

## Guide for Performing High Throughput Searches on Quest with Galaxy

1. Log into galaxy at: <http://galaxy.kelleher.northwestern.edu/> using your Northwestern email (e.g alexandra.vannispenn@northwestern.edu) and password.

The screenshot displays the Galaxy web interface. On the left is a sidebar with a 'Tools' section containing a search bar and a list of tool categories: Analyze Data, Kelleher Testing, Get Data, Send Data, Lift-Over, Text Manipulation, Filter and Sort, Join, Subtract and Group, Convert Formats, Extract Features, Fetch Sequences, Fetch Alignments, Get Genomic Scores, NGS TOOLBOX BETA, Statistics, NGS: QC and manipulation, Graph/Display Data, NGS: RNA Analysis, Phenotype Association, GalaxyP, Metabonomics, and Workflows. The main content area is titled 'Standardized Searching in Four Simple Steps' and lists four steps: 1. Upload published standard search workflow, 2. Upload data, 3. Create dataset collection, and 4. Click run!. Below the steps is a paragraph about the National Resource for Translational and Developmental Proteomics (NRTDP) and its role in disease research. At the bottom of the main area are logos for Northwestern Proteomics, NUIT (Northwestern University Information Technology), and Northwestern University. On the right side of the interface is a 'History' panel showing 'Unnamed history' with 0 bytes and a message: 'This history is empty. You can load your own data or get data from an external source.'

Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of EIC at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NIGMS, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.

## 2. Orient yourself on the page.

The screenshot shows the Galaxy web interface. A red box highlights the top navigation bar, which contains seven subheadings: **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Admin**, **Help**, and **User**. Another red box highlights the left sidebar, which contains a search bar and a list of tools categorized under **Analyze Data**, **Workflows**, and **Metabonomics**. A third red box highlights the right sidebar, which contains a search bar and a list of history items. A fourth red box highlights the main content area, which displays a standardized search workflow and a list of tools.

Tools Panel

Seven subheadings

History Panel

- There are seven subheadings that link to different pages.
- The History Panel is on the right. When data is uploaded/analysis is performed, each output generates a dataset. These datasets are stored by Galaxy in the History Panel.
- All of the available tools are located in the blue Tools Panel on the left.

## 3. Create a new history. Go to the history column and click on the . A drop down menu will appear. Click on “Create New.”

The screenshot shows the Galaxy web interface with the history panel open. A red box highlights the gear icon in the top right corner of the history panel. Another red box highlights the dropdown menu that appears, which contains the following options: **Create New**, **Copy History**, **Copy Datasets**, **Share or Publish**, **Extract Workflow**, **Dataset Security**, **Resume Paused Jobs**, **Collapse Expanded Datasets**, **Unhide Hidden Datasets**, **Delete Hidden Datasets**, **Show Structure**, **Export Citations**, **Export to File**, **Delete**, and **Import from File**.

