Genomics Overview/WGS and Metagenomics

Martin Wiedmann

Gellert Family Professor of Food Safety

Department of Food Science

Cornell University, Ithaca, NY

E-mail: mw16@cornell.edu

607-254-2838

Outline

- **Why are we here?**
- Background on tools: from PFGE to WGS
- Case studies on WGS use for outbreak and in-plant environmental investigations
- Metagenomics
- Other WGS and NGS applications in dairy

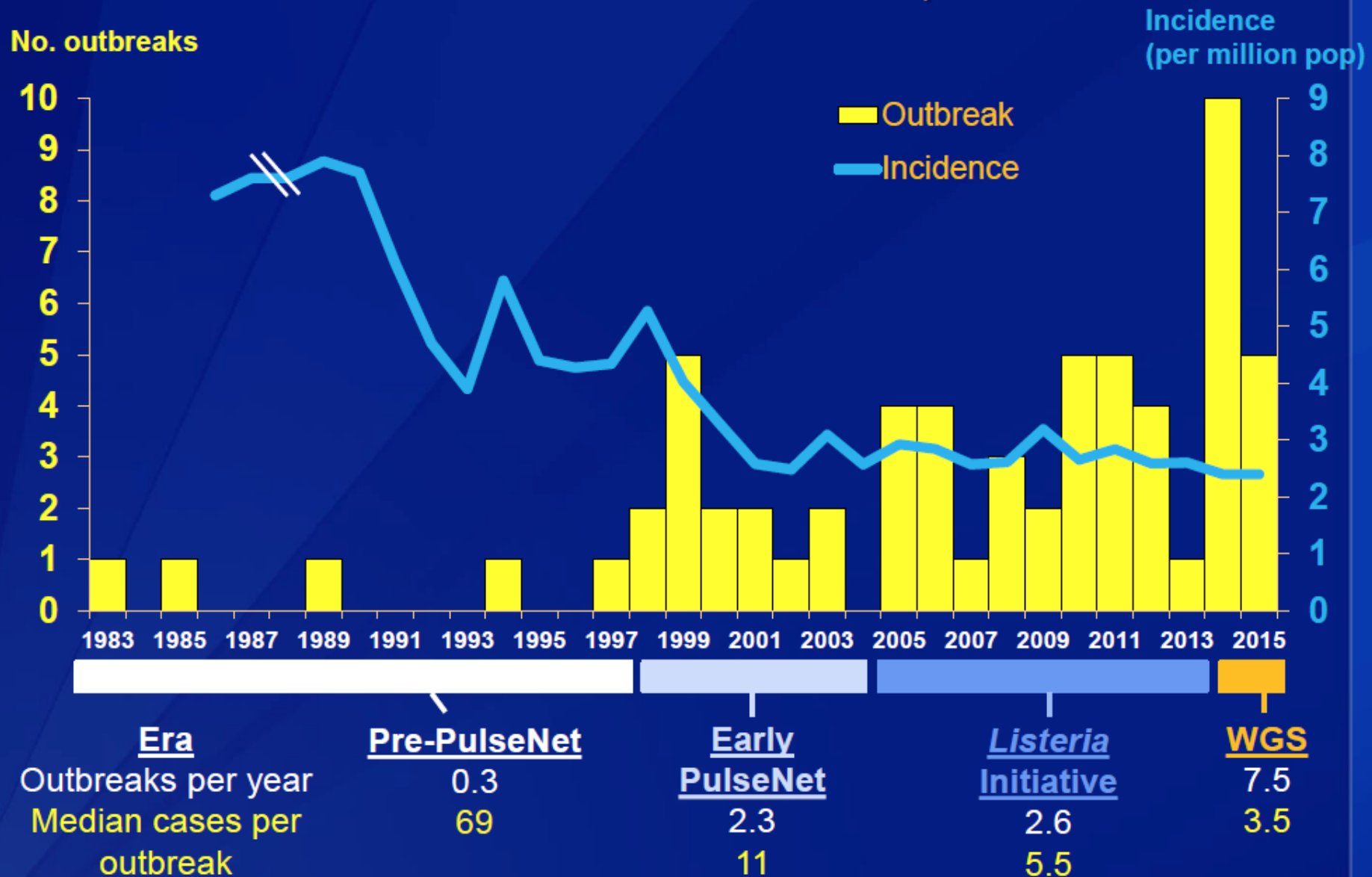# Food Safety News
Breaking news for everyone's consumption

CDC/FDA Partnership Targets Whole Genome Sequencing of Listeria Monocytogenes

By Brian Sauders | November 27, 2013

In a prior APHLTech blog post (NGS in Action: FDA's Genome TRAKR Network), Victor Waddell of the Arizona State Public Health Laboratory described the newly formed network of laboratories formed by the U.S. Food and Drug Administration (FDA). Known collectively as Genome TRAKR, the member laboratories perform whole genome sequencing (WGS) on bacterial foodborne pathogens isolated primarily from food and environmental sources.

On Sept. 1, 2013, the Centers for Disease Control and Prevention (CDC) began a partnership with the FDA Genome TRAKR network to utilize the network to conduct WGS of all Listeria monocytogenes collected from reported human illness cases in the United States. This effort leverages public health resources to evaluate and

Listeria Outbreaks and Incidence, 1983-2015

No. outbreaks

Incidence (per million pop)

- Outbreak
- Incidence

Era	Pre-PulseNet	Early PulseNet	*Listeria Initiative*	WGS
Outbreaks per year	0.3	2.3	2.6	7.5
Median cases per outbreak	69	11	5.5	3.5

Courtesy: Amanda Conrad, Outbreak Preparedness and Response Branch

Multistate Outbreak of Listeriosis Linked to Soft Cheeses Distributed by Karoun Dairies, Inc.

Posted September 23, 2015 3:45 PM ET

- 24 people infected with one of the closely related *Listeria* strains have been reported from 9 states since August 8, 2010.
  - 22 people were hospitalized. Five illnesses were pregnancy-related; one resulted in a fetal loss. One death was reported from Ohio.

  - FDA isolated *Listeria monocytogenes* from two environmental samples collected in September 2015 from the Central Valley Cheese, Inc. manufacturing facility in Turlock, California. Central Valley Cheese, Inc. manufactures cheese for Karoun Dairies. Whole genome sequencing showed that the two isolates are closely related genetically to isolates from ill people.

In addition, whole genome sequencing showed that 5 Listeria isolates collected in 2010 from the same facility were also closely related genetically to isolates from ill people.

March 2015: Listeriosis cases linked to Blue Bell ice cream

```
284881 tcttcaagtg caatatgaaa tttttgcttc aatgtagttc tagcgacacg ctcttgcatg
284941 ataatttgtt ccatgacttg agttaaatct tccacaattt ctgcttccgt tttgcccaga
285001 gttacttgat ttgagacctg gaaaatatta cttgcaggca tgctaccttc tccgtatata
285061 cctcttacca caaaacctag gctcctaatc gcttcaatca cacttttaat cctttttgtt
285121 gttacaagtc ccggtaaatg taccatcaca gaagcgcgca tcccagtccc aatattagtc
285181 acacaactcg tcaaatatcc aaactcttta tcaaaagcat aactaagctt ctcttccaca
285241 taaccatcaa tttgcaacgc tgcctctaac gcatcgaata atctcaagcc aggcgtcata
285301 cactgtatcc gcaaatgatc ttcttcattc aacataatac tcacattttc ttcctcattt
285361 aaaagaaccg caccatattc actcttattc atcatataag gactaattaa atgtttttcc
285421 accaaaagag ccttttgtaa caaagaaatt tgattcattt taatcaaaat aaaattatca
285481 tcaaaaactt cggcaatatt atcgacaatc tcttccttct gctcataaat cggaaactgt
285541 tcatccttca aattcctagc aagcctaata cgtgagctta aaacgacatc atcgtcatca
285601 ccagcatttt ctaaccaaga actaagccgc ggttcaaata cattcattta tcctcacctc
285661 cagctttcaa agccttgatt tcatcacgga caacagcagc ttcctcaaac tcctcatttt
285721 taactagctg tgctaacttt tcttgtaaac gactaatttc attttgtaca ccttccgctc
285781 gttcaacttc tgctggaacc ttaccaatat gttttttata tccattttgc actcgaccaa
285841 taatcggtac taactgtgcc tcaaaagccg aataacactc cggacatcca actcgattag
285901 tctttaaaaa ctcttcaaaa ctaagctggc aattctcaca gacaacctct ttctgcgtat
285961 tttcctcttt ttgcaaaagc gttaagaaat caagagcaac ctcactaaaa gtgtccatcg
286021 ctttaactaa atcttttttct gaggagagag cttcgtccgt agcacaattc tcgcataggt
286081 acaatgactc cactttccca agctcattta attgttgcaa ggctataact gctttattttt
286141 ctccacatct ttgacaaatc atcgtcatca caccttaatc atatttcaaa gcaactaaca
286201 tcgcctctaa aatcctactt ctcaaaatat ctctatccgg taaaggtaaa attaaaactt
286261 cacggtccaa tgccgcaacc attaatcttg cttctttctt cgttattact tcttcttcaa
286321 gcaacctcaa aataacatct tcagaaaatg attgagagac cttttttatca tgaaccattg
286381 aaatgattgc ttctaacaat tgaagtttat cattcatttt cactttaata atccgaatat
286441 atccaccacc accacgtttg ctttcaacaa tataccctcg ttccattgta aatctagtgt
286501 ttattacata gttaatttgt gaaggtacac attcaaactt atctgcaatt tcacttcttt
286561 taatttcaac tgcttcacta gattccaata cttgtttcaa ataagcttct ataatatcag
286621 aaatattttt cattataagg aatctccttt ctgttttttg caaatcattg actttgacta
286681 tttttgactt taattatact agatatcccc acctattagc aatttattta tcttttacaa
286741 acaaaaaaag ccctaaataa ttaggacttt tcaaattgcc cggcagcgac ctactctcgc
286801 aggggaagc ccccaactac cattggcgca gagaagctta actaccgtgt tcgggatggg
286861 aacgggtgtg acctt
//
```

Draft Genome Sequences of 33 *Salmonella enterica* Clinical and Wildlife Isolates from Chile

Magaly Toro,[a,b] Patricio Retamal,[c] Marc Allard,[a] Eric W. Brown,[a] Peter Evans,[a] Narjol Gonzalez-Escalona[a]

Center for Food Safety and Applied Nutrition, Division of Microbiology, Food and Drug Administration, College Park, Maryland, USA[a]; Department of Nutrition and Food Science, University of Maryland, College Park, Maryland, USA[b]; Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, Chile[c]

The data provided will help in understanding the differences between *Salmonella* strains isolated from different countries and continents, improving traceback investigations for foodborne-related outbreak events. Moreover, these new draft ge-

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

Log in

Pathogen Detection BETA

Learn More

ⓘ View the recent webinar: 'Introducing the Pathogen Detection Isolates Browser'.

NCBI Pathogen Detection integrates bacterial pathogen genomic sequences originating in food, environmental sources, and patients. It quickly clusters and identifies related sequences to uncover potential food contamination sources, helping public health scientists investigate foodborne disease outbreaks.

About

FAQ

Browser Factsheet

Antimicrobial Resistance Factsheet

Antimicrobial Resistance

Contributors

Find isolates now!

Data Resources

Examples:

1. Search for isolates encoding a mobile colistin resistance gene and a KPC beta-lactamase

Isolates Browser

Feedback

e to search

10:18 PM
3/11/2019

Outline

- Why are we here?
- **Background on tools: from PFGE to WGS**
- Case studies on outbreak and in-plant environmental investigations
- Metagenomics
- Other WGS and NGS applications in dairy

Bacterial Genomes

- – Typically between 2 and 10 million nucleotides (2,000 to 10,000 genes)
- – Pan-genome: All genes in a species
  - • Core genome: ~3,000–5,000 genes, present in most strains of a given species
  - • Accessory genome: up to thousands of genes, not always present

- Structure
  - – Double-stranded DNA
  - – Usually single, circular chromosome

Strain-specific genes

Accessory

Core

Accessory

Nature Reviews | Genetics
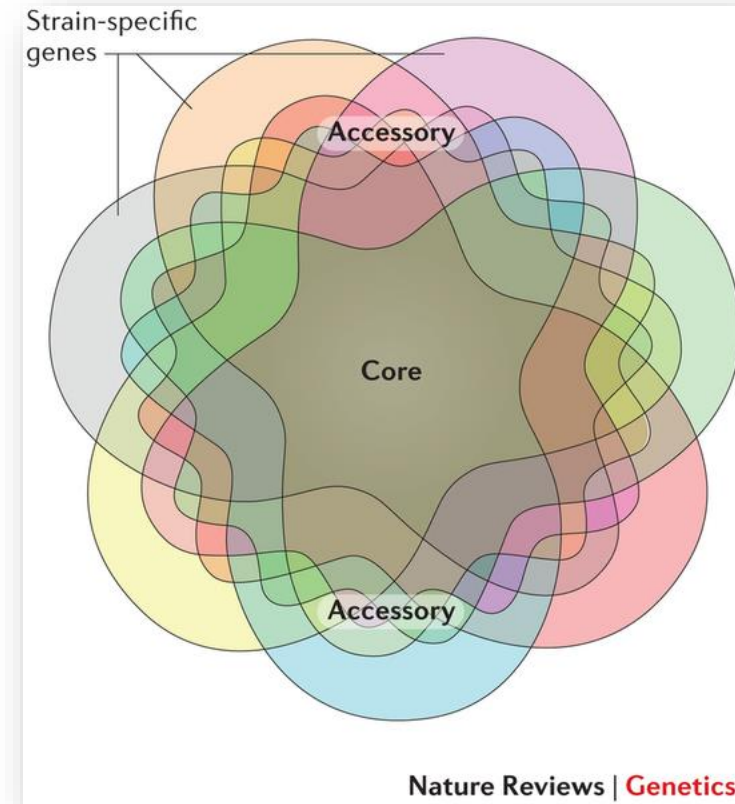
DNA-based strain typing methods using restriction enzyme-based methods

