# Application of Whole Genome Sequencing (WGS) to Diagnosis of Drug Resistance in Tuberculosis

## Global Consortium for Drug-resistant TB Diagnostics (GCDD)

*Faramarz Valafar*

*faramarz@sciences.sdsu.edu*
*http://informatics.sdsu.edu/*

*Biomedical Informatics Research Center (BMIRC)*
*Office: GMCS 625*
*San Diego State University*

GCDD
Global Consortium for Drug-resistant TB Diagnostics

# Molecular Diagnostics for Drug Resistant TB (DRTB)

Rapid Molecular Tests:

- Hain
- GeneXpert
- Pyrosequencing

Features:

- Inexpensive
- Fast
- Highly localized
  - Usually evaluate the presence of point mutations

# Considerations for Use of Knowledge Gained from Molecular Testing

1. Phenotypic prediction based on discovered single nucleotide polymorphisms (SNPs)

## 1. Prevalence of a SNP

| Table 3: Prevalence of gyrA mutation among GCDD Fluoroquinolone resistant isolates | | |
|---|---|---|
| Mutation | Isolates | Sensitivity |
| 94 A→G | 121 | 41.44% |
| 94 G→A | 41 | 14.04% |
| 90 C→T | 84 | 28.77% |
| 91 T→C | 6 | 2.05% |
| 88 G→T | 4 | 1.37% |
| 74 G→C& 75 C→G | 1 | 0.34% |
| 105 T→G & 112 G→C | 2 | 0.68% |
| No* gyrA mutation | 33 | 11.30% |
| * No mutation in QRDR except for Codon 95 | | |

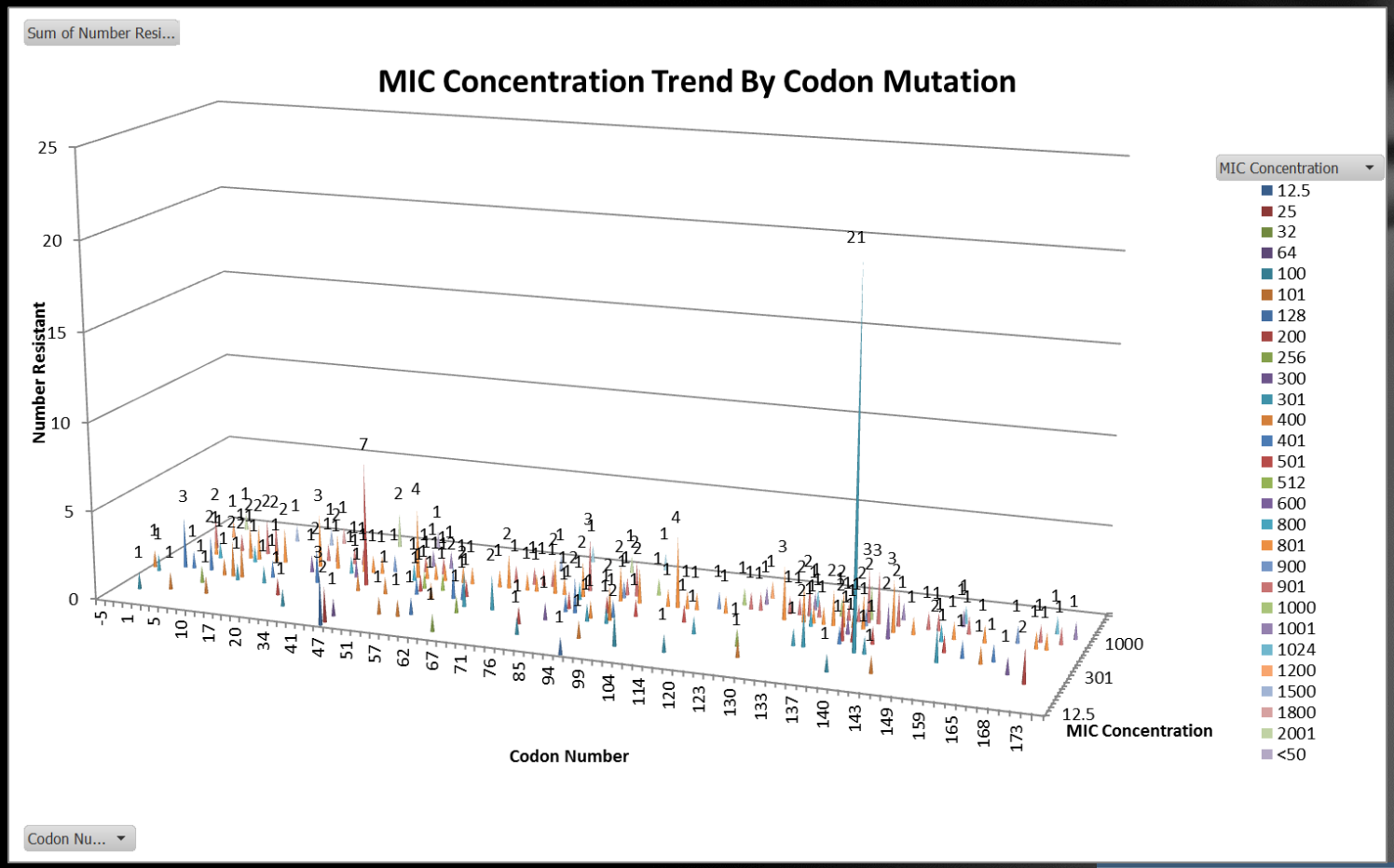
## 2. Regional Differences in Prevalence

