
Using Galaxy-P Documentation

Release 0.1

John Chilton, Pratik Jagtap

October 26, 2015

1	Introduction	1
2	Galaxy-P 101 - Building Up and Using a Proteomics Workflow	3
2.1	What Are We Trying to Do?	3
2.2	Setting Up Your Galaxy-P Account	3
2.3	Organizing Your Windows	4
2.4	Loading a Search Database	4
2.5	Using Protein Database Downloader	5
2.6	Using the Database Merge Tool	6
2.7	Creating a Decoy Database	7
2.8	Peaklist Processing	10
2.9	X! Tandem Search	12
2.10	Peptide Prophet Processing of X! Tandem Search	15
2.11	Converting ProtXML to a Table	18
2.12	False Discovery Rate Analysis	20
2.13	Understanding Galaxy Histories	26
2.14	Converting Histories to Workflows	28
2.15	Applying Workflows to Your Data	39
3	FTP	51
3.1	Uploading Files	51
3.2	See Also	55
4	Multiple File Datasets	57
4.1	Introduction	57
4.2	Creating a Multiple File Dataset	57
5	A Simple Workflow using peptide-shaker	61
6	Indices and tables	69

Introduction

This guide covers intermediate and advanced usage of the Galaxy-P platform for proteomic and mass spec data analysis. For an gentle introduction to Galaxy, please check out the [Galaxy 101](#) maintained at Penn State.

Galaxy-P 101 - Building Up and Using a Proteomics Workflow

This section is heavily inspired by the [Galaxy 101](http://wiki.galaxyproject.org/GalaxyTeam) maintained at Penn State by the *core Galaxy team* <<http://wiki.galaxyproject.org/GalaxyTeam>>.

In this walkthrough we will introduce you to basics of Galaxy-P:

- Uploading database using Protein Database Downloader.
- Getting raw data from an external source.
- Peaklist processing including options and parameters.
- X! tandem search.
- PeptideProphet processing of X!tandem search.
- Converting PepXML to a table.
- Using ProteinProphet to process PeptideProphet results.
- FDR analysis.
- Performing simple data manipulation.
- Understanding Galaxy's history system.
- Creating and editing workflows.
- Applying workflows to your data.

2.1 What Are We Trying to Do?

We are trying to set up a search of one RAW file (your experimental data acquired on an LTQ/Orbitrap instrument) against a database using a search algorithm named X!tandem. The idea is to estimate number of valid matches by using PeptideProphet or ProteinProphet and FDR analysis. Once we have created this workflow we can apply this workflow to another dataset and get results. So let's try it...

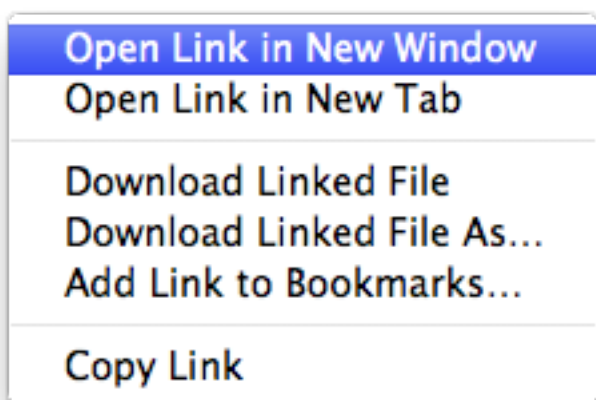
2.2 Setting Up Your Galaxy-P Account

Go to the User menu at the top of Galaxy-P interface and choose Register:

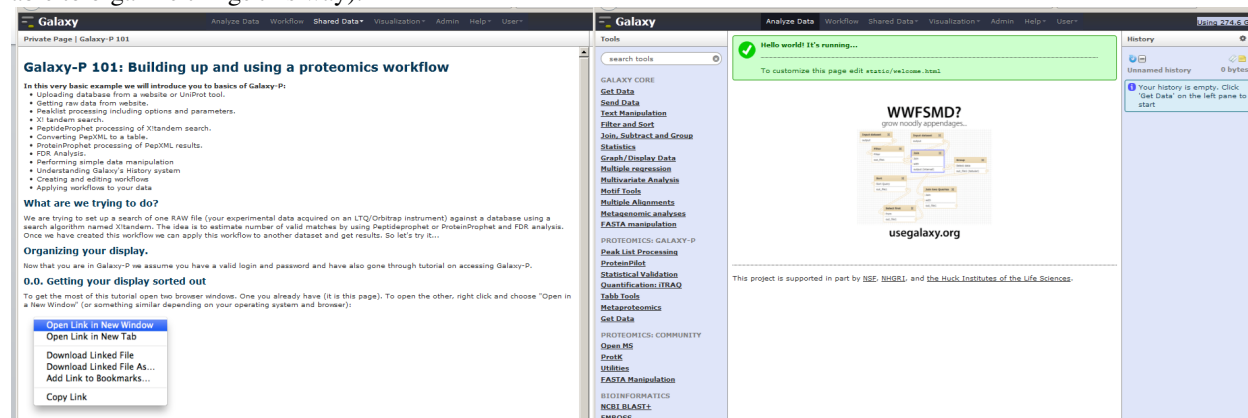
Follow the instructions as prompted.

2.3 Organizing Your Windows

To get the most of this tutorial open two browser windows. One you already have (it is this page). To open the other, right click and choose “Open in a New Window” (or something similar depending on your operating system and browser):



Then organize your windows as something like this (depending on the size of your monitor you may or may not be able to organize things this way):



2.4 Loading a Search Database

There are many options for how you can upload your search database (FASTA file with protein sequences). Three among these are:

- Protein Database Downloader.
- Use website link for the database (see this short tutorial).
- Upload database from the data library.

In this tutorial, we will explore using Protein Database Downloader for database search.

2.5 Using Protein Database Downloader

Now we get our search database. First thing we will do is to go to Tools → Proteomics: Galaxy-P → Get Data. Then click on Protein Database Downloader.

Get Data

- [PRIDE BioMart](#) Download proteomic data from the EBI Pride BioMart.
- [Protein Database Downloader](#)

Since we are going to analyze a human dataset, select human database along with isoforms.

Protein Database Downloader (version 0.1.0)

Download from:

UniProtKB

Taxonomy:

Homo sapiens (Human)

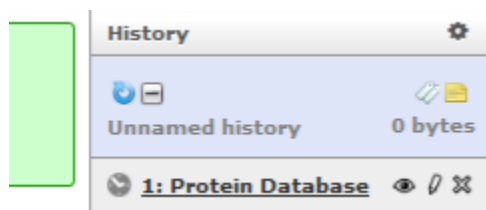
Include isoform data:



Execute

Clicking on “Execute” will download the human canonical isoform database.

After this you will see your first History Item in Galaxy’s right pane. It will go through gray (preparing) and yellow (running) states to become green:



For contaminants database, we will use the Protein database Downloader, select cRAP (contaminants) database and “Execute” much the same way you did for the human UniProt database.

Protein Database Downloader (version 0.1.0)

Download from:







cRAP (contaminants)

Execute

Output

A FASTA file containing the specified protein sequences.

After this you will see your second History Item in Galaxy’s right pane.

Unnamed history	67.1 MB
<u>2: Protein Database</u>	  
<u>1: Protein Database</u>	  

2.6 Using the Database Merge Tool

In order to merge these two FASTA files in your history, go to Tools → FASTA Manipulation → Merge FASTA Databases.

FASTA Manipulation

- Create decoy databases
Creates a random protein database with similar properties to a real protein database


Click on “Add new Input FASTA File(s)” – twice – so that your set up for merging files looks like this:

Merge FASTA Databases (version 0.1)

Input FASTA File(s)

Input FASTA File(s) 1


FASTA File:

1: Protein Database 

Remove Input FASTA File(s) 1

Input FASTA File(s) 2

FASTA File:
















2: Protein Database 

Remove Input FASTA File(s) 2

Add new Input FASTA File(s)

Execute

Click on Execute and you will be able to see your third History Item in Galaxy’s right pane.

History 	
  Unnamed history 67.1 MB  	
 <u>3: Merged FASTA from data 1 and data 2</u>	  
<u>2: Protein Database</u>	  
<u>1: Protein Database</u>	  

2.7 Creating a Decoy Database

Our next step is to create a decoy database out of the merged file (3rd history item on the list).

For this, go to Tools -> FASTA Manipulation -> 'Create Decoy Database'.

- **Create Decoy Database (reverse)**
Creates a decoy search database by adding reverse sequences to an existing database.

Ensure that history item 3 shows up in the "FASTA Input:" box and that the box for "Include original entries in output database:" is checked. Change Decoy prefix to **REV_**.

Your parameters for creating decoy database (reverse) tool should look like this.

Create Decoy Database (reverse) (version 0.1.0)

FASTA Input:
3: Merged FASTA from.. and data 1

Include original entries in output database:
☒

Specify advanced decoy frequency options?:
No, just create database with one reversed sequence for every input sequence

Decoy Prefix:
REV_

Execute

Click on "Execute" to generate the fourth item in your history list. This is the FASTA database that we will be using for our search.

4: Create Decoy Database (reverse) on data 3

3: Merged FASTA from data 2 and data 1

2: ftp://ftp.thegpm.org/ftp/cRAP/crap.fasta

1: http://www.uniprot.org/uniprot/?query=organism%3a9606+k eyword%3a1185&force=yes&format=fasta&include=yes

Now we will rename the history items to "Human UniProt", "cRAP", "Merged Human UniProt cRAP" and "Target_Decoys_Human_Contaminants" by clicking on the Pencil icon adjacent to each item. Also we will rename history to "Galaxy-P 101" (or whatever you want) by clicking on "Unnamed history" so everything looks like this:

Galaxy-P 101		158.4 MB
4:	<u>Target Decoy Human Contaminants) on data 3</u>	👁️ ✎️ ✕
3:	<u>Merged Human UniProt cRAP</u>	👁️ ✎️ ✕
2:	<u>cRAP</u>	👁️ ✎️ ✕
1:	<u>Human UniProt</u>	👁️ ✎️ ✕

Please feel free to explore tabs – Convert Format, Datatype, Permissions while you are editing the attributes. This is especially important while troubleshooting for steps that fail wherein a datatype has not been set properly or needs to be changed for subsequent steps.

Attributes Convert Format Datatype Permissions

Edit Attributes

Name:

Target_Decoy_Human_Contaminants

Info:

Annotation / Notes:

None

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build:

Click to Search or Select

Save

Auto-detect

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

The next step is to upload RAW file.

There are a few options on how you can upload your spectral data (RAW file acquired on a LTQ/Orbitrap) * Upload from your computer using your web browser. * Upload from your computer using the Galaxy-P FTP server. * Use website link for the RAW file (something we will use for this tutorial). * Import data from the data library.

While there are many ways of uploading a RAW file, we will use a weblink to upload a fractionated human salivary RAW file.

In order to upload a RAW file, you can go to “Upload File” in Tools section and type in [https://netfiles.umn.edu/users/pjagtap/Galaxy-P 101/Raw101.RAW](https://netfiles.umn.edu/users/pjagtap/Galaxy-P%20101/Raw101.RAW) and “Execute”.

Upload File (version 1.1.3)

File Format:

Auto-detect

Which format? See help below

File:

 Browse...

TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large file

URL/Text:

Here you may specify a list of URLs (one per line) or paste the contents of a file.

Files uploaded via FTP:

File	Size
------	------

*Your FTP upload directory contains no files.*This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at **lo**

Convert spaces to tabs:

☐ Yes

Use this option if you are entering intervals by hand.

Genome:

Execute

This will give us our 5th history item :

Galaxy-P 101	158.4 MB
5: https://netfiles.umn.edu/users/pjagtap/Galaxy-P 101/Raw101.RAW	
4: Target Decoy Human Contaminants) on data 3	
3: Merged Human UniProt cRAP	
2: cRAP	
1: Human UniProt	

We can change the name of the RAW file to Raw101.RAW by using the pencil icon button.

Attributes Convert Format Datatype Permissions

Edit Attributes

Name:

Info:

Annotation / Notes:

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

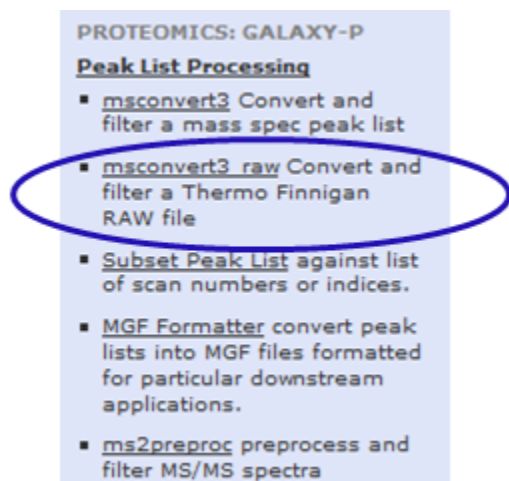
Database/Build:

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

2.8 Peaklist Processing

In the next step we will process the Raw file into a peaklist in mzml format so that it can be searched using X!tandem.

For this go to Tools → Proteomics: Galaxy-P → Peak List Processing → msconvert3 raw



Explore the features by clicking on Use Filtering. However, for this tutorial we are going to use default settings for msconvert.

Click on “Execute” by ensuring that the settings appear as shown below:

msconvert3_raw (version 0.1.0)

Thermo Finnigan RAW Input:
5: Raw101.RAW

Output Type:
mzML (indexed)

Use Filtering?:
☐

Advanced Settings:
Use Defaults

Execute

We have our sixth history item in the history list now.

History			
 Galaxy-P 101		 267.5 MB	
 6: msconvert3_raw on data 5			
5: Raw101.RAW			
4: Target Decoy Human Contaminants) on data 3			
3: Merged Human UniProt cRAP			
2: cRAP			
1: Human UniProt			

Change the name of the mzml file to “Peaklist Raw101”. Note that this peaklist is in mzml format.

Attributes Convert Format Datatype Permissions

Edit Attributes

Name:

Info:

msconvert --mzML -o
output input.RAW -c filters
fixme:heap:HeapSetInforma...

Annotation / Notes:

None

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build:

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

2.9 X! Tandem Search

For setting up a search of peaklist (Peaklist Raw101) against FASTA database (Target_Decoys_Human_Contaminants) using X!tandem, go to Tools→ ProtK (under PROTEOMICS: COMMUNITY) → X!Tandem MSMS Search.

