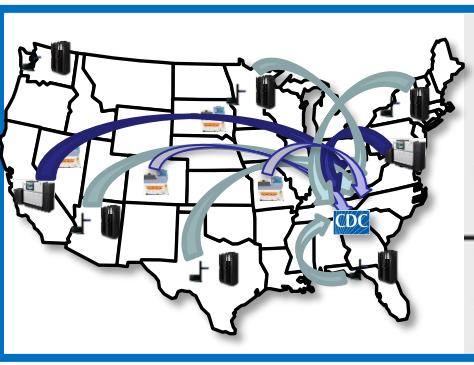
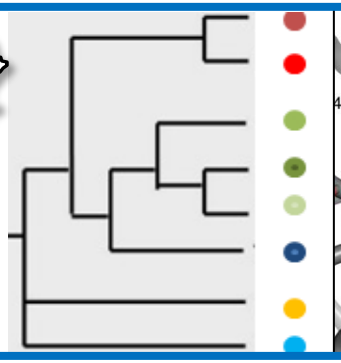
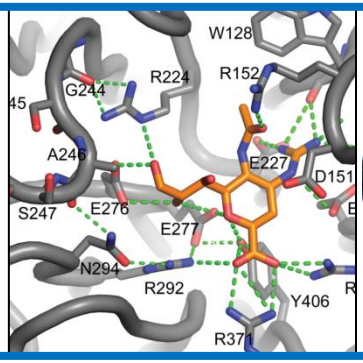
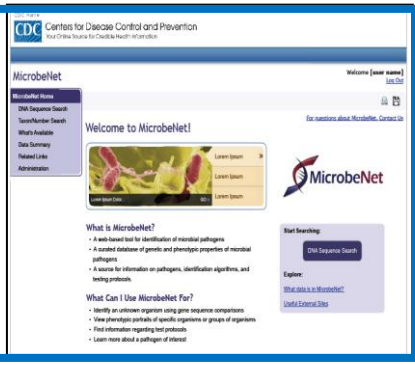
CDC PUBLIC HEALTH GRAND ROUNDS

Advanced Molecular Detection for Improving Public's Health







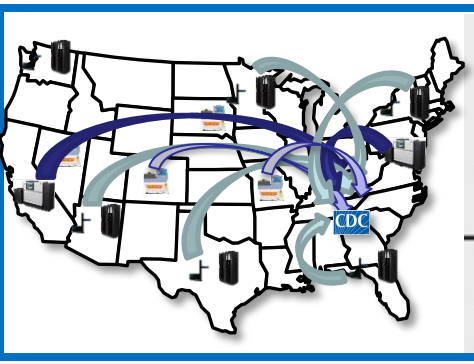


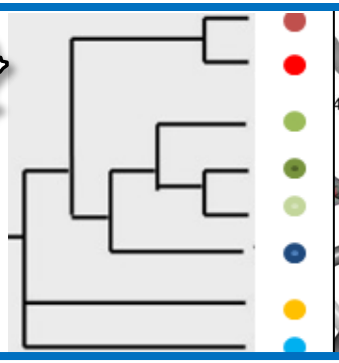
Accessible version: https://youtu.be/OYhnZ_yNRNo

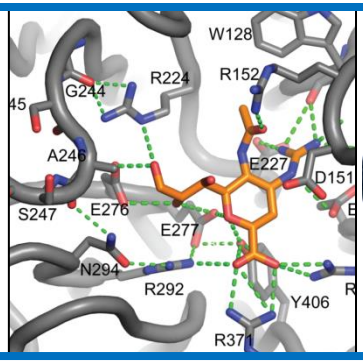
December 3, 2013

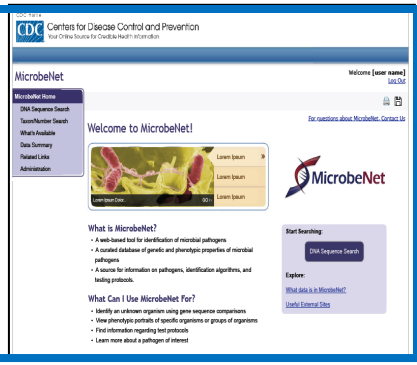
U.S. Department of Health and Human Services
Centers for Disease Control and Prevention

What is Advanced Molecular Detection?









Duncan MacCannell, PhD

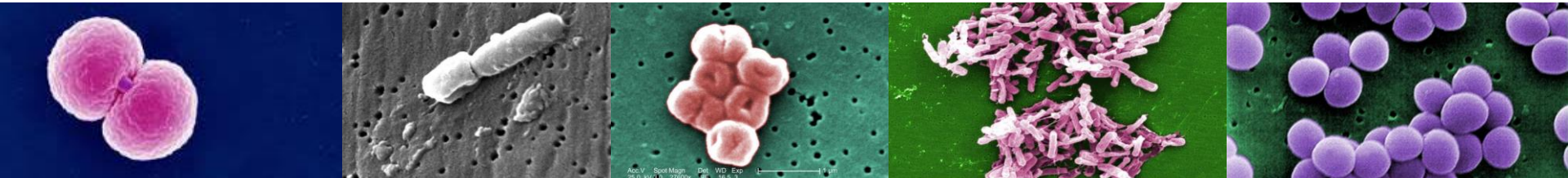
Senior Advisor for Bioinformatics

National Center for Emerging and Zoonotic Infectious Diseases, CDC

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

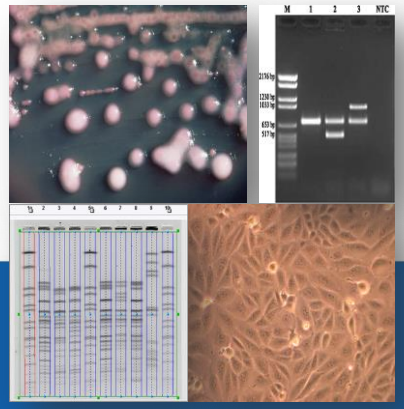
Emerging Threats

- **Pathogens**
  - Highly resistant pathogens in health care settings
  - Pathogens that jump from animals to humans
  - Globalization and the emergence of new and virulent pathogens
  - 5 pathogens are resistant to nearly all available drug treatments
- **Economic cost**
  - 1 in 6 Americans (48 million people) get sick from contaminated food products each year
    - Cost: $77B/year in healthcare treatment, workplace, and other economic losses
  - Each year, flu alone costs businesses ~$10.4B in direct costs for adult hospitalizations and outpatient visits

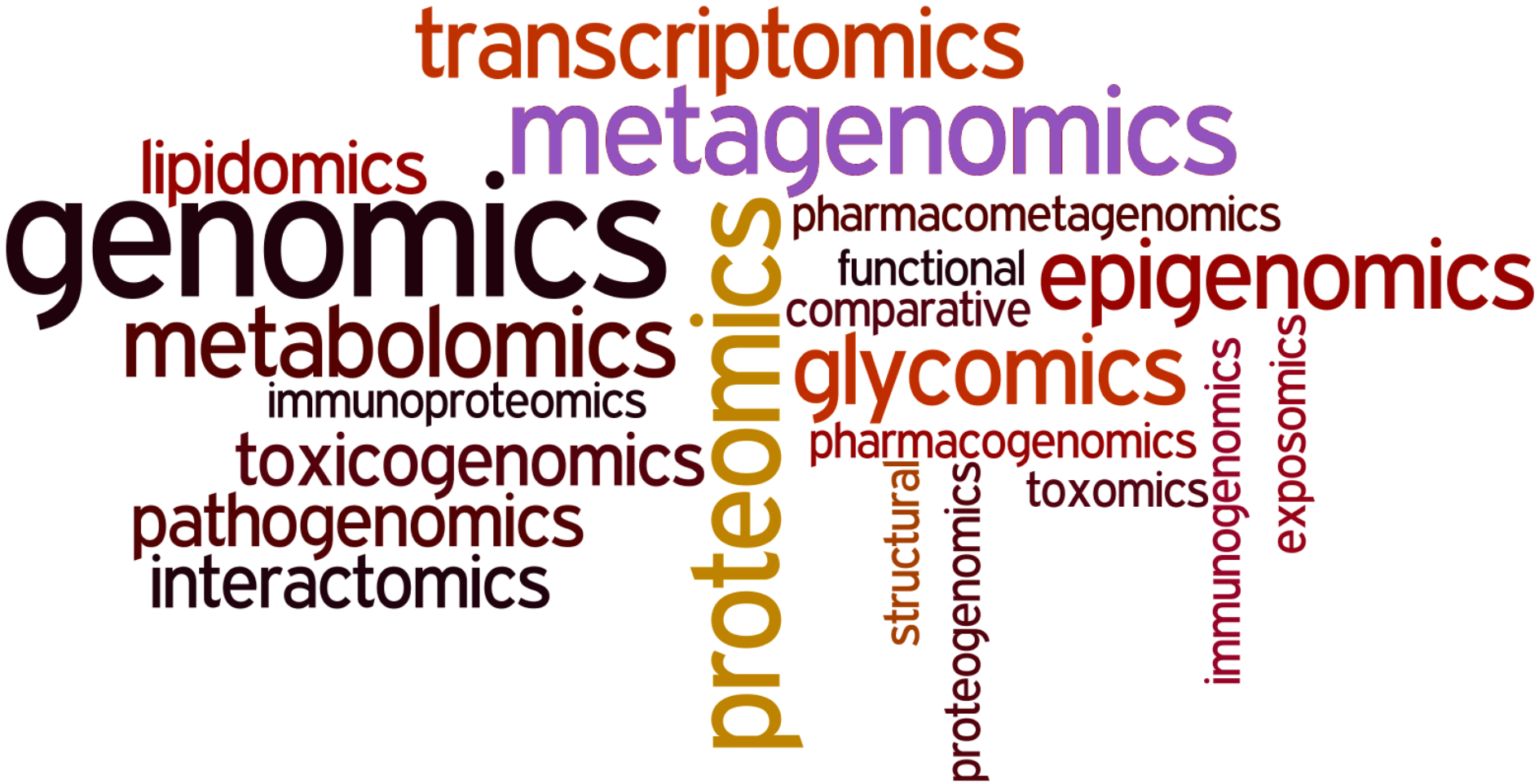


Challenges and Opportunities

- **Decades-old methods of detecting microbes and understanding their effect on human health are limiting our ability to respond to infectious disease threats**
- **The era of "-omics" and bioinformatics**
  - Allows improvements in cost, throughput, and accuracy
  - Represents a new paradigm for public health laboratories



The "-omics" at CDC



Metagenomics: Study of the genetic material of all the microbes in a specimen

Genomics: Study of the genetic material of an organism

Transcriptomics: Study of the transcriptome, the set of all RNA molecules in one cell or a population of cells

