



Advanced Molecular Detection

Southeast Region Bioinformatics



BaseSpace CLI

August 19, 2021

BPHL-SEbioinformatics@flhealth.gov

Outline



BaseSpace CLI



Installing BaseSpace CLI



Interacting with your data via BaseSpace CLI



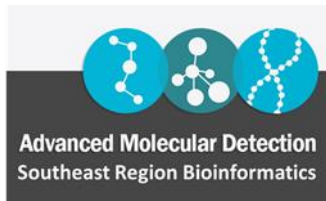
Data Transfers from BaseSpace to HiPerGator



Available pipelines and tools on HiPerGator

BaseSpace Command Line Interface (CLI)

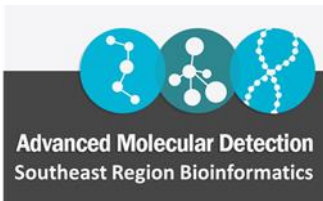
- Work with your data in BaseSpace via command-line from HiPerGator
- Overview and documentation
 - <https://developer.basespace.illumina.com/docs/content/documentation/cli/cli-overview>
- CLI examples/use cases
 - <https://developer.basespace.illumina.com/docs/content/documentation/cli/cli-examples>
 - Retrieve basic run and project information
 - Create biosamples, projects, etc.
 - Download files
 - Upload files
 -and more



Benefits of BaseSpace CLI

- Directly transfer data from BaseSpace to HiPerGator for analysis
 - Command line and Galaxy
- No need to download files locally and then reupload to HiPerGator
 - Avoid slow downloads to your local machine due to state firewalls

Advanced Molecular Detection
Southeast Region Bioinformatics



Install and Configure BaseSpace CLI

- Login to HiPerGator via Putty or a terminal
 - You will be in your home directory (/home/<user>/)
- Make a new directory, called “bin”

```
[usr@login]$ mkdir $HOME/bin  
[usr@login]$ ls
```

#If you see “bin”, then your new directory was successfully created.

- Type (or copy/paste) and execute the following:
 - Note – Control-C and Control-V do not work for copy/paste in Putty. To paste, right click once you’ve copied from your source. To copy, highlight text (that’s it).

```
[usr@login]$ wget "https://launch.basespace.illumina.com/CLI/latest/amd64-linux/bs" -O $HOME/bin/bs
```

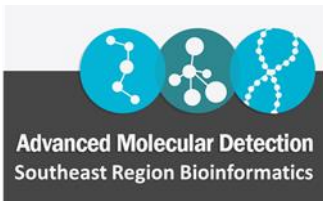
Install and Configure BaseSpace CLI

- Make the downloaded binary executable.

```
[usr@login]$ chmod u+x $HOME/bin/bs
```



Advanced Molecular Detection
Southeast Region Bioinformatics



Authenticate

- Authenticate your BaseSpace account

```
[usr@login]$ bs auth
```

- Follow the prompt by going to the displayed URL and follow the instructions. You will have to input your BaseSpace username and password.
 - If you receive a server error/warning, just wait and try again later.
 - You only have to authenticate once per Workgroup. If you change your BaseSpace password after authentication, the BaseSpace CLI will still continue to work.
 - If you have more than one Workgroup in your BaseSpace account, you need to authenticate for each Workgroup. Your default Workgroup will automatically be configured with the above command. To authenticate an additional workgroup, use the following command:

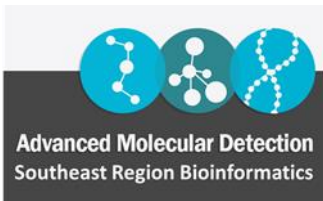
```
[usr@login]$ bs auth -c <name_of_second_non-default_workgroup>
```

Installation Complete!

- You can now start to work directly with your BaseSpace data and transfer data to HiPerGator.
- Go to <https://developer.basespace.illumina.com/docs/content/documentation/cli/cli-examples> to see a list of all commands and options for use cases.



