



AFRL-RH-WP-TR-2019-0018

**EVALUATING AN UPPER RESPIRATORY DISEASE
PANEL ON THE PORTABLE MinION SEQUENCER**

**Wanda J. Lyon
Zachary K. Smith
Brian Geier**

Human Signatures Branch

**James Baldwin
Clarise R. Starr
USAFSAM**

**APRIL 2019
Interim Report**

Distribution A: Approved for public release.

See additional restrictions described on inside pages

**AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
AIRMAN SYSTEMS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RH-WP-TR-2019-0018 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signature//

WANDA LYON, Work Unit Manager
Human Signatures Branch
Airman Systems Directorate
711th Human Performance Wing
Air Force Research Laboratory

//signature//

RICHARD D. SIMPSON, DR-IV, DAF
Human-Centered ISR Division
Airman Systems Directorate
711th Human Performance Wing
Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE				<i>Form Approved OMB No. 0704-0188</i>	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YY) 24-04-19		2. REPORT TYPE Interim Technical Report		3. DATES COVERED (From - To) 01-01-18 - 04-24-19	
4. TITLE AND SUBTITLE Evaluating an Upper Respiratory Disease Panel on the Portable MinION Sequencer				5a. CONTRACT NUMBER In-House	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 00000F	
6. AUTHOR(S) Wanda J. Lyon Ph.D., Zachary K. Smith, Brian Geier, James Baldwin Ph.D, and Clarise R. Starr Ph.D.				5d. PROJECT NUMBER 0000	
				5e. TASK NUMBER 00	
				5f. WORK UNIT NUMBER HOSN	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) 711 th Human Performance Wing, USAFSAM, Wright Patterson Air Force Base OH				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory 711 th Human Performance Wing Airman Systems Directorate Human-Centered ISR Division Human Signatures Branch Wright-Patterson AFB, OH 45433				10. SPONSORING/MONITORING AGENCY ACRONYM(S) 711 HPW/RHXB	
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S) AFRL-RH-WP-TR-2019-0018	
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution A; Approved for public release					
13. SUPPLEMENTARY NOTES 88ABW-2018-3416, 3 July 2018 (bioRxiv journal - doi: https://doi.org/10.1101/436600)					
14. ABSTRACT The MinION nanopore sequencer was released to community testers for evaluation using a variety of sequencing applications. The MinION was used to evaluate upper respiratory disease infections and was found to have tremendous potential for field use. In this study, we tested the ability of the MinION to accurately identify and differentiate clinical bacterial and viral samples via targeted sequencing and whole genome sequencing. The current nanopore technology has limitations with respect to error rate but has steadily improved with development of new flow cells and library kits. Upper respiratory disease organisms were successfully sequenced, identified and differentiated down to the strain level with 87-98% alignment against our reference genome database. This study offers evidence of the utility of sequencing for identify and differentiate both viral and bacterial species present within clinical samples.					
15. SUBJECT TERMS MinION nanopore sequencing, Portable device, next generation sequencing, microbiome					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT: SAR	18. NUMBER OF PAGES 31	19a. NAME OF RESPONSIBLE PERSON (Monitor) Wanda J. Lyon 19b. TELEPHONE NUMBER (Include Area Code) N/A
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			

TABLE OF CONTENTS

Abstract 1

Introduction..... 1

Materials and Methods..... 4

 Nucleic acid extraction..... 4

 Targeted RT-PCR and PCR 5

 MinION library preparation and sequencing 5

 Rapid 1D libraries 7

 Rapid 1 D libraries..... 7

 Illumina MiSeq library preparation and sequencing 7

 Sequence data analysis 8

Results..... 9

DISCUSSION..... 11

References..... 22

Abstract

The MinION nanopore sequencer was released to community testers for evaluation using a variety of sequencing applications. The MinION was used to evaluate upper respiratory disease infections and was found to have tremendous potential for field use. In this study, we tested the ability of the MinION to accurately identify and differentiate clinical bacterial and viral samples via targeted sequencing and whole genome sequencing. The current nanopore technology has limitations with respect to error rate but has steadily improved with development of new flow cells and library kits. Upper respiratory disease organisms were successfully sequenced, identified and differentiated down to the strain level with 87-98% alignment against our reference genome database. This study offers evidence of the utility of sequencing for identify and differentiate both viral and bacterial species present within clinical samples.

Introduction

Acute respiratory infections (ARI) have been a significant source of disease and nonbattle injury among military forces for centuries. Although mortality from ARIs is low in military populations, the impact of these infections on military readiness in terms of lost work-hours and convalescence is high. Furthermore, novel ARIs have been a periodic source of great morbidity and mortality, such as during the 1918 A/H1N1 (“Spanish”) influenza epidemic, the 2002-2003 severe acute respiratory syndrome (SARS) epidemic in Asia and North America, and the 2009-2010 swine-origin A/H1N1 global pandemic. Periodic outbreaks of adenovirus and similar infections have plagued military training centers in recent years, control of which has been hampered until very recently by the unavailability of vaccines [13, 21]. Supporting the notion

that viral infections play a major role in training and readiness, other studies have demonstrated the rates of ARI in trainees averaged 80%, with hospitalization rates at 20% [15, 16, 36]. Despite aggressive vaccination programs, ARI has been noted to be the cause of 70% of U.S. Air Force pilot groundings during Operation Desert Storm [27]. Although influenza is usually the pathogen that captures the attention of the world, other respiratory pathogens have recently garnered the attention of the military. Adenoviruses cause a wide array of symptoms, from conjunctivitis, gastroenteritis, and mild respiratory infections to life-threatening pneumonias in young adults with blood stream and brain infections in immunocompromised hosts. The Middle East respiratory syndrome-coronavirus (MERS-CoV) outbreak in the Middle East [40] along with the H7N9 influenza outbreak in China [42] has focused our attention on new and emerging pathogens that may be introduced into immunologically naïve populations or recombined into novel forms through human-human or human-animal transmission.

