Genomic analysis of **Pathogen X**

Andrew Rambaut | University of Edinburgh

THE UNIVERSITY
of EDINBURGH

What can pathogen genomics tell us?

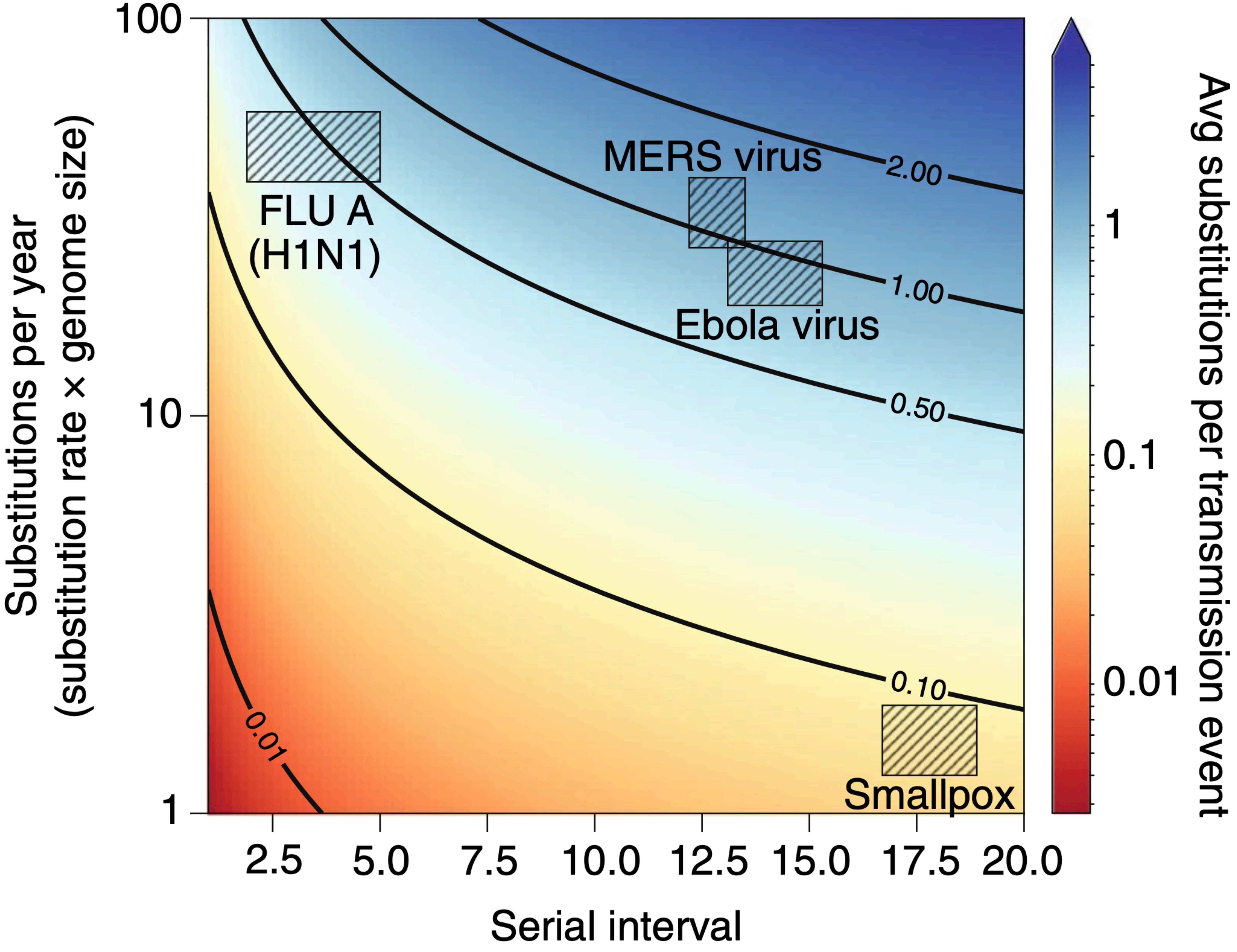
- Evolution is generally not predictable
- Evolution encodes a source of information about the history and epidemiology of the pathogen in its genome
- The greatest utility of pathogen genomics is situational awareness
    - An initial assessment of an outbreak when information is sparse
    - Help inform response to prevent outbreaks from becoming pandemics
- A pandemic is a collection of outbreaks and epidemics

What can pathogen genomics tell us about a new outbreak?

- What is the agent?
- What are related viruses - what can we assume based on these?
- How long has it been spreading?
- How fast is it spreading now?
- What is the source?

What can pathogen genomics tell us about a new outbreak?

- What is the agent?
- What can we assume based on related viruses?



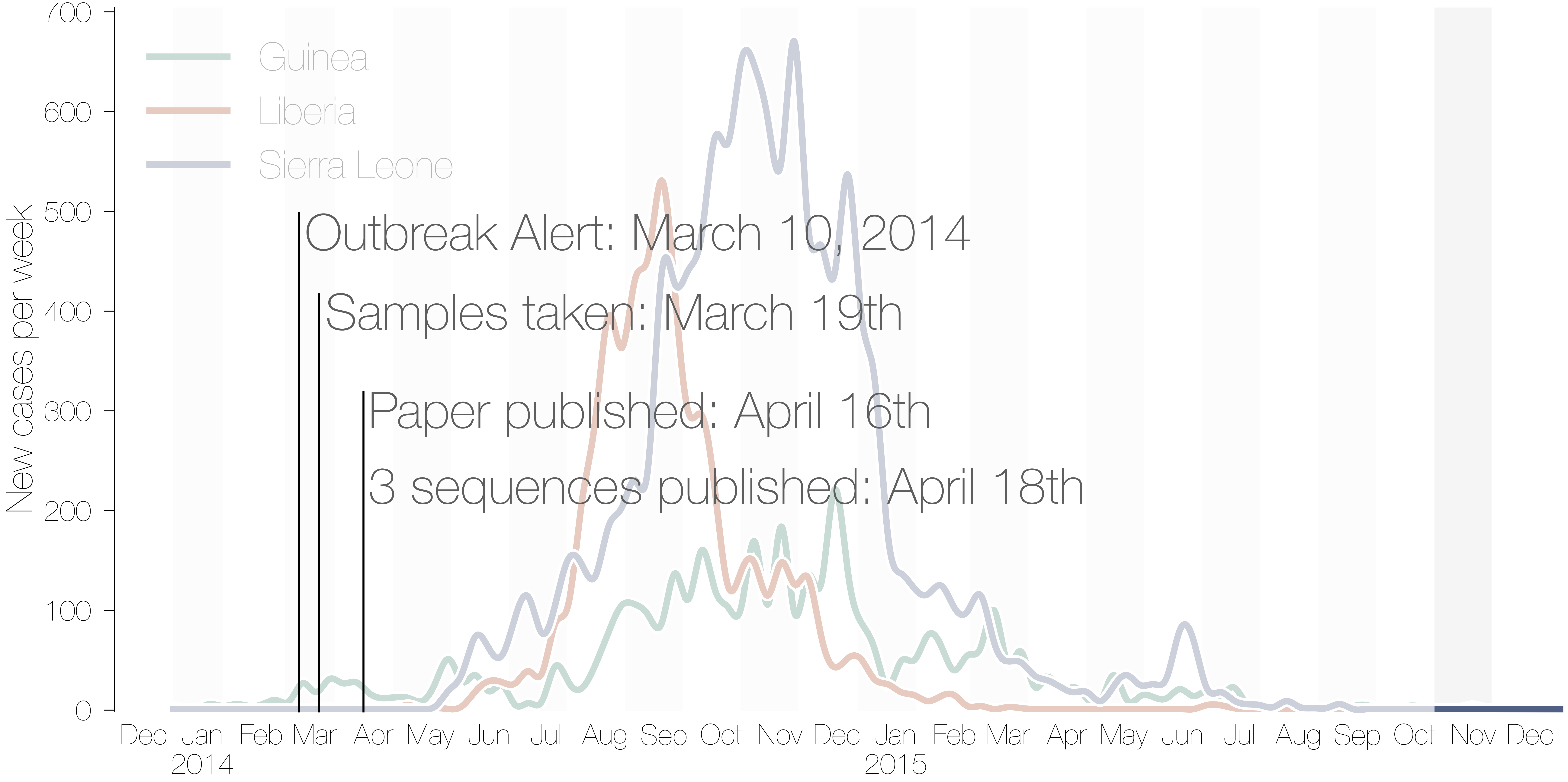
Grubaugh et al (2019) *Nat Micro* | https://doi.org/10.1038/s41564-018-0296-2

What can pathogen genomics tell us about a new outbreak?

- What is the agent?
- What can we assume based on related viruses?
- How long has it been spreading?
- How fast is it spreading now?
- What is the source?

Ebola virus (EBOV) epidemic 2014-2016 (2021?)

