



Docker for Public Health Bioinformatics

Week 3 Exercise

Contributing a Dockerfile to StaPH-B/docker-builds

Monday June 5th, 2023

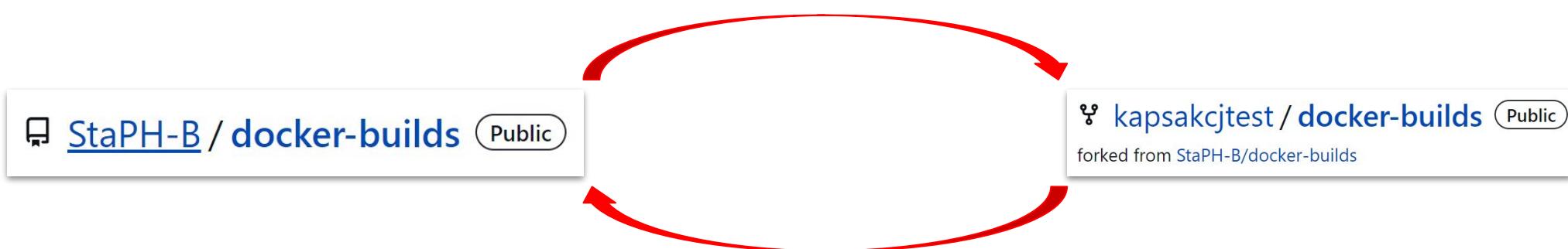
Curtis Kapsak, MS & Frank Ambrosio, MS | Theiagen Genomics

Homework!

- Now that we've learned strategies for write dockerfiles, let's put our knowledge to the test and write a new dockerfile.
- Let's share our dockerfiles & images with the community & contribute to the StaPH-B docker-builds project
 - <https://github.com/StaPH-B/docker-builds>

Outline of steps

1. Fork the StaPH-B/docker-builds repository to your personal GitHub account
2. Create new GitPod workspace with your forked repo. Use workspace to create & test new dockerfile. Commit code changes to your repo
3. Create Pull Request via GitHub so that you can contribute your new code back to StaPH-B/docker-builds repo



Background info

- StaPH-B/docker-builds github repository houses the Dockerfile code (& other code): <https://github.com/StaPH-B/docker-builds>

The screenshot shows the GitHub repository page for 'StaPH-B / docker-builds'. The repository is public and has 19 branches and 0 tags. The 'Code' tab is selected. A recent merge pull request from 'erinyoung' is highlighted. The repository's purpose is described as Dockerfiles and documentation on tools for public health bioinformatics. It uses Docker, Dockerfiles, bioinformatics, containers, and singularity. The repository has 136 stars, 11 watchers, and 89 forks.

Search or jump to... Pull requests Issues Codespaces Marketplace Explore

StaPH-B / docker-builds Public

Code Issues (33) Pull requests (12) Actions Projects (1) Security Insights

master 19 branches 0 tags Go to file Add file ▾

erinyoung Merge pull request #670 from StaPH-B/cjk-pangolin-update ... 1d0273f 4 days ago 2,302 commits

.github Updating finding new dockerfiles and building them to test (#499) 6 months ago

abricate Kgl abricate insaflu (#490) 7 months ago

any2fasta/0.4.2 Updated with Erin's recommendations 4 months ago

ariba/2.14.4 fixed ariva 2.14.4 dockerfile by installing specific version of pysam... 3 years ago

artic-ncov2019-epi2me/0.3.10 added line in artic-ncov2019-epi2me dockerfile so conda env is always... 2 years ago

artic-ncov2019-medaka Added specific git commit to Dockerfile 3 years ago

artic-ncov2019-nanopolish Added specific git commit to Dockerfile 3 years ago

artic-ncov2019 added dockerfile for artic-ncov2019 1.3.0-dev branch with latest meda... 2 years ago

About

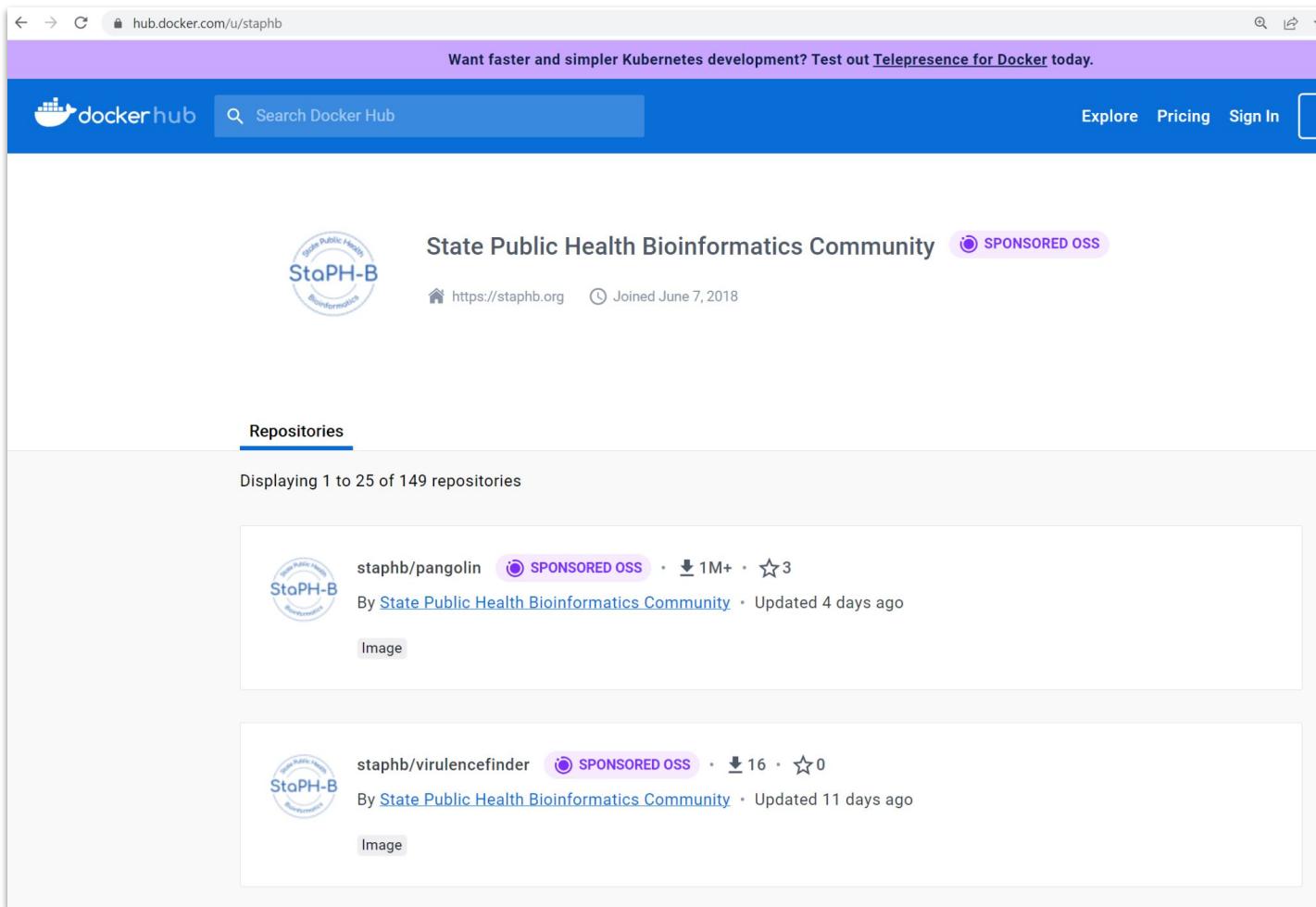
Dockerfiles and documentation on tools for public health bioinformatics

docker dockerfiles bioinformatics containers singularity

Readme GPL-3.0 license Activity 136 stars 11 watching 89 forks Report repository

Background info

- Docker images are hosted on both [dockerhub](#) and [quay.io](#)



Background info

- Each docker image has its own repository on dockerhub & quay.io

The screenshot shows a web browser displaying the Docker Hub website at hub.docker.com/r/staphb/pangolin. The page is for the Docker image `staphb/pangolin`, which is described as a "Software package for assigning SARS-CoV-2 genome sequences to global lineages". The repository was updated 4 days ago by the [State Public Health Bioinformatics Community](#). The Dockerfile and source code are available on [Github](#). A "Docker Pull Command" button contains the command `docker pull staphb/pangolin`. The "Overview" tab is selected, showing a brief description of the repository's purpose and instructions for contributing.

Want faster and simpler Kubernetes development? Test out [Telepresence for Docker](#) today.

Explore Pricing Sign In Register

staphb/pangolin SPONSORED OSS ☆ Pulls 1M+

By [State Public Health Bioinformatics Community](#) • Updated 4 days ago

Software package for assigning SARS-CoV-2 genome sequences to global lineages

Image

Overview Tags

docker-builds

This repository contains the Dockerfiles and other assorted files necessary for building Docker images for a variety of programs used by members of the StaPH-B (State Public Health Lab Bioinformatics) consortium. The purpose of this repository is to provide a centralized location for Docker images that is easily accessible for users, with clear documentation on how the containers were built and how to use them.

