

Genome sequencing and its use in public health responses to COVID-19

John-Sebastian Eden

Centre for Virus Research, The Westmead Institute for Medical Research, Westmead, NSW 2145, Australia;
and Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney Medical School, The University of Sydney,
Sydney, NSW 2006, Australia. Tel.: +61 2 8627 1817; Email: js.eden@sydney.edu.au

Abstract. Human history has been shaped by the heavy burden of infectious disease pandemics. Yet, despite the bitter lessons learned from history, even those in living memory such as the 1918 influenza pandemic and HIV/AIDS epidemic, COVID-19 stands unique in the sudden, immense health and economic impacts to the global human population. While the costs have been great, and the long-term consequences are still being revealed, the urgent need for action has also brought forward rapid developments and innovations to combat COVID-19 and better prepare us for future infectious disease outbreaks. One such area has been the widespread adoption of whole genome sequencing to inform public health responses. Genome sequencing during the COVID-19 pandemic has become key to tracking the spread of SARS-CoV-2 at all scales, to such a degree that terms such as genomics, mutations, variants and clusters are now common vernacular to politicians, health officials and the general public. This article provides a commentary on the genesis and evolution of SARS-CoV-2 genome sequencing, and its critical on-going role in the public health response to the COVID-19 pandemic.

Received 22 February 2021, accepted 10 March 2021, published online 9 April 2021

On 11 January 2020, Professor Eddie Holmes of the University of Sydney posted a tweet that provided the world with the first publicly available genome of SARS-CoV-2. The sequence shared by Professor Holmes came from the epicentre of the initial outbreak in Wuhan, with the work performed by a team led by Professor Zhang Yongzhen at the Shanghai Public Health Clinical Center, pioneers of RNA sequencing methods for pathogen discovery¹. While we do not yet know when the virus first entered the human population, the sharing of this genome represents the moment when the clock started ticking in the race to develop the tools needed to effectively respond to this unprecedented disease event. Within days, new diagnostic tests for SARS-CoV-2 were validated and shared² and the first designs of a number of vaccines that are now available for use entered early-stage testing³. The enormous, immediate impact of sharing this data highlights the wealth of information encoded in pathogen genomes, particularly for understanding their origins and potential to cause disease. Genomes have continued to be sequenced in order to monitor the genetic diversity and variation that has accumulated, allowing for the tracing of on-going evolution and the spread of the virus as the pandemic continues.

One year on, more than half a million genome sequences have been generated and shared on public databases including the Global Initiative on Sharing Avian Influenza Data (GISAID) and National Center for Biotechnology Information (NCBI) GenBank. This incredible number means that SARS-CoV-2 is already the most

sequenced virus in history (in terms of genome number), surpassing even influenza virus which has been under heavy genomic surveillance for decades. This is even more impressive given that sequencing virus genomes is not straightforward as viruses are rarely isolated and only make up a minuscule fraction of nucleic acid in a sample. RNA sequencing of viral genomes is possible, indeed it was the method by which SARS-CoV-2 was discovered and first sequenced. However, unbiased, direct sequencing approaches such as this generally lack sensitivity and are expensive. Moreover, they do not scale, as high sequencing depth is required as most of the RNA in a sample is derived from the host and the microflora present in the respiratory tract.

As an alternative, amplicon-based enrichments methods offer a simple and effective way to amplify the viral genome before library preparation and sequencing. One of the first and most widely used approaches is the ARTIC protocol that uses pools of 400 bp amplicons to tile across the virus genome. These are then sequenced using Oxford Nanopore Technologies (ONT) platforms such as the MinION (see <https://artic.network/ncov-2019>). The COVID-19 ARTIC protocol and its genome assembly software RAMPART were adapted from an approach developed during the West African Ebola virus epidemic (2013-16) that provided access to highly portable and near real-time genomic sequencing⁴. During the Ebola epidemic more than 5% of known cases were sequenced⁵, and these data were instrumental in understanding the transmission networks

fuelling the ongoing spread of the virus and ultimately controlling the outbreak. It was here that the potential of wide-spread genomic surveillance in public health responses was realised.