ProtK

- [InterProphet](#) Combine Peptide Prophet results from multiple search engines
- [OMSSA MSMS Search](#) Run an OMSSA MS/MS Search
- [Peptide Prophet](#) Calculate Peptide Prophet statistics on search results
- [Protein Prophet](#) Calculate Protein Prophet statistics on search results
- [X!Tandem MSMS Search](#) Run an X!Tandem Search

Set up search parameters for X!tandem as shown below. Ensure that you have the right FASTA file (4: Tar-

get_Decoys_Human_Contaminants), MSMS File (6: Peaklist Raw101) and click on Oxidation M for variable modifications. Ensure that you have 2 Missed Cleavages allowed, Trypsin as an enzyme, Fragment ion tolerance at 0.5 Da and Precursor ion tolerance at 10.0 ppm. The parameters file should look like below:

X!Tandem MSMS Search (version 1.0.0)

Database source:

Your Upload File

Uploaded FASTA file:

4: Target_Decoy_Huma...s on data 3

MSMS File:

6: Peaklist Raw101

An mzML file with MS/MS data

Variable Modifications:

Carbamidomethyl C
Glycocapture-N
Oxidation M

Hold the appropriate key while clicking to select multiple items

Custom Variable Modifications

You can specify a modification when present in a motif. For instance, 0.998@N!{P}[ST] is a deamide

Add new Custom Variable Modifications

Fixed Modifications:

Carbamidomethyl C
Glycocapture-N
Oxidation M

Hold the appropriate key while clicking to select multiple items

Custom Fixed Modifications

You can specify a modification when present in a motif. For instance, 0.998@N!{P}[ST] is a deamide

Add new Custom Fixed Modifications

Missed Cleavages Allowed:

2

Allow peptides to contain up to this many missed enzyme cleavage sites

Enzyme:

Trypsin

Fragment ion tolerance:

0.5

Fragment Ion Tolerance in Daltons

Precursor ion tolerance:

10.0

Precursor Ion Tolerance (Da or ppm)

Precursor Ion Tolerance Units:

ppm

Enable semi-cleavage rules:

☒

Match peptides where one end or the other does not conform to specified enzyme rule.

Allow multi-isotope search:

☐

This allows peptide candidates in windows around -1 Da and -2 Da from the acquired mass to be considered if not corrected to the monoisotopic mass.

Use K-Score Algorithm:

☒

The k-score algorithm is a Tandem score plug-in implementing the COMET score function as described in [1].

Thresholds:

ISB k-Score defaults

Set various X! Tandem threshold values.

Your history list should look like this now.

History	
Galaxy-P 101	419.9 MB
7: X!Tandem vs Target Decoy Human Contaminants.Peaklist	👁️ 0 ✕
Raw101.Peaklist	
Raw101.pepXML	
6: Peaklist Raw101	👁️ 0 ✕
5: Raw101.RAW	👁️ 0 ✕
4: Target Decoy Human Contaminants	👁️ 0 ✕
3: Merged Human UniProt cRAP	👁️ 0 ✕
2: cRAP	👁️ 0 ✕
1: Human UniProt	👁️ 0 ✕

2.10 Peptide Prophet Processing of X! Tandem Search

In this step we will process PepXML results from X! Tandem search to provide peptide probability scores for further analysis.

For this, go to Tools→ ProtK (under PROTEOMICS: COMMUNITY) → ‘Peptide Prophet’.


ProtK

- [InterProphet](#) Combine Peptide Prophet results from multiple search engines
- [OMSSA MSMS Search](#) Run an OMSSA MS/MS Search
- [Peptide Prophet](#) Calculate Peptide Prophet statistics on search results
- [Protein Prophet](#) Calculate Protein Prophet statistics on search results
- [X!Tandem MSMS Search](#) Run an X!Tandem Search

The Peptide Prophet parameters should be specified as follows:

Peptide Prophet (version 1.1.0)

Raw Search Results:

7: X!tandem vs Targe..w101.pepXML 

These files will typically be outputs from omssa or xtandem search tools

Expect true positives to have a glyco capture motif:

☐

Use icat information:

☐

Use phospho information:

☐

Use pI information:

☐

Use hydrophobicity / RT information:

☐

Use accurate mass binning:

☐

Don't use NTT model:

☐

Don't use NMC model:

☐

Use Gamma distribution to model the negatives:

☐

Applies only to X!Tandem results

Only use Expect Score as the discriminant:

☐

Applies only to X!Tandem results. Helpful for data with homologous top hits e.g. phospho or glyco

Force fitting:

☐

Bypasses automatic mixture model checks and forces fitting of a mixture model

Allow multiple instrument types:

☐

Warning instead of exit with error if instrument types between runs is different

Maldi data:

☐

Decoy Prefix:

decoy

Identifier prefix of decoy entries in search database.

Manually specify database:

☐

This may be needed for pepXML files generated outside of Galaxy or ProtK, in most cases you will not need to set this.


Execute

Change the name of the Peptide Prophet file to “peptide_prophet Raw101.pep.xml” by using the pencil icon.

History	
 Galaxy-P 101  420.9 MB	
8: <u>peptide_prophet</u> <u>Raw101.pep.xml</u>	  
7: <u>X!Tandem vs Target Decoy H</u> <u>uman Contaminants.Peaklist</u> <u>Raw101.Peaklist</u> <u>Raw101.pepXML</u>	  
6: <u>Peaklist Raw101</u>	  
5: <u>Raw101.RAW</u>	  
4: <u>Target Decoy Human Contami</u> <u>nants</u>	  
3: <u>Merged Human</u> <u>UniProt cRAP</u>	  
2: <u>cRAP</u>	  
1: <u>Human UniProt</u>	  

In this step, we will use ProteinProphet to process PeptideProphet results from X!tandem search to provide protein probability scores for further analysis.

For this, go to Tools → ProtK (under PROTEOMICS: COMMUNITY) → Protein Prophet.

Protein Prophet (version 1.0.0)**Peptide Prophet Results:** 

These files will typically be outputs from peptide prophet or interprophet

Inputs are from iProphet:☐**Don't apply Occam's razor:**☐

When selected no attempt will be made to derive the simplest protein list explaining observed peptides

Use group weights:☐

Check peptide's total weight (rather than actual weight) in the Protein Group against the threshold

Normalize NSP using Protein Length:☐**Use the log of probability in the confidence calculations:**☐**Use the EM to compute probability given the confident:**☐**Consider all possible peptides in the database in the confidence model:**☐**Report results for unmapped proteins:**☐**Use Expected Number of Ion Instances to adjust the peptide probabilities prior to NSP adjustment:**☐**Do NOT use peptide degeneracy information when assessing proteins:**☐**Minimum peptide prophet probability for peptides to be considered:****Minimum percentage of independent peptides required for a protein:****Execute**

This will generate 9th history item in the list.

2.11 Converting ProtXML to a Table

In this step we will convert Protein Prophet results to a tabular format so that they can be viewed or processed for further analysis.

For this, go to Tools -> Utilities (under PROTEOMICS: COMMUNITY) -> Convert ProtXML to Tabular. Ensure that the latest file in the history (protXML file – 9th in the list).

Convert ProtXML to Tabular (version 0.1.0)

Prot XML Input:

9: protein_prophet.p..xml.protXML

Execute

Clicking on Execute will give us our 10th file in our history.

Galaxy-P 101 1.0 GB

10: Convert ProtXML to
Tabular on data 9

9: protein_prophet.peptide_prophet Raw101.pep.xml.protXML
prot XML Search Results
format: protxml, database: 2

Explore the icons in the generate the file to see ('eye') the data OR to download ('floppy disk'), view details ('i' sign) OR rerun the analysis with changed parameters (blue 'rerun' icon). You can also create graphs and visualize using the 'graph' icon.

To 'see the data in main window panel.

To change name, attributes, datatype of the file.

To delete the file.

To download the file.

To view details.

To rerun the process.

To create graph and visualize.

10: Convert ProtXML to
XLS on data 9

388 lines, 245 comments
format: protxml.csv, database: 2

1. Entry Number	2. Group	Probability	3
entry no.	group	probability	pept:
peptide	sequence		
1	1.0000	ap LYSC_HUMAN ,tr F8V	
et-fastadb.cgi?Ref=tr F8VV32 F8VV32_1			
t/tpp/cgi-bin/peptidexml_html2.pl?Pe			

2.12 False Discovery Rate Analysis

In order to calculate FDR at the peptide level, we will first convert PeptideProphet file to a tabular format.

For this, go to Tools → ProtK (under PROTEOMICS: COMMUNITY) → Convert PepXML to Table. Ensure that the peptide prophet file in the history (peptide prophet file – 8th in the list) is highlighted.

PepXML to Table (version 1.0.1)

Input File:

8: peptide_prophet Raw101.pep.xml

A pepXML file

Execute

Convert a pepXML file to Tab delimited text

Change the name of this 11th file in history list to add “Table...” as shown in the image below.

Attributes

Convert Format

Datatype

Permissions

Edit Attributes

Name:

Table peptide_prophet Raw101.

Info:

Annotation / Notes:

None

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build:

Click to Search or Select

Number of comment lines:

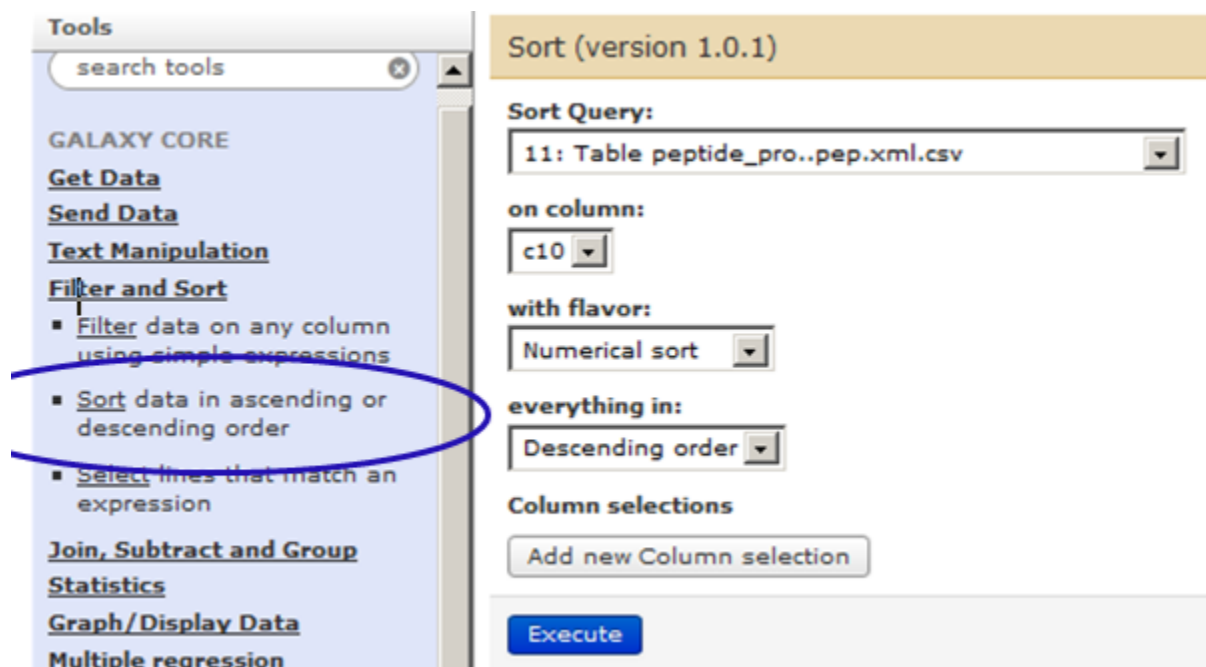
Save

Auto-detect

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

Next step is to sort the table with descending peptide probability scores. For this go to Tools → Filter and Sort → Sort. Ensure that table (csv file – 11th in the list) is highlighted.

Also select c10 (10th column which has peptide probability values) for sorting in descending order.



Click on 'Execute'. This will create you 12th list in history.

To compute FDR on this file, go to Tools → Statistical validation (under PROTEOMICS: GALAXY-P) → Compute False discovery Rate (FDR). Ensure that the sorted table (tsv file – 12th in the list) is highlighted. Also select c1 (1st column which has identifiers) for sorting in descending order. The parameters for this processing should look as below:

galaxy.msi.umn.edu/#

Galaxy

Tools

- FASTA manipulation**
- PROTEOMICS: GALAXY-P
- Peak List Processing**
- ProteinPilot
- Statistical Validation**
 - Mayu Select Select identified PSMs at specified FDR (must have been searched with decoy database).
 - Compute False Discovery Rate (FDR) and append it to tabular data.**
- Quantification: iTRAQ
- Tabb Tools
- Metaproteomics
- Get Data
- PROTEOMICS: COMMUNITY
- Open MS
- ProtK
 - Create Decoy Database (random) Creates a random protein database with similar properties to

Compute False Discovery Rate (FDR) (version 0.1)

Input Table:
12: Sort on data 11

How are decoys specified?:
By identifier prefix

Decoy Prefix:
REV_

Column containing identifiers.:
c1

Specify score column? (Not yet implemented!):
☐ Allows more precise handling of tied hits.

FDR Type:
Global (conservative)

Scaling:
2.0
Scaling factor used for FDR calculations, for a balanced database:

Execute

This will give you an output with a column with false discovery rate. You can use the 'eye' icon to visualize your data.

You can use the text manipulation tools such as 'Add column' and 'cut' in order to visualize your data better.

Go to Tools → Text manipulation → Add column. Ensure that the correct file (number 13 in the list) is chosen as a dataset. Change Iterate column to yes and click on execute.

Add column (version 1.0.0)

Add this value:
1

to Dataset:
13: Sort on data 11 with FDR
Dataset missing? See TIP below

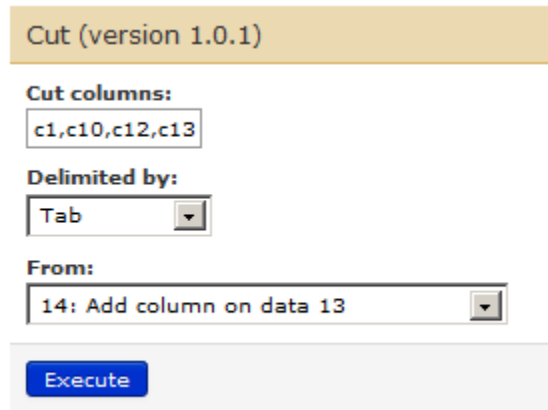
Iterate?:
YES

Execute

This will give you your 14th history item. Use the eye icon to confirm that a column has been added at the end of the file.

Next, we will cut columns 1, 10 12 and 13 from history dataset number 14. For this go to Tools → Text manipulation → Cut. Ensure that the correct file (number 14 in the list) is chosen as a dataset. Type in c1,c10,c12,c13 in 'cut

columns' box and press execute.



The screenshot shows the 'Cut' tool interface in Galaxy-P. At the top, a yellow header bar contains the text 'Cut (version 1.0.1)'. Below this, the 'Cut columns:' label is followed by a text input field containing 'c1,c10,c12,c13'. The 'Delimited by:' label is followed by a dropdown menu currently set to 'Tab'. The 'From:' label is followed by a dropdown menu showing '14: Add column on data 13'. At the bottom, there is a blue 'Execute' button on a light gray background.

This will give you 15th history dataset.