If you would like to contribute with a Docker image or improve upon the existing images, please fork the repository, make your changes/additions, and submit a pull request. If you are having an issue with an existing image, please submit an issue. We welcome any and all feedback! [See more details on how to contribute here](#)

Docker User Guide

Docker Pull Command

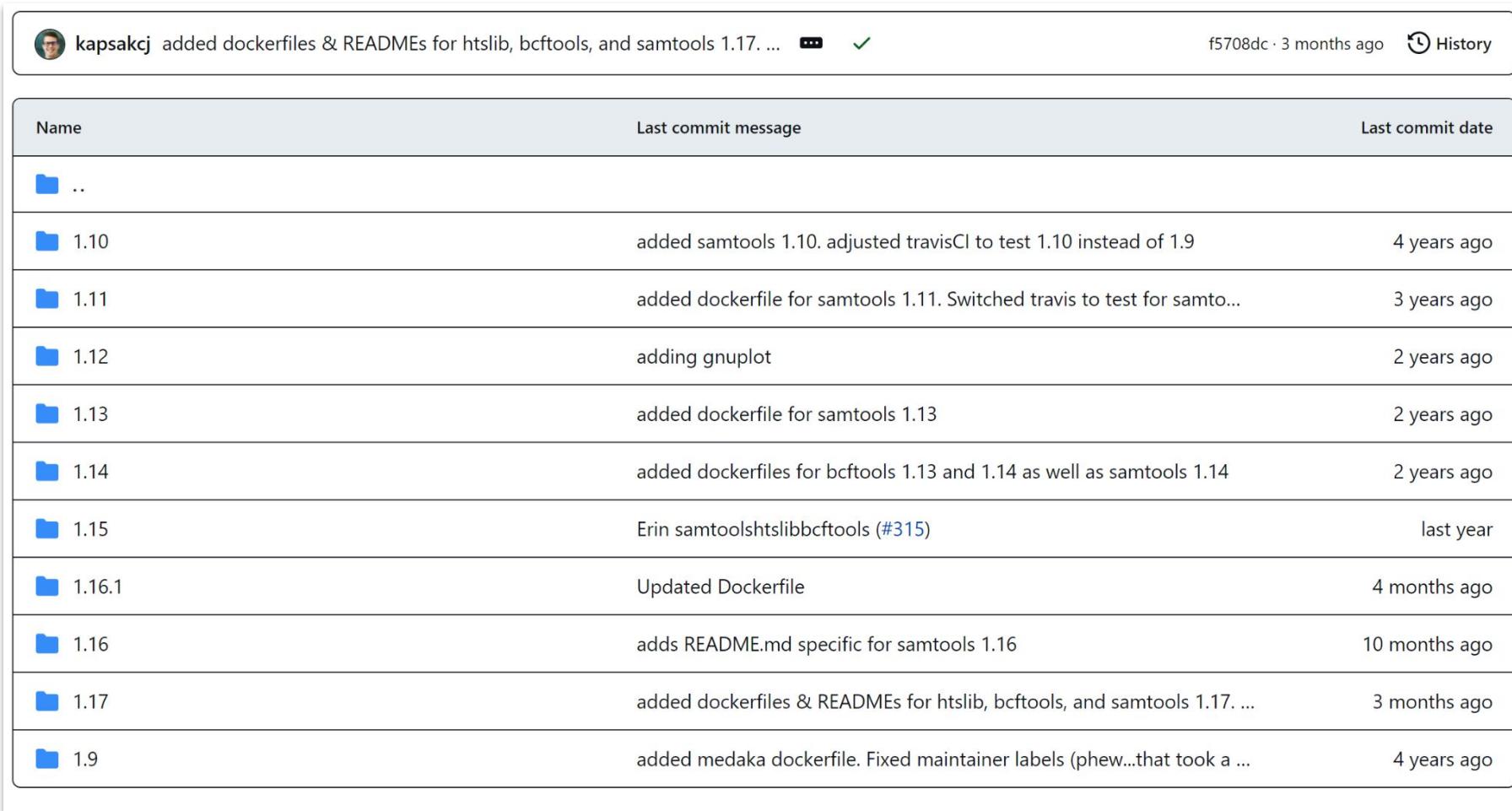
```
docker pull staphb/pangolin
```

Source Repository

Github StaPH-B/docker-builds

Background info

- Each tool has a single directory, and subdirectories are usually divided by that tool's version



The screenshot shows a GitHub commit history for a repository. At the top, a commit from user 'kapsakcj' is shown, adding dockerfiles & READMEs for htslib, bcftools, and samtools 1.17. The commit was made 3 months ago and includes a link to the history.

Name	Last commit message	Last commit date
..		
1.10	added samtools 1.10. adjusted travisCI to test 1.10 instead of 1.9	4 years ago
1.11	added dockerfile for samtools 1.11. Switched travis to test for samto...	3 years ago
1.12	adding gnuplot	2 years ago
1.13	added dockerfile for samtools 1.13	2 years ago
1.14	added dockerfiles for bcftools 1.13 and 1.14 as well as samtools 1.14	2 years ago
1.15	Erin samtoolshtslibbcftools (#315)	last year
1.16.1	Updated Dockerfile	4 months ago
1.16	adds README.md specific for samtools 1.16	10 months ago
1.17	added dockerfiles & READMEs for htslib, bcftools, and samtools 1.17. ...	3 months ago
1.9	added medaka dockerfile. Fixed maintainer labels (phew...that took a ...	4 years ago

Background info

- Each tool version subdirectory usually has a Dockerfile, a README.md and sometimes additional files

The screenshot shows a GitHub repository interface. At the top, there's a navigation bar with a repository icon, a dropdown menu showing "master", the repository name "docker-builds / samtools / 1.17 /", a search bar with "Go to file", an "Add file" button, and a three-dot menu. Below this is a commit history card for a commit by "kapsakcj" made 3 months ago. The commit message is "added dockerfiles & READMEs for htslib, bcftools, and samtools 1.17. ...". The commit hash is "f5708dc". To the right of the commit message are "History" and a "History" link. Below the commit card is a table listing files in the "1.17" directory. The columns are "Name", "Last commit message", and "Last commit date". The table contains three rows: a folder named "..." (Last commit message: "added dockerfiles & READMEs for htslib, bcftools, and samtools 1.17. ...", Last commit date: "3 months ago"), a "Dockerfile" (Last commit message: "added dockerfiles & READMEs for htslib, bcftools, and samtools 1.17. ...", Last commit date: "3 months ago"), and a "README.md" (Last commit message: "added dockerfiles & READMEs for htslib, bcftools, and samtools 1.17. ...", Last commit date: "3 months ago"). Below the table is a "README.md" file content editor with a "pencil" icon and a three-dot menu. The content of the README.md file is titled "samtools container" and includes a "Main tool:" section with two bullet points: "https://www.htslib.org/" and "GitHub".

Name	Last commit message	Last commit date
...	added dockerfiles & READMEs for htslib, bcftools, and samtools 1.17. ...	3 months ago
Dockerfile	added dockerfiles & READMEs for htslib, bcftools, and samtools 1.17. ...	3 months ago
README.md	added dockerfiles & READMEs for htslib, bcftools, and samtools 1.17. ...	3 months ago

samtools container

Main tool:

- <https://www.htslib.org/>
- GitHub

Step 1: Create a fork

- Fork the StaPH-B/docker-builds repository to your personal account

The screenshot shows a GitHub repository page for 'StaPH-B / docker-builds'. The top navigation bar includes links for Pull requests, Issues, Codespaces, Marketplace, and Explore. Below the navigation is a search bar and a 'Fork' button, which is highlighted with a red box. The main content area displays the repository's code, issues (33), pull requests (12), actions, projects (1), security, and insights. On the left, there are buttons for switching between branches ('master'), viewing 19 branches, and 0 tags. The right side features an 'About' section with a description of Dockerfiles and documentation for tools for public health bioinformatics, and a list of tags: docker, dockerfiles, bioinformatics, containers, and singularity. A 'Readme' link is also present.

github.com/StaPH-B/docker-builds

Search or jump to...

Pull requests Issues Codespaces Marketplace Explore

StaPH-B / docker-builds Public

Edit Pins Watch 11 Fork 89 Starred 136

Code Issues (33) Pull requests (12) Actions Projects (1) Security Insights

master ▾ 19 branches 0 tags Go to file Add file ▾ Code ▾

erinyoung Merge pull request #670 from StaPH-B/cj... 1d0273f 4 days ago 2,302 commits

.github Updating finding new dockerfiles and building them t... 6 months ago

abricate Kgl abricate insaflu (#490) 7 months ago

any2fasta/0.4.2 Updated with Erin's recommendations 4 months ago

About

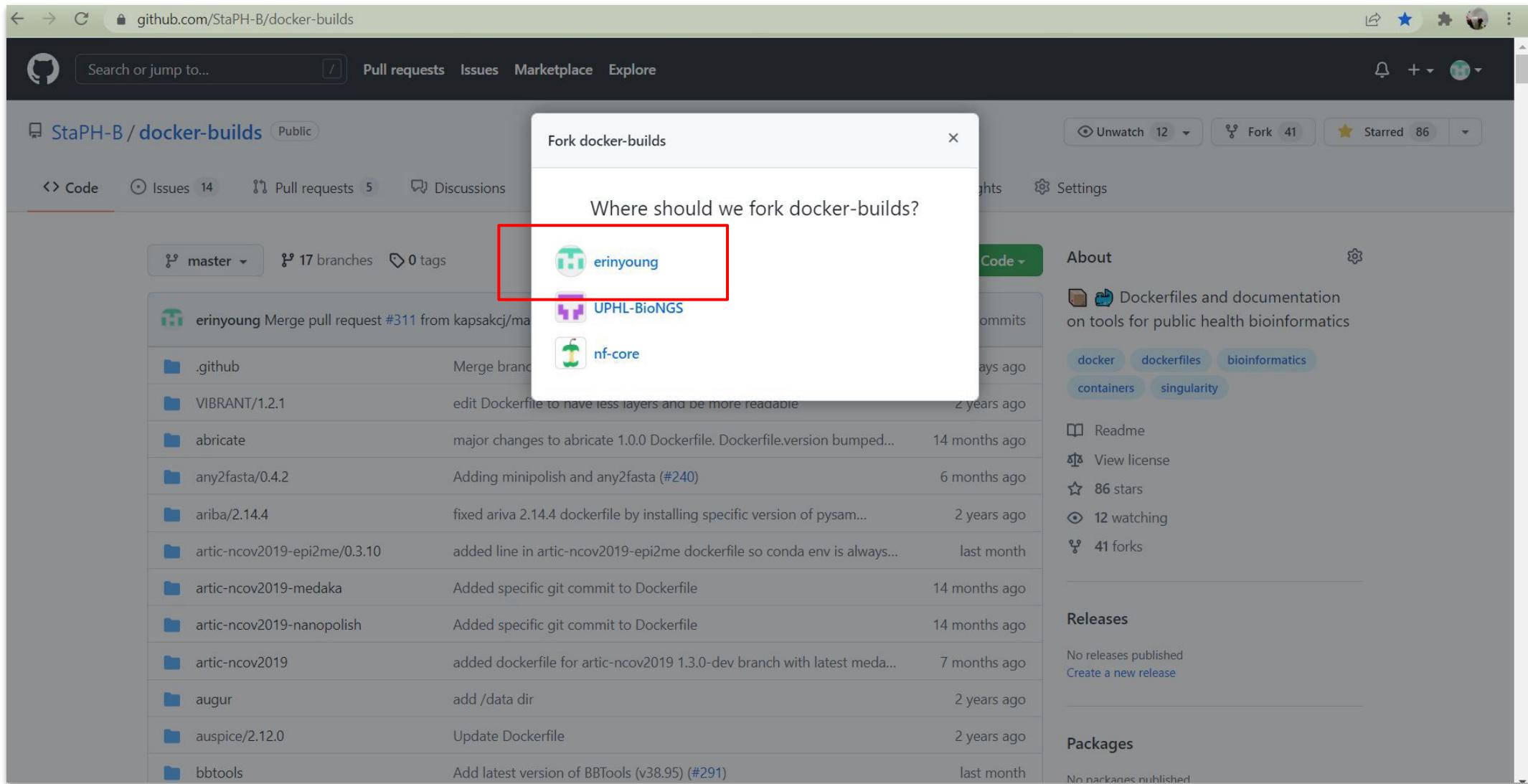
Dockerfiles and documentation on tools for public health bioinformatics

