



GenEpi BioTrain – Block 2 Wave 3

Case study: diarrhoeagenic *E. coli*

Carolina Silva Nodari

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

May 2024

Intended Learning Objectives

Specific objectives of this session:

1. Apply acquired knowledge on the genomics of diarrheagenic *E. coli* (DEC) to characterize a pathogenic strain
2. Analyse genomic data to identify genomic clusters of DEC
3. Combine epidemiological and genomic information to identify the source of contamination in an outbreak scenario

A local hospital in your country informs the public health agency that in the last three weeks a total of 12 patients aged 1-9 years old were hospitalized due to **bloody diarrhoea**. Five of them have also developed **haemolytic-uremic syndrome (HUS)**. Stool samples were collected from all patients and only ***E. coli*** strains were identified.

Based on your knowledge about the **pathogenicity potential of *E. coli***, do you believe that the strains identified in the stool samples are **responsible for the symptoms** presented by the patients?

Can you name a pathotype that is often associated to this clinical manifestations?

A local hospital in your country informs the public health agency that in the last three weeks a total of 12 patients aged 1-9 years old were hospitalized due to **bloody diarrhoea**. Five of them have also developed **haemolytic-uremic syndrome (HUS)**. Stool samples were collected from all patients and only ***E. coli*** strains were identified.

Based on your knowledge about the **pathogenicity potential of *E. coli***, do you believe that the strains identified in the stool samples are **responsible for the symptoms** presented by the patients?

Yes

Can you name a pathotype that is often associated to this clinical manifestations?

Enterohemorrhagic *E. coli* (EHEC)

The 12 *E. coli* strains identified in the stool samples were sent to the national reference laboratory for *Escherichia coli* of your country, where **WGS of *E. coli*** strains is routinely done.

A few days later, you are informed that the sequencing results were made available on **Enterobase**, in a **public workspace** called “**TP_Ecoli_GenEpiBioTrain_Part1**”.

Access the correct database on Enterobase and **load the workspace**.

To access the workspace:

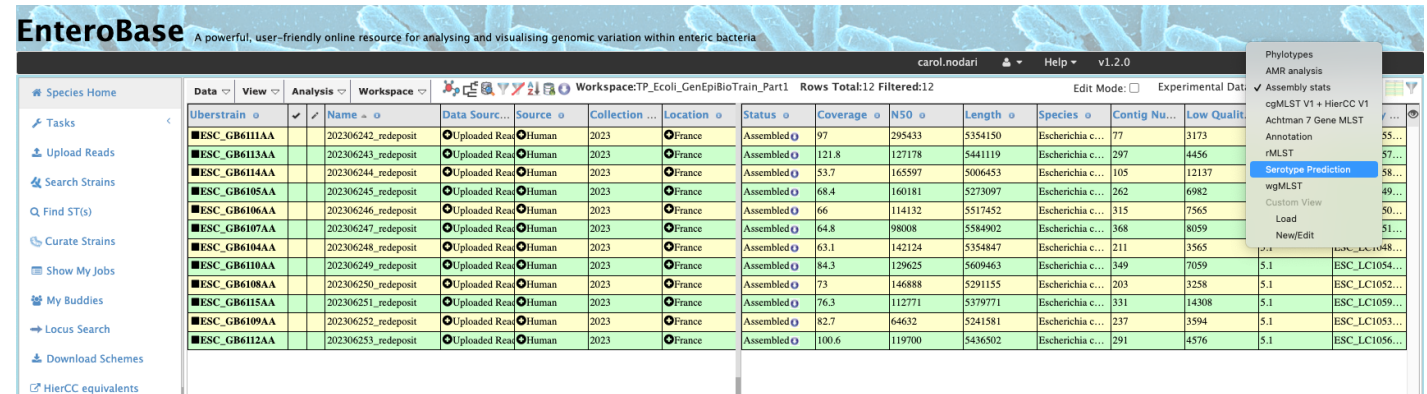
1. Go to EnteroBase -
<https://enterobase.warwick.ac.uk/>
2. Select and load the workspace -
https://enterobase.warwick.ac.uk/species/ecoli/sea_rch_strains?query=workspace:113646

Navigate through the different **Experimental Data** available for these genomes and answer the following questions:

1. How many different serotypes were identified in the strains?
 - a. 7
 - b. 8
 - c. 10
 - d. 12
2. Do you identify any EHEC among the sequenced strains? How many? Which serotype do they belong to?

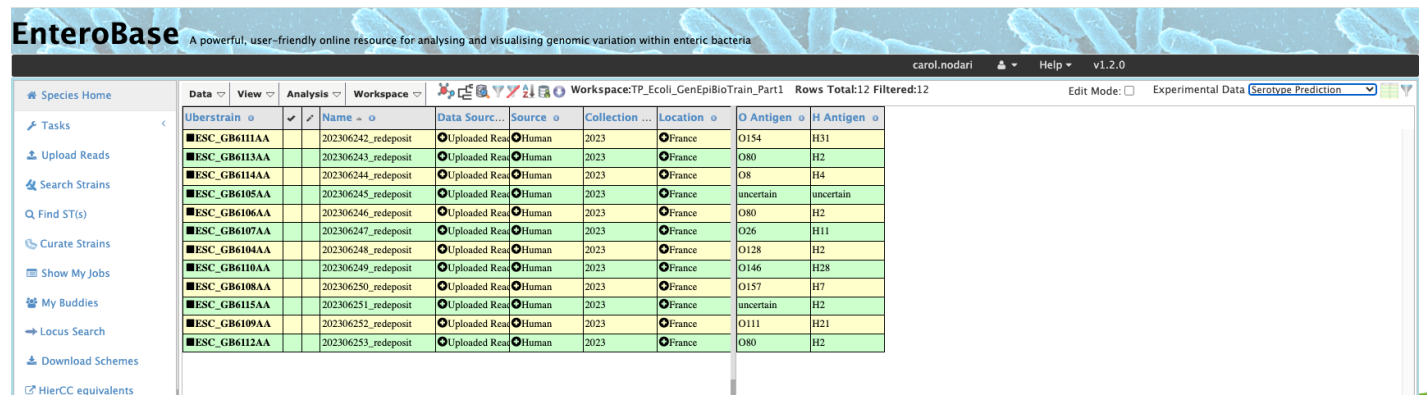
To determine the serotypes:

1. In the Experimental Data menu (top right corner), choose "Serotype Prediction"



The screenshot shows the EnteroBase web interface. The top navigation bar includes "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 table displays genomic data for various strains. The "Analysis" menu is open, and "Serotype Prediction" is selected. The table columns include: Uberstrain, Name, Data Source, Source, Collection, Location, Status, Coverage, N50, Length, Species, Contig Nu., and Low Quality.

2. Identify how many combinations of different O and H antigens are observed in the genomes.



The screenshot shows the EnteroBase web interface with the "Serotype Prediction" menu selected. The table displays genomic data for various strains, including columns for "O Antigen" and "H Antigen".