Next-generation sequencing (NGS) technologies are now capable of providing a whole genome sequence or metagenomics for a wide range of organisms [11, 12, 14, 17, 18, 39, 45]. NGS is being applied with increasing frequency in clinical microbiology laboratories to detect and characterize pathogens, thereby replacing the currently used Sanger sequencing methods [3, 41, 42]. Sanger sequencing has been a reliable and robust method that has served clinical laboratories well for over several decades; however, it is labor intensive, slow, and not easily adapted for processing large genomes or large samples sets. The characterization of upper respiratory disease (URD) viruses would certainly benefit from the use of whole genome sequencing (WGS) and targeted sequencing, providing us with tools for surveillance of highly diverse viral genomes [26, 31, 33]. Such data can assist in vaccine development, detection of

anti-viral drug resistance, or the identification of new emerging re-assorted viral genomes. These advances may help to mitigate the morbidity and mortality of influenza pandemics or seasonal viral epidemics.

Oxford Nanopore's MinION may therefore provide us with new opportunities to track infectious disease, for example, rapid sequencing of viral genomes would allow us to track early phase influenza pandemics, or determine what coronavirus virus genotypes are present during outbreaks. In addition, sequencing will allow us an opportunity to gain insight into viral genetic drift (single nucleotide polymorphism), emergence of new strains such as SARS-like coronavirus or MERS-CoV, and transmission patterns. To date, a number of pathogenic viral studies have been done with the MinION [5, 10, 23, 26, 28, 37, 46]. Metagenomic NGS is particularly attractive for surveillance of febrile illness because the approach can broadly detect viruses, bacteria, and parasites in clinical samples in the field [7, 13, and 40]. Although currently limited by a sample-to-answer turnaround time of >24 hours for benchtop sequencers, we report in this study that unbiased pathogen detection using both targeted and whole genomic sequencing can be realized with actionable results for clinical diagnostics [4, 6, 23, 29, 31, 33, 37, 43, 44 and 46] and public health [4, 10, 29] within 10 h.

NGS platforms such as the Illumina MiSeq and Life Technologies Ion Torrent have been extensively used to provide WGS data for viruses, potentially within 2-3 days of the receipt of a sample [40]. Therefore, there is clearly a place for a portable NGS instrument that can be deployed into the field for surveillance using available field ready kits, single use flow cells (Flongle) and sequencing reagents that are comparable to cost of benchtop sequencer. In

addition, analysis of sequencing data is accomplished in real time using an internet-connected laptop. The MinION has been successfully used for WGS and targeted sequencing [8, 22, 42], WGS is well suited for the investigation of URD species that high levels of genetic diversity and are difficult to access using targeted sequencing methods[23, 37 and 46]. The purpose of this study was to examine the efficacy of using the MinION to sequence URD organisms present in URD clinical samples.

Materials and Methods

Nucleic acid extraction

Influenza A virus H1N1 (VR1736), parainfluenza virus 1-3 (VR94, VR92, and VR93), coronavirus 229E and OC43 isolates (VR-740 and VR1558), *Mycoplasma pneumoniae* (ATCC 29342), *Haemophilus influenzae* virus (51907), and human adenovirus (VR-1603) were obtained from American Type Culture Collection (ATCC). The nucleic acid isolated from these organisms were used as control nucleic acid (NA) for library preparations and was added as internal control during library preparations. RNA/DNA were extracted from 200 μ L viral supernatants or from clinical nasal washes using the Maxwell total viral nucleic acid kits (Promega, Madison, WI) and eluted with 50 μ L of RT-PCR molecular grade water (Thermo Fisher Scientific, Waltham, MA). De-identified positive nasal washes were obtained from the clinical laboratory at Wright Patterson Air Force Base. Clinical isolates used in this study were influenza A virus, influenza B virus, parainfluenza viruses 1-3, adenovirus, and coronavirus OC043.

Targeted RT-PCR and PCR

Targeted RNA PCR was done per using primers listed in Table 1 using the SuperScript III OneStep RT-PCR system (Thermo Fisher Scientific) per the manufacturer's instructions. Briefly, 10 ng of DNA-free RNA or RNA-free DNA was used for each 50- μ L reaction. All PCR reactions were accomplished with 1.0 μ l of each primer at 10 μ M. RT-PCR thermocycling parameters were as follows: 30 min at 50°C, 2 min at 95°C, and then 35 cycles of 30 s at 95°C, 30 s at 52°C, and 75 s at 72°C, followed with a final extension at 68°C for 5 min. PCR of NA from bacteria or DNA viruses was accomplished using AccuStart II PCR Supermix (Quanta Biosciences, Beverly, MA) as recommended by the manufacturer. Cycling conditions were as follows: 3 minutes at 95°C denaturing, 25 cycles of 30 seconds at 94°C, 30 seconds at 55°C, and 90 seconds at 72°C, followed with a final extension at 72°C for 3 minutes. DNA or cDNA PCR reactions were purified using 1.5X Agencourt Ampure XP beads as described by the manufacturer (Beckman Coulter, Indianapolis, IN) and were quantified using a Qubit 4.0 fluorometer) using either DNA high sensitivity kits. (Thermo Fisher Scientific).

