

CryptoGenotyper Tutorial

Easy Analysis of *Cryptosporidium* Sanger Sequences in Galaxy

CryptoGenotyper is a bioinformatics tool developed by the National Microbial Laboratory (NML) of the Public Health Agency of Canada (PHAC). This tool analyzes raw Sanger sequencing data for the two most common *Cryptosporidium* gene targets: SSU/18S rRNA and *gp60*. This quick and easy to use tool can identify the species and subtype of your samples, create a sample Fasta file, and more. Datasets are compared against an internal reference database and outputs are created in standard nomenclature. This tool is available on the public Galaxy servers, [Galaxy Europe](#) and [VEuPathDB](#), or as a standalone [Bioconda](#) package.

Please cite the following publication if you find this subtyping tool useful in your work.

Yanta, C.A., Bessonov, K., Robinson, G., Troell, K., Guy, R.A. (2021) CryptoGenotyper: A new bioinformatics tool for rapid *Cryptosporidium* identification. *Food and Waterborne Parasitology*, 23:e00115. doi: [10.1016/j.fawpar.2021.e00115](https://doi.org/10.1016/j.fawpar.2021.e00115)

This tutorial will give you a walkthrough of how the CryptoGenotyper can be used and how to interpret the results in Galaxy.

There are 5 main Sections to this Tutorial.

- 1: Uploading Sanger Sequencing Data**
- 2: Preparing Datasets for Run**
- 3: Running the CryptoGenotyper Tool**
- 4: Using CryptoGenotyper Workflows**
- 5: Analyzing Results**

1) Uploading Sanger Sequencing Data

The CryptoGenotyper accepts raw Sanger sequencing chromatogram data files in .ab1 file format, representing the SSU rRNA or *gp60* gene targets. This data can correspond to three different sequencing read formats:

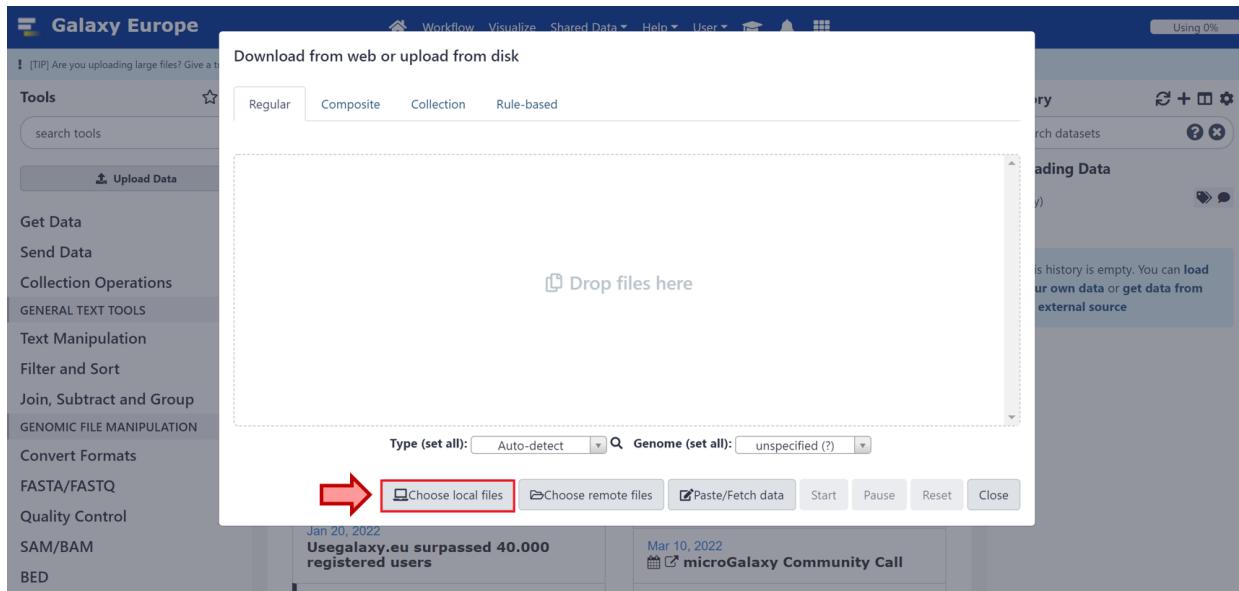
- 1) Forward (5'-3')
- 2) Reverse (3'-5')
- 3) Contig (forward and reverse reads)

This Sanger sequencing data must first be uploaded into Galaxy using the **Upload Data** tool on the left hand side bar on the home screen.