A number of other amplicon-based SARS-CoV-2 genome enrichment strategies have been developed with longer amplification products that are compatible with both ONT and Illumina platforms^{6,7}. The variety of methods and sequencing platforms used to generate the hundreds of thousands of publicly-available SARS-CoV-2 genomes reflects the local sequencing and bioinformatic capacity of public health and research labs. Despite this diversity, most sequencing protocols have been shown to be robust and reliable⁸. In Australia, the majority of SARS-CoV-2 genomic sequencing has been performed using Illumina-based platforms. This is partly due to the leveraging of the existing capacity for whole genome sequencing that has been developed in major microbiology public health labs such as the Centre for Infectious Diseases & Microbiology – Public Health at Westmead and the Microbiological Diagnostic Unit Public Health Laboratory in Melbourne. These programs were originally developed to investigate foodborne and nosocomial outbreaks, and for priority respiratory pathogens such as *Mycobacterium tuberculosis*. Importantly, this highlights how previous investment in pathogen genomics public health programs have proven to be invaluable in the effective responses to COVID-19 in Australia.

While genome sequencing offers important insights into the potential sources of viral infections and the relationships of viruses amongst infected individuals, the overall benefit to public health policy is greatest when genomics is integrated closely with diagnostic services and public health epidemiology teams. More often than not, genomics provides confirmatory data for infections where other information is already available, such as symptom onset, viral load, and travel and contact history. However, during the investigations of cases with unclear infection sources, genome sequencing may clarify the probable source of infection in cases where epidemiological links cannot be determined. The ability to link cryptic infections to established local clusters, or alternatively, ruling out cases linked to the cluster, is one the main benefits of genome sequencing. Compared to other parts of the world, community transmission has remained low in Australia, therefore a major priority for public health responses has been to identify all sources of local transmission. During the first wave of infections in Australia, and before the implementation of border control measures on 20 March 2020, the large number of returned travellers from countries across the globe meant that the diversity of SARS-CoV-2 strains was high, which aided in the identification of sources of local spread as multiple lineages and clusters were present and readily identifiable^{9,10}. While border control measures have remained

in-place and community transmission effectively eliminated, return travellers undergoing mandatory quarantining remain important sources of potential local transmission through infected quarantine workers. Large scale outbreaks from the virus escaping quarantine hotels have occurred and are an on-going concern particularly in the face of an ever-evolving virus.

In late 2020, a number of SARS-CoV-2 variants arose independently that have been found to carry novel and shared mutations in the spike protein that alter apparent transmissibility and/or antigenicity. These Variants of Concern (VOC) include the B.1.1.7, B.1.351 and P.1 viruses first described in the United Kingdom, South Africa and Brazil, respectively, and have had major impacts on seeding new epidemics even in areas with previously high seroprevalence¹¹. Furthermore, these VOCs have been spreading globally and replacing established lineages. There is particular concern for the B.1.351 VOC since several vaccines have been shown to have reduced efficacy against this variant¹². The value of genomic sequencing here has been firstly, in the initial identification of these VOCs through active surveillance programs, and secondly, in monitoring their prevalence moving forward. Following the roll-out of SARS-CoV-2 vaccines, genome sequencing has a vital role in monitoring for VOCs or any potential vaccine escape variants, and ultimately informing vaccine strain composition. While we are still learning of the prospects of long-term vaccine immunity to SARS-CoV-2, the rapid emergence of antigenically novel variants means it is likely that, similar to the influenza virus, future vaccines for SARS-CoV-2 will require updating in line with circulating diversity.

The COVID-19 pandemic has led to growth in viral genome sequencing and greater synergy between genomics and epidemiology to better inform public health responses. As a society, this is important as the burden of emerging and novel infectious diseases is one we are likely to carry for some time.

Conflicts of interest

The author declares no conflicts of interest.

