

A collaborative approach to improving representation in viral genomic surveillance

Paul Y. Kim^{1*}, Audrey Y. Kim¹, Jamie J. Newman², Eleonora Cella³, Thomas C. Bishop⁴, Peter J. Huwe⁵, Olga N. Uchakina⁵, Robert J. McKallip⁵, Vance L. Mack⁶, Marnie P. Hill⁶, Ifedayo Victor Ogungbe⁷, Olawale Adeyinka⁷, Samuel Jones⁸, Gregory Ware⁹, Jennifer Carroll⁹, Jarrod F. Sawyer⁹, Kenneth H. Densmore⁹, Michael Foster², Lescia Valmond¹, John Thomas¹, Taj Azarian³, Krista Queen⁹, Jeremy P. Kamil^{9,10*}

OPEN ACCESS

Citation: Kim PY, Kim AY, Newman JJ, Cella E, Bishop TC, Huwe PJ, et al. (2023) A collaborative approach to improving representation in viral genomic surveillance. PLOS Glob Public Health 3(7): e0001935. <https://doi.org/10.1371/journal.pgph.0001935>

Editor: Srinivasa Rao Mutheneni, CSIR-Indian

Institute of Chemical Technology, INDIA

Received: February 28, 2023

Accepted: June 5, 2023

Published: July 19, 2023

Copyright:© 2023 Kim 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.

Data Availability Statement: The data that support the findings of this study are publicly available from GISAID with the identifier(s) doi.org/10.55876/gis8.221011tg and doi.org/10.55876/gis8.220901cu.

Funding: The work reported in this publication was supported by The Rockefeller Foundation under grant number 2021 HTH 010 GA-S to PYK, JJN, JPK. PYK acknowledges the Office of the Provost at GSU for funding. JJN acknowledges the School of Biological Sciences at Tech for funding. JPK
RESEARCH ARTICLE

1 Department of Biological Sciences, Grambling State University, Grambling, LA, United States of America, **2** School of Biological Sciences, Louisiana Tech University, Ruston, LA, United States of America, **3** Burnett School of Biomedical Sciences, University of Central Florida, Orlando, FL, United States of America, **4** Physics and Chemistry Programs, Louisiana Tech University, Ruston, LA, United States of America, **5** Mercer University School of Medicine, Macon, GA, United States of America, **6** Mercer Medicine, Macon, GA, United States of America, **7** Department of Chemistry, Jackson State University, Jackson, MS, United States of America, **8** Health Services Center, Jackson State University, Jackson, MS, United States of America, **9** Center of Excellence for Emerging Viral Threats, Louisiana State University Health Shreveport, Shreveport, LA, United States of America, **10** Department of Microbiology and Immunology, Louisiana State University Health Shreveport, Shreveport, LA, United States of America

* kimp@gram.edu (PYK); jeremy.kamil@lsuhs.edu (JPK)

Abstract

The lack of routine viral genomic surveillance delayed the initial detection of SARS-CoV-2, allowing the virus to spread unfettered at the outset of the U.S. epidemic. Over subsequent months, poor surveillance enabled variants to emerge unnoticed. Against this backdrop, long-standing social and racial inequities have contributed to a greater burden of cases and deaths among minority groups. To begin to address these problems, we developed a new variant surveillance model geared toward building 'next generation' genome sequencing capacity at universities in or near rural areas and engaging the participation of their local communities. The resulting genomic surveillance network has generated more than 1,000 SARS-CoV-2 genomes to date, including the first confirmed case in northeast Louisiana of Omicron, and the first and sixth confirmed cases in Georgia of the emergent BA.2.75 and BQ.1.1 variants, respectively. In agreement with other studies, significantly higher viral gene copy numbers were observed in Delta variant samples compared to those from Omicron BA.1 variant infections, and lower copy numbers were seen in asymptomatic infections relative to symptomatic ones. Collectively, the results and outcomes from our collaborative work demonstrate that establishing genomic surveillance capacity at smaller

academic institutions in rural areas and fostering relationships between academic teams and local health clinics represent a robust pathway to improve pandemic readiness.

additionally acknowledges support from the U.S. National Institutes of Health under grants numbers P20GM121307-04S1 and P20GM134974; from Fast Grants, a part of Emergent Ventures at the Mercatus Center, George Mason University, under Fast Grant #2239. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. 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.

Introduction

In late 2019, a newly emergent betacoronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was identified as the etiological agent associated with an outbreak of viral pneumonia in Wuhan, China [1, 2]. The outbreak quickly grew to become a global public health emergency, now commonly referred to as the COVID-19 pandemic. Between November 2020 and January 2021, the first variants of concern (VOC) Alpha, Beta, and Gamma, were detected in England, South Africa, and Brazil and Japan, respectively. All three carried convergent spike (S) N501Y substitutions, while Beta and

Gamma shared substitutions S:K417N/T and S:E484K. It is now appreciated that SARS-CoV-2 variants can be more transmissible [3], differ in pathogenicity [4, 5], and/or escape immunity afforded by vaccination or by infection with earlier variants [6, 7]. Although the Omicron lineage has displaced Delta [8], new Omicron sublineages continue to emerge [9], certain of which harbor concerning combinations of S substitutions that escape neutralizing antibodies, e.g., F486V, F486P, R346T, K444T, N460K, F490S [10–16]. After multiple waves of Omicron descendent lineages, the emergency phase of the pandemic has wound down, but COVID-19 remains largely uncontrolled in virtually all geographies. Despite a greatly reduced mortality rate, the virus remains a threat, particularly to older individuals, and moreover, continues to evolve [17]. Ongoing microbial genomic surveillance is necessary not only to promptly identify new SARS-CoV-2 variants, but also to detect and inform public health responses to other microbial threats, such as influenza viruses, monkeypox virus, countless pre-emergent bat-borne viruses, and antimicrobial resistance in bacteria [18–22].

Starting with the sharing on 10 Jan 2020 of the first “novel coronavirus 2019” genome [23], scientists have closely tracked the evolution of SARS-CoV-2. Viral genomic surveillance activities sprung up to sequence the genomes of circulating viruses, usually from patient nasal swabs, sharing their data via GISAID (www.gisaid.org) [24], and/or the international nucleotide sequence database collaboration (INSDC) [25]. These practices collectively enable scientists around the world to track viral evolution across space and time while keeping tabs on mutations that can compromise the utility of diagnostic assays or negatively impact the clinical efficacy of vaccines or monoclonal antibody therapeutics. Effective surveillance relies on prompt viral whole genome sequencing of recent samples across representative geographies and populations, as well as rapid, ‘low latency,’ sharing of resulting data. Accordingly, investments in viral whole genome sequencing have increased greatly during the pandemic, with vastly more whole genome sequences of SARS-CoV-2 being shared than any virus in history.

At the outset of the pandemic, SARS-CoV-2 genome sequencing efforts were patchy at best and in many geographies non-existent [26–28]. Although scientists in the U.K. and South Africa swiftly established exemplary viral genomic surveillance programs, in the U.S., SARS-CoV-2 genome data initially came in large part from ad hoc efforts that often relied on discretionary funds or support from private foundations and philanthropies [29–31]. The emergence in late 2020 of the first variants of concern prompted the U.S. Centers for Disease Control and Prevention and the National Institutes of Health to invest in the development of large-scale nationwide genomic surveillance. Despite this, many states in the Midwest and Southern U.S. remain poorly represented in the genomic surveillance data [32], even though their residents have been disproportionately impacted by COVID-19 [33–35]. This ruralurban disparity was compounded during the pandemic for racial and ethnic minorities [36, 37] who already bore a greater burden of disease [38]. The failure to adequately track emerging variants in a representative fashion may have exacerbated existing disparities in allocation of public health resources, such as rapid antigen tests and monoclonal antibody therapies.

To address this lack of representation, Grambling State University, Louisiana Tech University, and Louisiana State University Health Shreveport established a viral genomic surveillance hub to reach several vulnerable and underserved populations in north Louisiana

[39]. Soon after, we partnered with Mercer University School of Medicine in Georgia, Jackson State University in Mississippi, as well as with community health centers in each of our areas to increase SARS-CoV-2 genome sequencing volume and representativeness. These efforts resulted in a new viral genomic surveillance network serving a total of 16 counties or parishes across these 3 southern U.S. states (Fig 1).

Our primary goals were to improve sampling from rural communities and minorities, to empower underrepresented minorities conducting the scientific work of viral genomic surveillance, and to provide individuals who donate samples for viral genome sequencing the opportunity to do so with consent and an understanding of how their samples would be used to track viral variants. Most viral genomic surveillance programs sequence viral genomes from patient swabs under a ‘medical waste’ exemption, without obtaining the patient’s consent. In participatory surveillance projects, however, community members and volunteers collaborate to collect data, such as reporting weekly influenza symptoms [40]. Echoing this concept, at our project sites in Louisiana we educated patients about viral genomic surveillance and variant tracking, and provided them the opportunity to decide whether or not to donate their samples for sequencing.

The COVID-19 pandemic has also underscored the importance of contextual metadata accompanying genomic data. Incomplete metadata confounds comparative analyses and impedes data-driven policymaking [41, 42]. Thus, an additional objective was to collect more complete metadata to maximize the usefulness of the shared sequences. We used self-reported patient metadata to assess relationships between viral gene copies (as an index of viral load) and variant type, vaccination status, and symptomatology.

Here, we summarize the results of our work to improve representation and expand viral genomic surveillance in Louisiana, Georgia, and Mississippi.

