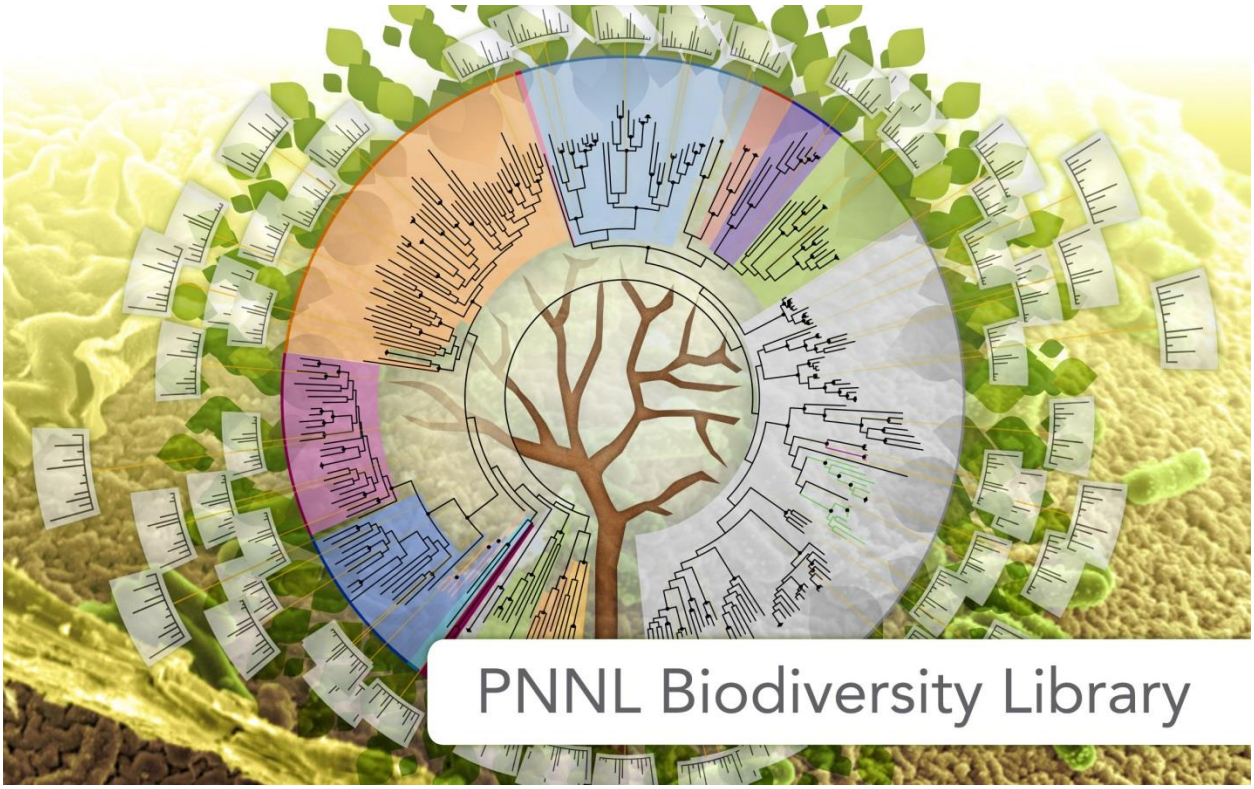


Tutorial for the PNNL Biodiversity Library Skyline Plugin

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. For information about the Skyline program, including how to download, please visit

https://skyline.gs.washington.edu/labkey/project/home/begin.view?

Further information about the Biodiversity Library available at:

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

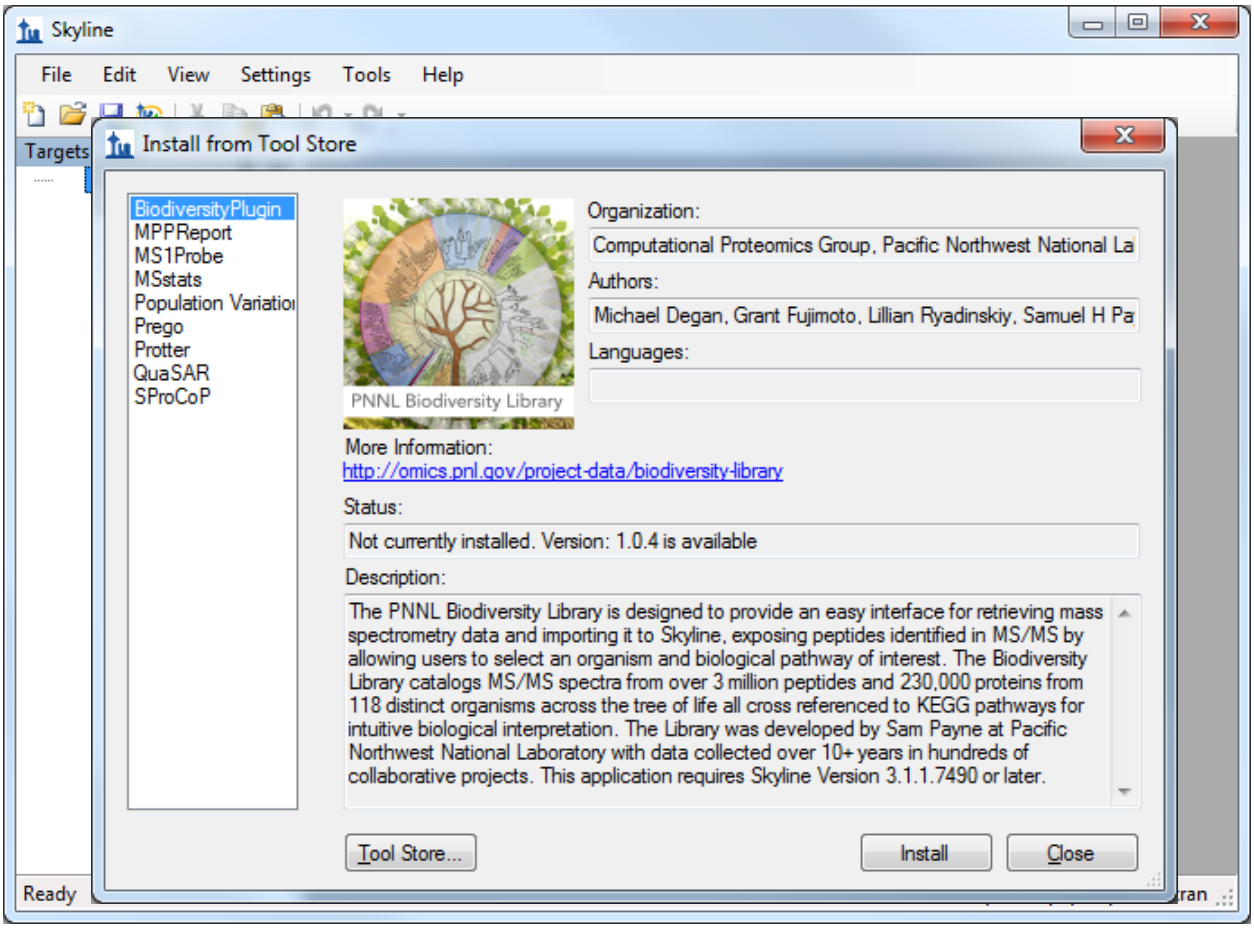
Program Goal

The goal of the plug-in is to help users browse the large spectral libraries from the PNNL Biodiversity Library and import relevant data into Skyline. Additional functionality allows users to import their own data into the plugin for easy navigation and loading into Skyline.