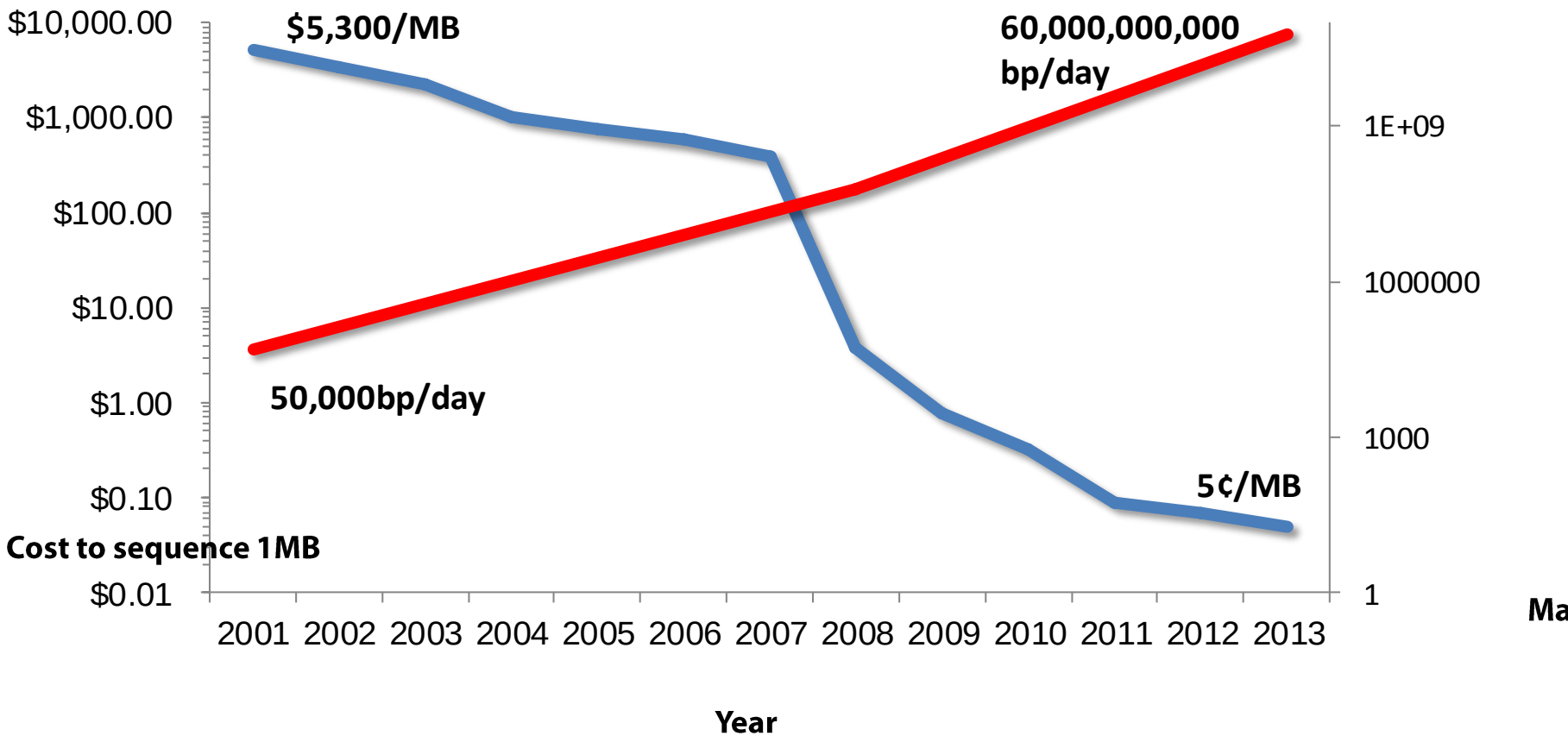
Examples of Advanced Molecular Detection Technologies at CDC

- **Genomics, metagenomics, and transcriptomics**
  - Next generation sequencing: high-throughput deep sequencing
  - RNA sequencing
  - Functional/comparative genomics, metagenomics
  - New, massively-parallel PCR- and array-based methods
- **Proteomics**
  - MALDI-TOF (matrix-assisted laser desorption/ionization–time of flight)

Bioinformatics and advanced data analytics

The challenge is no longer doing the testing but having the computational infrastructure needed to **manage, analyze, and interpret the data**

Next Generation Sequencing
A Fundamental Change in Molecular Microbiology



Source: NHGRI, http://www.genome.gov/sequencingcosts
Bp, Base pairs
MB, Million bases

Next Generation Sequencing Advantages

- **Advantages**
  - To learn about the microbe's phylogenetic history
    - "Where has it been—where it came from"
  - Try to predict "where it is going"
    - Is it on its way to become resistant to antimicrobial drugs?
    - Will its potential for human-to-human transmission change?
- **Examples of practical applications: Identification of**
  - Microbial clusters
  - Genetic markers for virulence and antimicrobial resistance
  - Evidence of bioengineering
  - New pathogens

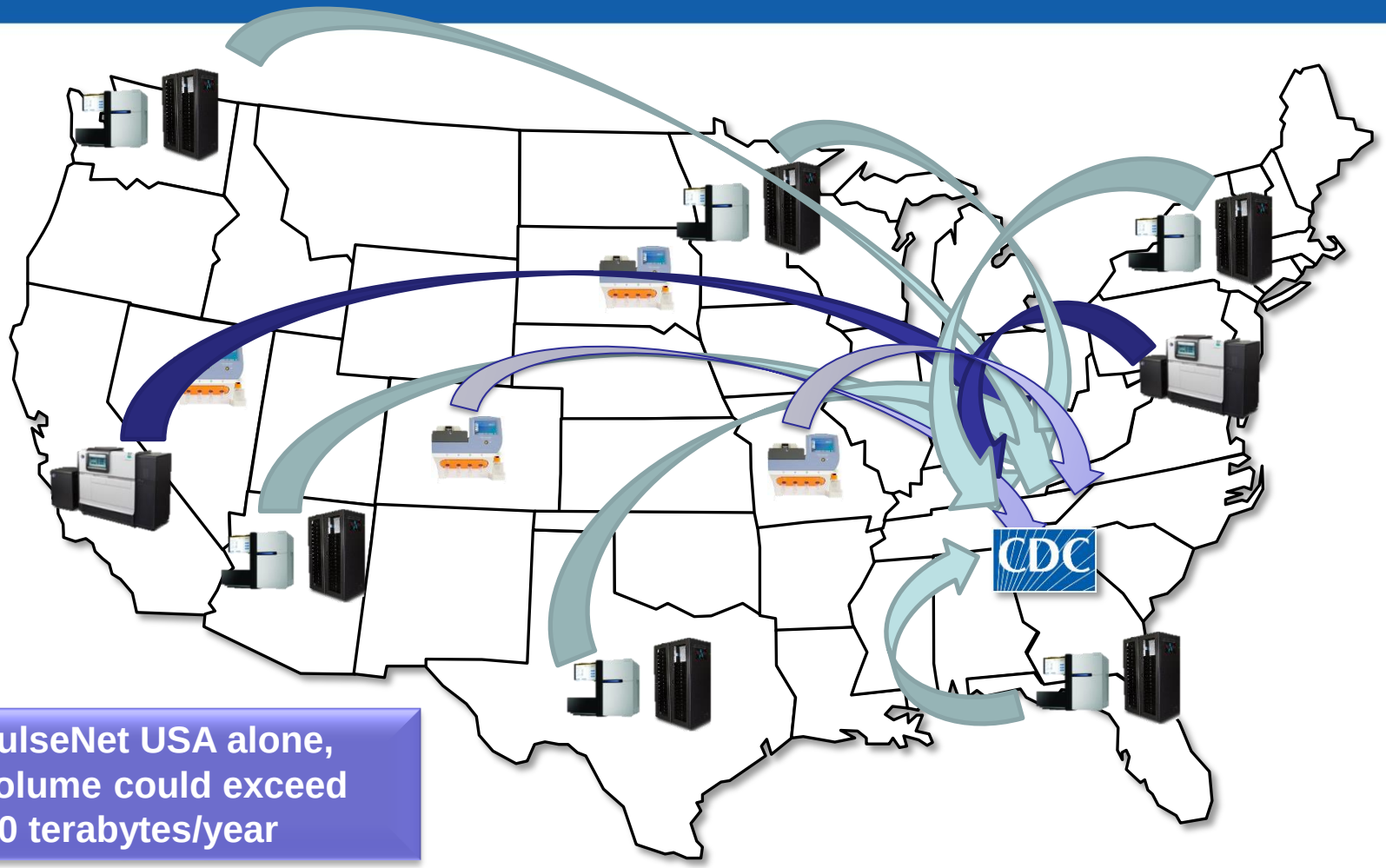
National Human Genome Research Institute, http://www.genome.gov/sequencingcosts

Next Generation Sequencing Challenges

- **Sheer volume of generated sequence data**
  - Management and storage
  - Analysis and statistical interpretation
    - Technical reproducibility
  - Public health application of laboratory information
- **Constant evolving of new technologies and protocols**
- **New skills and capabilities in the public health workforce**
  - Bioinformatics
  - Laboratory technology
  - Data science

http://www.genome.gov/sequencingcosts

A Larger Challenge
Integrating AMD into Public Health Surveillance



For PulseNet USA alone, data volume could exceed 250 terabytes/year

www.cdc.gov/pulsenet

AMD Initiative: A New Landscape for Combating Infectious Diseases

- **Proposed in the President's 2014 budget**
- **A major enhancement of CDC's current microbiology and bioinformatics capabilities to more accurately and rapidly (and with decreased cost)**
  - Diagnose infectious diseases
  - Investigate and control outbreaks
  - Understand transmission patterns
  - Determine antimicrobial resistance
  - Develop and target prevention measures, including vaccines

http://www.cdc.gov/AMD

AMD Initiative: Critical Components

Leverage emerging **molecular sequencing** technologies and **bioinformatics** to enable CDC to build and integrate critical capacities at the **national and state** levels to enhance infectious disease prevention and control

AMD In Action: Outbreak Investigation

Human
Animal
Environmental
Samples

IDENTIFY

Same pathogen?
Does it fit the clinical syndrome?
Is it present in all your cases?
Is it absent in your controls?

SUBTYPE

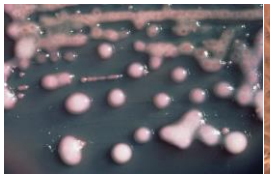
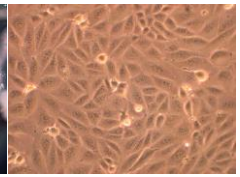
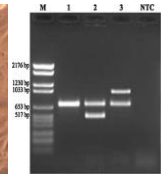
Same strain?
If different, how different?
Chain of transmission?