History			
Galaxy-P 101		1.0 GB	
15: <u>Cut on data 14</u>			
14: <u>Add column on data 13</u>			
13: <u>Sort on data 11 with FDR</u>			
12: <u>Sort on data 11</u>			
11: <u>Table peptide_prophet Raw101.pep.xml.csv</u>			
10: <u>Convert ProtXML to Tabular on data 9</u>			
9: <u>protein_prophet.peptide_prophet.X!tandem vs Target Decoy Human Contaminants.Peaklist Raw010.Peaklist Raw101.pepXML.pep.xml.protXML</u>			
8: <u>peptide_prophet Raw101.pep.xml</u>			
7: <u>X!tandem vs Target Decoy Human Contaminants.Peaklist Raw010.Peaklist Raw101.pepXML</u>			
6: <u>Peaklist Raw101</u>			
5: <u>Raw101.RAW</u>			
4: <u>Target Decoy Human Contaminants on data 3</u>			
3: <u>Merged Human UniProt cRAP</u>			
2: <u>CRAP</u>			
1: <u>Human UniProt</u>			

If you click on the Visualize / scatterplot icon you will see a scatterplot controls in the central window.

Data Controls

Plot Controls

Statistics

Use the following controls to change the data used by the chart. Use the 'Draw' button to render (or re-render) the chart with the current settings.

Data column for X:

column 3

Data column for Y:

column 3

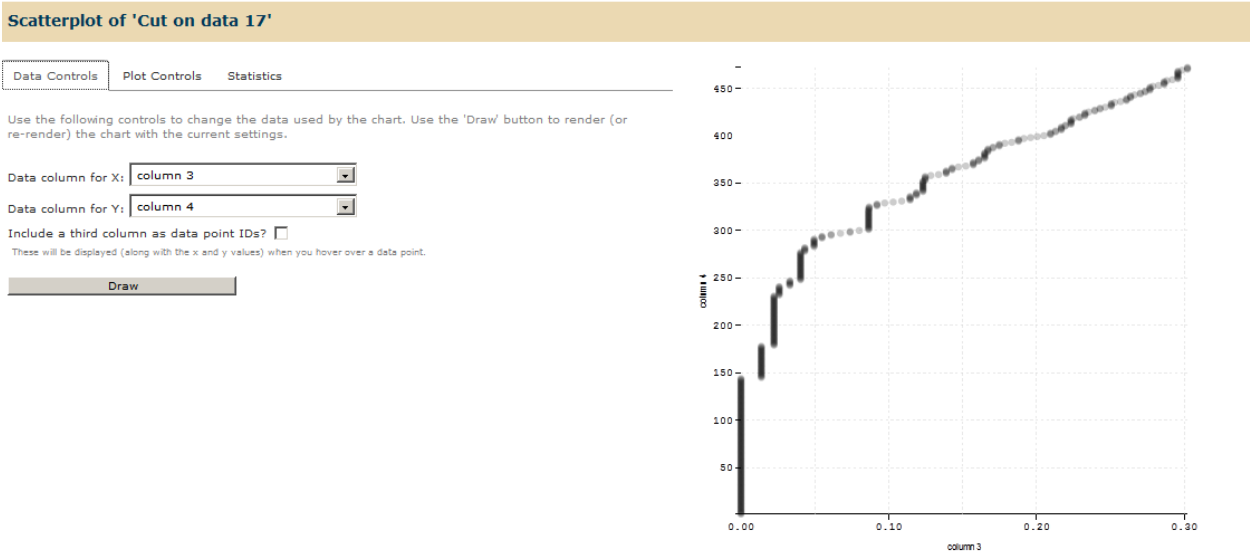
Include a third column as data point IDs?

☐

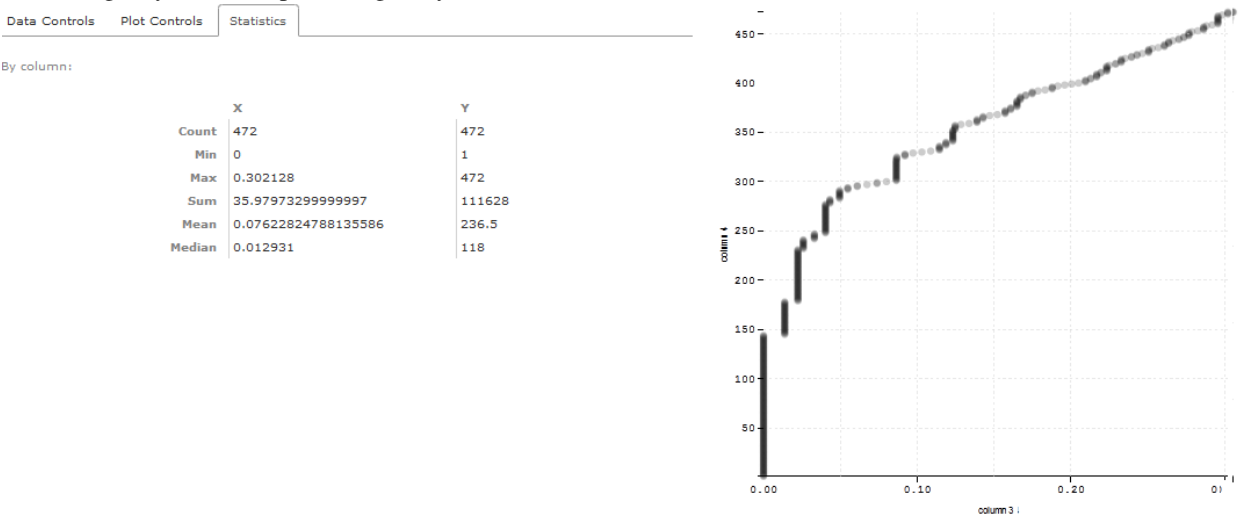
These will be displayed (along with the x and y values) when you hover over a data point.

Draw

Keep the Data column X: column 3 and change Data column Y: column 4. Click on Draw.

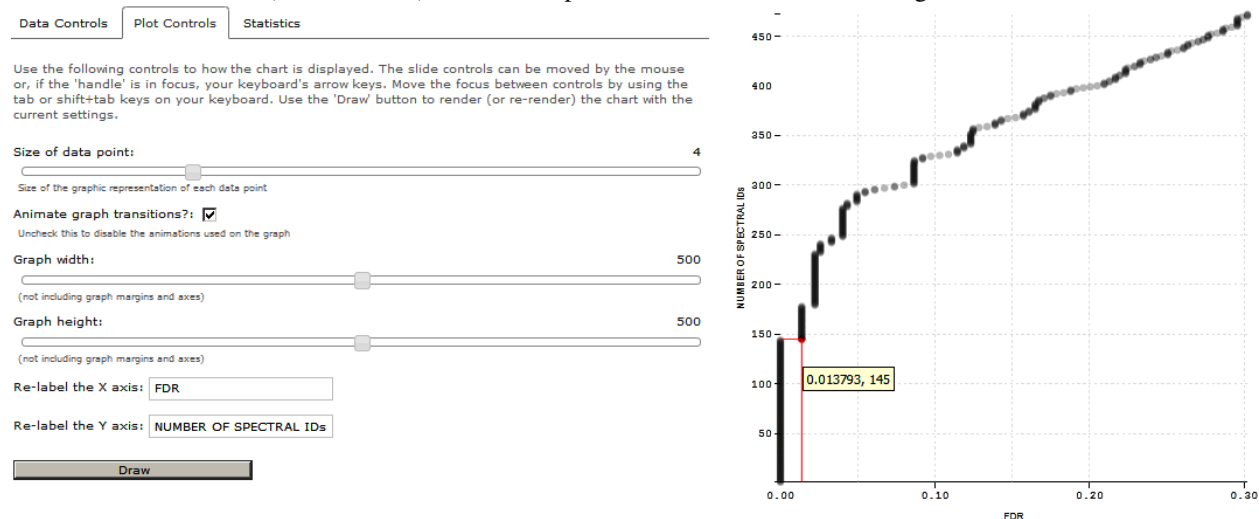


This will give you a ROC plot that gives you statistics and number of identifications at various levels of FDR.



Go to “Plot Controls” Tab and change the controls on your chart. Relabel X axis as “FDR” and Y-axis as ‘NUMBER OF SPECTRAL IDS’. Click on Draw to render the chart on current settings.

If you scroll on the graph, you will be able to see that 144 spectra were identified in this dataset before the first reverse match was encountered (FDR of 1.3%). Thus, 144 spectra were identified below 1% global FDR.



2.13 Understanding Galaxy Histories

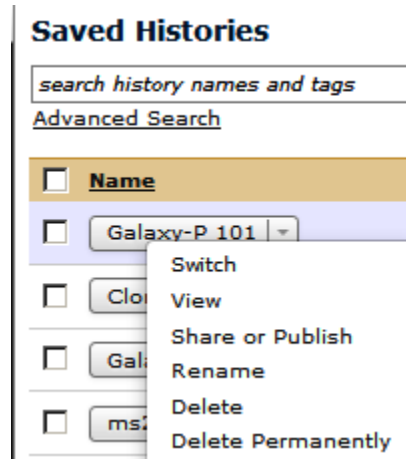
In Galaxy, your analyses live in histories such as this one:

History			
Galaxy-P 101		1.0 GB	
<u>15: Cut on data 14</u>			
<u>14: Add column on data 13</u>			
<u>13: Sort on data 11 with FDR</u>			
<u>12: Sort on data 11</u>			
<u>11: Table peptide prophet Raw101.pep.xml.csv</u>			
<u>10: Convert ProtXML to Tabular on data 9</u>			
<u>9: protein prophet.peptide prophet. X!tandem vs Target Decoy Human Contaminants.Peaklist Raw010.Peaklist Raw101.pepXML.pep.xml.protXML</u>			
<u>8: peptide prophet Raw101.pep.xml</u>			
<u>7: X!tandem vs Target Decoy Human Contaminants.Peaklist Raw010.Peaklist Raw101.pepXML</u>			
<u>6: Peaklist Raw101</u>			
<u>5: Raw101.RAW</u>			
<u>4: Target Decoy Human Contaminants on data 3</u>			
<u>3: Merged Human UniProt cRAP</u>			
<u>2: CRAP</u>			
<u>1: Human UniProt</u>			

Histories can be very large, you can have as many histories as you want, and all history behavior is controlled by the Options button on the top of the History pane:



















Many of the options here are self explanatory. If you create a new history, your current history does not disappear. If you would like to list all of your histories just choose Saved Histories and you will see a list of all your histories in the central window pane:



2.14 Converting Histories to Workflows

One of the history options listed is very special. It allows you to easily convert existing histories into analysis workflows. Why would you want to create a workflows out of a history? To redo the analysis again with minimal clicking.

Lets take a look at the history again:

History 	
 	 
Galaxy-P 101	1.0 GB
<u>15: Cut on data 14</u>	  
<u>14: Add column on data 13</u>	  
<u>13: Sort on data 11 with FDR</u>	  
<u>12: Sort on data 11</u>	  
<u>11: Table peptide prophet Raw101.pep.xml.csv</u>	  
<u>10: Convert ProtXML to Tabular on data 9</u>	  
<u>9: protein prophet.peptide prophet. X!tandem vs Target Decoy Human Contaminants.Peaklist Raw010.Peaklist Raw101.pepXML.pep.xml.protXML</u>	  
<u>8: peptide prophet Raw101.pep.xml</u>	  
<u>7: X!tandem vs Target Decoy Human Contaminants.Peaklist Raw010.Peaklist Raw101.pepXML</u>	  
<u>6: Peaklist Raw101</u>	  
<u>5: Raw101.RAW</u>	  
<u>4: Target Decoy Human Contaminants on data 3</u>	  
<u>3: Merged Human UniProt cRAP</u>	  
<u>2: CRAP</u>	  
<u>1: Human UniProt</u>	  

You can see that this history contains all steps of our analysis. So by building this history we have actually built a complete record of our analysis with Galaxy preserving all parameter settings applied at every step. Wouldn't it be nice to just convert this history into a workflow that we'll be able to execute again and again? This can be done by clicking on Options button and selecting Extract Workflow option:



The center pane will change as shown below and you will be able to choose which steps to include/exclude and how to name the newly created workflow. In this case I named it “Galaxy-P 101 Workflow”:

The following list contains each tool that was run to create the datasets in your current history. Please select those that you wish to include in the workflow.

Tools which cannot be run interactively and thus cannot be incorporated into a workflow will be shown in gray.

Workflow name
Galaxy-P 101 Workflow

Create Workflow Check all Uncheck all

Tool	History items created
Upload File <i>This tool cannot be used in workflows</i>	1: Human UniProt <input checked="" type="checkbox"/> Treat as input dataset
Upload File <i>This tool cannot be used in workflows</i>	2: cRAP <input checked="" type="checkbox"/> Treat as input dataset
Merge FASTA Databases <input checked="" type="checkbox"/> Include "Merge FASTA Databases" in workflow	3: Merged Human UniProt cRAP
Create Decoy Database (reverse) <input checked="" type="checkbox"/> Include "Create Decoy Database (reverse)" in workflow	4: Target_Decoy_Human_Contaminants
Upload File <i>This tool cannot be used in workflows</i>	5: Raw101.RAW <input checked="" type="checkbox"/> Treat as input dataset
msconvert3_raw <input checked="" type="checkbox"/> Include "msconvert3_raw" in workflow	6: Peaklist Raw101
XITandem MSMS Search <input checked="" type="checkbox"/> Include "XITandem MSMS Search" in workflow	7: XITandem_vs_Target_Decoy_Human_Contaminants.Peaklist Raw101.Peaklist Raw101.pepXML
Peptide Prophet <input checked="" type="checkbox"/> Include "Peptide Prophet" in workflow	8: peptide_prophet Raw101.pep.xml
Protein Prophet <input checked="" type="checkbox"/> Include "Protein Prophet" in workflow	9: protein_prophet.peptide_prophet Raw101.pep.xml.protXML
Convert ProtXML to XLS <input checked="" type="checkbox"/> Include "Convert ProtXML to XLS" in workflow	10: Convert ProtXML to XLS on data 9
PepXML to Table <input checked="" type="checkbox"/> Include "PepXML to Table" in workflow	11: Table peptide_prophet Raw101.pep.xml.csv
Sort <input checked="" type="checkbox"/> Include "Sort" in workflow	12: Sort on data 11
Compute False Discovery Rate (FDR) <input checked="" type="checkbox"/> Include "Compute False Discovery Rate (FDR)" in workflow	13: Sort on data 11 with FDR
Add column <input checked="" type="checkbox"/> Include "Add column" in workflow	14: Add column on data 13

Once you click Create Workflow you will get the following message: "Workflow 'Galaxy-P 101 Workflow' created from current history". But where did it go? Click on Workflow link at the top of Galaxy interface and you will a list of all workflows with "Galaxy-P 101 Workflow" listed at the top:

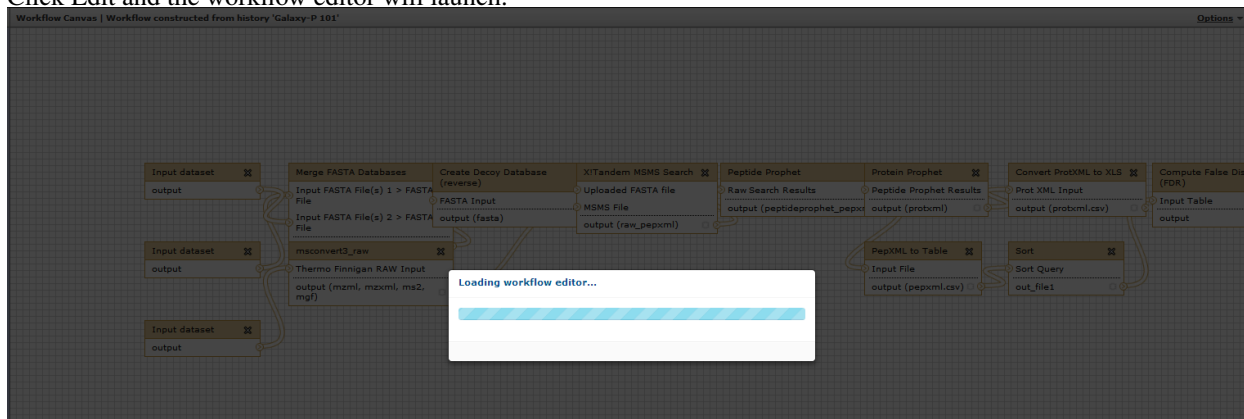
Analyze Data Workflow Shared Data Visualization Admin Help User

Workflow 'Galaxy-P 101 Workflow' created from current history.

