COllaborative Management Platform for detection and Analyses of (Re-) emerging and foodborne outbreaks in Europe

Infectious disease detection in the era of next generation sequences: opportunities, challenges, and the COMPARE project

Prof. Marion Koopmans, (Erasmus Medical Center, the Netherlands)
@MarionKoopmans
Prof Frank Aarestrup (DTU, Denmark)

This project has received funding from the *European Union's Horizon 2020 research and innovation program* under grant agreement 643476

Severe Respiratory Illness Associated with Enterovirus D68 – Multiple States, 2014

Recommend Tweet Share

HAN HEALTH ALERT NETWORK

Distributed via the CDC Healt

September 12, 2014, 17:00 ET

CDCHAN-00369

This is an official
CDC HEALTH ADVISORY

Enterovirus 68

MERS CoV

World Health Organization

Middle East respiratory syndrome coronavirus (MERS-CoV)

13 June 2014

Update on MERS-CoV transmission from animals to humans, and interim recommendations for at-risk groups

Avian influenza
H7N9 China
H5N1 Egypt

Statement on the 1st meeting of the IHR Emergency Committee on the 2014 Ebola outbreak in West Africa

WHO statement
8 August 2014

WHO statement on the meeting of the International Health Regulations Emergency Committee concerning the international spread of wild poliovirus

WHO statement
5 May 2014

Infectious disease situation 2015

- Dynamics of common infectious diseases are changing
  - Demographic change, population density, anti vaccine movement, AMR, etc.
- New diseases emerge frequently
  - Deforestation, population growth, health systems inequalities, travel, trade, climate change
- Effects are difficult to predict due to complexity of problems
- Public health and clinical response depend on global capacity for disease surveillance

Global capacity for emerging infectious disease detection

Emily H. Chan[a,b], Timothy F. Brewer[c,d], Lawrence C. Madoff[c,e], Marjorie P. Pollack[c], Amy L. Sonricker[a,b], Mikaela Keller[a,b,f], Clark C. Freifeld[a,b], Michael Blench[g], Abla Mawudeku[g], and John S. Brownstein[a,b,d,f,1]

[a]HealthMap, Children's Hospital Informatics Program, Harvard–Massachusetts Institute of Technology Division of Health Sciences and Technology, Boston, MA 02215; [b]Division of Emergency Medicine, Children's Hospital Boston, Boston, MA 02215; [c]ProMED-mail, International Society for Infectious Diseases, Brookline, MA 02446; [d]Departments of Medicine and Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada H3A 1A2; [e]Department of Medicine, University of Massachusetts Medical School, Worcester, MA 01655; [f]Department of Pediatrics, Harvard Medical School, Boston, MA 02215; and [g]Global Public Health Intelligence Network, Health Portfolio Operations Centre, Centre for Emergency Preparedness and Response, Public Health Agency of Canada, Ottawa, ON, Canada K0A 0K9

Edited by Burton H. Singer, University of Florida, Gainesville, FL, and approved October 29, 2010 (received for review May 10, 2010)

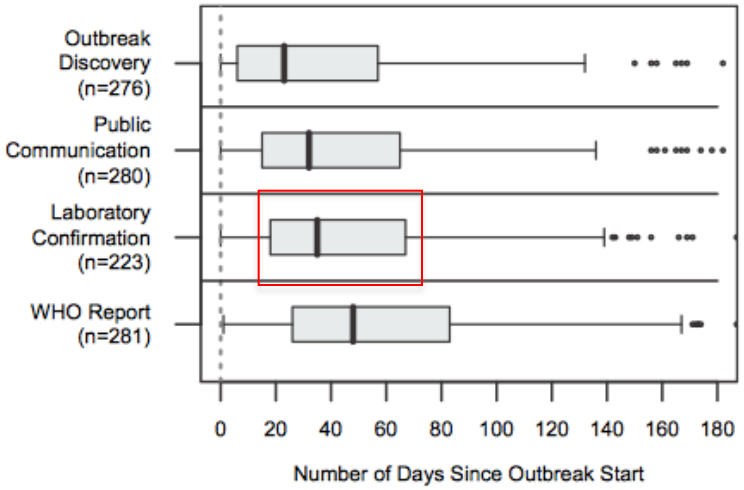


Fig. 2. Box plots of the median time between estimated outbreak start and various outbreak milestones for a subset of WHO-confirmed outbreaks, 1996–2009. Public communication refers to the earliest date of the public being informed about the existence of cases. WHO report refers to the date of WHO's *Disease Outbreak News* report about the outbreak. Some extreme outliers are not shown. *n*, sample size.

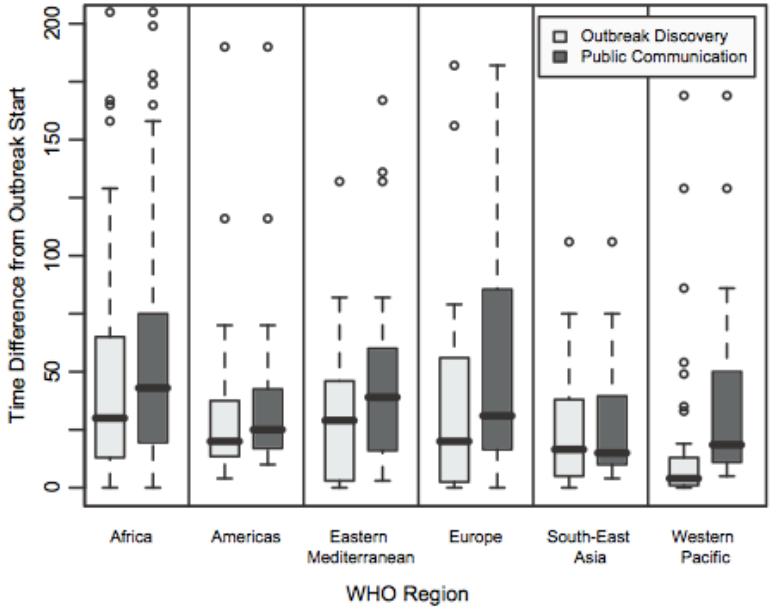
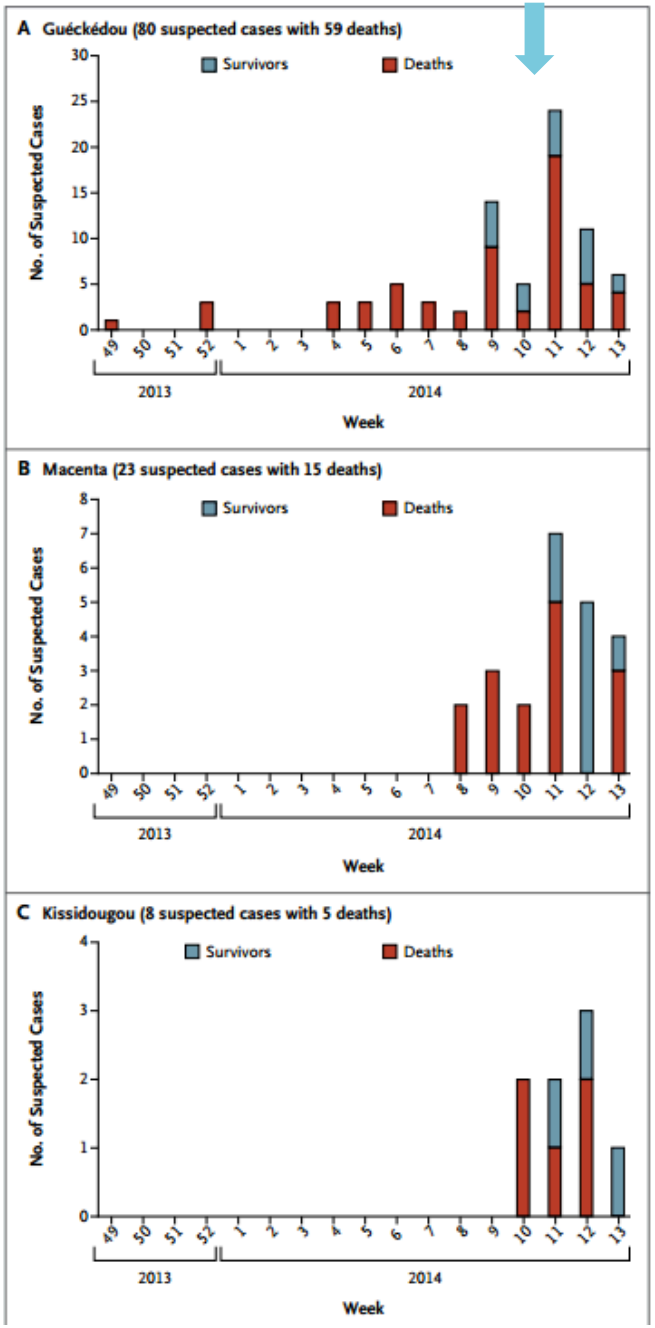


Fig. 4. Box plots of the median time difference from estimated outbreak start to outbreak discovery and public communication about the outbreak for selected WHO-verified outbreaks, 1996–2009, across various WHO regions. Extreme outliers are not shown.

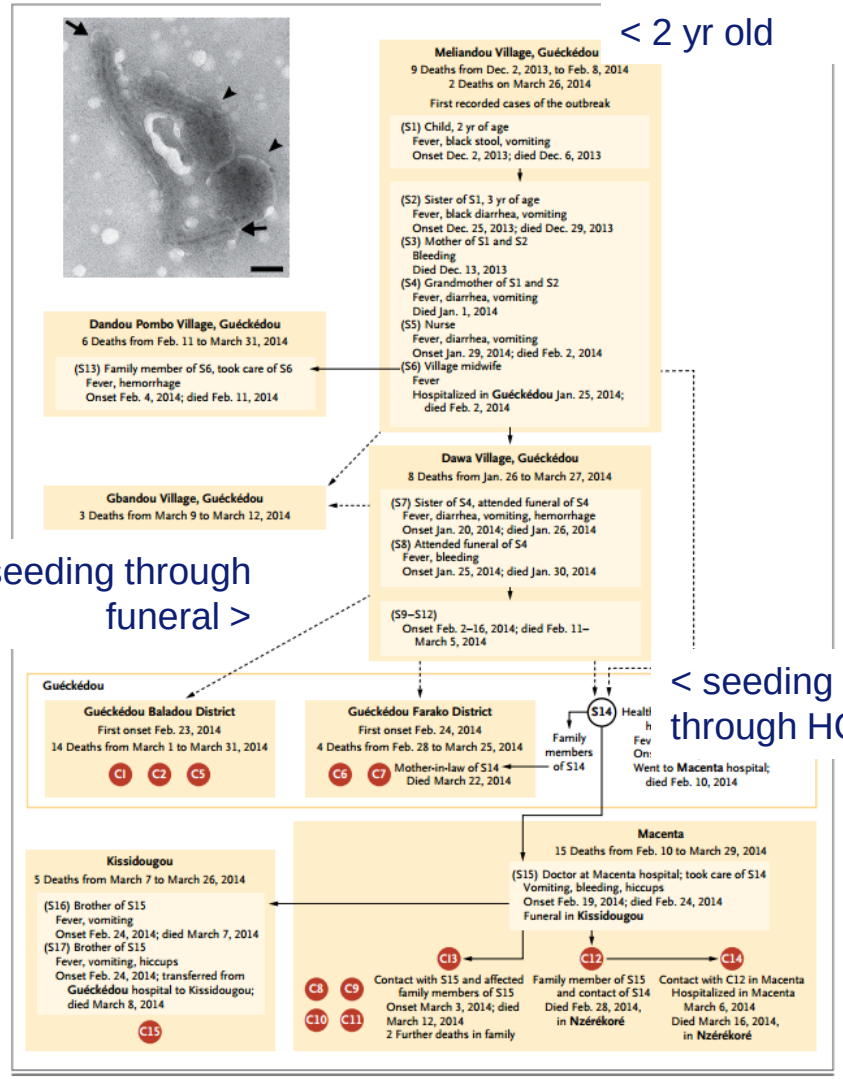
Start outbreak EBOV

- March 10, 2014 notification unknown disease characterized by fever, severe diarrhea, vomiting and high fatality rate in Guéckédou and Macenta in Guinea.
- March 22, EVD reported by Guinea to WHO.
- March 27, EVD suspected cases in Liberia and Sierra Leone related to outbreak in Guinea.
- April 3d: ZEBOV Dx

Diagnosis



< 2 yr old

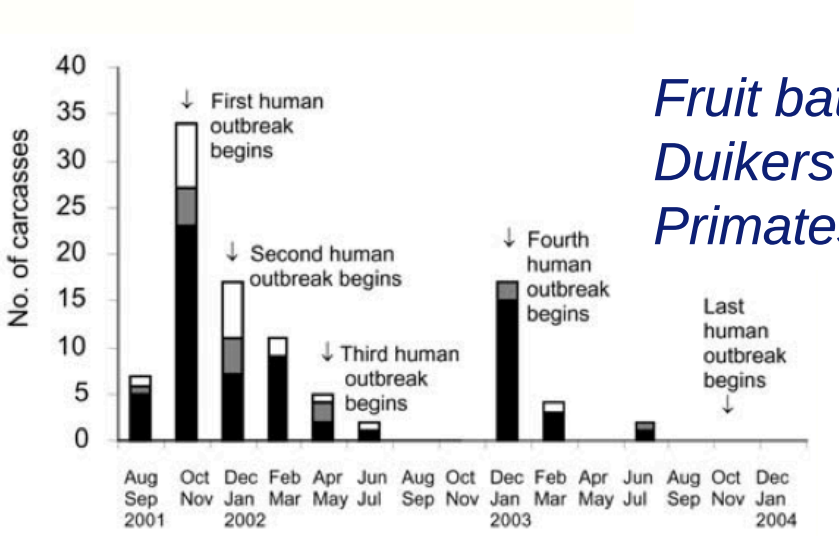


seeding through funeral >

< seeding through HCW

Animal surveillance, Gabon 2001-3

Figure 2. Field watertight clothes equipped with air filtration equipment, used for high-risk wild animal necropsy. Odzala National Park, Republic of Congo, June 2003. Photo: P. Rouquet.