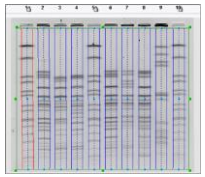
CHARACTERIZE

Important features?
Are there characteristics affecting its pathogenicity, transmissibility, and resistance to drugs that could explain its emergence or affect public health response?

CULTURE, PHENOTYPIC and MOLECULAR

Timeframe: DAYS/WEEKS



AMD

Timeframe: HOURS

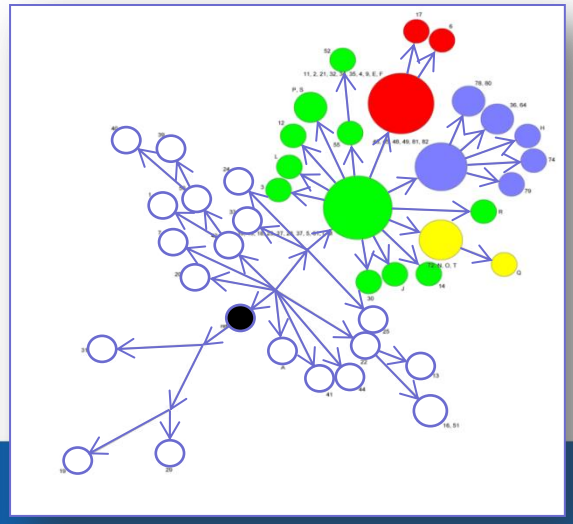
AMD Initiatives

1. Improve identification and detection of pathogens
2. Adapt new diagnostics to meet evolving public health needs
3. Help states meet future reference testing needs in a coordinated manner
4. Implement enhanced, sustainable, and integrated laboratory information systems
5. Develop prediction, modeling, and early recognition tools

1. Improve Identification and Detection of Pathogens

- **Expand capacity for rapid DNA sequencing and molecular characterization of pathogens**
  - Improve capabilities for data analysis and interpretation
- **Develop tools for genome-scale molecular epidemiology, and help CDC apply them to its laboratory and surveillance activities**

- **Outcome**
  - Rapid progress toward modernizing PulseNet and other critical laboratory-based surveillance systems

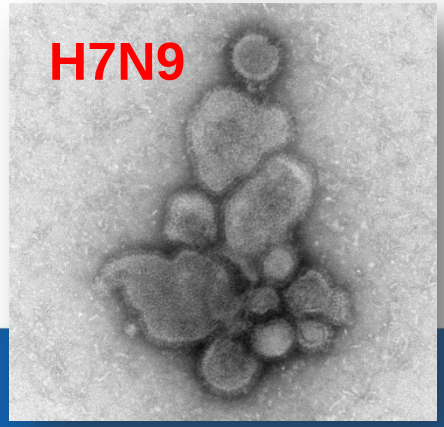


Today's example:
Detection of TB outbreaks
J Posey

2. Adapt New Diagnostics to Meet Evolving Public Health Needs

- **Establish leading capability to adapt the next generation of rapid, semi-automated, point-of-need molecular tests to meet evolving public health requirements**

- **Outcome**
  - Enhance CDC's ability to detect outbreaks early
  - Develop new test during outbreaks
  - Better characterize infectious disease threats



Today's example:
Detection of resistance in flu viruses
M Shaw

3. Help States Meet Future Reference Testing Needs in a Coordinated Manner

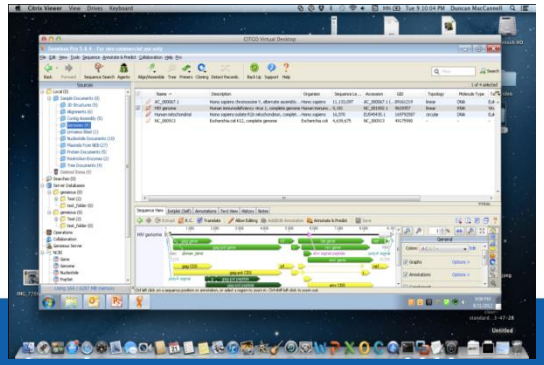
- **Help state and local public health laboratories to transition from culture-based methods to new molecular technologies**
- **Expand capacity for rapid DNA sequencing and molecular characterization to state/local level and improve capabilities for data analysis and interpretation**

- **Outcome**
  - More effective and better integrated outbreak response activities

Today's example:
Reference Databases and MicrobeNet™
J McQuiston

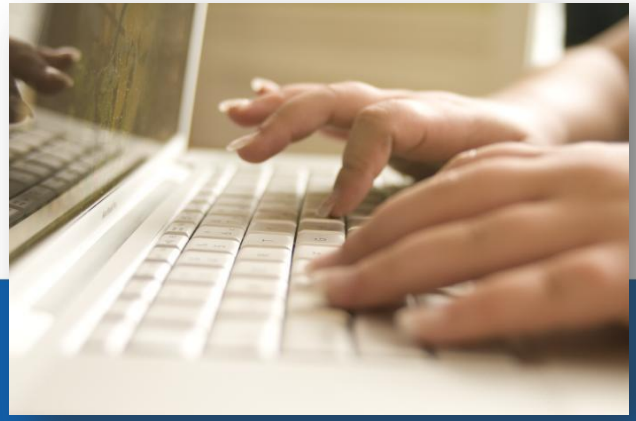
AMD Initiative: Strategic Investments (1)

- **Core scientific infrastructure**
  - Critical laboratory and bioinformatics infrastructure at CDC, state/local public health laboratories, and key overseas laboratories
    - Sequencers, mass-spectrometer, other instrumentation, reagents
    - High-performance computing, scientific workstations
    - Data storage, networking; data integration, knowledge management
    - Service contracts, software licensing, etc.



AMD Initiative: Strategic Investments (2)

- **Workforce development**
  - Training for CDC and public health laboratories workforce
    - Bioinformatics, genomics, -omics
  - New or retooled fellowship programs
    - Genomics
    - Bioinformatics: Partnership with Association of Public Health Laboratories
  - Recruitment of new staff and skill sets
    - Bioinformaticians, data scientists, laboratory specialists



AMD Initiative: Strategic Investments (3)

- **Consortia, partnerships, and alignment of efforts**
  - Academic institutions
  - State and Federal: NIH, FDA, DHS, DoD, DoE/National Laboratories
  - Nonprofit/Nongovernmental organizations
  - International community
  - Commercial/for-profit
- **Pilot projects with state, local, and other partners**
  - Outbreak detection, investigation and response
  - Leverage existing laboratory-based surveillance systems

NIH, National Institutes of Health; FDA, Food and Drug Administration
DHS, Department of Homeland Security
DoD, Department of Defense, DoE, Department of Energy

Advanced Molecular Detection Initiative

IMPROVED DETECTION

Enhanced recognition of emerging microbial threats and antimicrobial resistance

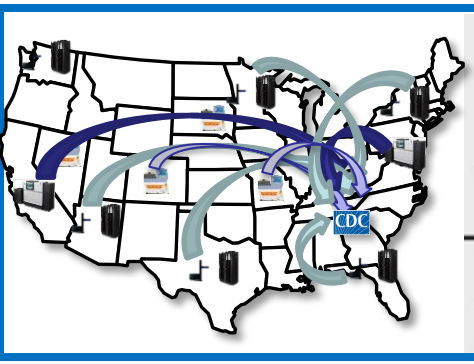
Better targeting of proven prevention strategies and development of new ones

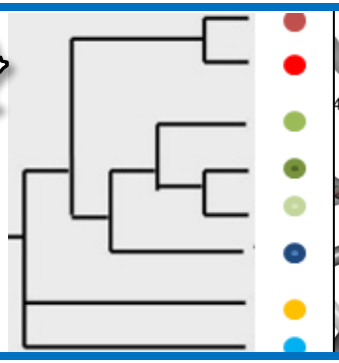
IMPROVED SURVEILLANCE

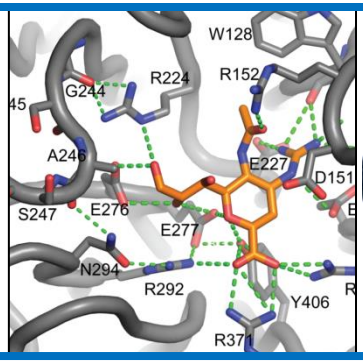
Improved surveillance information on the transmission of infections and the extent and spread of outbreaks

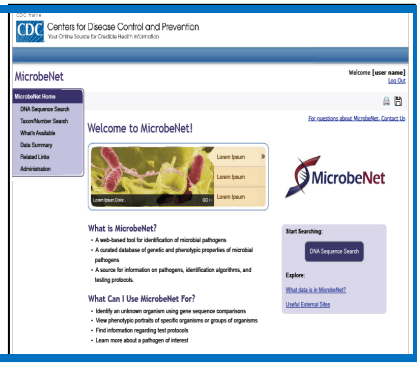
Faster, more effective control efforts

Investigation of Tuberculosis Outbreaks Using Whole Genome Sequencing









James Posey, PhD

Lead, Applied Research Team

Division of Tuberculosis Elimination

National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, CDC

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

National TB Genotyping Service 2004–present

- **Goal**
  - Genotype one *Mycobacterium tuberculosis* isolate from each culture-positive TB case
  - PCR-based methods used
    - Spoligotyping: Variation in direct repeat locus
    - MIRU-VNTR: Variation at 24 distinct loci
- **Results**
  - >80,000 isolates genotyped
  - ~95% coverage in recent years
  - 2010–2012
    - 14,429 genotypes
    - 50% of isolates are in one of 2,520 clusters

