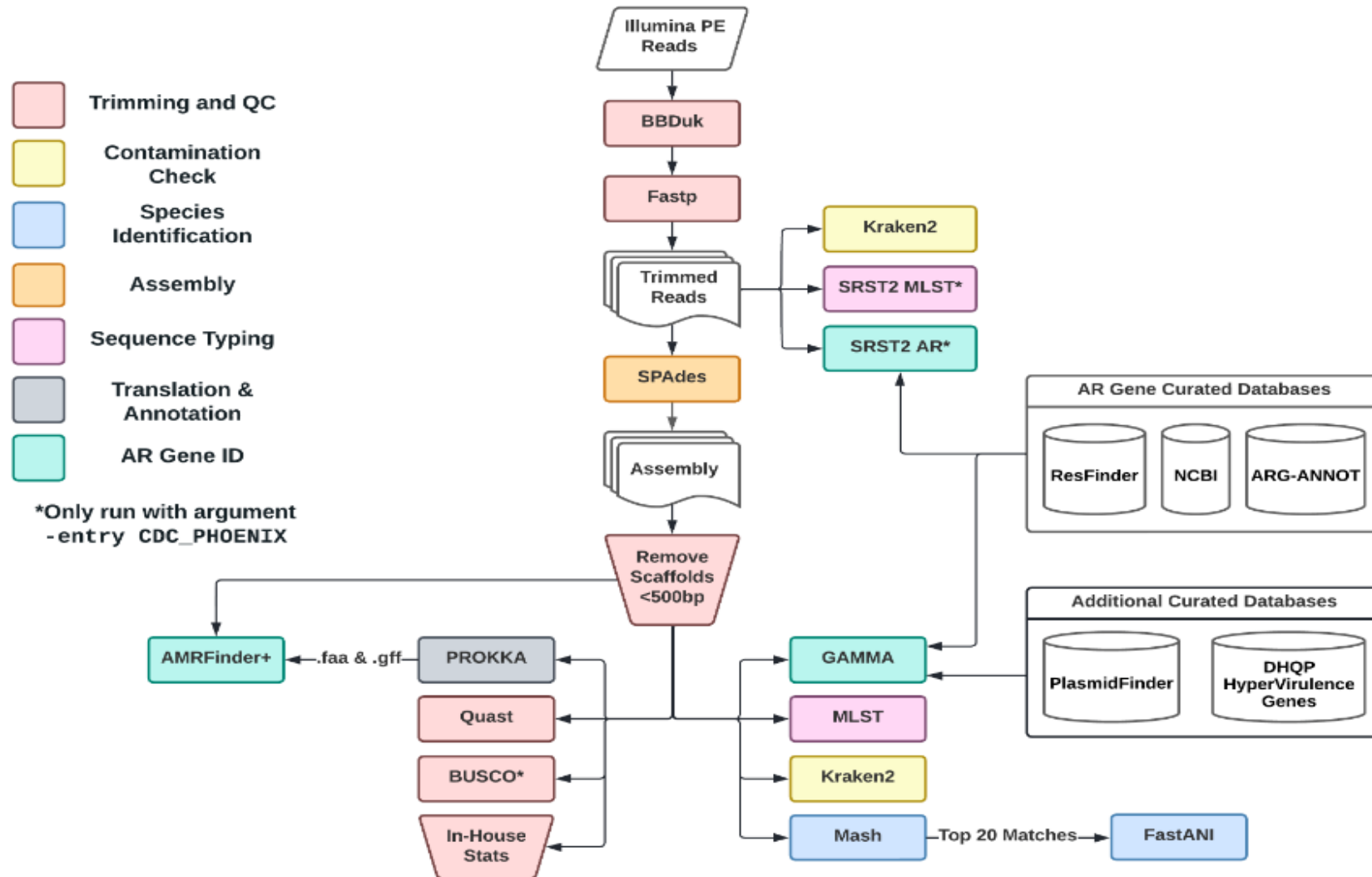
- Short Read Sequence Typing for Bacterial Pathogens

MultiQC

- A visualization tool that generates a single HTML report summarizing all samples in your project
- Results generated by MultiQC collate pipeline QC from supported tools e.g., FastQC

Pipeline Summary

PHoeNix v1.0.0 Workflow



Results

AutoSaveOff

Book1 - Excel

Search (Alt+Q)

