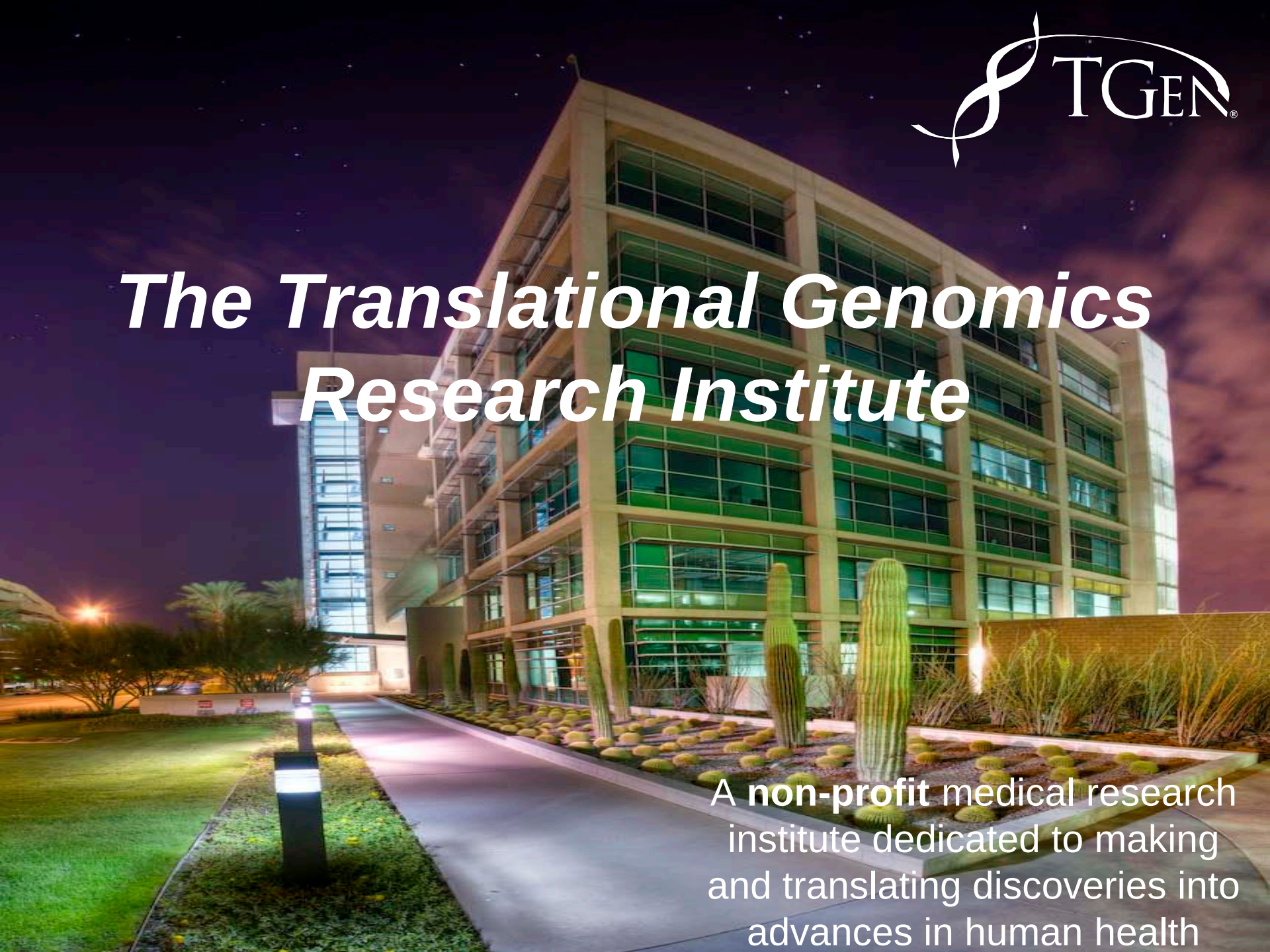
# Next Generation of Molecular Epidemiology and Diagnostics

David M. Engelthaler

TGen North



# *The Translational Genomics Research Institute*

A **non-profit** medical research institute dedicated to making and translating discoveries into advances in human health

# Divisions of TGen

Jeffrey Trent Ph.D.
President and CSO

- Neurogenomics
- Cancer and Cell Biology
- Genetic Basis of Human Disease
- Computational Biology
- Clinical Translational Genomics (Cancer)
- Diabetes and Metabolic Diseases
- Pharmaceutical Genomics
- Pathogen Genomics
  - TGen North

TGEN®
NORTH

Diagnostic ♦ Analytic ♦ Forensic ♦ Epidemiologic
Translational Pathogen Genomic Research

**We have the ability, therefore, we have the responsibility**

TGEN NORTH

Diagnostic ♦ Analytic ♦ Forensic ♦ Epidemiologic
Translational Pathogen Genomic Research

Profile:

- Est. April 13, 2007
- Size: ~11,000 sqft
- Staff size: 28
  - PhD, MS, MBA, MD, MPH, BS
- Funded Projects: 15-20

TGEN NORTH

TGEN NORTH

# TGen North

- Vision:
  - Diagnostic ♦ Analytic ♦ Forensic ♦ Epidemiologic

  Translational Pathogen Genomic Research

- Mission:

  "Advancing public health, clinical medicine, and biodefense through targeted microbial genomics"

- Values:

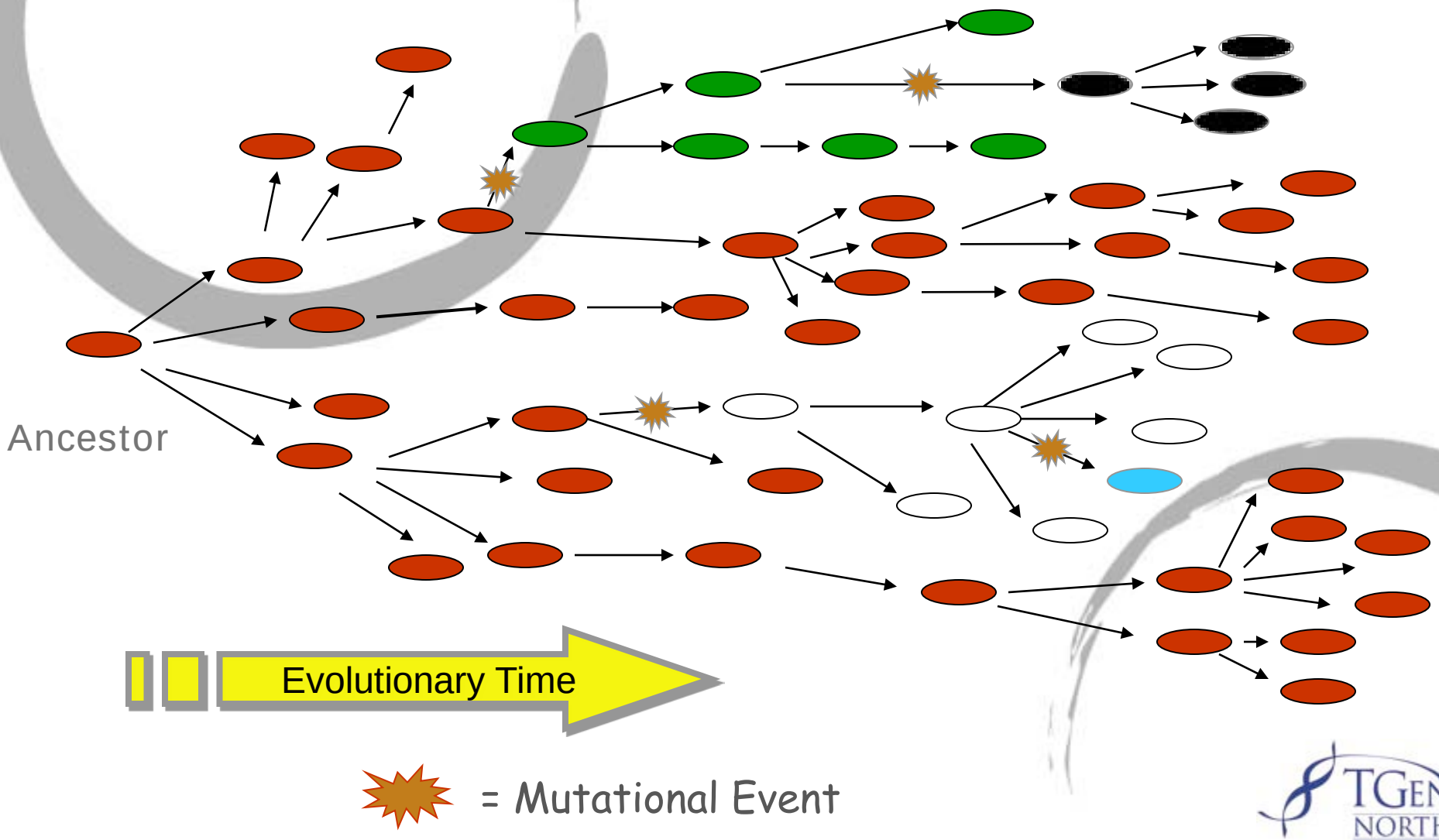
  Team, Leadership, Responsibility, Change, Growth, Vision

TGen NORTH

# Getting to Next Gen

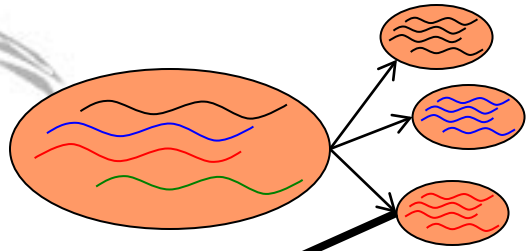
# Pathogen Genomic Analysis Concepts

# Genomic Diversity



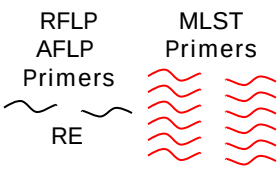
# How best to analyze samples?

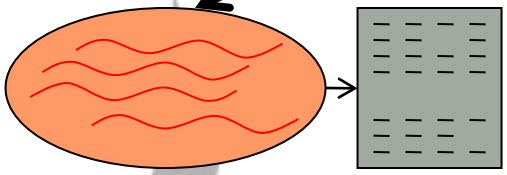
Direct culture of mixed population

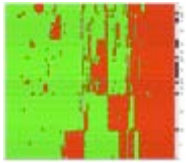


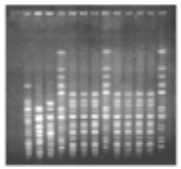
Isolated organisms speciated by morph, biochem, molecular

Genotype isolates

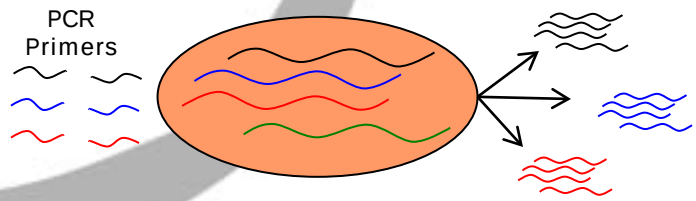




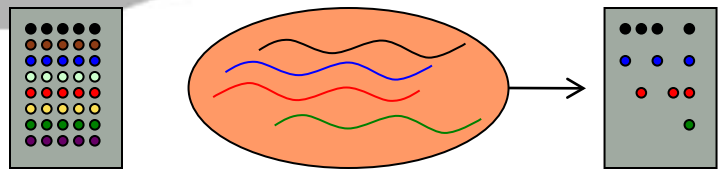


Detection in mixed population



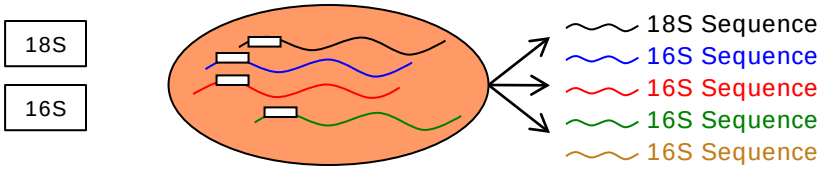

Highly sensitive, specific, quantitative analysis

Microarray



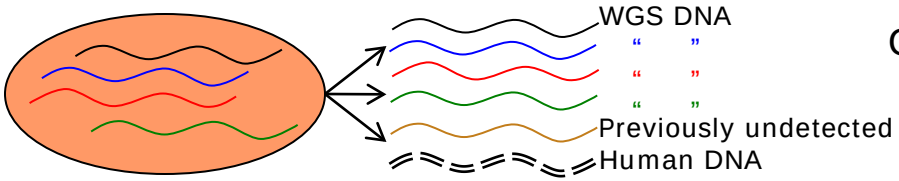
Detection and genotyping based on select SNP profile

