ORIGINAL ARTICLE

J Bacteriol Virol. Vol 49. No 3. September 2019; 49(3): 115–123
https://doi.org/10.4167/jbv.2019.49.3.115
eISSN 2093-0249

JBV

Rapid Whole-genome Sequencing of Zika Viruses using Direct RNA Sequencing

Jung Heon Kim[1§], Jiyeon Kim[1,2], Bon-Sang Koo[3], Hanseul Oh[3], Jung-Joo Hong[3] and Eung-Soo Hwang[1,2*]

[1]Institute of Endemic Diseases, Seoul National University Medical Research Center, Seoul, 03080, Korea
[2]Department of Microbiology and Immunology, Seoul National University College of Medicine, Seoul, 03080, Korea
[3]National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology, Cheongju, Korea

Corresponding
Eung-Soo Hwang, MD, PhD
Department of Microbiology and Immunology, Seoul National University College of Medicine, Seoul, 03080, Republic of Korea and Institute of Endemic Diseases, Seoul National University Medical Research Center, Seoul, 03080, Republic of Korea
Phone : +82-2-740-8307
Fax : +82-2-740-8331
E-mail : hesss@snu.ac.kr

§Jung Heon Kim and Jiyeon Kim contributed equally to this work

Received : August 18, 2019
Revised : August 28, 2019
Accepted : September 4, 2019

No potential conflict of interest relevant to this article was reported.

Zika virus (ZIKV) is one of the pathogens which is transmitted world widely, but there are no effective drugs and vaccines. Whole genome sequencing (WGS) of viruses could be applied to viral pathogen characterization, diagnosis, molecular surveillance, and even finding novel pathogens. We established an improved method using direct RNA sequencing with Nanopore technology to obtain WGS of ZIKV, after adding poly (A) tails to viral RNA. This established method does not require specific primers, complimentary DNA (cDNA) synthesis, and polymerase chain reaction (PCR)-based enrichment, resulting in the reduction of biases as well as of the ability to find novel RNA viruses. Nanopore technology also allows to read long sequences. It makes WGS easier and faster with long-read assembly. In this study, we obtained WGS of two strains of ZIKV following the established protocol. The sequenced reads resulted in 99% and 100% genome coverage with 63.5X and 21,136X, for the ZIKV PRVABC59 and MR 766 strains, respectively. The sequence identities of the ZIKV PRVABC59 and MR 766 strains for each reference genomes were 98.76% and 99.72%, respectively. We also found that the maximum length of reads was 10,311 bp which is almost the whole genome size of ZIKV. These long-reads could make overall structure of whole genome easily, and WGS faster and easier. The protocol in this study could provide rapid and efficient WGS that could be applied to study the biology of RNA viruses including identification, characterization, and global surveillance.

Key Words: Whole genome sequencing (WGS), Zika virus (ZIKV), Direct RNA sequencing

INTRODUCTION

Zika virus (ZIKV) is a member of the family Flaviviridae and was originally isolated in the Zika forest of Uganda (1, 2). ZIKV is transmitted by various mosquitoes including *Aedes aegypti* and *Aedes albopictus* (3). The symptom of ZIKV infection typically last for one week and is similar to those of dengue or chikungunya viral infection such as conjunctivitis, rash, malaise, headache, muscle and joint pain, and fever (4, 5). Similar to maternal cytomegalovirus infection (6), congenital infection of ZIKV can cause fetus microcephaly, malformations, and other complications such as preterm birth and miscarriage (7, 8). Neurologic complications of ZIKV infection have been reported in adults and children in the Pacific and the Americas, including Guillain-Barré syndrome, neuropathy and myelitis (9, 10). Because there is no vaccine for ZIKV, thus far, understanding characteristics and pathogenesis of the virus is necessary to develop effective

diagnosis, treatment, and prevention.

