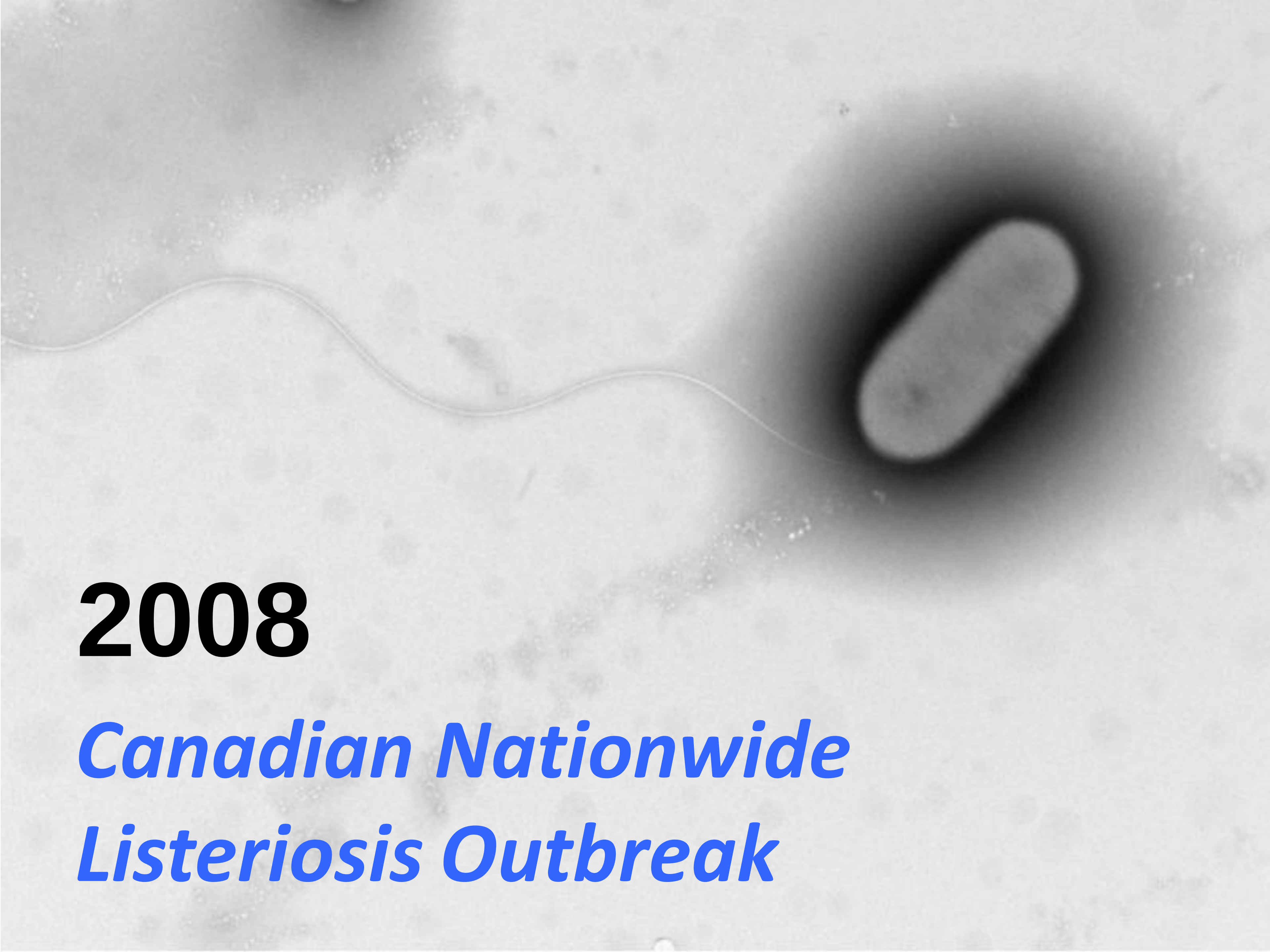
Canada's IRIDA Project for Infectious Disease Genomic Epidemiology

Gary Van Domselaar, PhD
Chief, Bioinformatics, NML – PHAC - Canada
EFSA 20th Colloquium
June 16 2014

PROTECTING CANADIANS FROM ILLNESS

Public Health Agency of Canada   Agence de la santé publique du Canada

Canada

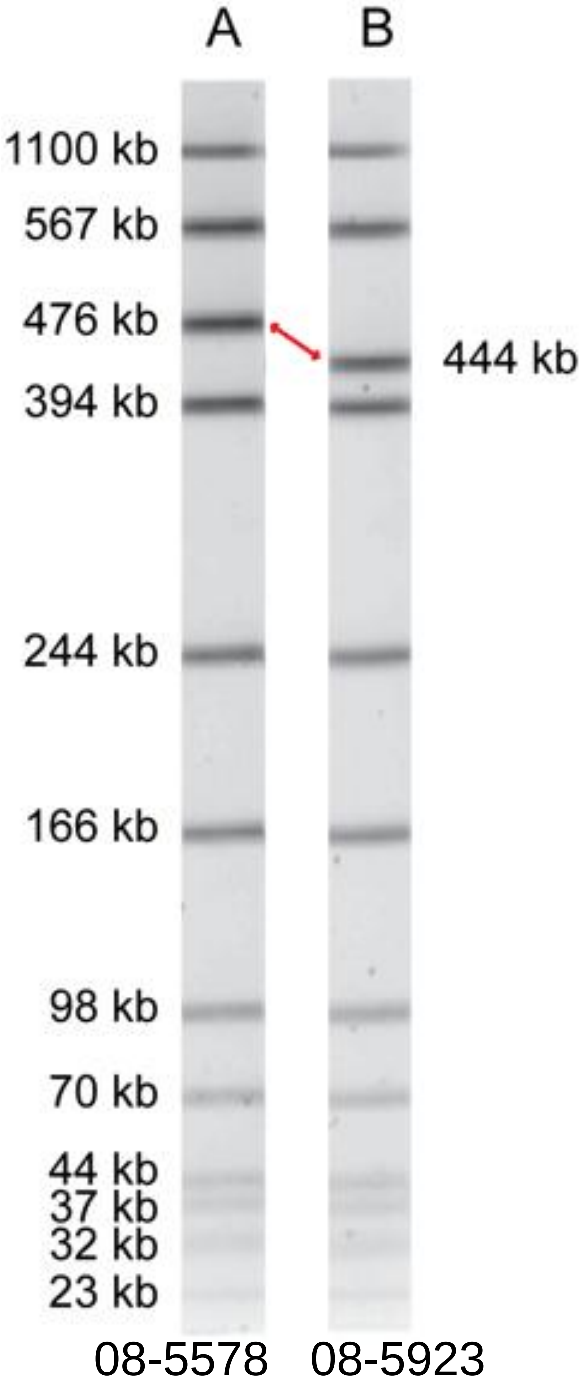


2008

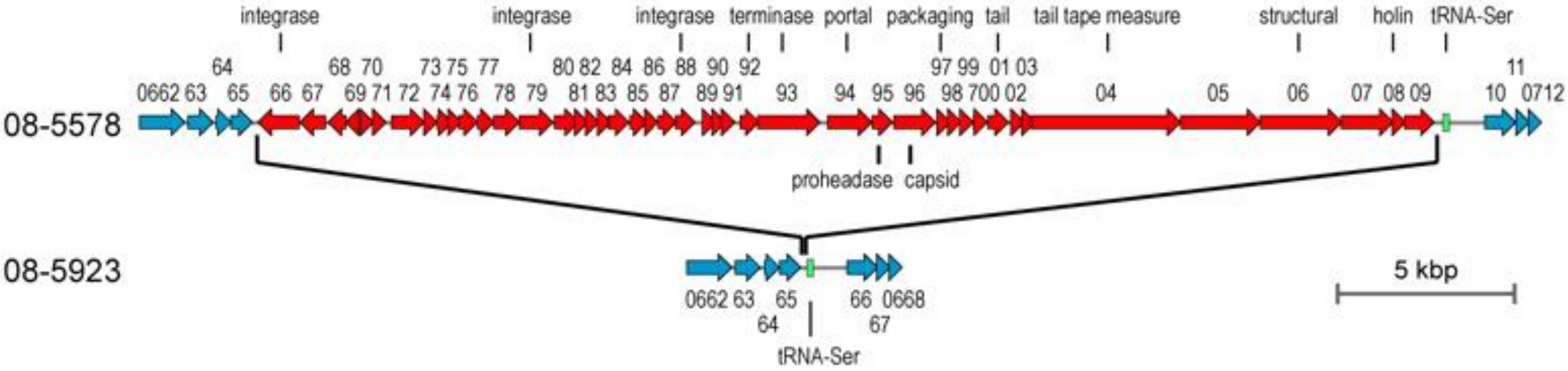
Canadian Nationwide Listeriosis Outbreak

