AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

VIROLOGY

Rapid and Sensitive Direct Detection and Identification of Poliovirus from Stool and Environmental Surveillance Samples by Use of Nanopore Sequencing

Alexander G. Shaw,[a] Manasi Majumdar,[b] Catherine Troman,[a] Áine O'Toole,[c] Blossom Benny,[d] Dilip Abraham,[d] Ira Praharaj,[e] Gagandeep Kang,[d] Salmaan Sharif,[f] Muhammad Masroor Alam,[f] Shahzad Shaukat,[f] Mehar Angez,[f] Adnan Khurshid,[f] Nayab Mahmood,[f] Yasir Arshad,[f] Lubna Rehman,[f] Ghulam Mujtaba,[f] Ribqa Akthar,[f] Muhammad Salman,[f] Dimitra Klapsa,[b] Yara Hajarha,[b] Humayun Asghar,[g] Ananda Bandyopadhyay,[h] Andrew Rambaut,[c] Javier Martin,[b] Nicholas Grassly[a]

[a]Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom
[b]Division of Virology, National Institute for Biological Standards and Control (NIBSC), Herts, United Kingdom
[c]Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Edinburgh, United Kingdom
[d]Division of Gastrointestinal Sciences, Christian Medical College, Vellore, India
[e]Division of Epidemiology & Communicable Diseases, Indian Council of Medical Research, New Delhi, India
[f]Department of Virology, National Institute for Health, Islamabad, Pakistan
[g]World Health Organization Eastern Mediterranean Regional Office, Amman, Jordan
[h]Bill and Melinda Gates Foundation, Seattle, Washington, USA

Javier Martin and Nicholas Grassly are joint senior authors.

ABSTRACT Global poliovirus surveillance involves virus isolation from stool and environmental samples, intratypic differential (ITD) by PCR, and sequencing of the VP1 region to distinguish vaccine (Sabin), vaccine-derived, and wild-type polioviruses and to ensure an appropriate response. This cell culture algorithm takes 2 to 3 weeks on average between sample receipt and sequencing. Direct detection of viral RNA using PCR allows faster detection but has traditionally faced challenges related to poor sensitivity and difficulties in sequencing common samples containing poliovirus and enterovirus mixtures. We present a nested PCR and nanopore sequencing protocol that allows rapid (<3 days) and sensitive direct detection and sequencing of polioviruses in stool and environmental samples. We developed barcoded primers and a real-time analysis platform that generate accurate VP1 consensus sequences from multiplexed samples. The sensitivity and specificity of our protocol compared with those of cell culture were 90.9% (95% confidence interval, 75.7% to 98.1%) and 99.2% (95.5% to 100.0%) for wild-type 1 poliovirus, 92.5% (79.6% to 98.4%) and 98.7% (95.4% to 99.8%) for vaccine and vaccine-derived serotype 2 poliovirus, and 88.3% (81.2% to 93.5%) and 93.2% (88.6% to 96.3%) for Sabin 1 and 3 poliovirus alone or in mixtures when tested on 155 stool samples in Pakistan. Variant analysis of sequencing reads also allowed the identification of polioviruses and enteroviruses in artificial mixtures and was able to distinguish complex mixtures of polioviruses in environmental samples. The median identity of consensus nanopore sequences with Sanger or Illumina sequences from the same samples was >99.9%. This novel method shows promise as a faster and safer alternative to cell culture for the detection and real-time sequencing of polioviruses in stool and environmental samples.

KEYWORDS enterovirus, environmental surveillance, nanopore sequencing, poliovirus, stool

Resurgent wild-type poliovirus transmission in Pakistan and Afghanistan and expanding outbreaks of vaccine-derived poliovirus (VDPV) in Africa pose the risk of reversing progress made by the Global Polio Eradication Initiative. Effective control of

Citation Shaw AG, Majumdar M, Troman C, O'Toole Á, Benny B, Abraham D, Praharaj I, Kang G, Sharif S, Alam MM, Shaukat S, Angez M, Khurshid A, Mahmood N, Arshad Y, Rehman L, Mujtaba G, Akthar R, Salman M, Klapsa D, Hajarha Y, Asghar H, Bandyopadhyay A, Rambaut A, Martin J, Grassly N. 2020. Rapid and sensitive direct detection and identification of poliovirus from stool and environmental surveillance samples by use of nanopore sequencing. J Clin Microbiol 58:e00920-20. https://doi.org/10.1128/JCM.00920-20.

Editor Yi-Wei Tang, Cepheid

Copyright © 2020 Shaw et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Alexander G. Shaw, a.shaw@imperial.ac.uk.

Received 29 April 2020
Returned for modification 26 May 2020
Accepted 29 June 2020
Accepted manuscript posted online 1 July 2020
Published 24 August 2020

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

these outbreaks using the oral poliovirus vaccine relies on rapid detection of poliovirus in stool samples from children with acute flaccid paralysis (AFP) or from environmental surveillance (ES) samples. Detection of poliovirus is not confirmed until a genetic sequence from the VP1 region is obtained, which allows circulating virus to be distinguished from vaccine strains. Identification of serotype 2 VDPV is of particular importance due to the global withdrawal of serotype 2 oral poliovirus vaccine (OPV) in April 2016, which has led to a growing cohort of children without immunity to serotype 2 poliovirus (1).

The current gold standard method of poliovirus detection involves culturing the virus using susceptible cell lines, performing reverse transcription-PCR (RT-PCR) of RNA extracted from cell culture supernatant to determine serotype, distinguishing Sabin from non-Sabin polioviruses (intratypic differentiation [ITD]), and subsequent sequencing by traditional (Sanger) methods (2). However, this method is slow and leads to delays between sample collection and sequencing result, compromising the speed and effectiveness of any vaccination response. For example, during 2016 to 2018, the median time to obtain a sequence needed for case confirmation varied from 3 to 6 weeks in countries in Africa and Asia with VDPV or wild-type poliovirus outbreaks (3). These delays reflect not only the complex and time-consuming cell culture steps but also, in many countries, the need to ship samples internationally to laboratories with sequencing facilities. These methods have other limitations, including challenges related to biocontainment of poliovirus grown in culture, virus mutation during growth in cell culture, and competitive exclusion of viruses in samples containing poliovirus mixtures, leading to a lack of detection of minority or low fitness variants (e.g., ES samples containing vaccine [Sabin] and vaccine-derived polioviruses).

Direct molecular detection of polioviruses in stool or ES samples using RT-PCR could allow the rapid detection of poliovirus without the need for cell culture. However, quantitative RT-PCR capable of amplifying all polioviruses through the use of degenerate (pan-poliovirus) primers has historically reported low sensitivity compared with that of cell culture (4). The use of serotype- or strain-specific primers improves sensitivity, but multiplex assays with many primers would be required to identify all polioviruses, and the resulting products are too short for informative sequencing (5). Quantitative PCR in the absence of confirmatory sequencing also risks false-positive results because of cross-contamination or nonspecific amplification when using degenerate primers (6).

A sensitive, nested RT-PCR for the direct detection of polioviruses and other enteroviruses from clinical material has been developed and recommended by the World Health Organization (WHO) for enterovirus surveillance, but this method generates only partial VP1 sequences that cannot be used to confirm poliovirus infection (7). More recently, a nested RT-PCR based on amplification of the entire capsid region using pan-enterovirus primers, followed by amplification of the full VP1 using degenerate pan-poliovirus primers, was shown to be highly sensitive for the detection of poliovirus in stool (4). However, sequencing of the VP1 RT-PCR product using the traditional (Sanger) method is problematic for samples containing >1 poliovirus because of its inability to accurately sequence mixed samples. Similarly, nonspecific amplification of VP1 from other enteroviruses, which are highly prevalent in stool in low-income countries and practically ubiquitous in ES samples, limits the applicability of this approach.

