Customize the Genome Browsers produced by G-OnRamp

Yating Liu

12/2018

Table of Contents

[1. Introduction 2](#_Toc530579425)

[2. Modify G-OnRamp workflow 2](#_Toc530579426)

[2.1 Big picture 3](#_Toc530579427)

[2.2 Modify tool parameters 4](#_Toc530579428)

[2.3 Add an evidence track to the Hub Archive Creator 5](#_Toc530579429)

[2.4 Show or hide an output dataset 10](#_Toc530579430)

[2.6 Save the changes to the workflow 11](#_Toc530579431)

[3. Exercise: edit the workflow to run RNA-Seq analysis on two different RNA-Seq samples 12](#_Toc530579432)

[3.1 Make a copy of the current workflow 12](#_Toc530579433)

[3.2 Edit the “G-OnRamp: Dbia3\_RNA-Seq” workflow 12](#_Toc530579434)

[3.3 Upload your datasets and run the workflow 18](#_Toc530579435)

[4. Remove an evidence track from the Hub Archive Creator 21](#_Toc530579436)

[5. Add custom tracks to the Genome Browser 22](#_Toc530579437)

[5.1 Upload the custom track file to the “G-OnRamp: Dbia3\_RNA-Seq” History 23](#_Toc530579438)

[5.2 Run Hub Archive Creator to create a new UCSC Assembly Hub with the custom track 24](#_Toc530579439)

# 1. Introduction

In addition to running the G-OnRamp workflow with default settings, one of the key features of Galaxy is the ability to modify an existing workflow (*e.g.*, change tool parameters, add or remove tools) using the Workflow Canvas. From this tutorial, you will learn how to:

* Modify tool parameters
* Modify the workflow by adding and removing tools
* Add or remove evidence tracks from the Hub Archive Creator
* Customize the workflow to run RNA-Seq analysis on two samples
* Add a custom track to the Hub Archive Creator

Note that this tutorial assumes that the reader is already familiar with the basic concepts of Galaxy and of G-OnRamp. It will modify the “**G-OnRamp: D. biarmipes F element**” workflow we have previously created in the “Introduction to G-OnRamp Walkthrough”.

**We will use the G-OnRamp virtual machine to illustrate how you can modify the G-OnRamp workflow. Open VirtualBox and then launch the G-OnRamp virtual machine. (See the** “Virtual Machine Installation Walkthrough” for details on how to install and run the G-OnRamp virtual machine.) Once the startup process for the G-OnRamp virtual machine is complete, the G-OnRamp Galaxy instance will be available at [http://192.168.56.18](http://192.168.56.18/).

# 2. Modify G-OnRamp workflow

**Log into your account on the G-OnRamp Galaxy instance at** [http://192.168.56.18](http://192.168.56.18/). **Click on the “Workflow” menu item at the top menu bar to access the list of available workflows. Click on the down** arrow for the “**G-OnRamp: D. biarmipes F element**” workflow and click on “**Copy**” to create a copy the G-OnRamp workflow for editing (Figure 1).

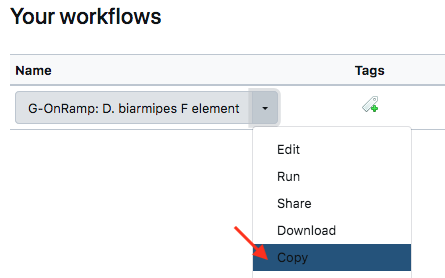


Figure 1: Use the “Copy” option to create a new copy of the G-OnRamp workflow.

Rename the new copy of the workflow (i.e. “Copy of G-OnRamp: D. biarmipes F element”) to “**Customized G-OnRamp**” using the “**Rename**” option in the drop-down menu. **Click on the down** arrow for the “Customized G-OnRamp” workflow and click on “**Edit**” to open the Workflow Canvas (Figure 2).

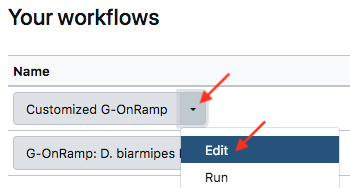


Figure 2: Click on the “Edit” option to open the “Customized G-OnRamp” workflow in the Workflow Canvas.

## 2.1 Big picture

The entire workflow is shown in Figure 3. Each box represents a tool. The “>“ symbol on the left side of the box denotes an input dataset for the tool. The “>“ symbol on the right side of the box denotes the output dataset produced by the tool. A tool could have multiple input datasets (*e.g.*, HISAT2) and output datasets (*e.g.*, Augustus). The “noodles” between the tools correspond to how data are processed by the different tools within the workflow.

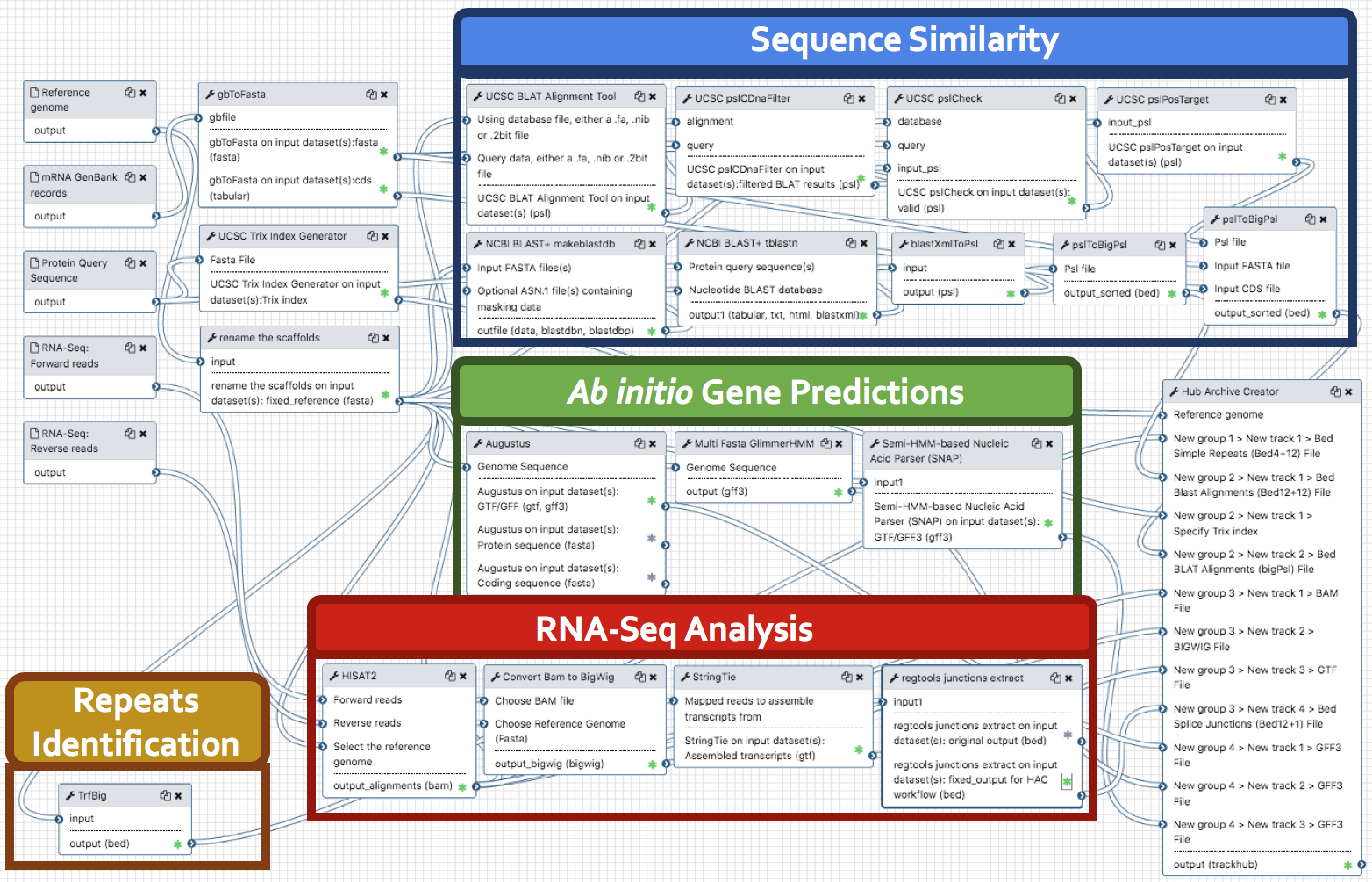


Figure 3: The entire G-OnRamp workflow shown in the Workflow Canvas

Each noodle shows how the output dataset from one tool is used as the input dataset for another tool. For example, Figure 4 shows the connection between the “rename the scaffolds” tool and the “Augustus” gene predictor. The output from the “rename the scaffolds” tool serves as the “Genome Sequence” input for Augustus.

