

Defining the uniqueness of isolates using the NCBI Pathogen Detection system

Using the NCBI Pathogen Detection system a request for the isolates under study can be submitted to the database. The screenshots shown below display the results for the four Russian isolates described in our study (available under URL link https://www.ncbi.nlm.nih.gov/pathogens/isolates/#SLR1_8250%20OR%20SLR1_7627%20OR%20SLR1_8245%20OR%20SLR1_8094). To identify the relatedness of the studied isolates to other isolates in the NCBI Pathogen Detection system, a 50-SNP single-linkage clustering was automatically created (Figure S1). When the “SNP cluster” field remains empty, it indicates that the system did not find isolates in the database with ≤ 50 SNP differences. For the isolates in our study, i.e. SLR1_7627, SLR1_8094, and SLR1_8250 no closely related isolates were found in the databases. However, isolates SLR1_8245 and SLR1_8239 were similar (0 SNP differences) and they formed a SNP cluster named PDS000093458.2 (Figure S2). These two isolates SLR1_8245 and SLR1_8239 belonged to the same salmonellosis outbreak (El’ban village in the Khabarovsk Territory) (Figure S2). So, we concluded that our isolates did not have a close relationship (at the ≤ 50 SNP threshold) to other isolates in the NCBI Pathogen Detection system or to each other – with the exception of isolates SLR1_8245 and SLR1_8239.

Search

SLR1_8250 OR SLR1_7627 OR SLR1_8245 OR SLR1_8094

✕

🔍

Share

Save

Saved Searches

Watched Isolates

Filters

Matched Clusters

#	Organism groups	SNP cluster	Matched isolates	Matched clinical isolates	Matched environmental isolates	Total isolates	Minimal min-diff	Minimal min-same	Latest update	count:1
1	Salmonella enterica	PDS000093458.2	1	1	0	2	n/a	0	2021-08-03	

Matched Isolates

Page 1 of 1

Records per Page 20

Choose columns

Download

Hide plus AMR genotypes

Expand all

Cross-browser selection

Displaying 1 - 4 of 4

#	Organism group	Strain	Isolate identifiers	Isolate	Create date	Location	Isolation type	SNP cluster	Min-same	Min-diff	BioSample	Assembly	AMR genotypes	Computed types
1	Salmonella enterica	SLR1_7627	SLR1_7627 SR57105757	PDT000826652.1	2020-08-30	Russia	clinical		n/a	n/a	SAHNI15666300	GCA_014334235.1	Complete (11) aadA1 aph(3')-Ib aph(6)-Id Show all 11 genes	Serotype: Bovismorbificans Antigen formula: 8:r:1,5
2	Salmonella enterica	SLR1_8094	SLR1_8094 SR57105758	PDT000826651.1	2020-08-30	Russia	clinical		n/a	n/a	SAHNI15666301	GCA_014334215.1	Complete (14) aac(3)-Ilg aac(6)-IIfc aph(3')-Ib Partial (1) ere(A) Show all 15 genes	Serotype: Enteritidis Antigen formula: 9:g,m:-
3	Salmonella enterica	SLR1_8245	SLR1_8245 SR59556581	PDT001100276.2	2021-07-31	Russia	clinical	PDS000093458.2	0	n/a	SAHNI20345806	GCA_019428465.1	Complete (5) catA1 mcr-1.1 mdsA Show all 5 genes	Serotype: Enteritidis Antigen formula: 9:g,m:-
4	Salmonella enterica	SLR1_8250	SLR1_8250 SR57105755	PDT000826653.1	2020-08-30	Russia	clinical		n/a	n/a	SAHNI15666299	GCA_014334255.1	Complete (3) mcr-1.1 mdsA mdsB	Serotype: Enteritidis Antigen formula: 9:g,m:-

Figure S1. Results of a similarity check (threshold ≤ 50 SNPs) of our analysed isolates with isolates contained in the NCBI Pathogen Detection system using URL https://www.ncbi.nlm.nih.gov/pathogens/isolates/#SLR1_8250%20OR%20SLR1_7627%20OR%20SLR1_8245%20OR%20SLR1_8094

