

# Translating Sequencing to the Public Health Laboratory

David M. Engelthaler, PhD

Associate Professor

Director, TGen North

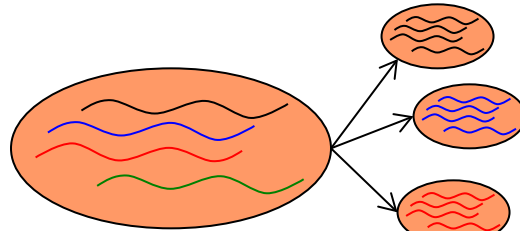
Translational Genomics Research Institute



# PH & Clinical Microbiology – Pre-Next Gen

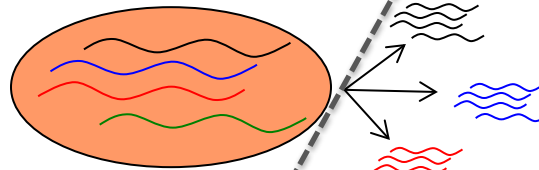
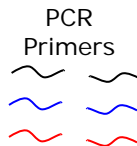
## DIRECT SAMPLE ANALYSIS

Direct culture of mixed population



Isolated organisms speciated by morph, biochem, molecular, mass spec

PCR Detection in mixed population



Highly sensitive, specific, quantitative analysis

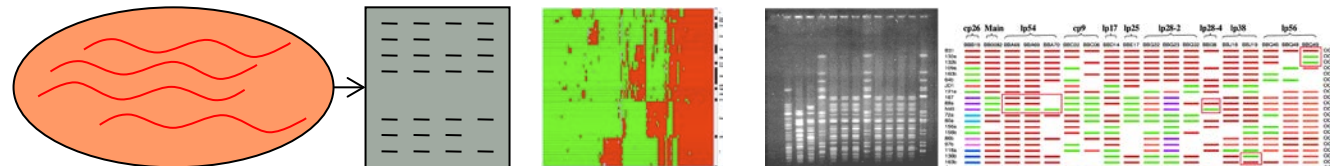
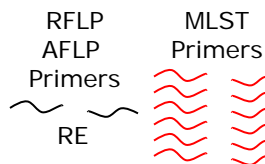
## ISOLATE ANALYSIS

Phenotype Analysis

Antibiotic Resistance



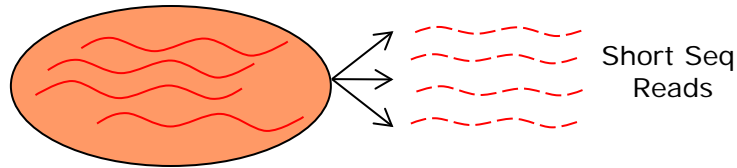
Genotype Analysis



# PH & Clinical Microbiology – Next Gen Era

## ISOLATE ANALYSIS

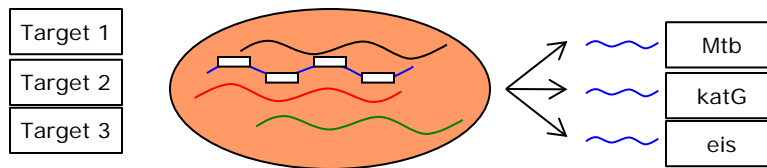
Whole Genome Sequencing



Genotyping/Phylogenetics  
Ab res, Virulence  
Expression

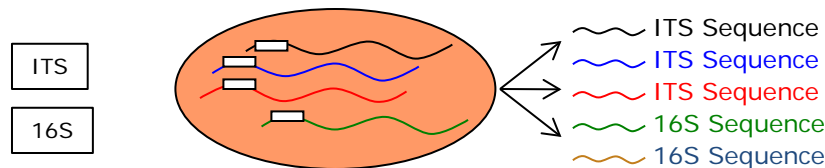
## DIRECT SAMPLE ANALYSIS

Amplicon Sequencing



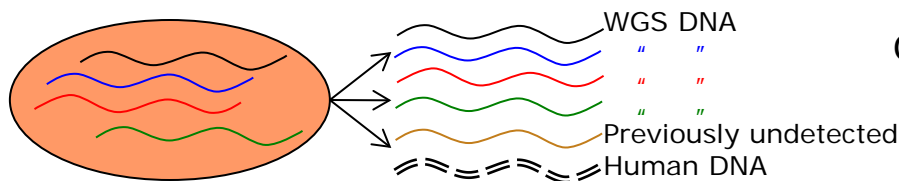
Highly Multiplexed sequencing of target amplicons

Mycobiome/  
Microbiome Profiling



NextGen sequencing of amplified ITS/16S

Whole Sample Sequencing

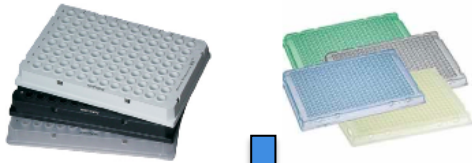


Detection  
Genotyping/Phylogenetics  
Metagenomics  
Ab res, Virulence  
Expression  
Host genome

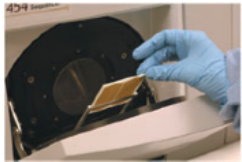
# Next Gen Sequencing

## Capillary electrophoresis (Sanger)

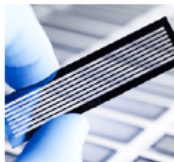
Between 96 and 384 samples  
(76 - 308 Kb/run)



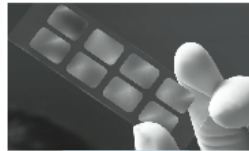
**454**  
400,000 samples  
(120Mb / run)



