The New York State Department of Health Experience with Whole Genome Sequencing: Promises, Paths and Realities

William Wolfgang
Wadsworth Center NYSDOH
2014 APHL

Promises-

- Improve detection of transmission paths and outbreak sources.
- Rapid and comprehensive detection of drug resistance and virulence factors from isolates.
- Rapid and comprehensive pathogen characterization from primary samples.
- Observe short term microbial evolution.
- Characterization of complex microbial communities.

- More rapid turn around times.
- Improved benefit to cost ratio.

Paths to Fulfillment - PHLs are the natural venue for investigating various paths

- Resources for retrospective and prospective analyses.
- Historical data for evaluation of new methods.
- QA/QC
- Topic expertise.
- Understanding what’s Practical.
- Need!!!!

Paths- Getting started

A need exists-

- Improve cluster resolution for *Salmonella* serovar Enteritidis.

In 2011 NGS looked promising-

- Affordable bench top sequencers.
- Papers were being published.
  - Retrospective studies of a *Salmonella* Montevideo outbreak.

Wadsworth buys an Ion Torrent PGM☺

Established a collaboration with:

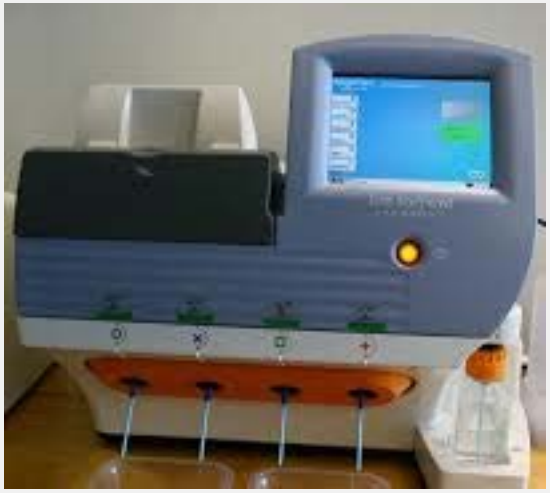
- **Martin Wiedmann and Henk Den Bakker at Cornell.**
- **Marc Allard, Eric Brown, Errol Strain, and Ruth Timme at the FDA.**
- **John Fontana, Stacey Kinney at the CT Dept. of Health**

Proof of principle retrospective study on a Cannoli associated outbreak.

- Sequenced **93 *S*. E. genomes**.

Wadsworth NGS Core Facility

Patrick Van Roey Ph. D. Director
Joe Wade Ph. D. Scientific Director



Wadsworth Ion torrent

Sequencing:
Mathew Shudt
Zhen Zhang
Melissa Leisner
Danielle Loranger
Charles MacGowan

Bioinformatics:
Mike Palumbo
Pascal LaPierre

Genometrakr MiSeq

NGS operations at the Wadsworth Core Facility

- 4 - MiSeq runs per week on two machines.
- 1 - Ion Torrent PGM run per week.
- 8 - Public Health Genomics Center projects.

- NGS strategies.
  - WGS
  - Amplicon
  - RNAseq

- Bioinformatics.
  - Custom built pipelines from publicly available software.
  - CLC Genomics workbench.

Public Health Genomics - projects

Project	NGS strategy	Sequencer
Salmonella surveillance	WGS	MiSeq
TB drug resistance and surveillance	WGS	MiSeq
Mosquito microbiome and West Nile Virus	16S RNA amplicons	MiSeq
Improved subtyping of HCV	RNAseq	Ion Torrent
Adenovirus characterization	WGS, denovo assembly	Ion Torrent MiSeq
C. Botulinum source tracking	WGS	MiSeq
Rabies virus intra-host evolution	WGS	Ion Torrent
Cystic Fibrosis	Amplicon	MiSeq DX

Real-time surveillance of *Salmonella* Enteritidis

Bacterial Molecular Typing Laboratory

William Wolfgang Ph. D.

