



METHOD ARTICLE

Sequencing the pandemic: rapid and high-throughput processing and analysis of COVID-19 clinical samples for 21st century public health [version 1; peer review: awaiting peer review]

Megan L Folkerts ¹, Darrin Lemmer¹, Ashlyn Pfeiffer¹, Danielle Vasquez¹, Chris French¹, Amber Jones¹, Marjorie Nguyen¹, Brendan Larsen², W. Tanner Porter¹, Krystal Sheridan¹, Jolene R. Bowers¹, David M. Engelthaler¹

¹Pathogen Genomics Division, Translational Genomics Research Institute, Flagstaff, AZ, 86005, USA

²Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, 85721, USA

V1 First published: 26 Jan 2021, 10:48
<https://doi.org/10.12688/f1000research.28352.1>

Latest published: 26 Jan 2021, 10:48
<https://doi.org/10.12688/f1000research.28352.1>

Open Peer Review

Reviewer Status AWAITING PEER REVIEW

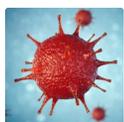
Any reports and responses or comments on the article can be found at the end of the article.

Abstract

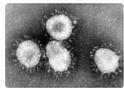
Genomic epidemiology has proven successful for real-time and retrospective monitoring of small and large-scale outbreaks. Here, we report two genomic sequencing and analysis strategies for rapid-turnaround or high-throughput processing of metagenomic samples. The rapid-turnaround method was designed to provide a quick phylogenetic snapshot of samples at the heart of active outbreaks, and has a total turnaround time of <48 hours from raw sample to analyzed data. The high-throughput method was designed for semi-retrospective data analysis, and is both cost effective and highly scalable. Though these methods were developed and utilized for the SARS-CoV-2 pandemic response in Arizona, U.S, and we envision their use for infectious disease epidemiology in the 21st Century.

Keywords

Genomic epidemiology, SARS-CoV2, targeted genomics, sequencing methods, phylogenetics,



This article is included in the [Disease Outbreaks gateway](#).



This article is included in the **Coronavirus** collection.

Corresponding author: Megan L Folkerts (mfolkerts@tgen.org)

Author roles: **Folkerts ML:** Conceptualization, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Lemmer D:** Data Curation, Formal Analysis, Investigation, Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Pfeiffer A:** Data Curation, Investigation; **Vasquez D:** Investigation; **French C:** Data Curation, Formal Analysis, Software; **Jones A:** Investigation; **Nguyen M:** Investigation; **Larsen B:** Supervision; **Porter WT:** Formal Analysis, Visualization, Writing – Review & Editing; **Sheridan K:** Investigation; **Bowers JR:** Conceptualization, Data Curation, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Engelthaler DM:** Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was funded in part by the NARBHA Institute and the Arizona Department of Health Services. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2021 Folkerts ML *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Folkerts ML, Lemmer D, Pfeiffer A *et al.* **Sequencing the pandemic: rapid and high-throughput processing and analysis of COVID-19 clinical samples for 21st century public health [version 1; peer review: awaiting peer review]** F1000Research 2021, 10:48 <https://doi.org/10.12688/f1000research.28352.1>

First published: 26 Jan 2021, 10:48 <https://doi.org/10.12688/f1000research.28352.1>

Introduction

With the advent of rapid and inexpensive next-generation sequencing, genomic epidemiology has proven to be an invaluable resource for the elucidation of disease outbreaks. Extending beyond traditional shoe-leather approaches, rapid-turnaround sequencing methods have allowed researchers to quickly gain insight into the genetic nature of pathogens at the heart of active outbreaks¹⁻⁶. By monitoring pathogen evolution over the course of an outbreak, large-scale genomics have the potential to allow for transmission mapping for infection control and prevention^{7,8}, to distinguish independent cases from those part of active clusters⁹, and to identify epidemiological patterns in time and space on both local and global scales^{7,10-12}.

The most recent example of this has been the collaborative genomic efforts mounted in response to the SARS-CoV-2 outbreak. Not long after the initial cases were identified, whole-genome sequencing quickly established the etiologic agent as a novel coronavirus¹³, the origins of which have since been elucidated¹⁴. Following the rapid spread of SARS-CoV-2, current next-generation technology and analysis pipelines allowed viral sequencing to take place on an unprecedented global scale, with collaborative consortia forming world-wide for the specific purpose of tracking and monitoring the pandemic¹⁵⁻¹⁷.

In most instances, including with the SARS-CoV-2 outbreak, genomic epidemiology has provided a retrospective view of pathogen spread and evolution well after the information is useful in the public health response to the outbreak^{1,2}. Genomic epidemiology should guide contemporaneous outbreak control measures, but can only do so if the data are generated and interpreted in real-time quickly enough to inform a response¹⁸. As technology has advanced, the potential exists to move beyond providing a retrospective genomic snapshot of an outbreak months after its occurrence, to providing actionable data in real-time for current outbreaks within hours after cases are identified⁴. Real-time genomic tracking has already proven valuable in a number of instances, including the recent West-African Ebola outbreak^{3,18}.