Tools Panel

Seven subheadings

History Panel

History Lists

Unnamed history

0 bytes

History Lists

Shared Histories

History Shared with Me

CURRENT HISTORY

Create New

Copy History

Copy Datasets

Share or Publish

Extract Workflow

Dataset Security

Resume Paused Jobs

Collapse Expanded Datasets

Unhide Hidden Datasets

Delete Hidden Datasets

Show Structure

Export Citations

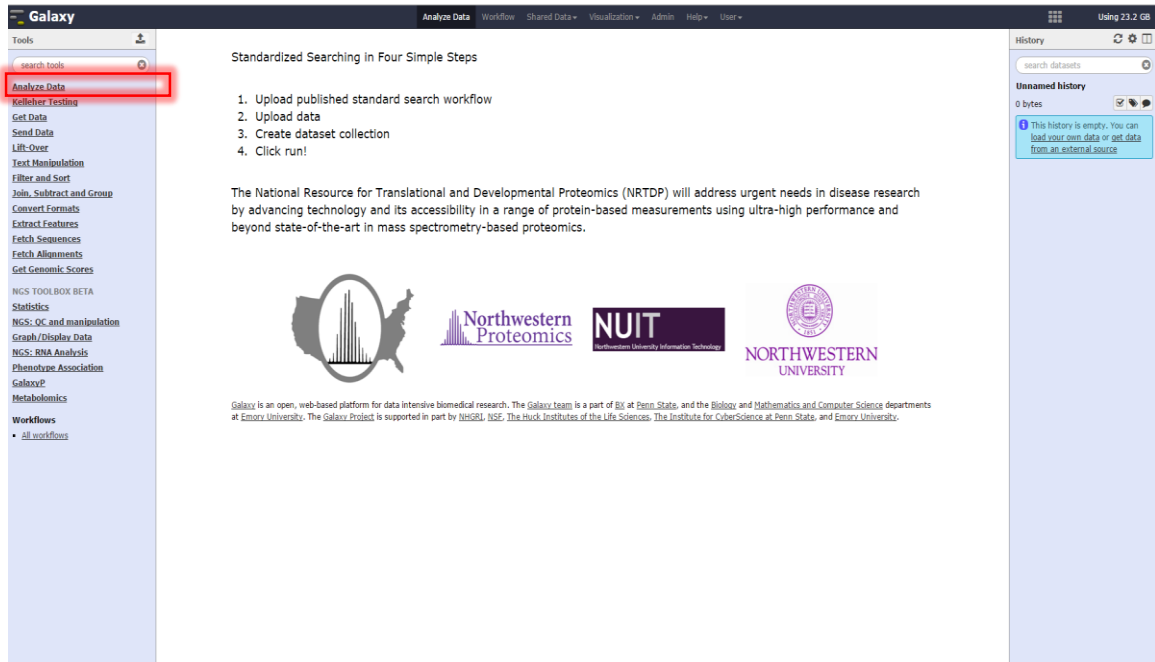
Export to File

Delete

OTHER ACTIONS

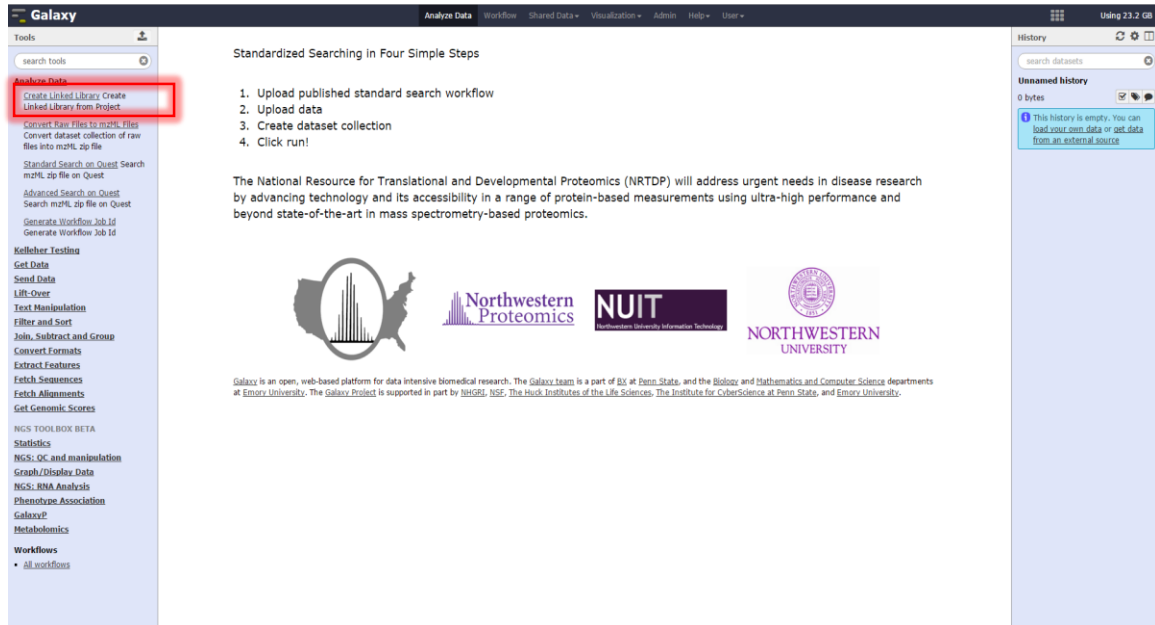
Import from File

4. Go to Tools Panel and click on “Analyze Data.” A menu containing all the available NRTDP tools will appear.



