

# Galaxy for virologist training Exercise 1: Introduction to Galaxy

---

<b>Title</b>	<b>Galaxy</b>
<b>Training dataset:</b>	None
<b>Questions:</b>	<ul style="list-style-type: none"><li>• How do I create a fasta reference for Crimea Congo?</li><li>• How many nucleotides has each fragment of Crimea Congo genome?</li></ul>
<b>Objectives:</b>	<ul style="list-style-type: none"><li>• Familiarize with Galaxy website</li><li>• Understand the Galaxy's history</li><li>• Learn how to upload data in Galaxy</li><li>• Learn how to visualize data in Galaxy</li><li>• Learn how to run tools in Galaxy</li><li>• Learn how to create a workflow</li><li>• Learn how to load a workflow in Galaxy</li></ul>

---

**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. It features a top navigation bar with links for Workflow, Visualize, Shared Data, Help, Login or Register, and a scratchbook icon. A yellow border highlights the top navigation. A blue border highlights the left sidebar containing a list of tools categorized under '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'. A red border highlights the central content area which includes a 'COVID-19 Research!' section with text about SARS-CoV-2 data analysis, a 'News' section listing tool updates for various dates, and an 'Events' section listing various conferences and workshops. A green border highlights the right sidebar titled 'History' which shows an 'Unnamed history' entry and a message indicating it is currently empty.

Where you have 4 different elements:

1. The first one in yellow is the Title panel with the buttons:
  - o Home (house): To go to the home page in Spanish
  - o Workflows: To go to the workflow manager
  - o Visualize: Displays the visualization manager and options
  - o Share Data: Displays the sharing options
  - o Help: Displays all the help menu available
  - o Login or Register
  - o Galaxy Training Materials (graduation cap): Displays de Galaxy Trainings list
  - o 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.

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. **⚠** 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.

**COVID-19 Research!**

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the [Galaxy SARS-CoV-2 portal](#). We mirror **all public SARS-CoV-2 data** from ENA in a [Galaxy data library](#) for your convenience. The Galaxy community has created [COVID-19 dedicated training materials](#). Please check our recent activities for more details.

If you need help submitting your data to public archives, like ENA, please [get in touch](#). We will support you in sharing your data.

"Anyone, anywhere in the world should have free, unhindered access to not just my research, but to the research of every great and enquiring mind across the spectrum of human understanding." – Prof. Stephen Hawking

**News**

- 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](#)
- Oct 16, 2021 [Training Infrastructure Feedback from Dr. Theodora Tsirka](#)

**Events**

- Nov 2, 2021 - Nov 23, 2021 [Galaxy 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 [NRZ-Authent Expertinnen- und Expertenworkshop](#)
- Nov 18, 2021 [OPEN CHAT](#)

Welcome to Galaxy, please log in

Public Name or Email Address

Password

Forgot password? Click [here](#) to reset your password.

[Login](#)

[elxir LOGIN](#)

Don't have an account? [Register here.](#)

**COVID-19 Research!**

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the [Galaxy SARS-CoV-2 portal](#). We mirror **all public SARS-CoV-2 data** from ENA in a [Galaxy data library](#) for your convenience. The Galaxy community has created [COVID-19 dedicated training materials](#). Please check our [recent activities](#) for more details.

If you need help submitting your data to public archives, like ENA, please [get in touch](#). We will support you in sharing your data.

"Anyone, anywhere in the world should have free, unhindered access to not just my research, but to the research of every great and enquiring mind across the spectrum of human understanding." – Prof. Stephen Hawking

**News**

- 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](#)

[OPEN CHAT](#)

The screenshot shows two pages side-by-side. On the left, the 'Create a Galaxy account' form is displayed, requiring fields for Email Address, Password, Confirm password, and Public name. A note specifies that public names must be at least three characters long and contain only lowercase letters, numbers, dots, underscores, and dashes. A 'Create' button is present. Below the form, a link says 'Already have an account? Log in here.' On the right, the 'GDPR Compliance Documentation' page is shown, featuring sections for 'ToS & PP' (with links to Privacy Policy and Terms of Service), 'GDPR Documentation' (with links to Your Rights Under the GDPR, Legitimate Interest Analyses, Data Storage and Access, and Data Processing Activities Register), and 'Contact' information for the Bioinformatics Group at Albert-Ludwigs-University Freiburg.

**Optional:** Logging into TiaaS server: [https://usegalaxy.eu/join-training/virologist\\_isciii](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:

1. Click the new-history (+) icon at the top of the history panel.
  - o 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
2. 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 three screenshots illustrate the process of renaming a history. The first shows the initial state with an 'Unnamed history' entry. The second shows the 'Haz clic para cambiar el nombre del historial' (Click to change history name) dialog box open over the history entry. The third shows the renamed history titled 'Crimea Congo Reference Genon'.