Here, we present a method for the direct detection and identification of polioviruses from stool and ES samples using a nested PCR followed by nanopore (MinION) sequencing of the poliovirus VP1 region. This approach allows rapid sequencing library preparation following RNA extraction (<14 h) and subsequent, simultaneous sequencing of multiple samples using a barcoded DNA library. Sequences are demultiplexed and analyzed in real-time to provide immediate identification of poliovirus. We validated the method using artificial mixtures of enteroviruses and assessed sensitivity and specificity against culture and quantitative PCR using stool and ES samples from Pakistan and India.

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

MATERIALS AND METHODS

Creation of artificial virus mixtures. We used virus stocks of Echovirus 7 (E7), Enterovirus-A71 (A71), Enterovirus-D68 (D68), and serotype 1 Sabin poliovirus (PV1) containing 4.10×10^9, 4.39×10^8, 1.26×10^7, and 4.00×10^9 median cell culture infectious doses ($CCID_{50}$s)/ml, respectively. We prepared an even mixture by adding 20 μl of cell culture supernatant of E7, 200 μl of EV71, 1,000 μl of D68, and 20 μl of PV1. To create an uneven sample, we mixed 20 μl of each stock and made the volume 500 μl. Sample 3 was a pure culture of PV1, and sample 4 was a trivalent oral poliovirus vaccine standard (8). We also tested novel oral poliovirus vaccine candidates that have been engineered for greater stability than the Sabin vaccine (2 candidates for each serotype), including novel serotype 2 vaccine (nOPV2) that is scheduled for use in response to serotype 2 VDPV outbreaks under WHO Emergency Use Listing in 2020 (9, 10).

Stool and environmental samples. Stool samples were collected during acute flaccid paralysis (AFP) surveillance in Pakistan between October 2019 and February 2020. These samples are collected routinely by the Global Polio Eradication Initiative to detect transmission of wild-type poliovirus or VDPVs that rarely emerge as a result of reversion of the Sabin strains contained within the live-attenuated OPV. Samples may also contain Sabin strains themselves, originating from the recent use of OPV. We tested 155 samples found to be positive during this period by standard cell culture and ITD for wild-type 1 poliovirus (33); circulating serotype 2 VDPV (12); or Sabin poliovirus serotype, 1, 2, and/or 3 (111).

We additionally tested stored fecal aliquots from 1- to 4-year-old Indian infants participating in a clinical trial of inactivated poliovirus vaccine conducted in 2013 and who had received bivalent OPV containing serotype 1 and 3 poliovirus (11). In 2013 and 2014, these samples were tested using a singleplex reverse transcription-quantitative PCR (qRT-PCR) for Sabin poliovirus 1 and 3 (12) and a subset by cell culture following the standard WHO protocol (13). We retrieved all 217 samples from this subset that were positive for serotype 1 or 3 poliovirus by qPCR and 49 randomly selected samples that were negative by qPCR and culture for inclusion in the current study.

Sewage samples collected by the grab method in 2015 and 2017 in Pakistan were also used. One-liter raw sewage specimens were processed using two-phase separation to generate sewage concentrate, followed by the poliovirus isolation cell culture algorithm following WHO guidelines (14). Results from six culture flasks were amalgamated using aliquots drawn from a single sample (3 ml in total). Sewage concentrates were also used for direct PCR and sequencing using nanopore, Illumina, or Sanger methods.

Institutional ethics approval was not required because samples are anonymized and free of personally identifiable information. Samples from Pakistan were collected as part of the WHO/Pakistan Government's official polio surveillance process.

RNA extraction from stool and environmental sample concentrates. Viral RNA extracted from stool and environmental samples needed to yield long PCR products (~4,000 bp). Based on previous experience with different commercially available RNA extraction kits (see Appendix S1 in the supplemental material), we established that both Roche and Qiagen kits were suitable (see Appendix S1). A Roche kit (catalog number 11858882001) was used for RNA extractions from artificial mixtures and sewage samples (using 200 μl of sewage concentrate as input). For stool samples, 1 g of feces was added to 5 ml of phosphate-buffered saline (PBS). Chloroform extractions were then performed as described in reference 13, and 140 μl of the suspension was used for RNA extraction using the QIAamp viral RNA minikit in Pakistan and QIAamp 96 virus QIAcube high-throughput (HT) kit in India.

PCR amplification. PCR products for Sanger, Illumina, and nanopore (MinION) sequencing were obtained using a combination of protocols. RT-PCR was performed using the Superscript III one-step RT-PCR system (Invitrogen). cDNA was first synthesized with a single primer being added and the second primer added for the PCR step. RT-PCR products spanning the enterovirus capsid region or whole poliovirus genome were generated using primer combinations shown in Data Set S1 and Fig. S1 in the supplemental material. Poliovirus VP1 products were generated by nested PCR by first amplifying the capsid region by RT-PCR with pan-enterovirus (pan-EV) primers, followed by a second PCR using pan-poliovirus (pan-PV) primers Q8/Y7 or Q8/Y7R or serotype-specific VP1 primers (Data Set S1; Fig. S1). DreamTaq DNA polymerase (Thermo Scientific) or LongAmp *Taq* polymerase (New England BioLabs [NEB]) was used for the second PCR.

Sequencing library preparation for nanopore sequencing. As initial tests demonstrated strong amplification of VP1 using the pan-PV Q8/Y7 primers, we sought to reduce the time and cost for nanopore sequencing library preparation by modifying these primers with either the addition of a barcode adaptor (VP1-BCA) or barcode (VP1-barcode) to allow immediate multiplexing of samples (Fig. 1). PCR products underwent a standard preparation for nanopore sequencing with the LSK-109 kit with barcoding using the EXP-PBC096 kit where necessary. Products amplified with adaptor primers entered the library preparation protocol at the cleanup step after adaptor ligation. Products amplified with the barcoded primers entered the protocol at the cleanup step after the barcoding PCR. A complete protocol is documented in Appendix S2 in the supplemental material and at https://www.protocols.io/groups/poliovirus-sequencing-consortium. Sequencing was performed on Oxford Nanopore Technologies (ONT) MinION Mk1B sequencers using R9.4 flow cells. Sequencing run conditions and outputs are described in Table S1 in the supplemental material.

Pan-EV capsid and pan-PV whole-genome PCR products from the artificial mixtures and ES samples underwent the standard preparation for nanopore sequencing with the LSK-108 kit with barcoding using the EXP-PBC001 kit.

Sanger sequencing of poliovirus VP1 nested PCR products. Amplified poliovirus VP1 products from nested PCR were purified using the QIAquick PCR purification kit (Qiagen) and sequenced using an ABI Prism 3130 genetic analyzer (Applied Biosystems) using standard protocols.

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

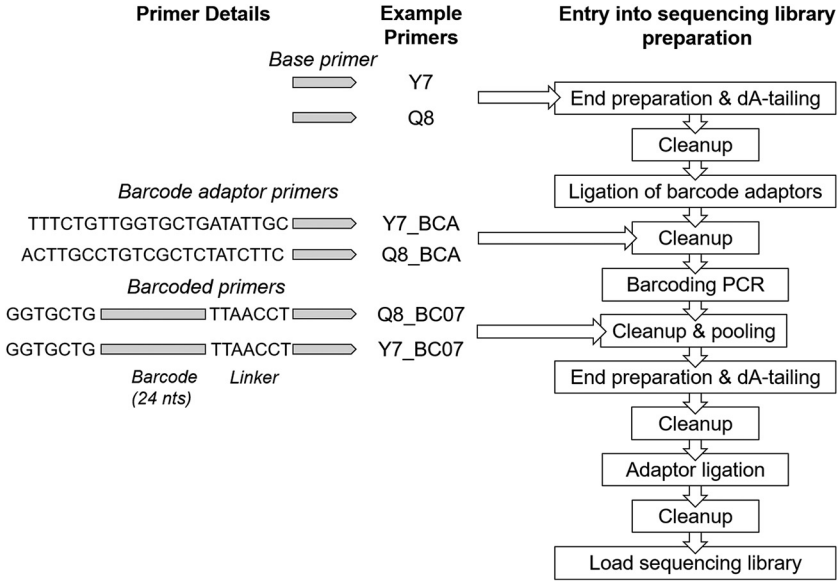


FIG 1 The formulation of primers for MinION sequencing and the entry of the various PCR products into the sequencing library preparation protocol. In addition to the basic primers, we have tested the addition of barcode adaptor (BCA) primers or barcodes and flanking regions (BC# primers, with # being the barcode number with reference to nanopore PCR ligation barcode sequences).