Ebolavirus epidemic 2014-2016: 28,646 recorded cases 11,323 recorded deaths



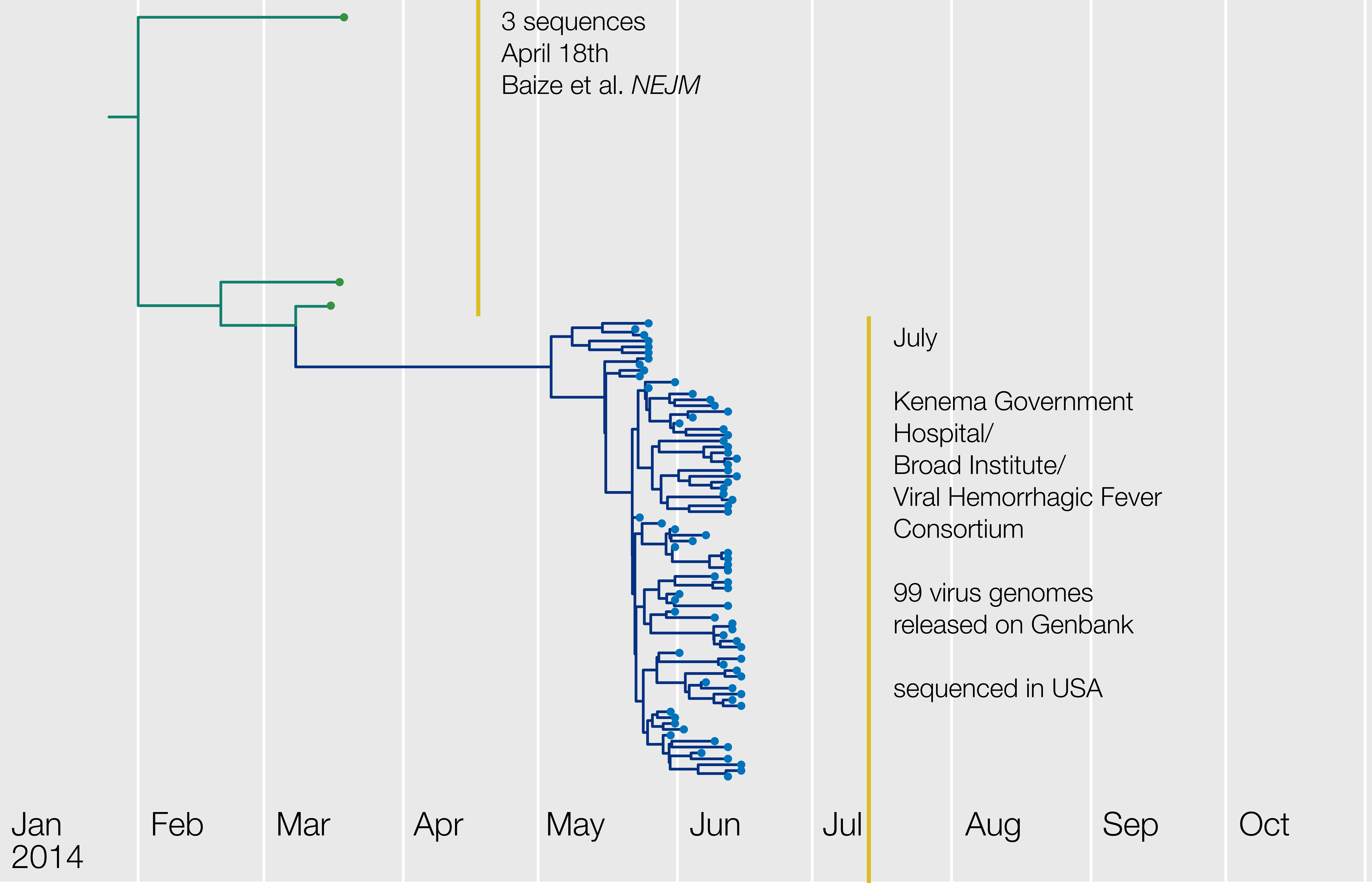
Outbreak Alert: March 10, 2014

Samples taken: March 19th

Paper published: April 16th

3 sequences published: April 18th

Baize et al (2014) *N Engl J Med*



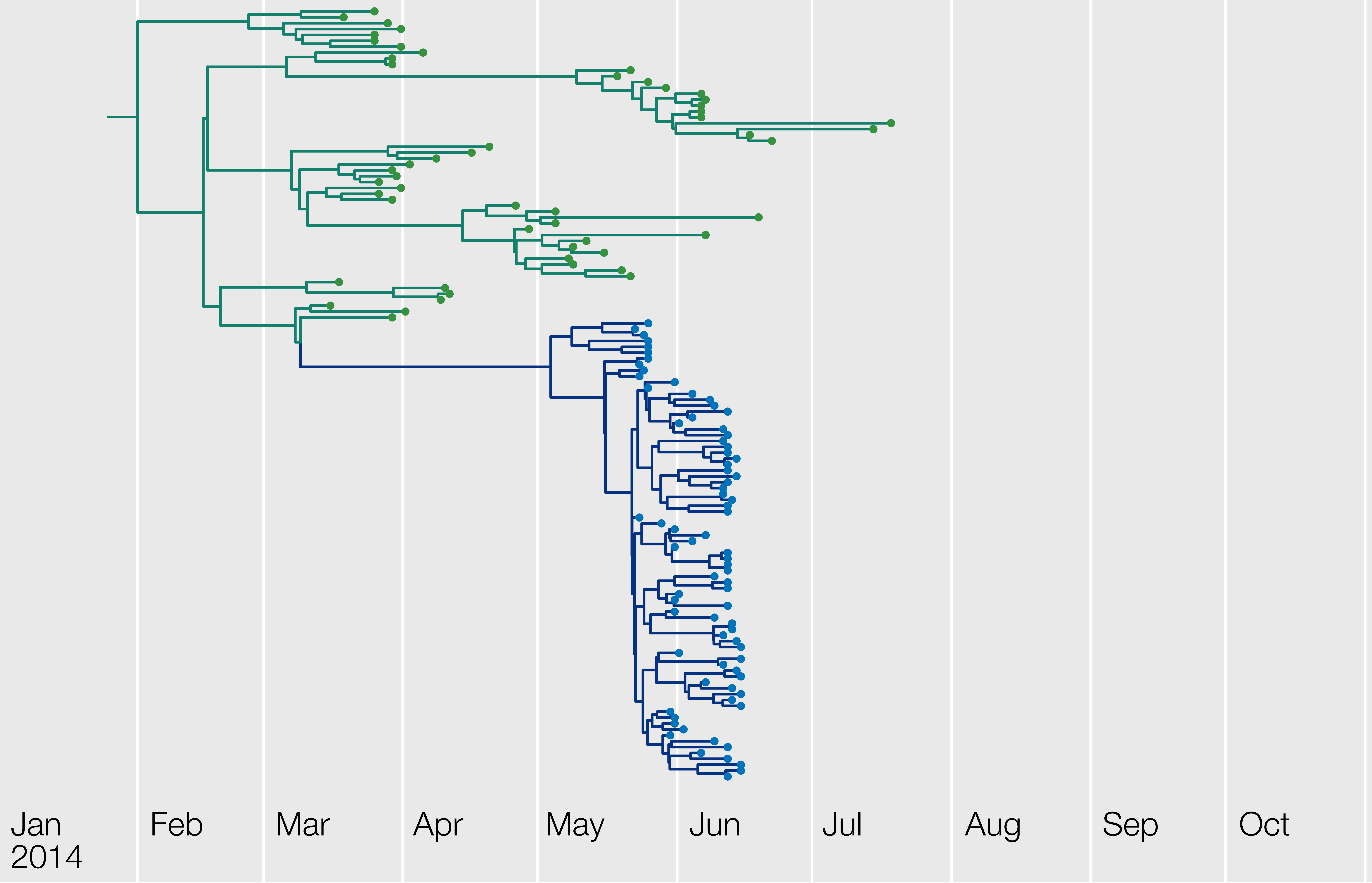
3 sequences
April 18th
Baize et al. *NEJM*

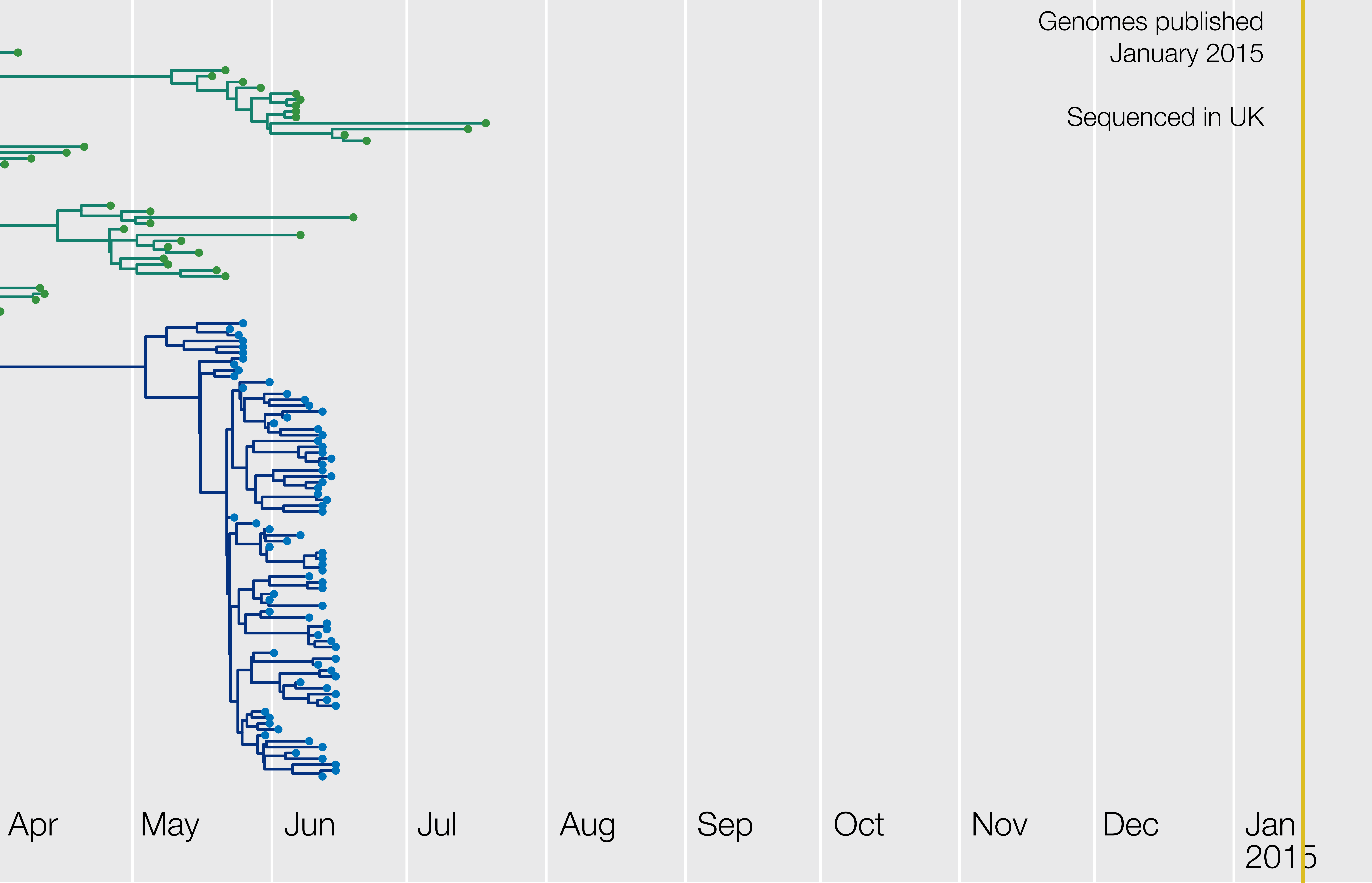
July

Kenema Government
Hospital/
Broad Institute/
Viral Hemorrhagic Fever
Consortium

99 virus genomes
released on Genbank

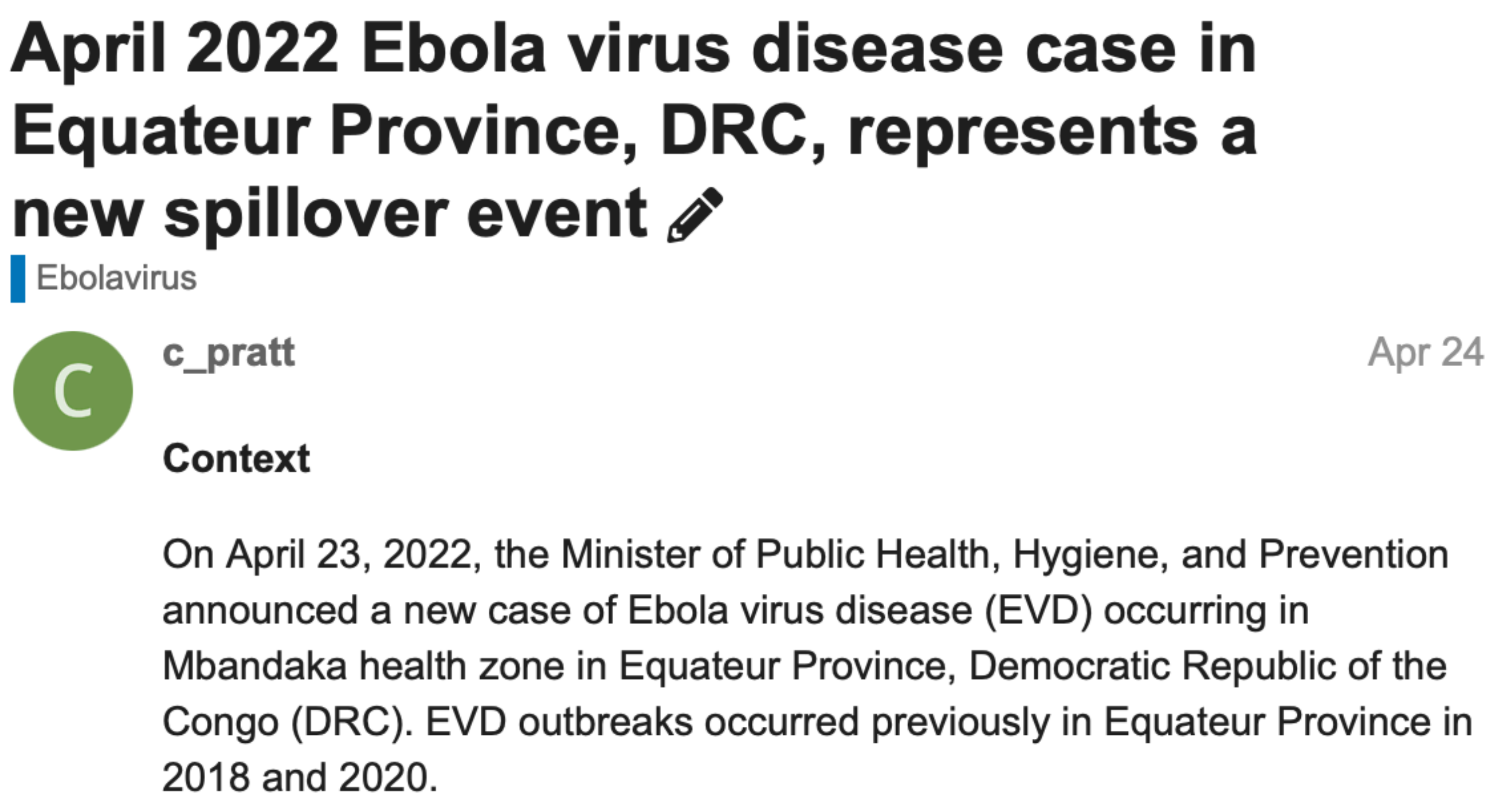
sequenced in USA


Jan
2014
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct


Genomes published
January 2015
Sequenced in UK
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2015

2022 Ebola virus outbreaks in DRC
Sequencing and analysis by INRB, Kinshasa

- Case sampled: 21st April 2022
- Analysis posted 24th April: https://virological.org/t/795
- Interpretation - a new spillover event from non-human animals



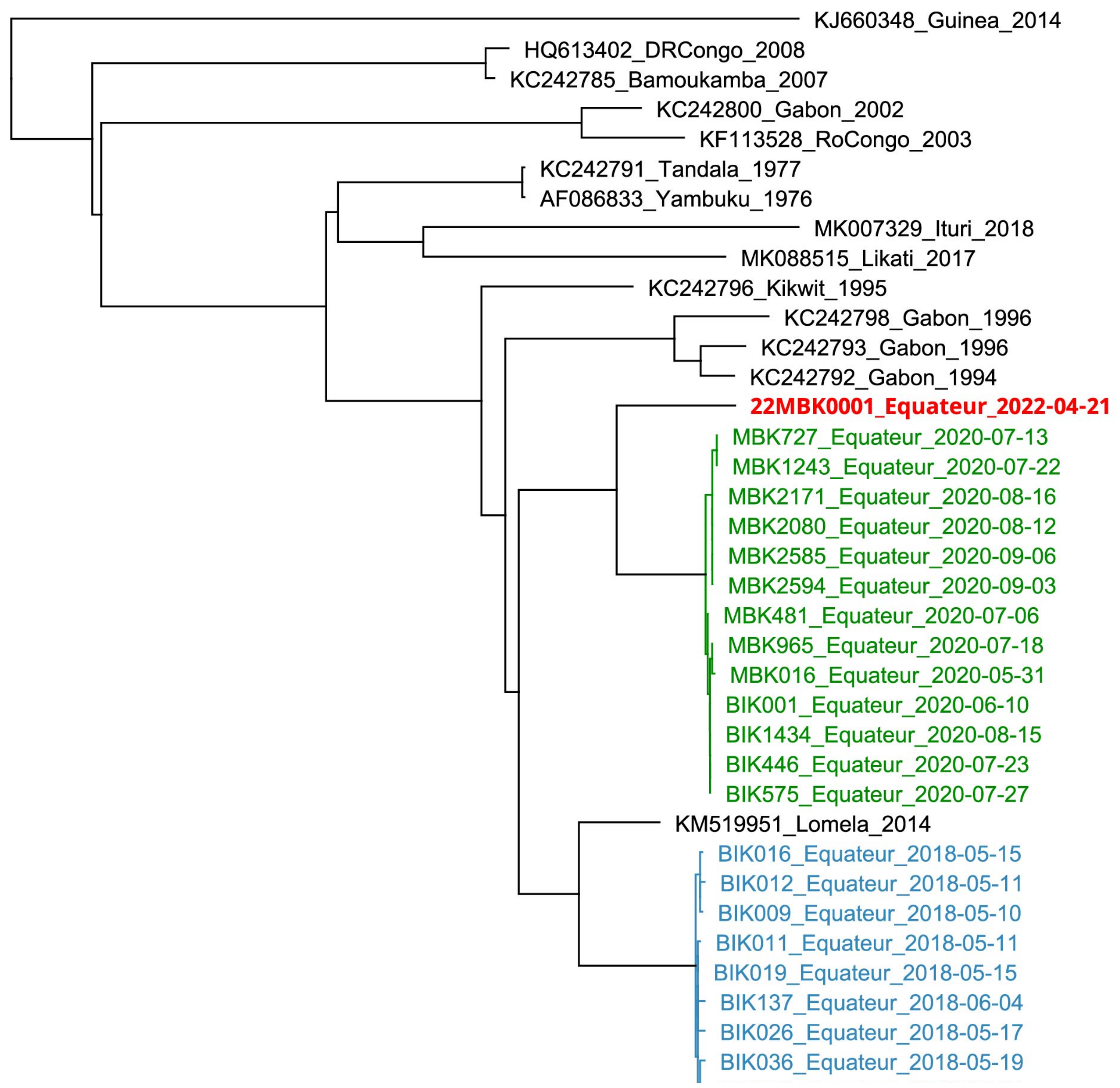
April 2022 Ebola virus disease case in Equateur Province, DRC, represents a new spillover event ✎

