Sequencing SARS-CoV-2 from Antigen Tests

Ashley Nazario-Toole¹, Holly M. Nguyen¹, Hui Xia¹, Dianne N. Frankel²,³, John W. Kieffer²,³, and Thomas F. Gibbons¹

¹ 59th Medical Wing, Clinical Investigations and Research Support Laboratory, Joint Base San Antonio-Lackland, TX, USA
² Trainee Health Surveillance, 559th Medical Group, THLS, Joint Base San Antonio-Lackland, TX, USA
³ Trainee Health, 559th Medical Group, THLS, Joint Base San Antonio-Lackland, TX, USA

* Corresponding author: thomas.f.gibbons6.civ@mail.mil (TG)
¶ These authors contributed equally to this work.

Disclaimer: The views expressed are those of the author and are not necessarily the official views of, or endorsed by, the U.S. Government, the Department of Defense, or the Department of the Air Force. No Federal endorsement of the platforms to produce this research data is intended. The voluntary, fully informed consent of the subjects used in this research was obtained as required by 32 CFR 219 and DoDI 3216.02_AFI 40-402.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Genomic surveillance empowers agile responses to SARS-CoV-2 by enabling scientists and public health analysts to issue recommendations aimed at slowing transmission, prioritizing contact tracing, and building a robust genomic sequencing surveillance strategy. Since the start of the pandemic, real time RT-PCR diagnostic testing from upper respiratory specimens, such as nasopharyngeal (NP) swabs, has been the standard. Moreover, respiratory samples in viral transport media are the ideal specimen for SARS-CoV-2 whole-genome sequencing (WGS). In early 2021, many clinicians transitioned to antigen-based SARS-CoV-2 detection tests, which use anterior nasal swabs for SARS-CoV-2 antigen detection. Despite this shift in testing methods, the need for whole-genome sequence surveillance remains. Thus, we developed a workflow for whole-genome sequencing with antigen test-derived swabs as an input rather than nasopharyngeal swabs. In this study, we use excess clinical specimens processed using the BinaxNOW™ COVID-19 Ag Card. We demonstrate that whole-genome sequencing from antigen tests is feasible and yields similar results from RT-PCR-based assays utilizing a swab in viral transport media.

Introduction

Early in the pandemic Mercatelli and Giorgi predicted a low mutation rate for SARS-CoV-2 based on whole-genome sequencing of 48,635 specimens [1]. Continued genomic surveillance has revealed a worrying mutation rate of $3.7 \times 10^{-6}$ per nucleotide per cycle [2]. Furthermore, the rate of mutation varies across the SARS-CoV-2 genome, with several sites exhibiting recurrent mutations that, due to strong positive selection, emerged independently a minimum of three times in strains sequenced across the globe [3]. Ongoing genomic surveillance of SARS-CoV-2 is critical for identifying emerging variants [4,5]; driving changes to public health policies [6]; and confirming cases of reinfection [7].
Variants

The CDC defines a variant as a viral genome that has one or more mutations that differentiate it from other variants in circulation [8]. Variants may have different potential impacts on public health. Variants of interest (VOIs) require monitoring. Variants of concern (VOCs) may affect treatment, transmission, or disease severity. Variants of high consequence (VOHC) have significantly reduced effectiveness of prevention measures or therapeutics relative to existing or previous variants. Fortunately, no VOHCs have been identified to date. The following VOCs highlight the challenges faced by clinicians and public health officials: B.1.1.7 (UK) increased transmission, P.1 (Brazilian) reduced serum neutralization, and B.1.351 (South African) reduced vaccine efficacy in clinical trials conducted by Pfizer and Novavax [9–12]. Increased transmission of regional strain variants may prompt enacting or increasing social distancing, mask wearing, and/or travel restriction.

