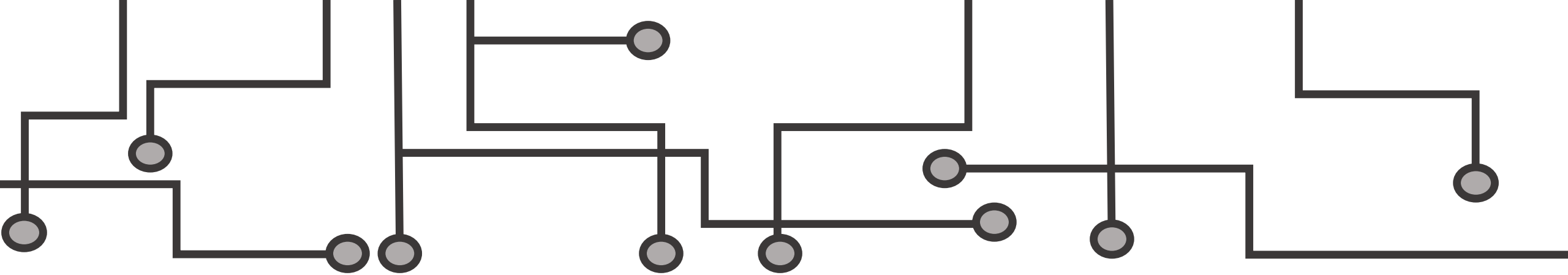


MEDICAL MICROBIOLOGY AND INFECTIOUS DISEASES CODING WORKSHOP

Presents

INTRODUCTION TO CONDA

INSTRUCTED BY

Jillian Rumore, PhD Candidate

INFORMATION FOR PARTICIPANTS

All workshops are being recorded and posted to the MMID Coding Workshop - YouTube

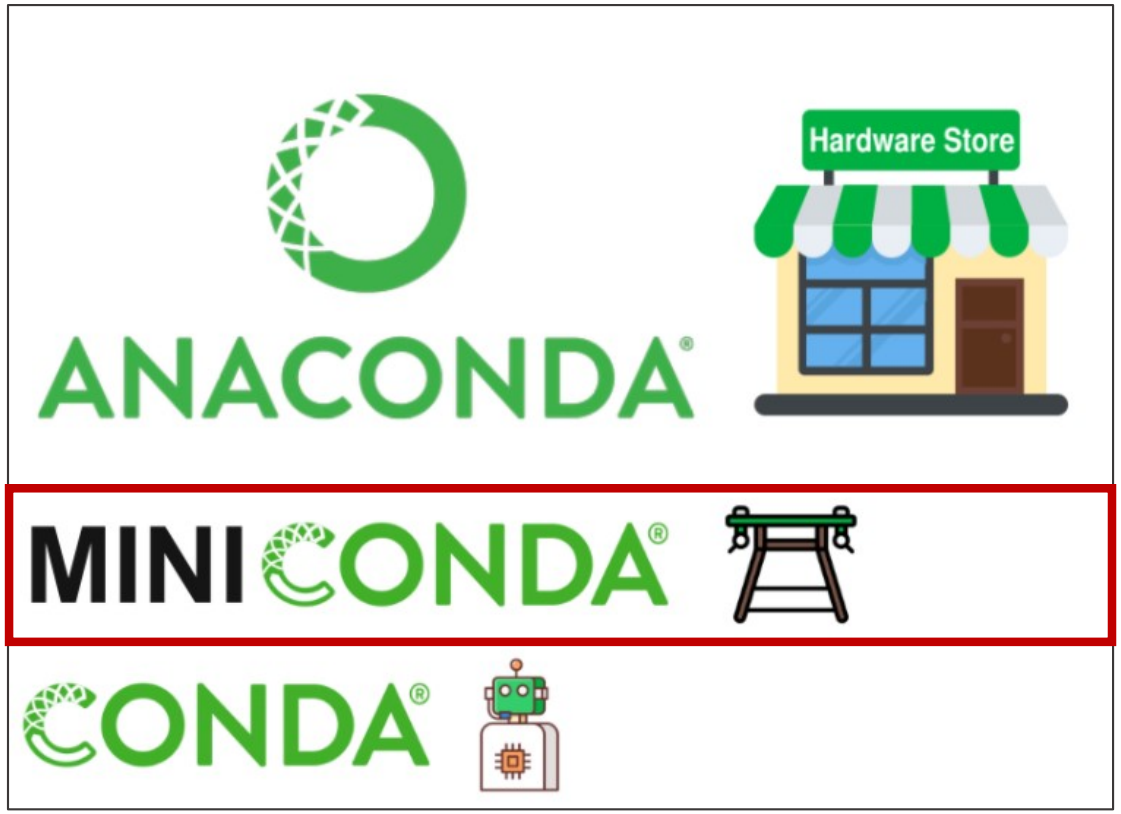
Please hold questions until the end of the workshop.

Q&A session will not be recorded.

LEARNING OBJECTIVES

1. Develop a basic understanding of Conda
2. Navigate and create Conda environments
3. Find packages and install them
4. BONUS - Run a package in Conda

WHAT IS CONDA?



- Package and environment management system
- Runs on Windows, macOS, Linux
- Language agnostic

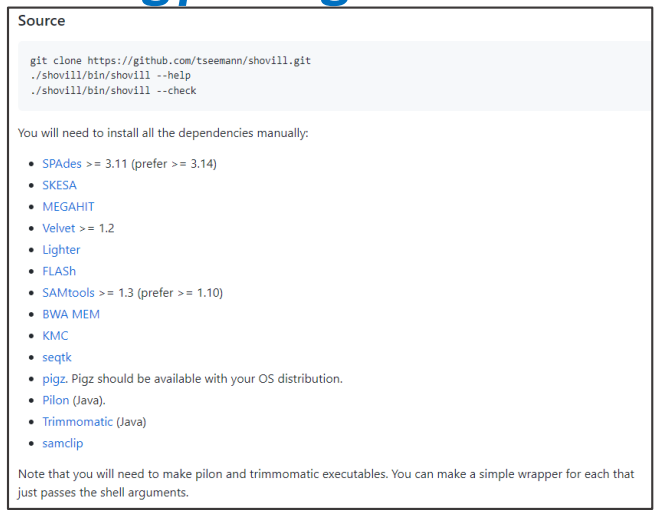
Image from https://www.mrdbourke.com/get-your-computer-ready-for-machine-learning-using-anaconda-miniconda-and-conda/

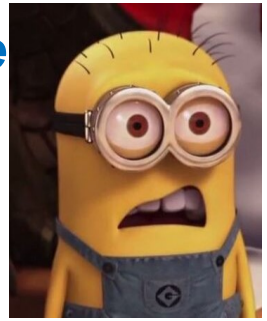
WHY USE CONDA?

