Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis

Alexander L. Greninger[1,2], Samia N. Naccache[1,2,#], Scot Federman[1,2,#], Guixia Yu[1,2], Placide Mbala[3,6], Vanessa Bres[4], Doug Stryke[1,2], Jerome Bouquet[1,2], Sneha Somasekar[1,2], Jeffrey M. Linnen[4], Roger Dodd[5], Prime Mulembakani[6], Bradley S. Schneider[6], Jean-Jacques Muyembe[3], Susan L. Stramer[5], Charles Y. Chiu[1,2]

Affiliations:

[1]Department of Laboratory Medicine, University of California, San Francisco, CA, USA 94107

[2]UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA, USA, 91407

[4]Institut National de Recherche Biomédicale, Democratic Republic of the Congo, Kinshasa, Africa

[5]Hologic, Inc., Bedford, MA 01730

[5]American Red Cross, Gaithersburg, MD 2087

[6]Metabiota, Inc., San Francisco, CA 94104

[7]Department of Medicine, Division of Infectious Diseases, University of California, San Francisco

[#]These authors contributed equally to this work.

ABSTRACT

We report unbiased metagenomic detection of chikungunya virus (CHIKV), Ebola virus (EBOV), and hepatitis C virus (HCV) from four human blood samples by MinION nanopore sequencing coupled to a newly developed, web-based pipeline for real-time bioinformatics analysis on a computational server or laptop(MetaPORE). At titers ranging from 10^7-10^8 copies per milliliter, reads to EBOV from two patients with acute hemorrhagic fever and CHIKV from an asymptomatic blood donor were detected within 4 to 10 minutes of data acquisition, while lower titer HCV virus ($1x10^5$ copies per milliliter) was detected within 40 minutes. Analysis of mapped nanopore reads alone, despite an average individual error rate of 24% [range 8-49%], permitted identification of the correct viral strain in all 4 isolates, and 90% of the genome of CHIKV was recovered with >98% accuracy. Using nanopore sequencing, metagenomic detection of viral pathogens directly from clinical samples was performed within an unprecedented <6 hours sample-to-answer turnaround time and in a timeframe amenable for actionable clinical and public health diagnostics.

Background

Acute febrile illness has a broad differential diagnosis and can be caused by a variety of pathogens. Metagenomic next-generation sequencing (NGS) is particularly attractive for diagnosis and public health surveillance of febrile illness because the approach can broadly detect viruses, bacteria, and parasites in clinical samples by uniquely identifying sequence data[1, 2]. Although currently limited by a sample-to-answer turnaround time of >20 hr (Fig. 1), we and others have reported that unbiased pathogen detection using metagenomic NGS can generate actionable results in timeframes relevant to clinical diagnostics [3-6] and public health [7, 8]. As sequence reads are generated in parallel and not in series, real-time analysis by second-generation platforms such as Illumina and Ion Torrent has been hampered by the need to wait until a sufficient read length has been achieved for diagnostic pathogen identification.

Nanopore sequencing is a third-generation sequencing technology that has two key advantages over second-generation technologies – longer reads and the ability to perform real-time sequence analysis. To date, the longer nanopore reads have enabled scaffolding of prokaryotic and eukaryotic genomes and sequencing of bacterial and viral cultured isolates[9-12], but the platform's capacity for real-time metagenomic analysis of clinical samples has not yet been leveraged. As of mid-2015, the MinION nanopore sequencer is capable of producing at least 100,000 sequences with an average read length of 5 kB, in total producing up to 1 Gb of sequence in 24 hours on one flow cell [13]. Here we present using nanopore sequencing for metagenomic detection of viral pathogens from clinical samples with sample-to-answer turnaround times of under 6 hours. We also present MetaPORE, a real-time, web-based sequence analysis and visualization tool for pathogen identification from nanopore data.

Materials and Methods

MAP program

Since July 2014, our lab has participated in the MinION Access Program (MAP), an early access program for beta users of the Oxford Nanopore MinION. Program participants receive free flow cells and library preparation kits for testing and validation of new protocols and applications on the MinION platform. During our time in the MAP program, we have seen significant progress in quality control and sequencing yield of flow cells (Table 1).

Nucleic acid extraction

Total nucleic acid was extracted from 400 μL of a chikungunya virus (CHIKV)-positive serum sample previously collected from an asymptomatic blood donor during the 2014 CHIKV outbreak in Puerto Rico (Chik1) [14]. The serum sample was inactivated in a 1:3 ratio of TRIzol LS (Life Technologies, Carlsbad, CA, USA) at the American Red Cross prior to shipping to University of California, San Francisco (UCSF). Direct-zol RNA MiniPrep (Zymo Research, Irvine, CA, USA) was used for nucleic acid extraction, including on-column treatment with

Turbo DNAse (Life Technologies) for 30 min at 37°C to deplete human host genomic DNA.

For the Ebola virus (EBOV) specimens, total nucleic acid was extracted using a QIAamp Viral RNA kit (Qiagen, Valencia, CA, USA) from 140 μL of whole blood from two patients with suspected Ebola hemorrhagic fever from a 2014 outbreak in the Democratic Republic of the Congo (DRC) (Ebola1 and Ebola2). RNA was extracted at Institut National de Recherche Biomédicale in Kinshasa, DRC, preserved using RNAstable (Biomatrica, San Diego, CA, USA), and shipped at room temperature to UCSF. Upon receipt, the extracted RNA sample was treated with 1 μL Turbo DNase (Life Technologies), followed by clean-up using the Direct-zol RNA MiniPrep Kit (Zymo Research).

For the HCV sample, an HCV-positive serum specimen at a titer of $1.6x10^7$ cp/mL (HepC1) was diluted to $1x10^5$ cp/mL using pooled negative serum. Total nucleic acid was then extracted from 400uL of serum using the EZ1 Viral RNA kit followed by treatment with Turbo DNase at 30 min at 37°C and clean-up using the RNA Clean and Concentrator Kit (Zymo Research).

Molecular confirmation of viral infection

A previously reported TaqMan quantitative reverse-transcription PCR (qRT-PCR) assay targeting the EBOV NP gene was used for detection of EBOV and determination of viral load. The assay was run on a Stratagene MX300P real-time PCR instrument and performed using the TaqMan Fast Virus 1-Step Master Mix (Life Technologies) in 20 μL total reaction volume (5 μL 4x Taqman mix, 1ul sample extract), with 0.75uM of each primer (F565 5'-TCTGACATGGATTACCACAAGATC-3', R640 5'-GGATGACTCTTTGCCGAACAATC-3') and 0.6uM of the probe (p597S 6FAM-AGGTCTGTCCGTTCAA-MGBNFQ). Conditions for the qRT-PCR were as follows: 50°C, 10minutes / 95°C, 20 s followed by 45 cycles of 95°C 3s / 60°C, 30 s. Viral copy number was calculated by standard curve analysis with a plasmid vector containing the EBOV amplicon. The first EBOV sample analyzed by nanopore sequencing (Ebola1) corresponded to the Ebola virus/*H.sapiens*-wt/COD/2014/Lomela-Lokolia16 strain, while the second Ebola sample (Ebola2) corresponded to the Ebola virus/*H.sapiens*-wt/COD/2014/Lomela-LokoliaB11 strain. The CHIKV-positive sample was identified and quantified using a transcription-mediated amplification (TMA) assay (Hologic, Bedford, MA, USA) as previously described [14]. HCV was quantified using the FDA-approved Abbott RealTi*me* RT-PCR assay as performed in the UCSF Clinical Microbiology Laboratory on the Abbott Molecular m2000 system.

Construction of metagenomic amplified cDNA libraries

To obtain ≥1 μg of metagenomic cDNA library required for the nanopore sequencing protocol, randomly amplified cDNA was generated using a primer-extension pre-amplification method (Round A/B) as described previously[15-17]. Of note, this protocol has been extensively tested on clinical samples for metagenomic pan-pathogen detection of DNA and RNA viruses, bacteria, fungi, and parasites [4, 6, 15, 17, 18]. Briefly, in Round A, RNA was reversetranscribed with SuperScript III Reverse Transcriptase (Life Technologies,) using Sol-PrimerA (5'-GTTTCCCACTGGAGGATA-N_9-3'), followed by second-strand DNA synthesis with Sequenase DNA polymerase (Affymetrix, Santa Clara, CA, USA). Reaction conditions for Round A were as follows: 1 µL of Sol-PrimerA (40 pmol/µl) was added to 4 µl of sample RNA, heated at 65°C for 5 minutes, then cooled at room temperature for 5 minutes. Five µL of SuperScript Master Mix (2 µl 5X First-Strand Buffer, 1 µL water, 1 µL 12.5 mM dNTP mix, 0.5 µL 0.1M DTT, 0.5 µL SS III RT) was then added and incubated at 42°C for 60 minutes. For second strand synthesis, 5 µL of Sequenase Mix #1 (1 µL 5X Sequenase Buffer, 3.85 µL ddH_2O, 0.15 µL Sequenase enzyme) was added to the reaction mix and incubated at 37°C x 8 min, followed by addition of Sequenase Mix #2 (0.45 µl Sequenase Dilution Buffer, 0.15 µl Sequenase Enzyme) and a second incubation at 37°C x 8 min. Round B reaction conditions were as follows: 5 µL of Round A-labeled cDNA was added to 45 µL of KlenTaq master mix per sample (5 µL 10X KlenTaq PCR buffer, 1 µL 12.5 mM dNTP, 1 µL 100 pmol/µl Sol-PrimerB (5'-GTTTCCCACTGGAGGATA-3'), 1 µL KlenTaq LA (Sigma-Aldrich, St Louis, MO), 37 µL ddH_2O). Reaction conditions for the PCR were as follows: 94°C for 2 minutes; 25 cycles of 94°C for 30 sec, 50°C for 45 sec, 72°C for 60 sec; 72°C for 5 minutes.

