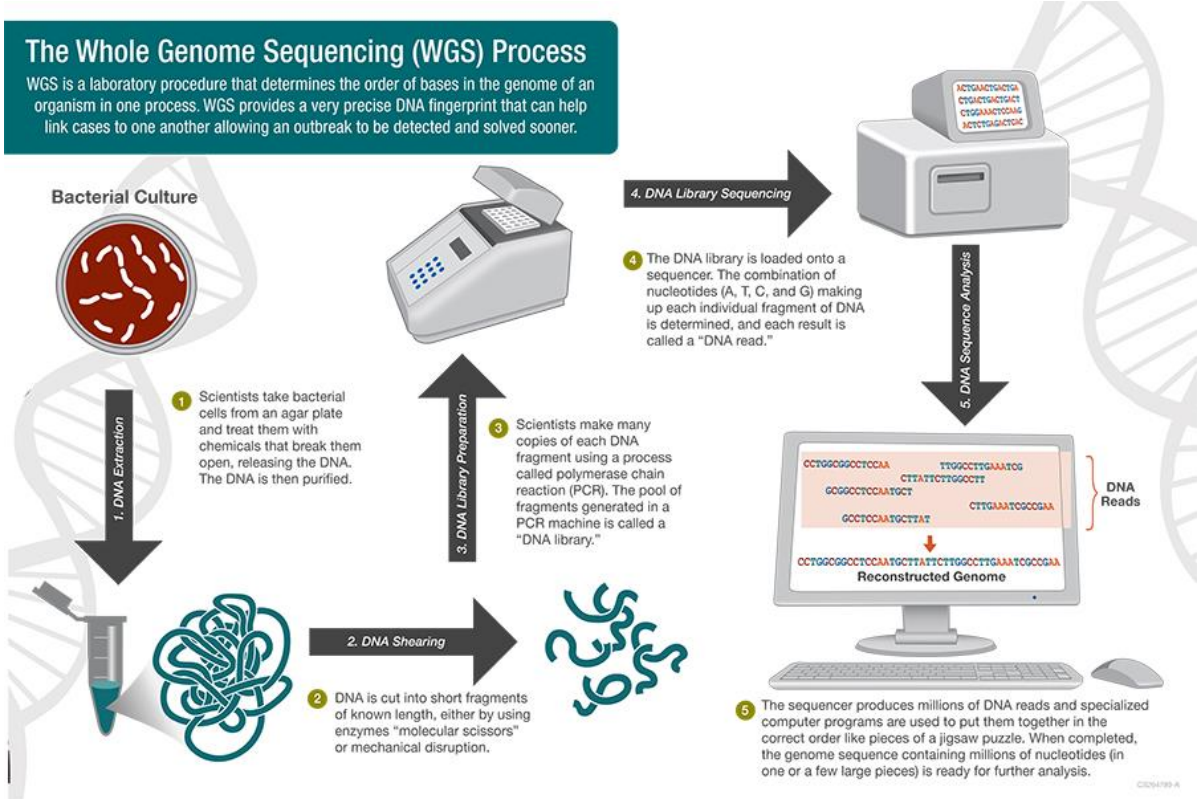
Next Generation Sequencing As a Tool in Foodborne Disease Surveillance And Outbreak Investigation – Challenges and Opportunities

Peter Gerner-Smidt, MD, DSc
Branch Chief

EFSA 2018 CONFERENCE
SCIENCE · FOOD · SOCIETY
Parma - Italy, 18-21 September 2018

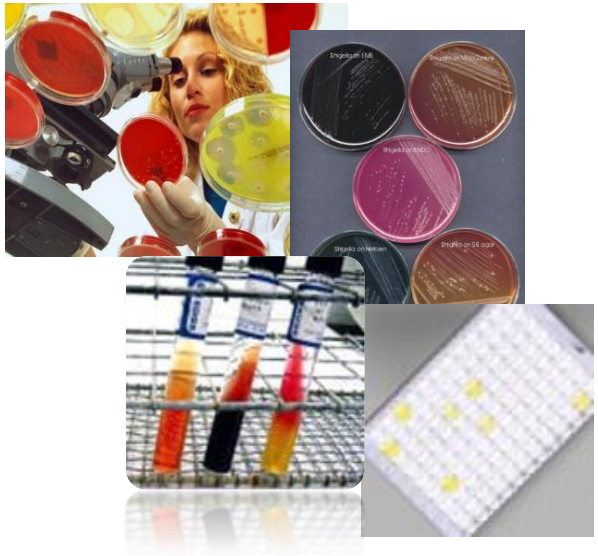
Next Generation Sequencing (NGS) ~ Whole Genome Sequencing (WGS)



Why The Hype About WGS?

