

Portable Proteomics Pipeline (P3)

Use Case: FTP

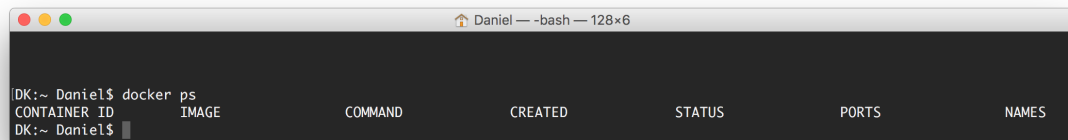
Daniel Kristiyanto (daniel.kristiyanto@pnnl.gov)

June 24, 2016

This is a use case/manual for a P3 pipeline with files retrieved from Pride database.

1. Make sure Docker engine is installed and running.

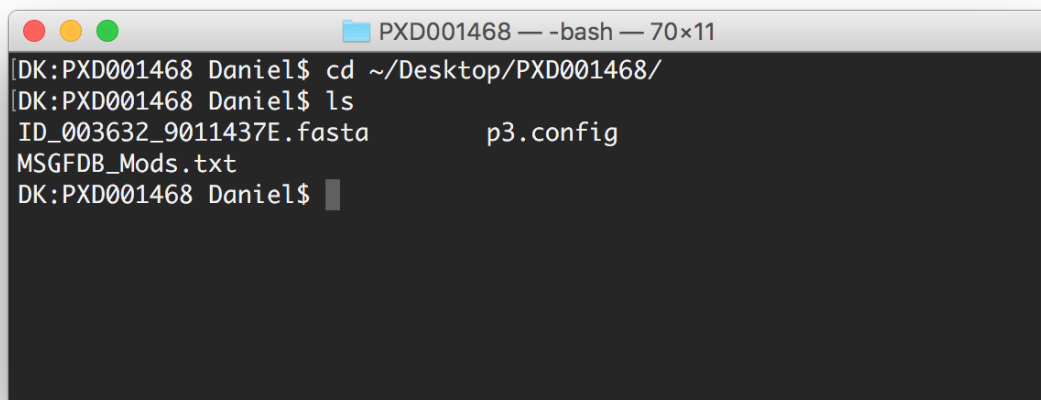
```
docker ps
```



```
DK:~ Daniel$ docker ps
CONTAINER ID    IMAGE    COMMAND    CREATED    STATUS    PORTS    NAMES
DK:~ Daniel$
```

2. Download a sample p3.config and MSGFDB_Mods.txt file and put it to a directory. e.g. ~/Desktop/PXD001468

```
cd ~/Desktop/PXD001468
```



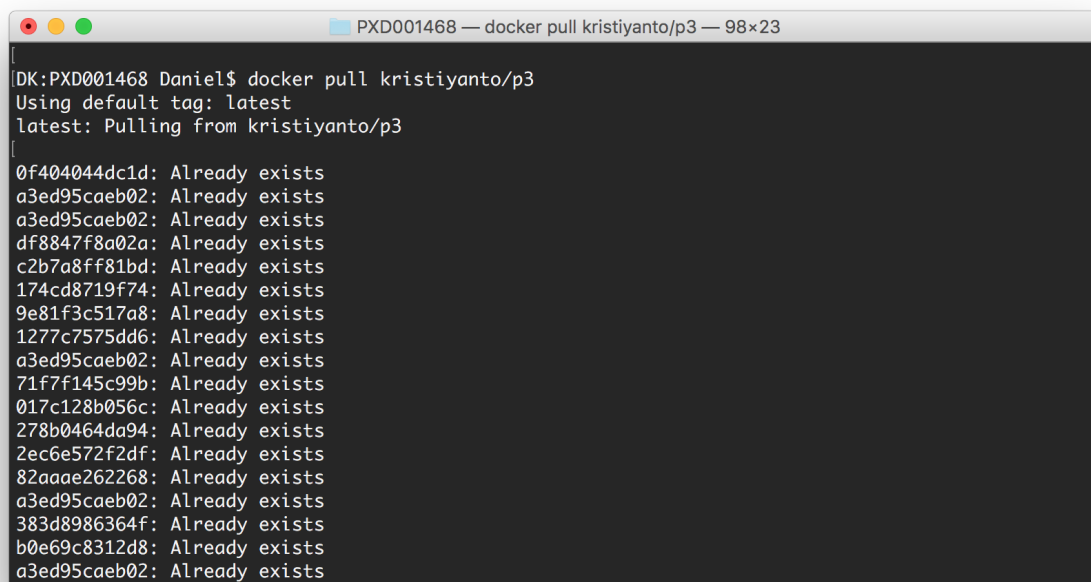
```
DK:PX001468 Daniel$ cd ~/Desktop/PXD001468/
DK:PX001468 Daniel$ ls
ID_003632_9011437E.fasta    p3.config
MSGFDB_Mods.txt
DK:PX001468 Daniel$
```