This is not to discredit the value of large-scale, retrospective studies. While rapid-turnaround genomics may prove essential for outbreak containment, retrospective studies will continue to be necessary to track pathogen evolution, gauge success of public health interventions, and to evaluate pathogen/host movement and behavior. With the recent SARS-CoV-2 pandemic, retrospective studies have so-far proven successful in identifying the timing and sources of outbreaks on a local¹⁹ and global scale^{5,20}, in evaluating the effectiveness of early interventions⁵, and in identifying super-spreader events²¹. Thus, in addition to real-time monitoring, high-throughput, cost-effective sequencing and analysis are needed to gain a better understanding of pandemics.

Here, we report two Illumina-based sequencing and analysis strategies for either real-time monitoring or large-scale, high-throughput targeted genomic sequencing of complex samples.

Though these strategies were developed for use with the current SARS-CoV-2 pandemic, we envision their potential use in any situation in which a genomic response is needed.

Methods

Sample information

Remnant nasopharyngeal swab specimens or extracted RNA were obtained from, or received by, the TGen North Clinical Laboratory (TNCL) in Flagstaff, AZ. All samples had previously tested positive for SARS-CoV-2 by RT-PCR.

RNA extraction

RNA was extracted using the MagMax Viral Pathogen II kit and a Kingfisher Flex automated liquid handler (ThermoFisher Scientific), with a DNase treatment incorporated to maximize viral RNA recovery from low viral-burden samples, defined as having an RT-PCR cycle-threshold (Ct) value above 33.0. These methods allowed for the rapid, scalable processing of a small number to hundreds of samples at once with minimal personnel, and prevented RNA extraction from becoming a bottleneck to overall throughput (<https://doi.org/10.17504/protocols.io.bnkhmct6>). Remnant RNA was obtained from TNCL, and had been extracted following their FDA emergency use authorization protocol for the diagnosis of COVID-19.

Targeted amplification

SARS-CoV2 RNA was amplified for both of the sequencing methods described below following the nCoV-2019 sequencing protocol V.1²² and using the ARTIC v3 primer set²³. Adapters were added to the resulting amplicons by one of the following means described below. The full sample processing workflow, starting with raw RNA and ending with deliverable data, is illustrated in [Figure 1](#) for reference.

Rapid-turnaround adapter addition and sequencing

For samples requiring immediate attention, e.g. those from patients potentially involved in an active outbreak, adapters were added with the DNA Prep kit (Illumina) as previously described²⁴. Amplicons were sequenced on the MiSeq platform, using a Nano 500 cycle kit with v2 chemistry (Illumina) (<https://doi.org/10.17504/protocols.io.bnnbmdan>).

High-throughput adapter addition and sequencing

In instances where retrospective data were needed from large numbers of samples, adapters were added with the plex-Well384 kit (Seqwell). Samples were multiplexed in batches of 1,152 and sequenced on a NextSeq 550 with v2 chemistry and 150 X 150 bp reads (Illumina). When batch sizes were not large enough to fill a NextSeq run, samples were sequenced on a MiSeq, with V3 chemistry (Illumina) (<https://doi.org/10.17504/protocols.io.bnkimcue>).

Data processing and analysis

Virus genome consensus sequences were built using the Amplicon Sequencing Analysis Pipeline (ASAP)^{25,26}. First, reads were adapter-trimmed using *bbduk*²⁷, and mapped to the Wuhan-Hu-1 genome²⁸ with *bwa mem*²⁹ using local alignment with soft-clipping. Bam alignment files were then processed to

