

CDC PUBLIC HEALTH GRAND ROUNDS



Accessible Version: <https://www.youtube.com/watch?v=jpXROWP6V74>

The Emerging Role of Pathogen Genomics in Public Health

January, 21 2020



U.S. Department of
Health and Human Services
Centers for Disease
Control and Prevention

Continuing Education Information

Continuing education: www.cdc.gov/getce

- After creating a TCEO account, click the “Search Courses” tab on the left and use “Public Health Grand Rounds” as a keyword search.
- All PHGR sessions eligible for CE should display, select the link for today’s session and then Continue button. **Course Access Code is PHGR10.**
- CE expires **February 21, 2020** for live and **February 21, 2022** for Web On Demand courses.
- Issues regarding CE and CDC Grand Rounds, email: tceo@cdc.gov

Disclosure Statement

CDC, our planners, presenters, and their spouses/partners wish to disclose they have no financial interests or other relationships with the manufacturers of commercial products, suppliers of commercial services, or commercial supporters.

Content will not include any discussion of the unlabeled use of a product or a product under investigational use.

CDC did not accept commercial support for this continuing educational activity.

Brand names, images, and trademarks are used for identification purposes only.

Public Health Grand Rounds Resources

Send comments or questions to:
grandrounds@cdc.gov

[youtube.com/user/
CDCStreamingHealth](https://youtube.com/user/CDCStreamingHealth)

Access full
PHGR sessions &
Beyond the Data

facebook.com/CDC

Like CDC's Facebook
page to stay
informed on all
things public health

Visit our website at: www.cdc.gov/grand-rounds

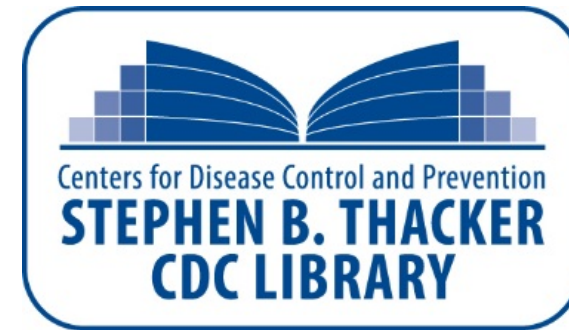
Additional Resources

Beyond The Data

**“Take home” messages in a short podcast at:
cdc.gov/grand-rounds**



scienceclips
CDC



**Scientific publications about
this topic at:
cdc.gov/library/sciclips**

Email grandrounds@cdc.gov with any questions or for help locating the additional resources

Must Reads

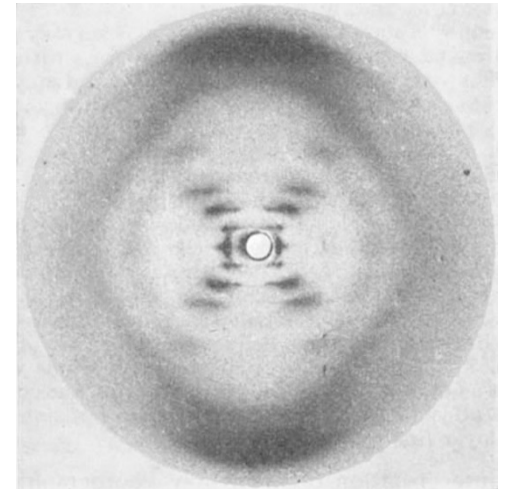
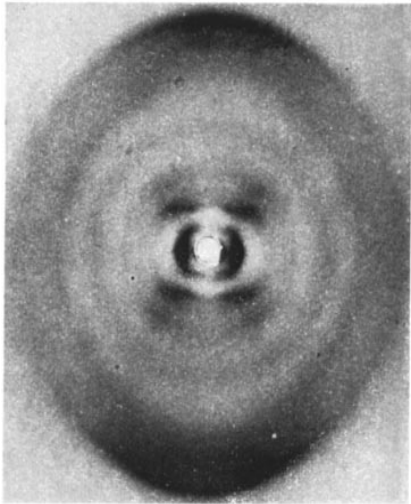
April 25, 1953

NATURE

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.



Today's Speakers and Contributors



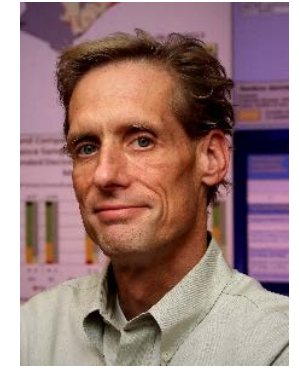
Duncan MacCannell, PhD



Heather Carleton, MPH PhD



Lauren Turner, PhD



Gregory Armstrong, MD

Acknowledgments

Luciann Draper
Luis Luque
Paula Eriksen
Steve Mann
Ann Gallaty

Alex Alvarez
Charlotte Dugan
Patrick Brady
Nathelia Barnes
Tom Race

Muin Khoury
William Thomas
Marta Gwinn
Elizabeth Neuhaus
Scott Sammons

Upcoming Programs of Interest

February 18, 2020

Public Health Grand Rounds

Measles

March 17, 2020

Public Health Grand Rounds

Laboratory Response Network

CDC PUBLIC HEALTH GRAND ROUNDS



The Emerging Role of Pathogen Genomics in Public Health

January 21, 2020



U.S. Department of
Health and Human Services
Centers for Disease
Control and Prevention

CDC PUBLIC HEALTH GRAND ROUNDS



The Emerging Role of Pathogen Genomics in Public Health

January 21, 2020



U.S. Department of
Health and Human Services
Centers for Disease
Control and Prevention

Introduction to Advanced Molecular Detection and Pathogen Genomics

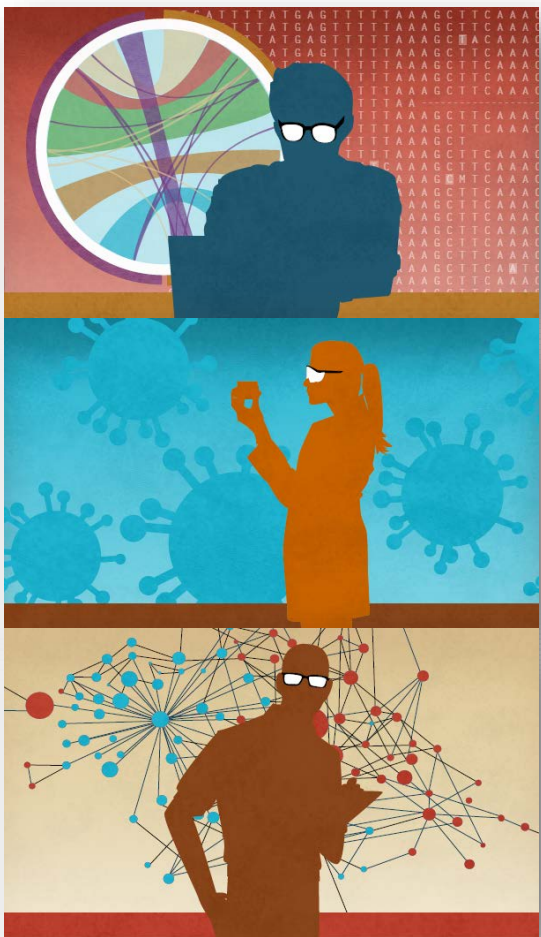


Duncan MacCannell, PhD

Chief Science Officer,

Office of Advanced Molecular Detection
Centers for Disease Control and Prevention

Advanced Molecular Detection (AMD)



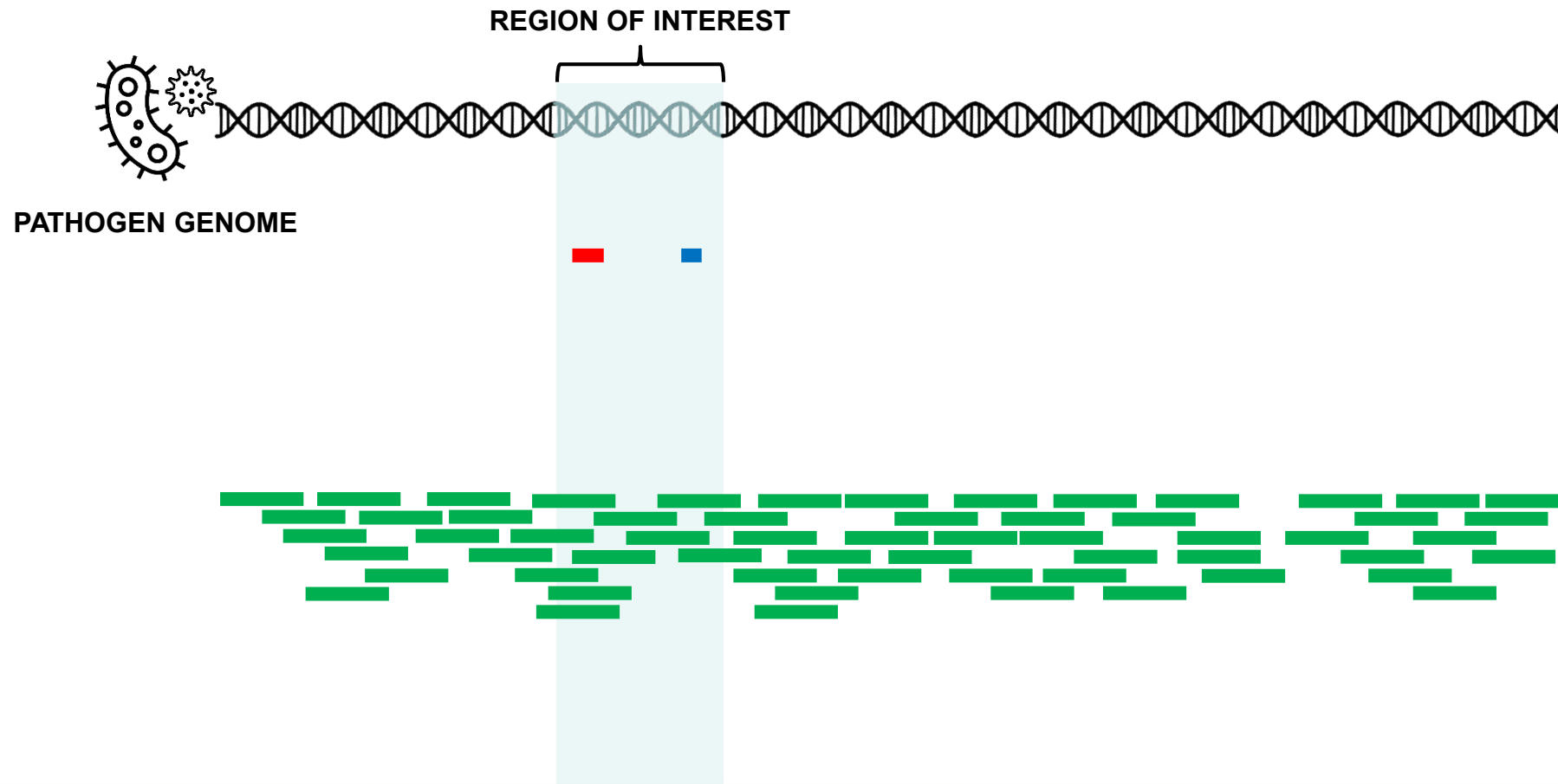
➤ AMD program established by Congress in 2014

- \$30 million per year public health laboratory innovation program
- Focus on transformational laboratory tech and scientific computing

➤ Principal objectives of AMD

- Improve **pathogen detection and characterization**
- Enable **new diagnostic methods** to meet public health needs
- Support **genomic and bioinformatics** needs in the US public health system
- Implement enhanced, sustainable, **integrated information systems**
- Develop tools for prediction, modeling and **early recognition of emerging infectious threats**

