Superbugs, Emerging Pathogens
and the
Evolution of Clinical Genomics

David Alexander

OAHPP Research Rounds
November - 2010

Superbugs, Emerging Pathogens and the Evolution of Clinical Genomics

The OAHPP and Research at the Public Health Laboratories

DNA Sequencing: From Gels to Genomes

Surveillance: Molecular Epidemiology of Pandemic Influenza A

Testing for Emerging pathogens:
- *Streptococcus pseudopneumoniae*
- *Mycobacterium xenopi*

Pathogen Discovery
- NDM-1-producing Enterobacteria

Future Applications
- Microbiomes and Metagenomics

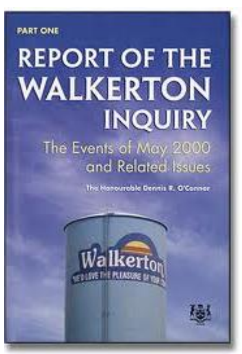
Summary and Questions

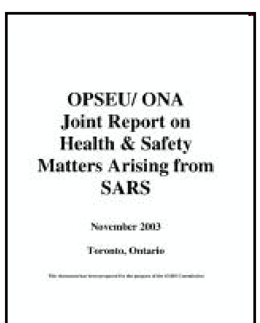
Infectious Disease in Ontario

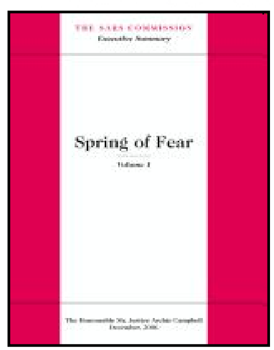
Ontario in the '00s

2000: **'Walkerton' *E.coli* O157:H7 Outbreak** → 2500 ill, 7 deaths
2003: **SARS** → 251 confirmed cases, 44 deaths
2005: **'Seven Oaks' Legionella Outbreak** → 127 cases, 21 deaths
2008: **'Maple Leaf' Listeriosis Outbreak** → 22/38 cases, 9/15 deaths
2009: **Swine Flu/H1N1 Influenza Pandemic** → 7000 confirmed cases, 128 deaths

These crises exposed cracks in Ontario's Public Health System:




OPSEU/ ONA Joint Report on Health & Safety Matters Arising from SARS
November 2003
Toronto, Ontario



- Funding
- Surveillance
- Testing Capacity
- Public Health Research
- Infection Control
- Communication
- Funding

Bill 171 and the OAHPP

Owing to the inertia, apathy and the lack of legislative powers conferred on health authorities but little can be said regarding [these deaths] except to again emphasize the great and ever pressing claim made upon municipal and county authorities for some early and active move on their part to make some in the direction of providing help and assistance for those suffering from [infectious disease] in some one or other of its many forms.

Annual Report of the Ontario Dept of Public Health - 1907

Health Systems Improvements Act, 2007 (Bill 171)

↓

Ontario Agency for Health Protection and Promotion

Bill 171 and the OAHPP

Bill 171: Schedule K ⇨ Objects of the OAHPP

(a) to provide scientific and technical advice and support to the health care system and the Government of Ontario in order to protect and promote the health of Ontarians and reduce health inequities;

(b) to develop, disseminate and advance public health knowledge, best practices, and research in the areas of population health assessment, infectious diseases, health promotion, chronic diseases, injury prevention, and environmental health;

(d) to develop, collect, use, analyse and disclose data, including population health, surveillance and epidemiological data, across sectors, including human health, environmental, animal, agricultural, education, community and social services and housing sectors, in a manner that informs and enhances healthy public policy

(e) to undertake, promote and co-ordinate public health research in co-operation with academic and research experts as well as the community;

(i) to undertake research related to evaluating the modes of transmission of febrile respiratory illnesses and the risk to health workers;

The OAHPP and Research at the Public Health Laboratories

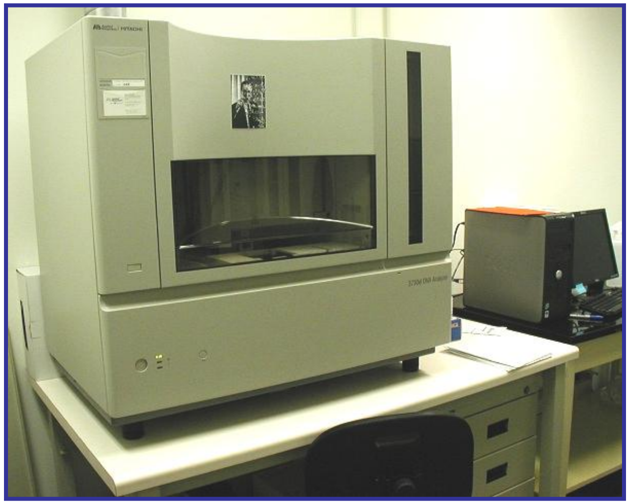
Advanced Diagnostics and Research Group

- Revived in 2006 ⇨ 'Molecular Research'
- Current Staffing
  - 5 Medical Microbiologists (MD's with ID training and many responsibilities)
  - 1 Clinical Microbiologist (PhD with ID training)
  - 5 Molecular Microbiologists (PhD trained scientists)
  - 7 Research Technicians (MSc/PhD trained)
  - <10 Student Researchers (undergrad → PhD trained)

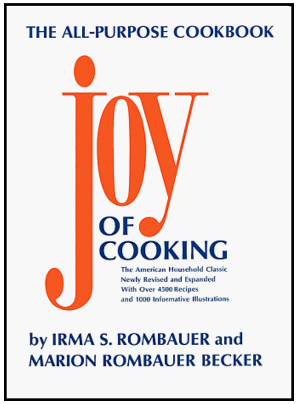
- Emphasis is on 'Clinical Research' including
  - 'Knowledge Generation' activities
  - Surveillance and Epidemiology activities
  - Test Development, Evaluation, Validation
  - Clinical Testing
- Diverse areas of study ⇨ Virology, Bacteriology, Mycology, Parasitology, Genomics!

