Simplified point-of-care full SARS-CoV-2 genome sequencing using nanopore technology

Anton Pembaur[1], Erwan Sallard[2], Patrick Philipp Weil[1], Jennifer Ortelt[3], Parviz Ahmad-Nejad[3], Jan Postberg[1,†]

[†]corresponding author

[1]Clinical Molecular Genetics and Epigenetics, Faculty of Health, Centre for Biomedical Education & Research (ZBAF), Witten/Herdecke University, Alfred-Herrhausen-Str. 50, 58448 Witten, Germany

[2]Institute of Virology and Microbiology, Faculty of Health, Centre for Biomedical Education & Research (ZBAF), Witten/Herdecke University, Stockumer Str. 10, 58453 Witten, Germany

[3]HELIOS University Hospital Wuppertal, Institute of Medical Laboratory Diagnostics, Centre for Clinical & Translational Research (CCTR), Witten/Herdecke University, Heusnerstr. 40, 42283 Wuppertal, Germany

E-Mail addresses:

Anton Pembaur: anton.pembaur@uni-wh.de

Erwan Sallard: erwan.sallard@uni-wh.de

Patrick Philipp Weil: patrick.weil@uni-wh.de

Jennifer Ortelt: jennifer.ortelt@helios-gesundheit.de

Parviz Ahmad-Nejad: parviz.ahmad-nejad@helios-gesundheit.de

Jan Postberg: jan.postberg@uni-wh.de

Abstract

Background: The scale of the ongoing SARS-CoV-2 pandemic warrants the urgent establishment a global decentralized surveillance and warning system to recognize local outbreaks and the emergence of novel variants-of-concern. Among the available deep-sequencing technologies, nanopore-sequencing could be an important cornerstone, since it is mobile, scalable and acquisition investments are comparably low. Therefore, streamlined and efficient nanopore-sequencing protocols need to be developed and optimized for SARS-CoV-2 variants identification, in particular for smaller hospital laboratories with lower throughput.

Results: We adapted and simplified existing workflows using the 'midnight' 1,200 bp amplicon split primer sets for PCR, which produce tiled overlapping amplicons covering almost all of the SARS-CoV-2 genome. Subsequently, we applied the Oxford Nanopore Rapid barcoding protocol and the portable MinION Mk1C sequencer in combination with the ARTIC bioinformatics pipeline. We tested the simplified and less time-consuming workflow on confirmed SARS-CoV-2-positive specimens from clinical routine and identified pre-analytical parameters, which may help to decrease the rate of sequencing failures. Duration of the complete pipeline was approx. 7 hrs for one specimen and approx. 11 hrs for 12 multiplexed barcoded specimens.

Conclusions: The adapted protocol contains less processing steps. Diagnostic CT values are the principal criteria for specimen selection. Subsequent to diagnostic qRT-PCR, multiplex library preparation, quality controls, nanopore sequencing and the bioinformatic pipeline can be completely conducted within one working-day.

Keywords

(+)RNA genome sequencing, COVID-19 surveillance, variant-of-concern (VOC)

Background

The scale of the ongoing SARS-CoV-2 pandemic warrants the urgency of establishing a global decentralized warning system to recognize local outbreaks and the emergence of novel variants-of-concern (VOC). In particular, surveillance of SARS-CoV-2 evolution aims to recognize VOCs with accelerated transmission rates, increased infectiousness and those potentially evolving immune escape mutations, and which require adaptations in containment and vaccination strategies (Cobey, Larremore et al. 2021, Gupta 2021, Harvey, Carabelli et al. 2021, Winger and Caspari 2021). Moreover, the search for the zoonotic origin of SARS-CoV-2 from comparative analyses of genomic data is an ongoing issue with relevance for the early recognition of future outbreak scenarios (Andersen, Rambaut et al. 2020, V'Kovski, Kzel et al. 2021).

Among the available deep-sequencing technologies, nanopore-sequencing could be an important cornerstone, since it is mobile, scalable and acquisition investments are comparably low. Further, the use of nanopore sequencing devices is associated with relatively low demands on IT infrastructure. In

70 the past they were already involved actions of genome surveillance, e.g. the 2015 Ebola outbreak in
71 Liberia, Guinea and sierra Leone (Quick, Loman et al. 2016), and since then there has been a
72 substantial technological progress. However, at least for smaller hospital laboratories with lower
73 throughput, it is desirable to develop protocols as streamlined as possible.
74 Nanopore sequencing allows the sequencing of either DNA or RNA (Garalde, Snell et al. 2018), and
75 does not require PCR amplification. Furthermore the technique has the potential of producing very
76 long, continuous reads, which theoretically allows to sequence in only one read the 29.903nt long
77 (+)RNA genome of SARS-CoV-2 (Taiaroa, Rawlinson et al. 2020), or its deriving cDNA after reverse
78 transcription (Figure 1).
79 With the aims of facilitating implementation in routine diagnostics with lower specimens throughput
80 and of simplifying the workflow, we tested modifications of existing ARTIC protocols for SARS-
81 CoV-2 full length (+)RNA genome sequencing (Freed, Vlkova et al. 2020, Li, Wang et al. 2020,
82 Tyson, James et al. 2020). Furthermore, we tested the simplified and less time-consuming workflow
83 on confirmed SARS-CoV-2-positive specimens from clinical routine and identified pre-analytical
84 parameters, which may help to decrease the rate of sequencing failures.
85
86
87 **Methods**
88 *Nucleic acids isolation*
89 Specimens included nasopharyngeal swabs (Xebios Diagnostics), which underwent routine COVID-19
90 diagnostic testing. RNA purification was performed via magnetic beads (Seegene NIMBUS/Tanbead).
91 Alternatively, total RNA was purified from 250 µL liquid specimen using 750 µL QIAzol lysis
92 reagent (Qiagen, Cat. No. 158845) upon manufacturer's recommendations, or using silica columns
93 (QiaAmp Viral Mini Kit, Qiagen).
94
95 *RT-PCR, quality assessments and library preparation*
96 A split primer set from the ARTIC protocol was used for 2 multiplex PCR reactions. To avoid overlaps
97 during multiplex PCR, each single-tube PCR reaction generates consecutively tiled, non-overlapping
98 1,200 bp amplicons. Mixed together after PCR, both resulting complementary amplicon mixtures cover
99 almost the entire SARS-CoV-2 genome (Freed, Vlkova et al. 2020).
100 Combined reverse transcription and amplification of multiple 1,200 bp amplicons (RT-PCR) was
101 performed in single tube 20 µL reactions using the Luna One-Step RT-qPCR Kit (NEB; E3005). For
102 RT-PCR, 8 µL of purified template RNA were used for each reaction. From 100 µM primer pools, 1
103 µL was used in each reaction. Reverse transcription was performed at 55°C for 30 minutes, followed
104 by incubation at 95°C for one minute. Then 34 cycles (pool 1) or 30 cycles (pool 2) of denaturation at
105 95°C for 20 seconds and annealing and extension in one step at 60°C for 210 seconds were performed.
106 A final extension was performed at 65°C. During the implementation phase amplicon sizes and DNA

