

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:

Overview

Organisms

Pathways

Select Proteins by Pathway

Review and export to Skyline

PNNL Biodiversity Library

The PNNL Biodiversity Library is designed to provide an easy interface for retrieving mass spectrometry data. This data can be exported into Skyline to assist in SRM assay design or DIA data analysis. The tool exposes peptides identified in MS/MS by allowing users to select an organism and biological pathway of interest. In total the Biodiversity Library catalogs MS/MS spectra from 2 million peptides and 200,000 proteins from 118 distinct organisms across the tree of life. All proteins are cross referenced to KEGG pathways for intuitive biological interpretation. The Library was developed by Sam Payne at [Pacific Northwest National Laboratory](http://www.pnl.gov) with data collected over 10+ years in hundreds of collaborative projects.

The wizard helps users browse data using the following steps:

- 1) Select an Organism.
- 2) Select pathways on interest.
- 3) Curate protein identifications.
- 4) Review and export data to Skyline.

Source: [Biodiversity Library v2.1.6](#) Tuesday, May 19, 2015
Funding: US DOE, [Biological and Environmental Research](#)
Contact: [Samuel Payne](mailto:Samuel.Payne@pnnl.gov)

U.S. DEPARTMENT OF ENERGY

Next >

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.

The screenshot displays the PNNL Biodiversity Library interface. On the left is a navigation sidebar with buttons for 'Overview', 'Organisms', 'Pathways', 'Select Proteins by Pathway', and 'Review and export to Skyline'. The main content area is titled '1) Select an Organism.' and contains two panels: 'Phylogeny Explorer' and 'Organism search'. The 'Phylogeny Explorer' panel shows a hierarchical tree of taxonomic levels, with 'Mycobacterium tuberculosis H37Rv' highlighted in a red box. The 'Organism search' panel is a scrollable list of various bacterial species names.

1) Select an Organism.
Data in the Biodiversity Library is organized by the organism. To retrieve data, please select an organism of interest, either through the Phylum/Class explorer or through the Organism search box. If you want to analyze data from multiple organisms, please export the data one organism at a time.

Phylogeny Explorer

- Archaea
- Bacteria
 - Acidobacteria
 - Actinobacteria
 - Actinobacteria
 - Actinobacteridae
 - Coriobacteridae
 - Gardnerella
 - Mycobacterium
 - Mycobacterium tuberculosis H37Rv**
 - Alphaproteobacteria
 - Bacteroidetes
 - Chlorobi
 - Chloroflexi
 - Crenarchaeota
 - Cyanobacteria
 - Deinococcus-Thermus
 - Euryarchaeota
 - Fibrobacteres/Acidobacteria Group
 - Firmicutes
 - Proteobacteria
 - Spirochaetes
 - Thermotogae
- Eukaryota

Organism search

- Acidiphilium cryptum JF-5
- Actinosynnema mirum DSM 43827
- Anabaena variabilis ATCC 29413
- Anaeromyxobacter dehalogenans 2CP-C
- Anaplasma phagocytophilum HZ
- Arthrobacter sp. FB24
- Bacillus anthracis str. Ames
- Bacillus anthracis str. Sterne
- Bacillus subtilis subsp. subtilis str. 168
- Bartonella henselae str. Houston-1
- Borrelia burgdorferi B31
- Brachy bacterium faecium DSM 4810
- Burkholderia mallei ATCC 23344
- Candidatus Chloracidobacterium thermophilum B
- Candidatus Pelagibacter ubique HTCC1062
- Caulobacter crescentus CB15
- Cellulomonas flavigena DSM 20109
- Cenarchaeum symbiosum A
- Chlorobium tepidum TLS
- Chloroflexus aurantiacus J-10-fl
- Cryptobacterium curtum DSM 15641
- Cyanothece sp. ATCC 51142
- Cyanothece sp. PCC 7424
- Cyanothece sp. PCC 7425
- Cyanothece sp. PCC 7822
- Cyanothece sp. PCC 8801
- Cyanothece sp. PCC 8802
- Dehalococcoides mccartyi 195
- Deinococcus radiodurans R1
- Desulfovibrio alaskensis G20
- Desulfovibrio desulfuricans ND132

< Previous Next >

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.

