Practical introduction to Singularity containers

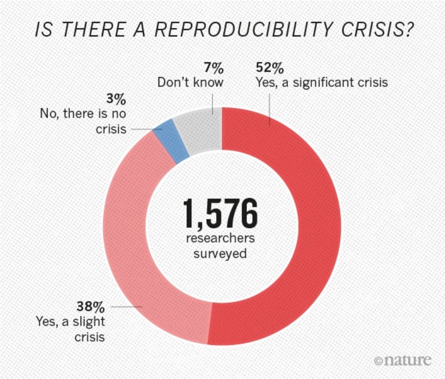
November 11, 2021

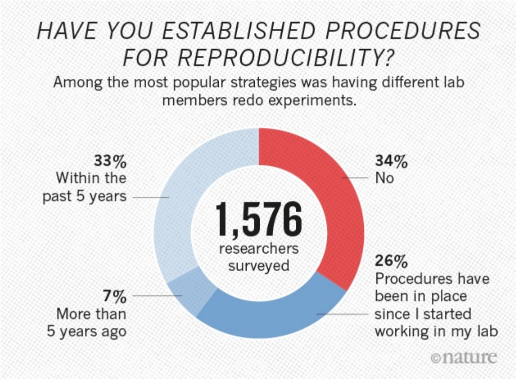
Tom Harrop
tharrop@unimelb.edu.au

melbournebioinformatics.org.au

🐦 @MelBioInf

Reproducibility





Jenny Bryan https://youtu.be/7oyiPBjLAWY & Nature News, 10.1038/533452a

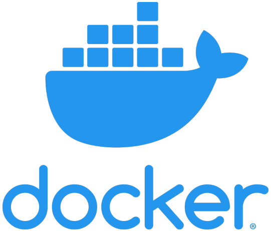
Analysis practices

Guidelines:

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

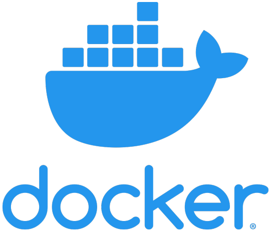
- Data management:
  - see Gayle Philip’s slides
- Project management, source control:
  - Jessica Chung’s slides
- Workflow management:
  - Bobbie Shaban’s slides
- Reproducible computing:
  - use containers

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 Nar
```

Get the version you need

On my computer:

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

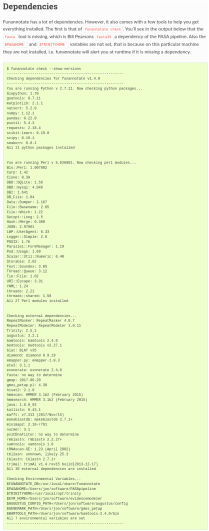
```
samtools 1.13
Using htslib 1.13+ds
Copyright (C) 2021 Genome Research Ltd.
```

Using a container:

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

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

Avoid tricky software installations



Dependencies

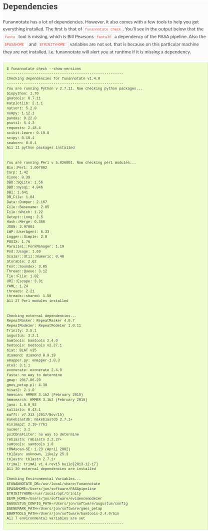
Funannotate has a lot of dependencies. However, it also comes with a few tools to help you get everything installed. The first is that of `funannotate check`. You'll see in the output below that the `fasta` tool is missing, which is Bill Pearsons `fasta36` a dependency of the PASA pipeline. Also the `$PASAHOME`` and `$TRINITYHOME`` 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.