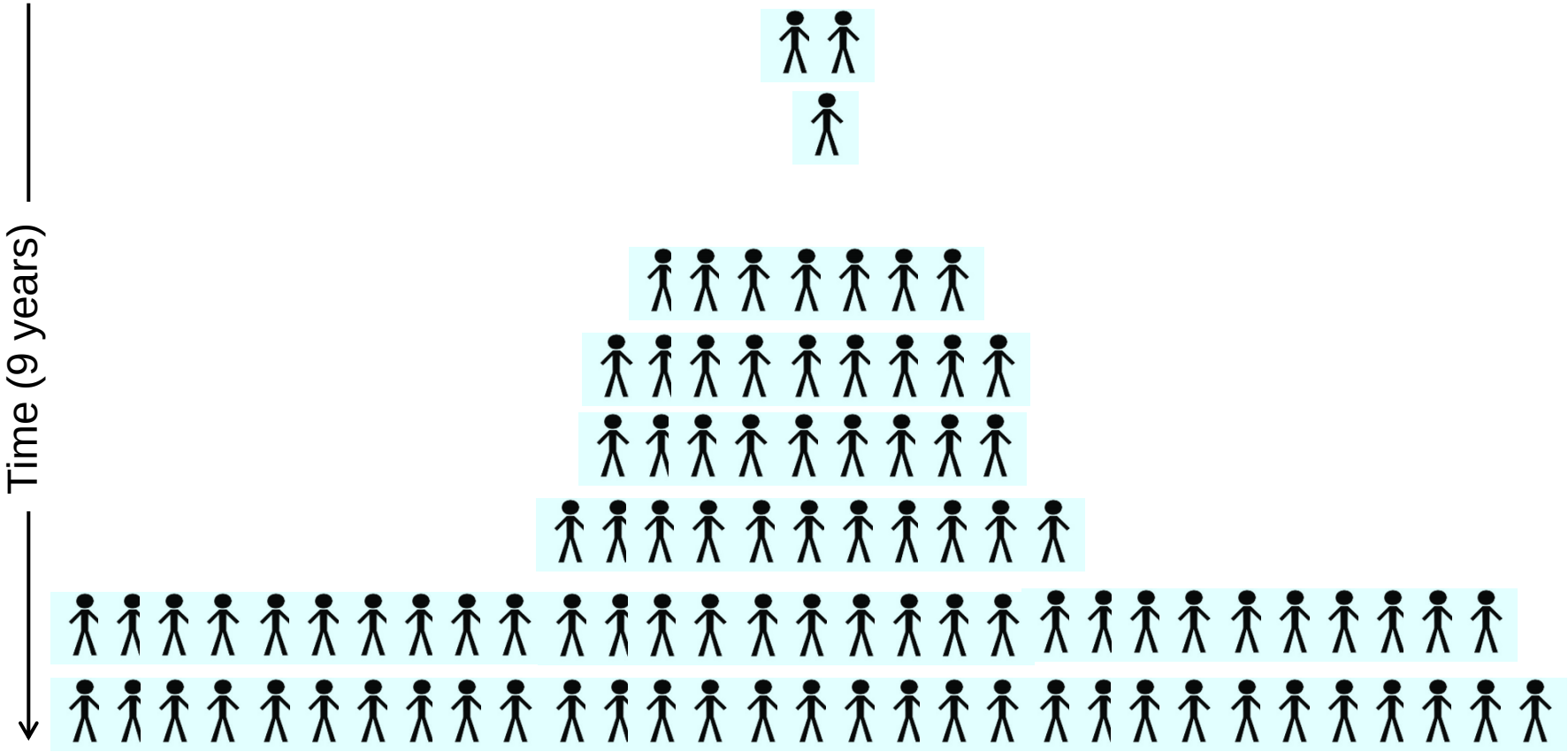
PCR, Polymerase chain reaction
MIRU-VNTR, Mycobacterial interspersed repetitive unit-variable number of tandem repeats

Conventional Genotyping Complemented with Whole Genome Sequencing

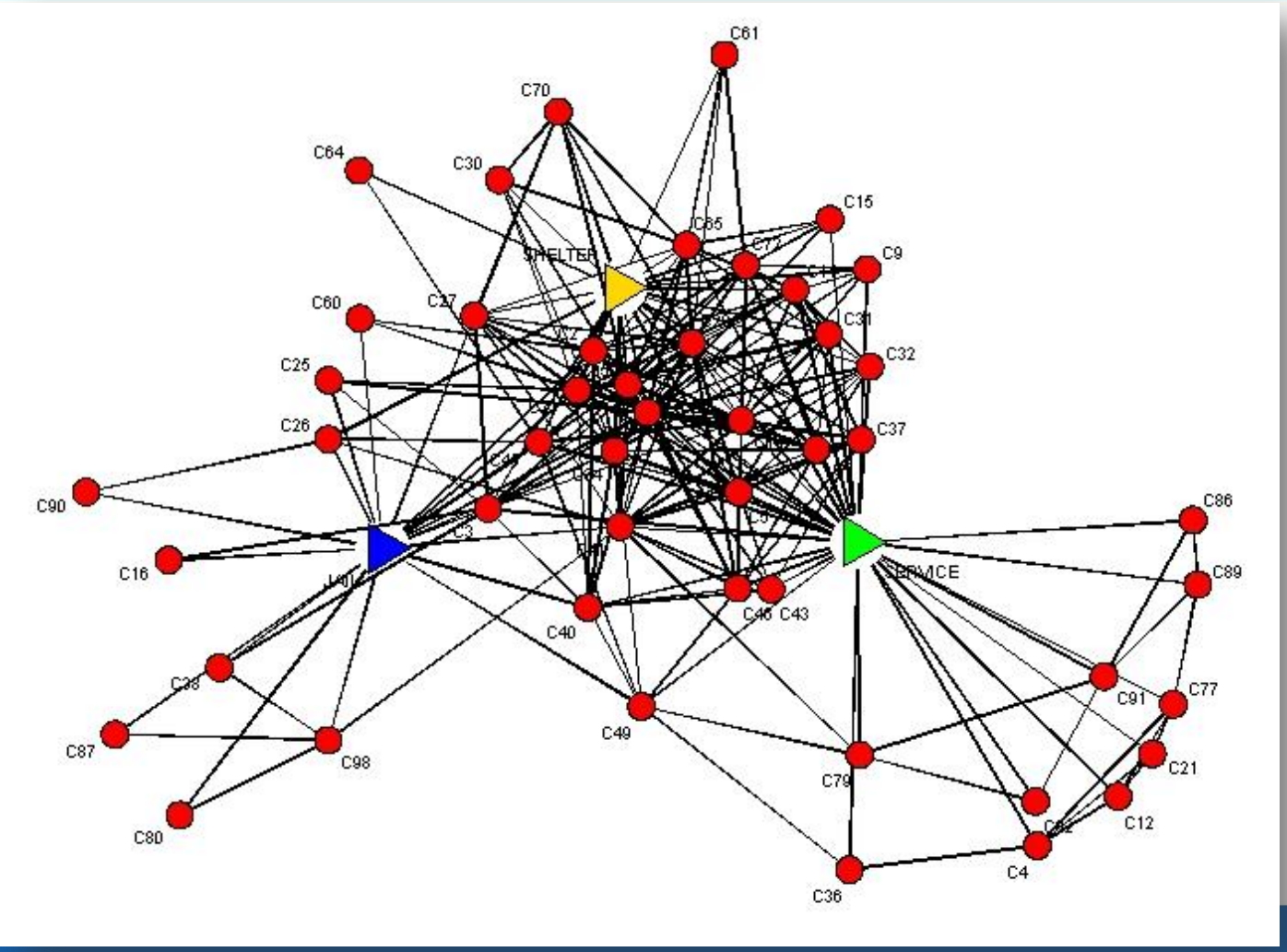
- **Conventional genotyping**
  - Represents <1% of the pathogen's genome
  - Targets are stable over long periods of time
  - Data is difficult to interpret when highly related strains remain circulating in community for long periods of time
- **Whole genome sequencing**
  - Represents ~90% of the pathogen's genome
  - Mutations accumulate at the rate of 0.5 SNP/year