Recently, whole genome sequencing (WGS) of viruses has become an efficient tool for pathogen characterization, diagnosis, molecular surveillance, and even finding new pathogens (11-13). Next generation sequencing (NGS) is one of the most powerful techniques for WGS. Complimentary DNA (cDNA) synthesis and enrichment of the target viruses for short-read NGS could be carried out by polymerase chain reaction (PCR) with random hexamers or specific primers (14). It is especially difficult to design and optimize specific primers for RNA viruses because they have highly variable sequences depending on their various strains, resulting in many biases and artifacts (15). Nanopore technology is long-read NGS from Oxford. Many studies have established WGS based on amplification using the long-read Nanopore technology for various RNA viruses including Ebola virus, ZIKV, yellow fever virus, and West Nile virus compare to the short-read NGS (16-18). Even Nanopore technology could not avoid amplification biases since it also used amplicons. Recently, several studies have tried to use direct RNA sequencing for the rapid detection of multiple RNA viruses (19) as well as WGS for RNA viruses (16). This method using direct RNA sequencing with Nanopore technology for WGS of single-stranded RNA viruses should be modified during the RNA library preparation steps, because there are no poly (A) tails on its RNA.

In this study, we established an improved method to obtain WGS of single-stranded RNA viruses using ZIKV with direct RNA sequencing of Nanopore technology. Our modified method combines 1) purification and concentration of viruses, 2) human ribosomal RNA (rRNA) depletion, 3) poly (A) tailing, 4) direct RNA sequencing by Nanopore technology, and 5) analysis of sequencing with gene alignment software to make consensus sequences. We compared the obtained two strains of ZIKVs (PRVABC59 and MR 766) with those from GenBank and confirmed the accuracy of the WGS.

MATERIALS AND METHODS

Cells and Viruses cultivation

Vero E6 (ATCC CRL-1586) cells were maintained in Dulbecco's minimal essential medium (DMEM; Hyclone) supplemented with 10% fetal bovine serum (FBS; Hyclone) and 1% penicillin/streptomycin (Gibco). Vero E6 cells were infected with ZIKV MR 766 strain (ATCC VR-1838) and ZIKV PRVABC59 (ATCC VR-1843) with 0.1 multiplicity of infection (MOI). The supernatant was harvested after 5days post infection (dpi). Viruses derived from the same batch were used for all experiments. Viral titers were measured using a conventional plaque assay and tissue culture infective dose 50 (TCID50) assay; Virus copy numbers were determined by revers transcription (RT)-quantitative polymerase chain reaction (qPCR).

Virus purification, RNA extraction, and cDNA synthesis

ZIKVs were purified by ultra-centrifugation (Beckman Coulter) on a 20% sucrose cushion for 3 h at 100,000 x g with SW28. Viral RNA was extracted using the Viral RNA Mini Kit (Qiagen) according to the manufacture's protocol. For RT-qPCR, cDNA synthesis was performed using the SuperScript IV First-Strand Synthesis System (Life Technologies), following the manufacture's protocol with random hexamers.

Quantification of viruses by Reverse Transcription (RT)-PCR

For the primer and probe design, NS5 gene sequences of ZIKV NS5(MR 766 and PRVABC59 strain) were obtained from NCBI GenBank database. The primers and probe were designed using primer express 3.0.1 (Thermo Fisher Scientific). Real-time PCRs were performed using a Taqman MasterMix (Applied Biosystems) containing 10 pmol forward (5-CCAGGAGGAARGATGTAYGCA-3) and reverse (5-TTCTCCAGATCAAACYTRCTAA TG-3) primers, 10 pmol Taqman probe (FAM-5-ATGACACTGCTGGCTGGGACACCC-3-TAMRA), and varying concentrations of ZIKV cDNA with ABI Quant studio 5 (Thermo Fisher Scientific). ZIKV-specific primers and Taqman probes were synthesized commercially (Bioneer). Real-time PCR

was performed under standard reaction conditions, as recommended by Applied Biosystems; 40 cycles of denaturation for 10 sec at 95 °C and annealing and extension for 60 sec at 60 °C.