concentrations were routinely checked by agarose gels or by microvolume electrophoresis (Agilent Bioanalyzer, Agilent DNA 12000 kit). Thereafter, amplicons from primer pools 1 and 2 were quantified by fluorimetry (Promega Quantus) and then mixed at equal concentrations. Library preparation was done using the Rapid Barcoding Sequencing Kit (Oxford nanopore; SQK-RBK004) upon manufacturers recommendations.

Nanopore sequencing

Sequencing was performed on a MinION Mk1C with the options 'basecalling' and 'demultiplexing' being enabled. As output format, FAST5 and FASTQ files were chosen. Sequencing time was set for 72 hours as default. Sequencing was stopped after reaching at least 10 megabases for each barcode.

Bioinformatics

The Oxford Nanopore MinION software includes guppy bascalling, adapter trimming and barcode identification and yields FASTQ and FAST5 files for each barcode with a quality filter. Then these barcode-sorted files are fed in the interARTIC pipeline. Except otherwise stated, the Nanopolish algorithm was routinely used.

Installation and usage of interARTIC pipeline was done following the developers' instructions: https://psy-fer.github.io/interARTIC/installation/. For faster analysis, an activation of all available threads in the advanced settings was done. This pipeline performs read filtering, alignments and returns a consensus FASTA file as well as coverage charts for visualization are generated (Figure 2B-H). Consensus FASTA files were uploaded to the Nextstrain webapp (https://clades.nextstrain.org) to perform phylogenetic analyses (Hadfield, Megill et al. 2018).

Data availability

FASTQ files and assembled FASTA formatted consensus sequences are available (https://t1p.de/minion-seqdata)

Results and Discussion

We defined milestones to achieve protocol simplification in order to reduce hands-on sample time and working step numbers. 1. We tested whether specimens can be directly taken from residual diagnostic specimens in 96-deepwell-plates obtained using magnetic beads (Seegene NIMBUS/Tanbead). For comparison we applied RNA purification protocols using silica columns (QiaAmp Viral Mini Kit, Qiagen) or guanidinium isothiocyanate (GITC) for RNA extraction (QIAzol lysis reagent, Qiagen). 2. To test whether reverse transcription can be successfully primed by SARS-CoV-2-specific primers, which are subsequently used for multiplex 1,200 bp amplicon amplification, purified RNA from different extraction protocols was used for cDNA synthesis and successive multiplex PCR with the 2

144 split 1,200 bo amplicon primer pools in single tube reactions 3. We aimed on improving the efficiency
145 of the nanopore sequencing workflow through exploitation of the onboard Guppy basecalling
146 capability of the Oxford Nanopore MinION Mk1C device. Moreover, implementation of the
147 interARTIC interface should help to avoid command line-based bioinformatics analyses as far as
148 possible in order to provide a user-friendly and efficient analysis pipeline.
149
150 *Protocol implementation considering differences in viral loads and influences of RNA extraction*
151 *protocols*
152 For protocol implementation we made use of serially diluted specimens of purified RNA from
153 patients' samples, which exhibited low cycle threshold (CT) value (CT=16-18) after routine
154 diagnostics RT-qPCR (N gene, RdRP gene; Seegene). Using the quantitative reference sample
155 Ch07470 for calibration we determined that a CT=25 corresponded to a SARS-CoV-2 copy number of
156 1.0x 10^6. Routinely used dilution factors were 2^0-2^{-10} in order to cover several CT value magnitudes
157 and to simulate different amounts of viral loads (CT=16 → approx. 5.12x 10^8 viral copies; CT=18 →
158 approx. 1.28x 10^8 viral copies). In terms of RNA yield and 1,200 bp amplicon PCR performance the
159 magnetic beads-based RNA purification protocol outperformed slightly the GITC method as well as
160 the column-based protocol, but we did not observe differences in read and coverage quality between
161 these different isolation methods (data not shown). Since direct sampling from 96-deepwell plates
162 allowed us to directly exploit residual specimens, which remained during routine RT-qPCR
163 diagnostics, we decided to focus on magnetic beads-based RNA purification during further protocol
164 development. Moreover, it shortened and simplified the workflow. After semiquantitative or
165 quantitative RT-PCR using multiplex primer pools 1 and 2 in separate single tube reactions for
166 combined reverse transcription of the SARS-CoV-2 (+)RNA and amplification of 1,200 bp amplicons,
167 band intensities exhibited strong dependence on viral loads. Moreover, after 32 PCR cycles band
168 intensities using primer pool 1 were weaker when compared with primer pool 2 (Figure S1). We
169 determined that 34 PCR cycles for primer pool 1 and 30 cycles for primer pool 2 were a good
170 compromise.
171
172 *Surveillance of multiplex nanopore sequencing and multiple reuses of flow cells*
173 Using the Rapid Barcoding Kit (SQK-RBK004, Oxford Nanopore) allowed us quick barcoding of the
174 PCR products without further purification steps being required and their immediate use for nanopore
175 sequencing. As a standard, we used 12 barcoded libraries for multiplex sequencing on R9 flow cells.
176 In contrast to the Oxford Nanopore MinION Mk1B, the MinION Mk1C device features onboard
177 guppy basecalling. In combination with the Rapid Barcoding Kit, we exploited this opportunity for
178 real-time surveillance of basecalling and demultiplexing for each of the 12 multiplexed samples per
179 run. This enabled us to recognize the exact time point at which a reading depth of approx. 10 Mbp per
180 barcode was achieved. Here, the viral load influence (simulated by serially diluted samples) was also