That was then…

DNA Core Lab



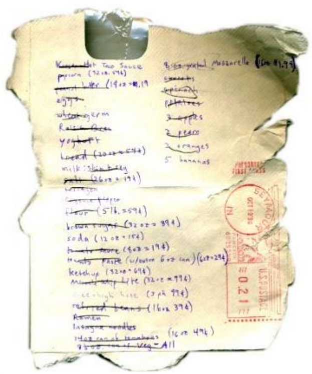
Genomics 101



⇨ **Cookbook:** Recipes *plus* Cooking Tips, Menus, Food History…
Durable (but can accumulate stains, changes over time)

Genomics ⇨ DNA

The collection of hereditary data in an organism
Durable data storage (stable - but not static)



⇨ **Shopping List:** Essential Recipe Information
Temporary (destroyed in pockets/purse)

Transcriptomics ⇨ RNA

Temporary data storage

⇨ **Food:** End Product of the Recipe - NOT paper and text

Proteomics ⇨ Protein

Antigens and Enzymes

DNA Sequencing

DNA ⇨ Information Storage Medium

Sequencing ⇨ Collect, Read and Use Information

Identification: What is the title of the cookbook?

⇨ 16S rRNA gene sequencing for identification of bacterial species

Classification: What kind of recipes are in the cookbook?

⇨ multilocus sequence analysis for 'genotyping' bacterial strains

Comparison: How does this cookbook compare to my other cookbooks?
Are the ingredients and recipes similar or different?

⇨ SNP detection = look for 'typos' in the cookbook

⇨ In/Del detection = look for recipes that have been inserted/deleted

DNA Sequencing: From Gels to Genomes

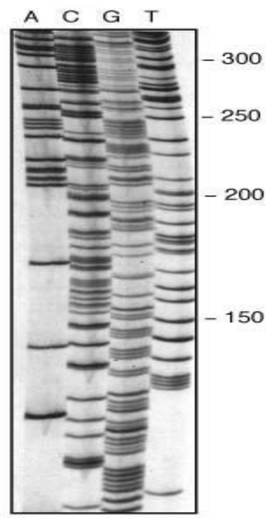
DNA Sequencing: 1975

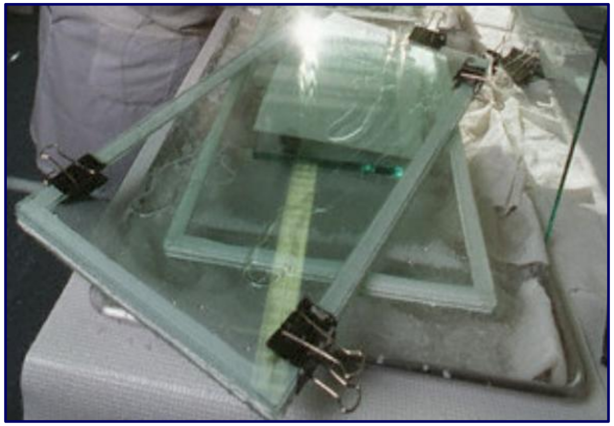
Allan Maxam and **Walter Gilbert** → 'chemical sequencing'

Frederick Sanger → 'chain terminator/dideoxy sequencing'

- Technically demanding → radiolabelled nucleotides, large sequencing plates
- Slow → few hundred bases/run (if you're really good!)
- Expensive → $1/base

but Sanger did complete the first DNA genome sequence: phiX174 (5386 bases)





DNA Sequencing: From Gels to Genomes

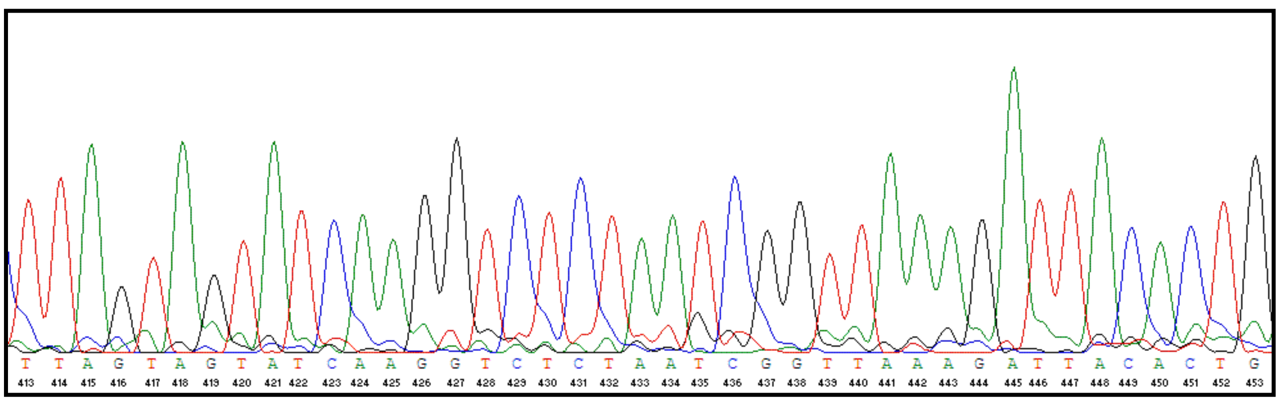
Automated Capillary Electrophoresis: 1985 - present

Modified version of Sanger sequencing → uses fluorescent ddNTPs: **Not Radioactive!**

First automated sequencer developed by **Leroy Hood** (1986)

First commercial sequencer: Applied Biosystems ABI 370 (1987)

- Technically simple → sequencing kits!
- Rapid → 500+ bases/sample and up to 384 samples/run
- Inexpensive → less than $0.01/base



DNA Sequencing: From Gels to Genomes

Next-Generation Sequencing: 2005 - present

Illumina (Solexa) and **SOLiD** sequencing: 'short read' technologies

Roche/454: 'long read' pyrosequencing

- Technically demanding and labour intensive → 3 days prep work/run
- Massive Throughput → 400,000,000+ bases/run
- Very Inexpensive → less than $0.00002/base

Bioinformatics is the bottleneck!

