Check for updates

Performance of COVIDSeq and Swift Normalase Amplicon SARS-CoV-2 Panels for SARS-CoV-2 Genome Sequencing: Practical Guide and Combining FASTQ Strategy

Hosoon Choi,[a] Munok Hwang,[a] Dhammika H. Navarathna,[b] Jing Xu,[a] Janell Lukey,[b] Chetan Jinadatha[c,d]

[a]Department of Research, Central Texas Veterans Health Care System, Temple, Texas, USA
[b]Department of Pathology and Laboratory Medicine Services, Central Texas Veterans Health Care System, Temple, Texas, USA
[c]Department of Medicine, Central Texas Veterans Health Care System, Temple, Texas, USA
[d]College of Medicine, Texas A&M University, Bryan, Texas, USA

Along with real-time reverse transcriptase PCR (RT-PCR) diagnostic testing, whole-genomic sequencing (WGS) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been an irreplaceable tool for epidemiological evaluation, and genomic sequence analysis has provided essential information in the development of antiviral therapeutics and vaccines (1, 2). Good sequencing coverage needs to be obtained to get information: the default threshold for lineage call at Pangolin is at least 50% of non-N bases (3). Obtaining sufficient WGS coverage from low-viral-load samples is challenging and often not subject to WGS in many labs. However, WGS of low-viral-load patient samples will enable a more complete picture of viral transmission and viral evolution (4).

Here, we report the evaluation of two SARS-CoV-2 amplicon-based library prep kits, COVIDSeq with ARTIC v3 primers (Illumina) and Swift Normalase amplicon SARS-CoV-2 panels (SNAP) (Swift Biosciences) (5–7) with digital PCR and WGS using a total of 121 COVID-19-positive samples with threshold cycle (C_T) values between 11 and 45 (BD Max SARS-CoV-2 reverse transcriptase quantitative PCR [RT-qPCR] assay; please refer to supplementary material for library preparation and sequencing). About one-third (42) of samples were low viral load ($C_T > 30$). Out of 121 samples, >95% genome coverage was obtained for 107 samples by SNAP and 89 samples by COVIDSeq. This sequence yield exceeds many previous SARS-CoV-2 sequencing reports, especially for low-viral-load samples.

Through testing of various viral load samples along with digital PCR, we were able to obtain a tentative cutoff value of sample viral load above which one can expect to acquire an informative sequencing outcome. Sequence coverage of >95% was obtained by using SNAP for all of the samples with $C_T \leq 35$ and by COVIDSeq for 97% of samples with $C_T \leq 30$. Sample RNA quantitation obtained using digital PCR provided more precise cutoff values. The quantitative digital PCR (TaqPath COVID combo kit [Thermo Fisher] analysis with QIAcuity [Qiagen]) cutoff values for obtaining >95% coverage are 10.5 copies/μL for SNAP and 147 copies/μL for COVIDSeq (Fig. 1). The median sequence read depth was 2,913× and 3,578× with SNAP and COVIDSeq, respectively. All samples >750× coverage from SNAP achieved >95% sequence coverage. Read depth of >2,800× is sufficient to have >95% sequence coverage by COVIDSeq.