Advanced Molecular Detection
Southeast Region Bioinformatics



BaseSpace CLI Commands and Use Case Examples

- bs list

```
[usr@login]$ bs list run --terse    #list run IDs
```

```
[usr@login]$ bs list project      #list project names and size in default workspace (in table format)
```

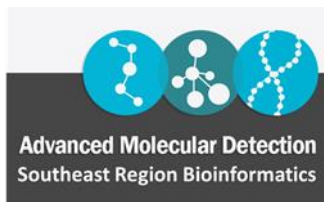
```
[usr@login]$ bs -c <workgroup_name> list project    #lists projects and size for specified, non-default workgroup (in table format) (can use -c for any bs option and sub-option)
```

```
[usr@login]$ bs list project --template='{{.Name}}'    #list just project names (non-table format)
```

```
[usr@login]$ bs list project --filter-term=<projectname> --template='{{.Id}}'    #Get project ID based on project name
```

```
[usr@login]$ bs list biosample --help    #list all flag options (can use for any bs option)
```

```
[usr@login]$ bs list biosample --project-name=<projectname> --template='{{.Name}}'    #list all biosamples for a particular project name
```



BaseSpace CLI Commands and Use Case Examples

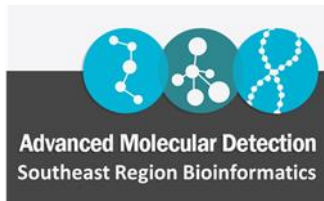
- bs download

```
[usr@login]$ bs download project -n <projectname> --extension=fastq.gz -o <output_dir> #Download just the fastqs from a project using project name
```

```
[usr@login]$ bs download project -i <projectID> --extension=fastq.gz -o <output_dir> #Download just the fastqs from a project using project ID
```

```
[usr@login]$ bs download biosample -n <biosampleID> -o <output_dir> #Download data from a specific biosample
```

Advanced Molecular Detection
Southeast Region Bioinformatics



BaseSpace CLI Commands and Use Case Examples

- bs create and bs upload

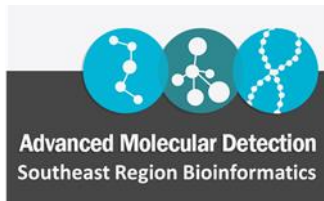
```
[usr@login]$ bs create project -n <projectname> #Create new project in BaseSpace
```

```
[usr@login]$ bs create biosample -n <biosamplename> -p <projectname> #Create new biosample in project
```

```
[usr@login]$ bs upload dataset -p <projectID> <fastqR1> <fastqR2> #Upload a pair of fastq files to a project; creates biosample automatically
```

```
[usr@login]$ bs upload dataset -p <projectID> --recursive <fastqs_dir> #Upload an entire directory of fastqs to a project
```

Southeast Region Bioinformatics

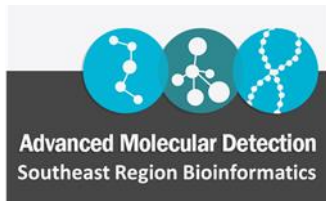




Current Analysis Options

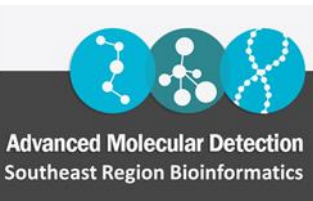
Scripts and pipelines available for state PHL use on HiPerGator

Southeast Region Bioinformatics



Pipelines

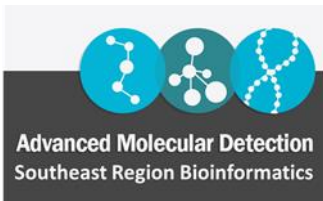
- FLAQ (FLorida Assembly Quality)
 - Generates annotated, de novo assemblies and quality metrics for WGS of bacterial species
- FLAQ-AMR (FLAQ - Antimicrobial Resistance)
 - Generates annotated, de novo assemblies and quality metrics for WGS of bacterial species
 - Determines species ID
 - Determines ST using MLST schemes from PubMLST
 - Identifies AMR genes, virulence genes, and plasmids
 - Performs serotyping of *Salmonella* and *E. coli*, if applicable (more species to come)
- To come
 - FLAQ-Lp (FLAQ – *Legionella pneumophila*)
 - Includes CDC's Lp Species ID Tool and Lp Serotyping Tool
 - FLAQ-USC (FLAQ – Unknown Species Classification)