- Simplifies installation and management of packages
  - Download packages (including dependencies) with a single line of code

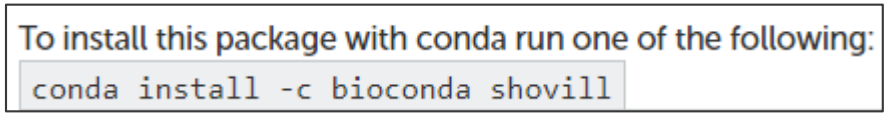
- Promotes reproducibility
  - Packages required to complete a certain task are contained in the same environment
  - Share environments with others

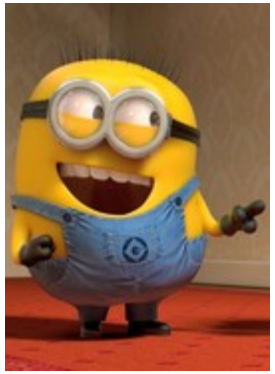
Installing packages from source





Installing packages using Conda

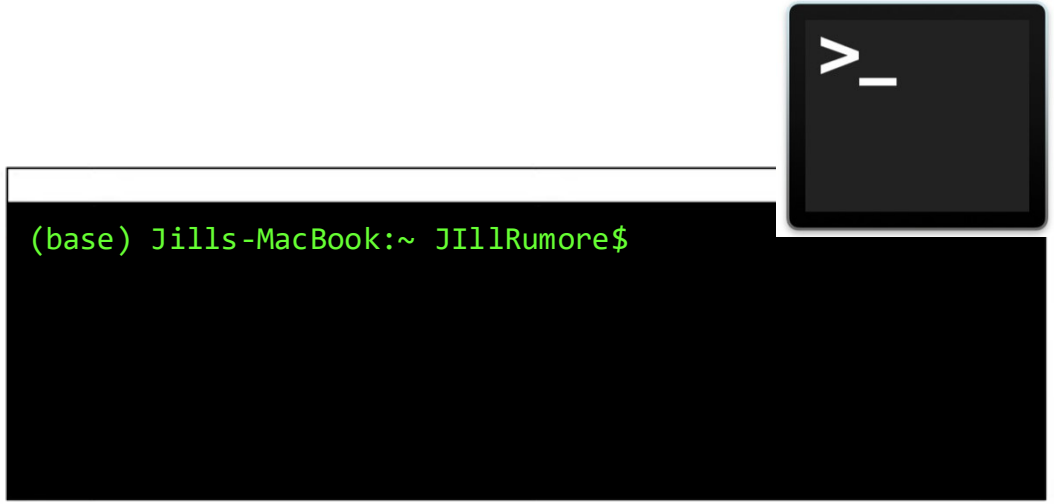




Images from https://despicableme.fandom.com/wiki/

GETTING STARTED

- **Open a new terminal**



GETTING STARTED

- **Verify Conda installed properly**

OPTION#1

COMMAND: `conda --version`

```
[jrumore@waffles ~]$ conda --version
conda 4.10.3
```

OPTION#2

COMMAND: `conda info`

```
[jrumore@waffles ~]$ conda info

     active environment : None
       user config file : /home/CSCScience.ca/jrumore/.condarc
 populated config files : /opt/miniconda3/.condarc
                          /home/CSCScience.ca/jrumore/.condarc
          conda version : 4.10.3
    conda-build version : not installed
         python version : 3.7.3.final.0
       virtual packages : __linux=3.10.0=0
                          __glibc=2.17=0
                          __unix=0=0
                          __archspec=1=x86_64
       base environment : /opt/miniconda3  (read only)
      conda av data dir : /opt/miniconda3/etc/conda
  conda av metadata url : None
          channel URLs : https://conda.anaconda.org/bioconda/linux-64
                         https://conda.anaconda.org/bioconda/noarch
                         https://conda.anaconda.org/conda-forge/linux-64
                         https://conda.anaconda.org/conda-forge/noarch
                         https://repo.anaconda.com/pkgs/main/linux-64
                         https://repo.anaconda.com/pkgs/main/noarch
                         https://repo.anaconda.com/pkgs/r/linux-64
                         https://repo.anaconda.com/pkgs/r/noarch
         package cache : /opt/miniconda3/pkgs
                         /home/CSCScience.ca/jrumore/.conda/pkgs
      envs directories : /home/CSCScience.ca/jrumore/.conda/envs
                         /Drives/P/conda_envs
                         /Drives/X/public/conda_envs
                         /opt/miniconda3/envs
             platform : linux-64
           user-agent : conda/4.10.3 requests/2.22.0 CPython/3.7.3 Linux/3.10.0-1160.49.1.el7.x86_
64 centos/7.9.2009 glibc/2.17
                UID:GID : 100359:100359
             netrc file : None
           offline mode : False
```

NAVIGATING CONDA

- **Get help with CONDA commands**

COMMAND: `conda --help`

```
[jrumore@waffles ~]$ conda --help
usage: conda [-h] [-V] command ...

conda is a tool for managing and deploying applications, environments and packages.

Options:

positional arguments:
  command
    clean        Remove unused packages and caches.
    compare      Compare packages between conda environments.
    config       Modify configuration values in .condarc. This is modeled
                 after the git config command. Writes to the user .condarc
                 file (/home/CSCScience.ca/jrumore/.condarc) by default.
    create       Create a new conda environment from a list of specified
                 packages.
    help         Displays a list of available conda commands and their help
                 strings.
    info         Display information about current conda install.
    init         Initialize conda for shell interaction. [Experimental]
    install      Installs a list of packages into a specified conda
                 environment.
    list         List linked packages in a conda environment.
    package      Low-level conda package utility. (EXPERIMENTAL)
    remove       Remove a list of packages from a specified conda environment.
    uninstall    Alias for conda remove.
    run          Run an executable in a conda environment. [Experimental]
    search       Search for packages and display associated information. The
                 input is a MatchSpec, a query language for conda packages.
                 See examples below.
    update       Updates conda packages to the latest compatible version.
    upgrade      Alias for conda update.

optional arguments:
  -h, --help     Show this help message and exit.
  -V, --version  Show the conda version number and exit.
```

- **Get help with a specific CONDA command**

COMMAND: `conda COMMAND --help`

```
[jrumore@waffles ~]$ conda create --help
usage: conda create [-h] [--clone ENV] [-n ENVIRONMENT | -p PATH] [-
c CHANNEL]
                    [--use-local] [--override-channels]
                    [--repodata-fn REPODATA_FNS] [--strict-channel-p
riority]
                    [--no-channel-priority] [--no-deps | --only-deps
]
                    [--no-pin] [--copy] [-C] [-k] [--offline] [-d] [
--json]
                    [-q] [-v] [-y] [--download-only] [--show-channel
-urls]
                    [--file FILE] [--no-default-packages] [--dev]
                    [package_spec [package_spec ...]]

Create a new conda environment from a list of specified packages. To
 use the created environment, use 'conda activate envname' look in t
hat directory first.  This command requires either the -n NAME or -p
 PREFIX option.

Options:

Target Environment Specification:
  -n ENVIRONMENT, --name ENVIRONMENT
                        Name of environment.
  -p PATH, --prefix PATH
                        Full path to environment location (i.e. pref
ix).

Examples:

    conda create -n myenv sqlite
```

ALL OPTIONS

USAGE

REQUIRED OPTIONS

EXAMPLE

CORE COMMANDS: `create, list, search, install, update, remove`

CONDA ENVIRONMENTS

- **A curated collection of packages that have been installed**
  - Environment = *collection of packages*
  - Packages = *code that someone has written for a specific purpose*

List all conda environments available to you

COMMAND: `conda env list`

NO ENVIRONMENTS

```
(base) C:\Users\rumor>conda env list
# conda environments:
#
base                  *  C:\Users\rumor\miniconda3
```

MANY ENVIRONMENTS

```
[jrumore@waffles conda_workshop]$ conda env list
# conda environments:
#
albacore                 /Drives/P/conda_envs/albacore
albacore-2.2.7           /Drives/P/conda_envs/albacore-2.2.7
albacore-2.3.0           /Drives/P/conda_envs/albacore-2.3.0
albacore-2.3.3           /Drives/P/conda_envs/albacore-2.3.3
albacore-dev             /Drives/P/conda_envs/albacore-dev
artemis-18.1.0           /Drives/P/conda_envs/artemis-18.1.0
artic-1.1.0_rc2-0        /Drives/P/conda_envs/artic-1.1.0_rc2-0
artic-1.1.1              /Drives/P/conda_envs/artic-1.1.1
artic-ncov2019           /Drives/P/conda_envs/artic-ncov2019
sra-tools_v2.10.8        /home/CSCScience.ca/jrumore/.conda/envs/sra-tools_v2.10.8
base                  *  /opt/miniconda3
```

