

Galaxy for virologist training

Exercise 1: Introduction to Galaxy

| | |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Galaxy |
| Training dataset: | None |
| Questions: | <ul style="list-style-type: none">How do I create a fasta reference for Crimea Congo?How many nucleotides has each fragment of Crimea Congo genome? |
| Objectives: | <ul style="list-style-type: none">Familiarize with Galaxy websiteUnderstand the Galaxy's historyLearn how to upload data in GalaxyLearn how to visualize data in GalaxyLearn how to run tools in GalaxyLearn how to create a workflowLearn how to load a workflow in Galaxy |
| Estimated time: | 1h 15 min |

When we have to do a bioinformatic analysis using a reference genome, we need to provide **just one reference file**. The problem with segmented genomes, such as Crimea Congo's, is that we have one different file for each fragment in the databases. So here we are going to learn how to load the different segments of a genome in Galaxy and concatenate them in order to create a unique fasta file that can be used for further analyses. Also, we are going to learn how to count the number of sequences in a multifasta file, and the number of nucleotides in each sequence in a fasta file.

1. Galaxy website

First of all go to [Galaxy Web Server in Europe](#) and you will see a display such as this one:

The screenshot shows the Galaxy Europe web interface. The top navigation bar includes links for Workflow, Visualize, Shared Data, Help, Login or Register, and a search bar. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button, followed by a list of genomic file manipulation tools: Get Data, Send Data, Collection Operations, GENERAL TEXT TOOLS, Text Manipulation, Filter and Sort, Join, Subtract and Group, GENOMIC FILE MANIPULATION, Convert Formats, FASTA/FASTQ, Quality Control, SAM/BAM, BED, VCF/BCF, Nanopore, and COMMON GENOMICS TOOLS. The main content area features a 'COVID-19 Research!' box with information about SARS-CoV-2 data and a quote from Prof. Stephen Hawking. Below this are sections for 'News' and 'Events'. The 'News' section lists tool updates for UseGalaxy.eu: 'Nov 6, 2021 UseGalaxy.eu Tool Updates for 2021-11-06', 'Oct 30, 2021 UseGalaxy.eu Tool Updates for 2021-10-30', 'Oct 23, 2021 UseGalaxy.eu Tool Updates for 2021-10-23', and 'Oct 18, 2021 Training Infrastructure Feedback from Dr. Theodora Tsirka'. The 'Events' section lists various hackathons and developer conferences: 'Nov 2, 2021 - Nov 23, 2021 Forces 2021', 'Nov 8, 2021 - Nov 12, 2021 ELIXIR BioHackathon Europe', 'Nov 11, 2021 Galaxy Developer Roundtable: Separated data PVC: How it works and potential missed implications', and 'Nov 16, 2021 - Nov 17, 2021 5. NRZ-Authent Expertinnen- und Expertenworkshop'. The right sidebar is titled 'History' and shows an 'Unnamed history' section with a note that it is empty and can be loaded from an external source.

Where you have 4 different elements: 1. The first one in yellow is the Title panel with the buttons: - Home (house): To go to the home page in Spanish - Workflows: To go to the workflow manager - Visualize: Displays the visualization manager and options - Share Data: Displays the sharing options - Help: Displays all the help menu available - Login or Register - Galaxy Training Materials (graduation cap): Displays de Galaxy Trainings list - Enable/Disable scratchbook (9 squares) 2. The left side panel in blue with all the tools in this Galaxy mirror 3.

Table of Contents

- o [1. Galaxy website](#)
 - [Sign up/Login:](#)
- o [2. Galaxy's history](#)
- o [3. Loading data:](#)
- o [5. Edit and Visualize your data:](#)
 - [Visualization](#)
 - [Edition](#)
- o [6. Run tools](#)
 - [Search](#)
 - [Run tools](#)
 - [Running jobs](#)
 - [Visualize results](#)
- o [7. Furtherly process your data](#)
 - [Results visualization](#)
 - [Share results](#)
- o [8. History management](#)
- o [9. Workflows](#)
 - [Creating workflows](#)
 - [Editing workflows](#)
 - [Sharing workflows](#)
 - [Importing workflows](#)
 - [Running workflows](#)

Central panel in red, which will let you run analyses and view outputs 4. Right panel in green, with the history record.

Sign up/Login:

The first thing we would do is to sign up, so you can save your history. To do that, you should follow the next steps: 1. Select Login or Register in the header panel 2. Select **Register here**. 3. Fill in the registration information. :warning: Use an email you can access now, because it will ask you to confirm your e-mail address. 4. Log into your e-mail, and verify your Galaxy account. 5. Log in with your credentials.

This screenshot shows the Galaxy Europe web interface. The top navigation bar includes links for Workflow, Visualize, Shared Data, Help, and a red-highlighted 'Login or Register' button. The left sidebar contains a 'Tools' section with a search bar and upload button, followed by categories like Get Data, Send Data, Collection Operations, and various genomic file manipulation tools. The central content area features a 'COVID-19 Research!' box with text about SARS-CoV-2 analysis and a quote from Stephen Hawking. Below this are sections for 'News' (with tool updates for Nov 6, Oct 30, Oct 23, Oct 18, and Oct 16) and 'Events' (with Force 2021, ELIXIR BioHackathon Europe, Galaxy Developer Roundtable, and NRZ-Authent Expertinnen- und Expertenworkshop). The right sidebar shows a 'History' section for an unnamed history record, indicating it is empty and prompting the user to load their own data or get data from an external source.

This screenshot shows the Galaxy Europe login page. It features a 'Welcome to Galaxy, please log in' message, input fields for 'Public Name or Email Address' and 'Password', a 'Forgot password?' link, and a 'Login' button. Below the login form is a 'COVID-19 Research!' box with the same content as the previous screenshot. To the right is a 'News' section with tool updates for Nov 6, Oct 30, Oct 23, Oct 18, and Oct 16, and an 'OPEN CHAT' button at the bottom.

Please register only one account - we provide this service free of charge and have limited computational resources. Multi-accounts are tracked and will be subjected to account termination and data deletion.

Create a Galaxy account

Email Address

Password

Confirm password

Public name

Your public name is an identifier that will be used to generate addresses for information you share publicly. Public names must be at least three characters in length and contain only lower-case letters, numbers, dots, underscores, and dashes ('.,_,-').

Create

Already have an account? [Log in here.](#)

GDPR

GDPR Compliance Documentation

For UseGalaxy.eu

ToS & PP

- [Privacy Policy](#)
- [Terms of Service](#)

GDPR Documentation

- [Your Rights Under the GDPR](#)
- [Legitimate Interest Analyses](#)
- [Data Storage and Access](#)
- [Data Processing Activities Register](#)

Contact

Bioinformatics Group

contact@usegalaxy.eu

Department of Computer Science
Albert-Ludwigs-University Freiburg
Georges-Köhler-Allee 106
79110 Freiburg
Germany

Optional: Logging into TiaaS server: https://usegalaxy.eu/join-training/virologist_isciii

2. Galaxy's history

Now select the [Home](#) button and return to the home page. We are going to learn how to manage the history, which is in the right panel. To do this, we will follow these steps:

- Click the new-history (+) icon at the top of the history panel.
 - If the new-history is missing:
 - Click on the galaxy-gear icon (History options) on the top of the history panel
 - Select the option Create New from the menu
- Click once on **Unnamed history** which is the title of your history and type a new meaningful name for it. In our case it would be good **Crimea Congo Reference Genome**. Then type **Enter** on the keyboard and the new name will be set.

