Problem definition

Raquel Tobes

Marina Manrique

Eduardo Pareja-Tobes

2013-08-26

content

- intro
- problem
- background & general concepts
- guidelines
- week plan
- teams

Problem

Design a system to rapidly characterize and identify the pathogen responsible for an outbreak

Based on
NGS and
Cloud Computing

Background & general concepts

- what is an outbreak?

- impact

- outbreaks and NGS

What is an outbreak?

"Outbreak is a term used in epidemiology to describe an occurrence of disease greater than would otherwise be expected at a particular time and place."

Wikipedia

What is an outbreak?

Normally caused by an infectious agent
Virus
Bacteria
Fungi

What is an outbreak?

Healthcare associated
Community acquired

What is an outbreak?

What is an outbreak?

What is an outbreak?

What is an outbreak?

Antibiotic resistant bugs

High-risk clones

What is an outbreak?

CONCISE ARTICLE

H. Linde · F. Wagenlehner · B. Strommenger ·
I. Drubel · J. Tanzer · U. Reischl · U. Raab ·
C. Höller · K. G. Naber · W. Witte · F. Hanses ·
B. Salzberger · N. Lehn

Healthcare-associated outbreaks and community-acquired infections due to MRSA carrying the Panton-Valentine leucocidin gene in southeastern Germany
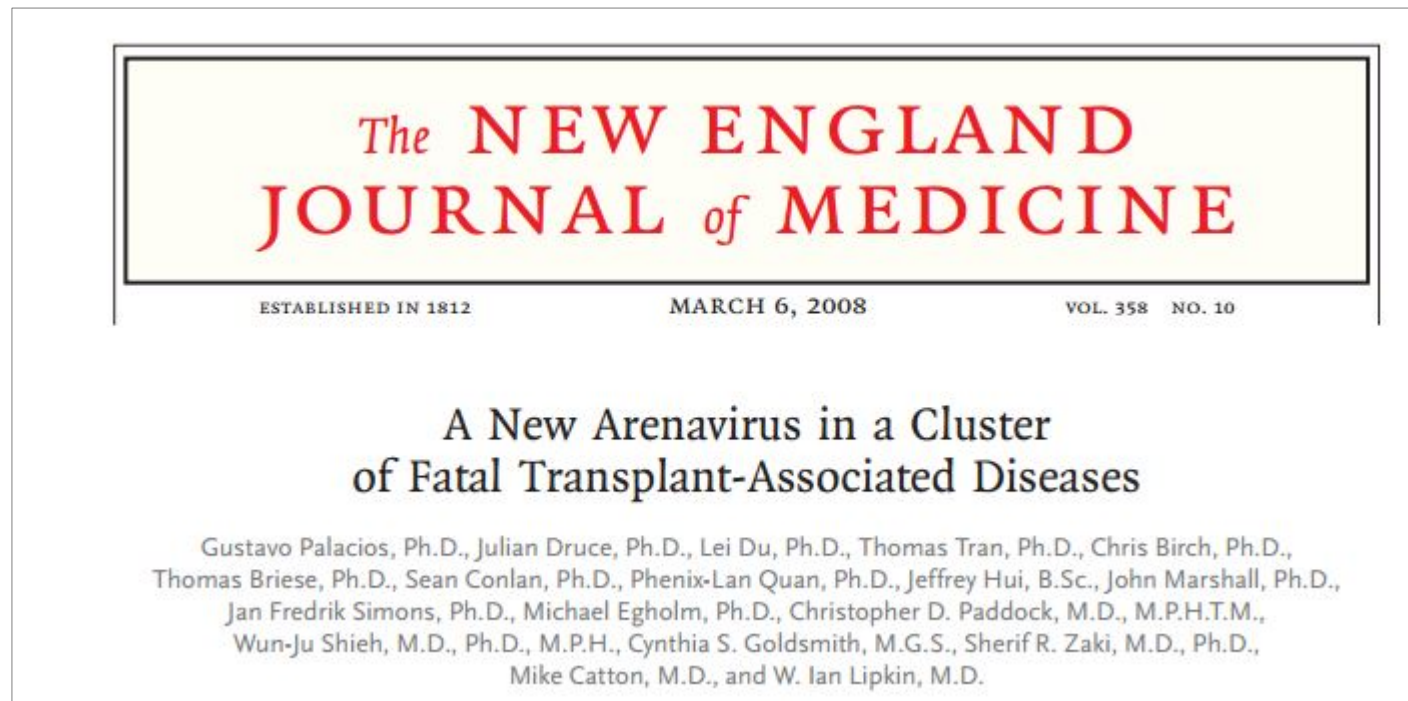
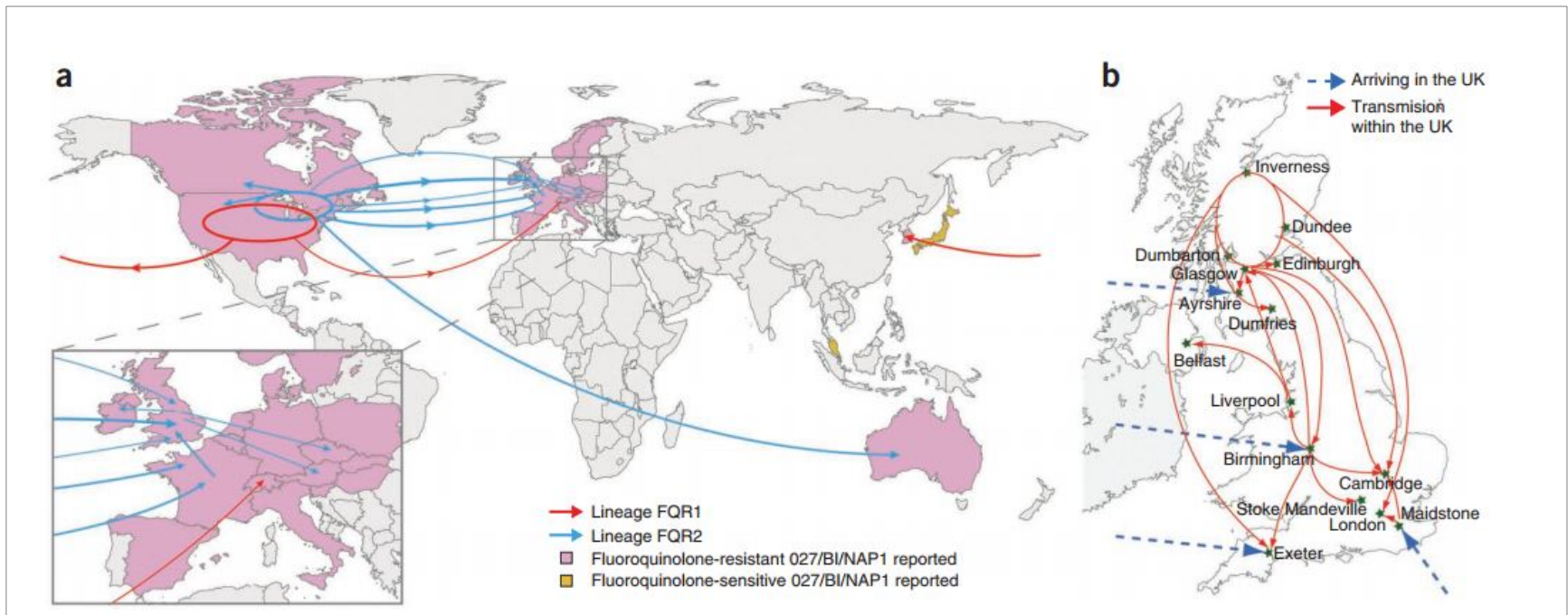# What is an outbreak?