docker dockerfiles bioinformatics
containers singularity

Readme

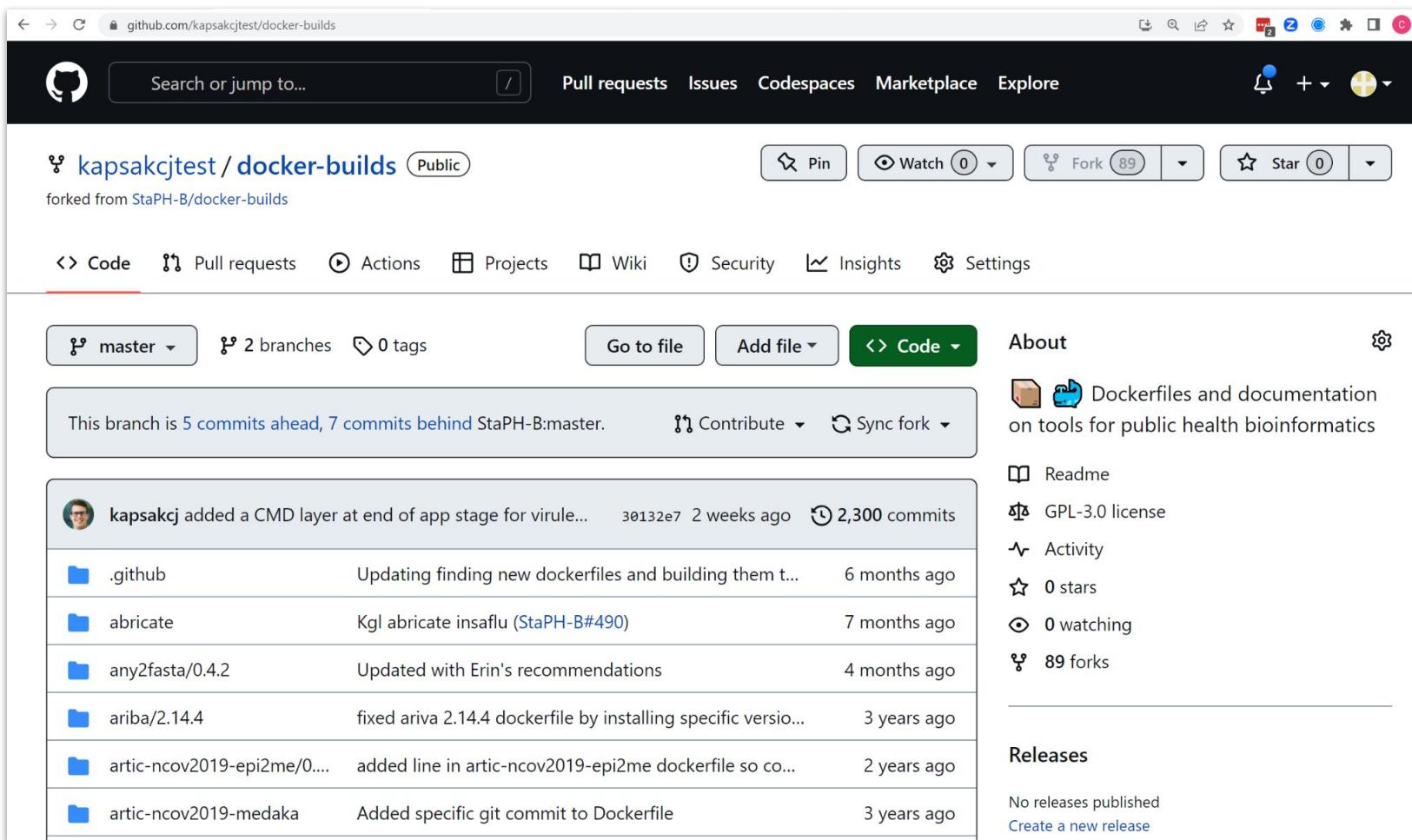
Step 1: Create a fork

- Fork the StaPH-B/docker-builds repository to your personal account



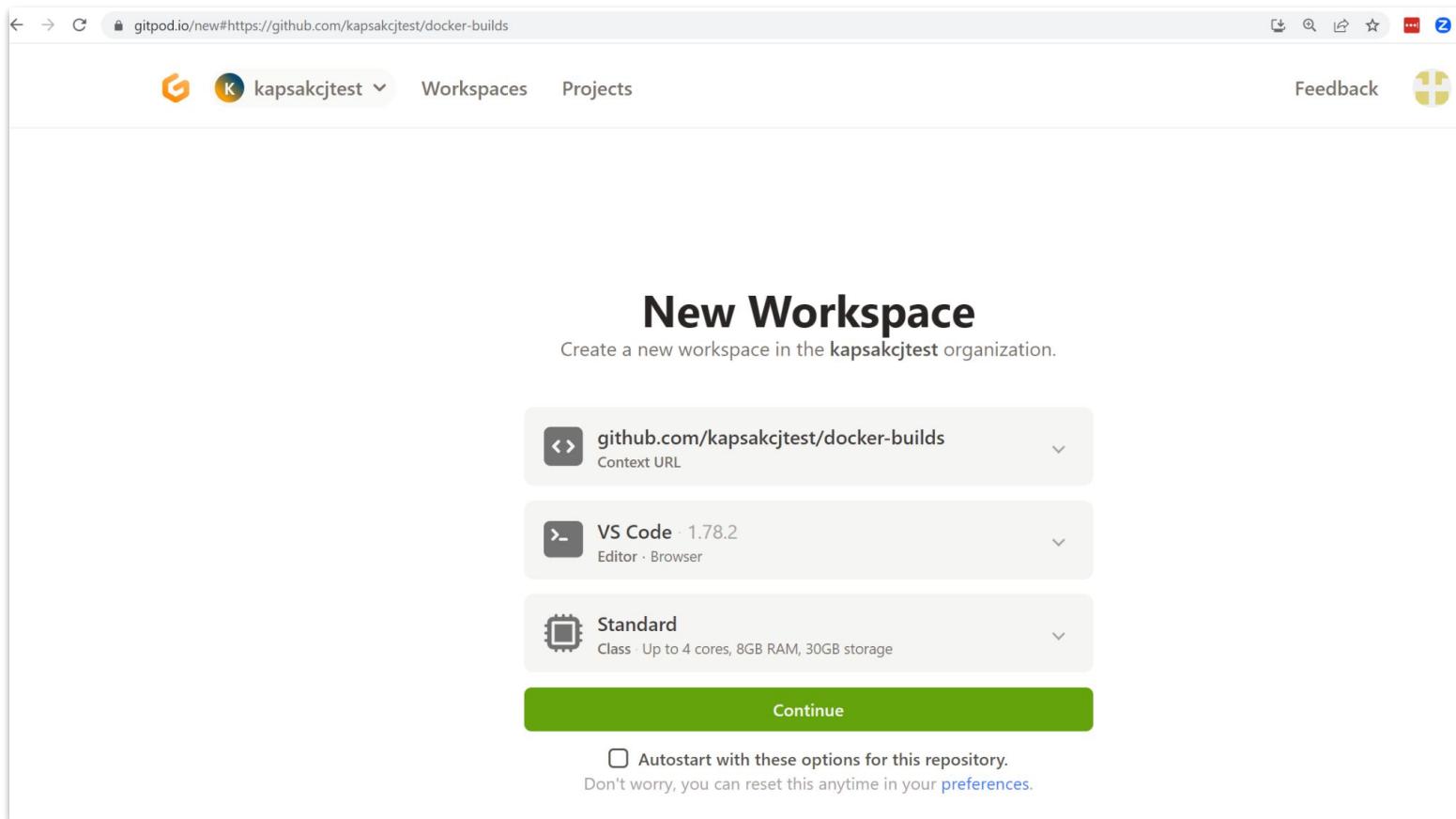
Step 1: Create a fork

- Now you have your own copy of the repository where you can freely make changes, without affecting the source repo



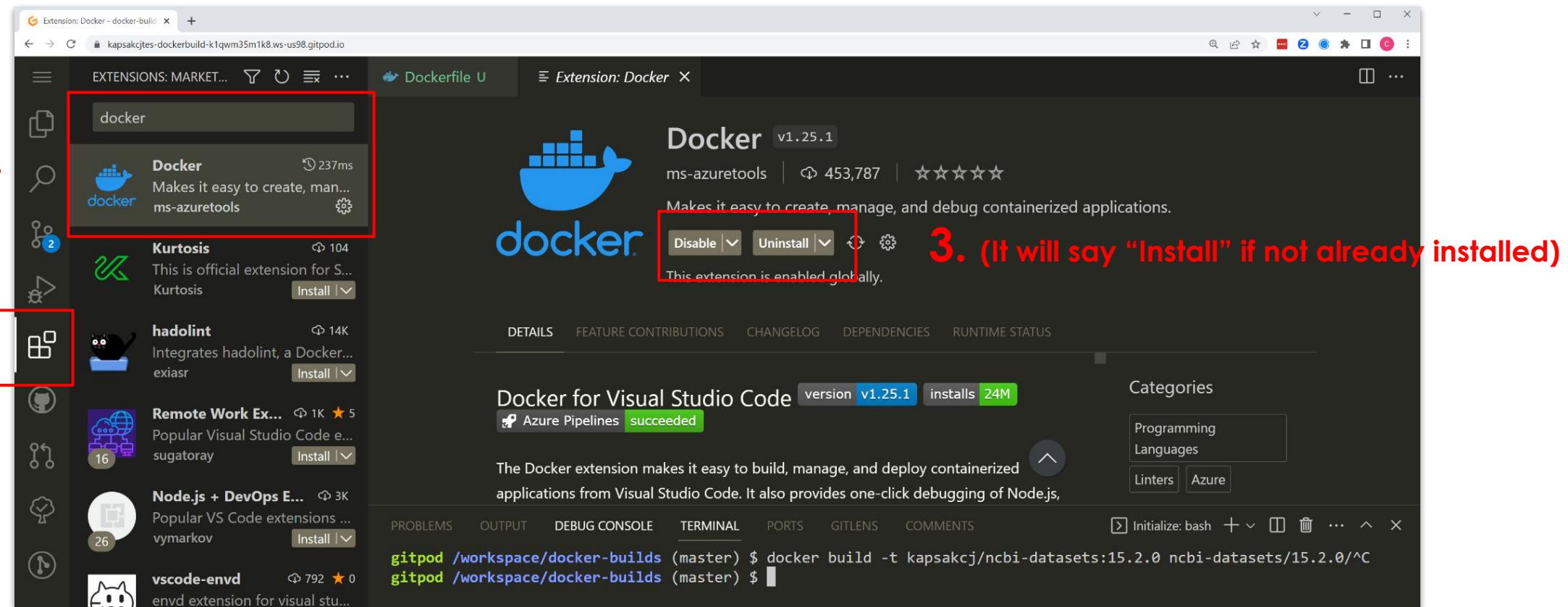
Step 2: Create new GitPod workspace

- Create a new GitPod workspace using your forked repo as the URL context. The URL will be specific to your github username. <https://github.com/<your-username>/docker-builds>
- example: <https://github.com/kapsakcjtest/docker-build>