| Table 4: Prevalence of rrs A1401G mutation among GCDD "injectable" resistant isolates | | | |
|---|---|---|---|
| | $AMK^R$ | $CAP^R$ | $KAN^R$ |
| India | 85% | 89% | 81% |
| Moldova | 65% | 63% | 26% |
| Philippines | 100% | 100% | 81% |
| South Africa | 91% | 92% | 88% |

| Table 5: Prevalence of eis promoter C-12T mutation among GCDD "injectable" resistant isolates | | | |
|---|---|---|---|
| | $AMK^R$ | $CAP^R$ | $KAN^R$ |
| India | 0% | 0% | 0% |
| Moldova | 10% | 25% | 54% |
| Philippines | 0% | 0% | 0% |
| South Africa | 0% | 0% | 0% |

# Considerations for Use of Knowledge Gained from Molecular Testing

1. Phenotypic prediction based on discovered single nucleotide polymorphisms (SNPs)

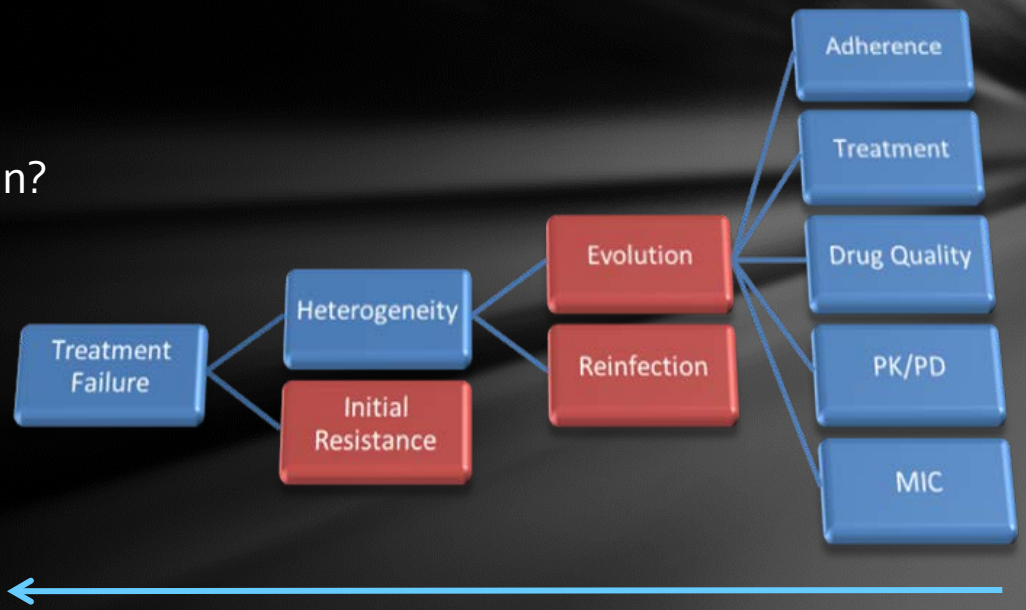


# Considerations for Use of Knowledge Gained from Molecular Testing

1. Phenotypic prediction based on discovered single nucleotide polymorphisms (SNPs)
    1. Consideration: Variable MIC for individual point mutations
        1. Likely causes:
            1. Highly localized consideration of genetic variation.
            2. Not all resistance conferring mutations are known
            3. Multiple functional pathways could cause resistance
        2. Partial solution: Combination of point mutations can help