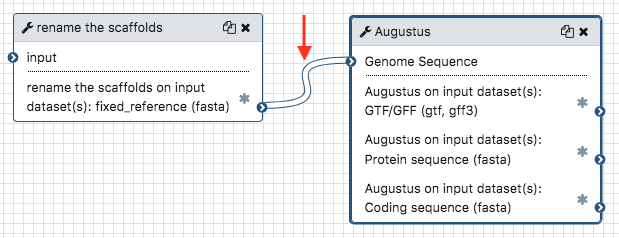


Figure 4: The connection between the output of the “rename the scaffolds” tool and the input of the “Augustus” tool (red arrow).

## 2.2 Modify tool parameters

You can click on each tool in the Workflow Canvas to learn more about the tool (i.e. what it does and how to use it). You can use the “Details” panel on the right to examine and change the tool parameters (Figure 5). If the analysis workflow was originally derived from a History, then the settings for each tool within the workflow will reflect the parameters used in that analysis. For example, because we used paired-end RNA-Seq data in the “Introduction to G-OnRamp Walkthrough”, the “**Individual paired reads**” option is selected in the HISAT2 tool. If you have unpaired RNA-Seq data, you will need to select the “Individual unpaired reads” option under the “Single end or paired reads” field. **(Note that you do not need to change the parameter here because we will use paired-end RNA-Seq data in this walkthrough.)**

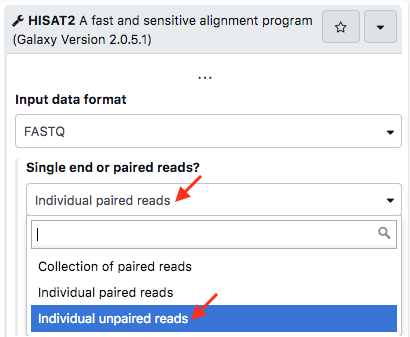


Figure 5: Click on HISAT2 and use the Details panel to edit the parameter settings for the HISAT2 tool.

If there is an arrow on the left side of a parameter, then the parameter could be set when you run the workflow (i.e. set at Runtime). For instance, if you open the Details panel of the Augustus tool, you will see an up arrow on the left side of the “**Model Organism**” parameter. Click on the up arrow icon to hide the field and to set this parameter at runtime (Figure 6).

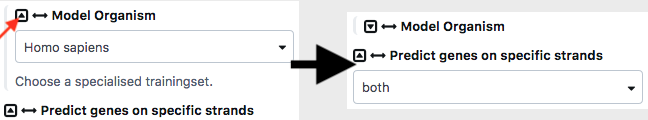


Figure 6: Click on the up arrow icon on the left side of the “Model Organism” field to set this parameter at runtime (red arrow).

## 2.3 Add an evidence track to the Hub Archive Creator

You can also use the Workflow Canvas to add or delete a tool. For example, if you want to add the results from WindowMasker (which identifies simple repeats and low complexity sequences) to the G-OnRamp workflow, you can use the search field in the “Tools” panel to search for “**WindowMasker**”. The tools WindowMasker\_ustat and WindowMasker\_mkcounts will appear in the Tools panel (Figure 7). These tools correspond to the two stages (ustat and mk\_counts, respectively) used by WindowMasker to identify repeats in DNA sequences. Click on both links to add the **WindowMasker\_ustat** and **WindowMasker\_mkcounts** tools to the Workflow Canvas (Figure 8).

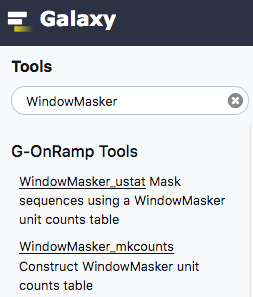


Figure 7: Use the search field in the Tools panel to search for the WindowMasker tools.

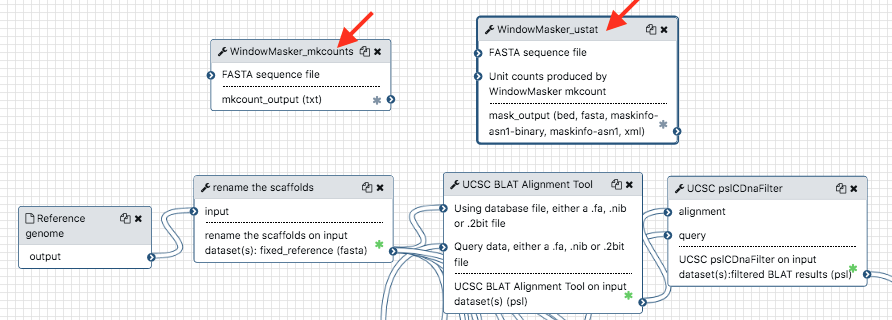


Figure 8: The WindowMasker\_ustat and WindowMasker\_mkcounts tools are added to the Workflow Canvas (red arrows).

The next step would be to incorporate the WindowMasker\_ustat and WindowMasker\_mkcounts tools with the rest of the G-OnRamp workflow. This is accomplished by specifying the input and output datasets for these WindowMasker tools.

The first stage of the WindowMasker analysis is mk\_counts (WindowMasker\_mkcounts), which constructs a unit counts table for a genome assembly. The unit counts correspond to the frequency of short sequences with length k (k-mers) in the genome assembly. There is a “>“ symbol on the left side of the “**FASTA sequence file**” field in WindowMasker\_mkcounts, which indicates that this tool requires the genome sequences in FASTA format as an input dataset. Here we will use the genome sequences that have been renamed by “rename the scaffolds” tool, which shortens the scaffold names so that they are less than 32 characters. (Older versions of the UCSC Genome Browser imposed a 31-character limit on the scaffold names.)

To establish a new connection between the “rename the scaffolds” and the WindowMasker\_mkcounts tools, click on the “>“ symbol on the right side of “**output (fasta)**” in the “rename the scaffolds” tool and drag it to the “>“ symbol on the left side of “**FASTA sequence file**” in the WindowMasker\_mkcounts tool. As you drag the connection, the connection will appear as a green “noodle” (Figure 9). When you release the mouse next to the “FASTA sequence field” field on the left side of the WindowMasker\_mkcounts box, it will establish the connection between these two tools.

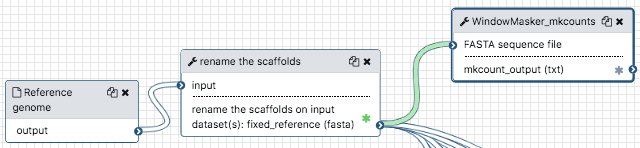


Figure 9: Connect the output from the “rename the scaffolds” tool to the input of the WindowMasker\_mkcounts tool.

The second stage of the WindowMasker analysis is ustat (WindowMasker\_ustat), which uses the unit counts produced by WindowMasker\_mkcounts to mask repetitive regions within a DNA sequence. The WindowMasker\_ustat tool requires a FASTA sequence file as an input dataset. Repeat the steps above to create a connection between “**output (fasta)**” of the “rename the scaffolds” tool with the “**FASTA sequence file**” input for the WindowMasker\_ustat tool (red arrow in Figure 10).