The First 25 Genomes: 1995 - 2000

ORGANISM	JOURNAL	DATE
Haemophilus influenzae Rd (KW20)	Science 269,496-512	1995-07-28
Mycoplasma genitalium G-37	Science 270,397-403	1995-10-20
Methanocaldococcus jannaschii DSM 2661	Science 273,1058-1073	1996-09-28
Mycoplasma pneumoniae M129	Nucleic Acids Research 24, 4420-4449	1996-11-15
Saccharomyces cerevisiae S288C	Nature 387,5-105	1997-05-29
Helicobacter pylori 26695	Nature 388,539-547	1997-08-07
Escherichia coli K-12, MG1655	Science 277,1453-1474	1997-09-05
Methanothermobacter thermoautotrophicus	J Bacteriology 179,7135-7155	1997-11-10
Bacillus subtilis subtilis 168	Nature 390,249-256	1997-11-20
Archaeoglobus fulgidus VC-16	Nature 390,364-370	1997-11-27
Borrelia burgdorferi B31	Nature 390,580-586	1997-12-11
Aquifex aeolicus VF5	Nature 392,353-358	1998-03-26
Pyrococcus horikoshii OT3	DNA Research 5,55-76	1998-04-30
Mycobacterium tuberculosis H37Rv	Nature 393,537-544	1998-06-11
Treponema pallidum pallidum Nichols	Science 281,375-388	1998-07-17
Chlamydia trachomatis D/UW-3/CX (sv D)	Science 282,754-759	1998-10-23
Rickettsia prowazekii Madrid E	Nature 396,133-140	1998-11-12
Caenorhabditis elegans	Science 282,2012-2018	1998-12-11
Helicobacter pylori J99	Nature 397,176-180	1999-01-14
Chlamydophila pneumoniae CWL029	Nature Genetics 21,385-389	1999-04-10
Aeropyrum pernix K1	DNA Research 6,83-101	1999-04-30
Thermotoga maritima MSB8	Nature 399,323-329	1999-05-27
Deinococcus radiodurans USUHS (R1)	Science 286,1571-1577	1999-11-19
Campylobacter jejuni jejuni NCTC 11168	Nature 403,665-668	2000-02-10
Neisseria meningitidis MC58	Science 287,1809-1815	2000-03-10

Source: Genomes OnLine Database

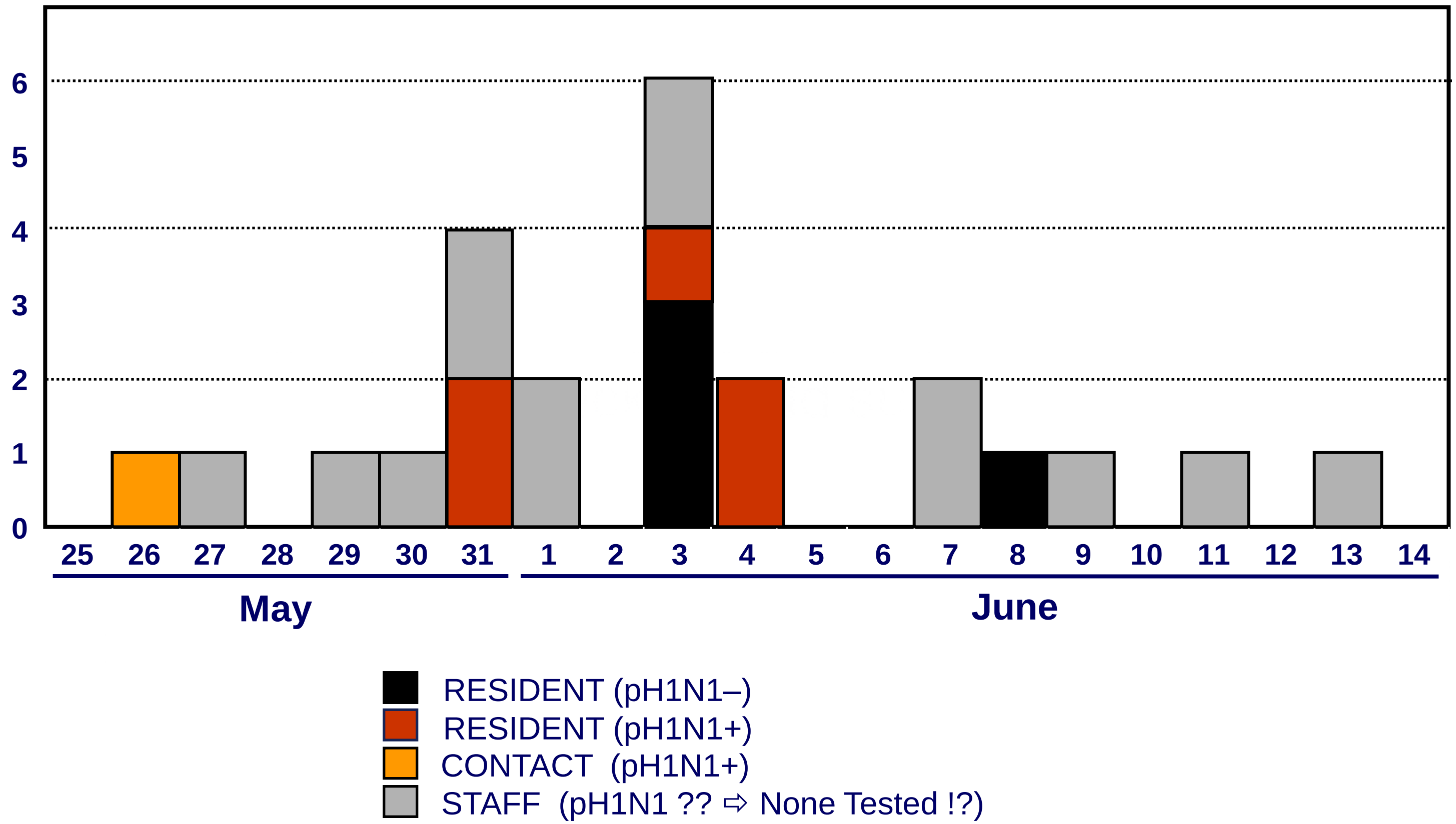
The Latest 25 Genomes: August/September 2010

ORGANISM	JOURNAL	DATE
Saccharomyces cerevisiae S288c	Unpublished	2010-09
Ferrimonas balearica PAT, DSM 9799	Unpublished	2010-09-24
Vulcanisaeta distributa IC-017, DSM 14429	Unpublished	2010-09-22
Halomonas elongata DSM 2581	**Environmental Microbiology** (in press)	2010-09-17
Methanoplanus petrolearius DSM 11571	Unpublished	2010-09-17
Sulfurimonas autotrophica OK10, DSM 16294	Unpublished	2010-09-15
Spirochaeta thermophila DSM 6192	Unpublished	2010-09-10
Dickeya dadantii 3937	Unpublished	2010-09-10
Clostridium cellulovorans 743B, ATCC 35296	**J Bacteriology** 192, 901-902	2010-09-02
Bacillus subtilis spizizenii W23	Unpublished	2010-08-31
Edwardsiella tarda FL6-60	Unpublished	2010-08-26
Mycoplasma hyorhinis HUB-1	Unpublished	2010-08-23
Candidatus Regiella insecticola LSR1	**Environmental Microbiology** (in press)	2010-08-19
Helicobacter pylori 908	Unpublished	2010-08-16
Methanothermobacter marburgensis Marburg	Unpublished	2010-08-11
Caldicellulosiruptor obsidiansis OB47	Unpublished	2010-08-11
Butyrivibrio proteoclasticum B316	**PLoS ONE** 5(8):e11942	2010-08-10
Staphylococcus aureus aureus JKD6159	Unpublished	2010-08-09
Acidilobus saccharovorans 345-15	**Appl Environ Microbiology** 76, 5652-5657	2010-08-06
Corynebacterium pseudotuberculosis 1002	Unpublished	2010-08-06
Oxalobacteraceae symbiotic bacterium CARI	Unpublished	2010-08-05
Corynebacterium pseudotuberculosis C231	Unpublished	2010-08-05
Acetohalobium arabaticum Z-7288, DSM 5501	**SIGS** 3(1): 57-65	2010-08-05
Clostridium saccharolyticum WM1, DSM 2544	Unpublished	2010-08-05
Thermosediminibacter oceani DSM 16646	Unpublished	2010-08-05