Ebolavirus

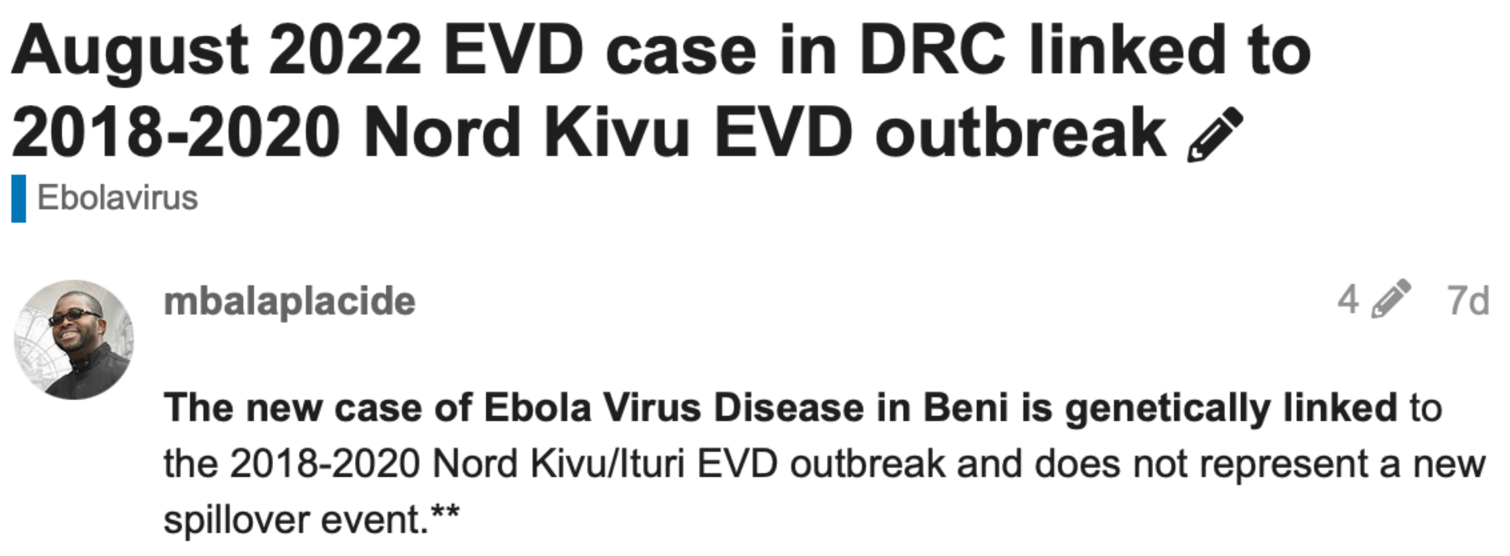
C **c_pratt** Apr 24

Context

On April 23, 2022, the Minister of Public Health, Hygiene, and Prevention announced a new case of Ebola virus disease (EVD) occurring in Mbandaka health zone in Equateur Province, Democratic Republic of the Congo (DRC). EVD outbreaks occurred previously in Equateur Province in 2018 and 2020.



- Case sampled: 15th August, 2022
- Analysis posted 22nd August: https://virological.org/t/795
- Interpretation:
  Genetically linked to the 2018-2020 Nord Kivu/Ituri EVD outbreak
  Does not represent a new spillover event

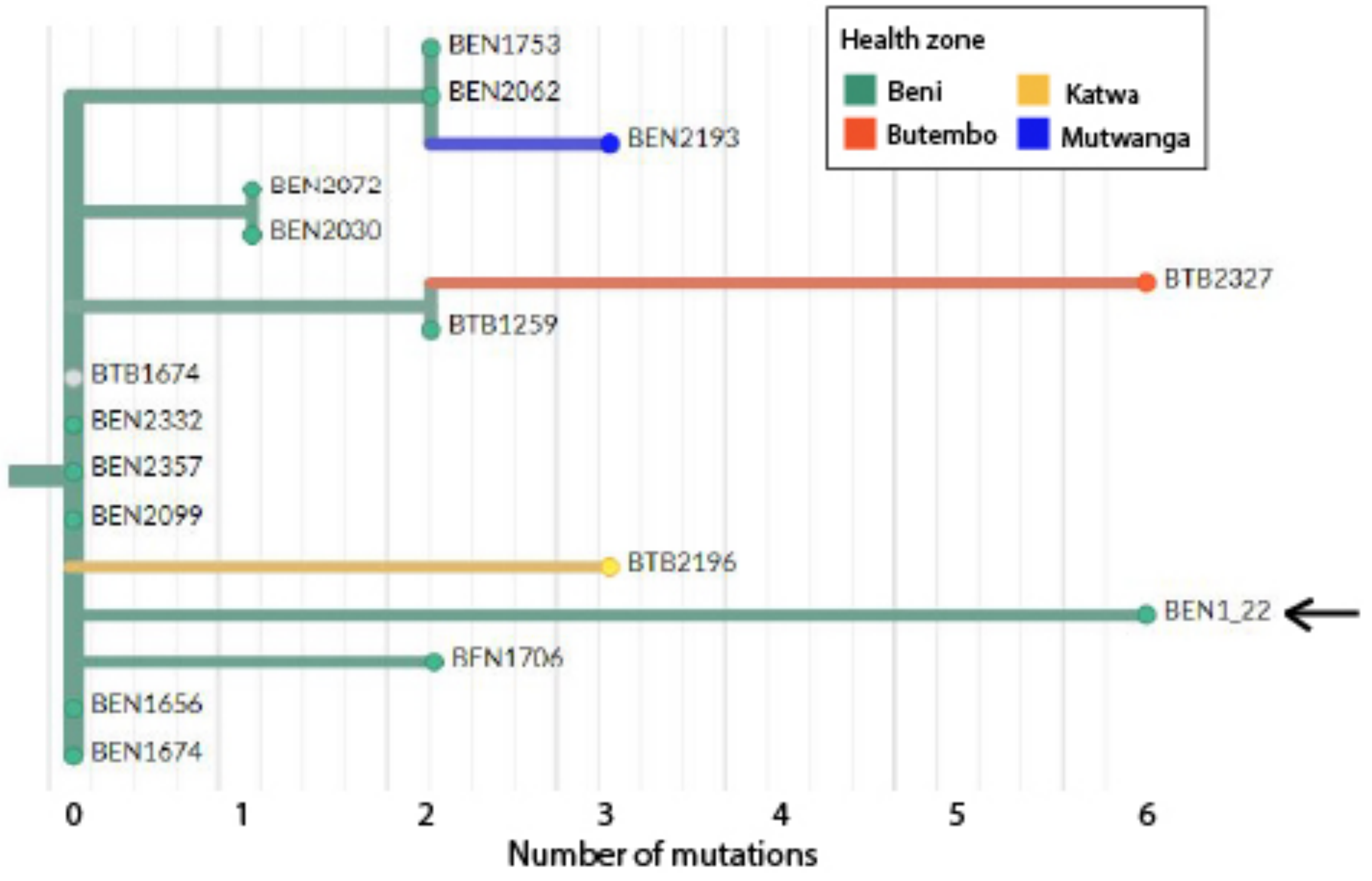


August 2022 EVD case in DRC linked to 2018-2020 Nord Kivu EVD outbreak ✎

Ebolavirus

mbalaplacide 4 ✎ 7d

The new case of Ebola Virus Disease in Beni is genetically linked to the 2018-2020 Nord Kivu/Ituri EVD outbreak and does not represent a new spillover event.**

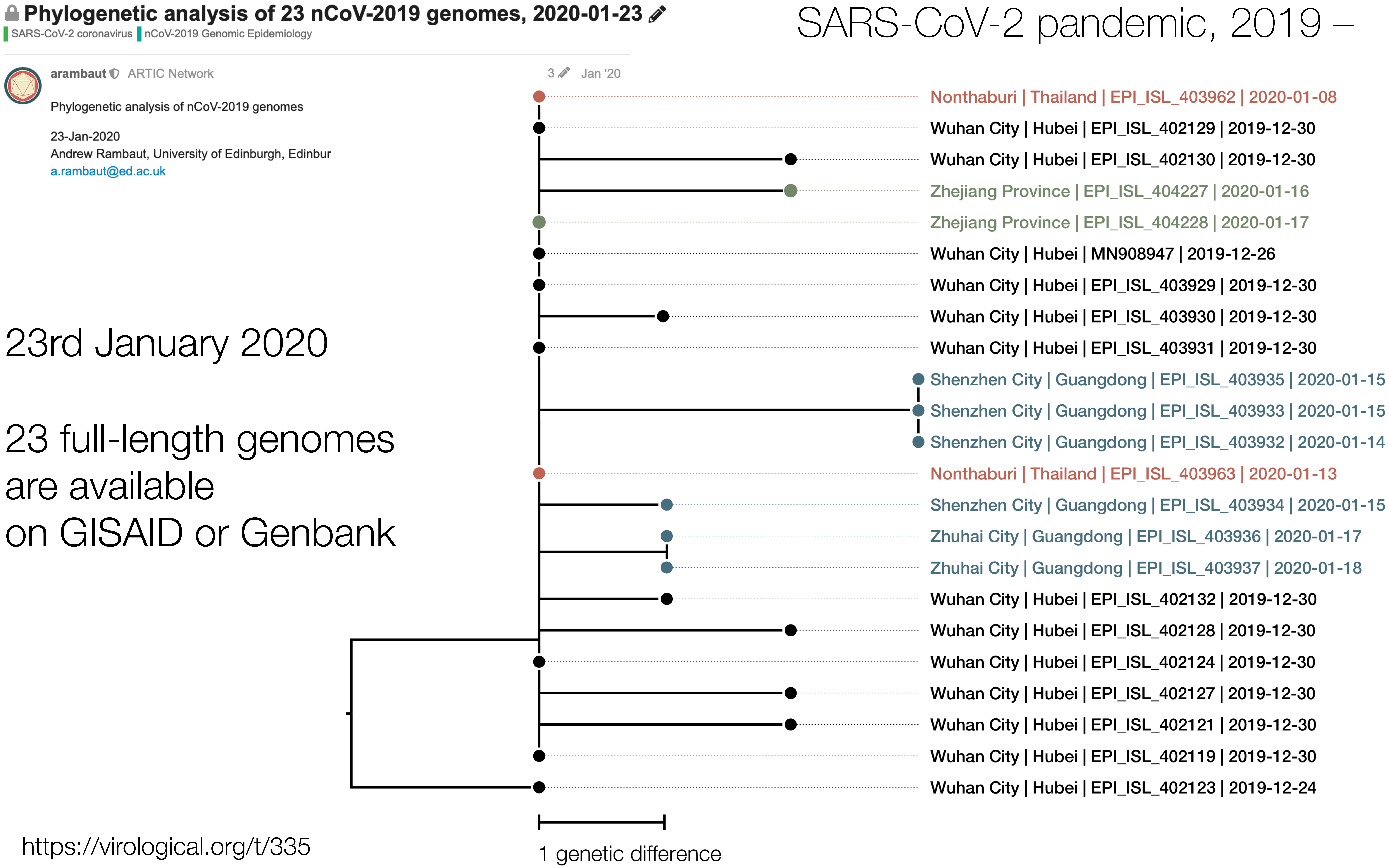


What can pathogen genomics tell us about a new outbreak?

- How long has it been spreading?
- How fast is it spreading now?

SARS-CoV-2 pandemic
2019 – ?