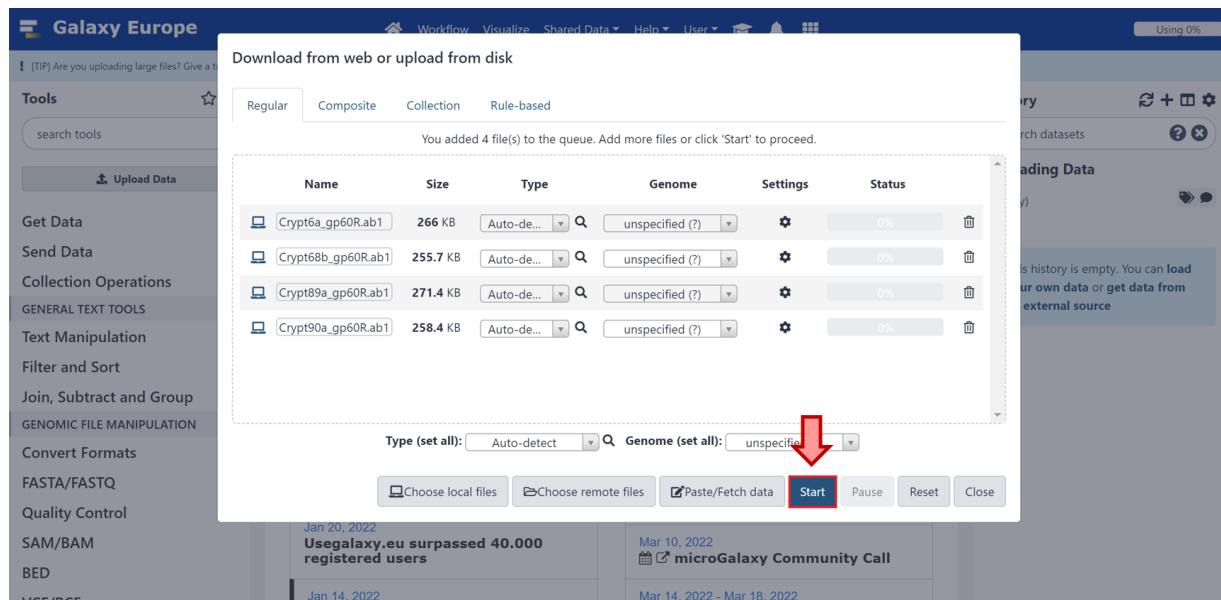
The screenshot shows the Galaxy Europe interface. On the left, a sidebar lists various tools under categories like Tools, Collection Operations, and General Text Tools. The 'Upload Data' button is highlighted with a red box and a red arrow pointing to it from the bottom-left. The main content area features a COVID-19 Research section, a News feed with two items, and an Events feed with one item. The right side shows a History panel which is currently empty. The top navigation bar includes links for Workflow, Visualize, Shared Data, Help, User, and a search bar.

A screen will appear to download the data from the web or upload from your computer. If the Sanger sequences are stored on your computer, select **Choose local files** and choose the files you wish to upload to Galaxy for analysis.

**Note: these files must be in the .ab1 file format*



The file browser will appear. Select all Sanger chromatograms that are to be analyzed and select **Open**. You should then see all the files listed. To upload into Galaxy, select **Start**.



After the status for all files have reached 100% (green), you can select **Close**. When the upload is finished on the Galaxy server, the files will appear green in the **History panel**.

The screenshot shows the Galaxy Europe web interface. A central modal window titled "Download from web or upload from disk" displays a table of uploaded files. The table has columns for Name, Size, Type, Genome, Settings, and Status. All four files listed (Crypt6a_gp60R.ab1, Crypt6b_gp60R.ab1, Crypt89a_gp60R.ab1, Crypt90a_gp60R.ab1) have a status of 100% and a green checkmark icon. Below the table are search and filter options: "Type (set all): Auto-detect", "Genome (set all): unspecified (?)", and buttons for "Choose local files", "Choose remote files", "Paste/Fetch data", "Start", "Pause", "Reset", and "Close". A red arrow points to the "Close" button. In the background, the main Galaxy interface shows the "History" panel with a list of datasets and a legend explaining the color coding for file status.

Name	Size	Type	Genome	Settings	Status
Crypt6a_gp60R.ab1	266 KB	Auto-de...	unspecified (?)		100% ✓
Crypt6b_gp60R.ab1	255.7 KB	Auto-de...	unspecified (?)		100% ✓
Crypt89a_gp60R.ab1	271.4 KB	Auto-de...	unspecified (?)		100% ✓
Crypt90a_gp60R.ab1	258.4 KB	Auto-de...	unspecified (?)		100% ✓

Legend

- Green = job finished
- Orange = job running
- Grey = job waiting to run
- Red = error

Files are uploaded once they are highlighted green.

2) Preparing Datasets for Run

If you have only one forward or reverse chromatogram you wish to run please proceed to **Section 3.**

Before further analysis can occur on multiple chromatograms, the data will need to be organized based on the relationships within the data. Three scenarios you may have are as follows:

- A) One forward and one reverse chromatogram- build a dataset pair
- B) Multiple forward OR reverse chromatograms- build a dataset list
- C) Multiple forward AND reverse pairs of chromatograms- build a list of dataset pairs

Scenario A: Building a Dataset Pair

Upload your forward and reverse chromatograms as described in **Section 1.**

Once the datasets have loaded select the checkbox **Operations on Multiple Datasets**.

Then select the checkboxes for your forward and reverse datasets.

The screenshot shows the Galaxy Europe web interface. On the left, there's a sidebar with 'Tools' (including 'crypto') and 'Upload Data'. The main workspace has two sections highlighted with red boxes: 'Step 1: Select "Operations on Multiple Datasets"' and 'Step 2: Select forward and reverse datasets'. In Step 1, there's a dropdown menu for 'Type of Sequences' with 'Paired Sequencing File(s)' selected. In Step 2, there's a dropdown menu for 'Output header line in the report?' with 'Yes' selected. To the right, the 'History' panel shows an unnamed history with two datasets: '2: Crypt80a_SSUR.ab1' and '1: Crypt80a_SSUF.ab1'. Red arrows point from the text labels 'Step 1: Select "Operations on Multiple Datasets"' and 'Step 2: Select forward and reverse datasets' to their respective sections in the interface.

Click the **For all selected...** button, then select **Build Dataset Pair**.