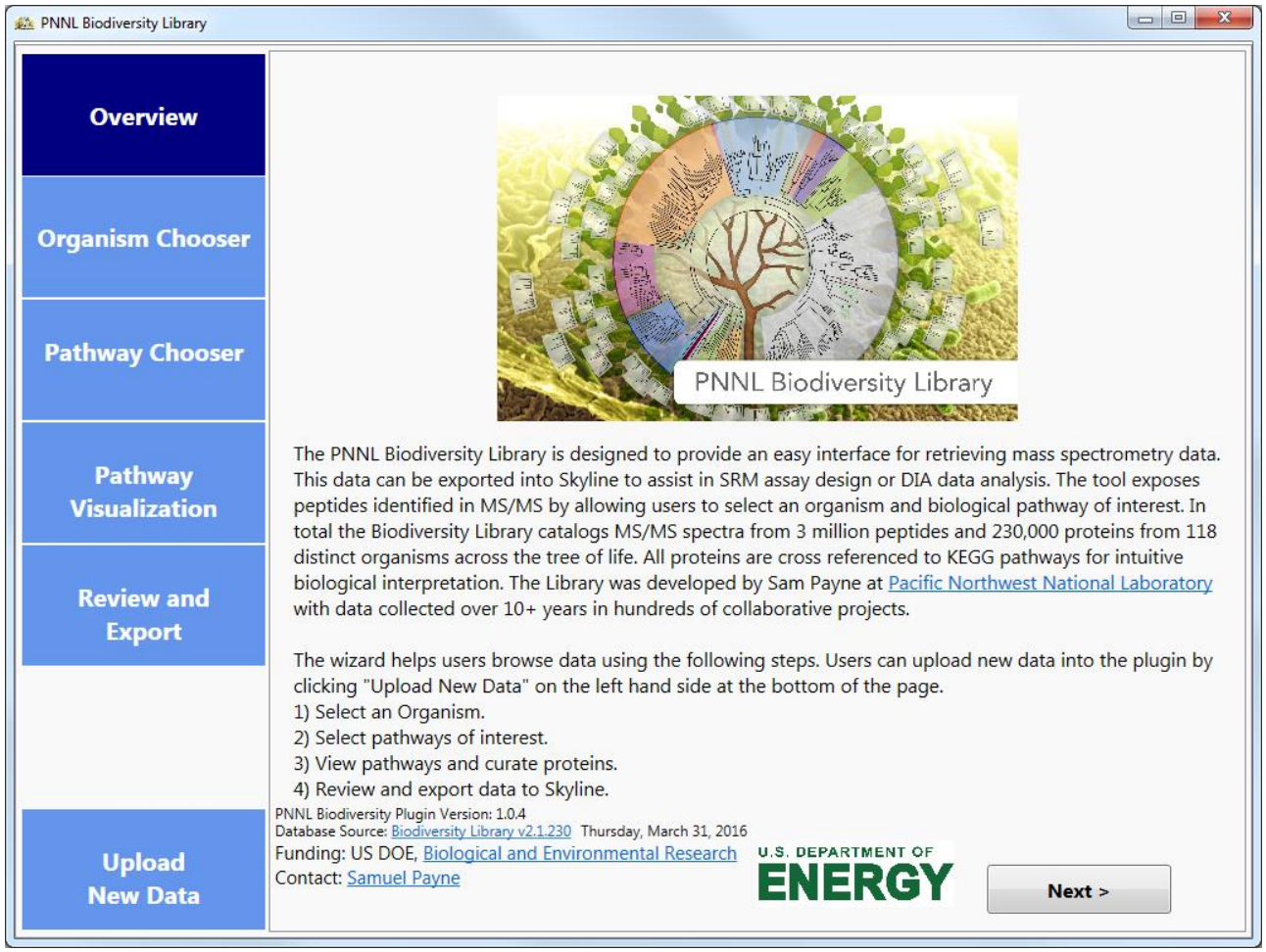
Table of Contents

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:



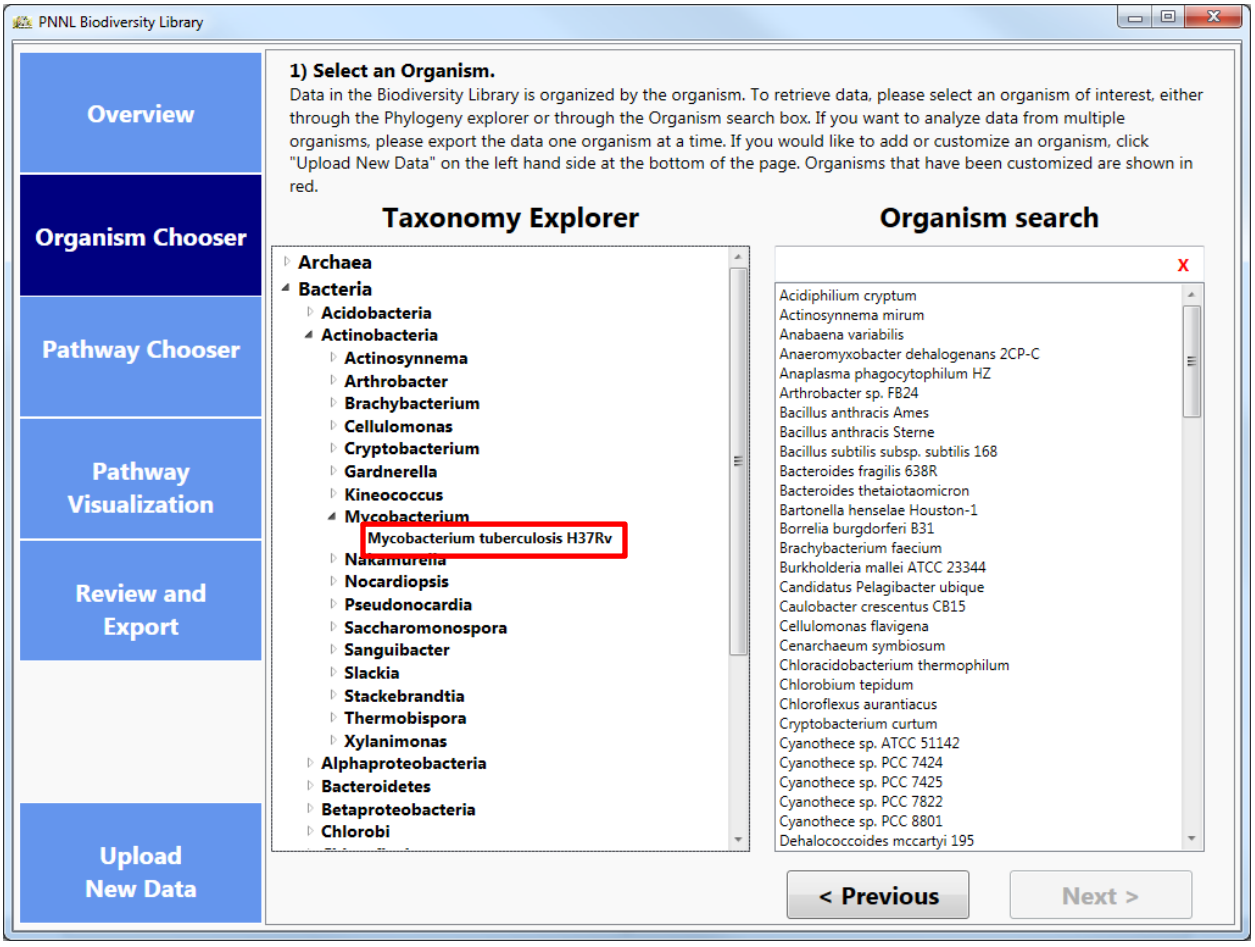
The plug-in is a wizard that walks users through the process of importing data into Skyline.

Select an Organism:

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

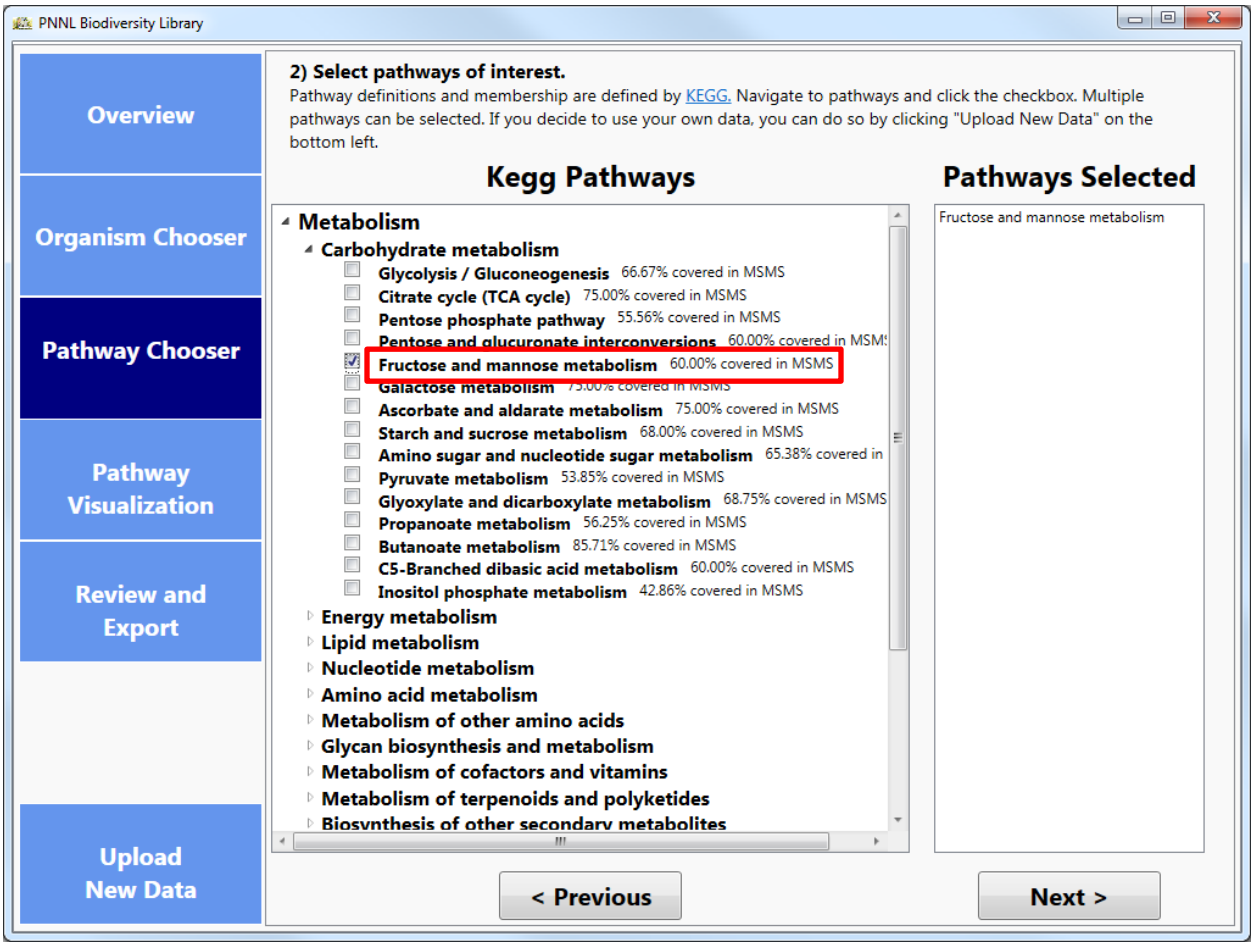
The Taxonomy Explorer organizes organisms through scientific naming convention, i.e. Kingdom, Phylum, Class and then individual strain of the organism. Human data is additionally broken down by tissue.

The Organism search (right panel) allows for users to type in a name and auto-filters the dataset to matching organisms.



Select Pathways of Interest:

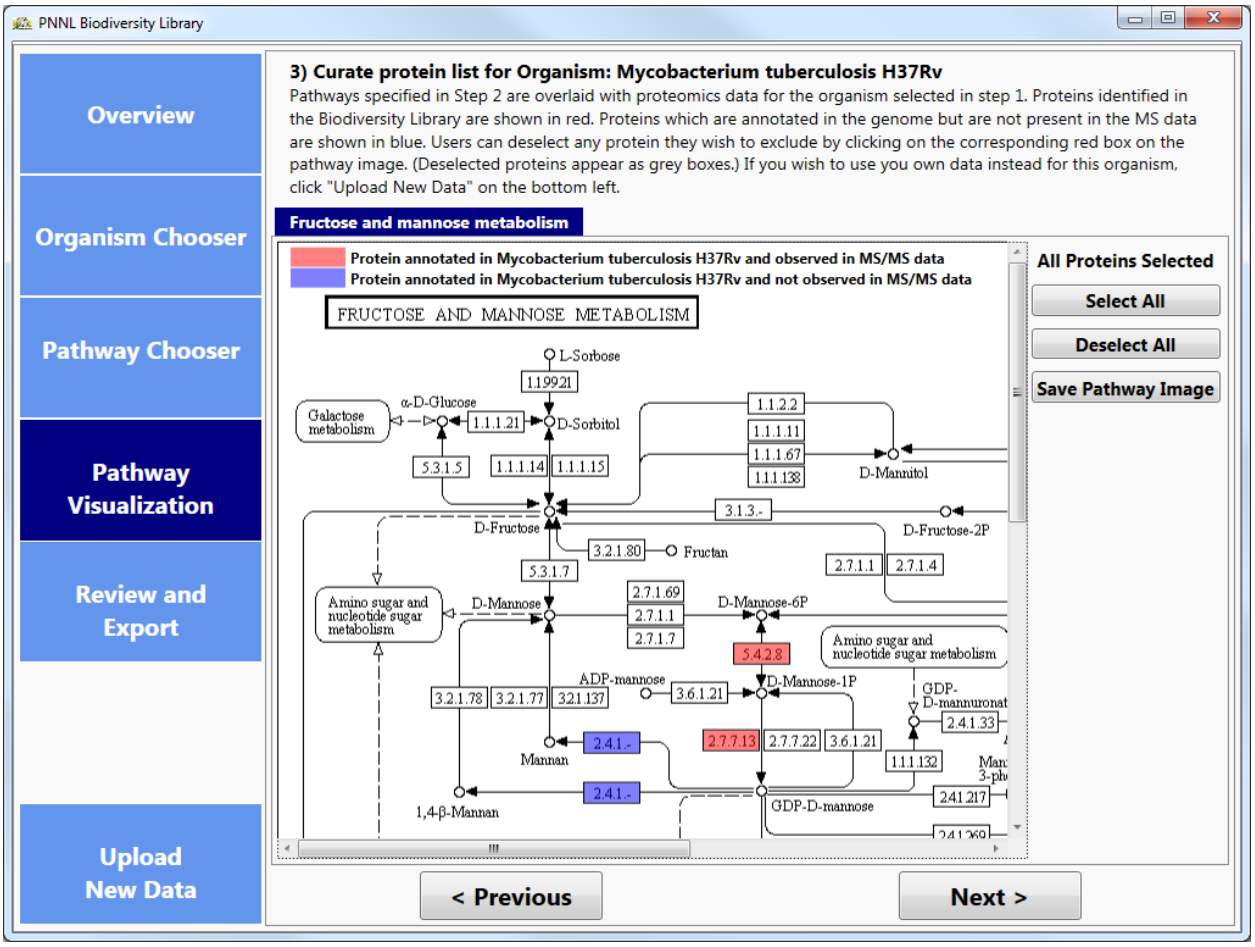
After you have selected an organism, you then choose the KEGG pathways you are interested in investigating. The coverage metric is dynamically calculated for each Organism/Pathway combination and shows the percentage of proteins in a pathway for which the Library contains proteomic data. You can select as many pathways as you desire.



Select proteins of interest through KEGG Pathway maps:

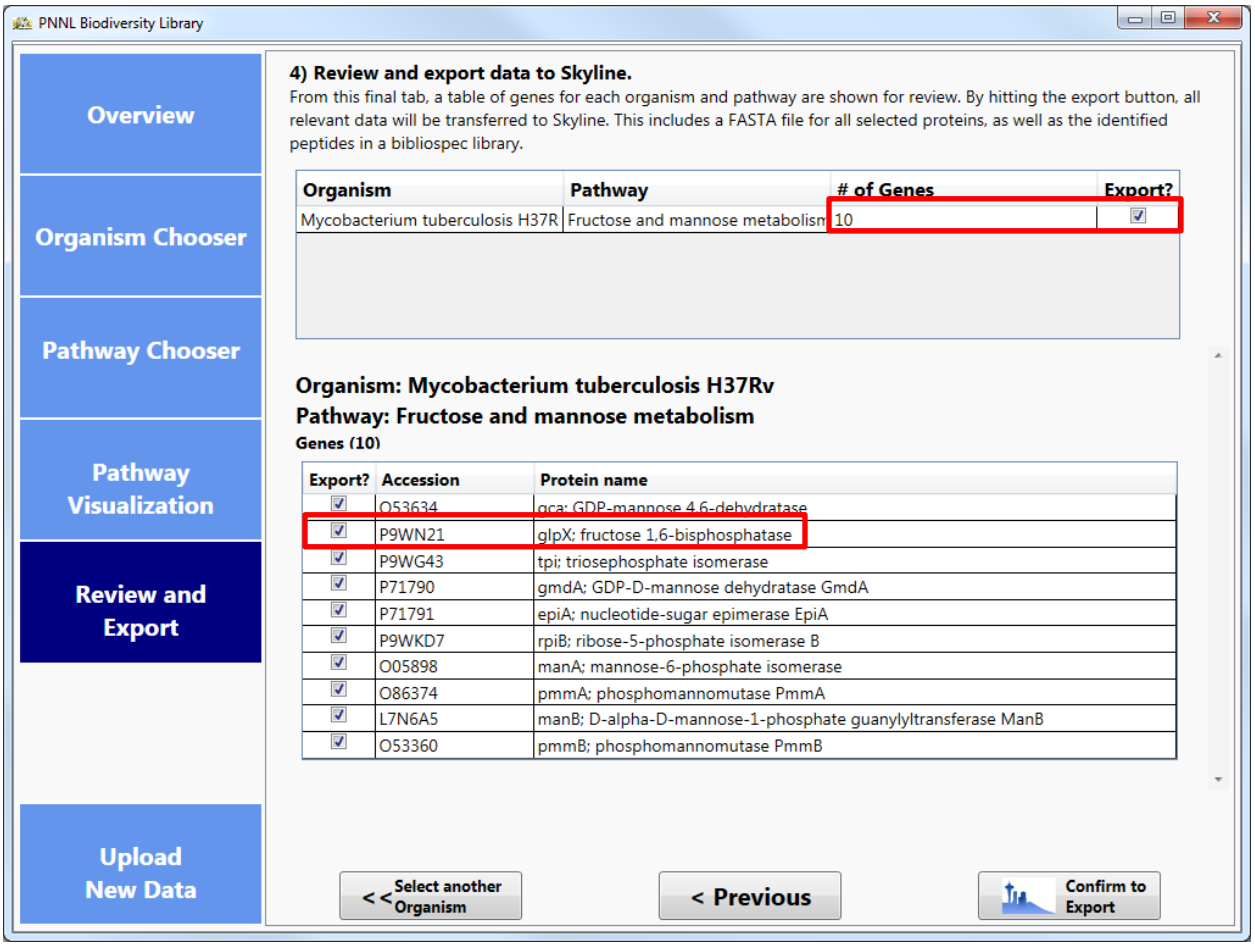
The images for the pathways you selected will be created dynamically, highlighting all the KEGG orthologs for the organism in the pathway. If that protein has been observed in the spectral library, it will be highlighted initially as red; unobserved proteins are highlighted in blue. From here, you can select the proteins which are of interest to you in the pathway in two ways, either by clicking on individual 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.



Review and export:

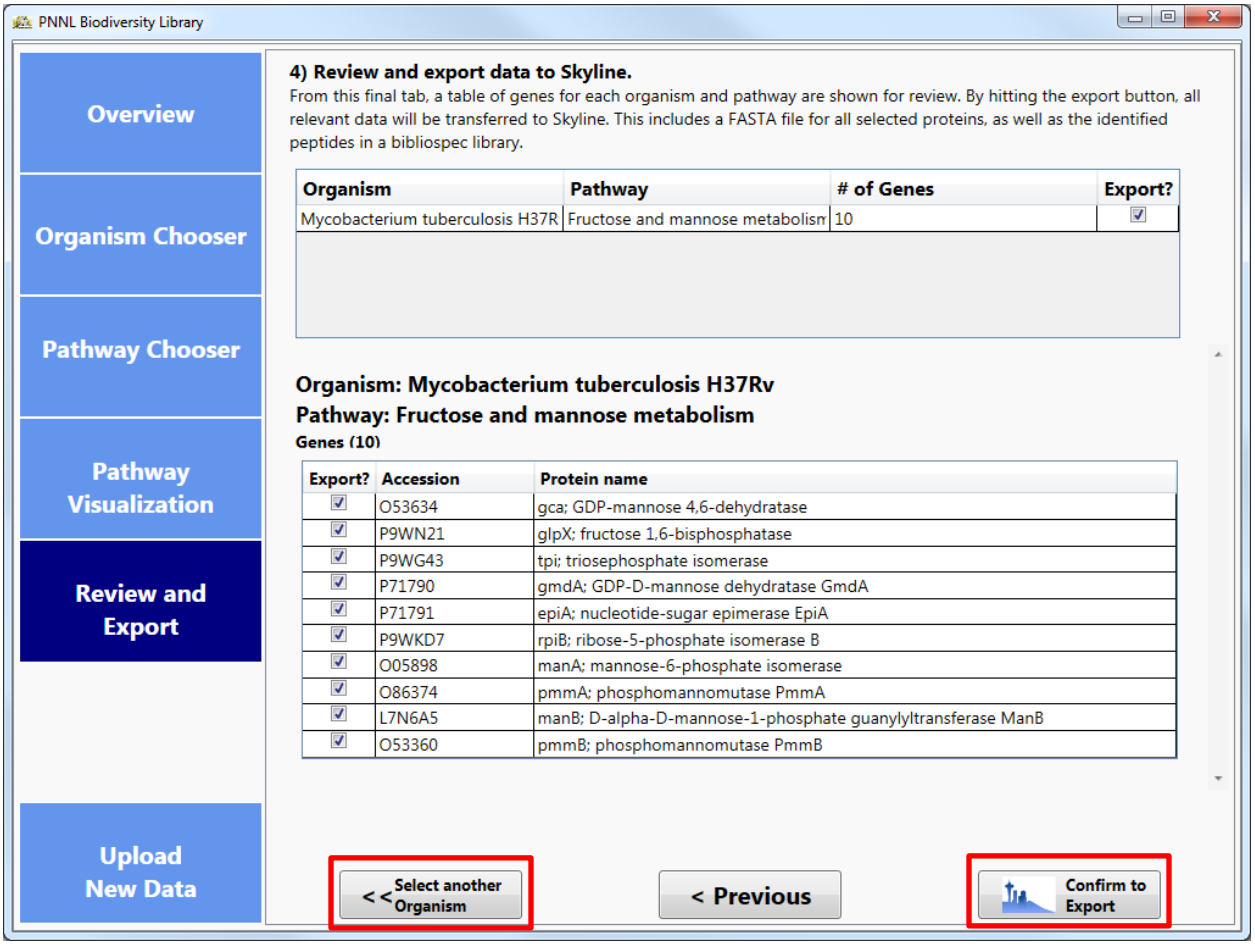
On the Review and Export tab, the application will display the total number of genes selected from all of your Organism/Pathway combinations as well as an individual breakdown of the proteins. If there are Organism/Pathway combinations you are no longer interested in, or individual proteins which you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



Export data to Skyline:

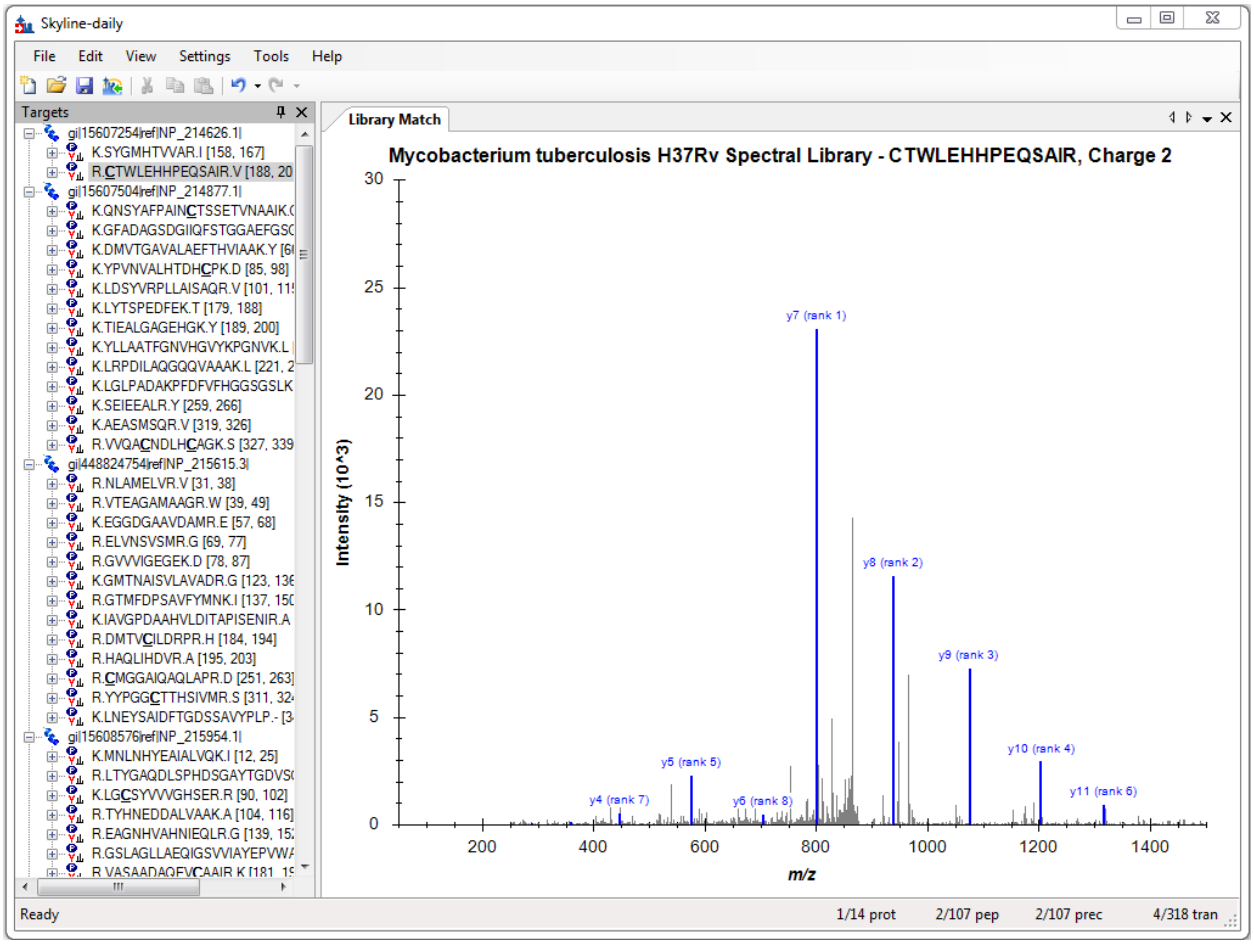
Once you are satisfied with your selection, you can then either select another organism and repeat the wizard steps, or export the data to Skyline.

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/). Both of these files (FASTA and Bibliospec library) are then imported to Skyline.



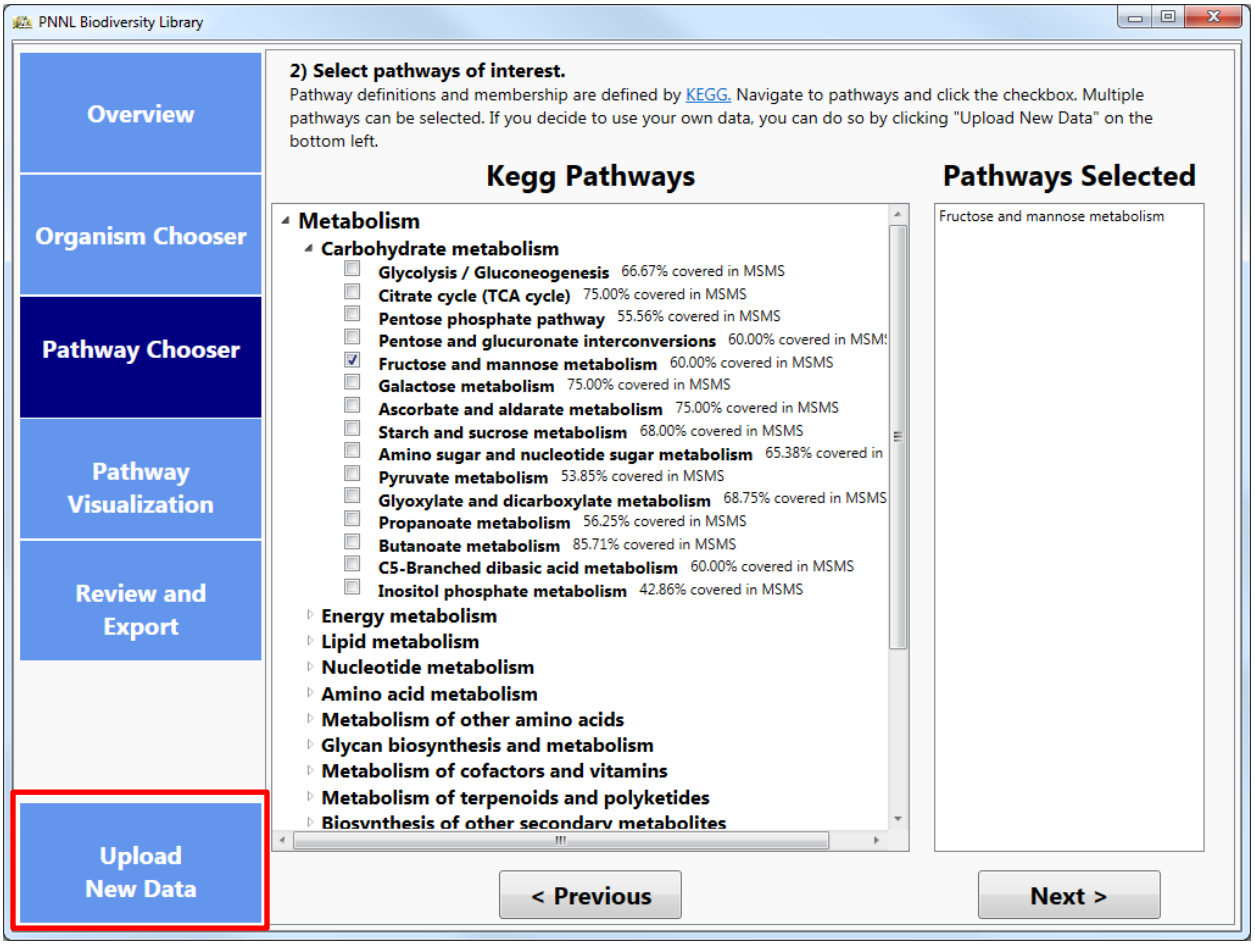
View spectral data in Skyline:

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

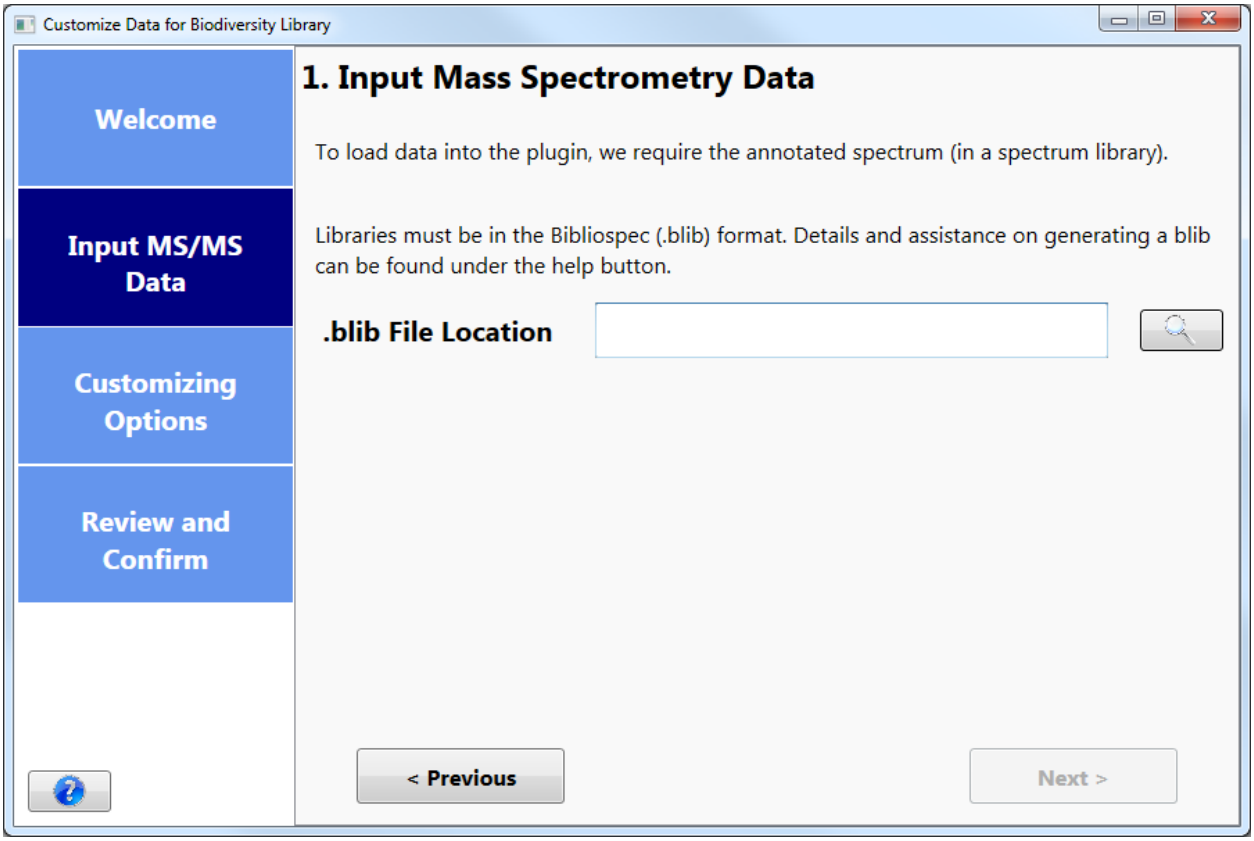


Customizing the plugin with your data:

The Biodiversity Plugin enables users to update the MS/MS data to include personal data and use the same pathway oriented interface to load their data directly into Skyline. Users start customizing their data by selecting the **Upload New Data** button at the bottom left of the window. This starts a new wizard to guide users through the upload process. Users can choose to upload new data at any time.



Input Mass Spec Data:



Only one file is required, a Bibliospec formatted spectral library. Please consult the online documentation for details about building a Bibliospec library. To build a library from mzIdentML results, and example command line is:

~> BlibBuild.exe -c 0.9999 E:\path\to\PSM_results.mzid E:\path\to\Library.blib

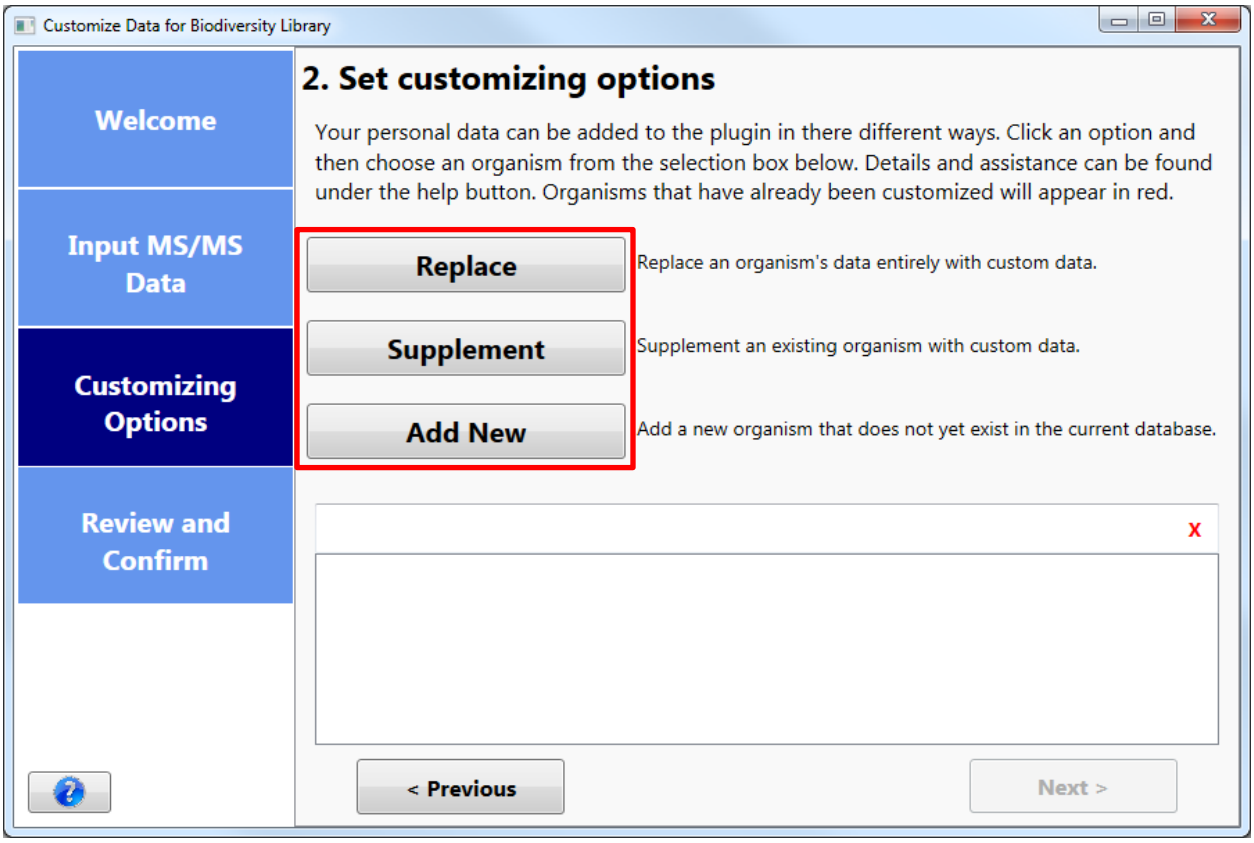
It is then necessary to filter the Blib file to contain ONE and only one PSM per peptide. Therefore, you must filter the above library file, for example:

~> BlibFilter.exe -b 1 E:\path\to\Library.blib E:\path\to\FilteredLibrary.blib

Customizing Options:

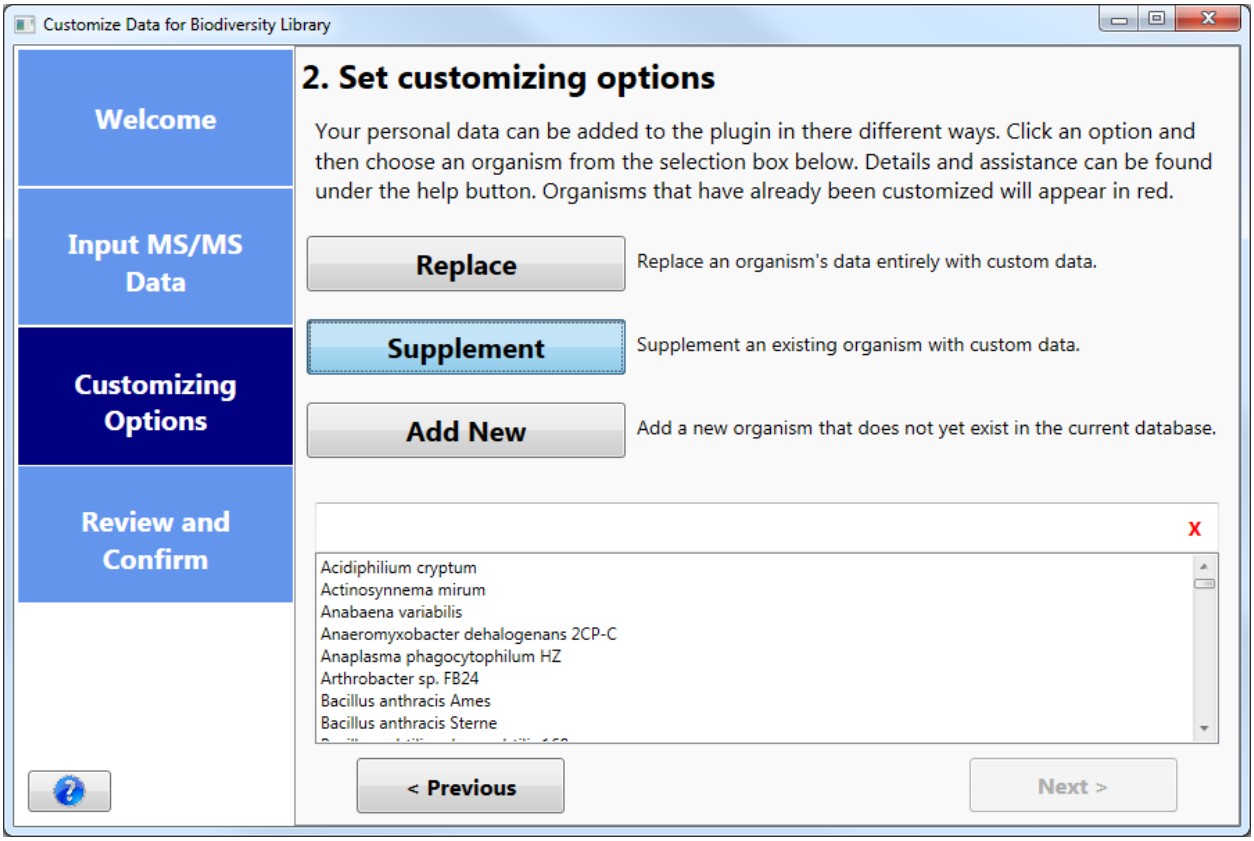
There are three options for updating organism data:

- Replace the existing organism entirely with custom data.
- Supplement an existing organism with additional data.
- Add a new organism that does not yet exist in the current database.



