Introduction to G-OnRamp Walkthrough

Yating Liu

07/2017

# 1. Introduction

The goal of this walkthrough is to illustrate the use of the G-OnRamp workflow to create a Genome Browser. This walkthrough assumes that the reader has some basic familiarity with Galaxy. [See the “Galaxy 101” walkthrough (<https://galaxyproject.org/tutorials/g101/>)] on the Galaxy web site for a detailed introduction to Galaxy]. The G-OnRamp instance that we will use to do the analysis is available at <http://cloud5.galaxyproject.org> (Note that this instance will only be available during the G-OnRamp workshop.). From this tutorial, you will learn how to:

* Create an account and login
* Get data from Data Libraries
* Import the G-OnRamp workflow into your workspace
* Run the workflow.

# 2. Register and login

## 2.1 Create an account

First of all, you need to create and log into your Galaxy account in order to access the full functionality of Galaxy. Open a web browser and navigate to the G-OnRamp instance at <http://cloud5.galaxyproject.org>. (Note that this instance will only be available during the G-OnRamp workshop.)If you have already created an account on the G-OnRamp instance, you can skip this section and jump to section 2.2 Login.

To create an account, click on “User” at the menu bar and then click on “Register” (Figure 1).

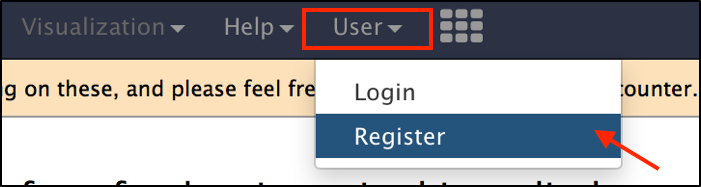


Figure : Create a user account on the G-OnRamp Galaxy instance.

The workspace will display the registration form (Figure 2). In addition to 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 see a green message box indicating that you are now logged into your new account. You can click on “Return to the home page” link to return to the home page (Figure 3).

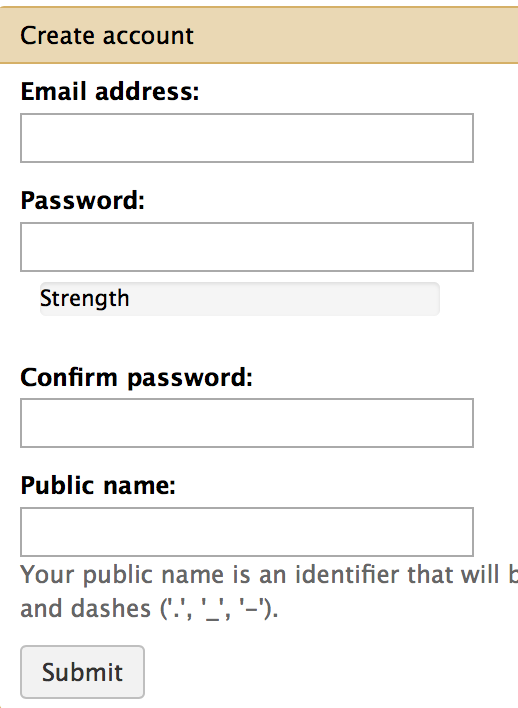


Figure : Registration form for creating a new Galaxy account.

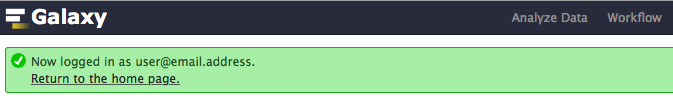


Figure 3: After you click on the “Submit” button, a green message box will appear indicating that you are now logged into your account (i.e. your email address). Click on the "Return to the home page" link to go back to the main page.

## 2.2 Login

When you return to G-OnRamp later, you can log in with your email address and password. First click on the “User” menu item on the menu bar, and then choose “Login” (Figure 4).

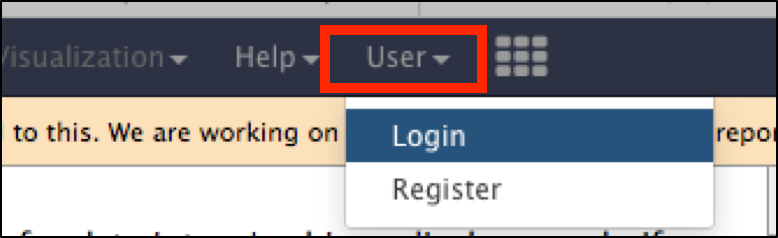


Figure : Open the login form from the “User” menu item on the main menu bar.