# Considerations for Use of Knowledge Gained from Molecular Testing

2. Important information that a broader look at the genome could provide:
   2. Treatment Failure:
      2. exogenous secondary infection?
      3. Endogenous evolution?
      4. Heteroresistance?
   3. Population Genetics:
      2. Metagenomic analysis
      3. Population dynamics
   4. Contact Tracing:
      2. Identification of point of original and any secondary infection(s)
      3. Early identification of an outbreak

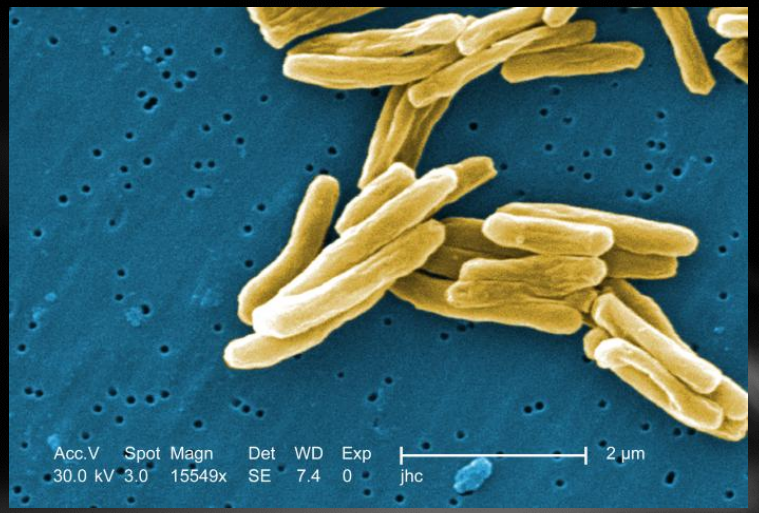


# Whole Genome Sequencing (WGS) as a Tool for a Broader Genomic Perspective

Process:

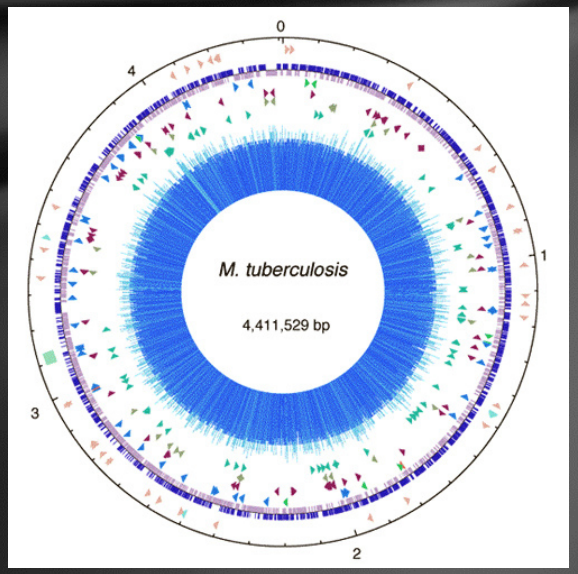
1. Sample Culture
2. DNA Extraction
3. Library Preparation
   1. Often includes (or is followed by) Amplification
4. WGS
   1. Illumina: Genome Analyzer, HiSeq, MiSeq, etc.
   2. Roche: 454
   3. Life Technologies: Ion Torent
   4. Pacific Biosciences: RS
   5. Etc.
5. Bioinformatics

