Nanopore sequencing: a rapid solution for infectious disease epidemics

Ying Cao[1,2], Jing Li[3,5], Xin Chu[4], Haizhou Liu[1,2], Wenjun Liu[3,5] & Di Liu[1,2,5*]

[1]Center for Emerging Infectious Diseases, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China;
[2]Research Group of Computer Virology, Center for Bacteria and Virus Resources and Application, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China;
[3]Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China;
[4]Bureau of Major R&D Programs, Chinese Academy of Sciences, Beijing 100864, China;
[5]University of Chinese Academy of Sciences, Beijing 100049, China

Emerging and re-emerging infectious diseases have given rise to a large number of human infections, morbidity, and heavy economic burden, including the Middle East respiratory syndrome caused by a coronavirus in 2012, global influenza pandemic caused by the H7N9 influenza A virus in 2013, Ebola epidemic in West Africa in 2014, and Lassa fever epidemic in Nigeria in 2019. The healthcare war against viruses deserves constant surveillance due to the continuous emergence of new viruses and rapid evolution of existing viruses (Liu and Liu, 2017).

Rapid identification of the causative agents of infectious diseases and access to their genomes are essential for establishing therapeutic and preventive measures. In addition, obtaining a high-quality genome sequence is essential for the analyses of phylogeny and genetic variation. Traditionally, pathogen identification and further research have utilized Sanger sequencing technology and/or next-generation sequencing technology (dominantly with Illumina® platforms), such as the phylogenetic analysis of H7N9 influenza virus isolated from patients in Guangzhou (Deng et al., 2017), which typically include amplification of target nucleic acids by PCR. While next-generation sequencing technology has been extensively used in pathogen identification, the short read lengths and GC bias (Pop and Salzberg, 2008) have limited its performance in subsequent genome assembly, annotation, and other bioinformatics analyses.

Third-generation sequencing (TGS) technology, including the Oxford Nanopore Technologies (ONT) platform and the Pacific Biosciences platform, have been widely applied in animal, plant, bacterial, and viral samples for the study of genome assembly, analysis of epigenetic markers, in transcriptomics, and in metagenomics (Mondal et al., 2018; Shin et al., 2018; Keller et al., 2018). Compared to Pacific Biosciences (PacBio) platforms, Oxford Nanopore platforms produce longer reads (as long as 2 Mb in length) and are handier, more affordable, and applicable in real-time detection of samples in epidemic areas with relatively poor sanitary conditions. Considering these advantages, the Oxford Nanopore platforms are much better suited for on-site sequencing and genetic analysis of epidemics.

The Nanopore sequencing platforms sequence DNA or RNA in one run. Nanopore DNA sequencing technology was developed first, with lower error rates (the per-nucleotide error rate is ~5%–15% (Luo et al., 2019)) and higher throughput than direct Nanopore RNA sequencing (the per-nucleotide error rate is approximately 17.02% (Harel et al., 2019)). During the Ebola outbreak in Guinea, Quick et al. designed an entire sequencing system based on the MinION

*Corresponding author (email: liud@wh.iov.cn)

platform of the ONT Company, which could be accommodated in a single ordinary test bed, for real-time genome monitoring of the epidemic (Quick et al., 2016). Even in an environment with resource-limited areas, it can produce results in 24 h or less, after receiving a positive Ebola sample, and the sequencing process takes only 15–60 min, providing a good example for epidemic surveillance. Both Illumina and Nanopore platforms were used in this study, and it was a preliminary demonstration of the broad application prospects of Nanopore platforms in epidemic research.

Another representative application of Nanopore was in the recent Lassa fever epidemic. Since January 2019, 327 cases (324 confirmed and three suspected) and 72 deaths from Lassa fever (caused by the Lassa fever virus, LASV) have been reported in just 40 days from 20 states and federal capital districts of Nigeria. Kafetzopoulou et al. used the MinION platform to sequence 120 LASV-positive samples, including plasma, breast milk, and cerebrospinal fluid (Kafetzopoulou et al., 2019). The sequencing results obtained by the MinION platform showed LASV to have a higher proportion of reads in each sequencing sample, with a maximum of 42.9% and an average frequency of 4.26%. The higher virus content allowed them to assemble enough genome sequences (>70%), making phylogenetic analysis of genomic fragments possible. Phylogenetic analysis of all recent LASV sequences, including the unpublished sequences from previous years and virus sequences available in GenBank, revealed diversification and correlation with the early strains of LASV. Use of the MinION platform in this epidemic allowed quick understanding of the same. Subsequently, the results of phylogenetic analysis were transmitted to Nigerian authorities and the World Health Organization, which alleviated public panic regarding human-to-human transmission in a short period of time.