Ribosomal RNA (rRNA) depletion

Human rRNA was depleted from the samples using NEBNext rRNA depletion kit (NEB) according to the manufacturer's instructions. Briefly, rRNA was removed from samples following treatment with rRNA removal solution. DNA was also removed from samples following treatment with Dnase I. rRNA and DNA-depleted samples were purified with magnetic beads (NEB).

Quantification, purification, and concentration of RNA

RNA was quantified using a Nanodrop Spectrophotometer (ND-1000, Thermo Fisher Scientific). AMPureXP beads (Beckman Coulter) was used for purification, and concentration of RNA following the manufacturer's instructions. Briefly, 1.8X volume of AMPureXP beads was added to RNA and incubated for 5 min. The sample was placed onto the magnetic rack and the supernatant was discarded. RNA was washed with fresh 80% ethanol, and eluted in nuclease-free water (Sigma).

Poly (A) tailing of RNA

After rRNA depletion, purified and concentrated RNA was modified by Poly(A)-tailing reaction using *E. coli* Poly(A) Polymerase (EPAP) (NEB) according to the manufacturer's protocol. Briefly, purified RNA was treated with 2 μl 10X *E. coli* Poly(A) Polymerase Reaction Buffer, 2 μl ATP (10 mM), and 1 μl EPAP at 37 °C for 30 min. Poly(A)-tailed RNA was then purified and eluted using AMPureXP beads (Beckman Coulter) as described above. The cleaned Poly(A)-tailed RNA was used as input for RNA library preparation.

Library preparation and direct RNA sequencing using Nanopore technology

Library preparation and direct RNA sequencing was performed according to the Direct RNA Sequencing Protocol using SQK-RNA002 kit (ONT). R9.4/FLO-MIN106 flow cell (ONT) and MinION (ONT) were used for sequencing of the RNA.

Data analysis

Base-calling was performed using Albacore version 2.3.4 (ONT) and EPI2ME version 2019.7.9 (ONT). The sequenced reads were modified using Porechop version 0.2.4 to remove adapter sequences and mapped to reference genomes (ZIKV MR 766 and PRVABC59 strains) to make the consensus sequences and calculate the coverage using Geneious Prime version 2019.0.4 (Geneious).

RESULTS

To establish a method of WGS for RNA viruses using direct RNA sequencing, we performed the following four procedures with the ZIKV PRVABC59 strain: 1) purification of the virus, 2) library preparation of RNA, 3) sequencing with Nanopore technology, and 4) data analysis for WGS. Then, we applied the established protocol to another ZIKV MR766 strain.

The first run was performed following the procedure. In the first procedure, 10ml supernatant of PRVAB 59 (4.2 x10^6 pfu/ml) was centrifuged at 2,000 x g, at 4 °C for 30 min to remove cell debris. The supernatant was ultra-centrifuged to purify the viruses. rRNA depletion was then performed to remove human rRNA. In the second procedure, poly (A) tailing was performed to add poly (A) tails at the end of the viral RNAs. Because direct RNA sequencing using Nanopore technology

requires poly (A) tails at the end of RNA, this step is necessary to perform WGS using Nanopore technology. Then, RNA library preparation was achieved to ligase adapters on the viral RNAs. In the third procedure, sequencing was performed to read long sequences. In the last procedure, base-calling and analysis of data were performed to assemble reads and make consensus sequences.

We obtained only 407 reads of ZIKV sequences with 94.32% of sequence identity and 98% of coverage with 9.6X from the first run (Table 1 and Fig. 1A). Although 407 reads were not enough to cover full genomes, we could make the consensus sequence because of long-reads with Nanopore technology. Human rRNA was removed successfully with centrifugation and rRNA depletion steps. But poly (A) tailing was performed to non-viral microorganisms as well as ZIKV, resulting in few reads of ZIKVs (Fig. 2A).