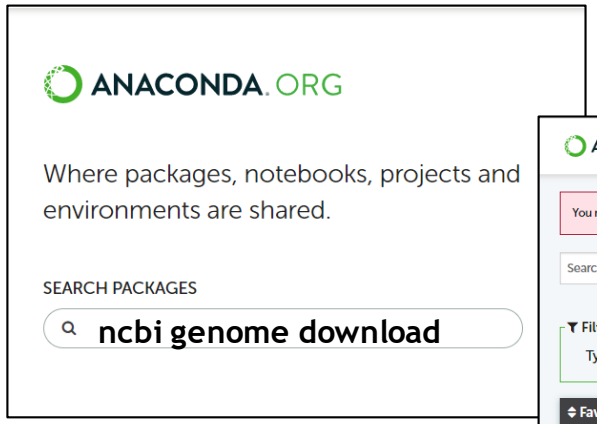
Installed by system administrators in shared drive

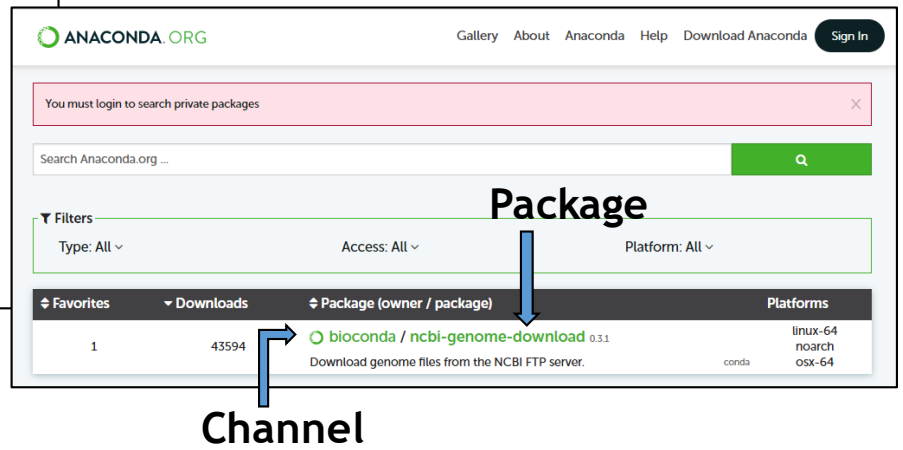
Helpful Tip

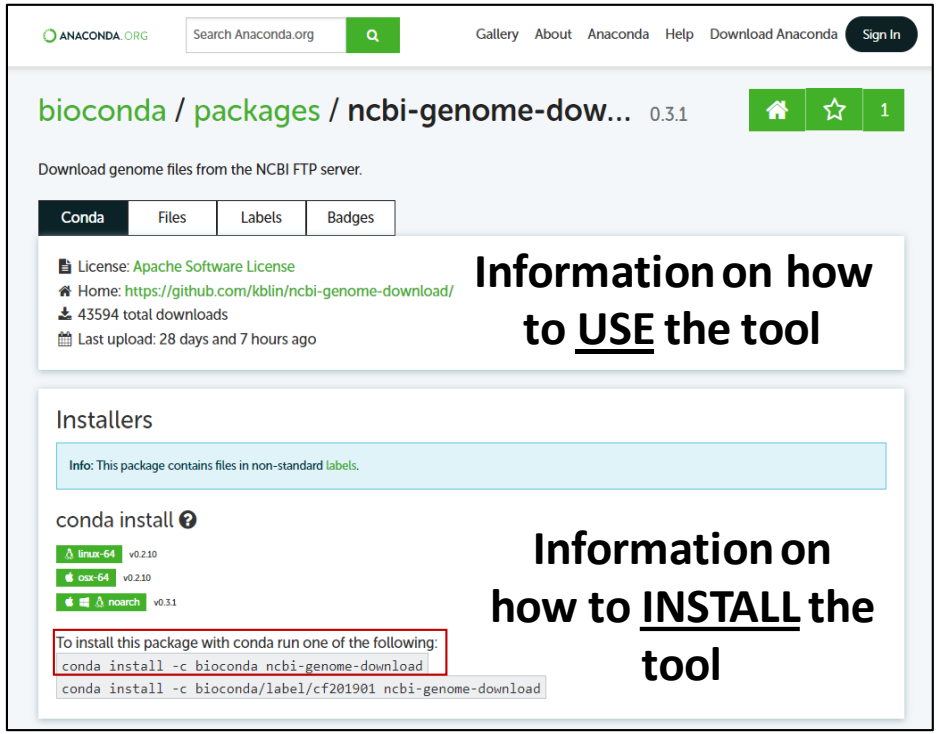
- An asterisk (*) will appear beside the active environment.

SEARCH FOR CONDA PACKAGES - ONLINE

- **Find the newest version of a tool**
  - Online repository: Anaconda Cloud (https://anaconda.org/)







Helpful Tip

- Only install packages from trusted sources.
- Bioconda and Conda-forge are reliable channels.

SEARCH FOR CONDA PACKAGES - TERMINAL

- **Check what versions of the package are available**

Search for a specific package

COMMAND: `conda search PACKAGENAME`

```
(conda_workshop) [jrumore@waffles ~]$ conda search ncbi-genome-download
Loading channels: done
# Name                       Version           Build  Channel
ncbi-genome-download           0.2.6             py_1  bioconda
ncbi-genome-download           0.2.7             py_0  bioconda
ncbi-genome-download           0.2.8           py27_1  bioconda
ncbi-genome-download           0.2.8           py35_1  bioconda
ncbi-genome-download           0.2.8           py36_1  bioconda
ncbi-genome-download           0.2.8             py_0  bioconda
ncbi-genome-download           0.2.9           py27_0  bioconda
ncbi-genome-download           0.2.9           py35_0  bioconda
ncbi-genome-download           0.2.9           py36_0  bioconda
ncbi-genome-download          0.2.10           py27_0  bioconda
ncbi-genome-download          0.2.10           py36_0  bioconda
ncbi-genome-download          0.2.10           py37_0  bioconda
ncbi-genome-download          0.2.11             py_0  bioconda
ncbi-genome-download          0.2.12             py_0  bioconda
ncbi-genome-download           0.3.0    pyh864c0ab_1  bioconda
ncbi-genome-download           0.3.0    pyh9f0ad1d_0  bioconda
ncbi-genome-download           0.3.1    pyh5e36f6f_0  bioconda
```

NAME: Package name

VERSION: Version number of the package

BUILD: Python version the package is made for

CHANNEL: Location where packages are stored; no listed channel = default channel

CREATING A NEW CONDA ENVIRONMENT

- **Creating environments and installing packages can be achieved in two ways:**

#1: Create a new CONDA environment THEN install the package [2 STEPS]

#2: Create a new CONDA environment AND install the package [1 STEP]

Helpful Tips

- Create a NEW Conda environment for each package to avoid dependency conflicts.
- Give the Conda environment a descriptive name; DO NOT use special characters or spaces.
- NEVER INSTALL packages in your base environment.
- CANNOT INSTALL packages in shared environments (Waffles users).