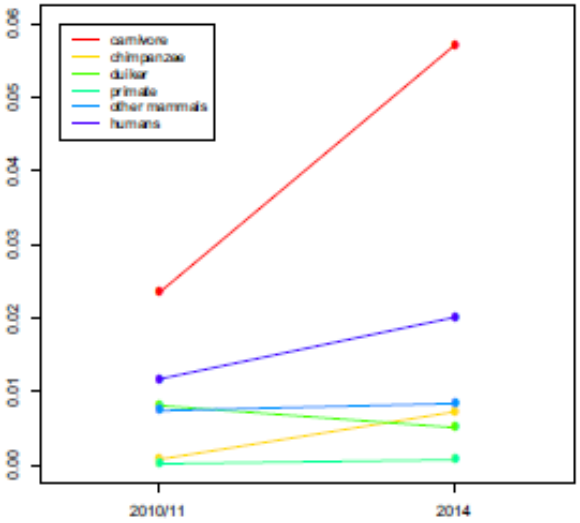
Fruit bats
Duikers
Primates

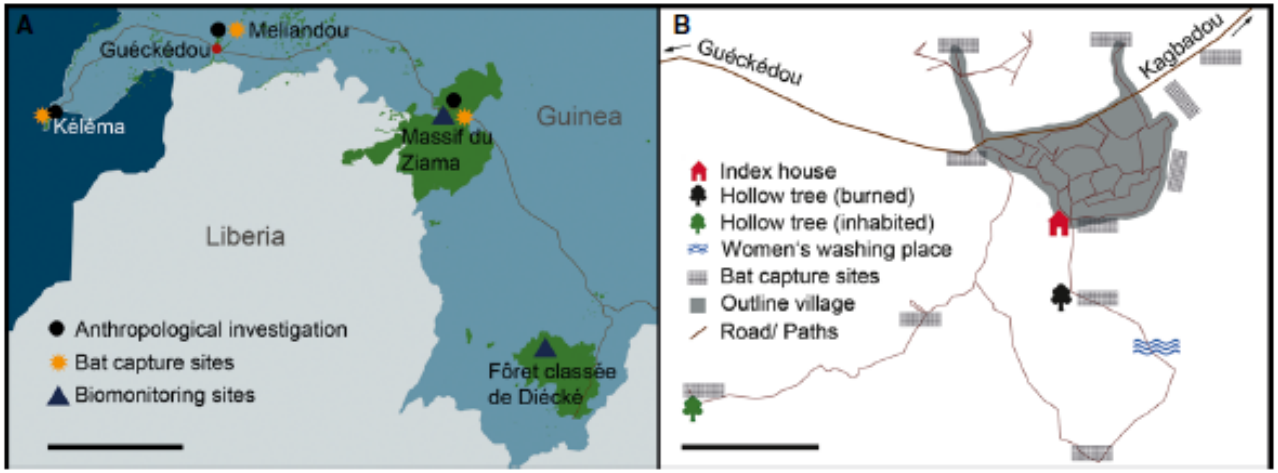
Figure 4. Temporal distribution of carcasses found in the forest straddling the border between Gabon and the Republic of Congo (2001–2003). Two peaks of mortality were observed: the first occurred in the Ekata region (Gabon) from November to December 2001 and the second from December 2002 to February 2003 in the Lossi gorilla sanctuary (Republic of Congo).

Outbreaks in animals detected prior to (4/5) human disease outbreaks

Convincing evidence for bushmeat related introductions

Rouquet et al., 2005

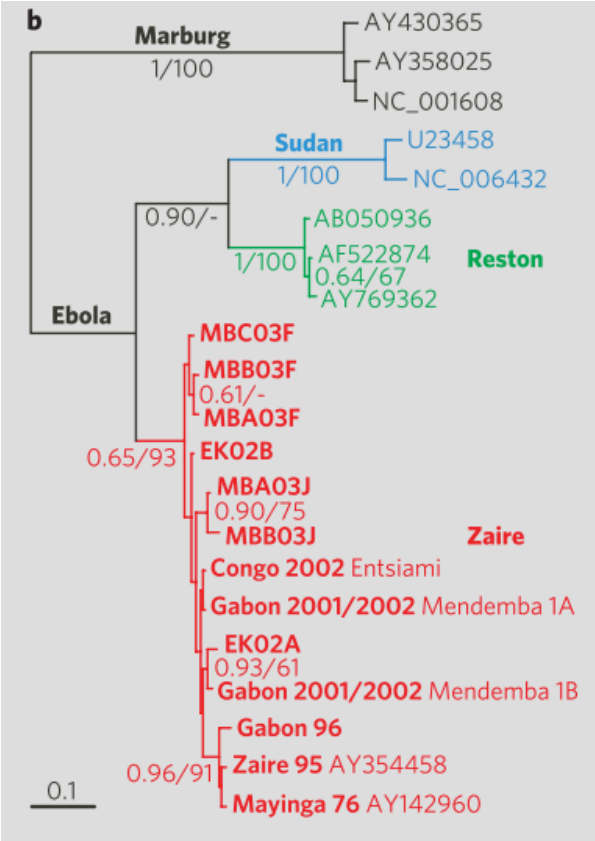


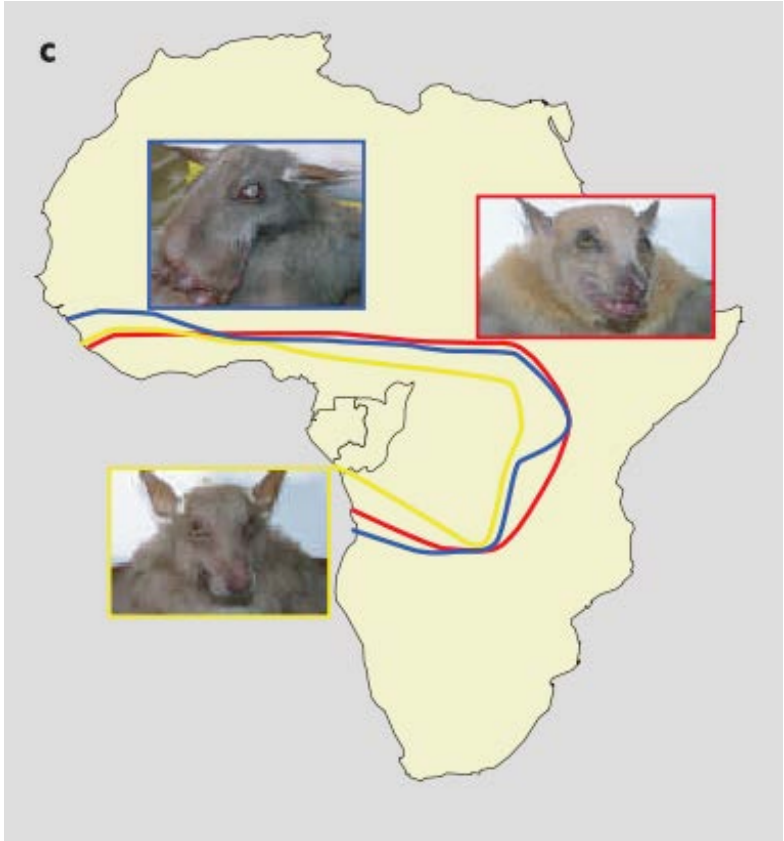


Single zoonotic event in Meliandou, bat-borne, followed by human2human transmisison

Saez et al., EMBO Mol Med, 2014

Fruit Bats as reservoir for EBOV





Overlapping ecological niche
No symptoms
Infection cyclical
Potential source of introduction into West Africa

Leroy et al., 2005

Potential under-reporting of Ebola

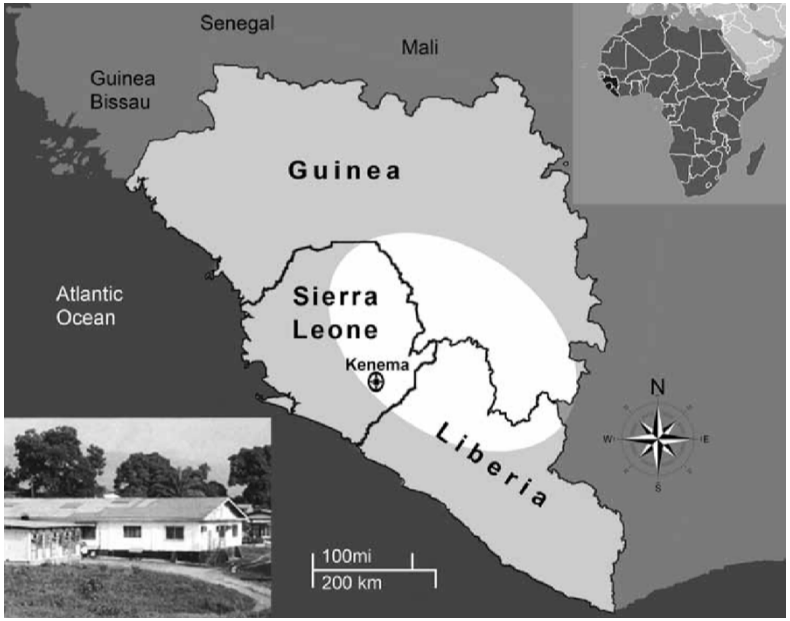


Table 2. Patients' antibody reactions to arthropod-borne and hemorrhagic fever virus antigens, Lassa Diagnostic Laboratory, Kenema, Sierra Leone, October 2006–October 2008*

Virus	No. positive /total (%)	No. IgM only positive/total (%)
Dengue	11/253 (4.3)	6/250 (2.4)
West Nile	7/253 (2.8)	3/250 (1.2)
Yellow fever	5/201 (2.5)	5/201 (2.5)
Rift Valley fever	5/253 (2.0)	5/253 (2.0)
Chikungunya	10/253 (4.0)	5/253 (2.0)
Ebola	19/220 (8.6)	18/219 (8.2)
Marburg	8/220 (3.6)	7/219 (3.2)
Crimean-Congo hemorrhagic fever	0/220	Not tested
Total	65/253 (25.7)	49/253 (19.4)

Schoepp et al., 2014

EBOV disease detection issues

- Delayed diagnosis
- Lack of understanding of possible zoonotic threats from wildlife
- Lack of surveillance recognizing unusual clinical syndromes
- Lack of routine and reference laboratory capacity for evaluation of (unusual) disease
- Change in outbreak profile from rural to urban > rapid spread