181 observable during sequencing. decreased viral loads let to a considerable decrease of passed reads
182 (Figure 2A). A manual stop of sequencing followed by flow cell washing (Flow Cell Wash Kit, EXP-
183 WSH004, Oxford Nanopore) allowed us to reuse a single flow cell for a series of 3 sequencing runs,
184 each using 12 multiplexed barcoded libraries. This specific scenario resulted in costs of approx. 40
185 USD per sample. It needs to be considered that theoretically the possible numbers of reuses depends
186 largely on the duration of the sequencing, which itself depends mainly on the number of used
187 barcodes. Further, if a MinION Mk1B is considered for this purpose, a similar surveillance
188 functionality could be achieved using RAMPART on a dedicated LINUX environment.
189
190 *Assembly of full SARS-CoV-2 genomes and pathogen genome data analyses*
191 For mapping and full length SARS-CoV-2 genome assembly, we used the FASTQ files resulting from
192 Guppy basecalling in order to generate FASTA formatted consensus sequence files. We used the
193 ARTIC pipeline through a graphical user interface (interARTIC; https://github.com/Psy-
194 Fer/interARTIC). Once installed, this is an easy to use and relatively fast pipeline with only five
195 minutes hands-on time, which enables the use of the Nanopolish or Medaka algorithms for
196 simultaneous analyses of multiplexed barcoded samples. As an example, Figure 2B shows the
197 complete and deep coverage of the complete SARS-CoV-2 genome after the combination of pools 1
198 (light blue) and 2 (pink). This demonstrates that all contained 1,200 bp amplicon were specifically and
199 efficiently amplified during the combined RT-PCR reaction (Figure 2B).
200 To compare the interARTIC and Geneious Prime pipelines, we used exactly the same FASTQ file
201 from the same sample shown in Figure 2B (https://go.geneious.com/video/how-to-assemble-
202 coronavirus-genomes). Phylogenetic analyses with Nextstrain using the FASTA consensus files
203 obtained from the interARTIC or the Geneious Prime pipelines resulted in considerably different
204 phylogenetic distances in clade 20I, showing that the bioinformatics pipeline influences the result
205 (Figure S2). the interARTIC pipeline proved superior in terms of coverage and sequencing depth.
206 Notably, within the interARTIC pipeline both options, the 'Nanopolish' and 'Medaka' algorithms
207 performed equally well with respect to consensus sequence quality, but 'Medaka' was considerably
208 faster.
209 We observed that the consensus sequences returned by the Nanopolish and Medaka algorithms contain
210 numerous unsolved regions (for which only 'N's are indicated). Interestingly, the regions unsolved by
211 one algorithm were generally solved by the other, so we assumed that the two consensus sequences
212 could be combined to produce a 'super-consensus' with improved variant prediction value. We
213 developed a Python code that merges the Nanopolish and Medaka consensus sequences and generates
214 the corresponding variant calling file. As expected, the 'super-consensus' contained less unsolved
215 regions when compared with the Nanopolish and Medaka consensus files alone. In addition, it retained
216 the high quality variants, which were identified by both algorithms alone, thus decreasing the numbers
217 of false-positive positions (probably caused by sequencing errors) (Figure 4A). Furthermore, the exact

218 location of variants compared to the reference sequence is more accurately identified by our code than
219 by Nanopolish or Medaka (Figure 4B). Consequently, our consensus-merging code improves the
220 quality of variant calling and highlights the complementarity of Nanopolish and Medaka for nanopore-
221 sequencing of SARS-CoV-2 and others. Nevertheless, this quality increase comes at the cost of
222 information loss such as the number of reads per variant or other metadata which were initially
223 generated by Nanopolish and Medaka and are not transferred to the 'super-consensus'.
224 Serial input RNA dilutions or, respectively, viral load influenced the depth of sequencing (Figure 2C-
225 H). For the output of high-quality consensus sequences in the FASTA file format a coverage threshold
226 of 20 was used as default. We generally observed that this could be reached when a SARS-CoV-2 titer
227 of 4x 10^6 was given. Viral copy numbers Lower than 4x 10^6 were associated with incompletely
228 assembled SARS-CoV-2 genomes. However, despite incomplete coverage in those cases enough
229 informative sequence data could be obtained for phylogenetic analyses from several low copy number
230 samples. Thus, we provide here a convincing line of evidence that the copy number-normalized CT
231 values of diagnostic RT-qPCR can be used as the criterion of sequencing success.
232 Single or batch high-quality consensus FASTA formatted sequences were used for phylogenetic tree
233 visualization and variant calling using the Nextstrain webapp (https://clades.nextstrain.org) (Hadfield,
234 Megill et al. 2018). In our hands, the obtained sequences could faithfully be assigned to specific clades
235 in the reference tree (Figure 3). Again, an influence of viral load was observed. As a result of serial
236 sample dilutions, we observed deviating phylogenetic distances within the clade, wherein specimens
237 classification occurred (Figure 3A), which eventually could lead to incorrect clade association. The
238 use of specimens from diagnostic routine with viral copy numbers higher than approx. 4x 10^6
239 apparently led to their faithful association with different clades, which were clades 20I and 19A in the
240 shown example (Figure 3B). The Nanopolish and Medaka consensus as well as the merged super-
241 consensus could be reliably associated to the corresponding clade (20I in the shown example). The
242 Medaka consensus mapped at a greater distance than the Nanopolish consensus, probably because the
243 Medaka algorithm does not correct frameshifts, while the super-consensus had an intermediary
244 distance between the two other consensus (Figure 4C).
245
246 Taken together, the main achievements of an optimized workflow are: 1. Purified RNA from SARS-
247 CoV-2-positive patients can be directly taken from residual diagnostic specimens in 96-deepwell-
248 plates; 2. cDNA synthesis and successive multiplex PCR with 2 split primer pools can be performed in
249 single tube reactions. Since cDNA synthesis is primed by SARS-CoV-2-specific primers for 1,200 bp
250 amplicon amplification, there is no need for use of unspecific hexanucleotide priming. 3. Onboard
251 Guppy basecalling with the Oxford Nanopore MinION Mk1C device and implementation of the
252 interARTIC led to a further reduction of working steps and hands-on time (Figure 5). Implementation
253 in smaller hospital laboratories with lower specimens' throughput can be easily done at moderate
254 costs.