WindowMasker\_ustat also needs the unit counts table produced by WindowMasker\_mkcounts as an input. Connect the “**mkcount\_output**” for the WindowMasker\_mkcounts tool with the “**Unit counts produced by WindowMasker mkcount**” input for the WindowMasker\_ustat tool (blue arrow in Figure 10).

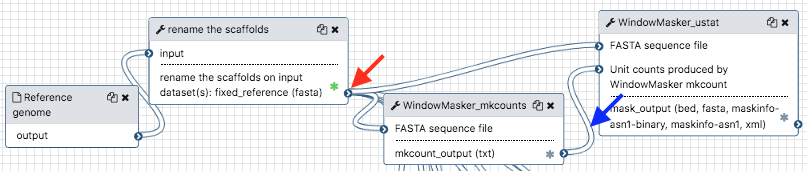


Figure 10: Connect the “output (fasta)” from the “rename the scaffolds” tool to the “FASTA sequence file” inputs for the WindowMasker\_mkcounts and WindowMasker\_ustat tools (red arrow). In addition, connect the “mkcount\_output (txt)” from the WindowMasker\_mkcounts tool to the “Unit counts produced by WindowMasker mkcount” input for the WindowMasker\_ustat tool (blue arrow).

Using the same approach, you can connect the output of the WindowMasker\_ustat tool to the Hub Archive Creator. However, because all the input connections to the Hub Archive Creator are already connected to the output connections from the other tools, we need to add another input connection to the Hub Archive Creator before we can create the connection with the output from the WindowMasker\_ustat tool.

Click on the Hub Archive Creator box in the Workflow Canvas and then examine the Details panel on the right. Scroll down to the end of the first group (“**Repeats**”) and click on “**Insert New track**” to create the WindowMasker track (Figure 11).



Figure 11: Click on the “Insert New Track” button in the Details pazel to create a new input connection for the Hub Archive Creator.

In order to establish a connection between two tools, the datatype of the output dataset from the first tool must be the same as the datatype of the input dataset for the second tool. You can see the output format of WindowMasker by clicking on WindowMasker\_ustat and open the Details panel. In this case, the output format for WindowMasker\_ustat is **BED** (Figure 12).

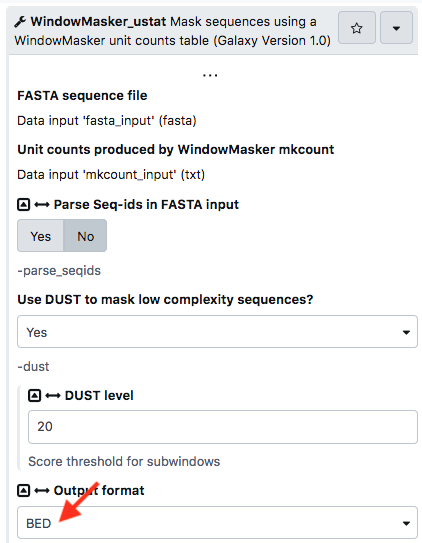


Figure 12: Open the Details panel of the WindowMasker\_ustat tool and verify that the output format is set to BED.

Consequently, we need to set the format for the new input connection in the Hub Archive Creator to “**BED**” (Figure 13, red arrow). We will use “**BED Generic**” as the Bed Choice (Figure 13, blue arrow). You can specify the label for the track as “**WindowMasker**” and select a color (*e.g.*, light blue) for that track (Figure 13, purple arrows).

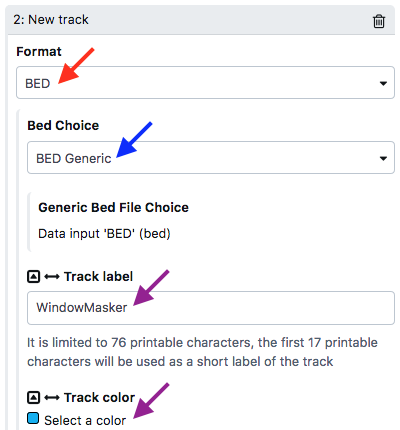
z

Figure 13: Select the “BED” format in the drop-down menu, then change the “Track label” to “WindowMasker” and the “Track color” to “light blue”.

A “**New group 1 > New track 2 > Generic Bed File Choice**” entry will appear in the Hub Archive Creator tool (Figure 14).

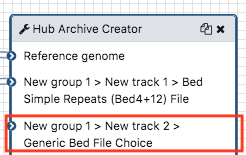


Figure 14: The Hub Archive Creator with a new BED input connection (red rectangle)

After you have created the new BED input connection, you can connect the output of the WindowMasker\_ustat tool to the input of the Hub Archive Creator. Click on the “>“ symbol next to the “**mask\_output**” field in the WindowMasker\_ustat tool, drag it to the “>“ symbol next to the “**New group 1 > New track 2 > Generic Bed File Choice**” entry in the Hub Archive Creator tool and then release the mouse (Figure 15).

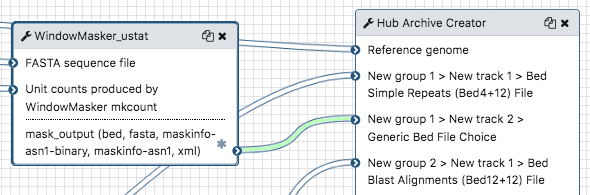


Figure 15: Connect the output of WindowMasker\_ustat to the Hub Archive Creator

## 2.4 Show or hide an output dataset

To simplify the display in the History panel, the output of each tool is hidden from your History by default. To show a dataset in the History, you can mark the dataset as a workflow output by clicking on the “\*” symbol. All unmarked datasets will be hidden from your History. For example, StringTie can produce up to eight output files. However, because only the “**output\_gtf (gtf)**” output is marked (the “\*” symbol is in dark orange) in the workflow, only the GTF file will appear in your History after you run the workflow (Figure 16). The other ten output datasets will be hidden from the History. This feature is particularly useful when you are working with large workflows that produce many temporary datasets.

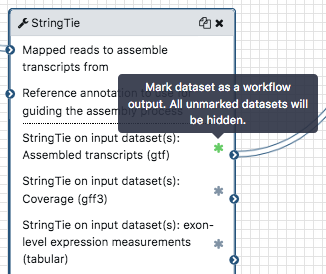


Figure 16: The orange star next to “output\_gtf (gtf)” output for the StringTie tool indicates that this dataset has been marked as a workflow output.

You can show the outputs from WindowMasker\_mkcounts and WindowMasker\_ustat tools in the History by clicking on the “\*” symbols next to “**mkcount\_output (txt)**” and “**mask\_output**” (Figure 17).

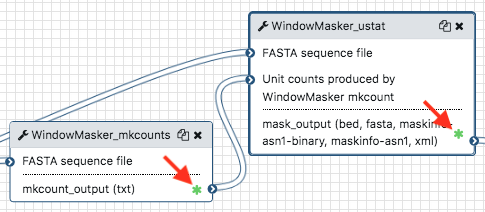


Figure 17: Show the outputs from the WindowMasker tools in the History by clicking on the “\*” symbols next to the output datasets (red arrows).

## 2.6 Save the changes to the workflow

Remember to save your changes before you leave the Workflow Canvas page. Click on the “**Save**” floppy disk icon at the top right corner of Workflow Canvas (Figure 18).

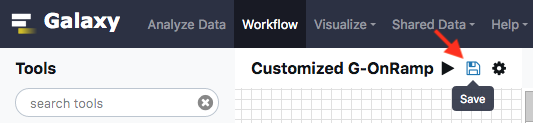


Figure 18: Save the changes that you have made to the workflow

# 3. Exercise: edit the workflow to run RNA-Seq analysis on two different RNA-Seq samples

## 3.1 Make a copy of the current workflow

Make a copy of “Customized G-OnRamp” workflow and then rename the new workflow as “**G-OnRamp: Dbia3\_RNA-Seq**” (Figure 19). Click on the drop-down menu for the new workflow and select “**Edit**” to go to the Workflow Canvas.