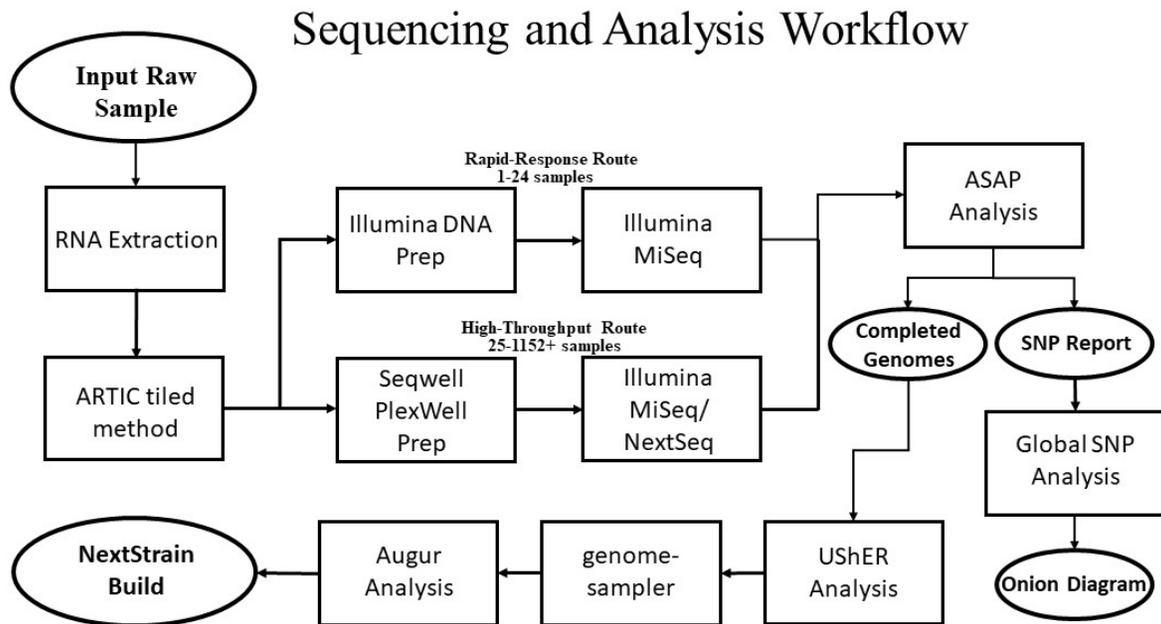


Figure 1. General workflow for the processing of complex samples. Basic workflow for processing of either high-throughput or rapid-response samples.

generate the consensus sequence and statistics on the quality of the assembly by the following: 1) Individual basecalls with a quality score below 20 were discarded. 2) Remaining basecalls at each position were tallied. 3) If coverage $\geq 10X$ and $\geq 80\%$ of the read basecalls agreed, a consensus basecall was made. 4) If either of these parameters were not met, an ‘N’ consensus call was made. 5) Deletions within reads, as called during the alignment, were left out of the assembly, while gaps in coverage (usually the result of a missing amplicon) were denoted by lowercase ‘n’s. Only consensus genomes covering at least 90% of the reference genome with an average depth of $\geq 30X$ were used in subsequent analyses. Consensus genomes generated using these methods include those from Ladner *et al.*¹⁹. Statistics reported for each sample included: total reads, number of reads aligned to reference, percent of reads aligned to reference, coverage breadth, average depth, and any SNPs and INDELs found in $\geq 10\%$ of the reads at that position.

Phylogenetic inference

Rapid phylogenetic inference was used when it was critical to rapidly answer two initial genomic epidemiological questions: i) are the samples in this set closely related, and ii) to what samples in the public database are these most closely related? To answer the former, SNP comparisons were made among a sample set of interest and output in text format directly from the ASAP results, foregoing the computational time of generating phylogenetic trees.

To answer the latter, a database of all SNPs in the GISAID global collection¹⁶ was generated and periodically updated by downloading all the genomes and filtering them for quality

and completeness, then aligning to the WuHan-Hu-1 reference to identify any variants, as described³⁰. The list of all variants and metadata for each sample were then put into a relational database for fast querying. The list of SNPs common to samples in a given set was then compared to this global SNP database using a custom script to determine how many GISAID genomes shared all or a subset of the SNPs from the set. A stacked Venn diagram to illustrate globally-shared SNPs was constructed using the statistical package “R” (Version 4.0.2)³¹ and the “ggplot2”³² and “ggforce”³³ R-packages.

For subsequent, more sophisticated phylogenetic analyses, phylogenetic trees were constructed in NextStrain¹⁷, with genomes from GISAID subsampled by uploading our genomes of interest to the UCSC UShER (UShER: Ultrafast Sample placement on Existing tRee) tool³⁴, identifying relevant genomes (those most related to our genomes of interest), and further reducing that set of genomes when necessary with genome-sampler³⁵.

Consent

Samples used were remnant, de-identified samples from a clinical diagnostics lab. No ethical approval was required for their inclusion in this study.

Results

High-throughput workflow

For the Seqwell workflow, turnaround time from raw sample to sequence data was approximately 72 hours, and to phylogenetic inference approximately 10 hours, for 1,152 samples (Table 1). Cost per sample was ~\$50, which is comparable to similar sample prep methods (Table 1)³⁶. Success rates, defined

as the percentage of samples with greater than 90% genome coverage, were similar to other previously described methods^{36,37}. Of a random subset of 897 samples analyzed, approximately 83% of samples with a Ct <33 yielded genomes with ≥90%

breadth of coverage of the reference genome, i.e., a complete genome, with an average breadth of coverage of the reference genome of 91%. (Figure 2, Table 2). Failure to obtain a complete genome when Ct was <33 was attributed to sample

Table 1. Cost and labor analysis of two Illumina-based prep methods for complex sample processing. Workflow metrics for Illumina’s DNA prep method, and Seqwell’s plexWell method.

*Listed is the number of samples that can be processed within the specified turnaround time, on a single sequencing run. This is not necessarily the upper limit of either processing system.