In addition, combining FASTQ files obtained from two kits improved the sequencing coverage and read depth significantly (please refer to supplementary material for combining FASTQ strategy). Combining FASTQ files was performed for 52 samples in which coverage is <95% or samples with lineage call discrepancy by either kit. By combining FASTQ files, >95% was obtained from all 21 samples with $C_T \leq 30$ and from 22 out of 31 samples with
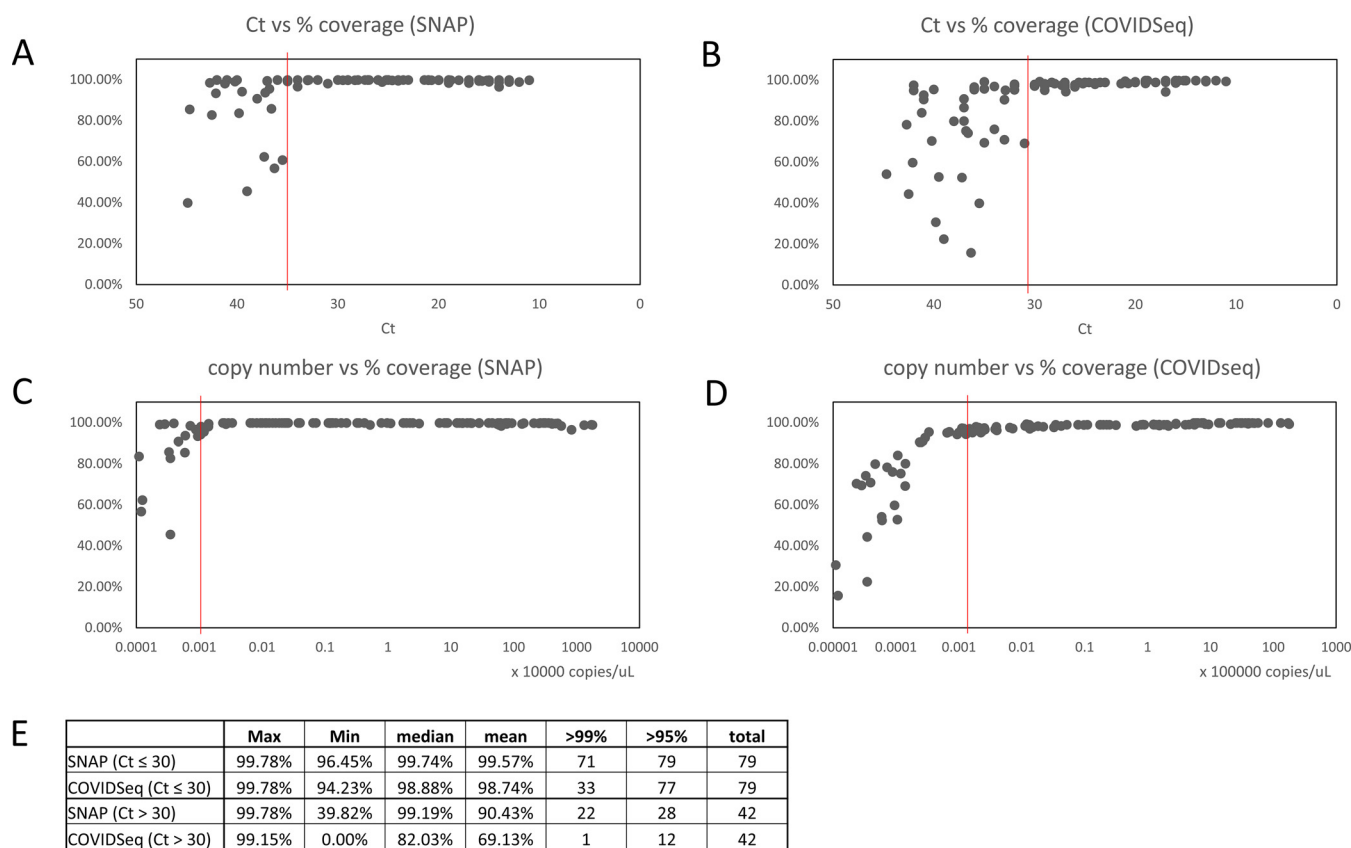
Address correspondence to Chetan Jinadatha, chetan.jinadatha@va.gov.

FIG 1 Correlation between viral loads and percent coverage. (A) Correlations between C_T values and percent coverage of SNAP protocol. 100% of samples sample with >95% coverage above the cutoff values. (B) Correlations between C_T values and percent coverage of COVIDSeq protocol. 97% of samples sample with >95% coverage above the cutoff values. (C) Correlations between copy number obtained by digital PCR and percent coverage of SNAP protocol. 100% of samples sample with >95% coverage above the cutoff values. (D) Correlations between copy number obtained by digital PCR and percent coverage of COVIDSeq protocol. 100% of samples sample with >95% coverage above the cutoff values. Red bars are proposed C_T cutoff values with which all of the samples located to the right side produced >95% genome coverage. (E) Summary of sequencing coverage of SNAP protocol and COVIDSeq protocol.

$C_T > 30$ (ranges C_T, ~31 to ~42.7) (Fig. 2). The strategy to combine FASTQ files could also be effective in the case of amplicon loss (8), which could occur due to novel mutations in emerging variants like Omicron.