Then enter your email address and password (Figure 5).

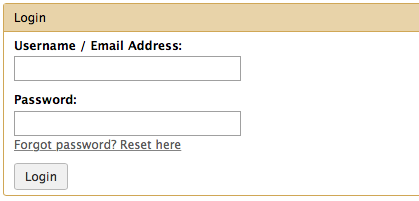


Figure : Enter the email address you have previously used to create your Galaxy account and your account password to log into Galaxy.

# 3. Import G-OnRamp workflow

The G-OnRamp workflow has been shared with you in the Shared Data. Click on “Shared Data” in the menu bar and select “Workflows” in the drop-down menu to access all the shared workflows available on the G-OnRamp instance (Figure 6).

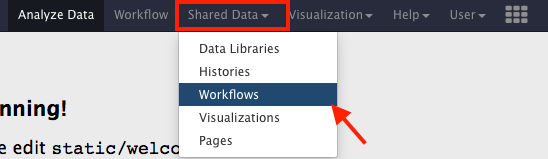


Figure 6: Click on "Shared Data" and then select "Workflows" in the drop-down menu.

Find the “G-OnRamp workflow for UCSC” workflow item and click on the down arrow to open the drop-down menu. Click on the “Import” option to import this workflow into your Galaxy account (Figure 7).

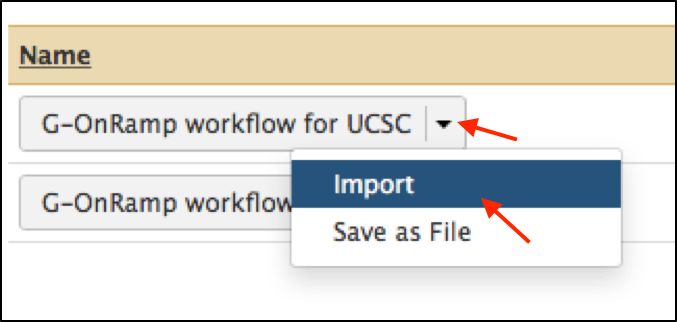


Figure : Click on the down arrow next to the “G-OnRamp workflow for UCSC” label and then select “import” to import this workflow.

You can access the imported workflow via the “Workflow” menu item on the menu bar. The “imported: G-OnRamp workflow for UCSC” workflow will appear under the “Your workflows” section. 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 workflow or click on the down arrow to collapse a drop-down menu, and then select “Rename” (Figure 8).

