

OPINION PIECE

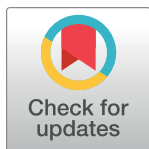
New rules for genomics-informed COVID-19 responses—Lessons learned from the first waves of the Omicron variant in Australia

Ashleigh F. Porter^{1*}, Norelle Sherry², Patiyan Andersson², Sandra A. Johnson², Sebastian Duchene¹, Benjamin P. Howden^{1,2}

1 Department of Microbiology and Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia, **2** Microbiological Diagnostic Unit Public Health Laboratory, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

☞ These authors contributed equally to this work.

* ashleigh.porter@unimelb.edu.au



OPEN ACCESS

Citation: Porter AF, Sherry N, Andersson P, Johnson SA, Duchene S, Howden BP (2022) New rules for genomics-informed COVID-19 responses—Lessons learned from the first waves of the Omicron variant in Australia. *PLoS Genet* 18(10): e1010415. <https://doi.org/10.1371/journal.pgen.1010415>

Editor: Dana C. Crawford, CWRU: Case Western Reserve University, UNITED STATES

Published: October 13, 2022

Copyright: © 2022 Porter et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the Medical Research Future Fund (MRF9200006) and the National Health and Medical Research Council Australia (Investigator Grant GNT1196103 to BPH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

During the COVID-19 pandemic, phylodynamics and phylogeography have been launched into the spotlight as tools to model the spread of the SARS-CoV-2 virus. In Australia, we have relied on genomic epidemiology (and associated derived parameters such as viral growth rate, reproductive number, and estimated sampling proportion) to inform public health policy changes [1]. This was possible due to the high proportion of SARS-CoV-2 cases sequenced in Australia throughout 2020 and 2021, where we maintained low burdens of both cases and deaths. The recent Omicron ‘waves’ experienced in Australia and globally, combined with the relaxation of public health restrictions, has seen a significant jump in Australia’s case numbers, rising to the top 10 globally in newly reported cases and deaths in August and September 2022 (<https://covid19.who.int/table>). With around 10-fold more cases per day in 2022 compared to the previous year, our sequencing strategy has had to adapt along with the virus. Here, we emphasise how in the ‘COVID-normal’ future, the way we sequence during high-case load settings can optimise our application of phylogenomic methods to sufficiently inform the COVID-19 pandemic response.

The rise of the Omicron variant

During the first stage of the pandemic (January 2020 to November 2021), a comprehensive genomics-informed response was possible in Australia, due to the low case numbers of COVID-19. The low case numbers also enabled one of the highest genome sequencing rates globally (up to 80% of Victorian cases sequenced in the “second wave” in 2020, [Fig 1](#)) [1]. However, the introduction of the Omicron variant in late 2021, coinciding with the lifting of restrictions, led to an exponential increase in cases ([Fig 1](#)). We anticipate there was a significant number of undiagnosed cases (e.g. PCR tests underestimating true transmission [3]), during the peak of the Omicron wave, due to the overwhelming demand for testing, which inundated the established diagnostic PCR testing systems and created a shortage of rapid antigen tests. Furthermore, half of the reported COVID-19 cases were being counted from positive rapid antigen tests ([Fig 1](#)), which were not able to be sequenced. The ongoing Omicron wave has had significant local and global impacts, underlining the necessity to prepare for future variants, including Omicron subvariants and other emerging variants of concern (VOCs) that are not manageable by current public health control methods.

