

WGS of Cholera: Building Global Capacity for Molecular Subtyping of Enteric Bacteria through PulseNet International

15 April 2019 – Annecy, France

Maryann Turnsek and Michele Parsons

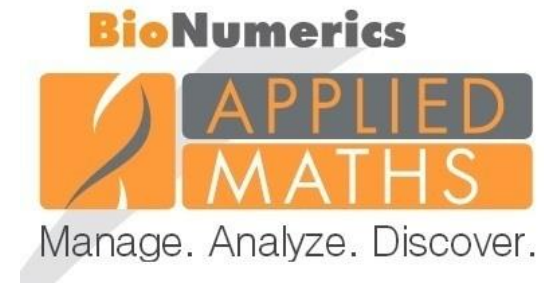
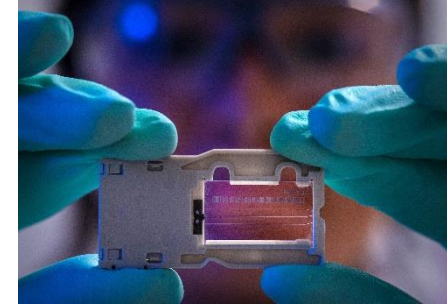
Centers for Disease Control and Prevention, Atlanta USA

Objectives

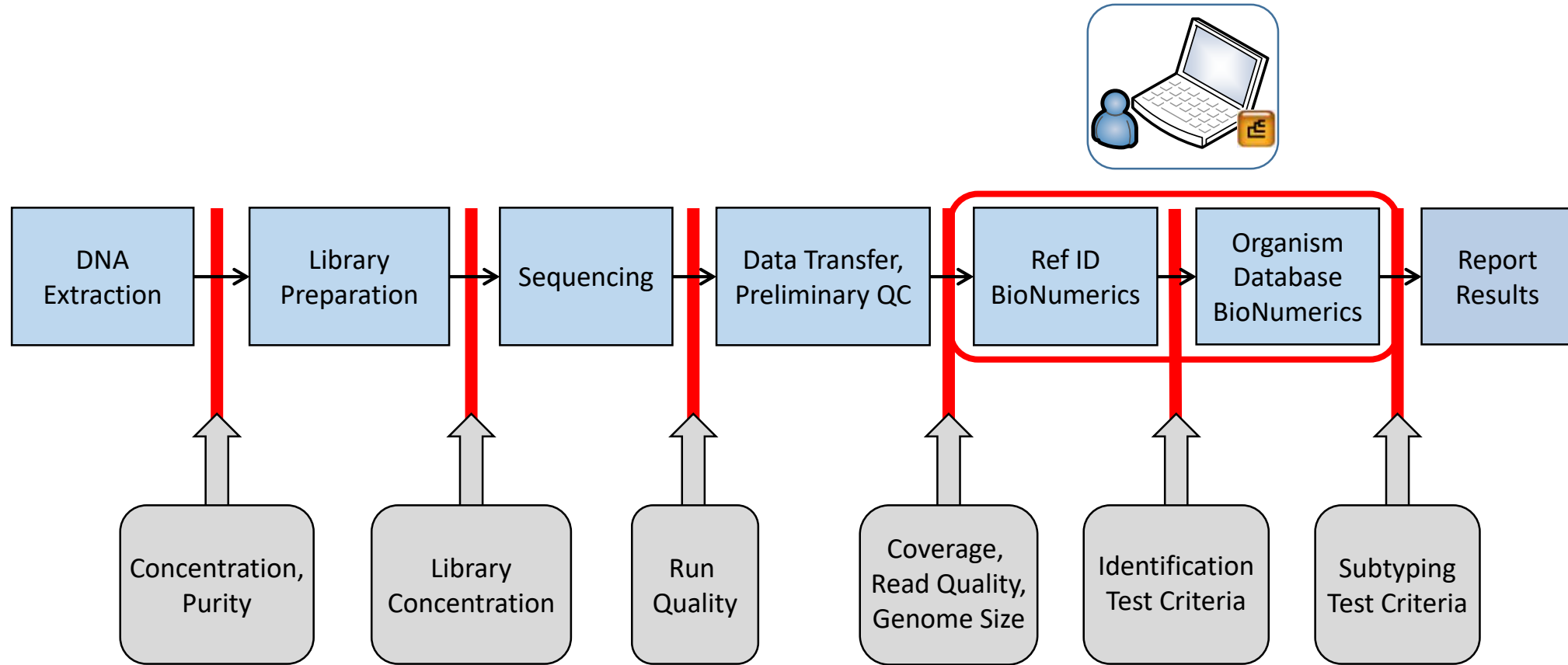
- Provide an overview of approaches and activities in PulseNet International to sequence enteric bacteria, including *Vibrio cholerae*
- Discuss challenges for implementation and data sharing regarding WGS