- Include PFGE and ribotyping

- Detect changes (mutations) in restriction enzyme recognition sites (6 – 8 bp in length) and length variations in sequences between recognition sites
  - Single bp mutation can potentially create a different subtype
  - Only small fraction of genome is "probed" for variation

Listeria monocytogenes EGD-e

chromosome
2944528 bp

Listeria monocytogenes EGD-e

chromosome
2944528 bp
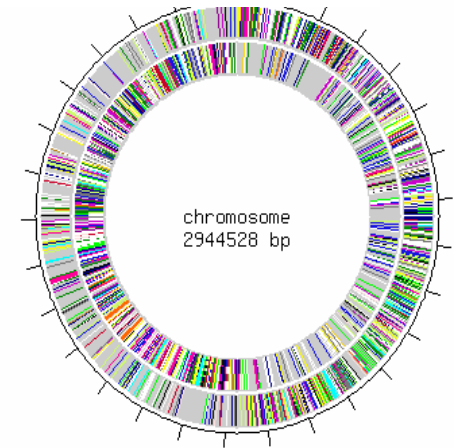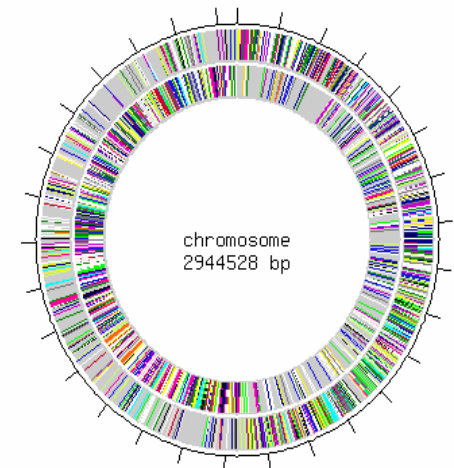
Pulsed Field Gel Electrophoresis
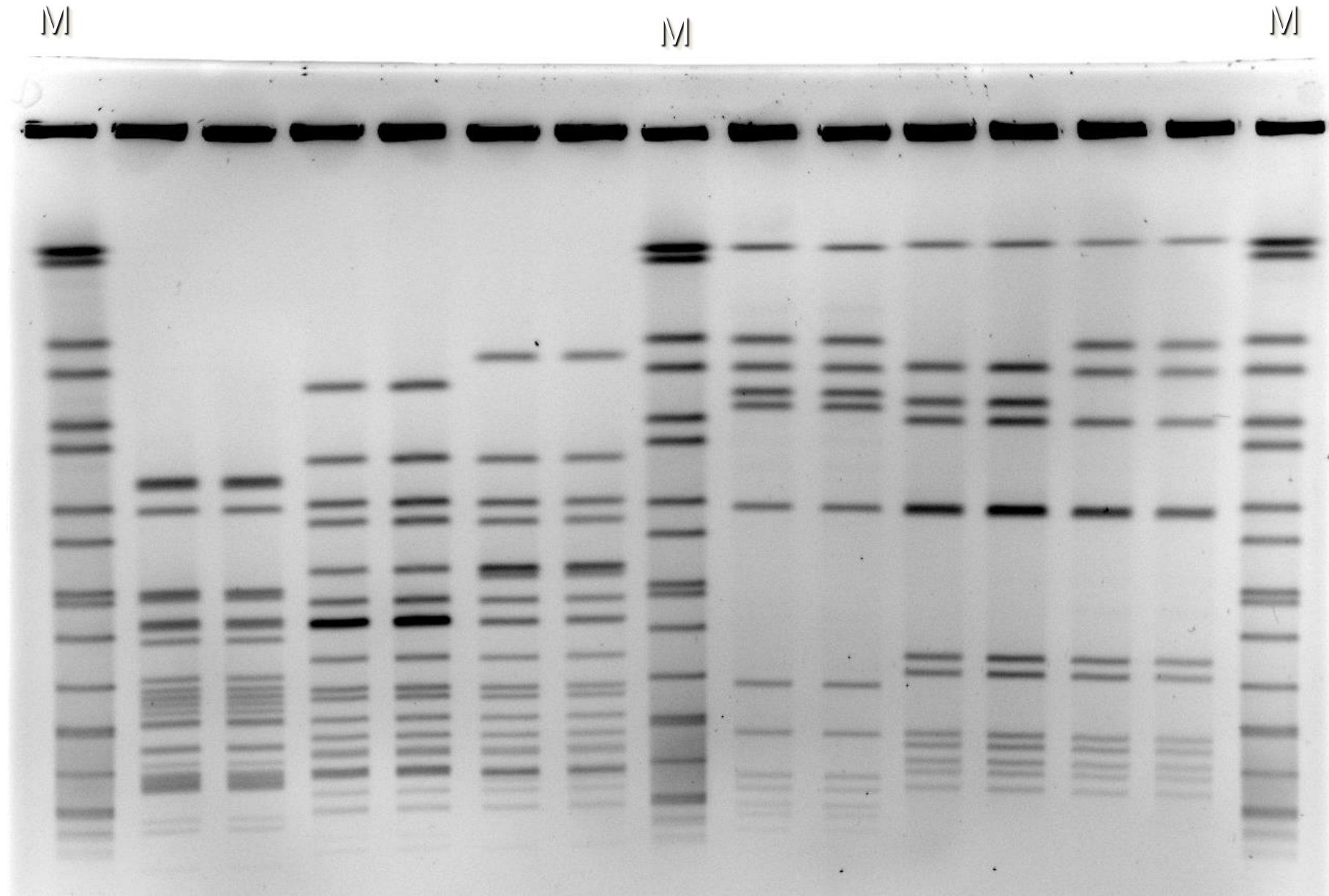
Challenges with use of PFGE as a subtyping method in outbreak investigations

- Two isolates may show the same PFGE type even though they are genetically distinct
  - PFGE only interrogates small part of the genome
- Two isolates may show "slightly" (?? - the "3-band rule") different PFGE patterns despite sharing a very recent common ancestor
  - Could be due to lateral genes transfer, loss of plasmid, rearrangements, point mutations etc.

DNA sequencing-based subtyping
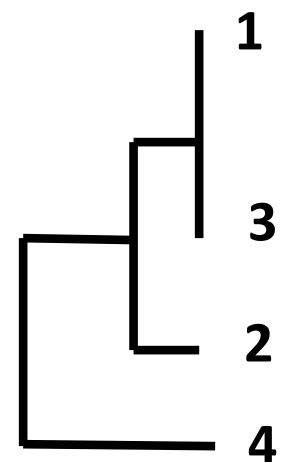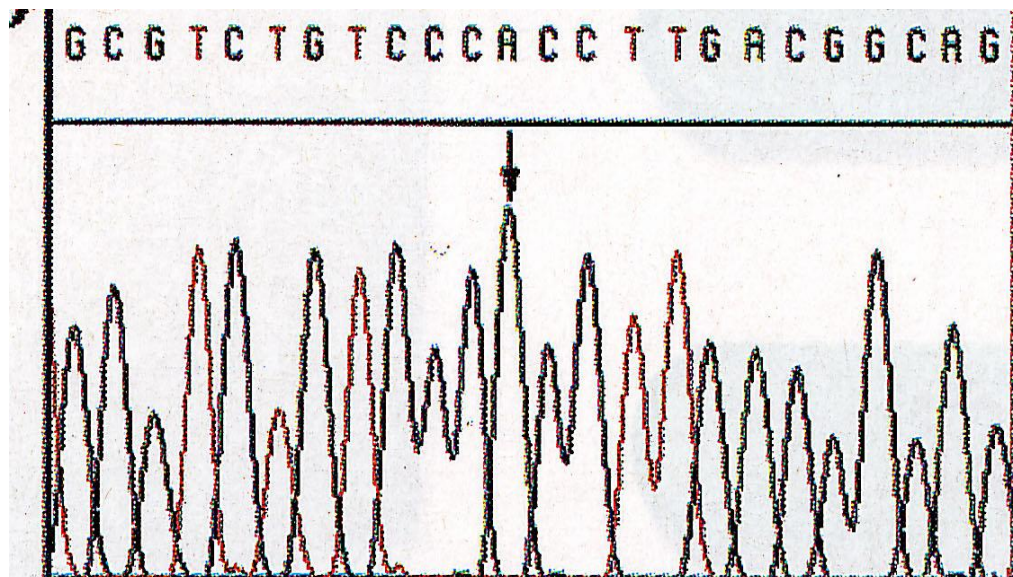
Isolate 1   AACATGCAGACTGACGATTCGACGTAGGCTAGACGTTGACTG
Isolate 2   AACATGCAGACTGACGATTCGTCGTAGGCTAGACGTTGACTG
Isolate 3   AACATGCAGACTGACGATTCGACGTAGGCTAGACGTTGACTG
Isolate 4   AACATGCATACTGACGATTCGTCGAAGGCTAGACGTTGACTG

SNP: single nucleotide polymorphism

Different sequencing-based methods

- MLST (multi-locus sequence typing):
  - Typically sequencing of 7 genes (only 600 to 700 nt parts of genes are typically sequenced)
- Whole genome sequencing
  - Complete genome is sequenced
  - Can be done by different methods, but today is typically done by "next gen sequencing" methods (NGS)
- "NGS" can be used for applications other than WGS:
  - **Metagenomics** as an important application
  - NGS include 2nd gen methods (Illumina, sequences genomes in short pieces, typically <500 nt) and 3rd gen methods (e.g., MinIon, sequences large fragments, >10,000 nt)

Different approaches to analyzing WGS data

- High quality SNP
  - Data analysis available through FDA and CDC pipelines; also available through BioNumerics
- MLST, including wgMLST and cg MLST
  - Data analyses available in BioNumerics; selected MLST schemes are also available through free an public platforms
    - wgMLST: CDC, PHE, PulseNet International
    - cgMLST: Institute Pasteur

Analysis of genome wide SNPs (wgSNPs)

- Identifies all high confidence SNPs over whole genome (approx. 3 to 5 million nucleotides)
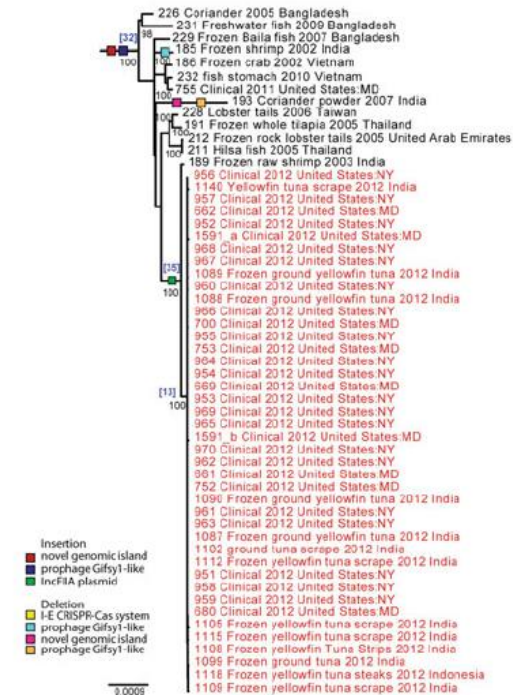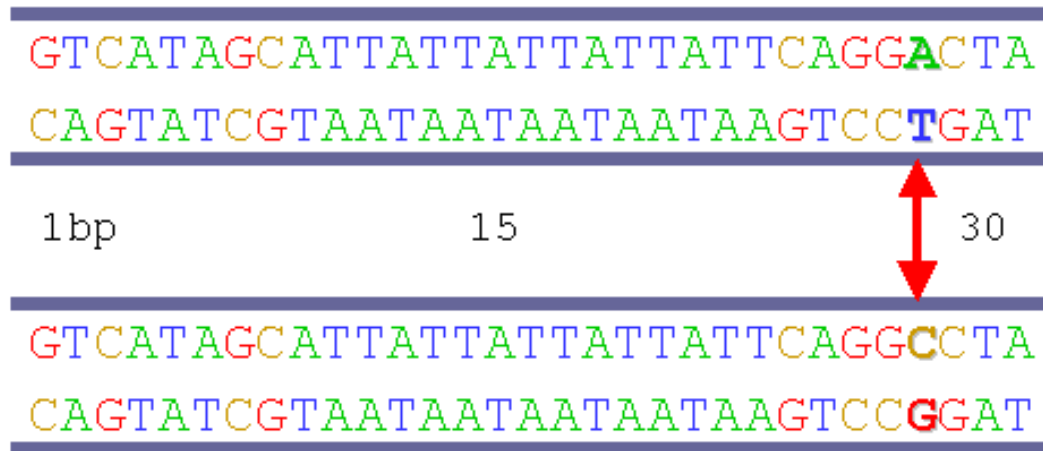
SNP data analysis: overview

```
Isolate 1 AACATGCAGACTGACGATTCGACGTAGGCTAGACGTTGACTG
Isolate 2 AACATGCAGACTGACGATTCGTCGTAGGCTAGACGTTGACTG
Isolate 3 AACATGCAGACTGACGATTCGACGTAGGCTAGACGTTGACTG
Isolate 4 AACATGCATACTGACGATTCGTCGAAGGCTAGACGTTGACTG
```
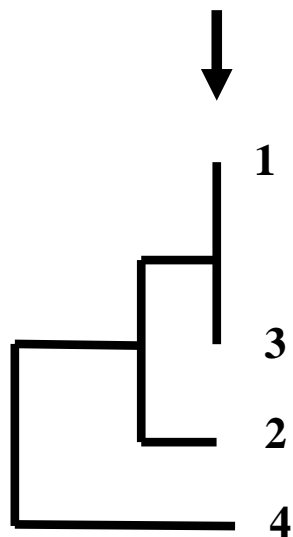