Sequence library preparation for Illumina sequencing. Pan-EV capsid and pan-PV VP1 nested PCR products from ES samples were additionally sequenced using Illumina next-generation sequencing. Sequencing libraries were prepared using Nextera XT reagents and sequenced on a MiSeq instrument using a 2 × 250-mer paired-end v2 flow cell and manufacturer's protocols (Illumina).

Bioinformatics. Nanopore data were drawn from 20 individual MinION sequencing runs, each of which is documented in Table S1. Fast5 files were base-called using Albacore v1.8 (pan-PV and pan-EV experiments) or Guppy v3.4 (remaining experiments) with graphics processing unit (GPU) acceleration on a Precision 7540 Dell laptop running Ubuntu 18 LTS with 32 GB RAM, a 16 core Intel Xeon central processing unit (CPU), and Nvidia RTX3000 GPU. Real-time analysis of VP1 reads was performed using a pipeline customized for the identification and mapping of PV nanopore data and then visualized using RAMPART (15). Once sufficient coverage of the VP1 region for each demultiplexed sample was confirmed using RAMPART, consensus sequences were generated from the base-called reads using a custom pipeline build on a Snakemake framework (16). In brief, reads are binned by sample and matched against a custom database comprising poliovirus and nonpolio enterovirus (NPEV) VP1 regions drawn from NCBI Nucleotide (https://www.ncbi.nlm.nih.gov/nuccore/) and the NIAID Virus Pathogen Database and Analysis Resource (ViPR) (17). To score poliovirus presence, a minimum of 50 reads was required within a bin, which is above the minimum threshold of sequencing reads required to give an accurate consensus, as established with simulated data (see Fig. S2 in the supplemental material). A read-correction ("polishing") module, which uses mafft v7 (18), minimap2 v2.17 (19), racon v1.4.7 (20), medaka v0.10.0 (21), and python code for quality control and curation, creates a consensus sequence for each virus population. The installable, documented pipeline can be found online at https://github.com/polio-nanopore/realtime-polio.

VDPV and other Sabin-related poliovirus sequences were not included in the VP1 reference database used by RAMPART because of the potential for spurious matches by nanopore-sequenced Sabin poliovirus, given the sequencing error rate before polishing and consensus sequence generation. Instead, sequencing reads with the closest match to the reference Sabin poliovirus VP1 sequences for each serotype were classified as "Sabin-related' polioviruses. Consensus sequences were subsequently generated for these Sabin-related reads to distinguish VDPV from Sabin polioviruses.

We additionally developed a bioinformatics pipeline to distinguish mixtures of polioviruses, including VDPV and Sabin polioviruses of the same serotype. Raw reads were clustered at 85% identity with vsearch v2.7.0 (22) and aligned to Sabin references using minimap2 v2.17. Variant calling was performed within clusters using medaka v0.10.0. Raw reads were then mapped to the variant sites to confirm linkage, and a maximum likelihood tree of the representative variants was generated using the phangorn package in R v3.4.3 (23, 24). Trees were cut at a height of 0.1, a consensus sequence for each subtree was generated, and any identical consensus sequences were merged. Further development of this pipeline is ongoing and documented online at https://github.com/polio-nanopore/realtime-polio. This same pipeline was used to detect potential cross-contamination of samples during RNA extraction or PCR that could result in a mixture of poliovirus reads for a single reaction.

Nanopore sequences from pan-EV capsid and pan-PV complete genome amplicons generated from artificial mixtures and ES samples were demultiplexed and trimmed using Porechop v0.2.3, and reads

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

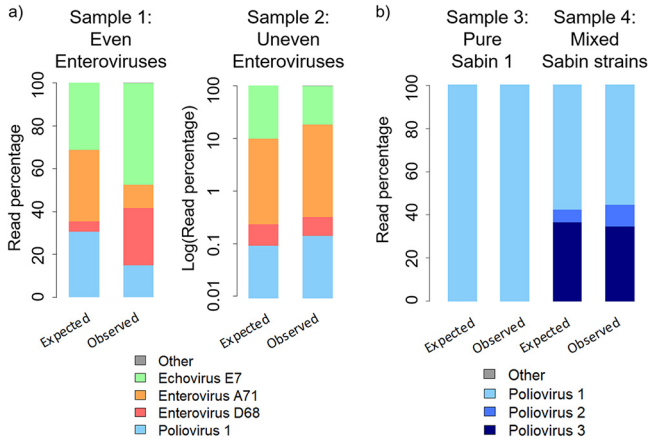


FIG 2 Identification of enterovirus and poliovirus mixtures by nanopore sequencing of the entire capsid or full genome. The observed abundance of sequencing reads is compared with that expected based on the known composition of artificial enteroviruses mixtures (a) and poliovirus mixtures (b). For each sample, the left bar shows the expected outcome according to the viral titers, and the right bar shows the observed results through comparison of the sequencing reads with the VP1 database. The results for sample 2 have been plotted on a log scale.

containing internal adaptors were discarded. Reads were screened to retain only those of the expected length, ensuring a minimum of 90% coverage (3,650 to 4,200 nucleotides for the capsid and 7,000 to 7,700 nucleotides for the nearly full-length genome). Reads were mapped with blastn (25) to the custom VP1 database.

Filtered reads from Illumina sequencing of pan-EV and nested VP1 products generated with pan-PV Q8/Y7 primers were mapped to the VP1 database, and PV consensus sequences were extracted. These sequence data were processed and analyzed using a custom workflow with the Geneious R10 software package (Biomatters) as described before (26).

Phylogenetic trees showing the relatedness of VP1 sequences generated using different sequencing approaches were constructed in MEGAX (27) using the neighbor-joining method from evolutionary distances estimated under the Tamura-Nei substitution model using the maximum composite likelihood method (28).

Data availability. Demultiplexed sequencing data are available at the European Nucleotide Archive under accession number PRJEB37193.

RESULTS

Identification of enteroviruses and polioviruses in artificial mixtures and ES samples using one-step RT-PCR and nanopore sequencing. Nanopore sequencing of the enterovirus capsid generated by RT-PCR using pan-EV primers identified all viruses included in the artificial mixtures, with read counts broadly reflecting the known composition of the mixtures (Figure 2a). Sabin poliovirus 1 was detected even when the virus constituted only 0.1% of the viral mixture. Discrepancy between the proportion of reads by serotype and the known mixture of the even sample may derive from the pooling of products from the two pan-EV primer sets, with overrepresentation of species B and D enteroviruses (E7 and EV-D68). Nanopore sequencing of the pan-PV full-genome product generated by RT-PCR of artificial poliovirus mixtures generated sequences for nearly the entire genome, with the abundance of reads closely matching that expected (Figure 2b).

We sequenced pan-EV RT-PCR entire capsid products from eight ES samples collected in Pakistan during 2015 and 2017 using nanopore and compared them with assembled reads for the same region generated through Illumina sequencing. The distribution of enterovirus serotypes was highly similar for the two sequencing methods (Fig. 3). Poliovirus was found in all eight samples by nanopore sequencing; yet, attributed reads formed only a small percentage of the total reads (0.002% to 6.65%; median, 0.03%). Whole-capsid poliovirus contigs were produced only for two of the eight ES samples from the Illumina reads, representing 0.08% and 5.47% of total reads.

