



BioContainers Bonanza

Amanda Cooksey

February 22, 2019



Previous webinars

Webinar: Make Your Life (and Analyses) Easier with Containers

By Julian Pistorius

Webinar: Going Places with Your Container

By Tyson Swetnam



Who am I?

- Biologist
- Genomics/transcriptomics
- Working with containers for about 2 years
- Use BioContainers to integrate tools in the CyVerse Discovery Environment (DE)



What we will cover today

- Review of container basics
- Bioconda
- BioContainers
- How to integrate your BioContainer as a tool in the DE



Benefits of containers

- Platform independent
 - Laptop, server, virtual machines, HPC, Windows, Mac, Linux
- Portable
 - Build it once; run it anywhere
- Reproducible analyses
 - Always the same version of your tool and all its dependencies
- Scalable
 - Vertically and horizontally



Review of terminology

- **Image:** self-contained, read-only 'snapshot' of your applications and packages, with all their dependencies
- **Container:** a running instance of your image
- **Image registry:** a storage and content delivery system, holding named images, available in different tagged versions
- **Docker:** a program that runs and handles life-cycle of containers and images
- **CyVerse tool:** Software program that is integrated into the back end of the DE for use in DE apps
- **CyVerse app:** graphic interface of a tool made available for use in the DE

Dockerfiles and recipes

- Images are built from files that specify software, environmental variables, files to add, metadata
- Two containerization programs
 - Docker--built from Dockerfiles
 - Singularity--built from Singularity recipes
- Not to be confused with Bioconda recipes
 - Used to build Bioconda packages (not containers)
 - Consists of two files: meta.yaml, build.sh



BIOCONDA[®]

- a channel of the conda package manager
- a repository of recipes hosted on GitHub
- over 600 contributors and 450 members who add, modify, update and maintain the recipes
- a repository of **more than 6000** bioinformatics packages
- **Each package added to Bioconda also has a corresponding Docker BioContainer automatically created and uploaded to Quay.io**



BIOCONTAINERS

- Community-driven project
- Infrastructure and basic guidelines to create, manage and distribute Bioinformatics containers
- Focus in proteomics, genomics, transcriptomics and metabolomics
- Provide containers ready to be used by the bioinformatics community



- Registry of container images
- You can create your own free public repositories
 - Account not necessary to use BioContainers
- Integrated with Bioconda GitHub repository for automated builds
- Every Bioconda package (more than 6000) has a BioContainer here
- BioContainers are Docker-based
- All BioContainers are publicly available and free
- <https://quay.io/>

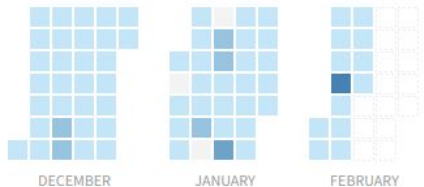


[Next step](#)





Repository Activity



Recent Repo Builds

No builds have been run for this repository.

Description

No repository description has been set

Pull this container with the following Docker command:

```
docker pull quay.io/biocontainers/porechop
```



Repository Tags

[Compact](#) [Expanded](#)

1 - 7 of 7


TAG

LAST MODIFIED ↓

SECURITY SCAN

SIZE

EXPIRES

MANIFEST

☐ 0.2.3_seqan2.1.1--py36h2d50403_3

7 months ago

☐ Unsupported

50.1 MB

Never

SHA256 65f1cbe96399


☐ 0.2.3_seqan2.1.1--py35h2d50403_3

7 months ago

☐ Unsupported

49.3 MB

Never

SHA256 8ea41e99346a


☐ 0.2.3_seqan2.1.1--py35_2

a year ago

☐ Unsupported

47.8 MB

Never

SHA256 5c0403963e69


☐ 0.2.3_seqan2.1.1--py36_2

a year ago

☐ Unsupported

48.4 MB

Never

SHA256 3fb07cfb2451


☐ 0.2.3_seqan2.1.1--py36_1

a year ago

☐ Unsupported

48.4 MB

Never

SHA256 1e3f4794cc59


☐ 0.2.3_seqan2.1.1--py35_1

a year ago

☐ Unsupported

47.8 MB

Never

SHA256 22504bb6deae


☐ 0.2.3_seqan2.1.1--0

a year ago

☐ Unsupported

48.4 MB

Never

SHA256 4bf6abaebe85





Repository Tags

[Compact](#)
[Expanded](#)

