

Microbial Genome Sequencing For Food Safety

Dr. Matthew Stasiewicz

mstasie@Illinois.edu

4/4/2019 - 1:15 pm

Core Questions

- Why should we care about genome sequencing?
 - Federal agencies now detect an unprecedented number outbreaks
 - Can link food/environmental isolates to clinical isolates without prior knowledge
- Will this really impact me/my company?
 - It could. 'Bad luck' is becoming more common
- Can a non-expert actually use genome sequencing?
 - Yes. Critical steps are getting automated.
- What (positive) impacts might WGS have on the food industry?
 - Improved environmental monitoring and pathogen profiling

Why should we care about WGS?

Federal agencies can detect an unprecedented number of outbreaks

How: Nearly perfect resolution for DNA-based subtyping

Who: US Federal Agencies

Non-Regulatory

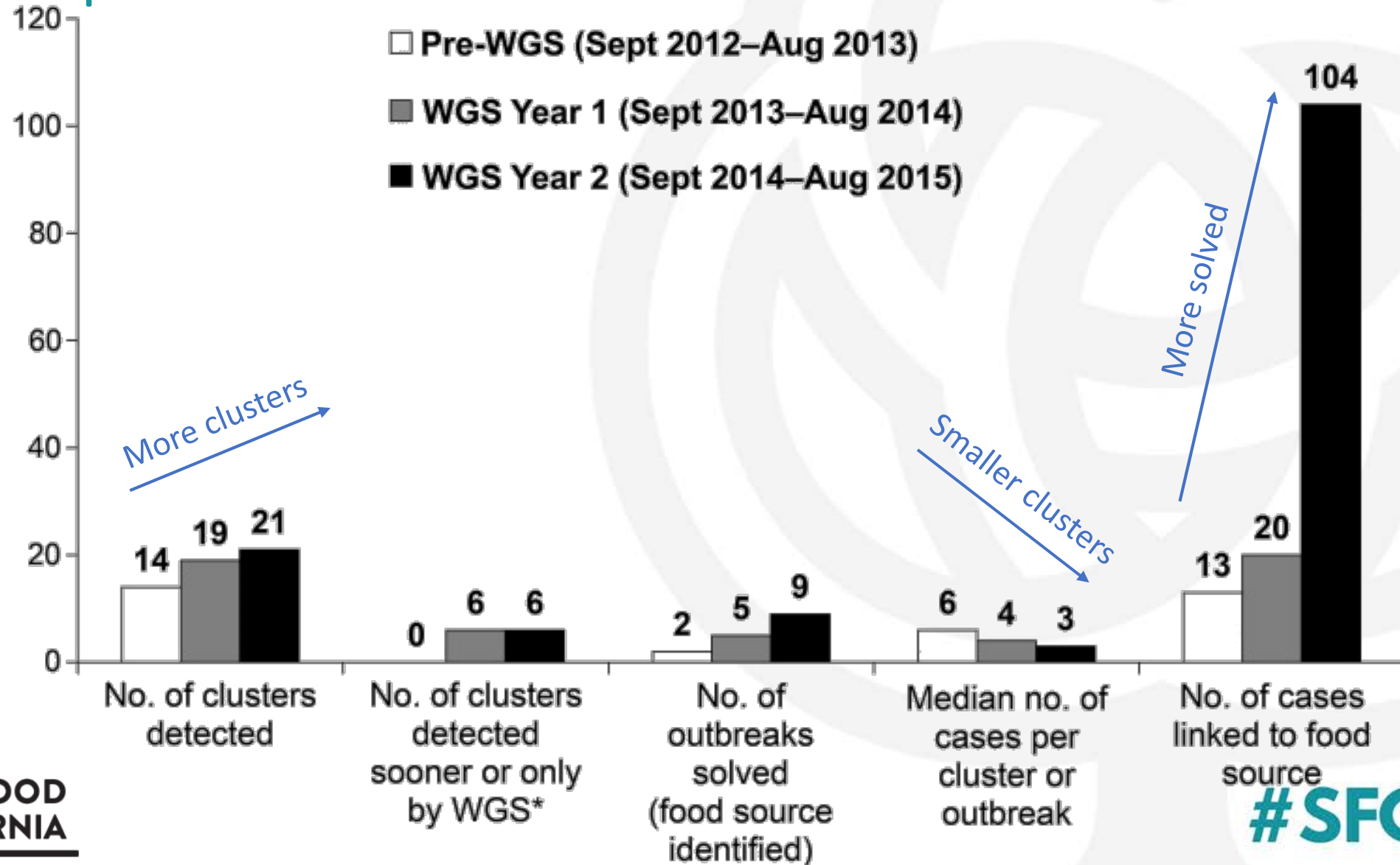
- CDC
 - Sequences clinical isolates
 - Disease surveillance
 - Outbreak detection and investigation

Began 2013
Now almost ALL multi-state FBD outbreak investigations use genome sequencing

Regulatory

- FDA & USDA (FSIS)
 - Sequences food & environmental isolates
 - Inspection
 - Enforcement
 - Investigation
 - Recall / Traceback

WGS improved listeriosis outbreak detection



Example Outbreaks

- Improved Outbreak Detection
 - Link distinct bacterial strains
 - Link historic isolates
 - Link isolates from routine inspections

Cheese: Multi-Strain

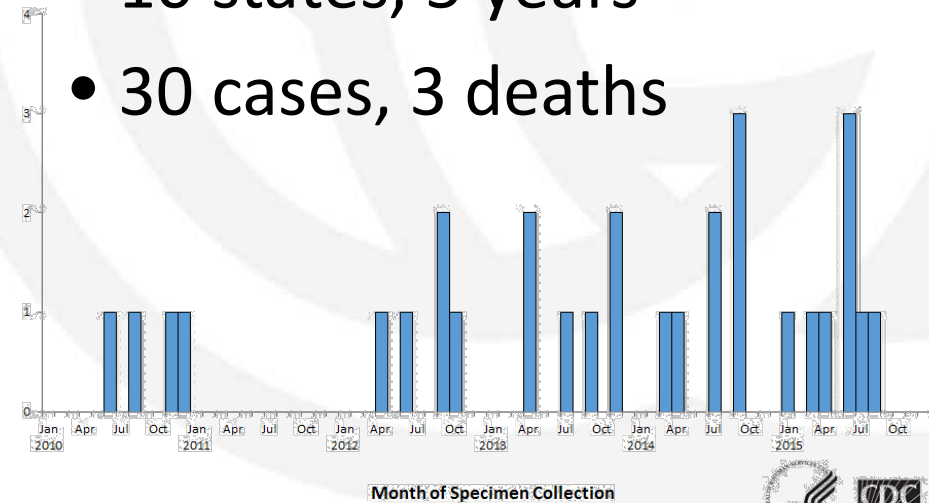
Improved outbreak detection

- Resolve multi-strain outbreaks



Soft cheese, Sep. 2015

- 5 *L. mono* strains closely related by WGS
- 10 states, 5 years
- 30 cases, 3 deaths



#SFC2019

Cheese: Multi-Strain

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

Posted September 23, 2015 3:45 PM ET

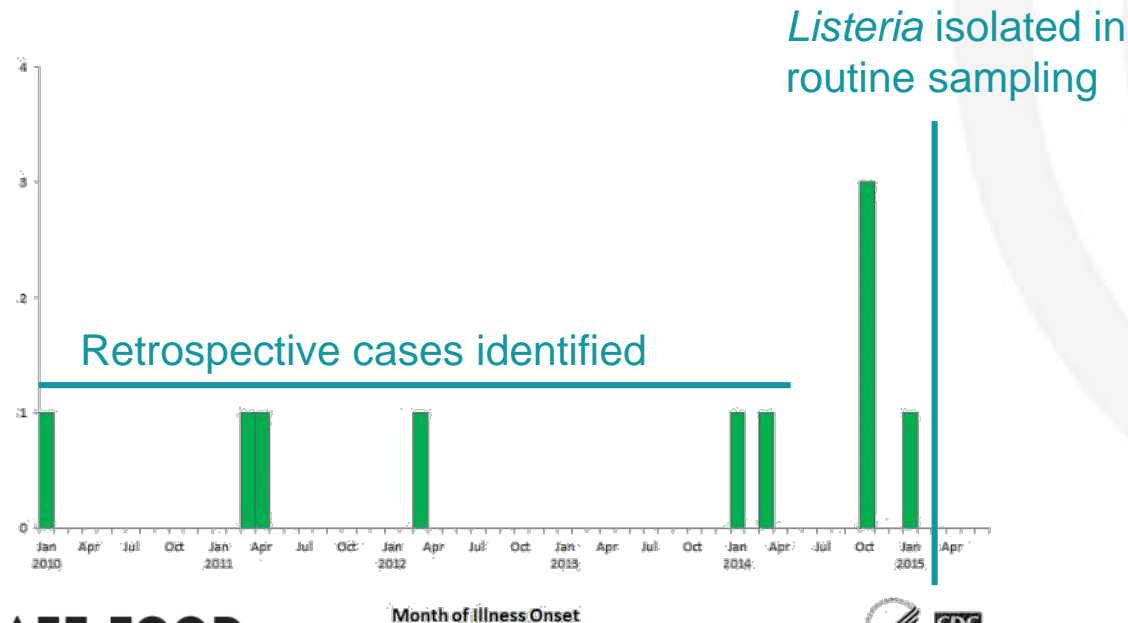
- 24 people infected with one of the closely related *Listeria* strains have been reported from 9 states since August 8, 2010.
 - 22 people were hospitalized. Five illnesses were pregnancy-related; one resulted in a fetal loss. One death was reported from Ohio.
 - FDA isolated *Listeria monocytogenes*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

Ice Cream: Routine Inspection & Retrospective

Improved outbreak detection

- Link routine inspection isolates to human illness



Ice-cream, 2015



Why Should We Care?

