Introduction to G-OnRamp Walkthrough

Yating Liu

07/2016

# 1. Introduction

The overall goal of this walkthrough is to illustrate the use of the G-OnRamp workflow to create a genome browser. The walkthrough will cover topics including how to create an account and login, how to import the G-OnRamp workflow into your workspace, how to modify the workflow by changing tool parameters and adding new tools, and how to run the workflow. This walkthrough assumes that the reader has some basic familiarity with Galaxy. (See the “Overview of Galaxy” presentation for a detailed introduction to Galaxy). Note that all of the screenshots in this walkthrough are based on the Galaxy instance running on the GWU public server at <https://galaxy.colonialone.gwu.edu/new_galaxy/>. The G-OnRamp instance that we will use to do analysis is running on the Cloud. There might be some minor differences between the user interface of the GWU Galaxy server and the G-OnRamp instance.

The links to the G-OnRamp instance and the G-OnRamp workflow that we will use in this walkthrough are available in the “G-OnRamp\_links.txt” file. You can find this file in the “Intro\_G-OnRamp\_walkthrough” folder on your Desktop.

# 2. Register and login

## 2.1 Create an account

Open the “G-OnRamp\_links.txt” file inside the “Intro\_G-OnRamp\_walkthrough” folder on your Desktop. Copy the web address next to the “G-OnRamp” entry into the address bar of your web browser and then press “Enter” to navigate to the G-OnRamp instance.

First of all, you need to create and log into your Galaxy account in order to access the full functionality of Galaxy. To create an account, click on “User” at the menu bar and then click on “Register” (Figure 1).

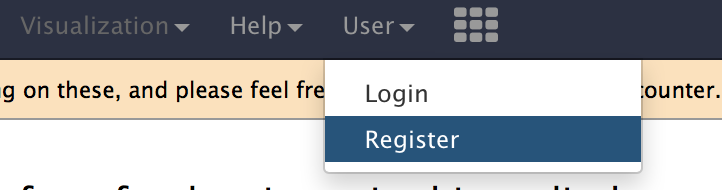


Figure : Create a user account on the G-OnRamp Galaxy instance

The workspace will display the registration form (Figure 2). Besides entering your email address and your password, you will also need to specify a public name. The public name is an identifier that will be assigned to all the Galaxy resources (e.g., histories, workflows) that you share with other users. The public name must be at least three characters in length and contains only lower-case letters, numbers, and the '-' character. Click on the “Submit” button to create your account. If the new account has been created successfully, you will be logged into your new account and will be redirected to the home page.

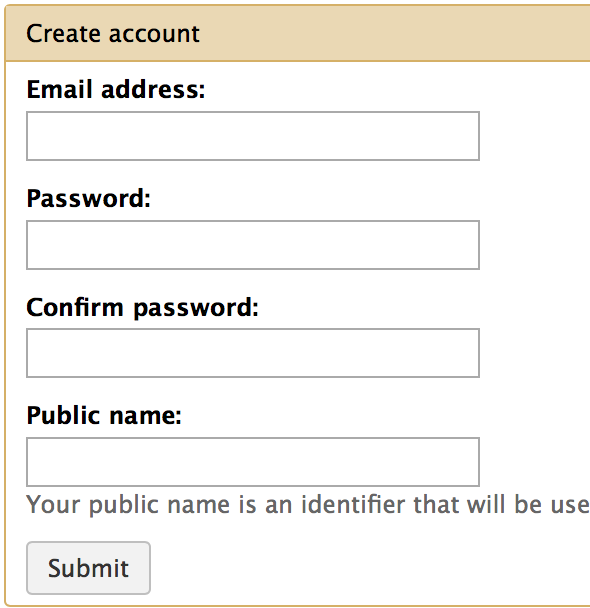


Figure : Registration form for creating a new Galaxy account

## 2.2 Login

When you return to G-OnRamp later, you can log in with your email address and password. First click on “User” at the menu bar, and then choose “Login” (Figure 3).

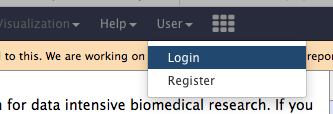


Figure : Open the login form from the “User” menu at the main menu bar

Then enter your email address and password (Figure 4).

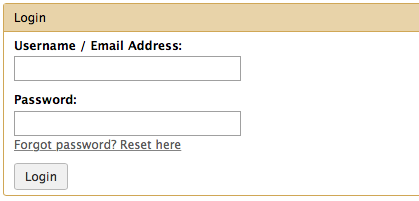


Figure : Enter the email address that you used to create your Galaxy account and your account password to log into Galaxy

# 3. Import G-OnRamp workflow

The G-OnRamp workflow has been shared with you via a link named “G-OnRamp workflow” in the “G-OnRamp\_links.txt” file. Copy the link into the address bar of your web browser and press “Enter” to navigate to the workflow page. Open the link and click on the green “+” button at the top right corner to import the workflow into your Galaxy account (Figure 5).