MinION library preparation and sequencing

This work was completed as part of the ONT MinION early access program. During this time the flow cells, flow cell chemistries and library kits changed numerous times during the two-year study. The ONT MinION Sequencing kit that was used to prepare libraries as shown in Table 3. DNA or cDNA libraries were sequenced with various flow cells throughout the study including versions R7, R7.3, and Spot-ON 9.4. The subsequent increase in flow cell number indicates improvements in their flow cells, with Spot-On R9.4 flow cell being the latest improvement. In addition, the ONT MinION library reagents were changed as kit improvements were made during this study (Kit: SQK-MAP0005 for the R7.03 flow cell and Kits: SQK-LSK- 209, SQK-

RAD 004 (Rapid) for R9.4 or Spot-ON R9.4 flow cells). Barcoding of the DNA fragments was accomplished by using the barcoding kit (EXP-PBC001). Libraries kit used for each of the sequencing libraries are shown in Table 2 and 3, as well as, the amount of starting material and flow cells type. The Oxford Nanopore Technologies (ONT; Oxford Nanopore Technologies, Oxford, United Kingdom) MinION Sequencing kit were used to prepare DNA and RNA libraries as described by the manufacturer. Briefly, DNA or cDNA was sheared at 6,000 rpm in a g-TUBE (Covaris). The fragmented DNA was repaired using NEBNext FFPE repair mix (New England BioLabs [NEB]), Ipswich, MA). After 15 min of incubation at 20°C with FFPE repair buffer and FFPE repair mix, the repaired DNA was purified using AMPure XP beads in a 1:1 ratio, washed with 80% ethanol, and eluted with nuclease-free water. The DNA fragments were end repaired and deoxyadenosine (dA) tailed using the NEBNext end repair/dA tailing module (NEB). DNA was mixed with DNA CS (positive-control strand), Ultra II end prep reaction buffer and enzyme mix and incubated for 5 min at 20°C and 5 min at 65°C in a thermocycler. DNA was purified using AMPure XP beads in a 1:1 ratio, washed with 80% ethanol, and eluted with nuclease-free water.

End-prepped DNA was incubated with hairpin adapter and NEB blunt/TA ligase master mix (NEB) for 15 min at room temperature, followed by addition of hairpin tether and 10 min of incubation at room temperature. The libraries were incubated with bead binding buffer-washed My One streptavidin C1 beads (Invitrogen, Carlsbad, CA) in a 1:1 ratio for 5 min at room temperature, washed with bead binding buffer, and eluted with elution buffer prior to loading.

buffer, and eluted with elution buffer prior to loading and were quantified using a Qubit 4.0 fluorometer and DNA high sensitivity kits (Life Technologies). The libraries were transferred to a low-binding 1.5-mL Eppendorf tubes and placed on ice until loaded onto the flow cell.

Rapid 1D libraries

Rapid 1 D libraries (ONT R9 version; SQK-RAD004) **were prepared** using whole genomic NA from clinical isolates. Whole genome products were sheared with a Covaris g-tube (Woburn, MA) with fragmentation sizing to approximately 5 kb. Two hundred nanograms of DNA was tagged by incubation with fragmentation mix (FRM) for 1 min at 30°C followed by 1 min at 75°C in a thermocycler. Tagmented DNA was then adapter ligated by incubation with rapid adapter mix (RAD) and NEB blunt/TA ligase master mix for 5 min at room temperature and directly loaded for sequencing as described by the manufacturer.

ONT sequencing

A 48-hour sequencing protocols were initiated using the MinION control software, MinKNOW versions 1.1-1.3.24, for flow cells 7.3 and MinKNOW versions 1.4.6.-2.4.5 for R9 flow cells, respectively. Read event data were base-called by the Metrichor agent (version 0.46.1.9, which was later updated to EPI2ME) using WIMP application (What In My Pot; 1.2.2 rev 1.5) with appropriate 2D and 1D workflow scripts [20].

Illumina MiSeq library preparation and sequencing

Targeted RT-PCR amplification was completed as described earlier using the same NA extracted for the MinION sequencing experiments. The primers used for targeted RT-PCR are listed in

Table 1. Libraries for Illumina MiSeq sequencing were prepared using the Nextera XT library prep kit (Illumina, San Diego, CA). For each library, 1 ng of DNA was added to 10 ul of RSB buffer and 5 ul of ATM. This reaction mixture was incubated at 55°C for 5 min to tagment the DNA and then neutralized with 5 ul of NT buffer. Tagmented DNA fragments were indexed with unique barcode primers and PCR amplified in a 50 ul reaction mixture as follows: 72°C for 3 min, 95°C for 30 s, and 12 cycles of 95°C for 10 s, 55°C for 30 s, and 72°C for 30 s, followed by 72°C for 5 min. PCR cleanup was performed by addition of 30 ul of AMPure XP beads (Beckman Coulter, Brea, CA), followed by 7 min of incubation and then two washings with 80% ethanol and elution into 35 ul of RSB. Nextera libraries for all isolates were multiplexed and sequenced in a paired-end 300-cycle mode on an Illumina MiSeq instrument using 600-cycle reagent kit v. 3 (Illumina). Library preparations were quantified using a Qubit 4.0 fluorometer and DNA high sensitivity kit (Thermo Fisher Scientific).

Sequence data analysis

Read data were extracted from the native HDF5 format (Fast 5) and converted into FASTA using poretools [28] and sequence reads were assessed for quality and trimmed with Nanofilt [9]. Quality reads with a q-score lower than 4 were omitted. Kraken and Bracken bioinformatics packages were used for alignment against a database comprised of fungal, viral and bacterial genomes [2, 30, 44, 47]. All genomes, viral and bacterial were downloaded from NCBI Ref Seq (May, 2015) and were used to build the Kraken database (version 0.10.4) with k=31. Bracken (Bayesian Reestimation of Abundance after classification with Kraken) estimates species abundances in metagenomic samples. Bracken [30, 35] was used to reassign alignments to the species level and visualization was accomplished with Krona [35] using the highest number of hits (lowest e-

value) or Metrichor's WIMP application. Illumina reads were assessed for quality using FastQC and trimmed with Trimmomatic 0.36 using the following settings: trim reads: Leading: 3, Trailing: 3, and Sliding Window: 4:15. Our in-house pipeline aligns the reads using the Kraken and Bracken and the data is visualized with KRONA [35]. Burrows-Wheeler Alignment MEM (BWA; aligned to reference and Human GR2) and SAM Tools was used to extract statistics and error rate per 100 bases [24, 25]. Sequence reads were filtered for human sequences and removed prior to visualization. Tables 2 and 3 list the analysis tools used for each data set. Read percentage identity is defined as $100 * \text{matches} / (\text{matches} + \text{deletions} + \text{insertions} + \text{mismatches})$. Fraction of reads aligned are defined as $(\text{alignment length} + \text{insertions} - \text{deletions}) / (\text{alignment length} + \text{unaligned length} - \text{deletions} + \text{insertions})$.