Source: Genomes OnLine Database

Molecular Epidemiology of Pandemic Influenza A

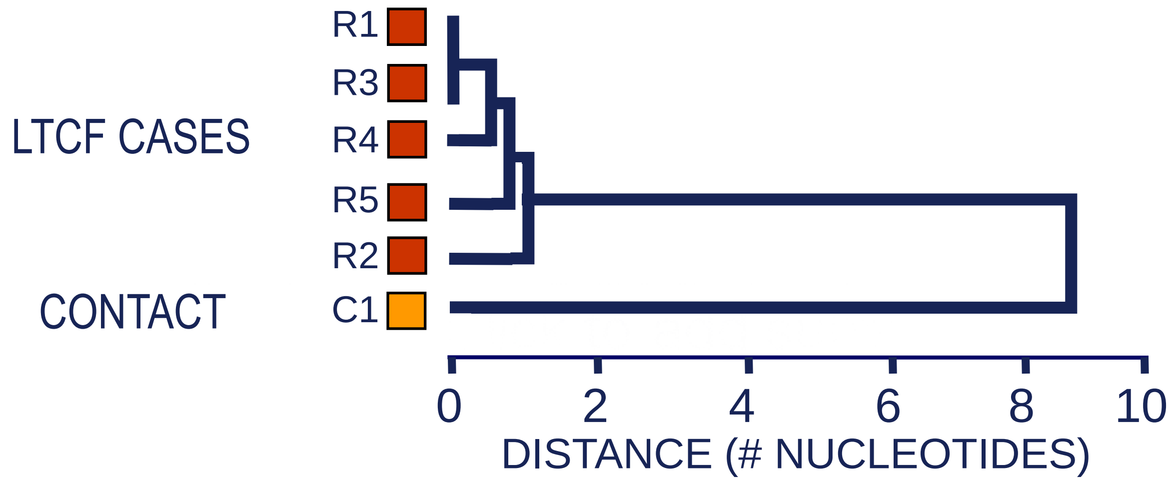
Respiratory Outbreak Within a Long-Term Care Facility



Alexander et al. 2010. Infect Cntrl Hosp Epi. 31(12):1300-2

Molecular Epidemiology of Pandemic Influenza A

Transmission of pH1N1 Within a Long-Term Care Facility



Case Identified by Contact Tracing was NOT the source of the outbreak
No Staff or Visitor Data Available for Comparison

Alexander et al. 2010. Infect Cntrl Hosp Epi. 31(12):1300-2

Evolution of Oseltamivir Resistance

Multiple pH1N1 Isolates from Single Individual

- Original Isolate = Oseltamivir Sensitive
- Intermediate Isolate = Oseltamivir Resistant
- Final Isolate = Oseltamivir Resistant

Resistance Due to single nucleotide change in Neuraminidase (NA) gene

C ⇨ T = Histidine ⇨ Tyrosine (H275Y)

But whole genome sequencing also revealed

- Multiple, additional mutations accumlulated over time ⇨ May impact Fitness and/or Virulence
- Intermediate Isolate included mixture of wildtype and mutant viruses

H1N1 Gene	PB2	HA	HA	NA	NA
Original Amino Acid	Ser	Ser	Val	His	Ser
Original Nucleotide	A	T	T	C	G
Intermediate Nucleotide	Mixed	Mixed	Mixed	T	Mixed
Final Nucleotide	G	C	C	T	A
Final Amino Acid	Gly	Pro	Ala	Tyr	Asn

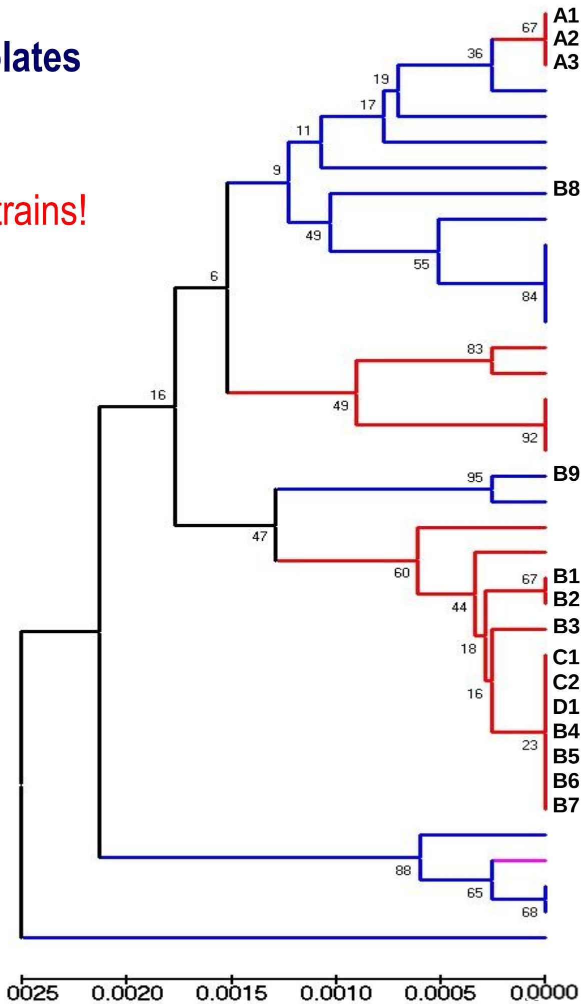
H275Y

Molecular Epidemiology of Pandemic Influenza A

Whole Genome Sequencing of pH1N1 Isolates

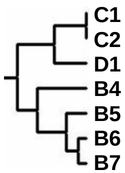
- 35 isolates from Northern Ontario
- Temporal Diversity
  - ⇨ 1st Wave and 2nd Wave = Different Strains!
- Geographic Diversity
  - ⇨ Same Town = Multiple Strains
  - ⇨ Different Towns = Different Strains