1 - 7 of 7




TAG	LAST MODIFIED ↓	SECURITY SCAN	SIZE	EXPIRES	MANIFEST	
<input type="checkbox"/> 0.2.3_seqan2.1.1--py36h2d50403_3	7 months ago	Unsupported	50.1 MB	Never	SHA256 65f1cbe96399	Fetch Tag
<input type="checkbox"/> 0.2.3_seqan2.1.1--py35h2d50403_3	7 months ago	Unsupported	49.3 MB	Never	SHA256 8ea41e99346a	
<input type="checkbox"/> 0.2.3_seqan2.1.1--py35_2	a year ago	Unsupported	47.8 MB	Never	SHA256 5c0403963e69	
<input type="checkbox"/> 0.2.3_seqan2.1.1--py36_2	a year ago	Unsupported	48.4 MB	Never	SHA256 3fb07cfb2451	
<input type="checkbox"/> 0.2.3_seqan2.1.1--py36_1	a year ago	Unsupported	48.4 MB	Never	SHA256 1e3f4794cc59	
<input type="checkbox"/> 0.2.3_seqan2.1.1--py35_1	a year ago	Unsupported	47.8 MB	Never	SHA256 22504bb6deae	
<input type="checkbox"/> 0.2.3_seqan2.1.1--0	a year ago	Unsupported	48.4 MB	Never	SHA256 4bf6abaebe85	

QUAY Explore Tour Tutorial Pricing search Sign in

Repositories

Repository Tags

Image Format: (Select Image Format)

- Docker Pull (by tag)
- Docker Pull (by digest)
- Squashed Docker Image
- rkt Fetch

TAG	LAST MODIFIED	SECURITY SCAN	SIZE	CACHED	MANIFEST
<input type="checkbox"/> 0.2.3_seqan2.1.1--py36h2d50403_3	7 months ago	Unsupported	50.1 MB	Never	SHA256 65f1cbe96399
<input type="checkbox"/> 0.2.3_seqan2.1.1--py35h2d50403_3	7 months ago	Unsupported	49.3 MB	Never	SHA256 8ea41e99346a
<input type="checkbox"/> 0.2.3_seqan2.1.1--py35_2	a year ago	Unsupported	47.8 MB	Never	SHA256 5c8403963e69
<input type="checkbox"/> 0.2.3_seqan2.1.1--py36_2	a year ago	Unsupported	48.4 MB	Never	SHA256 3fb07cfb2451
<input type="checkbox"/> 0.2.3_seqan2.1.1--py36_1	a year ago	Unsupported	48.4 MB	Never	SHA256 1e3f4794cc59
<input type="checkbox"/> 0.2.3_seqan2.1.1--py35_1	a year ago	Unsupported	47.8 MB	Never	SHA256 22504bb6deae
<input type="checkbox"/> 0.2.3_seqan2.1.1--0	a year ago	Unsupported	48.4 MB	Never	SHA256 4bf6abaebe85

Fetch Tag:  0.2.3_seqan2.1.1--py36h2d50403_3 ×

Image Format:



Docker Pull (by tag)



Command:

```
>_ docker pull quay.io/biocontainers/porechop:0.2.3_seqan2.1.1--py36h2d50403_3
```

Copy Command

Close



Pull the image

```
[amcooksey@rogue ~]$ sudo docker pull quay.io/biocontainers/porechop:0.2.3_seqan2.1.1--py36h2d50403_3
[sudo] password for amcooksey:
0.2.3_seqan2.1.1--py36h2d50403_3: Pulling from biocontainers/porechop
a3ed95caeb02: Already exists
b0dc45cd432d: Already exists
9466b3513669: Already exists
ddd482ea7b54: Already exists
4d69f833b9d8: Already exists
e7c454e5167d: Already exists
e38092b005c0: Already exists
f879b42dfe2b: Already exists
9417599398f7: Pull complete
Digest: sha256:65f1cbe96399eff89df55169f25d2b52f46115f9d4080c388fdeb7b22dc76b30
Status: Downloaded newer image for quay.io/biocontainers/porechop:0.2.3_seqan2.1.1--py36h2d50403_3
```