- WGS is transforming microbiology, replacing numerous traditional methods with one in a single efficient workflow

  Identification – Serotyping – Virulence profiling – Resistance profiling – Subtyping – and much more



Reference Characterization by WGS:

'One Shot' Characterization Of STEC

Genus/Species: *Escherichia coli*

Serotype: O104:H4

Pathotype: Shiga toxin-producing and enteroaggregative *E. coli* (STEC/EAEC)

Virulence profile: *stx2a, aggR, aggA, sigA, sepA, pic, aatA, aaiC, aap*

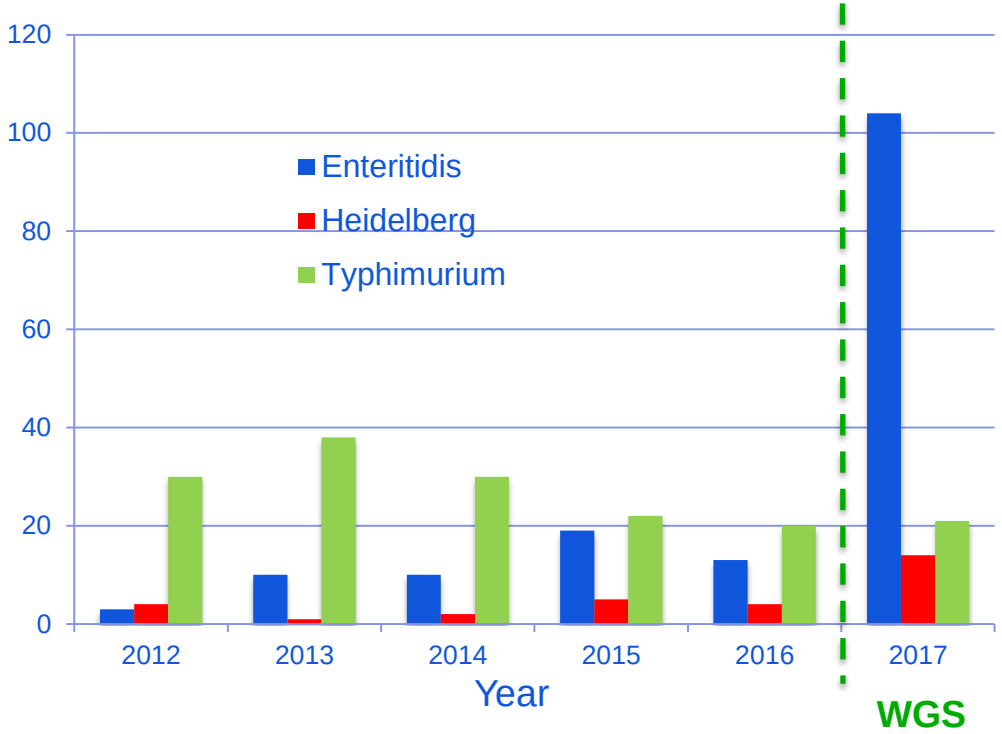
Sequence Type: ST678

Allele code: 102.45.26.35.3

Antimicrobial resistance genes: bla_{TEM-1}, $bla_{CTX-M-15}$, *strAB, sul2, tet(A)A, dfrA7*

Salmonella outbreaks in Canada

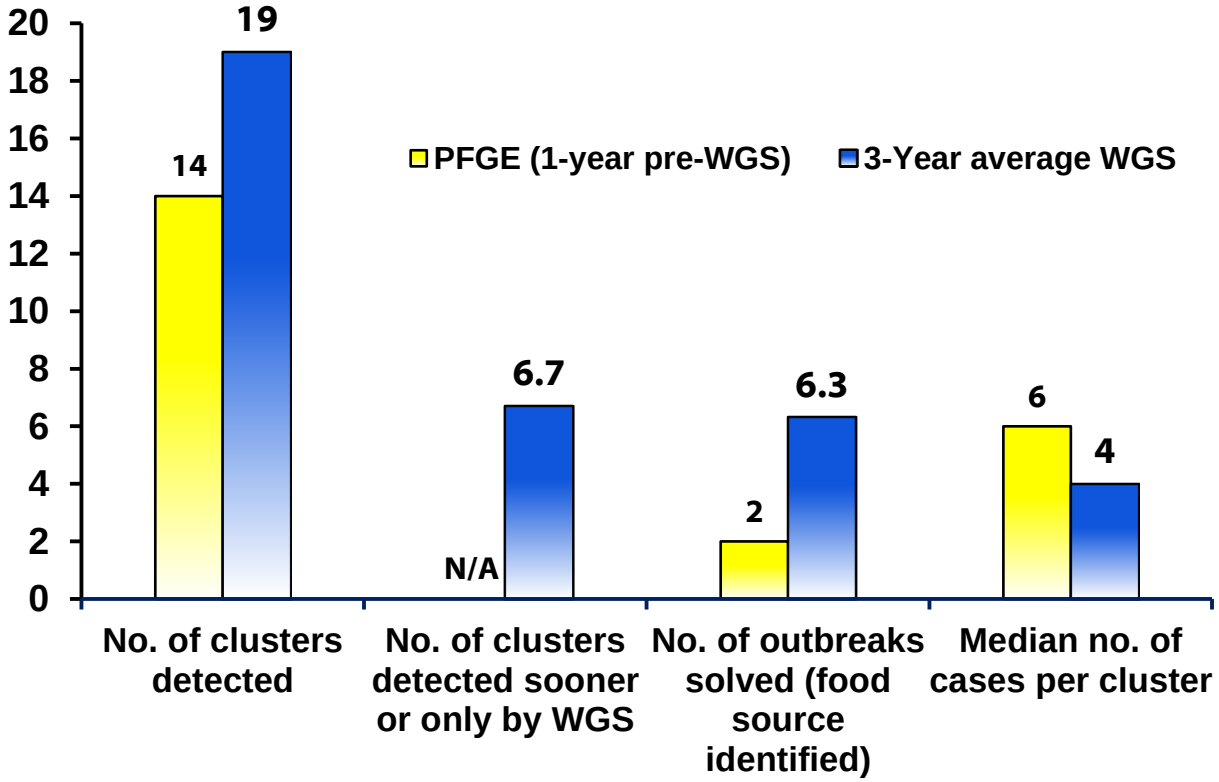
Number of Salmonella outbreaks detected with laboratory data