Sequencing in Public Health



\$1,000/Mbase
<800bp/target

SANGER SEQUENCING

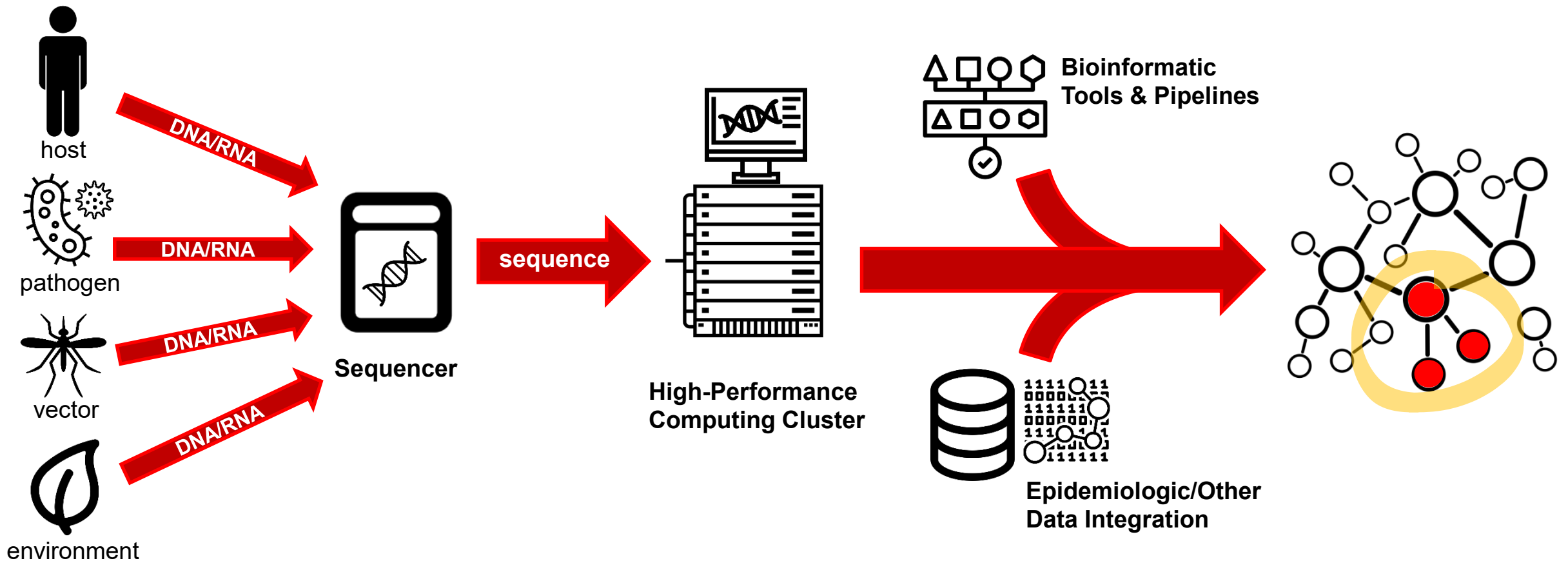


\$0.01/Mbase
75-400bp; millions

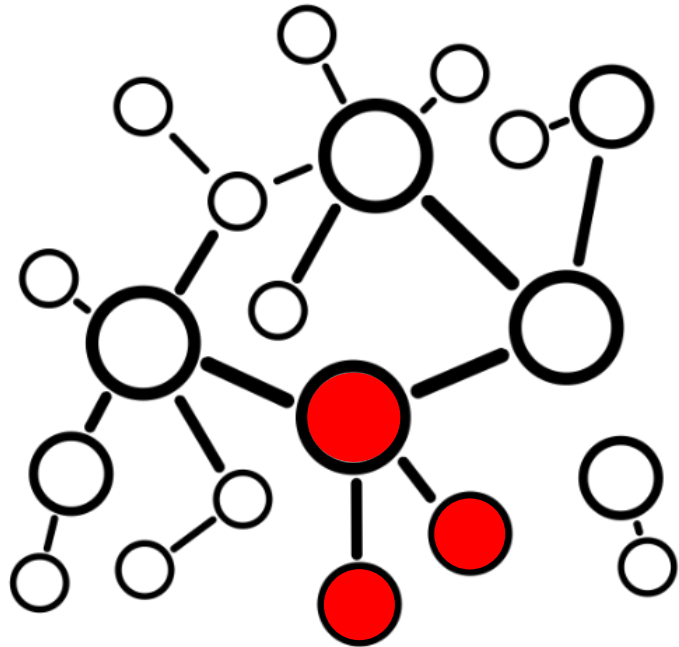
NEXT-GENERATION SEQUENCING

bp = basepairs

How is NGS Used in Public Health?



What Information can be Obtained from Pathogen Genomes?



➤ Pathogen Genomics

- Reference identification
- High-resolution genotyping
 - Clustering and phylogenetics
- Functional annotation and characterization
 - Virulence, antimicrobial resistance, signatures
 - Phenotype prediction
- Diagnostic development and validation
- Minor population variants, quasi-species
- Microbial ecology and diversity

AMD Projects and Applications: 2019



Building Capacity and
Workforce



Exploring New
Technologies



Enhancing Surveillance



Identifying Vector-borne
Diseases



Developing Faster Tests



Uncovering Emerging
Threats



Improving Vaccines



Battling HIV and STDs



Tracking Global Health



Mapping Environmental
Threats



Strengthening Food
Safety



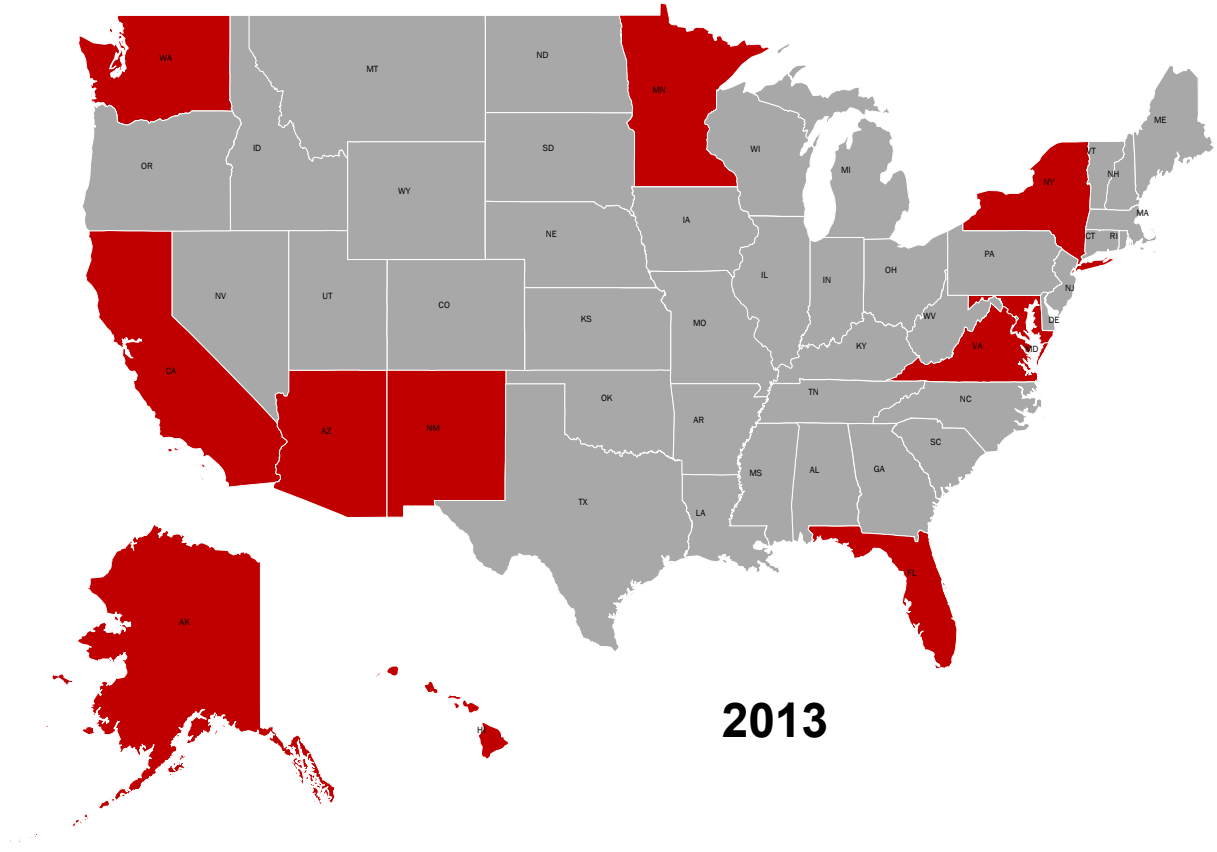
Combating HAIs and
AMR

On the Front Lines: Transforming Influenza Surveillance



Sequencing instrument

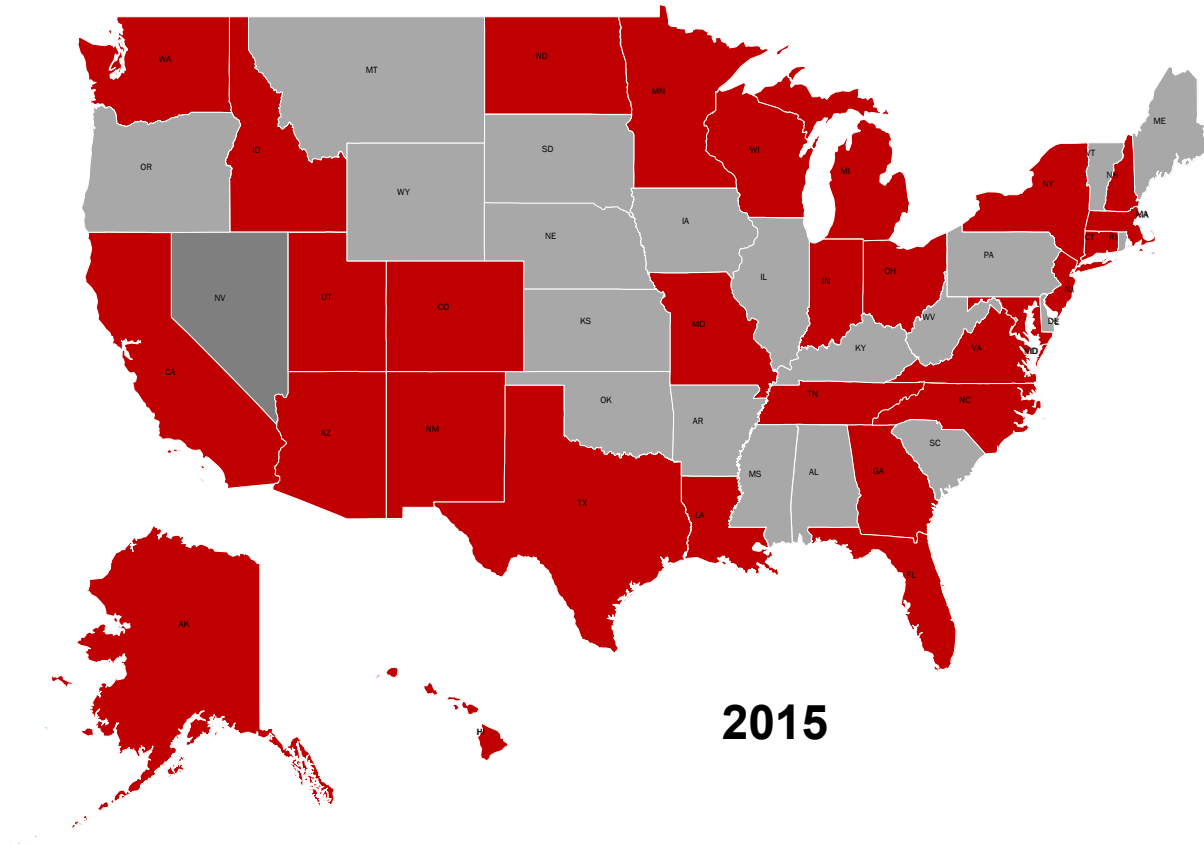
NGS Capacity in State Public Health Laboratories (2013–2018)