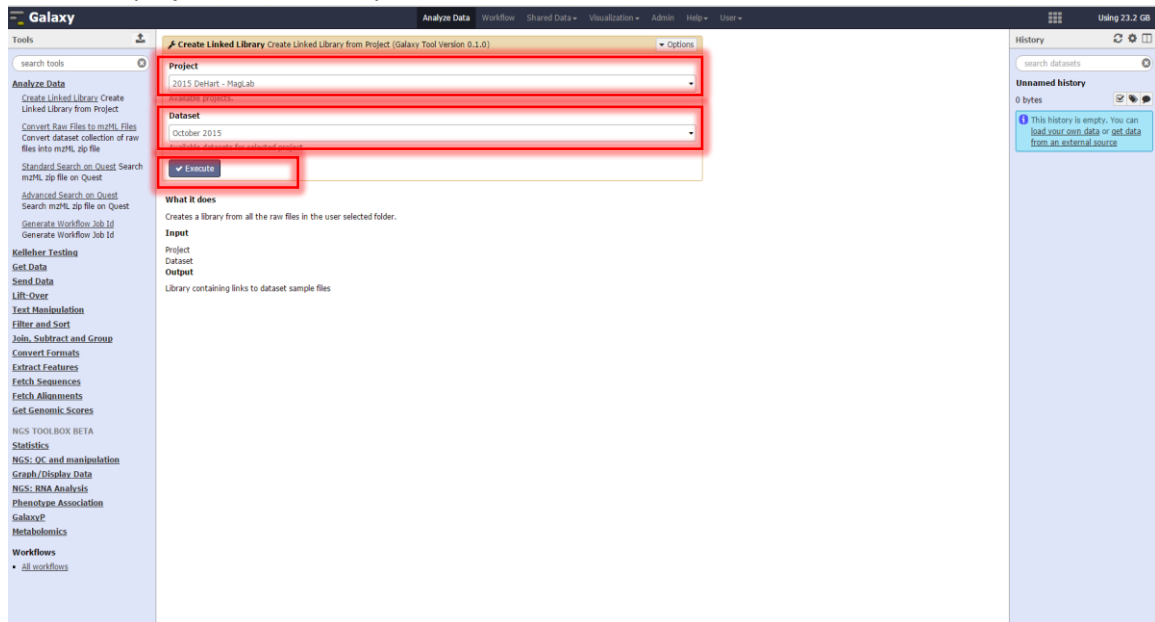
The screenshot shows the Galaxy web interface. The 'Tools' panel on the left has a search bar and a list of tool categories. The 'Analyze Data' category is highlighted with a red box. The main content area displays 'Standardized Searching in Four Simple Steps' with a numbered list: 1. Upload published standard search workflow, 2. Upload data, 3. Create dataset collection, 4. Click run!. Below this is a paragraph about the NRTDP project and logos for Northwestern Proteomics, NUIT, and Northwestern University. The right sidebar shows a 'History' panel with a search bar and a message that the history is empty.