Courtesy C. Nadon, Public Health Agency of Canada

Real-time WGS Improves Laboratory Surveillance

Listeria Metrics

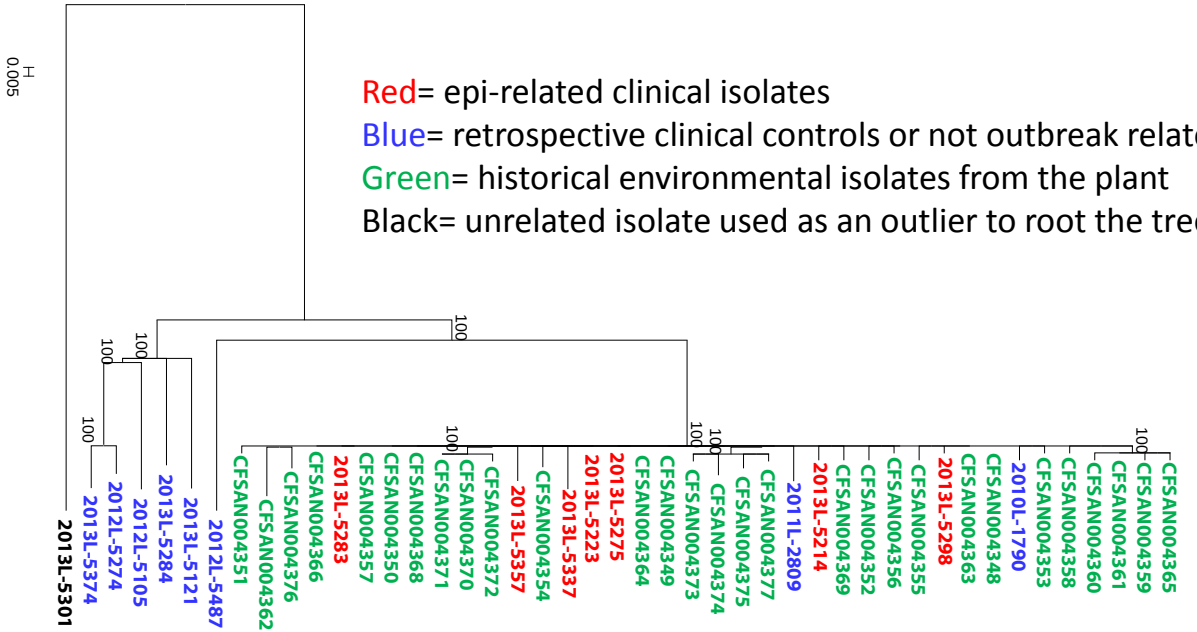


Courtesy Amanda Conrad, CDC Outbreak Response & Preparedness Branch

Listeria Outbreak Linked to Artisan Cheese (2013)

hqSNP

Historical isolates from the plant environment added to the comparison (courtesy FDA/CFSAN)



Red= epi-related clinical isolates

Blue= retrospective clinical controls or not outbreak related

Green= historical environmental isolates from the plant

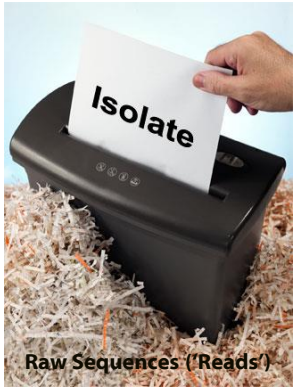
Black= unrelated isolate used as an outlier to root the tree

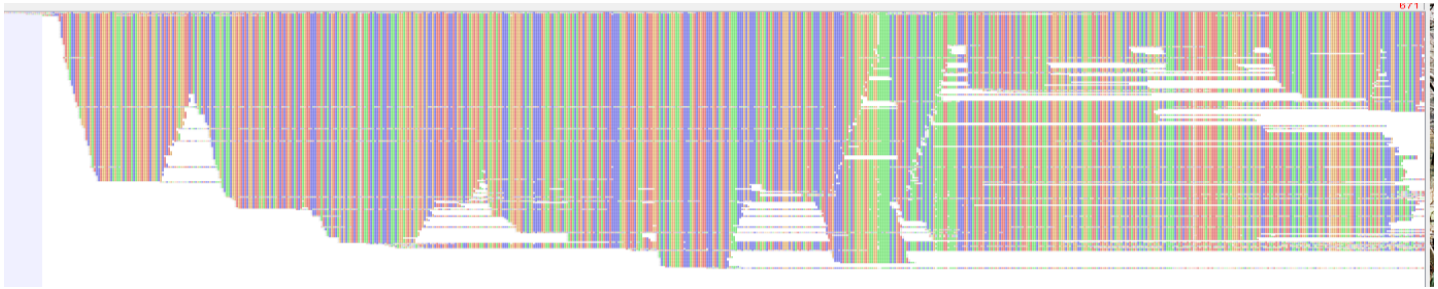
How WGS Influence Outbreak Investigations

