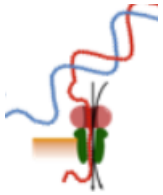
Understanding health and wellness through microbiome research

WGS analysis

Designing a metagenome project and Sample collection

- Sample collection
- Metadata collection

- Collect the samples in sterilized tubes and transport the laboratory in dry ice immediately after collection, else keep it stored in -20 and then transfer in dry ice
- Human Samples: Fill out a detailed questionnaire on the metadata and collect all relevant information



- DNA extraction
- sequencing

- Use the appropriate DNA extraction kit and remove the debris or eukaryotic cells before extraction

Saliva samples

Using Saliva collector

Skin/scalp/ oral samples

Using cotton swabs soaked in solutions

Fecal samples

Using sterile collector

Metadata

	Clinical data			Meat in g/week			Fish	Vegetables	Fruits
Sample ID	State	Age (yrs)	BMI	Red meat (pork, beef, veal, venison)	White meat (chicken, turkey)	Total (all meat, i.e. red, white and other)	g/week	g/week	g/week
31766	controls	73	32	137	137	274	200	1400	1400
31537	controls	70	31	137	137	274	200	1400	3500
31600	controls	70	30	69	69	137	200	1400	100
31428	controls	69	30	0	0	0	0	1400	100
530167	controls	72	30	274	0	274	0	1400	1400
530315	controls	75	31	69	69	137	200	1400	1000
530050	controls	71	29	137	69	206	0	1000	1000

Sample workflow for metagenomic sample sequencing

Sample collection

Transportation

Storage

Whole Metagenomic Sequencing

No PCR amplification of specific genes

Metagenomic DNA extraction

Library Preparation

16S rRNA Sequencing

16S rDNA Amplification

Indexing samples for Multiplexing

Sequencing

Data analysis

Interpretation of Results

Understanding the format raw sequencing reads

FASTQ file

What is a FASTQ File?

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*((((***+))%%%++)(%%%%).1***-+*''))**55CCF>>>>>>CCCCCCC65
```

Sequence header

Separator b/w sequence and quality scores

Sequence

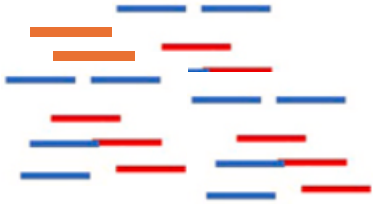
Quality scores in ASCII characters

Each character represents a numerical value: a so-called Phred score, encoded via a single letter encoding.

