C G T A C G T A

A C G T A C G T

How bioinformatics helped crack the Amerithrax case

Adam M. Phillippy

May 7, 2019





The Forefront of Genomics[®]

Acknowledgements

- Disclaimer
 - All information presented here was taken from public sources
 - I was but a small, often unaware, cog in this investigation
 - I am speaking for myself, not the government

TIGR Microbial Genomics

• Jacques Ravel, Claire Fraser, Tim Read

TIGR Bioinformatics

- <u>Steven Salzberg</u>, <u>Mihai Pop</u>, Martin Shumway
- (presentation adapted from S. Salzberg)

• FBI, USAMRIID, NAU





The case

- September 29, 2001
 - Robert Stevens, American Media photo editor, falls ill
 - Diagnosed with anthrax on October 3, 2001
- By end of October
 - NBC News, NY Post, US Senate letters
 - 4 deaths (1 Florida, 2 DC, 1 NYC)
- In total

NHGRI

- At least 5 mailings
- 22 victims, 5 deaths (1 Conn.)
- 35 contaminated mailrooms
- 7 of 26 Capitol Hill buildings positive
- 7 years, 600,000+ FBI investigator hours





The letters



NIH FBI.gov

Searching for the Leahy letter





Bacillus anthracis





1876: First bacterium conclusively demonstrated to cause disease, Robert Koch

1881: First effective live bacterial vaccine, Louis Pasteur

1987: Soviet production capacity estimated at 5,000 tons per year



Art Friedlander, monkey spleen, inhalational anthrax

State of the art *B.a.* typing in 2000



Keim et al. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within Bacillus anthracis. J. Bacteriology (2000) NHGRI

NIH

Typing the attack strain

- Paul Keim's VNTR analysis
 - The attack strain was "Ames"
 - A higher resolution typing was impossible at the time
- Sequence Ames and the attack strain?
- From the FBI report:

"The tests needed to conduct such an analysis of *Bacillus anthracis* did not even exist in 2001. It was, as one world-renowned expert later called it, 'Star Wars stuff."





Microbial evolution simplified





Outbreak attribution





Sequence the genomes



Read *et al.* Comparative genome sequencing for discovery of novel polymorphisms in Bacillus anthracis. Science (2002)

NIH NHGRI

The future of outbreaks in 2002

Comparative Genome Sequencing for Discovery of Novel Polymorphisms in *Bacillus anthracis*

Timothy D. Read¹, Steven L. Salzberg¹, Mihai Pop¹, Martin Shumway¹, Lowell Umayam¹, Lingxia Jiang¹, Erik Holtzapple¹, Joseph D. Busch², Kimothy L. Smith², James M. Schupp², Daniel Solomon², Paul Keim², Claire M. Fraser^{1,*}

= Author Affiliations

¹ The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. ² Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA.

"These results demonstrate that genome-based analysis of microbial pathogens will provide a powerful new tool for investigation of infectious disease outbreaks."

A few chromosomal differences found between Florida and Ames Porton found, though likely caused by plasmid curing



Sequence the genomes





Sequence, assemble, compare





"Oh \$#!t" – Claire Fraser (Wired 2011)

Florida

Ames Ancestor





The critical clue: morphotypes





NIH

Biodefense solutions to protect our nation

Patricia Worsham and Terry Abshire

Rasko et al. Bacillus anthracis comparative genome analysis in support of the Amerithrax investigation. PNAS (2010) NHGRI

Morphotype hypothesis

Batches of B. anthracis

spores



Rasko *et al.* Bacillus anthracis comparative genome analysis in support of the Amerithrax investigation. *PNAS* (2010)

Sequence all the things!





Again!?



Morph A (Leahy letter)

Ames Ancestor





Is the assembly correct?

- $P(genome|reads) \propto P(reads|genome)$
 - Uniform coverage from random shearing
 - Reads are consistent with the assembly
 - Modulo some expected sequencing error and bias





Genome assembly forensics

- Foundation for GAGE, QUAST, etc.
- Reference based
 - Nucmer contig alignment
 - Annotate all variants

Reference free

- Mate-pairs and CE-statistic
- k-mer repeat analysis
- Depth of coverage
- Heterozygous SNPs
- Nucmer read mapping



Wild type Morph A

Morph A puzzle solved





Rasko et al. Bacillus anthracis comparative genome analysis in support of the Amerithrax investigation. PNAS (2010)

Wild type Morph A

Morph A assay developed





AA Porton LL10 PL10 DL10



duplications triplication

Rasko *et al.* Bacillus anthracis comparative genome analysis in support of the Amerithrax investigation. *PNAS* (2010)

From phenotype to genotype

- Morph A
 - Variable tandem expansion of 16S start
- Morph B
 - Intergenic SNP upstream of Spo0F
 - Phosphotransferase of the sporulation initiation phosphorelay
- Morph C/D
 - Variable truncations of sensor histidine kinase
 - Phosphorylation of Spo0F
- Morph E
 - 9 & 21bp in-frame deletions in response regulator
 - Dephosphorylation of Spo0F
- Morphs stable in subculture
- Not observed in parallel Ames Ancestor cultures
- Used as a metagenomic fingerprint of the source













Assay design and screen

- FBI screened >1,000 samples
 - 4 assays validated by FBI Laboratory
- Only 8 samples positive for all 4
 - Morphs A1, A3, D, & E
- All traced to flask RMR-1029
 - The "Dugway" spores
 - 35 different production runs
 - 164 L of spore prep distilled to 1 L
 - Held at USAMRIID in Frederick, MD



RMR-1029 flask



Biodefense solutions to protect our nation



Case closed?

- August 6, 2008. FBI officials explained why charges were about to be brought against anthrax researcher Dr. Bruce Ivins, who took his own life before those charges could be filed.
- "The Amerithrax Task Force conducted more than 9,100 interviews, executed more than 70 searches, and followed leads across six continents. Along the way, a brand new science was developed that ultimately led to the big break in the case."
- "That science—creating a DNA equivalent of a fingerprint—allowed investigators to pinpoint the origins of the anthrax."

http://www.justice.gov/archive/amerithrax/docs/amx-investigative-summary.pdf



Amerithrax was groundbreaking

Microbial forensics

- Trace the source of an outbreak via whole genomes
 - "MUMmer3" Kurtz, Phillippy, Delcher, Salzberg

Assembly validation

- Genome assemblies are hypotheses to be tested
 - "Genome assembly forensics" Phillippy, Schatz, Pop, Salzberg

Read mapping: 2nd mention ever?

- Whole-genome SNPs, Indels, and SVs
 - "Comparative genome assembly" Pop, Phillippy, Delcher, Salzberg

Rare variant analysis

• Everything is a metagenome if you go deep enough



Outbreak tracing in 2016





NIH

NHGRI



ebola.nextstrain.org



Nextstrain: real-time tracking of pathogen evolution. Hadfield et al. bioRxiv (2017)



Reconstructed history of the West African Ebola virus epidemic

https://youtu.be/j4Ut4krp8GQ