- **Improved case definitions in outbreaks**
  - Apparent PFGE clusters are not single-source outbreaks or are pseudo-clusters
  - Isolates with same PFGE patterns may be unrelated
  - Isolates with different PFGE patterns may be related
- **Increase confidence in the link between human and product isolates**
- **Link historical cases to a current outbreak investigation**
- **Characterize the ecology of long-term pathogen reservoirs in the food chain**

The Basics of WGS

- “Massive parallel sequencing”
- The whole genome sequenced in small random pieces (‘shotgun sequencing’, 25- >1000 bp) multiple times (‘coverage’)
- Four major & different sequence technologies
  - Each with different strengths and weaknesses
- ‘Coverage’ usually 20- several 100 X





The Basics of WGS

- Assembling and annotating the sequence
  - Solving the puzzle using an ‘assembler’ software

‘Reference -Based Assembly

‘*de novo* Assembly’

  - Assembled in 1- 200 (- 500) fragments (‘contigs’)
  - MANY DIFFERENT ASSEMBLY SOFTWARES- None are perfect
  - Each make different errors

Two High-Discrimination Analytical Approaches

- Nucleotide level analysis
  - Single Nucleotide Polymorphisms (SNPs)
  - 'Like assessing all the letters in a book'
    - Difficult to standardize between laboratories
- Gene level
  - Multi-Locus Sequence Typing (MLST, cg/wgMLST)
  - 'Like assessing all the words in a book'
    - Can be standardized between laboratories
- Many different pipelines and schemes for sequence analysis
  - No two pipelines provide the exact same results!
    - Results generated by different pipelines can NOT directly be compared
  - But each generates **reproducible results**

Quality control – quality control – QUALITY CONTROL

- Assuring WGS quality within one institution is fairly easy
  - WGS works very well for national surveillance with centralized analysis
- No international quality standards exist
- No international consensus on the use of specific pipelines

What About Global Surveillance of Foodborne Infections?

- A foodborne infection on one continent may have its source on a different continent
- International outbreaks are common

Analytical Tools in Public Domain

Center for Genomic Epidemiology

Home | Organization | Project | Services | Contact

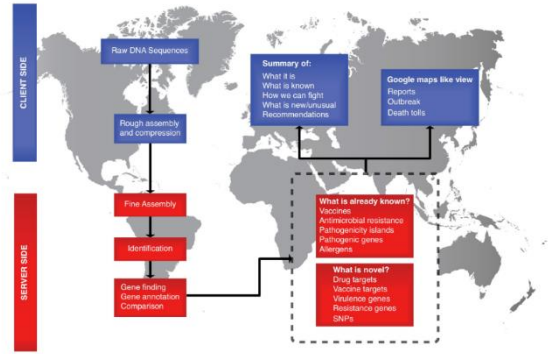
Services

Pipeline:

- Identifcation of acquired antibiotic resistance genes. Bacterial Analysis Pipeline (BAP)

Phenotyping:

- Identifcation of acquired antibiotic resistance genes. ResFinder
- Identifcation of functional metagenomic antibiotic resistance determinants. ResFinderFG
- Identifcation of acquired antibiotic resistance genes using Kmers. KmerResistance
- Prediction of a bacteria's pathogenicity towards human hosts. PathogenFinder
- Identifcation of acquired virulence genes. VirulenceFinder
- Determination of



Welcome to the Center for Genomic Epidemiology

INNUENDO

Project definition | The Team | News and Events | Links | Innuendo Sequence DataBase | Teaching

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

Log in

Health > Pathogen Detection

Pathogen Detection BETA

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.

Find isolates now!

Examples:

1. Search for isolates encoding a mobile colistin resistance gene and a KPC beta-lactamase search: AMR_genotypes:mcr* AND AMR_genotypes:blaKPC*
2. Search for Salmonella isolates from the USA search: geo_loc_name: USA AND taxgroup_name:"Salmonella enterica"

Learn More

About
FAQ
Browser Factsheet
Antimicrobial Resistance Factsheet
Antimicrobial Resistance
Contributors

Data Resources

Isolates Browser
Antimicrobial resistance reference gene database

WGS for Foodborne Disease Surveillance in The Global Context

Fast, precise, simple communication and easy sharing of data is key in outbreak investigations