Acknowledgements

This research did not receive any specific funding. The author thanks Drs Jen Kok and Bethany Horsburgh for constructive feedback on the manuscript.

References

1. Wu, F. *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269. doi:10.1038/s41586-020-2008-3
2. Corman, V.M. *et al.* (2020) Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* **25**, 2000045. doi:10.2807/1560-7917.ES.2020.25.3.2000045

3. Jackson, L.A. *et al.* (2020) An mRNA vaccine against SARS-CoV-2 – preliminary report. *N. Engl. J. Med.* **383**, 1920–1931. doi:[10.1056/NEJMoa2022483](https://doi.org/10.1056/NEJMoa2022483)
4. Quick, J. *et al.* (2016) Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232. doi:[10.1038/nature16996](https://doi.org/10.1038/nature16996)
5. Dudas, G. *et al.* (2017) Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* **544**, 309–315. doi:[10.1038/nature22040](https://doi.org/10.1038/nature22040)
6. Eden, J.-S. *et al.* (2020) An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. *Virus Evol.* **6**, veaa027. doi:[10.1093/ve/veaa027](https://doi.org/10.1093/ve/veaa027)
7. Freed, N.E. *et al.* (2020) Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biol. Methods Protoc.* **5**, bpaa014. doi:[10.1093/biomethods/bpaa014](https://doi.org/10.1093/biomethods/bpaa014)
8. Bull, R.A. *et al.* (2020) Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat. Commun.* **11**, 6272. doi:[10.1038/s41467-020-20075-6](https://doi.org/10.1038/s41467-020-20075-6)
9. Seemann, T. *et al.* (2020) Tracking the COVID-19 pandemic in Australia using genomics. *Nat. Commun.* **11**, 4376. doi:[10.1038/s41467-020-18314-x](https://doi.org/10.1038/s41467-020-18314-x)
10. Rockett, R.J. *et al.* (2020) Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nat. Med.* **26**, 1398–1404. doi:[10.1038/s41591-020-1000-7](https://doi.org/10.1038/s41591-020-1000-7)
11. Sabino, E.C. *et al.* (2021) Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **397**, 452–455. doi:[10.1016/S0140-6736\(21\)00183-5](https://doi.org/10.1016/S0140-6736(21)00183-5)
12. Wu, K. *et al.* (2021) Serum neutralizing activity elicited by mRNA-1273 vaccine – preliminary report. *N. Engl. J. Med.* doi:[10.1056/NEJMc2102179](https://doi.org/10.1056/NEJMc2102179)

Biography



Dr John-Sebastian Eden is a research scientist at the Westmead Institute for Medical Research and senior research fellow in the Sydney Medical School, University of Sydney. Dr Eden leads the Viromics research group that uses genomics to better understand the origins and evolution of human and zoonotic pathogens.

MICROBIOLOGY AUSTRALIA, 2021, 42, 46
https://doi.org/10.1071/MA20058_CO

Therapeutics for COVID-19: established and in development

Kasha P Singh, Joe Sasadeusz, Sharon R Lewin and Jennifer Audsley

The authors advise that on page 219 of their published article (*Microbiology Australia*, Volume 41, Issue 4, pages 217–223, doi:[10.1071/MA20058](https://doi.org/10.1071/MA20058)), in the first sentence under the heading ‘Repurposed drugs with cellular targets used for viral entry: ACE2 and TMRPSS2 inhibitors’, ‘acetylcholinesterase (ACE)-2 receptor’ should read ‘angiotensin-converting enzyme (ACE)-2 receptor’. The correct text is shown here:

SARS-CoV-2 uses the angiotensin-converting enzyme (ACE)-2 receptor for cell entry, and serine protease TMPRSS2 for S-protein priming, both potential targets for antiviral intervention³¹.

Also, the label for RdNP inhibitors is misplaced in Figure 2: it belongs at the region of RNA replication.

BE HEARD

We're working towards a more informed society.

Have your say now.

www.theasm.org.au