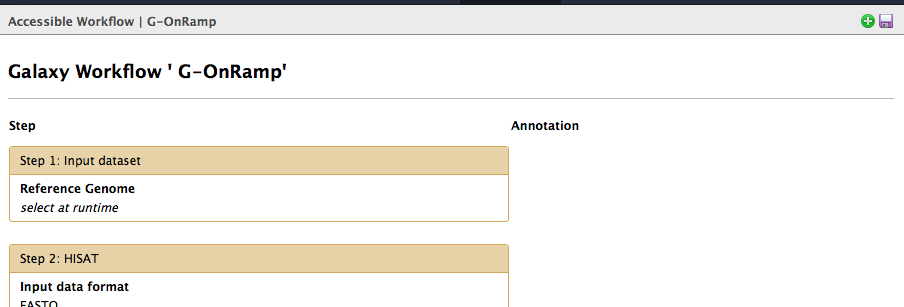


Figure : Click on the green “+” icon to import the “G-OnRamp” workflow into your workspace

You can access the imported workflow via the “Workflow” menu item on the menu bar. The “imported: G-OnRamp” workflow will appear under the “Your workflows” section. When you right click (control-click on the Mac) on the workflow or click on the down arrow, a context-sensitive menu will appear where you can edit, run, share or download, copy, rename, view, and delete the workflow (Figure 6).

To help us keep track of the different workflows, we will change the name of the workflow to “G-OnRamp: D. biarmipes F element”. Right click on the “Imported:G-OnRamp” workflow and then select “Rename”.

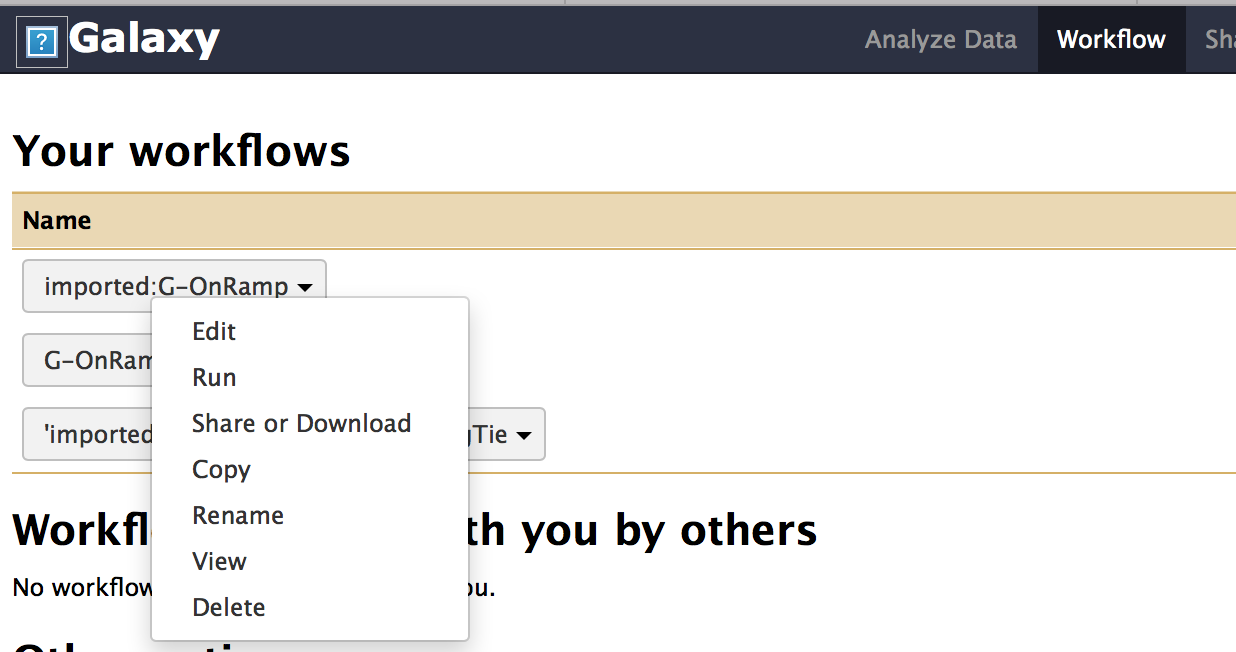


Figure : Context-sensitive menu that specifies the different ways that you can manipulate the imported workflow

Enter “**G-OnRamp: D. biarmipes F element**” into the “Workflow Name” field and then click on “Rename” to change the name of the workflow (Figure 7).

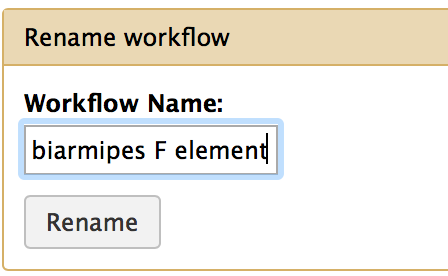


Figure : Rename the workflow

# 3. Run the G-OnRamp workflow

## 3.1 Upload your datasets

The test datasets that we will use in this walkthrough are available in the “Datasets” folder within the Intro\_G-OnRamp folder on your Desktop. There are four test datasets in the folder: the reference genome sequence from the *Drosophila biarmipes* Muller F element, a collection of *Drosophila melanogaster* protein sequences, and the forward and reverse paired-end reads from contig16 of *D. biarmipes* RNA-Seq data. Click on the “Get Data” link in the Tools panel and then click on the “Upload File” link to upload all the datasets to your Galaxy account (Figure 8).