## 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/data/S_DQ  
133507.fasta  
https://raw.githubusercontent.com/BU-  
ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/data/M_EU  
037902.fasta  
https://raw.githubusercontent.com/BU-  
ISCIII/galaxy_virologist_training/one_week_4day_format/exercises/data/L_EU  
044832.fasta
```

4. 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
5. Select **Start**
6. When everything is green in the screen, select **Close**

The screenshot shows the Galaxy web interface. On the left, a sidebar lists 'Tools' like 'Upload Data', 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS' (Text Manipulation, Filter and Sort, Join, Subtract and Group), and 'GENOMIC FILE MANIPULATION' (Convert Formats, FASTA/FASTQ, Quality Control, SAM/BAM, BED, VCF/BCF, Nanopore). Below these are 'COMMON GENOMICS TOOLS' (Operate on Genomic Intervals). A central panel displays a '24 hours downtime starting from November 22nd, 2021 at 17:45 pm CET' message. Another section titled 'COVID-19 Research!' provides information about SARS-CoV-2 analysis. A quote by Stephen Hawking is also present. The right side features a 'History' section for the 'Crimea Congo Reference Genome' and a message about an empty history. A news feed and event calendar are also shown.

## Descargar de la red o cargar desde disco

The screenshot shows the 'Upload Data' dialog. It has tabs for 'Regular', 'Composite', 'Collection', and 'Rule-based'. The 'Regular' tab is selected. A message says 'You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.' Below is a table with columns: Name, Size, Type, Genome, Settings, and Status. A row shows a file named 'New File' with size '386 b', type 'Auto-de...', genome '----- Additional ...', settings gear icon, status '0%', and a trash bin icon. A large text input box contains URLs for downloading data from GitHub. At the bottom are buttons for 'Elegir archivos locales', 'Choose remote files', 'Paste/Fetch data' (which is highlighted with a blue border), 'Start' (highlighted with a blue border), 'Pause', 'Reset', and 'Close'.

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 web interface with three main panels:

- Left Panel (History):** Shows a search bar "buscar conjuntos de datos" and a list titled "Crimea Congo Reference Genome" with 3 items: "3 shown" (empty), "3: L\_EU044832.fasta", "2: M\_EU037902.fasta", and "1: S\_DQ133507.fasta".
- Middle Panel (Data Preview):** Shows a search bar "buscar conjuntos de datos" and a list titled "Crimea Congo Reference Genome" with 3 items: "3 shown" (empty), "3: L\_EU044832.fasta", "2: M\_EU037902.fasta", and "1: S\_DQ133507.fasta".
- Right Panel (Data View):** Shows a search bar "buscar conjuntos de datos" and a list titled "Crimea Congo Reference Genome" with 3 items: "3 shown" (19.3 KB). The items are listed as follows:
  - 3: L\_EU044832.fasta (highlighted in green)
  - 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	  

**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

```
>DQ133507.1 Crimean-Congo hemorrhagic fever virus strain Kosovo Hoti segment S, complete sequence
TCTCAAAGAACACGTGCGCTTACGCCAACAGTGTTCCTCTGAGTGTCTCAAAATGAAACAAAGAT
GAGGGTGAACAGCAAAAGATGAGATGAAACAAATGGTTGGAGAGTTAAAAAGGGAAATGACTTATGGCA
CTTCACAAACTACTCCCTTGGAGAATGACCAAAATCTGGAATAGTTTGTTCCAGATGCCAG
CCCCACTGTGATGACAGAACAGACTCCATCTATGCATCGGCTCTGGTGGAAACAACAAAGTTCTGCA
CCCCATATGATGATGCTTGGCTGAGCTACTGGCATTTGAGAAGAAGGGCTTGGATGTTTGGAGAGA
ATTCAAGGAAACCAATCTGGATGAGACTATGCTGAGCTGAAGGTTGATGTTCCAAAATAGAACAA
ACTTGCCAAATTACACACGGCTGCTCAAGTGGAGGAAGGACATAGGTTTCTGCTTAATGCAACACAG
GCAGCCTTAAGCACAAAGTCTTGGCAAAATATAAGCTCTGGGAATTGTAATGCTGTTAAAGAAA
TGCTGTGAGACATGATTGAGAGGAGAAATTAAATCTCAACAGGGGGGTGATGAAATTCACCGGGCC
AGTGACGCCGTGAAATTGGAGTGTGCAAGGAGATTGTAAGGAAATGACTACATGCTCAATCCA
CTTTGGGGGAGCATCAACAAATCAGGCCGTCAAGGAATGACACTTGTGCAACAGGCCCTGGCAAGCTTG
CAAGAGACCGAGGGAAAGGAGTCTTGAGCAAGAACAGACCTGGTGGACCTTAATGGTATTGG
CAAGCACACAGGAGCAAGTTGACAAAGCAAGTGGCAACAGCATGAAATACAACACTCTTAAGACATTTGC
AAAGCACAAAGGTTTATAAAATTCAATGCTCTTCGACAAAGTGCACAGATTAACACTCTTCA
GCTCGTTTACTGGCTCTACAGGCCGGTGTGACTCCAGAGACCTTCCAACTGTCACAGTTCTTCTT
CGAAGCTGGGGAGCAGCCAAAGGGGACCAAAAAAAATGAAAAAGGACACTCTGACACTCCAAATGAAGTGG
GGGAAGAAGCTTTAATGAGCTCTTGGCTGATGAACTCTTCAGCAGAACAGAACTCATGCACTCTGGCTG
TGTGACAGCCGTAGAAATCAGTGAAATGGTGTCTGCTTGGAAATCCCTGTTGCAATCTGATGTA
CGCTGCTCAGGGATCTGGACATACAAAGTCATTCACCTTGGACAAAGCACAGAACAAACACCA
TGTGCGCAAGACAAATTGTCAAATTATTGAAATCCAAAATGAACTTACAGGATTAACATCAGGACATGGACATTG
TAGCCTCTGAGCACTGCTGACCAATCCCTTGTGGCAAGCAGTCCTCATTCCAAAATGCTTACAAATG
CAAGGGCAATGCTTACAGTTGACTTAAATCACGTTTATTTAATGCTTATATAATGCTGTTTGTCAA
GCATCAGTACTACAGTTGACTTAAATCACGTTTATTTAATGCTTATATAATGCTGTTTGTCAA
TTTATCTGCTATTTCAATTAACTTAAAGGGCTGTGCGGCAACGATATCTTGAGA
```

**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	  

2. Another way is to select the name of the file to see the first five lines of the 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 sequences  
formato: **fasta**, base de datos: ?

display with IGV local