Selecting an Organism:

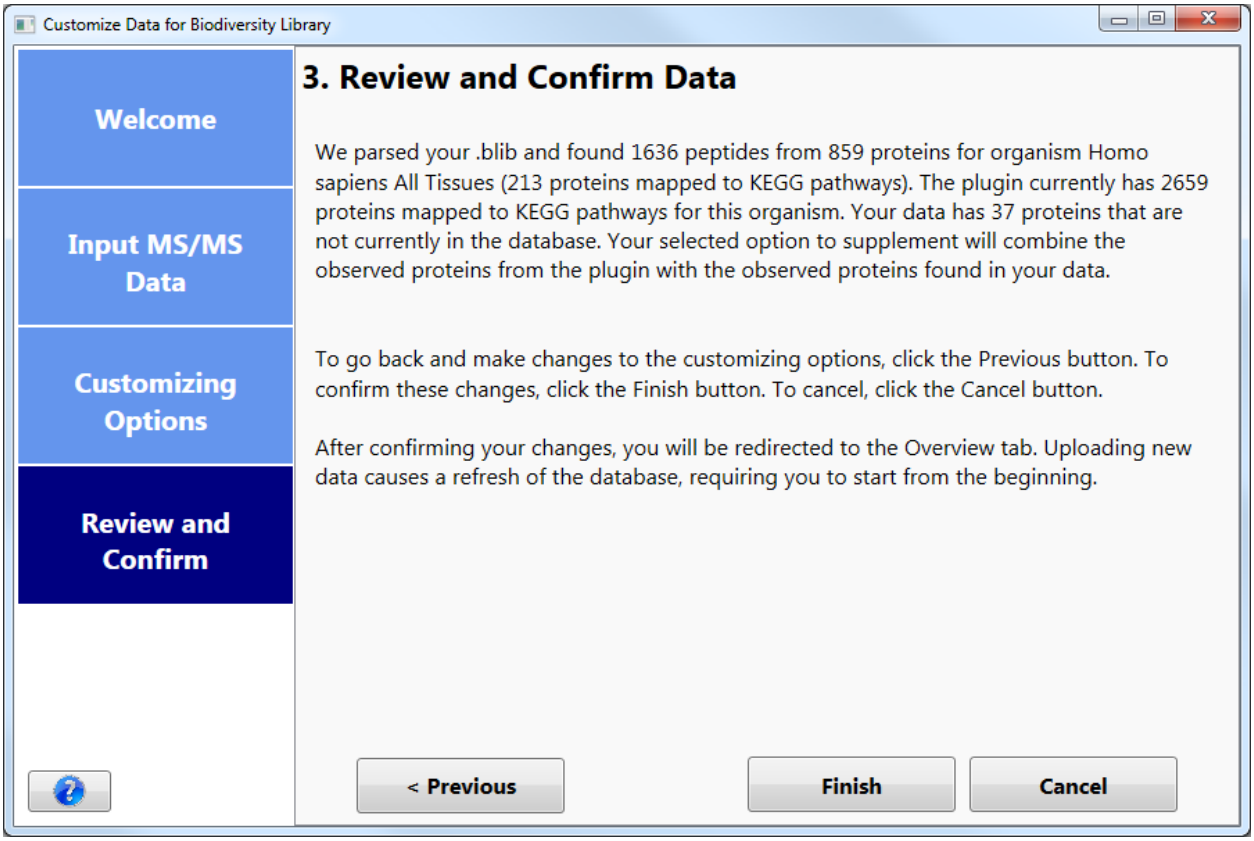
Once the customizing option has been selected, the list box at the bottom of the window will populate with organisms to search and choose from.



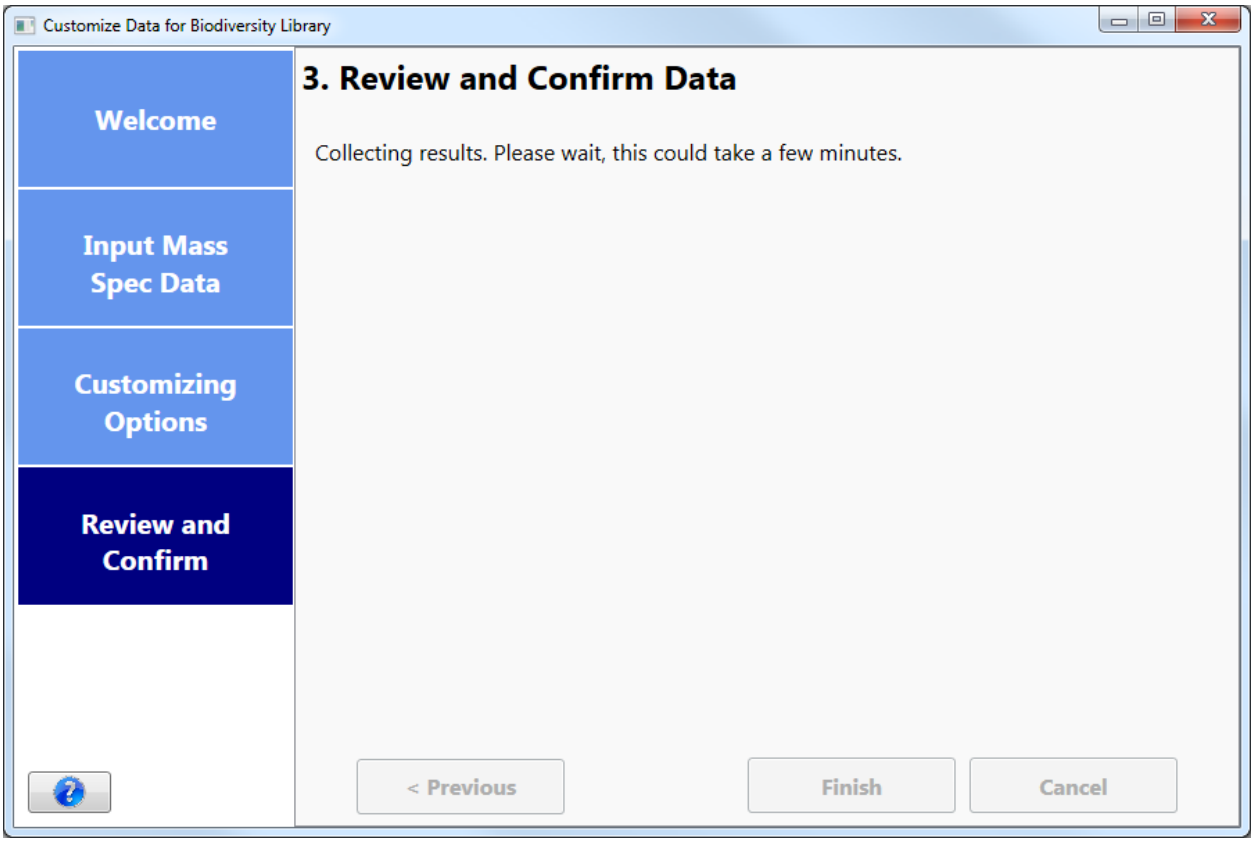
Once you have selected an organism, the next button will enable. Once you click the **Next** button, the wizard will begin collecting the results. Users will be navigated to the Review and Confirm tab where they will wait for their results to appear.

