AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

COMMENTARY

# Collaboration between Clinical and Academic Laboratories for Sequencing SARS-CoV-2 Genomes

**Jeremy Wang,[a] Shawn E. Hawken,[b] Corbin D. Jones,[a] Robert S. Hagan,[c] Frederic Bushman,[d] John Everett,[d] Melissa B. Miller,[b] Kyle G. Rodino[e]**

[a]Department of Genetics, Center for Gastrointestinal Biology and Disease, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[b]Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA
[c]Division of Pulmonary Diseases and Critical Care Medicine, Department of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA
[d]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[e]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

**ABSTRACT** Genomic sequencing of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) continues to provide valuable insight into the ever-changing variant makeup of the COVID-19 pandemic. More than three million SARS-CoV-2 genome sequences have been deposited in Global Initiative on Sharing All Influenza Data (GISAID), but contributions from the United States, particularly through 2020, lagged the global effort. The primary goal of clinical microbiology laboratories is seldom rooted in epidemiologic or public health testing, and many laboratories do not contain in-house sequencing technology. However, we recognized the need for clinical microbiologists to lend expertise, share specimen resources, and partner with academic laboratories and sequencing cores to assist in SARS-CoV-2 epidemiologic sequencing efforts. Here, we describe two clinical and academic laboratory collaborations for SARS-CoV-2 genomic sequencing. We highlight roles of the clinical microbiologists and the academic laboratories, outline best practices, describe two divergent strategies in accomplishing a similar goal, and discuss the challenges with implementing and maintaining such programs.

Beginning in the fall of 2020, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) lineages emerged globally showing evidence for greater transmissibility and disease severity and decreased treatment efficacy (1). Since then, SARS-CoV-2 variants of concern (VOC) have swept the globe, displacing parent SARS-CoV-2 strains and, in the case of the Delta variant (B.1.617.2/AY.*), risen to dominance in many countries. In the United States, Delta now accounts for >99% of all SARS-CoV-2 (2). Increased positivity rates as a consequence of VOC transmission have led to public health interventions, such as the revision of masking guidelines and vaccine mandates (3). However, widespread transmission of SARS-CoV-2 VOC has implications that extend beyond increased case counts. For example, the efficacy of SARS-CoV-2 monoclonal antibody treatment (MAb) and vaccines and the integrity of diagnostic tests are in jeopardy if regions of the genome encoding their targets are altered.

Variants emerge when viruses containing mutations that occur during normal RNA virus replication spread in a population (4). Mutations can occur in antigenic regions of the viral genome, such as in the SARS-CoV-2 spike protein that mediates viral attachment to host cells. The spike protein is the primary target of neutralizing antibodies and vaccines. Thus, immunity after natural infection and vaccination, as well as the efficacy of MAb treatment, may be affected by mutations in the spike coding region (4–6). Already, variants have been recognized that demonstrate potential or observed resistance to MAb treatments, including bamlanivimab, casirivimab, imdevimab, and etesevimab. The FDA has revoked (bamlanivimab) or modified recommendations on their use with severe

**Editor** Romney M. Humphries, Vanderbilt University Medical Center



Address correspondence to Melissa B. Miller, Melissa.Miller@unchealth.unc.edu, or Kyle G. Rodino, Kyle.Rodino@pennmedicine.upenn.edu.

The authors declare no conflict of interest.

*The views expressed in this article do not necessarily reflect the views of the journal or of ASM.*

**Accepted manuscript posted online** 5 January 2022

**Published** 16 March 2022

COVID-19 to include health care provider monitoring of data on currently circulating variants to guide treatment decisions (2, 7, 8). Similarly, SARS-CoV-2 genomic data have already identified several variants with observed or potentially reduced neutralization by postvaccination sera. These data have led to calls for the development of vaccines targeting current variants and long-term strategies to deploy future vaccines to protect against variants that have not yet emerged (9).

Variant tracking is also required for monitoring of the efficacy of diagnostic and surveillance testing for SARS-CoV-2. The FDA has warned that diagnostic SARS-CoV-2 tests have reduced efficacy for the detection of some SARS-CoV-2 variants (10). Mutations that occur at genome target sites for SARS-CoV-2 diagnostics can result in false-negative results, imperiling patient care, case identification, and public health tracking. If a variant has a mutation in a diagnostic target which renders the test ineffective or less sensitive, diagnostic laboratories may be blind to circulating strains, disrupting reporting of positive cases to public health authorities. Monitoring mutations that may impact commercial tests is crucial for maintaining accurate diagnostics in the setting of emerging variants (11). In addition, sequencing samples with negative results from patients with a high clinical suspicion for COVID-19 may identify variants that would otherwise evade detection (12).

Strategies to track current circulation and emergence of variants require robust real-time genomic surveillance data. The use of such data requires the reporting of linked patient metadata to state and national public health authorities. No standardized pipeline exists for genomic data generation, analysis, and reporting at the state and federal level. Throughout the pandemic, the United States has lagged other countries in the proportion of cases sequenced (13). By early 2021, the U.S. SARS-CoV-2 genome sequences in online repositories represented less than 2% of all reported cases. There were vast regional differences in cases sequenced, in part because analysis took place in academic medical centers (14). Although the CDC implemented programs to enhance genomic surveillance, these programs only slightly increased the proportion of cases sequenced in the United States (National SARS-CoV-2 Strain Surveillance, ~750 samples/week) or put the onus on commercial and local public health/hospital laboratories to perform sequencing and variant reporting (14–16).