The screenshot shows the Galaxy Europe web interface. On the left, the 'Tools' sidebar lists various bioinformatics tools like 'nhmmer', 'hmmscan', 'hmmsearch', 'jackhmmer', 'hmmscan', and 'phmmmer'. In the center, the 'CryptoGenotyper' tool is selected, with its parameters set: 'Marker' to 'SSU rRNA', 'Type of Sequences' to 'Contig', and 'Paired Sequencing File(s)' dropdown showing 'No ab1 dataset collection available.'. Below these are options for 'Output header line in the report?' (Yes), 'Email notification' (off), and a checked 'Execute' button. On the right, the 'History' panel shows two datasets: '2: C' and '1: C'. A red arrow points to the 'For all selected...' button in the context menu that appears when clicking on the '1: C' dataset. Another red arrow points to the 'Build Dataset Pair' option in the same context menu.

IMPORTANT Ensure your forward chromatogram is in the row labeled (**forward:**) and your reverse chromatogram is in the row labeled (**reverse:**). If they are not in the correct row, select the **Swap** button.

Give it a memorable name, then select **Create collection**.

This screenshot shows the 'Create a collection from a pair of datasets' dialog. It contains fields for 'forward' and 'reverse' chromatograms, both currently set to 'forward: Crypt80a_SSUF.ab1'. A red arrow points to the 'Swap' button. Another red arrow points to the 'Name:' field, which is filled with 'Crypt80a_SSU'. A third red arrow points to the 'Create collection' button at the bottom right of the dialog.

Your dataset pair is now ready to be run. You can proceed to **Section 3**.

Scenario B: Building a Dataset List

Upload your forward OR reverse chromatograms as described in **Section 1**.

Once the datasets have loaded select the checkbox **Operations on Multiple Datasets**.

Then select the checkboxes for your forward or reverse datasets.

The screenshot shows the Galaxy Europe web interface. On the left, a sidebar lists various tools and operations. In the center, there's a news section with 'COVID-19 Research!' and an 'UPDATE 2' notice about maintenance. On the right, a 'History' panel displays a list of datasets with checkboxes. A red box highlights the 'Operations on multiple datasets' checkbox in the history panel, and another red arrow points to the 'Select all datasets' button in the update notice. The history list contains four entries:

- 4: Crypt90a_gp60R.ab1 (checked)
- 3: Crypt89a_gp60R.ab1 (checked)
- 2: Crypt68b_gp60R.ab1 (checked)
- 1: Crypt6a_gp60R.ab1 (checked)

Click the **For all selected...** button, then select **Build Dataset List**.

The screenshot shows the Galaxy Europe web interface. On the left, there's a sidebar with various tools like 'Search tools', 'Upload Data', 'Get Data', 'Send Data', etc. The main area has sections for 'COVID-19 Research', 'UPDATE 2 - Not any more limited computing capacity on next 25-26.01.2022', 'News', and 'Events'. On the right, the 'History' panel lists four datasets. A red arrow points to the 'For all selected...' button, and another red arrow points to the 'Build Dataset List' option in the dropdown menu.

Give it a memorable name, then select **Create collection**.

The screenshot shows the 'Create collection from a list of datasets' dialog box. It lists several datasets under 'Start over' (e.g., Crypt90a_gp60R.ab1, Crypt89a_gp60R.ab1, Crypt68b_gp60R.ab1, Crypt6a_gp60R.ab1). A red arrow points to the 'Name:' input field where 'gp60 reverse datasets' is typed. Another red arrow points to the 'Create collection' button at the bottom right.

Your dataset list is now ready to be run. You can proceed to **Section 3**.

Scenario C: Building a List of Dataset Pairs

Upload your forward AND reverse chromatograms as described in **Section 1**.

Once the datasets have loaded select the checkbox **Operations on Multiple Datasets**.

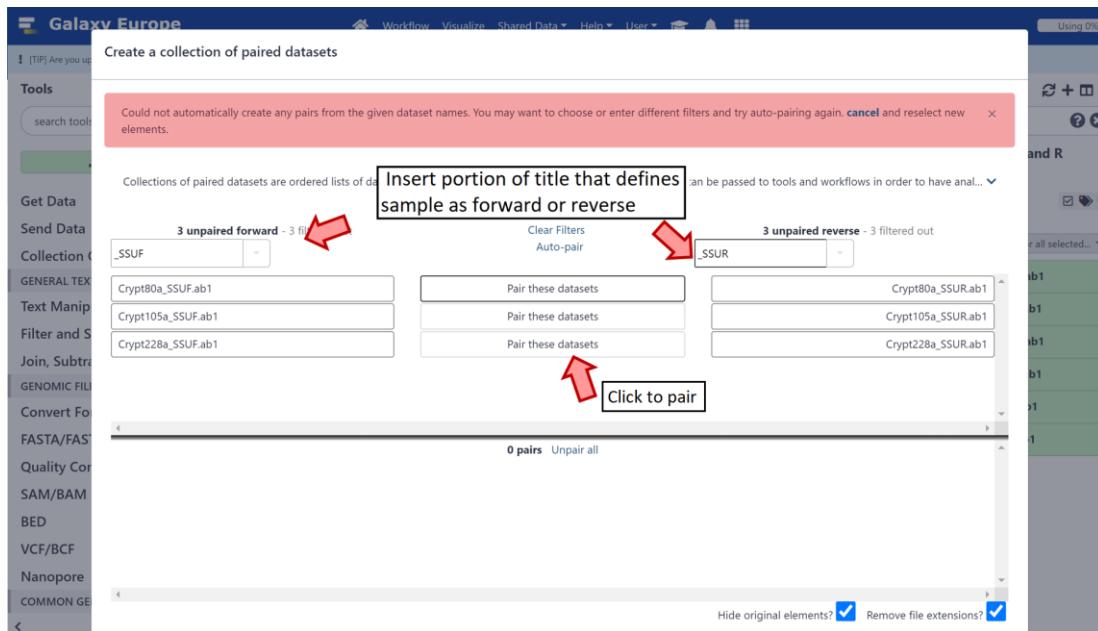
Then select the checkboxes for all datasets.

