



OPEN

Universal primers for rift valley fever virus whole-genome sequencing

Kwan Woo Kim^{1,2,9}, Banseok Lee^{2,3}, Sujeong Eom^{2,3}, Donghoon Shin^{2,3}, Changwoo Park^{4,5,6}, Seil Kim^{4,5,7} & Hana Yi^{2,3,8}

Rift Valley fever (RVF) is a mosquito-borne zoonotic disease causing acute hemorrhagic fever. Accurate identification of mutations and phylogenetic characterization of RVF virus (RVFV) require whole-genome analysis. Universal primers to amplify the entire RVFV genome from clinical samples with low copy numbers are currently unavailable. Thus, we aimed to develop universal primers applicable for all known RVFV strains. Based on the genome sequences available from public databases, we designed eight pairs of universal PCR primers covering the entire RVFV genome. To evaluate primer universality, four RVFV strains (ZH548, Kenya 56 (IB8), BIME-01, and Lunyo), encompassing viral phylogenetic diversity, were chosen. The nucleic acids of the test strains were chemically synthesized or extracted via cell culture. These RNAs were evaluated using the PCR primers, resulting in successful amplification with expected sizes (0.8–1.7 kb). Sequencing confirmed that the products covered the entire genome of the RVFV strains tested. Primer specificity was confirmed via *in silico* comparison against all non-redundant nucleotide sequences using the BLASTn alignment tool in the NCBI database. To assess the clinical applicability of the primers, mock clinical specimens containing human and RVFV RNAs were prepared. The entire RVFV genome was successfully amplified and sequenced at a viral concentration of 10^8 copies/mL. Given the universality, specificity, and clinical applicability of the primers, we anticipate that the RVFV universal primer pairs and the developed method will aid in RVFV phylogenomics and mutation detection.

Rift Valley fever (RVF) is a zoonotic disease affecting both animals and humans. The causal pathogen, a mosquito-borne virus, was first reported in young sheep and humans in the Rift Valley of Kenya in 1931¹. Although initially confined to sub-Saharan Africa, the disease has been reported in other regions beyond the African continent since 2000, including Saudi Arabia, Yemen, and China^{2,3}. Typically, dozens to hundreds of cases of RVF infection occur every 2–3 years. In cases where livestock are affected by this disease, a characteristic symptom known as the “abortion storm” leads to high mortality rates and miscarriage rates. The spread of RVF causes significant economic damage to livestock production and rural farming communities^{4,5}.

The incubation period of the disease is approximately 2–6 days, and mortality rates in humans are estimated to be 0.5–2.0%⁶. The majority of patients exhibit mild symptoms, including fever, dizziness, and severe weight loss, and naturally recover within 2–7 days. However, symptoms such as shock or encephalitis can occur in severe cases. The main biological vectors of RVF, through which it is transmitted to both livestock and humans, include more than 30 mosquito species belonging to the genera *Aedes*, *Culex*, *Anopheles*, *Eretmapodites*, *Mansonia*, and *Coquillettidia*⁷. Despite the serious nature of RVF infection in both humans and livestock, there are currently no effective vaccines or treatments available for this disease. Veterinary vaccines and attenuated vaccines have

¹Department of Public Health Sciences, Graduate School, Korea University, Seoul, Republic of Korea. ²Interdisciplinary Program in Precision Public Health, Korea University, Seoul, Republic of Korea. ³Integrated Biomedical and Life Science, Graduate School, Korea University, Seoul, Republic of Korea. ⁴Microbiological Analysis Team, Group for Biometrology, Korea Research Institute of Standards and Science (KRISS), Daejeon, Republic of Korea. ⁵Convergent Research Center for Emerging Virus Infection, Korea Research Institute of Chemical Technology (KRICT), Daejeon, Republic of Korea. ⁶Department of Agricultural Biotechnology, Seoul National University, Seoul, Republic of Korea. ⁷Department of Bio-Analysis Science, University of Science and Technology, Daejeon, Republic of Korea. ⁸School of Biosystems and Biomedical Sciences, Korea University, Seoul, Republic of Korea. ⁹Present address: Center for Study of Emerging and Re-emerging Viruses, Korea Virus Research Institute, Institute for Basic Science (IBS), Daejeon, Republic of Korea. ✉email: stapler@kriss.re.kr; hanayi@korea.ac.kr

been developed, but vaccines for human use have not been developed yet⁴. Hence, in 2018, RVF was designated as a “Disease X” priority disease by the World Health Organization (WHO)⁸.