We provide a detailed protocol here, which includes the Python code and corresponding command line: dx.doi.org/10.17504/protocols.io.bvstn6en

Conclusions

The adapted protocol contains less processing steps. Diagnostic CT values are the principal criteria for specimen selection. After diagnostic qRT-PCR, multiplex library preparation, quality controls, nanopore sequencing and the bioinformatic pipeline can be completely conducted within one working-day.

List of abbreviations

(+)RNA (sense polarity [single-stranded] ribonucleic acid), RdRP (RNA-dependent RNA polymerase), RT-qPCR (reverse transcription-quantitative polymerase chain reaction, SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)

Declarations

Ethics approval and consent to participate

For the collection and use of specimens at Helios University Hospital Wuppertal (North Rhine-Westphalia, Western Germany) we obtained approval of the Witten/Herdecke University Ethics board (No. 160/2020). All work has been conducted according to the principles expressed in the Declaration of Helsinki.

Competing interests

There are no conflicting interests for none of the authors, which need declaration.

Funding

The study was financed by own institutional means.

References

Andersen, K. G., A. Rambaut, W. I. Lipkin, E. C. Holmes and R. F. Garry (2020). "The proximal origin of SARS-CoV-2." Nat Med **26**(4): 450-452.

Cobey, S., D. B. Larremore, Y. H. Grad and M. Lipsitch (2021). "Concerns about SARS-CoV-2 evolution should not hold back efforts to expand vaccination." Nat Rev Immunol **21**(5): 330-335.

Freed, N. E., M. Vlkova, M. B. Faisal and O. K. Silander (2020). "Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding." Biol Methods Protoc **5**(1): bpaa014.

Garalde, D. R., E. A. Snell, D. Jachimowicz, B. Sipos, J. H. Lloyd, M. Bruce, N. Pantic, T. Admassu, P. James, A. Warland, M. Jordan, J. Ciccone, S. Serra, J. Keenan, S. Martin, L. McNeill, E. J. Wallace, L. Jayasinghe, C. Wright, J. Blasco, S. Young, D. Brocklebank, S. Juul, J. Clarke, A. J. Heron and D. J. Turner (2018). "Highly parallel direct RNA sequencing on an array of nanopores." Nat Methods **15**(3): 201-206.

Gupta, R. K. (2021). "Will SARS-CoV-2 variants of concern affect the promise of vaccines?" Nat Rev Immunol **21**(6): 340-341.
Hadfield, J., C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford and R. A. Neher (2018). "Nextstrain: real-time tracking of pathogen evolution." Bioinformatics **34**(23): 4121-4123.
Harvey, W. T., A. M. Carabelli, B. Jackson, R. K. Gupta, E. C. Thomson, E. M. Harrison, C. Ludden, R. Reeve, A. Rambaut, C.-G. U. Consortium, S. J. Peacock and D. L. Robertson (2021). "SARS-CoV-2 variants, spike mutations and immune escape." Nat Rev Microbiol **19**(7): 409-424.
Li, J., H. Wang, L. Mao, H. Yu, X. Yu, Z. Sun, X. Qian, S. Cheng, S. Chen, J. Chen, J. Pan, J. Shi and X. Wang (2020). "Rapid genomic characterization of SARS-CoV-2 viruses from clinical specimens using nanopore sequencing." Sci Rep **10**(1): 17492.
Quick, J., N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, J. A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouedraogo, B. Afrough, A. Bah, J. H. Baum, B. Becker-Ziaja, J. P. Boettcher, M. Cabeza-Cabrerizo, A. Camino-Sanchez, L. L. Carter, J. Doerrbecker, T. Enkirch, I. G. G. Dorival, N. Hetzelt, J. Hinzmann, T. Holm, L. E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C. H. Logue, A. Mazzarelli, S. Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E. Pallash, L. V. Patrono, J. Portmann, J. G. Repits, N. Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R. L. Yemanaberhan, E. G. Zekeng, R. Trina, A. Bello, A. A. Sall, O. Faye, O. Faye, N. Magassouba, C. V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, A. Camara, M. Gerard, G. Bado, B. Baillet, D. Delaune, K. Y. Nebie, A. Diarra, Y. Savane, R. B. Pallawo, G. J. Gutierrez, N. Milhano, I. Roger, C. J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D. Turner, G. Pollakis, J. A. Hiscox, D. A. Matthews, M. K. O'Shea, A. M. Johnston, D. Wilson, E. Hutley, E. Smit, A. Di Caro, R. Woelfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keita, A. Rambaut, P. Formenty, S. Gunther and M. W. Carroll (2016). "Real-time, portable genome sequencing for Ebola surveillance." Nature **530**(7589): 228-232.
Taiaroa, G., D. Rawlinson, L. Featherstone, M. Pitt, L. Caly, J. Druce, D. Purcell, L. Harty, T. Tran, J. Roberts, M. Catton, D. Williamson, L. Coin and S. Duchene (2020). "Direct RNA sequencing and early evolution of SARS-CoV-2." bioRxiv: 2020.2003.2005.976167.
Tyson, J. R., P. James, D. Stoddart, N. Sparks, A. Wickenhagen, G. Hall, J. H. Choi, H. Lapointe, K. Kamelian, A. D. Smith, N. Prystajecky, I. Goodfellow, S. J. Wilson, R. Harrigan, T. P. Snutch, N. J. Loman and J. Quick (2020). "Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore." bioRxiv: 2020.2009.2004.283077.
V'Kovski, P., A. Kratzel, S. Steiner, H. Stalder and V. Thiel (2021). "Coronavirus biology and replication: implications for SARS-CoV-2." Nat Rev Microbiol **19**(3): 155-170.
Weil, P. P., J. Hentschel, F. Schult, A. Pembaur, B. Ghebremedhin, O. Mboma, A. Heusch, A. C. Reuter, D. Muller, S. Wirth, M. Aydin, A. C. W. Jenke and J. Postberg (2021). "Combined RT-qPCR and pyrosequencing of a Spike glycoprotein polybasic cleavage motif can uncover pediatric SARS-CoV-2 infections associated with heterogeneous presentation." Mol Cell Pediatr **8**(1): 4.
Winger, A. and T. Caspari (2021). "The Spike of Concern-The Novel Variants of SARS-CoV-2." Viruses **13**(6).