# WGS *Mycobacterium tuberculosis* (*Mtb*)




## What to consider:

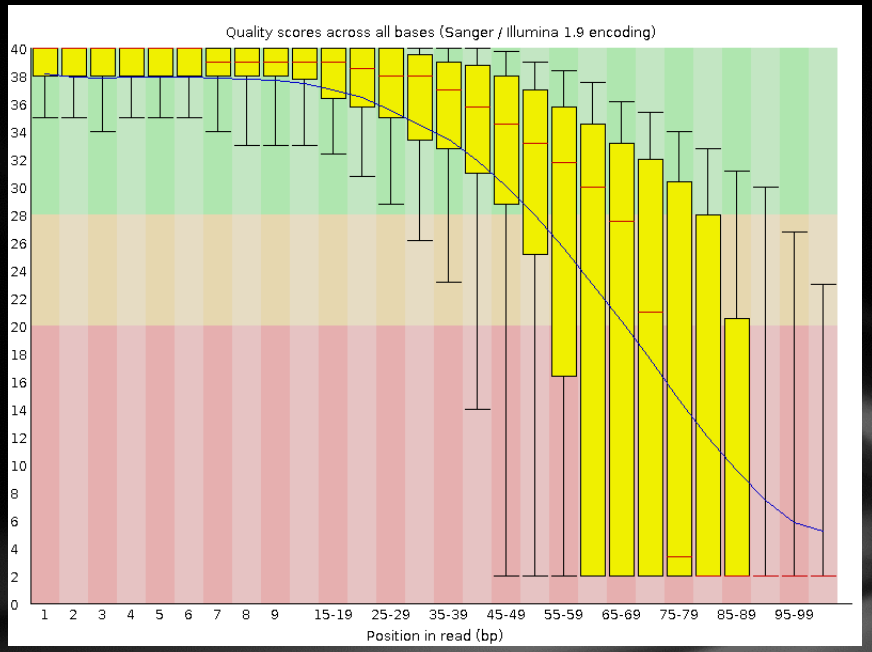
1. Mtb has a circular genome
   1. 4.4 million bps
   2. Unbalanced genome (62.5% GC content)
      1. Will suffer from GC bias in the amplification step
2. WGS usually means mass processing
   1. Illumina Hiseq: needs 96 isolates
   2. Slow growth rate can become a big problem