The emergence of VOC has made it crucial to track emerging variants at local levels in order to facilitate real-time response to increased case counts, monitor diagnostic tests, and inform SARS-CoV-2 treatment decisions. Recently, there has been a federal push to increase sequencing capacity in the United States, with the CDC initially investing $200 million. The focus has been partnerships with commercial and academic laboratories and issuing guidance for standardizing reporting of SARS-CoV-2 sequencing data to public health authorities (14, 17). Additionally, in April 2021, the Biden administration announced that $1.7 billion would be provided to support sequencing and bioinformatics infrastructure for monitoring SARS-CoV-2 variants (18). This federal support for increasing sequencing capacity came with an initial disbursement of between $1 million and $17 million to individual states to support these efforts (18). Although support through federal funding is an excellent first step toward improving genomic surveillance in the United States, most public health laboratories have limited or no capacity for genome sequencing or analysis. Building a robust and responsive genomic surveillance system from the ground up is an expensive and time-consuming undertaking. The ever-changing SARS-CoV-2 pandemic has shown that surveillance cannot wait. In the interim, local partnerships between clinical diagnostic laboratories and academic laboratories with next-generation sequencing (NGS) capacity and bioinformatics expertise are crucial to keep pace with the SARS-CoV-2 pandemic.

## THE ROLE OF CLINICAL MICROBIOLOGISTS

Few clinical microbiology laboratories have the in-house capability or capacity for high-throughput SARS-CoV-2 surveillance sequencing. Collaborating with academic laboratories or university core sequencing facilities with existing equipment and

bioinformatics support is a substitute for this infrastructure gap. Here, we highlight the roles of clinical microbiologists in such partnerships.

**Regulatory requirements, safety, and quality.** Genomic sequencing occurs almost exclusively on residual SARS-CoV-2 diagnostic specimens, making the clinical laboratory a key supporter of epidemiologic and public health initiatives. A clinical laboratory must abide by regulatory requirements when transferring residual clinical samples to non-Clinical Laboratory Improvement Amendments (CLIA) academic laboratories, including maintaining a log of samples shared, specimen deidentification, and other data security measures as defined by the appropriate Institutional Review Board (IRB) approval or exemption. When transferring samples to nonclinical labs, it is also important to address biosafety. Academic laboratories or sequencing cores may have a wide range of experience in handling infectious samples. The clinical microbiologist should offer guidance on appropriate sample handling, ensuring the necessary biosafety equipment (e.g., biosafety cabinet) is available and that sample inactivation occurs appropriately. Similarly, clinical microbiologists can offer advice on workflow and process control, which is gained through the rigors of testing in the CLIA-compliant environment and can benefit the consistency of results in the academic laboratory. A robust, repeatable process is needed to scale with demand and provide sustainability of SARS-CoV-2 sequencing results. This process is particularly important for workflow compatibility if long-term goals include moving the developed assay to the clinical laboratory.

**Identifying samples of significant interest.** While the bulk of SARS-CoV-2 sequencing is performed in an unbiased fashion (i.e., randomly selecting samples to provide a snapshot of circulating variants), there are reasons to target specific samples. Examples include investigations into suspected outbreaks, severe cases in vaccinated individuals, or samples with abnormal test performance (e.g., unusual variance between cycle threshold values of multitarget assays). Notification of these events can come from a variety of sources, including infection preventionists, clinical services, public health agencies, or from within the clinical laboratory. All highlight avenues of communication that are frequently established with the clinical laboratory that may not be in place with the academic laboratory or sequencing core. Additionally, as such conversations may require a review of prior test results, interpretation in the context of clinical history, or an assay quality assurance investigation including troubleshooting with commercial entities, the clinical microbiologist is best qualified to serve as the intermediary by fielding such requests, evaluating, and following up with results as appropriate.

**Reporting and patient-level information.** A challenge of nonclinical, epidemiologic sequencing of SARS-CoV-2 is balancing the perceived clinical need (curiosity) for individualized result reporting while maintaining the appropriate level of patient anonymity across the spectrum of consumers. This dilemma was simplified with the release of Centers for Medicare and Medicaid Services (CMS) guidance on patient-level reporting of non-CLIA SARS-CoV-2 sequencing results, which allowed only for individual reporting to public health agencies and specifically prohibited the return of results to patients and providers (19). At both of our institutions, samples are anonymized prior to transfer to our academic partners and deidentified metadata are uploaded to the appropriate public databases (e.g., Global Initiative on Sharing All Influenza Data [GISAID] and NCBI) and in aggregate to our publicly available SARS-CoV-2 sequencing dashboards, namely, University of North Carolina at Chapel Hill (UNC) (http://unc.cov2seq.org/) and University of Pennsylvania (Penn) (https://microb120.med.upenn.edu/data/SARS-CoV-2/). Even for clinical colleagues and hospital administration, these aggregate data reports provide sufficient information to inform testing strategies or policies on transmission mitigation and educating staff and patients on the current pandemic makeup. We advocate for the clinical microbiologist to be active in these conversations and assist in translating these data for institutional colleagues and policy makers, as interpretation of genomic sequencing data may ultimately impact clinical laboratory operations. In cases where genomic data need to be reconnected to patient information for public health reporting, we have relied on the clinical microbiologist