Figures and Figure legends

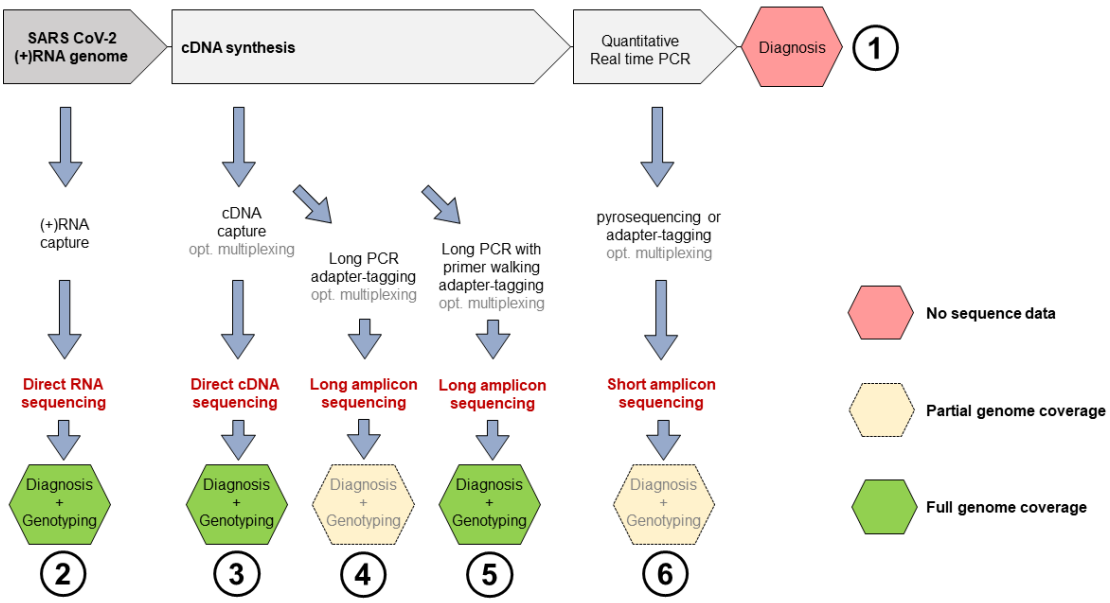


Figure 1. Nanopore sequencing options (2-6) compatible with diagnostic qPCR pipelines (1). The pyrosequencing option in path (6) requires biotin-tagged primers for qPCR (Weil, Hentschel et al. 2021).

