



GenEpi BioTrain – Block 2 Wave 3

Gene-by-gene approach Part 2: cgMLST applications using EnteroBase

Carolina Silva Nodari

French National Reference Center for *E. coli*, *Shigella* and *Salmonella*

May 2024

Intended Learning Objectives

Specific objectives of this session:

1. Comprehend the principles and applications in epidemiological surveillance and outbreak investigation of gene-by-gene approaches, including cgMLST and hierarchical clustering
2. Correlate previously acquired knowledge on genome sequencing and the biology of *E. coli* and *Shigella* with the tools available on EnteroBase and their importance for pathogen characterisation
3. Understand how the different features present on EnteroBase can be used for epidemiological surveillance and outbreak investigation

Outline

This session consists of the following elements

1. Important concepts applied to EnteroBase
2. Submitting and searching genomes on EnteroBase
3. “Experimental Data” on *E. coli*, *Shigella* and *Salmonella*
4. “Custom View”
5. Tree visualization and data export
6. Data sharing

EnteroBase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

Log In Register Help v1.2.0

EnteroBase currently contains 1,068,988 bacterial strains

Salmonella

Total strains: 472,517

Assemblies in progress: 427

Schemes: rMLST, Achtman 7 Gene MLST, cgMLST V2 + HierCC V1, wgMLST

Database Home

Escherichia/Shigella

Total strains: 304,717

Assemblies in progress: 16

Schemes: rMLST, wgMLST, Achtman 7 Gene MLST, cgMLST V1 + HierCC V1

Database Home

Mycobacterium

This database is hosted by DSMZ-German Collection of Microorganisms and Cell Cultures GmbH

DSMZ Leibniz Institute
DSMZ-German Collection of Microorganisms and Cell Cultures GmbH

Database Home

Streptococcus

Total strains: 101,836

Assemblies in progress: 11

Schemes: cgMLSTv1

Database Home

Clostridioides

Total strains: 30,698

Assemblies in progress: 1

Schemes: wgMLST, rMLST, cgMLST V1 + HierCC V1

Database Home

Vibrio

Total strains: 18,190

Assemblies in progress: 0

Schemes: rMLST, Vibrio cgMLST + HierCC

Database Home

Helicobacter

Total strains: 11,841

Assemblies in progress: 1

Schemes: rMLST

Database Home

Yersinia

Total strains: 8,766

Assemblies in progress: 1

Schemes: rMLST, wgMLST, Achtman 7 Gene, McNally 7 Gene, cgMLST V1 + HierCC V1

Database Home

Moraxella

Total strains: 2,800

Assemblies in progress: 0

Schemes: rMLST, Achtman 7 Gene

Database Home

“A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria”

Includes several databases with various tools, sometimes specific for each pathogen

Theory and important concepts on EnteroBase

Hierarchical Clustering (HierCC)

cgMLST

- Evaluation of the allelic variation of a combination of genes belonging to the core genome
- *E. coli/Shigella*: 2513 alleles
- *Salmonella* spp.: 3002 alleles
- Most genomes present a unique allelic profile (cgST)

HierCC

- Classification of the relationship of a given genome based on the allelic distance to an ancestral strain, already present in the database
- 13 classification levels
- HC0 – no difference
- HC5 – maximum 5 allele differences (genomic surveillance)
- ...
- HC2350 – different species within the *Escherichia* genus
- HC2850 – sub-species of *Salmonella*

Zhou, Charlesworth & Achtman, *Bioinformatics*. 2021;37(20):3645-3646.

Hierarchical Clustering (HierCC)

Escherichia/Shigella

- HC0 – identical strains
- HC5 – most used threshold for genomic surveillance
- HC100 – most serotypes of *E. coli*
- HC400 – some clonal groups
- HC1100 – clonal complexes - equivalent to 7-MLST CCs
- HC2000 – superlineages
- HC2350 – species

Zhou, Charlesworth & Achtman, *Bioinformatics*. 2021;37(20):3645-3646.

HierCC

- HC0 – identical strains
- HC5 – most used threshold for genomic surveillance
- HC50 – evolutive group
- HC2000 – superlineages (~ serotype)
- HC2850 – sub-species

Submitting and searching genomes on EnteroBase

EnteroBase currently contains 1,068,988 bacterial strains

Salmonella

Total strains: 472,517

Assemblies in progress: 427

Schemes: rMLST, Achtman 7 Gene MLST, cgMLST V2 + HierCC V1, wgMLST

Database Home

Escherichia/Shigella

Total strains: 304,717

Assemblies in progress: 16

Schemes: rMLST, wgMLST, Achtman 7 Gene MLST, cgMLST V1 + HierCC V1

Database Home

Mycobacterium

This database is hosted by DSMZ-German Collection of Microorganisms and Cell Cultures GmbH

DSMZ Leibniz Institute
DSMZ-German Collection of Microorganisms and Cell Cultures GmbH

Database Home

Streptococcus

Total strains: 101,836

Assemblies in progress: 11

Schemes: cgMLSTv1

Database Home

Clostridioides

Total strains: 30,698

Assemblies in progress: 1

Schemes: wgMLST, rMLST, cgMLST V1 + HierCC V1

Database Home

Vibrio

Total strains: 18,190

Assemblies in progress: 0

Schemes: rMLST, Vibrio cgMLST + HierCC

Database Home

Helicobacter

Total strains: 11,841

Assemblies in progress: 1

Schemes: rMLST

Database Home

Yersinia

Total strains: 8,766

Assemblies in progress: 1

Schemes: rMLST, wgMLST, Achtman 7 Gene, McNally 7 Gene, cgMLST V1 + HierCC V1

Database Home

Moraxella

Total strains: 2,800

Assemblies in progress: 0

Schemes: rMLST, Achtman 7 Gene

Database Home

Submitting genomes

- Start with the basics: choose the **database** that corresponds to the pathogen of interest (submitting in the wrong database will make your data fail QC)
- Important:** you must have an account and be logged-in to be able to upload and download data to EnteroBase

Species Home

Tasks

Upload Reads

Search Strains

Find ST(s)

Curate Strains

Show My Jobs

My Buddies

Locus Search

Download Schemes



Escherichia/Shigella



304,717

Total Strains

Search Strains | Find ST(s) | Upload Reads



0

Jobs Running

View Jobs



17

My Strains

Curate Strains



Mine:43
Shared:0
Public:36

79

Work Spaces

Load Workspace



Database curators

- Eva Litrup, Microbiology, SSI
- Richard Stark, Bioinformatics RTP, University of Warwick
- Scott Beatson, School of Chemistry and Molecular Biosciences, The University of Queensland
- Francois-Xavier Weill, CNR des Salmonella, Institut Pasteur
- Carolina Nodari, Recherche et d'Expertise des Bactéries Pathogènes Entériques, Institut Pasteur
- Kat Holt, Dept Infection Biology, LSHTM

Genome submission

When uploading genomes, it is possible to include a variety of metadata, including country, date of isolation and niche source (human, food, environment), and a minimal of information is required to be able to submit genomes in the database

Enterobase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

[Add Metadata](#) [Uploaded Reads](#)

Data ▾ Edit ▾ View ▾ Submit Data Contact Email Release Period (Months) 0 Rows Total:1 Errors:3

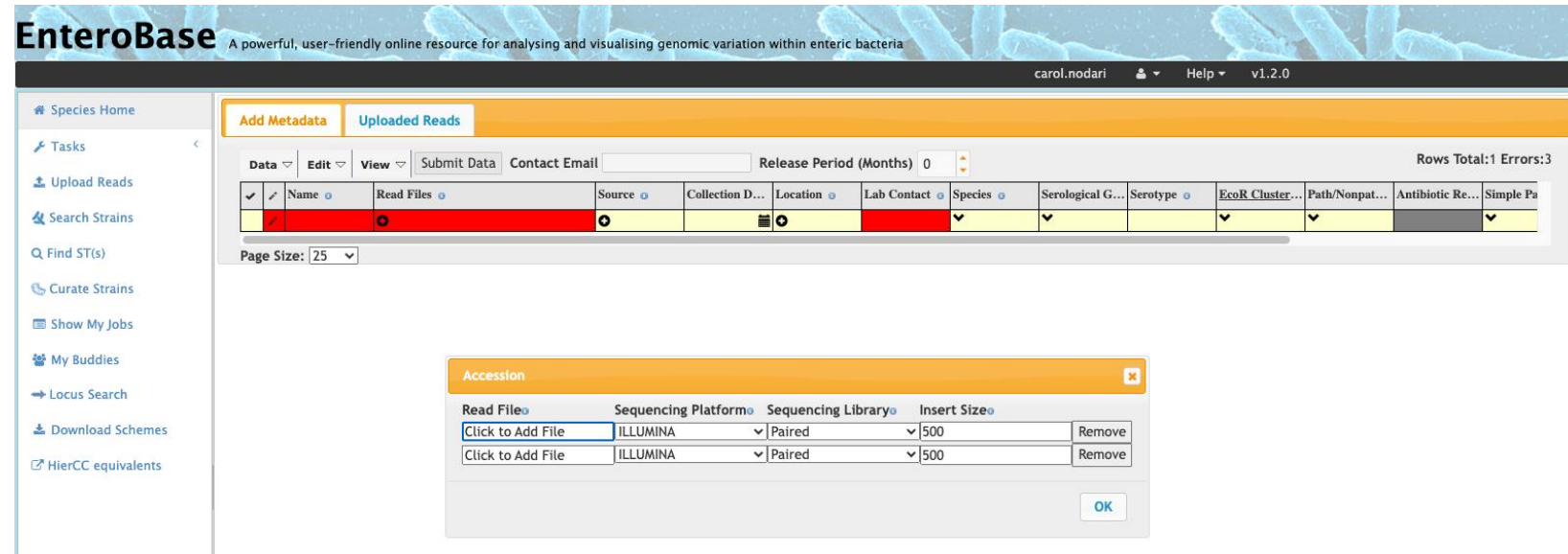
✓	✎	Name ▾	Read Files ▾	Source ▾	Collection D...	Location ▾	Lab Contact ▾	Species ▾	Serological G...	Serotype ▾	EcoR Cluster...	Path/Nonpat...	Antibiotic Re...	Simple Patho ▾	Simple Disea...	Comment ▾	Disease ▾
✓	✎							▼	▼		▼	▼		▼			▼

Page Size: 25 ▾

[Species Home](#)
[Tasks](#)
[Upload Reads](#)
[Search Strains](#)
[Find ST\(s\)](#)
[Curate Strains](#)
[Show My Jobs](#)
[My Buddies](#)
[Locus Search](#)
[Download Schemes](#)
[HierCC equivalents](#)

Genome submission

- When uploading genomes, it is possible to include a variety of metadata, including country, date of isolation and niche source (human, food, environment), and a minimal of information is required to be able to submit genomes in the database.



