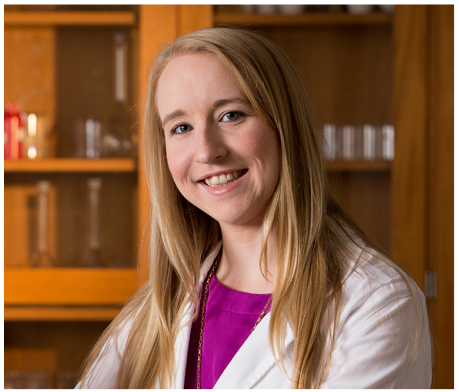
REP Codes, NCBI Pathogen Detection, and SEDRIC

Jack Marr, MPH

June 24. 2025

Introductions

Lauren Hudson, PhD



- Research Assistant Professor
- Dept. of Food Science
- University of Tennessee
- lkhudson@utk.edu

Jack Marr, MPH

- Epidemiologist
- Foodborne and Enteric Diseases
- Tennessee Department of Health
- jack.marr@tn.gov

Agenda

- REP Codes
  - Understanding REP codes
  - Local REP codes, investigation, and monitoring

BREAK

- NCBI Pathogen Detection
  - Cluster detection
  - Notification set-up
- SEDRIC
  - Cluster detection
  - Historical outbreaks and serotypes/species
- Next session, Discussion

Learning Objectives

REP codes	Explain what REP codes are
	Understand benefits of tracking and investigating REP codes
	Identify, investigate document local REP codes
NCBI	Utilize the NCBI Pathogen Detection Portal for cluster investigation
	Set up email notifications for new isolates in cluster
SEDRIC	Utilize SEDRIC for cluster detection and investigation
	Utilize SEDRIC to obtain historical context for rare serotypes
Overall	Choose the appropriate cluster detection/identification tool (e.g., NCBI Pathogen Detection, SEDRIC) to use in different contexts

Reoccurring, Emerging, and Persisting Enteric Bacterial Strains

What are REP Codes?

- "Group of bacteria, closely related by whole genome sequencing, that continues to cause illness over time"
  - Intentionally vague
- Since 2019, CDC has used the PulseNet system to identify and monitor REP strains of several major pathogens, including:
  - *Salmonella enterica*
  - *E. coli* O157:H7
  - *Listeria monocytogenes*
  - *Campylobacter jejuni*

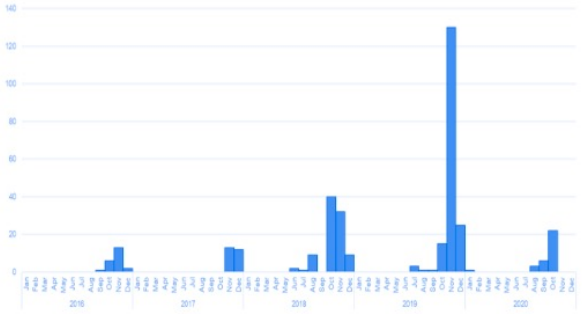
What are REP Codes?

- Typical PH response – acute, investigation tied to single source, low genetic diversity to be specific
- ✨Not! Like! Other! Bacteria!💅
  - **Reoccurring**: Repeatedly causing acute outbreaks separated by periods with no/low illness
  - **Emerging**: Previously novel or rare strains
    - (Show potential to) increase in cases
  - **Persisting**: Causing illnesses consistently over a long period of time
- Category may not be straight forward or change over time

https://www.cdc.gov/foodborne-outbreaks/php/rep-surveillance/index.html

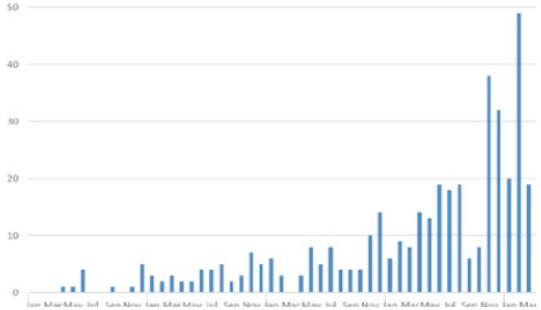
REP Strains

Reoccurring Strain



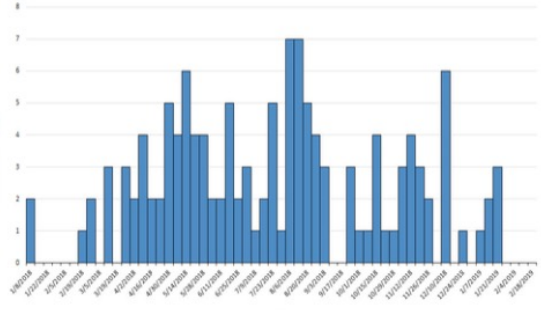
Strain that periodically causes a substantial number of illnesses, typically in outbreaks, separated by periods when it is not isolated from people or it causes very few illnesses

Emerging Strain



Strain that causes illnesses that have increased in frequency, or have the potential to increase in frequency, over time

Persisting Strain

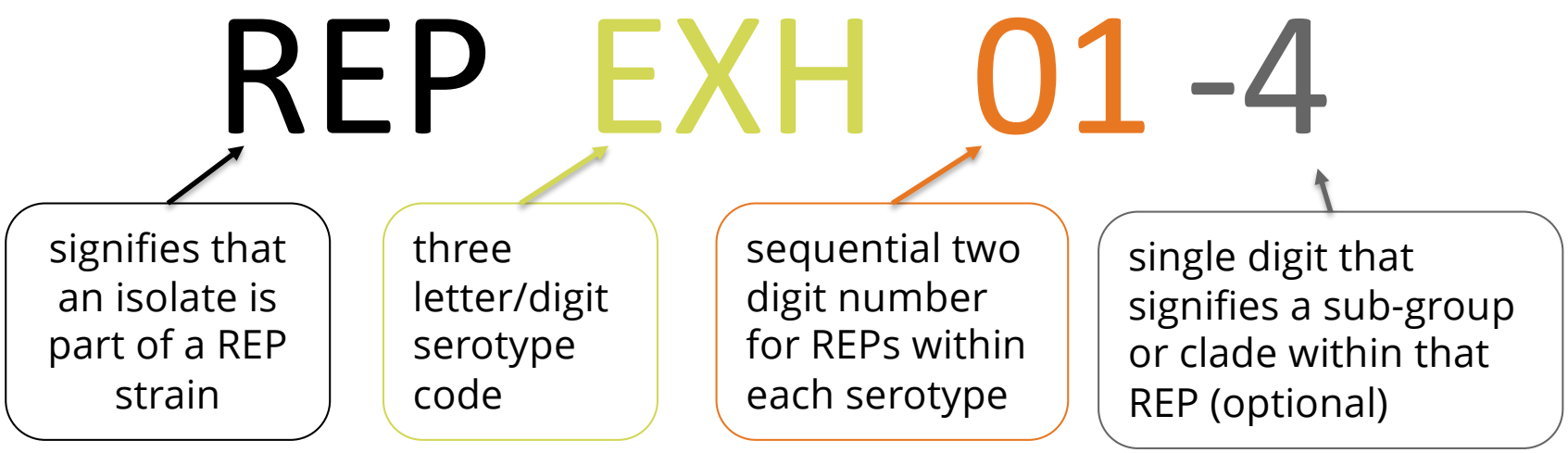


Strain that causes illnesses consistently over time, although the frequency of illnesses may fluctuate