**TABLE 1** Platform comparison[a]

| Parameter | Data by platform | |
|---|---|---|
| | Illumina | Oxford Nanopore |
| Capital costs[b] | $250,000 (NextSeq) | $1,000 (MinION + computer) |
| Consumable cost per genome | $43.98 | $19.60 |
| RNA extraction materials cost per genome | $11.04 | $3.39 |
| Total cost per genome[c] | $55.02 | $22.99 |
| Turnaround time[d] | 4 days | 21 h |
| Optimum no. of samples per sequencing run | >250 | 96 |

[a]Consumables costs assume that optimal batch size is used for each platform and reflect only the experiences of our respective programs. Realized costs are institution specific depending on equipment and reagents.
[b]Cost reflects equipment used. Alternative platforms may be more comparable in price.
[c]Cost does not include labor.
[d]Turnaround time includes RNA extraction through the construction of the genome sequence and lineage/clade assignment. ONT turnaround time assumes that sequencing is run with real-time basecalling.

for this role. At both of our institutions, the clinical microbiologist serves as the holder of the linkage file, maintaining separation of protected health information (PHI) from the academic laboratory, but allowing patient-level data to be linked for public health purposes, as approved by our respective IRBs. At the current time, a compelling use case for clinically reportable SARS-CoV-2 genomic data is absent. However, we advocate that the clinical microbiologist should remain engaged with these requests and continuously evaluate potential clinical needs. As experts in diagnostics, clinical microbiologists should rationalize testing strategies and justify potential benefits or illustrate current shortcomings.

## THE ROLE OF ACADEMIC LABORATORIES OR GENOMIC CORES

The missions of clinical and academic cores are substantially different. Clinical sequencing is focused narrowly and controlled tightly in both assay selection and implementation. Academic cores, in contrast, are constantly adapting their approaches to the latest technologies and experimental ideas of the researchers they support. Thus, academic and clinical laboratories are kept separate and distinct. The urgent challenge of SARS-CoV-2 strain characterization, however, showed that the complementary strengths of academic sequencing facilities and clinical laboratories could be used to develop assays rapidly and effectively to fill public health needs.

Academic cores typically have the equipment, expertise, and staff to pivot rapidly and tackle a new assay and scale it quickly. Most of the equipment (i.e., robotics, sequencers, and other assays) in academic cores are general purpose. Robotics platforms, for instance, are reprogrammed routinely to accommodate new protocols. Academic centers also host a variety of sequencing platforms, which facilitates finding the right platform at the right scale for an assay. At UNC, for example, several different sequencers were investigated before it was determined that the Oxford Nanopore Technologies (ONT) platform provided the best fit to the turnaround time (TAT), accuracy, and scale needed. Furthermore, the availability of both MinION and GridION platforms at the UNC academic core allowed the team to adjust the scale of the assays rapidly and provide consistently rapid TAT (Table 1). The Penn team found the widely used Illumina technology most convenient, primarily based on the availability of equipment and familiarity with adapting the workflow for multiple applications.

Many large academic cores have staff scientists who routinely assess new and emerging technologies. This experience allows them to rapidly implement and assess recently published assays. For SARS-CoV-2, the urgency of the need for effective sequencing solutions resulted in a bevy of preprints, new kits, and reported best approaches to sequencing and detecting viral strain variation. Investigators and the core staff were able to work quickly and effectively through these approaches to find those that met the needs of both the research and the clinical communities. As demand drove the need for increased sequencing capacity, highly trained core staff could be redirected to support the SARS-

CoV-2 assay work without the need to recruit and hire new staff, which is limited in the clinical setting. Similarly, as demand waned, these staff could be refocused to other work without an institutional loss of knowledge.

As with the wet-bench laboratories, academic cores typically have or work with a team of bioinformaticians to support processing and analysis of data. While the genome of SARS-CoV-2 is small, and the data sets produced by sequencing were small compared with those generated for human and animal model studies, the downstream processing needed to be highly specific. The on-site staff again were able to redirect their efforts to investigating and supporting the best analysis approaches. Additionally, either local or cloud-based solutions are already available at academic cores. At both UNC and Penn, bioinformatics experts used existing infrastructure to support and scale SARS-CoV-2 bioinformatics without the need to purchase additional hardware.

## WORKFLOW EXAMPLES AND BEST PRACTICES

The workflows presented are examples from the SARS-CoV-2 sequencing programs at UNC and Penn. Other methods could also meet the need and have been used at other institutions.

**Sequencing platforms.** The dominant platforms for routine amplicon-based sequencing of SARS-CoV-2 are Illumina and Oxford Nanopore Technologies (ONT) instruments. Both have been adopted worldwide for surveillance sequencing of patient-derived samples. Several trade-offs exist between these technologies, with the most salient being capital cost of the sequencer(s), throughput, cost per sample, and turnaround time (Table 1). We discuss these factors and common use cases below.