To improve efficiency of adding poly (A) tails only to the viruses, the second run was performed following procedure above except for the centrifugation step. The centrifugation in the second run was performed at 17,000 x g, at 4 °C for 20 min to remove cell debris as well as other microorganism such as bacteria and fungi. The analyzed data is shown in Table 1, resulting in 2,688 reads of ZIKV sequences with 98.76% of sequence identity and 99% of genome coverage with 63.5X (Table 1 and Fig. 1B). Accuracy and coverage were improved in the second run, suggesting high-speed centrifugation is important step to enhance efficiency. Although the ZIKV portion of total reads in the second run was increased compared to the first run, it should be increased even more to improve accuracy and coverage. (Fig. 2B),

Next, the ZIKV MR766 strain was performed following this protocol. In this third run, 60ml of high tittered virus supernatant was used ($3.0x10^7$ pfu/ml) to increase amount of viral RNAs. The analyzed data is shown in Table 1, resulting in 253,313 reads of ZIKV sequences with 99.73% of sequence identity and 100% of genome coverage with 21,136X (Table1 and Fig. 1C). Accuracy and coverage were significantly improved in the third run and the ZIKV portion of total reads was also dramatically increased (Fig. 2C), suggesting sufficient virus is necessary to increase efficiency. Surprisingly, we found many long-reads in this third run. The maximum length of reads in this third run was 10,311 bp which is similar to the whole genome size (10,772 bp) of ZIKV (Fig. 3). These long-reads could make WGS faster and easier.

To confirm whether the step of high-speed centrifugation could affect the efficiency of viral isollation, reverse transcription quantitative PCR (RT-qPCR) was performed with 250 μl of harvested samples from each procedure; 1) the original viral supernatant, 2) the viral supernatant after high-speed centrifugation, 3) the pellet after high-speed centrifugation of the viral supernatant, and 4) the viral supernatant after ultra-centrifugation. About 3.0 X 10^7 copies of ZIKV were detected in the original viral supernatant. The detected ZIKV in the viral supernatant after high-speed centrifugation was similar to that in the original viral supernatant. In addition, few viruses were detected in the pellet after high-speed centrifugation, suggesting that the viruses were not lost in the high-speed centrifugation. The detected viruses were significantly reduced after ultra-centrifugation. These results showed that only ultra-centrifugation affected the concentration and isolation of the virus not high-speed centrifugation (Fig. 4).

Table 1. Sequence data after alignment

	# Read	Read length (bp)		Reads mapping zika ref. genome	Genome coverage (%)	Identity % vs ref.genome		Mean depth	Genome size (bp)
		Min	Max			MK713748.1	MK105975.1		
Zika PRVABC59_1	407	97	1,699	263	98	94.32		9.6X	10,471
Zika PRVABC59_2	2,688	47	2,905	1,587	99	98.76		63.5X	10,583
Zika MR766	253,313	26	10,311	253,313	100		99.72	21,136X	10,772