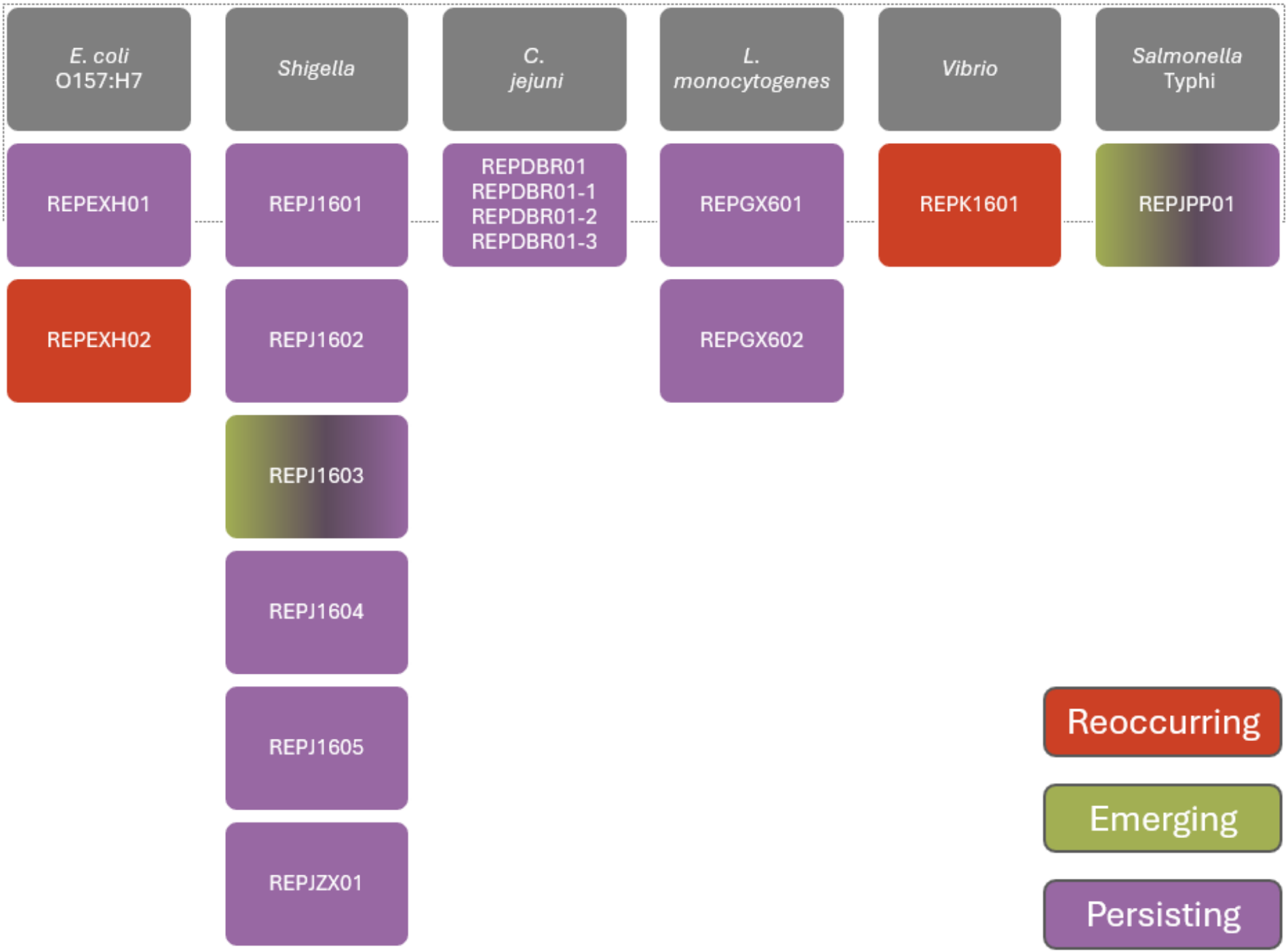
Image: Roundtable Discussion of Reoccurring, Emerging, and Persisting Strains (REPs), InFORM Conference, 2024

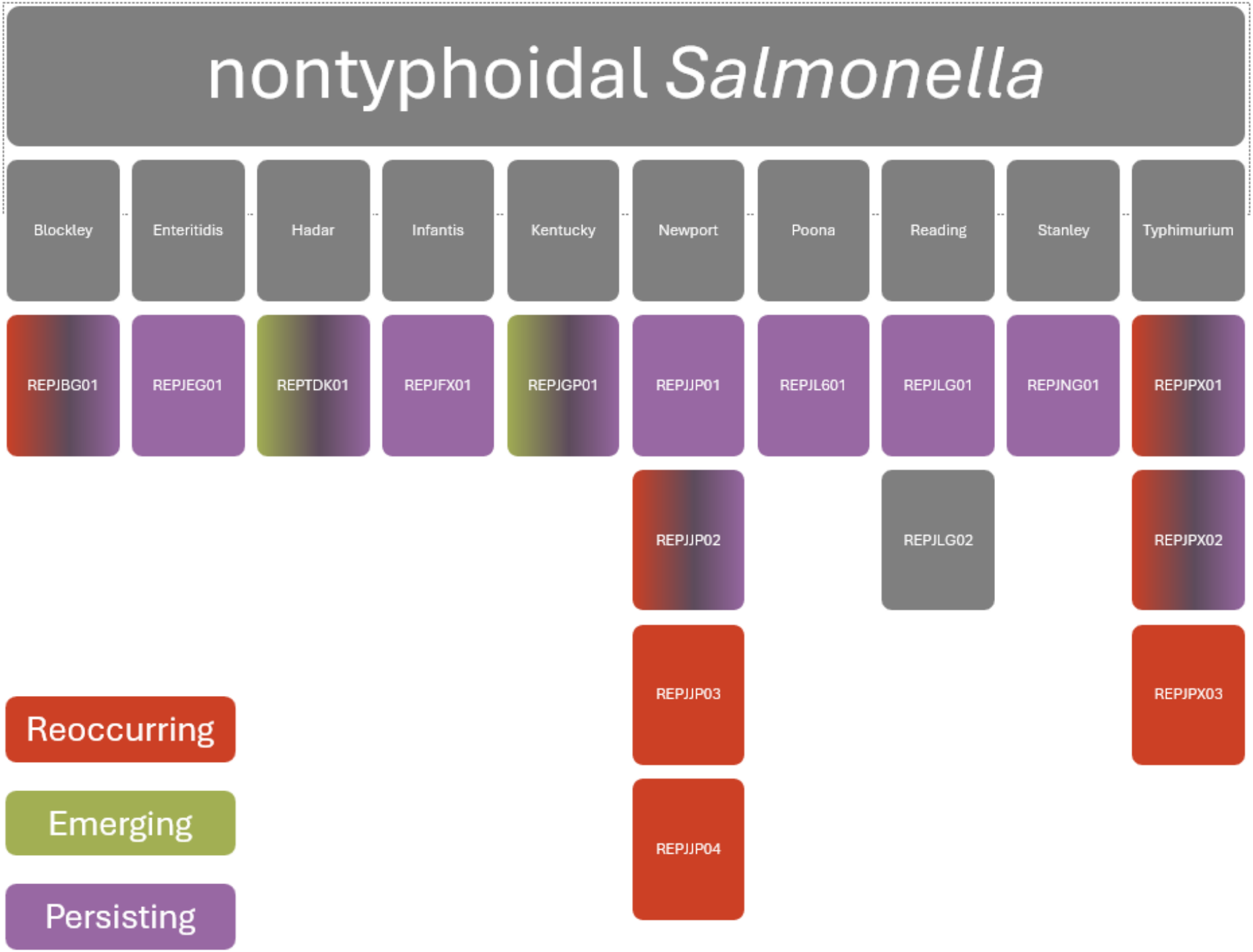
REP Strains

- Identifying a strain as a REP strain is based on many factors, including:
  - the number of illnesses, if increasing
  - the characteristics of the strain (e.g., multidrug resistance, high virulence, increased transmissibility)
  - if the strain is the cause of large or frequent outbreaks



https://www.cdc.gov/foodborne-outbreaks/php/rep-surveillance/index.html


E. coli O157:H7
Shigella
C. jejuni
L. monocytogenes
Vibrio
Salmonella Typhi
REPEXH01
REPEXH02
REPJ1601
REPJ1602
REPJ1603
REPJ1604
REPJ1605
REPJZX01
REPDBR01
REPDBR01-1
REPDBR01-2
REPDBR01-3
REPGX601
REPGX602
REPK1601
REPJPP01
Reoccurring
Emerging
Persisting


nontyphoidal Salmonella
Blockley
Enteritidis
Hadar
Infantis
Kentucky
Newport
Poona
Reading
Stanley
Typhimurium
REPJBG01
REPJEG01
REPTDK01
REPJFX01
REPJGP01
REPJJP01
REPJL601
REPJLG01
REPJNG01
REPJPX01
REPJJP02
REPJLG02
REPJPX02
REPJJP03
REPJPX03
REPJJP04
Reoccurring
Emerging
Persisting