Metagenomics




454 Next Gen sequencing of amplified 18S/16S

Whole metagenome sequencing




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

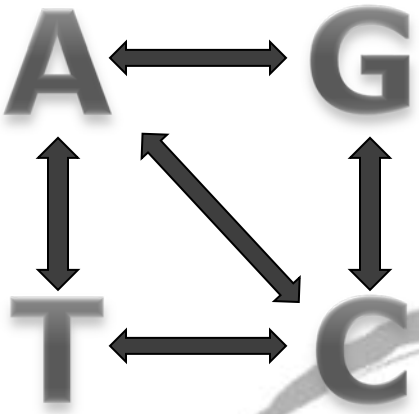
# Next Gen Sequencing

**Capillary electrophoresis (Sanger)**

350 Kb/run
~0.1 Genomes

# SNP
# Single Nucleotide Polymorphism

- The mutation of one DNA nucleotide into another
- Some SNPs cause change in protein structure; some cause no changes at all
- All are useful for monitoring populations structure



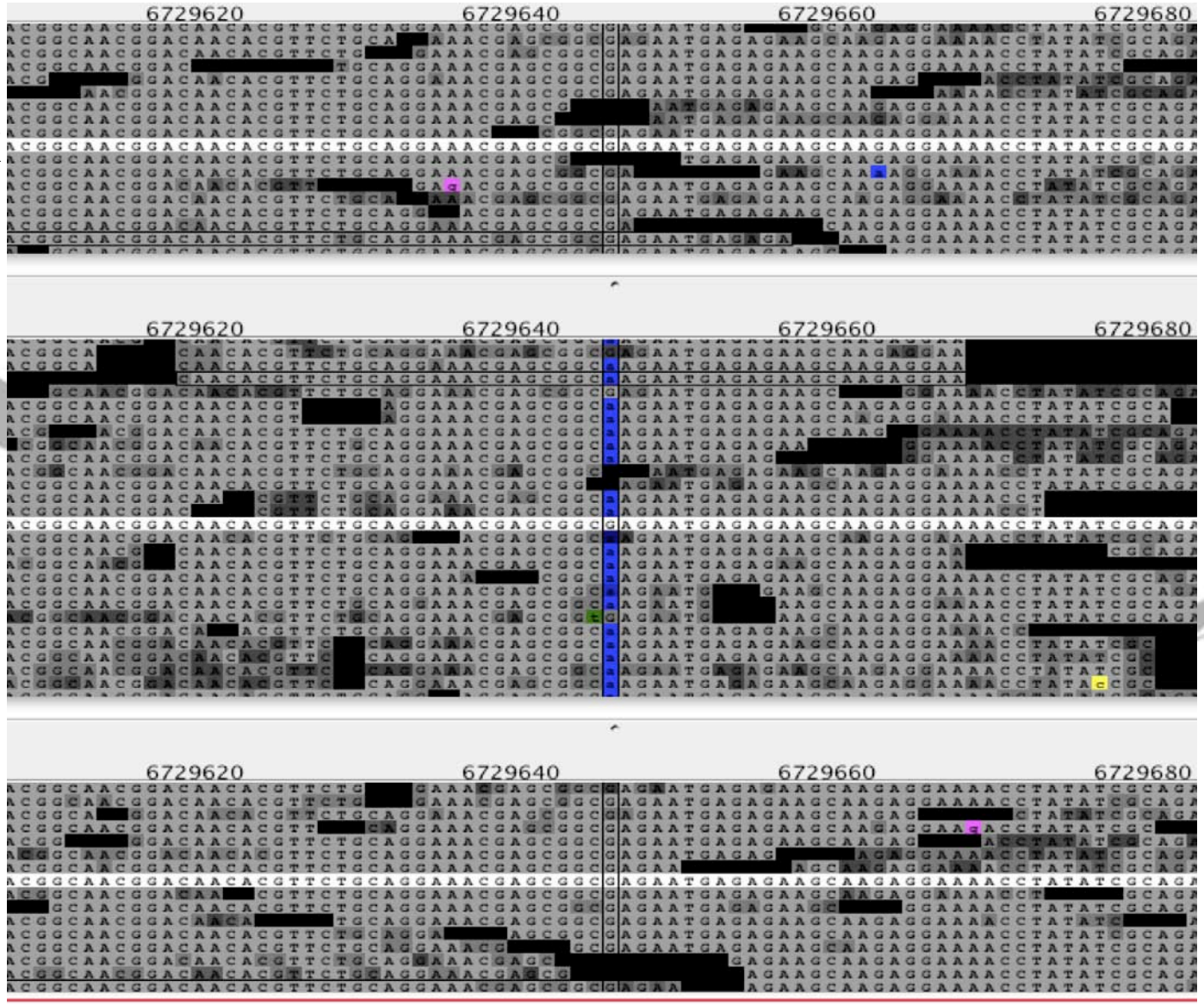

TGEN NORTH

# Finding SNPs

Sample X

Sample Y

Sample Z




TGen NORTH

# Our Hypothesis

**Next Gen sequencing technology can have immediate impacts molecular epidemiology and clinical medicine**

# Our Translational Strategy

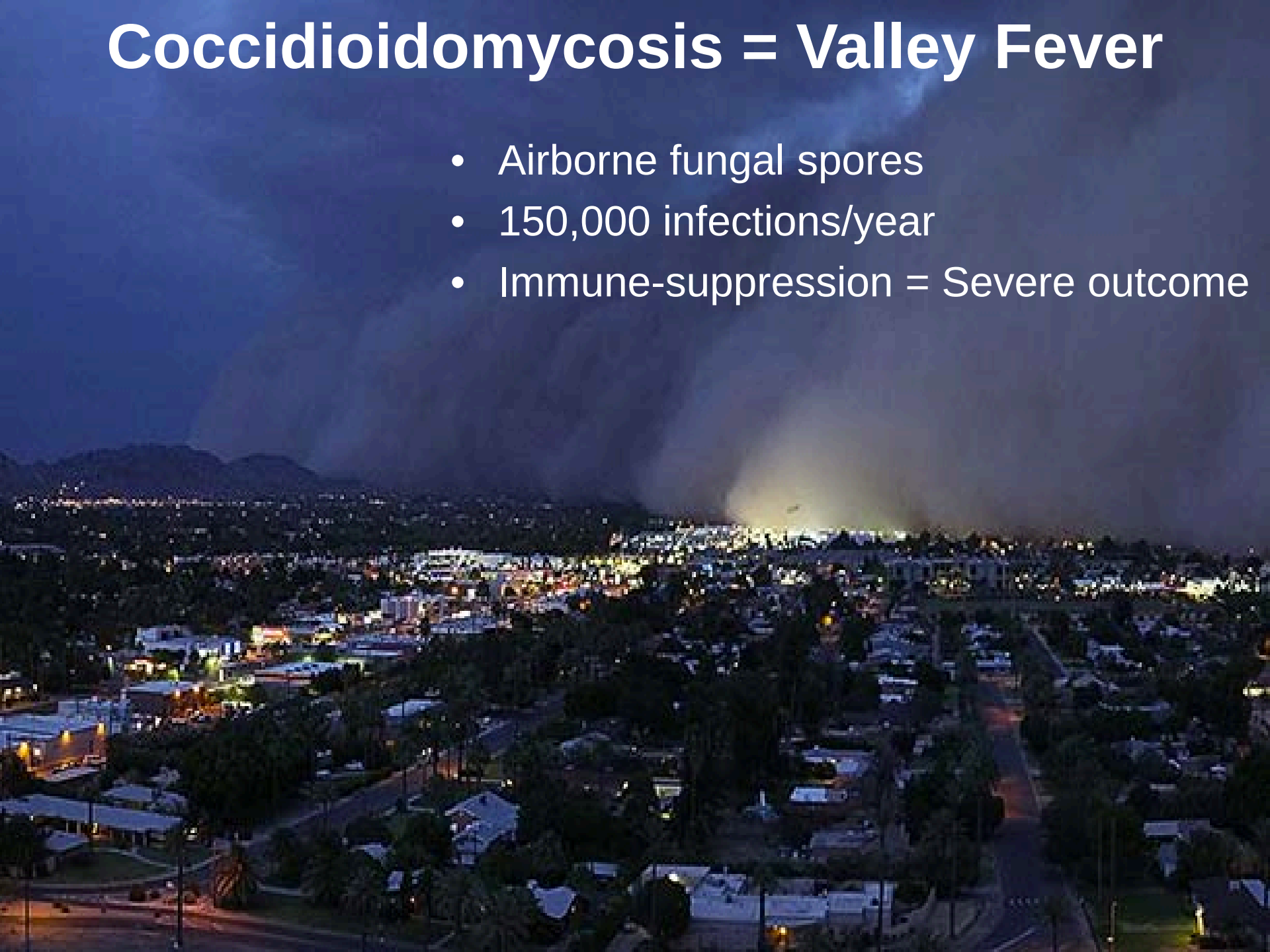
**Apply next gen to ongoing outbreak investigations and in complex clinical sample analysis**

TGEN NORTH

# First Stop:
# Cocci Outbreak

# Coccidioidomycosis = Valley Fever

- Airborne fungal spores
- 150,000 infections/year
- Immune-suppression = Severe outcome



# Valley Fever Outbreak

- Three organ recipients from a common donor from L.A. all developed severe Valley Fever – two died
- No samples from donor existed for testing (donor: homeless patient found unresponsive on the street)
- CDC needed evidence to determine if all three were infected by transplants
- CDC contacted TGen North for assistance with investigation

TGen NORTH

# Valley Fever Sequencing

- *Coccidioides immitis* - 28.9 Mb genome
- 50bp Single End SOLiD (Carpten Lab)
- Many months of bioinformatic analysis
  - Numerous repetitive elements (17%)
  - Highly recombining
  - Limited prior knowledge