- Powerful Tool for Public Health
 - Improved outbreak detection
 - Links isolates from
 - Distinct strains
 - Historic cases
 - Routine Inspection

Will this really impact me/my company?

It could. 'Bad luck' is becoming more common

Dole and Listeriosis

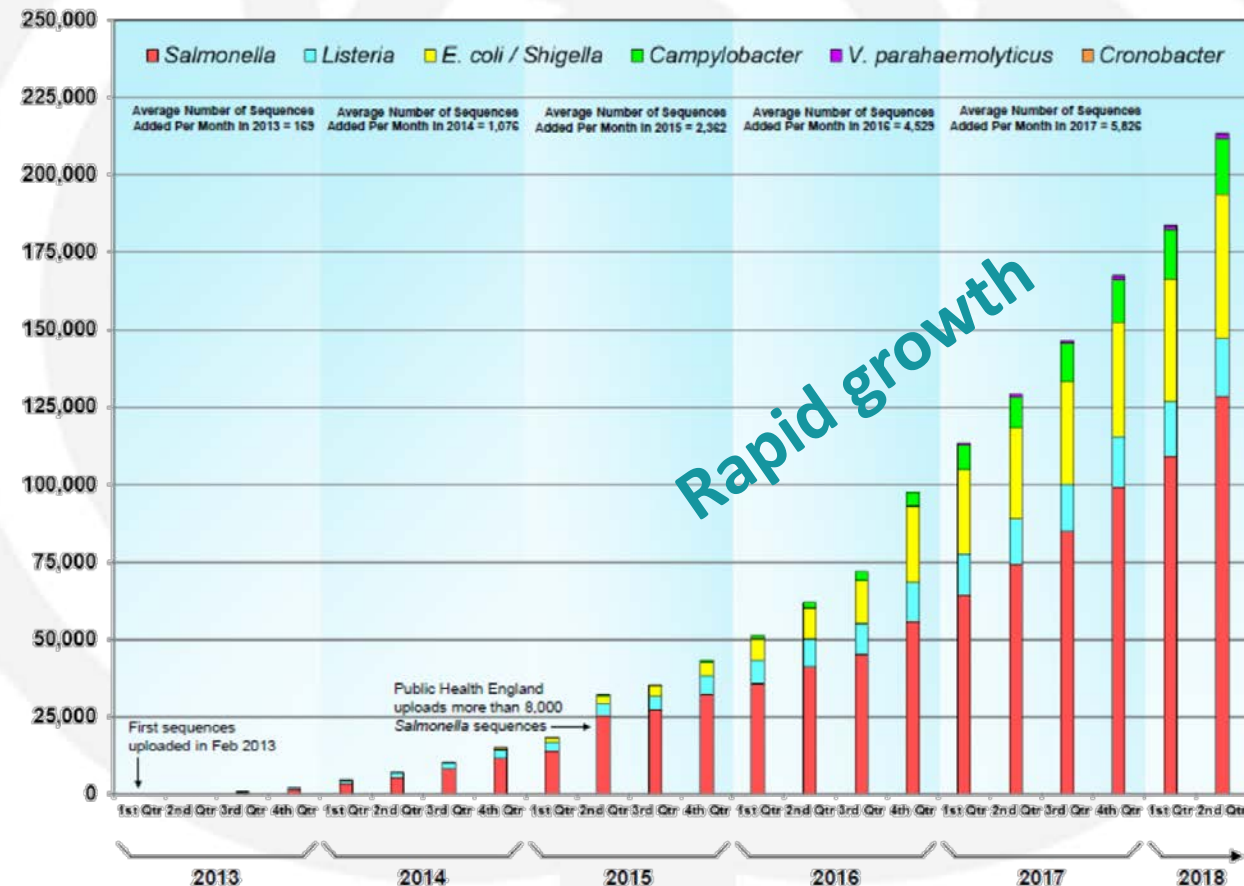
- Multistate (& Country) Outbreak
 - Packaged Salads; Springfield OH
 - 19 ill & hospitalized (+3 Can.), 1 death. 9 states
- Routine sampling link
 - Ohio Dpt. of Ag. retail salad -> clinical match
- Costs:
 - Shut down plant ~ 4 months
 - Criminal investigation
 - \$25.5 mil recall & closure
 - 2016 revenue down \$100 mil (of \$4.5 billion)



Why should we care?

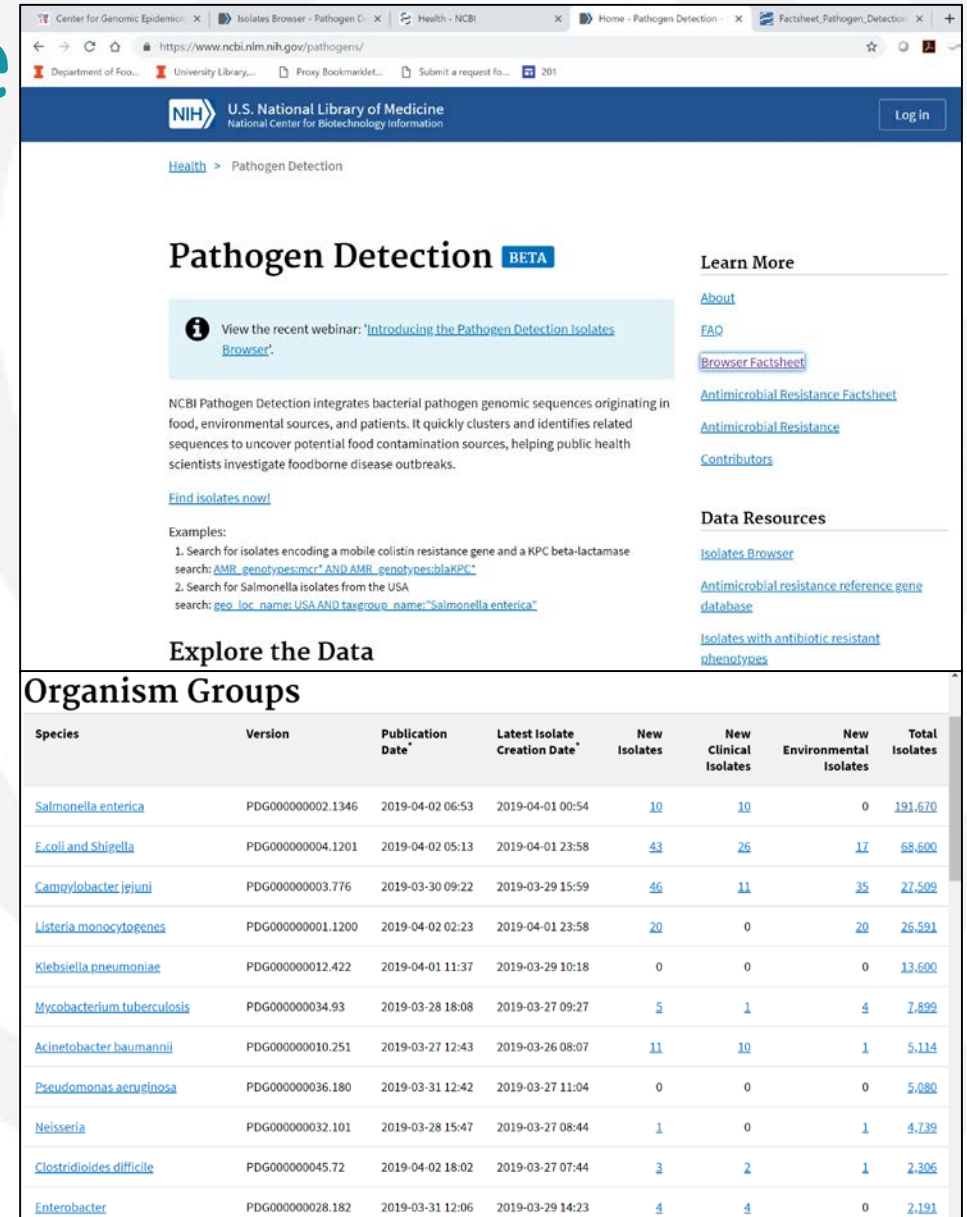
- Rapidly expanding database of clinical, food, and environmental isolates
 - US and International
 - Monitored for potential outbreaks
- Smaller outbreaks are being detected
 - Therefore smaller contamination events represent an enterprise risk
- **Public data**

Total Number of Sequences in the GenomeTrakr Database



Publicly Available Database and Analysis Tool

- NCBI Pathogen Detection Portal
- Capabilities
 - Analyzes short read sequence data
 - Reference-base and *de novo* analysis
 - SNP comparison
 - Cluster: < 50 SNPs
- Updated daily



The screenshot shows the NCBI Pathogen Detection portal. The main heading is "Pathogen Detection BETA". Below it, there is a description of the tool and a list of "Data Resources" including "Isolates Browser", "Antimicrobial resistance reference gene database", and "Isolates with antibiotic resistant phenotypes". The "Organism Groups" table is displayed below, listing various bacterial species with their respective version, publication date, latest isolate creation date, and counts for new isolates, clinical isolates, environmental isolates, and total isolates.