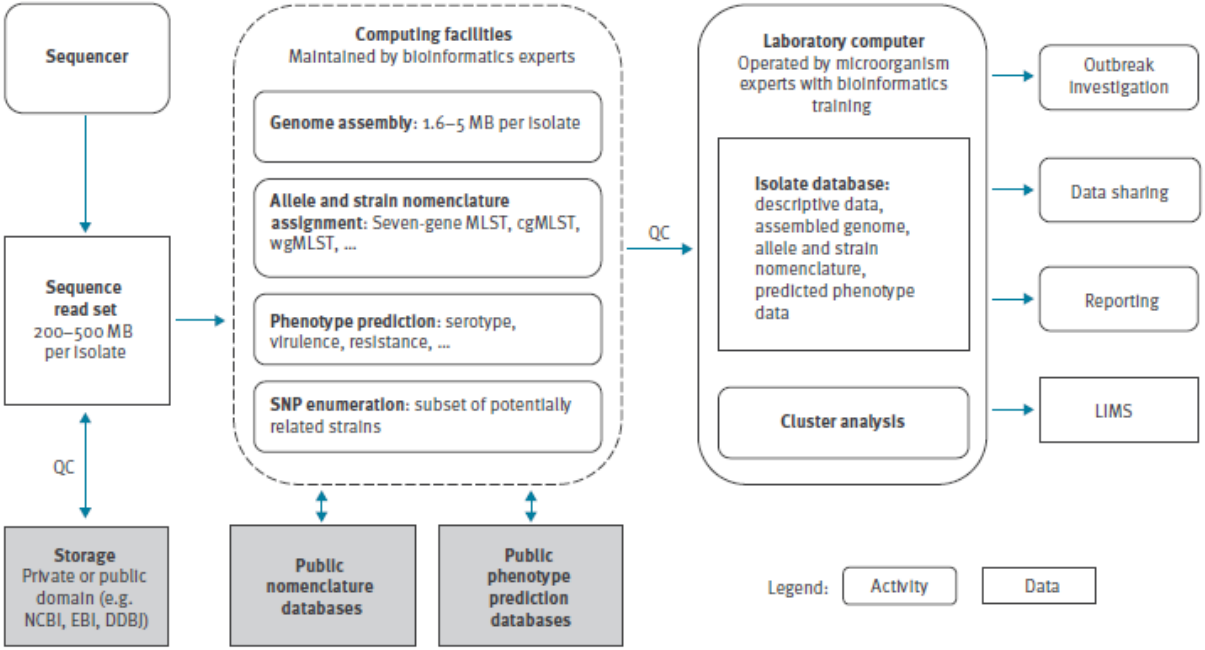
- **Standardized/harmonized and validated generation of results**
- **Results in standardized format**
- **Low volume format**
  - **to accommodate slow internet speeds**
  - **no need to go back to raw data**
- **Solutions must be PRACTICAL and NOT necessarily PERFECT 'If it works, it is good enough'**

PulseNet International
USA, Canada, Europe, Asia Pacific, Africa, Middle East, Latin America and the Caribbean

International Standardization and Harmonization Of WGS For Surveillance Of Foodborne Pathogens

The PulseNet Model

Nadon C, Van Walle I, *et al*. PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill. 2017;22(23):pii=30544. DOI: http://dx.doi.org/10.2807/1560-7917.ES.2017.22.23.30544*

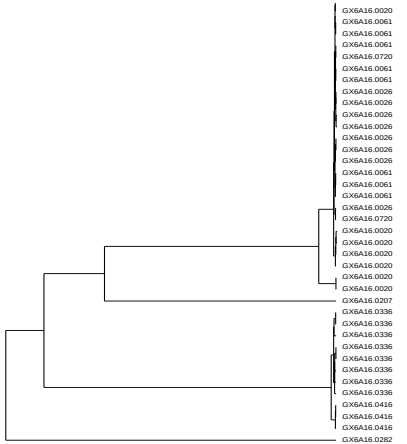


PulseNet International
USA, Canada, Europe, Asia Pacific, Africa, Middle East, Latin America and the Caribbean

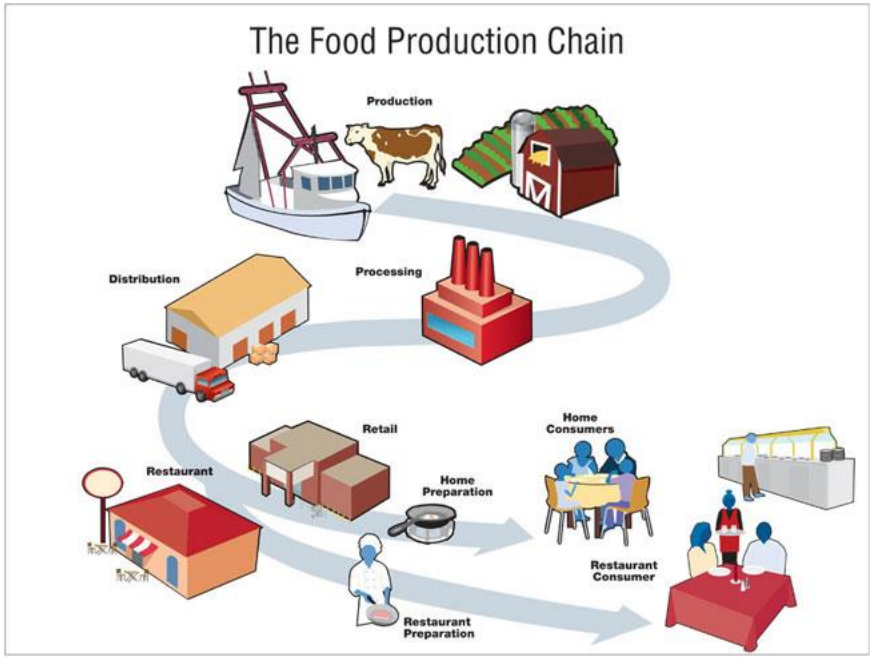
The Challenge of Data Interpretation

With WGS, How Close Is Close?

