Key differences between viral and bacterial genomics

Alicia Arnott
Deputy Head of Epidemiology, VIDRL

Doherty Institute

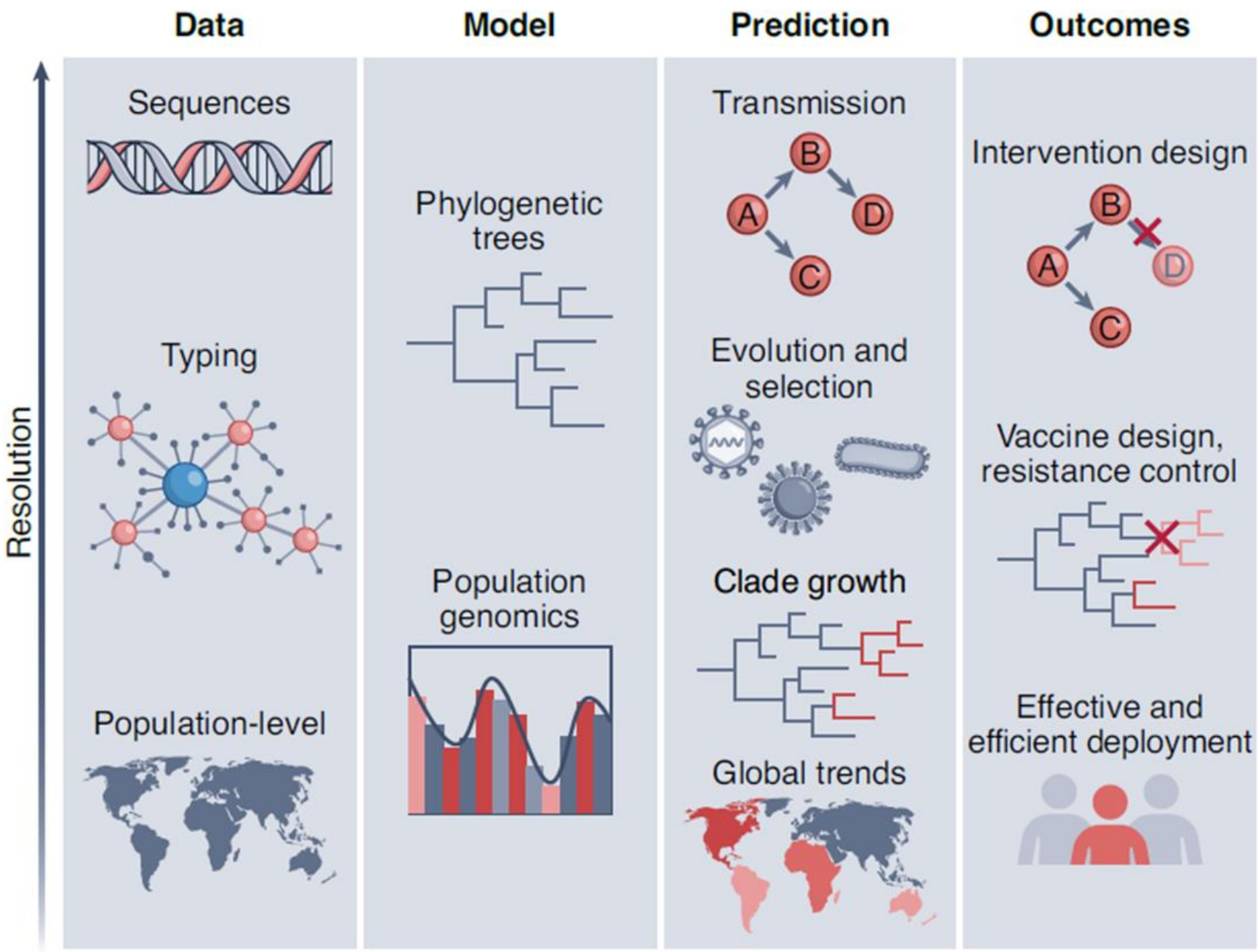
THE UNIVERSITY OF MELBOURNE

The Royal Melbourne Hospital

A joint venture between The University of Melbourne and The Royal Melbourne Hospital

Introduction

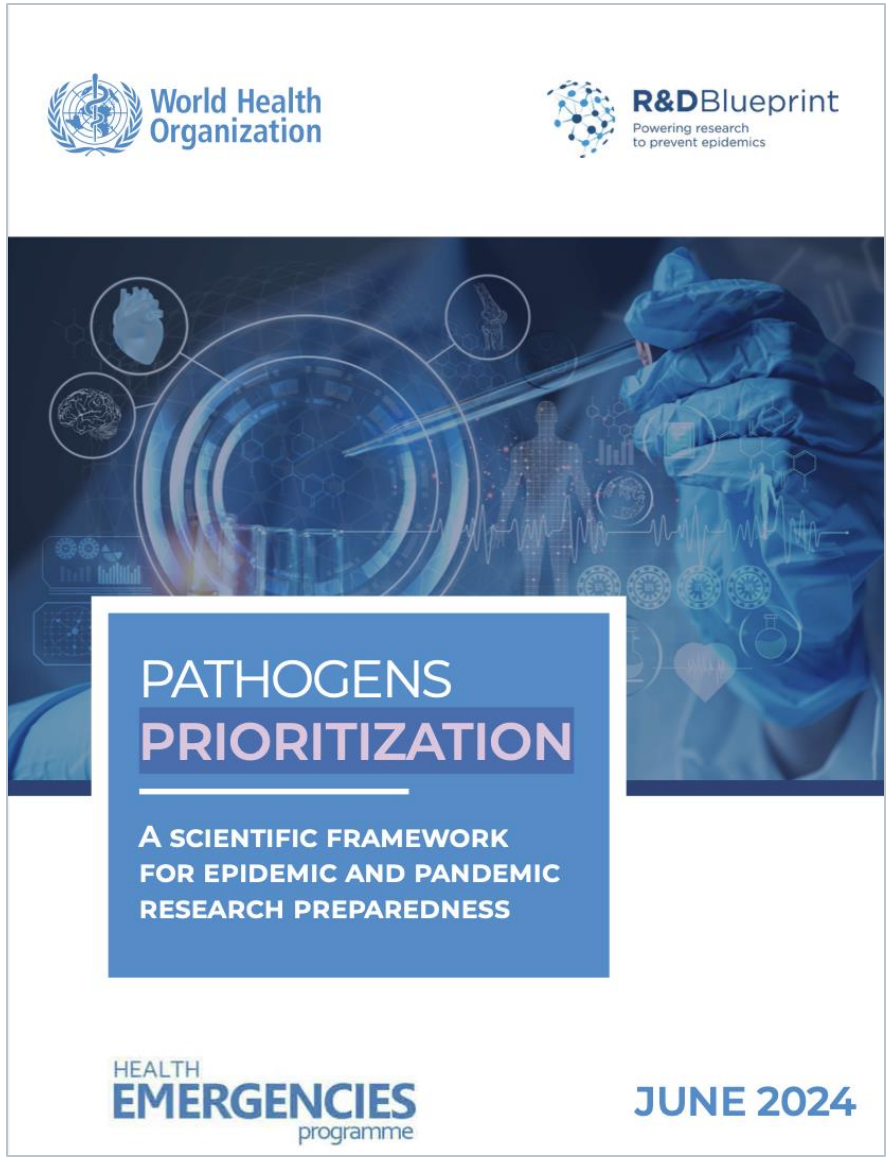
- Translational public health genomics is increasingly being applied to investigate pathogens of public health importance
- Genomics was being utilised for the public health management of infectious diseases prior to the pandemic
- Pandemic highlighted the public health utility of pathogen genomics, increasing global implementation efforts



Stockdale et al., *Nature Microbiology:* vol 7, November 2022

Post-pandemic pathogen prioritization

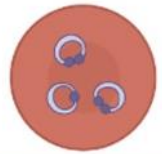
- The number of pathogens of public health concern is very large, while the resources for disease research and development is limited
- In 2022, the WHO employed a global panel of experts to identify which pathogen families/diseases pose the greatest public health risk due to their epidemic potential and/or whether there is no or insufficient countermeasures