Results

Targeted amplicon sequencing using the MinION

A panel of URD viruses (influenza A virus, influenza B virus, parainfluenza virus 1, parainfluenza virus 2, parainfluenza virus 3, human adenovirus D, coronavirus OC043 and bacterial strains (*M. pneumoniae* and *H. influenzae*) were used in this study. The same NA from each sample was used to generate libraries for both Oxford Nanopore MinION and Illumina MiSeq. All of the organisms listed were successfully sequenced on the MinION and identified correctly using the WIMP application, which uses a cloud-based reference databank. Figure 1 shows a representative example of the data analysis for an influenza A H1N1 isolate listed in Table 1. In addition, BWA MEM and SAM Tools was used to verify the identification of each isolates listed in Tables 2 [24, 25]. BWA MEM alignment of the reads were in agreement with Metrichor's identification for each of the nasal wash samples listed in Table 2.

Fast 5 files generated by MinKNOW were converted to FASTQ using Poretools [28] then the FASTQ files were FASTQC, trimmed, aligned with Kraken and visualized with KRONA using our in-house pipeline [30, 35]. For example, Table 2 shows that sequence reads generated from Influenza A H1N1 cDNA resulted in 114 high quality reads and 104 of these reads aligned (87.3%) to Influenza A H1N1 (Table 2). MinION sequence data for viruses Influenza B, Parainfluenza 1 (PAIV1) and Coronavirus OC043 had a percent identity of 98.7, 18.6 and 91.2, respectively (Table 2). Interestingly, there were small differences between the sequencing data quality in terms of alignment error. Influenza A H1N1, Influenza B, Parainfluenza 1 (PAIV1) and Coronavirus OC043 had alignment error of 1.15E-01, 1.54E-01, 2.74E-02 and 1.37E-01, respectively (Table 2). The Nanopore kits used early in this study resulted a low number of sequence reads, which is probably due improper NA fragment to adapter ratios during ligation step in the library preparations.

Multiplexing Targeted PCR Libraries

NA extracted from clinical nasal washes and the isolated RNA (200 ng/ul) or DNA (200 ng/ul) were used in library preparation. Direct cDNA sequencing kit was used to prepare the RNA for sequencing. The individual libraries were barcoded and mixed in equal molar concentrations prior to being loaded onto a R9.4 flow cell. The percent-alignment of reads to reference genomes ranged from 92-96% (Table 3). The number of reads were somewhat different among the tested organisms and probably due to an error in quantification of the input library prior to pooling. Future barcoding pooling protocols should include a Tapestation or Bioanalyzer (Agilent) quantification in terms of fragment length for more accurate calculation of the nM concentration prior to pooling. Metrichor's cloud base-calling WIMP application correctly identified each

organisms in the library pool (Figure 3F). All of the URD organisms listed in Table 3 were reevaluated using our Kraken/Bracken pipeline to verify Metrichor's identification and visualized with KRONA (Figure 3A-E).

1D Rapid libraries

H. influenzae and Coronavirus 229E libraries were prepared using ONT's rapid 1D library preparation kit. These data show that rapid 1D kit is a suitable replacement kit for previously used ONT's 2D kits. Sequencing runs were ran 17.5 hours and the total base yield for an isolate of *H. influenzae* was 44.9 million with a quality score centered on 6 (Figure 4). *H. influenzae* was correctly identified using the cloud based Metrichor's WIMP application (Figure 4A) and the read alignment was verified with our Kraken/Bracken pipeline and visualized with KRONA (Figure 4B). Coronavirus 229E was also successfully identified using Metrichor's WIMP application (S1). The coverage was calculated for both the *H. influenzae* and Coronavirus 229E with the formula $C=LN/G$, where C = coverage, L = average sequence read length, average Length N = number of reads, and G = the genome size. *H. influenzae* and Coronavirus 229E genome coverage was approximately 186X and 170X, respectively.

DISCUSSION

Sequencing of URD viruses and bacteria was accomplished on the MinION nanopore sequencer using both whole genome and amplicon sequencing. Having the ability to sequence amplicons and whole genomes in a mobile, small-footprint platform is attractive when collecting and analyzing samples in the field, and these qualities are also desirable as sequencing methods move toward environmental and surveillance applications. These methods, data, and results represent a practical and novel application for utilizing MinION nanopore sequencing technology. The data

in this study were generated with both R7.3 and 9.4 chemistry flow cells, which accurately identified a panel of URD organisms. As the MinION technology improves, sequencing will generally become cheaper, faster, and more accurate as demonstrated in this study where there was improvements in the error rate as new flow cells and kits were released resulting in improved alignment percentages >92% (Tables 2 and 3).

Base-calling analysis for the MinION nanopore sequencer is an area of active development, as well as flow cell configurations, protocols for kits, and associated kit chemistries. The best quality sequence data are termed 2D reads, while 1D reads are from the template strand. Data quality has improved and was demonstrated in this study with each improvement in kit chemistry and flow cell type. This study shows that the platform can be applied for rapid microbial detection with 3-hour run time, providing sufficient data to generate a reliable identification. As sequencing yield, quality, and turnaround times continue to improve, it is anticipated that nanopore sequencing will challenge mainstays such as PCR for identification of unknown samples.