```
>DQ133507.1 Crimean-Congo hemorrhagic fever virus
TCTCAAAGAACACGTGCCCTTACCCCCACAGTGTCTCT
GAGGTGAACAGCAAAGATGAGATGAACAAATGGTTGAGGA
CTTCACAAACTCCTACTCCCTTGCGAGAATGTACCAAT
GCCACTGATGATGCACAGAAGGACTCCATCTATGCATCGG
```

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
TCTCAAAGAACACGTGCCCTTACCCCCACAGTGTCTCT
GAGGTGAACAGCAAAGATGAGATGAACAAATGGTTGAGGA
CTTCACAAACTCCTACTCCCTTGCGAGAATGTACCAAT
GCCACTGATGATGCACAGAAGGACTCCATCTATGCATCGG
```

- **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, subtype L, isolate CCU/95/1, complete genome
TCTCAAAGAACACGTGCCGTTACGCCAACAGTGTTCTCTGAGGTGAACAGCAAAGATGAGATGAACAAATGGTTGAGGA
CTTCACAAACTCCTACTCCTTGCAGAATGTACCAATCGCCACTGATGATGCACAGAAGGACTCCATCTATGCATCGG
```

- *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
TCTCAAAGAACACGTGCCGTTACGCCAACAGTGTTCTCT
GAGGTGAAACAGCAAGATGAGATGAACAAATGGTTGAGGA
CTTTCACAAACTCCTACTCTTTGCAGAATGTACCAAATC
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	Data Fetch	History
search tools		buscar conjuntos de datos
Upload Data		Crimea Congo Reference Genome
Get Data		3 shown
Send Data		19.3 KB
Collection Operations		
GENERAL TEXT TOOLS		?
Text Manipulation		display with IGV local
Filter and Sort		>DQ133507.1 Crimean-Congo hemorrhagic fever
Join, Subtract and Group		TCTCAAAGAACACGTGCCGTTACGCCAACAGTGTTCTCT
GENOMIC FILE MANIPULATION		GAGGTGAAACAGCAAGATGAGATGAACAAATGGTTGAGGA
Convert Formats		CTTTCACAAACTCCTACTCTTTGCAGAATGTACCAAATC
FASTA/FASTQ		CGCCACTGATGATGCACAGAAGGACTCCATCTATGCATCGG
Quality Control		
SAM/BAM		
BED		
VCF/BCF		
Nanopore		
COMMON GENOMICS TOOLS		
Operate on Genomic Intervals		

- **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
TCTCAAAGAACACGTGCCGTTACGCCACAGTGTTCTCT
GAGGTGAACAGCAAAGATGAGATGAACAAATGGTTGAGGA
CTTCACAAACTCTACTCTTTGCAGAATGTACCAATG
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 visualizations

**Editor**  
Manually edit text

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

**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
TCTCAAAGAACACGTGCCGTTACGCCACAGTGTTCTCT
GAGGTGAACAGCAAAGATGAGATGAACAAATGGTTGAGGA
CTTCACAAACTCTACTCTTTGCAGAATGTACCAATG
CGCCACTGATGATGCACAGAAGGACTCCATCTATGCATCGG
```

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

The screenshot shows the Galaxy interface with the following details:

- History**: Shows 3 items: 3: L\_EU044832.fasta, 2: M\_EU037902.fasta, and 1: S\_DQ133507.fasta.
- Search Bar**: "buscar conjuntos de datos" with a help icon (?) and a close button (X).
- Summary Panel**:
  - Crimea Congo Reference Genome**
  - Genome**
  - 3 shown
  - 19.3 KB
  - Icons for edit, delete, and other actions.
- Sequence View Panel**:
  - 3: L\_EU044832.fasta, 2: M\_EU037902.fasta, 1: S\_DQ133507.fasta (with edit icons)
  - 1 sequences
  - formato: **fasta**, base de datos: ?
  - Icons for IGV local, copy, download, and help (?).
  - Text: "display with IGV local".
  - Sequence content:

```
>DQ133507.1 Crimean-Congo hemorrhagic fever virus, subtype C, isolate 90/1, 5' end
TCTCAAAGAACACGTGCCGTTACGCCAACAGTGTTCTCT
GAGGTGAACAGCAAAGATGAGATGAACAAATGGTTGAGGA
CTTCACAAACTCCTACTCCTTGCAGAATGTACCAATG
CGCCACTGATGATGCACAGAAGGACTCCATCTATGCATCGG
```

**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

**Attributes**

Name: S\_DQ133507.fasta

Info:

Annotation:

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

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.

**⚠ Select **Save** button to save the changes.**

We are going to rename the files as shown here:

History

buscar conjuntos de datos

**Crimea Congo Reference Genome**

3 shown

19.3 KB

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)**.

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**

concatenate 1

Upload Data

Show Sections

VCFgenotype-to-haplotype: Convert genotype-based phased alleles into haplotype alleles

LC/MS matching: Annotation of MS peaks using matching on a spectra database.

