**Building Singularity Image**

**1. installation of Singularity**:

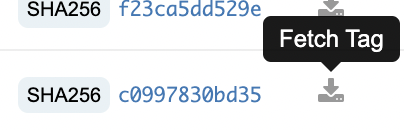
**1.1 Option 1. Installation Tutorial**

<https://docs.sylabs.io/guides/3.0/user-guide/installation.html>

**1.2 Option 2. Installation Using Docker**

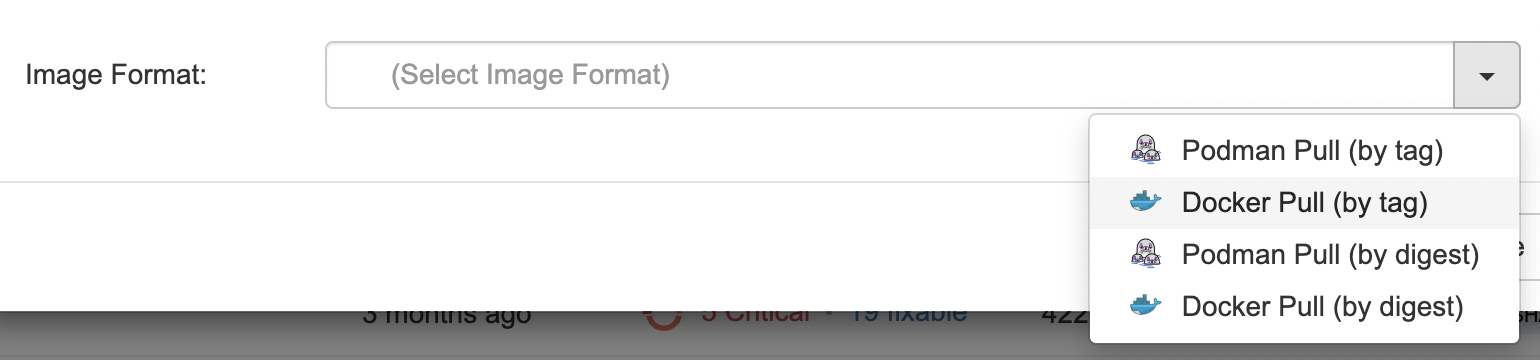
-If Docker is already installed onto your system, then you can install singularity by pulling through Docker

1.Go to [quay.io](https://quay.io/repository/singularity/singularity?tab=tags).



2.Pick the version of Singularity that you wish to install (you may have trouble installing versions past 3.10) and click fetch tag (the small gray button on the very right side of the screen)

* -slim is much smaller than the other versions and provides the same functions



3.Select Docker (pull by tag) and copy the resulting tag

4.Go to the Linux Directory that you wish to install Singularity in and paste the tag

