September 2024
Issue 63

The Broad Street Pump

A Centre for Infectious Diseases and Microbiology - Public Health (CIDM-PH) and Sydney Institute for Infectious Diseases (Sydney ID) publication

Combined visualization of genomic and epidemiological data for outbreaks

Carl J.E. Suster[*1,2,3], *Anne E. Watt* [2,3,4], *Qinning Wang* [1,4], *Sharon C.-A. Chen* [1,2,3,4], *Jen Kok* [1,2,3,4], *Vitali Sintchenko* [1,2,3,4]

Affiliations

1. Centre for Infectious Diseases and Microbiology – Public Health, Westmead Hospital, Westmead, New South Wales, Australia
2. Sydney Infectious Diseases Institute, The University of Sydney, Sydney, New South Wales, Australia
3. School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
4. Centre for Infectious Diseases and Microbiology Laboratory Services, Institute of Clinical Pathology and Medical Research, NSW Health Pathology, Westmead, New South Wales, Australia

https://doi.org/10.1101/2024.04.02.24305229

In epidemiological investigations, pathogen genomics can provide insights and test epidemiological hypotheses that would not have been possible through traditional epidemiology. Tools to synthesize genomic analysis with other types of data are a key requirement of genomic epidemiology. We propose a new 'phylepic' visualization that combines a phylogenomic tree with an epidemic curve. The combination visually links the molecular time represented in the tree to the calendar time in the epidemic curve, a correspondence that is not easily represented by existing tools. Using an example derived from a foodborne bacterial outbreak, we demonstrated that the phylepic chart communicates that what appeared to be a point-source outbreak was in fact composed of cases associated with two genetically distinct clades of bacteria. We provide an R package implementing the chart. We expect that visualizations that place genomic analyses within the epidemiological context will become increasingly important for outbreak investigations and public health surveillance of infectious diseases.

Inside this issue

Dr Carl Suster
CIDM-PH Postdoctoral Fellow
carl.suster@health.nsw.gov.au

CIDM centre for infectious diseases & microbiology Public Health
NSW GOVERNMENT Health Western Sydney Local Health District
THE UNIVERSITY OF SYDNEY

For investigations of infectious disease outbreaks, high-resolution characterisation and clustering of pathogen genomes using whole genome sequencing (WGS) has proven to be a powerful complementary tool to epidemiological investigations [1,2]. A epidemiological investigation will often involve establishing a case definition to demarcate the outbreak, geospatial and temporal clustering, case follow-up and further investigations to establish transmission pathways or common reservoirs [3]. Genomic epidemiology draws together bioinformatic analysis from pathogen genomic sequences and epidemiological data to provide context and support to inferences of transmission. Synthesis of genomic and epidemiological data is required to perform these inferences, however this can be challenging given the complexity of the underlying data [4]. These challenges can be compounded when genomic and epidemiological investigations are conducted by discrete organisations (for example public health laboratories and local public health services respectively) with different information contexts and objectives.

The epidemic curve, a histogram of cases binned along the time axis, is a key representation relied upon in disease control. It is straightforward to prepare from line listing data, supports both data exploration and statistical epidemiology [5], and is straightforward to interpret. Genomic epidemiology instead relies upon phylogenomic trees, which display a detailed summary of the relationships between the included genomes under the particular assumptions of the underlying evolutionary models. Phylogenomic trees are prepared using specialized tools that implement the evolutionary models and efficiently search for the optimal tree (e.g. IQ-TREE, http://www.iqtree.org/). Trees are an essential representation of the salient information contained in a set of genomes that support genomic epidemiology and public health surveillance. Visually extracting this information relies upon knowledge of the conventions of tree drawing, and some familiarity with the process of their preparation. Users of trees may therefore require dedicated training so as to minimize the risk of misinterpretation.

We propose a new visualization called the 'phylepic' chart that synthesizes epidemic curves and phylogenomic trees. The structure of the phylepic chart facilitates visual linkage of the molecular time represented in the tree with the epidemiological time represented in the epidemic curve [6]. For illustration, we adapted a dataset from the investigation of a foodborne point-source outbreak in New South Wales (NSW), Australia. Bacterial isolates were collected during the outbreak period and underwent WGS. The resulting sequences were analysed alongside historical sequences collected in NSW using conventional bioinformatic pipelines to produce a core genome phylogenomic tree. For the purpose of illustration, we have pruned the resulting phylogeny and added random noise to the dates associated with each isolate. The chart is presented in Figure 1.