**Time from raw sample to analyzed data

Sequencing method	Samples able to be processed*	Cost per Sample	Personnel Needed	Turnaround time**
Illumina DNA Prep	24	\$93.54	1	48hrs
Seqwell plexWell	1152	\$50.55	4-6	82hrs

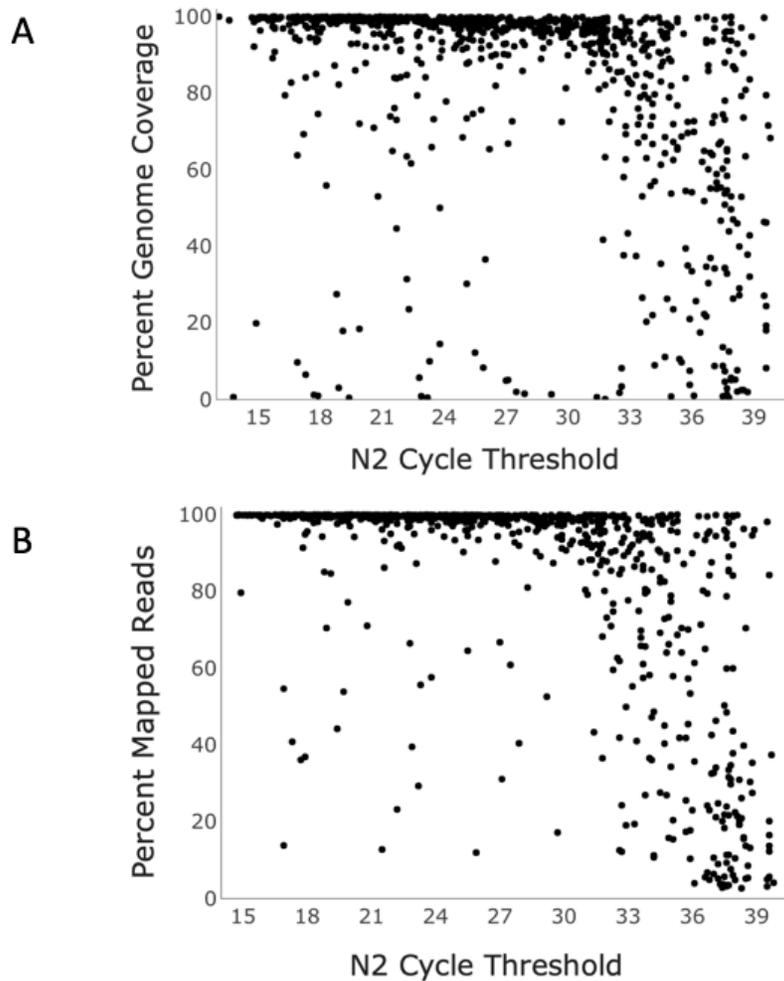


Figure 2. Sequencing outcomes of SARS-CoV2-positive samples processed using Seqwell’s plexWell method. A. Nucleocapsid-2 Ct value vs percent genome coverage and **B.** percent total reads mapped to SARS-CoV-2 for 897 SARS-CoV-2-positive samples sequenced using Seqwell’s plexWell 384 system.

preparation error, sample degradation, or poor sequencing run metrics. Beyond a Ct of 33, success rate dropped drastically. Between Ct values of 33 and 35, complete genome success rate was ~41% (Table 3), with an average breadth of coverage of 78%. Above a Ct of 35, ~18% of samples yielded a complete genome, and breadth of coverage dropped below 57% (Table 2).

Uniform depth of genome coverage across samples was targeted when pooling samples for sequencing, but depth generally decreased as Ct increased. Average depth of coverage was approximately 2766X through a Ct of 33. Past 33, coverage dropped off sharply, with an average depth of 1221X between Ct values of 33-35. Between Cts of 35 and 37, coverage was 716X, and above a Ct of 37, coverage dropped to 576X (Table 2).

Similarly, percent of total reads aligning to the SARS-CoV-2 reference genome decreased as Ct increased (Table 2, Figure 2). Up to a Ct of 30, most (~96%) of reads mapped to SARS-CoV-2. From Cts of 30 to 35, this noticeably decreased, and continued to drop as Cts rose. Beyond a Ct of 37, average percentage of reads aligning dropped to 38%.

Rapid-Turnaround Workflow

Turnaround time for the Illumina DNA prep method was significantly faster than the high-throughput plexWell system. It took less than 48 hours to go from raw sample to deliverable, analyzed data (Table 1), and this time could potentially be reduced further by reducing cycle numbers per sequencing run.

Both the plexWell and DNA Prep methods use a tagmentation system for adapter addition, rather than a ligation-based approach. This results in adapters being added ~50 bp or more from the end of overlapping amplicons generated during gene-specific PCR. A second PCR step amplifies only the tagged regions, resulting in final libraries of ~300bp. These approaches negate the need for primer trimming prior to alignment.

Generating consensus genomes and a SNP report from the sequence data, which takes approximately 15 minutes for a small (<=24 samples) dataset, quickly shows whether the samples are part of the same outbreak or transmission network. To quickly find a potential origin of a cluster or sample (assuming genomes in the public domain were collected prior to the

Table 2. Sequence metrics of SARS-CoV2-positive samples processed using Seqwell's plexWell method. Average sequencing metrics for various RT-PCR cycle threshold values of samples sequenced using the plexWell 384 system, and either an Illumina MiSeq or an Illumina NextSeq550.