ONT sequencing platforms offer an alternative to traditional sequencing-by-synthesis with several advantages and disadvantages. Nanopore sequencing produces long reads (up to megabases) with a mean error rate around 5%. Unlike Illumina, these errors are dominated by short indels, occurring most often in homopolymer stretches. Nanopore sequencing produces reads asynchronously and continuously, enabling real-time data acquisition. Sequence data are generated and can be analyzed immediately, and sequencing can be terminated as soon as enough data are generated. These features lead to a faster turnaround time than is possible with sequencing-by-synthesis platforms. In our hands at UNC, a single flow cell produces enough data for up to 96 samples in under 12 hours. The very low capital investment for the MinION sequencer ($1,000) contributed to its rapid and broad adoption early in the pandemic to perform routine genomic surveillance near the point of collection. A single MinION/GridION flow cell is cost effective for 12 to 96 samples at a time, further reducing the complexity and cost associated with sequencing surveillance in low- and medium-throughput settings, including academic medical centers. The MinION, as opposed to the ONT GridION and PromethION systems, must be attached to a sufficiently powerful computer to enable real-time basecalling and minimize turnaround time. A computer sufficient to perform real-time basecalling for a single MinION can be purchased or purpose built for less than $1,000 (20, 21).

The Illumina method is efficient for larger batches and is the approach favored at Penn. The cost of sequencing instruments is much higher than that for the MinION, but the instruments allow sequencing of larger batches. Typically, ~96 specimens and controls are included in a batch and several batches are combined for sequencing on a NextSeq instrument. Illumina has instruments that permit both smaller (MiSeq and MiniSeq) and larger (NovaSeq) batches. For use of instruments with a larger capacity, upstream steps, such as sample acquisition and processing, often become limiting. Thus, filling up large batches can be slow and progress limiting so that the midcapacity NextSeq instrument is a good fit.

**Data generation pipeline.** Consistent processing and rigorous quality control are critical in both molecular biology protocols and computational analysis to produce reliable, unbiased data for clinical interpretation and local and global public health efforts. To this end, many efficient and reproducible protocols have been developed to sequence SARS-CoV-2 genomes from clinical samples. The most widely used

noncommercial assay is that initially developed by the ARTIC Network (22). The traditional ARTIC protocol amplifies the SARS-CoV-2 genome in 98 partially overlapping segments in 2 nonoverlapping primer pools.

At UNC, the resulting amplicons of ~400 bp each can be sequenced on either Oxford Nanopore or Illumina platforms. In support of this method, a variety of laboratory protocols have been implemented for RNA extraction, reverse transcription, PCR, and library preparation to increase throughput, improve genome recovery, and reduce consumables costs and prep time. Our sequencing and analysis pipeline has evolved as technologies, best practices, and needs have changed. For routine surveillance of known positive samples (primarily nasal or nasopharyngeal swabs), we implemented variations of the ARTIC protocol depending on materials/reagent availability, viral titer, and batch size. For smaller batches (e.g., ≤24 samples), we use one of a range of longer amplicon panels—derived from the full ARTIC set—depending on the sample titer. Longer amplicon tiles produce more even coverage and avoid primer dropouts due to sequence divergence better than panels with more primers, but they require significantly higher starting concentrations of viral RNA. In general, for a cycle threshold ($C_T$) of <30, we use a subset of ARTIC primers targeting ~1.2-Kbp amplicons (23). For a $C_T$ of <20 to 25, our experience is that amplicons of 3 to 5 Kbp can be amplified reliably and further reduce coverage variation, but they are seldom practical for even moderate numbers of samples. For these longer amplicon libraries, we use a transposase-based barcoding kit for Nanopore sequencing, further reducing the time to obtain the genome sequence compared with ligation-based multiplexing. In particular, the hands-on time required for the "rapid" long-amplicon library prep is often almost half that of the full ligation prep required for standard ~400-bp amplicons. For large batches (i.e., 25 to 96 samples) or those with a mixture of low and high $C_T$ (up to ~35), we default to the ARTIC V4 amplicon set followed by "native" ligation barcoding that allows for efficient batch processing and maximizes the recovery of low-titer samples.

At Penn, the ARTIC V4 primers and POLAR protocol were used for all samples (24). Samples were analyzed if they achieved a cycle of threshold of <28 from various swab-based platforms and <20 from saliva-based testing on the Advanta Dx assay (Fluidigm, San Francisco, CA) because these values correlated with acquiring adequate quality sequence and appropriate coverage.

**Minimum quality.** Complete and accurate genomes are necessary for downstream analyses, including identification of mutations, lineage classification, and phylogenetic analysis. Accuracy is typically considered a function of the read depth at each locus and completeness the proportion of the genome meeting this coverage threshold.

At UNC, 20× is a widely used coverage threshold that ensures high consensus accuracy and was implemented in our pipeline (25). Downstream analyses vary somewhat in the proportion of the genome required to make accurate inference. For confident identification of the Pango lineage (and WHO variant classification)—a primary endpoint for clinical and public health usage—this threshold is as low as 70% (30% missing sites/Ns), matching the default threshold for maximum ambiguous loci in the Pangolin lineage inference software. For many aggregate analyses, more conservative thresholds are often used, namely, up to 99%. At UNC, a threshold of 7,000 missing sites (~25%) was used for taking a genome through downstream analysis and submission to public repositories. While clade/lineage assignments can be inaccurate for less complete genome sequences, Pangolin output and confidence values are evaluated carefully to exclude poorly supported or indeterminate lineage calls before reporting. These thresholds (20× over 75% of the genome) are achievable typically for samples with sufficient material ($C_T$, <30). The typical throughput of a MinION/GridION flow cell, namely, ~5 Gbp for a 12-hour run, equates to an average depth of ~1,700× across 96 samples.

At Penn, genomes were accepted for further analysis if they achieved 95% coverage with at least 5 reads per base. Average coverage was much higher, but quality control focused on the weakest part of the data for each genome.

