



wellcome trust

sanger
institute

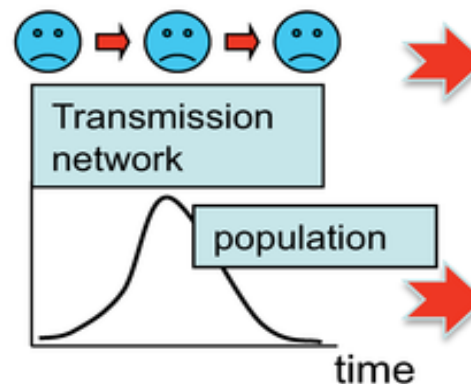
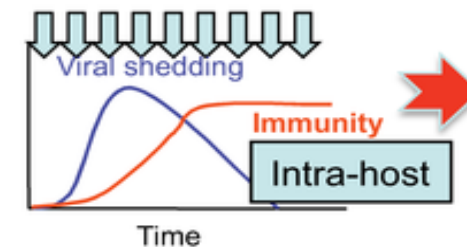
**‘The transformation of viral epidemiology
and clinical studies by next-generation
sequencing’**

Utility of virus genomes

- Sequencing virus genomes
 - The individual
 - National epidemic tracking
 - Zoonotic chatter
 - The beginning of an outbreak
- Endemic disease



Sampling scale



Metapopulation/regional/
global

ACTIONABLE &
INTERPRETABLE
INFORMATION

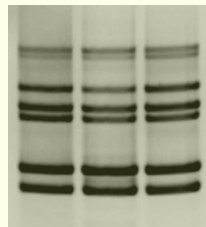
Genome sequencing 2016



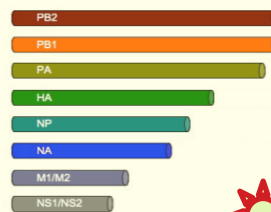
1. Samples & phenotype



2. DNA/RNA



3. Genome Amplification or enrichment



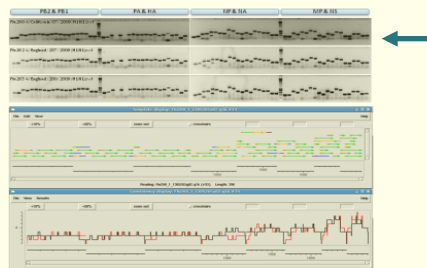
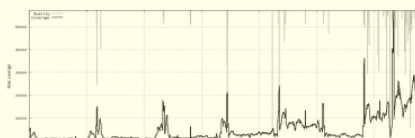
4. MiSeq



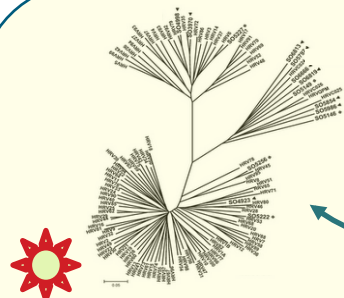
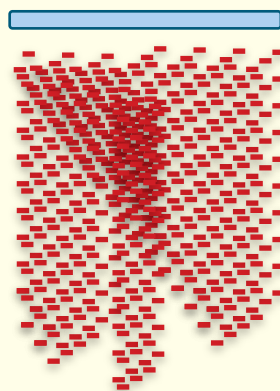
4. Ion Torrent



4. MinIon



5. Assembly + QA/QC

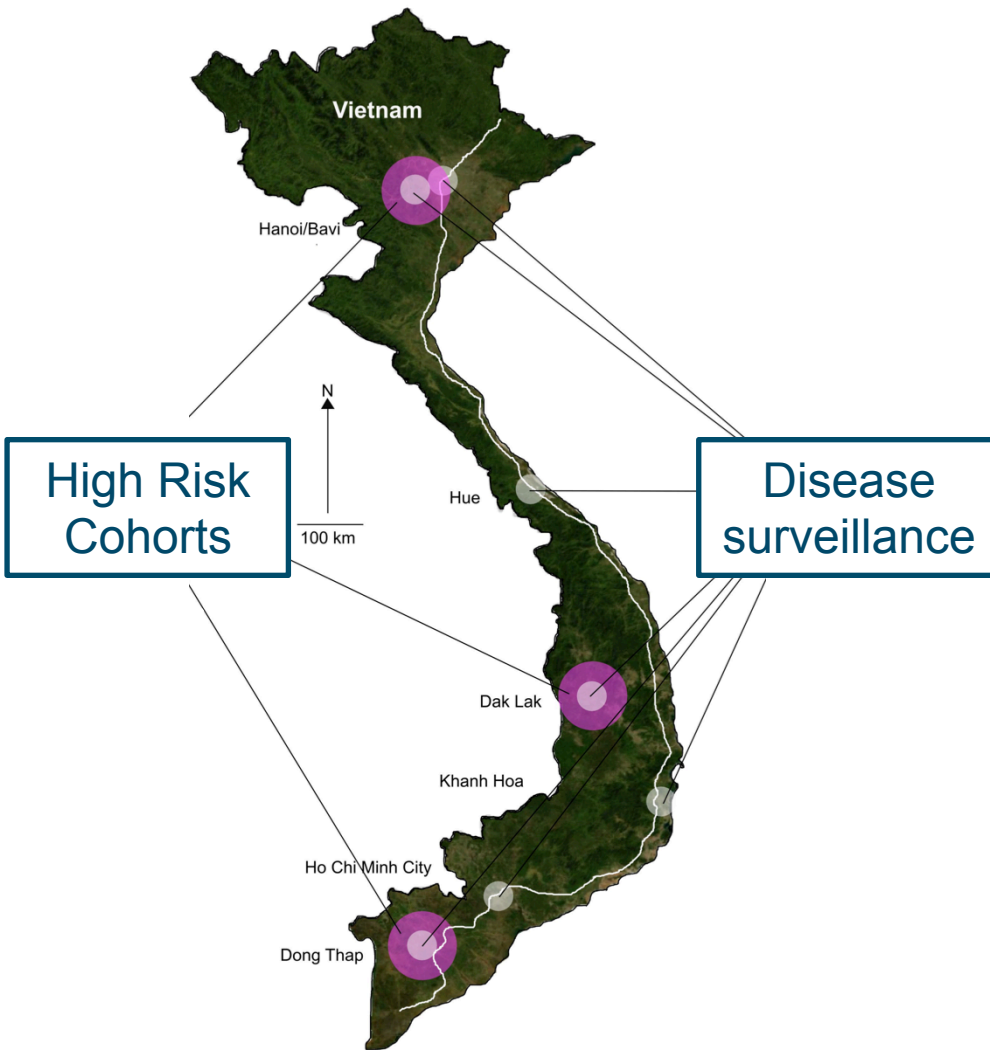


6. Analysis + data integrations

- **Zoonotic chatter – Rotaviruses in Vietnam**
- **The beginnings of human transmission – MERS CoV**
- **Major outbreaks – Ebola virus**
- **Infection control in the UK – Influenza virus**

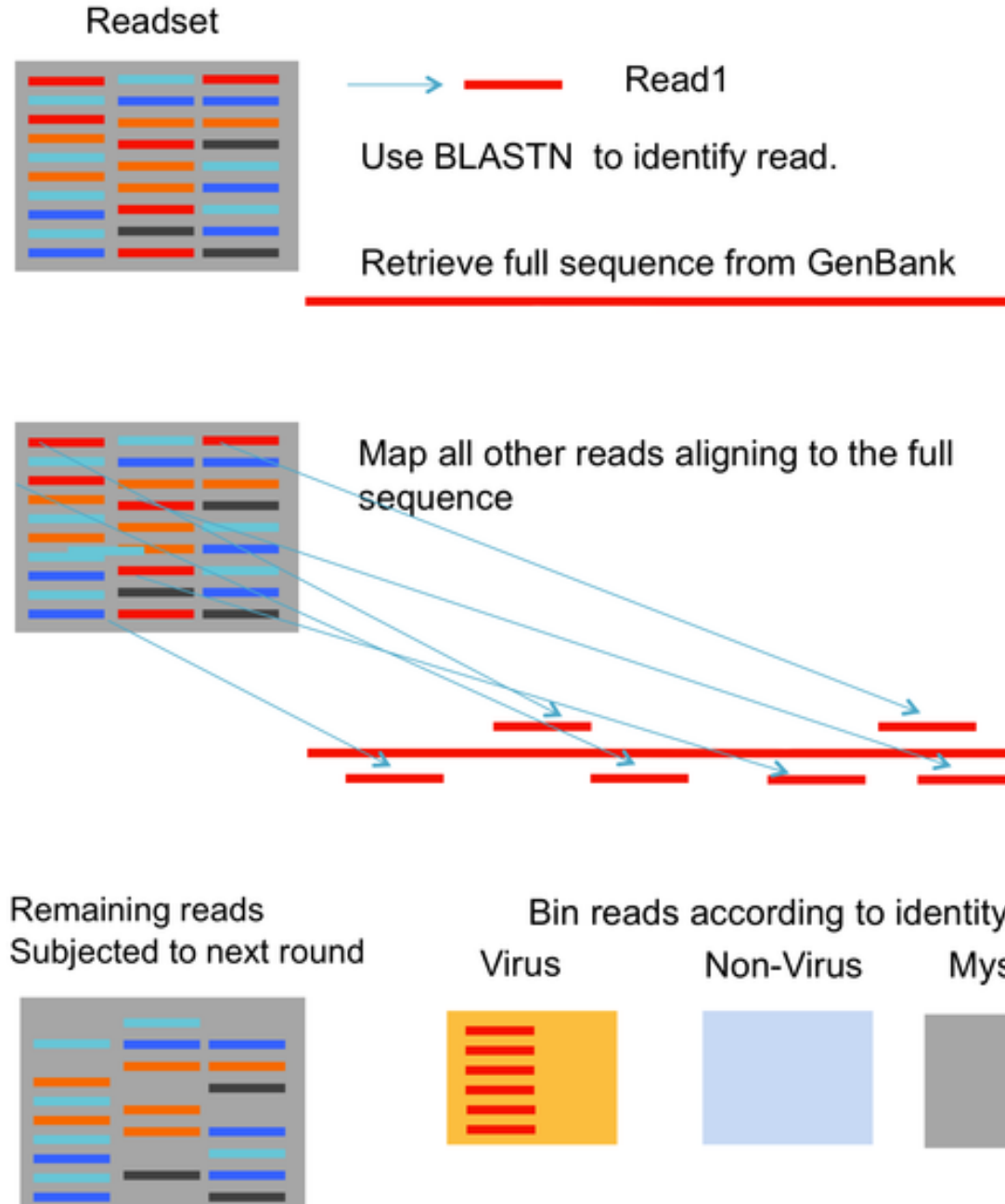
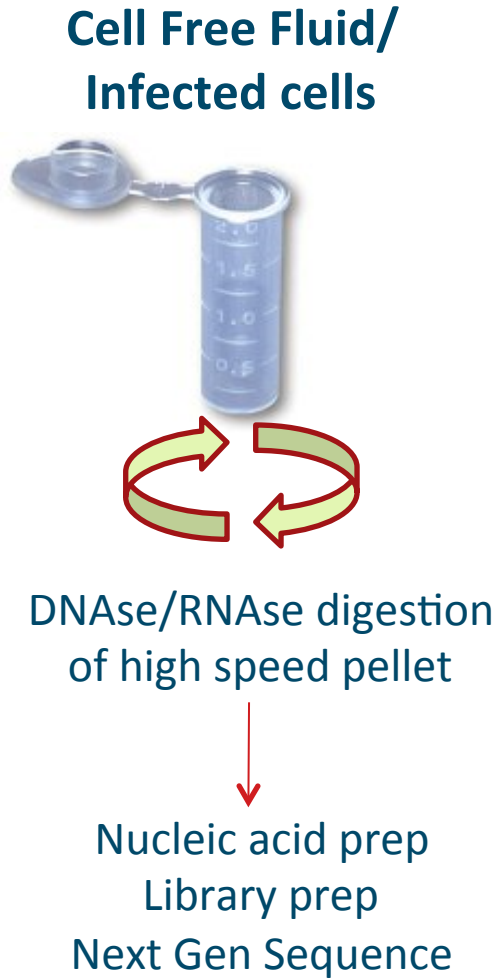
Sampling frameworks - zoonosis