Thsaliki, Lakshmi

FileHomeInsertPage LayoutFormulasDataReviewViewHelp

PasteCutCopyFormat Painter

Clipboard

Calibri11A^A^

Font

BBIU

Font

Alignment

General

Number

Styles

Cells

Editing

Sensitivity

G6

22

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	ID	Auto_QC	Warning	Estimated	Genome	Assembly	#_of_Scaff	GC_%	Species	Taxa_Conf	Taxa_Sour	Kraken2	Kraken2	MLST_Sch	MLST_1	MLST_Sch	MLST_2	GAMMA	GAMMA	(AMRFinde	Hyperviru	Plasmid	Auto_QC	Failure_Reas
2	JB1220012	PASS	0	153.1	3989752	0.9913 (.24	63	47.48	Vibrio ch	98.20% A	ANI_REFS	Vibrio(98	Vibrio(10	vcholerae	-	-	-	varG_NG	almG_AE003852.1	catB9_NG_047621.1				
3	JB1220013	PASS	0	71.31	5004923	0.9954 (.05	36	46.88	Vibrio vul	98.19% A	ANI_REFS	Vibrio(96	Vibrio(99	vvulnificu	-	-	-							
4	JB1220013	PASS	0	53.08	5126957	0.9748 (.55	51	44.59	Vibrio alg	98.58% A	ANI_REFS	Vibrio(95	Vibrio(99	vibrio	-	-	-	blaCARB-4	catC_JPIP01000949.1					
5	JB1220013	PASS	0	50.31	5125348	0.9745 (.56	48	44.58	Vibrio alg	98.57% A	ANI_REFS	Vibrio(96	Vibrio(99	vibrio	-	-	-	blaCARB-4	catC_JPIP01000949.1					
6	JB1220013	PASS	0	130.7	1826008	0.9839 (.3	22	37.96	Haemoph	99.94% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST56	-	-		hmrM_L42023.1					
7	JB1220013	PASS	2	97.11	1834063	0.9882 (.24	20	38.12	Haemoph	97.58% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST567	-	-		hmrM_L42023.1					
8	JB1220013	PASS	0	129.7	2022970	1.0900 (.18	39	38.29	Haemoph	99.76% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST3	-	-	blaTEM-1	hmrM_L42023.1					
9	JB1220013	PASS	0	173.5	1974327	1.0638 (.13	75	38.45	Haemoph	99.99% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST3	-	-		hmrM_L42023.1					
10	JB1220013	PASS	0	147.7	1899522	1.0235 (.45	28	38.05	Haemoph	98.27% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST208	-	-		hmrM_L42023.1					
11	JB1220013	PASS	0	169.6	1868794	1.0069 (.14	104	38.09	Haemoph	99.96% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST124	-	-		hmrM_L42023.1					
12	JB1220013	PASS	0	63	4587284	1.0201 (.3	61	48.27	Vibrio nav	98.33% A	ANI_REFS	Vibrio(96	Vibrio(99	-	-	-	-							
13	JB1220013	PASS	0	39.15	5443211	0.9711 (.62	114	57.14	Klebsiella	99.92% A	ANI_REFS	Klebsiella	Klebsiella	klebsiella	ST17	-	-	ampH_CP	aac(6')-Ib_1_M21682,aadA1_NC	Col440I_1,pKPC-CAV1193_1,IncN				
14	JB1220013	PASS	2	42.93	4284141	1.0632 (.13	57	39.47	Proteus m	99.55% A	ANI_REFS	Proteus(8	Proteus(9	-	-	-	-	blaKPC-2	aac(6')-33_NG_047267.1,aadA1	IncC_1				
15	JB1220013	PASS	2	39.03	1815033	0.9780 (.46	23	37.94	Haemoph	99.79% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST1218	-	-		hmrM_L42023.1					
16	JB1220013	FAIL	2	25.51	7079404	1.0693 (.95	135	65.79	Pseudom	99.96% A	ANI_REFS	Pseudom	Pseudom	paeruginc	ST111	-	-	blaOXA-3	aac(6')-29;parC_S87L,gyrA_T83I				coverage_below_30(25	
17	JB1220013	PASS	1	52.49	7359583	1.1117 (.15	165	65.57	Pseudom	99.90% A	ANI_REFS	Pseudom	Pseudom	paeruginc	ST111	-	-	blaOXA-2	aac(6')-29;gyrA_T83I,parC_S87L					
18	JB1220013	PASS	0	127.4	1814680	0.9778 (.46	23	37.97	Haemoph	99.96% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST124	-	-		hmrM_L42023.1					
19	JB1220013	PASS	0	135.1	1835926	0.9892 (.22	28	38.01	Haemoph	99.95% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST124	-	-		hmrM_L42023.1					
20	JB1220014	PASS	0	149.4	1907038	1.0275 (.57	48	38.04	Haemoph	98.28% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST84	-	-		hmrM_L42023.1					
21	JB1220014	PASS	1	117.2	1889428	1.0181 (.37	31	38.01	Haemoph	99.77% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST107	-	-	blaTEM-1	hmrM_L42023.1					
22	JB1220014	FAIL	3	124.9	3341091	1.8002 (.16	1411	36.32	Haemoph	99.76% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	ST3	-	-	blaTEM-1	aac(6')-Ie_NG_047055.1,bleO	repUS12			STDev_above_2.58(16	
23	JB1220014	PASS	0	108.9	2089247	0.9749 (.27	75	51.87	Neisseria	99.18% A	ANI_REFS	Neisseria	Neisseria	(neisseria	ST2578	-	-	penA_1_A	farB_AE00 foIP_R228S					
24	JB1220014	FAIL	1	105.2	3171960	1.7091 (.14	1375	38.16	Haemoph	99.33% A	ANI_REFS	Haemoph	Haemoph	hinfluenz	-	-	-	blaTEM-1	hmrM_L42023.1					STDev_above_2.58(14

Sheet1



Link to PHoeNix

GitHub link for CDC PHoeNix pipeline: [Home · CDCgov/phoenix Wiki · GitHub](#)

Next Week

- We'll go over editing the HPC config file for running on HPG
- We'll have a quick demo and time for questions and troubleshooting

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?

bphl-sebioinformatics@flhealth.gov

Lakshmi Thsaliki, MS

Bioinformatician

Lakshmi.Thsaliki@flhealth.gov

Molly Mitchell, PhD

Bioinformatician

Molly.Mitchell@flhealth.gov

Sarah Schmedes, PhD

Bioinformatics Supervisor

Sarah.Schmedes@flhealth.gov