The first screenshot shows the 'History' panel with a single entry labeled 'Unnamed history' and a message indicating it is empty. The second screenshot shows the same panel after the user has typed 'Haz clic para cambiar el nombre del historial' (Click to change the history name) into the entry field. The third screenshot shows the history renamed to 'Crimea Congo Reference Genon'.

The screenshot shows the 'History' panel with the history entry renamed to 'Crimea Congo Reference Genon'. A message at the bottom indicates the history is empty and provides options to load data or get data from an external source.

3. Loading data:

Now we are going to load the data. In this case we are going to use the Crimea Congo reference genome. Crimea Congo's genome is composed of 3 segments, each with its own code:

- S segment: DQ133507
- M segment: EU037902
- L segment: EU044832

In order to load these fragments in Galaxy we have to follow these steps: 1. In the left side panel, select **Upload Data** 2. In the new panel select **Paste/Fetch Data** 3. Then copy the following block of text:

```
https://raw.githubusercontent.com/BU-ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/Day1/data/S_DQ133507.fasta
https://raw.githubusercontent.com/BU-ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/Day1/data/M_EU037902.fasta
https://raw.githubusercontent.com/BU-ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/Day1/data/L_EU044832.fasta
```

1. Now, in the **Download data from the web by entering URLs (one per line) or directly paste content.** square, paste the text you copied before
2. Select **Start**
3. When everything is green in the screen, select **Close**

The screenshot shows the Galaxy web interface. On the left, the sidebar has 'Upload Data' selected. The main area shows a message about 24 hours of downtime starting November 22nd, 2021. Below it is a 'COVID-19 Research!' box with links to Galaxy SARS-CoV-2 portal and ENA. A news section quotes Prof. Stephen Hawking. The right sidebar shows an empty history for the 'Crimea Congo Reference Genome' dataset.

Descargar de la red o cargar desde disco

The 'Upload Data' dialog box is open. Step 2 highlights the 'Paste/Fetch data' input field where the URLs were pasted. Step 3 highlights the 'Start' button. Step 4 highlights the 'Close' button. The dialog also shows tabs for 'Regular', 'Composite', 'Collection', and 'Rule-based' upload types, and a note about adding files to the queue.

With this, our data is loading into Galaxy. You can see that each job is given a different number, so you can keep track of the order of your jobs with it.

The jobs can have three different states: 1. Waiting: Your jobs will have a grey color and a clock on their left side. In this state your jobs are waiting to enter in the Galaxy server. 2. Running: Your jobs will have an orange color and rotatory dots on their left side. In this state your jobs are running in the Galaxy server. 3. Done: Your jobs will have a green color. Your data is ready to be used.

History

buscar conjuntos de datos

Crimea Congo Reference Genome

3 shown

(empty)

- 3: L_EU044832.fasta
- 2: M_EU037902.fasta
- 1: S_DQ133507.fasta

History

buscar conjuntos de datos

Crimea Congo Reference Genome

3 shown

(empty)

- 3: L_EU044832.fasta
- 2: M_EU037902.fasta
- 1: S_DQ133507.fasta

History

buscar conjuntos de datos

Crimea Congo Reference Genome

3 shown

19.3 KB

- 3: L_EU044832.fasta
- 2: M_EU037902.fasta
- 1: S_DQ133507.fasta

5. Edit and Visualize your data:

Visualization

Now we can start using our data. First of all, we are going to see how these fasta files look like. There are different ways to do this: 1. Select the :eye: icon in the right to the file name. For the first time, our center panel has changed, and now it displays the content inside the fasta file.

History

buscar conjuntos de datos

Crimea Congo Reference Genome

3 shown

19.3 KB

- 3: L_EU044832.fasta
- 2: M_EU037902.fasta
- 1: S_DQ133507.fasta

1. Another way is to select the name of the file to see the first five lines of the file.

The screenshot shows the BioEdit software interface. The top right corner displays the word "History" and icons for search, add, and settings. Below this is a search bar with the placeholder "buscar conjuntos de datos" and a question mark icon. The main area is titled "Crimea Congo Reference Genome". It shows three entries:

- 3 shown
- 19.3 KB
- 3: L_EU044832.fasta
- 2: M_EU037902.fasta
- 1: S_DQ133507.fasta

Each entry has a detailed view button (eye), edit button (pencil), and delete button (X). The bottom section shows the sequence details for "1: S_DQ133507.fasta":

formato: **fasta**, base de datos: ?

display with IGV local

```
>DQ133507.1 Crimean-Congo hemorrhagic fever virus, complete genome
```

TCTCAAGAACACCGTCCGGCTTACGCCCAACAGTGTCTCTGAGGTGAACAGAACAGATGAGATGAACAAATGGTTGAGGAATCTTTCACAAACTCTACTCCTTTGCGAGAATGTACCAAATGCCCACTGATGACACAGAAGGACTCATCTATGCATCGG

When we display this file summary, we obtain additional options to process this file:

- Save: Allows you to save your files locally

History

buscar conjuntos de datos

Crimea Congo Reference Genome

3 shown

19.3 KB

| | |
|---------------------|--|
| 3: L_EU044832.fasta | |
| 2: M_EU037902.fasta | |
| 1: S_DQ133507.fasta | |

1 sequences
formato: **fasta**, base de datos: ?

display with IGV local

```
>DQ133507.1 Crimean-Congo hemorrhagic fever virus, isolate DQ133507, complete genome
TCTCAAAGAACACGTGCCGCTTACGCCACAGTGTCTCT
GAGGTGAACAGCAAAAGATGAGATGAAACAAATGGTTGAGGA
CTTTCACAAACTCCTACTCTTTGCGAGAATGTACCAAATG
CGCCACTGATGATGCACAGAAGGACTCCATCTATGCATCGG
```

- **Copy link:** copies the link of the data to your clipboard.

History

buscar conjuntos de datos

Crimea Congo Reference Genome

3 shown

19.3 KB

| | |
|---------------------|--|
| 3: L_EU044832.fasta | |
| 2: M_EU037902.fasta | |
| 1: S_DQ133507.fasta | |

1 sequences
formato: **fasta**, base de datos: ?

display with IGV local

```
>DQ133507.1 Crimean-Congo hemorrhagic fever virus, isolate DQ133507, complete genome
TCTCAAAGAACACGTGCCGCTTACGCCACAGTGTCTCT
GAGGTGAACAGCAAAAGATGAGATGAAACAAATGGTTGAGGA
CTTTCACAAACTCCTACTCTTTGCGAGAATGTACCAAATG
CGCCACTGATGATGCACAGAAGGACTCCATCTATGCATCGG
```

- **View details:** Shows a new window in the center panel with additional information about the sample.

