

# Cluster Detection at CDC: The Life of a PulseNet Cluster



Lisa Theobald  
September 2009



# Overview

## ■ Cluster Detection

- Local and National
- Verification of Data
- Pattern Naming

## ■ Response

- Cluster Code Assignment
- Line Lists, Frequency Graphs, and CDC Team Responses

## ■ Follow-up

- Cluster Reports, Epi Requests, and Cluster Conclusion

## ■ Data Interpretation Issues

- Similar patterns
- Local vs. National Databases
- Interpreting Patterns within Outbreaks



# Cluster Detection

## ■ Local

- Perform cluster search within local database
- Compare to National Database
- Report local clusters of 3 or more human isolates (above normal pattern baseline frequency) on CDC Team
- Begin epi investigation

## ■ National

- Perform cluster search within National Database *or* respond to local CDC Team posting
- Report multi-state clusters that indicate an increase in a particular PFGE pattern
- Report regional increases above national baseline
- Report non-human matches to 3 or more human isolates

# Verification of Data

- **Serotype Confirmation**
  - PFGE clustering
  - “Unconfirmed Serotype”
- **Proper Band Marking**
  - SOP PND04: Gel Analysis Guidelines
- **Sufficient Demographic Information**
  - Age, sex, source state, etc.
- **Correct Data Format**
  - Dates: YYYY-MM-DD; source site; serotype; etc.

# Pattern Naming

- Step 1: Select patterns to be named
  - Database Managers first pull up patterns to check band markings and that the PFGE image is satisfactory
  
- Step 2: Compare to Unique Pattern List
  - The Unique Pattern List contains at least one example of each pattern in the National Database
  - Newly uploaded patterns are compared to the Unique Pattern List in order to assign pattern numbers
  
- Step 3: Name the patterns
  
- Step 4: Confirm pattern names
  - Putting “Confirmed” in PFGE-*enzyme*-Status field



# Pattern Naming

## Automated Pattern Naming:

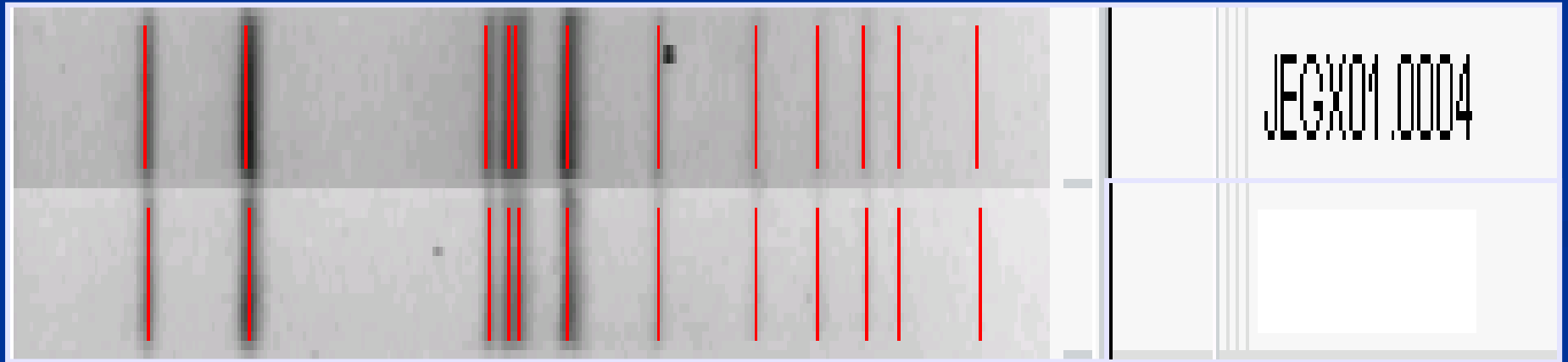
- Once a pattern has been submitted to the National Database, a pattern name will be assigned by the automated pattern naming system in BioNumerics (this is no longer used for *Salmonella*)
  - @ = new pattern assigned by automated system
  - & = existing pattern, but assigned by automated system
- DO NOT REPORT UNCONFIRMED PATTERN NUMBERS
- Pattern names can *only* be verified by CDC Database Managers

# Pattern Naming

- Patterns for all clusters and outbreaks and all *Campylobacter*, *Listeria*, *E. coli*, and the top 25 serotypes of *Salmonella* should be named within 2 working days
  - Exceptions: database manager(s) are out; database team is short-staffed; unusually high submissions and/or clusters; serotype is not filled in
- Patterns are also compared to the entire database
  - Used to verify serotype or name when no serotype is entered