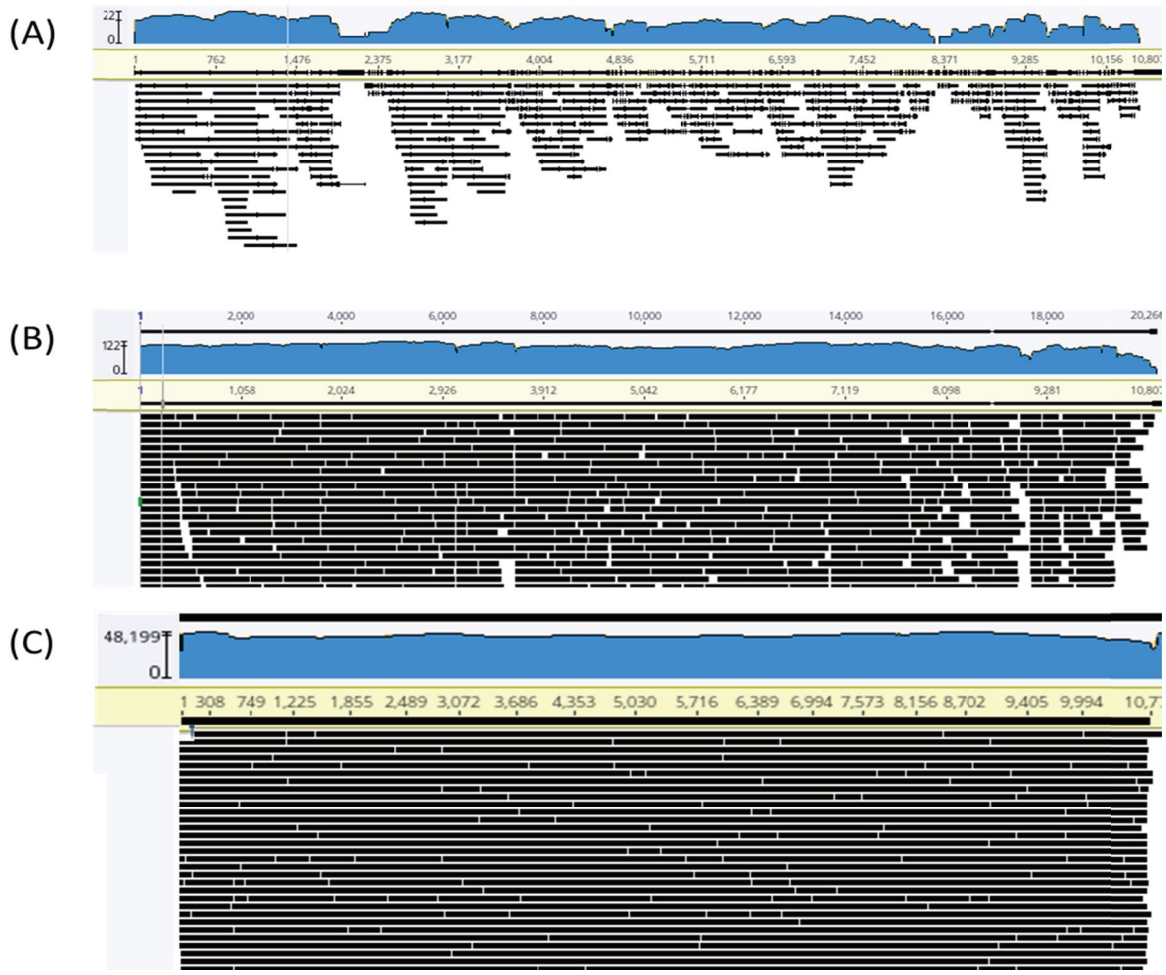

Figure 1. Sequence alignment display on the reference ZIKV genome with Geneious Prime for three runs. (A) Sequence alignment for ZIKV PRVABC59 strain from the first run. (B) Sequence alignment for ZIKV PRVABC59 strain from the second run.(C) Sequence alignment for ZIKV MR 766 strain from the third run.