The screenshot shows a web application window titled "PNNL Biodiversity Library". On the left is a vertical navigation menu with five items: "Overview", "Organisms", "Pathways", "Select Proteins by Pathway", and "Review and export to Skyline". The "Pathways" item is currently selected and highlighted in dark blue.

The main content area is titled "2) Select pathways of interest." and includes a sub-header "Kegg Pathways". Below this is a list of pathways under the "Metabolism" category. Each pathway has a checkbox and a coverage percentage. The "Fructose and mannose metabolism" pathway is checked and highlighted with a red border. The coverage for this pathway is "80.00% covered in MSMS".

Other pathways listed include:

- Glycolysis / Gluconeogenesis (100.00% covered in MSMS)
- Citrate cycle (TCA cycle) (100.00% covered in MSMS)
- Pentose phosphate pathway (88.89% covered in MSMS)
- Pentose and glucuronate interconversions (80.00% covered in MSMS)
- Galactose metabolism (100.00% covered in MSMS)
- Ascorbate and aldarate metabolism (75.00% covered in MSMS)
- Starch and sucrose metabolism (96.00% covered in MSMS)
- Amino sugar and nucleotide sugar metabolism (92.00% covered in MSMS)
- Pyruvate metabolism (96.15% covered in MSMS)
- Glyoxylate and dicarboxylate metabolism (100.00% covered in MSMS)
- Propanoate metabolism (100.00% covered in MSMS)
- Butanoate metabolism (100.00% covered in MSMS)
- C5-Branched dibasic acid metabolism (100.00% covered in MSMS)
- Inositol phosphate metabolism (100.00% covered in MSMS)

Below the list are several collapsed categories: Energy metabolism, Lipid metabolism, Nucleotide metabolism, Amino acid metabolism, Metabolism of other amino acids, Glycan biosynthesis and metabolism, Metabolism of cofactors and vitamins, Metabolism of terpenoids and polyketides, and Biosynthesis of other secondary metabolites.

On the right side of the main content area, there is a box titled "Pathways Selected" which currently contains the text "Fructose and mannose metabolism".

At the bottom of the interface are two buttons: "< Previous" and "Next >".

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.

3) Curate protein list for Organism: Mycobacterium tuberculosis H37Rv
Pathways specified in Step 2 are overlaid with proteomics data for the organism selected in step 1. Proteins identified in the Biodiversity Library are shown in red. Proteins which are annotated in the genome but are not present in the MS data are shown in blue. Users can deselect any protein they wish to exclude by clicking on the corresponding red box on the pathway image. (Deselected proteins appear as grey boxes.)

Fructose and mannose metabolism

Legend:
[Red Box] Protein annotated in Mycobacterium tuberculosis H37Rv and observed in MS/MS data
[Blue Box] Protein annotated in Mycobacterium tuberculosis H37Rv and not observed in MS/MS data

FRUCTOSE AND MANNOSE METABOLISM

Enzymes shown in the pathway include: 1.1.1.21, 1.1.1.14, 1.1.1.15, 1.1.1.11, 1.1.1.67, 1.1.1.38, 3.1.3.-, 3.2.1.80, 2.7.1.1, 2.7.1.4, 5.3.1.5, 5.3.1.7, 2.7.1.69, 2.7.1.1, 2.7.1.7, 5.4.2.8, 3.2.1.78, 3.2.1.77, 3.2.1.137, 3.6.1.21, 2.7.7.13, 2.7.7.22, 3.6.1.21, 1.1.1.132, 2.4.1.-, 2.4.1.-, 2.4.1.217, 2.4.1.269, 2.4.1.33, 1.1.1.132, 2.4.1.33, 2.4.1.217, 2.4.1.269.

Metabolites shown include: L-Sorbose, α-D-Glucose, D-Sorbitol, D-Fructose, Fructan, D-Mannitol, D-Fructose-2P, D-Mannose, D-Mannose-6P, D-Mannose-1P, GDP-D-mannuronat, GDP-D-mannose, 1,4-β-Mannan, Mannan, ADP-mannose, and Man-3-ph.

Buttons: < Previous, Next >

Right Panel: All Proteins Selected, Select All, Deselect All, Save Pathway Image

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.

4) Review and export data to Skyline.
From this final tab, a table of genes for each organism and pathway are shown for review. By hitting the export button, all relevant data will be transferred to Skyline. This includes a FASTA file for all selected proteins, as well as the identified peptides in a bibliospec library.