*Ct value reported is that for the nucleocapsid-2 gene

**Uniform depth of coverage was targeted for all samples

Average Cycle Threshold Value (n)*	Average % aligned	Average % coverage	Average depth**	% samples > 90% coverage
<20 (194)	96.64	92.36	3371.58	87.63
20-25 (210)	96.91	92.61	2848.11	85.24
25-30 (169)	95.67	91.11	2733.37	84.62
30-33 (110)	88.58	87.18	2113.09	74.55
33-35 (71)	74.68	77.85	1221.05	40.85
35-37 (52)	60.20	56.91	716.43	17.31
37+ (91)	38.28	52.71	576.09	18.68

Table 3. Sequence metrics for SARS-CoV2-positive samples processed using Illumina's DNA Prep method. Average sequencing metrics for various RT-PCR cycle threshold values of samples sequenced using the Illumina DNA Prep system and an Illumina MiSeq.

*Ct value reported is that for the nucleocapsid-2 gene

**Uniform depth of coverage was targeted for all samples

Average Cycle Threshold Value (n)*	Average % aligned	Average % coverage	Average depth**	% samples > 90% coverage
<25 (6)	96.56	94.53	1503.00	83.33
25-30 (8)	91.92	87.60	1115.25	50%
30-37 (3)	56.46	62.83	851.21	0%

sample(s) of interest), a global SNP database can be used. Constructing or updating a global genome SNP database takes several hours, but it can be done prior to sample sequencing (and regularly). Running the script to query the global SNP database for particular SNPs of interest takes mere seconds.

To rapidly visualize results of the global SNP database query, a stacked Venn diagram (aka, “onion diagram”) visually describes the hierarchical nature, i.e., the parsimony, of SNPs found in the SARS-CoV-2 genomes (Figure 3), and is easily generated using the methods described above or an alternative tool.

Although fewer data are available from the Illumina DNA prep method, as it was primarily used to process 5 or fewer samples at a time, data suggest that it performed slightly worse than the Seqwell system, with odds of obtaining a complete genome dropping to 0 at Ct values greater than 30 (Table 3).

Discussion

The recent SARS-CoV-2 outbreak has highlighted the need for real-time sequencing and data analysis capacity in the face of active pandemics, as well as high-throughput sequencing and analysis strategies for comprehensive retrospective analysis and evaluation. We describe two different strategies that can be used in combination with Illumina platforms for the rapid or high-throughput interrogation of samples involved in disease

outbreaks. Though the results reported here are specific to the SARS-CoV-2 outbreak, these methods could conceivably be modified and applied to many other situations in which a genomic epidemiological response is needed.

The Illumina DNA prep-SNP-comparison analysis method is effective in providing rapid sequence data and genomic epidemiology information for small numbers of samples (<48 hours from raw sample to rough phylogenetic placement). Though throughput is limited by both the number of available indices for multiplexing and the nature of the protocol itself, this method has the advantage of being scalable to small numbers of samples, and can be performed in the course of several hours for small sample subsets.

The Seqwell plexWell system provides a scalable, cost-effective, and high-throughput method of processing thousands of samples with minimal laboratory personnel (Table 1). As the plexWell protocol calls for the pooling of samples at an initial step in the adapter tagmentation process, hundreds of samples were able to be taken through the later steps of the protocol by a single individual in an 8-hour timeframe. This, coupled with the availability of thousands of index combinations, allowed for a high-throughput, cost-effective means of processing large numbers of samples on a weekly basis with minimal laboratory personnel and infrastructure.