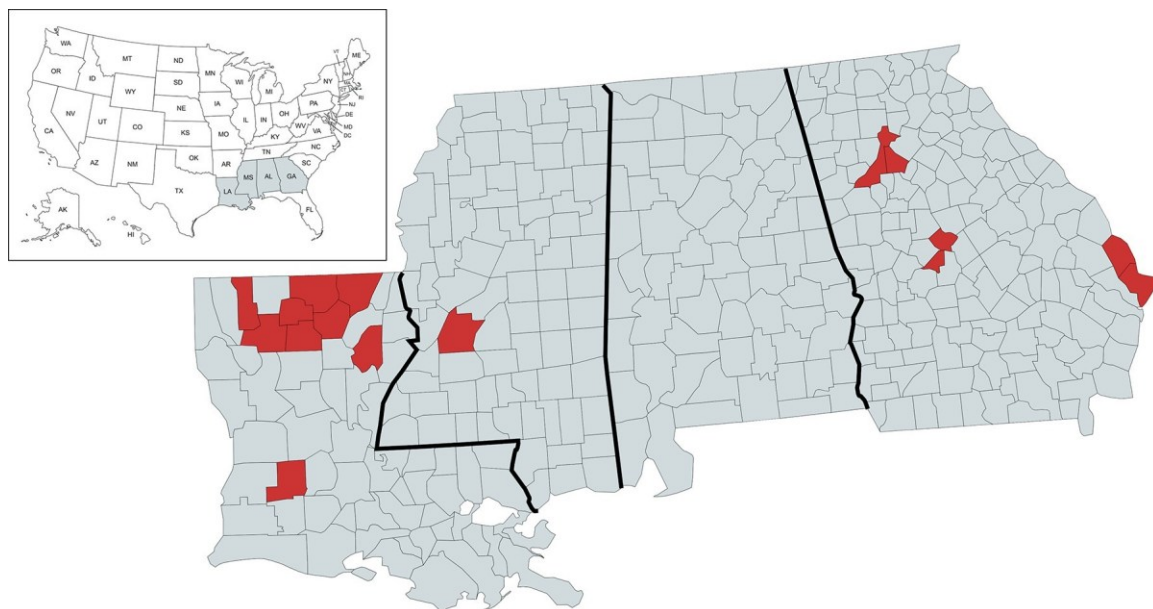


Fig 1. Geographic coverage of SARS-CoV-2 genomic surveillance in Louisiana, Georgia, and Mississippi. Map of the surveillance region with parishes (Louisiana) or counties (Georgia, Mississippi) where at least one specimen was sequenced by the network indicated in red. Map created using MapChart and republished under a CC BY license with permission from Minas Giannakas, original copyright 2014.

<https://doi.org/10.1371/journal.pgph.0001935.g001>

Results and discussion

Improving representativeness in genomic surveillance

In Louisiana, our six partner clinics collected 405 rapid antigen-positive clinical specimens from the residents of nine parishes (Allen, Bienville, Franklin, Jackson, Lincoln, Morehouse, Ouachita, Union, and Webster) between 22 July 2021 and 12 April 2022 (Fig 1). We worked closely with clinicians at the collection sites to gain informed consent from patients and collected metadata on patient demographics and vaccination status for more than 90% of the specimens (Table 1).

In these nine parishes, between 24.2% (Ouachita) to 82.9% (Union) of the residents live in rural areas as defined by the U.S. Census Bureau [43]. Rural areas may be especially vulnerable to the pandemic because they tend to have less capacity for diagnostic testing and healthcare services [44, 45] while being home to a poorer, older, health-compromised population [46]. Bienville and Franklin parishes in particular ranked among the 20 poorest counties out of 300 identified as having the highest COVID-19 mortality rates in recent analysis of 3,200 U.S. counties and county equivalents (e.g., parishes) [47]. Notably, the majority of the other parishes in the region were likewise characterized as low income and high mortality.

The disproportionate impact of COVID-19 on racial and ethnic minorities is well documented [48–50]. Nearly 60% of our donors self-identified as non-white or of mixed heritage (two or more races), compared to approximately 41% non-white overall in the sampled parishes per U.S. Census data [51]. Therefore, our data reflects enhanced representation of minority groups. In our view, this nominal oversampling nonetheless comports with equitable genomic surveillance because it compensates [52] for the overall disproportionately fewer samples from both rural and minority populations in the U.S. as a whole. Notably, 8.4% of our specimens were from Hispanic donors who make up just 3.1% of the northeast Louisiana population but were far more likely to contract COVID-19, accounting for 13.8% of overall cases in the region (up to 60% in some parishes) according to the Louisiana Department of Health COVID-19 dashboard [53]. Our partnership with The Health Hut mobile clinic, which was staffed with Spanish-speaking health care workers, contributed to improved representativeness among this population.

Overall, we detected SARS-CoV-2 RNA in 389 of the 405 specimens collected in Louisiana, using the U.S. CDC's real-time RT-PCR diagnostic assay. We carried out whole viral genome sequencing on these PCR-positive samples, ultimately recovering viral genome data from 272 (70%) of them. These data represented 1.32% of the 20,617 genomes shared via GISAID from Louisiana during the reported period.

Accelerating regional genomic surveillance

After successfully establishing relationships with local clinics and developing a pipeline for exchanging specimens and data across our network in Louisiana, we felt confident sharing our model with other peer institutions across the south. Mercer University School of Medicine in Georgia and Jackson State University in Mississippi both joined our efforts, building on relationships to engage their own communities in genomic surveillance.

In Georgia, COVID-19 positive patient swabs were collected by Mercer Medicine Student Health Services in Bibb, Chatham, DeKalb, Effingham, and Fulton counties and from a community clinic located in Peach County that is operated by Mercer Medicine ([Fig 1](#)). Roughly 81% of the Georgia specimens were obtained from counties outside of the Atlanta metropolitan area. Approximately 48% of Mercer donors did not self-identify as white. After testing positive via PCR analysis, specimens were sent to the Center of Excellence for Emerging Viral

Table 1. Characteristics of study donors in Louisiana (22 July 2021–12 April 2022).

CHARACTERISTICS	NUMBER OF DONORS (%)*
	n = 405
Median age (range)	22 (1–86) years
Sex	
Male	180 (44.4)
Female	215 (53.21)
Not disclosed	10 (2.5)
Race/ethnicity	
White	99 (24.4)
Black	205 (50.6)
Hispanic	34 (8.4)
Asian	4 (<1)
Pacific Islander	1 (<1)
Alaska Native	1 (<1)
Other	1(<1)
Not disclosed	66 (16.2)
PCR	
Negative	15 (3.7)
Positive	389 (96)
Indeterminate	1 (<1)
Vaccination	
No	209 (51.6)
Yes	170 (42)
Partial †	26
Full ‡	113
Booster §	18
Not specified ¶	13
Not disclosed	26 (6.4)
Previous infection	

No	321 (79.3)
Yes	41 (10.1)
Not disclosed	43 (10.6)
Symptomatic	
No	52 (12.8)
Yes	336 (83)
Not disclosed	17 (4.2)

*unless otherwise indicated

Vaccination

† one dose mRNA

‡ two dose mRNA or one dose Janssen

§ three dose mRNA

¶ specific type and date(s) of vaccination unknown

<https://doi.org/10.1371/journal.pgph.0001935.t001>

Threats (CEVT) at Louisiana State University Health Shreveport (LSUHS) for Illumina sequencing.

Among other achievements, our collaboration with Mercer identified two globally emergent Omicron sub-variants that were being monitored closely by the international community due to their concerning patterns of spike substitutions and rapid spread [54, 55]. These included the first documented case in Georgia and the 40th overall U.S. case of the BA.2.75 variant, as well as the sixth case in Georgia of BQ.1.1, one of the most antibody-evasive variants that had been described to date [54]. Overall, Mercer generated 703 SARS-CoV-2 genomes, representing 1.8% of the total data available in GISAID from Georgia between 13 September 2021, when the first Mercer sample was collected up to 10 October 2022.

At our Mississippi site, the Jackson State University (JSU) team made use of the JSU Health Services Center as its sample collection site (Fig 1). The JSU Health Services Center served as the primary COVID-19 testing site for students, staff, and visitors of the university and supported the surrounding community's testing and vaccination efforts. The participation of JSU extended our project's network into a predominantly black community that has long been subjected to environmental injustices and health disparities, including the recent collapse of its city water system [56]. One hundred percent of the 33 samples submitted for sequencing at the CEVT were from non-white donors.

Building modern DNA sequencing capacity for local genomic surveillance

The specimens we collected in Louisiana were initially submitted to the LSUHS CEVT in Shreveport for sequencing until we built local sequencing capacity at Grambling State University (GSU) and Louisiana Tech (Tech). By sequencing locally, we improved our latency or turnaround time from sample collection to release on GISAID, from a median latency of 24 days to 5 days at GSU and 12 days at Tech when processing contemporary specimens (i.e., excluding results from archived samples) (Fig 2B). Through this approach, we identified

the first confirmed case of Omicron in northeast Louisiana, a region comprising 12 parishes and 323,000 residents, within 2 days of sample collection [57]. We then shared the variant data along with weekly wastewater surveillance results and other COVID-19-related health information through a dashboard developed for the north Louisiana community (www.nla-health.com/dashboard).

The results of sequencing including investigation of viral genome coverage and copy number, and phylogenetic analyses are described below.

Viral genome coverage

Plotting the breadth of coverage for Oxford Nanopore Technologies (Nanopore) sequencing reveals that samples with a reverse transcriptase quantitative PCR (RT-qPCR) threshold cycle (Ct) or quantification cycle (Cq) for viral RNA detection of approximately Cq 30 reliably resulted in complete genome coverage in our hands (Fig 3). In some cases, however, full viral genomes were recovered from samples with higher Cq values (indicative of fewer viral genomes in the starting material), and the upper limit in our setting was approximately Cq 35. These observations are congruent with what is seen in the literature across various 'next generation sequencing' (NGS) methodologies for SARS-CoV-2 genome sequencing [58]. Nonetheless, some groups using the Nanopore platform have reported high coverage consensus genome sequences from specimens with Cq values as high as 39 [59]. Variability in primer design and amplification parameters, library preparation method, and sequencing run time can influence results. For example, longer amplicons reduce the likelihood of primer interactions [60] and primer mismatches due to mutations [61], generally outperforming shorter amplicons [62, 63]. Shorter amplicons, however, can recover more complete genomes from specimens where the viral load is low or the RNA is degraded [64, 65].