	1	2	3	4
1	0			
2	1	0		
3	0	1	0	
4	3	2	3	0

Whole genome multilocus sequence typing (wgMLST)

- Allows for simpler analysis and clear naming of subtypes
- Performs comparison on a gene by gene level

	Isolate A	Isolate B	Isolate C
Gene 1	1	1	1
Gene 2	8	8	12
Gene 3	5	5	2
Etc.			
Gene 1,005	4	4	4
wgMLST type	**A**	**A**	**B**

XbaI   *SpeI*
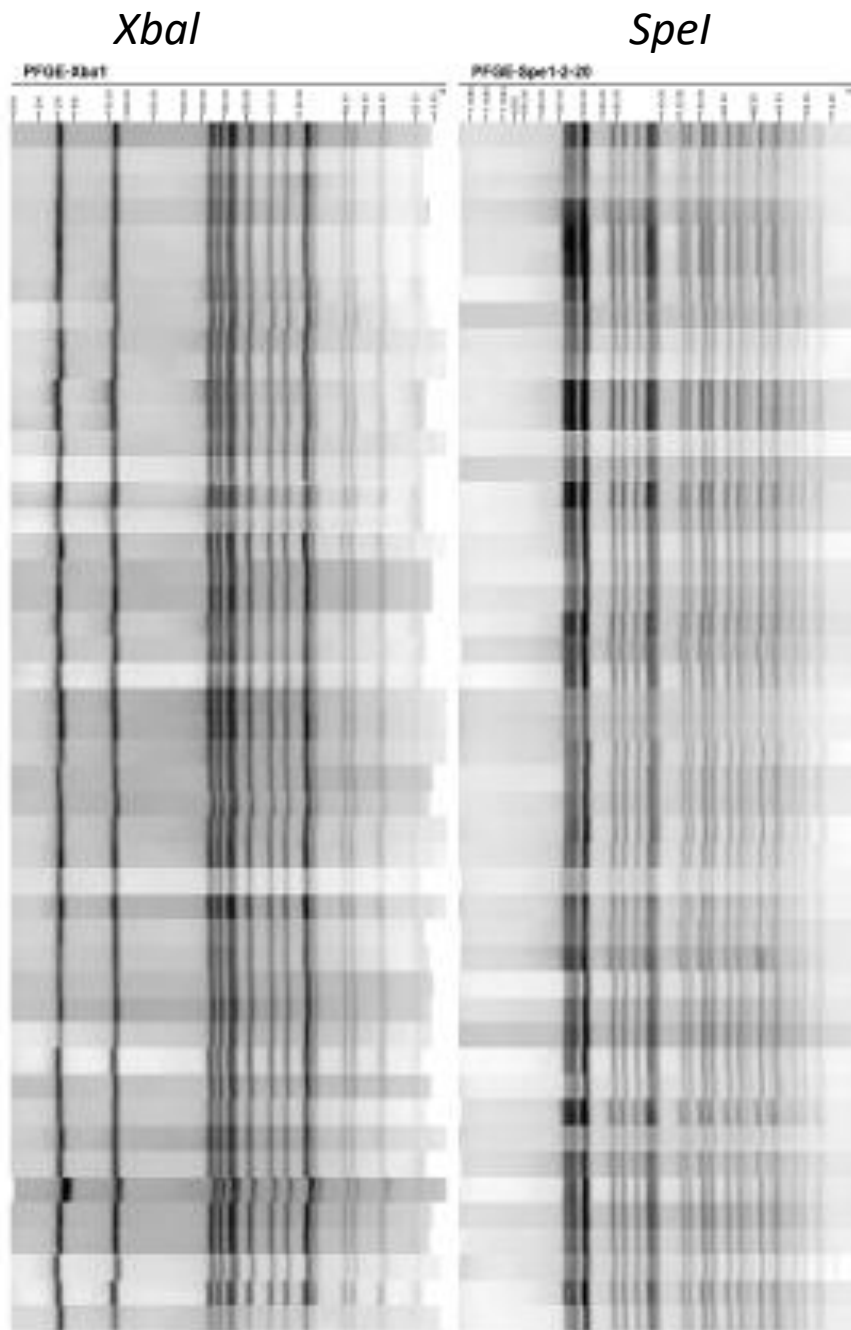
Includes isolates form Salmonella outbreak linked to sausages (Rhode Island) and isolates from pistachios

Den Bakker et al. 2011. AEM.

Tip-dated maximum clade credibility tree based on SNP data for 47 Montevideo isolates
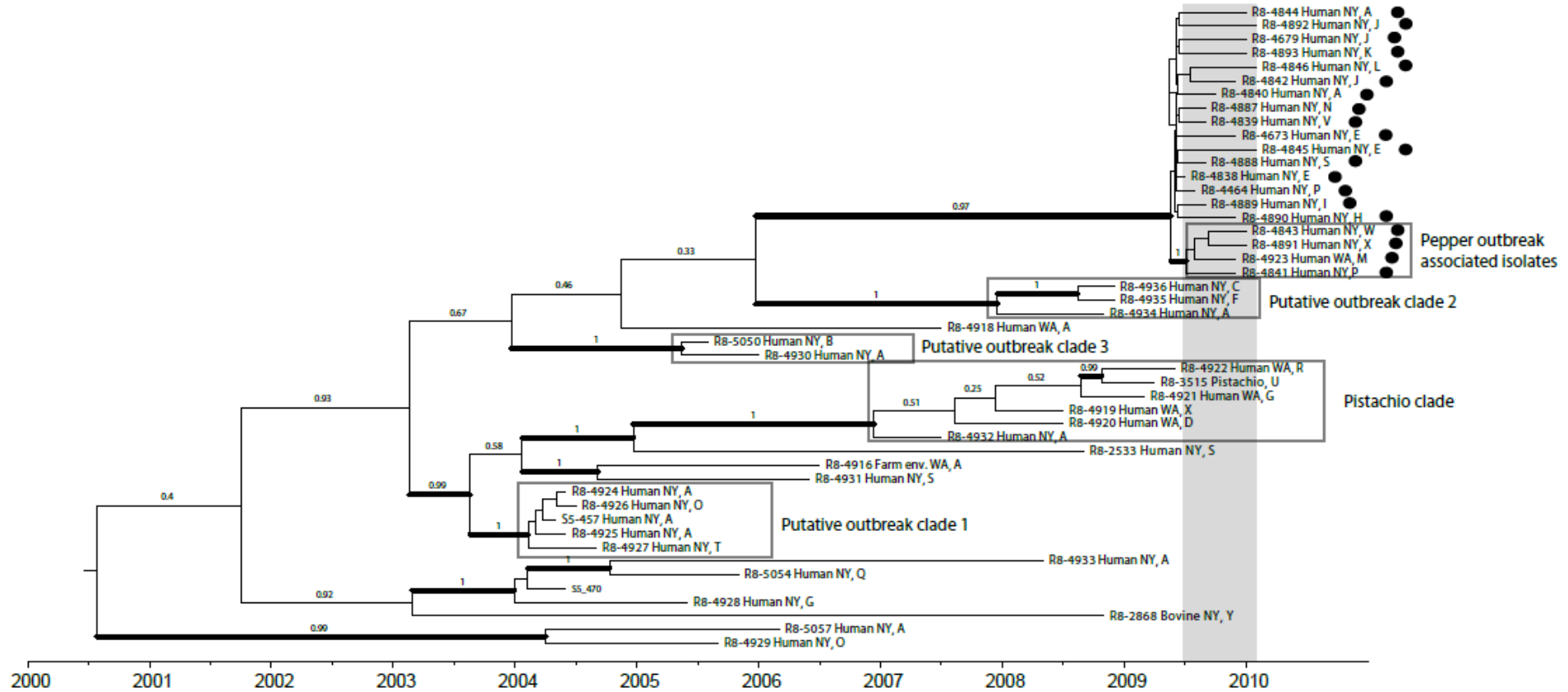
Outline