SARS-CoV-2 pandemic, 2019 –



Phylogenetic analysis of 23 nCoV-2019 genomes, 2020-01-23

SARS-CoV-2 coronavirus | nCoV-2019 Genomic Epidemiology

arambaut ARTIC Network 3 Jan '20

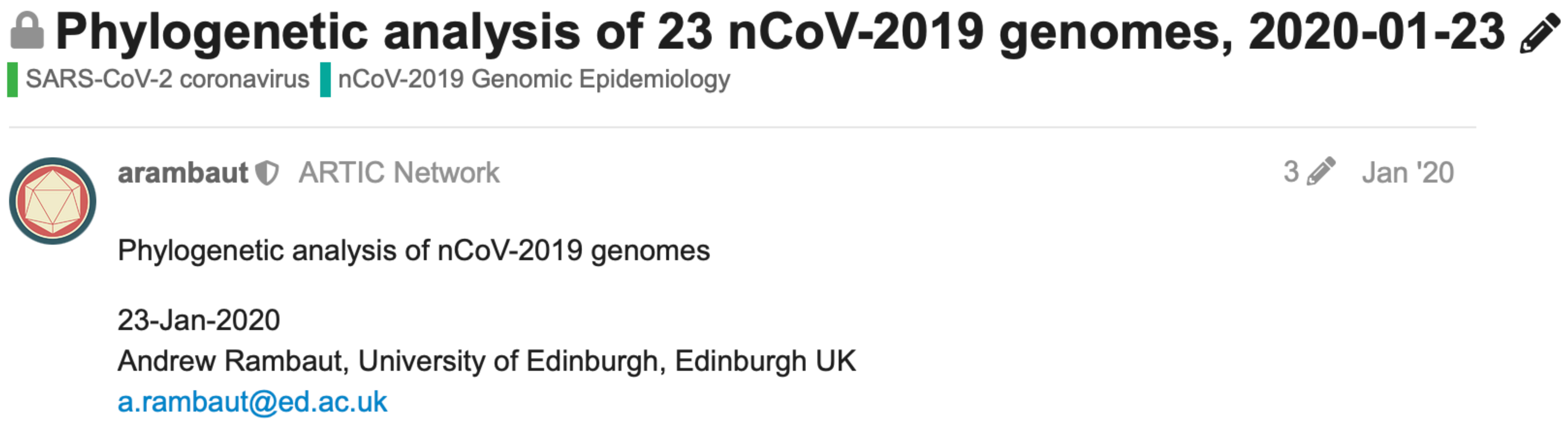
Phylogenetic analysis of nCoV-2019 genomes

23-Jan-2020
Andrew Rambaut, University of Edinburgh, Edinbur
a.rambaut@ed.ac.uk

23rd January 2020

23 full-length genomes
are available
on GISAID or Genbank

https://virological.org/t/335



Phylogenetic analysis of 23 nCoV-2019 genomes, 2020-01-23

SARS-CoV-2 coronavirus nCoV-2019 Genomic Epidemiology

arambaut ARTIC Network 3 Jan '20

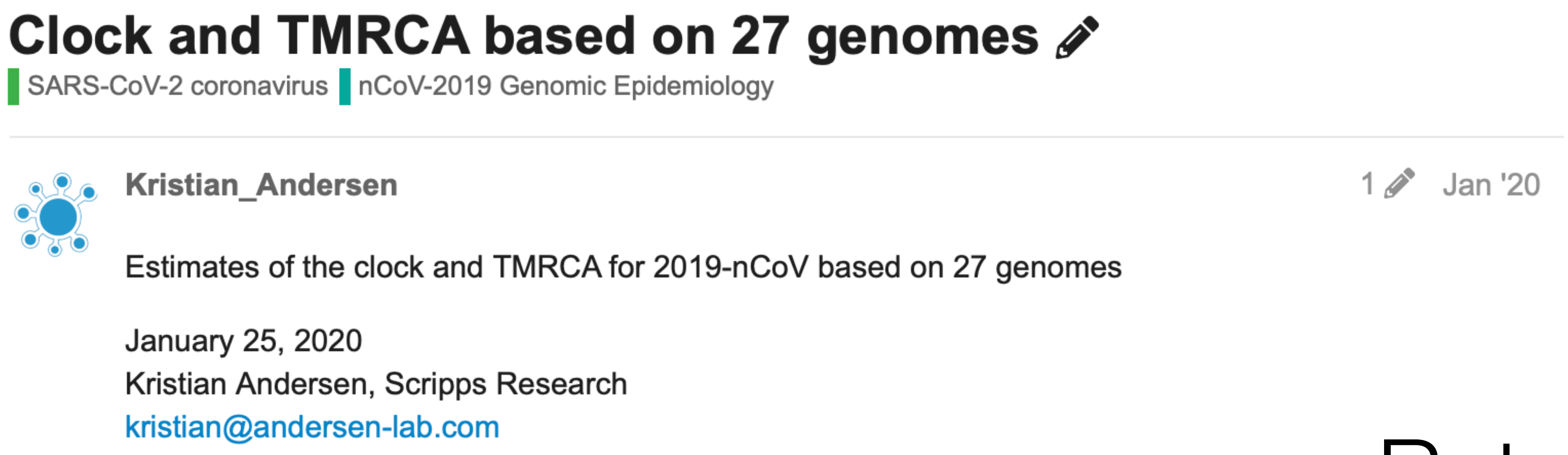
Phylogenetic analysis of nCoV-2019 genomes

23-Jan-2020
Andrew Rambaut, University of Edinburgh, Edinburgh UK
a.rambaut@ed.ac.uk

Assumed rate subst/site/year	Estimated date of MRCA	95% interval
1x10^{-3}	29-Nov-2019	08-Nov-2019 – 16-Dec-2019
0.5x10^{-3}	30-Oct-2019	18-Sep-2019 – 04-Dec-2019

Virus	Estimated rate x10-3 subst/site/year	Reference
SARS-CoV	0.80 – 2.38	Zhao et al. 2004 [2]
MERS-CoV	0.63 [0.14 – 1·1]	Cotten et al. 2013 [3]
	1.12 [0.88 – 1.37]	Cotten et al. 2014 [4]
	0.96 [0.83 – 1.09]	Dudas et al. 2018 [5]
HCoV-OC43	0.43 [0.27 – 0.60]	Vijgen et al. 2005 [6]

https://virological.org/t/335



Clock and TMRCA based on 27 genomes

Kristian_Andersen

Estimates of the clock and TMRCA for 2019-nCoV based on 27 genomes

January 25, 2020
Kristian Andersen, Scripps Research
kristian@andersen-lab.com

25th January 2020

27 full-length genomes
are available
on GISAID or Genbank

Rate of evolution

Median	95% HPD
1.1×10^{-3}	4.0×10^{-6} - 5.5×10^{-3}

Date of most recent common ancestor

Median	95% HPD
02 Dec 2019	01 Oct 2019 - 22 Dec 2019

https://virological.org/t/335

Phylodynamic analysis of SARS-CoV-2 | Update 2020-03-06

verity.hill ARTIC Network

Phylogenetic analysis of SARS-CoV-2 genomes

06-Mar-2020

Rate of evolution

Data	Estimated rate	95% interval
12-Feb, 75 genomes	$0.92x10^{-3}$	$0.33x10^{-3}$ – $1.46x10^{-3}$
24-Feb, 86 genomes	$0.80x10^{-3}$	$0.14x10^{-3}$ – $1.31x10^{-3}$
04-March, 116 genomes	$1.04x10^{-3}$	$0.71x10^{-3}$ – $1.40x10^{-3}$

Date of most recent common ancestor

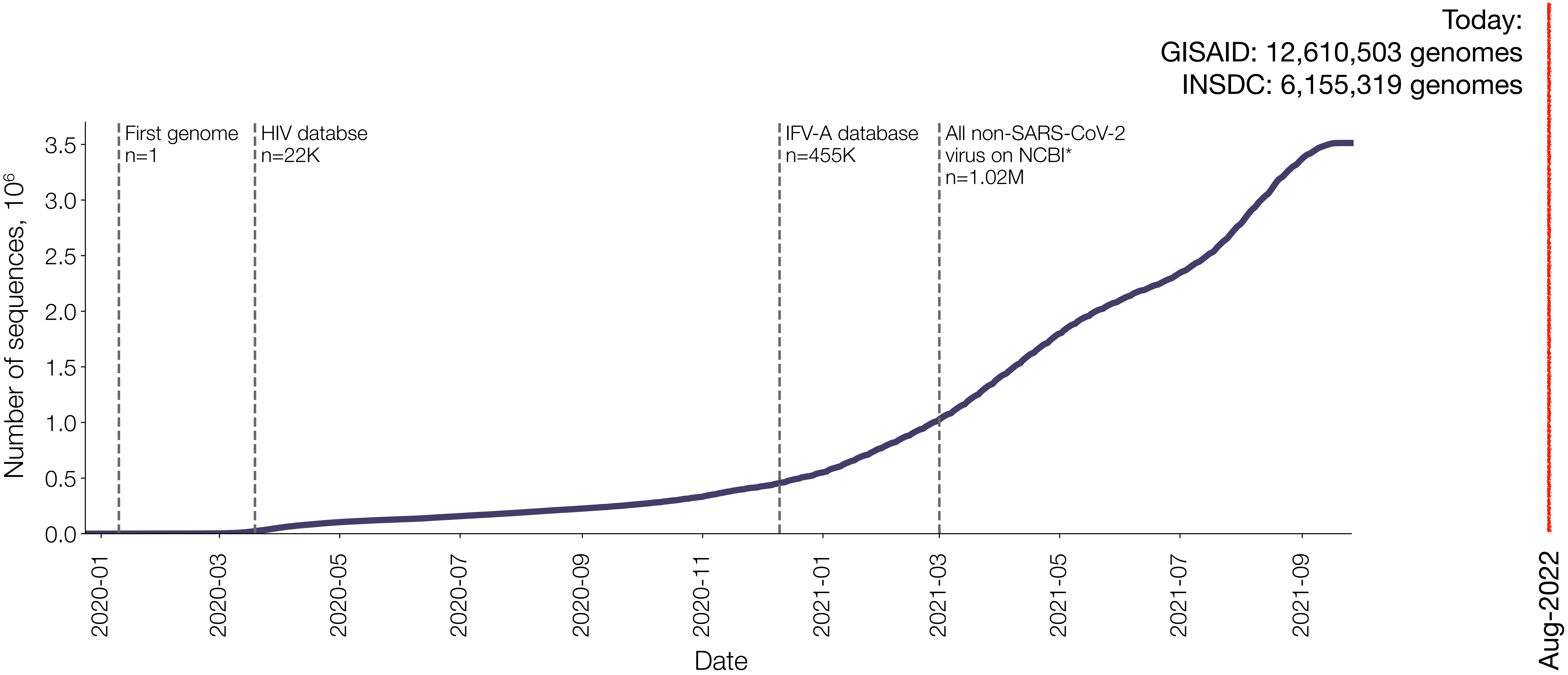
Data	Estimated MRCA	95% interval
12-Feb, 75 genomes	29-Nov-2019	28-Oct-2019 – 20-Dec-2019
24-Feb, 86 genomes	17-Nov-2019	27-Aug-2019 – 19-Dec-2019
04-March, 116 genomes	03-Dec-2019	16-Nov-2019 – 17-Dec-2019

Epidemic growth rate

Data	Doubling time (days)	95% interval
12-Feb, 75 genomes	6.2	4.1 – 12.3
24-Feb, 86 genomes	7.2	4.7 – 16.3
04-March, 116 genomes	6.8	5.0 – 10.3