Nested capsid/VP1 PCR barcoded primer development. To increase the specificity and sensitivity of our method, we performed a nested PCR, using pan-EV entire

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

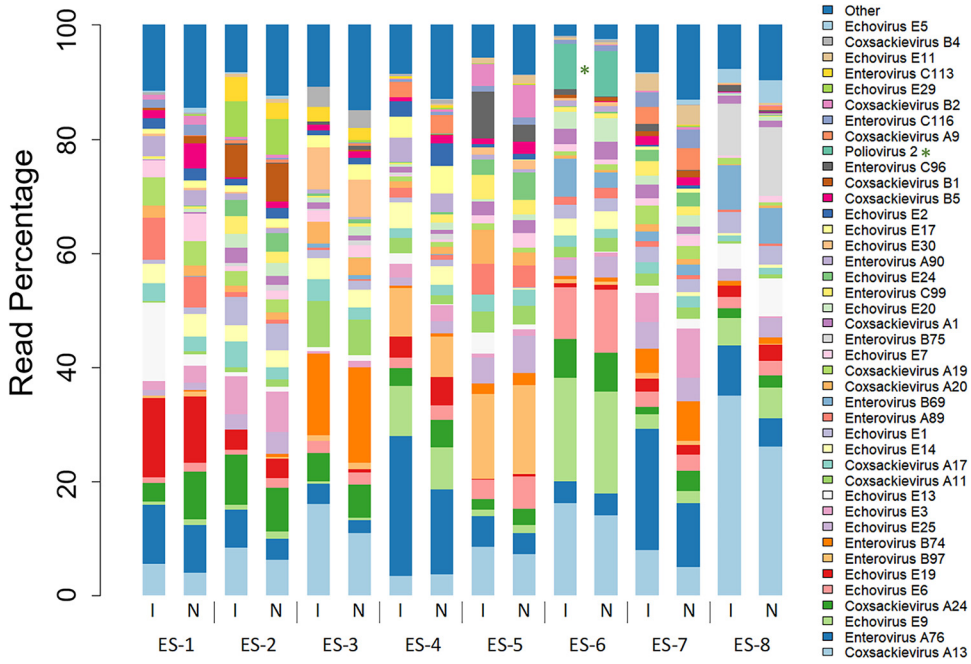


FIG 3 Enteroviruses present in eight ES samples from Pakistan analyzed by nanopore (N) and Illumina MiSeq (I) sequencing of pan-EV PCR products. Sequencing reads attributed to the 90% most abundant viruses are shown, with the remaining identified reads being grouped into "other." Detections of poliovirus serotypes have been starred.

capsid RT-PCR with Cre-R and 5′NCR primers in an RT-PCR, followed by pan-PV PCR using Q8/Y7 VP1 primers (Data Set S1). Modification of the VP1 primers with the addition of ONT barcodes or the barcode adaptor sequence gave a slight decrease in sensitivity but was equivalent to the loss of product that occurred during the additional library preparation steps when using the unadapted primers (see Appendix S3 in the supplemental material). Equivalence of the VP1-BCA and VP1-barcode primers was further confirmed by amplification of VP1 from the pan-EV products of 20 ES samples, which yielded DNA at >25 ng/μl suitable for sequencing for 18/20 (90%) and 20/20 (100%) of samples for each primer set, respectively (Fisher's exact test, $P = 0.487$).

Nested PCR and nanopore sequencing of stool samples. Nested PCR with barcoded Q8/Y7 primers and nanopore sequencing of the 155 stool samples collected during 2019 and 2020 in Pakistan detected 90.9% (30/33) of the wild-type 1 polioviruses that were identified by cell culture and Sanger sequencing, 92.5% (37/40) of serotype 2 Sabin and VDPVs, and 88.3% (106/120) of Sabin 1 and 3 polioviruses (Table 1). One wild-type poliovirus and one serotype 2 VDPV were detected in samples that were not positive for these viruses by culture and Sanger sequencing (giving specificities compared with culture of 99.2% and 99.3% for these virus types, respectively). Serotype 1, 2, and 3 Sabin polioviruses were detected in 14 samples that were not positive for these

TABLE 1 Detection and sequencing of poliovirus in 155 Pakistan stool samples using nanopore sequencing compared with cell culture[a]

	Nanopore sequencing results (no. of samples)								
Culture result	Sabin 1	Sabin 2	VDPV2	Sabin 3	Wild-type 1	Sabin 1 + 3	Sabin 1 + Sabin 3 + VDPV2	Wild-type 1 + Sabin 1	Negative
Sabin 1	**6**	0	0	0	0	4	0	0	1
Sabin 2	0	**28**	0	0	0	0	0	0	1
VDPV2	0	1	**9**	0	0	0	0	0	1
Sabin 3	0	0	0	**24**	1	5	1	0	2
Wild-type 1	0	0	0	0	**29**	1	0	1	2
Sabin 1+3	0	0	0	2	0	**32**	0	0	4

[a]Concordant results by nanopore sequencing and culture are highlighted in bold.

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

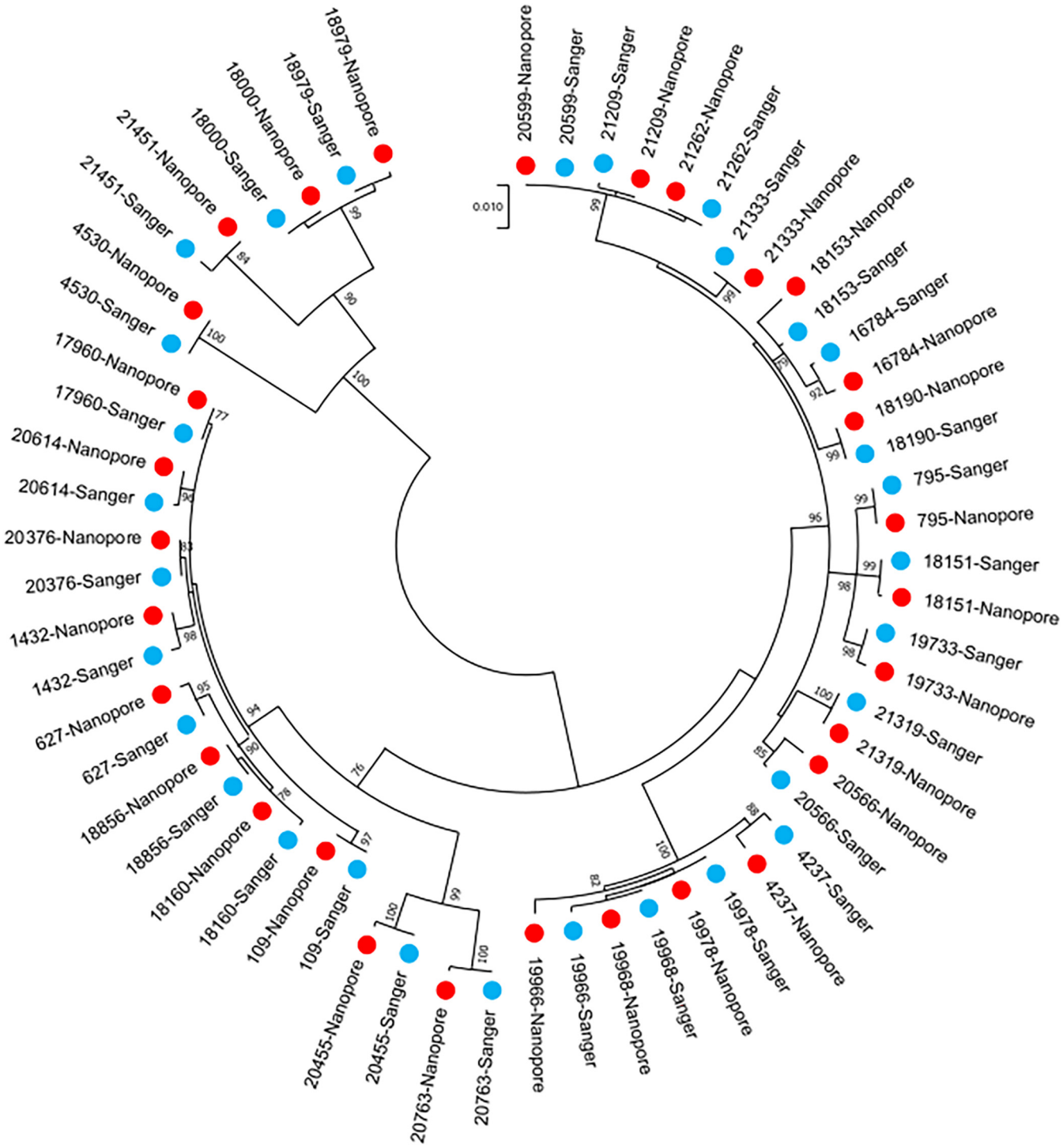


FIG 4 Relatedness of wild-type serotype 1 polioviruses identified in stool samples in Pakistan using alternative methods. VP1 sequences based on direct PCR and nanopore sequencing (red) are compared with Sanger-sequenced cell culture isolates (blue).

viruses by cell culture (specificity, 95.6%). The variant-calling pipeline identified contamination in 16 samples (14 in the last 2 sequencing runs), and these reads were removed. For the 30 samples where wild-type 1 poliovirus was identified by both culture and nanopore sequencing, the median identity for the VP1 region was 99.9% (range, 99.0% to 100%) (Fig. 4). For the samples where serotype 2 Sabin polioviruses or VDPVs were identified by both methods, the median identity for the VP1 region was 100% (range, 99.8% to 100%).