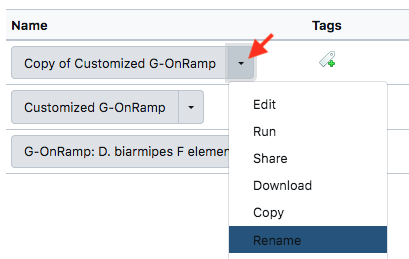


Figure 19: Copy the current workflow and rename it as “G-OnRamp: Dbia3\_RNA-Seq”

## 3.2 Edit the “G-OnRamp: Dbia3\_RNA-Seq” workflow

We will modify the workflow to process RNA-Seq data from two samples (*D. biarmipes* RNA-Seq paired-end reads from adult females and adult males). You can run the RNA-Seq analysis tools on each sample within a single workflow and then view the evidence tracks for both samples on the UCSC Genome Browser. To reduce the amount of time required for this analysis, we will use only a small subset of the RNA-Seq data from the original samples (i.e. the “shallow” datasets).

Below are some hints on how to construct this workflow:

First, you need to add an additional set of tools, including HISAT2, StringTie, Convert Bam to BigWig, and regtools junctions extract, in order to perform the RNA-Seq analysis for the two samples.

Second, you need to modify the parameter settings for HISAT2. Select the “**Individual paired reads**” option under the “Single end or paired reads?” field (Figure 20). Select the “**Use a genome from history**” option under the “Source for the reference genome to align against” field (Figure 21).

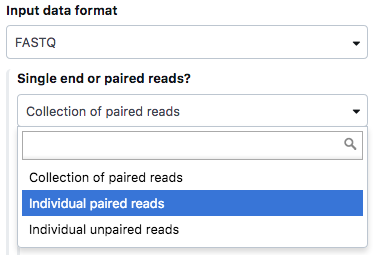


Figure 20: In the “Details” panel of the HISAT2 tool, choose the “Individual paired reads” option under the “Single end or paired reads?” field.

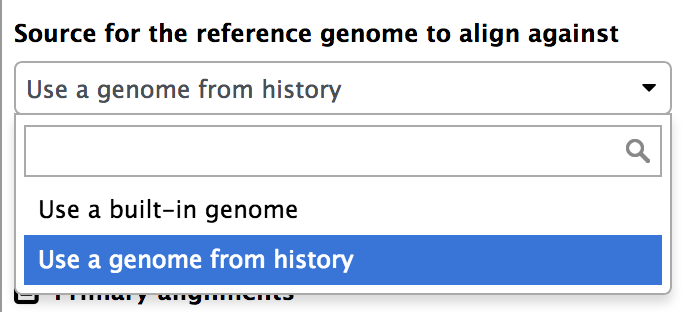


Figure 21: Choose the “Use a genome from history” option for the “Source for the reference genome to align against” field.

Third, you need to add two “Input dataset” to the workflow for the forward and reverse reads files from the second RNA-Seq sample. Click on the “**Inputs**” header in the Tools panel to expand the section. Click on the “**Input dataset**” link twice to add two input datasets to the workflow (Figure 22).

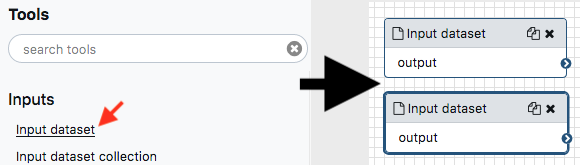


Figure 22: Click on the “Inputs” header in the “Tools” panel and then click on the “Input dataset” link twice to add two new input datasets to the workflow (i.e. for the forward and reverse reads of an additional RNA-Seq sample).

In order to differentiate the forward and reverse reads for the two RNA-Seq samples, you need to change the labels of the four RNA-Seq input datasets in the workflow. Click on each input dataset and edit their step label on the “Details” panel (Figure 23). Change the labels of the RNA-Seq input datasets in the original workflow to “**Male shallow: Forward reads**” and “**Male shallow: Reverse reads**”, respectively. Change the labels of the two new input datasets to “**Female shallow: Forward reads**” and “**Female shallow: Reverse reads**” (Figure 24).

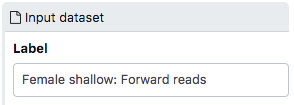


Figure 23: Change the step labels of the RNA-Seq input datasets to indicate the expected data file. For example, you can change the label of the input dataset for the forward reads of the adult females RNA-Seq sample to “Female shallow: Forward reads”.

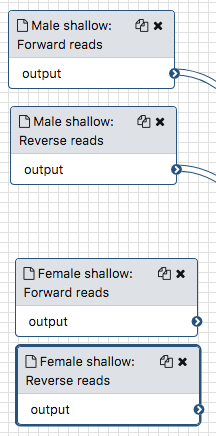


Figure 24: Label the four RNA-Seq input datasets as “Male shallow: Forward reads”, “Male shallow: Reverse reads”, “Female shallow: Forward reads”, and “Female shallow: Reverse reads”.

Connect the outputs from the female shallow datasets to the “**Forward reads**” and “**Reverse reads**” inputs of the new HISAT2 tool, respectively. Connect the output from the “rename the scaffolds” tool to the “**Select the reference genome**” input for the new HISAT2 tool (Figure 25).

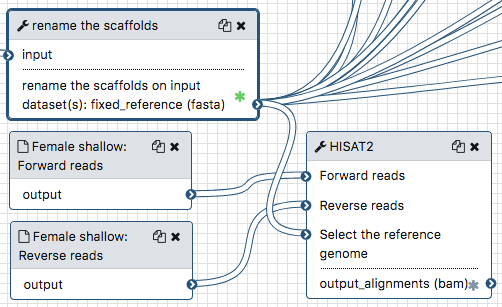


Figure 25: Connect the output of “Female shallow: Forward reads” with the “Forward reads” input of the HISAT2 tool. Connect the output of “Female shallow: Reverse reads” with the “Reverse reads” input. Connect the “fixed\_reference (fasta)” of “rename the scaffolds” with the “Select the reference genome” input.

Connect the “**output\_alignments (bam)**” from HISAT2 with the inputs for Convert Bam to BigWig, regtools junctions extract, and StringTie. You should also mark the outputs from these tools as workflow outputs by clicking on the “\*” symbol (Figure 26). These outputs will be used by the Hub Archive Creator to create evidence tracks on the UCSC Assembly Hub.

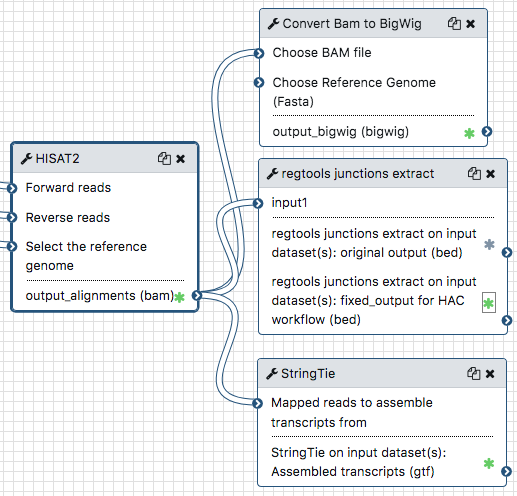


Figure 26: Connect the output alignments from HISAT2 to the inputs of Convert Bam to BigWig, regtools junctions extract, and StringTie. Click on the “\*” symbol next to “output\_alignments (bam)”, “output\_bigwig (bigwig)”, “fixed\_output for HAC workflow (bed)”, and “Assembled transcripts (gtf)” to mark them as workflow outputs.

The Convert Bam to BigWig tool also needs the reference genome as its input. Here we connect the “**fixed\_reference (fasta)**” from the “rename the scaffolds” tool to the “**Choose Reference Genome (Fasta)**” of the “Convert Bam to BigWig” tool (Figure 27).