Species	Version	Publication Date*	Latest Isolate Creation Date*	New Isolates	New Clinical Isolates	New Environmental Isolates	Total Isolates
Salmonella enterica	PDG000000002.1346	2019-04-02 06:53	2019-04-01 00:54	10	10	0	191,670
E.coli and Shigella	PDG000000004.1201	2019-04-02 05:13	2019-04-01 23:58	43	26	17	68,600
Campylobacter jejuni	PDG000000003.776	2019-03-30 09:22	2019-03-29 15:59	46	11	35	27,509
Listeria monocytogenes	PDG000000001.1200	2019-04-02 02:23	2019-04-01 23:58	20	0	20	26,591
Klebsiella pneumoniae	PDG000000012.422	2019-04-01 11:37	2019-03-29 10:18	0	0	0	13,600
Mycobacterium tuberculosis	PDG000000034.93	2019-03-28 18:08	2019-03-27 09:27	5	1	4	7,899
Acinetobacter baumannii	PDG000000010.251	2019-03-27 12:43	2019-03-26 08:07	11	10	1	5,114
Pseudomonas aeruginosa	PDG000000036.180	2019-03-31 12:42	2019-03-27 11:04	0	0	0	5,080
Neisseria	PDG000000032.101	2019-03-28 15:47	2019-03-27 08:44	1	0	1	4,739
Clostridioides difficile	PDG000000045.72	2019-04-02 18:02	2019-03-27 07:44	3	2	1	2,306
Enterobacter	PDG000000028.182	2019-03-31 12:06	2019-03-29 14:23	4	4	0	2,191

I'll show two examples

From <<https://www.ncbi.nlm.nih.gov/pathogens/about/>>

Ice Cream Outbreak: Genome in Database

- Remember that outbreak?

NCBI Resources How To

BioSample BioSample Advanced

Display Settings: Full Send to:

Pathogen: environmental/food/other sample from *Listeria monocytogenes*

Identifiers BioSample: SAMN03463710; Sample name: CFSAN030685; SRA: SRS896150

Organism [Listeria monocytogenes](#)
cellular organisms; Bacteria; Firmicutes; Bacilli; Bacillales; Listeriaceae; Listeria

Package [Pathogen: environmental/food/other: version 1.0](#)

Attributes

strain	CFSAN030685
collected by	Kansas Department of Agriculture
collection date	2015-03-20
isolation source	chocolate ice cream
geographic location	USA:KS
latitude and longitude	missing
host disease	missing
isolate name alias	KS-15037-10

BioProject [PRJNA215355](#) *Listeria monocytogenes*
Retrieve [all samples](#) from this project

Submission [CFSAN](#), Justin Payne; 2015-04-07

If you know where to look...
Agencies do

Real-time addition of **new** isolates

Filters

Columns

Selected : 1

Download

Page 1 of 4

20

View 1 - 20 of 61

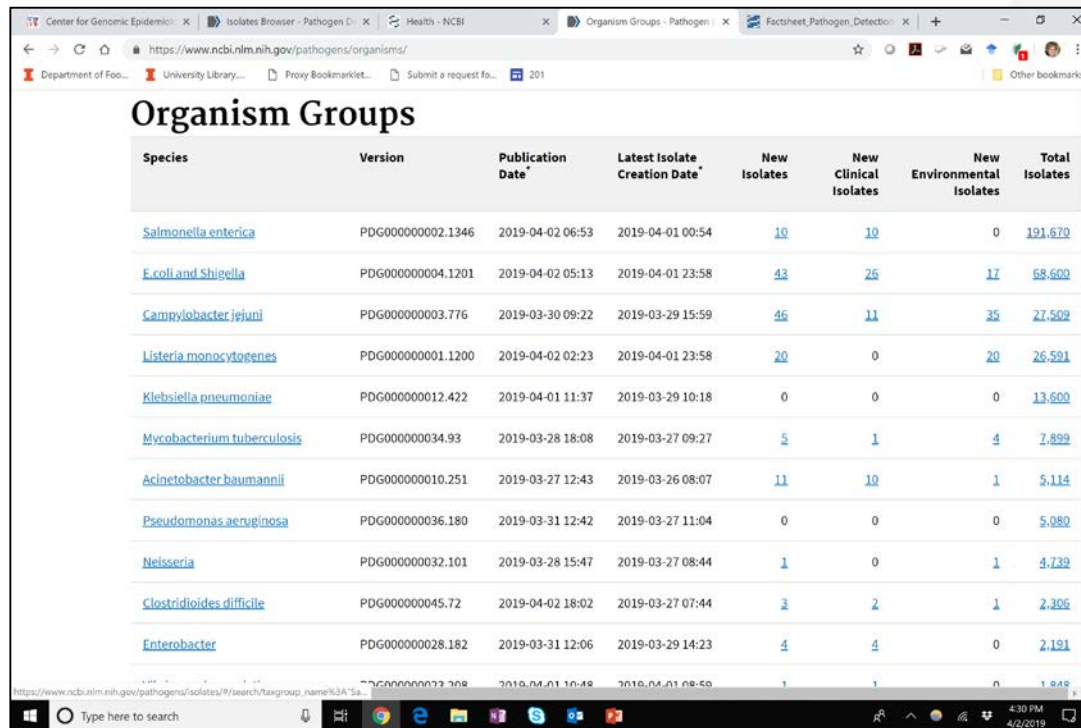
#	<input type="checkbox"/>	Organism C	Strain	Isolate	Create	Locati	Isolation So	Isolation type	SNP clust	Min	Min	Bio Sample	Assembly	K-mer gro	Sero	Host	AST p
1	<input type="checkbox"/>	Listeria monocytogenes	PNUSA	PDT000044	2014-12-07	USA		clinical	PDS000000	9	11	SAMN03253355		PDG000000			
2	<input type="checkbox"/>	Listeria monocytogenes	PNUSA	PDT000063	2015-04-30	USA		clinical	PDS000000	6	9	SAMN03567978		PDG000000			
3	<input type="checkbox"/>	Listeria monocytogenes	PNUSA	PDT000063	2015-04-30	USA		clinical	PDS000000	3	6	SAMN03567979		PDG000000			
4	<input type="checkbox"/>	Listeria monocytogenes	2013L-5169	PDT000065	2015-06-02	USA		clinical	PDS000000	4	8	SAMN03754733		PDG000000			
5	<input type="checkbox"/>	Listeria monocytogenes	2010L-1686	PDT000065	2015-06-02	USA		clinical	PDS000000	3	5	SAMN03755213		PDG000000			
6	<input type="checkbox"/>	Listeria monocytogenes	CFSAN	PDT000052	2015-04-07	USA:KS	chocolate ice cream	environmental	PDS000000	0	5	SAMN03463707		PDG000000			
7	<input type="checkbox"/>	Listeria monocytogenes	CFSAN	PDT000052	2015-04-07	USA:KS	chocolate ice cream	environmental	PDS000000	0	5	SAMN03463701		PDG000000			
8	<input type="checkbox"/>	Listeria monocytogenes	CFSAN	PDT000052	2015-04-07	USA:KS	chocolate ice cream	environmental	PDS000000	0	5	SAMN03463703		PDG000000			