Comparison of the nanopore and short-read sequencing data (Illumina MiSeq) showed that there was agreement in major taxonomic units identified. Therefore, the methodologies described in this study demonstrate that both whole genome sequencing and targeted sequencing can be used for rapid, accurate, and efficient detection of microbial and viral diversity in clinical samples. The use of WGS is also beneficial when studying viral genetic rearrangements and genetic drift in viruses [19, 23]. In addition, these data also support the utilization of targeted sequencing over deep sequencing for field applications where surveillance is needed.

Although long reads generated from the nanopore sequencer were found to have relatively higher error rates when compared with benchtop sequencing reads. However, MinION sequencing has been shown to have significant advantages, such as portability, low cost, and real-time detection. Nevertheless, sequencing of low titer samples, such as nasal washes, with amplicon sequencing are more desirable over culturing and will provide a rapid method for identification of organisms in clinical samples. This study demonstrated that a low number of long reads are sufficient for accurate identification of small genomes and new flow cells and chemistries (Spot-On 9.4) showed improvements in the number of errors in sequencing data (Table 2 and 3).

The rapid 1D kit is a single stranded library composed of template strands was used to evaluate rapid processing metagenomic samples. The use of rapid 1D library demonstrates it is functionally a game changer when rapid identification of pathogens is needed for tracking acute infections or during surveillance. This approach is especially beneficial when identification is difficult using conventional cell culture techniques or in some cases when targeted methods are not available. Establishing a system for rapid microorganism identification via metagenomic sequencing seems pertinent, especially for field applications.

CONCLUSIONS

Having the ability to sequence in a mobile, small-footprint platform is attractive when collecting and analyzing samples in the field, and these qualities are also desirable as sequencing methodologies moves towards a smaller footprint sequencer (SmidgION). Portable MinION and SmidgION sequencing technologies would have significant advantages for its direct use in field

applications, which allows for rapid sample-to-answer capabilities and negating the need to ship samples back to the laboratory.

Availability of supporting data

The raw sequence data sets are available at:

https://figshare.com/articles/combined_fastq_zip/7172600

Table 1. Primers used for targeted RT-PCR

Organism	Gene (URD Panel Name or Source)	Forward Primer 5' to 3'	Reverse Primer 5' to 3'	Instrument Used
Adenovirus				
Influenza A H1N1	CDC ^a Hemagglutinin (HA)	CAACCAAAAATGAAAG	CCGTCCAGTAGTARTTRATTCT	MinION & MiSeq
Influenza A	polymerase segment 2 (PB2)	CCTCTGGGATTCCTTTCGTCAGTC	GAGAAGTTCGGYGGIAGICTTTG	MiSeq
Influenza A H3N2	CDC (HA)	TGTA AACGACGGCCAGTAAAGC AGGGGATAATTCTA	CAATAKATGCTTATTC	MinION & MiSeq
Influenza B	CDC (HA)	GCAGACCATTTTCTAA	ACCACACTTTTTGAGG	MinION & MiSeq
Parainfluenza strain 1	URD panel G13 23/25 (Baldwin et al. [1])	ACIGCTATIGCTTGATTGTCTCCTT G	CCAAGAGGGGGTATAGAIGG	MiSeq & Minion
Parainfluenza strain 2	URD panel G2 23/25 (Baldwin et al. [1])	TCTTGTTGTA ACTGCAATCGCCTG	CAAGGGAGGTATTGAAGGCCTAT	MinION
Parainfluenza strain 3	URD panel G13 23/25 (Baldwin et al. [1])	TCTTGTTGTA ACTGCAATCGCCTG	CCAAGAGGGGGTATAGAIGG	MinION
Coronavirus 229E	Spike protein (Silva et al. [39])	TACCCTCCGACTTTGCATTC	TACTGCACCCACAAAAGCAC	MinION
Coronavirus HKU14	Spike protein (Silva et al. [39])	TGCCTATTGCACCAGGAGTC	TCAGCCATGTCAGGTGTTAC	MinION & MiSeq
<i>Mycobacterium pneumoniae</i>	16 S Quast et al. [41]	AGAGTTTGATCMTGGC	GACGGGCGGTGWGTRCA	MinION

^a <https://www.CDC.gov>

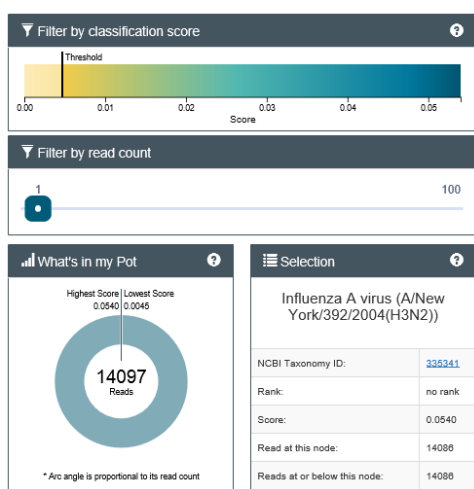
Table 2. Targeted Sequencing of URD samples using the MiSeq and MinION

Descriptor	Influenza A H1N1	Influenza A H1N1	Influenza B	Influenza B	Parainfluenza 1	Parainfluenza 1	Coronavirus OC043	Coronavirus OC043
Flow cell version	MinION R7.3	MiSeq 600 V3	MinION R7.3	MiSeq 600 V3	MinION R7.3	MiSeq 600 V3	MinION R7.3	MiSeq 600V3
Base calling version/ alignment	Metrichor & BWA/SAM Tools Kraken/ Bracken	Illumina Basespace MiSeq/ BWA/SAM Tools	Metrichor & BWA/SAM Tools Kraken / Bracken	Illumina Basespace MiSeq/ BWA/SAM Tools	Metrichor BWA/SAM Tools Kraken/ Bracken	Illumina Basespace MiSeq BWA/SAM Tools	Metrichor BWA/SAM Tools Kraken/ Bracken	Illumina Basespace MiSeq/ BWA/SAM Tool
Library protocol version and amount of starting material	SQK-MAP005 1000 ng	Nextera XT 1.0 ng	SQK-MAP005 1000 ng	Nextera XT 1.0 ng	SQK-LSK109 1000 ng	Nextera XT 1.0 ng	SQK-LSK109 1000 ng	Nextera XT 1.0 ng
Target gene	HA	HA	HA	HA	HA	HA	Spike protein	Spike protein
PCR amplicon base pair length	763	763	886	866	200	150	400	400
Total number of reads passing QC reads	114	202701	7339	230049	404	396725	114	612294
Total number of aligned reads	104	176744	7244	212748	75	365403	104	597202
100 base error rate	1.3E-01	3.63E-02	1.54E-01	7.89E-02	1.01E-01	2.74E-02	1.37E-01	3.47E-3
Percent alignment	87.3	97.5	98.7	93.2	18.6	92.1	91.2	97.5