Emerging Illness – REPJPP01

HIGHLIGHTS

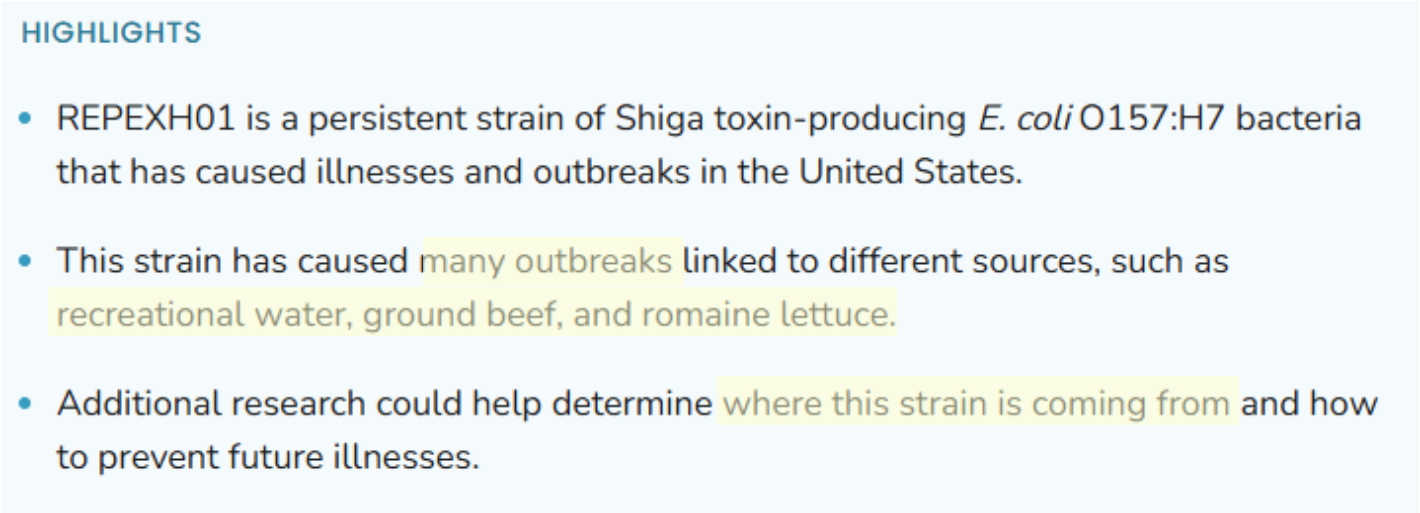
- REPJPP01 is an emerging strain of extensively drug-resistant *Salmonella* Typhi, the bacteria that cause typhoid fever. This strain has caused illnesses in the United States since 2018.
- REPJPP01 has caused illness in travelers to Pakistan and neighboring countries and in people who did not leave the United States.
- Getting vaccinated against typhoid fever before international travel – especially to South Asia – and following safe eating and drinking habits while traveling can prevent illness.

What makes this "emerging"?

- ID'ed in Pakistan, now found in more countries including US not related to travel
- Antibiotic resistance – ceftriaxone initially, now some XDR

Did you know: bacteria in REPJPP01 are within 13 alleles?

Persistent Illness – REPJPP01



HIGHLIGHTS

- REPEXH01 is a persistent strain of Shiga toxin-producing *E. coli* O157:H7 bacteria that has caused illnesses and outbreaks in the United States.
- This strain has caused many outbreaks linked to different sources, such as recreational water, ground beef, and romaine lettuce.
- Additional research could help determine where this strain is coming from and how to prevent future illnesses.

What makes this "persistent"?

- Multiple outbreaks
- Variety of sources
- Unclear where strain is coming from

Did you know: REPEXH01 includes 634 PN cases and 14 outbreaks!

Knowledge Check

REP stands for:

A. Repeating, Emerging, and Persisting

B. Repeating, Enteric, and Persisting

C. Reoccurring, Emerging, and Persisting

D. Reoccurring, Emerging, Problematic

Knowledge Check

REP codes include:

A. Three letter or digit serotype code
B. R, E, or P for type of REP code
C. Two-digit number for strains within each serotype
D. A and B
E. A and C

Knowledge Check

True or False: REP strains can belong to more than one classification (i.e. reoccurring, emerging, persisting)

True!

TN REP Strain - Shiga toxin-producing *E. coli* Cluster with unknown link, 2023 to 2024*

Local REP Codes

- Currently, CDC monitors and tracks all REP codes in US through SEDRIC
- States can use the core principles of REP codes for themselves!
  - Reoccurring STEC cluster in TN

SALM1.0 - 6788.1.1.4x | SALM1.1-
58.68.73.1993 | SALM1.1-
58.68.73.1993.2457 | SALM1.1-
58.68.73.1993.2457.109548 | SALM1.1-
58.68.73.1993.2457.114195 | SALM1.1-
58.68.73.1993.2457.14658 | SALM1.1-
58.68.73.1993.2457.164384 | SALM1.1-
58.68.73.1993.2457.17700 | SALM1.1-
58.68.73.1993.2457.18082 | SALM1.1-
58.68.73.1993.2457.18439 | SALM1.1-
58.68.73.1993.2457.25526 | SALM1.1-
58.68.73.1993.2457.3078 | SALM1.1-
58.68.73.1993.2457.38755 | SALM1.1-
58.68.73.1993.2457.42783 | SALM1.1-
58.68.73.1993.2457.49319 | SALM1.1-
58.68.73.1993.2457.56160 | SALM1.1-
58.68.73.1993.2457.5954 | SALM1.1-
58.68.73.1993.2457.8056 | SALM1.1-

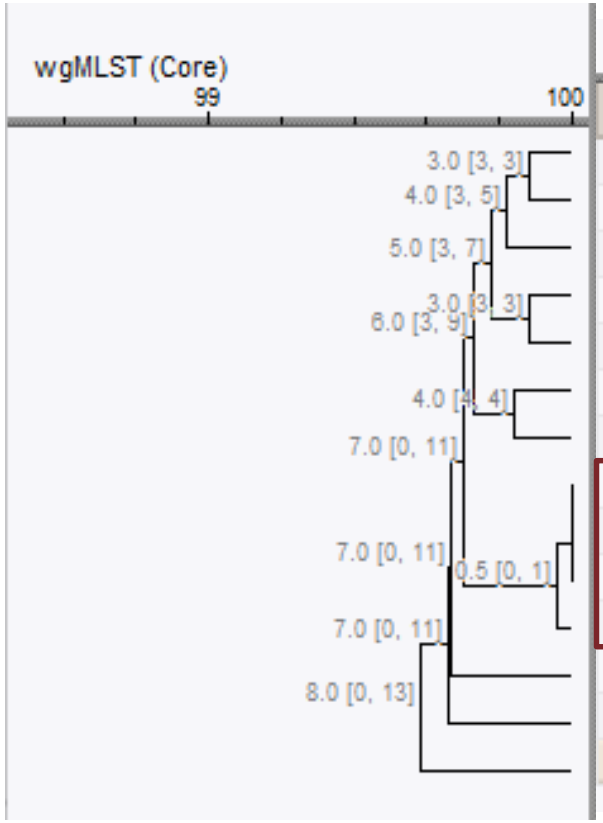
58.68.73.1993.2457.8114 | SALM1.1-
58.68.73.1993.2457.8376 | SALM1.1-
58.68.73.1993.2457.89476 | SALM1.1-
58.68.73.1993.2457.89534 | SALM1.1-
58.68.73.1993.2457.89929 | SALM1.1-
58.68.73.1993.2457.90628 | SALM1.1-
58.68.73.1993.2457x | SALM1.1-
58.68.73.1993.2457x.14658 | SALM1.1-
58.68.73.1993.2457x.164384 | SALM1.1-
58.68.73.1993.2457x.18082 | SALM1.1-
58.68.73.1993.2457x.213038 | SALM1.1-
58.68.73.1993.2457x.3078 | SALM1.1-
58.68.73.1993.2457x.47093 | SALM1.1-
58.68.73.1993.2457x.5954 | SALM1.1-
58.68.73.1993.2457x.8376 | SALM1.1-
58.68.73.1993.33058

REPJPP01 allele codes

Notification and Review

- TDH was notified in May 2023, of a hospitalized STEC O157 case who developed hemolytic uremic syndrome (HUS) and subsequently passed away
  - Classified as a public health incident
- Review of cases in BN
  - Had the same or similar allele code to the 2023 case, but **not highly related (within 13 alleles)**
    - Multiple commonalities, particularly age
- In the late summer of 2023, received more cases with the same allele code including an outbreak at a daycare