- Tree based on Protein Sequences
- Genetic Diversity > Antigenic Diversity



Genetically Identical Cluster
All Cases from Same Week and Town A

Town B: Multiple Cases
Different Times = Different Strains



Amino Acid Identity
≠
Nucleotide Identity

1st Wave (April-July 2009)
2nd Wave (Oct-Nov 2009)
2010

Emerging Pathogens: *Streptococcus pseudopneumoniae*

S. pseudopneumoniae

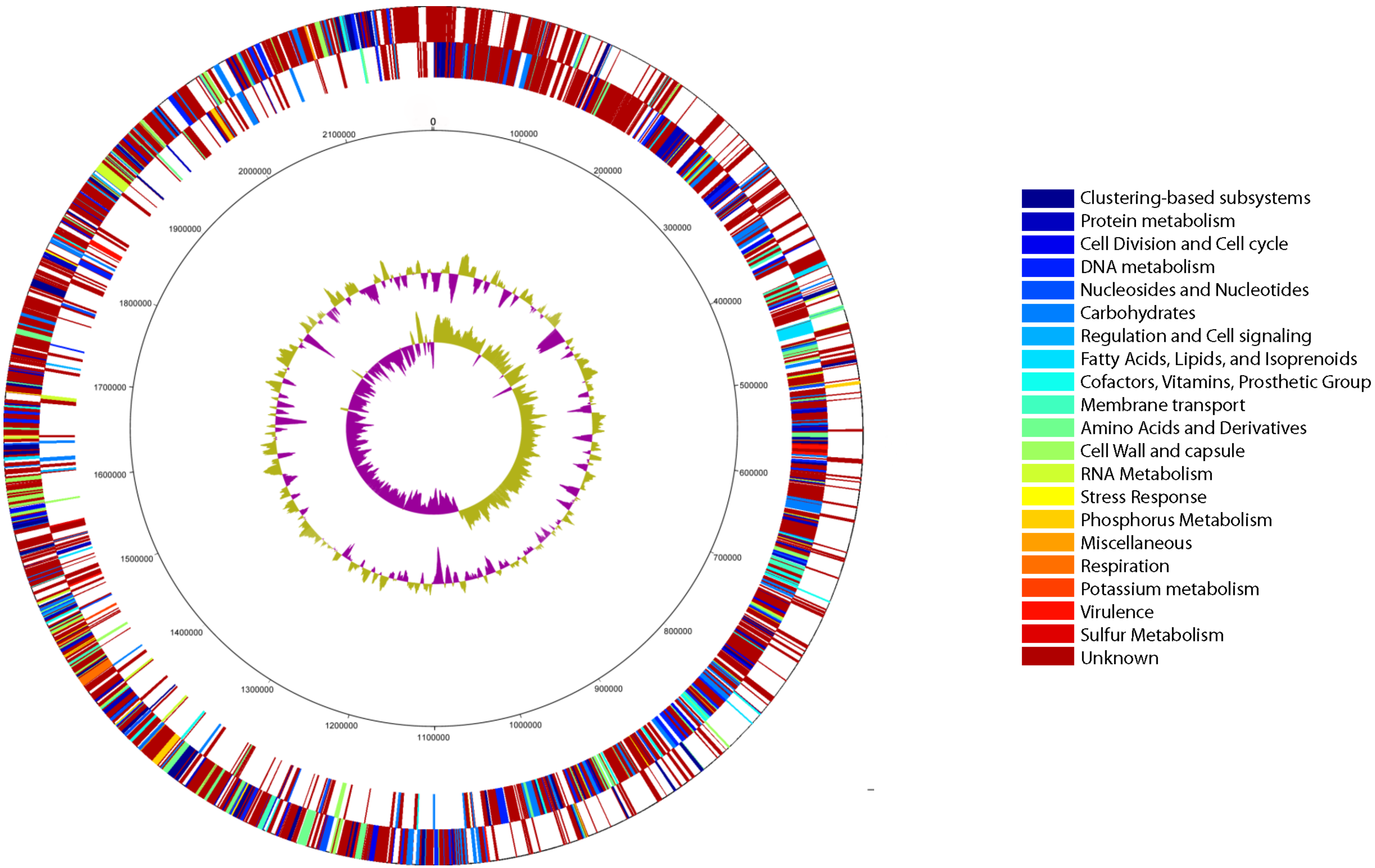
- Species established in 2004 (Arbique et al. 2004. J Clin Micro. 42:4686-4696)
- Genetically and phenotypically similar to *S. pneumoniae and S. mitis*

Test	*S.pneumo*	*S. pseudopneumo*	*S. mitis*
PHENOTYPIC			
Hemolysis	α-hemolytic	α-hemolytic	α-hemolytic
Optochin (5% CO_2)	Sensitive	Resistant	Resistant
Optochin (O_2)	Sensitive	Sensitive	Resistant
Bile Solubility	Soluble	Not Soluble	Not Soluble
GENETIC			
AccuProbe Pneumococcus	*S.pneumo*	*S.pneumo*	Not *pneumo*
Pneumolysin (*ply*) Sequencing	*S.pneumo*	*S.pneumo*	Not *pneumo*
16S rRNA Sequencing	*S.pneumo*	*S.pseudo*	*S.mitis*

