

Tracking outbreak dynamics using within-host variation

All viral populations within a host comprise dominant and minor genomic variants. Although most attention is focused on dominant genomic variants (that is, lineage-defining changes), minor variants can also be transmitted between hosts and contribute to viral evolution. These underlying patterns of within-host variation can paint a detailed epidemiological picture of chains of transmission across small outbreaks. Dominant and minor genomic variants can be analysed and used to compare related samples using next-generation sequencing, which provides high coverage and good resolution. When combined with sufficient epidemiological data, the analysis of within-host variants allows researchers to robustly infer the dynamics of viral transmission. Reporting in *Virus Evolution* in 2021, San and colleagues used bioinformatic methods combined with genomic and epidemiological data to examine within-host diversity and its role in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission dynamics. We selected this paper to highlight key work done by a team in South Africa during the COVID-19 pandemic that demonstrates best practice in genomics studies, and exemplifies the benefits of doing science locally rather than sending samples abroad.

Much of the focus during the SARS-CoV-2 pandemic has been on tracking lineage emergence and antigenic changes in the SARS-CoV-2 genome to guide the public-health response to surges in cases. This was usually conducted using consensus (dominant variant) sequence data, which can be handled at scale and allows for real-time analyses. However, within-host variation has an important role in the evolution of viruses such as influenza. Due to the tremendous sequencing effort, there have been many studies describing the same for SARS-CoV-2. San and colleagues revisited two nosocomial outbreaks in Kwazulu-Natal province of South Africa, which they had previously characterized by epidemiological and phylogenetic methods, and re-sequenced samples taken at

the time of the outbreak using the ARTIC amplicon method with Illumina sequencing. The authors chose to re-examine these outbreaks to gain a greater understanding of the selection pressures acting on virus populations, in the hope that it would help to prevent similar surges in infections in the future.

Unlike previous analyses of these two outbreaks, the authors sequenced two biological replicates, to assess potential biases introduced by the sequencing steps and to measure contamination. This level of technical rigour ensured a higher level of certainty in the identification of minor variants, which are generally more sensitive to error than the dominant variants used to build phylogenetic trees. The authors next applied bottleneck analysis (estimating the number of virus particles transmitted to a new host) to their validated dataset to characterize SARS-CoV-2 transmission dynamics.

Their analysis leverages a mathematical model that accounts for variant calling thresholds (detection method) and the stochastic nature of viral replication dynamics in both the source and recipient (what is being detected). This is deemed superior to the more common 'mutation counting' methods that fail to account for variant calling errors. The authors confirmed that their methods improved the tracking of transmission events during outbreaks. Their combined within-host diversity and bottleneck analyses improve the resolution of outbreak transmission events between hosts and provide evidence that transmission of minor variants is common in SARS-CoV-2. Additionally, their analysis shed light on transmission events in the second outbreak that were not identified by phylogenetic analysis alone.

The approach developed by San and colleagues has clear advantages but is not without caveats. Limitations include the difficulty of distinguishing between the transmission of minor variants and recurrent mutation of the same minor variant in independent hosts. Also, the elimination of false-positive variants can result in a loss of true variants from the data.

Using the newly characterized within-host transmission events, the authors reconstructed outbreak dynamics, including details of previously established epidemiological transmission links that are not supported by the new within-host variant analysis. The authors obtained better resolution by providing insights into both chains of infection and directions of transmission. The authors used a minor variant minimum frequency threshold of 3%, which produced low bottleneck estimates consistent with the biological characteristics of SARS-CoV-2. The authors' approach is supported by other reports in this area of research (M. A. Martin & K. Koelle. *Sci. Transl. Med.* **13**, eabh1803; 2021), showing that bottleneck estimates are sensitive to the minor variant minimum frequency thresholds used to determine minor variant contribution to a transmission event.

Overall, San and colleagues identify an effective use of the SARS-CoV-2 mutational landscape that, when integrated with bottleneck analyses, can retrospectively track viral transmission chains. This case study could lead to improvements in the resolution of within-host variations and future outbreak investigations. The authors nicely used local collaborations and networks to provide an illuminating insight into transmission dynamics that could be repeated the world over, given the resources.

I'ah Donovan-Banfield^{1,2}✉ & Edward Cunningham-Oakes^{1,3}✉

¹Department of Infection Biology and Microbiomes, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK. ²NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Liverpool, UK.

³NIHR Health Protection Research Unit in Gastrointestinal Infections, Liverpool, UK.

✉ e-mail: I.Donovan-Banfield@liverpool.ac.uk; Edward.Cunningham-Oakes@liverpool.ac.uk

Original article: San, J. E. et al. Transmission dynamics of SARS-CoV-2 within-host diversity in two major hospital outbreaks in South Africa. *Virus Evol.* **7**, veab041 (2021).

Acknowledgements

I.D.-B. is affiliated to the National Institute for Health and Care Research Health Protection Research Unit (NIHR HPRU)

in Emerging and Zoonotic Infections at the University of Liverpool in partnership with the UK Health Security Agency (UKHSA), in collaboration with Liverpool School of Tropical Medicine and the University of Oxford. I.D.-B. is based at the University of Liverpool. The views expressed are those of the author(s) and not necessarily those of the NIHR, the

Department of Health and Social Care or the UK Health Security Agency. E.C.-O. is affiliated to the NIHR HPRU in Gastrointestinal Infections at the University of Liverpool in partnership with the UKHSA, in collaboration with the University of Warwick. E.C.-O. is based at the University of Liverpool. The views expressed are those of the author(s) and

not necessarily those of the NIHR, the Department of Health and Social Care or the UKHSA.

Competing interests

The authors declare no competing interests.