SNP, Single nucleotide polymorphism

Large TB Cluster



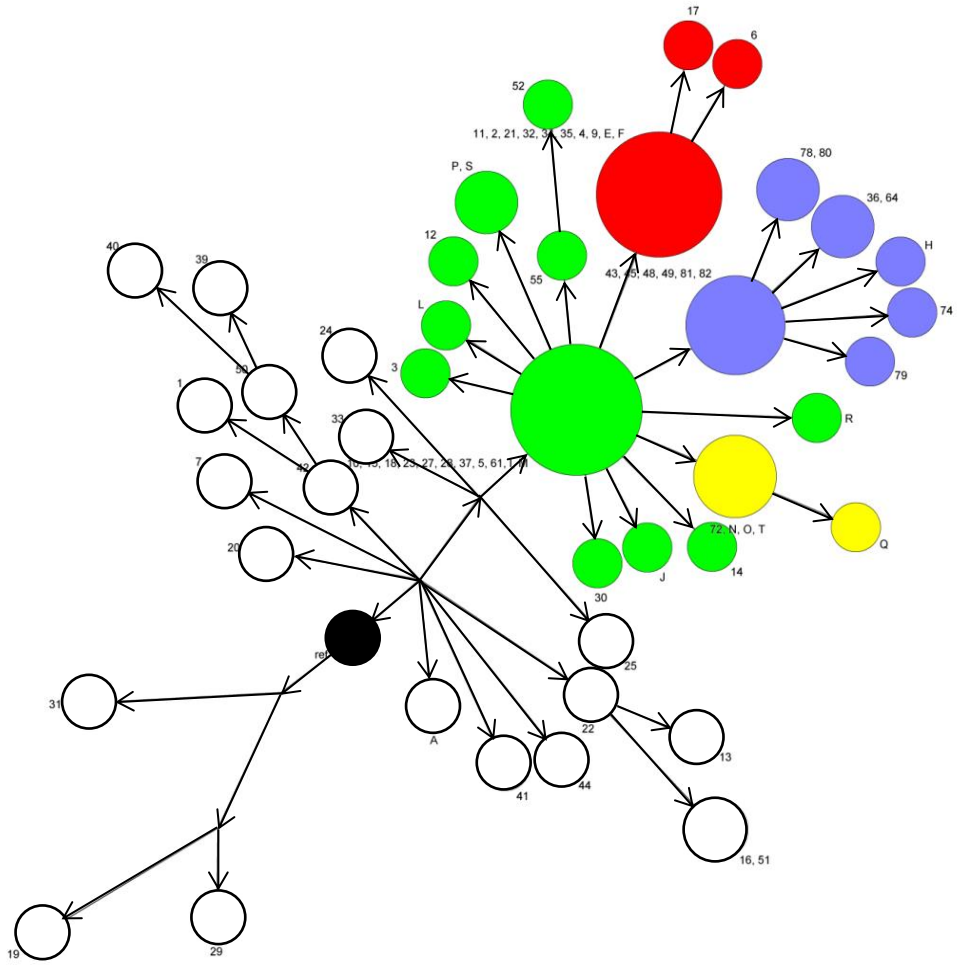
"Social Network" of the Patients



Patients

Transmission sites

Whole Genome Sequencing Results
Answers About Exposure and Transmission

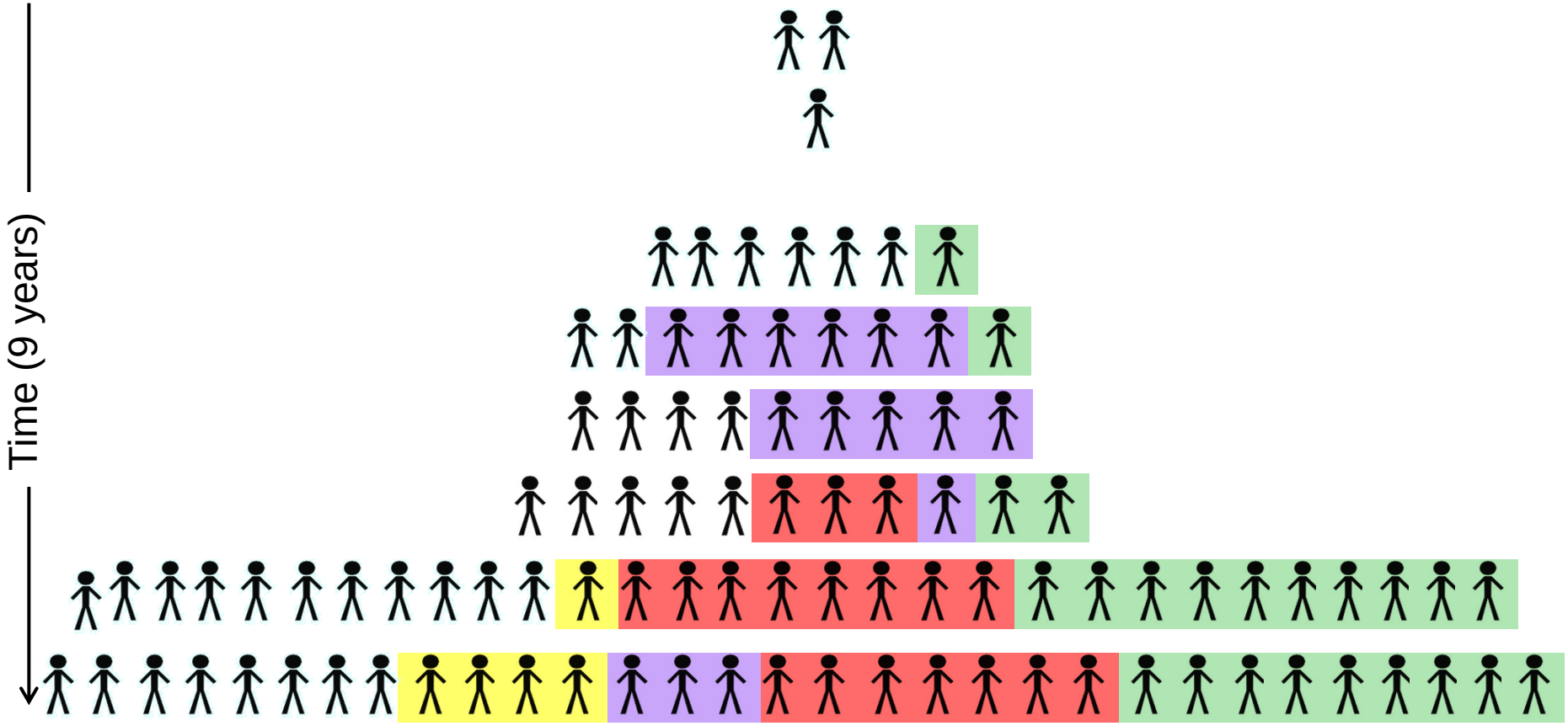


72 Isolates
102 SNPs
41 Genomic types
4 Clusters

Branch length = Number of differences between 2 genomic types

Circle size = Number of isolates within a genomic type

Large TB Cluster



Pairing Whole Genome Sequencing and Epidemiological Data

- Identified cases that were part of transmission chains
- Identified cases that were remotely exposed
- Confirmed sites where transmission likely occurred
- Identified sites where transmission was not likely to have occurred

Now: Retrospective analysis
In the future: Real-time analysis

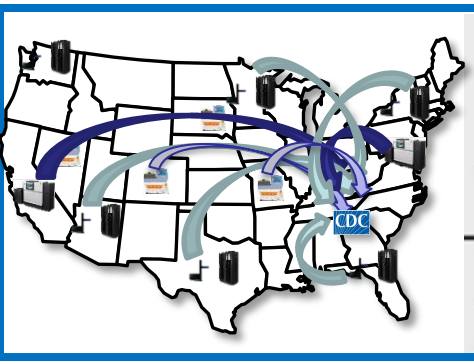
From Knowledge to Action

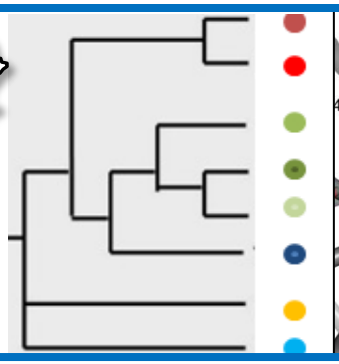
Focus and prioritize efforts during outbreak investigations

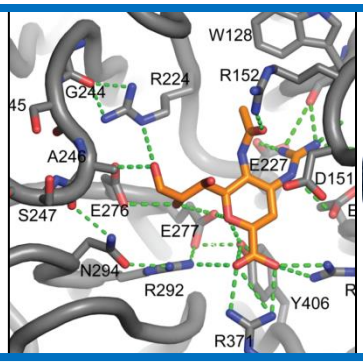
Expand contact investigations around identified source cases

Prioritize TB screening at those sites where transmission was confirmed

Detecting Drug Resistance in Novel Influenza Viruses







Michael Shaw, PhD

Associate Director for Laboratory Science

Influenza Division

National Center for Immunization and Respiratory Diseases, CDC

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

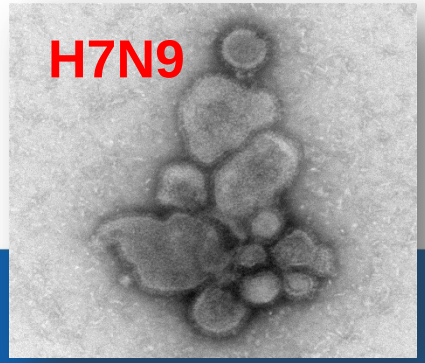
Influenza Antiviral Drugs

- **2 classes of drugs are FDA-approved for treatment of influenza infections in the United States**
  - M2 inhibitors: Amantadine and Rimantadine
  - NA inhibitors (NAIs): Oseltamivir and Zanamivir
    - Peramivir is approved in Japan, South Korea, and China
    - Laninamivir is approved in Japan
- **Drug susceptibility monitoring and assessment**
  - Monitoring: An integral part of virological surveillance
  - Assessment: Based on sequence analysis and/or phenotypic testing
- **Laboratory correlates for clinically relevant resistance to NAIs have not yet been established**

NA, Neuraminidase

Recognition of Novel Influenza A H7N9

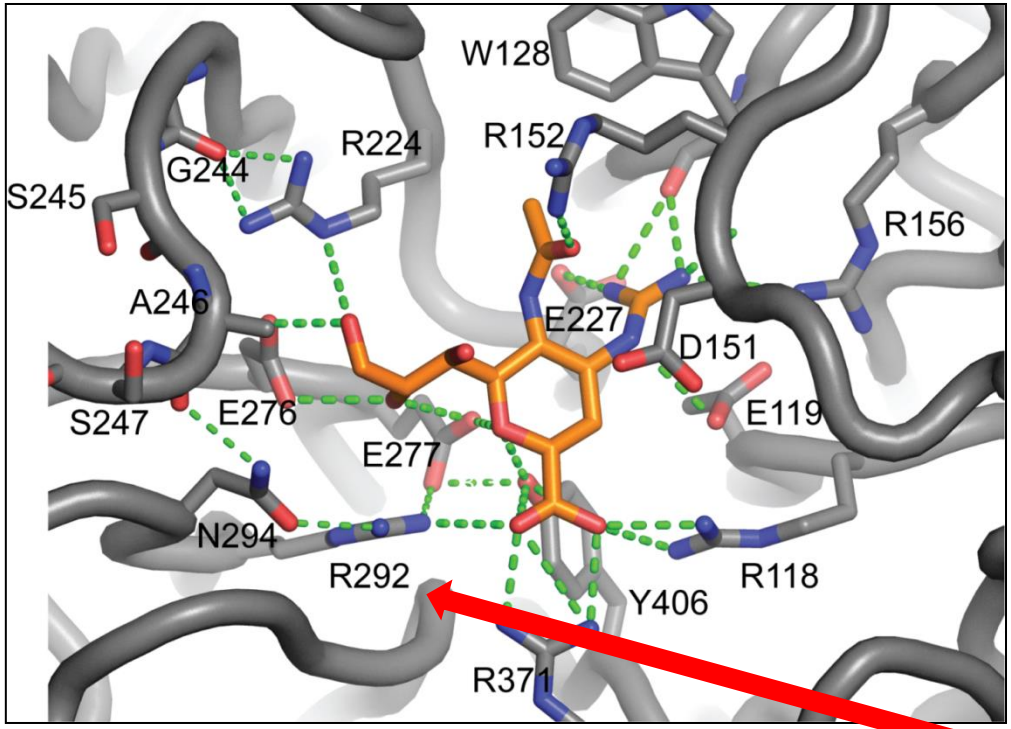
- **On March 31, 2013, the China Health and Family Planning Commission notified WHO of 3 cases of human infection with influenza A (H7N9)**
  - All 3 cases presented with respiratory tract infection with progression to severe pneumonia and breathing difficulties
  - 2 deaths, 1 patient in critical condition
  - China CDC (Beijing) immediately released full genome sequence information fo



Laboratory Investigation During Response

- **Partial or complete genetic sequences from 54 cases have been posted**
  - All 8 genes are from avian origin
- **Genetic changes in the sequences have been associated with adaptations leading to**
  - Enhanced virus binding to mammalian respiratory cells
  - Enhanced replication in mammalian respiratory cells
  - Increased severity of infection
- **Antiviral susceptibility**
  - Overall, susceptible to oseltamivir and zanamivir
  - Viruses from 6 cases have genetic markers for resistance to oseltamivir and zanamivir; clinical relevance is unknown
  - Resistant to amantadine and rimantadine

Key Neuraminidase Molecular Markers



NA	119	136	151	222	224	247	274	276	292	294	371

Recombinant NAs Used for Antiviral Susceptibility Assessment

Application of Genetic Sequence Information

- **Proteomics: Transient expression system for the neuraminidase (NA) enzyme using H7N9 framework**
  - Allows mutations to be introduced into the NA
  - Mutations R292K, E119V, I222R, I222K, and other mutations were detected in H7N9 viruses recovered from drug-treated patients in China
- **Expressed NA can be evaluated for susceptibility to licensed antiviral drugs**
  - R292K conferred significant resistance

NA, Neuraminidase

Susceptibility of A/Taiwan/1 (H7N9) to NA Inhibitors

- Recovered from a patient after oseltamivir treatment with standard dose (5 days) + double dose (5 days)
- R292K was detected and reported
- Viruses passed in eggs, were tested in pyrosequencing assays

A/Taiwan/1	R292K (%)	I222R (%)	I222K (%)	E119V (%)
Passage 1	30	53	6	6
Passage 2	9	45	4	33

Chang SY, et al. Lancet 2013;381:1621
Nguyen HT, et al. Antiviral Therapy 2012;17:159-73
Egg isolate (E1), kindly provided by Taiwan CDC, was propagated US CDC