Run the container

```
[amcooksey@rogue racon]$ sudo docker run --rm -v $(pwd):/working-dir -w /working-dir --entrypoint="porechop" quay.io/biocontainers/porechop:0.2.3_seqan2.1.1--py36h2d50403_3 -i SRR6059710.fastq -o porechop_out.fastq
```

```
sudo \  
docker run \  
--rm \  
-v $(pwd):/working-dir \  
-w /working-dir \  
--entrypoint="porechop" \  
quay.io/biocontainers/porechop:0.2.3_seqan2.1.1--py36h2d50403_3 \  
-i SRR6059710.fastq \  
-o porechop_out.fastq
```

‘entrypoint’ is very important:
BioContainers don’t have built-in entrypoints
It tells the container what to run



Container is running

Loading reads

SRR6059710.fastq
543,374 reads loaded

Looking for known adapter sets

4,990 / 10,000 (49.9%)



Trimming adapters from read ends

SQK-NSK007_Y_Top: AATGTACTTCGTTCA GTTACGTATTGCT
SQK-NSK007_Y_Bottom: GCAATACGTA ACTGAACGAAGT
SQK-MAP006_Y_Top_SK63: GGTGTTTCTGTTGGTGCTGATATTGCT
SQK-MAP006_Y_Bottom_SK64: GCAATATCAGCACCAACAGAAA
PCR_1_start: ACTTGCCTGTCGCTCTATCTTC
PCR_1_end: GAAGATAGAGCGACAGGCAAGT
PCR_tail_1_start: TTAACCTTTCTGTTGGTGCTGATATTGC
PCR_tail_1_end: GCAATATCAGCACCAACAGAAAGGTTAA
PCR_tail_2_start: TTAACCTACTTGCCTGTCGCTCTATCTTC
PCR_tail_2_end: GAAGATAGAGCGACAGGCAAGTAGGTTAA

Saving trimmed reads to file

Saved result to /working-dir/porechop_out.fastq



List contents of working directory

```
[amcooksey@rogue racon]$ ls -l
total 11350140
-rw-r----- 1 amcooksey iplant-everyone 346188054 May  8  2018 concat_reads.fastq
-rw-r----- 1 amcooksey iplant-everyone   23424 May  8  2018 miniasm_cat_output.fasta
-rw-r----- 1 amcooksey iplant-everyone   11745 May  8  2018 minimap_cat_rnd2_out.paf
-rw-r--r-- 1 root      root      838803132 Feb 14 12:23 porechop_out.fastq
-rw-r--r-- 1 amcooksey iplant-everyone 9579801472 May 14  2018 SRR6059708.fastq
-rw-r--r-- 1 amcooksey iplant-everyone 857704006 May 14  2018 SRR6059710.fastq
```

At this point

Run your BioContainer anywhere you can run Docker

- Laptop
- CyVerse Atmosphere (or other cloud service)
- HPC systems with Singularity

Webinar: Going Places with Your Container by Tyson Swetnam

- building a Singularity container from an existing Docker container
- running docker containers on CyVerse's Atmosphere cloud-computing platform



CyVerse Discovery Environment

- Provide graphic interface for common analysis tools
- DE tools are Dockerized
- Users can integrate tools
- And build apps
- CyVerse tools and apps
 - Keep them private
 - Share them with collaborators
 - Make them publicly available



Tool Integration in the DE





Authentication Service



Username:

Password:

LOGIN



HTTPS cyverse.org

Links to CyVerse Resources

[Additional Information](#)

[Service Status](#)

[Learning Center](#)

[About CyVerse](#)



[DE](#)

[Next step](#)





Apps

Apps ▾ Workflow ▾ Share ▾ Refresh

Search Apps

Manage Tools Switch View

Categories

My Apps

Topic

Operation

HPC

Apps under development

Favorite Apps

My public apps

Shared with me

Favorite Apps

Filter: All

Name

Integrated by

Rating

System

No apps to display!



Manage Tools

Tools ▾ Share ▾ Refresh All Search tools...