# Pattern Naming

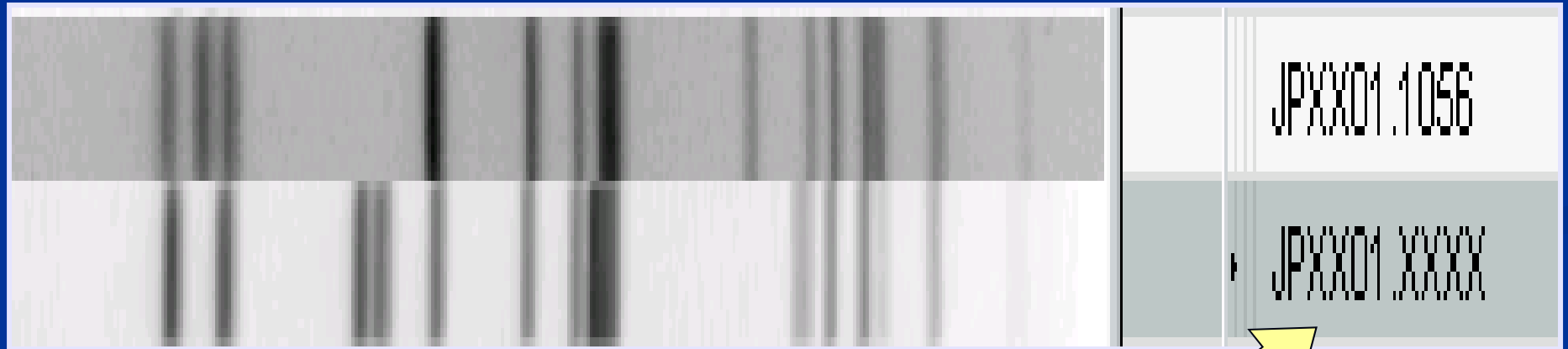
**Pattern seen before:** assigned the corresponding pattern name





# Pattern Naming

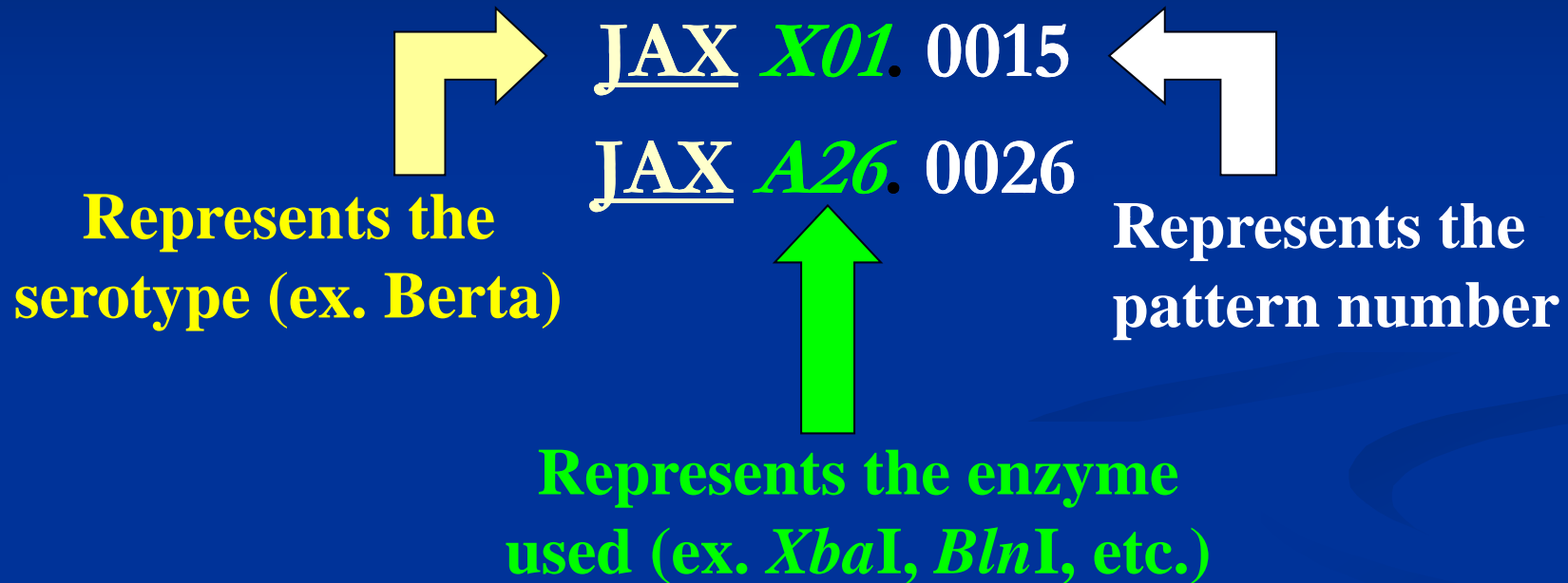
**New pattern:** given the next available name and goes into the Unique Pattern List as a representative



**The new isolate will be named according to the next available unused pattern number, and becomes a part of the Unique Pattern List**

# Pattern Naming

Every pattern name in the database has **3** components:



Pattern numbers do NOT represent the time the pattern was submitted, the frequency of the pattern in the database, nor genetic relatedness to other patterns

# Overview

- Cluster Detection
  - Local and National
  - Verification of Data
  - Pattern Naming
- Response
  - Cluster Code Assignment
  - Line Lists, Frequency Graphs, and CDC Team Responses
- Follow-up
  - Cluster Reports, Epi Requests, and Cluster Conclusion
- Data Interpretation Issues
  - Similar patterns
  - Local vs. National Databases
  - Interpreting Patterns within Outbreaks



# Cluster Code Assignment

- CLUSTERS of cases identified by PulseNet are investigated by epidemiologists
  - If epidemiologic links are found between cases, the cluster is classified as an OUTBREAK
- Each Cluster is given a “Cluster Code”

**0904MIJPX-1**

Year Month State Serotype Code Cluster #



# Cluster Code Assignment

## A code is given if:

- National or regional pattern frequency for human patterns in the past 60/120 days  $\geq 2x$  background frequency
- Anytime there is an epi link. Examples:
  - specific age/gender distribution for a cluster
  - epi link identified in posting (cruise ship, travel exposure, same institution, etc.)