Relying on the epidemiological data alone, the silhouette of the epidemic curve in Figure 1 is consistent with a single point-source outbreak. The epidemic curve is coloured by genomic cluster membership. The presence of two genomic clusters suggests that there are two distinct strains associated with cases occurring in the same period. The phylogenomic tree provides more complete genomic context of the outbreak by providing detailed information about the inferred relatedness of bacterial sequences. The tree reveals that the most recent common ancestor of the two outbreak clusters (Clust-04 and Clust-05) is also the common ancestor of several other clusters and singletons in the historical dataset. This provides evidence that Clust-04 and Clust-05 arose from separate introductions of the pathogen. From the tree alone the outbreak is not evident as there is no representation of the temporal proximity of cases. Only the combination of the two lines of evidence leads to the conclusion of distinct introductions.

The calendar grid in the lower right of Figure 1 shows a visual projection of the tree onto a time axis representing the date. For this particular chart the time axis was restricted to the outbreak period, meaning that for isolates collected before mid-August there is no calendar tile. To distinguish these from cases with missing metadata, triangles are drawn at the edges of the calendar to indicate out-of- range values. The epidemic curve and calendar grid are binned by epidemiological week. The phylepic chart is most readable when the tree is relatively small and there are not too many date bins, both of which are reasonable restrictions for the intended application in outbreak investigation. The bin width can be decreased to a single day or increased to another interval depending on the appropriate resolution for an investigation.

While we have chosen a foodborne bacterial pathogen, phylepic charts are equally compatible with any other pathogens such as respiratory viruses. The tree need not necessarily be built from the consensus genome or a core genome as in our example; it is sufficient that a tree of some sort is prepared that is meaningful in the context of the investigation. It remains important to convey any caveats and limitations of this analysis alongside the results to ensure correct interpretation.