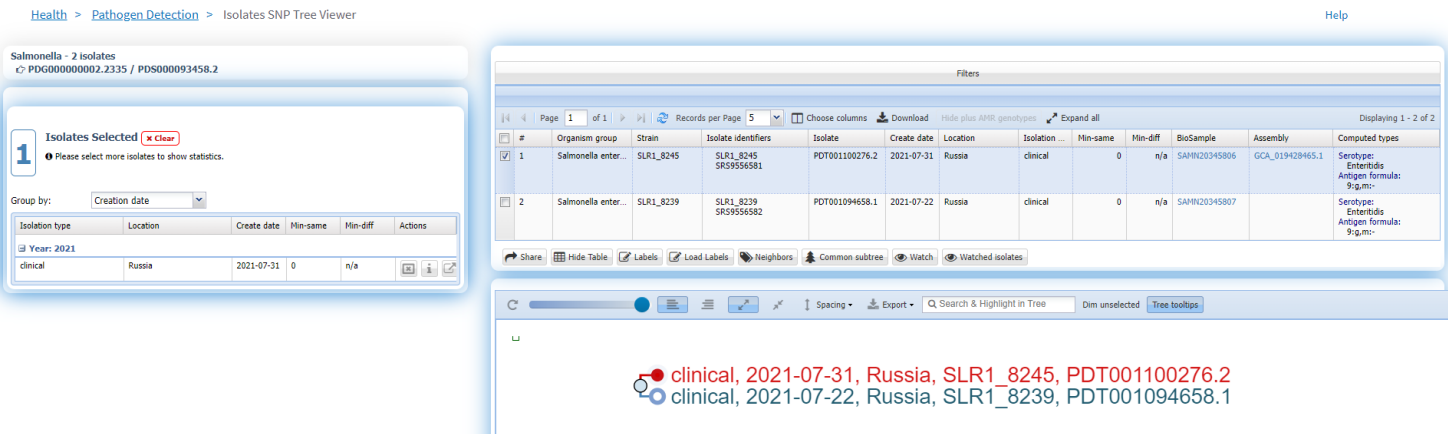


Figure S2. Screenshot of SNP cluster PDS000093458.2. Isolates *S. Enteritidis* SLR1_8245 and SLR1_8239 were isolated from the same outbreak and have 0 SNP differences.

Definition of uniqueness of isolates using EnteroBase system.

For each of the isolates EnteroBase system automatically calculated cluster numbers based on algorithms of hierarchical clustering where HC0 (indistinguishable), HC2 (two loci differences), HC5 (five loci differences), HC10 (differ in 10 loci), and HC20 (differ in 20 loci) etc. (Figure S3). For isolates SLR1_7627 (HC0_238586), SLR1_8250 (HC0_238575), SLR1_8245 (HC0_278706) HC0 cluster number was the same at levels HC2, HC5, HC10, and HC20, and for isolate SLR1_8094 HC0_236467 was the same at levels HC2, HC5, HC10.

EnteroBase														
Salmonella														
Log In Register Help v1.1.2														
Species Home														
Tasks														
Upload Reads														
Search Strains														
Find ST(s)														
Workspace:None Rows Total:4 Filtered:4														
Experimental Data cgMLST V2 + HierCC V1														
Uberstrain	ST	HC0 (indist...	HC2	HC5	HC10	HC20	HC50	HC100	HC200	HC400				
SLR1_8094	238579	236467	236467	236467	236467	2	2	2	2	2				
SLR1_7627	238586	238586	238586	238586	238586	499	499	499	499	499				
SLR1_8250	238575	238575	238575	238575	238575	87	87	12	12	12				
SLR1_8245	278706	278706	278706	278706	278706	87	87	12	12	12				

Figure S3. Studied isolates and their sequence types (ST) and cluster numbers (HC0, HC2, HC5, HC10, HC20 etc.) calculated based on algorithms of hierarchical clustering of cgMLST data in the EnteroBase system.

To show that the database did not contain any isolate that was identical to our isolates (at least with a difference of 10 loci) we performed separate searching of values of HC10 cluster numbers of each studied isolate in EnteroBase. The example of the query

Search all Strains of Salmonella

Predefined Search

Ignore Legacy Data Only Editable Strains Show Failed Assemblies Show Sub Strains

Strain Metadata AND OR Experimental Data

Experiment Type cgMLST V2 + HierCC V1

Data Type HC10

Operator contains

Value 236467

Clear

Cancel Submit

For each HC10 value we received only one isolate (HC10_238586 - SLR1_7627, HC10_236467 - SLR1_8094, and HC10_238575 - SLR1_8250) (Figure S4).

Uberstrain	ST	HC0 (indist...	HC2	HC5	HC10	HC20	HC50	HC100	HC200	HC400				
SLR1_7627	238586	238586	238586	238586	238586	238586	499	499	499	499				
SLR1_8094	238579	236467	236467	236467	236467	2	2	2	2	2				
SLR1_8250	238575	238575	238575	238575	238575	238575	87	87	12	12				
SLR1_8239	278706	278706	278706	278706	278706	278706	87	87	12	12				
SLR1_8245	278706	278706	278706	278706	278706	278706	87	87	12	12				

Figure S4: Results of individual queries using different HC10 values (HC10 = 238586, HC10 = 236467, and HC10 = 238575).

As expected (see supplementary figure S2), for HC10 278706 we received two outbreak isolates SLR1_8245 and SLR1_8239.

We concluded that we did not reveal any close relationship of our isolates to each other or to other isolates (with the exception of SLR1_8245 and SLR1_8239) by core genome MLST in the EnteroBase as our isolates showed at least 10 allele differences by cgMLST.