CREATING A NEW CONDA ENVIRONMENT #1

STEP #1A – CREATE A NEW ENVIRONMENT

OPTIONS
-y → Do not ask for confirmation
-n → Name

Create a new Conda environment

COMMAND: `conda create -y -n ENVNAME`

SLURM COMMAND: `sbatch -c NCPUS --mem MEMORY -p PARTITION --wrap="conda create -y -n ENVNAME"`

COMMAND

```
(base) C:\Users\rumor>conda create -y -n conda_workshop
Collecting package metadata (current_repodata.json): done
Solving environment: done


## Package Plan ##

  environment location: C:\Users\rumor\miniconda3\envs\conda_workshop



Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#     $ conda activate conda_workshop
#
# To deactivate an active environment, use
#
#     $ conda deactivate


(base) C:\Users\rumor>_
```

SLURM COMMAND (WAFFLES USERS)

```
[jrumore@waffles ~]$ sbatch -c 1 --mem 4G -p NMLResearch --wrap="conda create -y -n conda_workshop"
Submitted batch job 8288733
```

CREATING A CONDA ENVIRONMENT #1

STEP #1B: VERIFY THE NEW ENVIRONMENT WAS CREATED

1

List Conda environments

COMMAND: `conda env list`

```
(base) C:\Users\rumor>conda env list
# conda environments:
#
base                  *  C:\Users\rumor\miniconda3
conda_workshop           C:\Users\rumor\miniconda3\envs\conda_workshop
```

OR

Search for a specific environment

COMMAND: `conda env list | grep 'ENVNAME'`

```
[jrumore@waffles ~]$ conda env list | grep 'conda_workshop'
conda_workshop           /home/CSCScience.ca/jrumore/.conda/envs/conda_workshop
```

2

Activating a Conda environment

COMMAND: `source activate ENVNAME`

```
[jrumore@waffles ~]$ source activate conda_workshop
(conda_workshop) [jrumore@waffles ~]$
```

OR

`conda activate ENVNAME`

3

List the contents of the active environment

COMMAND: `conda list`

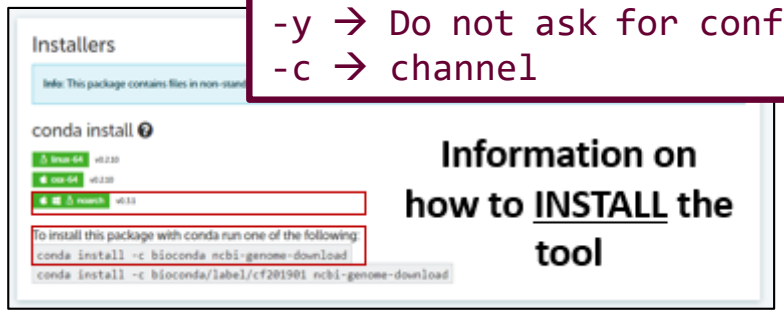
```
(conda_workshop) [jrumore@waffles ~]$ conda list
# packages in environment at /home/CSCScience.ca/jrumore/.conda/envs/conda_workshop:
#
# Name                    Version                   Build  Channel
```

→ **EMPTY ENVIRONMENT**

CREATING A CONDA ENVIRONMENT #1

STEP #2A – INSTALL PACKAGES

OPTIONS
-y → Do not ask for confirmation
-c → channel



Install the package in the active environment

COMMAND: `conda install –y –c CHANNEL PACKAGENAME`

SLURM COMMAND: `sbatch -c NCPUS --mem MEMORY -p PARTITION --wrap="conda install -y -c CHANNEL PACKAGENAME"`

COMMAND

```
(conda_workshop) C:\Users\rumor>conda install -y -c bioconda ncbi-genome-download
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: C:\Users\rumor\miniconda3\envs\conda_workshop

  added / updated specs:
    - ncbi-genome-download

The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    sqlite-3.37.0              |       h2bbff1b_0         785 KB
    ------------------------------------------------------------
                                           Total:         785 KB

The following NEW packages will be INSTALLED:

  appdirs            pkgs/main/noarch::appdirs-1.4.4-pyhd3eb1b0_0
  brotlipy           pkgs/main/win-64::brotlipy-0.7.0-py39h2bbff1b_1003
  ca-certificates    pkgs/main/win-64::ca-certificates-2021.10.26-haa95532_2
  certifi            pkgs/main/win-64::certifi-2021.10.8-py39haa95532_0
  cffi               pkgs/main/win-64::cffi-1.15.0-py39h2bbff1b_0
  charset-normalizer pkgs/main/noarch::charset-normalizer-2.0.4-pyhd3eb1b0_0
  cryptography       pkgs/main/win-64::cryptography-36.0.0-py39h21b164f_0
  idna               pkgs/main/noarch::idna-3.3-pyhd3eb1b0_0
  ncbi-genome-downl~ bioconda/noarch::ncbi-genome-download-0.3.1-pyh5e36f6f_0
  openssl            pkgs/main/win-64::openssl-1.1.1l-h2bbff1b_0
  pip                pkgs/main/win-64::pip-21.2.4-py39haa95532_0
  pycparser          pkgs/main/noarch::pycparser-2.21-pyhd3eb1b0_0
  pyopenssl          pkgs/main/noarch::pyopenssl-21.0.0-pyhd3eb1b0_1
  pysocks            pkgs/main/win-64::pysocks-1.7.1-py39haa95532_0
  python             pkgs/main/win-64::python-3.9.7-h6244533_1
  requests           pkgs/main/noarch::requests-2.26.0-pyhd3eb1b0_0
  setuptools         pkgs/main/win-64::setuptools-58.0.4-py39haa95532_0
  six                pkgs/main/noarch::six-1.16.0-pyhd3eb1b0_0
  sqlite             pkgs/main/win-64::sqlite-3.37.0-h2bbff1b_0
  tqdm               pkgs/main/noarch::tqdm-4.62.3-pyhd3eb1b0_1
  tzdata             pkgs/main/noarch::tzdata-2021e-hda174b7_0
  urllib3            pkgs/main/noarch::urllib3-1.26.7-pyhd3eb1b0_0
  vc                 pkgs/main/win-64::vc-14.2-h21ff451_1
  vs2015_runtime     pkgs/main/win-64::vs2015_runtime-14.27.29016-h5e58377_2
  wheel              pkgs/main/noarch::wheel-0.37.0-pyhd3eb1b0_1
  win_inet_pton      pkgs/main/win-64::win_inet_pton-1.1.0-py39haa95532_0
  wincertstore       pkgs/main/win-64::wincertstore-0.2-py39haa95532_2

Downloading and Extracting Packages
sqlite-3.37.0        | 785 KB    | ############################################################################ | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done

(conda_workshop) C:\Users\rumor>
```

SLURM COMMAND (WAFFLES USERS)

```
(conda_workshop) [jrumore@waffles ~]$ sbatch -c 1 --mem 4G -p NMLResearch --wrap="conda install -y -c bioconda ncbi-genome-download"
Submitted batch job 8315089
```

Helpful Tip

- Make sure the desired environment is activated BEFORE installing packages.
- Active environment will appear in rounded brackets beside your user name.

CREATING A CONDA ENVIRONMENT #1

STEP #2B – VERIFY THE PACKAGE INSTALLED

List the contents of the active environment

COMMAND: `conda list`

```
(conda_workshop) C:\Users\rumor>conda list
# packages in environment at C:\Users\rumor\miniconda3\envs\conda_workshop:
#
# Name                    Version                   Build  Channel
appdirs                   1.4.4              pyhd3eb1b0_0
brotlipy                  0.7.0           py39h2bbff1b_1003
ca-certificates           2021.10.26           haa95532_2
certifi                   2021.10.8       py39haa95532_0
cffi                      1.15.0          py39h2bbff1b_0
charset-normalizer        2.0.4              pyhd3eb1b0_0
cryptography              36.0.0          py39h21b164f_0
idna                      3.3                pyhd3eb1b0_0
ncbi-genome-download      0.3.1              pyh5e36f6f_0    bioconda
openssl                   1.1.1l               h2bbff1b_0
pip                       21.2.4          py39haa95532_0
pycparser                 2.21               pyhd3eb1b0_0
pyopenssl                 21.0.0             pyhd3eb1b0_1
pysocks                   1.7.1           py39haa95532_0
python                    3.9.7                h6244533_1
requests                  2.26.0             pyhd3eb1b0_0
setuptools                58.0.4          py39haa95532_0
six                       1.16.0             pyhd3eb1b0_0
sqlite                    3.37.0               h2bbff1b_0
tqdm                      4.62.3             pyhd3eb1b0_1
tzdata                    2021e                hda174b7_0
urllib3                   1.26.7             pyhd3eb1b0_0
vc                        14.2                 h21ff451_1
vs2015_runtime            14.27.29016          h5e58377_2
wheel                     0.37.0             pyhd3eb1b0_1
win_inet_pton             1.1.0           py39haa95532_0
wincertstore              0.2             py39haa95532_2

(conda_workshop) C:\Users\rumor>
```