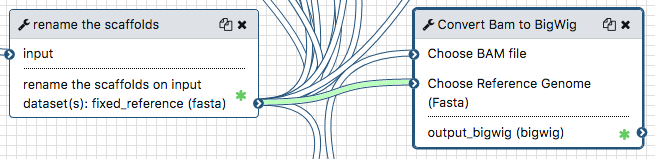


Figure 27: Use output (fasta) from the “renamed the scaffold” tool as the “Choose Reference Genome (Fasta)” input for the “Convert Bam to BigWig” tool

In order to incorporate the outputs from HISAT2, Convert Bam to BigWig, regtools junctions extract, and StringTie into the UCSC Assembly Hub, we need to add input connections to the Hub Archive Creator. To add these connections, click on the Hub Archive Creator box in the Workflow Canvas and then examine the “Details” panel on the right. Scroll down to end of the third group “**RNA-Seq Analysis**” and click on the “**Insert New track**” button four times to create four new input connections.

Configure the Format field for the four new evidence tracks as follows:

1. Select the Format “**BAM**” for the “**output\_alignments (bam)**” dataset from HISAT2
2. Select the Format “**BIGWIG**” for the “**output\_bigwig (bigwig)**” dataset from the “Convert BAM to BigWig” tool
3. Select the Format “**BED**”, and the Bed Choice “**BED Splice junctions (bed12+1 / spliceJunctions.as)**” for the “**fixed\_output (bed)**” dataset from the “regtools junctions extract” tool
4. Select the Format “**GTF**” for the “**output\_gtf (gtf)**” dataset from the StringTie tool

You can add a “Track label” to the new tracks to provide additional details on the expected inputs to the Hub Archive Creator. For example, you can label the new tracks “**Female shallow Sequence Alignment**”, “**Female shallow Sequence Coverage**”, “**Female shallow Splice Junctions**”, and “**Female shallow StringTie Transcripts**”. You might also want to customize the Track labels for the previous set of RNA-Seq analysis tracks by adding “**Male shallow**” to the beginning of the labels.

You can also customize the color of each track by clicking on “select a color”. For example, you can assign the red color to the Male shallow RNA-Seq sample, and the blue color to the Female shallow RNA-Seq sample (Figure 28).

Finally, we will connect the marked output from each tool with their corresponding connections in the Hub Archive Creator (Figure 29).

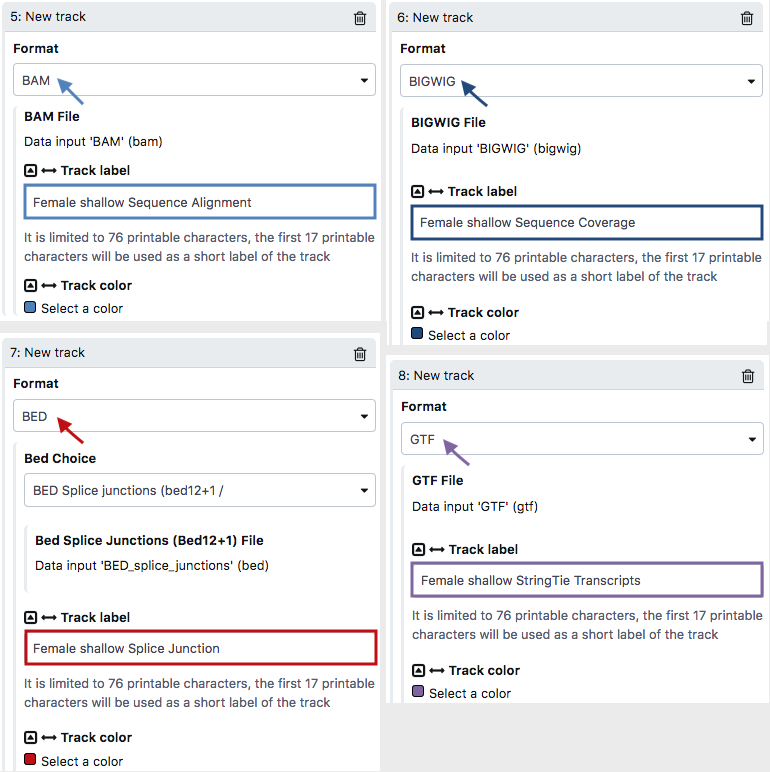


Figure 28: Insert four new tracks to the “RNA-Seq Analysis” group to add input connections for the outputs from HISAT2, Convert Bam to BigWig, regtools junctions extract, and StringTie to the Hub Archive Creator. Specify the corresponding format (color arrows) and track labels (color boxes) for each track. You can change the track color by clicking on “Select a color”.

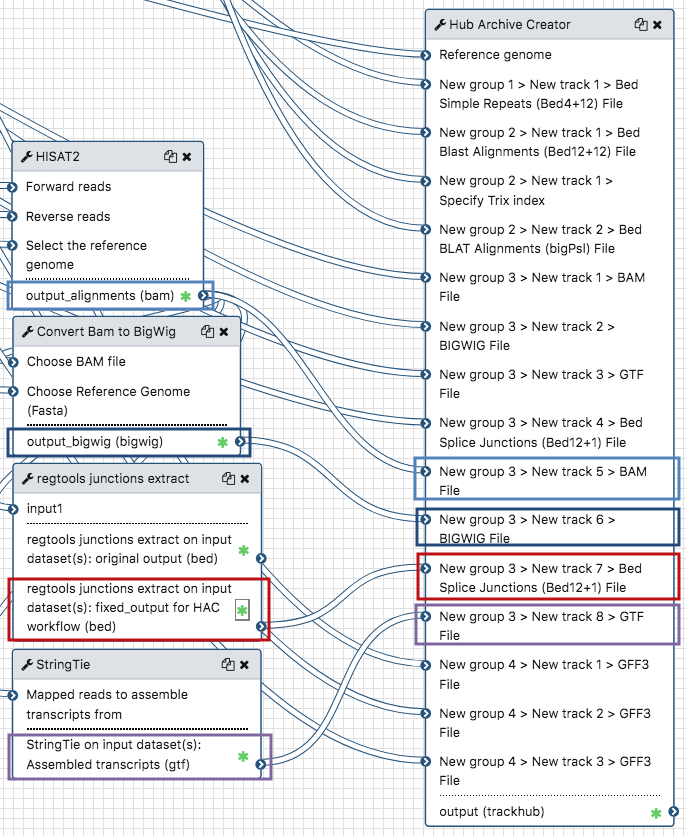


Figure 29: Connect the marked output from each tool (color boxes) to the corresponding new input connections (boxes with the same colors as their corresponding inputs) in the Hub Archive Creator.

Don’t forget to save the changes before you leave the Workflow Canvas.

## 3.3 Upload your datasets and run the workflow

To help us keep track of the different analyses, we will create a new History and change the name of the History to “**G-OnRamp: Dbia3\_RNA-Seq**”.

The reference genome sequence and the informant transcript and protein sequence files are available in the Data Libraries. Click on the “**Shared Data**” menu item at the top menu bar and then select “**Data Libraries**” in the drop-down menu. Click on the “**Intro to G-OnRamp**” link, and then select the checkbox next to the reference genome sequence (**dbia3.fa**), *D. melanogaster* protein sequences (**dmel-hits-translation-r6.11.fa**), and the *D. melanogaster* RNA GenBank records (**dmel-mrna-chrom4.gb.txt**). Click on the “**Export to History**” button and then select the “**as Datasets**” option (Figure 30). Click on the “**Import**” button to begin the data transfer.

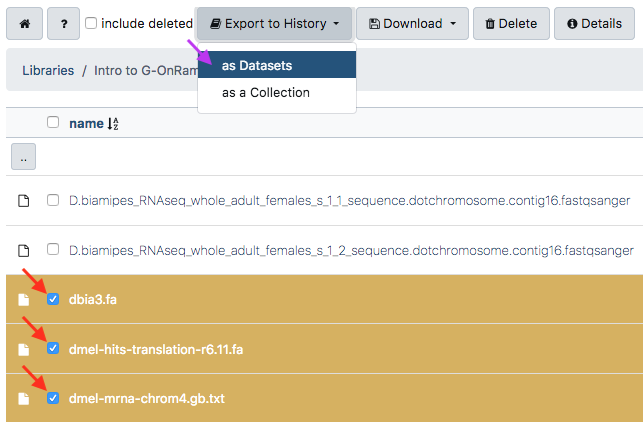


Figure 30: Transfer the *D. biarmipes* genome sequence and the *D. melanogaster* protein and transcript records (red arrows) from the “Intro to G-OnRamp” Data Library to the “G-OnRamp: Dbia3\_RNA-Seq” History as separate Datasets (purple arrow).