Specimens with low viral loads can show uneven genome coverage owing to poor amplification of certain amplicons. Such coverage problems can be overcome to some extent by optimizing PCR and sequencing protocols. For instance, Lagerborg *et al.* found that increasing

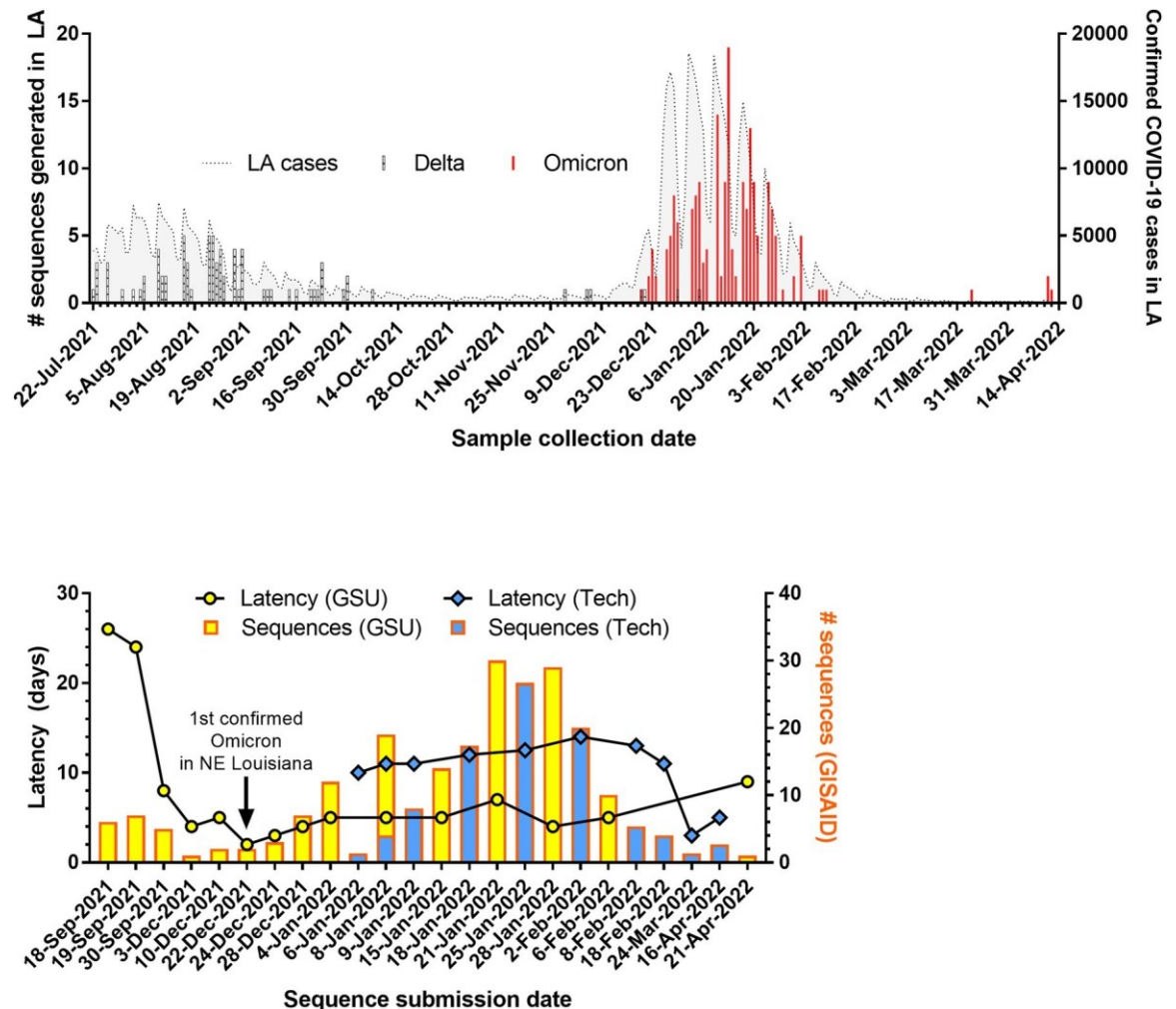


Fig 2. SARS-CoV-2 genomic surveillance in Louisiana. (A) Number of variants sequenced over time in Louisiana overlaid with number of confirmed COVID-19 cases in the U.S. (B) Number of sequences uploaded and latency for new sequencing entities established at Grambling State University (GSU) and Louisiana Tech University (Tech).

<https://doi.org/10.1371/journal.pgph.0001935.g002>

PCR cycles from 35 to 40 resulted in more uniform coverage for Illumina-based ARTIC sequencing of low viral load specimens while still avoiding PCR artifacts [65]. Compared to the Nanopore rapid barcoding kit used for library preparation in our study, the more laborintensive Nanopore ligation sequencing kit can yield greater sequencing depth and better (*de novo*) assembly of the SARS-CoV-2 genome [66]. Of course, longer run times can also increase the number of mapped reads to enhance recovery of complete genomes.

Viral gene copies

The fast-spreading Omicron variant swiftly displaced the Delta variant in our sampling, with a minimal period of overlap (Fig 2A). We find that the mean Cq value of Omicron specimens (28.87, 95% CI 28.13–29.61) was significantly higher ($p < 0.0001$) than that of Delta specimens (24.09, 95% CI 22.51–25.66). This corroborates early reports of lower viral gene

copy number determined by RT-qPCR [67, 68] or lower infectious viral load determined by focus forming assay [69] in Omicron infections compared to Delta infections.

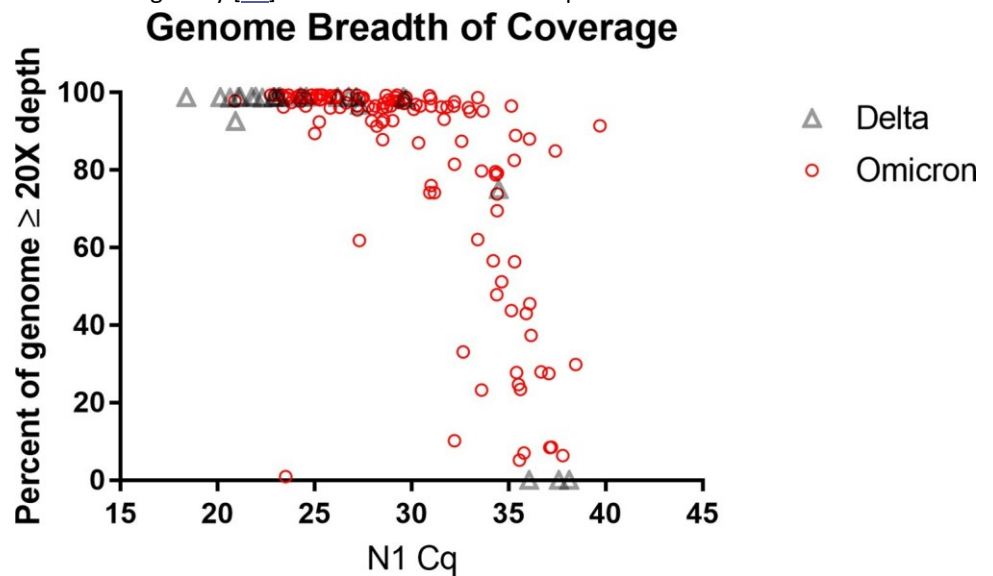


Fig 3. Decreasing genome breadth of coverage in samples with high Cq values. Higher Cq values were observed in Omicron ($n = 115$) than in Delta ($n = 24$). All but one Delta specimen was amplified for 32 cycles and all Omicron specimens were amplified for 35–40 cycles.

<https://doi.org/10.1371/journal.pgph.0001935.g003>

One possible explanation for the observed difference in gene copy number/imputed viral load is that during the initial Omicron surge, a greater proportion of the population may have acquired some degree of immunity through vaccination or prior infection. Some reports support the idea of attenuated viral load with vaccination [70–72], while others found no significant difference in viral genome copies between unvaccinated and vaccinated individuals [73–75].

Having collected vaccination and prior infection metadata from 379 and 362 of the 405 donors respectively, we compared immune history with viral gene copy numbers. We found no statistically significant differences in the mean Cq value between unvaccinated individuals with no previous infection ($n = 80$) and fully vaccinated individuals with or without a previous infection ($n = 80$) (Fig 4A).

Several groups have found no correlation between RT-qPCR quantification of viral genes and symptomatology [76–79]. Two studies reported higher Cq values in asymptomatic individuals ($n = 5$ or 3) compared to symptomatic individuals ($n = 14$ or 9) but the sample sizes were limited [80, 81]. In a larger study, asymptomatic individuals ($n = 2179$) exhibited higher Cq values compared to symptomatic individuals ($n = 739$); however, the authors concluded that the statistically significant difference of 0.71 cycles was not clinically meaningful [82]. In our dataset, the mean Cq value of asymptomatic donors (31.36, 95% CI 29.7–33.01; $n = 30$) was more than 3 cycles higher ($p = 0.0006$) than that of symptomatic donors (28.04, 95% CI 27.28–28.8; $n = 160$) (Fig 4B).

The asymptomatic donors in this study were mostly college students who tested positive during baseline surveillance on the GSU campus, so the age distribution is highly skewed compared to symptomatic donors. There are conflicting reports on the relationship between patient age and viral load, but one of the largest cross-sectional studies found that viral load tends to increase with age [83]. To account for a possible age effect, we compared asymptomatic donors who were age 30 or younger ($n = 28$) with symptomatic donors in the same age group ($n = 93$) and found that the asymptomatic group still had lower ($p = 0.0057$) mean viral gene copy number (31.78, 95% CI 30.17–33.39 vs 28.55, 95% CI 27.39–29.7).