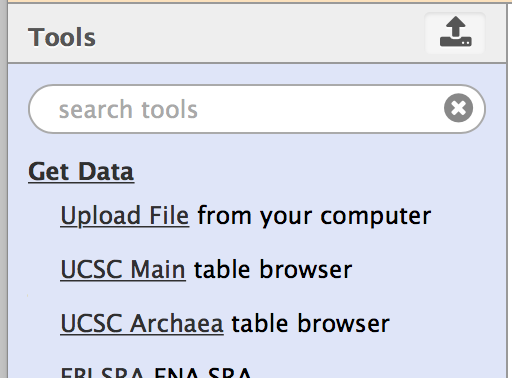


Figure : Click on the "Get Data" and "Upload File" in the Tools panel

The Upload File dialog box will appear in your workspace. Click on “Choose local file” and then navigate to the “Datasets” folder. Press and hold the “Shift” key while clicking on all the files in this folder and then click on the “Choose” button (Figure 9).

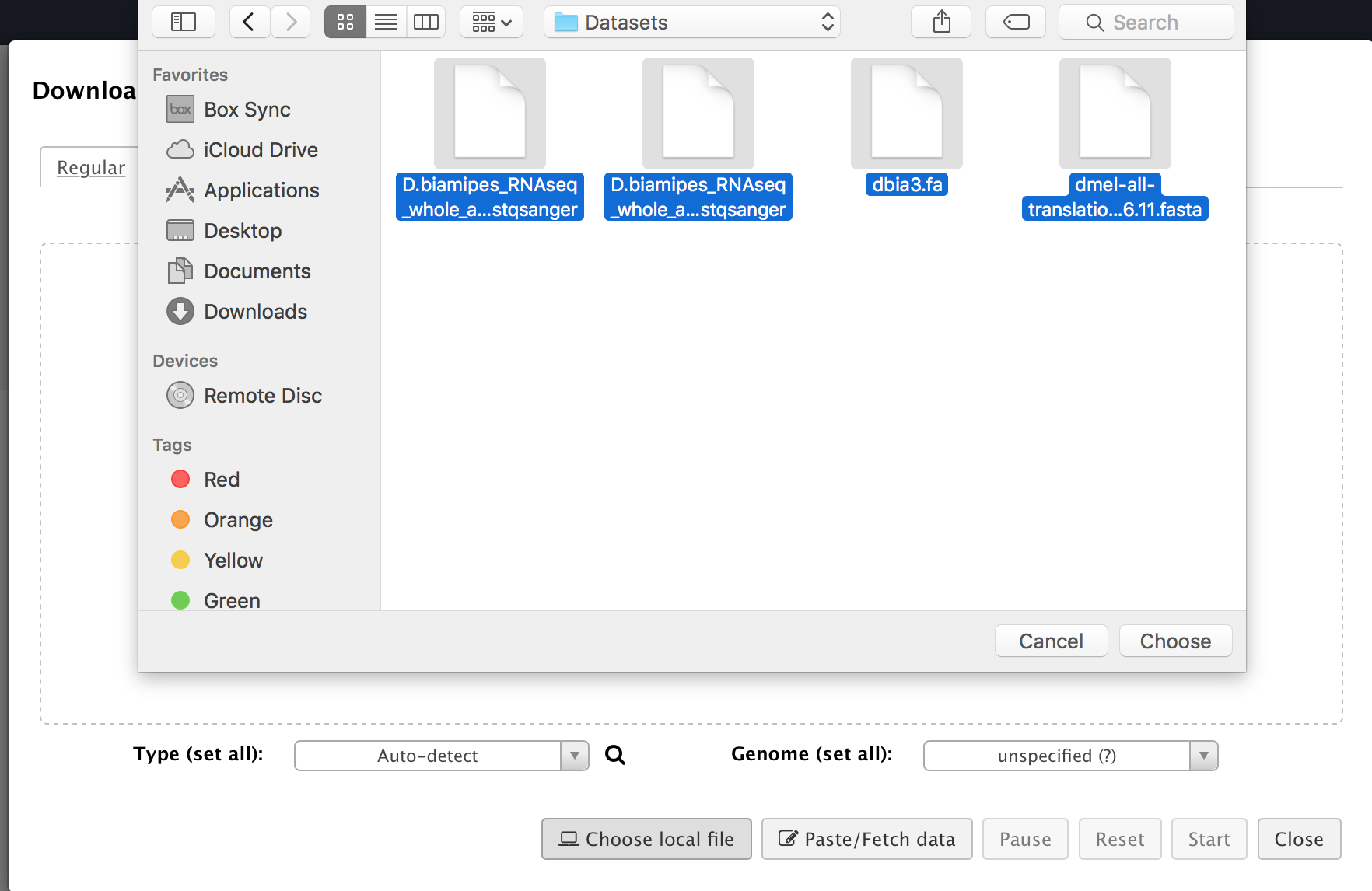


Figure : Select all the files in the "Datasets" folder

The four files will appear in the upload dialog box. Click on the “Start” button to start uploading these files to your History (Figure 10). Click on the “Close” button to dismiss the upload file dialog. The files will be added to your History and you will need to wait for the files to be uploaded to the G-OnRamp server.

