

Metagenomic sequencing - foodborne pathogens

Identify and track foodborne pathogens using long-read metagenomic sequencing

AIM

The aim is to identify and monitor foodborne pathogens using long-read metagenomic sequencing. This will enhance food safety surveillance and improve outbreak management by offering detailed insights into the microbial composition, virulence factors, and antibiotic resistance of pathogens

Outline:

1. Prepare Galaxy and data
2. Preprocessing
 1. Quality Control and Preprocessing
 2. Host read filtering
3. Taxonomy Profiling
4. Gene-based pathogen identification
 1. Assembly
 2. Antimicrobial Resistance Genes
 3. Virulence Factor identification
5. Allele-based pathogen identification
 1. Variant Calling or SNP Calling
 2. Mapping Depth and Coverage
 3. Consensus Genome Building
6. Pathogen Detection Samples Aggregation and Visualisation
 1. Heatmap
 2. Phylogenetic Tree building

Metagenomic sequencing - foodborne pathogens

7. Conclusion

OBJECTIVES

1. **Pathogen Detection:** To efficiently detect and identify a wide range of foodborne pathogens from contaminated food samples, including bacteria, viruses, fungi, and parasites.
2. **Genomic Profiling:** To fully characterize the genomes of foodborne pathogens, focusing on virulence factors, antimicrobial resistance genes, and other genomic elements influencing pathogenicity.
3. **Tracking Evolution:** To track the genetic evolution of foodborne pathogens over time, identifying mutations or gene transfers that could alter virulence or resistance.
4. **Source Tracking and Epidemiology:** To trace the origins of foodborne outbreaks by matching pathogen strains in food products with clinical or environmental isolates using high-resolution genomic data.
5. **Antibiotic Resistance Analysis:** To identify antibiotic resistance genes in pathogens to facilitate more precise interventions and management strategies during outbreaks.

Metagenomic sequencing - foodborne pathogens

Prepare Galaxy and data:

Data upload:

- Create a new history for this analysis
- Rename the history

The screenshot shows the India Galaxy workbench interface. The left sidebar includes links for Home, Upload, Tools, Workflows, Workflow Invocations, Visualization (which is selected), Histories, History (disabled), Notifications, and Settings. The main content area displays a "Saved Visualizations" section with options like Trackster, Audio player, Bar Horizontal (NVD3), Bar Horizontal Stacked (NVD3), and Bar Stacked (NVD3). A central banner reads "This is the Indian galaxy server, Welcome Home!". Below it, text about the workbench's purpose and training resources is displayed. The right side features a "History" panel titled "Pathogen detection from (direct Nanopore) sequencing data". It shows a message indicating the history is empty and provides options to load data or get data from an external source.

Import datasets:

- Import the following samples via link from [Zenodo](#) or Galaxy shared data libraries:
- Rename the files to `Barcode10` and `Barcode11` respectively
- Create a collection named `Samples` that includes both datasets (`Barcode10` and `Barcode11`)

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India web interface. On the left, there's a sidebar with icons for Upload, Tools, Workflows, Workflow Invocations, Visualization (which is selected), Histories, Notifications, and Settings. The main area has a title "Upload from Disk or Web". Below it, tabs for Regular, Composite, Collection, and Rule-based are shown. A large text input field contains URLs for two files: "https://zenodo.org/record/11222469/files/Barcode10_Spike2.fastq.gz" and "https://zenodo.org/record/11222469/files/Barcode11_Spike2b.fastq.gz". At the bottom of the dialog are buttons for "Type (set all)", "Reference (set all)", and file selection options like "Choose local file", "Choose remote files", and "Paste/Fetch data". To the right, a history panel shows datasets: "Barcode11_Spike2b.fastq" and "Barcode10_Spike2.fastq". The top right corner indicates "Using 16%".

The screenshot shows the Galaxy India "Welcome Home!" page. The left sidebar is identical to the previous screenshot. The main content area features a large heading "This is the Indian galaxy server, Welcome Home!". Below it, a paragraph introduces the server: "Namaste All! Welcome to the India Galaxy workbench, a comprehensive set of tools and workflows dedicated to accelerate your bioinformatics analyses focusing Bharat. This workbench is built on the Galaxy framework thanks to Galaxy Europe we are hosting the seeds of Indian Galaxy to facilitate more wider user base." It also mentions events and training webinars at www.galaxyproject.in. Two links are provided: "1. Training" and "2. Appreciation". Under "Training", it says: "For knowledge and about practical implementation about Galaxy framework there is a wide variety of collection of tutorials available. We strongly recommend to use the resources provided by [Galaxy Training Network \(GTN\)](#). Also we endorse to contribute and develop training materials of data analysis based on Galaxy ([Batut et al., 2017](#)).". Under "Appreciation", it says: "We sincerely appreciate the efforts of all individuals and their organizations ([Bioclues](#), [TMS Foundation](#)) for supporting the Indian Galaxy Instance." On the right, a history panel shows two datasets: "Barcode11_Spike2b.fastq.gz" and "Barcode10_Spike2.fastq.gz". The top right corner indicates "Using 16%".

Metagenomic sequencing - foodborne pathogens

Pathogen detection from
(direct Nanopore)
sequencing data

The screenshot shows a software interface for sequencing data management. At the top, it displays "617 MB" of available space. Below this, there are two main entries in a list:

- 2: Barcode11_Spike2b.fastq.gz** (with a green background)
- 1: Barcode10_Spike2.fastq.gz** (with a blue background)

Each entry includes a small icon, a pencil icon for editing, and a trash can icon for deletion. A large blue button labeled "#Barcode11Spike2b" is positioned below the first entry, and another blue button labeled "#Barcode10Spike2" is below the second.

The screenshot shows the India Galaxy workbench interface. On the left, there is a sidebar with various navigation options: Upload, Tools, Workflows, Workflow Invocations, Visualizations (which is currently selected), Histories, History (dropdown), Notifications, and Settings. The URL at the bottom is <https://indiausegalaxy.eu/>.

The main content area features a heading "This is the Indian galaxy server, Welcome Home!" followed by a brief welcome message and links for training and appreciation. Below this, there is a section titled "Training" with a paragraph about the Galaxy framework and a link to the Galaxy Training Network (GTN). There is also a "Appreciation" section with a paragraph of thanks to supporting organizations.

On the right side, there is a "History" panel and a context menu for two selected datasets:

- 2: Barcode11_Spike2b.fastq.gz** (selected)
- 1: Barcode10_Spike2.fastq.gz**
- #Barcode11Spike2b**

The context menu options include:

- With 2 selected...
- Hide
- Delete
- Delete (permanently)
- Build Dataset List** (highlighted)
- Build Dataset Pair
- Build List of Dataset Pairs
- Build Collection from Rules
- Change Database/Build
- Change data type
- Add tags
- Remove tags

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy web interface with the title "Galaxy India". The left sidebar includes options like Upload, Tools, Workflows, Workflow Invocations, Visualization (which is selected), Histories, Notifications, and Settings. The main area is titled "Create a collection from a list of datasets". It lists two datasets: "Barcode11_Spike2b.fasta.gz" and "Barcode10_Spike2.fasta.gz", each with a "Discard" button. A "Name:" field contains "Samples" with a checked "Hide original elements?" checkbox. A "Create collection" button is at the bottom right. A note on the right side provides information about the Galaxy framework and training resources. Below this note, a section titled "Appreciation" expresses gratitude to Bioclues and TMS Foundation.

Pathogen detection from
(direct Nanopore)
sequencing data

This screenshot shows a collection named "Samples" in the Galaxy interface. The collection contains 2 datasets and has a size of 617 MB. It has 1 location, 4 versions, and a refresh icon. The collection is highlighted in green. A "Settings" gear icon is visible above the collection list. The collection details show "5: Samples" and "a list with 2 datasets".

Metagenomic sequencing - foodborne pathogens

Preprocessing:

- Quality Control:** Tools like FastQC, NanoPlot, and MultiQC are used to assess the quality of raw sequencing reads both before and after preprocessing.
- Trimming and Filtering:** Porechop is employed to remove adapters, and Fastp filters out low-quality and short reads.
- Host Read Removal:** Minimap2 aligns and removes host genome sequences (such as Gallus gallus), and Kraken2, along with the Kalamari database, helps classify and eliminate other host sequences. Krakentools is then used to extract non-host reads for further analysis.

Pre-Processing:

- Import the workflow into Galaxy
- Run Workflow 1: Nanopore Preprocessing workflow using the following parameters

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

- Left Sidebar:** Includes icons for Upload, Tools, Workflows, Workflow Invocations, Visualization (selected), Histories, History (dropdown), Notifications, and Settings.
- Top Bar:** Galaxy India, Home, Workflow, Visualize, Data, Help, User, Logout, Help, and a search bar.
- Central Area:**
 - Import workflow:** A form with fields for "Archived file or url", "GA4GH servers", and "TRS ID". Below it are sections for "Import from a Galaxy workflow export URL or a workflow file" (with a "Archived Workflow URL" input) and "Archived Workflow File" (with a "nanopore_preprocessing.ga" input and a "Browse" button). A "Import workflow" button is at the bottom.
 - Right Panel:** A "History" panel showing two datasets:
 - Barcode11_Spike2b.fastq.gz (#Barcode11Spike2b)
 - Barcode10_Spike2.fastq.gz (#Barcode10Spike2)

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India web interface. On the left, a sidebar navigation includes: Upload, Tools, Workflows, Workflow Invocations, **Visualization** (selected), Histories, Notifications, and Settings. The main content area has tabs for Home, Workflow, Visualize, Data, Help, User, and a bell icon. The "Workflows" tab is active.

Workflows: A search bar at the top allows filtering by name or update time. It shows two workflows: "Nanopore Preprocessing (imported from uploaded file)" (edited less than a minute ago) and "Find exons with the highest number of features" (edited about 1 month ago). Both workflows have tags: #Collection, #microGalaxy, #PathoGFAIR, #Nanopore, and #IWC.

Visualizations: A search bar for visualizations is present. The "Saved Visualizations" section lists five items: Trackster, Audio player, Bar Horizontal (NVD3), Bar Horizontal Stacked (NVD3), and Bar Stacked (NVD3).

History: A sidebar on the right shows a history of datasets. Two datasets are listed under "Pathogen detection from (direct Nanopore) sequencing data": "Barcode11_Spike2b.fastq.gz" and "Barcode10_Spike2.fastq.gz".

This screenshot shows the "Nanopore Preprocessing (imported from uploaded file)" workflow form. The "samples_profile" input field is optional and currently empty. The "collection_of_all_samples" input field is required and has a dropdown menu set to "accepted formats". A message states: "No data datasets available". Below the form, a link says "Expand to full workflow form".

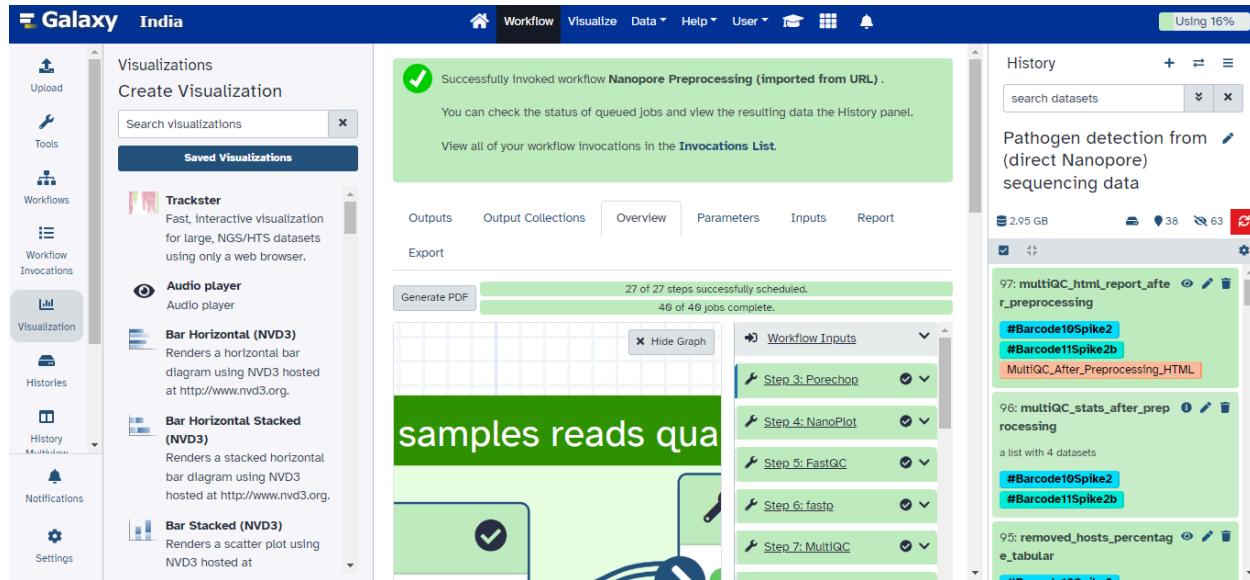
The left sidebar and top navigation are identical to the previous screenshot. The main content area shows the workflow form with its inputs and validation messages.

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India interface. On the left, the navigation bar includes options like Upload, Tools, Workflows, Workflow Invocations, Visualization (which is selected), Histories, Notifications, and Settings. The main area displays the "Workflow: Nanopore Preprocessing (imported from URL) (version: 1)". This workflow has two inputs: "samples_profile" (optional, set to "PacBio/Oxford Nanopore read to reference mapping (-Hk19) (map-pb)") and "collection_of_all_samples" (set to "5: Samples"). The accepted formats are listed as fastq or fastq.gz. A note states: "based on the lab preparation of the samples during sequencing, there should be a sample profile better than the other, to be chosen as an optional input to Minimap2. e.g. PacBio/Oxford Nanopore For more details check: <https://github.com/lh3/minimap2?tab=readme-ov-file#use-cases>". The right panel shows the History section with a dataset named "5: Samples".

This screenshot shows the same Galaxy India interface after the workflow has been invoked. The main message in the center says "Successfully Invoked workflow Nanopore Preprocessing (imported from URL). You can check the status of queued jobs and view the resulting data in the History panel. View all of your workflow invocations in the Invocations List." Below this, the "Outputs" tab is selected, showing a graph titled "Waiting to complete Invocation 1.". The graph indicates 18 of 27 steps successfully scheduled and 3 of 29 jobs complete. The workflow inputs are shown as Step 3: Porechop, Step 4: NanoPlot, and Step 5: FastQC. The right panel shows the History section with multiple datasets: "73: collection_of_preprocessed_samples" (a list with 2 datasets), "72: quality_retained_all_reads" (Barcode10Spike2 and Barcode11Spike2b), and "67: kraken2_with_kalamri_database_output" (a list with 2 datasets).