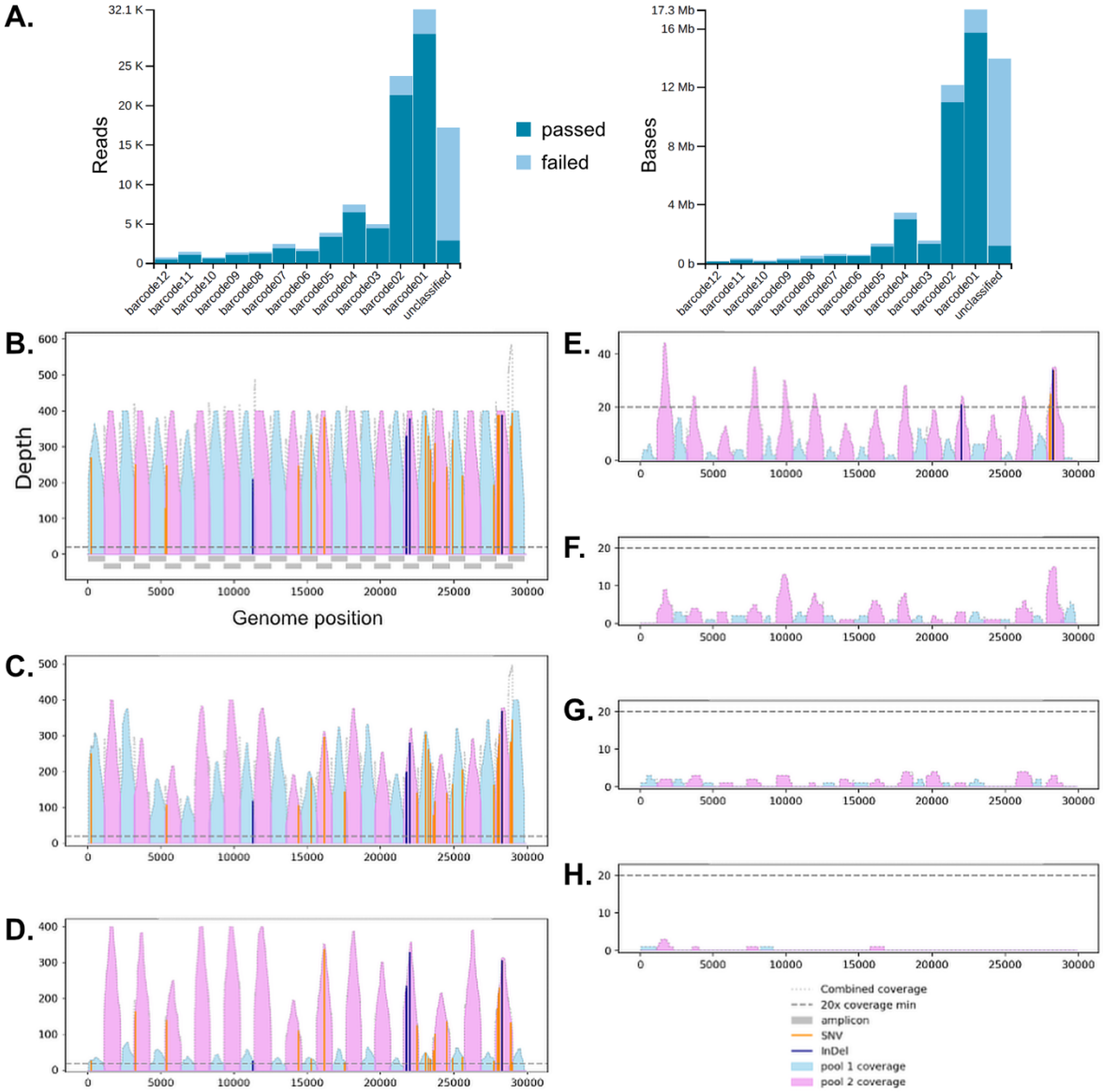


Figure 2. Surveillance using nanopore sequencing (A.), and effects of specimens dilution on the SARS-CoV-2 genome coverage (B.), whereby viral copy numbers were 1.28x 10^6 (B.); 5.12x 10^8 (C.); 2.56x 10^8 (D.); 4x 10^6 (E.); 2x 10^6 (F.); 1x 10^6 (G.); 5x 10^5 (H.). In (A.) increasing barcode numbers (X-axis) correspond to decreasing viral titers.

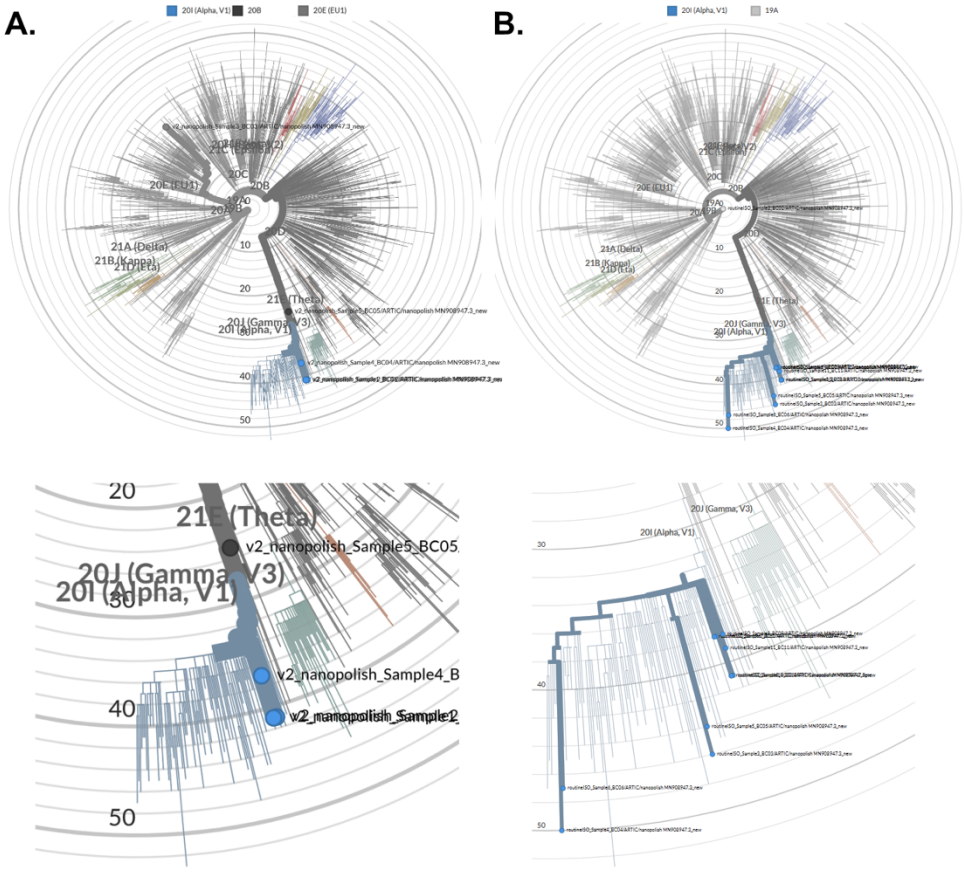
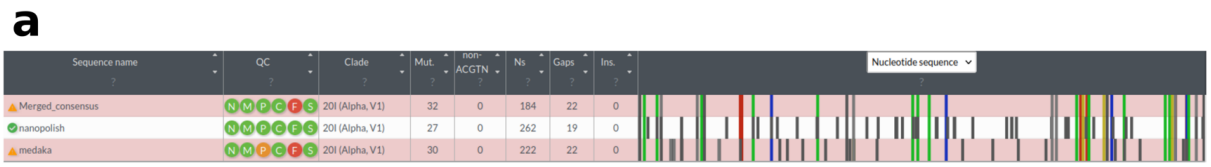
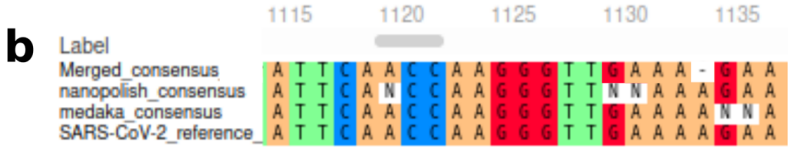