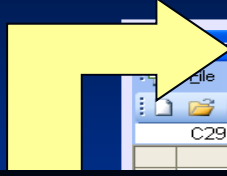
## A code is not given, but cluster reported to CDC epis

- 3+ human cases match a food, but not above baseline
  - Line list reported to epis





# Line Lists



**File Name  
=  
Cluster Code**

	A	B	C	D	E	F	G	H	I
	LabID	Serotype	PFGE-XbaI-pattern	Outbreak	SourceCountry	SourceState	SourceCounty	SourceCity	SourceCountry
07005601	CO	Typhimurium	JPXX01.1260	0707MLJJPX-1c	USA	CO			
07005607	SD	Typhimurium	JPXX01.1260	0707MLJJPX-1c	USA	SD			
07005636	WY	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	0707MLJJPX-1c	USA	WY	Albany	Laramie	
07005602	IN	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	0707MLJJPX-1c	USA	IN	Marion	Unknown	
07006196	CO	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	0707MLJJPX-1c	USA	CO			
07006200	IL	Typhimurium	JPXX01.1260	0707MLJJPX-1c	USA	IL		Rosemo	
07006200	NJ	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	0707MLJJPX-1c	USA	NJ	Middlesex	Perth Ar	
07005131	NY	Typhimurium	JPXX01.1260	0707MLJJPX-1c	USA	NY	Washington	GLENS	
070051000	ME	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	0707MLJJPX-1c	USA	ME	WASHINGTON	MACHIA	
07005175	IL	Typhimurium	JPXX01.1260	0707MLJJPX-1c	USA	IL	Cook	Chicago	
12	IN	07ENT0007	IN	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	USA	IN	Elkhart	Elkhart
13	IN	07ENT0019	IN	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	USA	IN	Marion	Unknown
14	MA	07PF0927	MA	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	USA	MA	PLYMOUTH	
15	WY	WY022564	WY	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	USA	WY	Natrona	Casper
16	OK	07OKE0645	OK	Typhimurium	JPXX01.1260	USA	OK	Tulsa	Tulsa
17	MN	E2007001496	MN	Serotype pending	JPXX01.1260	USA	NE		
18	NJ	701062	NJ	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	USA	NJ	Warren	Hacketts
19	OH	2007306191	OH	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	USA	OH	Montgomery	Dayton
20	CT	02029850	CT	Typhimurium	JPXX01.1260	USA	CT	Hartford	Rocky H
21	MA	07PF0977	MA	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	USA	MA	WORCESTER	
22	NC	NC2007-000885	NC	Typhimurium	JPXX01.1260	USA		Davie	Mocksvi
23	NY	BAC0700005294	NY	Typhimurium	JPXX01.1260	USA	NY	Monroe	ROCHE:
24	OH	2007306290	OH	Typhimurium var. O 5 - (Copenhagen)	JPXX01.1260	USA	OH	Warren	Springbc
25	NY	BAC0700005348	NY	Typhimurium	JPXX01.1260	USA	NY	Ontario	CANAN
26	NY	BAC0700005422	NY	Typhimurium	JPXX01.1260	USA	NY	Monroe	ROCHE:
27									
28									
29									
30									
31									
32									
33									
34									