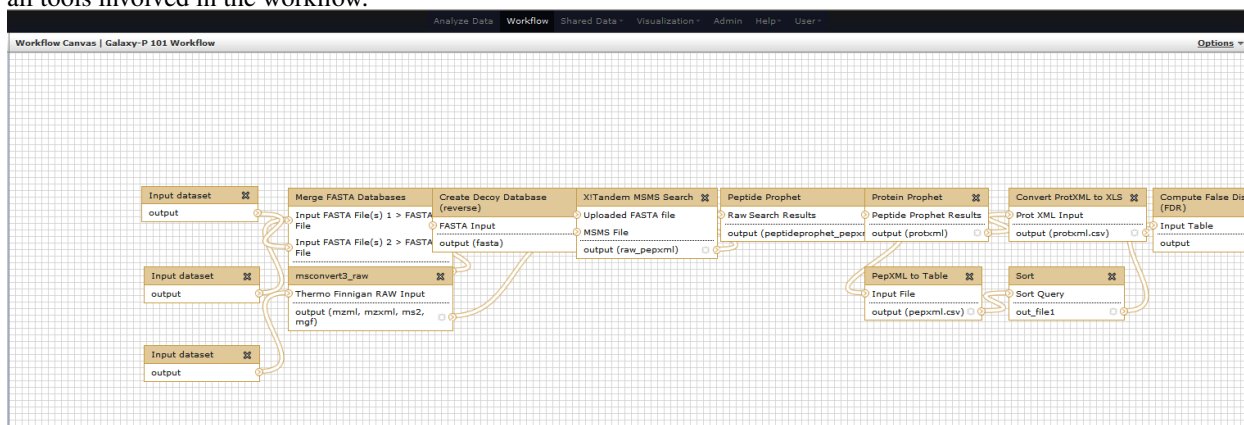
If you click on a triangle adjacent to the workflow's name you will see the following dialogue:



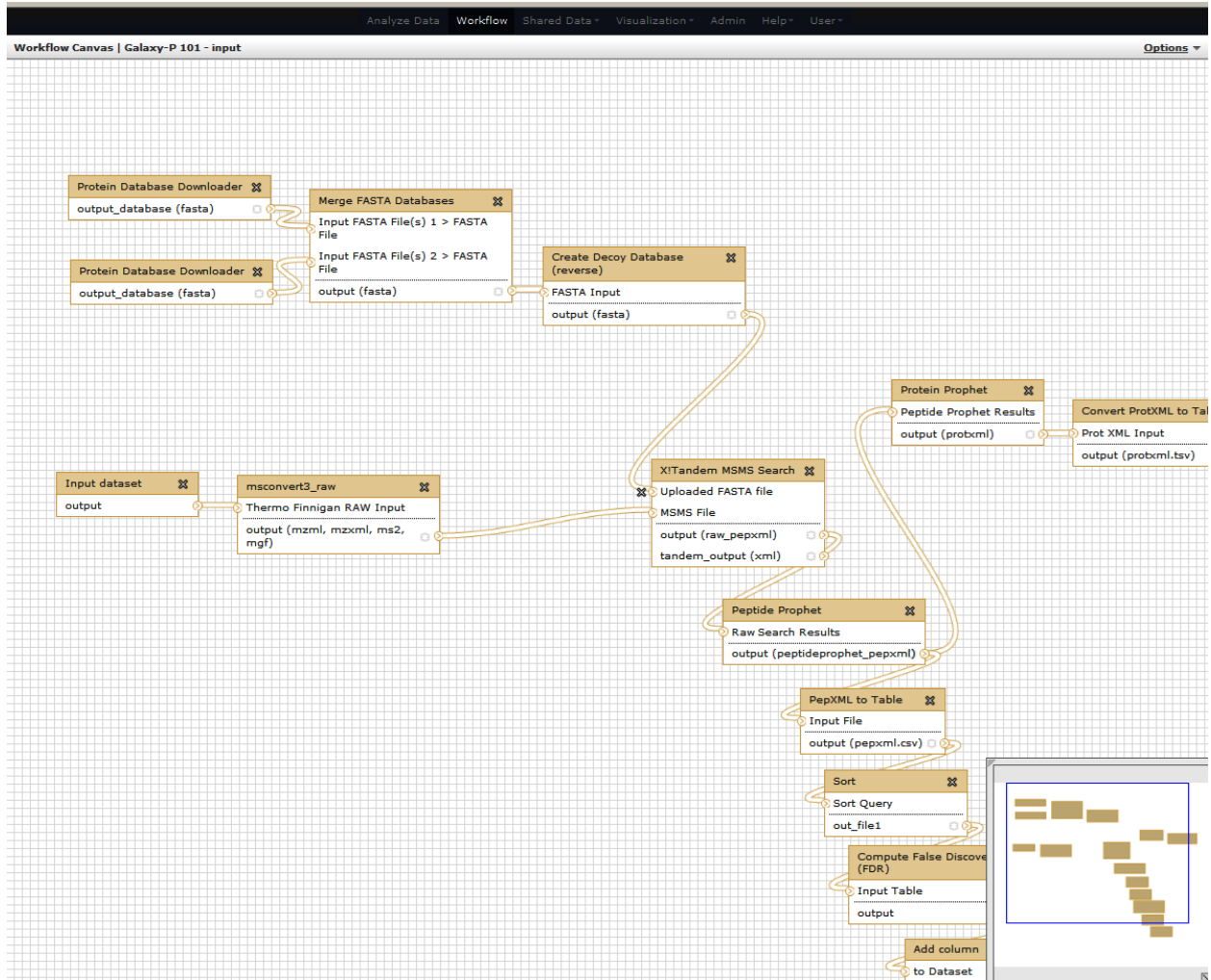
Click Edit and the workflow editor will launch.



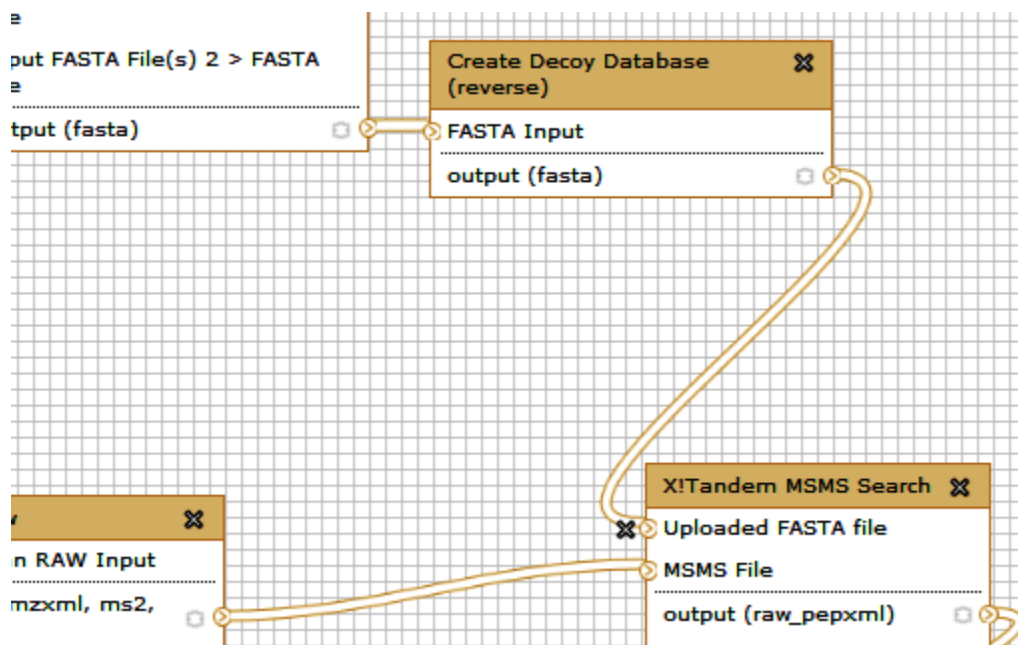
It will allow you to examine and change settings of this workflow as shown below. Note that you can click on each box so that you can see parameters of this tool on the right pane. This is how you can view and change parameters of all tools involved in the workflow.



You can also reorganize your workflow so that it makes intuitive sense. For example, I have compartmentalized this workflow as follows. However, as in any form of art – this is not the only way of representing your workflow. Re-architecture as the best way you wish to.

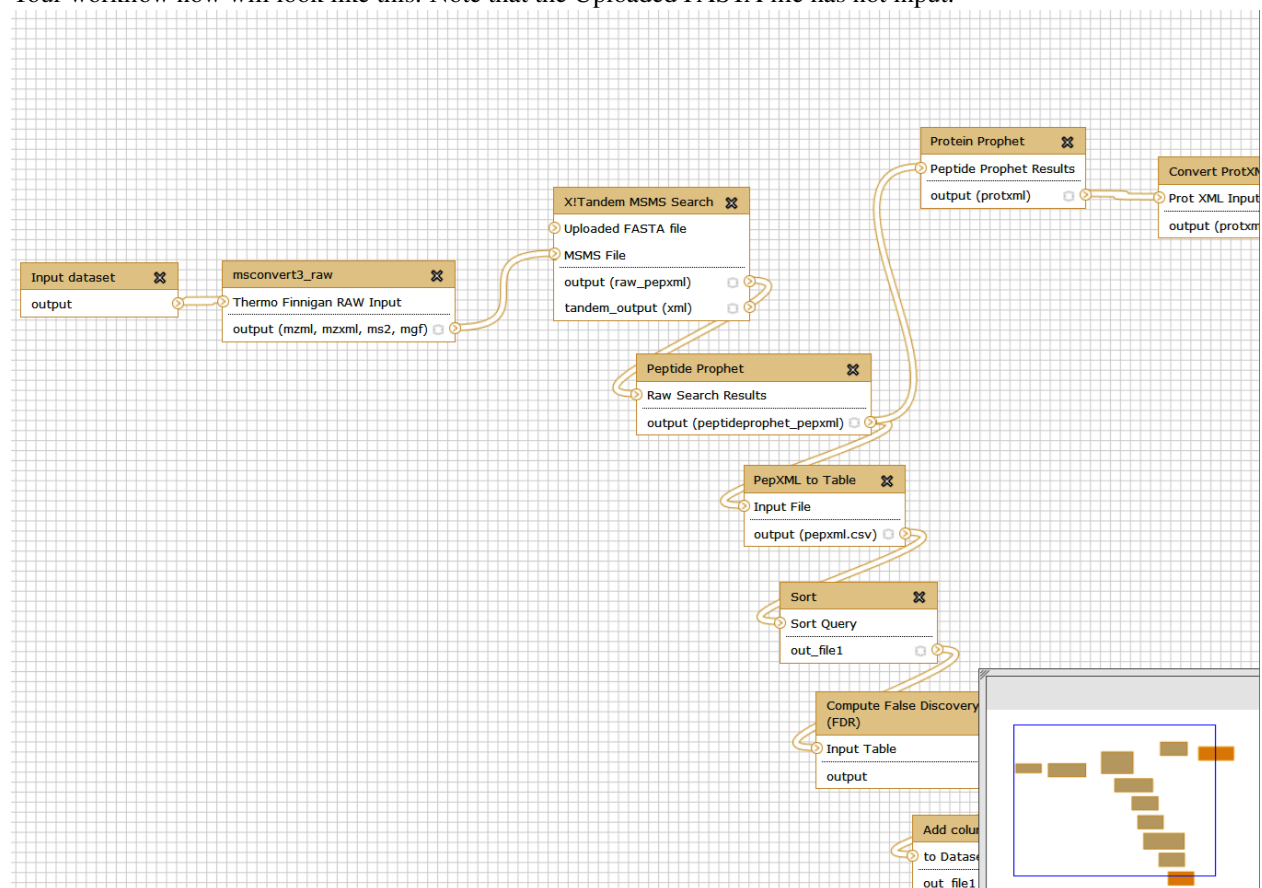


For this tutorial, we are going to reformat the workflow by replacing the earlier part of database creation with an input database. For this we excise the part where uploaded FASTA file merges with the X! Tandem search step.



You can delete all the steps involving FASTA database creation. Delete one step at a time...

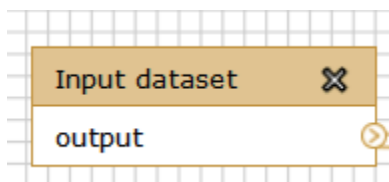
Your workflow now will look like this. Note that the Uploaded FASTA file has not input.



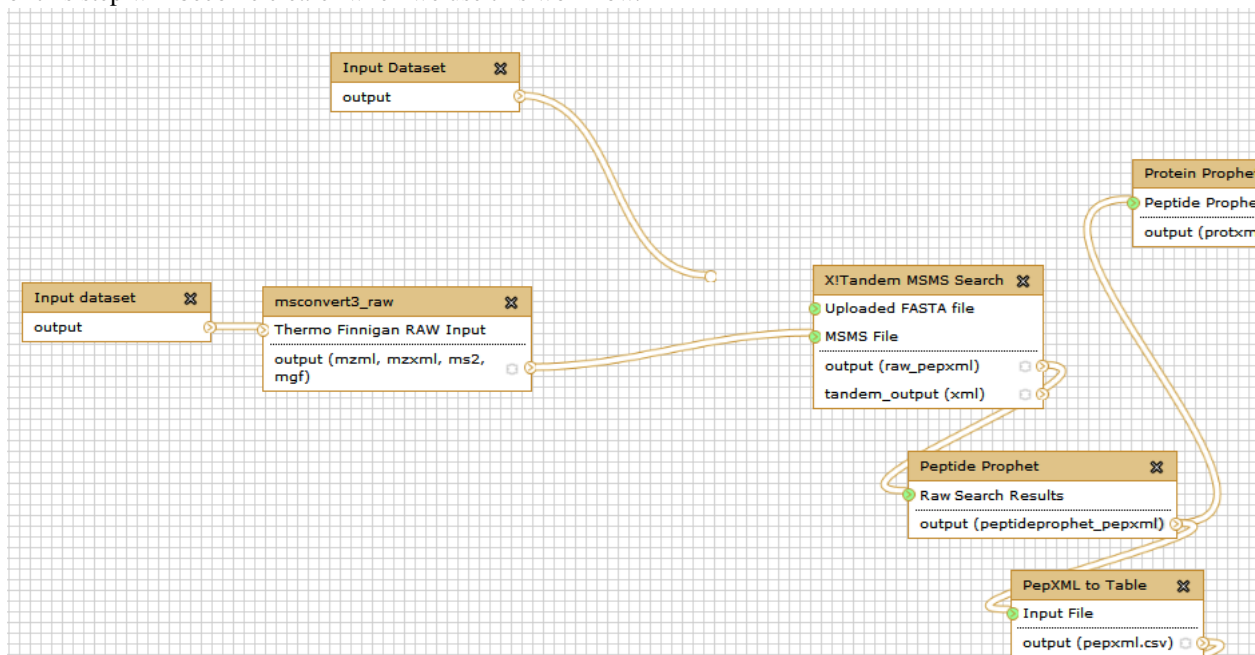
You can now scroll down to the left bottom corner of the screen and click on inputs. 'Input dataset' option will show up. Click on it.