- Why are we here?
- Background on tools: from PFGE to WGS
- **Case studies on outbreak and in-plant environmental investigations**
  - **Case study 1: outbreak detection**
  - Case study 2: subtyping of food or environmental isolates
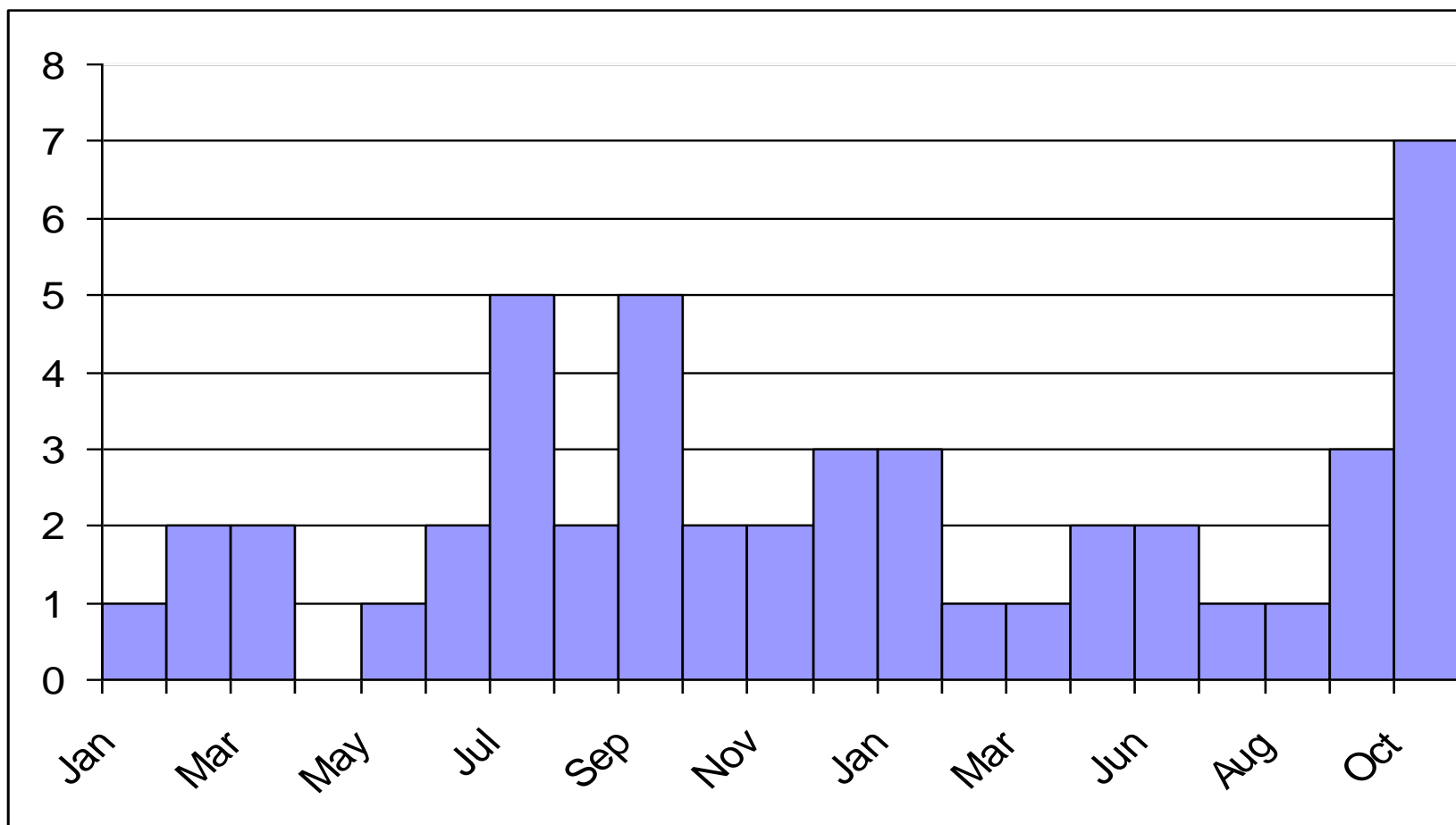
Subtyping results
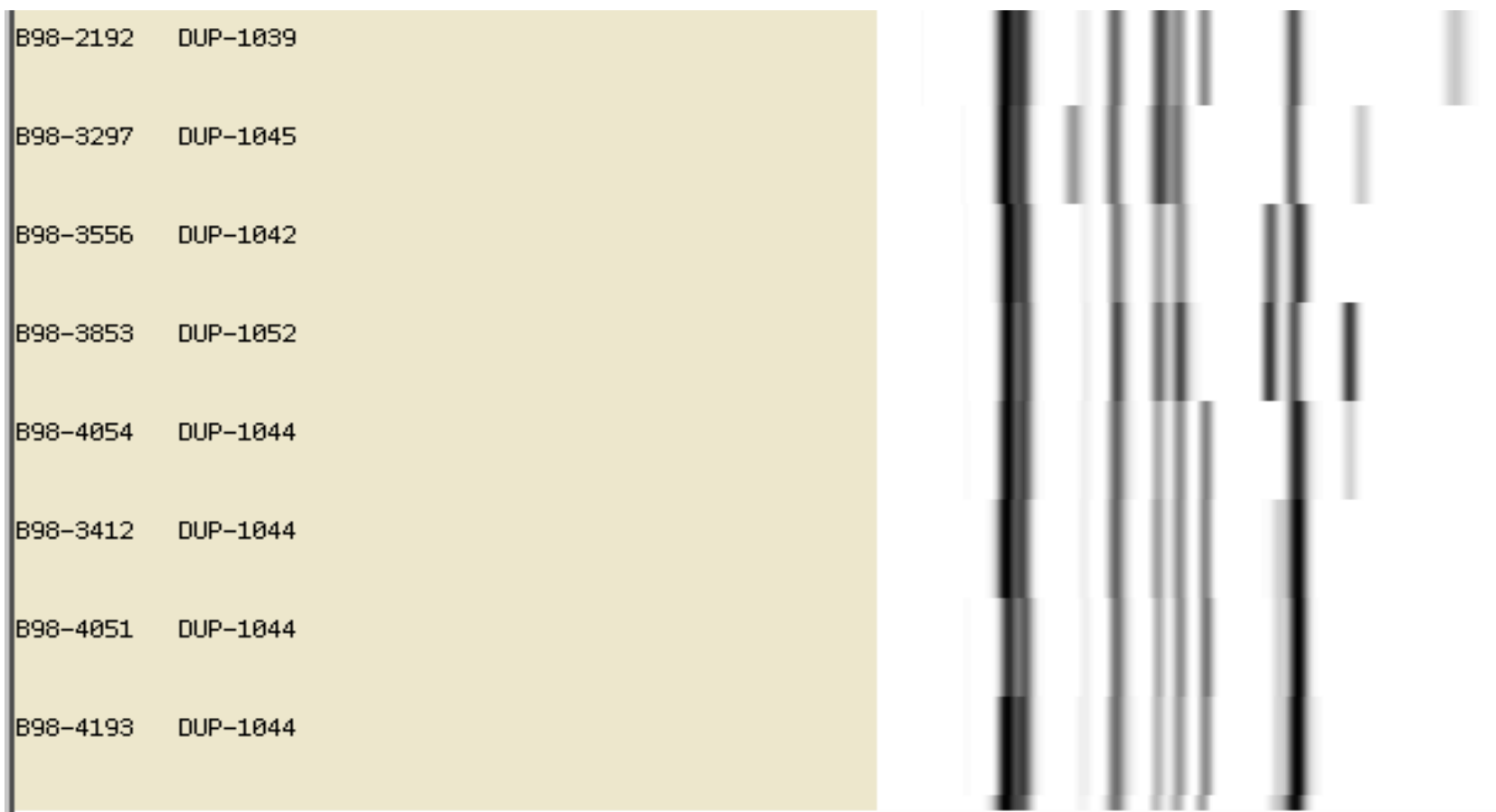
Epidemic curve for 1/97 - 2/99 in NYS
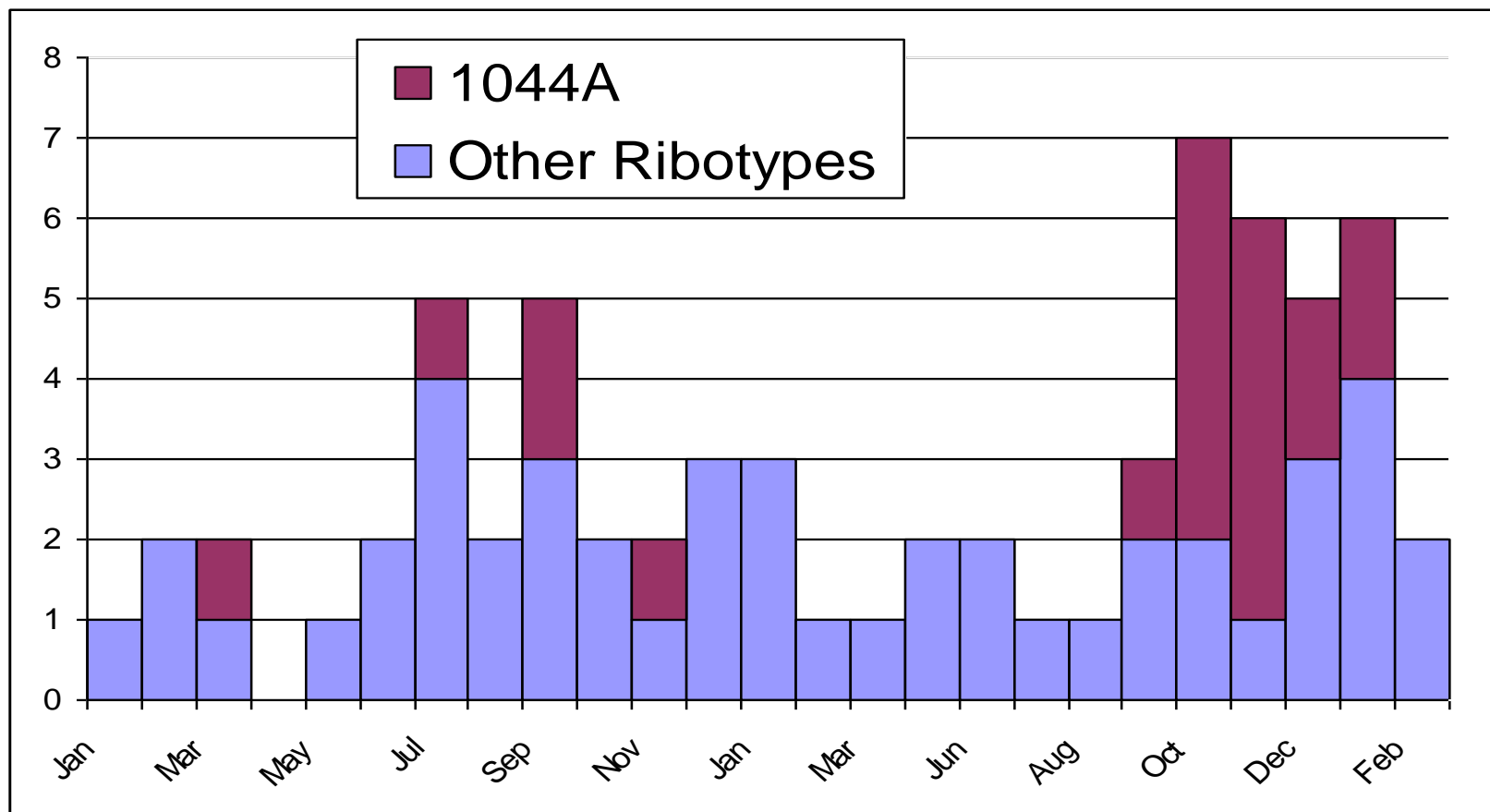
Exposure*	Cases exposed/total (%)	Controls exposed/total (%)	OR	95 % CI	P value*
First questionnaire					
Onion	10/17 (59)	11/21 (52)	1·3	0·3–5·8	0·69
Lettuce	13/18 (72)	14/21 (67)	1·3	0·3–6·6	0·71
Potato salad	9/16 (56)	11/21(52)	1·2	0·3–5·4	0·81
Chicken	17/17 (100)	17/21 (81)	Undef.	0·6–undef.	0·08
Turkey	9/18 (50)	6/19 (32)	2·2	0·5–10·5	0·25
Beef	17/18 (94)	10/21 (48)	18·7	1·9–456·6	0·002
Frankfurters, cooked	16/18 (89)	6/19 (32)	17·3	2·4–160·0	<0·0004
Frankfurters, any	16/18 (89)	8/21 (38)	13·0	1·9–111·6	0·001
Ham	9/14 (64)	5/21 (24)	5·8	1·0–35·4	0·02
American cheese	13/17 (76)	16/21 (76)	1·0	0·2–5·9	0·98
Mozzarella	9/17 (53)	7/21(33)	2·2	0·5–10·6	0·22
Cheddar cheese	10/16 (63)	9/21 (43)	2·2	0·5–10·7	0·24

Data sources for food history comparsions

- Case-control studies
- Case-case studies
  - May identify incorrect likely food sources if sporadic cases are not matched to outbreak cases
- Historical food consumption data

Investigation of the Outbreak

Epidemiologic evidence indicates that romaine lettuce is a likely source of this outbreak.

In interviews, ill people answered questions about the foods they ate and other exposures in the week before they became ill. Eleven (79%) of 14 people interviewed reported eating romaine lettuce. This percentage is significantly higher than results from a survey [PDF – 787 KB] of healthy people in which 47% reported eating romaine lettuce in the week before they were interviewed. Ill people reported eating different types of romaine lettuce in several restaurants and at home.

Whole genome sequencing (WGS) results showed that the *E. coli* O157:H7 strain isolated from ill people in this outbreak is closely related genetically to the *E. coli* strain isolated from ill people in a 2017 outbreak linked to leafy greens in the United States and to romaine lettuce in Canada ⬀ . The current outbreak is not related to a recent multistate outbreak of *E. coli* O157:H7 infections linked to romaine lettuce. People in the spring outbreak were infected with *E. coli* O157:H7 bacteria with a different DNA fingerprint.

Foodborne Diseases Active Surveillance Network (FoodNet)

Population Survey

Atlas of Exposures, 2006-2007

The Power of WGS: Single Case Linked to Pork Purchased from Live Animal Market

- ❏ **August 2014: Preterm infant born by emergency C-section; has sepsis and requires intubation**
- ❏ ***Listeria* isolated from blood and endotracheal tube**
- ❏ **Grandmother had purchased freshly-slaughtered pig ~1 week earlier; samples taken of leftover pork**

wgMLST (<All Characters>)

100

Pork isolate
Pork isolate
Blood isolate
Endotracheal isolate

1.5[0-3]
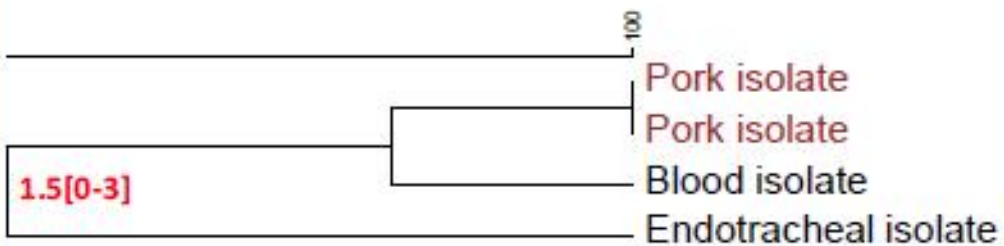
Sporadic case
- Food source identification

The theoretical background

- Bacteria divide asexually: Bacterial populations can be seen as large populations of "identical twins"

- Mutation rate during replication is low: extremes of the suggested mutation rates range from 2.25×10^{-11} to 4.50×10^{-10} per bp per generation

  - **With a genome size of around 5 Million bp per bacterial genome (5×10^6) between approx. 450 and 9,000 generations are needed for a single SNP difference**

  - Eyre et al. estimated evolutionary rate of 0.74 SNVs per successfully sequenced genome per year for C. difficile (N. Engl. J. Med. 2013)

    - "Whole-genome sequencing … identified 13% of cases that were genetically related (≤2 SNVs) but without any evidence of plausible previous contact through a hospital, residential area, or family doctor."

- **Unknown bacterial generation time in different environments complicates interpretation**

  - How often does Salmonella multiply in a dry facility (per year)??

Real world observations

BMC Genomics

BioMed Central

Research article

Open Access

Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment

Renato H Orsi[1], Mark L Borowsky[2,7], Peter Lauer[3], Sarah K Young[2], Chad Nusbaum[2], James E Galagan[2,4], Bruce W Birren[2], Reid A Ivy[1], Qi Sun[5], Lewis M Graves[6], Bala Swaminathan[6] and Martin Wiedmann*[1]

bacterial genome evolution in natural environments is limited. We thus performed full genome analyses on four *Listeria monocytogenes*, including human and food isolates from both a 1988 case of sporadic listeriosis and a 2000 listeriosis outbreak, which had been linked to contaminated food from a single processing facility. All four isolates had been shown to have identical

Results: The two *L. monocytogenes* isolates from 1988 and the two isolates from 2000 had highly similar genome backbone sequences with very few single nucleotide (nt) polymorphisms (1 – 8 SNPs/isolate; confirmed by re-sequencing). While no
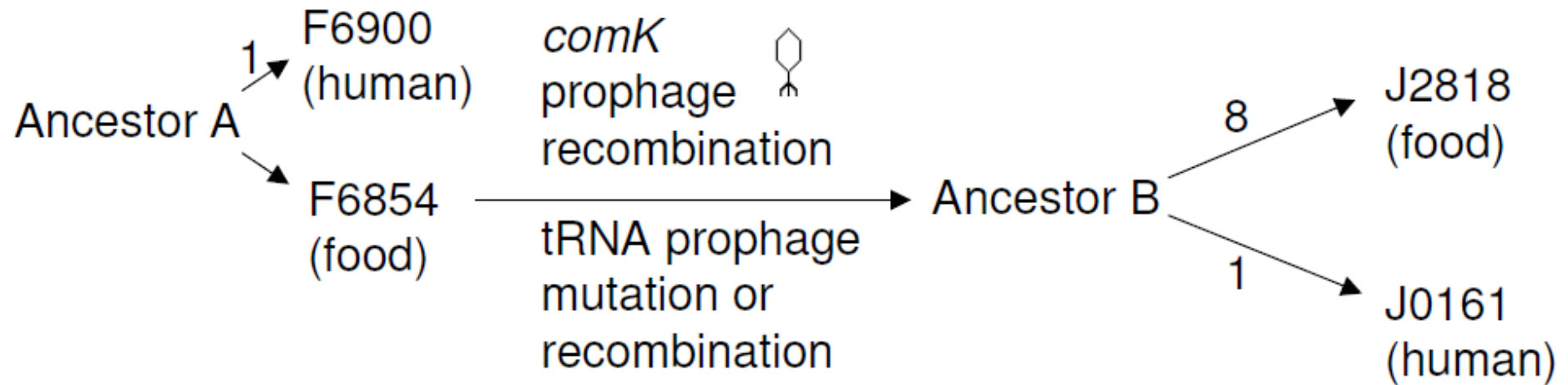
Figure 3
Schematic of the putative evolutionary history of the *L. monocytogenes* strain in the food facility between 1988 and 2000. Numbers on the arrows represent new mutations. Ancestor A is the ancestor of F6854 and F6900 (the food and human isolate, respectively, from the sporadic case in 1988) and Ancestor B is the ancestor of J0161 and J2818 (the food and human isolate, respectively, from the outbreak in 2000).

