MetaProteomeAnalyser – FEM lab SOP

Feb 2017

by C Thorn

1. **Running MPA**

Should be a simple shortcut on desktop that you double click, but if not then. Open terminal.

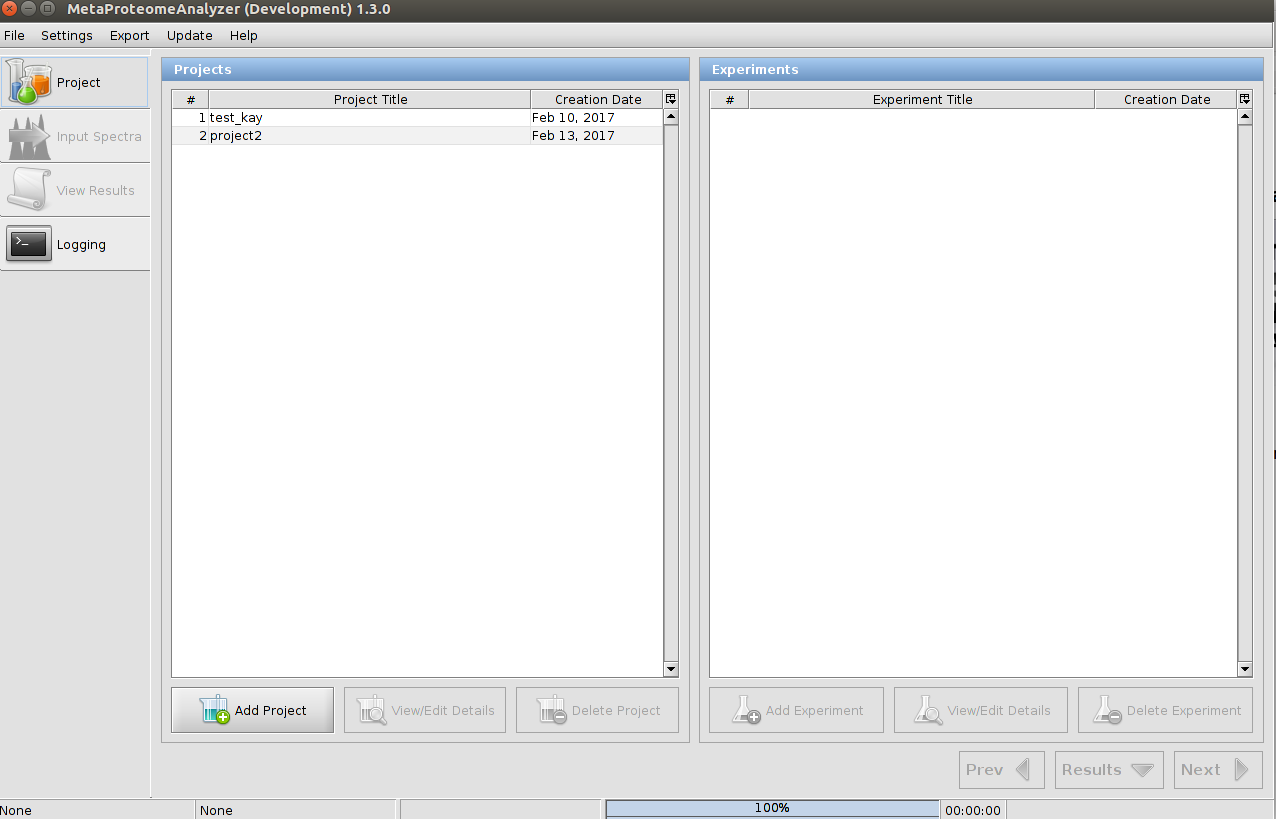
*cd scratch/metaprot/*

And then start the sh script:

*sh StartMPA.sh*

And MPA should open

2. **Getting started**



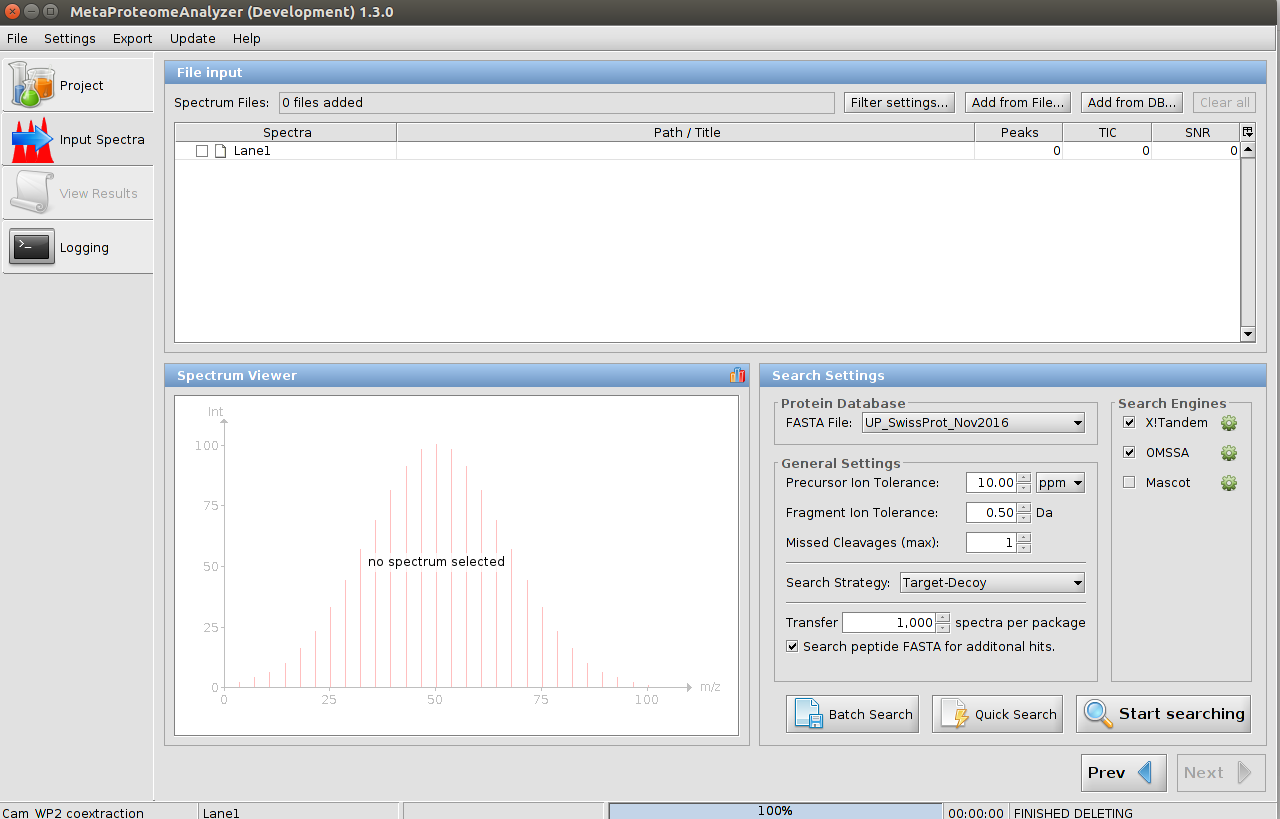
Home page will look like this. Start by clicking ‘Add Project’. Choose a name and ‘save’.

(The project would be the name of the work package, for example.)

Click on newly made project and then ‘Add experiment’. Typically this will be the equivalent of one biological replicate, ie one lane on an SDS gel (even if that lane was analysed in multiple chunks)

Then click ‘Next’ and proceed to upload page

3. **Uploading spectra files.**

These should be .mgf files. If you have .dat files and want to use Mascot (and have paid for the licence :o ) then see appendix ‘Mascot files’

There are two ways to do this:

a) old way: click ‘Add from file’ on top right of screen. Navigate to your files and select all mgf files for that one SDS lane/bio rep. Click ‘open’ & upload will begin.

Wait for it to finish loading your spectra. Click in the little white box by the name of your spectra files in order to select all the spectra (ie in order to ignore the filter setting applied by MPA – see Appendix 2 on FDR if you want to know more)

Then click ‘Start Searching’

