

## Portable Nanopore Sequencing for Viral Surveillance

Carl T. Wittwer\*

The recent *Nature* article (1) entitled, “Real-Time, Portable Genome Sequencing for Ebola Surveillance,” reports nanopore sequencing of the Ebola virus from 142 subjects near the end of the outbreak in West Africa from March to October 2015. The nanopore sequencer (MinION, Oxford Nanopore Technologies) has a mass of 87 g and plugs into a laptop computer by way of a USB 3.0 interface and cable. The consumable flow cell has up to 512 channels available for sequencing, each of which can simultaneously read single strands of DNA and their ligated complementary strands as “2D” reads, providing correlated data on both strands of DNA. The nanopores are typically made up of proteins like  $\alpha$ -hemolysin that are pore-forming bacterial toxins embedded in an electrically insulating membrane. A voltage between the 2 sides of the membrane attracts DNA to the pore in an electrolyte solution. The DNA is captured by a polymerase near the pore that paces the rate of transport to 30–280 bases/s. The passage of single-stranded DNA affects the pore current, generating electrical “squiggle signals” that are sequence-specific. In simple terms, this is the Coulter principle applied to molecules rather than cells. In half of the cases, sufficient reads (25 2D reads for each amplicon) were obtained in <1 h of instrument time with a mean passing rate of 36.3%. The mean read rate was 138 amplicons/min, and in some cases, only 15 min of sequencing was required. Custom offline acquisition software was provided by the company so that Internet communication was not initially required. Nanopore sequencing is conceptually elegant and data acquisition is fast. However. . .

Before sequencing, sample transport and preparation are required. To prevent delays of air transport of samples, a portable 50-kg laboratory was transported in airline luggage to Guinea. As samples became available locally, they were transported to the portable laboratory. For each sample, RNA was extracted and reverse-transcriptase PCR performed by standard methods, including 3.5 h of thermal cycling. Typically, 11 amplicons (mean length 1765 bp) covering 96.8% of the 18959-bp Ebola ge-

nome were used for library preparation. Library generation included fluorescence quantification, equimolar pooling, end repair, 3'-dA tailing, adaptor ligation with linkage of complementary PCR strands for 2D reads, final purification, and quantification. Excluding PCR, the presequencing process required 8 different commercial kits and 2–3 h (estimated from kit user's guides) of manual processing. Including PCR, the minimum preparation time was about 6 h from receipt of a sample at the portable laboratory. Priming, calibration, and quality control of the sequencer can be performed in parallel to the sample preparation steps. Each sample was processed individually to minimize the time to result.

After data acquisition, the squiggle signals need to be converted to a base sequence. The bioinformatics were complicated enough that the data required uploading by the Internet to the Cloud and downloading for central processing in the UK. Because the nanopore sequencer does not measure single bases, the highest probability of 6-mer subsequences was determined by hidden Markov modeling, a method enabled by input from both the authors and the company. Indeed, the data pipeline developed appears to be a fine example of successful academic/industrial collaboration. The pipeline processing time depended on the availability and speed of the Internet. In one case, the sample to result interval, including remote bioinformatics, was <24 h, although the protocol was more usually performed in 2 working days. The authors conclude that real-time, remote genomic surveillance can be established rapidly in resource-limited settings to monitor infectious disease outbreaks.

