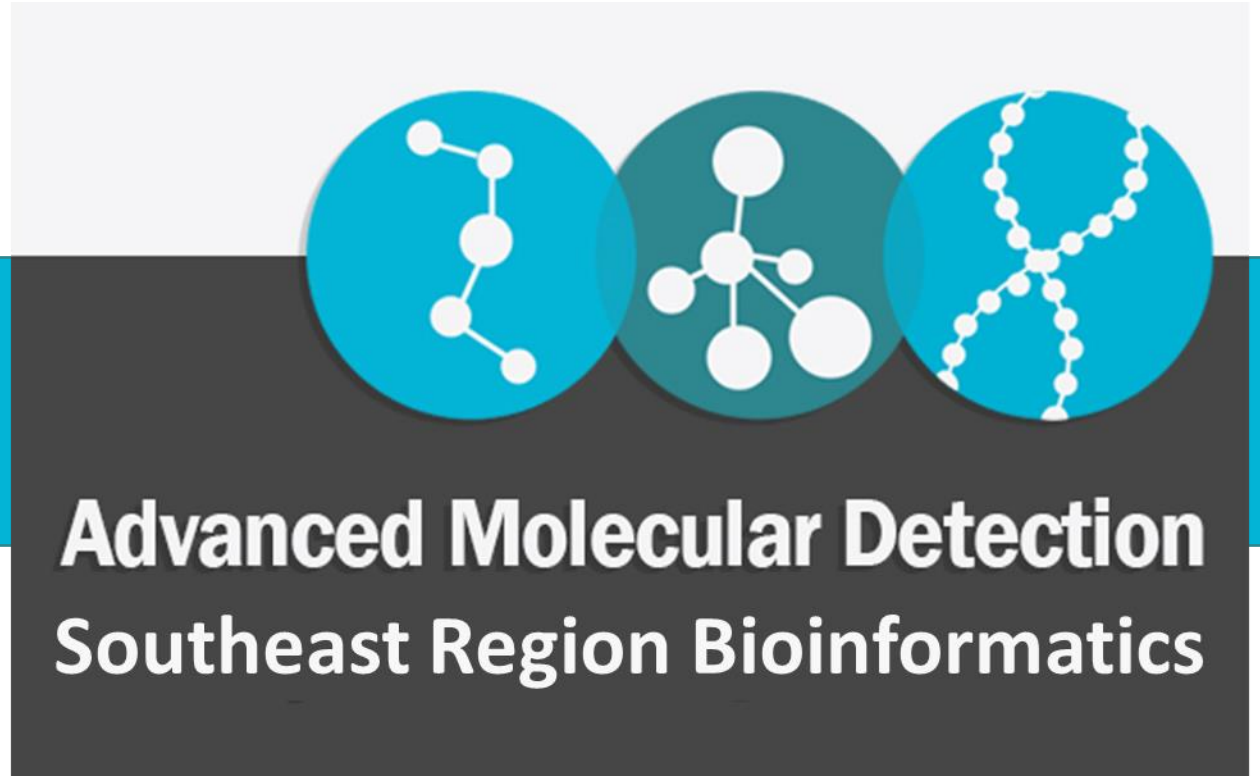



**Office Hours – Session 2**

03/20/2023

# Outline

Introductions

ELC Updates/Reminders

Agenda

What is Conda?

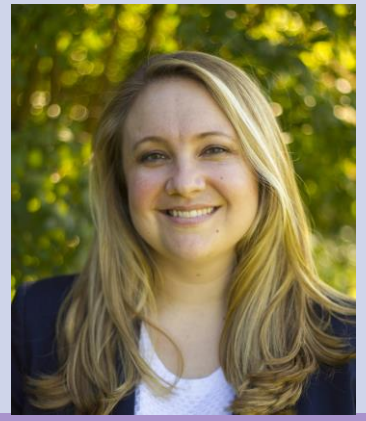
Importance of Conda

Downloading & Installing Conda

Creating & Managing Environments

Questions

# BRR Team



**Sarah Schmedes, PhD**
Lead Bioinformatician
Sarah.Schmedes@flhealth.gov

Molly Mitchell
Bioinformatician
Molly.Mitchell@flhealth.gov

**Lakshmi Thsaliki, MS**
Bioinformatician
Lakshmi.Thsaliki@flhealth.gov

For any BRR requests, please send an email to all three team members above (IT is fixing our group email)
Based on the requests one of us from the team will respond as soon as possible.