The screenshot shows the Galaxy Europe web interface. On the left, there's a sidebar with various tools like 'Upload Data', 'Get Data', 'Send Data', and 'Collection Operations'. In the main area, there's a 'COVID-19 Research!' section and a 'News' section with recent updates. On the right, a 'History' panel lists datasets under 'CryptoGenotyper F and R'. A red arrow points to the 'Operations on multiple datasets' checkbox in the top right corner of the history panel. Another red arrow points to the 'Select All Datasets' button below it. The history list contains six datasets, all of which are checked.

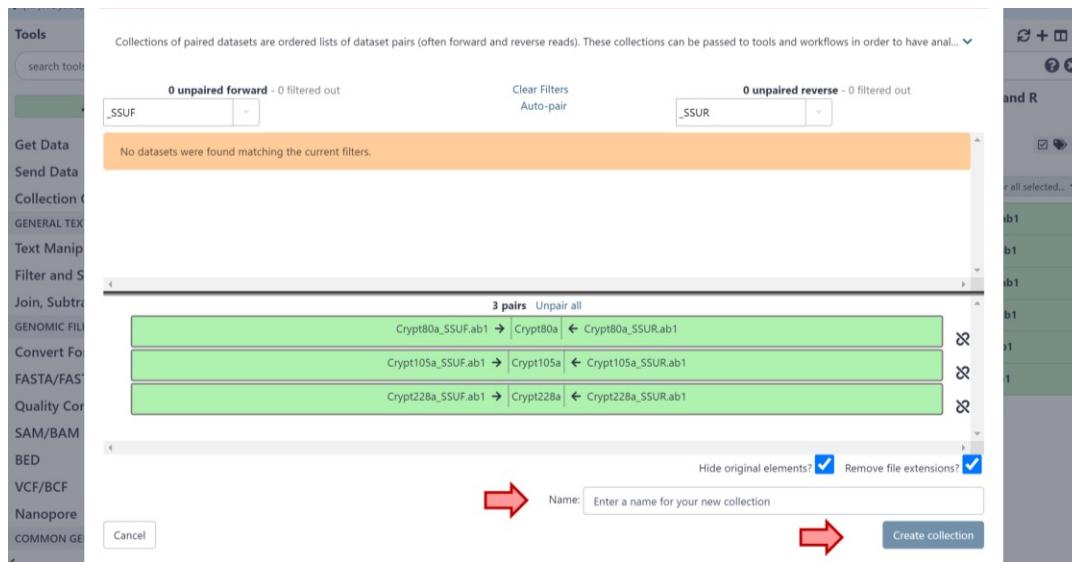
Click the **For all selected...** button, then select **Build List of Dataset Pairs**.

This screenshot shows the same Galaxy Europe interface as above, but with a different focus. A red arrow points to the 'For all selected...' button in the 'History' panel. Another red arrow points to the 'Build List of Dataset Pairs' option in the context menu that appears when clicking on one of the selected datasets. The context menu also includes other options like 'Hide datasets', 'Unhide datasets', and 'Delete datasets'.

Insert the portion of the title that defines each dataset as either forward or reverse. Then click **Pair these datasets**. If they were automatically paired correctly you can skip this step.



Hide original elements? and **Remove file extensions?** will be autoselected. Once paired, give the collection a memorable name, then select **Create collection**.



Your list of dataset pairs is now ready to be run. Proceed to **Section 3**.

3) Running the CryptoGenotyper Tool

Once the datasets are uploaded and formatted, you can now run the CryptoGenotyper tool to analyze your results.

The three following scenarios can be used for either SSU rRNA or *gp60* datasets; simply change the **Marker** designation to the gene of interest.

Scenario A: One forward and reverse chromatogram formatted into a dataset pair

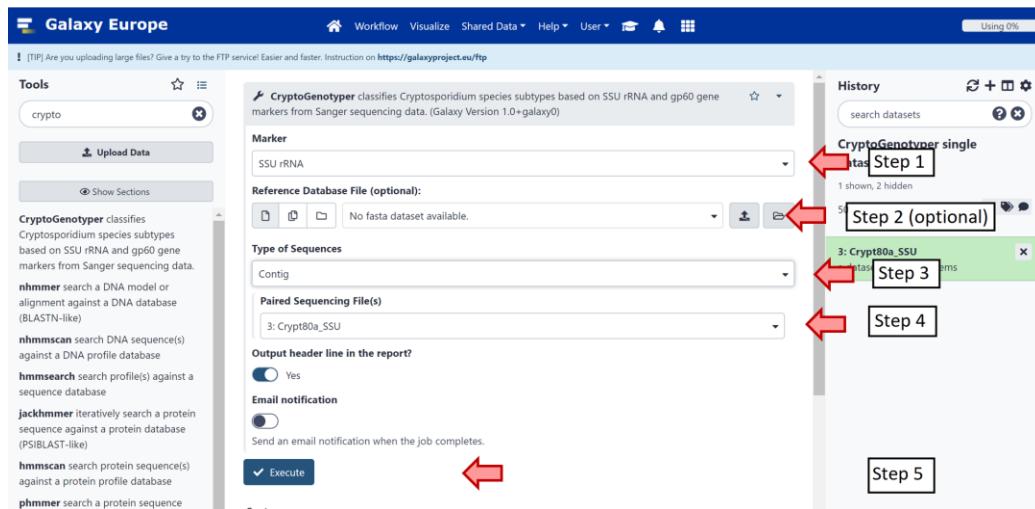
Step 1: Select the correct **Marker** (SSU rRNA or *gp60*).