History

buscar conjuntos de datos

Crimea Congo Reference Genome

3 shown

19.3 KB

| | |
|---------------------|--|
| 3: L_EU044832.fasta | |
| 2: M_EU037902.fasta | |
| 1: S_DQ133507.fasta | |

1 sequences
formato: **fasta**, base de datos: ?

display with IGV local

```
>DQ133507.1 Crimean-Congo hemorrhagic fever virus, isolate DQ133507, complete genome
TCTCAAAGAACACGTGCCGCTTACGCCACAGTGTCTCT
GAGGTGAACAGCAAAAGATGAGATGAAACAAATGGTTGAGGA
CTTTCACAAACTCCTACTCTTTGCGAGAATGTACCAAATG
CGCCACTGATGATGCACAGAAGGACTCCATCTATGCATCGG
```

Galaxy Europe

Flujo de Trabajo Visualizar ▾ Datos Compartidos ▾ Ayuda ▾ Usuario ▾

Using 0%

On November 22nd at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

- search tools
- Upload Data**
- Get Data
- Send Data
- Collection Operations
- GENERAL TEXT TOOLS
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- GENOMIC FILE MANIPULATION
- Convert Formats
- FASTA/FASTQ
- Quality Control
- SAM/BAM
- BED
- VCF/BCF
- Nanopore
- COMMON GENOMICS TOOLS
- Operate on Genomic Intervals

Data Fetch

Dataset Information

| | |
|------------------------|----------------------------------------------------------------------------------|
| Number | 1 |
| Name | S_DQ133507.fasta |
| Created | Tuesday Nov 16th 12:20:58 2021 UTC |
| Filesize | 1.8 KB |
| Dbkey | ? |
| Format | fasta |
| File contents | contents |
| History Content API ID | 11ac94b70d0bb33aa55c3ad9c03ccdf5 |
| History API ID | afae927a40305960 |
| UUID | 334627ae-0e82-4ac0-8513-7d91457216c1 |
| Full Path | /data/dnb03/galaxy_db/files/3/4/dataset_334627ae-0e82-4ac0-8513-7d91457216c1.dat |

Tool Parameters

| Input Parameter | Value |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| request_version | 1 |
| request_json | {"space_to_tab": false, "to_posix_lines": true, "targets": [{"destination": {"type": "hdas"}, "elements": [{"url": "https://raw.githubusercontent.com/BU-ISCII/galaxy_virologist_training/one_week_4day_format/exercises/Day1/data/S_DQ133507.fasta", "src": "url", "dbkey": "?", "ext": "auto", "hashes": [], "in_place": false, "purge_source": false, "object_id": "76714756"}, {"url": "https://raw.githubusercontent.com/BU-ISCII/galaxy_virologist_training/one_week_4day_format/exercises/Day1/data/L_EU044832.fasta", "src": "url", "dbkey": "?", "ext": "auto", "hashes": [], "in_place": false, "purge_source": false, "object_id": "76714757"}, {"url": "https://raw.githubusercontent.com/BU-ISCII/galaxy_virologist_training/one_week_4day_format/exercises/Day1/data/M_EU037902.fasta", "src": "url", "dbkey": "?", "ext": "auto", "hashes": [], "in_place": false, "purge_source": false, "object_id": "76714758}], "name": ""}, "auto_decompress": true, "check_content": true} |

History

buscar conjuntos de datos

Crimea Congo Reference Genome

3 shown

19.3 KB

3: L_EU044832.fasta
2: M_EU037902.fasta
1: S_DQ133507.fasta

1 sequences
formato: fasta, base de datos: ?

display with IGV local

```
>DQ133507.1 Crimean-Congo hemorrhagic fever
TCTCAAGAACACGTCGGCTTACGCCAACAGTGTTCTT
GAGGTGAACGAAAGATGAGATGAAATGGTTGAGGA
CTTTCACAAACTCTACTCCTTTGCGAGATGTACCAAAT
CGCCACTGTATGCACAGAAAGGACTCATCTATGCATCGG
```

- **Visualize this data:** As we said before in the theory, in the visualization panel you have all the options of visualization allowed in Galaxy, but not all of them fit your data. With this button, you can see which visualization options are better for your type of data.

History

buscar conjuntos de datos

Crimea Congo Reference Genome

3 shown

19.3 KB

3: L_EU044832.fasta
2: M_EU037902.fasta
1: S_DQ133507.fasta

1 sequences
formato: fasta, base de datos: ?

display with IGV local

```
>DQ133507.1 Crimean-Congo hemorrhagic fever
TCTCAAGAACACGTCGGCTTACGCCAACAGTGTTCTT
GAGGTGAACGAAAGATGAGATGAAATGGTTGAGGA
CTTTCACAAACTCTACTCCTTTGCGAGATGTACCAAAT
CGCCACTGTATGCACAGAAAGGACTCATCTATGCATCGG
```

Galaxy Europe

Flujo de Trabajo Visualizar ▾ Datos Compartidos ▾ Ayuda ▾ Usuario ▾

Using 0%

On November 22nd at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

- search tools
- Upload Data**
- Get Data
- Send Data
- Collection Operations
- GENERAL TEXT TOOLS
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- GENOMIC FILE MANIPULATION
- Convert Formats
- FASTA/FASTQ
- Quality Control
- SAM/BAM
- BED
- VCF/BCF
- Nanopore
- COMMON GENOMICS TOOLS
- Operate on Genomic Intervals

search visualizations

Editor
Manually edit text

Multiple Sequence Alignment
The MSA viewer is a modular, reusable component to visualize large MSAs interactively on the web.

History

buscar conjuntos de datos

Crimea Congo Reference Genome

3 shown

19.3 KB

3: L_EU044832.fasta
2: M_EU037902.fasta
1: S_DQ133507.fasta

1 sequences
formato: fasta, base de datos: ?

display with IGV local

```
>DQ133507.1 Crimean-Congo hemorrhagic fever
TCTCAAGAACACGTCGGCTTACGCCAACAGTGTTCTT
GAGGTGAACGAAAGATGAGATGAAATGGTTGAGGA
CTTTCACAAACTCTACTCCTTTGCGAGATGTACCAAAT
CGCCACTGTATGCACAGAAAGGACTCATCTATGCATCGG
```

- **Help:** Displays help about the tool used to generate the data.