```
!"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHI
|    |    |    |    |    |    |    |    |
0....5...10...15...20...25...30...35...40
|    |    |    |    |    |    |    |    |
worst.............................best
```

From Sequencing Reads to Biological Insights

Objective: To process raw reads and extract meaningful biological information using a structured metagenomic pipeline



📥 **Starting Point:**

•FASTQ files

(From sequencer or public database)

🎯 **What to do: Extract All Possible Biological Conclusions, such as:**

- 🦠 **Which microbes are present?** (Taxonomy)
- ⚙️ **What are they doing?** (Functions/pathways)
- 💊 **Are there antibiotic resistance genes?**
- 🧬 **Any virulence genes or mobile elements?**
- 🌱 **How diverse is the microbial community?**
- 🔄 **How do samples differ from each other?**
- 🧠 **What can we infer about health/disease?**

🛠️ **How to do?**

"Step by step, We will understand about various tools to reach these conclusions"

How Do We Analyze WGS Data?

There are **two main ways** to analyze metagenome sequencing data:

🌐 1. Web-Based Tools (GUI)

- No coding needed
- Beginner-friendly interface
- Ideal for **exploration, learning, and quick analysis**
- Examples: Galaxy, MG-RAST, KBase

🖥️ 2. Command-Line Tools (CLI)

- More flexible and powerful
- Essential for **large-scale** and **customized analyses**
- Runs on local or server environments (Linux)

We will start by learning tools using **web platforms** to get familiar with the concepts.

Starting with Galaxy: A GUI-Based Microbiome Analysis Platform

- Open-source, web-based platform for biomedical data analysis
- No command-line experience needed
- Integrated tools for **quality control**, **taxonomic profiling**, **functional annotation**, etc.

🌎 Access

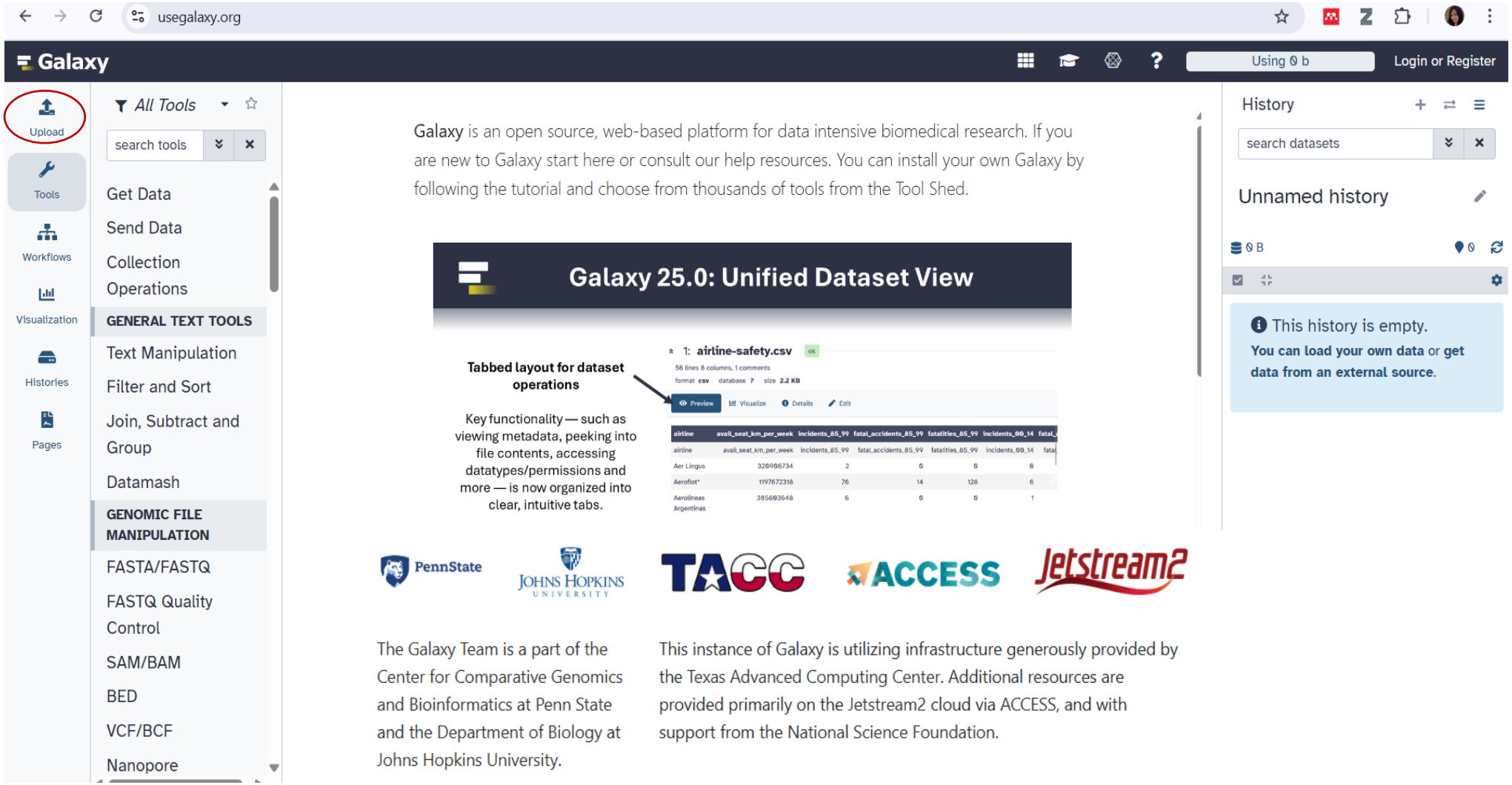
- Public servers: usegalaxy.org

The Galaxy Interface

- Three main panels

  Left: Available Tools,

  Middle: View your data and run tools, **Right:** Full record of your analysis **history**



How to find publicly available data

Listen

Research Article

Landscape of flavonoid metabolism in human gut microbiome and its association with health and disease

Deepika Pateriya , Vishnu Prasoodanan P. K. , Joy Scaria & **Vineet K. Sharma**

Article: 2520788 | Received 24 Feb 2025, Accepted 05 Jun 2025, Published online: 25 Jun 2025

Cite this article https://doi.org/10.1080/29933935.2025.2520788 Check for updates

Full Article | Figures & data | References | Supplemental | Citations | Metrics | © Licensing | Reprints & Permissions

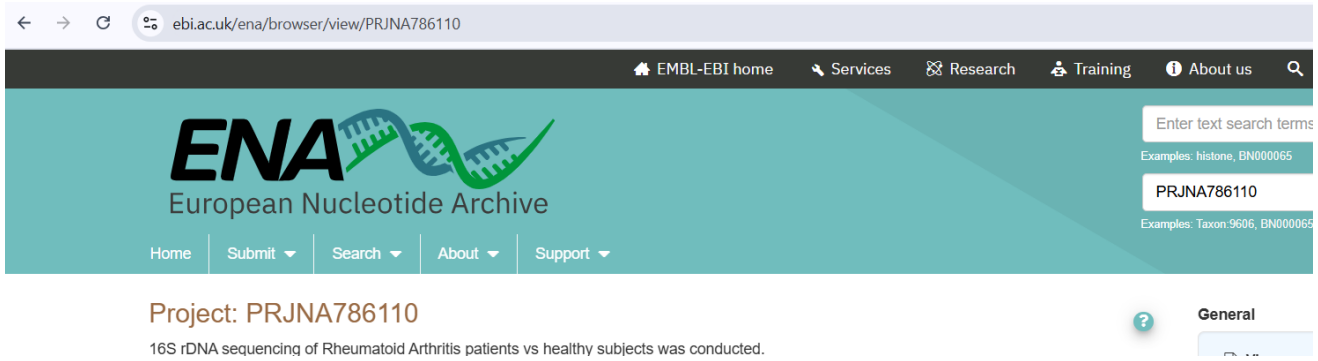
In this article

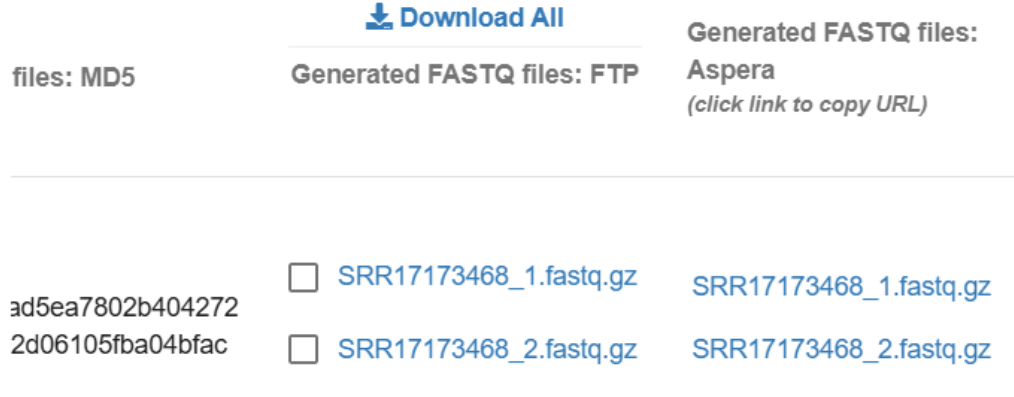
- Results
- Discussion
- Materials and methods

Data availability statement

Metagenomic data used in this paper are publicly available under the following BioProject IDs: PRJNA400072, PRJNA421881, PRJNA531273, PRJEB7774, PRJEB10878, PRJNA268964, PRJNA278393, PRJNA485056, and PRJNA397112.

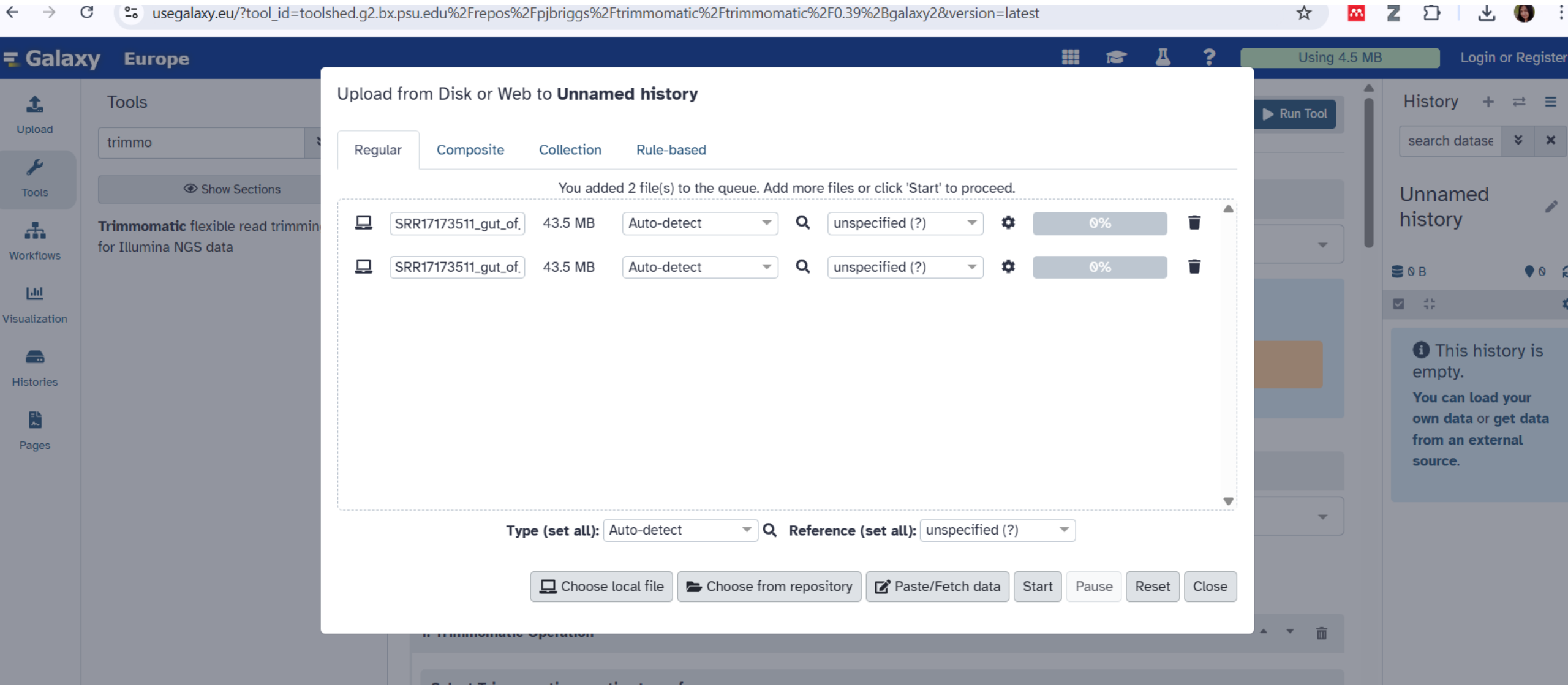
https://www.ebi.ac.uk/ena/browser/home



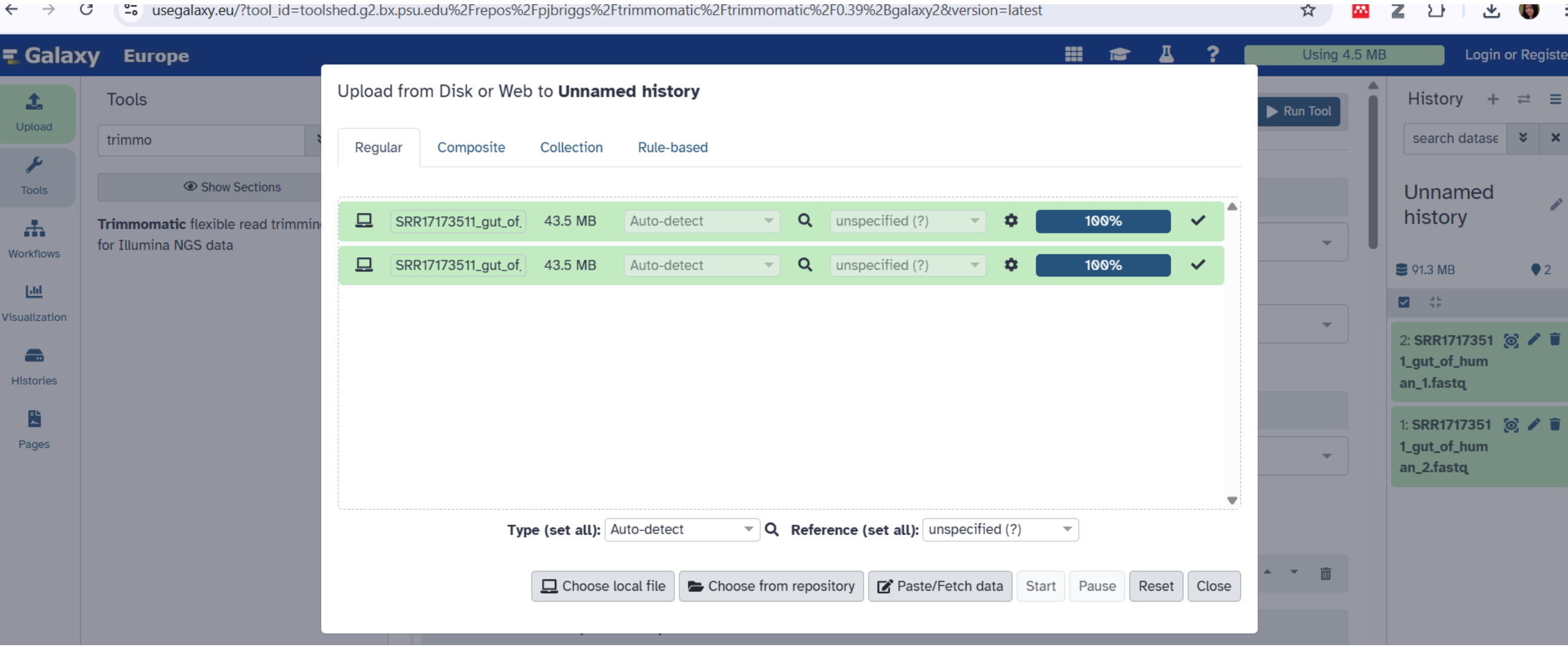


Analysing raw sequence data using Galaxy microbiome analysis platform

- Importing raw sequence data
- Paired-end reads in this case



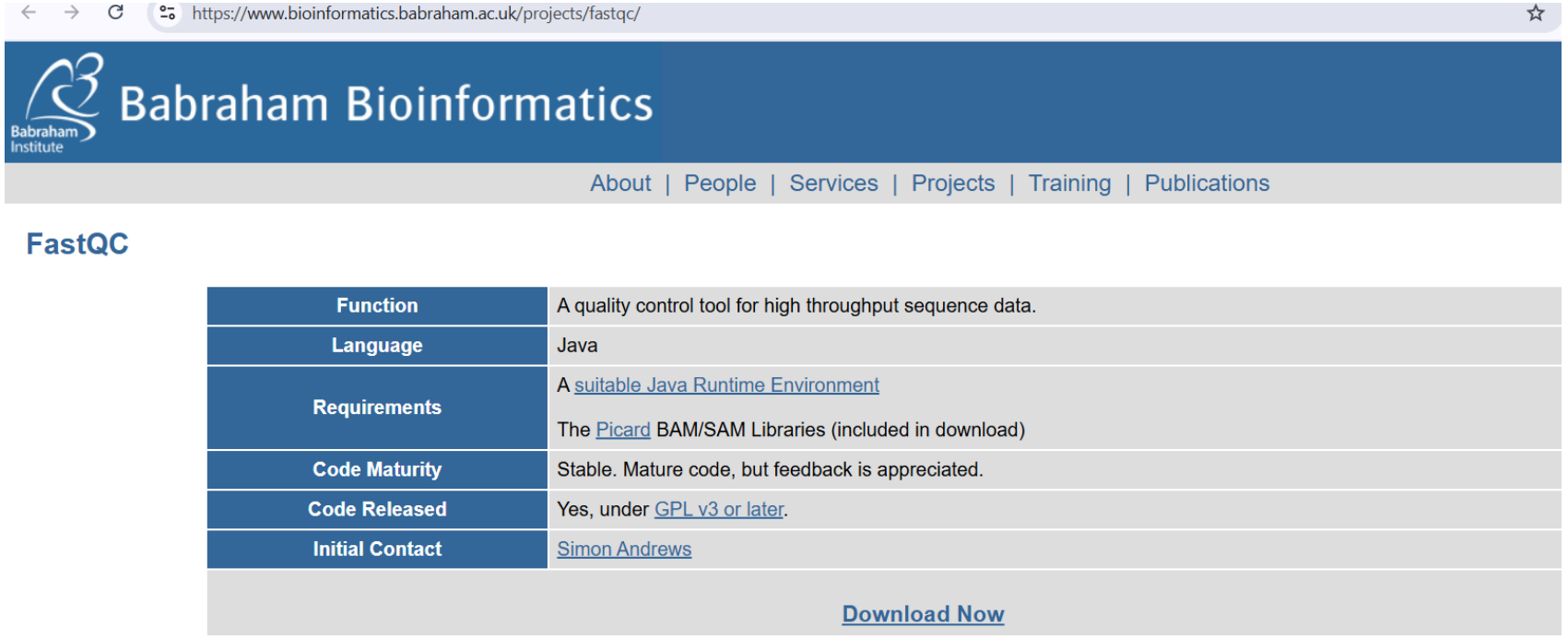
- Data import successful
- Start analysis with data cleaning



Objective: Check the quality of raw reads before any downstream analysis

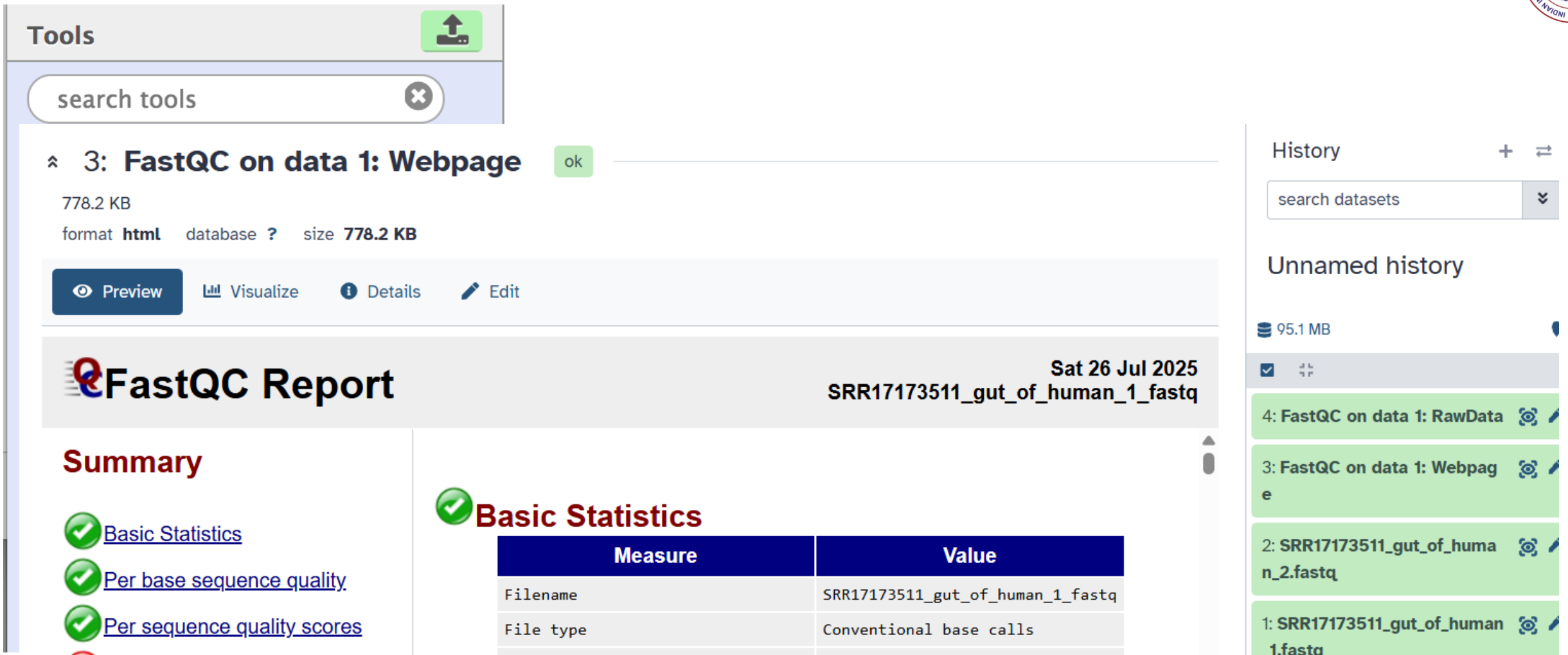
FastQC: A quality control tool for high throughput sequence data

This tool provides an easy way to perform a quality control check on sequence data coming from high-throughput sequencing pipelines

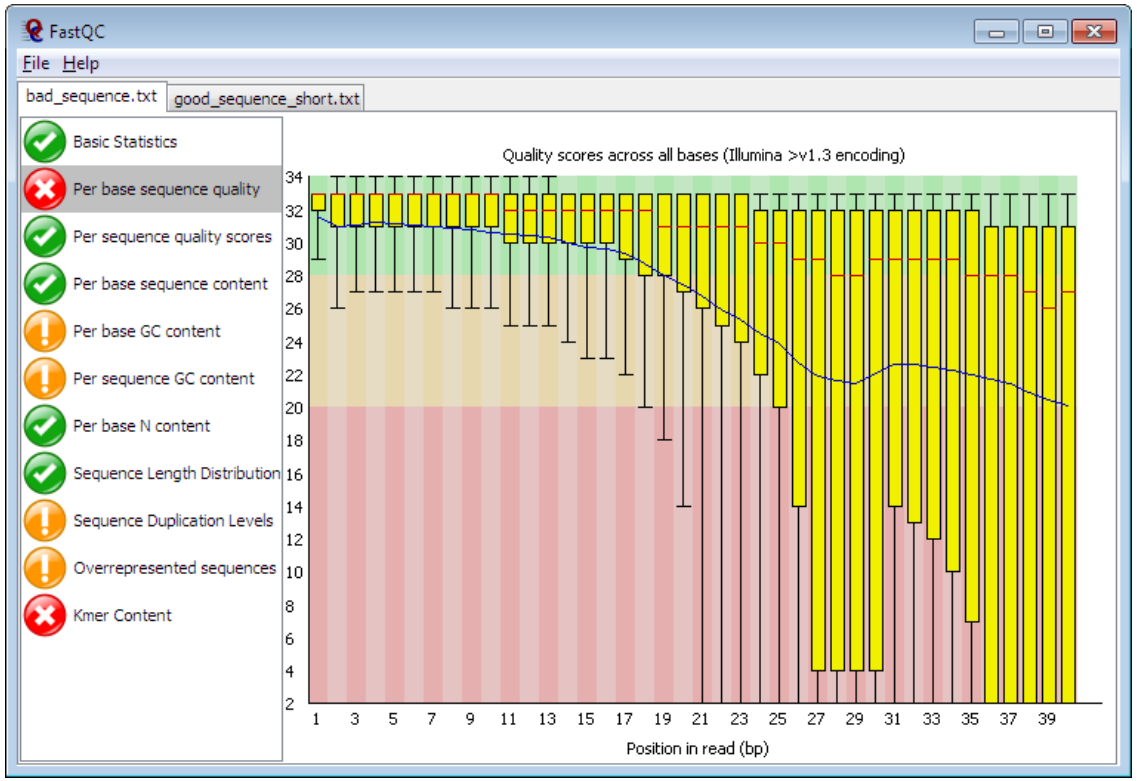


Visit: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Type FastQC in the search box on top left hand corner

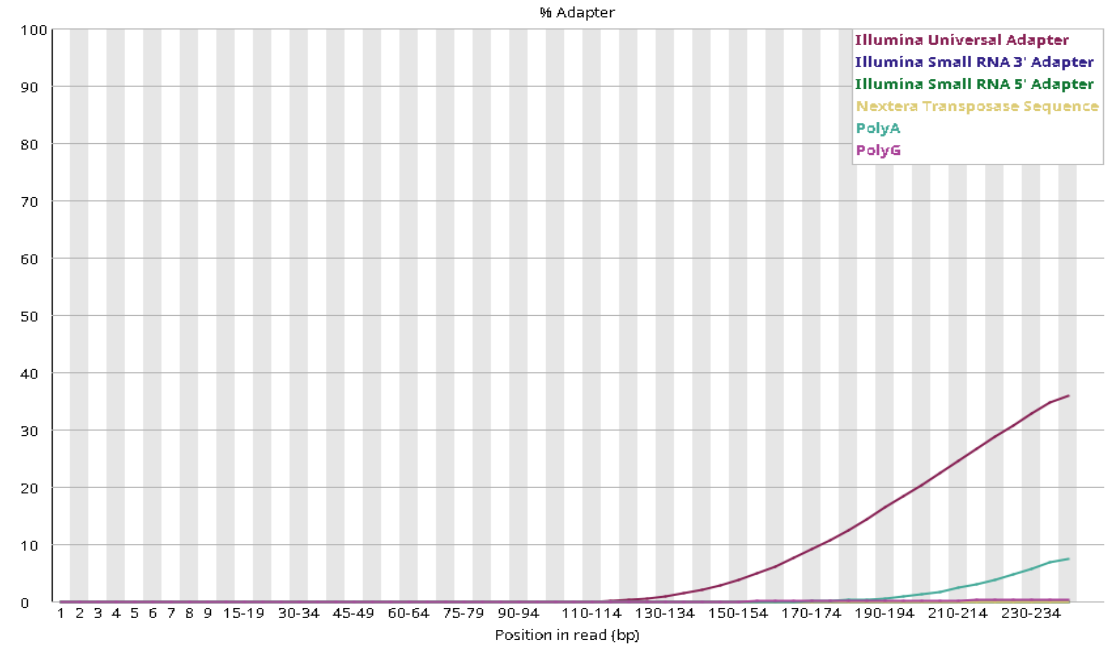


Quality control checks on raw sequence data (FastQC)

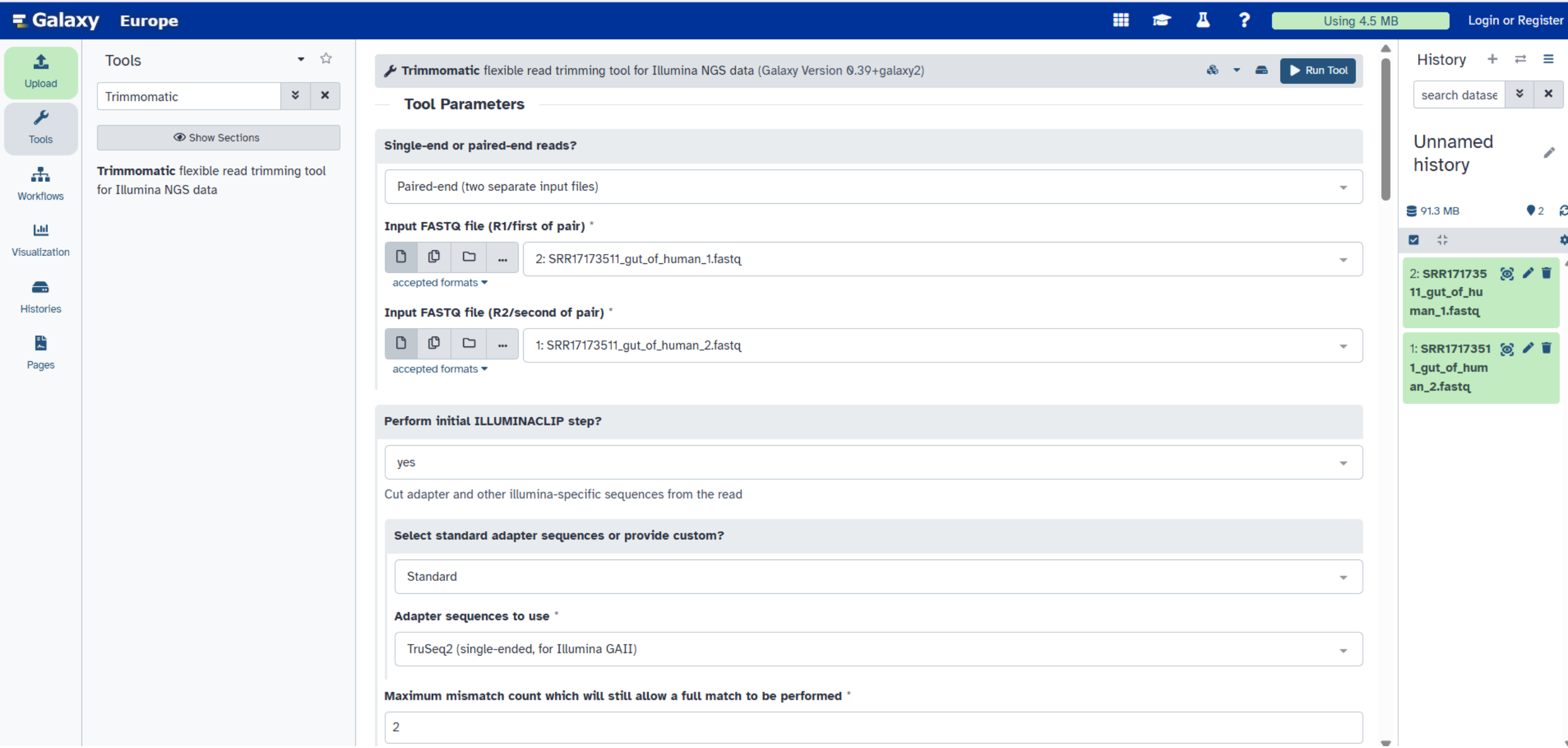


- During library preparation, synthetic adapters are ligated onto DNA fragments to enable sequencing
- If these adapter sequences are not fully removed, they will remain in our reads as technical artifacts

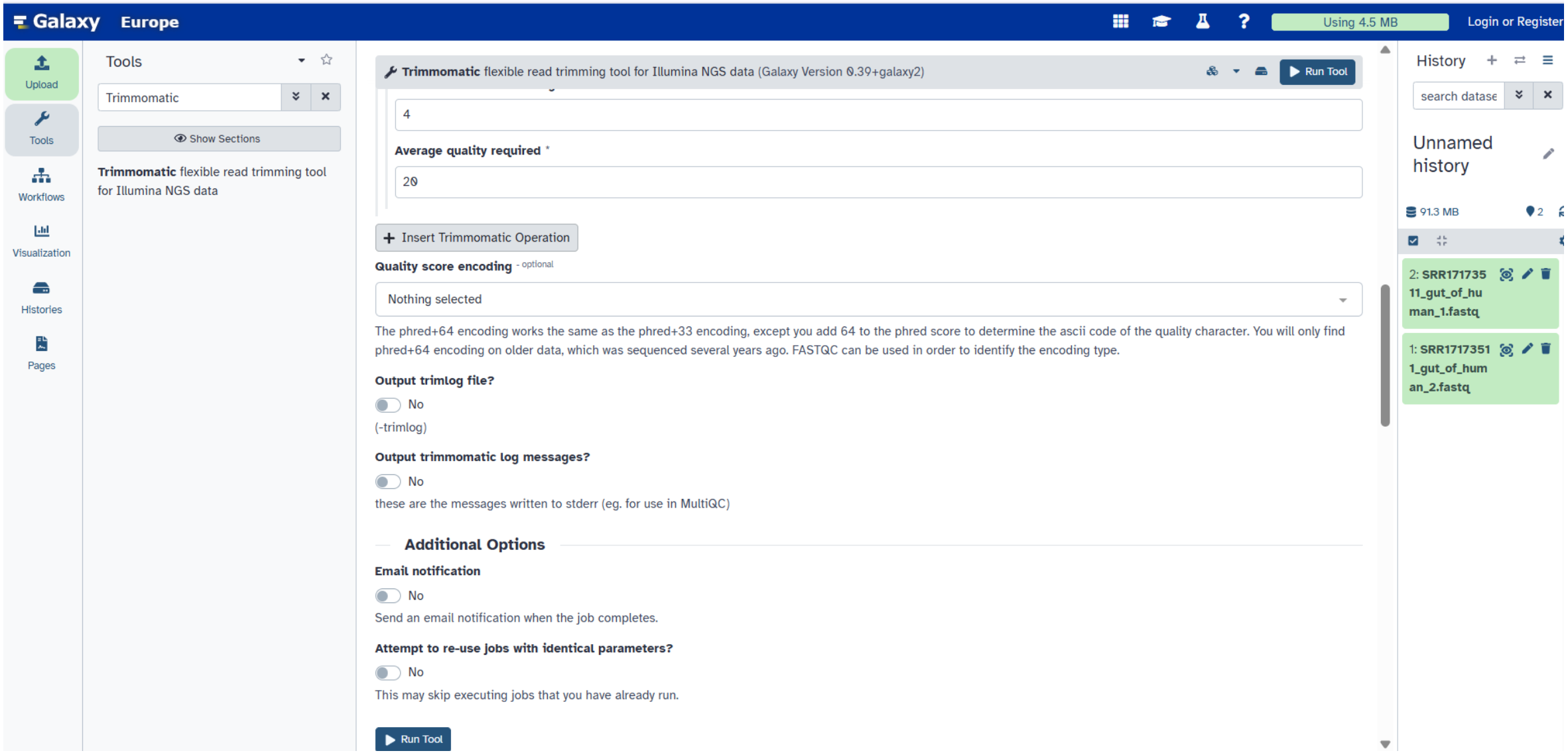
Adapter Content



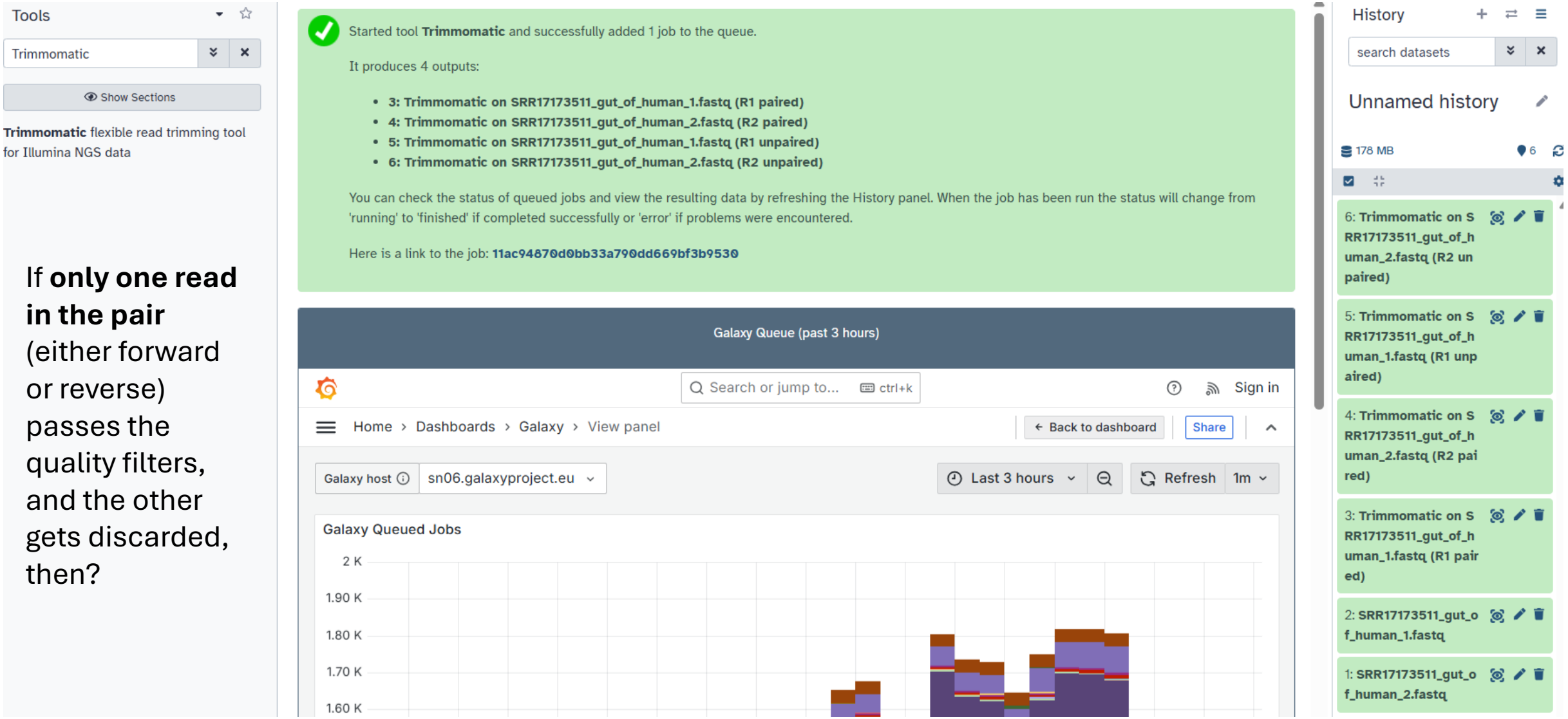
Running Trimmomatic using galaxy platform



Running Trimmomatic: Setting various parameters

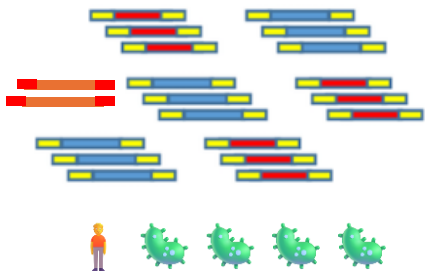


Running Trimmomatic: Job submitted, Job completed

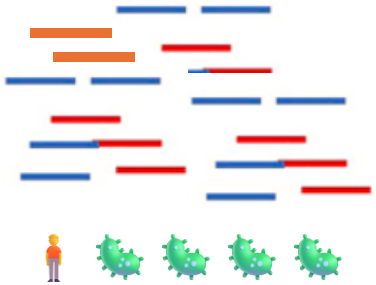


If **only one read in the pair** (either forward or reverse) passes the quality filters, and the other gets discarded, then?

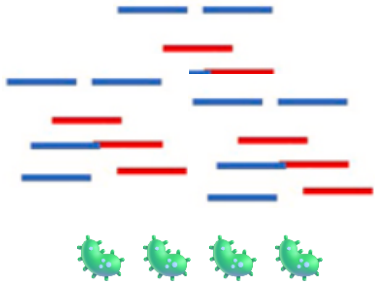
From FASTQ to Good Quality Reads — But Wait... There is Human DNA!



Raw read fragments with adapters



After Adapter and Quality Trimming : Adapter ends are removed, But still a mix of microbial and host (human) DNA



we need these reads to focus solely on microbial content (clean reads)

🔬 **Objective**

•To **remove non-microbial (host) DNA, human DNA** in human gut microbiome studies.

•Ensures **downstream analyses** (e.g., taxonomic profiling, assembly) are **not biased** by host sequences.

Common Tools

•**Bowtie2** (most widely used)

  •Aligns reads to human genome (e.g., hg38)

•**BWA** (Burrows-Wheeler Aligner)

  •Alternative for alignment

•**KneadData:** Wrapper around **Bowtie2**; automates trimming + host removal

Removing host DNA contamination using Bowtie2

nature › nature methods › brief communications › article

Brief Communication | Published: 04 March 2012

Fast gapped-read alignment with Bowtie 2

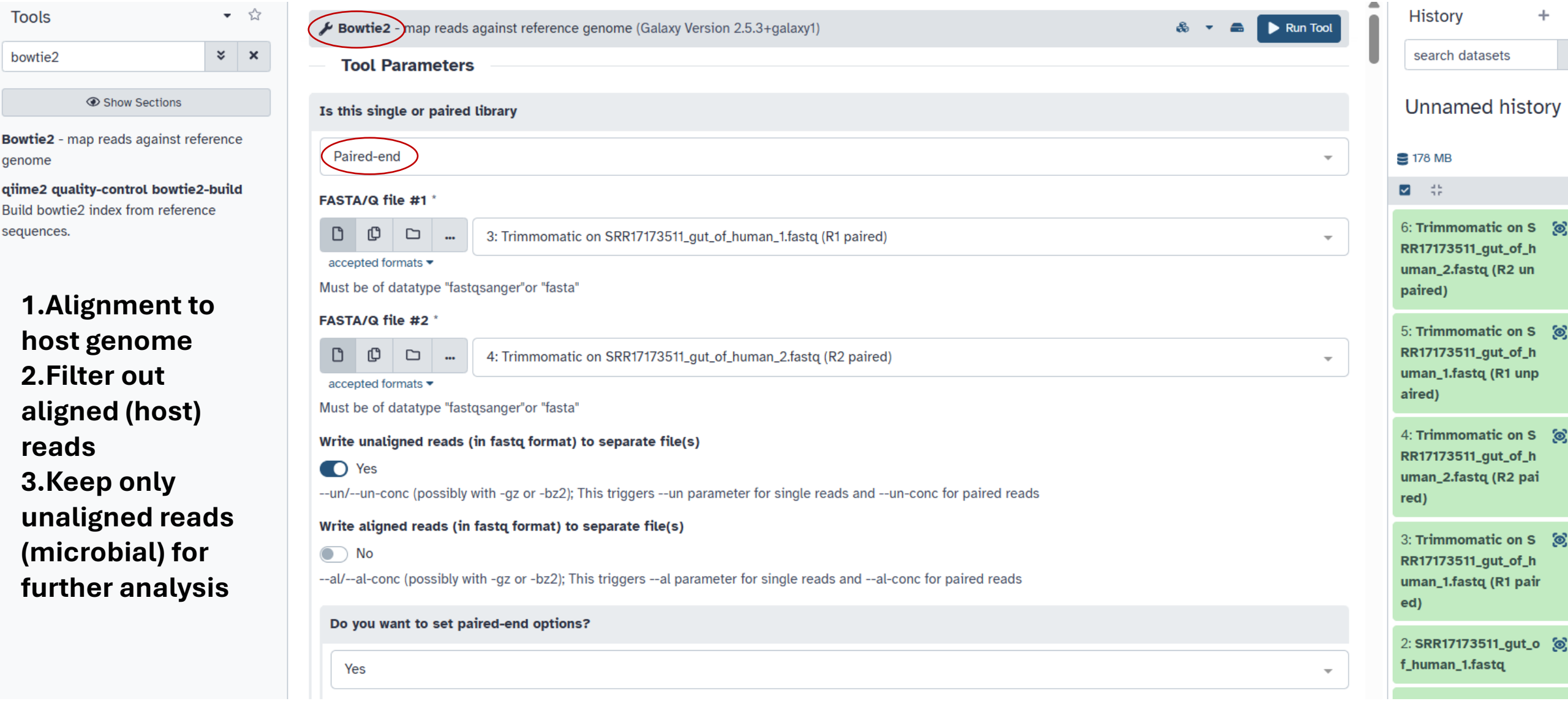
Ben Langmead & Steven L Salzberg

Nature Methods **9**, 357–359 (2012) | Cite this article

109k Accesses | **48k** Citations | **176** Altmetric | Metrics

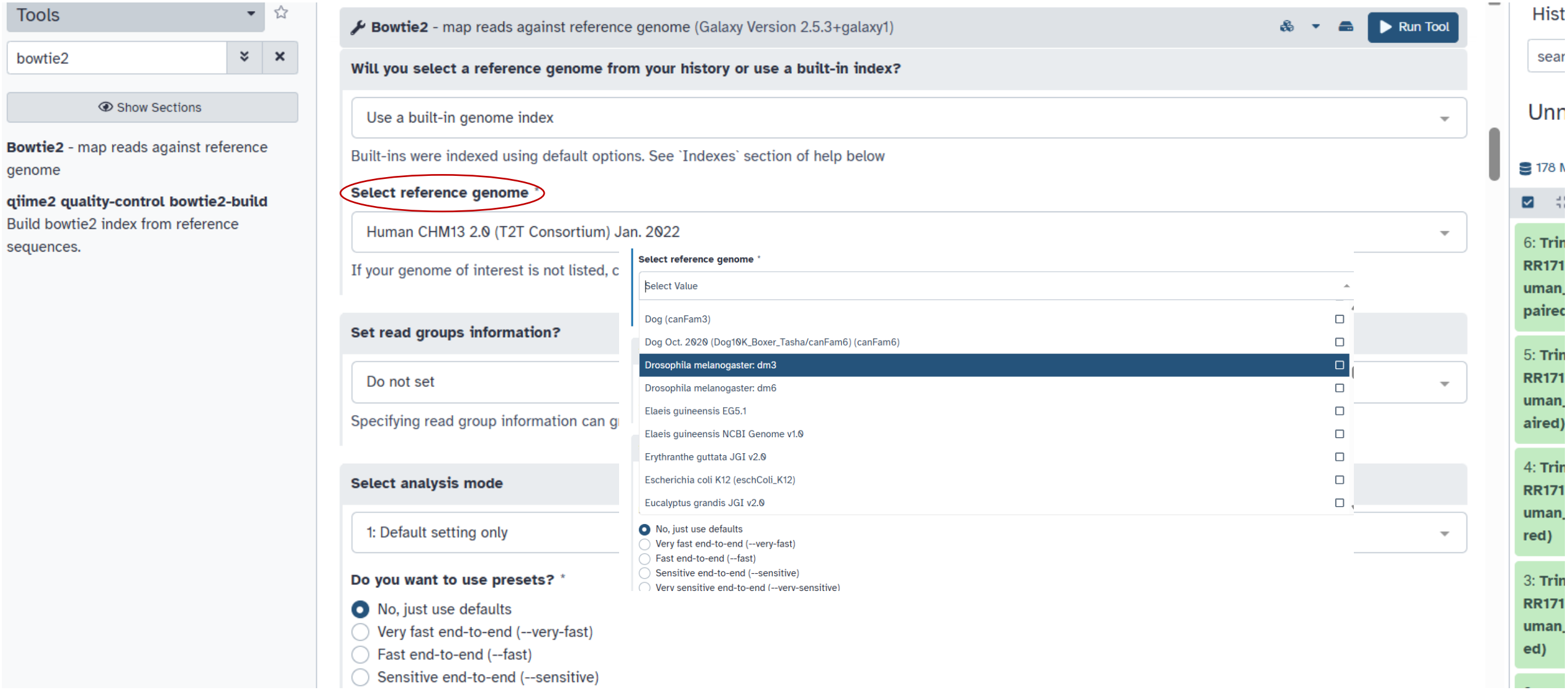
- **How it works**: Uses fast and memory-efficient alignment to map sequencing reads to the host genome.
- **Input**: Quality-trimmed FASTQ files + host genome index.
- **Output**:
- Aligned reads (host-contaminated)
- **Unaligned reads** (microbial reads kept for downstream analysis)

Removing host DNA contamination using Bowtie2 in Galaxy



1.Alignment to host genome
2.Filter out aligned (host) reads
3.Keep only unaligned reads (microbial) for further analysis

Removing host DNA contamination: Selecting reference genomes



Bowtie2 running.....

Tools

bowtie2

Show Sections

Bowtie2 - map reads against reference genome

qiime2 quality-control bowtie2-build
Build bowtie2 index from reference sequences.

Started tool **Bowtie2** and successfully added 1 job to the queue.

It produces 3 outputs:

- **7: Bowtie2 on data 4 and data 3: unaligned reads (L)**
- **8: Bowtie2 on data 4 and data 3: unaligned reads (R)**
- **9: Bowtie2 on data 4 and data 3: alignments**

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.

Here is a link to the job: **11ac94870d0bb33aa803271fdf9d65f2**

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 Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de.NBI Freiburg.

History

search dat

Unname

178 MB

9: Bowti
a 4 and dat
ments

8: Bowti
a 4 and dat
ned reads (

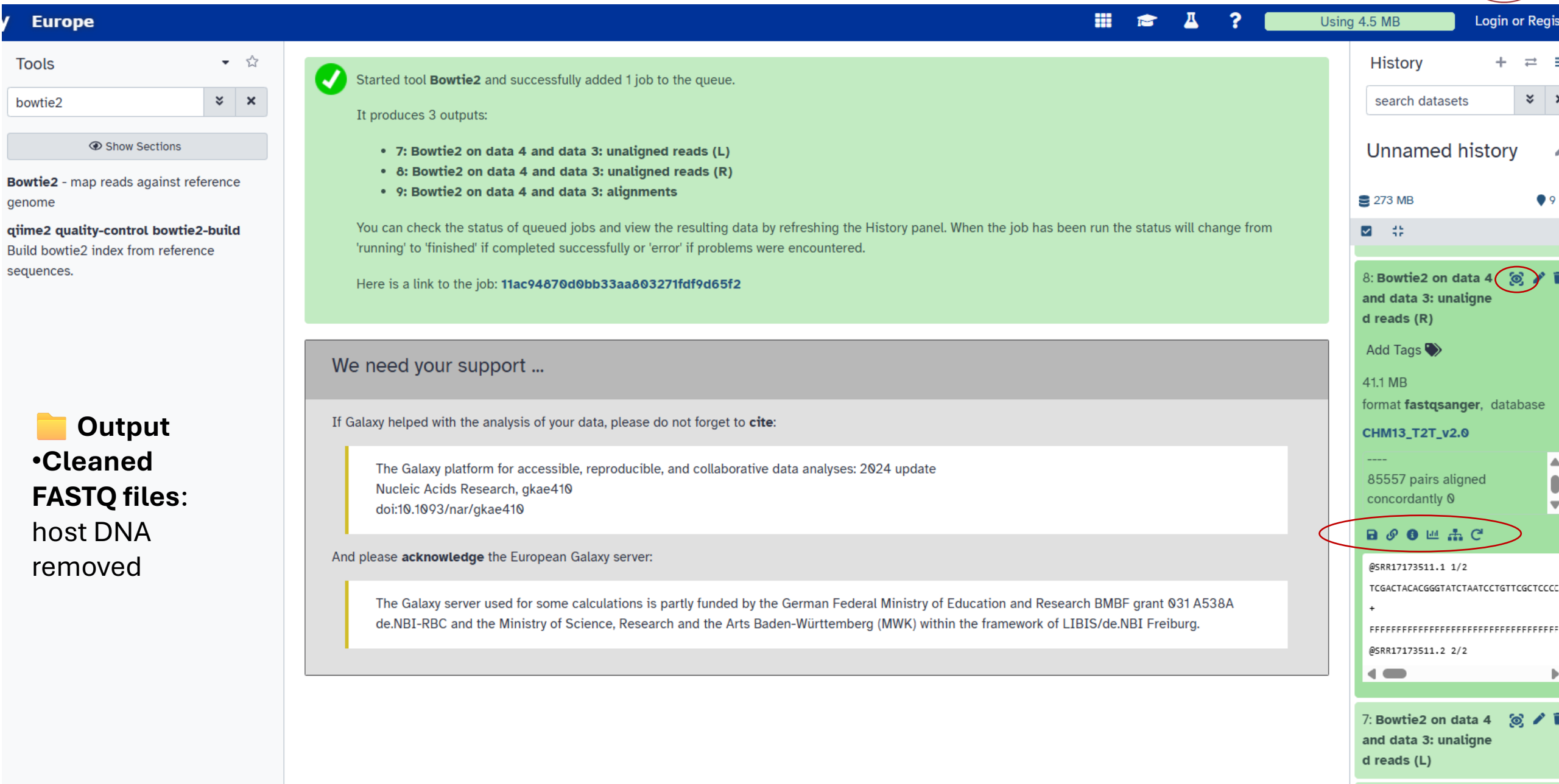
7: Bowti
4 and data
ed reads (L

6: Trimmom
RR17173511
uman_2.fast
paired)

5: Trimmom
RR17173511
uman_1.fast
aired)

4: Trimmom

Saving processed files



📁 **Output**
•Cleaned FASTQ files: host DNA removed

Running Bowtie2 using CLI

 # download ready to use bowtie2 database of human host genome GRCh38 (hg38)