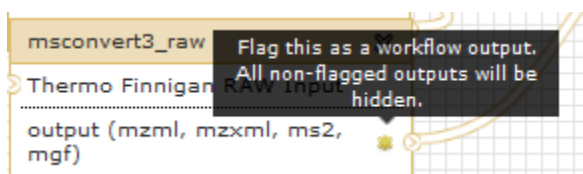
Once this is done, 'input dataset' box will appear in the central Window pane.



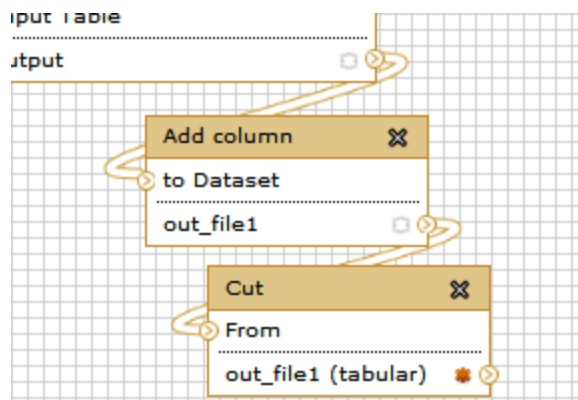
Position the Input dataset pane in such a way that you can join the output arrow on this box with the input arrow for the X!tandem search. In other words, we are providing a FASTA database input for the X!tandem search. The significance of this step will become clearer when we use this workflow.



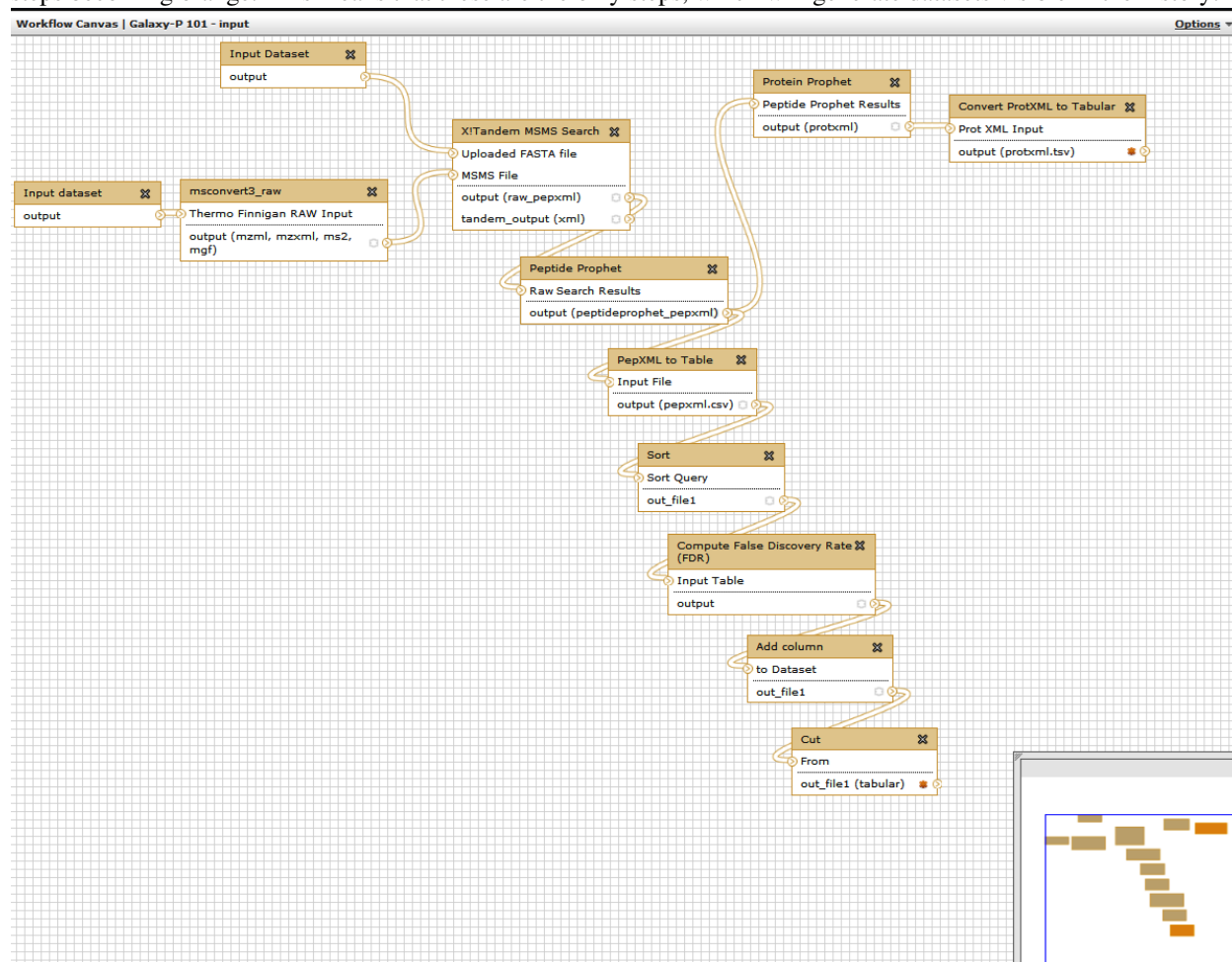
Below is mentioned one of the many things that you can do with workflows. When workflow is executed one is usually interested in the final product and not in the intermediate steps. These steps can be hidden by mousing over a small asterisk in the lower right corner of every tool box:



Yet there is a catch. In a newly created workflow all steps are hidden by default and default behavior of Galaxy is that if all steps of a given workflow are hidden, then nothing gets hidden in the history. This may be counterintuitive, but this is done to decrease the amount of clicking if you do want to hide some steps. So in our case if we want to hide all intermediate steps with the exception of the last one we will click that asterisk in last step of the workflow:



Once you do this the representation of the workflow in the bottom right corner of the editor will change with these steps becoming orange. This means that these are the only steps, which will generate datasets visible in the history:



Right now both inputs to the workflow look exactly the same. This is a problem as will be very confusing which input should be FASTA files and which should be RAW file. In your workflow you will see that the top input dataset connects to the X!tandem search, so it must correspond to the Target Decoy database. If you click on this box you will be able to rename the dataset in the right pane:

Details
<div>Input dataset</div> <p>Name:</p> <input type="text" value="Target Decoy FASTA"/>
<div> <div>Edit Step Attributes</div> <div> <p>Annotation / Notes:</p> <div> Target Decoy FASTA </div> <p>Add an annotation or notes to this step; annotations are available when a workflow is viewed.</p> </div> </div>

... and Thermo RAW file (second input).

Details
<div>Input dataset</div> <p>Name:</p> <input type="text" value="Thermo RAW File"/>
<div> <div>Edit Step Attributes</div> <div> <p>Annotation / Notes:</p> <div> RAW file from ThermoFinnigan LTO/ Orbitrap instrument. </div> <p>Add an annotation or notes to this step; annotations are available when a workflow is viewed.</p> </div> </div>

Feel free to annotate as many steps as you can so that it can be easier for you to revisit and understand the workflow or easier to share it with others.

Renaming outputs

Finally let's rename the workflow's output. For this click on one of the last datasets ("Convert ProtXML to Tabular") and in the Edit Step Actions dialogue box select "Rename Dataset".

Details

Tool: Convert ProtXML to XLS

Version: 0.1.0

Prot XML Input
Data input 'input' (prot.xml)

Edit Step Actions

Rename Dataset

output

Create

Rename Dataset on output

New output name:
Protein Table.

This action will rename the result dataset.

Add actions to this step; actions are applied when this workflow step completes.

Edit Step Attributes

Annotation / Notes:

Add an annotation or notes to this step; annotations are available when a workflow is viewed.

Similarly, for the other output dataset (“Cut”) , click on the box and in the Edit Step Actions dialogue box select “Rename Dataset”.

Details

Tool: Cut

Version: 1.0.1

Cut columns: ▼
c2,c10,c12,c13

Delimited by: ▼
Tab

From
Data input 'input' (txt)

Edit Step Actions

Rename Dataset ▼
out_file1 ▼ Create

Rename Dataset
on out_file1 ✕

New output name:
FDR| ROC curve.

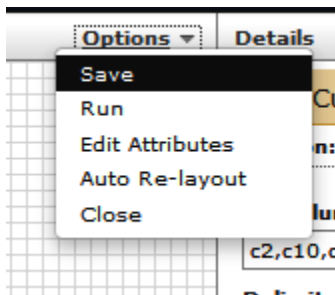
This action will rename the result dataset.

Add actions to this step; actions are applied when this workflow step completes.

Remember you can highlight as many “outputs” as you want to and rename them for sake of a more complete and shareable annotation.

Save! It is important...

Now let's save the changes we've made by clicking Options (top of the center pane) and selecting Save:

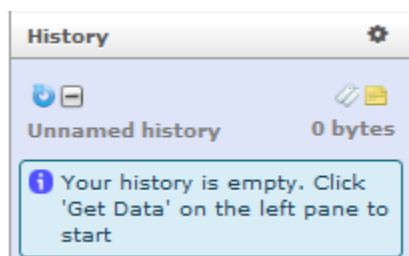


2.15 Applying Workflows to Your Data

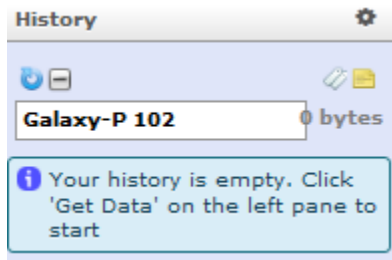
Let us use this workflow on another Raw file. For this go to Analyze Data → History options icon and select create new history.



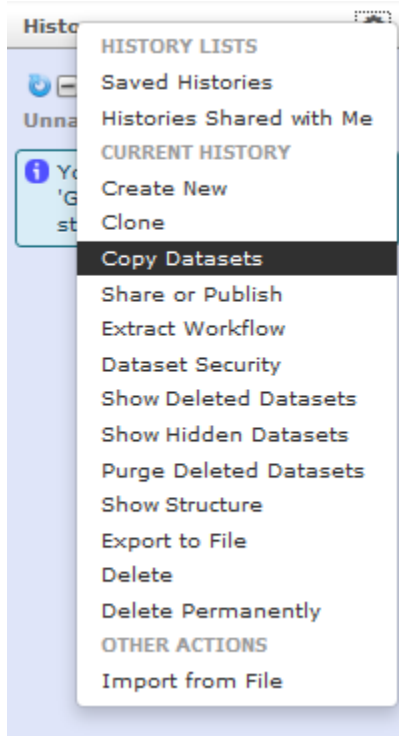
This will create a new history with no datafiles.



Let us name this history Galaxy-P 102.



We will transfer some of the files from our Galaxy-P 101 history to Galaxy-P 102 history. For this go to History options icon and click on copy datasets.



This will open up a new central window pane. Transfer the target-decoy FASTA file from Galaxy P-101 (Source History) to Galaxy-P 102 (Destination history).

i Copy any number of history items from one history to another.

Source History:

5: Galaxy-P 101 (current history)

- ☐ 1: Human UniProt
- ☐ 2: CRAP
- ☐ 3: Merged Human UniProt cRAP
- ☒ 4: Target_Decoys_Human_Contaminants on data 3
- ☐ 5: Raw101.RAW
- ☐ 6: Peaklist Raw101
- ☐ 7: X!tandem vs Target Decoy Human Contaminants.Peaklist Raw010.Peaklist Raw101.pepXML
- ☐ 8: peptide_prophet Raw101.pep.xml
- ☐ 9: protein_prophet.peptide_prophet.X!tandem vs Target Decoy Human Contaminants.Peaklist Raw010.Peaklist Raw101.pepXML.pep.xml.protXML
- ☐ 10: Convert ProtXML to Tabular on data 9
- ☐ 11: Table peptide_prophet Raw101.pep.xml.csv
- ☐ 12: Sort on data 11
- ☐ 13: Sort on data 11 with FDR
- ☐ 14: Add column on data 13
- ☐ 15: Cut on data 14

Destination History:

2: Galaxy-P 102

[Choose multiple histories](#)

— OR —

New history named:

Copy History Items

This will transfer the target-decoy FASTA file in your current history.

History ⚙

Galaxy-P 102
116.0 MB

1:
Target_Decoys_Human_Contaminants on data 3

Click on Galaxy-P 102 history and let us download another raw file from the following link:

[https://netfiles.umn.edu/users/pjagtap/Galaxy-P 101/Raw102.RAW](https://netfiles.umn.edu/users/pjagtap/Galaxy-P%20101/Raw102.RAW)

Upload File (version 1.1.3)

File Format:

Which format? See help below

File:

TIP: Due to browser limitations, uploading files larger than 2G

URL/Text:

Here you may specify a list of URLs (one per line) or paste th

Files uploaded via FTP:

File

Your FTP upload directory contains no files.

This Galaxy server allows you to upload files via FTP. To uplo

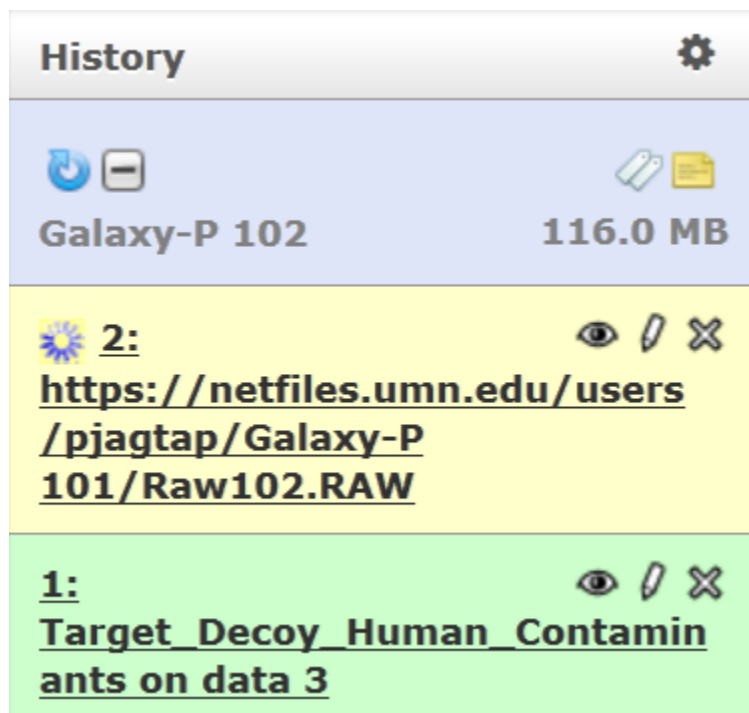
Convert spaces to tabs:


☐ Yes





Use this option if you are entering intervals by hand.