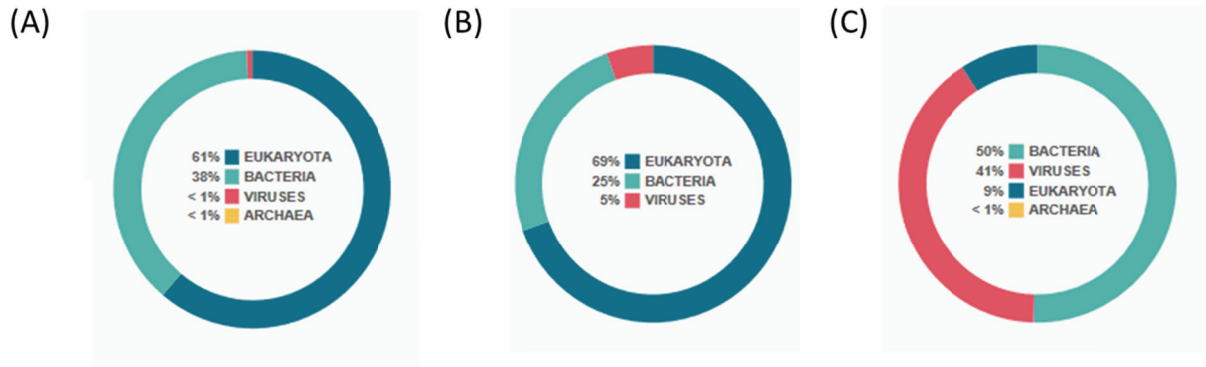

Figure 2. Analysis with EPI2ME. EPI2ME is a cloud-based data analysis platform of nanopore data in real-time. It shows the portion of reads for each microorganism. (A) The portion of reads for ZIKV PRVABC59 strain from the first run. (B) The portion of reads for ZIKV PRVABC59 strain from the second run. (C) The portion of reads for ZIKV MR 766 strain from the third run.

Figure 3. Long-reads with Nanopore technology. Several long sequences of ZIKV MR 766 were aligned to the reference genome. The maximum long sequence was 10,311bp.

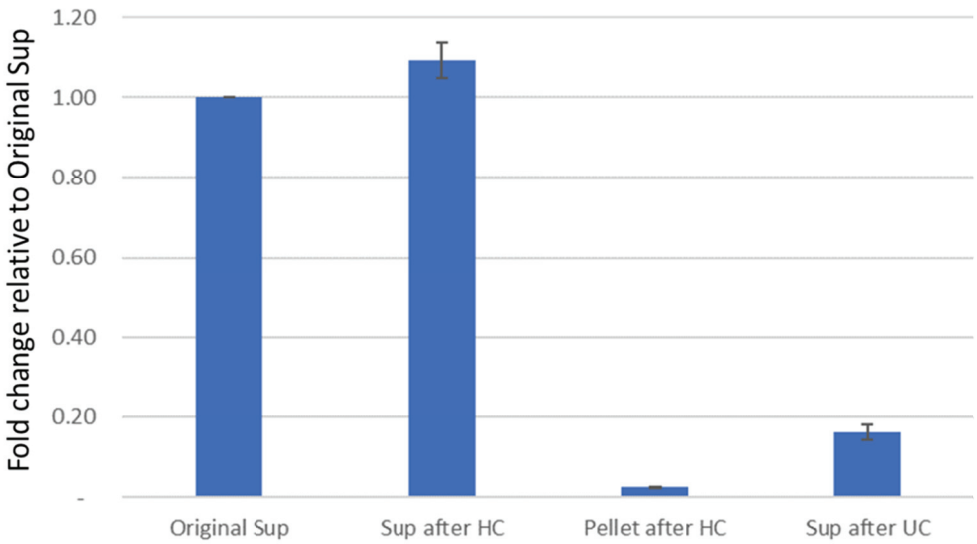

Figure 4. Relative quantification of ZIKV. Original Sup represents the original viral supernatant, Sup after HC represents the viral supernatant after high-speed centrifugation, Pellet after HC represents the pellet after high-speed centrifugation of the viral supernatant, and Sup after UC represents the viral supernatant after ultra-centrifugation. Bars indicate SEM from three experiments.

DISCUSSION

Clinical samples may contain not only human cells but also non-viral microorganisms such as bacteria or fungi. The rRNA depletion step removed parts of human RNAs but not all (Fig. 2A and B). Additionally, this step could not remove RNAs from non-viral microorganisms. Since the poly(A) tailing reaction was also targeting non-viral RNA, we modified the protocol to add the step of high-speed centrifugation (17,000 x g) to get rid of non-viral microorganisms, resulting in a significant reduction of non-viral targets (Table 1, Fig. 1C, and Fig. 2C). After direct RNA sequencing using Nanopore technology, we also set up the easy and fast analysis method to assemble and make the consensus sequence using Geneious Prime. We generated WGS of two strains of ZIKV, PRVABC59 and MR 766, without any cDNA synthesis and PCR-based enrichment. This established protocol (Fig. 5) can be applied to any viral RNA studies by providing an efficient method to perform WGS using direct RNA sequencing without any specific primers. This protocol could contribute to finding novel RNA viruses because cDNA synthesis and specific primers are not required. Recently, RNA epigenetics could be involved in gene expression and diseases (20) as well as viral replication and host interaction (21). Because RNA sequencing (RNA-seq) is an important tool for studying RNA epigenetics (RNA base modifications), the protocol in this study could be useful for detecting RNA base modifications in RNA viruses.