**Solexa**  
32M samples  
(2Gb / run)



**Solid**  
40M samples  
(2-6 Gb / run)



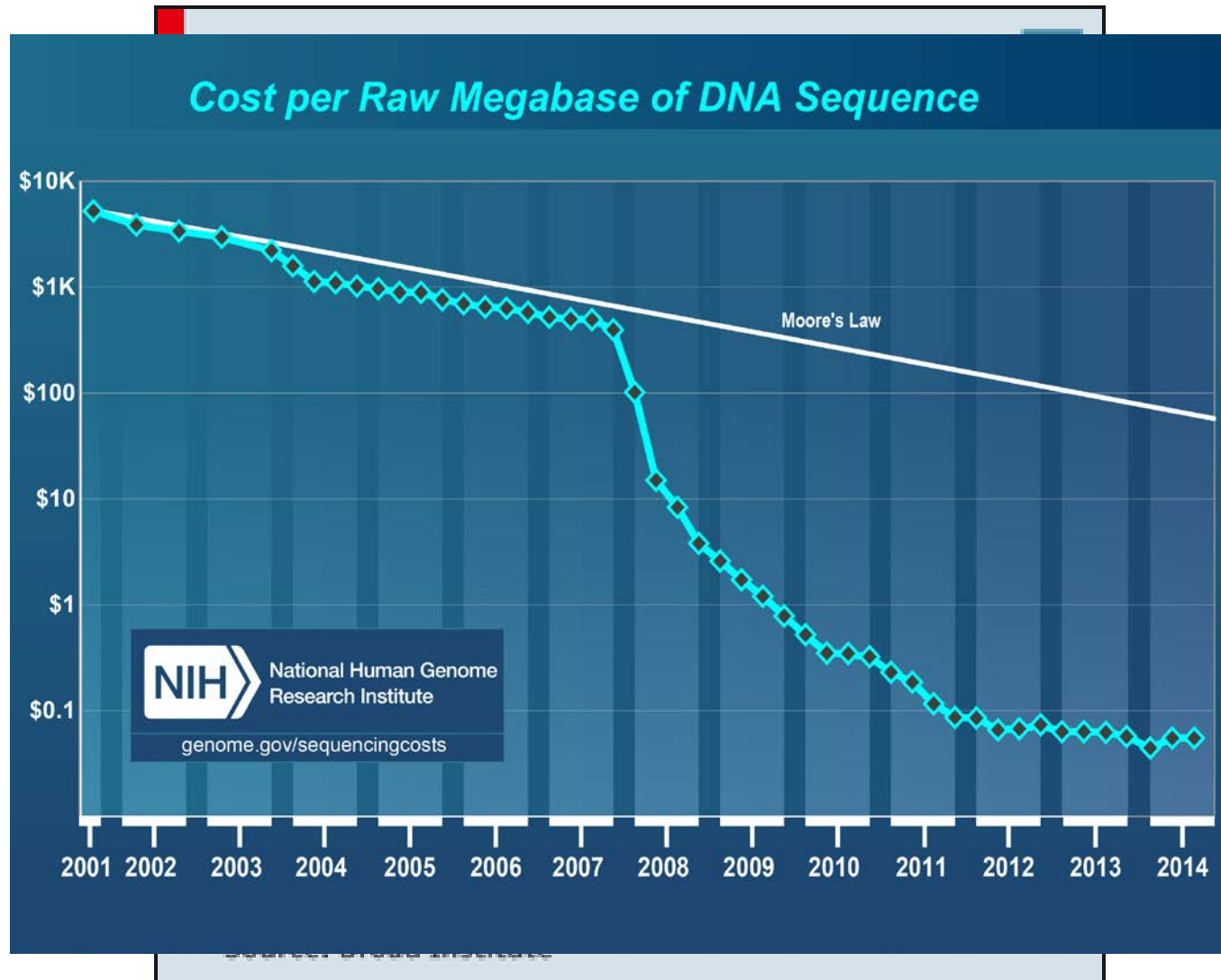
4GB – 20 hrs



1GB – 4.5hrs

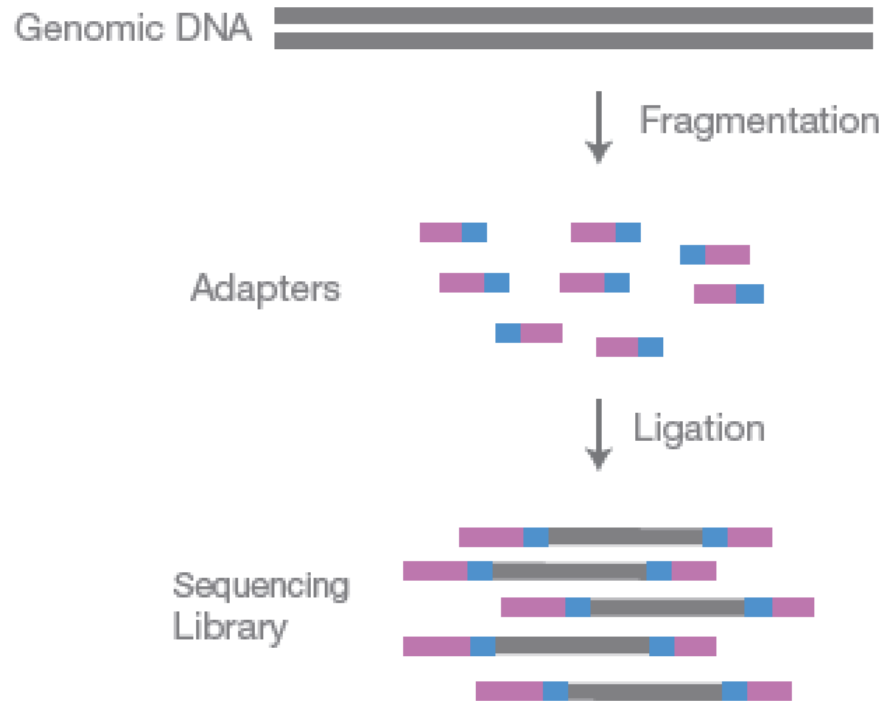


# Moore's Law vs. Carlson's Curve



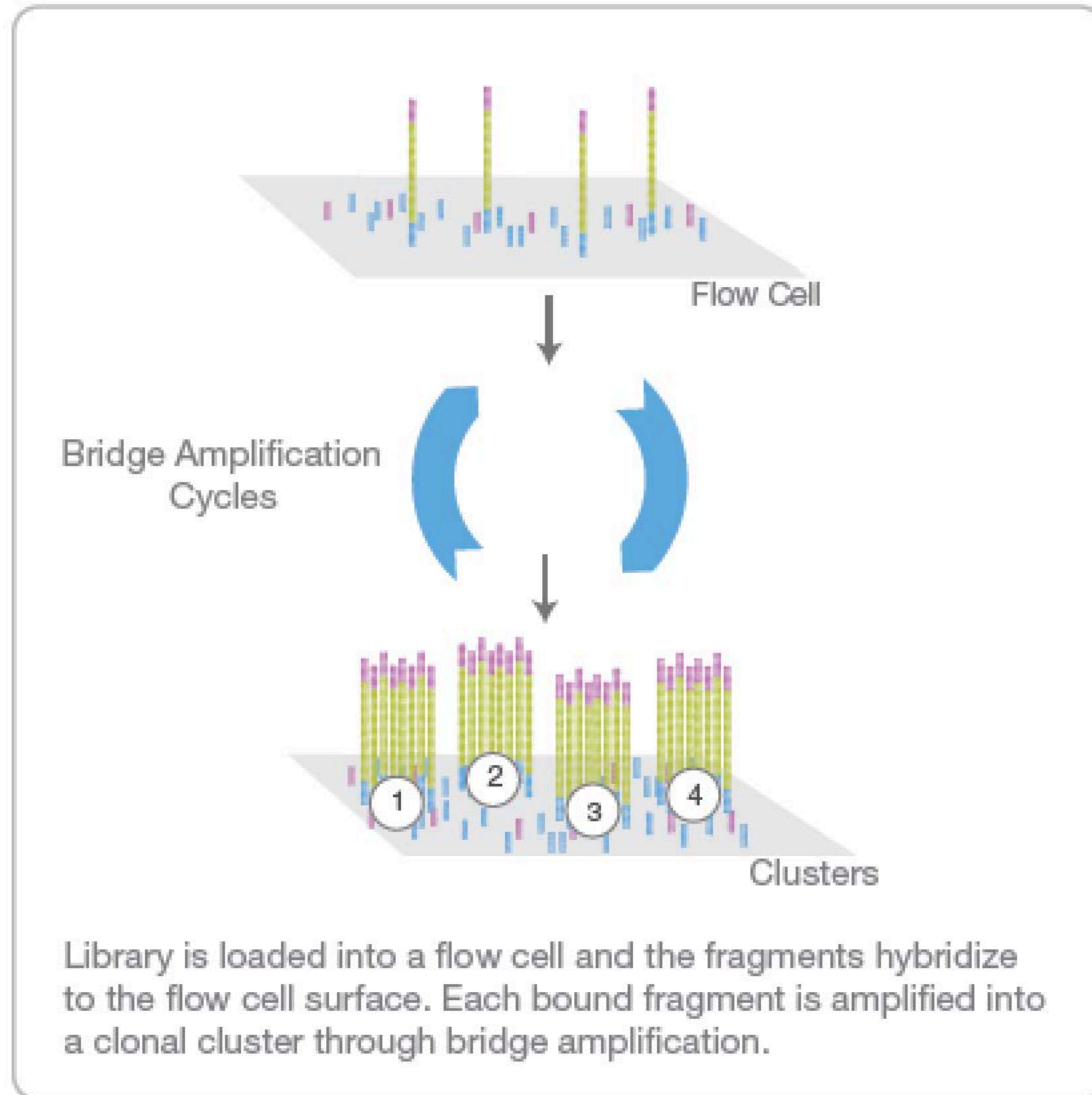
# Sequence Process

## Library Prep



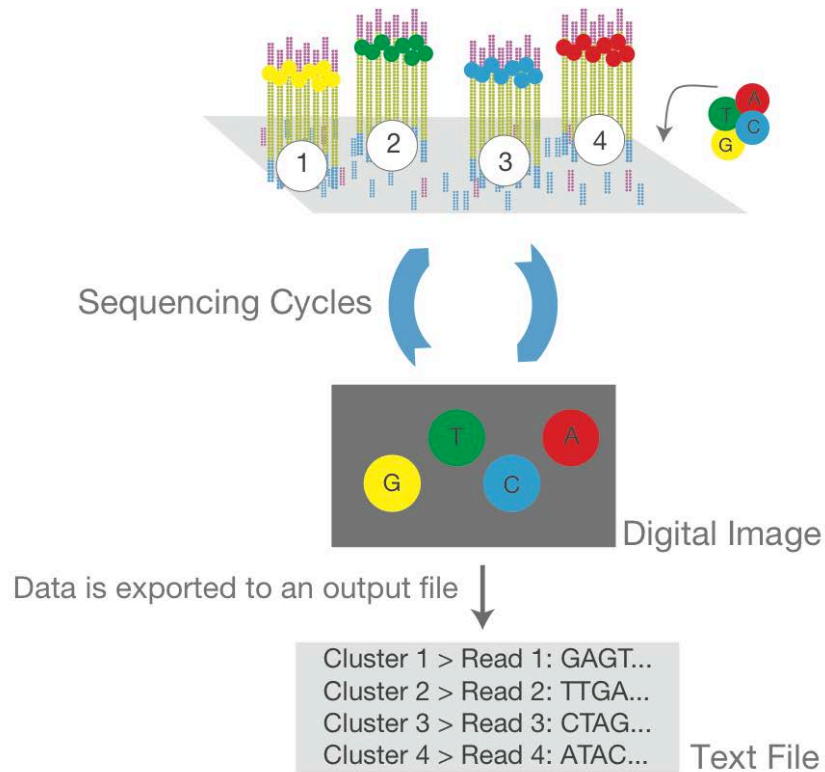
NGS library is prepared by fragmenting a gDNA sample and ligating specialized adapters to both fragment ends.

# Sequence Process Amplification



# Sequence Process

## Sequencing

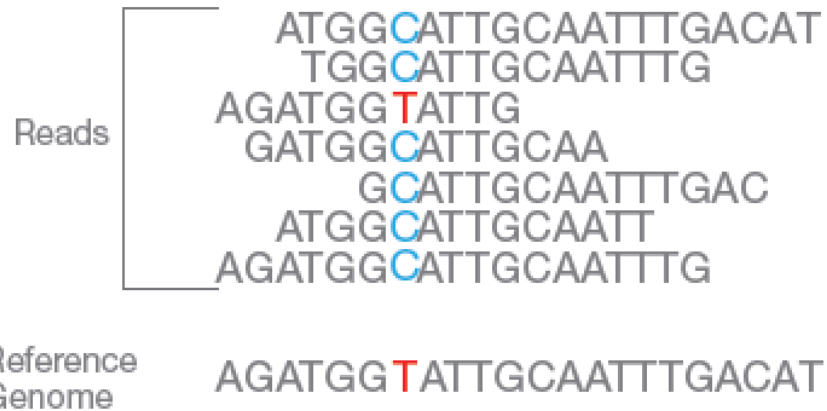


Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated “n” times to create a read length of “n” bases.



# Sequence Process

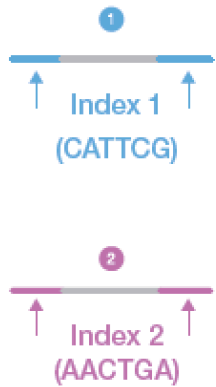
## Data Analysis



Reads are aligned to a reference sequence with bioinformatics software. After alignment, differences between the reference genome and the newly sequenced reads can be identified.

# Multiplexing

A  
Library Preparation



B  
Pool



C  
Sequence



Sequence Output  
to Data File

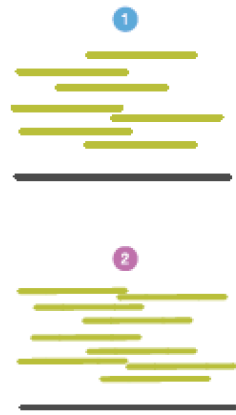
```
CATT CGACGGATCG  
AACTGAGTCCGATA  
AACTGATCGGATCC  
CATT CGTGGCAGTC  
AACTGAACCTGATG  
AACTGAGATTACAA  
CATT CGCAGTTCATT  
CATT CGAACTTCGA
```

D  
Demultiplex

Diagram D shows the demultiplexed reads, separated by index. Index 1 reads are in blue and yellow, and Index 2 reads are in purple and yellow.

```
1  
CATT CGACGGATCG  
CATT CGTGGCAGTC  
CATT CGCAGTTCATT  
CATT CGAACTTCGA  
  
2  
AACTGAGTCCGATA  
AACTGATCGGATCC  
AACTGAACCTGATG  
AACTGAGATTACAA
```

E  
Align



- Library 1 Barcode
- Library 2 Barcode
- Sequencing Reads
- DNA Fragments
- Reference Genome

# WGS Bioinformatic Toolkit

## SNP Callers

SolSNP  
Varscan

## Assemblers

Velvet  
oases  
phrap  
SSPACE  
edena v3  
Amos  
MIRA

## Next-Gen Aligners

bwa  
STAMPY  
Novoalign  
NGen

## Whole Genome Aligners

Mauve  
MUMmer  
Mugsy

## Phylogenetic Software

circos  
arb  
mothur  
FigTree  
MetaTree  
FastTree  
RaxML  
PAUP 4.0b10  
MEGA

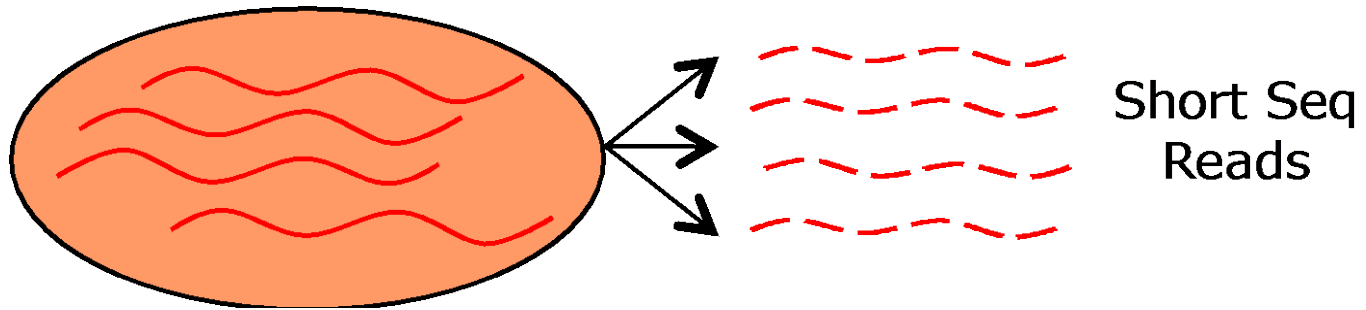
## Next-Gen Misc Tools

sickle  
htseq-count  
DESeq  
Qiime  
GATK  
Bedtools  
FASTX Toolkit  
Picard  
IGV  
SAMTools  
CLC Genomic Workbench

## Whole-Genome Misc Tools

glimmer  
transeq  
Jalview  
MPI-BLAST

# Whole Genome Analysis



Disease Outbreaks

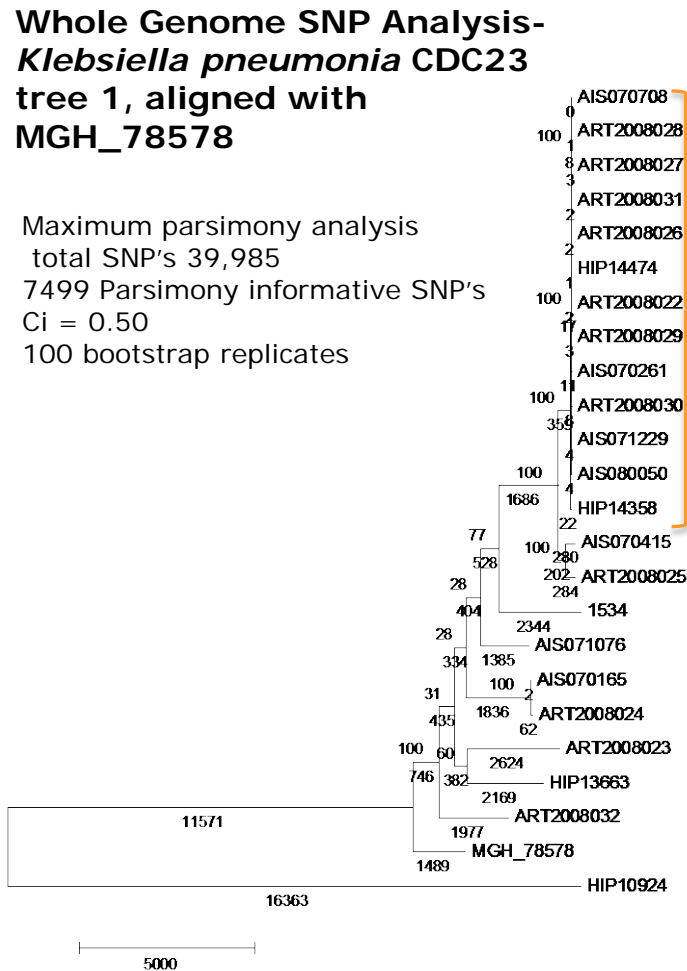
Emerging Infectious Diseases

**Molecular Epi**  
Genotyping and  
Disease Outbreaks

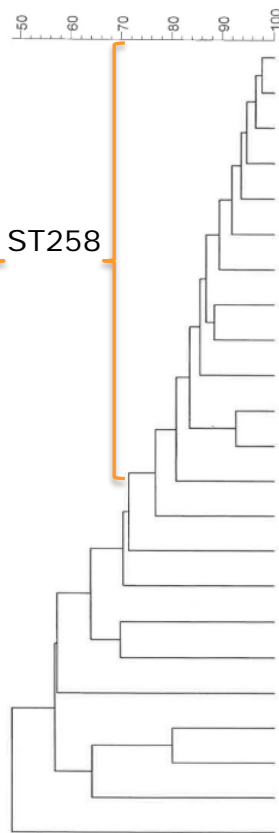
# Genotyping, Pop Gen and *Klebsiella*

## Whole Genome SNP Analysis- *Klebsiella pneumoniae* CDC23 tree 1, aligned with MGH\_78578

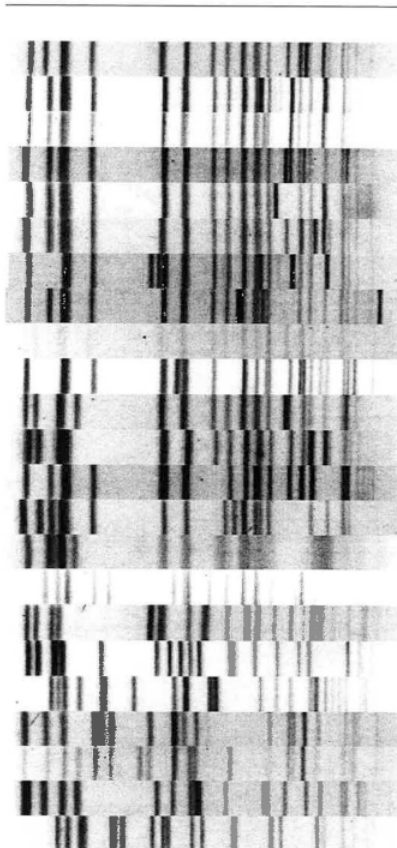
Maximum parsimony analysis  
total SNP's 39,985  
7499 Parsimony informative SNP's  
Ci = 0.50  
100 bootstrap replicates



Dice (Tol 1.5%-1.5%) (H>0.0% S>0.0%) [0.0%-100.0%]  
PFGE-Xbal-H9812



PFGE-Xbal-H9812



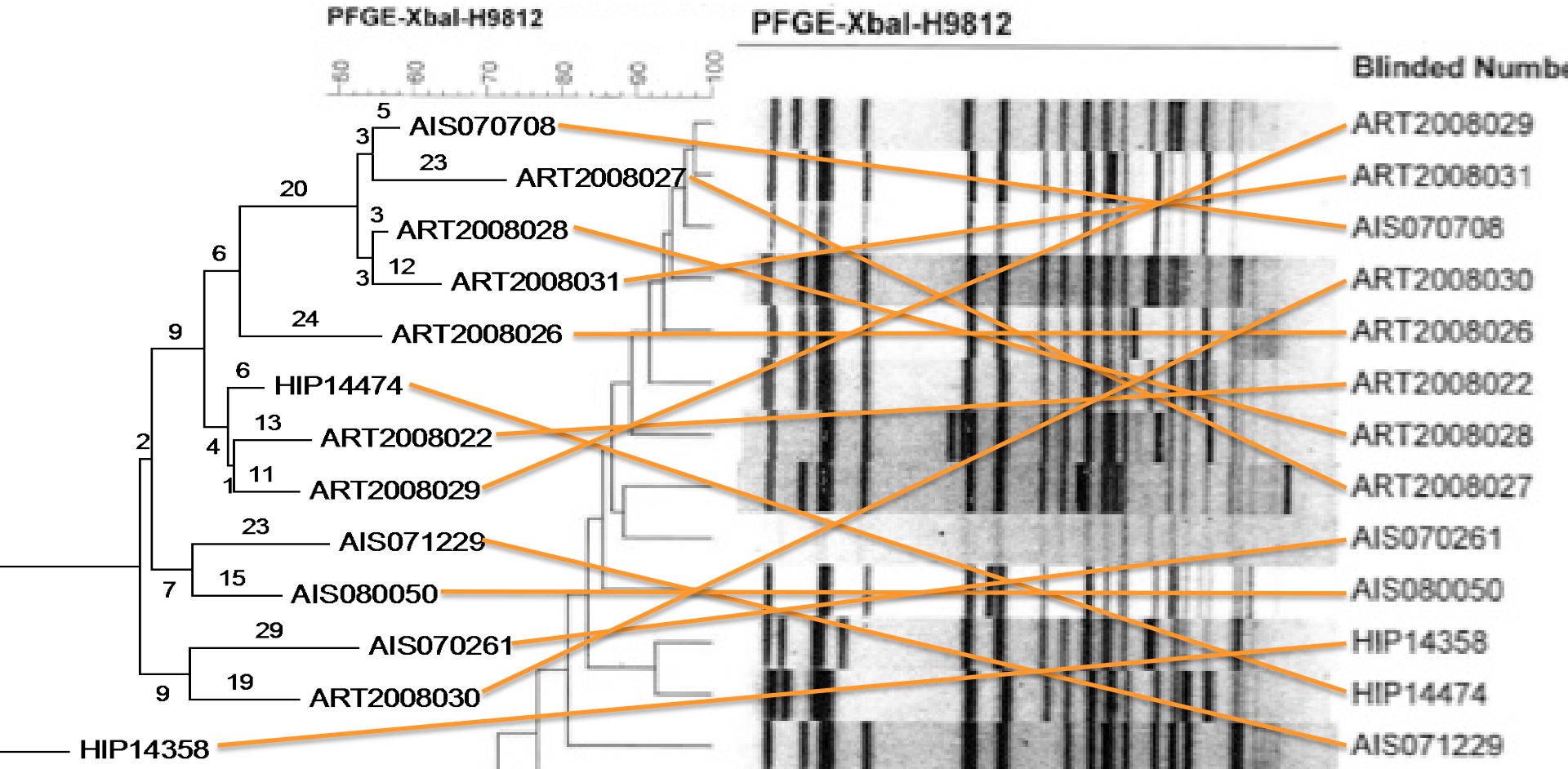
Blinded Number	MLST	Year	State	City
ART2008029	258	2006	DE	Syrma
ART2008031	258	2007	CO	Greeley
AIS070708	258	2007	NM	Albuquerque
ART2008030	258	2006	GA	Atlanta
ART2008026	258	2006	NJ	Livingston
ART2008022	258	2005	PA	Hershey
ART2008028	258	2006	AZ	Phoenix
ART2008027	258	2006	AZ	Phoenix
AIS070261	258	2007	Israel	Tel Aviv
AIS080050	258	2008	IL	Chicago
HIP14358	258	2005	NY	NY
HIP14474	258	2005	PA	Philadelphia
AIS071229	258	2007	MA	Brighton
ART2008025	21	2006	MD	Jefferson City
AIS070415	259	2007	CA	Richmond
1534	37	1996	NC	Durham
HIP13663	234	2004	NY	NY
ART2008032	45	2007	VA	Charlottesville
AIS071076	101	2007	India	Vellatore
ART2008024	14	2006	MO	St. Louis
AIS070165	14	2008	MI	Grand Rapids
ART2008023	228	2005	MD	Baltimore
HIP10924	257	2001	PA	Pittsburgh