Crimea Congo Reference Genome

3 shown

19.3 KB

3: L_EU044832.fasta

2: M_EU037902.fasta

1: S_DQ133507.fasta

1 sequences
formato: **fasta**, base de datos: ?

display with IGV local

```
>DQ133507.1 Crimean-Congo hemorrhagic fever virus, isolate DQ133507, partial genome
TCTAAAGAAACACGTGCCCTAACGCCACAGTGTCTCTGAGGTGAACGCAAAGATGAGATGAAACAAATGGTTGAGGAGCTTTCACAAACTCTACTCCTTTGGAGAATGTACCAATTGCCACTGATGATGCACAGAAGGACTCCATCTATGCATCGG
```

Note: If you select again in the file name, the summary disappears

Edition

Now we are going to rename all the fasta files we uploaded to Galaxy. To do this, we have to click in the pencil icon that appears next to each file name. This will display a new central window with the different edition options for each file:

Crimea Congo Reference Genome

3 shown

19.3 KB

3: L_EU044832.fasta

2: M_EU037902.fasta

1: S_DQ133507.fasta

Galaxy Europe

Flujo de Trabajo Visualizar Datos Compartidos Ayuda Usuario Using 0%

On November 22nd at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Edit dataset attributes

Attributes Convert Datatypes Permissions

Editar atributos

Name: S_DQ133507.fasta

Info

Annotation

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build

----- Additional Species Are Below -----

History

Crimea Congo Reference Genome

3 shown

19.3 KB

3: L_EU044832.fasta

2: M_EU037902.fasta

1: S_DQ133507.fasta

This screen allows you to perform different things. Starting from the right:

- Set permissions: Allows you to manage the access and permissions of the selected file, for the different users registered.

- Datatype: Allows you to change the datatype of the existing dataset, but not modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.
- Convert: Allows you to create a new dataset with the contents of this dataset, converted to a new format.
- Change the attributes: Allows you to rename the file, and add some additional information.

:warning: Select **Save** button to save the changes.

We are going to rename the files as shown here:

The screenshot shows the Galaxy History interface. At the top, there are buttons for History, New Dataset, and Settings. Below that is a search bar labeled "buscar conjuntos de datos". The main area displays a "Crimea Congo Reference Genome" dataset with the following details:

- 3 shown
- 19.3 KB
- Three datasets listed:
 - 3: L_fragment.fasta
 - 2: M_fragment.fasta
 - 1: S_fragment.fasta

6. Run tools

Now we are going to use the fasta files uploaded to Galaxy to run tools. To run tools we have to:

Search

1. Search the tool in the search tab. We want to concatenate the fasta files, so we are going to search for **concatenate** in the bar.
2. Select the tool we want to use. In this case **Concatenate datasets tail-to-head (cat)**.

The screenshot shows the Galaxy Europe interface. The search bar at the top contains the text "concatenate". On the left, a sidebar lists various tools under categories like "Tools", "VCFgenotype-to-haplotype!", "LC/MS matching", "Concatenate images", and "Concatenate datasets tail-to-head". The "Concatenate datasets tail-to-head (cat)" tool is highlighted with a blue border and numbered "2". A message box on the right says "24 hours downtime starting from November 24th, 2021 at 17:45 pm CET". The main content area shows the "Crimea Congo Reference Genome" dataset with the three concatenated FASTA files. The right side of the screen shows the Galaxy History interface with the same dataset and files.

Run tools

When we select the tool we are going to see the tool's options in the center panel. We are going to see different information about the tool we want to run.

:warning: These options are tool specific. This means each tool has its own options.

1. Tool name, version and options to save and share the tool
2. The input dataset options: - We can select data from the history - Upload data from a collection - Upload a dataset (the upload dataset pop up will appear) - Brows a dataset (you can brows dataset from the history)
3. Insert new dataset blocks (no

need in our case) 4. Execute button 5. Tool information: - :warning: - What it does

- Examples - Citaiton

To concatenate the samples, we will follow the wollowing steps: 1. In *Datasets to concatenate*: - Press *Ctrl* key in your keyboard - Select the three fasta files **while still pressing the Ctrl key**. 2. Press execute

The screenshot shows the Galaxy Europe web interface. On the left, a sidebar lists various tools: VCFgenotype-to-haplotype, LC/MS matching, Concatenate images, Concatenate FASTA alignment by species, Concatenate datasets tail-to-head, Concatenate datasets tail-to-head (cat), AXT to concatenated FASTA, Concatenate two BED files, FASTA Merge Files and Filter Unique Sequences, and bcftools concat. The main panel displays the 'Concatenate datasets tail-to-head (cat)' tool. In the 'Datasets to concatenate' section, three files are listed: 3: L_fragment.fasta, 2: M_fragment.fasta, and 1: S_fragment.fasta. A blue box highlights the 'Ctrl' key being pressed. Below this, there is an 'Execute' button, which is also highlighted with a red box. To the right, the 'History' panel shows a new entry: 'Crimea Congo Reference Genome' with three datasets: 3: L_fragment.fasta, 2: M_fragment.fasta, and 1: S_fragment.fasta.

Running jobs

Once we have pressed **Execute**, a new central panel window will appear and our job will be in queue process: 1. In the top of the panel (blue) you have a summary of what we've just run. In our case 3 input datasets have are involved in a single process, with a unique output. 2. In the foot of the panel (red) you have some recommendations from Galaxy on how to process your data after the process we have just run. 3. In the history (yellow) we have now a new entry, which is the number 4, with the results of our job. Galaxy names jobs according to the used tool and the input dataset.

The screenshot shows the Galaxy Europe web interface after the job has been executed. The 'Tools' sidebar remains the same. The main panel now displays the results of the 'Concatenate datasets' tool. It shows a green checkmark and the message: 'Executed Concatenate datasets and successfully added 1 job to the queue.' It details the inputs: '1: S_fragment.fasta', '2: M_fragment.fasta', and '3: L_fragment.fasta'. It also shows the output: '4: Concatenate datasets on data 3, data 2, and data 1'. Below this, a 'Tool recommendation' section suggests other tools: heatmap2, Query Tabular, Sort, UMI-tools extract, Manipulate AnnData, Molecule to fingerpr..., Closed-reference OTU..., FastQC, Trimmomatic, and Bowtie2. The 'History' panel on the right shows the completed job: '4: Concatenate datasets on data 3, data 2, and data 1' with the same three datasets.

Visualize results

Whenever our job is green, we can see the results by clicking in the :eye: icon. Now we can see the three sequences for the segments, headers included, in a unique fasta file.

The screenshot shows the Galaxy Europe interface with the following details:

- Tools:** concatenate
- History:** 4 shown (Crimea Congo Reference Genome)
- Dataset 1 (L1):** >00133507.1 Crimean-Congo hemorrhagic fever virus strain Kosovo Hoti segment S, complete sequence
- Dataset 2 (L2):** >EU037902.1 Crimean-Congo hemorrhagic fever virus strain Kosovo Hoti segment M, complete sequence
- Dataset 3 (L3):** >EU037902.1 Crimean-Congo hemorrhagic fever virus strain Kosovo Hoti segment L, complete sequence
- Dataset 4 (L4):** 4: Concatenate datasets on data 3, data 2, and data 1

Now we are going to rename the fasta file as follows: 1. Click on the :pencil: icon
2. Write **Crimea Congo Ref Genome** in the **Name** square 3. Press **Save**

The screenshot shows the Galaxy Europe interface with the following details:

- Tools:** concatenate
- History:** 4 shown (Crimea Congo Reference Genome)
- Dataset 1 (L1):** Crime Congo Ref Genome
- Dataset 2 (L2):** L_fragment.fasta
- Dataset 3 (L3):** M_fragment.fasta
- Dataset 4 (L4):** S_fragment.fasta

First Question Answer

- How do I create a fasta reference for fragmented Crimea Congo genome?