Antiviral Susceptibility Assay A/Taiwan/1/2013 with 4 Mutations

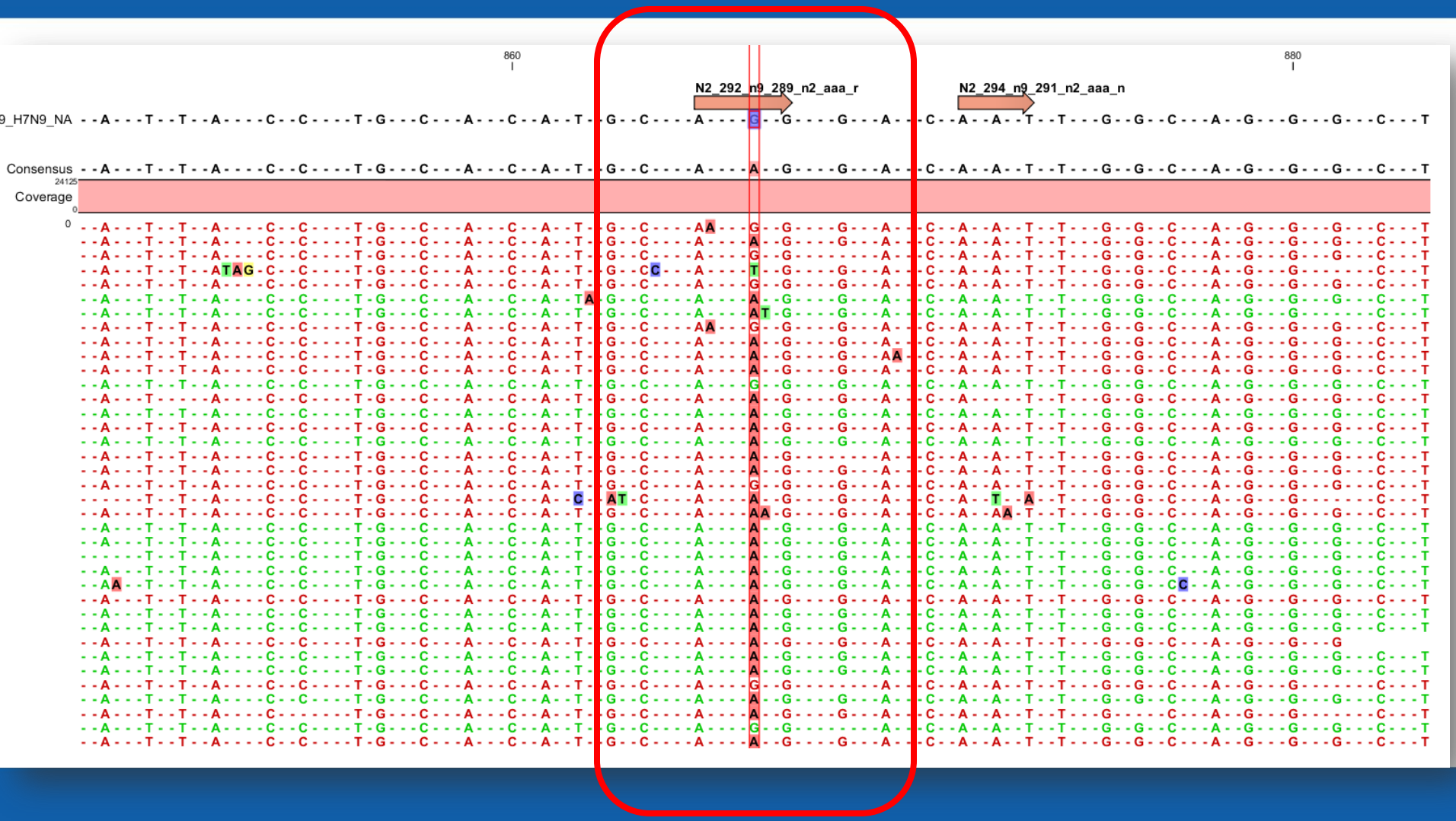
Mutation	Oseltamivir*	Zanamivir*	Peramivir	Laninamivir
R292K	>1000	55	>1000	22
I222K	32	8	6	13
I222R	3	12	2	14
E119V	84	9	1	2

*** Licensed for use in the United States**

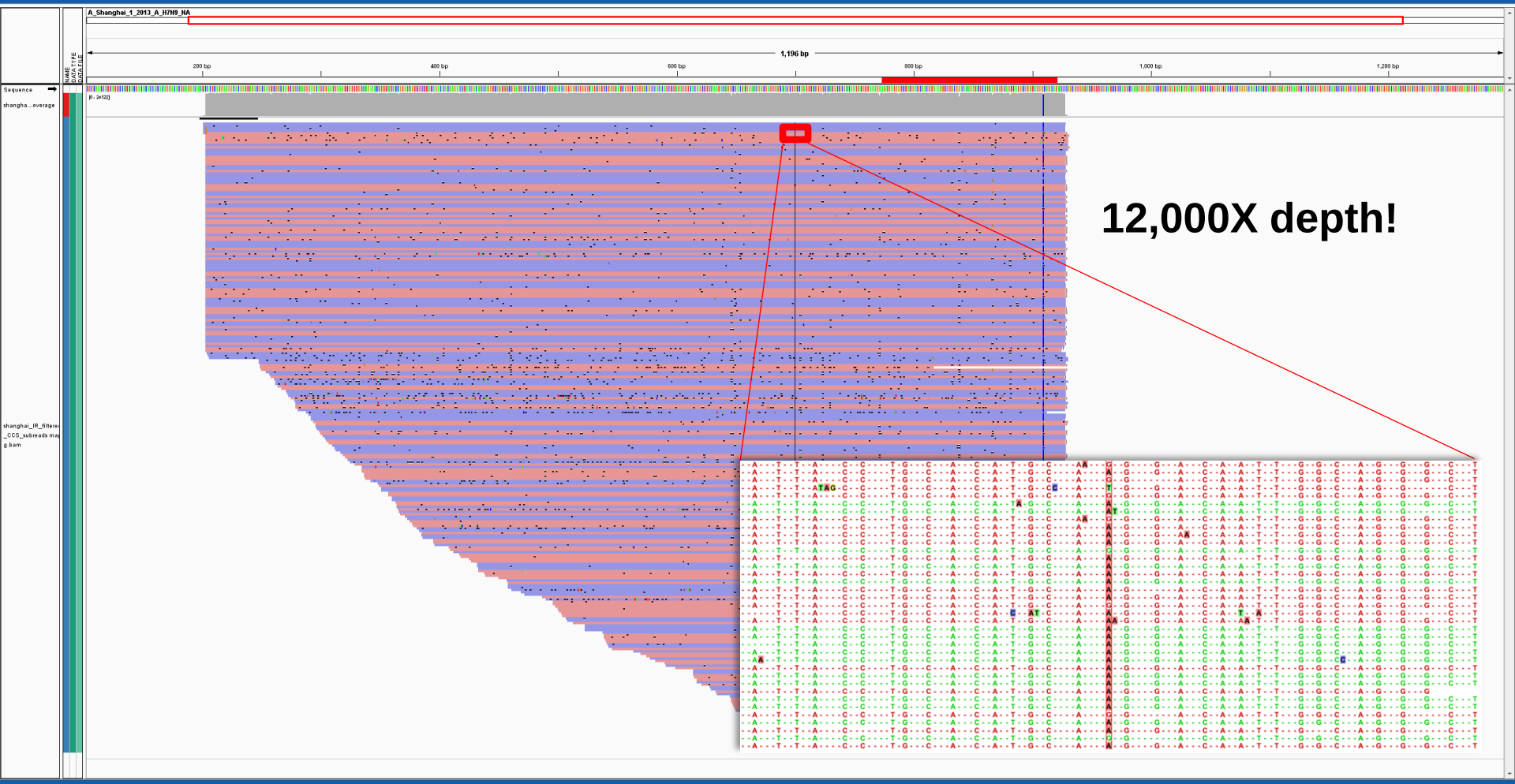
Normal:	<10 fold
Reduced:	10–100 fold
Highly reduced:	>100 fold

Sleeman K, et al. (submitted for publication)

Detect Nucleotide Changes Resulting from Laboratory Passage



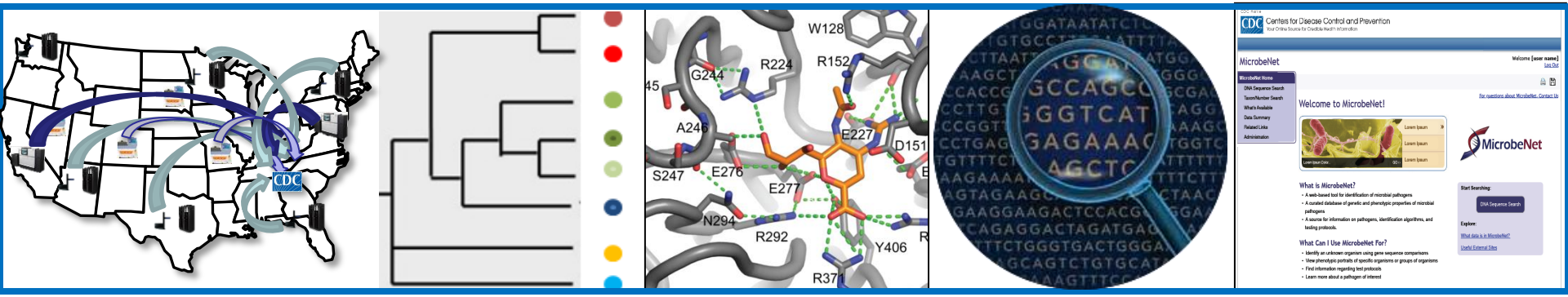
Deep Sequencing is Really Deep for Influenza



Public Health Implications

- **Genetic sequence data**
  - Can rapidly identify genetic changes that may impact the effectiveness of antiviral agents
- **Proteomics and analysis of the variant population**
  - Allows monitoring for mutations that might herald the emergence of new phenotypes affecting transmissibility, viral fitness, and vaccine design
- **Deep sequence analysis of variant populations within a single clinical specimen**
  - Allows detection of linked mutations that might affect the progression and severity of disease within an individual patient

MicrobeNet™: Connecting Public Health Laboratories



John R. McQuiston, PhD

Team Lead, Special Bacteriology Reference Laboratory

Division of High Consequence Pathogens and Pathology

National Center for Emerging and Infectious Diseases, CDC

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

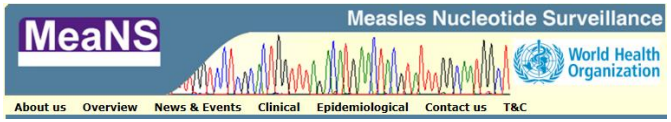
Examples of Existing Databases

- **GenBank**
  - At the National Center for Bioinformatics Information/NIH
  - An annotated collection of all publicly available DNA sequences http://www.ncbi.nlm.nih.gov/genbank