Figure 3. Comparison on the evolutionary distance calculation, when data from the same specimen was processed by 2 different bioinformatics pipelines. Phylogenetic tree visualization was done using the Nextstrain open-source platform for pathogen genome data analyses (Hadfield, Megill et al. 2018). We used serial dilutions of identical specimens (A.), or a selection of samples from different individuals from clinical routine (B.) for phylogenetic analyses. Below the trees a magnification of clade 20I is shown, wherein most specimens grouped (A,B).





consensus	position	mutation
Nanopolish	1120	A -> G
	1130	GA -> G
Medaka	1135	GA -> G
merged	1133	A -> -

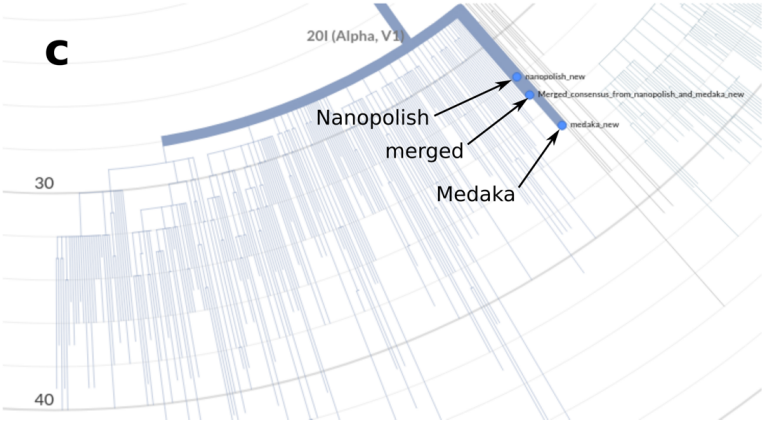


Figure 4. Comparison of the consensus sequences returned by the Nanopolish and Medaka algorithm and the merged consensus for a nanopore sequencing run corresponding to a viral load of 1.28x10**6. **(a)** Alignment of the Nanopolish and Medaka consensus as well as the merged consensus on the SARS-CoV-2 reference genome by the Nextstrain program. The merged consensus conserved all of the high quality mutations that mapped to the known variant 20I, while the non-matching mutations (likely sequencing errors) of the Nanopolish or Medaka consensus were mostly lost. **(b)** Comparison of variant solving by Nanopolish, Medaka and our code. The multiple alignment was performed by MAFFT online tool with the three consensus sequences and the SARS-CoV-2 reference genome (MN908947.3). The table summarizes the variants identified by the three algorithms in the region studied as reported in the .vcf file they return. Contrary to the Nanopolish and Medaka consensus, the merged consensus solved the entire region and led to accurate variant calling. **(c)** Clade mapping and phylogenetic distances calculated by Nextstrain for the three consensus sequences.

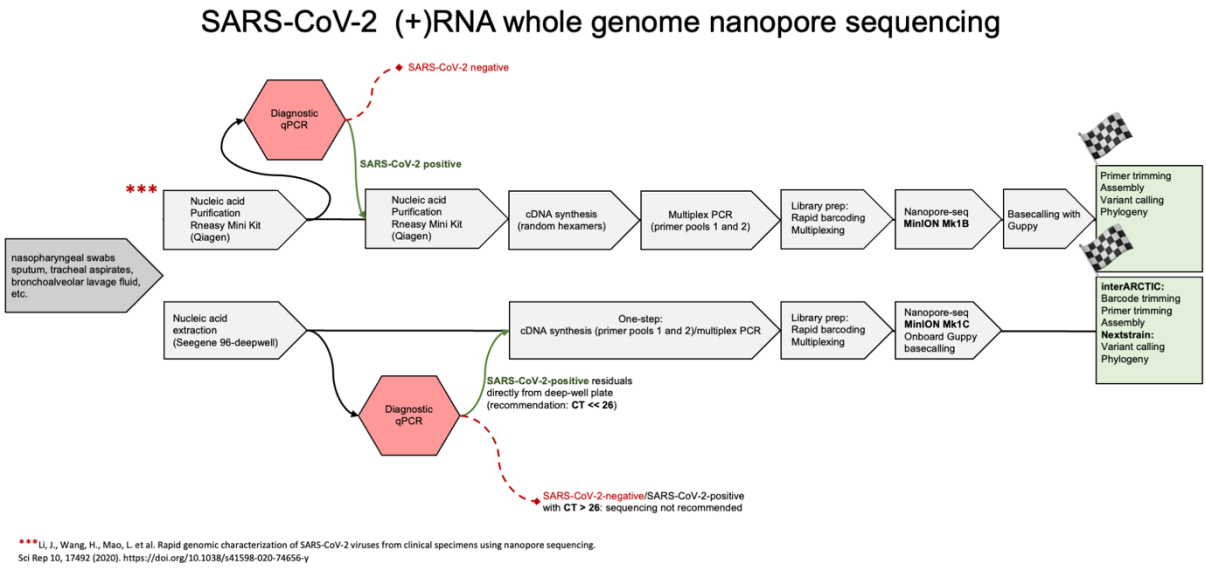


Figure 5. Comparison of a rapid SARS-CoV-2 whole (+)RNA genome nanopore sequencing pipeline (Freed, Vlkova et al. 2020) with the novel simplified workflow, whose main achievements are: (1) Purified RNA from SARS-CoV-2-positive patients can be directly taken from residual diagnostic specimens in 96-deepwell-plates; (2) cDNA synthesis and successive multiplex PCR with 2 primer pools can be performed in single tube reactions. Since cDNA synthesis is primed by SARS-CoV-2-specific primers for 1,200 bp amplicon amplification, there is no need for use of unspecific hexanucleotide priming. (3) Onboard Guppy basecalling with the Oxford Nanopore MinION Mk1C device and implementation of the interARTIC led to a further reduction of working steps and hands-on time.