TGEN NORTH

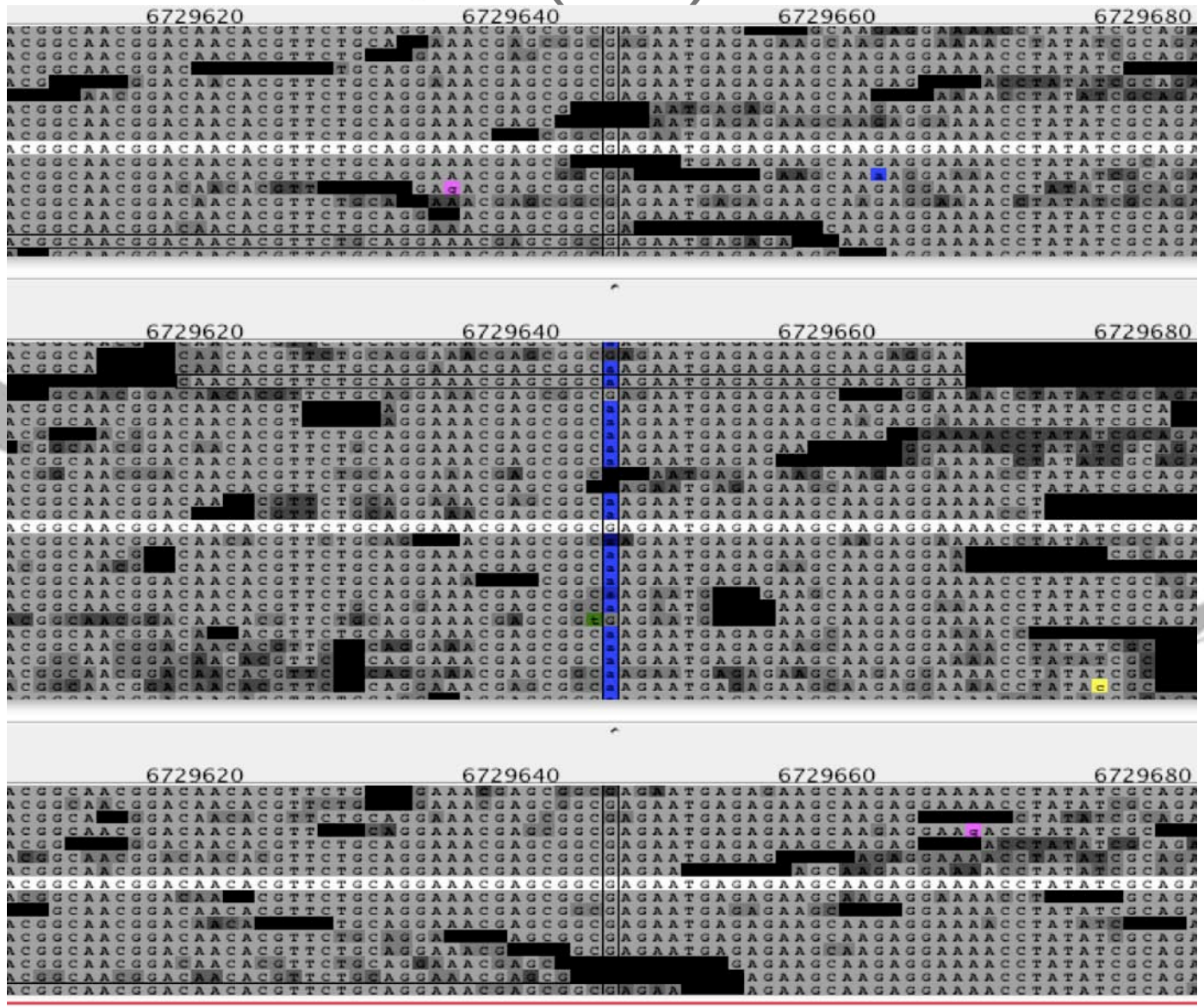
# Among-cluster isolate SNPs

(n=3)

Pt. X

Pt. Y

Pt. Z

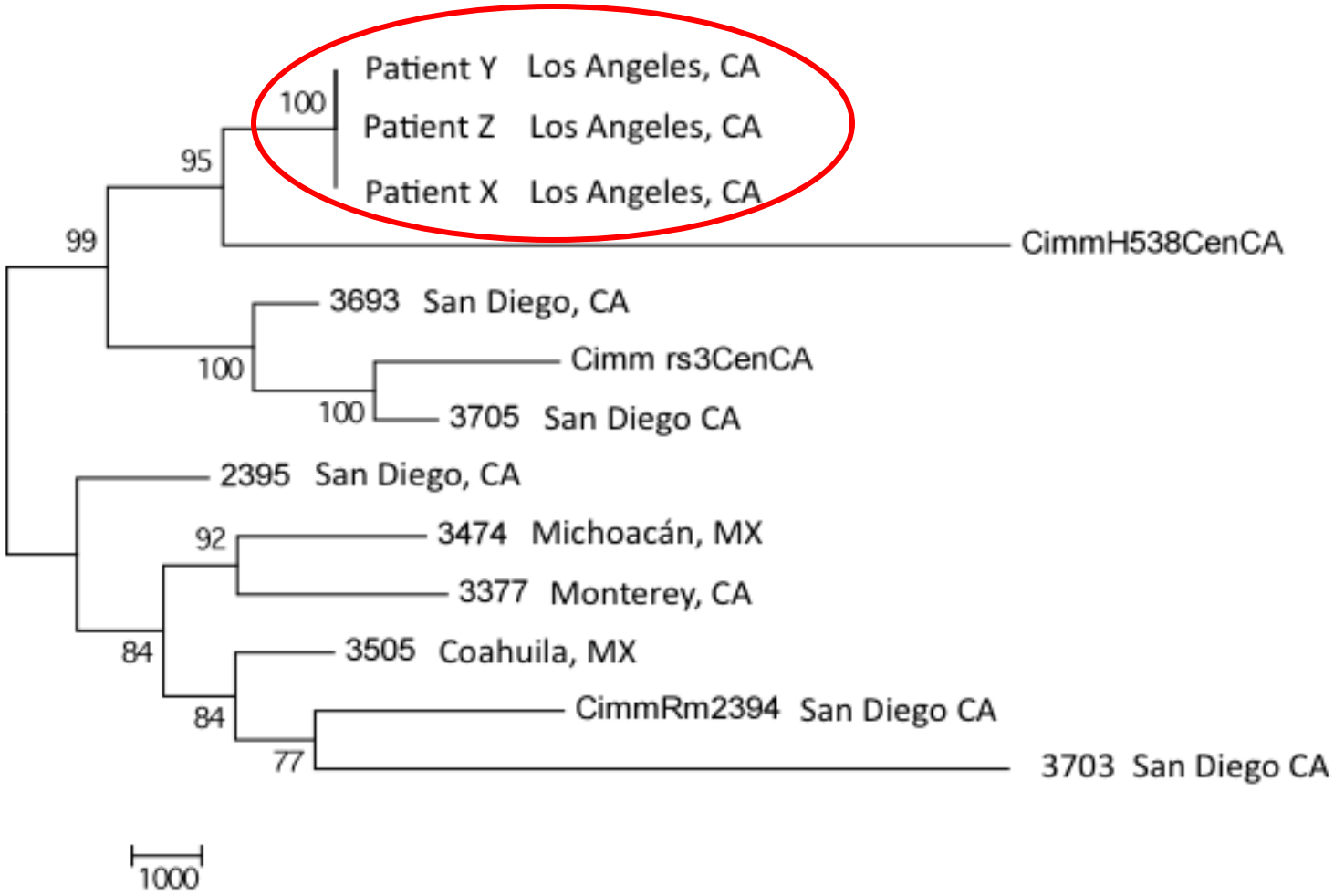



Engelthaler et al 2010

TGen NORTH

# Maximum Parsimony Phylogenetic analysis of 13 *C. immitis* genomes

32,695 SNPs, C.I. = .63




Engelthaler et al 2010

# Next Stop:
# Crypto … “The Killer Fungus”

Infectious diseases on msnbc.com

# Deadly airborne fungus in Oregon set to spread

The new, rare strain has killed 1 in 4 infected, researchers say

Jump to text below

TODAY

# The 'Killer Fungus': Should We Be Scared?

By ALICE PARK Friday, Apr. 23, 2010

More Related

- Across Asia, Dengue Fever Cases Reach Record Highs

Sponsored Links

Penny Stock Soaring 3000%
Sign up for Free to
out what the next
3000% Stock Wi
www.PennyStock

TODAY: Ap
iPads & Ma

# Killer Lung Fungus Hits Northwest

SPORES LODGE IN TISSUE; KILL 25% OF THOSE INFECTED

By Mary Papenfuss, Newser Staff
Posted Apr 23, 2010 2:00 AM CDT | Updated Apr 23, 2010 7:11 AM CDT

STORY COMMENTS (23) Do you like this story? Like

The spores of Cryptococcus gatti lodge in the lungs of those infected. (Shutterstock)

(NEWSER) – A virulent, deadly strain of a fungus that settles in human lungs has bee discovered in the US Northwest.

Cryptococcus gatti—or C. gatti—first spread from Canada into the US 5 years ago, but has recently developed into a particularly powerful strain in Oregon and is expected to move next to Washington, Idaho, and California. Nearly a quarter of people infected have died, reports National Geographic. "As the range of the organism expands and the number of cases increases, it's becoming a concern," a Duke University expert tells CNN.

The rare tropical fungus has generally infected only those whose immune systems compromised, but the new str individuals. Scien in th

MY TAKE ON THIS STORY: CLICK BELOW TO VOTE

HILARIOUS 2% | DEPRESSING 2% | SCARY 92%

Home > Blogs > Tech Talk

Tech Talk

April 26, 2010 10:20 AM

# First "Killer Bees," Now "Killer Fungus"

Posted by Charles Cooper 1 comments

Share 17 submit to digg Tweet 0 Share E-mail Pr

It may sound like a John Belushi routine from Saturday Night Live, but scientists are exp spread of what they describe as a potentially lethal fungus making its way down the Paci So far, the biggest outbreak has taken place on Vancouver Island in British Columbia, b past the order into Oregon where it's become a "a major source of illness in the region PLoS Pathogens.

So far, five of 21 people contracting the fungus in the U.S. have died.

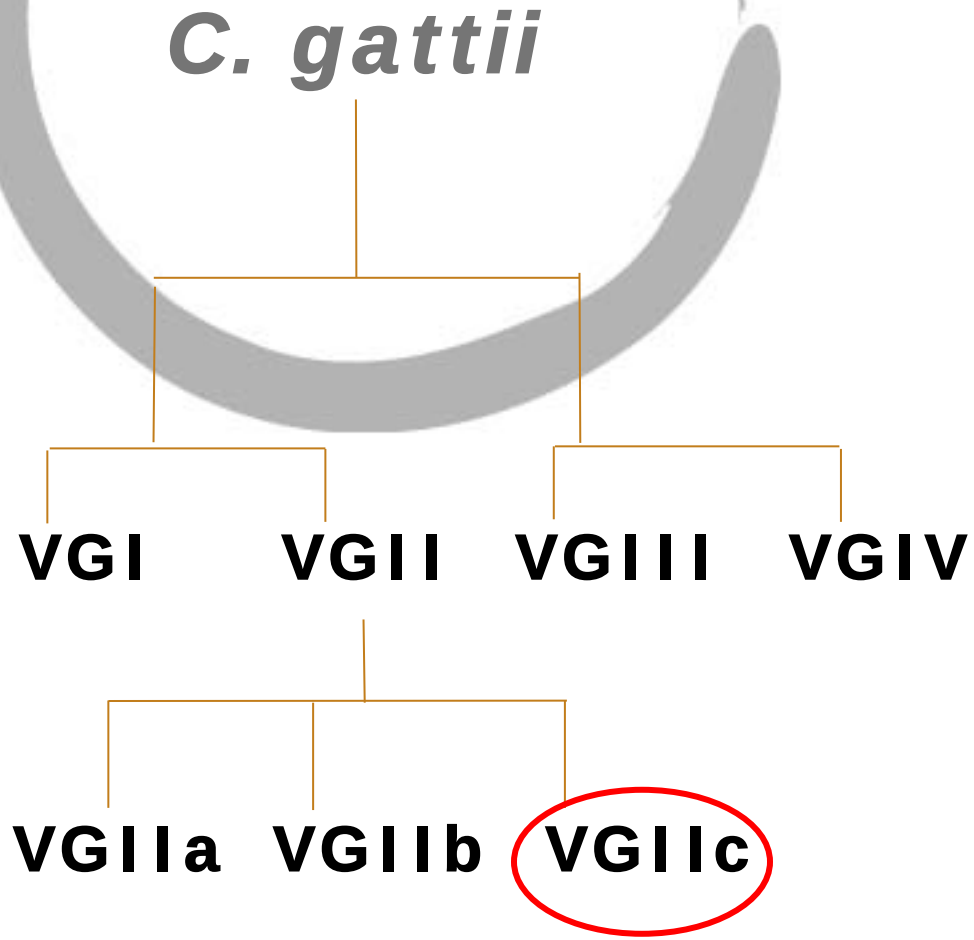
"This novel fungus is worrisome because it appears to be a threat to otherwise hea of Duke University in North Carolina, who led the study.

The fungus, known as *Cryptococcus gattii*, was previously found in the tropics, where it has frequently been a with Eucalyptus trees. But *C. gattii* was found in Vancouver Island in 1999 and has since been linked to disease i

# *Cryptococcus gattii* in Pacific Northwest

- *C. gattii* – tree/soil fungus; can infect inmmune-competent
- Emerged on Vancouver Island in 1999
- New outbreak US: human cases in Oregon and Washington
- US outbreak seems to be caused by new strain (VGIIc) with a fatality rate of 25%

# Groupings within *C. gattii*




RFLP, AFLP, MLST

MLST

TGEN NORTH

# Epidemiological Questions For TGen

Is the current outbreak truly from a new strain introduced to PNW?