```
$ funannotate check --show-versions
-------------------------------------------------------
Checking dependencies for funannotate v1.4.0
-------------------------------------------------------
You are running Python v 2.7.11. Now checking python packages...
biopython: 1.70
goatools: 0.7.11
matplotlib: 2.1.1
natsort: 5.2.0
numpy: 1.12.1
pandas: 0.22.0
psutil: 5.4.3
requests: 2.18.4
scikit-learn: 0.19.0
scipy: 0.19.1
seaborn: 0.8.1
All 11 python packages installed


You are running Perl v 5.026001. Now checking perl modules...
Bio::Perl: 1.007002
Carp: 1.42
Clone: 0.39
DBD::SQLite: 1.56
DBD::mysql: 4.046
DBI: 1.641
DB_File: 1.84
Data::Dumper: 2.167
File::Basename: 2.85
File::Which: 1.22
Getopt::Long: 2.5
Hash::Merge: 0.300
JSON: 2.97001
LWP::UserAgent: 6.33
Logger::Simple: 2.0
POSIX: 1.76
Parallel::ForkManager: 1.19
Pod::Usage: 1.69
Scalar::Util::Numeric: 0.40
Storable: 2.62
Text::Soundex: 3.05
Thread::Queue: 3.12
Tie::File: 1.02
URI::Escape: 3.31
YAML: 1.24
threads: 2.21
threads::shared: 1.58
All 27 Perl modules installed


Checking external dependencies...
RepeatMasker: RepeatMasker 4.0.7
RepeatModeler: RepeatModeler 1.0.11
Trinity: 2.5.1
augustus: 3.2.1
bamtools: bamtools 2.4.0
bedtools: bedtools v2.27.1
blat: BLAT v35
diamond: diamond 0.9.19
emapper.py: emapper-1.0.3
ete3: 3.1.1
exonerate: exonerate 2.4.0
fasta: no way to determine
gmap: 2017-06-20
gmes_petap.pl: 4.30
hisat2: 2.1.0
hmmscan: HMMER 3.1b2 (February 2015)
hmmsearch: HMMER 3.1b2 (February 2015)
java: 1.8.0_92
kallisto: 0.43.1
mafft: v7.313 (2017/Nov/15)
makeblastdb: makeblastdb 2.7.1+
minimap2: 2.10-r761
nucmer: 3.1
pslCDnaFilter: no way to determine
rmblastn: rmblastn 2.2.27+
samtools: samtools 1.8
tRNAscan-SE: 1.23 (April 2002)
tbl2asn: unknown, likely 25.3
tblastn: tblastn 2.7.1+
trimal: trimAl v1.4.rev15 build[2013-12-17]
All 30 external dependencies are installed

Checking Environmental Variables...
$FUNANNOTATE_DB=/usr/local/share/funannotate
$PASAHOME=/Users/jon/software/PASApipeline
$TRINITYHOME=/usr/local/opt/trinity
$EVM_HOME=/Users/jon/software/evidencemodeler
$AUGUSTUS_CONFIG_PATH=/Users/jon/software/augustus/config
$GENEMARK_PATH=/Users/jon/software/gmes_petap
$BAMTOOLS_PATH=/Users/jon/software/bamtools-2.4.0/bin
All 7 environmental variables are set
-------------------------------------------------------
```

- lots of dependencies
- not all dependencies are listed

funannotate.readthedocs.io

Avoid tricky software installations



Dependencies

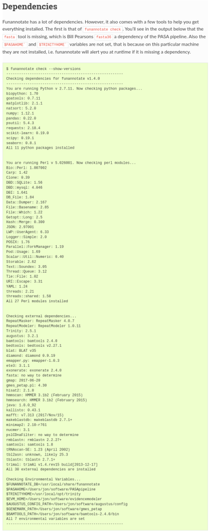
Funannotate has a lot of dependencies. However, it also comes with a few tools to help you get everything installed. The first is that of `funannotate check`. You'll see in the output below that the `fasta` tool is missing, which is Bill Pearsons `fasta36` a dependency of the PASA pipeline. Also the `$PASAHOME`` and `$TRINITYHOME`` 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.