**Cluster Code**

**Date Line List  
Created**

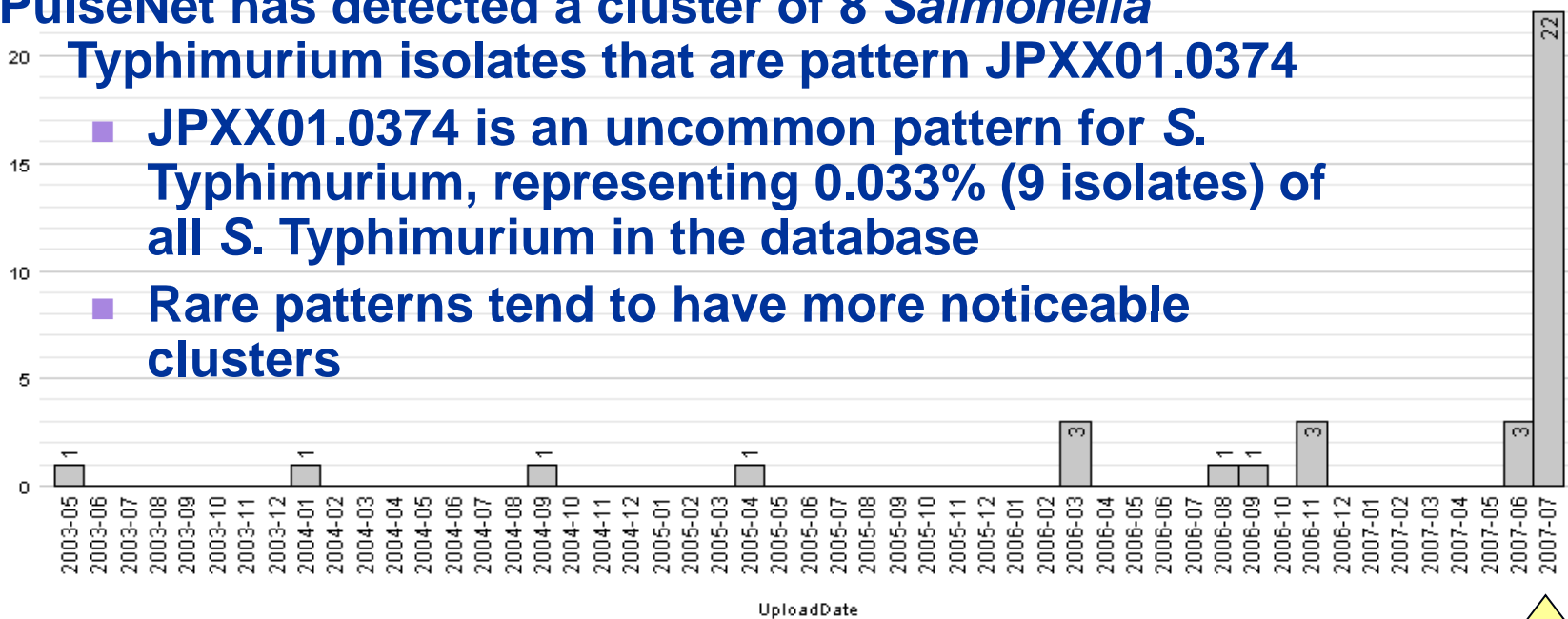


# Pattern Frequencies

## Example 1

PulseNet has detected a cluster of 8 *Salmonella* Typhimurium isolates that are pattern JPXX01.0374

- JPXX01.0374 is an uncommon pattern for *S. Typhimurium*, representing 0.033% (9 isolates) of all *S. Typhimurium* in the database
- Rare patterns tend to have more noticeable clusters



An obvious increase



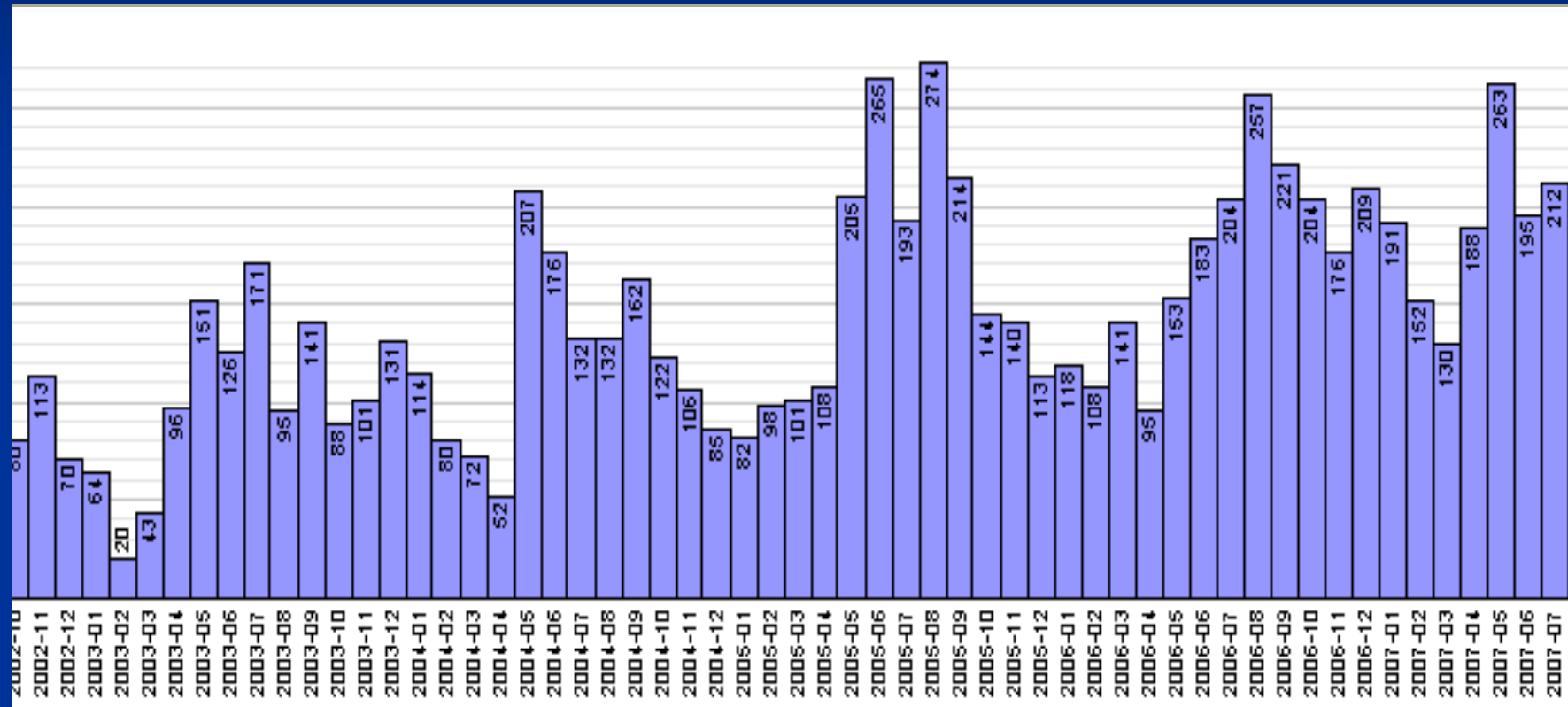
# Pattern Frequencies

## Example 2

- A local PulseNet participant has detected a cluster of 8 *Salmonella* Enteritidis isolates that are pattern JEGX01.0004
  - JEGX01.0004 is the MOST common pattern for *S. Enteritidis*, representing 45% (~9,500 patterns) of all *S. Enteritidis* in the database
  - More common patterns tend to have less obvious clusters, and are often defined by an increase above baseline

# Pattern Frequencies

## *Salmonella* Enteritidis “Pattern 4”



**Common Patterns: Detecting true increases can be difficult**

# CDC Team Response from CDC

! 354.6. [0803NHJPX-1c \(JPXX01.0275\) NH Typhimurium](#)

 [Nehal Patel](#)

04/01/08 03:19 PM

Hello!