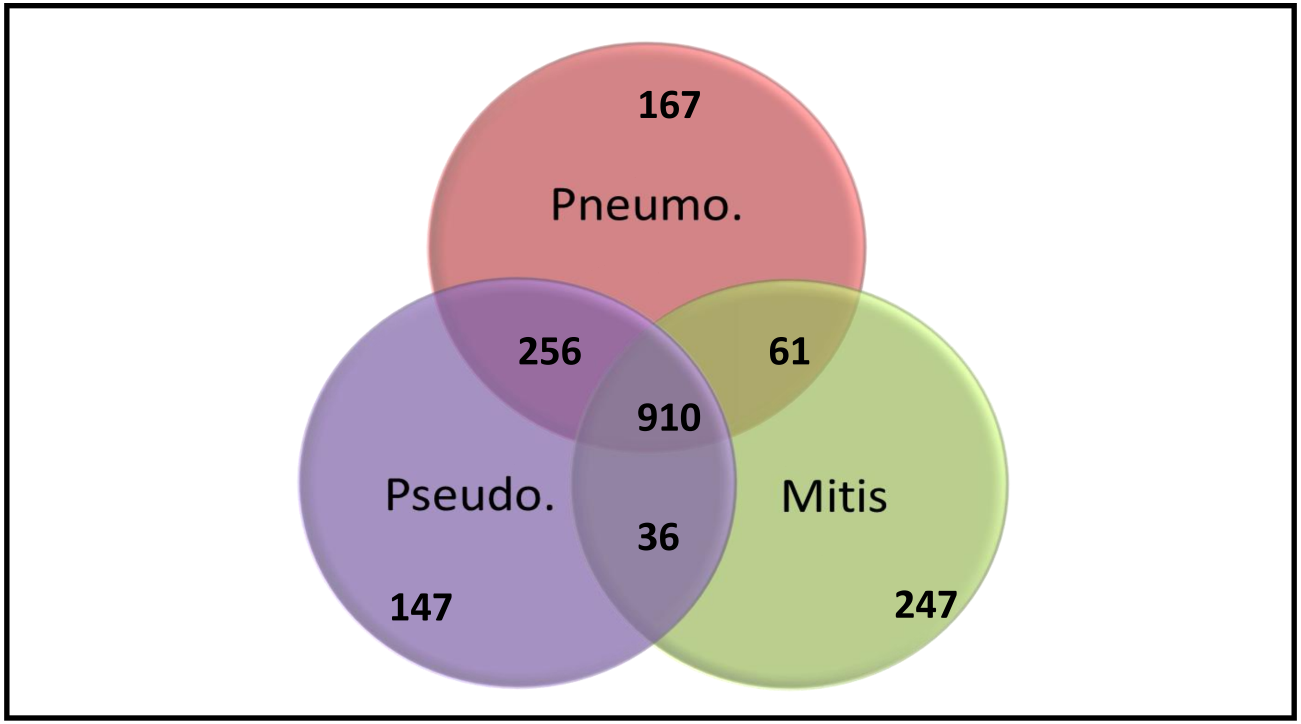
Clinical Confusion!

Is the clinical isolate a nasty *S. pneumoniae* or (mostly) harmless *S. mitis* ?
Is *S. pseudopneumoniae* a professional pathogen or common commensal?

Emerging Pathogens: *Streptococcus pseudopneumoniae*



Emerging Pathogens: *S. pseudopneumoniae* vs other Streptococci



	Pseudo.	Mitis (NCTC 12261)	Pneumo. (TIGR4)
Size	2,201,903 bp	2,045,857 bp	2,160,842 bp
CDS	2359	2172	2301
RNA	43	97	70

Emerging Pathogens: *Mycobacterium xenopi*

Mycobacterium xenopi

- Non-Tuberculous Mycobacterial species → TB-like respiratory disease
- Second most common NTM in Ontario → 800 isolates per year
- Originally identified (1959) as a frog pathogen (*Xenopus laevis*)
- Species associated with hot water systems/showers
- More common in Great Lakes region than elsewhere in Canada?

	M. tuberculosis	*M. xenopi*
Disease	Tuberculosis	Tuberculosis-like
Isolates/Year	1000 Isolates	800 Isolates
Cases/Year	500-600 Cases	Unknown
Treatment	Clear Guidelines	No Clear Guidelines
Treatment Duration	6 Months (up to 2 yrs)	Months - Years
Drug Resistance	Some (1% MDR)	No Routine Testing
Surveilance	Mandatory Reporting	No Surveillance
Suveillance Tools	Many Genotyping Tools	No Genotyping Tools

Emerging Pathogens: *Mycobacterium xenopi*

Genome Sequencing and Surveillance

- Sequenced *M.xenopi* ATCC 19250^T → Type Strain (from *Xenopus laevis*)
- Use genome data for development of molecular surveillance tools
- Sequence additional strains for which clinical data exists

Surveillance Questions

- Are animal (e.g. frog) and environmental isolates the same as human isolates?
- How much variation among human isolates?
- Is there an association between *M. xenopi* genotype and disease phenotype?

Study Outcomes

- Determine if ongoing *M. xenopi* surveillance is required
- Evaluate genetic diversity of *M. xenopi* strains in Ontario
- Determine if genotyping can predict disease phenotype/severity

Antibiotic Resistant Superbugs: NDM-1

'New Delhi metallo-ß-lactamase'

- Novel antibiotic-resistance gene found in *Escherichia coli* and *Klebsiella* spp.
- Confers high level resistance to carbapenems and other ß-lactam antibiotics
- Associated with travel and medical treatment in India, Pakistan and Bangladesh

K.pneumoniae **strain GN529:** First NDM-1 producing strain isolated in Ontario

- August 16th: specimen recovered from patient with history of travel/hospitalization in India
- August 19th: phenotypic testing revealed resistance to carbapenems as well as
  ALL tested aminoglycosides, quinolones, tetracycline and sulphonamides
  ONLY susceptible to tigecycline and colistin
- August 20th: presence of NDM-1 confirmed with a novel PCR assay
- September 3rd: Whole genome sequence obtained

NDM-1 localized to large plasmid (1 of 3 !) and exhibits novel architecture

Tijet, Alexander et al. 2010. EID *(submitted)*

	ANTIBIOTIC	MIC (µg/ml)	STATUS	ALLELE
Penicillins and Cephalosporins	Ampicillin	≥256	Resistant	$bla_{SHV\text{-}11}$ $bla_{TEM\text{-}1}$ $bla_{CTX\text{-}M\text{-}15}$ $bla_{OXA\text{-}1}$
	Amox/Clav	≥32/16	Resistant	
	Pip/Tazo	≥128/4	Resistant	
	Cefazolin	≥4	Resistant	
	Cefoxitin	≥256	Resistant	
	Ceftazidime	≥256	Resistant	
	Cefotaxime	≥256	Resistant	
	Cefepime	48	Resistant	
Carbapenems	Ertapenem	32	Resistant	$bla_{NDM\text{-}1}$
	Meropenem	≥32	Resistant	
	Imipenem	≥32	Resistant	
Aminoglycosides	Amikacin	≥256	Resistant	*armA* *aadA2* *strA-strB*
	Gentamicin	≥256	Resistant	
	Tobramycin	≥16	Resistant	
Quinolones	Nalidixic acid	≥32	Resistant	*qnrS1* $gyrA^{M}$ $parC^{M}$
	Ciprofloxacin	≥32	Resistant	
	Levofloxacin	≥8	Resistant	
Tetracycline	Tetracycline	≥16	Resistant	*tetA*
Glycylcyclines	Tigecycline	1	Sensitive	
Polymyxins	Colistin	0.5	Sensitive	
Sulfonamides	Co-trimoxazole	4/76	Resistant	*sull, dfrA12*
Macrolides	Macrolides	n/a	Resistant	*mph2lmel*