```
wget https://genome-idx.s3.amazonaws.com/bt/GRCh38_noalt_as.zip
unzip GRCh38_noalt_as.zip
```

 # run bowtie2 mapping

 # using --un-conc-gz to get gzip compressed output files; 8 processors)

 # use alignment mode "local" especially in case raw-reads are not adapter/quality trimmed

```
bowtie2 -p 8 -x GRCh38_noalt_as \
  -1 SAMPLE_R1.fastq.gz \
  -2 SAMPLE_R2.fastq.gz \
  --very-sensitive-local \
  --un-conc-gz \
  SAMPLE_host_removed \
  > SAMPLE_mapped_and_unmapped.sam
```

http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml#output-options

 # bowtie2 results (gz files without gz ending)

```
ls
```

```
SAMPLE_host_removed.1
SAMPLE_host_removed.2
...
```

🧮 Track Your Data at Every Step!

Seqkit: A fast, easy-to-use toolkit to inspect your FASTQ/FASTA files

SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation

Wei Shen, Shuai Le, Yan Li ✉, Fuquan Hu ✉

Published: October 5, 2016 • https://doi.org/10.1371/journal.pone.0163962

https://doi.org/10.1371/journal.pone.0163962

Read statistics after each step

Raw reads

file	format	type	num_seqs	sum_len	min_len	avg_len	max_len
R1_001.fastq.gz	FASTQ	DNA	112,252,406	16,837,860,900	150	150	150
R2_001.fastq.gz	FASTQ	DNA	112,252,406	16,837,860,900	150	150	150

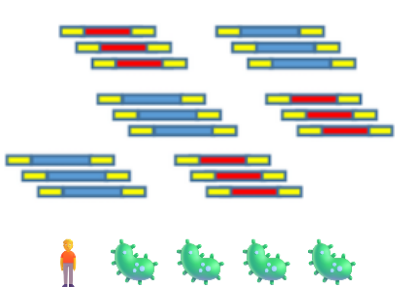
Trimmed reads

file	format	type	num_seqs	sum_len	min_len	avg_len	max_len
R1_paired.fastq.gz	FASTQ	DNA	111,854,222	16,218,290,876	50	145.0	150
R2_paired.fastq.gz	FASTQ	DNA	111,854,222	162,130,845,088	50	144.9	150

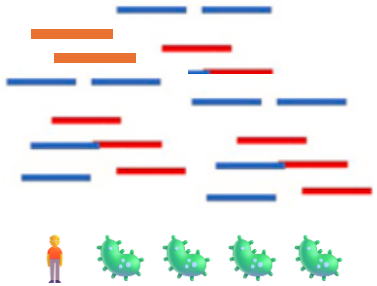
Clean reads

file	format	type	num_seqs	sum_len	min_len	avg_len	max_len
BM_OUT_1.fastq.gz	FASTQ	DNA	111,814,087	16,212,501,865	50	145.0	150
BM_OUT_2.fastq.gz	FASTQ	DNA	111,814,087	16,207,292,619	50	144.9	150

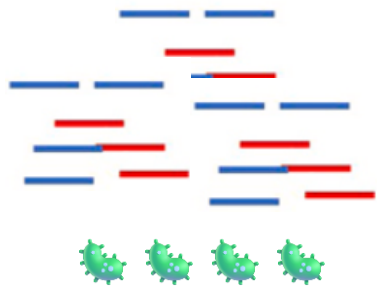
From FASTQ to Good Quality Reads — What next?



Raw read



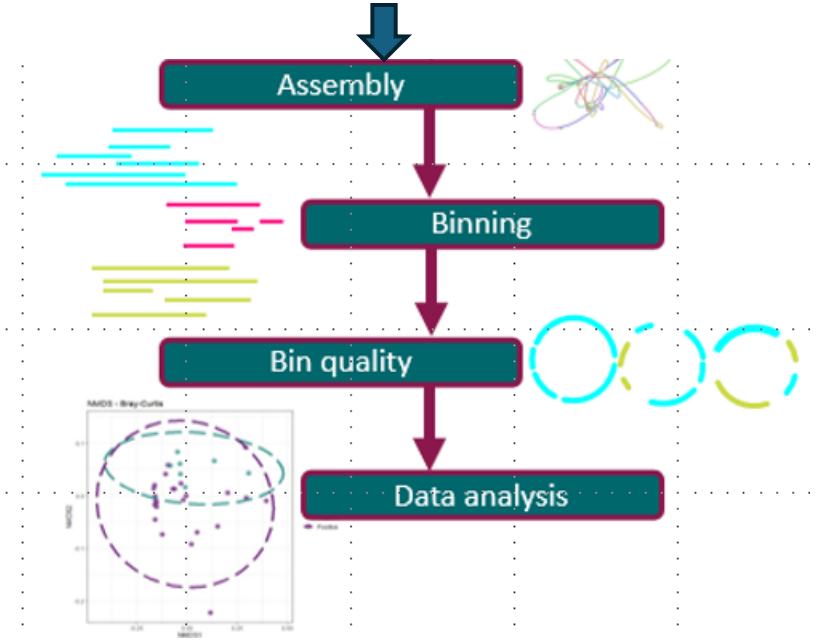
Quality Trimmed reads



Clean reads

ACGCTGCTAGCTAGTGCTCTTTGATCGGCGTAGGATCGCTCAGCTTCGCG.....
GCTAGTGCTGATCTCTGATATGCTCGTATAGCGTAGCGGCTGAGAG.....
.....
TCGCGCTGATAGTCGCTCGAGATCAGATCGCTATATTGCGCGTATATCG.....
CTCTAGGATCTCTCTGAGATCTTCTGCCGCGTATACGCATAACGC.....

Directly analyze quality-filtered reads without assembly



Read-based Analysis of Cleaned Metagenomic Reads

- Directly analyze quality-filtered reads without assembly.
- **Advantages**: Faster, less memory-intensive, and suitable for large-scale comparisons.

Taxonomic Profiling (To identify and quantify microbial taxa in the sample)

- **Tools**:
  - Kraken2 – k-mer-based, ultra-fast https://ccb.jhu.edu/software/kraken2/
  - MetaPhlAn – marker gene-based, high resolution https://huttenhower.sph.harvard.edu/metaphlan/
  - Centrifuge – space-efficient for large databases https://ccb.jhu.edu/software/centrifuge/
  - Kaiju – protein-level classification, good for viruses https://www.nature.com/articles/ncomms11257

Functional Profiling (To identify microbial functions (e.g., pathways, enzyme activities)

- **Tools**:
  - HUMAnN: Works with MetaPhlAn profiles; pathway-level output https://github.com/biobakery/humann
  - DIAMOND: fast protein alignment

Widely used taxonomy assignment tools

Article | Open access | Published: 23 February 2023

Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4

Aitor Blanco-Míguez, Francesco Beghini, Fabio Cumbo, Lauren J. McIver, Kelsey N. Thompson, Moreno Zolfo, Paolo Manghi, Leonard Dubois, Kun D. Huang, Andrew Maltez Thomas, William A. Nickols, Gianmarco Piccinno, Elisa Piperni, Michal Punčochář, Mireia Valles-Colomer, Adrian Tett, Francesca Giordano, Richard Davies, Jonathan Wolf, Sarah E. Berry, Tim D. Spector, Eric A. Franzosa, Edoardo Pasolli, Francesco Asnicar, ... Nicola Segata + Show authors

Nature Biotechnology **41**, 1633–1644 (2023) | Cite this article

70k Accesses | **722** Citations | **138** Altmetric | Metrics

Short Report | Open access | Published: 28 November 2019

Improved metagenomic analysis with Kraken 2

Derrick E. Wood, Jennifer Lu & Ben Langmead

Genome Biology **20**, Article number: 257 (2019) | Cite this article

129k Accesses | **5052** Citations | **140** Altmetric | Metrics

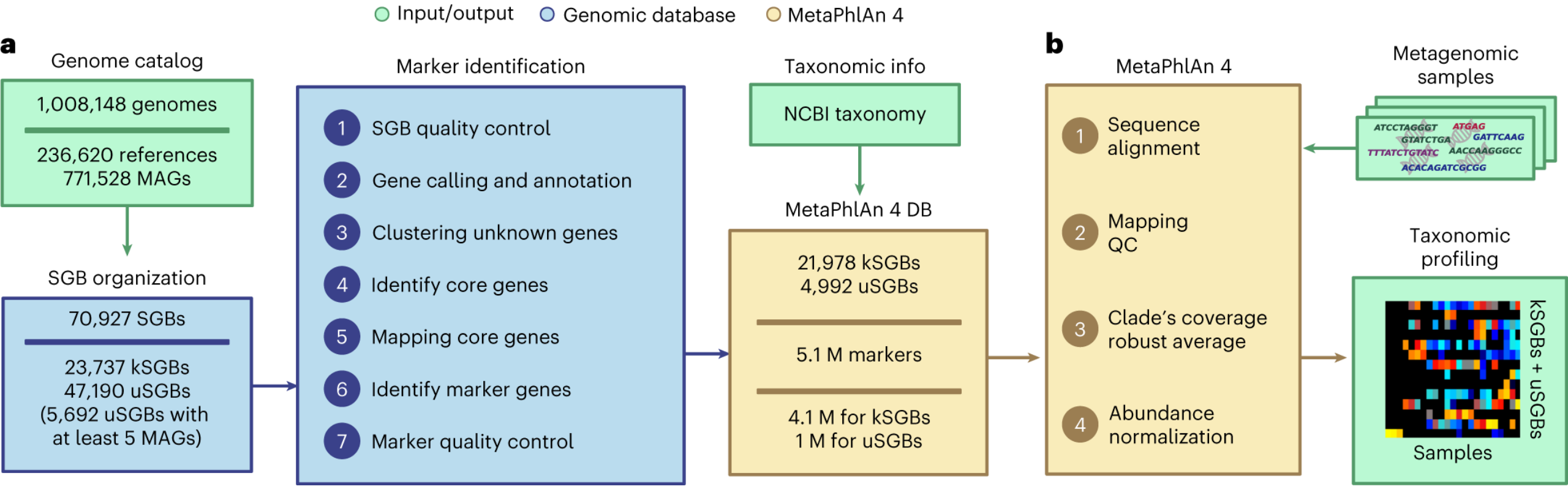
MetaPhlAn4

- **Approach**: Uses a curated set of **~1 .6 million clade-specific marker genes**
- **Database**: Built from reference genomes across bacteria, archaea, viruses, and eukaryotes
- **Output**: Relative abundance of taxa (not raw read counts)

Kraken2

- **Approach**: **K-mer-based classification** using a hash of all k-mers in reference genomes
- **Database**: Large, comprehensive (e.g., RefSeq), can be customized
- **Output**: Read-level assignments → can be summarized for abundance

Taxonomy assignment using MetaPhlAn4

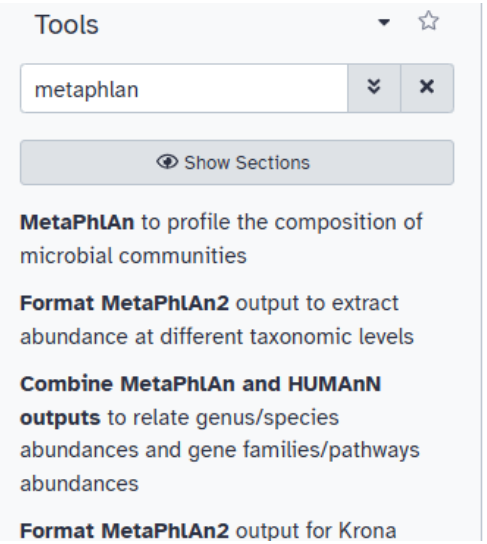


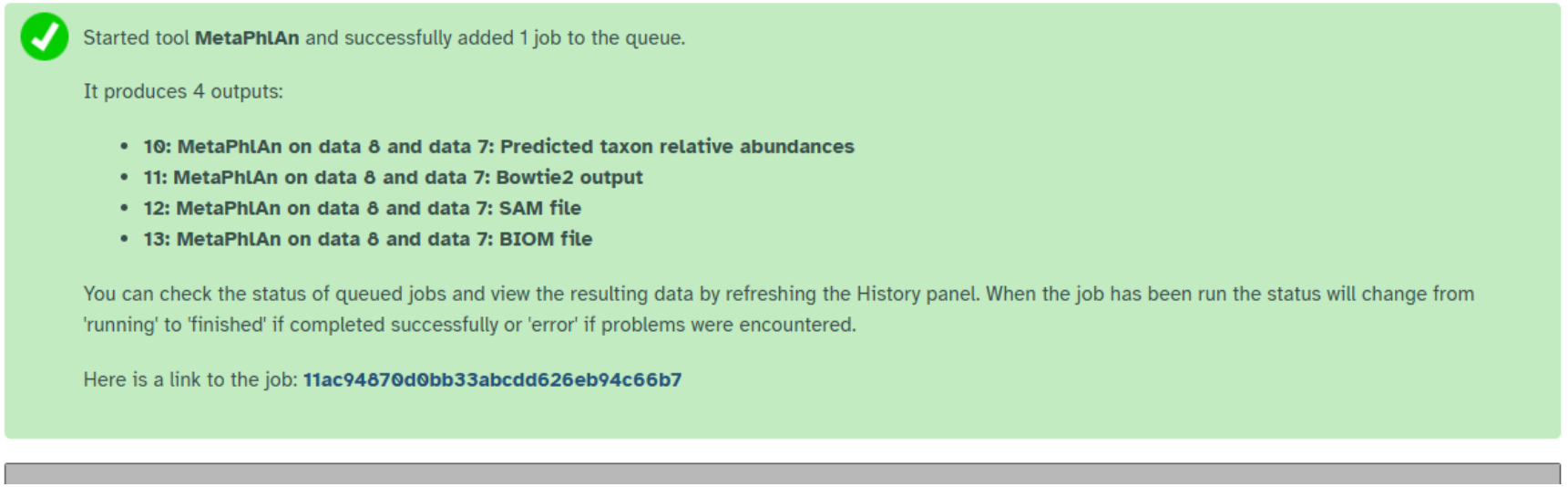
Taxonomy assignment using MetaPhlAn using Galaxy server

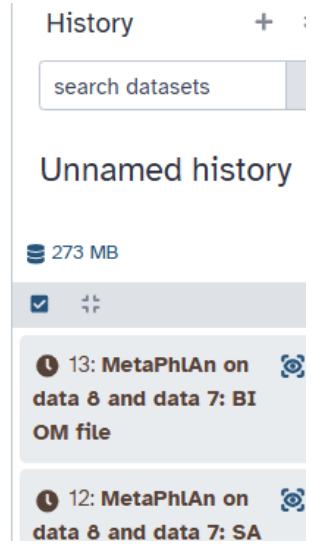
1.MetaPhlAN tool with

- "Input file" to the imported file
- "Database with clade-specific marker genes"

This step may take a couple of minutes.







MetaPhlAn results

files are generated:

•A tabular file with the community structure

```
#39999588 reads processed
#SampleID       Metaphlan_Analysis
#clade_name     NCBI_tax_id     relative_abundance      additional_species
k__Bacteria     2       100.0
k__Bacteria|p__Firmicutes       2|1239  62.55403
k__Bacteria|p__Bacteroidetes    2|976   34.22134
k__Bacteria|p__Proteobacteria   2|1224  2.10971
k__Bacteria|p__Actinobacteria   2|201174        0.96485
k__Bacteria|p__Bacteria_unclassified    2|      0.15007
k__Bacteria|p__Firmicutes|c__Clostridia 2|1239|186801   53.79254
k__Bacteria|p__Bacteroidetes|c__Bacteroidia     2|976|200643    34.22134
k__Bacteria|p__Firmicutes|c__CFGB4806   2|1239|  3.11065
k__Bacteria|p__Firmicutes|c__Erysipelotrichia   2|1239|526524   2.66124
k__Bacteria|p__Firmicutes|c__Negativicutes      2|1239|909932   2.22649
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria     2|1224|28216    1.09711
k__Bacteria|p__Proteobacteria|c__Deltaproteobacteria    2|1224|28221    0.99145
k__Bacteria|p__Actinobacteria|c__Coriobacteriia 2|201174|84998  0.59071
k__Bacteria|p__Actinobacteria|c__Actinomycetia  2|201174|1760   0.37415
k__Bacteria|p__Firmicutes|c__CFGB38642  2|1239|  0.27135
k__Bacteria|p__Firmicutes|c__Bacilli    2|1239|91061    0.25411
k__Bacteria|p__Bacteria_unclassified|c__CFGB76763       2||     0.15007
k__Bacteria|p__Firmicutes|c__Firmicutes_unclassified    2|1239|  0.13332
```

Simplified MetaPhlAn output

3: MetaPhlAn on data 2 and data 1: Predicted taxon relative abundances

ok

75 lines 4 columns, 5 comments

format **tabular** database **?** size **3.8 KB**

Preview | Visualize | Details | Edit

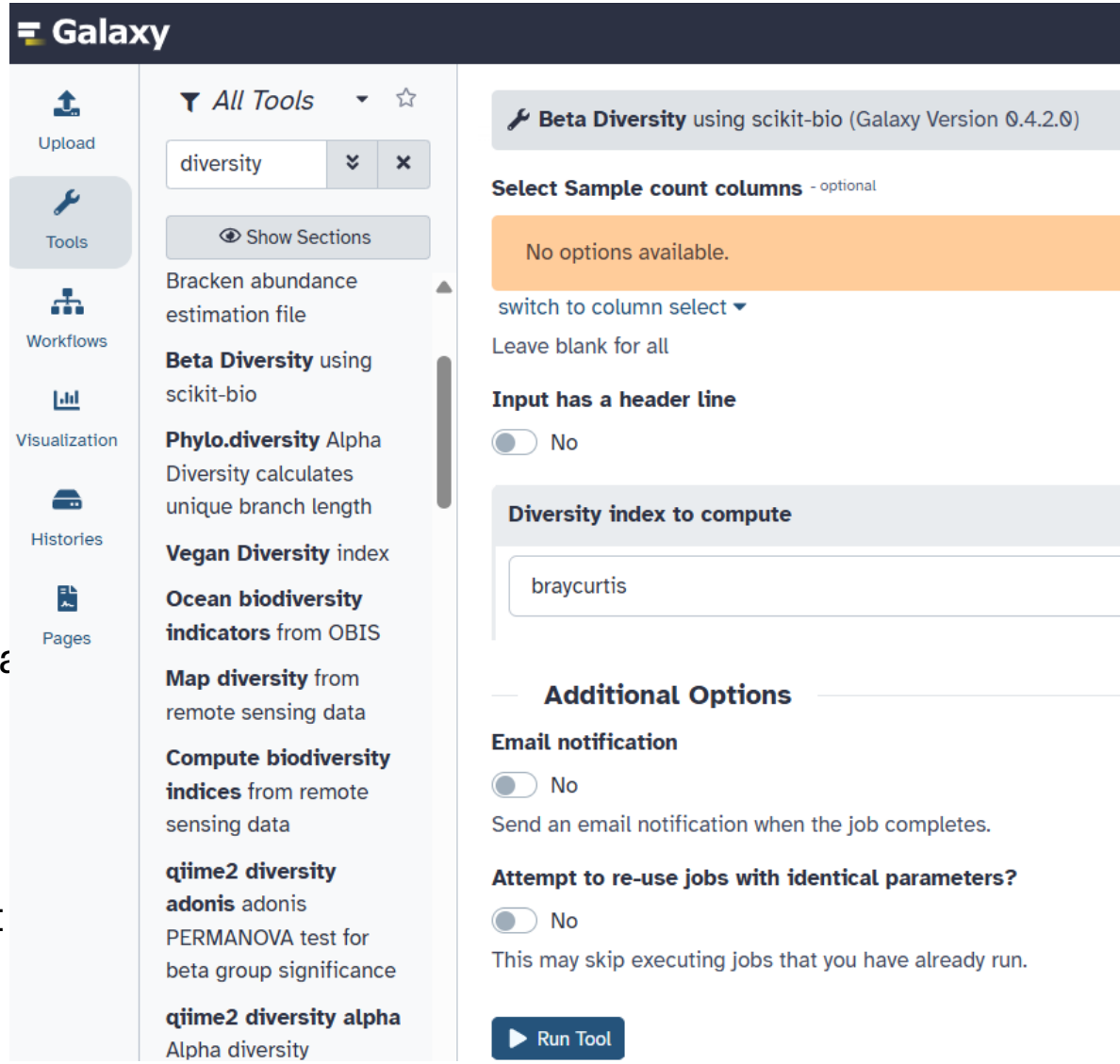
Column 1	Column 2	Column 3	Column 4
#mpa_vOct22_CHOCOPhlAnSGB_202212			
#/usr/local/bin/metaphlan in_f,in_r --input_type fastq --read_min_len 70 --bt2_ps very-sensitive --min_mapq_val 5 --bowtie2db /cvmfs/data.galaxyproject.org/byhand/metaphlan/mpa_vOct22_CHOCOPhlAnSGB_202212 --index mpa_vOct22_CHOCOPhlAnSGB_202212 -t rel_ab --tax_lev s --min_cu_len 2000 --add_viruses --stat tavg_g --stat_q 0.2 --perc_nonzero 0.33 --avoid_disqm --sample_id_key SampleID --sample_id Metaphlan_Analysis -o /anvil/scratch/x-xcgalaxy/main/staging/69512792/outputs/dataset_e23b4e5c-f2fb-419f-bc8c-e9f577ce3bcd.dat --bowtie2out bowtie2out -s /anvil/scratch/x-xcgalaxy/main/staging/69512792/outputs/dataset_ffc8cadf-2e0d-4879-8ea5-edb3014a75a6.dat --biom /anvil/scratch/x-xcgalaxy/main/staging/69512792/outputs/dataset_540122ff-fe8b-4903-ad62-fc638d1ba04d.dat --nproc 29 --offline			
#7681771 reads processed			
#SampleID	Metaphlan_Analysis		
#clade_name	NCBI_tax_id	relative_abundance	additional_species
s__Klebsiella_pneumoniae	573	24.03482	
s__Ruminococcus_bromii	40518	19.69272	
s__Firmicutes_bacterium_AF16_15	2292885	8.71585	
s__Roseburia_inulinivorans	360807	5.40871	

Alpha Diversity (Within-sample diversity)
•Measures how diverse a single sample is

Beta Diversity (Between-sample diversity)
•Measures how different microbial communities are between samples

Differential Abundance Analysis
•Compare species abundance between groups (e.g., hea vs. CRC)

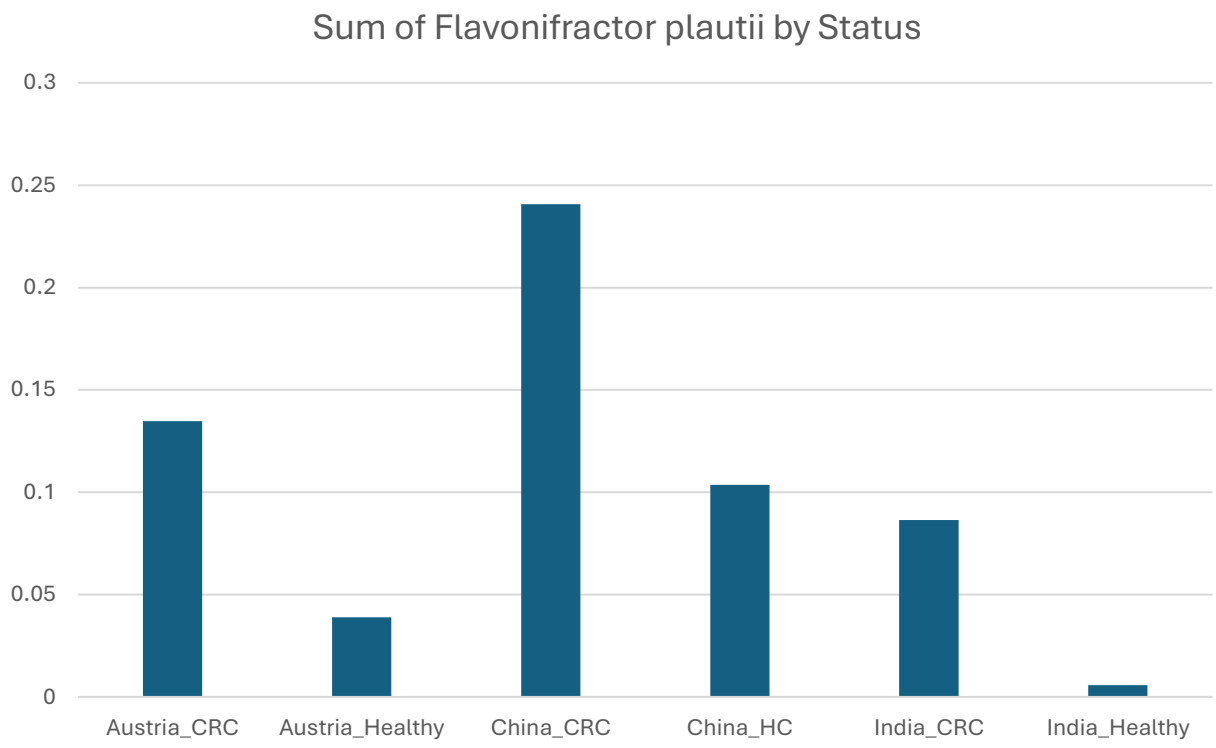
Input for Machine Learning Models
•Use species abundance for classification (e.g., predict disease status)



Which species high in healthy/disease

Status	Sample ID	Flavonifractor plautii
Austria_CRC	ERR688508	0.00027
Austria_CRC	ERR688512	0.00178
Austria_CRC	ERR688519	6.00E-05
Austria_Healthy	ERR688506	0.00016
Austria_Healthy	ERR688507	0.0003
China_CRC	ERR1018306	0.01234
China_CRC	ERR1018307	0.00057
China_HC	ERR1018268	0.00068
India_CRC	SRR8865599	0.02399
India_CRC	SRR8865600	0.00023
India_Healthy	HAG	6.00E-05
India_Healthy	HAH	6.00E-05

Status	Abundance of Flavonifractor plautii
Austria_CRC	0.1348
Austria_Healthy	0.03893
China_CRC	0.24071
China_HC	0.10358
India_CRC	0.08646
India_Healthy	0.00577



Understanding What Microbes Do: Functional Profiling

Objective:

"To move beyond 'who is there' to 'what they can do' — by identifying metabolic pathways, enzymes, and gene functions encoded in the microbiome."

Why?

•Functional profiling reveals **metabolic potential** and **biological roles** of microbes.

•Useful in understanding **health/disease associations**, e.g., inflammation, nutrient-utilizing genes, or antibiotic resistance

Functional Profiling using HUMAnN

HUMAnN: Functional Profiling Using MetaPhlAn Outputs (https://github.com/biobakery/human)

- Profile gene families and pathways from microbial communities using a taxonomic guide
- Uses **MetaPhlAn taxonomic profile** to map functional content

•Performs:

- •Gene family profiling (e.g., UniRef)
- •Pathway reconstruction (e.g., MetaCyc pathways)

•Outputs:

- •**Gene family abundances**
- •**Pathway abundances and coverage**

•Output files: .tsv tables usable in downstream analysis or visualization.