This cluster of *S. Typhimurium* posted by NH has been given the cluster code 0803NHJPX-1c. The PFGE XbaI Pattern associated with this cluster is JPXX01.0275, comprising 0.36%(163/45000) of *S. Typhimurium* isolates in the database.

In the last 60 days, this pattern has been seen 6 times from the following: AR(2), CT, NH(2), and NY.

Nehal Patel

PulseNet Database Administration Team

Phone: (404) 639-1256

PulseNet: (404) 639-4558

**Updates will include updated epi information, updated line lists, etc.**

## Information included in reply posted to CDC Team:

1. Subject changed to format: <Outbreak Code> (<pattern number>)\_<LabID of posting lab>\_<Organism and/or Serotype>
2. Name of serotype and state that originally posted
3. Cluster Code assigned
4. Pattern assigned and frequency
5. Indistinguishable isolate(s) in the last 60 days (120 for *Listeria*) with linelist



# CDC Team Response from CDC

- CDC will respond to CDC Team postings within 2 business days.
  - If cluster code is assigned, a line list will be posted along with the cluster response. Labs listed as having indistinguishable patterns should follow-up locally.
  - If cluster code is not assigned, posting lab should continue to monitor the PFGE pattern locally for additional signs of increase.

# Overview

- Cluster Detection
  - Local and National
  - Verification of Data
  - Pattern Naming
- Response
  - Cluster Code Assignment
  - Line Lists, Frequency Graphs, and CDC Team Responses
- Follow-up
  - Cluster Reports, Epi Requests, and Cluster Conclusion
- Data Interpretation Issues
  - Similar patterns
  - Local vs. National Databases
  - Interpreting Patterns within Outbreaks



# Cluster Reports

- The Database Team meets every Monday morning to discuss recent and ongoing clusters
  - New clusters are assigned a status of **Active (PN)**
  - PulseNet Canada and VetNet databases are checked against Active (PN) clusters before reporting to epis
  - **Active (PN)** clusters are reported to CDC Epis, who decide which to include in **Active** investigation

**Regardless of cluster status, labs should continue to investigate locally and post additional matches not included in initial line list**

# Cluster Reports

Sample cluster report that is sent to the CDC epis after meeting

## Salmonella

JKA

0808OHJPX-1c. S. Typhimurium posted by OH

XbaI: JPXX01.1126 (0.13%)

BlnI: JPXA26.0029 (0.13%%)

60 Days: AK, ME, NM(2), NYC, OH(5), PA(4), VT

Source:

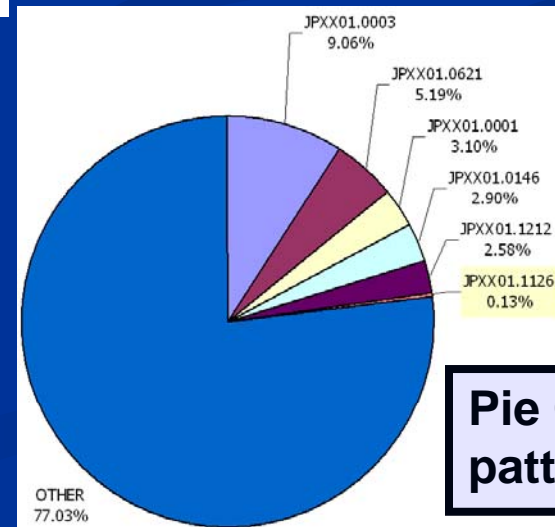
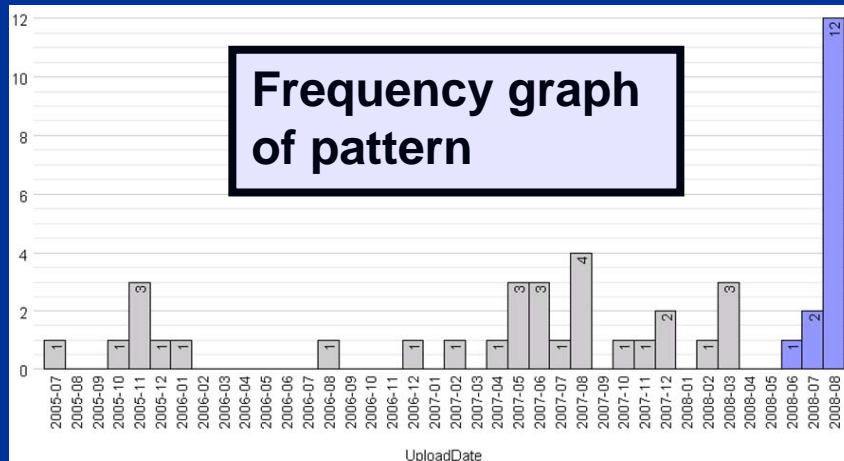
Notes no recent matches in PNC

USDA/FDA Nothing recent in VetNet

MLVA Info:

**Epi Update: Sent to Thai-An Nguyen for follow-up**

- Outbreak Code, serotype, posting lab
- Patterns involved and %
- States involved
- Notes, USDA/FDA/MLVA info
- Epi Update