Metagenomic sequencing - foodborne pathogens



Quality Control and Preprocessing:

- **FastQC:** Generates reports to assess read quality.
- **NanoPlot:** Visualizes long-read sequencing data.
- **Porechop:** Trims adapter sequences.
- **Fastp:** Filters low-quality and short reads.
- **MultiQC:** Aggregates and compares quality reports from various stages

Metagenomic sequencing - foodborne pathogens

Inspect the HTML two outputs of MultiQC for `Barcode10` before and after preprocessing tagged `MultiQC_Before_Preprocessing` and `MultiQC_After_Preprocessing`

MultiQC Report Before PreProcessing

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2024-10-17, 12:03 based on data in:

/data/jwd02f/main/074/562/74562384/working/multiqc_WDir

Welcome! Not sure where to start? [Watch a tutorial video](#)

(6:06)

History + ≡ ⌂

search datasets

Pathogen detection from (direct Nanopore) sequencing data

2.95 GB 38 63

32: bam_map_to_host

a list with 2 datasets

31: multiQC_html_report_before_processing

#Barcode10Spike2
#Barcode11Spike2b
MultiQC_Before_Preprocessing.HTML

30: multiQC_stats_before_processing

a list with 3 datasets

#Barcode10Spike2

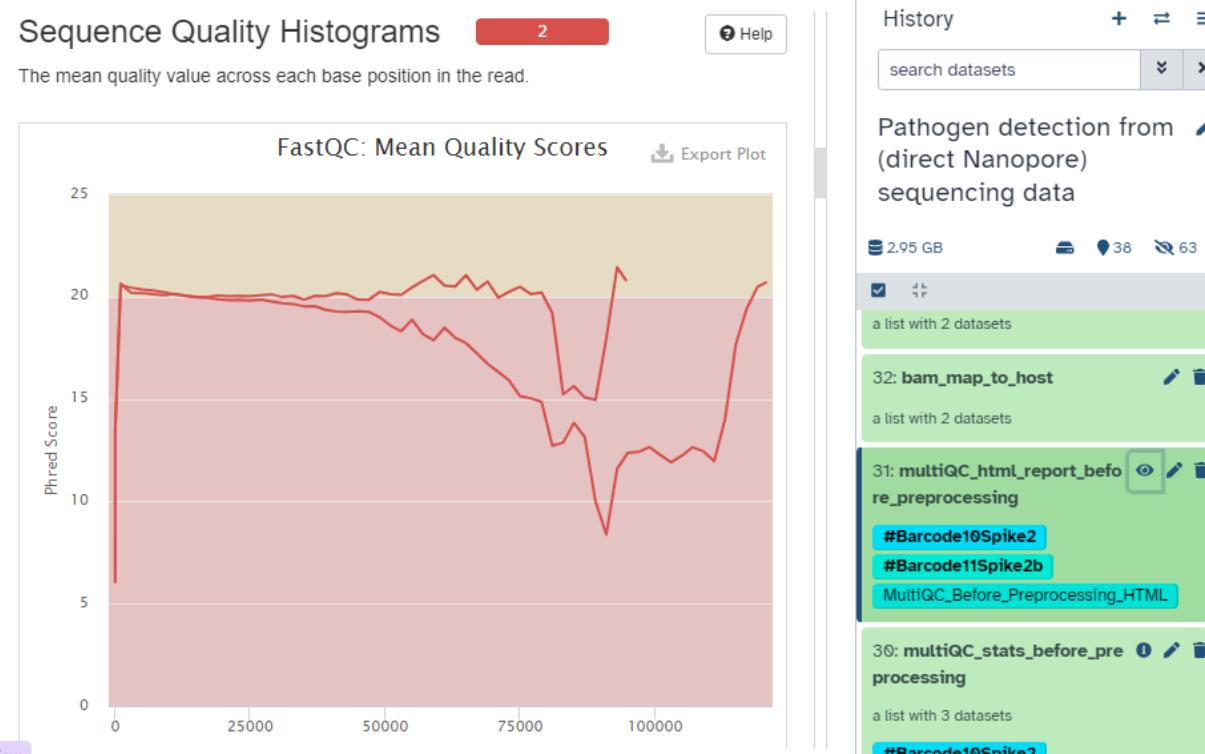
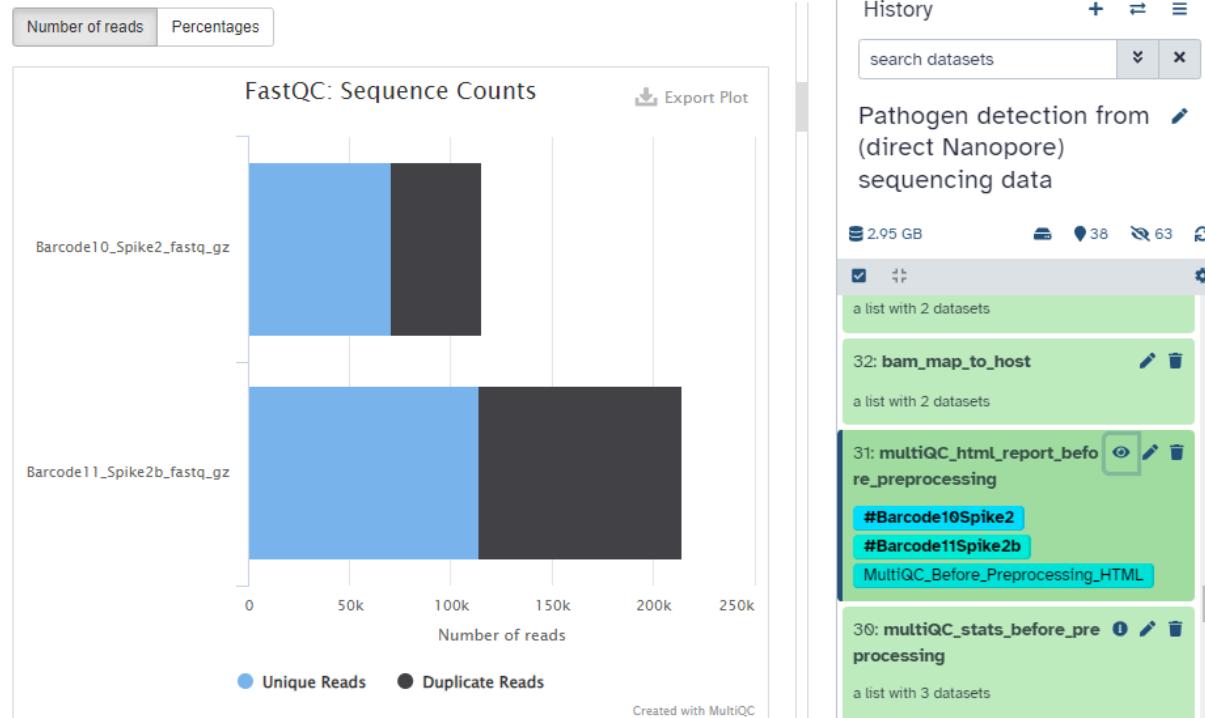
General Statistics

Sample Name	% Dups	% GC	Length	M Seqs
Barcode10_Spike2_fastq_gz	38.4%	47%	1 402 bp	0.1
Barcode11_Spike2b_fastq_gz	46.8%	51%	3 516 bp	0.2

FastQC

preview

Metagenomic sequencing - foodborne pathogens



Metagenomic sequencing - foodborne pathogens

FastQC: Per Sequence Quality Scores

Count

Mean Sequence Quality (Phred Score)

Created with MultiQC

search datasets ✖

Pathogen detection from (direct Nanopore) sequencing data

2.95 GB 38 63

a list with 2 datasets

32: bam_map_to_host

a list with 2 datasets

31: multiQC_html_report_befo
re_preprocessing

#Barcode10Spike2

#Barcode11Spike2b

MultiQC_Before_Preprocessing_HTML

30: multiQC_stats_before_pre
processing

a list with 3 datasets

#Barcode10Spike2

MultiQC

MultiQC Report After PreProcessing

MultiQC

MultiQC Report After PreProcessing

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2024-10-17, 12:54 based on data in:
/data/jwd02f/main/074/563/74563277/working/multiqc_wDir

Welcome! Not sure where to start? [Watch a tutorial video](#) (6:06)

don't show again ✖

General Statistics

preview

History + ☰

search datasets ✖

Pathogen detection from (direct Nanopore) sequencing data

2.95 GB 38 63

97: multiQC_html_report_aft
er_preprocessing

#Barcode10Spike2

#Barcode11Spike2b

MultiQC_After_Preprocessing_HTML

96: multiQC_stats_after_prep
rocessing

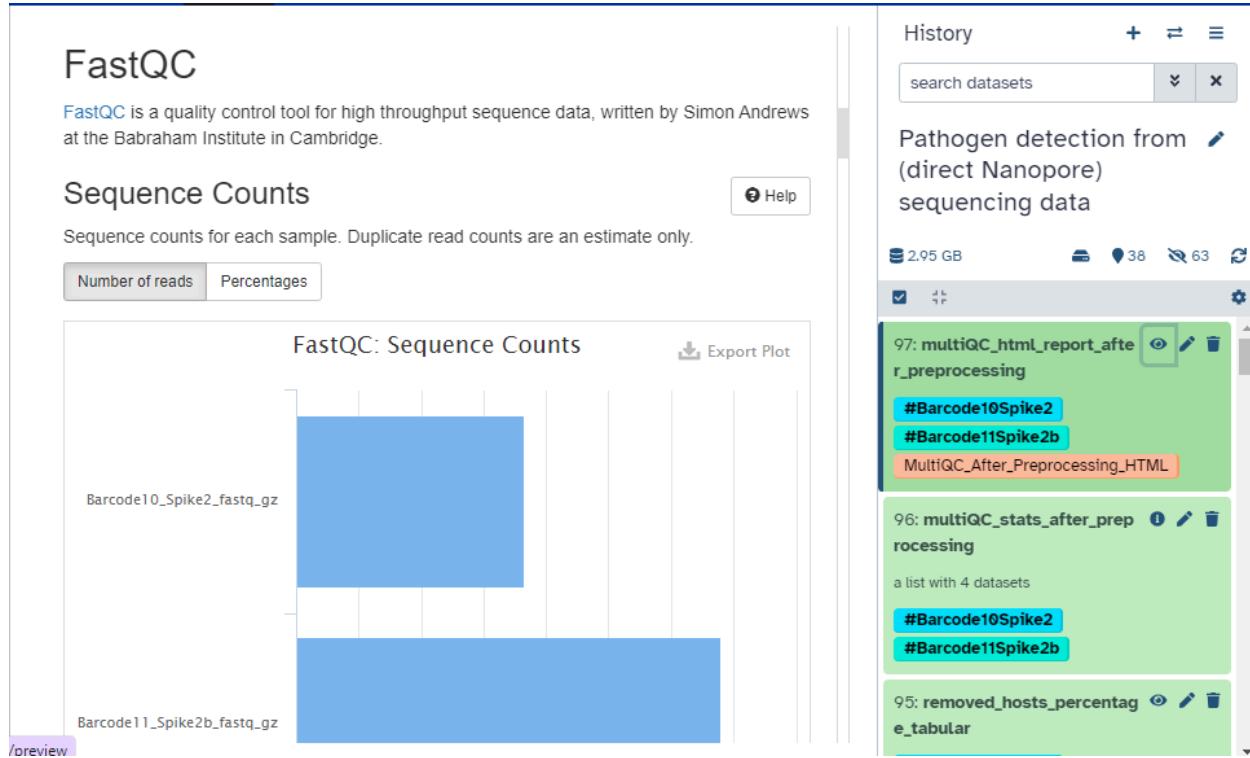
a list with 4 datasets

#Barcode10Spike2

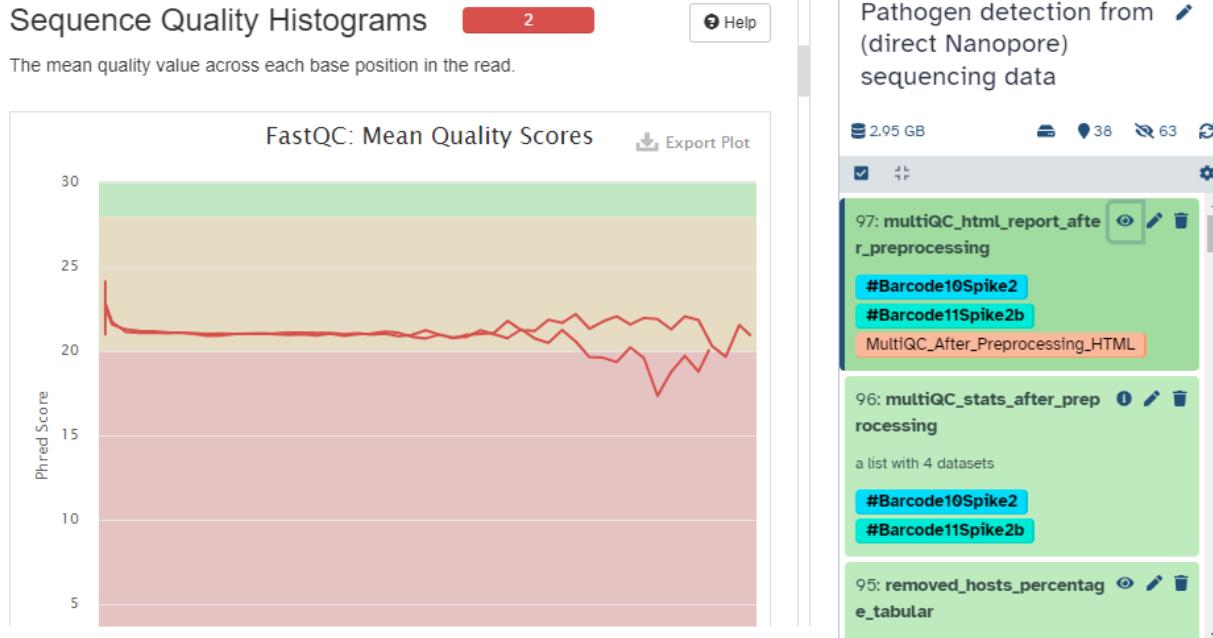
#Barcode11Spike2b

95: removed_hosts_percentag
e_tabular

Metagenomic sequencing - foodborne pathogens



Metagenomic sequencing - foodborne pathogens



Host read filtering:

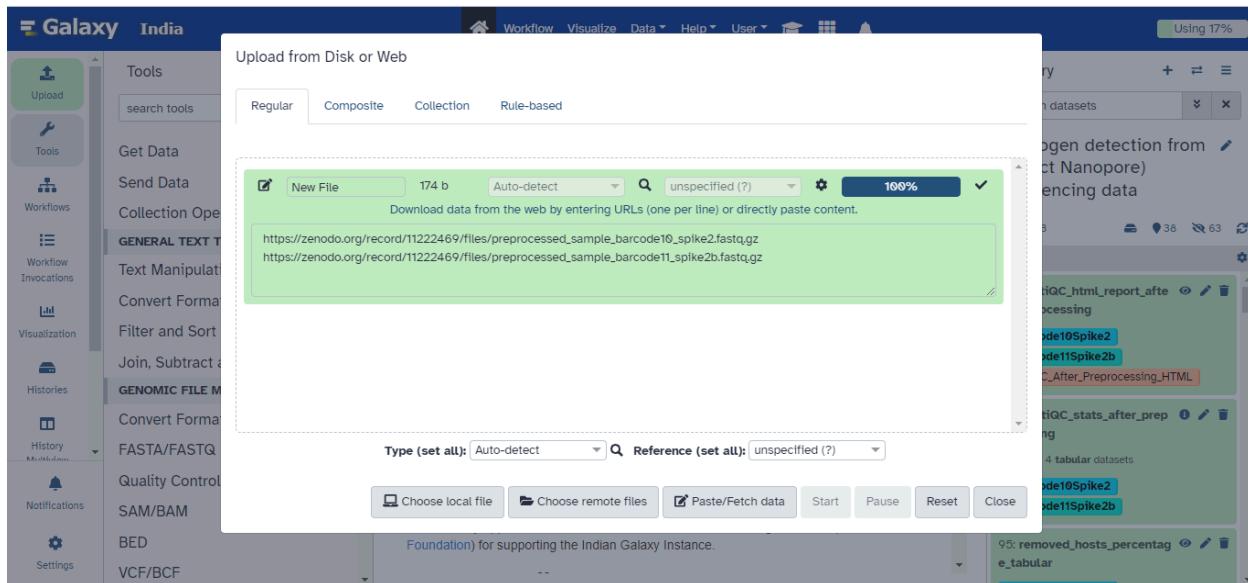
- **Primary Host Removal:** Aligns to the chicken genome (*Gallus gallus*) using Minimap2, retaining only non-host reads.
- **Further Contamination Check:** Utilizes Kraken2 and the Kalamari database to remove remaining host or contaminant sequences, such as mitochondrial reads from other food hosts.

Metagenomic sequencing - foodborne pathogens

Samtools fastx on the **mapped** (output of Split BAM by reads mapping status tool), then inspect the output for **Barcode10**. If you are a short version taker then inspect the output named **host_sequences_fastq**

Data upload:

- Import the quality processed samples fastqsanger files via link from [Zenodo](#) or the Shared Data library:
- Rename datasets to **Barcode10** and **Barcode11** respectively
- Create a collection named **collection** of preprocessed samples from the two imported datasets



Metagenomic sequencing - foodborne pathogens

Galaxy India

Workflow Visualize Data Help User Home Notifications

Using 17%

Upload Tools Workflows Workflow Invocations Visualization Histories History Notifications Settings

Tools search tools

Get Data Send Data Collection Operations

GENERAL TEXT TOOLS

- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group

GENOMIC FILE MANIPULATION

- Convert Formats
- FASTA/FASTQ
- Quality Control
- SAM/BAM
- BED
- VCF/BCF

Edit Dataset Attributes

Name: barcode10

Info: uploaded fastq.gz file

Annotation: optional

Database/Build: optional

unspecified (?)