Step 2: Create new GitPod workspace

- Open your new GitPod workspace
- Before getting started, add the “Docker” VSCode extension (author: ms-azuretools) to your GitPod environment
- Click on the Extension button on left
- search for “Docker”, click on the first option, and click the install button



Step 2: Create new GitPod workspace

- Write & test your new dockerfile in GitPod using the editor and `docker build` command in the terminal
- Example:

adding new dockerfile
for NCBI datasets v15.2.0

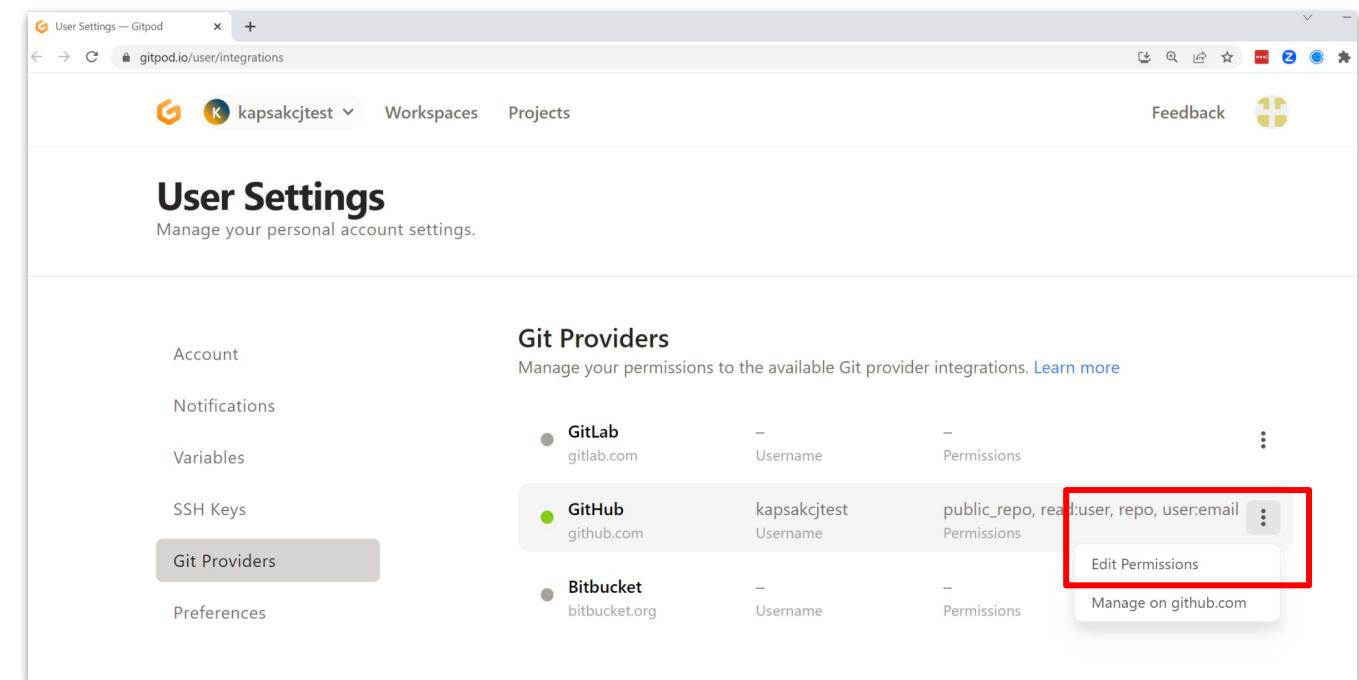
The screenshot shows a GitPod workspace interface. On the left is the Explorer sidebar, which lists several Docker builds and a specific ncbi-datasets folder containing subfolders from 13.31.0 to 14.20.0, each with its own Dockerfile and README.md. The ncbi-datasets/15.2.0 folder is currently selected. The main area is the Dockerfile editor, displaying a Dockerfile for version 15.2.0. The terminal at the bottom shows the command `gitpod /workspace/docker-builds (master) \$ docker build -t kapsakcj/ncbi-datasets:15.2.0 ncbi-datasets/15.2.0` being run.

```
FROM ubuntu:jammy as app
ARG DATASETS_VER="15.2.0"
LABEL base.image="ubuntu:jammy"
LABEL dockerfile.version="1"
LABEL software="NCBI's datasets and dataformat"
LABEL software.version="${DATASETS_VER}"
LABEL description="Downloads biological sequence data from NCBI"
LABEL website="https://www.ncbi.nlm.nih.gov/datasets/docs/v1/"
LABEL license="https://github.com/ncbi/datasets/blob/master/pkgs/ncbi-datasets-cli/LICENSE.md"
LABEL maintainer="Erin Young"
LABEL maintainer.email="eriny@utah.gov"
LABEL maintainer2="Curtis Kapsak"
LABEL maintainer2.email="kapsakcj@gmail.com"
# unzip isn't needed for datasets/dataformat, but it is often used after downloading files with datasets
# RUN apt-get update && apt-get install -y --no-install-recommends curl
```

gitpod /workspace/docker-builds (master) \$ docker build -t kapsakcj/ncbi-datasets:15.2.0 ncbi-datasets/15.2.0

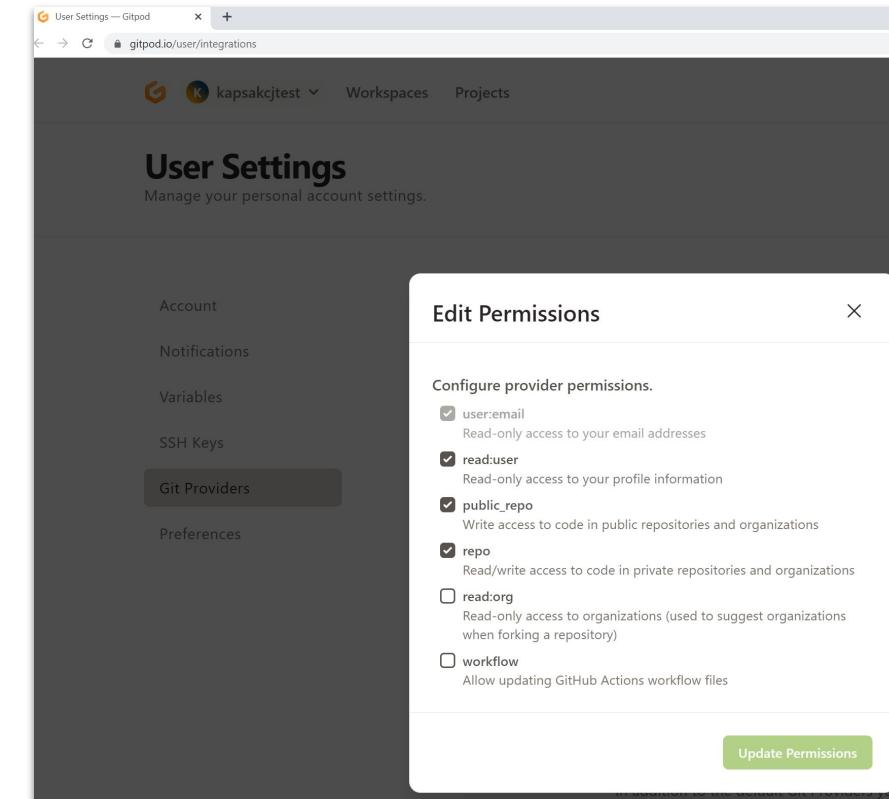
Step 2: Create new GitPod workspace

- Once your dockerfile builds successfully & passes the tests, save the file & commit the changes to your repo
- NOTE: You will have to give GitPod permissions to push commits on your behalf. Grant it permissions by navigating to <https://gitpod.io/user/integrations>
- Click three dots to the right of GitHub settings, then click “Edit Permissions”



Step 2: Create new GitPod workspace

- Once your dockerfile builds successfully & passes the tests. Save the file & commit the changes to your repo
- NOTE: You will have to give GitPod permissions to push commits on your behalf. Grant it permissions by navigating to <https://gitpod.io/user/integrations>
- Click three dots to the right of GitHub settings, then click “Edit Permissions”
- Check the boxes for:
 - read:user
 - public_repo
 - repo
- Click Update Permissions button