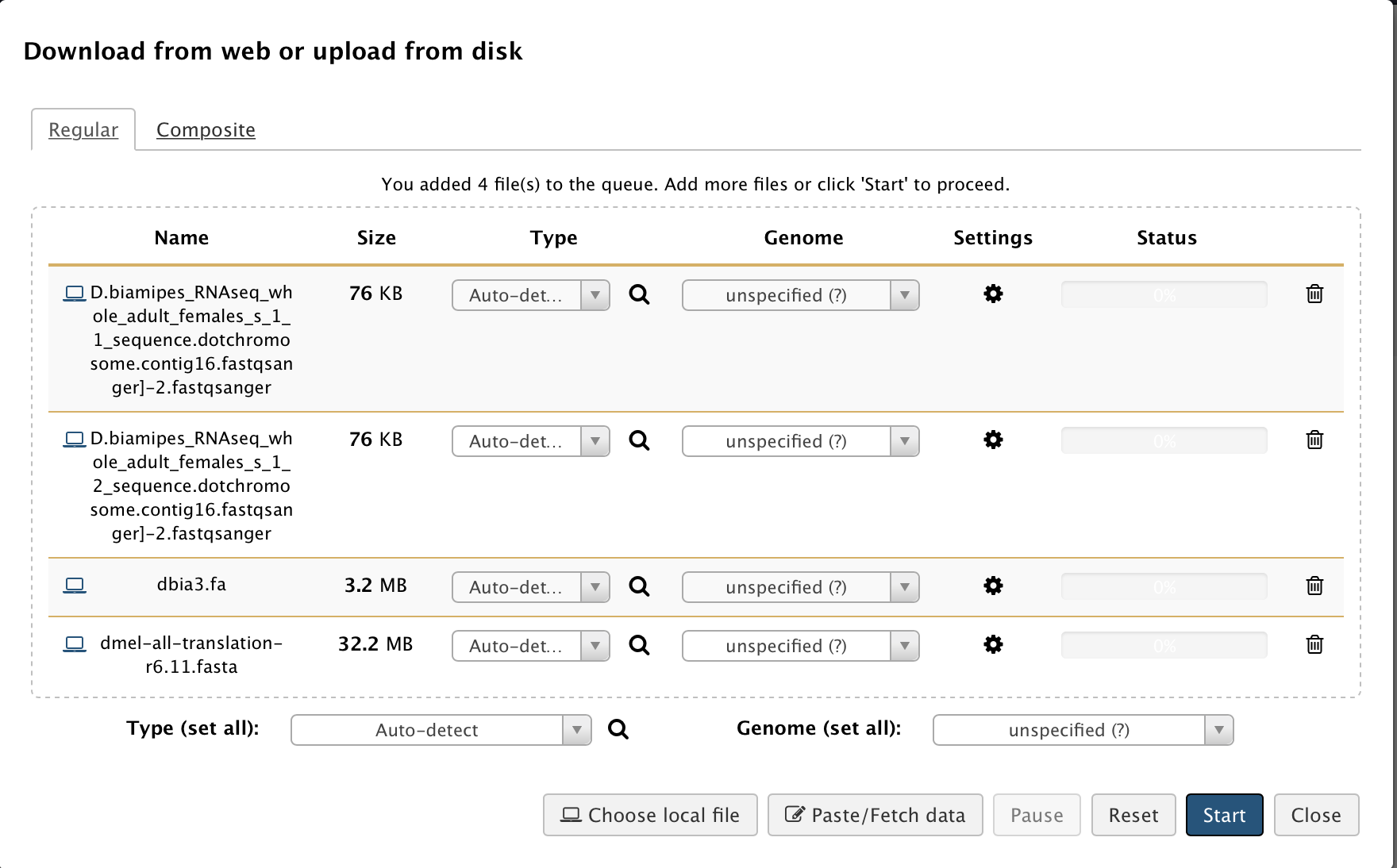


Figure : Click on the "Start" button to begin the file upload

You can review the status of the file upload in your History panel (Figure 11). The grey color indicates that the files are waiting in the queue. The yellow color indicates that the files are in the process of being uploaded. The green color indicates that the files have been uploaded to the server.

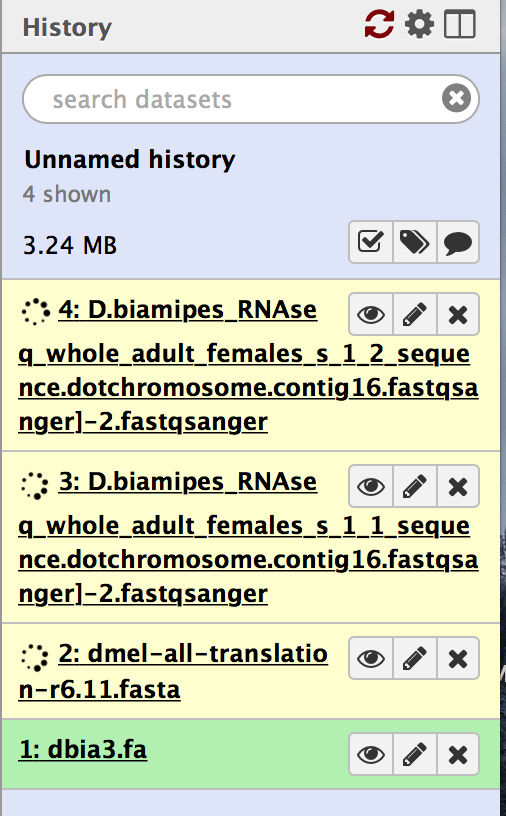


Figure : The colors in the History panel reflect the file upload status

## 3.2 Run the workflow

Wait until all four items in the History panel turn green. Once all of your datasets have been uploaded to the server, you can use the G-OnRamp workflow to analyze your datasets. However, before running the G-OnRamp workflow, you need to verify that Galaxy has assigned the correct data format to each of the file. In this case, Galaxy incorrectly assigned the “dmel-all-translation-r6.11.fasta” file to the sdf format instead of the fasta format. Therefore, you need to manually edit its datatype by clicking on “pencil” icon at the top right corner of this dataset (Figure 12).