Save Auto-detect

History

Pathogen detection from (direct Nanopore) sequencing data

3.23 GB 40 63

103: barcode11
102: barcode10
Add Tags Display

53.6 MB format fastq.gz, database ?
uploaded fastq.gz file

Save Auto-detect

https://india.usegalaxy.eu/datasets/4838ba20a6d8676519e40156ed4b5f13/review

Galaxy India

Workflow Visualize Data Help User Home Notifications

Using 17%

Upload Tools Workflows Workflow Invocations Visualization Histories History Notifications Settings

Tools search tools

Get Data Send Data Collection Operations

GENERAL TEXT TOOLS

- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group

GENOMIC FILE MANIPULATION

- Convert Formats
- FASTA/FASTQ
- Quality Control
- SAM/BAM
- BED
- VCF/BCF

Edit Dataset Attributes

Attributes updated.

Name: barcode11

Info: uploaded fastq.gz file

Annotation: optional

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

Database/Build: optional

unspecified (?)

Save Auto-detect

History

Pathogen detection from (direct Nanopore) sequencing data

3.23 GB 40 63

103: barcode11
102: barcode10
Add Tags Display

53.6 MB format fastq.gz, database ?
uploaded fastq.gz file

Save Auto-detect

Metagenomic sequencing - foodborne pathogens

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy web interface for creating a dataset collection. On the left, a sidebar lists various tools and data types: Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, Notifications, and Settings. The main area is titled "Create a collection from a list of datasets". It displays two items: "barcode11" and "barcode10", each with a "Discard" button. A checkbox "Hide original elements?" is checked. Below this, a text input field is set to "collection of preprocessed samples". To the right, there's a "Create collection" button. A modal window titled "Convert Formats" is open, listing options: FASTA/FASTQ, Quality Control, SAM/BAM, BED, and VCF/BCF. Under "Database/Build" (optional), it says "unspecified (?)". Buttons for "Save" and "Auto-detect" are present. On the far right, a preview pane shows a FASTQ.gz file with sequence data:

```
@S1984528-cbea-4574-a2bd-4d1f16b424ed runid=a  
=barcode10 parent_read_id=S1984520-cbea-4574-  
TATATCATTAAACCGCTCAGCCGCTCGGCCAGITCAGGCCGGTA/  
+  
,0(66$33&+9::8877'{{{{;96$3338).4;-03423%`%
```

Metagenomic sequencing - foodborne pathogens

History + ⌂ ⌄

search datasets

Pathogen detection from (direct Nanopore) 🖊

sequencing data

History Size

3.23 GB 39 67

106: collection of preprocessed samples 🖊 ⏷

a list with 2 datasets

97: multiQC_html_report_after_preprocessing 🌐 🖊 ⏷

#Barcode10Spike2
#Barcode11Spike2b
MultiQC_After_Preprocessing_HTML

96: multiQC_stats_after_prep 🌐 🖊 ⏷

rocessing

a list with 4 tabular datasets

#Barcode10Spike2
#Barcode11Spike2b

Metagenomic sequencing - foodborne pathogens

Taxonomy Profiling:

- **Taxonomic Assignment:** Kraken2 is used to assign taxonomy from the kingdom to species levels, identifying potential pathogens and mixed infections.
- **Visualization:** The taxonomic composition is visualized to detect possible pathogens.

Taxonomy Profiling and visualization:

- Import the workflow into Galaxy
- Run Workflow 2: Taxonomy Profiling and Visualization with Krona

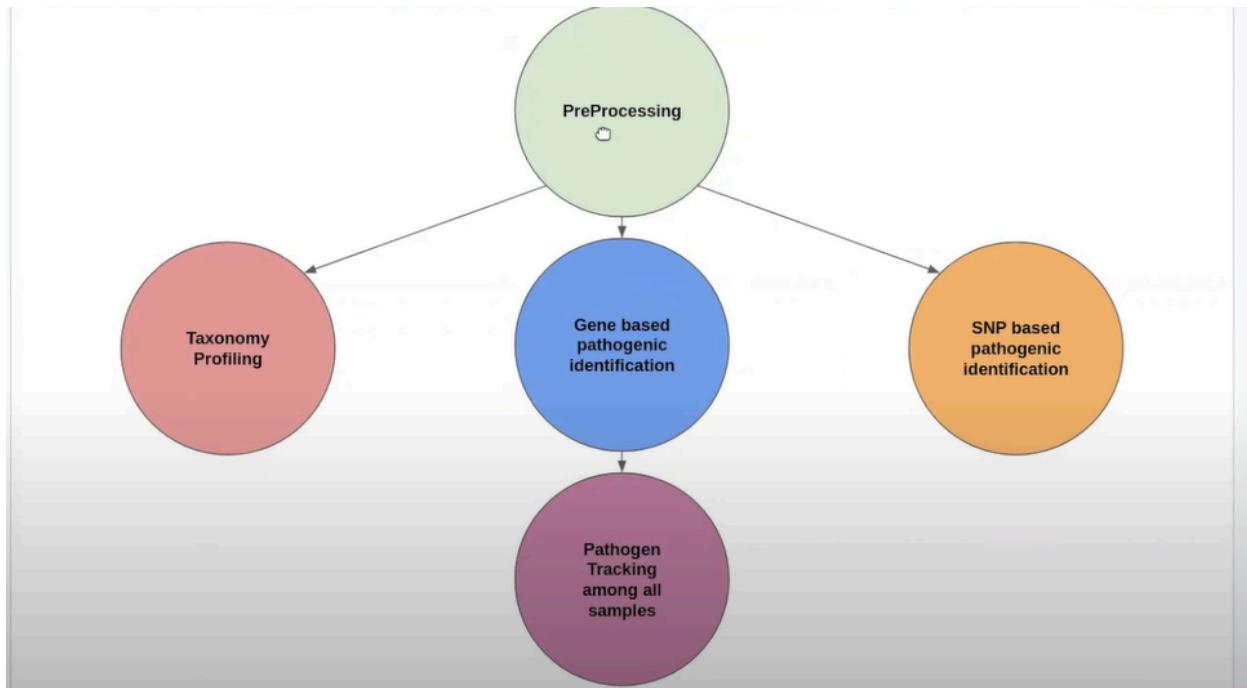
The screenshot shows the Galaxy web interface with a blue header bar. The left sidebar contains links for Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, Notifications, and Settings. The main content area has a title 'Import workflow'. It includes two input fields: 'Archived file or url' and 'GA4GH servers', and a 'TRS ID' field. Below these is a section for 'Import from a Galaxy workflow export URL or a workflow file'. It has a 'Archived Workflow URL' input field and a note: 'If the workflow is accessible via a URL, enter the URL above and click Import.' There is also a 'Archived Workflow File' input field with a 'Browse' button and a note: 'If the workflow is in a file on your computer, choose it and then click Import.' At the bottom is a 'Import workflow' button. To the right is a 'History' panel showing three datasets: '97: multiQC_html_report_after_r_processing', '96: multiQC_stats_after_processing', and '95: removed_hosts_percentage.tabular'. The top right corner of the interface shows 'Using 17%'.

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India web interface. On the left, the navigation bar includes 'Upload', 'Tools' (selected), 'Workflows', 'Workflow Invocations', 'Visualization', 'Histories', 'History', 'Notifications', and 'Settings'. The main area is titled 'Import workflow' and contains fields for 'Archived file or url' (with a placeholder 'GA4GH servers' and 'TRS ID'), 'Archived Workflow URL' (empty), and 'Archived Workflow File' (containing 'taxonomy_profiling_and_visualisation_with_krona.ga'). A 'Browse' button is available for local files. Below these are buttons for 'Import workflow' and 'Import workflow from File'. To the right, the 'History' panel shows three entries: '97: multiQC_html_report_after_preprocessing' (containing '#Barcode10Spike2', '#Barcode11Spike2b', and 'MultiQC_After_Preprocessing_HTML'), '96: multiQC_stats_after_processing' (containing '#Barcode10Spike2' and '#Barcode11Spike2b'), and '95: removed_hosts_percentage_table'.

The screenshot shows the Galaxy India web interface. The left sidebar is identical to the previous screenshot. The main area is titled 'Workflows' and displays two workflows: 'Taxonomy Profiling and Visualization with Krona (imported from uploaded file)' (edited less than a minute ago) and 'Nanopore Preprocessing (imported from URL)' (edited about 9 hours ago). Both workflows have tags: '#Collection', '#microGalaxy', '#PathoGFAIR', '#IWC', '#Nanopore', and '#IWC'. Below the workflows, a success message states: 'Workflow "Taxonomy Profiling and Visualization with Krona (Imported from uploaded file)" imported successfully.' The history panel on the right is also visible.

Metagenomic sequencing - foodborne pathogens



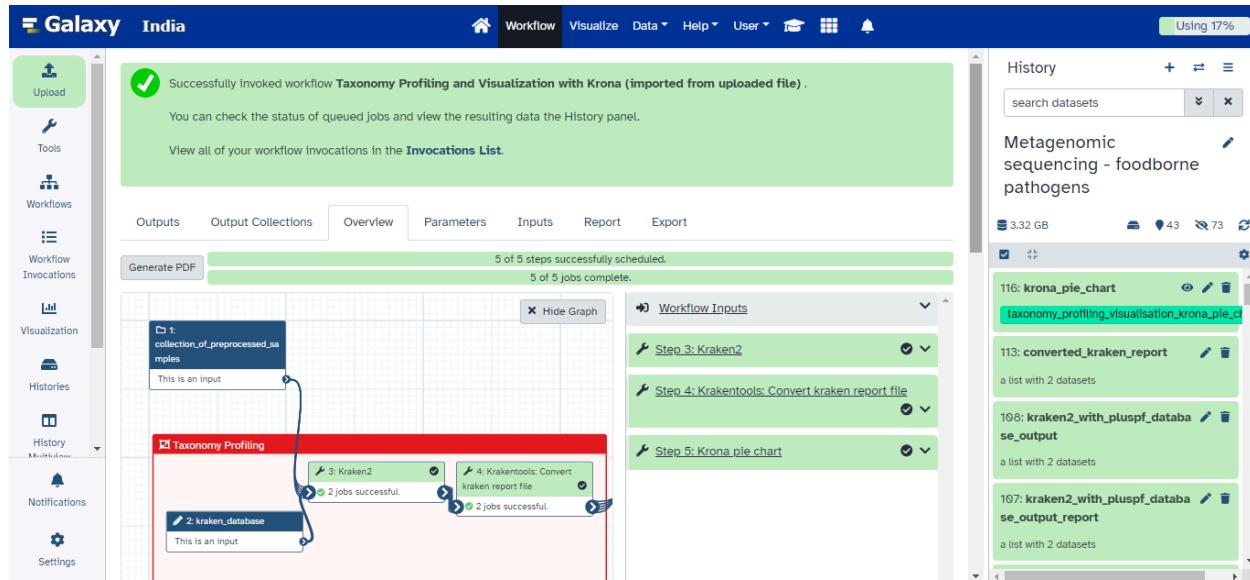
The screenshot shows the Galaxy platform interface for 'India'. The top navigation bar includes 'Workflow', 'Visualize', 'Data', 'Help', 'User', and a search bar. The left sidebar contains links for 'Upload', 'Tools', 'Workflows', 'Invocations', 'Visualization', 'Histories', 'Notifications', and 'Settings'. The main workspace displays a workflow titled 'Workflow: Taxonomy Profiling and Visualization with Krona (imported from uploaded file) (version: 1)'. This workflow consists of several steps:

- collection_of_preprocessed_samples**: A collection of preprocessed samples (106 items) accepted formats: FASTQ.
- kraken_database**: Prebuilt Refseq Indexes: PlusPF (Standard plus protozoa and fungi) (Version: 2022-06-07 - Downloaded: 2022-09-04T165121Z).
- kraken_database_for_user_to_choose**: A dropdown menu for selecting a kraken database.

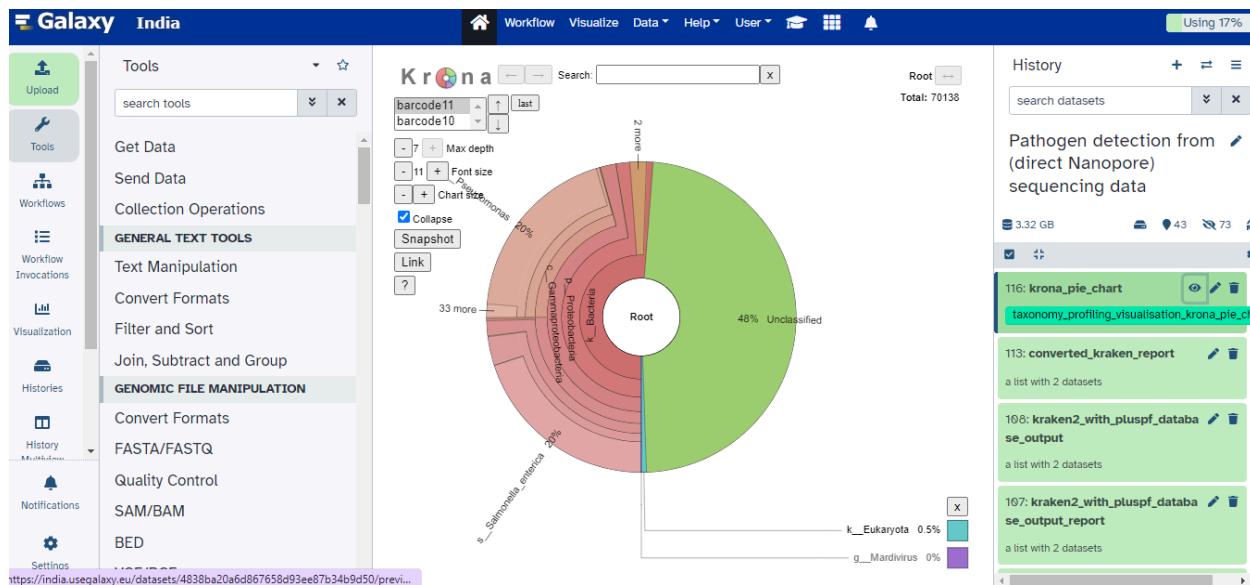
The right side of the screen shows the 'History' panel, which lists three datasets:

- 106: collection of preprocessed samples (3.23 GB)
- 97: multiQC_html_report_after_r_preprocessing (#Barcode10Spike2, #Barcode11Spike2b, MultiQC_After_Preprocessing_HTML)
- 96: multiQC_stats_after_preprocessing (#Barcode10Spike2, #Barcode11Spike2b)

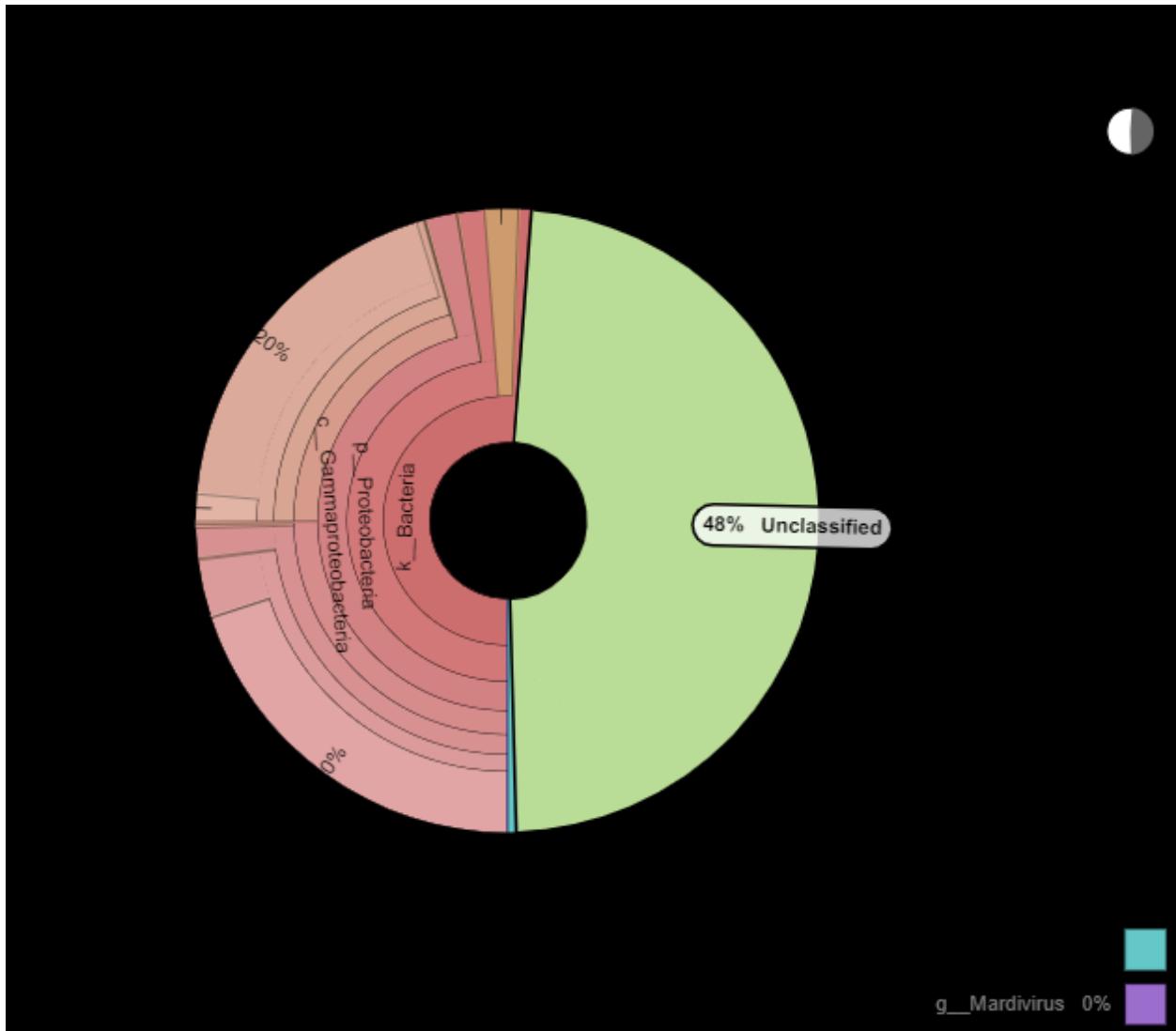
Metagenomic sequencing - foodborne pathogens