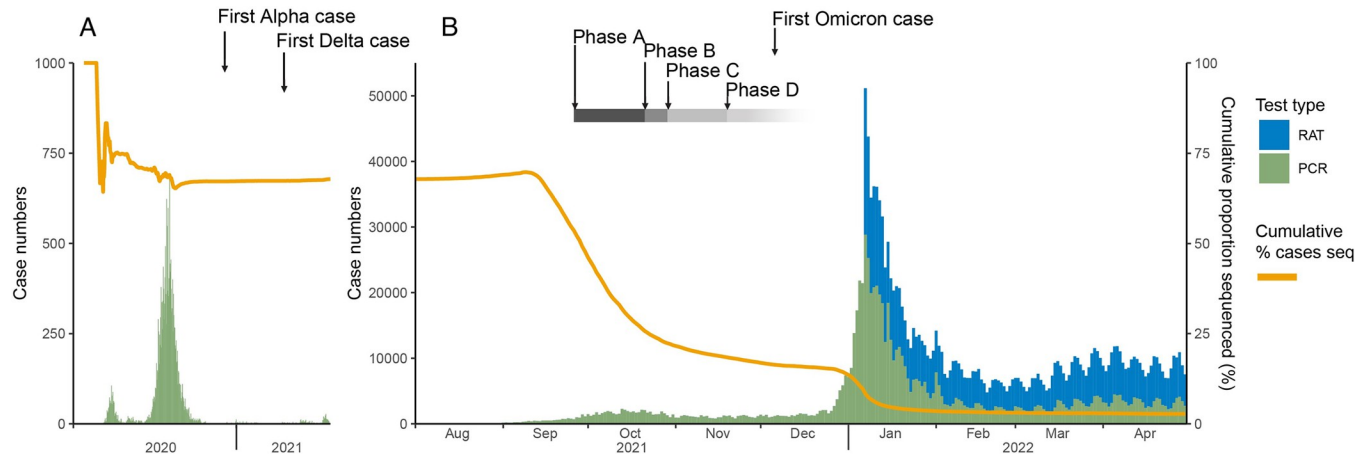


Fig 1. Epidemiological curve of SARS-CoV-2 in Australia, demonstrating the rapid rise in cases in 2022 upon emergence of the Omicron variant, and the diminishing cumulative proportion of cases sequenced. Panel A has a reduced scale to visualise the limited case numbers during January 2020 to July 2021 inclusive. Panel B includes data from August 2021 to April 2022 inclusive. The case numbers (Y-axis, left) obtained from PCR tests and rapid antigen tests (RAT) are respectively shown in green and blue. The cumulative percentage of cases sequenced (Y-axis, right) is visualised as an orange line. The emergence of variants of concern (Alpha, Delta and Omicron), as well as the 4 phases of the Victorian Government’s “Roadmap” [2] for easing COVID-19 restrictions, are shown at the top of the curve. The Victorian Roadmap is an example of jurisdiction-managed responses to the COVID-19 pandemic. The roadmap included four phases (A-D) that gradually eased restrictions on travel and social distancing measures, partly based on the percentage of the population fully vaccinated. Phase A focused on returning students to the classroom, whereas Phase B and C guided the return to work and travel upon reaching vaccination targets of 70% and 80% of the population (16+) fully vaccinated, respectively. The final stage, Phase D, was reached when 80% of the population (12+) was fully vaccinated, allowing the restrictions to ease and align with Australia’s National COVID-19 response.

<https://doi.org/10.1371/journal.pgen.1010415.g001>

Considering this, we discuss the ripple effect of the Omicron wave in Australia, as well as which sequencing strategies will be most effective for modelling the COVID-19 pandemic, both locally and globally. We highlight Australia’s focus on genomic epidemiology, a field focused on understanding the spread of a pathogen through a population by combining genomic data and epidemiological metadata. Specifically, we discuss phylodynamics, which utilises epidemiological and genomic data to explore how evolution and epidemiology drive phylogenetic patterns [4], and phylogeography, which focuses on the processes that determine the spatial distribution and spread of observed lineages [5]. Although estimates from these methods have had a large impact on pandemic response, there are three major issues Australia (and the rest of the world) will be facing:

- i. We cannot sustain our previous rates of sequencing, especially with highly transmissible VOCs and elevated case numbers.
- ii. Even if we could sequence every positive case, our models are unable to incorporate the full Australian dataset, let alone global sequence data.
- iii. Genome surveillance alone may not be sufficiently informative to produce meaningful epidemiological estimates.

Therefore, we argue here that *more genomes* do not necessarily mean *better results*.

As there has never been such an extensive pathogen genomics dataset (and such a high degree of public interest) to fully utilise genomic epidemiological methods, our major question for our long-term COVID-normal future is: how sustainable will it be to continue to generate this data, and more importantly, is it beneficial for the global response?

Sustainable sequencing to track the COVID-19 pandemic

Whilst public health and social measures, quarantine restrictions and vaccination have all been utilised in past and current pandemics, the COVID-19 pandemic is the first to employ

genomic sequencing on a massive global scale. Genomics has provided a major advantage to pandemic control by improving our understanding of the underlying transmission dynamics and evolution of SARS-CoV-2 in near-to-real time. Australia's initial approach to COVID-19 management involved strict non-pharmacological interventions (border closures, travel restrictions and social distancing measures) which resulted in elimination of the virus in Australia for several months between 2020 and 2021 (Fig 1). This tactic was coupled with coordinated diagnostic testing and sequencing efforts, resulting in approximately 50% of known cases nationally having a sequenced genome. In both Australia and New Zealand, where the proportion of cases sequenced has been substantial, we have been able to use the data at the "macro" level (studying global evolution of the virus, emergence of VOCs and informing public health policies) and at the "micro" level (inferring local transmission networks and the impact of public health interventions on genomic clusters) [1,6–9]. For example, genome data unequivocally traced most cases from a large outbreak in Victoria to a single hotel quarantine breach [1], leading to major policy changes in how Australia manages hotel quarantine facilities. Amendments to hotel quarantine included improvements to training (specifically for infection control practices), restriction of employment (to reduce community spread from employees working multiple jobs) and increased testing of employees. Furthermore, the impact of these changes can be observed, with a vast improvement to hotel quarantine escape risk in the period after policy change [1].

However, upon the easing of restrictions in late 2021 [1], the rapid spread of Omicron quickly overloaded the established testing capability and stretched the capacity of healthcare services. Due to the rapid rise in case numbers, the SARS-CoV-2 sequencing proportion was markedly reduced (Fig 1). Although a high-level national sequencing strategy was rapidly developed for Australia, the plan was dedicated to managing the public health response. The strategy aimed to balance sequencing of priority groups (determined by public health requirements) with surveillance of community infections, but implementation varied between jurisdictions. This was due to variable capacities, logistics and pandemic stage within jurisdictions (for example, one state did not experience community transmission until many months after the other jurisdictions).

In reconsidering our sequencing strategies and looking forward, we believe that the sequencing strategy could be further optimised from a modelling perspective to utilise our resources effectively [10]. Attempting to sequence the previously high proportion of cases is now unfeasible, and furthermore, we have found that "inconsequential sequences" (such as closely related sequences from a household-wide infection) provide diminishing returns for guiding the response. When looking towards the future, we will require a system that will sustainably sequence a proportion of positive COVID-19 cases during periods of high case numbers. Importantly, our strategy also needs to balance background, community-level "representative" sequencing along with "focused" sequencing (e.g. returning travellers), to ensure we are gathering the full diversity of lineages as new VOCs continue to emerge and spread (as seen with the sub-variants of the Omicron variant). We also note that SARS-CoV-2 sequencing strategies should *continue* to evolve as case numbers and the public health landscape changes, hence any sequencing strategies should continue to be re-evaluated during their implementation.