5. Click on “Create a Linked Library.”

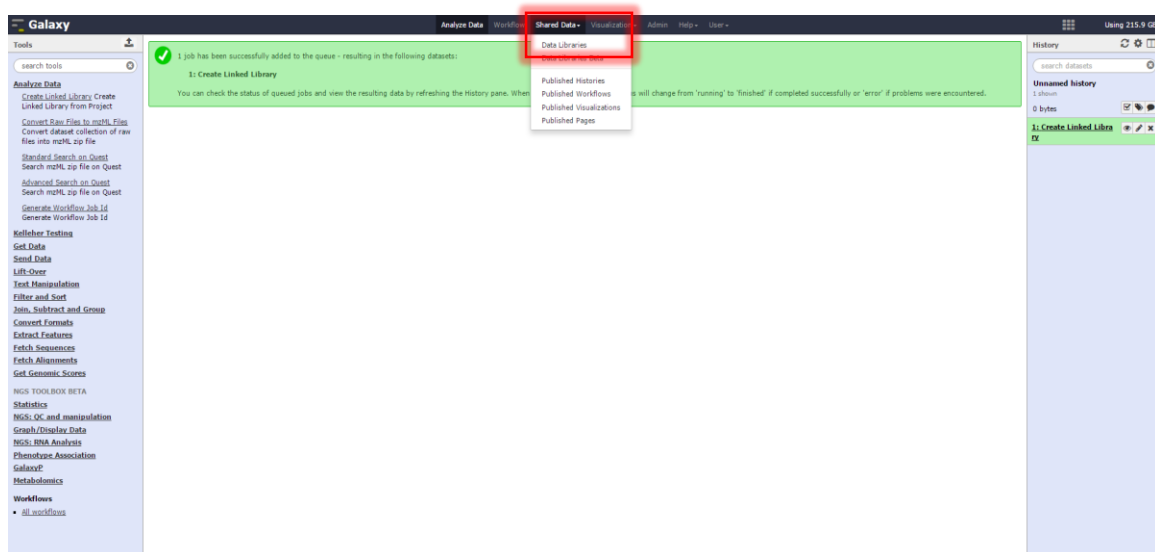


The screenshot shows the Galaxy web interface. The 'Tools' panel on the left has a search bar and a list of tool categories. The 'Analyze Data' category is highlighted with a red box, and the 'Create a Linked Library' option is highlighted within it. The main content area displays 'Standardized Searching in Four Simple Steps' with a numbered list: 1. Upload published standard search workflow, 2. Upload data, 3. Create dataset collection, 4. Click run!. Below this is a paragraph about the NRTDP project and logos for Northwestern Proteomics, NUIT, and Northwestern University. The right sidebar shows a 'History' panel with a search bar and a message that the history is empty.

6. Select the project and dataset you would like to search. Click “Execute.”



7. Click on the “Shared Data” subheading and “Data Libraries” (indicated by the red box) to view available linked library.



8. The Data Libraries screen displays available data libraries. Click on the desired data set library.

**Galaxy** Analyze Data Workflow Shared Data Visualization Admin Help User Using 23.2 GB

**Data Libraries**

search dataset name, info, message, dbkey  q

[Advanced Search](#)

Data library name	Data library description
2015 DeHart - MagLab:December 2015	

9. Once you click on the data library you will move to a screen that lists the dataset raw files. Select files to import to history by checking their checkboxes. If you would like to select all the files check the square button next to “Name” at the top of the page. At the bottom of the page there is a “For selected datasets” dropdown menu. Set the dropdown next to “Import to current history” and click “Go.”

**Galaxy** Analyze Data Workflow Shared Data Visualization Admin Help User Using 23.2 GB

**Data Library "2015 DeHart - MagLab:December 2015"** Add datasets Add folder Library Actions

<input checked="" type="checkbox"/> Name	Message	Data type	Date uploaded	File size
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f1_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:08:52 2015 (UTC)	585.8 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f2_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:09:06 2015 (UTC)	577.3 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f3_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:08:57 2015 (UTC)	657.0 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f4_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:08:49 2015 (UTC)	753.7 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f5_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:08:57 2015 (UTC)	853.1 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f6_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:08:51 2015 (UTC)	984.8 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f7_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:08:59 2015 (UTC)	958.2 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f8_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:09:06 2015 (UTC)	1.1 GB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f1_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:08:45 2015 (UTC)	760.4 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f1_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param_025q.raw		RAII	Tue Dec 8 18:08:39 2015 (UTC)	795.0 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f1_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param_360MS1.raw		RAII	Tue Dec 8 18:09:02 2015 (UTC)	730.3 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f2_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:08:44 2015 (UTC)	1.0 GB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f2_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param_025q.raw		RAII	Tue Dec 8 18:08:50 2015 (UTC)	1.1 GB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f2_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param_360MS1.raw		RAII	Tue Dec 8 18:09:07 2015 (UTC)	1021.1 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f3_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:08:53 2015 (UTC)	1.0 GB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f3_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param_025q.raw		RAII	Tue Dec 8 18:08:45 2015 (UTC)	1.1 GB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f3_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param_360MS1_attempt3.raw		RAII	Tue Dec 8 18:09:08 2015 (UTC)	997.9 MB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f4_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param.raw		RAII	Tue Dec 8 18:08:51 2015 (UTC)	1.1 GB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f4_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param_025q.raw		RAII	Tue Dec 8 18:08:08 2015 (UTC)	1.1 GB
<input type="checkbox"/> LCA_CD_DLD1paw_10per_f4_freshd13X_IPAgradientv1_1000A_Bum_PLRPS_at_param_360MS1.raw		RAII	Tue Dec 8 18:08:50 2015 (UTC)	1.1 GB

For selected datasets: Import to current history Go

Tip: You can download individual library datasets by selecting "Download this dataset" from the context menu (triangle) next to each dataset's name.