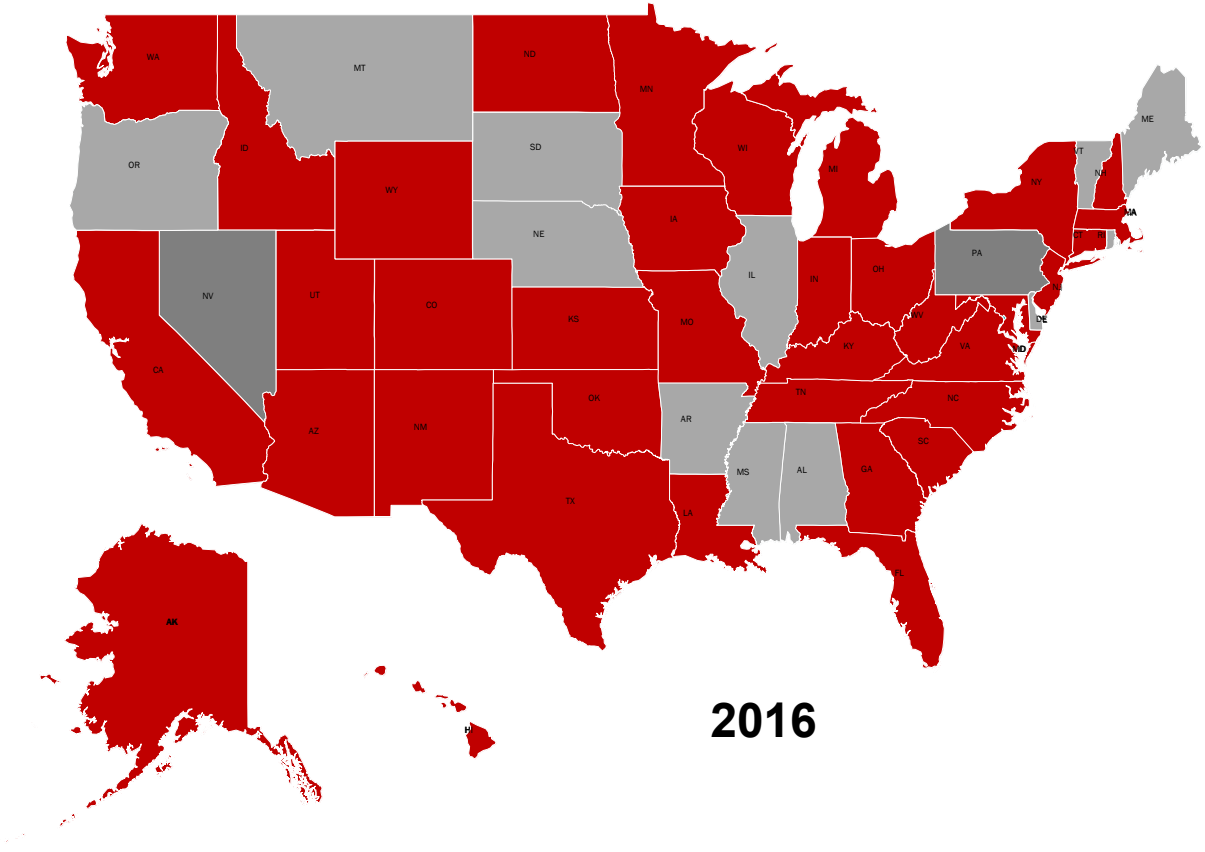
2013

www.aphl.org/aboutAPHL/publications/Documents/ID_NGSSurveyReport_52015.pdf

NGS Capacity in State Public Health Laboratories (2013–2018)

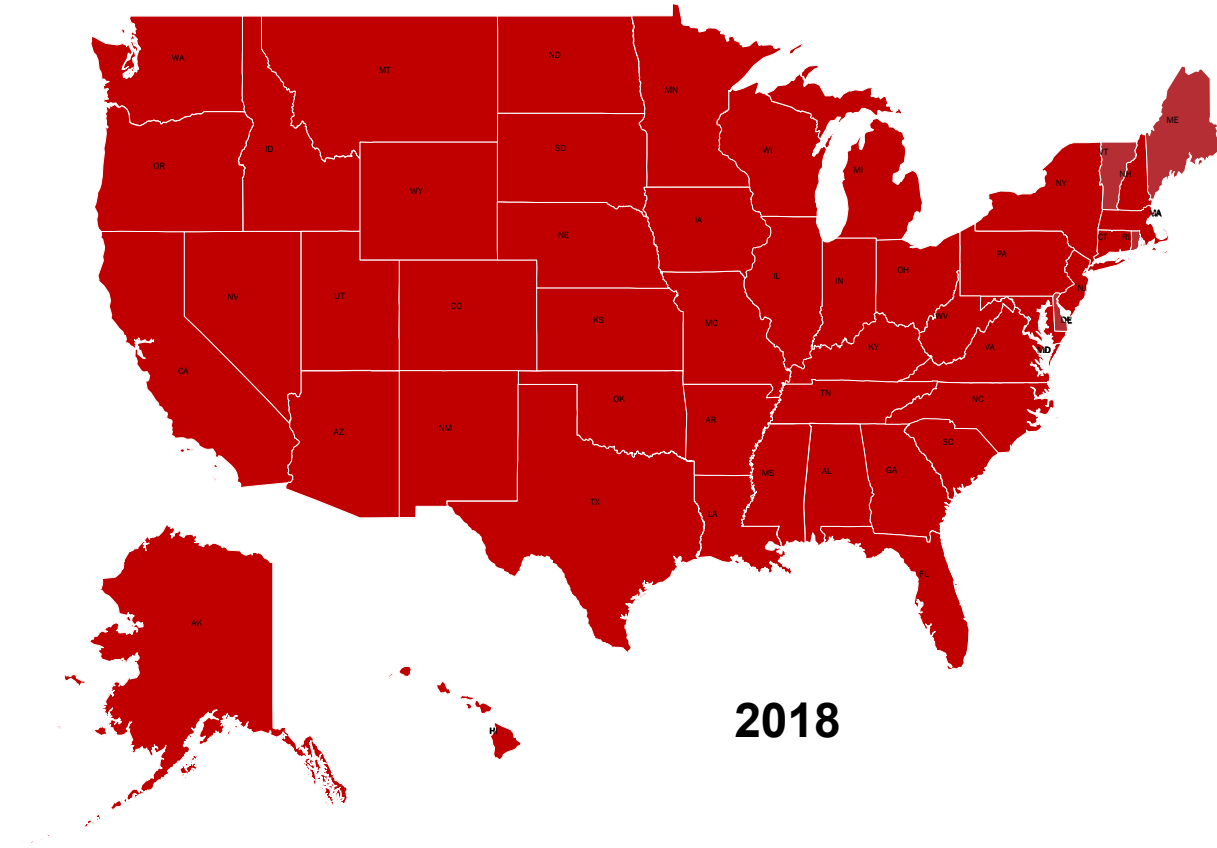


NGS Capacity in State Public Health Laboratories (2013–2018)



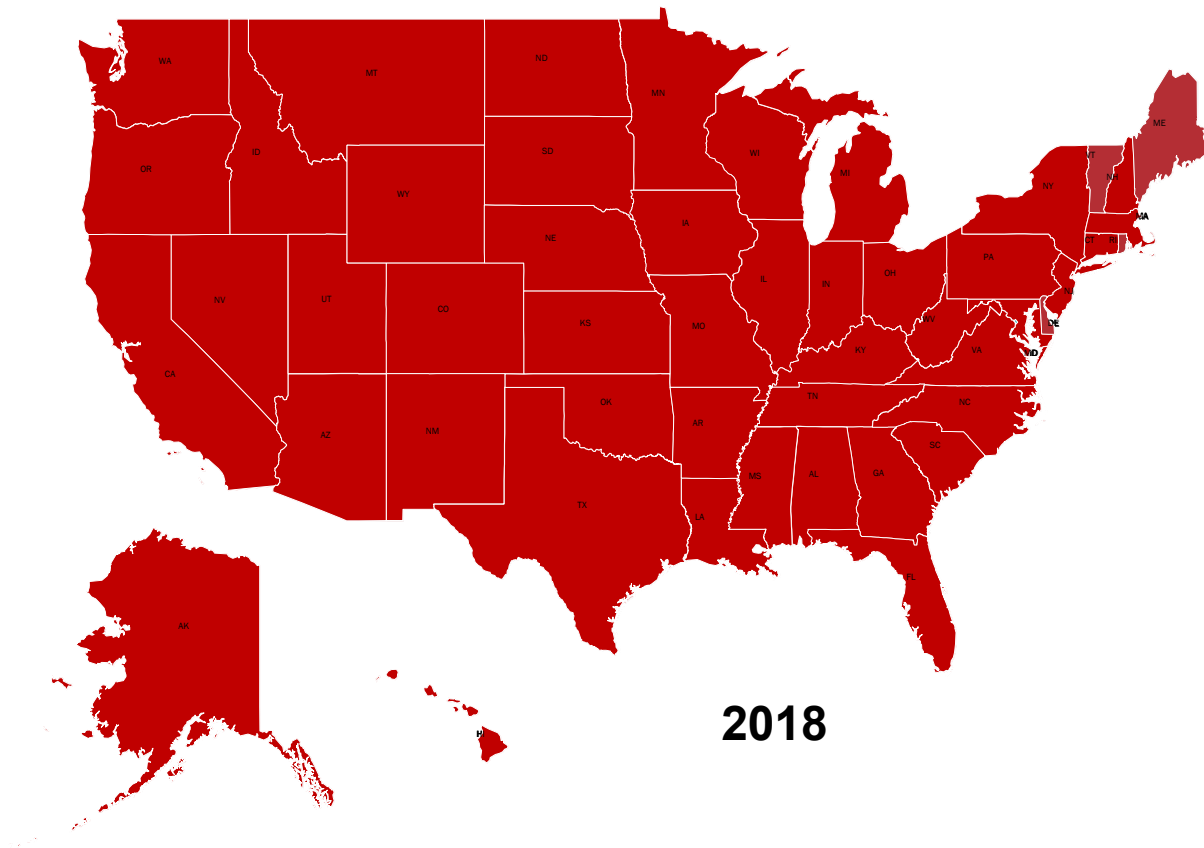
www.aphl.org/aboutAPHL/publications/Documents/ID_NGSSurveyReport_52015.pdf

NGS Capacity in State Public Health Laboratories (2013–2018)



www.aphl.org/aboutAPHL/publications/Documents/ID_NGSSurveyReport_52015.pdf

NGS Capacity in State Public Health Laboratories (2013–2018)



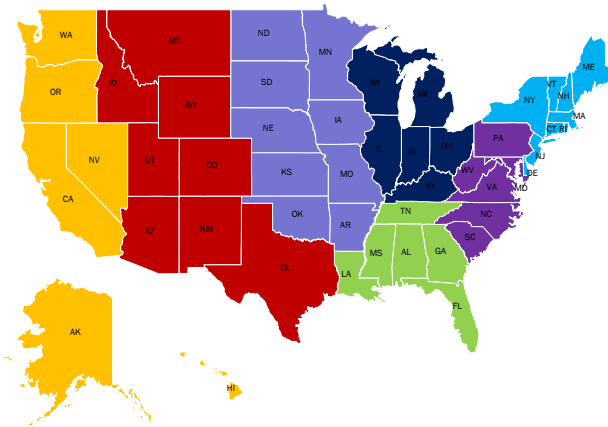
➤ A growing list of applications

- PulseNet
- Healthcare-associated infections
- Bacterial meningitis
- Cryptococcus
- Cyclospora
- **Viral hepatitis**
- Gonococcus
- Influenza
- **Legionella**
- Malaria
- Streptococcus
- Tickborne diseases
- **Tuberculosis**
- To be determined

Building Sustainable Public Health Bioinformatics Capacity



CDC/APHL Bioinformatics Fellows, 2019 Cohort



AMD Bioinformatics Resource and Training Regions

- **Workforce development (recruitment and training)**
 - CDC and APHL Bioinformatic Fellowship Program
 - AMD bioinformatics regional resources and training Leads
 - Building community: state public health bioinformaticians (StaPH-B)
- **Engage across the public health, academic, and private sector communities.**
- **Bridge the gap between existing open source bioinformatics software development and public health users and applications.**

AMD: In Summary

- **AMD program**
 - Working to foster multidisciplinary laboratory innovation in public health
 - Helping to drive shift toward open data, standards, methodological transparency and collaboration.
- **With financial and technical support from CDC and key partners, NGS capacity has spread nationwide over the past 6 years.**
- **HPC infrastructure, bioinformatics workforce capacity, quality, and data management remain critical challenges.**