Preparation of nanopore sequencing libraries

Amplified cDNA from Round B was purified using AMPure XP beads (Beckman Coulter, Brea, CA), and 1 µg DNA was used as input into Oxford

Nanopore Genomic DNA MAP-003 Kits (Chik1, Ebola1) and MAP-004 Kits (HepC1, Ebola2) for generation of MinION Oxford Nanopore-compatible libraries[9, 11]. Briefly, the steps include (1) addition of control lambda phage DNA, (2) end-repair with the NEBNext End Repair Module, (3) 1X AMPure purification, (4) dA-tailing with the NEBNext dA-tailing Module, (5) ligation to protein-linked adapters HP/AMP (Oxford Nanopore Technologies, Oxford, UK) using the NEBNext QuickLigation Module x 10 min at room temperature, (6) purification of ligated libraries using magnetic His-Tag Dynabeads (Life Technologies), and (7) elution in 25 µL buffer (Oxford Nanopore Technologies). Lambda phage DNA was not added during preparation of the Ebola2 sample library.

Nanopore sequencing

Nanopore libraries were run on an Oxford Nanopore MinION flow cell after loading 150 µL sequencing mix (6 µL library, 3 µL fuel mix, 141 µL BP mix) per the manufacturer's instructions. The Chik1 and Ebola1 samples were run consecutively on the same flow cell, with an interim wash performed using Wash-Kit-001 (Oxford Nanopore).

Illumina sequencing

For the Chik1 and Ebola1 specimens, amplified round B cDNA was purified using AMPure XP beads (Beckman Coulter) and 2 ng used as input into the Nextera XT Kit (Illumina). After 13 cycles of amplification, Illumina library concentration and average fragment size were determined using the Agilent Bioanalyzer. Sequencing was performed on an Illumina MiSeq using 150 nucleotide (nt) single-end runs and analyzed using the SURPI computational pipeline (UCSF, CA, USA) [15].

MetaPORE bioinformatics pipeline

We developed a custom bioinformatics pipeline for real-time pathogen identification and visualization from nanopore sequencing data (MetaPORE), available at https://github.com/chiulab/MetaPORE. The MetaPORE pipeline consists of a set of Linux shell scripts, Python programs, and JavaScript / HTML code, and was tested and run on an Ubuntu 14.10 computational server with 64 cores and 512 GB memory. Alternatively, MetaPORE was tested and run on a laptop (4 cores, 32GB RAM, Ubuntu 14.10) by restricting the identification database to viral sequences, rather than all of NCBI nt.

Raw FAST5/HDF files from the MinION instrument are base-called using the Metrichor 2D Basecalling v1.14 pipeline (Metrichor). The MetaPORE pipeline continually scans the Metrichor download directory for batch analysis of downloaded sequence reads. For each batch of files (collected every time 200 reads are downloaded in the download directory or ≥2 min of elapsed time), the 2D read or either the template or complement read, depending on which is of higher quality, is converted into a FASTQ file using HDF5 Tools (https://www.hdfgroup.org/HDF5/doc/RM/Tools.html). The *cutadapt* program is then used to trim Sol-PrimerB adapter sequences from the ends of the reads[19].

Next, the nucleotide BLAST (BLASTn) aligner is used to computationally subtract host reads [15, 20] aligning to the human fraction of the National Center for Biotechnology Information (NCBI) nucleotide collection database (NT database, downloaded March 2015) at word size 11 and e-value cutoff of 10^{-5}. On our 64 core machine, the remaining, non-human reads are then aligned by BLASTn to the entire NCBI NT database using the same parameters. On a laptop, the non-human reads are aligned to the viral fraction of the NCBI NT database. Reads that hit this viral database are then aligned by BLASTn to NCBI nt. For each read, the single best hit by e-value is retained. The NCBI GenBank gene identifier assigned to the best hit is then annotated by taxonomic lookup of the corresponding lineage, family, genus, and species [15].

For real-time visualization of results, a graphical user interface was developed for the MetaPORE pipeline. A "live" taxonomic count table is displayed as a donut chart using the CanvasJS (http://canvasjs.com) graphics suite, with the chart refreshing every 30 sec (Supplementary Data, Movie 1). For each viral family, genus, and species, the "top hit" is chosen to be the reference sequence with the greatest number of aligned reads, with priority given to reference sequences in the following order: (1) complete genomes; (2) complete sequence, or (3) partial sequences / individual genes. Coverage maps are generated in MetaPORE by mapping all aligned reads at a given taxonomic level (species, genus, or family) to the 'top hit" using LASTZ v1.02 [21], with interactive visualization provided using a custom web program that accesses the HighCharts JavaScript library (http://www.highcharts.com). A corresponding interactive pairwise identity plot is generated by using SAMtools [22] to calculate the consensus FASTA sequence from the coverage map, followed by pairwise 100 bp sliding-window comparisons of the consensus to the reference sequence using the BioPython implementation of the Needleman-Wunsch algorithm [23, 24]. For purposes of comparison, the MetaPORE pipeline was also run on a subset of 100,000 reads from parallel Illumina MiSeq data corresponding to the Chik1, Ebola1, and Ebola2 samples.

Phylogenetic Analysis

The overall CHIKV phylogeny (Fig. 2E, inset) consisted of all 188 near-complete or complete genome CHIKV sequences available in the NCBI NT database as of March 2015. A subphylogeny including the MiSeq- and nanopore-sequenced Puerto Rico strain PR-S6 presented here and previously[14], as well as additional Caribbean CHIKV strains and other representative members of the Asian-Pacific clade, was also analyzed (Fig. 2E). The EBOV phylogeny (Fig. 3E) consisted of the newly MiSeq- and nanopore-sequenced Ebola strain Lomela-LokoliaB11 from the 2014 DRC outbreak [25] and other representative EBOV strains, including strains from the 2014-2015 West African outbreak[8, 26]. Sequences were aligned using the MAFFT algorithm[27], and phylogenetic trees were constructed using the MrBayes algorithm [28] in the Geneious software package [29].

RESULTS

Example 1: High-titer CHIKV (Flowcell #1)

To test the ability of nanopore sequencing to identify metagenomic reads from a clinical sample, we first analyzed a plasma sample harboring high-titer CHIKV and previously sequenced on an Illumina MiSeq platform (Fig. 2B) [14]. The plasma sample corresponded to an asymptomatic blood donor who had screened positive for CHIKV infection during the 2014 outbreak in Puerto Rico (strain PR-S6), with a calculated viral titer of 9.1x10^7 copies/mL.

A read aligning to CHIKV, the 96th read, was sequenced within 6 min (Fig. 2B) and detected by BLASTN alignment to the NCBI NT database within 8 min of data acquisition, demonstrating an overall sample-to-detection turnaround time of <6 hr (Fig. 1). After early termination of the sequencing run at the 2 hr 15 min time point, 556 of 19,452 total reads (2.85%) were found to align to CHIKV (Fig. 2B and C). The individual CHIKV nanopore reads had an average length of 455 nucleotides (nt) [range 126-1477] and average percent identity of 79% [range 51-92%] to the most closely matched reference strain, a CHIKV strain from the neighboring British Virgin Islands (KJ451624) (Table 1). When only high-quality 2D pass reads were included, 346 of 5,139 (6.7%) reads aligned to CHIKV, comparable to the proportion of CHIKV reads (248,677 of 3,235,096, 7.7%) identified by corresponding metagenomic sequencing on the Illumina MiSeq (Fig. S1).

Mapping of the 556 nanopore reads aligning to CHIKV to the assigned reference genome (KJ451624) showed recovery of 90% of the genome at 3X coverage and 98% at 1X coverage. Notably, despite high individual mapped error rates in the nanopore reads, ranging from 8 to 49%, 97-99% identity to the reference genome (KJ451624) was achieved across contiguous regions with at least 3X coverage (Fig. 2D). Furthermore, phylogenetic analysis revealed co-clustering of the CHIKV genomes independently assembled from MinION nanopore and Illumina MiSeq reads on the same branch within the Caribbean subclade (Fig. 2E). Overall, a large proportion of reads (55%) in the error-prone nanopore data remained unidentifiable, while other aligning reads aside from CHIKV corresponded to human, lambda phage control spike-in, uncultured bacterial, or other eukaryotic sequences (Fig. 2C).