#SFC2019

Demonstrate Query Relevant to This Group

Salmonella

192,000 total isolates

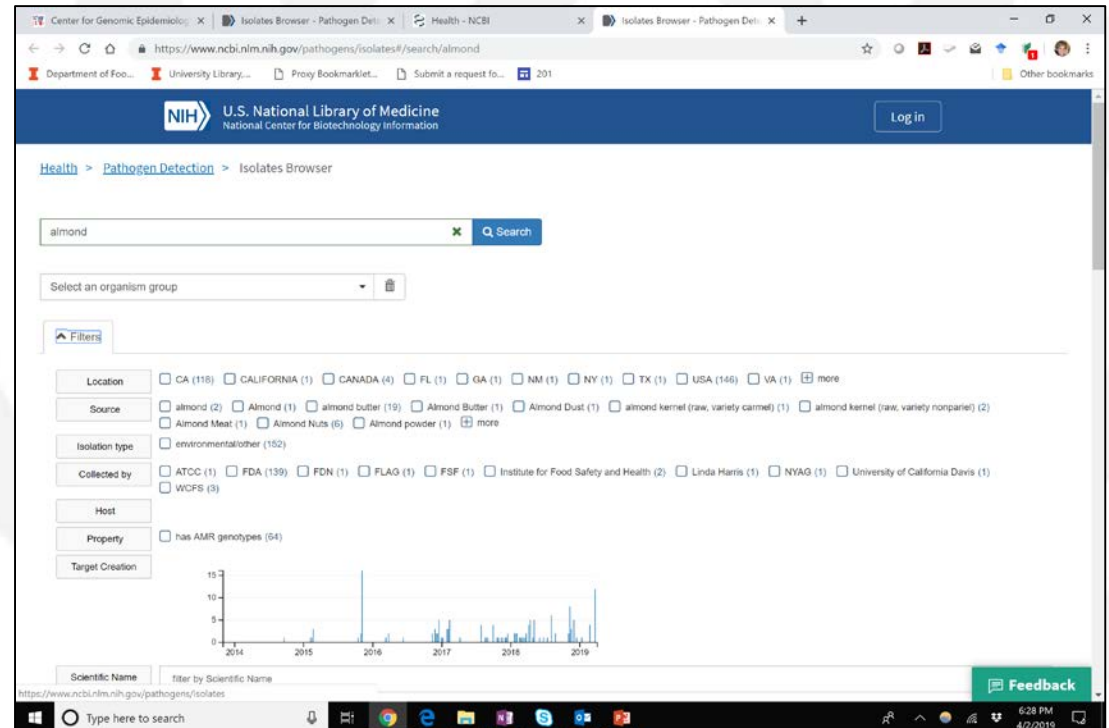


The screenshot shows the 'Organism Groups' table on the NCBI Pathogen Detection Isolates Browser. The table lists various bacterial species with their respective version numbers, publication dates, latest isolate creation dates, and counts for new isolates, new clinical isolates, new environmental isolates, and total isolates.


Species	Version	Publication Date*	Latest Isolate Creation Date*	New Isolates	New Clinical Isolates	New Environmental Isolates	Total Isolates
Salmonella enterica	PDG000000002.1346	2019-04-02 06:53	2019-04-01 00:54	10	10	0	191,670
E.coli and Shigella	PDG000000004.1201	2019-04-02 05:13	2019-04-01 23:58	43	26	17	68,600
Campylobacter jejuni	PDG000000003.776	2019-03-30 09:22	2019-03-29 15:59	46	11	35	27,509
Listeria monocytogenes	PDG000000001.1200	2019-04-02 02:23	2019-04-01 23:58	20	0	20	26,591
Klebsiella pneumoniae	PDG000000012.422	2019-04-01 11:37	2019-03-29 10:18	0	0	0	13,600
Mycobacterium tuberculosis	PDG000000034.93	2019-03-28 18:08	2019-03-27 09:27	5	1	4	7,899
Acinetobacter baumannii	PDG000000010.251	2019-03-27 12:43	2019-03-26 08:07	11	10	1	5,114
Pseudomonas aeruginosa	PDG000000036.180	2019-03-31 12:42	2019-03-27 11:04	0	0	0	5,080
Neisseria	PDG000000032.101	2019-03-28 15:47	2019-03-27 08:44	1	0	1	4,739
Clostridioides difficile	PDG000000045.72	2019-04-02 18:02	2019-03-27 07:44	3	2	1	2,306
Enterobacter	PDG000000028.182	2019-03-31 12:06	2019-03-29 14:23	4	4	0	2,191
Staphylococcus aureus	PDG000000023.208	2019-04-01 10:48	2019-04-01 08:50	1	1	0	1,949

Almonds

152 isolates. Most CA, FDA, Enviro.







1 Cluster




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

Log in

Health > Pathogen Detection > Isolates Browser




almond   Search


Select an organism group  


 Filters  Expand All  Download


24 environment cluster with
425 total, including clinical

Matched clusters

	Page 1 of 16 							3 	View 1 - 3 of 47
#	Organism groups	SNP cluster	Matched isolates	Matched clinical isolates	Matched environmental isolates	Total isolates	Minimal min-diff	Latest update	
1	Salmonella enterica	PDS000002780.59	7	0	7	90	0	2018-11-08	
2	Salmonella enterica	PDS000001617.20	1	0	1	22	0	2017-07-31	
3	Salmonella enterica	PDS000002770.176	24	0	24	425	1	2019-03-25	








Page 1 of 16 

3 

View 1 - 3 of 47

Matched isolates

	 Choose Columns 	Page 1 of 8 												20 	View 1 - 20 of 152	
#	Organism Group	Strain	Serovar	Isolate	Create Date	Location	Isolation Source	Isolation type	Host	SNP cluster	Min-s	Min-d	BioSample	Assembly	AMR genotypes	K-mer group
1	Salmonella enterica	CFSAN06		PDT000482408.1	2019-03-25	USA:CA	Finished Almonds	environmental		PDS000032387.4	1	3	SAMN103969			PDG000000002.1346
2	Salmonella enterica	CFSAN06		PDT000482369.1	2019-03-25	USA:CA	Raw Almonds	environmental		PDS000002770.1	9	7	SAMN103959			PDG000000002.1346
3	Salmonella enterica	CFSAN06		PDT000482332.1	2019-03-25	USA:CA	Raw Almonds	environmental		PDS000027731.6	1	n/a	SAMN103959			PDG000000002.1346
4	Salmonella enterica	CFSAN06		PDT000482329.1	2019-03-25	USA:CA	Finished Almonds	environmental			n/a	n/a	SAMN103959			PDG000000002.1346

Automated comparisons: SNPs, timeline

Salmonella - 425 isolates
PDG000000002.1346 / PDS000002770.176

24 Isolates Selected ✕ Clear

Distance between selected isolates :
minimum=0 SNPs, maximum= 59 SNPs, average=36 SNPs

Target creation date range :
2013-09-04 to 2019-03-25

Filter within selected isolates Group

isolates mindiff

- 2019 - 6 Isolate(s)
- 2018 - 13 Isolate(s)
- 2017 - 4 Isolate(s)
- 2013 - 1 Isolate(s)

environmental/other USA:CA 2013-09-04 5

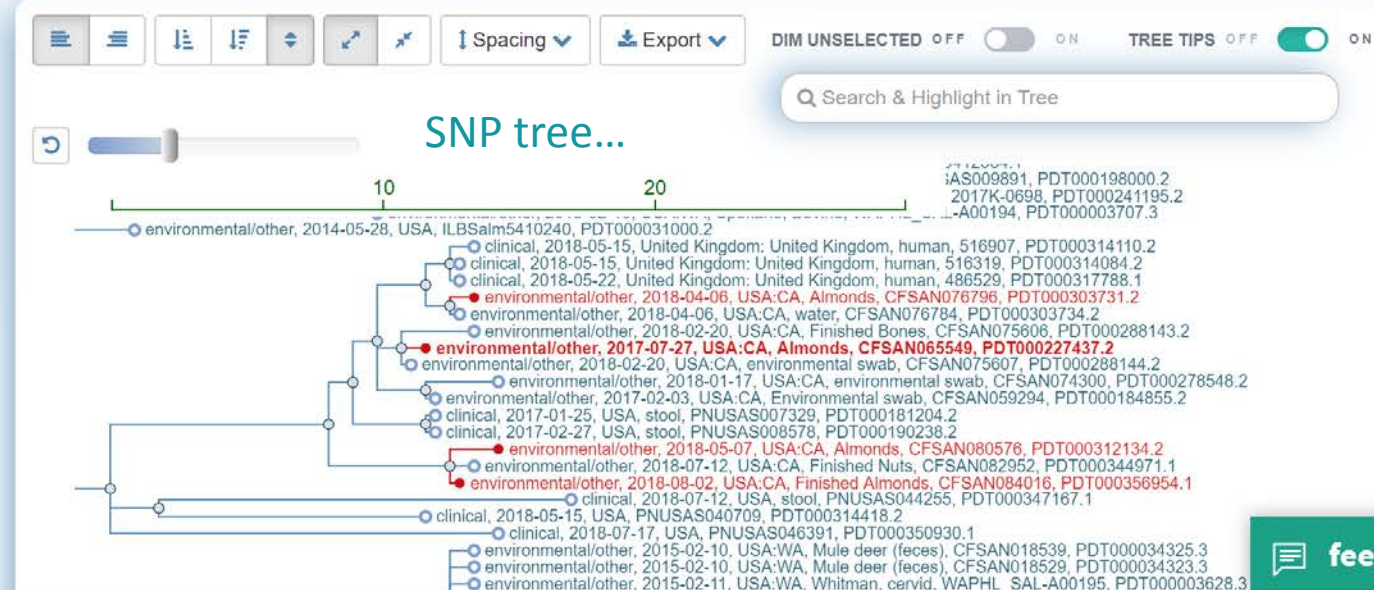
Timecourse...

Filters Share Hide Table Labels Neighbors Links to genomes... Download

Choose Columns Page 1 of 85 View 1 - 5 of 425

#	Strain	Serovar	Isolate	Create Date	Location	Isolation Source	Isolation type	Host	Min-s	Min-d	BioSample	Assembly
1	CFSAN06		PDT000482241.1	2019-03-25	USA:CA	Almonds	environmental/		0	7	SAMN103956	
2	CFSAN06		PDT000482247.1	2019-03-25	USA:CA	Almonds	environmental/		1	8	SAMN103956	
3	CFSAN06		PDT000482328.1	2019-03-25	USA:CA	Almonds	environmental/		0	7	SAMN103956	
4	CFSAN06		PDT000482369.1	2019-03-25	USA:CA	Raw Almonds	environmental/		9	7	SAMN103956	
5	CFSAN06		PDT000227437.2	2017-07-27	USA:CA	Almonds	environmental/		1	5	SAMN073428	

Choose Columns Page 1 of 85 View 1 - 5 of 425



More Zooming



What if **you** collected a *Salmonella* that fell in one of these clusters?