Contact: wwolfgan@wadsworth.org

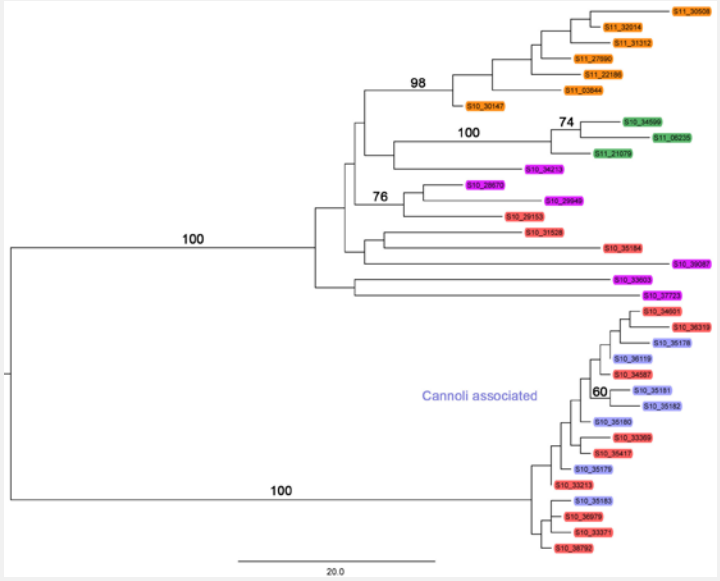
Dianna Schoonmaker-Boop

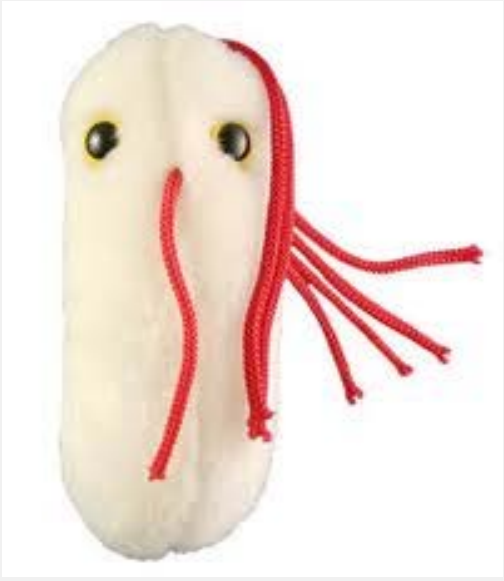
Deborah Baker

Lisa Thompson

Charles MacGowan

Melissa Leisner





Salmonella surveillance

Need: Improve outbreak detection of *Salmonella* Enteritidis.

Approach: Sequence all SE in as received in real-time to test feasibility.

- WGS
- MiSeq and Ion torrent
- snp based phylogenetics to detect clusters

Current findings: Feasible - Many more clusters are detected.

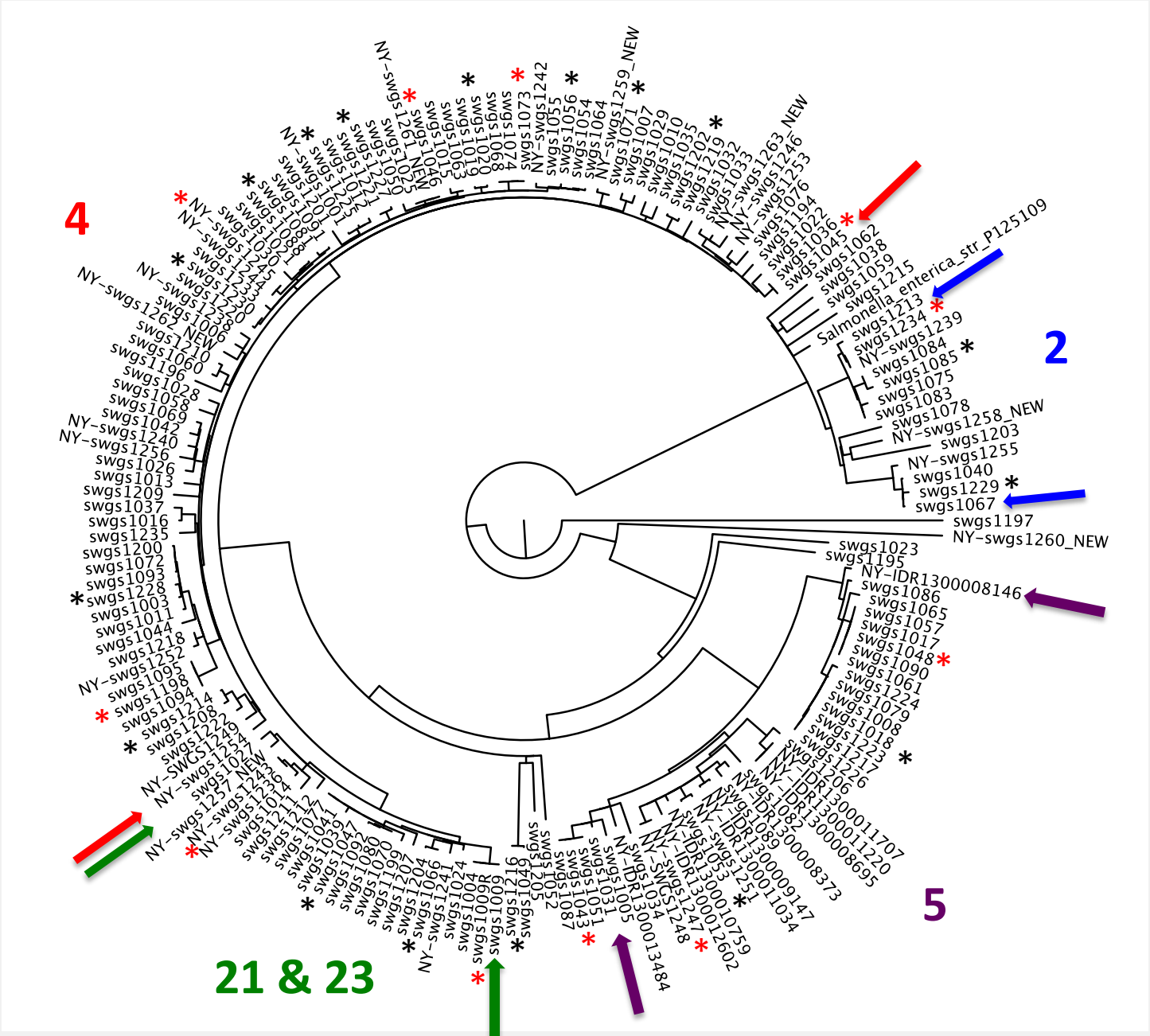
Outcomes: Improved outbreak source detection.