# Cluster Reports

- A bundle file with representatives of every cluster for the calendar year is posted to CDC Team each Friday (organism-specific)
  - Save the bundle file in your local database each week to see if you have any matches to existing clusters
- In addition to the bundle file, a report of **Active** and **Active (PN)** clusters is posted to CDC Team each Friday
  - If you have matches to a cluster in the weekly bundle, refer to the Active Cluster Report to see if that cluster is **Active**, **Active (PN)**, or **Closed**
  - If your matching isolates are not already included on the Active Cluster Report, post your matches to CDC Team for that cluster
  - **REMINDER: DO NOT** reply to emails from CDC Team—post **DIRECTLY** on CDC Team





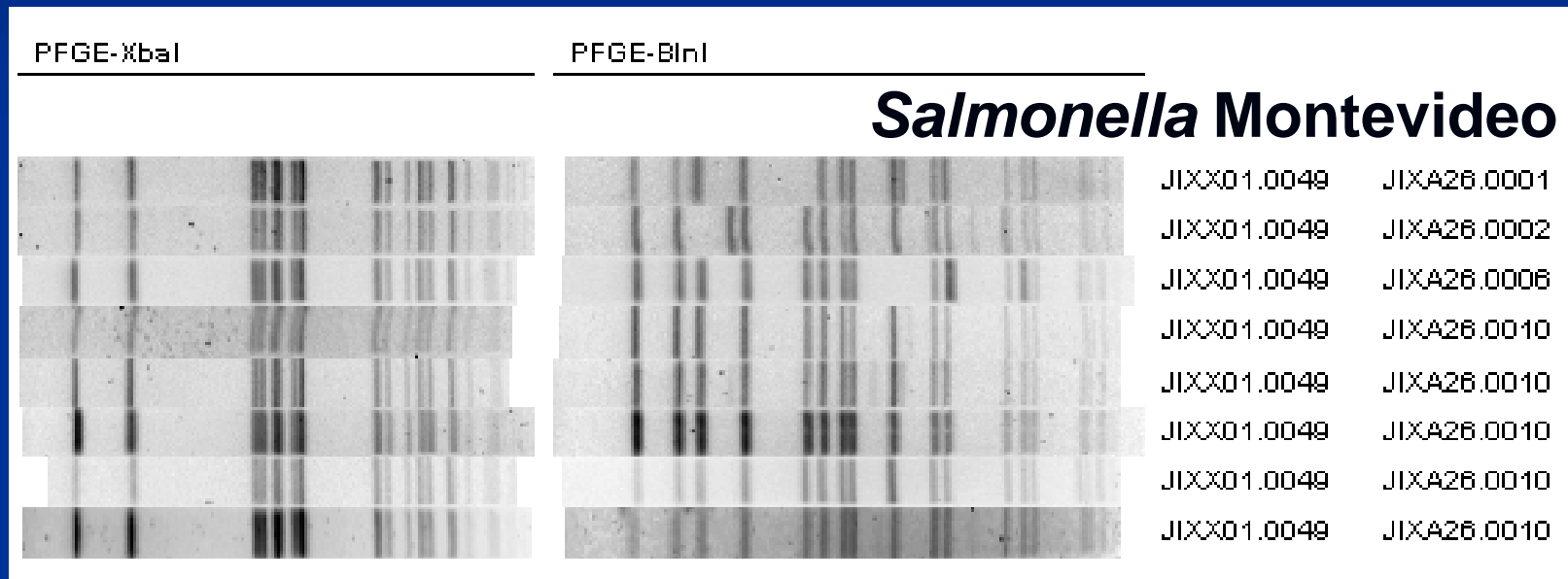
# Epi Requests: Additional Information

- CDC epis communicate with Public Health Lab epis; Public Health Lab epis and PulseNet lab participants should communicate at the state level
- Second enzyme results
- MLVA results
- VetNet matches
- International matches
  - CDC has direct access to PulseNet Canada databases
    - Client-only: compare bundle file of USA patterns to online Canada database
  - PulseNet participants abroad (PulseNet Latin America, Europe, Asia Pacific, and Middle East)



# Second Enzyme Results

## Example



Indistinguishable 1<sup>st</sup> enzyme patterns: need 2<sup>nd</sup> enzyme results to distinguish clusters

# Second Enzyme Results

- Adds to the discrimination of all organisms when using PulseNet protocols
- Can limit the need for epi follow-up of likely unrelated cases
- Running fingerprint profiles of both enzymes on the initial gel is often cost-effective and saves time.

# MLVA Analysis

- Sequence-based subtyping
- Can further discriminate common PFGE patterns through highly variable target sequences
- Data may be epidemiologically more relevant than PFGE data
- Results more straightforward
- Currently MLVA results are housed in databases separate from PFGE; however, the ultimate goal is to have them in combined databases

# Cluster Conclusion

- Investigation Ends
  - With weekly active cluster report from epis, PulseNet is notified when a cluster has been closed from the epi side
  - Summary posted to CDC Team for large outbreaks
  - Cluster thread is archived and removed from CDC Team after 3 weeks of inactivity
    - CDC PulseNet has Word documents of all archived postings