Example 2: High-titer Ebola virus (Flowcell #1)

We next attempted to replicate our metagenomic detection result on the nanopore sequencer with a different virus by testing a whole-blood sample from a patient with Ebola hemorrhagic fever during the August 2014 outbreak in the DRC (Ebola1, strain Lomela-Lokolia16) [25]. To conserve flowcells, the same nanopore flowcell used to run the Chik1 sample was washed and stored overnight at 4°C, followed by nanopore sequencing of the Ebola1 sample (viral titer of 1.0x10^7 copies/mL by real-time qRT-PCR). Only 41 of 13,090 nanopore reads (0.31%) aligned to EBOV, as compared to 20,820 of 2,743,589 reads (0.76%) at 117X coverage for the Illumina MiSeq (Fig. 3B and D; Fig S1). The decrease in relative number (41 versus 556) and percentage (0.31% versus 2.85%) of target viral nanopore reads in the Ebola1 relative to Chik1 sample is consistent with the lower levels of viremia (1.0x10^7 versus 9.1x10^7 copies/mL) and high host background (whole blood versus plasma) (Fig. 2C). Nonetheless, the first read aligning to EBOV was detected in a similar timeframe as in the Chik1 sample, sequenced within 8 min and detected within 10 min of data acquisition. EBOV nanopore reads were 359 nt in length on average [range 220-672 nt], with a mean individual pairwise identity of 78% [range 56-88%]. Nevertheless, the majority of Ebola sequences (31 of 41, 76%) were found to align to the most closely matched strain of EBOV in the NT reference database, strain Lomela-Lokolia16 (Fig. 2D).

Despite washing the flowcell between the two successive runs, 7 CHIKV reads were recovered during the Ebola1 library sequencing, suggesting the potential for carryover contamination. CHIKV reads were not present in the corresponding Illumina MiSeq run (Fig. S1-B), confirming that the source of the contamination derived from the Chik1 library that was run on the same flow cell as the Ebola1 library.

Example 3: Moderate-titer HCV (Flowcell #2)

Our previous experiments revealed both the total number of metagenomic reads and proportion of target viral reads at a given titer that could be obtained from a single MinION flowcell, and showed that the proportion of viral reads obtained by metagenomic nanopore and MiSeq sequencing was comparable. Thus, we projected that the minimum concentration of virus that could be reproducibly detected using our current metagenomic protocol would be 1x10^5 copies/mL. An HCV-positive clinical sample (HepC1) was diluted in negative control serum matrix to a titer of 1x10^5 copies/mL and processed for nanopore sequencing using an upgraded library preparation kit (MAP-004). After four consecutive runs on the same flowcell with repeat loading of the same metagenomic HepC1 library (Fig. 3A), a total of 85,647 reads were generated, of which only 6 (0.0070%) aligned to HCV (Fig. 3B). Although the entire series of flowcell runs lasted for >12 hr, the first HCV read was sequenced within 34 min (Fig. 3B), enabling BLASTn detection within 36 min of data acquisition. Given the low titer of HCV in the HepC1 sample and hence low corresponding fraction of HCV reads in the nanopore data, the vast majority (96%) of viral sequences identified corresponded to the background lambda phage spike-in. Importantly, although only 6 HCV reads were identified by nanopore sequencing, all 6 reads aligned to the correct genotype, genotype 1b (Fig. 3D).

Example 4: High-titer Ebola virus with real-time MetaPORE analysis (Flowcell #3)

To enable real-time analysis of nanopore sequencing data, we combined our BLASTn analysis with monitoring and user-friendly web visualization into a real-time bioinformatics pipeline for pathogen detection named MetaPORE. We tested MetaPORE by sequencing a nanopore library (Ebola2) constructed using the upgraded MAP-004 kit and corresponding to a whole blood sample from a patient with suspected Ebola hemorrhagic fever during the 2014 DRC outbreak. Four consecutive runs of the Ebola2 library on the same flowcell over 34 hr, yielded a total of 335,044 reads, of which 593 (0.18%) aligned to EBOV (141 of 6009, or 2.3%, of 2D pass reads). Notably, the first EBOV read was sequenced 44 sec after data acquisition and correctly detected in ~3 min by MetaPORE (Fig. 4B; Supplementary Information: Movie). A total of 3 EBOV reads, mapping across the genome and confirming unambiguous detection of the virus, were identified using MetaPORE within 9 min of data acquisition (Supplementary Information: Movie). In the corresponding Illumina MiSeq data, 1,778 reads (0.93%) out of a total of 192,158 reads aligned to EBOV by BLASTn, comparable to the proportion of EBOV reads in nanopore 2D pass data. Nanopore read coverage across the EBOV genome was relatively uniform with at least 1 read mapping to >88% of the genome and areas of zero coverage also seen with much higher-coverage Illumina MiSeq data (Fig. 4D). The detection of EBOV by real-time metagenomic nanopore sequencing was confirmed by qRT-PCR testing of the clinical blood sample, which was positive for EBOV at an estimated titer of 7.64 x10^7 cp/ml. Phylogenetic analysis of the Ebola2 genome independently recovered by MinION nanopore and Illumina MiSeq sequencing revealed that nanopore sequencing alone was capable of pinpointing the correct EBOV outbreak strain and country of origin (Fig. 4E).

DISCUSSION

Unbiased point-of-care testing for pathogens by rapid metagenomic sequencing has the potential to radically transform infectious disease diagnosis in clinical and public health settings. In this study, we sought to demonstrate the potential of the nanopore instrument for metagenomic pathogen identification in clinical samples by coupling an established assay protocol with a new, real-time sequence analysis pipeline. To date, high reported error rates (10-30%) and relatively low-throughput (<100,000 reads per flow cell) have hindered the utility of nanopore sequencing for analysis of metagenomic clinical samples [9, 11]. Prior work on infectious disease diagnostics using nanopore has focused on rapid PCR amplicon sequencing of viruses and bacteria[11], or real-time sequencing of pure bacterial isolates in culture, such *Salmonella* in a hospital outbreak [12]. To our knowledge, this is the first time that nanopore sequencing has been used for real-time metagenomic detection of pathogens in complex, high-background clinical samples in the setting of human infections. Here we also sequenced a microbial genome to high accuracy (>98% identity) directly from a metagenomic clinical sample and not from culture, using only multiple overlapping, albeit error-prone, nanopore reads and without resorting to the use of a secondary platform such as an Illumina MiSeq for sequence correction (Fig. 2D).

Real-time analysis is critical for time-critical sequencing applications such as outbreak investigation[7] and metagenomic diagnosis of life-threatening infections in hospitalized patients[3, 4, 6]. NGS analysis for clinical diagnostics is currently performed after sequencing is completed, analogous to how PCR products were analyzed by agarose gel electrophoresis in the 1990s. To date, most clinical PCR assays have been converted to a real-time format that reduces "hands-on" laboratory technician time and effort, decreases overall sample-to-answer turnaround times, and potentially yields quantitative information. Notably, our nanopore data suggest that very few reads are needed to provide an unambiguous diagnostic identification, despite high individual per read error rates of 10-30%. The ability of nanopore sequence analysis to accurate identify viruses to the species and even strain / genotype level is facilitated by the high specificity of sequence data, especially with the longer target viral reads achievable by nanopore (Table 1, 391 bp average length) versus second-generation sequencing.

Although the overall turnaround time from metagenomic sample-to-detection has been reduced to <6 hr, many challenges remain for routine implementation of this technology in clinical and public health settings. Improvements to make library preparation faster and more robust are critical, including automation and optimization of each step in the protocol. We also looked only at clinical samples at moderate to high titers of 10^5-10^8 copies / mL, and the sensitivity of metagenomic nanopore sequencing at lower titers remains unclear at current achievable sequencing depths. Standard wash protocols appear inadequate to prevent cross-contamination when reusing the same flow cell, as CHIKV reads were identified in the downstream Ebola1 sample sequence run. One solution may be to perform only one nanopore sequencing run per flow cell for clinical diagnostic purposes, akin to how disposable cartridges are used for clinical quantitative PCR testing on a Cepheid GenXpert instrument to prevent cross-contamination[30]. Another is to uniquely barcode individual samples at the cost of added time and effort. Finally, the current accuracy of a single nanopore read will most likely be insufficient to allow confident species identification of bacteria, fungi, or parasites, which have much larger genomes and shared conserved genomic regions than viruses. Single-nucleotide resolution will also be required for detection of antimicrobial resistance markers [31], which is difficult to achieve from relatively low-coverage metagenomic data [32]. These limitations can potentially be overcome by target enrichment methods such as capture probes to increase coverage, improvements in nanopore sequencing technology, or more accurate base-calling and alignment algorithms for nanopore data [33, 34].