What if the **FDA** did?
What if the **CDC** did?

Genome Sequencing Challenges & Opportunities

When are two isolates the same or different?

Can we find identical isolates in different locations?

What does this mean for environmental monitoring?

How can a company engage with this analysis?

Some of the challenges

- Identical bacteria (100% whole genome match) can be found in different places that can be potential sources of foodborne disease outbreaks
 - So: Is identical enough?
- Minor changes (a “few” SNPs) can occur quickly, e.g. during growth in an infected human
 - So: How close is close enough?
- What about epi?

A somewhat hypothetical case study

- Three people with listeriosis
 - The *Listeria monocytogenes* isolates differ by 2 SNPs
- All three 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 observation

Appl Environ Microbiol. 2015 Sep 1;81(17):6024-37. doi: 10.1128/AEM.01049-15. Epub 2015 Jun 26.



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.

clonal spread within a state (7 clades), or (iii) spanning multiple states (5 clades). Isolates that formed one of 11 deli-specific clades differed by a median of 10 SNPs or fewer. Isolates from 12 putative persistence events had significantly fewer SNPs (median, 2 to 22 SNPs) than between isolates of the same subtype from other delis (median up to 77 SNPs), supporting persistence of the strain. **In 13 events, nearly indistinguishable isolates (0 to 1 SNP) were found across multiple delis.** No individual genes were

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

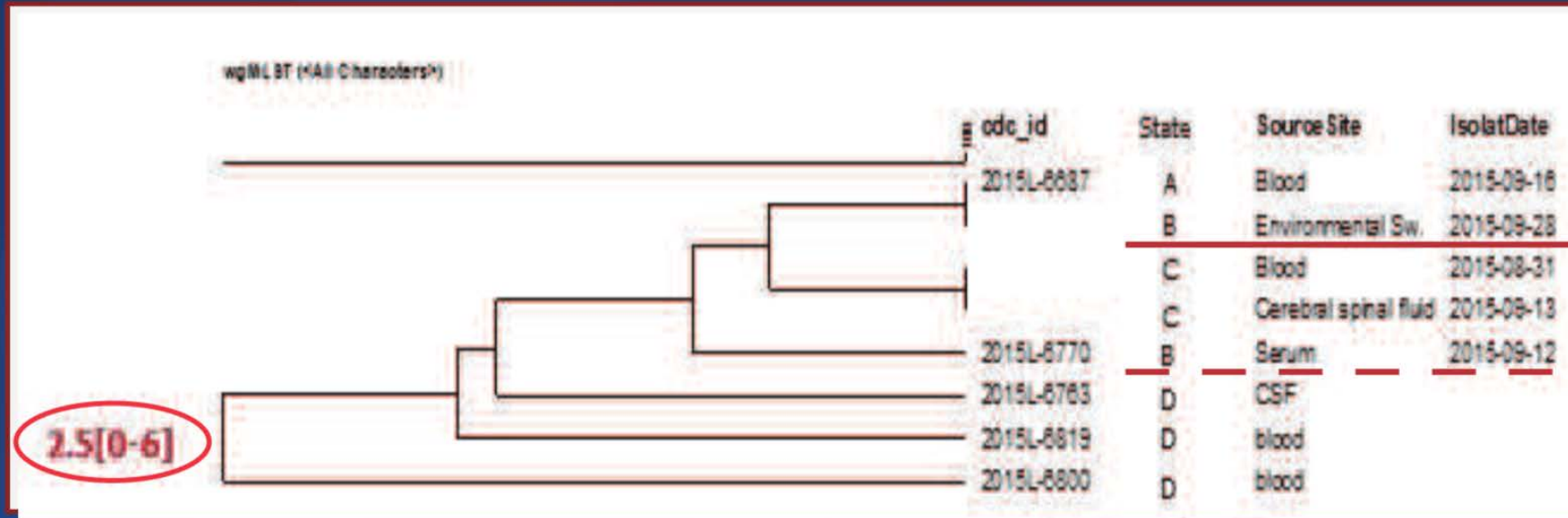
US Regulatory Position (Dr. Peter Gerner-Smidt's slides from 2016):

- A WGS match between a food isolate and a clinical isolate does NOT mean that the food caused the patient's illness
- They likely share an ancestor somewhere in the food production chain
- Epidemiological and traceback investigation remain critical

But wouldn't this still be the starting point of investigation?

A WGS match between a food isolate and a clinical isolate does NOT mean that the food caused the patient's illness

Listeria Cluster



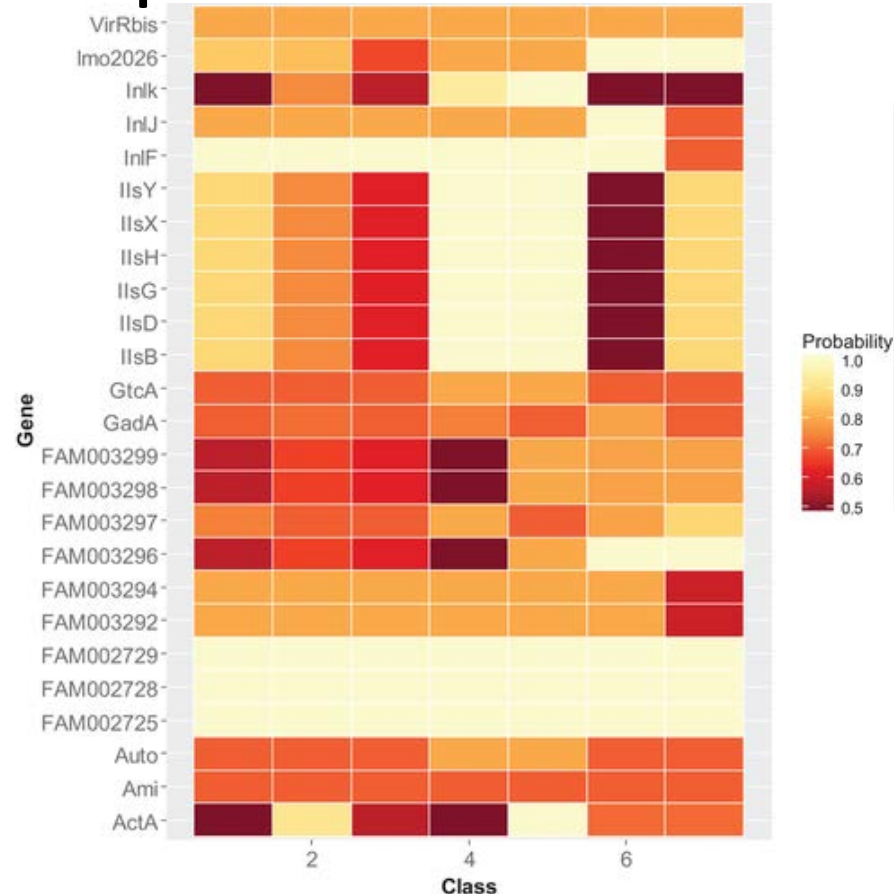
They likely share an ancestor somewhere in the food production chain
Epidemiological and traceback information remains critical

Dr. Peter Gerner-Smidt
Chief, Enteric Disease
Laboratory Branch,
CDC.
4th Asia-Pacific
International Food
Safety Conference,
Penang, Malaysia,
October 11- 13, 2016

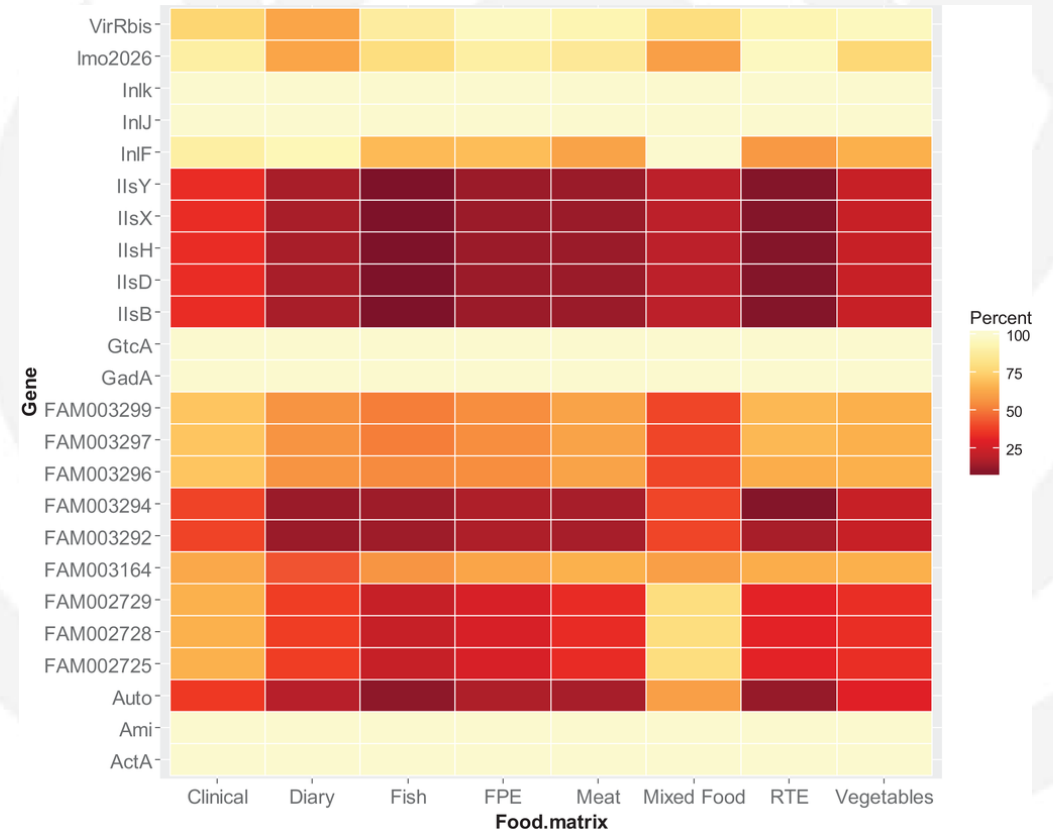
Individual Strain Risk Assessment

What about the strain
you just found?