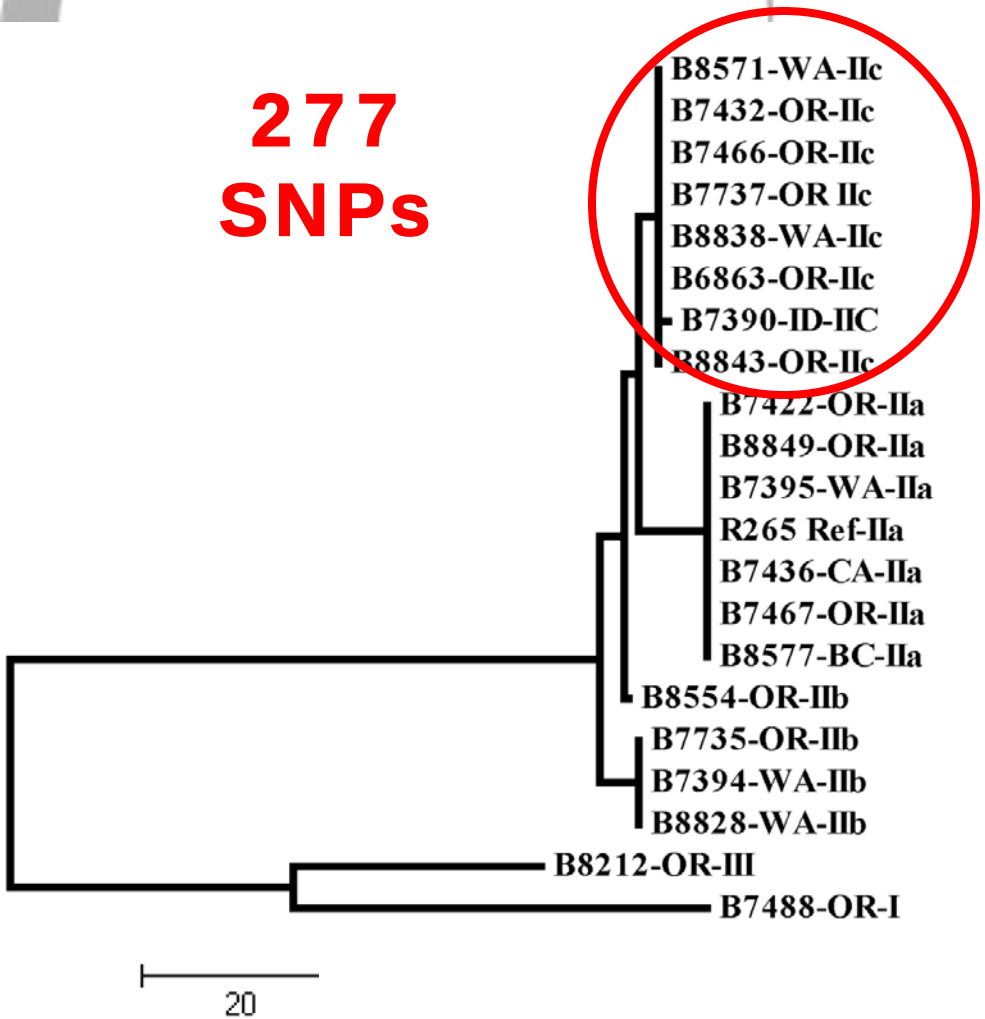
Can we distinguish between isolates from outbreak cluster?