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

Stool samples containing serotype 1 or 3 Sabin poliovirus in India likely had suffered degradation of viral RNA as a result of storage for 6 years. Nonetheless, 69% (149/217) yielded >50 reads of the correct virus, and this value increased to 91% (109/120) among those with at least 100 copies of the target sequence at the time of original quantitative PCR testing (see Appendix S4 in the supplemental material).

Nested PCR and nanopore sequencing of novel oral poliovirus vaccine candidates. Nested PCR using Q8/Y7 primers generated a VP1 product for all eight novel oral poliovirus vaccine candidates, and consensus sequences generated from nanopore reads were 100% identical to Illumina sequences generated for the same samples.

Nested PCR and nanopore sequencing of ES samples. Nanopore sequencing using nested PCR with pan-PV VP1 Q8/Y7 primers resulted in a median of 87.1% of VP1 enterovirus reads mapping to poliovirus (range, 27.4% to 100.0%) and result in just 6.8% (range, 0.1% to 95.9%) when using the more degenerate Q8/Y7R primer. Consensus building, variant calling, and results from the more specific barcoded Q8/Y7 primers allowed the identification of poliovirus mixtures in these samples, including wild-type, vaccine (Sabin), and vaccine-derived polioviruses corresponding to different serotypes (Fig. 5). These results were almost identical to those obtained by Illumina sequencing of VP1 (Q8/Y7) nested PCR products and Sanger sequencing of VP1 nested PCR products generated with serotype-specific primers (Fig. 4). Nanopore sequencing detected eight of nine wild-type 1 polioviruses identified by Sanger or Illumina sequencing of direct VP1 products and showed very similar detection of Sabin-related polioviruses (Pearson's phi coefficients were 0.89, 1.0, and 1.0 for serotypes 1, 2, and 3, respectively, for nanopore versus Sanger or Illumina results [which were identical]). In addition, it was possible to distinguish serotype 2 VDPV and Sabin polioviruses from nanopore sequencing data using our bioinformatics pipeline, which is not possible for Sanger or Illumina reads (the latter because they are short and it is difficult to confirm mutations are shared by a single virus). All three direct detection and sequencing methods gave somewhat different results from those generated by Sanger sequencing of cell culture isolates, which is based on combined results of up to 6 positive flasks per sample, with direct detection identifying fewer polioviruses (e.g., 8 or 9 versus 15 wild-type polioviruses and a lower prevalence of Sabin polioviruses).

Wild-type poliovirus nanopore consensus sequences were nearly identical to those obtained from the same sample aliquot by serotype-specific VP1 PCR and Sanger or Illumina sequencing (median identity to Sanger sequences of 100%; range, 99.9% to 100.0%; $n = 8$). These VP1 sequences were also very closely related to those obtained from corresponding cell culture flasks using Sanger with a median identity of 100.0% (range, 98.7% to 100.0%) (Fig. 6; see Table S2 in the supplemental material). Similar results were obtained for Sabin 2 and VDPV2 sequences compared with Sanger or Illumina sequencing results (median identity to Sanger sequences of 100.0%; range, 99.9% to 100.0%; $n = 26$).

Real-time visualization of demultiplexed sequencing results. The sequencing data generated through the nested VP1 protocol were analyzed in real-time using RAMPART (Fig. 7). This analysis included live demultiplexing and mapping of each read to Sabin or wild-type polioviruses or NPEVs, allowing an immediate decision about when sufficient reads had been generated for each sample to permit ending of the sequencing run, preserving the MinION flow cell chemistry for further sequencing runs. At the end of each run, the custom bioinformatics pipeline created a report for each sample that describes the number of reads mapping to each reference and the corresponding consensus sequence.

DISCUSSION

Here, we present a culture-independent method for the rapid detection and sequencing of poliovirus from stool and ES samples. Using barcoded primers in a nested PCR, we were able to proceed from sample to sequence in 3 days, with real-time identification of polioviruses in each sample during nanopore (MinION) sequencing using RAMPART (15). The overall sensitivity of this method to detect wild-type or

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

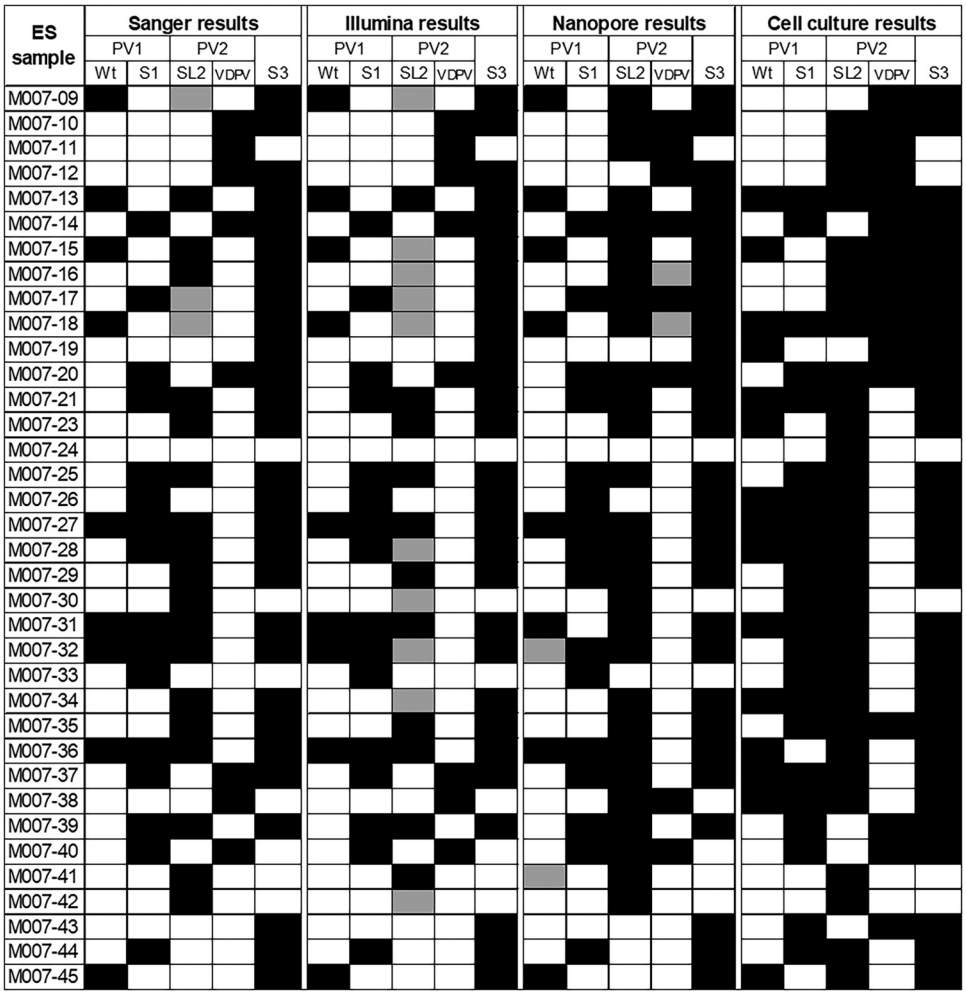


FIG 5 Polioviruses detected in Pakistan ES samples by cell culture or direct PCR followed by sequencing using three different methods. Direct detection of poliovirus in 36 environmental samples by nested PCR and nanopore sequencing compared with Illumina sequencing of the nested PCR product, serotype-specific PCR, and Sanger sequencing ("Sanger") or Sanger sequencing of cell culture isolates (up to 6 flasks per sample). A black cell indicates detection of the virus. Sabin-related detections have been clarified as either Sabin (S) or VDPVs where possible. For Sanger and Illumina results, gray indicates the detection of Sabin poliovirus and the possibility of a VDPV based on multiple peaks and SNP analysis, respectively. For nanopore results, gray indicates detection of the virus but at a low read threshold.

