Title

Application of CRISPR-Based Human and Bacterial Ribosomal RNA Depletion for SARS-CoV-2 Shotgun Metagenomic Sequencing

Journal

American Journal of Clinical Pathology, 159(2)

ISSN

0002-9173

Authors

Cerón, Stacey
Clemons, Nathan C
von Bredow, Benjamin
et al.

Publication Date

2023-02-01

DOI

10.1093/ajcp/aqac135

Peer reviewed

Application of CRISPR-Based Human and Bacterial Ribosomal RNA Depletion for SARS-CoV-2 Shotgun Metagenomic Sequencing

Stacey Cerón, PhD, Nathan C. Clemons, PhD, Benjamin von Bredow, PhD, and Shangxin Yang, PhD, D(ABMM)

From the Department of Pathology and Laboratory Medicine, UCLA David Geffen School of Medicine, Los Angeles, CA, USA.

ABSTRACT

Objectives: The aim of this study is to evaluate the effectiveness of a CRISPR-based human and bacterial ribosomal RNA (rRNA) depletion kit (JUMPCODE Genomics) on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) shotgun metagenomic sequencing in weakly positive respiratory samples.

Methods: Shotgun metagenomics was performed on 40 respiratory specimens collected from solid organ transplant patients and deceased intensive care unit patients at UCLA Medical Center in late 2020 to early 2021. Human and bacterial rRNA depletion was performed on remnant library pools prior to sequencing by Illumina MiSeq. Data quality was analyzed using Geneious Prime, whereas the identification of SARS-CoV-2 variants and lineages was determined by Pangolin.

Results: The average genome coverage of the rRNA-depleted respiratory specimens increased from 72.55% to 93.71% in overall samples and from 29.3% to 83.3% in 15 samples that failed to achieve sufficient genome coverage using the standard method. Moreover, rRNA depletion enhanced genome coverage to over 85% in 11 (73.3%) of 15 low viral load samples with cycle threshold values up to 35, resulting in the identification of genotypes.

Conclusion: The CRISPR-based human and bacterial rRNA depletion enhanced the sensitivity of SARS-CoV-2 shotgun metagenomic sequencing, especially in low viral load samples.

INTRODUCTION

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to more than 518 million infections and 6.25 million deaths worldwide.[1] Surges of cases have been correlated with the dominance of SARS-CoV-2 variants that display a higher infectivity rate and/or an ability to evade the immune response.[2,3] Genomic surveillance has been instrumental in the identification of novel variants as well as monitoring transmission dynamics and viral evolution.[4] Various methods have been developed for SARS-CoV-2 sequencing: amplicon, hybrid capture, and shotgun metagenomics.[5]

The basis of amplicon-based sequencing relies on primers that target specific regions of the viral genome in a multiplex tiling polymerase chain reaction (PCR) assay prior to sequencing.[5] The most common targeted amplicon approach for SARS-CoV-2 sequencing was designed by the ARTIC Network.[6] However, constant modifications to the primers and the protocol must be done to account for amplicon dropout, viral mutations, and overall

KEY POINTS

- Shotgun metagenomics is an unbiased sequencing method that requires a higher abundance of viral RNA and sequencing depth for genomic analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), making this approach challenging for specimens with lower viral loads.
- A CRISPR-based human and bacterial ribosomal RNA (rRNA) depletion on pooled libraries significantly increased genome coverage that had otherwise not achieved a sufficient depth of more than 85% with the standard method.
- With rRNA depletion, sufficient genome coverage can be achieved in most of the weakly positive samples with Ct values up to 35, which leads to successful identification of SARS-CoV-2 genotypes.

KEY WORDS

SARS-CoV-2; COVID-19; Shotgun metagenomics; Sequencing; Genomic surveillance

Am J Clin Pathol XXXX 2022;XX:1-0
HTTPS://DOI.ORG/10.1093/AJCP/AQAC135

Received: July 30, 2022
Accepted: October 3, 2022

Corresponding author: Shangxin Yang, PhD, D(ABMM); shangxinyang@mednet.ucla.edu.