Kitchell et al. AAC 2009

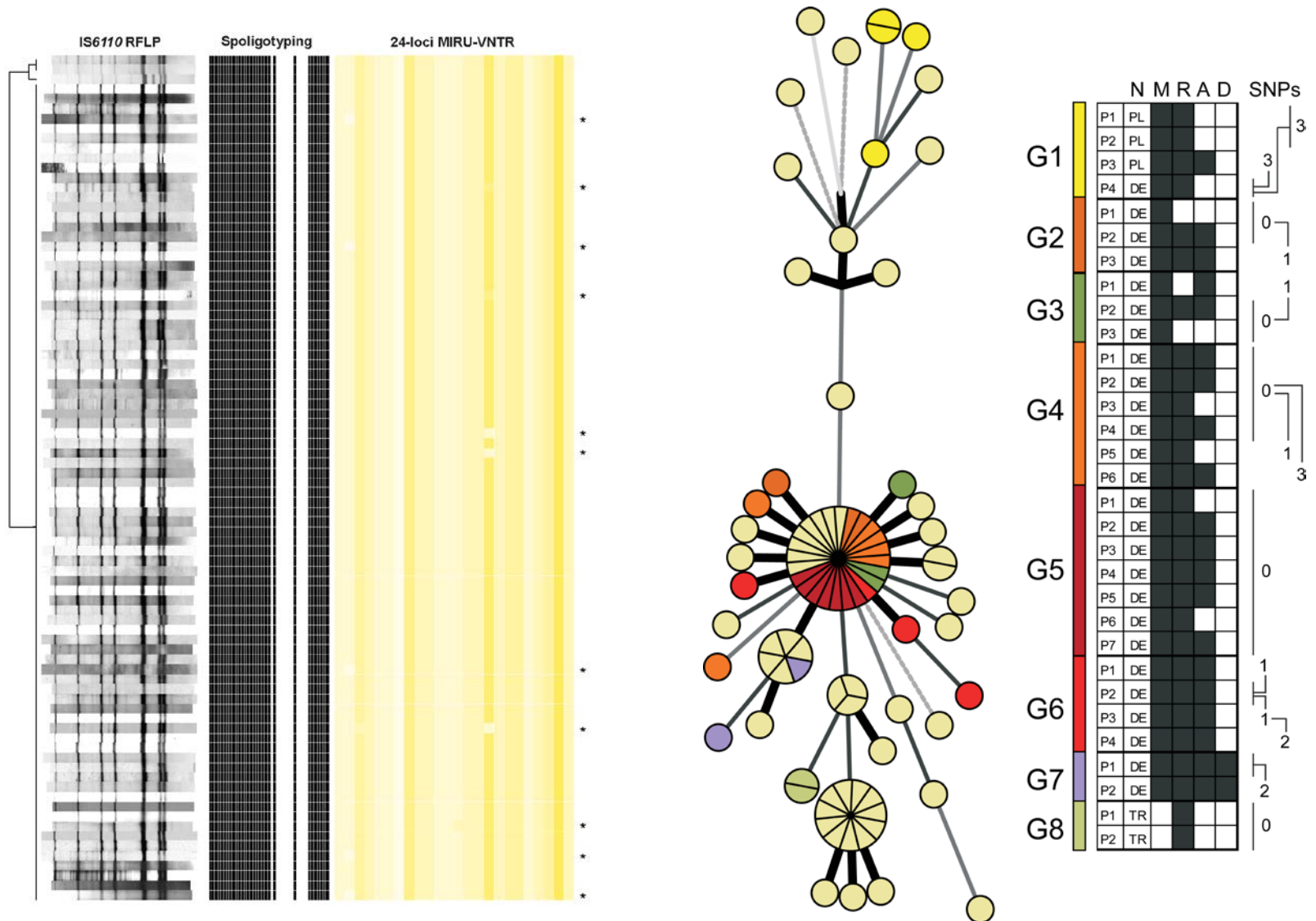
Bowers et al. 2012 (unpublished)

# WGST vs PFGE

## Klebsiella pneumoniae ST258

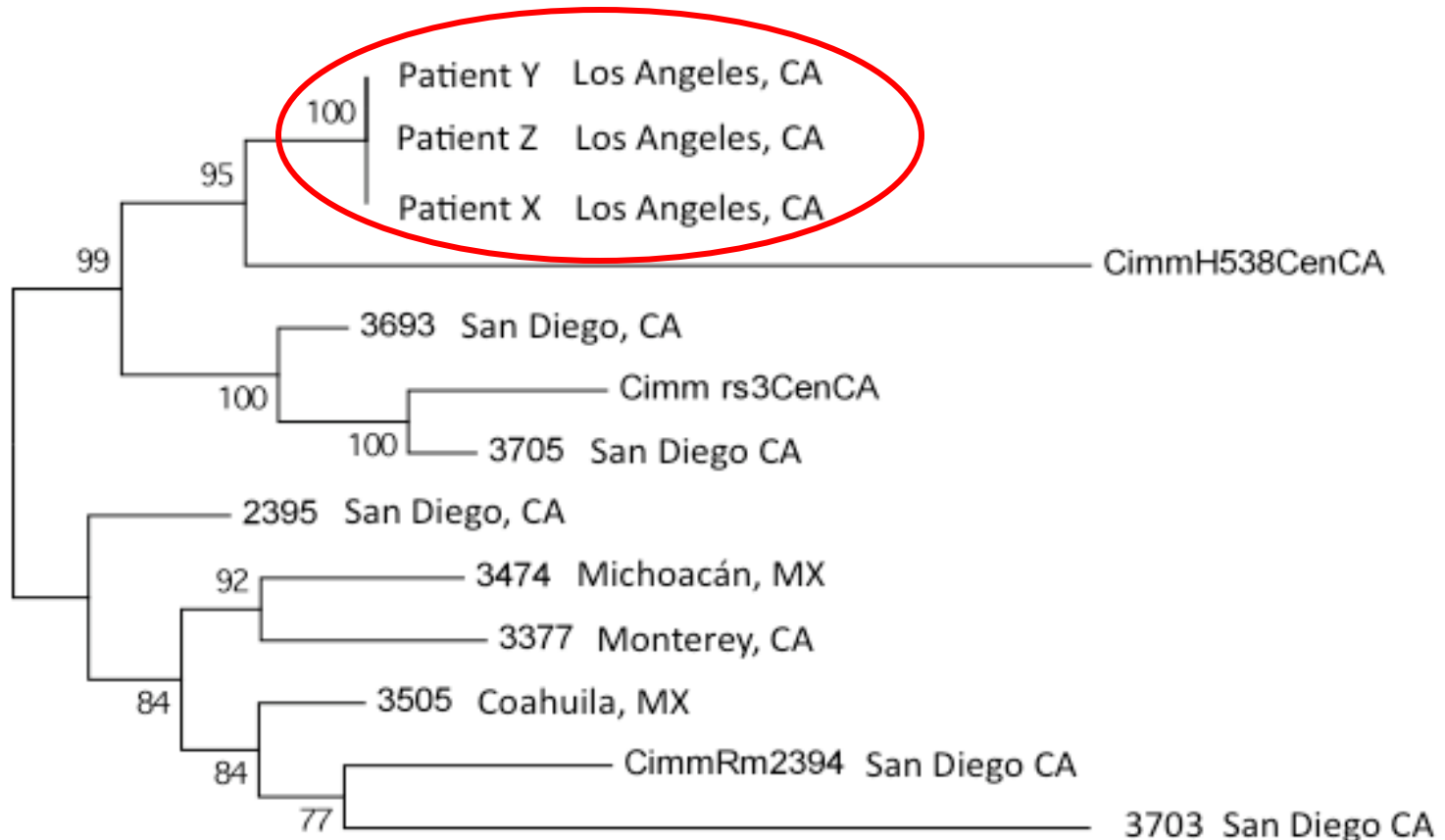


# *Mtb* Molecular Epi – Contact Tracing





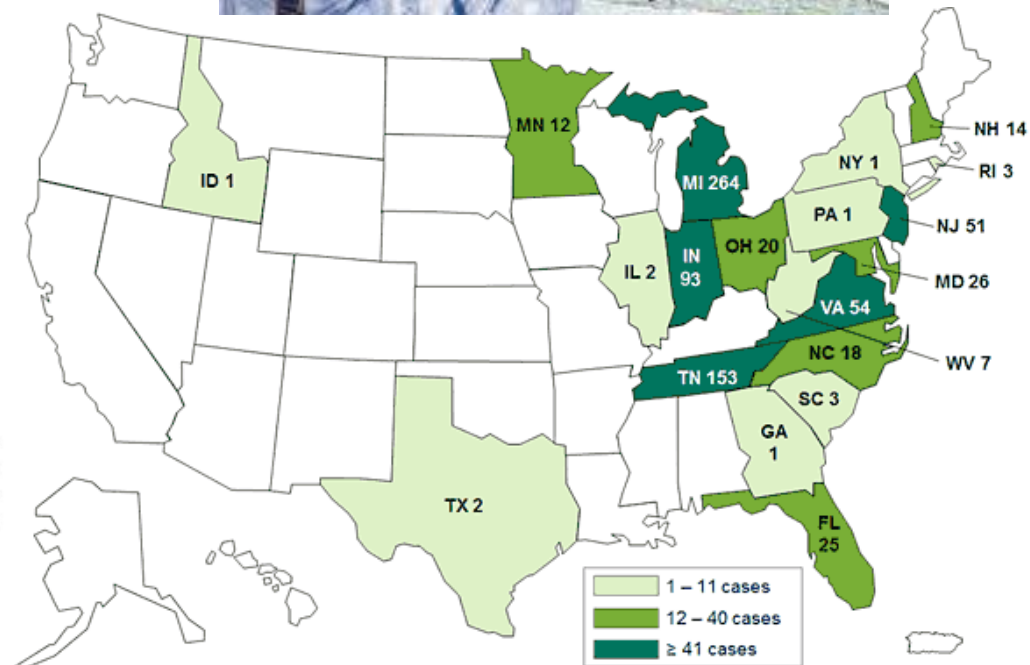
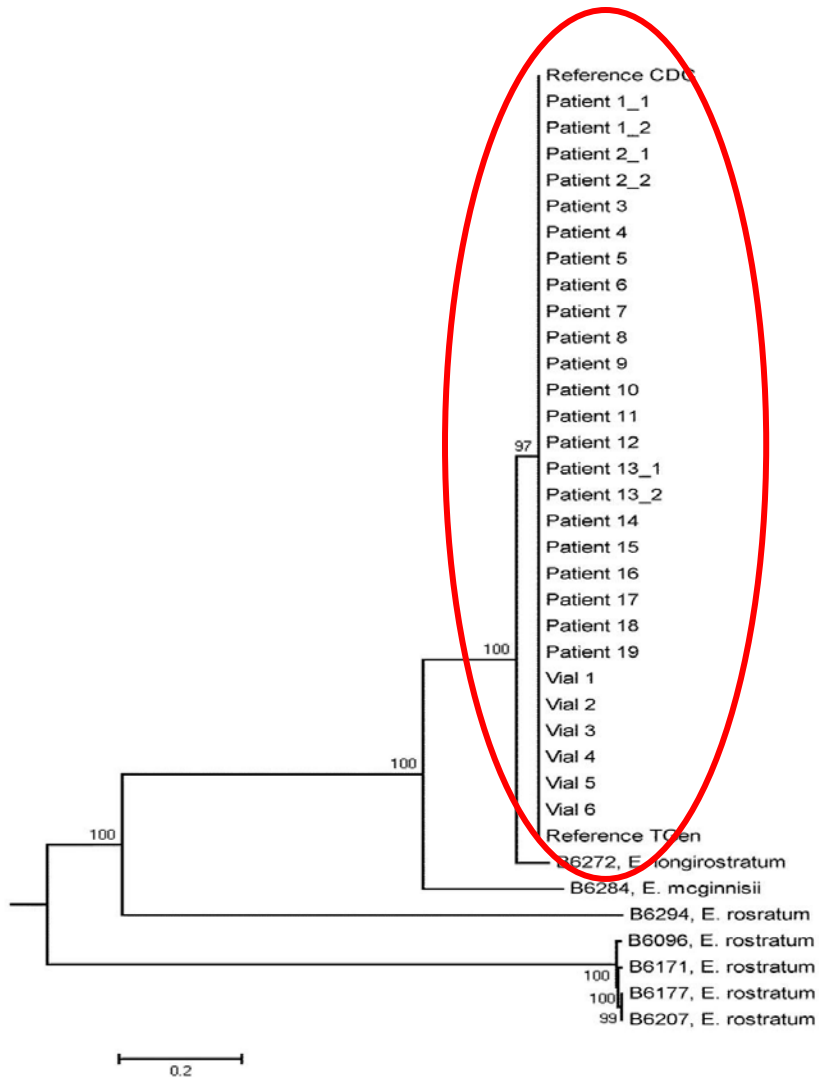
# Transplant Outbreak with *Coccidioides immitis*