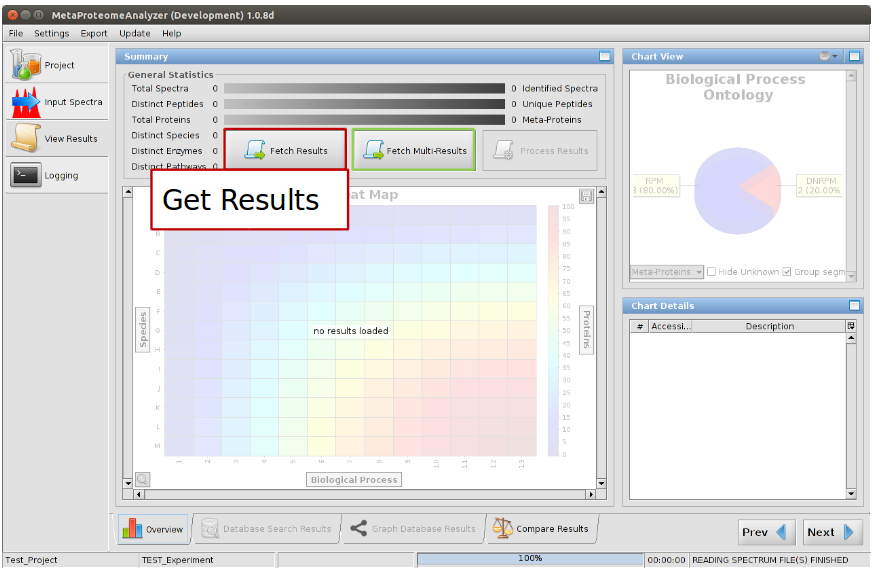
This will typically take hours to ~1 day/night.

b) Click, ‘Quick search’ button & choose files in same way. Is meant to be more efficient than old way if you have big files. If you do use this, you need to have your search settings already chosen, before you choose your files, as spectra searches as it uploads them. See next section on search settings.

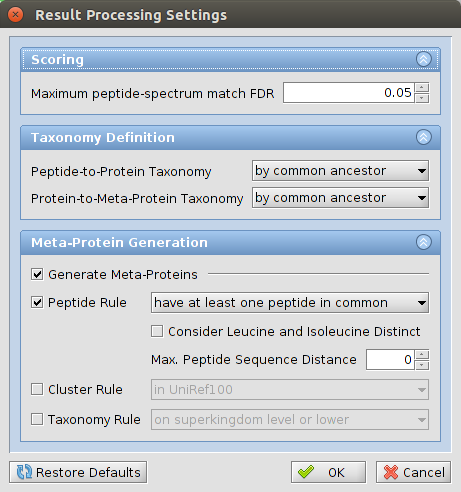
**4. Search settings**

This depends on your Mass Spec run and the information should be available from your sequencing centre. If you are unable to get this information for any reason, just leave the default settings

4. **Loading and analysing results**



Wait for results to download click ‘process results’ button & the following window will appear.



In here you will have the choice of how you want to select for your metaproteins and also what false detection rate (FDR) you will you use. If unsure, keep defaults but make a note of what you chose. For more details on what FDR to choose, see appendix II.

You can also choose here how you want to group your metaproteins. There is information about the different metaprotein generation methods in the supplementary material of the MPA release publication.

As a start, just leave the defaults as they are.

**Adding a metagenome as a search database**

Using the JGI or similar, find a metagenome from a similar sample type to your own (eg temperate grassland) and download the .faa or .fasta file.

Try this link (you’ll need to make a free account with JGI in order to download):

https://img.jgi.doe.gov/cgi-bin/m/main.cgi

Once file is unzipped, copy it into the folder > scratch/metaprot/data/fasta/

Once you have downloaded your file, open it and check the prefix to the peptides. They should look like one of these:

UNIPROTSPROT (">sp|"),

UNIPROTTREMBL (">tr|"),

NCBIGENBANK (">gi|"),

NCBIREFERENCE (">ref|"),

METAGENOME1 (">generic|"),

SILICO\_PEPTIDE (">pep|"),

METAGENOME2 (">|"),

METAGENOME3 (">");

If they don’t, it probably means it is not compatible with MPA and the search won’t be able to find any results.

Loading the file:

In the ‘Update’ dropdown menu at the top of MPA, choose ‘Add Database’

Navigate to your file, click open and MPA will start doing its thing. This can take some hours, depending on metagenome, so a good one to leave running overnight.

Rothamstead metagenomes

<ftp://ftp-adn.ec-lyon.fr/Metasoil-datasets/>

**Changing between databases**

Navigate to

scratch>metaprot

Open file called config.properties

Should open in text editor

On about fourth line of text, change text after

dbName:

to the name of the database you want to change to. At the moment (feb 2017), we have two databases:

1) all data run prior to January 2017 on a database called **metaprot\_olddata**.

2) Data from after January 2017 is stored in database called **metaprot**

Appendix I

Mascot search

- do Mascot searches on the Mascot server (wherever that is)

- get the \*.dat files, which contain Mascot results

- load those files using "Batch Search"-button (name of button as of Feb 2017 – will soon be changed though)

- continue through MPA as normal.

Appendix II

False discovery rate