```
$ funannotate check --show-versions
-------------------------------------------------------
Checking dependencies for funannotate v1.4.0
-------------------------------------------------------
You are running Python v 2.7.11. Now checking python packages...
biopython: 1.70
goatools: 0.7.11
matplotlib: 2.1.1
natsort: 5.2.0
numpy: 1.12.1
pandas: 0.22.0
psutil: 5.4.3
requests: 2.18.4
scikit-learn: 0.19.0
scipy: 0.19.1
seaborn: 0.8.1
All 11 python packages installed


You are running Perl v 5.026001. Now checking perl modules...
Bio::Perl: 1.007002
Carp: 1.42
Clone: 0.39
DBD::SQLite: 1.56
DBD::mysql: 4.046
DBI: 1.641
DB_File: 1.84
Data::Dumper: 2.167
File::Basename: 2.85
File::Which: 1.22
Getopt::Long: 2.5
Hash::Merge: 0.300
JSON: 2.97001
LWP::UserAgent: 6.33
Logger::Simple: 2.0
POSIX: 1.76
Parallel::ForkManager: 1.19
Pod::Usage: 1.69
Scalar::Util::Numeric: 0.40
Storable: 2.62
Text::Soundex: 3.05
Thread::Queue: 3.12
Tie::File: 1.02
URI::Escape: 3.31
YAML: 1.24
threads: 2.21
threads::shared: 1.58
All 27 Perl modules installed


Checking external dependencies...
RepeatMasker: RepeatMasker 4.0.7
RepeatModeler: RepeatModeler 1.0.11
Trinity: 2.5.1
augustus: 3.2.1
bamtools: bamtools 2.4.0
bedtools: bedtools v2.27.1
blat: BLAT v35
diamond: diamond 0.9.19
emapper.py: emapper-1.0.3
ete3: 3.1.1
exonerate: exonerate 2.4.0
fasta: no way to determine
gmap: 2017-06-20
gmes_petap.pl: 4.30
hisat2: 2.1.0
hmmscan: HMMER 3.1b2 (February 2015)
hmmsearch: HMMER 3.1b2 (February 2015)
java: 1.8.0_92
kallisto: 0.43.1
mafft: v7.313 (2017/Nov/15)
makeblastdb: makeblastdb 2.7.1+
minimap2: 2.10-r761
nucmer: 3.1
pslCDnaFilter: no way to determine
rmblastn: rmblastn 2.2.27+
samtools: samtools 1.8
tRNAscan-SE: 1.23 (April 2002)
tbl2asn: unknown, likely 25.3
tblastn: tblastn 2.7.1+
trimal: trimAl v1.4.rev15 build[2013-12-17]
All 30 external dependencies are installed

Checking Environmental Variables...
$FUNANNOTATE_DB=/usr/local/share/funannotate
$PASAHOME=/Users/jon/software/PASApipeline
$TRINITYHOME=/usr/local/opt/trinity
$EVM_HOME=/Users/jon/software/evidencemodeler
$AUGUSTUS_CONFIG_PATH=/Users/jon/software/augustus/config
$GENEMARK_PATH=/Users/jon/software/gmes_petap
$BAMTOOLS_PATH=/Users/jon/software/bamtools-2.4.0/bin
All 7 environmental variables are set
-------------------------------------------------------
```

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

funannotate.readthedocs.io

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 `fasta` tool is missing, which is Bill Pearsons `fasta36` a dependency of the PASA pipeline. Also the `$PASAHOME`` and `$TRINITYHOME`` 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.

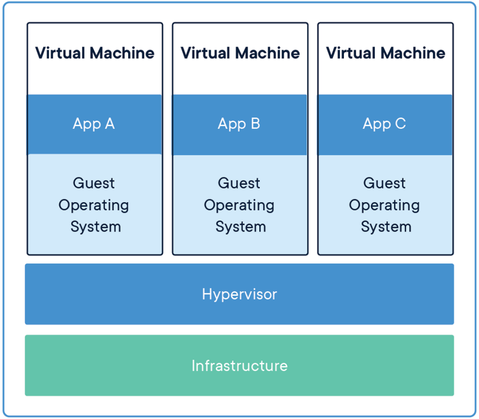
```
$ funannotate check --show-versions
-------------------------------------------------------
Checking dependencies for funannotate v1.4.0
-------------------------------------------------------
You are running Python v 2.7.11. Now checking python packages...
biopython: 1.70
goatools: 0.7.11
matplotlib: 2.1.1
natsort: 5.2.0
numpy: 1.12.1
pandas: 0.22.0
psutil: 5.4.3
requests: 2.18.4
scikit-learn: 0.19.0
scipy: 0.19.1
seaborn: 0.8.1
All 11 python packages installed