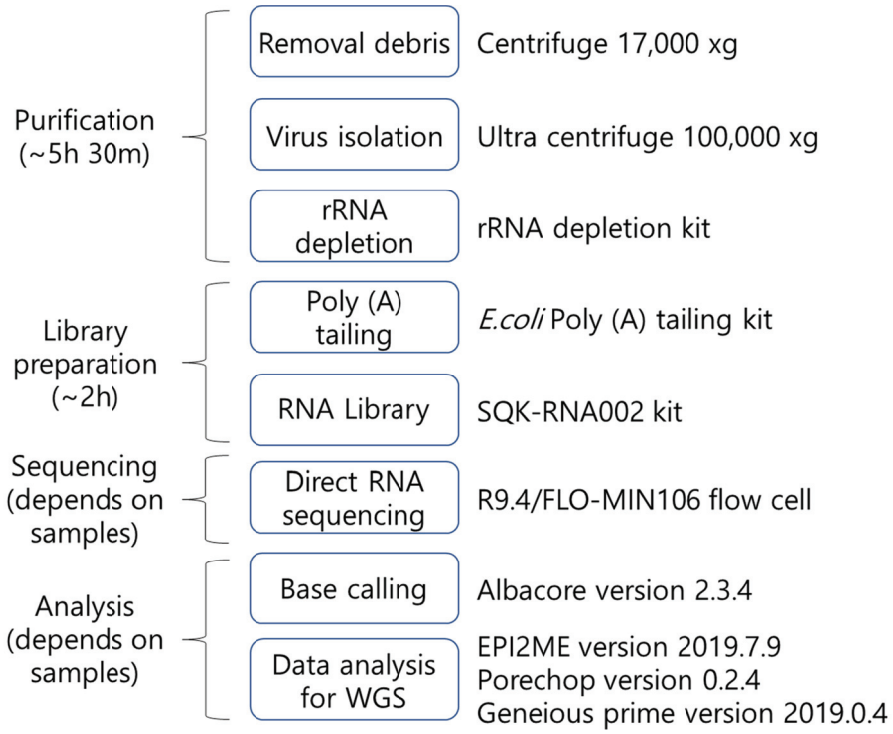

Figure 5. The work flow for WGS of single-strand RNA virus using direct RNA sequencing with Nanopore technology. High-speed centrifugation was important to improve the efficiency of WGS and direct RNA sequencing with Nanopore technology could make WGS faster and easier with long-reads and without specific primers and PCR-based enrichment

Even though, adding the step of high-speed centrifugation reduced amount of non-viral microorganisms, some bacteria still were remained (Fig. 2C). And we also recognize that this protocol requires large amounts of virus to improve coverage and depth of the target. We confirmed that the results from the MR766 strain ($1.8x10^9$ pfu) had higher identity, coverage, and depth than that of the PRVABC59 strain ($4.2 x10^7$ pfu). Improvement of these limitation should be studied further. In conclusion, the protocol in this study could provide fast and efficient WGS using direct RNA sequencing that could be applied to study biology of RNA viruses including identification, characterization, and global surveillance.

Acknowledgments

This work was supported by a grant from the KRIBB Research Initiative Program (KGM4571922), Republic of Korea.

REFERENCES

1) Malone RW, Homan J, Callahan MV, Glasspool-Malone J, Damodaran L, Schneider Ade B, et al. Zika Virus: Medical Countermeasure Development Challenges. PLoS Negl Trop Dis 2016;10:e0004530.

2) Sikka V, Chattu VK, Popli RK, Galwankar SC, Kelkar D, Sawicki SG, et al. The emergence of zika virus as a global health security threat: A review and a consensus statement of the INDUSEM Joint working Group (JWG). J Glob Infect Dis 2016;8:3-15.

