Tutorial for the PNNL Biodiversity Library Skyline Plugin

Developed By: Michael Degan, Grant Fujimoto, Lillian Ryadinskiy, Samuel H Payne

Contact Information: Samuel Payne (Samuel.Payne@pnnl.gov)

Michael Degan (Michael.Degan@pnnl.gov)

Grant Fujimoto (Grant.Fujimoto@pnnl.gov)

This tutorial is to help users install and use the PNNL Biodiversity Library plugin with Skyline. Troubleshooting help will follow at the end of the tutorial. We note that this plugin does not work with Windows XP and must be used with Skyline version 3.1.1.7490 or greater.
Further information about the Biodiversity Library available at:

http://omics.pnl.gov/project-data/biodiversity-library

Installation:
The PNNL Biodiversity Library Plugin for Skyline can be installed through the Skyline tool store interface.
Tools -> Tool Store -> PNNL Biodiversity Library

Running the Plugin:

Select an Organism:

You are able to select your desired organisms in one of two ways, either through the Phylogeny Explorer or by directly searching for the organism by name.

The Phylogeny Explorer organizes organisms through scientific naming convention, i.e. Kingdom, Phylum, Class and then individual strain of the organism. Human data is additionally broken down by tissue.

The Organism search (right panel) allows for users to specifically enter a name.

.

Select Pathways of Interest:

After you have selected an organism, you then choose the KEGG pathways you are interested in investigating within that organism. Alongside the pathway, the coverage metric is dynamically calculated for each Organism/Pathway combination. You can select as many of these pathways as you desire.

Select proteins of interest through KEGG Pathway maps:

The images for the pathways you selected will be created dynamically, highlighting all the KEGG orthologs for the organism in the pathway. If that ortholog is part of the set of data which has been observed in the MS/MS space, it will be highlighted initially as red, while if it wasn't it will be highlighted blue. From here, you can select the proteins which are of interest to you in the pathway in two ways, either by clicking on individual ortholog boxes or by clicking the select/deselect all buttons on the right hand side of the screen. If an ortholog does become deselected it will then be highlighted as grey to show that the information will not be pushed to the next step of the process.

Additionally, you can save the pathway image by clicking the button on the right-hand side of the screen, below the Deselect All button. This will save the image for the pathway along with your current selection of orthologs as a .png file with a resolution of 300 dpi.

Fine control of proteins to export:

On the Review and Export tab, the application will display the total number of genes selected for review from all of your Organism/Pathway combinations of interest as well as an individual breakdown of the proteins. If there are Organism/Pathway combinations you are no longer interested in, or individual proteins which you do not wish to export to Skyline, you can prevent these from being exported by simply unchecking the box for the row(s) you are no longer interested in.

Export data to Skyline:

Once you are satisfied with your selection of combinations and/or proteins, you can then either select another organism to follow the same workflow, with its own selection of pathways and proteins, or export the data of interest to Skyline.

Once you confirm the export to skyline, the application will create a FASTA for the proteins you indicated and download the Spectral Library files for the organisms of interest from the Biodiversity Library's Repository on MassIVE (ftp://massive.ucsd.edu/MSV000079053/library/). Once all of the information is downloaded, it is then imported to Skyline for you to use in your experiments.

View your data in Skyline:

The plugin will automatically propagate the information downloaded into Skyline, where it uses your peptide settings to create split the proteins imported into peptides and uses the Spectral Library downloaded through the plugin to perform library matching.
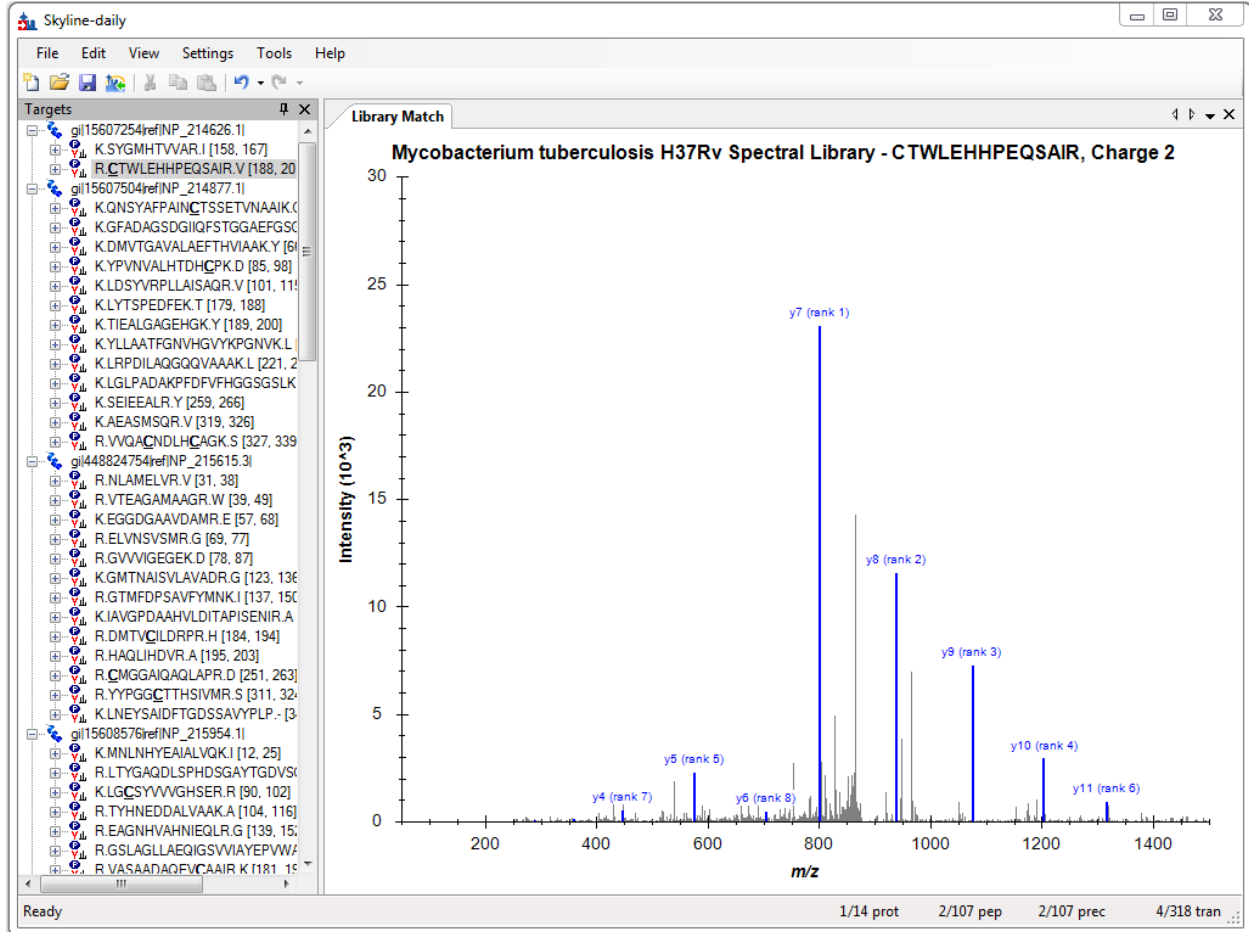
Troubleshooting