COVID-19 and the new era of genomic epidemiology

The COVID-19 pandemic has caused a major shift in the field of genomic epidemiology and surveillance, as reviewed previously [11–13]. In the last decade, increased computational power has enabled scientists to analyse large and complex viral datasets, such as HIV, Ebola,

influenza and Zika viruses [14–16]. However, the volume of genomic data generated for the COVID-19 pandemic has vastly outstripped past epidemics—for example, the current count of SARS-CoV-2 genomes is at least five orders of magnitude greater than the sequencing effort for decades-long history of past outbreaks, such as swine flu. Although complex phylogenetic models have been developed before the COVID-19 pandemic (such as those developed for Ebola surveillance) [17], these methods have not previously been applied to such a sizeable dataset.

The power of applying genomic epidemiological models to the SARS-CoV-2 dataset has recently been thoroughly reviewed [18], demonstrating how the combination of genomic data with epidemiological metadata, such as travel history [19] or healthcare associated infections [20], has allowed us to prepare appropriate outbreak responses in real-time. Phylogenetic modelling has been used to explore the transmission, spatial dispersal and epidemiology of SARS-CoV-2 [21], for example; it has been used to distinguish community transmission from novel importations in China [22], the emergence and spread of new variants in the USA and South Africa [23–25], the impact of public health interventions on emerging lineages in Brazil [26], and detailed outbreak transmission dynamics across Australia and New Zealand [1,6,7,9]. However, there are several major issues of applying these models to the SARS-CoV-2 global sequence dataset.

An overwhelmingly large and complicated dataset

Firstly, our models are computationally unable to incorporate all the sequence data available. This is complicated by a connected issue, the quality of the data. Although this is not unique to the SARS-CoV-2 dataset, data gathered from large public databases (such as GISAID) is notoriously poorly formatted (with formatting inconsistency and errors, along with missing metadata) and can range dramatically in sequence quality (i.e. large gaps that create issues in comparative analyses). All these factors can complicate downstream analysis, contributing to the infamous “bioinformatics bottleneck” [27].

Widespread sampling bias throughout dataset

There has been a noticeable bias in SARS-CoV-2 sampling, especially spatiotemporal bias [18,27,28]. The bulk of genomes sequenced are sourced from the UK, Europe and North America, which account for almost 90% of all genomes in GISAID. This sampling bias contributes to the difficulty of drawing comparisons from different models and spatiotemporal datasets [28]. Presently, there are no guidelines available for utilising the immense (and sometimes overwhelming) dataset, however, commonly used approaches to attempt to reduce sampling bias include subsampling the available genomes geographically and temporally [26] or using global diversity via the Nextstrain “backbone” [29].

The sampling bias present in the SARS-CoV-2 dataset is incompatible with most phylogeographic methods, including discrete trait ancestral reconstruction [30]. However, this has led to innovative solutions—the integration of additional information such as individual travel history, transportation data and epidemiological data into extended phylogeographic models. Importantly, the inclusion of this new information suggested alternative hypotheses not apparent using only genomic and geographic data [18]. Although these novel methods have proven powerful, the availability of the necessary metadata is not consistent—for example, even if case travel history metadata is collected, it is not shared on public databases.

Incompatibility between datasets

Finally, although it is tempting to explore the spatiotemporal spread of SARS-CoV-2, several complications arise when trying to review the “global” findings. Not only is the

epidemiological data dependent on each location (e.g. population size and structure), but the sampling proportion varies dramatically between countries. As an example of two countries using different approaches, teams from Brazil and the USA both combined genomic epidemiology with two different types of movement data (travel pattern analysis and flight data) [31,32]. However, the dataset that was used in these models would be biased by the fact that North America had a much higher sampling proportion (1.5%) compared to South America (0.3%). Also, critically, the location where the COVID-19 case was sequenced does not equate to location where it was acquired (a variable which would benefit greatly from collection and storage of travel history metadata on sequence-sharing platforms).

Generating results from different datasets with the same model does not necessarily mean we can draw comparisons—for example, it is not reasonable to contrast the reproductive number of the same variant (e.g. Omicron) circulating in two different locations, as differences in population dynamics (including population size, density, vaccination status, susceptibility due to previous VOC infection, and social distancing measures) will dramatically bias results. However, we can reduce this bias with thorough metadata collection (i.e. vaccination coverage). Ideally, each dataset requires a unique approach designed specifically for the hypothesis being tested, particularly for methodologies that involve a variety of elements, such as phylogeography. When looking forward, novel tools include the use of negative controls in discrete phylogeographic models [33], and the potential to use both sequence data and “non-sequenced” data (positive cases that did not undergo whole genome sequencing) [34].

How do we collect our “ideal” dataset?

As we have hinted at previously, even if we committed to collecting every sequence of a COVID-19 infection, the advantages gained from high-volume sequence data would presently not equal the investment. However, if we focused our resources towards collecting high quality data—both a sufficient level of “representative” and “focused” sequencing, and comprehensive metadata (Fig 2), this is where we might see real benefits to our downstream analysis. Based on a modelling perspective, we require a minimum of 10–100 sequences to inform a phylogenetic tree. To inform public health management of the COVID-19 pandemic, 10% of cases is sufficient to obtain meaningful phylogenetic resolution (due to the lack of genetic diversity within SARS-CoV-2 lineages). However, in high case-load scenarios (such as those we have seen in the Omicron wave), where it would be unreasonable to sequence such a high proportion of cases, we recommend focusing on representative sequencing of a smaller percentage of cases (1–2%) [35] (Fig 3).

With this dataset, we could extend past the basic genomic epidemiological methods and focus on the outputs from complex, informative phylodynamic and phylogeographic models. As seen in Fig 2, there are many other sources of data that can be gathered to benefit genomic epidemiology, such as patient metadata (vaccination status, treatment history). Additionally, we could benefit from early detection of outbreaks by setting up continuous surveillance systems, such as wastewater surveillance, serological surveys, and sampling zoonotic spillover events in wildlife, livestock and domestic animals. Not only will including these streams of data into our models reduce the necessity to sequence a large proportion of COVID-19 cases, but evidence suggests that combining sequence data and metadata could be more informative for informing public health measures [19,36].