Can we come up with a SNP cut-off: The Listeria example
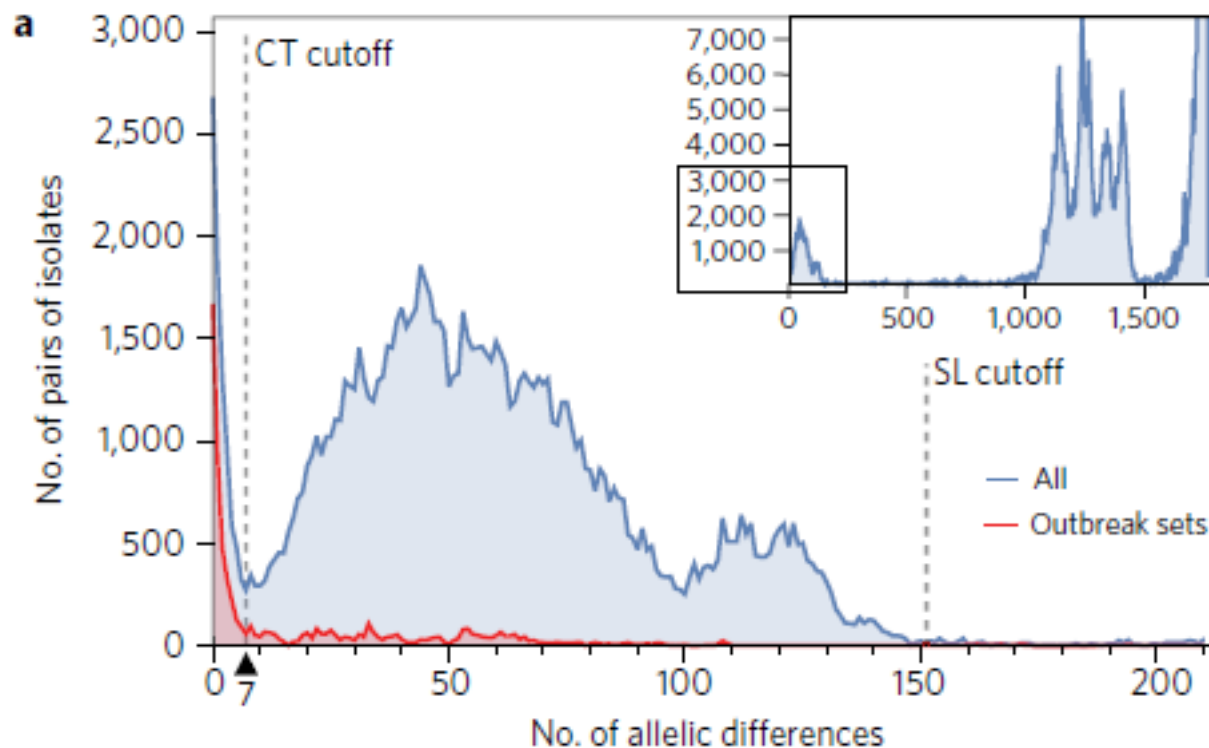
A number of studies suggests <u>average</u> divergence <u>by approx. 0.5 to 1 SNPs or alleles</u> per year

Moura et al., Nature Microbiology. 2016

A somewhat hypothetical case study

- Three people with listeriosis; the *Listeria monocytogenes isolates* differ by 2 SNPs
- All three people purchased cheese in super market B

A somewhat hypothetical case study

- Three people with listeriosis; the *Listeria monocytogenes isolates* differ by 2 SNPs

- All three people purchased deli meat in retail deli B

- *L. monocytogenes* with WGS type that differs by 1 - 3 SNPs from the human isolates is found in a drain in the retail deli B

A somewhat hypothetical case study

- Three people with listeriosis; the *Listeria monocytogenes isolates* differ by 2 SNPs

- All three people purchased deli meat in retail deli B

- *L. monocytogenes* with WGS type that differs by 1 - 3 SNPs from the human isolates is found in a drain in the retail deli B

- Subsequent work shows that additional *L. monocytogenes* isolates with WGS type that differs by 1 - 3 SNPs from the human isolates are found in 2 more retail delis in other states

Real world observations

AEM FULL TEXT

Whole genome sequencing allows for improved identification of persistent Listeria monocytogenes in food associated environments.

Stasiewicz MJ[1], Oliver HF[2], **Wiedmann M**[3], den Bakker HC[4].

33   supporting persistence of the strain.   In 13 events, nearly indistinguishable isolates (0-1 SNP)

34   were found across multiple delis.   No individual genes were enriched among persistent isolates

In one case, isolates with < 3 SNP differences were found in retail delis in there different states

Ecological prevalence, genetic diversity, and epidemiological aspects of *Salmonella* isolated from tomato agricultural regions of the Virginia Eastern Shore

Rebecca L. Bell[1]*, Jie Zheng[1], Erik Burrows[1], Sarah Allard[1†], Charles Y. Wang[1], Christine E. Keys[1], David C. Melka[1], Errol Strain[1], Yan Luo[1], Marc W. Allard[1], Steven Rideout[2] and Eric W. Brown[1]

PFGE pattern (pattern 25, JJPX01.0011). The strains with the smallest number of SNP differences to the clinical isolates in Clade 2a are two isolates from AREC, one from goose feces isolated in September 2010 (CFSAN000929) and one from creek sediment (CFSAN000927) isolated in August of the same year. These two AREC isolates appear to be sisters with

North American Migration Flyways
(with Principal Routes)

Atlantic Flyway
Mississippi Flyway
Central Flyway
Pacific Flyway
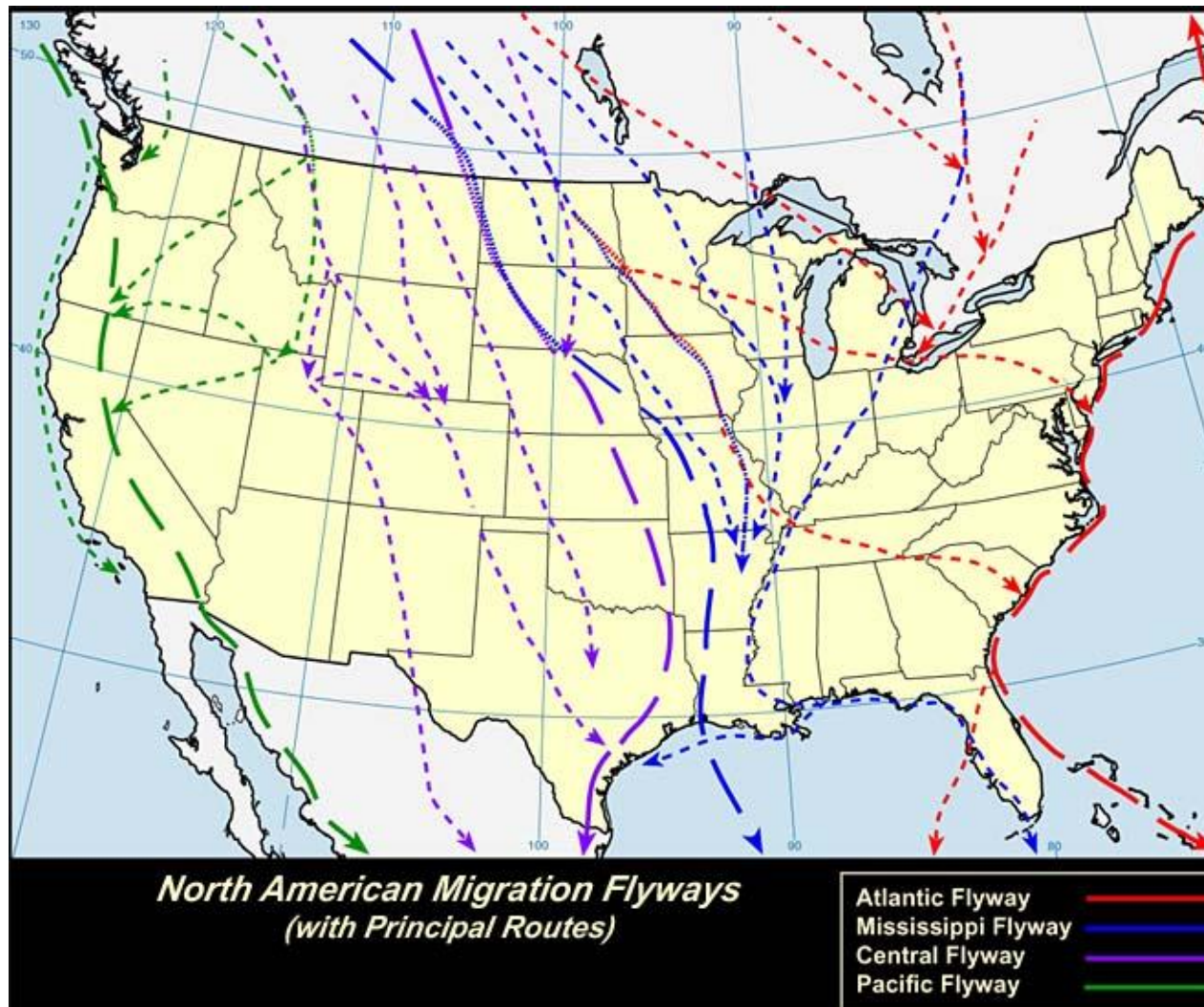
Protracted Outbreak of Salmonella Newport Infections Linked to Ground Beef: Possible Role of Dairy Cows -- 21 States, 2016-17

In January 2017, CDC identified a cluster of _Salmonella enterica_ serotype Newport infections with isolates sharing an indistinguishable pulsed-field gel electrophoresis (PFGE) pattern, JJPX01.0010 (pattern 10).

Epidemiologic Investigation: 106 cases were identified in 21 states. Most illnesses (72%) were reported from southwestern states, including AZ (30), CA (25), NM (14), and TX (7). Illness onset from 4 Oct 2016, through 19 Jul 2017. Among 65 interviewed patients, 52 (80%) reported eating ground beef at home in the week before illness began. This percentage was significantly higher than the 2006-2007 FoodNet Population Survey. Among the 52 patients who ate ground beef at home, 31 (60%) reported that they bought it or maybe bought it from multiple locations of 2 national grocery chains, and 21 (40%) reported that they bought ground beef from locations of 15 other grocery chains.

Traceback Investigation: USDA-FSIS conducted traceback on ground beef for 11 patients who provided shopper card records or receipts. Approx. 20 ground beef suppliers belonging to at least 10 corporations were identified; 10 of the 11 records traced back to 5 company A slaughter/processing establishments, 7 of 11 traced back to five company B slaughter/processing establishments, and 4 of 11 traced back to 2 company C slaughter/processing establishments.

Product and Animal Testing: Opened samples of ground beef from 3 patients' homes were collected. All were purchased from 1 of 2 national grocery chains that had been identified by a majority of patients. One sample, collected from ground beef removed from its original packaging, yielded the outbreak strain.

The outbreak strain was also isolated from four NM dairy cattle. One was collected from a spontaneously aborted fetus in July 2016, and one was isolated from feces from a young calf in November 2016. The 3rd isolate was identified by searching the USDA-APHIS NVSL database for Salmonella Newport isolates collected from cattle in AZ, CA, TX, NM, and WI; the only Newport pattern 10 isolate identified was from a fecal sample from a NM dairy cow collected during November 2016. The 4th isolate was from a USDA-FSIS routine cattle fecal sample collected at a TX slaughter establishment in 2016; FSIS determined the sample was from a dairy cow and identified the NM farm of origin. Officials were not able to identify the farm or farms of origin for the dairy cows associated with the other 3 samples or whether the 4 dairy cows were associated with a single farm.

Laboratory Investigation: SNP analysis showed that 106 clinical isolates were closely related to each other genetically, to the 4 dairy cattle isolates, and to the leftover ground beef isolate (range = 0-12 SNP differences), suggesting that the Salmonella bacteria found in patients, ground beef, and dairy cattle all shared a common source.

Public Health Response: Because the USDA-FSIS traceback investigation did not converge on a common production lot of ground beef or a single slaughter/processing establishment, and no ground beef in the original packaging yielded the outbreak strain, a recall of specific product was not requested. A public warning was not issued to consumers because specific, actionable information was not available (e.g., a specific brand or type of ground beef).

A hypothetical cases study

- July 2010: cheese from creamery X found positive for *L. monocytogenes* when a sample collected at a supermarket was tested by government lab
  - Isolates was subsequently characterized by WGS
- June 2017: Human isolates "matches" July 2010 isolate
  - Both isolates are DNI (**D**arn **N**ear **I**dentical); i.e., 3 SNP difference
- What if the person reports having eaten cheeses labeled "creamery X"? What if the person did not report eating cheese at all? etc.