The RVF virus (RVFV) genome comprises three single-stranded RNA segments designated as large, medium, and small segment. The large segment (L; 6.4 kb) encodes an RNA-directed RNA polymerase⁹. The medium segment (M; 3.8 kb) encodes two envelope glycoproteins (Gn and Gc) and the NSm protein^{10–12}. The small segment (S; 1.6 kb), which is characterized by ambisense polarity, encodes a structural nucleoprotein (N) in the negative sense and a small nonstructural protein (NSc) in the positive sense¹³.

Since its initial discovery, RVFV has spread across a wide geographical range, and, to date, only a single serotype has been identified¹⁴. Approximately 184 full-length genomes of RVFV have been deposited in the Virus Pathogen Resource (ViPR) database; they have low genetic diversity and differ by less than 5% and 2% at the nucleotide (nt) and amino acid (aa) levels (M segment: nt, 5%; aa, 2%; and L and S segments: nt, 4%; aa, 1%), respectively. However, phylogenetic tree analysis based on the whole genome has revealed several distinct genetic lineages (A to O) of RVFV¹⁵. This diversity is believed to be attributable to the reassortment of the segmented genome^{16,17}, and despite high genome sequence similarity, RVFV is phylogenetically branched into 15 lineages¹⁵.

RVFV diagnostic methods include quantitative real-time PCR (qRT-PCR)¹⁸, multiplex PCR-based microarrays¹⁹, and recombinase polymerase amplification (RPA)²⁰. The genes *NSm*, *NSs*, and *Gc* are used for qRT-PCR^{18,21,22}. Although these methods are considerably valuable for the rapid and accurate diagnosis of viral infections, the methods are of limited utility with respect to characterizing specific mutations in a genome-wide manner. Viral mutations are of concern because they can influence the transmissibility of the virus, severity of infection, and the efficacy of vaccines. This highlights the need for the rapid identification of mutations, as exemplified by the mutations in the SARS-CoV-2 virus, which has had considerable effects on virus transmission and vaccine efficacy^{23–25}. Consequently, the WHO recommends whole-genome analysis for the accurate diagnosis of infectious viruses and the identification of associated mutations. Sequencing of viral genomes requires the initial isolation of viral agents from clinical samples preserved in cell culture, which takes several days. To obtain genome information immediately, it is necessary to extract and sequence viral nucleic acids directly from clinical samples. However, the amount of viral nucleic acids in clinical samples is generally insufficient for full-genome analysis; hence, initial sample amplification is required. For example, the ARTIC nCoV-2019 Amplicon Panel has been used for the whole-genome analysis of SARS-CoV-2^{26,27}. MERS-CoV and the alphacoronavirus universal primer panel has been developed for genome surveillance^{28,29}. At present, however, no universal primers are available for the amplification of RVFV nucleic acid.

The RVFV genome does not show specific regions where mutations are known to frequently occur. The mutations found in the RVFV genome are a result of random single-nucleotide polymorphisms (SNPs) rather than from specific segments. As mentioned earlier, due to the lack of serotype differentiation in RVFV, complete genomic information is crucial for distinguishing different strains^{14,30}. Therefore, the development of universal primers for RVFV could serve as a useful tool for monitoring RVFV mutations and determining the genetic diversity of the virus. Currently, the RVFV genome is being characterized by the following methods: Genome sequences are determined via sequence-independent single primer amplification (SISPA) and primer walking with Sanger sequencing^{31,32}. Target sequences of the nonstructural (*Ns*) gene and glycoprotein (*Gn*) gene are used for phylogenetic analyses³³.