1000

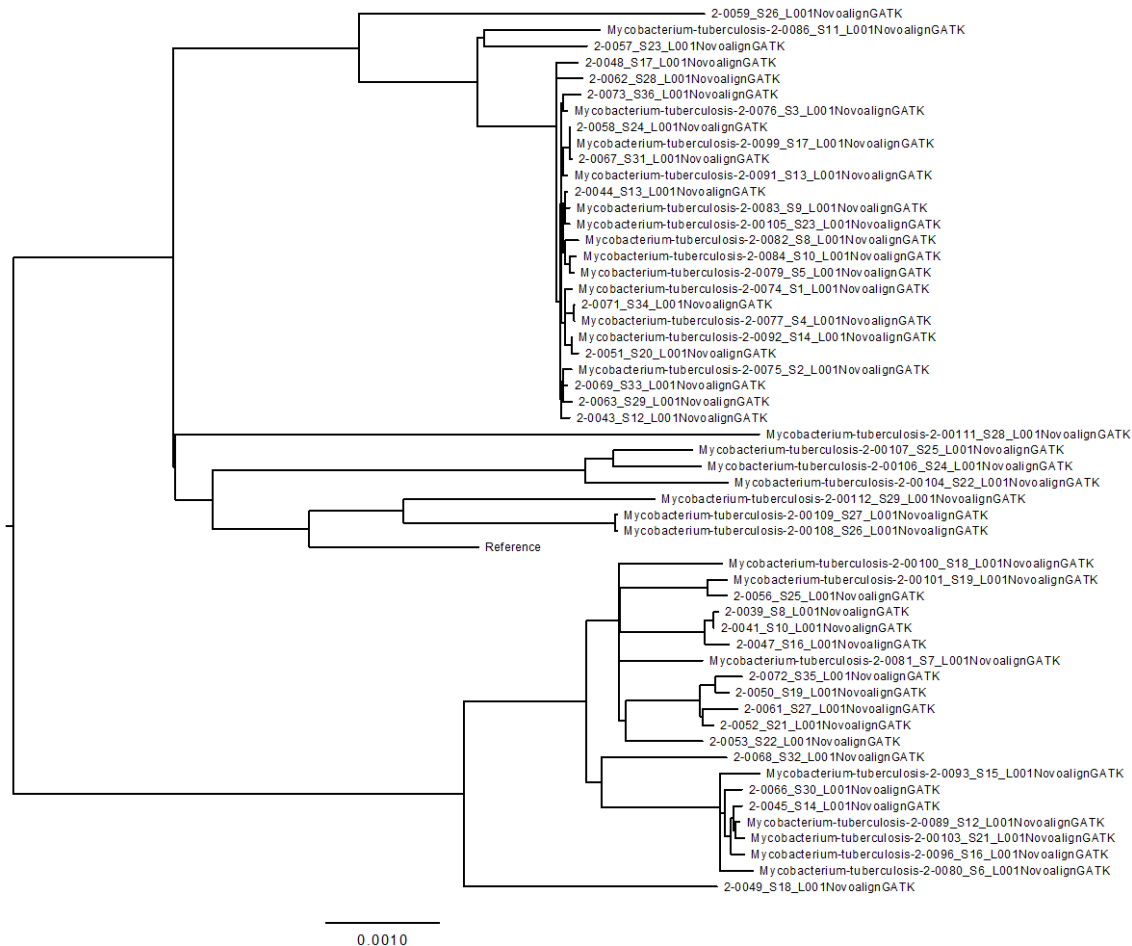
Engelthaler et al 2010

# Exserohilum meningitis from injectable steroid



# WGST and TB

## High quality SNP reference database and phylogeny



# **Molecular Epi**

## Emerging Infectious Diseases

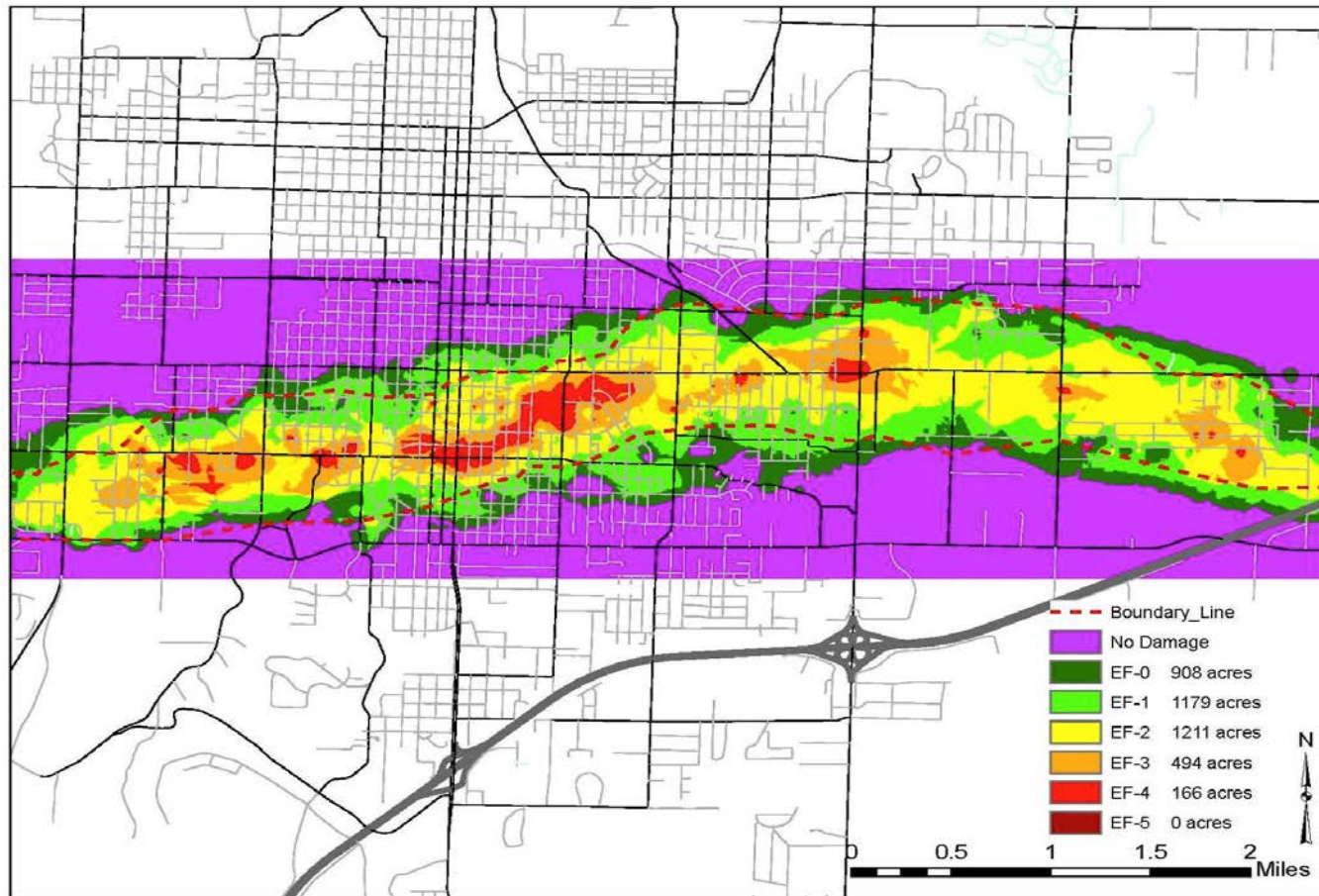
# **Apophysomyces**

Joplin, Missouri

May, 2011

# Killer Tornado

- EF5 Tornado
- 156 Killed
- 1,100 Hospitalized
- 30-50% of Joplin destroyed





# May 22, 2011

## Joplin, MO: Deep Skin Wound Fungal Infection Cases

Health and  
Senior Services  
MISSOURI DEPARTMENT

### Legend

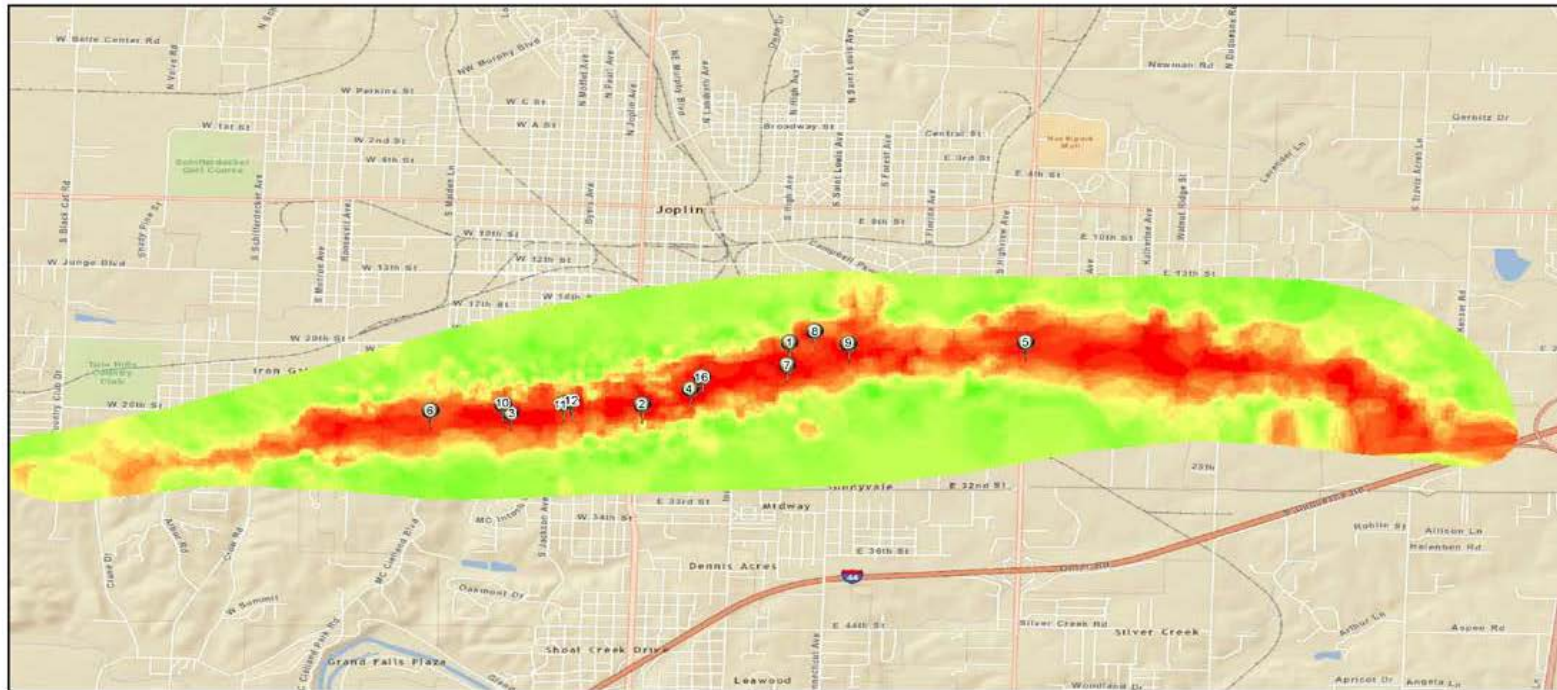
- Structure Damage
- Catastrophic Damage
  - Limited Damage
  - ICU Study

### Location Map



0 0.5 1 Miles

Data Sources: USACE, CDC, FEMA  
Missouri Department of Health and Senior Services  
CAITS/IOGI  
AMG  
September 12, 2011  
M-CERT/Joplin Tornado/Control Study





# Day 10



# Day 24



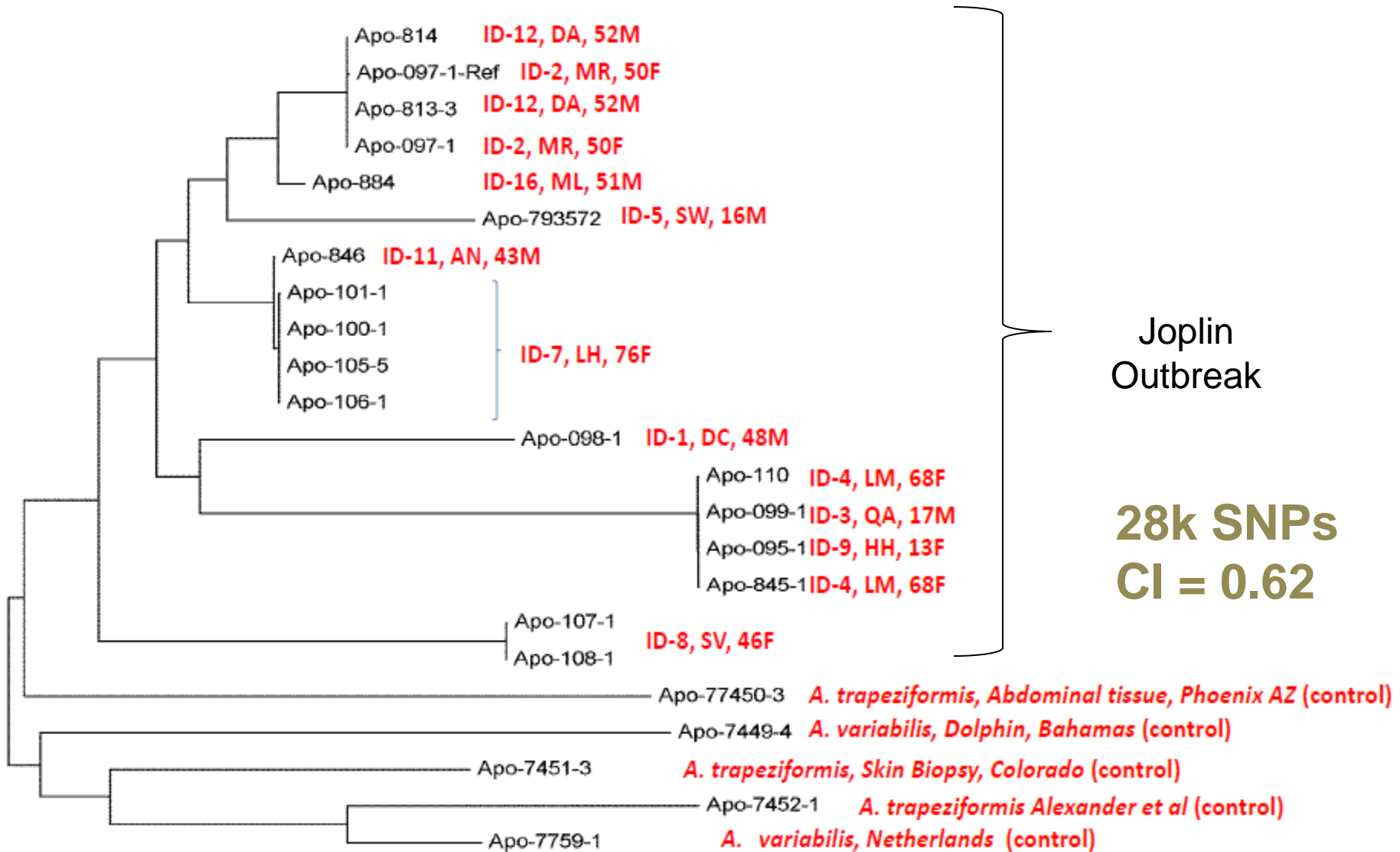
Photo Courtesy Gina Weddle, RN, MSN, CPNP-AC, The Children's Mercy Hospital

