Tutorial 12
MetaPhlAn and HUMAnN

Zoe Sucato, Catherine Nguyen, Tobenna Oduah

MetaPhlAn [5]

MetaPhlAn stands for **Meta**genomic **Phl**ogenetic **An**alysis.

- Computational tool for profiling the composition of microbial communities from metagenomic shotgun sequencing data.
- Relies on unique clade-specific marker genes.
  - From ~17,000 reference genomes:
    - 13,500 bacterial and archeal genomes
    - 3,500 viral genomes
    - 110 eukaryotic genomes
- Uses bowtie2.

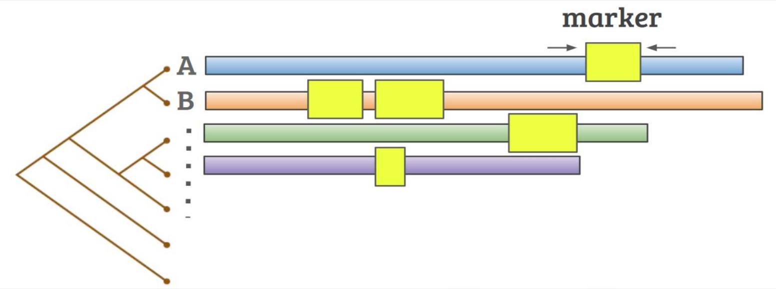
Definitions

Clade: Group of organisms believed to have evolved from a common ancestor on a phylogenetic tree.

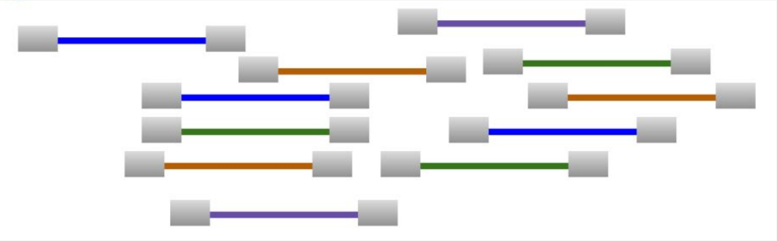
Clade-specific marker: Coding sequences that are strongly conserved within the clade's genomes and do not possess substantial local similarity with any sequence outside the clade.

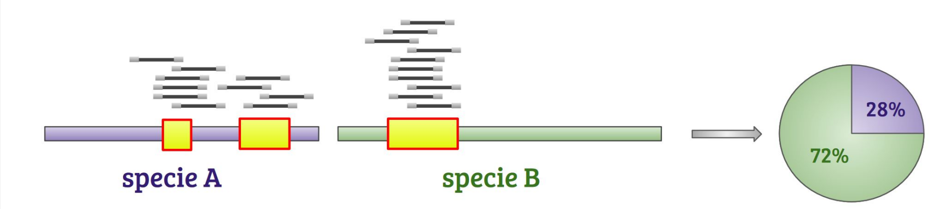
MetaPhlAn Algorithm [2]

Database of reference genomes and their relationships, with identified clade-specific markers



Sequenced sample





Reads mapped to marker genes

HUMAnN [3]

HUMAnN stands for the **H**MP **U**nified **M**etabolic **A**nalysis **N**etwork.

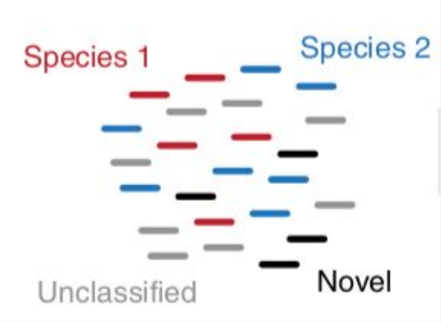
- A method for profiling the abundance of microbial metabolic pathways, including other molecular functions from metagenomic/metatranscriptomic sequencing data.

Definitions

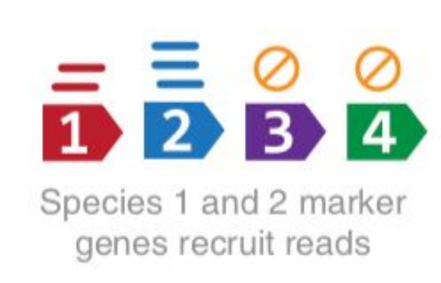
HMP: Human Microbiome Project. An initiative to research and understand the microbial components of the human genetic landscape and how that translates to health-related norms and concerns.

HUMAnN Algorithm [6]

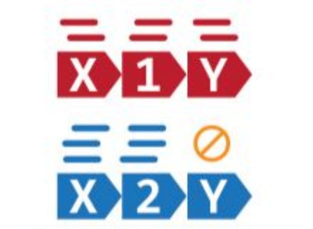
Sequenced reads

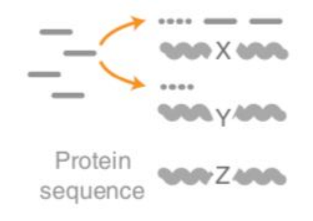


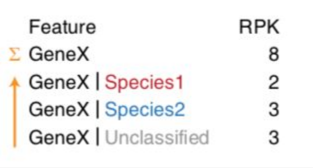
Reads get assigned to species (MetaPhlAn)



Genes matched to function







Feature	RPK
Σ GeneX	8
GeneX | Species1	2
GeneX | Species2	3
GeneX | Unclassified	3

MetaPhlAn

Which microbes are there?

HUMAnN

What can the microbes do?

Demo

Original Plans… Changed

Originally meant to use Biobakery Workflows.

Could not use Docker directly, so we used Singularity to run a Docker image.

Many issues were encountered:

- Pre-existing Singularity Container did not have the tools (wmgz) installed.
- Creating a new Singularity container with the Docker image failed due to incorrect dependencies.
- Using pip install was not viable since many modules were installed and some failed to install.

New Plan

Use Galaxy/Hutlab! [4]

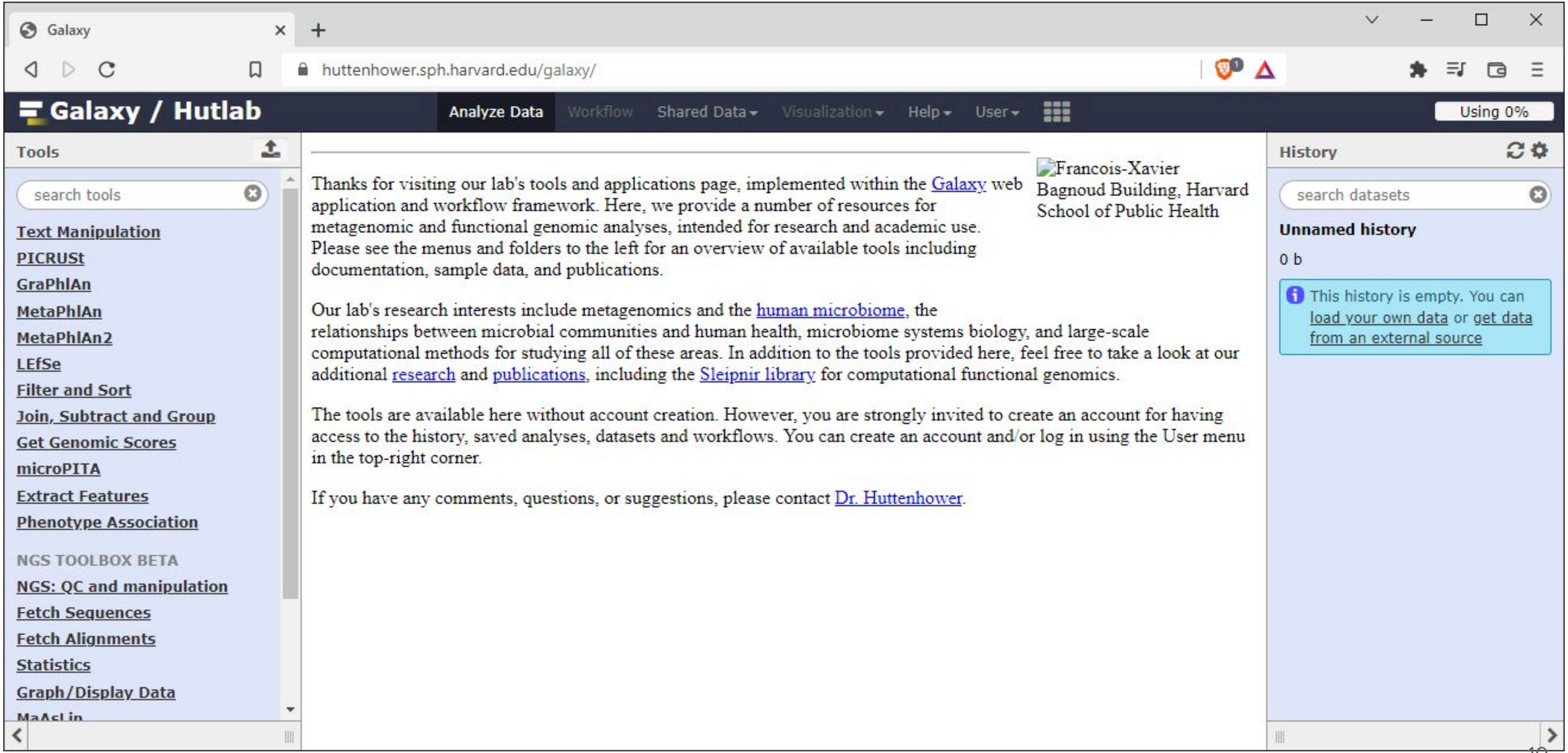
Galaxy / Hutlab

All online; no package installs.