In the above studies, the MinION platform was mostly used for sequencing DNA or reverse transcribed cDNA. Recently, the Nanopore sequencing technology has been found to also allow direct RNA sequencing without the need for reverse transcription or PCR amplification, which enables the study of viral genome in its original state (Garalde et al., 2018). Since the description of direct RNA sequencing method, it has been used in the study of genomes of influenza virus, coronaviruses, and many more (Keller et al., 2018; Viehweger et al., 2018). Keller et al. modified the direct RNA sequencing technology of MinION platform, replacing the oligo T sequence in the adaptor with a universal primer for influenza virus to enrich it in the samples (Keller et al., 2018). Sequencing experiments successfully obtained an influenza viral genome with genome coverage of 100% and a sequencing depth of 1,377 to 5,161. The successful practice of enriching target sequences, using genomic features to design linker sequences, provides a reference for the study of other viruses. Recently, Wongsurawat et al. have used a

mixture of negative and positive single-stranded (ss) RNA viruses as an example, including Mayaro virus, Chikungunya virus, and Zika virus, to complete the sequencing of multiple ssRNA viruses in a single reaction (Wongsurawat et al., 2019). In this study, the researchers obtained sequences of the same length as reference genomes, with identity up to 97%. Results of this laboratory test demonstrated the possibility of applying direct RNA sequencing to detect viruses during epidemics.

Nanopore sequencing technology is used as an emerging and promising technology for viral transcriptomes, structural variation, and genome re-sequencing studies (Depledge et al., 2019; Gallagher et al., 2018). Due to the limitation of read length in the sequencing technology, full-length genome of the virus cannot be included in one read. Hence, quasispecies are often used as research units to study viral phylogeny and genetic variation. Virus quasi-species generally refer to highly correlated, yet not identical, and extremely large dynamic population. Therefore, use of a consensus sequence to represent all information of a quasi-species is inaccurate. Although the Oxford Nanopore platforms have the potential, they are not ready for the genome study of different individuals in a virus population owing to their poor single-base accuracy.

At present, single-base accuracy of the Nanopore sequencer is approximately 85% (Cretu Stancu et al., 2017), and accuracy of the corrected consensus sequence is 97% (Jain et al., 2018). For organisms such as animals or plants with large genomes and relatively low variability, increasing sequencing depth can reduce the impact of error rates on the results of the study. However, impact of high error rate on viruses, which have small genomes with high mutation rates, is not negligible, and this limits the use of Nanopore technology in virus research. Oxford Nanopore Technologies Limited and many other researchers have been working to improve the accuracy of sequencing results using the Nanopore equipment, reagents, and subsequent analytical algorithms.

It is foreseeable that on-site detection and diagnosis of emerging and re-emerging infectious diseases, based on Nanopore sequencing technology, will be a general trend in the future. However, there are still some issues that need to be addressed. First, since the samples are collected and detected on site, the proportion of pathogens tends to be low in the samples. Compared to the NGS platform, Minion sequencing has a greater impact on the pathogen content in the sample because there is no bridge PCR amplification process. At the same time, more host information will also affect subsequent bioinformatics analysis. Therefore, a method for targeted enrichment of gene sequences is required from the aspect of sample processing, gene extraction, and genomic library construction. In addition to the proportion of pathogens, higher sequencing error rates also make it difficult to

use MinION platform, especially direct RNA sequencing platform) to detect unknown pathogens. While efforts have been made to upgrade the Nanopore technology, new algorithms are still required to address the error rate of the Nanopore sequencing platforms from base calling and subsequent polish, assembly, and other aspects. With continuous improvement of sequencing quality and read length of the Nanopore sequencing platforms, it will be increasingly useful in the treatment of epidemics in the future.