Overview of approach for sequencing *Vibrio cholerae*

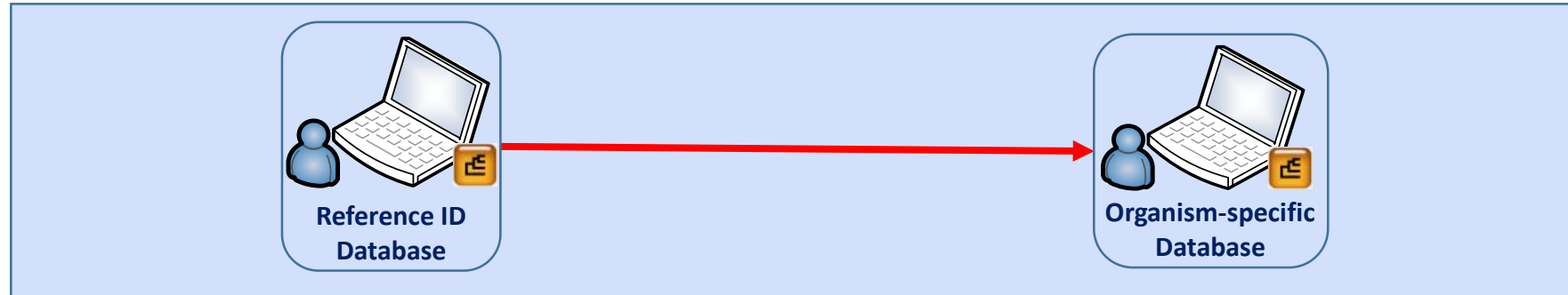
- Applications for sequence data
 - Outbreaks
 - Surveillance
 - Research/special studies
- Preferred methods for WGS
 - Illumina chemistry and platforms
 - BioNumerics 7.6 or higher for analysis
- Managing sequence data
 - Raw reads maintained by CDC
 - Data uploaded in real time to Sequence Read Archive (SRA) at NCBI
 - Fastq files, limited metadata
 - *Vibrio* bioproject identifier: PRJNA266293



Standardized WGS Workflow in Enteric Diseases Lab Branch (EDLB), CDC



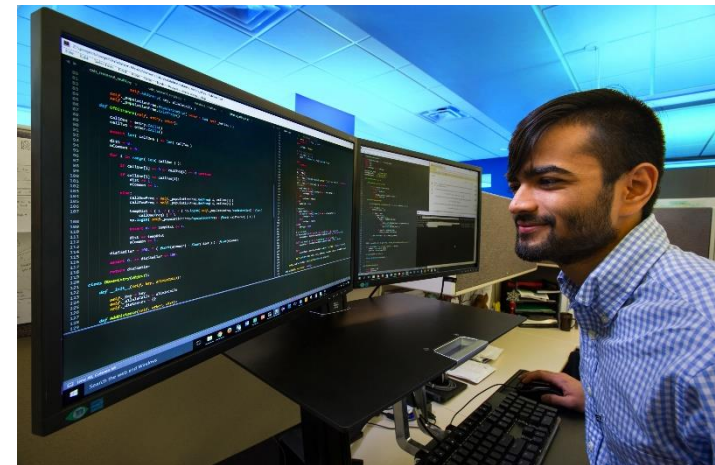
BioNumerics version 7.6



- Assembly
 - Trimming, quality checks
 - Average Nucleotide Identity (ANI)
 - 'O' group determination
 - Serotype, biotype
 - Virulence profiling
 - Resistance prediction
 - cgMLST/wgMLST for surveillance
- ✓ *Vibrio* allele database is being validated and will be finalized in 2020.

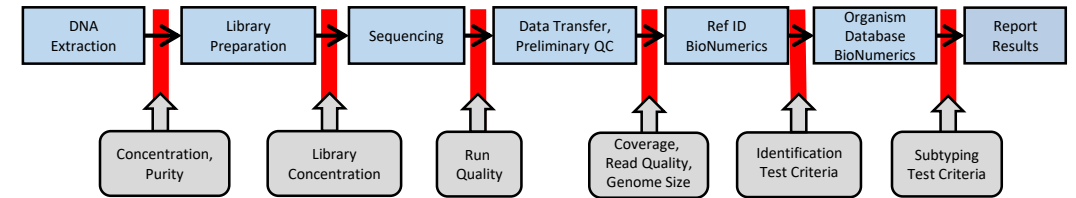
Why use MLST to analyze and compare bacterial genomes?