CONCLUSION

Our results indicate that unbiased metagenomic detection of viral pathogens from clinical samples with a sample-to-answer turnaround time of <6 hours and real-time bioinformatics analysis is feasible with nanopore sequencing. We demonstrate unbiased detection of diagnostic EBOV sequence in under four minutes after the Oxford Minion nanopore initiated sequencing. This technology will be particularly desirable for enabling point-of-care genomic analyses in the developing world, where critical resources, including reliable electric power, laboratory space, and computational server capacity, are often severely limited. MetaPORE, the real-time sequencing analysis platform developed here, is web-based and able to be run on a laptop. As sequencing yield, quality, and turnaround times continue to improve, we anticipate that third-generation technologies such as nanopore sequencing will challenge clinical diagnostic mainstays such as PCR and TMA testing, fulfilling the dream of an unbiased, point-of-care test for infectious diseases.

FIGURE LEGENDS

Figure 1. Metagenomic sequencing workflow for MinION nanopore sequencing compared to Illumina MiSeq sequencing. The turnaround time for sample-to-detection nanopore sequencing, defined here as the cumulative time taken for nucleic acid extraction, reverse transcription, library preparation, sequencing, MetaPORE bioinformatics analysis, and detection, was under 6 hr, while Illumina sequencing took over 20 hr. The time differential is accounted for by increased times for library quantitation, sequencing, and bioinformatics analysis with the Illumina protocol. *assumes a 12-hr 50 nucleotide (nt) single-end MiSeq run of ~12-15 million reads, with 50 nt the minimum read length needed for accurate pathogen identification. **denotes estimated average SURPI bioinformatics analysis run length for MiSeq data[15]. The stopwatch is depicted as a 12-hr clock.

Figure 2. Metagenomic identification of CHIKV and EBOV from clinical blood samples by nanopore sequencing. (A) Timeline of sequencing runs on flowcell #1 with sample reloading, plotted as a function of elapsed time in hr since the start of flowcell sequencing. **(B)** Cumulative numbers of all sequenced reads (black line) and target viral reads (red line) from the Chik1 run (left) or Ebola1 run (right), plotted as a function of individual sequencing run time in min. **(C)** Taxonomic donut charts generated using the MetaPORE bioinformatics analysis pipeline from the Chik1 run (left) and Ebola1 run (right). The total number of reads analyzed is shown in the center of the donut. **(D)** Coverage and pairwise identity plots generated in MetaPORE by mapping reads aligning to CHIKV (left, Chik1 run) or EBOV (right, Ebola1 run) to the closest matching reference genome in the NT database (asterisk). **(E)** Whole-genome phylogeny of CHIKV. Representative CHIKV genome sequences from the Asian-Pacific clade, including the Puerto Rico PR-S6 strain recovered by nanopore and MiSeq sequencing, or all 188 near-complete or complete CHIKV genome sequences available in the NCBI NT database as of March 2015 (inset), are included. Branch lengths are drawn proportionally to the number of nucleotide substitutions per position, and support values are shown for each node. Abbreviations: CHIKV, chikungunya virus; EBOV, Ebola virus; Chik1, chikungunya virus, strain PR-S6 sample; Ebola1, EBOV, strain Lomela-Lokolia16 sample.

Figure 3. Metagenomic identification of HCV from a clinical serum sample by nanopore sequencing. (A) Timeline of sequencing runs on flowcell #2 with HepC1 sample reloading, plotted as a function of elapsed time in hr since the start of flowcell sequencing. **(B)** Cumulative number of all sequenced reads (black line) and HCV viral reads (red line), plotted as a function of individual sequencing run time in min. **(C)** Taxonomic donut charts generated using the MetaPORE bioinformatics analysis pipeline. The total number of reads analyzed is shown in the center of the donut. **(D)** Coverage and pairwise identity plots generated in MetaPORE by mapping reads aligning to HCV to the closest matching reference genome in the NT database. Abbreviations: HCV, hepatitis C virus; HepC1: hepatitis C virus, genotype 1b sample.

Figure 4. Metagenomic identification of EBOV from a clinical blood sample by nanopore sequencing and MetaPORE real-time bioinformatics analysis. Nanopore data generated from the Ebola2 library and sequenced on flowcell #3 was analyzed in real-time using the MetaPORE bioinformatics analysis pipeline, and compared to corresponding MiSeq data. **(A)** Timeline of nanopore sequencing runs on flowcell #3 with sample reloading, plotted as a function of elapsed time in hr since the start of flowcell sequencing. **(B)** Cumulative numbers of all sequenced reads (black line) and target viral reads (red line) from the nanopore run (left) or MiSeq run (right), plotted as a function of individual sequencing run time in min. **(C)** Taxonomic donut charts generated by real-time MetaPORE analysis of the nanopore reads (left) and post-run analysis of the

MiSeq reads (right). The total number of reads analyzed is shown in the center of the donut. Note that only a subset of Illumina reads (n=100,000) were analyzed using the MetaPORE pipeline. **(D)** Coverage and pairwise identity plots generated by MetaPORE analysis of nanopore reads (left) and MiSeq reads (right), generated by mapping reads aligning to EBOV to the closest matching reference genome in the NT database (asterisk). Abbreviations: EBOV, Ebola virus; Ebola2, EBOV, strain Lomela-LokoliaB11 sample.

Supplemental Figure 1. MetaPORE analysis of Illumina MiSeq data from samples containing CHIKV and EBOV. (A) Taxonomic donut charts generated from MiSeq Chik1 run data (left) and Ebola1 run data (right) using the MetaPORE bioinformatics analysis pipeline. The total number of MiSeq reads analyzed is shown in the center of the donut. Note that only a subset of reads (n=100,000) were analyzed. **(D)** Coverage and pairwise identity plots generated in MetaPORE by mapping CHIKV reads from the Chik1 run (left) or EBOV reads from the Ebola1 run (right) to the closest matching reference genome in the NT database.

Movie. MetaPORE Real-Time Bioinformatics Analysis and Visualization (clip 0:19 - 10:13, 9.8 min) Detection of EBOV by metagenomic nanopore sequencing and real-time MetaPORE bioinformatics analysis. Raw FAST5 files are uploaded to the Metrichor cloud-based analytics platform for 2D basecalling (right panel). After downloading from Metrichor, basecalled FAST5 reads are collected in batches of 200 reads and automatically processed in real-time by MetaPORE. Reads corresponding to human host background are computationally subtracted by BLASTn alignment to the human portion of the NCBI NT database. Following subtraction, remaining reads are then aligned to NCBI NT using BLASTn for taxonomic identification. The closest matched reference genomes on the basis of hit frequency are shown in donut plots that are updated each minute in real-time (left panel). Note that the first EBOV read from the Ebola2 sequencing run is detected 3 min 6 sec (3:26) after the start of sequence acquisition (0:19). (clip 10:21 - 11:51, 1.5 min) Web-based, interactive coverage map and pairwise identity plots, generated in real-time by MetaPORE, enable zooming, highlighting of individual values, outputting of relevant statistical data, and exporting of the graphs in various formats. The plots shown in the movie correspond to the analyzed data after completion of nanopore sequencing.

Data Availability

Nanopore and MiSeq sequencing data, along with sample metadata, have been submitted to NCBI under the following GenBank accession numbers: Sample metadata along with nanopore and MiSeq sequencing data have been submitted to the NCBI for each sample under the following SRA Study accessions: Ebola virus/H.sapiens-wt/COD/2014/Lomela-Lokolia16 (SRP057409), Ebola virus/H.sapiens-wt/COD/2014/Lomela-LokoliaB11 (SRS933322), Chik1 (SRP057410), and HepC1 (SRP057418).

SOURCE OF FUNDING

This study is supported in part by a grant from the National Institutes of Health (R01-HL105704) (CYC) and an UCSF-Abbott Viral Discovery Award (CYC).

COMPETING INTERESTS

CYC is the director of the UCSF-Abbott Viral Diagnostics and Discovery Center (VDDC) and receives research support in pathogen discovery from Abbott Laboratories, Inc. JL and VB are employees of Hologic, Inc. P. Mbala and P. Mulembakani and BS are employees of Metabiota, Inc.

AUTHOR’S CONTRIBUTIONS:

ALG, and CYC directed the study. ALG performed all nanopore experiments; SNN, GY, SS, and JB provided cDNA libraries and generated Illumina libraries. SF and ALG constructed the MetaPORE pipeline; SF,DS, CYC generated real time visualization; ALG, SF, SNN validated MetaPORE pipeline. VB, JML, RD, and SS provided chikungunya sample; PM, BSS, JJM provided ebola sample. ALG, SF, SNN, and CYC wrote the manuscript.