Table 3. Multiplexed Targeted RT-PCR or PCR Amplification

Descriptor	Human Adenovirus Clinical Isolate	Parainfluenza Strain 1 Clinical Isolate	Parainfluenza Strain 2 Clinical Isolate	Parainfluenza Strain 3 Clinical Isolate	<i>M. pneumoniae</i> ATCC 29342
Flow cell version	R9.4	R9.4	R9.4	R9.4	R9.4
Base calling version	Metrichor 2D version 1.107 RNN	Metrichor 2D version 1.107 RNN	Metrichor 2D version 1.107 RNN	Metrichor 2D version 1.107 RNN	Metrichor 2D version 1.107 RNN
Library protocol version/starting material	SQK- LSK209 2D DNA 1000 ng	SQK-LSK209 2D cDNA 1000 ng	SQK-LSK209 2D cDNA 1000 ng	SQK-LSK209 2D cDNA 1000 ng	SQKLSK209 2D DNA 1000 ng
Average base pair 2D read length	2423	2483	2994	2639	6351
Total number of 2D pass reads	1763	1400	1785	656	2281
Per base error rate	1.09E-01	1.01E-01	9.53E-02	9.08E-02	1.29E-.01
Percent aligned identity	95.9	91.7	94.8	92.5	93.3

A



B

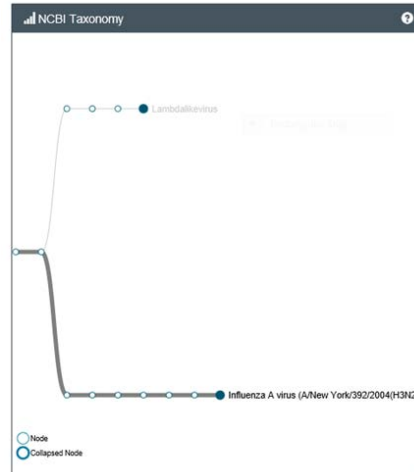
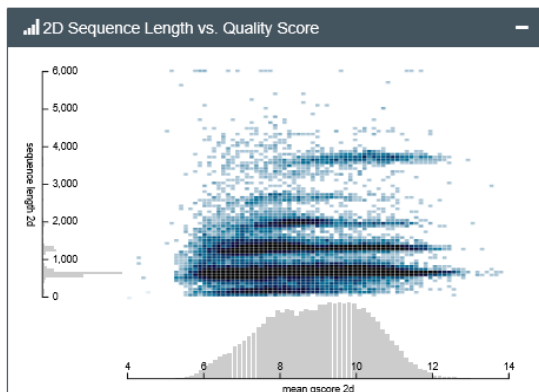


Figure 1. Target amplification of influenza A virus using the H2N2 primers listed in Table 1. Metrichor WIMP application was used for identification of the Influenza A H3N2 virus.

A



B

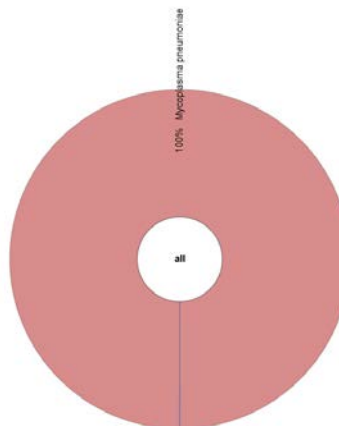
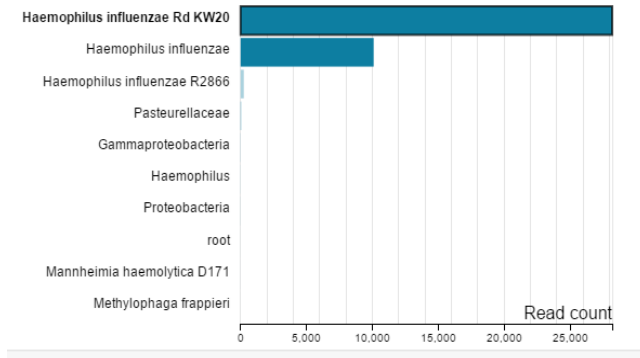


Figure 2. *M. pneumoniae* 2D library identification using Metrichor and Bracken. (A) 2D sequence (template + complement strands) as a function of quality score. (B) Bracken was used to align the reads and top hit abundance reads were visualized with KRONA



Figure 3. Multiplexing libraries of URD organisms. (A-E) Barcoded URD reads were aligned with Bracken and visualized with KRONA. (F) Number of reads for each barcode after analysis with the Metrichor 2D base calling application.