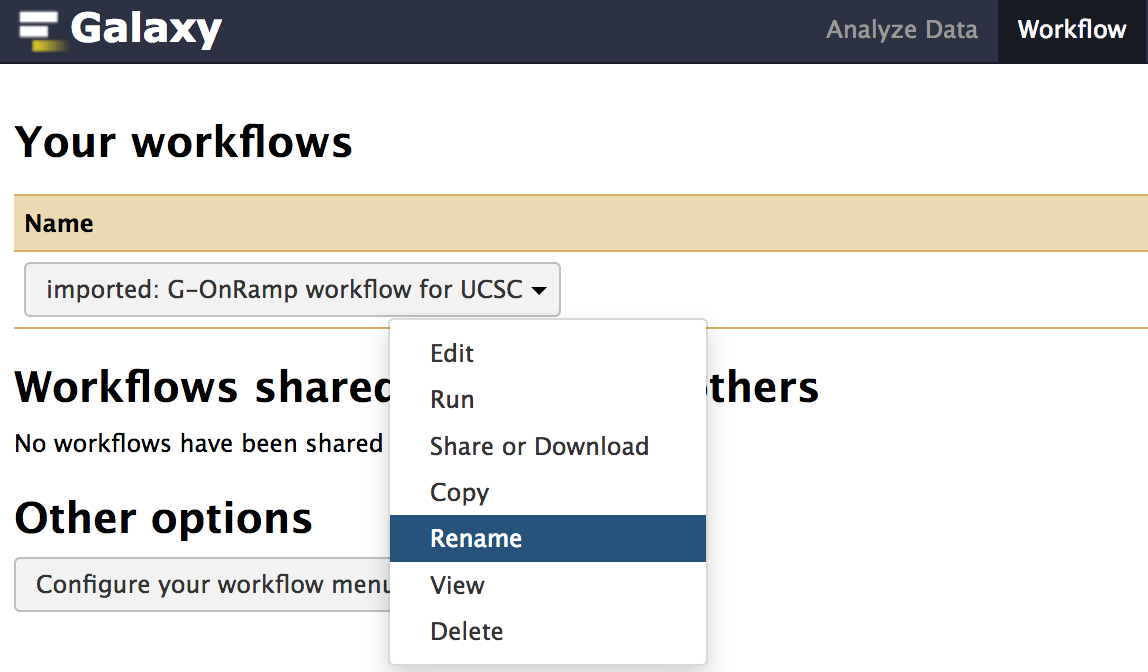


Figure 8: A drop-down menu which shows the different ways that you can manipulate the imported workflow. Click on the “Rename” option to rename the workflow.

Enter “**G-OnRamp: D. biarmipes F element**” into the “Workflow Name” field and then click on the “Rename” button to change the name of the workflow (Figure 9). Go back to the home page by clicking on the “Analyze Data” menu item on the menu bar.

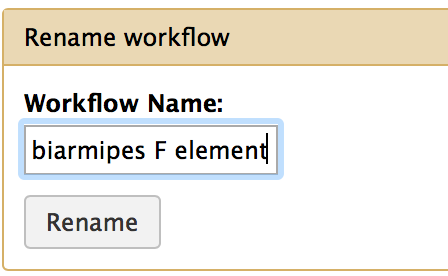


Figure : Rename the workflow.

To help us keep track of the different Histories, we will change the name of the current History to “G-OnRamp: D. biarmipes F element”. Click on the “Unnamed history” title in the History panel, type “**G-OnRamp: D. biarmipes F element**” and then press Enter (Figure 10).

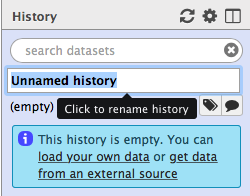


Figure 10: Click on the "Unnamed history" title on the History panel. Change the name of the History to “G-OnRamp: D. biarmipes F element” and then press Enter.

# 3. Run the G-OnRamp workflow

## 3.1 Copy the datasets from Data Libraries

The test datasets that we will use in this walkthrough are available in the Data Libraries. Click on the “Shared Data” menu item on the menu bar and then select “Data Libraries” in the drop-down menu (Figure 11). Click on the “**Intro\_G-OnRamp**” link. You will see five datasets in the folder: the genome sequences from the *Drosophila biarmipes* Muller F element (dbia3.fa), a collection of *Drosophila melanogaster* protein sequences (dmel-hits-translation-r6.11.fa), a collection of *D. melanogaster* RNA GenBank records (dmel-mrna-chrom4.gb.txt), and the *D. biarmipes* RNA-Seq paired-end reads that have previously been mapped to scaffold\_16 (forward reads file: “D.biamipes\_RNAseq\_whole\_adult\_females\_s\_1\_1\_sequence.dotchromosome.contig16.fastqsanger” and reverse reads file: “D.biamipes\_RNAseq\_whole\_adult\_females\_s\_1\_2\_sequence.dotchromosome.contig16.fastqsanger”).

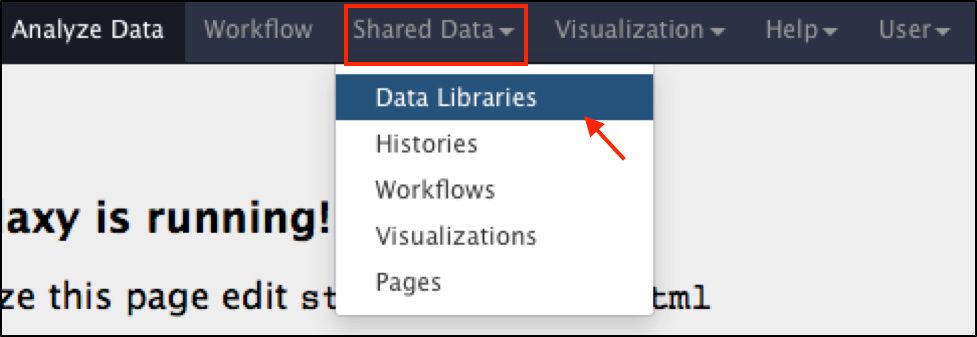


Figure : Click on the "Shared Data" menu item on the top menu bar and then click on "Data Libraries" in the drop-down menu to access the test datasets.