Uberstrain	Name	Data Source	Source	Collection	Location	O Antigen	H Antigen
ESC_GB6111AA	202306242_redeposit	Uploaded Read	Human	2023	France	O154	H31
ESC_GB6113AA	202306243_redeposit	Uploaded Read	Human	2023	France	O80	H2
ESC_GB6114AA	202306244_redeposit	Uploaded Read	Human	2023	France	O8	H4
ESC_GB6105AA	202306245_redeposit	Uploaded Read	Human	2023	France	uncertain	uncertain
ESC_GB6106AA	202306246_redeposit	Uploaded Read	Human	2023	France	O80	H2
ESC_GB6107AA	202306247_redeposit	Uploaded Read	Human	2023	France	O26	H11
ESC_GB6104AA	202306248_redeposit	Uploaded Read	Human	2023	France	O128	H2
ESC_GB6110AA	202306249_redeposit	Uploaded Read	Human	2023	France	O146	H28
ESC_GB6108AA	202306250_redeposit	Uploaded Read	Human	2023	France	O157	H7
ESC_GB6115AA	202306251_redeposit	Uploaded Read	Human	2023	France	uncertain	H2
ESC_GB6109AA	202306252_redeposit	Uploaded Read	Human	2023	France	O111	H21
ESC_GB6112AA	202306253_redeposit	Uploaded Read	Human	2023	France	O80	H2

Navigate through the different **Experimental Data** available for these genomes and answer the following questions:

1. How many different serotypes were identified in the strains?

- a. 7
- b. 8**
- c. 10
- d. 12

O Antigen o	H Antigen ...
O154	H31
O80	H2
O8	H4
uncertain	uncertain
O80	H2
O26	H11
O128	H2
O146	H28
O157	H7
uncertain	H2
O111	H21
O80	H2

8 distinct combinations of O and H antigens were observed:

O8:H4, O26:H11, O80:H2, O111:H21, O128:H2, O146:H28, O154:H31, and O157:H7.

For one and two strains the H and O antigens could not be identified, respectively.

Navigate through the different **Experimental Data** available for these genomes and answer the following questions:

2. Do you identify any EHEC among the sequenced strains? How many? Which serotype do they belong to?

To determine the pathotypes:

1. In the Experimental Data menu (top right corner), choose "Phylotypes"

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

carol.nodari Help v1.2.0

Workspace: TP_Ecoli_GenEpiBioTrain_Part1 Rows Total: 12 Filtered: 12

Phylotypes

AMR analysis

Assembly stats

cgMLST V1 + HierCC V1

Achtman 7 Gene MLST

Annotation

rMLST

Serotype Prediction

wgMLST

Custom View

Load

New/Edit

Uberstrain	Name	Data Source	Source	Collection	Location	Status	Coverage	N50	Length	Species	Contig Nu...	Low Qualit
■ESC_GB6111AA	202306242_redeposit	Uploaded Res	Human	2023	France	Assembled	97	295433	5354150	Escherichia c...	77	3173
■ESC_GB6113AA	202306243_redeposit	Uploaded Res	Human	2023	France	Assembled	121.8	127178	5441119	Escherichia c...	297	4456
■ESC_GB6114AA	202306244_redeposit	Uploaded Res	Human	2023	France	Assembled	53.7	165597	5006453	Escherichia c...	105	12137
■ESC_GB6105AA	202306245_redeposit	Uploaded Res	Human	2023	France	Assembled	68.4	160181	5273097	Escherichia c...	262	6982
■ESC_GB6106AA	202306246_redeposit	Uploaded Res	Human	2023	France	Assembled	66	114132	5517452	Escherichia c...	315	7565
■ESC_GB6107AA	202306247_redeposit	Uploaded Res	Human	2023	France	Assembled	64.8	98008	5584902	Escherichia c...	368	8059
■ESC_GB6104AA	202306248_redeposit	Uploaded Res	Human	2023	France	Assembled	63.1	142124	5354847	Escherichia c...	211	3565
■ESC_GB6110AA	202306249_redeposit	Uploaded Res	Human	2023	France	Assembled	84.3	129625	5609463	Escherichia c...	349	7059
■ESC_GB6108AA	202306250_redeposit	Uploaded Res	Human	2023	France	Assembled	73	146888	5291155	Escherichia c...	203	3258
■ESC_GB6115AA	202306251_redeposit	Uploaded Res	Human	2023	France	Assembled	76.3	112771	5379771	Escherichia c...	331	14308
■ESC_GB6109AA	202306252_redeposit	Uploaded Res	Human	2023	France	Assembled	82.7	64632	5241581	Escherichia c...	237	3594
■ESC_GB6112AA	202306253_redeposit	Uploaded Res	Human	2023	France	Assembled	100.6	119700	5436502	Escherichia c...	291	4576

2. Check the column "Pathovar" to identify the EHEC genomes

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

carol.nodari Help v1.2.0

Workspace: TP_Ecoli_GenEpiBioTrain_Part1 Rows Total: 12 Filtered: 12

Phylotypes

AMR analysis

Assembly stats

cgMLST V1 + HierCC V1

Achtman 7 Gene MLST

Annotation

rMLST

Serotype Prediction

wgMLST

Custom View

Load

New/Edit

Uberstrain	Name	Data Source	Source	Collection	Location	Clermont T...	Clermont T...	flmH (flmT...	Pathovar	Stx1	Stx2	ipaH	plnV	ST
■ESC_GB6111AA	202306242_redeposit	Uploaded Res	Human	2023	France	E	E	flmH31	E. coli - STEC	+	-	-	-	-
■ESC_GB6113AA	202306243_redeposit	Uploaded Res	Human	2023	France	A	A	flmH54	E. coli - EHEC	-	+	-	-	-
■ESC_GB6114AA	202306244_redeposit	Uploaded Res	Human	2023	France	C	C	flmH25	E. coli - ETEC/STEC	-	+	-	-	+
■ESC_GB6105AA	202306245_redeposit	Uploaded Res	Human	2023	France	D	D	flmH331	E. coli - EHEC	-	+	-	-	-
■ESC_GB6106AA	202306246_redeposit	Uploaded Res	Human	2023	France	A	A	flmH54	E. coli - EHEC	-	+	-	-	-
■ESC_GB6107AA	202306247_redeposit	Uploaded Res	Human	2023	France	B1	B1	flmH440	E. coli - EHEC	+	+	-	-	-
■ESC_GB6104AA	202306248_redeposit	Uploaded Res	Human	2023	France	B1	B1	flmH32	E. coli - STEC	+	+	-	-	-
■ESC_GB6110AA	202306249_redeposit	Uploaded Res	Human	2023	France	G	F	flmH222	E. coli - STEC	-	+	-	-	-
■ESC_GB6108AA	202306250_redeposit	Uploaded Res	Human	2023	France	E	E	flmH82	E. coli - EHEC	-	+	-	-	-
■ESC_GB6115AA	202306251_redeposit	Uploaded Res	Human	2023	France	B1	B1	flmH25	E. coli - EHEC	+	+	-	-	-
■ESC_GB6109AA	202306252_redeposit	Uploaded Res	Human	2023	France	B1	B1	flmH149	E. coli - STEC	-	+	-	-	-
■ESC_GB6112AA	202306253_redeposit	Uploaded Res	Human	2023	France	A	A	flmH54	E. coli - EHEC	-	+	-	-	-

Navigate through the different **Experimental Data** available for these genomes and answer the following questions:

2. Do you identify any EHEC among the sequenced strains? How many? Which serotype do they belong to?

Seven genomes were identified as EHEC.