# Day 102

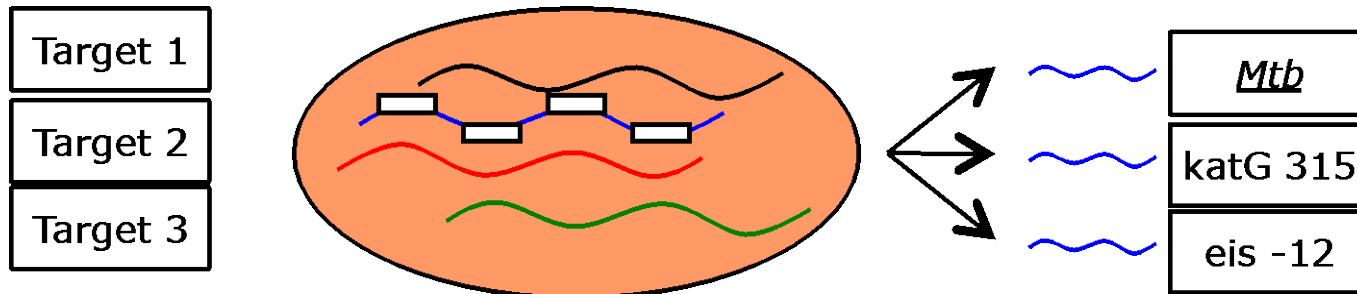


Photo Courtesy Gina Weddle, RN, MSN, CPNP-AC, The Children's Mercy Hospital

# *Apophysomyes* spp.



# Amplicon Sequencing for Drug Sensitivity Testing in *Mycobacterium tuberculosis*



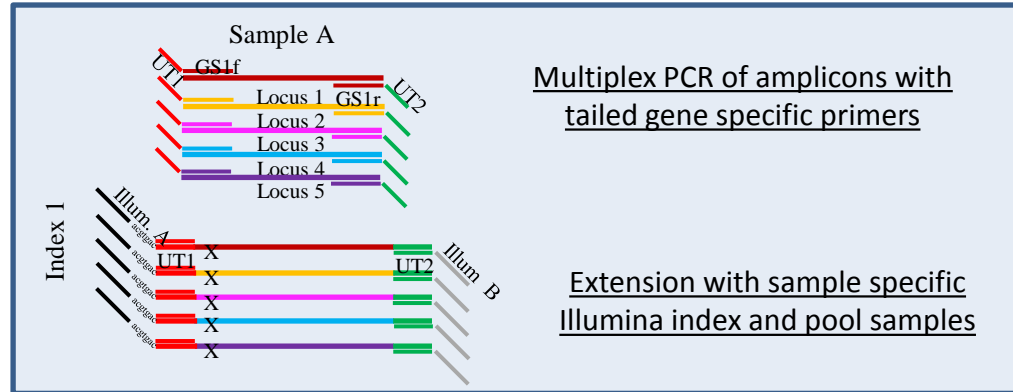
# Mtb Antibiotic Resistance

- Focusing on 6 genes
  - *katG* – Isoniazid resistance
  - *rrs* – Kanamycin/Amikacin/Capreomycin resistance
  - *inhA* promoter – Isoniazid resistance
  - *gyrA* – Moxifloxacin/Ofloxacin resistance
  - *eis* promoter – Kanamycin/Amikacin/Capreomycin resistance
  - *rpoB* – Rifampicin resistance
- Focusing on 36 mutations identified in *Rodwell et al 2013 JCM*

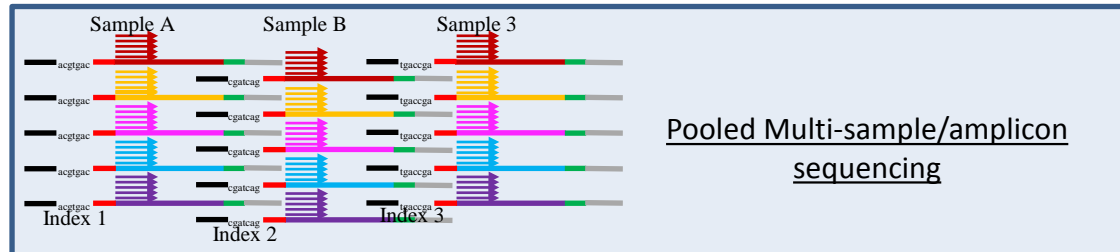


# Highly Multiplex Amplicon Sequencing

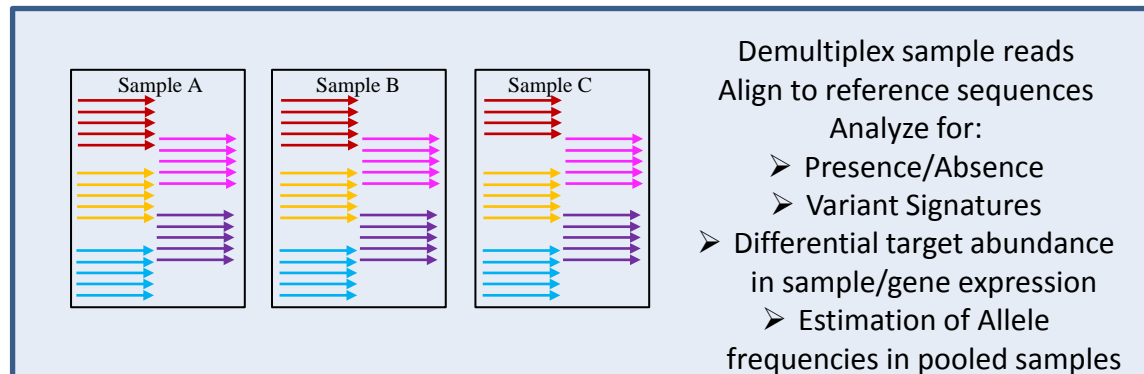
**Next-Gen Ready  
MultiLocus  
Amplification**



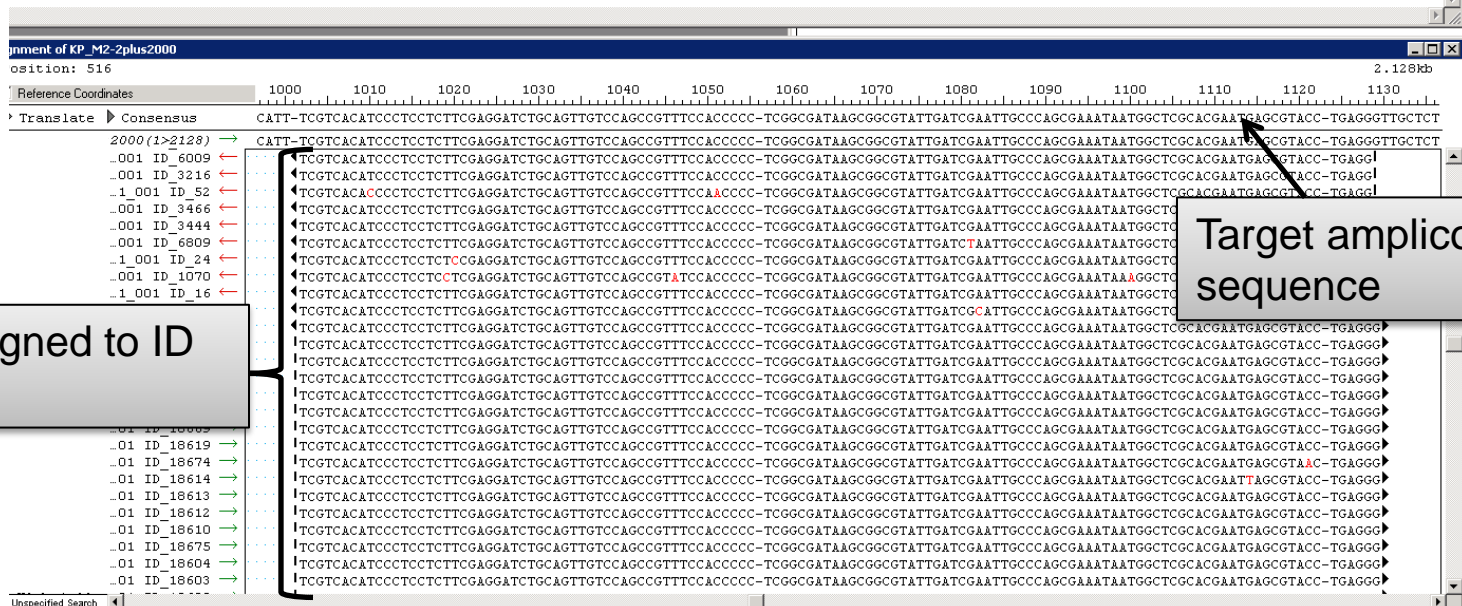
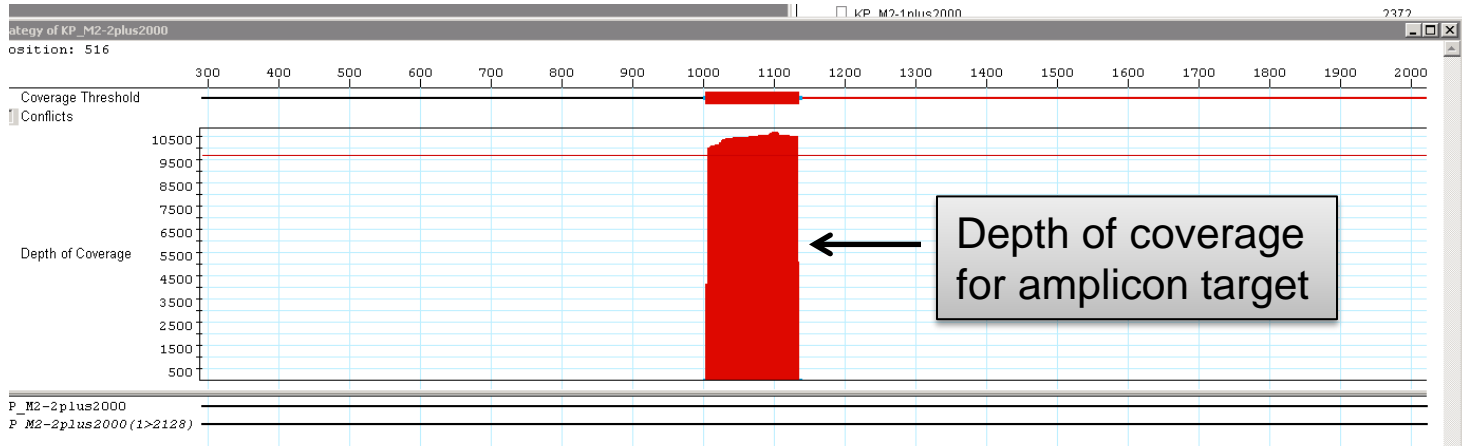
**Rapid Short Read  
Sequencing**