in this case, p3.config must have the following fields:

```
REPO = FTP
FTP1 = (e.g: ftp://ftp.pride.ebi.ac.uk/pride/data/archive/2015/06/PXD001468)
RUN_MSGF = (e.g: YES)
METHOD = (e.g: SPECTRUM_COUNT)
```

If other PrideID is desired, change the PRIDEID option in the p3.config. Ensure that the corresponding PrideID have the required files (*.fasta, *.mzml, etc.).

3. Download / update P3 from Dockerhub, and run the image.



```
PXD001468 — docker pull kristiyanto/p3 — 98x23
DK:PXD001468 Daniel$ docker pull kristiyanto/p3
Using default tag: latest
latest: Pulling from kristiyanto/p3
0f404044dc1d: Already exists
a3ed95caeb02: Already exists
a3ed95caeb02: Already exists
df8847f8a02a: Already exists
c2b7a8ff81bd: Already exists
174cd8719f74: Already exists
9e81f3c517a8: Already exists
1277c7575dd6: Already exists
a3ed95caeb02: Already exists
71f7f145c99b: Already exists
017c128b056c: Already exists
278b0464da94: Already exists
2ec6e572f2df: Already exists
82aaae262268: Already exists
a3ed95caeb02: Already exists
383d8986364f: Already exists
b0e69c8312d8: Already exists
a3ed95caeb02: Already exists
```

```
eval $(docker-machine env default)
docker pull kristiyanto/p3
docker run --rm -v ~/Desktop/PXD001468:/root/data kristiyanto/p3
```

```
PXD001468 — docker run --rm -v ~/Desktop/PXD001468:/root/data kristiyanto/p3 — 98x23
220c94cc3e8a: Pull complete
362f45fb0f4b: Pull complete
61e53057a09d: Pull complete
b3301924c329: Pull complete
Digest: sha256:55da5153dacf0de5d2e806c7dc31f34876a28ecf2b3e16d57821f9c405f86f2f
Status: Downloaded newer image for kristiyanto/p3:latest
DK:PXD001468 Daniel$ docker run --rm -v ~/Desktop/PXD001468:/root/data kristiyanto/p3
Bioconductor version 3.1 (BiocInstaller 1.18.3), ?biocLite for help
Loading required package: methods
[1] "Species: Homo sapiens (Human)"
[1] "Chick JM, Kolippakkam D, Nusinow DP, Zhai B, Rad R, Huttlin EL, Gygi"
[2] "SP. A mass-tolerant database search identifies a large proportion of"
[3] "unassigned spectra in shotgun proteomics as modified peptides. Nat"
[4] "Biotechnol. 2015 Jun 15"
[1] "ftp://ftp.pride.ebi.ac.uk/pride/data/archive/2015/06/PXD001468"
Starting at: 2016-06-24 20:52:17.152871
PXD001468
Reading from PRIDE REPOSITORY. Pride ID: PXD001468
Downloading from ftp://ftp.pride.ebi.ac.uk/pride/data/archive/2015/06/PXD001468
Downloading...
Downloading...
b1906_293T_proteinID_01A_QE3_122212.mzXML
```

if necessary, another image can be run simultaneously (in a different terminal windows).

```
docker run --rm -v ~/Desktop/PXD001468:/root/data kristiyanto/p3
```

4. Once processes completed, the directory will be populated with:

- Mass Spectrometry files downloaded from Pride repository (*.mzid, *.mzml)
- MSGF+ output files (*.mzid, *.canon, *.revCat.cnlp, *.revCat.csarr, *.revCat.cseq)
- Output table (*.txt)
- Output file as R objects (*.rda)

If the process stop returning error: * Make sure all the required files are in the folder. E.g: Fasta files sometimes are not provided in the FTP/Pride repository and must be provided manually. * Remove all *.tmp files before re-run the containers.