To identify the serotypes of the EHEC genomes, there are several options:

1. **Navigate** back and forth **between the two menus** to note down the serotypes of the EHEC genomes;

To identify the serotypes of the EHEC genomes, there are several options:

1. **Navigate** back and forth between the two menus to note down the serotypes of the EHEC genomes;
2. **Filter** the genomes in the workspace by pathotype and move back to the "Serotype prediction" to check the serotypes of the filtered genomes

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

carol.nodari Help v1.2.0

Workspace: TP_Ecoli_GenEpiBioTrain_Part1 Rows Total: 12 Filtered: 12

Experimental Data | Phylotypes

Uberstrain	Name	Data Source	Source	Collection	Location	Clermont T...	Clermont T...	fimH (fimT...	Pathovar	Stx1	Stx2
ESC_GB6113AA	202306242_redeposit	Uploaded Read	Human	2023	France	E	E	fimH31	E. coli - STEC	+	-
ESC_GB6113AA	202306243_redeposit	Uploaded Read	Human	2023	France	A	A	fimH54	E. coli - EHEC	-	+
ESC_GB6114AA	202306244_redeposit	Uploaded Read	Human	2023	France	C	C	fimH25	E. coli - ETEC/STEC	-	+
ESC_GB6105AA	202306245_redeposit	Uploaded Read	Human	2023	France	D	D	fimH331	E. coli - EHEC	-	+
ESC_GB6106AA	202306246_redeposit	Uploaded Read	Human	2023	France	A	A	fimH54	E. coli - EHEC	-	+
ESC_GB6107AA	202306247_redeposit	Uploaded Read	Human	2023	France	B1	B1	fimH440	E. coli - EHEC	+	+
ESC_GB6104AA	202306248_redeposit	Uploaded Read	Human	2023	France	B1	B1	fimH32	E. coli - STEC	+	+
ESC_GB6110AA	202306249_redeposit	Uploaded Read	Human	2023	France	G	F	fimH222	E. coli - STEC	-	+
ESC_GB6108AA	202306250_redeposit	Uploaded Read	Human	2023	France	E	E	fimH82	E. coli - EHEC	-	+
ESC_GB6115AA	202306251_redeposit	Uploaded Read	Human	2023	France	B1	B1	fimH25	E. coli - EHEC	+	+
ESC_GB6109AA	202306252_redeposit	Uploaded Read	Human	2023	France	B1	B1	fimH149	E. coli - STEC	-	+
ESC_GB6112AA	202306253_redeposit	Uploaded Read	Human	2023	France	A	A	fimH54	E. coli - EHEC	-	+

Filter Data

FieldName: Pathovar

Contains: EHEC

E. coli - EHEC

Filter

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

carol.nodari Help v1.2.0

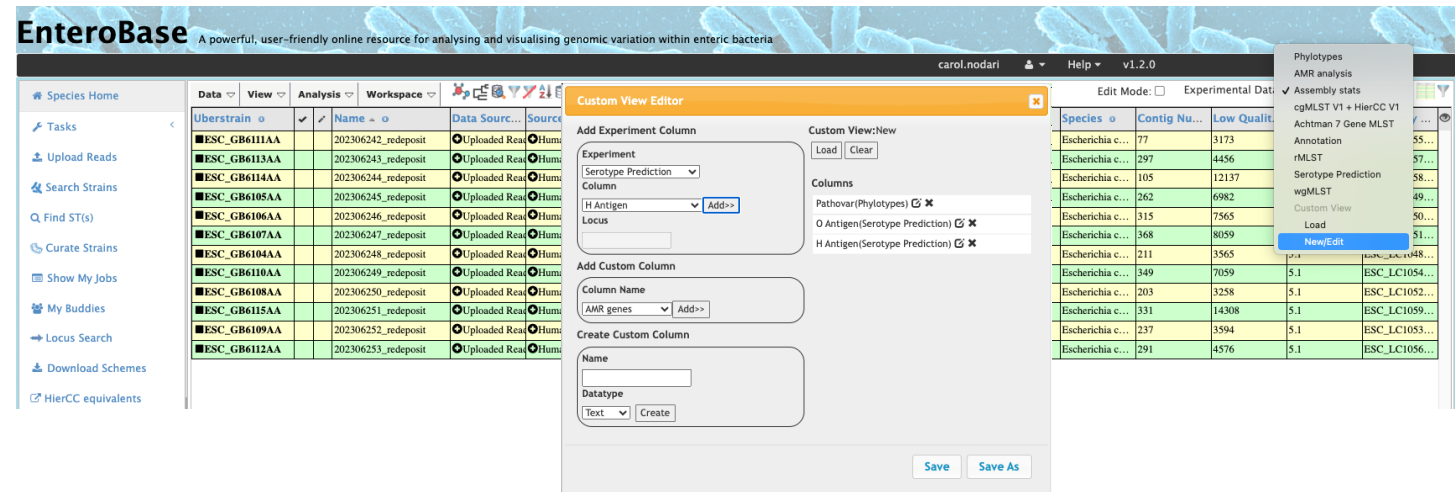
Workspace: TP_Ecoli_GenEpiBioTrain_Part1 Rows Total: 12 Filtered: 7