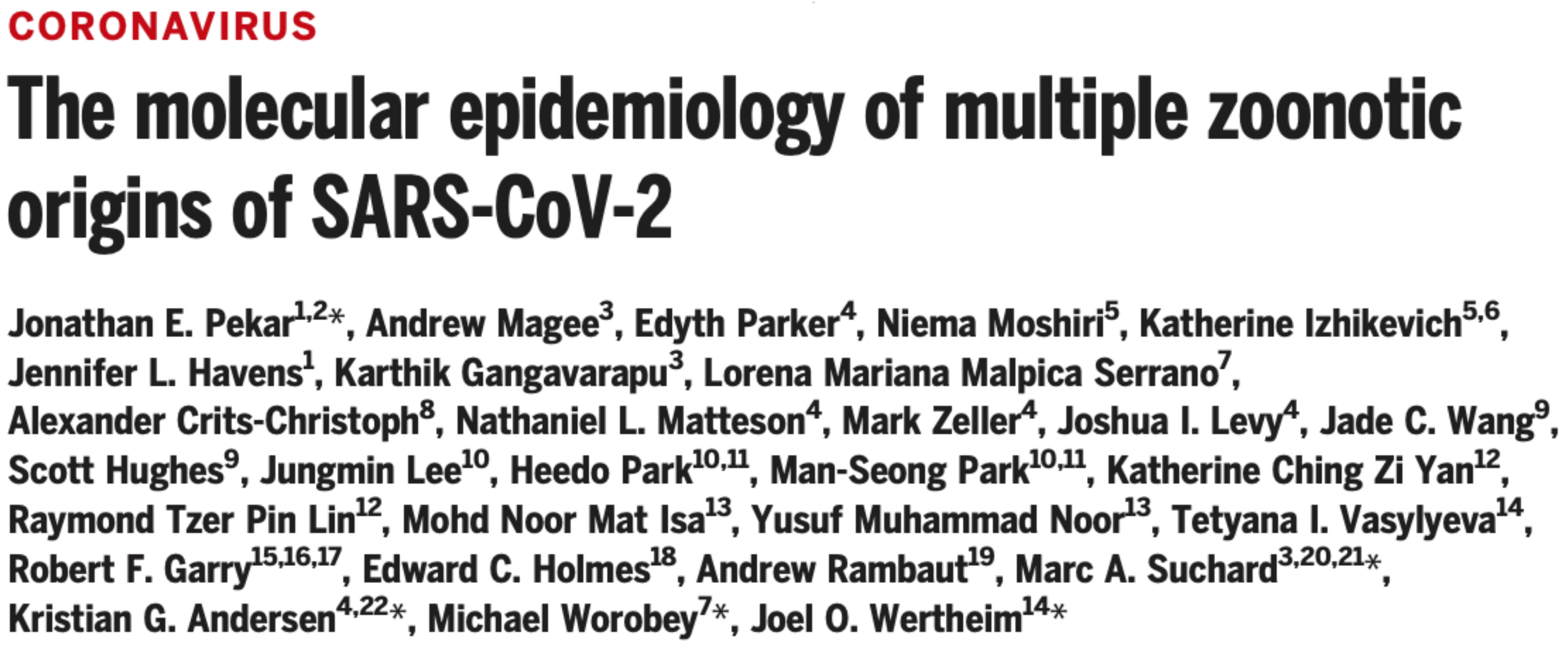
https://virological.org/t/420

Today:
GISAID: 12,610,503 genomes
INSDC: 6,155,319 genomes



*non-partial virus records >1K bp

figure by Áine O'Toole

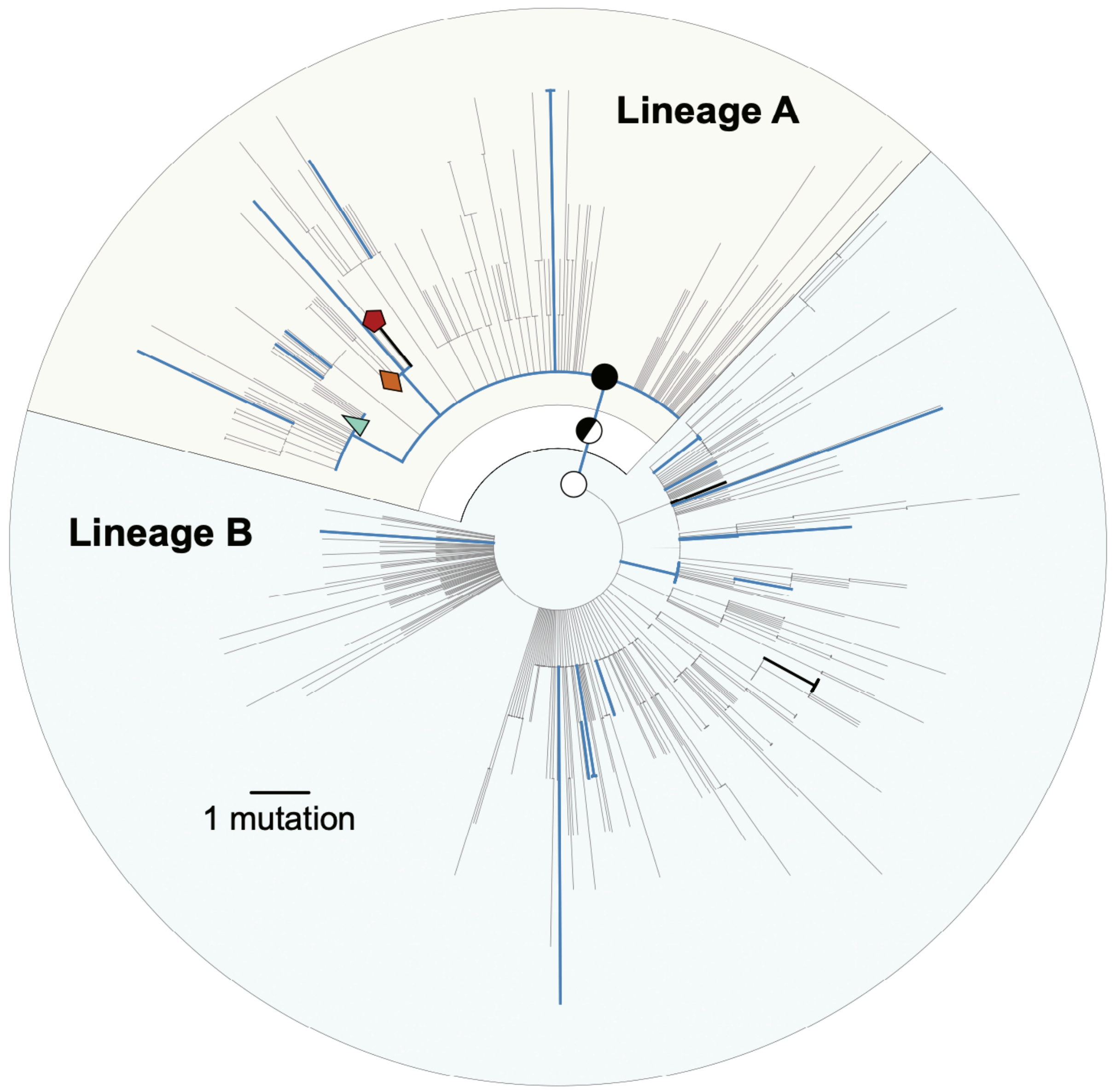


CORONAVIRUS

The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2

Jonathan E. Pekar[1,2]*, Andrew Magee[3], Edyth Parker[4], Niema Moshiri[5], Katherine Izhikevich[5,6], Jennifer L. Havens[1], Karthik Gangavarapu[3], Lorena Mariana Malpica Serrano[7], Alexander Crits-Christoph[8], Nathaniel L. Matteson[4], Mark Zeller[4], Joshua I. Levy[4], Jade C. Wang[9], Scott Hughes[9], Jungmin Lee[10], Heedo Park[10,11], Man-Seong Park[10,11], Katherine Ching Zi Yan[12], Raymond Tzer Pin Lin[12], Mohd Noor Mat Isa[13], Yusuf Muhammad Noor[13], Tetyana I. Vasylyeva[14], Robert F. Garry[15,16,17], Edward C. Holmes[18], Andrew Rambaut[19], Marc A. Suchard[3,20,21]*, Kristian G. Andersen[4,22]*, Michael Worobey[7]*, Joel O. Wertheim[14]*

747 genomes
Sampled from December 2019
to February 2020



Date of most recent common ancestor

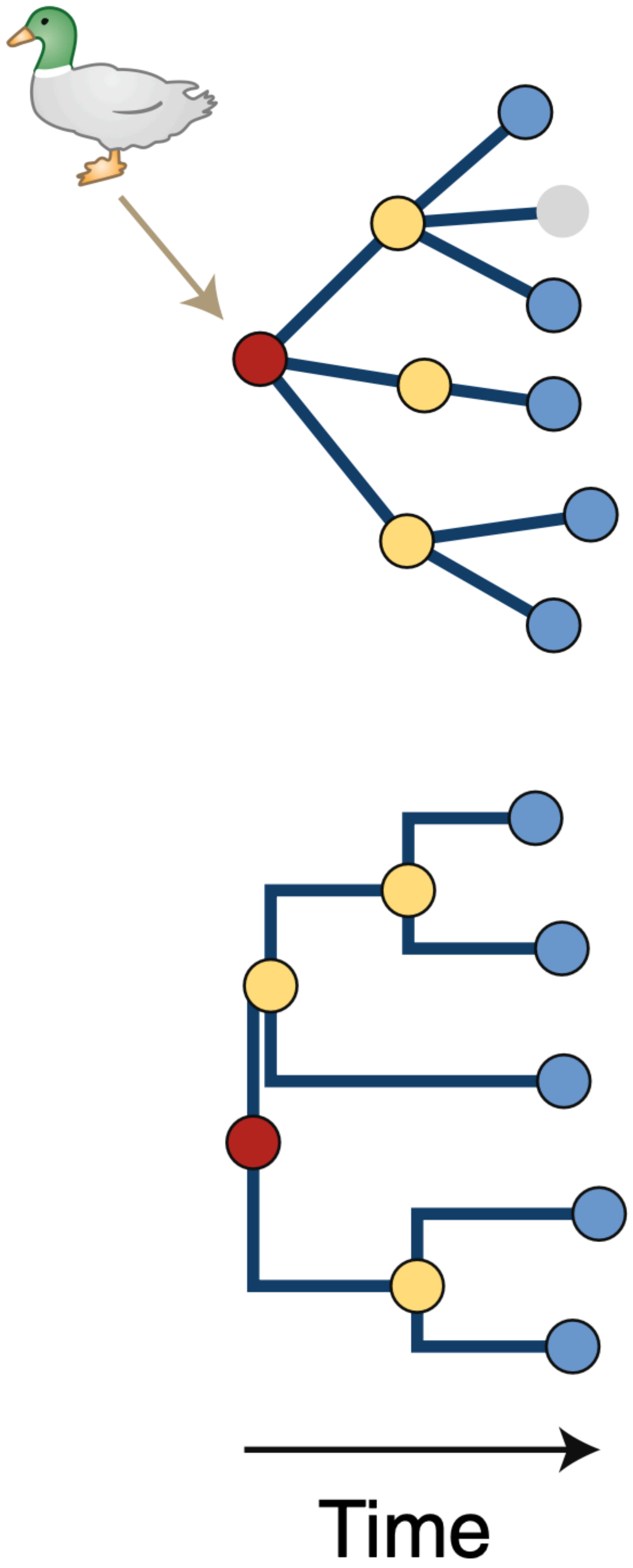
11 December 2019
95% interval: 25 November to 12 December

Pekar et al., Science 377, 960–966 (2022)

What can pathogen genomics tell us about a new outbreak?

- What is the agent?
- What can we assume based on related viruses?
- How long has it been spreading?
- How fast is it spreading now?
- What is the source?

outbreak or zoonosis?



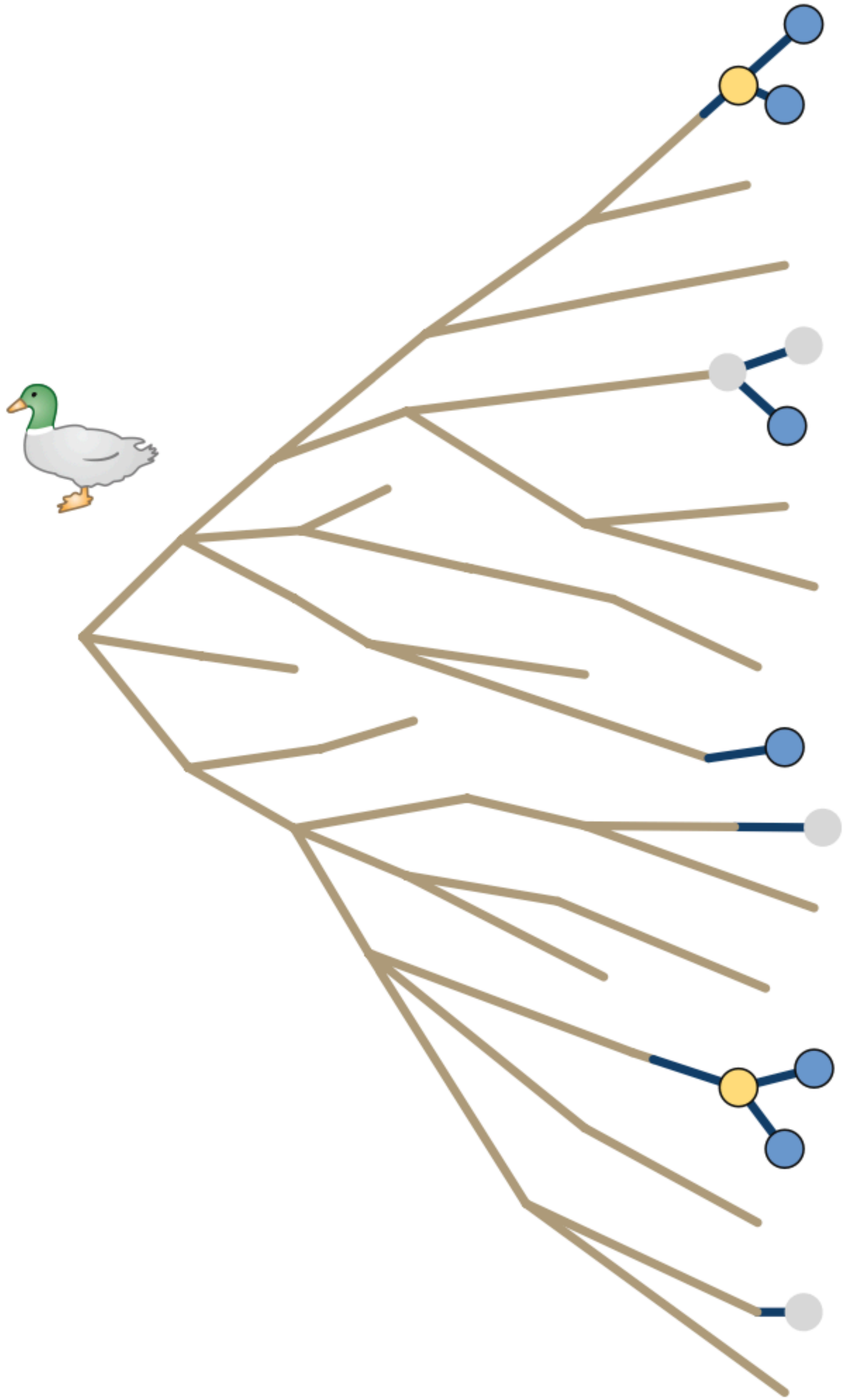
- 🔵 sampled case
- 🟡 inferred case
- 🔴 primary case
- ⚪ unobserved case

- — human to human transmission
- — reservoir transmission

outbreak or zoonosis?

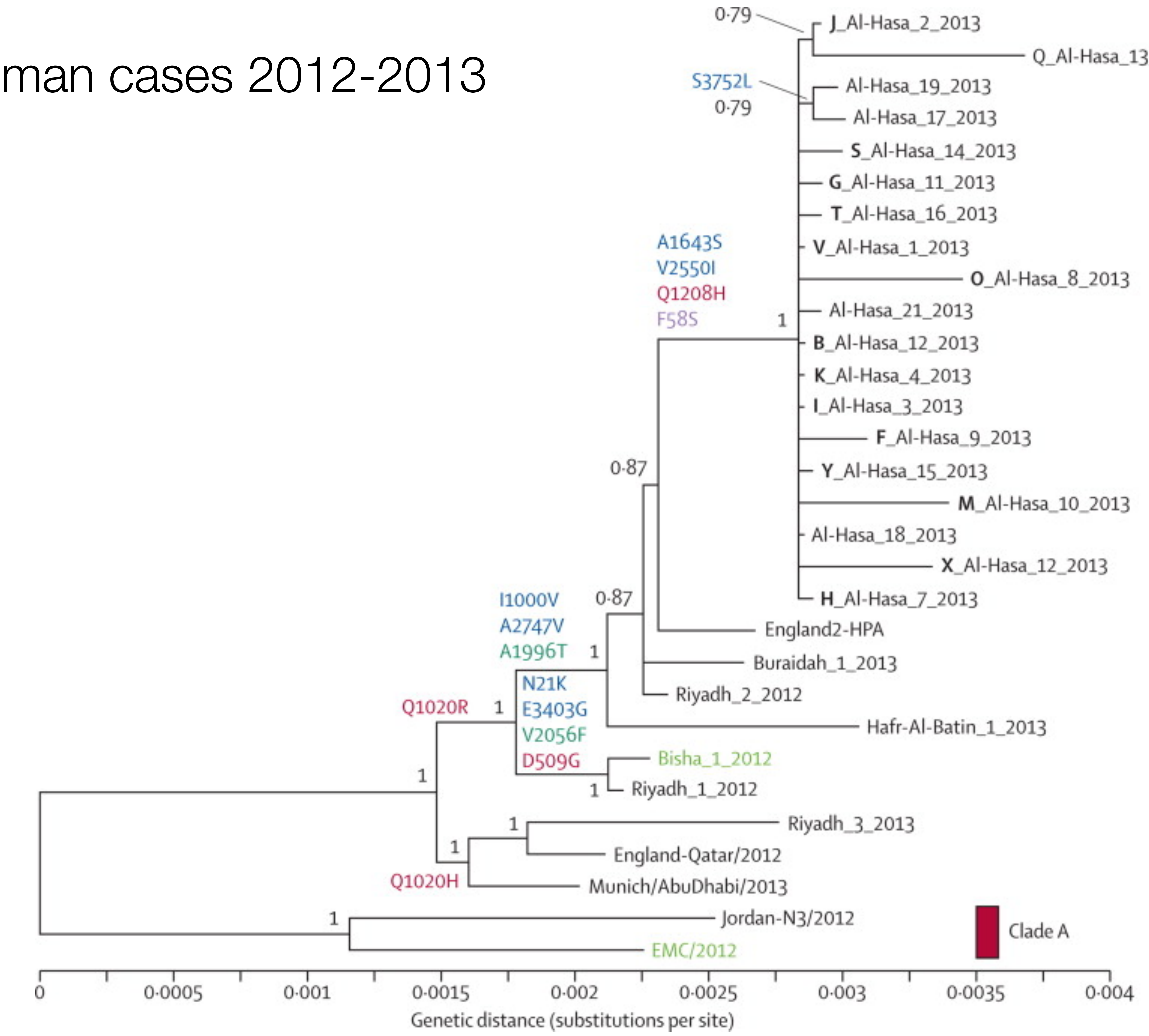
- Reservoirs are diverse (by definition)
- Are human clusters diverse?