Additionally, although there has been much work on COVID-19 research, one area of improvement we could all benefit from is *connections*. There are several elements involved in making progress towards a COVID-normal future (overviewed in Fig 2). Examples include breaking down barriers between organisations, building links between research and public

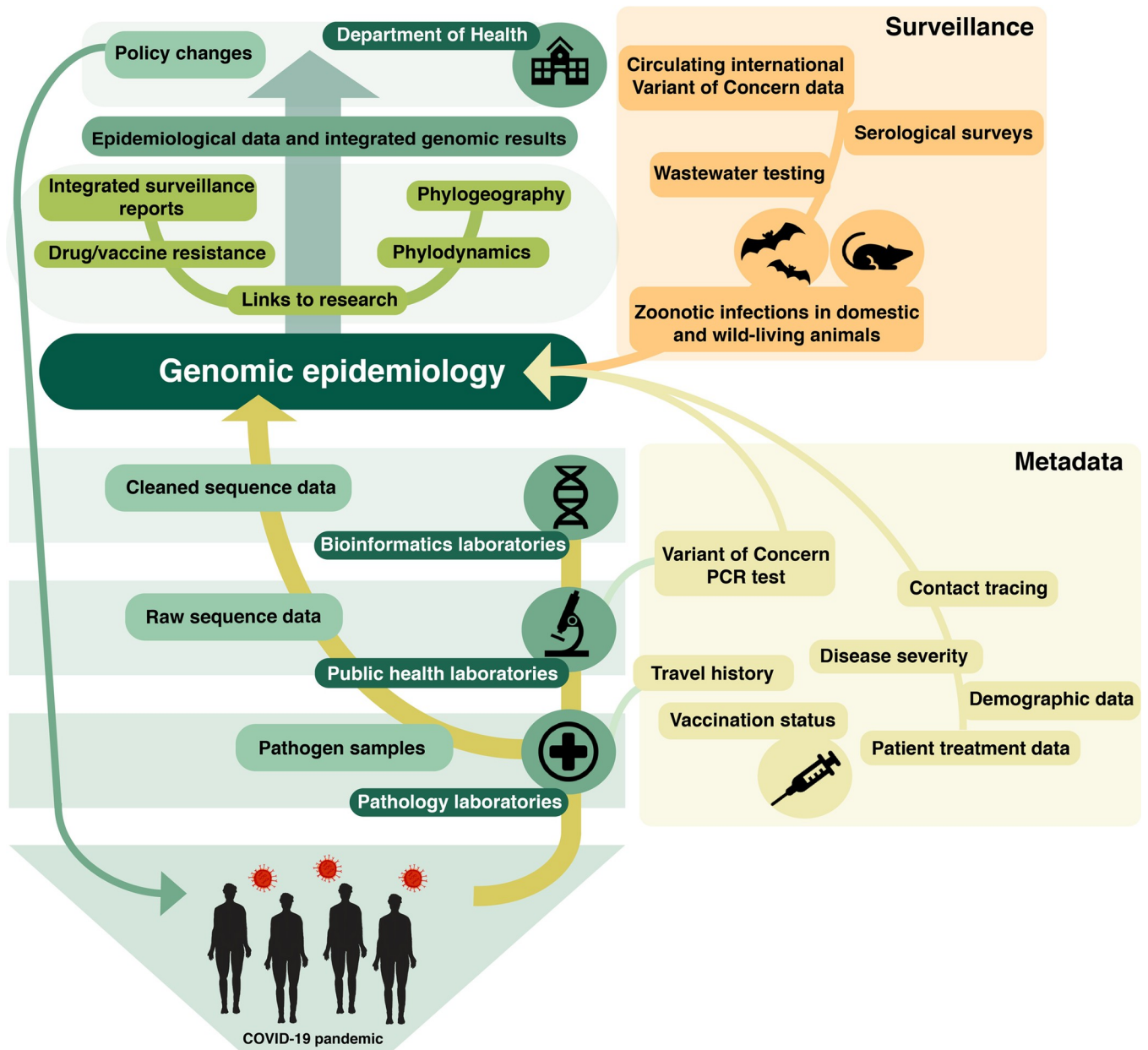


Fig 2. Visualisation of different sources of COVID-19 data, including sequence data and metadata, and how they feed into genomic epidemiology and link to research and policy making decisions. Here we highlight the flow and collaboration within the Australian COVID-19 pandemic response network, between public and private pathology, public and private pathology laboratories, bioinformatics laboratories, research groups and the health departments. The infographic on the left represents the pipeline of sequence data, sourced from samples collected from the public, feeding into genomic epidemiology and phylogenomic tools. The coloured bubbles on the right represent additional sources of data, such as epidemiological metadata (yellow bubbles) or global surveillance data (orange bubbles).

<https://doi.org/10.1371/journal.pgen.1010415.g002>

health groups, establishing governance mechanisms to facilitate data distribution, establishing effective communication between disciplines, and forming channels to share data [17]. Importantly, building these multidisciplinary connections would not come at a major cost—instead, we can build on the processes and platforms that have already been developed, which have been supported by investments that have already been made for managing the COVID-19 pandemic.