Applications of Next-Generation Sequencing to Foodborne Disease Prevention



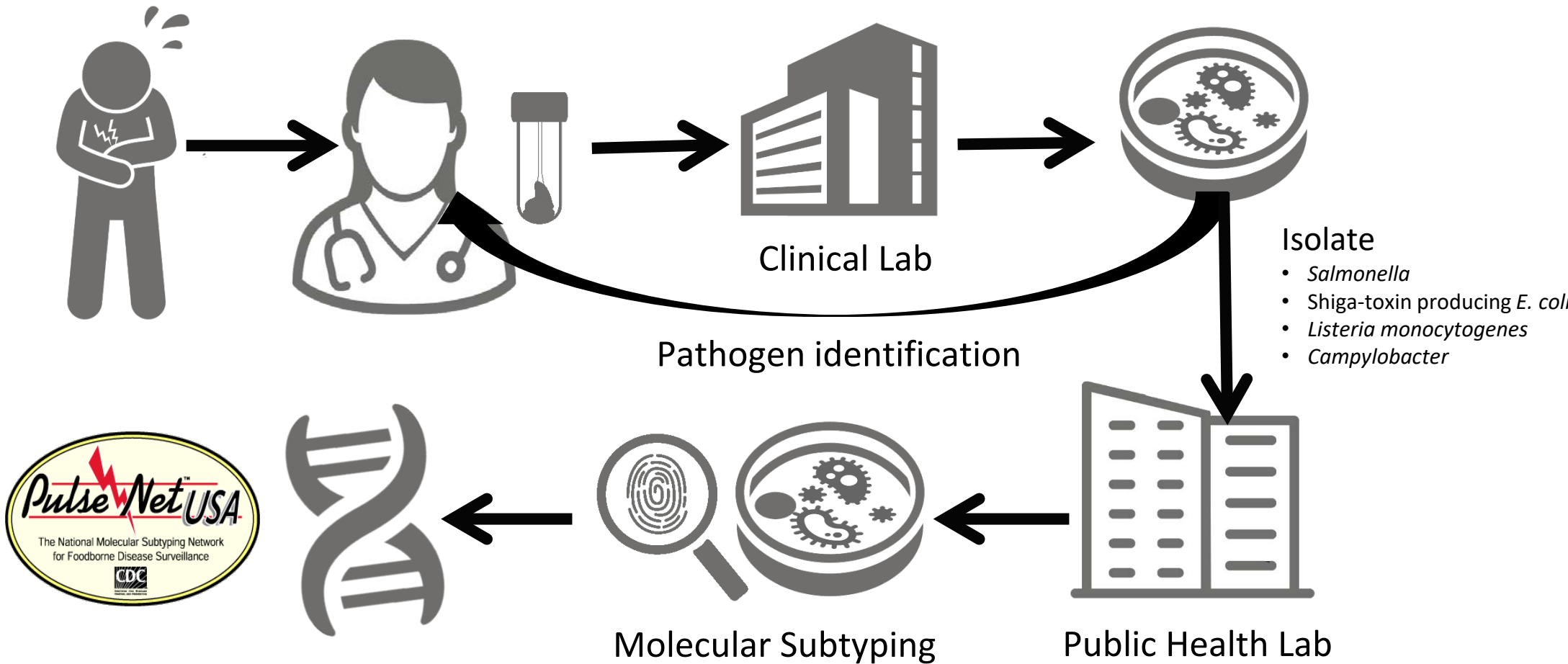
Heather Carleton, MPH, PhD

Bioinformatics and Metagenomics Team Lead

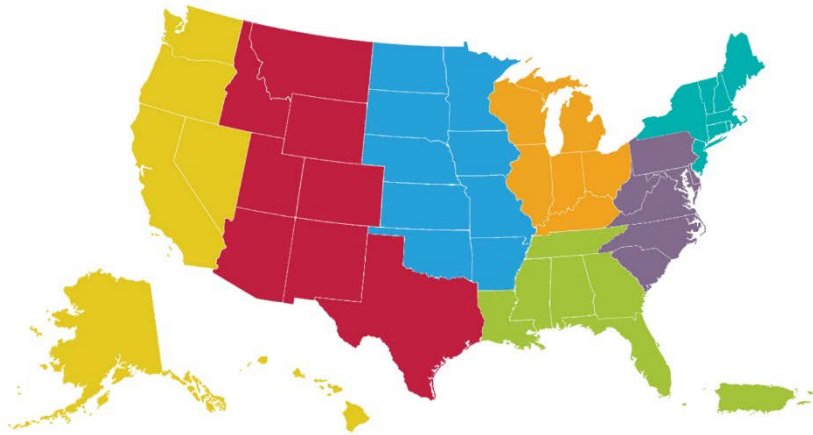
Enteric Diseases Laboratory Branch

Division of Foodborne, Waterborne, and Environmental Diseases

Foodborne illness: From Patient to Public Health Surveillance



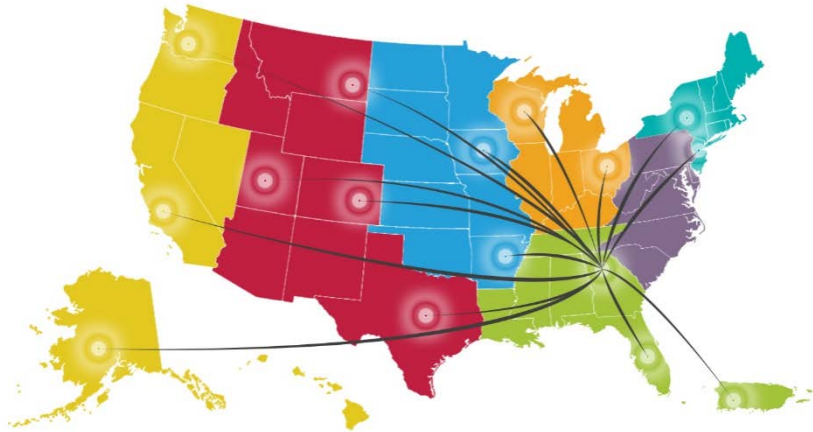
What is PulseNet?



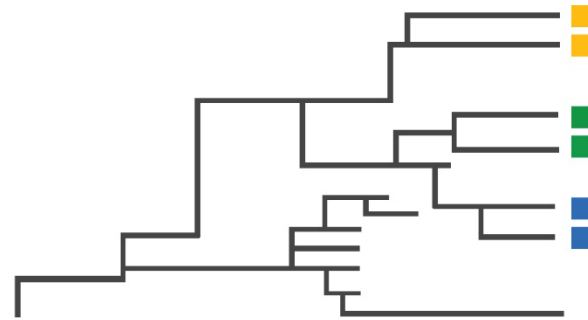
```
TGGTATTTATGCTGCTGGAGGTTCCACAA  
3CCGCGTGCGGAGCCATATTTATTCCCG  
>CGCATACGCCAGGATAATCCAGCATAAA  
3TTCTTTCCGTATGTTTTGCGTTCATTTAA  
GCTGCTGGAGGTTCGTGGTATTTCCACAA  
TTTTGCGTTGAGCCATATTTATTCCCGTC  
>GCTTTATCGCCTAGACAAAACCTGCATAA  
3CCCCGCCGTGTCGGGA A ATGTGTTCAT  
ACCTGCATAAATTGCTTTATCGCCTAGAI  
GGGA A ATGTGTTCAATTA ACATTGCCCC
```

**Public health labs generate
molecular fingerprints using WGS**

What is PulseNet?

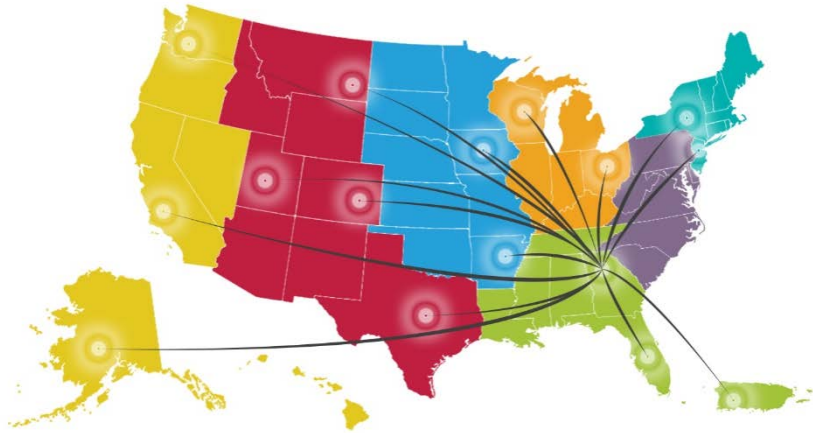


Data from pathogens are transmitted to centralized database at CDC in Atlanta, Georgia

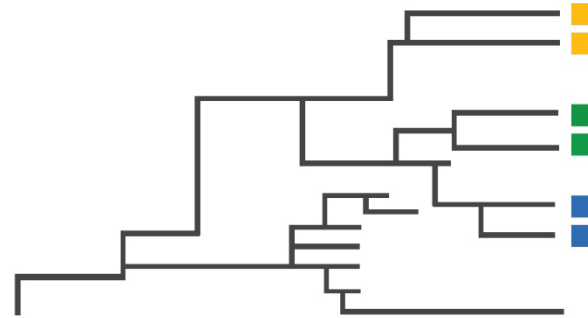


PulseNet Monitors for clusters of illnesses with the same molecular “fingerprint”

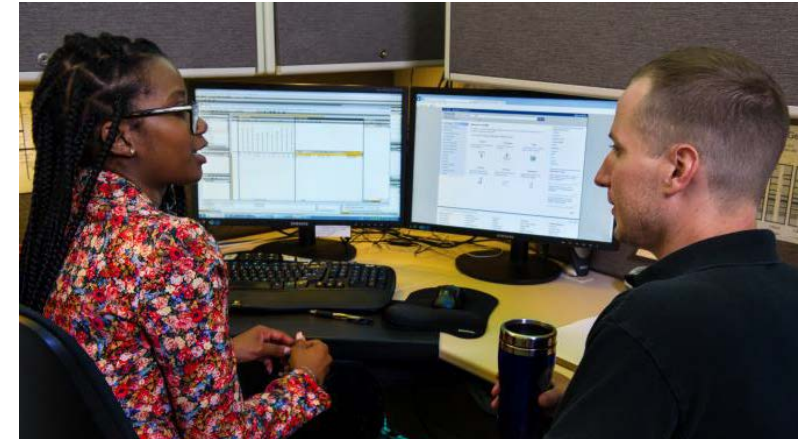
What is PulseNet?



Data from pathogens are transmitted to centralized database at CDC in Atlanta, Georgia

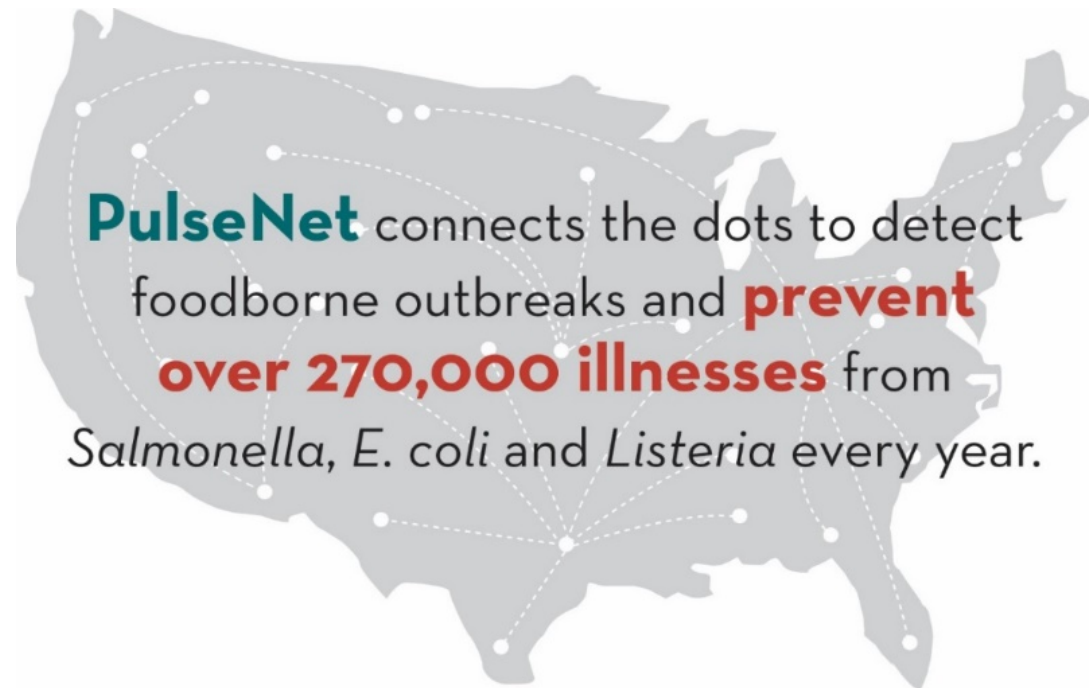


PulseNet Monitors for clusters of illnesses with the same molecular “fingerprint”



Then tells epidemiologists about clusters to investigate

PulseNet 1996-2020: Disease Prevention and Cost Savings

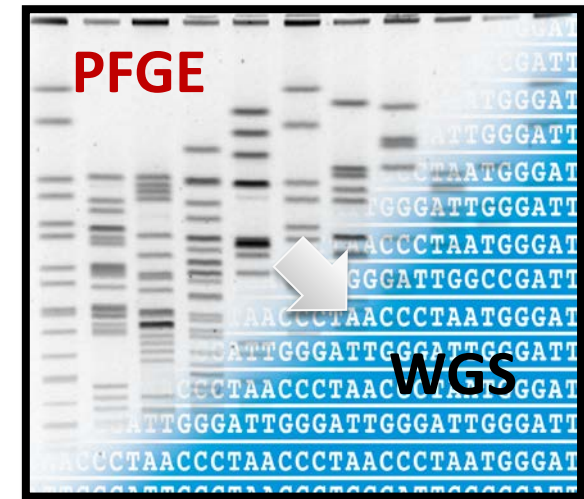


Every year **PulseNet** saves at least **half a billion dollars** in medical costs and lost productivity.



