Tripal Galaxy

Execution of Scientific Workflows for Tripal-Based Community Databases

Shawna Spoor, Connor Wytko, Ming Chen, Abdullah Almsaeed, Bradford Condon, Heidi Hough, Nic Herndon, Nick Mills, Margaret Staton, Jill Wegrzyn, Alex Feltus, Dorrie Main, Stephen P. Ficklin
Galaxy is an open source online platform for scientific data integration and analysis.

- Makes computational biology accessible to everyone
- Allows for multi-step computational analyses
  - simple linear process
  - browser based user interface
The Tripal Galaxy module allows Tripal-based sites to use the Galaxy project. That means:

- Analytical tools for users within your website
  - Consistent user interface
  - The power of high performance computing on your website
The module can be used in two ways:

- Using a graphical user interface
The module can be used in two ways:

- Through the application programming interface (API)
Tripal-Galaxy Docs

Tripal Galaxy Extension Module

- Install Tripal Galaxy
  - Installation
  - Set Permissions
- Option 1: Automatic Webform Creation
  - Site-wide Files
  - Adding a Remote Galaxy Server
  - Adding a New Workflow
  - Set User Quotas
- Option 2: Power an Application
- Workflow Settings
  - Customizing Emails
- Monitoring Usage
  - The Job Queue
  - Usage Stats
- Additional Settings
- Executing Jobs and Tasks
  - Automatic Job Execution
  - Automatic Job Status Checking
  - Automatic Housekeeping
  - Manually Execute Jobs
- Tripal Galaxy API
  - Tripal Galaxy API Overview
    - Adding a Galaxy Server
    - Get a List of Galaxies
    - Connect to Galaxy
    - Test if a Galaxy server is accessible.
    - Add a Workflow to Tripal
    - Prepare a Workflow for Execution
    - Get a History Name
    - Create a History and Upload Files
    - Invoke a Workflow
    - Check the status of a workflow submission
    - Retrieving Results from Galaxy

https://tripal-galaxy.readthedocs.io/en/latest/
Tripal Galaxy API Overview

Here code examples and brief descriptions are provided for common applications of the Tripal Galaxy API. Not all arguments for all API functions are thoroughly described. Please see the Tripal Galaxy API Functions page for specific details.

Adding a Galaxy Server

The Tripal Galaxy module allows a site administrator to add a new Galaxy Server to Tripal using an online web interface. However, you can programmatically add a new Galaxy server using the `tripal_galaxy_add_galaxy` function. For example:

```php
// Get the current user.
global $user;

// Prepare the information needed for the server.
$data = array(
    'servername' => 'Use Galaxy Public',
    'description' => 'The public server provided by the Galaxy Project',
    'url' => 'https://usegalaxy.org/',
    'username' => 'fakeuser',
    'api_key' => '48e6d1213e5dd6d187e6811b38b6de51',
    'uid' => $user->uid,
);

// Add the Galaxy server.
$galaxy = tripal_galaxy_add_galaxy($data);
```

A use case when you may want to programmatically add a new Galaxy server is if you want to provide your own interface for allowing others to add new servers besides the site admin.

Get a List of Galaxies

You can retrieve a list of all Galaxy servers known to Tripal by calling the `tripal_galaxy_get_galaxies` function.

```php
// Get the list of galaxies.
$galaxies = tripal_galaxy_get_galaxies();

// Iterate through that list to perform actions on each, or retrieve
// information.
foreach ($galaxies as $galaxy) {
    // Get the galaxy ID
    $galaxy_id = $galaxy->galaxy_id;

    // Do stuff here.
    }
```
The blend4php library was developed as part of the Tripal Gateway Project as stand-alone PHP bindings to the Galaxy RESTful API. It allows any PHP-based web application to communicate with a remote Galaxy instance.

https://github.com/galaxyproject/blend4php
Module installation can be done using:

- **Drush**: `drush pm-enable tripal_galaxy`
- **GUI**: `admin/modules/install`
- **Git clone**:
  ```
  git clone git@github.com:tripal/tripal_galaxy.git
  ```
Using Tripal Galaxy: Setting it up
Using Tripal Galaxy: Adding a Galaxy Instance

Galaxy Server Name *

Galaxy Server

Please provide the name of the remote Galaxy Server

Description

A Galaxy server on my local machine.

URL *

localhost:8080

The URL for the remote Galaxy server.