Reinfection

Although exceedingly rare, reinfection has been documented and is a threat to vulnerable populations. Although most suspected cases of reinfection were a resurgence of the same viral strain that initially infected the patient, other cases demonstrate reinfection with genetically distinct genomes [7,13]. Genome surveillance enables us to monitor for cases of reinfection and discern whether these cases are associated with particular variants.

Specimen Sources

Initially, SARS-CoV-2 tests relied on nasopharyngeal swabs placed in viral transport media (VTM) followed by RT-PCR [14–16]. In May 2020, the FDA approved the first emergency use authorization for an antigen test of SARS-CoV-2 [17]. In early 2021, the DoD began using antigen testing on asymptomatic active duty and DoD civilian personnel [18,19]. Antigen tests meet the World Health Organization’s (WHO) definition of an Affordable, Selective and Sensitive, User-
friendly, Rapid and Robust, Equipment-free, Deliverable to end-users (ASSURED) diagnostic tool [20,21]. Thus, antigen testing has been a suitable tool for global monitoring of new SARS-CoV-2 cases. Currently, most pipelines for WGS rely on upper respiratory tract clinical diagnostic samples with high viral loads [22,23]. However, antigen tests use anterior nasal (AN) swabs. NP and AN swabs differ in the location from which specimen is derived from the patient [24–26]. Additionally, NP swabs are stored in up to 3 milliliters (mL) of viral transport media or buffered saline solutions whereas AN swabs are inserted into an antigen card with only a few drops of proprietary extraction buffer [16,25]. Finally, after collection, samples in viral transport media are stored in freezers; where antigen cards may be stored at a range of temperatures. All of these pre-analytical factors present concerns regarding the quality of specimen to be used for whole-genome sequencing.

**Sequencing considerations**

To continue whole-genome sequences at institutions that have adopted antigen testing, communication with clinicians is required to minimize pre-analytical factors that may lead to sample degradation or discarding of samples. Institutions that choose to adopt antigen-based whole genome sequence pipelines must work with clinicians and institutional review boards (IRBs) to ensure that clinical specimens are retained for genome surveillance purposes. Clinical researchers must also work with clinicians to ensure proper storage of antigen cards prior to transportation to the sequencing laboratory.

Here, we tested various collection and storage parameters to optimize sample preparation for SARS-CoV-2 whole genome sequencing from Ag-cards. We demonstrate that it is possible to produce high quality genomic surveillance data from antigen-derived AN swabs. Specifically, we validated PCR-based whole genome sequencing from AN swabs from the BinaxNOW™ COVID-19 Ag Card (Abbott Diagnostics Scarborough).
Results

Workflow

Our optimized workflow, shows how de-identified, excess clinical specimens of positive BinaxNOW™ COVID-19 Ag Cards are processed (Figure 1). Per CDC [18,27] and BMBL 6th Edition [28] guidelines for handling SARS-CoV-2, the antigen cards are handled in a biosafety cabinet in a biosafety level 2 (BSL2) room. Flocked swabs are removed from the antigen card and placed into 500 µL of DNA/RNA Shield in a 15 mL conical tube. Swabs are then broken at the breakpoint, the tube cap securely fastened, and the sample vortexed vigorously for 15-30 seconds. Samples are stored at 4°C until RNA extraction. Viral load is determined using a modified version of the CDC research use only (ROU) real-time RT-PCR assay, which targets the viral nucleocapsid (CDC 2019-nCoV N1) and human RNase P (RP). Samples with an N1 cycle threshold (C_T) less than 25 have sufficient quantity of SARS-CoV-2 RNA for sequencing library preparation (Paragon Genomics CleanPlex Flex SARS-CoV-2 Panel). Prepared libraries are checked using fragment analysis and a library quality ratio score is calculated as previously described [29]. Briefly, quality ratio scores are calculated by dividing the concentration [ng/µL] of target amplicons (fragments 250 base pairs (bp) to 350 bp) by the concentration of nonspecific bands (50 base pairs (bp) to 190 bps). Samples with a quality score greater than or equal to one are sequenced using Illumina’s MiSeq or NextSeq500 systems. Finally, sequence data is processed through a user-defined bioinformatics pipeline.