Select all the datasets by selecting the checkbox before the “name” column header. Click on “to History” button to copy all the files to the current History (Figure 12). You will see a pop-up window titled “Import into History”. Click on the “Import” option to import the datasets into the current History (Figure 13).

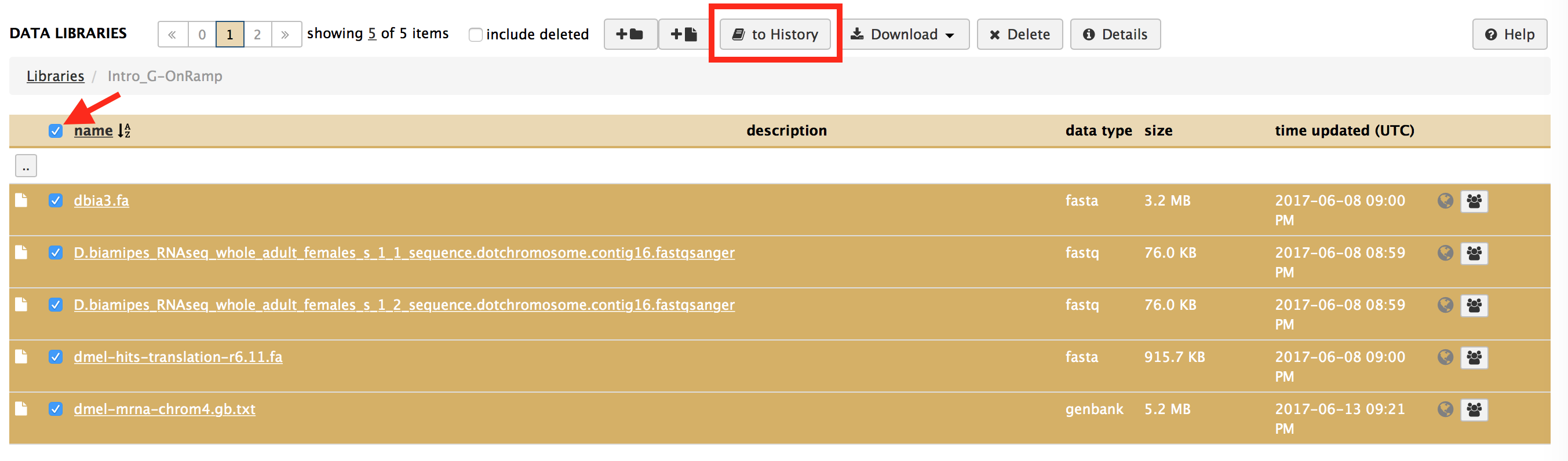


Figure : Select the checkbox next to the “name” column header to select all five datasets. Alternatively, you can select the checkbox next to each dataset individually.

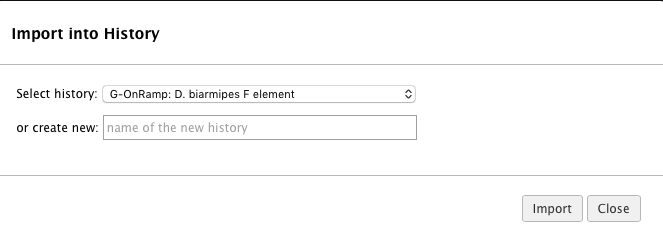


Figure : After choosing the datasets and clicking on the “to History” button, you will see the "Import into History" window. Click on the "Import" button to import the selected datasets into current History.

A green message box will appear once the data import is complete. Click on the “Analyze Data” menu item on the top menu bar to navigate back to the home page. The five files from the Data Libraries will appear in the History panel (Figure 14).

