# APHL/CDC TB Whole Genome Sequencing Project

## Perspective from a Pilot Laboratory

Ohio Department of Health

Kevin Sohner, BS Microbiology
Ohio Department of Health Laboratory

10th National Conference on Laboratory Aspects of Tuberculosis
April 19, 2017

# Project Background

- **Discriminatory power** of conventional methods (i.e. spoligo and MIRU-VNTR) is sometimes insufficient
- WGS data result in **more focused targeting** compared to genotyping for outbreak investigations
- **Demand** for WGS to support cluster investigations **exceeds the capacity** of DTBE LB

# Project Background

October 29, 2015 - Funding Opportunity Announcement (FOA): Establishment of *Mycobacterium tuberculosis* Complex Whole Genome Sequencing (WGS) Reference Centers.

Options

1. Provide capacity for DTBE cluster investigations
2. Prospective universal WGS surveillance for a particular jurisdiction
3. A combination of option 1 and 2

Six month project to encourage development new methods for *M.tb* DNA extraction and whole genome sequencing optimization

# Challenges

- Methodology
- Staffing/Workload/Cost
- Infrastructure
- Turnaround Time
- Instrument Reliability
- Data Storage/Management
- Bioinformatics Support
- Utility of  MTBC WGS data for prospective samples

# Methodology – Culture & Inactivation

| State | Sample Source | # Samples / Month | Culture Method | Inactivation Method |
|---|---|---|---|---|
| 1 | CDC/ 1 | 48 | LJ | 80°C for 30 min (Water Bath) |
| 2 | CDC | 16 | MGIT | 95°C for 30 min (Dry Heat) |
| 3 | CDC/ 3 | 64 | 7H9 | No Inactivation Used |
| 4 | 4 | 64 | MGIT | 80°C for 60 min (Dry Heat) |
| 5 | CDC/5 | 16 | 7H9/MGIT | 80°C for 60 min (Dry Heat) |

# Methodology – Isolation and Quantification

| State | DNA Isolation Method | Quantification |
|---|---|---|
| 1 | CTAB 65°C | Qubit® dsDNA BR Assay Kit<br>and<br>DNA purity is checked using Nanodrop |
| 2 | FastPrep-24™ 5G Tissue Homogenizer and InstaGene™ Matrix (Bio-Rad) | Qubit® dsDNA BR Assay Kit |
| 3 | FastPrep-24™ 5G Tissue Homogenizer and InstaGene™ Matrix (Bio-Rad) | Qubit® dsDNA BR Assay Kit |
| 4 | FastPrep-24™ 5G Tissue Homogenizer and InstaGene™ Matrix (Bio-Rad) | Qubit® dsDNA BR Assay Kit |
| 5 | In-house validated bead tube | Qubit® dsDNA BR Assay Kit (Nanodrop) |

# Methodology – Library Preparation

| State | Library Prep | PCR Cycles | DNA Template | Sequencing Kit |
|---|---|---|---|---|
| 1 | Nextra XT Kit 2x250 bp | 14 | 12pM | MiSeq Kit V2 500 or V3 600 |
| 2 | Nextra XT Kit 2x250 bp | 14 | 10pM | MiSeq Kit V2 500 |
| 3 | Nextra XT Kit 2x250 bp | 15 | 16pM | MiSeq Kit V2 500 |
| 4 | Nextra XT DNA Sample Preparation Kit | 15 | 11-15pM | MiSeq Kit V2 500 |
| 5 | Nextra XT Kit PulseNet 96 samples | 15 | 13pM | MiSeq Kit V2 500 |

# Staffing/Workload/Cost

Goal: To determine the cost per sample and sequence run to carry out NGS for MTBC.

What was needed

- Number of samples/run & Number of runs/month
- Equipment costs
- Consumables
- Overhead
- Staff costs

# Cost per Run—Crunching the numbers

- Proposed
  - Estimate from RFP Application
- Calculated (Submitted by Site)
  - Submitted data from sites including: Staff, Equipment, Consumables and Overhead
- Corrected (APHL)
  - Standardizing the data—only costs associated with sequencing

# Startup Costs-Equipment & Staff

| | 1 (CDC Only) | 2 (CDC Only) | 3 | 4 | 5 | AVG |
|---|---|---|---|---|---|---|
| Runs/ Month | 2 | 1 | 4 | 5 | 1 | 2.6 |
| Samples/ Run | 16 | 16 | 16 | 15 | 16 | n/a |
| Startup Costs | $205,000 | $236,000 | $5,000 | $165,000 | $118,000 | $158,000 |