**Bioinformatic  
amplicon sequence  
analysis**



# Amplicon Sequencing from Clinical Sample



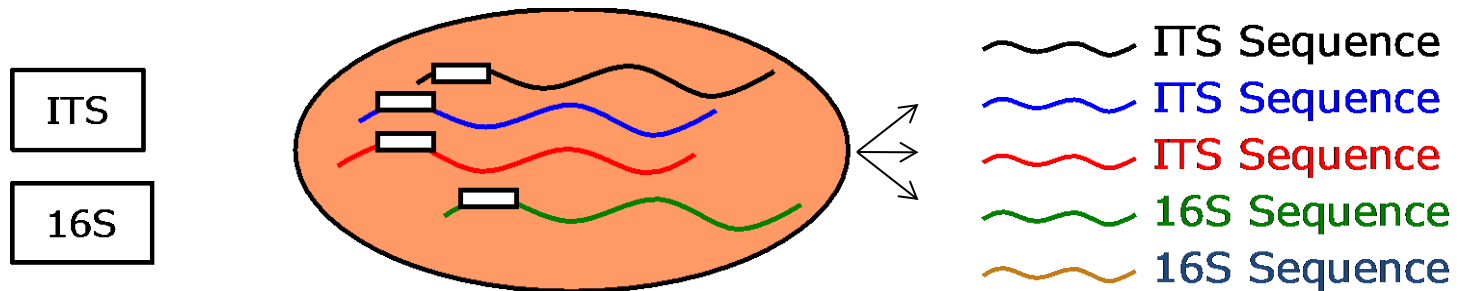


# Next Gen DST

Table. Summary of Next Gen DST results from clinical specimens

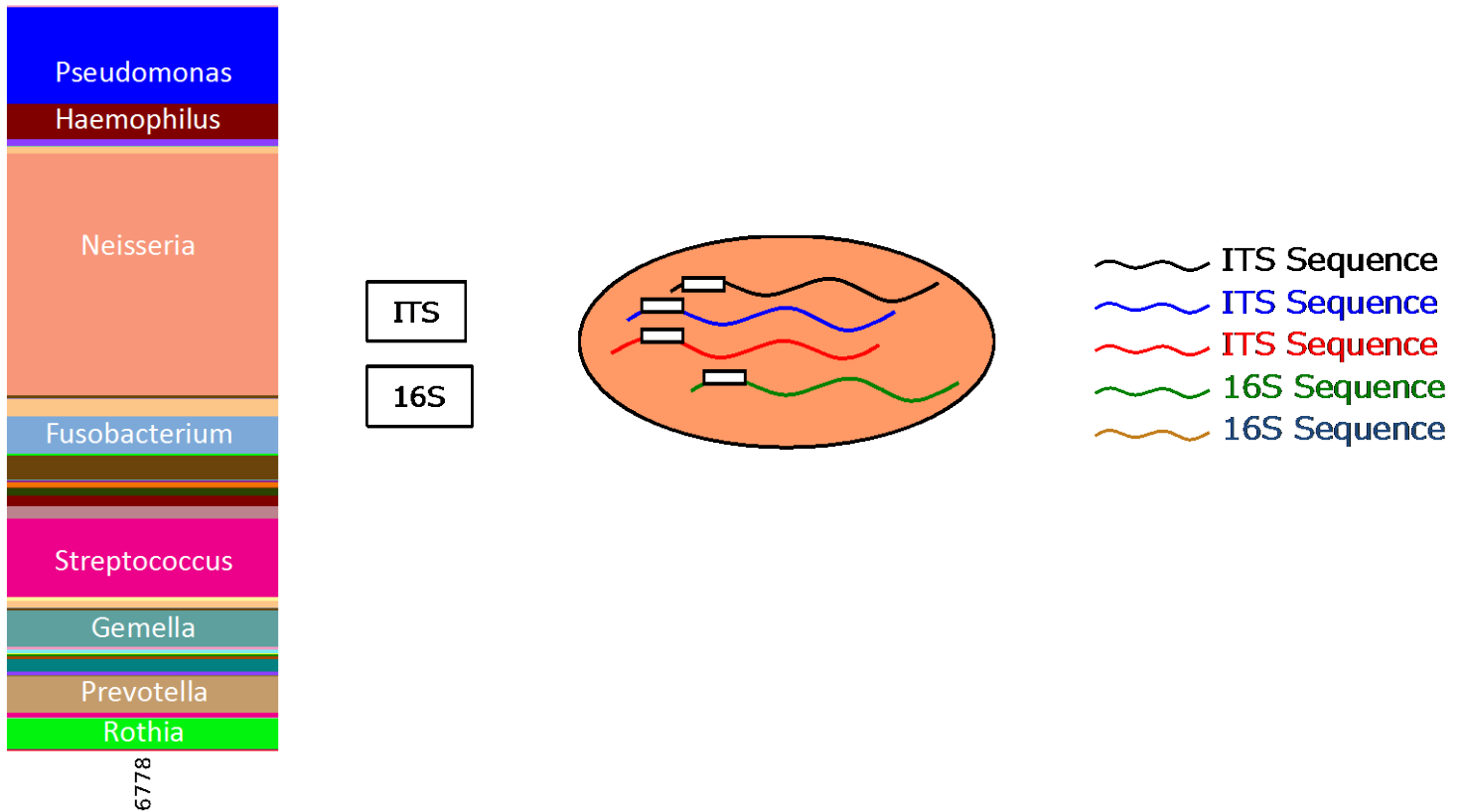
Next Gen DST (no.)	Phenotypic DST result (no.)		Accuracy analysis results (%)		
	R <sup>a</sup>	S <sup>a</sup>	Correlation	Sensitivity	Specificity
<b>INH (171)</b>					
Mutations Detected	94	0	96.5%	94.0%	100.0%
No mutations	6 <sup>b</sup>	71			
<b>RIF (172)</b>					
Mutations Detected	79	0	97.7%	95.2%	100.0%
No mutations	4	89			
<b>KAN (166)</b>					
Mutations Detected	51	5	95.8%	96.2%	95.6%
No mutations	2	108			
<b>MOX (172)</b>					
Mutations Detected	13	0	98.8%	86.7%	100.0%
No mutations	2	157			
<b>OFX (172)</b>					
Mutations Detected	12	1	98.3%	85.7%	99.4%
No mutations	2	157			

# Amplicon sequencing for Microbiomics

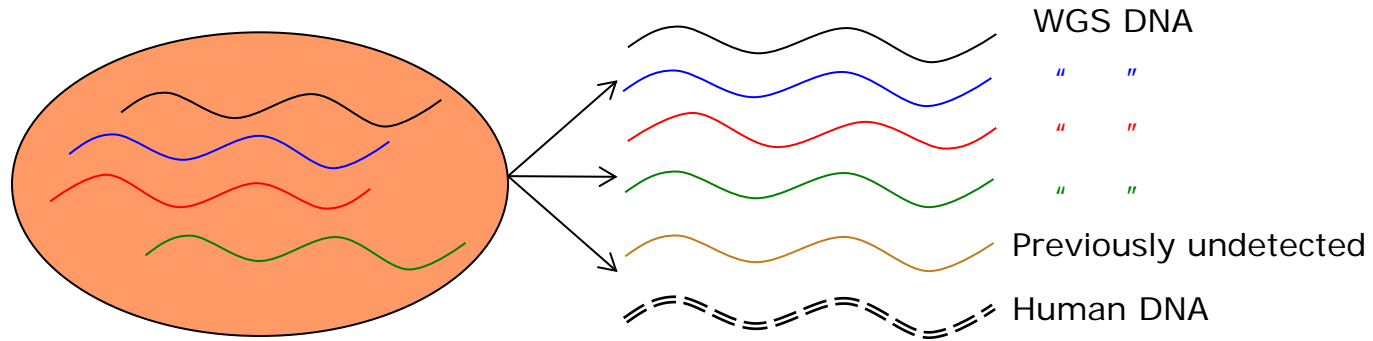


ITS and 16S analysis

# 16S Analysis of CF Patients



# Whole Metagenome Sequencing



Detection  
Genotyping/Phylogenetics

**Metagenomics**

Ab res, Virulence

Expression

Host genome

# Whole Metagenome Sequencing

Known Players

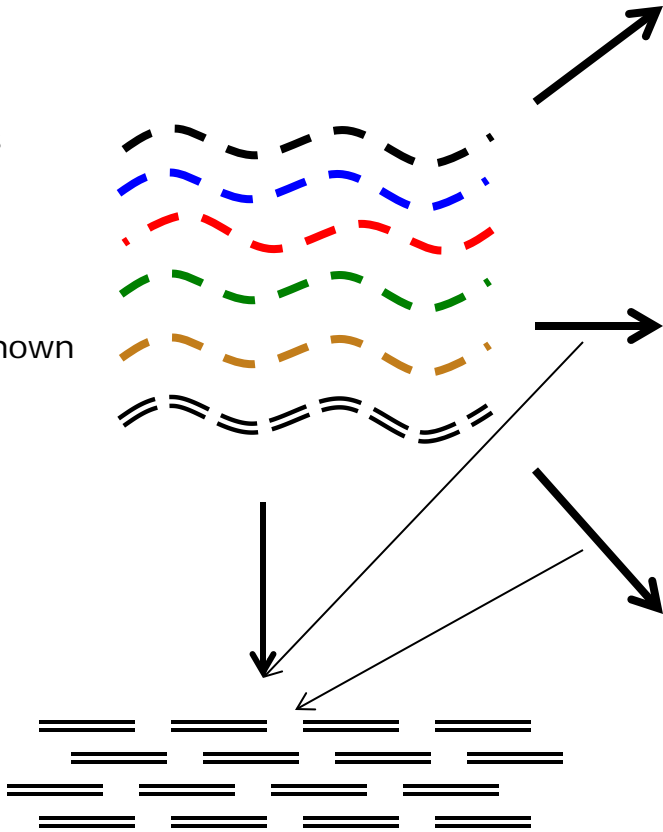
" "

" "

" "