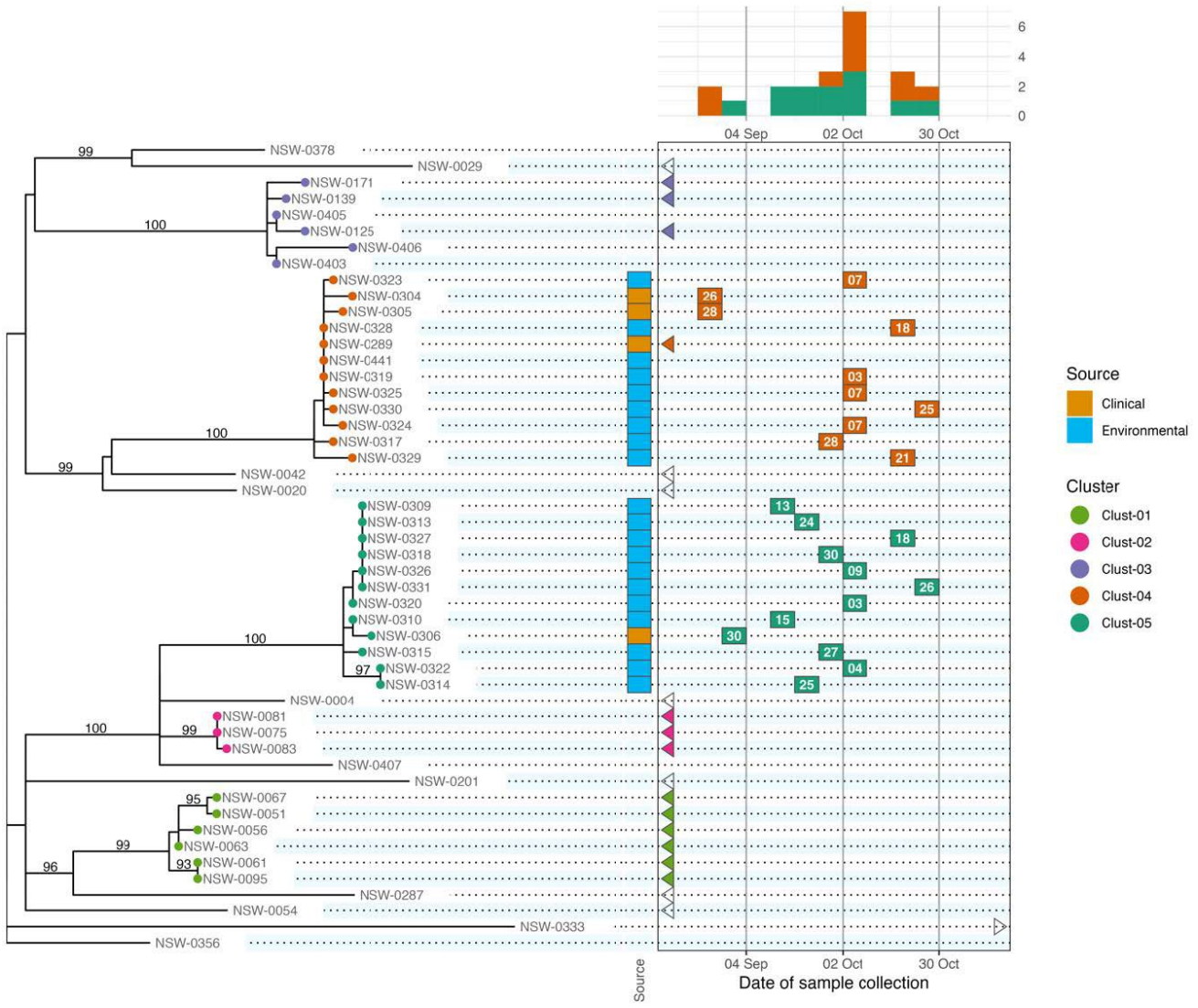

Figure 1. *A phylepic chart using an illustrative dataset adapted from a foodborne enteric pathogen outbreak in NSW. The panel on the left shows the phylogenomic tree with tips colours by genomic cluster membership. Numbers on branches are bootstrap support values expressed as a percentage. The panel on the upper right shows the epidemic curve using date of sample.*

To communicate relevant epidemiological context, most popular tools for visualizing trees (e.g. the ggtree R package, https://yulab-smu.top/treedata-book/, or the ETE toolkit Python library, http://etetoolkit.org/) allow tips to be annotated with an array of coloured boxes or symbols. This can work well for categorical data or binary such as the isolation source of samples, results of laboratory assays, or presence of antimicrobial resistance, toxins, or virulence factors. These annotations can be incorporated into the phylepic chart, as shown with the coloured tiles indicating clinical and environmental sample sources in Figure 1. Additional columns can be incorporated to represent other categorical or quantitative data as required. With some care, it is possible to use existing software libraries such as those mentioned above to produce a chart similar to Figure 1.

Dates associated with each case (reporting, sample collection, symptom onset) are key data for generating epidemiological hypotheses. These dates do not directly relate to the branch lengths in phylogenomic trees. Instead, the axis extending from the root of the tree to its tips can be thought of as molecular time, reflecting the evolutionary distances predicted by the model. It is possible to express branch lengths in units of time using an inferred molecular clock rate, and by incorporating temporal information as explicit constraints in the model [7]. With a time-scaled tree of this sort, an alternative visualization is made possible where an epidemic curve is vertically aligned with the tree [6]. This can be helpful to show how population-level dynamics over a longer period correspond to structural features of the tree, however it is less suitable for high resolution outbreak investigation, since the dates associated with individual cases are not shown.

We have implemented the phylepic chart in an R package. Our package uses the ggplot2 library [8] as a framework, the ggraph library (https://ggraph.data-imaginist.com) for drawing the tree, and the cowplot library (https://wilkelab.org/cowplot) for aligning panels. We have designed the package so that minimal code is required to implement simple visualizations such as Figure 1, while full customisation and annotation of the plots is possible using the underlying frameworks. It takes as input a prebuilt phylogeny (in any commonly used tree format such as Newick) and tabular data containing at minimum a column of dates associated with each sequence in the tree. Compared to existing software libraries for annotating phylogenetic trees, the package implements routines for date handling (including automatic binning by week) and ensuring data matching and consistency across panels.

As genomic epidemiology becomes a routine part of public health outbreak investigations, effective integration of genomic and epidemiological evidence to support public health and clinical responses remains a key challenge. Communication and visualization tools will assist in conveying complex genomic data into investigation and response contexts [9,10]. Our experience with using visualizations similar to phylepic charts in public health reporting has shown that they can help epidemiologists and other public health professionals to navigate phylogenomic trees by relating genomic features to more familiar details of the cases under investigation.

References

1. Armstrong GL, et al. Pathogen Genomics in Public Health. New England Journal of Medicine Massachusetts Medical Society, 2019; 381: 2569–2580.

2. Baker KS. Microbe hunting in the modern era: reflecting on a decade of microbial genomic epidemiology. Current Biology 2020; 30: R1124–R1130.

3. World Health Organization. Framework and toolkit for infection prevention and control in outbreak preparedness, readiness, and response at the national level. Geneva: World Health Organization, 2021.