Challenges: Creating a network, metadata, backward compatibility, costs.

Time to implementation:

- Partial: 1 year
- Full: 2 to 4 years

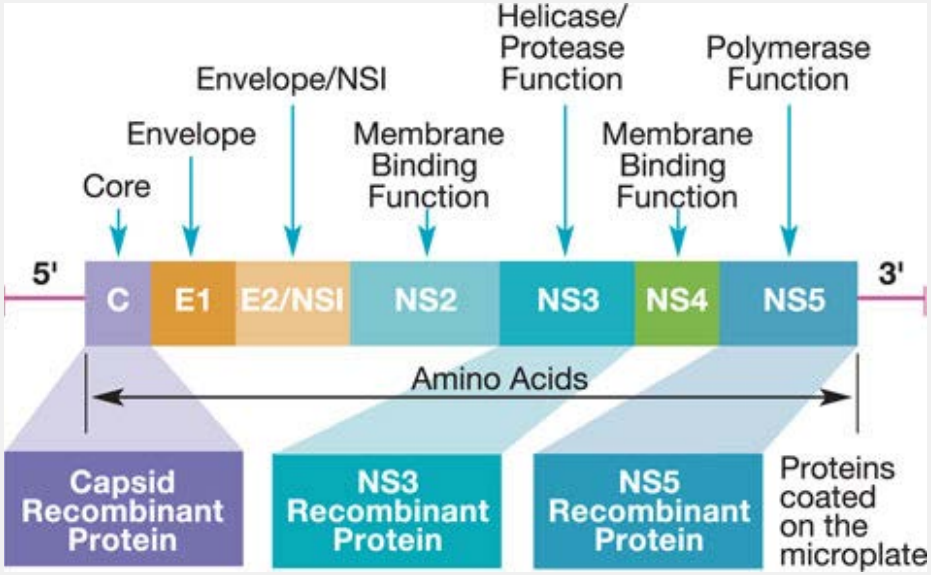
We see lots of clusters



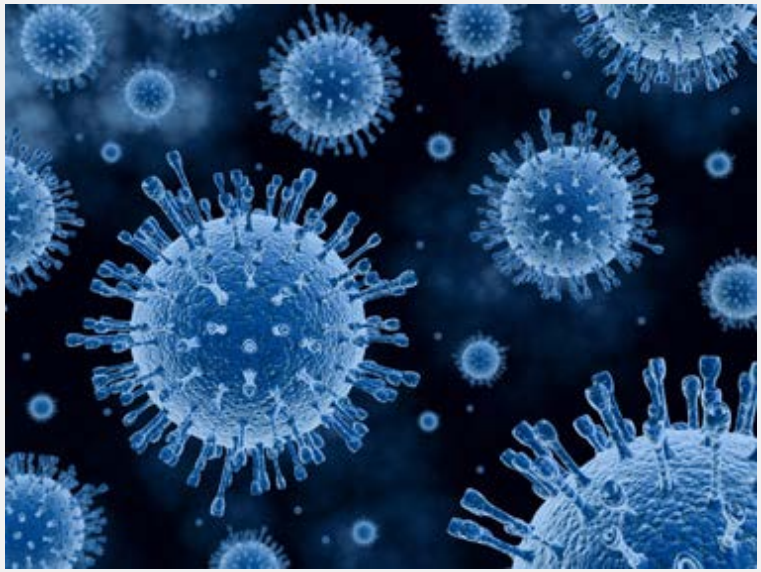
Improved subtyping and molecular epidemiology for Hepatitis C Virus (HCV)

Bloodborne Virus Laboratory

Monica Parker Ph.D.
Kathy Chou Ph.D
Contact: yxc10@notes.health.state.ny.us
Renee Hallack



BioRad



EMCDDA

Improved subtyping and molecular epidemiology for Hepatitis C Virus (HCV)

Need:

- Accurate identification of transmission source during outbreaks.
- Accurate genotype and subtype classification.
- Commercially available genotype assays provides limited subtype information.