PFGE variation detected by

PulseNet Canada

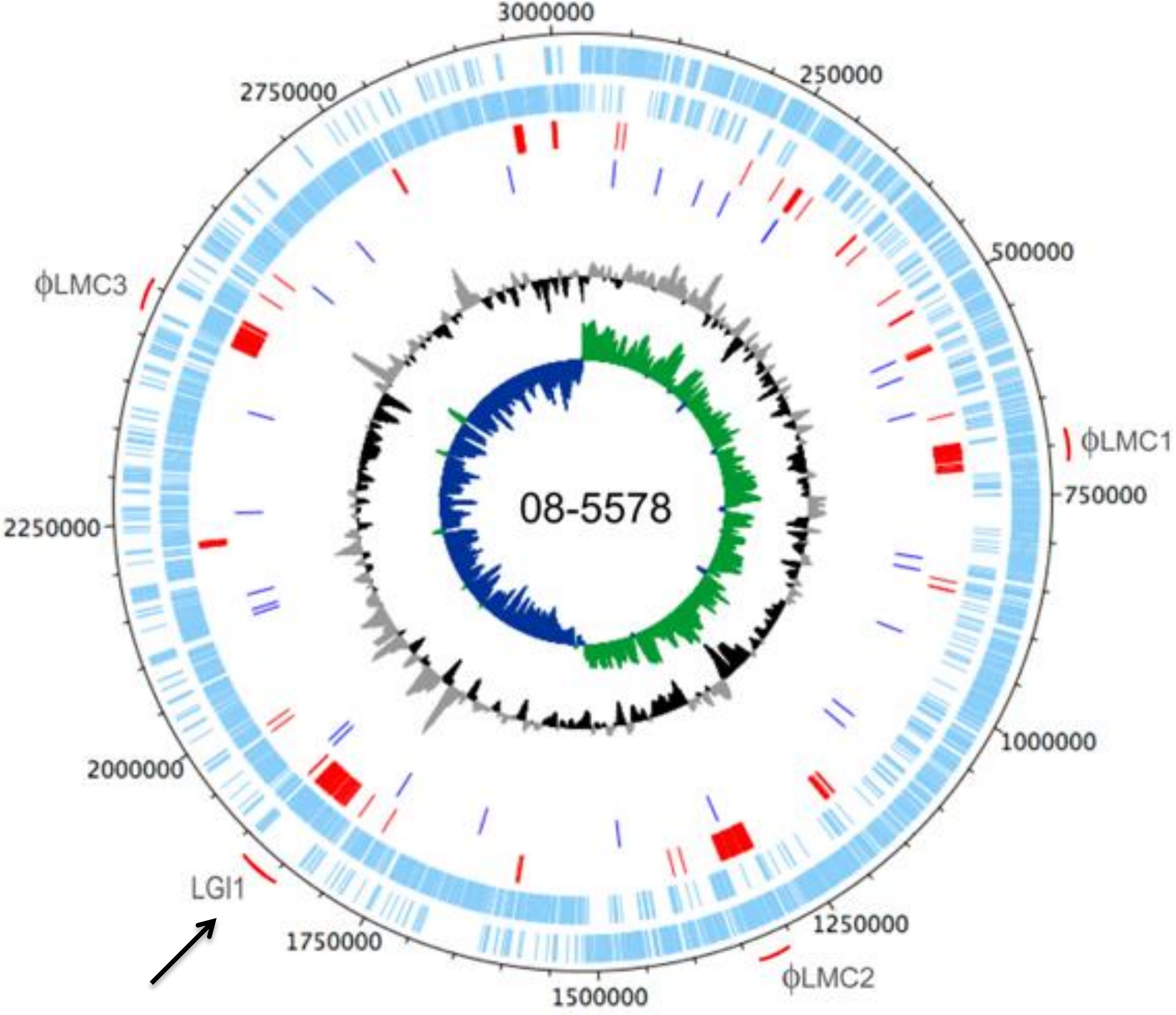


Prophage accounts for PFGE variation

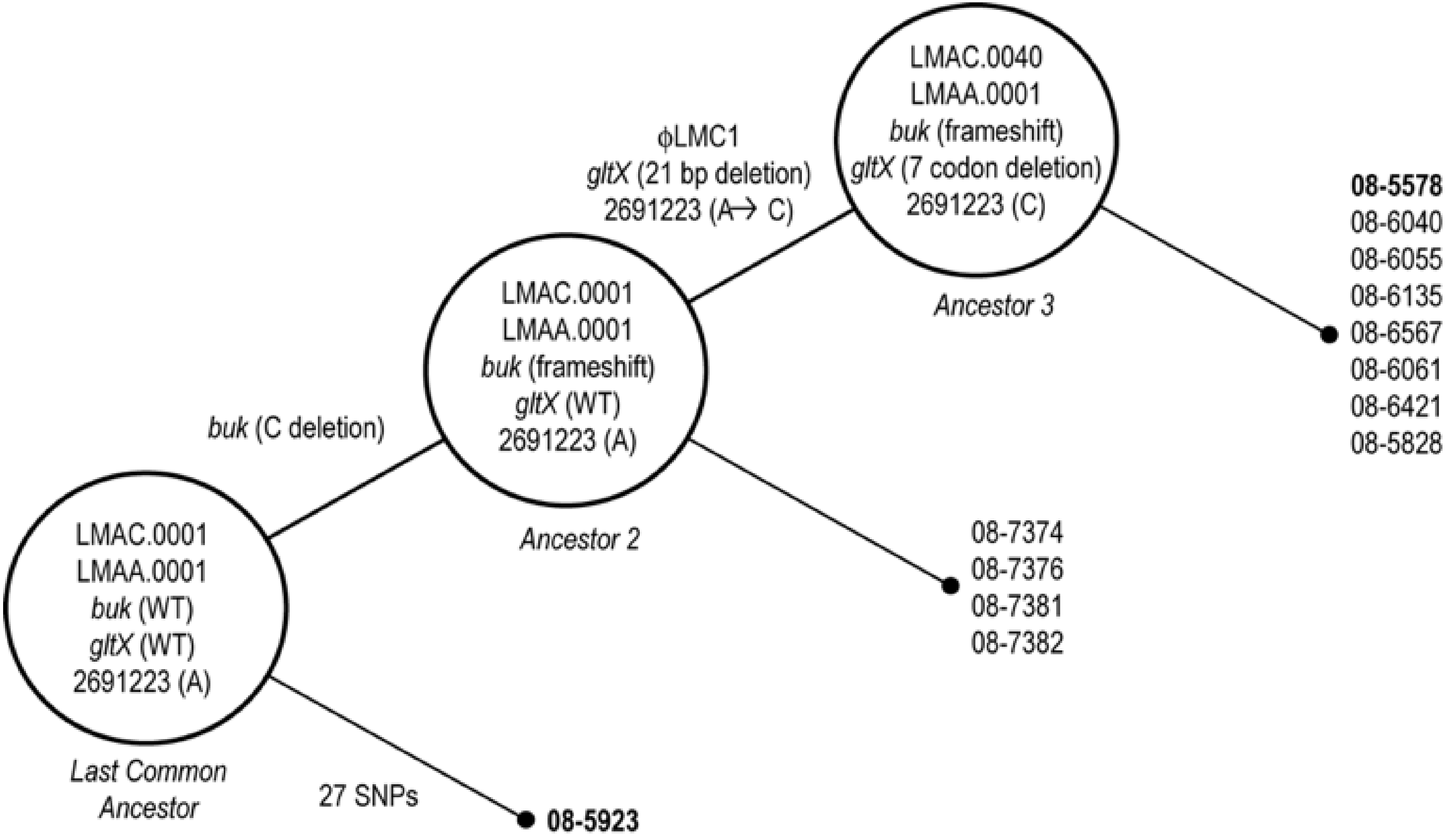


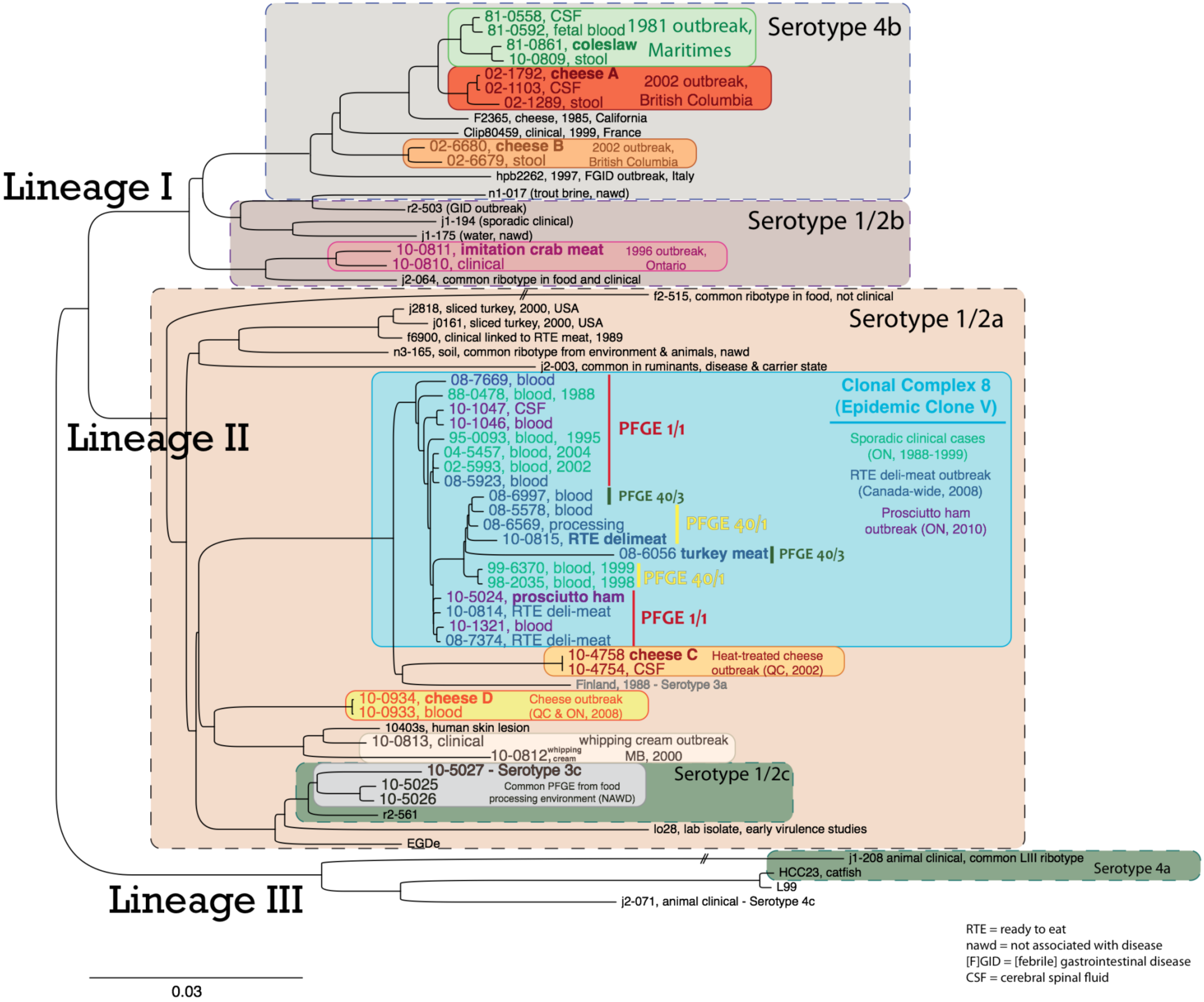
Circular Chromosomal Map

Listeria monocytogenes 08-5578

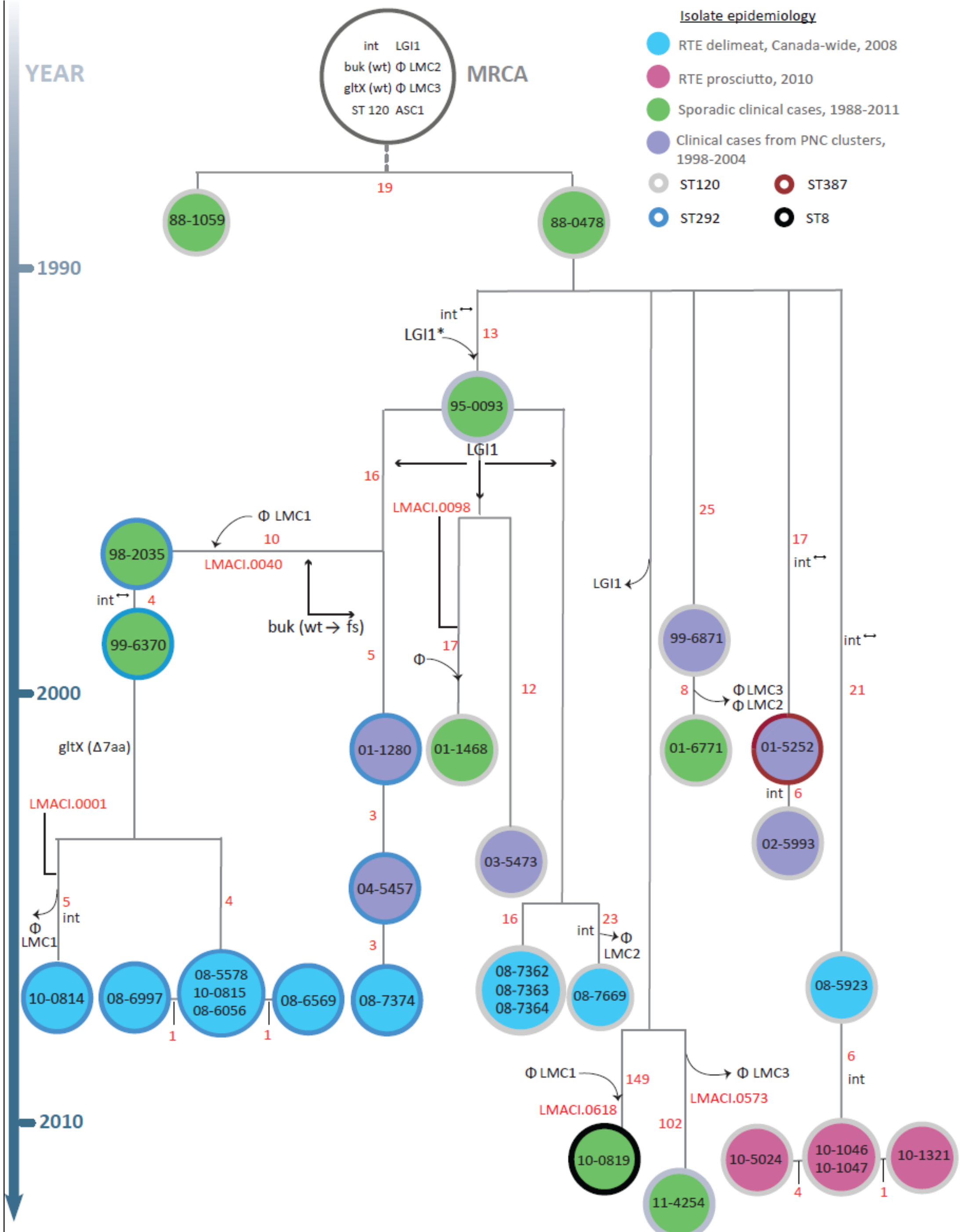


SNPs and InDels inform **Evolution**





Lineage I
Serotype 4b
81-0558, CSF
81-0592, fetal blood
81-0861, coleslaw
10-0809, stool
1981 outbreak, Maritimes
02-1792, cheese A
02-1103, CSF
02-1289, stool
2002 outbreak, British Columbia
F2365, cheese, 1985, California
Clip80459, clinical, 1999, France
02-6680, cheese B
02-6679, stool
2002 outbreak, British Columbia
hpb2262, 1997, FGID outbreak, Italy
n1-017 (trout brine, nawd)
Serotype 1/2b
r2-503 (GID outbreak)
j1-194 (sporadic clinical)
j1-175 (water, nawd)
10-0811, imitation crab meat
10-0810, clinical
1996 outbreak, Ontario
j2-064, common ribotype in food and clinical
Lineage II
Serotype 1/2a
f2-515, common ribotype in food, not clinical
j2818, sliced turkey, 2000, USA
j0161, sliced turkey, 2000, USA
f6900, clinical linked to RTE meat, 1989
n3-165, soil, common ribotype from environment & animals, nawd
j2-003, common in ruminants, disease & carrier state
Clonal Complex 8 (Epidemic Clone V)
Sporadic clinical cases (ON, 1988-1999)
RTE deli-meat outbreak (Canada-wide, 2008)
Prosciutto ham outbreak (ON, 2010)
08-7669, blood
88-0478, blood, 1988
10-1047, CSF
10-1046, blood
95-0093, blood, 1995
04-5457, blood, 2004
02-5993, blood, 2002
08-5923, blood
PFGE 1/1
08-6997, blood
PFGE 40/3
08-5578, blood
08-6569, processing
10-0815, RTE delimeat
PFGE 40/1
08-6056 turkey meat
PFGE 40/3
99-6370, blood, 1999
98-2035, blood, 1998
PFGE 40/1
10-5024, prosciutto ham
10-0814, RTE deli-meat
10-1321, blood
08-7374, RTE deli-meat
PFGE 1/1
10-4758 cheese C
10-4754, CSF
Heat-treated cheese outbreak (QC, 2002)
Finland, 1988 - Serotype 3a
10-0934, cheese D
10-0933, blood
Cheese outbreak (QC & ON, 2008)
10403s, human skin lesion
10-0813, clinical
10-0812, whipping cream
whipping cream outbreak MB, 2000
Serotype 1/2c
10-5027 - Serotype 3c
10-5025
10-5026
Common PFGE from food processing environment (NAWD)
r2-561
lo28, lab isolate, early virulence studies
EGDe
Lineage III
j1-208 animal clinical, common LIII ribotype
HCC23, catfish
Serotype 4a
L99
j2-071, animal clinical - Serotype 4c
0.03
RTE = ready to eat
nawd = not associated with disease
[F]GID = [febrile] gastrointestinal disease
CSF = cerebral spinal fluid


