



# Cancer Analysis Workflow

Yes we can run it on Irma with Singularity containers

---

Maxime Garcia

2017-08-29



# What is Singularity?



- Docker-like containers technology
- Specific for HPC environment

# What is Singularity?



- Docker-like containers technology
- Specific for HPC environment
- Without the root user security problem

# What is Singularity?



- Docker-like containers technology
- Specific for HPC environment
- Without the root user security problem
- Supported by Nextflow

# What is Singularity?



- Docker-like containers technology
- Specific for HPC environment
- Without the root user security problem
- Supported by Nextflow
- Can pull containers from Docker-hub

## Why would we want to do that?

---

- For a better control

## Why would we want to do that?

---

- For a better control
- of every tool used
- of every version used

## Why would we want to do that?

---

- For a better control
- of every tool used
- of every version used
- For a better reproducibility

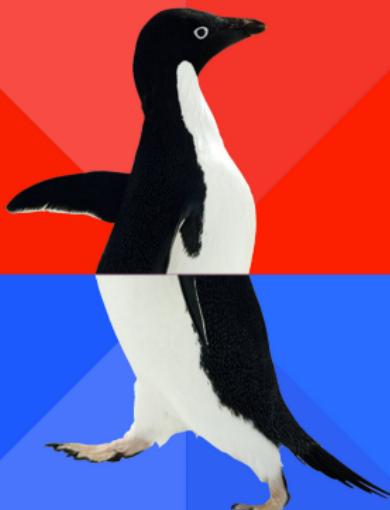
## Why would we want to do that?

---

- For a better control
- of every tool used
- of every version used
- For a better reproducibility
- Containerization is reproducible and portable

We can do everything!

WE CAN USE ANY VERSION  
OF ANY TOOL WE WANT



WE ARE RESPONSIBLE FOR  
EVERY TOOL WE NEED

# The CAW-containers repository

README.md



## Containers for CAW

[release v1.1](#) [license MIT License](#) [nextflow >0.23.2](#) [chat on gitter](#) [build passing](#)

CAW is a complete open source pipeline to detect somatic variants from WGS data developed at the [National Genomics Infrastructure](#) at [SciLifeLab Stockholm](#), Sweden.

For further information/help contact: [maxime.garcia@scilifelab.se](mailto:maxime.garcia@scilifelab.se) or [szilveszter.juhos@scilifelab.se](mailto:szilveszter.juhos@scilifelab.se), or join the gitter chat: [gitter.im/SciLifeLab/CAW](https://gitter.im/SciLifeLab/CAW)

### Documentation

This repo comes with documentation about the different containers, found in the `doc/` directory:

1. Documentation about containers
2. Documentation about building

### Author

• Maxime Garcia

---

# The One Script

---

SciLifeLab/CAW-containers

- 24 different Docker containers for all the different processes

# The One Script

---

SciLifeLab/CAW-containers

- 24 different Docker containers for all the different processes
- One Script to build them all,

# The One Script

---

SciLifeLab/CAW-containers

- 24 different Docker containers for all the different processes
- One Script to build them all,

One Script to push them

# The One Script

---

SciLifeLab/CAW-containers

- 24 different Docker containers for all the different processes
- One Script to build them all,

One Script to push them

- One Script to pull them all,

# The One Script

---

SciLifeLab/CAW-containers

- 24 different Docker containers for all the different processes
- One Script to build them all,

One Script to push them

- One Script to pull them all,

and with Singularity run them

## Inside the script

---

SciLifeLab/CAW-containers/main.nf

```
docker build -t $repository/$container:$tag \
$baseDir/containers/$container/.

docker push $repository/$container:$tag

singularity pull --name $container-${tag}.img \
docker://$repository/$container:$tag
```



- Nextflow natively support Singularity



- Nextflow natively support Singularity
- Easy configuration



- Nextflow natively support Singularity
- Easy configuration
- Automatic pull of the containers

# Easy configuration

SciLifeLab/CAW/configuration/singularity.config

```
singularity {
    autoMounts = true
    enabled = true
}
process {
    $RunFastQC.container      = 'docker://maxulysse/fastqc:1.1'
    $RunFreeBayes.container   = 'docker://maxulysse/freebayes:1.1'
    $RunGenotypeGVCFs.container = 'docker://maxulysse/gatk:1.1'
    $RunManta.container       = 'docker://maxulysse/runmanta:1.1'
    $RunMultiQC.container     = 'docker://maxulysse/multiqc:1.1'
    $RunMutect1.container     = 'docker://maxulysse/mutect1:1.1'
}
```

# Easy configuration

SciLifeLab/CAW/configuration/singularity.config

```
singularity {
    autoMounts = true
    enabled = true
}
process {
    $RunFastQC.container      = 'docker://maxulysse/fastqc:1.1'
    $RunFreeBayes.container   = 'docker://maxulysse/freebayes:1.1'
    $RunGenotypeGVCFs.container = 'docker://maxulysse/gatk:1.1'
    $RunManta.container       = 'docker://maxulysse/runmanta:1.1'
    $RunMultiQC.container     = 'docker://maxulysse/multiqc:1.1'
    $RunMutect1.container     = 'docker://maxulysse/mutect1:1.1'
}
process.container = 'shub://MaxUlysse/shubcontainer'
```

## What to do on a secure cluster?

---

- How about running CAW with Singularity on Bianca or Irma?