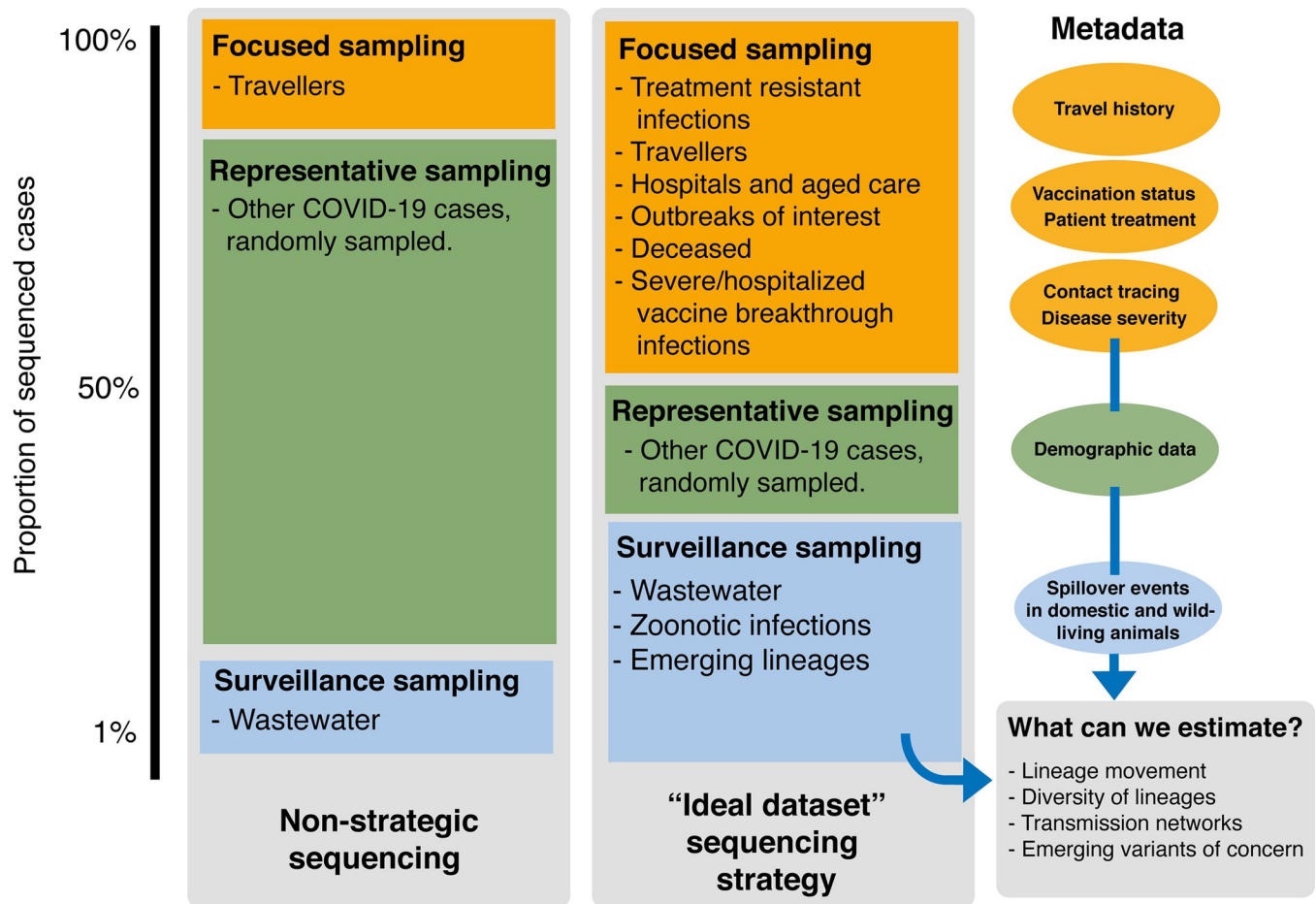


Fig 3. Illustration of the current “non-strategic” sequencing used for monitoring the COVID-19 pandemic, in comparison to the plan we have outlined as our “ideal dataset”. The two main streams of sequencing we have described, focused and representative sampling, are represented by the orange and green boxes, respectively. These streams are shown along with surveillance sampling, represented by the blue box. The size of each box represents the proportion of sequenced genomes being generated from each stream. The associated metadata with each stream is represented by a coloured bubble in the right panel. Following the blue arrows, we overview the parameters we can estimate from the “ideal dataset” sequencing strategy, combined with metadata.

<https://doi.org/10.1371/journal.pgen.1010415.g003>

Future directions for genomics-informed surveillance of COVID-19 and imminent pandemics

As we anticipate future “waves” to occur with the emergence of novel VOCs (as part of the continued evolution of SARS-CoV-2) we need to develop proactive strategies to optimally use genomics for managing the COVID-19 pandemic. Emerging VOCs may demonstrate increased transmissibility and immune evasion, including vaccine breakthrough as seen with the Omicron variant [37,38], presenting challenges for tracking cases and maintaining sequencing levels. New challenges are constantly developing, such as Omicron subvariants BA.2, BA.4 and BA.5 evolving the capability to re-infect despite vaccination or previous SARS-CoV-2 infection [39], and the rise of SARS-CoV-2 Omicron recombinants (a result of recombination via co-infection with multiple lineages) [40]. Furthermore, with global movement now returning to normal levels, nations must prepare for the circulation of a diversity of SARS-CoV-2 lineages and emerging VOCs, as well as multiple introductions from global locations. To dedicate our resources towards capturing the diversity of lineages, identifying VOCs,

and rapidly detecting new outbreaks, we need an effective plan to strategically gather data (both genomic sequencing and metadata, as seen in Fig 2).

Hubs and other organisational structures that support continual genomic sequencing for SARS-CoV-2 surveillance have been beneficial during the pandemic (as seen with UKHSA), and moving forward, supporting equitable access across the globe should be a priority. These “hubs” should include multidisciplinary teams, including scientists with backgrounds in microbiology, molecular biology, epidemiology and phylogenetics, as well as clinicians and public health experts. Although sequencing and analyses are performed in each Australian jurisdiction, a centralised platform has been established to share national sequences and limited metadata (e.g. quarantine status, travel history), termed AusTrakka [41]. This platform has proven especially useful for managing SARS-CoV-2 surveillance across Australia. Other countries could benefit from utilising this model locally, but we could all benefit from applying this model on a global scale. Importantly, these hubs would support equitable access to genomic sequencing, which is crucial for countries with limited resources.