Viruses pose greatest global pandemic threat







	RNA Viruses	DNA Viruses	Bacteria	Fungi	Protozoa[a] (Parasites)
Pandemic Potential	**High**	**Moderate**	**Moderate, historically high**	**Low**	**Low**
Features	Continual threat due to RNA genomes prone to mutation; fast viral reproduction and transmission; propensity to cause respiratory infections with airborne potential	Less prone to mutations than RNA viruses and transmission less likely to be airborne	Slower reproduction than viruses, less mutation prone, but able to transfer fitness genes; Threat now reduced due to fast detection and well-developed treatments	Nontraditional pandemic threat due to slow speed of spread and effective antifungals, though multidrug resistant organisms are increasing	Pandemic threat is low due to spread through vector, however malaria is estimated to contribute to half of all human deaths throughout history
Speed of Transmission	Fast	Fast–Medium	Medium	Slow	Slow–Medium
Mode of Transmission	Airborne Respiratory, Droplet, contact with infected fluids	Droplet, contact with infected fluids or materials (fabric)	Droplet (respiratory); Contact with infected humans or animals	Airborne spores; Contact with infected humans or animals; Environmental	Generally through an intermediate host (vector)
Detection Methods	PCR-based (hours) Select Antigens (hours–days)	PCR- based (hours) Antibodies (hours–days)	Bacterial culture (days), PCR-based (hours)	Fungal culture (days) Select antigen tests	Microscopy (hours) PCR-based (hours)
Countermeasures	Masks for respiratory transmission, antivirals, vaccines if available	Vaccines, standard precautions[b]	Vaccines, antibiotics, improved sanitation	Antifungals, control of environmental fungal threats	Barriers between hosts and humans, elimination of reservoirs, anti-parasitic medications
Examples of Pandemic Pathogens	SARS-CoV1 (SARS) SARS-CoV2 (COVID-19) Ebola (Hemorrhagic Fever) HIV (AIDS) Influenza (Spanish, Swine, Avian flus)	*Variola major* (Smallpox)	*Yersinia pestis* (Bubonic Plague) *Vibrio cholera* (cholera) *Mycobacterium tuberculosis* (TB)	No human pandemics caused by fungi; widespread fungal disease affecting bats, frogs; fungi posited to contribute to dinosaur die-off	*Plasmodium spp* (malaria) *Trypanasoma lewisi* (widespread death of the Christmas Island Rat)

https://sciencepolicyreview.org/wp-content/uploads/securepdfs/2021/08/Matzko_Preventing_Pandemics.pdf

Viral pathogens pose greatest pandemic threat

- In 2022, the priority diseases were:
  - COVID-19
  - Crimean-Congo haemorrhagic fever
  - Ebola virus disease and Marburg virus disease
  - Lassa fever
  - Middle East respiratory syndrome coronavirus (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS)
  - Nipah and henipaviral diseases
  - Rift Valley fever
  - Zika
  - "Disease X"*

	2017	2018	2024		
Family	**Priority Pathogens**	**Priority Pathogens**	**PHEIC risk**	**Priority Pathogens**	**Prototype Pathogens**
Adenoviridae			Low-Medium		Recombinant Mastadenovirus
Adenoviridae			Low-Medium		Mastadenovirus blackbeardi serotype 14
Anelloviridae			Low		
Arenaviridae	Arenaviral hemorrhagic fevers including Lassa Fever	Lassa Fever virus	High	Mammarenavirus lassaense	Mammarenavirus lassaense
Arenaviridae			High		Mammarenavirus juninense
Arenaviridae			High		Mammarenavirus lujoense
Astroviridae			Low		Mamastrovirus virginiaense
Bacteria			High	*Vibrio cholerae serogroup 0139*	
Bacteria			High	*Yersinia Pestis*	
Bacteria			High	*Shigella dysenteriae serotype 1*	
Bacteria			High	*Salmonella enterica non typhoidal serovars*	
Bacteria			High	*Klebsiella pneumoniae*	
Bornaviridae			Low		Orthobornavirus bornaense
Coronaviridae	Middle East Respiratory Syndrome Coronavirus	Middle East Respiratory Syndrome Coronavirus	High	*Subgenus* Merbecovirus	*Subgenus* Merbecovirus
Coronaviridae	Other highly pathogenic coronaviral diseases such as Severe Acute Respiratory Syndrome	Severe Acute Respiratory Syndrome	High	*Subgenus* Sarbecovirus	*Subgenus* Sarbecovirus
Filoviridae	Filoviral diseases Ebola	Ebola virus disease	High	Orthoebolavirus zairense	Orthoebolavirus zairense
Filoviridae	Filoviral diseases Marburg	Marburg virus disease	High	Orthomarburgvirus marburgense	
Filoviridae			High	Orthoebolavirus sudanense	
Flaviviridae	Zika virus	Zika virus	High	Orthoflavivirus	Orthoflavivirus
Flaviviridae			High	Orthoflavivirus denguei	Orthoflavivirus denguei
Flaviviridae			High		Orthoflavivirus encephalitidis
Flaviviridae			High		Orthoflavivirus nilense
Hantaviridae			High	Orthohantavirus sinnombreense	Orthohantavirus sinnombreense
Hantaviridae			High	Orthohantavirus hantanense	
Hepadnaviridae			Low		Orthohepadnavirus hominoidei genotype C

	2017	2018	2024		
Family	**Priority Pathogens**	**Priority Pathogens**	**PHEIC risk**	**Priority Pathogens**	**Prototype Pathogens**
Hepeviridae			Low		Paslahepevirus balayani genotype 3
Herpesviridae			Low		
Nairoviridae	Crimean Congo Haemorrhagic Fever	Crimean Congo Haemorrhagic Fever	High	Orthonairovirus haemorrhagiae	Orthonairovirus haemorrhagiae
Orthomyxoviridae			High	Alphainfluenzavirus Influenzae H1	Alphainfluenzavirus Influenzae H1
Orthomyxoviridae			High	Alphainfluenzavirus Influenzae H2	
Orthomyxoviridae			High	Alphainfluenzavirus Influenzae H3	
Orthomyxoviridae			High	Alphainfluenzavirus Influenzae H5	Alphainfluenzavirus Influenzae H5
Orthomyxoviridae			High	Alphainfluenzavirus Influenzae H6	
Orthomyxoviridae			High	Alphainfluenzavirus Influenzae H7	
Orthomyxoviridae			High	Alphainfluenzavirus Influenzae H10	
Papillomaviridae			Low		
Paramyxoviridae	Nipah and related henipaviral diseases	Nipah and henipaviral diseases	High	Henipavirus nipahense	Henipavirus nipahense
Parvoviridae			Low		Protoparvovirus carnivoran
Peribunyaviridae			Low		Orthobunyavirus oropoucheense
Phenuiviridae	Severe Fever with Thrombocytopenia Syndrome		High	Bandavirus dabieense	Bandavirus dabieense
Phenuiviridae	Rift Valley Fever	Rift Valley Fever	High		Phlebovirus riftense
Picobirnaviridae			Low		Orthopicobirnavirus hominis
Picornaviridae			Medium	Enterovirus coxsackiepol	
Picornaviridae			Medium		Enterovirus alphacoxsackie 71
Picornaviridae			Medium		Enterovirus deconjucti 68
Pneumoviridae			Low-Medium		Metapneumovirus hominis
Polyomaviridae			Low		
Poxviridae			High	Orthopoxvirus variola	
Poxviridae			High		Orthopoxvirus vaccinia
Poxviridae			High	Orthopoxvirus monkeypox	Orthopoxvirus monkeypox
Retroviridae			Medium	Lentivirus humimdef1	Lentivirus humimdef1
Rhabdoviridae			Low		Genus Vesiculovirus
Sedoreoviridae			Low		Genus Rotavirus
Spinareoviridae			Low		Orthoreovirus mammalis
Togaviridae			High	Alphavirus chikungunya	Alphavirus chikungunya
Togaviridae			High	Alphavirus venezuelan	Alphavirus venezuelan
Pathogen X	Pathogen X	Pathogen X		Pathogen X	

https://www.who.int/publications/m/item/pathogens-prioritization-a-scientific-framework-for-epidemic-and-pandemic-research-preparedness