A few more things to add

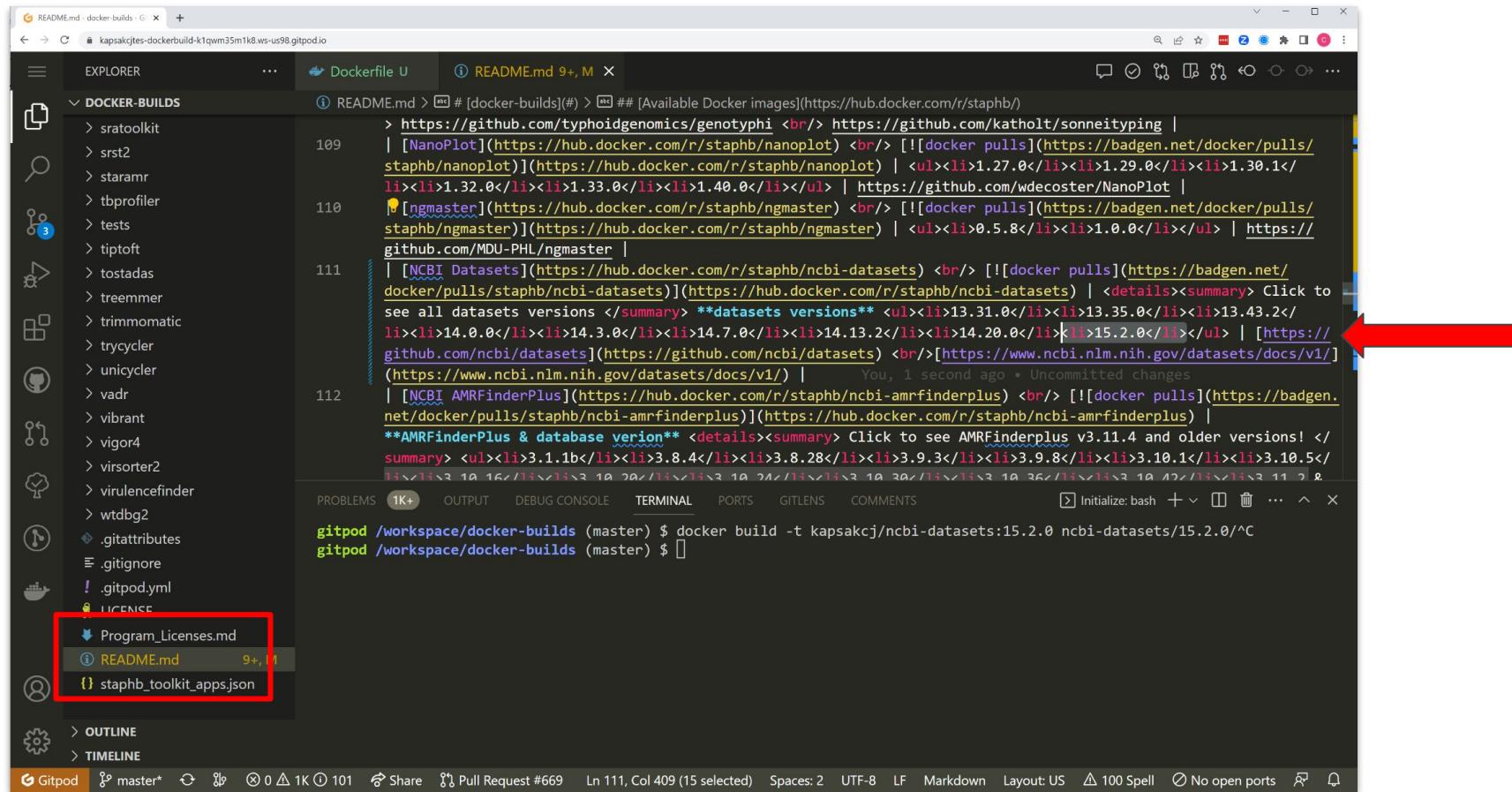
1. StaPH-B docker-builds requires a few more changes to be made alongside a dockerfile.
2. Please update:
 - a. The main **/README.md** with the new version (or a new row in the table for a new tool)
 - b. **/Program_Licences.md**
 - c. Add a short **README.md** for the tool
 - i. example: **ncbi-datasets/15.2.0/README.md**
 - ii. You can copy&paste the copy from the previous version if it exists

A few more things to add

1. Update the main `/README.md` with the new version (or a new row in the table for a new tool)
 - a. For tools already in the repo, add a new bullet point in the table with the new version.
 - b. For a tool that is new to the repo, add a new line.
 - i. Pro tip: copy and paste an existing line, and update the contents

A few more things to add

1. Update the main `/README.md` with the new version (or a new row in the table for a new tool)



```
> https://github.com/typoidgenomics/genotypiph |  
| [NanoPlot](https://hub.docker.com/r/staphb/nanoplot) <br/> [!docker pulls](https://badgen.net/docker/pulls/  
staphb/nanoplot)](https://hub.docker.com/r/staphb/nanoplot) | <ul><li>1.27.0</li><li>1.29.0</li><li>1.30.1</li>  
<li>1.32.0</li><li>1.33.0</li><li>1.40.0</li></ul> | https://github.com/wdecoester/NanoPlot |  
| [ngmaster](https://hub.docker.com/r/staphb/ngmaster) <br/> [!docker pulls](https://badgen.net/docker/pulls/  
staphb/ngmaster)](https://hub.docker.com/r/staphb/ngmaster) | <ul><li>0.5.8</li><li>1.0.0</li></ul> | https://  
github.com/MDU-PHL/ngmaster |  
| [NCBI Datasets](https://hub.docker.com/r/staphb/ncbi-datasets) <br/> [!docker pulls](https://badgen.net/  
docker/pulls/staphb/ncbi-datasets)](https://hub.docker.com/r/staphb/ncbi-datasets) | <details><summary> Click to  
see all datasets versions </summary> **datasets versions** <ul><li>13.31.0</li><li>13.35.0</li><li>13.43.2</li>  
<li>14.0.0</li><li>14.3.0</li><li>14.7.0</li><li>14.13.2</li><li>14.20.0</li><li>15.2.0</li></ul> | [https://  
github.com/ncbi/datasets](https://github.com/ncbi/datasets) <br/> [<a href="https://www.ncbi.nlm.nih.gov/datasets/docs/v1/">https://www.ncbi.nlm.nih.gov/datasets/docs/v1/] | You, 1 second ago + Uncommitted changes  
| [NCBI AMRFinderPlus](https://hub.docker.com/r/staphb/ncbi-amrfinderplus) <br/> [!docker pulls](https://badgen.  
net/docker/pulls/staphb/ncbi-amrfinderplus)](https://hub.docker.com/r/staphb/ncbi-amrfinderplus) |  
**AMRFinderPlus & database verion** <details><summary> Click to see AMRFinderplus v3.11.4 and older versions! </summary>  
<ul><li>3.1.1b</li><li>3.8.4</li><li>3.8.28</li><li>3.9.3</li><li>3.9.8</li><li>3.10.1</li><li>3.10.5</li>  
</ul>
```

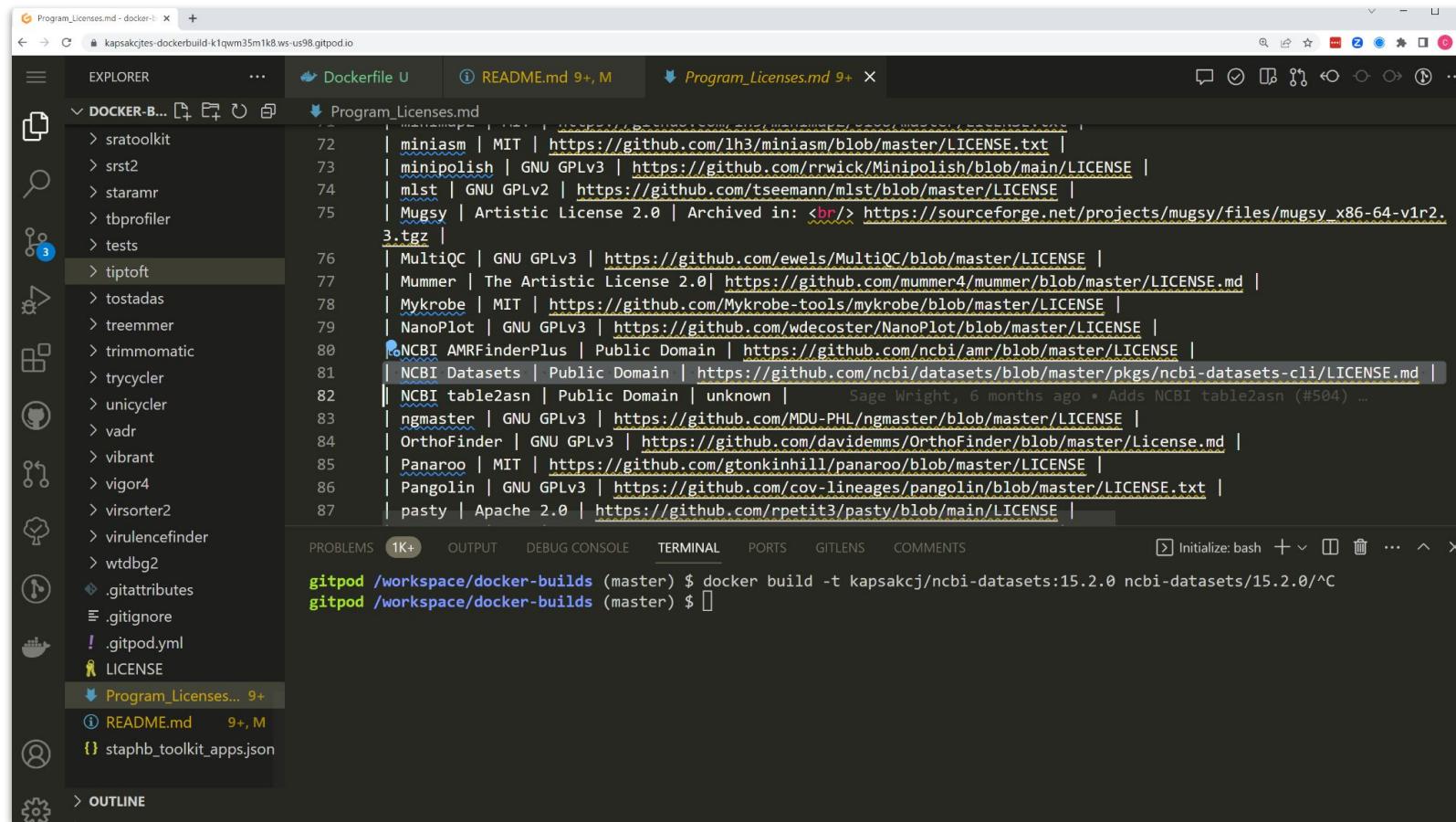
gitpod /workspace/docker-builds (master) \$ docker build -t kapsakcj/ncbi-datasets:15.2.0 ncbi-datasets/15.2.0/^C
gitpod /workspace/docker-builds (master) \$

Gitpod master* 0 1K 101 Share Pull Request #669 Ln 111, Col 409 (15 selected) Spaces: 2 UTF-8 LF Markdown Layout: US △ 100 Spell No open ports

How to add a new bullet point
for the new version in the table

A few more things to add

1. Ensure **/Program_Licences.md** lists the license info for the tool
 - a. NCBI-datasets is already added, no need to adjust this file