TGEN NORTH

# Maximum parsimony tree - 20 isolates

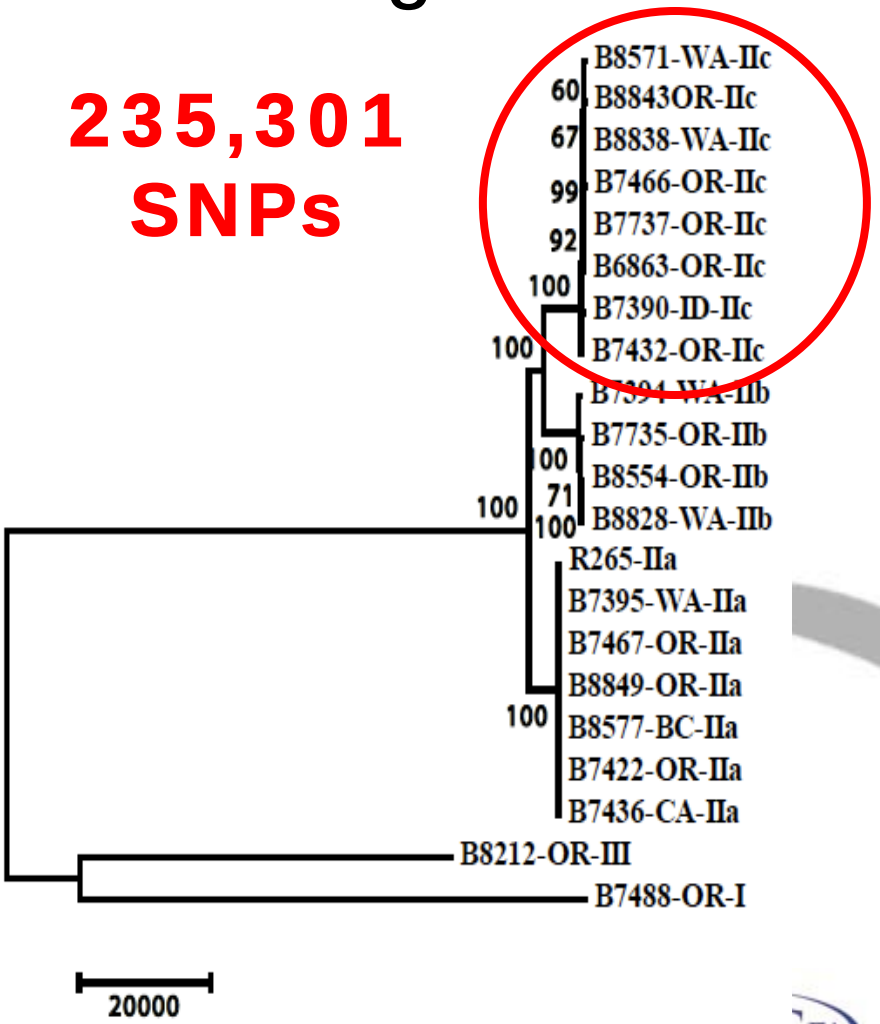
## MLST

**277 SNPs**




## Whole genome

**235,301 SNPs**



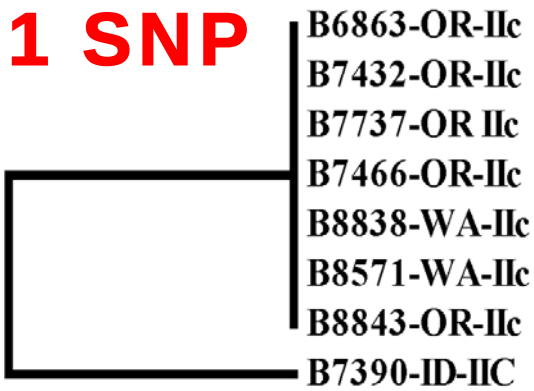

Gillece et al 2011

# Maximum parsimony - VGIIc isolates

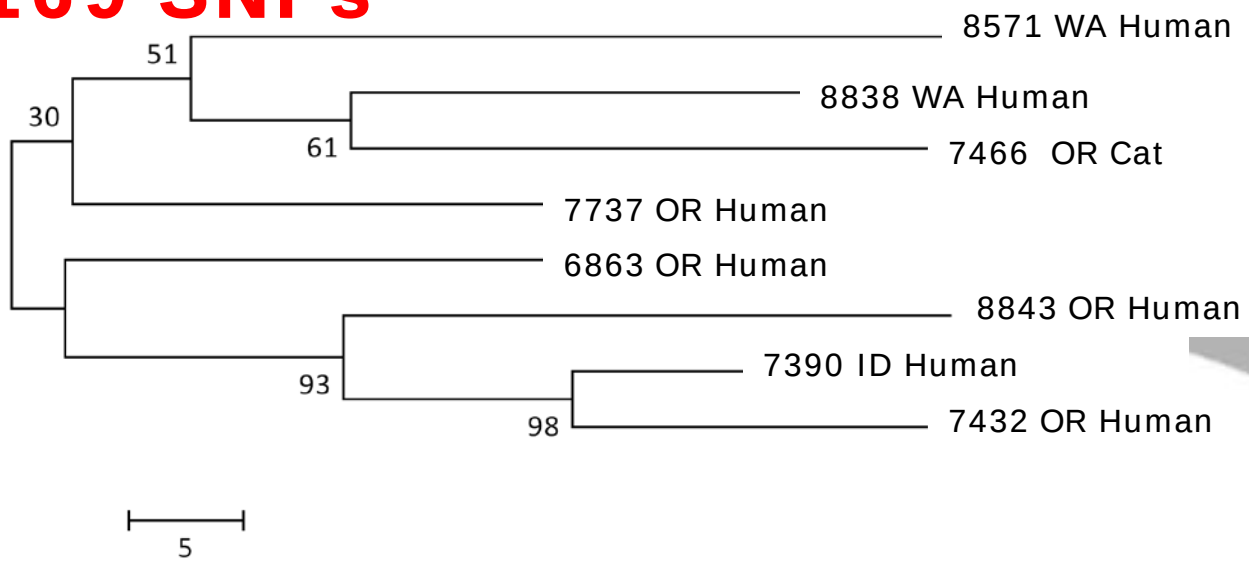
MLST

Whole genome

**1 SNP**

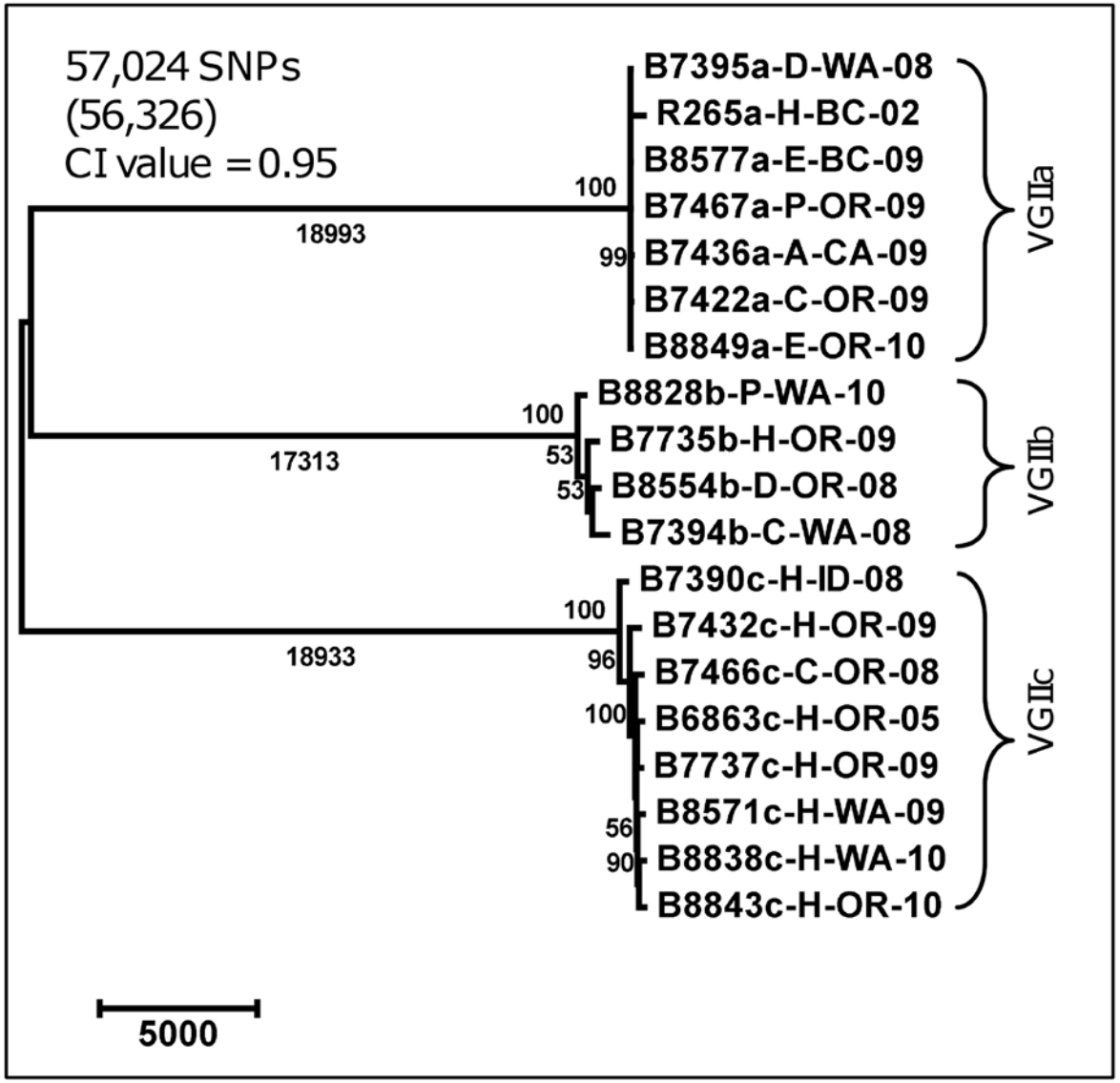



**169 SNPs**




Gillece et al 2011

# Crypto Family Tree in PNW




Gillece et al 2011

# Crypto Initial Findings

- Whole Genome Sequence analysis showed distinct new populations of Cryptococcus in the Pacific Northwest
- Looking at these genomes will allow us to find possible genetic causes of increased death and disease
- Analysis has led us to develop new tests to rapidly identify and fingerprint Crypto strains

TGEN NORTH

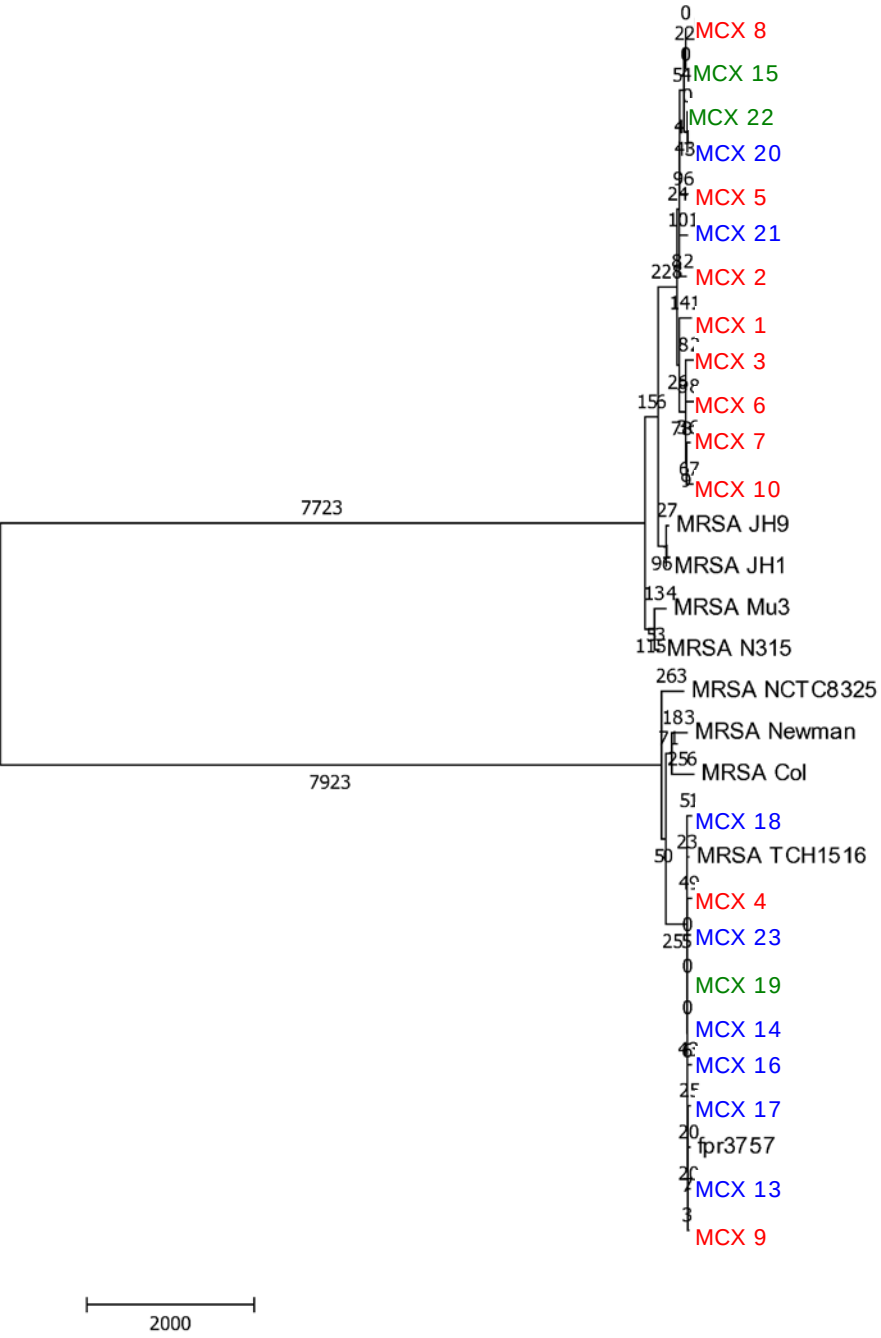
# MRSA Outbreak

Suspect Health Care Provider Associated Infections

# Arizona Outbreak

- AZ Hospital Staff identified increased MRSA infections associated with HCP
- 10 isolates taken from HCP and patients
- 10 isolates taken from random patients
- Isolates sent to TGen North for WGST Analysis