Genomics is an essential tool for public health management of infectious diseases

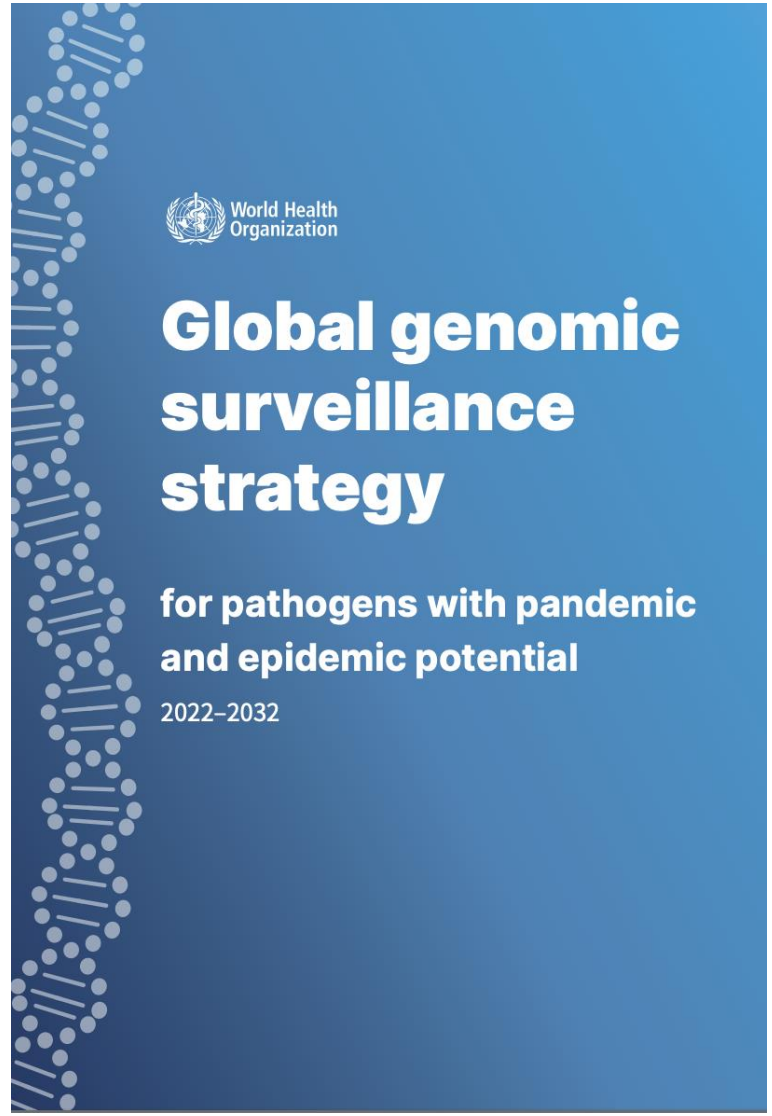
- WGS can overcome important limitations of conventional epidemiological and laboratory methods:

Epidemiological

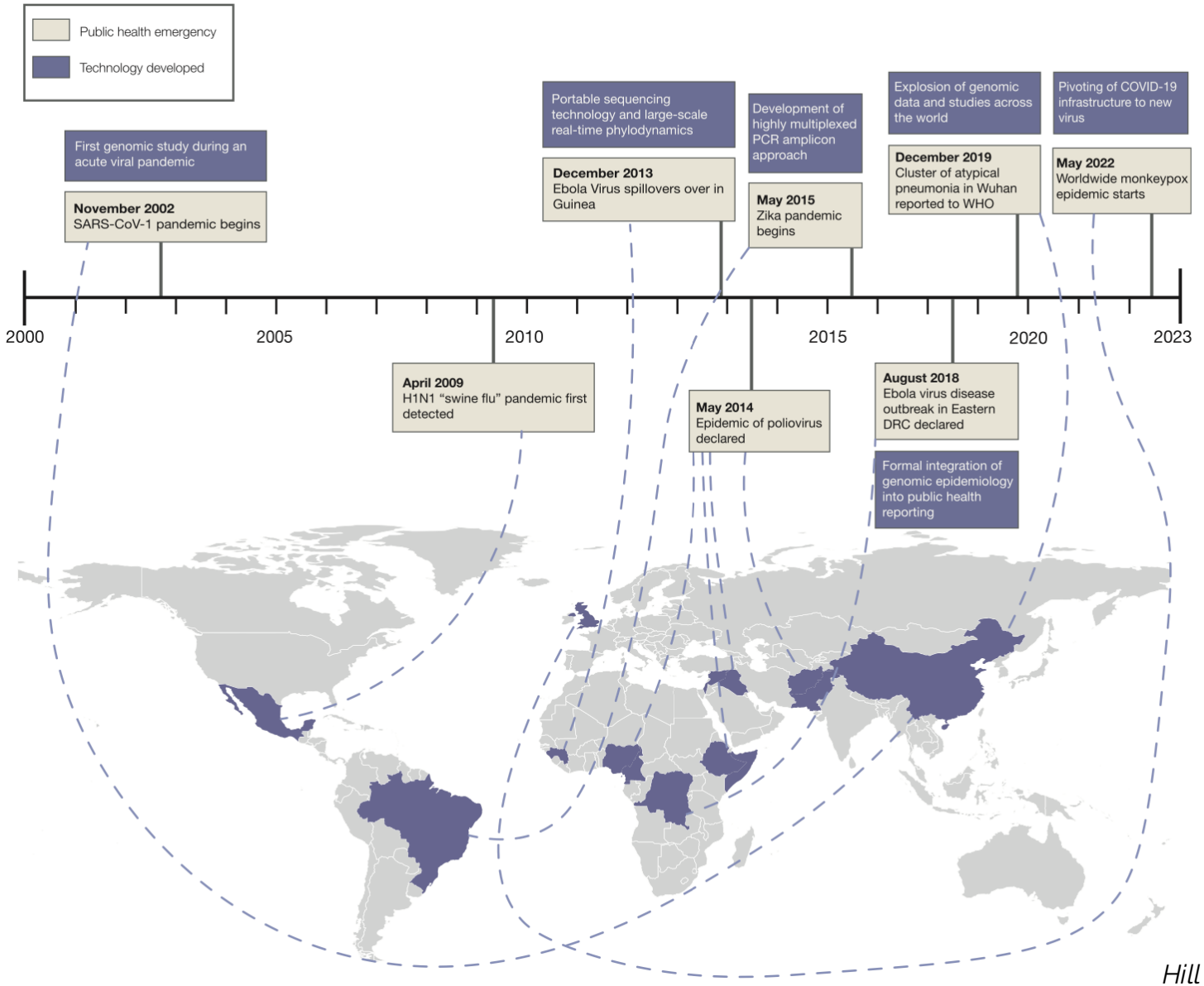
- Poor case recall
- Case lost to follow-up
- Deliberately misleading information
- Missing cases
- Confirm/refute tenuous epidemiological links

Laboratory

- Multiple tests required to answer different questions about the pathogen
- Often cannot definitively determine whether pathogens are identical