Tip: Several compression options are available for downloading multiple library datasets simultaneously:

- gzip: Recommended for fast network connections
- bzip2: Recommended for slower network connections (smaller size but takes longer to compress)
- zip: Not recommended but is provided as an option for those who cannot open the above formats

10. You will get this confirmation.

**Galaxy** Analyze Data Workflow Shared Data Visualization Admin Help User Using 23.2 GB

**Data Library "2015 DeHart - MagLab:December 2015"** Add datasets Add folder Library Actions

20 datasets imported into 1 history: Unnamed history

11. Return to home page so you can view your history by clicking on the Galaxy Icon at the top of the page. You will notice that your history has been populated with green boxes. A successful import is indicated by the selected files appearing in the history panel.

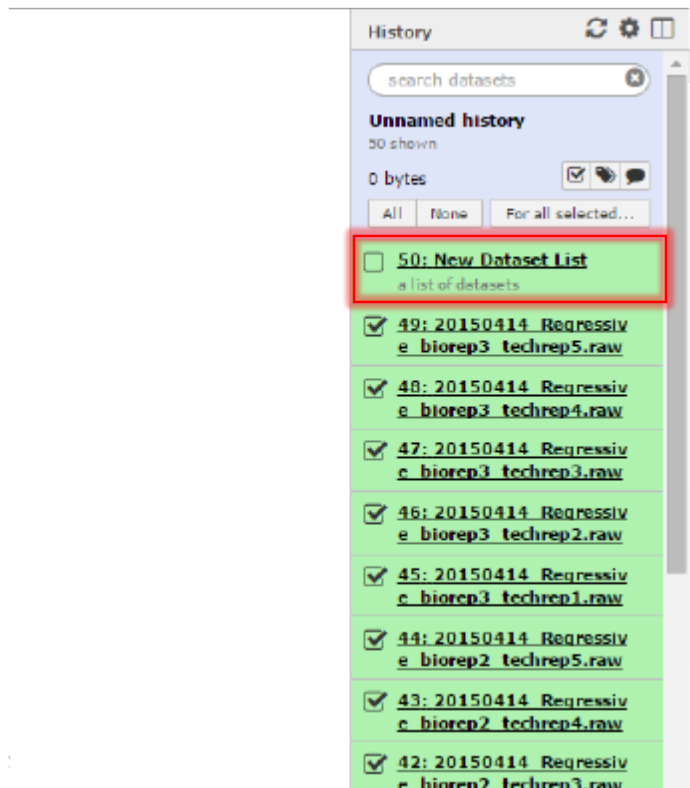
The screenshot shows the Galaxy web interface. On the left is a sidebar with navigation links like 'Tools', 'Workflows', and 'History'. The main content area displays 'Standardized Searching in Four Simple Steps' with a list of instructions: 1. Upload published standard search workflow, 2. Upload data, 3. Create dataset collection, 4. Click run! Below this is a paragraph about the National Resource for Translational and Developmental Proteomics (NRTDP) and logos for Northwestern Proteomics, NUI, and Northwestern University. On the right, a 'History' panel is visible, showing a list of datasets with checkboxes and icons. The datasets are listed in a table format with columns for dataset ID, name, and size.

12. Next create the dataset list. To do this, move to the History Panel.

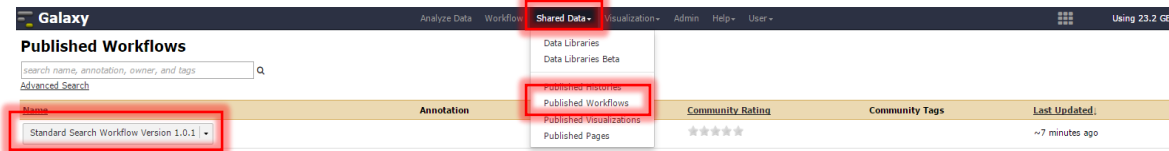
- a. Click on the checkbox.
- b. Check all the imported raw files
- c. Click For All Selected
- d. Select Build Dataset List

The screenshot shows the 'History' panel in Galaxy. It displays a list of datasets with checkboxes. The 'All' checkbox is selected, and the 'For all selected...' button is highlighted. Below the list, the 'Build Dataset List' option is selected in the context menu. The datasets listed include 'e biorep3 techrep2.raw', '45: 20150414 Regressiv e biorep3 techrep1.raw', and '44: 20150414 Regressiv e biorep2 techrep5.raw'.