Approach:

- Whole genome RNA-seq.
- Ion torrent PGM; 316 Chip.
- Reference based subtype classification and phylogenetic analyses.

Current findings: Pilot study establishes feasibility.

Outcomes:

- Comprehensive genomic data on circulating HCV strains in real time.
- Improved resolution of transmission pathways.
- Vaccine and drug development.

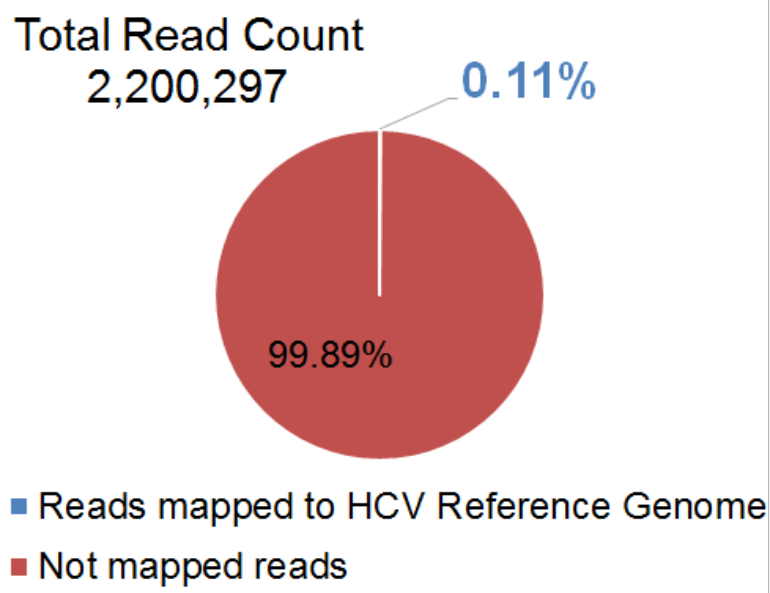
Challenges:

- Restricted to primary samples (plasma).
- Highly variable genome.

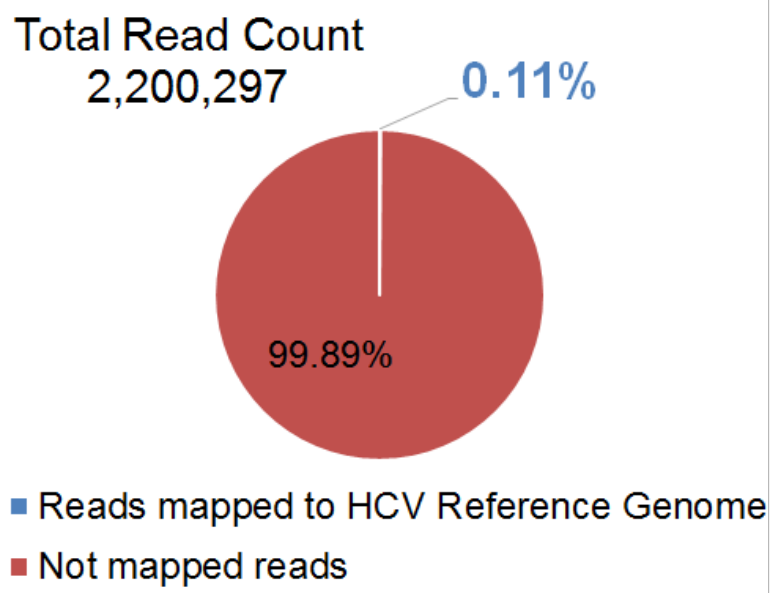
Time to implementation:

- Feasibility study: 1 year (in progress).
- Implementation: 2 – 4 years for outbreak investigations.

Only 0.11% of reads from patient serum map back to a reference HCV genome



Alignment to reference genotypes reveals patient was infected with Type 1b HCV



Reference Genome (~9.6kb)	Total Read Count	Average Coverage	Fraction of Reference Covered
1a	1228	22.69	0.69
1b	**2504**	**31.97**	**0.99**
2a	419	3.37	0.27
2b	246	2.79	0.20
3a	518	4.71	0.31
4a	340	4.64	0.41
5a	466	4.22	0.32
6a	299	3.77	0.35

CLC Bio (CLC Genomic Workbench 7.0.3)

- "Average Coverage" and "Fraction of Reference Covered," indicate this patient is likely to be infected by HCV 1b subtype.
- **Sanger sequencing corroborates this finding.**

Evaluation of NGS as a diagnostic tool for *Mycobacterium tuberculosis*

Bacterial Disease Laboratory

Kim Musser Ph. D.

Contact: musser@wadsworth.org

Vincent Escuyer Ph. D.