You are running Perl v 5.026001. Now checking perl modules...
Bio::Perl: 1.007002
Carp: 1.42
Clone: 0.39
DBD::SQLite: 1.56
DBD::mysql: 4.046
DBI: 1.641
DB_File: 1.84
Data::Dumper: 2.167
File::Basename: 2.85
File::Which: 1.22
Getopt::Long: 2.5
Hash::Merge: 0.300
JSON: 2.97001
LWP::UserAgent: 6.33
Logger::Simple: 2.0
POSIX: 1.76
Parallel::ForkManager: 1.19
Pod::Usage: 1.69
Scalar::Util::Numeric: 0.40
Storable: 2.62
Text::Soundex: 3.05
Thread::Queue: 3.12
Tie::File: 1.02
URI::Escape: 3.31
YAML: 1.24
threads: 2.21
threads::shared: 1.58
All 27 Perl modules installed


Checking external dependencies...
RepeatMasker: RepeatMasker 4.0.7
RepeatModeler: RepeatModeler 1.0.11
Trinity: 2.5.1
augustus: 3.2.1
bamtools: bamtools 2.4.0
bedtools: bedtools v2.27.1
blat: BLAT v35
diamond: diamond 0.9.19
emapper.py: emapper-1.0.3
ete3: 3.1.1
exonerate: exonerate 2.4.0
fasta: no way to determine
gmap: 2017-06-20
gmes_petap.pl: 4.30
hisat2: 2.1.0
hmmscan: HMMER 3.1b2 (February 2015)
hmmsearch: HMMER 3.1b2 (February 2015)
java: 1.8.0_92
kallisto: 0.43.1
mafft: v7.313 (2017/Nov/15)
makeblastdb: makeblastdb 2.7.1+
minimap2: 2.10-r761
nucmer: 3.1
pslCDnaFilter: no way to determine
rmblastn: rmblastn 2.2.27+
samtools: samtools 1.8
tRNAscan-SE: 1.23 (April 2002)
tbl2asn: unknown, likely 25.3
tblastn: tblastn 2.7.1+
trimal: trimAl v1.4.rev15 build[2013-12-17]
All 30 external dependencies are installed