- The application is taking an unusually long time while generating a FASTA from my selection of proteins.
  - o Depending on your Skyline settings, Skyline could be seeing that there are proteins incoming from this FASTA which do not contain peptides according to the current filter settings. While this window is open from Skyline, our application is waiting for a response. To allow the information to be fully exported, select "Keep" so that the proteins will export and the .blib file(s) which download afterwards will provide a library match and populate the peptides for these new proteins.
- No proteins appeared as red in the KEGG pathway maps.
  - o This could be due to two possibilities:
    - ▪ The organism contains proteins which should be observed in MS/MS space on this pathway, but were not observed in experimental results. These ortholog boxes will still be highlighted as blue.
    - ▪ The organism does not contain proteins which should be observed in MS/MS space on this pathway. In this case, no ortholog boxes will be highlighted on the image at all.
- Spectral Library doesn't download.
  - o Due to the server based location of the Spectral libraries, our application does require a stable internet connection to acquire these files. If you have confirmed that your internet connection is stable, the issue is most likely due to the repository on MassIVE being inaccessible. Please contact either ccms-web@proteomics.ucsd.edu or ccms-web@cs.ucsd.edu for further assistance with their FTP connection.
- No proteins were downloaded from NCBI.
  - o Due to the application creating FASTA information dynamically based on protein selection, our application does require a stable internet connection to acquire these files from NCBI at runtime. If you have confirmed that your internet connection is stable, the issue is most likely due to an issue with NCBI's FTP connection. Please send an email directly to info@ncbi.nlm.nih.gov for further assistance with their FTP connection.
- My favorite protein was not found or exported to Skyline.
  - o There are a number of reasons why a specific protein was not pushed into skyline.
    - ▪ Skyline peptide settings (missed cleavages)
    - ▪ Protein not observed in spectral library
    - ▪ Protein is not functionally classified by KEGG
  - o To manually insert new protein:
    - ▪ File -> Insert -> FASTA and copy your FASTA snippet in (Skyline will automatically populate the protein with peptides based on your digestion settings)
    - ▪ Edit -> Insert -> Proteins and fill out the table.
  - o To manually insert a peptide go to Edit -> Insert -> Peptides and fill out the table.
- No proteins were imported into Skyline
  - o This could be due to two possibilities:
    - ▪ The plugin was unable to obtain FASTA information for all proteins selected for export. NCBI or your local computer may be experiencing connectivity issues.
    - ▪ No proteins were selected for export from the plugin. The Spectral Library for the organism(s) examined still should export, but a message will display informing you that no proteins were selected and thus no FASTA was created

- Error Messages:
  - My version of Skyline is out of date.
    - Our application requires Skyline to be version 3.1.1.7490 or later. To update to the most recent version, please go to [Skyline Daily's webpage](#) to download the latest version of Skyline.
  - Unable to establish connection to NCBI to acquire FASTA for your organisms.
    - The plugin querys NCBI in real time and requires a stable internet connection. NCBI or your local computer may be experiencing connectivity issues.
  - MassIVE Server Unreachable
    - The plugin downloads the spectral library files from the UCSD MassIVE FTP server and requires a stable internet connection. MassIVE or your local computer may be experiencing connectivity issues.
- How do I add my data to the Biodiversity Library?
  - The Biodiversity Library is curated by Samuel Payne at Pacific Northwest National Laboratory. For more information about your data being visible in the Library, please contact Samuel.Payne@pnnl.gov.