Supplementary Information

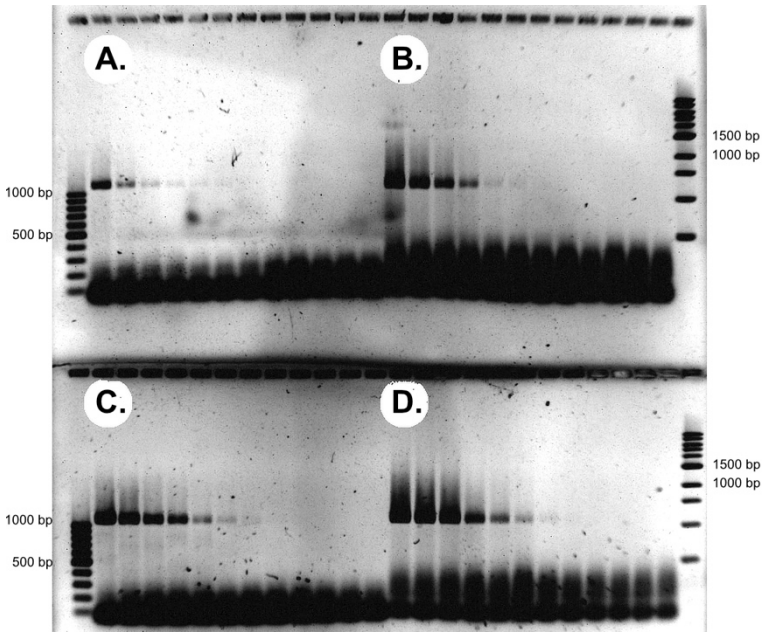


Figure S1. The influence of viral load on the amplification of 1,200 bp amplicon was analysed by agarose gel electrophoresis using samples, where RNA was serially diluted prior to cDNA synthesis and semiquantitative multiplex PCR (A, B) or quantitative multiplex PCR (C, D). Multiplex primer pool 1 (A, C) and pool 2 (B, D) were used in separate single tube reactions. (A-D.) From left to right: Decreasing viral loads (dilution factors 2^{0}-2^{-10} plus no template control).

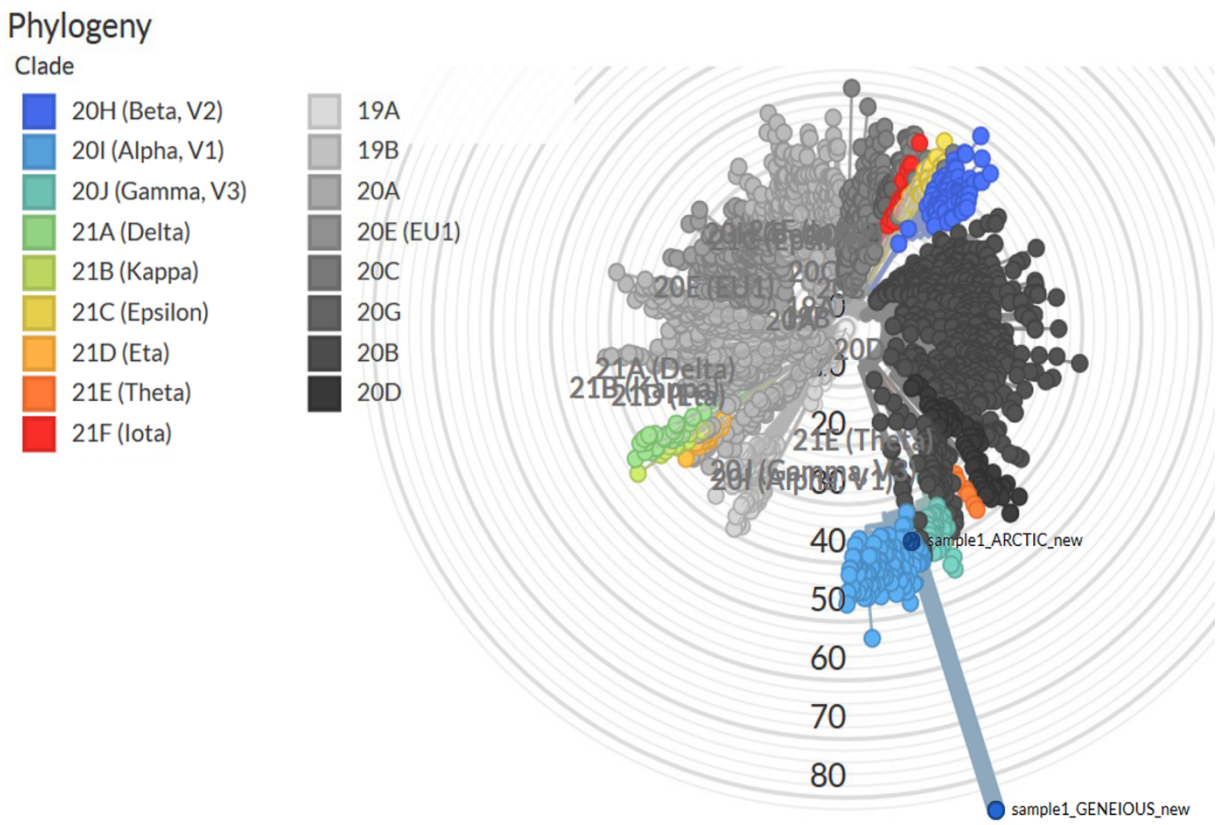


Figure S2. Comparison on the phylogenetic distance, when data from the same specimen was processed by 2 different bioinformatics pipelines Phylogenetic tree visualization was done using the Nextstrain web app (https://clades.nextstrain.org) for pathogen genome data analyses (Hadfield, Megill et al. 2018).