Tanya Halse

Stop TB UK

NGS for *Mycobacterium tuberculosis* for drug resistance

Need: 800-1000 TB cases each year in NYS; ~ 10 % are drug resistant.

- Currently: 12 real-time PCR and pyrosequencing reactions assess drug resistance.
- All require reagents, technical time and supervisor review and reporting.

Approach: Evaluate the cost and benefit of a single NGS assay.

- WGS.
- Amplicon?
- MiSeq.

Current findings: ~**43** genomes are complete.

- NGS data correlate or improve information on these strains of MTB.
- Testing of primary samples is in progress.

Outcomes:

- More cost effective and decreased hand on time.
- Sensitive and specific detection as well as improved resistance prediction.

Challenges:

- Needs to work on primary specimens.

NGS for strain typing Beijing family MTB to improve molecular epidemiology

Need: Current typing methods work poorly for the Beijing strain of MTB.

Approach: Test WGS as a replacement for the existing multiple strain typing assays.

- WGS
- MiSeq
- Reference based SNP phylogenetic analysis.

Current findings: Promising tool, improved resolution is observed.

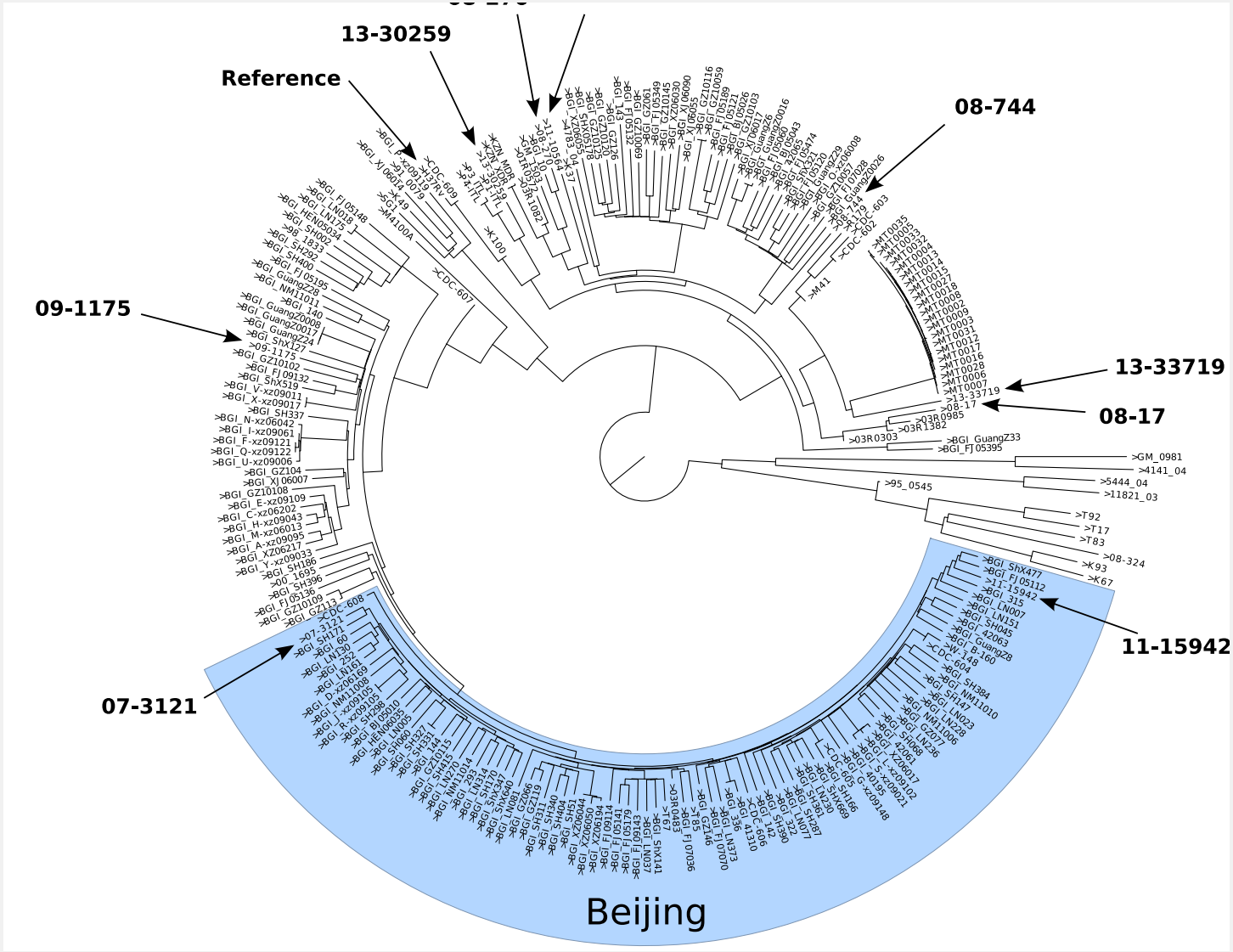
Outcomes: Improve surveillance for Beijing and other closely related MTB strains.