Extension:

Really local (ICU)

Worldwide

What is an outbreak?

Local

The **NEW ENGLAND JOURNAL** *of* **MEDICINE**

ESTABLISHED IN 1812   MARCH 6, 2008   VOL. 358   NO. 10

A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases

Gustavo Palacios, Ph.D., Julian Druce, Ph.D., Lei Du, Ph.D., Thomas Tran, Ph.D., Chris Birch, Ph.D., Thomas Briese, Ph.D., Sean Conlan, Ph.D., Phenix-Lan Quan, Ph.D., Jeffrey Hui, B.Sc., John Marshall, Ph.D., Jan Fredrik Simons, Ph.D., Michael Egholm, Ph.D., Christopher D. Paddock, M.D., M.P.H.T.M., Wun-Ju Shieh, M.D., Ph.D., M.P.H., Cynthia S. Goldsmith, M.G.S., Sherif R. Zaki, M.D., Ph.D., Mike Catton, M.D., and W. Ian Lipkin, M.D.
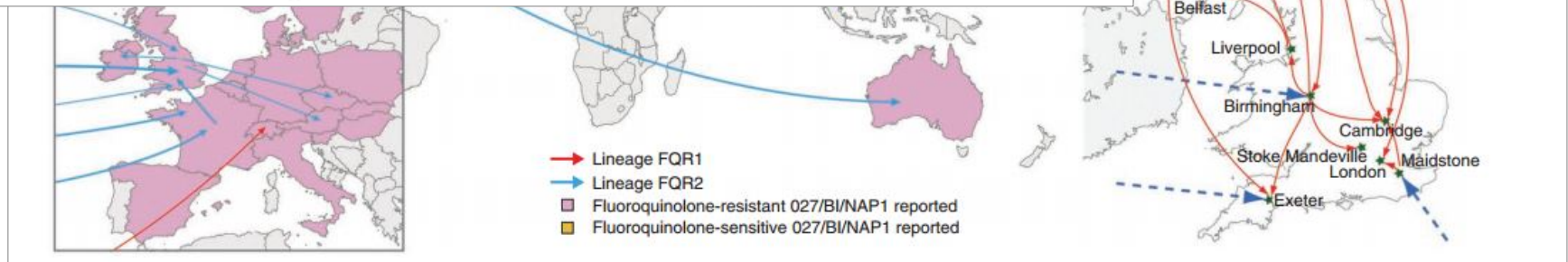
What is an outbreak?

Worldwide

What is an outbreak?

Worldwide

nature
genetics

Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*

Miao He[1], Fabio Miyajima[2,3], Paul Roberts[2,3], Louise Ellison[1], Derek J Pickard[1], Melissa J Martin[4], Thomas R Connor[1], Simon R Harris[1], Derek Fairley[5], Kathleen B Bamford[6,7], Stephanie D'Arc[6,7], Jon Brazier[8], Derek Brown[9], John E Coia[9], Gill Douce[9], Dale Gerding[10], Hee Jung Kim[11], Tse Hsien Koh[12], Haru Kato[13], Mitsutoshi Senoh[13], Tom Louie[14], Stephen Michell[15], Emma Butt[15], Sharon J Peacock[1,16–18], Nick M Brown[17,18], Tom Riley[19], Glen Songer[20], Mark Wilcox[21], Munir Pirmohamed[2,3], Ed Kuijper[22], Peter Hawkey[23], Brendan W Wren[4], Gordon Dougan[1], Julian Parkhill[1] & Trevor D Lawley[1]

- – – → Arriving in the UK
- ——→ Transmission within the UK

→ Lineage FQR1
→ Lineage FQR2
▨ Fluoroquinolone-resistant 027/BI/NAP1 reported
▨ Fluoroquinolone-sensitive 027/BI/NAP1 reported

Impact

A real impact on public health

Impact

Avian flu H1N1

CDC estimates

- 61 million people infected with 2009 H1N1
- 274,000 2009 H1N1-related hospitalizations
- **12,470 2009 H1N1-related deaths**

http://www.cdc.gov/h1n1flu/estimates_2009_h1n1.htm

Impact

Multistate Fungal Meningitis Outbreak

Current outbreak in the States

Healthcare associated infection. Non contagious

http://www.cdc.gov/hai/outbreaks/meningitis.html

Impact

Multistate Fungal Meningitis Outbreak

At-A-Glance

- Status: Ongoing Investigation
- Infection: Fungal
- Facility Type: Outpatient Setting
- Case Count: 749*
- States: 20*
- Deaths: 63*
- Laboratory Results

* Case counts will be updated September 5, 2013.