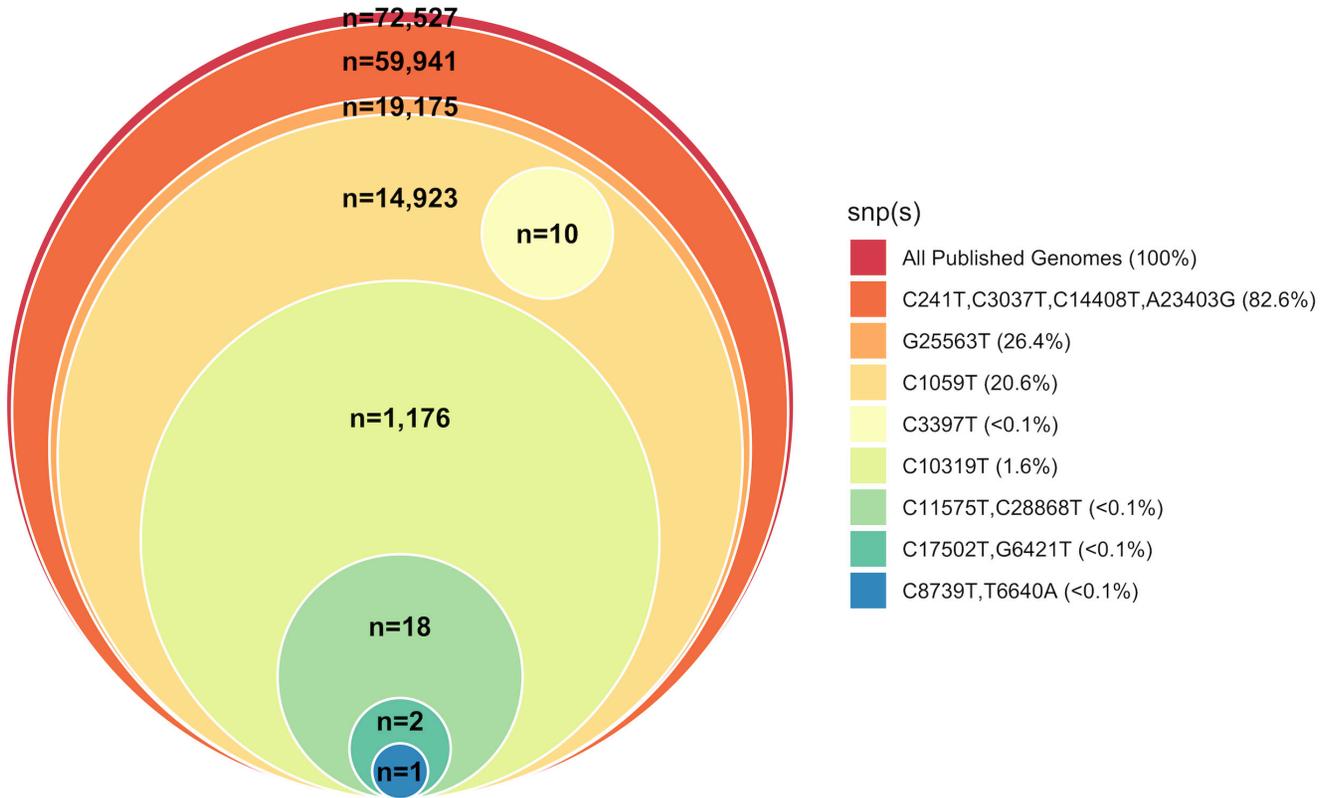


Figure 3. Phylogenetic relationship of SARS-CoV2-positive samples to Wuhan-Hu-1 reference strain. Stacked Venn “onion” diagram indicating the hierarchical nature of cluster-specific SNPs relative to the reference strain in global collection of SARS-CoV2 samples.

Typical analyses of virus genomes include phylogenetic tree construction to understand transmission patterns. Nextstrain¹⁷ and GISAID¹⁶ have been crucial to the SARS-CoV-2 scientific community for global and local epidemiologic understanding, and tools for smart subsampling (e.g. genome-sampler³⁵) are now necessary with the growth of the public databases. However, reconstructing phylogenies, especially paired with finding relevant subsets, takes time, and is often overkill for initial, time-sensitive public health needs. We employ a simple, rapid analysis method and visualization meant as a quick-look to determine relatedness among a sample set of interest and/or relatedness of a sample or set to the entire public database of genomes. Because of the novelty of SARS-CoV-2 and its low rate of recombination, merely comparing the low number of SNPs across samples without applying a phylogenetic model or program is often enough to answer initial questions about COVID-19 transmission, e.g. whether samples in a given set are closely related, and/or which samples in the global database are most closely related to a given sample set. Our SNP queries followed by generation of a stacked Venn diagram (onion diagram) offer a much faster alternative or antecedent to complete phylogenetic analysis. Other pathogens or situations where SNP numbers are expected to be very low may also benefit from these rapid analysis methods.

For retrospective studies, time can allow for more robust phylogenetic analyses including smart subsamplers³⁵, such as NextStrain¹⁷ and other commonly used phylogenetic tools; however their employment can significantly add time to a rapid response. The UShER tool³⁴, which can rapidly place genomes onto an existing SARS-CoV-2 phylogenetic tree, can greatly speed-up the final analysis. Parsing the output of UShER generates a subset of public genomes that are phylogenetically close to the samples of interest. This reduces the input dataset to subsamplers such as genome-sampler³⁵, which significantly reduces the computation time for further subsampling based on geography and time, which in turn significantly reduces the computation time for a NextStrain analysis.

Each of the two methods have their limitations. We observed a reduced success rate of the rapid Illumina DNA Prep method over the high-throughput Seqwell system, as evidenced by both decreased overall breadth of coverage and decreased success of obtaining complete genomes at Ct values above 30. It should be noted, however, that rapid prep study was conducted on a limited sample set, using samples from active outbreak clusters that were shipped from long distances through varying ambient temperatures. Samples used to evaluate the Seqwell system were obtained locally and processed entirely in-house. This difference in handling, coupled with the sample size difference, may in part account for the differences in results in the two prep methods. Also, at ~\$100/sample in reagent costs, the rapid Illumina DNA Prep method is less cost-effective (Table 1). And though the turnaround for the protocol is ~1.5 days, faster sequencing is achievable through other methods, such as through the use of long-read Nanopore sequencing^{18,36}. Nanopore technology, however, has the disadvantage of having a higher per-base error rate when compared to short-read sequencing methods¹⁸, thus its lack of accuracy may outweigh

any potential time savings, particularly in situations where relatively few nucleotide variants can radically alter phylogenetic placement such as for SARS-CoV-2. Also, the SNP-comparison analysis is rapid and robust because of the relatively low numbers of SNPs so far documented in the SARS-CoV-2 genome, due to the virus's novelty, the lack of recombination, and the unmatched robustness of the global SARS-CoV-2 genome database. Though this method could be applied to other types of outbreaks, rapid, precise phylogenetic placement will rely on these same factors.