Potential added value of genome sequencing

1. Zaire-like virus
2. Single introduction
3. Two introductions into Guinea

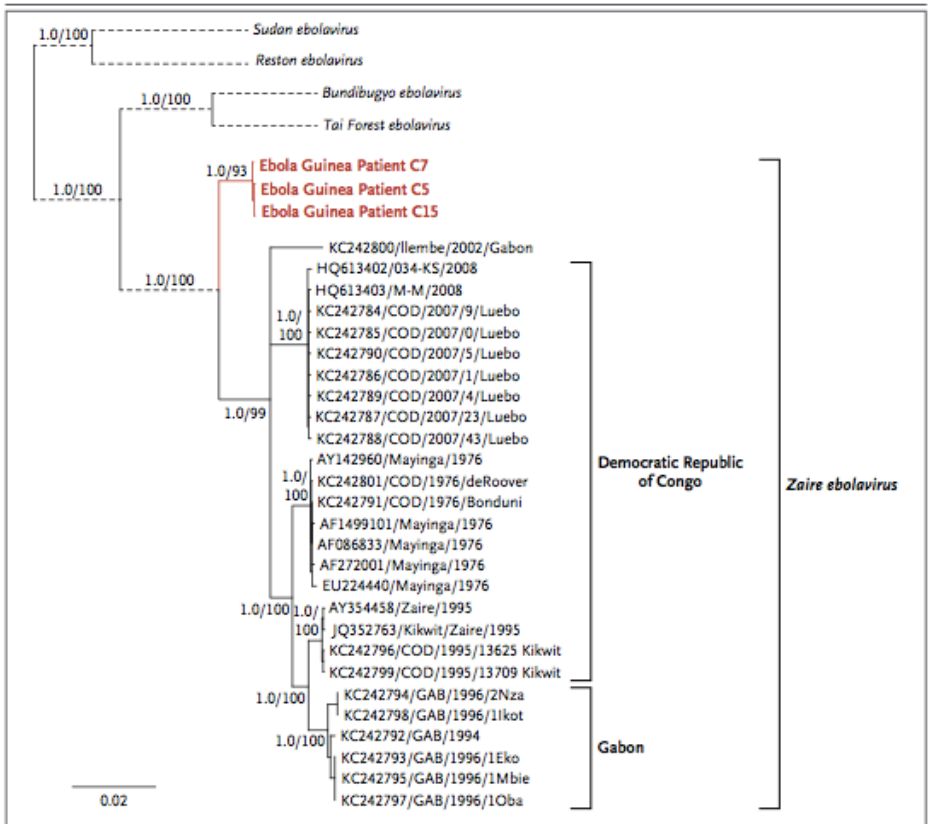
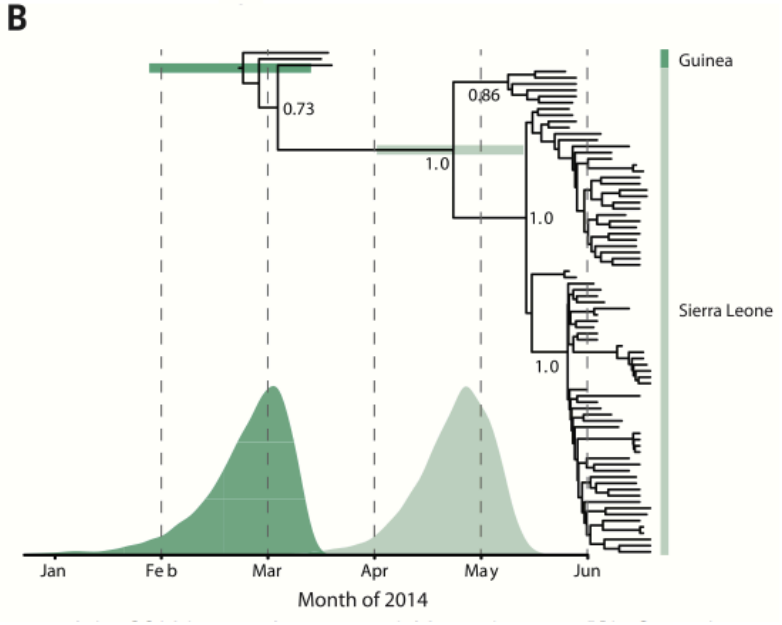
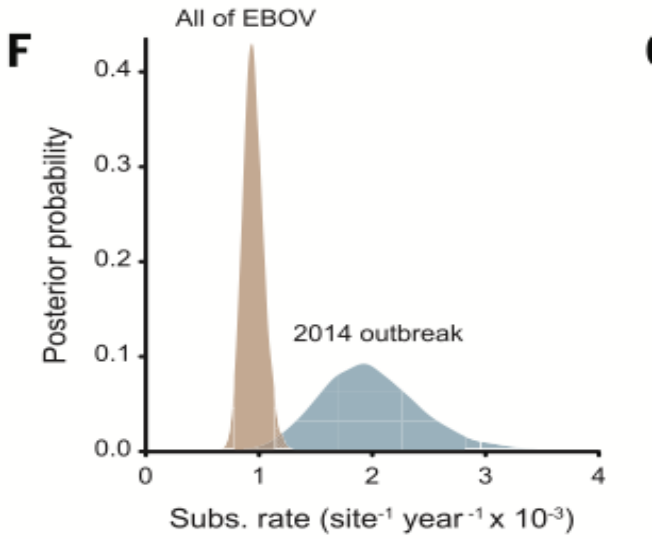
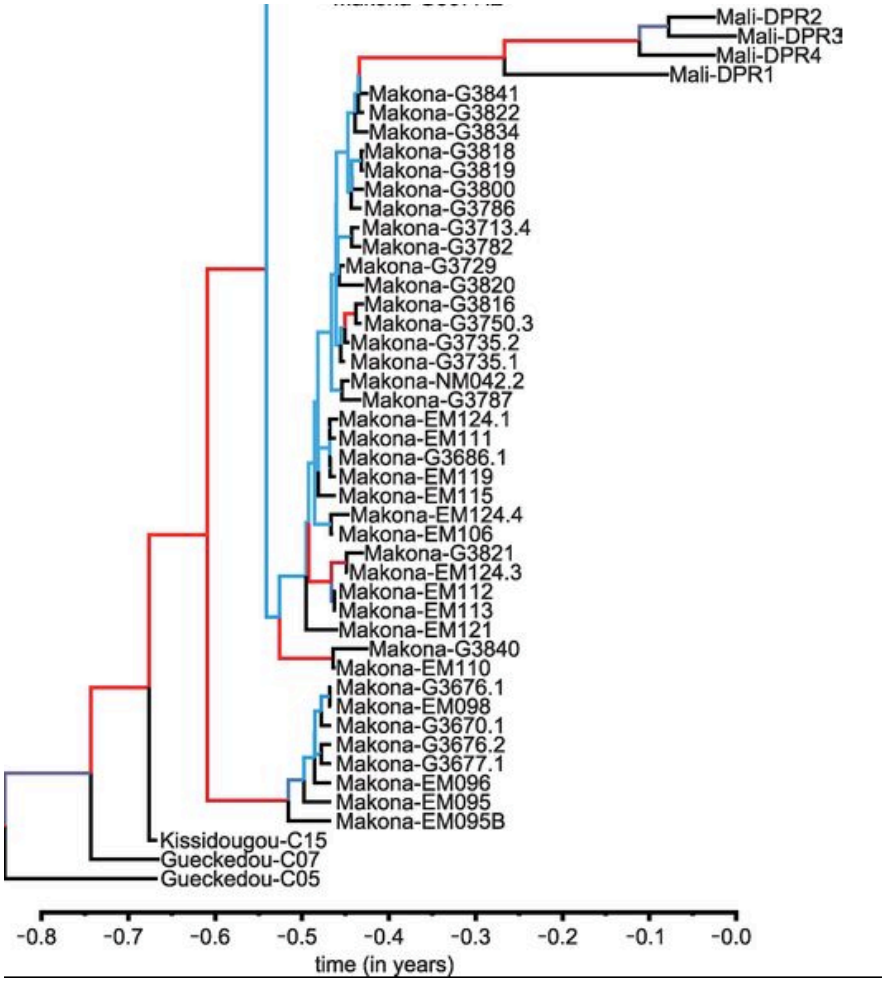


Figure 3. Phylogenetic Analysis of the Ebolavirus Genus, Including the EBOV Strains from Guinea.
The phylogenetic tree was inferred with the use of the Bayesian Markov Chain Monte Carlo method. A second tree that was inferred for the same set of sequences with a maximum-likelihood method confirmed the Bayesian tree (data not shown). Bayesian posterior probabilities and bootstrap percentages (1000 replicates of the maximum-likelihood tree) are shown on the branches. For clarity of presentation, the branches for the non-EBOV species were shortened and condensed (dashed branches). The GenBank accession number, strain designation, country of origin, and year of isolation are indicated on the EBOV branches. The EBOV Guinea strain is available from the European Virus Archive (www.european-virus-archive.com).

Baize et al., October 2014







- Diversification of virus
- Geographic lineages > tracking possible
- Effect of mutations on molecular Dx, therapeutics, vaccines unknown

Gire et al., 2014; Hoenen et al., 2015

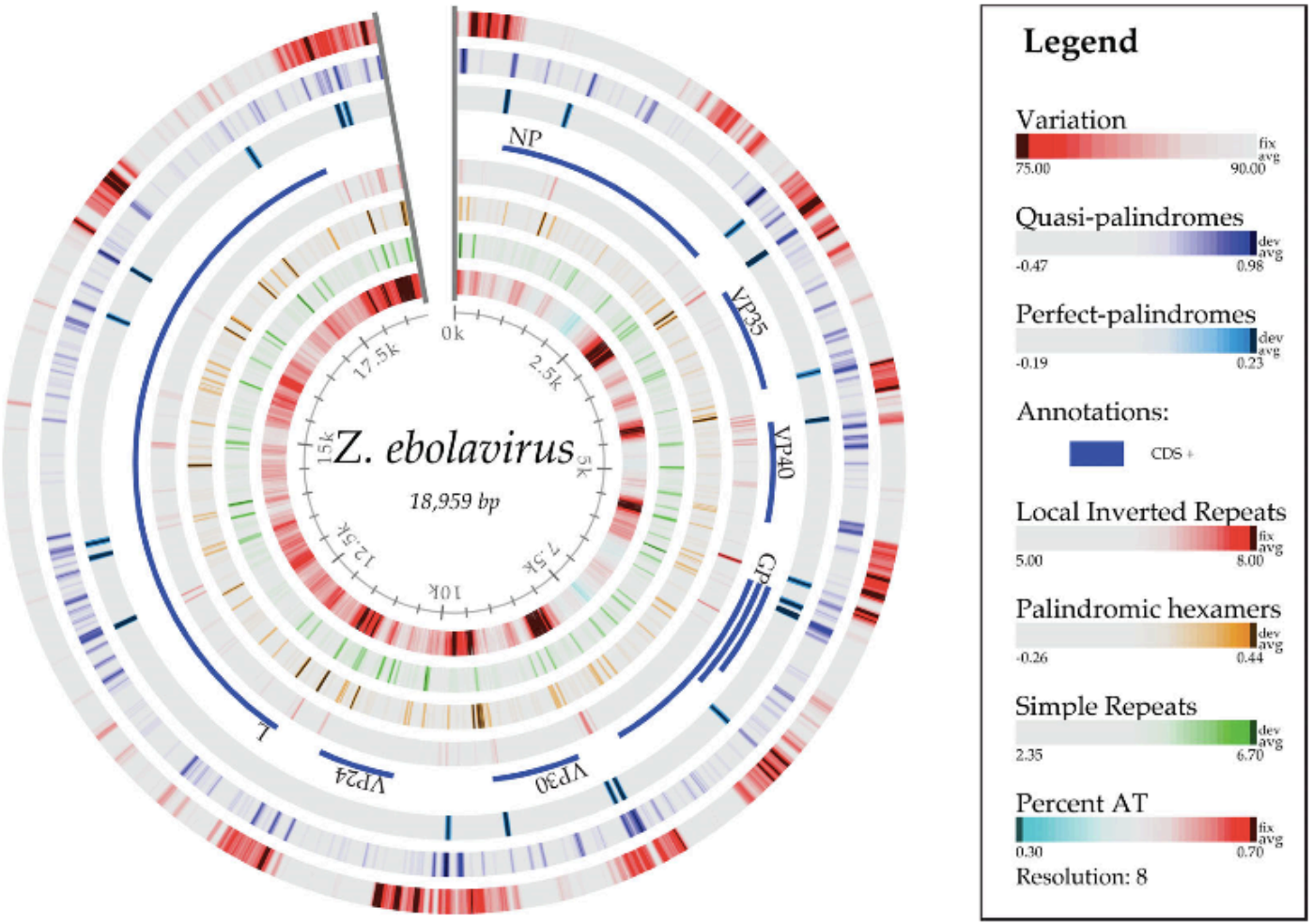
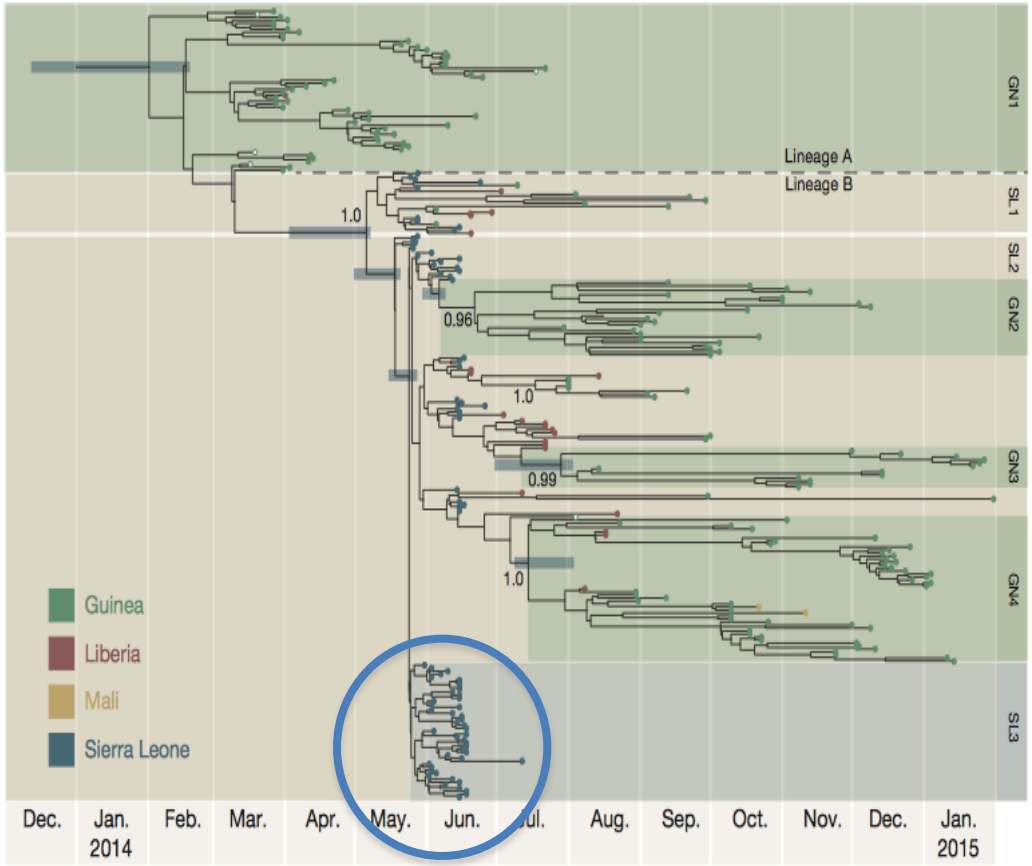


Figure 5. Atlas of the genome of ebolavirus KJ660347, showing, from the outer ring inwards, variations within 84 other ebolavirus genomes, structural cruciforms and palindromes (van Noort *et al.* 2003), the coding sequences, local inverted repeats, palindromic hexamers, simple repeats and AT content. The conservation percentage (%) is defined as the number of genomes with the same letter on a multiple sequence alignment normalized to range from 0 to 100% for each site along the chromosome of Ebola KJ660347.