- No isolates 100% identical
- WGS data are contiguous



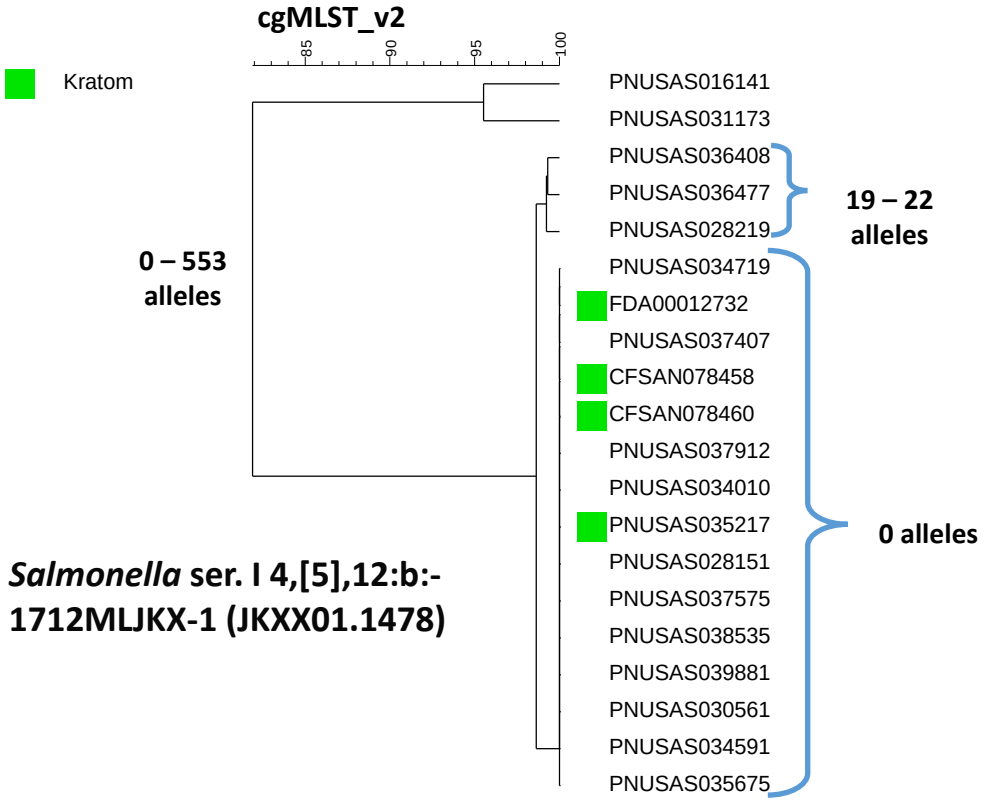
- Epidemiological data and other metadata more critical than ever for WGS data interpretation



Salmonella Outbreak Associated with Kratum Consumption/Use in the U.S. 2018

- ~ 200 cases
- 6 serotypes:
  - **I 4,[5],12:b:-**
  - Thompson
  - **Okatie**
  - Javiana
  - Heidelberg
  - Weltevreden

Kratom, Thang, Kakuam, Thom, Ketom, and Biak

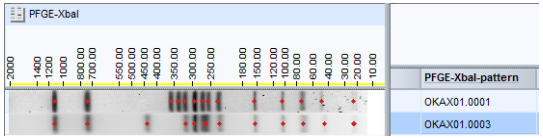


- ***Salmonella* ser. I 4,[5],12:b:- 1712MLJKX-1 (JKXX01.1478)**

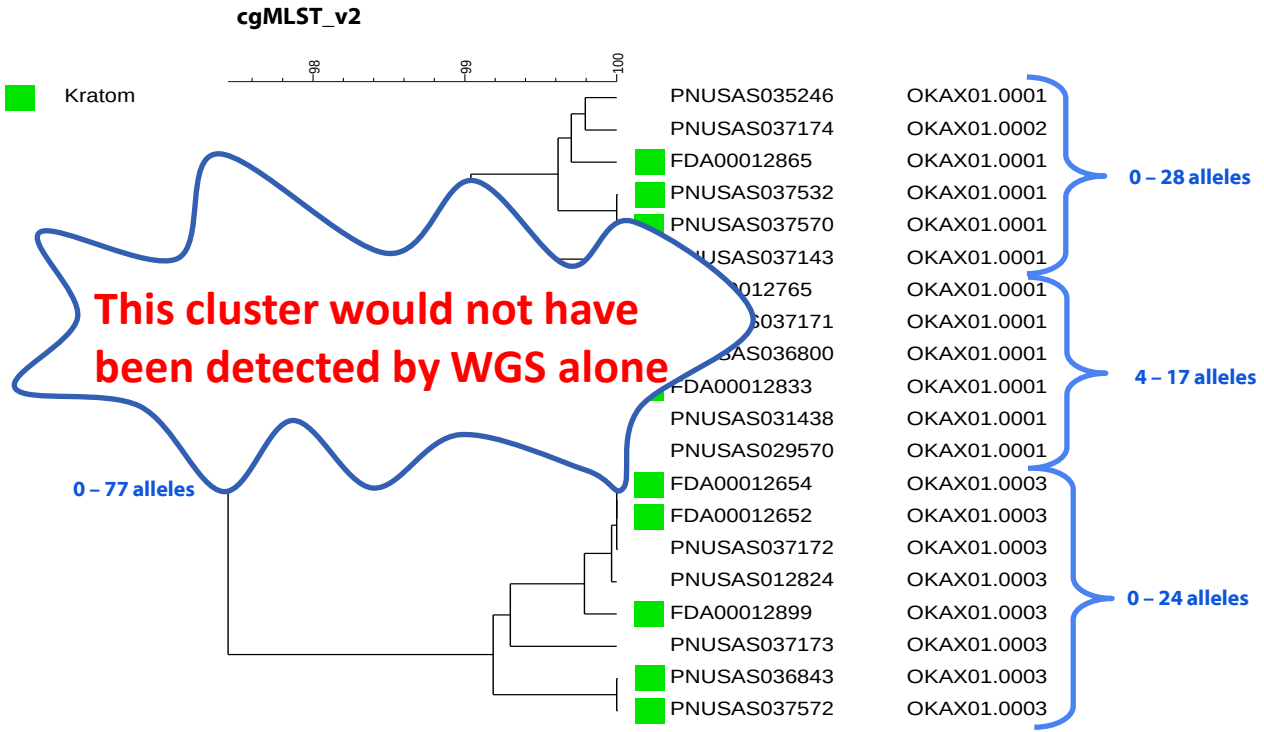
The methods used in the analysis of this sequence data are preliminary and remain under validation.