Outline

- Why are we here?
- Background on tools: from PFGE to WGS
- **Case studies on outbreak and in-plant environmental investigations**
  - Case study 1: outbreak detection
  - **Case study 2: subtyping of food or environmental isolates**
- Metagenomics
- Other WGS and NGS applications in dairy

2000 US outbreak - Environmental persistence of *L. monocytogenes*

- 1988: one human listeriosis case linked to hot dogs produced by plant X
- 2000: 29 human listeriosis cases linked to sliced turkey meats from plant X

From the Centers for Disease Control and Prevention

Leads From the Morbidity and Mortality Weekly Report
Atlanta, Ga

Multistate Outbreak of *Salmonella* Serotype Agona Infections Linked to Toasted Oats Cereal—United States, April-May, 1998

209 cases

Information as of May 13, 2008 (FINAL Update)

Click Here for Advice to Consumers   **28 cases**

CDC is collaborating with public health officials in multiple states across the United States and with the U.S. Food and Drug Administration (FDA) to investigate a multi-state outbreak of *Salmonella* Agona infections. An investigation that includes interviews of persons with *Salmonella* Agona infections and comparison of the DNA fingerprints suggests that cereal from Malt-O-Meal unsweetened Puffed Rice Cereals and unsweetened Puffed Wheat Cereals is likely related to these illnesses.

Implications for food industry

- January 2015: cheese from facility A found positive for *Salmonella* when a sample collected at a supermarket was tested by government lab
  - Regulatory action?

Implications for food industry

- January 2015: cheese from facility A found positive for *Salmonella* when a sample collected at a supermarket was tested by government lab

- December 2017: environmental sample from facility A found positive for *Salmonella* when samples were collected and tested by a regional government lab
  - Regulatory action?

The real world

FDA re-inspected and re-sampled the SM Fish facility from August 15, 2016 to September 9, 2016 and learned that the firm's cleaning and sanitation procedures were unsuccessful in solving its environmental *Listeria* contamination. Testing results showed that *Listeria* was detected in 12 out of the 116 locations swabbed throughout the facility, including on a direct food contact surface. Other locations found to harbor the bacteria were non-food contact surfaces that are in sufficient proximity to the food and food contact surfaces to create an increased risk of contaminating the food, particularly considering inspection observations. Whole genome sequencing matched some of the *Listeria* findings genetically to samples collected during the June/July 2016 inspection, as well as to samples collected during the 2015 inspection, indicating that at least three strains of *Listeria* have been consistently present in this facility during a two-year period.

Case study

- FDA found *Listeria monocytogenes* in X environmental samples from facility A

Health > Pathogen Detection > Isolates Browser

| Find one or more isolates ... | 🔍 **Search** |

| Listeria monocytogenes | ▼ | 🗑 |

⌃ Filters ③ ✖

Location	☐ AK (2) ☐ AR (2) ☐ AZ (3) ☐ BRENHAM (2) ☐ CA (154) ☐ CO (13) ☐ CT (12) ☐ DE (12) ☐ FL (27) ☐ GA (5) ☐ HI (7) ☐ ID (15) ☐ IL (43) ☐ IN (10) ☐ KS (7) ☐ KY (30) ☐ LA (9) ☐ MA (26) ☐ MD (10) ☐ ME (18) ☐ MI (26) ☐ MN (6) ☐ MO (13) ☐ NH (15) ☐ NJ (11) ☐ NM (42) ☑ NY (109) ☐ OH (45) ☐ OR (5) ☐ PA (7) ☐ RI (21) ☐ SC (61) ☐ TN (24) ☐ TURKEY (1) ☐ TX (44) ☐ USA (1,081) ☐ UT (7) ☐ VA (13) ☐ VT (1) ☐ WA (125) ☐ WI (73) ⊟ less
Source	☐ basil, loose (1) ☐ basil, packaged (1) ☐ bianco, brie-style cheese (1) ☐ Cashew Butter (2) ☐ cheese (4) ☐ cilantro, loose (1) ☐ cold smoked salmon slices (1) ☐ cold smoked white fish (2) ☐ env. sponges (4) ☐ environmental sponges (10) ☐ Environmental (95) ☐ environmental samples (4) ☐ environmental sponge (20) ☐ environmental sponge samples (12) ☐ environmental sponge sticks (3) ☑ environmental swab (109) ☐ Environmental Swab (63) ☐ environmental swabs (2) ☐ Environmental Swabs (47) ☐ fresh cheese (5) ☐ frozen corn (3) ☐ frozen mixed vegetables (1) ☐ frozen oriental blend vegetables (2) ☐ gouda cheese wheel (2) ☐ grated cheese (2) ☐ patient (1) ☐ queso fresco (1) ☐ queso fresco cheese (1) ☐ Raw Milk Cheese (2) ☐ ricotta cheese (1) ☐ salmon trimmings (1) ☐ scallion cream cheese (1) ☐ smoked salmon (1) ☐ smoked salmon, vacuum packed (1) ☐ swab (21) ☐ Swab (2) ☐ vaccum packed pickled herring (1) ⊟ less
Collected by	☑ FDA (109) ☐ NY (4)
Host	
Property	☐ has AMR genotypes (109)
Target Creation	
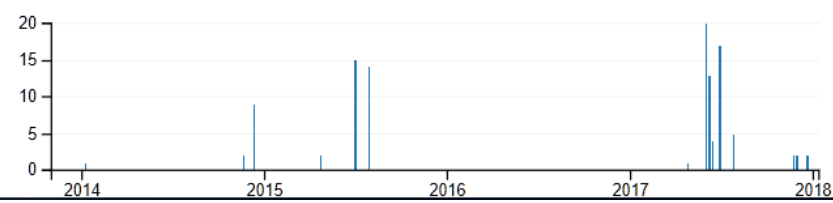
11	Listeria monocytogenes	FDA00011944	PDT000221089.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014208.1	19	n/a	SAMN0725547	PDG000000001.829	fosX lin	2017-06-15	SRR5758415
12	Listeria monocytogenes	FDA00011946	PDT000221090.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014208.1	16	n/a	SAMN0725547	PDG000000001.829	fosX lin	2017-06-15	SRR5758416
13	Listeria monocytogenes	FDA00011950	PDT000221088.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014209.1	4	n/a	SAMN0725546	PDG000000001.829	fosX lin	2017-06-15	SRR5758414
14	Listeria monocytogenes	FDA00011935	PDT000221079.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014208.1	13	n/a	SAMN0725548	PDG000000001.829	fosX lin	2017-06-14	SRR5758405
15	Listeria monocytogenes	FDA00011937	PDT000221081.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014208.1	41	n/a	SAMN0725548	PDG000000001.829	fosX lin	2017-06-14	SRR5758407
16	Listeria monocytogenes	FDA00011938	PDT000221085.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014208.1	13	n/a	SAMN0725547	PDG000000001.829	fosX lin	2017-06-14	SRR5758411
17	Listeria monocytogenes	FDA00011942	PDT000221076.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014210.1	17	n/a	SAMN0725547	PDG000000001.829	fosX lin	2017-06-14	SRR5758402
18	Listeria monocytogenes	FDA00011934	PDT000221083.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014210.1	4	n/a	SAMN0725548	PDG000000001.829	fosX lin	2017-06-14	SRR5758409
19	Listeria monocytogenes	FDA00011943	PDT000221084.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014210.1	4	n/a	SAMN0725547	PDG000000001.829	fosX lin	2017-06-14	SRR5758410
20	Listeria monocytogenes	FDA00011940	PDT000221077.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014209.1	4	n/a	SAMN0725547	PDG000000001.829	fosX lin	2017-06-14	SRR5758403
21	Listeria monocytogenes	FDA00011936	PDT000221082.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014209.1	11	n/a	SAMN0725548	PDG000000001.829	fosX lin	2017-06-14	SRR5758408
22	Listeria monocytogenes	FDA00011933	PDT000221112.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014209.1	9	n/a	SAMN0725548	PDG000000001.829	fosX lin	2017-06-14	SRR5758447
23	Listeria monocytogenes	FDA00011941	PDT000221078.1	2017-06-27	USA:NY	environmental swab	environmental/other		n/a	n/a	SAMN0725547	PDG000000001.829	fosX lin	2017-06-14	SRR5758404
24	Listeria monocytogenes	FDA00011939	PDT000221080.1	2017-06-27	USA:NY	environmental swab	environmental/other		n/a	n/a	SAMN0725547	PDG000000001.829	fosX lin	2017-06-14	SRR5758406
25	Listeria monocytogenes	FDA00011930	PDT000221106.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014210.1	37	n/a	SAMN0725548	PDG000000001.829	fosX lin	2017-06-13	SRR5758440
26	Listeria monocytogenes	FDA00011931	PDT000221113.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014210.1	49	n/a	SAMN0725548	PDG000000001.829	fosX lin	2017-06-13	SRR5758448
27	Listeria monocytogenes	FDA00011932	PDT000221114.1	2017-06-27	USA:NY	environmental swab	environmental/other	PDS000014210.1	40	n/a	SAMN0725548	PDG000000001.829	fosX lin	2017-06-13	SRR5758482

Organism	Strain	Isolate	Create date	Isolation source	Isolation type	SNP cluster	Minimum SNP distance within same source type	Minimum SNP difference across different source types
Listeria monocytogenes	FDA00011944	PDT000221089.1	6/27/2017	environmental swab	environmental/other	PDS000014208.1	19	n/a
Listeria monocytogenes	FDA00011946	PDT000221090.1	6/27/2017	environmental swab	environmental/other	PDS000014208.1	16	n/a
Listeria monocytogenes	FDA00011935	PDT000221079.1	6/27/2017	environmental swab	environmental/other	PDS000014208.1	13	n/a
Listeria monocytogenes	FDA00011937	PDT000221081.1	6/27/2017	environmental swab	environmental/other	PDS000014208.1	41	n/a
Listeria monocytogenes	FDA00011938	PDT000221085.1	6/27/2017	environmental swab	environmental/other	PDS000014208.1	13	n/a
Listeria monocytogenes	FDA00011950	PDT000221088.1	6/27/2017	environmental swab	environmental/other	PDS000014209.1	4	n/a
Listeria monocytogenes	FDA00011940	PDT000221077.1	6/27/2017	environmental swab	environmental/other	PDS000014209.1	4	n/a
Listeria monocytogenes	FDA00011936	PDT000221082.1	6/27/2017	environmental swab	environmental/other	PDS000014209.1	11	n/a
Listeria monocytogenes	FDA00011933	PDT000221112.1	6/27/2017	environmental swab	environmental/other	PDS000014209.1	9	n/a
Listeria monocytogenes	FDA00011942	PDT000221076.1	6/27/2017	environmental swab	environmental/other	PDS000014210.1	17	n/a
Listeria monocytogenes	FDA00011934	PDT000221083.1	6/27/2017	environmental swab	environmental/other	PDS000014210.1	4	n/a
Listeria monocytogenes	FDA00011943	PDT000221084.1	6/27/2017	environmental swab	environmental/other	PDS000014210.1	4	n/a
Listeria monocytogenes	FDA00011930	PDT000221106.1	6/27/2017	environmental swab	environmental/other	PDS000014210.1	37	n/a
Listeria monocytogenes	FDA00011931	PDT000221113.1	6/27/2017	environmental swab	environmental/other	PDS000014210.1	49	n/a
Listeria monocytogenes	FDA00011932	PDT000221114.1	6/27/2017	environmental swab	environmental/other	PDS000014210.1	40	n/a
Listeria monocytogenes	FDA00011941	PDT000221078.1	6/27/2017	environmental swab	environmental/other		n/a	n/a
Listeria monocytogenes	FDA00011939	PDT000221080.1	6/27/2017	environmental swab	environmental/other		n/a	n/a

What if governments find a link with our business? How can we prepare ourselves?