YEAR
1990
2000
2010
MRCA
int LGI1
buk (wt) Φ LMC2
gltX (wt) Φ LMC3
ST 120 ASC1
Isolate epidemiology
RTE delimeat, Canada-wide, 2008
RTE prosciutto, 2010
Sporadic clinical cases, 1988-2011
Clinical cases from PNC clusters, 1998-2004
ST120
ST387
ST292
ST8
19
88-1059
88-0478
int
LGI1*
13
95-0093
LGI1
16
LMACI.0098
Φ LMC1
10
98-2035
LMACI.0040
int
4
99-6370
buk (wt → fs)
5
17
Φ
12
25
17
int
LGI1
99-6871
int
8
Φ LMC3
Φ LMC2
21
gltX (Δ7aa)
01-1280
01-1468
01-6771
01-5252
int
6
02-5993
LMACI.0001
3
03-5473
04-5457
5
int
Φ
LMC1
4
3
16
int
23
Φ
LMC2
10-0814
08-6997
08-5578
10-0815
08-6056
08-6569
08-7374
08-7362
08-7363
08-7364
08-7669
08-5923
1
1
Φ LMC1
149
LMACI.0618
Φ LMC3
LMACI.0573
102
6
int
10-0819
11-4254
10-5024
10-1046
10-1047
10-1321
4
1

The 2010 Cholera Outbreak in Haiti: How Science Solved a Controversy

Fabini D. Orata[1], Paul S. Keim[2,3], Yan Boucher[1]*

[1] Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada, [2] Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, Arizona, United States of America, [3] Pathogen Genomics Division, Translational Genomics Research Institute, Flagstaff, Arizona, United States of America

The 2010 Earthquake and Cholera Outbreak in Haiti

On January 12, 2010, a catastrophic 7.0 magnitude earthquake struck Haiti, affecting 3,500,000 people [1,2]. This severely damaged an already marginal public sanitation system, creating ideal conditions for outbreaks of major infectious diseases. In October 2010, nine months after the earthquake, an outbreak of cholera started, which quickly spread all across the country [3]. As of January 7, 2014, 8,534 deaths and 697,256 cholera cases have been reported by the Haitian Ministry of Public Health and Population [4]. Prior to 2010, there was no reported history of cholera in Haiti, despite devastating outbreaks in the Caribbean region in the 19th century [5]. Many wondered where the cholera in Haiti came from. Two hypotheses as to its origin were presented. The climatic hypothesis argued that nonpathogenic *Vibrio cholerae*, indigenous in the coastal waters of Haiti, was given the right environmental circumstances and evolved into a pathogenic strain [6]. On the other hand, the human transmission hypothesis suggested that cholera was introduced to Haiti by individuals infected in a foreign country.

Cholera and *V. cholerae*

Cholera, caused by *V. cholerae*, is a disease characterized by very severe diarrhea and dehydration, which can lead to death in less than 48 hours if left untreated. Cholera is treatable through oral rehydration salt solutions, intravenous fluids, or antibiotics, depending on severity [7]. Ingestion of contaminated water is the main vehicle for human infection. The principal virulence determinant is the potent cholera toxin, encoded by the *ctxAB* genes on the bacteriophage CTXϕ [8] found in toxigenic *V. cholerae* genomes. The toxin, together with other virulence factors, leads to the harmful effects of the *V. cholerae* infection (Figure 1). These auxiliary virulence factors are encoded in clusters of genes called genomic islands, which are acquired by environmental *V. cholerae* through horizontal gene transfer [9] (Figure 1). It is also important to note that infection can be asymptomatic, and these cases play a major role in the transmission of the disease [10]. *V. cholerae* is of major public health concern because of its potential to cause pandemics. Seven such pandemics have been recorded since 1817, when cholera first spread beyond the Indian subcontinent, all presumably caused by *V. cholerae* belonging to the O1 serogroup. *V. cholerae* of the classical biotype dominated the previous six pandemics and was replaced by the El Tor biotype in the currently ongoing seventh pandemic, which originated in Southeast Asia in 1961 [7]. In 1992, a new serogroup of *V. cholerae*, O139, was first identified after causing cholera epidemics in India and Bangladesh [11]. Cholera has been eliminated from industrialized countries by efficient water and sewage treatments but not in less-developed countries with poor water sanitation.

Initial Studies Support the Human Transmission Hypothesis

Rumors spread on October 27, 2010, pointing blame for the outbreak at the United Nations Stabilization Mission in Haiti (MINUSTAH) troops from Nepal who had recently set up camp in Meille, a small village 2 km south of Mirebalais (Figure 2A). This followed revelations by news reporters showing improper sewage waste disposal in the camp [12,13].

The stool samples collected by the Haiti National Public Health Laboratory from cholera patients at the start of the outbreak were sent to the Centers for Disease Control and Prevention (CDC) for analysis. On November 13, the CDC reported that *V. cholerae* El Tor O1 was isolated from the samples and independent isolates were indistinguishable by multiple rapid phenotypic and molecular characterization methods, suggesting that a single strain caused the outbreak and was likely introduced into Haiti in one event [14].

A study by Piarroux and colleagues made use of all available epidemiological data, checking hospital records, doing field surveys, and applying statistics for spatiotemporal analysis, to trace the source and spread of the outbreak [15]. The findings of their investigation confirmed the news reporters' claims. Based on all gathered evidence, they put together a likely scenario – the MINUSTAH camp contaminated the Meille tributary with fecal matter by their unsanitary practice of sewage drainage, and the Meille tributary connects downstream to the Latem River that goes through the town of Mirebalais, the site of the first reported cholera case [16]. The Latem River eventually connects to the Artibonite River, the longest as well as the most important river that spans Haiti (Figure 2A). The movement and spread of cholera in the early onset of the epidemic was closely linked to proximity with the Artibonite River.

Citation: Orata FD, Keim PS, Boucher Y (2014) The 2010 Cholera Outbreak in Haiti: How Science Solved a Controversy. PLoS Pathog 10(4): e1003967. doi:10.1371/journal.ppat.1003967

Editor: Joseph Heitman, Duke University Medical Center, United States of America

Published April 3, 2014

Funding: This work was supported by the Integrated Microbial Biodiversity Program of the Canadian Institute for Advanced Research (CIFAR) and the Natural Sciences and Engineering Research Council of Canada (NSERC) to FDO and YB. PSK was supported by the Northern Arizona University Cowden Endowment for Microbiology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: yboucher@ualberta.ca

SINGLE (FOREIGN) SOURCE HYPOTHESIS

http://solutionshaiti.blogspot.com/

Outbreak Isolates Sequenced (9 Total)

Jan 26

3 cases late Oct

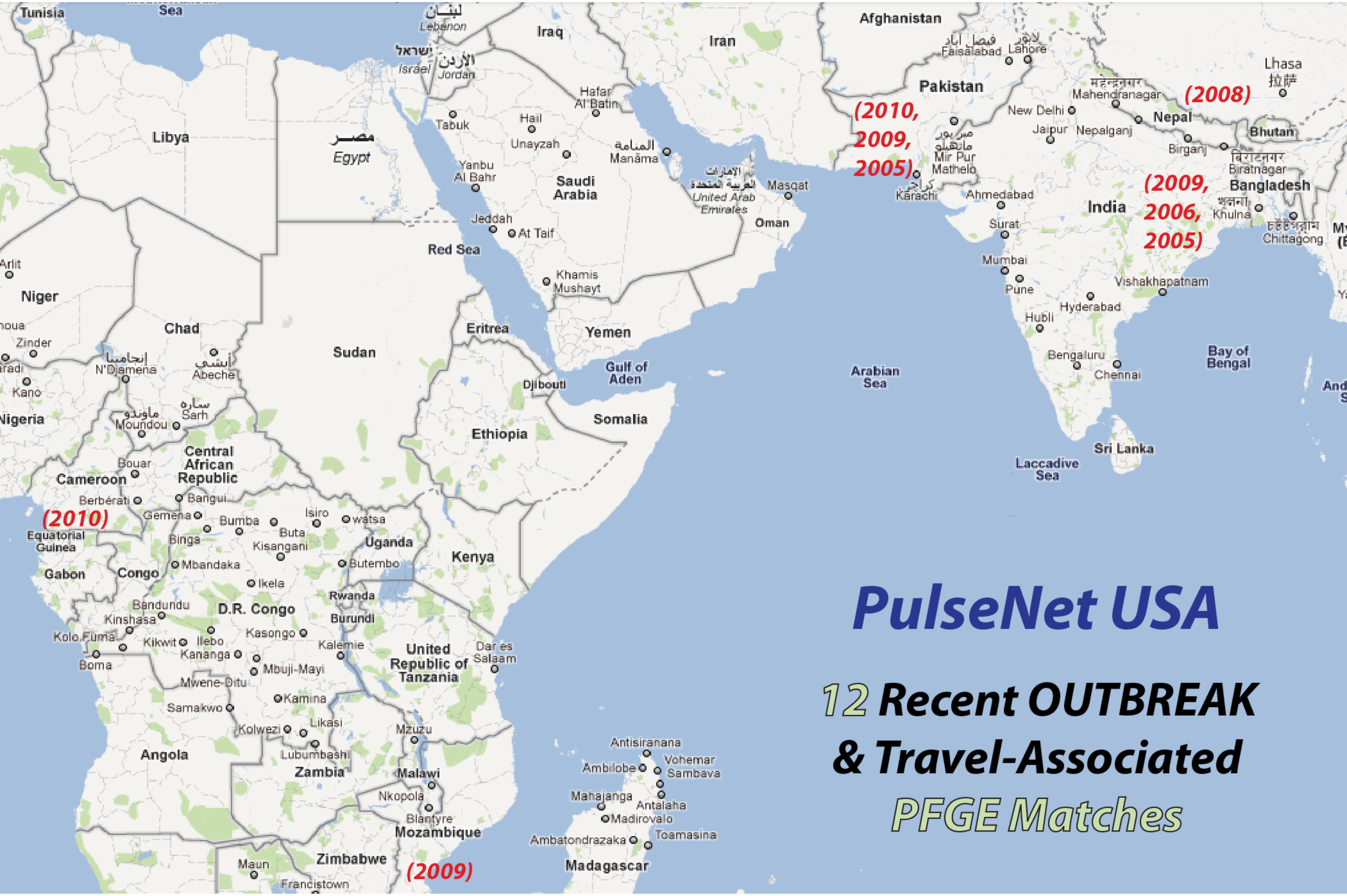
Case 1 Oct. 19

PFGE Variants

MINU STAH

2011

Nov 27



PulseNet USA

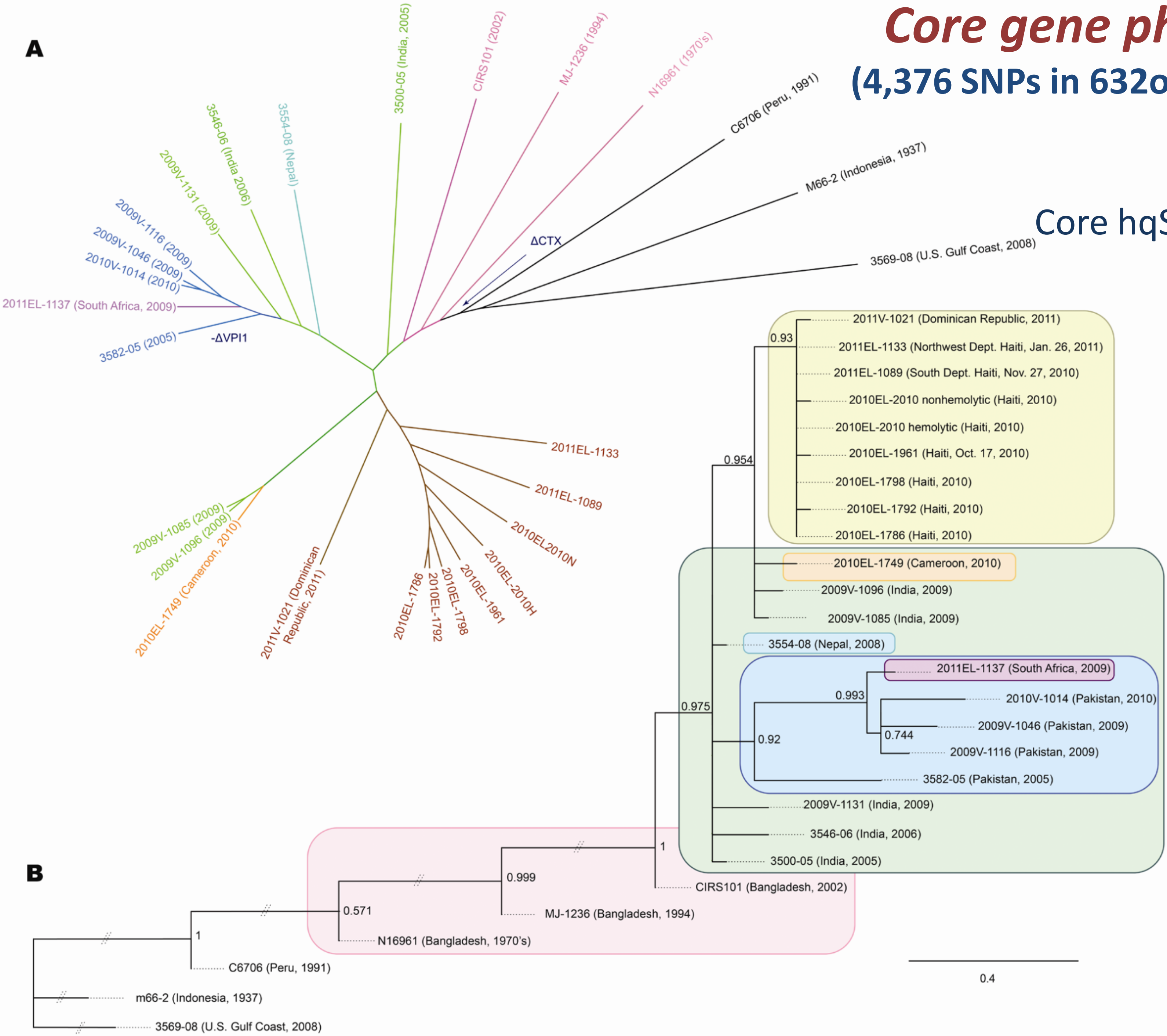
12 Recent OUTBREAK & Travel-Associated *PFGE Matches*