The work of My Phan and Matt Cotten in collaboration with Stephen Baker

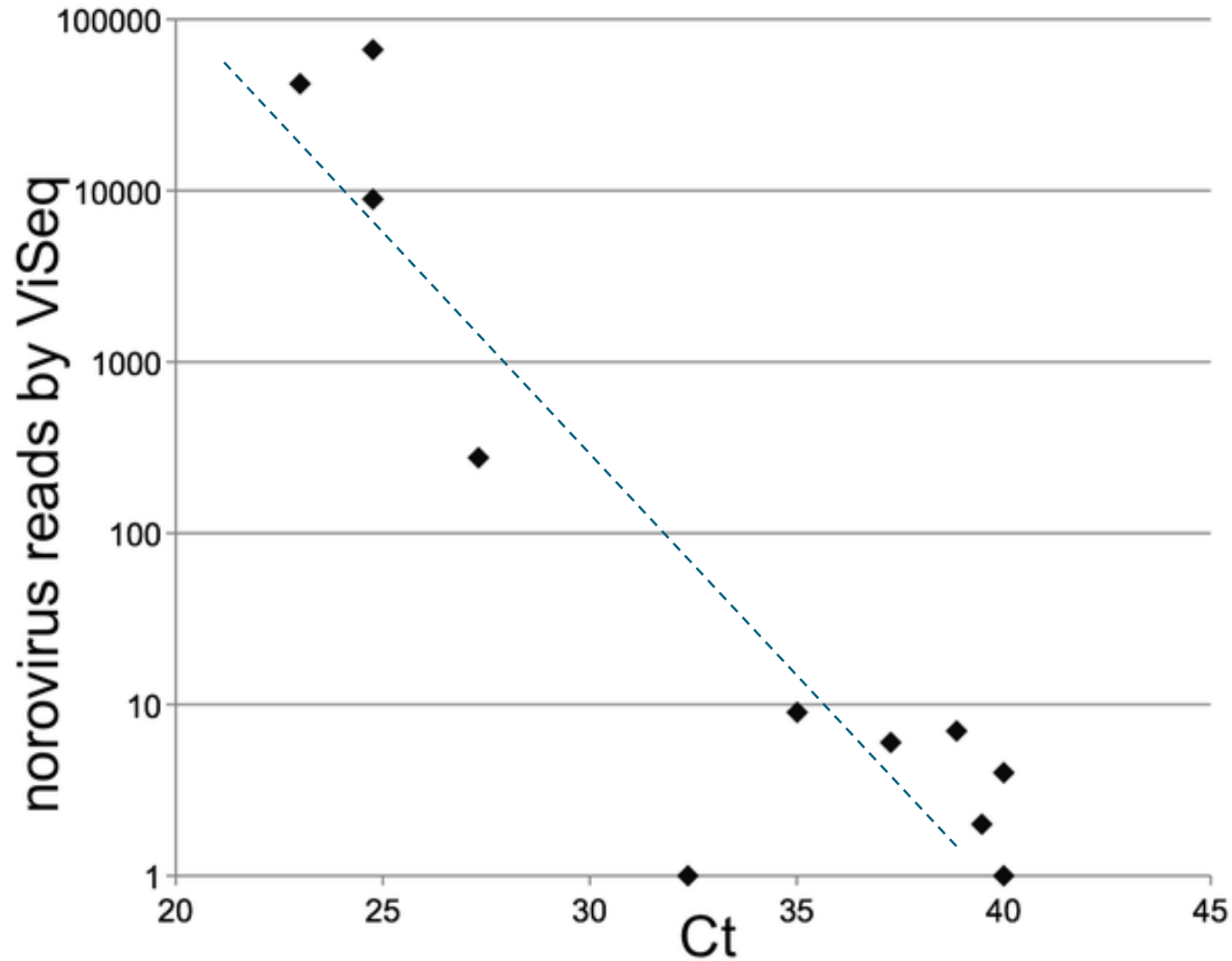


Pilot work of Wellcome Trust VIZIONS

Agnostic Virus Genome Detection



ViSeq method Sensitivity - Norovirus



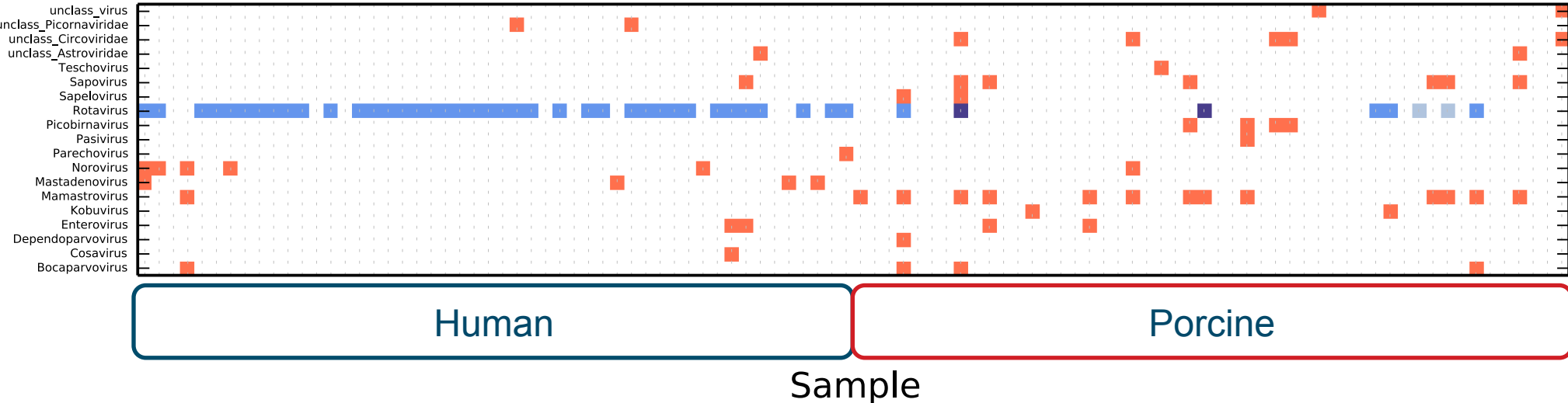
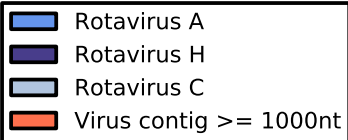
Enteric virus content of human, porcine fecal samples

Criteria to be a virus:

A contiguous assembled sequence of length x (i.e. <1000bp)

A minimum read depth per base for the contig. (i.e.100)

Virus contents in all samples

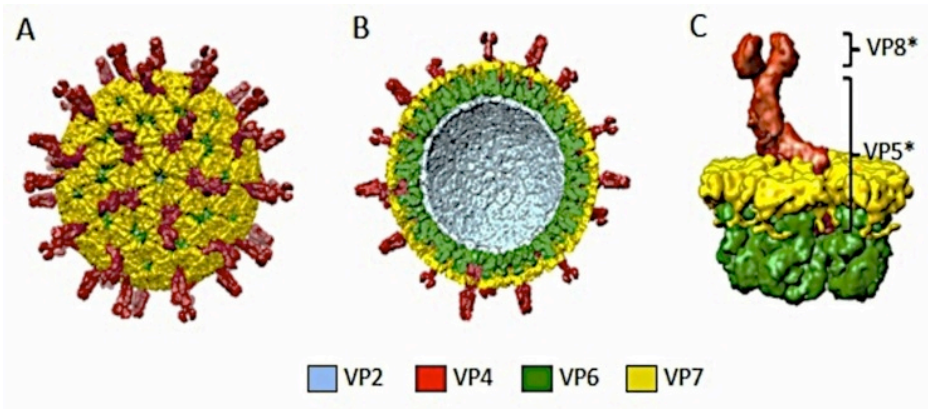


Rotavirus genomes

Reoviridae – Rotavirus

For Group A rotaviruses

Major classification based on G (VP7, seg 9) and P (VP4, seg 4) - targets of neutralizing antibodies



Extended classification

Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx

27 37 17 9 9 8 18 10 12 15 11

Major Human Group A

G1P[8], G2P[4], G3P[8], G4P[8], G9P[8]

on a genotype constellation

I1-R1-C1-M1-A1-N1-T1-E1-H1











or

I2-R2-C2-M2-A2-N2-T2-E2-H2

Rotavirus genome segment reassortment

		Genotype constellation										Host count		
		VP7	VP4	VP6	VP1	VP2	VP3	NSP1	NSP2	NSP3	NSP4	NSP5	Human	Pig
Mono-infection	G1	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	33	0	
	G2	P[4]	I2	R2	C2	M2	A2	N2	T2	E2	H2	12	0	
	G3	P[8]	I1	R1	C1	M1	A1	N1	T1	E1	H1	2	0	
	G4	P[6]	I1	R1	C1	M1	A8	N1	T7	E1	H1	1	1	
	G4	P[6]	I1	R1	C1	M1	A8	N1	T1	E1	H1	0	6	
	G5	P[13]	I5	R1	C1	M1	A8	N1	T7	E1	H1	0	1	
	G11	P[13]	I5	R1	C1	M1	A8	N1	T7	E1	H1	0	1	
	G9	P[23]	I5	R1	C1	M1	A8	N1	T1	E1	H1	0	1	
	G5	P[13]	I5	R1	C1	M1	A8	N1	T7	x	H1	0	1	
	G1	P[8]	I2	R2	C2	M2	A2	N2	T2	E2	H2	3	0	
	G2	P[8]	I2	R2	C2	M2	A2	N2	T2	E2	H2	1	0	
G1	P[8]	x	R1	C1	M1	A1	N1	T1	E1	H1	1	0		
Mixed infection	G1/G2	P[8]/P[4]	I1/I2	R1	C1/C2	M1/M2	A1/A2	N1/N2	T1/T2	E1/E2	H1/H2	1	0	
	G1/G2	P[8]/P[4]	I1/I2	R1/R2	C1	M1/M2	A1/A2	N1/N2	T1/T2	E1/E2	H1	1	0	
	G1/G2	P[8]/P[4]	I1/I2	R1/R2	C1/C2	M1/M2	A1/A2	N1/N2	T1/T2	E2	H2	1	0	
	G1/G2	P[8]/P[4]	I1/I2	R1/R2	C1/C2	M1/M2	A1/A2	N1/N2	T1/T2	E1/E2	H1	1	0	
	G1/G2	P[8]/P[4]	I1/I2	R1/R2	C1/C2	M1/M2	A1/A2	N1/N2	T1/T2	E1/E2	H1/H2	1	0	
	G1/G2	P[8]/P[4]	I1/I2	R1	C1/C2	M1	A1/A2	N1/N2	T1/T2	E1/E2	H1/H2	1	0	
	G1	P[8]	I1	R1	C1	M1	A1	N1/N1	T1	E1	H1	1	0	
	G1/G4	P[8]	I1	R1	C1	M1	A1/A8	N1/N1	T1/T1	E1/E1	H1/H1	0	1	
	G9/G11	P[13]/P[23]	I5/I5	R1/R1	C1	M1	A8/A8	N1/N1	T1/T7	E1/E1	H1	0	1	