Challenges: DNA extraction, backward compatibility may be needed.

Time to implementation (for both):

- Partial: 1 year
- Full: 2-3 years if successful

MTB TREE


13-30259
Reference
08-744
09-1175
13-33719
08-17
11-15942
07-3121
Beijing

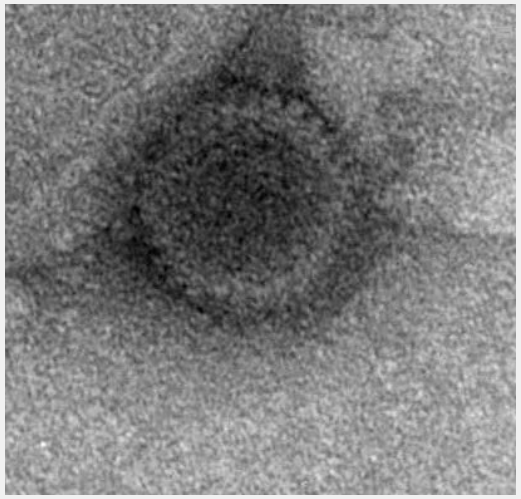
Genetic characterization of human Adenovirus (HAdV)

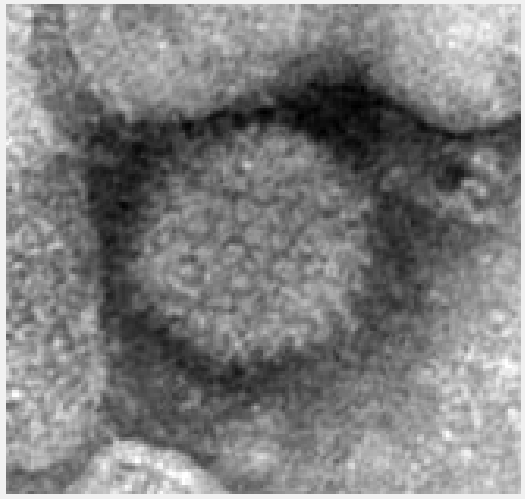
Virology Laboratory

Kirsten St. George Ph. D.

Daryl Lamson

Contact: dml20@health.state.ny.us





HAdV

Need: Comprehensive genetic characterization for outbreak investigations, virulence factor detection, and vaccine development.

Approach: Propagate to high titer and purify HAdV DNA.

- WGS.
- MiSeq or Ion Torrent.
- *De novo* assembly, phylogenetic analysis and annotation.

Current findings: Annotated 50 genomes.

- Identified and analyzed a novel intertypic recombinant group D HAdV.
- Aided in resolving the investigation of HAdV8 outbreaks.

Outcomes: High resolution relatedness analysis for outbreak investigations and detailed genetic characterization of strains.

Challenges: Bioinformatics, data storage and computing power.