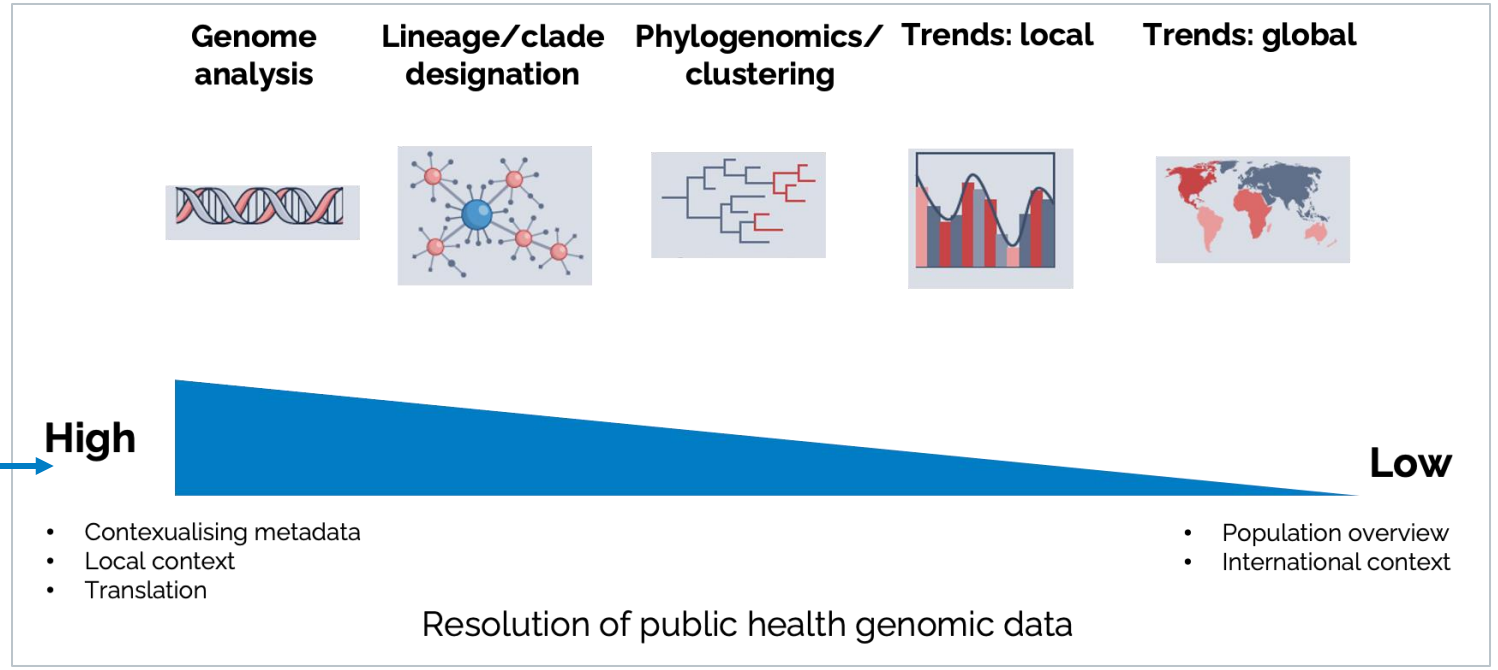


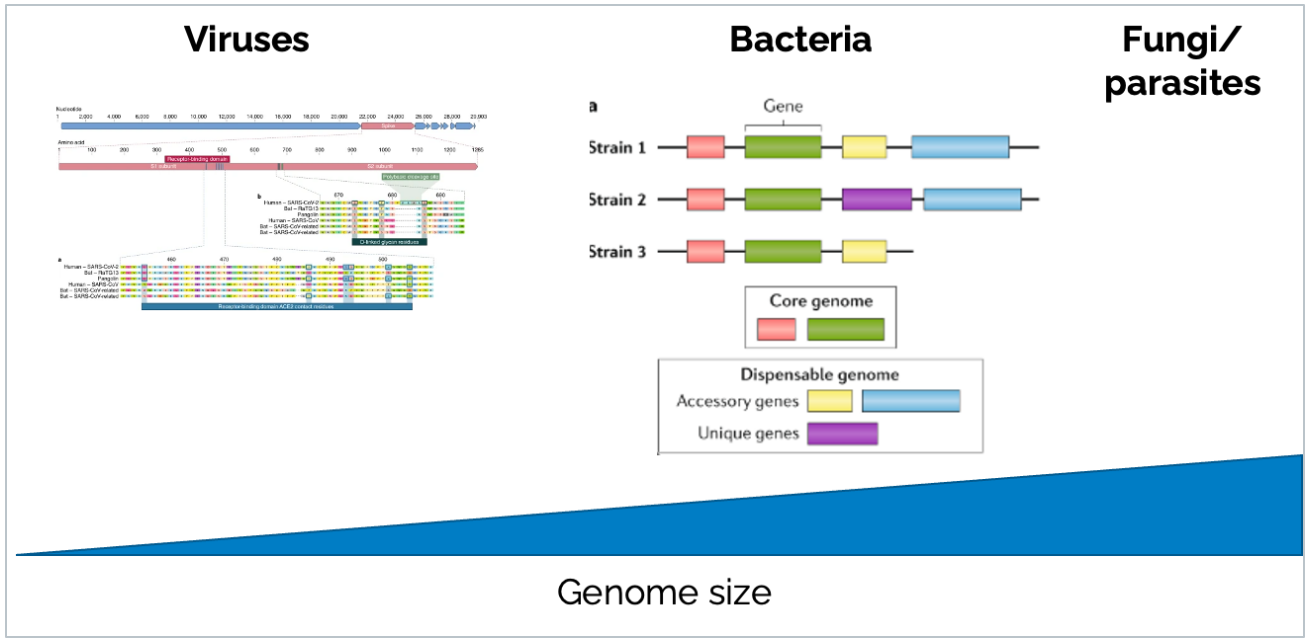
Genomics for investigation of viral outbreaks over time



Hill et al., 2023, Cell Host & Microbe

How you compare sequences differs by **bug** and **question**





How does sequencing and analysis differ between viruses and bacteria?

Key differences between viral and bacterial pathogens

Viral and bacterial genomes typically differ in the type of nucleic acid they contain, their size, and how they are organised:

Factor	Virus	Bacteria
Genetic material	DNA or RNA	DNA
Typical genome size	+	++
Genome complexity	+	++
Genome organisation	Single/double stranded	Circular
Plasmids present	No	Yes
Primary detection methods	Molecular	Culture/molecular