As a result, all 79 samples with $C_T \leq 30$ achieved genome coverage >95%. For samples with low viral load, $C_T > 30$ (ranges C_T, ~31 to ~42.7), genome coverage >95% was obtained for 33 samples out of 42 samples with SNAP only or combining COVIDSeq and SNAP. Overall, SNAP produced better coverage and depth on moderate- and low-titer samples; all of the samples with $C_T \leq 35$ or >10.5 copies/μL achieved >95% sequence coverage. Combining FASTQ files from COVIDSeq and SNAP also increased sequence coverage and read depth for low-titer samples.
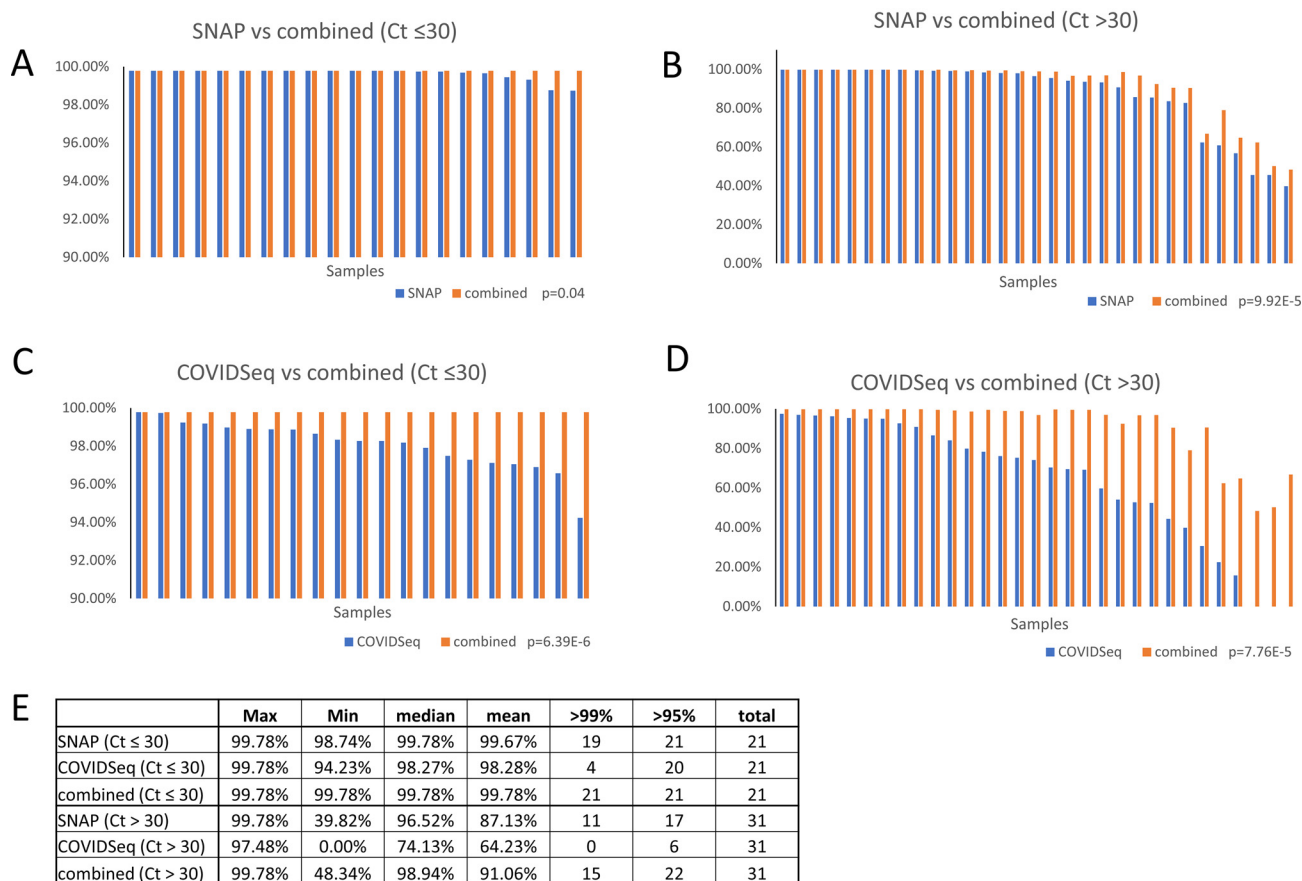
SUPPLEMENTAL MATERIAL

Supplemental material is available online only.
SUPPLEMENTAL FILE 1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

FIG 2 Increases of percent coverage by combined FASTQ. (A) Comparison of percent coverage between SNAP protocol and combined FASTQ of low-C_T samples. (B) Comparison of percent coverage between SNAP protocol and combined FASTQ of high-C_T samples. (C) Comparison of percent coverage between COVIDSeq protocol and combined FASTQ of low-C_T samples. (D) Comparison of percent coverage between COVIDSeq protocol and combined FASTQ of high C_T samples. (E) Summary of sequencing coverage of SNAP protocol, COVIDSeq protocol, and combined FASTQ.