7. Furtherly process your data

Now that we have our concatenated fasta file, we can check that everything is fine by scrolling down the genome, and checking that the three fragments are fine, or we can use another tool to count the number of sequences in a fasta file, and the number of nucleotides in each sequence.

To do this, we are going to: 1. Search **fasta** in the tool square. 2. Select **Fasta Statistics Display summary statistics for a fasta file** 3. In *fasta or multifasta file* select **multiple data set** 4. With *Ctrl* key pressed, select the 3 fragments and the multifasta file 5. Press **Start** button.

Galaxy Europe

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

fasta 1

Edit dataset attributes

Attributes updated.

Attributes Convert Datatypes Permissions

Editor atributos Auto-detect Save

Name
Crimea Congo Ref Genome

Info

Annotation

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build
----- Additional Species Are Below -----

History

buscar conjuntos de datos

Crimea Congo Reference

Genome

4 shown
38.59 KB

4: Crimea Congo Ref Genome
3: L_fragment.fasta
2: M_fragment.fasta
1: S_fragment.fasta

Galaxy Europe

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

fasta 1

Fasta Statistics Display summary statistics for a fasta file. 2

FastaCLI Appends decoy sequences to FASTA files

bedtools GetFastaBed use intervals to extract sequences from a FASTA file

bedtools MaskFastaBed use intervals to mask sequences from a FASTA file

NormalizeFasta normalize fasta

Fasta or multifasta file

1 2 This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

Genome size estimate (optional)

Email notification

No 3

Execute

Fasta Stats Displays the summary statistics for a fasta file.

Outputs in tabular form:
Lengths: n50, min, max, median and average
Number of base pairs: A, C, G, T, N, Total and Total_not_N

History

buscar conjuntos de datos

Crimea Congo Reference

Genome

4 shown
38.59 KB

4: Crimea Congo Ref Genome
3: L_fragment.fasta
2: M_fragment.fasta
1: S_fragment.fasta

Now we have 4 jobs running, because this tool will run one statistics process for each fasta file we selected.

Galaxy Europe

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

fasta

bcftools consensus Create consensus sequence by applying VCF variants to a reference fasta file

VCPrimers: Extract flanking sequences for each VCF record

SNP distance matrix Compute distance in SNPs between all sequences in a FASTA file

ConvertFastaToPrositeCSV Create Prosite CSV Input From a Protein FASTA

Filter fasta to remove sequences based on input criteria (filter_fasta)

Fasta Statistics Display summary statistics for a fasta file.

FastaCLI Appends decoy sequences to FASTA files

bedtools GetFastaBed use intervals to extract sequences from a FASTA file

bedtools MaskFastaBed use intervals to mask sequences from a FASTA file

NormalizeFasta normalize fasta

Executed Fasta Statistics and successfully added 4 jobs to the queue. 1

The tool uses 4 inputs:

- 1: S_fragment.fasta
- 2: M_fragment.fasta
- 3: L_fragment.fasta
- 4: Crimea Congo Ref Genome

It produces 4 outputs:

- 5: Fasta Statistics on data 1: Fasta summary stats
- 6: Fasta Statistics on data 2: Fasta summary stats
- 7: Fasta Statistics on data 3: Fasta summary stats
- 8: Fasta Statistics on data 4: Fasta summary stats

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

buscar conjuntos de datos

Crimea Congo Reference

Genome

8 shown
38.59 KB

8: Fasta Statistics on data 4: Fasta summary stats
7: Fasta Statistics on data 3: Fasta summary stats
6: Fasta Statistics on data 2: Fasta summary stats
5: Fasta Statistics on data 1: Fasta summary stats
4: Crimea Congo Ref Genome
3: L_fragment.fasta
2: M_fragment.fasta
1: S_fragment.fasta

Results visualization

Now we are going to see the statistics summary for each fasta file. To do this we

have to select the :eye: icon in each of the Fasta Statistics output.

For the **S fragment**, we are going to see the number of sequences inside the fasta file, and the number of nucleotides. We are going to:

1. Select the :eye: icon in the job with the name *Fasta Statistics on data 1: Fasta summary stats*
2. See the *num_bp* row, which corresponds to the number of nucleotides in the fasta file, 1673 in this case.
3. Check *num_seq*, corresponding to the number of sequences in the fasta file.

| | GC_content | 45.5 |
|--------------|------------|------|
| L50 | 1 | |
| L90 | 1 | |
| len_N50 | 1673 | |
| len_N90 | 1673 | |
| len_max | 1673 | |
| len_mean | 1673 | |
| len_median | 1673 | |
| len_min | 1673 | |
| num_A | 501 | |
| num_C | 375 | |
| num_G | 387 | |
| num_N | 0 | |
| num_T | 410 | |
| num_bp | 1673 | |
| num_bp_not_N | 1673 | |
| num_seq | 1 | |

Now we are going to repeat this process for the rest of the fasta files:

M fragment

- How many nucleotides are in M fragment?

L fragment

- How many nucleotides are in L fragment?

Crimea Congo Genome

- How many sequences and nucleotides are in the Crimea Congo reference genome?

Now we can answer the second question.

Second Question Answer

- How many nucleotides has each fragment of Crimea Congo genome?

Share results

Now that we know that the reference genome for the whole Crimea Congo virus is done correctly, we can use it as reference genome for further analysis in this same history, or save it to use it in our computer. To do so: 1. Select the name of the fasta you want to download: **4: Crimea Congo Ref Genome** 2. Select the **Save** button in the emerging panel.

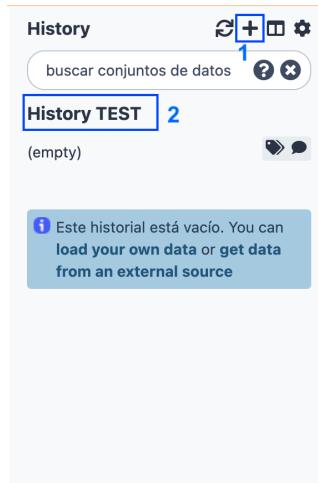
The screenshot shows the Galaxy Europe web interface. On the left, there's a sidebar with various tools categorized under 'Tools' (e.g., search tools, upload data), 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'GENOMIC FILE MANIPULATION', 'Convert Formats', 'FASTA/FASTQ', 'Quality Control', 'SAM/BAM', 'BED', 'VCF/BCF', 'Nanopore', and 'COMMON GENOMICS TOOLS'. The main content area has a yellow banner about a 24-hour downtime starting on November 24th. Below it is a green box for 'COVID-19 Research!' with a quote from Stephen Hawking. There are also sections for 'News' (with links to training infrastructure feedback and tool updates) and 'Events' (with links to FORCEs eScience course, Forces 2021, and NRZ-Authent Expertinnen- und Expertenworkshop). On the right, a 'History' panel is open, showing a list of items related to the 'Crimea Congo Reference Genome', including 'Fasta Statistics on data' entries and a sequence snippet.