- **Global Initiative on Sharing All Influenza Data**
  - http://platform.gisaid.org/epi3/frontend#4f52ba

- **Measles Nucleotide Surveillance (MeaNS)**
  - http://www.hpa-bioinformatics.org.uk/Measles/Public/Web_Front/main.php

- **Polio Nucleotide Surveillance**
  - Similar to MeaNS (under construction)



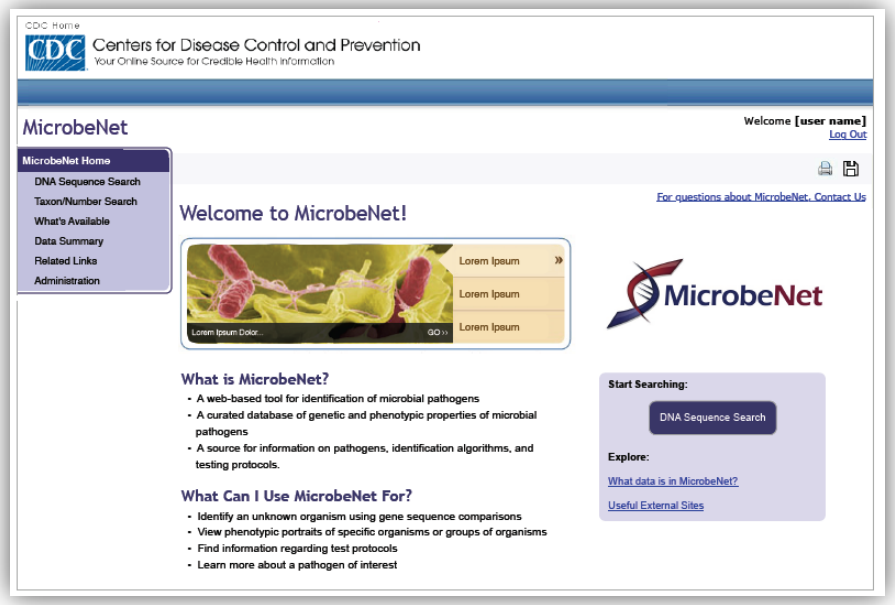
- **An online tool developed at CDC for the public health and clinical laboratories to access CDC-based technology and databases**
- **MicrobeNet™ will connect public health laboratories with CDC subject matter experts for help with identifying pathogens**
  - Each tool or module helps identify the genus or species level and also connects the user back to the supporting data on the diverse pathogens

Genus	**species**
Mycobacterium	***tuberculosis***

https://microbenet.cdc.gov

MicrobeNet™ Capabilities

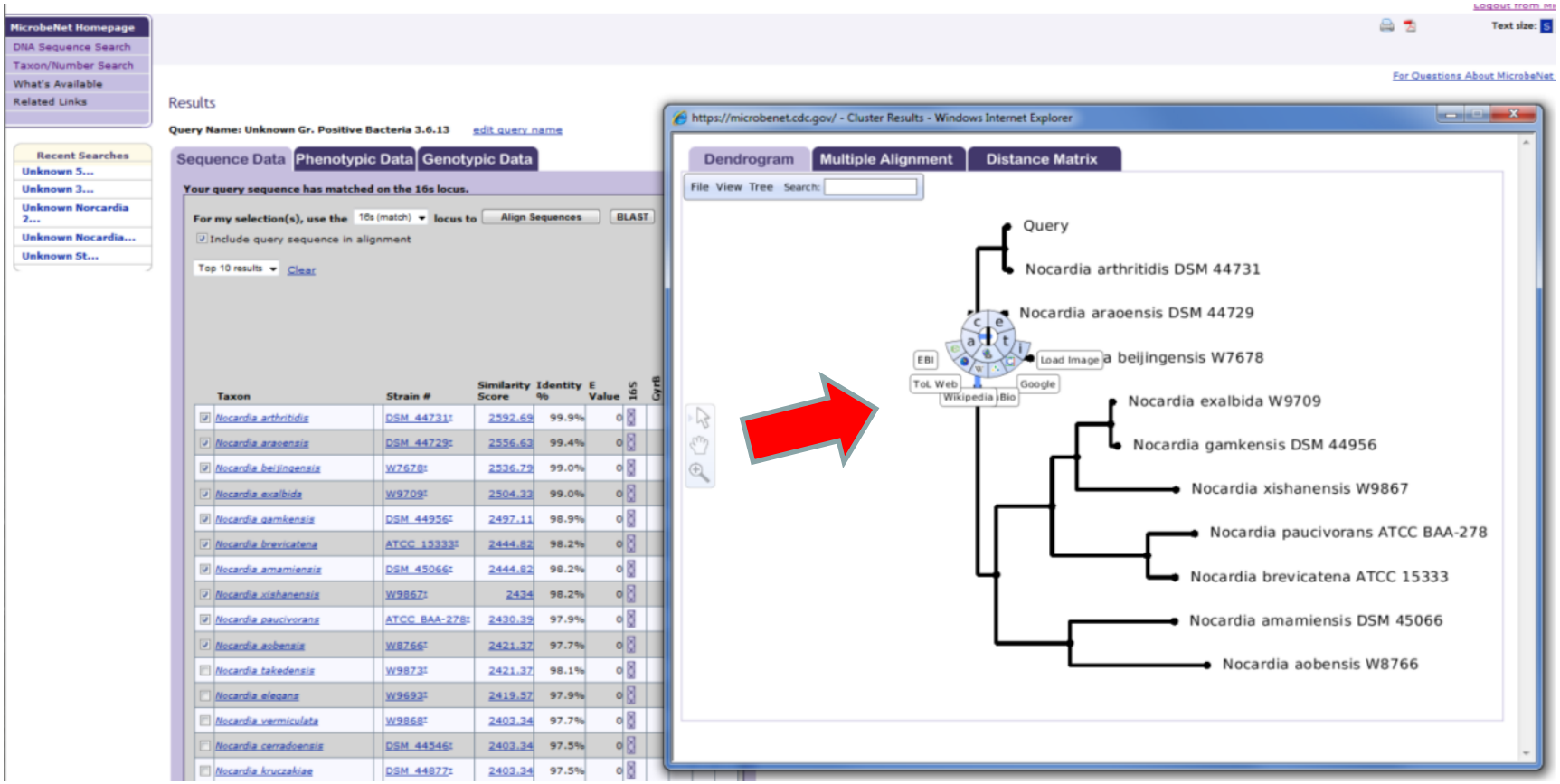
- **Current**
  - DNA sequence
- **Future**
  - Phenotypic data
  - Biochemical tests
    - Fermentation of sugar
    - Production of catalase
  - Proteomics
    - MALDI-TOF mass spectrometry
  - Genomics



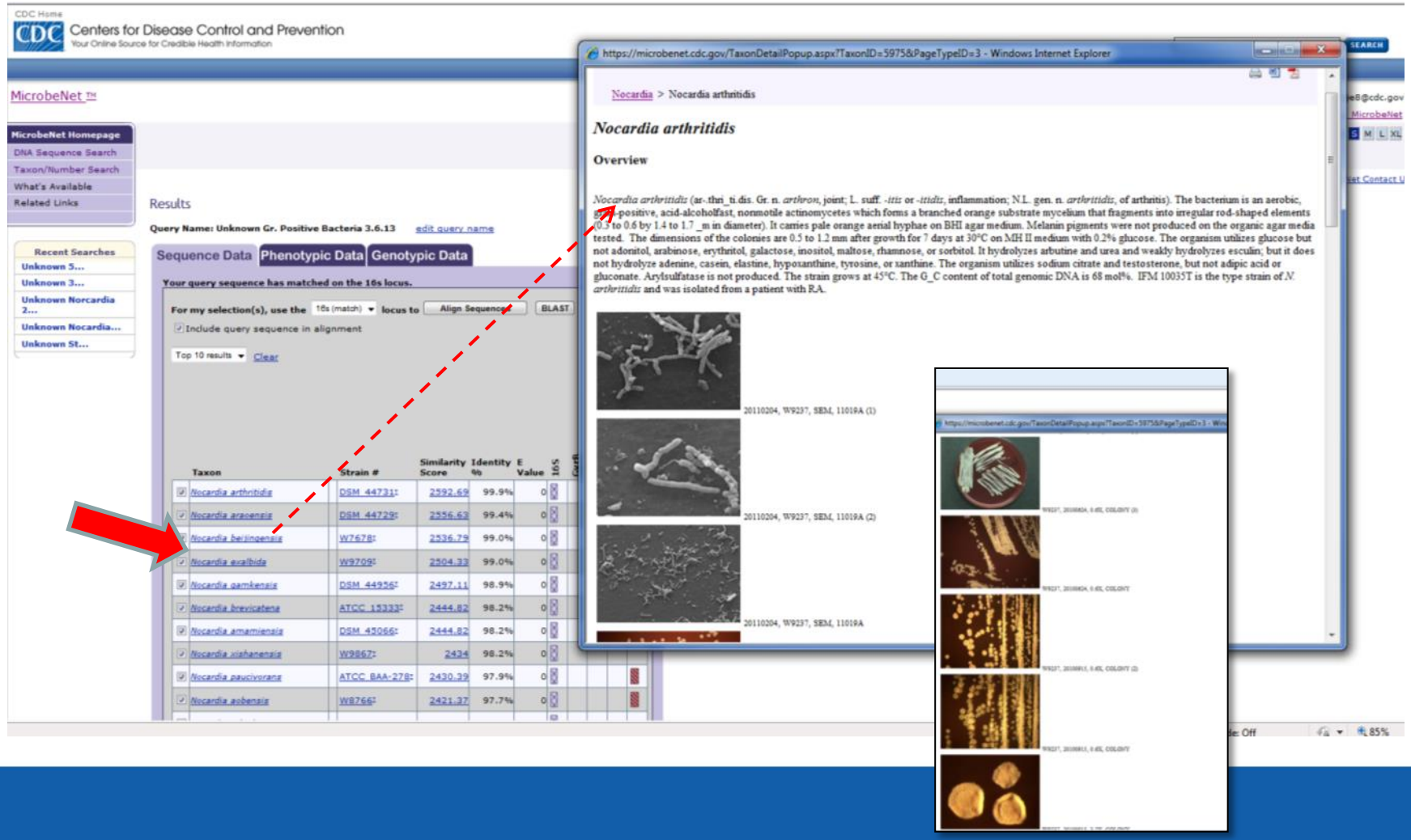
https://microbenet.cdc.gov

MALDI-TOF, matrix-assisted laser desorption/ionization–time of flight

Searching MicrobeNet™ to Identify an Unknown Sequence



Information in MicrobeNet™



Information in MicrobeNet™

Centers for Disease Control and Prevention
Your Online Source for Credible Health Information

MicrobeNet ™

MicrobeNet Homepage
DNA Sequence Search
Taxon/Number Search
What's Available
Related Links

Recent Searches
Unknown 5...
Unknown 3...
Unknown Norcardia 2...
Unknown Nocardia...
Unknown St...