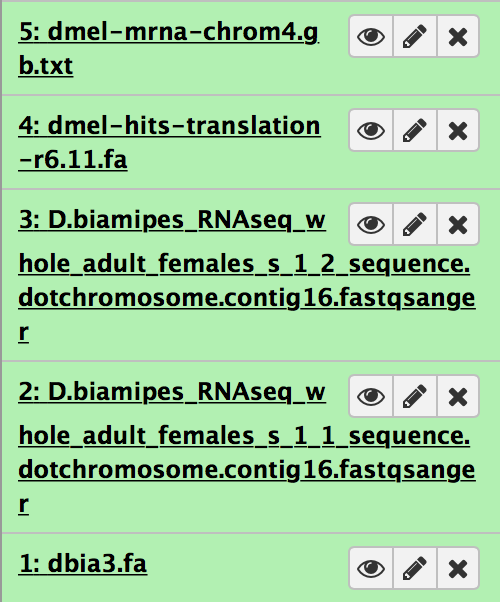


Figure 14: The five datasets were imported into the current History. The green color indicates that the data import was successful.

## 3.2 Run the workflow

Click on the “Workflow” menu item on the menu bar. 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).

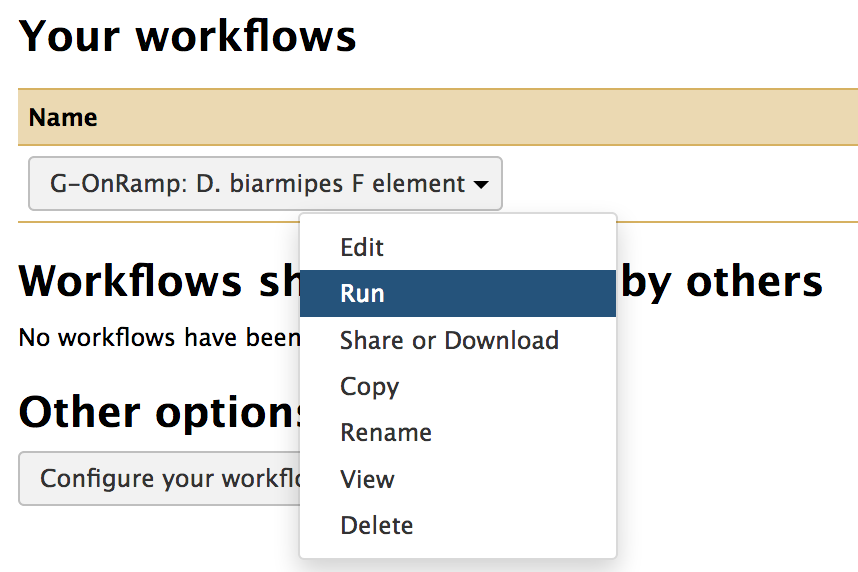


Figure : Click on the “Run” option in the drop-down menu for the “G-OnRamp: D. biarmipes F element” workflow to run the workflow.

The workflow consists of 25 steps. In order to run the workflow, you need to specify the datasets that should be used for the “Reference genome” (Figure 16), “mRNA GenBank records” (Figure 17), “Protein Query Sequence” (Figure 18), “RNA-Seq: Forward reads” and “RNA-Seq: reverse reads” (Figure 19). (Note that the step numbers in your workflow might differ from those shown in the Figures, but the titles should be consistent.)



Figure 16: Select the reference genome (dbia3.fa).



Figure : Select the *D. melanogaster* mRNA records (dmel-mrna-chrom4.gb.txt).



Figure 18: Select the protein sequences (dmel-hits-translation-r6.11.fa) that you would like to align against the assembly.

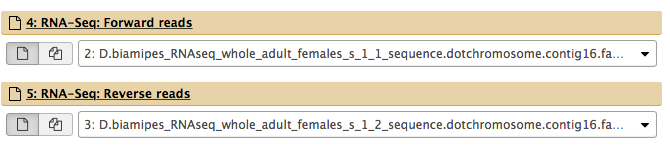


Figure : Select the forward and reverse reads from your RNA-Seq experiments.

You can also use this form to examine and edit the parameter settings for each step of the workflow. For example, you should specify the model organism for each gene prediction tool [Augustus (Figure 20), GlimmerHMM (Figure 21) and SNAP (Figure 22)].