However, it has limitations.

Example: Doesn’t make the output from one process nice for the next process

Go to the Galaxy/Hutlab Website.



Francois-Xavier Bagnoud Building, Harvard School of Public Health

Thanks for visiting our lab's tools and applications page, implemented within the Galaxy web application and workflow framework. Here, we provide a number of resources for metagenomic and functional genomic analyses, intended for research and academic use. Please see the menus and folders to the left for an overview of available tools including documentation, sample data, and publications.

Our lab's research interests include metagenomics and the human microbiome, the relationships between microbial communities and human health, microbiome systems biology, and large-scale computational methods for studying all of these areas. In addition to the tools provided here, feel free to take a look at our additional research and publications, including the Sleipnir library for computational functional genomics.

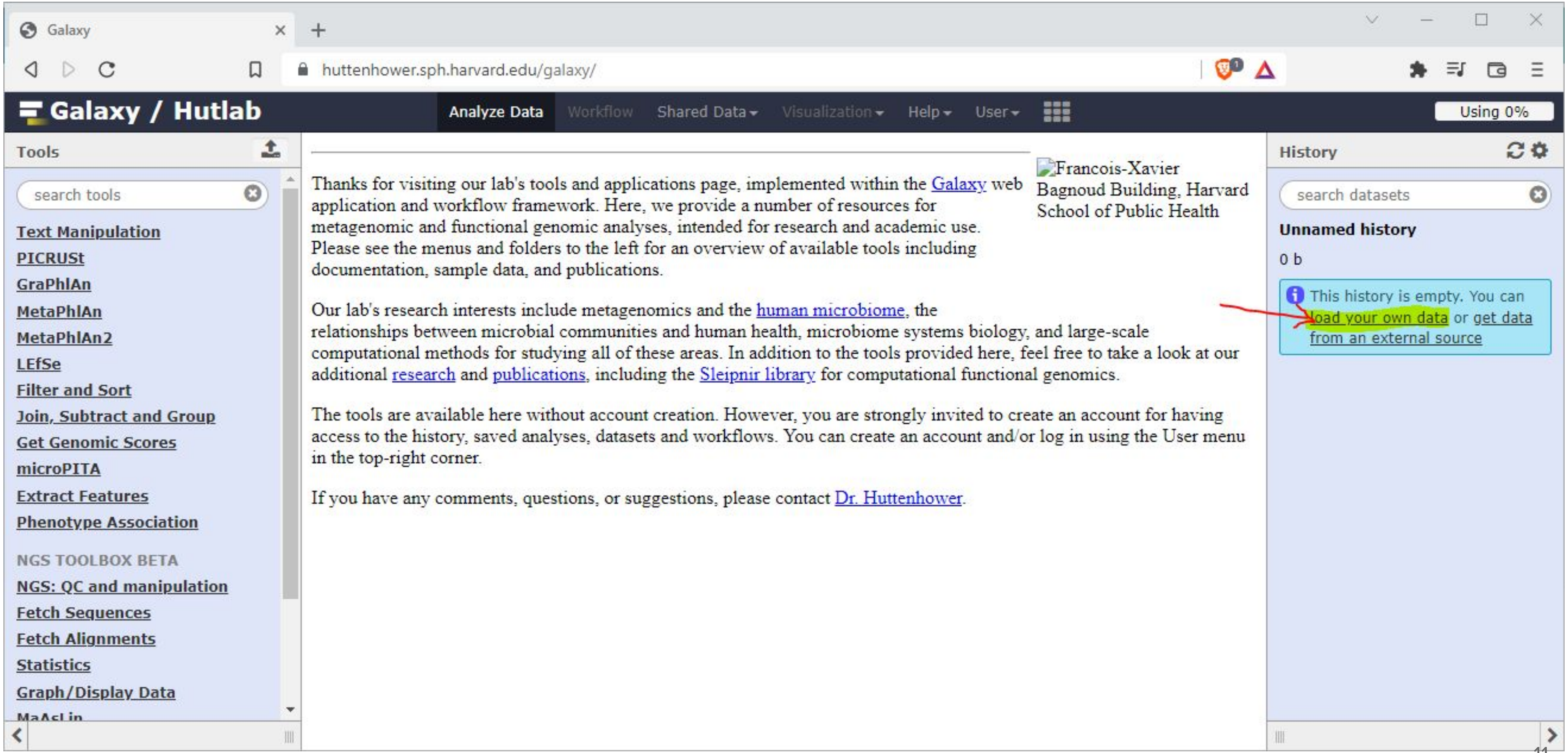
The tools are available here without account creation. However, you are strongly invited to create an account for having access to the history, saved analyses, datasets and workflows. You can create an account and/or log in using the User menu in the top-right corner.

If you have any comments, questions, or suggestions, please contact Dr. Huttenhower.

This history is empty. You can load your own data or get data from an external source

https://huttenhower.sph.harvard.edu/galaxy/

To upload your fasta file, click “load your own data”.



Francois-Xavier Bagnoud Building, Harvard School of Public Health

Thanks for visiting our lab's tools and applications page, implemented within the Galaxy web application and workflow framework. Here, we provide a number of resources for metagenomic and functional genomic analyses, intended for research and academic use. Please see the menus and folders to the left for an overview of available tools including documentation, sample data, and publications.

Our lab's research interests include metagenomics and the human microbiome, the relationships between microbial communities and human health, microbiome systems biology, and large-scale computational methods for studying all of these areas. In addition to the tools provided here, feel free to take a look at our additional research and publications, including the Sleipnir library for computational functional genomics.

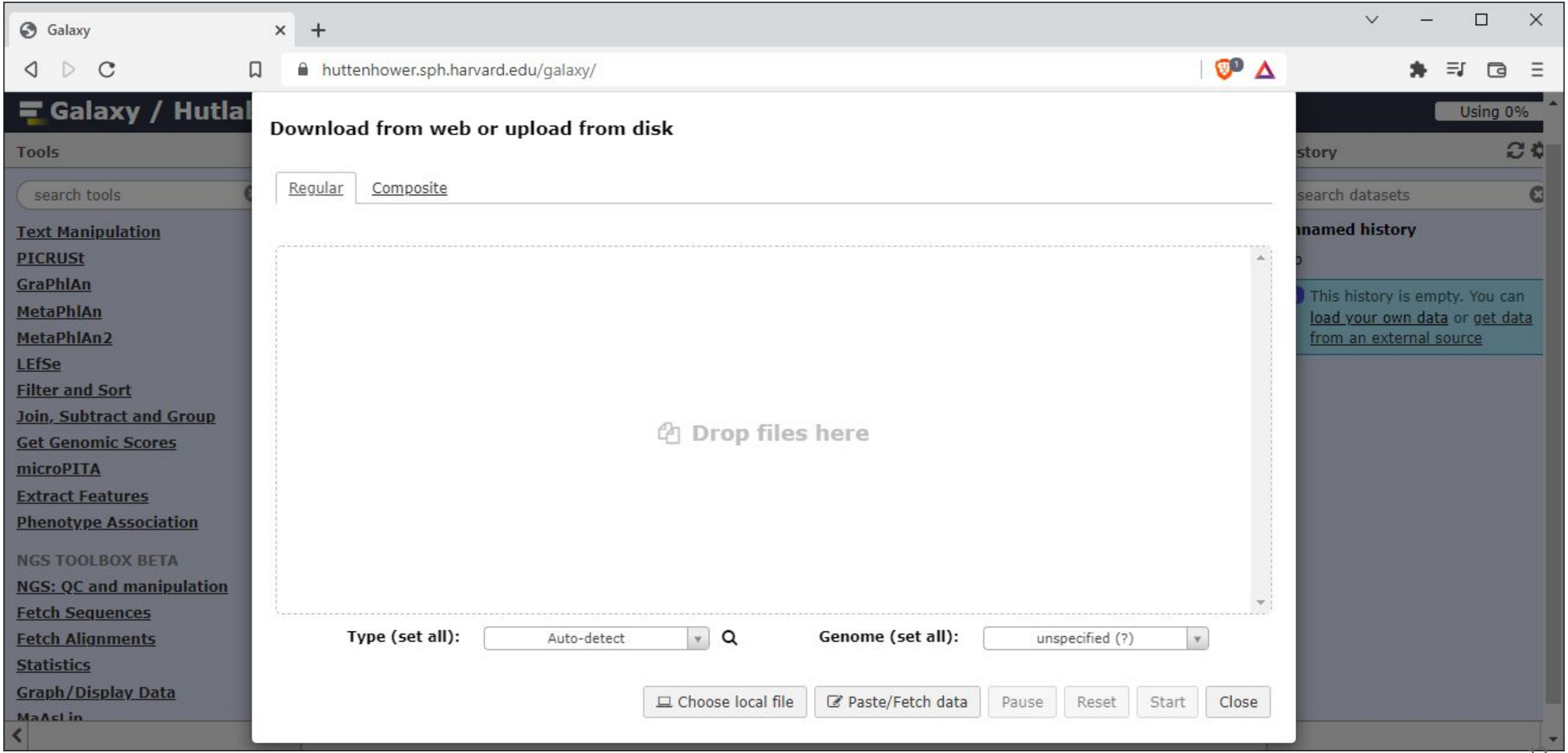
The tools are available here without account creation. However, you are strongly invited to create an account for having access to the history, saved analyses, datasets and workflows. You can create an account and/or log in using the User menu in the top-right corner.