TABLE 1 Next-Generation Sequencing Results of Clinical Coronavirus Disease 2019 Specimens[a]

			Standard Method			JUMPCODE					
Sample	Collection Year	Ct Value	No. of Mapped Sequences	Genome Coverage, %	Total SARS-CoV-2 Reads Mapped to Reference, %	No. of Mapped Sequences	Genome Coverage, %	Total SARS-CoV-2 Reads Mapped to Reference, %	Accomplished Coverage: Standard vs JUMPCODE	Lineage	Variant
1	2020	35.01	7	1.4	0.000	619	35.7	0.003	None	ND	ND
2	2020	11.45	10,134,775	100	74.205	5,982,310	100	68.040	Standard	B.1.2	Other
3	2020	25.31	13,919	100	1.869	85,591	100	70.056	Standard	B.1.2	Other
4	2020	32.86	310	19	0.049	2,945	100	1.991	JUMPCODE	B.1.596	Other
5	2020	21.54	1,941	92.8	2.224	57,253	100	5.286	Standard	B.1.2	Other
6	2020	12.25	6,969,074	100	56.594	4,812,369	100	91.136	Standard	B.1.243	Other
7	2020	16.40	4,334,894	100	49.919	3,197,762	100	77.648	Standard	B.1.2	Other
8	2020	24.50	692	46.5	5.843	6,976	67.3	0.306	None	ND	Other
9	2020	25.24	75	15.8	0.272	2,348	34	0.956	None	ND	ND
10	2020	27.72	2,511	74.3	0.019	15,553	100	0.559	JUMPCODE	AY.26	Other
11	2020	20.25	127,496	99.98	3.818	101,776	100	5.953	Standard	B.1.429	Epsilon
12	2020	22.76	2,263	97.7	1.865	44,545	99.6	3.797	Standard	B.1429	Epsilon
13	2021	21.15	31,422	99.9	4.084	27,504	100	7.693	Standard	B.1.429	Epsilon
14	2021	23.70	140,160	100	5.872	121,618	100	12.341	Standard	B.1.2	Other
15	2021	21.25	221,019	99.99	17.833	217,358	100	35.057	Standard	B.1.1.519	Other
16	2021	28.75	276	29.3	0.098	10,179.81	98	0.243	JUMPCODE	B.1.429	Epsilon
17	2021	28.00	10	2.7	0.059	13,926.06	99.4	0.100	JUMPCODE	B.1	Other
18	2021	15.90	50,944	100	3.589	30,020,996	100	91.948	Standard	B.1.595	Other
19	2021	29.15	849	39.8	0.073	5,843	100	1.114	JUMPCODE	B.1.2	Other
20	2021	23.60	9437.22	96.9	1.0289	16,017	100	5.3483	Standard	B.1.427	Epsilon
21	2021	19.00	265,952	100	18.735	8,530,834	100	26.128	Standard	B.1.429	Epsilon
22	2021	34.62	363	20.3	0.042	458	89.4	0.155	JUMPCODE	B.1	Other
23	2021	17.09	82,258	100	5.890	805,873	100	0.884	Standard	B.1.429	Epsilon
24	2021	19.17	73,623	100	0.704	101,031	100	11.179	Standard	B.1.2	Other
25	2021	19.15	49,669	100	61.856	572,001	99.9	43.943	Standard	B.1.427	Epsilon
26	2021	29.35	703	36.8	0.012	10,625	100	0.337	JUMPCODE	B.1.2	Other
27	2021	24.90	214	39.4	1.154	7,409	96.6	6.212	JUMPCODE	B.1.429	Epsilon
28	2021	15.68	1,212,228	100	59.121	43,433,986	100	90.008	Standard	B.1.429	Epsilon
29	2021	17.72	94,655	100	6.592	2,180,221	100	90.852	Standard	B.1.429	Epsilon
30	2021	35.06	723	40.5	0.025	3,939	100	0.648	JUMPCODE	B.1.2	Other
31	2021	18.05	74,959	100	6.964	1,897,700	100	35.432	Standard	B.1.429	Epsilon
32	2021	22.06	1,230	90.6	2.120	24,509	99.3	44.574	Standard	B.1.429	Epsilon

HTTPS://DOI.ORG/10.1093/AJCP/AQAC135

Table 1 (cont)

Sample	Collection Year	Ct Value	Standard Method			JUMPCODE					
			No. of Mapped Sequences	Genome Coverage, %	Total SARS-CoV-2 Reads Mapped to Reference, %	No. of Mapped Sequences	Genome Coverage, %	Total SARS-CoV-2 Reads Mapped to Reference, %	Accomplished Coverage: Standard vs JUMPCODE	Lineage	Variant
33	2021	28.45	287	12.3	0.789	1,545	37.3	0.065	None	ND	ND
34	2021	16.20	390	24.4	0.018	735	97.3	0.001	JUMPCODE	B.1.427	Epsilon
35	2021	24.70	4,362	85.8	0.447	8,992	100	7.291	JUMPCODE	B.1.427	Epsilon
36	2021	31.35	646	36.3	0.078	508	94.5	0.404	JUMPCODE	B.1.429	Epsilon
37	2021	19.15	31,875	99.9	14.623	984,768	100	31.303	Standard	B.1.429	Epsilon
38	2021	22.46	125,068	99.96	16.366	135,319	100	4.004	Standard	P.1.10	Gamma
39	2021	20.65	92,144	100	16.566	94,644	100	6.682	Standard	AY.25.1	Delta
40	2021	24.57	21,985	99.99	0.963	28,903	100	2.415	Standard	AY.103	Delta