4. Watt AE, et al. State-wide genomic epidemiology investigations of COVID-19 in healthcare workers in 2020 Victoria, Australia: Qualitative thematic analysis to provide insights for future pandemic preparedness. The Lancet Regional Health - Western Pacific 2022; 25: 100487.

5. Nishiura H. Methods to Determine the End of an Infectious Disease Epidemic: A Short Review. In: Chowell G, Hyman JM, eds. Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases. Cham: Springer International Publishing, 2016, p. 291–301.

6. Sintchenko V, Holmes EC. The role of pathogen genomics in assessing disease transmission. BMJ 2015; 350: h1314–h1314.

7. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum likelihood phylodynamic analysis. Virus Evolution 2018; 4: vex042.

8. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2nd ed. 2016. Cham: Springer International Publishing: Imprint: Springer, 2016.

9. Arnott A, et al. Documenting elimination of co-circulating COVID-19 clusters using genomics in New South Wales, Australia. BMC Research Notes 2021; 14: 415.

10. Argimón S, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microbial Genomics Microbiology Society, 2016; 2: e000093.

Staff Profile

Dr Rong Liu

Dr Rong Liu

CIDM-PH Health Economics Research Fellow

rong.liu@health.nsw.gov.au

Dr. Rong Liu joined the Centre for Infectious Disease and Microbiology - Public Health (CIDM-PH) as a Health Economics Research Fellow in August 2024, following the completion of her PhD at the George Institute for Global Health, University of New South Wales. Her doctoral research focused on the efficacy, safety, drivers, implementation strategies, feasibility, and cost-effectiveness of seasonal influenza vaccination for individuals with or at risk of cardiovascular diseases, particularly in China. Her findings demonstrated that influenza vaccination significantly reduces hospitalizations among people with cardiovascular diseases and identified a potentially cost-effective strategy to improve vaccine coverage among heart failure patients in China. Her PhD was funded by the Australian Government Research Training Program.

Dr. Liu holds a Master's degree in Global Health from the University of Washington and a Bachelor's degree in Preventive Medicine from Zhengzhou University. She has previously worked at a government think tank, a clinical research organization, and a virtual hospital, where she was involved in program evaluations, cardiovascular clinical trials, and economic evaluations. Her research interests include infectious triggers of chronic diseases, vaccination, intervention strategies, program evaluation, and economic evaluation. Passionate about enhancing health system efficiency and population health globally, Dr. Liu will leverage CIDM-PH's expertise in infectious diseases and advanced diagnostics to assess the cost-effectiveness of disease control and prevention measures for informed decision-making.

The evolving epidemiology of mpox in Africa, New South Wales and Australia following the global outbreak of 2022.

Rebecca Rockett [1,2,3,4], Jen Kok [1,2,3,4], Maurizio Stefani [4]

1. Centre for Infectious Diseases and Microbiology – Public Health, Westmead Hospital, Westmead, New South Wales, Australia
2. Sydney Infectious Diseases Institute, The University of Sydney, Sydney, New South Wales, Australia
3. School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
4. Centre for Infectious Diseases and Microbiology Laboratory Services, Institute of Clinical Pathology and Medical Research, NSW Health Pathology, Westmead, New South Wales, Australia

Mpox was first declared a Public Health Emergency of International Concern (PHEIC) by the World Health Organization (WHO) on 23 July 2022 after monkeypox virus (MPXV) infections were reported in multiple countries outside Africa, including Australia.[1] The epidemiologic features of this mpox outbreak from clade IIb (lineage B.1) viruses were different compared to previous clade I viruses as disease was spreading predominantly in men who have sex with men (MSM), rather than children or from zoonotic spillovers. Due to the subsequent decline in cases and no evidence of severe mortality from this outbreak, the PHEIC was rescinded on 11 May 2023. In parallel with the international experience, there was a decline of mpox cases in Australia from 144 cases in 2022 to 26 cases in 2023, which also led to the rescindment of mpox as a Communicable Disease Incident of National Significance (CDINS) on 25 November 2022, after it was first declared on 26 July 2022. The incidence of mpox cases in New South Wales (NSW) during this outbreak was also low, with 54 and 12 cases notified in 2022 and 2023, respectively.[1]

As reported in the publication by Stefani et al (available at: *https://www.pathologyjournal.rcpa.edu.au/article/S0031-3025(24)00217-4/pdf*),[2] during the period of 1 May 2022 to 5 September 2023, there were frequent detections of both herpes simplex virus (HSV) and varicella zoster virus (VZV) at our laboratory in samples where MPXV testing was requested, using an in-house developed multiplex assay targeting MPXV, HSV and VZV. In 2022, HSV and VZV were detected in 13.5% and 9.8% of samples, which is comparable to the MPXV detection rate of 12.5%. In 2023, there was a marked reduction in the requests for MPXV testing, and MPXV, HSV and VZV were detected in 0.9%, 17.5% and 13.7% of samples, respectively. This highlights the importance of considering HSV and VZV infections in persons with suspected mpox, particularly when the incidence of mpox was low.