If you have any comments, questions, or suggestions, please contact Dr. Huttenhower.

This history is empty. You can load your own data or get data from an external source

https://huttenhower.sph.harvard.edu/galaxy/

Drag your fasta file into this upload box.



Download from web or upload from disk

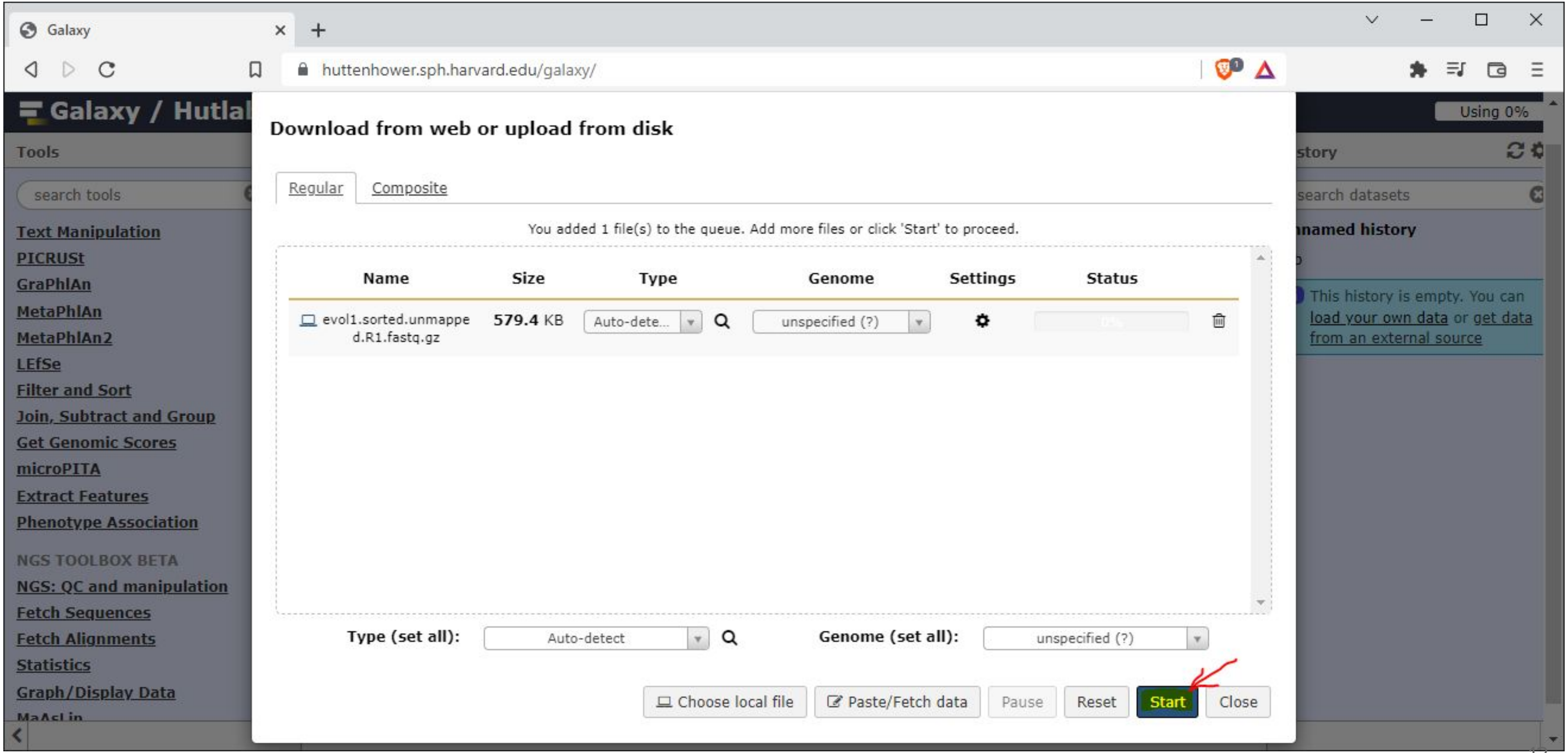
Type (set all): Auto-detect

Genome (set all): unspecified (?)

This history is empty. You can load your own data or get data from an external source

https://huttenhower.sph.harvard.edu/galaxy/

Click start to upload.



Download from web or upload from disk

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

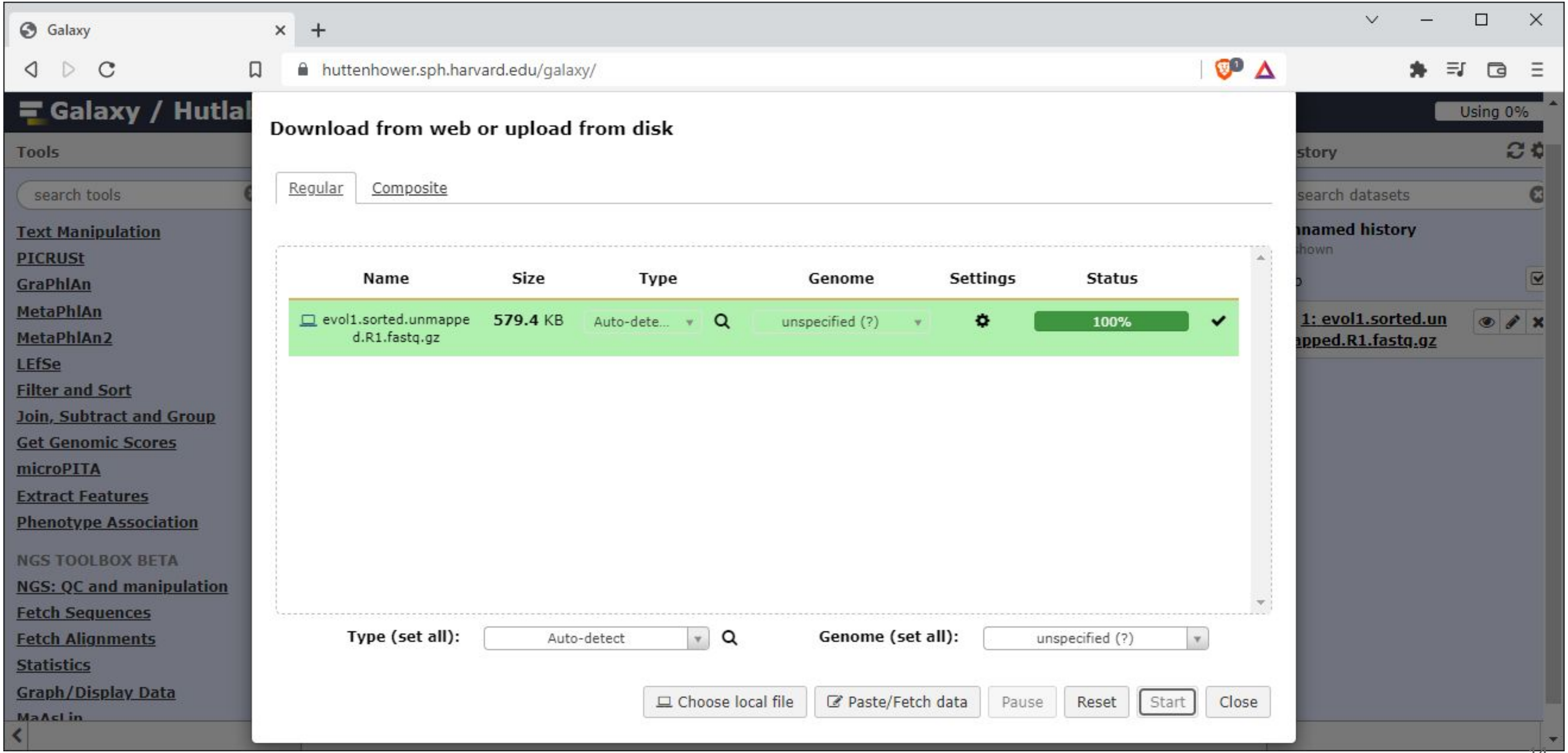
Name	Size	Type	Genome	Settings	Status
evol1.sorted.unmapped.R1.fastq.gz	579.4 KB	Auto-dete...	unspecified (?)		

Type (set all): Auto-detect

Genome (set all): unspecified (?)

https://huttenhower.sph.harvard.edu/galaxy/

Wait until status is 100%.