Checking Environmental Variables...
$FUNANNOTATE_DB=/usr/local/share/funannotate
$PASAHOME=/Users/jon/software/PASApipeline
$TRINITYHOME=/usr/local/opt/trinity
$EVM_HOME=/Users/jon/software/evidencemodeler
$AUGUSTUS_CONFIG_PATH=/Users/jon/software/augustus/config
$GENEMARK_PATH=/Users/jon/software/gmes_petap
$BAMTOOLS_PATH=/Users/jon/software/bamtools-2.4.0/bin
All 7 environmental variables are set
-------------------------------------------------------
```

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

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

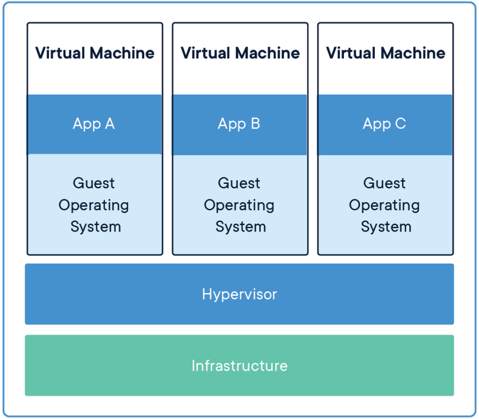
funannotate.readthedocs.io

What is a container?



docker.com

What is a container?



Virtual Machine

App A

Guest Operating System

Virtual Machine

App B

Guest Operating System

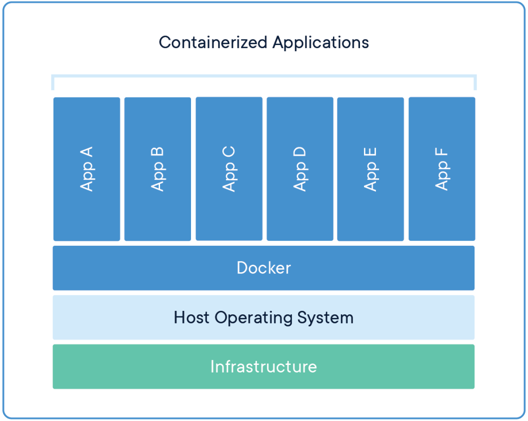
Virtual Machine

App C

Guest Operating System

Hypervisor

Infrastructure



Containerized Applications

App A

App B

App C

App D

App E

App F

Docker

Host Operating System

Infrastructure

docker.com

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*

* may have been fixed

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

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
- 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

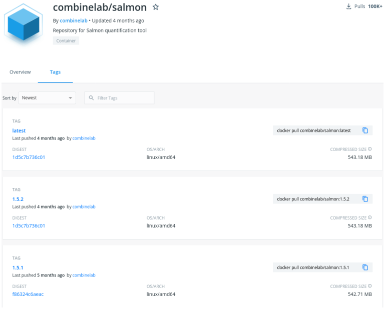
Snakefile:

```
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 ☆

Pulls **100K+**

By combinelab • Updated 4 months ago

Repository for Salmon quantification tool

Container

TAG
latest
Last pushed **4 months ago** by combinelab

```
docker pull combinelab/salmon:latest
```

DIGEST 1d5c7b736c01 · OS/ARCH linux/amd64 · COMPRESSED SIZE 543.18 MB

TAG
1.5.2
Last pushed **4 months ago** by combinelab

```
docker pull combinelab/salmon:1.5.2
```

DIGEST 1d5c7b736c01 · OS/ARCH linux/amd64 · COMPRESSED SIZE 543.18 MB

TAG
1.5.1
Last pushed **5 months ago** by combinelab

```
docker pull combinelab/salmon:1.5.1
```

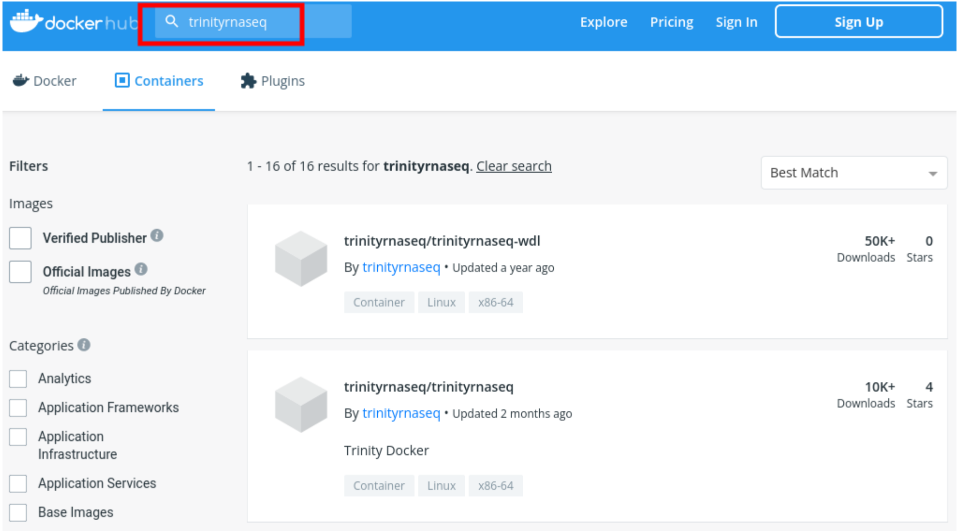
DIGEST f86324c6aeac · OS/ARCH linux/amd64 · COMPRESSED SIZE 542.71 MB

