

Communication

A Novel High-Throughput Nanopore-Sequencing-Based Strategy for Rapid and Automated S-Protein Typing of SARS-CoV-2 Variants

Gabriel E. Wagner ^{1,*}, Massimo G. Totaro ², André Volland ³, Michaela Lipp ¹, Sabine Saiger ¹, Sabine Lichtenegger ¹, Patrick Forstner ¹, Dorothee von Laer ³, Gustav Oberdorfer ² and Ivo Steinmetz ^{1,*}

- ¹ Diagnostic and Research Institute of Hygiene, Microbiology and Environmental Medicine, Medical University of Graz, 8010 Graz, Austria; michaela.lipp@medunigraz.at (M.L.); sabine.saiger@medunigraz.at (S.S.); sabine.lichtenegger@medunigraz.at (S.L.); patrick.forstner@medunigraz.at (P.F.)
- ² Department of Biochemistry, Graz University of Technology, 8010 Graz, Austria; massimo.totaro@tugraz.at (M.G.T.); gustav.oberdorfer@tugraz.at (G.O.)
- ³ Department of Hygiene, Microbiology and Public Health, Institute of Virology, Medical University of Innsbruck, 6020 Innsbruck, Austria; andre.volland@i-med.ac.at (A.V.); dorothee.von-laer@i-med.ac.at (D.v.L.)
- * Correspondence: gabriel.wagner-lichtenegger@medunigraz.at (G.E.W.); ivo.steinmetz@medunigraz.at (I.S.); Tel.: +43-316-385-73700 (I.S.)



Citation: Wagner, G.E.; Totaro, M.G.; Volland, A.; Lipp, M.; Saiger, S.; Lichtenegger, S.; Forstner, P.; von Laer, D.; Oberdorfer, G.; Steinmetz, I. A Novel High-Throughput Nanopore-Sequencing-Based Strategy for Rapid and Automated S-Protein Typing of SARS-CoV-2 Variants. *Viruses* **2021**, *13*, 2548. <https://doi.org/10.3390/v13122548>

Academic Editors:
Luis Martinez-Sobrido and
Fernando Almazan Toral

Received: 29 November 2021
Accepted: 16 December 2021
Published: 19 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Rapid molecular surveillance of SARS-CoV-2 S-protein variants leading to immune escape and/or increased infectivity is of utmost importance. Among global bottlenecks for variant monitoring in diagnostic settings are sequencing and bioinformatics capacities. In this study, we aimed to establish a rapid and user-friendly protocol for high-throughput S-gene sequencing and subsequent automated identification of variants. We designed two new primer pairs to amplify only the immunodominant part of the S-gene for nanopore sequencing. Furthermore, we developed an automated “S-Protein-Typer” tool that analyzes and reports S-protein mutations on the amino acid level including a variant of concern indicator. Validation of our primer panel using SARS-CoV-2-positive respiratory specimens covering a broad C_t range showed successful amplification for 29/30 samples. Restriction to the region of interest freed sequencing capacity by a factor of 12–13, compared with whole-genome sequencing. Using either the MinION or Flongle flow cell, our sequencing strategy reduced the time required to identify SARS-CoV-2 variants accordingly. The S-Protein-Typer tool identified all mutations correctly when challenged with our sequenced samples and 50 deposited sequences covering all VOCs (December 2021). Our proposed S-protein variant screening offers a simple, more rapid, and low-cost entry into NGS-based SARS-CoV-2 analysis, compared with current whole-genome approaches.

Keywords: SARS-CoV-2; next-generation sequencing; vaccine escape; S-protein; nanopore; surveillance; typing

1. Introduction

Emerging SARS-CoV-2 variants and their rapid spread represent major challenges for global vaccination strategies and health care systems [1,2]. Repeated waves of COVID cases highlight the relevance of tracking SARS-CoV-2 S-protein mutations, because of their importance for vaccine efficacy and therapeutic antibody therapy [3,4]. Thus far, identification of SARS-CoV-2 S-gene variants has relied on next-generation sequencing of the whole genome [5–7] or Sanger sequencing of PCR products [8]. RT-PCR has been used for rapid variant identification, screening for certain known mutations only [9,10].

For whole-genome sequencing, nanopore sequencing and the ARTIC pipeline have proven to be powerful tools to monitor genome alterations of circulating SARS-CoV-2 variants [5–7]. Here, we propose a nanopore-based strategy to enable fast and easy

molecular surveillance at low cost in the diagnostic laboratory, by reducing the sequenced region primarily to the S-protein's immunodominant S1 subunit (region of interest (ROI)). Furthermore, we provide a user-friendly analysis pipeline for broad application in a diagnostic setting.