- Stable nomenclature
- Computationally easier and faster than other methods
- No need for a reference strain



Why use BioNumerics Software for WGS data?

- Standardized way to manage and analyze sequence data
- ‘One shot’ reference characterization within customized organism-specific databases
- Does not require advanced bioinformatics skills



 DEPARTMENT OF HEALTH AND HUMAN SERVICES Public Health Service

Centers for Disease Control
and Prevention (CDC)
Atlanta GA 30333

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*

Customized for NCBI Submissions

NCBI submission settings

BioProject and organisation
Provide general information for all submissions: what BioProject do they belong to, and who is submitting.

BioProject accession:

Organisation name:

SPUID namespace:

Type:

Role:

Contact first name:

Contact last name:

Contact e-mail:

FTP upload directory:

< Back Next > Cancel

BioProject accessions:

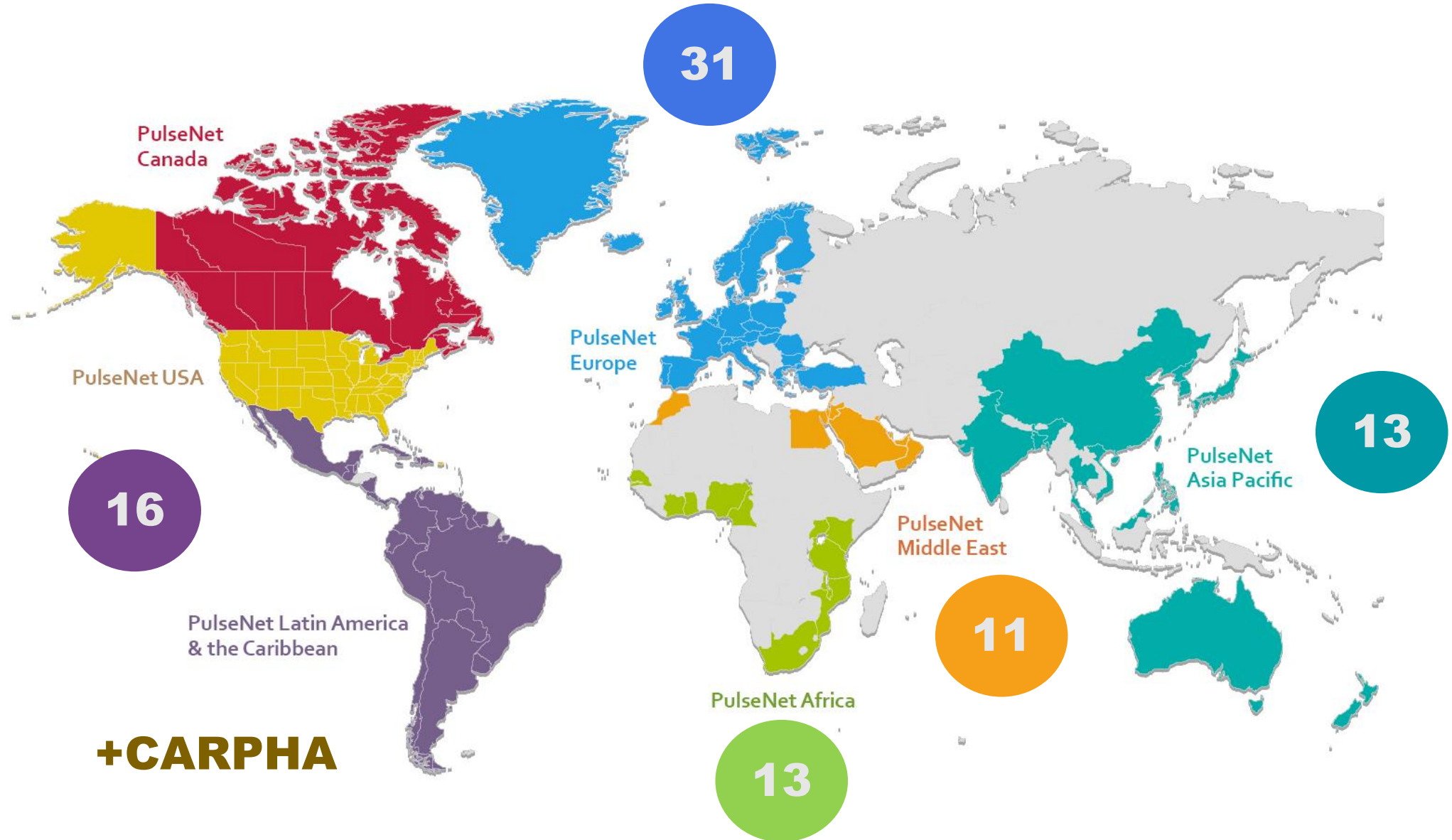
Campylobacter: PRJNA239251
E. coli and *Shigella*: PRJNA218110
Listeria: PRJNA212117
Salmonella: PRJNA230403
***Vibrio*: PRJNA266293**