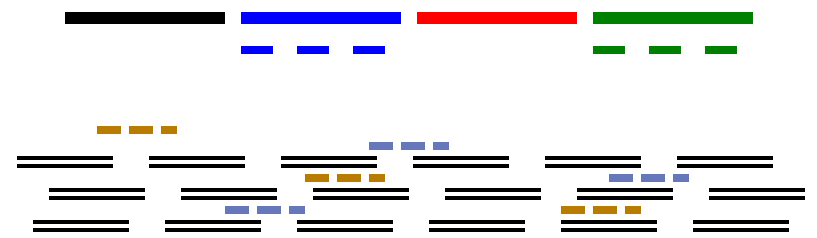
Previously Unknown  
Players

Human DNA

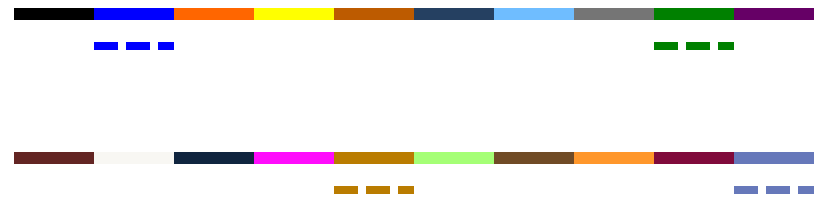


Strip out human reads  
*"in silico Hybridization"*

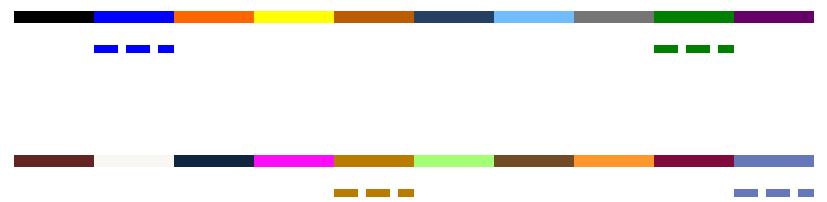
## Targeted BLAST



## Pan-Microbe BLAST



## Pan-Microbe Alignment



# Targeted BLAST – CF Metagenome

Alignment of all reads against know CF Threats

*S. aureus*

*P. aeruginosa*

*E. cloaca*

*K. pneumonia*

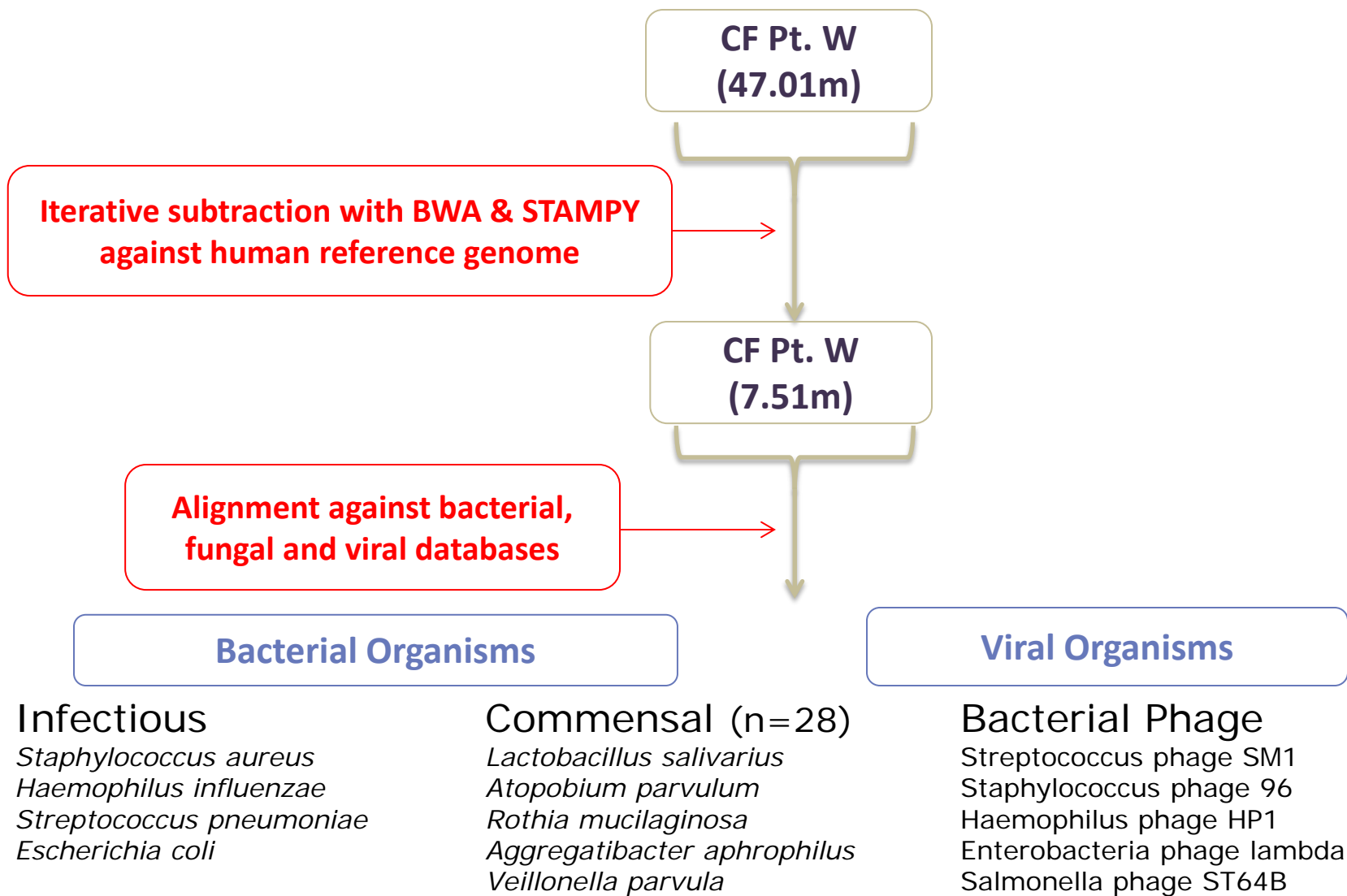
*H. influenzae*

*S. maltophilia*

*B. cepacia*

*A. baumannii*

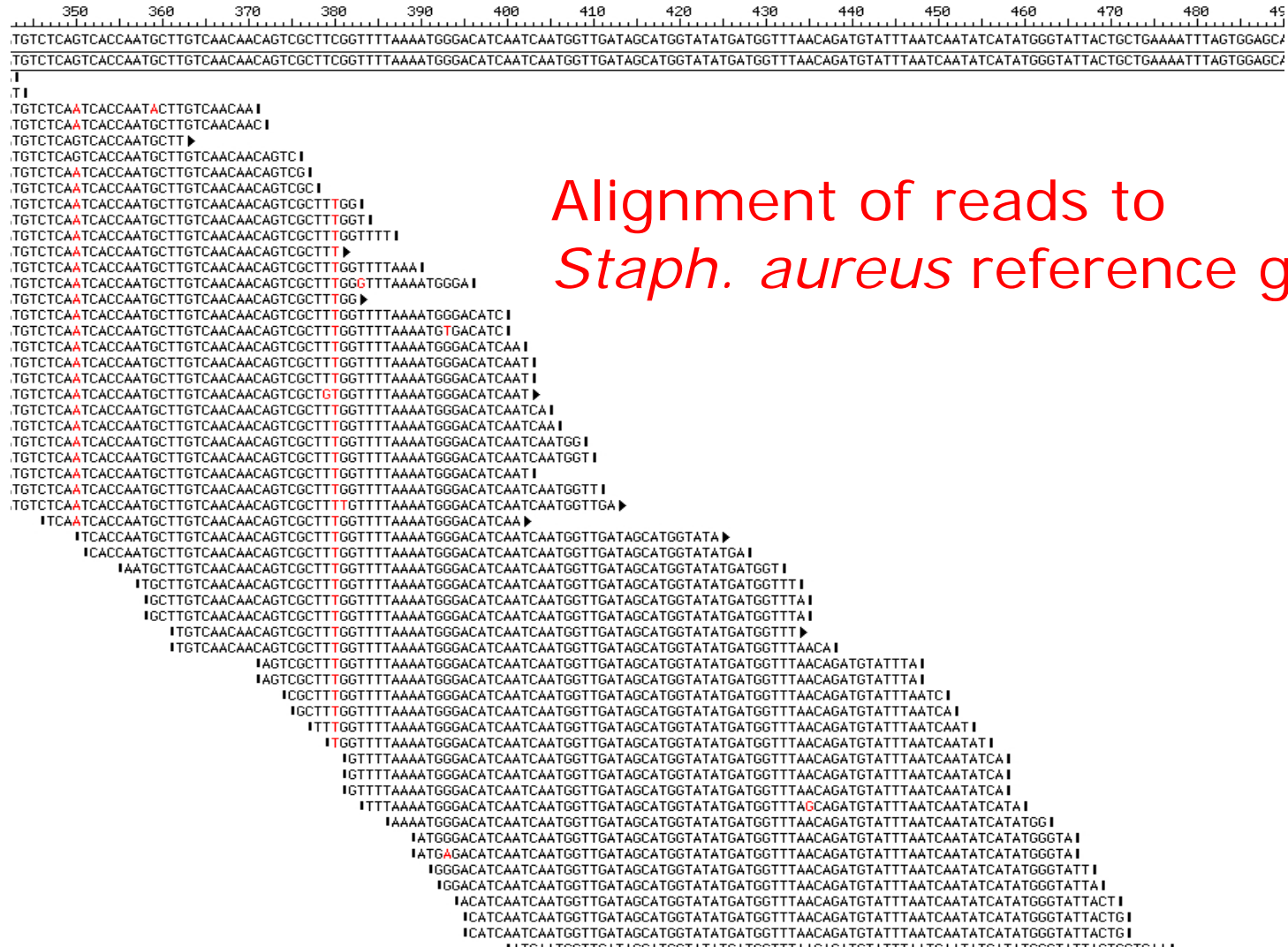
# Pan-Microbe Alignment



# Detecting Pathogens in CF

## *Staphylococcus aureus*

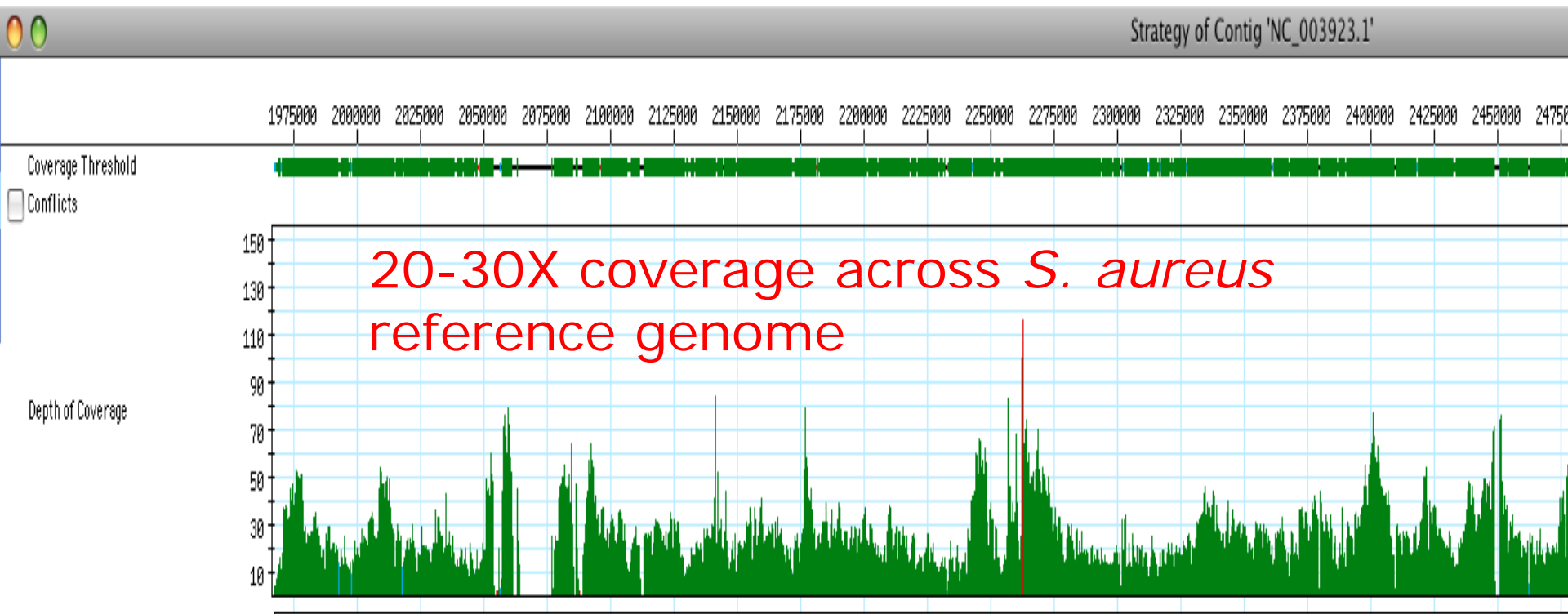
Alignment of Contig 'yqil1'





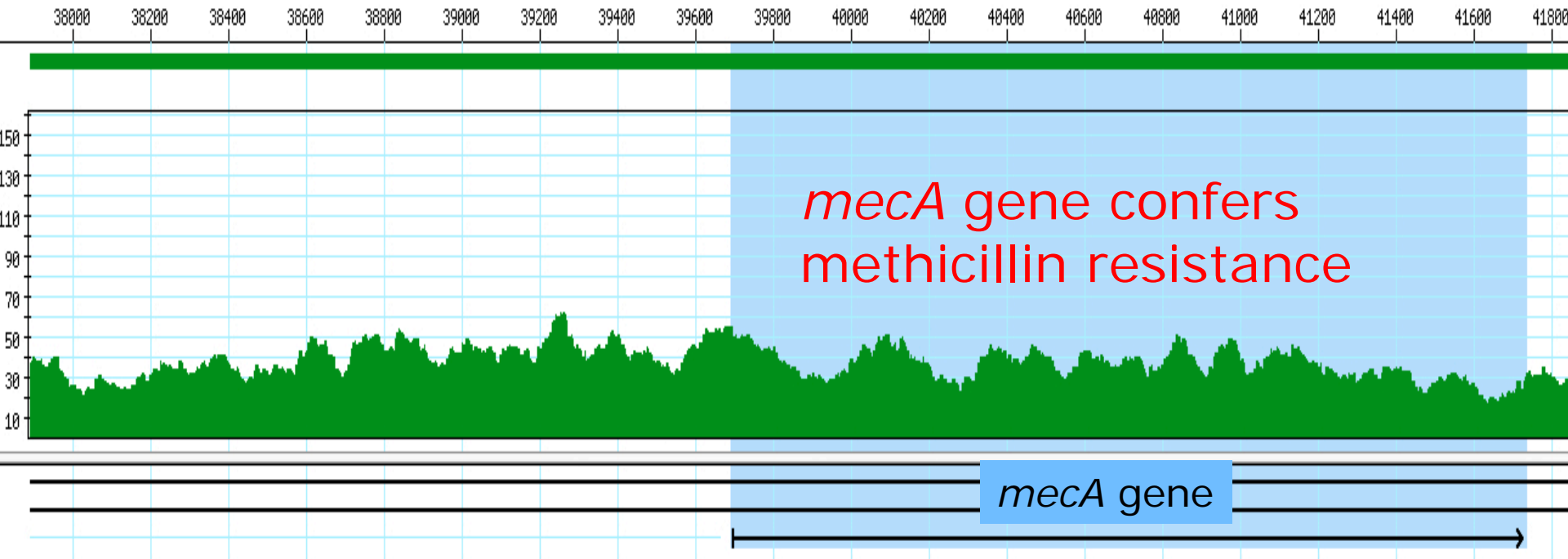
# Detecting Pathogens in CF

## *Staphylococcus aureus*

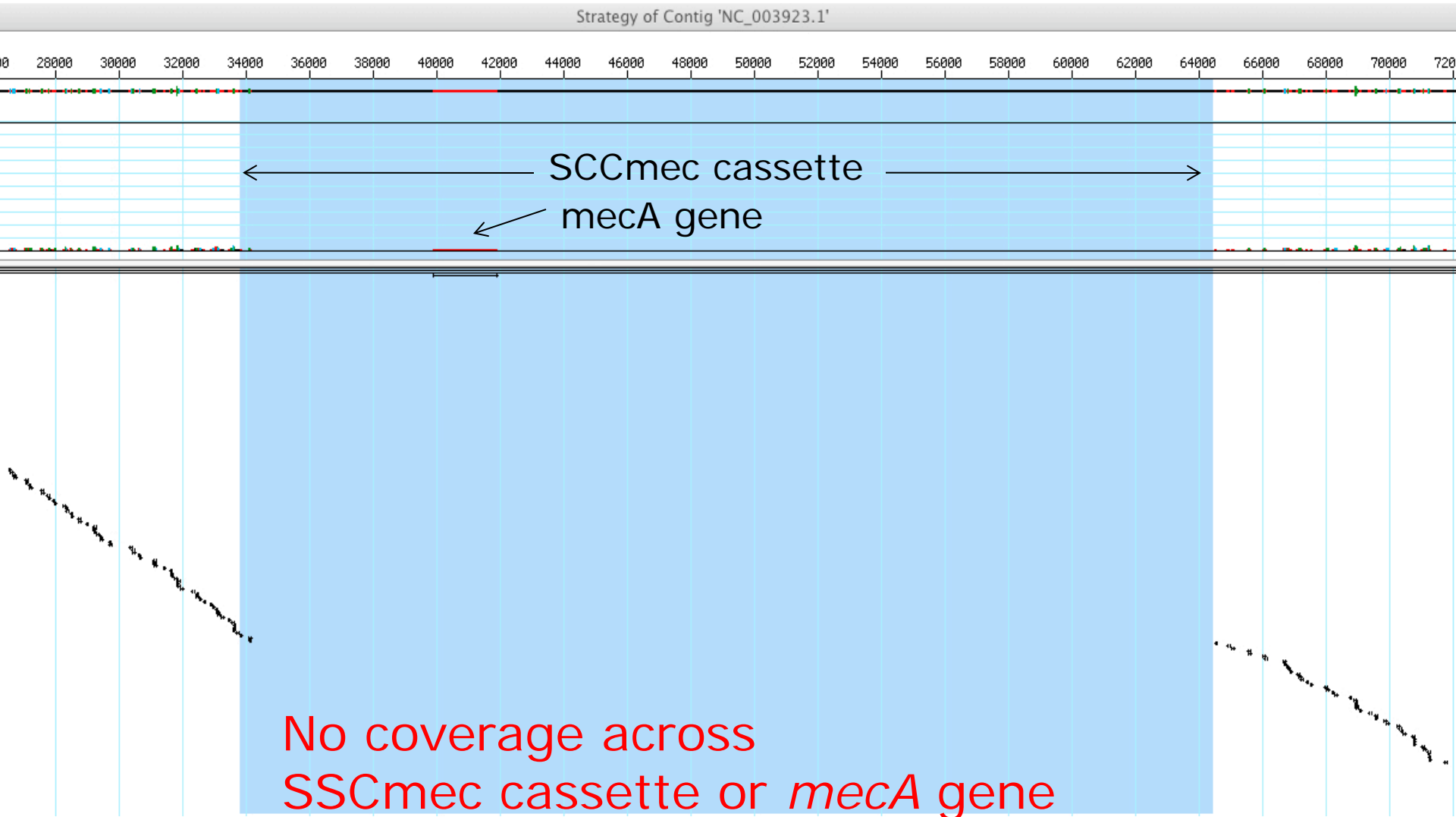


# Antibiotic Resistance in CF MRSA

Strategy of Contig 'NC\_003923.1'



# Antibiotic Susceptibility in CF MSSA not MRSA



# Genotyping a CF MRSA Sample

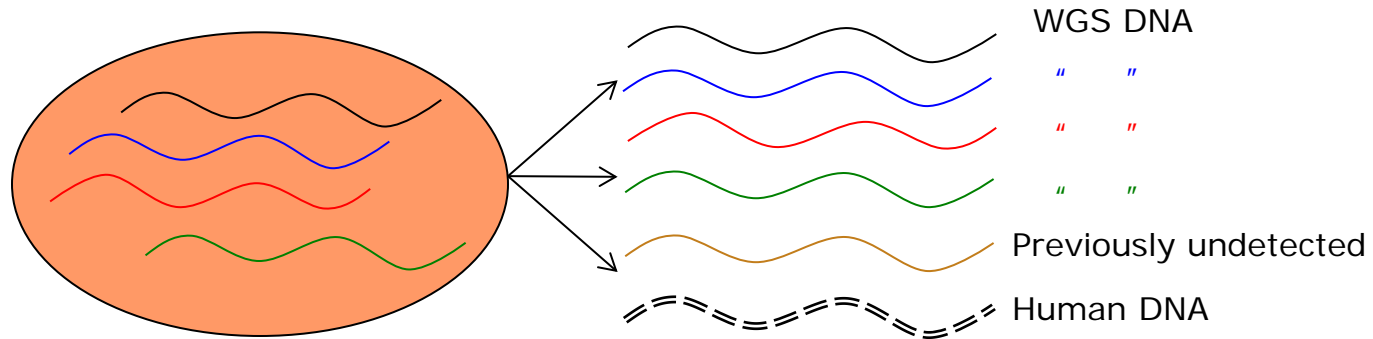
Selection: 834232 -> 834232 = 1

Alignment of Contig 'NC\_007793'