Concatenate images

Concatenate FASTA alignment by species

Concatenate datasets tail-to-head 2

Concatenate datasets tail-to-head (cat)

AXT to concatenated FASTA Converts an AXT formatted file to a concatenated FASTA alignment

Concatenate two BED files

FASTA Merge Files and Filter Unique Sequences Concatenate FASTA database files together

bcftools concat Concatenate or combine VCF/BCF files

https://usegalaxy.eu/tool\_runner?tool\_id=toolshed.o2bx.nsu.edu%2Frenos%2Frrma%2Ficmsmatching%2Ficmsmatching%2F3.0

**24 hours downtime starting from November 24th, 2021 at 17:45 pm CET**

On November 24th, 2021 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.

During this time we will be upgrading Galaxy to the latest version (21.09), migrating/upgrading the DB server and performing other maintenance.

Please take it into account in your job schedule.

**COVID-19 Research!**

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the [Galaxy SARS-CoV-2 portal](#). We mirror all [public SARS-CoV-2 data](#) from ENA in a [Galaxy data library](#) for your convenience. The Galaxy community has created [COVID-19 dedicated training materials](#). Please check our [recent activities](#) for more details.

If you need help submitting your data to public archives, like ENA, please [get in touch](#). We will support you in sharing your data.

"Anyone, anywhere in the world should have free, unhindered access to not just my research, but to the research of every great and enquiring mind across the spectrum of human understanding." – Prof. Stephen Hawking

**News**

Nov 13, 2021 Training Infrastructure feedback: FORCeS eScience course

Nov 13, 2021 UseGalaxy.eu Tool Updates for 2021-11-13

**Events**

Nov 2, 2021 - Nov 23, 2021 Forces 2021

Nov 16, 2021 - Nov 17, 2021 5. NRZ-Authent Expertinnen- und Expertenworkshop

Nov 18, 2021

OPEN CHAT

### 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. 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)
- Browse a dataset (you can browse dataset from the history)

3. Insert new dataset blocks (no need in our case)

4. Execute button

5. Tool information:

- !
- What it does
- Examples
- Citation

To concatenate the samples, we will follow the following 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: 'concatenate', 'Upload Data', 'Show Sections', '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. It has a 'Datasets to concatenate' section containing three files: '3: L\_fragment.fasta', '2: M\_fragment.fasta', and '1: S\_fragment.fasta'. A blue box highlights the 'Ctrl' key being held down over the file names. Below this, there's an 'Email notification' section with a 'No' radio button and a note about sending an email when the job completes. At the bottom is a large 'Execute' button with a checkmark, also highlighted with a blue box and the number '2'. A warning message below the button cautions against concatenating datasets of different kinds if they are not in the same format. To the right, the 'History' panel shows a new entry: '4: concatenated.fasta' (19.3 KB). The 'History' panel also includes sections for 'Crimea Congo Reference Genome' and a search bar.

## 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 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.

**Galaxy Europe**

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

[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**

concatenate

Upload Data

Show Sections

**VCFgenotype-to-haplotype:** Convert genotype-based phased alleles into haplotype alleles

**LC/MS matching:** Annotation of MS peaks using matching on a spectra database.

**Concatenate images**

**Concatenate FASTA alignment by species**

**Concatenate datasets tail-to-head (cat)**

**AXT to concatenated FASTA** Converts an AXT formatted file to a concatenated FASTA alignment

**Concatenate two BED files**

**FASTA Merge Files and Filter Unique Sequences** Concatenate FASTA database files together

**bcftools concat** Concatenate or combine VCF/BCF files

**concatenate** Executed **Concatenate datasets** and successfully added 1 job to the queue.

The tool uses 3 inputs:

- 1: S\_fragment.fasta
- 2: M\_fragment.fasta
- 3: L\_fragment.fasta

It produces this output:

- 4: Concatenate datasets on data 3, data 2, and data 1

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.

**Tool recommendation**

You have used tp\_cat tool. For further analysis, you could try using the following/recommended tools. The recommended tools are shown in the decreasing order of their scores predicted using machine learning analysis on workflows. Therefore, tools at the top may be more useful than the ones at the bottom. Please click on one of the following/recommended tools to open its definition.

Concatenate datasets

- heatmap2
- Query Tabular
- Sort
- UMI-tools extract
- Manipulate AnnData
- Molecule to fingerpr...
- Closed-reference OTU...
- FastQC
- Trimmomatic
- Bowtie2

**History**

buscar conjuntos de datos

**Crimea Congo Reference Genome**

4 shown

19.3 KB

1: 4: Concatenate datasets on data 3, data 2, and data 1

3: L\_fragment.fasta

2: M\_fragment.fasta

1: S\_fragment.fasta

## 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.

**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**

- concatenate

**Upload Data**

**Show Sections**

**VCFgenotype-to-haplotype:** Convert genotype-based phased alleles into haplotype alleles

**LC/MS matching:** Annotation of MS peaks using matching on a spectra database.

**Concatenate images**

**Concatenate FASTA alignment by species**

**Concatenate datasets tail-to-head**

**Concatenate datasets tail-to-head (cat)**

**AXT to concatenated FASTA:** Converts an AXT formatted file to a concatenated FASTA alignment

**Concatenate two BED files**

**FASTA Merge Files and Filter Unique Sequences:** Concatenate FASTA database files together

**bcftools concat**: Concatenate or combine VCF/BCF files

**History**

buscar conjuntos de datos

**Crimea Congo Reference Genome**

4 shown  
38.59 KB

**4: Concatenate datasets on data 3, data 2, and data 1**

**3: L\_fragment.fasta**

**2: M\_fragment.fasta**

**1: S\_fragment.fasta**

Now we are going to rename the fasta file as follows:

1. Click on the  icon
  2. Write **Crimea Congo Ref Genome** in the *Name* square
  3. Press **Save**

The screenshot shows the Galaxy Europe web interface. On the left, a sidebar lists various tools: concatenate, 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 'concatenate' tool is currently selected. The main area is titled 'Edit dataset attributes' and contains tabs for Attributes, Convert, Datatypes, and Permissions. The 'Attributes' tab is active, showing a 'Name' input field containing 'Crimea Congo Ref Genome' with a blue border around it. To the right of the input field are 'Auto-detect' and 'Save' buttons, with the 'Save' button also having a blue border. Below the name input is a large empty text area for 'Info'. Under 'Annotation', there is a note: 'Add an annotation or notes to a dataset; annotations are available when a history is viewed.' A 'Database/Build' dropdown menu shows '----- Additional Species Are Below -----'. On the far right, a 'History' panel displays four datasets: '4: Concatenate datasets on data 3, data 2, and data 1' (highlighted with a green background and labeled '1'), '3: L\_fragment.fasta', '2: M\_fragment.fasta', and '1: S\_fragment.fasta'. Each dataset has edit and delete icons next to it.

## 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**

**Editar 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**

Crimea Congo Reference Genome  
4 shown  
38.59 KB

4: Crimea Congo Ref Genome 2  
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. (Galaxy Version 1.0.3) 2

**fasta or multifasta file**

4: Crimea Congo Ref Genome  
3: L\_fragment.fasta  
2: M\_fragment.fasta  
1: S\_fragment.fasta 3

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

fast dataset to get statistics for.

**Genome size estimate (optional)**

Estimate of the genome size in bases. If specified, NG50 and LG50 will be calculated.

**Email notification**

No

Send an email notification when the job completes.

**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**

Crimea Congo Reference Genome  
4 shown  
38.59 KB

4: Crimea Congo Ref Genome 2  
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.

The screenshot shows the Galaxy Europe web interface. On the left, a sidebar lists various bioinformatics tools. In the center, a green box indicates that the 'Fasta Statistics' tool has been executed successfully, adding 4 jobs to the queue. The tool uses 4 inputs: S\_fragment.fasta, M\_fragment.fasta, L\_fragment.fasta, and Crimea Congo Ref Genome. It produces 4 outputs: Fasta Statistics on data 1-4. A note at the bottom says you can check the status of queued jobs and view resulting data by refreshing the History panel. On the right, the History panel shows the completed job details.

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

## 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.

The screenshot shows the Galaxy Europe interface. On the left, a sidebar lists various bioinformatics tools. In the center, the results of the 'GC\_content' tool are displayed, showing GC content values for different samples. The 'num\_bp' and 'num\_seq' rows are highlighted with red boxes. On the right, the History panel shows the completed job details for the Fasta Statistics tool.

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

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

## M fragment

- ▶ How many nucleotides are in M fragment?

5364 nt

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

**Tools** (Search term: `fasta`):

- GC\_content: 45.1
- L50: 1
- L90: 1
- len\_N50: 5364
- len\_N90: 5364
- len\_max: 5364
- len\_mean: 5364
- len\_median: 5364
- len\_min: 5364
- num\_A: 1651
- num\_C: 1235
- num\_G: 1186
- num\_N: 0
- num\_T: 1292
- num\_bp: 5364 (highlighted)
- num\_bp\_not\_N: 5364
- num\_seq: 1 (highlighted)

**History** (Search term: `buscar conjuntos de datos`):

- Crimea Congo Reference Genome** (8 shown, 39.39 KB)
  - 8: Fasta Statistics on dat a 4: Fasta summary stats
  - 7: Fasta Statistics on data 3: Fasta summary stats
  - 6: Fasta Statistics on dat a 2: Fasta summary stats
  - 5: Fasta Statistics on dat a 1: Fasta summary stats
  - 4: Crimea Congo Ref Gen ome
  - 3: L\_fragment.fasta
  - 2: M\_fragment.fasta
  - 1: S\_fragment.fasta

## L fragment

- ▶ How many nucleotides are in L fragment?

12150 nt

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

**Tools** (Search term: `fasta`):

- GC\_content: 41.4
- L50: 1
- L90: 1
- len\_N50: 12150
- len\_N90: 12150
- len\_max: 12150
- len\_mean: 12150
- len\_median: 12150
- len\_min: 12150
- num\_A: 3928
- num\_C: 2372
- num\_G: 2661
- num\_N: 0
- num\_T: 3189
- num\_bp: 12150 (highlighted)
- num\_bp\_not\_N: 12150
- num\_seq: 1 (highlighted)

**History** (Search term: `buscar conjuntos de datos`):

- Crimea Congo Reference Genome** (8 shown, 39.39 KB)
  - 8: Fasta Statistics on dat a 4: Fasta summary stats
  - 7: Fasta Statistics on dat a 3: Fasta summary stats
  - 6: Fasta Statistics on dat a 2: Fasta summary stats
  - 5: Fasta Statistics on dat a 1: Fasta summary stats
  - 4: Crimea Congo Ref Gen ome
  - 3: L\_fragment.fasta
  - 2: M\_fragment.fasta
  - 1: S\_fragment.fasta

## Crimea Congo Genome

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

19187 nt

	Value
GC_content	42.8
L50	1
L90	2
len_N50	12150
len_N90	5364
len_max	12150
len_mean	6395
len_median	5364
len_min	1673
num_A	6080
num_C	3982
num_G	4234
num_N	0
num_T	4891
num_bp	19187
num_bp_not_N	19187
num_seq	3

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, a sidebar lists various tools and operations. A central message box displays a maintenance notice for November 24th. Below it, a green box contains COVID-19 research information. To the right, there's a news section with two items, an events section with one item, and a history panel on the far right showing a list of recent histories.