WGS: laboratory procedure

Culture-independent WGS

Culture-dependent WGS

Isolate pathogen and extract nucleic acid

Library preparation and sequencing

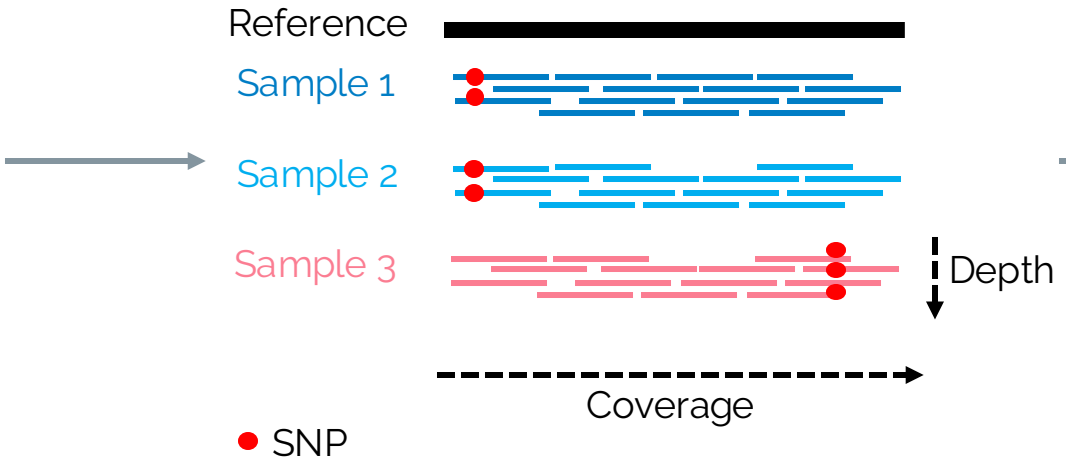
Generation of raw data

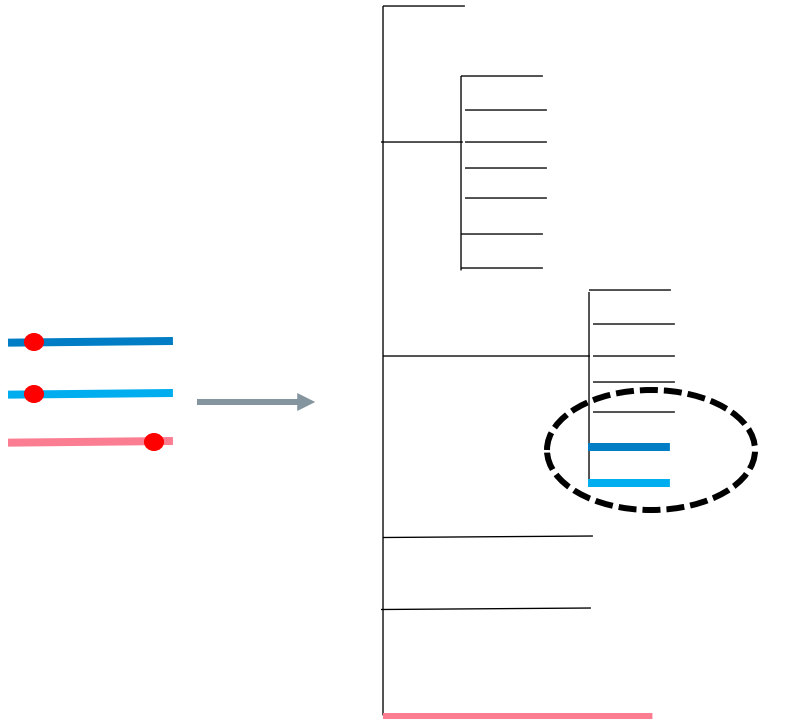
Key differences between viral and bacterial pathogens

Viral and bacterial genomes typically differ in how they are analysed and compared:

Factor	Virus	Bacteria
Whole genome alignment	Yes	No
Core genome/genes	No	Yes
Typing methods	Lineages/clades	Sequence types (STs), MLST, cgMLST
Phenotypic AMR	Extremely limited	Yes
Known AMR mutations	Extremely limited understanding: pathogen specific mutations	Often well understood: pathogen agnostic genes/pathogen specific mutations

WGS: data analysis





Data quality control

- Speciation
- Trimming

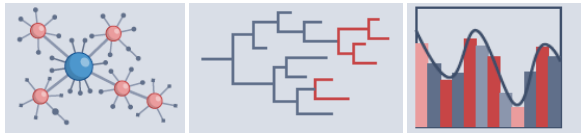
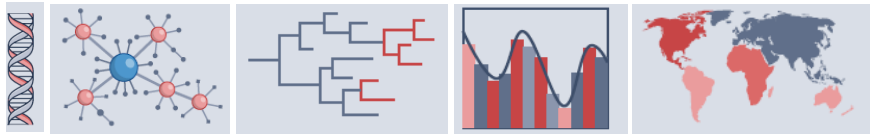
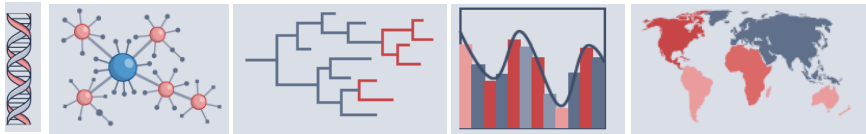
Mapping to reference and single nucleotide polymorphism (SNP) calling

Phylogenetic tree generation and cluster identification

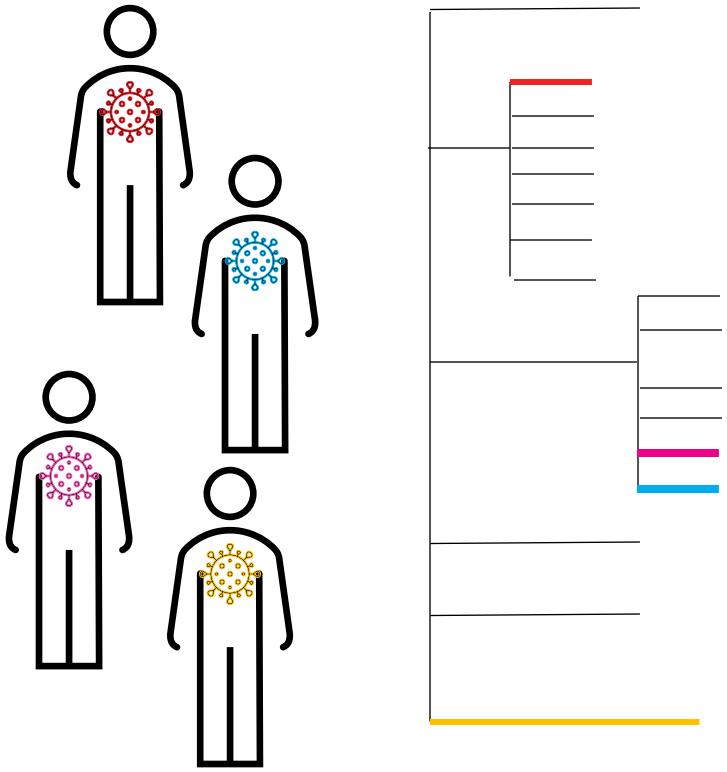
How does interpretation of genomic data differ between viruses and bacteria?

Typical public health objectives

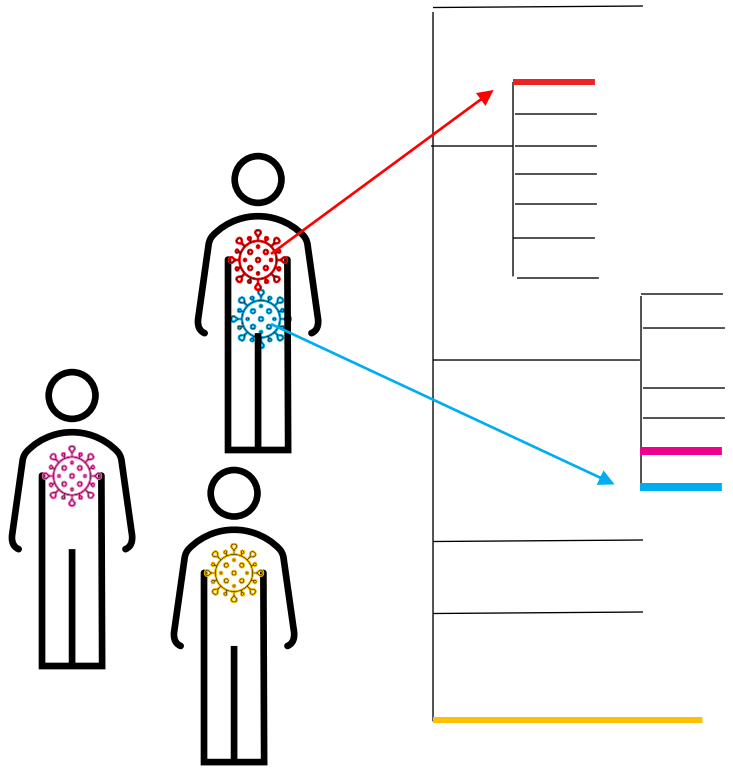
- Are cases linked/exposed to the same source?
- Is the viral lineage/sub-clade novel?
- Will the vaccines/diagnostic tests/antivirals still work?
- Is the lineage/sub-clade associated with severe disease?