	834630	834640	834650	834660	834670	834680	834690	834700	834710	834720	8347
Translate	AAGCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAATAATTTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
Consensus	AAGCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAATAATTTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
NC_007793	AAGCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAATAATTTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_16_5026_7408#ACA	AAGCTTAATTCGAAGTC										
ILLUMINA-3BDE4F_0001_6_22_7476_11082#AC	AAGCTTAATTCGAAGTC										
ILLUMINA-3BDE4F_0001_6_8_7374_2792#ACAG	AAGCTTAATTCGAAGTCG										
ILLUMINA-3BDE4F_0001_6_29_6207_18948#AC	AAGCTTAATTCGAAGTCGAAT										
ILLUMINA-3BDE4F_0001_6_12_6146_11415#AC	AAGCTTAATTCGAAGTCGAATTAAG										
ILLUMINA-3BDE4F_0001_6_7_16792_9053#ACA	AAGCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAAT										
ILLUMINA-3BDE4F_0001_6_24_13216_4170#AC	AAGCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAATAAT										
ILLUMINA-3BDE4F_0001_6_9_2727_19513#ACA	AAGCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAATAATTT										
ILLUMINA-3BDE4F_0001_6_4_12728_7936#ACA	AAGCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAATAATTTTGGATATTTA										
ILLUMINA-3BDE4F_0001_6_13_11125_8874#AC	AAGCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAATAATTTTGGATATTTTAA										
ILLUMINA-3BDE4F_0001_6_13_10328_15579#A	AAGCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAATAATTTTGGATATTTATTTTAATAGTGTCA										
ILLUMINA-3BDE4F_0001_6_7_16931_10919#AC	AAGCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAATAATTTTGGATATTTTATT										
ILLUMINA-3BDE4F_0001_6_27_3282_19579#AC	GCTTAATTCGAAGTCGAATTAAGATGTTGAGCTGTAATAATTTTGGATATTTATTTTAATAGTGTGCATCACAGCGT										
ILLUMINA-3BDE4F_0001_6_26_17748_7808#AC	TCGAAGTCGAATTAAGATGTTGAGCTGTAATAATTTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATA										
ILLUMINA-3BDE4F_0001_6_2_9257_12999#ACA	GAAGTCGAATTAAGATGTTGAGCTGTAATAATTTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATAAT										
ILLUMINA-3BDE4F_0001_6_5_9448_15867#ACA	TTAAGATGTTGAGCTGTAATAATTTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTT										
ILLUMINA-3BDE4F_0001_6_2_6886_7161#ACAG	TGTTGAGCTGTAATAATTTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAAT										
ILLUMINA-3BDE4F_0001_6_4_19420_9317#ACA	TTGAGCTGTAATAATTTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTA										
ILLUMINA-3BDE4F_0001_6_8_4240_9806#ACAG	GCTTAATTTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_18_14906_10924#A	TTATTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_16_1749_2817#ACA	TATTTGGATATTTATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_3_16348_18895#AC	TATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_19_18306_17005#A	ATTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_5_7641_12422#ACA	TTTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_25_15575_6782#AC	TTTAATAGTGTGCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_7_4383_19552#ACA	GTCATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_21_6277_7670#ACA	CATCACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_6_10878_18806#AC	CACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_10_5040_4192#ACA	ACAGCGTTAAAAATAATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_18_16817_13883#A	AATGCTTACTTTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_26_12794_10475#A	TTTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_15_1112_20895#AC	TTAAATTTAAAC										
ILLUMINA-3BDE4F_0001_6_17_18244_15645#A	TTAAAC										
ILLUMINA-3BDE4F_0001_6_28_10834_7098#AC	TTAAAC										
ILLUMINA-3BDE4F_0001_6_15_4077_6292#ACA	TAAAC										
ILLUMINA-3BDE4F_0001_6_11_17344_16619#A	AAC										
ILLUMINA-3BDE4F_0001_6_8_18354_17171#AC											
ILLUMINA-3BDE4F_0001_6_13_3570_7946#ACA											
ILLUMINA-3BDE4F_0001_6_3_6523_11451#ACA											
ILLUMINA-3BDE4F_0001_6_24_18973_17868#A											
ILLUMINA-3BDE4F_0001_6_20_15648_10161#A											
ILLUMINA-3BDE4F_0001_6_6_10678_18637#AC											

SNP helps define strain as Clonal Complex 8 or "USA300"

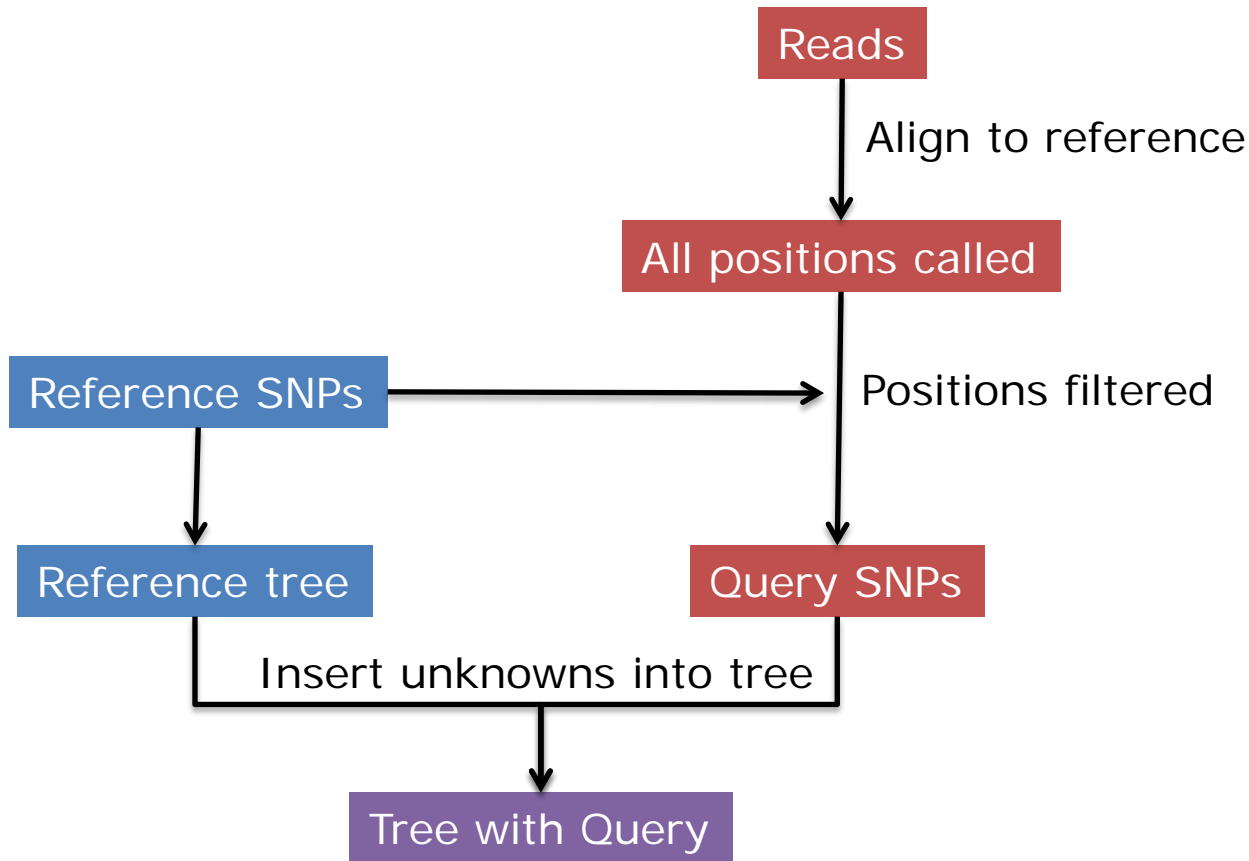
# Whole Metagenome Sequencing



Detection  
**Genotyping/Phylogenetics**

Metagenomics  
Ab res, Virulence  
Expression  
Host genome

# Straight from Sample - WG-FAST

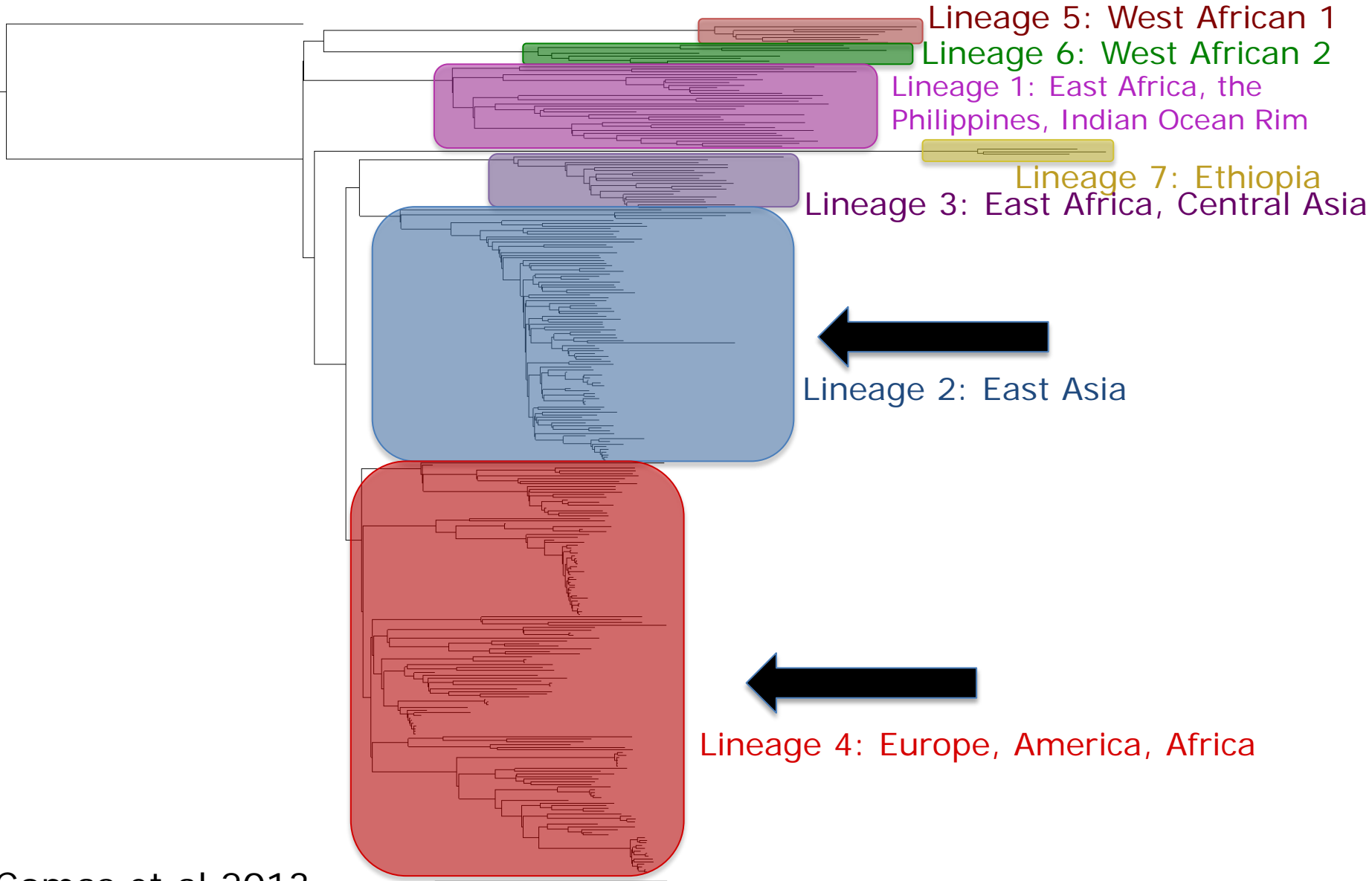


# Clinical Samples

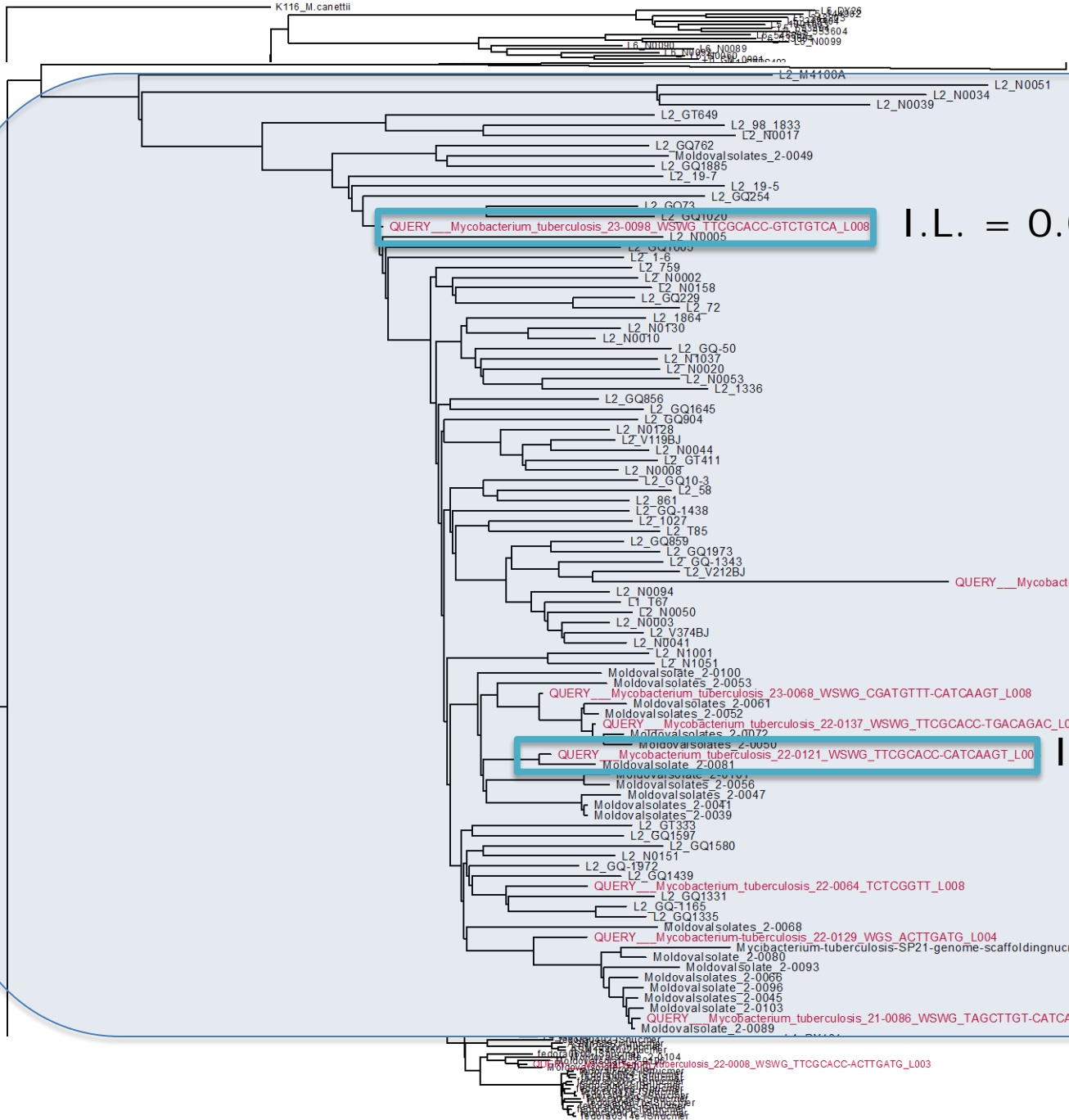
- Whole genome prep on sputa DNA extracts
- 26 remnant sputum samples
  - Crude sediment DNA extraction
  - Moldova
  - via Global Consort TB Dx Development
- Hiseq - Reads will be human, Mtb, and any other organisms present
- Compare to high quality reference tree



# Mtb Reference Phylogeny







I.L. = 0.008

I.L. = 0.999

# The Promise of Next Gen for PH and Clinical Laboratory

- Detection and Genotyping – Now
- Antimicrobial Resistance Profiling - Now
- Virulence Phenotyping - Now
- Complex Infections - Now
- Host Factors – Soon
- 3-8 Hour Sequencing – Now
- 15 min Sequencing – Soon
- Kit-based + Automation - Soon

# Collaborators

- TGen
  - Rebecca Colman
  - Elizabeth Driebe
  - James Schupp
  - Jason Sahl
  - Chandler Roe
  - Darrin Lemmer
  - Paul Keim
- AZ Dept Health Services
  - Victor Waddell
  - William Slanta
  - Ken Komatsu
- CDC
  - Mycotics, DHQP, Enterics, Vector
- Global Consortium - GCDD
- University of California San Diego
  - Timothy Rodwell
  - Antonino Catanzaro
- San Diego State Univ
  - Faramarz Valfar
- University of Arkansas
  - Donald Catanzaro
- Phthisiopneumology Institute (PPI), Republic of Moldova
  - Valeriu Crudu
- Yale University
  - Ted Cohen

**Thank You!**