Experimental Data | Serotype Prediction

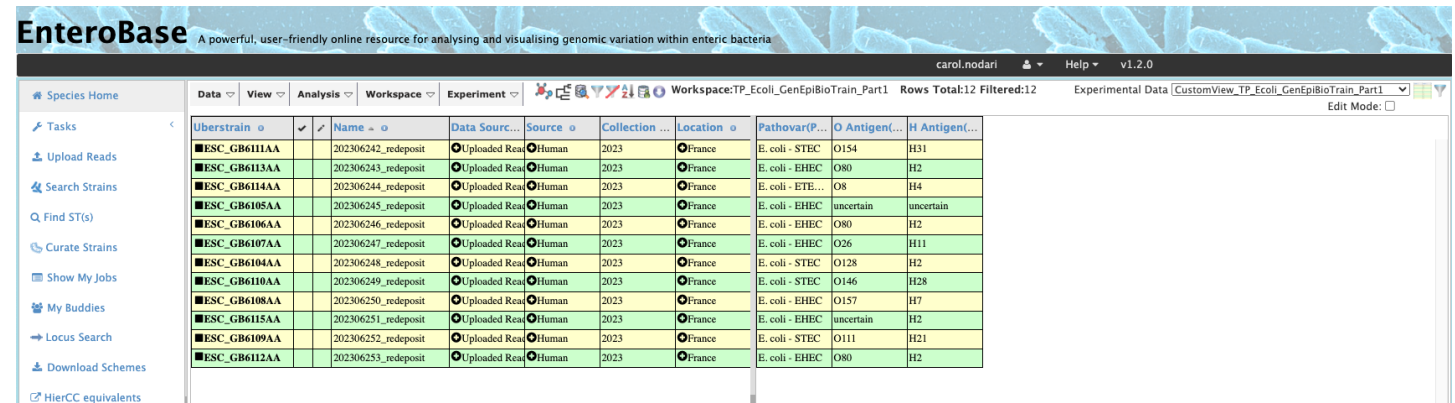
Uberstrain	Name	Data Source	Source	Collection	Location	O Antigen	H Antigen
ESC_GB6113AA	202306243_redeposit	Uploaded Read	Human	2023	France	O80	H2
ESC_GB6105AA	202306245_redeposit	Uploaded Read	Human	2023	France	uncertain	uncertain
ESC_GB6106AA	202306246_redeposit	Uploaded Read	Human	2023	France	O80	H2
ESC_GB6107AA	202306247_redeposit	Uploaded Read	Human	2023	France	O26	H11
ESC_GB6108AA	202306250_redeposit	Uploaded Read	Human	2023	France	O157	H7
ESC_GB6115AA	202306251_redeposit	Uploaded Read	Human	2023	France	uncertain	H2
ESC_GB6112AA	202306253_redeposit	Uploaded Read	Human	2023	France	O80	H2

To identify the serotypes of the EHEC genomes, there are several options:

1. **Navigate** back and forth between the two menus to note down the serotypes of the EHEC genomes;
2. **Filter** the genomes in the workspace by pathotype and move back to the "Serotype prediction" to check the serotypes of the filtered genomes
3. **Create a custom view** with the columns "Pathovar", "O antigen" and "H antigen" to visualize the data side by side (here you can also filter)



The screenshot shows the EnteroBase Custom View Editor interface. On the left, a table lists various EHEC strains (e.g., ESC_GB6111AA, ESC_GB6113AA) with their accession numbers and upload status. The main panel is titled "Custom View Editor" and contains sections for "Add Experiment Column" (with a dropdown for "Serotype Prediction" and a button "Add->"), "Add Custom Column" (with a text input for "Column Name" and a "Create" button), and "Create Custom Column" (with a text input for "Name" and a "Create" button). On the right, a table displays the results of the custom view, showing columns for "Species", "Contig Nu...", "Low Quality", and "Serotype Prediction".



The screenshot shows the EnteroBase Workspace interface. The top bar indicates "Workspace: TP_Ecoli_GenEpiBioTrain_Part1" and "Rows Total: 12 Filtered: 12". The main table displays columns for "Uberstrain", "Name", "Data Source", "Source", "Collection", "Location", "Pathovar", "O Antigen", and "H Antigen". The table lists 12 rows of EHEC data, including strains like ESC_GB6111AA, ESC_GB6113AA, and ESC_GB6114AA, with their respective serotypes and locations.

Navigate through the different **Experimental Data** available for these genomes and answer the following questions:

Pathovar(P...	O Antigen(...	H Antigen(...
E. coli - EHEC	O157	H7
E. coli - EHEC	O26	H11
E. coli - EHEC	O80	H2
E. coli - EHEC	O80	H2
E. coli - EHEC	O80	H2
E. coli - EHEC	uncertain	uncertain
E. coli - EHEC	uncertain	H2

2. Do you identify any EHEC among the sequenced strains? How many? Which serotype do they belong to?

Seven genomes were identified as EHEC. They belong to (at least) three serotypes: O157:H7, O26:H11, and O80:H2.

Navigate through the different **Experimental Data** available for these genomes and answer the following questions:

3. (Bonus question) Which are the virulence markers evaluated by EnteroBase that allow the identification of EHEC?
 - a. Presence of *eae* and absence of pInv
 - b. Presence of *stx1* and *stx2*
 - c. Absence of LT and ST
 - d. Presence of *stx1/stx2* and *eae*
4. (Bonus question) Which was the most frequent virulence marker evaluated by EnteroBase identified in isolates in this dataset?
 - a. *eae*
 - b. *stx1*
 - c. *stx2*
 - d. ST

Navigate through the different **Experimental Data** available for these genomes and answer the following questions:

3. (Bonus question) Which are the virulence markers evaluated by EnteroBase that allow the identification of EHEC?
 - a. Presence of *eae* and absence of pInv
 - b. Presence of *stx1* and *stx2*
 - c. Absence of LT and ST
 - d. Presence of *stx1* or *stx2* and *eae***
4. (Bonus question) Which was the most frequent virulence marker evaluated by EnteroBase identified in isolates in this dataset?
 - a. *eae*
 - b. *stx1*
 - c. *stx2***
 - d. ST

Pathovar o	Stx1 o	Stx2 o	ipaH o	pInv o	ST o	LT o	eae o
E. coli - STEC	+	+	-	-	-	-	-
E. coli - EHEC	-	+	-	-	-	-	+
E. coli - EHEC	-	+	-	-	-	-	+
E. coli - EHEC	+	+	-	-	-	-	+
E. coli - EHEC	-	+	-	-	-	-	+
E. coli - STEC	-	+	-	-	-	-	-
E. coli - STEC	-	+	-	-	-	-	-
E. coli - STEC	+	-	-	-	-	-	-
E. coli - EHEC	-	+	-	-	-	-	+
E. coli - EHEC	-	+	-	-	-	-	+
E. coli - ETEC/STEC	-	+	-	-	+	-	-
E. coli - EHEC	+	+	-	-	-	-	+

Without building a tree, can you identify any epidemiologically relevant **genomic cluster** among the genomes? Which of the following Experimental Data available on EnteroBase did you use for it?

- a. Serotype O80:H2
- b. Sequence Type ST301
- c. HC100 1952
- d. HC5 36540

(If have created a Custom View for the previous questions, try adding these new columns to it.)

If a genomic cluster was identified, how many isolates belonged to it?