REFERENCES

1. Pallen MJ: **Diagnostic metagenomics: potential applications to bacterial, viral and parasitic infections.** *Parasitology* 2014, **141:**1856-1862.
2. Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P: **Metagenomics for pathogen detection in public health.** *Genome Med* 2013, **5:**81.
3. Brown JR, Morfopoulou S, Hubb J, Emmett WA, Ip W, Shah D, Brooks T, Paine SM, Anderson G, Virasami A, et al: **Astrovirus VA1/HMO-C: an increasingly recognized neurotropic pathogen in immunocompromised patients.** *Clin Infect Dis* 2015, **60:**881-888.
4. Naccache SN, Peggs KS, Mattes FM, Phadke R, Garson JA, Grant P, Samayoa E, Federman S, Miller S, Lunn MP, et al: **Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing.** *Clin Infect Dis* 2015, **60:**919-923.
5. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan PL, Hui J, Marshall J, et al: **A new arenavirus in a cluster of fatal transplant-associated diseases.** *N Engl J Med* 2008, **358:**991-998.
6. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, Miller S, et al: **Actionable diagnosis of neuroleptospirosis by next-generation sequencing.** *N Engl J Med* 2014, **370:**2408-2417.
7. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, Khristova ML, Weyer J, Swanepoel R, Egholm M, et al: **Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa.** *PLoS Pathog* 2009, **5:**e1000455.
8. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, Jalloh S, Momoh M, Fullah M, Dudas G, et al: **Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak.** *Science* 2014, **345:**1369-1372.
9. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, Wain J, O'Grady J: **MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island.** *Nat Biotechnol* 2015, **33:**296-300.
10. Goodwin S, Gurtowski J, Ethe-Sayers S, Despande P, Schatz M, McCombie WR: **Oxford Nanopore Sequencing and de novo Assembly of a Eukaryotic Genome.** *bioRxiv* 2015, http://dx.doi.org/10.1101/013490.
11. Kilianski A, Haas JL, Corriveau EJ, Liem AT, Willis KL, Kadavy DR, Rosenzweig CN, Minot SS: **Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer.** *Gigascience* 2015, **4:**12.
12. Quick J, Ashton P, Calus S, Chatt C, Gossain S, Hawker J, Nair S, Neal K, Nye K, Peters T, et al: **Rapid draft sequencign and real-time**

nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biology* 2015, **16**.

13. **The MinION Access Programme - Community** [https://nanoporetech.com/community/the-minion-access-programme]

14. Chiu CY, Bres V, Yu G, Krysztof D, Naccache SN, Lee D, Pfeil J, Linnen JM, Stramer SL: **Emerging genomic assays for identification of chikungunya virus infection in blood donors from Puerto Rico, 2014.** *Emerging Infectious Diseases* 2015, **(in press)**.

15. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, Bouquet J, Greninger AL, Luk KC, Enge B, et al: **A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples.** *Genome Res* 2014, **24:**1180-1192.

16. Chen EC, Miller SA, DeRisi JL, Chiu CY: **Using a pan-viral microarray assay (Virochip) to screen clinical samples for viral pathogens.** *J Vis Exp* 2011.

17. Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G, Kim E, Pillai DR, Guyard C, Mazzulli T, et al: **A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America.** *PLoS One* 2010, **5:**e13381.

18. Greninger AL, Naccache SN, Messacar K, Clayton A, Yu G, Somasekar S, Federman S, Stryke D, Anderson C, Yagi S, et al: **A novel outbreak enterovirus D68 strain associated with acute flaccid myelitis cases in the USA (2012-14): a retrospective cohort study.** *Lancet Infect Dis* 2015.

19. Martin M: **Cutadapt removes adapter sequences from high-throughput sequenincg reads.** *EMBnetjournal* 2011, **17:**10-12.

20. Chiu CY: **Viral pathogen discovery.** *Curr Opin Microbiol* 2013, **16:**468-478.

21. Harris RS: **Improved Pairwise Alignment of Genomic DNA.** Pennsylvania State University, 2007.

22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25:**2078-2079.

23. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ: **Biopython: freely available Python tools for computational molecular biology and bioinformatics.** *Bioinformatics* 2009, **25:**1422-1423.

24. Needleman SB, Wunsch CD: **A general method applicable to the search for similarities in the amino acid sequence of two proteins.** *J Mol Biol* 1970, **48:**443-453.

25. Maganga GD, Kapetshi J, Berthet N, Kebela Ilunga B, Kabange F, Mbala Kingebeni P, Mondonge V, Muyembe JJ, Bertherat E, Briand S, et al: **Ebola virus disease in the Democratic Republic of Congo.** *N Engl J Med* 2014, **371:**2083-2091.

26. Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, Soropogui B, Sow MS, Keita S, De Clerck H, et al: **Emergence of Zaire Ebola virus disease in Guinea.** *N Engl J Med* 2014, **371:**1418-1425.
27. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** *Mol Biol Evol* 2013, **30:**772-780.
28. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17:**754-755.
29. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.** *Bioinformatics* 2012, **28:**1647-1649.
30. Blakemore R, Story E, Helb D, Kop J, Banada P, Owens MR, Chakravorty S, Jones M, Alland D: **Evaluation of the analytical performance of the Xpert MTB/RIF assay.** *J Clin Microbiol* 2010, **48:**2495-2501.
31. Fournier PE, Dubourg G, Raoult D: **Clinical detection and characterization of bacterial pathogens in the genomics era.** *Genome Med* 2014, **6:**114.
32. Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P: **A bioinformatician's guide to metagenomics.** *Microbiol Mol Biol Rev* 2008, **72:**557-578, Table of Contents.
33. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M: **Improved data analysis for the MinION nanopore sequencer.** *Nat Methods* 2015, **12:**351-356.
34. Loman NJ, Watson M: **Successful test launch for nanopore sequencing.** *Nat Methods* 2015, **12:**303-304.

Table 1

Experiment / Sample	Flow Cell #	Run #	# of Active Pores	Run time (min)	Total reads	Pass reads	Fail reads	Pass / fail rate	Target Virus*
Chik1	1	first run	345	138	19,452	5,139	14,313	35.9%	CHIKV
Ebola1	1	first run	105	1,022	13,090	1,831	11,259	16.3%	EBOV
HepC1	2	first run	171	122	10,305	729	9,877	7.4%	HCV
HepC1	2	reload #1	293	192	26,626	2,155	25,758	8.4%	HCV
HepC1	2	reload #2	256	298	32,212	1,207	31,289	3.9%	HCV
HepC1	2	reload #3	214	156	14,805	287	14,275	2.0%	HCV
Ebola2	3	first run	397	79	28,651	1,537	27,114	5.7%	EBOV
Ebola2	3	reload #1	426	222	95,861	2,899	92,962	3.1%	EBOV
Ebola2	3	reload #2	380	1,091	166,524	1,539	164,985	0.9%	EBOV
Ebola2	3	reload #3	200	1,357	44,272	34	44,238	0.1%	EBOV
TOTAL					451,798	17,357	436,070	4.0%	

Table S1. Run data for each flow cell.

*abbreviations: CHIKV, chikungunya virus; EBOV, Ebola virus; HCV, hepatitis C virus
**based on average pairwise identity of aligned viral reads to the most closely matched refere

Avg viral read length	Avg viral read error rate**
375	20.6% (8-49%)
317	22.0% (12-43%)
497	
511	
506	
479	33.1% (24-46%)
366	
377	
372	
326	22.3% (8-48%)
391	24.3% (8-49%)