2. Materials and Methods

DNA quantification mentioned in this publication was carried out using the Qubit dsDNA HS Assay Kit on a Qubit 4 fluorometer (Thermo Scientific, Waltham, MA, USA).

SARS-CoV-2 RNAs were isolated from nasopharyngeal or throat swabs using the MagMAX Viral/Pathogen Nucleic Acid Isolation Kit (Thermo Scientific) on a KingFisher Flex (Thermo Scientific Thermo Scientific, Waltham, MA, USA) according to manufacturer instructions.

The reported C_T values were obtained using the Cobas SARS-CoV-2 Kit on a Cobas 6800 system (Roche Diagnostics, Vienna, Austria), according to manufacturer instructions.

Sample amplification and library preparation for nanopore sequencing were carried out as described in the "PCR tiling of SARS-CoV-2 virus with rapid barcoding" protocol, based on a rapid barcoding kit (Oxford Nanopore, Oxford, UK). However, before barcoding, samples were adjusted to the same DNA concentration and, if stated, our S1-ROI primer panel (Table 1) was used for amplicon generation.

Table 1. S1-ROI primer set: Primer names and sequences to amplify the region of interest (immunodominant domain of the S-protein). Furthermore, the amplified region of the reference strain (Wuhan-Hu-1) and the factor of reduction, compared with whole-genome amplification, and hence theoretical reduction in sequencing time are denoted.

S1-ROI Primer Panel	
S1-ROI-1_Forward	TGCCACTAGTCTCTAGTCAG
S1-ROI-1_Reversed	CACAGTTGCTGATTCTCTTCC
S1-ROI-2_Forward	CAGATGCTGTAGACTGTGC
S1-ROI-2_Reversed	TGACTAGCTACACTACGTGC
Amplified region	21,585–23,631
Factor of reduction compared with the whole genome	~13.1

To assess the factor of sequencing time reduction, we used the S1-ROI panel and the whole-genome "midnight" panel [11], respectively, to amplify an otherwise identical sample from the same starting material. Subsequently, the amplified DNA was subjected to another PCR using the respective primer panel to generate two equally concentrated pools with sufficient material for the downstream experiment. Therefore, both pools were adjusted to 30 ng/ μ L, and each pool was used to generate 48 barcoded samples. The resulting 96 samples were sequenced as described below.

All libraries were sequenced on R9 flow cells on a MinION (Oxford Nanopore, Oxford, UK). Sequences were analyzed using the ARTIC pipeline [7,12] in combination with our newly developed S-Protein-Typer tool.

When the DNA sequences were fed to the S-Protein-Typer, it performed a translation and alignment operation, before identifying the amino acid differences, compared with the wild-type strain, and reporting them. The resulting mutations were then evaluated by a random forest classifier, which matched them with any of the known variants of concern (VOC).

Files for processing sequence data resulting from our primer panel, the developed S-Protein-Typer tool, and instructions for installation are available at <https://github.com/MassimoGregorioTotaro/s-protein-typer> (accessed on 18 December 2021).

Table 2. Sequencing summary of nanopore runs comparing our S1-ROI primer panel and the whole-genome “midnight” panel suggested by Oxford Nanopore: (A) a 20 min MinION run of a balanced pool of 48 S1-ROI and 48 Midnight samples each to assess the factor of time reduction; (B) six patient samples were sequenced with both panels on a MinION for 15 min to ensure applicability of the approach in case of original respiratory samples. Sequencing time was reduced compared to (A) due to the lower number of samples; (C) the library used in (B) was also sequenced on a Flongle to reconfirm the results in this setting and to assess overall Flongle output. As Flongle outputs are generally much lower, the flow cell was run till the end (18 h). Sequencing coverage was calculated by dividing the mean output per sample by the respective target (S1-ROI: 2.275 kbp; Midnight: 29.903 kbp). The factor of S1-ROI sequencing time reduction (sequencing coverage S1-ROI/sequencing coverage midnight), exemplifies the time reduction to obtain the same coverage.

	(A) MinION Balanced Samples		(B) MinION Patient Samples		(C) Flongle Patient Samples	
	S1-ROI	Midnight	S1-ROI	Midnight	S1-ROI	Midnight
Primer panel						
Number of samples	48	48	6	6	6	6
Mean output per sample [kbp]	603.44	602.82	4225.80	4656.64	1576.81	1792.32
SD [kbp]	257.29	257.18	1162.81	1541.23	420.77	355.29
Sequencing coverage	265	20	1857	156	693	60
Factor of S1-ROI sequencing time reduction	13		12		12	

For automated analysis of S-protein amino acid mutations, we developed a versatile tool (Figure 1B) that can handle data regardless of the sequencing technology used and that can be adapted to any viral protein with minor changes in the codebase. As input files, single and concatenated sequences in FASTA format are accepted. These are extensively elaborated before translation to ensure analytical robustness. Once the reading frame is identified, the mutations and predicted VOC are presented to the user. Our tool correctly identified all mutations of the respective variants when challenged with data from our collection and 50 deposited sequences, 10 of each VOC (as of December 2021: Alpha, Beta, Gamma, Delta, and Omicron); see Supplementary Material for corresponding NCBI Sequence Accession numbers.