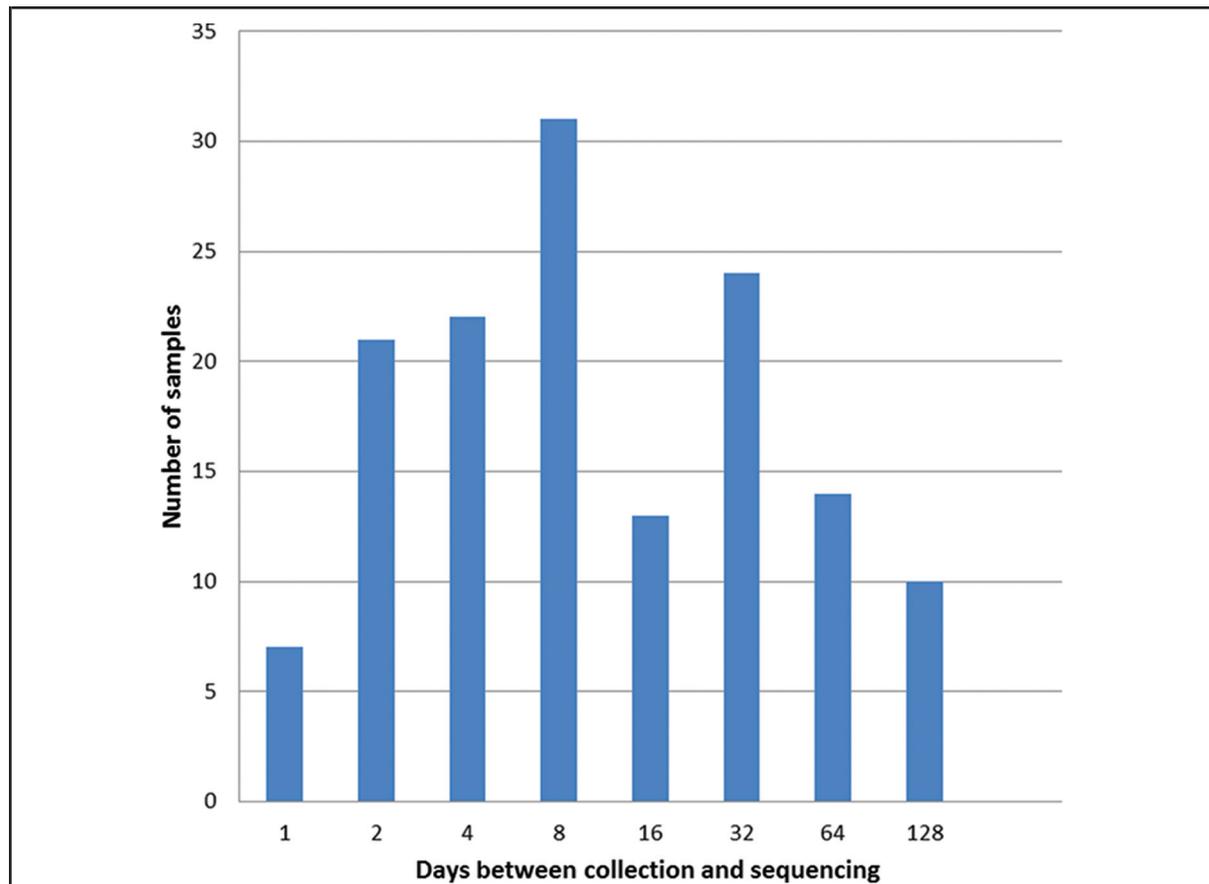
Several limitations of the sequencing system and study are apparent and many are discussed by the authors of the *Nature* article. The error rate of single nanopore reads is high, often exceeding 10% and making multiple reads necessary for consensus averaging, typically 25 2D reads. Insertions and deletions, as well as single-base runs of  $\geq 5$  bases are not accurately determined and their identification was not attempted. Some of the viral DNA (1.2%) was not analyzed because of primer blind spots, an oversight that could have been remedied by alternative primer placement. Another 2% of viral DNA at the 3' end was not sequenced. Some of the samples did not amplify well and required 17 shorter amplicons, presumably because of sample degradation. Twelve percent of samples (17/142) were excluded because of poor quality and/or low sensitivity.

Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, UT.  
\* Address correspondence to the author at: Department of Pathology, University of Utah Medical School, 50 N Medical Dr., Salt Lake City, UT 84132; Fax 801-581-6001; e-mail carl.wittwer@path.utah.edu.

Received May 9, 2016; accepted June 23, 2016.

Previously published online at DOI: 10.1373/clinchem.2016.256693

© 2016 American Association for Clinical Chemistry



**Fig. 1.** The number of days between sample collection and sequencing of 142 Ebola samples from West Africa between March and October 2015.

The time required for offsite analysis (the sequencing bioinformatics pipeline) is not included. The histogram cells are on a log scale in order to include all data. For example, the histogram cell labeled 16 includes values from 9 to 16. Some of the 142 samples with large intervals were retrospectively sequenced to provide additional epidemiological context.

As originally published, some of the sequencing dates between Supplementary Tables 2 and 4 in the *Nature* article were not consistent. Communication with the authors identified Supplementary Table 2 in the *Nature* article as correct. Using these data, the times between collection date and sequencing date are plotted in Fig. 1. Although 7 samples (5%) were processed in 1 day and 21 (15%) in 2 days, the median time was 6 days with a mean of 18.5 days. If the bioinformatics processing times were included, the intervals would be even longer. Viewed in this way, the data appears less “real-time” than the impression one gets from reading the article. Much of the delay is likely sample transport time and/or the fact that some of the samples were retrospectively sequenced. Bringing the laboratory to an outbreak is a great option, but outbreaks may cover a substantial geographical area, and sample transport is often slow in resource-limited areas. The authors accurately describe additional chal-

lenges for portable laboratories. Items best obtained on-site (rather than transported by air) include ethanol and an uninterrupted power supply for the PCR thermal cycler. Adequate Internet bandwidth for data upload was also a persistent problem. Eventually, with continued advances in bioinformatics processing, onsite analysis can be anticipated to eliminate this bottleneck.