A



B

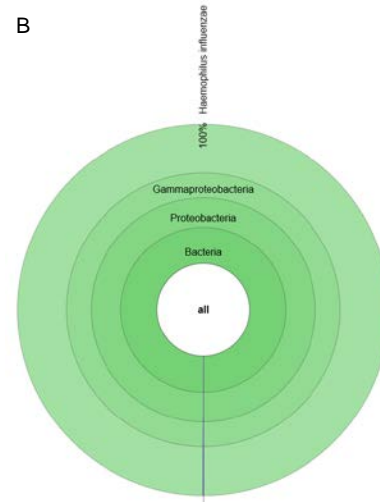
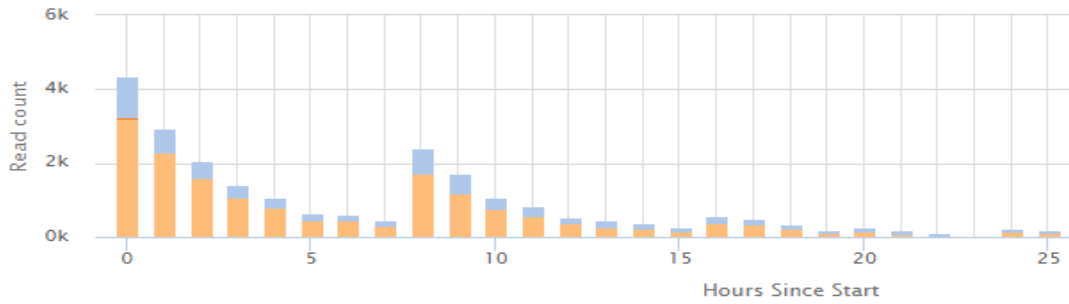
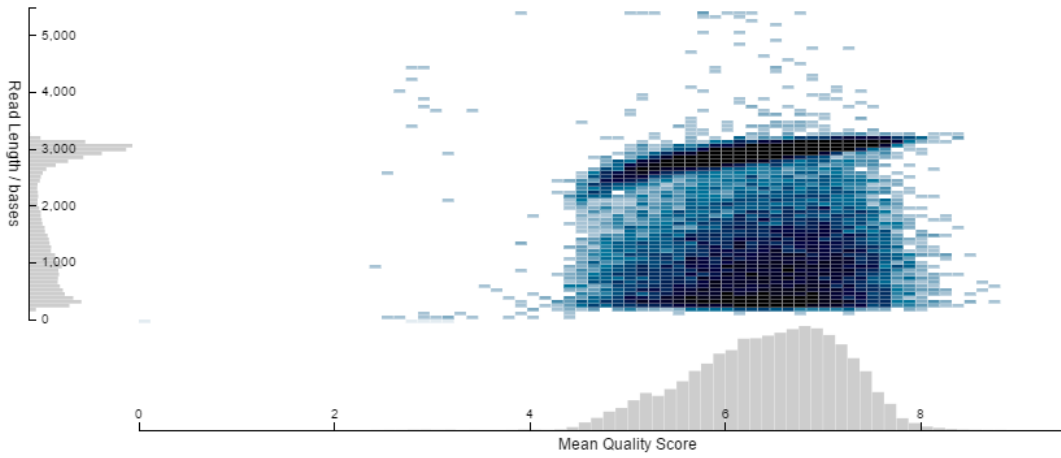


Figure 4. *H. influenzae* was prepared for sequencing using a rapid 1D method. (A) *H. influenzae* read alignment using Metrichor WIMP application. (B) *H. influenzae* identification were verified with Bracken and visualized with KRONA.

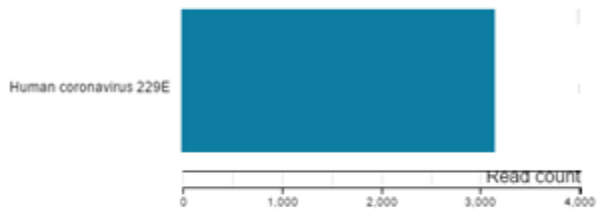
A



B



C



S1. Coronavirus 229E was prepared for sequencing using the rapid 1D library method. (A) Coronavirus 229E read count as a function of sequence time. (B) Read length as a function of quality score. (C) Coronavirus 229E identification using Metrichor WIMP application.

Acknowledgements

This work was supported by the Defense Health Program and the Chief Scientist Office at Wright Patterson Air Force Base, 711th Human Performance Wing, Ohio. The authors acknowledge the contribution of Mr. Craig Strapple for performing the bioinformatics analyses.

References

1. Baldwin JC, et al. 2010. Upper respiratory virus serotype panel for the pyrosequencer. Wright-Patterson AFB (OH): U.S. Air Force School of Aerospace Medicine; Technical Report AFRL-SA-WP-TR-2012-0002.
2. Bao Y, et al. The influenza virus resource at National Center for Biotechnology Information. *J. Virol* 2008; 82(2):596-601.
3. Bertelli, and Greub G. 2013. Rapid bacterial genome sequencing: methods and applications in clinical microbiology. *Clin Microbiol Infect.* 19(9):803-813.
4. Briese T, et al. 2009. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog.* 5(5):e1000455.
5. Brown B, Watson M, Minot S, Rivera M, and Franklin R. 2017. MinION nanopore sequencing of environmental metagenome; a synthetic approach. *GiaScience*, 1-10 6:
6. Brown JR, et al. 2015. *Astrovirus* VA1/HMO-C: an increasingly recognized neurotropic pathogen in immunocompromised patients. *Clin Infect Dis.* 60(6):881-888.
7. Chan JZ, Pallen MJ, Oppenheim B, and Constantinidou C. 2012. Genome sequencing in clinical microbiology. *Nat Biotechnol.* 30(11):1068-1071.