Step 2: (Optional) Choose a **Reference Database File** if you are not using the pre-installed database.

Step 3: Select **Type of Sequences** (in this example “Contig”).

Step 4: Choose your file under **Paired Sequencing File(s)** if it is not automatically selected.

Step 5: Click **Execute**



To learn about viewing and interpreting your results, proceed to **Section 5: Analyzing Results**.

Scenario B: Multiple forward OR reverse chromatograms in a dataset list

*To have outputs concatenated for bulk submissions, please go to **Section 4: Using CryptoGenotyper Workflows**

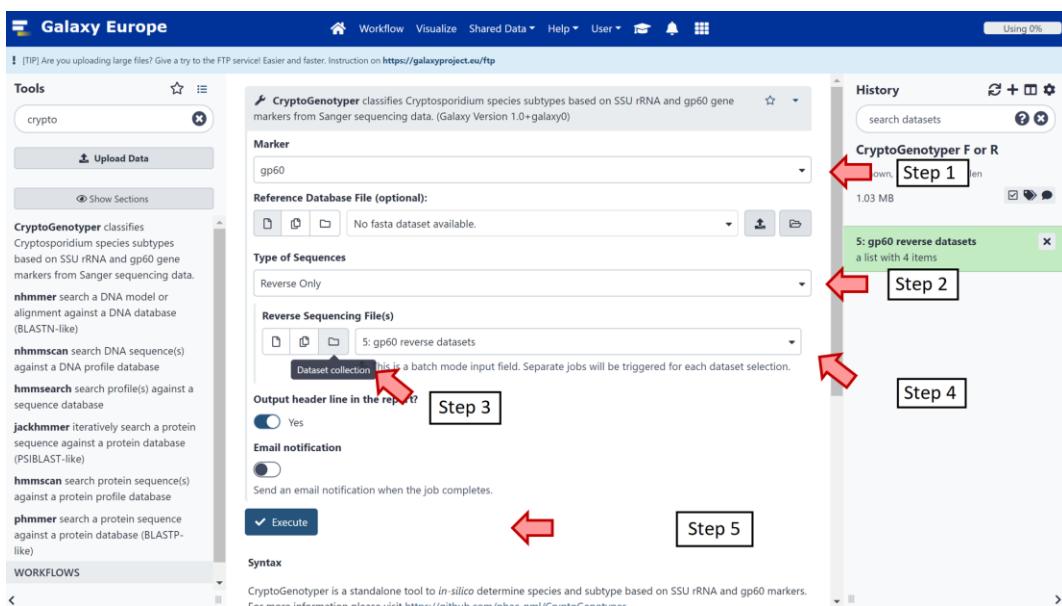
Step 1: Select the correct **Marker** (SSU rRNA or *gp60*).

Step 2: Select **Type of Sequences** (in this example “Reverse Only”).

Step 3: Click “**Dataset Collection**” Button (shown below)

Step 4: Choose your collection under **Reverse Sequencing File(s)** if it is not automatically selected.

Step 5: Click **Execute**



To learn about viewing and interpreting your results, proceed to **Section 5: Analyzing Results**.

Scenario C: Multiple forward AND reverse pairs of chromatograms in a list of dataset pairs

*To have outputs concatenated for bulk submissions, please go to **Section 4: Using CryptoGenotyper Workflows**

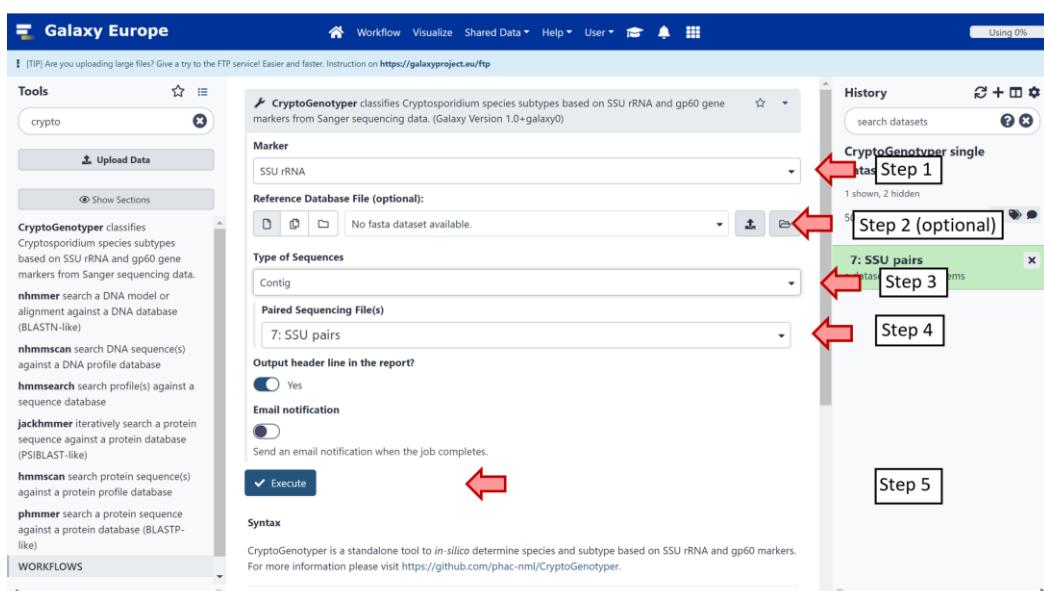
Step 1: Select the correct **Marker** (SSU rRNA or gp60).

