

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 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. At the top, there is a navigation bar with links for Home, Workflow, Visualize, Shared Data, Help, Login or Register, and a search bar. The main area is divided into several panels:

- Tools Panel (Blue):** Contains sections for 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.
- Central Panel (Red):** This panel contains the following sections:
 - COVID-19 Research:** A green box with text about SARS-CoV-2 analysis and a quote from Prof. Stephen Hawking.
 - News:** A list of recent news items:
 - 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
 - Oct 18, 2021: Training Infrastructure Feedback from Dr. Theodora Tsirka
 - Events:** A list of upcoming events:
 - 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
 - Nov 16, 2021 - Nov 17, 2021: 5. NRZ-Authent Expertinnen- und Expertenworkshop
- History Panel (Green):** Shows the user's history with the message: "This history is empty. You can load your own data or get data 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. Central panel in red, which will let you run analyses and view outputs 4. Right panel in green, with the history record.

Table of Contents

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

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. At the top, there is a navigation bar with links for Workflow, Visualize, Shared Data, Help, and a redboxed 'Login or Register' button. Below the navigation bar is a sidebar with various tool categories like Tools, Get Data, Send Data, Collection Operations, and several sections under GENERAL TEXT TOOLS. The main content area features a 'COVID-19 Research!' section with a green background, news items, and events. On the right side, there is a 'History' panel showing an 'Unnamed history' entry with a note that it is empty and can be loaded from an external source.

This screenshot shows the Galaxy Europe web interface with a focus on the login form. The page title is 'Galaxy Europe'. The main content area has a 'Welcome to Galaxy, please log in' message. It includes 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!' section with a green background, news items, and events. On the right side, there is a 'News' section with a green background, news items, and an 'OPEN CHAT' button.

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

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 screenshot shows the Galaxy History panel. At the top, there is a search bar labeled "buscar conjuntos de datos" and a set of icons. Below the search bar, the title "History" is displayed. Underneath the title, the text "Unnamed history" is shown, followed by "(empty)". A blue tooltip box appears over the history title, containing the message: "Este historial está vacío. You can load your own data or get data from an external source".

The screenshot shows the Galaxy History panel. The title "History" is at the top. Below it, the history is titled "Crimea Congo Reference Genon", which is highlighted with a blue border. Underneath the title, the text "(empty)" is visible. A blue tooltip box appears over the history title, containing the message: "Haz clic para cambiar el nombre del historial".

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

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**

Name	Size	Type	Genome	Settings	Status
New File	386 b	Auto-de...	----- Additional ...	⚙️	0%

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

2

3

4

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.

The screenshot shows the Galaxy interface with three completed FASTA files listed:

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

The first file, L_EU044832.fasta, is highlighted with a green background, indicating it is the currently selected or viewed item.

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.

The screenshot shows the Galaxy interface with the content of the L_EU044832.fasta file displayed. The eye icon next to the file name is highlighted with a blue border, indicating it was recently selected.

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

>DQ13507.1 Crimean-Congo hemorrhagic fever virus strain Kosovo Hoti segment S, complete sequence
TCTCAAAAGAACACGGCCCTTACGCCAACATGTTCTTCAGGATGCTGCNCAAATGGAAACAGACAG
GGAGGTACAGACGAAGATGAGATGACAATGGTTGAGGATTTAAAGAGGAATGACTTGTGGACA
CTTCTAAACAACTCTACTTCTTGGAAGATTAACAAATCTGGAAGATTGTTGGCAGATGCGCAG
GCCAGATGAGGAAAGAACTTCTTATGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAA
ATTCAGGAAACATCAATCTTGGGAACTTCTGCACTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAA
ACTTCGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAA
GCAGCCTTAAGAACAACTGCTCTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAA
TGTGTCAGCATGATGATAAGGGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAA
AGTGCACGCTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAA
CGCGACGACGGACGACGGACGACGGACGACGGACGACGGACGACGGACGACGGACGACGGACGACGGACGACGGACGACGGACG
AAAGCACAGACGCTTAAATAAAATTCTGCTCTGCAAGTCACAGCATGATAACAAACCTCTAAAGCACATTGCC
AAAGCACAGACGCTTAAATAAAATTCTGCTCTGCAAGTCACAGCATGATAACAAACCTCTAAAGCACATTGCC
GCTCTTAACTGGCTTACAGGCGGGTGTGACCTCGAGACCTTCCCACCTGTCTCACAGTCTCTT
CAACAGCTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAA
GGCTGCGGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAA
FCITGCAACCCGGCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAA
CGCTGCTCAGGGATCTGGACATACCAAGTCATCTGCAACCTCTGGACAGCACAGACGGACACACCCA
TGTCGAACAGACAACTGTCAATTATTTGAAATCCA AAAAACAGGTTTAAACATCACAGGACATGGAATT
TAGCTCTGGAGACCTCTGGACAACTTCTGCAACCTCTGGCAAGAGCTTCACTTCAAACTGGCTTCA
CAAGGGCAATGGCCTAATGGCCTAACATCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAACTTCTGAA
GCTACTACTTCACTGAGTTACTTAACTTCACTGTTTAACTTCACTGAGTTTCACTGAGTTTCACTGAGTTTCACTGAGTTTCACTGAGTT
TTTACTCTGCTAATCTTCACTTCAAAATCTTAAAGGCTGTGGCCGCAACGATATCTTGGAA