8. History management

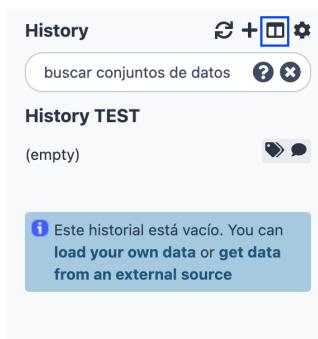
Now, we are going to learn how to manage the history. In this case, we created a new history record and, while we were doing our analysis, the steps we followed were recorded.

This history is saved in your account so you can create a new one for a new analysis, and access previous analysis later.

1. To create a new history, select the + button in the history panel.
2. Then, rename your new history to: **History TEST**



Now we have a clean history, but we have lost the previous history with the Crimea Congo results. To see the previous history, we have to access the history manager:



Now we can check out the previous history, with all the Crimea Congo results. We are going to remove the TEST history and go back to the Crimea Congo Ref Genome history to share it. 1. Select the dropdown icon :warning: be sure to

select the dropdown in the history you want to delete, not in the good one. 2.

Select **Delete** 3. Press **Switch to** in the Crimea Congo history 4. Select the

HOME icon

The screenshot shows the Galaxy Europe interface with the following details:

- Top Bar:** Galaxy Europe, Flujo de Trabajo, Visualizar, Datos Compartidos, Ayuda, Usuario, Using 0%.
- Message Bar:** [UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.
- Left Panel:** History TEST (empty). It says "Drag datasets here to copy them to the current history". A message "Este historial está vacío" is displayed.
- Right Panel:** Reference Genome. It lists several datasets:
 - 8: Fasta Statistics on data 4: Fasta summary stats
 - 7: Fasta Statistics on data 3: Fasta summary stats
 - 6: Fasta Statistics on data 2: Fasta summary stats
 - 5: Fasta Statistics on data 1: Fasta summary stats
 - 4: Crimea Congo Ref Genome
 - 3: L_fragment.fasta
 - 2: M_fragment.fasta
 - 1: S_fragment.fasta
- Buttons:** search histories, search all datasets, Create new, Switch to (highlighted with a red box).

Once we are finished, we can save our history in order to access this results

later, or to share them with other lab members. To do this, we are going to: 1.

Select the engine icon in the history 2. Select **Share or publish** 3. Select the option **Make History accessible**

The screenshot shows the Galaxy Europe interface with the following details:

- Top Bar:** History, +, gear icon (highlighted with a red box), buscar conjuntos de datos.
- Title:** Crimea Congo Reference Genome
- Content:** 8 shown, 39.39 KB.
- Actions:** History Actions (highlighted with a red box):
 - Copy
 - Compartir o Publicar (highlighted with a red box)
 - Mostrar Estructura
 - Extraer Flujo de Trabajo
 - Set Permissions
 - Make Private
 - Reanudar Trabajos en Pausa
- Dataset List:** 8: Fasta Statistics on data 4: Fasta summary stats, 7: Fasta Statistics on data 3: Fasta summary stats, 6: Fasta Statistics on data 2: Fasta summary stats, 5: Fasta Statistics on data 1: Fasta summary stats, 4: Crimea Congo Ref Genome, 3: L_fragment.fasta, 2: M_fragment.fasta, 1: S_fragment.fasta.

The screenshot shows the Galaxy Europe web interface. On the left, there's a sidebar with various tool categories like Tools, Get Data, Send Data, Collection Operations, and several sub-sections under GENERAL TEXT TOOLS and GENOMIC FILE MANIPULATION. The main area displays a history titled 'Crimea Congo Reference Genome'. It includes options to share or publish the history, with a checked checkbox for 'Make History accessible'. Below this, it says 'This History is currently accessible via link.' and provides a URL: <https://usegalaxy.eu/u/svarona/h/crimea-congo-reference-genome>. The history list shows items such as '8: Fasta Statistics on data' and '3: Fasta summary stats', along with other FASTA files and their respective statistics.

Now everyone with the link can access the history.

9. Workflows

Creating workflows

Now we are going to create a workflow so every time we input three fasta files with crimea congo fragments to this workflow, it will concatenate them into a unique fasta file and generate stats of them: 1. Select the engine icon in the history 2. Select **Extract workflow** 3. Check if every step is correct 4. Rename the workflow to: **Create Crimea Congo Reference Genome** 5. Select **Create workflow**

This screenshot shows the Galaxy History panel for the 'Crimea Congo Reference Genome' history. The history contains eight items, including FASTA files and their corresponding summary statistics. A context menu is open over the third item, '3: Fasta summary stats', with the 'Extraer Flujo de Trabajo' option highlighted. Other options in the menu include Copy, Compartir o Publicar, Mostrar Estructura, Set Permissions, Make Private, and Reanudar Trabajos en Pausa. To the right of the history list, there are sections for History Actions, Acciones de Conjuntos de Datos, Descargas, and Beta Features.

Now your workflow has been created so go to the workflow manager, where you can see the list of all your workflows.

Editing workflows

Now we are going to have a look to the workflow we created:

1. Select the name of the workflow **Create Crimea Congo Reference Genome**
2. Select **Edit**
- You will see all the squares corresponding to each of the workflow's processes.
4. Move them a little bit you can have a better look at it.
5. Go back to the workflow manager.

Galaxy Europe

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Using 0%

Tools

search tools

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

COMMON GENOMICS TOOLS

Search Workflows

+ Create Import

Create Crimea Congo Reference Genome

1 a few seconds ago

2 6 days ago

History

buscar conjuntos de datos

Crimea Congo Reference Genome

8 shown

39.39 KB

8: Fasta Statistics on data a 4: Fasta summary stats

7: Fasta Statistics on data 3: Fasta summary stats

6: Fasta Statistics on data a 2: Fasta summary stats

5: Fasta Statistics on data a 1: Fasta summary stats

4: Crimea Congo Ref Genome

3: L_fragment.fasta

2: M_fragment.fasta

1: S_fragment.fasta

Galaxy Europe

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Using 0%

Tools

search tools

Inputs

Get Data

Send Data

Collection Operations

Expression Tools

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

Quality Control

SAM/BAM

Create Crimea Congo Reference Genome

Name: Create Crimea Congo Reference Genc

Version: 1: Nov 17th 2021, 8 steps

Annotation:

These notes will be visible when this workflow is viewed.

License: Specify a license for this workflow.

Creator: Add a new creator - either a person or an organization.

Tags:

Apply tags to make it easy to search for and find items with the same tag.