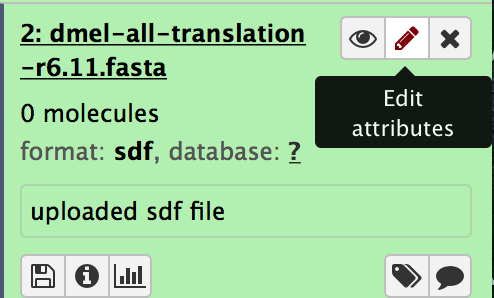


Figure : Edit the attributes of the dataset by clicking on the "pencil" button

Click on the Datatype tab, click on the down arrow under the “New Type” field, and then select “fasta” (Figure 13). Don’t forget to click on the “Save” button (Figure 14).

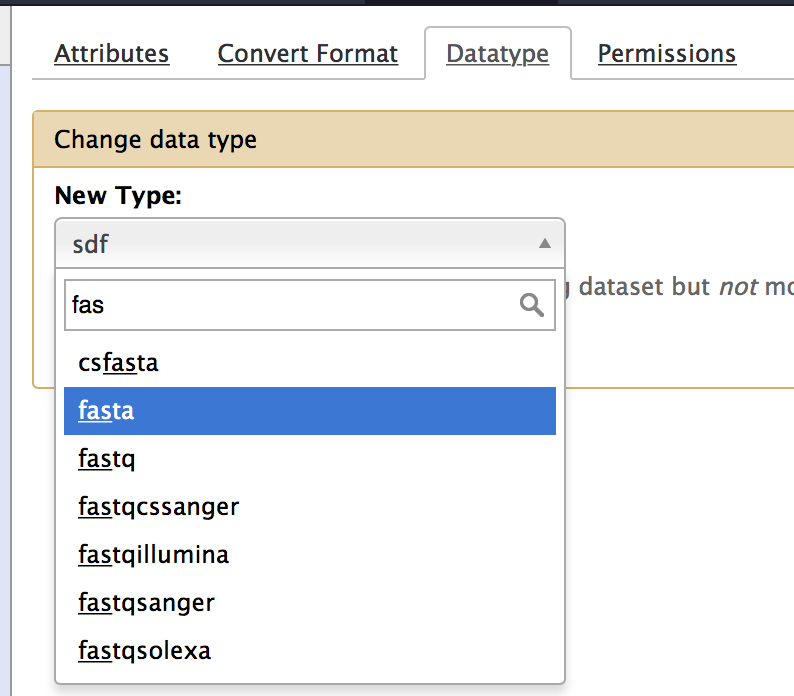


Figure : Change the datatype of the dmel-all-translation-r6.11.fasta dataset to fasta

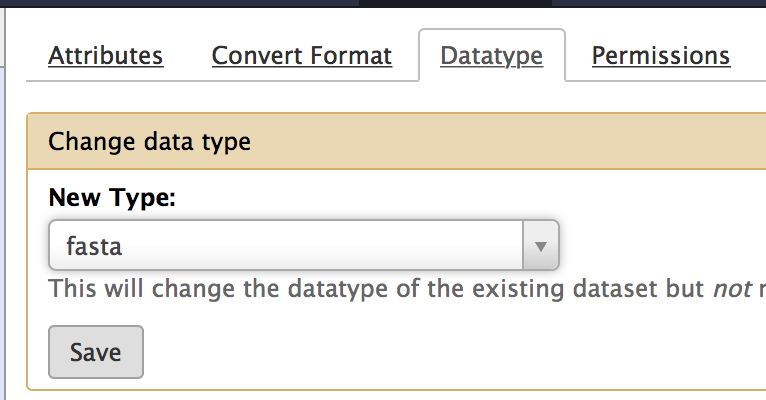


Figure : Save the new datatype for the dmel-all-translation-r6.11.fasta dataset

Click on the “Workflow” menu on the menu bar. Right click on the “G-OnRamp:D. biarmipes F element” workflow and then select “Run” from the drop-down menu to run the workflow (Figure 15).