Download from web or upload from disk

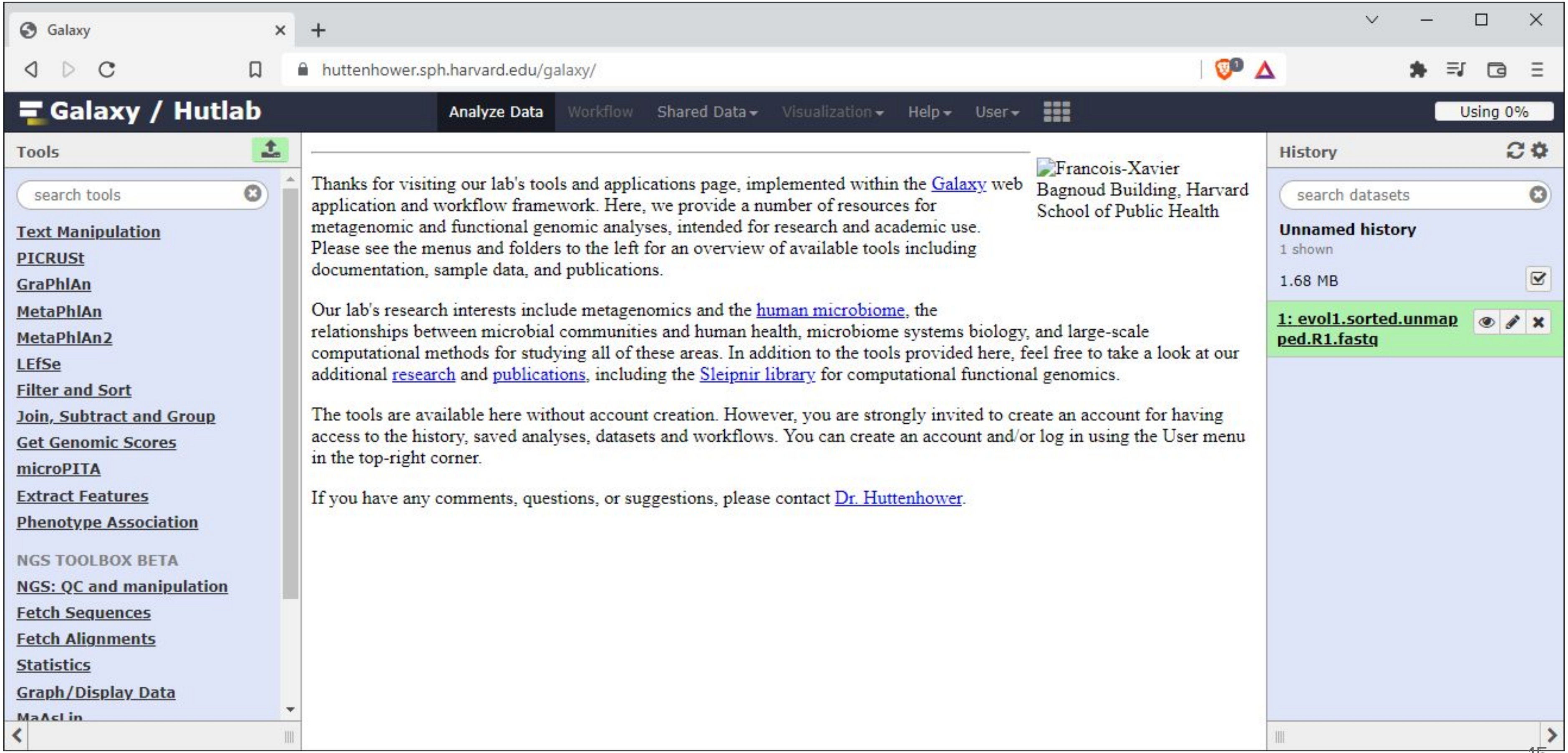
Name	Size	Type	Genome	Settings	Status
evol1.sorted.unmapped.R1.fastq.gz	579.4 KB	Auto-dete...	unspecified (?)	⚙	100% ✓

Type (set all): Auto-detect

Genome (set all): unspecified (?)

https://huttenhower.sph.harvard.edu/galaxy/

The uploaded file should appear in the right panel.



Francois-Xavier Bagnoud Building, Harvard School of Public Health

Thanks for visiting our lab's tools and applications page, implemented within the Galaxy web application and workflow framework. Here, we provide a number of resources for metagenomic and functional genomic analyses, intended for research and academic use. Please see the menus and folders to the left for an overview of available tools including documentation, sample data, and publications.

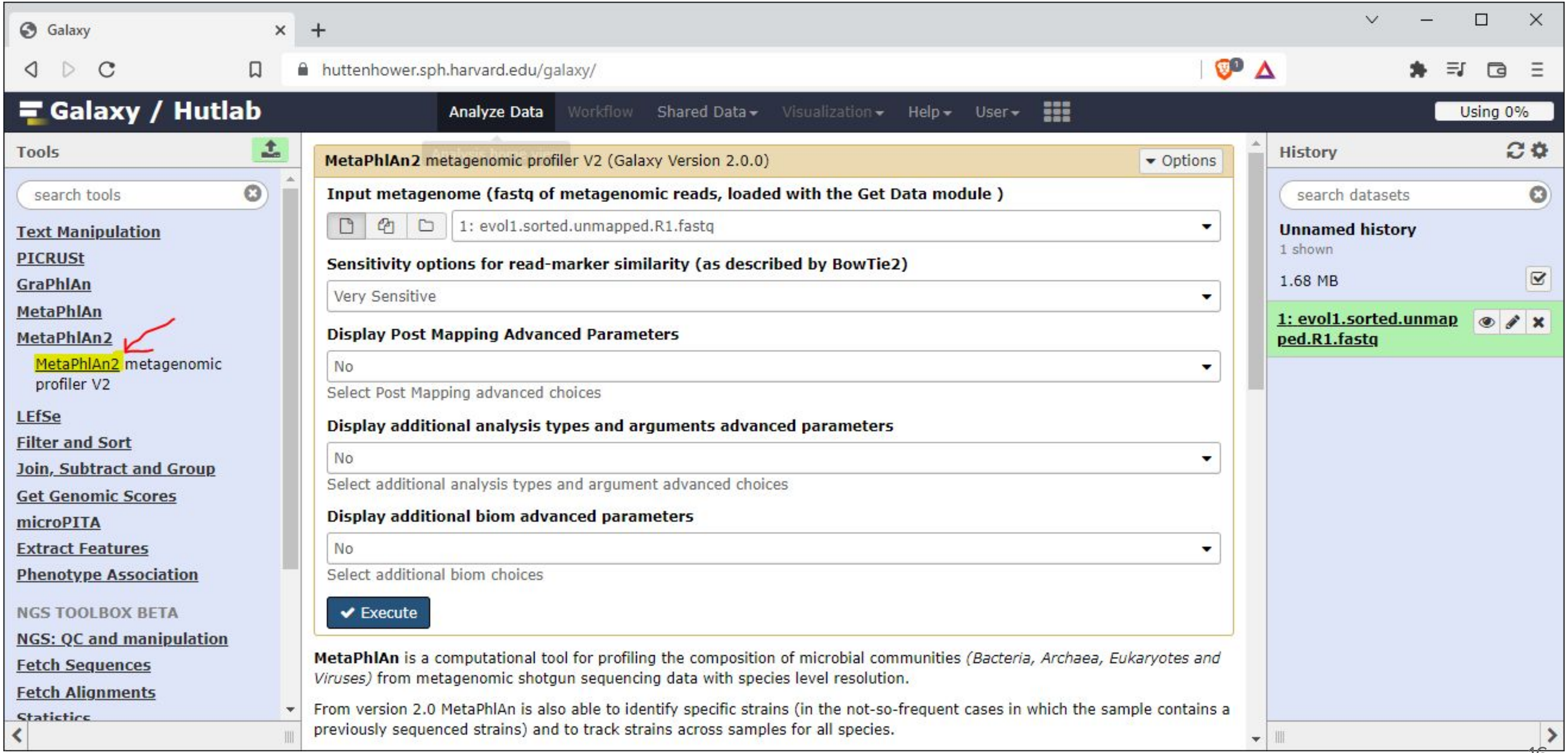
Our lab's research interests include metagenomics and the human microbiome, the relationships between microbial communities and human health, microbiome systems biology, and large-scale computational methods for studying all of these areas. In addition to the tools provided here, feel free to take a look at our additional research and publications, including the Sleipnir library for computational functional genomics.

The tools are available here without account creation. However, you are strongly invited to create an account for having access to the history, saved analyses, datasets and workflows. You can create an account and/or log in using the User menu in the top-right corner.

If you have any comments, questions, or suggestions, please contact Dr. Huttenhower.

https://huttenhower.sph.harvard.edu/galaxy/

Navigate to the MetaPhlAn2 profile from the left panel.



MetaPhlAn2 metagenomic profiler V2 (Galaxy Version 2.0.0)

Input metagenome (fastq of metagenomic reads, loaded with the Get Data module)
1: evol1.sorted.unmapped.R1.fastq

Sensitivity options for read-marker similarity (as described by BowTie2)
Very Sensitive

Display Post Mapping Advanced Parameters
No
Select Post Mapping advanced choices

Display additional analysis types and arguments advanced parameters
No
Select additional analysis types and argument advanced choices