Epi	Genomic
	
	
	
	

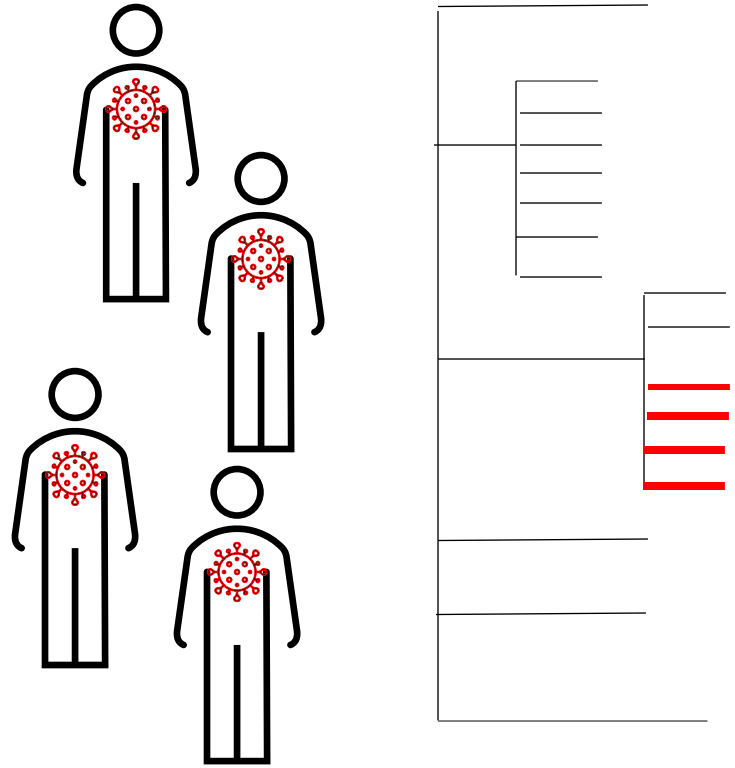
Not all viral pathogen populations are the same



One case = one distinct consensus

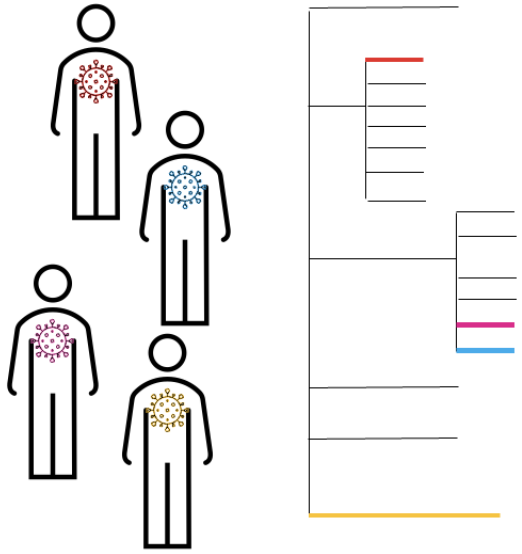


One case > one distinct consensus (i.e. Mpox)



One case = one identical consensus (i.e. beginning explosive transmission)

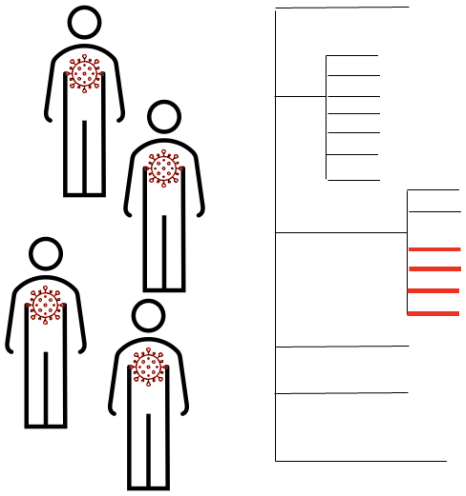
Key considerations in interpretation of standard genomic epidemiological data: resp virus



One case = one distinct consensus

- Directly linking cases must be interpreted in epidemiological context
- Comprehensive global genomic surveillance and high geographical representation permits geographical source attribution and understanding of currently circulating lineages
- Genomic relatedness between cases can be attributed to direct link(s) in the context of robust epidemiological evidence
- A genomic threshold for relatedness between cases has been determined, thus genomic clustering can be performed

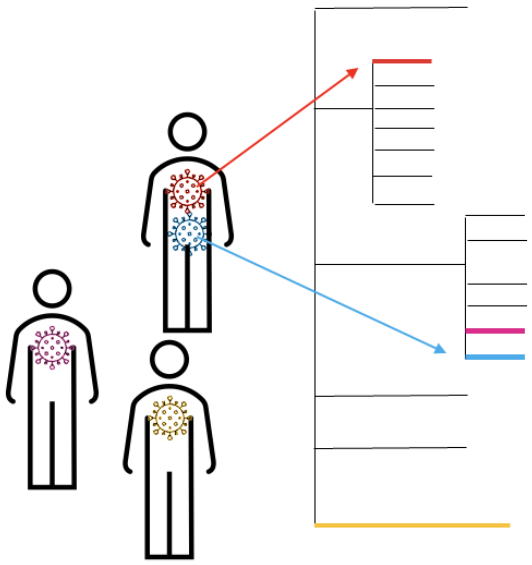
Key considerations in interpretation of genomic epidemiological data: emergent virus



One case = one identical consensus (i.e. beginning explosive transmission)

- Directly linking cases cannot be conducted using genomics alone and genomic data must be interpreted in epidemiological context
- Lack of global genomic surveillance and bias in geographical representation limits geographical source attribution and understanding of currently circulating lineages
- Genomic relatedness between cases should only be attributed to a common exposure(s) rather than direct link(s) in the absence of robust epidemiological evidence
- Within-host diversity: longitudinal sampling required to better understand
- A genomic threshold for relatedness between cases has not been determined. Genomic clustering may not be able to be performed

Key considerations in interpretation of complex genomic epidemiological data: Mpox



One case > one distinct consensus (i.e. Mpox)

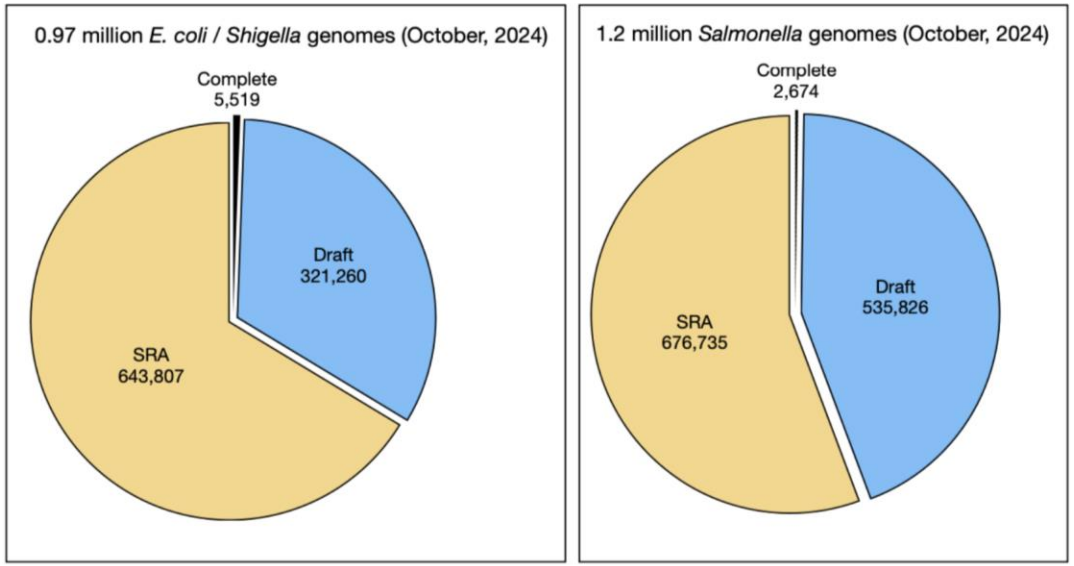
- Directly linking cases cannot be conducted using genomics alone and genomic data must be interpreted in epidemiological context
- Lack of global genomic surveillance and bias in geographical representation limits geographical source attribution and understanding of currently circulating lineages
- Genomic relatedness between cases should only be attributed to a common exposure(s) rather than direct link(s) in the absence of robust epidemiological evidence
- Within-host diversity: multiple sub-lineages and sequence diversity have been detected from different bodily sites within the same host, potentially related to complexity of transmission
- A genomic threshold for relatedness between cases has not been determined. Genomic relatedness is confounded by ongoing viral evolution, within and between hosts, and presence of large genome deletions which limits available genomic data for comparison.