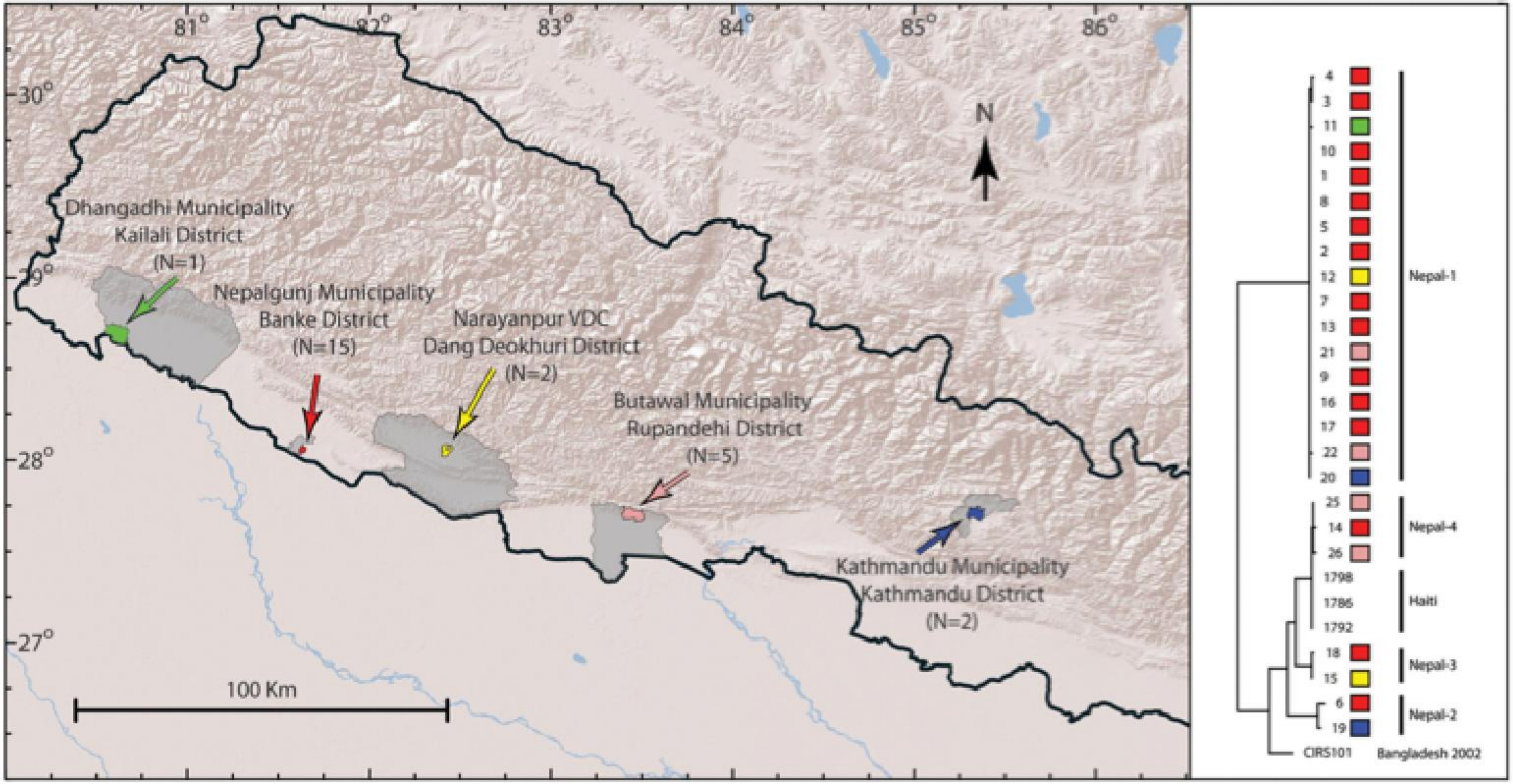
Core gene phylogeny

(4,376 SNPs in 632orthologues)

Tool(s):

Core hqSNP Pipeline

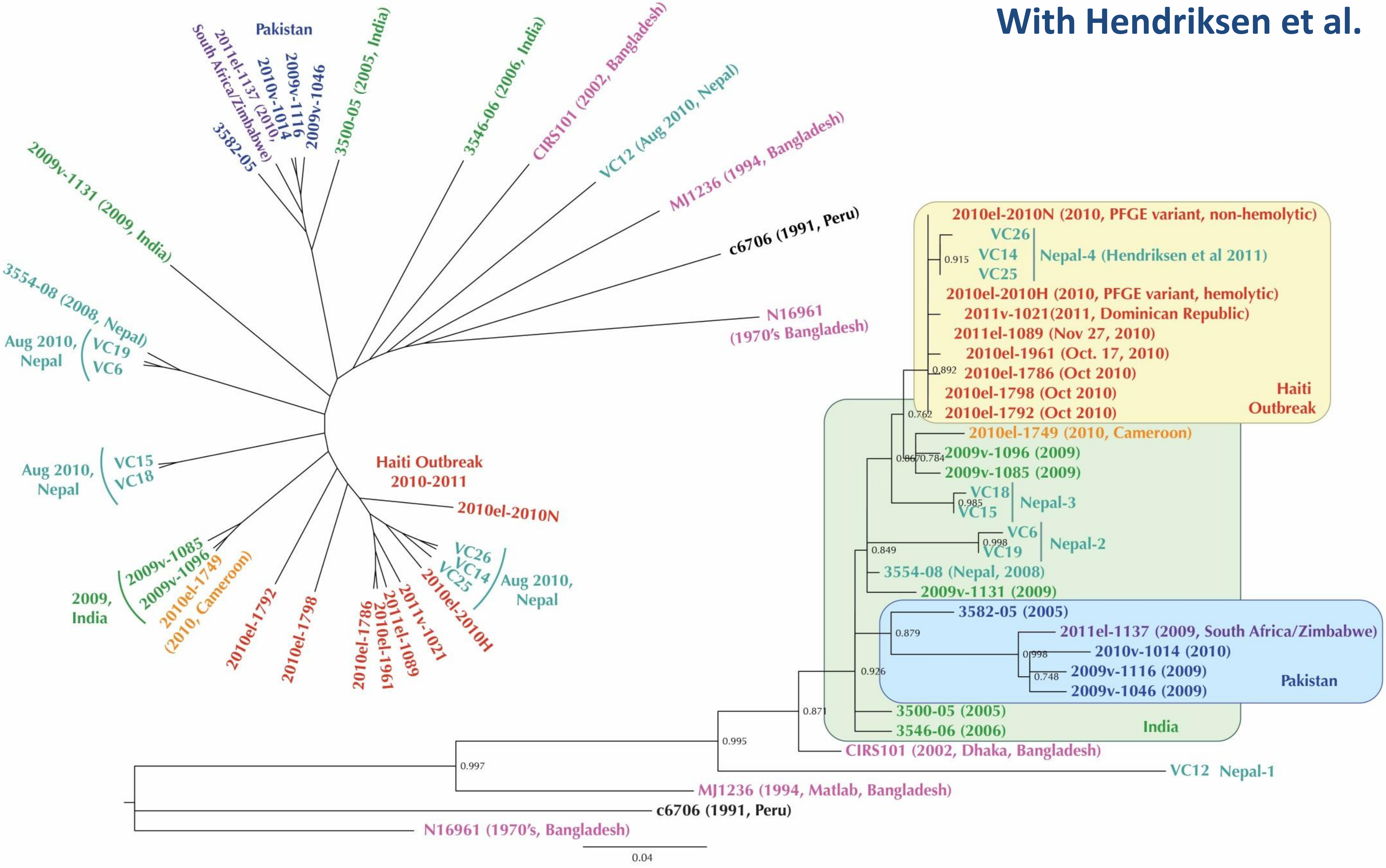




Hendriksen et. al. mBio. 2011

Core gene phylogeny

With Hendriksen et al.



	log P(model|data)	S.E.	Constant	Expanding	Exponential
Constant	-296.61	0.232	-	-4.243	-4.342
Expansion	-286.841	0.156	4.243	-	-0.099
exponential	-286.613	0.188	4.342	0.099	-

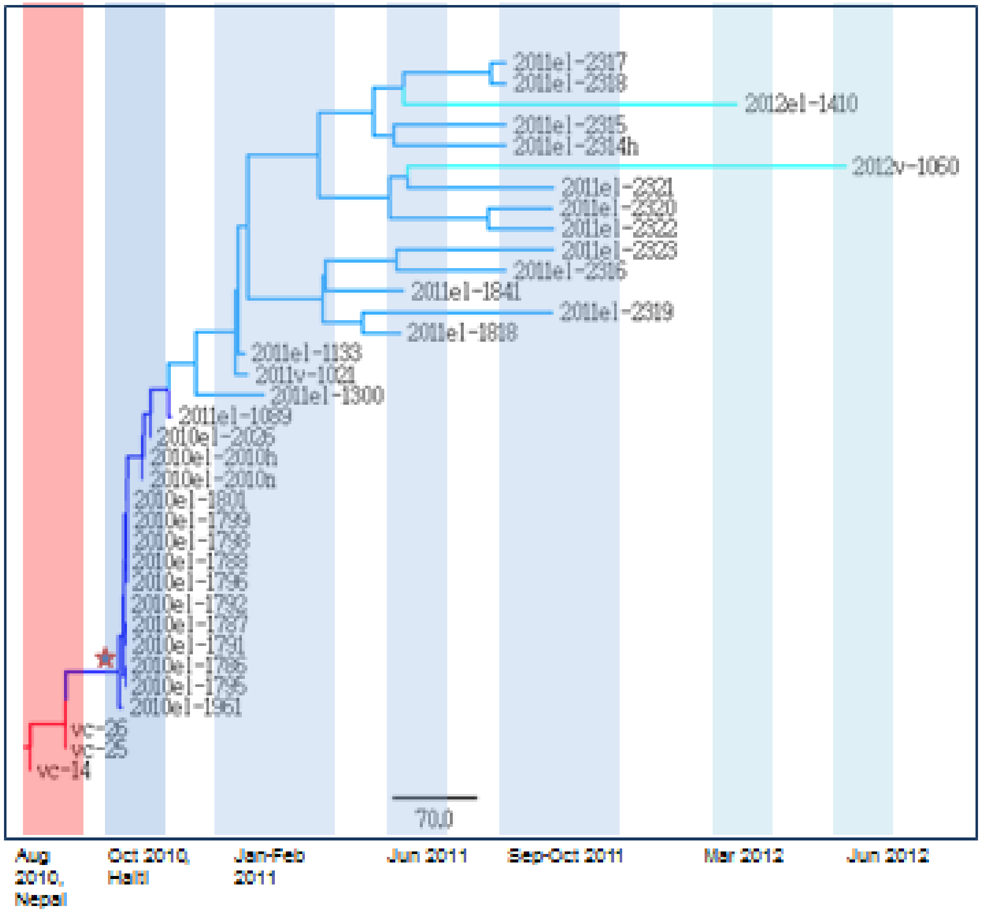
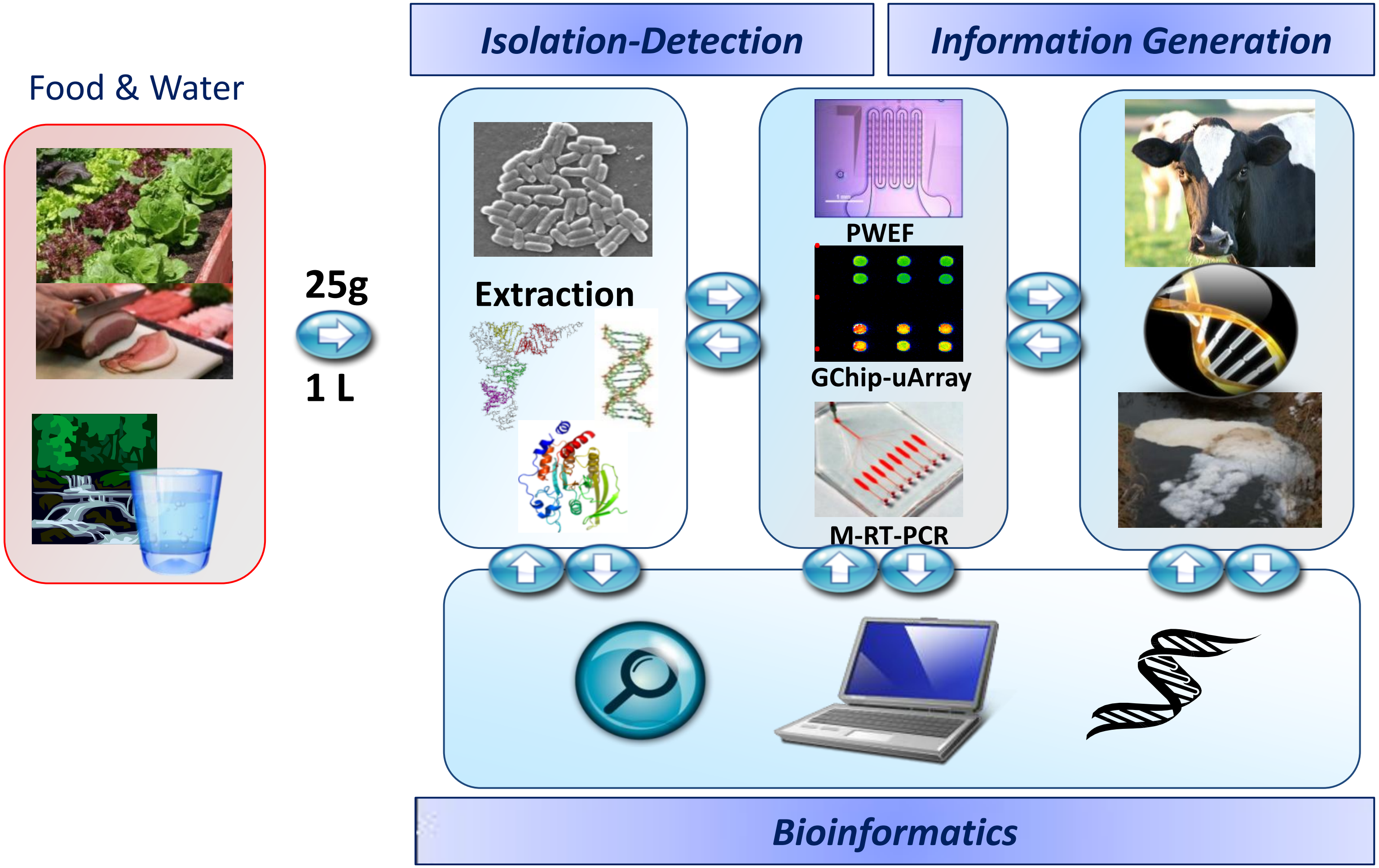