Reassortments

	Wa-like genotype (G1/G3/G4/G9-P[8]-I1-R1-C1-M1-A1-N1-T1-E1-H1)		G5
	DS-1-like genotype (G2/G8-P[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2)		G11
	G4-P[6]		I5
	A8-T7		P[13]
	G9-P[23]		x Sequence not determined

Strain by Host

- Human_Other countries
- Human_VNM
- Pig_Other countries
- Vaccine strains
- Vizion_Human
- Vizion_Pig
- Cow
- Mouse

Branch by genotype

- G1
- G2
- G3
- G4
- G5
- G6
- G9
- G11
- G26
- Other genotype

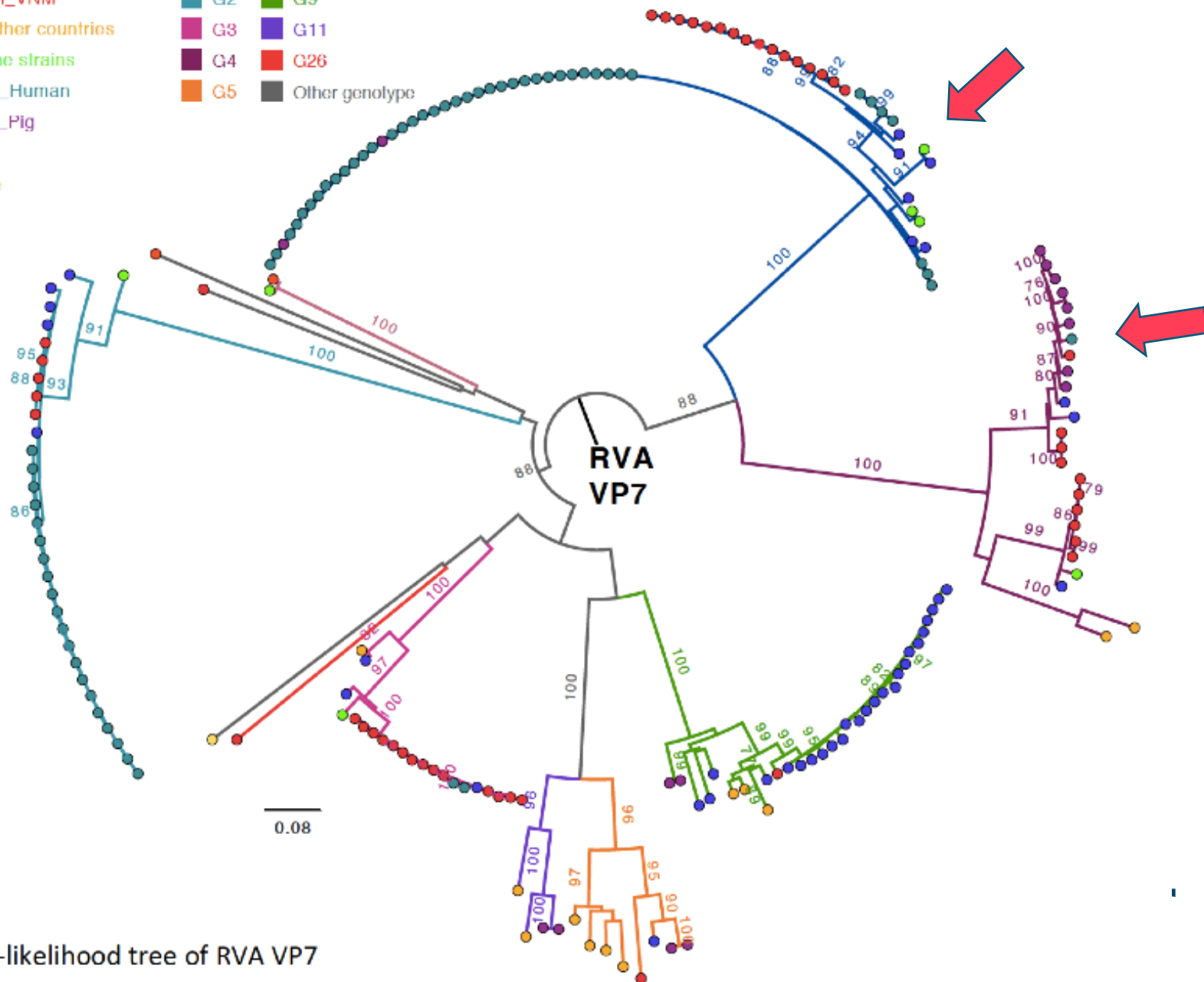
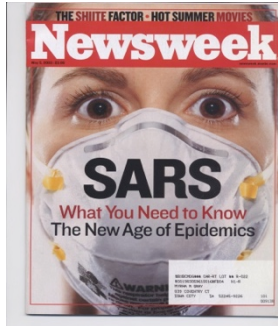


Figure 3A. Maximum-likelihood tree of RVA VP7

Virus genomes and zoonotic chatter

- **Human and animal contact areas**
 - Random sequencing
 - Risk groups
 - Random sampling of human clinical samples
 - Random sampling of veterinary samples
- **Linked to serology surveys and risk maps**
 - Virus genetic diversity and genome movement risk maps
- **Requires simple sample preparation, sequencing and assembly methods (commercial?)**

12 years of Virus Outbreaks



SARS
Coronavirus

H5N1
Influenza A

H1N1
Influenza A

H7N9
Influenza A

MERS
Coronavirus

Ebola
virus

2002-3

2003-4

2009-10+

2013

2012-5

2014-5

8273 cases
775 deaths

630 cases
375 deaths

global
~579 000 deaths

134 cases
44 deaths

1266 cases
388 deaths

28 103 cases
11 290 deaths

§CFR ~10%

CFR ~60%

CFR ~0.01%

CFR ~33%

CFR ~30%

CFR ~40%

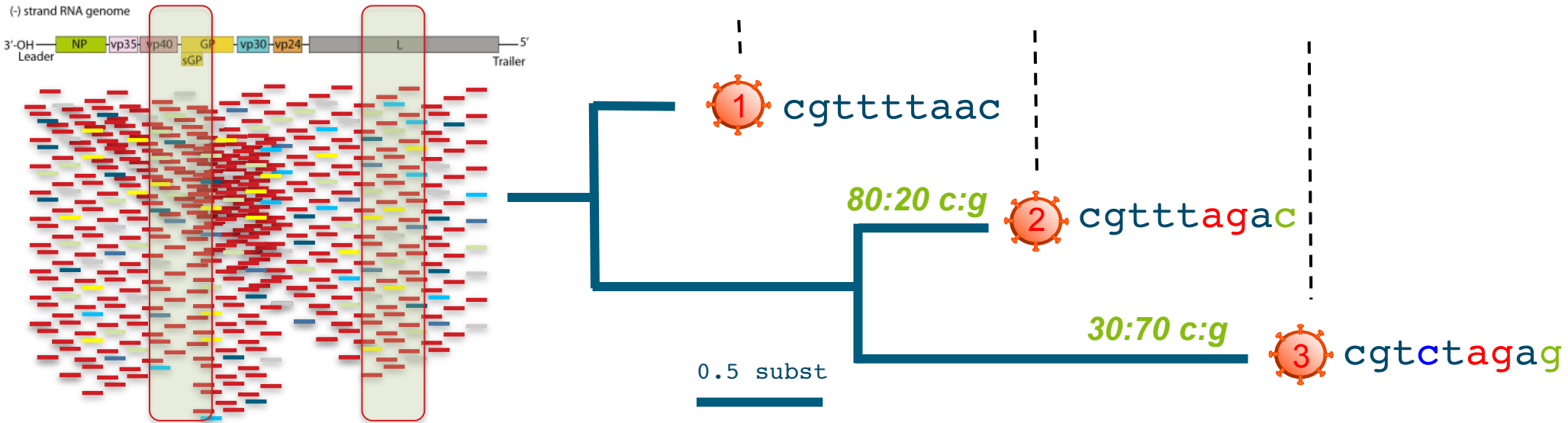
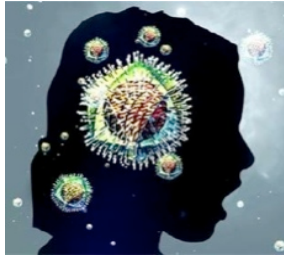
In an outbreak there should be a commitment to turn a portion of RESIDUAL diagnostic nucleic acid into a publically available pathogen genomes at **NO** additional cost to the country(s) experiencing the outbreak & that the data leads to

SHARED, ACTIONABLE & INTERPRETABLE INFORMATION

§Case fatality rate

Virus genomes & molecular epidemiology

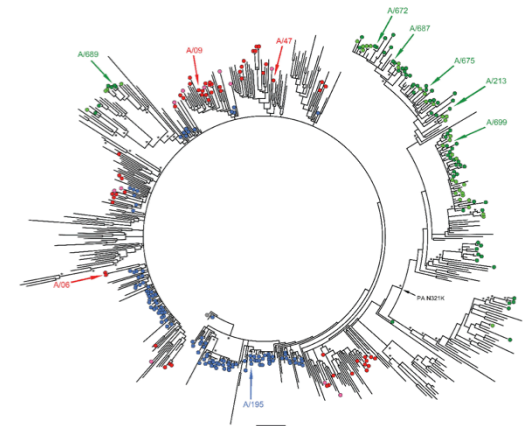
Phylodynamics, a term coined to denote the interplay between evolution and epidemiology when occurring on the same timescale