Case Investigations



Pat...	Serot...	Toxin_wgs	IsolatDate	Receiv...	So...	Sour...	Allele_code
STEC	O157:H7	stx2a; stx2c	2020-09-04	2020-10-06	Stool	Human	EC1.0 - 9.1.3.218
STEC	:H7	stx2a; stx2c	2022-08-18	2022-08-19	Stool	Human	EC1.0 - 9.1.3.218.40
STEC	:H7	stx2a; stx2c	2022-07-07	2022-07-13	Stool	Human	EC1.0 - 9.1.3.218
STEC	O157:H7	stx2a; stx2c	2021-09-17	2021-09-24	Stool	Human	EC1.0 - 9.1.3.218
STEC	O157:H7	stx2a	2022-07-13	2022-07-18	Stool	Human	EC1.0 - 9.1.3.218
STEC	:H7	stx2a; stx2c	2023-07-02	2023-07-06	Stool	Human	EC1.0 - 9.1.3.218
STEC	O157:H7	stx2a	2023-05-03	2023-05-04	Stool	Human	EC1.0 - 9.1.3.218
STEC	O157:H7	stx2a; stx2c	2023-09-03	2023-09-05	Stool	Human	EC1.0 - 9.1.3.218.55
STEC	O157:H7	stx2a; stx2c	2023-09-02	2023-09-07	Stool	Human	EC1.0 - 9.1.3.218.55
STEC	O157:H7	stx2a; stx2c	2023-08-31	2023-09-05	Stool	Human	EC1.0 - 9.1.3.218.55
STEC	O157:H7	stx2a; stx2c	2023-07-16	2023-07-18	Stool	Human	EC1.0 - 9.1.3.218
STEC	O157:H7	stx2a; stx2c	2024-01-25	2024-01-26	Stool	Human	EC1.0 - 9.1.3.218
STEC	:H7	stx2a	2022-08-03	2022-08-12	Stool	Human	EC1.0 - 9.1.3.218
STEC	O157:H7	stx2a; stx2c	2022-08-23	2022-08-24	Stool	Human	EC1.0 - 9.1.3

**Allele codes previous version*

Case Investigations

EC1.0 - 9.1.3.218 allele code* with 15 TN cases

- Four cases belong to a daycare-related outbreak

17/18 are in children

Middle and West Tennessee

12/17 were hospitalized, 9 had HUS

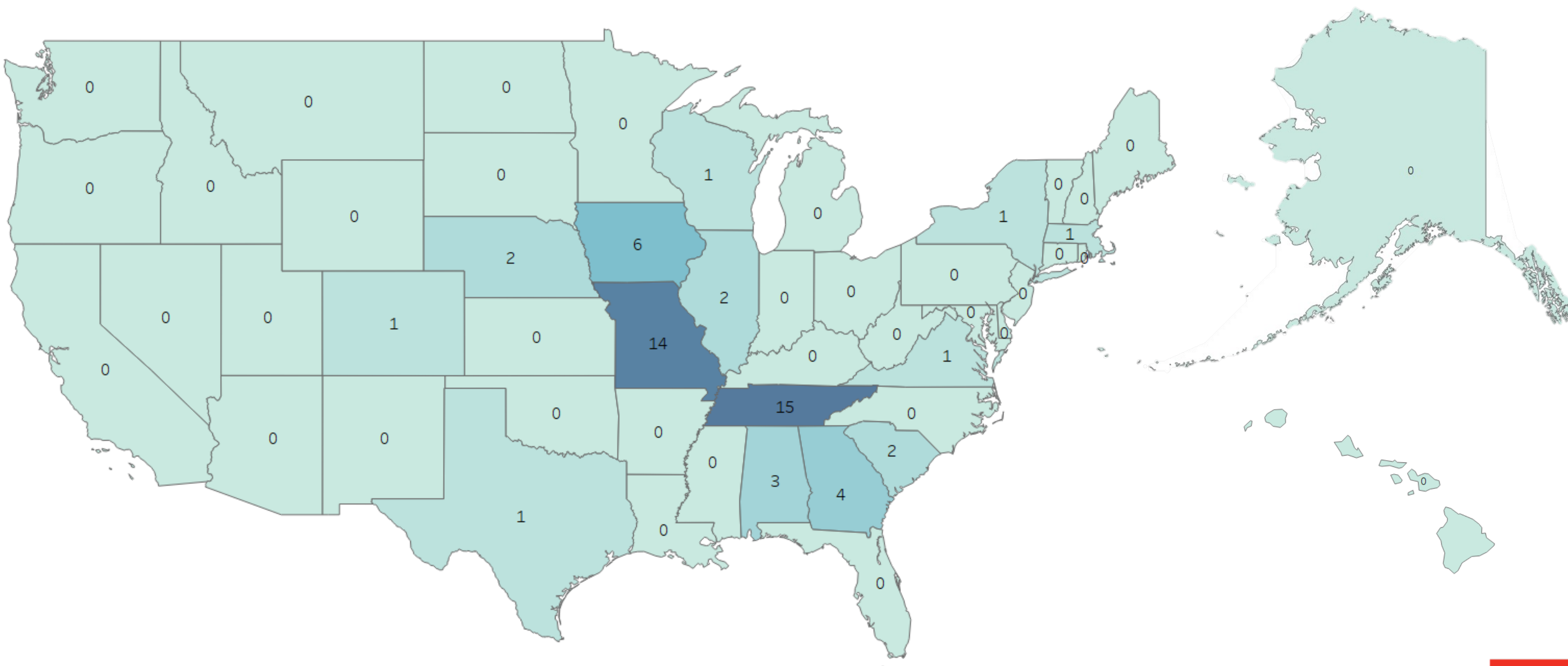
Two deaths

No signal or suspect source

**Allele codes previous version*

Out of State Cases

- As of 05/02/2024 there are 38 cases in 13 states
  - MO 14, IA 6, GA 4, AL 3, Il 2, NE 2, SC 2, CO 1, MA 1, NY 1, TX 1, VA 1, WI 1



Conclusions – STEC cases

- Local REP code?
  - ✅ Reoccurs yearly between June and September
  - ✅ Geographically centered on West/Mid TN, nearby states
  - ✅ Severe illness among consistent group

- Atypical investigation
  - Primarily data review mixed with outbreak investigation
  - Large allelic difference vs. more specific parameters

Monitoring & Investigating REP Strains

Monitoring

- Gain a deeper understanding of the strain
- Classification into genetic and epidemiological subclusters
- Collaborating with industry partners
- Line lists updated less frequently

Active follow-up

- If there is an increase in cases or new cases
- Collection of early epidemiological and traceback data (as needed)
- Informing partners of potential subclusters/ outbreaks
- Communicate with other states
- Line lists updated more frequently

Outbreak investigation

- Similar to traditional outbreak response
- Quickly gathering information, re-interviews
- Collaborating with regulatory agencies
- Line lists updated regularly

Why Investigate REP Strains?

- Track how REP strains change over time
- Collaborate on measures to reduce spread
- Better understand transmission routes and sources
  - Prior investigation of settings, other factors
- Streamline future outbreak investigations involving the same strain
  - Guiding hypothesis generation
  - Create focused or supplemental questionnaires and/or to prioritize questions about related exposures during interviews

Documentation and tracking

- Multistate REP codes tracked via internal REDCap
  - With all other local/MS clusters
- TN Local REP codes are not tracked
  - For future local REP codes – will likely implement similar system
  - Do keep records and files in usual secure drive
  - No naming convention yet
- Other state testimonials