**Informatics and analysis.** Consistent processing and rigorous quality control are equally important in data processing and bioinformatic analysis. Consistent and

transparent processing is critical; data quality issues resulting from low-titer samples, processing variation, and contamination are not always avoidable. A full analytical pipeline consists typically of initial read processing and genome assembly followed by variant and phylogenetic inference and reporting/visualization. Initial data processing steps, including basecalling, demultiplexing, and trimming sequencing adapters, barcodes, and primers, are generic read processing tasks that are performed commonly by academic sequencing cores. A representative and broadly applicable bioinformatic pipeline for sequence processing and assembly is the ARTIC network nCoV bioinformatics standard operating procedure (SOP) (26). The pipeline used at Penn is as previously described (27).

**Data sharing.** To support local and global public health efforts, and in accordance with the World Health Organization guidance, sequences should be shared publicly by submission to appropriate public databases (typically, GISAID and chosen International Nucleotide Sequence Database Collaboration (INSDC), such as NCBI GenBank) with corresponding metadata (28, 29). The public availability of SARS-CoV-2 genomic data in as near real-time as possible—in particular, forgoing an embargo before publication—continues to enable better identification and tracking of viral evolution and transmission patterns that inform public health decision-making.

To support surveillance at an academic health center, provide a resource depicting local SARS-CoV-2 variant makeup, and inform local and state public health agencies, both institutions produce regular reports on aggregate trends, including mutation frequencies and lineages. These results are made available publicly through a Web-based report and visualization tool that additionally present aggregate lineage trends, mutation tracking, and a phylogenetic tree to allow for a more detailed assessment of up-to-date sequence data, for example to identify local clusters (Fig. 1A and B).

## CHALLENGES

While academic-clinical laboratory partnerships highlight the success that can be achieved through collaboration, there are a number of challenges. The overlap of clinical diagnostics, public health, and research creates concerns related to safeguarding protected health information (PHI) and information technology security. When our institutions began these collaborations, no guidance existed regarding how or whether academic laboratories should report sequencing data, how it should be validated, and how it should be submitted to public health authorities. However, both of our institutions committed to SARS-CoV-2 sequencing based on our belief that it was the right thing to do for public health. Subsequently, CMS issued guidance confirming that non-CLIA-certified laboratories are allowed to perform SARS-CoV-2 sequencing on identified patient samples as long as patient-level reports are not issued to patients or providers. CMS, CDC, and the Association of Public Health Laboratories confirmed that non-CLIA laboratories should report patient-level sequencing data to public health authorities (19, 29, 30). If a laboratory reports patient-level sequencing data for a person's diagnosis or treatment, then it must be done in a CLIA-certified laboratory using a CLIA-validated test.

As mentioned above, the link between public health and research facilities can and should be the clinical laboratory. Clinical laboratories handle PHI and public health reporting on a routine basis. By using deidentified but linked identifiers on remnant patient samples, the risk of a confidentiality breach can be minimal when transferring specimens or data to research cores for sequencing or analysis. Secure networked shared drives can be used to transfer data back to the clinical laboratory so that variant sequence data can be linked to the patient and reported to public health authorities. Even though variant detection falls under the umbrella of public health, it is our opinion that Institutional Review Board approval or exemption should be sought to document the safeguards being used and the personnel who have access to PHI.

In recent months, some state health departments have pushed to have variant data reported by Electronic Laboratory Reporting (ELR), similar to SARS-CoV-2 diagnostic test results. While the data are likely more manageable on the public health side with ELR