8. Deamer D, Akeson M, and Branton D. 2016. Three decades of nanopore sequencing. *Nat Biotechnol.* 34(5):518-524.
9. De Coster W, Hert SD, Schultz T, Cruts M and Van Broeckhoven C. 2018. Nanopack: visualizing and processing long-read sequencing data. *Bioinformatic* 34(15):2666-2669.
10. Depledge DP et al. 2019. Direct RNA sequencing on nanopore arrays refines the transcriptional complexity of viral pathogen *Nature comm* (2019) 10:754 | <https://doi.org/10.1038/s41467-019-08734-9>
11. Gire SK, et al. 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science.* 345(6202):1369-1372.
12. Greninger AL et al. 2015. Rapid metagenomics identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genom Med*, 7:99-112.
13. Hoke CH Jr, Hawksworth A, and Snyder CE Jr. 2012. Initial assessment of impact of adenovirus type 4 and type 7 vaccine on febrile respiratory illness and virus transmission in military basic trainees, March 2012. *MSMR.* 19(3):2-4.
14. Imamura T, et al. 2011. Enterovirus 68 among children with severe acute respiratory infection, the Philippines. *Emerg Infect Dis.*17(8):1430-1435.
15. Forsyth BR, Bloom HH, Johnson KM, and Chanock RM. 1963. Patterns of illness in rhinovirus infections of military personnel. *N Engl J Med.* 269:602-606.
16. Gray GC, et al. 1999. Respiratory diseases among U.S. military personnel: countering emerging threats. *Emerg Infect Dis.* 5(3):379-385.
17. Ghedin E, et al. 2005 Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature.* 437 (7062):1162-1166.

18. Johns MC, 2009. Preliminary report: febrile acute respiratory disease caused by adenovirus type 14, U.S. Coast Guard Training Center, Cape May New Jersey, March 2009. *MSMR*. 16(5):10-11.
19. Harel N, Meir M Gophna U, and Stern A. 2019 sequencing complete genome of RNA viruses with MinION nanopore :finding associations between mutations bioRxiv Mar. 13, 2019; doi: <http://dx.doi.org/10.1101/575480>
20. Juul S, et al. 2015. What's in my pot? Real-time species identification on the MinION. <http://biorxiv.org/content/early/2015/11/06/030742>
21. Kaida A, et al. 2011. Enterovirus 68 in children with acute respiratory tract infections, Osaka, Japan. *Emerg Infect Dis*. 2011; 17(8):1494-1497.
22. Kilianski A, et.al. 2015. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *Gigascience*. 4:12.
23. Keller MW et al. 2018. Direct RNA sequencing of the coding complete Influenza A virus genome. *Sci.reports*, 8:14408.
24. Li H., et al. 2009. 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25, 2078-2079.
25. Li H. and Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics*, Epub. [PMID: [20080505](https://pubmed.ncbi.nlm.nih.gov/20080505/)]
26. Li L, et al. 2015. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *J Virol Methods*. 213:139-146.
27. Lim ML, Gabele SE, Wallace MR, Gray GC, and Earhart KC. 2004. Upper respiratory tract infections (URI). *Mil Med*. 169(4):15-16.

28. Loman NJ and Quinlan J. 2014. Poretools: a toolkit for analyzing nanopore sequence data
Bioinformatics 1:30(23)3399-3401.
29. Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled de novo
using only nanopore sequencing data. Nat Methods. 12(8):733-735.
30. Lu J, Breitwieser FP, Thielen P, Salzberg SL. 2017. Bracken: estimating species abundance
in metagenomics data. Peer J Computer Science 3:e104, doi:10.7717/peerj-cs.104
31. Madoui MA, et al. 2015. Genome assembly using Nanopore-guided long and error-free
DNA reads. BMC Genomics. 16:327.
32. Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P. 2013. Metagenomics for pathogen
detection in public health. Genome Med.5(9):81.
33. Naccache SN et al. 2015. Diagnosis of neuroinvasive astrovirus infection in an
immunocompromised adult with encephalitis by unbiased next-generation sequencing. Clin
Infect Dis.60 (6):919-923.
34. Nakamura S et al. 2009. Direct metagenomic detection of viral pathogens in nasal and fecal
specimens using an unbiased high-throughput sequencing approach. PLoS One. 4(1):e4219.
35. Ondov BD, Bergman NH, and Phillippy AM. 2011. Interactive metagenomics visualization
in web browser. BMC Bioinformatics 30: 12(1):385.
36. O'Shea MK, Pipkin C, Cane PA, and Gray GC. 2007. Respiratory syncytial virus: an
important cause of acute respiratory illness among young adults undergoing military
training. Influenza Other Respir Viruses. 1(5-6):193-197.
37. Rambo-Martin B L. et al. 2019. Mitigating pandemic risk with Influenza A virus field
Surveillance at a swine-human interface. bioRxiv Mar. 21, 2019; doi:
<http://dx.doi.org/10.1101/585588>

- 38.
39. Shin J et al. 2016. Analysis of mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Sci Rep.* ; 6:29681.
40. Silva CS et al. 2014. Human respiratory coronaviruses detected in patients with influenza-like illness in Arkansas, USA. *Virology*. S2. doi:10.4172/2161-0517.S2-004.
41. Van Boheemen S et al. 2012. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio.* 3(6):e00473-12.
42. Van Den Hoek S, Verhelst J, Vuylsteke M, and Saelens X. 2015. Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. *BMC Genomics.* 16:79
43. Vernikos G, Medini D, Riley DR, Tettelin H. 2015. Ten years of pan-genome analyses. *Curr Opin Microbiol.* 23:148-154.
44. Quast C, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools *Nucl. Acids Res.* 41 (D1): D590-D596.
45. Quick J, Quinlan AR, and Loman NJ. 2014. A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *Gigascience.* 3:22.
46. Quiñones-Mateu ME, Avila S, Reyes-Teran G, and Martinez MA. 2014. Deep sequencing: becoming a critical tool in clinical virology. *J Clin Virol.* 61(1):9-19.
47. Wongsurawat T et al. 2019 Rapid sequencing of multiple RNA viruses in their native form *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2019.00260>.
48. Wood DE, and Salzberg SL 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15:R46.

49. Yu H. 2013. Human infection with avian influenza A H7N9 virus: an assessment of clinical severity. *Lancet*. 382(9887):138-145.