Machine-Learning ID of Most Important Virulence Genes



Prevalence of Those Genes in Isolates of Different Courses



Genome Sequencing Challenges & Opportunities

When are two isolates the same or different?
Can we find identical isolates in different locations?
What does this mean for environmental monitoring?
How can a company engage with this analysis?

Web-Based Analysis

Center for Genomic Epidemiology

[Home](#)[Organization](#)[Project](#)[Services](#)[Contact](#)

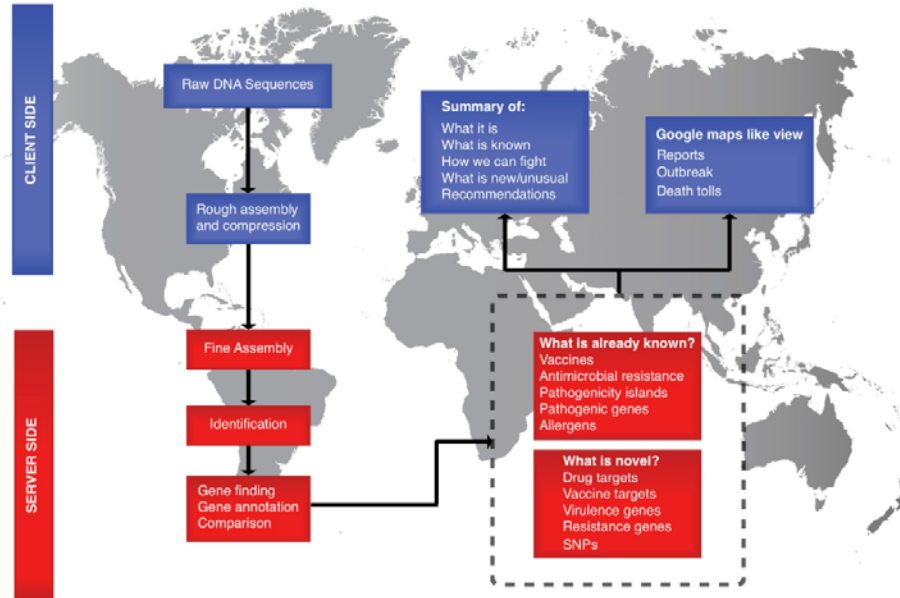
Services

Pipeline:

- Pipeline to analyse bacterial genomes.
[Bacterial Analysis Pipeline \(BAP\)](#)

Phenotyping:

- Identification of acquired antibiotic resistance genes.
[ResFinder](#)
- Identification of functional metagenomic antibiotic resistance determinants.
[ResFinderFG](#)
- Identification of genes and mutations leading to linezolid resistance
[LRE-finder](#)
- Identification of acquired antibiotic resistance genes using Kmers.
[KmerResistance](#)
- Prediction of a bacteria's pathogenicity towards human hosts.
[PathogenFinder](#)
- Identification of acquired virulence genes.
[VirulenceFinder](#)
- Determination of Restriction-Modification sites (based on [REBASE](#).)
[Restriction-ModificationFinder](#)
- SPIFinder identifies Salmonella Pathogenicity Islands
[SPIFinder](#)
- ToxFinder identifies genes involved in mycotoxin synthesis



Welcome to the Center for Genomic Epidemiology

The cost of sequencing a bacterial genome is \$50 and is expected to decrease further in the near future and the equipment needed cost less than \$150 000. Thus, within a few years all clinical microbiological laboratories will have a sequencer in use on a daily basis. The price of genome sequencing is already so low that whole genome sequencing will also find worldwide application in human and veterinary practices as well as many other places where bacteria are handled. In Denmark alone this equals more than 1 million isolates annually in 15-20 laboratories and globally up to 1-2 billion isolates per year. The limiting factor will therefore in the future not be the cost of the sequencing, but how to assemble, process and handle the large amount of data in a standardized way that will make the information useful, especially for diagnostic and surveillance.

The aim of this center is to provide the scientific foundation for future internet-based solutions where a central database will enable simplification of total genome sequence information and

News

New article: Rapid and precise alignment of raw reads against redundant databases with KMA
September 2018
[Link to article...](#)

Improved laboratory procedure and bioinformatics workflow for metagenomics-based identification of microorganisms: The workflow includes a curated reference genome database, reducing the identification of false-positives
April 2018
[Link to article...](#)

What Can We Learn from a Metagenomic Analysis of a Georgian Bacteriophage Cocktail?
December 2015
[Link to article...](#)

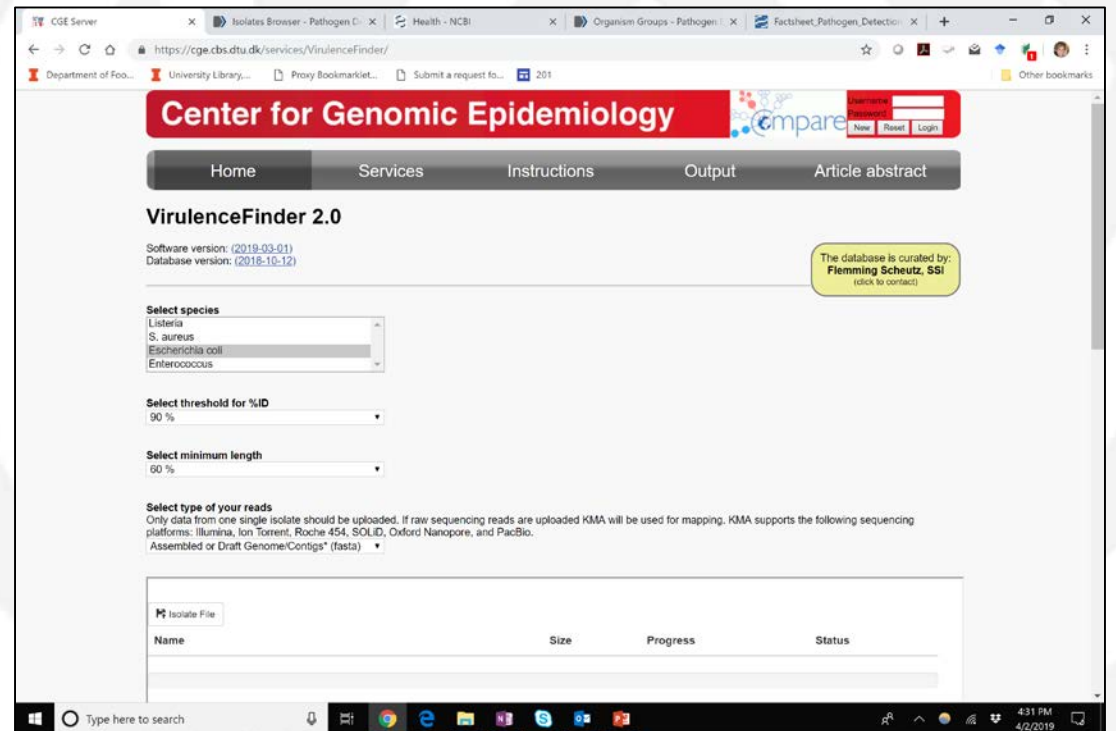
WGS typing is a superior alternative to conventional typing strategies
August 2015
In combination with other available WGS typing tools, E. coli serotyping can be performed solely from WGS data, providing faster and cheaper typing than current routine procedures. [Link to article...](#)

Introduction to microbial whole genome sequencing and analysis for clinical microbiologist
April 2015
We offer clinical microbiologists the possibility to learn how to use the tools for e.g. typing, identifying plasmids, antibiotic resistance and virulence genes and for phylogenetic analysis. [Sign up...](#)

Consortium to combat infectious disease outbreaks
January 2015
The COMPARE project has been funded with 20 million Euros from the EU. The Consortium consists of