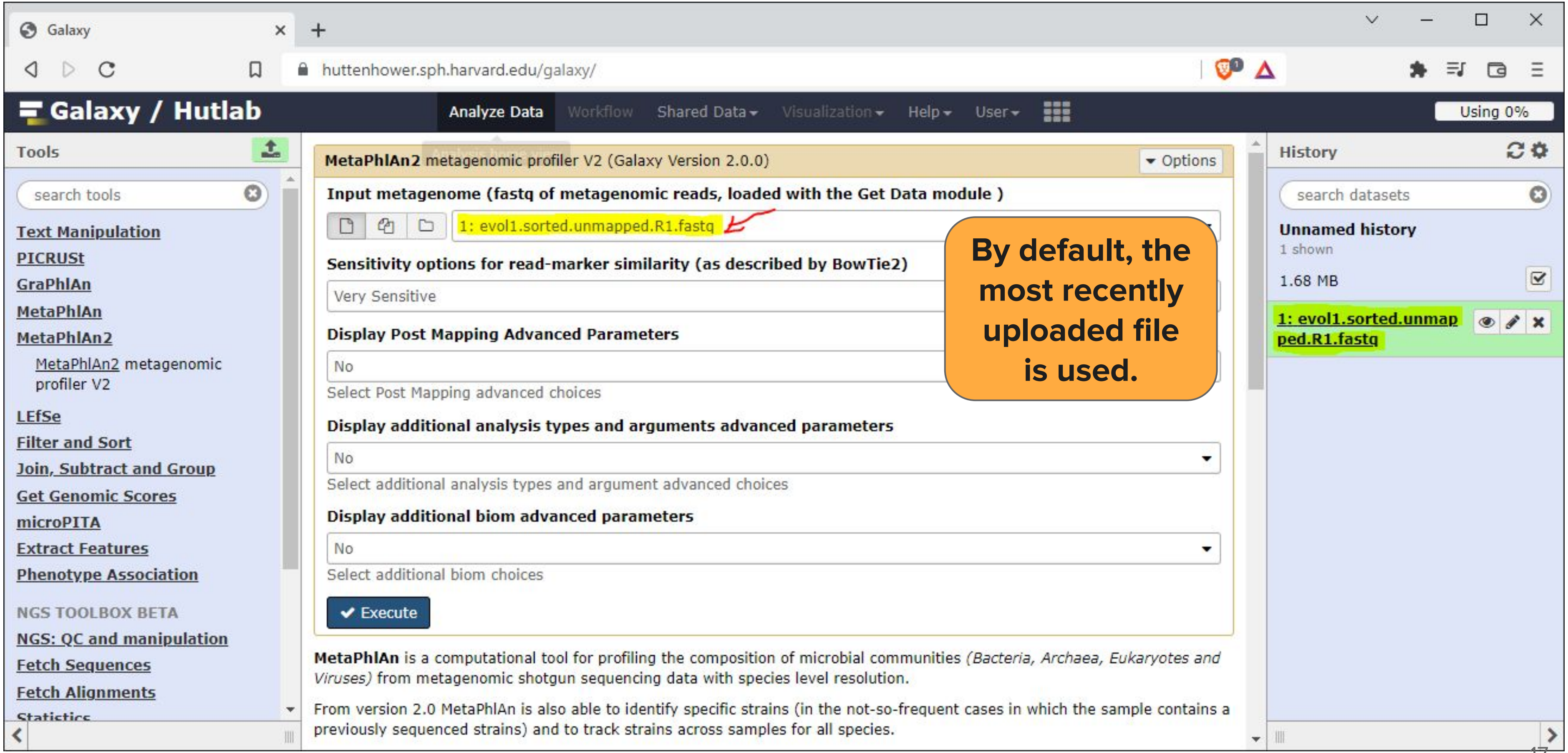
Display additional biom advanced parameters
No
Select additional biom choices

MetaPhlAn is a computational tool for profiling the composition of microbial communities *(Bacteria, Archaea, Eukaryotes and Viruses)* from metagenomic shotgun sequencing data with species level resolution.

From version 2.0 MetaPhlAn is also able to identify specific strains (in the not-so-frequent cases in which the sample contains a previously sequenced strains) and to track strains across samples for all species.

https://huttenhower.sph.harvard.edu/galaxy/

Make sure the correct input is set.



MetaPhlAn2 metagenomic profiler V2 (Galaxy Version 2.0.0)

Input metagenome (fastq of metagenomic reads, loaded with the Get Data module)

1: evol1.sorted.unmapped.R1.fastq

Sensitivity options for read-marker similarity (as described by BowTie2)

Very Sensitive

Display Post Mapping Advanced Parameters

No

Select Post Mapping advanced choices

Display additional analysis types and arguments advanced parameters

No

Select additional analysis types and argument advanced choices

Display additional biom advanced parameters

No

Select additional biom choices

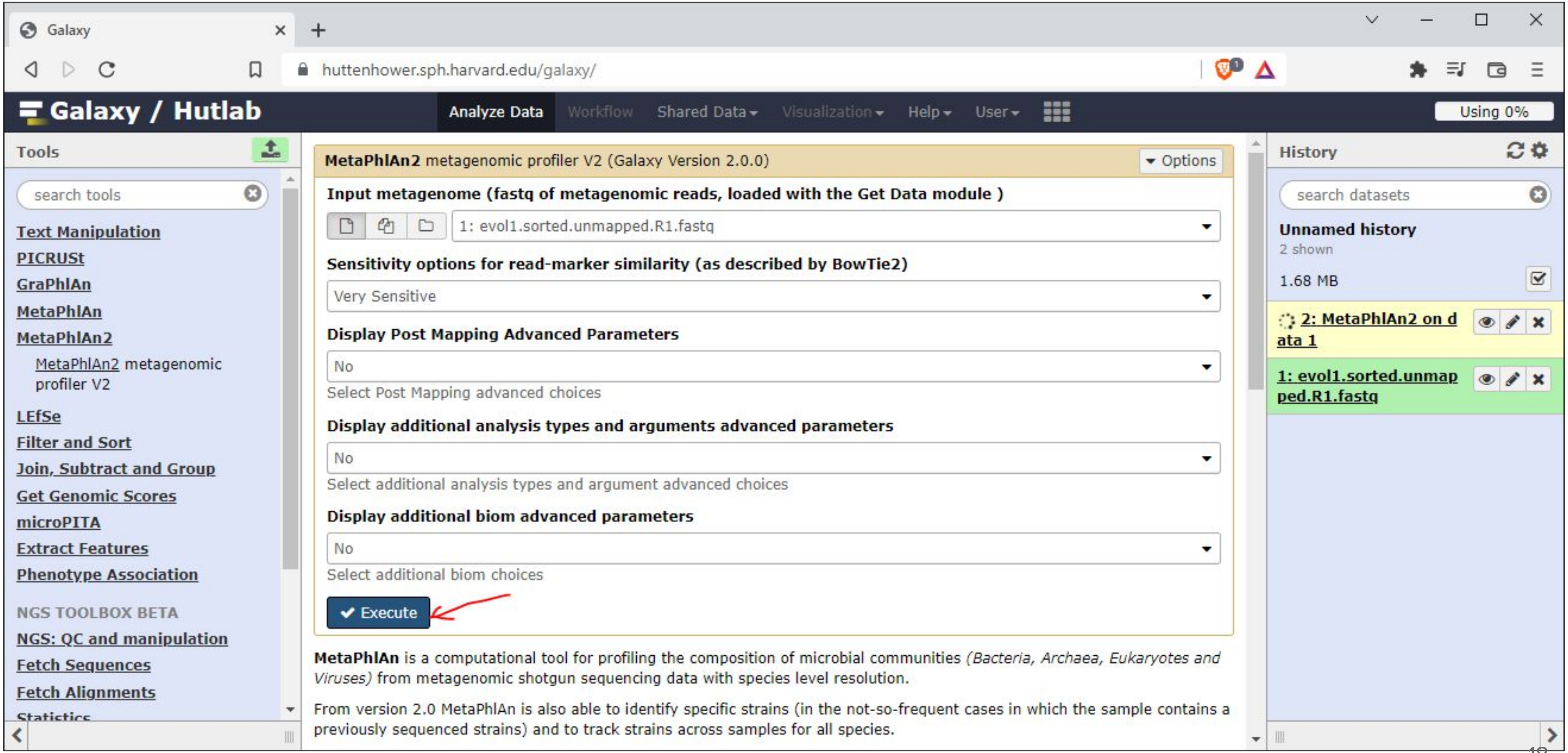
MetaPhlAn is a computational tool for profiling the composition of microbial communities *(Bacteria, Archaea, Eukaryotes and Viruses)* from metagenomic shotgun sequencing data with species level resolution.

From version 2.0 MetaPhlAn is also able to identify specific strains (in the not-so-frequent cases in which the sample contains a previously sequenced strains) and to track strains across samples for all species.

By default, the most recently uploaded file is used.

https://huttenhower.sph.harvard.edu/galaxy/

Keep default settings and click “Execute”.



MetaPhlAn2 metagenomic profiler V2 (Galaxy Version 2.0.0)

Input metagenome (fastq of metagenomic reads, loaded with the Get Data module)

1: evol1.sorted.unmapped.R1.fastq

Sensitivity options for read-marker similarity (as described by BowTie2)