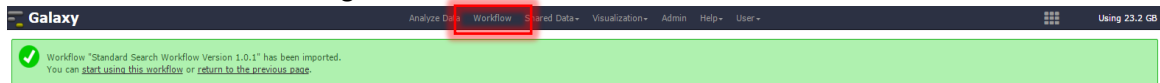
13. A new dataset collection will appear in the History Panel



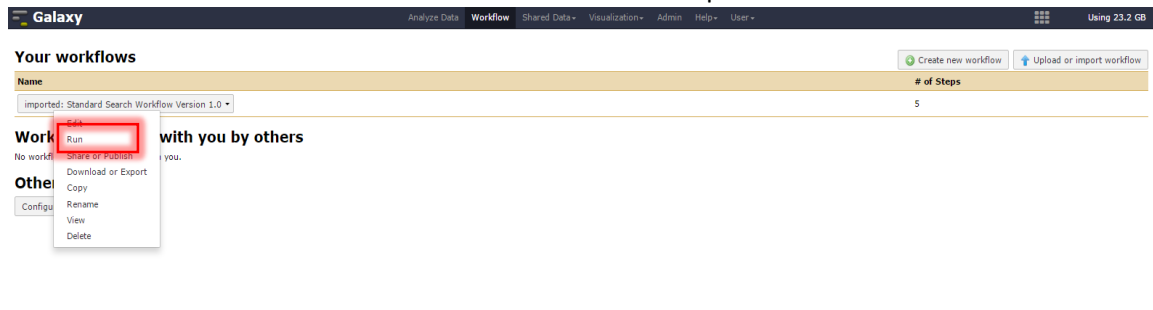
14. Navigate to Shared Data. Click on “Published Workflows” in the drop down menu. Import Published Standard Search Workflow 1.1.0 if you do not already have it by clicking on the down arrow of the “Standard Search Workflow Version 1.1.0” tab.



15. Click “Workflow” subheading.

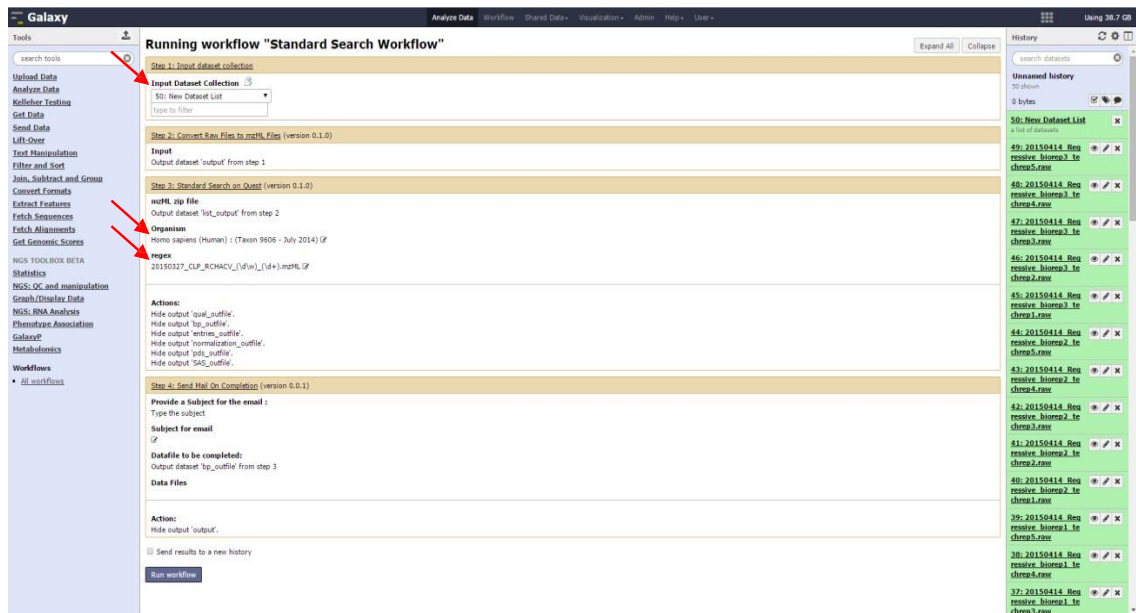


16. Click “Run” in Standard Search Workflow Version 1.1.0 dropdown.



17. Input search parameters and click “Run Workflow.”

- Select Dataset list just created.
- Select Organism to be searched.
- Insert regex provided by an administrator.
- Resolution
- Leave Build Version as Production