Salmonella Outbreak Associated with Kratum Consumption/Use in the U.S. 2018

***Salmonella* ser. Okatie**
OKAX01.0001, OKAX01.0003



Kratom, Thang, Kakuam, Thom, Ketom, and Biak

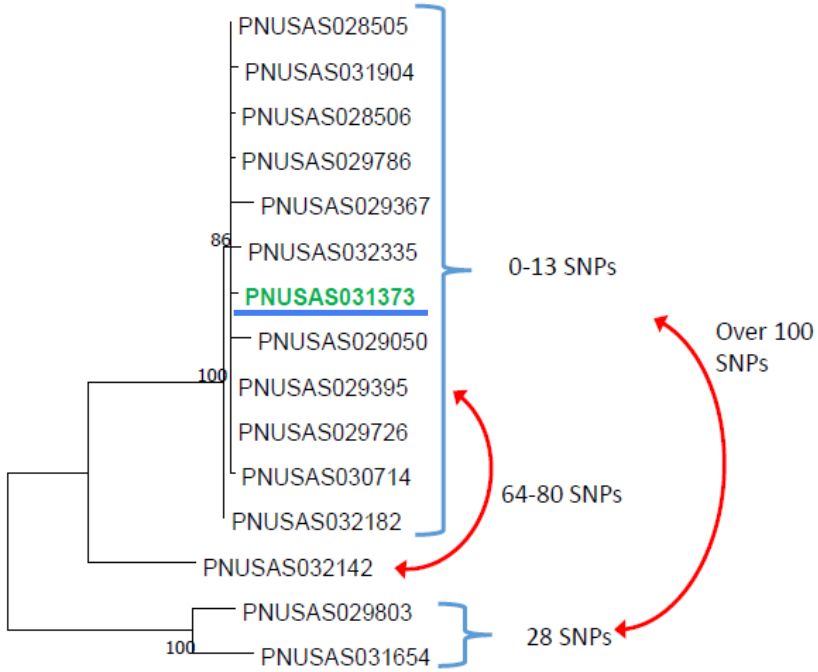


The methods used in the analysis of this sequence data are preliminary and remain under validation.

Don’t let the WGS data fool you!
All supporting information must always be considered

Salmonella ser. Typhimurium strain **from** Egg Nog clustering with isolates from outbreak associated with laboratory exposure, 2017

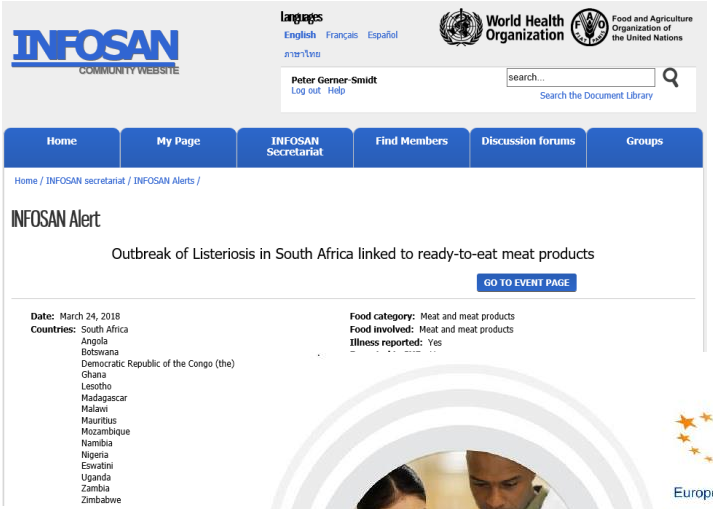
WGS_id	Key	SourceSite	UploadDate
PNUSAS032142		Stool	1/11/2018
PNUSAS031904		Stool	1/3/2018
PNUSAS028505		Stool	11/21/2017
PNUSAS032182		Stool	1/5/2018
PNUSAS029395		STOOL	11/22/2017
PNUSAS029726		STOOL	11/30/2017
PNUSAS029050		Stool	11/17/2017
PNUSAS031373		Egg Nog	12/22/2017
PNUSAS032335		stool	12/22/2017
PNUSAS028506		Stool	11/21/2017
PNUSAS029786		Stool	11/21/2017
PNUSAS029803		Stool	12/4/2017
PNUSAS031654		Stool	12/8/2017
PNUSAS030714		Stool	12/14/2017
PNUSAS029367		Stool	11/28/2017