4. Discussion

Here, we report a rapid and cost-saving approach for the detection of SARS-CoV-2 S-protein variants, by nanopore sequencing, based on sequencing only an ROI rather than the whole genome. Given that currently authorized SARS-CoV-2 vaccines rely almost exclusively on the S-protein’s antigenicity and the associated immune protection, the surveillance of its variants is of importance due to high relevance for potential vaccine escape and/or infectivity [13,14].

The observations summarized by Harvey et al. [4], indicating that most epitopes for neutralization belong to the receptor-binding domain (target of 90% of neutralizing antibody activity) and the amino-terminal domain, prompted us to design a specific “S1-ROI” primer set consisting of just two primer pairs. Amplicon length and overhangs also allow for Sanger sequencing, should a MinION be unavailable or time not be a priority.

Our ROI-focused approach frees sequencing capacity per isolate due to the reduction in the sequenced region, hence no sequencing capacity “is lost” to other parts of the genome. Therefore, as the DNA amount per flow cell is fixed, one can (I) increase the number of samples or (II) reduce sequencing time. This is demonstrated by our proof-of-concept experiments with balanced samples, showing that it is possible to obtain the same coverage by a factor of 13 faster using our panel. Moreover, our primer panel cuts initial cost substantially compared to the whole-genome panels (e.g., 4 primers instead of 58 “midnight” panel primers).

Validating our approach with six patient samples on a MinION and Flongle flow cell, respectively, showed very similar results with a factor of reduction of $\sim 12\times$ in both cases. The Flongle is the cheapest nanopore flow cell but comes with the lowest sequencing capacity. Nevertheless, the overall Flongle output (20.2 Mbp) we obtained would allow for 40 samples with $200\times$ coverage, effectively reducing its price/sample to ~ 1.5 EUR.

Notably, these advantages come at the cost of sacrificing information about mutations in other parts of the genome, which might contribute to virulence, transmissibility, etc. In such a case, however, the ROI can easily be extended by designing additional primers to complement our panel.

A straightforward and intuitive analysis of sequencing results is key for a broad application. Our automated tool identifies S-protein mutations, compared with the wild-type, and reports the alterations on the amino acid level. Thereby, the evaluator can easily spot combinations of adverse mutations in case of newly emerging variants. In combination with tools such as InterARTIC [15], easy reconstruction and analysis of the S-protein sequence can be obtained with minimal bioinformatics expertise once the pipeline is set up.

In conclusion, our strategy enables low threshold access to fast molecular surveillance of SARS-CoV-2 S-protein variants with regard to cost and bioinformatics knowledge. Such strategies will be essential during the next waves of the pandemic to intensify sequencing efforts for close monitoring of circulating variants, in Europe and especially in lower-income countries. Ultimately, the workflow and tools described can easily be adapted and implemented for other viruses, rendering it a general strategy for surveillance of variants of concern.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/v13122548/s1>, Table S1: NCBI sequence accession numbers used for validation of the S-Protein-Typer.

Author Contributions: Conceptualization, G.E.W., D.v.L., and I.S.; methodology, G.E.W., D.v.L., G.O., and I.S.; software, M.G.T., and G.O.; validation, G.E.W., A.V., M.L., S.S., D.v.L., G.O., and I.S.; formal analysis, G.E.W., M.G.T., A.V., D.v.L., and I.S.; investigation, G.E.W., A.V., M.L., S.S., P.F., G.O., and I.S.; resources, D.v.L., G.O., and I.S.; data curation, G.E.W., M.G.T., M.L., S.S., and P.F.; writing—original draft preparation, G.E.W., M.G.T., A.V., S.L., G.O., and I.S.; writing—review and editing, G.E.W., G.O., S.L., and I.S.; visualization, G.E.W., A.V., and M.G.T.; supervision, G.E.W., D.v.L., G.O., and I.S.; project administration, G.E.W. and I.S.; funding acquisition, D.v.L., G.O., and I.S. All authors have read and agreed to the published version of the manuscript.