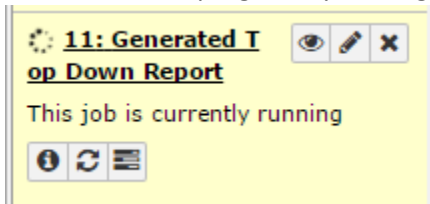


18. Search will be added to History Panel.





19. Monitor Search progress by clicking on check progress button in History Panel item.



20. Page refreshes every 30 seconds.

High Throughput Galaxy Search Using Quest

Date: 2015-08-26  
User: joseph.green@northwestern.edu  
Last Raw File Searched: 2015\_08\_09\_F15\_tech3

Search Progress

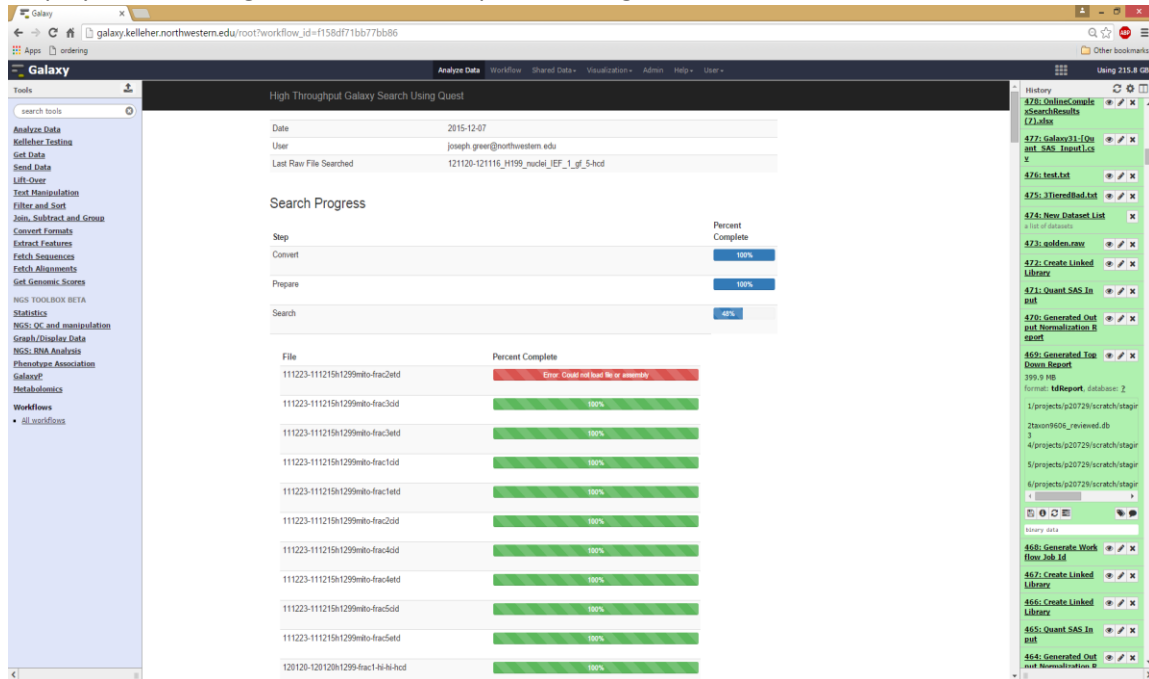
Step	Percent Complete
Convert	100%
Prepare	100%
Search	97%

File	Percent Complete
2015_08_09_F11_tech1	100%
2015_08_09_F11_tech2	100%
2015_08_09_F11_tech3	100%
2015_08_09_F12_tech1	100%
2015_08_09_F12_tech2	80%
2015_08_09_F12_tech3	60%
2015_08_09_F13_tech1	100%
2015_08_09_F13_tech2	100%
2015_08_09_F13_tech3	100%
2015_08_09_F14_tech1	100%
2015_08_09_F14_tech2	100%