To add new columns to an existing Custom View:

1. Go to "Custom View > New/Edit"
2. Load your Custom View
3. Add the new columns
4. Save and close the Custom View Editor
5. Go to "Custom View > Load" and choose the updated custom view

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

carol.nodari 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: TP_Ecoli_GenEpiBioTrain_Part1 Rows Total:12 Filtered:12

Uberstrain	Name	Data Source	Source	Collection	Location	Status	Coverage	NSO	Length	Species	Contig Nu...	Low Quality
ESC_GB6111AA	202306242_redeposit	Uploaded Real	Human	2023	France	Assembled	97	295433	5354150	Escherichia c...	77	3173
ESC_GB6113AA	202306243_redeposit	Uploaded Real	Human	2023	France	Assembled	121.8	127178	5441119	Escherichia c...	297	4456
ESC_GB6114AA	202306244_redeposit	Uploaded Real	Human	2023	France	Assembled	53.7	165597	5006453	Escherichia c...	105	12137
ESC_GB6105AA	202306245_redeposit	Uploaded Real	Human	2023	France	Assembled	68.4	160181	5273097	Escherichia c...	262	6982
ESC_GB6106AA	202306246_redeposit	Uploaded Real	Human	2023	France	Assembled	66	114132	5517452	Escherichia c...	315	7565
ESC_GB6107AA	202306247_redeposit	Uploaded Real	Human	2023	France	Assembled	64.8	98008	5584902	Escherichia c...	368	8059
ESC_GB6104AA	202306248_redeposit	Uploaded Real	Human	2023	France	Assembled	63.1	142124	5354847	Escherichia c...	211	3565
ESC_GB6110AA	202306249_redeposit	Uploaded Real	Human	2023	France	Assembled	84.3	129625	5609463	Escherichia c...	349	7059
ESC_GB6108AA	202306250_redeposit	Uploaded Real	Human	2023	France	Assembled	73	146888	5291155	Escherichia c...	203	3258
ESC_GB6115AA	202306251_redeposit	Uploaded Real	Human	2023	France	Assembled	76.3	112771	5379771	Escherichia c...	331	14308
ESC_GB6109AA	202306252_redeposit	Uploaded Real	Human	2023	France	Assembled	82.7	64632	5241581	Escherichia c...	237	3594
ESC_GB6112AA	202306253_redeposit	Uploaded Real	Human	2023	France	Assembled	100.6	119700	5436502	Escherichia c...	291	4576

Phylotypes
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: **Achtman 7 Gene MLST**

Column: **ST**

Locus:

Add Custom Column

Column Name: **AMR genes**

Create Custom Column

Name:

Datatype: **Text**

Custom View: **CustomView_TP_Ecoli_GenEpiBioTrain_Part1**

Load

Columns

- Pathovar(Phylotypes) ☒
- O Antigen(Serotype Prediction) ☒
- H Antigen(Serotype Prediction) ☒
- ST(Achtman 7 Gene MLST) ☒
- HC100(cgMLST V1 + HierCC V1) ☒
- HC5(cgMLST V1 + HierCC V1) ☒

Save Save As

Without building a tree, can you identify any epidemiologically relevant **genomic cluster** among the genomes? Which of the following Experimental Data available on EnteroBase did you use for it?

- a. Serotype O80:H2
- b. Sequence Type ST301
- c. HC100 1952
- d. HC5 36540**

(If have created a Custom View for the previous questions, try adding these new columns to it.)

Pathovar(Phylotypes) o	O Antigen(...	H Antigen(...	ST(Achtma...	HC100(cg...	HC5(cgMLS...
E. coli - STEC	O128	H2	4748	14844	84508
E. coli - EHEC	uncertain	uncertain	32	1918	84516
E. coli - EHEC	O80	H2	301	1952	254784
E. coli - EHEC	O26	H11	21	94	202927
E. coli - EHEC	O157	H7	11	567	100049
E. coli - STEC	O111	H21	40	56101	254782
E. coli - STEC	O146	H28	738	1086	254780
E. coli - STEC	O154	H31	1892	52128	84512
E. coli - EHEC	O80	H2	301	1952	36540
E. coli - EHEC	O80	H2	301	1952	36540
E. coli - ETEC/STEC	O8	H4	88	48027	251028
E. coli - EHEC	uncertain	H2	17	3	94839

If a genomic cluster was identified, how many isolates belonged to it? **2**

Build a **MS tree** and colour the nodes based on their **HC5** values to identify the genomic cluster.

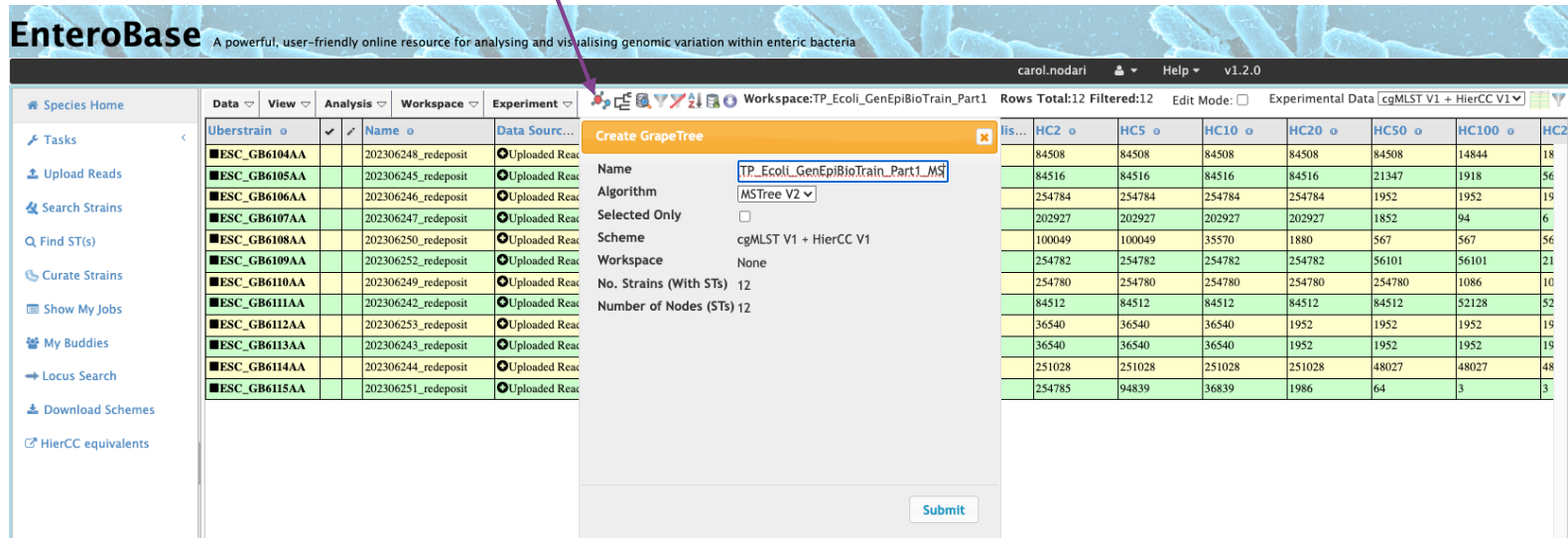
What is the **allelic distance** between the two genomes that are part of the cluster? What is the distance to the **other EHEC O80:H2** present in the dataset?

Two tips:

- a. Remember to allow pop-ups in your internet browser to make sure the tree will automatically load in a new window.
- b. You can zoom in and out and change the scale of the branches in the tree to be able to see the distances.

To build and colour the MS Tree:

1. Go to "Experimental Data > cgMLST V1 + HierCC V1"
2. Click on the **GrapeTree** icon
3. Choose a name and submit the tree.
4. After it has loaded, go to "Enterobase > **Import Fields**" and add the column "HC5" from the "cgMLST V1 + HierCC V1" Experiment.