Given the increasing complexity of the SARS-CoV-2 landscape and high case numbers, it is critical to establish and maintain consistent surveillance of SARS-CoV-2 on multiple fronts, including community sampling and wastewater detection. We should not ignore the potential for spillover and spillback events, along with the formation of viral reservoirs, in wild-living and domesticated animals. We emphasise a one health approach (and teams that reflect veterinary, epidemiological, and ecological knowledge) for surveillance of spillover events and potential reservoirs in animals (e.g. minks, mice and deer) [42–44]. In Australia, our wastewater surveillance system is at the forefront for early detection lineages or variants of concern, as seen with the identification of Omicron sub-variants BA.2, BA.4 and BA.5 in wastewater [45,46], alerting authorities to be on the lookout for a rise in cases and contributing to modelling to inform public health.

For surveillance of clinical samples, we need to consider the aims of sequencing in each setting to inform the local sequencing strategy, and utilise our resources most effectively for measuring the evolution and spread of the virus. In this current phase in Australia, SARS-CoV-2 sequencing primarily focuses on determining proportions of current variants, identifying new emerging variants, and identifying the introduction of new VOCs into the community. Part of the current strategy includes gathering data on disease severity to inform public health activities, including modelling. Secondary aims may include assessing the performance of diagnostic tests and drug therapies with new variants, and investigation of specific outbreaks or populations (e.g. healthcare-associated outbreaks or prolonged infections in immunosuppressed cases). We note that there is an inherent trade-off in the choice to focus on target populations versus the general community. Whilst over-representation of cases with severe disease provides more data about disease severity and healthcare utilisation, this focus reduces the sensitivity of detecting the emergence or introduction of new variants in the community, hence delaying the time to identification of these critical events.

To achieve the aims listed above, sequencing should ideally be continual, rather than sporadic, as “sequencing blitzes” only gather information from closely related sequences and provide no temporal overview. Instead, community-based genomic surveillance provides consistent coverage of SARS-CoV-2 evolution—both temporally and geographically [47]. Along with continual, strategic sequencing (overviewed in Fig 3), the collection of basic metadata is essential to the practical use of COVID-19 data, including sample collection date, symptom onset date, exposure site history, travel history, vaccination status, and previous COVID-19 history. This collection strategy will provide valuable information, (especially if shared on a data platform such as AusTrakka), particularly for outbreaks of interest (i.e. to focus on a hospital-associated outbreak).

We recommend two main streams of sequencing [10]:

1. Focused sampling: Confirmed cases (positive nucleic acid amplification tests) from target groups of interest (outbreaks, travellers, hospitalizations, aged care, deceased, severe/hospitalized vaccine breakthrough infections): sequence as much as possible and capture essential metadata on travel history, vaccination status, and/or hospital history.
2. Representative sampling: Confirmed cases (positive nucleic acid amplification tests): sequence randomly and collect relevant metadata to contextualise the background prevalence of genomic lineages circulating.

These two streams of sequencing will need to be combined with consistent surveillance (e.g. wastewater and zoonotic spillovers) and a strategy for collection of essential metadata (Fig 3). Furthermore, we recommend a central platform for integrated pathogen genomics epidemiology, exemplified by the use of AusTrakka in Australia [41].

A COVID-normal approach will require public health and research teams to focus on the emergence and transmission dynamics of SARS-CoV-2 lineages, along with associations to COVID-19 disease severity. Importantly, genomic epidemiology can assist with understanding the emergence and evolution of SARS-CoV-2 and transmission dynamics if sufficient sequence data and metadata is available. For example, as shown in Fig 3, past strategies for SARS-CoV-2 sequencing around the globe have closely resembled the left box, where most SARS-CoV-2 genomes have been collected from positive COVID-19 cases indiscriminately (excluding incoming travellers). Although there has been considerable genomic data collected in Australia, this approach, in conjunction with limited metadata, has hindered our ability to draw meaningful conclusions from the dataset as case numbers rise. However, we propose the application of the streams of sequencing outlined above, which would provide sufficient information to estimate many important parameters, even in scenarios where COVID-19 case-loads are high (Fig 3).

We can also supplement samples that were unable to be sequenced with supporting information, such as metadata (i.e. results from a VOC PCR test), to inform downstream analysis (Fig 2). With this strategy, we are confident that despite sequencing a smaller proportion of overall COVID-19 cases, the additional streams metadata will provide sufficient information to draw informative estimates (Fig 3).

Much of the long-term COVID-normal future will be informed by our ability to exploit genomic epidemiology through gathering data about SARS-CoV-2 (both at the sequence and metadata level) and *sharing* it. We believe a global, coordinated response for data collection and modelling will be essential, both for the ongoing COVID-19 pandemic and future infectious disease outbreaks.