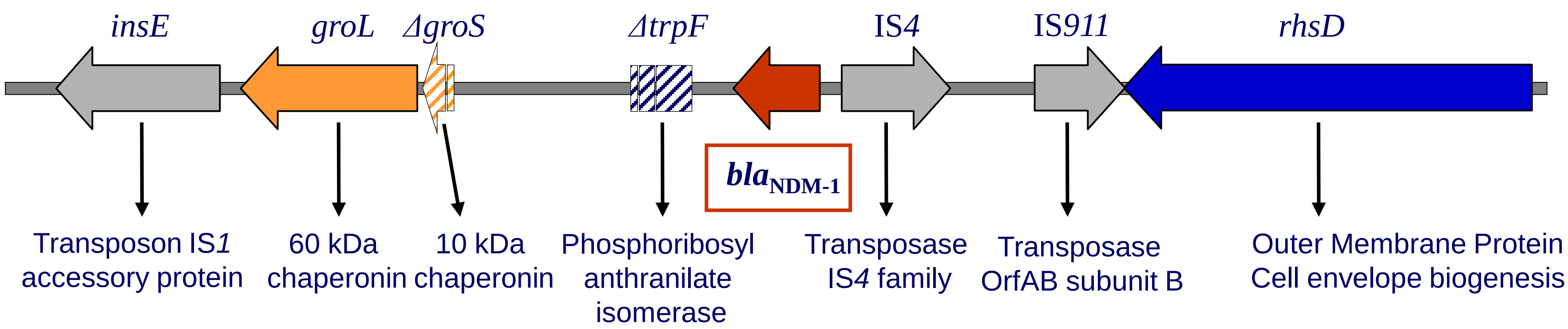
Detected by PCR*

*PCR is not part of routine testing

ALL Drug Resistance Loci Detected by Genome Sequencing

Antibiotic Resistant Superbugs: NDM-1 Architecture

K. pneumoniae GN529 plasmid (ST147)

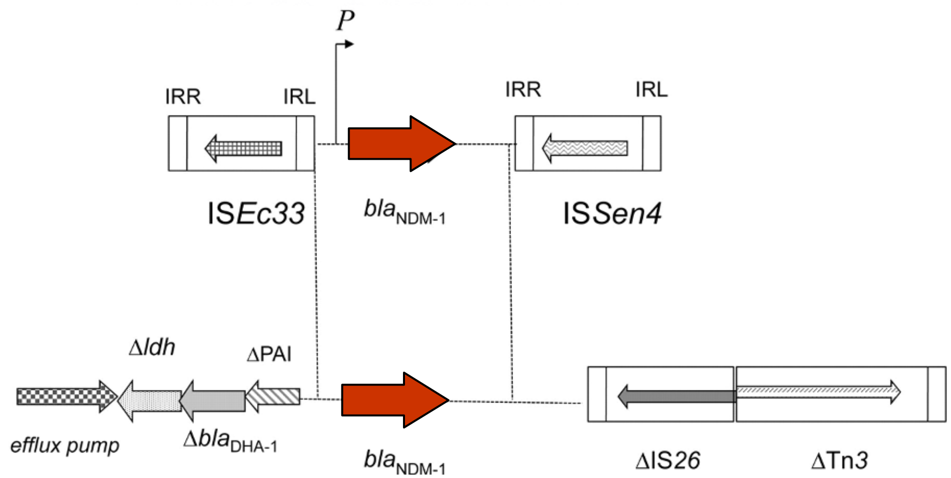


E. coli

(Poirel et al. AAC 2010. 54:4914-6)

K. pneumoniae (ST14)

(Yong et al: AAC 2009 53:5046-5054)



Microbiomes and Metagenomics

Microbial Ecosystems ⇨ The Microbiome

- In the 'real world' bacteria live in complex ecosystems ⇨ 100's of species mixed together
- Proper balance of microbial species is necessary for the ecosystem to function properly
- Unbalanced ecosystem ⇨ 'Disease'

Metagenomics

- Methods for surveying/monitoring changes in the microbiome
- Classical 'Metagenomics' ⇨ Culture… but <10% of bacteria can be grown in the lab!
- Molecular Metagenomics ⇨ PCR & Sequence → Not necessary to grow anything!

Metagenomics Projects

- 'Healthy' vs 'Sick' Microbiome ⇨ Intestinal Disease, UTIs, STIs
- Troubleshooting Molecular Tests ⇨ Is a non-target organism triggering false positive results ?
- Evaluation of Culture Media ⇨ Does enrichment media 'enrich'? Does selective media 'select'?

Clinical Genomics: Summary

Inexpensive, High Throughput DNA Sequencing is here!

Sequencing is a tool for Research *and* Clinical Labs

- ✓ Molecular Epidemiology
- ✓ Pathogen Surveillance
- ✓ Discovery of Drug Resistance Determinants and Virulence Factors
- ✓ Improved Diagnostics
- ✓ Metagenomics
- ✓ Potential to Sequence Host and Pathogen at the Same Time

Clinical Genomics Challenges

- ✓ Bioinformatics and Data Management
- ✓ Specimen Processing Pipeline and Testing Infrastructure
- ✓ Sequencing does not replace the need for Culture and Classical Microbiology

Clinical Genomics: Acknowledgments

DNA Core

Jennifer Ma
Keisha Warren
Weilong Hao

Influenza

Jonathan Gubbay
AliReza Eshaghi
Alex Marchand-Austin
Anne-Luise Winter
Peel Public Health