Cases and Deaths with Fungal Infections Linked to Steroid Injections

| State	Total Case Count	Meningitis Only	Meningitis + Paraspinal/Spinal Infection	Stroke w/out Lumbar Puncture Only	Paraspinal/Spinal Infection only	Peripheral Joint Infection Only	Paraspinal/Spinal Infection + Peripheral Joint Infection	Deaths
Florida (FL)	25	22	1	1	1	0	0	7
Georgia (GA)	1	1	0	0	0	0	0	0
Idaho (ID)	1	1	0	0	0	0	0	0
Illinois (IL)	2	2	0	0	0	0	0	0
Indiana (IN)	91	30	17	1	43	0	0	11
Maryland (MD)	26	23	1	0	2	0	0	3
Michigan (MI)	264	23	46	2	166	25	2	19
Minnesota (MN)	12	10	0	0	2	0	0	1
North Carolina (NC)	18	1	3	0	14	0	0	1
New Hampshire (NH)	14	9	0	0	0	5	0	0
New Jersey (NJ)	51	30	11	0	9	1	0	0
New York (NY)	1	0	0	0	1	0	0	0
Ohio (OH)	20	12	3	0	5	0	0	1
Pennsylvania (PA)	1	1	0	0	0	0	0	0
Rhode Island (RI)	3	1	1	0	1	0	0	0
South Carolina (SC)	3	2	0	0	1	0	0	0
Tennessee (TN)	153	22	57	3	69	2	0	15
Texas (TX)	2	2	0	0	0	0	0	0
Virginia (VA)	54	41	9	0	4	0	0	5
West Virginia (WV)	7	0	2	0	5	0	0	0
TOTAL	**749**	**233**	**151**	**7**	**323**	**33**	**2**	**63**

*Deaths reported are from all causes among persons who meet the case definition and may not be directly attributed to a fungal infection.

Case counts by state are based on the state where the procedure was performed, not the state of residence.

Persons with Fungal Infections Linked to Steroid Injections, by State
AUGUST 5, 2013 3:30 PM EST
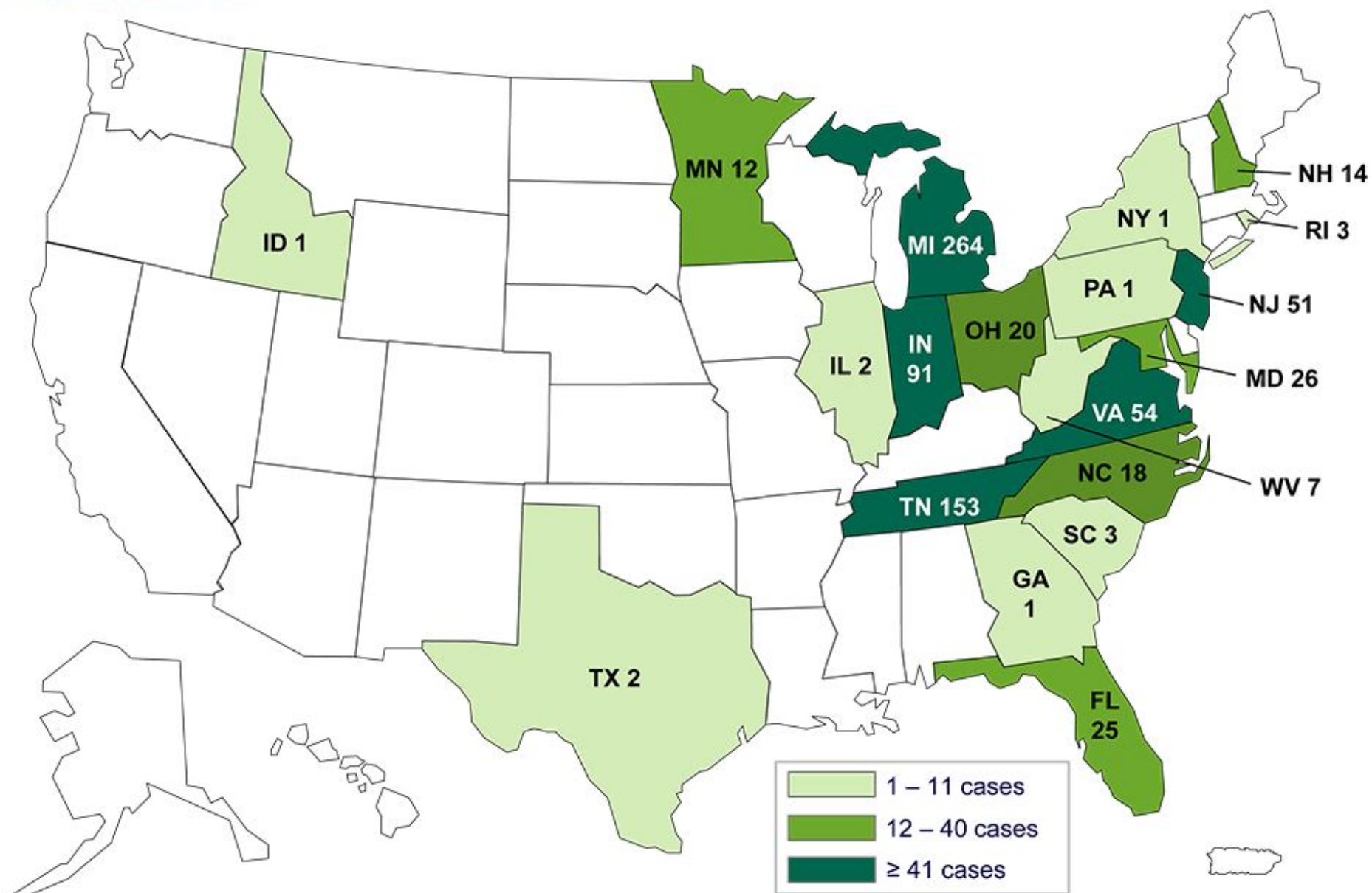
Impact

Multistate Fungal Meningitis Outbreak

Detection of Fungal DNA in Human Body Fluids and Tissues during a Multistate Outbreak of Fungal Meningitis and Other Infections

Lalitha Gade,[a] Christina M. Scheel,[a] Cau D. Pham,[a] Mark D. Lindsley,[a] Naureen Iqbal,[a] Angela Ahlquist Cleveland,[a] Anne M. Whitney,[b] Shawn R. Lockhart,[a] Mary E. Brandt,[a] Anastasia P. Litvintseva[a]

Mycotic Diseases Branch[a] and Bacterial Special Pathogens Branch,[b] National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

PCR and sequencing. Three primer pairs were used for nucleic acid amplification. (i) Broad-spectrum fungal primers ITS3 and ITS4 anneal within the conserved regions of 5.8S and 28S ribosomal DNA (rDNA) genes and amplify an ~350-bp fragment that includes the ITS2 region (ITS3, 5′-GCATCGATGAAGAACGCAGC; ITS4, 5′-TCCTCCGCTTAT TGATATGC) (24, 26). (ii) *Exserohilum*-specific primers were developed for this investigation and amplify the variable 230-bp region of ITS2 (Exs4F, 5′-GAAGAACGCAGCGAAATGCG; Exs4R, 5′-CCGAAAACCA GTAGGTCGGC). (iii) Positive control primers Beta2/Beta3 that amplify portions of the human β-globin gene (Beta2 [GH20], 5′-GAAGAGCCA AGGACAGGTAC; Beta3 [PC04], 5′-CAACTTCATCCACGTTCACC) (27). Each sample was processed using all three PCR primer pairs.