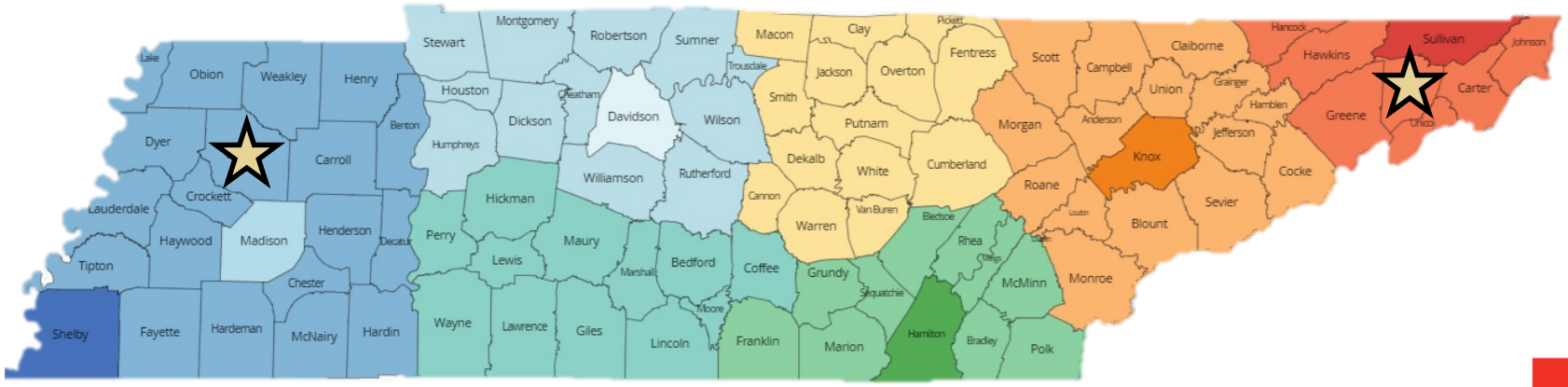
Documentation and tracking

Investigation Status:	Closed Change to CLOSED when investigation is complete.
Date Investigation Closed:	10-11-2023 Today M-D-Y
Reason Investigation Closed:	REP strain investigation complete
Date First Identified/CDC Notification:	09-22-2023 Today M-D-Y
Source of Notification:	CDC
Cluster Jurisdiction:	Multistate - CDC REP Strain
Organism:	Listeria
Species/Serotype:	monocytogenes
Serotype Notes:	possibly linked to queso fresco-type soft cheeses Expand
Allele Code:	LMO1.1 - 1.72.1.2
REP Code:	REPGX602

NCBI Pathogen Detection Tool with *Salmonella* Stanley Example

Cluster Identification

- *Salmonella* Stanley cluster, 0 alleles apart
  - Case A: 4-year-old boy in Northeast Tennessee
  - Case B: 58-year-old woman in West Tennessee
- Common exposure to small turtles
- NCBI and SEDRIC used to find more cases
  - Pulled cases within 20 SNPs in last year
  - One WA case with turtle exposure

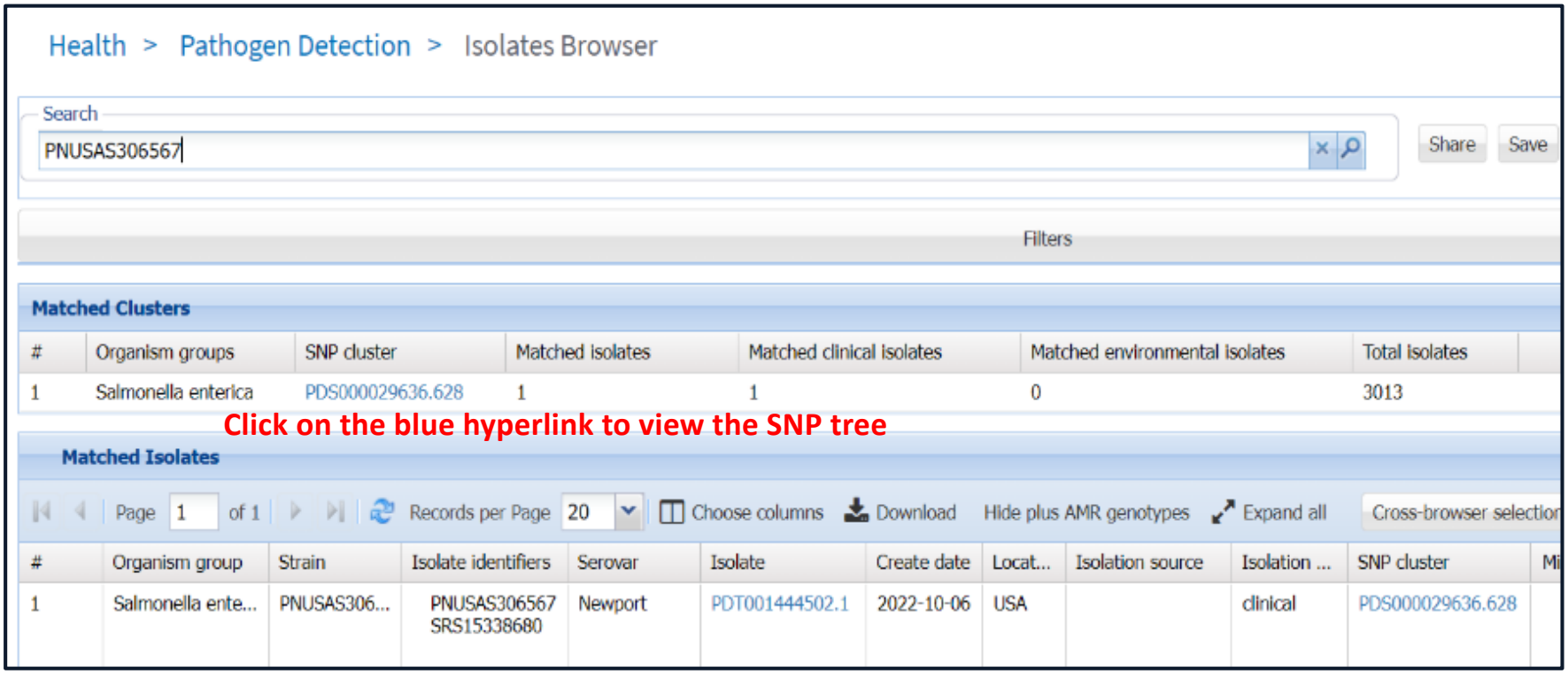


Using NCBI Pathogen Detection Isolates Browser

Search for cases in the National Center for Biotechnology Information (NCBI) Isolates Browser by using the **WGS ID** or **PNUSA #** into search bar to determine if there are any current or historical human, animal, or environmental on a national level are closely related to the cases in the cluster being investigated.