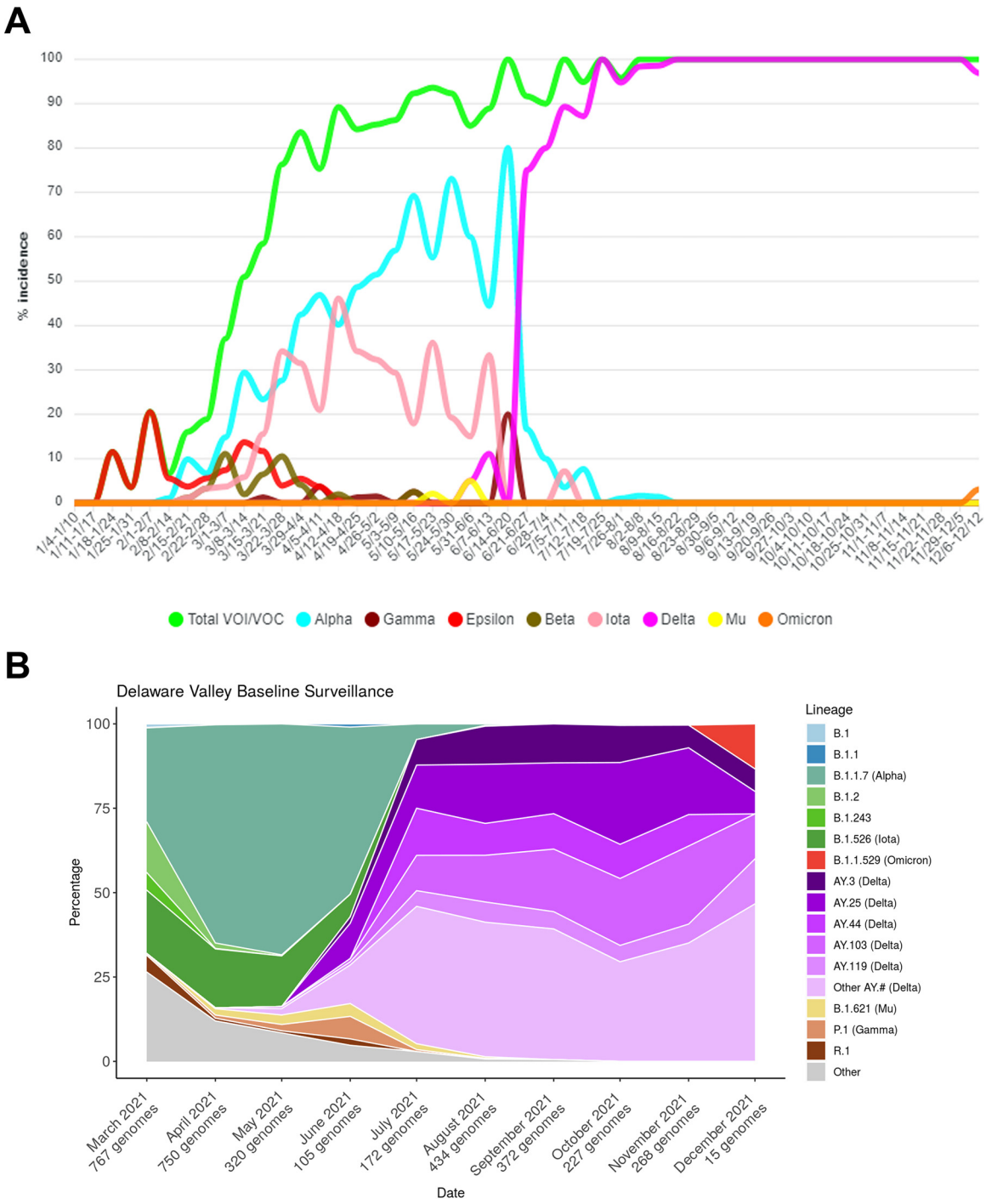


**FIG 1** (A) Trend of variants of interest/variants of concern (VOI/VOC) over time collected from UNC Medical Center as illustrated on the UNC surveillance sequencing dashboard (http://unc.cov2seq.org). (B) SARS-CoV-2 lineage trends of time for samples collected from the University of Pennsylvania Health System and collaborators as illustrated on the Penn Medicine SARS-CoV-2 surveillance sequencing dashboard (https://microb120.med.upenn.edu/data/SARS-CoV-2/).

submission, there are significant concerns from the diagnostic/research perspective. To report through ELR, the variant data (whether just the lineage result or actual sequence data) must be entered into the Electronic Medical Record (EMR), such as EPIC. The result is then linked to a patient record. Even if the result does not cross the interface for providers to see, it is available in the Laboratory Information System (LIS; i.e., EPIC Beaker). When identifiable research data are reported in the LIS, the results are available to anyone with access to the LIS or LIS report building. This availability of data is potentially a

violation of PHI protections. For large health care systems, there are hundreds to thousands of laboratory employees who would have access to this information, of whom many may not have the expertise to interpret data or have a consultant available to assist in interpretation. At both of our institutions, hospitals throughout our health system submit samples for genomic surveillance. We receive calls frequently from a laboratory or provider wanting to know a patient's variant result (which we do not release). If the result is in the LIS for the purpose of ELR, it becomes a clinical test, even if there is not a specific medical intervention associated with the result. However, the majority of laboratories have not performed a CLIA validation for SARS-CoV-2 sequencing and variant identification.

The conundrum of having patient-level sequencing data available for physicians is also complicated by the clinical meaning of the data. Clinical microbiology laboratories are not in the business of doing testing for testing sake. We are thoughtful about the tests we offer and the associated reporting so that the clinical interpretation is meaningful and results provide clinically actionable data. To date, there is not an example of a SARS-CoV-2 lineage that would alter patient care, so as of this writing, it is of no clinical value to report patient-level results. However, the possibility exists that eventually sequence data will provide insights into the activity of oral therapeutics or monoclonal antibody treatments as variants continue to emerge and more therapeutics are available. In the future, there may be scenarios in which it is clinically valuable to have lineage data, similar to when influenza A had both H3 (oseltamivir susceptible) and pre-2009 H1 (oseltamivir resistant) cocirculating. For this reason, the argument for the collaboration of clinical and research/core laboratories is strengthened. The sooner clinical laboratories are included in patient SARS-CoV-2 sequencing efforts, the easier it will be to transition if/when the time comes for a clinical test for SARS-CoV-2 variant reporting.

When thinking of a potential clinically reportable test, issues such as TAT and throughput will have to be considered. SARS-CoV-2 sequencing is not a 1-hour test that can be used simultaneously to detect virus and report a variant, which would be a clinically actionable time frame, when/if indicated. Sequencing laboratories usually get results in 48 to 96 h, but the reality is that sequencing is done weekly to optimize workflow and costs. The longer the time to result, the more limited the clinical utility of results. Nonetheless, sequencing efforts can help inform the development of more targeted diagnostic tests for variant detection, such as real-time PCR (12).