\$1 spent = \$70 saved!

Transition to Use of Whole Genome Sequencing in PulseNet



**Modified from Carleton and Gerner-Smidt (ASM Microbe July 2016)*

Transition to Use of Whole Genome Sequencing in PulseNet



Utility of WGS Data:

- Detecting outbreaks with more precision
- More effective outbreak investigations and trace back

**Modified from Carleton and Gerner-Smidt (ASM Microbe July 2016)*

Putting Together the Picture of a Foodborne Outbreak



**Molecular Surveillance
(PulseNet)**



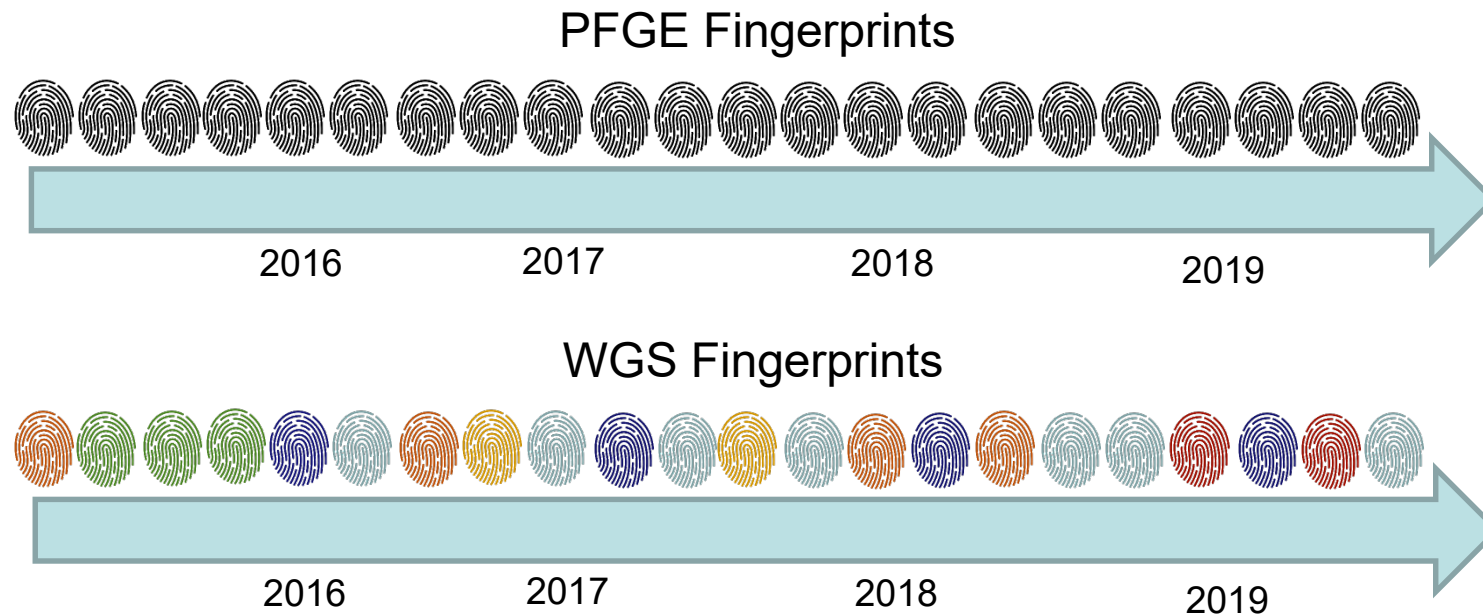
**Epidemiology
(case interviews)**



Trace back to source

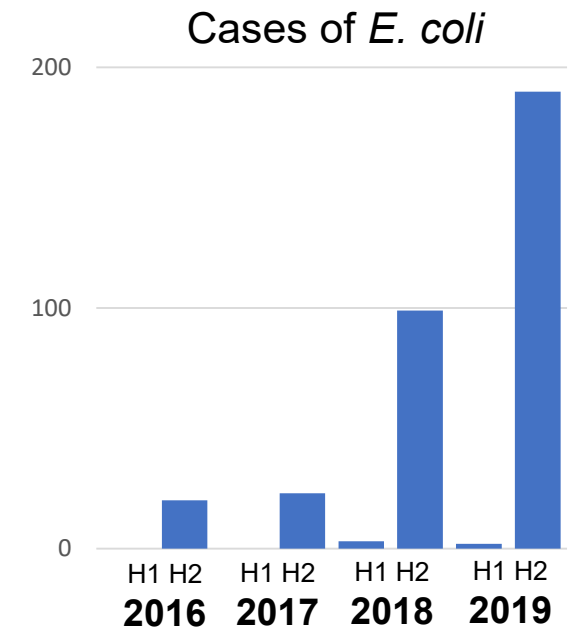
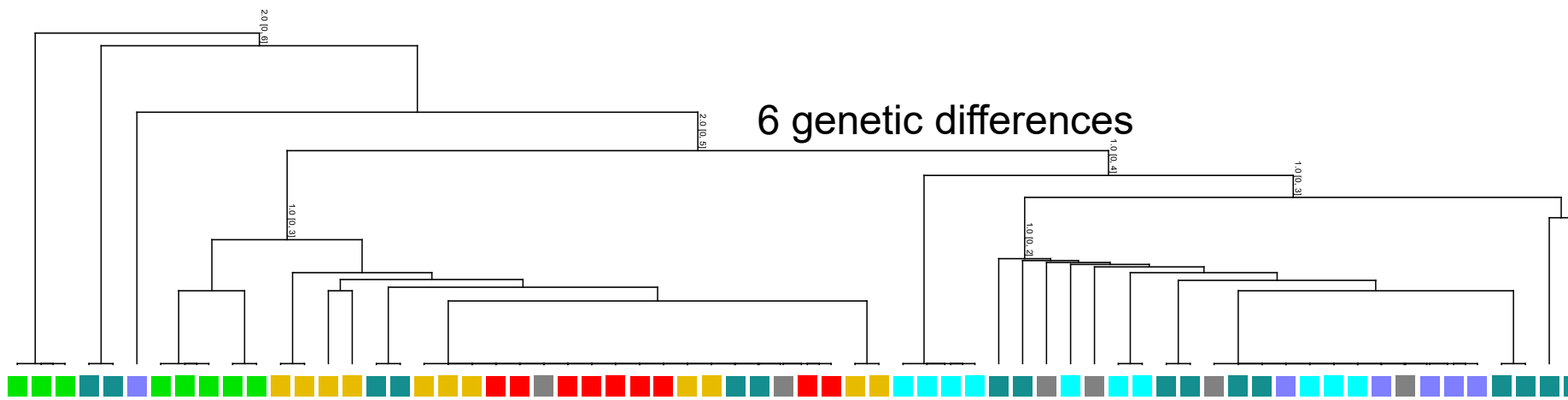
Using WGS to Track Reoccurring, Emerging, and Persisting Strains

WGS links cases that are associated with reoccurring, emerging, and persisting strains



Using WGS to Track Reoccurring, Emerging, and Persisting Strains

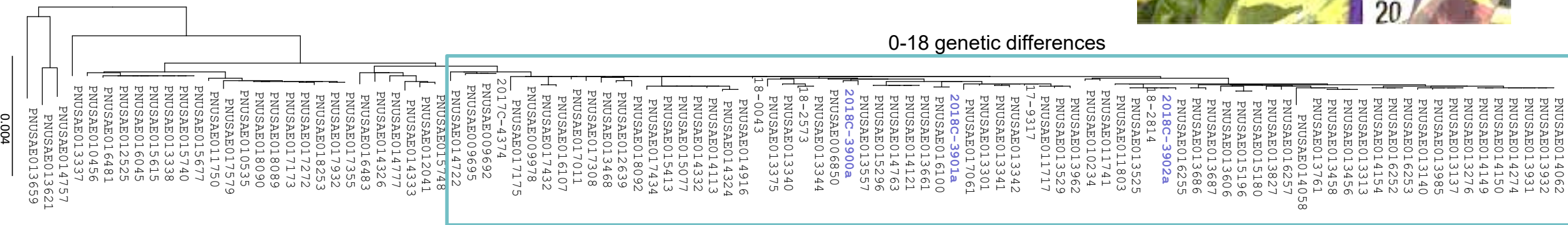
- WGS linked cases of *E.coli* O157:H7 that occurred over a 4-year period
- Linked to a single growing region



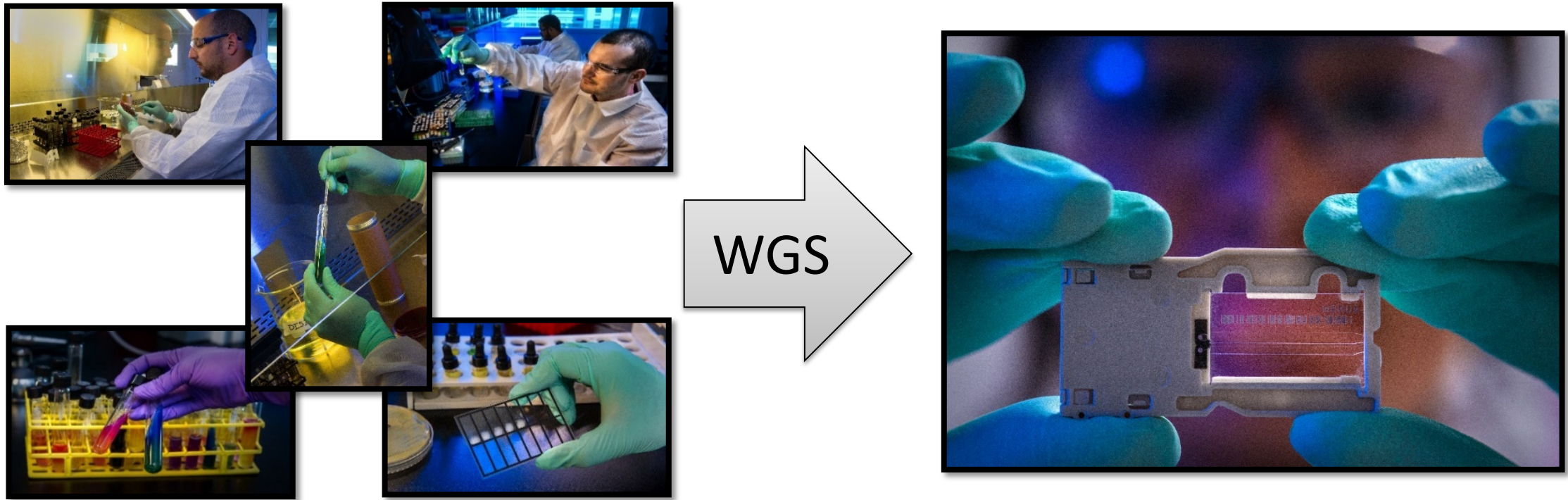
figures courtesy of Kane Patel

Using WGS to Help Outbreak Prevention Efforts

- *E. coli* O157 infections were linked to romaine lettuce from Yuma growing area
- *E. coli* O157 from water matched outbreak strains
- Led to new public health action
 - Requirement to sanitize water used for overhead irrigation during the past 21 days of production
 - Label bag with growing region and harvest date