CREATING A CONDA ENVIRONMENT #2

STEP #1A – CREATE ENVIRONMENT AND INSTALL PACKAGES

OPTIONS
-y → Do not ask for confirmation
-n → Name
-c → channel

Create a new environment and install packages

COMMAND: `conda create -y -n ENVNAME -c CHANNEL PACKAGENAME`

SLURM COMMAND: `sbatch -c NCPUS --mem MEMORY -p PARTITION --wrap="conda create -y -n ENVNAME -c CHANNEL PACKAGENAME"`

COMMAND

```
(base) jrumore@DESKTOP-DI4ORKU:~$ conda create -y -n conda_workshop -c bioconda ncbi-genome-download
```

SLURM COMMAND (WAFFLES USERS)

```
[jrumore@waffles ~]$ sbatch -c 1 --mem 4G -p NMLResearch --wrap="conda create -y -n conda_workshop -c bioconda ncbi-genome-
download"
Submitted batch job 8323047
```

CREATING A CONDA ENVIRONMENT #2

STEP #1B – VERIFY THE PACKAGE INSTALL

Activate the environment

COMMAND: `source activate ENVNAME` **OR** `conda activate ENVNAME`

```
(base) jrumore@DESKTOP-DI4ORKU:~$ conda activate conda_workshop
(conda_workshop) jrumore@DESKTOP-DI4ORKU:~$
```

List the contents of the environment

COMMAND: `conda list`

```
(conda_workshop) jrumore@DESKTOP-DI4ORKU:~$ conda list
# packages in environment at /home/jrumore/miniconda3/envs/conda_workshop:
#
# Name                    Version                   Build  Channel
_libgcc_mutex             0.1                         main
_openmp_mutex             4.5                         1_gnu
appdirs                   1.4.4              pyhd3eb1b0_0
brotlipy                  0.7.0           py39h27cfd23_1003
ca-certificates           2021.10.26           h06a4308_2
certifi                   2021.10.8       py39h06a4308_2
cffi                      1.14.6          py39h400218f_0
charset-normalizer        2.0.4              pyhd3eb1b0_0
cryptography              36.0.0          py39h9ce1e76_0
idna                      3.3                pyhd3eb1b0_0
ld_impl_linux-64          2.35.1               h7274673_9
libffi                    3.3                  he6710b0_2
libgcc-ng                 9.3.0               h5101ec6_17
libgomp                   9.3.0               h5101ec6_17
libstdcxx-ng              9.3.0               hd4cf53a_17
ncbi-genome-download      0.3.1              pyh5e36f6f_0    bioconda
ncurses                   6.3                  h7f8727e_2
```

CHANGING CONDA ENVIRONMENTS

- **To work in a DIFFERENT Conda environment, DEACTIVATE the current Conda environment, then ACTIVATE the new one.**

Deactivate a Conda environment

COMMAND: `conda deactivate`

```
(conda_workshop) [jrumore@waffles ~]$ conda deactivate
[jrumore@waffles ~]$
```

Activate the new Conda environment

COMMAND: `source activate ENVNAME` OR `conda activate ENVNAME`

```
[jrumore@waffles ~]$ source activate shovill
(shovill) [jrumore@waffles ~]$
```

Helpful Tip

- Active environment will appear in rounded brackets beside your user name.

INSTALLING A SPECIFIC PACKAGE VERSION

OPTIONS
-y: Do not ask for confirmation
-c: channel

Install a specific version in the active environment

COMMAND: `conda install -y -c CHANNEL PACKAGENAME=VERSION`

SLURM COMMAND: `sbatch -c NCPUS --mem MEMORY -p PARTITION --wrap="conda install -y -c CHANNEL PACKAGENAME=VERSION"`

COMMAND

```
(conda_workshop) [jrumore@waffles ~]$ conda install -y -c bioconda ncbi-genome-download=0.2.12
```

SLURM COMMAND

```
(conda_workshop) [jrumore@waffles ~]$ sbatch -c 1 --mem 4G -p NMLResearch --wrap="conda install -y -c bioconda ncbi-genome-download=0.2.12"
Submitted batch job 8333275
```

List contents of the active environment

COMMAND: `conda list`

```
(conda_workshop) [jrumore@waffles ~]$ conda list
# packages in environment at /home/CSCScience.ca/jrumore/.conda/envs/conda_workshop:
#
# Name                    Version                   Build  Channel
_libgcc_mutex             0.1                 conda_forge    conda-forge
_openmp_mutex             4.5                       1_gnu    conda-forge
appdirs                   1.4.4              pyh9f0ad1d_0    conda-forge
brotlipy                  0.7.0           py36h8f6f2f9_1001    conda-forge
ca-certificates           2021.10.8           ha878542_0    conda-forge
certifi                   2021.5.30        py36h5fab9bb_0    conda-forge
cffi                      1.14.6           py36hd8eec40_1    conda-forge
charset-normalizer        2.0.9              pyhd8ed1ab_0    conda-forge
cryptography              35.0.0           py36hb60f036_0    conda-forge
enum34                    1.1.10           py36h9f0ad1d_2    conda-forge
idna                      3.1                pyhd3deb0d_0    conda-forge
ld_impl_linux-64          2.36.1               hea4e1c9_2    conda-forge
libffi                    3.4.2                h7f98852_5    conda-forge
libgcc-ng                 11.2.0              h1d223b6_11    conda-forge
libgomp                   11.2.0              h1d223b6_11    conda-forge
libnsl                    2.0.0                h7f98852_0    conda-forge
libstdcxx-ng              11.2.0              he4da1e4_11    conda-forge
libzlib                   1.2.11            h36c2ea0_1013    conda-forge
ncbi-genome-download      0.2.12                    py_0    bioconda
```

Helpful Tip

- Make sure you have the correct environment activated before installing packages.

UPDATING CONDA PACKAGES

OPTIONS
-y: Do not ask for confirmation

Update a specific package in the active environment

COMMAND: `conda update -y PACKAGENAME`

SLURM COMMAND: `sbatch -c NCPUS --mem MEMORY -p PARTITION --wrap="conda update -y PACKAGENAME"`

COMMAND

```
(conda_workshop) [jrumore@waffles ~]$ conda update -y ncbi-genome-download
```

SLURM COMMAND

```
(conda_workshop) [jrumore@waffles ~]$ sbatch -c 1 --mem 4G -p NMLResearch --wrap="conda update -y ncbi-genome-download"
Submitted batch job 8333275
```

List the contents of the CONDA environment

COMMAND: `conda list`

```
(conda_workshop) [jrumore@waffles ~]$ conda list
# packages in environment at /home/CSCScience.ca/jrumore/.conda/envs/conda_workshop:
#
# Name                    Version                   Build  Channel
_libgcc_mutex             0.1                 conda_forge    conda-forge
_openmp_mutex             4.5                       1_gnu    conda-forge
appdirs                   1.4.4              pyh9f0ad1d_0    conda-forge
ca-certificates           2021.10.8           ha878542_0    conda-forge
certifi                   2016.9.26               py36_0    conda-forge
ld_impl_linux-64          2.36.1             hea4e1c9_2    conda-forge
libffi                    3.4.2              h7f98852_5    conda-forge
libgcc-ng                 11.2.0             h1d223b6_11    conda-forge
libgomp                   11.2.0             h1d223b6_11    conda-forge
libnsl                    2.0.0              h7f98852_0    conda-forge
libstdcxx-ng              11.2.0             he4da1e4_11    conda-forge
libzlib                   1.2.11            h36c2ea0_1013    conda-forge
ncbi-genome-download      0.3.1              pyh5e36f6f_0    bioconda
```

Helpful Tip

- Make sure you have the correct environment activated BEFORE updating packages.