Additional challenges exist related to funding sequencing efforts. Although national programs like CDC SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology and Surveillance (SPHERES) and state-level funding are available, not every laboratory has access to these funds. Clinical laboratories, in particular, are held to a fiscal year budget for new testing initiatives. The budget is tied closely to reimbursement, for which there is currently none specific to SARS-CoV-2 sequencing. Clinical budgets are already under pressure in the COVID-19 era, and it is difficult to obtain financial support for efforts that support public health and/or research efforts but have no patient-level impact or associated billing and reimbursement. Therefore, most clinical-academic SARS-CoV-2 sequencing collaborations rely on funding outside the health care system. Limited and uncertain funding impacts the number of specimens sequenced and the potential sustainability of these collaborations. However, our personal experiences highlight that internal funding can be secured when there is a shared need or common goal, particularly when filling the gap provides broadly beneficial information. Both sequencing programs were funded initially in a grassroots fashion, cobbling together multiple donations and contributions from a variety of departments, centers, and partners, including university offices with sources of philanthropic funding, that spanned the health systems and universities. Cumulatively, the contributions provided support and midrange sustainability to our efforts, ultimately allowing the time and data needed to secure external support.

In addition to funding, limitations in other resources, including personnel, reagents, and equipment, can impact the volume of sequencing that can be performed. Laboratories with limited resources or an overwhelming number of samples may opt to

sequence a fraction (e.g., 10%) or finite number of positive specimens per week. Others with fewer samples or increased capacity may be able to analyze a larger percentage of specimens. Restrictions in capacity will impact the accuracy in providing a snapshot of circulating variants or sensitivity in detecting an emerging variant. Modeling can be used to predict how changes in sampling or volume can impact the confidence in conclusions (31). It is our opinion that performing sequencing is the primary objective, with the ideal volume being secondary. Targets for sequencing capacity should be tailored to the specific institution and situation, maximizing value while sustainably managing resources.

## CONCLUSIONS

We highlight two examples of clinical-academic laboratory partnerships to increase SARS-CoV-2 sequencing and variant monitoring. Our experiences serve as a model for such collaborations but more importantly show the power of using existing expertise from both clinical and academic laboratories to bolster public health reporting. Individually, each laboratory (clinical or academic) would not have been able to develop robust, sustainable programs as quickly as the partnerships. The success of this model was due to the willingness of both parties to provide critical guidance early during assay development, from the flexibility, capacity, and expertise of the academic core and from the diagnostic, PHI, and public health reporting expertise of clinical microbiologists. As we look forward, we need to formalize the establishment of these partnerships to build upon existing public health infrastructure so that we can maintain a scalable surveillance program for emerging infectious diseases and be better prepared for the next pandemic.

## ACKNOWLEDGMENTS

We thank the individuals that provided support for these projects, both in the form of financial contributions and in advocating for our proposals. UNC is grateful for the support of the University Cancer Research Fund and the COVID-19 Research Accelerator Fund who initially funded this initiative and the NC Policy Collaboratory for ongoing funding. Penn is appreciative for the multiple sources of financial support (detailed online at https://microb120.med.upenn.edu/data/SARS-CoV-2/) that enabled our project and the ongoing funding as part of the Centers for Disease Control and Prevention SPHERES Broad Agency Announcement award and the Penn Center for Research on Coronaviruses and Other Emerging Pathogens.

We acknowledge the many members of both clinical and academic laboratories that have contributed to the success and sustainability of our programs. Finally, we thank our colleagues in public health for their willingness to partner with our programs, and we hope these connections will grow in the future.

## REFERENCES

1. Walensky RP, Walke HT, Fauci AS. 2021. SARS-CoV-2 variants of concern in the United States—challenges and opportunities. JAMA 325:1037–1038. https://doi.org/10.1001/jama.2021.2294.
2. Centers for Disease Control and Prevention. 2020. COVID data tracker. Centers for Disease Control and Prevention, Atlanta, GA. https://covid.cdc.gov/covid-data-tracker.
3. Centers for Disease Control and Prevention. 2021. COVID-19 vaccination program operational guidance. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/vaccines/covid-19/covid19-vaccination-guidance.html.
4. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, Consortium C-GU, Peacock SJ, Robertson DL, COVID-19 Genomics UK (COG-UK) Consortium. 2021. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol 19:409–424. https://doi.org/10.1038/s41579-021-00573-0.
5. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD. 2021. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe 29:463–476.e6. https://doi.org/10.1016/j.chom.2021.02.003.
6. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, Duprex WP. 2021. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science 371:1139–1142. https://doi.org/10.1126/science.abf6950.
7. United States Food and Drug Administration. 2021. FDA authorizes revisions to fact sheets to address SARS-CoV-2 variants for monoclonal antibody products under emergency use authorization. United States Food and Drug Administration, Silver Spring, MD. https://www.fda.gov/drugs/drug-safety-and-availability/fda-authorizes-revisions-fact-sheets-address-sars-cov-2-variants-monoclonal-antibody-products-under.
8. United States Food and Drug Administration. 2021. Coronavirus (COVID-19) update: FDA revokes emergency use authorization for monoclonal antibody bamlanivimab. United States Food and Drug Administration, Silver Spring, MD. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-revokes-emergency-use-authorization-monoclonal-antibody-bamlanivimab.
9. Krause PR, Fleming TR, Longini IM, Peto R, Briand S, Heymann DL, Beral V, Snape MD, Rees H, Ropero AM, Balicer RD, Cramer JP, Munoz-Fontela C, Gruber M, Gaspar R, Singh JA, Subbarao K, Van Kerkhove MD, Swaminathan S, Ryan MJ, Henao-Restrepo AM. 2021. SARS-CoV-2 variants and vaccines. N Engl J Med 385:179–186. https://doi.org/10.1056/NEJMsr2105280.

