Native MacOS installation of the ‘File Checker’, a.k.a. the import pipeline from *pdm\_utils*

Before beginning:

Recall that the SEA\_VM has a validation folder and a *validate* Terminal command to bring you into the validation folder and remind you how to invoke the import pipeline.

We will be emulating this structure on your Apple computer. The approach outlined below should work for any relatively recent MacOS operating system (10.11 or higher). Apple have recently changed the default shell from *bash* to *zsh*, which has minimal impact overall, but I’ll document how to account for either. If you have changed the shell to something else entirely, you’ll have to figure out how to compensate.

I highly recommend reading through the full instructions for each section before attempting it.

If you get stuck at any step, please do not proceed - contact me at [christian.gauthier@pitt.edu](mailto:christian.gauthier@pitt.edu), as the steps build on each other. An error in one will almost certainly result in errors in the next step(s).

Terminal commands will be indicated by *italicized, Menlo font*.

You should have received a few files in the same bundle you found this file in: config.txt, FlatfileQC.command, import\_table.csv, and an empty genomes folder. You should place these together in a new folder (for example /Users/your\_username/validation). This will be your equivalent to the SEA\_VM validation folder. Genomes to QC should go in the genomes folder, with a row in the import\_ticket.csv for each genome therein.

**NOTE**: if you open the import ticket with Excel, it will likely never work again, as Excel likes to put hidden characters that interfere with pdm\_utils ability to parse the file properly. Use TextEdit, or some other lightweight text editor to make changes to this file.

Let’s begin:

1. Figure out what your default shell is
   1. Open a Terminal (keep it open through the whole installation process)
   2. Run *echo $SHELL*
   3. If the output is /bin/bash, your shell is bash
   4. If the output is /bin/zsh, your shell is zsh
   5. If the output is something else, this document is not your friend
2. Install MySQL 5.7.31
   1. Download the [installer](https://downloads.mysql.com/archives/get/p/23/file/mysql-5.7.31-macos10.14-x86_64.dmg)
   2. Run the installer with all defaults. Enter your password if prompted.

**NOTE**: if your Mac does not let you install programs from anywhere, we will try to fix that for you. In your open Terminal, run *sudo spctl --master-disable* (enter your password when prompted). If you have administrative privileges on your computer, you should now be allowed to install any program you want without Apple stopping you (they may still try slowing you down). Now try installing again.

* 1. Make note of the temporary password generated by the MySQL installer (screenshot, or recent versions of MacOS may let you copy and paste from the popup window).
  2. Now we’re going to add MySQL to your system path. In your open Terminal:

bash users:

*nano ~/.bash\_profile*

*export PATH=’$PATH:/usr/local/mysql/bin’*

*Ctrl + X*

*Enter*

*Ctrl + O*

*source ~/.bash\_profile*

zsh users:

*nano ~/.zshrc*

*export $PATH=’$PATH:/usr/local/mysql/bin’*

*Ctrl + X*

*Enter*

*Ctrl + O*

*source ~/.zshrc*

**NOTE**: if your bash\_profile or zshrc file is not empty, use the arrow keys to move the cursor to the bottom of the file before proceeding. Ctrl + X, Enter saves the file. Ctrl + O closes nano and returns to the Terminal prompt. Source lets the terminal know about the changes we just made.

**NOTE**: if you’re copying and pasting, you’ll need to replace the quote characters by manually re-typing them. In nano you’ll have to use the arrow keys to navigate the cursor to the correct position.

* 1. In your System Preferences, there should now be a button at the bottom for MySQL. Click here, then click the button to ‘Start MySQL Server.’ The MySQL Server Instance should be set to “running.” You can also set MySQL to start automatically whenever your computer starts up by clicking the box.

Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

* 1. In your open Terminal:

*mysql -u root -p*

Enter the temporary password from before. It will hide what you type, so you won’t be able to check your spelling - copy and paste should work here. If successful, it should print some basic MySQL server information, then enter a mysql prompt.

*ALTER user ‘root’@’localhost’ IDENTIFIED BY ‘phage’;*

*FLUSH PRIVILEGES;*

*EXIT;*

**NOTE**: again if copying and pasting, all quote characters will need to be re-typed

manually, with the arrow keys being your means of navigating the cursor.