Click on the “**Analyze Data**” menu item on the menu bar to go back to the home page. The forward and reverse RNA-Seq reads from the adult females and adult males samples are available in the “**data**” directory of the exercise package:

* Adult females forward reads: **Dbia3\_adult\_females\_shallow\_1.fastq.zz**
* Adult females reverse reads: **Dbia3\_adult\_females\_shallow\_2.fastq.gz**
* Adult males forward reads: **Dbia3\_adult\_males\_shallow\_1.fastq.gz**
* Adult males reverse reads: **Dbia3\_adult\_males\_shallow\_2.fastq.gz**

Use the “**Upload File**” tool (available under “Get Data”) to transfer these RNA-Seq datasets to the G-OnRamp server (Figure 31).

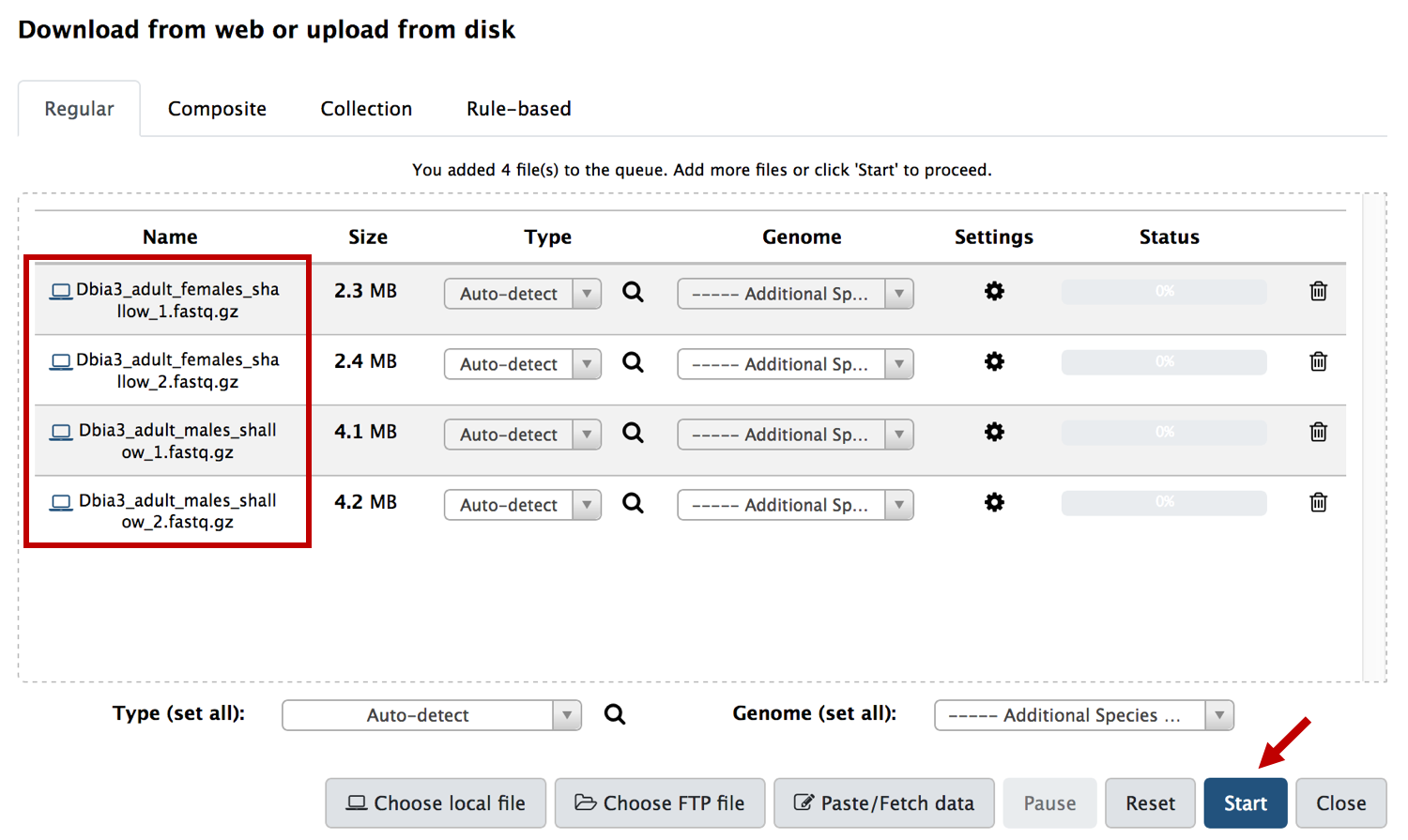


Figure 31: Use the “Upload File” tool to transfer the four RNA-Seq read files in the “data” directory of the exercise package to the G-OnRamp server. Click on the “Start” button (red arrow) to begin the transfer.

Once the file transfer is complete, click on the “**Workflow**” menu item at the top menu bar. Click on the down arrow next to the “**G-OnRamp: Dbia3\_RNA-Seq**” workflow and then select “**Run**”. Be sure to select the correct input datasets (Figure 32) and change the model organism parameter for the gene prediction tools (Figure 33). Remember to change the “UCSC Genome Browser assembly ID” (*e.g.*, to “**Dbia3 RNASeq Hub**”) in the “Hub Archive Creator” step.

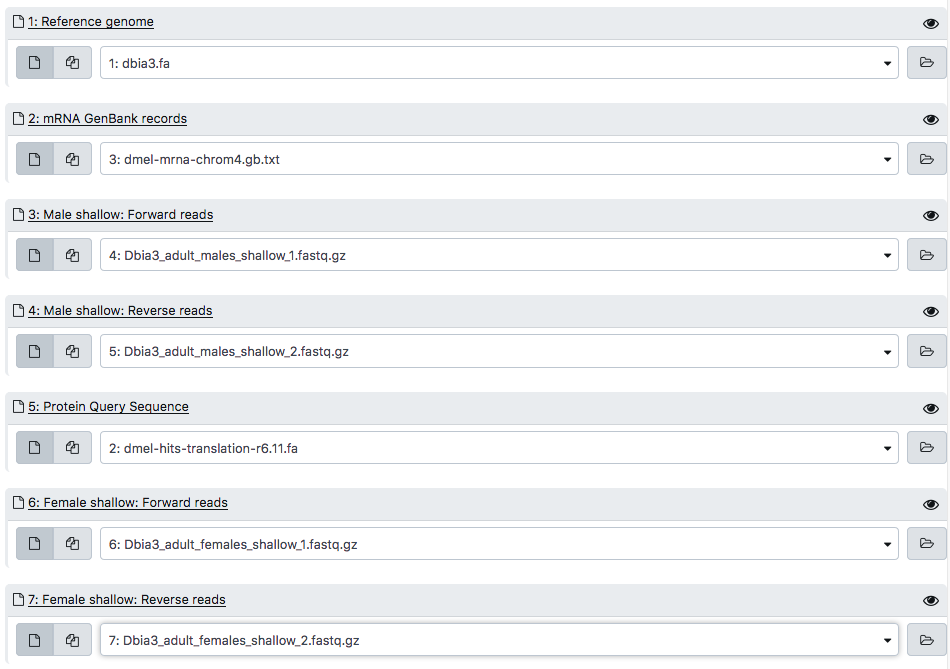


Figure 32: Make sure that you have selected the correct input datasets for the first seven steps of the workflow. (Note: because the step numbers depend on the layout of the workflow, the step numbers in your version of the workflow might not match the numbers shown in this figure.)