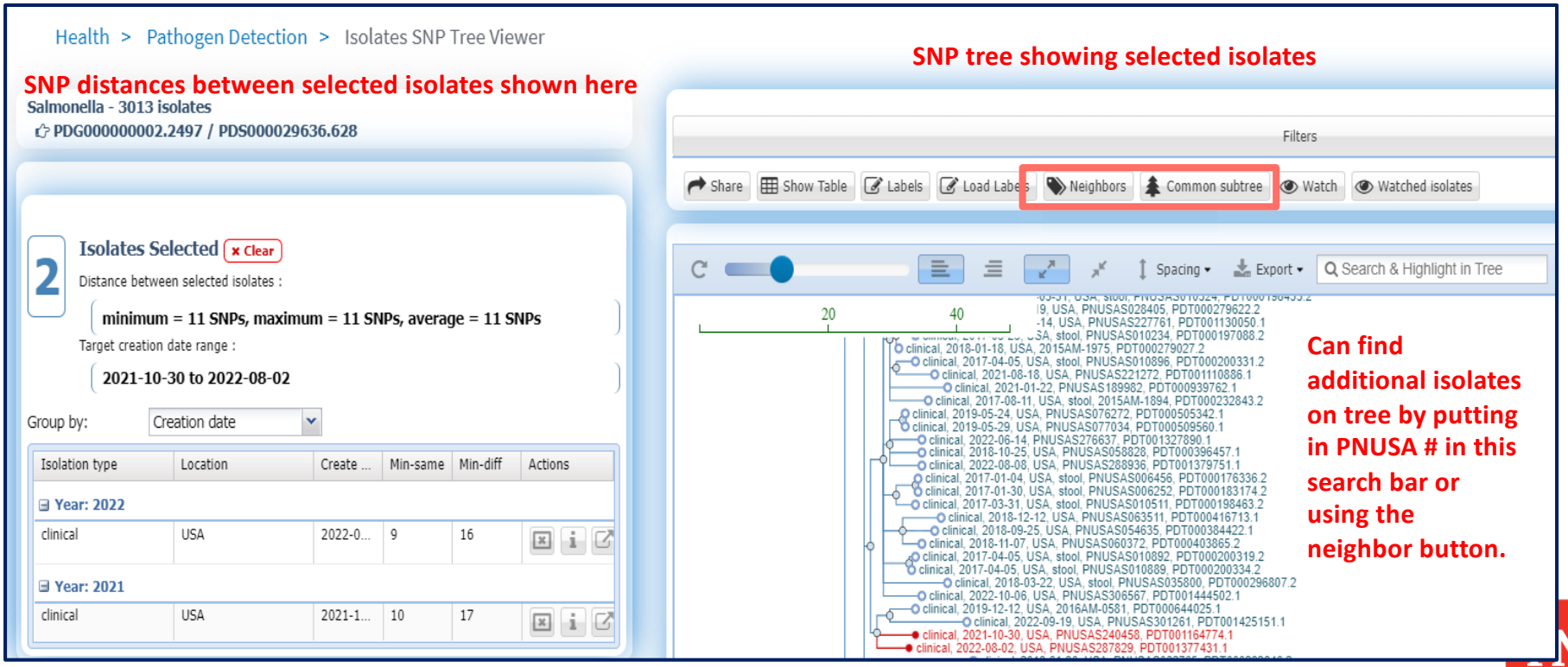
Lookup: PNUSAS266798, PNUSAS266800

Using NCBI Pathogen Detection Isolates Browser

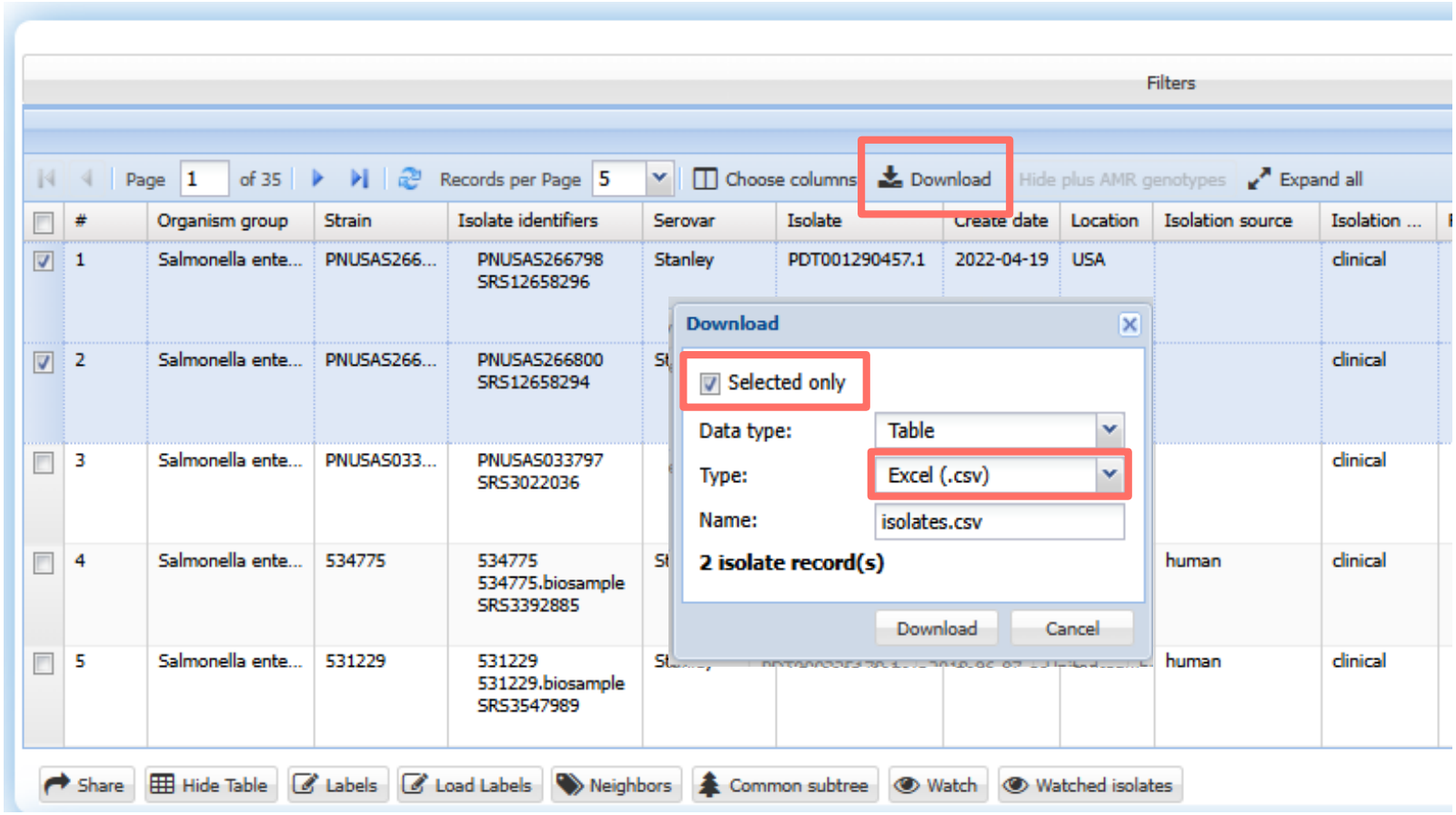


NCBI – Finding Neighbors

- Make a note of the PNUSA numbers of any isolates that are closely related in time and in SNP distance.
- The information in NCBI does not tell you what state an isolate (may say for environmental or animal isolates) is from.
- The neighbors button can automatically select closely related isolates.