SHARING CONDA ENVIRONMENTS - EXPORT

Reproducibility is a KEY objective of Conda

- Important that the same environment can be re-constructed across platforms and operating systems.

Share a CONDA environment

COMMAND: `conda env export --from-history > environment.yml`

SLURM COMMAND: `sbatch -c NCPUS --mem MEMORY -p PARTITION --wrap="conda env export --from-history > environment.yml"`

COMMAND

```
(conda_workshop) [jrumore@waffles conda_workshop]$ conda env export --from-history > environment.yml
```

SLURM COMMAND

```
(conda_workshop) [jrumore@waffles conda_workshop]$ sbatch -c 1 --mem 1G -p NMLResearch --wrap="conda env export --from-history > environment.yml"
Submitted batch job 8333275
```

Helpful Tip

- Make sure the desired conda environment is activated.
- `--from-history`: ensures the environment export is operating system agnostic.
- *environment.yml* file will be exported to the current working directory.

SHARING CONDA ENVIRONMENTS - INSTALL

- **Share *environment.yml* file with those who want to reproduce the environment**

environment.yml - Notepad

File Edit Format View Help

```
name: conda_workshop
channels:
  - bioconda
  - conda-forge
  - defaults
dependencies:
  - ncbi-genome-download
prefix: /home/CSCScience.ca/jrumore/.conda/envs/conda_workshop
```

Install the Conda environment

COMMAND: `conda env create -f environment.yml`

OPTIONS

`-f` → environment definition file

SLURM COMMAND: `sbatch -c NCPUS --mem MEMORY -p PARTITION --wrap="conda env create -f environment.yml"`

COMMAND

```
[jrumore@waffles conda_workshop]$ conda env create -f environment.yml
```

SLURM COMMAND

```
[jrumore@waffles Project_Jill_Rumore]$ sbatch -c 1 --mem 4G -p NMLResearch --wrap="conda env create -f ./conda_workshop/environment.yml"
Submitted batch job 8333275
```

Helpful Tip

- Make sure no conda environments are active.

REMOVING PACKAGES

OPTIONS
`-y` → do not ask for confirmation

Remove a Conda package from the active environment

COMMAND: `conda remove –y PACKAGENAME`

SLURM COMMAND: `sbatch -c NCPUS --mem MEMORY -p PARTITION --wrap="conda remove –y PACKAGENAME"`

COMMAND

```
(conda_workshop) [jrumore@waffles ~]$ conda remove -y ncbi-genome-download
```

SLURM COMMAND

```
(conda_workshop) [jrumore@waffles ~]$ sbatch -c 1 --mem 4G -p NMLResearch --wrap="conda remove -y ncbi-genome-download"
Submitted batch job 8335177
```

List the contents of the active environment

COMMAND: `conda list`

```
(conda_workshop) [jrumore@waffles ~]$ conda list
# packages in environment at /home/CSCScience.ca/jrumore/.conda/envs/conda_workshop:
#
# Name                    Version                   Build  Channel
_libgcc_mutex             0.1                 conda_forge    conda-forge
_openmp_mutex             4.5                       1_gnu    conda-forge
ca-certificates           2021.10.8           ha878542_0    conda-forge
certifi                   2021.5.30       py36h5fab9bb_0    conda-forge
ld_impl_linux-64          2.36.1              hea4e1c9_2    conda-forge
libffi                    3.4.2               h7f98852_5    conda-forge
libgcc-ng                 11.2.0              h1d223b6_11    conda-forge
libgomp                   11.2.0              h1d223b6_11    conda-forge
```

Helpful Tip

- Make sure you have the correct environment activated BEFORE removing packages.

REMOVING ENVIRONMENTS

Remove an entire Conda environment – DRY RUN

COMMAND: `conda env remove -d -n ENVNAME`

```
[jrumore@waffles ~]$ conda env remove -d -n conda_workshop

Remove all packages in environment /home/CSCScience.ca/jrumore/.conda/envs/conda_workshop:
```

OPTIONS
-y → do not ask for confirmation
-n → Name
-d → dry-run

Remove an entire Conda environment

COMMAND: `conda env remove -y -n ENVNAME`

SLURM COMMAND: `sbatch -c NCPUS --mem MEMORY -p PARTITION --wrap="conda env remove -y -n ENVNAME"`

COMMAND

```
[jrumore@waffles ~]$ conda env remove -y -n conda_workshop
```

SLURM COMMAND

```
[jrumore@waffles ~]$ sbatch -c 1 --mem 4G -p NMLResearch --wrap="conda env remove -y -n conda_workshop"
Submitted batch job 8339386
```

Helpful Tip

- Make sure the conda environment you want to remove is DEACTIVATED.
- Before removing an entire CONDA environment, it is STRONGLY recommended to DO A DRY RUN to make sure you are removing the correct environment.

REMOVING ENVIRONMENTS

- **Verify the environment was removed**

List Conda environments

COMMAND: `conda env list`

OR

Search for a specific environment

COMMAND: `conda env list | grep ‘ENVNAME’`

```
[jrumore@waffles ~]$ conda env list
canu_v2.1.1              /home/CSCScience.ca/jrumore/.conda/envs/canu_v2.1.1
checkM                   /home/CSCScience.ca/jrumore/.conda/envs/checkM
chewbbacca               /home/CSCScience.ca/jrumore/.conda/envs/chewbbacca
clustalW                 /home/CSCScience.ca/jrumore/.conda/envs/clustalW
conterminator            /home/CSCScience.ca/jrumore/.conda/envs/conterminator
ectyper0.8.0             /home/CSCScience.ca/jrumore/.conda/envs/ectyper0.8.0
ectyper0.8.1             /home/CSCScience.ca/jrumore/.conda/envs/ectyper0.8.1
```

```
[jrumore@waffles ~]$ conda env list | grep 'conda_workshop'
[jrumore@waffles ~]$
```

RUN A PACKAGE IN CONDA

STEPS:

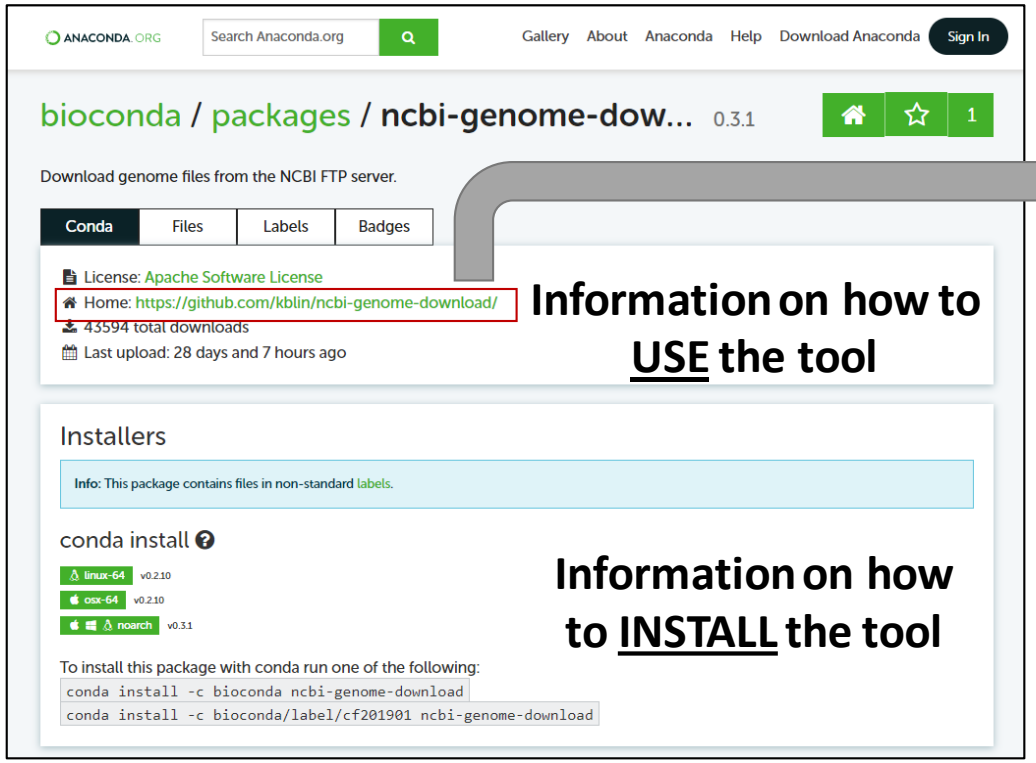
1. Determine what it is you want to accomplish.
2. Review the package documentation.
3. Activate the Conda environment containing the package of interest.
4. Set-up a project directory.
5. Run the package.
6. Verify the package ran successfully.