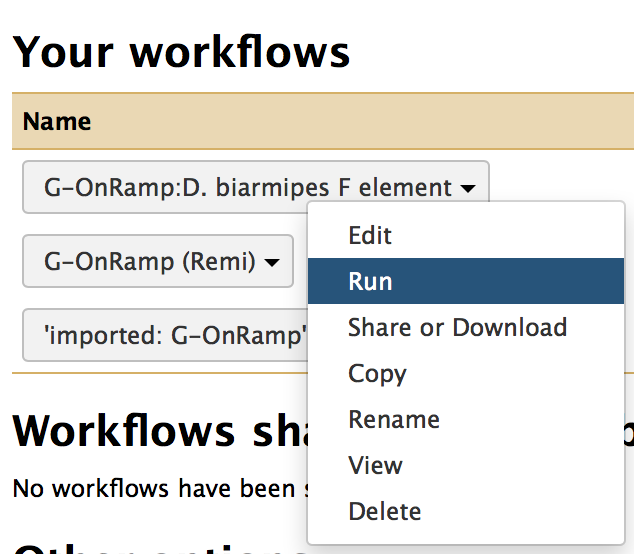


Figure : Click on the drop-down menu for the “G-OnRamp: D. biarmipes F element” workflow to run the workflow

The workflow consists of 11 different steps. In order to run the workflow, you need to specify the datasets that should be used in Step 1 (Figure 16), Step 4 (Figure 17), and Step 6 (Figure 18).

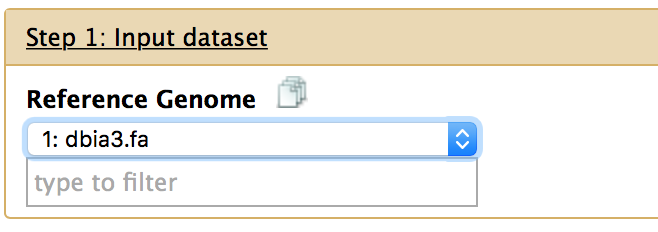


Figure : Select the reference genome for Step 1

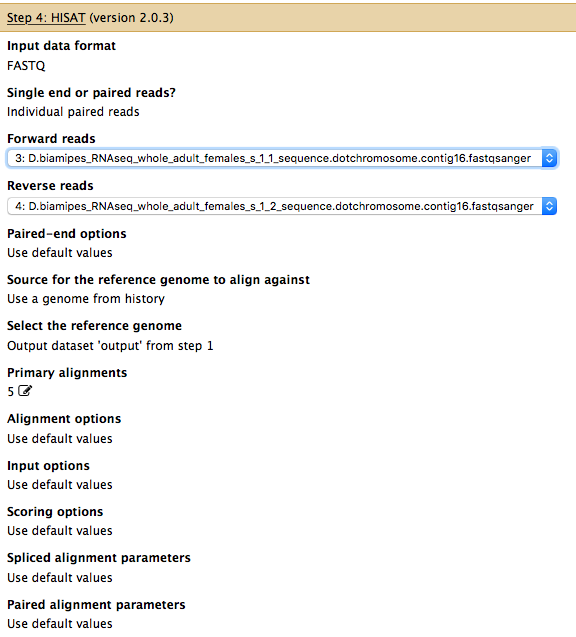


Figure : Select the forward and reverse reads from your RNA-Seq experiments for Step 4

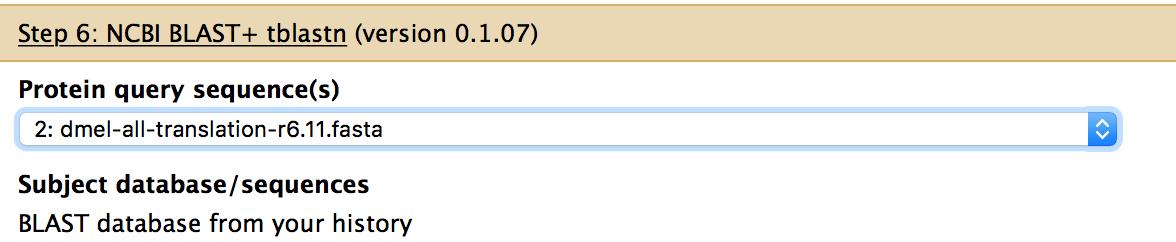


Figure : Select the protein sequence(s) that you would like to align against the assembly in Step 6

You can also use this form to examine and edit the parameter settings for each step of the workflow. Let’s run the workflow with the default settings this time. We will cover how to modify the G-OnRamp workflow in a subsequent walkthrough. (See the “Customize the Genome Browsers produced by G-OnRamp” walkthrough for details.)

Finally, scroll down to the bottom of the page. Check the box to send the results to a new History and then click on the “Run workflow” button to start the analysis (Figure 19) .

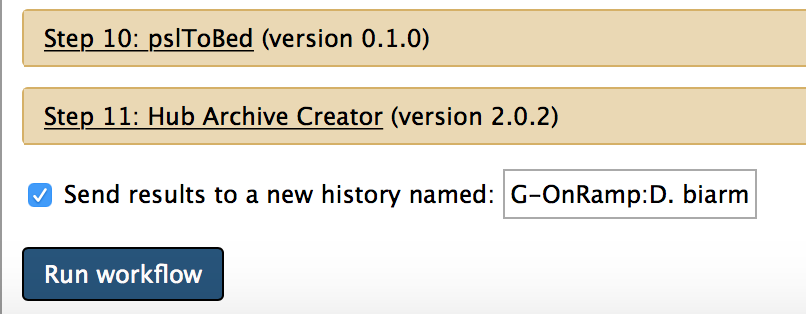


Figure : Check the box to send the workflow results to a new History and then click on the "Run workflow" button to start the analysis

## 3.3 View the results

In order to view your analysis results, you need to switch to the new History that we have just created. Click on the  button at the top right corner of the History panel (Figure 20).