- Some developers provide docker containers

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

hub.docker.com

You can search for images on hub.docker.com



Filters

Images

□ 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

trinityrnaseq/trinityrnaseq-wdl

By trinityrnaseq • Updated a year ago

Container Linux x86-64

50K+ Downloads 0 Stars

trinityrnaseq/trinityrnaseq

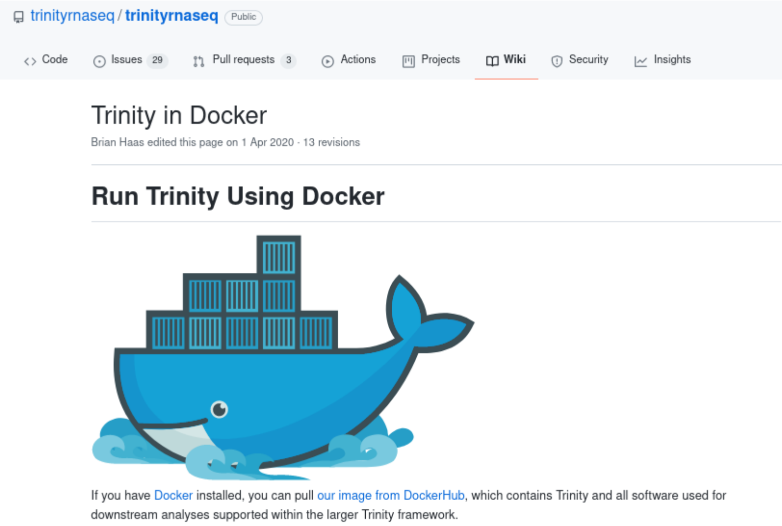
By trinityrnaseq • Updated 2 months ago

Trinity Docker

Container Linux x86-64

10K+ Downloads 4 Stars

Check with the software developer



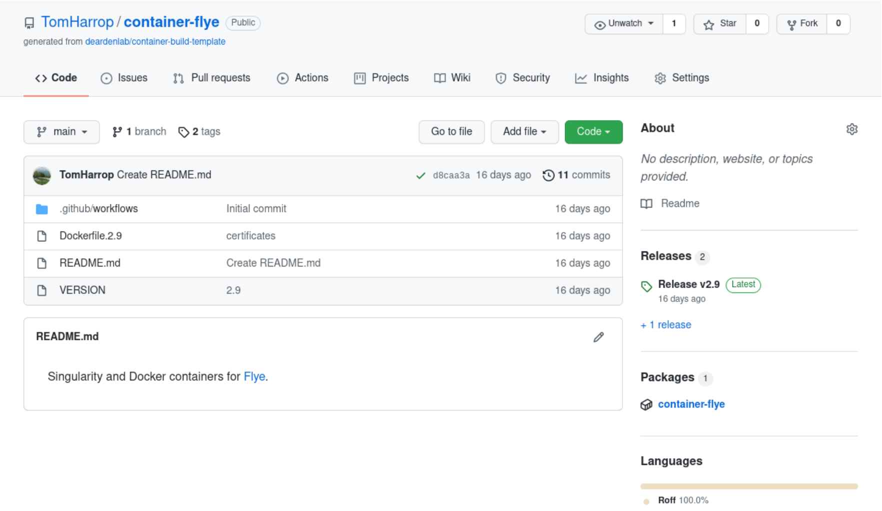
Trinity in Docker

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

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

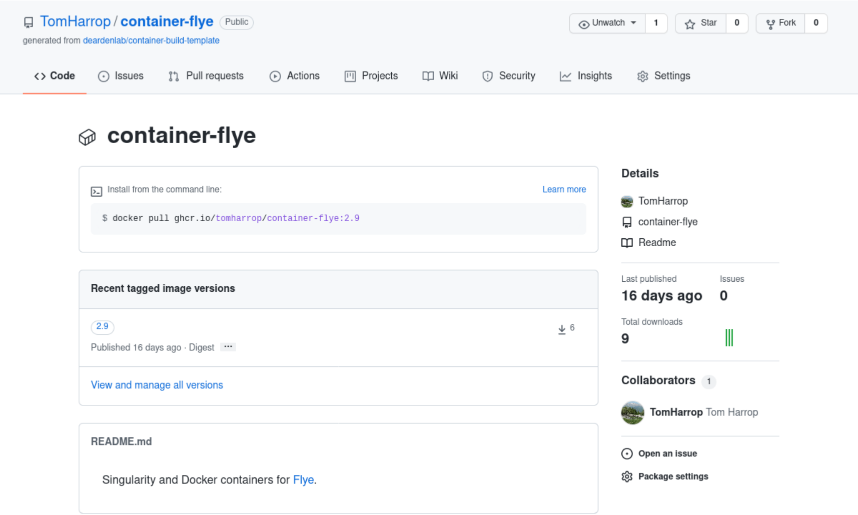
container-flye

Languages

● **Roff** 100.0%

TomHarrop/container-build-template

Build them yourself!



generated from deardenlab/container-build-template

container-flye

Install from the command line:

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

Recent tagged image versions

2.9 — 6 downloads

Published 16 days ago · Digest

Last published: 16 days ago

Issues: 0

Total downloads: 9

TomHarrop Tom Harrop

README.md

Singularity and Docker containers for Flye.

TomHarrop/container-build-template

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

See the DTU's license here

When to use containers

Modules:

Conda, venv, *etc.*:

Containers:

When to use containers

Modules:

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

- installed by the sysadmin
- properly configured
- easy

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

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:

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?

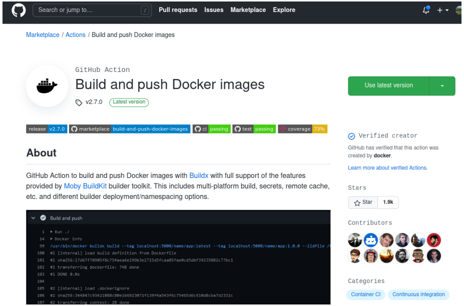
Get started with containers

- 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

Building containers on GitHub

TomHarrop/container-build-template



GitHub Action

Build and push Docker images

v2.7.0 Latest version

About

GitHub Action to build and push Docker images with Buildx with full support of the features provided by Moby BuildKit builder toolkit. This includes multi-platform build, secrets, remote cache, etc. and different builder deployment/namespacing options.

Build and push

```
▶ Run ./
▶ Docker info
/usr/bin/docker buildx build --tag localhost:5000/name/app:latest --tag localhost:5000/name/app:1.0.0 --iidfile /
#1 [internal] load build definition from Dockerfile
#1 sha256:17d67f70905f8c754aea6e299b3e2715d5fcaa05fae0cd5dbf39235082c77bc1
#1 transferring dockerfile: 74B done
#1 DONE 0.0s

#2 [internal] load .dockerignore
#2 sha256:344847c93421088c80e1b6923071f139f4a543f6c75465d6c610d8cba7d2332c
#2 transferring context: 2B done
```

Verified creator

GitHub has verified that this action was created by **docker**.

Learn more about verified Actions.

Stars

Contributors

Categories

Container CI

Continuous integration

Docker and Singularity example

TomHarrop/container-flye

- generally use the Docker container in Singularity
- some old versions can’t login to GHCR

Is this useful?

In your experience do developers appreciate pull requests so that everyone has access to the container from the official page - or do containers take a bit of time to maintain so developers would be less keen?

- question from Alistair Legione

- do enough people use these setups?
- is Conda preferred?
- PRs back to the software developer?