- + Add tool...
- + Request Tool
- Edit
- ✗ Delete
- ↑ Use in app

	Image name	Tag	Status
	kapeel/ballgown-r-package		Public
	kapeel/helitron		Public
	kapeel/sine	latest	Public
etup	bjoyce3/simprilyhtfilessetup	1.0	Public
IncReporter	gandli0806/incinereporter	1.0	Public

Add Tool

Tool Information

* Tool Name: porechop

Description: Porechop is a tool for finding and removing adapters from Oxford Nanopore reads.

* Version : 0.2.3

* Image name: quay.io/biocontainers/porechop

Tag: 0.2.3_seqan2.1.1--py36h2d50403_3

Docker Hub URL:

* Type: executable

* OSG Image Path:

Entrypoint: porechop

Working Directory:

UID:

Max CPU Cores:

Memory Limit: 16 GB

OK Cancel

Be sure to include
'entrypoint'

Add Tool



Tool porechop 0.2.3 added successfully! Your tool is subject to computing restrictions. Please click on the `i` button to view them.

OK

Manage Tools

Tools ▾ Share ▾ Refresh

	Image name	Tag	Status
	quay.io/biocontainers/porechop	0.2.3_s...	own

- + Add tool...
- + Request Tool
- Edit
- ✗ Delete
- ↑ Use in app

Create App

Apps

Save Preview Command Line Order

App Items

Section

Section

Files/Folders

Multiple Input Files

Add Delete

Files

Input File

File Browse

New App

Tool used:

porechop 0.2.3

* App name:

Enter a name for your new app

* App description:

Enter a description for your new app (<255 characters)

Section 1

Drag items here to include in this section.

Details:

Select an item from the ce

Command line view

porechop

porechop 0.2.3

Save
 Preview
 Command Line Order

App Items

Section

Section

Files/Folders

Multiple Input Files

Add
 Delete

Files

Input File

File
 Browse

porechop 0.2.3

Tool used:

porechop 0.2.3

* App name:

porechop 0.2.3

* App description:

Porechop is a tool for finding and removing adapters from Oxford Nanopore reads. Adapters on the ends of reads are trimmed off, and when a read has an adapter in its middle, it is treated as chimeric and chopped into separate reads. Porechop performs thorough alignments to effectively find adapters, even at low sequence identity.

Input

FASTQ or FASTA Nanopore read files :

Add
 Delete

Name

Details: *threads*

Integer Input label:

threads

Argument option:

-t

Default value:

Enter a default value to display

☒ Do not display this item in the app.
 ☐ Make this field required.
 ☒ Exclude this item if nothing is entered.

Tool tip text:

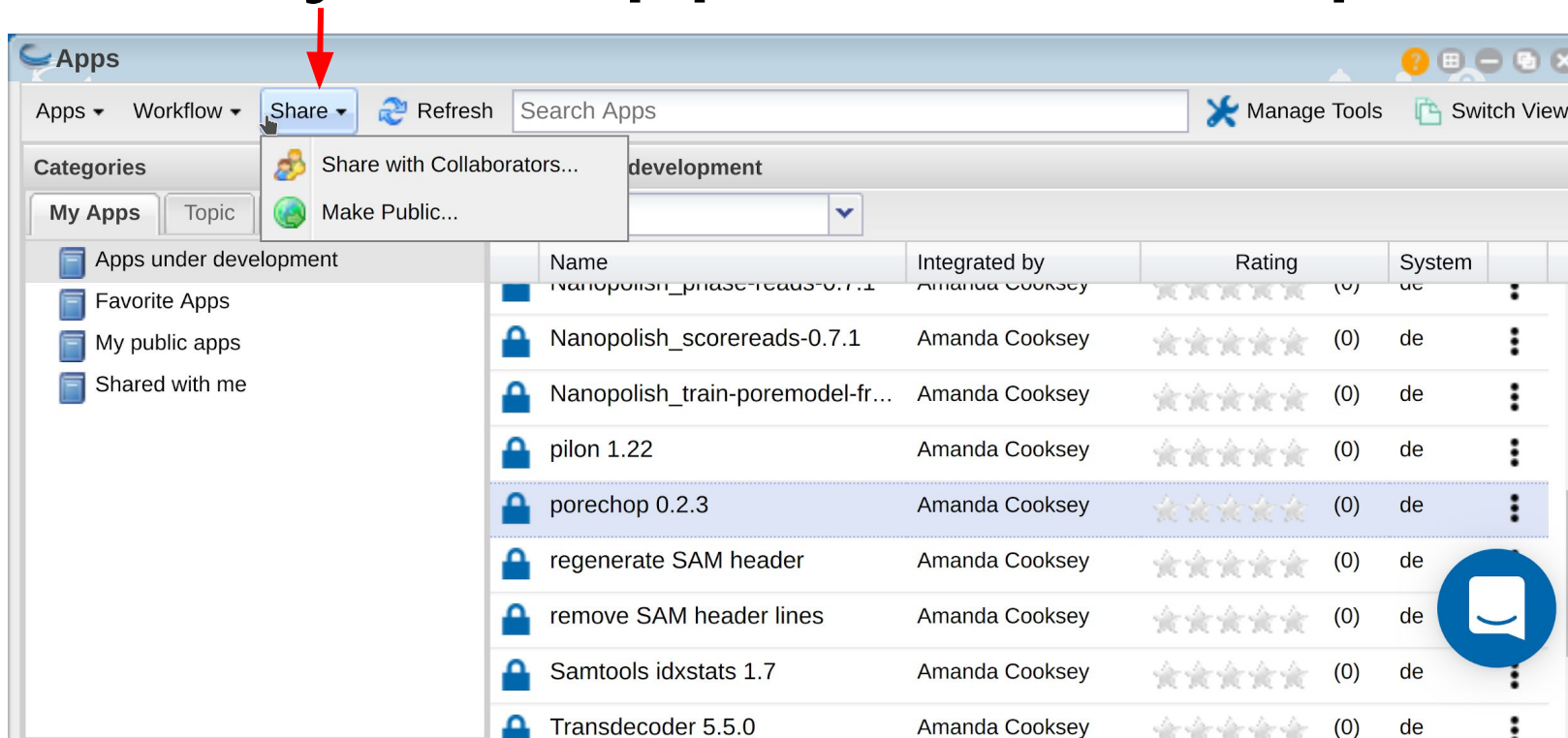
Enter tool tip here

Validation rules:

Command line view

```
porechop -i file -t --barcode_threshold --barcode_diff --adapter_threshold --check_reads --end_size --min_trim_size --extra_end_trim --end_threshold --middle_threshold --extra_middle_trim_good_side --extra_middle_trim_bad_side --min_split_read_size -o file -b file
```

Share your app or make it public



The screenshot shows the 'Apps' interface with a red arrow pointing to the 'Share' dropdown menu. The menu is open, showing two options: 'Share with Collaborators...' and 'Make Public...'. The 'Make Public...' option is highlighted. The interface also includes a 'Categories' sidebar, a 'Search Apps' bar, and a table of apps.

Name	Integrated by	Rating	System
Nanopolish_phase-reads-0.7.1	Amanda Cooksey	(0)	de
Nanopolish_scorereads-0.7.1	Amanda Cooksey	(0)	de
Nanopolish_train-poremodel-fr...	Amanda Cooksey	(0)	de
pilon 1.22	Amanda Cooksey	(0)	de
porechop 0.2.3	Amanda Cooksey	(0)	de
regenerate SAM header	Amanda Cooksey	(0)	de
remove SAM header lines	Amanda Cooksey	(0)	de
Samtools idxstats 1.7	Amanda Cooksey	(0)	de
Transdecoder 5.5.0	Amanda Cooksey	(0)	de

Steps to integrate a BioContainer in the Discovery Environment

1. Find BioContainers image at Quay.io
2. Pull your image with Docker
3. Run container with appropriate data to test
4. Add tool in DE
5. Build app for tool
6. Share your tool/app as you wish



Useful Links and References

- [Bioconda GitHub Repository](#)
- [BioContainers on GitHub](#)
- [Quay.io](#)
- [DockerHub](#)
- [Biocontainers.pro](#)
- [BioContainers Edu](#)





CyVerse is supported by the National Science Foundation under Grants No. DBI-0735191, DBI-1265383 and DBI-1743442.



Cold
Spring
Harbor
Laboratory