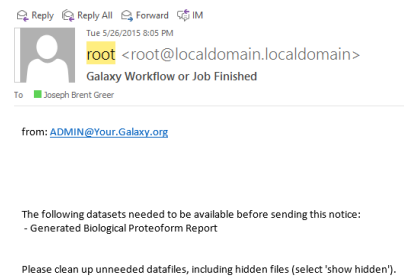
History

- 62: Duplicate List
- 61: New Dataset List
- 60: 20151110\_Hs732\_Gr 5.3.2mm
- 59: 20151110\_Hs732\_Gr 5.3.2mm
- 58: 20151110\_Hs732\_Gr 5.2.2mm
- 57: 20151110\_Hs732\_Gr 5.1.2mm
- 56: 20151110\_Hs732\_Gr 4.3.2mm
- 55: 20151110\_Hs732\_Gr 4.2.2mm
- 54: 20151110\_Hs732\_Gr 4.1.2mm
- 53: Generate Workflow Job 14
- 52: Create Linked Library
- 51: New Dataset List
- 50: 20151017\_T031\_16\_3 19\_EBA22\_Tech04\_Bio10 2.2mm
- 49: 20151017\_T032\_16\_3 19\_CS786.63\_Tech02\_Bio5 8.2mm
- 48: 20151017\_T031\_16\_3 19\_CS786.63\_Tech02\_Bio1 38.2mm
- 47: 20151017\_T030\_16\_3 19\_EBA22\_Tech02\_Bio10 2.2mm
- 46: 20151017\_T028\_16\_3 19\_EBA22\_Tech04\_Bio16

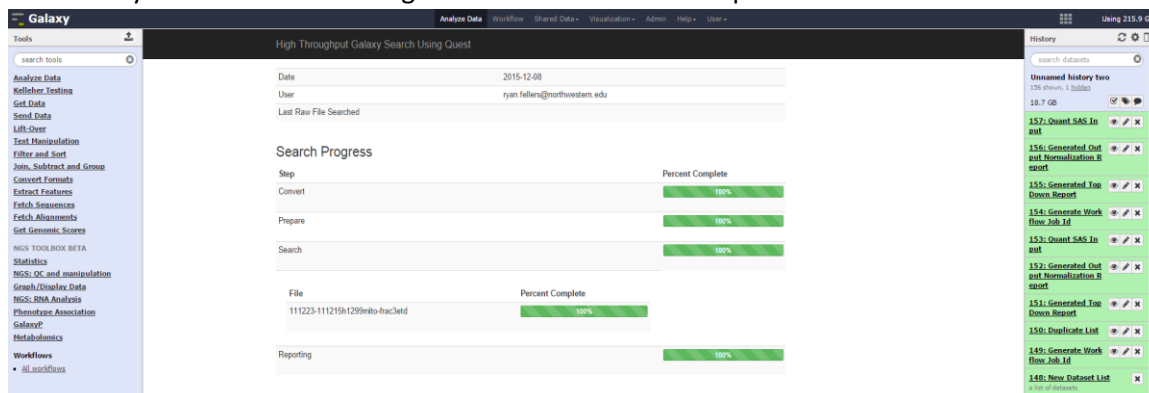
21. Display errors during search with descriptive message.



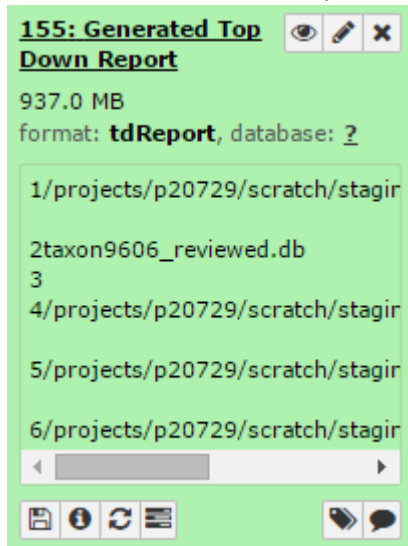
22. You will receive an email when the search completes.



23. The History Panel items will turn green when the search completes.



24. Download search results by clicking “download” button.



25. Open results in TD Viewer (<http://proteinaceous-apps.azurewebsites.net/TopDownViewer>)