Results

Query Name: Unknown Gr. Positive Bacteria 3.6.13 edit query name

Sequence Data | Phenotypic Data | Genotypic Data

Your query sequence has matched on the 16s locus.

For my selection(s), use the 16s (match) locus to Align Sequences BLAST

Include query sequence in alignment

Top 10 results Clear

Taxon	Strain #	Similarity Score	Identity %	E Value	16S	GyrB
Nocardia arthritidis	DSM 44731T	2592.69	99.9%	0		
Nocardia araoensis	DSM 44729T	2556.63	99.4%	0		
Nocardia beijingensis	W7678T	2536.79	99.0%	0		
Nocardia exalbida	W9709T	2504.33	99.0%	0		
Nocardia gamkensis	DSM 44956T	2497.11	98.9%	0		
Nocardia brevicatena	ATCC 15333T	2444.82	98.2%	0		
Nocardia ansamiensis	DSM 45066T	2444.82	98.2%	0		
Nocardia xishanensis	W9867T	2434	98.2%	0		
Nocardia paucivorans	ATCC BAA-278T	2430.39	97.9%	0		
Nocardia aobensis	W8766T	2421.37	97.7%	0		

Done

https://microbenet.cdc.gov/ - Strain Detail - Windows Internet Explorer

Nocardia arthritidis

Strain Number: DSM_44731T

Synonymous Strain Number(s): IFM_10035T, W9237T

Type Strain: Yes

Sequence | Phenotype | Data Sources | References

Panel: **Actinomycetes Panel**

MicrobeNet #:

Actinomycetes Panel

Antimicrobial Susceptibilities

Amikacin:	<0.25 (S)
Amoxicillin/Clavulanate:	1/0.5 (S)
Ampicillin:	2 (S)
Ceftriaxone:	2 (S)
Ciprofloxacin:	>8 (R)
Clarithromycin:	4 (I)
Imipenem:	>32 (R)
Linezolid:	2 (S)
Minocycline:	2 (I)
Trimetho/Sulfa:	0.25/4.8 (S)
Vancomycin:	>32 (R)

Acid production from Gordon Sugars

Adonitol:	-
Cellobiose:	-
D-Fructose:	+
D-Galactose:	-

https://microbenet.cdc.gov/ - Strain Detail - Windows Internet Explorer

Glycerol:	+
i-Erythritol:	-
i-myo-Inositol:	-
Lactose:	-
L-Arabinose:	-
L-Rhamnose:	-
Maltose:	-
Mannose:	-
Melibiose:	-
Raffinose:	-
Salicin:	+w
Sucrose:	-
Trehalose:	-

Growth Characteristics

Growth in Lysozyme:	+
Growth at 25°C:	+
Growth at 35°C:	+
Growth at 45°C:	+

Other Characteristics

Arylsulfatase production:	-
Utilization of citrate:	-
Nitrate reduction (Routine):	-
Utilization of acetamide:	-

Hydrolysis of

Adenine:	-
Casein:	-
Hypoxanthine:	-
Tyrosine:	-
Xanthine:	-
Esculin:	dark
Urea:	+

MicrobeNet™ and Genomics Reference Laboratory of the Future

Public Health Laboratory A

- Performs whole genome sequencing
- Selects tests they want to perform through MicrobeNet™

Tests Requested	Results of Tests
☑ Genus	*Mycobacterium*
☑ Species	*tuberculosis*
☐ Biochemical profile	
☑ AST profile	Rifampin (R) Isoniazid (R) Ethambutol (S) Pyrazinamide (S)
☐ SNP profile	
☑ Spoligotype	Spoligotype: 000000000030771
☐ MLST	

AST, Antimicrobial susceptibility testing
SNP, Single nucleotide polymorphism (mutation)
MLST, Multi-locus sequence typing

Current Searches for Unusual Pathogens
Jan–Sep 2013, 747 Searches

Nocardia
Neisseira
Brevibacterium
Corynebacterium
Gordonia
Klebsiella
Haemophilus

Current Searches for Unusual Pathogens Jan–Sep 2013

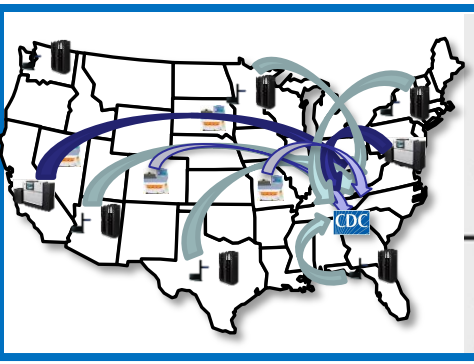
- **Who is using MicrobeNet™ now?**
  - 64 users from 28 state health departments
  - 3 international users from Brazil and Republic of Georgia
- **Training**
  - 2 workshops completed in February and April 2013
  - 1 workshop planned for December 2013
  - Collaborative effort with Association of Public Health Laboratories and National Laboratory Training Network

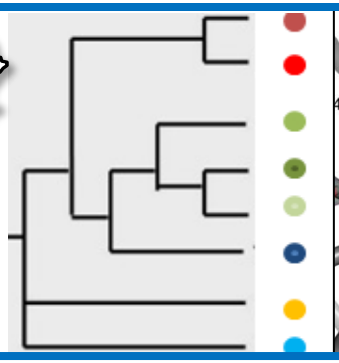
MicrobeNet™ Future Applications

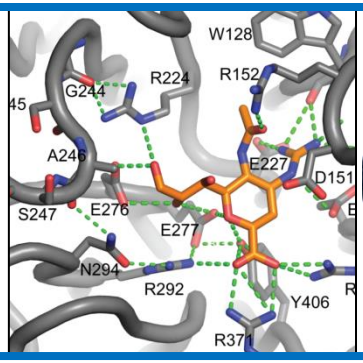
- **New tools (modules) for enhanced search capability**
  - Phenotypic: Fermentation of sugars, other biochemical tests
  - Gene sequence typing: E.g., MLST
  - Protein mass spectrometry: E.g., MALDI-TOF
- **Expanded databases for**
  - Fungal and viral pathogens
  - Additional bacterial pathogens (from current 200 to 3,000 species)
- **Whole genome search capability**
- **Geographic location capacity**
- **Enhanced communication with public health partners**
  - Contact information (CDC Branch) available for each species

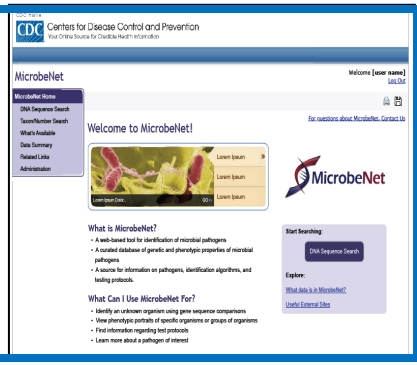
MLST, Multi-locus sequence typing
MALDI-TOF, Matrix-assisted laser desorption/ionization–time of flight

Advancing AMD









David M Engelthaler

Director, TGen North

Translational Genomics Research Institute, Flagstaff, Arizona

tgen CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

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

The Public Health Need for AMD and CDC

- **CDC is well placed to seize the momentum**
  - Advances in both genomic science and technology
- **Not if, but when … Nice To Have to Need To Have**
  - Public Health Needs: Genomics, proteomics, bioinformatics, and big data science and technology
- **AMD**
  - **Science evolution to public health revolution**

What Will It Take

- **Key pieces**
  - Genomics and proteomics cores
    - With development and improvement strategies
  - Advanced and accessible bioinformatics
    - Big data needs big investment
  - Willingness to adopt and adapt the state of the science
    - Example: Human microbiome to population microbiome
    - Next: Incorporation of the community microbiome into population health surveillance and response strategies

Leadership

- **AMD needs 21st century leadership**
  - A grave and growing public health need
    - CDC has always led the world in public health response
  - A culture of advanced science at CDC
    - Adapt and adopt the state of the science to once again create state-of-the-art public health
  - Commitment in the face of challenges
    - Leadership will drive this change, not funding

Good News!

- **CDC has the leadership to make this happen at all levels**
  - From program to director
- **The science and technology exists**
  - Rapid adoption, adaptation, and advancement is already occurring
- **AMD will be a reality**
  - Federal, state, and local public health

AMD Is Changing the Landscape for Combatting Infectious Diseases

"We are on the verge of revolutionizing our ability to diagnose infectious diseases, investigate and control outbreaks, understand transmission patterns, develop and target vaccines, and determine antimicrobial resistance—all with increased timeliness and accuracy and decreased costs"

=

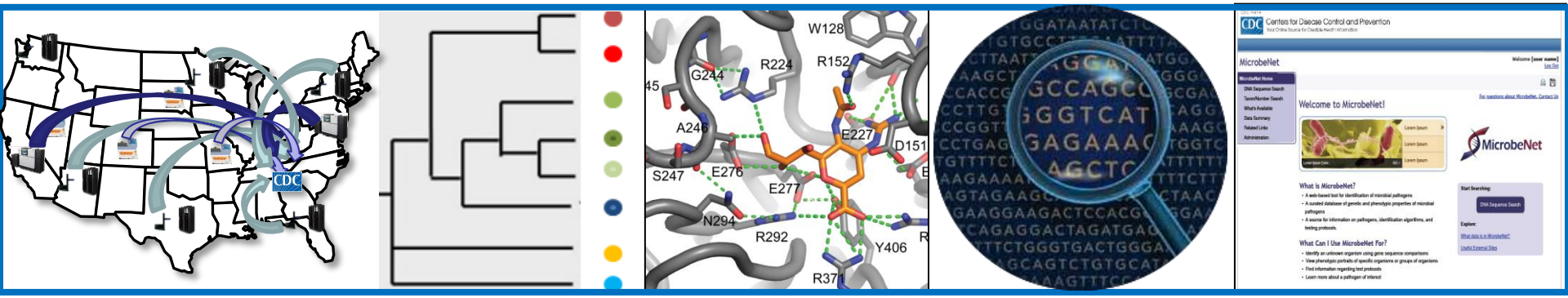
=

http://www.cdc.gov/AMD

The Way Forward

- **CDC AMD cores**
  - Proteomics
  - Genomics: Deep sequencing
  - Bioinformatics
- **A connected public health system**
  - Program expertise
  - Local support
  - Global vision and partnerships

The Future

- A global genomics perspective will allow us to better understand health, disease, and the human condition
- Through the AMD Initiative, CDC is poised to lead public health into the 21st century

Advanced Molecular Detection for Improving Public's Health