References

Cretu Stancu, M., van Roosmalen, M.J., Renkens, I., Nieboer, M.M., Middelkamp, S., de Ligt, J., Pregno, G., Giachino, D., Mandrile, G., Espejo Valle-Inclan, J., et al. (2017). Mapping and phasing of structural variation in patient genomes using nanopore sequencing. Nat Commun 8, 1326.

Deng, Y., Li, C., Han, J., Wen, Y., Wang, J., Hong, W., Li, X., Liu, Z., Ye, Q., Li, J., et al. (2017). Phylogenetic and genetic characterization of a 2017 clinical isolate of H7N9 virus in Guangzhou, China during the fifth epidemic wave. Sci China Life Sci 60, 1331–1339.

Depledge, D.P., Srinivas, K.P., Sadaoka, T., Bready, D., Mori, Y., Placantonakis, D.G., Mohr, I., and Wilson, A.C. (2019). Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. Nat Commun 10, 754.

Gallagher, M.D., Matejusova, I., Nguyen, L., Ruane, N.M., Falk, K., and Macqueen, D.J. (2018). Nanopore sequencing for rapid diagnostics of salmonid RNA viruses. Sci Rep 8, 16307.

Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods 15, 201–206.

Harel, N., Meir, M., Gophna U., and Stern, A. (2019). Sequencing complete genomes of RNA viruses with minION nanopore: finding associations between mutations. bioRxiv 575480, https://doi.org/10.1101/575480.

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 36, 338–345.

Kafetzopoulou, L.E., Pullan, S.T., Lemey, P., Suchard, M.A., Ehichioya, D. U., Pahlmann, M., Thielebein, A., Hinzmann, J., Oestereich, L., Wozniak, D.M., et al. (2019). Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. Science 363, 74–77.

Keller, M.W., Rambo-Martin, B.L., Wilson, M.M., Ridenour, C.A., Shepard, S.S., Stark, T.J., Neuhaus, E.B., Dugan, V.G., Wentworth, D. E., and Barnes, J.R. (2018). Direct RNA sequencing of the coding complete influenza A virus genome. Sci Rep 8, 14408.

Liu, W.J., and Liu, D. (2017). The triphibious warfare against viruses. Sci China Life Sci 60, 1295–1298.

Luo, R., Sedlazeck, F.J., Lam, T.W., and Schatz, M.C. (2019). A multi-task convolutional deep neural network for variant calling in single molecule sequencing. Nat Commun 10, 998.

Mondal, T.K., Rawal, H.C., Chowrasia, S., Varshney, D., Panda, A.K., Mazumdar, A., Kaur, H., Gaikwad, K., Sharma, T.R., and Singh, N.K. (2018). Draft genome sequence of first monocot-halophytic species *Oryza coarctata* reveals stress-specific genes. Sci Rep 8, 13698.

Pop, M., and Salzberg, S.L. (2008). Bioinformatics challenges of new sequencing technology. Trends Genet 24, 142–149.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A., et al. (2016). Real-time, portable genome sequencing for Ebola surveillance. Nature 530, 228–232.

Shin, H.S., Lee, E., Shin, J., Ko, S.R., Oh, H.S., Ahn, C.Y., Oh, H.M., Cho, B.K., and Cho, S. (2018). Elucidation of the bacterial communities associated with the harmful microalgae *Alexandrium tamarense* and *Cochlodinium polykrikoides* using nanopore sequencing. Sci Rep 8, 5323.

Wongsurawat, T., Jenjaroenpun, P., Taylor, M.K., Lee, J., Tolardo, A.L., Parvathareddy, J., Kandel, S., Wadley, T.D., Kaewnapan, B., Athipanyasilp, N., et al. (2019). Rapid sequencing of multiple RNA viruses in their native form. Front Microbiol 10, 260.

Viehweger, A., Krautwurst, S., Lamkiewicz, K., Madhugiri, R., Ziebuhr, J., Hölzer, M., and Marz, M. (2018). Nanopore direct RNA sequencing reveals modification in full-length coronavirus genomes. BioRxiv 483693, https://doi.org/10.1101/483693.