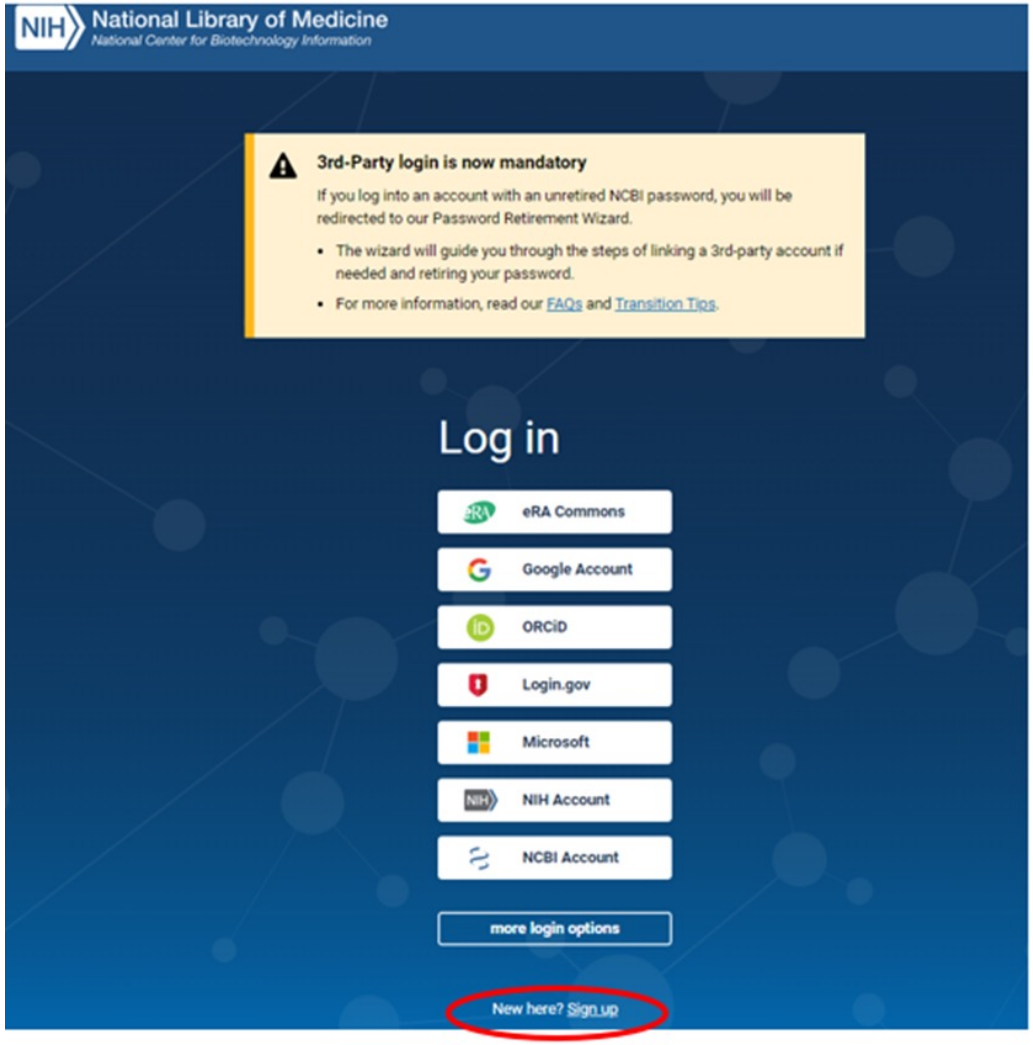


NCBI – Download Data



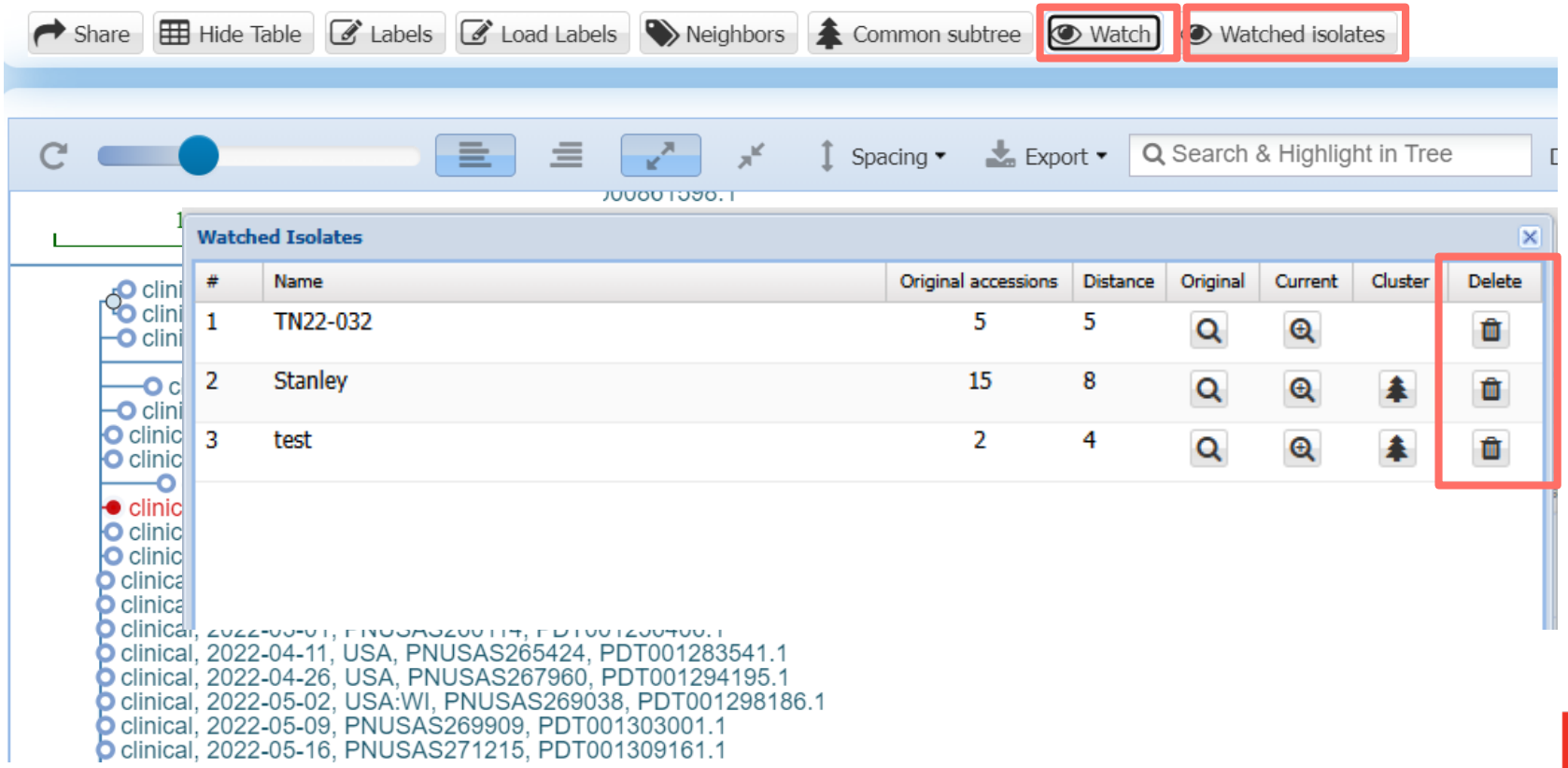
Using NCBI Pathogen Detection Isolates Browser

To get notified of new isolates added to a cluster in NCBI, please create an account first or sign in if you have an account.



Using NCBI Pathogen Detection Isolates Browser

- Above the tree there is an icon that looks like an eye and is labeled "**Watch**".
- Click it and then name the isolate/cluster a unique name so you're able to identify what cluster it is in reference to (ex. S. Liverpool cluster).
- Adjust the SNP distance to get notified of any isolates related to the one being watched within the threshold. Click "Save".



Knowledge Check

Isolates can be searched for in NCBI with:

A. PNUSA or WGS IDs

B. Lab accession number

C. Investigation ID

D. None of the above

Knowledge Check

True or False: Only one isolate can be searched for at a time in NCBI.

False!

Knowledge Check

What button does NOT appear in NCBI?

A. Neighbors
B. Watch
C. Watched Isolates
D. Show State Origin
E. Labels

SEDRIC Cluster Detection with *Salmonella* i4:i- Example

Background

- On October 15, 2024, the TN FED Program identified **2** genetically related cases of Salmonella I 4:i- residing in Shelby County
- By December 10, 2024, **9** cases were identified via NCBI:
  - 5 Memphis/Shelby Co. (MSR)
  - 1 Illinois
  - 1 Mississippi
  - 1 Ohio
  - 1 Texas

Laboratory Analysis

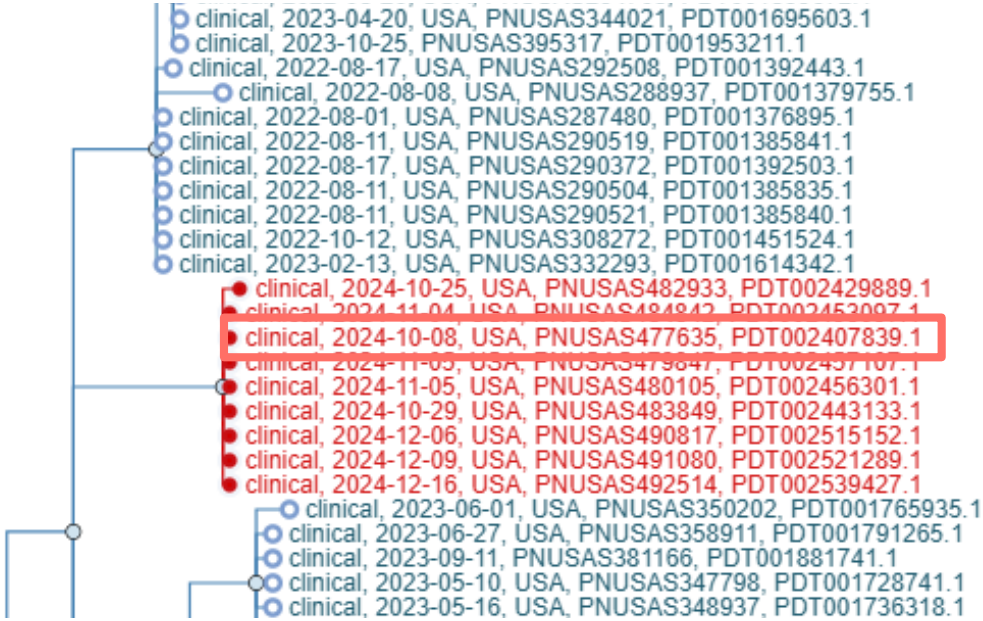
9 **Isolates Selected** × Clear

Distance between selected isolates :

minimum = 0 SNPs, maximum = 1 SNPs, average = 0 SNPs

Target creation date range :