Advanced Molecular Detection
Southeast Region Bioinformatics

# 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.**

Advanced Molecular Detection
Southeast Region Bioinformatics

# Agenda

**March 20** - Conda
**April 3** – Nextflow
**April 17** – **Cancelled**
**April 24** - PHoeNIx #1 (Installation and Dependencies) (**Rescheduled**)
**May 1** – PHoeNIx #2 (Demo)

Future Trainings
- ONT & FL's Flisochar pipeline
- StaPH-B Toolkit Programs/Pipelines
- AMRFinder+ (--organism)
- GISAID flagged SARS-CoV-2
- Git (git clone, etc.)
- PHA4GE Harmonization (AMR)
- Generating R figures
- …and more

Advanced Molecular Detection
Southeast Region Bioinformatics

# What is Conda?

- Python-based environment & package manager
- Creates reproducible analysis pipelines using crowd-sourced & version-controlled packages
- Conda creates self-contained modules that contain the necessary programs, etc. for completing a particular computing task
- Reduces difficulties with typical tool/package installation

**Environment:**

- A computing environment which is the collection of programs, language libraries, etc. in which a computer operates

**Package:**

- A package is a collection of software that may contain things such as programs (e.g., Python), programming libraries (e.g., Perl), or other useful tools.
- Conda combines packages to construct environments for doing complex tasks

Advanced Molecular Detection
Southeast Region Bioinformatics

# Importance of Conda

- Works on all major operating systems like Mac, Windows, & Linux, though not all packages are available for every operating system
- Helps keep your computing environment organized so you're less likely to end up with dependency issues
- Packages are version-controlled so you can easily switch versions if one isn't working or if a particular version is required for a pipeline
- Environments can be reproduced and shared leading to increased analytical veracity
- Works exceptionally well when combined with workflow managers (e.g. Nextflow)

Advanced Molecular Detection
Southeast Region Bioinformatics

# Installing Anaconda / Manage Conda Environments

- Download the installation script

```
$ wget https://repo.anaconda.com/archive/Anaconda3-2022.05-Linux-x86_64.sh
```

- Type **ls –althr –color=auto** to confirm the file downloaded, which displays in white color (and does not have executable permissions for the owner/user
- To add executable permissions to the file, type

```
$ chmod u + x  Anaconda3-2022.05-Linux-x86_64.sh
```

- Type **ls –althr –color=auto**, color should be changed to green from white and you'll see an "x" for owner

Advanced Molecular Detection
Southeast Region Bioinformatics

# Installing Anaconda

- To install Anaconda, execute the shell script given below

```
$ ./ Anaconda3-2022.05-Linux-x86_64.sh
```

- Follow all the prompts, scroll through the license agreement and hit enter which should take a several minutes for the process to complete.
- Answer yes and/or hit enter for all the prompts
- Close Putty/shell terminal & login again to initialize your base conda environment

Advanced Molecular Detection
Southeast Region Bioinformatics

# Using HPG Conda

- Instead of installing anaconda, you can use the conda module in HPG

```
$ module load conda
```

- Creating and naming an environment (-yp ‘your path’ to where you want the conda env installed)

```
$ conda create –yp /blue/bphl-<state>/<user>/<conda_envs>/<env_name>
```

```
$ conda activate /blue/bphl-<state>/<user>/<conda_envs>/<env_name>
```

Advanced Molecular Detection
Southeast Region Bioinformatics

# Updating Conda and Adding Channels

- Make sure to update conda regularly

```
$ conda update conda
```

- Adding addition channels - may already be installed (must do this the first time you use so you can install tools)

```
$ conda config --add channels defaults
$ conda config --add channels bioconda
$ conda config --add channels conda-forge
```