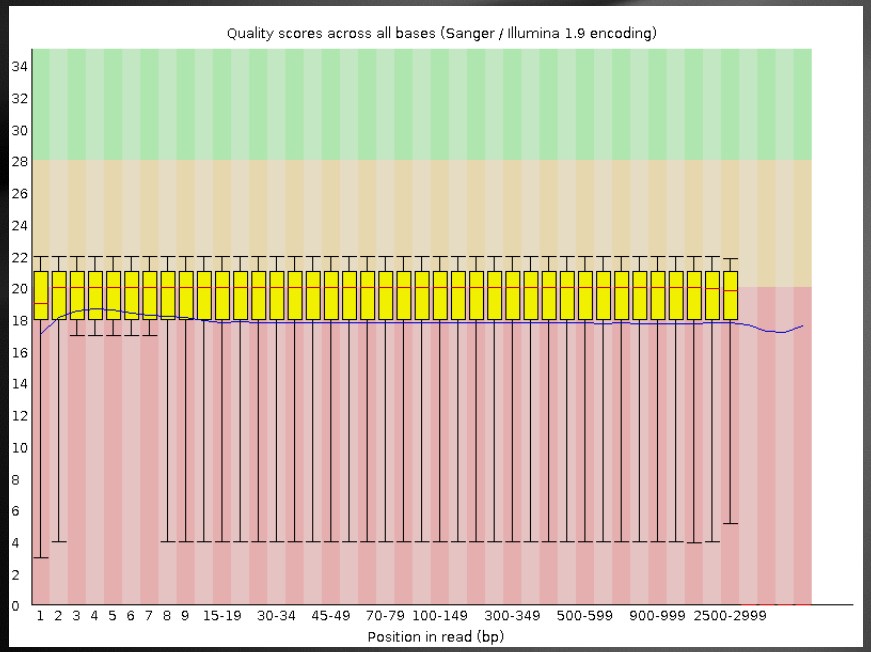


# GCDD's Approach to WGS

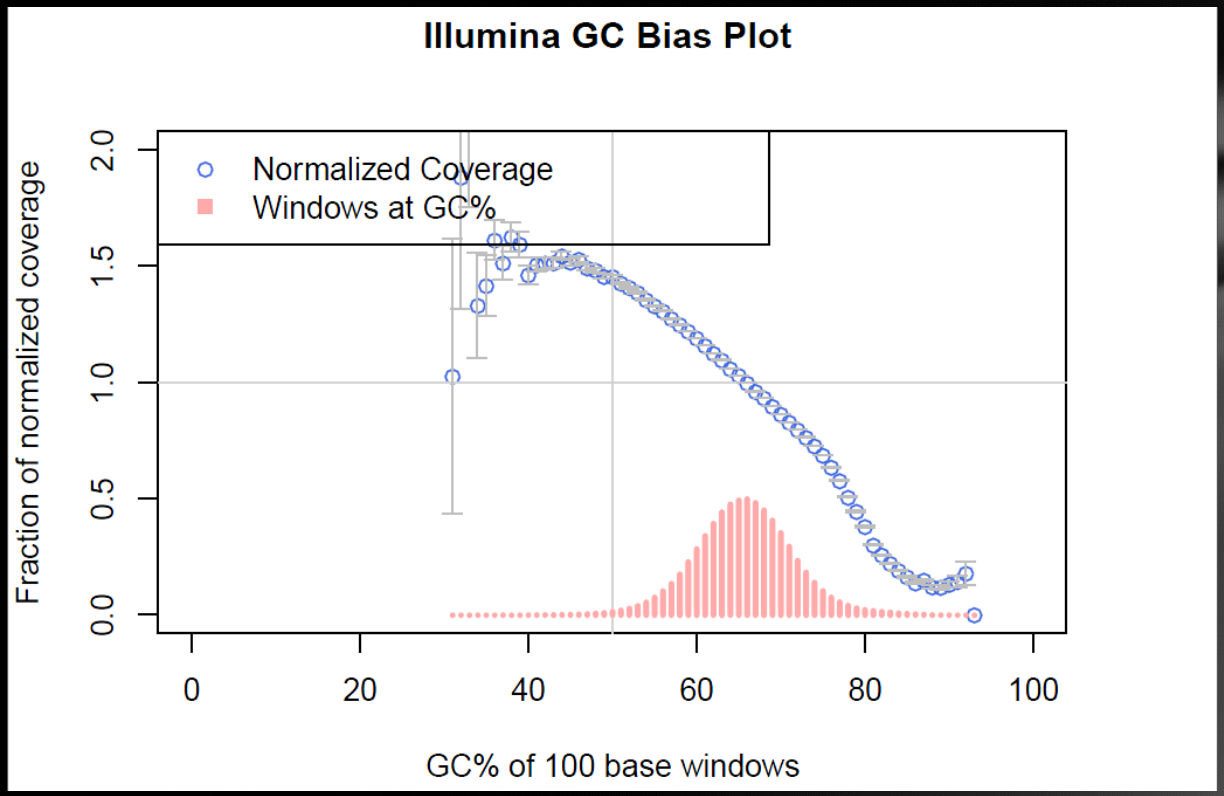
## Process:

1. Platform: Pacific Biosciences RS system
   1. Very long reads
      1. Easy de novo assembly
      2. Easy mapping to a reference genome
   2. No base quality score drop at the end of the read
      1. Needs much lower sequencing depth
   3. No systematic bias
      1. No GC Bias
   4. Can sequence one isolate at a time.
      1. Important for its utility in diagnostics