Figure S5. Molecular clock.

Canada’s GRDI Interdepartmental Pilot Project for Food & Water Safety



Isolation-Detection

Information Generation

Food & Water

25g

1 L

Extraction

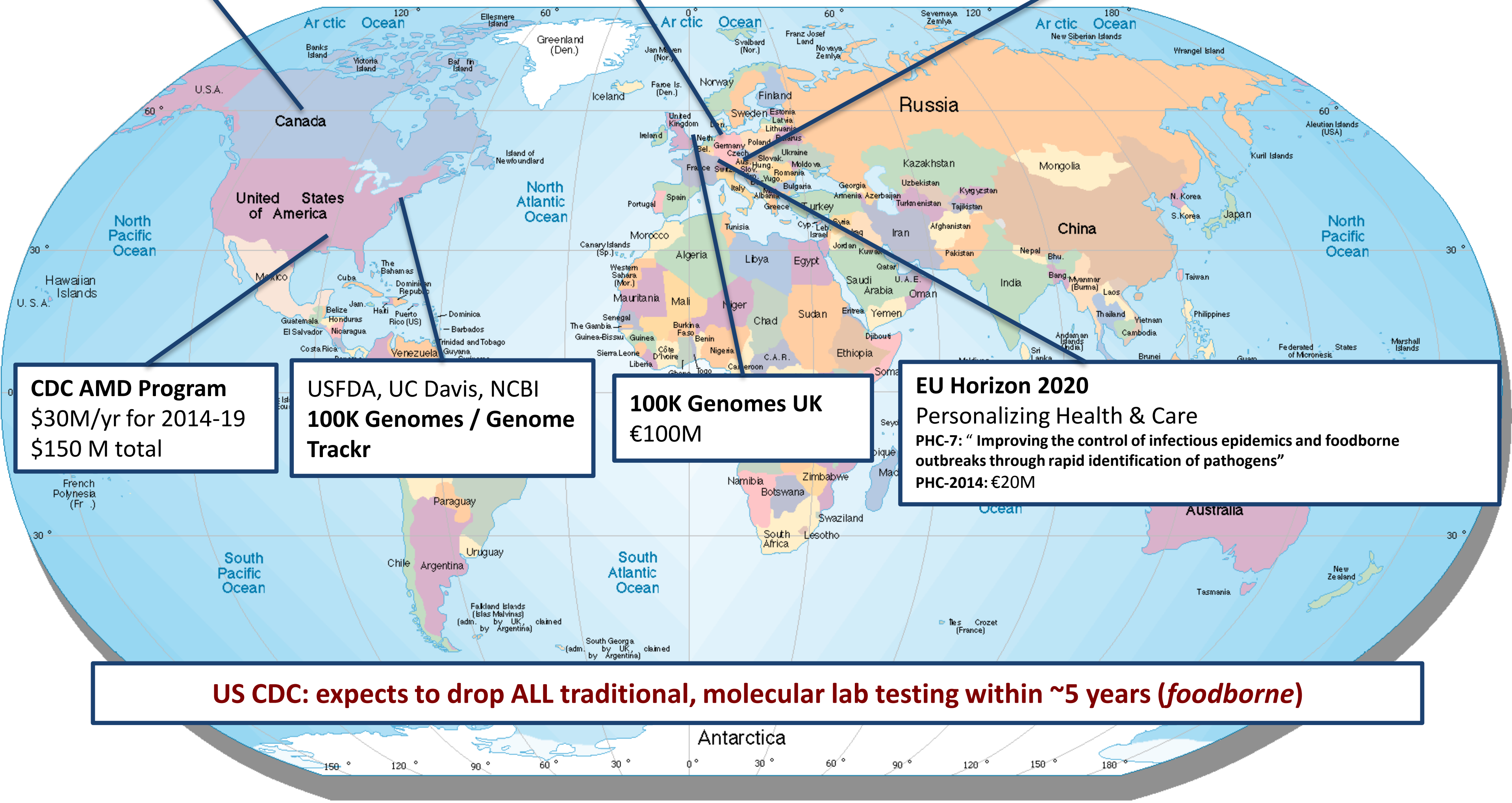
PWEF

GChip-uArray

M-RT-PCR

Bioinformatics

- Developing an integrated analytical system for ‘omic data querying & visualization
- Establishing a national enteric bacterial pathogens database
- Delivering novel detection platforms
- Modernizing source attribution & risk assessment methods



NML PHG Program
GRDI FWS: $7.2M over 4 y
IRIDA: $1.5M over 3 y

Danish Technical University
€6M PoC Global Data Sharing

Patho-NGen-Trace Consortium
€6M over 5y, WGS for Disease Surveillance

CDC AMD Program
$30M/yr for 2014-19
$150 M total

USFDA, UC Davis, NCBI
100K Genomes / Genome Trackr

100K Genomes UK
€100M

EU Horizon 2020
Personalizing Health & Care
PHC-7: " Improving the control of infectious epidemics and foodborne outbreaks through rapid identification of pathogens"
PHC-2014: €20M

US CDC: expects to drop ALL traditional, molecular lab testing within ~5 years (*foodborne*)

GRDI FWS Platform

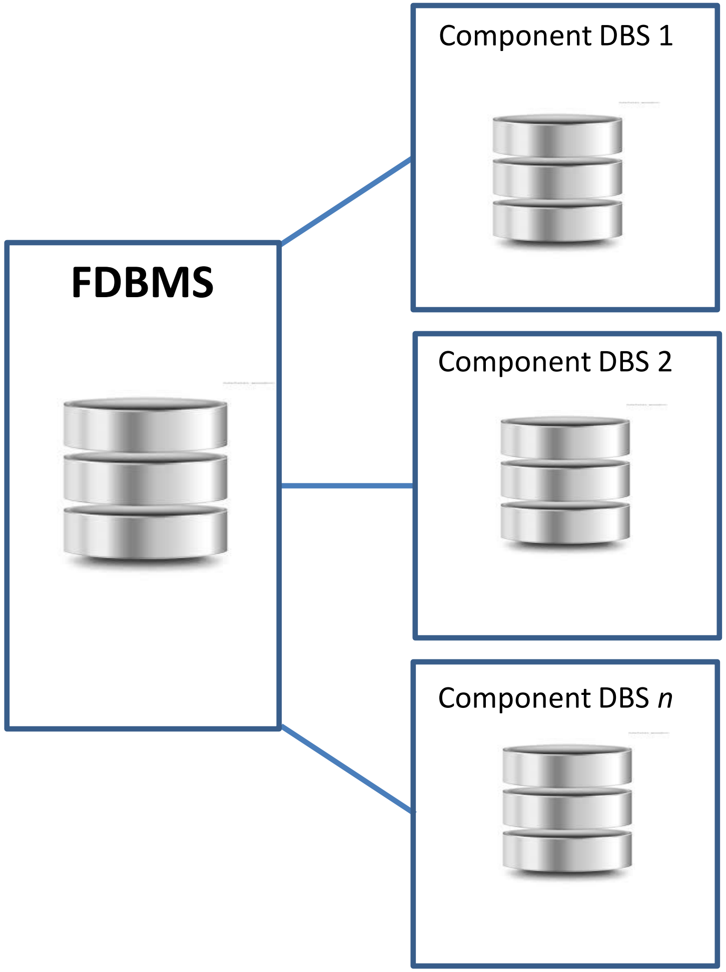
1. WGS database system and application ontology for infectious disease genomic epidemiology
2. Bioinformatics tools and pipelines for analysis of WGS data
3. *In silico* typing tools and linking of WGS data to microbial sequence typing databases
4. Training

open source

IRIDA Platform Expansion

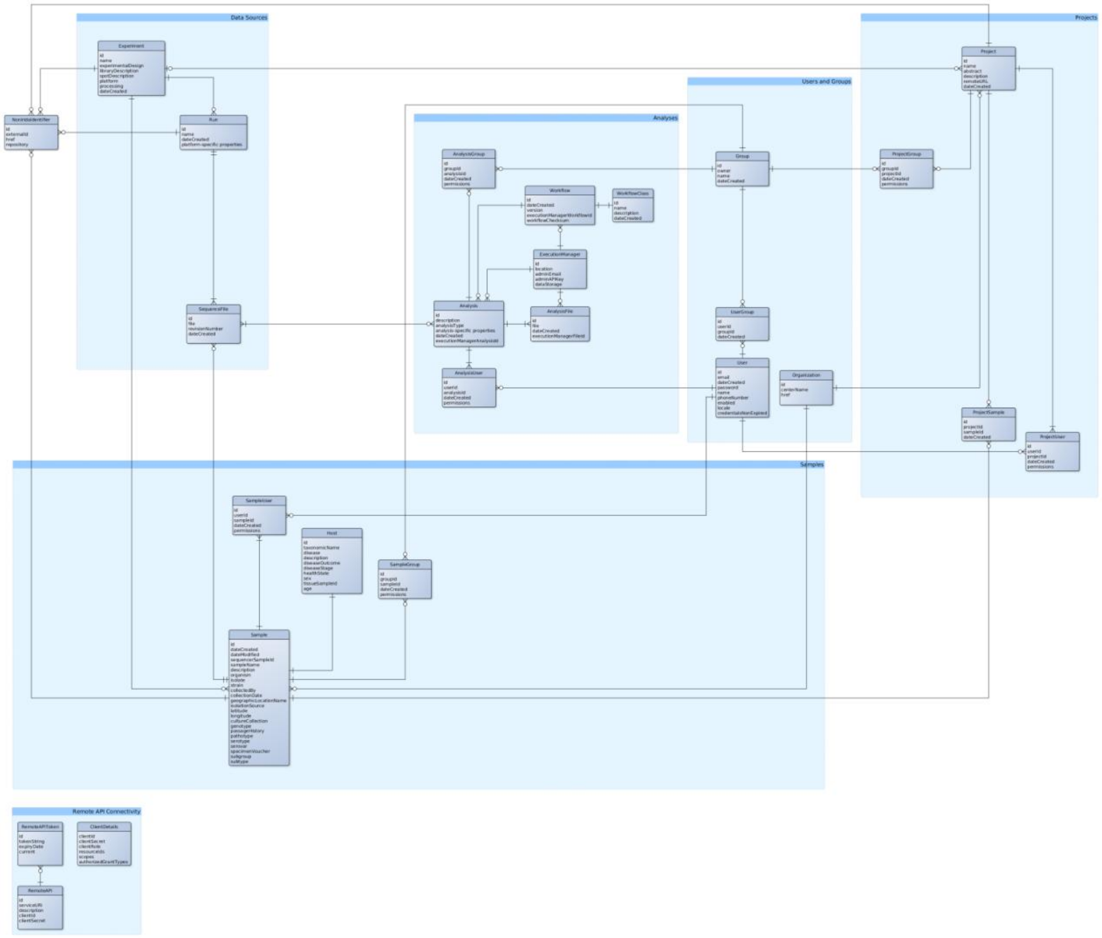
1. IRIDA: Integrated Rapid Infectious Disease Analysis
2. 1.56M, 3 year Genome Canada Large-Scale Applied Platform Grant
3. SFU / BCCDC / PHAC-NML / Dalhousie
4. Expands the GRDI platform to include additional epidemiology and pathogenomic analysis functionality, and additional training
5. Implementation in provincial public health labs
6. Training