## 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**

The screenshot shows the 'History TEST' panel. It has a header with a search bar and a 'buscar conjuntos de datos' button. Below is a list area with '(empty)' and a note: 'Este historial está vacío. You can load your own data or get data from an external source'. There are also icons for sharing and reporting.

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:

The screenshot shows the Galaxy History interface. At the top, there are buttons for History, Refresh, Create New, and Settings. Below that is a search bar with placeholder text 'buscar conjuntos de datos' and a help/x icon. The main area is titled 'History TEST' and contains the message '(empty)'. There are two small icons at the bottom right of the history list. A blue callout box in the center says: 'Este historial está vacío. You can load your own data or get data from an external source'.

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 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. The top navigation bar includes 'Galaxy Europe', a red 'Flujo de Trabajo' button, and other menu items like 'Visualizar', 'Datos Compartidos', 'Ayuda', 'Usuario', and 'Logout'. 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 the navigation is a search bar for histories and datasets. The main area shows a history named 'History TEST' which is currently empty. A dropdown menu is open over the history, with item '2' pointing to the 'Borrar' option. Another dropdown menu is open over the dataset list, with item '3' pointing to the 'Switch to' option. The dataset list itself contains several entries, each with edit and delete icons.

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 History interface. At the top, there are buttons for Refresh, Add, Delete, and Settings. Below that is a search bar with placeholder text "buscar conjuntos de datos" and a help/x icon. The main title is "Crimea Congo Reference Genome". It displays "8 shown" datasets and a total size of "39.39 KB". On the right, there are three small icons: a checkbox, a heart, and a speech bubble.

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

**History Actions**

- Copy
- Compartir o Publicar
- Mostrar Estructura
- Extraer Flujo de Trabajo
- Set Permissions
- Make Private
- Reanudar Trabajos en Pausa

**Acciones de Conjuntos de Datos**

- Copiar Conjuntos de Datos
- Contraer Conjuntos de Datos Expandidos
- Mostrar Conjuntos de Datos Ocultos
- Eliminar Conjuntos de Datos Ocultos
- Eliminar Definitivamente los Conjuntos de Datos Eliminados

**Descargas**

- Exportar las Citas de la Herramienta
- Exportar Historial a un Archivo

**Beta Features**

- Use Beta History Panel

The screenshot shows the Galaxy Europe interface. The left sidebar includes sections for Tools (with a search bar), Get Data, Send Data, Collection Operations, GENERAL TEXT TOOLS, GENOMIC FILE MANIPULATION, and COMMON GENOMICS TOOLS. The main area shows the "Share or Publish History 'Crimea Congo Reference Genome'" page. It has two radio buttons: "1 Make History accessible" (selected) and "2 Make History publicly available in Published Histories". Below this, it says "This History is currently accessible via link. Anyone can view and import this History by visiting the following URL:" followed by a URL field containing "url: https://usegalaxy.eu/u/svarona/h/crimea-congo-reference-genome". To the right, the Galaxy History interface is shown again, identical to the one in the first screenshot, displaying the same 8 datasets and details.

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**

**History Actions**

- Copy
- Compartir o Publicar
- Mostrar Estructura
- Extraer Flujo de Trabajo**
- Set Permissions
- Make Private
- Reanudar Trabajos en Pausa

**Acciones de Conjuntos de Datos**

- Copiar Conjuntos de Datos
- Contraer Conjuntos de Datos Expandidos
- Mostrar Conjuntos de Datos Ocultos
- Eliminar Conjuntos de Datos Ocultos
- Eliminar Definitivamente los Conjuntos de Datos Eliminados

**Descargas**

- Exportar las Citas de la Herramienta
- Exportar Historial a un Archivo

**Beta Features**

- Use Beta History Panel

**Tools**

search tools

**Workflow name** **1**  
Create Crimea Congo Reference Genome

**Create Workflow** **2**

**Check all** **Uncheck all**

**History items created**

- 1 S\_fragment.fasta
- 2 M\_fragment.fasta
- 3 L\_fragment.fasta
- 4 Crimea Congo Ref Genome
- 5 Fasta Statistics on data 1: Fasta summary stats
- 6 Fasta Statistics on data 2: Fasta summary stats

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

The screenshot shows the Galaxy Europe web interface. At the top, there's a navigation bar with links like 'Flujo de Trabajo' (Workflow), 'Visualizar', 'Datos Compartidos', 'Ayuda', 'Usuario', and a search bar. A message at the top states: '[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.'

The main area is titled 'Workflow "Create Crimea Congo Reference Genome" created from current history. You can edit or run the workflow.' To the left, a sidebar titled 'Tools' lists categories: '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). There's also a 'Upload Data' button.

To the right, the 'History' section shows the workflow steps under the title 'Crimea Congo Reference Genome'. It lists 8 steps:

- 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

Each step has edit and delete icons. The total size of the workflow is 39.39 KB.

## 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**
3. 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.

**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**

**Create Crimea Congo Reference Genome**

1

2

Tags Updated Sharing Bookmarked

a few seconds 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  
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.