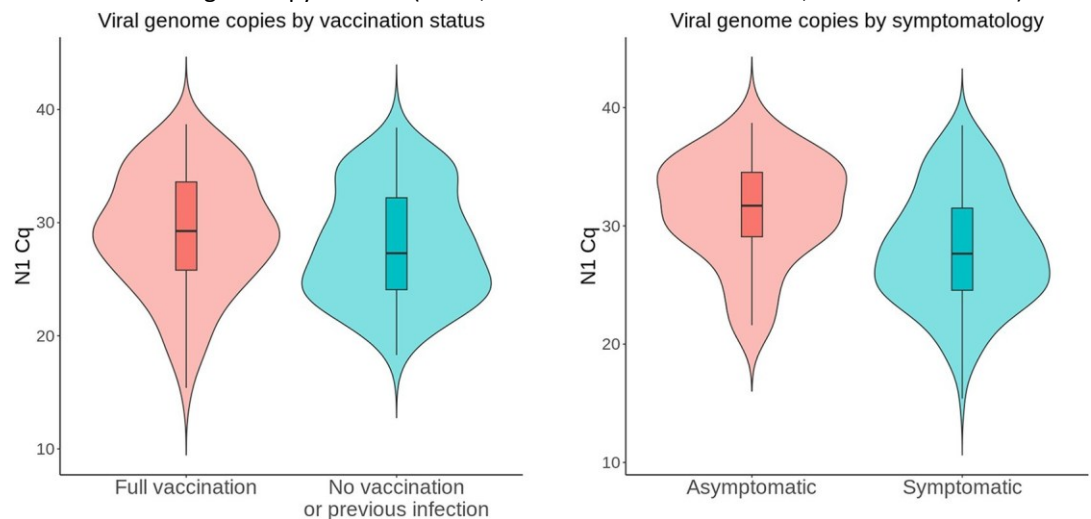


Fig 4. Viral gene copies by vaccination status and symptomatology. (A) No statistically significant differences ($p = 0.1191$) in viral gene copies were observed between unvaccinated donors without a previous infection ($n = 80$) and vaccinated donors with or without a previous infection ($n = 80$). (B) Viral gene copies were lower ($p = 0.0006$) in asymptomatic donors ($n = 30$) compared to symptomatic donors ($n = 160$).

<https://doi.org/10.1371/journal.pgph.0001935.g004>

Only Cq values derived at GSU were used in these analyses to avoid potential confounding by interlaboratory variation in PCR assays [84]. These analyses should be considered carefully within the limitations of our study which include a relatively small sample size and lack of longitudinal information on peak copy number, and also within the overarching context that copy number is an imprecise index of viral load [85].

Phylogenetic analysis of Louisiana and Georgia SARS-CoV-2 genomes

We performed independent phylogenetic analysis for Delta and BA.1.x SARS-CoV-2 viral populations from Louisiana and Georgia in the context of representative global samples. Time-scaled maximum likelihood phylogenies revealed that Delta and BA.1.x isolates from Louisiana and Georgia are interspersed throughout the global population suggesting multiple independent introductions (Fig 5). Often, these introductions resolve into distinct subclades indicative of ongoing transmission within a new location. Further, putative introductions were spatially diverse and included Europe and South America as well as other states throughout the US. These local epidemics were nested within the larger North American epidemic for both Delta and BA.1.x waves.

Conclusion

Collaborative community-centered genomic surveillance networks

As part of a broader initiative supported by the Rockefeller Foundation's Pandemic Prevention Institute, we established a genomic surveillance network focused on equitable viral genomic surveillance among more vulnerable and marginalized populations. To achieve this, we formed collaborations between institutions with very high research activity, primarily undergraduate universities, a graduate medical school, Historically Black Colleges and Universities, and local community health centers. Through this collaborative and decentralized approach, we sequenced viral genomes from respiratory specimens provided by patients in rural areas of Louisiana, Georgia, and Mississippi. These three states rank 41st, 43rd, and 45th respectively

● Africa ● Asia ● Europe ● North America ● Oceania ● South America
 ● GA genomes from this study ● LA genomes from this study

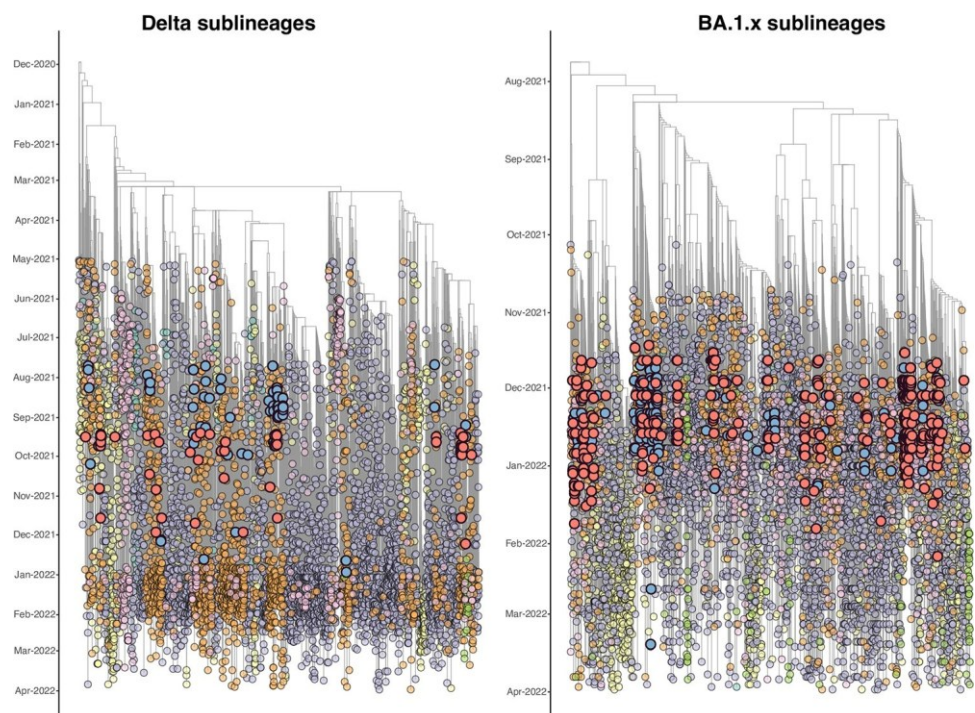


Fig 5. Time-resolved Maximum Likelihood (ML) phylogenies of Delta and Omicron (BA.1.x) sublineages. The genomes generated in this study are colored in red for Georgia and blue for Louisiana. Tree tips are colored according to their location (continent) and the color legend is on the top of the figure.

<https://doi.org/10.1371/journal.pgph.0001935.g005>

among U.S. states in the percentage of cumulative cases sequenced [32]. Thus, with only a modest investment from the Rockefeller Foundation and internal support from our institutions, we increased the viral genomic surveillance coverage from these states such that the sequenced samples better represented the ethnic, racial, and cultural diversity of the U.S. population. Our approach also enabled us to collect and share rich metadata, including vaccination status and county/parish-level geographical location, which is

particularly important when emerging variants such as BA.2.75 and BQ.1.1 are detected. In both the GSU and Tech initiatives, health care workers at partnering clinics briefly educated patients about how viral genomic surveillance is used to track emerging variants and provided patients with the opportunity to provide consent for their positive sample to be used for whole viral genome sequencing.

Residents of rural areas face significant barriers rooted in geographic isolation, scarcity of services, lower socioeconomic status and other social determinants of health, and even cultural constraints that impede them from getting the healthcare that they need [86]. During the pandemic, this disparity manifested as disproportionately high mortality rates in rural America, particularly among minorities [87]. In rural areas, so-called regional universities represent essential infrastructure that can provide not only educational opportunities but also access to testing and vaccination. These universities are uniquely positioned to gain trust and encourage participation in genomic surveillance because they often function as community anchors, vital to the local economy, civics, and culture [88]. A viral genomic surveillance strategy that leverages these strong community ties also offers a much-needed on-ramp for students from diverse groups to gain genomics research experience that will prepare them to combat misinformation in their own communities and participate in the genomics workforce.

Viral genomic surveillance at all scales

Oxford Nanopore Technologies (Nanopore) DNA sequencing has been used for real-time genomic surveillance during outbreaks of Ebola in West Africa [89], Zika [90] and Yellow Fever [91] in Brazil, and Lassa fever in Nigeria [92]. During the COVID-19 pandemic, many laboratories around the world used Nanopore sequencing to rapidly generate and share SARS-CoV-2 genome data. The Nanopore platform is especially suitable for smaller institutions because start-up costs and infrastructure requirements are minimal, operational costs are low, and the approach is highly scalable. A modestly equipped molecular biology laboratory can implement a Nanopore rapid barcoding “swab to data” workflow for less than \$3500 USD at current prices. Operational costs, excluding RNA extraction, are estimated at £18.91 (\$23.61 USD) per ligated barcode for 12 barcodes on a MiniON flow cell when using a wash kit to remove the previous library [93]. This corresponds well with our cost analysis, which *includes* RNA extraction, of \$20–\$25 USD per tagged barcode for 12 barcodes when the flow cell was washed and reused.

Working with raw genome sequencing data often requires significant computational resources and bioinformatics expertise. However, Nanopore-based viral genome sequencing can be performed on a consumer-grade personal computer, using readily available workflows that include cloud-based, “point-and-click” assembly and analysis pipelines. Although higher throughput facilities can multiplex 96 or more barcodes at a cost of less than £10 (\$12.50 USD) per barcode [93], the platform enables smaller laboratories to cost-effectively provide genomic surveillance on limited numbers of samples, thus enhancing the ability reach underserved communities and provide much needed coverage in under-sampled areas.

The democratization of pathogen genomics is nowhere better demonstrated than in Africa, where technological advances in viral genome sequencing have been coupled with

major investments by Africa CDC and its partners who rapidly expanded genomic surveillance across the continent [94]. These efforts in Africa resonate with some of our experiences in the rural Southern U.S., highlighting that locally based programs can close blind spots in viral genomic surveillance while significantly decreasing turnaround time. The success of these types of decentralized, community-based approaches is exemplified by our identification of the first confirmed case of Omicron (B.1.1.529, or 'BA.1') in northeast Louisiana, and by our early detection in Georgia of concerning emergent Omicron sublineages BA.2.75 and BQ.1.1.