Galaxy Europe

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Using 0%

Tools

search tools

Inputs

Get Data

Send Data

Collection Operations

Expression Tools

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Create Crimea Congo Reference Genome

Step Annotation:

Add an annotation or notes to this step. Annotations are available when a workflow is viewed.

Optional:

No

Format(s):

Leave empty to auto-generate filtered list at runtime based on connections.

Sharing workflows

Now we are going to share our workflow: 1. Select the name of the workflow

Create Crimea Congo Reference Genome 2. Select **Share** 3. Select **Make**

Workflow Accessible Via Link 4. There you have the link to share it 5. Go back to the workflow manager.

Galaxy Europe

Flujo de Trabajo Visualizar Datos Compartidos Ayuda Usuario Using 0%

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

- search tools
- + Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Search Workflows

+ Create Import

Create Crimea Congo Reference Genome

1

Name Tags Updated Sharing Bookmarked

2 minutes ago

6 days ago

History

buscar conjuntos de datos

Crimea Congo Reference Genome

8 shown

39.39 KB

8: Fasta Statistics on data a 4: Fasta summary stats

7: Fasta Statistics on data a 3: Fasta summary stats

6: Fasta Statistics on data a 2: Fasta summary stats

5: Fasta Statistics on data a 1: Fasta summary stats

4: Crimea Congo Ref Genome

3: L_fragment.fasta

2: M_fragment.fasta

1: S_fragment.fasta

Galaxy Europe

Flujo de Trabajo Visualizar Datos Compartidos Ayuda Usuario Using 0%

Go back to Workflows List

Workflow 'Create Crimea Congo Reference Genome'

Share

This workflow is currently restricted so that only you and the users listed below can access it.

Make Workflow Accessible via Link

Generates a web link that you can share with other people so that they can view and import the workflow.

Make Workflow Accessible and Publish

Makes the workflow accessible via link (see above) and publishes the workflow to Galaxy's Published Workflows section, where it is publicly listed and searchable.

You have not shared this workflow with any users yet.

Share with a user

Export

Download workflow as a file so that it can be saved or imported into another Galaxy server.

This workflow must be accessible. Please use the option above to "Make Workflow Accessible and Publish" before receiving a URL for importing to another Galaxy.

Create image of workflow in SVG format

Export to the www.myexperiment.org site.

myExperiment username: svarona

myExperiment password: *

Galaxy Europe

Flujo de Trabajo Visualizar Datos Compartidos Ayuda Usuario Using 0%

Go back to Workflows List

Workflow 'Create Crimea Congo Reference Genome'

Share

This workflow is currently **accessible via link**.
Anyone can view and import this workflow by visiting the following URL:
<https://usegalaxy.eu/u/svarona/w/create-crimea-congo-reference-genome>

Disable Access to Workflow Link

Disables workflow's link so that it is not accessible.

Publish Workflow

Publishes the workflow to Galaxy's Published Workflows section, where it is publicly listed and searchable.

You have not shared this workflow with any users yet.

Share with a user

Export

Download workflow as a file so that it can be saved or imported into another Galaxy server.

Use this URL to import the workflow directly into another Galaxy server:
<https://usegalaxy.eu/u/svarona/w/create-crimea-congo-reference-genome/json>
(Copy this URL into the box titled 'Workflow URL' in the Import Workflow page.)

Create image of workflow in SVG format

Export to the www.myexperiment.org site.

Importing workflows

Now we are going to import a Galaxy workflow. Remember that you cannot import your own workflow from your user, if you already have it. So copy my own workflow or one of your colleague's:

[https://usegalaxy.eu/u/svarona/w\(concat-f frags-reference-genome](https://usegalaxy.eu/u/svarona/w(concat-f frags-reference-genome)

1. Now paste the link in your browser's URL
2. There you have a summary of the workflow.
3. In the right side you have two buttons:
 - Left one to download the workflow
 - Right one (+) to import the workflow.
4. Go back to the Workflow manager and check if it is there.

The screenshot shows the Usegalaxy.eu Galaxy interface. On the left, a sidebar lists various tools and operations. A central panel displays a workflow titled "Concatenate fragments Reference Genome". The workflow consists of four steps: Step 1: Input dataset (S_fragment.fasta), Step 2: Input dataset (M_fragment.fasta), Step 3: Input dataset (L_fragment.fasta), and Step 4: Fasta Statistics (fasta or multifasta file). To the right, a "History" panel shows a list of datasets, including "Isolate Ast199" which contains 8 items. The top navigation bar includes links for "Galaxy | Europe | Accessible Workflow | Concatenate fragments Ref... - https://usegalaxy.eu/u/svarona/w(concat-frags-reference-genome)" and "https://usegalaxy.eu/u/svarona/w(concat-frags-reference-genome) - Búsqueda de Google".

Running workflows

Now we are going to learn how to run a workflow with new data. Crimea Congo's genome we already have is the one for the Kosovo Hoti strain. Now, we are going to obtain the Reference genome for isolate Ast199, with the following codes for their sequences:

- S segment: KX056052
- M segment: KX056051
- L segment: KX056050

1. Create a new history (as previously explained) named "Isolate Adt199"
2. Select the run icon in the workflow you want to run.
3. Now we have to upload the new fasta fragments. We are going to select the **Upload Data** icon and the pop-up seen before to upload data will appear:
 - In the new panel select **Paste/Fetch Data**

- Now, in the **Download data from the web by entering URLs (one per line) or directly paste content.** square, paste the text you copied before:

```
https://raw.githubusercontent.com/BU-ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/data/S_KX056052.fasta
https://raw.githubusercontent.com/BU-ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/data/M_KX056051.fasta
https://raw.githubusercontent.com/BU-ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/data/L_KX056050.fasta
```

- Select **Start**

- When everything is green in the screen, select **Cancel**

- Select browse datasets in the like icon for the S fragment
- Select the S fragment from the list
- Repeat steps 4 and 5 for fragments L and M so the resulting window is like the one in the picture.

7. Select **Run Workflow.**

The screenshot shows the Galaxy Europe interface. On the left, there's a sidebar with 'Tools' (search tools, upload data), 'Get Data', 'Send Data', 'Collection Operations', and sections for 'GENERAL TEXT TOOLS' (Text Manipulation, Filter and Sort, Join, Subtract and Group), 'GENOMIC FILE MANIPULATION', and 'Convert Formats'. The main area shows a search bar 'Search Workflows' and a list of workflows. One workflow, 'Create Crimea Congo Reference Genome', is expanded, showing it was created 'an hour ago'. Another workflow, 'FastQC', is listed as 'FastQC' with a creation date of '6 days ago'. To the right, there's a 'History' section with a message: 'Este historial está vacío. You can load your own data or get data from an external source'. A large blue number '3' is overlaid on the 'Run Workflow' button.