The high-throughput SeqWell prep system also has its drawbacks. The pooling strategy of the plexWell system prevents downscaling, thus small numbers of samples cannot be processed effectively with this system. The turnaround time of this method, when large sample numbers are processed, is not competitive³⁶. The majority of processing time is lost to the ARTIC portion of the protocol, however, and not specifically to the plexWell adapter addition. And though the combinatorial index system allows for thousands of samples to be multiplexed in a single sequencing run, the SARS pandemic has demonstrated that even this may not be sufficient to meet the data challenges presented by expansive disease outbreaks.

Despite overwhelming benefits of employing next generation technological advances in real-time during a public health emergency, challenges remain when using genomic epidemiology as a means of pandemic control and monitoring. It has been demonstrated that genomics are not, in and of themselves, sufficient to completely elucidate the mechanisms and transmission of all pathogen-transmitted disease, particularly when asymptomatic and mild infections are known to play a role in transmission³⁸ but are less likely to be identified and subsequently sequenced. Thus, the integration of genomic and traditional epidemiology is paramount to the success of this 21st century public health capability.

Data availability

Underlying data

Zenodo: Raw sequencing metrics from two different prep methods for obtaining SARS-CoV2 genomes, <http://doi.org/10.5281/zenodo.4309900>³⁹.

This project contains the following underlying data:

- Sequencing_metrics_IlluminaDNAprep.xlsx- data used for generation of Table 3
- Sequencing_metrics_Seqwellprep.xlsx- data used for generation of Figure 2 and Table 2

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

Acknowledgements

The authors would like to thank Hayley Yaglom for her helpful review of the first draft and the members of the Arizona COVID Genomics Union (ACGU), for their insight and support of this work.