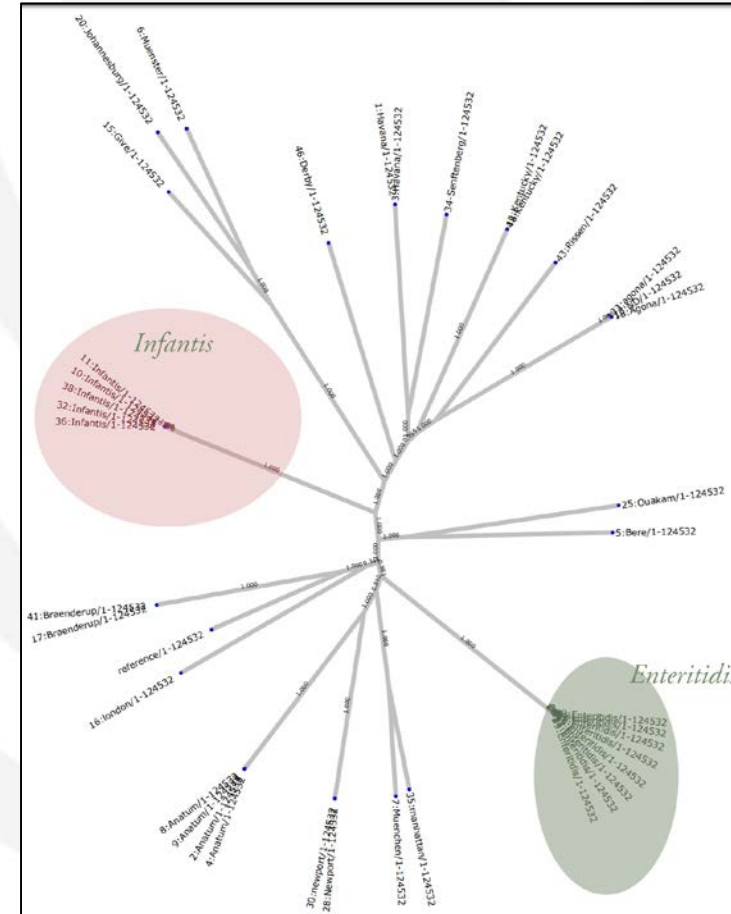
Portals to automate key analyses

- Web uploads to bioinformatic tools
- Actively maintained and updated with current
 - Methods
 - Databases
- Simplified inputs
 - Raw reads or basic assemblies
- Pipelines also available
 - So YOU could use this



Profiling *Salmonella* in Queretaro, MX with CGE tools

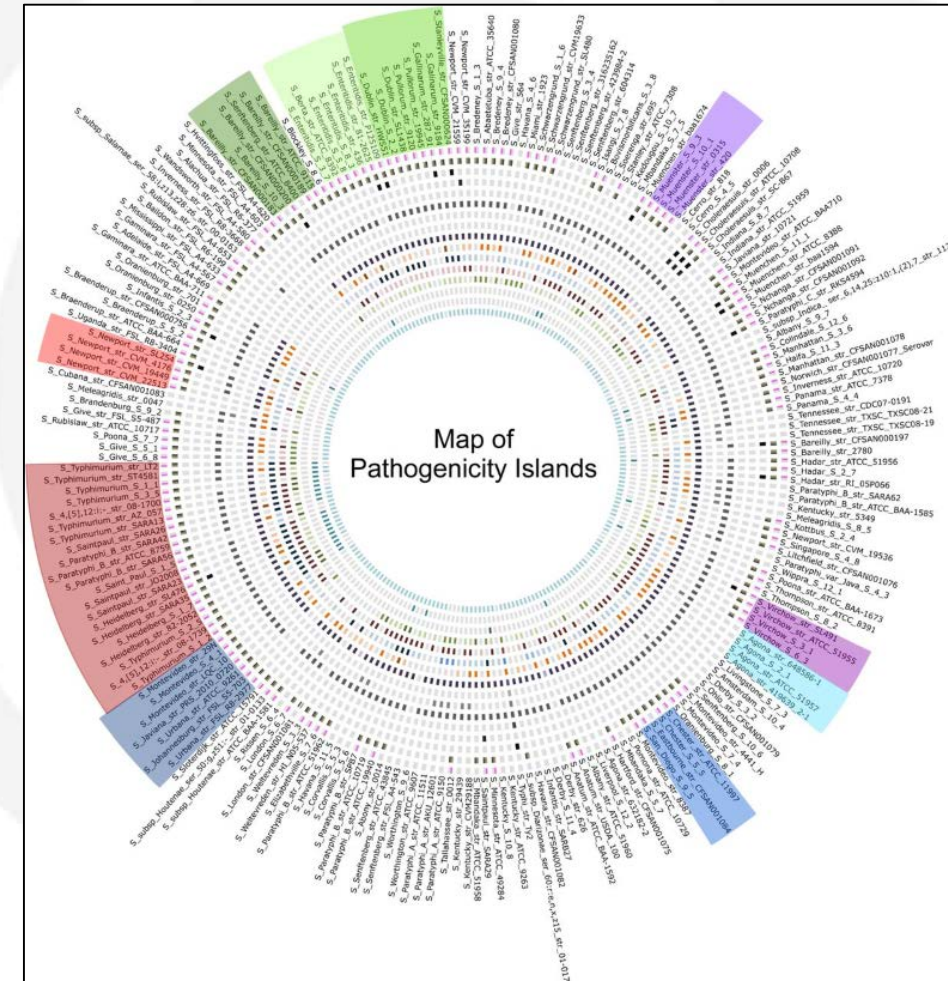
- Poultry-associated *Salmonella* isolates
 - Supermarket or open-markets
 - 4 years of collection
 - Culture-based isolation
 - Approx. 500 total isolates
- Genome Sequencing Process
 - UAQ extracts DNA, ships
 - Illinois sequences (HiSeq)
 - New student analyzes (< 1 mth)



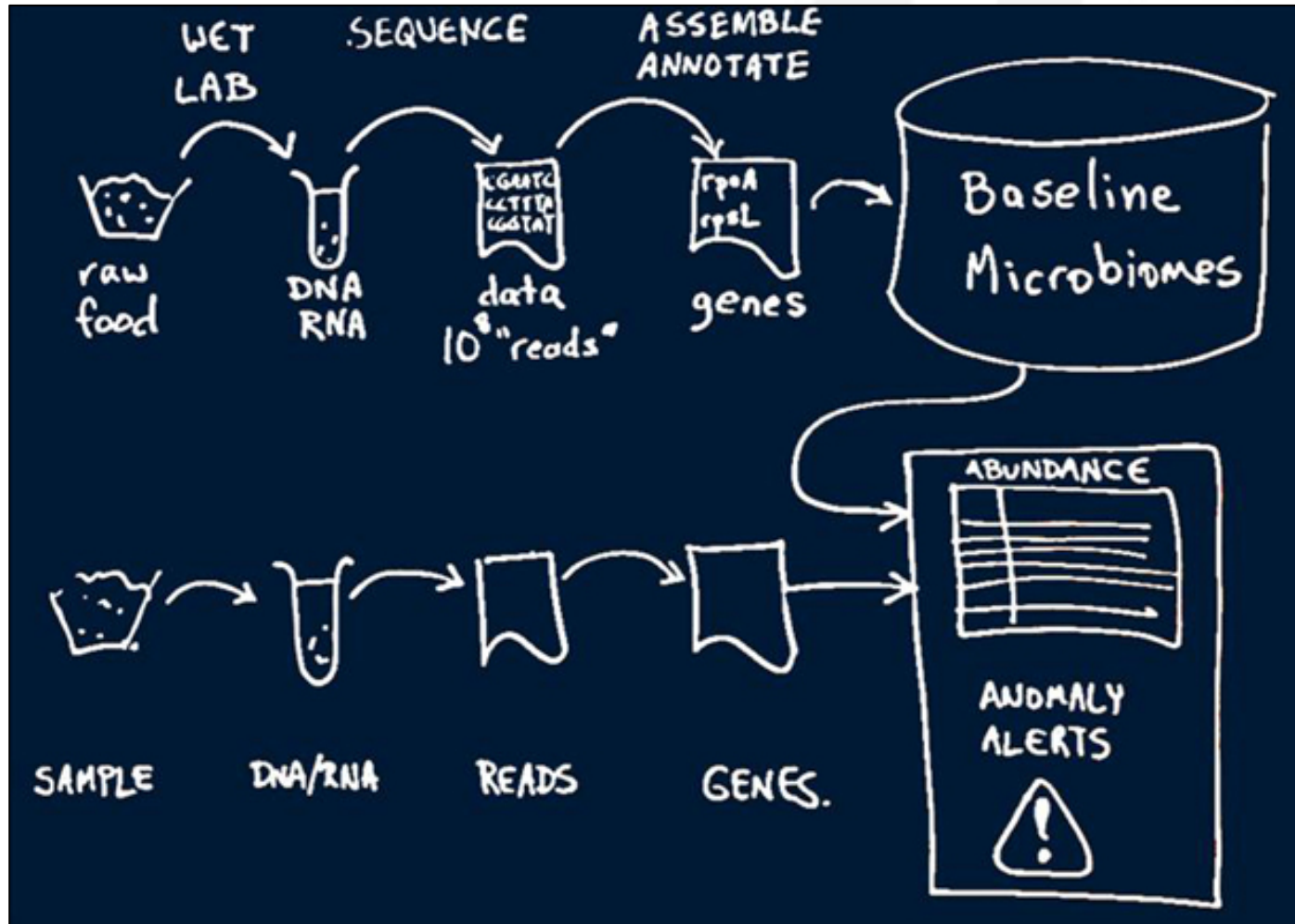
Profiling *Salmonella* in Queretaro, MX with CGE tools