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 with the 'History' tab selected. The main area displays a list of sequence files:

- Crimea Congo Reference Genome**
 - 3 shown
 - 19.3 KB
- 3: L_EU044832.fasta**
- 2: M_EU037902.fasta**
- 1: S_DQ133507.fasta**

Below the list, there is a button labeled "display with IGV local". The bottom part of the window shows the sequence content of the selected file (S_DQ133507.fasta) in a green-highlighted box:

```
>DQ133507.1 Crimean-Congo hemorrhagic fever virus complete genome
TCTCAAGAACACCGTCCGCTTACGCCAACAGTGTCTCT
GAGGTGAACAGCAAAAGTAGAGATGAACAAATGGTTGAGGA
CTTCACAAACTCTACTCCTTGTGAGAATGACCAAATC
CGCCACTGTGATGCACAGAAGGACTCATCTATGCATCGG
```

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 poly-A sequence
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:

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

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

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

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

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's a search bar labeled "buscar conjuntos de datos". Below it, the title "Crimea Congo Reference Genome" is displayed, followed by "3 shown" and a size indicator "19.3 KB". There are three items listed: "3: L_fragment.fasta", "2: M_fragment.fasta", and "1: S_fragment.fasta", each with a small icon set next to them.

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. On the left, the search bar contains "concatenate". A callout box labeled "1" points to the search bar. Another callout box labeled "2" points to the "Concatenate datasets tail-to-head (cat)" tool in the search results. The right side of the screen shows the Galaxy History interface with the same three datasets as the previous screenshot.

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 main panel displays a success 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 tools based on machine learning analysis: heatmap2, Sort, UMI-tools extract, Manipulate AnnData, Molecule to fingerpr..., Closed-reference OTU..., FastQC, Trimmomatic, and Bowtie2. The 'History' panel on the right shows the new job entry: '4: Concatenate datasets on data 3, data 2, and data 1' along with the three input 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 eye edit delete
3: L_fragment.fasta eye edit delete
2: M_fragment.fasta eye edit delete
1: S_fragment.fasta eye edit delete

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 2

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

Fasta Statistics Display summary statistics for a fasta file. (Galaxy Version 1.0.3)

fasta or multifasta file

1 eye edit delete 2 eye edit delete 3 eye edit delete

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

fastaa dataset to get statistics for.

Genome size estimate (optional)

Email notification
 No

Send an email notification when the job completes.

Execute 3

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 eye edit delete
3: L_fragment.fasta eye edit delete
2: M_fragment.fasta eye edit delete
1: S_fragment.fasta eye edit delete

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.

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

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.

	Value
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.

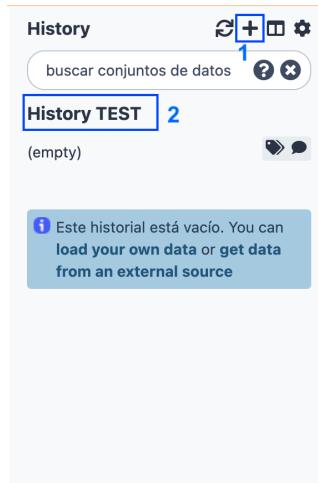
The screenshot shows the Galaxy Europe web interface. On the left, there's a sidebar with various tools categorized under 'Tools' (Search tools, Upload Data), 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS' (Text Manipulation, Filter and Sort, Join, Subtract and Group), 'GENOMIC FILE MANIPULATION' (Convert Formats, FASTA/FASTQ, Quality Control, SAM/BAM, BED, VCF/BCF, Nanopore), and 'COMMON GENOMICS TOOLS'. The main content area has sections for 'COVID-19 Research!', 'News' (Training Infrastructure feedback: FORGEs eScience course, UseGalaxy.eu Tool Updates for 2021-11-13), and 'Events' (Nov 2, 2021 - Nov 23, 2021: Forces 2021, Nov 16, 2021 - Nov 17, 2021: 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' (8 shown, 39.39 KB). Item 1 is '8: Fasta Statistics on data a 4: Fasta summary stats'. Item 2 is '2: 1. Crimea-Congo hemorrhagic fever display with IGV local'.

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 web interface. At the top, there's a banner with maintenance information. Below it, the main workspace has two panels: 'Historial actual' (left) and 'Reference Genome' (right). In the 'Historial actual' panel, there's a dropdown menu with 'Borrar' (Step 2) highlighted. Above the dropdown is a 'Switch to' button (Step 3), which is also highlighted with a blue 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**

This screenshot shows a history panel titled 'Crimea Congo Reference Genome'. It includes a toolbar with a gear icon (Step 1), a search bar, and a 'share conjuntos de datos' button (Step 2). The main area lists several datasets with edit and delete icons. At the bottom, there's a text input field containing 'share_history_1' (Step 3).

The screenshot shows the Galaxy Europe web interface. On the left, a sidebar lists various tool categories: 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. Under Tools, there's a 'search tools' input field and a 'Upload Data' button. The main content area displays a history titled 'Crimea Congo Reference Genome'. It includes a 'Share or Publish History' section with two radio buttons: 'Make History accessible' (selected) and 'Make History publicly available in Published Histories'. 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>. A blue box highlights this URL with the number '2'. To the right, a 'History' panel shows the contents of the history, which include 8 items: 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, and 1: S_fragment.fasta. Each item has edit and delete icons.

Now everyone with the link can access the 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>