vaccine-derived polioviruses in Pakistan was 90% compared with culture, ITD, and Sanger sequencing methods. Consensus sequences were near identical to those generated by Sanger sequencing of cell culture isolates for the same samples, indicating that accurate VP1 sequences can be generated by nanopore sequencing. Further optimization that we plan for this method has the potential to further increase sensitivity, such as testing duplicate extractions from a single stool sample and use of an RNA extraction control (e.g., bacteriophage Qβ).

Immediate identification of the poliovirus sequence allows downstream phylogenetic analysis and interpretation of epidemiological significance (29, 30). Faster generation of these results will allow earlier vaccination responses and containment of virus spread, which can have a substantially greater impact on poliomyelitis cases during an outbreak (31). Immediate availability of the poliovirus sequence allows potential sample contamination to be assessed and prevented—a challenge faced by direct detection methods relying on quantitative PCR, which requires subsequent additional PCR and sequencing reactions.

The high sensitivity of nested entire capsid/VP1 PCR for poliovirus in stool samples

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

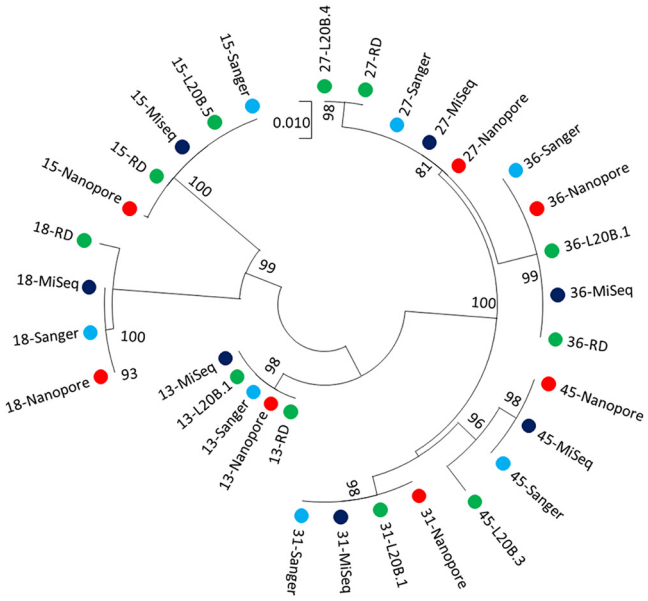


FIG 6 Relatedness of wild-type poliovirus 1 identified in ES samples in Pakistan using alternative methods. Consensus VP1 sequences derived by each sequencing platform (Sanger, MiSeq, and nanopore) are shown in addition to the VP1 sequences of corresponding culture isolates (L20B and RD, labeled according to cell line used in isolation). No wild-type poliovirus 1 sequence was isolated by culture for sample M007-23; hence, an isolate from a sample taken 2 months prior from the same ES site was used in the analysis (starred sequence).

has been previously reported (100% sensitivity for 84 stool extracts) (4). Combining this approach with barcoded VP1 primers and nanopore sequencing allows sequencing of poliovirus mixtures and avoids the problem of nonspecific amplification of NPEVs by the Q8/Y7 primers reported in reference 4 since any such amplicons are correctly classified as NPEVs by nanopore (unlike Sanger) sequencing.

While previous studies have successfully used Nanopore sequencing to target specific enteroviruses (32–34), the use of pan-EV and pan-PV primers allows mixtures of polioviruses and of polioviruses with other enteroviruses in stool and ES samples to be identified. In artificial mixtures, the distribution of sequencing reads was found to match the known composition of the mixture. Currently, identification of poliovirus mixtures requires laborious testing using multiple culture flasks, plaque isolation, or antibody blocking of specific serotypes (35). Nanopore sequencing of ES samples gave results that were consistent with current culture-based methods but detected fewer polioviruses overall. This may reflect the much smaller aliquot of sewage concentrate ($\sim$200 μl) undergoing RNA extraction compared with that inoculated in cell culture flasks (6 flasks with 0.5 ml each). Indeed, testing the same RNA extract with RT-PCR using serotype-specific primers gave almost identical results and sequences to the nested PCR and nanopore sequencing, indicating that viruses within that aliquot were accurately and sensitively identified. Furthermore, the current cell culture algorithm typically finds discordant results among flasks (36), in part due to a heterogenous distribution of viruses within samples. This suggests that testing multiple aliquots or achieving a greater sewage concentration would increase the sensitivity of direct molecular detection methods used to test ES samples.

Nanopore sequencing using MinION has a number of advantages over alternative next-generation sequencing approaches, which can also be applied to both culture supernatants and PCR products. These advantages include lower initial investment and running costs and greater portability and ease of use. This has facilitated rapid training and implementation in laboratories using this technology for the first time. In addition, generation of full-length reads allows shared mutations to be identified along the VP1 region, which is especially important for ES samples that may contain mixtures of polioviruses and