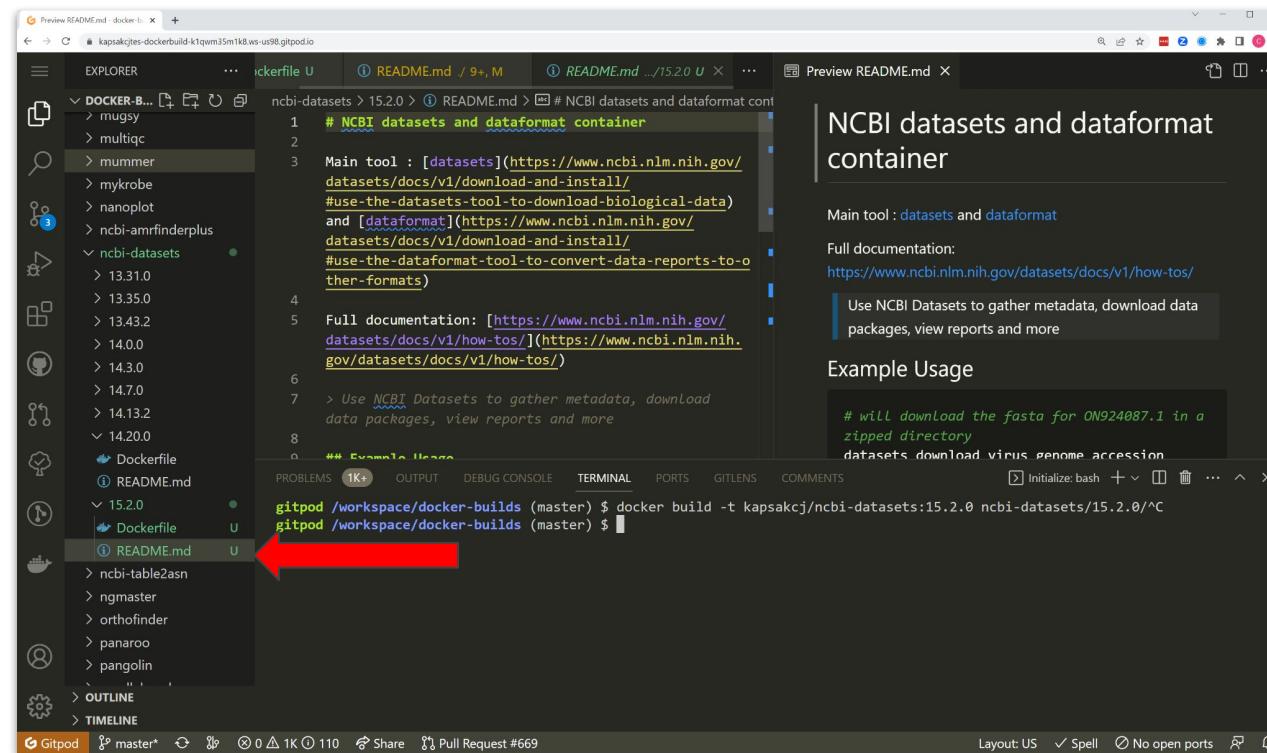
```
gitpod /workspace/docker-builds (master) $ docker build -t kapsakcj/ncbi-datasets:15.2.0 ncbi-datasets/15.2.0/^C
gitpod /workspace/docker-builds (master) $ [redacted]
```



A few more things to add

1. Add a short **README.md** for the tool
 - a. example: **ncbi-datasets/15.2.0/README.md**
 - b. The README.md can be re-used & updated from the previous version's README.md:

ncbi-datasets/14.20.0/README.md



```
# NCBI datasets and dataformat container
Main tool : [datasets](https://www.ncbi.nlm.nih.gov/datasets/docs/v1/download-and-install/#use-the-datasets-tool-to-download-biological-data) and [dataformat](https://www.ncbi.nlm.nih.gov/datasets/docs/v1/download-and-install/#use-the-dataformat-tool-to-convert-data-reports-to-other-formats)

Full documentation: [https://www.ncbi.nlm.nih.gov/datasets/docs/v1/how-tos/](https://www.ncbi.nlm.nih.gov/datasets/docs/v1/how-tos/)

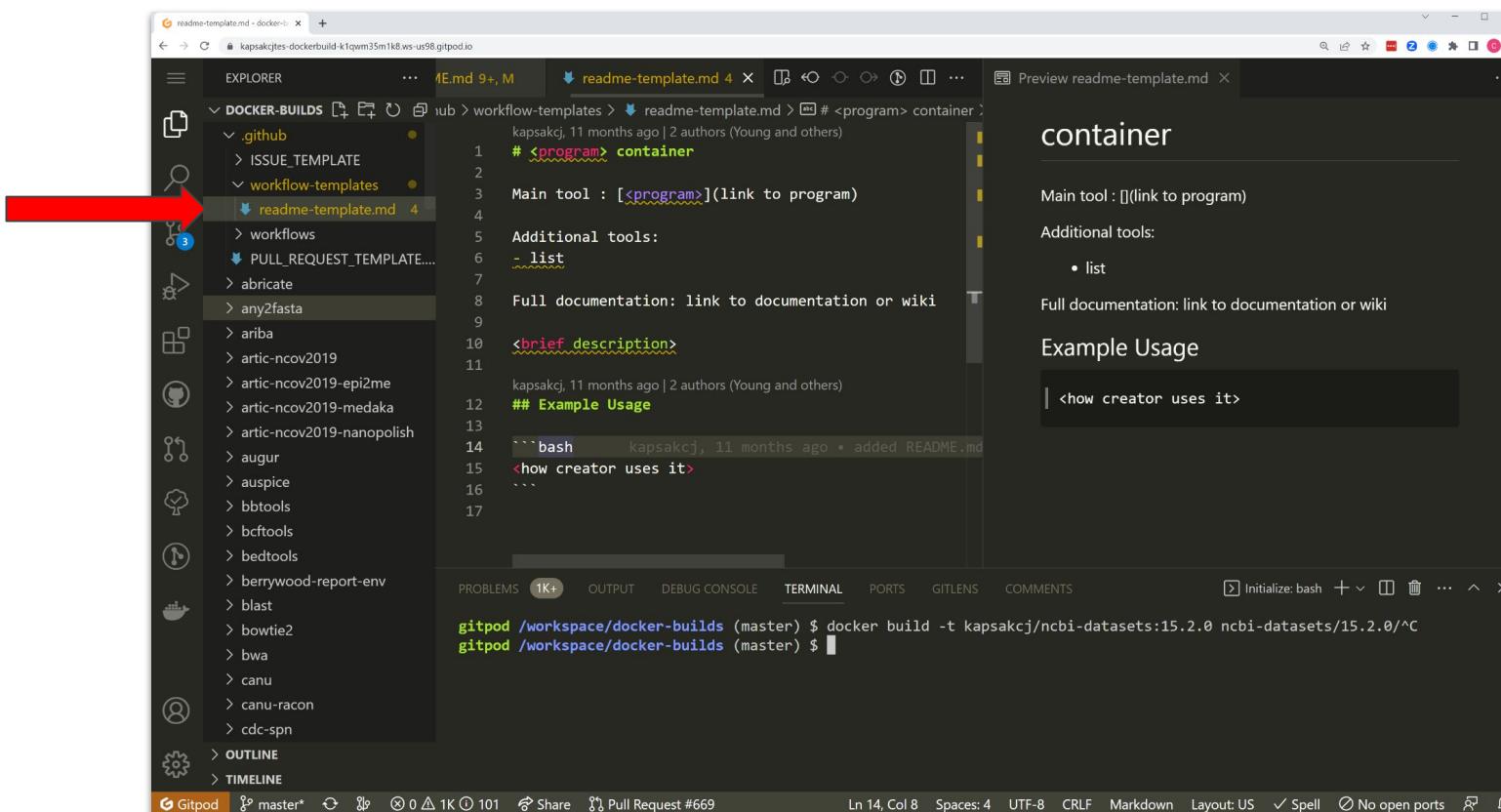
Use NCBI Datasets to gather metadata, download data packages, view reports and more

## Example Usage
# will download the fasta for ON924087.1 in a zipped directory
datasets.download.virus.genome.accession
```

The screenshot shows a Gitpod interface with the README.md file for the ncbi-datasets/14.20.0 repository. The file content is displayed in the main editor area, detailing the use of NCBI Datasets and DataFormat containers. A red arrow points to the README.md file in the Explorer sidebar, highlighting its location.

A few more things to add

1. Add a short **README.md** for the tool
 - a. If adding a new README.md, StaPH-B has a template available to get started
 - b. it is located at: <.github/workflow-templates/readme-template.md>



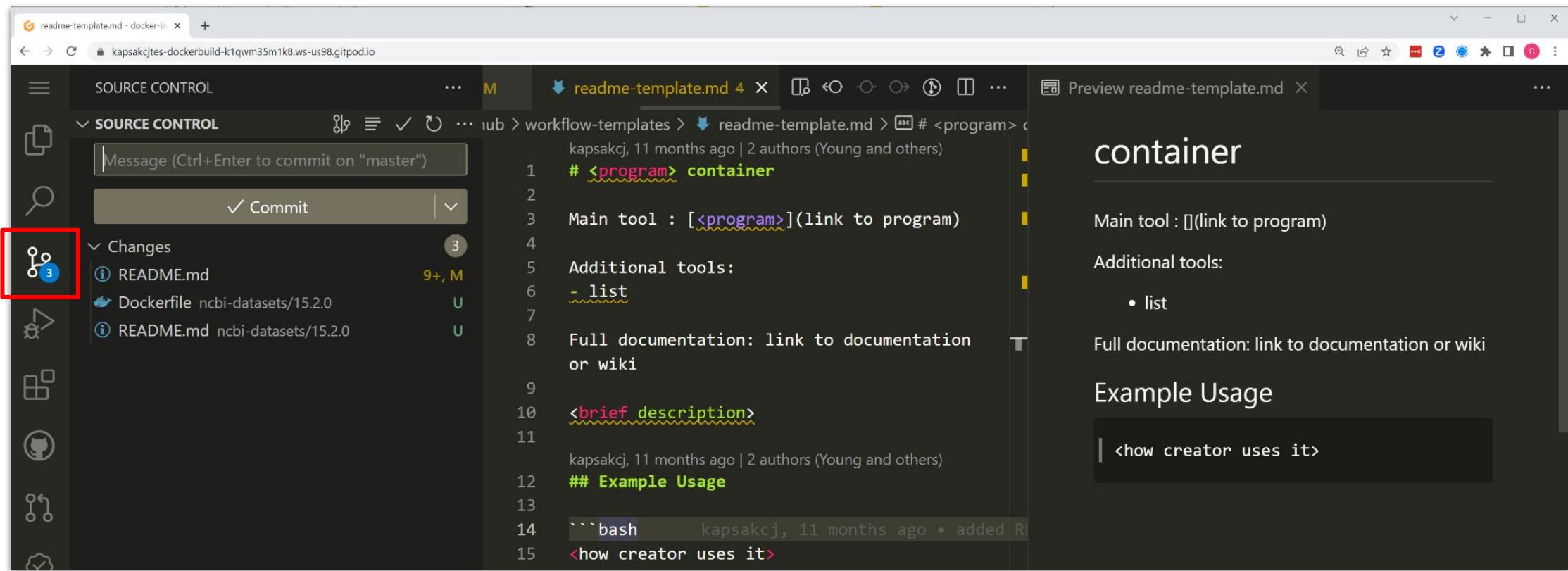
The screenshot shows a dark-themed instance of VS Code. In the left sidebar, under the 'DOCKER-BUILDS' section, there is a folder named 'github'. Inside 'github', there is a folder named 'workflow-templates'. Within 'workflow-templates', there is a file named 'readme-template.md'. A red arrow points to this file. The main editor area displays the content of 'readme-template.md'. The terminal at the bottom shows a command being run: 'gitpod /workspace/docker-builds (master) \$ docker build -t kapsakcj/ncbi-datasets:15.2.0 ncbi-datasets/15.2.0/^C'. The status bar at the bottom indicates the file is on 'master' branch, has 101 changes, and is in 'Pull Request #669'.