2024-10-08 to 2024-12-16

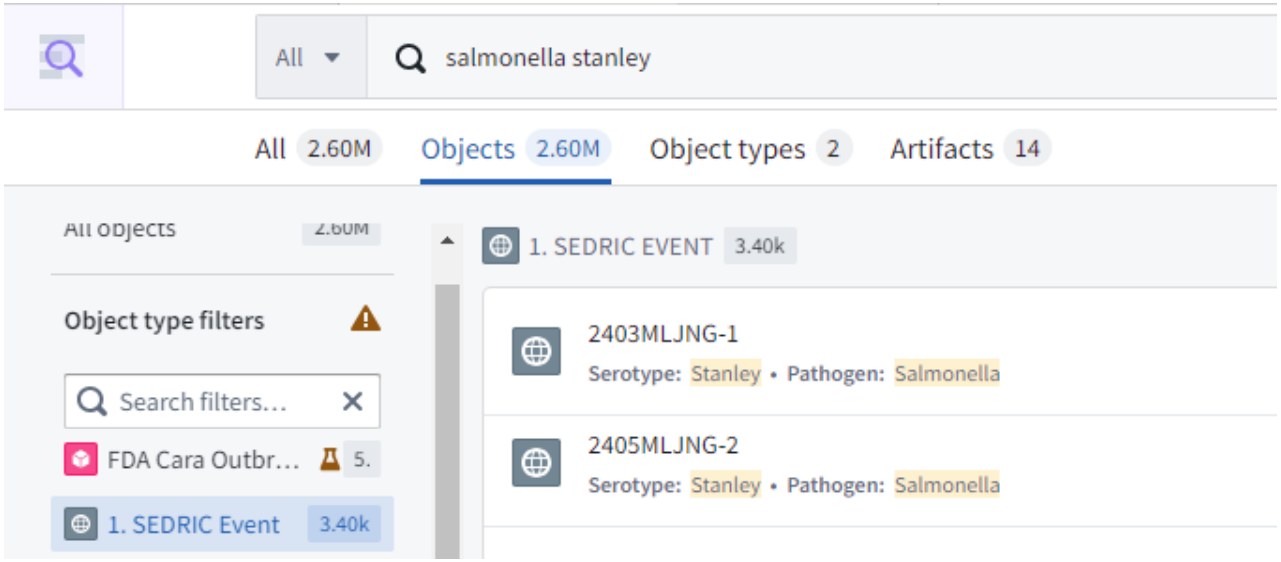


SEDRIC isolates

- Use PNUSA number
  - ID isolates, demos, basic information
  - Find state contact for additional information

SEDRIC

- How to find
  - Historical information, rare serotypes
    - Can search for a serotype or species with SEDRIC event filter to find historical outbreaks

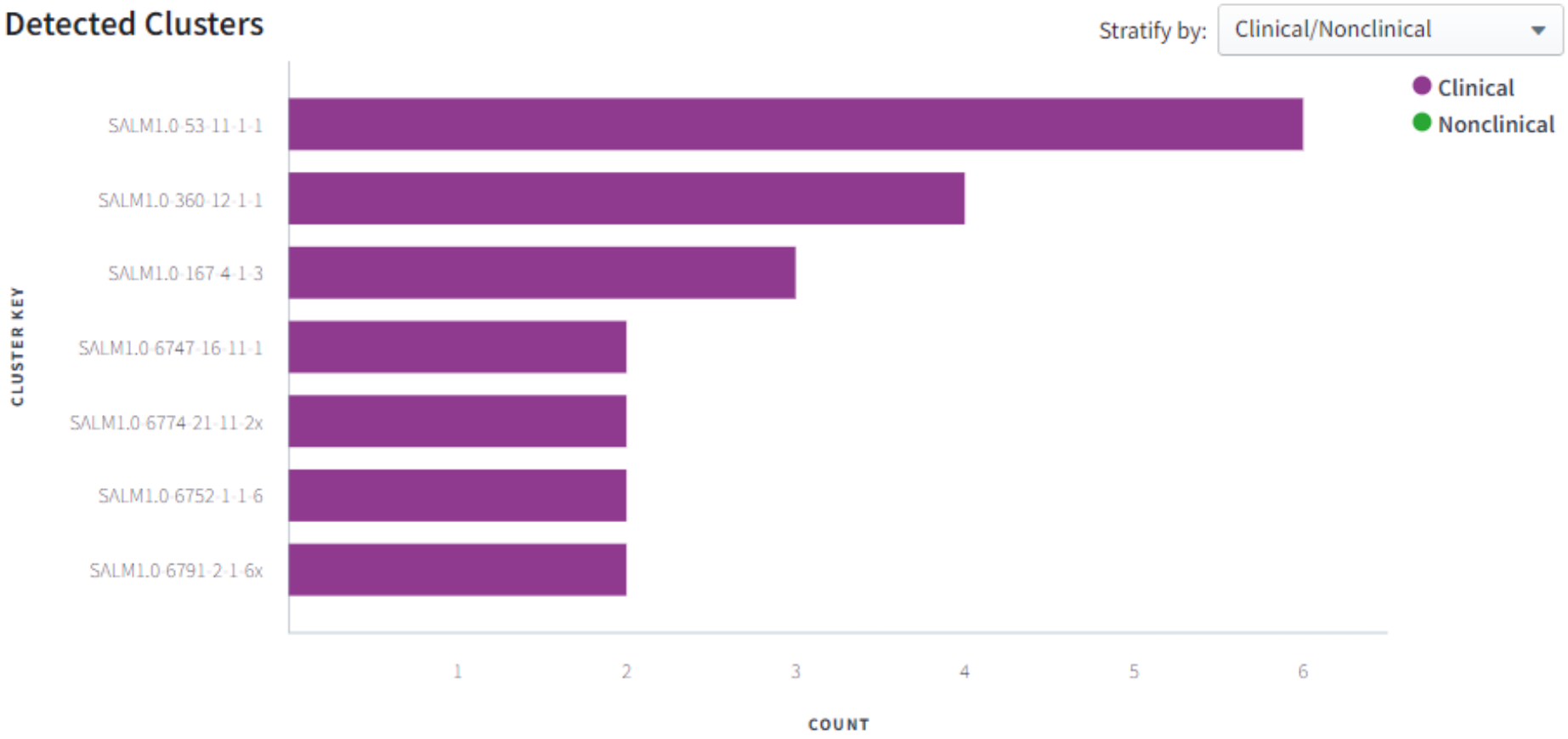


  - Other cases: use cluster detection tool!

SEDRIC Cluster Detection

- On main page, select SEDRIC, then Cluster Detection
- Settings
  - 3 month window with two isolates minimum
    - Min. of 4 allele digits for Salmonella, STEC
    - Min. of 5 allele digits for Salmonella for Enteriditis, Javiana, Newport, Typhimurium, I4i-, Campylobacter
    - X codes – Checking NCBI for these to identify sub-clusters, SNP threshold for x code clusters
- Threshold for investigation may increase/decrease based on season and capacity
  - Highly clonal serotypes/species = higher threshold

SEDRIC Cluster Detection



SEDRIC Cluster Detection

Selected Isolates

Key	Outbreak	Allele Code	Serotype	WGS ID	Isolation Date
TN[redacted]	2404MLAFR-1	SALM1.0 - 2871.1.1.19.1.1	Africana	PNUSAS432729	2024-04-09
TN[redacted]	2404MLAFR-1	SALM1.0 - 2871.1.1.19.1	Africana	PNUSAS436186	2024-04-22
TN[redacted]	2404MLAFR-1	SALM1.0 - 2871.1.1.19.1.1	Africana	PNUSAS437316	2024-04-27
TN[redacted]	2404MLAFR-1	SALM1.0 - 2871.1.1.19.1.1	Africana	PNUSAS437319	2024-04-03
TN[redacted]	2404MLAFR-1	SALM1.0 - 2871.1.1.19.1.1	Africana	PNUSAS437327	2024-04-26

Pros and cons

- Cons
  - Fewer clusters detected compared to PN 2.0
    - Typically highly clonal species/serotypes e.g. Enteriditis
    - Some shorter allele codes can be missed
  - Cannot detect *Shigella*, non-jejuni *Campylobacter* clusters
- Pros
  - Easier access than PulseNet!
  - Higher threshold, lower number of clusters to investigate
  - Can include additional states, demographic filters

NCBI and SEDRIC Investigation Summary

NCBI

- Search using PNUSA #s/WGS ID
- SNP-based dendrogram trees
- Used to find related national and environmental isolates (“neighbors”)
- Can “watch” isolate and get emails when new related isolates are added
- No state or demographic information, public access

SEDRIC

- Limited access
- Search using PNUSA numbers found in NCBI, serotype, or allele code
- No dendrogram trees; use cluster detection application with allele distance
- Great for finding related national, historical, and environmental isolates by allele code/serotype (if uncommon)
- Demographic, state and sometimes exposure information are available
- Find state or national contacts to find for additional data

Knowledge Check

What information can you find in SEDRIC?

A. Basic demographic data of isolates
B. Contact information
C. Isolates with a specific allele code
D. All of the above
E. None of the above

Knowledge Check

Connect the data system with units of genetic difference.

Final Session

- Tuesday July 1st 1-2PM EST
- Please complete survey!
- Topics
  - Data visualization
    - Tableau
    - SatScan
  - Communication
    - Communicating WGS results
    - Standard phrasing
  - PulseNet 2.0
    - How to get access

Questions?