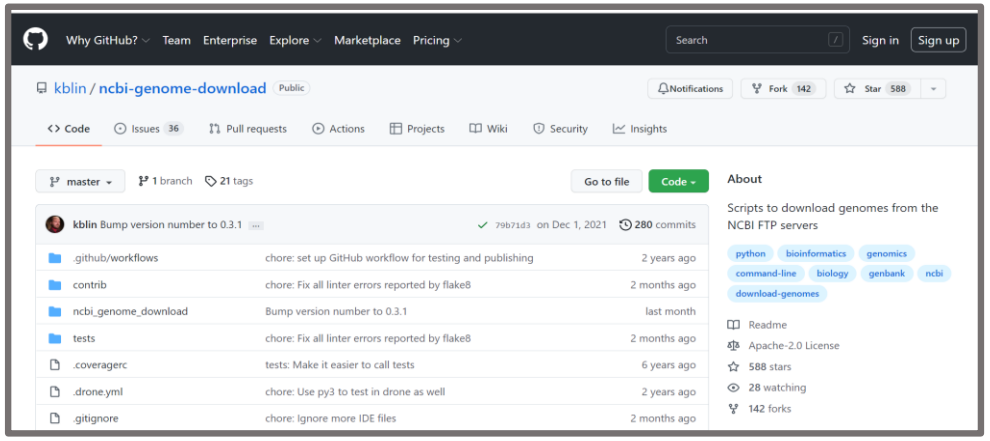
RUN A PACKAGE IN CONDA - EXAMPLE

STEPS:

1. Determine what you would like to accomplish.
   - *Download the complete genome in fasta format for the Helicobacter pylori reference strain 26695 (taxID 85962) from RefSeq.*

2. Review the tool documentation.





To download all viral RefSeq genomes in FASTA format, run:

```
ncbi-genome-download --formats fasta viral
```

To download only completed bacterial RefSeq genomes in GenBank format, run:

```
ncbi-genome-download --assembly-levels complete bacteria
```

To download a specific bacterial RefSeq genomes based on its NCBI taxonomy ID, run:

```
ncbi-genome-download --taxids 511145 bacteria
```

Note: The above command will download the RefSeq genome belonging to *Escherichia coli str. K-12 substr. MG1655*.

To get an overview of all options, run

```
ncbi-genome-download --help
```

RUN A PACKAGE IN CONDA - EXAMPLE

STEPS:

3A. Activate the conda environment.

```
[jrumore@waffles ~]$ source activate conda_workshop
(conda_workshop) [jrumore@waffles ~]$
```

3B. Review the package options in the

```
(conda_workshop) [jrumore@waffles ~]$ ncbi-genome-download --help
```

TASK

- *Download the complete genome in fasta format for the Helicobacter pylori reference strain 26695 (taxID 85962) from RefSeq.*

Options to include in command to run package

- `--formats`
- `--assembly-levels`
- `--taxids`
- `groups`

```
usage: ncbi-genome-download [-h] [-s {refseq,genbank}] [-F FILE_FORMATS]
                            [-l ASSEMBLY_LEVELS] [-g GENERA] [--genus GENERA]
                            [--fuzzy-genus] [-S STRAINS] [-T SPECIES_TAXIDS]
                            [-t TAXIDS] [-A ASSEMBLY_ACCESSIONS]
                            [-R REFSEQ_CATEGORIES]
                            [--refseq-category REFSEQ_CATEGORIES] [-o OUTPUT]
                            [--flat-output] [-H] [-P] [-u URI] [-p N] [-r N]
                            [-m METADATA_TABLE] [-n] [-N] [-v] [-d] [-V]
                            [-M TYPE_MATERIALS]
                            groups

positional arguments:
  groups                The NCBI taxonomic groups to download (default: all).
                        A comma-separated list of taxonomic groups is also
                        possible. For example: "bacteria,viral"Choose from:
                        ['all', 'archaea', 'bacteria', 'fungi',
                        'invertebrate', 'metagenomes', 'plant', 'protozoa',
                        'vertebrate_mammalian', 'vertebrate_other', 'viral']

optional arguments:
  -h, --help            show this help message and exit
  -s {refseq,genbank}, --section {refseq,genbank}
                        NCBI section to download (default: refseq)
  -F FILE_FORMATS, --formats FILE_FORMATS
                        Which formats to download (default: genbank).A comma-
                        separated list of formats is also possible. For
                        example: "fasta,assembly-report". Choose from:
                        ['genbank', 'fasta', 'rm', 'features', 'gff',
                        'protein-fasta', 'genpept', 'wgs', 'cds-fasta', 'rna-
                        fna', 'rna-fasta', 'assembly-report', 'assembly-
                        stats', 'all']
```

```
  -l ASSEMBLY_LEVELS, --assembly-levels ASSEMBLY_LEVELS
                        Assembly levels of genomes to download (default: all).
                        A comma-separated list of assembly levels is also
                        possible. For example: "complete,chromosome". Choose
                        from: ['all', 'complete', 'chromosome', 'scaffold',
                        'contig']
  -g GENERA, --genera GENERA
                        Only download sequences of the provided genera. A
                        comma-seperated list of genera is also possible. For
                        example: "Streptomyces coelicolor,Escherichia coli".
                        (default: [])
  --genus GENERA        Deprecated alias of --genera
  --fuzzy-genus         Use a fuzzy search on the organism name instead of an
                        exact match.
  -S STRAINS, --strains STRAINS
                        Only download sequences of the given strain(s). A
                        comma-separated list of strain names is possible, as
                        well as a path to a filename containing one name per
                        line.
  -T SPECIES_TAXIDS, --species-taxids SPECIES_TAXIDS
                        Only download sequences of the provided species NCBI
                        taxonomy IDs. A comma-separated list of species taxids
                        is also possible. For example: "52342,12325".
                        (default: [])
  -t TAXIDS, --taxids TAXIDS
                        Only download sequences of the provided NCBI taxonomy
                        IDs. A comma-separated list of taxids is also
                        possible. For example: "9606,9685". (default: [])
  -A ASSEMBLY_ACCESSIONS, --assembly-accessions ASSEMBLY_ACCESSIONS
                        Only download sequences matching the provided NCBI
                        assembly accession(s). A comma-separated list of
                        accessions is possible, as well as a path to a
                        filename containing one accession per line.
```

```
  -R REFSEQ_CATEGORIES, --refseq-categories REFSEQ_CATEGORIES
                        Only download sequences of the provided refseq
                        categories (default: all)
  --refseq-category REFSEQ_CATEGORIES
                        Deprecated alias for --refseq-categories
  -o OUTPUT, --output-folder OUTPUT
                        Create output hierarchy in specified folder (default:
                        /Drives/K/jrumore)
  --flat-output         Dump all files right into the output folder without
                        creating any subfolders.
  -H, --human-readable  Create links in human-readable hierarchy (might fail
                        on Windows)
  -P, --progress-bar    Create a progress bar for indicating the download
                        progress
  -u URI, --uri URI     NCBI base URI to use (default:
                        https://ftp.ncbi.nih.gov/genomes)
  -p N, --parallel N    Run N downloads in parallel (default: 1)
  -r N, --retries N     Retry download N times when connection to NCBI fails
                        (default: 0)
  -m METADATA_TABLE, --metadata-table METADATA_TABLE
                        Save tab-delimited file with genome metadata
  -n, --dry-run         Only check which files to download, don't download
                        genome files.
  -N, --no-cache        Don't cache the assembly summary file in
                        /home/CSCScience.ca/jrumore/.cache/ncbi-genome-
                        download.
  -v, --verbose         increase output verbosity
  -d, --debug           print debugging information
  -V, --version         print version information
```