Very Sensitive

Display Post Mapping Advanced Parameters

No

Select Post Mapping advanced choices

Display additional analysis types and arguments advanced parameters

No

Select additional analysis types and argument advanced choices

Display additional biom advanced parameters

No

Select additional biom choices

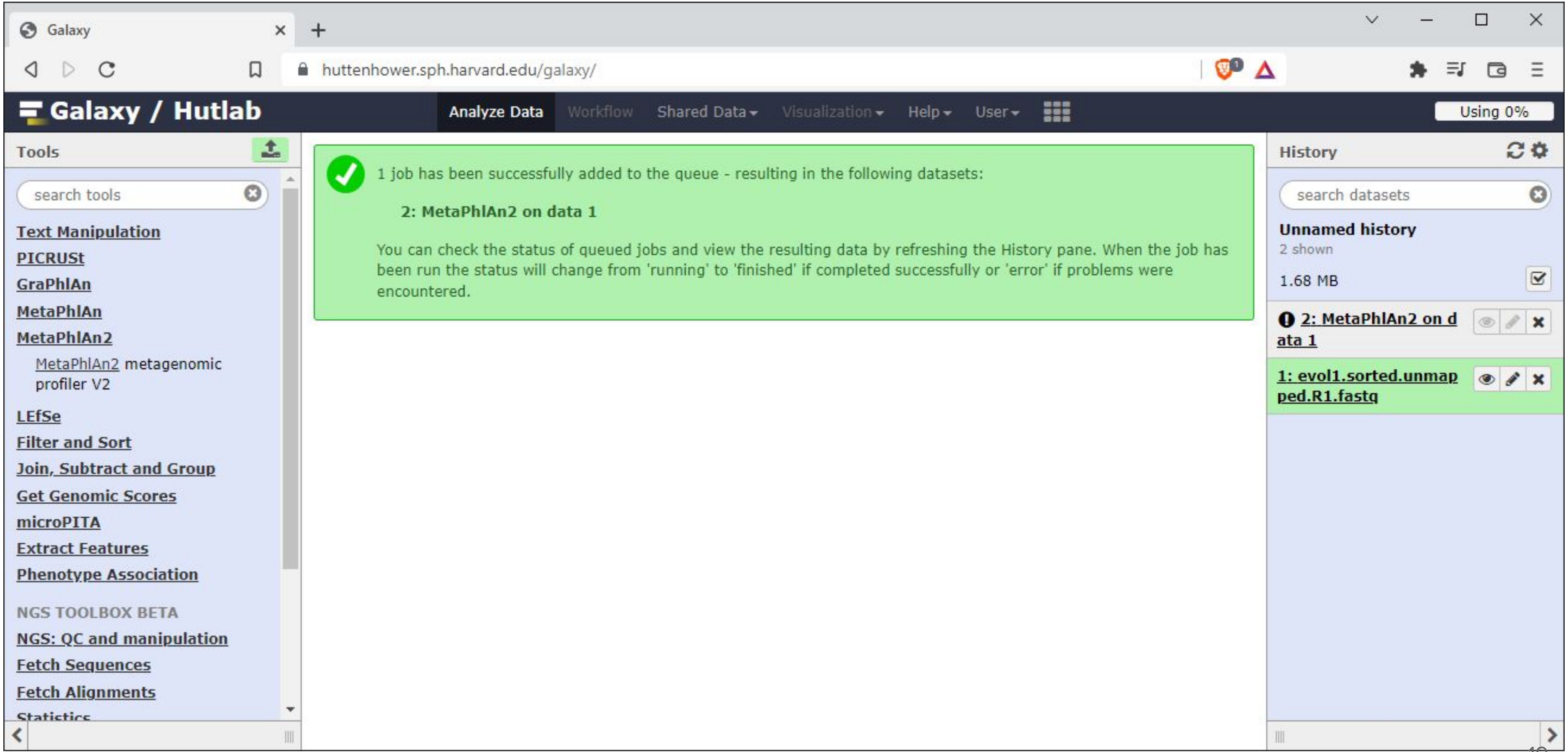
✔ Execute

MetaPhlAn is a computational tool for profiling the composition of microbial communities *(Bacteria, Archaea, Eukaryotes and Viruses)* from metagenomic shotgun sequencing data with species level resolution.

From version 2.0 MetaPhlAn is also able to identify specific strains (in the not-so-frequent cases in which the sample contains a previously sequenced strains) and to track strains across samples for all species.

https://huttenhower.sph.harvard.edu/galaxy/

Successful execution! Job is now added to the queue.



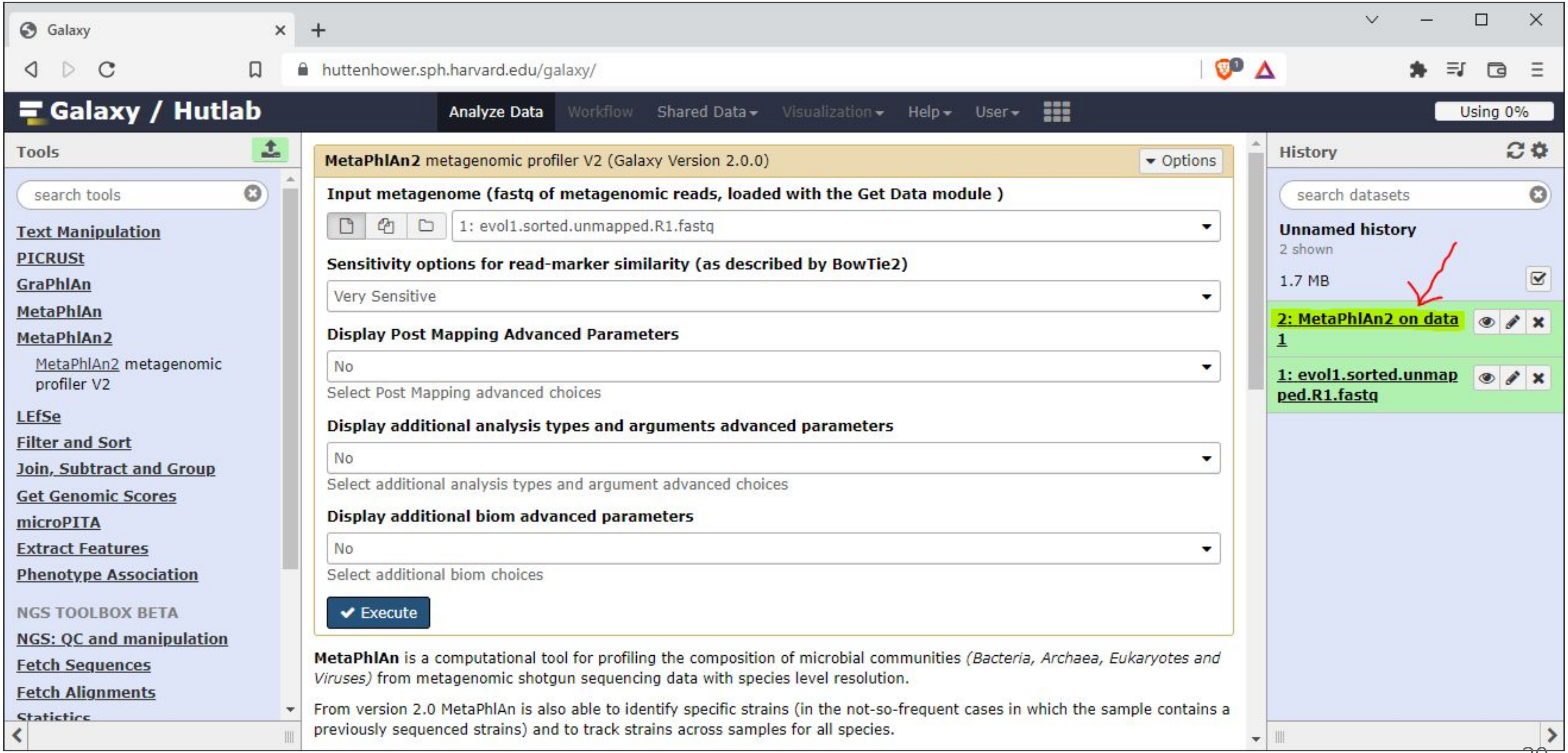
1 job has been successfully added to the queue - resulting in the following datasets:

2: MetaPhlAn2 on data 1

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

https://huttenhower.sph.harvard.edu/galaxy/

Result will appear green when it is completed.



MetaPhlAn2 metagenomic profiler V2 (Galaxy Version 2.0.0)

Input metagenome (fastq of metagenomic reads, loaded with the Get Data module)

1: evol1.sorted.unmapped.R1.fastq

Sensitivity options for read-marker similarity (as described by BowTie2)

Very Sensitive

Display Post Mapping Advanced Parameters

No

Select Post Mapping advanced choices

Display additional analysis types and arguments advanced parameters

No

Select additional analysis types and argument advanced choices

Display additional biom advanced parameters

No

Select additional biom choices

MetaPhlAn is a computational tool for profiling the composition of microbial communities *(Bacteria, Archaea, Eukaryotes and Viruses)* from metagenomic shotgun sequencing data with species level resolution.

From version 2.0 MetaPhlAn is also able to identify specific strains (in the not-so-frequent cases in which the sample contains a previously sequenced strains) and to track strains across samples for all species.