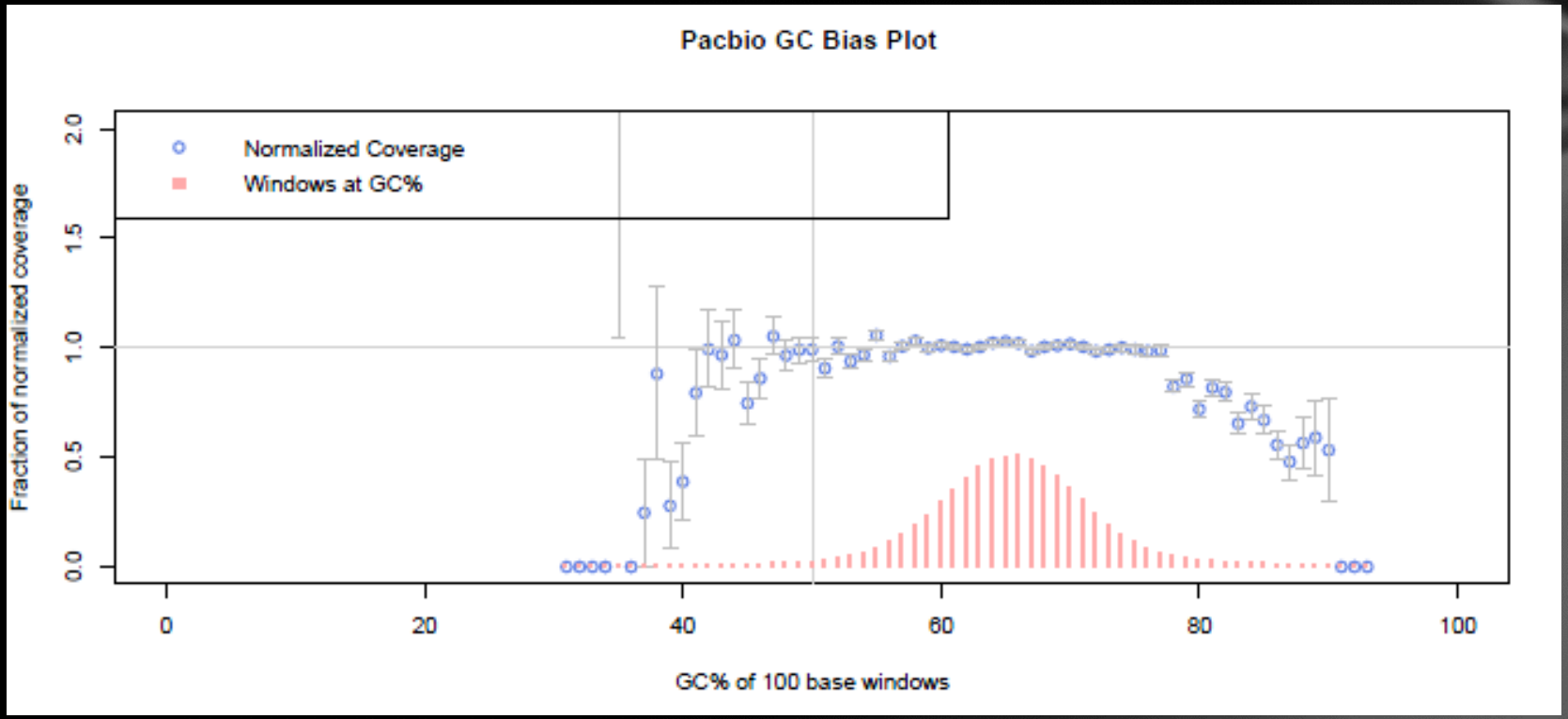


# Illumina's GC Bias Affecting Coverage Depth



# PacBio Coverage Depth



# GCDD's Approach to WGS

## What has been done:

1. Developed an in-house bioinformatic pipeline (PacDAP) for base and variant calling.
    1. Reliable base calling: PacDAP registers an uniform error rate of 50 in PHRED scale in base calling.
        1. Higher than all other sequencing platforms including Sanger, and
        2. Significantly higher than industry standard (score of 30)
    2. Can identify SNPs on a genomic scale
    3. Have identified 28,963 unstable loci in the genome

# GCDD's Approach to WGS

## What has been done:

4. Rapidly detects all major mutations associated with resistance to seven drugs:
   1. First line: Rifampicin, Isoniazid
   2. Three aminoglycosides: Amikacin, Kanamycin, and Capreomycin
   3. Two fluoroquinolones: Moxifloxacin and ofloxacin

2. 366 isolates from four countries have been sequenced:
   1. India, Moldova, the Philippines, South Africa
   2. Have identified 28,963 unstable loci in the genome

# GCDD's WGS Results, continued

## Results of PacDAP analysis:

1. Novel Markers of Resistance:

| Table 7. Number of the unannotated SNPs associated with drug resistance | |
|---|---|
| **Phenotype** | Novel SNPs |
| **$INH^R$** | 37 |
| **$RIF^R$** | 10 |
| **$MOX^R$/$OFX^R$** | 42 |
| **$AMK^R$** | 35 |
| **$CAP^R$** | 31 |
| **$KAN^R$** | 29 |

# GCDD's WGS Results, continued

## Results of PacDAP analysis:

2. Lineage: Have identified mutations on a genomic scale for determining lineage at a much finer scale than MIRU/Spoligo Typing
3. Outbreak boundaries: Identified specific markers for two outbreaks in India
4. Compensatory Markers
5. Precursor Markers
6. Host Specific Markers
7. Evolutionary Markers
8. Molecular Clock loci

**Table 8. Minimum number of SNPs identified in GCDD archive isolates in each category.**

| SNP Category | Count |
|---|---|
| Strain Specific | 24 |
| Outbreak Specific | 8 |
| Host Specific | 78 |
| Evolutionary Path Specific | 3 |
| Evolutionary State Specific | 11 |
| Compensatory | 3 |
| Precursor | 2 |

# Acknowledgements


*GCDD*

- Dr. Antonino Catanzaro
- Dr. Timothy Rodwell
- Dr. Richard Garfein
- Dr. Donald Catanzaro
- Lynn Jackson
- Collaborators from
  - India
  - Moldova
  - The Philipinnes
  - South Africa

*BMIRC*

- Sarah Ramirez Busby
- Ashu Chawla
- Carmela Chen
- Afif Elghraoui
- Min Soo Kim
- Jessica Torres
- Victoria Zadorozhny