Enterobase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

Uploaded Reads

Data Edit View Submit Data Contact Email Release Period (Months) 0 Rows Total:1 Errors:3

✓	Name	Read Files	Source	Collection D...	Location	Lab Contact	Species	Serological G...	Serotype	EcoR Cluster...	Path/Nonpat...	Antibiotic Re...	Simple Pa

Page Size: 25

Accession

Read File	Sequencing Platform	Sequencing Library	Insert Size	
Click to Add File	ILLUMINA	Paired	500	Remove
Click to Add File	ILLUMINA	Paired	500	Remove

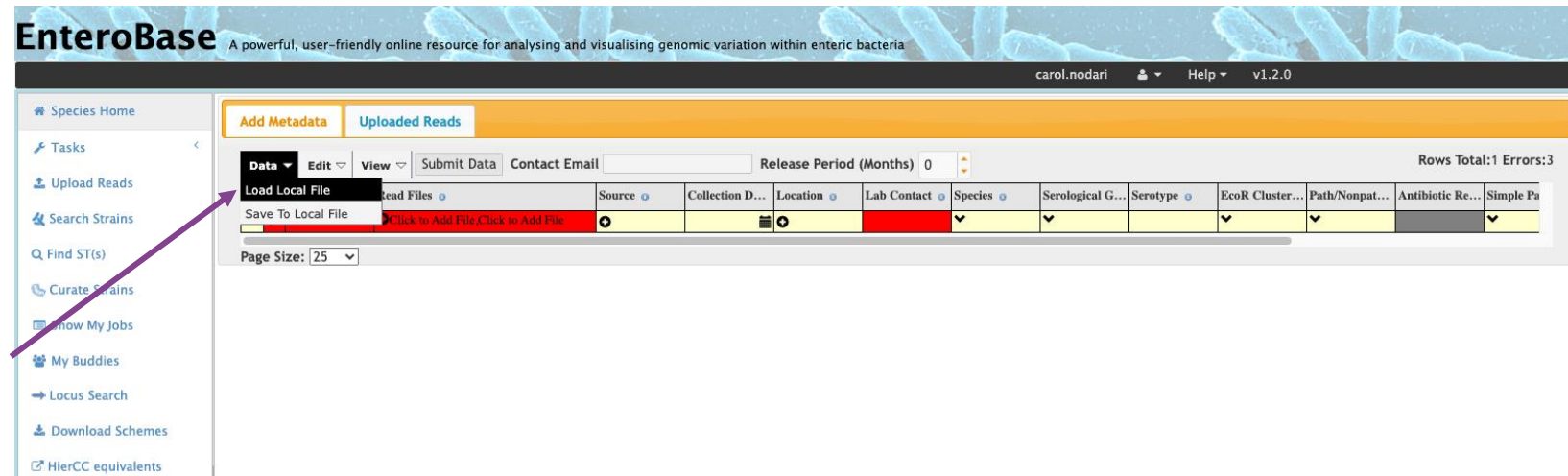
OK

- By clicking on the “read files” field, a window will pop up. One file should be chosen for each line.
- Important:** for the moment, it is only possible to submit Illumina genomic data.

Genome submission - batch upload

- Genomes and their associated metadata can be uploaded in batches.
- A tab-separated **.txt** file containing all the metadata, as well as the names of the read files can be uploaded using the "Data > Load local file" menu.
- Important:** in the "Read files" column, files names *.R1.fastq(.gz) and *.R2.fastq(.gz) must be separated by a **comma** and no white spaces.

	A	B	C	D	E	F	G	H	I
1	Name	Species	Serotype	Collection Year	Continent	Country	Source	Lab Contact	Read Files
2	202403595	Escherichia coli		2024	Europe	France	Human	F.X.Weill, Institut Pasteur	Ec202403595_S28.1.fastq.gz, Ec202403595_S28.2.fastq.gz
3	202403596	Escherichia coli		2024	Europe	France	Human	F.X.Weill, Institut Pasteur	Ec202403596_S29.1.fastq.gz, Ec202403596_S29.2.fastq.gz
4	202403597	Escherichia coli		2024	Europe	France	Human	F.X.Weill, Institut Pasteur	Ec202403597_S30.1.fastq.gz, Ec202403597_S30.2.fastq.gz
5									
6									



Enterobase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari | Help | v1.2.0

Species Home | Tasks | Upload Reads | Search Strains | Find ST(s) | Curate Strains | Show My Jobs | My Buddies | Locus Search | Download Schemes | HierCC equivalents

Add Metadata | **Uploaded Reads**

Data | Edit | View | Submit Data | Contact Email | Release Period (Months) 0 | Rows Total: 1 Errors: 3

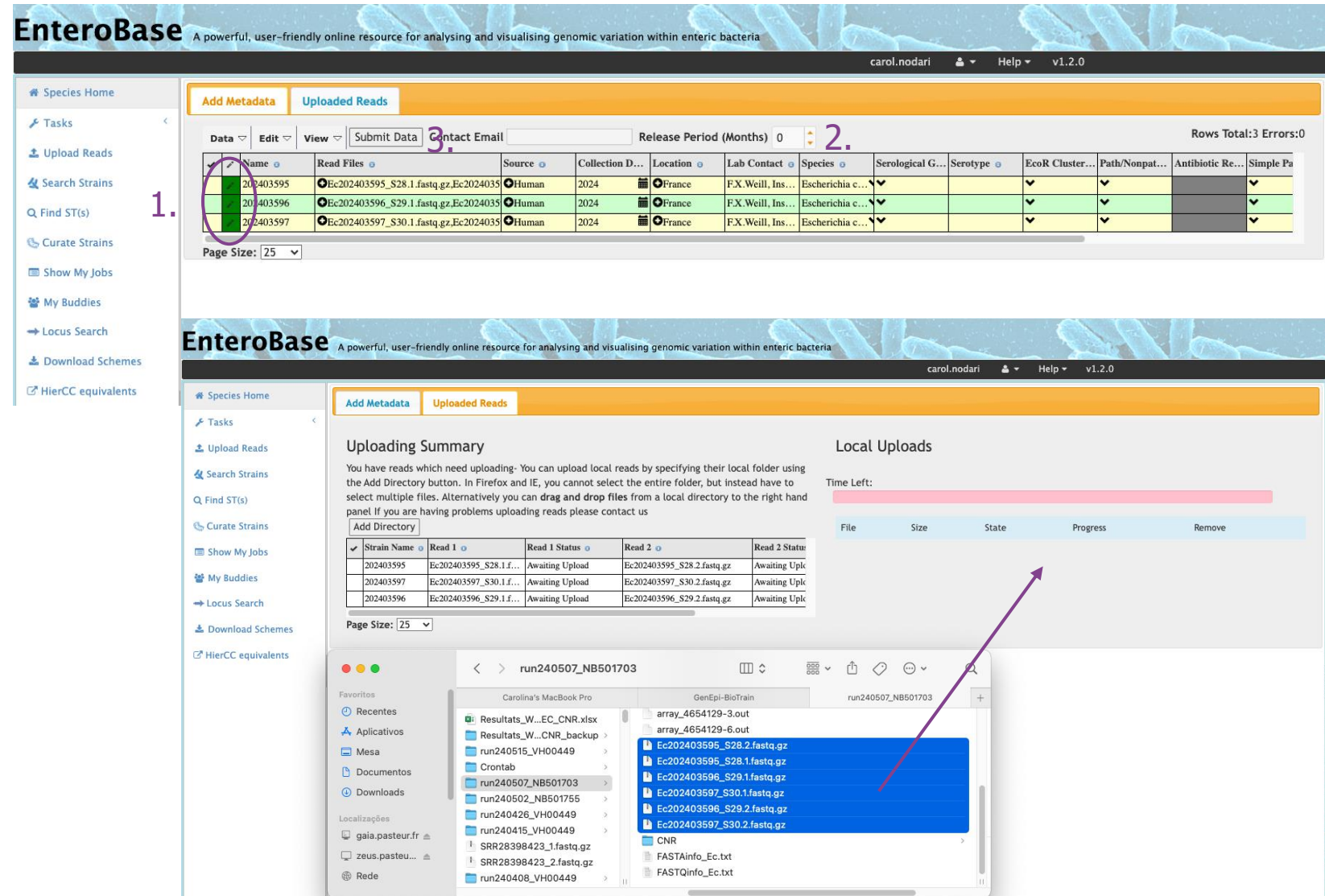
Load Local File | Read Files | Source | Collection D... | Location | Lab Contact | Species | Serological G... | Serotype | EcoR Cluster... | Path/Nonpat... | Antibiotic Re... | Simple Pa

Save To Local File | Click to Add File | Click to Add File

Page Size: 25

Genome submission - batch upload

1. If all data was correctly submitted, the second column will become green. Fields with errors will be marked in red.
2. Choose the time to make data public (up to 12 months).
3. Submit the metadata.
4. In the "Uploaded reads" tab you will be able to add the read files by choosing a folder containing them or by dragging and dropping the files directly on the right-side panel.



The screenshot displays the EnteroBase web interface. The top navigation bar includes the EnteroBase logo and the tagline "A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria". The user's name "carol.nodari" and version "v1.2.0" are shown in the top right. The left sidebar contains navigation links: Species Home, Tasks, Upload Reads, Search Strains, Find ST(s), Curate Strains, Show My Jobs, My Buddies, Locus Search, Download Schemes, and HierCC equivalents. The main content area is divided into two tabs: "Add Metadata" and "Uploaded Reads". The "Uploaded Reads" tab is active, showing a table of uploaded reads. The table has columns: Name, Read Files, Source, Collection Date, Location, Lab Contact, Species, Serological Group, Serotype, EcoR Cluster, Path/Nonpath, Antibiotic Resistance, and Simple Path. The first three rows of data are highlighted in green, indicating successful uploads. The first row is circled in red. The table also shows a "Page Size" of 25 and a "Rows Total: 3 Errors: 0". Below the table, there is an "Uploading Summary" section with instructions on how to upload reads. To the right of the summary is a "Local Uploads" section with a "Time Left" indicator and a table of local uploads. The table has columns: File, Size, State, Progress, and Remove. The table shows three files: "Ec202403595_S28.1.fastq.gz", "Ec202403596_S29.1.fastq.gz", and "Ec202403597_S30.1.fastq.gz". Below the local uploads section is a file explorer window showing the contents of a folder named "run240507_NB501703". The folder contains several files, including "array_4654129-3.out", "array_4654129-6.out", "Ec202403595_S28.1.fastq.gz", "Ec202403596_S29.1.fastq.gz", "Ec202403597_S30.1.fastq.gz", "Ec202403598_S29.2.fastq.gz", "Ec202403597_S30.2.fastq.gz", "CNR", "FASTAinfo_Ec.txt", and "FASTQinfo_Ec.txt". A red arrow points from the "Local Uploads" section to the file explorer window.