Federated Database Design

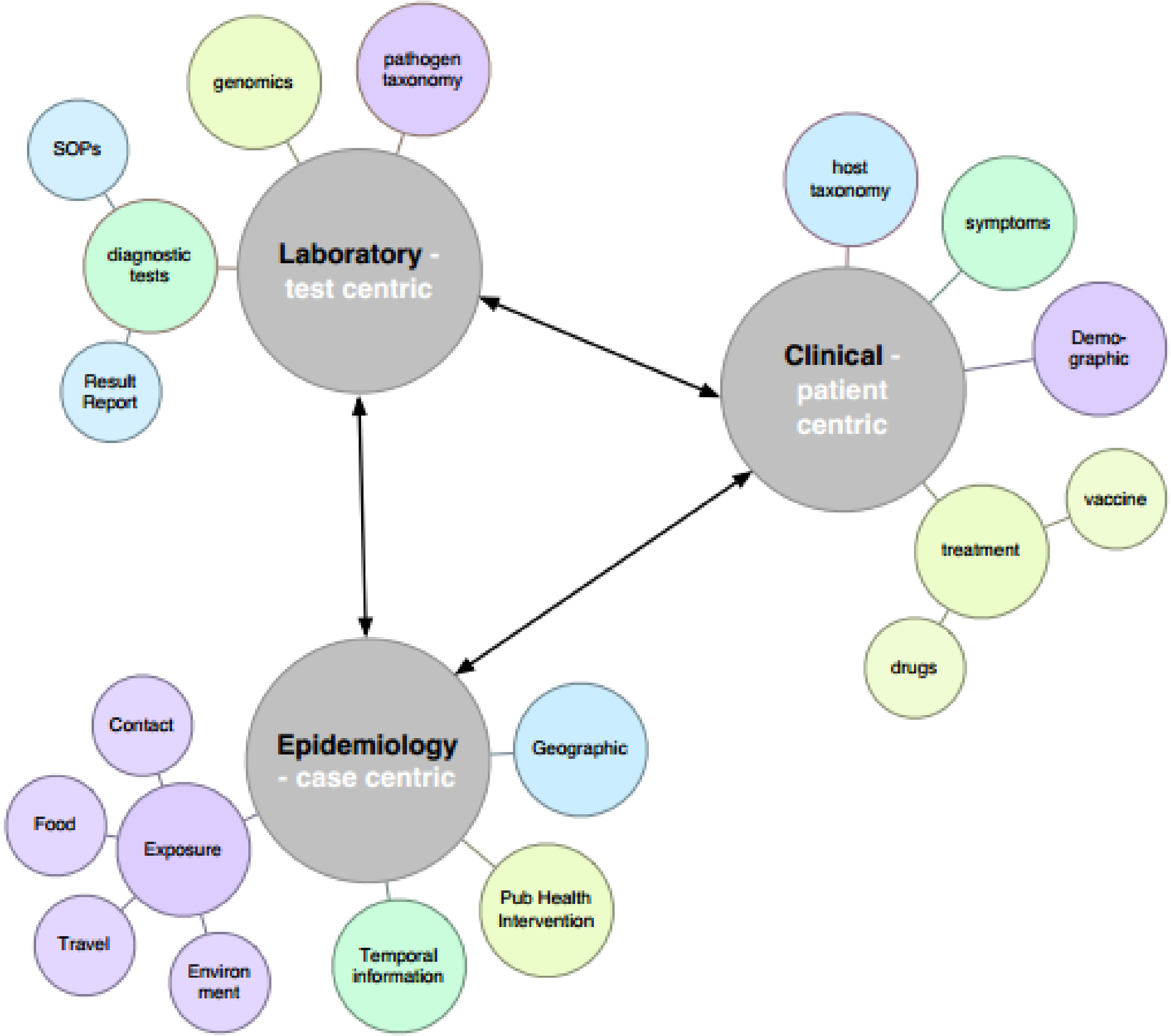


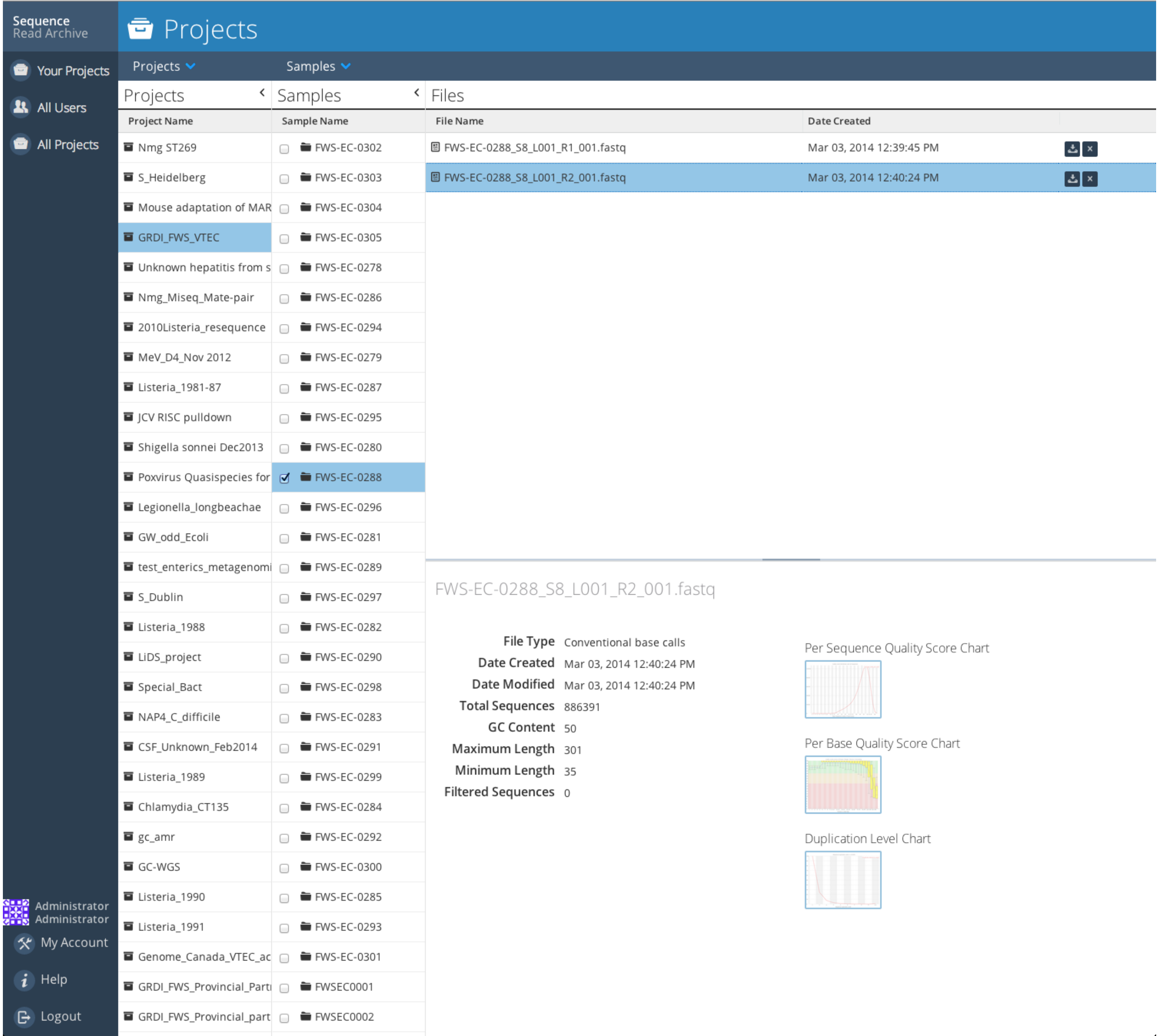
Data Model

- Based on NCBI
- Project
- Sample
- Data Sources
  - Sequence File
  - Experiment
  - Run
- Analysis
  - Workflow
  - Analysis File



Ontologies and Data Standards





Sequence Read Archive

Projects

Projects

Project Name
Nmg ST269
S_Heidelberg
Mouse adaptation of MAR
GRDI_FWS_VTEC
Unknown hepatitis from s
Nmg_Miseq_Mate-pair
2010Listeria_resequence
MeV_D4_Nov 2012
Listeria_1981-87
JCV RISC pulldown
Shigella sonnei Dec2013
Poxvirus Quasispecies for
Legionella_longbeachae
GW_odd_Ecoli
test_enterics_metagenomi
S_Dublin
Listeria_1988
LiDS_project
Special_Bact
NAP4_C_difficile
CSF_Unknown_Feb2014
Listeria_1989
Chlamydia_CT135
gc_amr
GC-WGS
Listeria_1990
Listeria_1991
Genome_Canada_VTEC_ac
GRDI_FWS_Provincial_Parti
GRDI_FWS_Provincial_part

Samples

	Sample Name
☐	FWS-EC-0302
☐	FWS-EC-0303
☐	FWS-EC-0304
☐	FWS-EC-0305
☐	FWS-EC-0278
☐	FWS-EC-0286
☐	FWS-EC-0294
☐	FWS-EC-0279
☐	FWS-EC-0287
☐	FWS-EC-0295
☐	FWS-EC-0280
☑	FWS-EC-0288
☐	FWS-EC-0296
☐	FWS-EC-0281
☐	FWS-EC-0289
☐	FWS-EC-0297
☐	FWS-EC-0282
☐	FWS-EC-0290
☐	FWS-EC-0298
☐	FWS-EC-0283
☐	FWS-EC-0291
☐	FWS-EC-0299
☐	FWS-EC-0284
☐	FWS-EC-0292
☐	FWS-EC-0300
☐	FWS-EC-0285
☐	FWS-EC-0293
☐	FWS-EC-0301
☐	FWSEC0001
☐	FWSEC0002

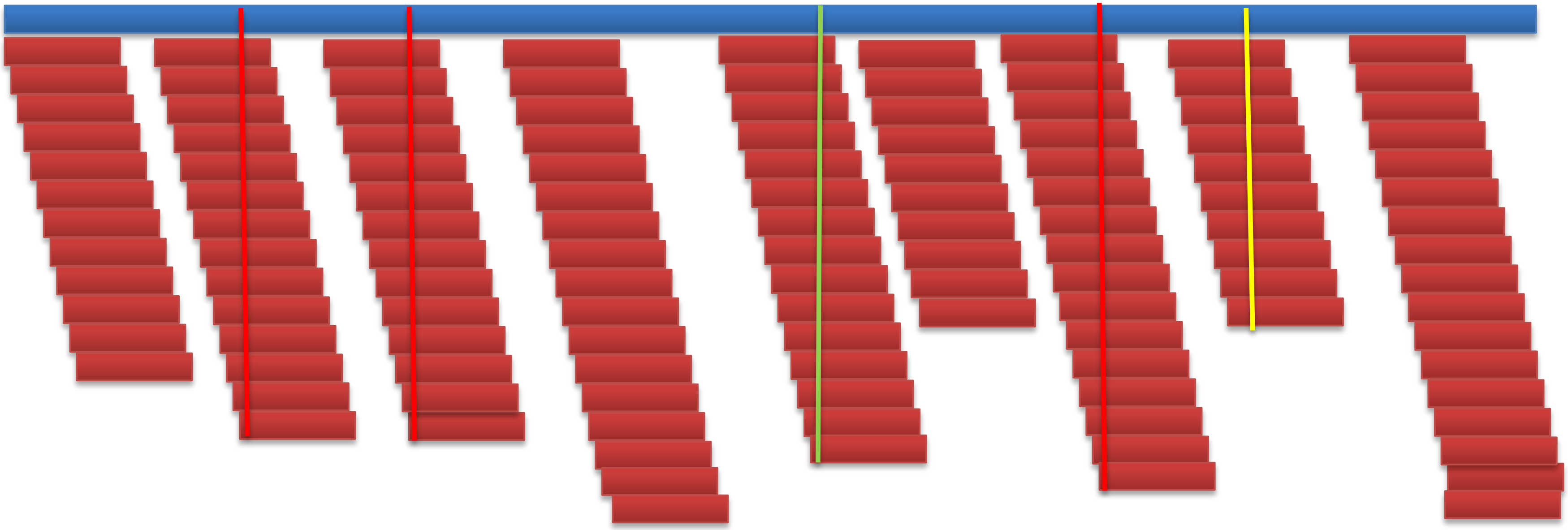
Files

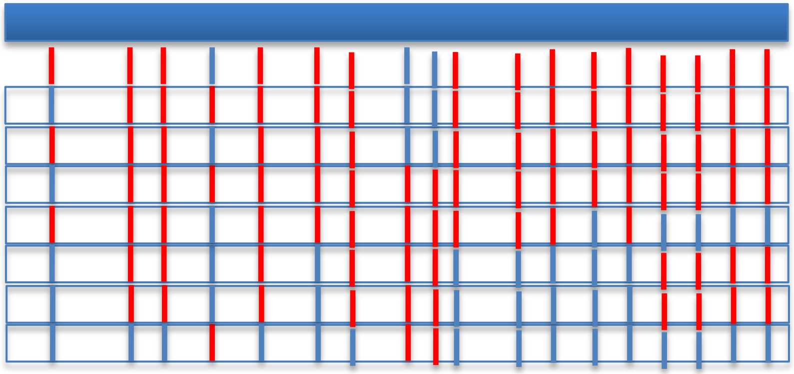
File Name	Date Created
FWS-EC-0288_S8_L001_R1_001.fastq	Mar 03, 2014 12:39:45 PM
FWS-EC-0288_S8_L001_R2_001.fastq	Mar 03, 2014 12:40:24 PM