Consensus genome

Samples

Samples/Diagnostics

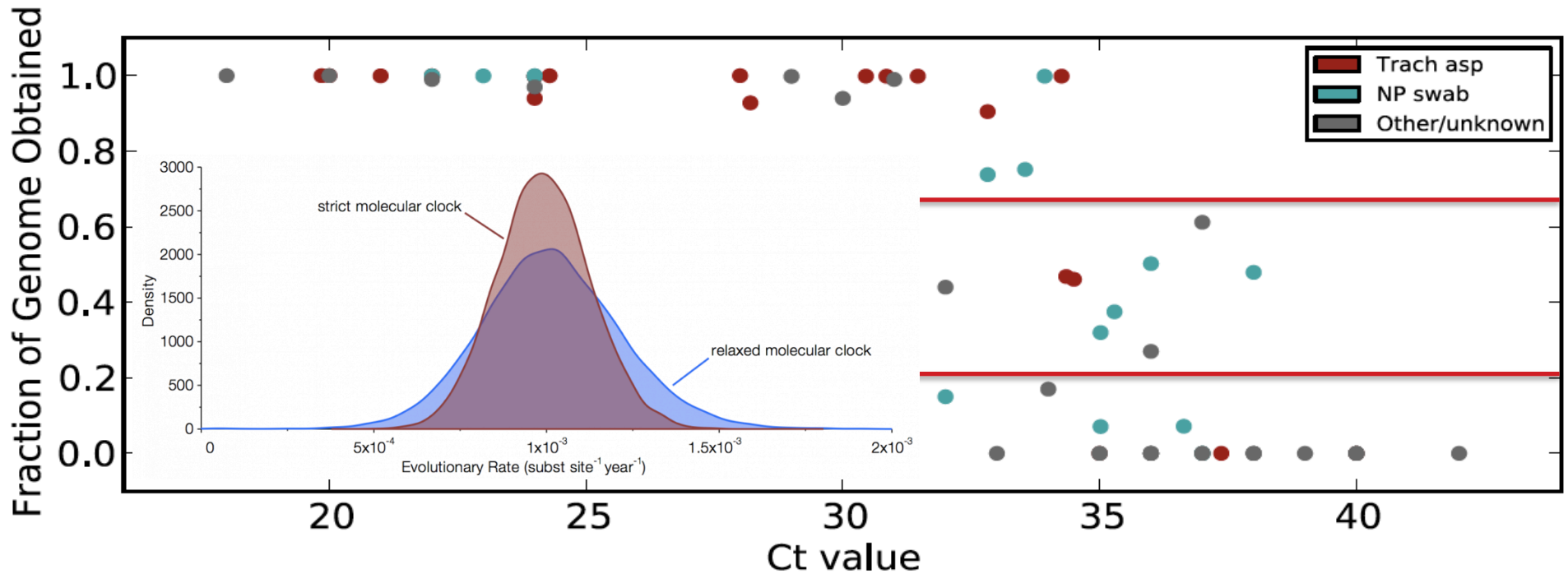
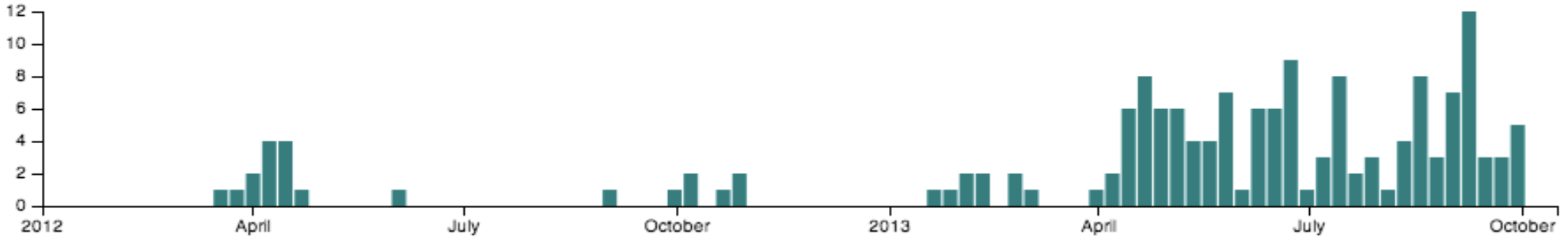


Clinical reporting

Molecular diagnostics

Clinically actionable/useful virus genome

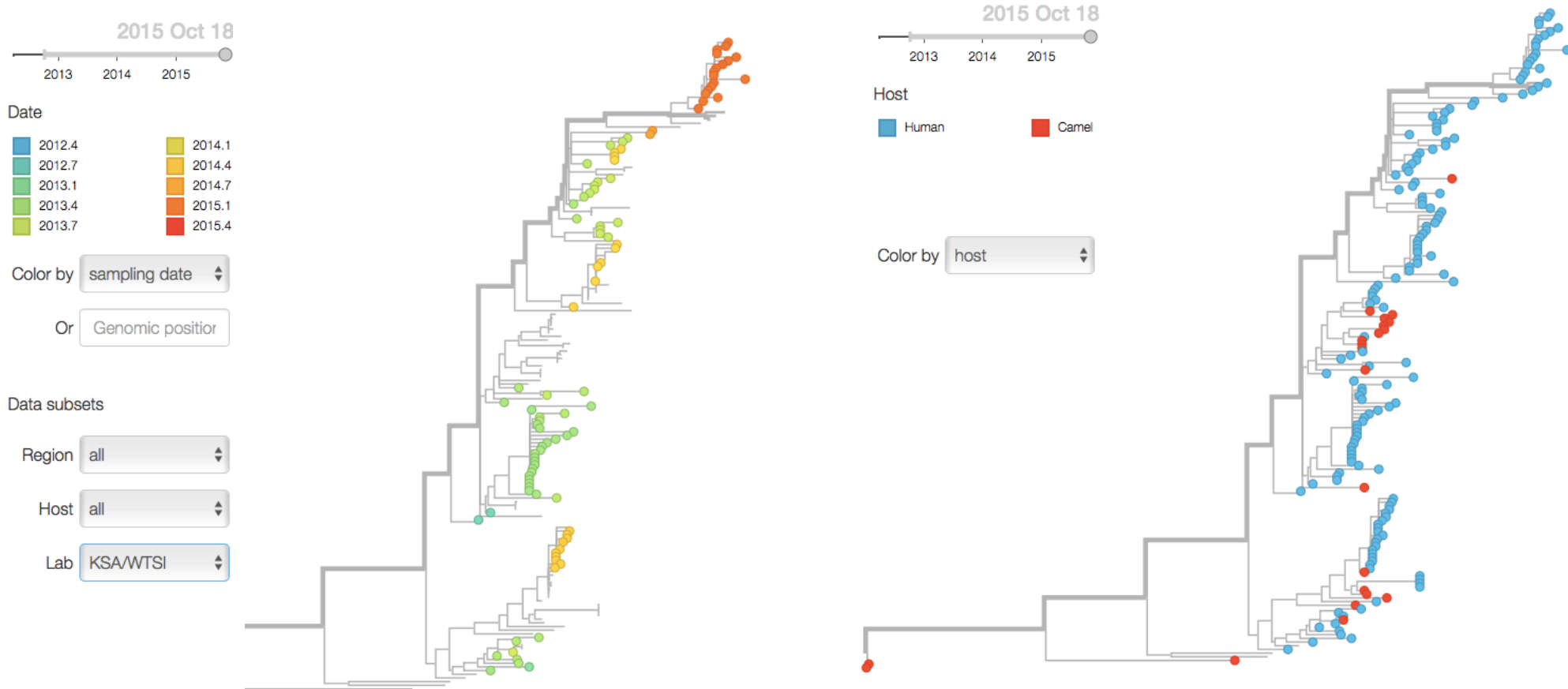
Cases by date



MERS-CoV – 30Kb RNA genome

Sharing data allows new insights

<http://mers.nextflu.org>

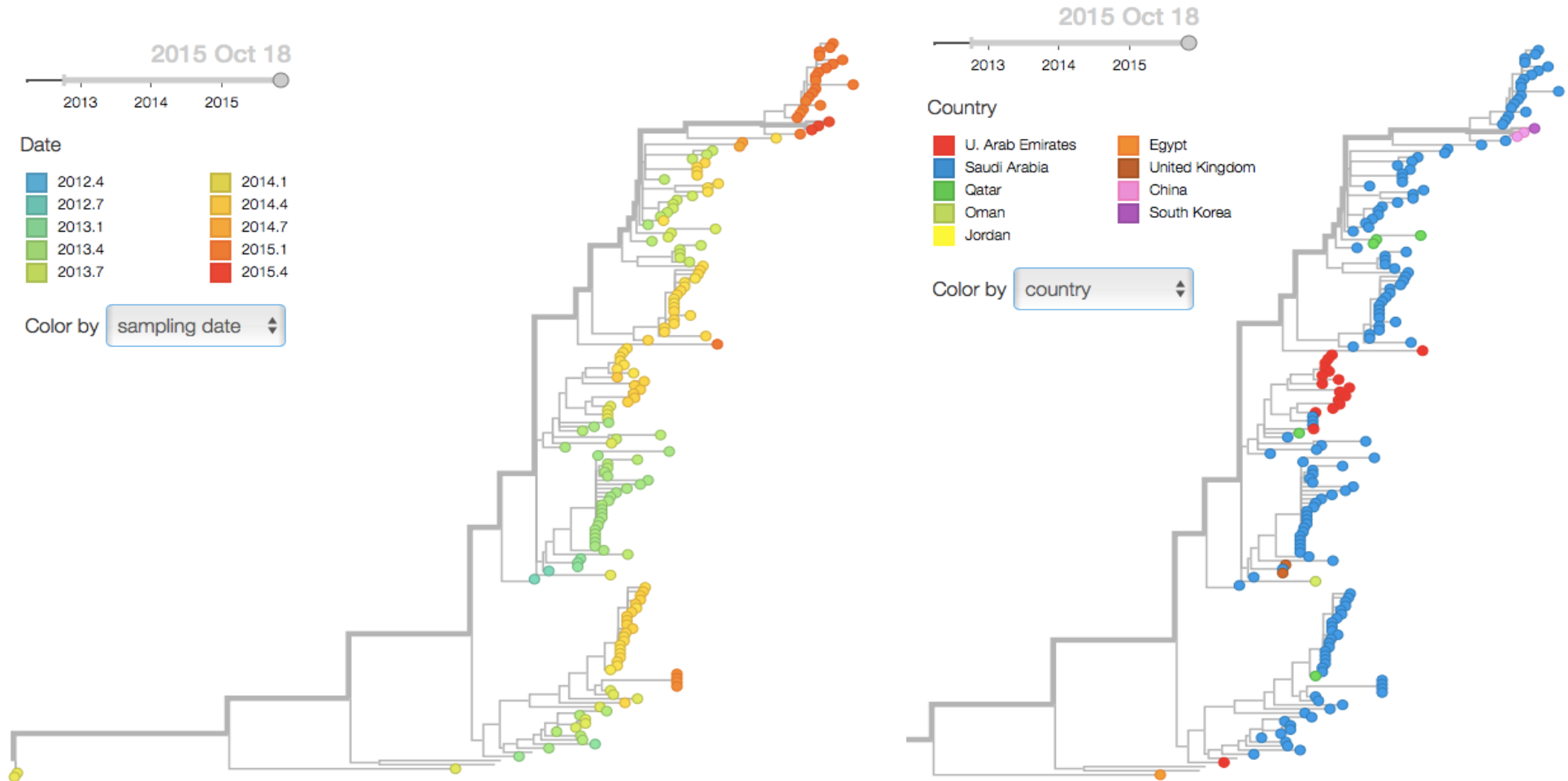


Assiri et al (2013), NEJM 369 (5), 407;