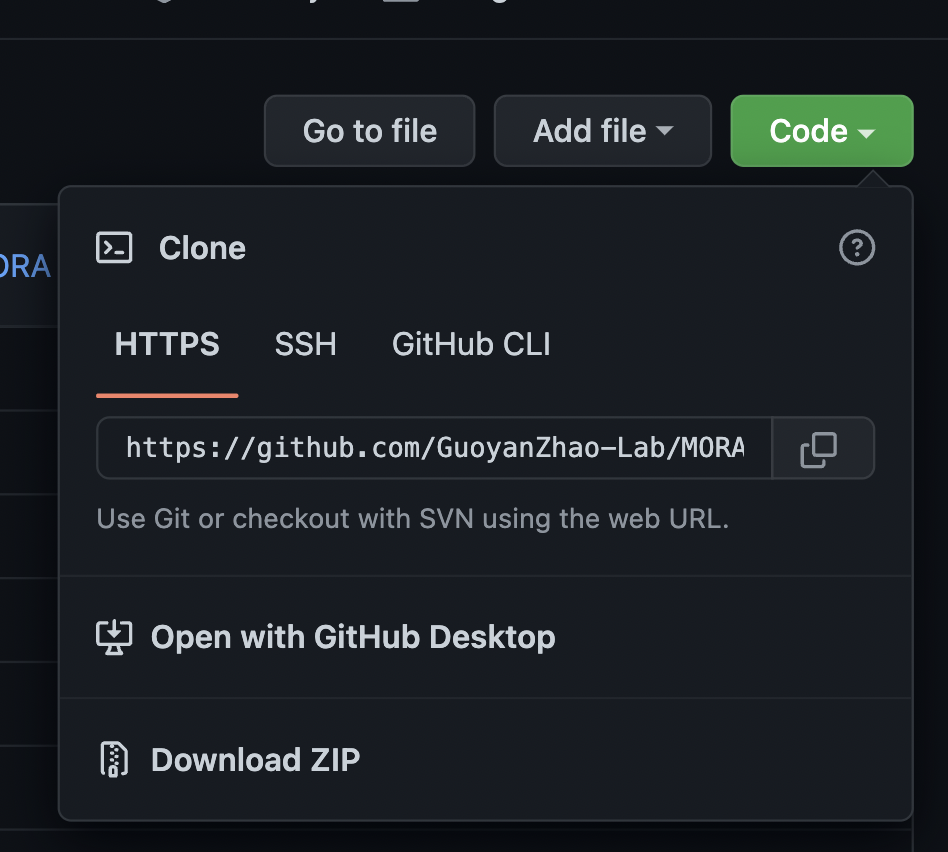
**2 General Tutorial** (recommended, but not required): <https://docs.sylabs.io/guides/3.0/user-guide/quick_start.html>

-If you wish to learn the basics of Singularity, use the above tutorial. Otherwise, follow the instructions on this Doc

**3 Downloading Necessary Files**

[Github Link](https://github.com/GuoyanZhao-Lab/MORA)

-Go to the Github Link and click the green button that says “Code”



-Copy the HTTPS link

-Navigate to the directory that you wish to place the files (ideally this would be the same folder

-run the command git clone [github link]

**4. Pipeline Directory Explained**

-the MORA\_Singularity directory contains the following 5 folders

Example Input:

-this folder contains the .fa and .con files that express both the specific genes we are examining and the background genes (which serve as a comparison to the specific genes)

MORA\_Pipeline\_v0.1:

This folder contains the following

-pipeline.sh: shell script that takes in user input before running the program

-MORA\_Singularity\_v0.1.pl: perl script that runs the entire pipeline

-all of the perl scripts and R scripts needed to run the pipeline

-all of the tools needed to run the pipeline (Patser and Alphabet)

-a README file

Database:

CISBP\_v2.00\_ForTesting:

Folder with smaller dataset for testing

CISBP\_v2.00\_HumanMouseRatCombined\_QCed\_DB\_Final: folder that contains .Matrix files that measures the relative occurrence of DNA nucleotides

-the other summarize the data in the matrix files in .txt files

Output

-this folder is currently empty except for a placeholder. When running the pipeline, the results will populate this folder

Temp

-this folder is currently empty except for a placeholder. When running the pipeline, many files that need to be temporarily stored will be placed here.

**5. Creating Definition File**

-create a definition file in the directory in which you want to create the singularity image (ideally this would be the parent directory of MORA\_Singularity

>Use the command touch def\_file.def

-This is the text for the original definition file, but a few edits must be made

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Bootstrap: docker

From: ubuntu:20.04

%files

/home/singularity/enrichment\_pipeline/MORA\_Singularity

%runscript

exec /home/louisli/tools/enrichment\_pipeline/MORA\_Singularity/MORA\_Pipeline\_v0.1/pipeline.sh

%post

apt-get update && apt-get install -y software-properties-common

add-apt-repository universe

apt-get update

apt-get install -y libparallel-forkmanager-perl

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**5.1 Instructions for replacing file path**:

5.1.1 %Ffles

-The file path to the MORA\_Singularity folder under %files must be replaced with the user’s unique path.