Impact

Multistate Fungal Meningitis Outbreak

This is *just* a real case of how an outbreak is tracked

Impact

Multistate Fungal Meningitis Outbreak

This is *just* a real case of how an outbreak is tracked

Where the detection is based on PCR and sequencing

But what if...

Impact

Multistate Fungal Meningitis Outbreak

-The agent is not well characterized? We don't have this set of primers or they don't work nicely

- The agent is spreading (even) much faster than expected? We need quicker ways to identify it

Impact

NGS and Cloud
Computing
could help here?

Outbreaks and NGS

Quickly characterize the pathogen and identify it

Outbreaks and NGS

Quickly characterize the pathogen and identify it

That's a fact today

Outbreaks and NGS

Modern clinical microbiology: new challenges and solutions

Pierre-Edouard Fournier, Michel Drancourt, Philippe Colson, Jean-Marc Rolain, Bernard La Scola and Didier Raoult

Abstract | In the twenty-first century, the clinical microbiology laboratory plays a central part in optimizing the management of infectious diseases and surveying local and global epidemiology. This pivotal role is made possible by the adoption of rational sampling, point-of-care tests, extended automation and new technologies, including mass spectrometry for colony identification, real-time genomics for isolate characterization, and versatile and permissive culture systems. When balanced with cost, these developments can improve the workflow and output of clinical microbiology laboratories and, by identifying and characterizing microbial pathogens, provide significant input to scientific discovery.

Outbreaks and NGS

Modern clinical microbiology: new challenges and solutions

Pierre-Edouard Fournier, Michel Drancourt, Philippe Colson, Jean-Marc Rolain, Bernard La Scola and Didier Raoult

Abstract | In the twenty-first century, the clinical microbiology laboratory plays a central part in optimizing the management of infectious diseases and surveying local and global epidemiology. This pivotal role is made possible by the adoption of rational sampling, point-of-care tests, extended automation and new technologies, including mass spectrometry for colony identification, real-time genomics for isolate characterization, and versatile and permissive culture systems. When balanced with cost, these developments can improve the workflow and output of clinical microbiology laboratories and, by identifying and characterizing microbial pathogens, provide significant input to scientific discovery.

Outbreaks and NGS

Microorganism	Location	Year	Reference
Carbapenem-resistant *Klebsiella pneumoniae*	USA	2011	112
Clostridium difficile	Worldwide	2013	113
Escherichia coli O104:H4	Germany	2011	114,115
Legionella pneumophila serogroup 1	United Kingdom	2013	116
Methicillin-resistant *Staphylococcus aureus* (MRSA)	United Kingdom	2009	117
Mycobacterium tuberculosis	Canada	2006–2008	118
Vibrio cholerae O1 biovar El Tor	Haiti	2010–2011	119
Arenavirus	Australia	2008	120
Bas-Congo virus	Democratic Republic of the Congo	2009	121
Influenza A virus H1N1	Worldwide	2009	122