The aim of this study was to develop universal RVFV primers that could be applied to RVFV amplicon-based whole-genome sequencing. To verify the applicability of the developed primers, we assessed the nucleic acids of four representative viral strains from different phylogenetic branches of RVFV. Additionally, mock clinical samples were prepared to evaluate the applicability of the developed methodology to the analysis of patient clinical samples.

Results

Selection of universal primers for RVFV

A total of 80 candidate universal primers were designed based on the publicly available genomes of RVFV strains. To select the most optimal universal primers, candidate primers were screened via PCR analysis using the four RVFV strains (ZH548, Kenya 56 (IB8), BIME-01, and Lunyo). Eight universal primer pairs for RVFV covering the entire RVFV genome segments and physical regions were selected (Table 1, Fig. 1). Gel electrophoresis analysis of the PCR amplicons revealed products of the expected size (0.8–1.7 kb) (Fig. 2). However, several unexpected phenomena were observed, listed as follows. (1) The M2 amplicon of the ZH548 strain appeared different from its theoretically expected size. However, Sanger sequencing of this specific band revealed that the target sequence was amplified with the correct size and sequence. (2) In all the four strains tested, we observed the generation of multiple bands in L2 and S2 amplicons. However, we decided to retain those primers due to the absence of alternative primer candidates for those regions. (3) The amplification efficiency of L1 decreased according to the decrease in viral RNA concentration. This was problematic because the lower amplification efficiency than that of other primer regions could result in the failure of sequencing of corresponding regions. Because the search for better primers for this region did not succeed, four additional primer pairs were designed and used together in the further analyses of this problematic region. (4) Evaluation of the potential applicability of multiplex tiling PCR revealed a lack of S-segment synthesis in the amplicon of the Lunyo strain (Supplementary Fig. 1), indicating that the selected primer pairs would not be suitable for multiplex tiling PCR.

Genome sequences determined from primer amplicons

For each of the four RVFV strains (ZH548, Kenya 56 (IB8), BIME-01, and Lunyo), purified amplification products were pooled at the same concentration and sequenced using the Oxford Nanopore Technologies (ONT; Oxford, United Kingdom) MinION Mk1C sequencer and the Illumina (San Diego, CA, USA) MiSeq platform.

Target region	Primers	Sequence (5' → 3')	Product size (bp)
RL01	RL01-1F	ACACAAAGGCGCCCAATCA	1,737
	RL01-1737R	AAACCCTGGAAGTGGAGAG	
RL02	RL02-1472F	CATCAGCCATCCTTRTCAG	1,957
	RL02-3428R	GCTGGGAAGCTGATTAGCA	
RL03	RL03-2962F	CCTGTGTGGACTTGTGC	2,066
	RL03-5027R	GATTCCTTTATCTCAACGTTTG	
RL04	RL04-4471F	TCACTGGATTCTATAGTTTAC	1,914
	RL04-6384R	ACACAAAGACCGCCCAATATT	
RM01	RM01-2F	ACACAAAGACGGTGCATTAAT	2,036
	RM01-2037R	AGGTATGTGCTATAACGTGG	
RM02	RM02-1650F	GATGATCAGTCAGTTAGCTC	2,196
	RM02-3845R	TCAAAGAGTTAGTTTAATTCCY	
RS01	RS01-1F	ACACAAAGACCCCTAGTG	818
	RS01-818R	AACCTCTARTCAACCTCAAC	
RS02	RS02-907F	TGGGCAGCCACTTAGGCT	764
	RS02-1671R	ACACAAAGCTCCCTAGAGAT	

Table 1. List of RVFV universal primers for whole genome sequencing.

