

# Practical introduction to Singularity containers

November 11, 2021

Tom Harrop  
[tharrop@unimelb.edu.au](mailto:tharrop@unimelb.edu.au)

[melbournebioinformatics.org.au](http://melbournebioinformatics.org.au)

 [@MelBioInf](https://twitter.com/MelBioInf)

# Reproducibility

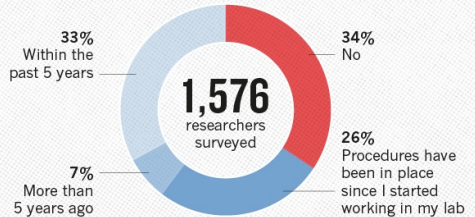
## IS THERE A REPRODUCIBILITY CRISIS?



©nature

## HAVE YOU ESTABLISHED PROCEDURES FOR REPRODUCIBILITY?

Among the most popular strategies was having different lab members redo experiments.



©nature

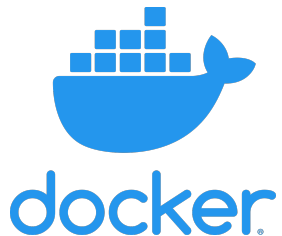
# Analysis practices

## Guidelines:

1. Don't modify raw data
2. Record the steps  
(with version control)
3. Capture the computing environment

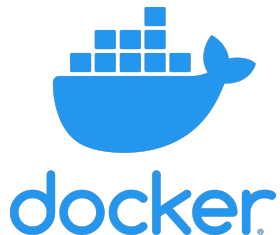
- 👤 Data management:
  - see Gayle Philip's slides
- 👤 Project management, source control:
  - Jessica Chung's slides
- 👤 Workflow management:
  - Bobbie Shaban's slides
- 👤 Reproducible computing:
  - use containers

# Why use containers?



**Motivation:** Use software without installing it

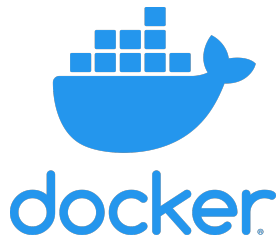
# Why use containers?



**Motivation:** Use software without installing it

**Benefits:** Portability, reproducibility

# Why use containers?



**Motivation:** Use software without installing it

**Benefits:** Portability, reproducibility

**Drawbacks:** Effort

# Use software without installing it

## On my computer:

```
# not provided for my OS  
guppy_basecaller --version
```

guppy\_basecaller: command not found

## Using a container:

```
singularity exec guppy_5.0.16.sif \  
guppy_basecaller --version
```

: Guppy Basecalling Software, (C) Oxford Nanopore

# Get the version you need

## On my computer:

```
# installed by my OS package manager  
samtools --version
```

```
samtools 1.11  
Using htlib 1.11-4  
Copyright (C) 2020 Genome Research Ltd.
```

## Using a container:

```
singularity exec samtools_1.12.sif \  
samtools --version
```

```
samtools 1.12  
Using htlib 1.12  
Copyright (C) 2021 Genome Research Ltd.
```



## Avoid tricky software installations

### Dependencies

Funannotate has a lot of dependencies. However, it also comes with a few tools to help you get everything installed. The first is that of `funannotate check`. You'll see in the output below that the `fastx` tool is missing, which is Bill Pearson's `fastx` a dependency of the PAGA pipeline. Also the `STAR` and `SRANIMATE` variables are not set, that is because on this particular machine they are not installed. I.e. Funannotate will alert you at runtime if it is missing a dependency.

[illegible]

🤖 lots of dependencies



🤖 not all dependencies are listed

## Avoid tricky software installations

### Dependencies

Funannotate has a lot of dependencies. However, it also comes with a few tools to help you get everything installed. The first is that of `funannotate check`. You'll see in the output below that the `fastx` tool is missing, which is Bill Pearson's `fastx` a dependency of the PAGA pipeline. Also the `BLAST` and `SRSTKIT` variables are not set, that is because on this particular machine they are not installed. I.e. Funannotate will alert you at runtime if it is missing a dependency.