Genome:





Clicking on Execute will add second file to this history.






History 

Galaxy-P 102 **116.0 MB**

 **2:**   

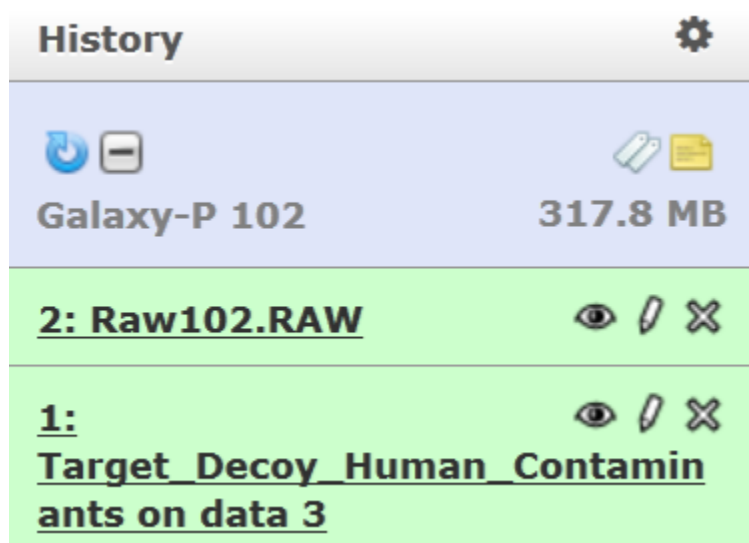
<https://netfiles.umn.edu/users/pjagtap/Galaxy-P101/Raw102.RAW>


1:   





Target_Decoys_Human_Contaminants on data 3

Rename the RAW file by using the pencil icon. Change the name to “Raw102”.




Now, for your rerun, your history template with 2 files (Galaxy-P 102) is ready.






History 

Galaxy-P 102 **317.8 MB**

2: Raw102.RAW   

1:   

Target_Decoys_Human_Contaminants on data 3

Go to Workflows tab and open the Galaxy-P 101 workflow and select “Run”.



Select appropriate files from your current history as inputs.

Step 1: Input dataset

RAW

RAW Input 

2: Raw102.RAW



type to filter

Step 2: Input dataset

Target Decoy FASTA

Target Decoy FASTA 

1: Target_Decoy_Huma..s on data 3



type to filter

Step 3: msconvert3 raw (version 0.1.0)

Step 4: X!Tandem MSMS Search (version 1.0.0)

Step 5: Peptide Prophet (version 1.1.0)

Step 6: PepXML to Table (version 1.0.0)

Step 7: Protein Prophet (version 1.0.0)

Step 8: Sort (version 1.0.1)

Step 9: Convert ProtXML to Tabular (version 0.1.0)

Step 10: Compute False Discovery Rate (FDR) (version 0.1)

Step 11: Add column (version 1.0.0)


46 Step 12: Cut (version 1.0.1) **Chapter 2. Galaxy-P 101 - Building Up and Using a Proteomics Workflow**

☐ Send results to a new history

You can choose to run this in the same history or create a history of “only outputs” in a new history from this analysis. For this tutorial, we will run it in the same history.

Click on Run Workflow. Luckily you do not have to wait as Galaxy will automatically start jobs once uploads have ended.


Analyze Data Workflow Shared Data Visualization Admin Help User


 Successfully ran workflow "Galaxy-P 101 - input". The following datasets have been added to the queue:


- 2: Raw102.RAW
- 1: Target_Decoy_Human_Contaminants on data 3
- 3: msconvert3_raw on data 2
- 4: XITandem_vs_Target_Decoy_Human_Contaminants on data 3.msconvert3_raw on data 2.msconvert3_raw on data 2.pepXML
- 5: peptide_prophet.XITandem_vs_Target_Decoy_Human_Contaminants on data 3.msconvert3_raw on data 2.msconvert3_raw on data 2.pepXML.pep.xml
- 6: peptide_prophet.XITandem_vs_Target_Decoy_Human_Contaminants on data 3.msconvert3_raw on data 2.msconvert3_raw on data 2.pepXML.pep.xml.csv
- 7: protein_prophet.peptide_prophet.XITandem_vs_Target_Decoy_Human_Contaminants on data 3.msconvert3_raw on data 2.msconvert3_raw on data 2.pepXML.pep.xml.protXML
- 8: Sort on data 6
- 9: Convert ProtXML to Tabular on data 7
- 10: Sort on data 6 with FDR
- 11: Add column on data 10
- 12: Cut on data 11


History 


   
Galaxy-P 102 317.8 MB



 [12: Cut on data 11](#)   


 [11: Add column on data 10](#)   



 [10: Sort on data 6 with FDR](#)   

 [9: Convert ProtXML to Tabular on data 7](#)   


 [8: Sort on data 6](#)   

 [7: protein prophet.peptide prophet.X!Tandem vs Target Decoy Human Contaminants on data 3.msconvert3 raw on data 2.msconvert3 raw on data 2.pepXML.pep.xml.protXML](#)   


 [6: peptide prophet.X!Tandem vs Target Decoy Human Contaminants on data 3.msconvert3 raw on data 2.msconvert3 raw on data 2.pepXML.pep.xml.csv](#)   

 [5: peptide prophet.X!Tandem vs Target Decoy Human Contaminants on data 3.msconvert3 raw on data 2.msconvert3 raw on data 2.pepXML.pep.xml](#)   

 [4: X!Tandem vs Target Decoy Human Contaminants on data 3.msconvert3 raw on data 2.msconvert3 raw on data 2.pepXML](#)   















 [3: msconvert3 raw on data 2](#)   

[2: Raw102.RAW](#)   

[1: Target Decoy Human Contaminants on data 3](#)   

Get coffee

As we mentioned above this will take some time, so go get coffee and then you will see this. Note that because all intermediate steps of the workflow were hidden, once it is finished you will only see the final dataset.

History 	
 	 
Galaxy-P 102	850.2 MB
<u>27: FDR ROC curve.</u>	  
<u>23: Protein Table.</u>	  
<u>3: Raw102.RAW</u>	  

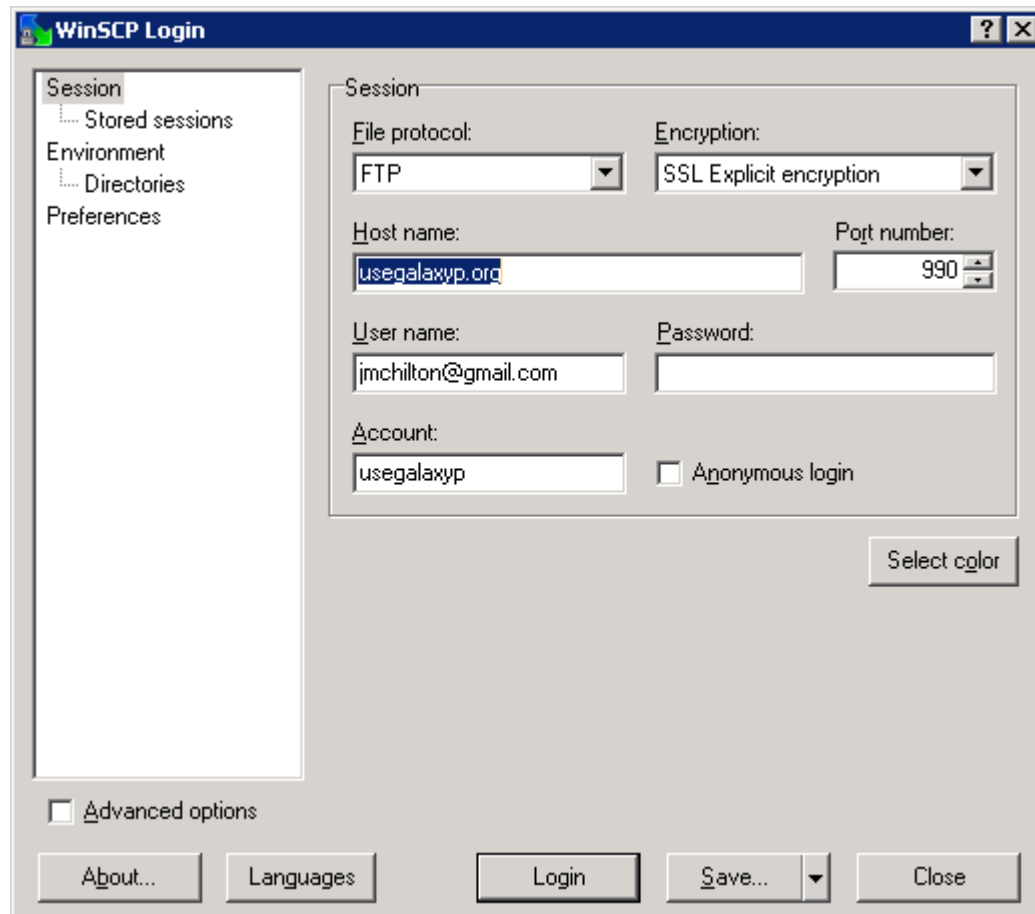
Galaxy provides the ability to upload files via the web interface (under the tool menu “Data Source” -> “Upload”), however this only allows one to upload one file at a time and web browsers generally limit uploads to 2 GB. Galaxy-P provides an FTP server that can be used to upload more and larger files.

3.1 Uploading Files

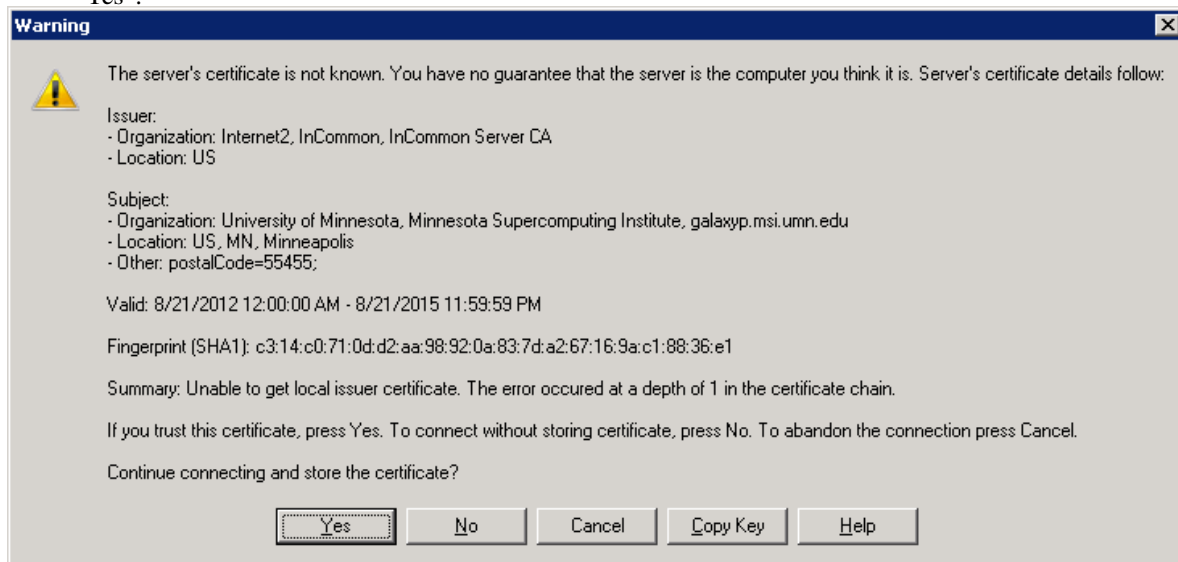
Property	Value
Hostname	usegalaxyp.org
Connection	FTPS / FTP + Encryption
Username	E-mail used to register Galaxy-P account.
Port	990

The following walkthrough demonstrates how to connect to Galaxy-P FTP server using [WinSCP](#) and upload RAW data files. WinSCP is demonstrated because it is a popular piece of freely available software, but many other tools could be used as long as they support FTP with encryption.

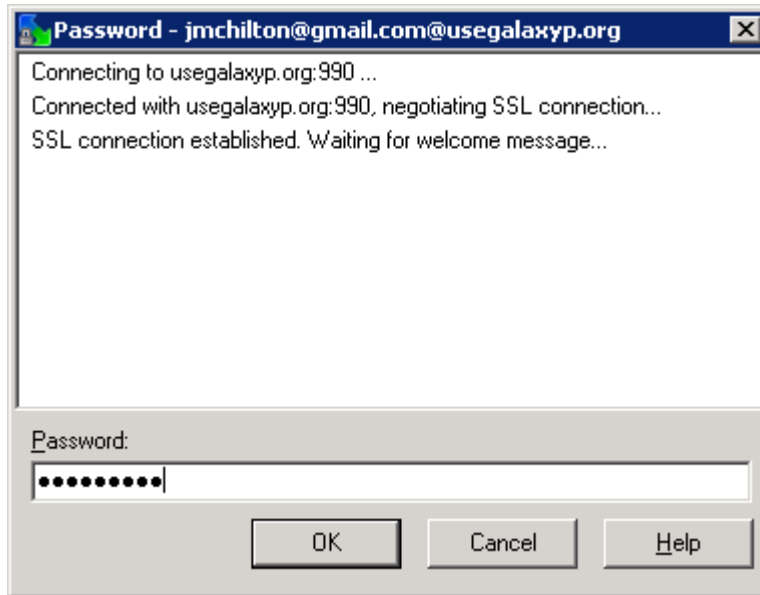
- Open WinSCP, specify connection information, and then click login. You may also want to give this connection a name and save it for later reuse.



- The first time you connect, you will likely be prompted to store the hosts SSL certificate, do this by clicking “Yes”.

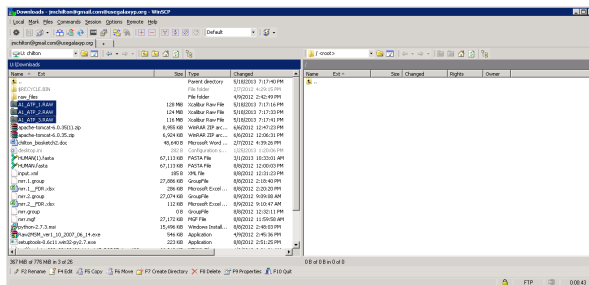


- When prompted for your password, please enter it and click “Okay”.