Outbreaks and NGS

EHEC German outbreak 2011

A real example of how NGS was used for characterizing the pathogen and designing primers to detect it
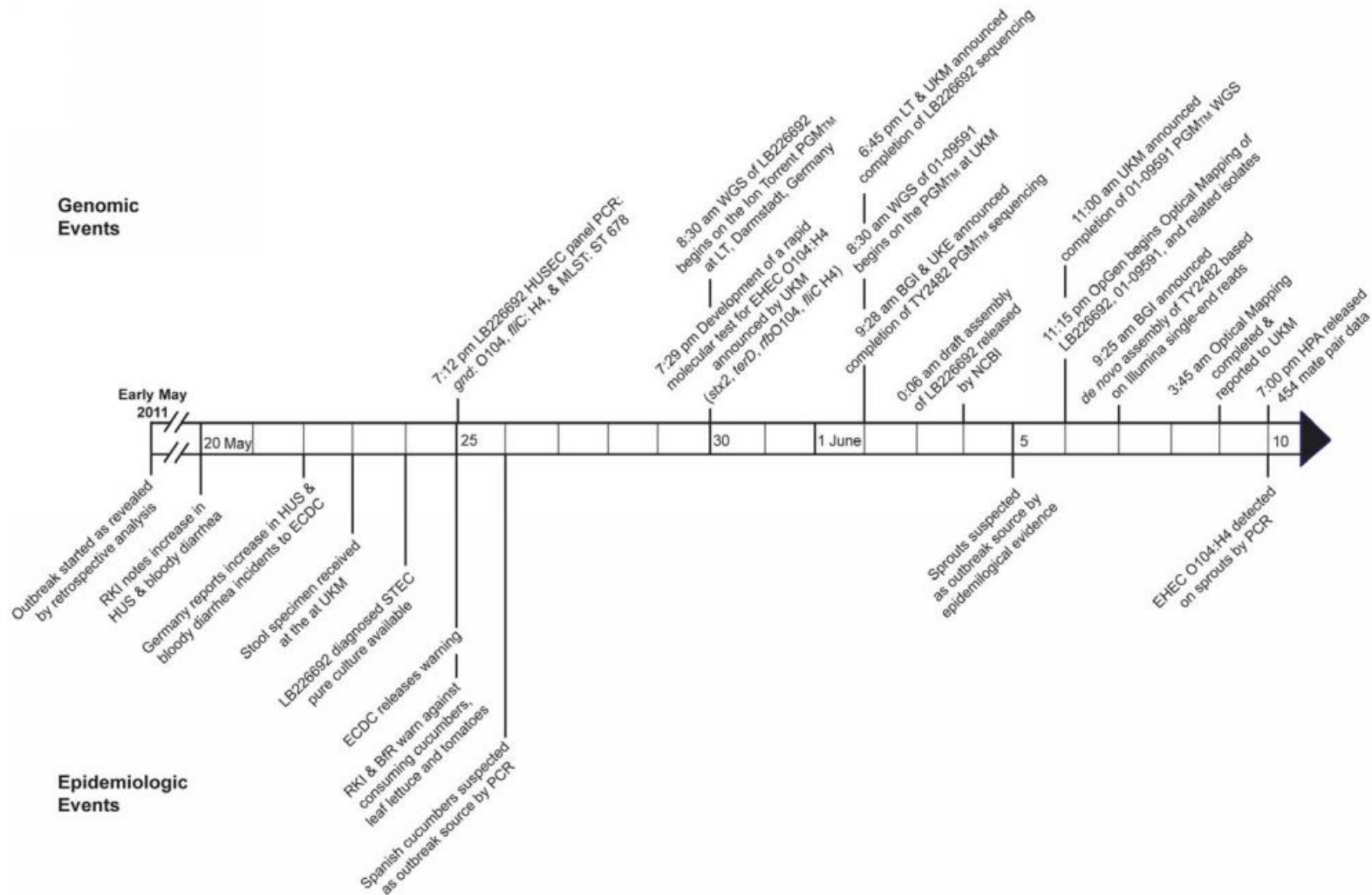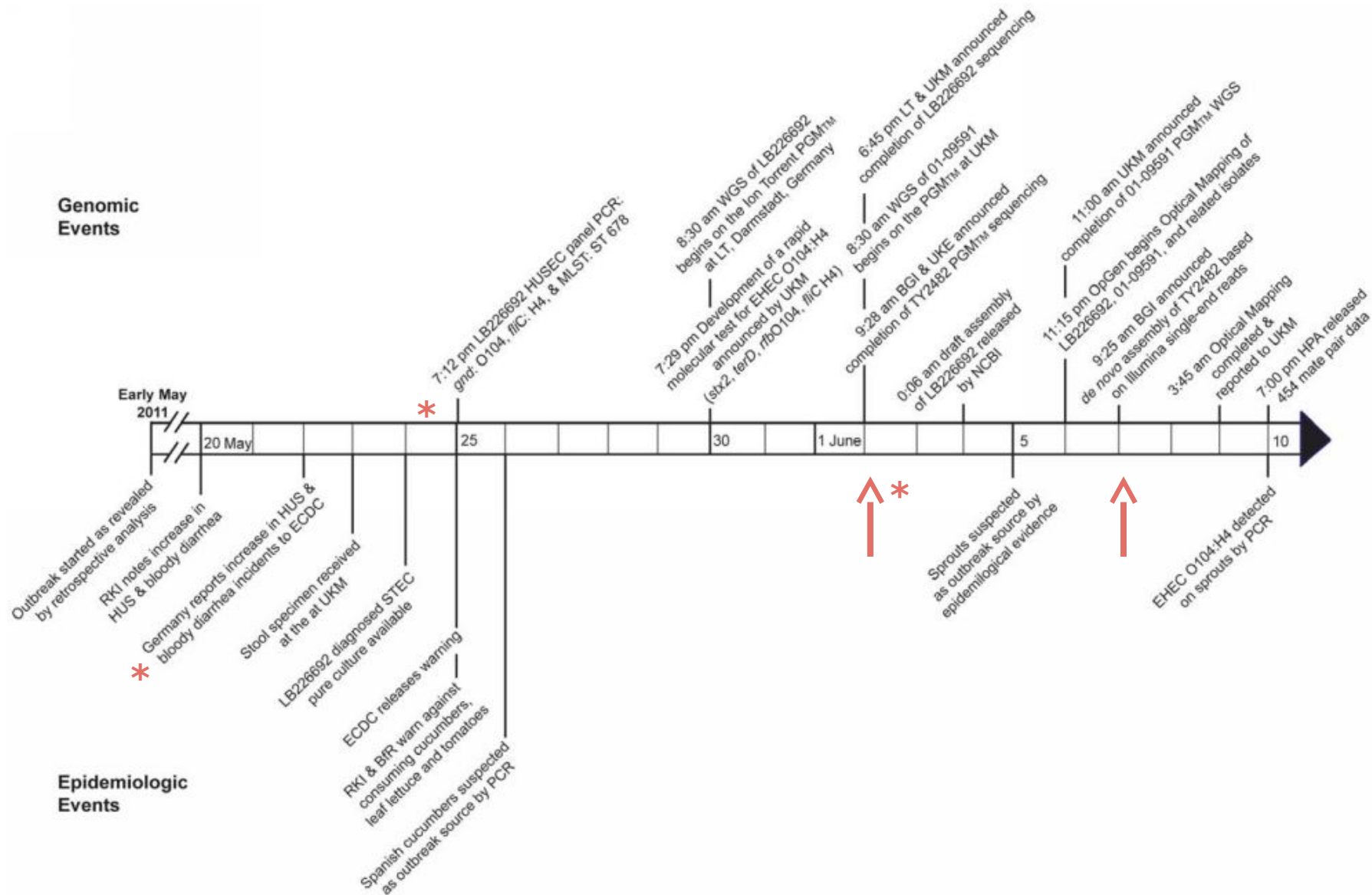
Outbreaks and NGS

Prospective Genomic Characterization of the German Enterohemorrhagic *Escherichia coli* O104:H4 Outbreak by Rapid Next Generation Sequencing Technology

Alexander Mellmann[1,9], Dag Harmsen[2*,9], Craig A. Cummings[3,9], Emily B. Zentz[4], Shana R. Leopold[1], Alain Rico[5], Karola Prior[2], Rafael Szczepanowski[2], Yongmei Ji[3], Wenlan Zhang[1], Stephen F. McLaughlin[3], John K. Henkhaus[4], Benjamin Leopold[1], Martina Bielaszewska[1], Rita Prager[6], Pius M. Brzoska[3], Richard L. Moore[4], Simone Guenther[5], Jonathan M. Rothberg[7], Helge Karch[1]

1 Institute of Hygiene, University Münster, Münster, Germany, 2 Department of Periodontology, University Münster, Münster, Germany, 3 Life Technologies, Foster City, California, United States of America, 4 OpGen, Gaithersburg, Maryland, United States of America, 5 Life Technologies, Darmstadt, Germany, 6 Robert Koch Institute, Wernigerode Branch, Wernigerode, Germany, 7 Ion Torrent by Life Technologies, Guilford, Connecticut, United States of America

Figure 1. Events timeline of German EHEC O104:H4 outbreak. Major events relating to the outbreak epidemiology (below arrow) and those

Genomic Events

- 7:12 pm LB226692 HUSEC panel PCR: *gnd*: O104, *fliC*: H4, & MLST: ST 678
- 8:30 am WGS of LB226692 begins on the Ion Torrent PGM™ at LT, Darmstadt, Germany
- 7:29 pm Development of a rapid molecular test for EHEC O104:H4 announced by UKM (*stx2*, *terD*, *rfbO*104, *fliC* H4)
- 6:45 pm LT & UKM announced completion of LB226692 sequencing
- 8:30 am WGS of 01-09591 begins on the PGM™ at UKM
- 9:28 am BGI & UKE announced completion of TY2482 PGM™ sequencing
- 0:06 am draft assembly of LB226692 released by NCBI
- 11:00 am UKM announced completion of 01-09591 PGM™ WGS
- 11:15 pm OpGen begins Optical Mapping of LB226692, 01-09591, and related isolates
- 9:25 am BGI announced de novo assembly of TY2482 based on Illumina single-end reads
- 3:45 am Optical Mapping completed & reported to UKM
- 7:00 pm HPA released 454 mate pair data