User Name

shawna

The user name for the Galaxy server. This username is used to launch all jobs by default. If this field is left blank then it is expected that the user has an account on the Galaxy server and will provide their username when executing workflows.

API Key

82171a8d96f6cc8c8e0e420c8f0617e0

The API key for the user name specified above. If this field is left blank then it is expected that the user will provide their own API key when submitting a job.

Submit  Test Connection
Using Tripal Galaxy: Adding a Workflow

### Add Workflows

<table>
<thead>
<tr>
<th>ID</th>
<th>WORKFLOW NAME</th>
<th>GALAXY SERVER</th>
<th>STATUS</th>
<th>ACTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are no workflows

---

Select Galaxy Server

---

<table>
<thead>
<tr>
<th>ADD WORKFLOW</th>
<th>WORKFLOW NAME</th>
<th>WORKFLOW ID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>differential-expression-deseq2-analysis (imported from uploaded file)</td>
<td>0a24b1f62a0cc04</td>
</tr>
<tr>
<td></td>
<td>wgcmc-analysis</td>
<td>1cd8e26b131e891</td>
</tr>
<tr>
<td></td>
<td>OrthoQuery_proteome (imported from uploaded file)</td>
<td>df7a10c02a5b08e</td>
</tr>
<tr>
<td></td>
<td>Test (imported from uploaded file)</td>
<td>5969b1f201f12ae</td>
</tr>
<tr>
<td></td>
<td>genome-annotation (imported from uploaded file)</td>
<td>597429621d6eb2b</td>
</tr>
</tbody>
</table>

Submit
The workflow was successfully added!

Sometimes these workflows need a bit of tweaking to make them more intuitive for users. Click the title of the workflow below to customize how it will appear to users. Consider improving:

- The workflow title.
- The instructions for the workflow.
- The "About this step" field for each tool to help the user understand the workflow.
- Each tool setting descriptions as desired.
## Using Tripal Galaxy: Job Queue

### Analysis Results

![Image of job queue interface](image)

**Workflow**

<table>
<thead>
<tr>
<th>ID</th>
<th>WORKFLOW</th>
<th>GALAXY SERVER</th>
<th>USER</th>
<th>SUBMISSION TIME</th>
<th>START TIME</th>
<th>END TIME</th>
<th>STATUS</th>
<th>RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>wgcna-analysis</td>
<td>Galaxy Server</td>
<td>administrator</td>
<td>10/14/2018 - 08:46</td>
<td>10/14/2018 - 08:46</td>
<td>10/14/2018 - 08:46</td>
<td>Completed</td>
<td>View</td>
</tr>
<tr>
<td>52</td>
<td>wgcna-analysis</td>
<td>Galaxy Server</td>
<td>administrator</td>
<td>10/14/2018 - 08:43</td>
<td>10/14/2018 - 08:43</td>
<td>10/14/2018 - 08:44</td>
<td>Error</td>
<td>View</td>
</tr>
</tbody>
</table>

**User**

<table>
<thead>
<tr>
<th>ID</th>
<th>WORKFLOW</th>
<th>GALAXY SERVER</th>
<th>USER</th>
<th>SUBMISSION TIME</th>
<th>START TIME</th>
<th>END TIME</th>
<th>STATUS</th>
<th>RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>wgcna-analysis</td>
<td>Galaxy Server</td>
<td>administrator</td>
<td>10/14/2018 - 08:46</td>
<td>10/14/2018 - 08:46</td>
<td>10/14/2018 - 08:46</td>
<td>Completed</td>
<td>View</td>
</tr>
<tr>
<td>52</td>
<td>wgcna-analysis</td>
<td>Galaxy Server</td>
<td>administrator</td>
<td>10/14/2018 - 08:43</td>
<td>10/14/2018 - 08:43</td>
<td>10/14/2018 - 08:44</td>
<td>Error</td>
<td>View</td>
</tr>
</tbody>
</table>
Using Tripal Galaxy: Usage Stats

**Top 10 Submitters**
The following bar chart shows the top 10 users that have submitted workflows.

- **administrator**: 6
- **john.doe**: 1
- **Penelope Van Strap**: 1

**Top 10 Workflows Submitted**
The following bar chart shows the top 10 workflows that have been submitted.

- **Differential Gene Expression (DESeq)**: 1
- **mapping-hisat2-single-end**: 1
- **Expression Network Construction (WGCNA)**: 6
Using Tripal Galaxy: Files

Upload files and make them available to all users.
Co-Expression Network Construction (WGCNA)

Step 1: Instructions

Gene Co-Expression Networks GCNs are often used to help identify groups of genes that are co-functional. The analysis requires a file containing a Gene Expression Matrix (GEM) containing gene expression levels derived from either RNA-seq or microarray data sets. Genes that tend to be expressed at similar levels across a set of samples are considered co-expressed. A correlation value (often Pearson's or Spearman's correlation) is assigned to each gene pair. WGNCA is an R-based package that performs pair-wise gene correlations and uses a method referred to as soft-thresholding to cull pair-wise correlation values that are non-significant. To learn more about WGCNA and co-expression network construction please see the WGCNA homepage.