Advanced Molecular Detection
Southeast Region Bioinformatics

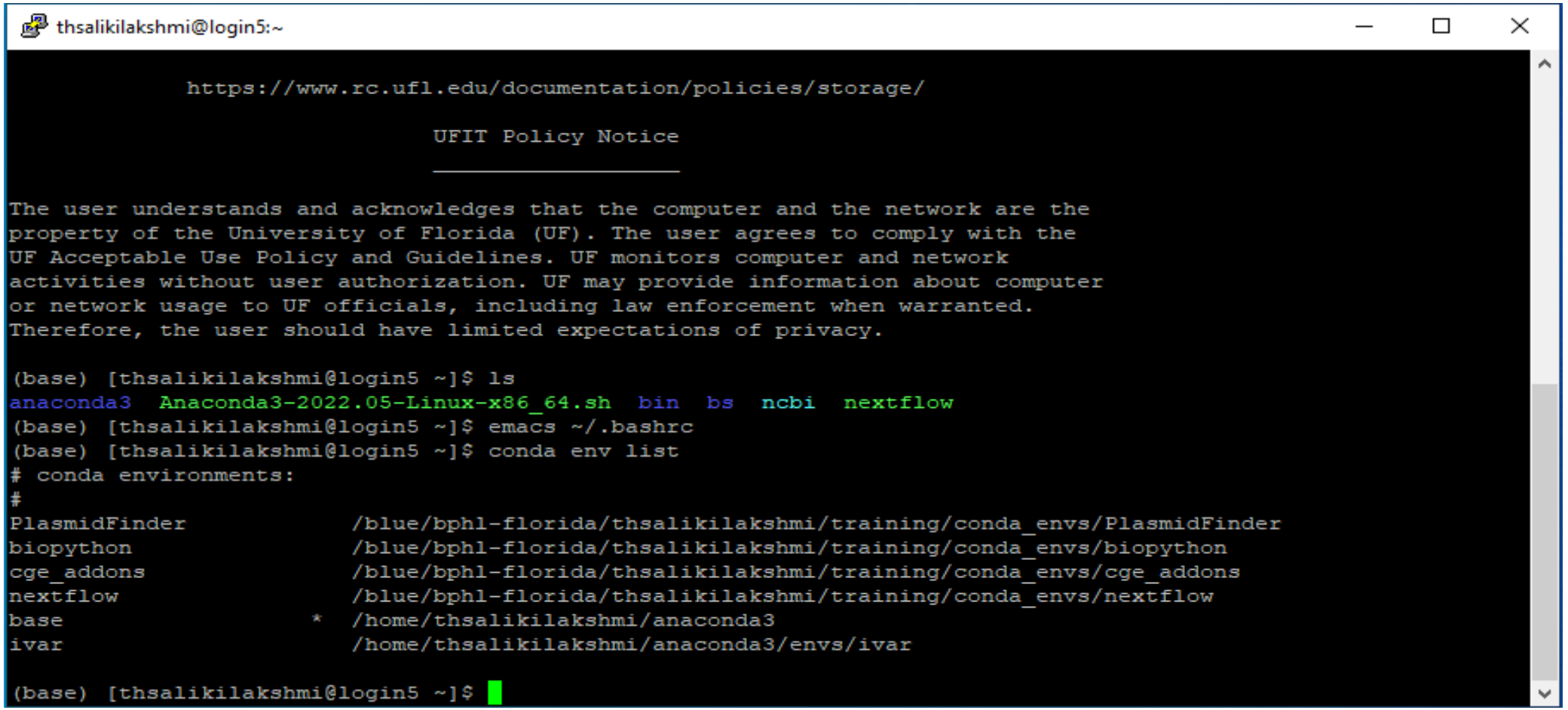
# Creating & Managing Conda Environments

- Environments are an integral part of conda-based workflows and also can be used to run individual programs in the command line
- They are customizable, reproducible, and shareable modules that contain the resources for a specific task or set of tasks
- Environments also help avoid dependency issues where required programs are incompatible with previously installed programs or program versions
- To display all the environments available along with their locations, type

```
$ conda env list
```

Advanced Molecular Detection
Southeast Region Bioinformatics

# Conda Environments List




Advanced Molecular Detection
Southeast Region Bioinformatics

# Conda Environments

- By default, an environment called base is created when installing & initializing Conda upon login. Base contains standard Python packages.
- To activate an environment in Conda, type

```
$ conda activate <environment name>
```