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

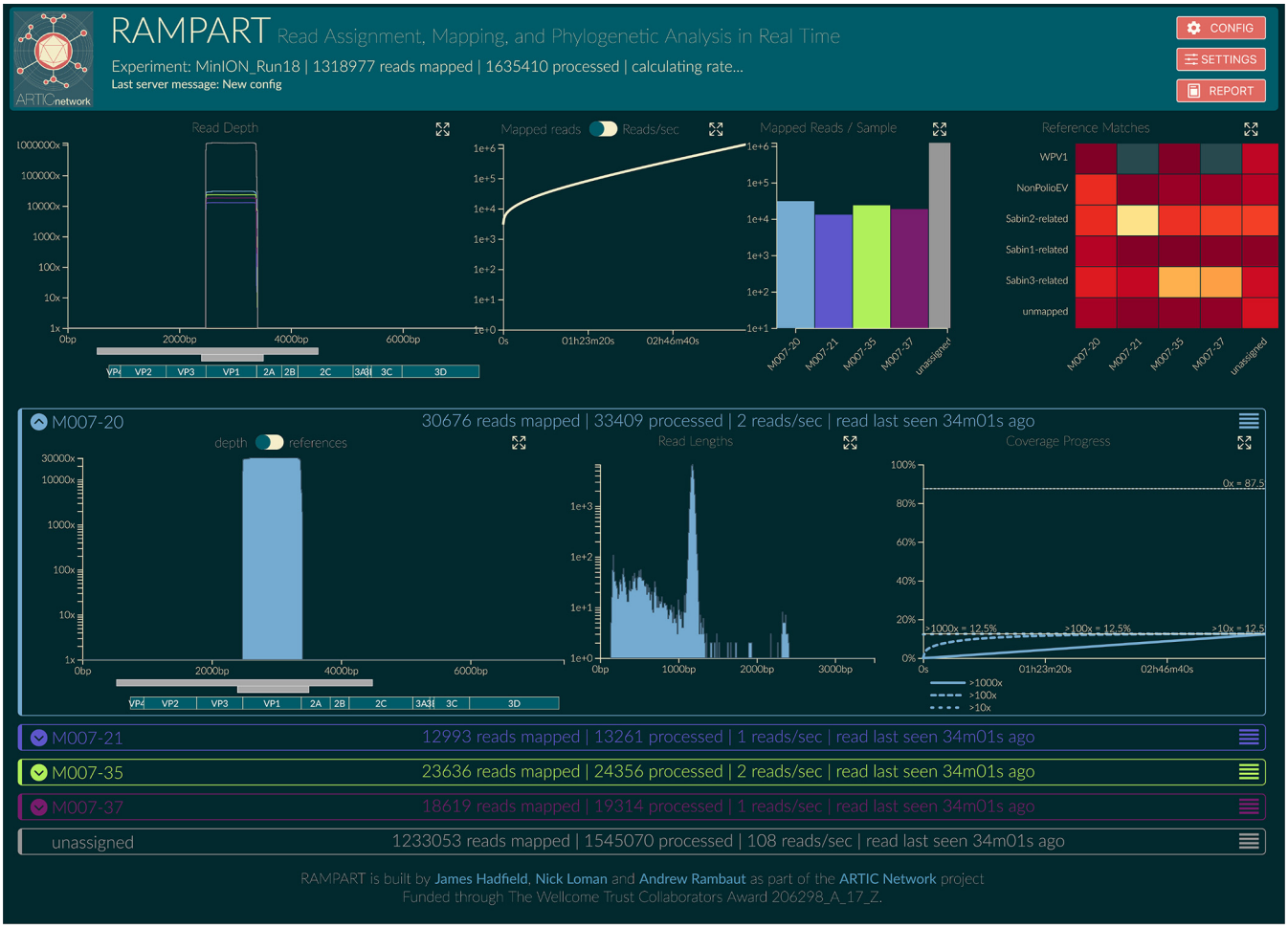


FIG 7 Illustration of RAMPART display for data from four ES samples. Top (left to right): (i) summary of coverage across the genome for all samples, amplicons, and genes of poliovirus; (ii) number of mapped reads plotted against time; (iii) number of mapped reads for each sample, with unassigned reads shown in gray; and (iv) heatmap showing the proportion of reads for each sample that map to either wild-type poliovirus, Sabin-related poliovirus or NPEVs or do not map ("unmapped"). Bottom: detailed information is be displayed for each sample. The panel header indicates the sample that has been processed, the rate of data processing, and the number of reads that have been mapped. Middle (left to right): (i) coverage across the genome and sample composition are displayed; (ii) read length distribution is shown, with a peak at 1,200 bases, the expected amplicon size; (iii) coverage of genome over time, which plateaus at 12.5% given that the amplicon spans only this proportion of the genome.

where short reads often cannot distinguish multiple Sabin-related viruses. A more detailed analysis of isolates using the genome-length pan-PV primer set would even allow the identification of viral recombination between polioviruses and human Enterovirus C serotypes, as the primer binding sites fall within shared nonstructural regions of the genome.

The nested PCR and nanopore sequencing protocol can be scaled up to allow for the processing of multiple samples simultaneously. While generating data for this article, groups of up to 96 samples were run together on a single MinION flow cell. This level of multiplexing still produced >15,000 reads per sample after just 4 hours of sequencing (using an R9.4.1 flow cell). Currently, the MinION flow cell chemistry can accommodate 48 hours of sequencing, and therefore, a single flow cell could in principle sequence >500 samples. In practice, we typically sequenced ~200 samples with a tail-off due to pore degradation.

The nested PCR and nanopore sequencing protocol has a number of limitations. First, individual sequence reads on the MinION are error prone (currently ~5% single-molecule error rate for R9 MinION flowcell). An accurate consensus can be generated from just 50 reads or more, but for samples containing mixtures of closely related polioviruses (i.e., Sabin and VDPV of the same serotype), an initial clustering step is required. Our current bioinformatics pipeline was able to achieve an accurate identification of Sabin, vaccine-derived,

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

and wild-type polioviruses in ES samples, but further validation is required. Sequencing errors may also affect the accurate determination of sample barcodes. However, by enforcing stringent matching criteria (double barcodes with >80% identity), we observe a negligible misclassification of reads (0.002%). Second, a concern with repeated usage of flow cells is carry-over from one run to the next. This can however be prevented using a DNase wash out and alternating sets of barcoded primers that would allow any carry-over to be identified (37). Third, our protocol generates only VP1 sequences, consistent with current methods. However, full-genome sequences may be of interest, and amplicons for the nearly full-length genome can be generated directly from the sample and sequenced on a MinION device using pan-poliovirus primers (38). This PCR is less sensitive than the nested pan-EV-VP1 protocol and in some cases may require culture-based or other methods of enrichment (e.g., see reference 39). Finally, careful handling of samples and laboratory processing are required to ensure optimal sensitivity and to avoid sample cross-contamination. Further work is required to develop laboratory quality control and quality assurance procedures, including a provision of standardized positive controls if this method is to be adopted more widely by the Global Polio Laboratory Network. We demonstrate however that the immediate availability of sequencing data allows identification of contaminants and their removal from data sets, thus minimizing false positives compared to other direct detection methods. Further optimization of the method may also be possible, for example through testing alternative VP1 primers that are specific to particular polioviruses (e.g., for improved detection of wild-type 1 or circulating serotype 2 VDPV in ES samples).

In conclusion, we have developed a nested PCR with time- and cost-saving barcoded primers that allows multiplex sample sequencing on a portable MinION sequencer. Using this method, only quantification, sample pooling, and the addition of adaptors is required prior to sequencing. This method was shown to be sensitive and to generate accurate poliovirus consensus sequences from stool samples. Complex poliovirus mixtures were identified in environmental samples, although further development of bioinformatic methods for the consistent identification of variants within ES samples is required. This method has the potential to replace the current cell culture-based testing for poliovirus and has a number of significant advantages over alternative direct molecular detection methods.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 1.2 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

We thank the members of the ARTIC network for helpful discussions in the development of this work. We thank James Hadfield for support and development of the RAMPART software package.

This work was funded by grants from the Bill and Melinda Gates Foundation (OPP1171890 and OPP1207299). The Artic Network is funded by the Wellcome Trust through Collaborative Award 206298/Z/17/Z. N.G. acknowledges support from the Medical Research Council (UK) Centre for Global Infectious Disease Analysis and a Royal Society Wolfson laboratory refurbishment scheme grant. A.R. acknowledges support from the European Research Council (grant agreement number 725422-ReservoirDOCS).

REFERENCES

1. Jorgensen D, Pons-Salort M, Shaw AG, Grassly NC. 2020. The role of genetic sequencing and analysis in the polio eradication program. Virus Evol 2020:veaa040. https://doi.org/10.1093/ve/veaa040.
2. World Health Organization. 2017. WHO polio laboratory manual/supplement 1: an alternative test algorithm for poliovirus isolation and characterization. World Health Organization, Geneva, Switzerland. http://polioeradication.org/wp-content/uploads/2017/05/NewAlgorithmForPoliovirusIsolationSupplement1.pdf.
3. Global Polio Eradication Initiative. 2020. Polio Information System (POLIS). World Health Organization, Geneva, Switzerland. https://extranet.who.int/polis/About.
4. Arita M, Kilpatrick DR, Nakamura T, Burns CC, Bukbuk D, Oderinde SB, Oberste MS, Kew OM, Pallansch MA, Shimizu H. 2015. Development of an efficient entire-capsid-coding-region amplification method for direct detection of poliovirus from stool extracts. J Clin Microbiol 53:73–78. https://doi.org/10.1128/JCM.02384-14.