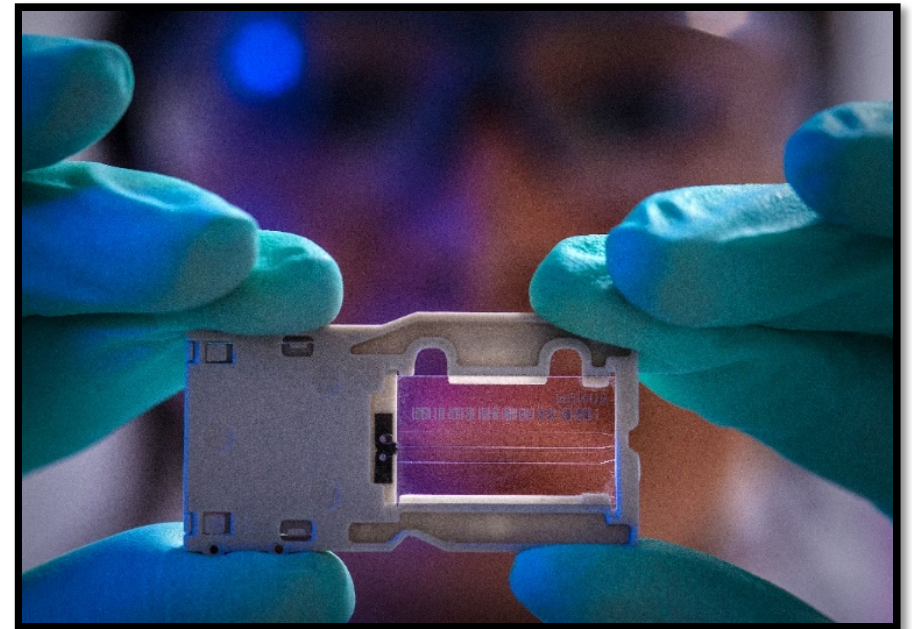
WGS Provides More Information Than Just Relatedness



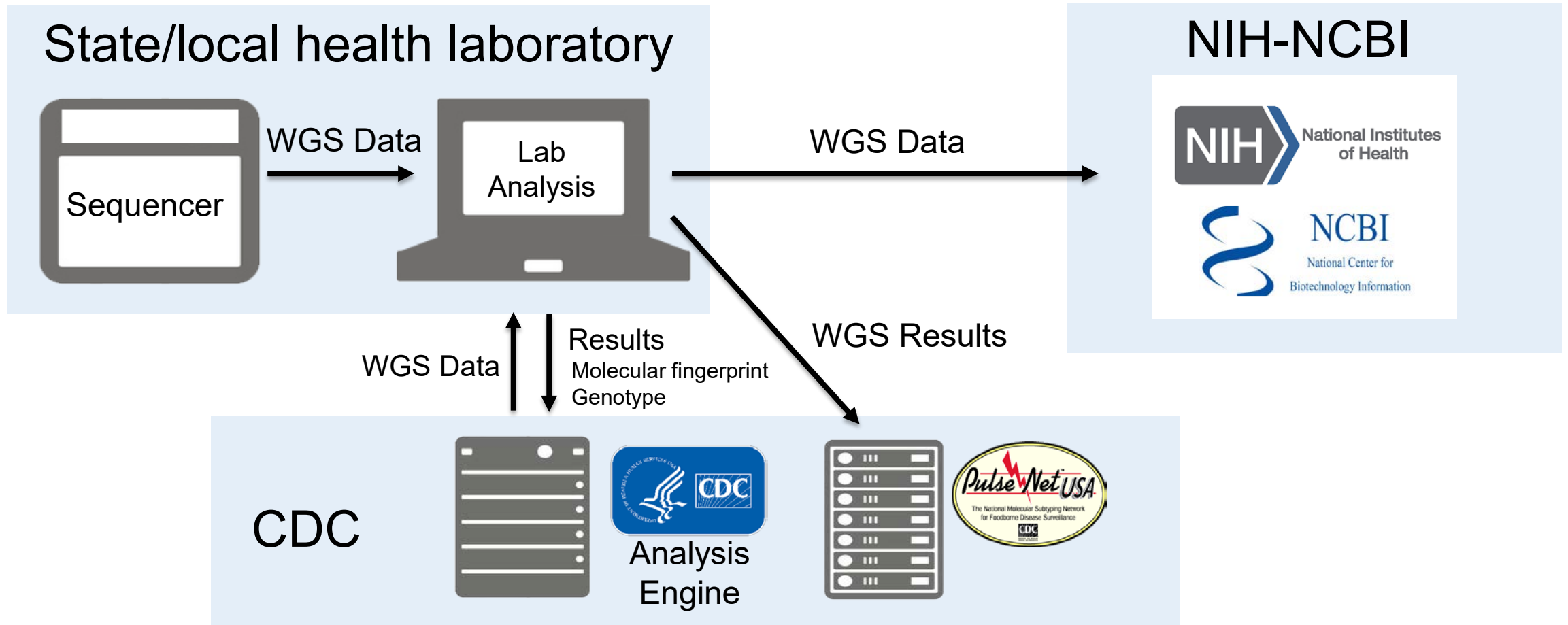
WGS Provides More Information Than Just Relatedness

➤ Genotyping provides

- Identification of organism
- Serotype (O and H genes; other surface marker genes)
- Antibiotic resistance genes and mutations – which can be associated with harder to treat infections
- Virulence genes – can lead to more severe infections



PulseNet Uses a Distributed Analysis System to Extract Information from WGS



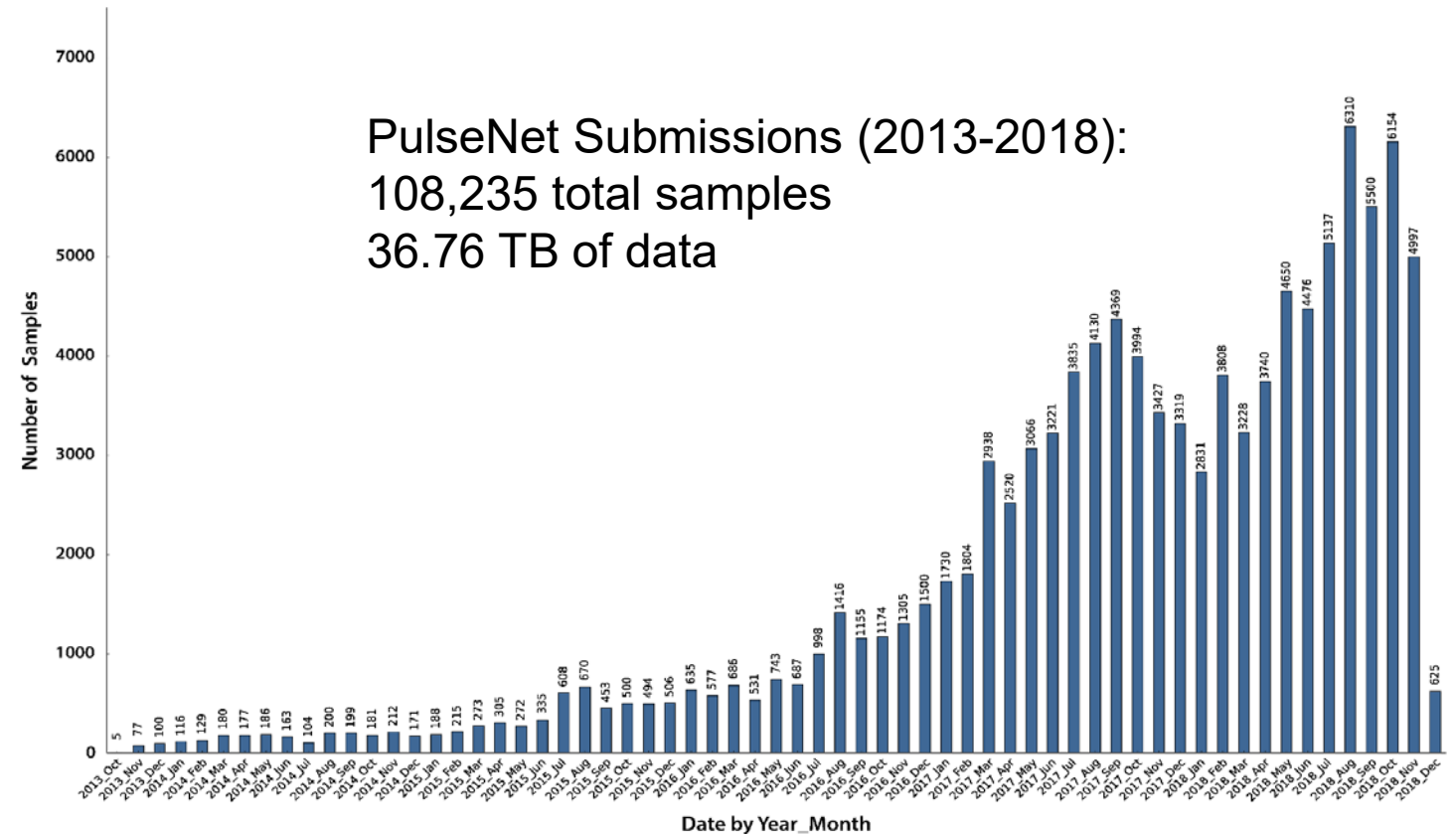
Challenges and Opportunities for WGS Analysis

➤ Challenges

- Amount of data
- Complex computing jobs
- Differing IT resources

➤ Opportunities

- Detect foodborne outbreaks with more precision
- Receive genotyping and molecular fingerprint results for local cluster detection



Mbps: megabits per second (1 Mbps = dialup internet, > 100 = “fast internet”)

Terabyte: one trillion bytes of data; Library of Congress adds 5 Terabytes of data per month to its archives

Challenges and Opportunities for WGS Analysis

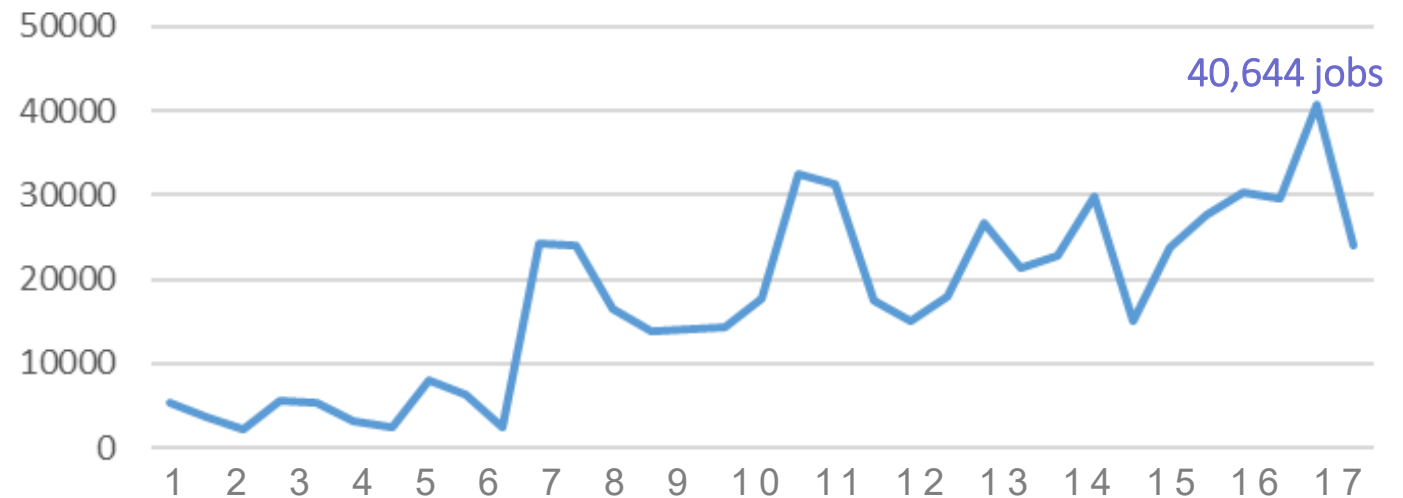
➤ Challenges

- Amount of data
- Complex computing jobs
- Differing IT resources

➤ Opportunities

- Detect foodborne outbreaks with more precision
- Receive genotyping and molecular fingerprint results for local cluster detection

Jobs per Week in 2019



Mbps: megabits per second (1 Mbps = dialup internet, > 100 = “fast internet”)

Terabyte: one trillion bytes of data; Library of Congress adds 5 Terabytes of data per month to its archives