3) Grard G, Caron M, Mombo IM, Nkoghe D, Mboui Ondo S, Jiolle D, et al. Zika virus in Gabon (Central Africa)--2007: a new threat from Aedes albopictus? PLoS Negl Trop Dis 2014;8:e2681.

4) Brasil P, Pereira JP Jr, Moreira ME, Ribeiro Nogueira RM, Damasceno L, Wakimoto M, et al. Zika Virus Infection in Pregnant Women in Rio de Janeiro. N Engl J Med 2016;375:2321-34.

5) Bearcroft WG. Zika virus infection experimentally induced in a human volunteer. Trans R Soc Trop Med Hyg 1956;50:442-8.

6) Ahlfors K, Ivarsson SA, Harris S. Report on a long-term study of maternal and congenital cytomegalovirus infection in Sweden. Review of prospective studies available in the literature. Scand J Infect Dis 1999;31:443-57.

7) Schuler-Faccini L, Ribeiro EM, Feitosa IM, Horovitz DD, Cavalcanti DP, Pessoa A, et al. Possible Association Between Zika Virus Infection and Microcephaly - Brazil, 2015. MMWR Morb Mortal Wkly Rep 2016;65:59-62.

8) Mlakar J, Korva M, Tul N, Popović M, Poljšak-Prijatelj M, Mraz J, et al. Zika Virus Associated with Microcephaly. N Engl J Med 2016;374:951-8.

9) Oehler E, Watrin L, Larre P, Leparc-Goffart I, Lastere S, Valour F, et al. Zika virus infection complicated by Guillain-Barre syndrome--case report, French Polynesia, December 2013. Euro Surveill. 2014;19.

10) Broutet N, Krauer F, Riesen M, Khalakdina A, Almiron M, Aldighieri S, et al. Zika Virus as a Cause of Neurologic Disorders. N Engl J Med 2016;374:1506-9.

11) Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. Nat Rev Microbiol 2017;15:183-92.

12) Grad YH, Newman R, Zody M, Yang X, Murphy R, Qu J, et al. Within-host whole-genome deep sequencing and diversity analysis of human respiratory syncytial virus infection reveals dynamics of genomic diversity in the absence and presence of immune pressure. J Virol 2014;88:7286-93.

13) Parker J, Chen J. Application of next generation sequencing for the detection of human viral pathogens in clinical specimens. J Clin Virol 2017;86:20-6.

14) Baronti C, Piorkowski G, Leparc-Goffart I, de Lamballerie X, Dubot-Pérès A. Rapid next-generation sequencing of

dengue, EV-A71 and RSV-A viruses. J Virol Methods 2015;226:7-14.

15) Marston DA, McElhinney LM, Ellis RJ, Horton DL, Wise EL, Leech SL, et al. Next generation sequencing of viral RNA genomes. BMC Genomics 2013;14:444.

16) Oude Munnink BB, Kik M, de Bruijn ND, Kohl R, van der Linden A, Reusken C, et al. Towards high quality real-time whole genome sequencing during outbreaks using Usutu virus as example. Infect Genet Evol 2019;73:49-54.

17) Bowden R, Davies RW, Heger A, Pagnamenta AT, de Cesare M, Oikkonen LE, et al. Sequencing of human genomes with nanopore technology. Nat Commun 2019;10:1869.

18) McNaughton AL, Roberts HE, Bonsall D, de Cesare M, Mokaya J, Lumley SF, et al. Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). Sci Rep 2019;9:7081.

19) Wongsurawat T, Jenjaroenpun P, Taylor MK, Lee J, Tolardo AL, Parvathareddy J, et al. Rapid Sequencing of Multiple RNA Viruses in Their Native Form. Front Microbiol 2019;10:260.

20) Garber K. Epigenetics comes to RNA. Science 2019;365:16-7.

21) Lichinchi G, Zhao BS, Wu Y, Lu Z, Qin Y, He C, et al. Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. Cell Host Microbe 2016;20:666-73.