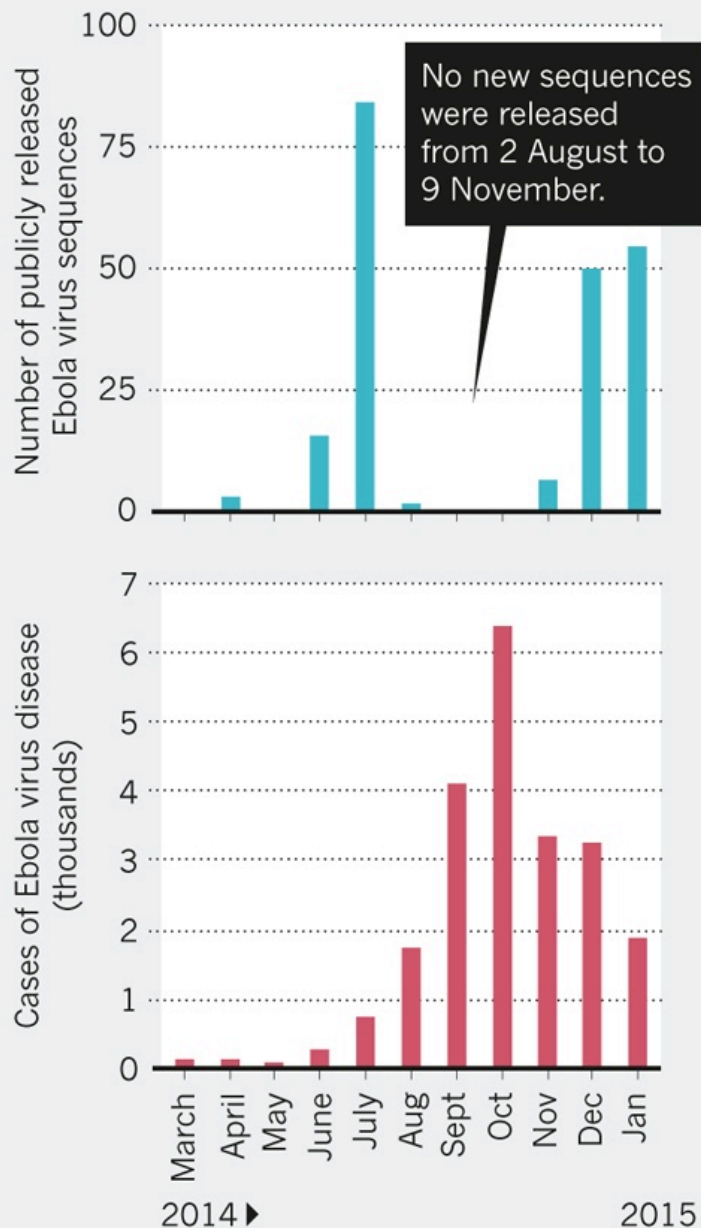
Cotten et al (2013), Lancet 382 (9909), 1993-2002;