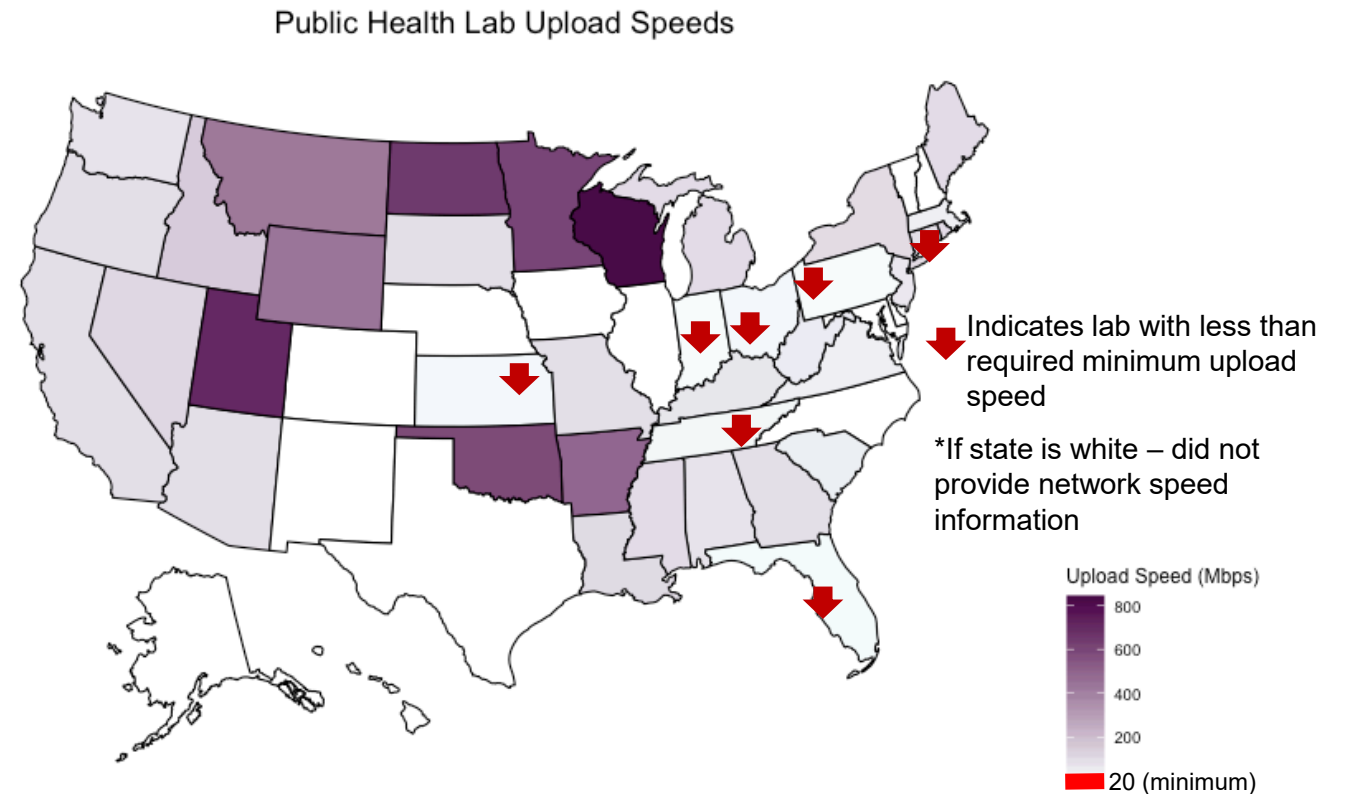
Challenges and Opportunities for WGS Analysis

➤ Challenges

- Amount of data
- Complex computing jobs
- Differing IT resources

➤ Opportunities

- Detect foodborne outbreaks with more precision
- Receive genotyping and molecular fingerprint results for local cluster detection



Mbps: megabits per second (1 Mbps = dialup internet, > 100 = “fast internet”)

Terabyte: one trillion bytes of data; Library of Congress adds 5 Terabytes of data per month to its archives

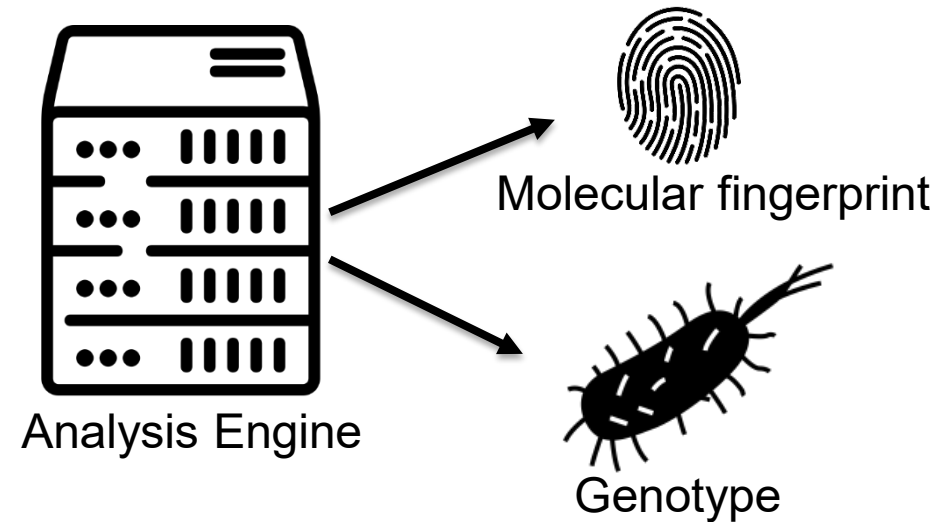
Challenges and Opportunities for WGS Analysis

➤ Challenges

- Amount of data
- Complex computing jobs
- Differing IT resources

➤ Opportunities

- Detect foodborne outbreaks with more precision
- Receive genotyping and molecular fingerprint results for local cluster detection



Mbps: megabits per second (1 Mbps = dialup internet, > 100 = "fast internet")

Terabyte: one trillion bytes of data; Library of Congress adds 5 Terabytes of data per month to its archives

The Future of PulseNet

- **With the transition to WGS, PulseNet will be able to detect outbreaks with more precision.**
- **This transition was made possible with initial investments by AMD.**
- **Beyond WGS, PulseNet is continuing to innovate by applying novel technologies like metagenomics to characterize foodborne pathogens directly from patient samples.**



State-level AMD Applications Across the Infectious Disease Spectrum



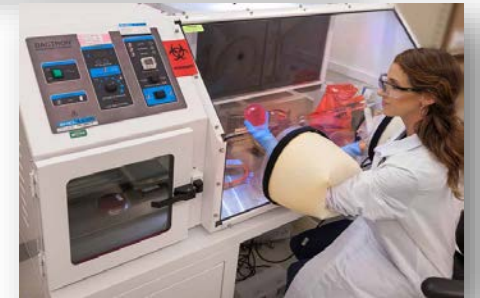
Lauren Turner, PhD

Foodborne and Advanced Pathogen Characterization Lead Scientist

Virginia Department of General Services, Division of Consolidated Laboratory Services

Division of Consolidated Laboratory Services (DCLS) Virginia's State Laboratory

- ❑ **The Commonwealth of Virginia's public health, environmental, agriculture, and consumer protection laboratory**
- ❑ **Serving local, state, and federal agencies**
- ❑ **Comprehensive testing services including**
 - ❑ Neonatal screening
 - ❑ Drugs of abuse testing
 - ❑ Microbiology testing
 - ❑ Food and water contamination monitoring
 - ❑ Metal and pesticide analyses
 - ❑ Chemical analyses



Infectious Disease Surveillance

The Not so Good Old Days

❑ **Pre-AMD surveillance and outbreak detection tools**

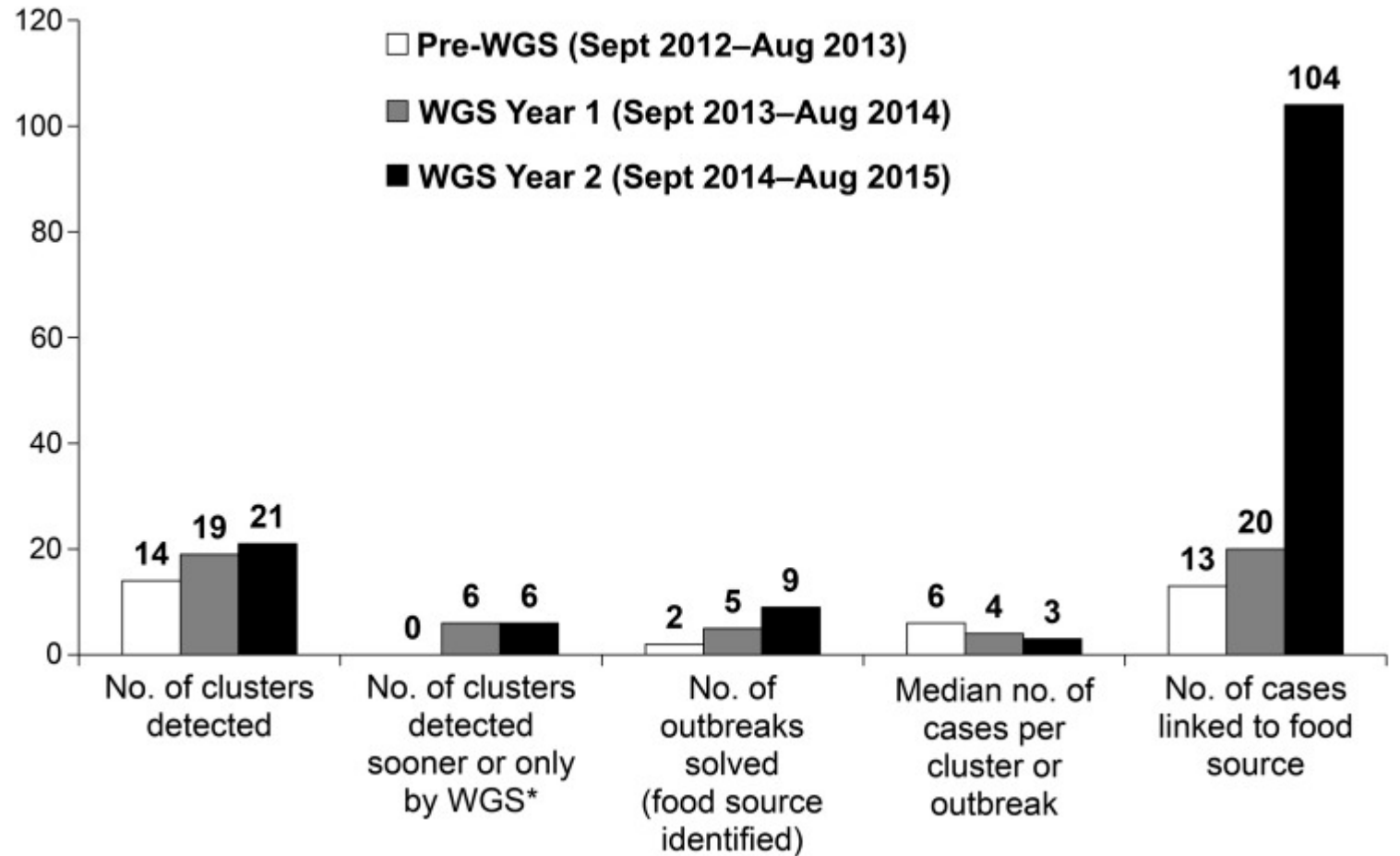
- ❑ Specimen culture and pathogen identification
- ❑ Serovar or serotype subgrouping
- ❑ Virulence factor genotyping and phenotyping
- ❑ DNA fingerprinting by pulsed-field gel electrophoresis (PFGE)
 - ❑ Foodborne, healthcare-associated and community acquired outbreaks; >25 different organisms

❑ **Constraints**

- ❑ Costly and time intensive
- ❑ Loss of institutional knowledge with aging workforce
- ❑ Limited detection with targeted DNA-methods
- ❑ Limited resolution of pathogen genetic differences or similarity

Advancing Public Health Laboratory Infectious Disease Surveillance

- ❑ **WGS improves accuracy**
- ❑ **Provides epidemiologists and regulatory officials with actionable results**
- ❑ **Improves use of laboratory resources**
- ❑ **Reduces turnaround time**



Advancing Pathogen Surveillance through Next-generation Sequencing (NGS)

Goals of Virginia DCLS












































- ❑ Pragmatic implementation of NGS to meet state and federal program needs
- ❑ Develop capacity for data analytics (bioinformatics)
- ❑ Use pathogen genomics bioinformatics tools
 - ❑ Pathogen identification
 - ❑ Pathogen genotyping
 - ❑ Early detection of possible outbreaks
 - ❑ Enhanced evaluation of suspected outbreaks
 - ❑ Outbreak tracking
 - ❑ Source attribution