Inspect the Kraken2 report for Barcode10



Metagenomic sequencing - foodborne pathogens



Gene-based pathogen identification:

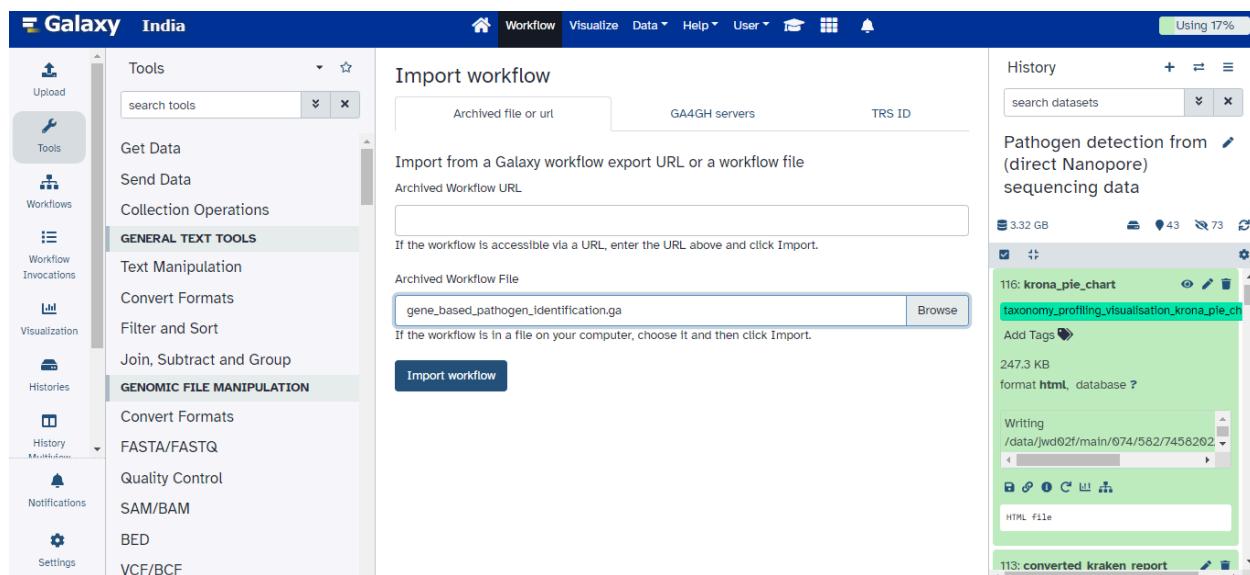
- **Genome Assembly:** Uses metaflye for contig assembly and Medaka for polishing. The assembly graph is visualized with Bandage.

Metagenomic sequencing - foodborne pathogens

- **Pathogenicity Gene Detection:** ABRicate is used to identify virulence factor (VF) and antimicrobial resistance (AMR) genes.
- **Output:** Produces FASTA and tabular files detailing identified genes for pathogen profiling and strain identification.

Gene based Pathogenic Identification:

- Import the workflow into Galaxy
- Run Workflow 3: Gene-based Pathogen Identification



Metagenomic sequencing - foodborne pathogens

This screenshot shows the Galaxy India interface. The left sidebar contains navigation links for Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, Notifications, and Settings. The main area displays a "Workflows" section with a search bar and filtering options. It lists several workflows:

- Gene-based Pathogen Identification (imported from uploaded file)
- Taxonomy Profiling and Visualization n with Krona (imported from uploaded file)
- Nanopore Preprocessing (imported from URL)
- Find exons with the highest number of features

The right sidebar shows a "History" panel with a list of datasets:

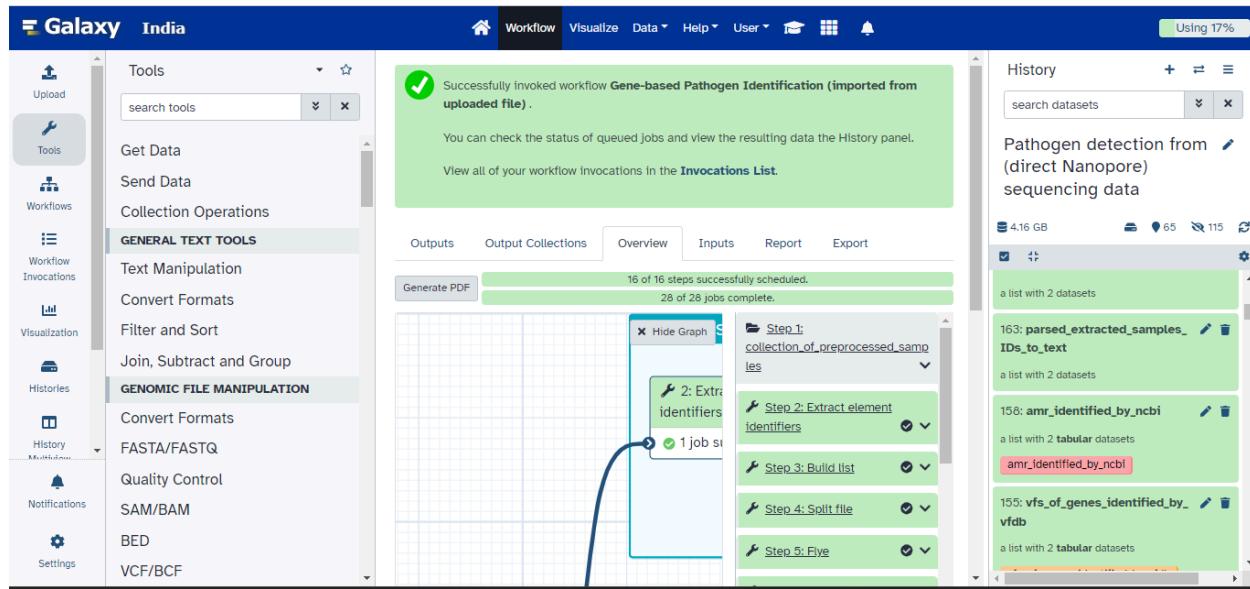
- 116: krona_pie_chart
- 247.3 KB format html, database ?
- Writing /data/jwd@2f/main/074/562/7458202
- HTML file
- 113: converted kraken report

This screenshot shows the Galaxy India interface with a different focus. The left sidebar is identical to the first screenshot. The main area displays a "Workflow: Gene-based Pathogen Identification (imported from uploaded file) (version: 1)" step. It shows a collection of preprocessed samples (106: collection of preprocessed samples) and a "Run Workflow" button.

The right sidebar shows a "History" panel with a list of datasets:

- se_output_report
- 106: collection of preprocessed samples
- 97: multiQC_html_report_after_preprocessing
- #Barcode10Spike2
- #Barcode11Spike2b
- MultiQC_After_Preprocessing_HTML
- 96: multiQC_stats_after_processing

Metagenomic sequencing - foodborne pathogens



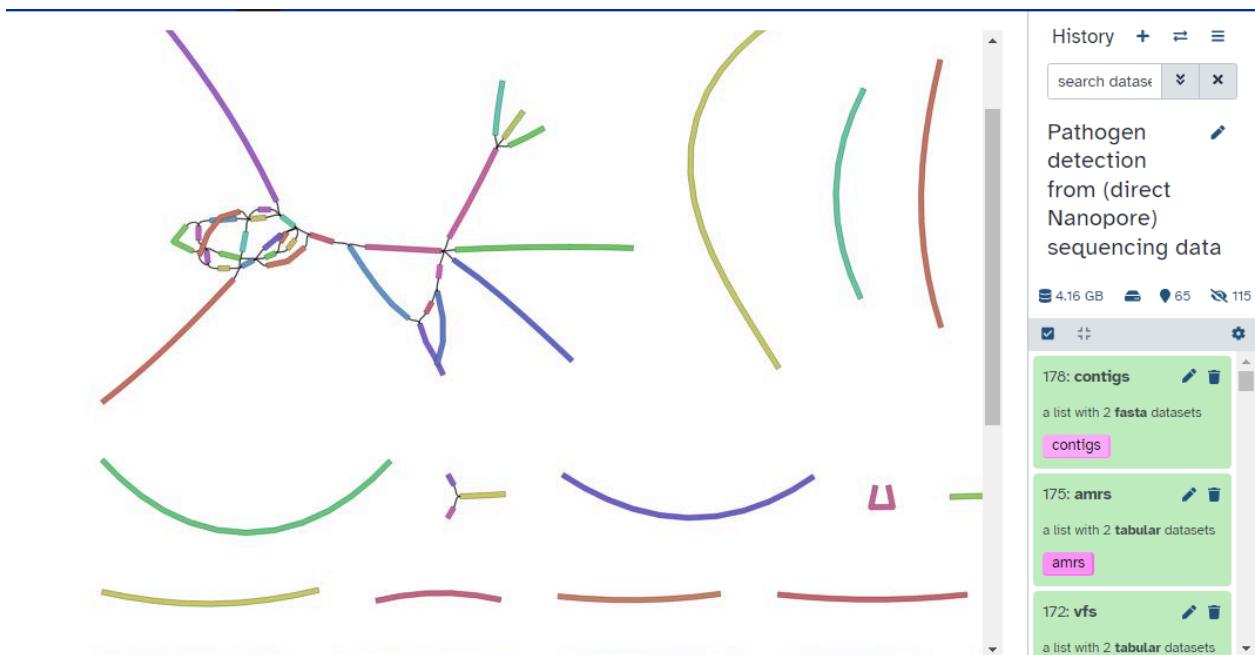
Assembly:

- **Assembly:** Metaflye or Flye creates longer contigs from reads.
- **Visualization:** Assembly graphs are visualized with Bandage.
- **Polishing:** Medaka improves accuracy by correcting assembly errors.

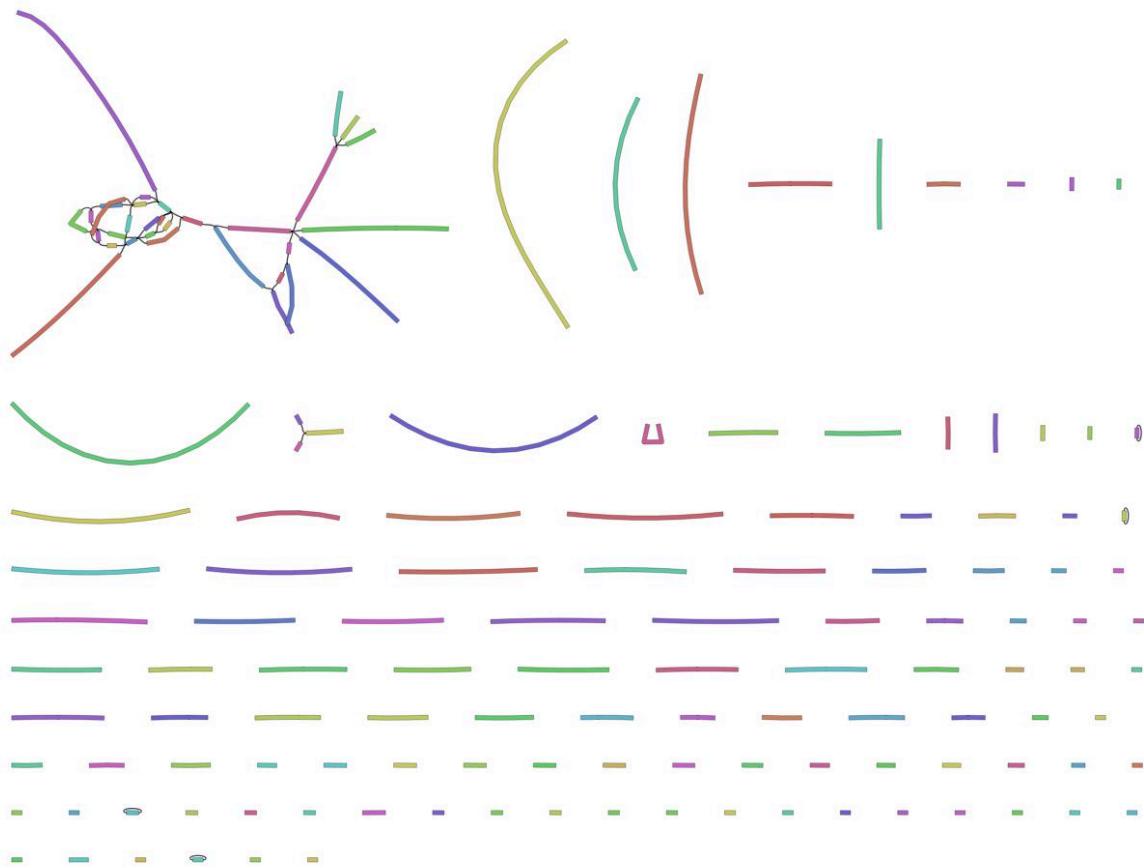
Inspect Flye and Medaka consensus pipeline output results for Barcode10

Metagenomic sequencing - foodborne pathogens

Bandage Image



Metagenomic sequencing - foodborne pathogens



Metagenomic sequencing - foodborne pathogens

Medaka consensus pipeline

This dataset is large and only the first megabyte is shown below.
[Show all](#) | [Save](#)

```
>contig_56
ACGGCGCTTCTGAAGGCCAGCTGAGCTGAAAATGGTTATCAAAGTTCAAGTCAGTCGGCACTTATGGCACGCCATGCAGACCAACATCCACCTGGCTATTGCGGGGGTGAC
>contig_147
ATCCGGCTTTACTGTTGCGCTGTAAACATTGTTTCCGCTGTTCATCCCCACCGTCTGGACGCCATGGCATTCTTAATACTCAACATAATCACGCTCAGAAGGAA
>contig_91
GCGCAGCGATATAAGCGCATGGCGGCAAATGGCAAACGGTAAAGCTTCCACAGTCAGGATTGTATCACATCAGCCATGGTAGCCTGACTCCAGCCCATCAGCCCCCAGCGT
>contig_37
TGGTAGGCTGCGCTGTGCTACCCACGTTAGACCGAAAACATTAAAGGTAGTCGCCCATCGACAATTATGTCGCCAGCCAGTTGCTCGCCGCCAATACGCTGCGCATG
>contig_129
CTGCAAGTATGCAACAAACTGCATTGCAAGATTGTCAGCTTCTTAATCGGGAAAGTGTACGATGCGTCTTACATGCGCTATTAGCGCTTGTACCCCTCCGAGCGTAT
>contig_103
CTTTTTCTTCTTCTTCAAAAAAAAAAAAAAAACAAAAAAACAAAAAAACAAAAAAACAAAAAAAGGGCTTTCTCTCTCTCTAAACGGGACCCCCCCCCCCCC
>contig_130
ACGGTTTATTGACCAATTGTTGTTATGTCGGCTAGCTGCCCTGTGATAAACGACGCCGGCCGGCGATTATCGGGCAGCGAAGGGAGACTATGACCATCACGCCGCTCCC
>contig_63
CAGTACCGTGCCTGCTGGCAAATCCAGCGCTGCCGAAAGGTGAAGGCTGATCTGGAAACGCTGCCGCGACTGCTGAAATGGCAGGGGAGCGTCCGCTGATCC
>contig_23
GACGTTCTCGAAGTCGGCCCTAGCGC AAAATGGCTCCGACAAACTAAACACCAGTGATGACGGCTAAAGGTACGATTACGGAGCTGCCAGAAAGGCAAATACCGC
>contig_43
TGAGCGCGCACACTGCGC AAAACTGCGCACGCTGGTATGTTGAGTAAGTAATGGGATCCACGCCAGCAGGAAGTCACGGCTCGCAGCTTATCTGCAAATCGTA
>contig_111
CGCGCCTGCTCGATAATGTTAACGTGACGACGCCAGCGCTGCCAGCTGCGCTCGCAGCTTATGCGGATTTTCGACACCGCTCGCAGTAACTGCGAT
>contig_13
CCACACTACCATCGCGCTACGGCGTTCACTCTGAGTCGGCATGGGTGAGGTGGACCCACCGCGTATTGCGCCAGGCAAATCTGTTCTTCAACCGCCACTCTGTA
>contig_119
TAATGGCGATAACAGTGAAGATGCGCGTTACGGCGGCGTCTGAGATTCAAATTTCTCTGATTGTTGGTACTAAAAAGGGATCCGAGATCCCTTTATCAGATAGTTA
3684c4880aea62f1/preview
```