Ct, cycle threshold; ND, not determined.
[a]Read length is ~250 base pairs.

improvements in workflow and sequencing quality.[6-8] In contrast, the hybrid capture–based method is able to detect SARS-CoV-2 by the hybridization of specific probes to target regions in the viral genome.[9,10] This technique is able to tolerate genomic variability slightly better than the amplicon-based method given that the capture of the target region is less specific.[10] However, certain modifications to this approach might be required to achieve higher sensitivity for specimens with a low viral burden.[11] Both the amplicon-based and the hybrid capture–based methods can be affected by the continuous genetic changes of the virus and require periodic modification and updates.

Shotgun metagenomics is an unbiased approach that is able to sequence all DNA/RNA present in a sample.[5] Several clinical applications of this method have included the identification of novel pathogens, monitoring disease outbreaks, and guidance for therapy.[5,12] Despite demonstrating clinical utility, shotgun metagenomics requires a higher abundance of microbial DNA/RNA or transcripts and higher sequencing depth to obtain quality data needed for genomic epidemiologic analysis.[12-14] Certain techniques have been successfully applied for the targeted enrichment of low-abundance species, including background depletion.[15]

In this study, we monitored SARS-CoV-2 variants and lineages in respiratory specimens acquired from solid organ transplant patients and deceased intensive care unit (ICU) patients in late 2020 to early 2021. The analysis of the initial run demonstrated that most of the positive specimens with a PCR cycle threshold (Ct) value less than 24 achieved sufficient genome coverage for analysis. To increase genome coverage in specimens with low viral burden, we evaluate the effectiveness of a commercial CRISPR-based human and bacterial ribosomal RNA depletion on SARS-CoV-2 shotgun metagenomic sequencing.

MATERIALS AND METHODS

From late 2020 to early 2021, remnant SARS-CoV-2 PCR-positive respiratory specimens, including nasopharyngeal swabs and bronchoalveolar lavage collected from solid organ transplant patients and deceased ICU patients at the UCLA Medical Center, were included in this study. The coronavirus disease 2019 (COVID-19) PCR methods have been previously described,[16] and samples were tested by one of the following two assays: Simplexa COVID-19 Direct Real-Time RT-PCR Assay (DiaSorin Molecular) and TaqPath COVID-19 RT-PCR Assay (Thermo Fisher Scientific). The positive respiratory samples had a range of Ct values between 11 and 36. Standard shotgun libraries were prepared using the NEBNext Ultra II RNA protocol (New England Biolabs) and sequenced on MiSeq (Illumina).

In subsequent runs, bacterial and human Cas9/single guide RNA ribonucleoprotein complex provided with the CRISPRclean Metatranscritpomic rRNA Depletion Kit (JUMPCODE Genomics) was used on pooled leftover libraries prior to sequencing, following the manufacturer's protocol. Following the CRISPR digestion of bacterial and human sequences in the complementary DNA libraries, 0.6× AMPure XP beads (Beckman Coulter Life Sciences) were