[illegible]

- 🤖 lots of dependencies
- 🤖 not all dependencies are listed
- 🤖 need specific versions
- 🤖 package managers don't always help

## Avoid tricky software installations

### Dependencies

Paranotate has a lot of dependencies. However, it also comes with a few tools to help you get everything installed. The first is that of `conda`. You'll see in the output below that the `twine` tool is missing, which is Bill Pearsons' `twine` a dependency of the PKA pipeline. Also the `shelvarc` and `SRCArchive` variables are not set, that is because on this particular machine they are not installed. i.e. Paranotate will alert you at runtime if it is missing a dependency.

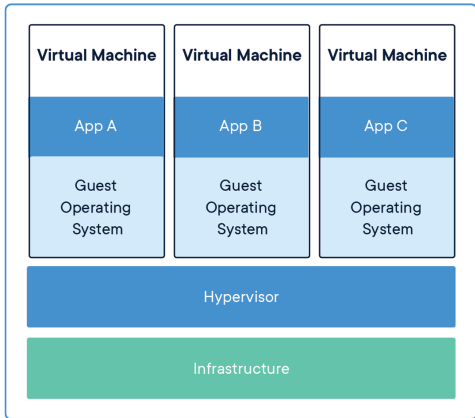
```
$ funannotate vdb check --show-versions
Checking dependencies for funannotate v3.0
=====
You are running Python 3.7.11. Now checking python packages
biopython: 1.78
cycler: 0.11.0
matplotlib: 3.1.1
networkx: 2.3.0
numpy: 1.32.1
pandas: 0.32.0
pillow: 5.4.3
requests: 2.18.4
scikit-learn: 0.20.0
scipy: 0.20.1
seaborn: 0.8.1
All 11 python packages installed
```

[illegible][illegible][illegible]

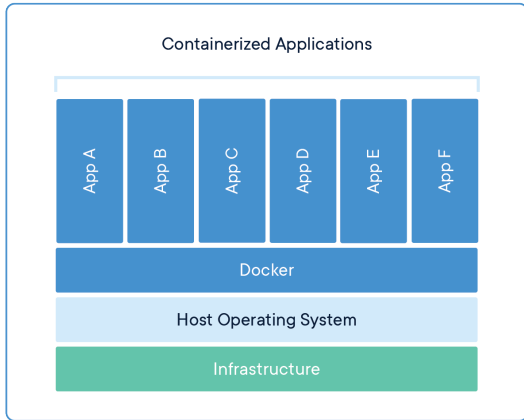
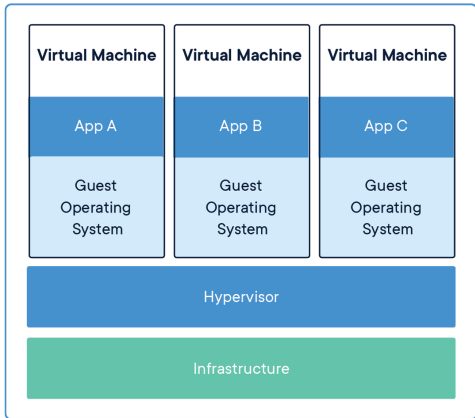
- 🤖 lots of dependencies
- 🤖 not all dependencies are listed
- 🤖 need specific versions
- 🤖 package managers don't always help

```
singularity run docker://nextgenusfs/funannotate
```

# What is a container?



# What is a container?



# Docker, or Singularity?

## Docker

- popular on cloud platforms
- effectively requires root to run\*
- doesn't respect job schedulers\*

\* may have been fixed

# Docker, or Singularity?

## Docker

- popular on cloud platforms
- effectively requires root to run\*
- doesn't respect job schedulers\*

\* may have been fixed

## Singularity

- runs on HPC
- popular in research / edu?
- **supports Docker containers**

# Docker, or Singularity?

## Docker

- popular on cloud platforms
- effectively requires root to run\*
- doesn't respect job schedulers\*