Not including sample transport, Internet data transfer, and bioinformatics analysis, nanopore sequencing as described required about 2.5 h of manual sample processing (RNA isolation and library preparation), 3.5 h of reverse transcriptase PCR, and 1 h to generate the electronic sequencing data. Many companies are working to automate sample processing before massively parallel sequencing. The first instruments to appear will likely handle many samples in parallel without much attention to the turnaround time, an acceptable match for instruments with slow sequence acquisition. However, when

data generation can occur in 15–60 min, as in nanopore sequencing, rapid sample processing becomes much more important and need only process single samples as they are received, a simpler goal for automation. The 3.5 h of reverse transcriptase PCR is an even larger bottleneck that is slow by today's standards. Thirty-five cycles of PCR can be performed in <15 s, albeit for a 60-bp product (2). The amplification of 2000 bp products should take <15 s/cycle with a rapid polymerase (Kapa 2G Fast, Kapa Biosystems) completing 45 cycles in 11.25 min. The required time for reverse transcription is estimated at 10–20 nucleotides/s (3), requiring about 3 min for 2000 bp, so the entire reverse transcriptase PCR should require <15 min if the temperature control of extreme PCR (2) were available in commercial systems.

Real-time analysis is used in molecular biology for DNA amplification that is continuously monitored and in nanopore sequence acquisition. In both cases, data can be analyzed during the process and provide feedback that may change the process, the simplest example being termination of data collection. Real-time science has also found its way into circulating tumor cell analysis, because multiple samples of peripheral blood can monitor the evolution of tumor progression over time (4). Similarly, real-time epidemic surveillance by nanopore sequencing can monitor disease evolution by establishing genetic links to previously infected individuals. Whereas in PCR and nanopore sequence acquisition, real-time monitoring takes place in minutes, tumor progression and epidemic surveillance typically occur over months to years. Nevertheless, the use of real-time science may be appropriate in both cases, because there is hope of strategic intervention to improve outcomes. If real-time science were defined as experimentation to publication in <4 months, "Real-time, portable genome sequencing for Ebola surveillance" makes the grade, because it was published online 95 days after the last sequencing run. The

112 authors, 6 co-first authors, and 49 contributing institutions completed writing, editing, and reviewing the manuscript for submission in 18 days after the last sequencing was performed. This is an awesome geoscientific accomplishment.

The constructive tension between centralized and distributed diagnostics continues. Centralized laboratories reduce cost by batching at the expense of turnaround time. Diagnostics close to the patient usually return results faster than reference laboratories but at increased cost. Perhaps the Ebola surveillance described here could have been performed at a commercial sequencing laboratory at lower overall cost, but this misses the vision and potential of tools like nanopore sequencing that will get faster and less expensive in the future. Already, molecular syndromic testing for 15–25 infectious agents in an hour is commonplace (5). Ultimately, the market and/or governments will decide how much rapid testing is worth. It depends on the situation.

**Author Contributions:** All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

**Authors' Disclosures or Potential Conflicts of Interest:** Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

**Employment or Leadership:** C.T. Wittwer, *Clinical Chemistry*, AACC.

**Consultant or Advisory Role:** None declared.

**Stock Ownership:** None declared.

**Honoraria:** None declared.

**Research Funding:** C.T. Wittwer, BioFire Diagnostics.

**Expert Testimony:** None declared.

**Patents:** C.T. Wittwer, patent number 20150118715.

## References

1. Quick JO, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;530:228–32.
2. Farrar JS, Wittwer CT. Extreme PCR: efficient and specific DNA amplification in 15–60 seconds. *Clin Chem* 2015; 61:145–53.
3. Whiting SH, Champoux JJ. Strand displacement synthesis capability of Moloney murine leukemia virus reverse transcriptase. *J Virol* 1994;68:4747–58.
4. De Luca F, Rotunno G, Salvianti F, Galardi F, Pestrin M, Gabelline S, et al. Mutational analysis of single circulating tumor cells by next generation sequencing in metastatic breast cancer. *Oncotarget* 2016;7:26107–19.
5. Buss SN, Leber A, Chapin K, Fey PD, Bankowski MJ, Jones MK, et al. Multicenter evaluation of the BioFire FilmArray gastrointestinal panel for etiologic diagnosis of infectious gastroenteritis. *J Clin Microbiol* 2015; 53:915–25.