RUN A PACKAGE IN CONDA - EXAMPLE

STEPS:

4. Setup a project directory to save the results.

```
(conda_workshop) [jrumore@waffles ~]$ cd /Drives/W/Projects/Project_Jill_Rumore/conda_workshop/
(conda_workshop) [jrumore@waffles conda_workshop]$ mkdir ncbi_download
(conda_workshop) [jrumore@waffles conda_workshop]$ cd ncbi_download/
(conda_workshop) [jrumore@waffles ncbi_download]$
```

5. Run the package.

TASK

- *Download the complete genome in fasta format for the Helicobacter pylori reference strain 26695 (taxID 85962) from RefSeq.*

Options to include in command to run package

- `--formats`
- `--assembly-levels`
- `--taxids`
- `groups`

COMMAND

```
(conda_workshop) [jrumore@waffles ~]$ ncbi-genome-download --formats fasta --assembly-levels complete --taxids 85962 bacteria
```

SLURM COMMAND

```
(conda_workshop) [jrumore@waffles ~]$ sbatch -c 1 --mem 4G -p NMLResearch --wrap="ncbi-genome-download --formats fasta --assembly-levels complete --taxids 85962 bacteria"
Submitted batch job 8568915
```

RUN A TOOL IN CONDA - EXAMPLE

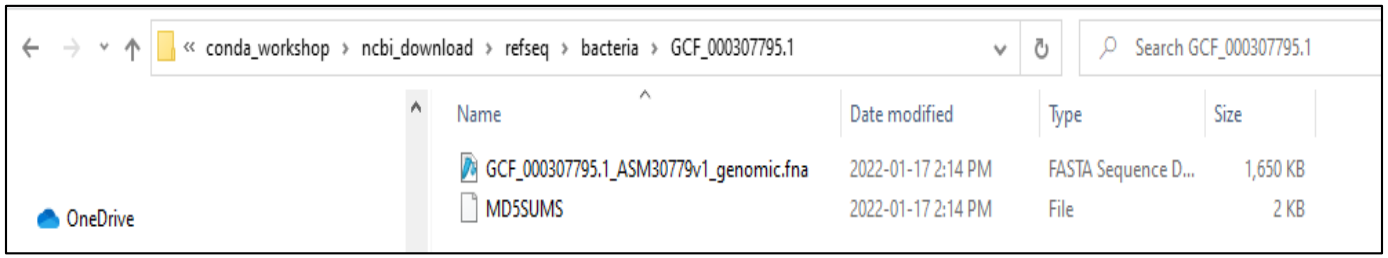
STEPS:

6. Verify the tool ran successfully.

TERMINAL

```
(conda_workshop) [jrumore@waffles ncbi_download]$ ls
refseq  slurm-8568915.out
(conda_workshop) [jrumore@waffles ncbi_download]$ cd refseq/
(conda_workshop) [jrumore@waffles refseq]$ ls
bacteria
(conda_workshop) [jrumore@waffles refseq]$ cd bacteria/
(conda_workshop) [jrumore@waffles bacteria]$ ls
GCF_000008525.1  GCF_000307795.1
(conda_workshop) [jrumore@waffles bacteria]$ cd GCF_000307795.1/
(conda_workshop) [jrumore@waffles GCF_000307795.1]$ ls
GCF_000307795.1_ASM30779v1_genomic.fna.gz  MD5SUMS
(conda_workshop) [jrumore@waffles GCF_000307795.1]$ gunzip GCF_000307795.1_ASM30779v1_genomic.fna.gz
(conda_workshop) [jrumore@waffles GCF_000307795.1]$ ls
GCF_000307795.1_ASM30779v1_genomic.fna  MD5SUMS
(conda_workshop) [jrumore@waffles GCF_000307795.1]$
```

FILE EXPLORER



TRY IT ON YOUR OWN!

STEPS:

1. Search for a package of your choice (https://anaconda.org/)
2. Review the package documentation (online and in the terminal).
3. Create a new conda environment and install the package.
4. Verify the environment was created and the package installed.
5. Set-up a project directory.
6. Activate the conda environment containing the package of interest.
7. Run the package.
8. Verify the package ran successfully.

FINAL THOUGHTS

- PRACTICE, PRACTICE, PRACTICE, and then PRACTICE some more
- Don’t be afraid to make mistakes
- Ask questions and/ or Google it!
- Be kind to your future self – take good notes

HELPFUL RESOURCES

https://www.youtube.com/watch?v=_VUynJ_CBJo

https://docs.conda.io/projects/conda/en/latest/user-guide/getting-started.html#managing-environments

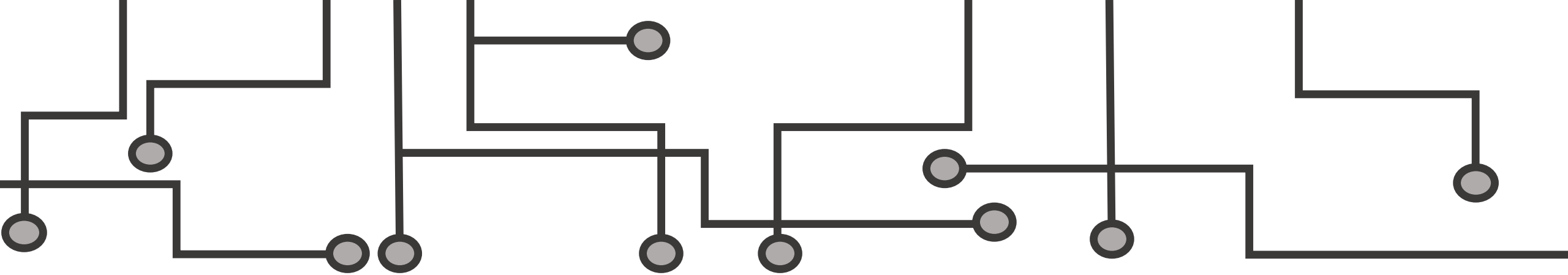
https://docs.conda.io/projects/conda/en/latest/commands.html

https://docs.conda.io/projects/conda/en/4.6.0/_downloads/52a95608c49671267e40c689e0bc00ca/conda-cheatsheet.pdf

https://slurm.schedmd.com/overview.html

https://slurm.schedmd.com/sbatch.html

https://help.dreamhost.com/hc/en-us/articles/115002768331-UNIX-commands-Decompressing-files



THANK YOU FOR ATTENDING!

The Q&A Session will now begin.

Please make sure to fill out the Exit Survey

We value your feedback!

More questions? Please email us at mmid.coding.workshop@gmail.com or post them to the workshop slack channel