Early May 2011

20 May | 25 | 30 | 1 June | 5 | 10

Epidemiologic Events

- Outbreak started as revealed by retrospective analysis
- RKI notes increase in HUS & bloody diarrhea
- Germany reports increase in HUS & bloody diarrhea incidents to ECDC
- Stool specimen received at the at UKM
- LB226692 diagnosed STEC pure culture available
- ECDC releases warning
- RKI & BfR warn against consuming cucumbers, leaf lettuce and tomatoes
- Spanish cucumbers suspected as outbreak source by PCR
- Sprouts suspected as outbreak source by epidemiological evidence
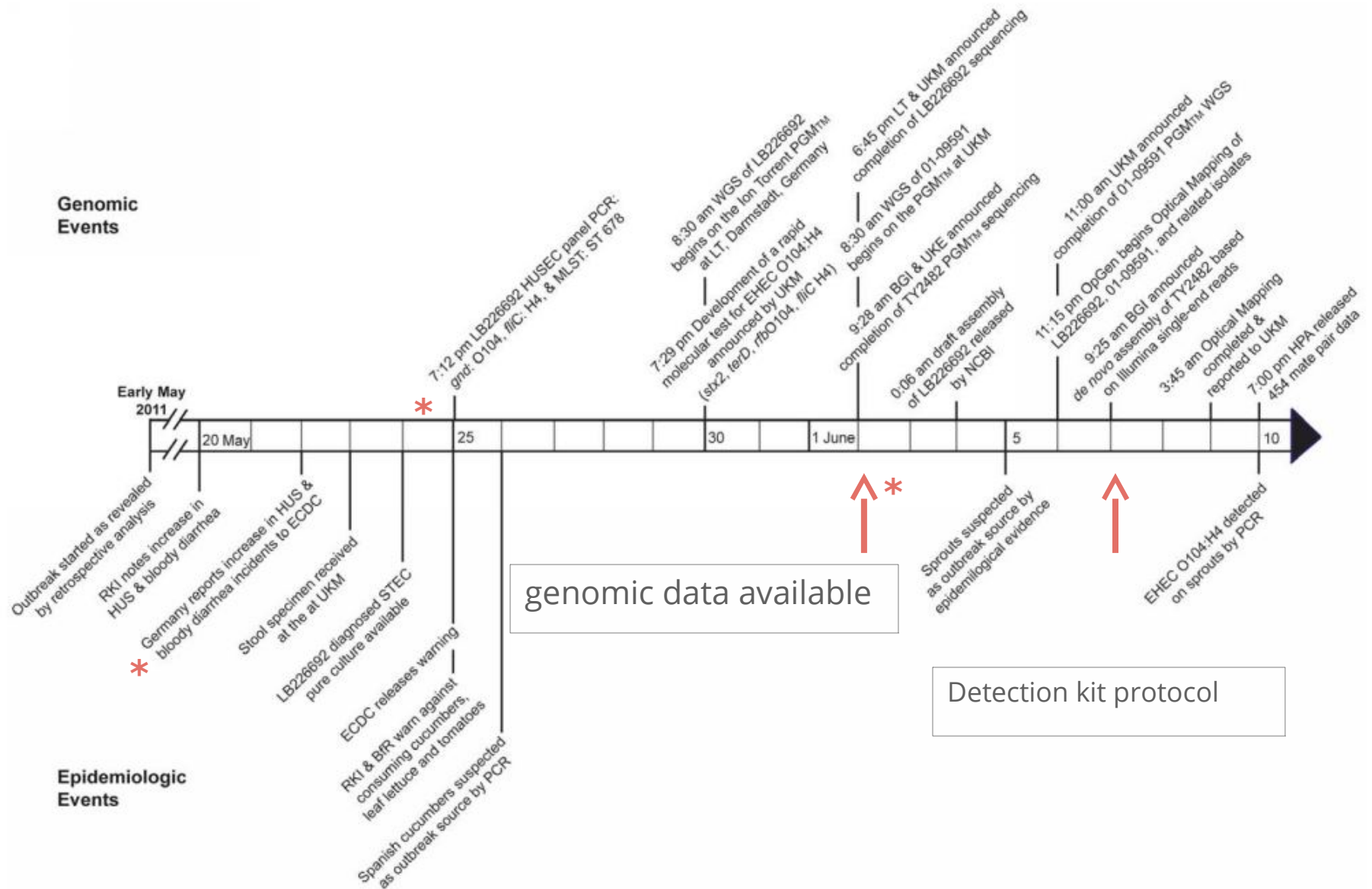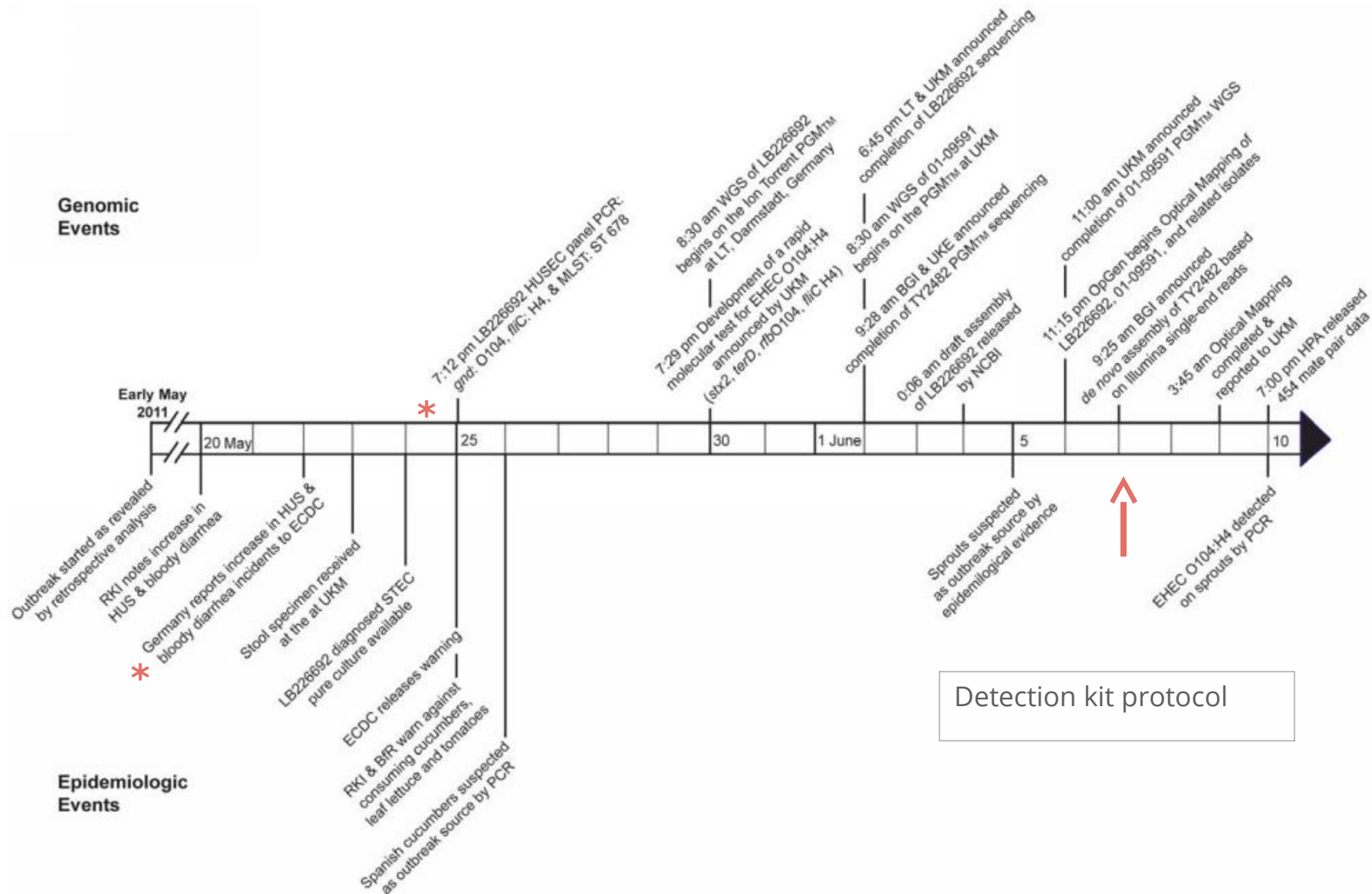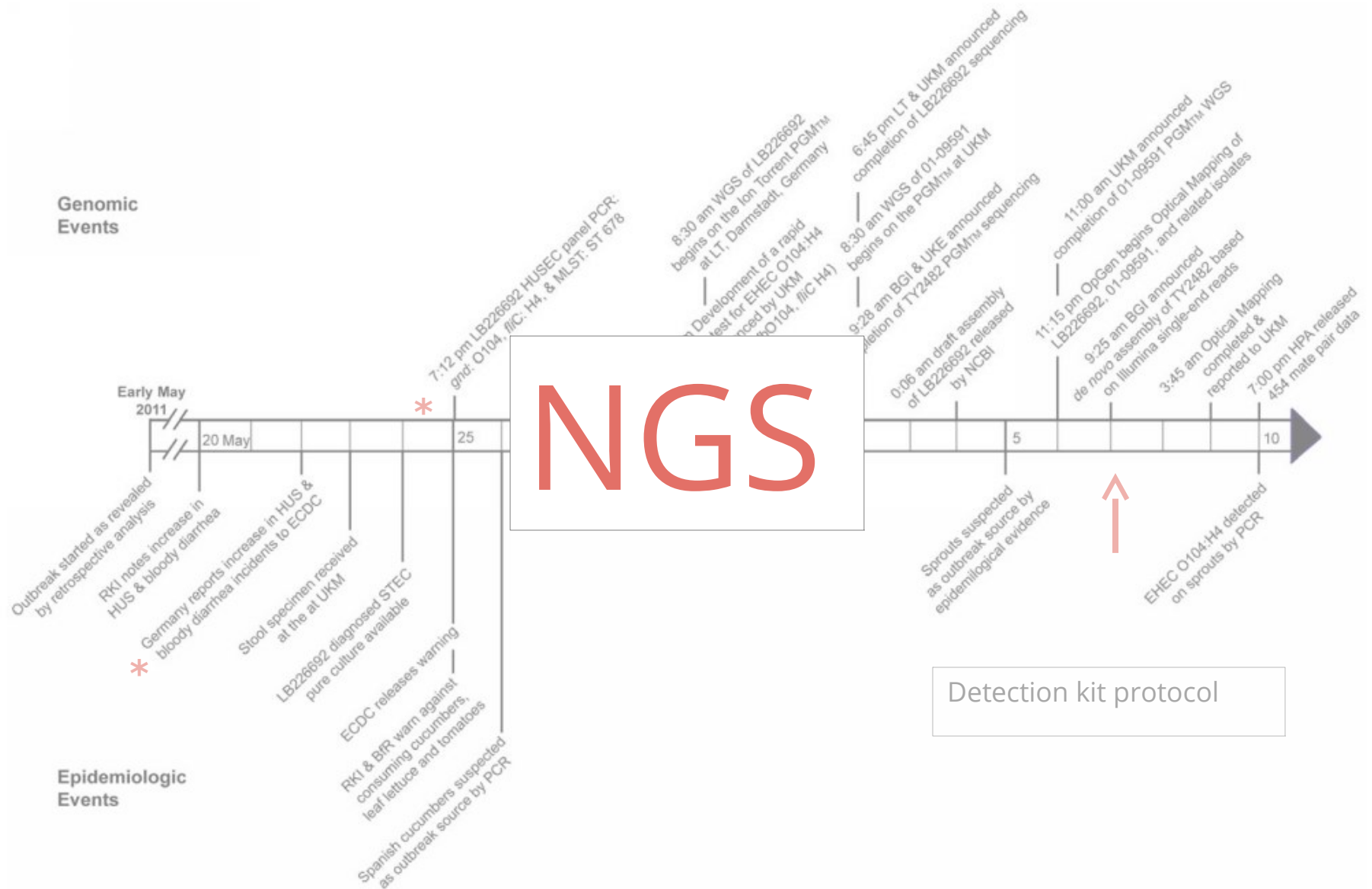- EHEC O104:H4 detected on sprouts by PCR

Figure 1. Events timeline of German EHEC O104:H4 outbreak. Major events relating to the outbreak epidemiology (below arrow) and those

Figure 1. Events timeline of German EHEC O104:H4 outbreak. Major events relating to the outbreak epidemiology (below arrow) and those