ence sequence

Figure 1

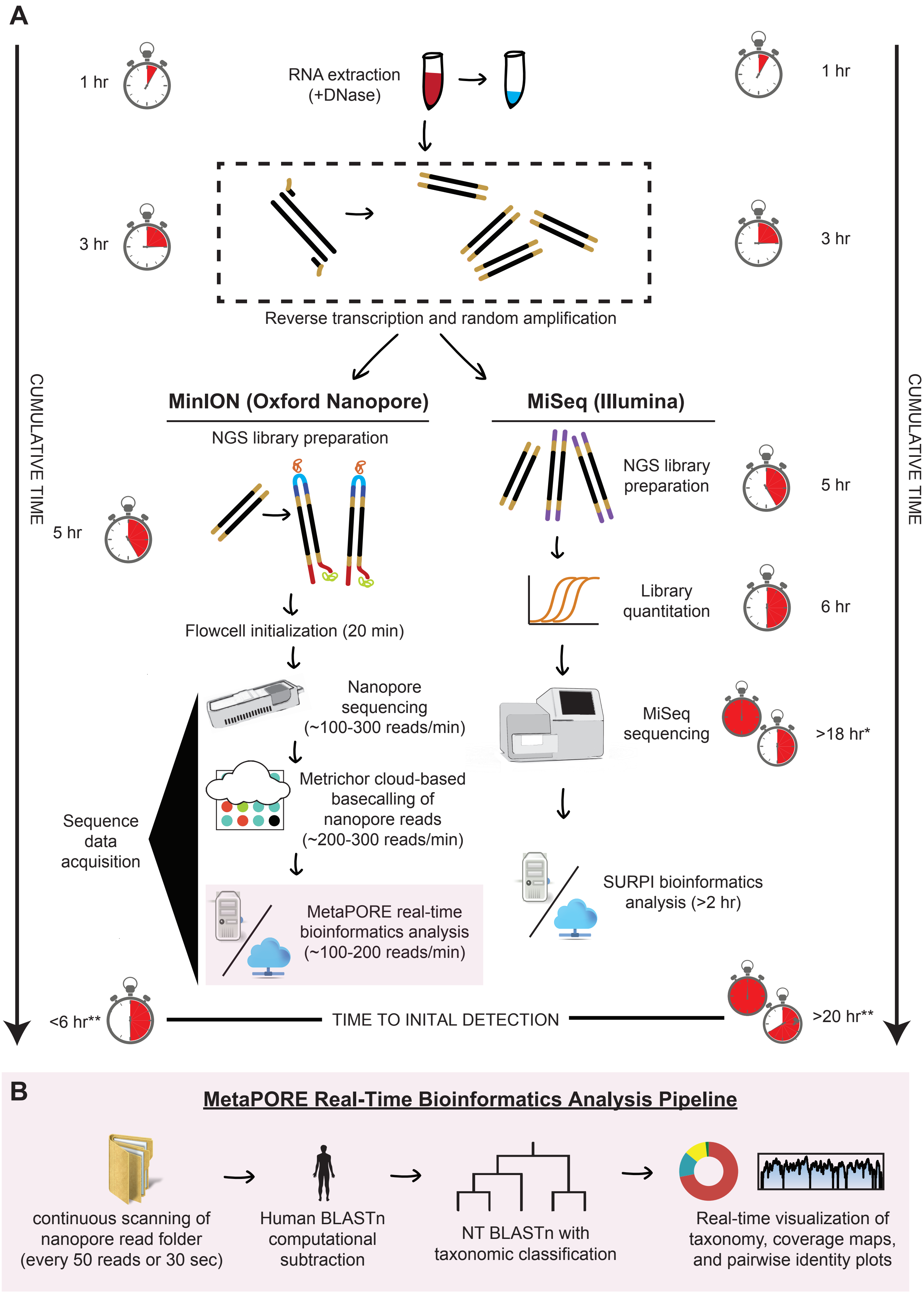

A
1 hr
RNA extraction
(+DNase)
1 hr
3 hr
3 hr
Reverse transcription and random amplification
CUMULATIVE TIME
MinION (Oxford Nanopore)
MiSeq (Illumina)
CUMULATIVE TIME
NGS library preparation
NGS library
preparation
5 hr
5 hr
Library
quantitation
6 hr
Flowcell initialization (20 min)
Nanopore
sequencing
(~100-300 reads/min)
MiSeq
sequencing
>18 hr*
Metrichor cloud-based
basecalling of
nanopore reads
(~200-300 reads/min)
Sequence
data
acquisition
SURPI bioinformatics
analysis (>2 hr)
MetaPORE real-time
bioinformatics analysis
(~100-200 reads/min)
<6 hr**
TIME TO INITAL DETECTION
>20 hr**
B
MetaPORE Real-Time Bioinformatics Analysis Pipeline
continuous scanning of
nanopore read folder
(every 50 reads or 30 sec)
Human BLASTn
computational
subtraction
NT BLASTn with
taxonomic classification
Real-time visualization of
taxonomy, coverage maps,
and pairwise identity plots

Figure 2

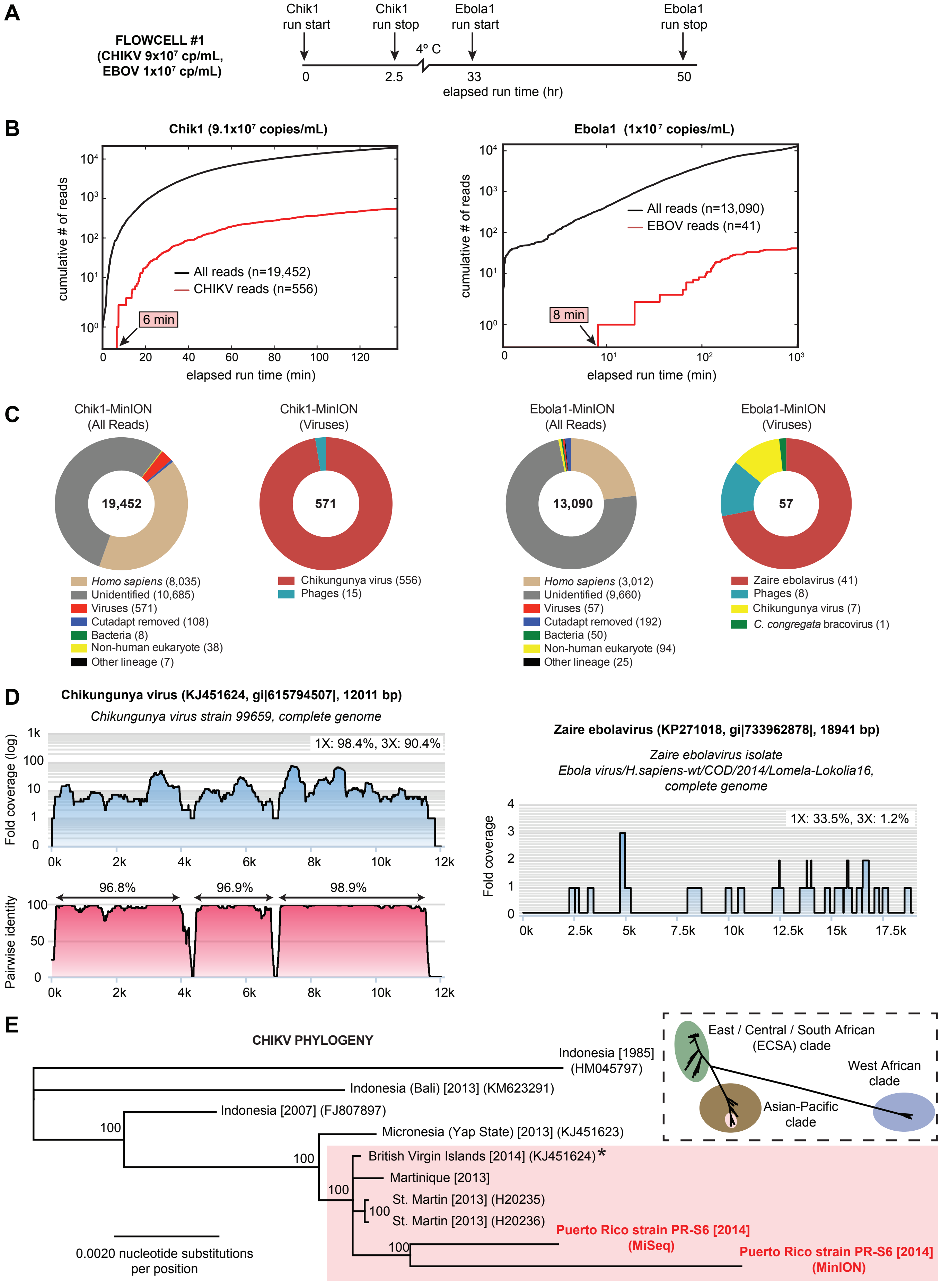

A
FLOWCELL #1
(CHIKV 9x10⁷ cp/mL,
EBOV 1x10⁷ cp/mL)
Chik1 run start
Chik1 run stop
4° C
Ebola1 run start
Ebola1 run stop
0
2.5
33
50
elapsed run time (hr)
B
Chik1 (9.1x10⁷ copies/mL)
cumulative # of reads
10⁰
10¹
10²
10³
10⁴
All reads (n=19,452)
CHIKV reads (n=556)
6 min
0
20
40
60
80
100
120
elapsed run time (min)
Ebola1 (1x10⁷ copies/mL)
All reads (n=13,090)
EBOV reads (n=41)
8 min
10¹
10²
10³
elapsed run time (min)
C
Chik1-MinION (All Reads)
19,452
Homo sapiens (8,035)
Unidentified (10,685)
Viruses (571)
Cutadapt removed (108)
Bacteria (8)
Non-human eukaryote (38)
Other lineage (7)
Chik1-MinION (Viruses)
571
Chikungunya virus (556)
Phages (15)
Ebola1-MinION (All Reads)
13,090
Homo sapiens (3,012)
Unidentified (9,660)
Viruses (57)
Cutadapt removed (192)
Bacteria (50)
Non-human eukaryote (94)
Other lineage (25)
Ebola1-MinION (Viruses)
57
Zaire ebolavirus (41)
Phages (8)
Chikungunya virus (7)
C. congregata bracovirus (1)
D
Chikungunya virus (KJ451624, gi|615794507|, 12011 bp)
Chikungunya virus strain 99659, complete genome
1X: 98.4%, 3X: 90.4%
Fold coverage (log)
0
1
10
100
1k
0k
2k
4k
6k
8k
10k
12k
96.8%
96.9%
98.9%
Pairwise identity
0
50
100
Zaire ebolavirus (KP271018, gi|733962878|, 18941 bp)
Zaire ebolavirus isolate
Ebola virus/H.sapiens-wt/COD/2014/Lomela-Lokolia16,
complete genome
1X: 33.5%, 3X: 1.2%
Fold coverage
0
1
2
3
4
0k
2.5k
5k
7.5k
10k
12.5k
15k
17.5k
E
CHIKV PHYLOGENY
Indonesia [1985] (HM045797)
Indonesia (Bali) [2013] (KM623291)
Indonesia [2007] (FJ807897)
Micronesia (Yap State) [2013] (KJ451623)
British Virgin Islands [2014] (KJ451624) *
Martinique [2013]
St. Martin [2013] (H20235)
St. Martin [2013] (H20236)
Puerto Rico strain PR-S6 [2014] (MiSeq)
Puerto Rico strain PR-S6 [2014] (MinION)
100
100
100
100
100
0.0020 nucleotide substitutions per position
East / Central / South African (ECSA) clade
West African clade
Asian-Pacific clade