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

carol.nodari | Help | v1.2.0

Workspace: TP_Ecoli_GenEpiBioTrain_Part1 Rows Total: 12 Filtered: 12 Edit Mode: Experimental Data [cgMLST V1 + HierCC V1]

Create GrapeTree

Name: TP_Ecoli_GenEpiBioTrain_Part1_MS

Algorithm: MSTree V2

Selected Only: ☐

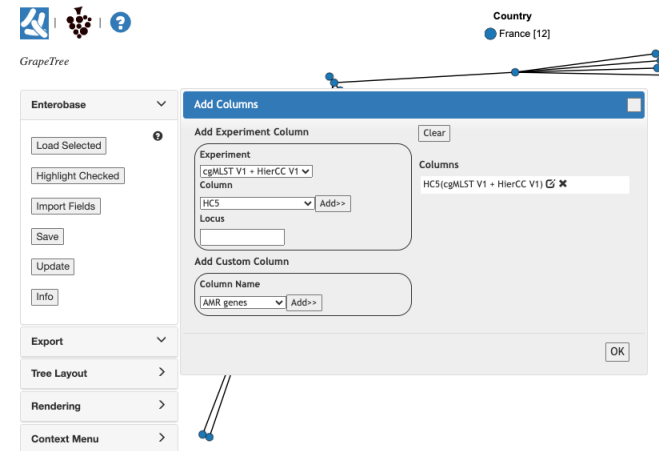
Scheme: cgMLST V1 + HierCC V1

Workspace: None

No. Strains (With STs): 12

Number of Nodes (STs): 12

Submit



Enterobase

Load Selected

Highlight Checked

Import Fields

Save

Update

Info

Export

Tree Layout

Rendering

Context Menu

Add Columns

Add Experiment Column

Experiment: cgMLST V1 + HierCC V1

Column: HC5

Add Custom Column

Column Name: AMR genes

OK

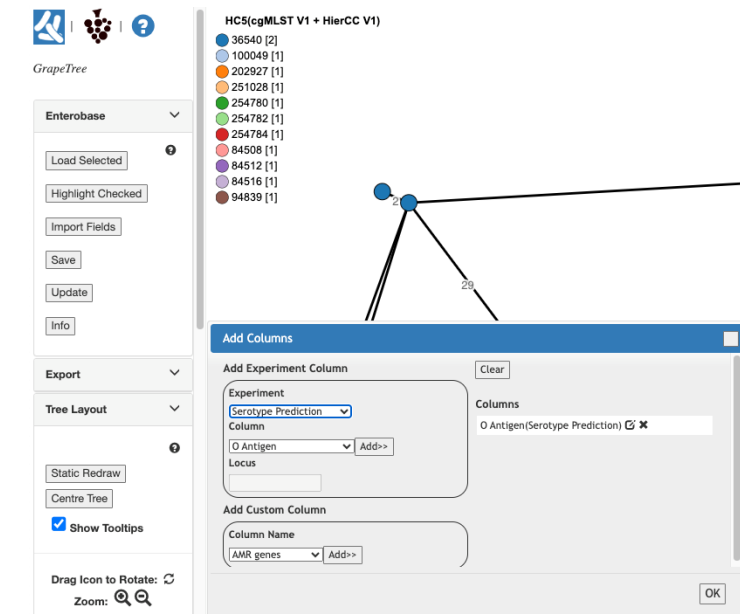
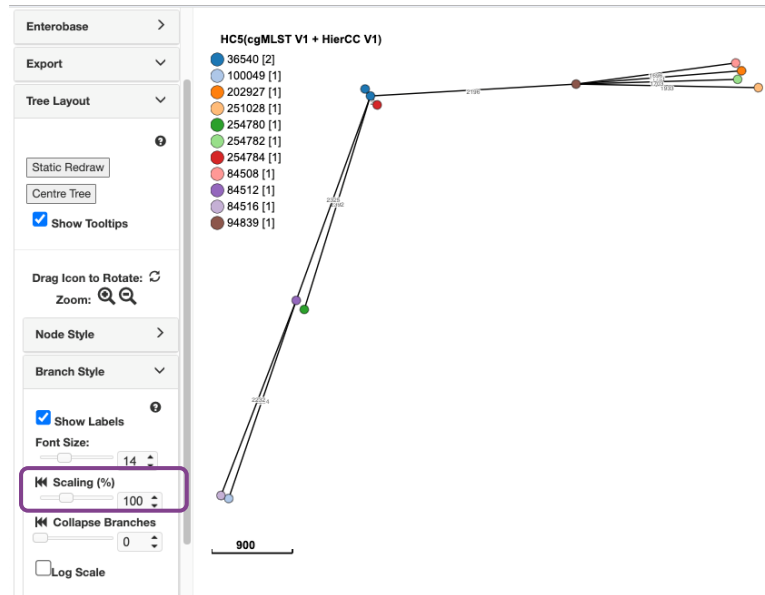
GrapeTree

Country: France [12]

To visualize allelic distances in the MS Tree:

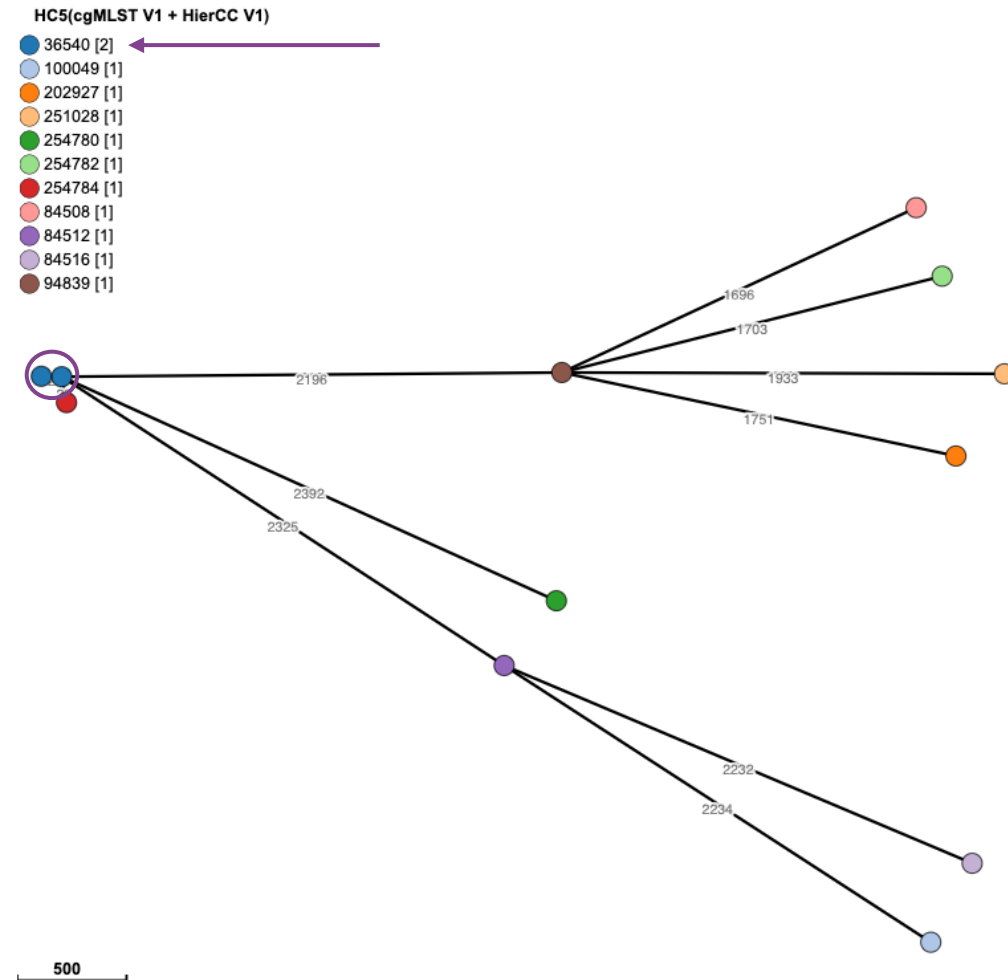
5. Go to "Tree layout > Branch Style" and check the "**Show Labels**" to see the **allelic distances** in the branches.
6. Change the **scaling** in the "Tree layout > Branch Style" to be able to see the distances in the shorter branches.