Final checklist before Pull Request (PR)

1. Ensure you have saved all updated files.
2. Commit all changes to your repository. You should have updated or added:
 - a. a Dockerfile:
 - i. **<tool>/<version>/Dockerfile**
 - b. a tool-specific README.md:
 - i. **<tool>/<version>/README .md**
 - c. **/README .md**
 - d. **/Program_licenses .md** (for new tools only)

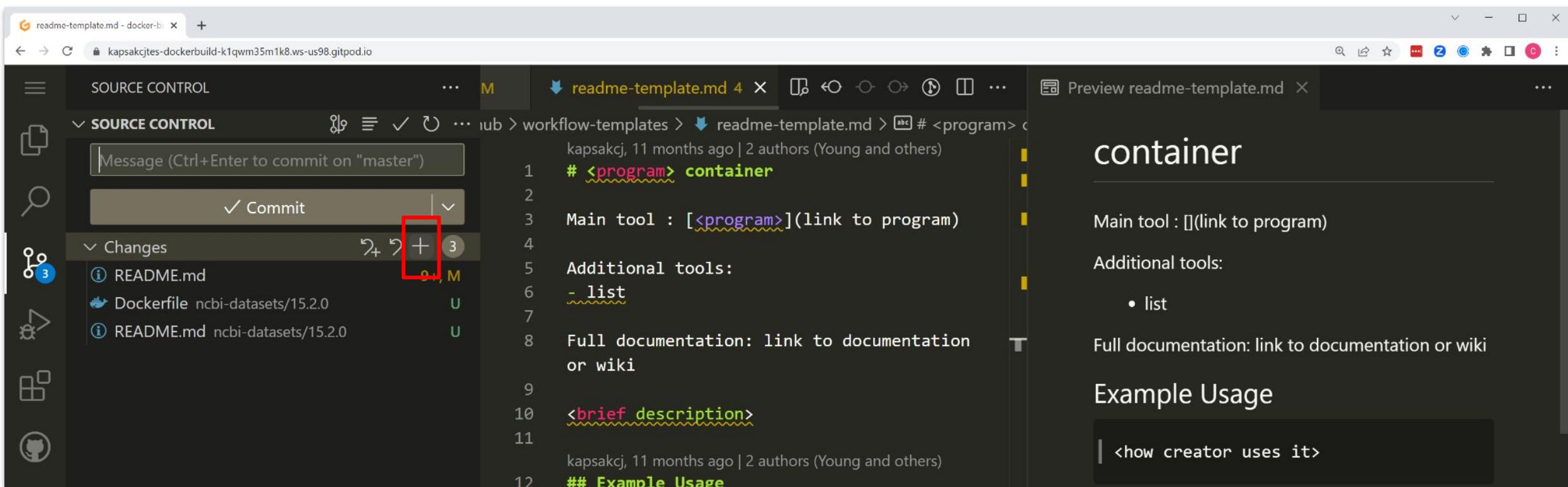
Final checklist before Pull Request (PR)

1. Commit all changes to your repository. You should have updated or added:
 - a. Click on “Source Control” button on left side



Final checklist before Pull Request (PR)

1. Commit all changes to your repository. You should have updated or added:
 - a. Stage all changes by clicking + button:



The screenshot shows a GitHub commit interface. On the left, there's a sidebar with icons for file, search, and other repository functions. The main area has tabs for 'readme-template.md' and 'Preview readme-template.md'. The preview window shows a section titled 'container' with a list of items. The commit interface includes fields for 'Message (Ctrl+Enter to commit on "master")' and a 'Commit' button. Below these are sections for 'Changes' and 'Commits'. A red box highlights the '+ Stage' button in the 'Changes' section.

```
readme-template.md - docker-build
readme-template.md 4 ×
Preview readme-template.md ×

SOURCE CONTROL
Message (Ctrl+Enter to commit on "master")
✓ Commit
Changes + Stage + 3
README.md
Dockerfile ncbi-datasets/15.2.0
README.md ncbi-datasets/15.2.0

readme-template.md 4 ×
Workflow templates > readme-template.md > # <program> container
kapsakcj, 11 months ago | 2 authors (Young and others)
1 # <program> container
2
3 Main tool : [<program>](link to program)
4
5 Additional tools:
6 - list
7
8 Full documentation: link to documentation or wiki
9
10 <brief description>
11
12 kapsakcj, 11 months ago | 2 authors (Young and others)
## Example Usage
```

container

Main tool : [](link to program)

Additional tools:

- list

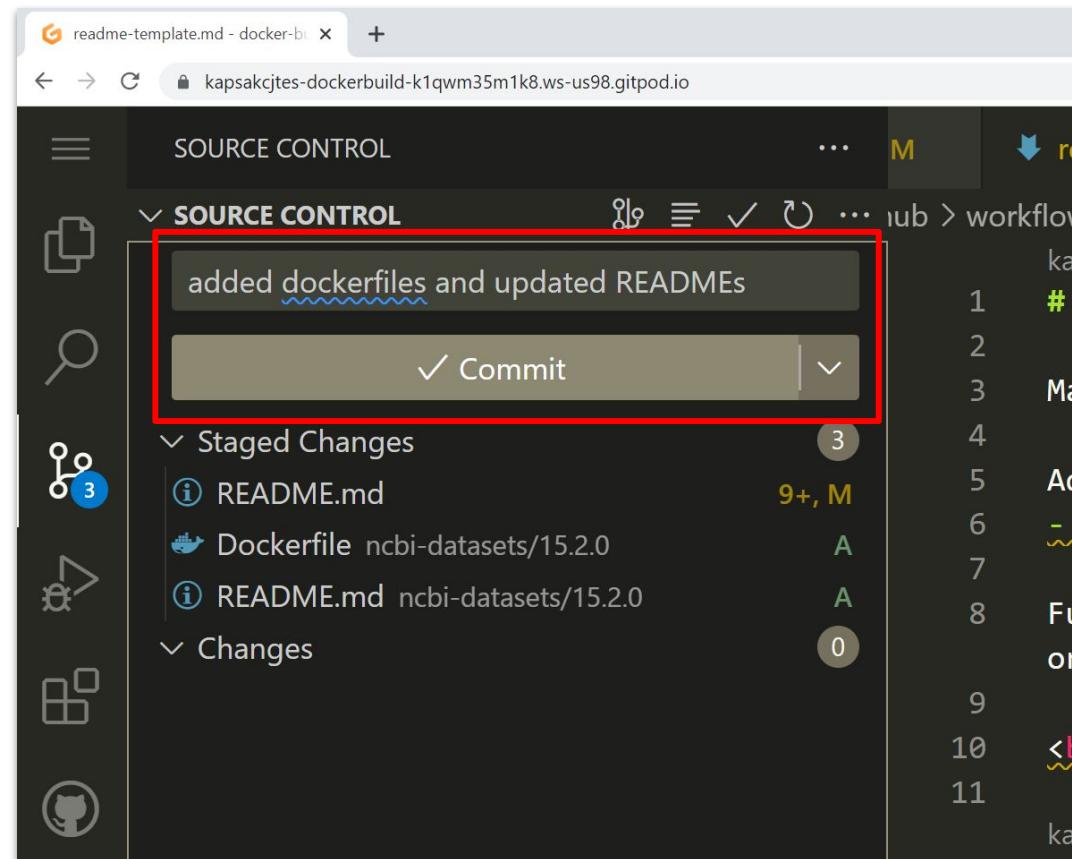
Full documentation: link to documentation or wiki

Example Usage

<how creator uses it>

Final checklist before Pull Request (PR)

1. Commit all changes to your repository. You should have updated or added:
 - a. Once changes are staged, write a short commit message
 - b. Click “Commit” button

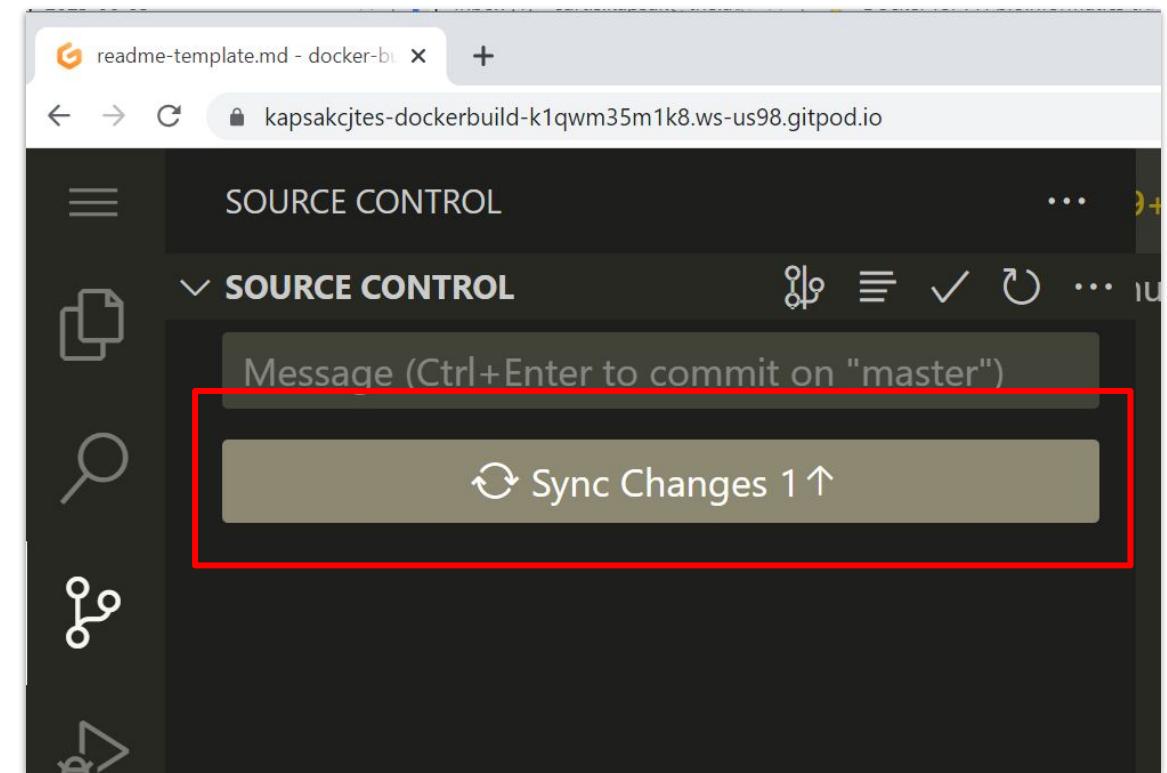


Final checklist before Pull Request (PR)