Name	Read Files	Source	Collection Date	Location	Lab Contact	Species	Serological Group	Serotype	EcoR Cluster	Path/Nonpath	Antibiotic Resistance	Simple Path
202403595	Ec202403595_S28.1.fastq.gz	Human	2024	France	F.X. Weill, Ins...	Escherichia c...						
202403596	Ec202403596_S29.1.fastq.gz	Human	2024	France	F.X. Weill, Ins...	Escherichia c...						
202403597	Ec202403597_S30.1.fastq.gz	Human	2024	France	F.X. Weill, Ins...	Escherichia c...						

File	Size	State	Progress	Remove
Ec202403595_S28.1.fastq.gz		Awaiting Upload		
Ec202403596_S29.1.fastq.gz		Awaiting Upload		
Ec202403597_S30.1.fastq.gz		Awaiting Upload		

Genome submission - batch upload

5. Start upload to begin transferring the data.
6. The submitted genomic data will be QCed, assembled, annotated and assigned a cgMLST and HierCC profile.
7. You will be able to follow the status of your submission on the "Curate Strains" and "Show My Jobs" menus.

Enterobase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

Species Home

Tasks

Upload Reads

Search Strains

Find ST(s)

Curate Strains

Show My Jobs

My Buddies

Locus Search

Download Schemes

HierCC equivalents

Add Metadata

Uploaded Reads

Uploading Summary

You have reads which need uploading- You can upload local reads by specifying their local folder using the Add Directory button. In Firefox and IE, you cannot select the entire folder, but instead have to select multiple files. Alternatively you can drag and drop files from a local directory to the right hand panel. If you are having problems uploading reads please contact us

Add Directory

Strain Name	Read 1	Read 1 Status	Read 2	Read 2 Status
202403595	Ec202403595_S28.1.f...	Local Upload	Ec202403595_S28.2.fastq.gz	Local Upload
202403597	Ec202403597_S30.1.f...	Local Upload	Ec202403597_S30.2.fastq.gz	Local Upload
202403596	Ec202403596_S29.1.f...	Local Upload	Ec202403596_S29.2.fastq.gz	Local Upload

Page Size: 25

Local Uploads

You have local reads which are ready to upload please the start button

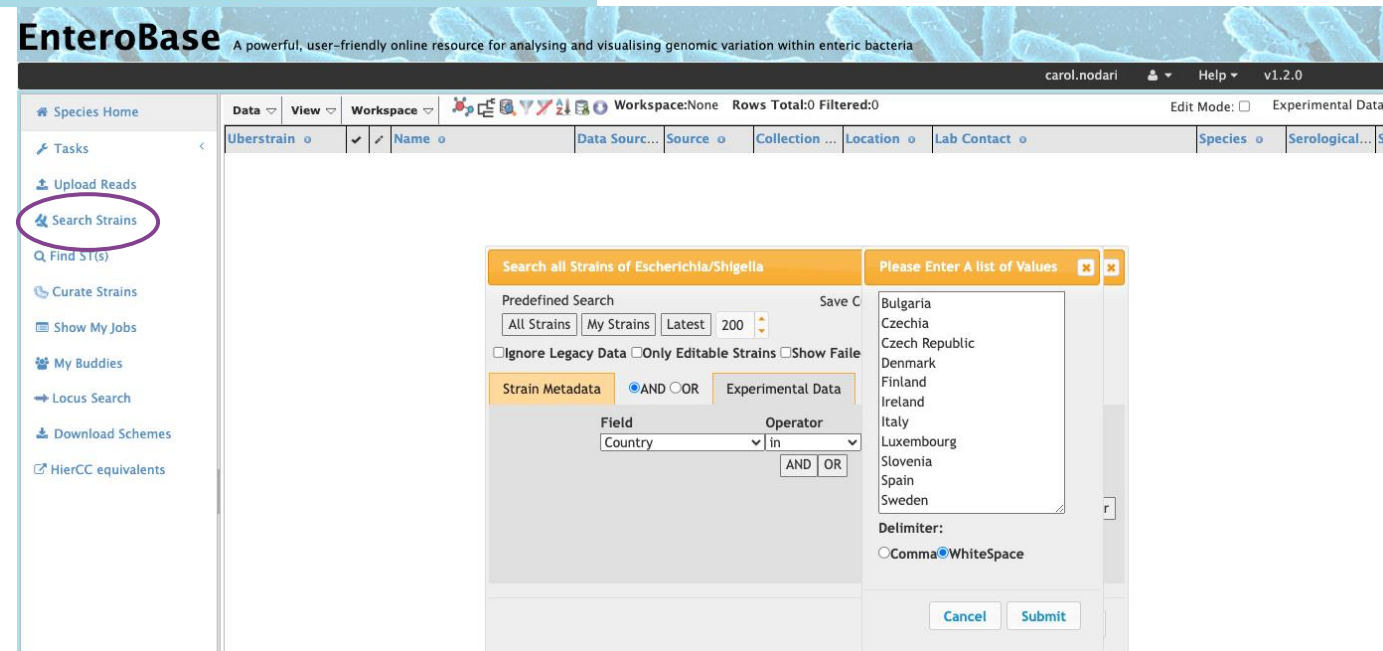
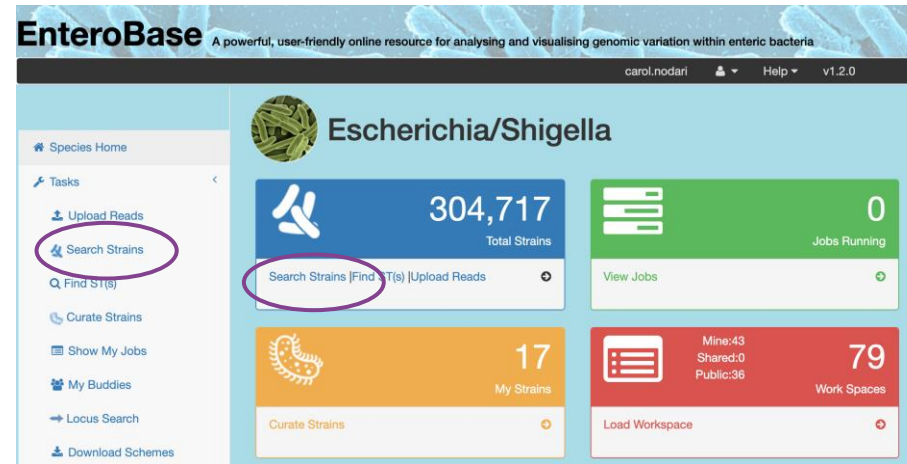
Start upload

Time Left:

File	Size	State	Progress	Remove
Ec202403595_S28.2.fastq.gz	143671228	Pending		
Ec202403595_S28.1.fastq.gz	142454052	Pending		
Ec202403596_S29.2.fastq.gz	126317985	Pending		
Ec202403597_S30.1.fastq.gz	95869683	Pending		
Ec202403596_S29.1.fastq.gz	125001805	Pending		
Ec202403597_S30.2.fastq.gz	97071692	Pending		

Searching genomes on EnteroBase

- Genomes of pathogens of interest present in **public databases** can be searched using Enterobase with the "Search Strains" tool.
- The research can be done based on strain **metadata** (isolation place or date, BioProject or SRA number on GenBank,...) or "**experimental data**", i.e., data derived from the genomes (ST, serotype, *E. coli* pathotype,...)
- Hint:** by using the operator "in", you can look into a **list** of values. Remember to write country or city names **in English**.



Searching genomes on EnteroBase

- Research results can be saved in “workspaces”, which can be shared later.
- You can also add the results of future searches in the same “workspace” using the “Add selected to workspace” option.
- The list of results of a search and/or a workspace are split in two parts: the left side contains the metadata, while the right side contains the experimental data.
- The experimental data shown by default are the “assembly status”, but several other options are available, with some of them being exclusive of a specific pathogen.

EnteroBase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari | Help | v1.2.0

Species Home | Tasks | Upload Reads | Search Strains | Find ST(s) | Curate Strains | Show My Jobs | My Buddies | Locus Search | Download Schemes | HierCC equivalents

Data View Workspace Workspace:None Rows Total:9027 Filtered:9027 Edit Mode: Experimental Data Assembly stats

Workspace: Empty

Workspaces: Mine | Shared | Public

Search:

Workspace Name: GenEpiBioTrain_Demo
Description:

Save

Strain	Metadata	Experimental Data	Status	Coverage
ESC_BA0269AA	2011-70-220-3	ERR769176	Assembled	1
ESC_BA1741AA	2011-70-220-2	ERR769175	Assembled	65.28
ESC_BA1744AA	2011-70-219-7	ERR769174	Assembled	45.67
ESC_BA1745AA	2011-70-219-4	ERR769173	Assembled	56.92
ESC_BA1748AA	2011-70-219-2	ERR769171	Assembled	48.3
ESC_BA1750AA	2011-70-217-4		Assembled	40.86
ESC_BA1751AA	2011-70-216-7		Assembled	49.25
ESC_BA1753AA	2011-70-216-4		Assembled	193
ESC_BA2107AA	2011-70-215-2		Assembled	189.86
ESC_BA2109AA	2011-70-214-10		Assembled	0
ESC_BA2111AA	2011-70-214-2		Assembled	164.81
ESC_BA2113AA	2011-70-214-1		Assembled	182.07
ESC_BA2117AA	2011-70-214-1		Assembled	221
ESC_BA2119AA	2011-70-214-1		Assembled	162
ESC_BA2125AA	2011-70-214-1		Assembled	182.07
ESC_BA2129AA	2011-70-214-1		Assembled	193.88
ESC_BA2131AA	2011-70-214-1		Assembled	316.4
ESC_BA2136AA	2011-70-214-1		Assembled	240.46
ESC_BA2138AA	2011-70-214-1		Assembled	300.99
ESC_BA2145AA	2011-70-214-1		Assembled	187.58
ESC_BA2147AA	2011-70-214-1		Assembled	121.83
ESC_BA2151AA	2011-70-214-1		Assembled	103.29
ESC_BA2153AA	2011-70-214-1		Assembled	313.66
ESC_BA2155AA	2011-70-214-1		Assembled	0
ESC_BA2157AA	2011-70-214-1		Assembled	269.56
ESC_BA2161AA	2011-70-214-1		Assembled	0

Searching genomes on EnteroBase

- Research results can be saved in “**workspaces**”, which can be shared later.
- You can also add the results of **future searches** in the same “workspace” using the “Add selected to workspace” option.
- The list of results of a search and/or a workspace are split in two parts: the **left side** contains the **metadata**, while the **right side** contains the **experimental data**.
- The experimental data shown by default are the “assembly status”, but several other options are available, with some of them being exclusive of a specific pathogen.