Overall, our work demonstrates the positive impact that universities can have within their communities, in particular the power of partnerships forged between academic teams and local community clinics to build engagement and foster participation while generating valuable data relevant to public health responses. We argue that policymakers and funding agencies should strongly support the formation of these types of nimble, local networks in global strategies to modernize viral genomic surveillance of common respiratory illnesses. To facilitate the expansion of community-based genomic surveillance programs, we have developed a guide entitled, *Building a Collaborative and Equitable Viral Genomic Surveillance Program: A Playbook for Researchers, Clinicians, Administrators, and Allies*, that offers a modular and flexible roadmap for universities and local clinics to set up their own viral genomic surveillance program (please contact the corresponding authors for information). When community-based programs that rapidly generate and share pathogen genome data become commonplace across geographies, the world will be better poised not only to promptly detect new pandemic threats but also to train a diverse and representative public health workforce for tomorrow.

Methods

Specimen collection

Patients testing positive ($n = 405$) on the BinaxNOW rapid antigen test (Abbott Laboratories; Abbott Park, IL) at the Foster Johnson Health Center in Grambling State University (Grambling, LA), The Health Hut (Ruston, LA), Louisiana Tech University (Ruston, LA), Michael Brooks Family Clinic (Ruston, LA), Minden Family Care Center (Minden, LA), and Serenity Springs Specialty Hospital (Ruston, LA) between 23 August 2021 and 12 April 2022 were recruited through an informed consent process. Remnant anterior nasal swabs used in the rapid test (The Health Hut, Michael Brooks Family Clinic, Louisiana Tech University, Minden Family Care Center, Serenity Springs Specialty Hospital) or freshly collected mid-turbinate nasal swabs (Foster Johnson Health Center) were preserved in 3 mL of viral transport media (VTM) at 4°C and later in the study in 1 mL of DNA/RNA Shield to inactivate the virus and allow for storage at room temperature.

Ethical considerations

Specimen collection in Louisiana clinics was conducted under 45 CFR 46.102(I)(2) as a public health surveillance activity. Consent was obtained in writing from all participants in Louisiana in accordance with Louisiana Tech University (Ruston, LA) IRB protocol HUC 21–106. Specimen collection in Georgia was conducted under Mercer University (Macon, GA) IRB protocol H2110206. Specimen collection in Mississippi was conducted under Jackson State University (Jackson, MS) IRB protocol 0097–22. Consent was not obtained from

participants in Georgia or Mississippi because the IRBs only approved collection of fully de-identified clinical remnant samples. Sequencing of samples at LSUHS CEVT was conducted under IRB protocol # STUDY00001445.

Illumina sequencing (LSUHS CEVT)

Total RNA was isolated from VTM within 72 hours of specimen collection using the Omega Bio-Tek Viral RNA Xpress kit (Norcross, GA) on an automated extraction platform. Presence of SARS-CoV-2 RNA was determined by using CDC primers and probes with LunaScript RT Supermix Kit (NEB) run on BioRad (Hercules, CA) C1000 real-time PCR machines. cDNA, PCR amplification, and library preparation was performed using the NEBNext ARTIC SARS-CoV-2 FS kit (NEB) according to the manufacturer's instructions. Libraries were quantified using Invitrogen Qubit BR dsDNA kit and size determined on an Agilent Tapestation 2200 using high sensitivity d1000 tapes (Santa Clara, CA). Libraries were loaded at 10.5 pM on an Illumina (San Diego, CA) MiSeq version 2, 300 cycle kit on an Illumina MiSeq instrument. Consensus sequences were generated with an in-house data analysis pipeline using bwa, samtools, and ivar.

Nanopore sequencing (GSU and Tech)

Total RNA was extracted from VTM or DNA/RNA Shield within 48 hours of specimen collection via the QIAmp Viral RNA Minikit (Qiagen; Germantown, MD) or Zymo Quick-DNA/RNA Viral MagBead Kit (Zymo Research; Irvine, CA) according to the manufacturers' protocols. Briefly, 500 uL of VTM was centrifuged at 5,000 g for 10 minutes to pellet cells and debris and either 140 uL of supernatant was extracted in 560 uL of Buffer AVL containing 560 ug of carrier RNA (QIAmp) or 200 uL of supernatant was extracted in 400 uL of Viral DNA/RNA Buffer followed by addition of 10 uL MagBinding Beads (Zymo). After washing the spin column or beads with the manufacturers' respective wash buffers, total RNA was eluted from the column in 60 uL of Buffer AVE (QIAmp) or from the beads in 30 uL of DNase/RNase-free water (Zymo) and immediately used for RT-qPCR and library preparation. The presence of SARS-CoV-2 was confirmed using the SARS-CoV-2 Research Use Only qPCR Primer and Probe Kit (Integrated DNA Technologies, IDT; Coralville, IA) and qScript XLT One-Step RTqPCR ToughMix (Quantabio; Beverly, MA) according to the CDC 2019-nCoV Real-Time RT-PCR Diagnostic Panel procedure.

The sequencing library was prepared using the 1200 bp amplicon "midnight" primer set (versions 1 or 2) [95] following the protocol by Freed and Silander [61]. Briefly, the RNA was reverse transcribed using the LunaScript RT Supermix Kit (New England BioLabs, NEB; Ipswich, MA). Tiled 1200 bp amplicons were generated using midnight primers (IDT) with Q5 Hot Start High-Fidelity 2X Master Mix (NEB) for 32 cycles (LA-GSU1 to LA-GSU19) or up to 40 cycles (LA-GSU20 to LA-GSU148 and LA-TECH1 to LA-TECH68) of multiplex PCR amplification. The two overlapping amplicon pools for each specimen were combined and quantified using the Qubit dsDNA HS Assay Kit (Invitrogen; Carlsbad, CA). A total of 75–150 ng of PCR product per specimen was barcoded using the Nanopore Rapid Barcoding Kits SQK-RBK004 or SQK-RBK110.96 (Oxford Nanopore Technologies, ONT; Oxford, UK). The barcoded samples were pooled and purified using AMPure XP beads at a 1:1 ratio of sample: beads (Beckman Coulter; Brea, CA). The final libraries (350–800 ng DNA) were sequenced on an R9.4.1 flow cell using a MinION Mk1B or Mk1C device, basecalled in real-time via MinKNOW, and

analyzed using the wf-artic workflow (medaka, minimap2, bcftools, samtools, nextclade, artic, pangolin) according to the ONT protocol.

Nanopore concordance with Illumina

We validated the performance of Nanopore MinION sequencing at GSU and Tech by sequencing a matched subset of 16 SARS-CoV-2 positive specimens using a well-established Illumina workflow at the CEVT. The results confirmed the accuracy of the Nanopore sequencing protocol and analysis pipeline in our setting with 100% consensus sequence identity in 13/16 of the samples and 99.9% sequence identity in 3/16 samples. The one discordant basecall was identical in the three samples: T by Illumina sequencing and G by Nanopore sequencing at 17,259, within the gene encoding the replicase accessory protein NSP13. The reference SARS-CoV-2 genome also contains a G at this position; the T mutation identified by Illumina sequencing results in a glutamic acid rather than an aspartic acid. The E1264D mutation identified in the samples sequenced with the Illumina protocol is well supported. Nanopore basecalling accuracy decreases in homopolymeric tracts [96], but there was not a homopolymer in this region, so the reason for this discordance remains unclear.

Importantly, all sequences were 100% concordant for the Pango lineages and GISAID clade assignments. The 16 samples sequenced by both protocols spanned a wide range of Cq values (15.6–32.4) but almost all samples, Nanopore or Illumina, generated > 98% genome coverage (Nanopore: 98.73%, Illumina: 99.67%). Illumina sequencing resulted in an average of 285 bases more sequence data than Nanopore, but this was mainly confined to the extreme 5' and 3' ends of the genome.

Phylogenetic analysis

SARS-CoV-2 genomes were downloaded from GISAID (www.gisaid.org) excluding the low coverage and incomplete records up to 30 April 2022, and subsampling carried out using subsampler to obtain a representative dataset [97]. Subsampling 4,199,772 (Delta sublineages) and 2,895,902 (BA.1.x sublineages) genomic sequences resulted in 5,882 (Delta sublineages) and 6,843 (BA.1.x sublineages) total sequences. Viral sequences were aligned using ViralMSA with default parameters [98] using Wuhan-1 (MN908947.3) reference and then manually curated with Aliview to designate the start site and trim terminal regions [99]. A maximum likelihood (ML) phylogeny was inferred using IQTREE v1.6.10 with the best-fit nucleotide substitution model identified by the Model Finder function [100]. Statistical support for nodes was determined using 1,000 bootstrap replicates. TreeTime [101] was used to transform this ML tree topologies into a dated tree using a constant mean rate of 8.0×10^{-4} nucleotide substitutions per site per year, after the exclusion of outlier sequences. The phylogeny was visualized using the *ggtree* package in RStudio with tips colored by collection source (continent).

Statistical analysis

Statistical analysis was performed using Prism version 7.05 (GraphPad; San Diego, CA). Continuous variables between two groups were tested for normal distribution (D'Agostino &

Pearson test) and homogeneity of variance (F test) then compared using an independent t-test with two-tails and alpha set to 0.05.

Supporting information

S1 Table. EPI_SET ID for sequences from this initiative.

(PDF)

S2 Table. EPI_SET ID for subsampled set of global sequences.

(PDF)

Acknowledgments

The authors deeply thank the healthcare professionals at Foster Johnson Health Center (Dawn Holmes), The Health Hut (Jackie White, Chelsea Streets), Michael Brooks Family Clinic (Natalie Wise, Taylor Schutzman, Whitney Walker), Minden Family Care Center (Beau Burton), Serenity Springs Specialty Hospital (Tonya Monk), Mercer Medicine (Certified Medical Assistants), and Jackson State University's Health Services Center (Kimberly Bullock) for recruiting patients, gaining informed consent, and/or collecting the clinical specimens. We thank Maarten Van Diest (Louisiana State University Health Shreveport) for assistance with data curation. We are grateful to the Louisiana Department of Health for their efforts to facilitate enhanced viral genomic surveillance as a public health activity. We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. The full acknowledgement table can be found in [S1](#) and [S2](#) Tables. We gratefully acknowledge the contribution of the GISAID Data Curation Team to enable rapid sharing of genomic data. We acknowledge the Louisiana Biomedical Research Network and its Bioinformatics, Biostatistics, and Computational Biology Core for providing computational resources.