- Do a Mock outbreak investigation for every plan yearly (not just a mock recall)

- Have data available that show that you have validated cleaning and sanitation procedures, so that you can limit recalls

- Have access to different subtyping tools, so you can proof you identified the root cause

  - In many cases you will need to show, with subtyping evidence, that you found the root cause of an issue to allow you to re-start a facility

Outline

- Why are we here?
- Background on tools: from PFGE to WGS
- Case studies on WGS use for outbreak and in-plant environmental investigations
- **Metagenomics**
- Other WGS and NGS applications in dairy

Metagenomcis

- Metagenomics is the characterization of all genetic material in a sample (for example milk or cheese sample)
    - 16S metagenomics: only sequencing of 16S rDNA
        - Easy to do and specific for bacteria
        - Limited discriminatory power (e.g., can't differentiates *L. monocytogenes* and *L. innocua*)
    - Shutgun metagenomics: sequences all DNA in a sample
        - May detect DNA from dead organisms
        - Interference in samples with large amounts of host DNA (e.g., milk)
- Allows for comprehensive characterization that can be used for QA and troubleshooting
    - Already regularly used to identify causes of quality issues
    - Considerable potential for monitoring of fermentations
- Increasingly used and applied by US FDA, regulatory agencies worldwide, and industry
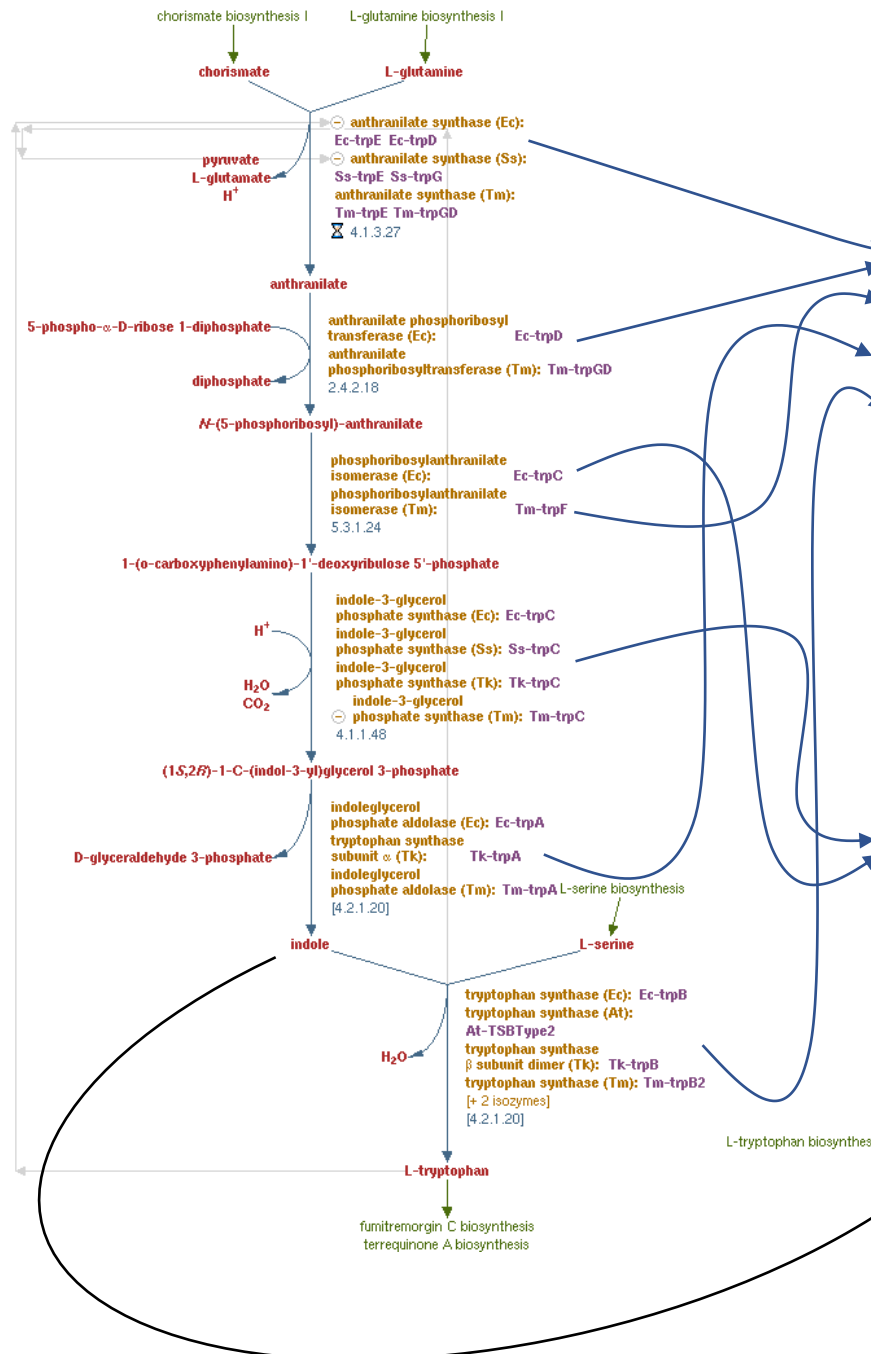
IBM: Sequencing the Food Supply Chain

http://www.research.ibm.com/client-programs/foodsafety/
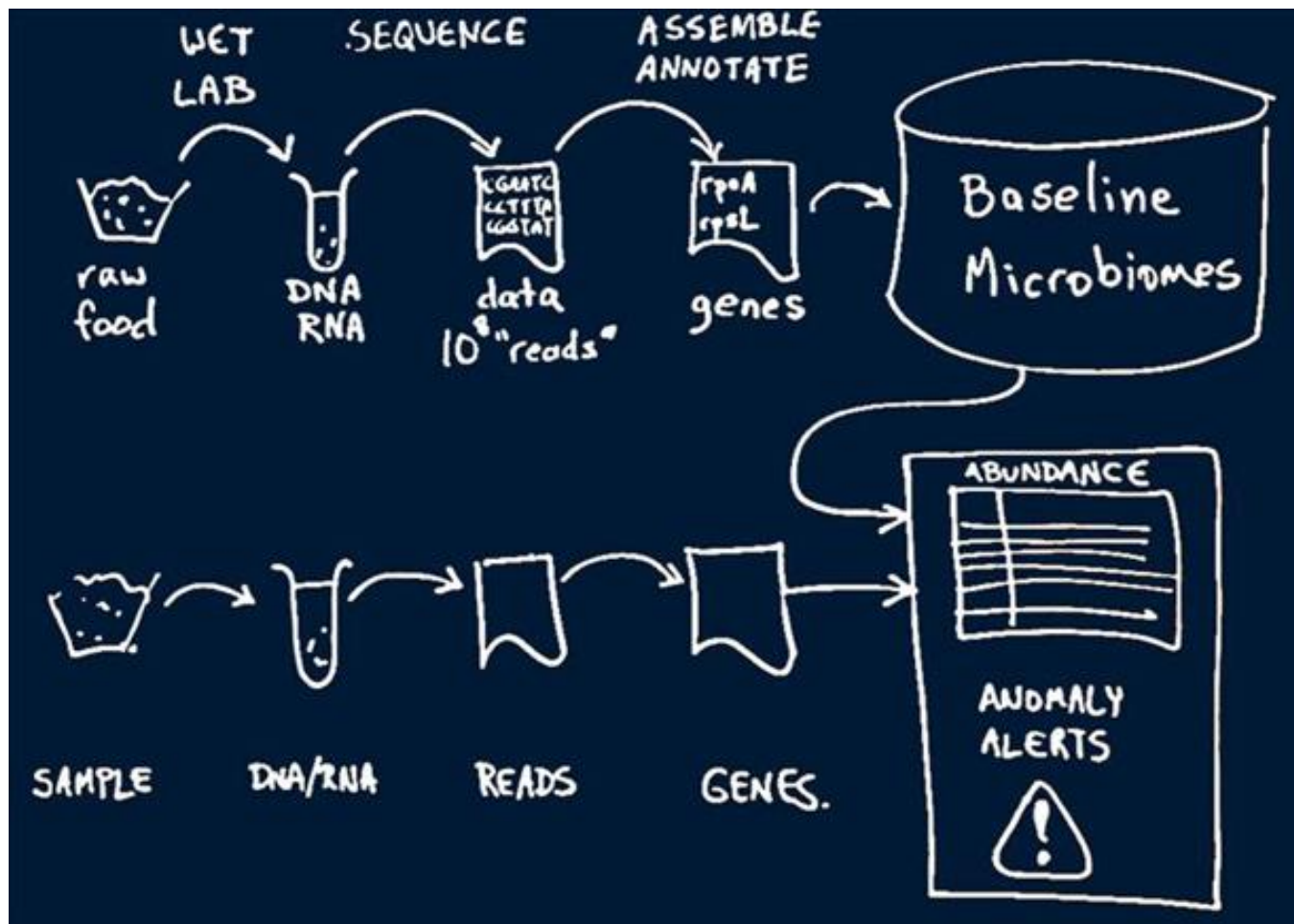
50 shades of gray: *Pseudomonas* causes gray discoloration in HTST milk

Key gene unique to Pseudomonas causing gray (and blue) color defects

| Gene name	Annotation
trpD	Anthranilate phosphoribosyltransferase
trpF	N-(5'-phosphoribosyl)anthranilate isomerase
trpA	Tryptophan synthase alpha chain
trpB	Tryptophan synthase beta chain
iolG	Inositol 2-dehydrogenase/D-chiro-inositol 3-dehydrogenase
mdh	Malate dehydrogenase
	Peptidase M
	Xylose Isomerase Domain-Containing Protein
degT	Pleiotropic regulatory protein
	Oxidoreductase Domain Protein
trpC	Indole-3-glycerol phosphate synthase
solR	Transcriptional activator protein solR
rhtB	Homoserine/homoserine lactone efflux protein
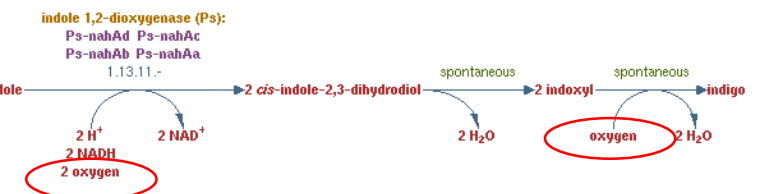
This reaction requires 3 oxygens!!

Could this be the future

- Auditor conducts audit of supplier facility X; collects 5 samples of ingredient (e.g., pepper), which are subsequently characterized by metagenomic analysis

- Subsequently, metagenomic analysis is conducted, at a risk-based frequency, of lots received at customer

- Lots that show "substantial" deviation will trigger further analyses (and possibly other actions?)
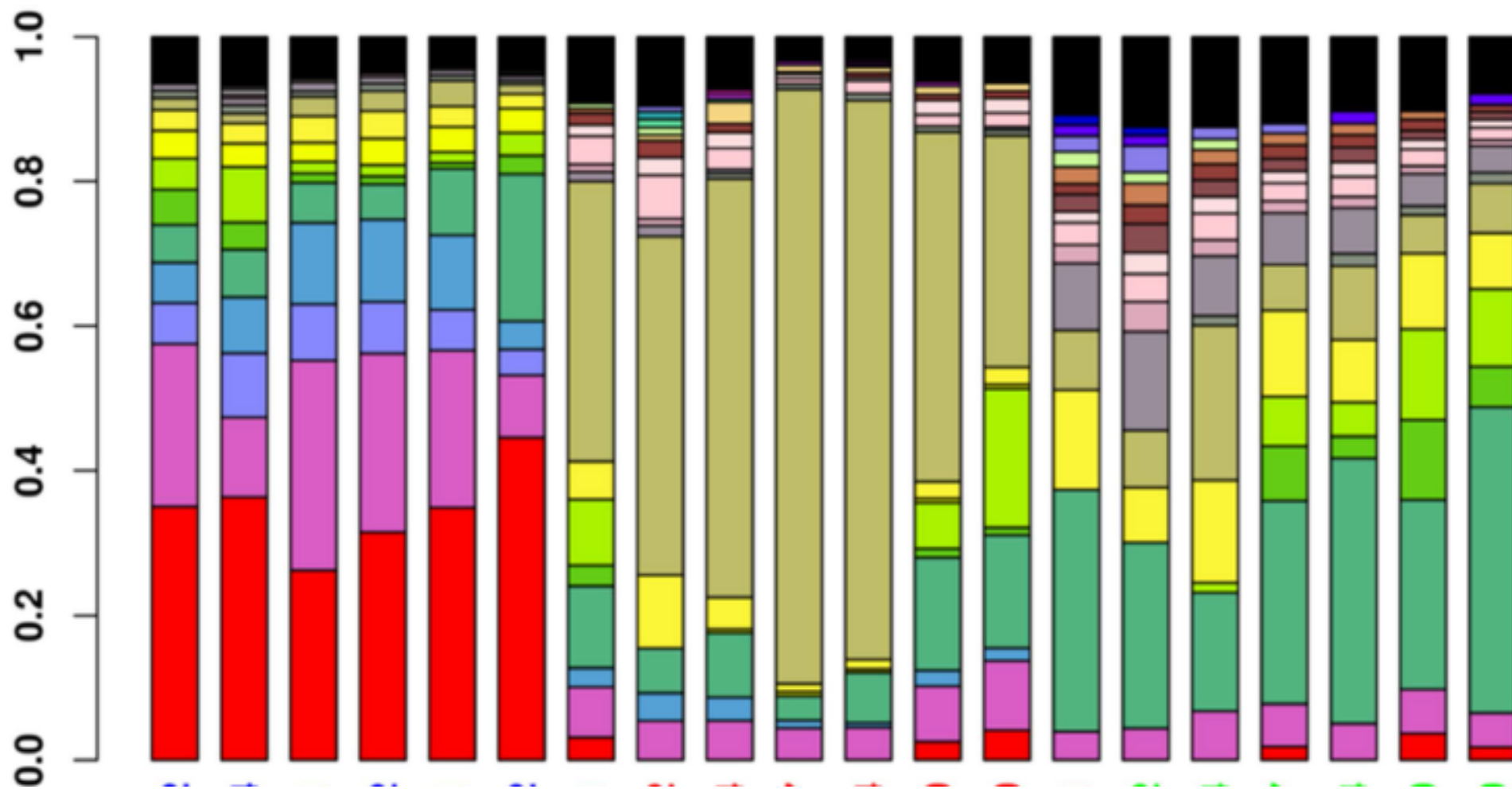
Outline

- Why are we here?
- Background on tools: from PFGE to WGS
- Case studies on WGS use for outbreak and in-plant environmental investigations
- Metagenomcis
- **Other WGS and NGS applications in dairy**

Improved characterization of foodborne pathogens

- WGS data will replace a suite of previous tests as WGS can be used to predict:
  - Serotype
  - Antimicrobial resistance
  - Presence of gene that will allow organisms to cause disease
- Impact includes more rapid identification of unusual and difficult to identify bacteria

Consequences of unreliable differentiation between pathogens and non-pathogens can be costly

The New Zealand Herald

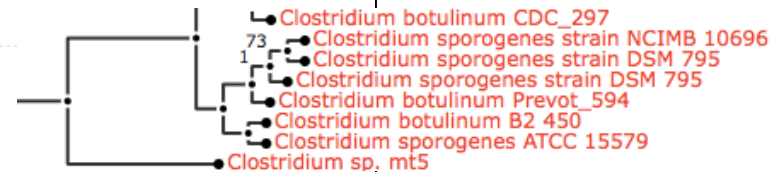**Fonterra botulism crisis was false alarm**

By Jamie Gray

Fonterra CEO Theo Spierings. Photo / Richard Robinson

Fonterra chief executive Theo Spierings defended the company's multimillion-dollar recall of dairy product thought to contain a botulism-causing bacteria, despite it proving to be a false alarm.

