CDC’s Advanced Molecular Detection (AMD) Program and Public Health

Scott Sammons
Technology Officer
Office of Advanced Molecular Detection

National Center for Emerging and Zoonotic Infectious Diseases
Office of Advanced Molecular Detection
NGS: A Disruptive Technology

Cost to sequence 1 million basepairs

- **$5,300/MB**
- **$5/MB**
- **50,000bp/day**
- **60,000,000,000 bp/day**

**VOLUME OF RAW DATA**

- **Kilobytes** (10^3 bytes)
- **Terabytes** (10^12 bytes)

**Maximum output per instrument**

- **2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013**
Drivers for Innovation and Change

- Faster time-to-answer.
  - More relevant, actionable information for public health response.
- Higher resolution/accuracy.
  - More specific, useful information from your data.
- Automate-able.
  - Different criteria for outbreak response/surveillance applications.
- Objective and reproducible results.
- Epidemiologic/clinical concordance.
- Cost effectiveness.
  - Comprehensive data capture can reduce downstream testing cost.
  - Cost per test for many of these technologies is extremely competitive, relative to conventional methods, and is decreasing.
WGS and Pathogen Genomics: Advantages

- It’s universal...
  - DNA/RNA sequencing workflows and approaches can be applied to a wide range of pathogenic organisms.

- It’s fundamental...
  - Genomics is a cornerstone for other “omic” approaches
  - Sequence databases starting point for assay devel./validation.

- It’s objective...
  - Sequence-based methods avoid subjectivity of phenotypic or fragment-based approaches. Volume of data → internal controls.

- It’s (relatively) future proof...
  - Comprehensive sequencing captures the features you know about, and those you don’t. Quality may change, but the sequence will not.
  - This makes it possible to back-test future approaches/targets on the data you collect today.
WGS and Genomic Epidemiology: Limitations

- It lacks standardization...
  - WGS is a rapidly-evolving technology space, both in terms of sequencing and analytics.
  - Standards and mechanisms for data/metadata analysis, storage and exchange remain under active debate and development.

- Comprehensive databases are still being built...
  - Without a useful baseline understanding of pathogen features/diversity, interpretation may be limited.
  - Need curated and comprehensive epi-linked reference databases.

- Many analyses require specialized bioinformatics infrastructure and staff.
  - Bioinformaticists, DBAs, programmers, system administrators, etc.
  - Technical and computational complexity of tasks can vary widely.

- Data management, retention and release. Storage. LIMS.
OUTBREAK INVESTIGATION

IDENTIFY

- Same pathogen?
- Does it fit the clinical syndrome?
- Is it present in all your cases?
- Is it absent in your controls?
- Biology/ecology of the organism?
...

SUBTYPE

- Same strain? How different?
- Does the clustering fit the hypothetical scenario? Eg: Timeframe? Spatial? Chain of transmission?

CHARACTERIZE

- Important features?
- Are there characteristics that could explain emergence or affect public health response?

CULTURE/PHENOTYPIC

MOLECULAR

Timeframe: DAYS

MOLECULAR

PHENOTYPIC

MOLECULAR

GENOMICS, Other -OMICS

RESPONSE

IDENTIFY/CHARACTERIZE

SUBTYPE/CHARACTERIZE

INFORM PREVENTION STRATEGIES

Human Animal Environmental Samples

Timeframe: DAYS/WEEKS
Advanced Molecular Detection (AMD)

2011: “Bioinformatics Blue Ribbon Panel”
2014: Congress establishes AMD initiative

• 5-year, $30m-per-year modernization program
• Core goals:
  o Improving *pathogen detection and characterization*
  o Developing *new diagnostics* to meet public health needs
  o Supporting genomic and *bioinformatics* needs in the US public health system
  o Implementing enhanced, sustainable, *integrated information systems*
  o Developing tools for prediction, modeling and *early recognition of emerging infectious threats*
AMD’s Focus

- NGS & bioinformatics: transforming pathogen detection and characterization
  - At CDC
  - In state and local health laboratories
- Other technologies: MALDI-TOF, optical mapping
- Infrastructure: storage, networking, high-performance computing, cloud
- Workforce development
- Collaboration
## Highest Impact Areas in Near Future

<table>
<thead>
<tr>
<th>Domain</th>
<th>Objective</th>
</tr>
</thead>
</table>
| Bacterial foodborne pathogens (PulseNet, 80+ labs) | • Move from PFGE to WGS  
• Address the threat from culture-independent diagnostic tests (CIDT) |
| Tuberculosis                                | • Move from MIRU/VNTR typing to WGS  
• Develop means of WGS directly from sample                                                        |
| Influenza                                   | • Move from traditional typing of isolates to WGS directly from samples  
• Develop cloud-based information system                                                         |
| Hospital-acquired and antimicrobial resistant organisms | • Develop standards for typing  
• Develop methods for assessing relatedness in outbreaks  
• Develop microbiome disruption index  
• Move GC typing from traditional phenotyping to WGS                                                |
AMD Portal: Standardizing Data Flow

State PHL

Phl User Community

NCBI

Cdc User Community

Cdc High Performance Computing CoE

HPC Clusters

Storage Array (3.6PB)