EditExecution of this workflow is supported by Galaxy.

This workflow uses the following tools:

- **WGCNA: preprocessing (version 1.0.2)**: Data cleaning and preprocessing.
- **WGCNA: construct network (version 1.0.2)**: Construct gene network.
- **WGCNA: eigengene visualization (version 1.0.2)**: Eigengene visualization.
Using Tripal Galaxy: Running a Workflow

Description

At this step the Gene Expression Matrix (GEM) is provided. It is a comma-separated (CSV) format file where each line of the file represents a gene and each "column" represents a single RNA-seq sample. The data consists of expression level quantification values.

Step 2: Data File

Data File *

Please provide a data file. You may either select an existing file or upload a new file. Note: you have currently used 4.71MB of your 100GB quota.

User Uploaded Files

Select Your File

expression_data2.csv (4.71MB)

Please select your file if it already exists on this site. Only files of the following types are listed: csv.

Upload a New File

File Upload

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>Upload Progress</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choose File</td>
<td>No file chosen</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If your files do not already exist above please upload as many pair of files as desired.
Hello, Galaxy is running!

To customize this page edit static/welcome.html

Configuring Galaxy » Installing Tools »

Take an interactive tour: Galaxy UI History Scratchbook

Galaxy is an open platform for supporting data intensive research. Galaxy is developed by The Galaxy Team with the support of many contributors.

The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.
Analysis Results

Analysis Name
Co-Expression Network Construction (WGCNA)

- Submission Details

Results
All files associated with the analysis, including input files and all output files are listed below. If an error occurred during creation of any file it will be noted. Expand the field for the result below to view the status, peek or download result files.

1: expression_data2.csv (ok)
2: trait_data2.csv (ok)
3: WGCNA: preprocessing (ok)
4: R workspace: WGCNA preprocessing (ok)
5: WGCNA: construct_network (ok)
6: R workspace: WGCNA construct_network (ok)
7: WGCNA: eigengene visualization (ok)

WGCNA: eigengene visualization

Download Results
WGCNA: eigengene visualization (1.43MB)

View Results
WGCNA: eigengene visualization
Note: It may take a few moments for the results page to load.
Results Viewer

Viewing results for WGCNA: eigengene visualization

Network heatmap
Tripal Galaxy in use: Hardwood Genomics

https://www.hardwoodgenomics.org/

https://github.com/statonlab/galaxy-workflows

https://github.com/statonlab/aurora-galaxy-tools
Forest trees are long-lived and immobile individuals that serve as ideal models to assess population structure and adaptation to the environment. Despite the availability of comprehensive data, the researchers who study them are challenged to integrate data describing genotype, phenotype, and the environment. Towards this goal, the web application Cartogratree was designed and implemented as an open repository and open-source analytic web-based framework for all three.
Tripal Galaxy in use: Cartogratreec

Your contact details are connected to your user account. Please use the following form to update or modify your profile details. These details are viewable to the general public.

Edit details - View details

The following table provides a list of analytical workflows that you have submitted for execution on this site. Click the drop down in the Actions column to view results, view your submission, re-run or delete it.

<table>
<thead>
<tr>
<th>Submission Status</th>
<th>Workflow Name</th>
<th>Submission Time</th>
<th>Start Time</th>
<th>End Time</th>
<th>Status</th>
<th>Actions</th>
</tr>
</thead>
</table>
NSF Award #1443040

Further Information:
Tripal Galaxy Docs:
http://tripal.info/tutorials/v3.x/galaxy

Tripal Galaxy github:
https://github.com/tripal/tripal_galaxy

Galaxy:
https://galaxyproject.org/

Tripal:
http://tripal.info/

Hardwood Genomics:
https://www.hardwoodgenomics.org/

TreeGenes:
https://treenomesdb.org