TGEN NORTH




# AZ MRSA Hospital Outbreak Investigation

10 HCP Associated Isolates
10 Random Patient Isolates
3 Repeat Isolates

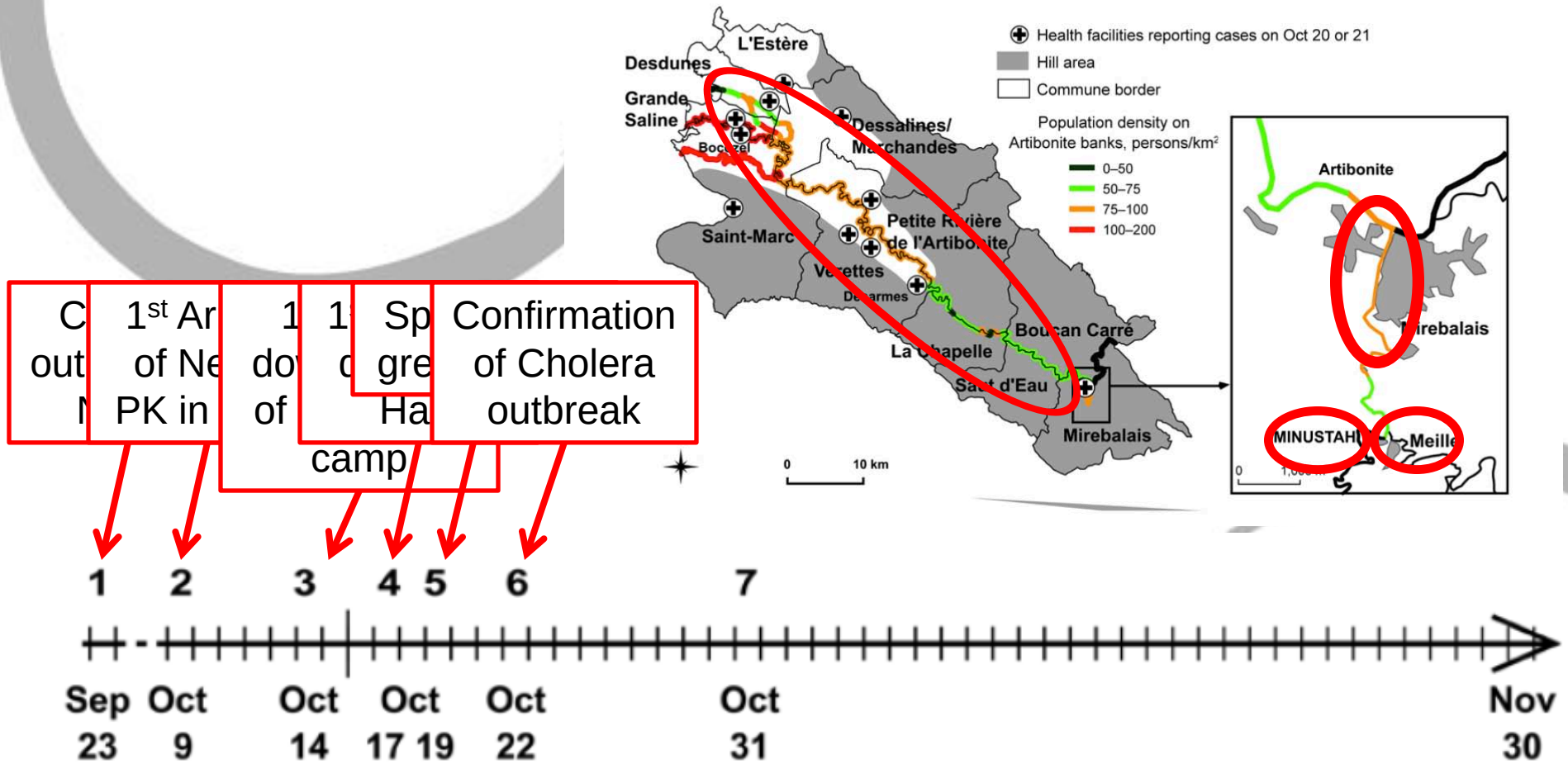
# Cholera in Haiti

# Crises in Haiti

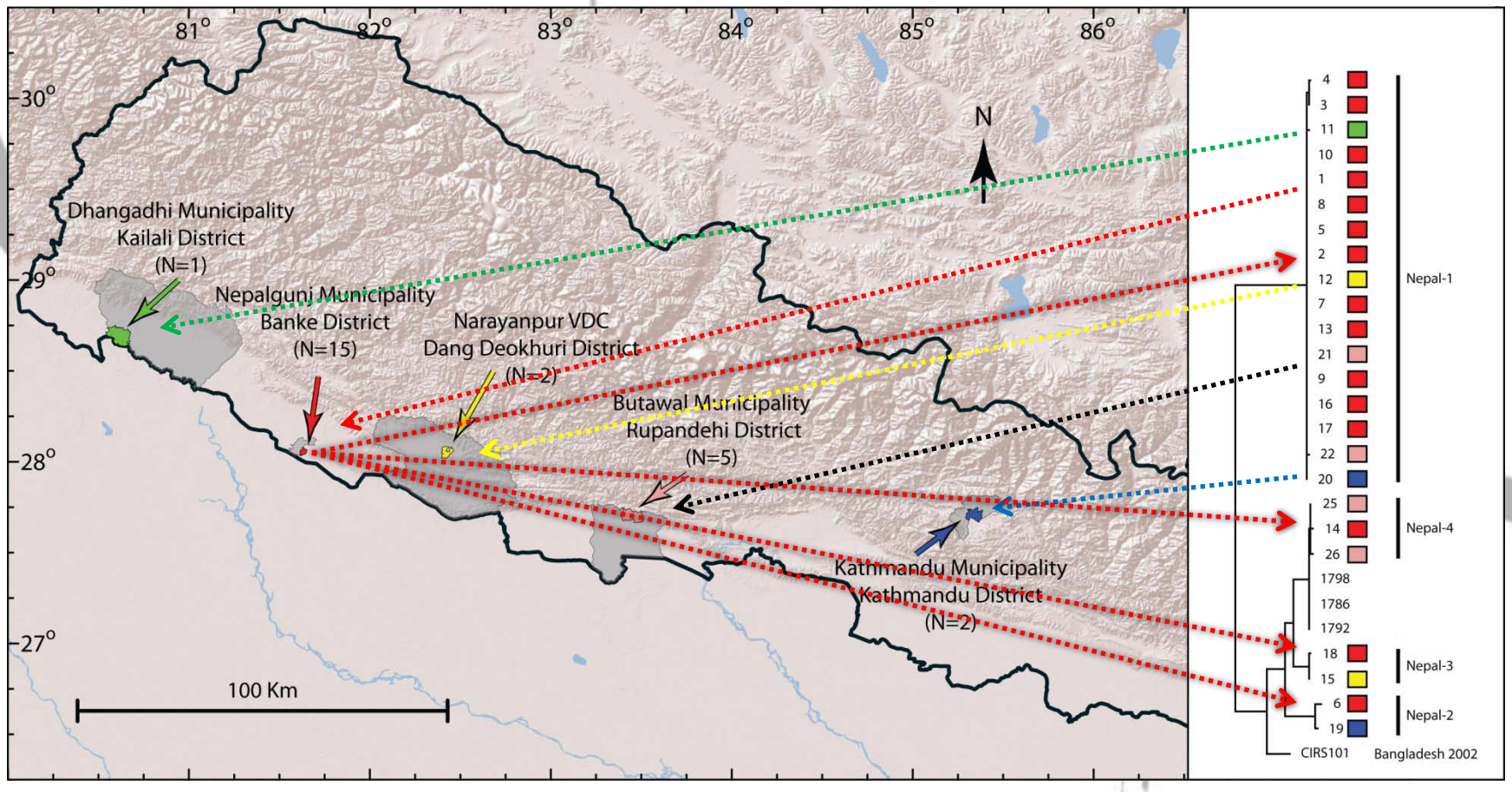
- Jan 12, 2010 - 7.0 $M_w$ earthquake hit Haiti
- By Jan 24 – 52 aftershocks, 316K dead, 300K injured
- Oct 10 2010 – Cholera confirmed in Haiti
- By Feb 21, 2011 - >4600 deaths, 250K cases of cholera in Haiti

TGEN NORTH

# Epi Points to Nepal Peacekeepers



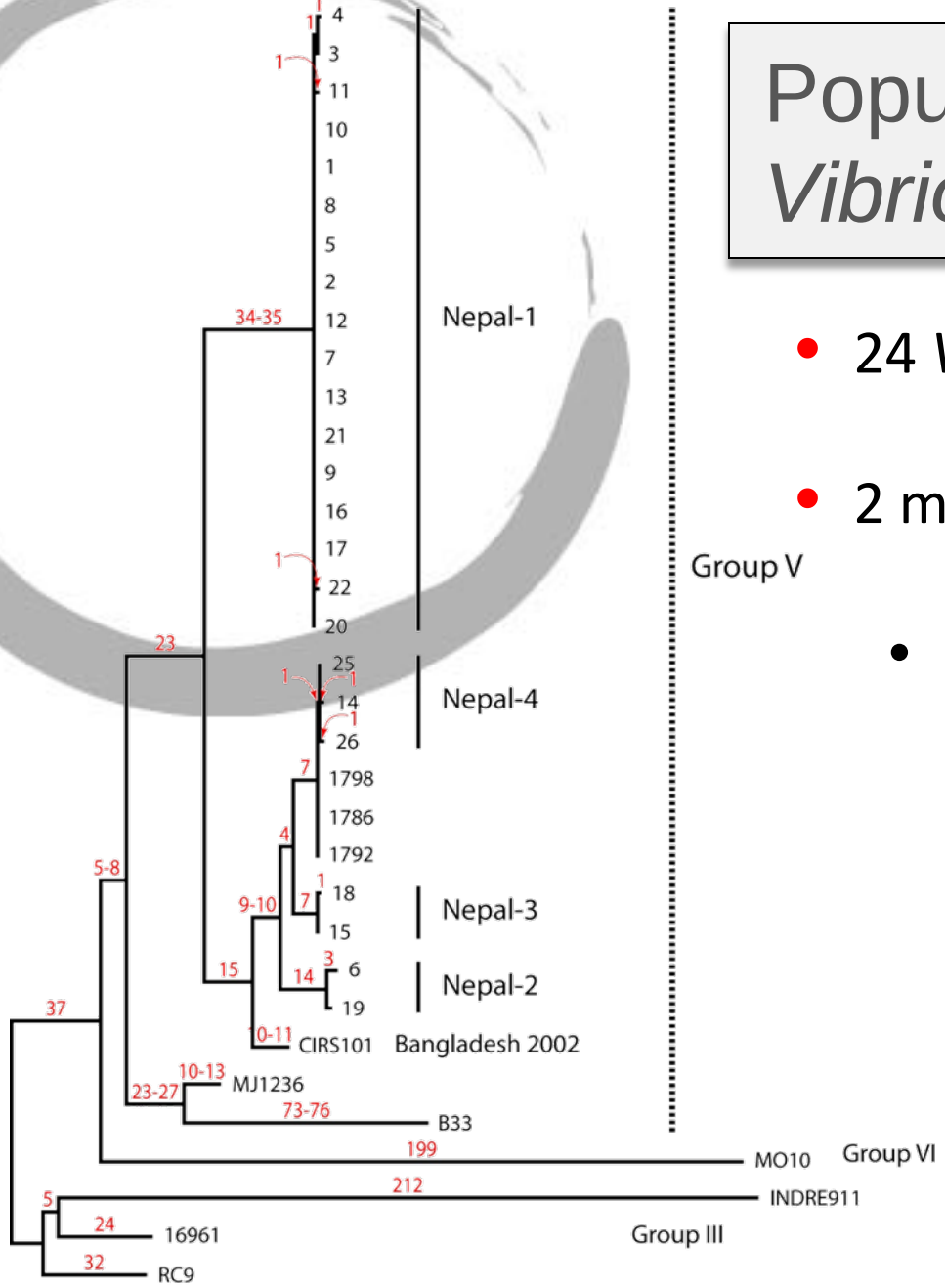
# Phylogeography of Nepalese *V. cholerae*




Rene S. Hendriksen, L. Price, D. Engelthaler, P. Keim, Geeta Shakya, Frank M. Aarestrup, & others
In press - *mBio*

TGEN NORTH

# Population Genetics of *Vibrio cholera* in Nepal




- 24 *V. cholera* Genomes
- 2 major Nepalese Clades
  - Nepal-1 has low diversity


Rene S. Hendriksen, L. Price, D.Engelthaler, P. Keim,
Geeta Shakya, Frank M. Aarestrup, & others
In press (*mBio*)

Danish Technical University - Copenhagen
TGen – Flagstaff, Arizona
National Public Health Laboratory, Katmandu


TGEN NORTH

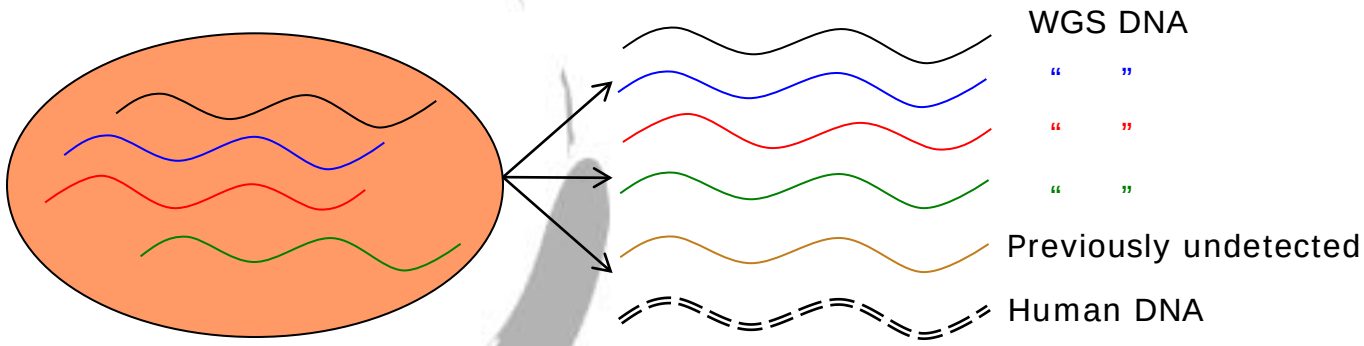
# Last Stop:
# Cystic Fibrosis and Infection

# CF and Respiratory Infections

- Chronic colonization and infection are the primary causes of morbidity and mortality
- Microbial culturing every 3 to 6 months
- The steady progression towards resistance is inevitable

TGEN NORTH

# Whole Metagenome Sequencing




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

TGEN NORTH

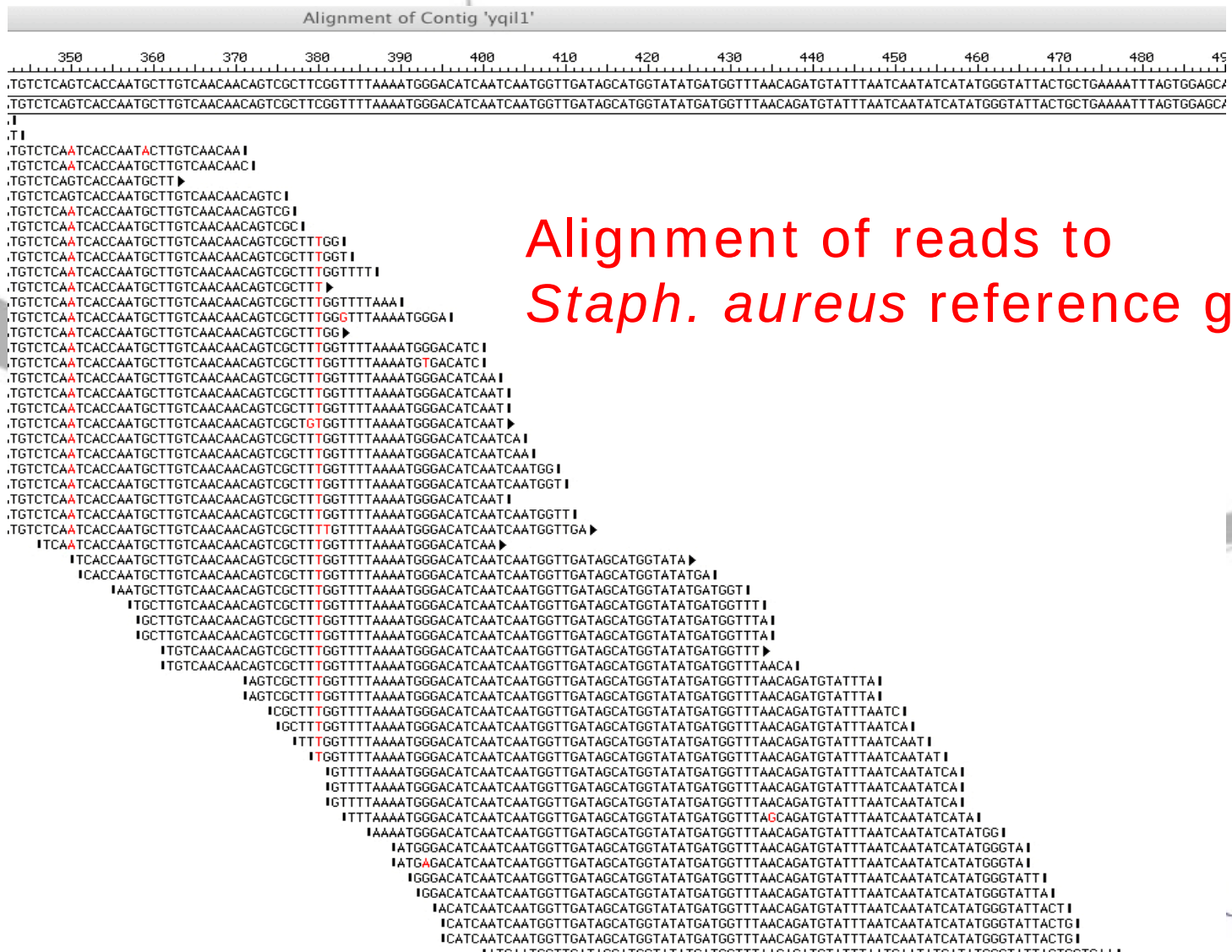
# CF Sequencing

- Single lane/sample in Illumina 75-100bp PE
- 3.5-6.0 Gbp total
- Alignment of all reads against know CF Threats