(The correct file path is **not** the direct file path to the folder).

-make sure that the

-Find the file path to the parent directory of the directory that contains the definition file. (we will call this directory A)

-Replace that file path with “home/singularity”

-navigate to the MORA\_Singularity folder

-all directories beyond directory A that lead to MORA\_Singularity must be appended to “home/singularity”

Example:

-the MORA\_Singularity folder has the file path /home/louisli/tools/enrichment\_pipeline/MORA

-the definition file is in /home/louisli/tools/enrichment\_pipeline

-the parent directory to this folder is /home/louisli/tools (This is Directory A)

-Replace /home/louisli/tools with home/singularity

-To get from Directory A to MORA\_Singularity we must add enrichment\_pipeline/MORA\_Singularity to the file path

-Therefore the correct file path will be home/singularity + enrichment\_pipeline/MORA\_Singularity = home/singularity/enrichment\_pipeline/MORA\_Singularity

5.1.2 %runscript

-The file path to the shell script under %runscript must also be changed to that of the user (use actual file path for this). Simply find the file path to pipeline.sh

**6.Building Singularity Image**

-If the definition file was edited correctly and the files were put into the correct places, then creating the image should be relatively simple

**6.1 Getting the Correct Build Command:**

docker run --privileged --rm -v ${PWD}:/home/singularity quay.io/singularity/singularity:v3.5.3-slim build /home/singularity/[path to desired folder]/pipeline.sif /home/singularity/[path to desired folder]/def\_file.def

-Ideally, the path to the definition file and the singularity image would be the same

-use the same rules of finding the file path as for finding the correct file path in the definition file under the %files section

-in summary, the parent directory of the directory that contains the definition file will be replaced with “home/singularity.” The name of the folder containing the definition file should be appended to this

Example:

docker run --privileged --rm -v ${PWD}:/home/singularity quay.io/singularity/singularity:v3.5.3-slim build /home/singularity/enrichment\_pipeline/pipeline.sif /home/singularity/enrichment\_pipeline/def\_file.def

-Definition file is located in the directory /home/louisli/tools/enrichment\_pipeline

-/home/louisli/tools becomes home/singularity.

-/enrichment\_pipeline is added to this followed by /pipeline.sif

NOTE: if you wish to create the singularity image in a different location than the definition file, you can do so by having a different file path for the .sif file.

**6.2 Running the Command**:

-Go to the parent directory of the directory in which the image will be created (Directory A from the section above)

-Run the command that you built in the previous subsection

If Errors Occur, Consider the Following:

-Were the Files copied in the correct location?

-Do the names of the files match what was put into the definition file?

-Does the name of my definition file match that of the given command?

-Was the command run in the correct location?

**Installing and Running the Pipeline**

This document will outline the various ways in which to install and run the MORA\_Singualarity\_v0.1 pipeline for motif enrichment analysis.

**7. Option 1: Downloading .tgz file**

-If one was sent the MORA\_Singularity\_v0.1.tgz file, they can download and unzip it to get access to all the pipeline’s code

NOTE: the .tgz file does not contain the Singularity image and thus, the user will have to enter a custom command into the command line.

**7.1 Unzipping the .tgz file**

1.Download the .tgz file

2.If the file is not in the desired location, use the mv command to move the .tgz file to the desired directory

mv [file path to MORA\_Singularity\_v0.1.tgz][desired\_directory]

3. Run the following command:

tar –xvzf MORA\_Singularity\_v0.1.tar.gz –C [directory containing .tgz file]

At this point, a folder called MORA\_Singularity\_v0.1 should appear with all of the subdirectories and files needed to run the directory

**7.2 Finding the Proper Command**

At this point, the User has all the necessary files for running the pipeline. However, each user will have to find the command that allows the program to run on their directory