Funding: M.G.T. and G.O. were supported by a grant from the Austrian Science Fund (FWF): P 30826.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Singh, J.; Rahman, S.A.; Ehtesham, N.Z.; Hira, S.; Hasnain, S.E. SARS-CoV-2 variants of concern are emerging in India. *Nat. Med.* **2021**, *27*, 1131–1133. [[CrossRef](#)] [[PubMed](#)]
2. Singh, J.; Samal, J.; Kumar, V.; Sharma, J.; Agrawal, U.; Ehtesham, N.Z.; Sundar, D.; Rahman, S.A.; Hira, S.; Hasnain, S.E. Structure–function analyses of new SARS-CoV-2 variants b.1.1.7, b.1.351 and b.1.1.28.1: Clinical, diagnostic, therapeutic and public health implications. *Viruses* **2021**, *13*, 439. [[CrossRef](#)] [[PubMed](#)]
3. Weisblum, Y.; Schmidt, F.; Zhang, F.; DaSilva, J.; Poston, D.; Lorenzi, J.C.; Muecksch, F.; Rutkowska, M.; Hoffmann, H.H.; Michailidis, E.; et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* **2020**, *9*, e61312. [[CrossRef](#)] [[PubMed](#)]
4. Harvey, W.T.; Carabelli, A.M.; Jackson, B.; Gupta, R.K.; Thomson, E.C.; Harrison, E.M.; Ludden, C.; Reeve, R.; Rambaut, A.; Consortium, C.-G.U.; et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424. [[CrossRef](#)] [[PubMed](#)]

5. Walker, A.; Houwaart, T.; Wienemann, T.; Vasconcelos, M.K.; Strelow, D.; Senff, T.; Hülse, L.; Adams, O.; Andree, M.; Hauka, S.; et al. Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, north-rhine westphalia, germany, february and march 2020. *Eurosurveillance* **2020**, *25*, 2000746. [[CrossRef](#)] [[PubMed](#)]
6. Bull, R.A.; Adikari, T.N.; Ferguson, J.M.; Hammond, J.M.; Stevanovski, I.; Beukers, A.G.; Naing, Z.; Yeang, M.; Verich, A.; Gamaarachchi, H.; et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat. Commun.* **2020**, *11*, 6272. [[CrossRef](#)] [[PubMed](#)]
7. Tyson, J.R.; James, P.; Stoddart, D.; Sparks, N.; Wickenhagen, A.; Hall, G.; Choi, J.H.; Lapointe, H.; Kamelian, K.; Smith, A.D.; et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* **2020**. [[CrossRef](#)]
8. Daniels, R.S.; Harvey, R.; Ermetal, B.; Xiang, Z.; Galiano, M.; Adams, L.; McCauley, J.W. A sanger sequencing protocol for SARS-CoV-2 S-gene. *Influenza Other Respir. Viruses* **2021**, *15*, 707–710. [[CrossRef](#)] [[PubMed](#)]
9. Boudet, A.; Stephan, R.; Bravo, S.; Sasso, M.; Lavigne, J.-P. Limitation of screening of different variants of SARS-CoV-2 by rt-pcr. *Diagnostics* **2021**, *11*, 1241. [[CrossRef](#)] [[PubMed](#)]
10. Wang, H.; Miller, J.A.; Verghese, M.; Sibai, M.; Solis, D.; Mfuh, K.O.; Jiang, B.; Iwai, N.; Mar, M.; Huang, C.; et al. Multiplex SARS-CoV-2 genotyping reverse transcriptase PCR for population-level variant screening and epidemiologic surveillance. *J. Clin. Microbiol.* **2021**, *59*, e0085921. [[CrossRef](#)] [[PubMed](#)]
11. Freed, N.E.; Vlkova, M.; Faisal, M.B.; Silander, O.K. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore rapid barcoding. *Biol. Methods Protoc.* **2020**, *5*, bpaa014. [[CrossRef](#)] [[PubMed](#)]
12. Quick, J.; Grubaugh, N.D.; Pullan, S.T.; Claro, I.M.; Smith, A.D.; Gangavarapu, K.; Oliveira, G.; Robles-Sikisaka, R.; Rogers, T.F.; Beutler, N.A.; et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **2017**, *12*, 1261–1276. [[CrossRef](#)] [[PubMed](#)]
13. Krammer, F. SARS-CoV-2 vaccines in development. *Nature* **2020**, *586*, 516–527. [[CrossRef](#)] [[PubMed](#)]
14. Creech, C.B.; Walker, S.C.; Samuels, R.J. SARS-CoV-2 vaccines. *JAMA* **2021**, *325*, 1318–1320. [[CrossRef](#)] [[PubMed](#)]
15. Ferguson, J.M.; Gamaarachchi, H.; Nguyen, T.; Gollon, A.; Tong, S.; Aquilina-Reid, C.; Bowen-James, R.; Deveson, I.W. Interartic: An interactive web application for whole-genome nanopore sequencing analysis of SARS-CoV-2 and other viruses. *bioRxiv* **2021**. [[CrossRef](#)] [[PubMed](#)]