Checking program versions
This is medaka 1.7.2

Program	Version	Required	Pass
bcftools	1.16	1.11	True
bgzip	1.16	1.11	True
minimap2	2.24	2.11	True
samtools	1.16.1	1.11	True
tabix	1.16	1.11	True

Aligning basecalls to draft

```
Creating fai index file /data/jwd05e/main/074/582/74582327/working/input_assembly.fa.fai
Creating mmi index file /data/jwd05e/main/074/582/74582327/working/input_assembly.fa.map-ont.mmi
[M::mm_idx_gen::0.503*1.00] collected minimizers
[M::mm_idx_gen::0.614*1.31] sorted minimizers
[M::main::0.702*1.27] loaded/built the index for 124 target sequence(s)
[M::mm_idx_stat] kmer size: 15; skip: 10; is_hpc: 0; #seq: 124
[M::mm_idx_stat::0.725*1.26] distinct minimizers: 1128205 (94.57% are singletons); average
occurrences: 1.065; average spacing: 5.349; total length: 6428917
[M::main] Version: 2.24-r1122
[M::main] CMD: minimap2 -I 166 -x map-ont -d
[data/jwd05e/main/074/582/74582327/working/input_assembly.fa.map-ont.mmi
[data/jwd05e/main/074/582/74582327/working/input_assembly.fa
[M::main] Real time: 0.790 sec; CPU: 0.943 sec; Peak RSS: 0.062 GB
[M::main::0.104*1.01] loaded/built the index for 124 target sequence(s)
[M::mm_mapopt_update::0.131*1.01] mid_occ = 10
[M::mm_idx_stat] kmer size: 15; skip: 10; is_hpc: 0; #seq: 124
[M::mm_idx_stat::0.148*1.00] distinct minimizers: 1128205 (94.57% are singletons); average
occurrences: 1.065; average spacing: 5.349; total length: 6428917
[M::worker_pipeline::7.235*7.42] mapped 37679 sequences
[M::main] Version: 2.24-r1122
[M::main] CMD: minimap2 -x map-ont --secondary=no -L --MD -A 2 -B 4 -O 4,24 -E 2,1 -t 12 -a
[data/jwd05e/main/074/582/74582327/working/input_assembly.fa.map-ont.mmi
[data/dnb10/galaxy_db/files/e/8/c/dataset_e8c2c578-26aa-4d31-9014-c27f8c210601.dat
[M::main] Real time: 7.561 sec; CPU: 53.980 sec; Peak RSS: 1.393 GB
[bam_sort_core] merging from 0 files and 12 in-memory blocks...
Running medaka consensus
```

Metagenomic sequencing - foodborne pathogens

Flye Consensus Fasta output results for Barcode10

This dataset is large and only the first megabyte is shown below.
[Show all](#) | [Save](#)

<< History: Pathogen detection from (direct N...
flye_consensus_fasta
a list with 2 **fasta** datasets
Download
1: barcode11
2: barcode10

```
>contig_10
GAAGGCAAACGACATCGAAATCGCATGATCAGCGATCCACAGCTTATAATA
TCAGGCAAACGACATCGAAATCGCATGATCAGCGATCCACAGCTTATAATA
CGCGCCAGCAGATTGCGACCAACGAGACAGCGCTAAAAATGTGCGCCCCAGTGA
ATCGACGCAAACCGTAGTGCAGAAAGGGCGTTGTCAGGGCTGTAGTCAAATA
GTTTCGCGCTCGAGGGCTCAGCGAGGCGAAAGATCGCACAGAACACGGCAG
TAAGGAATCCCCGGTGACATCGGACATGCTGGCGGAATACTGTCGCCAGAACGTC
ATGCCACGGCTAGTGTACGGTCAAGGTTGACAGCGGCTGGCTGTAACACGTC
TGACGGCGCAAGTTAAATACTGCTACTTGTACTCTGATAGCGGGGTTAATGGGG
CGTAGAGAACATCCACACTAAAGTTGAATAAGATAATACTCGAGGAAACAGCAGCAC
TCCGCGCTGAATCACTGGTAATCGCGCATGTCGACGGAGTCGAGGAGTAAGCGCCAAGC
CCCGGCGAGTTGAACACTTTAACGAGCATGGAGCCCGAGAACAGCGAACACTG
CAAGTCCCACATGGTCACTGACCGGATCATCGGTTGGCTAGGGCTGTTGAGAACACA
ACCCAGGTTGCTCACGCTTGGCTCGCGTACGGCATGTAATCTCGTTAACAGCA
TCGACAAACGACGCCGGGTAAGCGGCCATCGCGCGCACTCGCGCGCCCGAGCGTC
AGGGAGGGTAAATGTAGTGTCCAGCTGCGCCCCACGGTGGCGCCAGCGGAGT
TCCAGGAAATACATCGGATTAAACAGCTCCCCAGCGCAAATCGGGAAACGAGAT
CGCCGGTACCGCAATCGCTATTCTCACCGGCTCGCAGGGTACCGCACAGCGC
GGCGCGATGCCGACCGCATGCCAAACAGCACCGCCCGAGACATGCACTGGTTATGG
TCAGCCACAGCGCTGGCATAAAGCGCTGGCGATCTCTGGCAACCGGGCGAGCGGAC
ACCATCGACGGGGCAAATCGCGCTCACAGCATGGGTGATATGTCGAGAACCTGGCTG
CATGAAGCGCGTATCGAACGCCCCAGCTGAGCAGGACAGTTCTGACCAACTCGCGCTC
AGCTCGGGCGGCTATCGCGGGGGCTGCCGGCAACAAATGAAACAAATAAACAC
ACCGACACTGCCACAAATCACAGCTAGGAATTATCCAGCAGGGCTTGAGAACATAG
TTAAGCATCACAGAACCTTACCGGATGGCGAACAGCGCCATCCAGGCCATCG
CATTACTTAAATCGCATGTCAGGCTGAAACCGGTATCCGGCATAATCCAGAA
d43ad7bd2c54796/preview.
```

Antimicrobial Resistance Genes

- **ABRicate:** Searches contigs for AMR genes using the NCBI AMRFinderPlus database.
- **Output:** Provides information on contig names, AMR gene locations, and the substances they confer resistance to.
- **Tabular Manipulation:** Replacer tool adds sample IDs and contig locations for further analysis.

Metagenomic sequencing - foodborne pathogens

AMR genes found in Barcode10

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11
#FILE	SEQUENCE	START	END	STRAND	GENE	COVERAGE	COVERAGE_MAP	GAPS	%COVERAGE	%ID
barcode10	contig_112	1635	2810	+	tet(C)	1-1176/1176	=====	0/0	100.00	
barcode10	contig_113	8618	9458	-	sul1	1-840/840	=====	1/1	100.00	
barcode10	contig_113	10255	11082	-	blaOXA-2	1-828/828	=====	0/0	100.00	
barcode10	contig_128	18244	18665	+	tet(34)	1-421/465	=====	7/7	89.89	
barcode10	contig_137	172662	172964	+	aac(6')_Yersi	122- 428/432	...=====	6/8	69.68	
barcode10	contig_148	8618	9458	-	sul1	1-840/840	=====	1/1	100.00	
barcode10	contig_148	10255	11082	-	blaOXA-2	1-828/828	=====	0/0	100.00	

ib22d7de2ca23d06/preview

« History: Pathogen detectio...

amr_identifie d_by_ncbi edit

a list with 2 tabular datasets

amr_identified_by_ncbi

Download

1: barcode11 eye
amr_identified_by_ncbi

2: barcode10 eye
amr_identified_by_ncbi

AMR genes found in Barcode11

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11
#FILE	SEQUENCE	START	END	STRAND	GENE	COVERAGE	COVERAGE_MAP	GAPS	%COVERAGE	%IDE
barcode11	contig_4	2599788	2600068	-	aac(6')_Yersi	128- 431/447	...=====	6/11	66.44	
barcode11	contig_4	3855413	3855833	-	tet(34)	1-421/465	=====	4/4	90.11	

« History: Pathogen detectio...

amr_identifie d_by_ncbi edit

a list with 2 tabular datasets

amr_identified_by_ncbi

Download

1: barcode11 eye
amr_identified_by_ncbi

2: barcode10 eye
amr_identified_by_ncbi

Metagenomic sequencing - foodborne pathogens

Virulence Factor identification:

- **Objective:** To identify pathogenic bacteria in samples by detecting virulence factor (VF) genes.
- **Gene Identification:** Confirms the presence of pathogenic bacteria by detecting VF genes in assembled sequences.

Inspect VFs of genes Identified by VFDB output file from [Barcode10](#) and [Barcode11](#)

VFs of genes Identified by VFDB - Barcode10

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10
#FILE	SEQUENCE	START	END	STRAND	GENE	COVERAGE	COVERAGE_MAP	GAPS	%CC
barcode10	contig_10	55140	57273	-	slrP	1-	===== / =====	81/168	
						2298/2298			
barcode10	contig_100	2143	2424	+	fliD	9-290/360	===== / =====	2/2	
barcode10	contig_100	2493	3050	+	fliC	6-564/582	===== / =====	7/9	
barcode10	contig_100	3204	4044	+	motA	13-	===== / =====	10/11	
						850/888			

History: Pathogen detection from (direct Na...
vfs_of_genes_identified_by_vfdb
a list with 2 tabular datasets
vfs_of_genes_identified_by_vfdb

Download
1: barcode11
vfs_of_genes_identified_by_vfdb
2: barcode10
vfs_of_genes_identified_by_vfdb

Metagenomic sequencing - foodborne pathogens

VFs of genes Identified by VFDB - Barcode11								
Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9
#FILE	SEQUENCE	START	END	STRAND	GENE	COVERAGE	COVERAGE_MAP	GAPS
barcode11	contig_1	179191	181236	-	fpaA	401- 2448/2448	.===== =====	23/40
barcode11	contig_1	212180	213203	-	pscU	1-1024/1050	===== =====	8/12
barcode11	contig_1	213224	213988	-	pscT	1-765/789	===== =====	0/0
barcode11	contig_1	213985	214250	-	pscS	2-267/267	===== =====	2/4
barcode11	contig_1	214261	214907	-	pscR	1-647/654	===== =====	0/0

History: Pathogen detection from (direct Na...
vfs_of_genes_identified_by_vfdb
a list with 2 tabular datasets
vfs_of_genes_identified_by_vfdb
Download
1: barcode11
vfs_of_genes_identified_by_vfdb
2: barcode10
vfs_of_genes_identified_by_vfdb

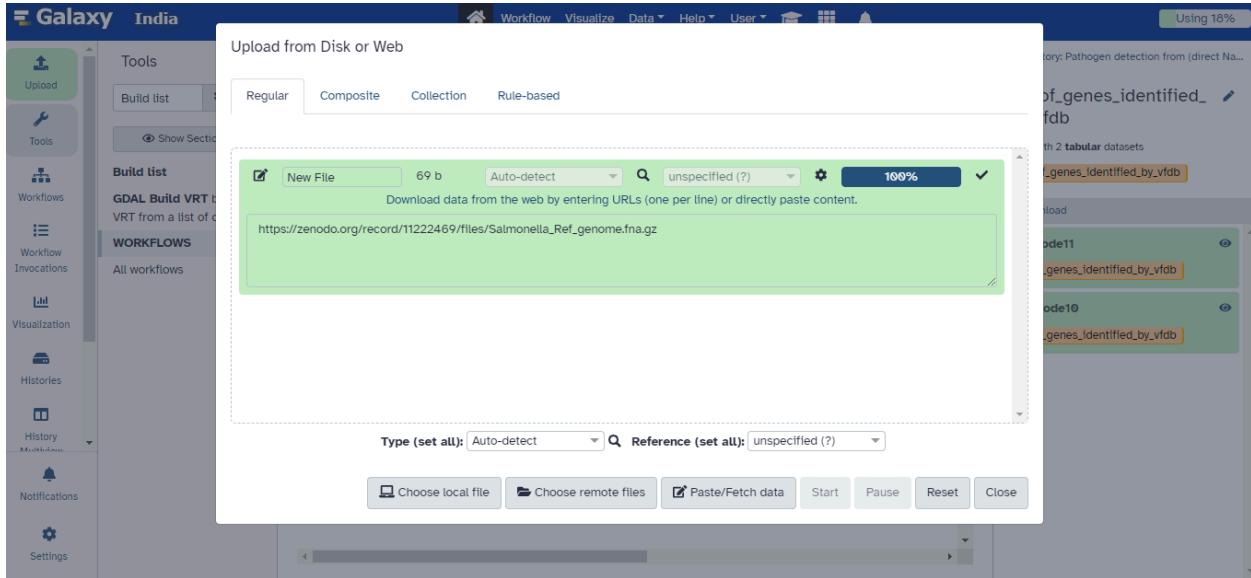
Allele-based pathogen identification:

- **Objective:** To identify pathogens by performing variant and SNP calling against a reference genome.
- **Mapping:** Samples are mapped to the *Salmonella enterica* reference genome to identify pathogenic variants.
- **Variant Analysis:** Identifies specific variants that indicate pathogenicity and reveals novel alleles.
- **Consensus Genome:** Builds consensus genomes for phylogenetic analysis of evolutionary relationships.

Data upload:

- Import a reference genome FASTA file via link from [Zenodo](#) or Galaxy shared data libraries

Metagenomic sequencing - foodborne pathogens



Metagenomic sequencing - foodborne pathogens

Pathogen detection
from (direct Nanopore)
sequencing data

4.16 GB 66 115

The screenshot shows a Galaxy workflow interface. At the top, there is a title "Pathogen detection from (direct Nanopore) sequencing data". Below the title, there are three status indicators: a blue folder icon for "4.16 GB", a location pin icon for "66", and a magnifying glass icon for "115". A gear icon is also present. The main area displays four green rectangular boxes, each representing a dataset:

- 181: Salmonella_Ref_genom e.fna.gz**: A green box containing a checkmark icon, a plus icon, and a minus icon. It includes edit and delete icons. Below the box, it says "a list with 1 **fasta** datasets" and has a pink button labeled "contigs".
- 178: contigs**: A green box containing a checkmark icon, a plus icon, and a minus icon. It includes edit and delete icons. Below the box, it says "a list with 2 **fasta** datasets" and has a pink button labeled "contigs".
- 175: amrs**: A green box containing a checkmark icon, a plus icon, and a minus icon. It includes edit and delete icons. Below the box, it says "a list with 2 **tabular** datasets" and has a pink button labeled "amrs".
- 172: vfs**: A green box containing a checkmark icon, a plus icon, and a minus icon. It includes edit and delete icons. Below the box, it says "a list with 2 **tabular** datasets".