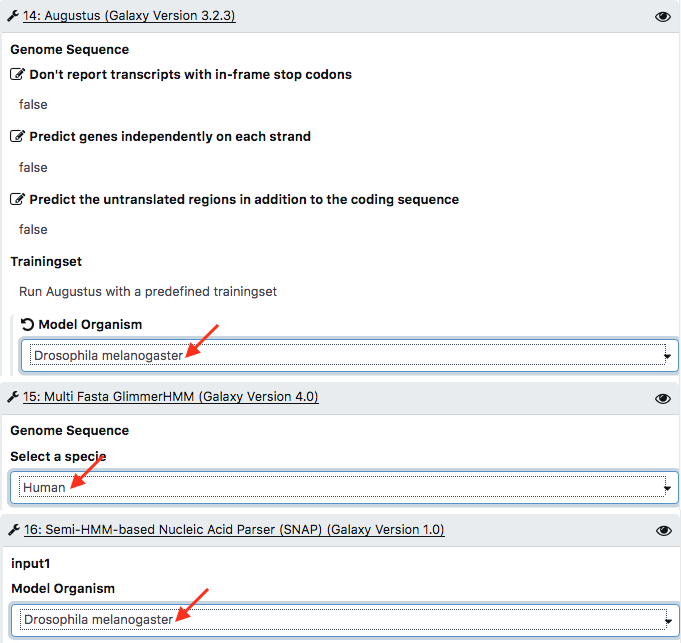


Figure 33: Change the model organism parameters for the gene prediction tools (i.e. Augustus, GlimmerHMM, and SNAP).

Click on the “**Run workflow**” button to launch the workflow. After all the steps in the modified G-OnRamp workflow have completed (which will take several minutes), the new *D. biarmipes* Assembly Hub will be available through the Hub Archive Creator step.

If you are running G-OnRamp on a public server (*e.g.*, launch G-OnRamp on the cloud via CloudLaunch), then you can click on the “**main**” link in the Hub Archive Creator step to view the new Assembly Hub on the UCSC Genome Browser (Figure 34).

If you are using the G-OnRamp virtual appliance, then the “main” link will not work because the G-OnRamp server is not publicly accessible. You can view the new Assembly Hub using the hosting options discussed in Section 5 of the “Introduction to G-OnRamp Walkthrough”.

If you are using the G-OnRamp virtual appliance, you can also view the sample Assembly Hub at <http://g-onramp.org/dbia3-rnaseq-hub>



Figure 34: View the Dbia3 genome assembly and the evidence tracks produced by the Hub Archive Creator on the UCSC Genome Browser.

# 4. Remove an evidence track from the Hub Archive Creator

Make a copy of “Customized G-OnRamp” workflow and rename the new workflow as “**G-OnRamp: Remove WindowMasker track**”. Click on the down arrow next on the new workflow and then select the “**Edit**” option to go to the Workflow Canvas.

To delete a tool from the workflow, click on the “x” at the top right corner of that tool. The tool and its connections will be removed from the workflow. For example, click on the “**x**” at the top right corners of the two WindowMasker tools to remove them from the workflow (Figure 35).

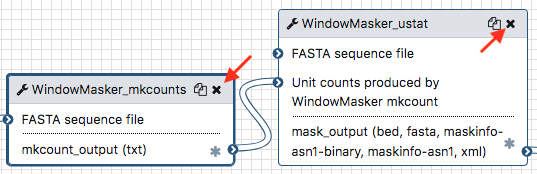


Figure 35: Click on the “x” at the top right corners of the WindowMasker\_mkcounts and the WindowMasker\_ustat tools to remove them from the workflow.

When you delete the WindowMasker tools, the connections between WindowMasker\_ustat and the Hub Archive Creator will also be removed. However, the “New group 1 > New track 2 > Generic Bed File Choice” entry in Hub Archive Creator will need to be removed manually.

Click on the Hub Archive Creator tool and scroll down to the “**Repeats**” group in the “Details” panel. Click on the “**Trash**” icon at the top right corner of the WindowMasker track to delete the track (Figure 36). Remember to save the changes to the workflow before you leave the Workflow Canvas.

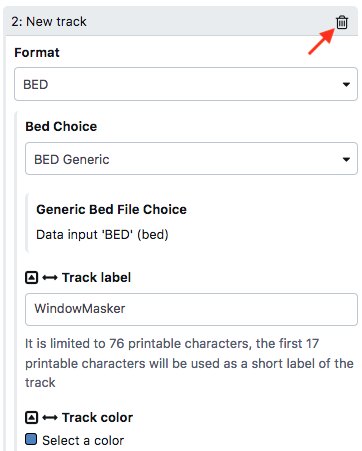


Figure 36: Click on the trash icon to remove the “WindowMasker” track from the Hub Archive Creator.

# 5. Add custom tracks to the Genome Browser

The G-OnRamp workflow performs analyses on the genome assembly and RNA-Seq reads to produce multiple complementary evidence tracks (i.e. sequence similarity, *ab initio* gene predictions, RNA-Seq analysis, repeats identification) for the UCSC Genome Browser and JBrowse. In the previous sections, we learned how to add the WindowMasker tools and the RNA-Seq analysis tools to the G-OnRamp workflow in order to run additional analyses and incorporate their results into the genome browser.

In addition to customizing the G-OnRamp workflow, you can also customize the genome browsers produced by G-OnRamp by adding custom track files directly (*e.g.*, incorporate results from other genome analysis tools such as MAKER). The steps for adding a custom track are 1) upload the custom track file to Galaxy; and 2) run the browser generator tool (**Hub Archive Creator** for generating UCSC Assembly Hub or **JBrowse Archive Creator** for generating JBrowse genome browsers). We will illustrate these steps by adding a RNA-Seq sequence coverage track from the *D. biarmipes* embryos (egg) sample to the UCSC Assembly Hub.

## 5.1 Upload the custom track file to the “G-OnRamp: Dbia3\_RNA-Seq” History

The BigWig file which contains the sequence read coverage for the *D. biarmipes* egg sample (“**D.biarmipes\_egg\_shallow\_sequence\_coverage.bigwig**”) is available in the “**data**” directory of the exercise package. Use the “**Upload File**” tool in the Tools panel (under the “Get Data” section) to transfer the file to the G-OnRamp server (Figure 37). The new History item will turn green after the file upload is complete.

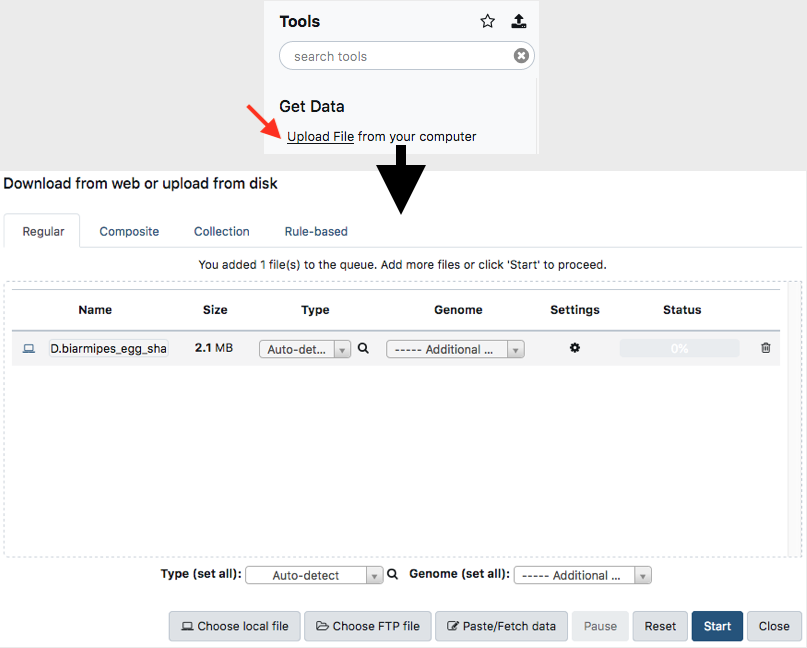


Figure 37: Access the “Upload File” tool from the Tools panel. Add the BigWig file to the upload queue via drag and drop or the “Choose local file” button. Click on the “Start” button to begin the file upload.