## Singularity

- runs on HPC
- popular in research / edu?
- **supports Docker containers**

\* may have been fixed

**Suggested workflow:** build (or obtain) your containers in Docker format and use Singularity to run them



# How to use Singularity

## 1. Install Singularity

- 🤖 not too hard to install locally
- 🤖 HPCs often have it installed
- 🤖 you can ask the admin to install it

e.g. on Spartan:

```
$ module load singularity/3.7.3  
$ singularity --version  
singularity version 3.7.3
```

# How to use Singularity

## 1. Install Singularity

- 🤖 not too hard to install locally
- 🤖 HPCs often have it installed
- 🤖 you can ask the admin to install it

e.g. on Spartan:

```
$ module load singularity/3.7.3
$ singularity --version
singularity version 3.7.3
```

## 2. Run a container

- 🤖 detailed instructions at [singularity.hpcng.org](https://singularity.hpcng.org)
- 🤖 basic commands is  
`singularity exec {container URL} \`  
`{command} {args}`
- 🤖 usually need some options, e.g.
  - `-B` binds paths to make files available in the container
  - `--nv` enables GPU support

# Workflow managers support containers

## Snakefile:

```
rule trim_adaptors:
    input:          'data/raw_reads.fastq'
    output:         'output/trimmed.fastq'
    container:      'docker://my_repos/trim_adaptors:1.7'
    shell:          'trim_adaptors --raw_reads={input} > {output}'

rule run_assembly:
    input:          'output/trimmed.fastq'
    output:         'output/assembly.fasta'
    container:      'docker://my_repos/tidy_assembler:2.9'
    shell:          'tidy_assembler --reads={input} > {output}'
```

# Workflow managers support containers

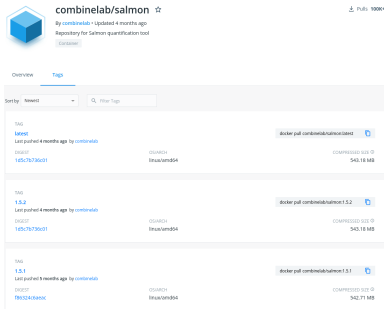
## Snakefile:

```
rule trim_adaptors:
    input:          'data/raw_reads.fastq'
    output:         'output/trimmed.fastq'
    container:      'docker://my_repos/trim_adaptors:1.7'
    shell:          'trim_adaptors --raw_reads={input} > {output}'
rule run_assembly:
    input:          'output/trimmed.fastq'
    output:         'output/assembly.fasta'
    container:      'docker://my_repos/tidy_assembler:2.9'
    shell:          'tidy_assembler --reads={input} > {output}'
```

## Run the workflow:

```
snakemake --use-singularity \
    --singularity-args ...
```

# Where to find containers



combinelab/salmon ☆

By combinelab • Updated 4 months ago  
Repository for Salmon quantification tool

1000+

Overview Tags

Sort by: Newest Filter tags


TAG	DIGEST	OS/ARCH	COMPRESSED SIZE
latest	1d5c7b739cd1	linux/amd64	543.18 MB
1.5.2	1d5c7b739cd1	linux/amd64	543.18 MB
1.5.1	f9b32418a00c	linux/amd64	542.71 MB



Some developers provide docker containers

```
singularity pull \  
docker://combinelab/salmon:1.1.0
```

# You can search for images on [hub.docker.com](https://hub.docker.com)

 **dockerhub**

[Explore](#) [Pricing](#) [Sign In](#) [Sign Up](#)

 [Docker](#)

 [Containers](#)

 [Plugins](#)

**Filters**

☐ **Verified Publisher**

☐ **Official Images**  
Official Images Published By Docker

**Categories**

☐ Analytics

☐ Application Frameworks

☐ Application Infrastructure

☐ Application Services

☐ Base Images

1 - 16 of 16 results for **trinityrnaseq**. [Clear search](#)