Organism	Pathway	# of Genes	Export?
Mycobacterium tuberculosis H37Rv	Fructose and mannose metabolism	14	<input checked="" type="checkbox"/>

Organism: Mycobacterium tuberculosis H37Rv
Pathway: Fructose and mannose metabolism
Genes (14)

Export?	Accession	Protein name
<input checked="" type="checkbox"/>	NP_214626	GDP-mannose 4,6-dehydratase
<input checked="" type="checkbox"/>	NP_214877	fructose-bisphosphate aldolase
<input checked="" type="checkbox"/>	NP_215615	fructose 1,6-bisphosphatase
<input checked="" type="checkbox"/>	NP_215954	triosephosphate isomerase
<input checked="" type="checkbox"/>	NP_216027	GDP-D-mannose dehydratase GmdA
<input checked="" type="checkbox"/>	NP_216028	nucleotide-sugar epimerase EpiA
<input checked="" type="checkbox"/>	NP_216545	6-phosphofructokinase PfkB
<input checked="" type="checkbox"/>	YP_177884	ribose-5-phosphate isomerase B
<input checked="" type="checkbox"/>	NP_217526	6-phosphofructokinase
<input checked="" type="checkbox"/>	NP_217629	phosphatase
<input checked="" type="checkbox"/>	NP_217772	mannose-6-phosphate isomerase
<input checked="" type="checkbox"/>	NP_217774	phosphomannomutase PmmA
<input checked="" type="checkbox"/>	YP_177951	D-alpha-D-mannose-1-phosphate guanylyltransferase ManB
<input checked="" type="checkbox"/>	NP_217875	phosphomannomutase PmmB

<< Select another Organism < Previous Confirm to Export

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.

The screenshot shows the 'PNNL Biodiversity Library' application window. On the left is a navigation sidebar with buttons for 'Overview', 'Organisms', 'Pathways', 'Select Proteins by Pathway', and 'Review and export to Skyline'. The main content area is titled '4) Review and export data to Skyline.' and contains a summary table and a detailed list of genes.

4) Review and export data to Skyline.
From this final tab, a table of genes for each organism and pathway are shown for review. By hitting the export button, all relevant data will be transferred to Skyline. This includes a FASTA file for all selected proteins, as well as the identified peptides in a bibliospec library.

Organism	Pathway	# of Genes	Export?
Mycobacterium tuberculosis H37Rv	Fructose and mannose metabolism	14	<input checked="" type="checkbox"/>