Step 2: (Optional) Choose a **Reference Database File** if you are not using the pre-installed database.

Step 3: Select **Type of Sequences** (in this example “Contig”).

Step 4: Choose your collection under **Reverse Sequencing File** if it is not automatically selected.

Step 5: Click **Execute**



Proceed to **Section 5: Analyzing Results** to learn more about analyzing your outputs.

4) Using CryptoGenotyper Workflows

There are six workflows available to make bulk submission easier:

gp6o- Forward Only

gp6o- Reverse Only

gp6o- Contigs

SSU rRNA- Forward Only

SSU rRNA- Reverse Only

SSU rRNA- Contigs

If you are using Galaxy VEuPathDB, the workflows are available under **Shared Data, Workflows** as shown below.

The screenshot shows the VEuPathDB Galaxy Site interface. At the top, there is a navigation bar with links for Analyze Data, Workflow, Visualize, Shared Data (with a red arrow pointing to it), User, and a grid icon. Below the navigation bar, there is a banner for VEuPathDB with an attention message about dataset purging. On the left, there is a sidebar with sections for Tools, VEU PATHDB APPLICATIONS (including VEuPathDB Export Tools, OrthoMCL Tools, RNA-Seq Tools), DATA TRANSFER (Globus Data Transfer, Get Data, Collection Tools, AtlasXomics tools), and a footer section. The main content area has a title "Welcome to the VEuPathDB Galaxy Site" and a subtitle "A free, interactive, web-based platform for large-scale data analysis". It also lists "With VEuPathDB Galaxy you can:" followed by two numbered steps. To the right, there is a "History" panel showing datasets like "Gp60_R" and a list of recent datasets: 38: Concatenate datasets on data 37, data 35, and others; 21: CryptoGenotyper:collection 19:reports; and 20: CryptoGenotyper:collection 19:fastas. A red arrow points to the "Workflows" link in the Shared Data menu.

Use the **Search Bar** or **Scroll Down** to the workflow you require.

The screenshot shows the Globus Genomics interface with the 'Published Workflows' section. On the left, there's a sidebar with various tools like VEuPathDB Applications, Data Transfer, and NGS Visualization. The main area displays a list of workflows. A red arrow points to the search bar at the top left. Another red arrow points to the dropdown menu next to the first workflow entry.

Name	Annotation	Owner	Community Rating
CryptoGenotyper v1.1 - SSU rRNA - contig (F and R)	christine-yanta-611733273		★★★★★
CryptoGenotyper v1.1 - SSU rRNA - reverse reads	christine-yanta-611733273		★★★★★
CryptoGenotyper v1.1 - SSU rRNA - forward reads	christine-yanta-611733273		★★★★★
CryptoGenotyper v1.1 - gp60 - forward reads	christine-yanta-611733273		★★★★★
CryptoGenotyper v1.1 - gp60 - contig (F and R)	christine-yanta-611733273		★★★★★
CryptoGenotyper v1.1 - gp60 - reverse reads	christine-yanta-611733273		★★★★★
DESeq2 Workflow for paired-end stranded reads (v.7)	eupathdb-1926010		★★★★★

On the right, there's a 'History' panel showing a list of recent datasets and workflows, such as 'Gp60_R' and '38: Concatenate dataset'. A red arrow points to the 'Run' button in the dropdown menu for the first workflow entry.

Select the **Down Arrow** next to the workflow of interest, then select either **Run** or **Import**.

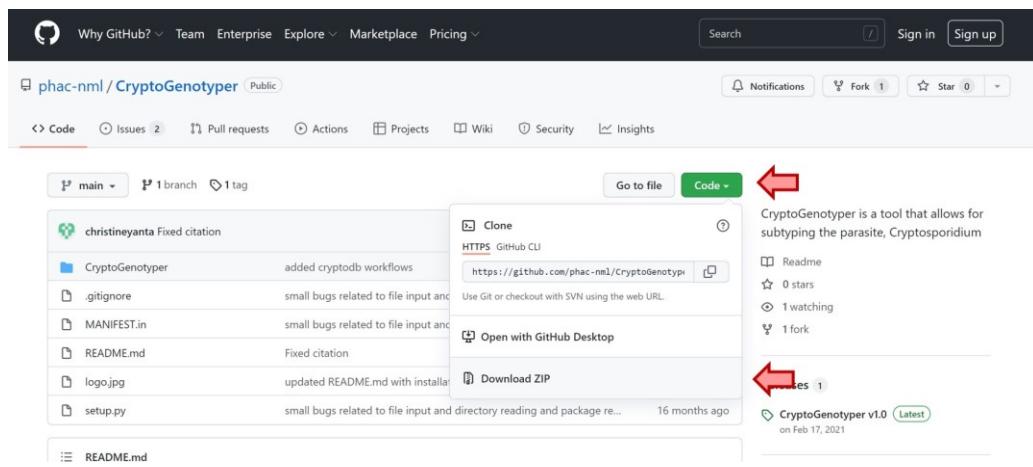
The screenshot shows the same interface as the previous one, but with a dropdown menu open next to the first workflow entry. A red arrow points to the dropdown menu itself. Another red arrow points to the 'Run' button within that menu.