To identify the other O80:H2 isolate, colour the nodes based on the O antigen (follow the same instructions as for the HC5).



Build a **MS tree** and colour the nodes based on their **HC5** values to identify the genomic cluster.

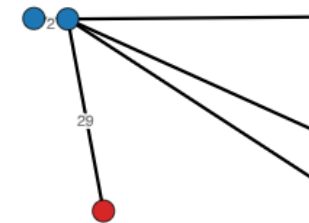
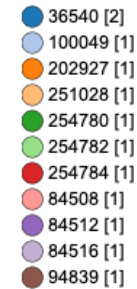
What is the **allelic distance** between the two genomes that are part of the cluster? What is the distance to the **other EHEC O80:H2** present in the dataset?



Build a **MS tree** and colour the nodes based on their **HC5** values to identify the genomic cluster.

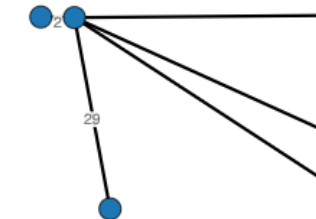
What is the **allelic distance** between the two genomes that are part of the cluster? What is the distance to the **other EHEC O80:H2** present in the dataset? **The two HC5_36540 have an AD = 2. The other O80:H2 is at an AD = 29 to the closest strain.**

HC5(cgMLST V1 + HierCC V1)



500

O Antigen(Serotype Prediction)



500

Considering that you identified a genomic cluster with **two cases of infection by an EHEC of serotype O80:H2 HC5_36540**, you decide to start an epidemiological investigation to determine if there was a **common source** of the *E. coli* strain.

Interviews with the parents of the two children affected by the outbreak strain were performed, and provided the following information:

Patient 1 (F, 4yo)	Patient 2 (M, 3yo)
Goes to kindergarten 1	Goes to kindergarten 2
Consumed minced meat from brand A at home 5 days before symptoms onset	Consumed minced meat from brand D at home 10 days before symptoms onset
Consumed pasteurized milk from brand B at home 10 days before symptoms onset	Consumed chocolate mousse made with eggs from a local farm 3 days before symptoms onset
Consumed smoked salmon from brand C 7 days before symptoms onset	Consumed cheese of unpasteurized milk from brand E 7 days before symptoms onset
No visits to farms	Visited a farm with cows 7 days before symptoms onset
Have a cat at home	Have a dog and a turtle at home
Came back from France 2 weeks before symptoms onset	No traveling abroad
Goes to public swimming pool 1 every week-end	Goes to public swimming pool 1 every week

Based on your knowledge of the **EHEC reservoirs**, which of these information could be considered a risk factor for EHEC infection? (Multiple answers possible)

Patient 1 (F, 4yo)	Patient 2 (M, 3yo)
Goes to kindergarten 1	Goes to kindergarten 2
Consumed minced meat from brand A at home 5 days before symptoms onset	Consumed minced meat from brand D at home 10 days before symptoms onset
Consumed pasteurized milk from brand B at home 10 days before symptoms onset	Consumed chocolate mousse made with eggs from a local farm 3 days before symptoms onset
Consumed smoked salmon from brand C 7 days before symptoms onset	Consumed cheese of unpasteurized milk from brand E 7 days before symptoms onset
No visits to farms	Visited a farm with cows 7 days before symptoms onset
Have a cat at home	Have a dog and a turtle at home
Came back from France 2 weeks before symptoms onset	No traveling abroad
Goes to public swimming pool 1 every week-end	Goes to public swimming pool 1 every week

Based on your knowledge of the **EHEC reservoirs**, which of these information could be considered a risk factor for EHEC infection? (Multiple answers possible)

Patient 1 (F, 4yo)	Patient 2 (M, 3yo)
Goes to kindergarten 1	Goes to kindergarten 2
Consumed minced meat from brand A at home 5 days before symptoms onset	Consumed minced meat from brand D at home 10 days before symptoms onset
Consumed pasteurized milk from brand B at home 10 days before symptoms onset	Consumed chocolate mousse made with eggs from a local farm 3 days before symptoms onset
Consumed smoked salmon from brand C 7 days before symptoms onset	Consumed cheese of unpasteurized milk from brand E 7 days before symptoms onset
No visits to farms	Visited a farm with cows 7 days before symptoms onset
Have a cat at home	Have a dog and a turtle at home
Came back from France 2 weeks before symptoms onset	No traveling abroad
Goes to public swimming pool 1 every week-end	Goes to public swimming pool 1 every week

Traceability efforts with the family and local food authorities allowed the identification of the production batches of the **minced meat from brands A and D** and samples were collected.

Surprisingly, both samples came back positive for ***stx2*-producing *E. coli***, but the clinical lab in charge of the analysis could **not serotype** the isolates.

A faecal swab of the one the **cows from the farm** visited by patient 2 also came back positive for ***E. coli* O80**, but **no PCR** for *stx* genes were performed.

All strains were then sent for sequencing to determine if they were genetically related to the clinical strains. The genomes were then submitted to Enterobase for comparison.

Using the information provided below, **retrieve the genomes** of the three EHEC strains from EnteroBase, and **save them in a workspace**.

- Minced meat from brand A: Strain name 12-2023_Strain13, Strain barcode ESC_GB9317AA
- Minced meat from brand D: Strain name 12-2023_Strain14, Strain barcode ESC_GB9313AA
- Faecal swab from cow: Strain name 12-2023_Strain16, Strain barcode ESC_GB9315AA

Recover the genomes of the **two clinical strains** that belong to the **outbreak** and add them to the **same workspace**. (Hint: go back to the public workspace "TP_Ecoli_GenEpiBioTrain_Part1", select the genomes of HC5 36540 and add them to your workspace).

Are any of the food/animal strains related to the outbreak?