```
# Pathway       demo_Abundance
UNMAPPED        6297.5273970005
UNINTEGRATED    3178.6342442721
UNINTEGRATED|unclassified       198.7273134329
PWY-4203: volatile benzenoid biosynthesis I (ester formation)   22.7546997826
PWY-4203: volatile benzenoid biosynthesis I (ester formation)|unclassified      22.7546997826
PWY490-3: nitrate reduction VI (assimilatory)   21.5240884774
PWY490-3: nitrate reduction VI (assimilatory)|unclassified      21.5240884774
HEME-BIOSYNTHESIS-II-1: heme b biosynthesis V (aerobic) 16.7796783348
HEME-BIOSYNTHESIS-II-1: heme b biosynthesis V (aerobic)|unclassified    16.7796783348
HEMESYN2-PWY: heme b biosynthesis II (oxygen-independent)       16.7796783348
HEMESYN2-PWY: heme b biosynthesis II (oxygen-independent)|unclassified  16.7796783348
PWY-6305: superpathway of putrescine biosynthesis       16.5975171231
```

Research Article

Landscape of flavonoid metabolism in human gut microbiome and its association with health and disease

Deepika Pateriya , Vishnu Prasoodanan P. K. , Joy Scaria & **Vineet K. Sharma**

Article: 2520788 | Received 24 Feb 2025, Accepted 05 Jun 2025, Published online: 25 Jun 2025

Cite this article https://doi.org/10.1080/29933935.2025.2520788 Check for updates

ABSTRACT

The positive effects of dietary flavonoids on health depend on their bioavailability in the human gut, where the flavonoid-modifying enzymes (FMEs) in gut bacteria play a crucial role in flavonoid metabolism. Thus, to comprehensively examine the role of FMEs in this process, we first constructed a database of potential FMEs containing 6,865 proteins. We identified homologs of these FMEs in gut bacterial genomes and reported species that can potentially modify flavonoids but were not previously known in this context. We examined the differential abundance of FMEs in the gut microbiomes of healthy and diseased individuals from Western and non-Western populations with distinct dietary habits. The differential enrichment of key FMEs between Western and non-Western populations and between disease and healthy samples highlights differences in gut flavonoid metabolism based on diet, population, and health status. This study reveals a

Related re

People also read

Two-week sup
adolescentis i
with lactose i

Monica Rama
Gut Microbes R
Published onlin

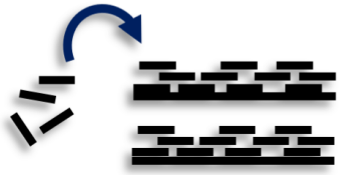
Novel cross-fe
metabolizing

Assembly-based Analysis of Cleaned Metagenomic Reads

1 Quality Control

- PCR duplicates removal
- Quality trimming
- Host removal
- Common contaminant removal

➢ **QC reads**



2 Assembly

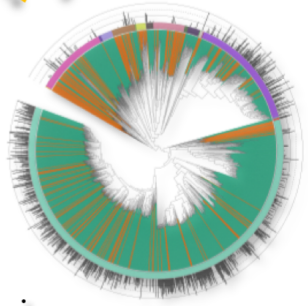
- Error correction
- Paired-end merging
- Assembly (metaSpades/megahit)
- Post-filtering

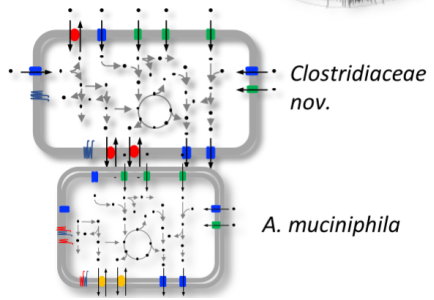
➢ **High-quality Scaffolds**

3 Genomic Binning

- Binning (metabat, maxbin2)
- Quality Assessment (checkM)
- Bin refining (DAS Tool)
- Dereplication (dRep)
- Quantification
- Robust taxonomic classification (GTDB)

➢ **Genomes**
➢ **Abundances**





4 Annotation

- Gene prediction (prodigal)
- Cluster redundant genes (linclust)
- Annotation (eggNOG)

➢ **Functional annotations**

Assembly-based Analysis

Assembling reads into longer contigs/scaffolds

Metagenome Assembly

•**Tools**: MEGAHIT, metaSPAdes

•**Goal**: Assemble reads to obtain larger fragments (contigs)

Binning into MAGs (Metagenome-Assembled Genomes)

•**Tools**: MetaBAT2, MaxBin2, CONCOCT

•**Goal**: Group contigs into draft genomes for individual microbial populations

Genome Annotation

•**Tools**: Prokka, eggNOG-mapper, DRAM

•**Goal**: Annotate genes, pathways, and functional elements on assembled genomes.

Comparative Genomics

•**Tools**: ANI tools, dRep

•**Goal**: Compare MAGs or contigs to reference genomes or across samples.

Novel Gene Discovery

•**Tools**: Gene prediction (Prodigal), HMM-based tools

•**Goal**: Identify new genes, biosynthetic gene clusters, etc.

Mobile Elements & Plasmids

•**Tools**: PlasFlow, MOB-suite

•**Goal**: Detect plasmids and horizontal gene transfer elements.

Step 1: Sequencing Reads = Shredded Sentences

You receive millions of tiny paper strips — just snippets of sentences from many mystery books.
🗣️ “Who are the authors?”
🧠 “What’s the story about?”
We don’t know yet!

Step 2: Assembly = Reconstructing Pages

You begin stitching the strips together into longer paragraphs or pages — these are your contigs.
🧩 It’s like putting puzzle pieces together without the box image!

Step 3: Binning = Grouping Pages into Books

You now try to group similar pages together — same language, same font, repeated names — into book piles.
📚 Each pile = a bin, probably from one book (one genome)

Step 4: MAGs = Full or Draft Books

Once a pile has **enough pages**, we say:
✅ “This is a complete or draft version of a book!”
That’s your MAG — Metagenome-Assembled Genome.
Some books are missing chapters (incomplete), some are messy (contaminated), and some are near perfect!

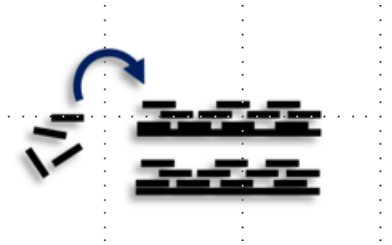
Metagenome Assembly

•**Why?** Enables access to entire genes, operons, or even full microbial genomes that aren't detectable from short reads alone

•**Goal:** Stitch short reads into longer sequences (contigs/scaffolds) for better context

•**Tools:**

- *MEGAHIT* – Fast and memory-efficient for large metagenomes (https://github.com/voutcn/megahit)
- *metaSPAdes* – Produces higher-quality assemblies for complex samples (http://cab.spbu.ru/software/spades)



JOURNAL ARTICLE

MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph FREE

Dinghua Li , Chi-Man Liu , Ruibang Luo , Kunihiko Sadakane , Tak-Wah Lam
Author Notes

Bioinformatics, Volume 31, Issue 10, May 2015, Pages 1674–1676,
https://doi.org/10.1093/bioinformatics/btv033
Published: 20 January 2015 **Article history**

Genome Res. 2017 May;27(5):824–834. doi: 10.1101/gr.213959.116

metaSPAdes: a new versatile metagenomic assembler

Sergey Nurk [1,4], Dmitry Meleshko [1,4], Anton Korobeynikov [1,2], Pavel A Pevzner [1,3]

Author information Article notes Copyright and License information
PMCID: PMC5411777 PMID: 28298430

Assembly Quality Check

Not all assemblies are equally good — we need to evaluate how **complete**, **accurate**, and **non-redundant** the contigs are.

Key Aspects to Assess:

- **N50**: Measures contiguity — higher N50 = longer contigs
- **Total assembly length**: Does it match expected metagenome size?
- **Number of contigs**: Fewer contigs = better assembly
- **Coverage depth**: Low-depth regions may be less reliable

Tools for Quality Check:

- **QUAST** – General assembly metrics and visual summaries
- **MetaQUAST** – Specifically for metagenomes