*S. aureus*
*P. aeruginosa*
*E. cloaca*
*K. pneumonia*
*H. influenzae*
*S. maltophilia*
*B. cepacia*
*A. baumanii*

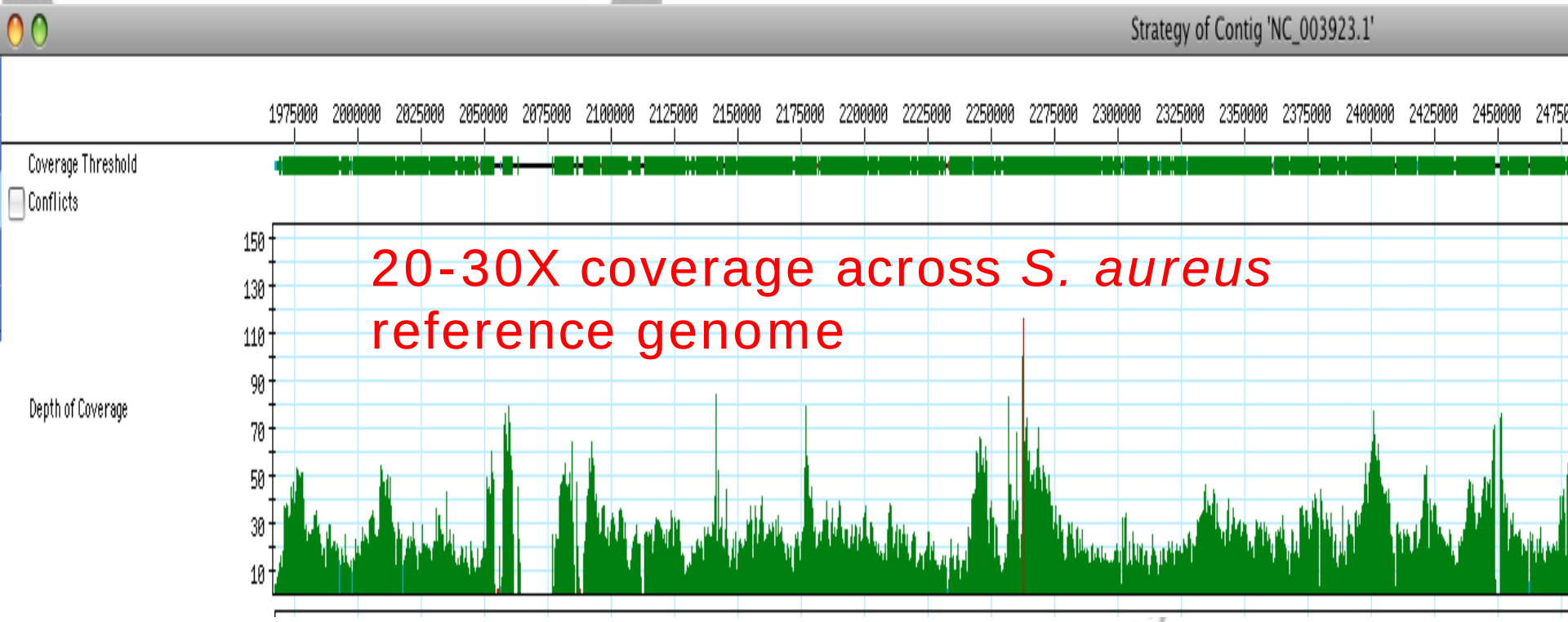
TGEN NORTH

# Detecting Pathogens in CF
# *Staphylococcus aureus*



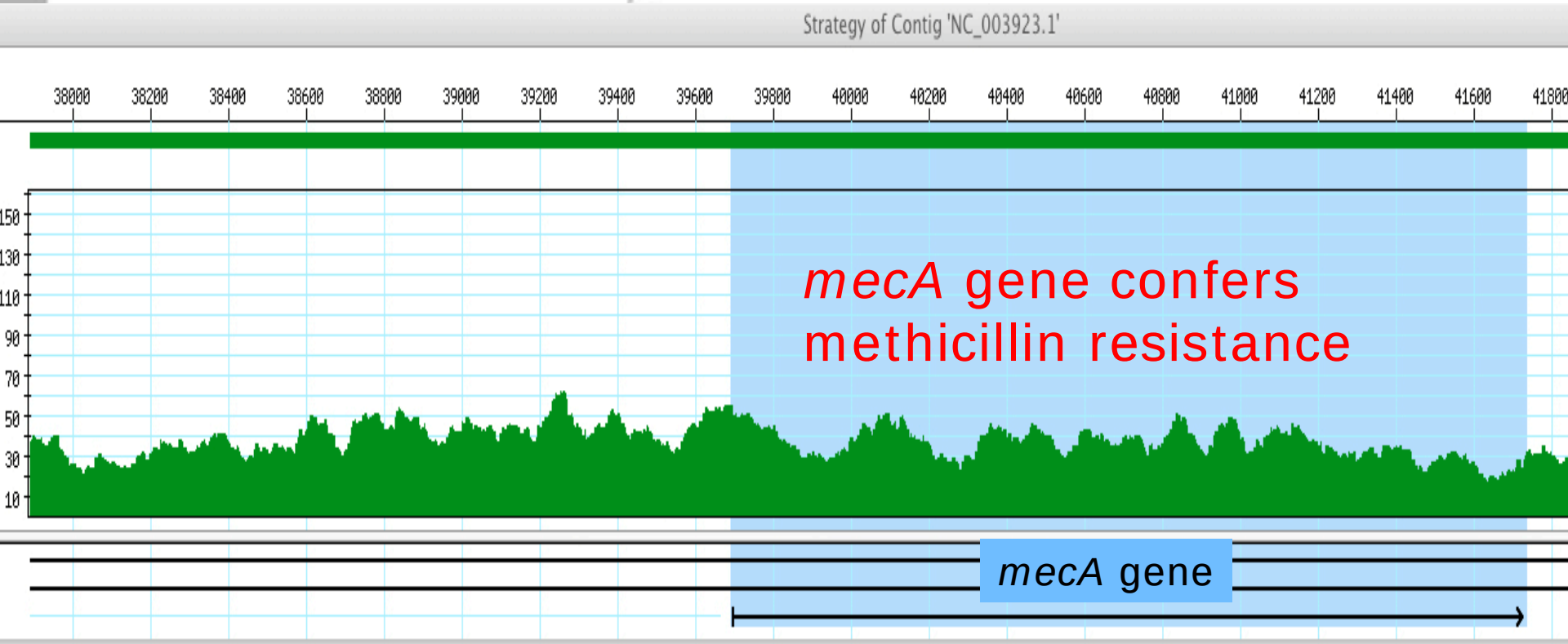

Alignment of reads to *Staph. aureus* reference genome

# Detecting Pathogens in CF *Staphylococcus aureus*

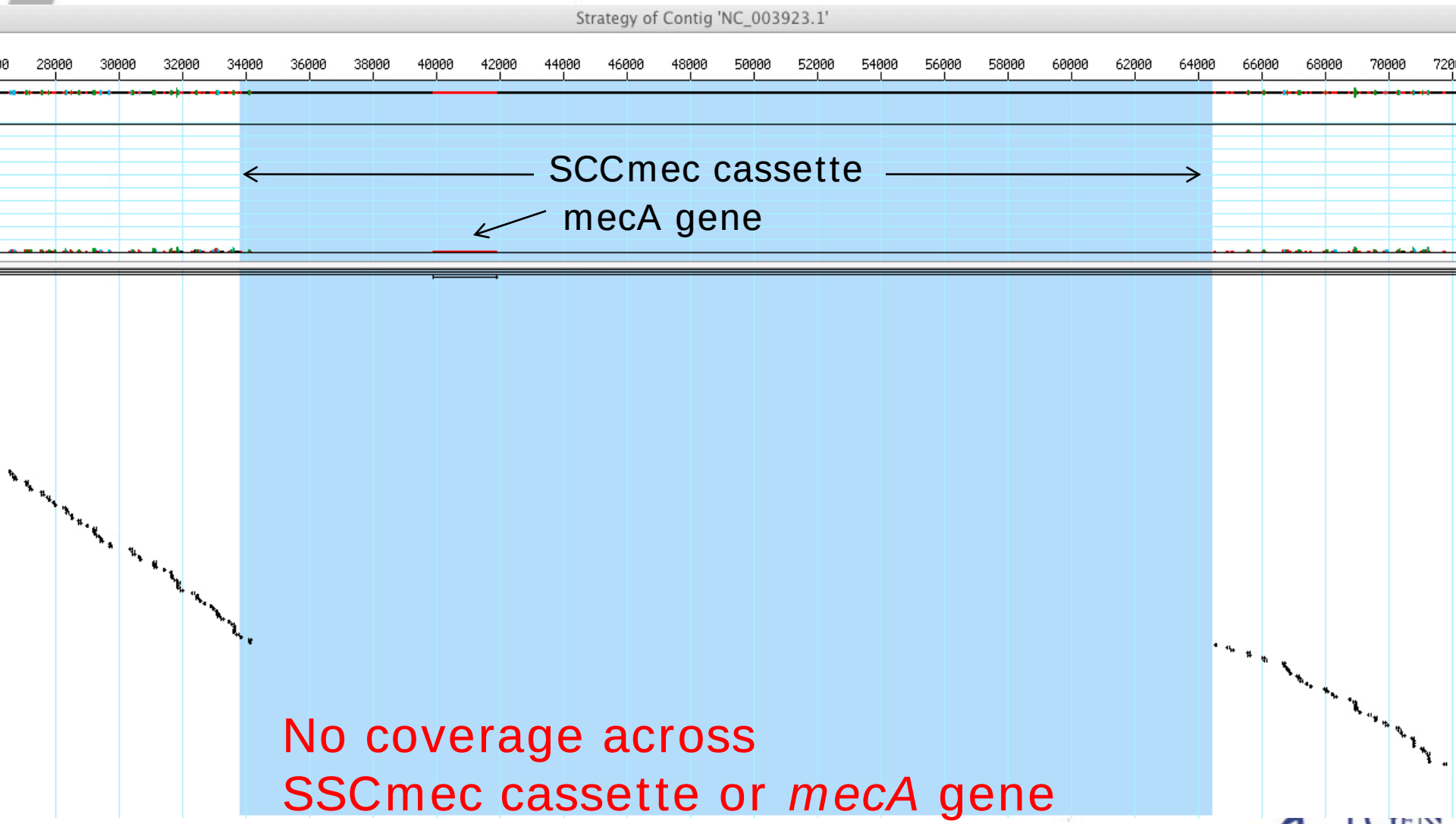



TGEN NORTH

# Antibiotic Resistance in CF MRSA

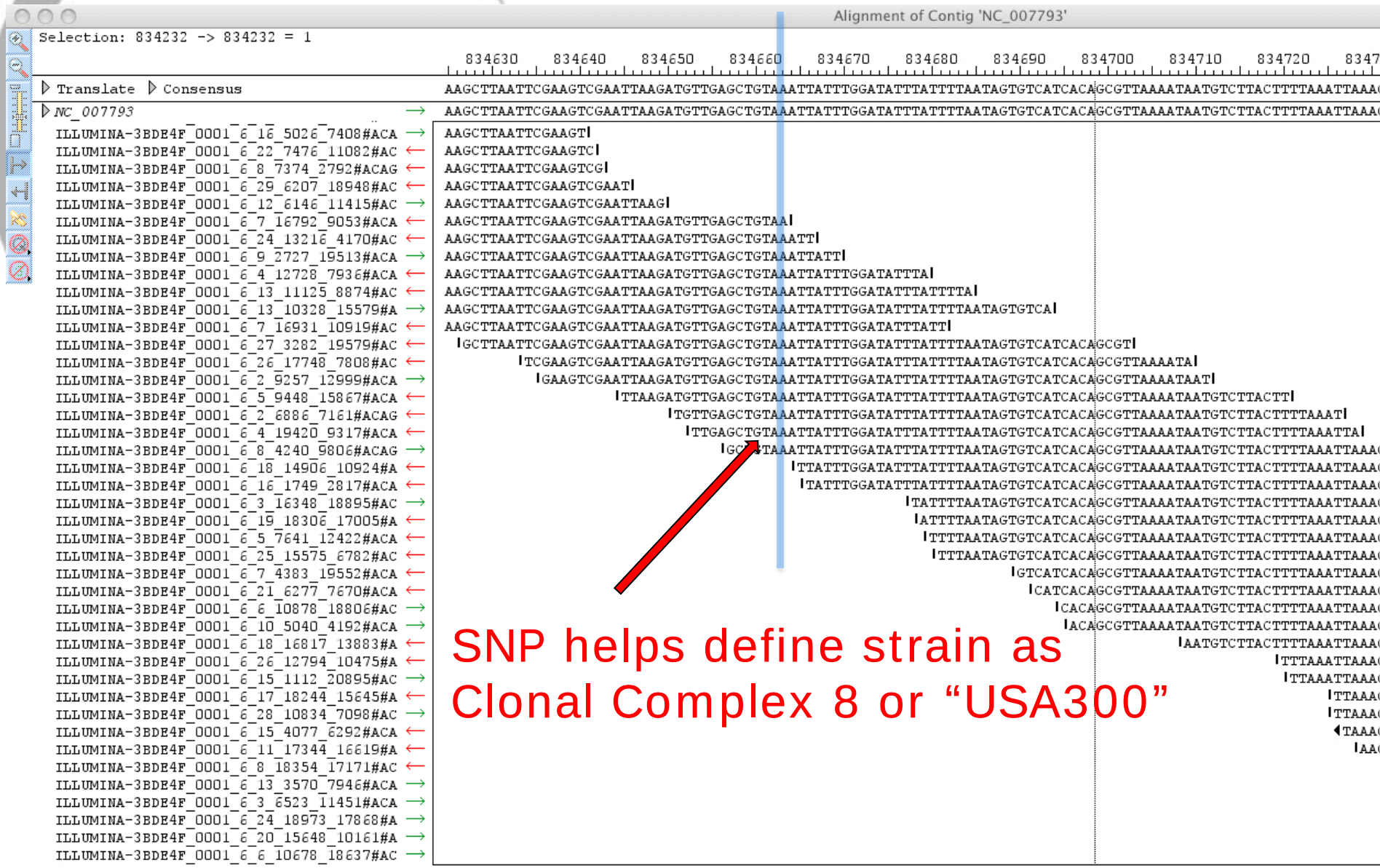


# Antibiotic Susceptibility in CF MSSA not MRSA

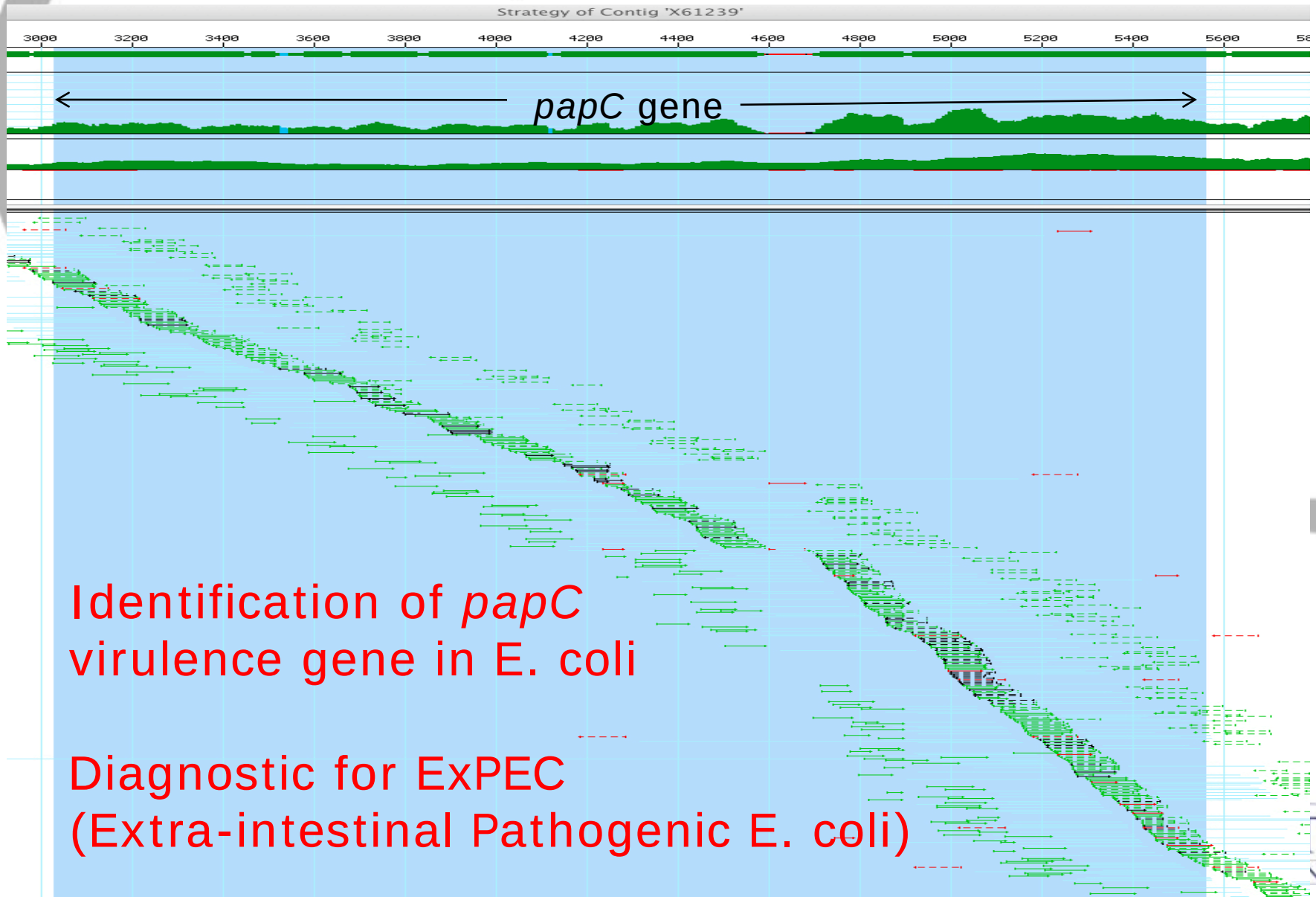



TGEN NORTH

# Genotyping a CF MRSA Sample



# Detecting Virulence in CF




Identification of *papC* virulence gene in E. coli

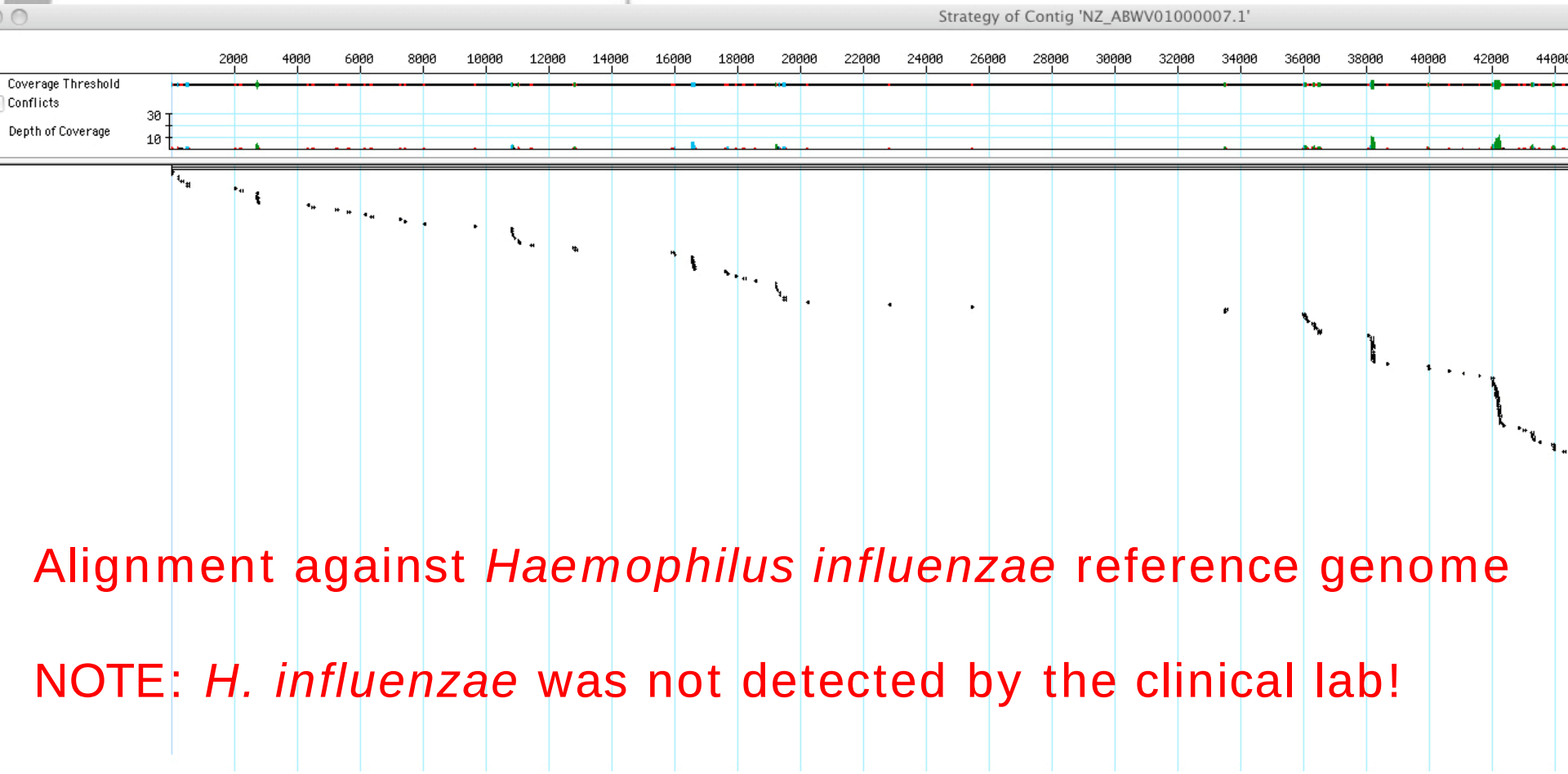
Diagnostic for ExPEC (Extra-intestinal Pathogenic E. coli)

# Polymicrobial infection in CF

Alignment to *Pseudomonas aeruginosa* reference genome

TGEN NORTH

# Polymicrobial Infection in CF (cont.)




Alignment against *Haemophilus influenzae* reference genome

NOTE: *H. influenzae* was not detected by the clinical lab!

# Final Thoughts

- Next gen is fundamentally transforming the way we look at pathogen genomics
  - And should change the way we look at clinical and public health analyses
- Advances in sequence technology are decreasing costs and increasing data
- Abundance of data will provide both the **pathway** and **barrier** to continued translation

TGEN NORTH

# Acknowledgements


- Entire TGen North Team
  - esp. Jim Schupp and John Gillece

    Paul Keim, Tricia O'Reilly, Elizabeth Driebe, Lance Price, Michael Bork, Maliha Aziz, Jolene Bowers, Jordan Buchhagen, Tania Contente, Becky Colman, Erin Kelley, Andrew Knadler, Arun Rawat, Richard Lester, David Smith, Carl Eberhard, Rose Rimza, Remy Hilsabeck, Jesse Trujillo, Zsuzsi Kovacs
- TGen HQ Team
  - esp. Carpten, Craig and Bioinfo Teams
- Arizona Department of Health Services
- CDC Mycology, Multiple State Health Depts
- Banner Health and Lab Sciences of AZ


TGen NORTH

# Thank You!

Sequence the World!