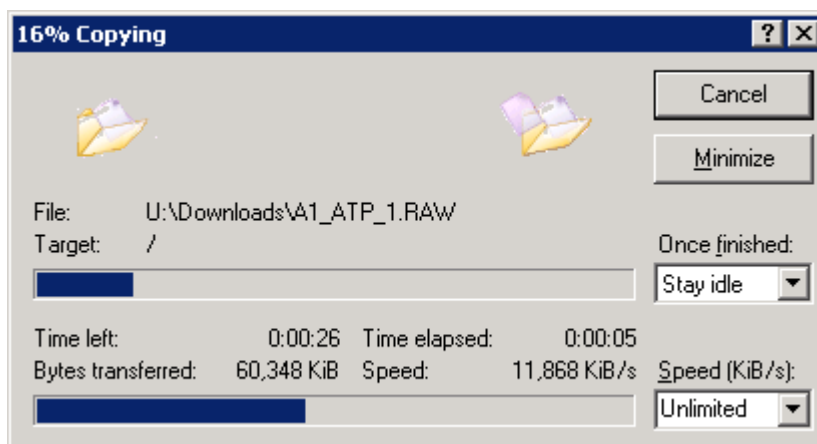


- If everything has gone well, you should now see two file browsers. The one on your left is your computer's files and the one on the right is your Galaxy-P staging area (which should be initially empty).

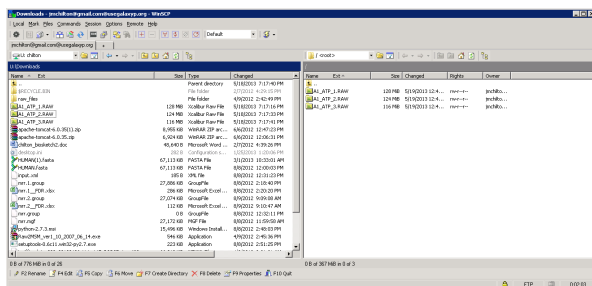
Using the left file browser, navigate to the files you wish to upload and select them.



- Drag and drop these files to the right panel to begin the transfer and wait as the files are transferred.

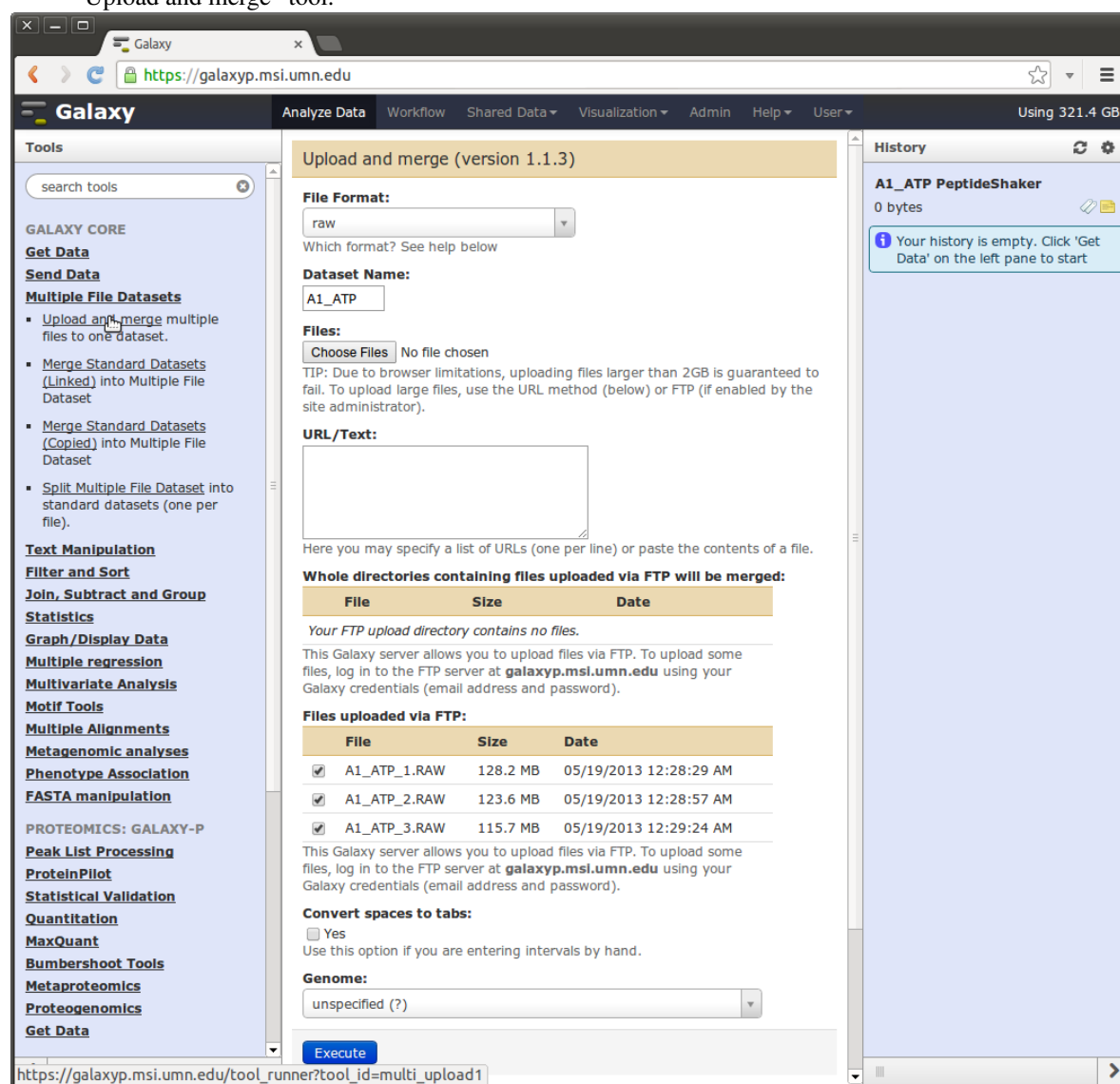


- Verify your files have been copied.



- These files may now be imported into a Galaxy history using the “Data Source” -> “Upload” tool. When filling out the upload information, instead of using the browser upload option simply check the uploaded files in the “Files uploaded by FTP” section.

Likewise, these a multiple file dataset can be created using these files and the “Multiple File Datasets” -> “Upload and merge” tool.



3.2 See Also

- [Galaxy Project Documentation on FTP Uploads](#)

Multiple File Datasets

4.1 Introduction

Traditional Galaxy workflows require a fixed number of input files and may produce a large number of intermediate files per input. Additionally, Galaxy renames datasets with each step, making tracking samples or fractions across long workflows very difficult. These issues together make Galaxy a problematic platform for workflows and application areas that require dealing with a large number of files. The prevalence of fractionated samples in mass spectrometry makes proteomics such an application area.

Galaxy-P however utilizes an extension to the core Galaxy framework called “Multiple File Datasets” designed to address these shortcomings in Galaxy. In simple terms, Galaxy-P can group many similar files into a single dataset (called a multiple file dataset). Normal Galaxy tools can then use these datasets and produce multiple file datasets of their own, operating on each item in parallel. In addition to keeping the number of datasets in the Galaxy manageable, this allows for the creation of workflows that can operate over any number of files and the individual files in the multiple file dataset are given consistent, trackable names across such complex workflows making sample tracking trivial.

4.2 Creating a Multiple File Dataset

Once you have an initial multiple file dataset, most tools when operating on these datasets will in return produce multiple file dataset outputs. So for most proteomic workflows, the first step is simply to create a multiple file RAW dataset or a multiple file peak list (e.g. mzML or mgf) dataset.

There are several ways to do this.

- One can use the “Multiple File Datasets” -> “Upload and Merge” tool along to create a multiple file datasets from files or a directory of files uploaded via FTP.

The screenshot displays the Galaxy web interface at <https://galaxyp.msi.umn.edu>. The top navigation bar includes links for Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, and User. The left sidebar contains a search bar and a list of tools categorized under 'GALAXY CORE' and 'PROTEOMICS: GALAXY-P'. The main panel shows the 'Upload and merge (version 1.1.3)' tool configuration. The 'File Format' is set to 'raw'. The 'Dataset Name' is 'A1_ATP'. The 'Files' section shows 'Choose Files' with a note that no file has been chosen. The 'URL/Text' section is empty. A table shows files uploaded via FTP:

File	Size	Date
<input checked="" type="checkbox"/> A1_ATP_1.RAW	128.2 MB	05/19/2013 12:28:29 AM
<input checked="" type="checkbox"/> A1_ATP_2.RAW	123.6 MB	05/19/2013 12:28:57 AM
<input checked="" type="checkbox"/> A1_ATP_3.RAW	115.7 MB	05/19/2013 12:29:24 AM

The right sidebar shows the 'History' panel with an empty list and a message: 'Your history is empty. Click 'Get Data' on the left pane to start'.

- One can use the “Multiple File Datasets” -> “Upload and Merge” tool and simply click the “Choose Files” button and select multiple file for upload.
- One can use the “Multiple File Datasets” -> “Upload and Merge” tool and paste in multiple URLs to have Galaxy download these files and create a multiple file dataset.

The screenshot shows the Galaxy web interface at localhost:8080. The main tool is 'Upload and merge (version 1.1.3)'. The interface is divided into three main sections:

- Left Sidebar (Tools):** Contains a search bar and a list of tool categories including GALAXY CORE, Text Manipulation, Statistics, Graph/Display Data, Multiple regression, Multivariate Analysis, Motif Tools, Multiple Alignments, Metagenomic analyses, FASTA manipulation, and PROTEOMICS: GALAXY-P.
- Main Workspace:**
 - File Format:** A dropdown menu set to 'raw'.
 - Dataset Name:** An empty text input field.
 - Files:** A section with a 'Choose Files' button and a message: 'No file chosen. TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).'.
 - URL/Text:** A text area containing three URLs:


```
ftp://ftp.pride.ebi.ac.uk/2013/04/PXD000141/A1_ATP_1_RAW
ftp://ftp.pride.ebi.ac.uk/2013/04/PXD000141/A1_ATP_2_RAW
ftp://ftp.pride.ebi.ac.uk/2013/04/PXD000141/A1_ATP_3_RAW
```
 - Whole directories containing files uploaded via FTP will be merged:** A section with a message: 'Your FTP upload directory contains no files. This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at localhost using your Galaxy credentials (email address and password).'.
 - Files uploaded via FTP:** A table showing three files:

File	Size	Date
090313_01.RAW	85.7 MB	05/29/2013 01:04:35 PM
090313_02.RAW	77.6 MB	05/29/2013 01:04:35 PM
090313_03.RAW	85.2 MB	05/29/2013 01:04:35 PM
 - Convert spaces to tabs:** A checkbox labeled 'Yes'.
- Right Sidebar (History):** Shows 'Unnamed history' with 0 bytes. A message states: 'Your history is empty. Click 'Get Data' on the left pane to start.'

A Simple Workflow using peptide-shaker

This section is a simple walkthrough of using Galaxy-P and multiple file datasets to analysis a collection of RAW files using `peptide-shaker`

The screenshot displays the Galaxy web interface at <https://galaxyp.msi.umn.edu>. The main tool being configured is 'Upload and merge (version 1.1.3)'. The configuration panel includes the following sections:

- File Format:** Set to 'raw'.
- Dataset Name:** Set to 'A1_ATP'.
- Files:** A 'Choose Files' button indicates no files are currently chosen. A tip notes that files larger than 2GB may fail due to browser limitations.
- URL/Text:** A text area for specifying URLs or pasting file contents.
- Whole directories containing files uploaded via FTP will be merged:** A table showing no files are currently merged.

File	Size	Date
Your FTP upload directory contains no files.		
- Files uploaded via FTP:** A table listing three files uploaded via FTP.

File	Size	Date
<input checked="" type="checkbox"/> A1_ATP_1.RAW	128.2 MB	05/19/2013 12:28:29 AM
<input checked="" type="checkbox"/> A1_ATP_2.RAW	123.6 MB	05/19/2013 12:28:57 AM
<input checked="" type="checkbox"/> A1_ATP_3.RAW	115.7 MB	05/19/2013 12:29:24 AM
- Convert spaces to tabs:** A checkbox labeled 'Yes' is selected.
- Genome:** Set to 'unspecified (?)'.

The right sidebar shows the 'History' panel with an entry 'A1_ATP PeptideShaker' (0 bytes). A message indicates that the history is empty and suggests clicking 'Get Data' to start.

Galaxy - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Galaxy Galaxy

https://usegalaxyp.org/root

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 734.8 MB

Tools

search tools

GALAXY CORE

[Get Data](#)

[Multiple File Datasets](#)

[Text Manipulation](#)

[Filter and Sort](#)

[Join, Subtract and Group](#)

[Statistics](#)

[Graph/Display Data](#)

[Multiple regression](#)

[Multivariate Analysis](#)

[Motif Tools](#)

[Multiple Alignments](#)

[Metagenomic analyses](#)

[Phenotype Association](#)

[FASTA manipulation](#)

PROTEOMICS: GALAXY-P

[Peak List Processing](#)

[Statistical Validation](#)

[Quantitation](#)

[MaxQuant](#)

[Bumbershoot Tools](#)

[Metaproteomics](#)

[Proteogenomics](#)

[Get Data](#)

PROTEOMICS: COMMUNITY

[Peptide Shaker](#)

[mzMatch](#)

[Open MS](#)

[ProtK](#)

[Utilities](#)

[Visualization](#)

[FASTA Manipulation](#)

[Adapt](#)

BIOINFORMATICS

[NCBI BLAST+](#)

[EMBOSS](#)

[Gene Prediction](#)

Welcome to Galaxy-P (beta).

Galaxy-P is a variant of [Galaxy](#) tailored toward proteomic data analysis.

Galaxy-P is running on disk that is not backed up, you should not assume data uploaded to or generated with Galaxy-P will be available in the future. Additionally, Galaxy-P is in very active development - things will break and things will change, so your patience is requested.

Tweets

Follow @usegalaxyp

The Galaxy-P Project @usegalaxyp 6 Nov
Just posted a screencast describing how to export data from TINT to Galaxy-P.
youtube.com/watch?v=2wxDJj...

The Galaxy-P Project @usegalaxyp 5 Nov
News about the Galaxy-P project and servers will be posted on this twitter account.

Load More

Tweet to @usegalaxyp

Getting Starts with Galaxy-P

A good place to start is the following Galaxy-P 101 page: [Building up and using a proteomics workflow](#).

The main Galaxy team's [screencasts page](#) has many useful screencasts and tutorials. While some are genomics specific, many will be useful for proteomics analysis in Galaxy-P.

Useful Links

- [Uploading databases using links](#).
- [Using the protein database downloader tool](#).
- [The Galaxy-P source code on Bitbucket](#)
- [The Galaxy-P issue tracker on Trello](#)

This project is funded by the [NSF](#).