EnteroBase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

Species Home Tasks Upload Reads Search Strains Find ST(s) Curate Strains Show My Jobs My Buddies Locus Search Download Schemes HierCC equivalents

Workspace: GenEpioBioTrain_Demo Rows Total: 9027 Filtered: 9027 Edit Mode: Experimental

Data	View	Analysis	Workspace	Data Source	Source	Collection	Location	Lab Contact	Status	Coverage
Uberstrain										
ESC_KB0261AA				2404T28976	Human	2024	Denmark	SSI	Assembled	75.4
ESC_KB0450AA				20240074	Human	2024	Spain	University Hospital Virgen Macarena	Assembled	31.3
ESC_KB0451AA				20240180	Human	2024	Spain	University Hospital Virgen Macarena	Assembled	48.2
ESC_BA0269AA				CFSAN029787	Human	15/4/2012	Italy	FDA	Assembled	1
ESC_BA1741AA				54228	Human	1/1/1947	Sweden	Wellcome Trust Sanger Institute	Assembled	65.28
ESC_BA1744AA				54178	Human	1/1/1945	Sweden	Wellcome Trust Sanger Institute	Assembled	45.67
ESC_BA1745AA				54179	Human	1/1/1944	Sweden	Wellcome Trust Sanger Institute	Assembled	56.92
ESC_BA1748AA				54184	Human	1/1/1945	Sweden	Wellcome Trust Sanger Institute	Assembled	48.3
ESC_BA1750AA				54185	Human	1/1/1945	Denmark	Wellcome Trust Sanger Institute	Assembled	40.86
ESC_BA1751AA				54190	Human	1/1/1945	Sweden	Wellcome Trust Sanger Institute	Assembled	49.25
ESC_BA1753AA				54210	Human	1/1/1943	Sweden	Wellcome Trust Sanger Institute	Assembled	193
ESC_BA2107AA				2011-70-220-3	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	189.86
ESC_BA2109AA				2011-70-220-2	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	0
ESC_BA2111AA				2011-70-219-9	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	164.81
ESC_BA2113AA				2011-70-219-7	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	182.07
ESC_BA2117AA				2011-70-219-4	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	221
ESC_BA2119AA				2011-70-219-2	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	162
ESC_BA2125AA				2011-70-217-4	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	182.07
ESC_BA2129AA				2011-70-216-7	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	193.88
ESC_BA2131AA				2011-70-216-4	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	316.4
ESC_BA2136AA				2011-70-215-2	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	240.46
ESC_BA2138AA				2011-70-214-10	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	300.99
ESC_BA2145AA				2011-70-214-2	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	187.58
ESC_BA2147AA				2011-70-43-1	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	121.83
ESC_BA2151AA				2011-70-42-3	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	103.29
ESC_BA2153AA				2011-70-42-2	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	313.66
ESC_BA2155AA				2011-70-42-1	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	0
ESC_BA2157AA				2011-70-41-11	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	269.56
ESC_BA2161AA				2011-70-41-4	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	0
ESC_BA2163AA				2011-70-41-2	Livestock	1/1/2011	Denmark	Justus Liebig University Giessen	Assembled	119.48

Phylogenies AMR analysis Assembly stats cgMLST V1 + HierCC V1 Achtman 7 Gene MLST Annotation rMLST Serotype Prediction wgMLST Custom View Load New/Edit

Page Size: 500 1-500 501-1000 1001-1500 1501-2000 2001-2500 2501-3000 3001-3500 3501-4000 4001-4500 4501-5000 GoTo: Record Go

“Experimental Data” on *E. coli*, *Shigella* and *Salmonella*

Experimental Data

- With the assembled genomes, a lot of data can be retrieved from the genomes.
- Among the features included in EnteroBase, we can find **serotype** prediction tools, which are specific for each pathogen.
- Recently, an **AMR analysis** using AMRFinderPlus has been implemented is currently been tested in small sets of genomes.
- Both 7 genes **MLST** and **cgMLST** can also be retrieved, as well as the **HierCC** information.
- Pathotypes** of *E. coli*, except for EAEC, and *Shigella* serogroups can also be identified in the menu "Phylotypes".

EnteroBase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

Workspace:None Rows Total:10609 Filtered:5230 Edit Mode: Experimental Data Serotype(SISTR1+SeqSero2)

Uberstrain	Name	Data Source	Source	Collection	Location	SISTR1 Ser...	SISTR1 Ser...	SISTR1 H1	SISTR1 H2	SeqSero2 S...	SeqSero2 S...	SeqSero2 ...	SeqSero2 ...
■SAL_KB9...	282	Uploaded Read	Human	2022	Italy	Typhimurium	B	i	1,2	Typhimurium	4	i	1,2
■SAL_LB2...	619	Uploaded Read	Human	2017	Slovenia	I 1,4,[5],12:b-	B	b	-	4:b-	4	b	-
■SAL_OB3...	1563	Uploaded Read	Human	2022	Spain	Enteritidis	D1	g,m	-	Enteritidis	9	g,m	-
■SAL_KB9...	13586	Uploaded Read	Human	2017	Italy	Infantis	C1	r	1,5	Infantis	7	r	1,5
■SAL_KB9...	18812	Uploaded Read	Human	2016	Italy	Choleraesuis	C1	c	1,5	Paratyphi C o...	7	c	1,5
■SAL_EA9...	24249	GCF_0005038	Human	2008	Ireland	Agona	B	f,g,s	-	Agona	4	f,g,s	-
■SAL_LA1...	00028816	ERR2719642	Human	2000	Denmark	Typhimurium	B	i	1,2	Typhimurium	4	i	1,2
■SAL_HA8...	42069	Uploaded Read	Human	2016	Denmark	I C1:z4,z23:-	C1	z4,z23	-	-	z4,z23	-	-
■SAL_KB9...	43587	Uploaded Read	Human	2017	Italy	Infantis	C1	r	1,5	Infantis	7	r	1,5
■SAL_FA56...	45605	SRR3417496	Human	2014	Luxembourg	Enteritidis	D1	g,m	-	Enteritidis	9	g,m	-
■SAL_KB9...	60228	Uploaded Read	Human	2016	Italy	Infantis	C1	r	1,5	Infantis	7	r	1,5
■SAL_KB9...	62768	Uploaded Read	Human	2017	Italy	Veneziana	F	i	e,n,x	Veneziana	11	i	e,n,x
■SAL_KB9...	68178	Uploaded Read	Human	2016	Italy	Infantis	C1	r	1,5	Infantis	7	r	1,5
■SAL_KB9...	72061	Uploaded Read	Human	2018	Italy	Infantis	C1	r	1,5	Infantis	7	r	1,5
■SAL_KB9...	72065	Uploaded Read	Human	2018	Italy	Infantis	C1	r	1,5	Infantis	7	r	1,5

EnteroBase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

Workspace:GenEpiBioTrain_Demo Rows Total:9027 Filtered:9027 Edit Mode: Experimental Data Serotype Prediction

Uberstrain	Name	Data Source	Source	Collection	Location	Lab Contact	O Antigen	H Antigen
■ESC_BA0269AA	CFSAN029787	SRR2019242	Human	15/4/2012	Italy	FDA	O96	H19
■ESC_BA1741AA	54228	ERR025724	Human	1/1/1947	Sweden	Wellcome Trust Sange	-	H16
■ESC_BA1744AA	54178	ERR025726	Human	1/1/1945	Sweden	Wellcome Trust Sange	-	H16
■ESC_BA1745AA	54179	ERR025727	Human	1/1/1944	Sweden	Wellcome Trust Sange	-	H16
■ESC_BA1748AA	54184	ERR025729	Human	1/1/1945	Sweden	Wellcome Trust Sange	-	H16
■ESC_BA1750AA	54185	ERR025730	Human	1/1/1945	Denmark	Wellcome Trust Sange	-	H16
■ESC_BA1751AA	54190	ERR025731	Human	1/1/1945	Sweden	Wellcome Trust Sange	-	H16
■ESC_BA1753AA	54210	ERR025732	Human	1/1/1943	Sweden	Wellcome Trust Sange	-	H16
■ESC_BA2107AA	2011-70-220-3	ERR769198	Livestock	1/1/2011	Denmark	Justus Liebig Universi	O23	H16
■ESC_BA2109AA	2011-70-220-2	ERR769197	Livestock	1/1/2011	Denmark	Justus Liebig Universi	-	H28
■ESC_BA2111AA	2011-70-219-9	ERR769196	Livestock	1/1/2011	Denmark	Justus Liebig Universi	O24	H4
■ESC_BA2113AA	2011-70-219-7	ERR769195	Livestock	1/1/2011	Denmark	Justus Liebig Universi	O33	H18
■ESC_BA2117AA	2011-70-219-4	ERR769193	Livestock	1/1/2011	Denmark	Justus Liebig Universi	-	H26
■ESC_BA2119AA	2011-70-219-2	ERR769192	Livestock	1/1/2011	Denmark	Justus Liebig Universi	O76	H32

Experimental Data

- With the assembled genomes, a lot of data can be retrieved from the genomes.
- Among the features included in EnteroBase, we can find **serotype** prediction tools, which are specific for each pathogen.
- Recently, an **AMR analysis** using AMRFinderPlus has been implemented is currently been tested in small sets of genomes.
- Both 7 genes **MLST** and **cgMLST** can also be retrieved, as well as the **HierCC** information.
- Pathotypes** of *E. coli*, except for EAEC, and *Shigella* serogroups can also be identified in the menu "Phylotypes".

EnteroBase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

Species Home Tasks Upload Reads Search Strains Find ST(s) Curate Strains Show My Jobs My Buddies Locus Search Download Schemes HierCC equivalents

Workspace:None Rows Total:10609 Filtered:5230 Edit Mode: Experimental Data CgMLST V2 + HierCC V1

Uberstrain	Name	Data Source	Source	Collection	Location	ST	HC0 (indis...)	HC2	HCS	HC10	HC20	HCS0	HC100
■SAL_KB9...	282	○Uploaded Read	○Human	2022	○Italy	296604	296604	296604	296604	296604	2	2	
■SAL_LB2...	619	○Uploaded Read	○Human	2017	○Slovenia	299085	299085	299085	299085	299085	165816	165816	
■SAL_OB3...	1563	○Uploaded Read	○Human	2022	○Spain	315120	315120	315120	1962	1962	87	87	87
■SAL_KB9...	13586	○Uploaded Read	○Human	2017	○Italy	296593	296593	296593	296593	296593	2398	2398	36
■SAL_KB9...	18812	○Uploaded Read	○Human	2016	○Italy	296588	296588	168357	168350	168350	79087	79087	79087
■SAL_EA9...	24249	○GCF_0005038	○Human	2008	○Ireland	5476	5476	5476	5476	5476	419	29	29
■SAL_LA1...	00028816	○ERR2719642	○Human	2000	○Denmark	85762	85762	85762	85762	85762	9921	9921	48
■SAL_HA8...	42069	○Uploaded Read	○Human	2016	○Denmark	57128	57128	57128	57128	57128	57128	57128	57128
■SAL_KB9...	43587	○Uploaded Read	○Human	2017	○Italy	296590	296590	296590	296590	296590	36	36	
■SAL_FA56...	45605	○SRR3417496	○Human	2014	○Luxembourg	514	514	514	511	511	397	397	12
■SAL_KB9...	60228	○Uploaded Read	○Human	2016	○Italy	296592	296592	296592	140078	21046	7828	36	36
■SAL_KB9...	62768	○Uploaded Read	○Human	2017	○Italy	296589	296589	296589	296589	296589	296589	296589	296589
■SAL_KB9...	68178	○Uploaded Read	○Human	2016	○Italy	296591	296591	296591	296591	41796	7828	36	36
■SAL_KB9...	72061	○Uploaded Read	○Human	2018	○Italy	296603	296603	296603	41796	7828	36	36	36
■SAL_KB9...	72065	○Uploaded Read	○Human	2018	○Italy	296595	296595	296595	41796	7828	36	36	36