Unnamed history

2 shown

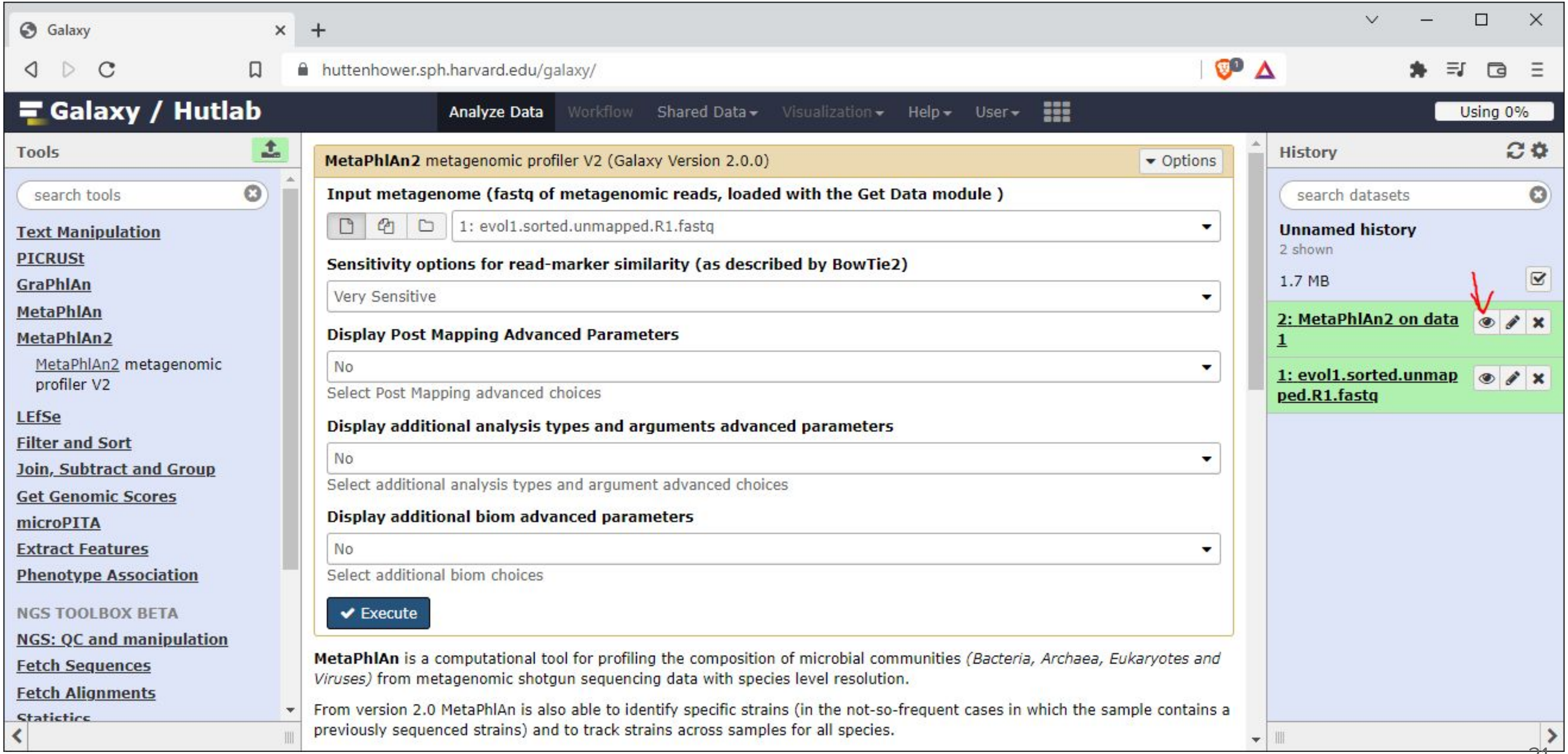
1.7 MB

2: MetaPhlAn2 on data 1

1: evol1.sorted.unmapped.R1.fastq

https://huttenhower.sph.harvard.edu/galaxy/

Click the eye icon to download and view result.



MetaPhlAn2 metagenomic profiler V2 (Galaxy Version 2.0.0)

Input metagenome (fastq of metagenomic reads, loaded with the Get Data module)

1: evol1.sorted.unmapped.R1.fastq

Sensitivity options for read-marker similarity (as described by BowTie2)

Very Sensitive

Display Post Mapping Advanced Parameters

No

Select Post Mapping advanced choices

Display additional analysis types and arguments advanced parameters

No

Select additional analysis types and argument advanced choices

Display additional biom advanced parameters

No

Select additional biom choices

MetaPhlAn is a computational tool for profiling the composition of microbial communities *(Bacteria, Archaea, Eukaryotes and Viruses)* from metagenomic shotgun sequencing data with species level resolution.

From version 2.0 MetaPhlAn is also able to identify specific strains (in the not-so-frequent cases in which the sample contains a previously sequenced strains) and to track strains across samples for all species.

Unnamed history

2 shown

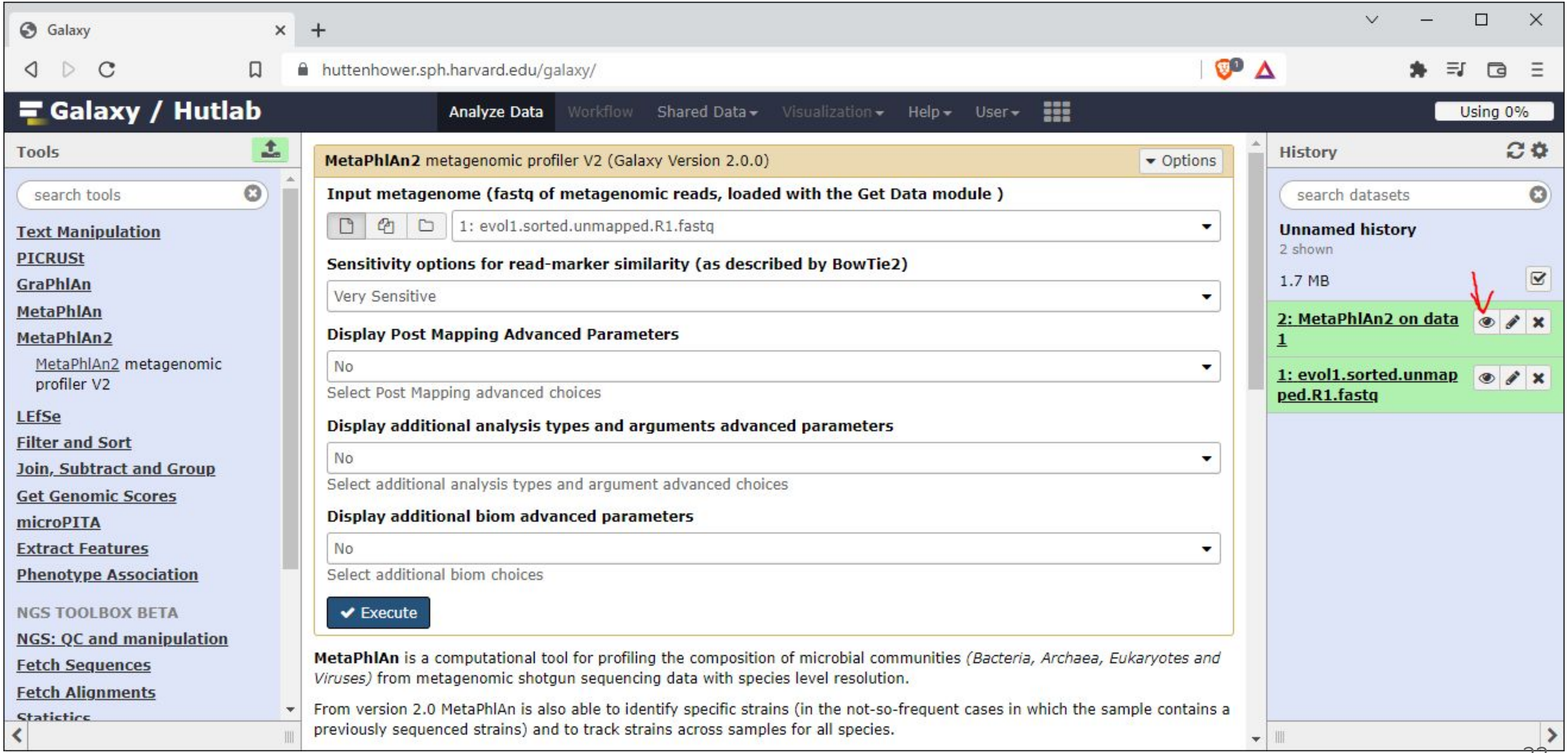
1.7 MB

2: MetaPhlAn2 on data 1

1: evol1.sorted.unmapped.R1.fastq

https://huttenhower.sph.harvard.edu/galaxy/

Click the eye icon to download and view result.



MetaPhlAn2 metagenomic profiler V2 (Galaxy Version 2.0.0)

Input metagenome (fastq of metagenomic reads, loaded with the Get Data module)

1: evol1.sorted.unmapped.R1.fastq

Sensitivity options for read-marker similarity (as described by BowTie2)

Very Sensitive

Display Post Mapping Advanced Parameters

No

Select Post Mapping advanced choices

Display additional analysis types and arguments advanced parameters

No

Select additional analysis types and argument advanced choices

Display additional biom advanced parameters

No

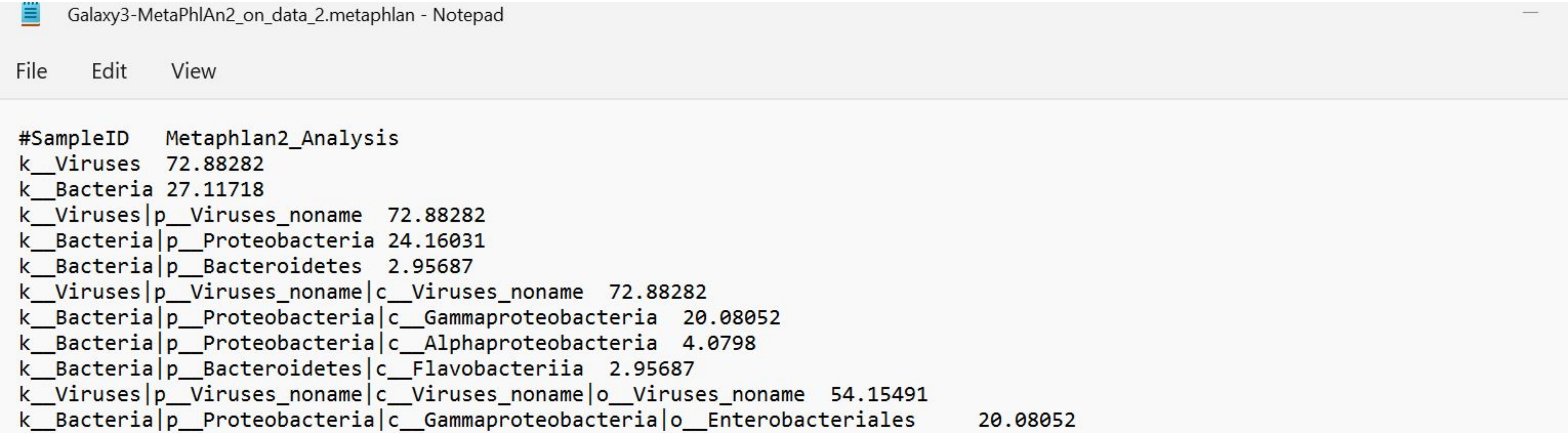
Select additional biom choices

MetaPhlAn is a computational tool for profiling the composition of microbial communities *(Bacteria, Archaea, Eukaryotes and Viruses)* from metagenomic shotgun sequencing data with species level resolution.