● sampled case
● inferred case
● primary case
● unobserved case
— human to human transmission
— reservoir transmission



MERS-CoV epidemic
2012 – ?

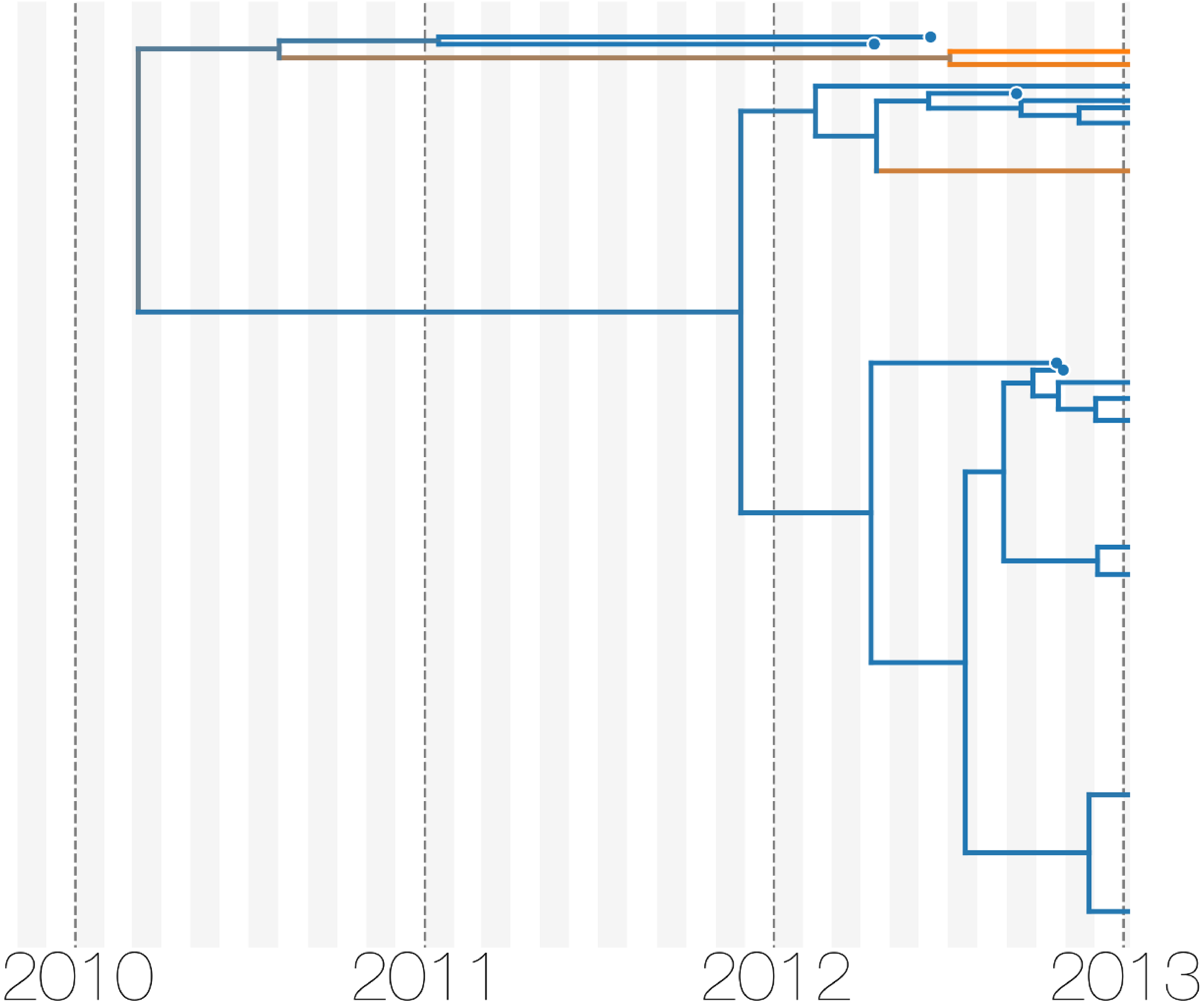
MERS-CoV human cases 2012-2013



Cotten et al (2013) *The* Lancet 382: 1993-2002 DOI: 10.1016/S0140-6736(13)61887-5

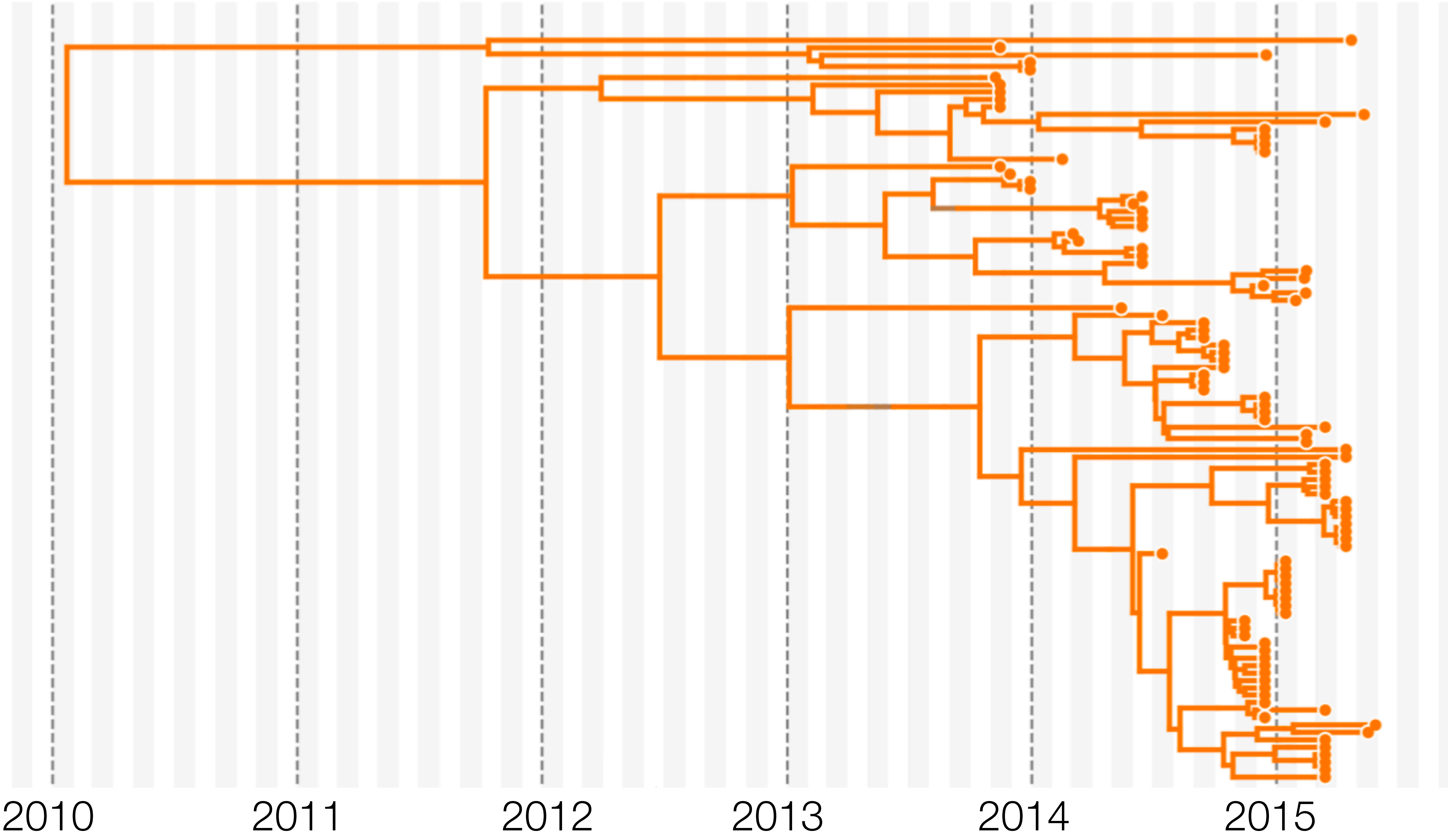
● human

- Circulating in humans since 2010?
- Unlikely - would be global
- high mortality - 8%