Jun et al., 2015

Incomplete information, many parties involved, no data sharing system



6 AUGUST

Caroll et al., August 2015

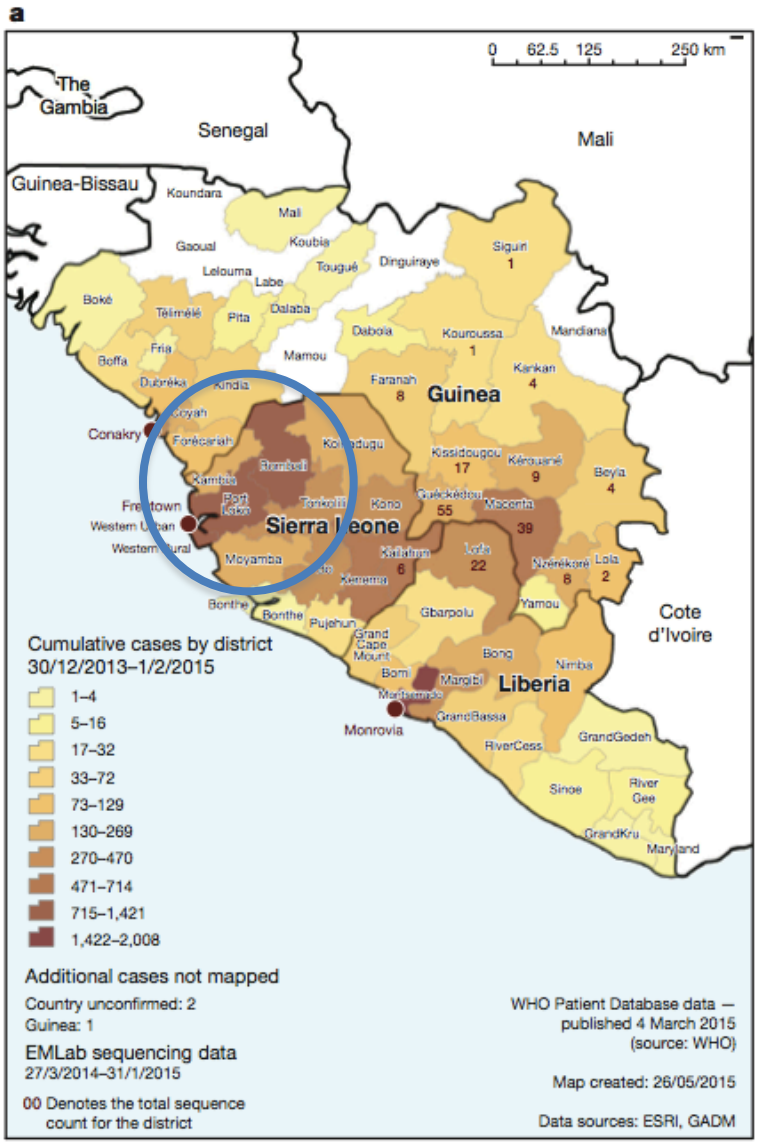
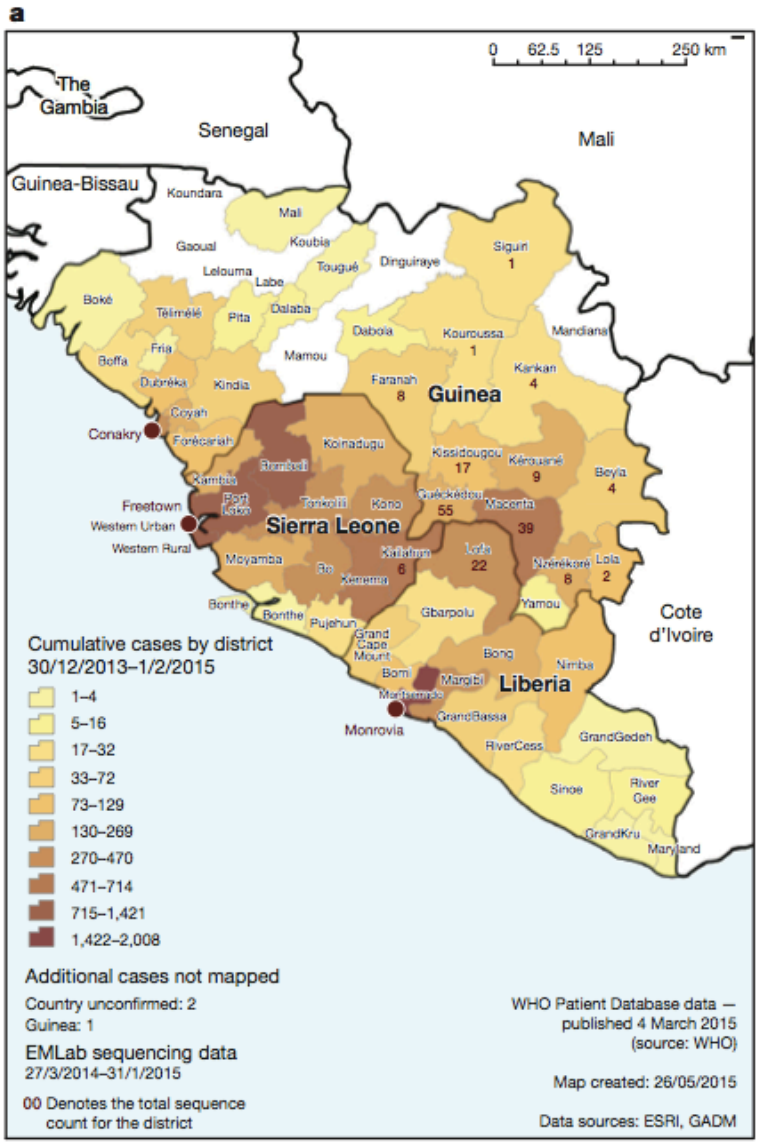


Figure 1 | Geographical location, sequence read depth, and read depth vs C_t value of patient samples. a, Geographical location of patient samples. The



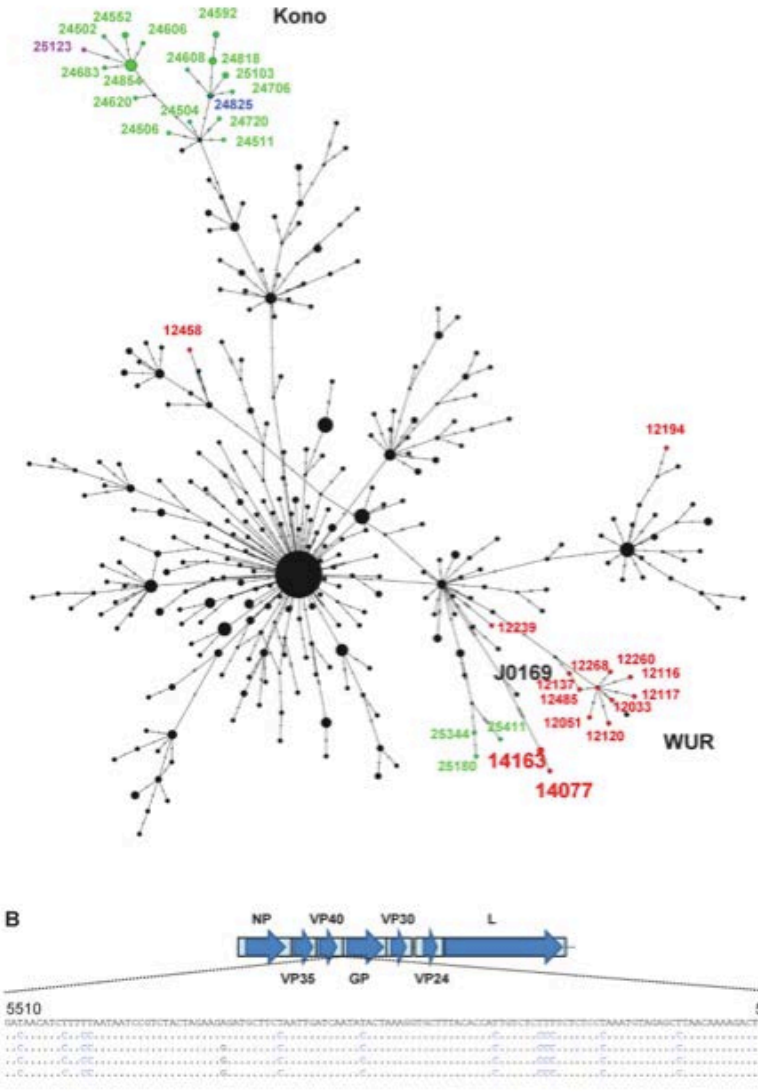
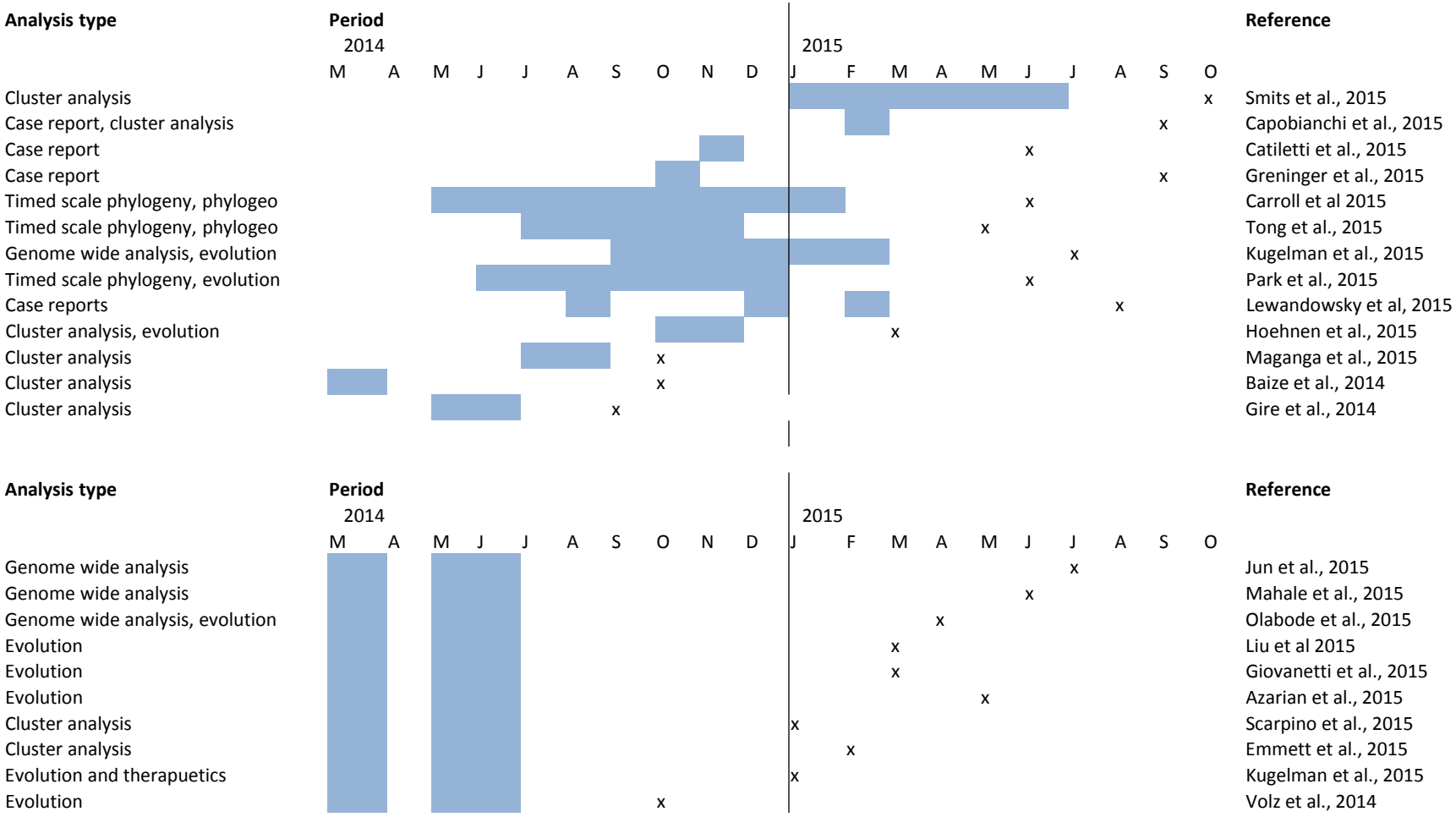
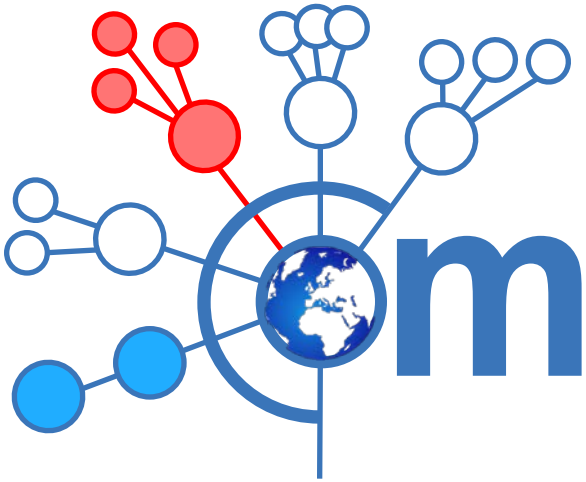


Figure 1 | Geographical location, sequence read depth, and read depth vs C_t value of patient samples. **a**, Geographical location of patient samples. The GP: glycoprotein; L: RNA-dependent RNA polymerase L; NP: nucleoprotein; VP: virus protein; WUR: Western Area Urban district.

Caroll et al., 2015; Smits et al., 2015

Timeliness of sequence-based disease detection and analysis





COmpare

COllaborative Management Platform for detection and Analyses of (Re-) emerging and foodborne outbreaks in Europe

A global platform for the sequence-based rapid identification of pathogens

Prof. Frank M. Aarestrup, coordinator (Technical University of Denmark)
Prof. Marion Koopmans, deputy coordinator (Erasmus Medical Center, the Netherlands)

This project has received funding from the *European Union's Horizon 2020 research and innovation program* under grant agreement 643476

Background

- Laboratory diagnostics increasingly rely on (pathogen) genomic information
- RNA / DNA are common across pathogens, therefore, methods to analyse pathogen genomes potentially are universal
- Next generation sequencing capacity is developing fast, and costs are becoming competitive

- ➢ Capturing NGS developments may provide a universal language that can be harnessed for early detection of outbreaks across disciplines and domains
- ➢ If the technology keeps developing, less equipped labs may leapfrog