Name	Annotation	Owner	Community Rating
CryptoGenotyper v1.1 - SSU rRNA - contig (F and R)	christine-yanta-611733273		★★★★★
CryptoGenotyper v1.1 - SSU rRNA - reverse reads	christine-yanta-611733273		★★★★★
CryptoGenotyper v1.1 - SSU rRNA - forward reads	christine-yanta-611733273		★★★★★

If you are using Galaxy Europe, download the workflows from Github as described below.

Download the six available workflows from Github: <https://github.com/phac-nml/CryptoGenotyper>

Click the green **Code** button, then select the **Download ZIP** option.



The example shown below for uploading and running bulk SSU rRNA contigs can be easily adapted to perform any other bulk runs with the six available workflows. Simply upload the required workflow, and run as described below.

To use your downloaded workflows, begin by uploading your bulk data and formatting as described in **Sections 1 and 2**.

Select the **Workflow** tab then select **Import**.

The screenshot shows the Galaxy Europe web interface. The top navigation bar includes links for Home, Workflow, Visualize, Shared Data, Help, User, and a menu icon. The 'Workflow' tab is currently selected. In the center, there's a search bar labeled 'Search Workflows' with a red arrow pointing to it. Below the search bar are buttons for '+ Create' and 'Import'. A tooltip above the 'Import' button says 'Chain tools into workflows'. To the right of the search area is a 'History' sidebar showing a workflow named 'CryptoGenotyper workflow F and R' with a size of 1.49 MB. At the bottom right, there's a green box labeled '7: SSU pairs' containing a list of items.

Browse and select the appropriate workflow file from your computer, then select **Import Workflow**. In this example SSU rRNA contigs are being used, so the Workflow that included SSU_rRNA_-_Contig_(F_and_R) in its title was imported.

This screenshot shows the 'Import Workflow' dialog box overlaid on the Galaxy Europe interface. The dialog has several sections: 'Import Workflow' (instructions to provide a Galaxy workflow export URL or a workflow file), 'Archived Workflow URL' (input field), 'Archived Workflow File' (input field with 'No file chosen' and a 'Browse' button, with a red arrow pointing to it), 'Import workflow' (button), 'Import Workflow from Configured GA4GH Tool Registry Servers (e.g. Dockstore)' (instructions and input field), and 'Import a Workflow from myExperiment' (link). The background shows the same Galaxy interface as the first screenshot, with the 'Workflow' tab selected.

Select the **Run Workflow** “play” button.

The screenshot shows the Galaxy Europe web interface. On the left, there's a sidebar with 'Tools' expanded, showing categories like 'GENERAL TEXT TOOLS' and 'GENOMIC FILE MANIPULATION'. In the center, a workflow titled 'SSU rRNA - Contig (F and R) - CryptoGenotyper ver 1.1 (imported from uploaded file)' is listed. To its right is a 'History' panel containing a dataset named 'CryptoGenotyper workflow F and R'. At the bottom right of the workflow card is a 'Run workflow' button, which has a red arrow pointing to it. A green box highlights the '7: SSU pairs' dataset in the history panel.

Ensure the correct collection of datasets is selected in **1: Input Dataset Collection**, then select **Run Workflow**.

This screenshot shows the configuration page for the workflow. It includes sections for 'History Options' (with a red arrow pointing to the 'Run Workflow' button), '1: Input dataset collection' (with a red arrow pointing to the '7: SSU pairs' dataset), and other parameters like 'Marker' and 'Reference Database File (optional)'. A green box highlights the '7: SSU pairs' dataset in the history panel.

Three outputs are created; Collection:Fasta, Collection:Report, and **Concatenated Datasets**.

Select the “eye” icon to view the concatenated file. Note that the concatenated file lacks column descriptors. Run a single dataset using the CryptoGenotyper tool to show the column titles.

The screenshot shows the Galaxy Europe web interface. On the left, the 'Tools' sidebar is open, displaying categories like 'GENERAL TEXT TOOLS', 'GENOMIC FILE MANIPULATION', and 'COMMON GENOMICS TOOLS'. In the center, a table displays three datasets: 'Crypt80a_forward', 'Crypt105a_forward', and 'Crypt228a_forward'. To the right, the 'History' panel shows the results of a 'CryptoGenotyper workflow F and R' run. It lists three items: '16: Concatenate datasets on data 15, data 13, and data 11' (with a red arrow pointing to it), '9: CryptoGenotyper:collection 7:reports' (with a red arrow pointing to it), and '8: CryptoGenotyper:collection 7:fastas'. A red arrow also points to the '16' item in the history list. The 'History' panel also indicates '4 shown, 6 hidden' and a total size of '1.49 MB'.

Continue to **Section 5: Analyzing Results** for tips to understand your data and potential Warning Messages.

5) Analyzing Results

A) Viewing Reports or Fasta Files

When there is only one Report and Fasta File output present, click the **View Data** button as shown below.