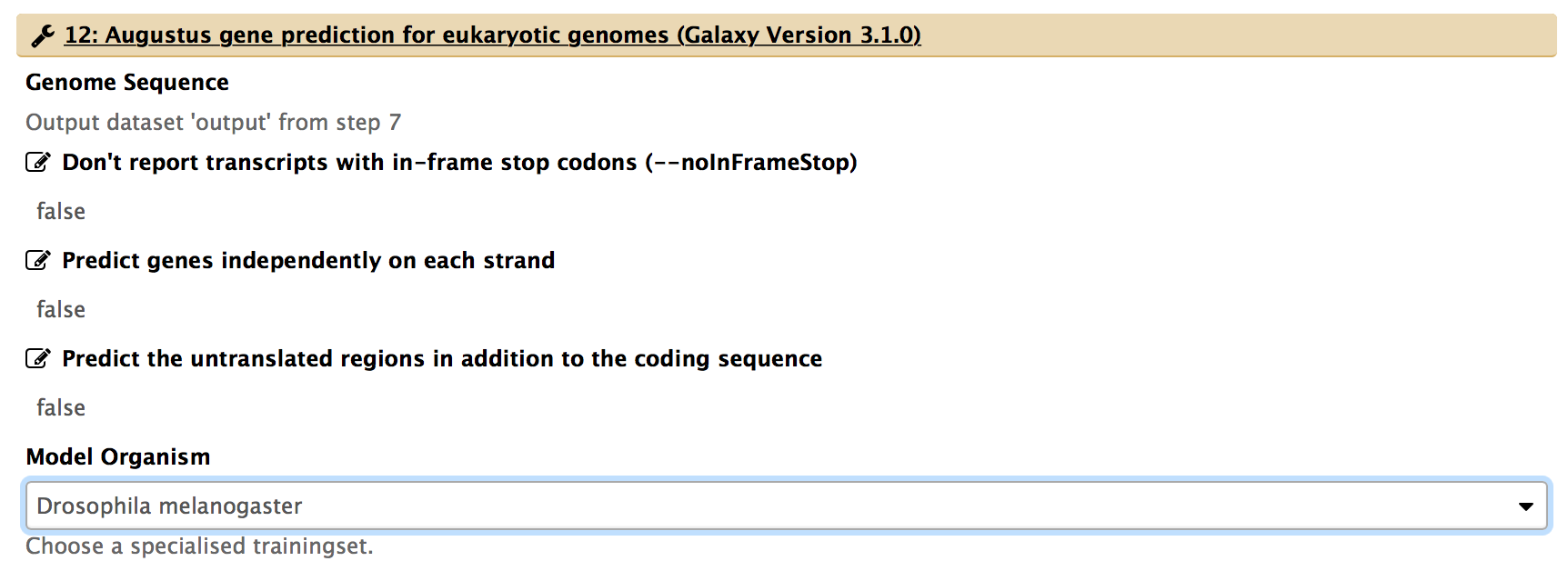


Figure : Choose "*Drosophila melanogaster*" as the model organism in the drop-down menu for Augustus.

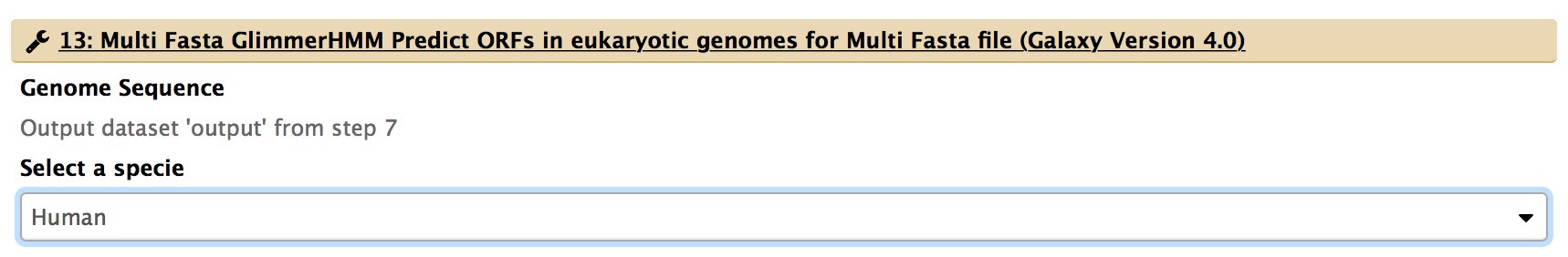


Figure : Select "Human" in the drop-down menu for GlimmerHMM.

