


## Department of Health

**Bioinformatics Regional Resources  
- Southeast Region -**



Sarah Schmedes, PhD, BRR Lead - Southeast Region  
Florida Bureau of Public Health Laboratories  
Jacksonville, FL  
December 6, 2019

---

---

---

---

---

---

---

---

## Outline

- CDC AMD Strategy
  - Bioinformatics resources
  - Workforce development
- Florida Bureau of Public Health Laboratories (FL BPHL)
- Southeast Bioinformatics Regional Resources (BRR)
  - BRR Lead
  - High Performance Computing (HPC)
  - Galaxy

2

---

---

---

---

---


---

---

---

## CDC AMD Strategy

- Increase and improve advanced molecular detection (AMD) capacity in state and local health departments
- Support and expand bioinformatics capabilities across the US
- Supported through 7 regions
  - Bioinformatics Regional Resource (BRR) Lead
  - Workforce Development (WFD) Lead (Training)



3

---

---

---

---

---

---


---

---

**Bioinformatics Regional Resource (BRR) Lead,  
Southeast Region**

Sarah Schmedes, PhD  
[Sarah.Schmedes@flhealth.gov](mailto:Sarah.Schmedes@flhealth.gov)  
(904) 791-1788

- Services:
  - Bioinformatics consultation
  - Pipeline development, data analysis support
  - Training support



4

---

---

---

---

---

---

---

---

**Bioinformatician (part-time),  
Southeast Region**

Taj Azarian, PhD  
Assistant Professor  
University of Central Florida  
[Taj.Azarian@ucf.edu](mailto:Taj.Azarian@ucf.edu)  
(407) 823-4682

- Services:
  - Bioinformatics consultation
  - Pipeline development, data analysis support

5

---

---

---

---

---

---

---

---

**Workforce Development (WFD)  
Lead, Southeast Region**

Tonia Parrott, PhD  
[Tonia.Parrott@dph.ga.gov](mailto:Tonia.Parrott@dph.ga.gov)  
(404) 327-7905

- Services:
  - Training
    - Workshops, courses
      - No cost to labs, including travel
  - Computational resources (training)

6

---

---

---

---

---

---

---

---

## NGS in Public Health

- Applications
  - Whole-genome sequencing
  - Targeted resequencing (enrichment/amplicon)
  - 16S and shotgun metagenomics
- Use cases
  - Is isolate-A related to Isolate-B?
  - Cluster/outbreak detection, surveillance
  - Antibiotic resistance or virulence factors of interest

7

---

---

---

---

---

---

---

---

## FL BPHL

### Molecular Division

- Team
  - 7 laboratory scientists
  - 1 bioinformatician
- Instrumentation
  - 4 MiSeqs, 2 iSeqs
- Sequencing application
  - PulseNet, ~5,000 isolates/year
  - Whole-genome sequencing (bacteria)
  - Targeted amplicon sequencing (viruses)



8

---

---

---

---

---

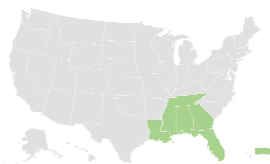
---

---

---

## Southeast Regional Resources

- Bioinformatics resources
  - BRR Lead
  - High Performance Computing (HPC)
    - Computational resources
    - Data storage
  - Local Galaxy instance



9

---

---

---

---

---

---

---

---

## Bioinformatics Regional Resource (BRR) Lead

### Consultation

- Public health bioinformatics
- Bioinformatics concepts
- Data management
- Data analysis strategy and implementation

### Data Analysis/Pipeline Development

- WGS, targeted-amplicon seq, metagenomics
- Linux and command-line
- Coding – Bash, Perl, R, Python
- Manage shared resources

### Training Support

- One-on-One
- Webinars
- Site visits

10

---

---

---

---

---

---

---

---

## Bioinformatics Regional Resource (BRR) Outreach

- Quarterly calls
- [Sarah.Schmedes@flhealth.gov](mailto:Sarah.Schmedes@flhealth.gov)
- (904) 791-1788
- Site visits, as needed
- Training support (in attendance)

11

---

---

---

---

---

---

---

---

## High Performance Computing (HPC)

- What is High Performance Computing (HPC)?
  - "Many computers all working together to provide higher performance"
- University of Florida's "HiPerGator"
  - UF Research Computing <https://www.rc.ufl.edu/about/>
  - " #3 Most power computer at a U.S. public university"
- HiPerGator Stats
  - 51,000 cores
  - 3 petabytes (PB) high-performance storage
  - Several other servers for additional services (i.e., Galaxy...)