Figure 1. Workflow. (A) BinaxNOW™ antigen cards are run per manufacturer guidelines. (B-C) Swabs are removed, placed in 500 µL DNA/RNA shield, broken at the breakpoint, capped and vortexed in a biosafety cabinet (23). (D) RNA is extracted. (E) RT-PCR is used to measure viral load as N1 C_T. (F) Library preparations are made for samples with C_T values less than 25. (G) Fragment analysis is used to check library quality and (H) samples with a quality score of equal
to or greater than 1 are sequenced. (I) An in-house bioinformatics pipeline is used to process sequence data.

**Sample Preparation Optimization**

To optimize the sample collection and preparation methods, we utilized a previously sequenced SARS-CoV-2 positive NP specimen with an N1 Cₜ of 12.28 (#5195) (Fig 2). First, we examined the potential for sample loss and degradation due to (1) exposure to the BinaxNOW™ proprietary extraction buffer and (2) storage conditions (Fig 2A). In this, and every following experiment, we included a standardized positive control, water extraction, and no template control. The positive control result is shown; however, the water extraction and template controls are not shown, due to their expected and observed lack of amplification. A BinaxNOW™ COVID-19 Ag Card swab was dipped into NP specimen # 5195 and then placed into 500 µL DNA/RNA shield (swab, no Ag test). For comparison, a second swab dipped into the same NP specimen, run through the antigen test for 15 min per manufacturer’s instructions, and then placed into DNA/RNA shield (swab). The N1 Cₜ of the “swab, no Ag test” sample was 15.75 while the N1 Cₜ of the “swab” sample was 15.82, indicating that little to no viral RNA was lost after 15 min exposure to the Ag-card extraction buffer. To test if prolonged storage of antigen cards might affect sample yield, a third swab was dipped into the NP specimen, run through the antigen test, and the whole card was stored at 4°C for 2 hours before the swab was placed in DNA/RNA Shield (swab & storage). Again, RT-PCR results revealed that there was no sample degradation, as observed through N1 Cₜ values of 15.82 (swab) and 15.60 (swab & storage). PCR-amplicon sequencing libraries were then prepared for each sample using the Paragon Genomics CleanPlex Flex SARS-CoV-2 Panel. For a library prep and sequencing control, we also prepared a library using commercially available SARS-CoV-2 genomic RNA (ATCC # VR-1986D, SARS-CoV-2 Isolate USA-WA1/2020). Library quality scores were calculated as previously described and samples were sequenced at 2x151 bp reads on the Illumina MiSeq system. N1 Cₜ, library quality ratio scores, 20X genome coverage...
(20 or more reads mapped per nucleotide), and viral PANGO lineages are shown in Figure 2B. As expected, the positive control VR-1986D was assigned to PANGO lineage A and all dipped Ag-card specimens were assigned to PANGO lineage B.1, the same lineage assigned to NP specimen 5195. IGV snapshots show that all samples map at greater than 99.5% at 20X coverage across the genome (Fig 2C). Together these finding indicated that sample exposure to Ag-card buffer and testing conditions do not impact the quality of RNA and subsequent sequencing steps.

**Figure 2.** Assay Development. (A) An NP specimen was used to evaluate the feasibility of obtaining viral RNA from Ag-card derived specimens. RT-PCR was carried out on specimens collected under each condition. (B) RT-PCR N1 and RP Ct values, sequence library quality scores, and viral PANGO lineage assignments of antigen card specimens and reference NP specimen #5195. (C) IGV screen shots of sample SARS-CoV-2 genome coverage. VR-1986D = Positive Control Genomic RNA from SARS-CoV-2, Isolate USA-WA1/2020.