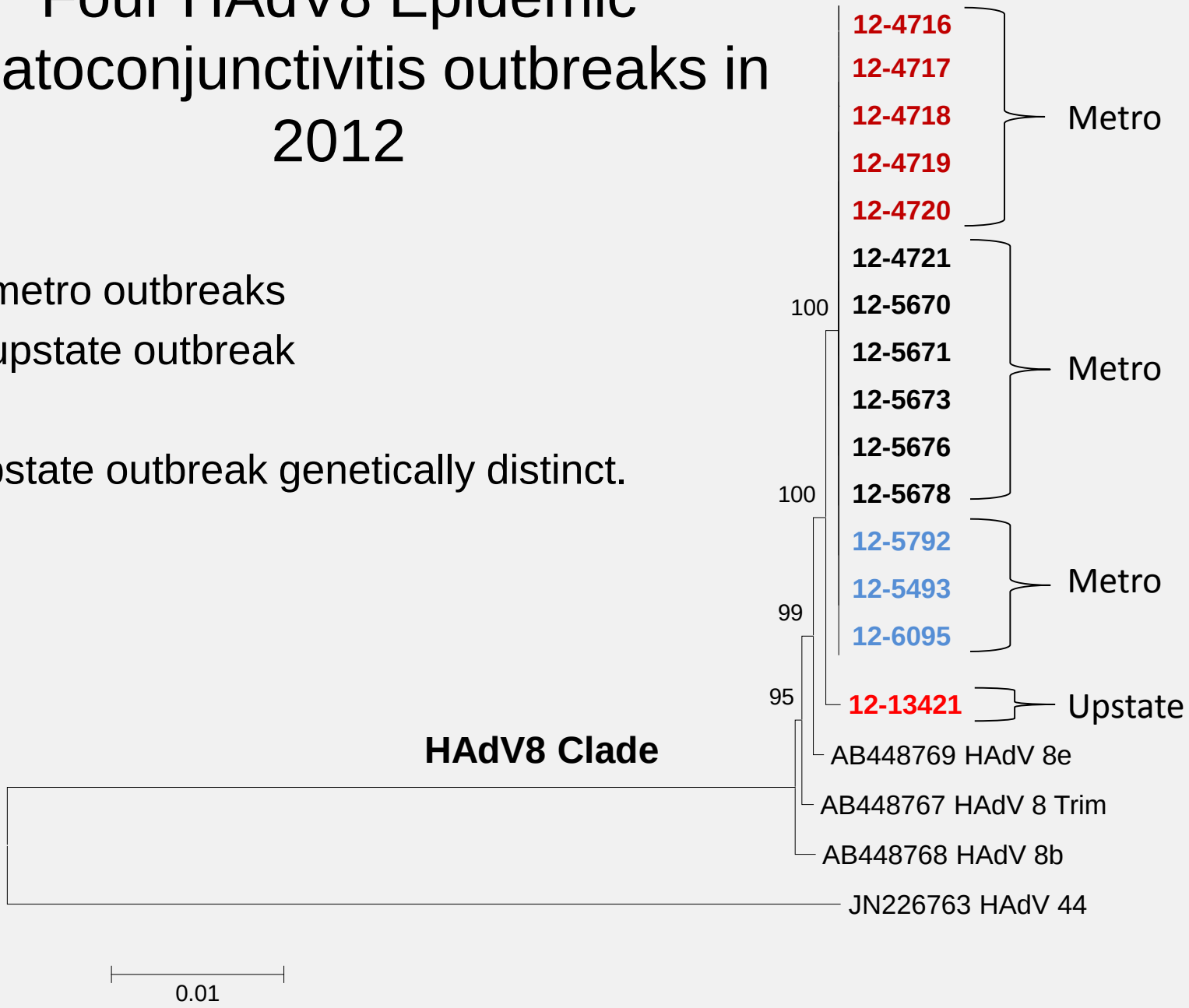
Time to implementation:

- Partial: 3 months
- Full: 1 year

Four HAdV8 Epidemic Keratoconjunctivitis outbreaks in 2012

- 3 metro outbreaks
- 1 upstate outbreak

- Upstate outbreak genetically distinct.



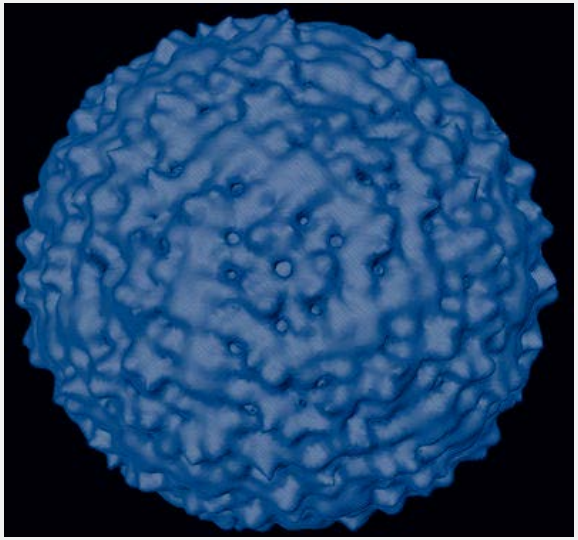
Mosquito microbiome and *West Nile virus*

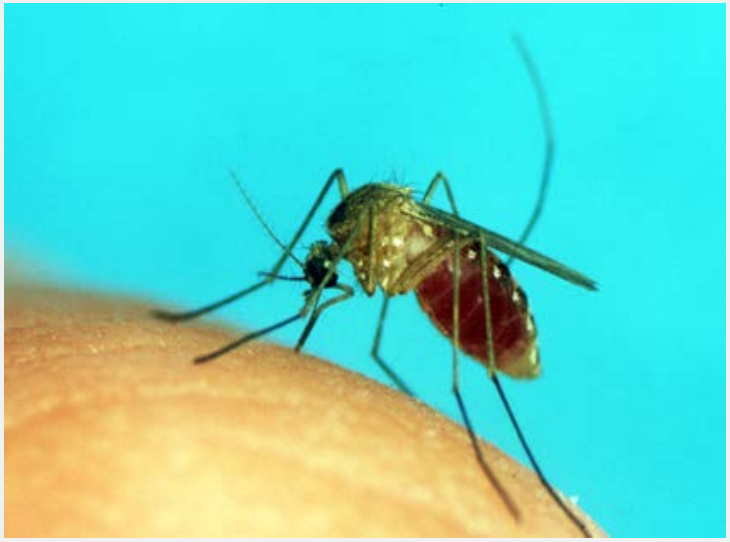
Arbovirology Laboratory

Alex Ciota Ph. D.

Contact: atc04@health.state.ny.us

Laura Kramer Ph. D.