Genomic Events

7:12 pm LB226692 HUSEC panel PCR: *gnd*: O104, *fliC*: H4, & MLST: ST 678

8:30 am WGS of LB226692 begins on the Ion Torrent PGM™ at LT, Darmstadt, Germany

7:29 pm Development of a rapid molecular test for EHEC O104:H4 announced by UKM (*stx2*, *terD*, *rfb*O104, *fliC* H4)

6:45 pm LT & UKM announced completion of LB226692 sequencing

8:30 am WGS of 01-09591 begins on the PGM™ at UKM

9:28 am BGI & UKE announced completion of TY2482 PGM™ sequencing

0:06 am draft assembly of LB226692 released by NCBI

11:00 am UKM announced completion of 01-09591 PGM™ WGS

11:15 pm OpGen begins Optical Mapping of LB226692, 01-09591, and related isolates

9:25 am BGI announced de novo assembly of TY2482 based on Illumina single-end reads

3:45 am Optical Mapping completed & reported to UKM

7:00 pm HPA released 454 mate pair data

Early May 2011

20 May 25 30 1 June 5 10

*

*

Outbreak started as revealed by retrospective analysis

RKI notes increase in HUS & bloody diarrhea

Germany reports increase in HUS & bloody diarrhea incidents to ECDC

Stool specimen received at the at UKM

LB226692 diagnosed STEC pure culture available

ECDC releases warning

RKI & BfR warn against consuming cucumbers, leaf lettuce and tomatoes

Spanish cucumbers suspected as outbreak source by PCR

Sprouts suspected as outbreak source by epidemiological evidence

EHEC O104:H4 detected on sprouts by PCR

Detection kit protocol

Epidemiologic Events

Figure 1. Events timeline of German EHEC O104:H4 outbreak. Major events relating to the outbreak epidemiology (below arrow) and those

Genomic Events

Early May 2011

Epidemiologic Events

NGS

Detection kit protocol

Figure 1. Events timeline of German EHEC O104:H4 outbreak. Major events relating to the outbreak epidemiology (below arrow) and those

Genomic characterization of the EHEC O104:H4 outbreak strain

Sequencing on the Ion Torrent PGM^{TM} sequencer was completed within 62 hours, leading to the public release of the draft assembly of outbreak strain LB226692 on June 3 (Fig. 1, Table S1). Sequence data of the closely related historical isolate 01-09591 was also generated while the outbreak was still occurring. Genome assemblies based on the PGM^{TM} reads showed that both of these HUS-causing strains (LB226692 and 01-09591) carry genes typically found in two different *E. coli* pathotypes, specifically EAEC and EHEC. Genome wide phylogenetic analysis based on core chromosomal ORFs (n = 1,144) demonstrated the close relationship of the LB226692 and 01-09591 strains to the previously sequenced EAEC strain 55989 (NCBI acc. no. NC_011478), and indicated that these strains are only distantly related to the commonly isolated EHEC serotypes (Fig. 3). However, unlike typical EAEC strains, both LB226692 and 01-09591 have an *stx$_2$*-harboring prophage integrated in *wrbA*, which is also the integration site for *stx$_2$*-phages in EHEC O157:H7 outbreak strains EDL933 [13] and Sakai (RIMD 0509952) [14]. The *wrbA* gene of EAEC 55989 is not occupied by a prophage. Furthermore, the IrgA homologue adhesin encoding gene (*iha*), which is responsible for adherence to epithelial cells and has been found in *eae*-negative STEC [15], is present in all three strains. In contrast to the two HUSEC041 complex strains, 55989 does not harbor the tellurite resistance encoding genes (*ter*). These characteristics led to the development of a rapid PCR-based test of *stx2*, O104 lipopolysaccharide (LPS) gene (*rfb$_{O104}$*), H4 flagellin-encoding gene (*fliC$_{H4}$*), and *terD* for the detection of the HUSEC041 complex [16].

http://bgiamericas.com/bgi-releases-a-complete-de-novo-e-coli-o104-genome-assembly-and-is-making-their-detection-kit-protocols-and-synthesized-primers-freely-available-to-worldwide-disease-control-and-research-agencies/

Outbreaks and NGS

EHEC German outbreak 2011

Detection kit protocols and synthesized primers available

In 16 days from the announcement of the outbreak

In only 6 days from the release of the first genomic data

Outbreaks and NGS

So yes,

NGS may be (and is) useful in this field in at least to steps:

1. Genome Characterization
2. Pathogen detection

Outbreaks and NGS

What about Cloud Computing?

Guidelines

NGS: your design should have an answer for

Rapid characterization of the pathogen causing the outbreak

- How would you characterize the pathogen? De novo assembly of the genome + annotation? MLST? Searching for virulence proteins exclusively? All, none?

- How quick you can do it once you have the sequences?

- Your design is scalable?

Guidelines

NGS: your design should have an answer for

Rapid identification of the pathogen in samples

- You have the agent characterized, how you'd identify it?

- The identification would be based on PCR? Whole genome sequencing?

Guidelines

Some general questions you should address in the design

- Which sequencing technology (or combination of them) would you use?

- Which samples requirements you would have? Could you work with clinical samples? really low DNA quantity with poor quality? Would you need a prior phase of pathogen isolation and growth?

- How long the whole process would take (wet lab + data analysis)?

Guidelines

The more realistic and detailed the better.

It's a design and research task.

There's not a unique correct solution for it

Week plan

	Mon 26	Tue 27	Wed 28	Thu 29	Fri 30
10:00 - 11:00	1 **T** welcome	6 **T/P** problem	11 **T** arch	16 **P** Q&A III	21 **P** present
11:00 - 11:30	*break*	*break*	*break*	*break*	*break*
11:30 - 12:30	2 **T** issues	7 **T** NGS	12 **P** nispero	17 **P** TW III	22 **P** present
12:30 - 14:00	*lunch*	*lunch*	*lunch*	*lunch*	*lunch*
14:00 - 15:30	3 **T** cloud what?	8 **P** statika	13 **P** bio4j	18 **P** TW IV	23 conclussions
15:30 - 15:45	*break*	*break*	*break*	*break*	--
15:45 - 16:45	4 **P** AWS I	9 **P** Q&A I	14 **P** Q&A II	19 **P** Q&A IV	--
16:45 - 17:15	*break*	*break*	*break*	*break*	--
17:15 - ??:??	5 **P** AWS II	10 **P** TW I	15 **P** TW II	20 **P** TW V	--

Teams

Team 1
Kim
Om
Andrea
Habib

Team 2
Fabian
Lizzy
Alexandre
Alexandra
Vedran

Team 3
Alexey
Jeannine
Jasmin
Somya

Data sets