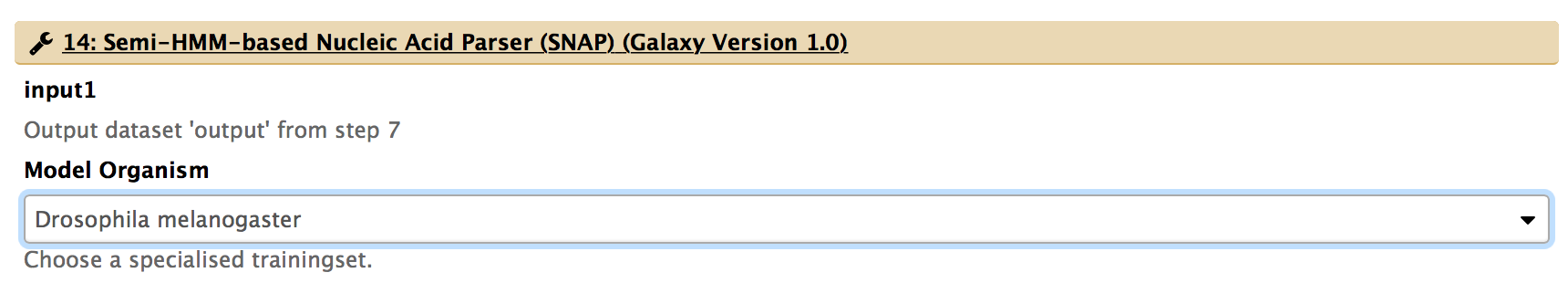


Figure : Choose "*Drosophila melanogaster*" in the drop-down menu for SNAP.

You can customize the name of the assembly hub produced by the Hub Archive Creator. The default name is set to “unknown”. We will change the name of the assembly hub to “Dbia3 Hub Assembly” (Figure 23).

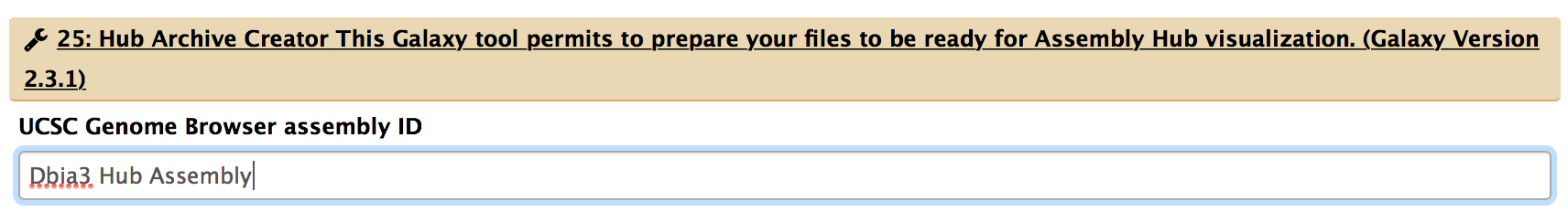


Figure : Specify the UCSC Genome Browser assembly ID.

Finally, click on the “Run workflow” button at the top right corner of the page to run the analysis (Figure 24).

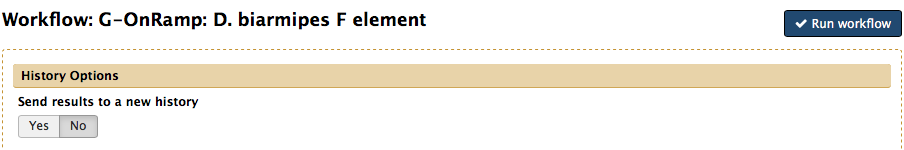


Figure : Click on the "Run workflow" button to start the analysis. By default, the workflow results will be sent to the current History.

## 3.3 View the results

After all the steps in the G-OnRamp workflow have completed (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 30) and then click on the “main” link next to the “display at Track Hub UCSC” field (Figure 25).

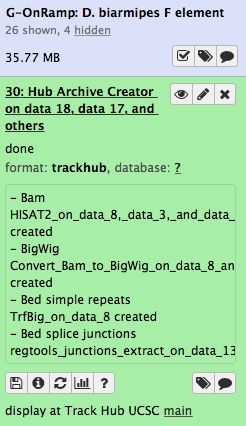


Figure : Click on the “main” link to view the *D. biarmipes* F element assembly hub on the UCSC Genome Browser.

Because the RNA-Seq dataset used in this walkthrough only contains the RNA-Seq reads that mapped to scaffold\_16, we will examine this scaffold using the UCSC Genome Browser. Enter “**scaffold\_16**” into the “Position/Search Term” field and then click on the GO button (Figure 26). Finally, you can then see the results of the G-OnRamp workflow as different evidence tracks on the UCSC Genome Browser (Figure 27).