EnteroBase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

Species Home Tasks Upload Reads Search Strains Find ST(s) Curate Strains Show My Jobs My Buddies Locus Search Download Schemes HierCC equivalents

Workspace:GenEpiBioTrain_Demo Rows Total:9027 Filtered:9027 Experimental Data CgMLST V1 + HierCC V1

Uberstrain	Name	Data Source	Source	Collection	Location	ST	HC0 (indis...)	HC2	HCS	HC10	HC20	HCS0
■ESC_KB0261AA	2404T28976	○Uploaded Read	○Human	2024	○Denmark	281348	281348	281348	281348	281348	281348	118194
■ESC_KB0450AA	20240074	○Uploaded Read	○Human	2024	○Spain	281647	281647	281647	281647	281647	3330	3288
■ESC_KB0451AA	20240180	○Uploaded Read	○Human	2024	○Spain	281646	281646	281646	281646	281646	281646	281646
■ESC_BA0269AA	CFSAN029787	○SRR2019242	○Human	15/4/2012	○Italy	15397	15397	15397	15397	15397	15397	2282
■ESC_BA1741AA	54228	○ERR025724	○Human	1/1/1947	○Sweden	14276	14276	14276	14276	14276	14276	14276
■ESC_BA1744AA	54178	○ERR025726	○Human	1/1/1945	○Sweden	14271	14271	14271	14271	14271	14271	7177
■ESC_BA1745AA	54179	○ERR025727	○Human	1/1/1944	○Sweden	14275	14275	14275	14275	14275	14275	14275
■ESC_BA1748AA	54184	○ERR025729	○Human	1/1/1945	○Sweden	20482	14272	14272	14272	14272	14272	14266
■ESC_BA1750AA	54185	○ERR025730	○Human	1/1/1945	○Denmark	14270	14270	14270	14270	14270	14270	7177
■ESC_BA1751AA	54190	○ERR025731	○Human	1/1/1945	○Sweden	14268	14268	14268	14268	14268	14268	7177
■ESC_BA1753AA	54210	○ERR025732	○Human	1/1/1943	○Sweden	105983	14266	14266	14266	14266	14266	14266
■ESC_BA2107AA	2011-70-220-3	○ERR769198	○Livestock	1/1/2011	○Denmark	14018	14018	14018	14018	14018	14018	14018
■ESC_BA2109AA	2011-70-220-2	○ERR769197	○Livestock	1/1/2011	○Denmark	14017	14017	14017	14017	14017	14017	14017
■ESC_BA2111AA	2011-70-219-9	○ERR769196	○Livestock	1/1/2011	○Denmark	14015	14015	14015	14015	14015	14015	14015

Experimental Data

- With the assembled genomes, a lot of data can be retrieved from the genomes.
- Among the features included in EnteroBase, we can find **serotype** prediction tools, which are specific for each pathogen.
- Recently, an **AMR analysis** using AMRFinderPlus has been implemented is currently been tested in small sets of genomes.
- Both 7 genes **MLST** and **cgMLST** can also be retrieved, as well as the **HierCC** information.
- Pathotypes** of *E. coli*, except for EAEC, and *Shigella* serogroups can also be identified in the menu "Phylotypes".

EnteroBase

A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari

Help

v1.2.0

Species Home

Data

View

Analysis

Workspace

Workspace:GenEpiBioTrain_Demo

Rows Total:9027 Filtered:423

Edit Mode:

Experimental Data

Phylotypes

Tasks

Upload Reads

Search Strains

Find ST(s)

Curate Strains

Show My Jobs

My Buddies

Locus Search

Download Schemes

HierCC equivalents

Uberstrain

Name

Data Source

Source

Collection

Location

Cler...

Cler...

fimH (fi...

Pathovar

Stx1

Stx2

ipaH

plnv

ST

LT

ee

ESC_CB4516AA

12597

Uploaded Read

Livestock

2013

Czechia

B1

B1

fimH32

E. coli - STEC

+

+

-

-

-

-

-

ESC_YA7447AA

100306

ERR8980237

Livestock

2016

Italy

D

D

fimH54

E. coli - STEC

-

+

-

-

-

-

-

ESC_YA7454AA

131435

ERR8980230

Livestock

2017

Italy

D

D

fimH54

E. coli - STEC

-

+

-

-

-

-

-

ESC_YA7450AA

144847

ERR8980234

Livestock

2017

Italy

D

D

fimH54

E. coli - STEC

-

+

-

-

-

-

-

ESC_ZA5996AA

144963

ERR9829669

Livestock

2016

Italy

D

D

fimH54

E. coli - STEC

-

+

-

-

-

-

-

ESC_YA7448AA

165504

ERR8980236

Livestock

2016

Italy

D

D

fimH54

E. coli - STEC

-

+

-

-

-

-

-

ESC_YA7449AA

230744

ERR8980235

Livestock

2016

Italy

D

D

fimH54

E. coli - STEC

-

+

-

-

-

-

-

ESC_YA7453AA

250923

ERR8980231

Livestock

2015

Italy

A

A

fimH24

E. coli - ETEC/STEC

-

+

-

-

+

-

-

ESC_YA7452AA

262715

ERR8980232

Livestock

2017

Italy

D

D

fimH54

E. coli - STEC

-

+

-

-

-

-

-

ESC_YA7446AA

307636

ERR8980238

Livestock

2015

Italy

D

D

fimH54

E. coli - STEC

-

+

-

-

-

-

-

ESC_ZA3983AA

20220536

Uploaded Read

2022

Spain

B1

B1

fimH32

E. coli - STEC

+

+

-

-

-

-

-

ESC_ZA3982AA

20220537

Uploaded Read

2022

Spain

C

C

fimH39

E. coli - STEC

+

-

-

-

-

-

-

ESC_VA3708AA

400702008

ERR4843472

Livestock

2015

Spain

C

C

fimH39

E. coli - ETEC/STEC

-

+

-

-

-

+

-

ESC_CB0800AA

400702008

SRR21990448

Human

2021

Spain

C

C

fimH39

E. coli - ETEC/STEC

-

+

-

-

-

+

-

ESC_VA3699AA

401402008

ERR4843481

Livestock

2015

Spain

B1

B1

fimH32

E. coli - STEC

-

+

-

-

-

-

-

ESC_CB0795AA

401402008

SRR21990443

Human

2021

Spain

B1

B1

fimH32

E. coli - STEC

-

+

-

-

-

-

-

ESC_VA3694AA

401402020

ERR4843486

Livestock

2015

Spain

C

C

fimH39

E. coli - STEC

-

+

-

-

-

-

-

ESC_MA6981AA

802802004

ERR3336384

Livestock

2015

Denmark

D

D

*

E. coli - STEC

+

-

-

-

-

-

-

ESC_CA4180AA

02W30354

Uploaded Read

Human

2015

Denmark

B1

B1

fimH32

E. coli - STEC

+

+

-

-

-

-

-

ESC_XA6261AA

03_A_12_5_5

SRR15115907

Livestock

7/3/2018

Ireland

B1

B1

fimH446

E. coli - STEC

+

+

-

-

-

-

-

ESC_XA6266AA

03_A_17_5_3

SRR15115912

Livestock

7/3/2018

Ireland

C

C

fimH39

E. coli - STEC

+

-

-

-

-

-

-

ESC_XA6265AA

03_A_20_4_5

SRR15115911

Livestock

7/3/2018

Ireland

B1

B1

fimH32

E. coli - STEC

-

+

-

-

-

-

-

ESC_XA6270AA

03_A_4_1_4

SRR15115916

Livestock

7/3/2018

Ireland

B1

B1

fimH446

E. coli - STEC

+

+

-

-

-

-

-

ESC_XA6269AA

03_A_4_5_5

SRR15115915

Livestock

7/3/2018

Ireland

B1

B1

fimH446

E. coli - STEC

+

+

-

-

-

-

-

ESC_XA6268AA

03_A_6_2_4

SRR15115914

Livestock

7/3/2018

Ireland

A

A

fimH1322

E. coli - STEC

-

+

-

-

-

-

-

ESC_XA6267AA

03_A_6_4_4

SRR15115913

Livestock

7/3/2018

Ireland

E

E

fimH158

E. coli - STEC

+

-

-

-

-

-

-

ESC_XA6243AA

04_A_11_5_1

SRR15115832

Livestock

26/3/2018

Ireland

B1

B1

fimH32

E. coli - STEC

-

+

-

-

-

-

-

ESC_XA6253AA

04_A_15_5_1

SRR15115899

Livestock

26/3/2018

Ireland

B1

B1

fimH1115

E. coli - STEC

+

-

-

-

-

-

-

ESC_XA6246AA

04_A_16_1_5

SRR15115892

Livestock

26/3/2018

Ireland

B1

B1

fimH38

E. coli - STEC

+

-

-

-

-

-

-

ESC_XA6252AA

04_A_24_1_1

SRR15115898

Livestock

26/3/2018

Ireland

R1

R1

fimH1115

E. coli - STEC

+

+

-

-

-

-

-

Page Size: 500

GoTo: Record

Go

“Custom View”

Custom View

- To facilitate navigating between the data provided in the different menus, it is possible to create **"Custom views"** with chosen information.
- It is also possible to create and add data to **"Custom columns"** that are not public and/or part of the analysis pipeline from Enterobase.
- Unless specified by the user, the custom views and information in the custom columns remain **private**.
- Custom views can be **saved** and reused in future workspaces and **shared** with other (trusted) users.