Best Match



**trinityrnaseq/trinityrnaseq-wdl**

By [trinityrnaseq](#) • Updated a year ago

Container

Linux

x86-64

50K+

0

DownloadsStars



**trinityrnaseq/trinityrnaseq**

By [trinityrnaseq](#) • Updated 2 months ago

Trinity Docker

Container

Linux

x86-64

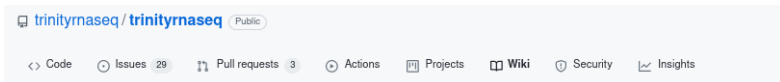
10K+

4

DownloadsStars

13 / 19

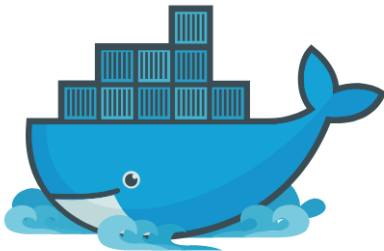
# Check with the software developer



## Trinity in Docker

Brian Haas edited this page on 1 Apr 2020 · 13 revisions

## Run Trinity Using Docker



If you have [Docker](#) installed, you can pull [our image from DockerHub](#), which contains Trinity and all software used for downstream analyses supported within the larger Trinity framework.

# Build them yourself!

TomHarrop / container-flye Public

generated from [deardenlab/container-build-template](#)

Unwatch 1

Star 0

Fork 0

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights


Settings

main 1 branch 2 tags





Go to file

Add file


Code

 **TomHarrop** Create README.md

✓ d8caa3a 16 days ago 11 commits

 .github/workflows	Initial commit	16 days ago
 Dockerfile.2.9	certificates	16 days ago
 README.md	Create README.md	16 days ago
 VERSION	2.9	16 days ago

README.md



Singularity and Docker containers for [Flye](#).

About

No description, website, or topics provided.

Readme

Releases 2

 **Release v2.9** Latest

16 days ago

[+ 1 release](#)

Packages 1


 [container-flye](#)

Languages

RoFF 100.0%



# Build them yourself!

 **TomHarrop** / **container-flye** Public

generated from [deardenlab/container-build-template](#)

Unwatch 1 Star 0 Fork 0

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

## container-flye

 Install from the command line:

[Learn more](#)

```
$ docker pull ghcr.io/tomharrop/container-flye:2.9
```

### Recent tagged image versions

[2.9](#)

↓ 6

Published 16 days ago · Digest [...](#)

[View and manage all versions](#)

### README.md

Singularity and Docker containers for [Flye](#).

### Details

 TomHarrop  
 container-flye  
 Readme

Last published Issues

**16 days ago** **0**

Total downloads

**9** 

### Collaborators 1

 TomHarrop Tom Harrop

[Open an issue](#)

[Package settings](#)

# Some barriers to container usage

- 👑 Building containers yourself can be painful
- 👑 Duplication of effort
- 👑 Some software **shouldn't** go in a container because of “unfortunate licensing issues”
  - DTU software e.g. rnammer, tmhmm
  - GATech: GeneMark
  - GIRInst's RepBase
- 👑 Getting Singularity installed

# When to use containers

**Modules:**

**Conda, venv, *etc.*:**

**Containers:**

# When to use containers

## Modules:

```
module load \  
samtools/1.12
```

## Conda, venv, etc.:

## Containers:

- 🤖 installed by the sysadmin
- 🤖 properly configured
- 🤖 easy

# When to use containers

## Modules:

```
module load \  
  samtools/1.12
```

- 👤 installed by the sysadmin
- 👤 properly configured
- 👤 easy

## Conda, venv, etc.:

```
conda install -c \  
  bioconda samtools
```

- 👤 latest packages
- 👤 testing in a specific environment

## Containers:

# When to use containers

## Modules:

```
module load \  
  samtools/1.12
```

- 👤 installed by the sysadmin
- 👤 properly configured
- 👤 easy

## Conda, venv, etc.:

```
conda install -c \  
  bioconda samtools
```

- 👤 latest packages
- 👤 testing in a specific environment

## Containers:

```
singularity run \  
  samtools.sif
```

- 👤 reproducibility
- 👤 collaboration
- 👤 portability
- 👤 performance?

# Get started with containers

## 👤 How I use containers:

1. **Write Docker recipes** (one tool, one recipe)
2. Remotely build Docker image store it permanently
3. Use with Singularity via a workflow manager

👤 Singularity has good documentation: [quick start](#)

👤 Docker documentation is ... extensive: [Best practices for writing Dockerfiles](#)

👤 Integrating with workflow managers is easy