Figure 3

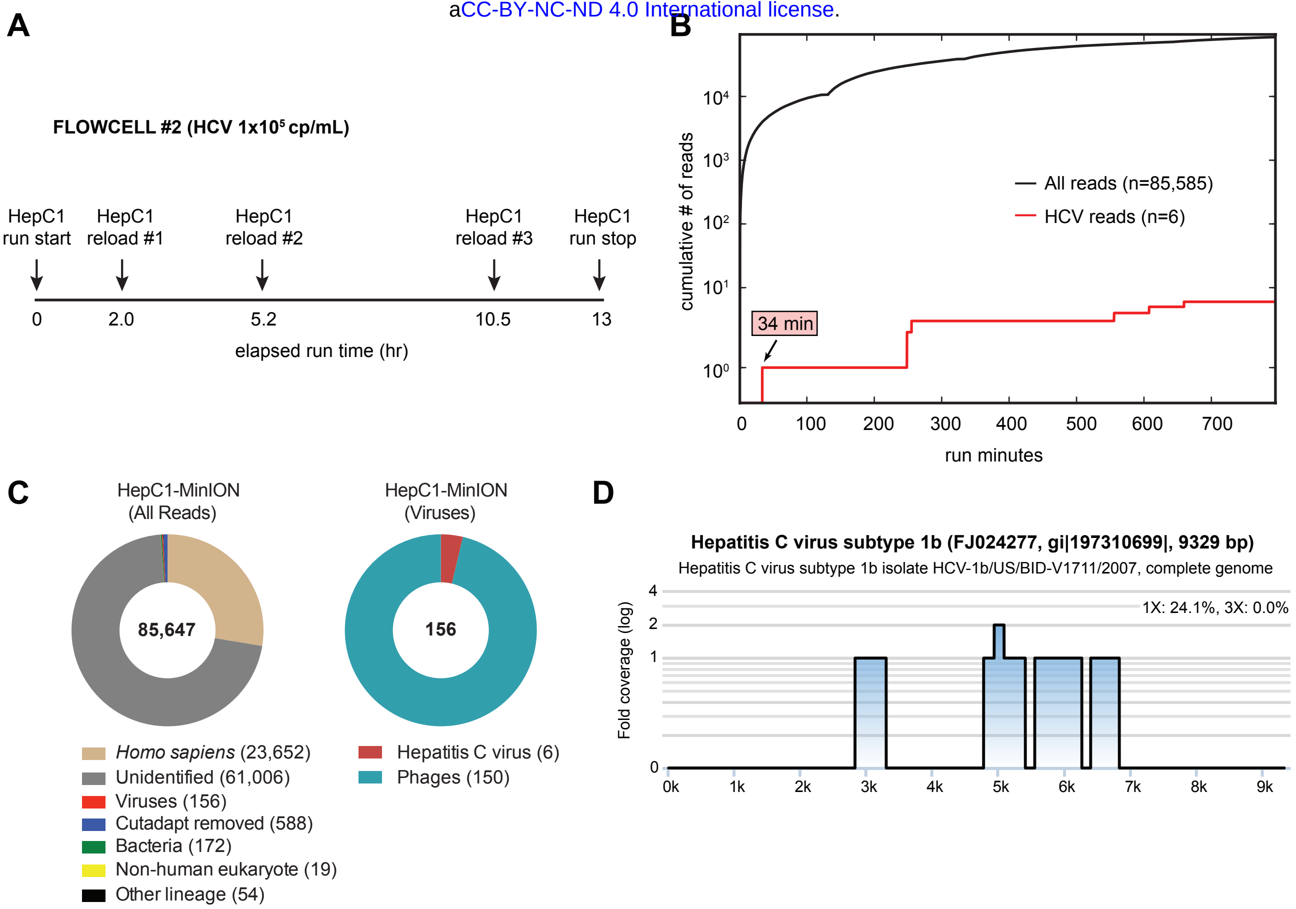
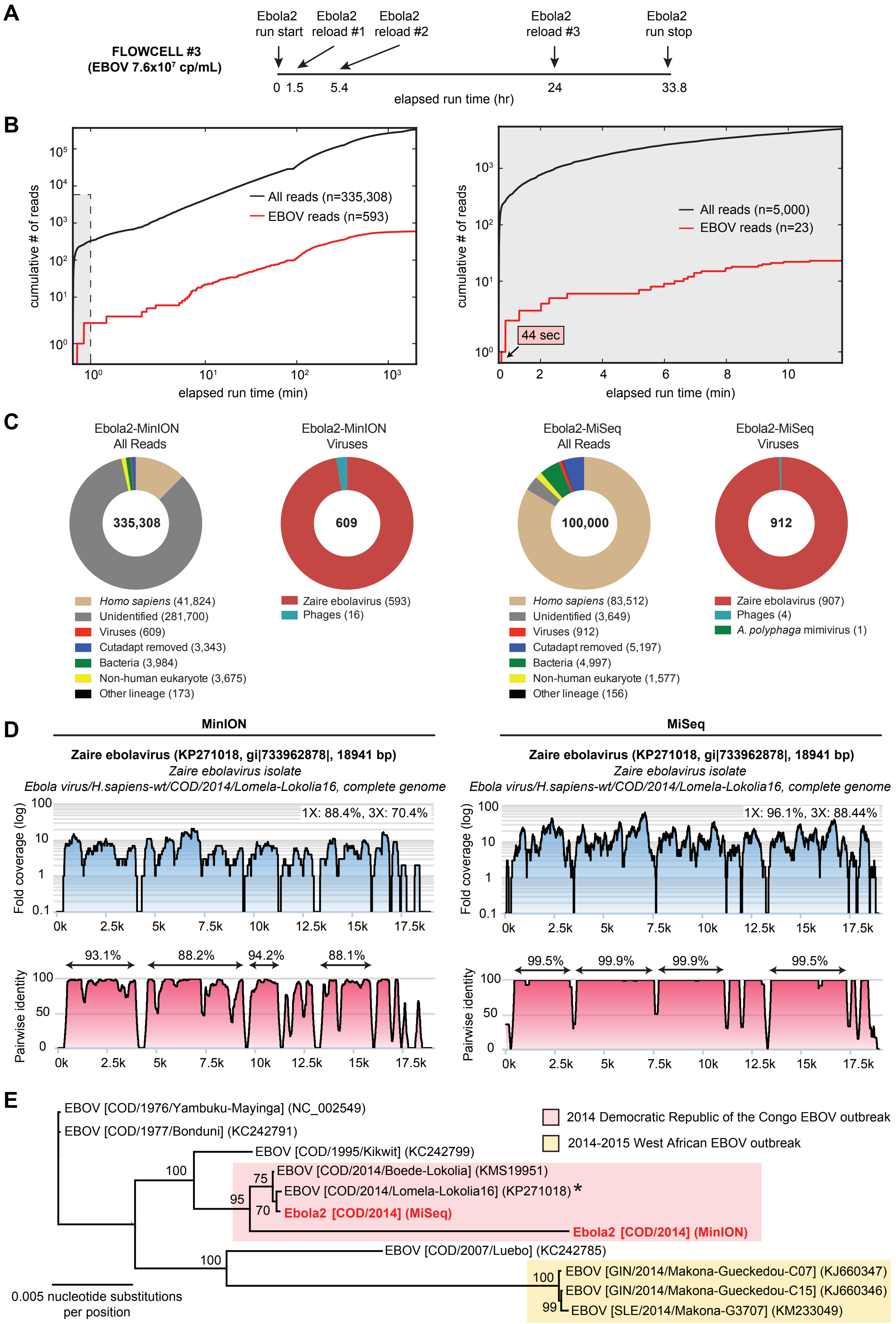
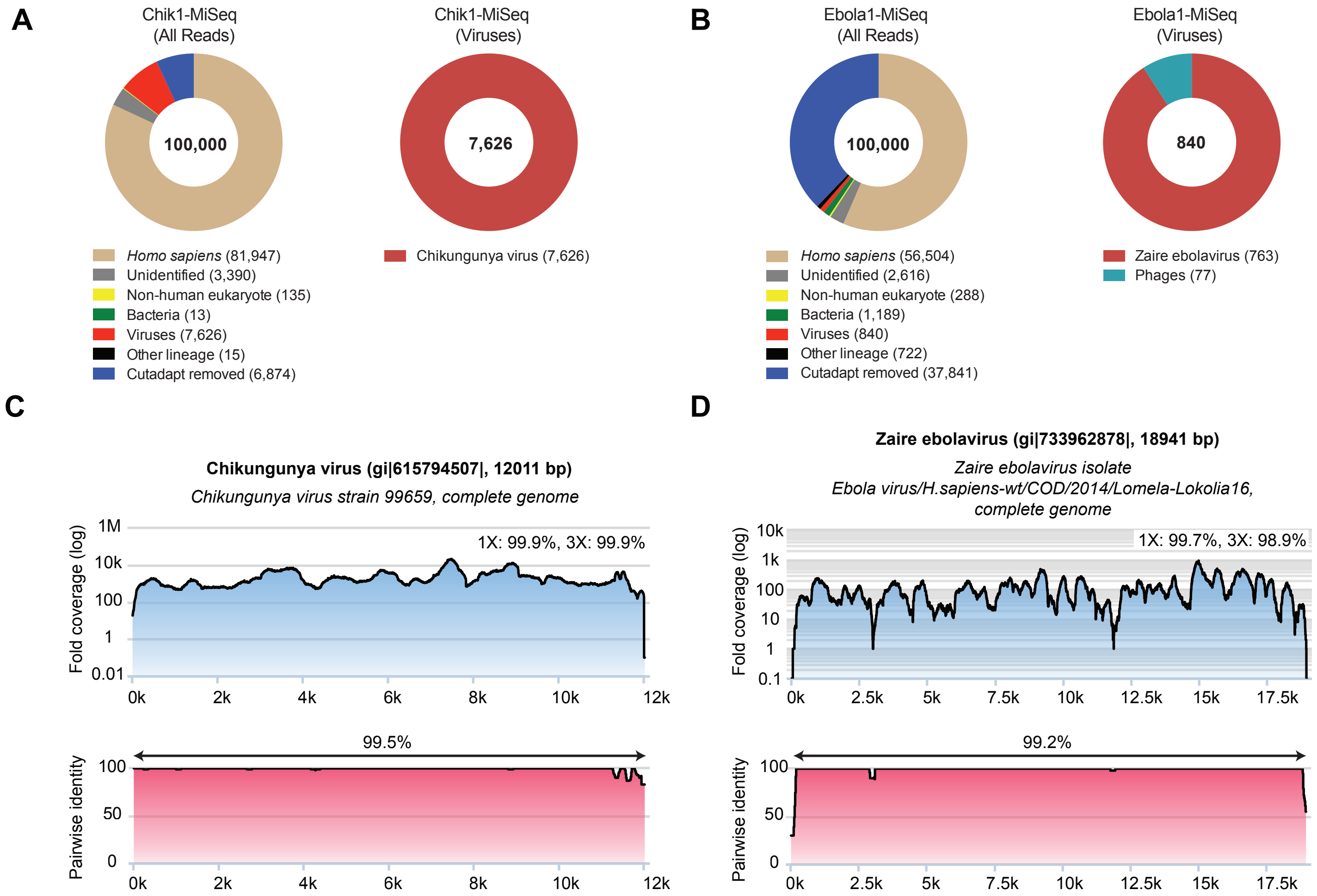