S. pseudopneumoniae

Dylan Pillai
Dea Shahinas
Rachel Lau
Andrew Wong

NDM-1

Roberto Melano
Samir Patel
Nathalie Tijet
Olga Lastovetska

M. xenopi

Frances Jamieson
Ted Marras
Pam Chedore & the TB Lab

Metagenomics

Anu Rebbapragada
Vanessa Allen
Gustavo Mallo
Stephen Perusini

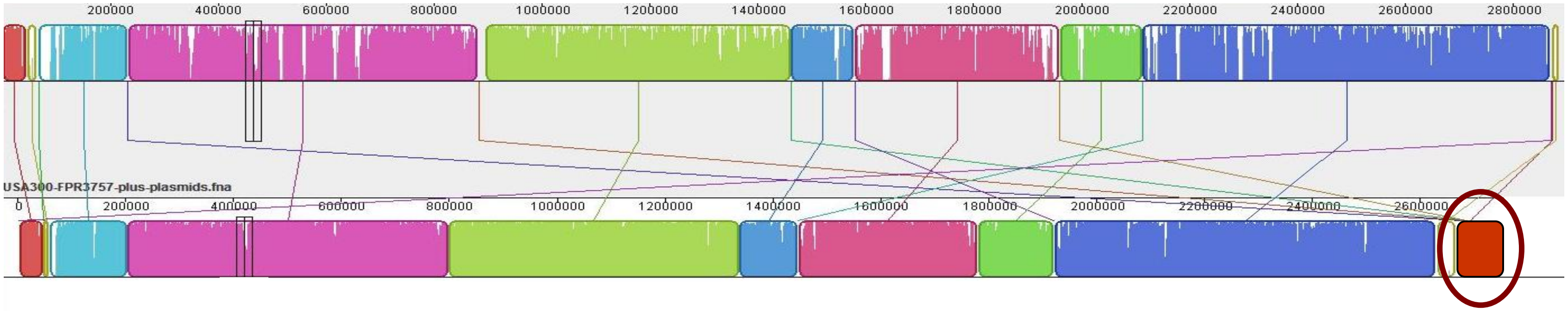
Antibiotic Resistant Superbugs: MMRSA

Clinical Isolate of MRSA

- **Phenotypically:** Mupirocin Resistant
- **Genotypically:** No Resistance Gene Detected → **PCR Failed**

Whole Genome Sequencing Performed

- Novel Strain exhibits extensive similarity to other MRSA
- **AND** includes 'extra' DNA encoding **Novel Mupirocin-Resistance Gene**
- ***mup*-2 gene** is on a plasmid and thus may be mobilized to other bacteria



'Extra DNA' including *mup*-2

DNA Core Lab

Roche GS-FLX/454 Genome Sequencer

- **Function:** Massively Parallel Pyrosequencing
  - ✓ Bacterial Genome Sequencing
    - ⇨ 2 staff
    - ⇨ 2 billion basepairs sequenced
    - ⇨ 20 genomes to date

- **Future:**
  - ✓ Metagenomics
  - ✓ Pathogen Discovery
  - ✓ Transcriptome Sequencing
  - ✓ Better Bioinformatics!

DNA Core Lab

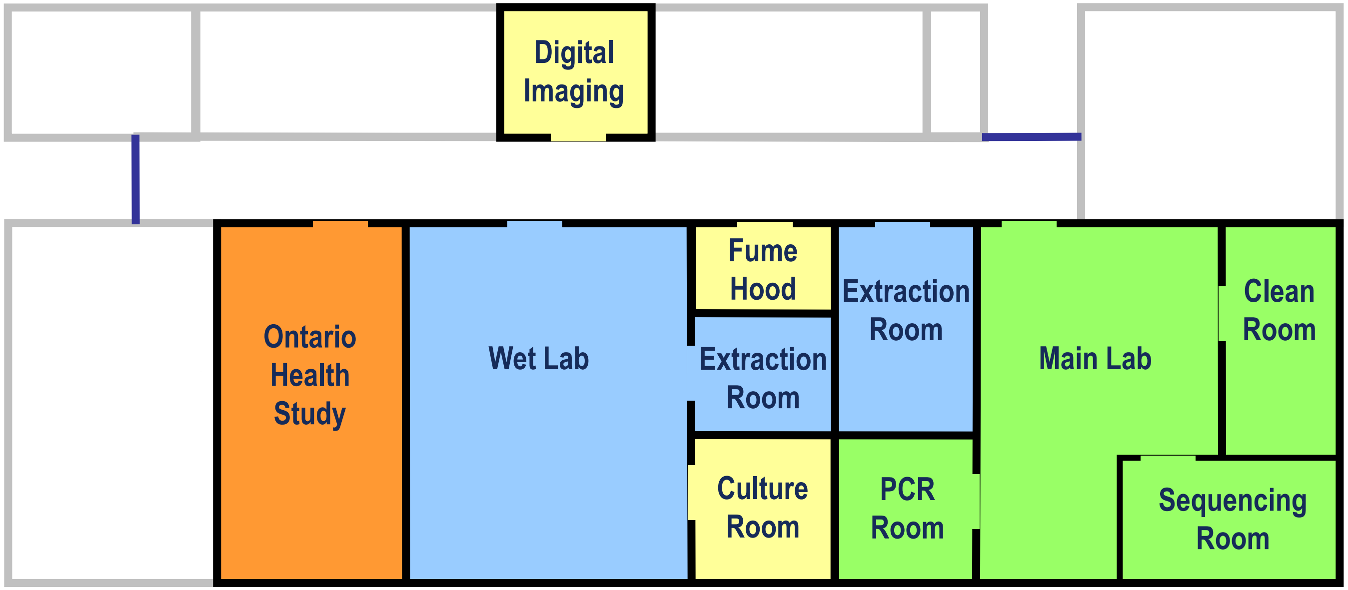
Applied Biosystems 3730xl Sequencer

- **Function:** Sanger DNA sequencing & Fragment Analysis
  - ✓ Flu Genome Sequencing (~100 genomes to date)
  - ✓ Mumps Genome Sequencing
  - ✓ MLSA
  - ✓ MIRU-VNTR (TB Genotyping)
  - ✓ 16S Sequencing

Transmission of Influenza A Pandemic (H1N1) 2009 Virus Within a Long-Term Care Facility in Ontario, Canada

	NUMBER	GENDER MALE	GENDER FEMALE	AGE MEDIAN (RANGE)	ILLNESS DURATION MEDIAN (RANGE)
ALL STAFF	220	13	207	43 yrs (23-64)	n/a
ALL RESIDENTS	180	50	130	88 yrs (33-100)	n/a
ALL CASES	23	5	18	41 yrs (23-100)	7 days (3-27)
STAFF CASES	14	2	12	33 yrs (23-59)	6 days (3-27)
RESIDENT CASES	9	3	6	70 yrs (33-100)	11 days (7-18)
pH1N1 CONFIRMED	5	2	3	56 yrs (33-88)	13 days (10-14)

101 Resources Road – EOMDC Labs



- DNA Core Lab
- H1N1 Seroprevelance Study
- Ontario Health Study
- Common Areas