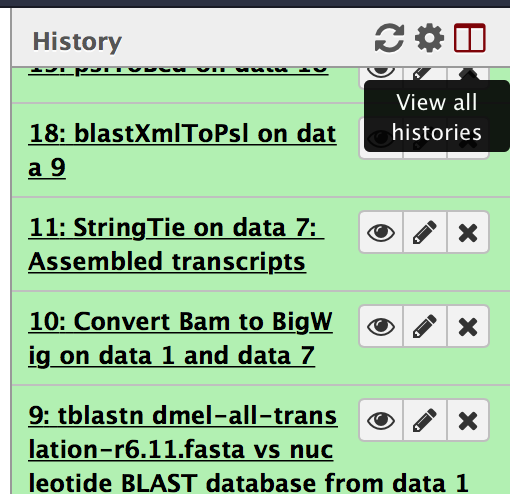


Figure : Click on the “View all histories” button at the top right corner of the History panel

This will open the history management page. Find the history that you have just created (i.e. G-OnRamp:D. biarmipes F element). Click on the “Switch to” button to use this as your Current History. Click on the “Done” button at the top left corner of the page to return to the home page (Figure 21).

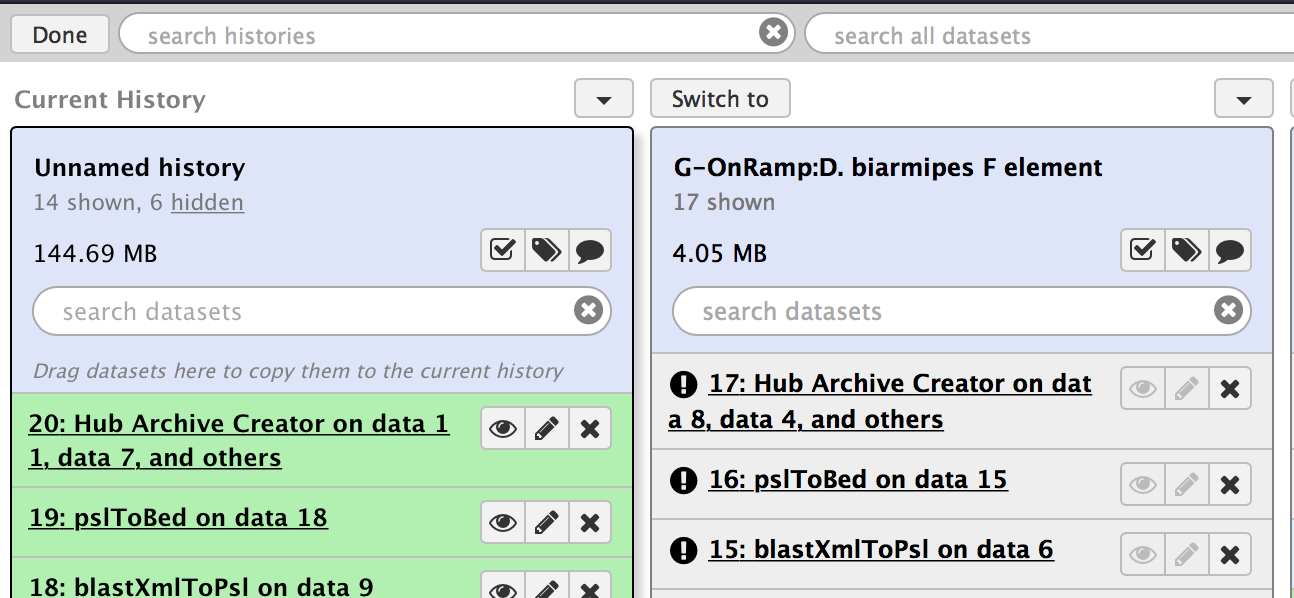


Figure : Switch to the History created by the “G-OnRamp: D. biarmipes F element” workflow.

Now you can see your analysis status and view all the results in the History panel (Figure 22).

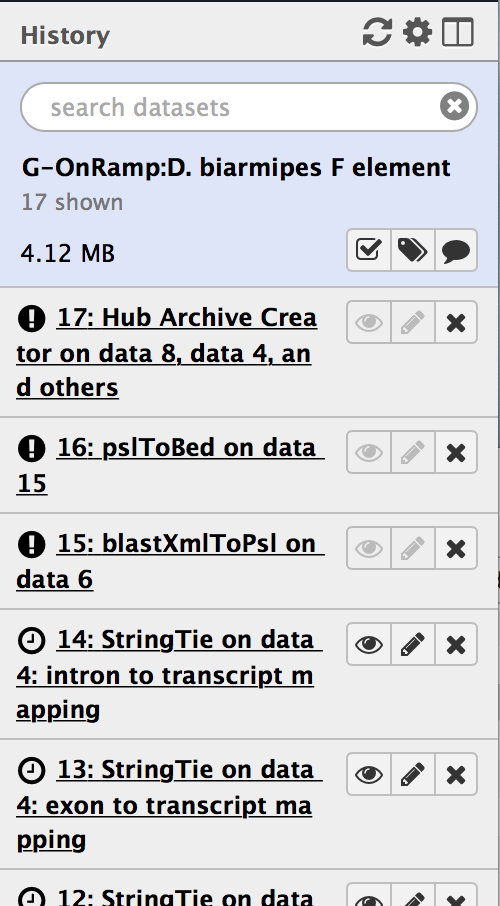


Figure : The “G-OnRamp:D. biarmipes F element” analysis in the History panel

After all the steps in the G-OnRamp workflow are complete (which will take a few minutes), you can view the genome browser for the *D. biarmipes* F element by expanding the “Hub Archive Creator” step (i.e. Step 17) and then click on the “main” link next to the “display at Track Hub UCSC” field (Figure 23).