12

---

---

---

---

---

---

---

---

## HPC for Public Health use

- What is the advantage of using HPC?
  - 1) Pay for resources you need (modify at any time) – **Free to states!!!!**
    - Large storage
    - Large number of CPUs and RAM
  - 2) Linux-based system
    - Required for most open-source bioinformatics tool (command-line)
  - 3) Tech support and system maintenance
  - 4) Access to pre-installed software
    - “No need to go through IT”!!!
  - 5) Direct data transfer from basespace

13

## HiPerGator

### Command-line



### Graphical User Interface (GUI) - Galaxy



14

## HiPerGator

### Command-Line

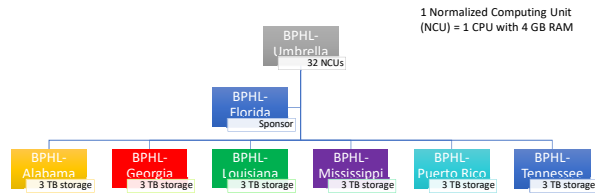
- Login through a terminal
- Access to HiPerGator storage
  - Transfer files via scp, sftp, directly from basespace, or NCBI
- Access to > 1,300 pre-installed applications
  - > 600 biological software
  - NCBI blast databases
- Tech and installation support

### Graphical User Interface (GUI) - Galaxy

- Login through web browser
- “Point and click” data analysis (no programming required)
- Upload data from local machine, via FTP, or from HiPerGator storage
- Access to the most widely used bioinformatics tools
- Data sharing
- Generate workflows (and share)

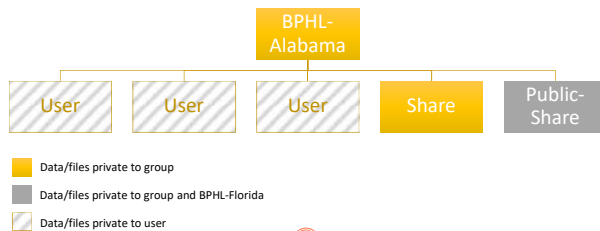
15

## HiPerGator – State groups



16

## State Group



17

## Share directory

- Private directory for state use only
- Public-share with BPHL-FL
  - State
    - Share select data/files
  - BPHL-FL
    - Share scripts and pipelines
    - Help troubleshoot by viewing files or data
    - Analyze select data
- Share workflows and data directly on HiPerGator
  - No need to mess with ftp sites or external drives!!!!

18

## Accessing HiPerGator - Galaxy

- GatorLink ID/HiPerGator account
  - Will send out a hand-out with instructions for user account set-up
- Login through internet browser
- Future webinar for in-depth details and walk-through of Galaxy

19

---

---

---

---

---

---

---

---

## Galaxy

- HiPerGator Galaxy instance
  - Use NCUs to run analyses
- Graphical User Interface (i.e., “click-and-point”)
- Use individual tools
- Create workflows
- Share workflows
- Share data