Pipelines for Comparative Genomics

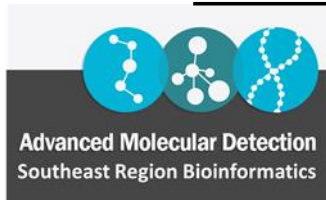
- FL-cgSNP (Core-Genome SNP analysis) – reference-free method
 - Pan-genome analysis (identifies core genes shared by all isolates) and generates a multiple sequence alignment
 - Identifies pair-wise SNPs between isolates and outputs a pairwise SNP matrix
 - Generates a maximum-likelihood phylogenetic tree
- hqSNP (High-Quality SNP analysis) – reference-based method
 - Uses CDC's Lyveset pipeline to identify hqSNPs
 - Generates a pairwise SNP matrix and maximum-likelihood phylogenetic tree
- Custom script to output annotated tree (based on metadata input)

Advanced Molecular Detection
Southeast Region Bioinformatics



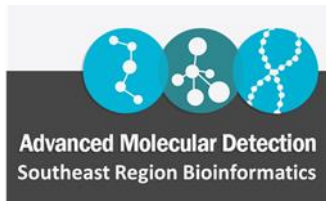
Scripts and Individual Tools

- Quick Species ID (screen against RefSeq database)
- Species ID, contamination check, and metagenomic classification
- Run CDC's Lp Species ID Tool and Lp Serotyping Tool locally
- Pull out gene sequence of interest from assembly (e.g., pull out AR genes of interest)
- Merge fastqs from multiple lanes on a NextSeq into one R1 and one R2 file (compatible with BioNumerics)
- Download fastqs from NCBI's Sequence Read Archive
- Run any tool individually
 - SeqSero2, mlst, abricate, etc.
- Run any tool or workflow in the Staph-B Toolkit (https://staph-b.github.io/staphb_toolkit/)
- Batch runs (i.e., run >2 analysis scripts at one time)
- **Custom scripts and pipelines as requested or needed**



Viral Pipelines and scripts

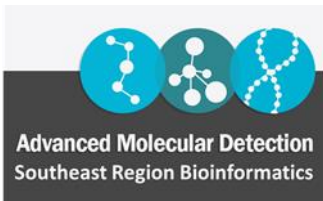
- FLAQ - SC2 (FLAQ – SARS-CoV-2)
 - Generates SARS-CoV-2 consensus assemblies from ARTIC V1, V2, or V3 targeted amplicon sequencing using Illumina (e.g., Nextera XT or Flex) and non-Illumina (e.g., PrimalSeq or MN Tailed) library prep
 - Outputs variant file and final report with quality metrics (including a PASS/FAIL quality flag based on public repository submission criteria)
 - Automatically generates flags if indels or internal stop codons are present, indicating the need for manual review.
- SC2 associated scripts to prepare/format assemblies for batch submissions to GISAID and NCBI's Genbank
- SC2-Correct-Assembly
 - Removes indels/SNPs that are likely PCR or sequencing artifacts/errors (requires prior manual review of mapped reads in IGV (or other program)
- Targeted Amplicon Variant Calling and Consensus Sequence Generation
 - For example, identify variants and generate a consensus sequence for HIV genotyping.
 - Will work for any target/amplicon (just need a reference sequence)



Analysis Request and/or support

- Email BPHL-SEbioinformatics@flhealth.gov
- One-on-one video conference sessions to set up your HiPerGator environment and walk-through the use of each pipeline/script needed
- Custom pipeline/script development as requested
- Scripts (and data, if applicable) can be shared through your /ufrc/bphl-<state>/public-share/ directory on HiPerGator

Advanced Molecular Detection
Southeast Region Bioinformatics

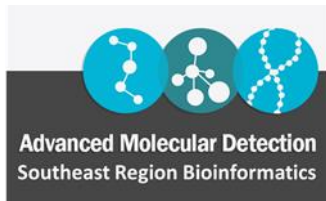


Future Webinars/Trainings

- SARS-CoV-2 Data Submissions
- SARS-CoV-2 Flagged Sample Review
- Requests???



Advanced Molecular Detection
Southeast Region Bioinformatics





Advanced Molecular Detection Southeast Region Bioinformatics

Questions???

BPHL-SEbioinformatics@flhealth.gov

Sarah Schmedes, PhD

Lead Bioinformatician

BRR Lead, Southeast Region

Jiaqi Li, PhD

Bioinformatician

Jason Blanton, PhD

Molecular Administrator

State Sequencing Coordinator