PulseNet International – Implementing WGS for Foodborne Disease Surveillance

- **Why are PulseNet International lab moving towards WGS?**
 - WGS is more precise and provides more information than current “gold standard” subtyping techniques, including pulsed-field gel electrophoresis (PFGE) and multi-locus variable number tandem repeat analysis (MLVA)
 - Data can be shared across all laboratories for routine surveillance, outbreak identification, antimicrobial resistance prediction and reference characterization

<https://www.cdc.gov/pulsenet/participants/international/wgs-vision.html>

PulseNet International



PulseNet International WGS Trainings

- Sequencing on the MiSeq Platform – Thailand, 2019



- Analysis methods – Kenya, 2019



Summary of WGS and Analysis Trainings Lead by CDC

Date	WGS wet lab	BioNumerics 7.6	Training site	Countries	#
November 2015		X	PNLAC meeting, La Paz Mexico	Chile, Paraguay, Costa Rica, Mexico, Peru and Columbia, Argentina, Ecuador	25
March 2016	X	X	CDC-Atlanta	Chile	1
November 2016		X	CDC-Atlanta	Costa Rica	3
March 2017	X		InDRE	Mexico, Cuba	10
March 2017		X	InDRE	Mexico	5
June 2017	X	X	CDC-Atlanta	Kenya	1
July-August 2017	X	X	Chile Institute of Public Health	Chile, Paraguay, Brazil, Peru and Argentina	9
November 2017	X	X	CDC-Atlanta	Argentina	2
June 2018	X	X	CDC-Atlanta	Lebanon and Israel	2
July 2018	X	X	KEMRI	Kenya	18
February 2019	X	X	CDC-Thailand	Thailand	15

Challenges Implementing WGS in Countries with Highest Cholera Burden

- **Availability of reagents** – limited ordering options (usually only regional distributors), procurement delays common
- **Instrument care** – sequencing platforms are sensitive, maintenance contracts are highly recommend, but expensive and require highly trained technicians
- **IT infrastructure** – improvements needed to handle / store volume of sequence data generated, ensure security of data
- **Availability of analytical tools** – limited “user friendly” and affordable options
- **Internet speeds** – slow internet speeds limit analysis, data transfer
- **High cost** – depending on Illumina platform and chemistry (capacity ranges from 2 – 40 genomes), \$1 000 - \$4 000 USD / run, \$100 - \$500 USD / sample (not including cost to analyze or store data)

Data sharing

- At trainings, we encourage standard policy for data sharing
 - ✓ Real time release to SRA in NCBI with limited metadata
- For requests, we share interpretation (e.g. phylogenetic tree)
 - ✓ Important to see what the data looks like
- From our experience, there is a great deal of mis-trust of genetic data being in the public realm
 - ✓ Lead by example

CDC Laboratory Contacts for GTFCC

- Michele Parsons, PhD
 - ypi5@cdc.gov
- Maryann Turnsek, BS
 - hud4@cdc.gov
- Cheryl Tarr, PhD
 - crt6@cdc.gov

Acknowledgements:

- Enteric Diseases Laboratory Branch (EDLB) in the Division of Foodborne, Waterborne and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, CDC
- Epidemiology, Informatics, Surveillance and Laboratory Branch (EISLB) in the Division of Global Health Protection, Center for Global Health, CDC

PulseNet International Contacts

- Jeny Concepción-Acevedo, PhD
 - Coordinator, PulseNet International
 - ypi5@cdc.gov
- Peter Gerner-Smidt, MD PhD
 - Chair, PulseNet International Steering Committee
 - plg5@cdc.gov