Enterobase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

Workspace: GenEpiBioTrain_Demo Rows Total: 9027 Filtered: 423 Edit Mode: Experimental Data

Species Home Tasks Upload Reads Search Strains Find ST(s) Curate Strains Show My Jobs My Buddies Locus Search Download Schemes HierCC equivalents

Uberstrain	Name	Data Source	Source	Collection	Location	Cler...	Cler...	fimH (fi...	Pathovar	Stx1	Stx2	ipaH
ESC_CB4516AA	12597	Uploaded Read	Livestock	2013	Czechia	B1	B1	fimH32	E. coli - STEC	+	+	-
ESC_YA7447AA	100306	ERR8980237	Livestock	2016	Italy	D	D	fimH54	E. coli - STEC	-	+	-
ESC_YA7454AA	131435	ERR8980230	Livestock	2017	Italy	D	D	fimH54	E. coli - STEC	-	+	-
ESC_YA7450AA	144847	ERR8980234	Livestock	2017	Italy	D	D	fimH54	E. coli - STEC	-	+	-
ESC_ZA5996AA	144963	ERR9829669	Livestock	2016	Italy	D	D	fimH54	E. coli - STEC	-	+	-
ESC_YA7448AA	165504	ERR8980236	Livestock	2016	Italy	D	D	fimH54	E. coli - STEC	-	+	-
ESC_YA7449AA	230744	ERR8980235	Livestock	2016	Italy	D	D	fimH54	E. coli - STEC	-	+	-
ESC_YA7453AA	250923	ERR8980231	Livestock	2015	Italy	A	A	fimH24	E. coli - ETEC/STEC	-	+	-
ESC_YA7452AA	262715	ERR8980232	Livestock	2017	Italy	D	D	fimH54	E. coli - STEC	-	+	-
ESC_YA7446AA	307636	ERR8980238	Livestock	2015	Italy	D	D	fimH54	E. coli - STEC	-	+	-
ESC_ZA3983AA	20220536	Uploaded Read		2022	Spain	B1	B1	fimH32	E. coli - STEC	+	+	-
ESC_ZA3982AA	20220537	Uploaded Read		2022	Spain	C	C	fimH39	E. coli - STEC	+	-	-
ESC_YA3708AA	400702008	ERR4843472	Livestock	2015	Spain	C	C	fimH39	E. coli - ETEC/STEC	-	+	-
ESC_CB0800AA	400702008	SRR21990448	Human	2021	Spain	C	C	fimH39	E. coli - ETEC/STEC	-	+	-
ESC_YA3699AA	401402008	ERR4843481	Livestock	2015	Spain	B1	B1	fimH32	E. coli - STEC	-	+	-

Phylogenetic tree: AMR analysis, Assembly stats, cgMLST V1 + HierCC V1, Achtman 7 Gene MLST, Annotation, rMLST, Serotype Prediction, wgMLST, Custom View, Load, New/Edit

Custom View Editor

Add Experiment Column

Experiment: cgMLST V1 + HierCC V1

Column: HC5

Locus:

Add Custom Column

Column Name: food

Create Custom Column

Name: food

Datatype: Text

Custom View: New

Load Clear

Columns

Pathovar(Phylogenetic) ✕

ST(Achtman 7 Gene MLST) ✕

O Antigen(Serotype Prediction) ✕

H Antigen(Serotype Prediction) ✕

HC5(cgMLST V1 + HierCC V1) ✕

age ✕

sexe ✕

travel ✕

food ✕

Custom columns

Save Save As

Custom View

- **Information** (metadata and/or data for custom columns) can be added in **batches** into EnteroBase.
- To allow correct correlation between the data and the strain, each strain on EnteroBase contains a unique identifier – a “**Barcode**”.
- **Strain barcodes** (which are different from assembly barcodes) can be found in the **left side panel** and **downloaded** to a local **.txt** file.
- After **updating** the local file with the information to be added, it can be **uploaded** again.

EnteroBase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari | Help | v1.2.0

Workspace: GenEpiBioTrain_Demo Rows Total: 9027 Filtered: 423

Experimental Data: STEC_GenEpiBioTrain_Demo

Species Home | Tasks | Upload Reads | Search Strains | Find ST(s) | Curate Strains | Show My Jobs | My Buddies | Locus Search | Download Schemes | HierCC equivalents

Data | Edit | View | Analysis | Workspace | Experiment | Tools

Search Strains: Save to Local File, Save Selected to File, My Strains, Download Selected Assemblies

Name	Barcode	Data Source	Source	Collection	Location	Pathovar(P...	ST(Achtm...	O Antigen(...	H Antigen(...	HCS(cgMLS...	age	sexe	travel	food
12597	ESC_CB4516AA	Uploaded Read	Livestock	2013	Czechia	E. coli - STEC	13869	O174	H8	219993	6	F	no	minced meat
100306	ESC_YA7447AA	ER8980237	Livestock	2016	Italy	E. coli - STEC	1	O139	H1	191878	55	M	Indonesia	ND
131435	ESC_YA7454AA	ER8980230	Livestock	2017	Italy	E. coli - STEC	1	O139	H1	191873	ND	ND	ND	ND
144847	ESC_YA7450AA	ER8980234	Livestock	2017	Italy	E. coli - STEC	1	O139	H1	191875	ND	ND	ND	ND
144963	ESC_ZA5996AA	ER89829669	Livestock	2016	Italy	E. coli - STEC	1	O139	H1	199736	ND	ND	ND	ND
165504	ESC_YA7448AA	ER8980236	Livestock	2016	Italy	E. coli - STEC	1	O139	H1	191876	ND	ND	ND	ND
230744	ESC_YA7449AA	ER8980235	Livestock	2016	Italy	E. coli - STEC	1	O139	H1	191877	ND	ND	ND	ND
250923	ESC_YA7453AA	ER8980231	Livestock	2015	Italy	E. coli - ETEC	10	O141	H4	191874	ND	ND	ND	ND
262715	ESC_YA7452AA	ER8980232	Livestock	2017	Italy	E. coli - STEC	1	O139	H1	191875	ND	ND	ND	ND
307636	ESC_YA7446AA	ER8980238	Livestock	2015	Italy	E. coli - STEC	1	O139	H1	191875	ND	ND	ND	ND
20220536	ESC_ZA3983AA	Uploaded Read		2022	Spain	E. coli - STEC	8381	O128	H2	196838	ND	ND	ND	ND
20220537	ESC_ZA3982AA	Uploaded Read		2022	Spain	E. coli - STEC	13245	O78	H4	196833	ND	ND	ND	ND
400702008	ESC_VA3708AA	Livestock		2015	Spain	E. coli - ETEC	88	O8	H9	166306	ND	ND	ND	ND
400702008	ESC_CB0800AA	SR21990448	Human	2021	Spain	E. coli - ETEC	88	O8	H9	166306	ND	ND	ND	ND

A	B	C	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	
1	Uberstrain	Name	Barcode	Pathovar(Phylotypes	ST(Achtm 7 Gene	O Antigen(Serotype	H Antigen(Serotype	HCS(cgMLST V1 + H	age	sexe	travel	food
2	ESC_CB4516AA	12597	ESC_CB4516AA	E. coli - STEC	13869	O174	H8	219993	6	F	no	minced meat
3	ESC_YA7447AA	100306	ESC_YA7447AA	E. coli - STEC	1	O139	H1	191878	55	M	Indonesia	ND
4	ESC_YA7454AA	131435	ESC_YA7454AA	E. coli - STEC	1	O139	H1	191873	ND	ND	ND	ND
5	ESC_YA7450AA	144847	ESC_YA7450AA	E. coli - STEC	1	O139	H1	191875	ND	ND	ND	ND
6	ESC_ZA5996AA	144963	ESC_ZA5996AA	E. coli - STEC	1	O139	H1	199736	ND	ND	ND	ND
7	ESC_YA7448AA	165504	ESC_YA7448AA	E. coli - STEC	1	O139	H1	191876	ND	ND	ND	ND
8	ESC_YA7449AA	230744	ESC_YA7449AA	E. coli - STEC	1	O139	H1	191877	ND	ND	ND	ND
9	ESC_YA7453AA	250923	ESC_YA7453AA	E. coli - ETEC/STEC	10	O141	H4	191874	ND	ND	ND	ND
10	ESC_YA7452AA	262715	ESC_YA7452AA	E. coli - STEC	1	O139	H1	191875	ND	ND	ND	ND
11	ESC_YA7446AA	307636	ESC_YA7446AA	E. coli - STEC	1	O139	H1	191875	ND	ND	ND	ND

EnteroBase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari | Help | v1.2.0

Workspace: GenEpiBioTrain_Demo Rows Total: 9027 Filtered: 423

Experimental Data: STEC_GenEpiBioTrain_Demo

Species Home | Tasks | Upload Reads | Search Strains | Find ST(s) | Curate Strains | Show My Jobs | My Buddies | Locus Search | Download Schemes | HierCC equivalents

Data | Edit | View | Analysis | Workspace | Experiment | Tools

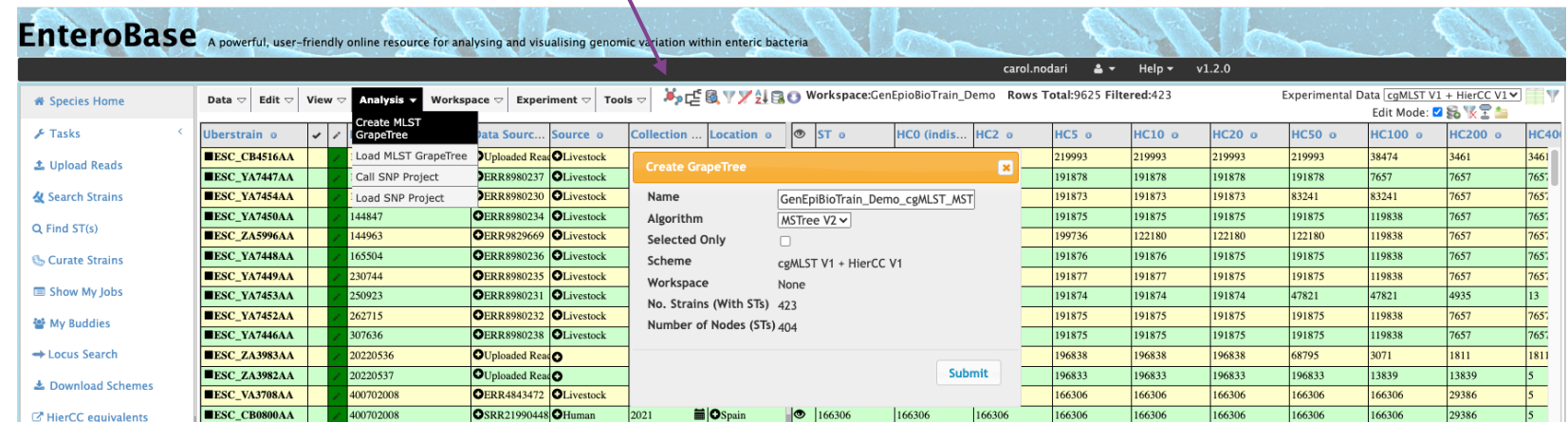
Search Strains: Save Changes, Download Data, Upload Data