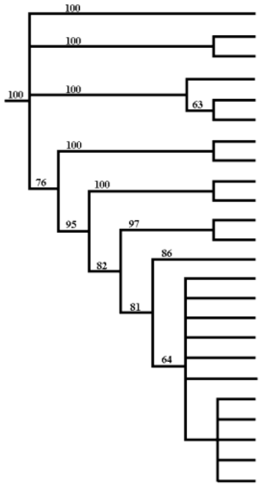
FWS-EC-0288_S8_L001_R2_001.fastq

File Type	Conventional base calls
Date Created	Mar 03, 2014 12:40:24 PM
Date Modified	Mar 03, 2014 12:40:24 PM
Total Sequences	886391
GC Content	50
Maximum Length	301
Minimum Length	35
Filtered Sequences	0

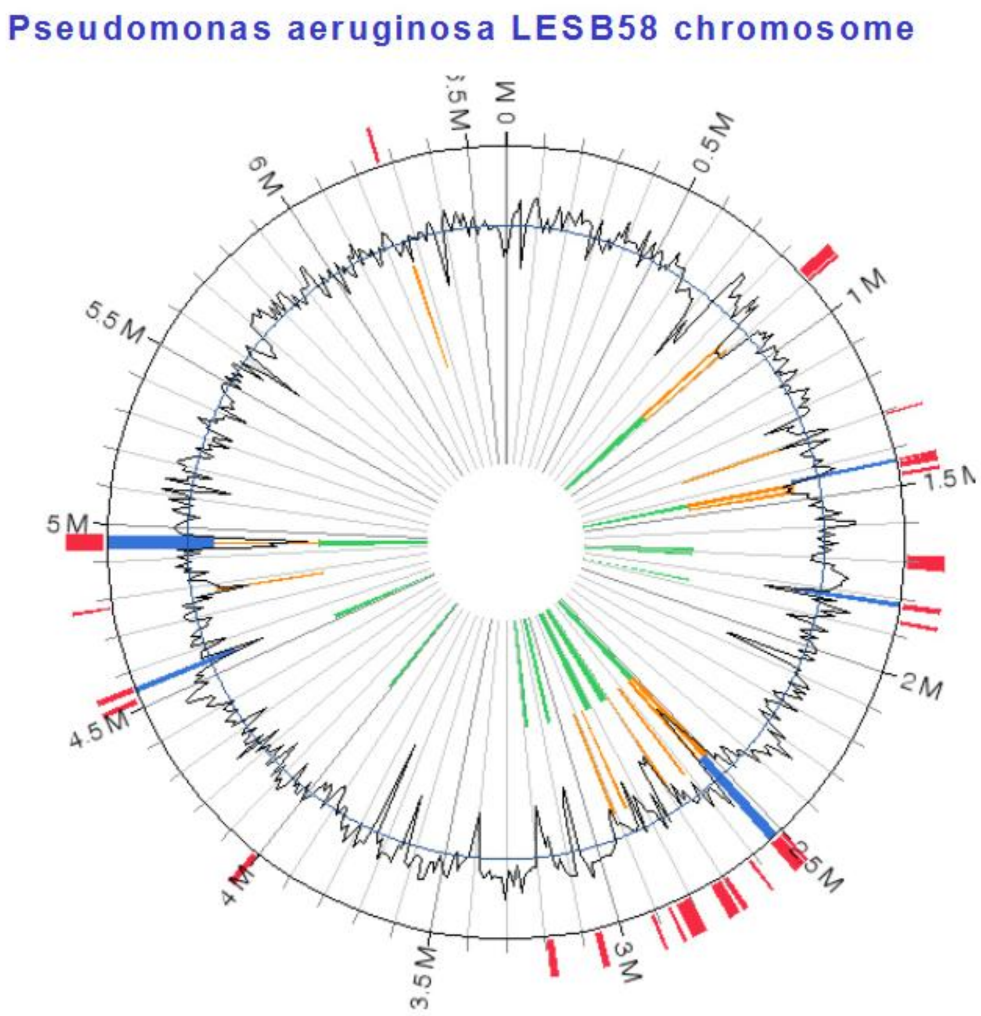
SNP Phylogenomics





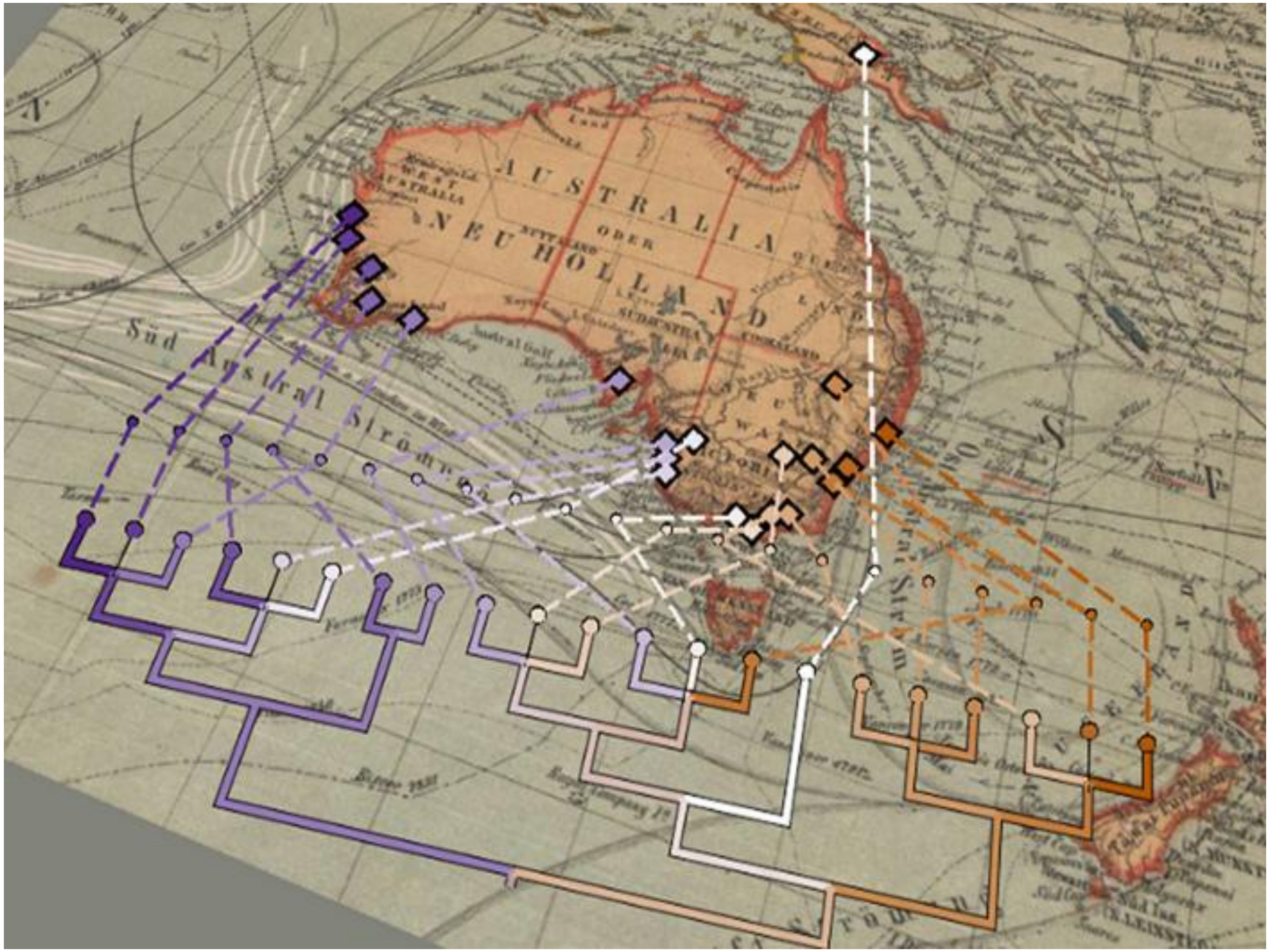


Genomic Island Detection



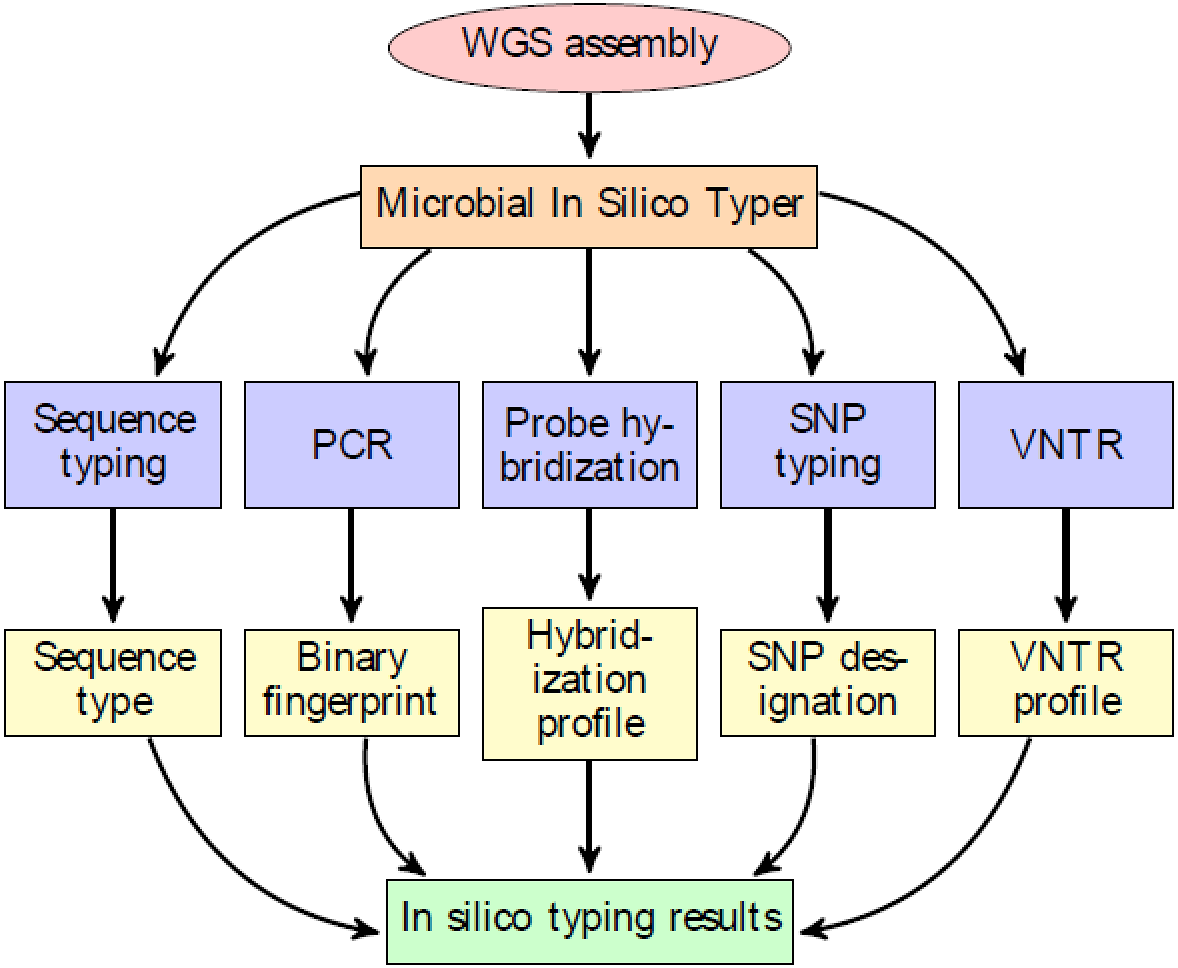
http://www.pathogenomics.sfu.ca/islandviewer