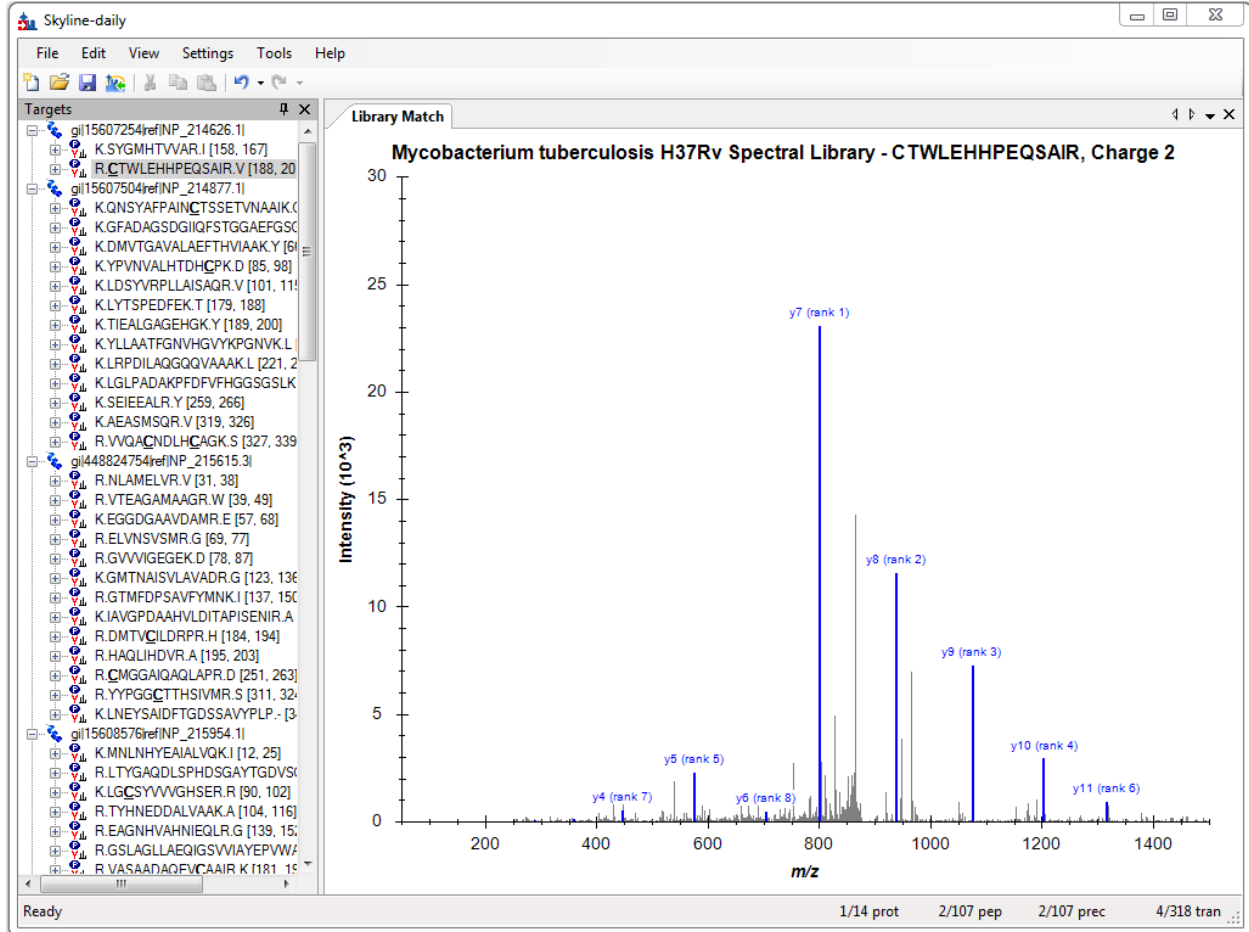
Organism: Mycobacterium tuberculosis H37Rv
Pathway: Fructose and mannose metabolism
Genes (14)

Export?	Accession	Protein name
<input checked="" type="checkbox"/>	NP_214626	GDP-mannose 4,6-dehydratase
<input checked="" type="checkbox"/>	NP_214877	fructose-bisphosphate aldolase
<input checked="" type="checkbox"/>	NP_215615	fructose 1,6-bisphosphatase
<input checked="" type="checkbox"/>	NP_215954	triosephosphate isomerase
<input checked="" type="checkbox"/>	NP_216027	GDP-D-mannose dehydratase GmdA
<input checked="" type="checkbox"/>	NP_216028	nucleotide-sugar epimerase EpiA
<input checked="" type="checkbox"/>	NP_216545	6-phosphofructokinase PfkB
<input checked="" type="checkbox"/>	YP_177884	ribose-5-phosphate isomerase B
<input checked="" type="checkbox"/>	NP_217526	6-phosphofructokinase
<input checked="" type="checkbox"/>	NP_217629	phosphatase
<input checked="" type="checkbox"/>	NP_217772	mannose-6-phosphate isomerase
<input checked="" type="checkbox"/>	NP_217774	phosphomannomutase PmmA
<input checked="" type="checkbox"/>	YP_177951	D-alpha-D-mannose-1-phosphate guanylyltransferase ManB
<input checked="" type="checkbox"/>	NP_217875	phosphomannomutase PmmB

At the bottom of the window, there are three buttons: '<< Select another Organism', '< Previous', and 'Confirm to Export' (with a bar chart icon). The first and last buttons are highlighted with red boxes in the image.

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.
 - 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.
 - 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.
 - 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.
 - 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.
 - 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
 - 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.
 - To manually insert a peptide go to Edit -> Insert -> Peptides and fill out the table.
- No proteins were imported into Skyline
 - 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:
 - When the Biodiversity Plugin encounters an error, the user is prompted with the option to send an error report. We highly encourage all users who encounter problems to send the error report along with a description of what lead up to the failure. These messages allow our developers to emulate the failure state so the problem can be diagnosed and fixed. We will try to address issues as soon as possible.
 - 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 queries 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@pnl.gov.