Name	Barcode	Data Source	Source	Collection	Location	Pathovar(P...	ST(Achtm...	O Antigen(...	H Antigen(...	HCS(cgMLS...	age	sexe	travel	food
12597	ESC_CB4516AA	Uploaded Read	Livestock	2013	Czechia	E. coli - STEC	13869	O174	H8	219993	6	F	no	minced meat
100306	ESC_YA7447AA	ER8980237	Livestock	2016	Italy	E. coli - STEC	1	O139	H1	191878	55	M	Indonesia	ND
131435	ESC_YA7454AA	ER8980230	Livestock	2017	Italy	E. coli - STEC	1	O139	H1	191873	ND	ND	ND	ND
144847	ESC_YA7450AA	ER8980234	Livestock	2017	Italy	E. coli - STEC	1	O139	H1	191875	ND	ND	ND	ND
144963	ESC_ZA5996AA	ER89829669	Livestock	2016	Italy	E. coli - STEC	1	O139	H1	199736	ND	ND	ND	ND
165504	ESC_YA7448AA	ER8980236	Livestock	2016	Italy	E. coli - STEC	1	O139	H1	191876	ND	ND	ND	ND
230744	ESC_YA7449AA	ER8980235	Livestock	2016	Italy	E. coli - STEC	1	O139	H1	191877	ND	ND	ND	ND
250923	ESC_YA7453AA	ER8980231	Livestock	2015	Italy	E. coli - ETEC	10	O141	H4	191874	ND	ND	ND	ND
262715	ESC_YA7452AA	ER8980232	Livestock	2017	Italy	E. coli - STEC	1	O139	H1	191875	ND	ND	ND	ND
307636	ESC_YA7446AA	ER8980238	Livestock	2015	Italy	E. coli - STEC	1	O139	H1	191875	ND	ND	ND	ND
20220536	ESC_ZA3983AA	Uploaded Read		2022	Spain	E. coli - STEC	8381	O128	H2	196838	ND	ND	ND	ND
20220537	ESC_ZA3982AA	Uploaded Read		2022	Spain	E. coli - STEC	13245	O78	H4	196833	ND	ND	ND	ND
400702008	ESC_VA3708AA	Livestock		2015	Spain	E. coli - ETEC	88	O8	H9	166306	ND	ND	ND	ND
400702008	ESC_CB0800AA	SR21990448	Human	2021	Spain	E. coli - ETEC	88	O8	H9	166306	ND	ND	ND	ND

Tree visualization and data export

cgMLST-based MS tree

- Different tools are available on Enterobase to visualize **phylogenetic relationships** between genomes in a **workspace**.
- It is possible to generate **MS trees** directly on Enterobase based on the **cgMLST** + HierCC data.
- For **epidemiological** purposes, it is recommended to build a **minimum-spanning tree** (MS tree).



Enterobase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

Workspace: GenEpiBioTrain_Demo Rows Total: 9625 Filtered: 423

Experimental Data [cgMLST V1 + HierCC V1] Edit Mode: [Icons]

Species Home Tasks Upload Reads Search Strains Find ST(s) Curate Strains Show My Jobs My Buddies Locus Search Download Schemes HierCC equivalents

Uberstrain Data Source Source Collection Location ST HC0 (Indis... HC2 HC5 HC10 HC20 HC50 HC100 HC200 HC400

■ ESC_CB4516AA Load MLST GrapeTree Uploaded Reads Livestock 219993 219993 219993 219993 38474 3461 3461

■ ESC_YA7447AA Call SNP Project ERR8980237 Livestock 191878 191878 191878 191878 7657 7657 7657

■ ESC_YA7454AA Load SNP Project ERR8980230 Livestock 191873 191873 191873 83241 83241 7657 7657

■ ESC_YA7450AA 144847 ERR8980234 Livestock 191875 191875 191875 191875 119838 7657 7657

■ ESC_ZA5996AA 144963 ERR9829669 Livestock 199736 122180 122180 122180 119838 7657 7657

■ ESC_YA7448AA 165504 ERR8980236 Livestock 191876 191876 191875 191875 119838 7657 7657

■ ESC_YA7449AA 230744 ERR8980235 Livestock 191877 191877 191875 191875 119838 7657 7657

■ ESC_YA7453AA 250923 ERR8980231 Livestock 191874 191874 191874 47821 47821 4935 13

■ ESC_YA7452AA 262715 ERR8980232 Livestock 191875 191875 191875 191875 119838 7657 7657

■ ESC_YA7446AA 307636 ERR8980238 Livestock 191875 191875 191875 191875 119838 7657 7657

■ ESC_ZA3983AA 20220536 Uploaded Reads 196838 196838 196838 68795 3071 1811

■ ESC_ZA3982AA 20220537 Uploaded Reads 196833 196833 196833 13839 13839 5

■ ESC_VA3708AA 400702008 ERR4843472 Livestock 166306 166306 166306 166306 166306 29386 5

■ ESC_CB0800AA 400702008 SRR21990448 Human 2021 Spain 166306 166306 166306 166306 166306 29386 5

Create MLST GrapeTree

Name GenEpiBioTrain_Demo_cgMLST_MST

Algorithm MSTree V2

Selected Only ☐

Scheme cgMLST V1 + HierCC V1

Workspace None

No. Strains (With STs) 423

Number of Nodes (STs) 404

Submit

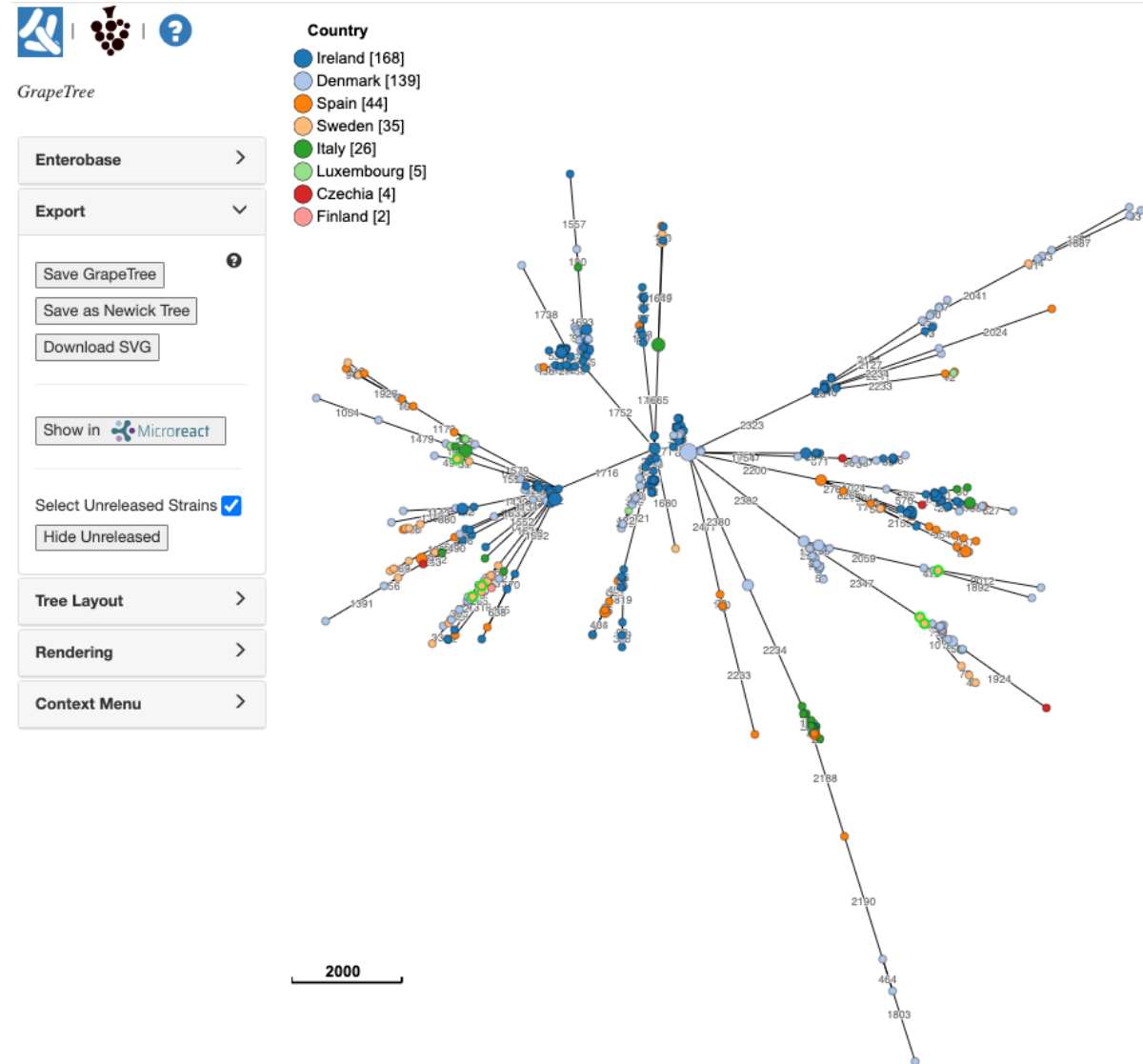
cgMLST-based MS tree

- By default, tree nodes are **coloured** using the “**Country**” metadata available. Other options are available using the menu “Tree Layout > Node style > Colour by”.
- To easily visualize **allele distances**, use “Tree Layout > Branch style > Show labels”.
- **Experimental and custom data** can also be used to colour nodes, by they first need to be imported using the “Enterobase > Import fields” button.



cgMLST-based MS tree

- Trees can also be **exported** in a variety of formats using the different the buttons in the “Export” menu.
- To be able to properly download the tree, **all strains present in the tree need to belong to you or be public**. To **check and remove any unreleased strains** from the tree, use “Select Unreleased Strains > Hide Unreleased”.
- Unfortunately, the option “Show in Microreact” is no longer functional.



SNP tree

- SNP trees can also be built with **a selection or all the genomes** from a workspace using the **dendrogram** icon.
- The **reference genome** used for the SNP calling **must be part of the workspace** and should preferably be a complete genome (but draft genomes can also be used).
- Since this job usually **takes some time**, it is possible to receive an email alert once it is completed.