**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 Genome

Version: 1: Nov 17th 2021, 8 steps

Annotation:

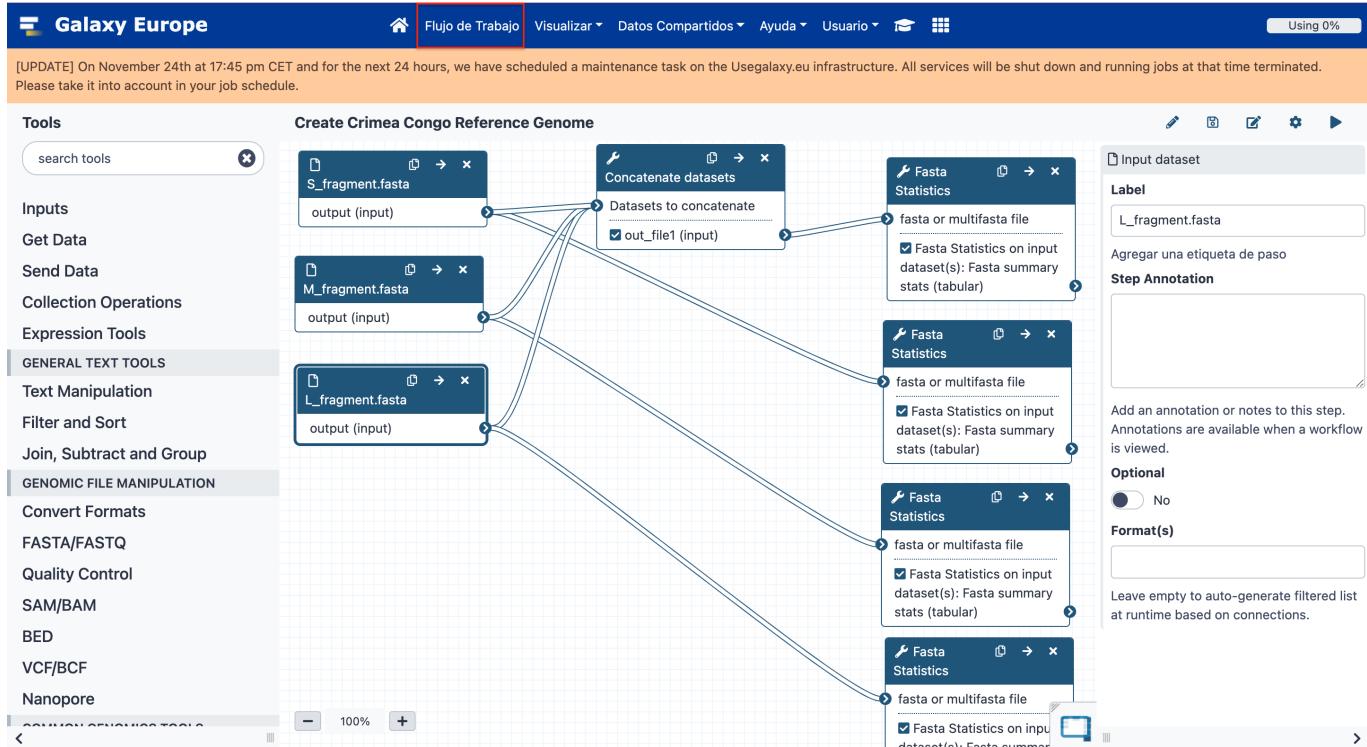
These notes will be visible when this workflow is viewed.

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.



## 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.

The screenshot shows the Galaxy Europe web interface. On the left, a sidebar lists various tools under categories like 'GENERAL TEXT TOOLS' and 'GENOMIC FILE MANIPULATION'. The main area displays a workflow named 'Create Crimea Congo Reference Genome'. A context menu is open over this workflow, with item 1 ('Share') highlighted. To the right, a 'History' panel shows a list of recent datasets, including 'Crimea Congo Reference Genome' and several Fasta files (L\_fragment.fasta, M\_fragment.fasta, S\_fragment.fasta). The top navigation bar includes links for 'Flujo de Trabajo', 'Visualizar', 'Datos Compartidos', 'Ayuda', 'Usuario', and a search bar.

This screenshot shows the details of the 'Create Crimea Congo Reference Genome' workflow. It includes sections for 'Share' (with a 'Make Workflow Accessible via Link' button highlighted), 'Export' (with a 'Download' button), and 'myExperiment' export options. The 'myExperiment' section requires a username ('svarona') and password ('\*\*\*\*\*').

**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](http://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](http://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-frags-reference-genome](https://usegalaxy.eu/u/svarona/w(concat-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:
  - o Left one to download the workflow
  - o 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 web interface. At the top, there's a navigation bar with links for 'Aplicaciones', 'Galaxy', and 'Otros marcadores'. Below the navigation is a search bar with the URL 'https://usegalaxy.eu/u/svarona/w(concat-frags-reference-genome)'. A message at the top of the page states: '!! [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.' To the left is 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 SERVICES TOOLS'. The main content area has sections for 'COVID-19 Research!', 'News' (Nov 13, 2021: Training Infrastructure feedback: FORCeS eScience course), and 'Events' (Nov 2, 2021 - Nov 23, 2021: Forces 2021, Nov 18, 2021: Galaxy Paper Cuts). On the right, there's a 'History' panel showing a list of datasets: 8: Fasta Statistics on data, 6: Fasta summary stats, 7: Fasta Statistics on data, 3: Fasta summary stats, 6: Concatenate datasets on data 1, data 2, and data 3, 5: Fasta Statistics on data, 2: Fasta summary stats, 4: Fasta Statistics on data, 1: Fasta summary stats, 3: L\_KX056050.fasta, 2: M\_KX056051.fasta, and 1: S\_KX056052.fasta. The bottom half of the page shows the 'Concatenate fragments Reference Genome' workflow. It has four steps: Step 1: Input dataset (S\_fragment.fasta, select at runtime), Step 2: Input dataset (M\_fragment.fasta, select at runtime), Step 3: Input dataset (L\_fragment.fasta, select at runtime), and Step 4: Fasta Statistics (fasta or multifasta file). To the right of the workflow, there are buttons for 'About this Workflow' (with download and import icons), 'Author' (svarona), 'Related Workflows' (All published workflows, Published workflows by svarona), 'Rating' (Community: 0 ratings, 0.0 average, Yours: 5 stars), and 'Tags' (Community: none, Yours: #). A large blue 'Download' button is also present.