```
(base) $ cat GM001_assembly_stats.tsv
Assembly                    GM001_spades_short
# contigs (>= 0 bp)         53519
# contigs (>= 1000 bp)      53519
# contigs (>= 5000 bp)      6640
# contigs (>= 10000 bp)     2594
# contigs (>= 25000 bp)     744
# contigs (>= 50000 bp)     275
Total length (>= 0 bp)      187894437
Total length (>= 1000 bp)   187894437
Total length (>= 5000 bp)   97263304
Total length (>= 10000 bp)  69578841
Total length (>= 25000 bp)  41998745
Total length (>= 50000 bp)  25697567
# contigs                   53519
Largest contig              437750
Total length                187894437
```

Bioinformatics. 2013 Feb 19;29(8):1072–1075. doi: 10.1093/bioinformatics/btt086

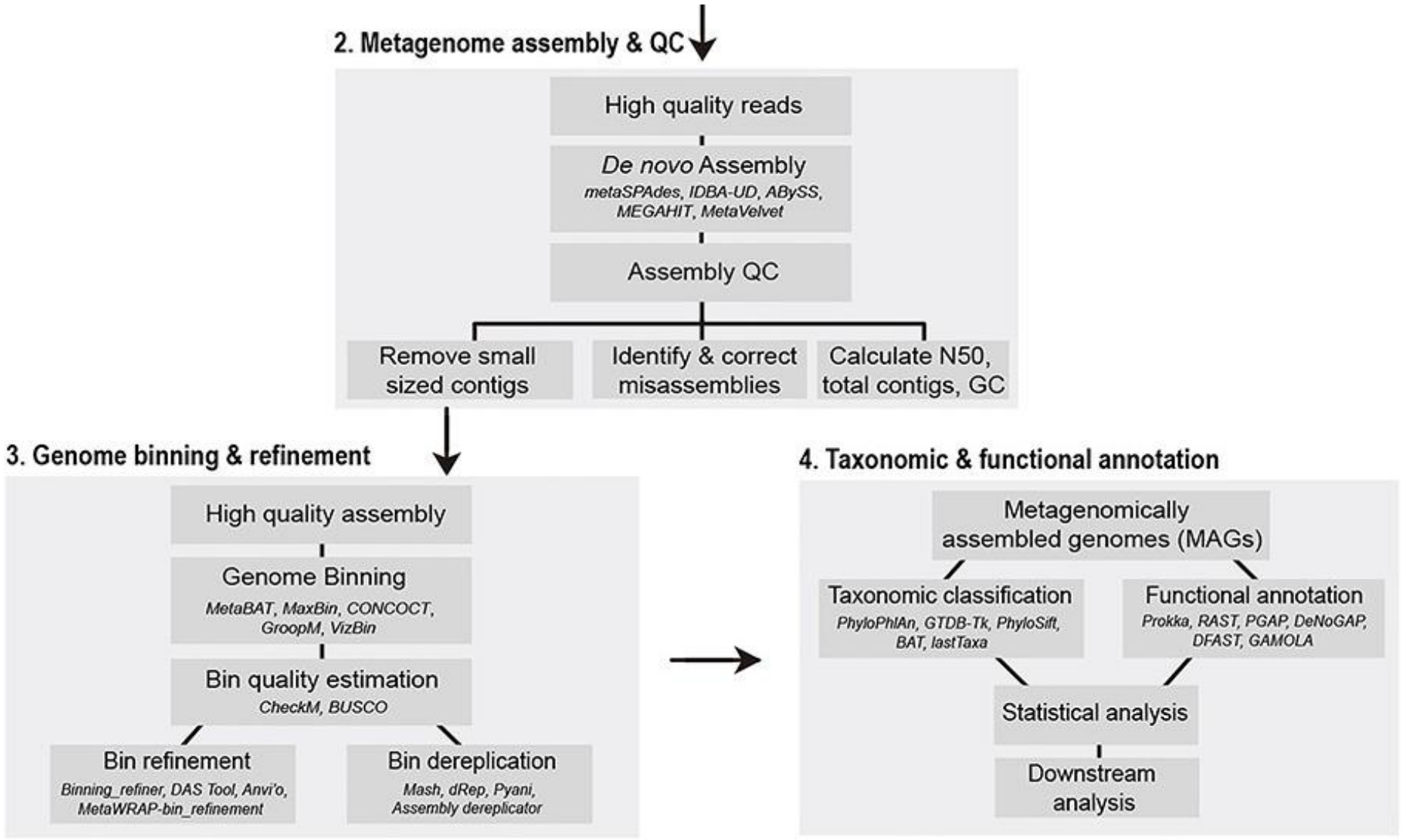
QUAST: quality assessment tool for genome assemblies

Alexey Gurevich [1,*], Vladislav Saveliev [1], Nikolay Vyahhi [1], Glenn Tesler [2]

Author information · Article notes · Copyright and License information

PMCID: PMC3624806 PMID: 23422339

https://doi.org/10.1093/bioinformatics/btt086


2. Metagenome assembly & QC
High quality reads
De novo Assembly
metaSPAdes, IDBA-UD, ABySS, MEGAHIT, MetaVelvet
Assembly QC
Remove small sized contigs
Identify & correct misassemblies
Calculate N50, total contigs, GC
3. Genome binning & refinement
High quality assembly
Genome Binning
MetaBAT, MaxBin, CONCOCT, GroopM, VizBin
Bin quality estimation
CheckM, BUSCO
Bin refinement
Binning_refiner, DAS Tool, Anvi'o, MetaWRAP-bin_refinement
Bin dereplication
Mash, dRep, Pyani, Assembly dereplicator
4. Taxonomic & functional annotation
Metagenomically assembled genomes (MAGs)
Taxonomic classification
PhyloPhlAn, GTDB-Tk, PhyloSift, BAT, lastTaxa
Functional annotation
Prokka, RAST, PGAP, DeNoGAP, DFAST, GAMOLA
Statistical analysis
Downstream analysis

Metagenomic Binning (Similar contigs are grouped together)

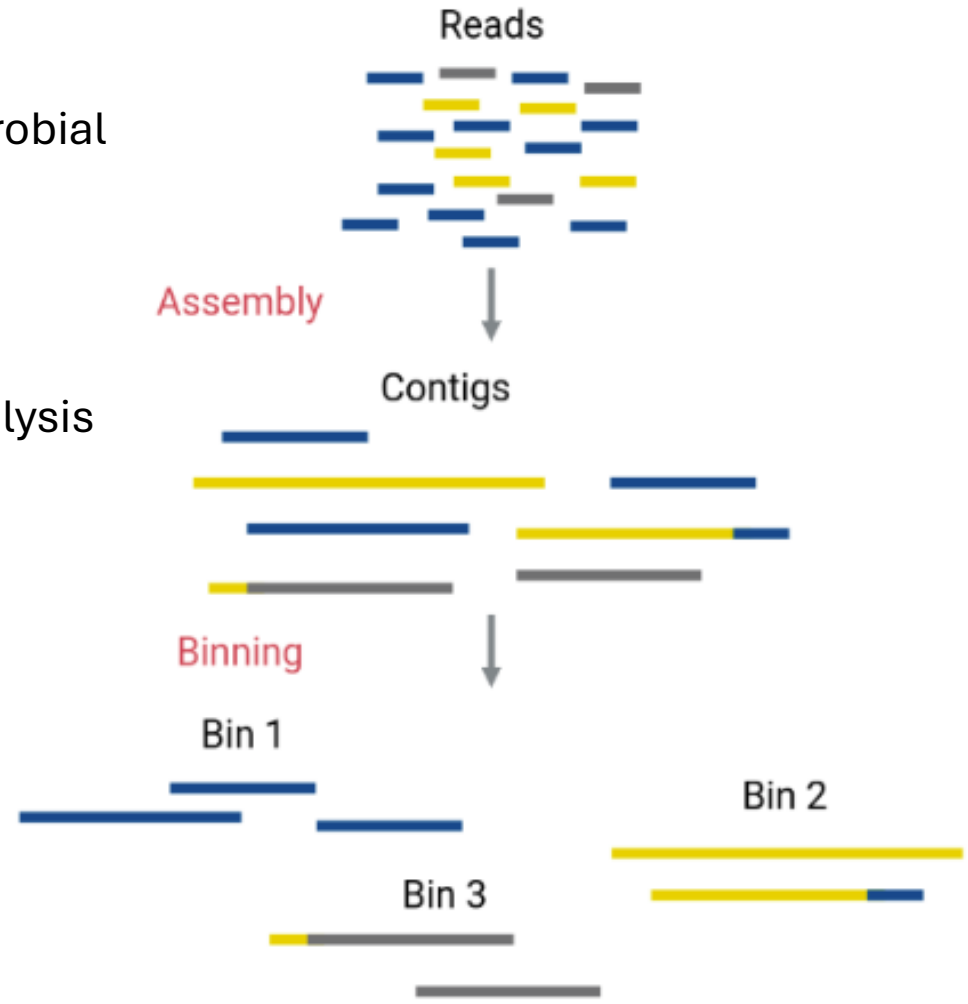
2. Binning into MAGs (Metagenome-Assembled Genomes)

•**Goal:** Cluster contigs into draft genomes representing distinct microbial species or strains

•**Tools:**

MetaBAT2, *MaxBin2*, *CONCOCT* – Use sequence composition and coverage patterns for grouping

•**Outcome:** High-quality draft genomes that allow species-level analysis and deeper functional exploration



MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies

Dongwan D Kang [1], Feng Li [2], Edward Kirton [1], Ashleigh Thomas [1], Rob Egan [1], Hong An [2], Zhong Wang [1,3,4,✉]

Editor: Joseph Gillespie

▸ Author information ▸ Article notes ▸ Copyright and License information

PMCID: PMC6662567 PMID: 31388474

https://bitbucket.org/berkeleylab/metabat

Metagenomically assembled genomes (MAGs)

🧱 What Are MAGs?

•MAGs are **draft microbial genomes** reconstructed directly from metagenomic data.

•Assembled by grouping contigs/scaffolds with:

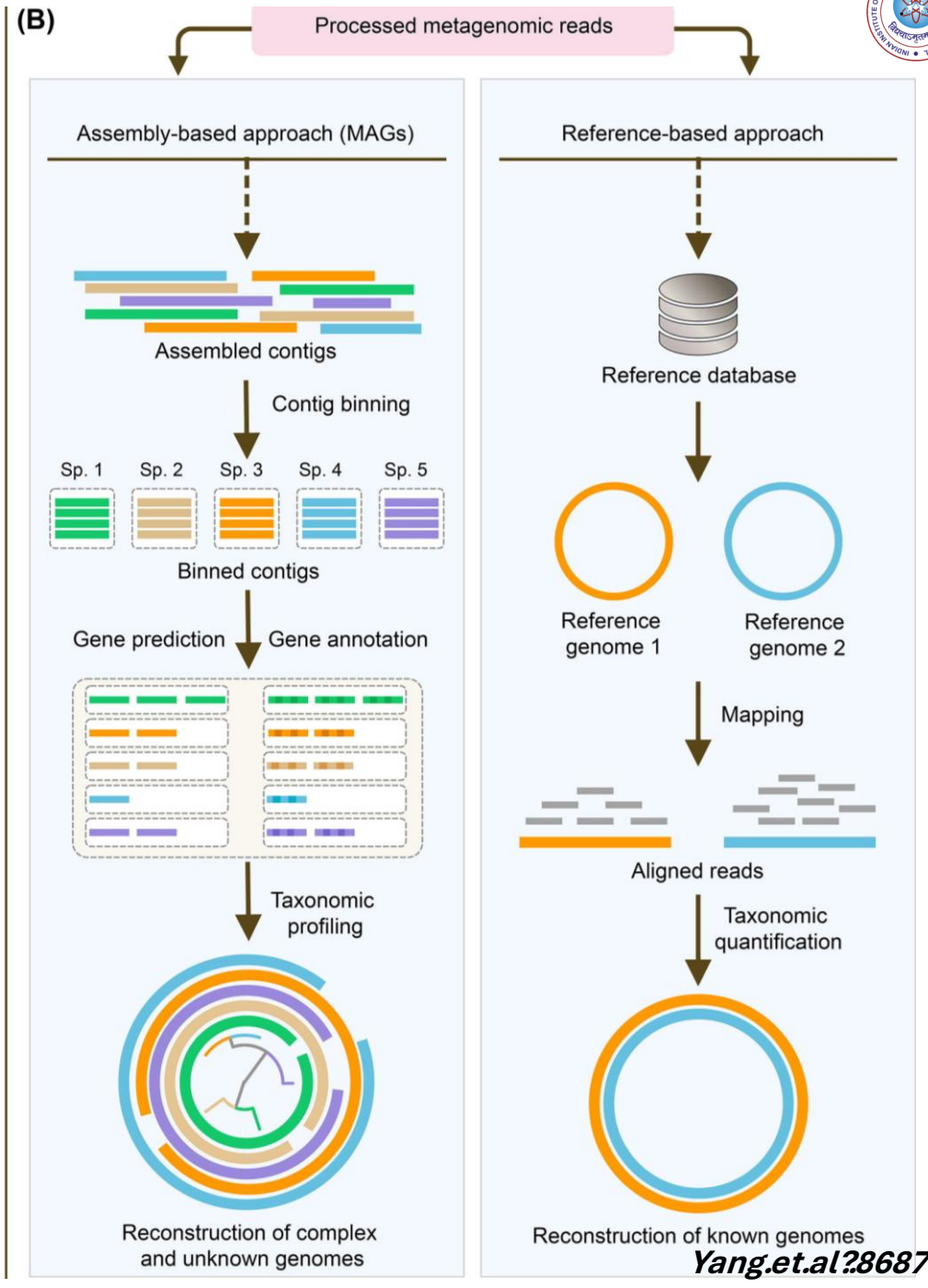
- Similar GC content
- Co-abundance patterns
- Tetranucleotide frequency

Why Are They Powerful?

•**Capture novel, uncultured microbes** missing from databases.

•Provide **both taxonomic and functional** insights.

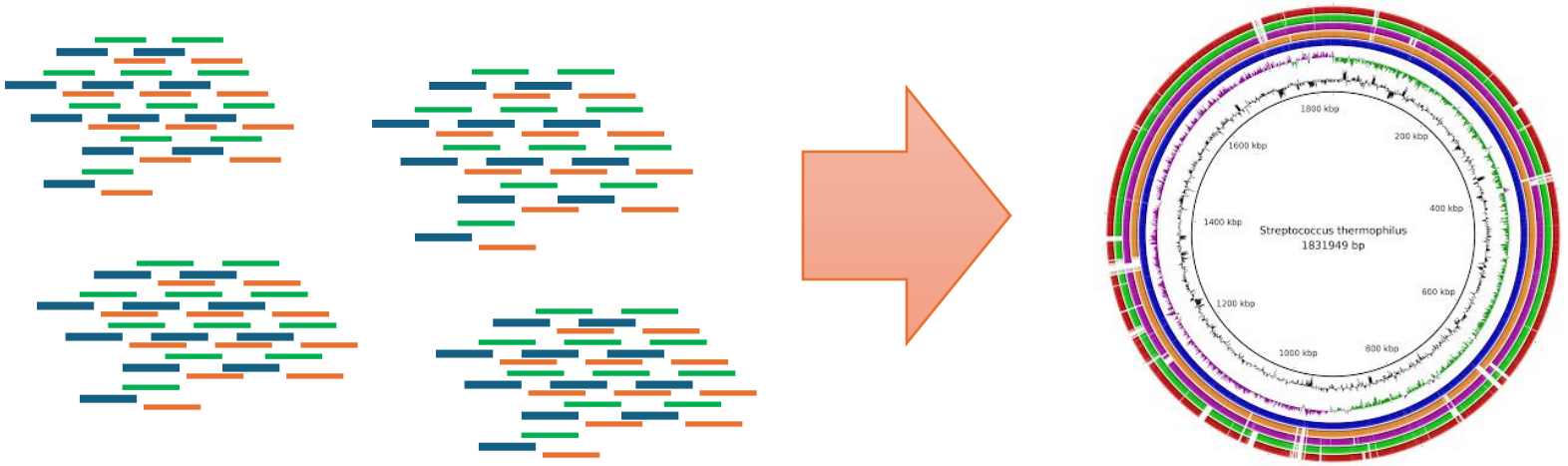
•Reveal the **genomic blueprint** of microbial players in complex ecosystems (e.g., gut, soil, ocean).



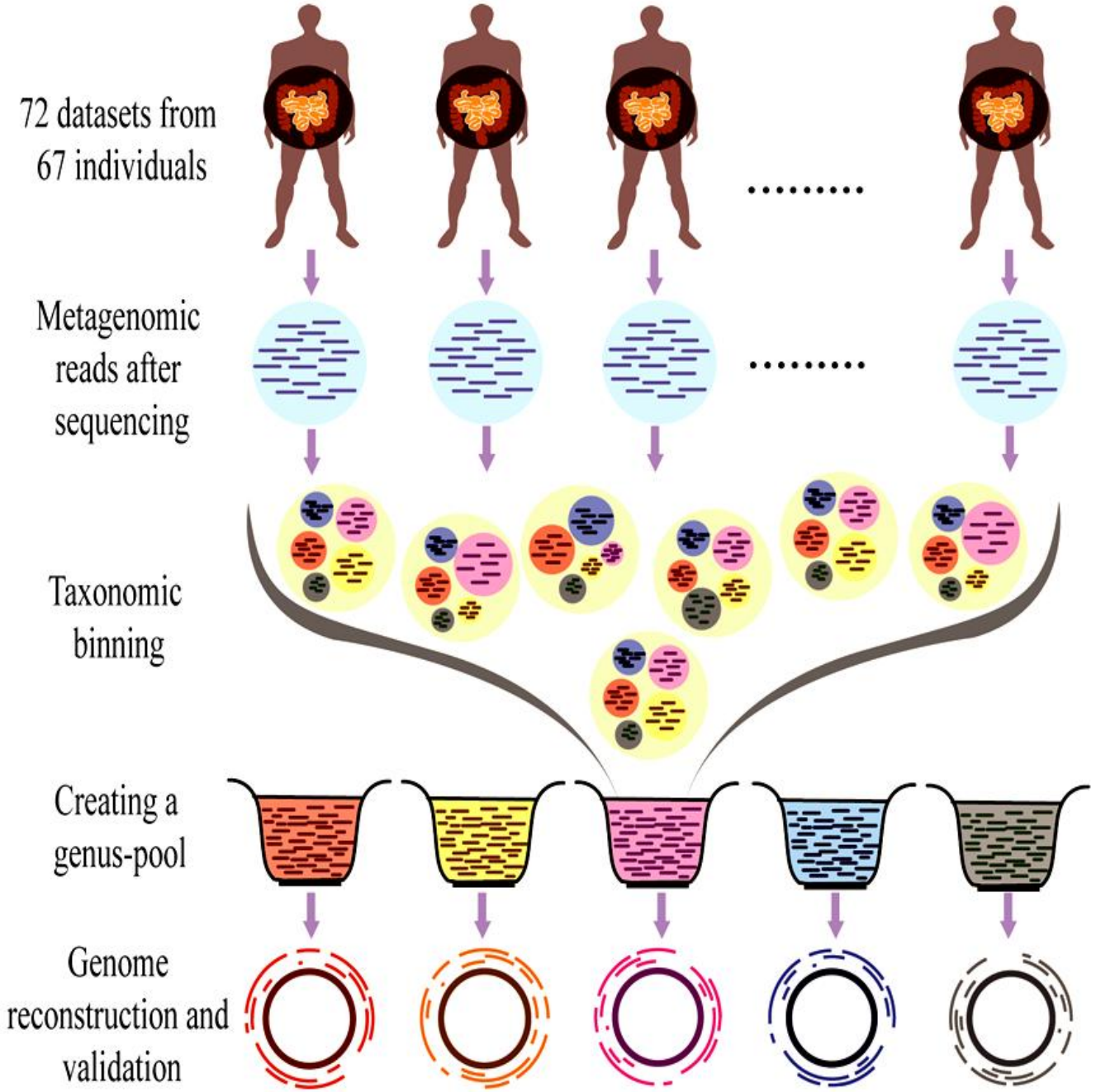
Yang.et.al?8687

What makes it so Difficult ??

- Not feasible to generate complete genomic assemblies of species from the metagenomic sequencing of a complex environment.
- Enormous microbial diversity, requires massive sequencing and cost to obtain a reasonable coverage for each species
- Mixture of metagenomic reads obtained from multiple samples of similar origin.



Binning-Assembly Approach



Result and Conclusion

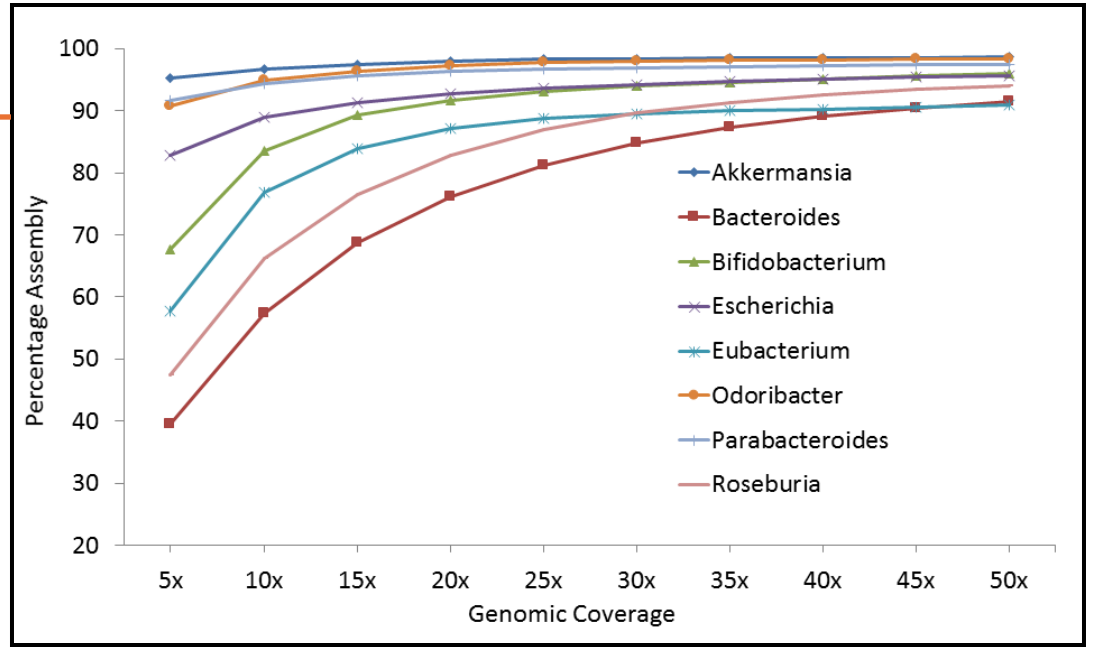
- Reconstruction of **1,156** bacterial and **279** viral genomes belonging to 219 bacterial and 84 viral families, respectively.

- For a total of **126 bacterial** and **11 viral** genomes, more than **90% complete draft genome** sequences could be reconstructed.

- Selected draft assembled genomes could be validated with **99.8% accuracy** using their ORFs.

- **25x coverage** is good enough for a high quality assembly.

- This approach along with **spiking** was useful in **improving the draft assembly** of a bacterial genome.



Genome	% Assembly	% Identity	% Complete ORFs
Odoribacter splanchnicus DSM 20712	99.17	97	99.94
Bacteroides thetaiotaomicron VPI 5482	98.89	92	99.93
Akkermansia muciniphila ATCC BAA 835	98.71	93	99.95
Parabacteroides distasonis ATCC 8503	98.69	97	99.81
Roseburia hominis A2 183	97.79	94	99.56
Bifidobacterium longum JCM 1217	97.14	94	99.80
Escherichia coli K 12 substr MDS42	95.72	95	99.84
Eubacterium siraeum V10Sc8a	95.07	93	99.84

Quality evaluation of MAGs

- MAGs must be **complete and clean** to be reliable.
- Tools like **CheckM** evaluate:
- ✅ **Completeness** – Are most essential genes present?
- ⚠️ **Contamination** – Are there genes from other organisms?
- Based on these, MAGs are rated as:
- 🥇 **High-quality** (e.g., >90% complete, <5% contamination)
- 🥈 Medium-quality, 🥉 Low-quality, or 🔧 Draft
- Only **high-quality bins** are used for downstream analysis.

CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes

Donovan H Parks [1], Michael Imelfort [1], Connor T Skennerton [1], Philip Hugenholtz [1,2], Gene W Tyson [1,3]

‣ Author information ‣ Article notes ‣ Copyright and License information

PMCID: PMC4484387 PMID: 25977477

https://doi.org/10.1101/gr.186072.114

Bin Id	Marker lineage	Completeness	Contamination	Strain heterogeneity
bin.5	f__Lachnospiraceae (UID1286)	99.28	0	0
bin.22	p__Actinobacteria (UID2112)	99.19	1.61	0
bin.17	o__Lactobacillales (UID374)	98.95	0	0
bin.16	o__Selenomonadales (UID1024)	98.75	0.1	100
bin.28	o__Clostridiales (UID1212)	98.66	2.01	33.33
bin.6	o__Selenomonadales (UID1024)	98.5	5.01	53.33
bin.9	f__Bifidobacteriaceae (UID1462)	98.34	2.88	28.57
bin.26	o__Clostridiales (UID1212)	97.99	1.01	50
bin.14	o__Clostridiales (UID1212)	97.99	1.01	0
bin.10	o__Lactobacillales (UID355)	97.91	3.69	87.5
bin.8	f__Lachnospiraceae (UID1286)	97.83	1.69	0

Taxonomic annotation of MAGs

What we want to know: (Assigning taxonomy):

➡️ *Who are these microbes?*

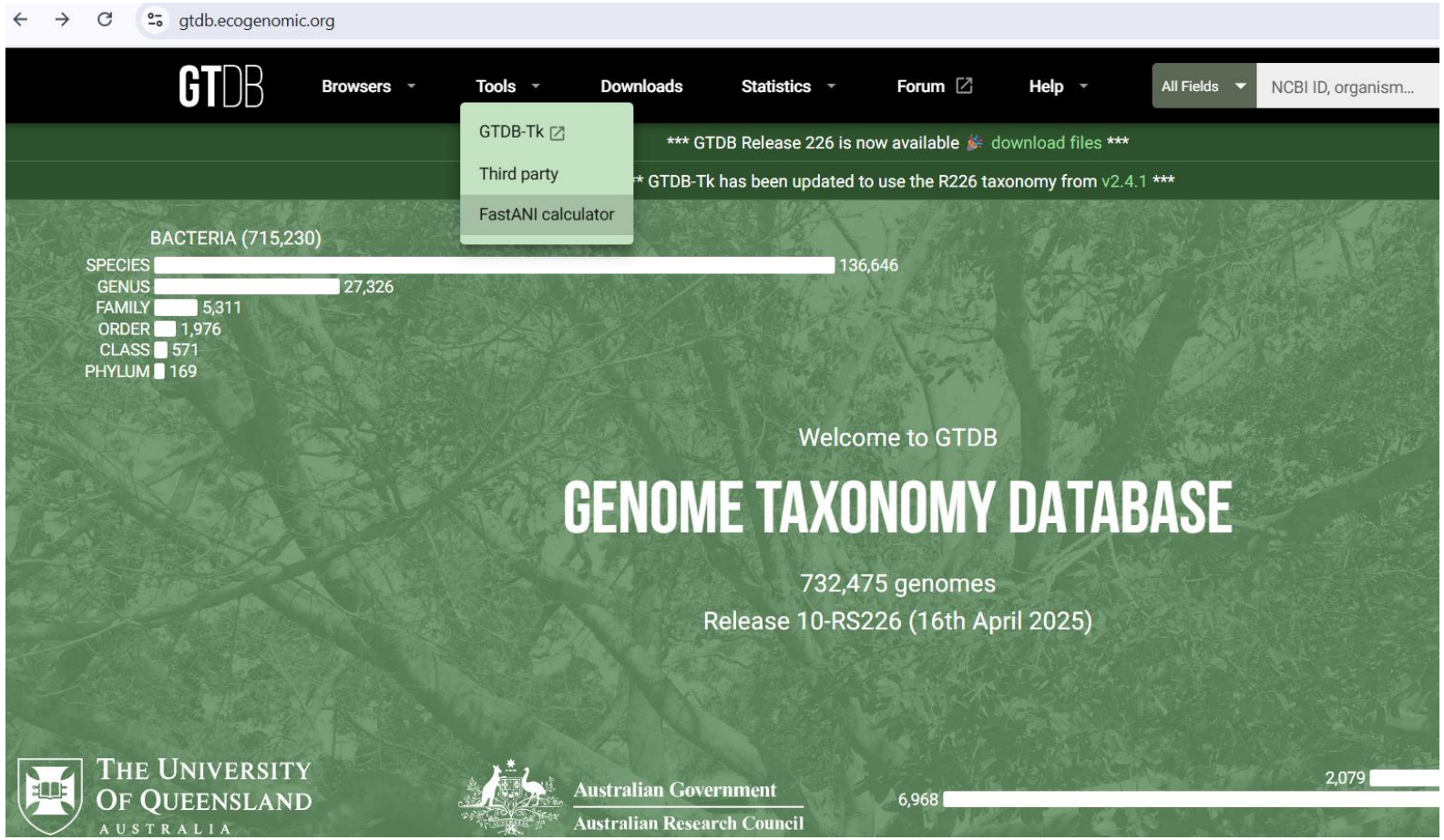
Tools:

🧪 **GTDB-Tk**

➡️ Looks for **universal marker genes**

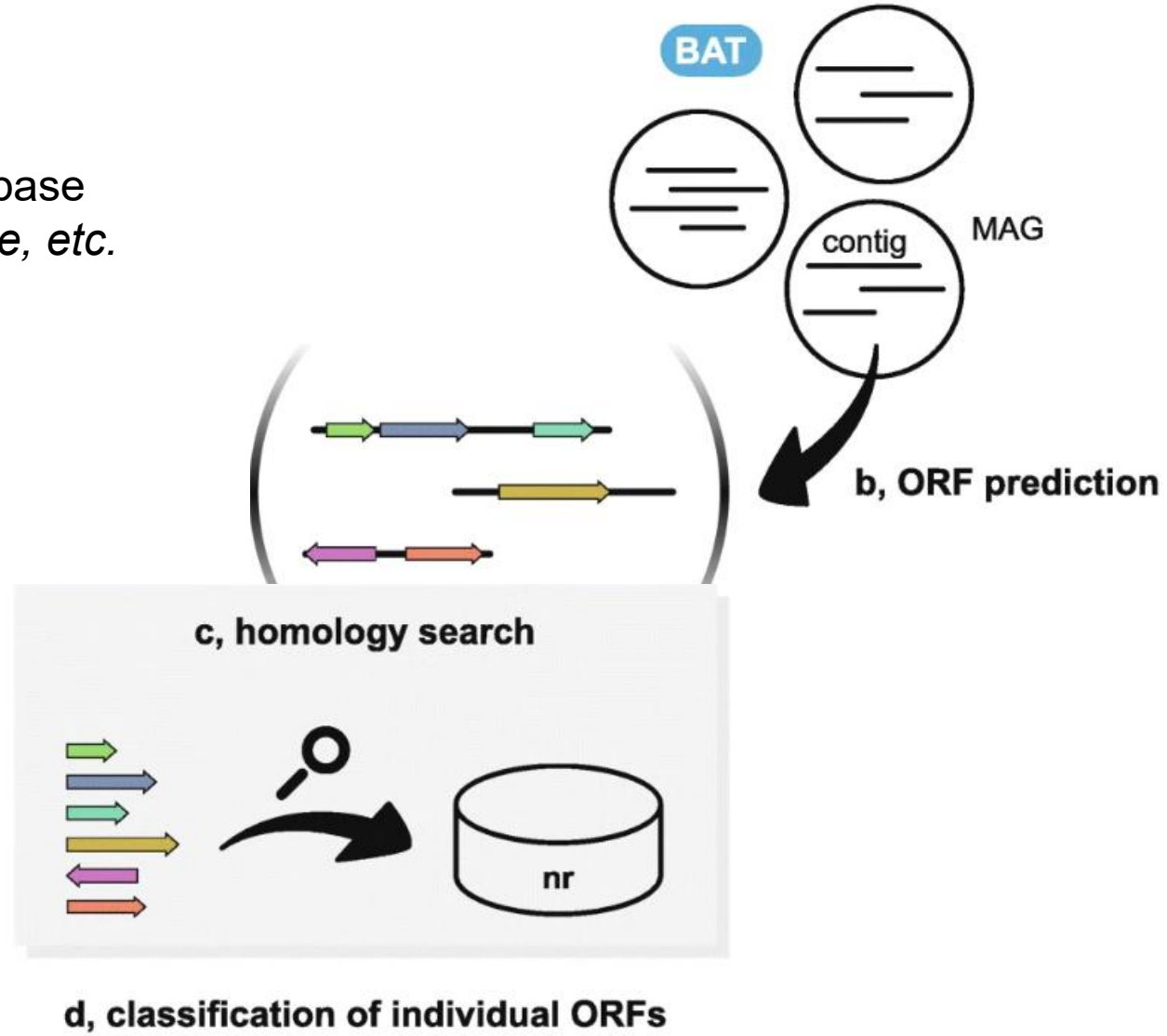
➡️ Matches to a **standard database of bacteria & archaea**

🔑 *Tells us "what species or genus this genome belongs to"*



Functional annotation of MAGs

🔬 **BAT (Bin Annotation Tool)**
➡️ Finds **genes in each MAG**
➡️ Compares them to **known proteins** in NCBI database
🔑 *Helps identify functions like metabolism, resistance, etc.*



Method | Open access | Published: 22 October 2019

Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT

F. A. Bastiaan von Meijenfeldt, Ksenia Arkhipova, Diego D. Cambuy, Felipe H. Coutinho & Bas E. Dutilh

Genome Biology **20**, Article number: 217 (2019) | Cite this article

25k Accesses | **435** Citations | **15** Altmetric | Metrics

https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1817-x

🌳 **PhyloPhlAn (Who's Related to Whom?)**
➡️ Builds **evolutionary trees** from genomes
➡️ Shows how each MAG is **related to others**
🔑 *Places unknown microbes into the tree of life*

🧩 **What You Give (Input)**
➡️ Your recovered MAGs (genomes from metagenome bins)

What It Uses (Database + Method)
➡️ Uses a **reference database** of known microbial genomes
➡️ Finds **marker genes** common across microbes
➡️ Builds a **phylogenetic tree** using these conserved genes

Article | Open access | Published: 19 May 2020

Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0

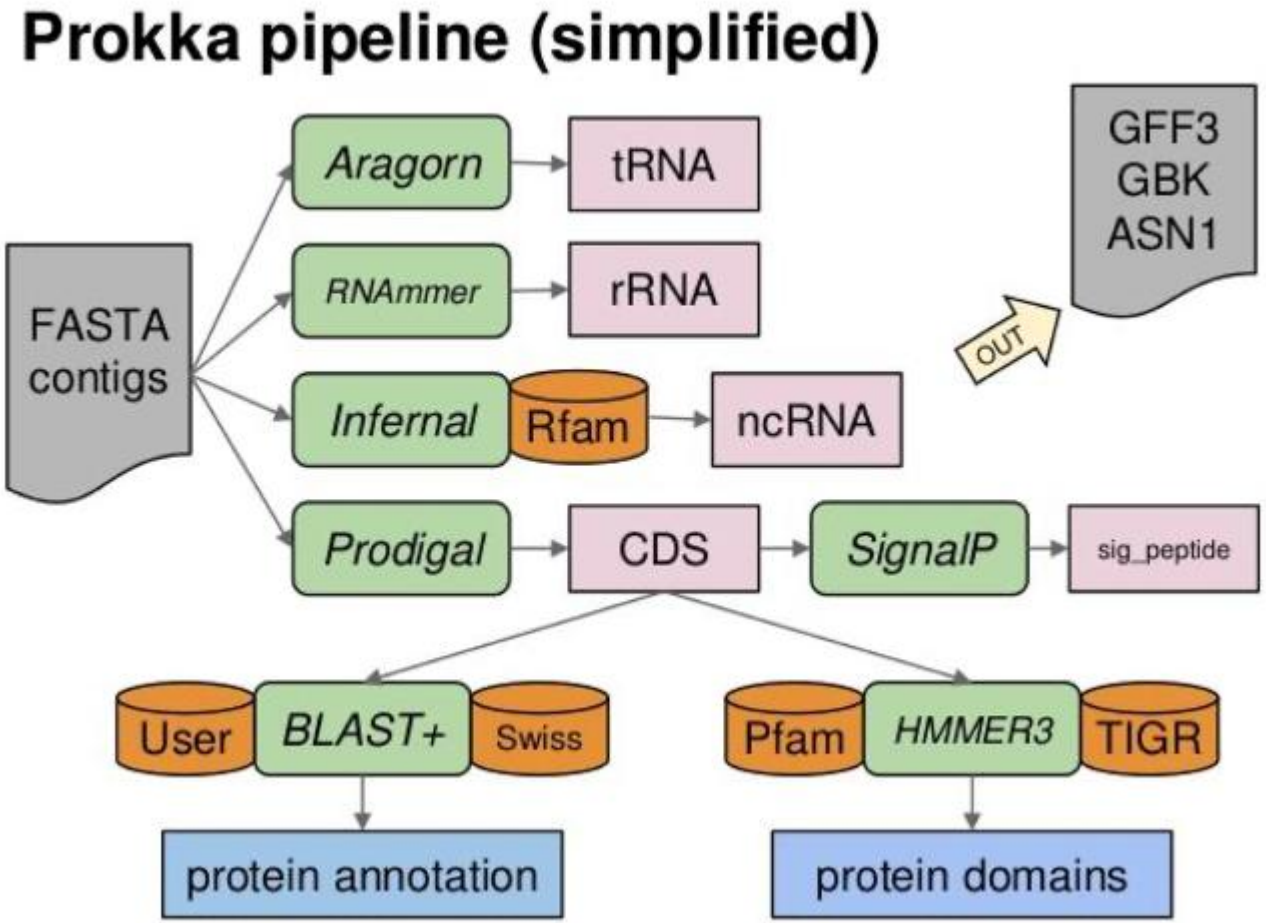
Francesco Asnicar, Andrew Maltez Thomas, Francesco Beghini, Claudia Mengoni, Serena Manara, Paolo Manghi, Qiyun Zhu, Mattia Bolzan, Fabio Cumbo, Uyen May, Jon G. Sanders, Moreno Zolfo, Evguenia Kopylova, Edoardo Pasolli, Rob Knight, Siavash Mirarab, Curtis Huttenhower & Nicola Segata

Nature Communications **11**, Article number: 2500 (2020) | Cite this article

38k Accesses | **636** Citations | **101** Altmetric | Metrics

Genome annotation

- **Goal:** Identify genes, assign functional roles, and link to biological pathways.
- **Tools:**
  - *Prokka* – Quick genome annotation using multiple databases
  - *eggNOG-mapper*– For pathway reconstruction and detailed function prediction
- **Why?** Translates raw sequences into biological meaning.



JOURNAL ARTICLE

Prokka: rapid prokaryotic genome annotation FREE

Torsten Seemann Author Notes

Bioinformatics, Volume 30, Issue 14, July 2014, Pages 2068–2069,
https://doi.org/10.1093/bioinformatics/btu153
Published: 18 March 2014 **Article history** ▾

https://doi.org/10.1093/bioinformatics/btu153

From Contigs to Functions — Gene Prediction

To identify genes from assembled contigs and link them to biological functions for deeper insights."

Why Gene Prediction?

•After assembling the metagenome, we need to **find genes hidden in the DNA**.
•These genes are the **functional units** that drive microbial activity.
•Once identified, we can:

- Annotate them with known **functions**
- Map to **pathways** (e.g., metabolism, resistance)
- Compare across samples or conditions

Tool : Prodigal