Dudas et al (2014) *eLife* **7**:e31257. DOI: 10.7554/eLife.31257

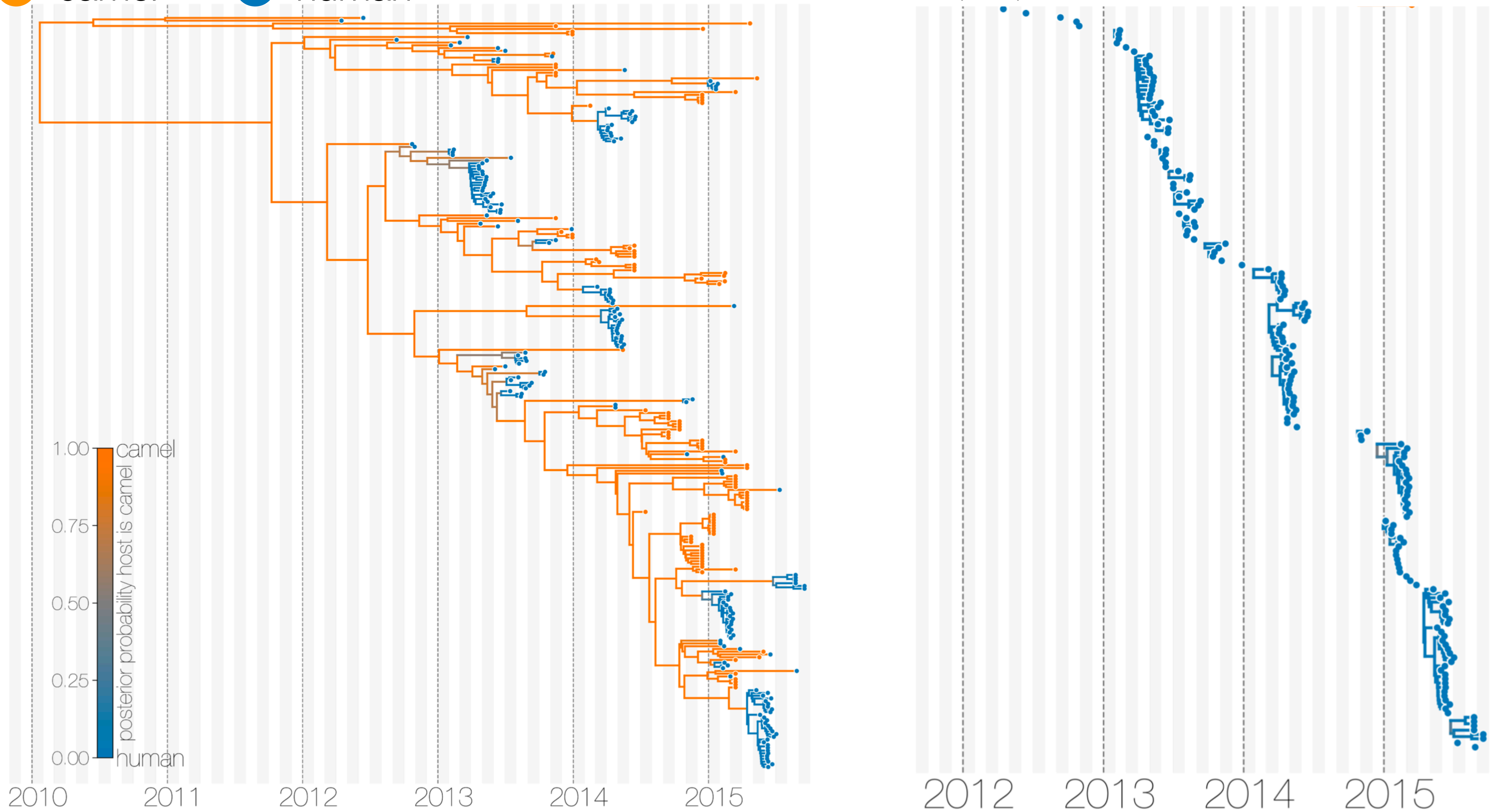
● Camel

Dudas et al (2014) *eLife* **7**:e31257. DOI: 10.7554/eLife.31257



● camel ● human

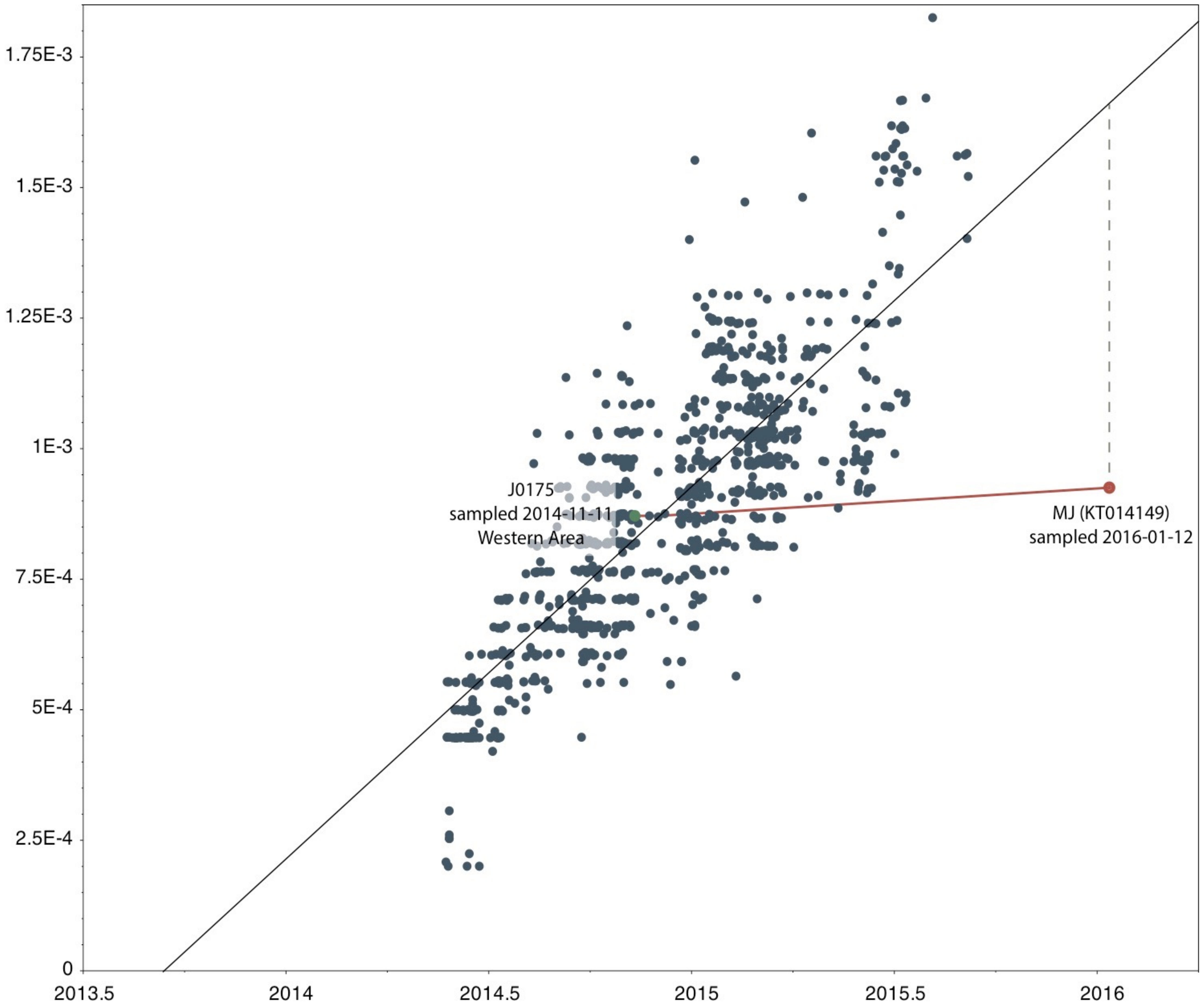
Dudas et al (2014) *eLife* **7**:e31257. DOI: 10.7554/eLife.31257



What can pathogen genomics tell us about a new outbreak?

- Evolution throws up surprises

Ebolavirus 2014 epidemic – Last case in Guinea, 2016



Sequencing revealed the genome was one nucleotide different to a sample isolated in Nov 2014

probably linked to persistence in a survivor

Diallo et al (2016) *Clin Infect Dis* **63**: 1353-1356 doi: 10.1093/cid/ciw601

Outbreak Feb 2021
Samples from 22nd Feb - 26th Feb

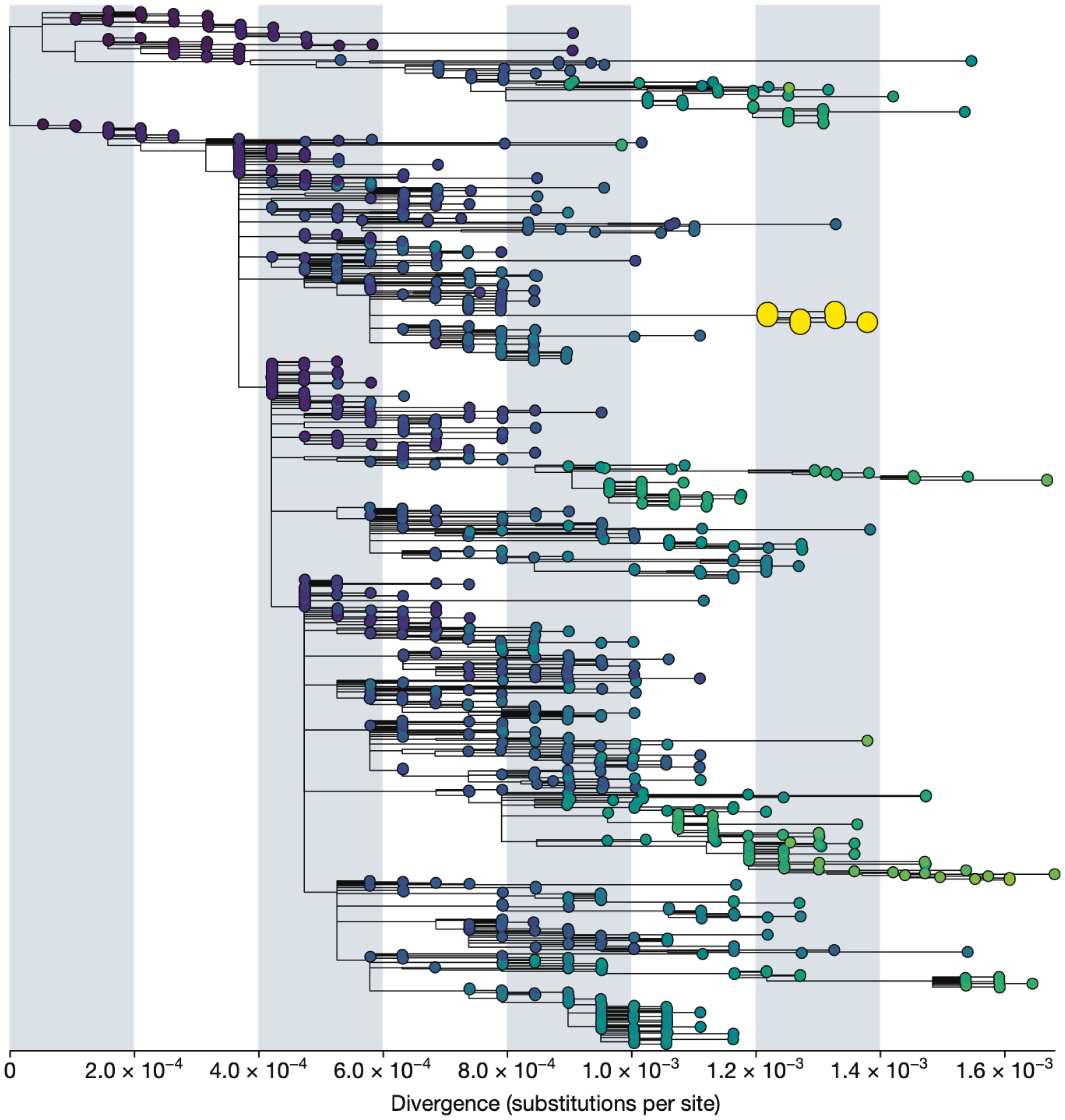
Analysis posted 12th March:
https://virological.org/t/651
https://virological.org/t/643
https://virological.org/t/642

Article

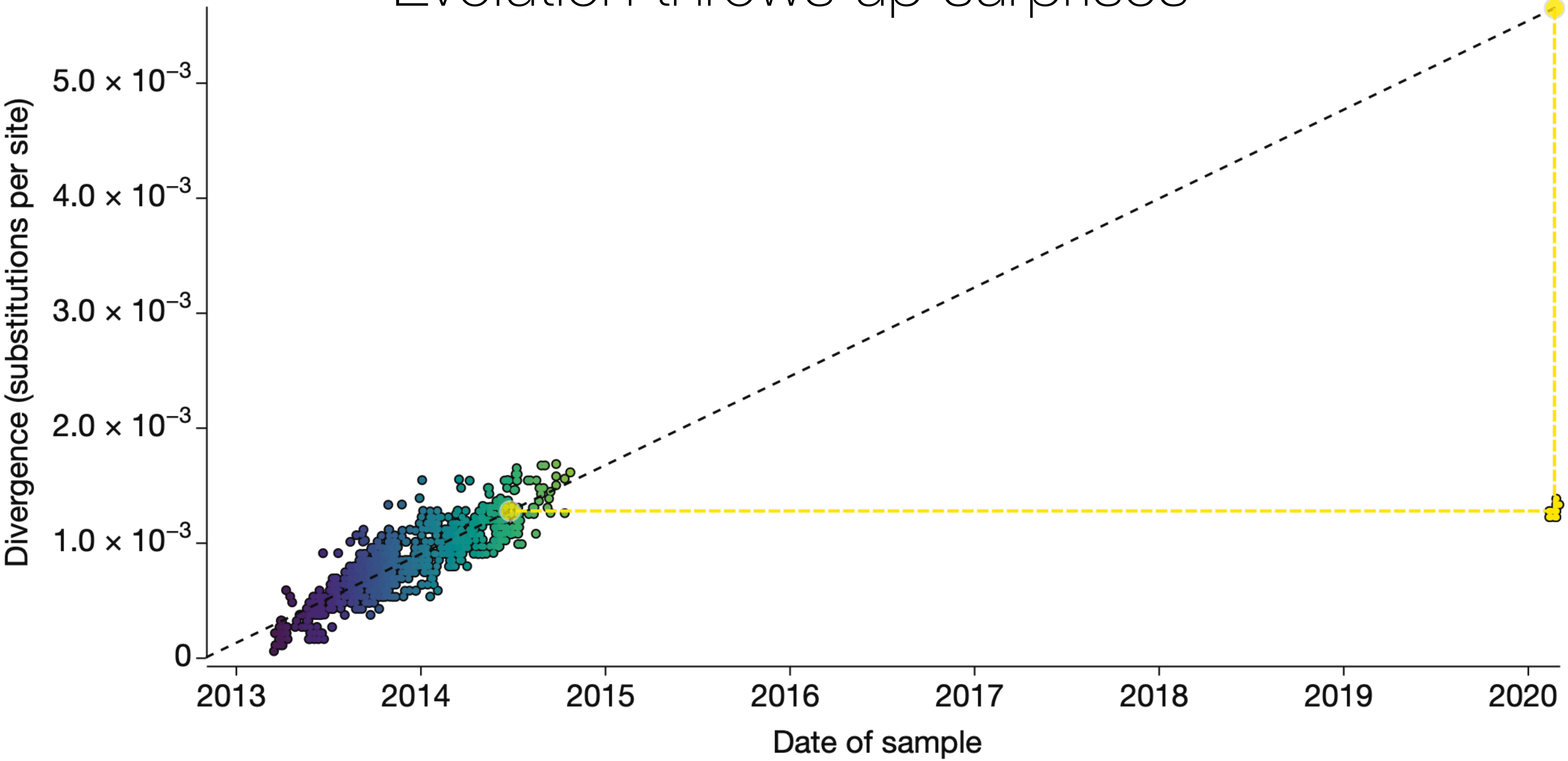
Resurgence of Ebola virus in 2021 in Guinea suggests a new paradigm for outbreaks

https://doi.org/10.1038/s41586-021-03901-9
Received: 6 April 2021
Accepted: 11 August 2021

Alpha Kabinet Keita[1,2,26] ✉, Fara R. Koundouno[3,4,26], Martin Faye[5,26], Ariane Düx[6,26], Julia Hinzmann[4,7,8,26], Haby Diallo[1], Ahidjo Ayouba[2], Frederic Le Marcis[1,2,9], Barré Soropogui[3], Kékoura Ifono[3,4], Moussa M. Diagne[5], Mamadou S. Sow[1,10], Joseph A. Bore[3,11], Sebastien Calvignac-Spencer[6], Nicole Vidal[2], Jacob Camara[3], Mamadou B. Keita[12], Annick Renevey[4,7], Amadou Diallo[5], Abdoul K. Soumah[1], Saa L. Millimono[3,4], Almudena Mari-Saez[6], Mamadou Diop[5], Ahmadou Doré[3], Fodé Y. Soumah[10], Kaka Kourouma[12], Nathalie J. Vielle[4,13], Cheikh Loucoubar[5], Ibrahima Camara[1], Karifa Kourouma[3,4], Giuditta Annibaldis[4,13], Assaïtou Bah[3], Anke Thielebein[4,7], Meike Pahlmann[4,7], Steven T. Pullan[8,11], Miles W. Carroll[8,11], Joshua Quick[14], Pierre Formenty[15], Anais Legand[15], Karla Pietro[16], Michael R. Wiley[16,17], Noel Tordo[18], Christophe Peyrefitte[5], John T. McCrone[19], Andrew Rambaut[19], Youssouf Sidibé[20], Mamadou D. Barry[20], Madeleine Kourouma[20], Cé D. Saouromou[20], Mamadou Condé[20], Moussa Baldé[10], Moriba Povogui[1], Sakoba Keita[21], Mandiou Diakite[22,23], Mamadou S. Bah[22], Amadou Sidibe[9], Dembo Diakite[10], Fodé B. Sako[10], Fodé A. Traore[10], Georges A. Ki-Zerbo[13], Philippe Lemey[24], Stephan Günther[4,7,13], Liana E. Kafetzopoulou[4,7,24], Amadou A. Sall[5], Eric Delaporte[2,25], Sophie Duraffour[4,7,13,27], Ousmane Faye[5,27], Fabian H. Leendertz[6,27], Martine Peeters[2,27], Abdoulaye Toure[1,12,27] & N'. Faly Magassouba[3,27]