Ohio Department of Health

# Cost Per Specimen

| | 1 (CDC Only) | 2 (CDC Only) | 3 | 4 | 5 | AVG |
|---|---|---|---|---|---|---|
| Runs/ Month | 2 | 1 | 4 | 5 | 1 | 2.6 |
| Samples/ Run | 16 | 16 | 16 | 15 | 16 | n/a |
| Proposed Cost/ Run | $2400 | $2268 | $4156 | $3150 | $2890 | $2973 |
| Per Specimen | $150 | $142 | $260 | $210 | $181 | $188 |
| Corrected Cost/Run | $4050 | $3892 | $3065 | $2738 | $3459 | $3441 |
| Per Specimen | $253 | $243 | $192 | $183 | $216 | $217 |

# Startup Costs

| Item | Cost |
|---|---|
| Benchtop Sequencer | ~$99,000 |
| Ancillary NGS Equipment | ~$45,000 |
| DNA Sequencing Supplies and Reagents* | ~$125.00/isolate |
| Service Agreements | ~$16,000/Sequencer |

* Estimate based on PulseNet data

# Costing--Findings

- Major cost of startup is equipment
- Actual cost was greater than estimated costs
- Labs with more runs per month achieved greater efficiency in cost savings
- Consumables are the major portion of cost

  ~ 48-68%

# Infrastructure

- Labs with existing infrastructure (ie sequencers) obtained through ELC AMD, PulseNet, or NARMS funding had lower start up costs and a shorter time to Mtb sequencing data during the pilot project.

# Turnaround Time

- Monthly Line Listed Reports Compiled
  - January-June 2016
- Specimens with incomplete data were excluded for the analysis
  - Exception: date fastq files received at CDC

Ohio Department of Health

# What turnaround times are we monitoring?


Ohio Department of Health

# Turnaround Times

| Monitored Intervals | Turnaround time Median (Mean) days | | | | | |
|---|---|---|---|---|---|---|
| | 1 (n=125) | 2 (n=32) | 3 (n=380) | 4 (n=286) | 5 (n=96) | Overall( n=919) |
| Isolate Rcvd to DNA Extraction | 26 (26) | 50 (49) | 27 (28) | 8 (12) | 21 (33) | 25 (25) |
| DNA Extraction to Sequencing | 18 (20) | 6 (7) | 8 (8) | 6 (6) | 12 (21) | 7 (11) |
| Sequencing to fastQ txfr | 0 (1) | 0 (0) | 1 (1) | 1 (2) | 2 (3) | 1 (1) |
| Isolate Rcvd to fastQ txfr | 44 (47) | 56 (56) | 36 (38) | 16 (20) | 41 (57) | 35 (37) |

Ohio Department of Health

# Turnaround Time-Findings

- Time from extraction to sequencing varied; Average Range: 6-21 days
- Minimal delays from sequencing to data transfer
- 96% of all specimens were sequenced and transferred within 2 months
  - Overall Average: 37 days, Avg. Range 20-57 days

Ohio Department of Health

# Instrument Reliability

- Many labs had issues with MiSeq reliability at some point in the pilot project.
- Problems were addressed fairly quickly by Illumina, although at least one site was down for 5 weeks.

# Data Storage

- At least one lab didn't have the necessary infrastructure to handle the large volumes of data produced by WGS.
- Data/network security was a difficult issue to address with IT staff
- ~20 GB per run; >1 GB per specimen
- 50 sequencing runs is about 1 terabyte!

# Bioinformatics Support

- Bioinformaticians : a relatively new breed; part molecular biologist, part computer scientist
- Few local resources at some state levels
- Create web-based analysis pipelines (Example - Mykrobe Predictor TB)

# Utility of MTBC WGS Data

- Less expensive ways to get at the same data?
- Turnaround time (we want real-time universal WGS and meaningful reports)
- Clinical vs Epidemiological data
- Consider assay replacement costs (PCR, 16S sequencing, genotyping, etc)

# Conclusions

1. Public Health Laboratories can provide increased capacity for WGS samples submitted to CDC for MTBC cluster investigations.
2. Successful sequencing derived from solid and liquid media (even frozen) specimens.
3. Concentration and purity of genomic DNA required for successful WGS was not as stringent as anticipated.
4. Turnaround times in the realm of clinical relevance are possible.
5. Cost per test can be offset/reduced by elimination of redundant assays and greater efficiency

# Acknowledgements

Anne Gaynor, APHL

Paul Zell, APHL

Division of TB Elimination, CDC

California Department of Public Health, MDL

Iowa State Hygienic Laboratory

Michigan Department of HHS-Bureau of Laboratories

New York State Dept. of Health-Wadsworth

Ohio Department of Health

# Contact Information

Kevin Sohner

Microbiology Supervisor

Ohio Department of Health

kevin.sohner@odh.ohio.gov

(614) 644-4668



Ohio Department of Health