References

- Gardy JL, Naus M, Amlani A, *et al.*: **Whole-Genome Sequencing of Measles Virus Genotypes H1 and D8 During Outbreaks of Infection Following the 2010 Olympic Winter Games Reveals Viral Transmission Routes.** *J Infect Dis.* 2015; **212**(10): 1574–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Coll F, Harrison EM, Toleman MS, *et al.*: **Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community.** *Sci Transl Med.* 2017; **9**(413): eaak9745.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Arias A, Watson SJ, Asogun D, *et al.*: **Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases.** *Virus Evol.* 2016; **2**(1): vew016.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gardy J, Loman NJ, Rambaut A: **Real-time digital pathogen surveillance - the time is now.** *Genome Biol.* 2015; **16**(1): 155.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Worobey M, Pekar J, Larsen BB, *et al.*: **The emergence of SARS-CoV-2 in Europe and North America.** *Science.* 2020; **370**(6516): 564–570.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Engelthaler DM, Valentine M, Bowers J, *et al.*: **Hypervirulent emm59 Clone in Invasive Group A Streptococcus Outbreak, Southwestern United States.** *Emerg Infect Dis.* 2016; **22**(4): 734–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rockett RJ, Arnott A, Lam C, *et al.*: **Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling.** *Nat Med.* 2020; **26**(9): 1398–1404.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Oude Munnink BB, Nieuwenhuijse DF, Stein M, *et al.*: **Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands.** *Nat Med.* 2020; **26**(9): 1405–1410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Scarpino SV, Iamarino A, Wells C, *et al.*: **Epidemiological and viral genomic sequence analysis of the 2014 ebola outbreak reveals clustered transmission.** *Clin Infect Dis.* 2015; **60**(7): 1079–82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dudas G, Carvalho LM, Bedford T, *et al.*: **Virus genomes reveal factors that spread and sustained the Ebola epidemic.** *Nature.* 2017; **544**(7650): 309–315.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bowers JR, Kitchel B, Driebe EM, *et al.*: **Genomic Analysis of the Emergence and Rapid Global Dissemination of the Clonal Group 258 Klebsiella pneumoniae Pandemic.** *PLoS One.* 2015; **10**(7): e0133727.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Oltean HN, Etienne KA, Roe CC, *et al.*: **Utility of Whole-Genome Sequencing to Ascertain Locally Acquired Cases of Coccidioidomycosis, Washington, USA.** *Emerg Infect Dis.* 2019; **25**(3): 501–506.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zhou P, Yang XL, Wang XG, *et al.*: **A pneumonia outbreak associated with a new coronavirus of probable bat origin.** *Nature.* 2020; **579**(7798): 270–273.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Andersen KG, Rambaut A, Lipkin WI, *et al.*: **The proximal origin of SARS-CoV-2.** *Nat Med.* 2020; **26**(4): 450–452.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Prevention CfDca: **SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance (SPHERES).** 2020.
[Reference Source](#)
- Elbe S, Buckland-Merrett G: **Data, disease and diplomacy: GISAID's innovative contribution to global health.** *Glob Chall.* 2017; **1**(1): 33–46.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hadfield J, Megill C, Bell SM, *et al.*: **Nextstrain: real-time tracking of pathogen evolution.** *Bioinformatics.* 2018; **34**(23): 4121–4123.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quick J, Loman NJ, Duraffour S, *et al.*: **Real-time, portable genome sequencing for Ebola surveillance.** *Nature.* 2016; **530**(7589): 228–232.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ladner JT, Larsen BB, Bowers JR, *et al.*: **An Early Pandemic Analysis of SARS-CoV-2 Population Structure and Dynamics in Arizona.** *mBio.* 2020; **11**(5): e02107–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Forster P, Forster L, Renfrew C, *et al.*: **Phylogenetic network analysis of SARS-CoV-2 genomes.** *Proc Natl Acad Sci U S A.* 2020; **117**(17): 9241–9243.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Miller SL, Nazaroff WW, Jimenez JL, *et al.*: **Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event.** *Indoor Air.* 2020.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quick J: **nCoV-2019 sequencing protocol V.1.** ARTICNetwork. 2020.
[Publisher Full Text](#)
- Tyson JR, James P, Stoddart D, *et al.*: **Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore.** *bioRxiv.* 2020.
[Publisher Full Text](#)
- Batty EM, Kochakarn T, Wangwiwatsin A, *et al.*: **Comparing library preparation methods for SARS-CoV-2 multiplex amplicon sequencing on the Illumina MiSeq platform.** *bioRxiv.* 2020.
[Publisher Full Text](#)
- Colman RE, Anderson J, Lemmer D, *et al.*: **Rapid Drug Susceptibility Testing of Drug-Resistant Mycobacterium tuberculosis Isolates Directly from Clinical Samples by Use of Amplicon Sequencing: a Proof-of-Concept Study.** *J Clin Microbiol.* 2016; **54**(8): 2058–67.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bowers JR, Lemmer D, Sahl JW, *et al.*: **Diagnostic Tool for Surveillance, Detection, and Monitoring of Klebsiella pneumoniae.** *J Clin Microbiol.* 2016; **54**(10): 2582–96.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- BBMap. 2014.
[Reference Source](#)
- Wu F, Zhao S, Yu B, *et al.*: **A new coronavirus associated with human respiratory disease in China.** *Nature.* 2020; **579**(7798): 265–269.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li H, Durbin R: **Fast and accurate long-read alignment with Burrows-Wheeler transform.** *Bioinformatics.* 2010; **26**(5): 589–95.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chan AP, Choi Y, Schork NJ: **Conserved Genomic Terminals of SARS-CoV-2 as Coevolving Functional Elements and Potential Therapeutic Targets.** *mSphere.* 2020; **5**(6): e00754–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- R: **A language and environment for statistical computing.** Version 4.0.2. R Foundation; 2020.
[Reference Source](#)
- Wickham H: **ggplot2: Elegant Graphics for Data Analysis.** Springer-Verlag; 2016.
[Reference Source](#)
- ggforce: **Accelerating 'ggplot2'.** Version R package version 0.3.2. 2020.
[Reference Source](#)
- Turakhia Y, Thornlow B, Hinrichs AS, *et al.*: **Ultrafast Sample Placement on Existing Trees (USHER) Empowers Real-Time Phylogenetics for the SARS-CoV-2 Pandemic.** *bioRxiv.* 2020.
[Publisher Full Text](#)
- Bolyen E, Dillon M, Bokulich N, *et al.*: **Reproducibly sampling SARS-CoV-2 genomes across time, geography, and viral diversity [version 2; peer review: 1 approved, 1 approved with reservations].** *F1000Res.* 2020; **9**: 657.
[Publisher Full Text](#)
- Paden CR, Tao Y, Queen K, *et al.*: **Rapid, Sensitive, Full-Genome Sequencing of Severe Acute Respiratory Syndrome Coronavirus 2.** *Emerg Infect Dis.* 2020; **26**(10): 2401–2405.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Maurano MT, Ramaswami S, Zappile P, *et al.*: **Sequencing identifies multiple, early introductions of SARS-CoV-2 to New York City Region.** *medRxiv.* 2020.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Meredith LW, Hamilton WL, Warne B, *et al.*: **Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study.** *Lancet Infect Dis.* 2020; **20**(11): 1263–1272.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Folkerts ML: **Raw sequencing metrics from two different prep methods for obtaining SARS-CoV2 genomes [Data set].** *Zenodo.* 2020.
<http://www.doi.org/10.5281/zenodo.4309900>

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research