# Overview

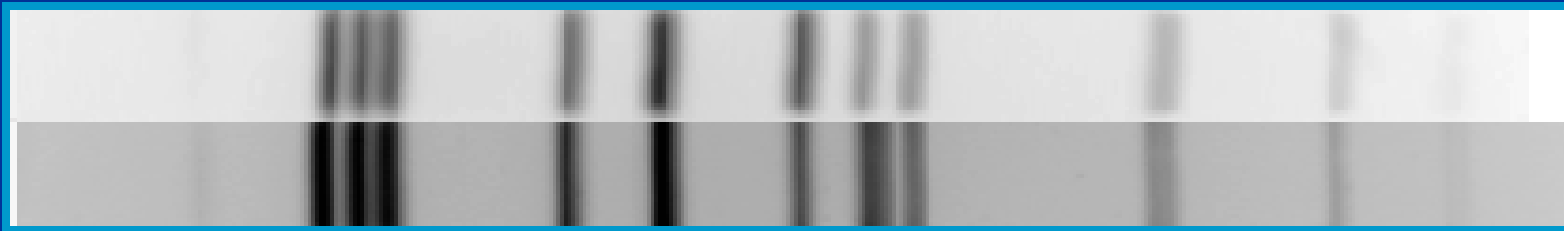
- Cluster Detection
  - Local and National
  - Verification of Data
  - Pattern Naming
- Response
  - Cluster Code Assignment
  - Line Lists, Frequency Graphs, and CDC Team Responses
- Follow-up
  - Cluster Reports, Epi Requests, and Cluster Conclusion
- Data Interpretation Issues
  - Similar patterns
  - Local vs. National Databases
  - Interpreting Patterns within Outbreaks



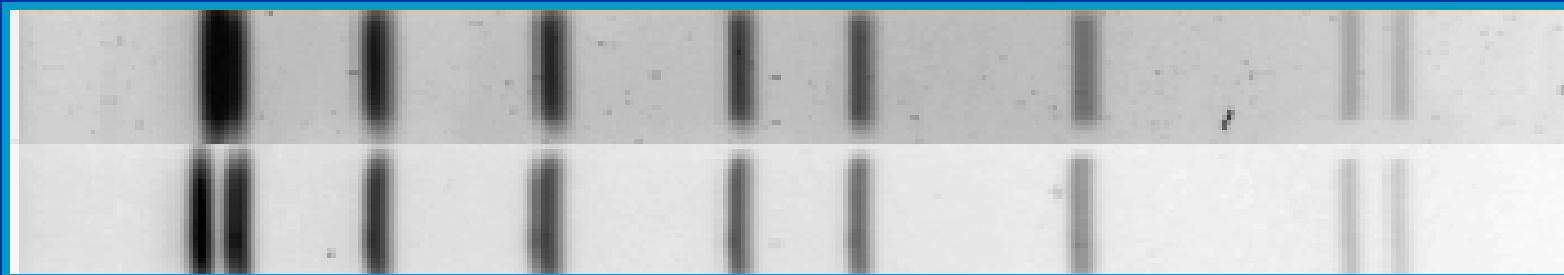
# Data Interpretation: Similar Patterns

- Patterns match at 100% in BioNumerics

## Example 1

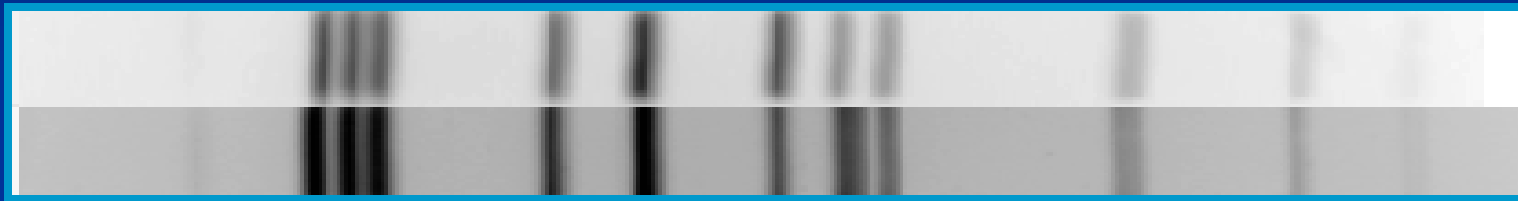


## Example 2



# Data Interpretation: Similar Patterns

## Example 1



- Patterns appear different upon further inspection
  - Bands marked the same
    - Marked incorrectly?
  - Doublet vs. Singlet



# Data Interpretation: Similar Patterns

## Example 2



- Patterns appear different upon further inspection
  - Bands marked the same
    - Marked incorrectly?
  - Spacing difference