Author Contributions

Conceptualization: Paul Y. Kim, Jamie J. Newman, Jeremy P. Kamil.

Data curation: Olga N. Uchakina, Vance L. Mack, Ifedayo Victor Ogungbe, Gregory Ware, Jennifer Carroll, Jarrod F. Sawyer, Kenneth H. Densmore, Michael Foster, Lescia Valmond, Krista Queen, Jeremy P. Kamil.

Formal analysis: Paul Y. Kim, Eleonora Cella, Taj Azarian.

Funding acquisition: Paul Y. Kim, Jamie J. Newman, Jeremy P. Kamil.

Investigation: Paul Y. Kim, Audrey Y. Kim, Jamie J. Newman, Olga N. Uchakina, Vance L. Mack, Ifedayo Victor Ogungbe, Olawale Adeyinka, Samuel Jones, Michael Foster, Lescia Valmond, John Thomas, Krista Queen, Jeremy P. Kamil.

Project administration: Paul Y. Kim, Jamie J. Newman, Peter J. Huwe, Robert J. McKallip, Marnie P. Hill, Ifedayo Victor Ogungbe, Gregory Ware, Jennifer Carroll, Kenneth H. Densmore, Krista Queen, Jeremy P. Kamil.

Resources: Thomas C. Bishop, Olga N. Uchakina, Vance L. Mack, Samuel Jones.

Software: Thomas C. Bishop, Vance L. Mack.

Supervision: Audrey Y. Kim, Peter J. Huwe, Robert J. McKallip.

Validation: Vance L. Mack.

Visualization: Eleonora Cella, Kenneth H. Densmore.

Writing – original draft: Paul Y. Kim.

Writing – review & editing: Paul Y. Kim, Audrey Y. Kim, Jamie J. Newman, Eleonora Cella, Thomas C. Bishop, Marnie P. Hill, Ifedayo Victor Ogungbe, Taj Azarian, Krista Queen, Jeremy P. Kamil.

References

1. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020; 579: 265–269. <https://doi.org/10.1038/s41586-020-2008-3> PMID: [32015508](#)
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020; 382: 727–733. <https://doi.org/10.1056/NEJMoa2001017> PMID: [31978945](#)
3. Brown CM, Vostok J, Johnson H, Burns M, Gharpure R, Sami S, et al. Outbreak of SARS-CoV-2 Infections, Including COVID-19 Vaccine Breakthrough Infections, Associated with Large Public Gatherings—Barnstable County, Massachusetts, July 2021. *MMWR Morb Mortal Wkly Rep*. 2021; 70: 1059–1062. <https://doi.org/10.15585/mmwr.mm7031e2> PMID: [34351882](#)
4. Sheikh A, McMenamin J, Taylor B, Robertson C, Public Health Scotland and the EAVE II Collaborators. SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness. *Lancet*. 2021; 397: 2461–2462.
5. Twohig KA, Nyberg T, Zaidi A, Thelwall S, Sinnathamby MA, Aliabadi S, et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *Lancet Infect Dis*. 2022; 22: 35–42. [https://doi.org/10.1016/S14733099\(21\)00475-8](https://doi.org/10.1016/S14733099(21)00475-8) PMID: [34461056](#)
6. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*. 2021; 596: 276–280. <https://doi.org/10.1038/s41586-021-03777-9> PMID: [34237773](#)
7. Planas D, Saunders N, Maes P, Guivel-Benhassine F, Planchais C, Buchrieser J, et al. Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. *Nature*. 2022; 602: 671–675. <https://doi.org/10.1038/s41586-021-04389-z> PMID: [35016199](#)
8. Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*. 2022; 603: 679–686. <https://doi.org/10.1038/s41586-022-04411-y> PMID: [35042229](#)
9. Tegally H, Moir M, Everatt J, Giovanetti M, Scheepers C, Wilkinson E, et al. Emergence of SARSCoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nat Med*. 2022; 28: 1785–1790. <https://doi.org/10.1038/s41591-022-01911-2> PMID: [35760080](#)
10. 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. *Nature*. 2022; 608: 593–602. <https://doi.org/10.1038/s41586-02204980-y> PMID: [35714668](#)

11. Wang Q, Guo Y, Iketani S, Nair MS, Li Z, Mohri H, et al. Antibody evasion by SARS-CoV-2 Omicron subvariants BA.2.12.1, BA.4 and BA.5. *Nature*. 2022; 608: 603–608. <https://doi.org/10.1038/s41586022-05053-w> PMID: [35790190](#)
12. Carabelli AM, Peacock TP, Thorne LG, Harvey WT, Hughes J, COVID-19 Genomics UK Consortium, et al. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol*. 2023; 21: 162–177. <https://doi.org/10.1038/s41579-022-00841-7> PMID: [36653446](#)
13. Cao Y, Jian F, Wang J, Yu Y, Song W, Yisimayi A, et al. Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. *Nature*. 2023; 614: 521–529. <https://doi.org/10.1038/s41586-022-05644-7> PMID: [36535326](#)
14. Miller J, Hachmann NP, Collier A-RY, Lasrado N, Mazurek CR, Patio RC, et al. Substantial Neutralization Escape by SARS-CoV-2 Omicron Variants BQ.1.1 and XBB.1. *N Engl J Med*. 2023; 388: 662–664. <https://doi.org/10.1056/NEJMc2214314> PMID: [36652339](#)
15. Planas D, Bruel T, Staropoli I, Guivel-Benhassine F, Porrot F, Maes P, et al. Resistance of Omicron subvariants BA.2.75.2, BA.4.6, and BQ.1.1 to neutralizing antibodies. *Nat Commun*. 2023; 14: 824. <https://doi.org/10.1038/s41467-023-36561-6> PMID: [36788246](#)
16. Qu P, Evans JP, Faraone JN, Zheng Y-M, Carlin C, Anghelina M, et al. Enhanced neutralization resistance of SARS-CoV-2 Omicron subvariants BQ.1, BQ.1.1, BA.4.6, BF.7, and BA.2.75.2. *Cell Host Microbe*. 2023; 31: 9–17.e3. <https://doi.org/10.1016/j.chom.2022.11.012> PMID: [36476380](#)
17. Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of SARSCoV-2. *Nat Rev Microbiol*. 2023. <https://doi.org/10.1038/s41579-023-00878-2> PMID: [37020110](#)
18. Isidro J, Borges V, Pinto M, Sobral D, Santos JD, Nunes A, et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nat Med*. 2022; 28: 1569–1572. <https://doi.org/10.1038/s41591-022-01907-y> PMID: [35750157](#)
19. Tian J, Sun J, Li D, Wang N, Wang L, Zhang C, et al. Emerging viruses: Cross-species transmission of coronaviruses, filoviruses, henipaviruses, and rotaviruses from bats. *Cell Rep*. 2022; 39: 110969. <https://doi.org/10.1016/j.celrep.2022.110969> PMID: [35679864](#)
20. Waddington C, Carey ME, Boinett CJ, Higginson E, Veeraraghavan B, Baker S. Exploiting genomics to mitigate the public health impact of antimicrobial resistance. *Genome Med*. 2022; 14: 15. <https://doi.org/10.1186/s13073-022-01020-2> PMID: [35172877](#)
21. Forde BM, Bergh H, Cuddihy T, Hajkiewicz K, Hurst T, Playford EG, et al. Clinical Implementation of Routine Whole-genome Sequencing for Hospital Infection Control of Multi-drug Resistant Pathogens. *Clin Infect Dis*. 2023; 76: e1277–e1284. <https://doi.org/10.1093/cid/ciac726> PMID: [36056896](#)
22. Jolly B, Scaria V. A distinct phylogenetic cluster of Monkeypox genomes suggests an early and cryptic spread of the virus. *J Infect*. 2023; 86: e24–e26. <https://doi.org/10.1016/j.jinf.2022.08.013> PMID: [35995311](#)
23. Edward C. Holmes on behalf of Professor Yong-Zhen Zhang and colleagues, Fudan University, Shanghai, China. Novel 2019 coronavirus genome. In: *Virological* [Internet]. 20 Jan 2020. Available: <https://virological.org/t/novel-2019-coronavirus-genome/319>
24. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall*. 2017; 1: 33–46. <https://doi.org/10.1002/gch2.1018> PMID: [31565258](#)
25. Arita M, Karsch-Mizrachi I, Cochrane G. The international nucleotide sequence database collaboration. *Nucleic Acids Res*. 2021; 49: D121–D124. <https://doi.org/10.1093/nar/gkaa967> PMID: [33166387](#)
26. Abbasi J. How the US Failed to Prioritize SARS-CoV-2 Variant Surveillance. *JAMA*. 2021; 325: 1380–1382. <https://doi.org/10.1001/jama.2021.3368> PMID: [33760030](#)
27. Brito AF, Semenova E, Dudas G, Hassler GW, Kalinich CC, Kraemer MUG, et al. Global disparities in SARS-CoV-2 genomic surveillance. *Nat Commun*. 2022; 13: 7003. <https://doi.org/10.1038/s41467022-33713-y> PMID: [36385137](#)