Phylogeography



http://kiwi.cs.dal.ca/GenGIS/Main_Page

Microbial In Silico Typing (MIST) Pipeline



WGS assembly

Microbial In Silico Typer

Sequence typing

PCR

Probe hybridization

SNP typing

VNTR

Sequence type

Binary fingerprint

Hybridization profile

SNP designation

VNTR profile

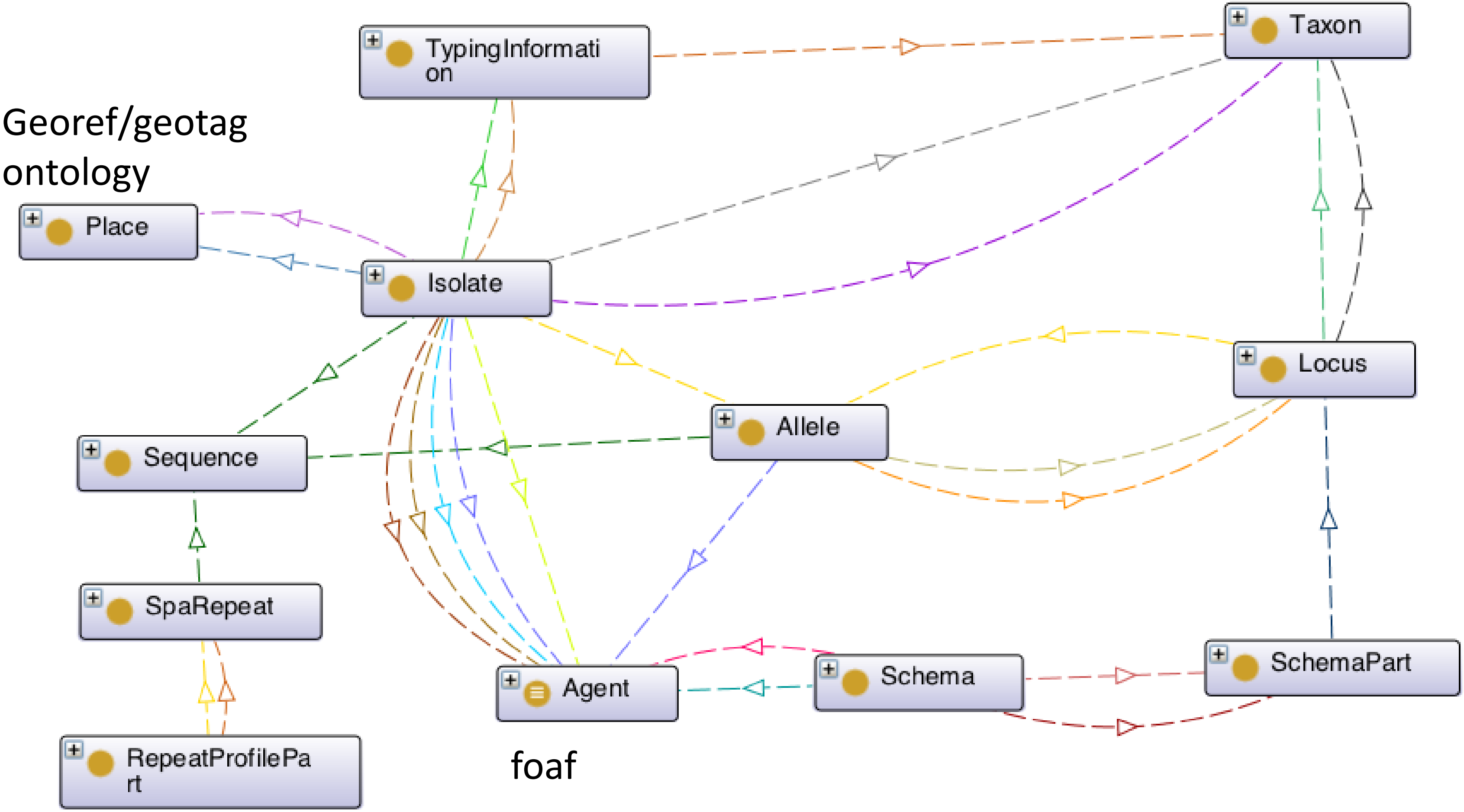
In silico typing results

TyPon

- **24 classes**
- **18 object properties**
- **26 data properties.**
- **5 datatypes**

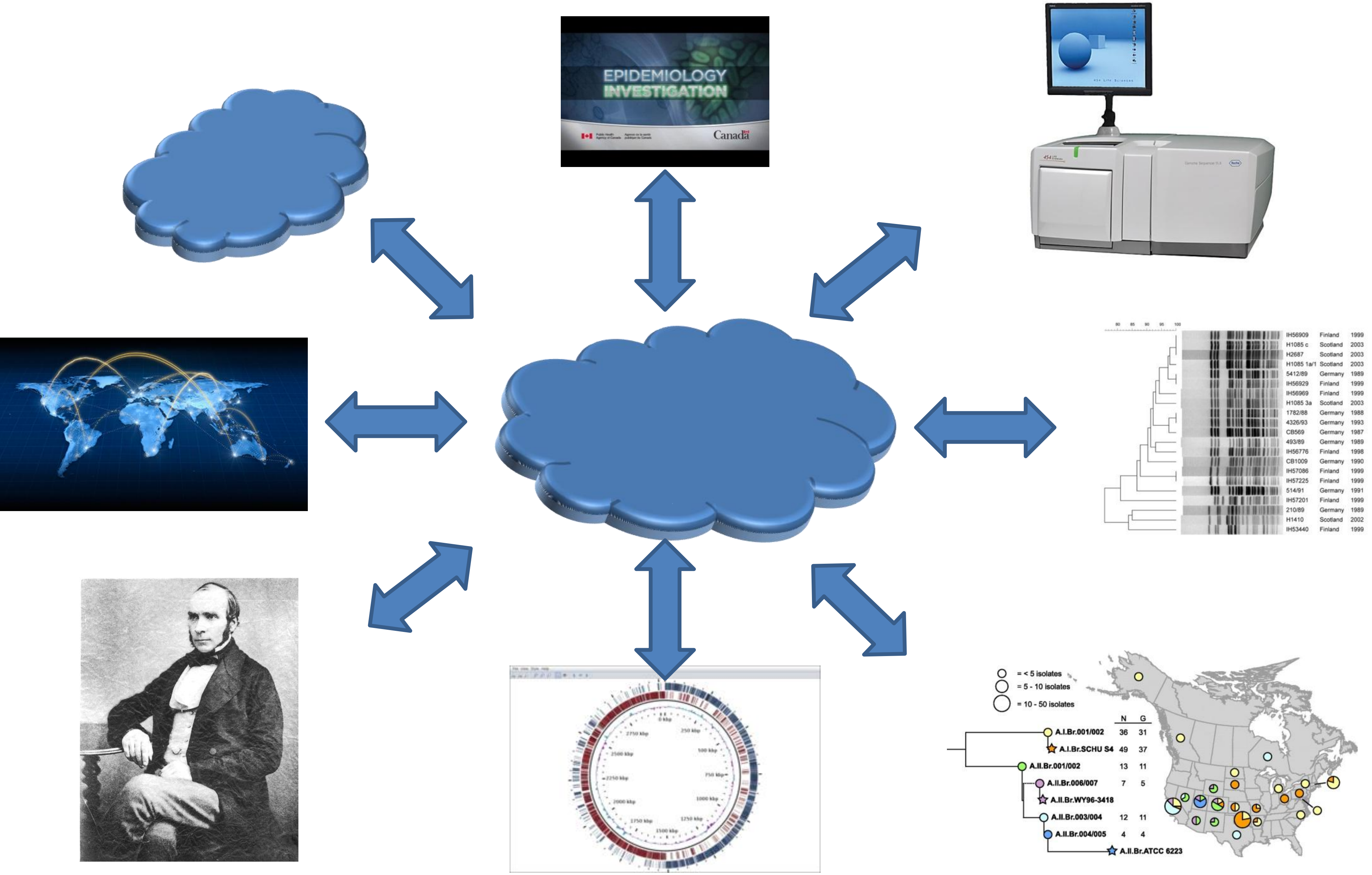
Typon:SBTM

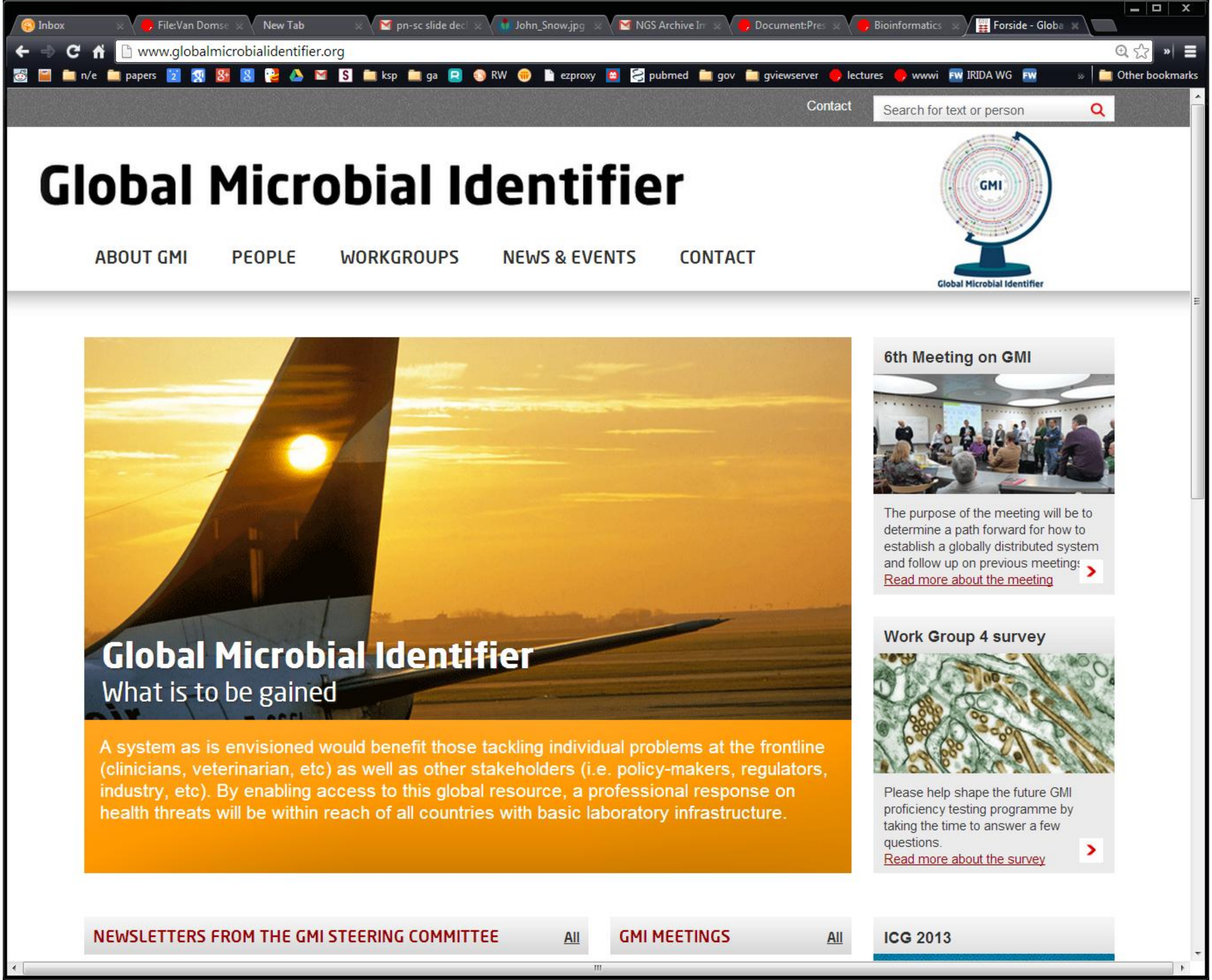
Uniprot



Georef/geotag ontology

foaf


EPIDEMIOLOGY
INVESTIGATION
Canada
80
85
90
95
100
IH56909 Finland 1999
H1085 c Scotland 2003
H2687 Scotland 2003
H1085 1a/1 Scotland 2003
5412/89 Germany 1989
IH56929 Finland 1999
IH56969 Finland 1999
H1085 3a Scotland 2003
1782/88 Germany 1988
4326/93 Germany 1993
CB569 Germany 1987
493/89 Germany 1989
IH56776 Finland 1998
CB1009 Germany 1990
IH57086 Finland 1999
IH57225 Finland 1999
514/91 Germany 1991
IH57201 Finland 1999
210/89 Germany 1989
H1410 Scotland 2002
IH53440 Finland 1999
= < 5 isolates
= 5 - 10 isolates
= 10 - 50 isolates
N G
A.I.Br.001/002 36 31
A.I.Br.SCHU S4 49 37
A.II.Br.001/002 13 11
A.II.Br.006/007 7 5
A.II.Br.WY96-3418
A.II.Br.003/004 12 11
A.II.Br.004/005 4 4
A.II.Br.ATCC 6223
0 kbp
250 kbp
500 kbp
750 kbp
1000 kbp
1250 kbp
1500 kbp
1750 kbp
2000 kbp
2250 kbp
2500 kbp
2750 kbp



Global Microbial Identifier

ABOUT GMI
PEOPLE
WORKGROUPS
NEWS & EVENTS
CONTACT

Global Microbial Identifier

What is to be gained

A system as is envisioned would benefit those tackling individual problems at the frontline (clinicians, veterinarian, etc) as well as other stakeholders (i.e. policy-makers, regulators, industry, etc). By enabling access to this global resource, a professional response on health threats will be within reach of all countries with basic laboratory infrastructure.

6th Meeting on GMI

The purpose of the meeting will be to determine a path forward for how to establish a globally distributed system and follow up on previous meeting

Read more about the meeting

Work Group 4 survey

Please help shape the future GMI proficiency testing programme by taking the time to answer a few questions.

Read more about the survey

NEWSLETTERS FROM THE GMI STEERING COMMITTEE All

GMI MEETINGS All

ICG 2013

Acknowledgments

Bioinformatics:

- Dr. Fiona Brinkman, SFU
- Dr. William Hsiao, BCCDC
- Dr. Robert Beiko
- Dr. Eduardo Taboada
- Dr. João Andre Carrico
- Dr. Ben Liang
- Dr. Natalie Knox
- Dr. Melanie Courtot
- Dr. Lynn Schriml
- Heather Kent
- Eric Enns
- Philip Mabon
- Franklin Bristow
- Thomas Matthews
- Shane Thiessen
- Paul Williams
- Josh Adam
- Aaron Petkau
- Matthew Stuart-Edwards
- Brendan Curran-Johnson
- Jessica Forbes
- Shivani Gupta
- Julie Shay
- Emma Griffiths
- Damion Dooley

Special Thanks:

- Dr. Celine Nadon
- Aleisha Reimer
- Lorilee Tescher
- Dr. Chrystal Berry
- Dr. Matt Gilmour
- Dr. Morag Graham
- Shaun Tyler
- Matt Walker
- Dr. Lee Katz
- Dr. Cheryl Tarr
- Dr. Peter Gerner-Smidt
- Bob Conarroe
- Harvey Ulrich
- Adam Olson