Allele based Pathogenic Identification:

- Import the workflow into Galaxy
- Run Workflow 4: Nanopore Allele-based Pathogen Identification

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India interface. The left sidebar includes 'Upload', 'Tools', 'Workflows', 'Workflow Invocations', 'Visualization', 'Histories', 'History', 'Notifications', and 'Settings'. The main area is titled 'Import workflow' with sections for 'Archived file or url' (GA4GH servers, TRS ID), 'Import from a Galaxy workflow export URL or a workflow file' (Archived Workflow URL), and 'Archived Workflow File' (selected 'allele_based_pathogen_identification (1).ga'). A 'Browse' button is available for local files. A 'Import workflow' button is at the bottom. The right sidebar shows a 'History' panel with datasets: 181: Salmonella_Ref_genome.fna.gz, 178: contigs (a list with 2 fasta datasets), 175: amrs (a list with 2 tabular datasets), and 172: vfs (a list with 2 tabular datasets). The top right corner shows 'Using 18%'.

The screenshot shows the Galaxy India interface. The left sidebar includes 'Upload', 'Tools', 'Workflows', 'Workflow Invocations', 'Visualization', 'Histories', 'History', 'Notifications', and 'Settings'. The main area is titled 'Workflows' with tabs for 'My workflows', 'Workflows shared with me', and 'Public workflows'. It features a search bar and filtering options ('Sort by: Name, Update time', 'Filter: Show deleted, Show bookmarked', 'Display: #, #'). Two workflows are listed: 'Allele-based Pathogen Identification (imported from uploaded file)' (edited less than a minute ago) and 'Gene-based Pathogen Identification (imported from uploaded file)' (edited about 15 hours ago). Both workflows have tags: '#Collection', '#microGalaxy', '#PathoGFAIR', '#IWC'. A 'Run workflow' button is present. The right sidebar shows a 'History' panel with datasets: 181: Salmonella_Ref_genome.fna.gz, 178: contigs (a list with 2 fasta datasets), 175: amrs (a list with 2 tabular datasets), and 172: vfs (a list with 2 tabular datasets). The top right corner shows 'Using 18%'.

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India web interface. The left sidebar contains navigation links: Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History (selected), Notifications, and Settings. The main area displays a workflow titled "Workflow: Allele-based Pathogen Identification (imported from uploaded file) (version: 1)". The workflow has two inputs: "collection_of_preprocessed_samples" (selected dataset 106) and "reference_genome_of_tested_strain" (selected dataset 181). The output section shows four datasets: "181: Salmonella_Ref_genom.e.fna.gz" (contigs), "176: contigs" (a list with 2 fasta datasets), "175: amrs" (a list with 2 tabular datasets), and "172: vfs" (a list with 2 tabular datasets). The top right corner shows "Using 18%".

Metagenomic sequencing - foodborne pathogens

Pathogen detection from (direct Nanopore) sequencing data

4.67 GB 91 147

238: number_of_variants_per_sample

237: Collapse Collection on data 236 and data 235

234: Select first on collection 23

233: mapping_mean_depth_per_sample

Variant Calling or SNP Calling:

- **Data Upload:** Upload the *S. enterica* reference genome from NCBI for mapping.
- **Mapping:** Reads are mapped to the reference genome using Minimap2.
- **Variant Identification:** Variants and SNPs are called using Clair3.
- **Indel Normalization:** Indels are normalized with bcftools.

Metagenomic sequencing - foodborne pathogens

- **Filtering:** Variants are filtered using SnpSift Filter or LoFreq.
- **Output:** A VCF file is generated containing variant data, and SnpSift Extract Fields is used to create a tabular report.

Map reads to the reference genome

The screenshot shows a bioinformatics pipeline interface. On the left, a terminal window displays the command used for mapping reads to a reference genome:

```
@HD VN:1.6 SO:coordinate  
@SQ SN:NC_003197.2 LN:4857450  
@SQ SN:NC_003277.2 LN:93933  
@PG ID:minimap2 PN:minimap2 VN:2.24- CL:minimap2 --q-occ-frac  
r1122  
@PG ID:samtools PN:samtools PP:minimap2 VN:1.14 CL:samtools sort -@8 -T /data/jwd02f/main/074/601/74601810/  
1acf2714-f9c2-  
46a4-afce-  
0220ea6fdbb4
```

On the right, a sidebar titled "map_with_minimap2" lists two datasets:

- 1: barcode11
- 2: barcode10

Each dataset entry includes a download icon and an eye icon.

Metagenomic sequencing - foodborne pathogens

History: Pathogen detection from (direct Nanopore) sequencing data

map_with_minimap2

a list with 2 **bam** datasets

Download

1: barcode11

2: barcode10

QNAME	FLAG	RNAME	POS	MAPQ	CIGAR
@HD VN:1.6 SO:coordinate					
@SQ SN:NC_003197.2 LN:4857450					
@SQ SN:NC_003277.2 LN:93933					
@PG ID:minimap2 PN:minimap2 VN:2.24- CL:minimap2 --q-occ-frac r1122					
@PG ID:samtools PN:samtools PP:minimap2 VN:1.14 CL:samtools sort -@8 -T /data/jwd02f/main/074/601/74601811/					
10adac22-e1f5-459a-a634-df3483564dd8	2048	NC_003197.2	1	60	7325H21M1D114M1D6M2I51M1I69M1D38M
1d52d6e9-8cfb-4e7d-bb95-61d0a56d2ea	2048	NC_003197.2	1	60	10646H33M2D9M1I19M3D55M1D12M1D38M

History: Pathogen detection from (direct Nanopore) sequencing data

claims

Pathogen detection from (direct Nanopore) sequencing data

a list with 2 **vcf_bgzip** datasets

Download

1: barcode11

2: barcode10

Loading...

Chrom	Pos	ID	Ref	Alt	Qual	Filter	Info	Format	data
##fileformat=VCFv4.2									
##FILTER=<ID=PASS,Description="All filters passed">									
##FILTER=<ID=LowQual,Description="Low quality variant">									
##FILTER=<ID=RefCall,Description="Reference call">									
##INFO=<ID=P,Number=0>Type=Flag,Description="Result from pileup calling">									
##INFO=<ID=F,Number=0>Type=Flag,Description="Result from full-alignment calling">									
##FORMAT=<ID=GT,Number=1>Type=String,Description="Genotype">									
##FORMAT=<ID=GQ,Number=1>Type=Integer,Description="Genotype Quality">									
##FORMAT=<ID=DP,Number=1>Type=Integer,Description="Read Depth">									
##FORMAT=<ID=AD,Number=R>Type=Integer,Description="Read depth for each allele">									
##FORMAT=<ID=PL,Number=G>Type=Integer,Description="Phred-scaled genotype likelihoods rounded to the closest integer">									
##FORMAT=<ID=AF,Number=1>Type=Float,Description="Estimated allele frequency in the range of [0,1]">									
##contig=<ID=NC_003197.2,length=4857450>									
##contig=<ID=NC_003277.2,length=93933>									
#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE

preview

Metagenomic sequencing - foodborne pathogens

```

Chrom Pos ID Ref Alt Qual Filter Info Format
##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters passed">
##FILTER=<ID=LowQual,Description="Low quality variant">
##FILTER=<ID=RefCall,Description="Reference call">
##INFO=<ID=P,Number=0,Type=Flag,Description="Result from pileup calling">
##INFO=<ID=F,Number=0,Type=Flag,Description="Result from full-alignment calling">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=AD,Number=R,Type=Integer,Description="Read depth for each allele">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="Phred-scaled genotype likelihoods rounded to the closest integer">
##FORMAT=<ID=AF,Number=1,Type=Float,Description="Estimated allele frequency in the range of [0,1]">
##contig=<ID=NC_003197.2,length=4857450>
##contig=<ID=NC_003277.2,length=93933>
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT

```

« History: Pathogen detection from (direct Nan...

clair3_merged_output

a list with 2 **vcf_bgzip** datasets

Download

1: barcode11	
2: barcode10	


```

Chrom Pos ID Ref Alt Qual Filter Info Format data
##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters passed">
##FILTER=<ID=LowQual,Description="Low quality variant">
##FILTER=<ID=RefCall,Description="Reference call">
##INFO=<ID=P,Number=0,Type=Flag,Description="Result from pileup calling">
##INFO=<ID=F,Number=0,Type=Flag,Description="Result from full-alignment calling">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=AD,Number=R,Type=Integer,Description="Read depth for each allele">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="Phred-scaled genotype likelihoods rounded to the closest integer">
##FORMAT=<ID=AF,Number=1,Type=Float,Description="Estimated allele frequency in the range of [0,1]">
##contig=<ID=NC_003197.2,length=4857450>
##contig=<ID=NC_003277.2,length=93933>
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT SAMPLE

```

« History: Pathogen detection from (direct Nan...

clair3_pileup_vcf

a list with 2 **vcf_bgzip** datasets

Download

1: barcode11	
2: barcode10	

Metagenomic sequencing - foodborne pathogens

Mapping Depth and Coverage:

- **Coverage:** The percentage of the reference genome covered by reads.
- **Depth:** The number of times each base is sequenced, which increases variant detection confidence.
- **Quality Control:** Ensures that no variants are missed, reducing false positives/negatives.
- **Tools:** Samtools depth and Samtools coverage are used for coverage analysis.

Samtools coverage output for `Barcode10`

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9
#rname	startpos	endpos	numreads	covbases	coverage	meandepth	meanbaseq	meanmapq
NC_003197.2	1	4857450	14869	4840833	99.6579	6.63781	21.3	53.9
NC_003277.2	1	93933	539	93924	99.9904	8.88245	21.3	57

Samtools coverage on collection 183
a list with 2 tabular datasets

Download

1: barcode11

2: barcode10

Metagenomic sequencing - foodborne pathogens

Consensus Genome Building:

- **Purpose:** To build consensus genomes for sample comparison and analysis.
- **Applications:** Essential for outbreak tracking (e.g., SARS-CoV-2), though less critical for slower-evolving pathogens like Salmonella.
- **Tool:** Bcftools consensus is used to generate consensus genomes.

Inspect the bcftools consensus output for [Barcode11](#)

The screenshot shows a bioinformatics tool interface with the following details:

- Warning Message:** "This dataset is large and only the first megabyte is shown below. Show all | Save"
- Sequence Data:** A large block of DNA sequence data representing the NC_003197.2 Salmonella enterica subsp. enterica serovar Typhimurium str. LT2 genome. The sequence starts with: >NC_003197.2 Salmonella enterica subsp. enterica serovar Typhimurium str. LT2, complete genome AGAGATTACGTCCTGGTCAAGAGGCCATGACAGAGGGATTGGTCAAATAAATATAGG...
- Tool Header:** "bcftools_consensus" (highlighted in green)
- Dataset List:** "a list with 2 fasta datasets"
 - 1: barcode11 (highlighted in green)
 - 2: barcode10
- Download Button:** "Download" (disabled)
- History:** "History: Pathogen detection from (direct Nan...)"
- Preview Link:** "6f796931b9a45b0/preview"

Metagenomic sequencing - foodborne pathogens

This screenshot shows a bioinformatics interface for metagenomic sequencing. At the top, a yellow warning box states: "This dataset is large and only the first megabyte is shown below." Below this are two links: "Show all" and "Save". The main content area displays a sequence of DNA bases (A, T, C, G) representing a bacterial genome. The sequence starts with: >NC_003197.2 Salmonella enterica subsp. enterica serovar Typhimurium str. LT2, complete genome AGAGATTACGTGTTGAAAGAGATCATGACAGGGGAATGGTTGAAAATATAATAGG... The sequence continues for several kilobases. To the right of the sequence is a sidebar titled "bcftools_consensus". It shows a list of datasets: "1: barcode11" and "2: barcode10", both of which are highlighted with a green background. There are also "Download" and "History" buttons.

Pathogen Detection Samples Aggregation and Visualisation:

- The workflow integrates data on AMR genes, virulence factors (VF), and genetic variants for pathogen analysis.
- **Heatmap:** Visualizes VF gene distribution across samples, identifying potentially pathogenic samples.
- **Phylogenetic Tree:** Relates samples based on shared genes, tracking evolutionary relationships. Together, these visualizations support public health responses and outbreak investigations, providing a clearer understanding of pathogen dynamics.

Metagenomic sequencing - foodborne pathogens

All Samples Analysis:

- Import the workflow into Galaxy
- Run Workflow 5: Pathogen Detection Samples Aggregation and Visualisation

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

- Header:** Galaxy India, Workflow, Visualize, Data, Help, User, Notifications, Using 18%.
- Left Sidebar:** Tools, Upload, Tools (selected), Workflows, Workflow Invocations, Visualizations, Histories, Notifications, Settings.
- Middle Panel:**
 - Import workflow:** Archived file or url, GA4GH servers, TRS ID.
 - Archived Workflow URL:** Input field for URL.
 - Archived Workflow File:** Input field containing "pathogen_detection_pathoGAI...ga" with a "Browse" button.
 - Buttons:** Import workflow.
- Right Panel:**
 - bcftools_consensus:** a list with 2 fasta datasets.
 - Download:** 1: barcode11, 2: barcode10.

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India web interface. The left sidebar contains navigation links for Upload, Tools, Get Data, Send Data, Collection, Operations, GENERAL TEXT TOOLS, GENOMIC FILE MANIPULATION, and Notifications. The main area displays a list of workflows under the heading "Workflows". The first workflow listed is "Pathogen Detection PathoGFAIR Samples Aggregation and Visualisation (imported from uploaded file)", which was edited less than a minute ago. It has tags #Collection, #microGalaxy, #PathoGFAIR, and #IWC. The second workflow is "Allele-based Pathogen Identification (imported from uploaded file)", edited about 26 hours ago, with tags #Collection, #microGalaxy, #PathoGFAIR, and #IWC. The third workflow is "Gene-based Pathogen Identification (imported from uploaded file)", edited 1 day ago, with tags #Collection, #microGalaxy, #PathoGFAIR, and #IWC. The fourth workflow is "Taxonomy Profiling and Visualization with Krona (imported from uploaded file)", edited 1 day ago, with tags #Collection, #microGalaxy, #PathoGFAIR, and #IWC. A sidebar on the right shows a history entry for "bcftools_consensus" with two datasets: barcode11 and barcode10.

Workflow: Pathogen Detection PathoGFAIR Samples Aggregation and Visualisation (imported from uploaded file) (version: 1)



Run Workflow

amr_identified_by_ncbi *



158: amr_identified_by_ncbi

accepted formats ▾

output_of_abricate_with_amrfinderncbi

vfs_of_genes_identified_by_vfdb *



155: vfs_of_genes_identified_by_vfdb

accepted formats ▾

output_of_abricate_with_vfdb

metadata - optional



Nothing selected

accepted formats ▾

samples_metadata

removed_hosts_percentage_tabular *



95: removed_hosts_percentage_tabular

accepted formats ▾

removed_hosts_percentage_tabular

mapping_mean_depth_per_sample *

Metagenomic sequencing - foodborne pathogens

Heatmap

- **Heatmap:** Displays the presence or absence of bacterial pathogen genes across samples, highlighting shared and unique genes. This approach aids in quickly identifying the prevalence of pathogen genes and relationships between samples.

Heatmap:

- Remove beginning with the following parameters:
- Group with the following parameters
- Filter empty datasets with the following parameters
- Column join (Galaxy version 0.0.3) with the following parameters
- Column Regex Find And Replace (Galaxy version 1.0.3) with the following parameters

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

- Header:** Galaxy India
- Left Sidebar:** Includes 'Upload', 'Tools' (selected), 'Workflows', 'Workflow Invocations', 'Visualizations', 'Histories', 'History' (selected), 'Notifications', and 'Settings'.
- Tool Panel:** Shows 'Remove beginning' selected under 'Tools'.
- Tool Configuration:**
 - Tool Parameters:** 'Remove first' is set to '1'. A note says: 'This is a batch mode input field. Individual jobs will be triggered for each dataset.'
 - Additional Options:** 'Email notification' is set to 'No'.
 - Help:** 'What it does' is visible.
- History Panel:** Shows a workflow named 'Pathogen detection from (direct Nanopore) sequencing data' with a dataset '155: vfs_of_genes_identified_by_vfdb'.

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy web interface for the India instance. On the left, the navigation bar includes links for Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History (selected), Notifications, and Settings. The main area displays a workflow run titled "Pathogen detection from (direct Nanopore) sequencing data". A green message box indicates that the tool "Remove beginning" has been started and added 2 jobs to the queue. Below this, a "Citizen Science Experiment!" box contains a "Citizen Science Project" section with instructions to identify the sex of marmalade hoverflies. To the right, the History panel lists several datasets and collections, including "2693: Remove beginning on collection 155", "2569: newick_genes_tree_graphs_collection", "2445: fasttree_nhx", and "2444: filtered_empty_datasets".