You should now have MySQL properly installed, and with the ‘root’ password set to ‘phage’ (same as every other SEA-PHAGES user, and the SEA\_VM...).

1. Install Miniconda3
   1. Download the [installer](https://repo.anaconda.com/miniconda/Miniconda3-py39_4.9.2-MacOSX-x86_64.sh). Depending on your browser, you may be prompted to open the file or save. Select “Save File.”

Graphical user interface, text, application

Description automatically generated

* 1. In your open Terminal:

*cd ~/Downloads && bash ./Miniconda3-py39\_4.9.2-MacOSX-x86\_64.sh*

Note: if you get the error message (-bash: bash: command not found), try pasting in the command

*export PATH=/usr/bin:/bin:/usr/sbin:/sbin:/usr/local/bin*

into terminal. Press Enter and then re-enter the command to install Miniconda3.

* 1. Hit Enter to review the license agreement. Hold Enter until prompted for ‘yes’ or ‘no’, to scroll through quickly. Or read it if you care to (Enter advances by one line). Enter ‘yes’ to agree to the license. Hit Enter to accept the default installation location. Enter ‘yes’ to allow conda initialization. When you are finished, close this terminal because you will be located in the Downloads folder. You can also navigate up to the main user directory.

**NOTE**: I might have got some specifics wrong here, but the gist is to just follow the instructions and accept defaults.

* 1. In your open Terminal:

bash users:

*source ~/.bash\_profile*

Note: if you get the error -bash: dirname: command not found, then you will need to do the following:

*nano ~/.bash\_profile*

Add in the following to the file

*export PATH=/usr/local/bin:/usr/bin:/bin:/usr/sbin:$PATH*

*Ctrl + X*

*Enter*

*Ctrl + O*

*source ~/.bash\_profile*

zsh users:

*source ~/.zshrc*

**NOTE**: this step will tell the shell what changes were made by the miniconda installer, namely addition of the *conda* command

* 1. In your open Terminal:

*conda config --set auto\_activate\_base false*

**NOTE**: this will prevent conda from loading the base environment every time the Terminal gets opened.

You should now have Miniconda3 properly installed.

1. Create conda environment and install *pdm\_utils*
   1. In your open Terminal

*conda create -n pdm\_utils python=3.8 -y*

*conda activate pdm\_utils*

*conda install -c bioconda -c conda-forge aragorn trnascan-se infernal -y*

*pip install pdm\_utils*

**NOTE**: if you get a compiler error, it will most likely be because you don’t have any developer tools installed yet. Apple were reasonably smart, and you should see a popup prompt you to install some components of the developer toolkit. Say yes, and wait for installation to complete, then you should be able to successfully re-run the failed command(s).

* 1. In your open Terminal:

*python3 -m pdm\_utils -h*

**NOTE**: here we are looking to see that pdm\_utils was installed properly. If it was, you should see a simple help menu that briefly shows the pipelines available through pdm\_utils.

You should now have the pdm\_utils toolkit installed, and the barebones components required to run the file checker (import pipeline).

1. Edit, and authorize running the FlatfileQC.command file
   1. In your open Terminal:

*cd /path/to/your/validation/folder*

*pwd*

copy the output to your clipboard - we’ll need it in a second

**NOTE**: your validation folder path should not have any spaces. My experience has been that it will not work if you have any spaces anywhere in your path.

**NOTE:** /path/to/your/validation/folder will need to be replaced by the real path to your validation folder (e.g. /Users/your\_username/validation)

* 1. In a Finder window, navigate to your validation folder. Right click the FlatfileQC.command, and open with TextEdit
  2. Do a search (Command + F) for *VALIDATION=*
  3. Immediately after the ‘=’, paste your validation folder’s path
  4. Save and quit TextEdit
  5. In your open Terminal:

*sudo chmod +x FlatfileQC.command*

You may be prompted to enter your computer’s password.

You should now be able to double-click to run the FlatfileQC.command file, which will perform all the steps to validate the flatfiles you put in the genomes folder. It will run interactively, and the import logs will be dumped to your validation folder.