References

1. Lane CR, Sherry NL, Porter AF, Duchene S, Horan K, Andersson P, et al. Genomics-informed responses in the elimination of COVID-19 in Victoria, Australia: an observational, genomic epidemiological study. *Lancet Public Health*. 2021; 6(8):e547–e56. Epub 2021/07/13. [https://doi.org/10.1016/S2468-2667\(21\)00133-X](https://doi.org/10.1016/S2468-2667(21)00133-X) PMID: 34252365; PubMed Central PMCID: PMC8270762.
2. Victorian Government. Australia. Victoria's Roadmap: Delivering the National Plan, 2021 [cited 2022 Aug 25]. Available from: <https://www.premier.vic.gov.au/sites/default/files/2021-09/210919%20-%20RoadmapDelivering%20the%20National%20Plan.pdf>.
3. Tosif S, Haycroft ER, Sarkar S, Toh ZQ, DL A-h, Donato C, et al. Virology and immune dynamics reveal high household transmission of ancestral SARS-CoV-2 strain. 2022.
4. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *science*. 2004; 303(5656):327–32. <https://doi.org/10.1126/science.1090727> PMID: 14726583

5. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS computational biology*. 2009; 5(9):e1000520. <https://doi.org/10.1371/journal.pcbi.1000520> PMID: 19779555
6. Seemann T, Lane CR, Sherry NL, Duchene S, Goncalves da Silva A, Caly L, et al. Tracking the COVID-19 pandemic in Australia using genomics. *Nat Commun*. 2020; 11(1):4376. Epub 2020/09/03. <https://doi.org/10.1038/s41467-020-18314-x> PMID: 32873808; PubMed Central PMCID: PMC7462846.
7. Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray K-A, et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nature medicine*. 2020; 26(9):1398–404. <https://doi.org/10.1038/s41591-020-1000-7> PMID: 32647358
8. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol*. 2020; 6(2):veaa061. Epub 2020/11/26. <https://doi.org/10.1093/ve/veaa061> PMID: 33235813; PubMed Central PMCID: PMC7454936.
9. Geoghegan JL, Ren X, Storey M, Hadfield J, Jelley L, Jefferies S, et al. Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. *Nat Commun*. 2020; 11(1):6351. Epub 2020/12/15. <https://doi.org/10.1038/s41467-020-20235-8> PMID: 33311501; PubMed Central PMCID: PMC7733492.
10. Communicable Diseases Genomics Network. CDGN, PHLN and CDNA Sampling Strategy for SARS-CoV-2 Gnostic Surveillance. 2021 [cited 2022 Aug 25]. Available from: https://www.health.gov.au/sites/default/files/documents/2021/11/cdgn-phln-and-cdna-sampling-strategy-for-sars-cov-2-genomic-surveillance_0.pdf.
11. Knyazev S, Chhugani K, Sarwal V, Ayyala R, Singh H, Karthikeyan S, et al. Unlocking capacities of genomics for the COVID-19 response and future pandemics. *Nature Methods*. 2022; 19(4):374–80. <https://doi.org/10.1038/s41592-022-01444-z> PMID: 35396471
12. Nelson MI, Thielen P. Coordinating SARS-CoV-2 genomic surveillance in the United States. *Virus Evolution*. 2021; 7(2):veab053.
13. Black A, MacCannell DR, Sibley TR, Bedford T. Ten recommendations for supporting open pathogen genomic analysis in public health. *Nature medicine*. 2020; 26(6):832–41. <https://doi.org/10.1038/s41591-020-0935-z> PMID: 32528156
14. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS pathogens*. 2014; 10(2):e1003932. <https://doi.org/10.1371/journal.ppat.1003932> PMID: 24586153
15. Hong SL, Dellicour S, Vrancken B, Suchard MA, Pyne MT, Hillyard DR, et al. In Search of Covariates of HIV-1 Subtype B Spread in the United States—A Cautionary Tale of Large-Scale Bayesian Phylogeography. *Viruses*. 2020; 12(2). Epub 2020/02/09. <https://doi.org/10.3390/v12020182> PMID: 32033422; PubMed Central PMCID: PMC7077180.
16. Grubaugh ND, Saraf S, Gangavarapu K, Watts A, Tan AL, Oidtman RJ, et al. Travel surveillance and genomics uncover a hidden Zika outbreak during the waning epidemic. *Cell*. 2019; 178(5):1057–71. e11. <https://doi.org/10.1016/j.cell.2019.07.018> PMID: 31442400
17. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature*. 2016; 530(7589):228–32. <https://doi.org/10.1038/nature16996> PMID: 26840485
18. Hill V, Ruis C, Bajaj S, Pybus OG, Kraemer MUG. Progress and challenges in virus genomic epidemiology. *Trends in Parasitology*. 2021; 37(12):1038–49. <https://doi.org/10.1016/j.pt.2021.08.007> PMID: 34620561
19. Lemey P, Hong SL, Hill V, Baele G, Poletto C, Colizza V, et al. Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nature communications*. 2020; 11(1):1–14.
20. Meredith LW, Hamilton WL, Warne B, Houldcroft CJ, Hosmillo M, Jahun AS, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *The Lancet Infectious Diseases*. 2020; 20(11):1263–71. [https://doi.org/10.1016/S1473-3099\(20\)30562-4](https://doi.org/10.1016/S1473-3099(20)30562-4) PMID: 32679081
21. Attwood SW, Hill SC, Aanensen DM, Connor TR, Pybus OG. Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. *Nature Reviews Genetics*. 2022. <https://doi.org/10.1038/s41576-022-00483-8> PMID: 35459859
22. Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, et al. Genomic epidemiology of SARS-CoV-2 in Guangdong province, China. *Cell*. 2020; 181(5):997–1003. e9. <https://doi.org/10.1016/j.cell.2020.04.023> PMID: 32359424
23. Washington NL, Gangavarapu K, Zeller M, Bolze A, Cirulli ET, Schiabor Barrett KM, et al. Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell*. 2021; 184(10):2587–94. e7. <https://doi.org/10.1016/j.cell.2021.03.052> PMID: 33861950