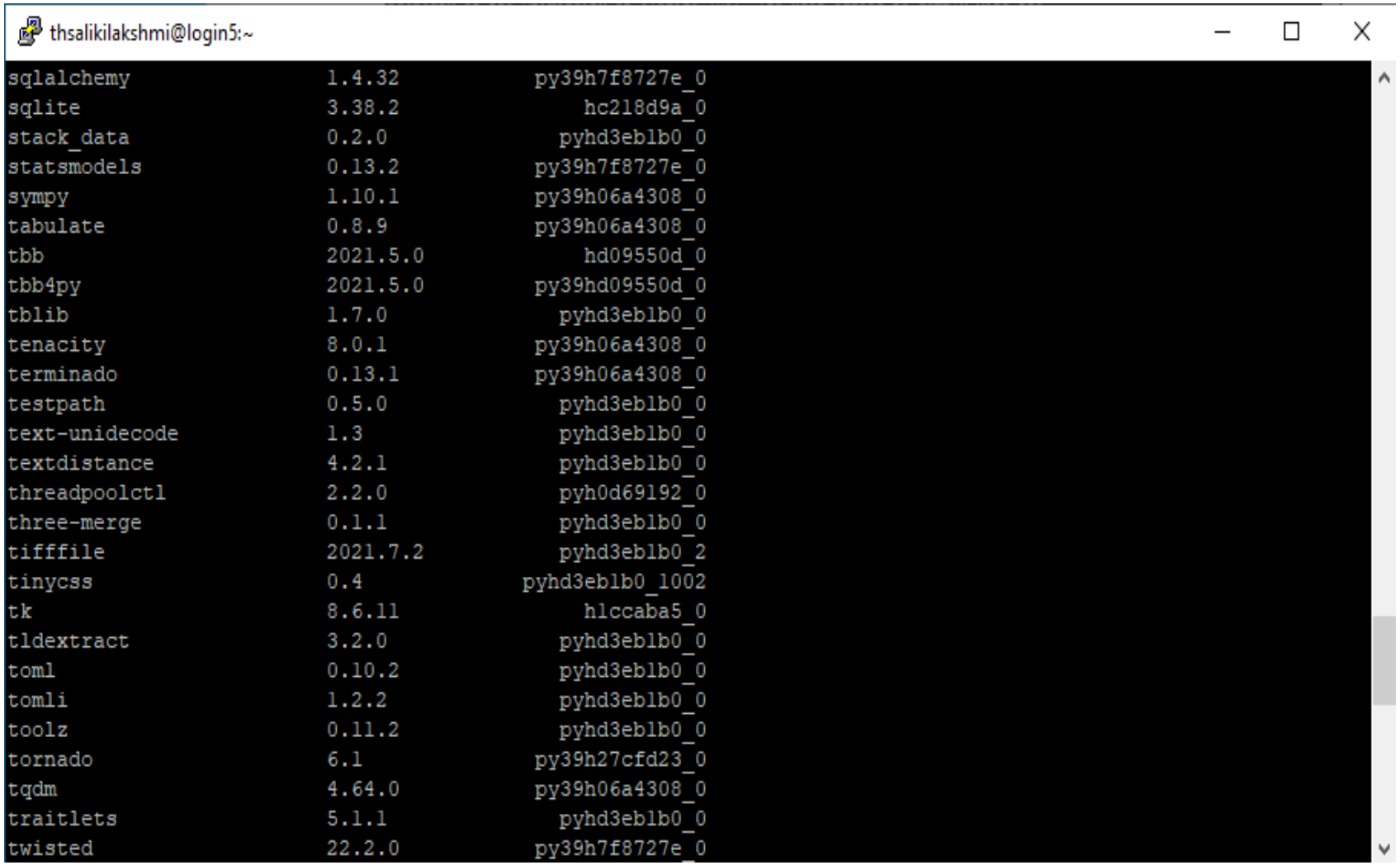
- As the base environment is loaded, to check what programs it contains, type

```
$ conda list
```

Advanced Molecular Detection
Southeast Region Bioinformatics

# Conda list

This screen appears which displays the name, version, and build channel




Advanced Molecular Detection
Southeast Region Bioinformatics

# Conda Environments

- To deactivate the environment, type

```
$ conda deactivate
```

- To create a new environment in conda, deactivate the current environment and then type

```
$ conda create -n <environment name>
```

- To delete an environment

```
$ conda remove --name <environment name> --all
```

- To delete an environment with an invalid path name (HPG conda module)

```
$ conda remove --all --prefix “/blue/bphl-<state>/<user>/conda_env/name_of_env”
```

Advanced Molecular Detection
Southeast Region Bioinformatics

# Mamba

- Reimplementation of conda written in C++
  - Faster friend
- Install your own version

```
$ conda install -c conda-forge mamba
```

- Or to use on HPG, load conda

```
$ module load conda
```

- Then, call mamba in place of conda

```
$ mamba create -n <environment name>
$ mamba activate <environment>
$ mamba install <dependencies> <package>
```

Advanced Molecular Detection
Southeast Region Bioinformatics

# Example: Shigatyper Environment

- Let's use shigatyper as an example (Shigatyper :: Anaconda.org)
  - "Quick and easy tool designed to determine Shigella serotype using Illumina or ONT reads"
- Create a shigatyper environment and install shigatyper with bioconda and conda-forge repository channel

```
$ conda create -n shigatyper -c conda-forge -c bioconda shigatyper
```