Cotten et al (2014), mBio 5(1), e01062-13

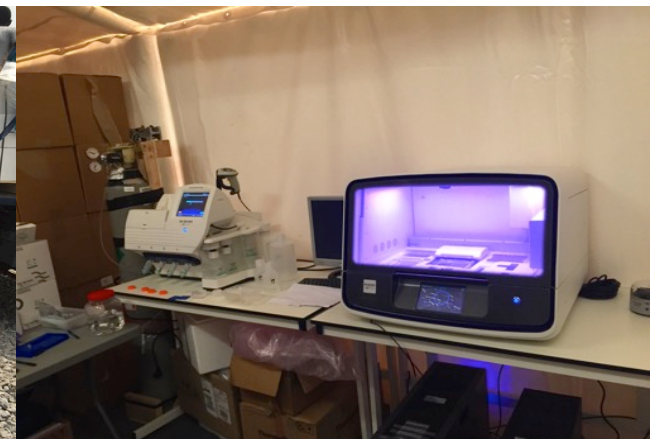
Shared data allows real-time updates



Ebola virus genome sustained local sequencing



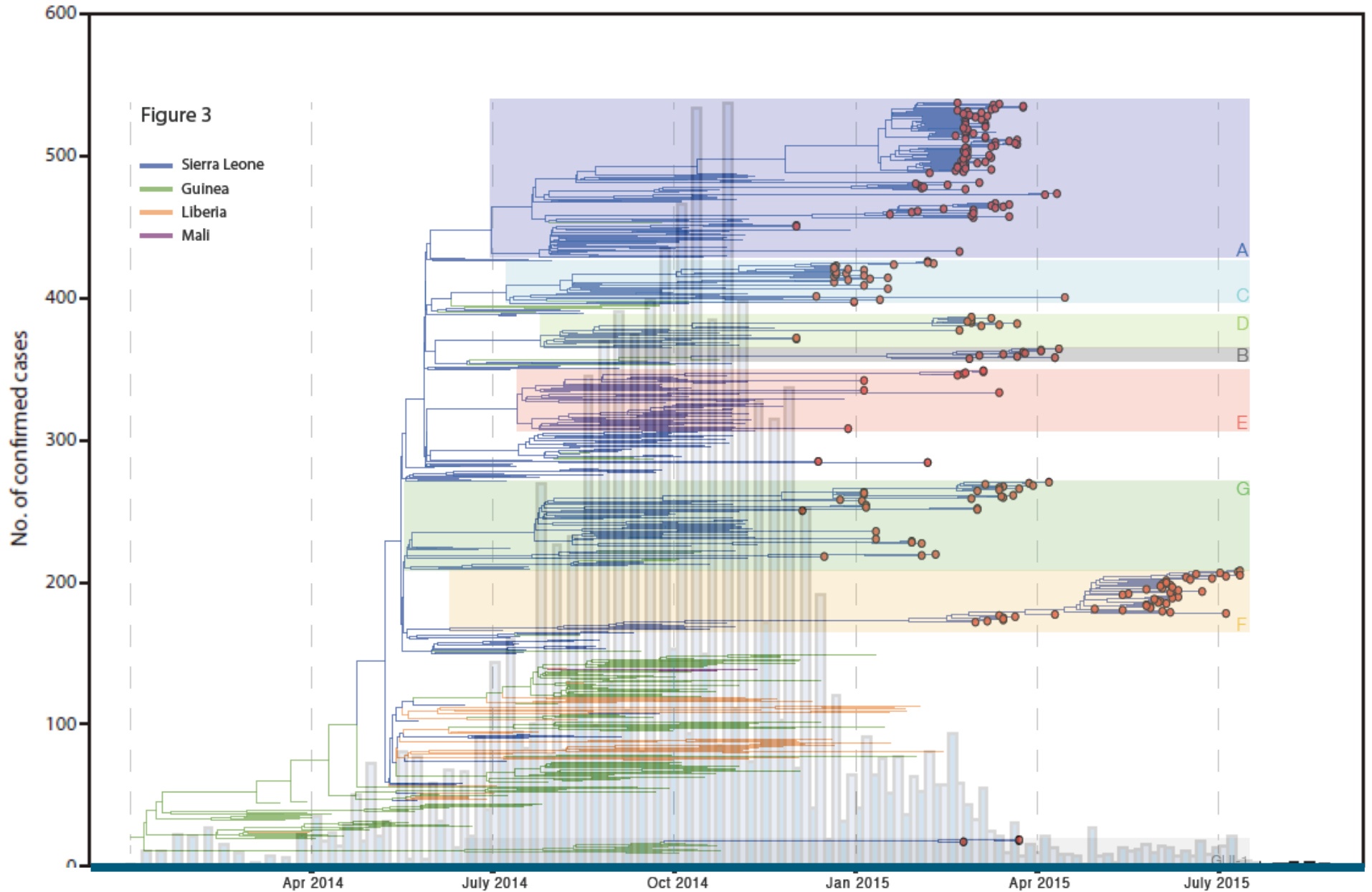
Move the data not the samples



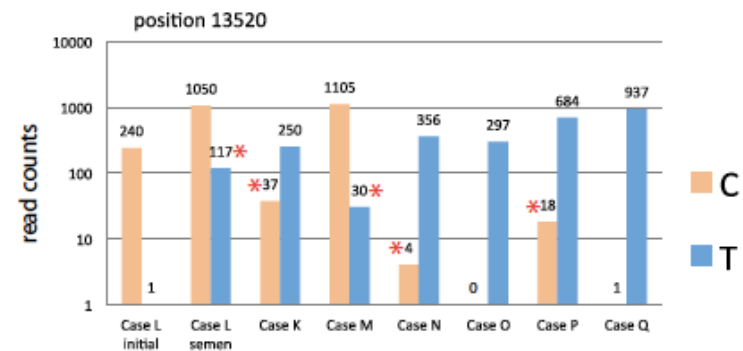
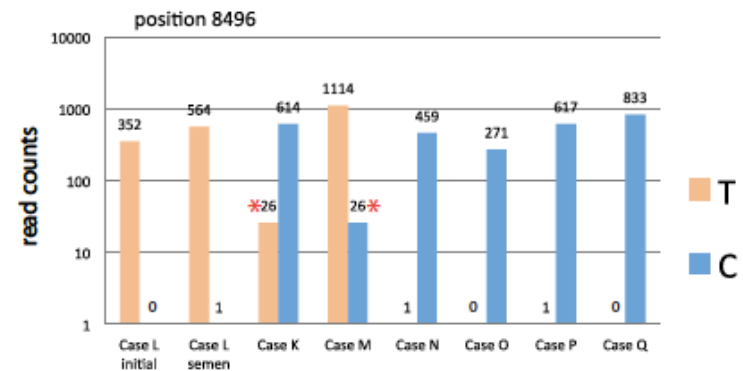
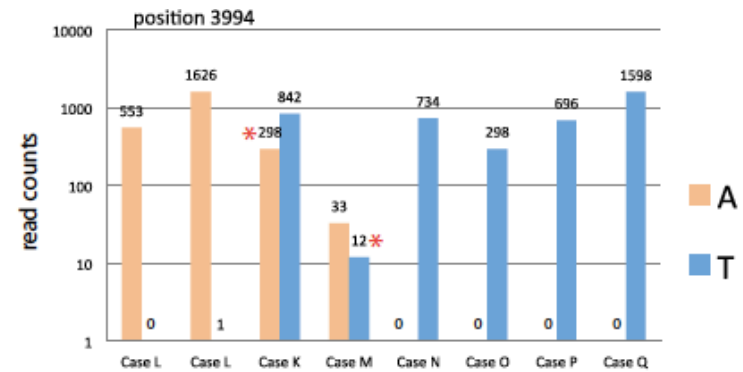
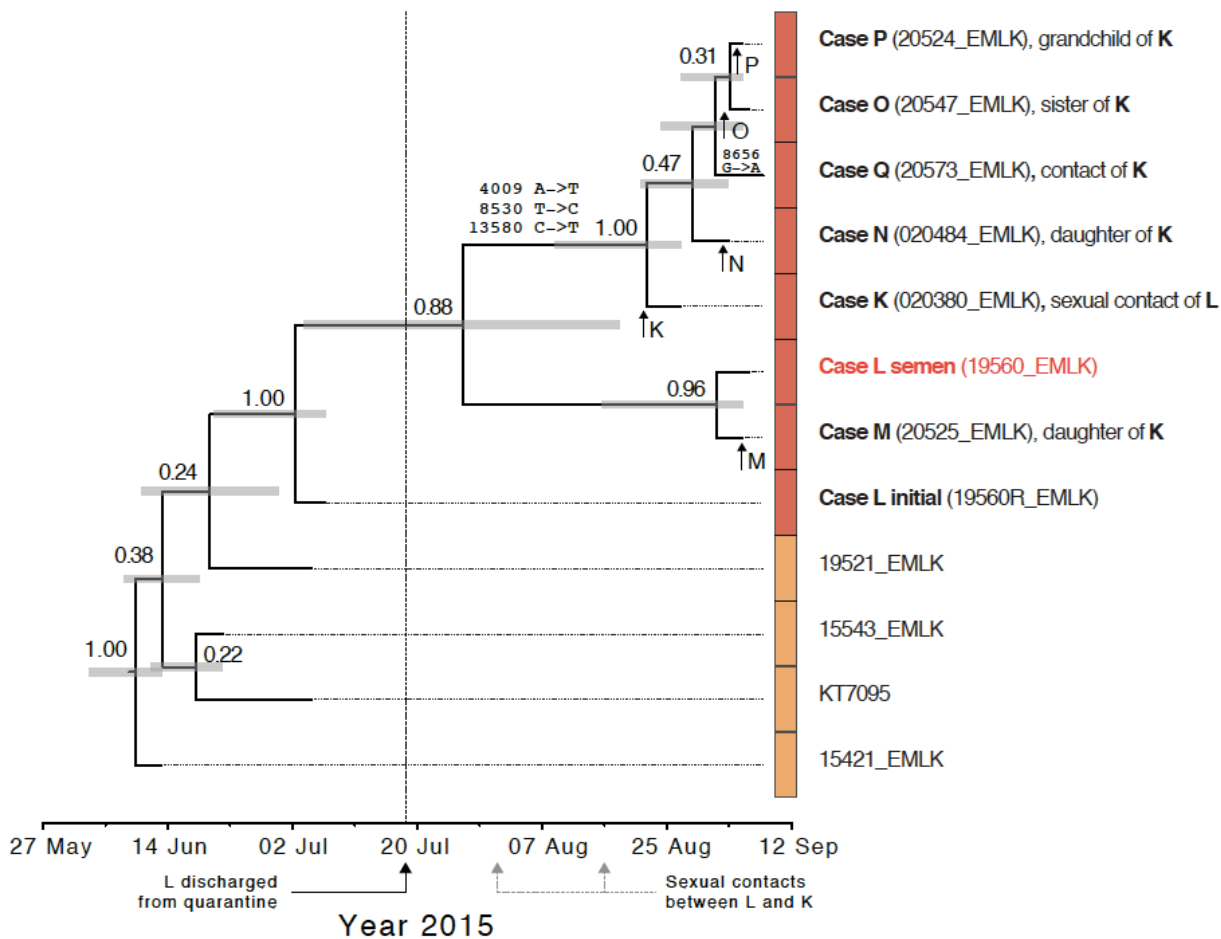
Ian Goodfellow, Armando Arias, Jia (Luca) Lu, Lucy Thorne, Matt Cotten, My Vu Tra Phan, Simon Watson, Andrew Rambaut, Dhamari Naidoo



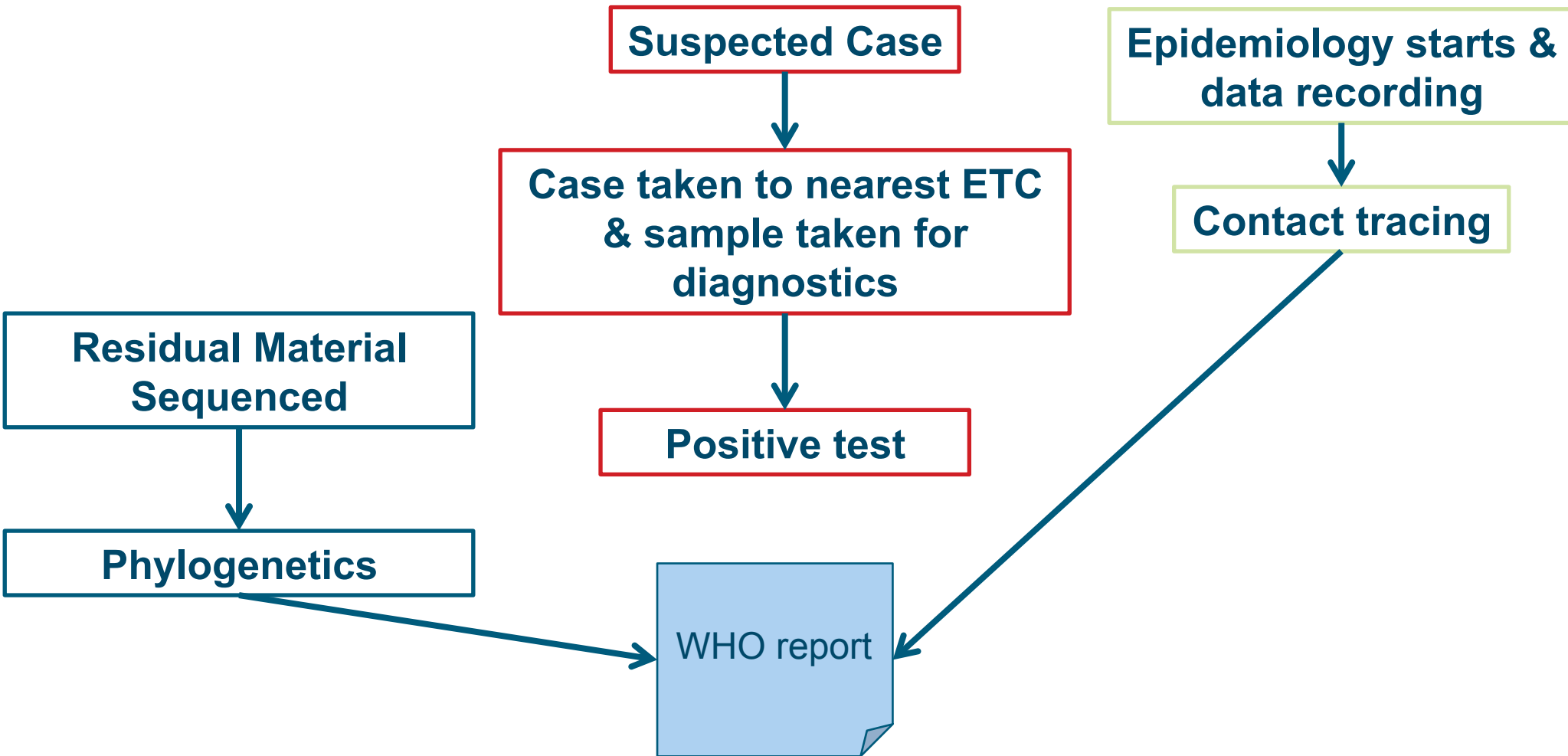
Phylogeography of Ebola virus 2014-5



Minority variants and transmission chains



Real time virus genomes



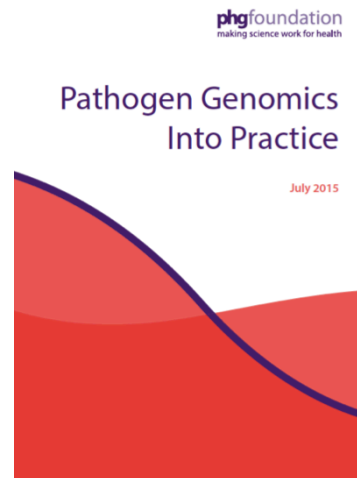
855 sequencing samples were processed between 16 April 2015 and 15 September 2015 yielding 614 EBOV genomes (72%).

Utility of virus genomes

Opportunities exist if virus genome sequences are routinely obtained from diagnostic samples during outbreaks or from routine diagnostic services.

In an integrated healthcare system pathogen genome sequence will allow evidence based infection control at different health care levels, will inform national epidemiology and will allow stratified patient management for treatments.