- Epidemiology
 - Predicts serotype (2 methods)
 - Creates SNP tree
 - Generates pretty tree
- Other Information
 - Antimicrobial resistance genes
 - Plasmids
 - Pathogenicity islands (SPIs)

Similar work could be done comparing isolates to sequences downloaded off GenomeTrakr



IBM: Consortium for Sequencing the Food Supply Chain



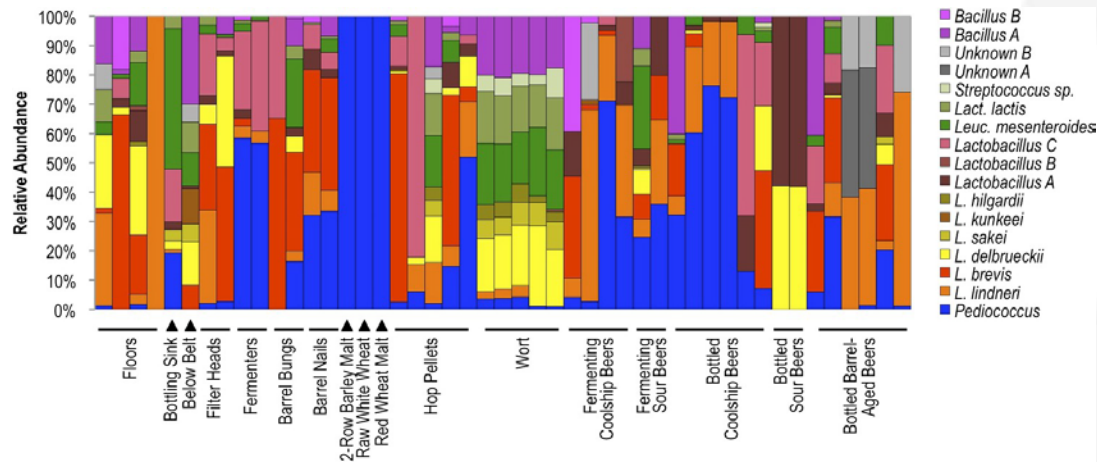
Goal = Microbiome based process control

Could this be just another service?

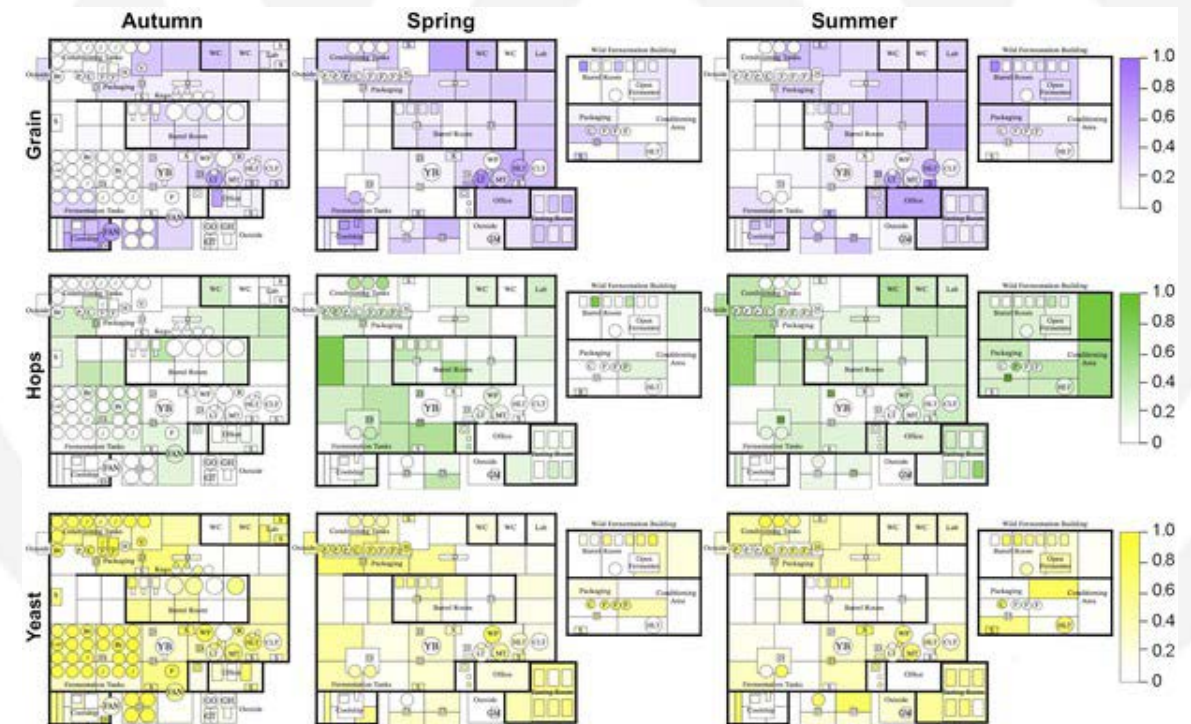


Neat Potential: Brewery Microbiome Baseline

Lactic Acid Bacteria Diversity Among Surfaces and Products



Relative contribution of ingredients to microbial communities



Big Questions:

How would a pathogen introduction shift the microbiome?

Would this provide enhanced safety?

Conclusions

Core Questions

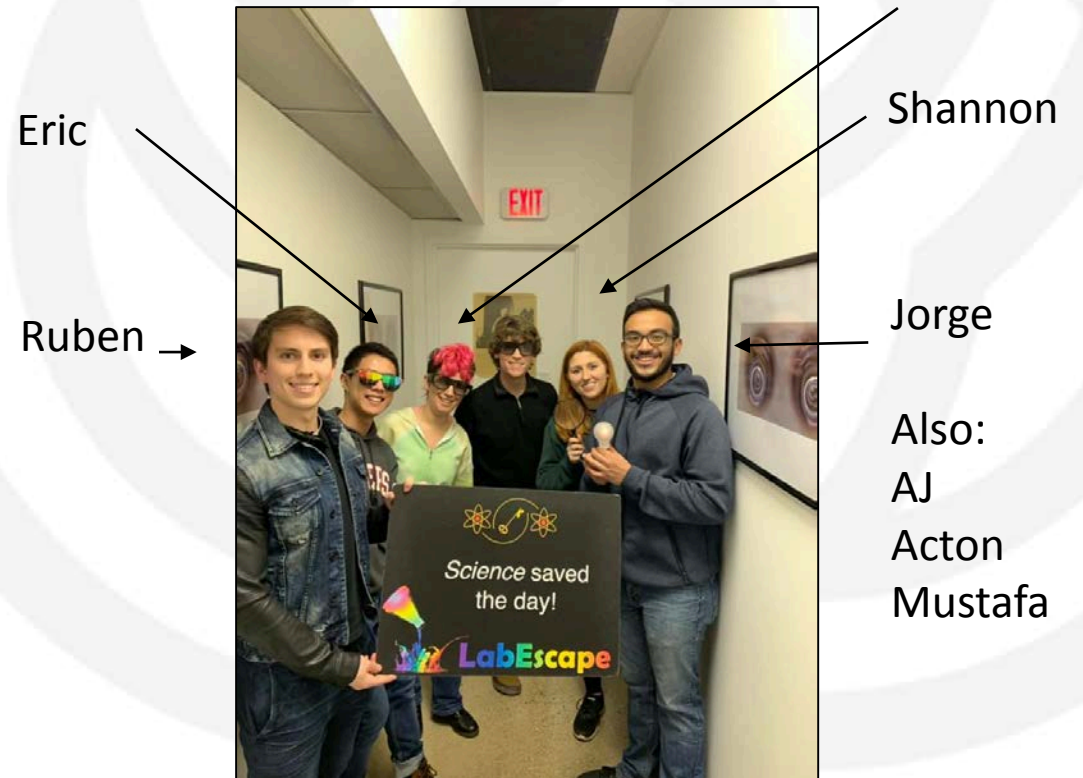
- Why should we care about genome sequencing?
 - Federal agencies now detect an unprecedented number outbreaks
 - Can link food/environmental isolates to clinical isolates without prior knowledge
- Will this really impact me/my company?
 - It could. 'Bad luck' is becoming more common
- Can a non-expert actually use genome sequencing?
 - Yes. Critical steps are getting automated.
- What (positive) impacts might WGS have on the food industry?
 - Improved environmental monitoring and pathogen profiling

Key Messages

Funding

- USDA Cooperative State Research, Education, and Extension Service Hatch project ILLU-698-903 directed by Stasiewicz
- Startup funds provided to Stasiewicz by the University of Illinois
- College of ACES, Office of International Programs. Global Academy 2018 funding

Thanks to my lab and wife Natalie (w)



Questions?

- Thank you for your attention
- I will be outside after the talk to discuss follow-up opportunities
- Feel free to contact me later
 - mstasie@Illinois.edu