Emerging Applications of Next-generation Sequencing

- ❑ **Began as primarily foodborne enteric bacterial pathogens**
 - ❑ Network-driven initiative (PulseNet)
 - ❑ Surveillance and outbreak detection
 - ❑ Transitioning from PFGE to WGS since 2014, beginning with *Listeria*
 - ❑ Heavy reliance on federal partners for bioinformatics analysis
- ❑ **Growing dependence on WGS data**
 - ❑ Foodborne pathogen surveillance solely by NGS in June 2019
- ❑ **Use of NGS expanding to tuberculosis, *Legionnaires' disease*, healthcare-associated infections, and more**

DCLS' AMD Capacity Building: Number of Trained Staff, Samples, and Public Health Applications

	2013	2014	2015	2016	2017	2018	2019
Staff trained to sequence							
Staff trained to extract NA							
Total samples sequenced	12	498	738	1063	2365	2252	2160
Instruments							
Bioinformatics staff							
GenomeTrakr							
PulseNet							
Tuberculosis							
HAI							

NA: Nucleic acid

HAI: Healthcare-associated infections

AMD Applications: *Listeriosis* Outbreak Source Tracking



CA: 1 case

MD: 7 cases

- CFSAN10069
- CFSAN10071
- CFSAN10070
- CFSAN10073
- CFSAN10074
- CFSAN10077
- CFSAN10068
- CFSAN10075
- CFSAN10072
- CFSAN10076
- 2014L-6025 - MD
- 2013L-5749 - MD
- 2014L-602 - MD
- 2013L-5594 - CA
- 2014L-6027 - MD
- 2013L-5752 - MD
- 2013L-5429



6 hqSNPs [2-18]

Nearest genetic neighbor
(663 hqSNPs different; different serotype)



0.1

hqSNP: high-quality single nucleotide polymorphism; median [range]

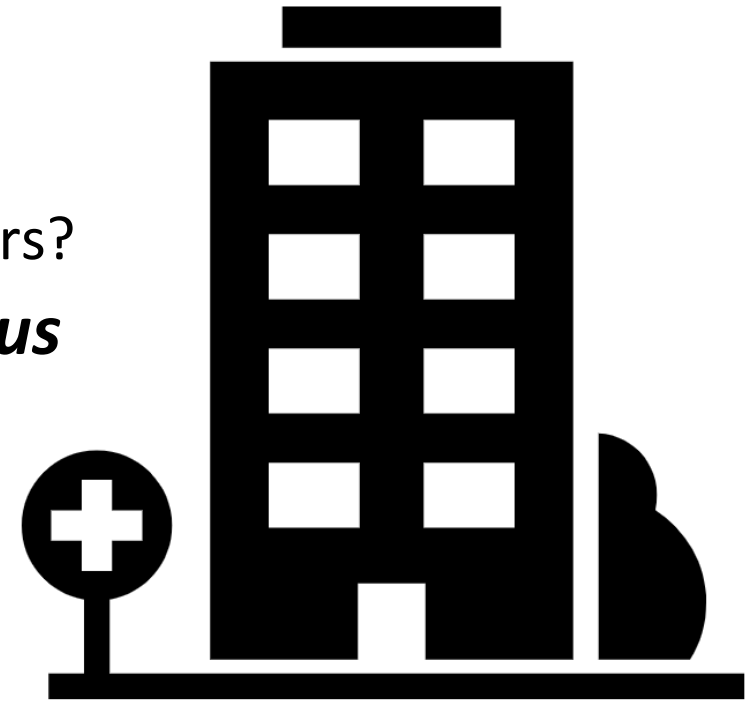
AMD Applications: Infection Transmission and Control in Healthcare Settings

❑ Investigating transmission of pathogens

- ❑ Is transmission occurring between facilities?
- ❑ Is this strain regionally or locally persistent?
- ❑ Is there asymptomatic infection among healthcare providers?

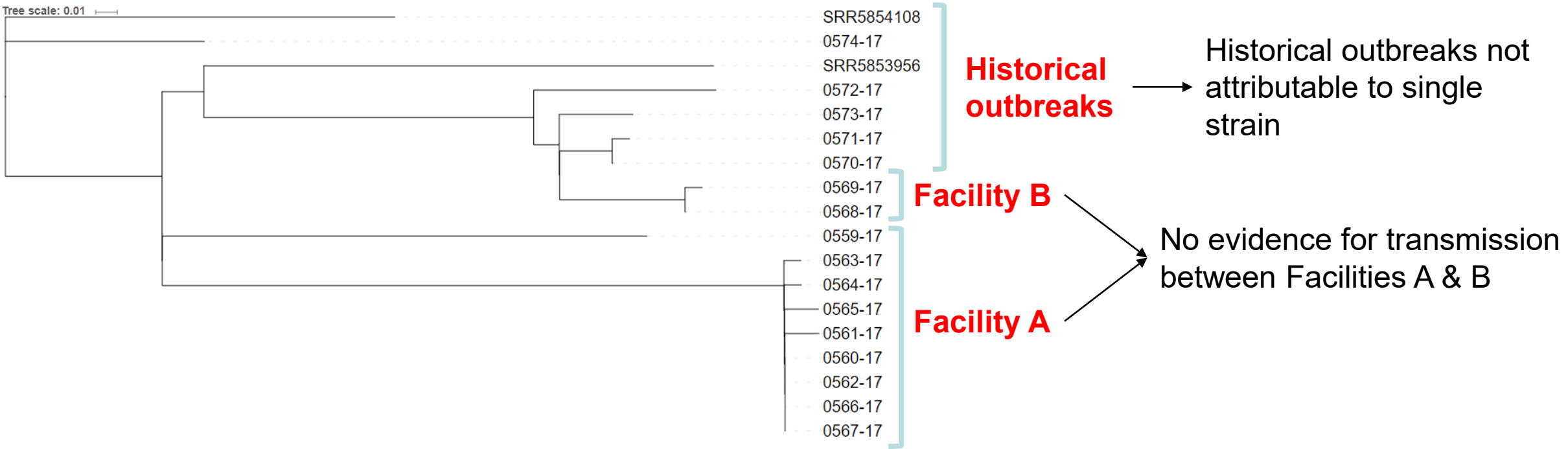
❑ NGS to assess severe invasive *Group A Streptococcus* (GAS) outbreaks

- ❑ Multiple facilities
- ❑ Regional occurrence
- ❑ 2013–2017



AMD Applications: Infection Transmission and Control in Healthcare Settings

Group A *Streptococcus* (GAS) Isolates from Clinical Facilities



AMD Applications: Pathogen Identification and Genotyping

- ❑ Applying predictive analytics for pathogen identification from sequence data
- ❑ Reference database comparison
 - ❑ Species identification
 - ❑ Genotyping serotype prediction

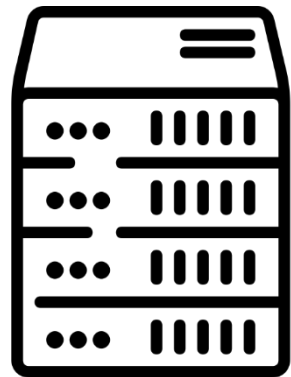
sample	r1_q	r2_q	est_genome_length	est_cvg	number_contigs	species_prediction	subspecies_predictions
isolate001	35.61	31.92	5367997	76.32	256	Escherichia_coli	O26:H11
isolate002	34.09	28.99	5417894	49.7	253	Escherichia_coli	O103:H2
isolate003	36.36	33.63	2931612	130.08	27	Listeria_monocytogenes	NA
isolate004	36.86	35.43	4754353	77.82	31	Salmonella_enterica	Heidelberg
isolate005	36.67	34.53	1686219	120.42	15	Streptococcus_pyogenes	emm89
isolate006	36.42	34.93	1802069	94.48	24	Streptococcus_pyogenes	emm1

Challenges to AMD Integration

- ❑ **Development of scientist knowledge, skills, and abilities**
 - ❑ Next-generation sequencing (NGS) methods and instrumentation
 - ❑ Bioinformatics analysis of NGS data
- ❑ **Building of NGS infrastructure**
 - ❑ Procurement, installation, and maintenance
- ❑ **Data storage and transfer solutions for terabytes of NGS data**
- ❑ **Computational resources for data analysis**
- ❑ **Policies and procedures to support expanding use of NGS and efficient analysis of data**
- ❑ **Funding to sustain and capitalize on AMD investments**

Opportunities Through AMD Integration

- ❑ **Many different approaches to achieve state capacity for AMD**
 - ❑ Investment in different IT structures and software applications for data analysis
 - ❑ Different avenues for computational resources



On-premises
hardware



Academic
partnerships



Cloud
computing

Opportunities Through AMD Integration

AMD Training & Bioinformatics Resource Support

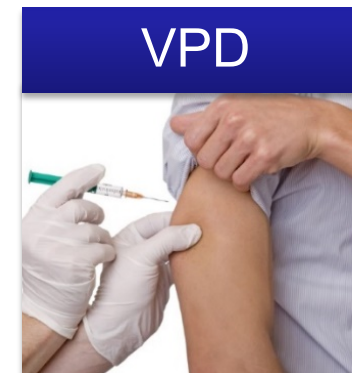
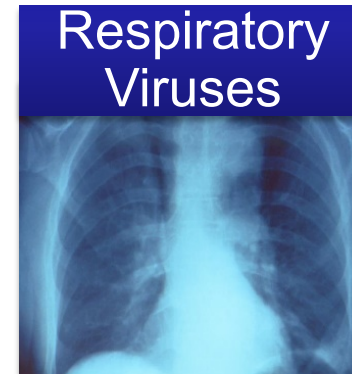
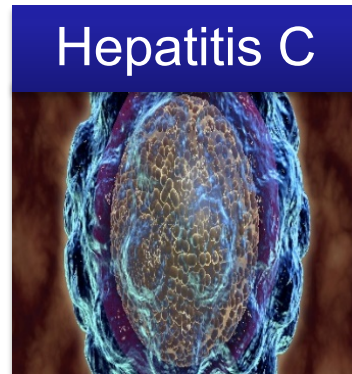
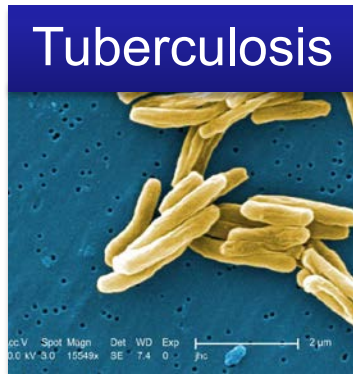


- ❑ **Modernization of public health lab testing**
- ❑ **Novel partnerships to strengthen state programs**
 - ❑ Regional training and bioinformatics resource support
 - ❑ University partnerships



Opportunities Through AMD Integration

States can tailor AMD to meet local disease control needs



VPD: vaccine-preventable disease

Future Virginia Laboratory Priorities and Needs

- Data integration and visualization**
- Data sharing agreements**
- Guidance for alignment with clinical laboratory requirements**
- Applying AMD to detect pathogens in clinical samples**
- Interdisciplinary workforce capacity building**

Where Does Pathogen Genomics Go from Here?



Gregory L. Armstrong, MD

Director, AMD Program

Centers for Disease Control and Prevention

	AMD before 2013	AMD in 2020
NGS capacity	One sequencer	Capacity in all infectious disease laboratories
Bioinformatics capacity	<10 bioinformaticians	~60 bioinformaticians; widespread experience and expertise in microbial genomics
Capacity in state/local health departments	NGS capacity in one laboratory	NGS capacity in all state and many local laboratories; 7 regional bioinformaticians and an increasing number of state-based bioinformaticians

- **Continued adaptation of NGS to public health priorities**
- **Emerging areas of interest**
 - New sequencing technologies
 - Metagenomics (sequencing directly from samples, rather than from cultured isolates)
 - Data integration and modernization

cdc.gov/amd
@cdc_amd

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.

Apply to Become a Bioinformatics Fellow

➤ Applications Due: **February 29, 2020**

The [APHL-CDC Bioinformatics Fellowship](#) aims to train and prepare bioinformaticians to apply their expertise within public health and design tools to aid existing public health personnel in the use of bioinformatics. Post-master's and postdoctoral-level professionals can apply their skills to a range of important and emerging public health problems, while gaining experience in their fields. Whether your specialty is metagenomics, algorithm/software development, microbial genomics, or another research area, we have a place for you!