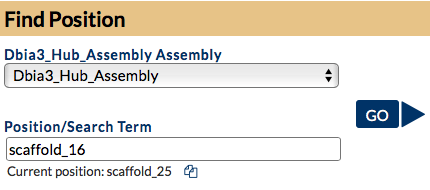


Figure 26: Specify the scaffold number (i.e. scaffold\_16) in the “Position/Search Term” field and then click on the “GO” button.

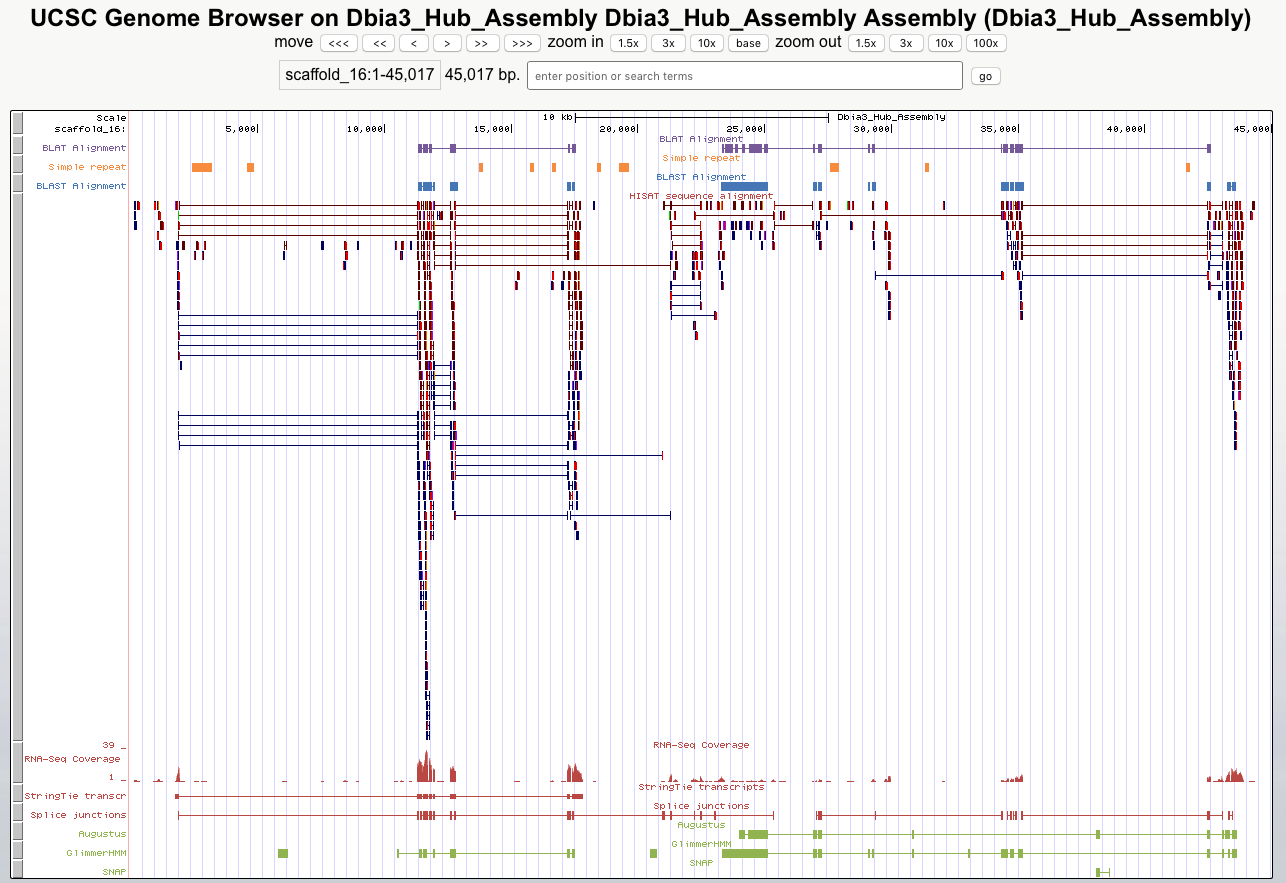


Figure 27: View the genome assembly and the evidence tracks produced by the Hub Archive Creator on the UCSC Genome Browser.