The screenshot shows the Galaxy Europe web interface. On the left, the 'Tools' sidebar is open, showing categories like 'GENERAL TEXT TOOLS', 'Text Manipulation', and 'Filter and Sort'. In the center, a table displays two sequence entries: 'Crypt80a_SSU_forward' and 'Crypt80a_SSU_reverse'. The right side shows the 'History' panel with a single dataset named 'CryptoGenotyper single dataset'. A red arrow points to the 'View data' button next to the dataset entry.

When there is a collection of Reports and Fasta Files, first select the **Report Collection** or **Fasta Collection** as shown below.

The screenshot shows the Galaxy Europe web interface. The 'Tools' sidebar is open, showing categories like 'GENERAL TEXT TOOLS', 'Text Manipulation', and 'Filter and Sort'. A green callout box highlights the 'COVID-19 Research!' section, which contains information about SARS-CoV-2 data analysis and a quote from Stephen Hawking. Below this, the 'News' and 'Events' sections are visible. On the right, the 'History' panel shows a collection of datasets: '9: CryptoGenotyper:collection 7:reports', '8: CryptoGenotyper:collection 7:fastas', and '7: SSU pairs'. Red arrows point to the 'View data' buttons for the 'reports' and 'fastas' collections.

Then click the **View Data** button of the Report or Fasta File you wish to view as shown below.

Galaxy Europe

[TIP] Are you uploading large files? Give a try to the FTP service! Easier and faster. Instruction on <https://galaxyproject.eu/ftp>

Sample Name	Type of Sequences	Mixed?	Species	Sequence
Crypt80a_forward	contig	Yes	C.parvum	ATCACATTAATGTGACATATCATTCAAGTTCTGACCTATCAGCTT,
		Yes	C.parvum	ATCACATTAATGTGACATATCATTCAAGTTCTGACCTATCAGCTT,

History

◀ Back to CryptoGenotyper F and R
CryptoGenotyper:collection 7:reports
 a list with 3 items

Scroll to the right to view the entire report.

Galaxy Europe

[TIP] Are you uploading large files? Give a try to the FTP service! Easier and faster. Instruction on <https://galaxyproject.eu/ftp>

Sample Name	Type of Sequences	Mixed?	Species	Sequence
Crypt80a_forward	contig	Yes	C.parvum	ATCACATTAATGTGACATATCATTCAAGTTCTGACCTATCAGCTT,
		Yes	C.parvum	ATCACATTAATGTGACATATCATTCAAGTTCTGACCTATCAGCTT,

Galaxy Europe

[TIP] Are you uploading large files? Give a try to the FTP service! Easier and faster. Instruction on <https://galaxyproject.eu/ftp>

Comments	Bit Score	Query Length (bp)	Query Coverage	E-value	Percent Identity	Accession Number
KGATACCGT	683.0	698	100%	0.0	99.3%	KT948751.1
	709.0	715	100%	0.0	99.7%	KM012040.1

Tools

◀ Back to CryptoGenotyper F and R
CryptoGenotyper:collection 7:reports
 a list with 3 items

Scroll to the right to see the remaining column details

B) Understanding **Report** Results

Not all columns will be present for each analysis type. **Subtype** and **Avg. Phred Quality** will be present only for *gp6O* datasets, while **Mixed?** will only be present for SSU rRNA datasets.

Sample Name: The name of the dataset that was uploaded.

Type of Sequences: This shows the sequencing read format of the dataset (Forward, Reverse, or Contig).

Mixed?: (Yes/No) If the chromatogram represents a mixed sample, 2 or more lines of data will be present.

Species: This is where the species name can be found, if the sample matched an entry in the database.

Subtype: Subtype will be displayed on *gp6O* datasets if a match is present.

Sequence: The transcribed sequence of the chromatogram.

Comments: Warnings will be present here. See the table below for explanations.

Avg. Phred Quality: A measure of nucleobase identification quality.

Bit Score: A measure of sequence similarity.

Query Length (bp): Length of the chromatogram in base pairs.

Query Coverage: The percentage of the input length that aligns with the database reference.

E-value: The number of nucleobase hits of similar quality likely to occur randomly.

Percent Identity: Percentage sequence match with database sequence.

Accession Number: The unique identifier of the database reference sequence.

***IMPORTANT:** The last six columns (**Bit Score**, **Query Length**, **Query Coverage**, **E-value**, **Percent Identity**, and **Accession Number**) are the results of performing a homology search (BLAST) of the chromatogram against the internal database. If there is not a 100% match further analysis is strongly advised, such as a BLAST against the nr database on NCBI.

C) CryptoGenotyper Warnings and their Explanations

Gene target	Warning message	Explanation
SSU and <i>gp60</i>	Could not analyze chromatogram. Please check manually	Program cannot interpret the chromatogram file
	Poor Sequence Quality. Check manually	The sequence has low Phred score and cannot decipher the data
	No blast hits.	The sequence does not represent a <i>Cryptosporidium</i> gene target in the database
NC		No comment
SSU specific	Average Phred Quality <10, could be other potential mixed seqs. Check manually	A <i>Cryptosporidium</i> species was identified, however due to low Phred quality, further mixed infections could be present.
<i>gp60</i> specific	Not all bases in repeat region had Phred quality ≥ 20	The repeat region is not of the highest quality. Often a result of mixed peaks or sequencing artifacts.
	Could not classify repeat region. Check manually.	The program could not classify the repeat region with certainty. Most likely a cause of sequencing artifacts or the sequence was trimmed into the repeat region.
	BLAST percent identity is less than 99.1%. Check manually in case of new <i>gp60</i> family	The program may have identified a new subfamily that is not described in the database. Check the sequence manually to confirm.