Mosquito microbiome and *West Nile virus*

Need: Identify microbial signatures associated with WNV infection and susceptibility in *Culex* mosquitoes.

Approach: 16s sequencing following bloodfeeding with or without WNV in colonized mosquitoes.

- MiSeq
- Qimme sequence analysis

Current findings: WNV exposure significantly alters bacterial populations in *Culex* mosquitoes.

Outcomes: Identify specific genera and/or signatures associated with infection and resistance in field mosquitoes.

Challenges: Individual variability in the field.

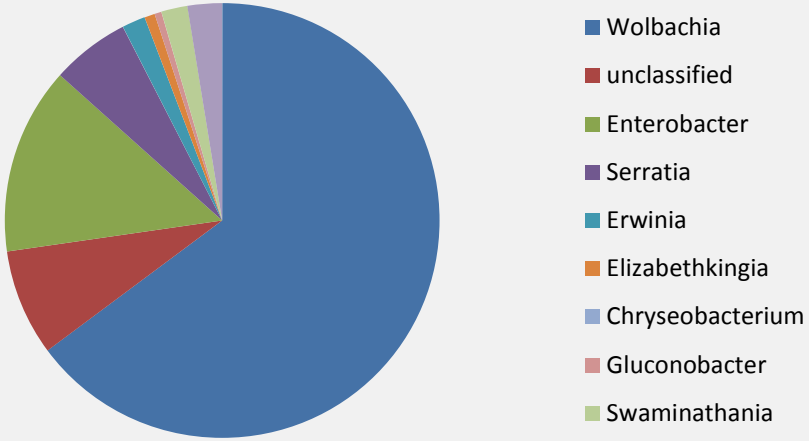
- Difficulty in assessing casual relationships.

Time to implementation:

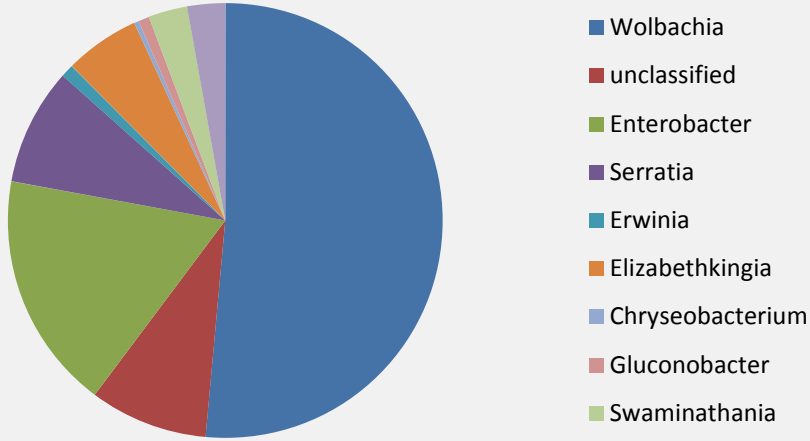
- Partial: 1 year
- Full: 2 to 4 years

Microbial signatures are altered with WNV exposure and infection status in mosquitoes

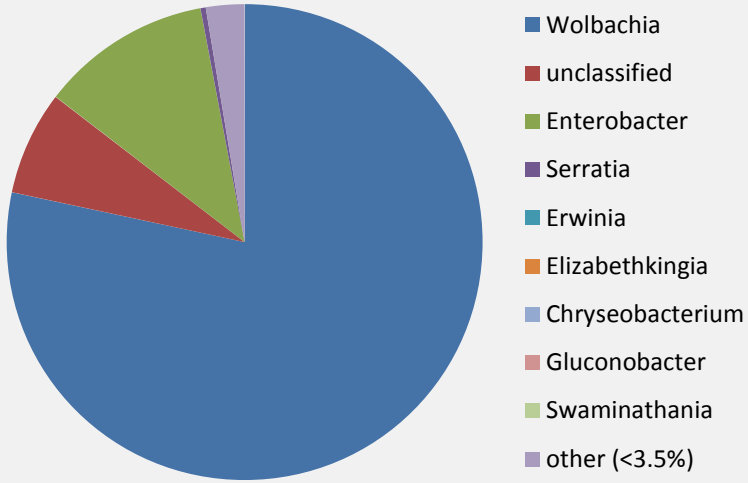
exposed-WNV neg



exposed-WNV pos



unexposed



Challenges for NGS

Increasing amounts of data.

Transitioning.

- Integration with previous methodologies.
- Integrating surveillance at a national level.

As sequencing technology and bioinformatics evolve:

- Need to maintain backward compatibility

Metadata, how much should be public.

- In real time?
- What elements?

Improving efficiency.

QA/QC.

Paying.

Summary

- Promises are real.
- So get a machine.
- The more paths that are explored, the more rapidly the technology will mature.
- Bioinformatics second.