The screenshot shows the Galaxy web interface for the India instance. The left sidebar is identical to the previous screenshot. The central area is focused on the "Group" tool configuration. The "Tool Parameters" section includes fields for "Select data" (set to "2693: Remove beginning on collection 155"), "Group by column" (set to "Column: 6"), and "Ignore case while grouping?" (set to "No"). Below these, there is an optional section for "Ignore lines beginning with these characters" with a list of symbols: >, @, +, <, *, and ^.

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India interface. On the left, the navigation bar includes options like Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History, Notifications, and Settings. The main area displays a success message for a 'Group' tool run:

Started tool Group and successfully added 2 jobs to the queue.

It produces 2 outputs:

- 2697: Group on data 2694
- 2698: Group on data 2695

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.

We need your support ...

If Galaxy helped with the analysis of your data, please do not forget to **cite**:

The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update
Nucleic Acids Research, gkae410
doi:10.1093/nar/gkae410

And please **acknowledge** the European Galaxy server:

The Galaxy server used for some calculations is partly funded by the German

The right side shows the History panel with several entries:

- 2696: Group on collection 2693
- 2693: Remove beginning on collection 155
- 2569: newick_genes_tree_graphs_collection
- vts_newick_genes_tree_graphs_collection
- 2445: fasttree_nhx

The screenshot shows the Galaxy India interface with the 'Tools' section selected. The 'Filter empty datasets' tool is currently active. The tool parameters include:

Tool Parameters

Input Collection *: accepted formats ▾ 2705: Group on collection 2702

Additional Options

Email notification: No

Send an email notification when the job completes.

Help

Synopsis

Removes empty elements from a collection.

This tool takes a dataset collection and filters out (removes) empty datasets. This is useful for continuing a multi-sample analysis when downstream tools require datasets to have content.

Collection (a list)
Sample 1 (100 Kb)

Collection (a list)
Sample 1 (100 Kb)

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India web interface. On the left, a sidebar navigation includes: Upload, Tools (selected), Workflows, Workflow Invocations, Visualization, Histories, History (selected), Notifications, and Settings. The main area displays a workflow history titled "Pathogen detection from (direct Nanopore) sequencing data". It shows four items: "2701: data 2698 and data 2697 (filtered empty datasets)" (a list with 2 datasets), "2696: Group on collection 2693" (a list with 2 datasets), "2693: Remove beginning on collection 155" (a list with 2 datasets), and "2569: newick_genes_tree_graphs_collection" (a list with 123 datasets). A green message box at the top indicates a successful tool run: "Started tool Filter empty datasets and successfully added 1 job to the queue. It produces this output: 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." Below this is a "Tool recommendation" section for the "PHD Comics" tool, which suggests other tools based on machine learning analysis.

This screenshot shows the configuration of the "Column join" tool in Galaxy India. The left sidebar is identical to the previous screenshot. The main area shows the "Tool Parameters" for "Column join on multiple datasets (Galaxy Version 0.0.3)". The "Tool Parameters" section includes: "Tabular files" (with a dropdown showing "2701: data 2698 and data 2697 (filtered empty datasets)"), "Identifier column" (set to "1"), "Number of header lines in each input file" (set to "0"), "Add column name to header" (radio button set to "Yes"), and "Fill character" (set to "0"). A note explains that setting "Number of header lines" to 0 adds a header line with column names. The right side of the screen shows the same workflow history as the first screenshot, listing the same four items.

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India web interface. On the left, the navigation bar includes 'Upload', 'Tools' (selected), 'Workflows', 'Workflow Invocations', 'Visualization', 'Histories', 'History' (selected), 'Notifications', and 'Settings'. The main area displays a 'Column join' tool run. A green success message states: 'Started tool Column join and successfully added 1 job to the queue.' It also says 'It produces this output:' followed by '2702: Column join on data 2700 and data 2699'. Below this, a note says: '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.' To the right, a 'Citizen Science Experiment!' section titled 'Citizen Science Project' asks users to identify the sex of marmalade hoverflies. It includes a photograph of several green flies and a link to Wikipedia. The 'History' panel on the right lists several runs: '2702: Column join on data 2700 and data 2699' (status: running), '2701: data 2698 and data 2697 (filtered empty datasets)' (status: finished), '2696: Group on collection 2693' (status: finished), and '2693: Remove beginning on collection 155' (status: finished). The top right corner shows 'Using 18%'.

This screenshot shows the Galaxy India interface with the 'Tools' tab selected. The left sidebar includes 'Upload', 'Tools' (selected), 'Workflows', 'Workflow Invocations', 'Visualization', 'Histories', 'History' (selected), 'Notifications', and 'Settings'. The central area shows a 'Column Regex Find And Replace' tool run. The 'Tool Parameters' section includes 'Select cells from' (set to '2702: Column join on data 2700 and data 2699'), 'using column' (set to 'Column: 1'), and a 'Check' section with a single entry: '1: Check' with 'Find Regex' set to '#KEY'. The 'History' panel on the right shows runs: '2702: Column join on data 2700 and data 2699' (status: running), '2701: data 2698 and data 2697 (filtered empty datasets)' (status: finished), '2696: Group on collection 2693' (status: finished), and '2693: Remove beginning on collection 155' (status: finished). The top right corner shows 'Using 18%'.

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy web interface with the following components:

- Left Sidebar:** Includes sections for Upload, Tools, Workflows, Workflow Invocations, Visualizations, Histories, History (with a dropdown menu), Notifications, and Settings.
- Top Bar:** Shows the Galaxy India logo, navigation links (Workflow, Visualize, Data, Help, User, Logout, Help, History, and a bell icon), and a status bar indicating "Using 18%".
- Central Panel:**
 - Tools Section:** Displays a "Column Regex Find And Replace" tool entry.
 - Message Box:** A green box indicates that a job has started successfully: "Started tool Column Regex Find And Replace and successfully added 1 job to the queue. It produces this output: 2703: Column Regex Find And Replace on data 2702". Below it, instructions for checking the status of queued jobs are provided.
 - Support Section:** A grey box titled "We need your support ..." contains information about citing Galaxy and acknowledging the European Galaxy server.
 - Bottom Message Box:** A white box at the bottom provides funding information: "The Galaxy server used for some calculations is partly funded by the German Federal Ministry of Education and Research BMBF grant 031A536A de.NBI".
- Right Panel:** A "History" panel showing a list of completed jobs:
 - 2703: Column Regex Find And Replace on data 2702
 - 2702: Column join on data 2700 and data 2699
 - 2701: data 2698 and data 2697 (filtered empty datasets)
 - 2696: Group on collection 2693
 - 2693: Remove beginning on collection 155The history panel also shows a total storage usage of 4.69 GB and 165 datasets.

Draw heatmap

Heatmap:

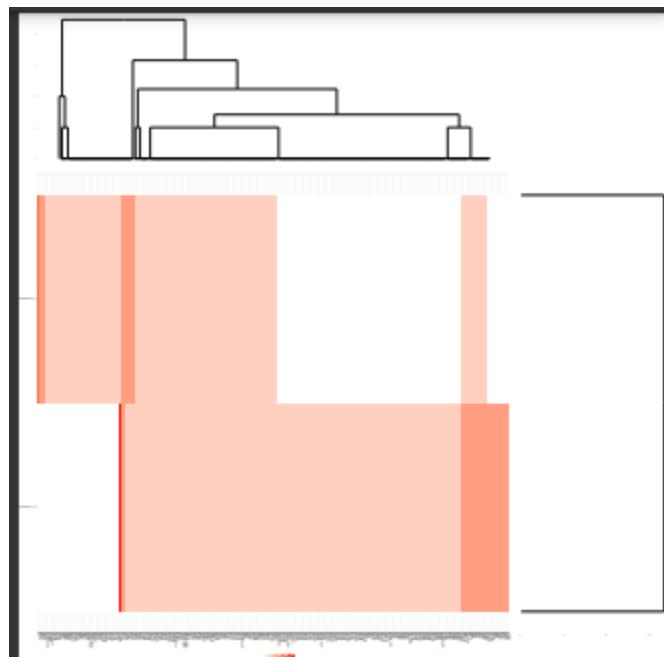
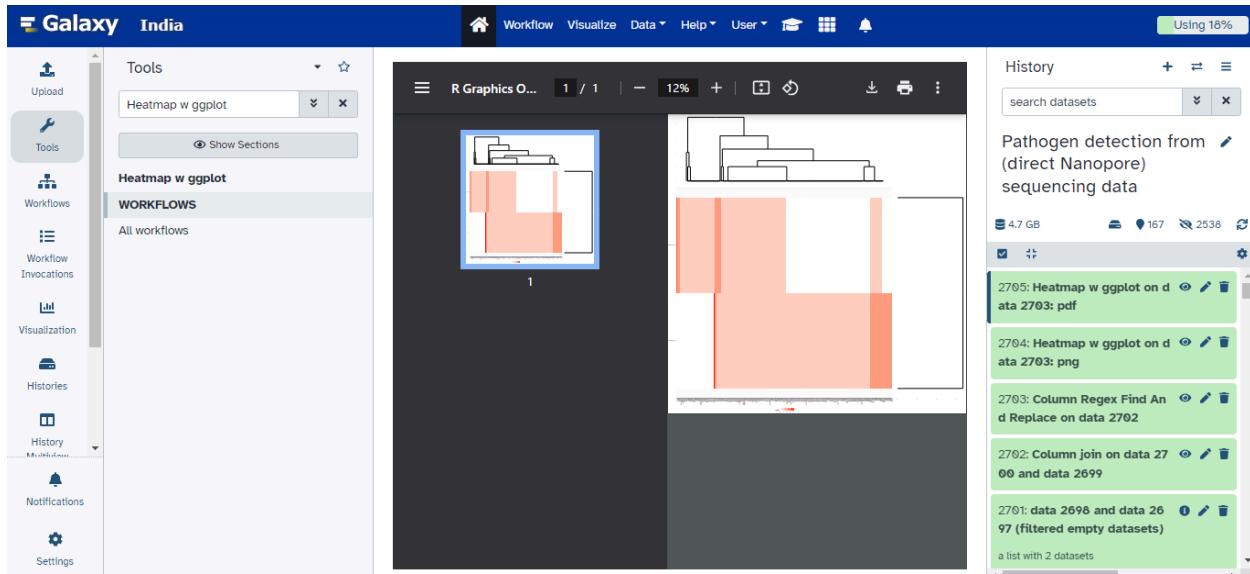
- Heatmap w ggplot (Galaxy version 3.4.0+galaxy0) with the following parameter

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy web interface for 'Heatmap w ggplot'. The left sidebar includes 'Upload', 'Tools' (selected), 'Workflows', 'Workflow Invocations', 'Visualization' (selected), 'Histories', 'History', 'Notifications', and 'Settings'. The main panel displays the 'Tool Parameters' for 'Heatmap w ggplot'. It shows a 'Select table' dropdown set to '2703: Column Regex Find And Replace on data 2702 accepted formats'. Under 'Select input dataset options', there is a dropdown for 'Dataset with header and row names' and a note about specific dataset input for reading. The 'Select column, for row names' dropdown is set to 'Column: 1' with a warning about header shifts. The 'Sample names orientation' section has 'vertical' selected. The 'Plot title' field contains 'Pathogenic Genes Per Samples'. The right panel shows the 'History' tab with a list of recent datasets, including 'Pathogen detection from (direct Nanopore) sequencing data'.

The screenshot shows the Galaxy interface after the 'Heatmap w ggplot' tool has been run. The left sidebar remains the same. The main panel now displays a green success message: 'Started tool Heatmap w ggplot and successfully added 1 job to the queue. It produces 2 outputs: 2704: Heatmap w ggplot on data 2703: png, 2705: Heatmap w ggplot on data 2703: pdf'. Below this message is a dark blue box labeled 'PHD Comics' with a 'Random' button and a small image icon. The right panel shows the 'History' tab with the same list of datasets as the previous screenshot.

Metagenomic sequencing - foodborne pathogens



Metagenomic sequencing - foodborne pathogens

Phylogenetic Tree Building

- **Phylogenetic Tree Construction:** Uses virulence factors (VFs) as markers to trace pathogen evolution across samples. ClustalW is used for multiple sequence alignment, FASTTREE builds the tree, and Newick Display visualizes it, aiding in phylogenetic analysis of unknown samples.

Extract the sequences of the VFs:

- Collapse Collection (Galaxy version 5.1.0) with the following parameters
- Collapse Collection (Galaxy version 5.1.0) with the following parameters
- Split by group (Galaxy version 0.6) with the following parameters
- Cut
- Remove beginning
- bedtools getfasta

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy web interface with the title "Galaxy India". The left sidebar includes links for Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History (selected), Notifications, and Settings. The main panel displays the "Collapse Collection" tool configuration. The "Tool Parameters" section specifies "Collection of files to collapse into single dataset" as "178: contigs" and "accepted formats" as "contigs". The "Keep one header line" option is set to "No". The "Prepend File name" option is set to "Yes", and the "Where to add dataset name" dropdown is set to "Same line and only once per dataset". The "Additional Options" section includes an "Email notification" toggle set to "No". On the right, the "History" panel lists several recent jobs, including "Heatmap w ggplot on data 2703: pdf" and "Column Regex Find And Replace on data 2702". The status bar at the top right indicates "Using 18%".

The screenshot shows the Galaxy web interface with the title "Galaxy India". The left sidebar includes links for Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History (selected), Notifications, and Settings. The main panel displays the "Collapse Collection" tool configuration, which has been successfully run. A green success message states: "Started tool Collapse Collection and successfully added 1 job to the queue. It produces this output: • 2706: Collapse Collection on data 160 and data 179". Below this, a "PHD Comics" section shows a "Random" button and a "Tool recommendation" section listing recommended tools like "Compute Replace". The "History" panel on the right lists several recent jobs, including "Collapse Collection on data 160 and data 179", "Heatmap w ggplot on data 2703: pdf", and "Column Regex Find And Replace on data 2702". The status bar at the top right indicates "Using 18%".

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India web interface. On the left, a sidebar includes options for Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History (selected), Notifications, and Settings. The main area displays the 'Collapse Collection' tool configuration. The 'Tool Parameters' section specifies 'Collection of files to collapse into single dataset' as '172: vfs'. Under 'Additional Options', 'Email notification' is set to 'No'. To the right, the 'History' panel lists several completed jobs, such as 'Collapse Collection on data 180 and data 179' and 'Heatmap w ggplot on data 2703: pdf'. A progress bar at the top indicates 'Using 18%'. The title bar shows 'Galaxy India'.

This screenshot shows the Galaxy India interface after the 'Collapse Collection' tool has been run. The main message in the center says: 'Started tool Collapse Collection and successfully added 1 job to the queue. It produces this output: 2707: Collapse Collection on data 174 and data 173'. Below this, a 'We need your support ...' section encourages citation and acknowledgment. The 'History' panel on the right shows the completed job '2707: Collapse Collection on data 174 and data 173'. The title bar shows 'Galaxy India'.

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy web interface for the 'Split by group' tool (Galaxy Version 0.6). The left sidebar includes links for Upload, Tools, Workflows, Workflow Invocations, Visualization, Histories, History (selected), Notifications, and Settings. The main panel displays the 'Tool Parameters' for 'Split by group'. Under 'File to split', the dataset '2707: Collapse Collection on data 174 and data 173' is selected. The 'on column' dropdown is set to 'Column: 13'. The 'Include header in splits?' option is checked ('Yes'). Under 'Additional Options', the 'Email notification' checkbox is unchecked ('No'). A 'Run Tool' button is at the bottom. To the right is a 'History' panel showing a list of jobs:

- 2707: Collapse Collection on data 174 and data 173
- 2706: Collapse Collection on data 180 and data 179
- 2705: Heatmap w ggplot on data 2703: pdf
- 2704: Heatmap w ggplot on data 2703: png
- 2703: Column Regex Find And Replace on data 2702
- 2702: Column join on data 2701

The screenshot shows the Galaxy web interface after the 'Split by group' tool has been run successfully. The main panel displays a green success message: 'Started tool Split by group and successfully added 1 job to the queue.' It also includes a note about citation and funding information. The right side shows the same history list as the previous screenshot.