28. Chen Z, Azman AS, Chen X, Zou J, Tian Y, Sun R, et al. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. *Nat Genet.* 2022; 54: 499–507. <https://doi.org/10.1038/s41588-02201033-y> PMID: [35347305](#)
29. National Academies of Sciences, Engineering, and Medicine; Division on Earth and Life Studies; Board on Life Sciences; Health and Medicine Division; Board on Health Sciences Policy; Committee on Data Needs to Monitor the Evolution of SARS-CoV-2. *Genomic Epidemiology Data Infrastructure Needs for SARS-CoV-2: Modernizing Pandemic Response Strategies.* Washington (DC): National Academies Press (US).
30. Else H. COVID “Fast Grants” sped up pandemic science. *Nature.* 2021. <https://doi.org/10.1038/d41586-021-02111-7> PMID: [34345037](#)
31. McCormick E. “We were in the dark”: why the US is far behind in tracking Covid-19 variants. *The Guardian.* 2 Feb 2021. Available: <https://www.theguardian.com/world/2021/feb/02/us-covid-varianttracking-cdc-research>
32. Centers for Disease Control and Prevention. COVID Data Tracker. In: COVID Data Tracker [Internet]. Atlanta, GA; [cited 1 Oct 2022]. Available: <https://covid.cdc.gov/covid-data-tracker>
33. Cuadros DF, Branscum AJ, Mukandavire Z, Miller FD, MacKinnon N. Dynamics of the COVID-19 epidemic in urban and rural areas in the United States. *Ann Epidemiol.* 2021; 59: 16–20. <https://doi.org/10.1016/j.annepidem.2021.04.007> PMID: [33894385](#)
34. Matthews KA, Ullrich F, Gaglioti AH, Dugan S, Chen MS, Hall DM. Nonmetropolitan COVID-19 Incidence and Mortality Rates Surpassed Metropolitan Rates Within the First 24 Weeks of the Pandemic Declaration: United States, March 1–October 18, 2020. *J Rural Health.* 2021; 37: 272–277. <https://doi.org/10.1111/jrh.12555> PMID: [33619806](#)
35. Mueller JT, McConnell K, Burow PB, Pofahl K, Merdjanoff AA, Farrell J. Impacts of the COVID-19 pandemic on rural America. *Proc Natl Acad Sci U S A.* 2021; 118: 2019378118. <https://doi.org/10.1073/pnas.2019378118> PMID: [33328335](#)
36. Cheng KJG, Sun Y, Monnat SM. COVID-19 Death Rates Are Higher in Rural Counties With Larger Shares of Blacks and Hispanics. *J Rural Health.* 2020; 36: 602–608. <https://doi.org/10.1111/jrh.12511> PMID: [32894612](#)
37. Henning-Smith C, Tuttle M, Kozhimannil KB. Unequal Distribution of COVID-19 Risk Among Rural Residents by Race and Ethnicity. *J Rural Health.* 2021; 37: 224–226. <https://doi.org/10.1111/jrh.12463> PMID: [32396220](#)
38. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Committee on Community-Based Solutions to Promote Health Equity, Baciu A, Negussie Y, et al. *The State of Health Disparities in the United States.* National Academies Press (US); 2017.
39. Houtman J, Glassman R, Shultz L, Rivera JM, Bass E, Bright R. Genomic Surveillance Is Essential To Track Covid-19 Variants in Both Unvaccinated and Vaccinated Populations. In: Rockefeller Foundation [Internet]. 1 Dec 2021 [cited 8 Jun 2022]. Available: <https://www.rockefellerfoundation.org/blog/genomic-surveillance-is-essential-to-track-covid-19-variants-in-both-unvaccinated-and-vaccinatedpopulations/>
40. McNeil C, Verlander S, Divi N, Smolinski M. The Landscape of Participatory Surveillance Systems Across the One Health Spectrum: Systematic Review. *JMIR Public Health Surveill.* 2022; 8: e38551. <https://doi.org/10.2196/38551> PMID: [35930345](#)
41. Gozashti L, Corbett-Detig R. Shortcomings of SARS-CoV-2 genomic metadata. *BMC Res Notes.* 2021; 14: 189. <https://doi.org/10.1186/s13104-021-05605-9> PMID: [34001211](#)
42. Schriml LM, Chuvochina M, Davies N, Eloë-Fadroshe EA, Finn RD, Hugenholtz P, et al. COVID-19 pandemic reveals the peril of ignoring metadata standards. *Scientific Data.* 2020; 7: 1–4.
43. Ratcliffe M, Burd C, Holder K, Fields A. *Defining Rural at the U.S. Census Bureau.* U.S. Census Bureau; 2016. Report No.: ACSGEO-1.
44. Iglehart JK. The Challenging Quest to Improve Rural Health Care. *N Engl J Med.* 2018; 378: 473–479. <https://doi.org/10.1056/NEJMhpr1707176> PMID: [29385378](#)
45. Johnston KJ, Wen H, Joynt Maddox KE. Lack Of Access To Specialists Associated With Mortality And Preventable Hospitalizations Of Rural Medicare Beneficiaries. *Health Aff.* 2019; 38: 1993–2002.

- <https://doi.org/10.1377/hlthaff.2019.00838> PMID: 31794307
46. Peters DJ. Community Susceptibility and Resiliency to COVID-19 Across the Rural-Urban Continuum in the United States. *J Rural Health*. 2020; 36: 446–456. <https://doi.org/10.1111/jrh.12477> PMID: 32543751
 47. A Poor People's Pandemic Report: Mapping the Intersections of Poverty, Race, and COVID-19. In: Poor Peoples Campaign [Internet]. Apr 2022 [cited 8 Jun 2022]. Available: <https://www.poorpeoplescampaign.org/pandemic-report/>
 48. Webb Hooper M, Na'poles AM, Pe'rez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. *JAMA*. 2020; 323: 2466–2467. <https://doi.org/10.1001/jama.2020.8598> PMID: 32391864
 49. Gross CP, Essien UR, Pasha S, Gross JR, Wang S-Y, Nunez-Smith M. Racial and Ethnic Disparities in Population-Level Covid-19 Mortality. *Journal of General Internal Medicine*. 2020. pp. 3097–3099. <https://doi.org/10.1007/s11606-020-06081-w> PMID: 32754782
 50. Mackey K, Ayers CK, Kondo KK, Saha S, Advani SM, Young S, et al. Racial and Ethnic Disparities in COVID-19-Related Infections, Hospitalizations, and Deaths: A Systematic Review. *Ann Intern Med*. 2021; 174: 362–373. <https://doi.org/10.7326/M20-6306> PMID: 33253040
 51. U.S. Census Bureau. Decennial Census P2: HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE. Available: [https://data.census.gov/table?q=race&q=040XX00US22\\$0500000&tid=DECENNIALPL2020.P2](https://data.census.gov/table?q=race&q=040XX00US22$0500000&tid=DECENNIALPL2020.P2)
 52. World Health Organization. Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health. World Health Organization; 2021 Jan.
 53. Louisiana Department of Health. Comparison of Percentage of COVID-19 Cases, Deaths, Vaccinations, and Populations by Race by Parish. In: Louisiana COVID-19 dashboard [Internet]. [cited 25 Apr 2023]. Available: <https://ldh.la.gov/Coronavirus/>
 54. Cao Y, Jian F, Wang J, Yu Y, Song W, Yisimayi A, et al. Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. *bioRxiv*. 2022. p. 2022.09.15.507787. <https://doi.org/10.1038/s41586-022-05644-7> PMID: 36535326
 55. Sheward DJ, Kim C, Fischbach J, Muschiol S, Ehling RA, Bjo'rkstro'm NK, et al. Omicron sublineage BA.2.75.2 exhibits extensive escape from neutralising antibodies. *bioRxiv*. 2022. p. 2022.09.16.508299. [https://doi.org/10.1016/S1473-3099\(22\)00663-6](https://doi.org/10.1016/S1473-3099(22)00663-6) PMID: 36244347
 56. Nwanaji-Enwerem JC, Casey JA. Naming Civic Health in Environmental Justice Discourse: The Jackson Water Crisis. *Lancet Reg Health Am*. 2022; 14. <https://doi.org/10.1016/j.lana.2022.100378> PMID: 36380985
 57. Bergeron N. Omicron Variant Identified in Lincoln Parish. *Ruston Daily Leader*. 23 Dec 2021: 1.
 58. Baker DJ, Aydin A, Le-Viet T, Kay GL, Rudder S, de Oliveira Martins L, et al. CoronaHIT: high-throughput sequencing of SARS-CoV-2 genomes. *Genome Med*. 2021; 13: 21. <https://doi.org/10.1186/s13073-021-00839-5> PMID: 33563320
 59. Yan Y, Wu K, Chen J, Liu H, Huang Y, Zhang Y, et al. Rapid Acquisition of High-Quality SARS-CoV-2 Genome via Amplicon-Oxford Nanopore Sequencing. *Viol Sin*. 2021; 36: 901–912. <https://doi.org/10.1007/s12250-021-00378-8> PMID: 33851337
 60. Itokawa K, Sekizuka T, Hashino M, Tanaka R, Kuroda M. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLoS One*. 2020; 15: e0239403. <https://doi.org/10.1371/journal.pone.0239403> PMID: 32946527
 61. Freed N, Silander O. SARS-CoV2 genome sequencing protocol (1200bp amplicon "midnight" primer set, using Nanopore Rapid kit) v2. *protocols.io*. ZappyLab, Inc.; 2021. <https://doi.org/10.17504/protocols.io.bwyppfvn>
 62. Brejova' B, Bors'ova' K, Hodorova' V, Čabanova' V, Gafurov A, Fričova' D, et al. Nanopore sequencing of SARS-CoV-2: Comparison of short and long PCR-tiling amplicon protocols. *PLoS One*. 2021; 16: e0259277. <https://doi.org/10.1371/journal.pone.0259277> PMID: 34714886
 63. Liu H, Li J, Lin Y, Bo X, Song H, Li K, et al. Assessment of two-pool multiplex long-amplicon nanopore sequencing of SARS-CoV-2. *J Med Virol*. 2022; 94: 327–334. <https://doi.org/10.1002/jmv.27336> PMID: 34524690