**Performance Comparison**

Given that we carried out our preliminary tests in a mock fashion by dipping swabs in a previously tested NP specimen, we next utilized COVID-19 positive Ag-cards to test the viability of sequencing from real specimens (excess clinical specimens collected under IRB-approved protocol FWH20200103E). Seven completed SARS-CoV-2 positive BinaxNOW™ COVID-19 Ag Cards were delivered to the lab and immediately processed as follows: Antigen cards were disassembled and the AN swabs and positive line of the lateral flow strips were separately placed into DNA/RNA shield, stored at 4°C for 48 hours, extracted, then amplified using RT-PCR (Fig 3A). We hypothesized that due to the nature of the lateral flow assay, viral RNA would be concentrated on the positive line of the strip. We found that swabs slightly outperformed the positive line by yielding higher viral loads in all samples. Specifically, swab N1 C\(_r\) values were,
on average, 1.57 \( C_T \)s lower than the line (± 1.16 \( C_T \)s) (p-value = 0.0116). Interestingly, host RNA levels were significantly higher in the swabs: RP was detected in swab specimens 4.43 cycles sooner than the line (± 1.30 \( C_T \)s) (p-value = 0.0001). Thus, relative to host RNA, viral RNA is enriched on the positive line. As we were able to detect comparable levels of viral RNA on swabs and positive lines, we next prepared libraries from three samples with N1 \( C_T \) < 25 from both the swab and positive line (Fig 3B and 3C). Regardless of the sample source, sequenced specimens had 20x genome coverage (20 or more reads per nucleotide) of greater than or equal to 99%, signifying that both the swab and positive line from antigen cards can be used for whole-genome sequencing with no loss in coverage.

**Figure 3.** Sample source performance comparison. (A) SARS-CoV-2 positive BinaxNOW™ COVID-19 Ag Cards were used to evaluate which part of the card, swab or lateral flow positive line, yielded the highest quantity and quality of viral RNA. Extracted RNA was tested for N1 and RP levels using RT-PCR. The average N1 and RP \( C_T \) values are plotted. n = 7 cards. Statistical analysis = Two-tailed, paired \( t \)-test, * p<0.05; *** p<0.0005 (B) N1 and RP Ct values, library quality scores and PANGO lineage assignments. (C) IGV screen shots of SARS-CoV-2 genome coverage.

**Discussion**

When the COVID-19 pandemic started, highly sensitive and specific real-time RT-PCR tests were the first to be developed, quickly becoming the mainstay of diagnostic testing [30,31]. However, RT-PCR tests require specialized kits, equipment, personnel, and laboratories. These factors lead to slower turnaround times as compared to recently released rapid diagnostic tests - which can be readily employed to screen populations such as basic military trainees. With more clinicians
and point-of-care testing sites adopting antigen-based SARS-CoV-2 tests, fewer RT-PCR specimens may be available to support genomic surveillance of viral variants worldwide. To facilitate future genomic surveillance efforts, we developed a workflow for SARS-CoV-2 whole-genome sequencing from antigen-based test specimens (Fig 1). We were able to recover viral RNA of sufficient quantity and quality for targeted sequencing of the SARS-CoV-2 genome (Fig 2). We found that the quality of genome sequences derived from Ag-test samples is comparable to RNA isolated from NP swabs collected for RT-PCR (Fig 2). Furthermore, a comparison of sample sources, antigen card swab vs. lateral flow assay (LFA) positive line, showed that viral RNA from both sources can generate high quality sequencing libraries (Fig 3). For ease of use and biosafety concerns, we recommend collecting the swab over the LFA positive line as both specimen types produced comparable libraries.