Examples of how genomic data can inform public health management of viral pathogens

Amount of available genomic data differs

- Viral sequencing was performed during the pandemic at a scale previously never achieved
  - Culture-independent sequencing
  - Smaller, less complex genome
  - Ease of specimen collection and transport

genomes high burden bacteria



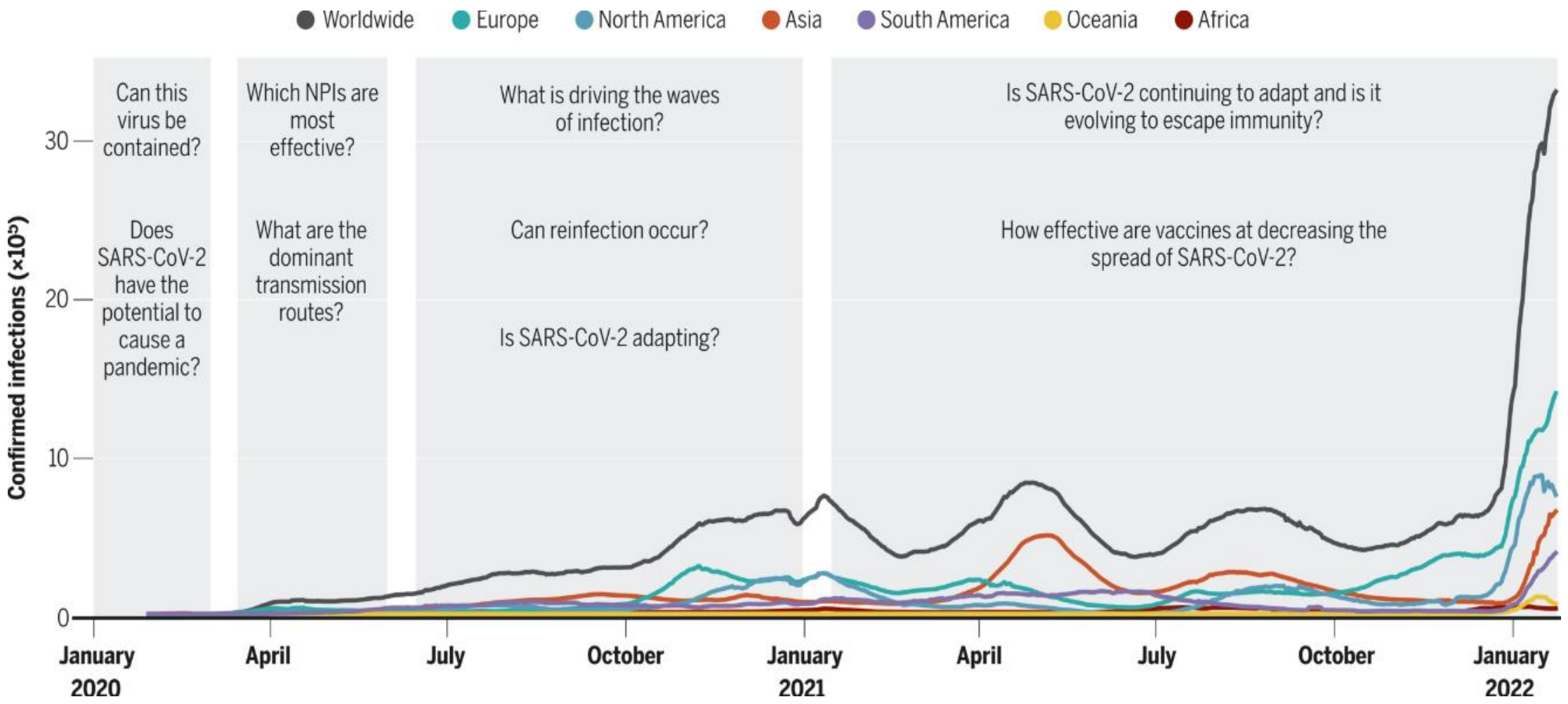
doi: https://doi.org/10.56367/OAG-045-11822

genomes high burden/emergent viruses

Pathogen	Time period	Genomes
SARS-CoV-2	2020-present	>20,000,000
Influenza	~2010-present	>500,000
Dengue	~2011-present	>16,000
Mpox	22-present	10,832

Source: https://gisaid.org and https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/

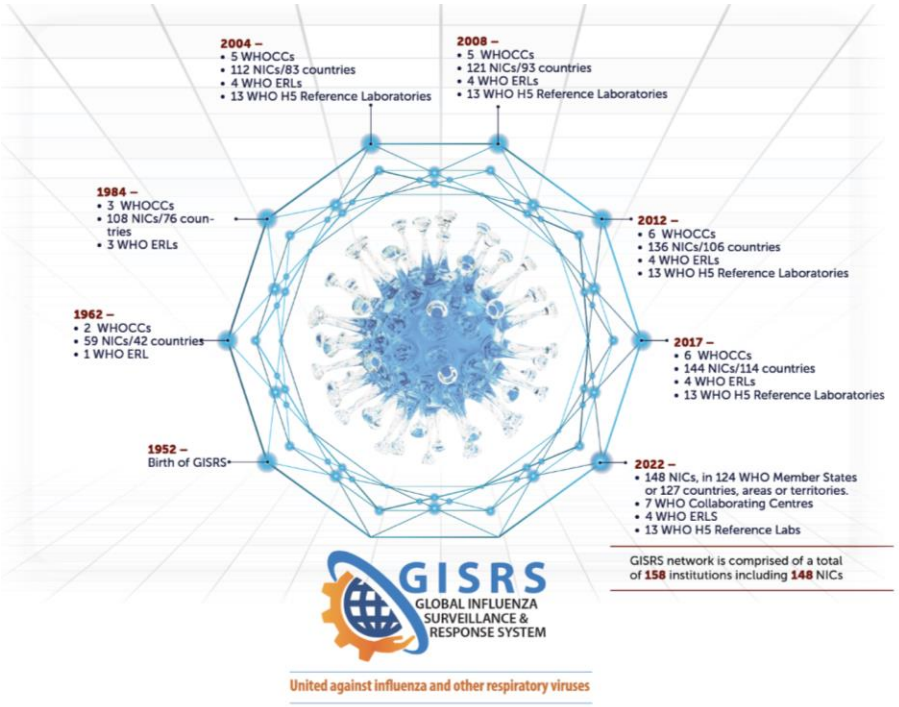
Epidemiological questions answered using genomics during the COVID-19 pandemic