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London

5. Laassri M, DiPiazza A, Bidzhieva B, Zagorodnyaya T, Chumakov K. 2013. Quantitative one-step RT-PCR assay for rapid and sensitive identification and titration of polioviruses in clinical specimens. J Virol Methods 189:7–14. https://doi.org/10.1016/j.jviromet.2012.12.015.
6. Oberste AS, Pallansch MA. 2005. Enterovirus molecular detection and typing. Rev Med Microbiol 16:163–171. https://doi.org/10.1097/01.revmedmi.0000184741.90926.35.
7. Nix WA, Oberste MS, Pallansch MA. 2006. Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. J Clin Microbiol 44:2698–2704. https://doi.org/10.1128/JCM.00542-06.
8. Cooper G, Christian P, Heath A. 2004. Report of a collaborative study to assess the suitability of a replacement International Standard for Oral Polio Vaccine. World Health Organization, Geneva, Switzerland. https://www.who.int/biologicals/publications/trs/areas/vaccines/polio/1992CollabOPV.pdf?ua=1.
9. Macadam AJ, Ferguson G, Stone DM, Meredith J, Knowlson S, Auda G, Almond JW, Minor PD. 2006. Rational design of genetically stable, live-attenuated poliovirus vaccines of all three serotypes: relevance to poliomyelitis eradication. J Virol 80:8653–8663. https://doi.org/10.1128/JVI.00370-06.
10. Van Damme P, De Coster I, Bandyopadhyay AS, Revets H, Withanage K, De Smedt P, Suykens L, Oberste MS, Weldon WC, Costa-Clemens SA, Clemens R, Modlin J, Weiner AJ, Macadam AJ, Andino R, Kew OM, Konopka-Anstadt JL, Burns CC, Konz J, Wahid R, Gast C. 2019. The safety and immunogenicity of two novel live attenuated monovalent (serotype 2) oral poliovirus vaccines in healthy adults: a double-blind, single-centre phase 1 study. Lancet 394:148–158. https://doi.org/10.1016/S0140-6736(19)31279-6.
11. John J, Giri S, Karthikeyan AS, Iturriza-Gomara M, Muliyil J, Abraham A, Grassly NC, Kang G. 2014. Effect of a single inactivated poliovirus vaccine dose on intestinal immunity against poliovirus in children previously given oral vaccine: an open-label, randomised controlled trial. Lancet 384:1505–1512. https://doi.org/10.1016/S0140-6736(14)60934-X.
12. Kilpatrick DR, Yang C-F, Ching K, Vincent A, Iber J, Campagnoli R, Mandelbaum M, De L, Yang S-J, Nix A, Kew OM. 2009. Rapid group-, serotype-, and vaccine strain-specific identification of poliovirus isolates by real-time reverse transcription-PCR using degenerate primers and probes containing deoxyinosine residues. J Clin Microbiol 47:1939–1941. https://doi.org/10.1128/JCM.00702-09.
13. World Health Organization. 2004. Polio laboratory manual, 4th ed. World Health Organization, Geneva, Switzerland. https://apps.who.int/iris/bitstream/handle/10665/68762/WHO_IVB_04.10.pdf.
14. World Health Organization. 2003. Guidelines for environmental surveillance of poliovirus circulation. 2003. World Health Organization, Geneva, Switzerland. http://polioeradication.org/wp-content/uploads/2016/07/WHO_V-B_03.03_eng.pdf.
15. ARTIC Network. 2020. Read assignment, mapping and phylogenetic analysis in real time (RAMPART). https://artic.network/rampart.
16. Koster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28:2520–2522. https://doi.org/10.1093/bioinformatics/bts480.
17. Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S, Gu Z, Zhou L, Larson CN, Dietrich J, Klem EB, Scheuermann RH. 2012. ViPR: an open bioinformatics database and analysis resource for virology research. Nucleic Acids Res 40:D593–D598. https://doi.org/10.1093/nar/gkr859.
18. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
19. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32:2103–2110. https://doi.org/10.1093/bioinformatics/btw152.
20. Vaser R, Sovic I, Nagarajan N, Sikic M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.
21. Oxford Nanopore Technologies. 2019. Medaka. Version 0.7.1. Oxford Nanopore Technologies, Oxford, UK. https://github.com/nanoporetech/medaka.
22. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584. https://doi.org/10.7717/peerj.2584.
23. Schliep KP. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593. https://doi.org/10.1093/bioinformatics/btq706.
24. R Core Team. 2018. R: a language and environment for statistical computing. https://www.R-project.org.
25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
26. Majumdar M, Klapsa D, Wilton T, Akello J, Anscombe C, Allen D, Mee ET, Minor PD, Martin J. 2018. Isolation of vaccine-like poliovirus strains in sewage samples from the United Kingdom. J Infect Dis 217:1222–1230. https://doi.org/10.1093/infdis/jix667.
27. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096.
28. Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023.
29. Alam MM, Sharif S, Shaukat S, Angez M, Khurshid A, Rehman L, Zaidi S. 2016. Genomic surveillance elucidates persistent wild poliovirus transmission during 2013–2015 in major reservoir areas of Pakistan. Clin Infect Dis 62:190–198. https://doi.org/10.1093/cid/civ831.
30. Akhtar R, Mahmood N, Alam MM, Naeem M, Zaidi SSZ, Sharif S, Khattak Z, Arshad Y, Khurshid A, Mujtaba G, Rehman L, Angez M, Shaukat S, Mushtaq N, Umair M, Ikram A, Salman M. 2019. Genetic epidemiology reveals 3 chronic reservoir areas with recurrent population mobility challenging poliovirus eradication in Pakistan. Clin Infect Dis 2019:ciz1037. https://doi.org/10.1093/cid/ciz1037.
31. Blake IM, Martin R, Goel A, Khetsuriani N, Everts J, Wolff C, Wassilak S, Aylward RB, Grassly NC. 2014. The role of older children and adults in wild poliovirus transmission. Proc Natl Acad Sci U S A 111:10604–10609. https://doi.org/10.1073/pnas.1323688111.
32. Rames E, Macdonald J. 2018. Evaluation of MinION nanopore sequencing for rapid enterovirus genotyping. Virus Res 252:8–12. https://doi.org/10.1016/j.virusres.2018.05.010.
33. Beato MS, Marcacci M, Schiavon E, Bertocchi L, Di Domenico M, Peserico A, Mion M, Zaccaria G, Cavicchio L, Mangone I, Soranzo E, Patavino C, Camma C, Lorusso A. 2018. Identification and genetic characterization of bovine enterovirus by combination of two next generation sequencing platforms. J Virol Methods 260:21–25. https://doi.org/10.1016/j.jviromet.2018.07.002.
34. Wang J, Ke YH, Zhang Y, Huang KQ, Wang L, Shen XX, Dong XP, Xu WB, Ma XJ. 2017. Rapid and accurate sequencing of enterovirus genomes using MinION nanopore sequencer. Biomed Environ Sci 30:718–726. https://doi.org/10.3967/bes2017.097.
35. Kilpatrick DR, Iber JC, Chen Q, Ching K, Yang S-J, De L, Mandelbaum MD, Emery B, Campagnoli R, Burns CC, Kew O. 2011. Poliovirus serotype-specific VP1 sequencing primers. J Virol Methods 174:128–130. https://doi.org/10.1016/j.jviromet.2011.03.020.
36. Hovi T, Blomqvist S, Nasr E, Burns CC, Sarjakoski T, Ahmed N, Savolainen C, Roivainen M, Stenvik M, Laine P, Barakat I, Wahdan MH, Kamel FA, Asghar H, Pallansch MA, Kew OM, Gary HE, Jr., deGourville EM, El Bassioni L. 2005. Environmental surveillance of wild poliovirus circulation in Egypt–balancing between detection sensitivity and workload. J Virol Methods 126:127–134. https://doi.org/10.1016/j.jviromet.2005.02.002.
37. Oxford Nanopore Technologies. 2018. Nuclease Flush. Oxford Nanopore Technologies, Oxford, UK.
38. Stern A, Te Yeh M, Zinger T, Smith M, Wright C, Ling G, Nielsen R, MacAdam A, Andino R. 2017. The evolutionary pathway to virulence of an RNA Virus. Cell 169:35–46.e19. https://doi.org/10.1016/j.cell.2017.03.013.
39. Arita M. 2013. Development of poliovirus extraction method from stool extracts by using magnetic nanoparticles sensitized with soluble poliovirus receptor. J Clin Microbiol 51:2717–2720. https://doi.org/10.1128/JCM.00499-13.

Downloaded from http://jcm.asm.org/ on September 1, 2020 at Imperial College London