Reviewing and Confirming:

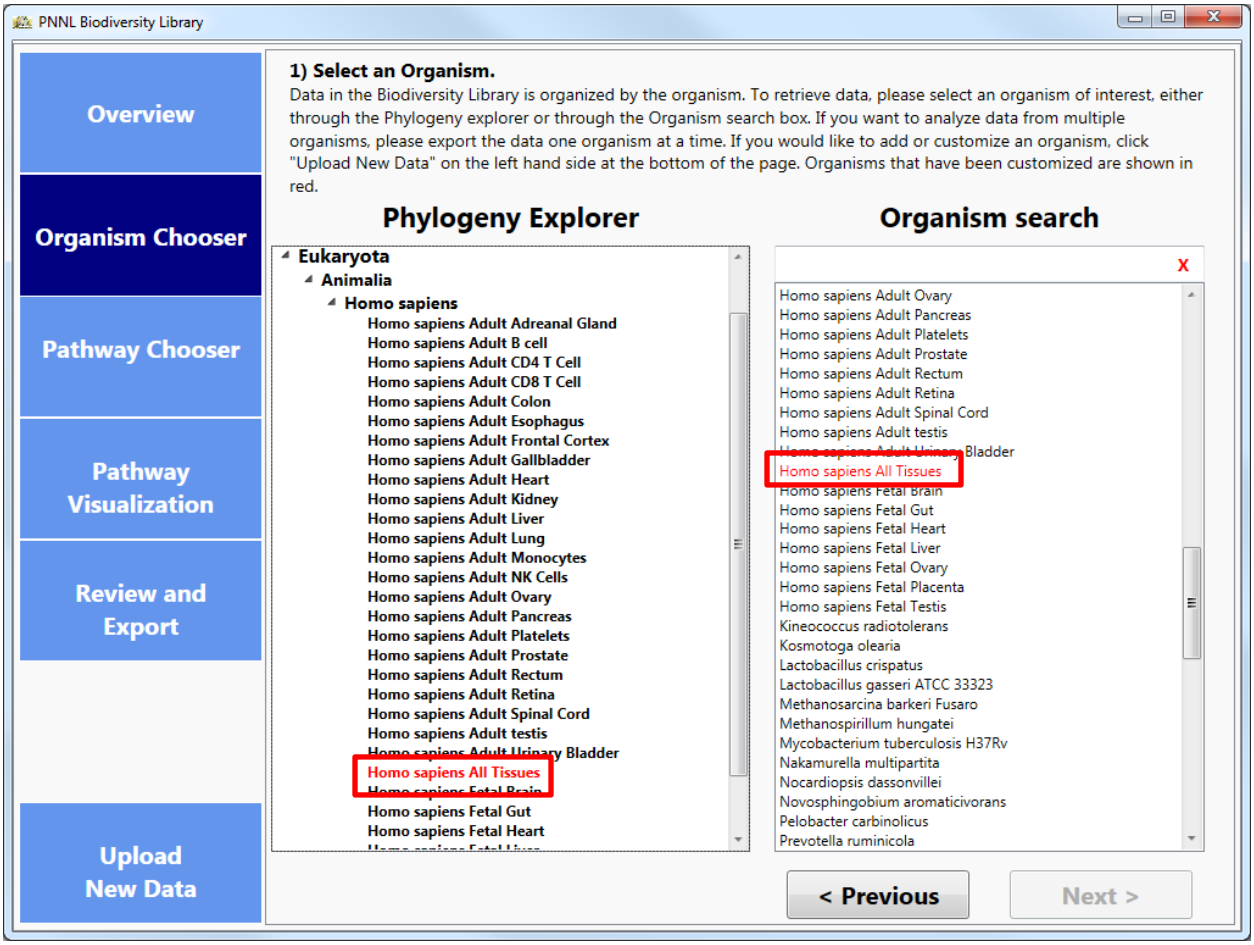
Once the results have been collected, a message will appear prompting you to confirm the results and finish the process by clicking the **Finish** button, cancel the process by clicking the **Cancel** button, or go back and modify your customizing options by clicking the **Previous** button.



Please note: Adding a new organism can take longer than Replacing or Supplementing. If the process takes longer than expected, a message (shown below) will appear, allowing you to wait there until the results are collected.

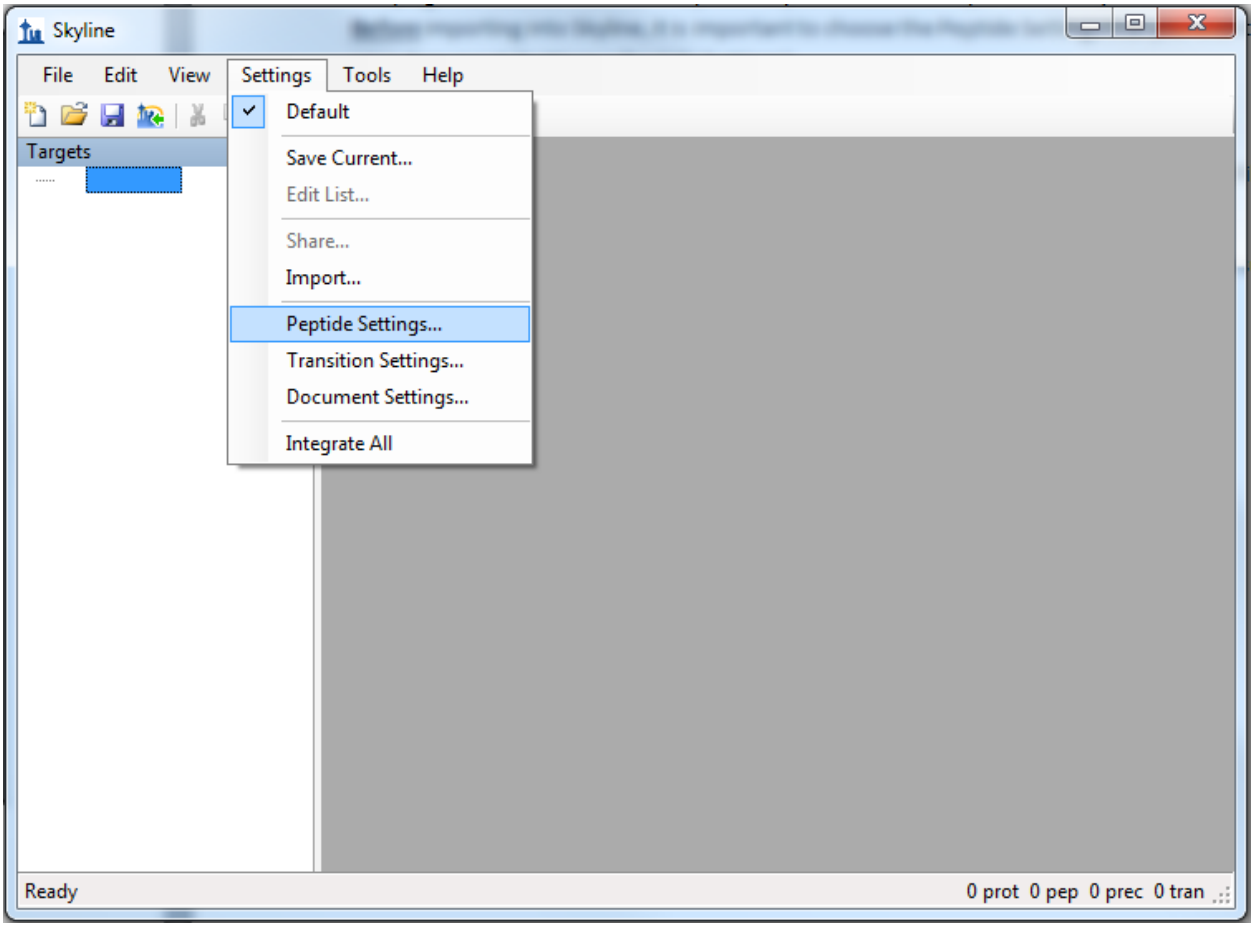


For your convenience, any organism that was customized by you will appear in red. Unmodified organisms will remain black.



Settings in Skyline

The plug-in creates a FASTA of the proteins you choose and imports that sequence file into Skyline. **Before** importing into Skyline, it is important to choose the Peptide Settings that you wish to use for your document (Settings > Peptide Settings).



After the FASTA has been imported, users will not be able to make changes in Digestion, Filter or Modifications tabs.

Additionally, it is recommended to set Transition Settings before importing the FASTA. Be sure to select precursor charge states (in the Filter Tab).

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
            - To add custom observed data, see the *Customizing the plugin with your data* section of the tutorial
        - 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@pnnl.gov.
- Observed protein count is zero when customizing data.
    - This could be due to two possibilities:
        - The organism you selected to update doesn't match the organism that was used in your results
        - The identifiers in your data are not Uniprot identifiers (e.g. RefSeq, Genbank, GI numbers)
        - The organism you selected does not use Uniprot as identifiers in KEGG