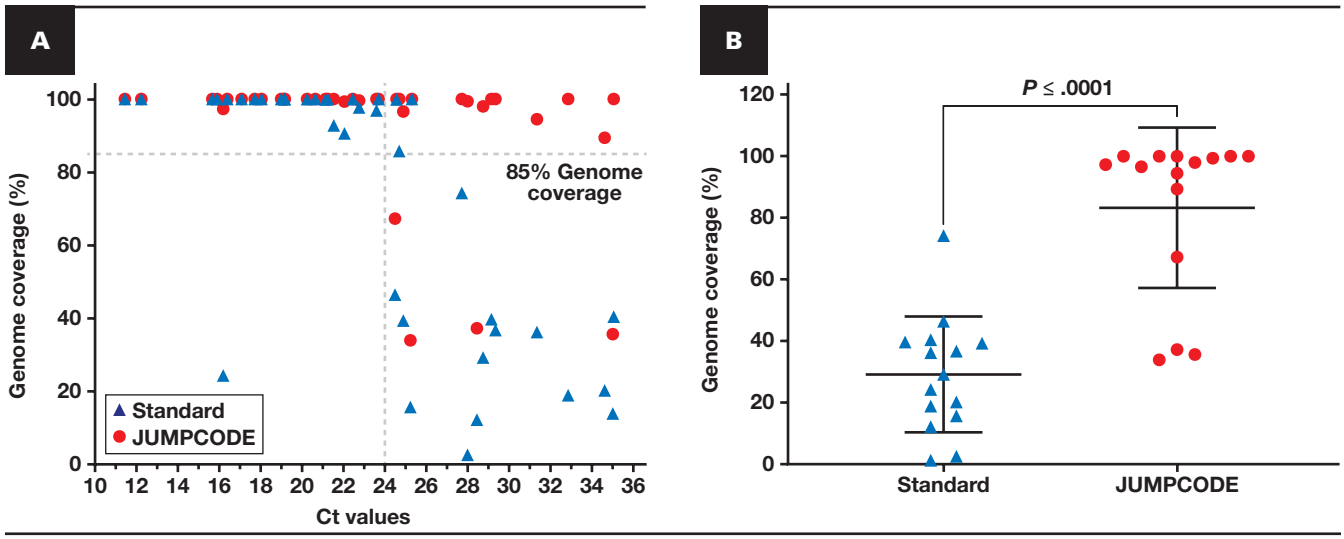

FIGURE 1 **A**, Overview of the genomic coverage percentage and cycle threshold values of clinical coronavirus disease 2019 (COVID-19) respiratory specimens. Blue triangles represent clinical specimens that underwent standard sequencing methods, whereas the red circles represent those that were treated with CRISPR-based ribodepletion prior to sequencing. **B**, Comparative assessment of the genomic coverage percentage of 15 clinical COVID-19 specimens before and after CRISPR-based ribodepletion. This subset of clinical specimens did not achieve sufficient genomic coverage (≥85%) using the standard method. Error bars represent standard deviation. Statistical significance was determined by a paired Student *t* test.

used to remove cleaved sequences prior to library enrichment by PCR amplification (13 cycles). Subsequently, AMPure XP beads were once again used to select DNA fragments more than 500 base pairs. The fragment sizes and relative concentration of the DNA fragments were confirmed with the 1000 Bioanalyzer (Agilent) prior to sequencing.

Metrics for quality control of the sequences included having a Q30 score of 85% or higher and a cluster passing filter of 85% or more. Genetic analysis was conducted with Geneious Prime on sequences with a genome coverage of 85% or more with at least 5× depth. Variants and lineages were identified with the open-source platform, Pangolin COVID-19 Lineage Assigner.[17]

This study was reviewed by the UCLA Human Research Protection Program and received an institutional review board exemption.

RESULTS

To determine SARS-CoV-2 lineages and variants, 40 respiratory specimens with a range of Ct values (11.45-35.06) were sequenced and analyzed. As expected, 22 of 23 clinical specimens with a Ct value of 24 or less but only 3 of 17 samples with a Ct value more than 24 achieved sufficient genome coverage (≥85%) using the standard shotgun metagenomic methods **TABLE 1** and **FIGURE 1A**. Genetic analysis demonstrated that the genome coverage increased in most human and bacterial ribosomal RNA (rRNA)–depleted specimens, with the overall genome coverage average of the entire cohort improving from 72.55% to 93.71% **FIGURE 1A**, **TABLE 1**.

Sufficient genome coverage was achieved in 11 (73.3%) of 15 rRNA-depleted (Ct ranged from 16.20 to 35.06) specimens that had otherwise not obtained a genome coverage of 85% or more using the standard method **TABLE 1**. Notably, the average genome coverage in these 15 samples improved significantly from 29.3% to 83.3% **FIGURE 1B**. Moreover, the average Ct of samples achieving sufficient genome coverage also increased from 19.80 to 22.58. It appears that both increased percentage of total reads mapped to the reference genome and improved genome coverage evenness (based on the visualization of the read mapping on individual samples; data not shown) contributed to the overall genome coverage enhancement **TABLE 1**. The increased genome coverage was essential for the identification of SARS-CoV-2 variants and lineages. As seen in **TABLE 1**, the Epsilon variant (B.1.427/429) was found in 47% of our samples, which also coincided with the predominant variant at that time in Los Angeles County, California (approximately 50% Epsilon).[18]

DISCUSSION

Global surveillance by sequencing has been instrumental in the discovery of emerging pathogens as well as monitoring transmission, pathogenesis, and disease.[19] Although various sequencing methods have been implemented for genetic surveillance of SARS-CoV-2, amplicon sequencing is one of the most common approaches.[20] However, a major disadvantage is the constant modifications required on the primer designs and workflow to accommodate continuous viral mutations.[7] For genomic surveillance, shotgun metagenomics is an efficient unbiased approach that is able to sequence all viral genetic elements in a specimen. Given that low-abundance transcripts might not be able to be sequenced effectively, modifications to deplete background genetic elements such as human and bacterial rRNA should be considered to enhance low viral reads.[5]