## 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*
4. Select browse datasets in the  like icon for the S fragment
  5. Select the S fragment from the list
  6. 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, 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, and Convert Formats. The main area displays a workflow history. A search bar at the top says "Search Workflows". Below it, a table lists two workflows: "Create Crimea Congo Reference Genome" (updated an hour ago) and "FastQC" (updated 6 days ago). To the right, a "History" section shows an entry for "Isolate Ast199" with a status of "2" and "(empty)". A message box indicates that the history is empty and suggests loading your own data or getting data from an external source.

This screenshot shows the configuration page for the "Create Crimea Congo Reference Genome" workflow. The left sidebar is identical to the previous screenshot. The main area is titled "Workflow: Create Crimea Congo Reference Genome". It contains three input fields: "S\_fragment.fasta" (No txt, fastq.gz, fasta.gz, genbank.gz, tabular.gz or fasta dataset available), "M\_fragment.fasta" (No fasta, txt, fastq.gz, fasta.gz, genbank.gz or tabular.gz dataset available), and "L\_fragment.fasta" (No fasta, txt, fastq.gz, fasta.gz, genbank.gz or tabular.gz dataset available). A "Run Workflow" button is located to the right of the workflow title. The right side of the screen shows the same "History" section as the first screenshot.

## Descargar de la red o cargar desde disco

This screenshot shows the "File Queue" interface. At the top, there are tabs for "Regular" and "Composite", with "Regular" selected. A message below the tabs says "You added 1 file(s) to the queue. Add more files or click 'Start' to proceed." The main area is a table with columns: Name, Size, Type, Genome, Settings, and Status. One row is visible, showing "New File" with size "371 b", type "Auto-de...", genome "----- Additional ...", settings icon, and status "0%". Below the table, a box contains URLs for downloading files from GitHub: "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", and "https://raw.githubusercontent.com/BU-ISCIII/galaxy\_virologist\_training/one\_week\_4day\_format/exercises/data/L\_KX056050.fasta". A large blue box surrounds this URL block. At the bottom, there are buttons for "Type (set all):" (Auto-detect), "Genome (set all):" (Additional...), "Elegir archivos locales", "Choose remote files", "Paste/Fetch data" (highlighted with a blue box and labeled "1"), "Start" (highlighted with a red box and labeled "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    Flujos 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

Workflow: Create Crimea Congo Reference Genome

S\_fragment.fasta

M\_fragment.fasta

L\_fragment.fasta

Run Workflow

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 X

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 Logout 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 X
- Upload Data

**Get Data**

**Send Data**

**Collection Operations**

**GENERAL TEXT TOOLS**

**Text Manipulation**

**Filter and Sort**

**Join, Subtract and Group**

**Workflow: Create Crimea Congo Reference Genome**

1 Run Workflow 2

S\_fragment.fasta  
1: S\_KX056052.fasta

M\_fragment.fasta  
2: M\_KX056051.fasta

L\_fragment.fasta  
3: L\_KX056050.fasta

History

buscar conjuntos de datos ? X

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 Logout 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 X
- Upload Data

**Get Data**

**Send Data**

**Collection Operations**

**GENERAL TEXT TOOLS**

**Text Manipulation**

**Filter and Sort**

**Join, Subtract and Group**

**Invocation 1...**

Successfully invoked workflow **Create Crimea 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... X

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 ? X

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.

The screenshot shows the Galaxy Europe interface. On the left, there's a sidebar with various tool categories like Tools, Get Data, Send Data, Collection Operations, and GENERAL TEXT TOOLS. In the main area, a message says "Successfully invoked workflow Create Crimea Congo Reference Genome." Below it, a comic strip from PHD Comics is shown. To the right, the History pane lists several completed tasks, each with a green checkmark and a preview icon. The tasks include Fasta Statistics on data, Concatenate datasets, and Fasta summary stats for different datasets like L\_KX056050.fasta, M\_KX056051.fasta, and S\_KX056052.fasta.

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

The screenshot shows a PDF report titled "Workflow Execution Summary of Create Crimea Congo Reference Genome". The report includes sections for "Workflow Inputs" and "Workflow Outputs". Under "Workflow Inputs", it shows "Input Dataset: L\_fragment.fasta" with its sequence. The sequence starts with: >KX056050.1 Crimean-Congo hemorrhagic fever nairovirus isolate Ast199 segment L, complete genome. The sequence continues with several lines of DNA sequence. The report is signed off with "Identified 41e1c20a62fd6291". The history pane on the right is identical to the one in the first screenshot, showing completed tasks for the workflow.

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

## Note:

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