The opinions expressed here are those of the authors and do not represent the views of Department of Veterans Affairs or of Central Texas Veterans Health Care System, Temple, TX.

REFERENCES

1. Kames J, Holcomb DD, Kimchi O, DiCuccio M, Hamasaki-Katagiri N, Wang T, Komar AA, Alexaki A, Kimchi-Sarfaty C. 2020. Sequence analysis of SARS-CoV-2 genome reveals features important for vaccine design. Sci Rep 10: 15643. https://doi.org/10.1038/s41598-020-72533-2.

2. Dai W, Zhang B, Jiang XM, Su H, Li J, Zhao Y, Xie X, Jin Z, Peng J, Liu F, Li C, Li Y, Bai F, Wang H, Cheng X, Cen X, Hu S, Yang X, Wang J, Liu X, Xiao G, Jiang H, Rao Z, Zhang LK, Xu Y, Yang H, Liu H. 2020. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science 368:1331–1335. https://doi.org/10.1126/science.abb4489.

3. Rambaut A, Holmes EC, O'Toole A, Hill V, McCrone JT, Ruis C, Du Plessis L, Pybus OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5:1403–1407. https://doi.org/10.1038/s41564-020-0770-5.

4. Teran RA, Walblay KA, Shane EL, Xydis S, Gretsch S, Gagner A, Samala U, Choi H, Zelinski C, Black SR. 2021. Postvaccination SARS-CoV-2 infections among skilled nursing facility residents and staff members - Chicago, Illinois, December 2020-March 2021. Am J Transplant 21:2290–2297. https://doi.org/10.1111/ajt.16634.

5. Addetia A, Lin MJ, Peddu V, Roychoudhury P, Jerome KR, Greninger AL. 2020. Sensitive recovery of complete SARS-CoV-2 genomes from clinical samples by use of swift biosciences' SARS-CoV-2 multiplex amplicon sequencing panel. J Clin Microbiol 59:e02226-20. https://doi.org/10.1128/JCM.02226-20.

6. Bhoyar RC, Jain A, Sehgal P, Divakar MK, Sharma D, Imran M, Jolly B, Ranjan G, Rophina M, Sharma S, Siwach S, Pandhare K, Sahoo S, Sahoo M, Nayak A, Mohanty JN, Das J, Bhandari S, Mathur SK, Kumar A, Sahlot R, Rojarani P, Lakshmi JV, Surekha A, Sekhar PC, Mahajan S, Masih S, Singh P, Kumar V, Jose B, Mahajan V, Gupta V, Gupta R, Arumugam P, Singh A, Nandy APVR, Jha RM, Kumari A, Gandotra S, Rao V, Faruq M, Kumar S, Reshma GB, Varma GN, Roy SS, Sengupta A, Chattopadhyay S, Singhal K, Pradhan S, et al. 2021. High throughput detection and genetic epidemiology of SARS-CoV-2 using COVIDSeq next-generation sequencing. PLoS One 16:e0247115. https://doi.org/10.1371/journal.pone.0247115.

7. Bhoyar RC, Senthivel V, Jolly B, Imran M, Jain A, Divakar MK, Scaria V, Sivasubbu S. 2021. An optimized, amplicon-based approach for sequencing of SARS-CoV-2 from patient samples using COVIDSeq assay on Illumina MiSeq sequencing platforms. STAR Protoc 2:100755. https://doi.org/10.1016/j.xpro.2021.100755.

8. Sanderson T, Barrett JC. 2021. Variation at Spike position 142 in SARS-CoV-2 Delta genomes is a technical artifact caused by dropout of a sequencing amplicon. medRxiv. https://doi.org/10.1101/2021.10.14.21264847.