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy web interface for 'Galaxy India'. On the left, a sidebar navigation includes 'Upload', 'Tools' (selected), 'Workflows', 'Workflow Invocations', 'Visualization', 'Histories', 'History' (selected), 'Notifications', and 'Settings'. The main area displays the 'Cut' tool configuration for 'Cut columns from a table (Galaxy Version 1.0.2)'. The 'Tool Parameters' section has 'Cut columns' set to 'c2,c3,c4' and 'Delimited by' set to 'Tab'. Under 'From', a dropdown menu shows '2708: Split by group collection' with 'accepted formats' listed as 'fastq'. A note states: 'This is a batch mode input field. Individual jobs will be triggered for each dataset.' The 'Additional Options' section has 'Email notification' set to 'No'. A 'Run Tool' button is at the bottom. To the right is a 'History' panel showing a list of datasets: '2708: Split by group collection' (326 datasets), '2707: Collapse Collection on data 174 and data 173', '2706: Collapse Collection on data 160 and data 179', '2705: Heatmap w ggplot on data 2703: pdf', and '2704: Heatmap w ggplot on data 2703: png'. The top right corner shows 'Using 18%'. The status bar at the bottom indicates '4.71 GB'.

This screenshot shows the Galaxy interface after the 'Cut' tool has been run. The 'History' panel on the right lists the completed job '3035: Cut on collection 2708' (326 datasets). The main area shows a green success message: 'Started tool Cut and successfully added 326 jobs to the queue. It produces 326 outputs:' followed by a bulleted list of 326 individual cut operations. The status bar at the bottom indicates 'Using 18%'.

- 3036: Cut on data 2709
- 3037: Cut on data 2710
- 3038: Cut on data 2711
- 3039: Cut on data 2712
- 3040: Cut on data 2713
- 3041: Cut on data 2714
- 3042: Cut on data 2715
- 3043: Cut on data 2716
- 3044: Cut on data 2717
- 3045: Cut on data 2718
- 3046: Cut on data 2719
- 3047: Cut on data 2720
- 3048: Cut on data 2721
- 3049: Cut on data 2722
- 3050: Cut on data 2723
- 3051: Cut on data 2724
- 3052: Cut on data 2725
- 3053: Cut on data 2726
- 3054: Cut on data 2727
- 3055: Cut on data 2728
- 3056: Cut on data 2729
- 3057: Cut on data 2730
- 3058: Cut on data 2731

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy India web interface. On the left, a sidebar navigation includes: Upload, Tools (selected), Workflows, Workflow Invocations, Visualization, Histories, History (selected), Notifications, and Settings.

The main content area displays the "Remove beginning of a file" tool from Galaxy Version 1.0.0. The "Tool Parameters" section includes:

- Remove first ***: A text input field containing "1".
- from ***: A dropdown menu showing "3035: Cut on collection 2706" and "accepted formats". A tooltip indicates: "This is a batch mode Input field. Individual jobs will be triggered for each dataset."

The "Additional Options" section includes:

- Email notification**: A toggle switch set to "No". A tooltip says: "Send an email notification when the job completes."

The "Help" section contains the tool's description: "This tool removes a specified number of lines from the beginning of a dataset."

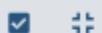
The top right corner shows "Using 18%" and the bottom right corner shows the "History" panel with a list of recent workflows:

- 3035: Cut on collection 2706 (a list with 326 datasets)
- 2706: Split by group collection (a list with 326 datasets)
- 2707: Collapse Collection on data 174 and data 173
- 2706: Collapse Collection on data 180 and data 179
- 2705: Heatmap w ggplot on data 2703: pdf

Metagenomic sequencing - foodborne pathogens

Pathogen detection from  (direct Nanopore) sequencing data

 4.71 GB  172  3516 



3362: Remove beginning on collection   3035

a list with 326 tabular datasets

3035: Cut on collection   2708

a list with 326 tabular datasets

2708: Split by group collection   

n

a list with 326 tabular datasets

2707: Collapse Collection on    data 174 and data 173

Metagenomic sequencing - foodborne pathogens

The screenshot shows the Galaxy web interface with the title "Galaxy India". The left sidebar includes "Upload", "Tools" (selected), "Workflows", "Workflow Invocations", "Visualization", "Histories", "History" (selected), "Notifications", and "Settings". The main area displays the "Edit Collection Attributes" dialog for a collection named "Pathogen detection from (direct Nanopore) sequencing data". The dialog shows a message: "Datatype changed successfully." and a dropdown menu for "New Type" set to "bed". The right sidebar shows the "History" panel with several recent operations:

- 3362: Remove beginning on collection 3035 (a list with 326 tabular datasets)
- 3035: Cut on collection 2708 (a list with 326 tabular datasets)
- 2708: Split by group collection (a list with 326 tabular datasets)
- 2707: Collapse Collection on data 174 and data 173

The screenshot shows the Galaxy web interface with the title "Galaxy India". The left sidebar includes "Upload", "Tools" (selected), "Workflows", "Workflow Invocations", "Visualization", "Histories", "History" (selected), "Notifications", and "Settings". The main area displays the configuration for the "bedtools getfasta" tool. The "Tool Parameters" section shows the input dataset "3362: Remove beginning on collection 3035" and a note: "This is a batch mode input field. Individual jobs will be triggered for each dataset. (-bed)". The "Choose the source for the FASTA file" section shows the "History" dropdown selected, and the "FASTA file" dropdown set to "2706: Collapse Collection on data 180 and data 179". A note at the bottom states: "Use the 'name' column in the BED file and the coordinates for the FASTA headers in the output FASTA file". The right sidebar shows the "History" panel with several recent operations:

- a list with 326 tabular datasets
- 2707: Collapse Collection on data 174 and data 173
- 2706: Collapse Collection on data 180 and data 179
- 2705: Heatmap w ggplot on data 2703: pdf
- 2704: Heatmap w ggplot on data 2703: png
- 2703: Column Regex Find An

Metagenomic sequencing - foodborne pathogens



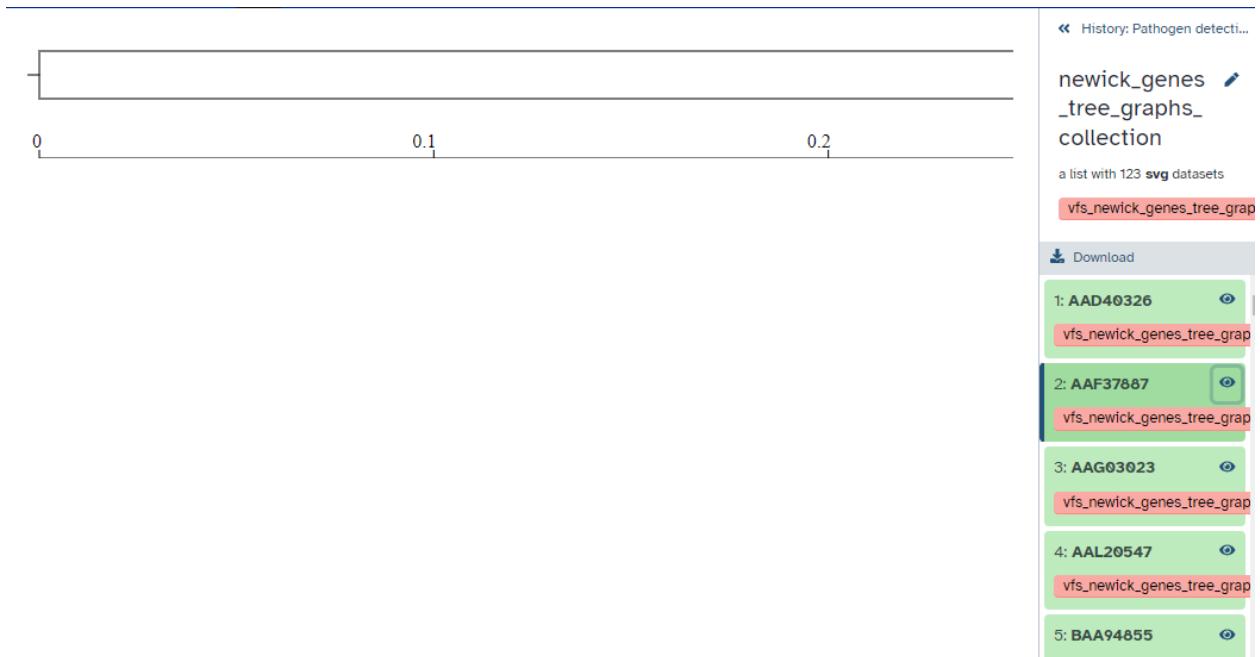
Phylogenetic Tree building:

- ClustalW (Galaxy version 2.1+galaxy1) with the following parameters
- FASTTREE (Galaxy version 2.1.10+galaxy1) with the following parameters:
- Newick Display (Galaxy version 1.6+galaxy1) with the following parameters

Metagenomic sequencing - foodborne pathogens

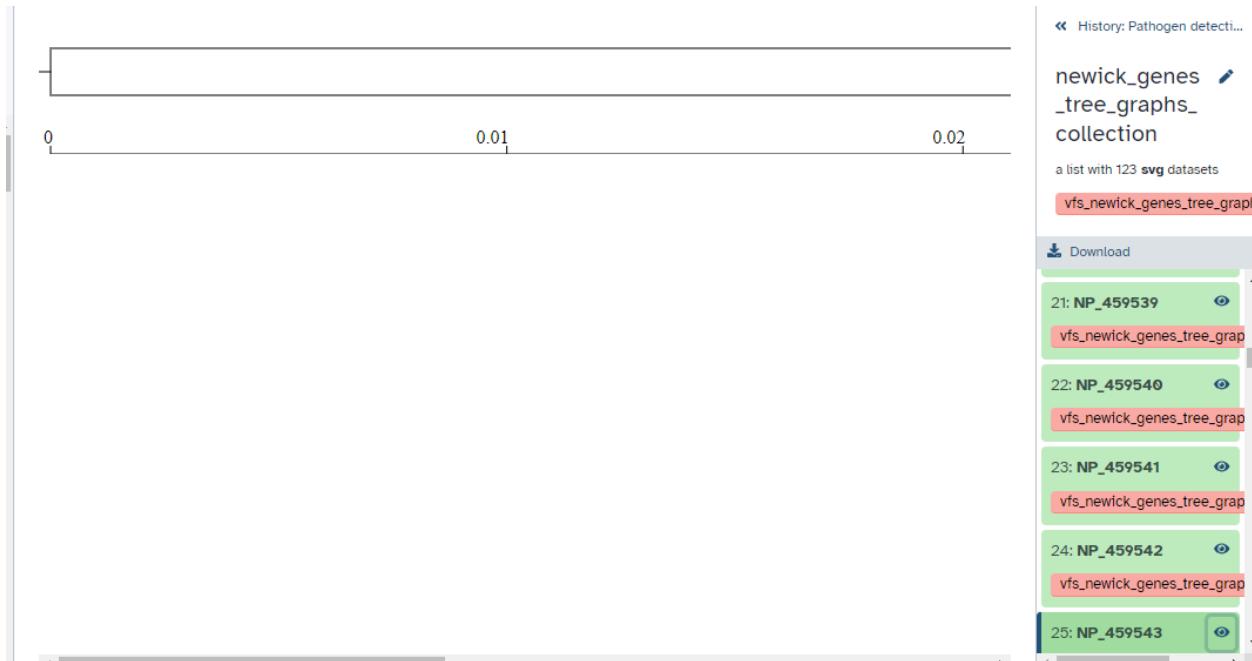
IDs: AAF37887 and NP_459543 look like. To access that go to the output of Newick

AAF37887



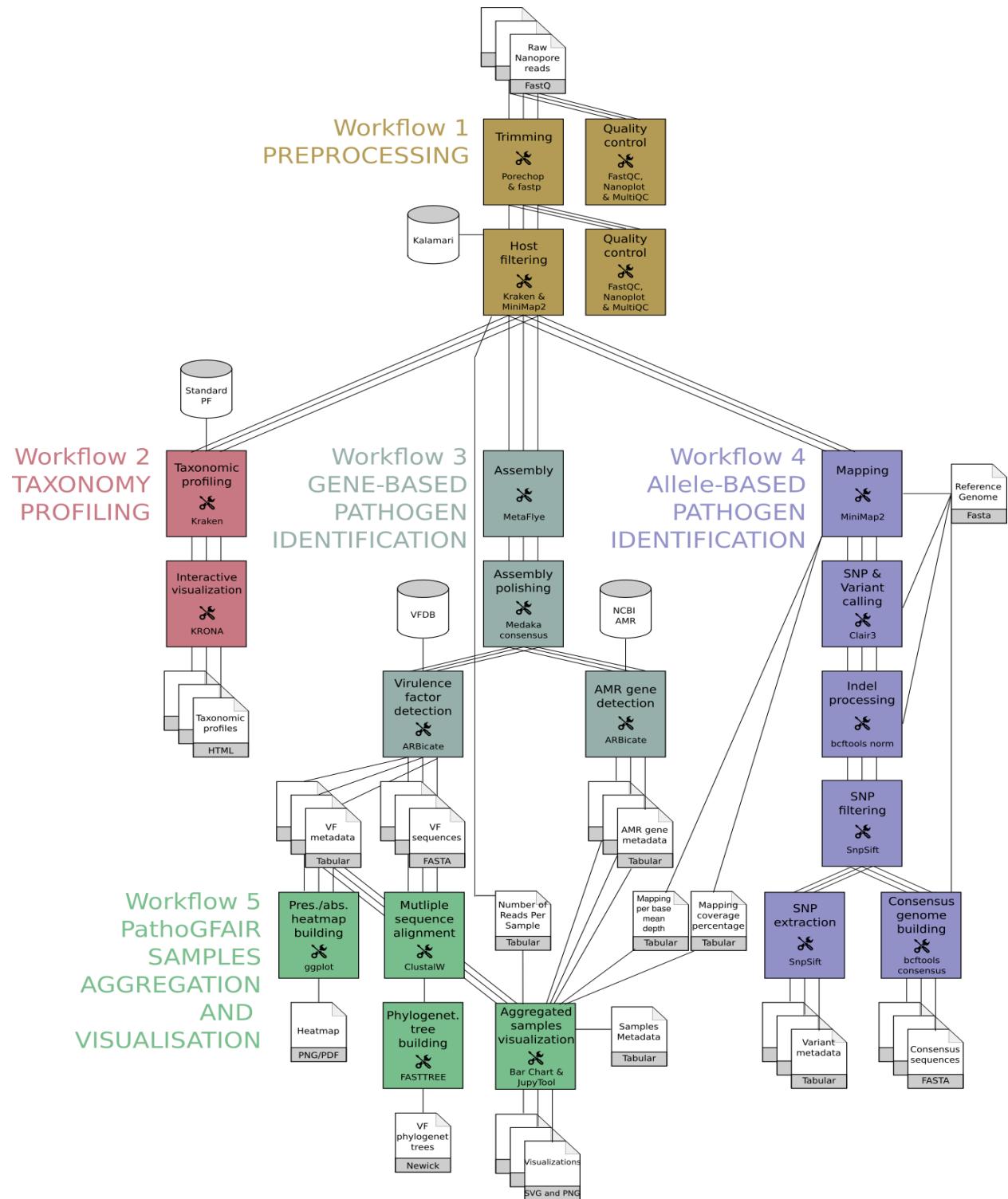
Metagenomic sequencing - foodborne pathogens

NP_459543



Metagenomic sequencing - foodborne pathogens

Conclusion:



Metagenomic sequencing - foodborne pathogens

The presented workflow offers a robust, end-to-end metagenomic sequencing pipeline for the identification and characterization of foodborne pathogens. By utilizing long-read sequencing data, the pipeline allows for high-accuracy pathogen detection through several key workflows:

1. **Preprocessing:** Raw data undergoes trimming, quality control, and host-read filtering to remove noise and prepare for downstream analyses.
2. **Taxonomy Profiling:** A broad assessment of microbial composition is performed, highlighting the taxonomic identities present in the sample.
3. **Gene-Based and Allele-Based Pathogen Identification:** Both gene-based methods (focused on pathogen-specific genes) and allele-based approaches (through variant and SNP calling) provide precise pathogen identification, allowing for the detection of antimicrobial resistance genes and virulence factors.
4. **Consensus Genome Building:** Using variant calls, a consensus genome is reconstructed, ensuring the generation of a high-quality genome for further analyses.
5. **Sample Aggregation and Visualization:** Finally, the aggregation of results into visual outputs like heatmaps and phylogenetic trees allows for a clear, interpretable summary of pathogen spread, virulence, and resistance patterns.

Metagenomic sequencing - foodborne pathogens

Overall, the pipeline efficiently integrates multiple tools and methods to deliver an accurate, in-depth characterization of foodborne pathogens, providing valuable insights into pathogen taxonomy, resistance profiles, and potential public health implications. The final visualizations, including phylogenetic trees and heatmaps, support decision-making for food safety and epidemiological tracking.