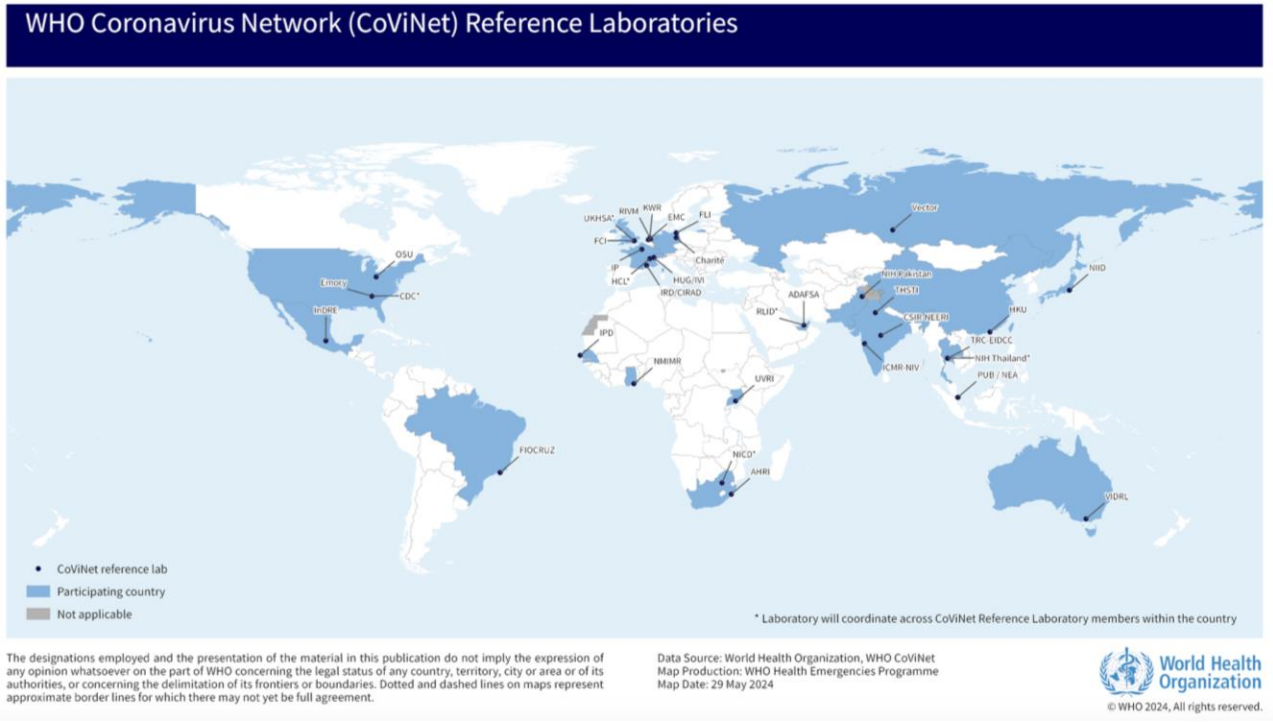


The changing epidemiology of SARS-CoV-2 | Science

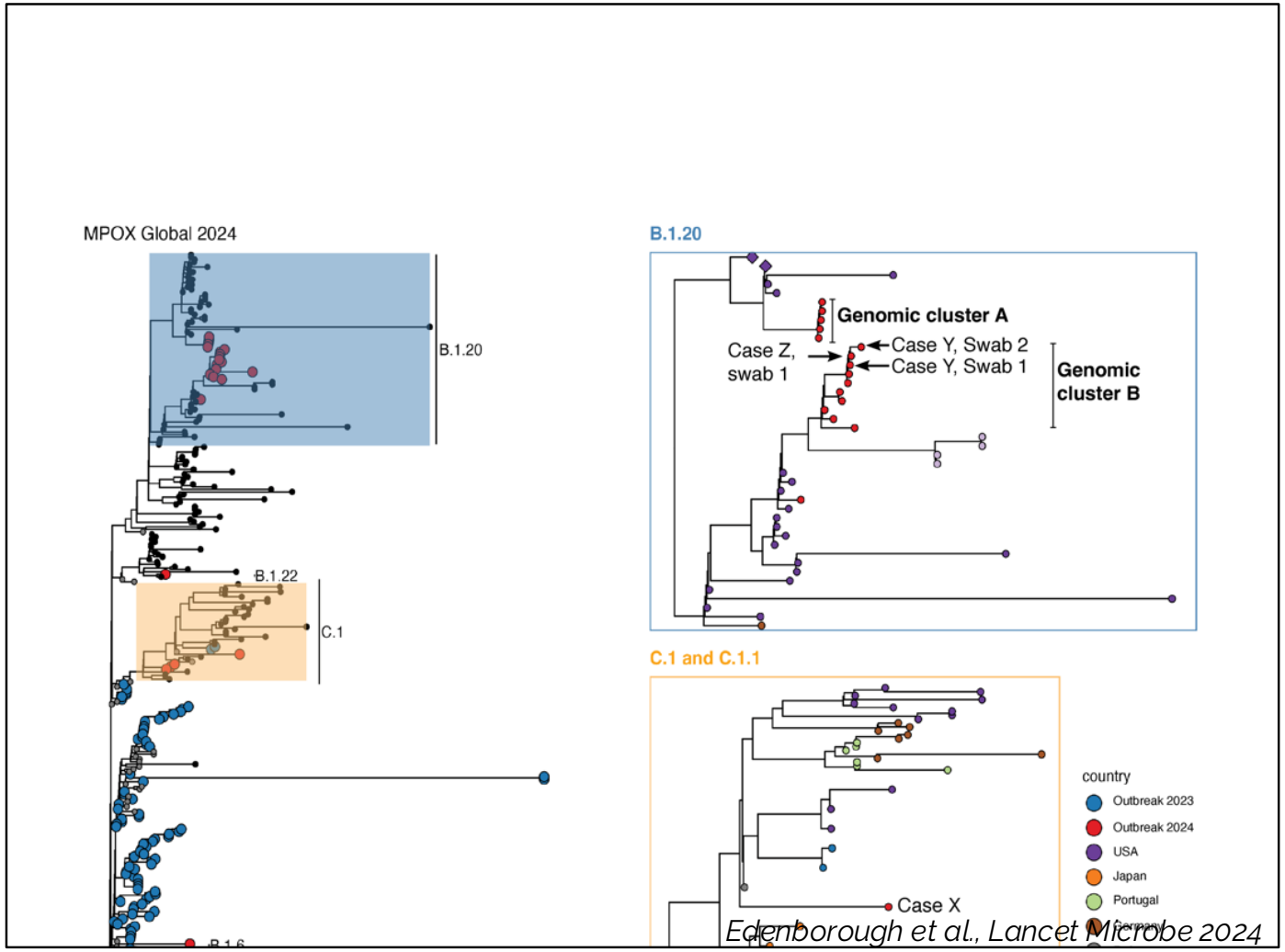
Integration of genomic data into established pathogen networks

- Viral pathogens will continue to cause Public Health Events of International Concern (PHEICs)
- WGS and genomic epidemiology increasingly integrated into established and emerging global networks focused on specific viruses:





Genomic data used to investigate emerging viral pathogens: Mpox Clade IIb

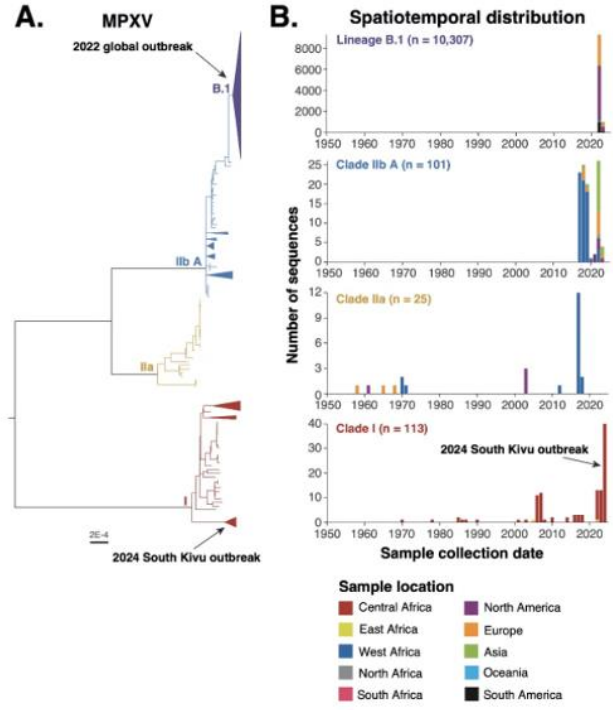


Edenborough et al., Lancet Microbe 2024

- In 2024, Mpox Clade IIb was reintroduced into Australia (Victoria) April 2024
- Five distinct Clade IIb sublineages detected amongst 19 cases
  - Multiple, concurrent introductions
- Limited global genomic surveillance
- Considerable diversity between sequences obtained from same case
  - Conventional genomic epi analysis very challenging

Genomic data used to investigate emerging viral pathogens: Mpox Clade Ib

Availability of genomic sequencing data



~50% from North America

~10,500 sequences on GISAID in total

653 genomes generated in 2024 in GISAID across all clades and countries

76 clade Ib genomes on GISAID

Otieno J et al, 2024, medRxi
https://doi.org/10.1101/2024.08.15.2431203

World Health Organization

HEALTH EMERGENCIES programme

Conclusions

- Translational public health genomics is a field that has grown exponentially since the COVID-19 pandemic
- How genomic data are generated and interpreted differs between viral and bacterial pathogens
- Genomics is increasingly utilised for routine surveillance and outbreak investigation of viral pathogens both locally and internationally
- Epidemiological context is essential to ensure accurate interpretation of genomic data
- Next up: introduction to dengue virus