•Specialized for **prokaryotic gene prediction**
•Fast and accurate, widely used in metagenomic pipelines
•Predicts **coding sequences (CDS)** and **translation start sites**

To assign biological meaning to predicted genes by linking them to known functions, orthologs, and pathways

Tool: eggNOG-mapper

•A tool for **fast, genome-wide functional annotation**.

•Uses the **eggNOG database** (evolutionary genealogy of genes: Non-supervised Orthologous Groups).

•Assigns:

- **Orthologs**
- **Gene Ontology (GO)** terms
- **KEGG** pathways
- **COG** functional categories
- **Enzyme Commission (EC)** numbers

[Predicted Genes (from Prodigal)] →

➡️ **eggNOG-mapper** →

✅ Assigns **function**, **pathway**, **role**, **ortholog group**

➡️ Output: Annotated gene table

Gene ID	COG	Function	KEGG Pathway	GO Terms
gene_01	COG0012	Translation	Ribosome	GO:0006412

Summary of Tools for Functional annotation

Tools	Types	Publications	Core algorithms	Websites
MG-RAST	Homology-based	Keegan et al. 2016	Parallelized BLAT	http://api.metagenomics.anl.gov/api.html
eggNOG-mapper	Homology-based	Huerta-Cepas et al. 2017	Hidden Markov Model	http://eggnog-mapper.embl.de
GhostKOALA	Homology-based	Kanehisa et al. 2016	GHOSTX (seed search method)	http://www.kegg.jp/blastkoala/
InterProScan	Motif-based	Quevillon et al. 2005	Phobius (Hidden Markov Model)	http://www.ebi.ac.uk/InterProScan/

Table adapted from Yang et al, 2021

Eggnog-mapper for functional annotation

JOURNAL ARTICLE

eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale

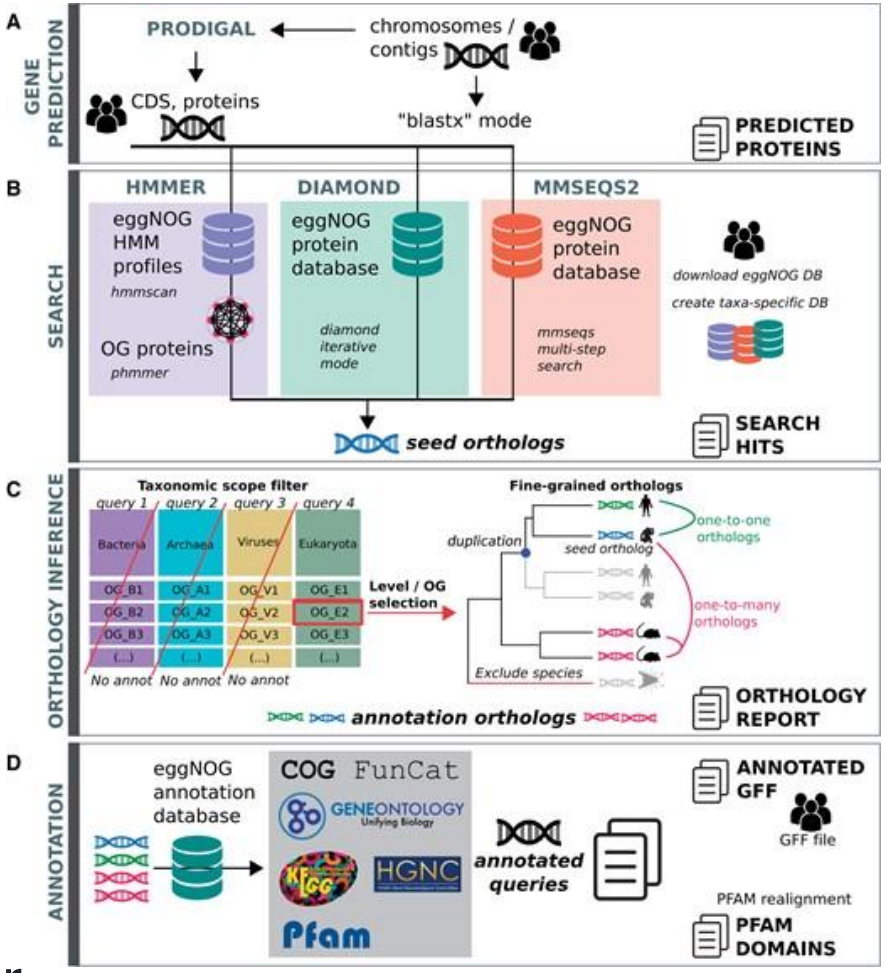
Carlos P Cantalapiedra, Ana Hernández-Plaza, Ivica Letunic, Peer Bork ✉,
Jaime Huerta-Cepas ✉

https://doi.org/10.1093/molbev/msab293

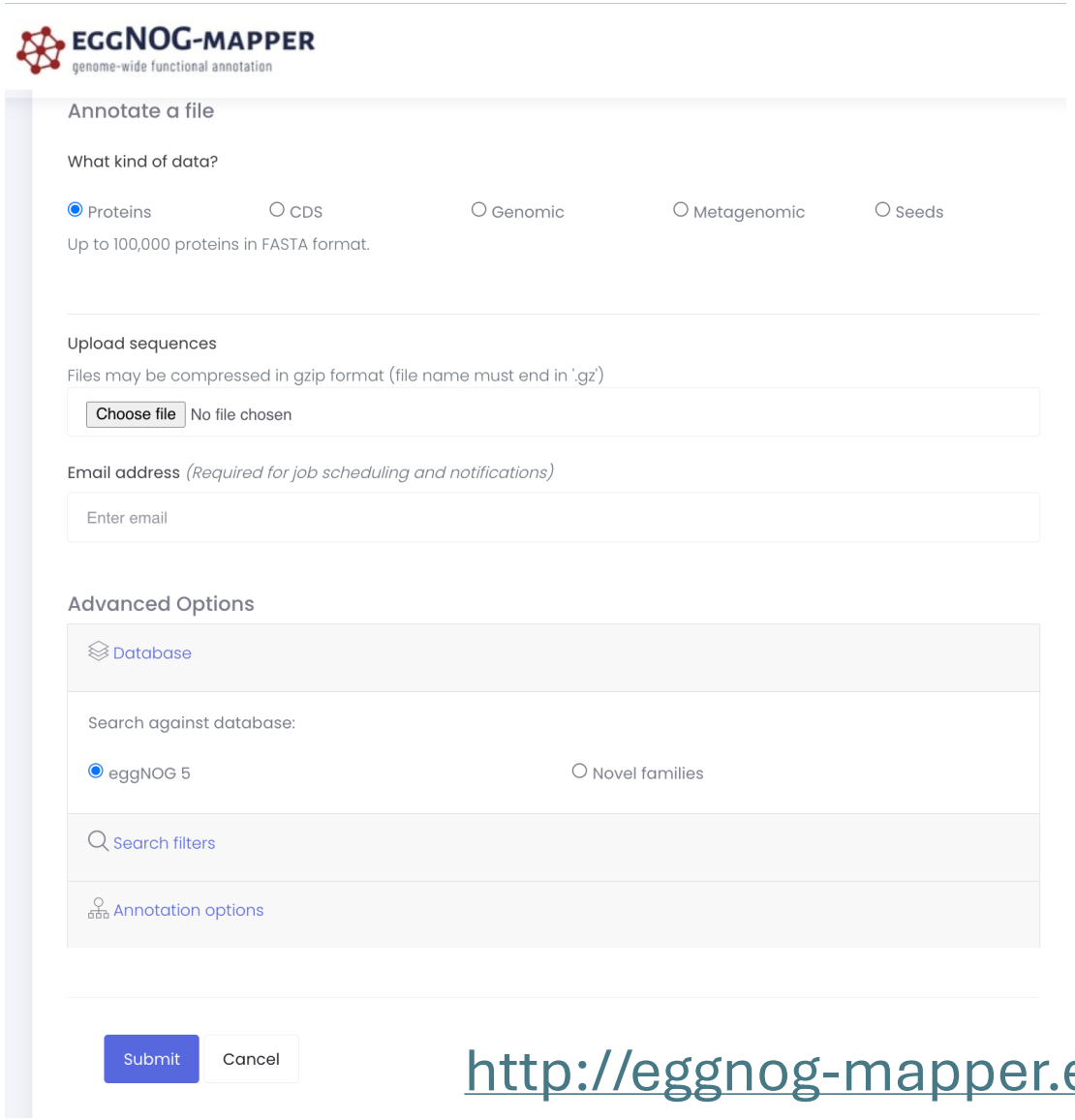
- *eggNog* database :
  - 5090 organisms, 2502 viruses.
  - 4.4 M orthologous groups annotated with COG category, Gene Ontolgy, EC number, Kegg orthologs and pathways, CAZy, PFAMs

EggNOG-mapper

- EggNOG-mapper is a tool for fast functional annotation of novel sequences.
- it uses precomputed orthologous groups and phylogenies from the eggNOG database (http://eggnog5.embl.de) to transfer functional information from fine-grained orthologs only.
- The use of orthology predictions for functional annotation permits a higher precision than traditional homology searches (i.e. BLAST searches), as it avoids transferring annotations from close paralogs.
- Common uses of eggNOG-mapper include the annotation of novel genomes, transcriptomes or even metagenomic gene catalogs.



Web-based interface



http://eggnog-mapper.embl.de/

query	seed_ortholog	evalue	score	eggNOG_OGs	max_annot_lvl	COG_category	Description	Preferred_name	GOs	EC	KEGG_ko	KEGG_Path
maker-406_1445_WH_illumina_pilon-augustus	42345.XP_008784385.1	4.89e-188	547.0	28KH1@1|root,2QSY8(	35493|Streptophyta	K	WRKY transcription factor	WRKY6	GO:0001067,GC	-	-	-
augustus-406_1445_WH_illumina_pilon-proce	42345.XP_008784479.1	2.19e-121	367.0	KOG0198@1|root,KOG	35493|Streptophyta	T	- STE kinases include homologs to sterile 7, sterile 1	-	-	-	ko:K20716	ko04016,m
maker-406_1445_WH_illumina_pilon-augustus	42345.XP_008790160.1	1.46e-181	520.0	28M37@1|root,2QTJX(	35493|Streptophyta	S	Anthranilate N-benzoyltransferase protein	-	GO:0003674,GC	2.3.1.133	ko:K13065	ko00940,k
maker-406_1445_WH_illumina_pilon-exonerat	4558.Sb02g002620.1	2.26e-17	91.7	28MYI@1|root,2QUH6	35493|Streptophyta	K	No apical meristem (NAM) protein	-	GO:0003674,GC	-	-	-
maker-706_1400_WH_illumina_pilon-augustus	42345.XP_008805098.1	0.0	965.0	2CMJX@1|root,2QQKV	35493|Streptophyta	G	Belongs to the glycosyltransferase 8 family	-	GO:0000003,GC	2.4.1.43	ko:K13648	ko00520,m
maker-706_1400_WH_illumina_pilon-augustus	42345.XP_008792746.1	2.35e-64	198.0	2AWEP@1|root,2RZYK(	35493|Streptophyta	U	May serve as docking site to facilitate the associati	MUB1	-	-	-	-
maker-706_1400_WH_illumina_pilon-augustus	42345.XP_008792748.1	0.0	1618.0	COG0542@1|root,KOG	35493|Streptophyta	O	Belongs to the ClpA ClpB family	-	GO:0005575,GC	-	ko:K03695	ko04213,m
maker-449_803_WH_illumina_pilon-augustus-(	4641.GSMUA_Achr8P23720_001	1.26e-53	174.0	COG0071@1|root,KOG	35493|Streptophyta	O	Belongs to the small heat shock protein (HSP20) fa	-	GO:0006950,GC	-	ko:K13993	ko04141,m
maker-449_803_WH_illumina_pilon-augustus-(	102107.XP_008242296.1	7e-150	449.0	COG5329@1|root,KOG	35493|Streptophyta	I	phosphoinositide phosphatase	-	GO:0003674,GC	-	-	-
maker-449_803_WH_illumina_pilon-augustus-(	42345.XP_008784359.1	5.81e-126	389.0	COG5329@1|root,KOG	35493|Streptophyta	I	SacI homology domain	-	GO:0003674,GC	-	-	-
augustus-449_803_WH_illumina_pilon-process	4641.GSMUA_Achr7P19810_001	2e-291	821.0	COG0438@1|root,KOG	35493|Streptophyta	M	Sucrose-cleaving enzyme that provides UDP-glucos	SUS1	GO:0000003,GC	2.4.1.13	ko:K00695	ko00500,k
augustus-449_803_WH_illumina_pilon-process	4641.GSMUA_Achr3P05810_001	4.67e-93	292.0	KOG1812@1|root,KOG	35493|Streptophyta	O	In Between Ring fingers	-	-	2.3.2.31	ko:K11975	-
maker-449_803_WH_illumina_pilon-augustus-(	4641.GSMUA_Achr3P05810_001	1.07e-82	259.0	KOG1812@1|root,KOG	35493|Streptophyta	O	In Between Ring fingers	-	-	2.3.2.31	ko:K11975	-
maker-449_803_WH_illumina_pilon-augustus-(	4641.GSMUA_Achr3P10120_001	2.57e-151	451.0	28J92@1|root,2QRMQ	35493|Streptophyta	K	helix loop helix domain	-	-	-	-	-
augustus-449_803_WH_illumina_pilon-process	4641.GSMUA_Achr9P14270_001	2.21e-76	236.0	KOG1812@1|root,KOG	35493|Streptophyta	O	IBR domain, a half RING-finger domain	-	-	2.3.2.31	ko:K11975	-
maker-449_803_WH_illumina_pilon-augustus-(	38727.Pavir.J11565.1.p	7.89e-20	87.8	COG5253@1|root,KOG	35493|Streptophyta	T	1-phosphatidylinositol-3-phosphate 5-kinase	-	GO:0000285,GC	2.7.1.150	ko:K00921	ko00562,k
maker-449_803_WH_illumina_pilon-exonerate	3712.Bo1g128650.1	4.81e-37	129.0	COG5253@1|root,KOG	35493|Streptophyta	T	1-phosphatidylinositol-3-phosphate 5-kinase	-	GO:0000285,GC	2.7.1.150	ko:K00921	ko00562,k
augustus-449_803_WH_illumina_pilon-process	42345.XP_008804527.1	1.01e-50	173.0	KOG4197@1|root,KOG	35493|Streptophyta	S	Protein of unknown function (DUF1685)	-	-	-	-	-
augustus-449_803_WH_illumina_pilon-process	42345.XP_008796915.1	1.09e-69	248.0	2CMHM@1|root,2QQD	35493|Streptophyta	S	Leucine-rich repeat	-	GO:0000902,GC	-	-	-
maker-449_803_WH_illumina_pilon-augustus-(	4432.XP_010278420.1	3.59e-76	245.0	28IMD@1|root,2QUZ4	35493|Streptophyta	S	Protein of unknown function, DUF547	-	-	-	-	-
maker-449_803_WH_illumina_pilon-augustus-(	4641.GSMUA_Achr7P15740_001	3.56e-138	413.0	28KG7@1|root,2QSXE(	35493|Streptophyta	K	MYB-CC type transfactor, LHEQLE motif	-	GO:0003674,GC	-	-	-
maker-449_803_WH_illumina_pilon-augustus-(	4641.GSMUA_Achr4P09650_001	3.1e-297	817.0	2CMCP@1|root,2QPZ6	35493|Streptophyta	G	Belongs to the glycosyl hydrolase 17 family	-	GO:0001871,GC	3.2.1.39	ko:K19893	ko00500,m
augustus-449_803_WH_illumina_pilon-process	29730.Gorai.011G106000.1	3.95e-69	224.0	KOG0198@1|root,KOG	35493|Streptophyta	T	Mitogen-activated protein kinase kinase kinase	-	GO:0000226,GC	-	-	-
maker-449_803_WH_illumina_pilon-augustus-(	42345.XP_008809099.1	3.95e-306	846.0	COG0318@1|root,KOG	35493|Streptophyta	I	AMP-binding enzyme C-terminal domain	-	GO:0001676,GC	6.2.1.12	ko:K01904	ko00130,k
augustus-449_803_WH_illumina_pilon-process	29730.Gorai.011G106000.1	3.77e-67	218.0	KOG0198@1|root,KOG	35493|Streptophyta	T	Mitogen-activated protein kinase kinase kinase	-	GO:0000226,GC	-	-	-
augustus-546_2426_WH_illumina_pilon-proce	4641.GSMUA_Achr5P19900_001	3.3e-69	216.0	28QVS@1|root,2QXIN(	35493|Streptophyta	S	Plant invertase/pectin methylesterase inhibitor	-	GO:0003674,GC	-	-	-
augustus-546_2426_WH_illumina_pilon-proce	4641.GSMUA_Achr5P19900_001	5.62e-70	219.0	28QVS@1|root,2QXIN(	35493|Streptophyta	S	Plant invertase/pectin methylesterase inhibitor	-	GO:0003674,GC	-	-	-
augustus-546_2426_WH_illumina_pilon-proce	4641.GSMUA_Achr4P29040_001	0.0	961.0	28I6U@1|root,2QRZJ@	35493|Streptophyta	S	Putative S-adenosyl-L-methionine-dependent met	-	GO:0003674,GC	-	-	-
augustus-546_2426_WH_illumina_pilon-proce	4641.GSMUA_Achr5P15240_001	5.42e-39	146.0	KOG0800@1|root,KOG	35493|Streptophyta	U	Autophagy-related protein	-	GO:0005575,GC	-	ko:K08331	ko04136,k
maker-546_2426_WH_illumina_pilon-augustus	4641.GSMUA_Achr4P24730_001	4.59e-131	407.0	COG5059@1|root,KOG	35493|Streptophyta	Z	Belongs to the TRAFAC class myosin-kinesin ATPase	-	GO:0000166,GC	-	ko:K11498	-
maker-546_2426_WH_illumina_pilon-augustus	4641.GSMUA_Achr5P19930_001	2.43e-76	246.0	COG5641@1|root,KOG	35493|Streptophyta	K	GATA transcription factor	-	GO:0000976,GC	-	ko:K21630	-
augustus-546_2426_WH_illumina_pilon-proce	42345.XP_008791643.1	3.05e-233	686.0	COG5059@1|root,KOG	35493|Streptophyta	Z	Belongs to the TRAFAC class myosin-kinesin ATPase	-	GO:0003674,GC	-	ko:K11498	-
augustus-546_2426_WH_illumina_pilon-proce	42345.XP_008779673.1	6.11e-30	120.0	28ITC@1|root,2QR4P@	35493|Streptophyta	S	Plant protein of unknown function	-	-	-	-	-
augustus-442_1224_WH_illumina_pilon-proce	4537.OPUNC05G15020.3	4.66e-133	392.0	COG0814@1|root,KOG	35493|Streptophyta	E	Transmembrane amino acid transporter protein	-	-	-	-	-
augustus-442_1224_WH_illumina_pilon-proce	42345.XP_008783335.1	4.16e-170	499.0	28I6U@1|root,2QQH0(	35493|Streptophyta	S	methyltransferase PMT26	-	GO:0003674,GC	-	-	-