From version 2.0 MetaPhlAn is also able to identify specific strains (in the not-so-frequent cases in which the sample contains a previously sequenced strains) and to track strains across samples for all species.

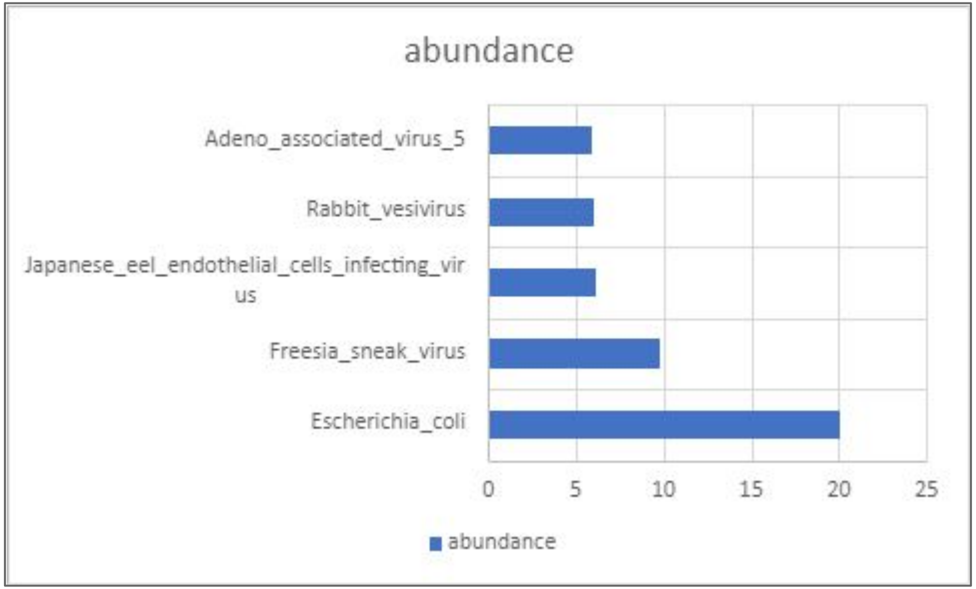
https://huttenhower.sph.harvard.edu/galaxy/

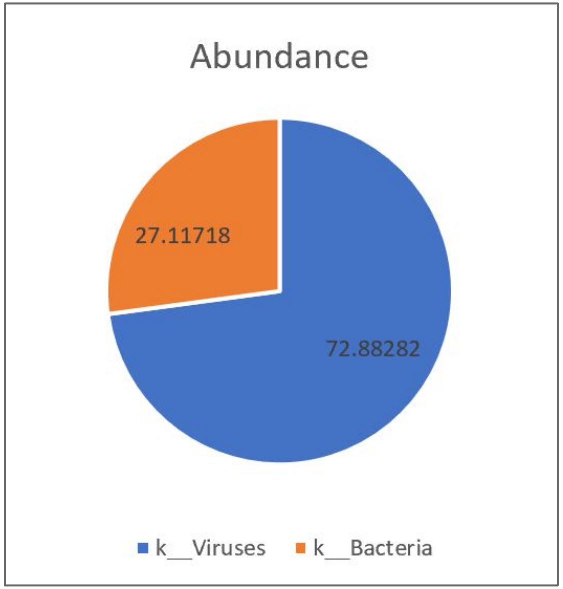
MetaphlAn Output Snippet



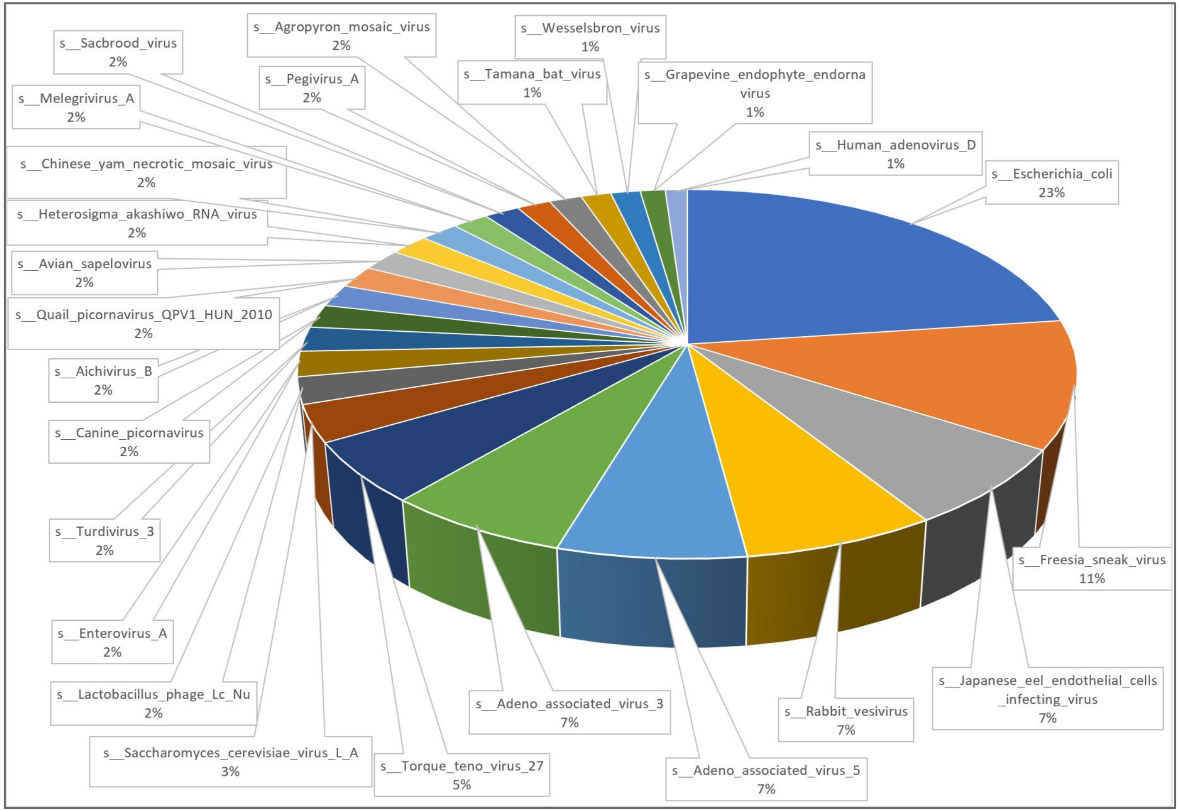
```
#SampleID	Metaphlan2_Analysis
k__Viruses	72.88282
k__Bacteria	27.11718
k__Viruses|p__Viruses_noname	72.88282
k__Bacteria|p__Proteobacteria	24.16031
k__Bacteria|p__Bacteroidetes	2.95687
k__Viruses|p__Viruses_noname|c__Viruses_noname	72.88282
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria	20.08052
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria	4.0798
k__Bacteria|p__Bacteroidetes|c__Flavobacteriia	2.95687
k__Viruses|p__Viruses_noname|c__Viruses_noname|o__Viruses_noname	54.15491
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacteriales	20.08052
```

Visualized Output





Species Detected



References

[1] “biobakery/biobakery_workflows.” n.d. GitHub. Accessed May 20, 2022. https://github.com/biobakery/biobakery_workflows.

[2] Borenstein Lab. n.d. “MetaPhlAn.” Accessed May 20, 2022. PowerPoint. http://borensteinlab.com/courses/TAU_CS_3116_B_19/presentations/7_MetaPhlan.pdf.

[3] The Huttenhower Lab. n.d. “biobakeryWorkflows.” The Huttenhower Lab. Accessed May 20, 2022. https://huttenhower.sph.harvard.edu/biobakery_workflows/.

[4] The Huttenhower Lab. n.d. “Galaxy / Hutlab.” Accessed May 20, 2022. https://huttenhower.sph.harvard.edu/galaxy/.

[5] The Huttenhower Lab. n.d. “MetaPhlAn2.” The Huttenhower Lab. Accessed May 20, 2022. https://huttenhower.sph.harvard.edu/metaphlan2/.

[6] Mehta, Subina, Pratik Jagtap, and Saskia Hiltemann. 2021. “Introduction to metatranscriptomics.” Galaxy Training! https://training.galaxyproject.org/archive/2021-10-01/topics/metagenomics/tutorials/metatranscriptomics/slides-plain.html.

Questions?