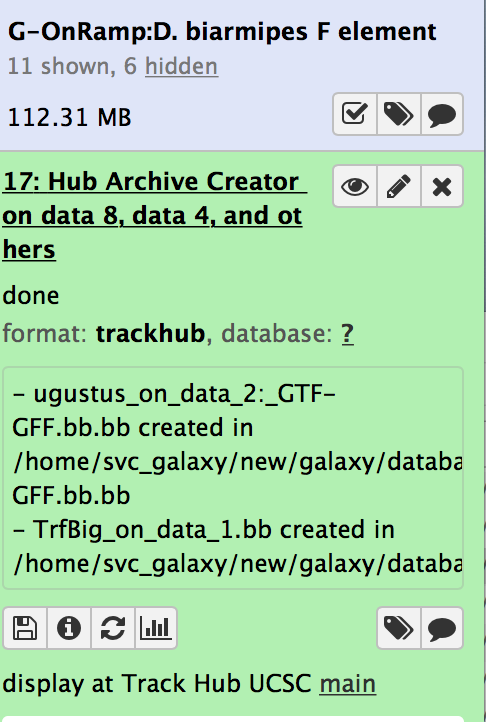


Figure : View results on the UCSC genome browser

Because the RNA-Seq dataset that we used in this analysis only contains the RNA-Seq reads that mapped to contig16, we will examine this contig using the UCSC Genome Browser. Enter “**contig16**” into the “Position/Search Term” field and then click on the GO button (Figure 24). Finally, you can then see the results of the G-OnRamp workflow as different evidence tracks on the UCSC genome browser (Figure 25).

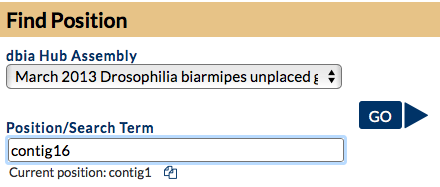


Figure : Specify the contig number in the “Position/Search Term” field and then click on GO

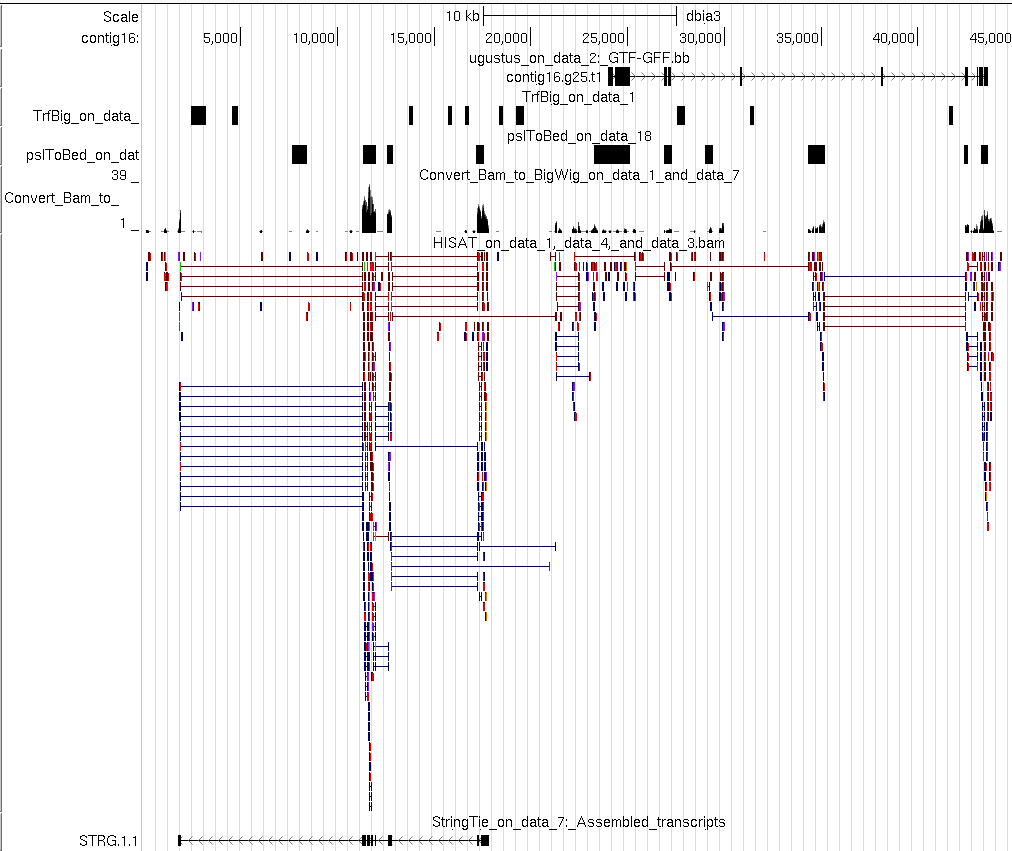


Figure : View the genome assembly and the evidence tracks produced by the Hub Archive Creator on the UCSC genome browser