maker-442_1224_WH_illumina_pilon-augustus	4641.GSMUA_Achr11P26330_001	3.1e-306	864.0	COG1874@1|root,KOG	35493|Streptophyta	G	Beta-galactosidase	-	GO:0000003,GC	-	-	-
augustus-442_1224_WH_illumina_pilon-proce	42345.XP_008783354.1	5.94e-246	687.0	COG0534@1|root,KOG	35493|Streptophyta	V	Belongs to the multi antimicrobial extrusion (MAT	-	GO:0003674,GC	-	ko:K03327	-
maker-442_1224_WH_illumina_pilon-augustus	4641.GSMUA_Achr6P18790_001	0.0	1022.0	KOG1278@1|root,KOG	35493|Streptophyta	U	Belongs to the nonaspanin (TM9SF) (TC 9.A.2) fami	-	-	-	ko:K17087	-
maker-442_1224_WH_illumina_pilon-exonerat	42345.XP_008780246.1	6.35e-36	122.0	2CZIP@1|root,2SAJ5@	35493|Streptophyta	S	Gibberellin regulated protein	-	GO:0001101,GC	-	-	-
augustus-442_1224_WH_illumina_pilon-proce	42345.XP_008782769.1	6.92e-69	221.0	KOG0198@1|root,KOG	2759|Eukaryota	T	Mitogen-activated protein kinase kinase kinase	-	GO:0000226,GC	-	-	-
maker-442_1224_WH_illumina_pilon-augustus	42345.XP_008793917.1	1.52e-99	313.0	28JYJ@1|root,2QVP0@	35493|Streptophyta	K	Auxin response factor	-	-	-	-	-
maker-442_1224_WH_illumina_pilon-augustus	4641.GSMUA_Achr5P01120_001	2.98e-123	368.0	2C3EN@1|root,2QQ00	35493|Streptophyta	S	Zinc finger protein	-	-	-	-	-

KEGG (Kyoto Encyclopedia for Genes and Genomes)

JOURNAL ARTICLE

KEGG: Kyoto Encyclopedia of Genes and Genomes

Minoru Kanehisa, Susumu Goto

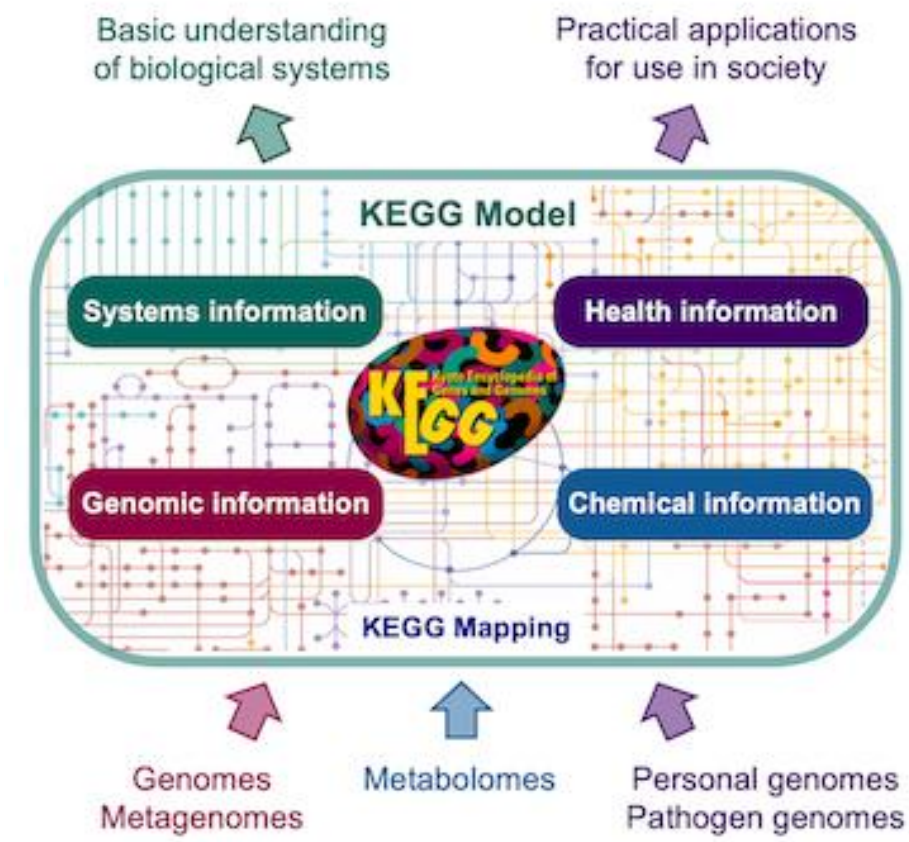
Nucleic Acids Research, Volume 28, Issue 1, 1 January 2000, Pages 27–30,
https://doi.org/10.1093/nar/28.1.27
Published: 01 January 2000

JOURNAL ARTICLE

KEGG: new perspectives on genomes, pathways, diseases and drugs

Minoru Kanehisa ✉, Miho Furumichi, Mao Tanabe, Yoko Sato, Kanae Morishima

Nucleic Acids Research, Volume 45, Issue D1, January 2017, Pages D353–D361,
https://doi.org/10.1093/nar/gkw1092
Published: 29 November 2016 **Article history** ▾



https://www.genome.jp/kegg/

GhostKOALA

https://www.kegg.jp/ghostkoala/

Automatic KO assignment and KEGG mapping service

BlastKOALA | GhostKOALA | KofamKOALA

KOALA job status 2023/12/11 16:42:27 (GMT+9)

	Blast	Ghost	Kofam
Number of jobs in the queue	2	2	2
Submission of last completed job	2023/12/11 16:13:31	2023/12/11 16:31:51	2023/12/11 16:08:39

GhostKOALA assigns K numbers to the user's sequence data by GHOSTX search against a nonredundant set of KEGG GENES [1]. This service will be re-evaluated due to the increasing size of the dataset, which is an order of magnitude larger than that for the new version of BlastKOALA. See Step-by-step Instructions.

Upload query amino acid sequences in FASTA format

Enter FASTA sequences

Or upload file: Choose File No file chosen

Your query data consisting of multiple amino acid sequences will be given K numbers by GhostKOALA. The file size of up to 300 MB with the limit of 500,000 sequences may be uploaded.

Enter KEGG GENES database file to be searched

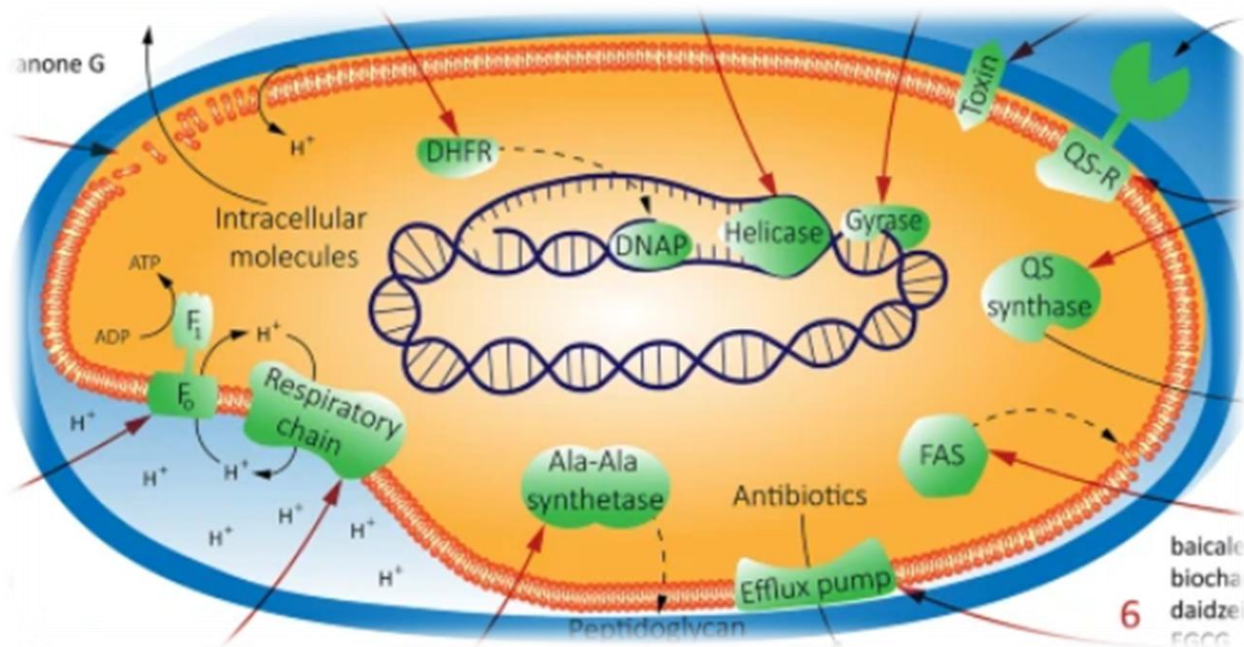
- genus_prokaryotes + virusess
- genus_prokaryotes + family_eukaryotes + viruses

Each of the nonredundant datasets, genus_prokaryotes and family_eukaryotes, contains about 10 million sequences.

Enter your email address

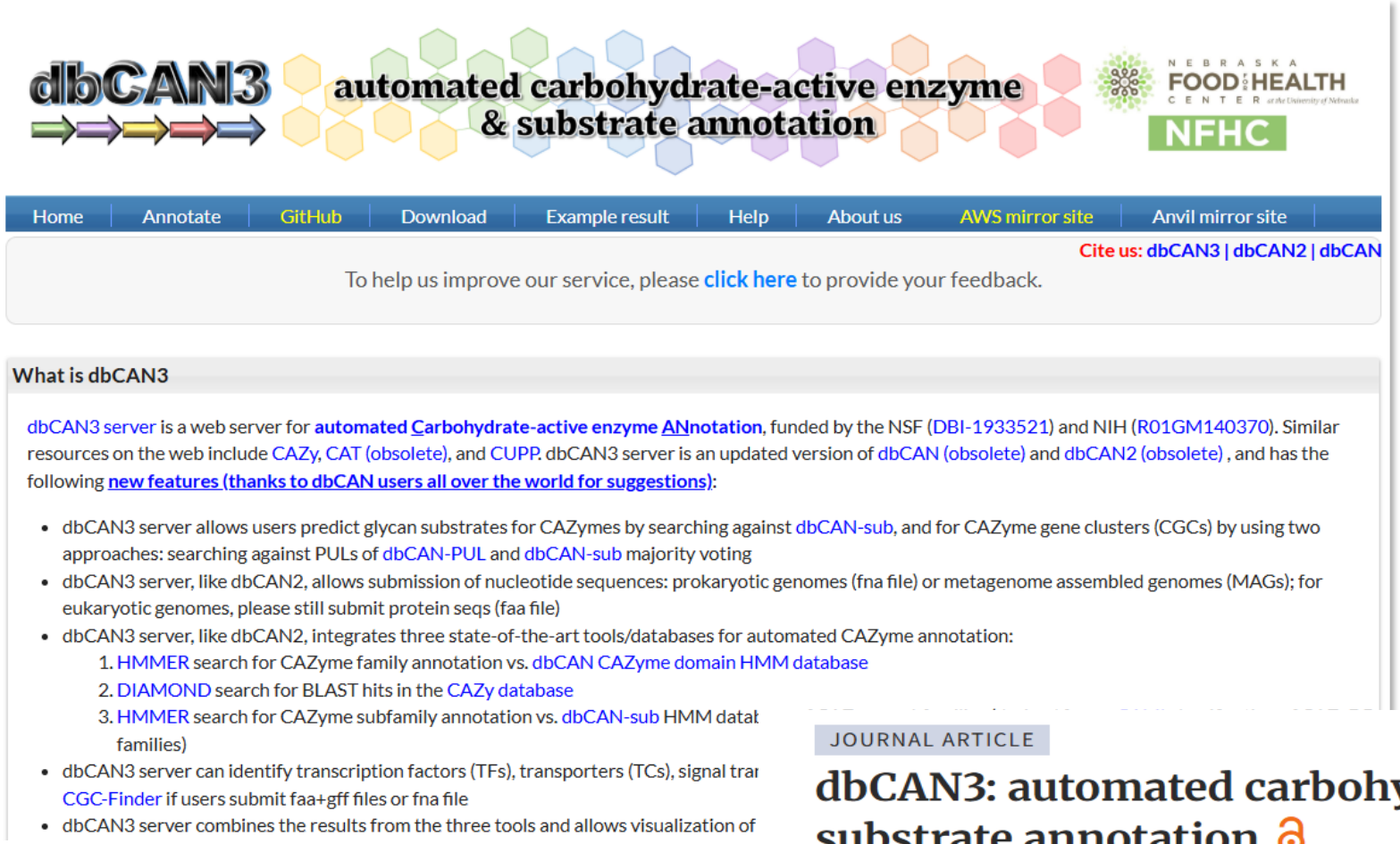
Request for email confirmation | Clear

Paste gene sequences here/ or upload file


anone G
H+
DHFR
Toxin
QS-R
Intracellular molecules
DNAP
Helicase
Gyrase
ATP
ADP
F1
Fo
H+
Respiratory chain
QS synthase
FAS
Ala-Ala synthetase
Antibiotics
Efflux pump
Peptidoglycan
6

Carbohydrate-metabolizing genes in Bacteria

bcb.unl.edu/dbCAN2/



Visit: https://bcb.unl.edu/dbCAN2/

JOURNAL ARTICLE

dbCAN3: automated carbohydrate-active enzyme and substrate annotation

Jinfang Zheng , Qiwei Ge , Yuchen Yan , Xinpeng Zhang , Le Huang , Yanbin Yin
Author Notes

Nucleic Acids Research, Volume 51, Issue W1, 5 July 2023, Pages W115–W121,
https://doi.org/10.1093/nar/gkad328
Published: 01 May 2023 **Article history**

Choose Sequence Type:

○ Protein sequence (example) ? ◉ Nucleotide sequence (example) ?

Choose Nucletotide Sequence Type: ?

◉ Complete/draft prokaryote genomes ○ mRNAs/CDSs/Metagenomes or short DNA seqs

Select Tools To Run:

☑ HMMER: dbCAN (E-Value < 1e-15, coverage > 0.35) ☐ DIAMOND: CAZy (E-Value < 1e-102) ☐ HMMER: dbCAN-sub (E-Value < 1e-15, coverage > 0.35) ☐ CGCFinder (Distance <= 2, signature genes = CAZyme+TC)?

Just paste some sequences here (note: only FASTA format please!!!)

Try example sequences

Or upload a fasta file including many sequences (Max File Size: 10MB)

Fasta file: Choose File MGYG000000099.fna

Submit

dbCAN3

automated carbohydrate-active enzyme & substrate annotation

NEBRASKA FOOD FOR HEALTH CENTER at the University of Nebraska
NFHC

Cite us: dbCAN3 | dbCAN2 | dbCAN

To help us improve our service, please click here to provide your feedback.

Your request was received and is **running** now. You will be notified by email once it is done if you entered an email.
You can also choose to stay at this page, which will be automatically refreshed every **5** seconds.

dbCAN3 automated carbohydrate-active enzyme & substrate annotation

NEBRASKA FOOD FOR HEALTH CENTER NFHC

Home | Annotate | GitHub | Download | Example result | Help | About us | AWS mirror site | Anvil mirror site

Cite us: dbCAN3 | dbCAN2 | dbCAN

To help us improve our service, please click here to provide your feedback.

Result of job: 20250726143838

Overview | HMMER: dbCAN

Download SignalP output | Download Prodigal predictions | Download CAZyme sequences | Download this table (keep those with # of Tools >=2 will give you best result; and use dbCAN domain assignment is recommended) ?

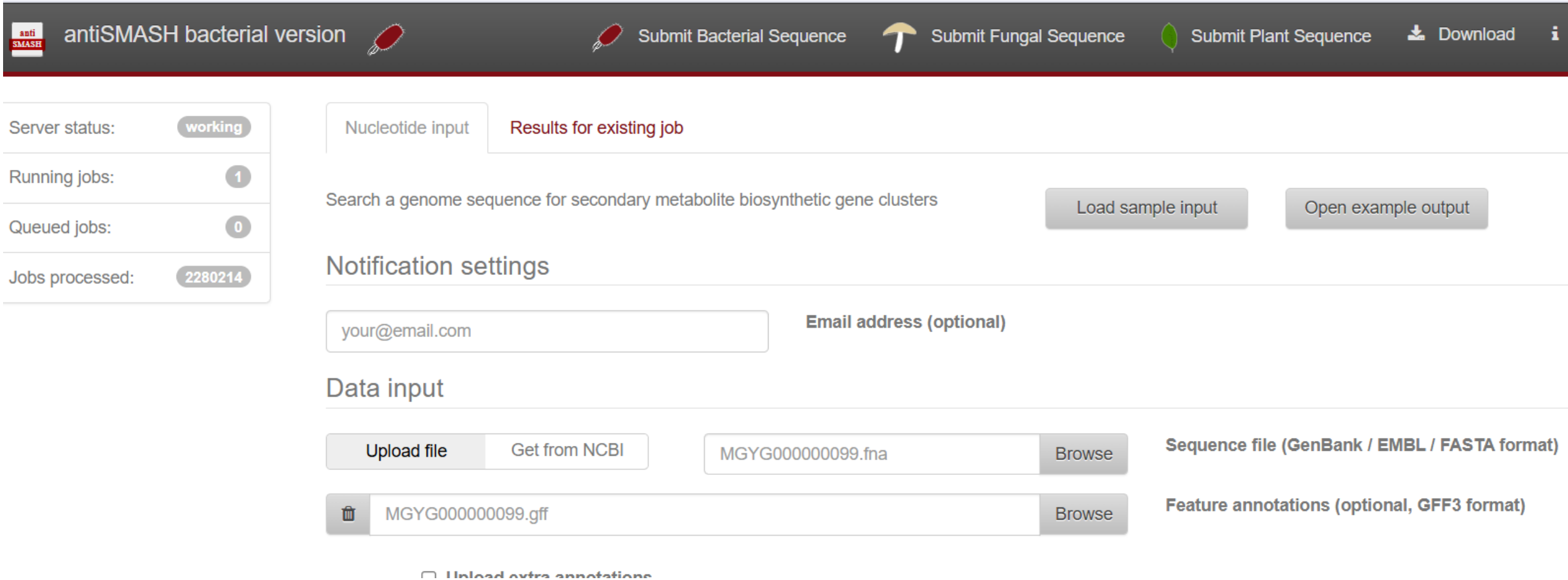
Show All entries Search:

Gene ID	EC#	HMMER	DIAMOND	dbCAN_sub	Signal Peptide	# of Tools
MGYG000000099_10_102	-	GT26(55-222)	-	-	N	1
MGYG000000099_10_68	-	GT119(10-375)	-	-	N	1
MGYG000000099_10_69	-	GT28(194-358)	-	-	N	1
MGYG000000099_13	-	GH188(7-222)	-	-	N	1

Gene ID	EC#	HMMER	dbCAN_sub	DIAMOND	Signalp	#ofTools
MGYG000000099_10_102	-	GT26(55-222)	-	-	N	1
MGYG000000099_10_68	-	GT119(10-375)	-	-	N	1
MGYG000000099_10_69	-	GT28(194-358)	-	-	N	1
MGYG000000099_13_10	-	GH188(7-222)	-	-	N	1
MGYG000000099_15_6	-	GH188(3-141)	-	-	N	1
MGYG000000099_15_9	-	GH179(17-342)	-	-	N	1
MGYG000000099_1_1005	-	GT2(5-123)	-	-	N	1
MGYG000000099_1_1034	-	GT1(12-386)	-	-	N	1
MGYG000000099_1_1071	-	GH4(3-179)	-	-	N	1
MGYG000000099_1_113	-	GT51(96-262)	-	-	N	1
MGYG000000099_1_17	-	GH77(10-489)	-	-	N	1
MGYG000000099_1_207	-	GT119(29-385)	-	-	N	1
MGYG000000099_1_326	-	GT76(91-274)	-	-	N	1
MGYG000000099_1_448	-	GH18(174-406)	-	-	N	1
MGYG000000099_1_912	-	GH25(51-223)	-	-	Y(1-67)	1
MGYG000000099_1_92	-	CE9(32-368)	-	-	N	1
MGYG000000099_1_955	-	GH31_10(247-697)+CBM35(850-979)+CBM35(989-1111)+CBM61(1123-1268)	-	-	Y(1-31)	1
MGYG000000099_1_956	-	GH31_7(356-817)+CBM35(955-1073)	-	-	Y(1-24)	1
MGYG000000099_1_959	-	GH31_7(121-589)	-	-	N	1
MGYG000000099_1_964	-	GH125(394-586)	-	-	N	1
MGYG000000099_1_965	-	GH65(301-675)	-	-	N	1
MGYG000000099_1_97	-	GT4(193-340)	-	-	N	1
MGYG000000099_2_251	-	GT4(185-336)	-	-	N	1
MGYG000000099_2_257	-	GT28(195-340)	-	-	N	1
MGYG000000099_2_332	-	CE4(56-167)	-	-	Y(1-38)	1
MGYG000000099_2_400	-	GH3(125-357)	-	-	Y(1-22)	1
MGYG000000099_2_448	-	GH188(3-199)	-	-	N	1
MGYG000000099_2_94	-	GT2(7-174)	-	-	N	1
MGYG000000099_3_259	-	CBM48(25-109)+GH13_9(176-480)	-	-	N	1
MGYG000000099_3_262	-	GT5(2-472)	-	-	N	1
MGYG000000099_3_263	-	GT35(93-806)	-	-	N	1
MGYG000000099_3_271	-	GT39(239-456)	-	-	N	1
MGYG000000099_3_363	-	GT2(49-274)	-	-	N	1
MGYG000000099_3_419	-	CBM54(198-312)	-	-	Y(1-36)	1
MGYG000000099_3_59	-	GT28(207-354)	-	-	N	1
MGYG000000099_4_97	-	GT2(8-155)	-	-	N	1
MGYG000000099_5_125	-	GH13_39(175-508)	-	-	N	1
MGYG000000099_5_126	-	GT2(7-173)	-	-	N	1
MGYG000000099_5_62	-	GT119(233-553)	-	-	N	1
MGYG000000099_6_154	-	CE7(5-316)	-	-	N	1
MGYG000000099_6_155	-	CE7(131-272)	-	-	N	1
MGYG000000099_6_156	-	GH177(53-274)	-	-	N	1
MGYG000000099_6_157	-	CBM40(58-221)+GH33(239-687)+CBM40(729-900)+CBM40(928-1087)	-	-	Y(1-33)	1
MGYG000000099_7_102	-	GT2(1477-1611)	-	-	N	1
MGYG000000099_7_141	-	CE4(47-168)	-	-	Y(1-34)	1
MGYG000000099_7_36	-	CE9(17-379)	-	-	N	1

Exploring Bacterial Secondary Metabolites

- Discover novel antibiotics or bioactive compounds
- Study metabolic potential of microbes



Server status:	working
Running jobs:	2
Queued jobs:	0
Jobs processed:	2280214

Status of job MGYG000000099.fna analysis

Submitted: Jul 27, 2025 01:23:39
Status: running: Comparing regions to reference database

Last status change: Jul 27, 2025 01:24:00

Important: If you did not provide an email address, please bookmark this page so you can access your antiSMASH results later.

If you specified an email address on the job submission page, you will be notified by email once the job is complete.

Running antiSMASH on a complete genome will take a couple of hours, and depending on the server load it will take a while for your job to start.

If you have found antiSMASH useful, please cite us. antiSMASH is free to use for everybody to use, even commercially. See our license and terms for details.

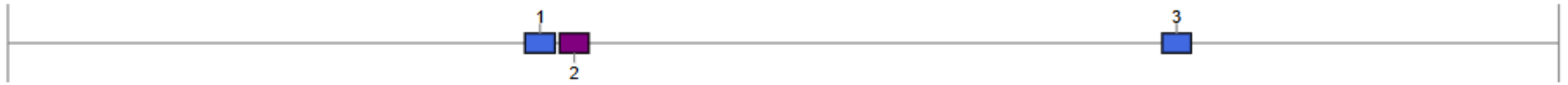
antiSMASH version 8.0.2

Select genomic region:

Overview | 1.1 | 1.2 | 1.3

Identified secondary metabolite regions using strictness 'relaxed'