# Local vs. National Databases

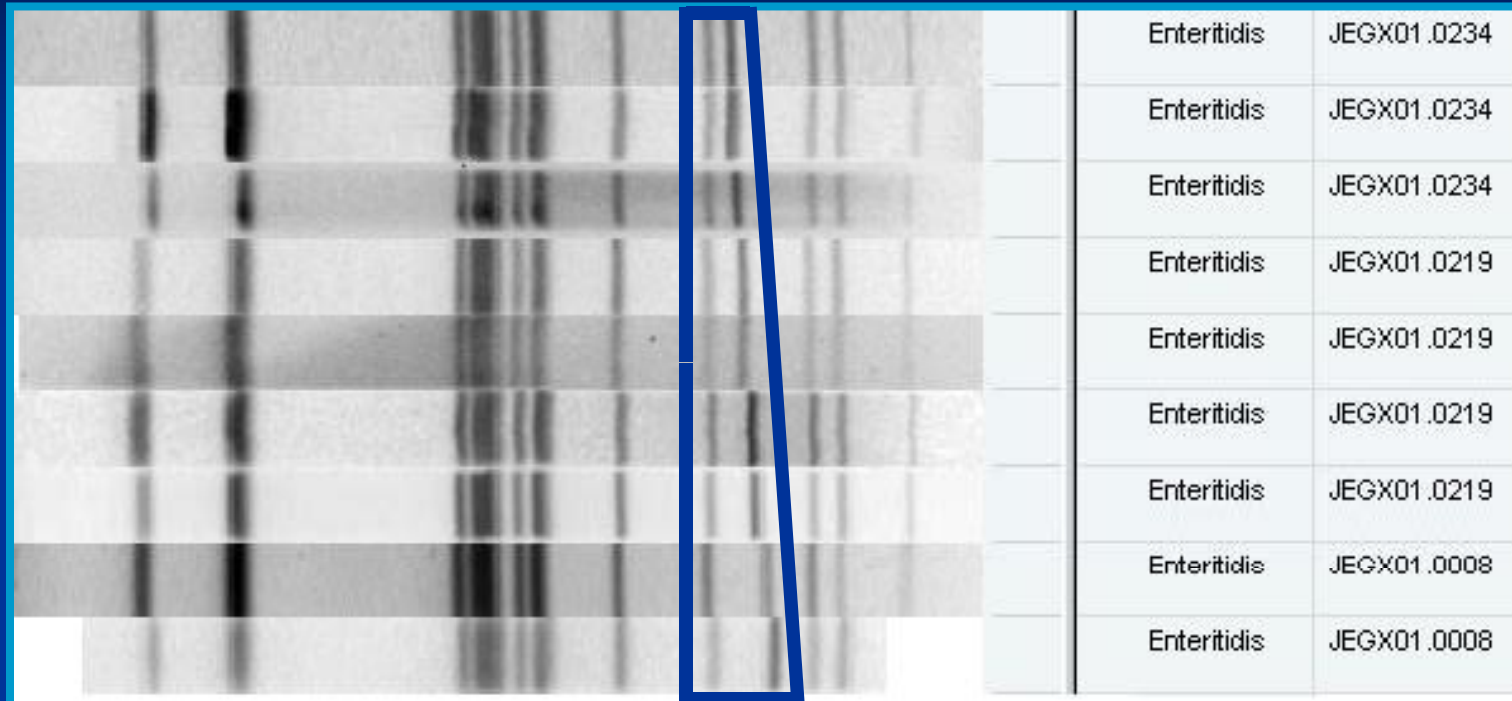
## Local

- Smaller, results from a few people
- Smaller representation of patterns
- Tends to be less diverse
- Sometimes different “common” patterns than seen nationally or in other regions
- Can only see local representation, no national perspective

## National

- Larger, results from many different labs
- Larger representation of patterns
- More diverse
- Different “common” patterns than seen locally
- See representation from all over, different regions

# Local vs. National Databases



Enteritidis	JEGX01.0234
Enteritidis	JEGX01.0234
Enteritidis	JEGX01.0234
Enteritidis	JEGX01.0219
Enteritidis	JEGX01.0219
Enteritidis	JEGX01.0219
Enteritidis	JEGX01.0219
Enteritidis	JEGX01.0008
Enteritidis	JEGX01.0008

- Example of what is often seen in a national database
  - Cannot always see the differences in the patterns
  - As more patterns are uploaded to the database, pattern names may change
  - If a PHL has a question about a pattern name they should call or email the database team—it can be explained why it is named that way

# Interpreting Patterns within Outbreaks: General Rules

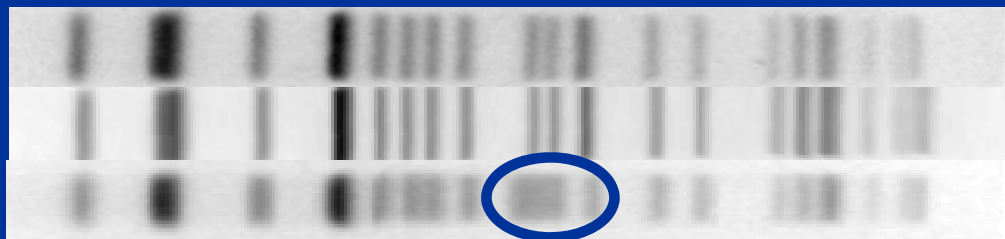
- Rerun isolates with unclear bands, poor resolution
  - Reproducibly different = different patterns
- Err on the side of calling things the same
- Additional tests may be necessary
- Sometimes pattern designations need to be reconsidered

# Interpreting Patterns within Outbreaks

- For surveillance (cluster detection), allow no differences for possible matches
- For outbreaks (case definition), differences may be acceptable
  - May be necessary for shigellosis infections (person to person transmission)

# Interpreting Patterns within Outbreaks

- State 1 and State 2 epi-linked isolates (common pattern)
- State 3 pattern looks similar but hard to determine if indistinguishable
- State 3 should repeat PFGE on isolate, run second enzyme



**Epi-linked: State 1**

**Epi-linked: State 2**

**Concurrent OB: State 3**

# Summary

- Cluster Detection
  - Local and National
  - Verification of Data
  - Pattern Naming
- Response
  - Cluster Code Assignment
  - Line Lists, Frequency Graphs, and CDC Team Responses
- Follow-up
  - Cluster Reports, Epi Requests, and Cluster Conclusion
- Data Interpretation Issues
  - Similar patterns
  - Local vs. National Databases
  - Interpreting Patterns within Outbreaks

# Questions?



**Thank you for your attention**  
The findings and conclusions in this presentation are those of the author and do not necessarily represent the views of the Centers for Disease Control and Prevention