20

---

---

---

---

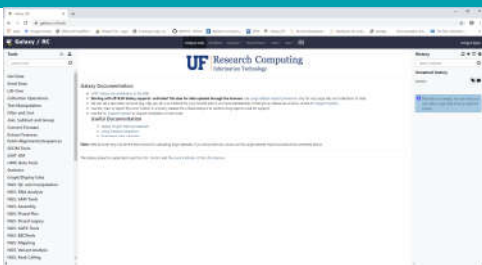
---

---

---

---

## Galaxy



21

---

---

---

---

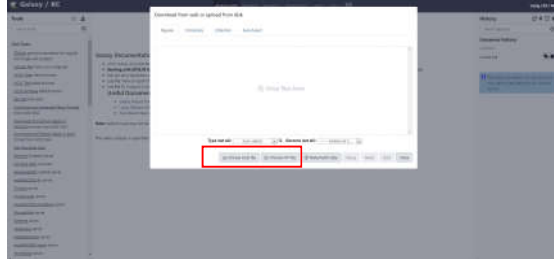
---

---

---

---

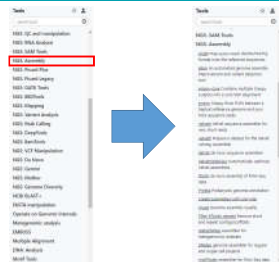
## Data upload



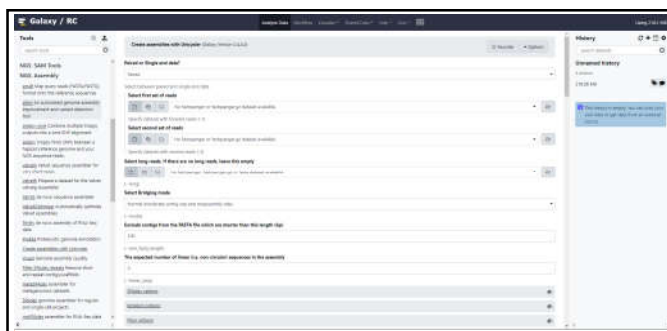
22

## Galaxy Tools

- Most commonly used NGS tools
- Submit installation request for new tools



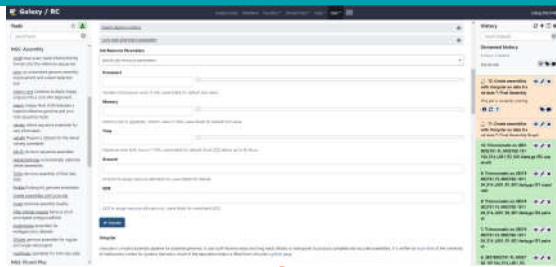
23



24



## Specify Job Resources



25

## Galaxy Workflows



26

## Accessing HiPerGator – command line

- Windows
  - Connect to HiPerGator (ssh client)
    - Putty or MobaXterm (free software)
  - File transfers (sftp)
    - WinSCP, FileZilla (free software)
- Linux or MacOS
  - Terminal (pre-installed)
    - Connecting and file transfers
- Future webinar for in-depth details and walk-through of HiPerGator

27

## HiPerGator

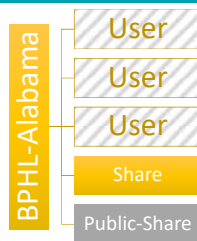
```

@schmuck@hyper2-
The user agrees to comply with Research Computing's policies.
https://www.rc.ufl.edu/services/procedures/
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/services/procedures/storage/protection/
OFFIT Policy Module
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
and network usage to UF officials, including law enforcement when warranted.
Therefore, the user should have limited expectations of privacy.
[User] [Enter] [Ctrl+C] [Exit]

```

28

## State Group



29

## Current analysis applications

- 600 pre-installed biological programs/software
- <https://github.com/CDCgov/> (additional tools than OAMD Portal)
- Custom scripts and pipelines
  - Automated data download from BaseSpace and NCBI databases
  - FASTQ quality control/trimming/filtering
  - De novo assembly
  - Read mapping
  - Variant detection (SNPs and Indels)
  - Species identification
  - Average nucleotide identity (ANI)
- Metagenomic classification
  - Abundance estimation
  - Contamination detection
- Gene annotation
- Pan-genome analysis
- Genome-wide SNP analysis
- Phylogenetic tree generation
- Phylogenetic tree annotation and editing

30

## Future applications/pipelines

- Antibiotic resistance gene detection
- Virulence gene detection
- Plasmid detection
- *E. coli* serotyping
- *Salmonella* serotyping

31

---

---

---

---

---

---

---

---

## Future Webinars

- HiPerGator - command line
- Galaxy Training
- Requests (???)

32

---

---

---

---

---

---

---

---

## Questions???

### Contacts:

Sarah Schmedes, PhD  
BRR Lead, Southeast Region  
[Sarah.Schmedes@flhealth.gov](mailto:Sarah.Schmedes@flhealth.gov)

Jason Blanton, PhD  
Molecular Administrator  
Whole Genome Sequencing Coordinator  
[Jason.Blanton@flhealth.gov](mailto:Jason.Blanton@flhealth.gov)

Tonia Parrott, PhD  
WFD Lead, Southeast Region  
[Tonia.Parrott@dph.ga.gov](mailto:Tonia.Parrott@dph.ga.gov)

Taj Azarian, PhD  
Bioinformatician (part-time)  
Assistant Professor, University of Central Florida  
[Taj.Azarian@ucf.edu](mailto:Taj.Azarian@ucf.edu)

33

---

---

---

---

---

---

---

---