History

PeptideShaker 101
367.4 MB

1: A1_ATP
Multifile dataset raw
format: m:raw, database: 2
Example Multiple File Dataset

A1_ATP_1.RAW (File 1)
A1_ATP_2.RAW (File 2)
A1_ATP_3.RAW (File 3)

Galaxy - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Galaxy

https://usegalaxy.org/root

pride raw files

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 734.8 MB

Tools

search tools

GALAXY CORE

[Get Data](#)

[Send Data](#)

[Multiple File Datasets](#)

[Text Manipulation](#)

[Filter and Sort](#)

[Join, Subtract and Group](#)

[Statistics](#)

[Graph/Display Data](#)

[Multiple regression](#)

[Multivariate Analysis](#)

[Motif Tools](#)

[Multiple Alignments](#)

[Metagenomic analyses](#)

[Phenotype Association](#)

[FASTA manipulation](#)

PROTEOMICS: GALAXY-P

[Peak List Processing](#)

[Statistical Validation](#)

[Quantitation](#)

[MaxQuant](#)

[Bumbershoot Tools](#)

[Metaproteomics](#)

[Proteogenomics](#)

[Get Data](#)

- PRIDE BioMart Download proteomic data from the EBI Pride BioMart.
- [Protein Database Downloader](#)

PROTEOMICS: COMMUNITY

[Peptide Shaker](#)

[mzMatch](#)

[Open MS](#)

[ProtK](#)

[Utilities](#)

[Visualization](#)

[FASTA Manipulation](#)

[Adapt](#)

Protein Database Downloader (version 0.2.0)

Download from:

UniProtKB

Taxonomy:

Mycobacterium tuberculosis H37Rv (MTB) [Not compl]

reviewed:

UniProtKB

Proteome Set:

Any

Include form data:

☐

Execute

Output

A FASTA file containing the specified protein sequences.

History

PeptideShaker 101

367.4 MB

1: A1_ATP

Multifile dataset raw

format: m:raw, database: 2

Example Multiple File Dataset

A1_ATP_1.RAW (File 1)

A1_ATP_2.RAW (File 2)

A1_ATP_3.RAW (File 3)

The screenshot shows the Galaxy web interface in Mozilla Firefox. The browser address bar displays `https://usegalaxy.org/root`. The Galaxy header includes navigation tabs: **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Help**, and **User**. The top right corner indicates 'Using 734.8 MB'.

Tools Panel (Left): A search bar is at the top. Below it, categories are listed: **GALAXY CORE** (Get Data, Send Data, Multiple File Datasets, Text Manipulation, Filter and Sort, Join, Subtract and Group, Statistics, Graph/Display Data, Multiple regression, Multivariate Analysis, Motif Tools, Multiple Alignments, Metagenomic analyses, Phenotype Association, FASTA manipulation), **PROTEOMICS: GALAXY-P** (Peak List Processing, Statistical Validation, Quantitation, MaxQuant, Bumpershoot Tools, Metaproteomics, Proteogenomics, Get Data), **PROTEOMICS: COMMUNITY** (Peptide Shaker, Create Target-Decoy Database, mzMatch, Open MS, ProtK), and **Peptide Shaker** (Peptide Shaker, Create Target-Decoy Database).

Create Target-Decoy Database (version 0.1.0) Panel (Center):

- FASTA Input:** A dropdown menu shows '2: Protein Database'.
- Execute:** A blue button.
- What it does:** 'Given an input database, this tool will produce a target-decoy database in the format required by PeptideShaker using dbtoolkit.'
- Citation:** 'For the underlying tool, please cite Martens et al. DBToolKit: processing protein databases for peptide-centric proteomics. Bioinformatics (2005) vol. 21 (17) pp. 3584-5. If you use this tool in Galaxy, please cite Chilton J, et al. <https://bitbucket.org/galaxy/galaxy-toolshed-peptideshaker>.'

History Panel (Right):

- PeptideShaker 101:** 367.4 MB.
- 2: Protein Database:** A dataset icon.
- 1: A1_ATP:** A dataset icon. Description: 'Multifile dataset raw format: m:raw, database: 2 Example Multiple File Dataset'. It lists three files: 'A1_ATP_1.RAW (File 1)', 'A1_ATP_2.RAW (File 2)', and 'A1_ATP_3.RAW (File 3)'.

Galaxy - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Galaxy

https://usegalaxy.org/root

pride raw files

Using 740.1 MB

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Tools

Get Data

Send Data

Multiple File Datasets

Text Manipulation

Filter and Sort

Join, Subtract and Group

Statistics

Graph/Display Data

Multiple regression

Multivariate Analysis

Motif Tools

Multiple Alignments

Metagenomic analyses

Phenotype Association

FASTA manipulation

PROTEOMICS: GALAXY-P

Peak List Processing

- msconvert3** Convert and filter a mass spec peak list
- msconvert3_raw** Convert and filter a Thermo Finnigan RAW file
- Subset Peak List** against list of scan numbers or indices.
- MGF Formatter** convert peak lists into MGF files formatted for particular downstream applications.
- ms2preproc** preprocess and filter MS/MS spectra
- Search Formatter** Use MzXML2Search to convert peak lists.

Statistical Validation

Quantitation

MaxQuant

Bumbershoot Tools

Metaproteomics

Proteogenomics

Get Data

PROTEOMICS: COMMUNITY

msconvert3_raw (version 0.2.0)

Thermo Finnigan RAW Input:

1: A1_ATP

Output Type:

mgf

Use Filtering?:

☐

Advanced Settings:

Use Defaults

Execute

What it does

Converts between various mass spectrometry peak list representations. Additional options such as filtering and/or precursor recalculation are available.

You can view the original documentation [here](#).

Citation

For the underlying tool, please cite ProteoWizard: Open Source Software for Rapid Proteomics Tools Development. Darren Kessner; Matt Chambers; Robert Burke; David Agus; Parag Mallick. Bioinformatics 2008; doi: 10.1093/bioinformatics/btn323.

If you use this tool in Galaxy, please cite Chilton J, et al. <https://bitbucket.org/galaxy/galaxy-toolshed-msconvert>

History

PeptideShaker 101

372.7 MB

3: Create Target-Decoy Database on data 2

2: Protein Database

1: A1_ATP

Multifile dataset raw
format: m:raw, database: 2
Example Multiple File Dataset

A1_ATP_1.RAW (File 1)
A1_ATP_2.RAW (File 2)
A1_ATP_3.RAW (File 3)

https://usegalaxy.org/tool_runner?tool_id=msconvert3_raw

The screenshot displays the Galaxy web interface in a Mozilla Firefox browser. The address bar shows `https://usegalaxy.org/root`. The top navigation bar includes links for File, Edit, View, History, Bookmarks, Tools, and Help. The main interface is divided into three panels:

- Tools Panel (Left):** A sidebar with a search bar and a list of tool categories. Under "PROTEOMICS: GALAXY-P", the "Peptide Shaker" tool is highlighted. Below it, a description states: "Peptide Shaker Perform protein identification combining X! Tandem and OMSSA (using SearchGUI) and PeptideShaker pipeline." Other tools listed include "Create Target-Decoy Database", "mzMatch", "Open MS", "ProTK", "Utilities", "Visualization", and "FASTA Manipulation".
- Tool Configuration Panel (Center):** The "Peptide Shaker (version 0.1.0)" configuration page. It includes the following settings:
 - Protein Database:** 3: Create Target-Decoy on data 2
 - Input Peak Lists (mgf):** 4: msconvert3_raw on data 1 (with a link to "select files directly | by multifile")
 - Precursor Ion Tolerance Units:** Parts per million (ppm)
 - Precursor Ion Tolerance:** 10.0
 - Fragment Tolerance (Daltons):** 0.5
 - Enzyme:** Trypsin
 - Maximum Missed Cleavages:** 2
 - Fixed Modification:** x carbamidomethyl c
 - Variable Modification:** x oxidation of m, x gtp desthiobiotinc12
 - Minimum Charge:** 2
 - Maximum Charge:** 4
 - Forward Ion:** b
 - Reverse Ion:** y
 - Specify Advanced Search Options:** (checkbox)
 - Specify Advanced PeptideShaker Processing Options:** (checkbox)
 - Specify Advanced PeptideShaker Filtering Options:** (checkbox)An "Execute" button is at the bottom of the configuration panel.
- History Panel (Right):** A list of previous tool runs. The top entry is "PeptideShaker 101" (664.1 MB). Below it are "4: msconvert3_raw on data 1", "3: Create Target-Decoy Database on data 2", "2: Protein Database", and "1: A1_ATP" (Multifile dataset raw, format: m:raw, database: 2, Example Multiple File Dataset). The "A1_ATP" entry shows a list of files: A1_ATP_1.RAW (File 1), A1_ATP_2.RAW (File 2), and A1_ATP_3.RAW (File 3).

The bottom of the browser window shows the URL `https://usegalaxy.org/tool_runner?tool_id=peptide_shaker`.

Galaxy - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Galaxy

https://usegalaxy.org/root

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 1.1 GB

Spectrum File	Identification File(s)	Precursor RT	Precursor m/z
A1_ATP.16058.16058.	A1_ATP_1 (File 1).omx	4779.4458	964.13
A1_ATP.8510.8510.1	A1_ATP_1 (File 1).omx	2403.5502	791.40087890625
A1_ATP.8366.8366.	A1_ATP_1 (File 1).omx	2357.8309	693.7
A1_ATP.17278.17278.	A1_ATP_1 (File 1).omx A1_ATP_1 (File 1).t.xml	5155.1521	950.45
A1_ATP.1993.1993.	A1_ATP_1 (File 1).omx	424.4967	981.53
A1_ATP.10870.10870.	A1_ATP_1 (File 1).omx	3146.9381	920.13
A1_ATP.6633.6633.	A1_ATP_1 (File 1).omx	1820.4371	741.93
A1_ATP.7348.7348.	A1_ATP_1 (File 1).omx	2041.1308	1159.9
A1_ATP.7331.7331.1	A1_ATP_1 (File 1).omx	2036.1117	719.6
A1_ATP.22984.22984.1	A1_ATP_1 (File 1).omx	6769.8229	587.56
A1_ATP.12992.12992.	A1_ATP_1 (File 1).omx	3821.3314	977.91
A1_ATP.11403.11403.	A1_ATP_1 (File 1).omx	3316.1163	993.13
A1_ATP.10047.10047.	A1_ATP_1 (File 1).omx	2885.758	957.52
A1_ATP.12494.12494.	A1_ATP_1 (File 1).omx	3663.9835	726.89
A1_ATP.14147.14147.1	A1_ATP_1 (File 1).omx	4186.682	518.99
A1_ATP.8474.8474.	A1_ATP_1 (File 1).omx	2392.1006	1045.87
A1_ATP.4615.4615.	A1_ATP_1 (File 1).omx	1225.8618	404.85
A1_ATP.11351.11351.	A1_ATP_1 (File 1).omx	3299.3848	918.19
A1_ATP.4388.4388.	A1_ATP_1 (File 1).omx	1158.5862	523.07
A1_ATP.11429.11429.	A1_ATP_1 (File 1).omx	3324.5418	1041.6
A1_ATP.4579.4579.	A1_ATP_1 (File 1).omx	1215.2906	472.54
A1_ATP.19966.19966.	A1_ATP_1 (File 1).omx	5971.9205	639.66
A1_ATP.6190.6190.	A1_ATP_1 (File 1).omx	1685.705	744.05
A1_ATP.13359.13359.	A1_ATP_1 (File 1).omx	3938.0828	1043.25
A1_ATP.11163.11163.	A1_ATP_1 (File 1).t.xml	3239.4479	666.84
A1_ATP.8033.8033.	A1_ATP_1 (File 1).omx	2253.6402	861.45
A1_ATP.5468.5468.1	A1_ATP_1 (File 1).omx	1472.6248	800.39
A1_ATP.8966.8966.	A1_ATP_1 (File 1).omx	2546.2684	976.17
A1_ATP.22172.22172.1	A1_ATP_1 (File 1).omx	6573.0315	727.456237792969
A1_ATP.15645.15645.	A1_ATP_1 (File 1).omx	4653.0733	776.43
A1_ATP.15020.15020.	A1_ATP_1 (File 1).omx	4459.4024	1039.87
A1_ATP.9314.9314.1	A1_ATP_1 (File 1).omx	2655.4786	824.43
A1_ATP.4952.4952.	A1_ATP_1 (File 1).omx	1324.9834	571.06
A1_ATP.7119.7119.	A1_ATP_1 (File 1).omx	1969.2179	705.05
A1_ATP.5902.5902.	A1_ATP_1 (File 1).omx	1599.1762	737.87
A1_ATP.7135.7135.	A1_ATP_1 (File 1).t.xml	1974.2127	880.49
A1_ATP.9994.9994.	A1_ATP_1 (File 1).omx	2868.7206	843.09
A1_ATP.13544.13544.1	A1_ATP_1 (File 1).omx	3995.9774	324.43
A1_ATP.8606.8606.	A1_ATP_1 (File 1).omx	2433.4604	728.79
A1_ATP.21233.21233.	A1_ATP_1 (File 1).omx	6342.1764	867.65
A1_ATP.12825.12825.	A1_ATP_1 (File 1).t.xml	3768.0869	797.47
A1_ATP.6806.6806.	A1_ATP_1 (File 1).omx	1873.5239	974.55
A1_ATP.10807.10807.1	A1_ATP_1 (File 1).omx	3126.8176	991.54
A1_ATP.11264.11264.	A1_ATP_1 (File 1).omx	3271.3843	1049.54
A1_ATP.9322.9322.	A1_ATP_1 (File 1).omx	2657.9808	903.1
A1_ATP.14312.14312.	A1_ATP_1 (File 1).omx	4239.0255	831.68
A1_ATP.4477.4477.	A1_ATP_1 (File 1).omx	1184.9677	343.44
A1_ATP.8417.8417.	A1_ATP_1 (File 1).omx	2374.0634	711.43

History

PeptideShaker 101
764.5 MB

8: PeptideShaker PSM
Report for data 3 and data 4

7: PeptideShaker
Protein Report for data 3 and data 4

6: PeptideShaker
Peptide Report for data 3 and data 4

5: PeptideShaker CPS
results for data 3 and data 4
96.9 MB
format: cps, database: 2
makeblastdb command: /opt/galaxy/tools/searchgui/1.13.1/share/java/SearchGUI-1.13.1/resources/makeblastdb/linux/makeblastdb -in dataset_7.dat Reindexing. Wed May 29 20:33:18 UTC 2013 Formatting dataset_7.dat for OMSSA. Building Building a new new

4: msconvert3_raw on
data 1

3: Create Target-Decoy
Database on data 2
8,196 sequences
format: fasta, database: 2

```
>tr|Q50655|Q50655_MYCTU Uncharacteri
MSTRQAAEADLAGKAAQYRPELARYAQRVMDLHE
MSRLSGYLTTPQARATFEAVLAKLAAPGAINPDHTE
GLLAGLRALIASGKLGQHNLFPVSIVVTTLTDLQF
AHYSPASGRYPQAI FDHGTPLALYHTRKIASPAQF
HHVTANTSTGRDITELTLACGPDNRLAEGKWTTH
```

2: Protein Database

1: A1_ATP
Multifile dataset raw
format: m:raw, database: 2

Indices and tables

- `genindex`
- `modindex`
- `search`