Influenza

TB

Etc.
Other AMD Foci at CDC

- **Diagnostics**
  - MicrobeNet—ID of unusual bacterial pathogens
  - Unexplained respiratory disease outbreaks
  - HCV infection classification
  - Digital PCR for hepatitis viral load

- **Malaria**
  - Anti-malarial resistance in plasmodia
  - Insecticide resistance in vectors

- **WGS for bacterial pathogen typing and inference of phenotype**
  - Pneumococcus, GAS, GBS
  - Meningococcus, *H. influenza*
  - *Legionella*
  - *B. anthracis*: forensics
  - *B. pertussis*: subtyping and investigation of emergence

- **Genomic typing of eukaryotic pathogens**
  - *Cyclospora, Cryptosporidium*
  - *Naegleria*
  - *Leishmania*
  - *Coccidioides*

- **Diagnosis and typing of viral pathogens**
  - Viral vaccine-preventable diseases
  - HIV
  - HCV (“GHOST” system)
  - Dengue & chikungunya

- **Pathogen discovery**
  - Application of NGS
  - Arboviruses
  - Filoviruses
Cross-Cutting Issues

- Workforce development
- Infrastructure
- Metagenomics—addressing culture-independent diagnostic testing
- Integration of genomic and epidemiologic data
Developing A Bioinformatics Workforce for Public Health

- Like microbiology and epidemiology, good bioinformatics is increasingly critical to good PH.

- Challenges with recruitment and retention
  - Skilled bioinformaticians are a highly sought-after resource.
  - Career bioinformaticians with relevant public health/microbiology experience are an extremely rare breed.
  - It is more common to find microbiologists with on-the-job bioinformatics experience -- expand/enable through training.
  - Limited state/federal workforce support (job series, competencies, training...) for new technical labor categories
  - Technical and informatics challenges (compute/storage/net)

- Interdisciplinary communication is vital: translating between program scientific needs and bioinformatics technical specialists. Team-based approaches, hub and spoke model.

- AMD training and capacity building for PHL: stay tuned!
Recent Applications
Whole Genome SNP Typing (WGST)

Reference Sequence/Genome

1. ACTAGA
2. ACTAGT
3. TCTACT
wgMLST / Binary Typing

- **SuperMLST**
  - Same concept as conventional MLST, but with 20+ loci
  - Each sequence is matched against a database of numbered alleles.
  - Example: (1,1,2,1,4,1,5,5,5,1,2,3,4,1,2,3,1,4,5,6,7,2,3,2,2,1)

- **Binary typing**
  - Presence or absence of specific virulence markers/sequences.
  - Example: Binary matrix (0,1,1,0,1,1,0,1,1,1,1,1,0)
National network of laboratories that conduct standardized molecular typing of food-borne bacterial pathogens.

Permits rapid detection of clusters of indistinguishable strains.

PFGE is robust and proven, but low-throughput, labor-intensive, subjective, and dependent on cultures.

Non-culture based diagnostics for foodborne pathogens are increasingly common in clinical practice.
Nationwide “Real-time” Listeriosis Surveillance using WGS

- Ongoing collaboration with FDA, NCBI, USDA and several state public health laboratories. Participation from several countries: France, UK, Denmark, Australia and Canada.

- Goal: Near “realtime” whole genome sequencing and analysis of isolates from all U.S. clinical cases of *Listeria monocytogenes* infection, as well as those from food/environmental sources.

- WGS within ~1 week of isolate receipt.

- Initiated: September 1, 2013.
Common Bioinformatics Platforms: Next Generation PulseNet

A number of other CDC programs are evaluating the use of BioNumerics infrastructure: mening, TB, GC, legionella, strep, HAIs, malaria, HIV
Listeria Cluster Metrics, Before and After WGS

- **No. of clusters detected**: Pre-WGS (09/12--08/13) = 14, WGS Year 1 (09/13--08/14) = 19, WGS Year 2 (09/14-08/15) = 21
- **No. of clusters detected sooner or only by WGS**: Pre-WGS = 0, WGS Year 1 = 6, WGS Year 2 = 6
- **No. of outbreaks solved (food source identified)**: Pre-WGS (09/12--08/13) = 1, WGS Year 1 (09/13--08/14) = 4, WGS Year 2 (09/14-08/15) = 9
- **Median no. of cases per cluster**:
  - Pre-WGS (09/12--08/13) = 6 cases
  - WGS Year 1 (09/13--08/14) = 4 cases
  - WGS Year 2 (09/14-08/15) = 3 cases
- **No. of cases linked to food source**:
  - Pre-WGS (09/12--08/13) = 6 cases
  - WGS Year 1 (09/13--08/14) = 16 cases
  - WGS Year 2 (09/14-08/15) = 93 cases

Courtesy: Brendan Jackson, Enteric Diseases Epidemiology Branch
Mycobacterium tuberculosis Cluster, State X

Phylogenetic tree based on conventional genotyping (spoligotyping/MIRU-VNTR)