COMPARE basics:

original call text

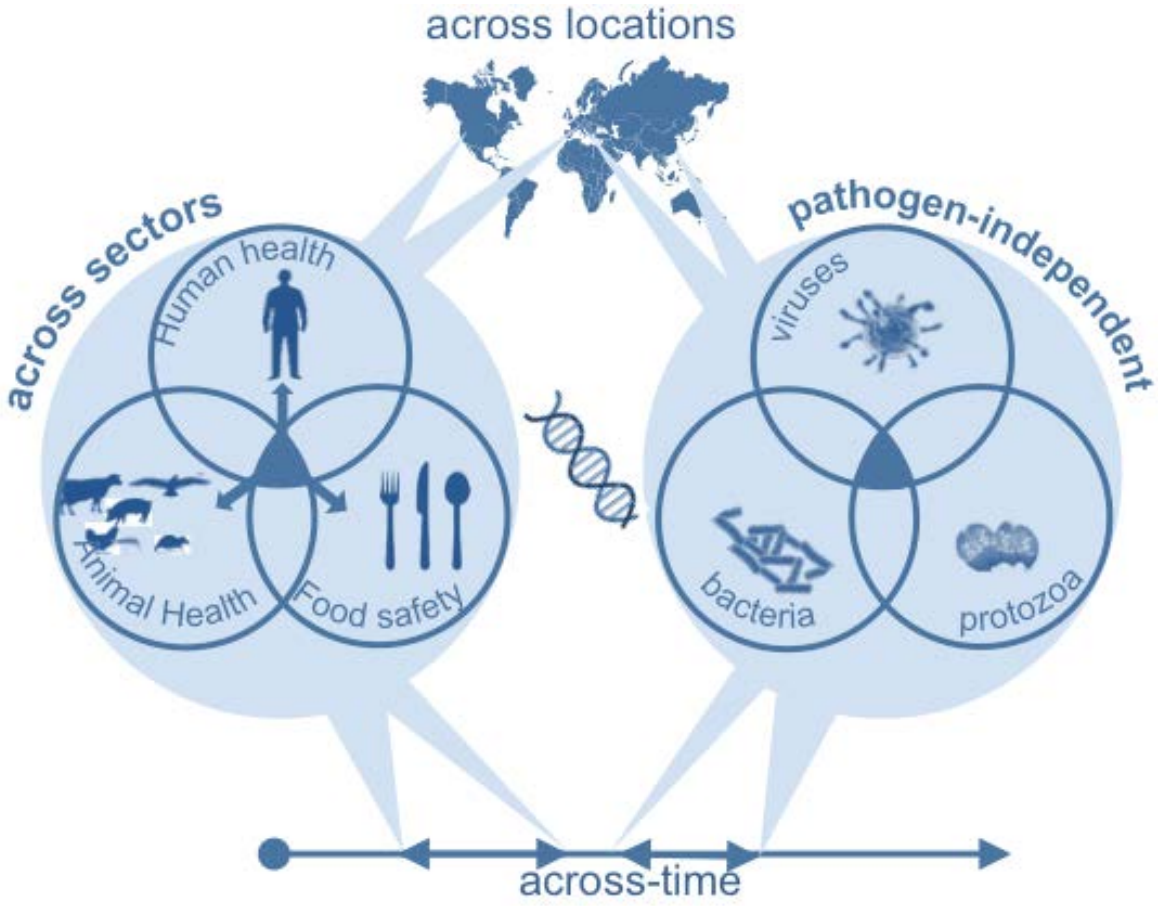
PHC 7 - 2014) Improving the control of infectious epidemics and foodborne outbreaks through rapid identification of pathogens (see also SC2)

Specific challenge: Human and animal health worldwide is increasingly threatened by potential epidemics caused by existing, new and emerging infectious diseases (including from antimicrobial resistant pathogens), placing a burden on health and veterinary systems, reducing consumer confidence in food, and negatively affecting trade, food chain sustainability and food security.

The increasing incidence and more rapid spread of such diseases are facilitated by modern demographic, environmental, technological and societal conditions. Many of these infections are zoonoses, necessitating an integrated, cross-border, “one health” approach to research and public health measures in the human and veterinary field, including the food chain.

Our vision:

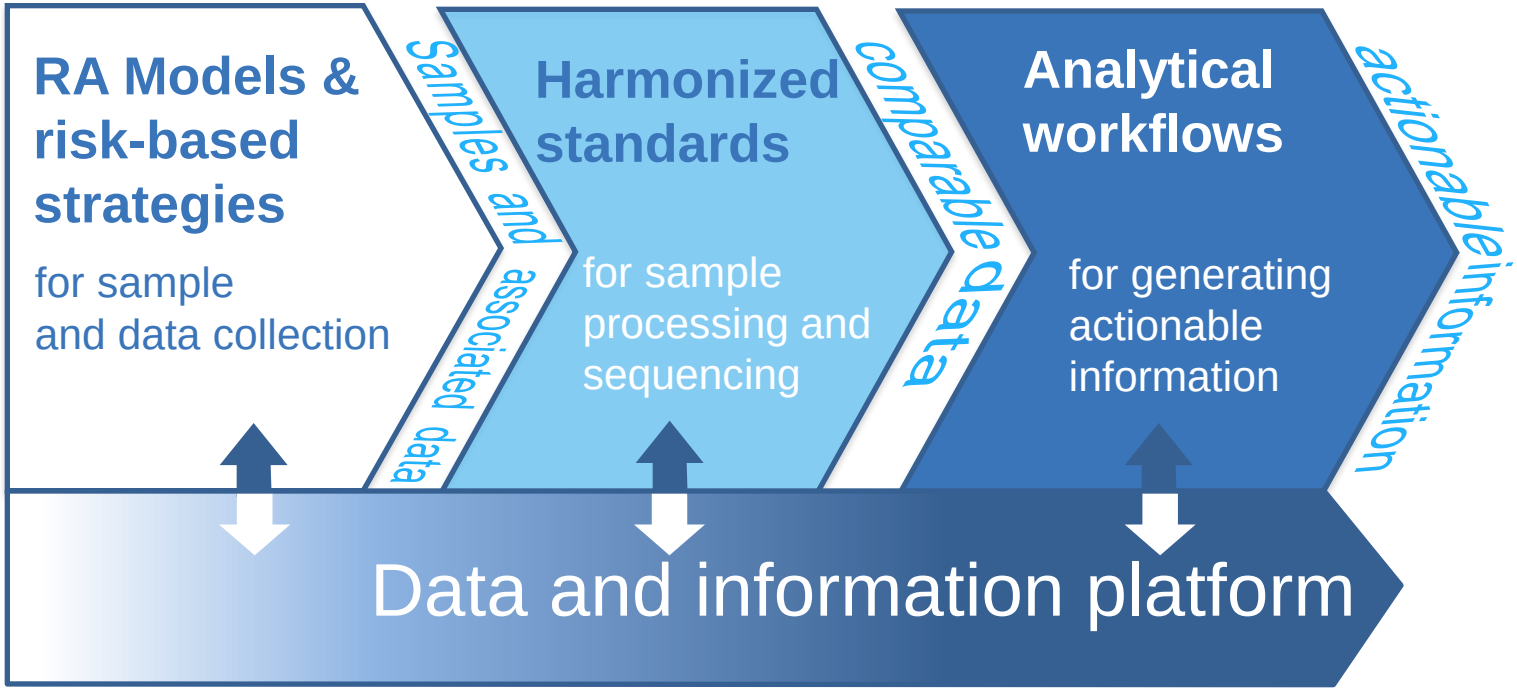
to build one system that serves all



COMPARE principles

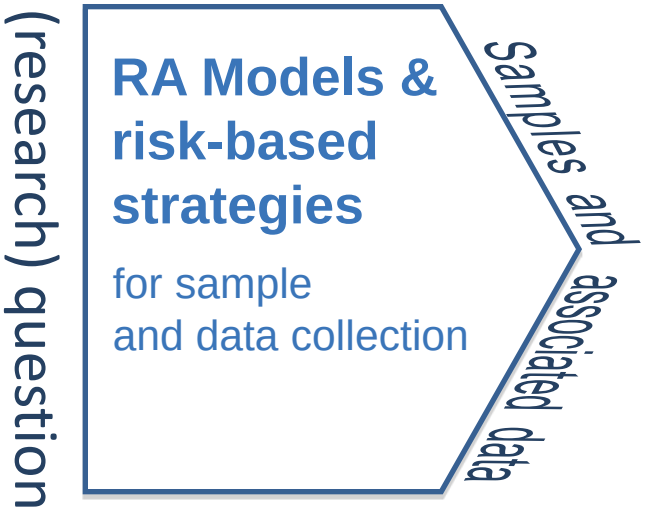
- COMPARE is a sector, domain and pathogen-**independent** system;
- analyzing **sequence-based pathogen data** in combination with **associated (clinical, epidemiological and other) data,**
- Building on **established infrastructures**
- COMPARE is a **user driven system**, designed with the information needs of its intended diverse group of future users and other stakeholders in mind;
- COMPARE will make **optimal use of existing and future complementary systems, networks and databases** ensuring compatibility where needed;
- COMPARE is a **flexible, scalable and open-source based** information-sharing platform.
- 1 December 2014 – 31 November 2019

Analytical framework and globally linked data and information sharing platform.



WP1

From question to samples and associated data:
Goal: to develop Risk Assessment (RA) models and risk-based sampling and data collection strategies for Next Generation Sequencing (NGS)-based analyses of food-borne and (re-) emerging infections



Emma Snary

WP-leader

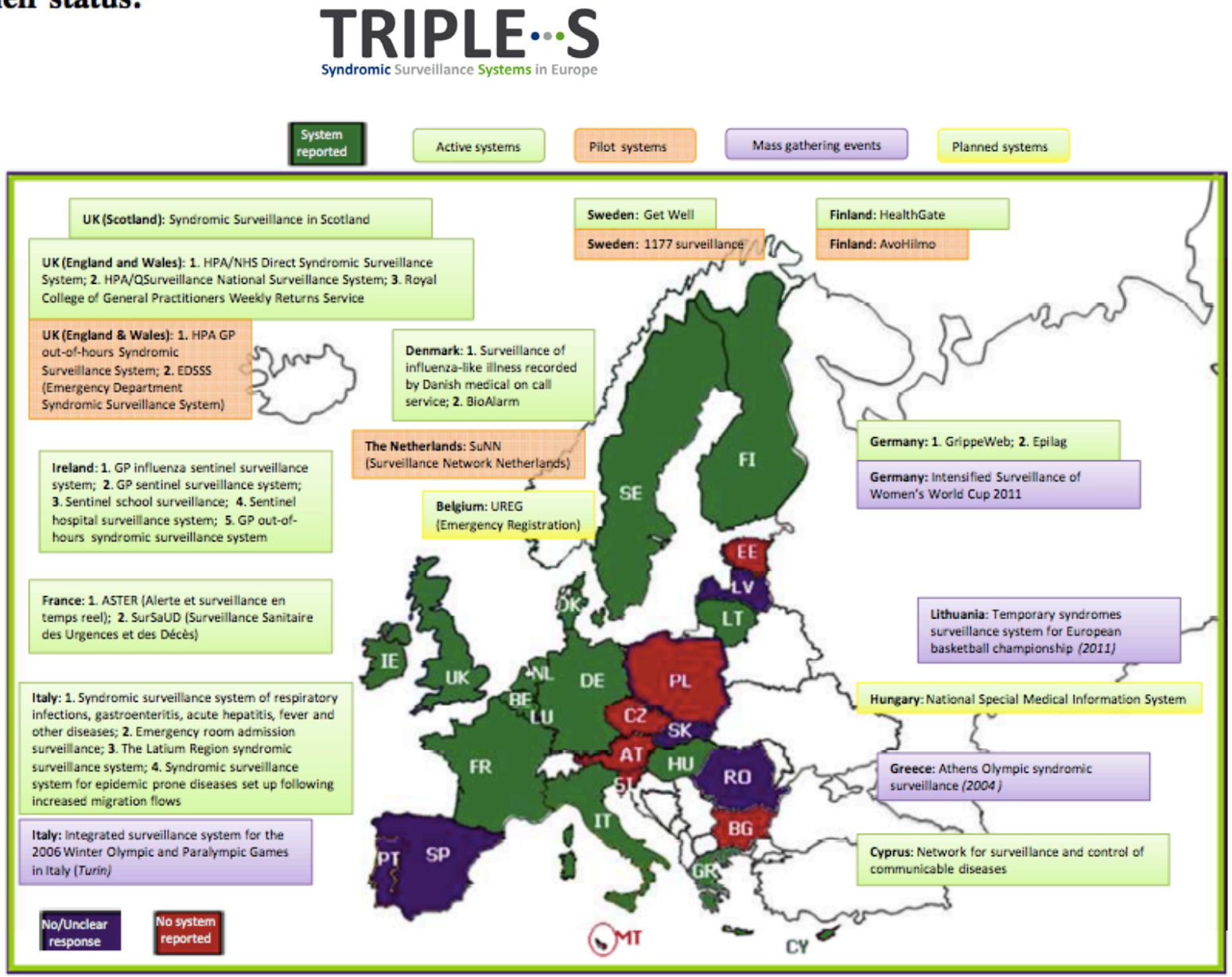
Christian Gortazar

WP co-leader

Scenario's:

- Novel airborne disease, respiratory syndrome
- Foodborne disease, enteric syndrome
- Vectorborne disease with fever/rash syndrome

Figure 1. Map of Europe showing names and locations of syndromic surveillance systems and their status.



Task 1.1: Generic framework

- Risk of infection in area **B** due to presence of hazard in area **A**

$$P\left(\begin{matrix}Presence\ of\\ hazard\ in\\ Area\ A\end{matrix}\right) P\left(\begin{matrix}Connection\\ Area\ A \rightarrow\\ Area\ B\end{matrix}\right) P\left(\begin{matrix}Suitable\\ environment\\ in\ Area\ B\end{matrix}\right) P\left(\begin{matrix}Transmission\\ in\ Area\ B\end{matrix}\right)$$

Entry | Exposure | Consequence

COMPARE is funded by the European Commission under Grant Agreement No 643476

Task 1.1: Generic framework

- Risk of infection in area **B** due to presence of hazard in area **A**

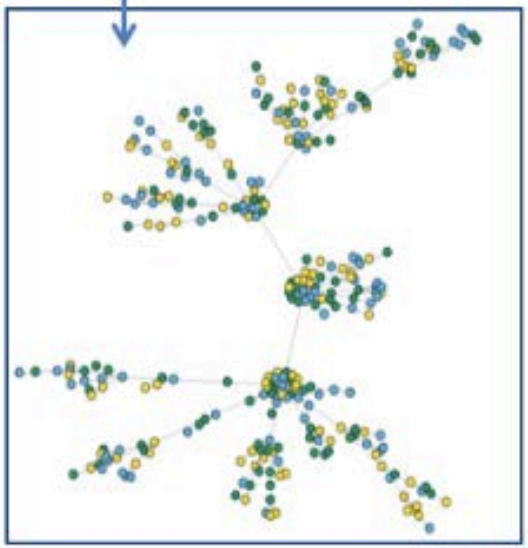
$$P\left(\begin{matrix}Presence\ of\\ hazard\ in\\ Area\ A\end{matrix}\right) P\left(\begin{matrix}Connection\\ Area\ A \rightarrow\\ Area\ B\end{matrix}\right) P\left(\begin{matrix}Suitable\\ environment\\ in\ Area\ B\end{matrix}\right) P\left(\begin{matrix}Transmission\\ in\ Area\ B\end{matrix}\right)$$