This screenshot shows the 'Workflow: Create Crimea Congo Reference Genome' form. It has three input fields: 'S_fragment.fasta' (with a note: 'No txt, fastq.gz, fasta.gz, genbank.gz, tabular.gz or fasta dataset available.'), 'M_fragment.fasta' (with a note: 'No fasta, txt, fastq.gz, fasta.gz, genbank.gz or tabular.gz dataset available.'), and 'L_fragment.fasta' (with a note: 'No fasta, txt, fastq.gz, fasta.gz, genbank.gz or tabular.gz dataset available.'). Below the inputs is a link 'Expand to full workflow form.'. To the right, there's a 'History' section with a message: 'Este historial está vacío. You can load your own data or get data from an external source'. A large blue number '3' is overlaid on the 'Run Workflow' button.

Descargar de la red o cargar desde disco

This screenshot shows the 'Download data from the web by entering URLs (one per line) or directly paste content.' section. At the top, there are dropdowns for 'Type (set all)' (set to 'Auto-detect') and 'Genome (set all)' (set to 'Additional ...'). Below is a text input field containing three URLs. At the bottom, there are buttons for 'Elegir archivos locales' (Local files), 'Choose remote files', 'Paste/Fetch data' (highlighted with a blue box and a red number '1'), 'Start' (highlighted with a red box and a red number '3'), 'Pause', 'Reset', and 'Cancel'.

Descargar de la red o cargar desde disco

Regular Composite

| Name | Size | Type | Genome | Settings | Status |
|----------|-------|------------|----------------------|----------|--------|
| New File | 371 b | Auto-de... | ----- Additional ... | | 100% ✓ |

Download data from the web by entering URLs (one per line) or directly paste content.

```
https://raw.githubusercontent.com/BU-ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/data/S_KX056052.fasta
https://raw.githubusercontent.com/BU-ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/data/M_KX056051.fasta
https://raw.githubusercontent.com/BU-ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/data/L_KX056050.fasta
```

Type (set all): Auto-detect Genome (set all): ----- Additional ...

Elegir archivos locales Choose remote files Paste/Fetch data Start Pause Reset Cancel

Galaxy Europe Flujo de Trabajo Visualizar Datos Compartidos Ayuda Usuario Using 0%

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

- search tools
- Workflow: Create Crimea Congo Reference Genome**
- Run Workflow

Get Data

S_fragment.fasta

M_fragment.fasta

L_fragment.fasta

History

buscar conjuntos de datos

Isolate Ast199

3 shown

19.26 KB

3: L_KX056050.fasta
2: M_KX056051.fasta
1: S_KX056052.fasta

Type to Search

| Label | Details | Time |
|----------------------------|---------|------------------|
| 3: L_KX056050.fasta | fasta | 2021-11-17 13:55 |
| 2: M_KX056051.fasta | fasta | 2021-11-17 13:55 |
| 1: S_KX056052.fasta | fasta | 2021-11-17 13:55 |

Upload Cancel

Galaxy Europe

Flujo de Trabajo Visualizar Datos Compartidos Ayuda Usuario Using 0%

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

- search tools
- Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Workflow: Create Crimean Congo Reference Genome

1 2 Run Workflow

S_fragment.fasta
1: S_KX056052.fasta

M_fragment.fasta
2: M_KX056051.fasta

L_fragment.fasta
3: L_KX056050.fasta

Expand to full workflow form.

History

buscar conjuntos de datos

Isolate Ast199

3 shown
19.26 KB

3: L_KX056050.fasta
2: M_KX056051.fasta
1: S_KX056052.fasta

Now our workflow is running, so we have to wait until every step is done to see the results.

Galaxy Europe

Flujo de Trabajo Visualizar Datos Compartidos Ayuda Usuario Using 0%

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

- search tools
- Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Workflow: Create Crimean Congo Reference Genome

Successfully invoked workflow Create Crimean Congo Reference Genome.
You can check the status of queued jobs and view the resulting data by refreshing the History pane, if this has not already happened automatically.

Invocation 1... Loading step state summary....
0 of 0 jobs complete (total number of jobs will change until all steps fully scheduled)...

Inputs Steps PHD Comics Random

History

buscar conjuntos de datos

Isolate Ast199

3 shown
19.26 KB

3: L_KX056050.fasta
2: M_KX056051.fasta
1: S_KX056052.fasta

Once the workflow is finished, we will see a window like this one, where all the datasets on the history are in green finished. Also, you can select the input and output dropdowns to see what has been run.

Galaxy Europe

Flujo de Trabajo Visualizar Datos Compartidos Ayuda Usuario Using 0%

[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.

Tools

- search tools
- Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Workflow: Create Crimean Congo Reference Genome

Successfully invoked workflow Create Crimean Congo Reference Genome.
You can check the status of queued jobs and view the resulting data by refreshing the History pane, if this has not already happened automatically.

View Report 1 8 of 8 steps successfully scheduled.
Download Biocompute Object 5 of 5 jobs complete.

Inputs Steps PHD Comics Random

History

buscar conjuntos de datos

Isolate Ast199

8 shown
39.31 KB

8: Fasta Statistics on data
a 6: Fasta summary stats
7: Fasta Statistics on data
3: Fasta summary stats
6: Concatenate datasets
on data 1, data 2, and data
a 3
5: Fasta Statistics on data
a 2: Fasta summary stats
4: Fasta Statistics on data
a 1: Fasta summary stats
3: L_KX056050.fasta
2: M_KX056051.fasta
1: S_KX056052.fasta

Galaxy also allows you to download a report in PDF format that looks like this:

The screenshot shows the Galaxy Europe web interface. On the left, there's a sidebar with 'Tools' (search tools, upload data), 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'GENOMIC FILE MANIPULATION', 'Convert Formats', 'FASTA/FASTQ', 'Quality Control', 'SAM/BAM', 'BED', 'VCF/BCF', 'Nanopore', and 'COMMON GENOMICS TOOLS' (Operate on Genomic Intervals). The main area has tabs for 'Flujo de Trabajo' (Workflow), 'Visualizar' (Visualize), 'Datos Compartidos' (Shared Data), 'Ayuda' (Help), 'Usuario' (User), and 'Galaxy' (Galaxy). A message at the top says '[UPDATE] On November 24th at 17:45 pm CET and for the next 24 hours, we have scheduled a maintenance task on the Usegalaxy.eu infrastructure. All services will be shut down and running jobs at that time terminated. Please take it into account in your job schedule.' Below this, a 'Galaxy Report' section shows a workflow step titled 'Create Crimea Congo Reference Genome' created on November 18, 2021, at 8:38. It includes an identifier (41e1c20a62fd6291) and a 'History' section for 'Isolate Ast199' with details like 8 shown results, 39.31 KB size, and various statistics and concatenation steps. The central part of the page displays the 'Workflow Execution Summary' for the 'Create Crimea Congo Reference Genome' workflow, listing 'Workflow Inputs' and 'Input Dataset: L_fragment.fasta'.

Finally we can have a look the the resulting stats in this history.

Note:

- This hands-on history URL: <https://usegalaxy.eu/u/svarona/h/crimea-congo-reference-genome>
 - This hands-in workflow URL: <https://usegalaxy.eu/u/svarona/w/concat-frags-reference-genome>