SHARED, ACTIONABLE & INTERPRETABLE INFORMATION



35 recommendations

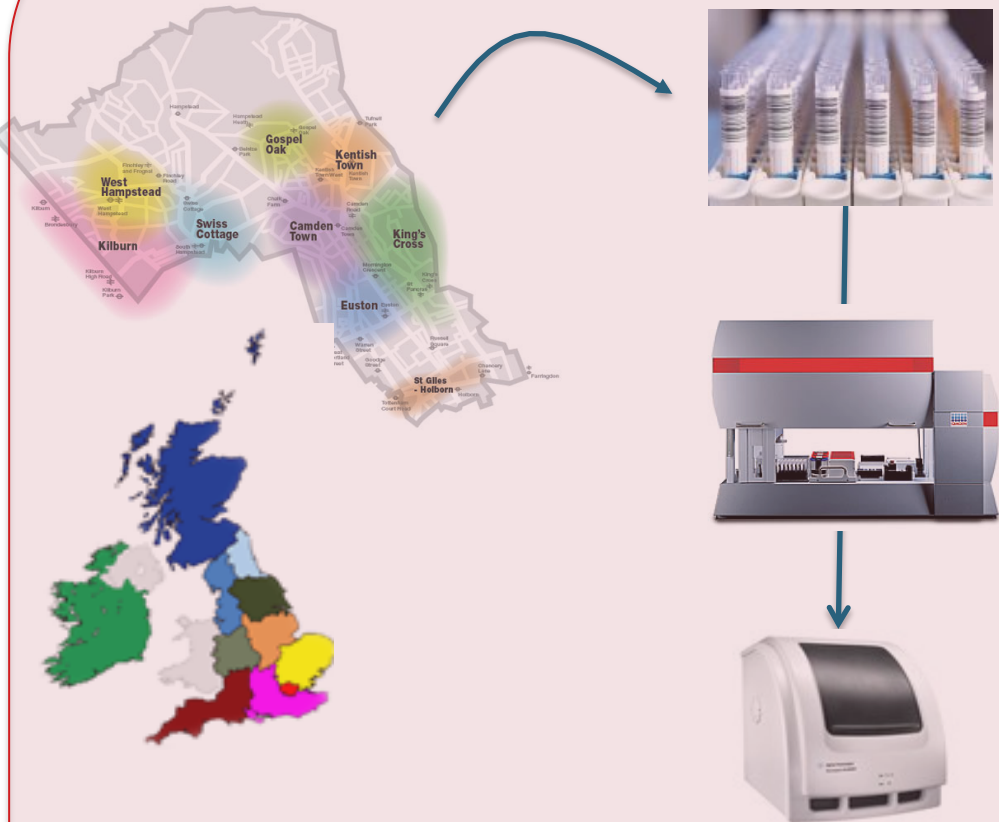


Infection response through virus genomics

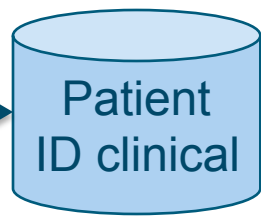
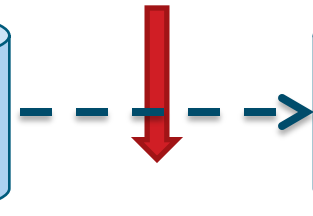
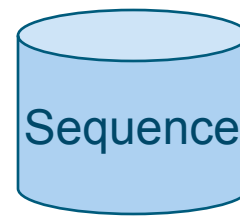
InfeCtion respONse through vIrus genomiCs



Samples/Requests



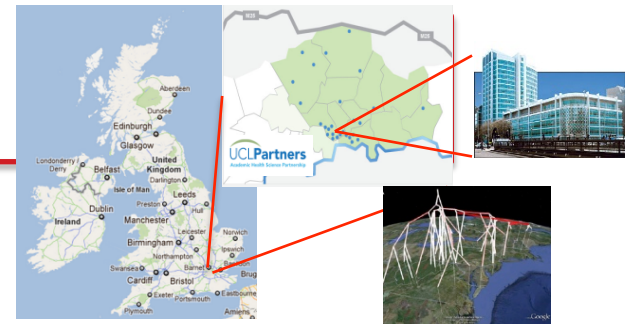
20 000 HIV,
HCV, IAV,
NV, MV



Clinical reporting

Diagnostics

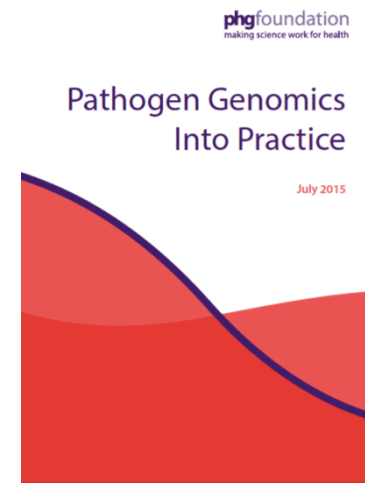
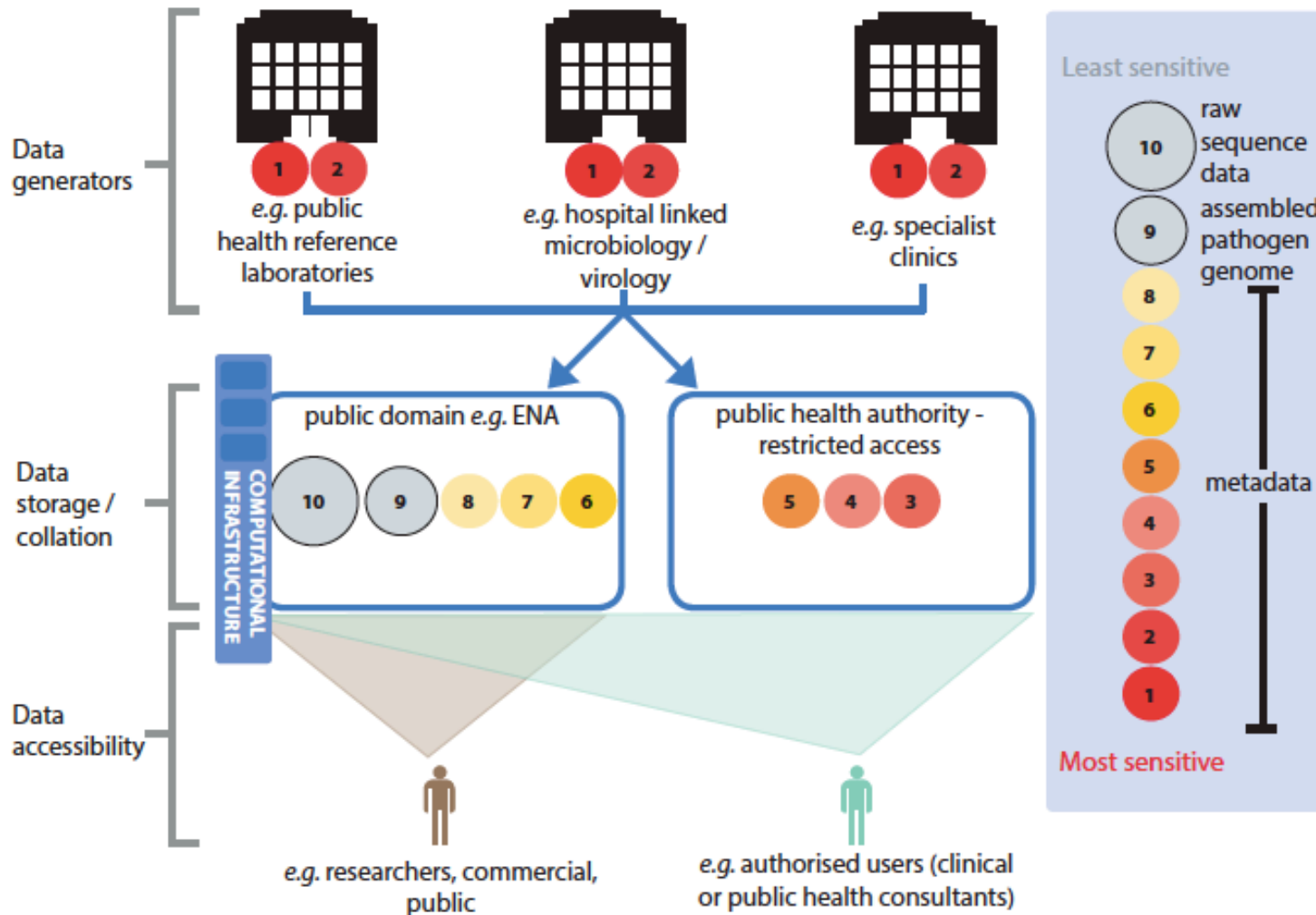
Clinical Virology Workflow



Interpretation

How to share data

Figure 15.3 A simplified vision of a two-tier data sharing strategy



The size of circles (not to scale) are indicative of the relative data storage burden (computational disc space), of the different subsets of data. Raw genomic data will consume the greatest disc space (therefore cost more to store than other data types), and so its longer term storage would be better suited in a consolidated repository build for high volume data storage.

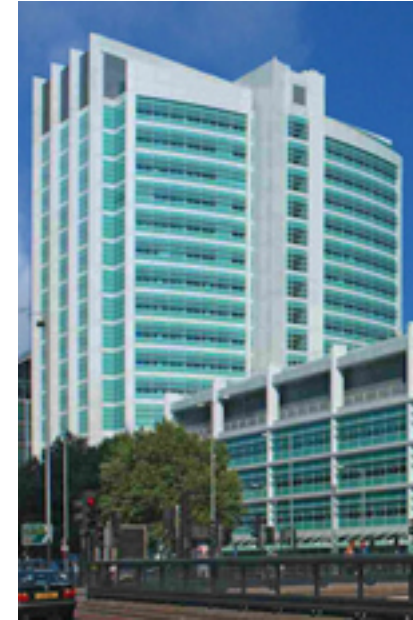
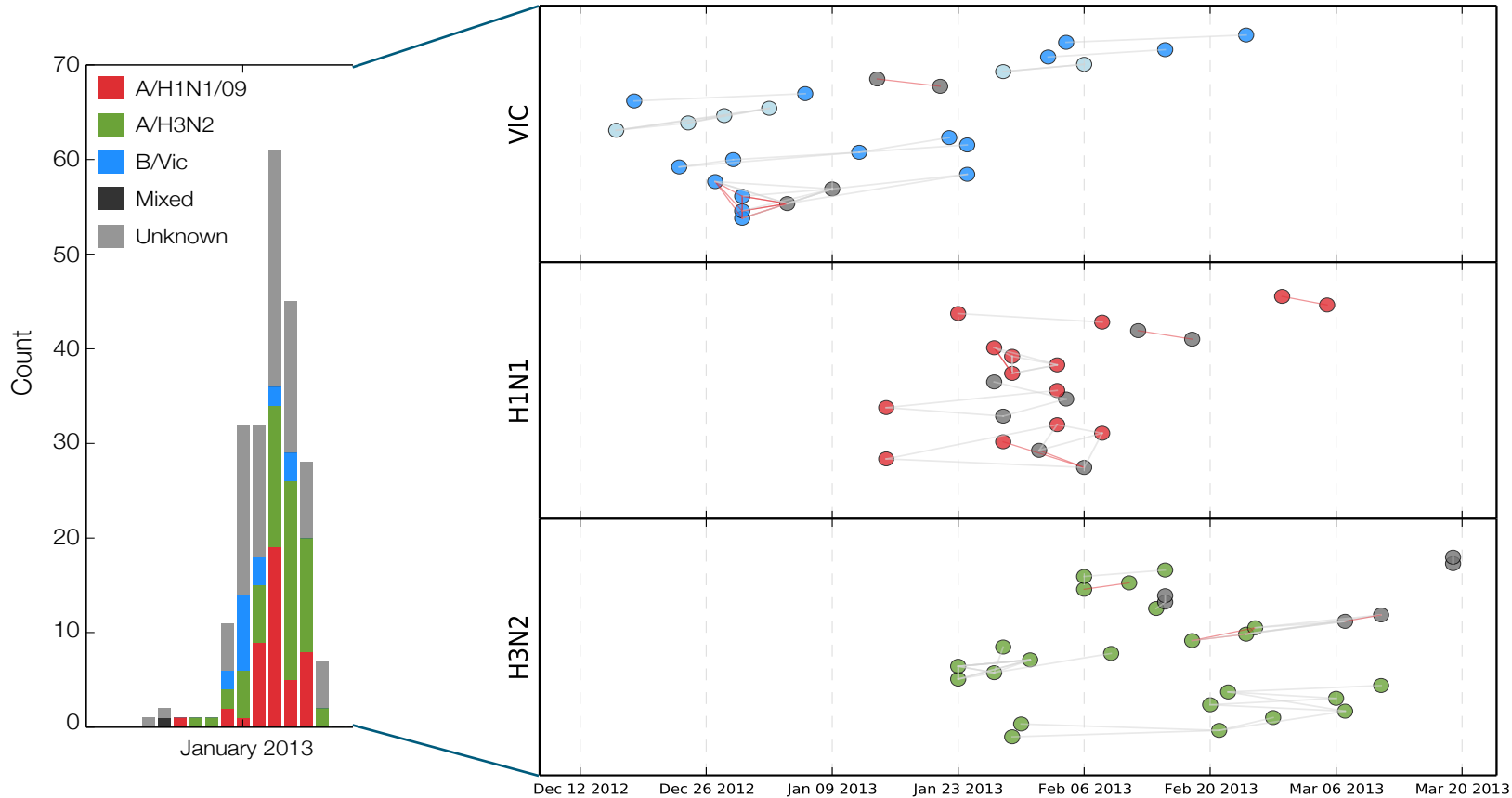
Mapping Genome sequencing to clinical data