~ 100 Patients

Phylogenetic tree based on WGS
IDU was a primary driver of transmission

Syringe-sharing allowed for explosive growth

### Contact Type

<table>
<thead>
<tr>
<th>Contact Type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDU Partner</td>
<td>60.4%</td>
</tr>
<tr>
<td>Social Contact</td>
<td>25.8%</td>
</tr>
<tr>
<td>Sex + IDU</td>
<td>9.0%</td>
</tr>
<tr>
<td>Sex Partner</td>
<td>5.8%</td>
</tr>
</tbody>
</table>

Unpublished preliminary data
Inferred HIV Transmission Network, Indiana, 2015

Best analogue of the network that HIV spread across

82.3% of contacts are related to IDU

<table>
<thead>
<tr>
<th>Contact Type</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDU Partner</td>
<td>60.4%</td>
<td>72.9%</td>
</tr>
<tr>
<td>Social Contact</td>
<td>25.8%</td>
<td>16.0%</td>
</tr>
<tr>
<td>Sex + IDU</td>
<td>9.0%</td>
<td>9.4%</td>
</tr>
<tr>
<td>Sex Partner</td>
<td>5.8%</td>
<td>1.7%</td>
</tr>
</tbody>
</table>

Unpublished preliminary data
Inferred HCV Transmission Network, Indiana, 2015  
(Analysis: GHOST)

- 240 HCV cases
- 19% infected with >1 HCV genotype
- 19 clusters of transmission
- Each cluster is shown in different color
- Unrelated cases are shown in dark blue
- One transmission cluster includes 91 cases (shown in orange)
Hepatitis C Outbreak, Healthcare Facility A
Quasispecies Analysis—the GHOST System

- 18 cases
- 154,233 reads
- 33,767 unique sequences.
- Each node is a unique sequence
- Different patients are shown in different colors
- Two sequences are linked if they differ at a single nucleotide position
Pertussis Incidence, United States, 1982 to 2012
Pertussis Outbreak, California, 2010
**Influenza Virus Processing Pipeline**

**Before NGS**

1. **Sample** → Batching & Shipping → Viral culture (typically, 2-3 passages) → Phenotyping, including antigenic typing → Antigenic subtyping → Sanger Sequencing

   - 1–2 weeks (longer if international)
   - Selection based on antigenic type
   - N ≈ 10,000
   - N ≈ 2000

**With NGS**

1. **Sample** → NGS → Viral culture (typically, 2-3 passages) → Phenotyping, including antigenic typing

   - Selection based on sequence
   - N > 10,000
   - N ≈ 2000
   - 1 week
   - 3-4 weeks

**Advantages**

- Faster
- Cheaper
- More samples
- More data
- Better data
Identification of Yellow Fever by next generation sequencing

Dr. Laura McMullan and Dr. Stuart Nichol (DHCPP/VSPB)

November 23, 2010

Mystery Ebola-Like Illness Kills 13 in Uganda

Gregory Bronch, Global Post

KAMPALA, Uganda — Thirteen people have died in northeastern Uganda from a mysterious disease. The World Health Organization and the Centers for Disease Control are investigating the outbreak of the disease that appears to be similar to the Ebola and Marburg hemorrhagic viruses.

The Ministry of Health in Uganda on Wednesday reported 13 deaths from a disease that resembles Ebola and Marburg hemorrhagic viruses. The epidemic is in the remote Acholi district.

The disease has the same symptoms as Ebola and Marburg hemorrhagic viruses, and the Ministry of Health is investigating the cause of the disease. The disease has not been identified as Ebola or Marburg hemorrhagic viruses.

The Ministry of Health is working with the World Health Organization and the Centers for Disease Control to investigate the outbreak. The Ministry of Health has also set up a quarantine area in the Acholi district.

The World Health Organization and the Centers for Disease Control are working with the Ministry of Health to investigate the outbreak. The Ministry of Health is working to contain the disease and prevent its spread.
Clinical case of progressive vaccinia infection from smallpox vaccination of an immune compromised patient

Pox antiviral ST-246 administered which targets pox gene F13L, a major envelope protein which mediates production of extracellular virus

Oral ST-246 given daily and vaccination site sampled over 3 week period
Control swab prior to ST-246
2 weeks after ST-246

C > T

869

943

T > A

Number of Results:

Reference Sequence Position

Variation %
3 weeks after ST-246

C > T
869

T > A
943
Developing A Bioinformatics Workforce for Public Health

- Like microbiology and epidemiology, good bioinformatics is increasingly critical to good PH.

- Challenges with recruitment and retention
  - Skilled bioinformaticians are a highly sought-after resource.
  - Career bioinformaticians with relevant public health/microbiology experience are an extremely rare breed.
  - It is more common to find microbiologists with on-the-job bioinformatics experience -- expand/enable through training.
  - Limited state/federal workforce support (job series, competencies, training...) for new technical labor categories
  - Technical and informatics challenges (compute/storage/net)

- Interdisciplinary communication is vital: translating between program scientific needs and bioinformatics technical specialists. Team-based approaches, hub and spoke model.

- AMD training and capacity building for PHL: stay tuned!
Application deadline for 2016-7 class: **April 1, 2016**

http://www.aphl.org/bioinformatics
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.