The Ministry for Primary Industries said it had received results confirming that the bacteria found in the whey protein concentrate (WPC) manufactured by Fonterra was not the botulism-causing clostridium botulinum.

As it turned out, the organism was confirmed as clostridium sporogenes, which is not capable of producing botulism causing toxins, the ministry said.

Clostridium botulinum CDC_297
73 — Clostridium sporogenes strain NCIMB 10696
1 — Clostridium sporogenes strain DSM 795
Clostridium sporogenes strain DSM 795
Clostridium botulinum Prevot_594
Clostridium botulinum B2 450
Clostridium sporogenes ATCC 15579
Clostridium sp. mt5

- Suspected botulism-causing bacteria identified in whey products
- ~1,000 tones of products recalled in 7 countries
- No disease cases
- Detailed strain characterization confirmed **species misclassification**

INTERNATIONAL JOURNAL OF **SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY**

Clostridium tepidum sp. nov., a close relative of *Clostridium sporogenes* and *Clostridium botulinum* Group I

Anatoly P. Dobritsa,[1,*] Kirthi K. Kutumbaka,[1] Kirsten Werner,[1] Martin Wiedmann,[2] Aaron Asmus[3] and Mansour Samadpour[1]

Abstract

Obligately anaerobic, Gram-stain-positive, spore-forming bacteria indistinguishable by pulsed-field gel electrophoresis were isolated from non-dairy protein shakes in bloated bottles. One of the isolates, strain IEH 97212[T], was selected for further study. The strain was closely related to *Clostridium sporogenes* and *Clostridium botulinum* Group 1 based on 16S rRNA gene sequence similarities. Phylogenetic analysis also showed that strain IEH 97212[T] and strain PE (=DSM 18688),

Standard genetic differentiation among *B. cereus* group species is not reliable

Miller et al., 2016

RESEARCH ARTICLE

Open Access

CrossMark
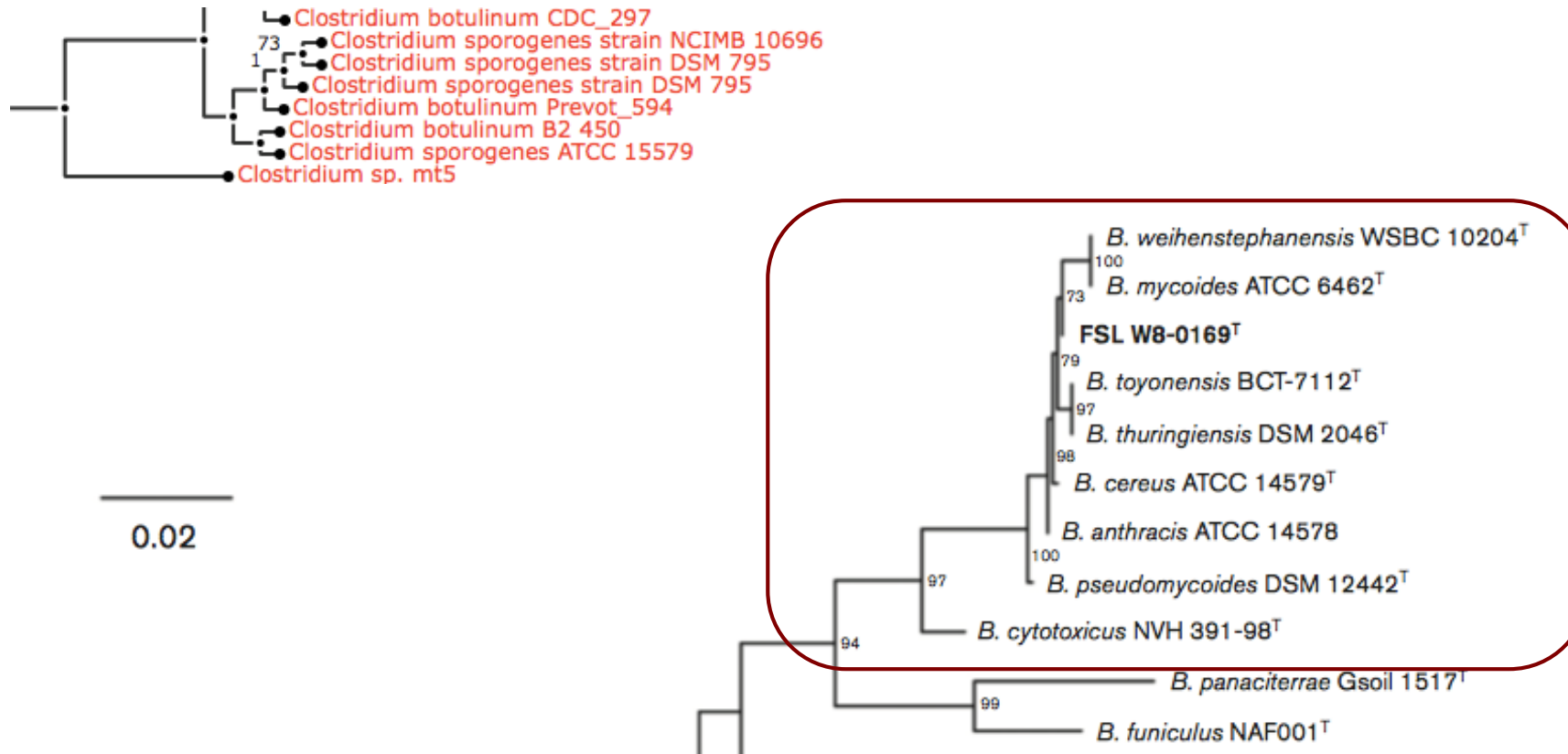
Production of hemolysin BL by *Bacillus cereus* group isolates of dairy origin is associated with whole-genome phylogenetic clade

Jasna Kovac, Rachel A. Miller, Laura M. Carroll, David J. Kent, Jiahui Jian, Sarah M. Beno and Martin Wiedmann[*]

AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

Rapid, High-Throughput Identification of Anthrax-Causing and Emetic *Bacillus cereus* Group Genome Assemblies via BTyper, a Computational Tool for Virulence-Based Classification of *Bacillus cereus* Group Isolates by Using Nucleotide Sequencing Data

Laura M. Carroll, Jasna Kovac, Rachel A. Miller, Martin Wiedmann
Department of Food Science, Cornell University, Ithaca, New York, USA

Challenges – a few examples

Report: Human DNA found in hot dogs

After analyzing hot dogs from 75 different brands, Clear Foods discovered human DNA in 2% of the samples studied. Rachel Holt (@ItsRachelHolt) dishes what else the study found. Buzz60

DNA blunder creates phantom serial killer

Police admit they wasted 15 years hunting for the 'Woman Without a Face'

She was a mysterious serial killer known as the "The Woman Without a Face" and detectives across Europe spent more than 15 years doing their utmost to bring her to justice for at least six brutal murders and a string of break-ins. Yesterday, however, they were forced to admit that she probably didn't exist.

The only clues that "The Woman Without a Face" left behind at 40 different crime scenes were DNA traces. These were collected on cotton swabs, supplied to the police in a number of European countries. Now police investigators have established that in all probability the DNA had not been left by their quarry but by a woman working for the German medical company supplying the swabs, who had inadvertently contaminated them.

German police who had been leading the hunt said they had probably been involved in one of the longest and most perplexing wild goose chases in criminal history. "This is a very embarrassing story," admitted police spokesman Josef Schneider.

Take home messages

- WGS and other genomics tools will be and already are "game changers" in many areas of microbial food safety and quality, including
  - Detection of more and smaller foodborne disease outbreaks
  - Better identification of pathogen and microbial persistence (and "unhygienic conditions")
  - More rapid detection of new pathogens
  - Improved characterization of (potential) foodborne pathogen isolates
  - *Metagenomics* approaches for QA, fraud detection, identification of spoilage root causes etc.
- WGS and other genomics tools are not magic bullets
  - There will be a continued need for (good) epidemiology to identify outbreak sources
    - Also tremendous need for understanding of potentially complex contemporary supply chains

Concluding thoughts on (industry) needs

- Have access to expertise in WGS and metagenomics
  - Need to have experts that understand the biology
- Assess your vulnerabilities
- Be prepared to see more outbreaks traced back to source
  - raw meat
  - Salmonella Enteritidis
- Make sure you have the data to keep recalls small
- We need some sort of safe harbor that will encourage industry use of WGS

Slides for detailed presentations

Analysis of genome wide SNPs (wgSNPs)

- Identifies all high confidence SNPs over whole genome (approx. 3 to 5 million nucleotides)
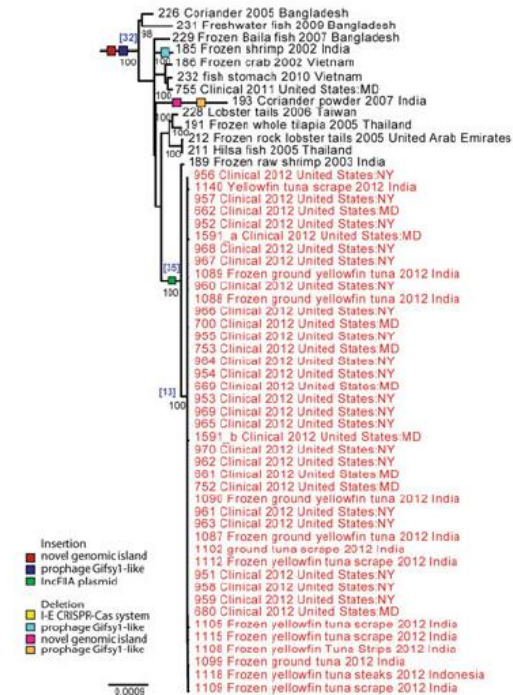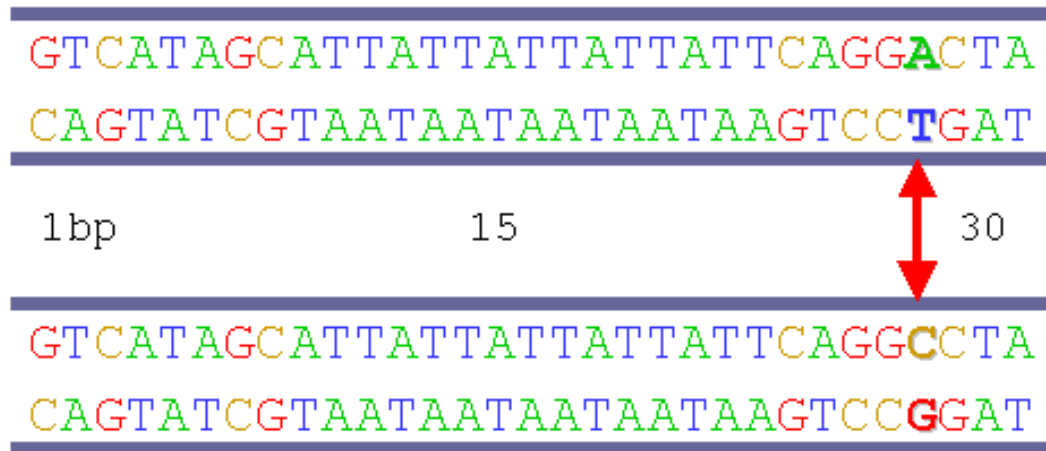
Whole genome multilocus sequence typing (wgMLST)

- Allows for simpler analysis and clear naming of subtypes
- Performs comparison on a gene by gene level

	Isolate A	Isolate B	Isolate C
Gene 1	1	1	1
Gene 2	8	8	12
Gene 3	5	5	2
Etc.			
Gene 1,005	4	4	4
wgMLST type	**A**	**A**	**B**

Microbial evolution 101 – mechanisms of change

Insertion or deletion ("indel")

ACCCTCTAGTAGTAGCA

ACCATCTAG . . . TAGCA   ACCCTCTAGTAGTAGTAGCA

3 differences (?) and one "genetic event"

What is a SNP?

Single Nucleotide Polymorphism (SNP)

```
ATGTTCCTC sequence
ATGTTGCTC reference
```

*phylogentically informative differences

Insertion or Deletion (Indel)

```
ATGTTCCCTC sequence
ATGTTC-CTC reference
```

*differences not used in high quality SNP (hqSNP) analysis

Where to call a SNP?

- Not all SNP pipelines are equal – where you call SNPs will affect the total SNP count

- SNPs relevant for phylogenetic analysis are vertically transmitted, not horizontally, so horizontal genetic elements like phages can be masked
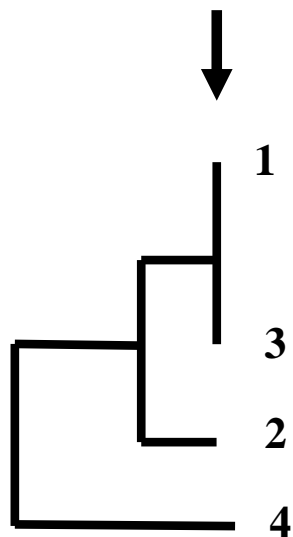
Mask mobile elements

-do no consider SNPs in this location

Only call SNPs in genes