Enterobase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari Help v1.2.0

Workspace: GenEpiBioTrain_Demo Rows Total: 9625 Filtered: 423

Create SNP Project

Name: GenEpiBioTrain_snp_tree_2024-05-12

Reference Genome: 52148

Selected Only: ☐

Min. % Presences: 95

Show Annotations: ☐

Email when finished: ☒

User Defined SNP List: None

Pre Existing SNPs: None

Workspace: None

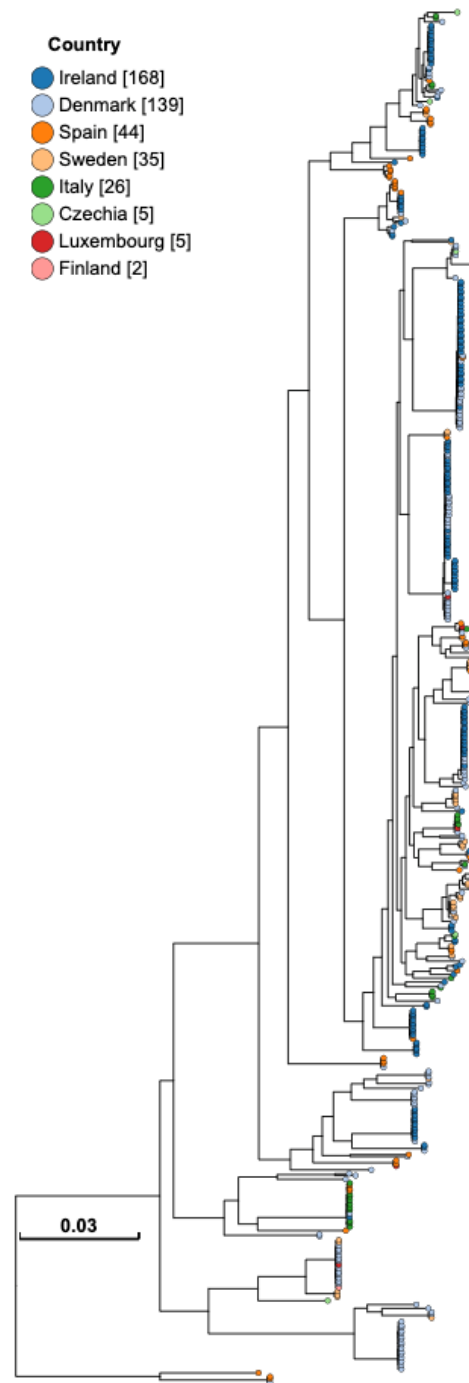
No. Strains (With Genomes): 423

Submit

Überstrain	Name	Data Source	Source	Collection	Location	ST	HC0 (indis...)	HC2	HC5	HC10	HC20
■ESC_CB4516AA	12597	Uploaded Read	○Livestock	2013	○Czechia	219993	219993	219993	219993	219993	219993
■ESC_YA7447AA	100306	○ERR8980237	○Livestock	2016	○Italy	191878	191878	191878	191878	191878	191878
■ESC_YA7454AA	131435	○ERR8980230	○Livestock	2017	○Italy	191873	191873	191873	191873	191873	191873
■ESC_YA7450AA	144847	○ERR8980234	○Livestock								
■ESC_ZA5996AA	144963	○ERR9829669	○Livestock								
■ESC_YA7448AA	165504	○ERR8980236	○Livestock								
■ESC_YA7449AA	230744	○ERR8980235	○Livestock								
■ESC_YA7453AA	250923	○ERR8980231	○Livestock								
■ESC_YA7452AA	262715	○ERR8980232	○Livestock								
■ESC_YA7446AA	307636	○ERR8980238	○Livestock								
■ESC_ZA3983AA	20220536	○Uploaded Read	○Livestock								
■ESC_ZA3982AA	20220537	○Uploaded Read	○Livestock								
■ESC_VA3708AA	400702008	○ERR4843472	○Livestock								
■ESC_CB0800AA	400702008	○SRR21990448	○Human								
■ESC_VA3699AA	401402008	○ERR4843481	○Livestock								
■ESC_CB0795AA	401402008	○SRR21990443	○Human								
■ESC_VA3694AA	401402020	○ERR4843486	○Livestock								
■ESC_MA6981AA	802802004	○ERR3336384	○Livestock								
■ESC_CA4180AA	02W30354	○Uploaded Read	○Human								
■ESC_XA6261AA	03_A_12_5_5	○SRR15115907	○Livestock								
■ESC_XA6266AA	03_A_17_5_3	○SRR15115912	○Livestock								
■ESC_XA6265AA	03_A_20_4_5	○SRR15115911	○Livestock	7/3/2018	○Ireland	184346	184346	184346	184346	184346	184346

SNP tree

- Node colouring and tree exports work similarly to the MS trees.
- It is also possible to **download a SNP matrix** (similar to a .vcf file) showing the location of the SNPs in each genome present in the tree.
- SNP projects on EnteroBase **do not remove recombining regions**, so it is not unusual to observe abnormally long branches.



EnteroBase A powerful, user-friendly online resource for analysing and visualising genomic variation within enteric bacteria

carol.nodari | Help | v1.2.0

SNP Project: GenEpiBioTrain_snp_tree_2024-05-25

Summary	
Strains	423
Reference Genome	52148
Min % Sites Present	95
No. Variant Sites	396668
Status	Complete

View

- Grape Tree
- Dendrogram
- Accessory Genome
- Microreact Project

Downloads

nwk file

identifier

- ☒ Strain Name
- ☐ Strain Barcode
- ☐ Default (Assembly Barcode)

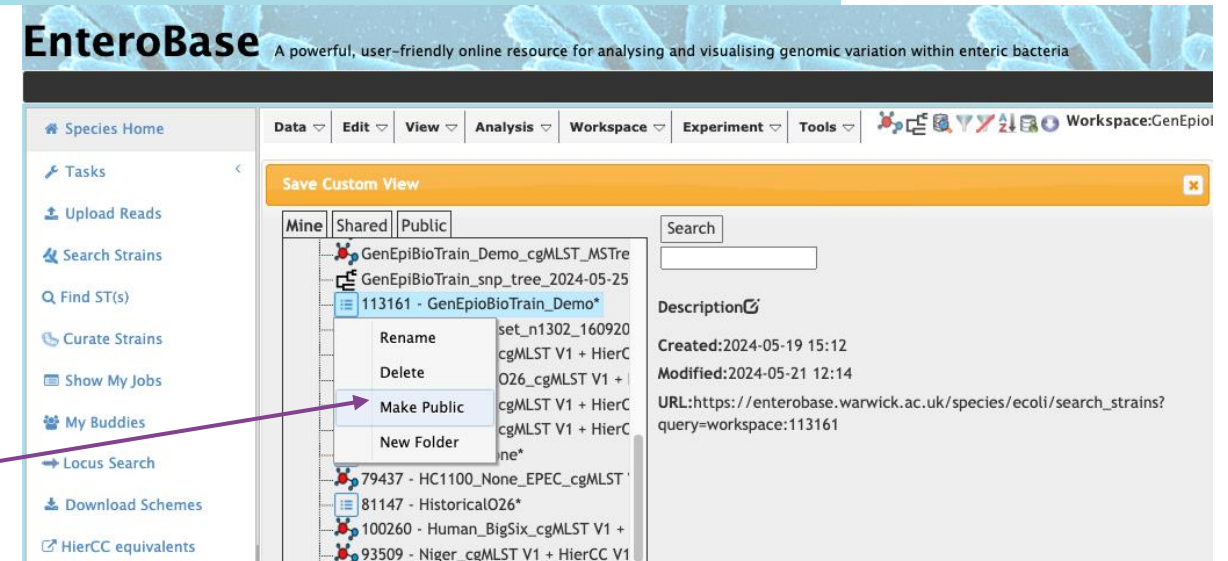
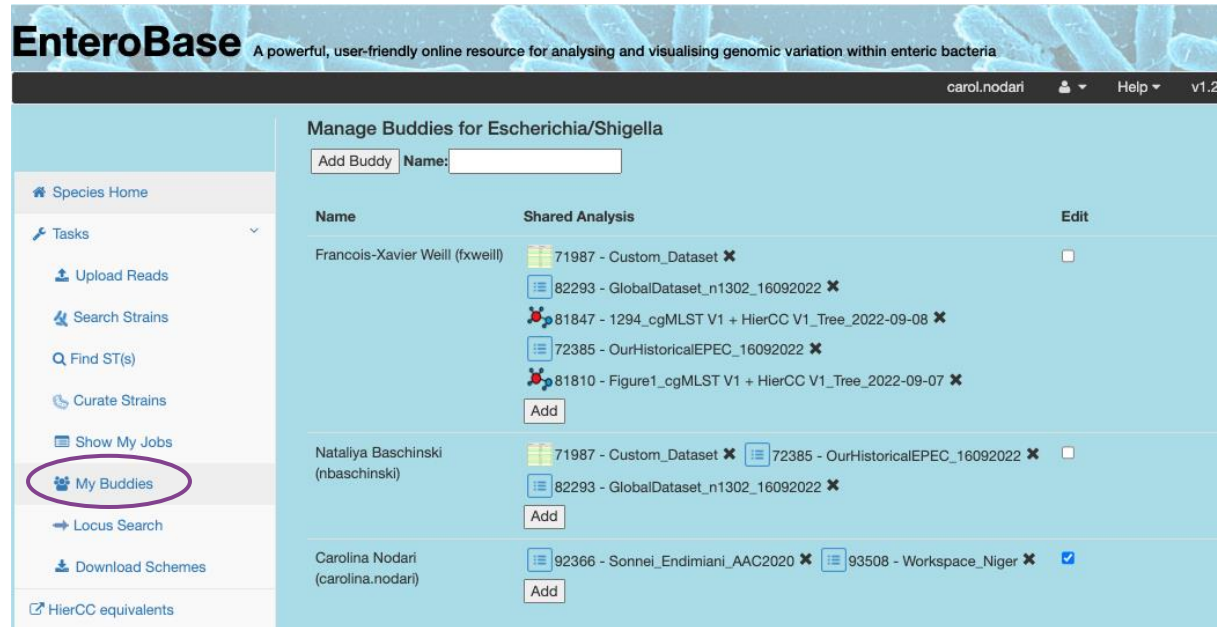
matrix

- vcf File
- mapping File

Data sharing

Data sharing

- **Workspaces, trees** and **custom views**, even when **private/unreleased**, can be shared with trusted user using the “**buddies**” system from EnteroBase.
- Depending on the rights given to the user, they can **visualize** and/or **edit** information, including **strain metadata**.
- It is also possible to make workspaces and tree **public to all EnteroBase users** by right-clicking on a given project in the “Workspace > Load” menu.



In summary

- Enterobase is a free web-based tool for visualization and analysis of genomes from different bacterial pathogens;
- It includes a back-end pipeline of quality control and genome assembling, as well as a large set of analysis tools;
- It hosts and curates the cgMLST schemes of different foodborne pathogens like *E. coli*, *Shigella* and *Salmonella*;
- The HierCC schemes provides a uniform nomenclature for cgMLST analysis, facilitating the exchange of information for epidemiological purposes;
- Data can be shared and exported in a simple way with trusted users as well as made publicly.

References

- Zhou Z, Alikhan NF, Mohamed K, Fan Y; Agama Study Group, Achtman M. The Enterobase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome Res. 2020;30(1):138-152.
doi:10.1101/gr.251678.119
- Zhou Z, Alikhan NF, Sergeant MJ, et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res. 2018;28(9):1395-1404.
doi:10.1101/gr.232397.117
- Zhou Z, Charlesworth J, Achtman M. HierCC: a multi-level clustering scheme for population assignments based on core genome MLST. Bioinformatics. 2021;37(20):3645-3646.
doi:10.1093/bioinformatics/btab234
- <https://enterobase.readthedocs.io/en/latest/index.html>

Acknowledgements

The creation of this training material was commissioned by ECDC to Institut Pasteur with the direct involvement of Carolina Silva Nodari and the contribution of the French National Reference Center for *Escherichia coli*, *Shigella* and *Salmonella* team.