Antigen card specimens present a few limitations when compared to NP swabs collected in viral transport media (VTM). First, in the workflow presented here, antigen test swabs are stored in 500 µL of DNA/RNA shield while NP swabs are stored in up to 3 mL of VTM. Multiple RNA extractions can be carried out from VTM specimens after the initial round RT-PCR testing but, antigen swab-derived viral RNA can only be extracted once. Next, variants with novel or interesting mutations can be cultured from VTM; enabling downstream biochemical or viral neutralization assays. In contrast, antigen tests immediately expose the specimen to an extraction buffer which disrupts the viral membrane, releases viral RNA, and enables the presentation of viral nucleocapsid antigens to anti-nucleocapsid antibodies on the LFA positive line. Thus, viruses collected post antigen-test are most likely non-culturable, however we did not evaluate cultivability.

Here we only attempted to sequence from the BinaxNOW™ antigen test, but other antigen tests may also yield viable specimens for whole-genome sequencing. This work is a first step for future
studies examining the sequencing utility of specimens collected for other antigen or rapid molecular tests.

Materials and Methods

Specimen Acquisition

The BinaxNOW™ COVID-19 Ag Card (Abbott) was used to test basic medical trainees for SARS-CoV-2. Excess clinical specimens were de-identified, the entire card placed in a plastic biohazard bag, and samples transferred to the Clinical Investigations and Research Support (CIRS) laboratory. Excess clinical specimens were obtained under an institutional review board (IRB) exempt protocol (IRB reference number FWH20200103E).

Specimen Preparation

The antigen card was disassembled in a biosafety cabinet. The swab was placed in a 15 mL conical tube with 500 µL DNA/RNA Shield (Zymo Research, Catalog # R1100) and stored at 4°C until the RNA was extracted. The positive line of the LFA test strip was excised and placed in 500 µL of DNA/RNA shield and stored in a 2 mL cryovial at 4°C until the RNA was extracted.

RNA Extraction

Samples were extracted using the EZ1 Virus Mini Kit v2.0 (Qiagen, Catalog # 955134), per the manual. Briefly, 400 µL of sample was extracted and eluted into 60 µL AVE buffer (supplied in the kit). The following controls were run with each extraction: a positive control – a nasopharyngeal swab of SARS-CoV-2 diluted to achieve a C_T of approximately 25 - and a negative control consisting of 200 µL nuclease-free water and 200 µL of DNA/RNA Shield. The extracted RNA was frozen at -80°C or used immediately for RT-PCR.
RT-PCR

The following primers were utilized: N1 Forward, 5'-GACCCCCAAAATCAGCGAAAT-3', N1 Reverse, 5'-TCTGGTTACTGCCAGTTGAATCTG-3', N1 FAM probe, 5'-(FAM)ACCCCGCATTACGTTTGGTGGACC(3'-BHQ-1)-3', RP Forward, 5'-AGATTTGGACCTGCGAGCG-3', RP Reverse, 5'-GAGCGGCTGTCTCCACAAGT-3', and RP Cy5 Probe 5'-(Cy5)TTCTGACCTGAAGGCTCTGCGCG(3'-BHQ-3)-3'. 20 µL reactions (15 µL master mix + 5 µL RNA) were prepared using TaqPath™ 1-Step Multiplex Master Mix (No ROX) (ThermoFisher Cat. # A28523) and 20X primer/probe mix. The final primer concentrations per reaction were: N1 Forward & Reverse Primers (400 nM), N1 Probe (200 nM), RF Forward & Reverse Primers (200 nM), and RP Probe (100 nM). The plate was run on the QuantStudio 7 under the following conditions: 25°C for 2 min, 53°C for 10 min, 95°C for 2 min, and 45 cycles of 95°C for 3 sec then 60°C for 30 sec. Fluorescence was detected at the end of each 60°C cycle.