- Or the same environment can be made with the following three commands

```
$ conda create -n shigatyper

$ conda activate shigatyper

$ conda install -c conda-forge -c bioconda shigatyper
```

Advanced Molecular Detection
Southeast Region Bioinformatics

# References

GitHub - um-dang/conda_on_the_cluster: A brief tutorial introducing learners to the Conda environment and package management system

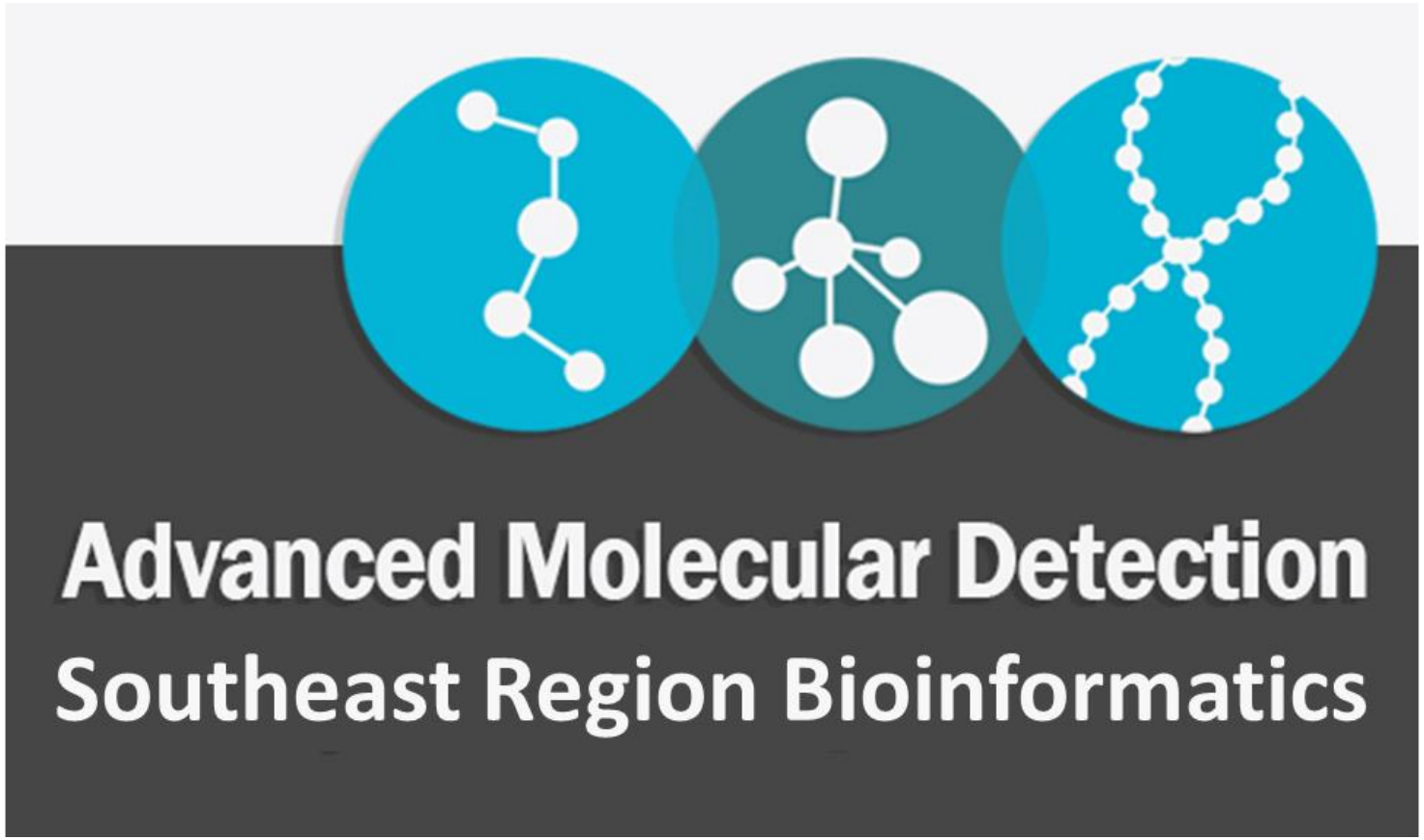
Conda cheat sheet conda-cheatsheet.pdf

Advanced Molecular Detection
Southeast Region Bioinformatics

# 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