- wgMLST coincides with hqSNP
- **Resfinder:** No resistance genes

The Challenge of Data Sharing

International Outbreak Investigations Using WGS



efsa
European Food Safety Authority

ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

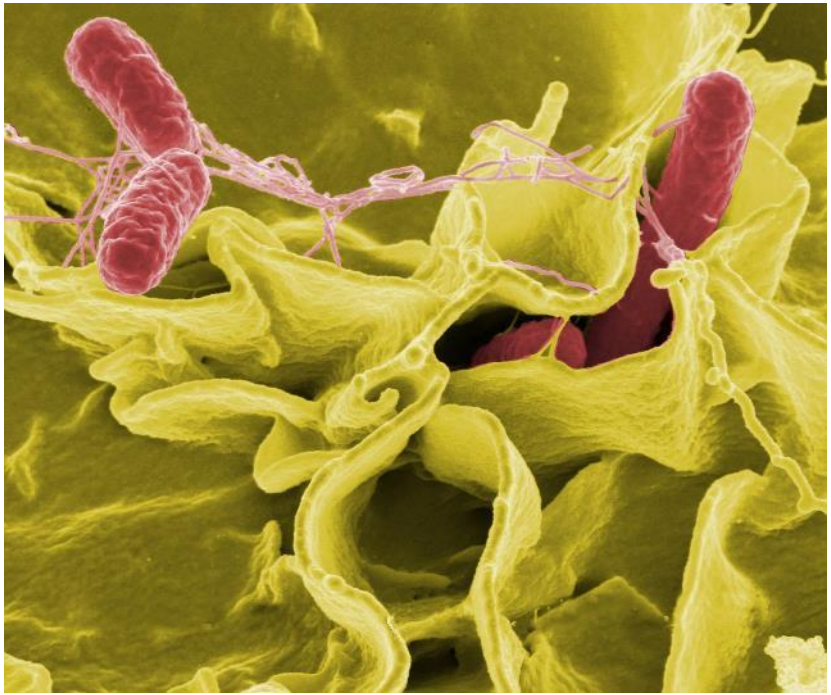
The Challenge of Data Sharing

- **WGS data should be publicly available in real time**
  - **SRA, ENA and the DNA Data Bank of Japan**
  - **Minimum epidemiological data – time, place and type of isolate**
- **Barriers**
  - **Ethics: Personal identifiable information**
  - **Intellectual property and other legal issues**
- **Food industry concerns**
  - **No "statute-of-limitations" on liability**
  - **No precise definition of "outbreak"**
  - **No international interpretation standards ➡ misinterpretation of data**
  - **Trade implications**

WGS: Concerns Remaining

- WGS turnaround time issues
  - Still long (~ 7 work days)
- Cost
- Cluster triage
  - Not resources to investigate all outbreaks
  - Which should be investigated?
- Culture-independent diagnostic testing (CIDT)
  - We are losing the isolates!

Coming Soon: Big Data to Improve Food Safety



- Pathogen characterization direct-from-specimen (faster) - Metagenomics
- Linking data from different sources, incl. non-lab data

  = More information to inform policy

But

- Privacy issues
- Regulatory hurdles
- Data capacity issues

Acknowledgements

Public Health
Agency of Canada

Disclaimers:

"The findings and conclusions in this presentation are those of the author and do not necessarily represent the official position of the Centers for Disease Control and Prevention"

"Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention or by the U.S. Department of Health and Human Services."

thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
merci
obrigado
dziękuję
bedankt
спасибо
go raibh maith agat
mochchakkeram
sukriya
kop khun krap
terima kasih
grazie
arigatô
dakujem
takk
tapadh leat
мерси
감사합니다
شكرا جزيلا
ありがとう
ευχαριστώ
dank je
mauruuru
hvala
sagolun
taiku
mersi
barka
welalin
tack
spas
kia ora
vinaka
blagodaram
misaotra
matondo
paldies
grazzi
mahalo
хвала
asante
manana
obrigada
tenki
murakoze
chokrane
mamnun
дякую
djiere dieuf
tau
akun
dankon
aciù
chnorakaloutioun
gratias ago
gràcies
sulpáy
sobodi
dèkuji
mèsi
didi madloba
najis tuke
kam sah hamnida
rahmat
tanemirt
rahmet
diolch
dhanyavadagalu
shukriya
mercé
trugarez
xièxie
kiitos
dankie
dhanyavad
nanni
nandri
bayarlalaa
gràçie
köszönöm
enkosi
Баярлалаа
faafetai lava
рахмат
спасибі