## What to do on a secure cluster?

---

- How about running CAW with Singularity on Bianca or Irma?
- No automatic pull of the containers

## What to do on a secure cluster?

---

- How about running CAW with Singularity on Bianca or Irma?
- No automatic pull of the containers
- The configuration is still easy

# Still an easy configuration

```
SciLifeLab/CAW/configuration/singularity-download.config
```

```
singularity {
    autoMounts = true
    enabled = true
}
process {
    $RunFastQC.container      = 'containers/fastqc-1.1.img'
    $RunFreeBayes.container   = 'containers/freebayes-1.1.img'
    $RunGenotypeGVCFs.container = 'containers/gatk-1.1.img'
    $RunManta.container       = 'containers/runmanta-1.1.img'
    $RunMultiQC.container     = 'containers/multiqc-1.1.img'
    $RunMutect1.container     = 'containers/mutect1-1.1.img'
}
```

## How to get the containers?

---

- Use the script that does the thing

## How to get the containers?

---

- Use the script that does the thing
- To pull all containers from Docker-hub into Singularity containers

## How to get the containers?

---

- Use the script that does the thing
- To pull all containers from Docker-hub into Singularity containers
- Transfer the containers to the secure cluster

## One last detail

---

- Do not forget to create the UPPMAX specific directories in the containers

## One last detail

---

- Do not forget to create the UPPMAX specific directories in the containers
- /pica

## One last detail

---

- Do not forget to create the UPPMAX specific directories in the containers
- /pica
- /proj

## One last detail

---

- Do not forget to create the UPPMAX specific directories in the containers
- /pica
- /proj
- /sw

# Let the magic happen

---

SciLifeLab/CAW/main.nf

```
nextflow run ~/CAW/main.nf --project project --step preprocessing \
--genome GRCh38 --sample sample.tsv -profile singularityLocal
```

# Using parallelization to do multiple tests

SciLifeLab/CAW/.travis.yml

```
^^Isudo: required
^^Ilanguage: java
^^Ijdk: openjdk8
^^Iservices:
^^I - docker
^^Ienv:
^^I - NXF_VER=0.25.6 SGT_VER=2.3.1 TEST=ANNOTATE PROFILE=singularity
^^I - NXF_VER=0.25.6 SGT_VER=2.3.1 TEST=ANNOTATE PROFILE=docker
^^I - NXF_VER=0.25.6 SGT_VER=2.3.1 TEST=RECALIBRATE PROFILE=singularity
^^I - NXF_VER=0.25.6 SGT_VER=2.3.1 TEST=RECALIBRATE PROFILE=docker
^^I - NXF_VER=0.25.6 SGT_VER=2.3.1 TEST=REALIGN PROFILE=singularity
^^I - NXF_VER=0.25.6 SGT_VER=2.3.1 TEST=REALIGN PROFILE=docker
^^I - NXF_VER=0.25.6 SGT_VER=2.3.1 TEST=MAPPING PROFILE=singularity
^^I - NXF_VER=0.25.6 SGT_VER=2.3.1 TEST=MAPPING PROFILE=docker

^^Iinstall: # Install Nextflow
^^I - "./scripts/install.sh --tool nextflow"

^^Iscript:
^^I - "./scripts/test.sh --profile $PROFILE --test $TEST --install --travisci"
```

# All your test are belong to us.

SciLifeLab / CAW  build passing

Current Branches Build History Pull Requests Build #433 More options 

✓ Pull Request #457 Update Manta and Strelka  #433 passed 

↳ Commit 2be368b   
↳ #457: Update Manta and Strelka   
↳ Branch master 

⌚ about 3 hours ago 

Maxime Garcia authored and committed 

Build Jobs

✓ # 433.1	 ↳ JDK: openjdk8	 NXF_VER=0.25.6 PROFILE=dockerTest TEST=ANNOTATION  22 min 12 sec	
✓ # 433.2	 ↳ JDK: openjdk8	 NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularit  23 min 57 sec	
✓ # 433.3	 ↳ JDK: openjdk8	 NXF_VER=0.25.6 PROFILE=dockerTest TEST=ANNOTATION  22 min 6 sec	
✓ # 433.4	 ↳ JDK: openjdk8	 NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularit  11 min 53 sec	
✓ # 433.5	 ↳ JDK: openjdk8	 NXF_VER=0.25.6 PROFILE=dockerTest TEST=RECALIBRATION  11 min 15 sec	
✓ # 433.6	 ↳ JDK: openjdk8	 NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularit  11 min 29 sec	
✓ # 433.7	 ↳ JDK: openjdk8	 NXF_VER=0.25.6 PROFILE=dockerTest TEST=REALIGNMENT  9 min 43 sec	
✓ # 433.8	 ↳ JDK: openjdk8	 NXF_VER=0.25.6 SGT_VER=2.3.1 PROFILE=singularit  14 min 13 sec	
✓ # 433.9	 ↳ JDK: openjdk8	 NXF_VER=0.25.6 PROFILE=dockerTest TEST=MAPPING  6 min 14 sec	

# The List of People Involved

---

Sebastian DiLorenzo

Jesper Eisfeldt

Maxime Garcia

Szilveszter Juhos

Max Käller

Malin Larsson

Marcel Martin

Monica Nistèr

Björn Nystedt

Pall Olason

Pelin Sahlén

Johanna Sandgren

Teresita Díaz De Ståhl

## Where to find us?

- We are on the SciLifeLab Slack  
#cancer-pipeline

## Where to find us?

- We are on the SciLifeLab Slack  
#cancer-pipeline
- We have a gitter channel  
<https://gitter.im/SciLifeLab/CAW>

## Where to find us?

- We are on the SciLifeLab Slack  
`#cancer-pipeline`
- We have a gitter channel  
<https://gitter.im/SciLifeLab/CAW>
- Our code is hosted on Github  
<https://github.com/SciLifeLab/CAW>

# Any questions?



# Any questions?