Evolution throws up surprises

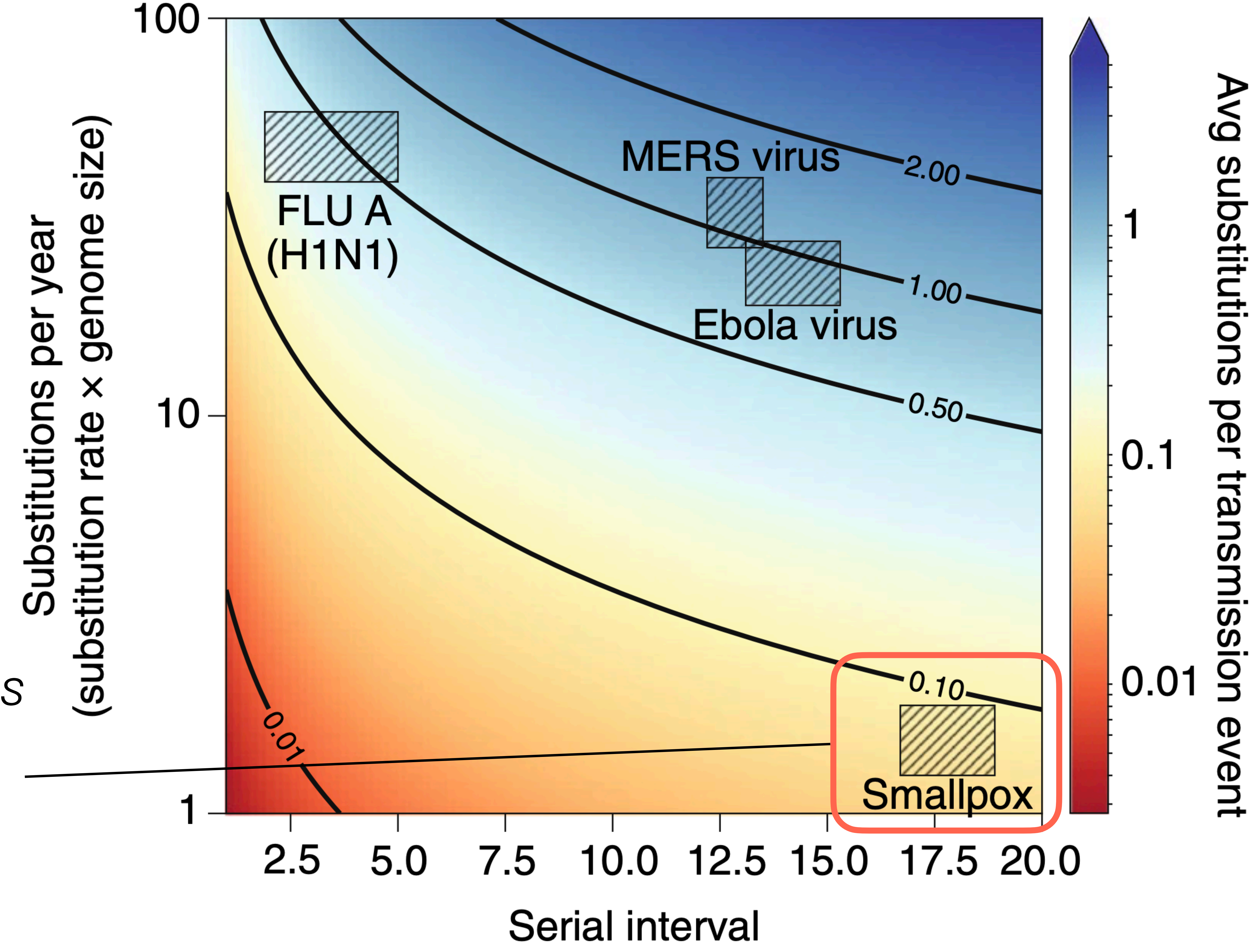


monkeypox virus (MPXV)
1970 – ?

Evolution throws up surprises

- What is the agent?
- What can we assume based on related viruses?

monkeypox virus (MPXV) = *Orthopoxvirus*
variola virus (VARV) evolutionary rate:
0.4 – 1.8 changes per year
(in a 200K genome)



Grubaugh et al (2019) *Nat Micro* | https://doi.org/10.1038/s41564-018-0296-2

Evolution throws up surprises

Monkeypox virus (MPXV)

First draft genome sequence of Monkeypox virus associated with the suspected multi-country outbreak, May 2022 (confirmed case in Portugal)

It's been a while since we've seen vborges — their last post was 10 months ago.

vborges 3 May 20

Joana Isidro[1], Vítor Borges[1], Miguel Pinto[1], Rita Ferreira[1], Daniel Sobral[1], Alexandra Nunes[1], João Dourado Santos[1], Maria José Borrego[3], Sofia Núncio[2], Ana Pelerito[2], Rita Cordeiro[2], João Paulo Gomes[1,*].

1 Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal
2 Emergency and Biopreparedness Unit, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal
3 National Reference Laboratory of Sexually Transmitted Infections, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal

- First genome sequenced from 2022 outbreak
- 20th May from Portugal
- Closest to genomes from 2018 from UK, Nigeria, Israel and Singapore
- 2–7 expected at *Orthopoxvirus* evolutionary rate

https://virological.org/t/799
Isidro et al (2022) *Nature Medicine* 28: 1569–1572

Evolution throws up surprises

Monkeypox virus (MPXV)

- 47 differences: 2–7 expected at *Orthopoxvirus* evolutionary rate
- Assumption that most cases are zoonotic in origin (rodent reservoir?)
- Evidence of rapid adaptation to human as a host?

Evolution throws up surprises

Monkeypox virus (MPXV)

- 47 differences: 2–7 expected at *Orthopoxvirus* evolutionary rate
- Assumption that most cases are zoonotic in origin (rodent reservoir?)
- Evidence of rapid adaptation to human as a host?

  Probably not - 42/47 mutations likely the result of human anti-virus protein APOBEC3

MPXV Clade IIb

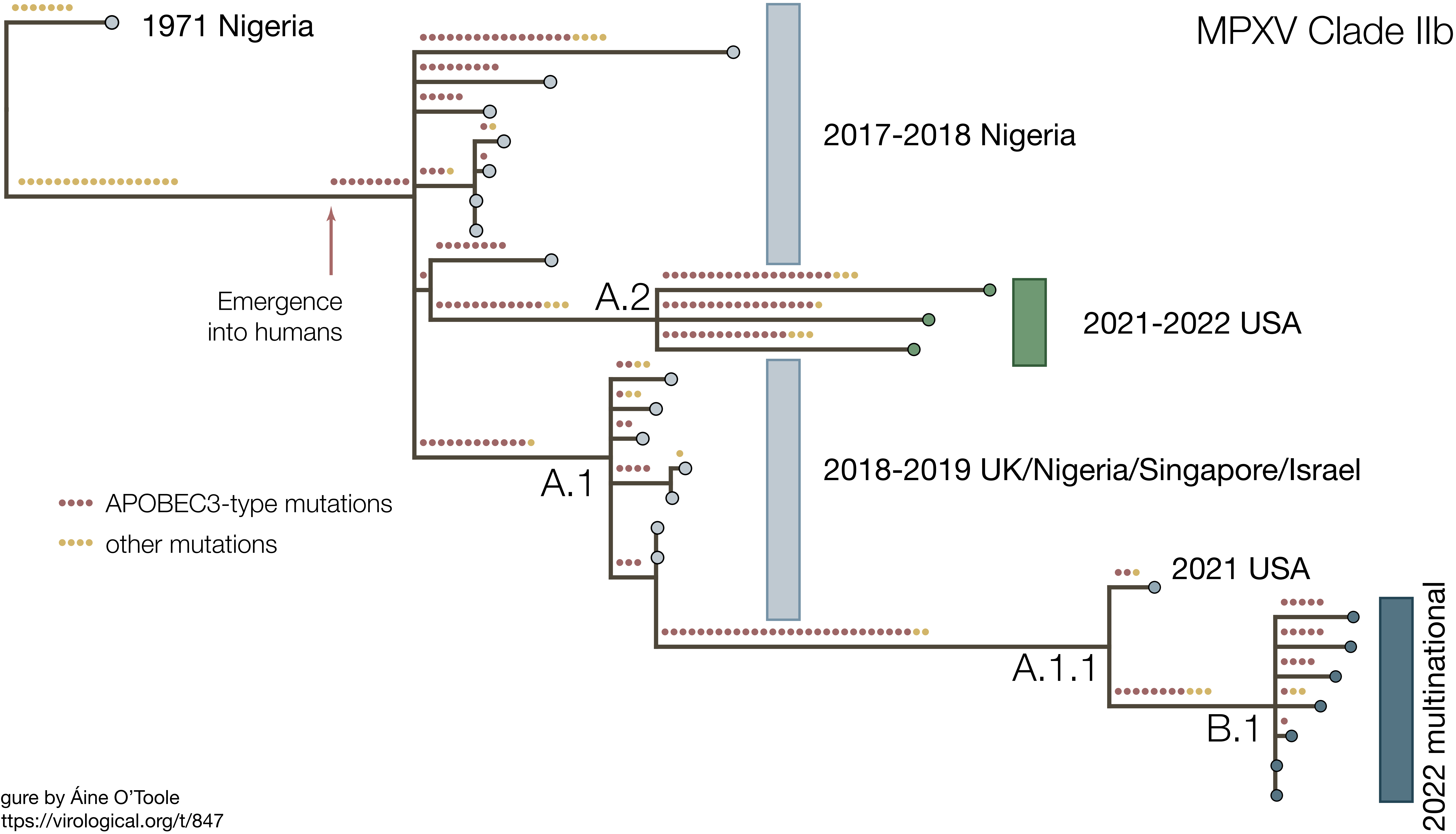
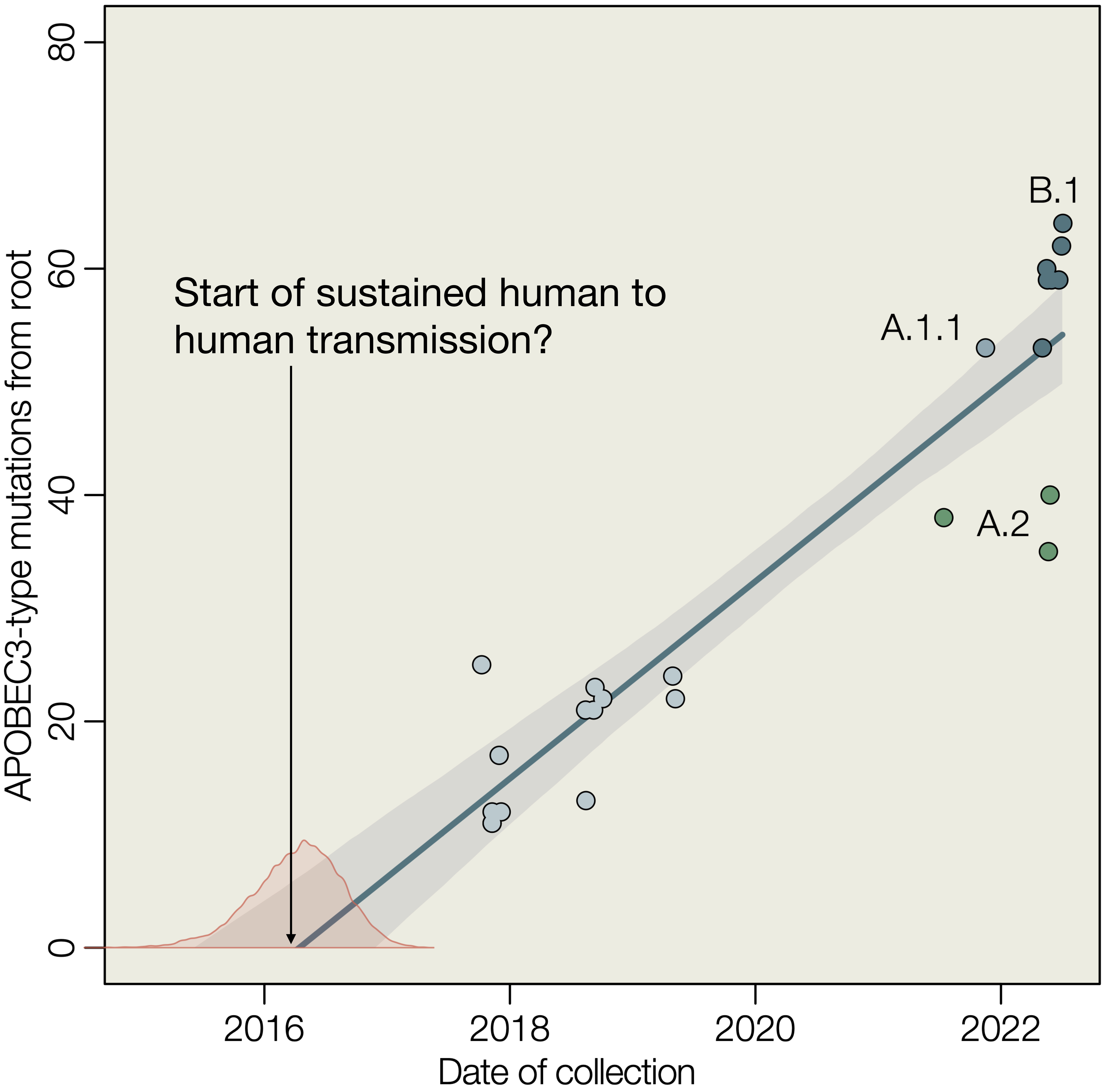


figure by Áine O'Toole
https://virological.org/t/847



https://virological.org/t/885

What does pathogen surveillance need?

- Genome sequences from a random sample of cases (or all cases).
- **Data-linkage** – dates of sampling, dates of onset, exposure to cases, exposure to animals, travel history.
- **Speed** – any interpretation will be about what was happening at the time the samples were collected. Can the information still be acted upon?
- **Expertise** – every virus behaves differently, many evolutionary processes at work. Interpretation of the data is always challenging.

What does pathogen surveillance need?

- Pathogen genomics rapidly becoming globally accessible as a technology

- **Expertise** – investment in pathogen bioinformatics, genomic epidemiology analysis and interpretation

University of Edinburgh, current and former:
Corey Ansley, Rachel Colquhoun, Gytis Dudas, Kate Duggan, Verity Hill, Ben Jackson, Danny Maloney, JT McCrone, Ifeanyi Omah, Áine O’Toole, Emily Scher, Xiaoyu Yu

ARTIC Network: http://artic.network/

COVID-19 Genomics UK Consortium: https://www.cogconsortium.uk