Library Preparation and Sequencing

Paragon Genomics’ CleanPlex® FLEX SARS-CoV-2 Panel (Cat. 918014) for Illumina platforms was used to prepare sequencing libraries (starting concentration of 10-50 ng RNA per sample). As a positive control, sequencing libraries were also prepped for VR-1986D, Genomic RNA from SARS-Related Coronavirus 2, Isolate USA-WA1/2020. Library quality and concentration was assessed via fragment analysis using Advanced Analytics’ High Sensitivity NGS Fragment Analysis Kit (Cat. DNF-474-0500). Library quality ratio scores (QRS) were determined by dividing the fragment analysis 250-350 bp peak concentration (ng/µL) by 50-190 bp peak concentration (ng/µL): excellent (QRS >10), Good (QRS 1.0 – 10), Fair (QRS <1 and >0.5), Poor (QRS < 0.5). Libraries with QRS > 1.0 were denatured and diluted to a final loading concentration of 10 pM following the Illumina MiSeq System Denature and Dilute Libraries Guide (Document # 15039740 v10), and then sequenced on the MiSeq system at 2 x 151 bp using the MiSeq v3, (600 cycle) kit (Illumina, Cat. MS-102-3003).
Bioinformatics

Illumina adaptor sequences were trimmed using the BaseSpace Onsite FASTQ Toolkit v1.0.0. Adapter trimmed FASTQ files were aligned to the SARS-CoV-2 reference genome (NC_045512.2) using Illumina's DRAGEN Bio-IT Platform. Primer sequences were removed using the Linux environment fgbio toolkit (bcftools) and a tab delimited file with primer genomic coordinates provided by Paragon Genomics. The DRAGEN was used to create variant call files from primer trimmed BAM files and consensus FASTA files were created using the fgbio toolkit. Genome coverage uniformity and mapping was visualized in IGV (BAM and VCF files).

Acknowledgements

The authors would like to thank the laboratories who provided the excess clinical specimens used in this research.

References


**Sample Acquisition and RNA Extraction**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess Clinical Specimen</td>
<td>Sample Preservation</td>
<td>Sample Mixing</td>
<td>RNA Extraction</td>
</tr>
</tbody>
</table>

**Quality Checks, Library Preparation, and Sequencing**

<table>
<thead>
<tr>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT-PCR</td>
<td>Library Preparation</td>
<td>Fragment Analyzer</td>
<td>Sequencing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trim Adaptors</td>
</tr>
</tbody>
</table>

**Figure 1.** Workflow. (A) BinaxNOW™ antigen cards are run per manufacturer guidelines. (B-C) Swabs are removed, placed in 500 µL DNA/RNA shield, broken at the breakpoint, capped, and vortexed in a biosafety cabinet (23). (D) RNA is extracted. (E) RT-PCR is used to measure viral load as N1 C_T. (F) Library preparations are made for samples with C_T values less than 25. (G) Fragment analysis is used to check library quality and (H) samples with a quality score of equal to or greater than 1 are sequenced. (I) An in-house bioinformatics pipeline is used to process sequence data.
**Figure 2.** Assay Development. (A) An NP specimen was used to evaluate the feasibility of obtaining viral RNA from Ag-card derived specimens. RT-PCR was carried out on specimens collected under each condition. (B) RT-PCR N1 and RP Ct values, sequence library quality scores, and viral PANGO lineage assignments of antigen card specimens and reference NP specimen #5195 (C) IGV screen shots of sample SARS-CoV-2 genome coverage. VR-1986D = Positive Control Genomic RNA from SARS-Cov-2, Isolate USA-WA1/2020.
Figure 3. Sample source performance comparison. (A) SARS-CoV-2 positive BinaxNOW™ COVID-19 Ag Cards were used to evaluate which part of the card, swab or lateral flow strip positive line, yielded the highest quantity and quality of viral RNA. Extracted RNA was tested for N1 and RP levels using RT-PCR. The average N1 and RP C_T values are plotted. n = 7 cards. Statistical analysis = Two-tailed, paired t-test, * p<0.05; *** p<0.0005 (B) N1 and RP C_T values, library quality scores and PANGO lineage assignments. (C) IGV screen shots of SARS-CoV-2 genome coverage.