Sample \longrightarrow Diagnostic identifier \longrightarrow Hospital/patient data

	Influenza	Norovirus	HIV	HCV
Total submitted lab numbers	407	20	360	152
Total records retrieved from Winpath.Results	406 (99.8%)	20 (100%)	360 (100%)	150 (98.7%)
Total identifiers found	406 (99.8%)	20 (100%)	240 (66.7%)	148 (97.4%)
Carecast data found	310 (76.2%)	14 (70%)	76 (21.1%)	30 (19.7%)
Obvious UCLH inpatient locations	357 (87.7)	14 (70%)	28 (7.8%)	2 (1.3%)
Patient stay records matches	240 (59%)	11 (55%)	20 (5.6%)	2 (1.3%)



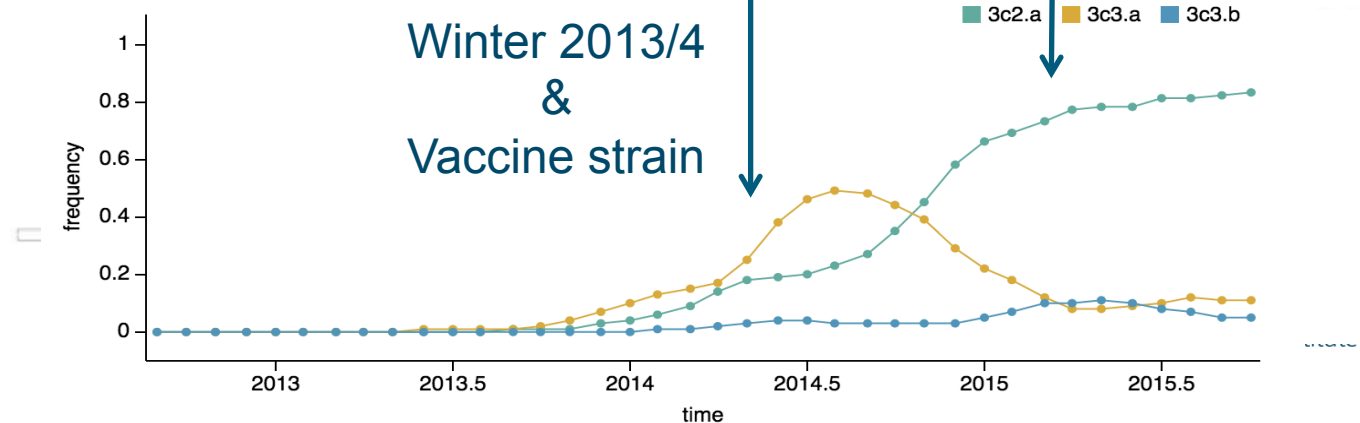
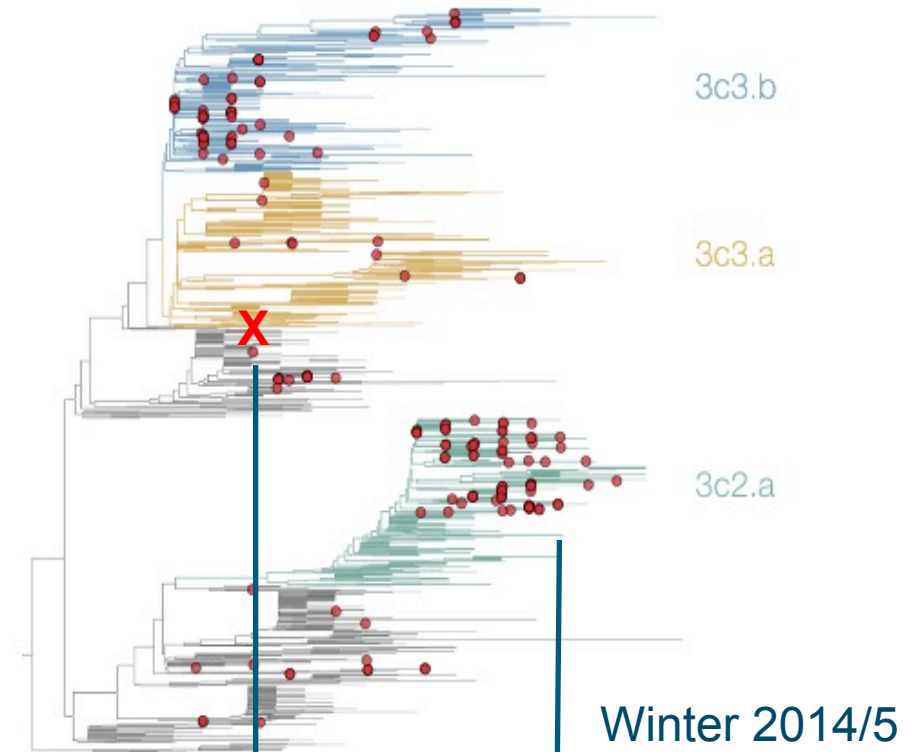
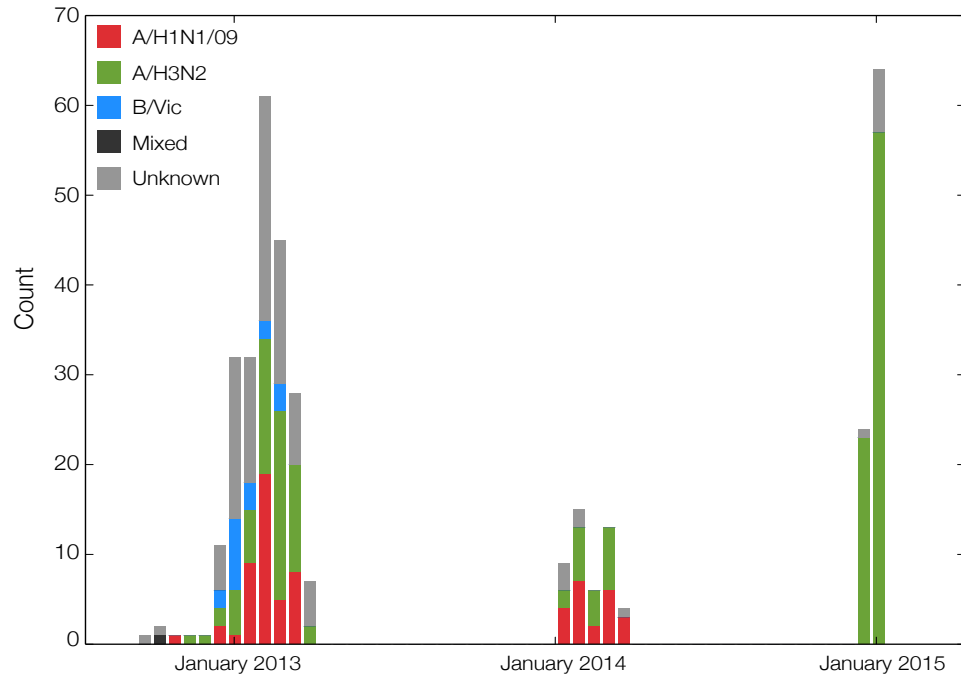
UK Influenza virus Hospital transmission chains



- **UCLH outbreaks driven by multiple introductions from community**
- **Limited patient-patient chains – ~1 in 7 cases initiate a chain of hospital transmission**

Early detection of vaccine mismatch

357 influenza genomes from UCLH & Barts Health over 3 consecutive influenza seasons



Conclusions

- **NGS for large scale and rapid virus genome sequencing is almost fit for purpose but need:**
 - Commercial sample to multiplex library
 - Accurate minority variant detection required
 - Stable computational pipelines
 - Linking to meta data
- **With appropriate sampling framework large scale sequencing can:**
 - Characterise a zoonotic reservoir
 - Identify zoonotic virus ‘chatter’
 - Inform outbreak control in the field and in hospitals

Acknowledgements

Sanger Institute

My VT Phan

Simon Watson

Matthew Cotton

University of Edinburgh

Andrew Rambaut

Gytis Dudas

UCL

Zisis Kozlakidis

Cambridge University

Ian Goodfellow

Armando Arias,

Jia (Luca) Lu,

Lucy Thorne

Dhamari Naidoo

WHO Emerging and Dangerous Pathogens Laboratory Network.

WHO Sierra Leone and Laboratory and Technical Working Group from MoHS

Centre for Disease Control (CDC), USA

Public Health England (PHE)

Kingdom of Saudi Arabia

Ministry of Health

Ziad Memish

Abdullah Al-Rabeeah, Hatem Makhdoom, Abdullah Assiri, Jaffar Al-Tawfiq,

Rafat Alhakeem, Hossam Madani, Fahad AlRabiah, Sami Al Hajjar, Ali

Albarrak, Hesham Flemban, Hanan Balkhy, Sarah Alsubaie



WP Leads

Paul Kellam
Andrew Hayward
Eleni Nastouli
Steven Morris

University College London

Laura Shallcross
Dan Frampton
Zisis Kozlakidis
Ellen Fragaszy
Tiziano Gallo Cassarino
Fatima Wurie
Anil Gunesh

BartsHealth

Duncan Clark
Jonathan Hubb
Graham Foster
Anthony Oliver

UCLH

Bridget Ferns
Elizabeth Gyimah
Jade Raffle
Pietro G. Coen

WT Sanger Institute

Simon Watson
Nick Grayson
Spela Binter
Stephanie Edwards
Swee Hoe Ong

PHE

Richard Myers
David James Allen
Richard Harris
David Brown
Tamyo Mbisa

MRC-CTU

David Dunn

Royal Free

Susan Hopkins
Daniel Webster

Tanzina Haque

William Rosenberg

University of Nottingham

William Irving
Jonathan Ball

University of Edinburgh

Andrew Leigh-Brown
Gonzalo Yerba

HCVRUk

John McLauchlan