7.2.1 Description of Command:

Note: Because each user’s file path is unique leading up to MORA\_Singularity\_v0.1, \* will be be put in front of each file path to denote the user’s path

**Command Template**

perl \*/MORA\_Singularity\_v0.1/MORA\_Singularity\_v0.1.pl <$in\_dir> <output\_dir> <$Specific\_gene\_file\_fa> <$FullBgGeneSeqFile> <$N\_simulation> <$Motif\_Database\_Dir> <$Motif\_Database\_File> <$temp\_dir> <$software path> <$software\_path> <step\_number>

$in\_dir: the path of the directory where all of the gene files are located

$output\_dir: the directory that the pipeline's output files get redirected to

$Specific\_gene\_file\_fa: the file of the specific genes in .fa format

$FullBgGeneSeqFile: the file path of the background genes in .fa format

$N\_simulation: the number of times a random number of genes is to be selected (also the number of .fa, .con, and ORI.txt files are produced). Optimal number is 100.

$Motif\_Database\_Dir: The path of the directory that contains all of the .matrix files

ex:

$Motif\_Database\_File: The path of the file containing matrix information

$temp\_dir: the directory that files will temporarily get written to

$software path: the path of the directory that contains the program

$step\_number: the step number that you want to run. Must be between 0 and 7. 0 will run all step

Sample Input:

perl \*/MORA\_Singularity\_v0.1/MORA\_Singularity\_v0.1.pl \*/MORA\_Singularity\_v0.1/ExampleInput/ \*/MORA\_Singularity\_v0.1/output/ \*/MORA\_Singularity\_v0.1/ExampleInput/JEM\_IRF1dep\_class1\_class3.fa \*/MORA\_Singularity\_v0.1/ExampleInput/BackgroundGene.fa 100 \*/MORA\_Singularity\_v0.1/database/CISBP\_v2.00\_ForTesting/CISBP\_v2.00\_HumanMouseRatCombined\_QCed\_DB\_Final \*/MORA\_Singularity\_v0.1/database/CISBP\_v2.00\_ForTesting/CISBP\_v2.00\_HumanMouseRatCombined\_QCed\_DB\_Final.txt / \*/MORA\_Singularity\_v0.1/temp \*/MORA\_Singularity\_v0.1/MORA\_Pipeline\_v0.1/ 0

**7.2.2 Adding Custom Data**

-If the user has their own data that they would like to add, they can do so. Although the user may place this data anywhere in the MORA\_Singularity\_v0.1 folder, information about where current files are will be provided below to assist the user with organization.

**\*/MORA\_Singularity\_v0.1/ExampleInput/**:Put information about genetics here. Ideally, .fa and .con files should be here