MGYG000000099_1



Region	Type	From	To	Similarity Confidence	Most similar known cluster
Region 1.1	ranthipeptide	369,879	391,456		
Region 1.2	terpene-precursor	394,736	415,617		
Region 1.3	cyclic-lactone-autoinducer	825,822	846,546		

No secondary metabolite regions were found in these records:

MGYG000000099_2
MGYG000000099_3
MGYG000000099_4
MGYG000000099_5
MGYG000000099_6
MGYG000000099_7
MGYG000000099_8
MGYG000000099_9

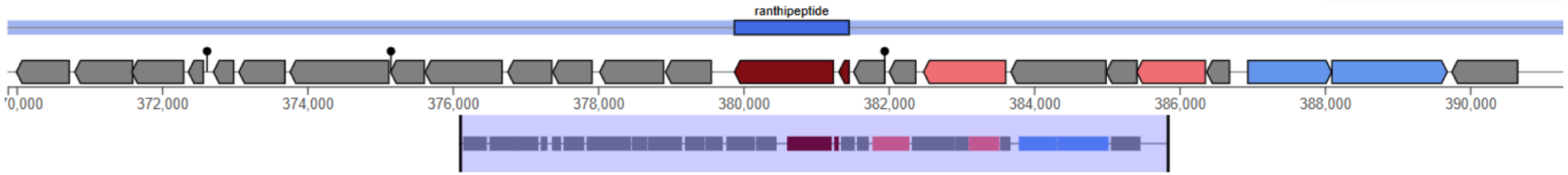
Select genomic region:

Overview | 1.1 | 1.2 | 1.3

MGYG000000099_1 - Region 1 - ranthipeptide

Location: 369,879 - 391,456 nt. (total: 21,578 nt) Show pHMM detection rules used

Download region SVG | Download region GenBank file



Legend:

core biosynthetic genes | additional biosynthetic genes | transport-related genes | regulatory genes | other genes | resistance | binding site

reset view | zoom to selection

Gene overview | Tailoring | KnownClusterBlast | SubClusterBlast | TFBS Finder

Gene/CDS overview

Filter:

Automatically zoom to filtered/selected features

Identifier	Product	Length NT	Length AA	Function	Sequence NT	Sequence AA	NCBI Blast
MGYG000000099_00396	Queuine tRNA-ribosyltransferase	1151	376	additional	Copy	Copy	BlastP
MGYG000000099_00397	Thiol-disulfide oxidoreductase ResA	1311	436	other	Copy	Copy	BlastP
MGYG000000099_00398	hypothetical protein	423	140	other	Copy	Copy	BlastP
MGYG000000099_00399	Thioredoxin reductase	945	314	biosynthetic-additional	Copy	Copy	BlastP
MGYG000000099_00400	Thioredoxin C-1	315	104	other	Copy	Copy	BlastP
MGYG000000099_00401	Choline transport ATP-binding protein OpuBA	1155	384	transport	Copy	Copy	BlastP

Gene details

MGYG000000099_00401
Choline transport ATP-binding protein OpuBA

Locus tag: MGYG000000099_00401
Protein ID: None
Gene: opuBA_1
Location: 386,943 - 388,097, (total: 1155 nt)

transport (smcogs) SMCOG1000: ABC transporter ATP-binding protein

NCBI BlastP on this gene
Blast against antiSMASH-database
MIBiG Hits
TransportDB BLAST on this gene

AA sequence: Copy to clipboard
Nucleotide sequence: Copy to clipboard

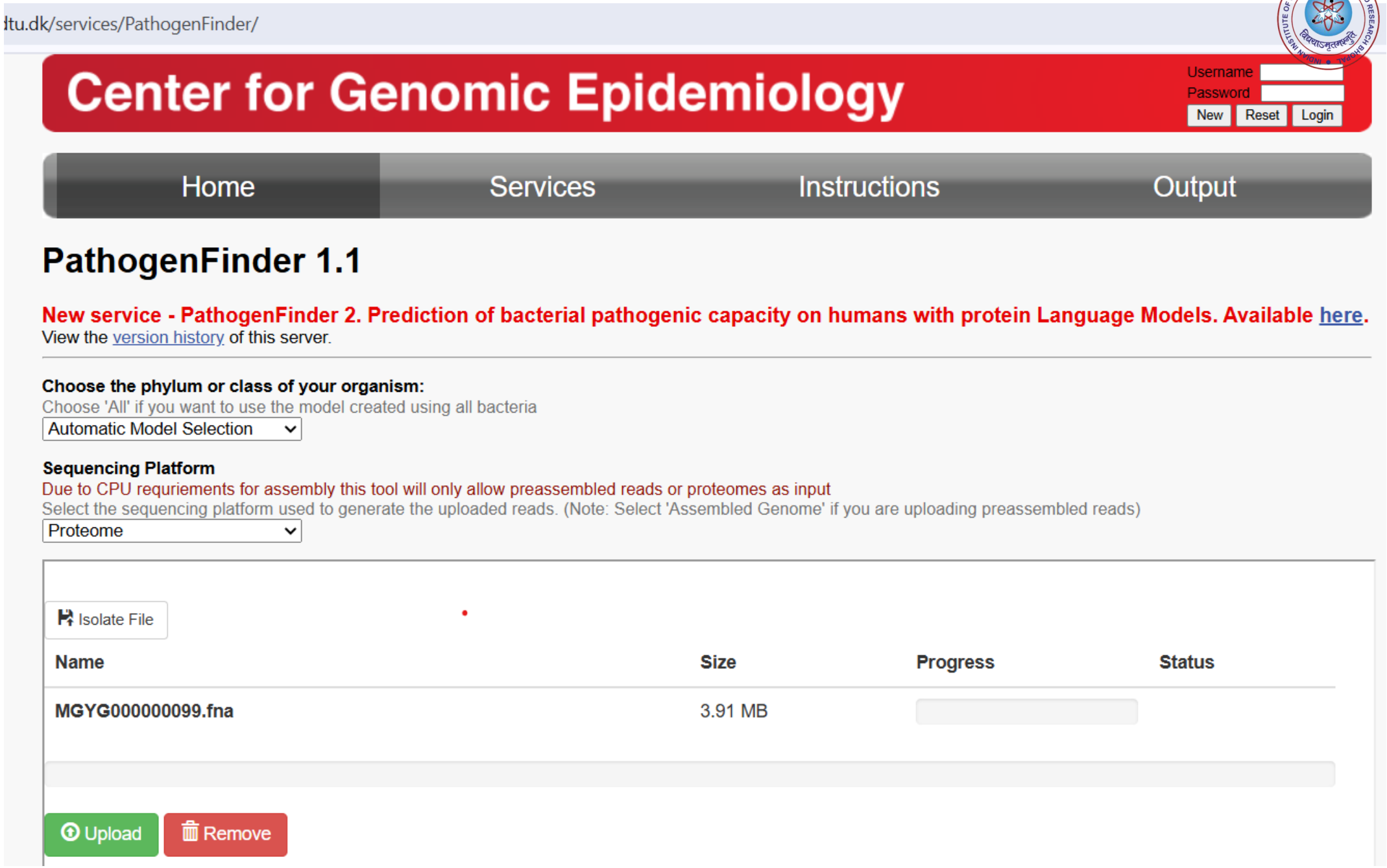
RREFinder | TFBS Finder

RRE predictions

ranthipeptide protocluster (369878...391456)
MGYG000000099_00392

http://dx.doi.org/10.1371/journal.pone.0077302

https://cge.food.dtu.dk/services/PathogenFinder/



Center for Genomic Epidemiology

Home | Services | Instructions | Output

The input organism was predicted as human pathogen

Probability of being a human pathogen	0.564
Input proteome coverage (%)	0.13
Matched Pathogenic Families	3
Matched Not Pathogenic Families	2

Sequences	3830
Total bpp	1179428
Longest seq	3127
Shortest seq	30
Avg seq lenght	307.0

Input Sequence: MGYG000000099_2_523 # 567188 # 569521 # -1 # ID=2_523;partial=00;start_type=ATG;rbs_motif=AGGAGG;rbs_spacer=11-12bp;gc_cont=0.517

Matched Family

PROJECT ID	ACCESSION ID	ORGANISMS	CLASS	PROTEIN FUNCTION	PROTEIN ID	%IDENTITY
18737	CP000837	Streptococcus suis GZ1, complete genome.	Lactobacillales	TrsE-like protein	ADE30946	87.52

Input Sequence: MGYG000000099_8_54 # 45751 # 46941 # -1 # ID=8_54;partial=00;start_type=ATG;rbs_motif=AGGAG;rbs_spacer=5-10bp;gc_cont=0.560

Matched Family

PROJECT ID	ACCESSION ID	ORGANISMS	CLASS	PROTEIN FUNCTION	PROTEIN ID	%IDENTITY
17419	CP001348	Clostridium cellulolyticum H10, complete genome.	Clostridia	transposase mutator type	ACL74802	83.08

MG RAST

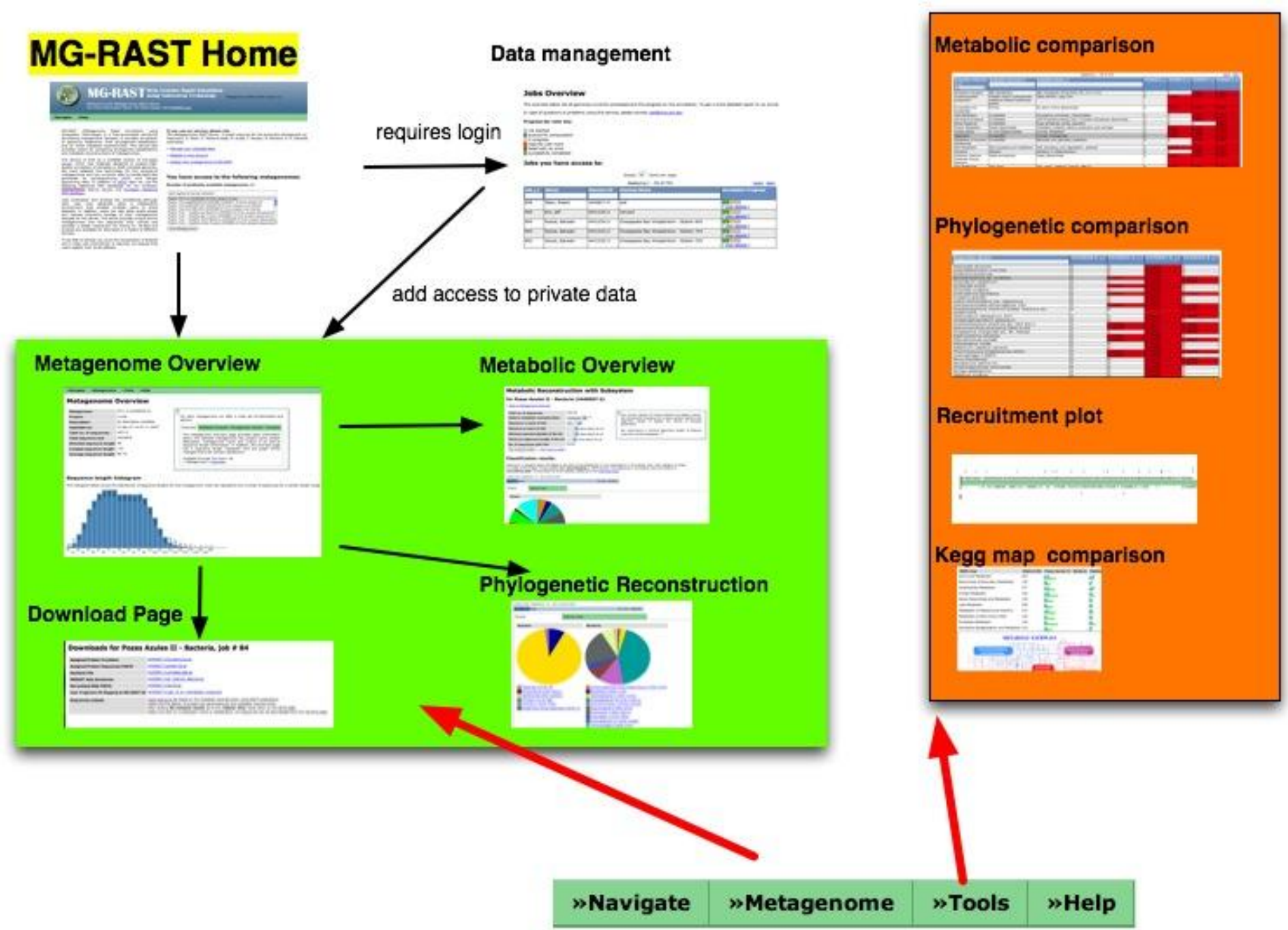
MG-RAST Tutorial

Getting started

- Web Page: http://metagenomics.anl.gov/
  - Register to run you own analysis
  - Publicly available datasets can be explored without registering.

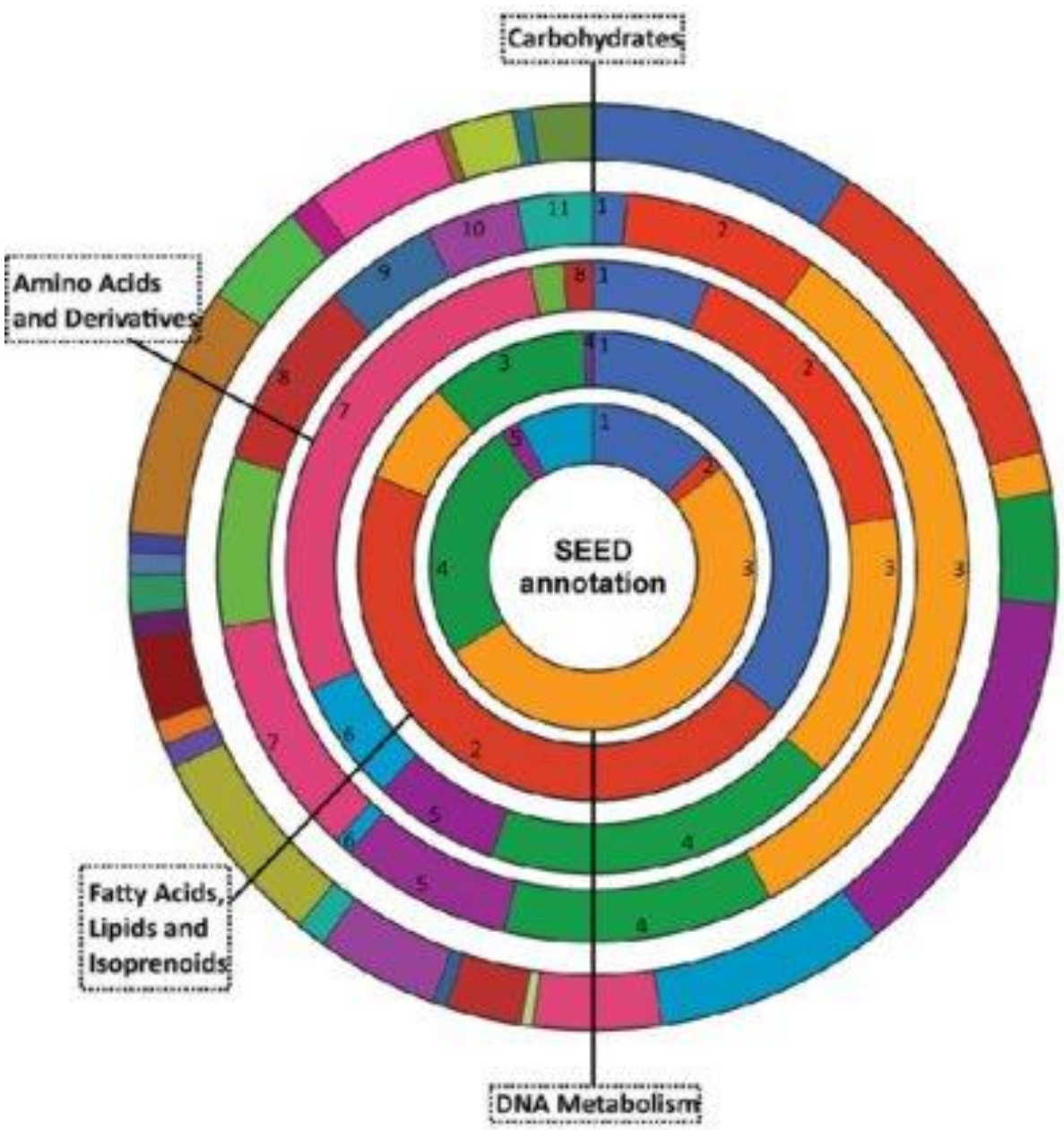
Web Interface

- **Home Page**
  - The home page has search box and options to upload, download and analyze data
  - Three icons on top right corner are available on all pages
  - The search box allows users to search publicly available datasets from the repository with extensive options to filter.
- **Search**
  - Simply click on the search button to browse the repository.
    - Will list all datasets; 20 records at time
    - The repository has 50,154 projects.
  - Provide specific search criteria to limit the results to you interest area.
    - The "field" dropdown lists the different features you can search against
    - Search all projects that has samples taken from "Nebraska"
    - Filter more to include samples collected from "nine-mile parairie"
    - The table shows the filtered results based on the supplied criteria.



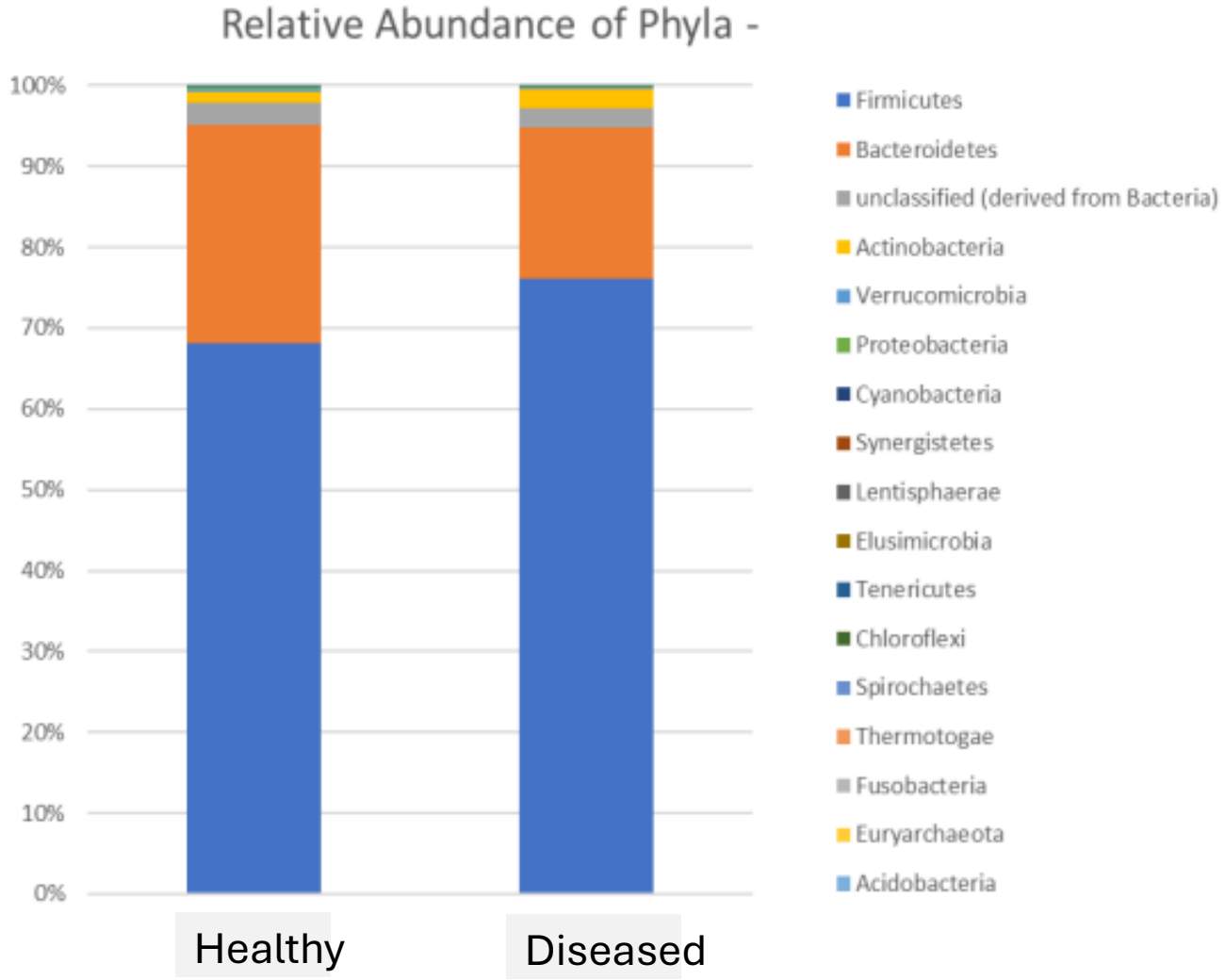
https://theseed.org/wiki/File:MG-Rast_Navigation_Overview.jpg

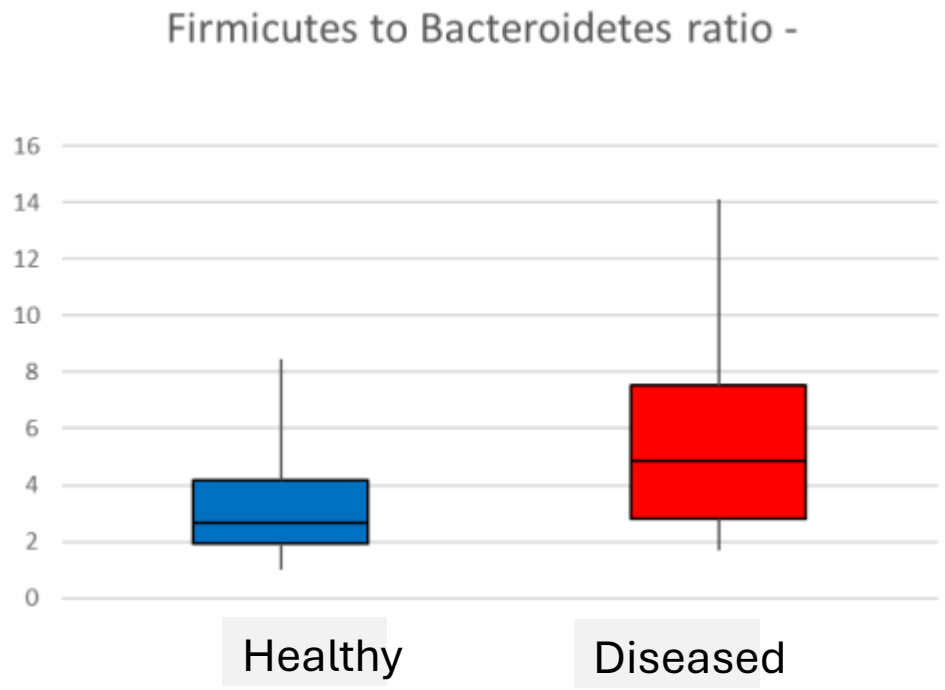
MG RAST annotation by SEED based approach



- Amino Acids and Derivatives
- Carbohydrates
- Cell Division and Cell Cycle
- Cell Wall and Capsule
- Clustering-based subsystems
- Cofactors, Vitamins, Prosthetic Groups, Pigments
- DNA Metabolism
- Dormancy and Sporulation
- Fatty Acids, Lipids and Isoprenoids
- Iron acquisition and metabolism
- Membrane Transport
- Metabolism of Aromatic Compounds
- Miscellaneous
- Motility and chemotaxis
- Nitrogen Metabolism
- Nucleosides and Nucleotides
- Phages, Prophages, Transposable elements, Plasmids
- Phosphorus Metabolism
- Photosynthesis
- Potassium Metabolism
- Protein Metabolism
- RNA Metabolism
- Regulation and Cell signaling
- Respiration
- Secondary Metabolism
- Stress Response
- Sulfur Metabolism
- Virulence, Disease and Defense

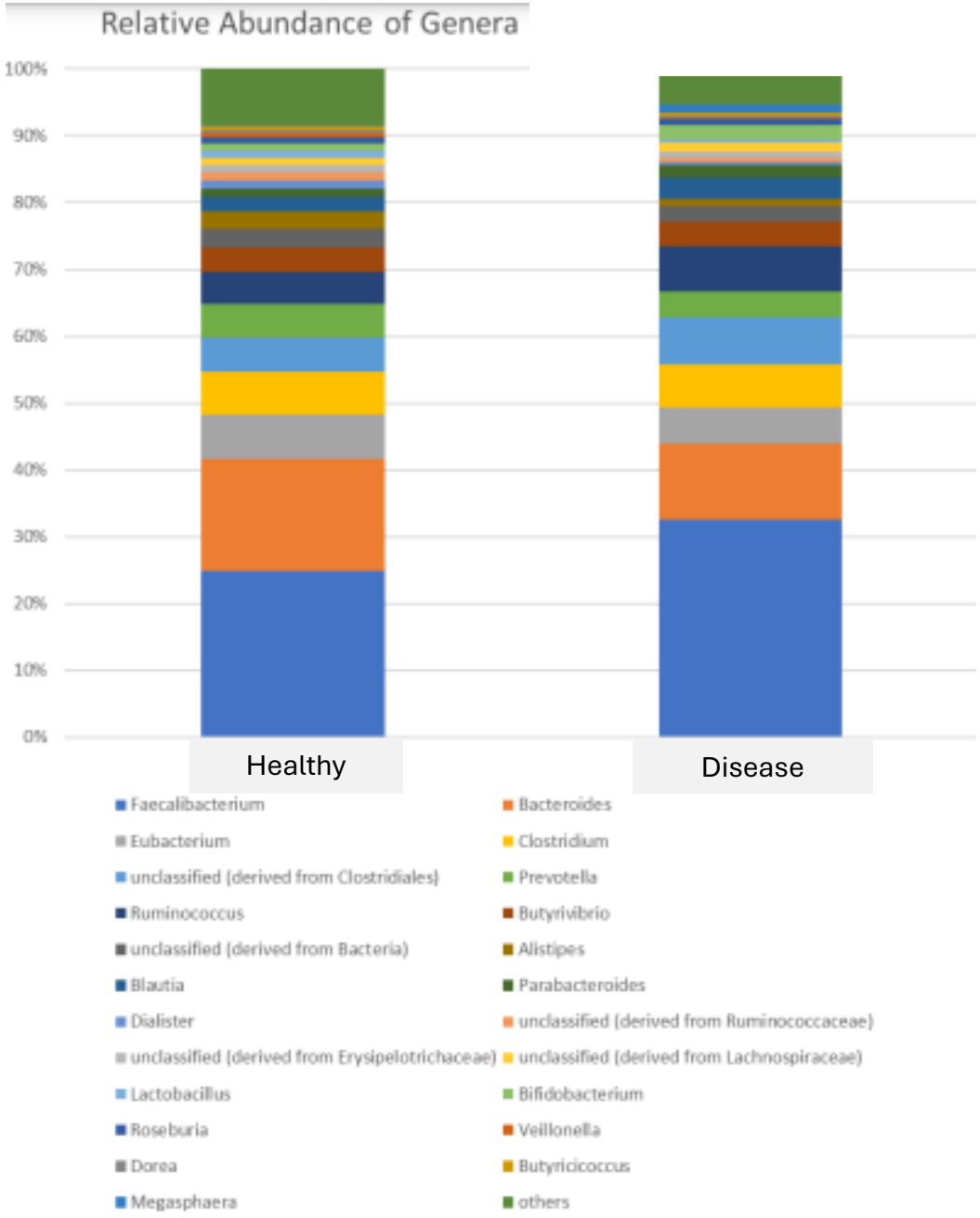
MG RAST Classification example

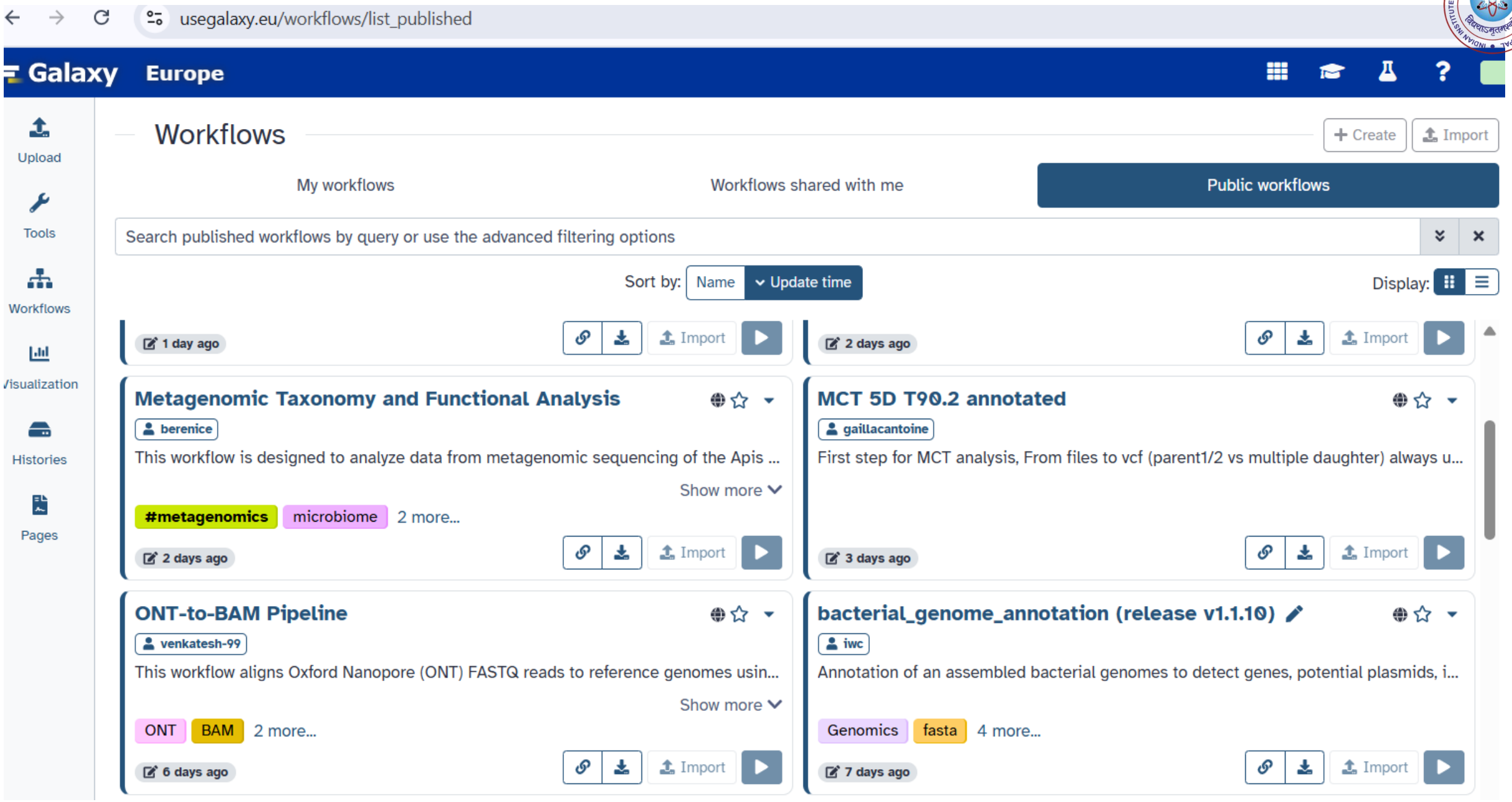




F/B ratio

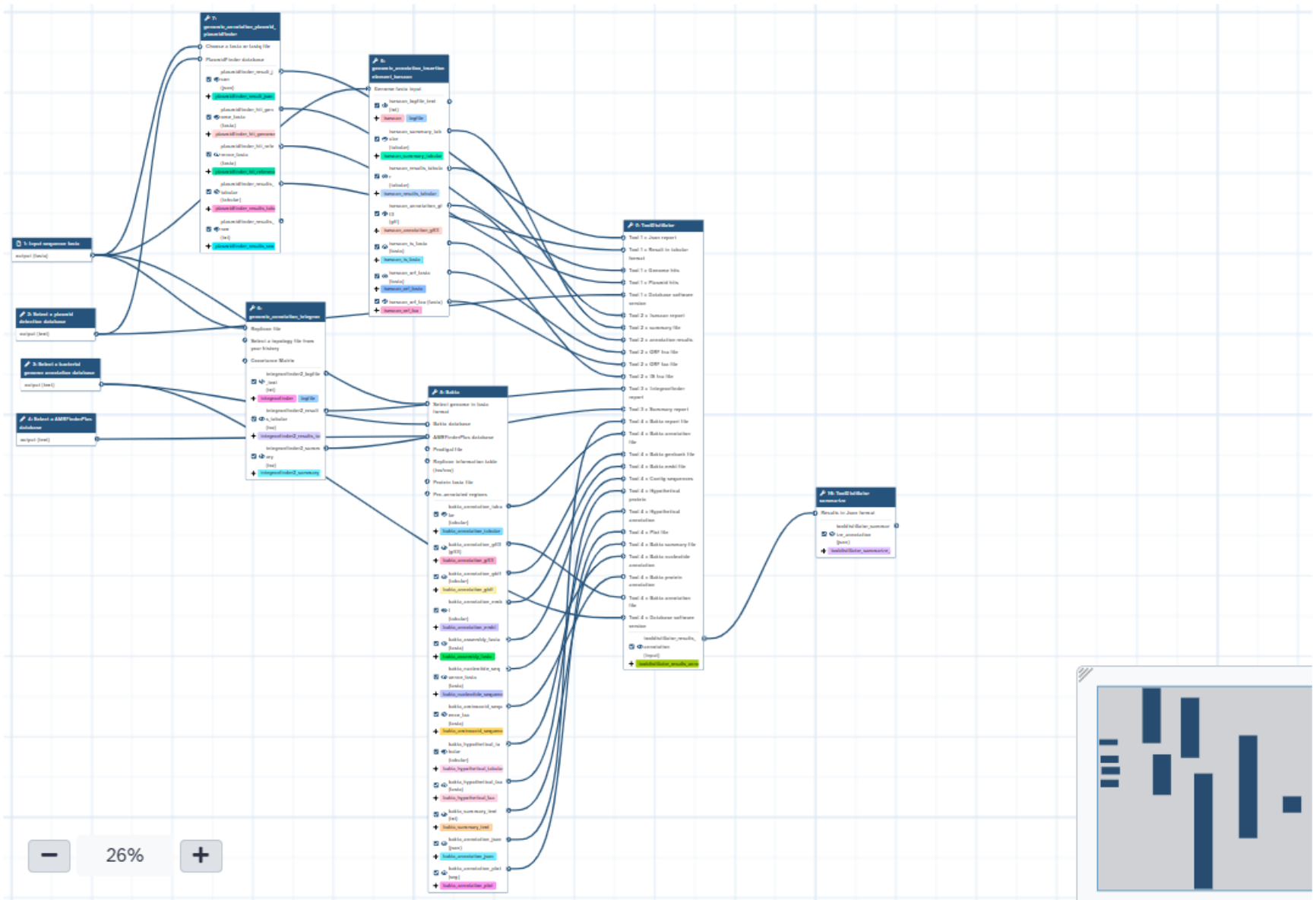
MG RAST Classification example




usegalaxy.eu/workflows/list_published
Galaxy Europe
Upload
Tools
Workflows
Visualization
Histories
Pages
Workflows
+ Create
Import
My workflows
Workflows shared with me
Public workflows
Search published workflows by query or use the advanced filtering options
Sort by: Name Update time
Display:
1 day ago
Import
2 days ago
Import
Metagenomic Taxonomy and Functional Analysis
berenice
This workflow is designed to analyze data from metagenomic sequencing of the Apis ...
Show more
#metagenomics
microbiome
2 more...
2 days ago
Import
MCT 5D T90.2 annotated
gaillacantoine
First step for MCT analysis, From files to vcf (parent1/2 vs multiple daughter) always u...
3 days ago
Import
ONT-to-BAM Pipeline
venkatesh-99
This workflow aligns Oxford Nanopore (ONT) FASTQ reads to reference genomes usin...
Show more
ONT
BAM
2 more...
6 days ago
Import
bacterial_genome_annotation (release v1.1.10)
iwc
Annotation of an assembled bacterial genomes to detect genes, potential plasmids, i...
Genomics
fasta
4 more...
7 days ago
Import

Workflow Preview

Download | Import



26%

About This Workflow

bacterial_genome_annotation (release v1.1.10) - Version 1

Author

iwc

All published Workflows by iwc

Creators

ABRomics
abromics-consortium
Pierre Marin
Clea Siguret

Description

Annotation of an assembled bacterial genomes to detect genes, potential plasmids, integrons and Insertion sequence (IS) elements.

Tags

Genomics | fasta | ABRomics | bacterial-genomics | Annotation | 1 more...

License

GNU General Public License v3.0 or later

Last Updated

Thursday Jul 17th 6:41:56 2025 GMT+5:30

We can clone the whole platform in our system and do all kind of analysis

http://wiki.galaxyproject.org/Admin/GetGalaxy

VERSION actually update version 2 mont

README.md

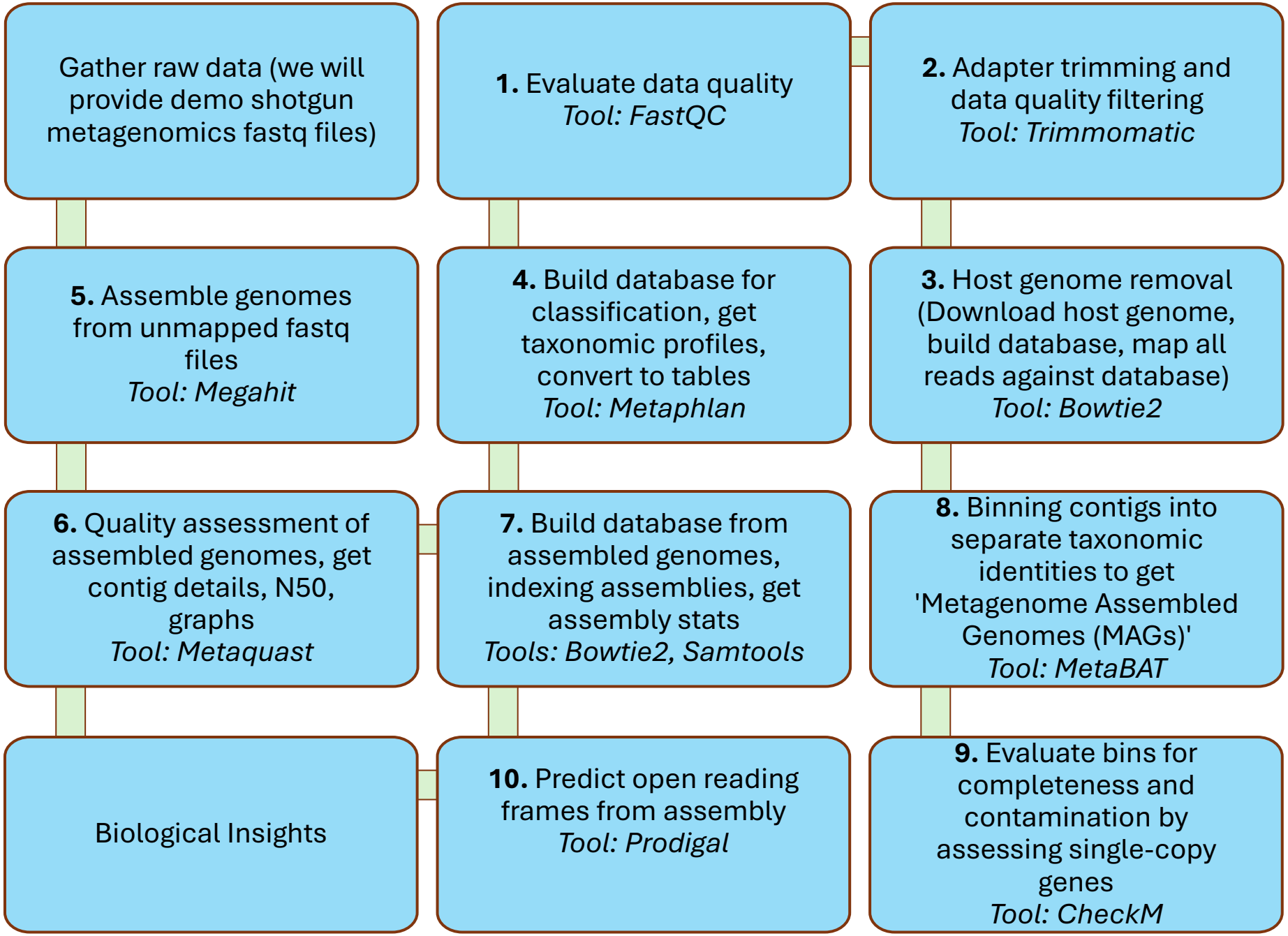
SqueezeMeta: a fully automated metagenomics pipeline, from reads to bins

- Find the SqueezeMeta paper at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.03349/full
- Find a second paper on how to analyse the output of SqueezeMeta at: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-03703-2
- Check some papers using SqueezeMeta: https://github.com/jtamames/SqueezeMeta/wiki/Some-papers-using-SqueezeMeta-(non-comprehensive-list)
- Make sure to check the wiki!

1. What is SqueezeMeta?

SqueezeMeta is a full automatic pipeline for metagenomics/metatranscriptomics, covering all steps of the analysis. SqueezeMeta includes multi-metagenome support allowing the co-assembly of related metagenomes and the retrieval of individual genomes via binning procedures. Thus, SqueezeMeta features several unique characteristics:

https://www.frontiersin.org/articles/10.3389/fmicb.2018.03349/full


Gather raw data (we will provide demo shotgun metagenomics fastq files)
1. Evaluate data quality
Tool: FastQC
2. Adapter trimming and data quality filtering
Tool: Trimmomatic
5. Assemble genomes from unmapped fastq files
Tool: Megahit
4. Build database for classification, get taxonomic profiles, convert to tables
Tool: Metaphlan
3. Host genome removal (Download host genome, build database, map all reads against database)
Tool: Bowtie2
6. Quality assessment of assembled genomes, get contig details, N50, graphs
Tool: Metaquast
7. Build database from assembled genomes, indexing assemblies, get assembly stats
Tools: Bowtie2, Samtools
8. Binning contigs into separate taxonomic identities to get 'Metagenome Assembled Genomes (MAGs)'
Tool: MetaBAT
Biological Insights
10. Predict open reading frames from assembly
Tool: Prodigal
9. Evaluate bins for completeness and contamination by assessing single-copy genes
Tool: CheckM