A
FLOWCELL #2 (HCV 1x10⁵ cp/mL)
HepC1 run start
HepC1 reload #1
HepC1 reload #2
HepC1 reload #3
HepC1 run stop
0
2.0
5.2
10.5
13
elapsed run time (hr)
B
cumulative # of reads
10⁴
10³
10²
10¹
10⁰
All reads (n=85,585)
HCV reads (n=6)
34 min
0
100
200
300
400
500
600
700
run minutes
C
HepC1-MinION (All Reads)
85,647
HepC1-MinION (Viruses)
156
Homo sapiens (23,652)
Unidentified (61,006)
Viruses (156)
Cutadapt removed (588)
Bacteria (172)
Non-human eukaryote (19)
Other lineage (54)
Hepatitis C virus (6)
Phages (150)
D
Hepatitis C virus subtype 1b (FJ024277, gi|197310699|, 9329 bp)
Hepatitis C virus subtype 1b isolate HCV-1b/US/BID-V1711/2007, complete genome
Fold coverage (log)
4
2
1
0
1X: 24.1%, 3X: 0.0%
0k
1k
2k
3k
4k
5k
6k
7k
8k
9k

Figure 4


A
FLOWCELL #3
(EBOV 7.6x10^7 cp/mL)
Ebola2 run start
Ebola2 reload #1
Ebola2 reload #2
Ebola2 reload #3
Ebola2 run stop
0 1.5 5.4 24 33.8
elapsed run time (hr)
B
cumulative # of reads
All reads (n=335,308)
EBOV reads (n=593)
elapsed run time (min)
All reads (n=5,000)
EBOV reads (n=23)
44 sec
C
Ebola2-MinION All Reads
335,308
Ebola2-MinION Viruses
609
Ebola2-MiSeq All Reads
100,000
Ebola2-MiSeq Viruses
912
Homo sapiens (41,824)
Unidentified (281,700)
Viruses (609)
Cutadapt removed (3,343)
Bacteria (3,984)
Non-human eukaryote (3,675)
Other lineage (173)
Zaire ebolavirus (593)
Phages (16)
Homo sapiens (83,512)
Unidentified (3,649)
Viruses (912)
Cutadapt removed (5,197)
Bacteria (4,997)
Non-human eukaryote (1,577)
Other lineage (156)
Zaire ebolavirus (907)
Phages (4)
A. polyphaga mimivirus (1)
D
MinION
MiSeq
Zaire ebolavirus (KP271018, gi|733962878|, 18941 bp)
Zaire ebolavirus isolate
Ebola virus/H.sapiens-wt/COD/2014/Lomela-Lokolia16, complete genome
Fold coverage (log)
1X: 88.4%, 3X: 70.4%
1X: 96.1%, 3X: 88.44%
Pairwise identity
93.1% 88.2% 94.2% 88.1%
99.5% 99.9% 99.9% 99.5%
0k 2.5k 5k 7.5k 10k 12.5k 15k 17.5k
E
EBOV [COD/1976/Yambuku-Mayinga] (NC_002549)
EBOV [COD/1977/Bonduni] (KC242791)
EBOV [COD/1995/Kikwit] (KC242799)
EBOV [COD/2014/Boede-Lokolia] (KMS19951)
EBOV [COD/2014/Lomela-Lokolia16] (KP271018) *
Ebola2 [COD/2014] (MiSeq)
Ebola2 [COD/2014] (MinION)
EBOV [COD/2007/Luebo] (KC242785)
EBOV [GIN/2014/Makona-Gueckedou-C07] (KJ660347)
EBOV [GIN/2014/Makona-Gueckedou-C15] (KJ660346)
EBOV [SLE/2014/Makona-G3707] (KM233049)
100 75 95 70 100 100 99
2014 Democratic Republic of the Congo EBOV outbreak
2014-2015 West African EBOV outbreak
0.005 nucleotide substitutions per position

Supp Figure 1


A
Chik1-MiSeq (All Reads)
100,000
Chik1-MiSeq (Viruses)
7,626
Homo sapiens (81,947)
Unidentified (3,390)
Non-human eukaryote (135)
Bacteria (13)
Viruses (7,626)
Other lineage (15)
Cutadapt removed (6,874)
Chikungunya virus (7,626)
B
Ebola1-MiSeq (All Reads)
100,000
Ebola1-MiSeq (Viruses)
840
Homo sapiens (56,504)
Unidentified (2,616)
Non-human eukaryote (288)
Bacteria (1,189)
Viruses (840)
Other lineage (722)
Cutadapt removed (37,841)
Zaire ebolavirus (763)
Phages (77)
C
Chikungunya virus (gi|615794507|, 12011 bp)
Chikungunya virus strain 99659, complete genome
Fold coverage (log)
1M
10k
100
1
0.01
1X: 99.9%, 3X: 99.9%
0k
2k
4k
6k
8k
10k
12k
99.5%
Pairwise identity
100
50
0
D
Zaire ebolavirus (gi|733962878|, 18941 bp)
Zaire ebolavirus isolate Ebola virus/H.sapiens-wt/COD/2014/Lomela-Lokolia16, complete genome
Fold coverage (log)
10k
1k
100
10
1
0.1
1X: 99.7%, 3X: 98.9%
0k
2.5k
5k
7.5k
10k
12.5k
15k
17.5k
99.2%
Pairwise identity
100
50
0