24. Happi AN, Ugwu CA, Happi CT. Tracking the emergence of new SARS-CoV-2 variants in South Africa. *Nature Medicine*. 2021; 27(3):372–3. <https://doi.org/10.1038/s41591-021-01265-1> PMID: 33723453
25. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*. 2020:2020.12.21.20248640. <https://doi.org/10.1101/2020.12.21.20248640>
26. Resende PC, Delatorre E, Gräf T, Mir D, Motta FC, Appolinario LR, et al. Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil. *Frontiers in Microbiology*. 2021;11. <https://doi.org/10.3389/fmicb.2020.615280> PMID: 33679622
27. Hodcroft EB, Zuber M, Nadeau S, Vaughan TG, Crawford KHD, Althaus CL, et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv: the preprint server for health sciences*. 2021:2020.10.25.20219063. <https://doi.org/10.1101/2020.10.25.20219063> PMID: 33269368.
28. Mavian C, Marini S, Prosperi M, Salemi M. A snapshot of SARS-CoV-2 genome availability up to April 2020 and its implications: data analysis. *JMIR public health and surveillance*. 2020; 6(2):e19170. <https://doi.org/10.2196/19170> PMID: 32412415
29. Tay JH, Porter AF, Wirth W, Duchene S. The emergence of SARS-CoV-2 variants of concern is driven by acceleration of the evolutionary rate. *medRxiv*. 2021:2021.08.29.21262799. <https://doi.org/10.1101/2021.08.29.21262799>
30. Ishikawa SA, Zhukova A, Iwasaki W, Gascuel O. A fast likelihood method to reconstruct and visualize ancestral scenarios. *Molecular biology and evolution*. 2019; 36(9):2069–85. <https://doi.org/10.1093/molbev/msz131> PMID: 31127303
31. Fauver JR, Petrone ME, Hodcroft EB, Shioda K, Ehrlich HY, Watts AG, et al. Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. *Cell*. 2020; 181(5):990–6.e5. <https://doi.org/10.1016/j.cell.2020.04.021> PMID: 32386545
32. Candido DS, Claro IM, De Jesus JG, Souza WM, Moreira FR, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*. 2020; 369(6508):1255–60. <https://doi.org/10.1126/science.abd2161> PMID: 32703910
33. Dellicour S, Lequime S, Vrancken B, Gill MS, Bastide P, Gangavarapu K, et al. Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework. *Nature communications*. 2020; 11(1):1–11.
34. Featherstone LA, Di Giullonardo F, Holmes EC, Vaughan TG, Duchêne S. Infectious disease phylodynamics with occurrence data. *Methods in Ecology and Evolution*. 2021; 12(8):1498–507.
35. European Centre for disease Prevention and Control. Guidance for representative and targeted genomic SARS-CoV-2 monitoring, 2021 [cited 2022 Aug 25]. Available from: <https://www.ecdc.europa.eu/en/publications-data/guidance-representative-and-targeted-genomic-sars-cov-2-monitoring#:~:text=For%20monitoring%2C%20ECDC%20recommends%20sequencing,or%202.5%25%20to%205%25>.
36. Lindsey BB, Villabona-Arenas CJ, Campbell F, Keeley AJ, Parker MD, Shah DR, et al. Characterising within-hospital SARS-CoV-2 transmission events using epidemiological and viral genomic data across two pandemic waves. *Nat commun*. 2022; 13(671)
37. Chen J, Wang R, Gilby NB, Wei G-W. Omicron (B. 1.1. 529): Infectivity, vaccine breakthrough, and antibody resistance. *ArXiv*. 2021.
38. Andrews N, Stowe J, Kirsebom F, Toffa S, Rieckard T, Gallagher E, et al. Effectiveness of COVID-19 vaccines against the Omicron (B. 1.1. 529) variant of concern. *MedRxiv*. 2021.
39. Cao Y, Yisimayi A, Jian F, Song W, Xiao T, Wang L, et al. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *bioRxiv*. 2022:2022.04.30.489997. <https://doi.org/10.1038/s41586-022-04980-y> PMID: 35714668
40. England P. SARS-CoV-2 variants of concern and variants under investigation in England. *Tech Brief*. 2022;39.
41. Hoang T, da Silva AG, Jennison AV, Williamson DA, Howden BP, Seemann T. AusTrakka: Fast-tracking nationalized genomics surveillance in response to the COVID-19 pandemic. *Nature Communications*. 2022; 13(1):865. <https://doi.org/10.1038/s41467-022-28529-9> PMID: 35165271
42. Munnink BBO, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science*. 2021; 371(6525):172–7. <https://doi.org/10.1126/science.abe5901> PMID: 33172935
43. Kuchipudi SV, Surendran-Nair M, Ruden RM, Yon M, Nissly RH, Nelli RK, et al. Multiple spillovers and onward transmission of SARS-Cov-2 in free-living and captive White-tailed deer (*Odocoileus virginianus*). *bioRxiv*. 2021:2021.10.31.466677. <https://doi.org/10.1101/2021.10.31.466677>

44. Griffin BD, Chan M, Taylor N, Mendoza EJ, Leung A, Warner BM, et al. SARS-CoV-2 infection and transmission in the North American deer mouse. *Nature Communications*. 2021; 12(1):3612. <https://doi.org/10.1038/s41467-021-23848-9> PMID: 34127676
45. ABC. New Omicron COVID subvariant, BA.4 or BA.5, found in Australia in Melbourne wastewater, 2022 [cited 2022 7 June]. Available from: <https://www.abc.net.au/news/2022-04-15/new-covid-omicron-variant-ba4-ba5-detected-australia/100994610>.
46. ABC. COVID-19 Omicron subvariant BA.2.12.1 detected in Australia in Victorian wastewater, 2022 [cited 2022 7 June]. Available from: <https://www.abc.net.au/news/2022-04-30/covid-19-omicron-subvariant-ba-2-12-1-detected-in-victoria/101027818>.
47. Lythgoe KA, Golubchik T, Hall M, House T, MacIntyre-Cockett G, Fryer H, et al. Lineage replacement and evolution captured by the United Kingdom Covid Infection Survey. medRxiv. 2022:2022.01.05.21268323. <https://doi.org/10.1101/2022.01.05.21268323>