10. United States Food and Drug Administration. 2021. Genetic variants of SARS-CoV-2 may lead to false negative results with molecular tests for detection of SARS-CoV-2—letter to clinical laboratory staff and health care providers. United States Food and Drug Administration, Silver Spring, MD. https://www.fda.gov/medical-devices/letters-health-care-providers/genetic-variants-sars-cov-2-may-lead-false-negative-results-molecular-tests-detection-sars-cov-2.
11. Rhoads DD, Plunkett D, Nakitandwe J, Dempsey A, Tu ZJ, Procop GW, Bosler D, Rubin BP, Loeffelholz MJ, Brock JE. 2021. Endemic SARS-CoV-2 polymorphisms can cause a higher diagnostic target failure rate than estimated by aggregate global sequencing data. J Clin Microbiol 59:e0091321. https://doi.org/10.1128/JCM.00913-21.
12. Greninger AL, Dien Bard J, Colgrove RC, Graf EH, Hanson KE, Hayden MK, Humphries RM, Lowe CF, Miller MB, Pillai DR, Rhoads DD, Yao JD, Lee FM. 2021. Clinical and infection control applications of SARS-CoV-2 genotyping: an IDSA/ASM consensus review document. J Clin Microbiol https://doi.org/10.1128/JCM.01659-21.
13. GISAID. 2021. GISAID—submission tracker global, on/privacy. https://www.gisaid.org/submission-tracker-global/.
14. Maxmen A. 2021. Why US coronavirus tracking can't keep up with concerning variants. Nature 592:336–337. https://doi.org/10.1038/d41586-021-00908-0.
15. Centers for Disease Control and Prevention. 2021. CDC's role in tracking variants. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/coronavirus/2019-ncov/variants/cdc-role-surveillance.html.
16. Centers of Disease Control and Prevention. 2021. SPHERES. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/coronavirus/2019-ncov/variants/spheres.html.
17. United States Department of Health and Human Services. 2021. Biden administration announces actions to expand COVID-19 testing. United States Department of Health and Human Services, Washington, DC. https://www.hhs.gov/about/news/2021/02/17/biden-administration-announces-actions-expand-covid-19-testing.html.
18. White House. 2021. Fact sheet: Biden administration announces $1.7 billion investment to fight COVID-19 variants. https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/16/fact-sheet-biden-administration-announces-1-7-billion-investment-to-fight-covid-19-variants/.
19. Center of Medicare and Medicaid Services. 2021. Does a facility that performs surveillance testing to identify SARS- CoV-2 genetic variants need a CLIA certificate? Center of Medicare and Medicaid Services, Baltimore, MD.
20. Oxford Nanopore Technologies. 2021. MinION Mk1B IT requirements. Oxford Nanopore Technologies, Oxford, United Kingdom. https://community.nanoporetech.com/requirements_documents/minion-it-reqs.pdf.
21. Peresini P, Boza V, Brejova B, Vinar T. 2021. Nanopore base calling on the edge. Bioinformatics 37:4661–4667. https://doi.org/10.1093/bioinformatics/btab528.
22. ARTIC Network. 2021. SARS-CoV-2. https://artic.network/ncov-2019.
23. Freed NE, Vlkova M, Faisal MB, Silander OK. 2020. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore rapid barcoding. Biol Methods Protoc 5:bpaa014. https://doi.org/10.1093/biomethods/bpaa014.
24. Glenn SHB, Durand NC, Mitra N, Godinez PS, Mahajan R, Blackburn A, Colaric ZL, Theisen JWM, Weisz D, Dudchenko O, Gnirke A, Rao SSP, Kaur P, Aiden EL, Presser AA. 2020. A rapid, low cost, and highly sensitive SARS-CoV-2 diagnostic based on whole genome sequencing. bioRxiv https://doi.org/10.1101/2020.04.25.061499.
25. Paden CR, Tao Y, Queen K, Zhang J, Li Y, Uehara A, Tong S. 2020. Rapid, sensitive, full-genome sequencing of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis 26:2401–2405.
26. Loman N, Rowe W, Rambaut A. 2021. ARTIC Network. https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html.
27. Everett J, Hokama P, Roche AM, Reddy S, Hwang Y, Kessler L, Glascock A, Li Y, Whelan JN, Weiss SR, Sherrill-Mix S, McCormick K, Whiteside SA, Graham-Wooten J, Khatib LA, Fitzgerald AS, Collman RG, Bushman F. 2021. SARS-CoV-2 genomic variation in space and time in hospitalized patients in Philadelphia. mBio 12:e03456-20. https://doi.org/10.1128/mBio.03456-20.
28. World Health Organization. 2021. WHO's code of conduct for open and timely sharing of pathogen genetic sequence data 2 during outbreaks of infectious disease. World Health Organization, Geneva, Switzerland.
29. Association for Public Health Laboratories. 2021. Recommendations for SARS-CoV-2 sequence data quality & reporting. Association for Public Health Laboratories, Silver Spring, MD.
30. Centers for Disease Control and Prevention. 2021. Guidance for reporting SARS-CoV-2 sequencing results. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/reporting-sequencing-guidance.html.
31. The University of Texas at Austin COVID-19 Modeling Consortium. 2021. Variant detection calculator. https://covid-19.tacc.utexas.edu/dashboards/variants/.