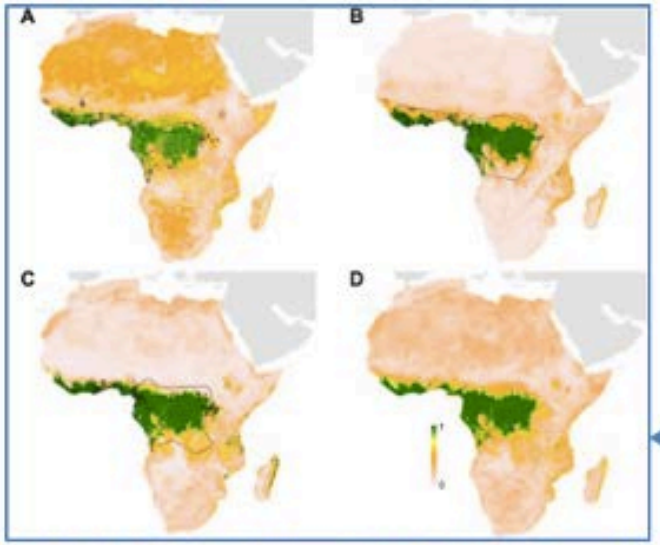
Summarising & visualising movement data

Entry

Exposure

Consequence





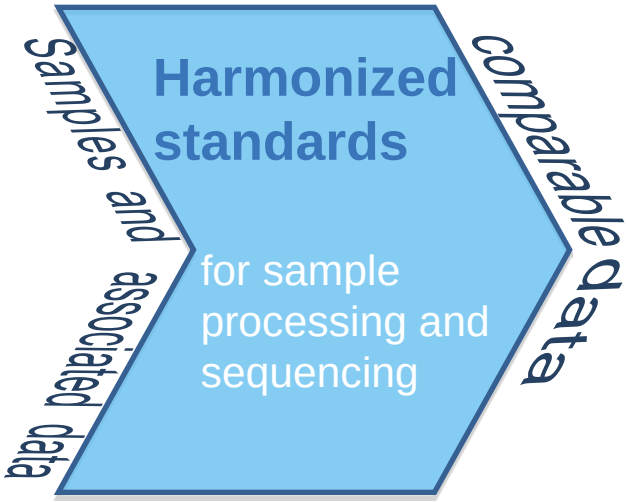
Ecological niche modelling

COMPARE is funded by the European Commission under Grant Agreement No 643476

WP2

From samples and associated data to comparable data

Goal: to develop harmonised analytical workflows for generation of high quality NGS data in combination with relevant metadata for pathogen detection and typing across sample types, pathogens and domains.



Martin Beer

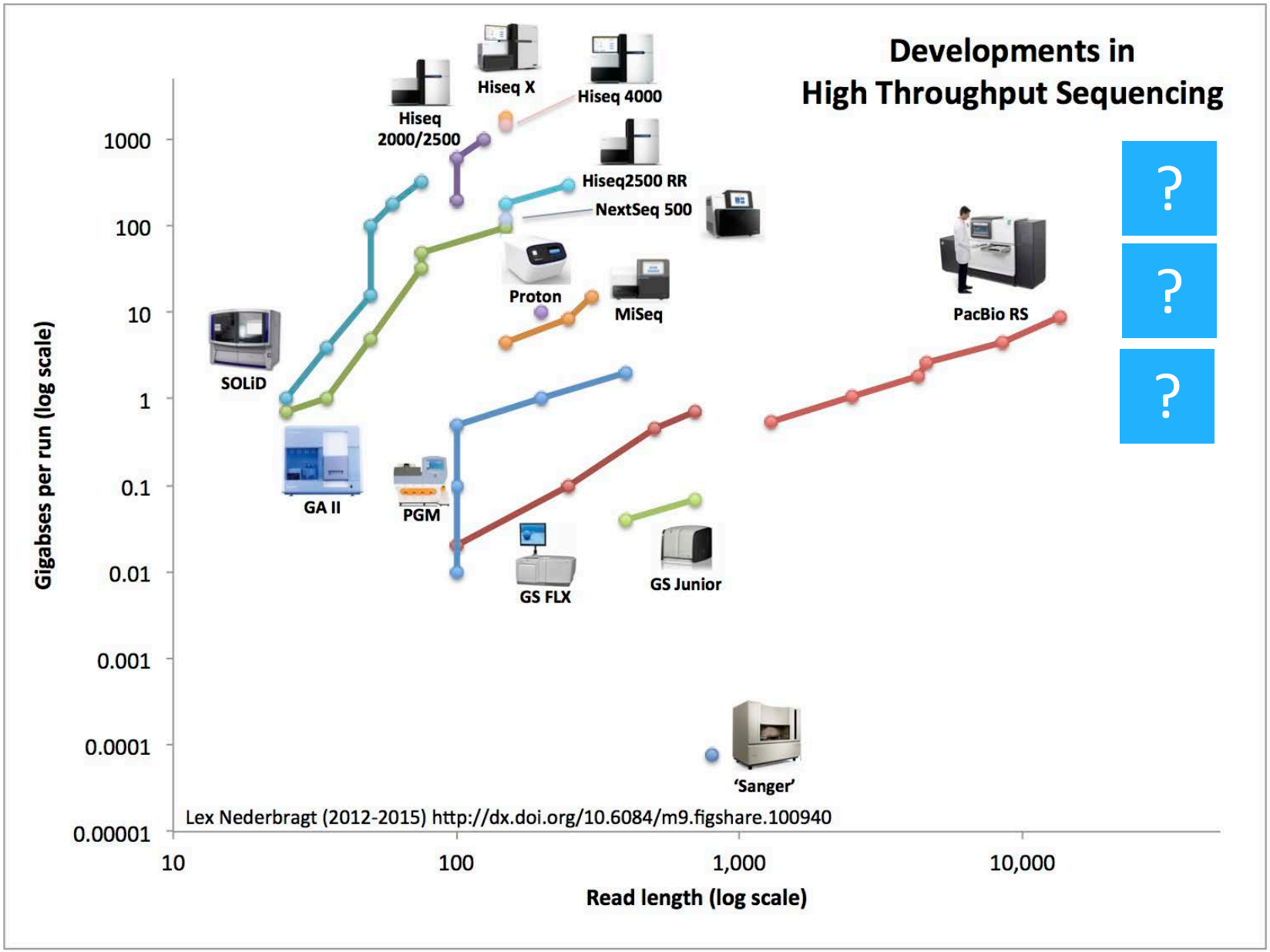
WP leader

Simone Caccio

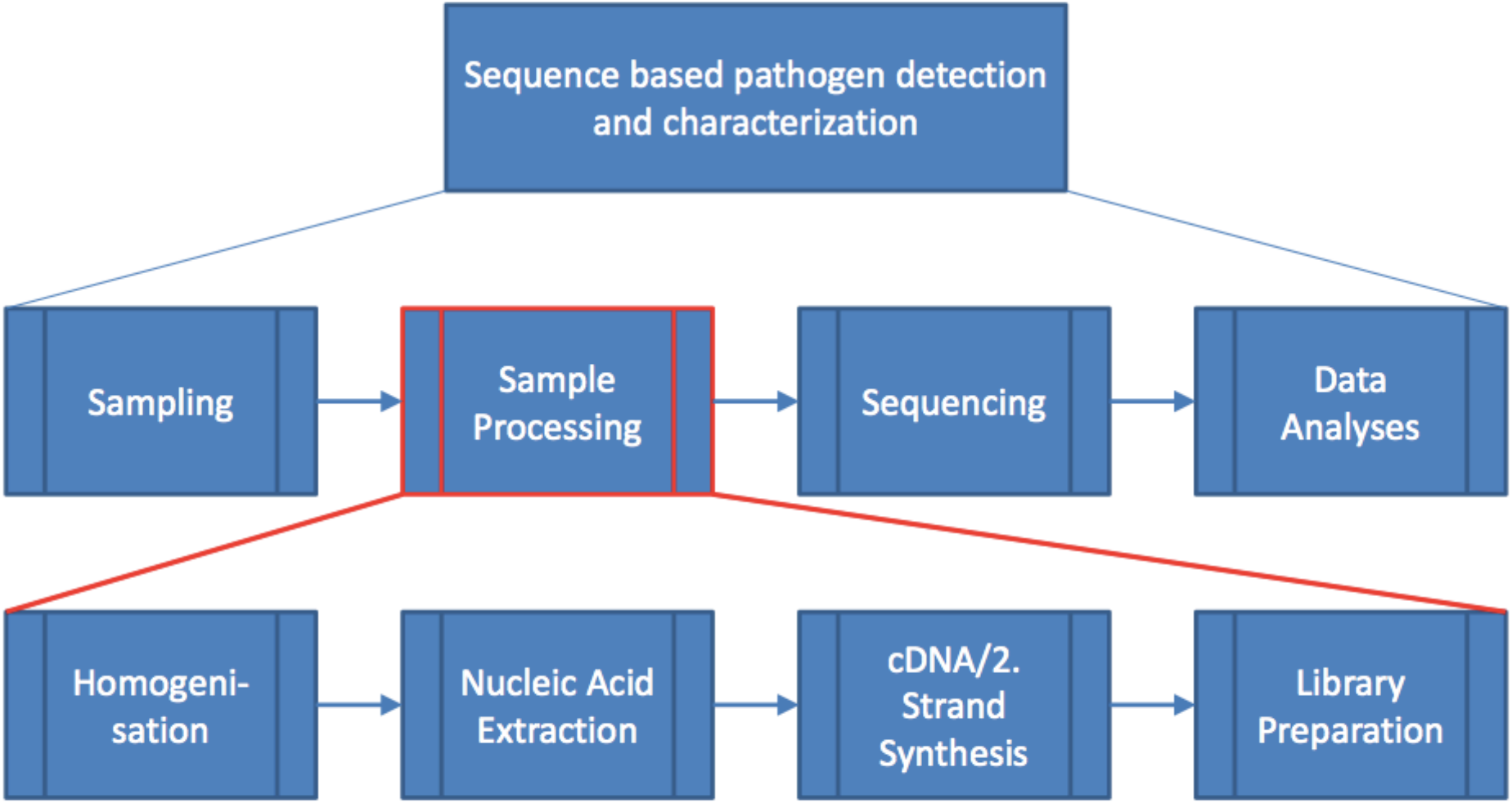
WP co-leader

Paul Kellam

WP co-leader


Developments in
High Throughput Sequencing
Hiseq X
Hiseq 4000
Hiseq
2000/2500
Hiseq2500 RR
NextSeq 500
Proton
MiSeq
PacBio RS
SOLiD
GA II
PGM
GS FLX
GS Junior
'Sanger'
?
?
?
Gigabses per run (log scale)
1000
100
10
1
0.1
0.01
0.001
0.0001
0.00001
10
100
1,000
10,000
Read length (log scale)
Lex Nederbragt (2012-2015) http://dx.doi.org/10.6084/m9.figshare.100940

Task 2: Standardised processes for sample processing (Dirk Höper/Simone Caccio)

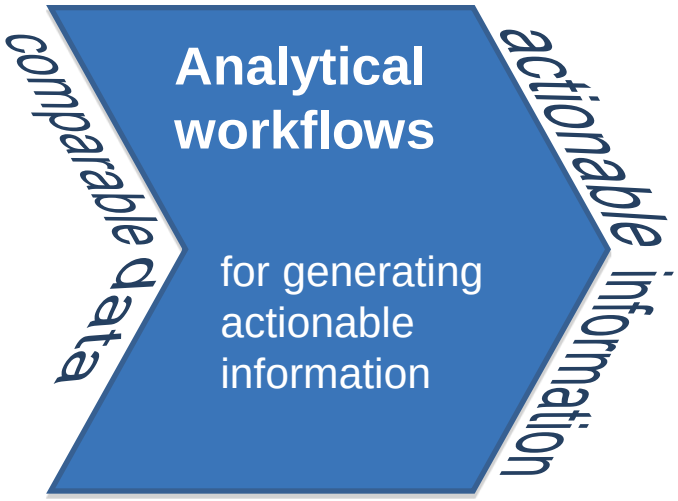


COMPARE is to be funded by the European Commission under Grant agreement No 643476

From comparable data to actionable information

Analytical workflows

"Actionable Information" is defined as information that enables users generating/receiving this information to take well-informed decisions and actions in pursuit of:



From comparable data to actionable information

Different users need different Analytical workflows

comparable data

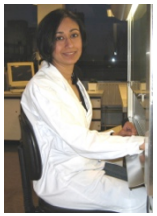
Analytical workflows

for generating actionable information

actionable information

Frontline diagnostics in human and veterinary clinical microbiology

Surbhi Malhortra — WP leader



Menno de Jong — WP co-leader

Anne Pohlmann — WP co-leader

Detection and analysis of foodborne outbreaks

Eva Møller-Nielsen — WP leader

Tine Hald — WP co-leader

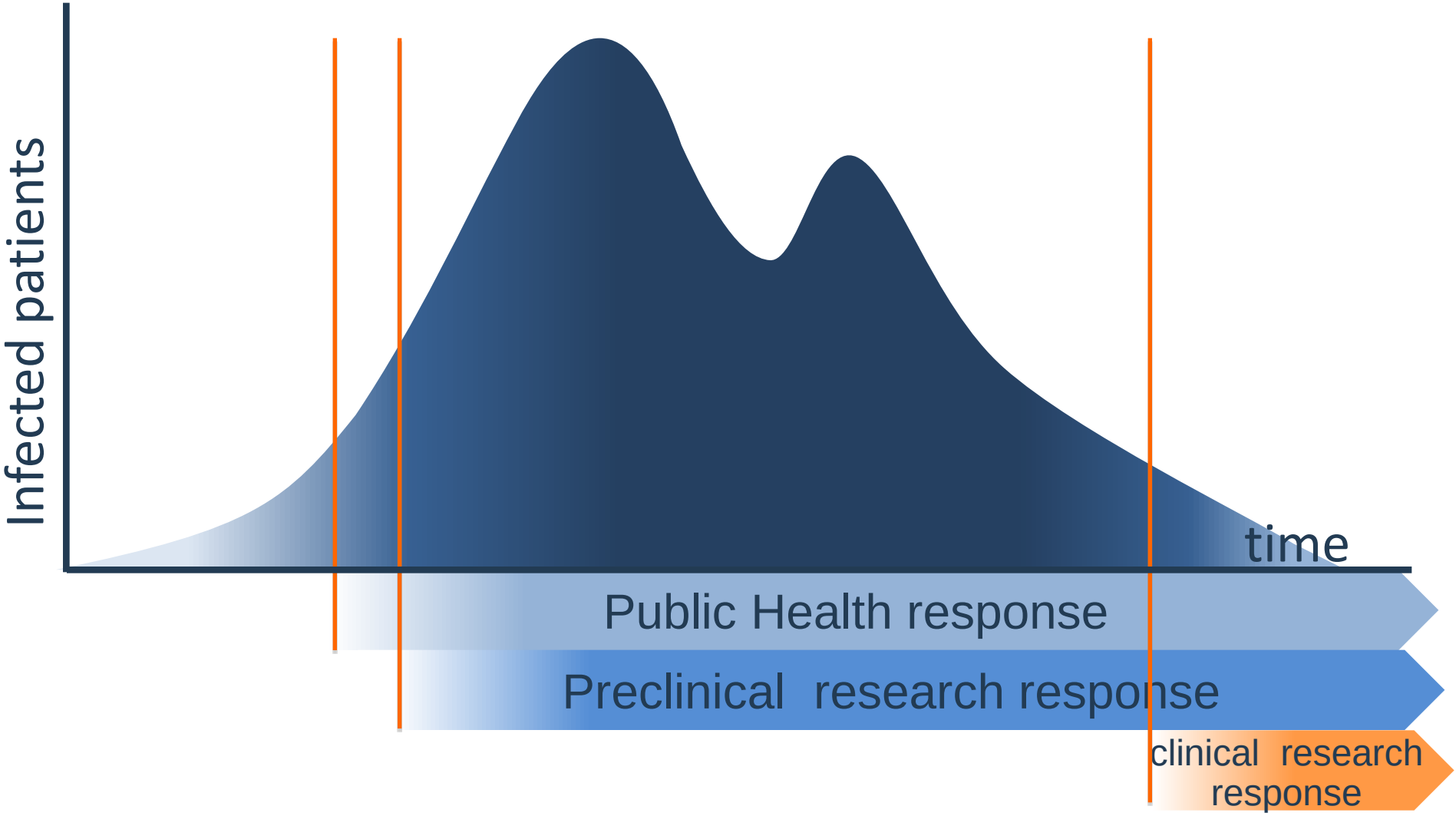
Anne Brissabois — WP co-leader

Detection and analysis of (re-) emerging outbreaks

Ron Fouchier — WP leader

Mark Woolhouse — WP co-leader

Clinical research response to ID outbreaks usually fragmented and too late



Epidemic preparedness research: European Union-supported efforts

- prediction
  - understanding emergence
  - surveillance
  - modelling
- early recognition and containment
  - surveillance
  - clinical awareness
  - infection control
- clinical research
  - pathogen & disease characterization
  - prevention & treatment
- funding
  - rapid responses

PREDEMICS
antigone
Empere

2009-2016
€ 36 M

Compare

2015-2020
€ 21 M

Platform foR European Preparedness Against (Re-)emerging Epidemics
PREPARE

2014-2019
€ 24 M

GloPID-R

2015-2020
€ 3 M

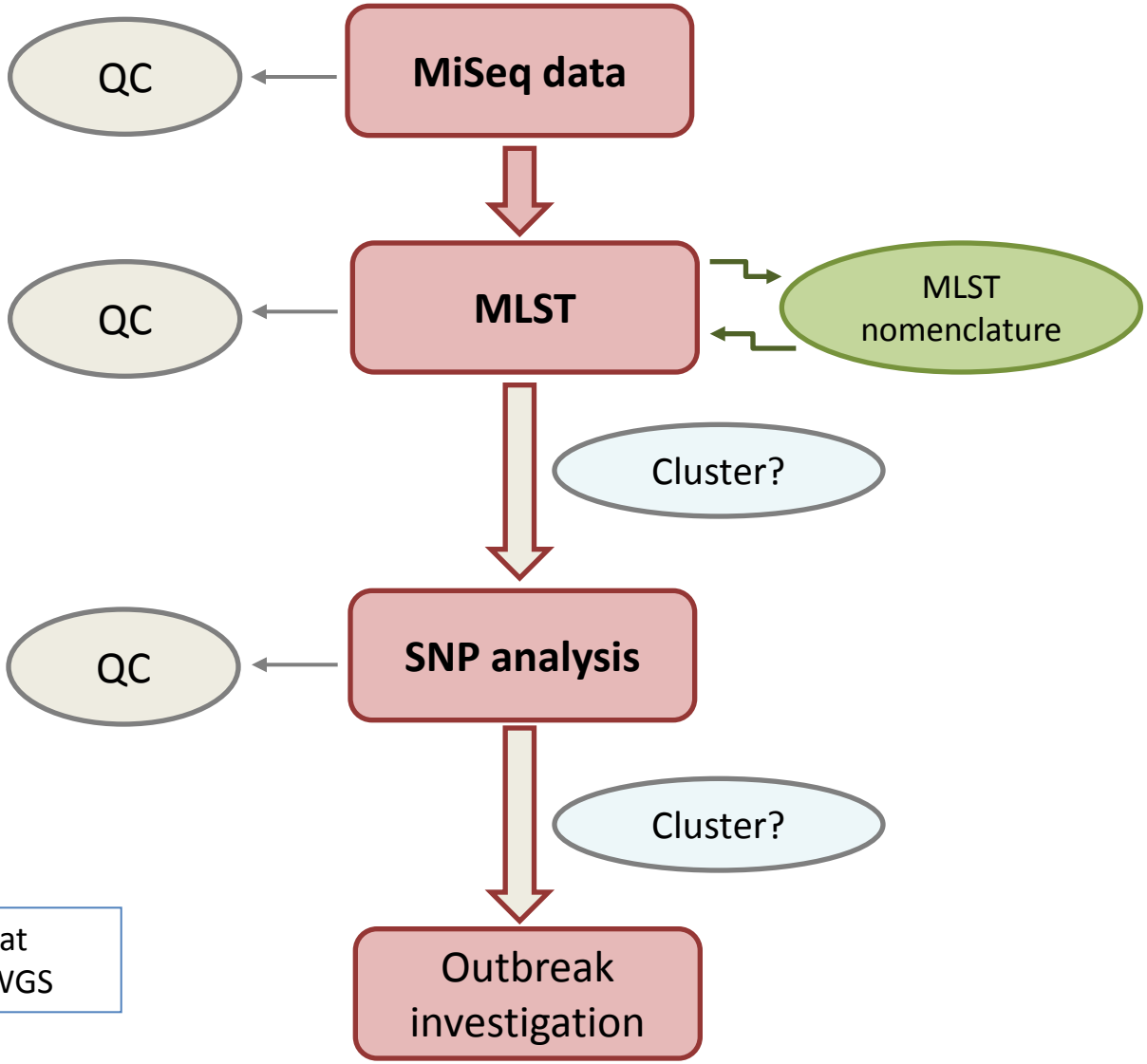
Increasing resolution of foodborne outbreak detection, Listeriosis surveillance in Denmark (courtesy of Eva Moller Nielsen)

Before September 2013:

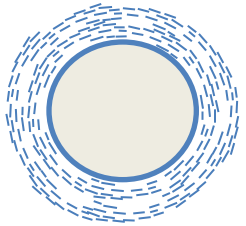
- PFGE all patient isolates
- Interview when suspected outbreak
- Food isolates from official control are stored, not typed

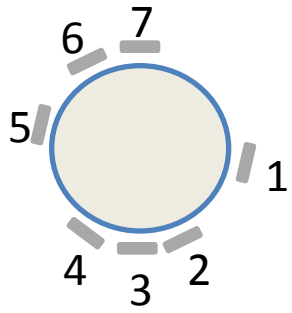
Improved surveillance 2013-14:

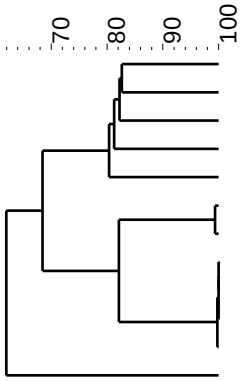
- WGS of patient isolates, weekly
- January 2014: follow-up on all patients (incl. interview when possible)
- June 2014: Isolates obtained by control visits by the food authority (FVST) are submitted for WGS at SSI



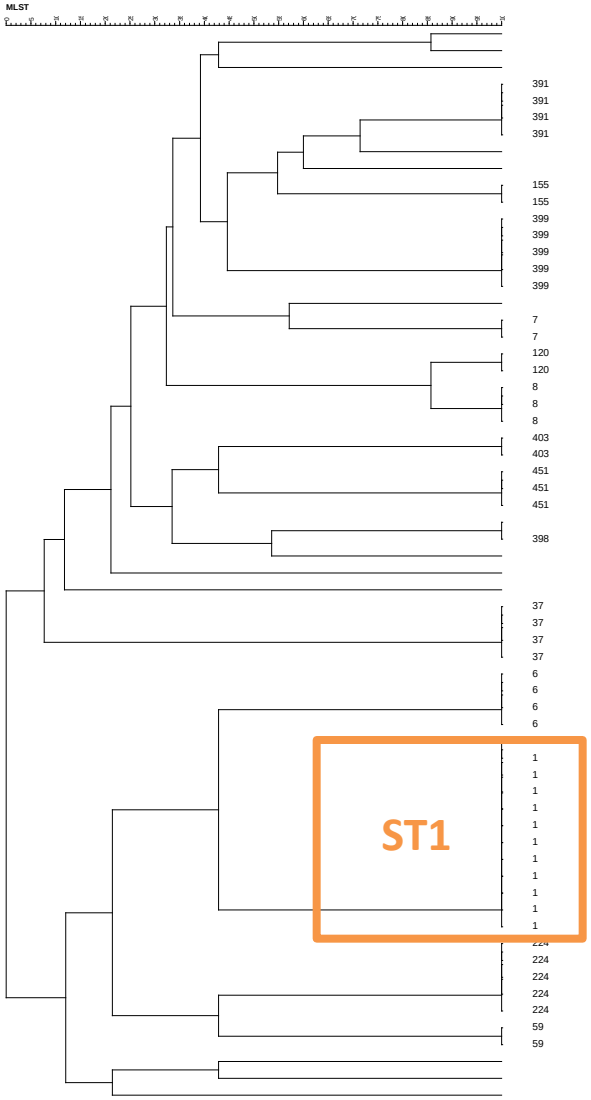
ssi-snp-pipeline at github.com/PHWGS



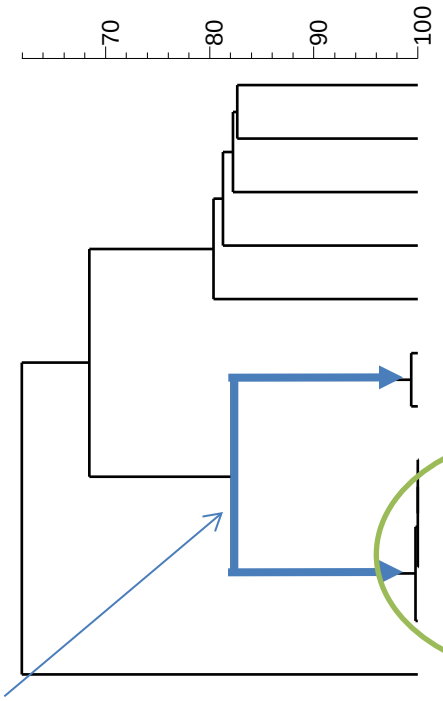




MLST tree



WGS tree: increased resolution



85 SNPs

Patient A
Patient A
Cluster
Cluster
Cluster
Cluster

4 patients
19 weeks
1 SNP

Date

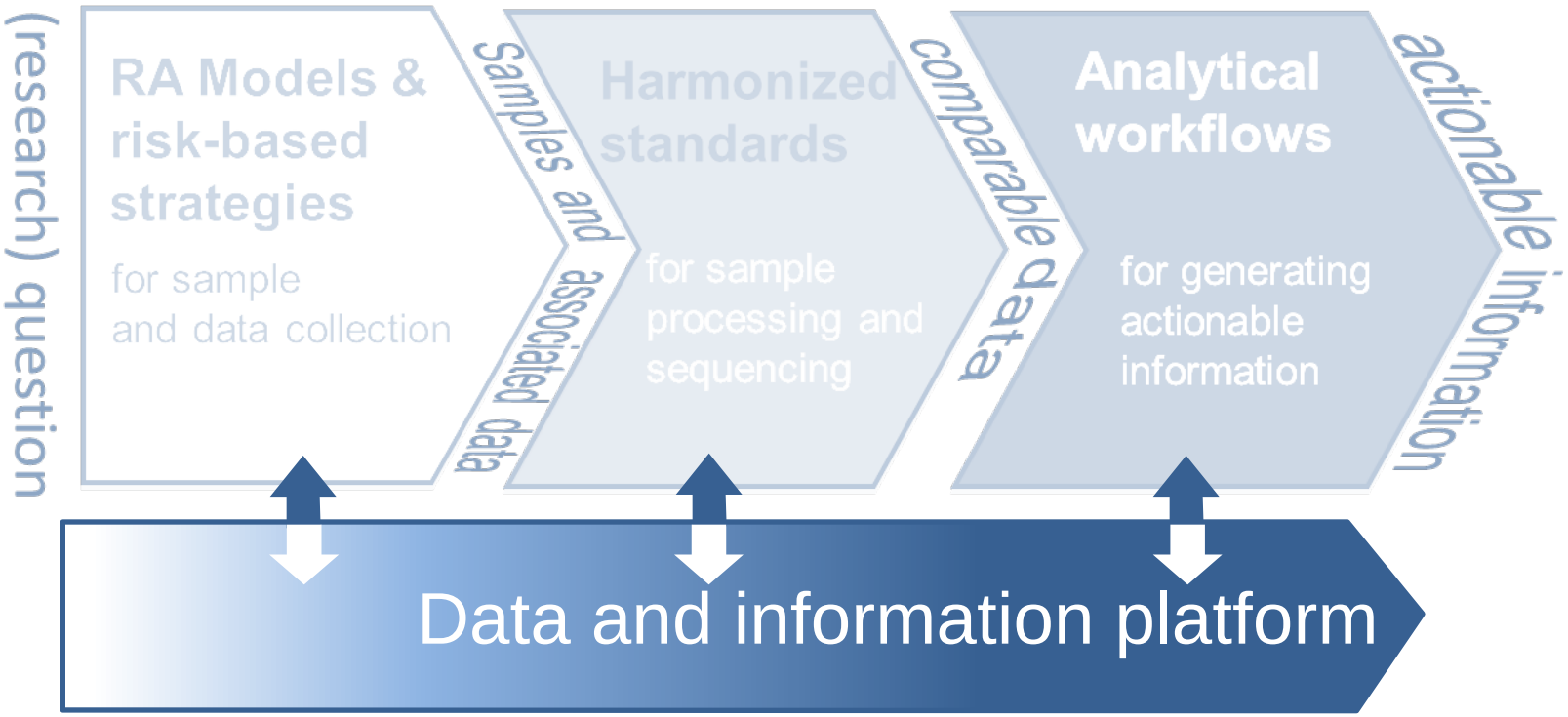
2014 Jan
2013 Marts
2013 Juli
2014 Marts
2013 Jan
2013 Jan
2013 April
2013 April
2013 Aug
2013 Sept
2013 Juni
2014 Jan

Recent competition on a ASM conference (20 Listeria isolates) WGS

- Do the product isolates from facility #1 match the environmental swabs from facility #1?
  - Easy: Pipeline (species, closest match, MLST followed by snp-tree). Around 2 hours to answer.