If the genome assembly contains scaffold names that are longer than 31 characters, the “rename the scaffolds” tool will either truncate or rename the scaffold names (depending on the user settings), and generate an output dataset with the name mappings between the original and the new scaffold names. You can use this name mapping file with the “rename the tracks” tool to rename the scaffolds in the custom track file. The “rename the tracks” tool supports custom track files in BED, GFF3, GTF, BAM, or BigWig formats.

## 5.2 Run Hub Archive Creator to create a new UCSC Assembly Hub with the custom track

The Hub Archive Creator is a Galaxy tool that is used to prepare your files for Assembly Hub visualization. Find the “**Hub Archive Creator**” in the Tools panel and open the tool. We can specify the name of the Assembly Hub under the “UCSC Genome Browser assembly ID” field, and then select the reference genome assembly under the “Reference genome” field. Evidence tracks are organized by track groups, which can be used to put multiple related evidence tracks together. Click on the “Insert New group” button to create the first track group. Specify the name of the track group in the “Group name” field. Click on the “Insert New track” button to add a new evidence track to the group (Figure 38).

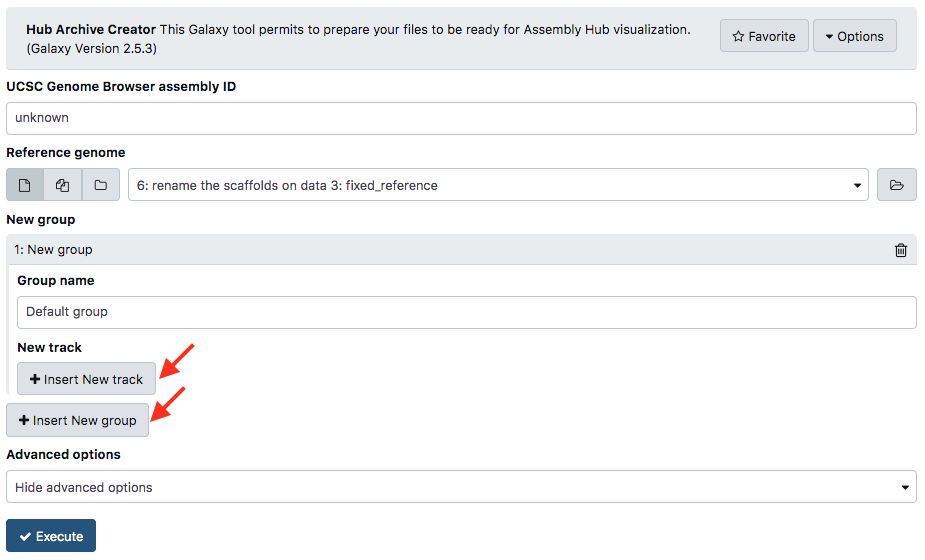


Figure 38: Configure the Hub Archive Creator by specifying the UCSC Genome Browser assembly ID and the reference genome. Use the “Insert New group” button to create a new track group, and use the “Insert New track” button to add an evidence track to a track group.

Since we want to include all the evidence tracks in the previous Assembly Hub, we will modify the previous configuration of Hub Archive Creator instead of adding all the evidence tracks one by one from scratch. Click on the History item produced by Hub Archive Creator (*e.g.,* “**41: Hub Archive Creator on data19, data 18, and others**”), and then click on the “**Run this job again**” icon (Figure 39). The configuration page will appear with the input datasets and parameters used in the previous analysis. (The configuration form might take several seconds to load.)

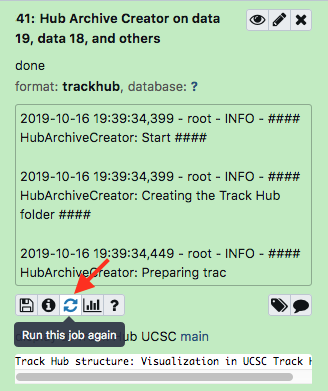


Figure 39: Click on the “Run this job again” icon (red arrow) to open the configuration page for the Hub Archive Creator with all the previous parameter settings.

The next step is to add the BigWig file from the *D. biarmipes* egg RNA-Seq sample to the “RNA-Seq Analysis” track group of the Hub Archive Creator. Find the “**RNA-Seq Analysis**” group in the track configuration form, and then scroll down to the end of this group. Click on the “**Insert New track**” button to add a new track configuration block (Figure 40).

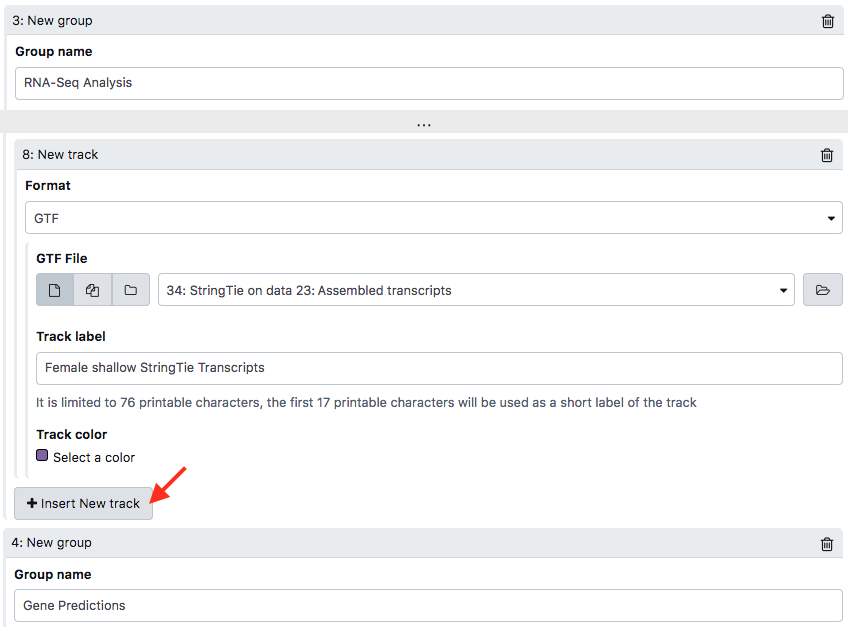


Figure 40: Scroll down to the end of the “RNA-Seq Analysis” group and then click on “Insert New track”

Select the “**BIGWIG**” option under the “Format” field. Select the custom track dataset (**D.biarmipes\_egg\_shallow\_sequence\_coverage.bigwig**) in the “BIGWIG file” field. Change the “Track label” to “**Egg shallow Sequence Coverage**”, and then click on “Select a color” and change the color of the track to orange (Figure 41).

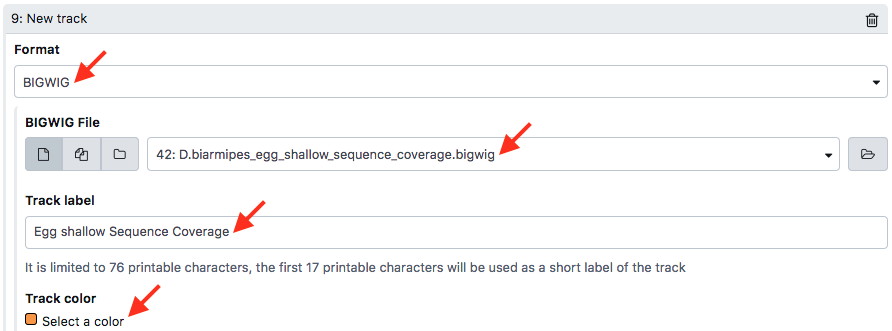


Figure 41: In the new track configuration block, select “BIGWIG” for the track file format and select the BigWig file for the *D. biarmipes* egg RNA-Seq sample as the input file. You can also customize the label and color for this new track.

Finally, scroll down to the end of the configuration blocks and click on the “**Execute**” button. The Hub Archive Creator will create a new Assembly Hub for *D. biarmipes* that includes the “Egg shallow Sequence Coverage” evidence track under the “RNA-Seq Analysis” group.