1. Commit all changes to your repository. You should have updated or added:
 - a. Once your commit has been made, click the “sync changes” button to send your commit to your forked GitHub repository

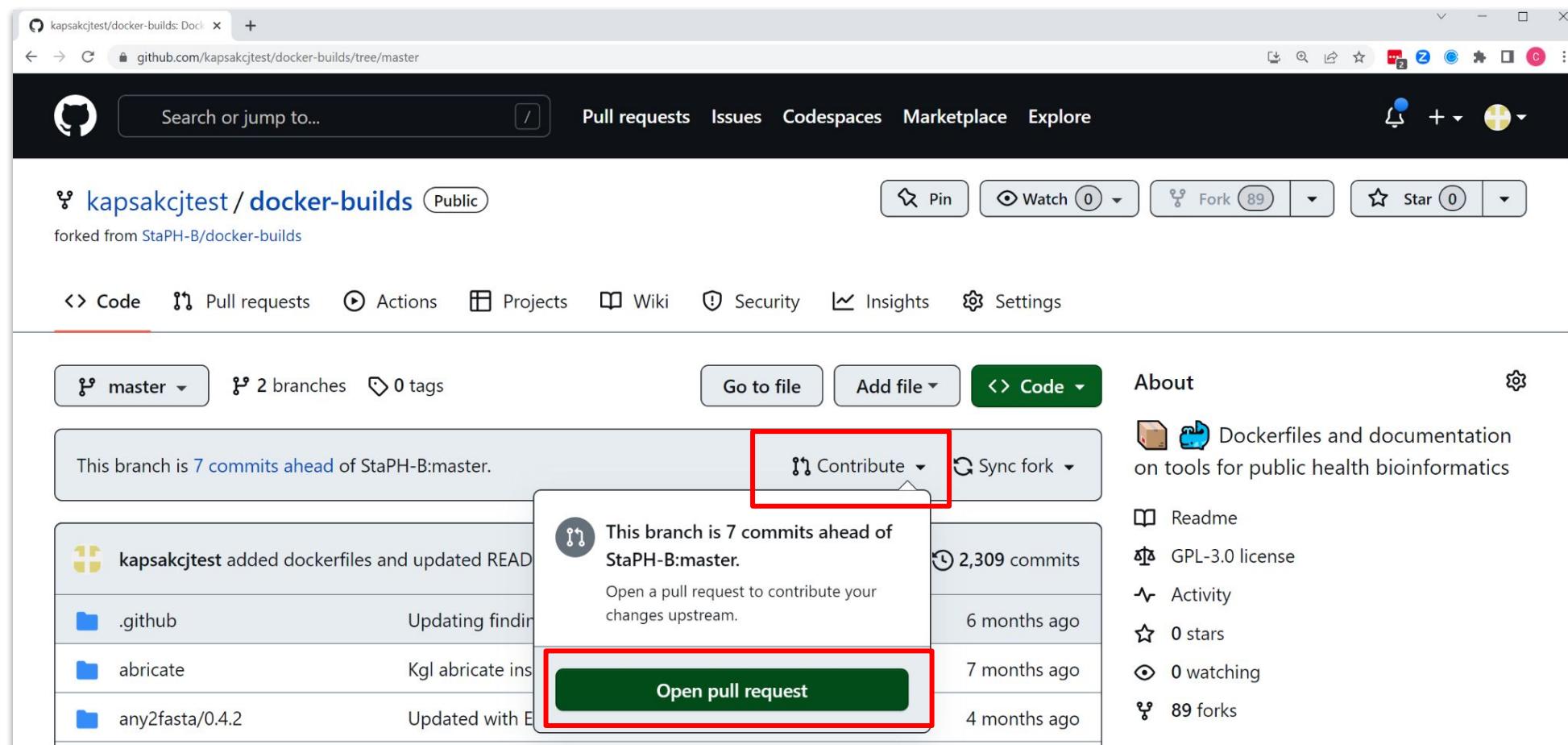
If you get a permission error,
you may need to grant GitPod
permission to push commits on your
behalf. See instructions on slide 15.

Go to <https://gitpod.io/user/integrations>
to grant permissions



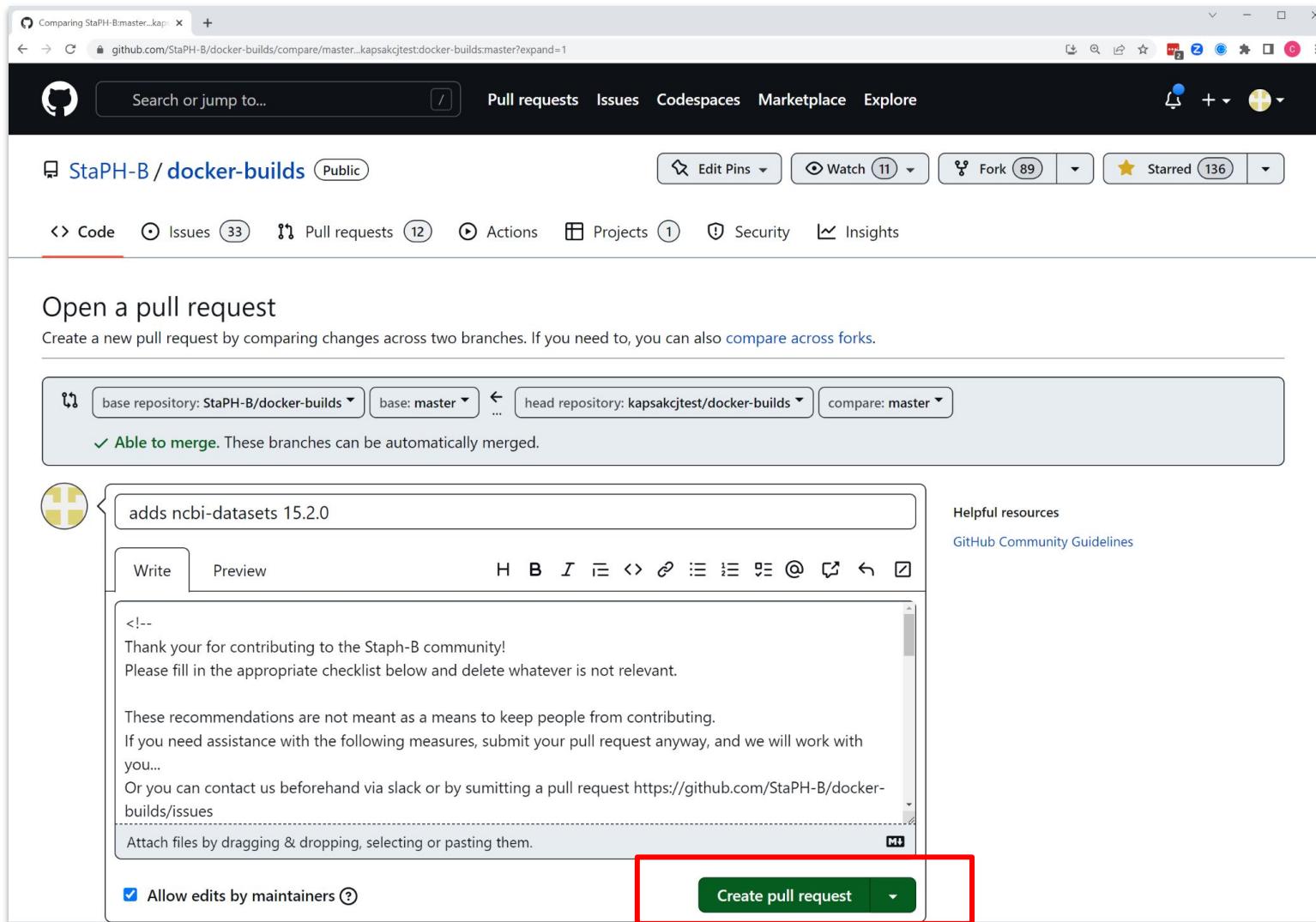
Create a Pull Request (PR)

1. Now that you've pushed your commits to your repo, it's time to create a pull request and contribute the code back to the original StaPH-B/docker-builds repo.
2. Navigate to your forked repo in your browser, and click "contribute" button to start a PR



Create a Pull Request (PR)

1. Fill out the PR form. You should have completed most of the checklist at this point.



Create a Pull Request (PR)

1. Congrats! You have started the PR process. High five! 
2. Now, either Curtis or one of the StaPH-B maintainers will review your code and offer suggestions for improvement. There may be some additional code changes required to ensure all automated tests pass. You can continue pushing commits on your forked repo in order to meet the requirements. The Pull Request will automatically update
3. Once we are satisfied with the PR, we will merge the changes into the original repo, and a docker image will be deployed to dockerhub and quay
4. Then anyone in the world will be able to access & use the docker image that you worked so hard to create. Woo!! 

Further reading & resources

- StaPH-B Github repo and docker hub account
 - <https://github.com/StaPH-B/docker-builds>
 - <https://hub.docker.com/u/staphb>
- Docker Documentation - a wealth of info here. Note that we use Docker Community Edition, as you have to pay for the Enterprise Edition
 - <https://docs.docker.com/>
- An awesome tutorial/workshop on docker for bioinformatics
 - <https://github.com/PawseySC/bio-workshop-18>
- Template for your Dockerfile
 - <https://github.com/StaPH-B/docker-builds/blob/master/dockerfile-template/Dockerfile>
- Some best practices
 - https://staphb.org/docker-builds/make_containers/
- Search for docker images and (sometimes) Dockerfiles here:
 - <http://hub.docker.com/>
 - <https://quay.io/>
- “What is Docker?”(~11 min)
 - https://www.youtube.com/watch?time_continue=1&v=aLipr7tTuA4

Acknowledgements

- MA DPH
- Members of StaPH-B & the docker-builds contributors & maintainers
 - Erin Young, UT PHL
 - Kelsey Florek, WI PHL
 - Kevin Libuit, Theiagen Genomics
 - Frank Ambrosio, Theiagen Genomics
 - many more awesome people!
 - StaPH-B docker-builds contributors:
<https://github.com/StaPH-B/docker-builds#authorsmaintainers>
- APHL
- CDC