- Do the product isolates match any other food/environmental isolates currently in the NCBI/SRA database under BioProjects PRJNA211456 or PRJNA215355?
  - Ca. 2 tera-bytes – depending on connection (1 to 10 Gb / h) from 200 to 2,000 hours
    - With 5 days to go we did not manage the challenge

The COMPARE platform

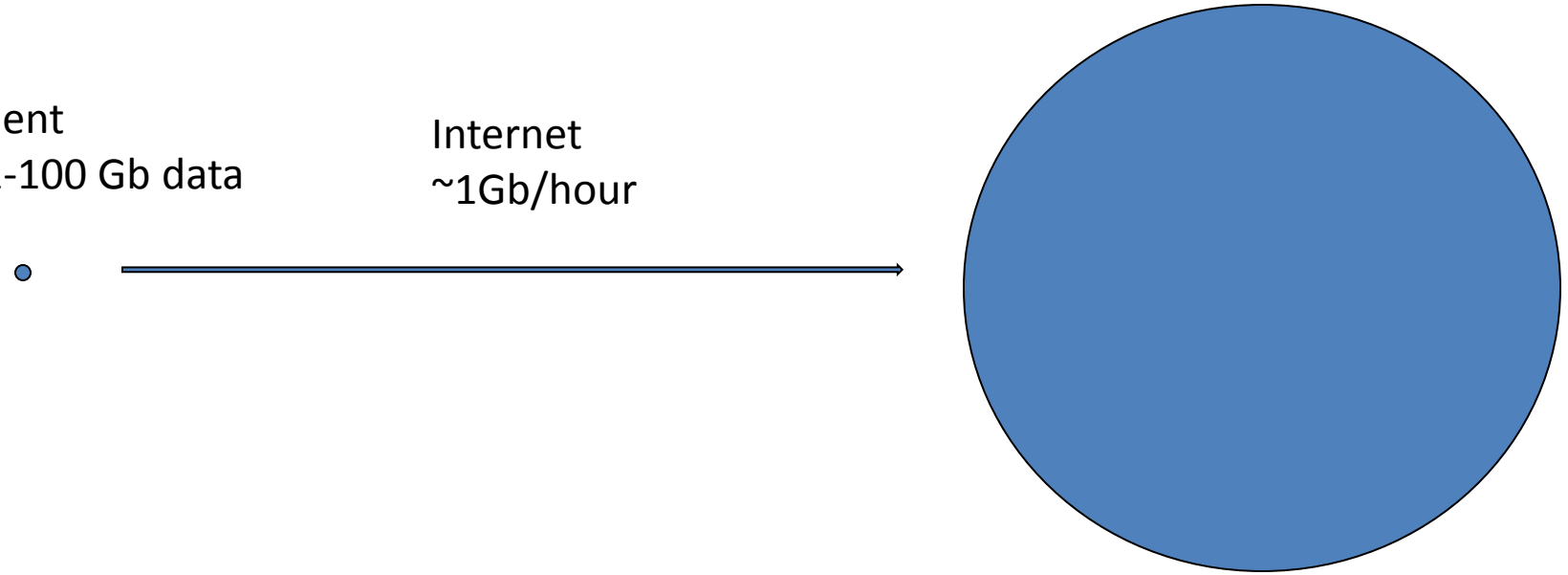


Data comparison problem

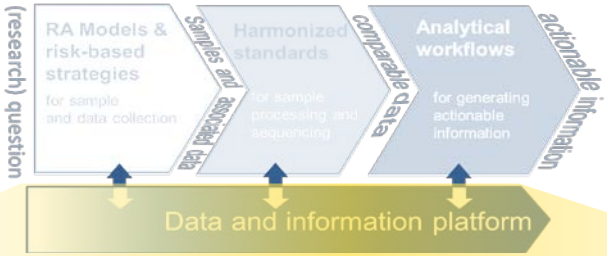
Global repositories
> 1-1000 Tb data

Client
~1-100 Gb data

Internet
~1Gb/hour



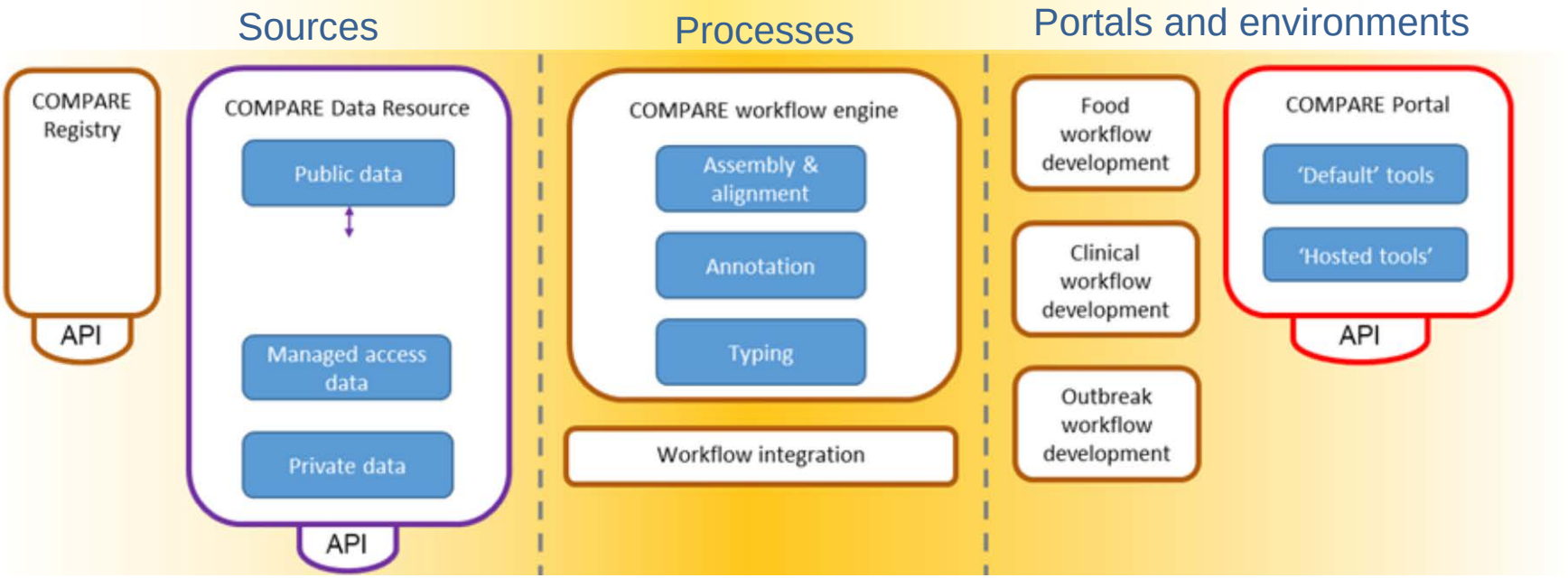
WP9 Information sharing platform



Guy Cochrane
WP leader

Ole Lund
WP co-leader

Istvan Csabai
WP co-leader



EMBL DTU elixir Data for Life

Building on the EU ESFRI Elixir, EMBL and DTU infrastructures

COMPARE

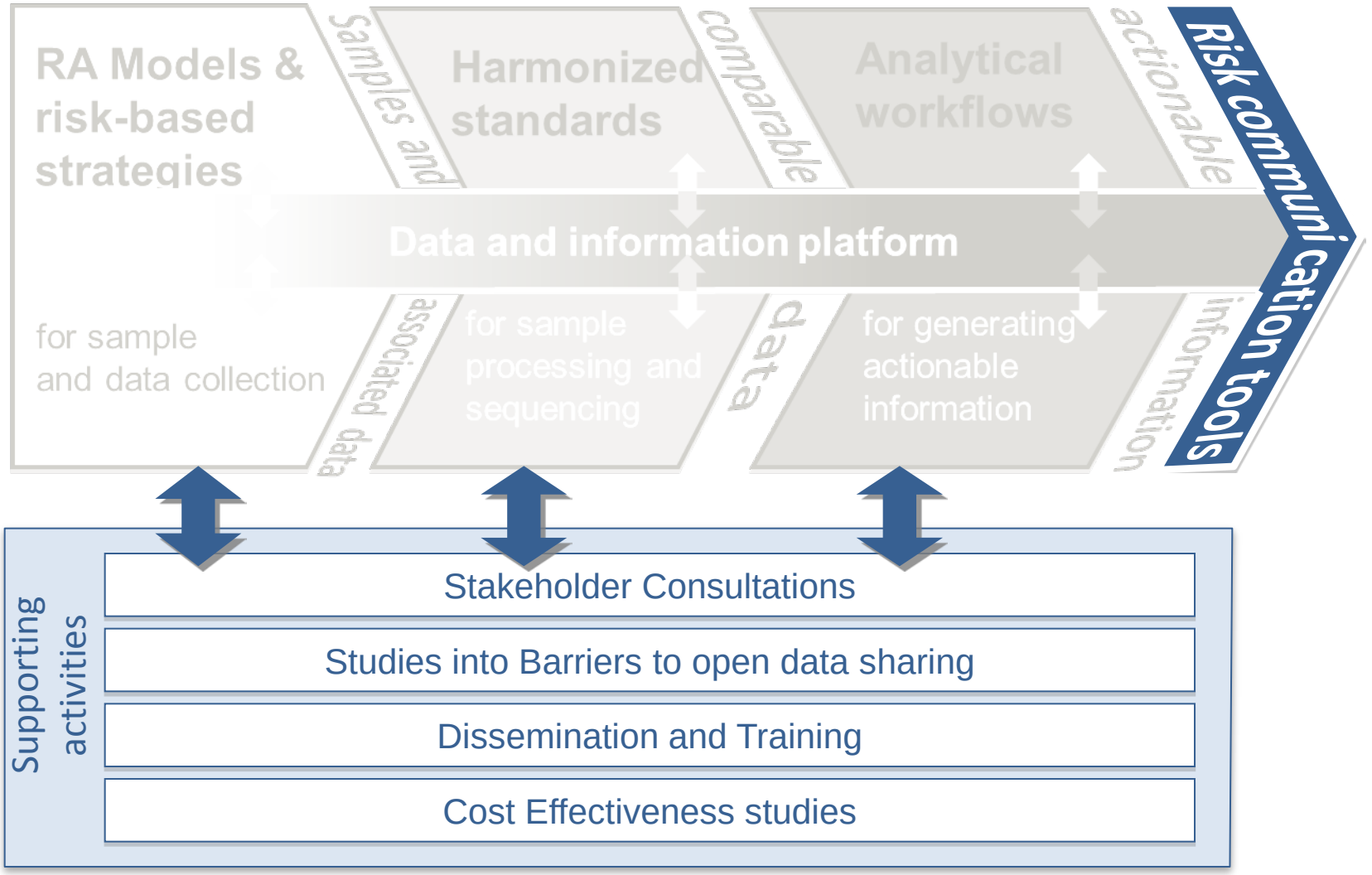
The **CO**llaborative **M**anagement **P**latform for detection and **A**nalyses of (**R**e-) emerging and foodborne outbreaks in **E**urope is a collaboration of 29 institutions with experience in outbreak detection and response in areas of human health, animal health and food safety.

COMPARE Reference Genomes

This COMPARE Reference Genomes page offers a curated selection of published reference sequences covering viral, bacterial and protozoan genomes. These sequences can be searched and retrieved via the following URLs as tagged records in the European Nucleotide Archive (ENA). The complete COMPARE Reference Genomes dataset can be retrieved via the following URL:

http://www.ebi.ac.uk/ena/data/xref/search?source=COMPARE-RefGenome

WPs 10-14: supporting activities

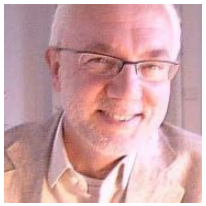


WP12 barriers to open data sharing

The willingness of all envisaged COMPARE users users to rapidly share their data with others is a crucial prerequisite for COMPARE having an impact on rapid pathogen/outbreak detection and mitigation. There are various barriers or bottlenecks to rapid and open sharing of sequence-based data and contextual metadata that influence the impact of COMPARE (e.g Publication priorities, Protection of Foreground, Exploitation of Foreground, fear of loss of control/capacity and capability gaps, Reputation/ Economic damages).

George Haringhuizen

WP leader



The goal of WP12 is to identify, clarify and, as far as feasible, develop practical solutions for Political, Ethical, Administrative, Regulatory and Legal (PEARL) barriers, that hamper the timely and openly sharing of data through COMPARE.

Conclusions

- WGS/NGS is rapidly entering diagnostic and public health arena, with near real time data generation
- Sequence platforms rapidly developing, cheaper, simpler
- Bottleneck at level of bioinformatics, particularly for intergroup comparison, national, international
- COMPARE aims to develop infrastructure and ICT to meet the coming demand
- In 1-2 years, we will be seeking partners for pilot projects

http://globalhealth.ie