64. Charre C, Ginevra C, Sabatier M, Regue H, Destras G, Brun S, et al. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evol.* 2020; 6: veaa075.
65. Lagerborg KA, Normandin E, Bauer MR, Adams G, Figueroa K, Loreth C, et al. DNA spike-ins enable confident interpretation of SARS-CoV-2 genomic data from amplicon-based sequencing. *bioRxiv.* 2021. <https://doi.org/10.1101/2021.03.16.435654> PMID: 33758855
66. Gonza'lez-Recio O, Gutie'rrez-Rivas M, Peiro'-Pastor R, Aguilera-Sepulveda P, Cano-Go'mez C, Jime'nez-Clavero MA', et al. Sequencing of SARS-CoV-2 genome using different nanopore chemistries. *Appl Microbiol Biotechnol.* 2021; 105: 3225–3234. <https://doi.org/10.1007/s00253-021-11250-w> PMID: 33792750
67. Hay JA, Kissler SM, Fauver JR, Mack C, Tai CG, Samant RM, et al. Viral dynamics and duration of PCR positivity of the SARS-CoV-2 Omicron variant. *medRxiv.* 2022. <https://doi.org/10.1101/2022.01.13.22269257>
68. Woodbridge Y, Amit S, Huppert A, Kopelman NM. Viral load dynamics of SARS-CoV-2 Delta and Omicron variants following multiple vaccine doses and previous infection. *medRxiv.* 2022. <https://doi.org/10.1038/s41467-022-33096-0> PMID: 36344489
69. Puhach O, Adea K, Hulo N, Sattonnet P, Genecand C, Iten A, et al. Infectious viral load in unvaccinated and vaccinated patients infected with SARS-CoV-2 WT, Delta and Omicron. *bioRxiv.* 2022. <https://doi.org/10.1101/2022.01.10.22269010>
70. Bramante CT, Proper JL, Boulware DR, Karger AB, Murray T, Rao V, et al. Vaccination Against SARS-CoV-2 Is Associated With a Lower Viral Load and Likelihood of Systemic Symptoms. *Open Forum Infect Dis.* 2022; 9: ofac066. <https://doi.org/10.1093/ofid/ofac066> PMID: 35392460
71. Elliott P, Haw D, Wang H, Eales O, Walters CE, Ainslie KEC, et al. Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the Delta variant. *Science.* 2021; 374: eabl9551. <https://doi.org/10.1126/science.abl9551> PMID: 34726481
72. Pritchard E, Matthews PC, Stoesser N, Eyre DW, Gethings O, Vihta K-D, et al. Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. *Nat Med.* 2021; 27: 1370–1378. <https://doi.org/10.1038/s41591-021-01410-w> PMID: 34108716
73. Acharya CB, Schrom J, Mitchell AM, Coil DA, Marquez C, Rojas S, et al. Viral Load Among Vaccinated and Unvaccinated, Asymptomatic and Symptomatic Persons Infected With the SARS-CoV-2 Delta Variant. *Open Forum Infect Dis.* 2022; 9: ofac135. <https://doi.org/10.1093/ofid/ofac135> PMID: 35479304
74. Chia PY, Ong SWX, Chiew CJ, Ang LW, Chavatte J-M, Mak T-M, et al. Virological and serological kinetics of SARS-CoV-2 Delta variant vaccine breakthrough infections: a multicentre cohort study. *Clin Microbiol Infect.* 2022; 28: 612.e1–612.e7. <https://doi.org/10.1016/j.cmi.2021.11.010> PMID: 34826623
75. Singanayagam A, Hakki S, Dunning J, Madon KJ, Crone MA, Koycheva A, et al. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. *The Lancet Infectious Diseases.* 2022. pp. 183–195. [https://doi.org/10.1016/s1473-3099\(21\)00648-4](https://doi.org/10.1016/s1473-3099(21)00648-4) PMID: 34756186
76. Lee S, Kim T, Lee E, Lee C, Kim H, Rhee H, et al. Clinical Course and Molecular Viral Shedding Among Asymptomatic and Symptomatic Patients With SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea. *JAMA Intern Med.* 2020; 180: 1447–1452. <https://doi.org/10.1001/jamainternmed.2020.3862> PMID: 32780793
77. Ra SH, Lim JS, Kim G-U, Kim MJ, Jung J, Kim S-H. Upper respiratory viral load in asymptomatic individuals and mildly symptomatic patients with SARS-CoV-2 infection. *Thorax.* 2021; 76: 61–63. <https://doi.org/10.1136/thoraxjnl-2020-215042> PMID: 32963115
78. Yang Q, Saldi TK, Gonzales PK, Lasda E, Decker CJ, Tat KL, et al. Just 2% of SARS-CoV-2-positive individuals carry 90% of the virus circulating in communities. *Proceedings of the National Academy of Sciences.* 2021. <https://doi.org/10.1073/pnas.2104547118> PMID: 33972412

79. Zuin M, Gentili V, Cervellati C, Rizzo R, Zuliani G. Viral Load Difference between Symptomatic and Asymptomatic COVID-19 Patients: Systematic Review and Meta-Analysis. *Infect Dis Rep*. 2021; 13: 645–653. <https://doi.org/10.3390/idr13030061> PMID: [34287354](#)
80. Han MS, Seong M-W, Kim N, Shin S, Cho SI, Park H, et al. Viral RNA Load in Mildly Symptomatic and Asymptomatic Children with COVID-19, Seoul, South Korea. *Emerg Infect Dis*. 2020; 26: 2497–2499. <https://doi.org/10.3201/eid2610.202449> PMID: [32497001](#)
81. Zhou R, Li F, Chen F, Liu H, Zheng J, Lei C, et al. Viral dynamics in asymptomatic patients with COVID-19. *Int J Infect Dis*. 2020; 96: 288–290. <https://doi.org/10.1016/j.ijid.2020.05.030> PMID: [32437933](#)
82. Lennon NJ, Bhattacharyya RP, Mina MJ, Rehm HL, Hung DT, Smole S, et al. Cross-sectional assessment of SARS-CoV-2 viral load by symptom status in Massachusetts congregate living facilities. *J Infect Dis*. 2021. <https://doi.org/10.1093/infdis/jiab367> PMID: [34255846](#)
83. Euser S, Aronson S, Manders I, van Lelyveld S, Herpers B, Sinnige J, et al. SARS-CoV-2 viral-load distribution reveals that viral loads increase with age: a retrospective cross-sectional cohort study. *Int J Epidemiol*. 2022; 50: 1795–1803. <https://doi.org/10.1093/ije/dyab145> PMID: [34999848](#)
84. Evans D, Cowen S, Kammel M, O'Sullivan DM, Stewart G, Grunert H-P, et al. The Dangers of Using Cq to Quantify Nucleic Acid in Biological Samples: A Lesson From COVID-19. *Clin Chem*. 2021; 68: 153–162. <https://doi.org/10.1093/clinchem/hvab219> PMID: [34633030](#)
85. Michalakakis Y, Sofonea MT, Alizon S, Bravo IG. SARS-CoV-2 viral RNA levels are not “viral load.” *Trends Microbiol*. 2021; 29: 970–972.
86. Douthit N, Kiv S, Dwolatzky T, Biswas S. Exposing some important barriers to health care access in the rural USA. *Public Health*. 2015; 129: 611–620. <https://doi.org/10.1016/j.puhe.2015.04.001> PMID: [26025176](#)
87. Lundberg DJ, Cho A, Raquib R, Nsoesie EO, Wrigley-Field E, Stokes AC. Geographic and Temporal Patterns in Covid-19 Mortality by Race and Ethnicity in the United States from March 2020 to February 2022. *medRxiv*. 2022. <https://doi.org/10.1101/2022.07.20.22277872> PMID: [35898347](#)
88. Orphan CM, McClure KR. An Anchor for the Region: Examining a Regional Comprehensive University's Efforts to Serve its Rural, Appalachian Community. *J Res Rural Educ*. 2019; 35: 1–19.
89. 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: 228–232. <https://doi.org/10.1038/nature16996> PMID: [26840485](#)
90. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc*. 2017; 12: 1261–1276. <https://doi.org/10.1038/nprot.2017.066> PMID: [28538739](#)
91. Faria NR, Kraemer MUG, Hill SC, Goes de Jesus J, Aguiar RS, Iani FCM, et al. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science*. 2018; 361: 894–899. <https://doi.org/10.1126/science.aat7115> PMID: [30139911](#)
92. Kafetzopoulou LE, Pullan ST, Lemey P, Suchard MA, Ehichioya DU, Pahlmann M, et al. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science*. 2019; 363: 74–77. <https://doi.org/10.1126/science.aau9343> PMID: [30606844](#)
93. Tyson JR, James P, Stoddart D, Sparks N, Wickenhagen A, Hall G, et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv*. 2020. <https://doi.org/10.1101/2020.09.04.283077> PMID: [32908977](#)
94. Tegally H, San JE, Cotten M, Moir M, Tegomoh B, Mboowa G, et al. The evolving SARS-CoV-2 epidemic in Africa: Insights from rapidly expanding genomic surveillance. *Science*. 2022; eabq5358. <https://doi.org/10.1126/science.abq5358> PMID: [36108049](#)
95. Freed NE, Vlкова M, Faisal MB, Silander OK. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biology Methods*

& Protocols. 2020; 5. Available:

<https://academic.oup.com/biomethods/article/5/1/bpaa014/5873518>

96. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. *PLoS One*. 2021; 16: e0257521. <https://doi.org/10.1371/journal.pone.0257521> PMID: [34597327](https://pubmed.ncbi.nlm.nih.gov/34597327/)
97. Alpert T, Brito AF, Lasek-Nesselquist E, Rothman J, Valesano AL, MacKay MJ, et al. Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. *Cell*. 2021. pp. 2595–2604.e13. <https://doi.org/10.1016/j.cell.2021.03.061> PMID: [33891875](https://pubmed.ncbi.nlm.nih.gov/33891875/)
98. Moshiri N. ViralMSA: massively scalable reference-guided multiple sequence alignment of viral genomes. *Bioinformatics*. 2021; 37: 714–716. <https://doi.org/10.1093/bioinformatics/btaa743> PMID: [32814953](https://pubmed.ncbi.nlm.nih.gov/32814953/)
99. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014; 30: 3276–3278. <https://doi.org/10.1093/bioinformatics/btu531> PMID: [25095880](https://pubmed.ncbi.nlm.nih.gov/25095880/)
100. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015; 32: 268–274. <https://doi.org/10.1093/molbev/msu300> PMID: [25371430](https://pubmed.ncbi.nlm.nih.gov/25371430/)
101. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol*. 2018; 4: vex042. <https://doi.org/10.1093/ve/vex042> PMID: [29340210](https://pubmed.ncbi.nlm.nih.gov/29340210/)