Figure 2 displays two GrapeTree phylogenetic trees. The left tree, titled "HC5(cgMLST V1 + HierCC V1)", shows three sources: Farm cow [1] (blue), Minced meat A [1] (light blue), and Minced meat D [1] (orange). The right tree, titled "Source", shows four sources: Farm cow [1] (blue), Minced meat A [1] (light blue), Minced meat D [1] (orange), and Patient 1 [1] (green). Both trees include a scale bar of 5 and a "GrapeTree" logo.

Using the information provided below, **retrieve the genomes** of the three EHEC strains from Enterobase, and **save them in a workspace**.

- Minced meat from brand A: Strain name 12-2023_Strain13, Strain barcode ESC_GB9317AA
- Minced meat from brand D: Strain name 12-2023_Strain14, Strain barcode ESC_GB9313AA
- Faecal swab from cow: Strain name 12-2023_Strain16, Strain barcode ESC_GB9315AA

Recover the genomes of the **two clinical strains** that belong to the **outbreak** and add them to the **same workspace**. (Hint: go back to the public workspace "TP_Ecoli_GenEpiBioTrain_Part1", select the genomes of HC5 36540 and add them to your workspace).

Are any of the food/animal strains related to the outbreak?

Yes, minced meat from brand D was contaminated with the outbreak strain.

Since the EHEC strain found in the minced meat from brand A was not related to the outbreak strain, new interviews with the parents of patient 1 were conducted to identify other potential sources of contamination, but no new evidences were found.

In the meantime, a **new patient 3** (M, 5 yo) affected by the outbreak strain was identified.

Since he also went to **kindergarten 1**, its kitchen was checked and **minced meat from brand D**, but from a different batch was found. There was a registry that both patients 1 and 3 had consumed meat from that batch in the kindergarten.

Microbiological analysis demonstrated that the meat was also contaminated by a ***stx2*-positive *E. coli*** (Strain name 12-2023_Strain20 – Strain barcode ESC_GB9314AA), which was also sequenced.

Add this new genome to your workspace and determine if the new meat sample was contaminated with the same strain.

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

Workspace:TP_Ecoli_GenEpiBioTrain_Part2

Rows Total:6 Filtered:6

Experimental Data

CustomView_TP_Ecoli_GenEpiBioTrain_Part2

Uberstrain

✓

✗

✕

Name

Data Sour...

Source

Collection ...

Location

Pathovar(P...

O Antigen(...

H Antigen(...

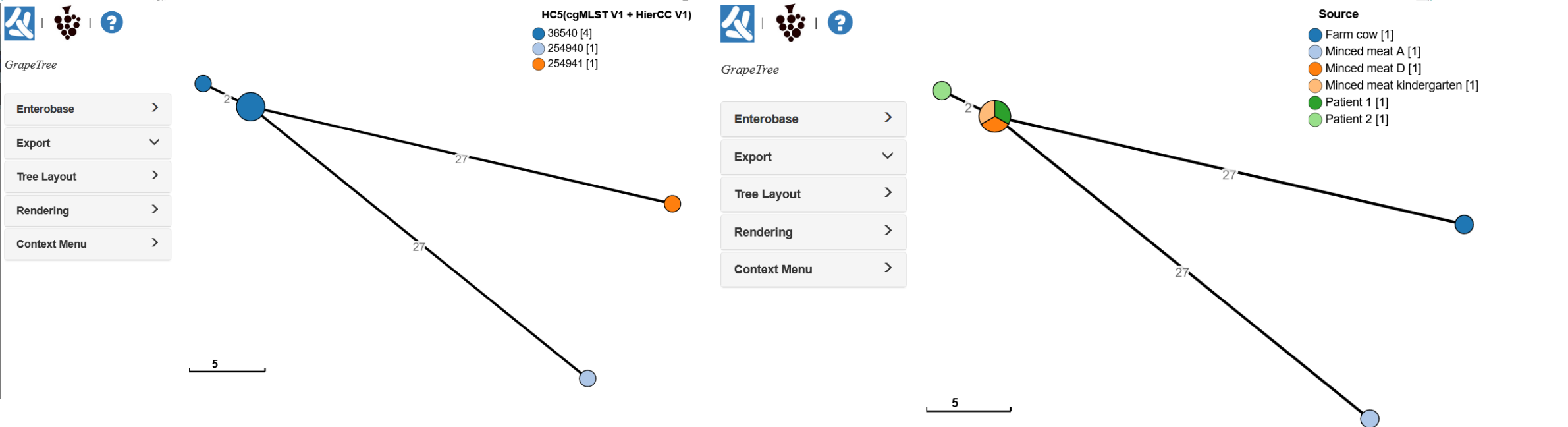
ST(Achtm...

HC100(cg...

HCS(cgMLS...

Source

■ESC_GB6112AA	202306253_redeposit	Uploaded Read	Human	2023	France	E. coli - EHEC	O80	H2	301	1952	36540	Patient 1
■ESC_GB6113AA	202306243_redeposit	Uploaded Read	Human	2023	France	E. coli - EHEC	O80	H2	301	1952	36540	Patient 2
■ESC_GB9313AA	12-2023_Strain14	Uploaded Read				E. coli - EHEC	uncertain	H2	301	1952	36540	Minced meat D
■ESC_GB9314AA	12-2023_Strain20	Uploaded Read				E. coli - EIEC...	uncertain	H2	301	1952	36540	Minced meat kindergarten
■ESC_GB9315AA	12-2023_Strain16	Uploaded Read				E. coli - EHEC	-	H2	301	1952	254941	Farm cow
■ESC_GB9317AA	12-2023_Strain13	Uploaded Read				E. coli - EHEC	uncertain	H2	301	1952	254940	Minced meat A



Since the EHEC strain found in the minced meat from brand A was not related to the outbreak strain, new interviews with the parents of patient 1 were conducted to identify other potential sources of contamination, but no new evidences were found.

In the meantime, a **new patient 3** (M, 5 yo) affected by the outbreak strain was identified.

Since he also went to **kindergarten 1**, its kitchen was checked and **minced meat from brand D**, but from a different batch was found. There was a registry that both patients 1 and 3 had consumed meat from that batch in the kindergarten.

Microbiological analysis demonstrated that the meat was also contaminated by a ***stx2*-positive *E. coli*** (Strain name 12-2023_Strain20 – Strain barcode ESC_GB9314AA), which was also sequenced.

Add this new genome to your workspace and determine if the new meat sample was contaminated with the same strain.

Yes, the new genome belonged to the same genomic cluster.

In summary

- Genomic surveillance allows a more refined characterization of diarrheagenic *E. coli* beyond serotype determination;
- Enterobase provides a variety of tools can be used for the genomic surveillance of EHEC;
- Knowing the reservoirs and contamination sources of the different pathotypes of *E. coli* can help targeting epidemiological investigation of potential risk factors;
- The surveillance of enteric *E. coli* must be homogeneous in all levels of the One Health framework, facilitating traceability and comparison of clinically relevant strains.

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.