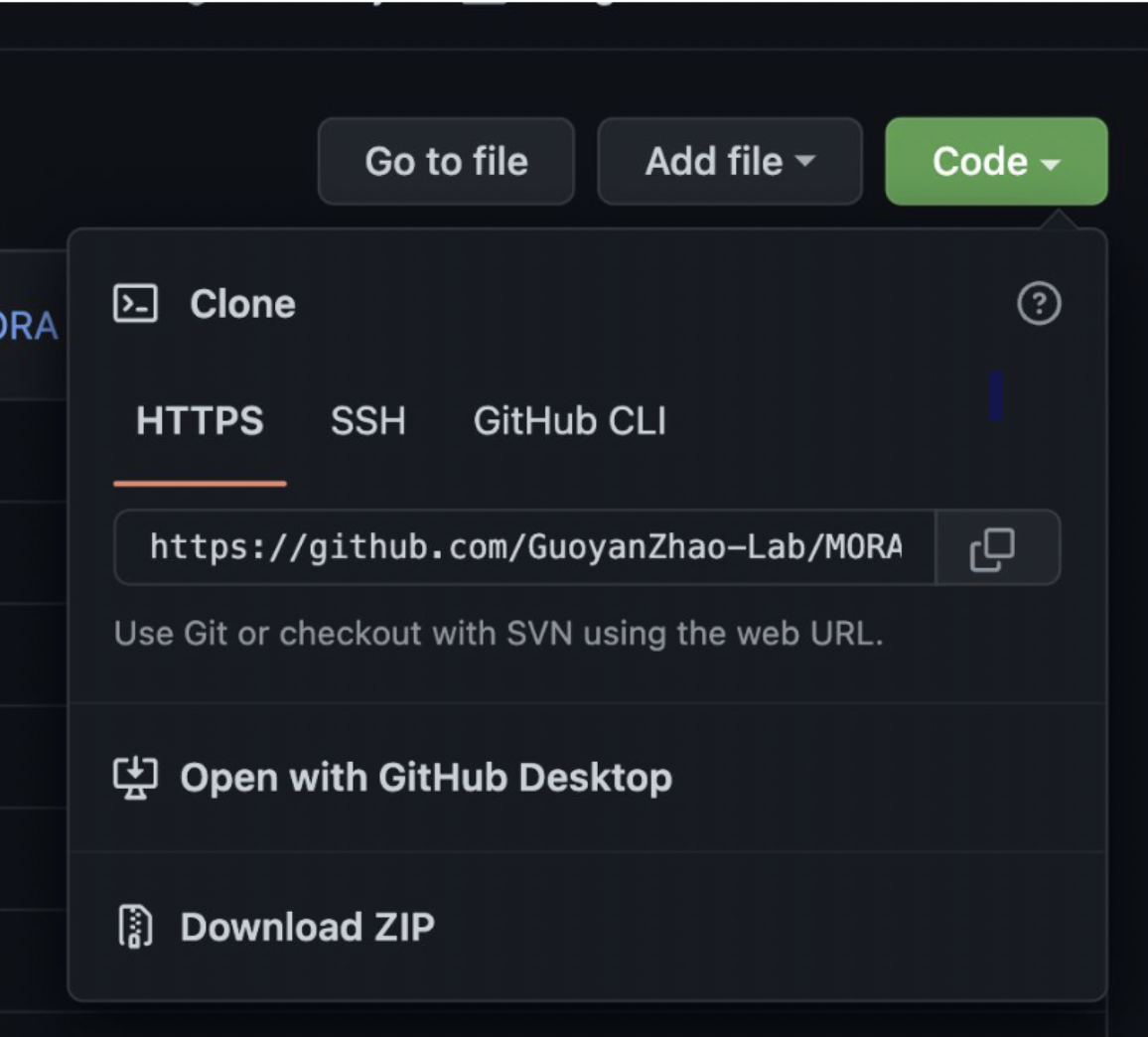
**\*/MORA\_Singularity\_v0.1/database/CISBP\_v2.00\_ForTesting:** put transcription motif data here. .txt files should go directly in this directory, while directories should be made for .matrix and .desc files

**8.Option 2: Using Github**

**8.1 Git Clone**

1.Go to the [Github Repository](https://github.com/GuoyanZhao-Lab/MORA-Singularity-)

2.click the green button that says “Code”



-copy the HTTPS text in the bar

3. Navigate to the directory that you wish to place the files

4. Run the command git clone [link]

**8.2 Running Singularity Image**

If you have singularity installed, you can run the pipeline.sif image. Singularity allows for a user to run a program and with its dependencies without having the files installed or giving a complete command.

1.Naviagate to the directory containing the singularity image

2.Run the command singularity run pipeline.sif

3.User will be asked to input the number of simulations followed by the step number.

-the ideal number of simulations is 100

-0 will run the whole pipeline, 1-7 will run a specific step number

**8.3 Running Through command prompt**

Using git clone will give the user access to all the files in the pipeline. This allows for the user to run the pipeline from the command line. Refer to section **1.2** for instructions on how to do this