Although the cases of mpox substantially decreased outside Africa after 2022, cases continued to rise within Africa, with 90% of cases occurring in the Democratic Republic of Congo (DRC). More recently, mpox was declared a PHEIC again by the WHO on 14 August 2022 due to the continued upsurge of cases in the DRC and its immediate neighbouring countries, in particular Burundi. Although there was ongoing mpox infection from clade I viruses in DRC after 2022, there has been a substantial increase in cases since the last quarter of 2023. There has also been the emergence of clade Ib viruses in South and then North Kivu provinces in the DRC, resulting in sustained human to human transmission. Clade Ib viruses mainly affect adults with transmission between household, heterosexual and healthcare contacts; and has higher mortality compared to clade IIb (lineage B.1) viruses (0.7% vs 0.19%). There have also been two cases of clade Ib viruses reported outside Africa (one each in Sweden and Thailand in August 2024). Mutations within the APOBEC3 region of the MPXV genome is thought to be responsible for the observed increased human to human transmission.[3]

Epidemiological and genomic data on clade Ib infections continue to emerge, the latter limited by the number of samples that are characterised genomically in Africa. Although the role of whole genome sequencing for MPXV in understanding disease severity, transmission, vaccine and antiviral efficacy is clear, its utility from a public health perspective continues to be debated. There has been a substantial rise in the number of mpox cases in Australia since April 2024 (572 cases as at 15 September 2024, of which 300 are in NSW).[1] Although there are no reported cases of clade Ib mpox infections in Australia at present, it is essential to ensure surveillance efforts are maintained to detect these viruses. Further global efforts are needed to limit the spread of mpox within Africa and internationally, including sustained diagnostic testing, vaccine availability and improved healthcare access.

References

1. Australian Government Department of Health and Aged Care. National Notifiable Disease Surveillance System. (available at: https://nindss.health.gov.au/pbi-dashboard/)

2. Stefani M, Ellem J, Jeoffreys N, et al. Frequent detection of herpes simplex virus and varicella zoster virus in samples submitted for monkeypox virus testing in New South Wales, Australia during the mpox outbreak 2022-2023. Pathology 2004; published ahead of print (available at: https://www.pathologyjournal.rcpa.edu.au/article/S0031-3025(24)00217-4/pdf).

3. Vakaniaki EH, Kacita C, Kinganda-Lusamaki E, et al. Sustained human outbreak of new MPXV clade I lineage in eastern Democratic Republic of Congo. Nat Med 2004; (available at: https://doi.org/10.1038/s41591-024-03130-3).

CONTACT US

CIDM-PH

Centre for Infectious Diseases and Microbiology Public Health

Westmead Hospital, ICPMR Building,
Level 3, Cnr Hawkesbury & Darcy Roads,
Westmead NSW 2145

PHONE:

(612) 8890 9870

WEBSITE:

www.wslhd.health.nsw.gov.au/CIDM-PH

E-MAIL:

WSLHD-CIDM-PH@health.nsw.gov.au

SYDNEY ID

Sydney Institute for Infectious Diseases

Westmead Hospital, Block K, Level 5

Cnr Hawkesbury & Darcy Roads
Westmead NSW 2145

WEBSITE:

www.sydney.edu.au/infectious-diseases-institute

E-MAIL:

infectious.diseases@sydney.edu.au

UPCOMING EVENTS

Sydney ID Annual Colloquium

New Horizons in Infectious Disease:
Vaccines, Advanced Therapeutics and Innovative Research IDeas

Thursday, 31 October 2024

12.45pm – 6pm (AEST)

University of Sydney, Camperdown

REGISTER HERE

CIDM-PH Annual Colloquium

Friday, 29 November 2024

9am – 4pm (AEST)

Westmead Education & Conference Centre, Westmead Hospital

PROGRAM & REGISTRATION AVAILABLE SOON

CIDM centre for infectious diseases & microbiology Public Health

THE UNIVERSITY OF SYDNEY