Our results demonstrate the effectiveness of the CRISPRclean Metatranscriptomic rRNA Depletion Kit in enhancing the genome coverage of SARS-CoV-2 sequencing in most samples. The depletion

of bacterial and human rRNA on clinical specimens with low viral loads enhanced the overall genome coverage sensitivity and thereby allowed for the identification of SARS-CoV-2 variants and lineages. In our critical patient cohort, Epsilon was identified as the predominant circulating variant with similar prevalence in the community during November 2020 to March 2021.[18]

In conclusion, we demonstrated that performing shotgun metagenomics with rRNA depletion can greatly improve sensitivity for genomic surveillance while maintaining its valuable capability of detecting novel variants without the need to constantly modify wet laboratory protocols.

Acknowledgments: This study was funded by UCLA Department of Pathology and Laboratory Medicine. The rRNA depletion reagents were provided by JUMPCODE Genomics.

REFERENCES

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* 2020;20:533-534.
2. Burioni R, Topol EJ. Assessing the human immune response to SARS-CoV-2 variants. *Nat Med.* 2021;27:571-572.
3. Hu J, Peng P, Cao X, et al. Increased immune escape of the new SARS-CoV-2 variant of concern Omicron. *Cell Mol Immunol.* 2022;19:293-295.
4. Chen Z, Azman AS, Chen X, et al. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. *Nat Genet.* 2022;54:499-507.
5. Chiara M, D'Erchia AM, Gissi C, et al. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. *Brief Bioinform.* 2021;22:616-630.
6. 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 Sep 4;2020.09.04.283077.
7. Davis JJ, Long SW, Christensen PA, et al. Analysis of the ARTIC Version 3 and Version 4 SARS-CoV-2 primers and their impact on the detection of the G142D amino acid substitution in the spike protein. *Microbiol Spectr.* 2021;9:e0180321.
8. Itokawa K, Sekizuka T, Hashino M, et al. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLoS One.* 2020;15:e0239403.
9. Gaudin M, Desnues C. Hybrid capture-based next generation sequencing and its application to human infectious diseases. *Front Microbiol.* 2018;9:2924.
10. Nagy-Szakal D, Couto-Rodriguez M, Wells HL, et al. Targeted hybridization capture of SARS-CoV-2 and metagenomics enables genetic variant discovery and nasal microbiome insights. *Microbiol Spectr* 2021;9:e0019721.
11. Xiao M, Liu X, Ji J, et al. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Med.* 2020;12:57.
12. Quince C, Walker AW, Simpson JT, et al. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol.* 2017;35:833-844.
13. Lam C, Gray K, Gall M, et al. SARS-CoV-2 genome sequencing methods differ in their abilities to detect variants from low-viral-load samples. *J Clin Microbiol.* 2021;59:e0104621.
14. Chen J, Hilt EE, Li F, et al. Epidemiological and genomic analysis of SARS-CoV-2 in 10 patients from a mid-sized city outside of Hubei, China in the early phase of the COVID-19 outbreak. *Front Public Health.* 2020;8:567621.
15. Lim YW, Haynes M, Furlan M, et al. Purifying the impure: sequencing metagenomes and metatranscriptomes from complex animal-associated samples. *J Vis Exp.* 2014;94:52117.
16. Price TK, Bowland BC, Chandrasekaran S, et al. Performance characteristics of severe acute respiratory syndrome coronavirus 2 RT-PCR tests in a single health system: analysis of >10,000 results from three different assays. *J Mol Diagn.* 2021;23:159-163.
17. O'Toole A, Scher E, Underwood A, et al. Assignment of epidemiological lineages in an emerging pandemic using the Pangolin tool. *Virus Evol.* 2021;7:veab064.
18. Yang S, Hemarajata P, Hilt EE, et al. Investigation of SARS-CoV-2 Epsilon variant and hospitalization status by genomic surveillance in a single large health system during the 2020-2021 winter surge in Southern California. *Am J Clin Pathol.* 2022;157:649-652.
19. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet.* 2018;19:9-20.
20. Nasir JA, Kozak RA, Aftanas P, et al. A comparison of whole genome sequencing of SARS-CoV-2 using amplicon-based sequencing, random hexamers, and bait capture. *Viruses.* 2020;12:895.

HTTPS://DOI.ORG/10.1093/AJCP/AQAC135