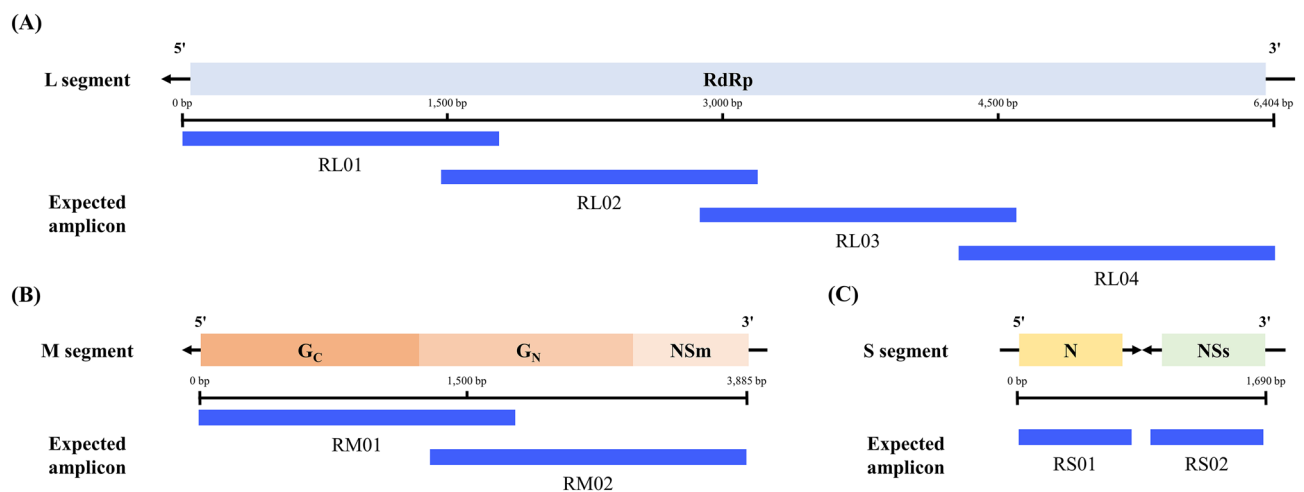


Figure 1. Position of the designed universal primers and the expected amplicons alongside the genome of RVFV. The graphical image of the (A) L segment, (B) M segment, and (C) S segment of the RVFV genome shows the length and gene contents of the RNA genome. The localization of expected PCR amplicons from the eight universal primers designed in this study are indicated as blue solid lines.

The sequencing statistics of MinION were as follows: (1) Strain ZH-548 produced 3.2 Gb (fast5 20 GB, 1,252,527 reads) of raw data and four linear contigs (17,132.8–237,223.1X sequencing depth) after assembly. (2) Kenya 56 (IB8) produced 574.34 Mb (fast5 3.36 GB, 198,768 reads) of raw data and four linear contigs (15,376–21,778.55X sequencing depth). (3) BIME-01 produced 4.22 Gb (fast5 26.1 GB, 1,434,586 reads) of raw data and four linear contigs (108,352.3–367,227.3X sequencing depth). (4) Lunyo produced 3.55 Gb (fast5 25.3 GB, 2,147,677 reads) of raw data and four linear contigs (43,571.9–280,659.9X sequencing depth). A mean Phred score of 13 (quality filtering > 8) and read lengths greater than 800–1600 bp (length filtering > 700 bp) were observed.

The sequencing statistics of MiSeq were as follows: (1) Strain ZH-548 produced 2.18 Gb (3,466,421 reads) of raw data and four linear contigs (99,154.9–285,104.7X sequencing depth) after assembly. (2) Kenya 56 (IB8) produced 2.57 Gb (4,089,726 reads) of raw data and four linear contigs (116,523.2–564,624.9X sequencing depth). (3) BIME-01 produced 2.15 Gb (3,420,329 reads) of raw data and four linear contigs (94,967.6–351,259.8X sequencing depth). (4) Lunyo produced 3.16 Gb (5,025,765 reads) of raw data and four linear contigs (83,987.5–762,540.4X sequencing depth). A mean Phred score of 30 was observed.

The MinION sequences of all four strains were successfully assembled via reference mapping. The mean sequencing read depth and coverage of MinION were 15,376–367,227X and 99.54–100%, respectively (Table 2). The MiSeq sequences of all four strains were successfully assembled based on reference mapping. Compared to the mapping method, the M segment of strain ZH548 was not assembled during the de novo assembly and showed lower genome coverage, thereby indicating that the reference mapping method is more suitable for

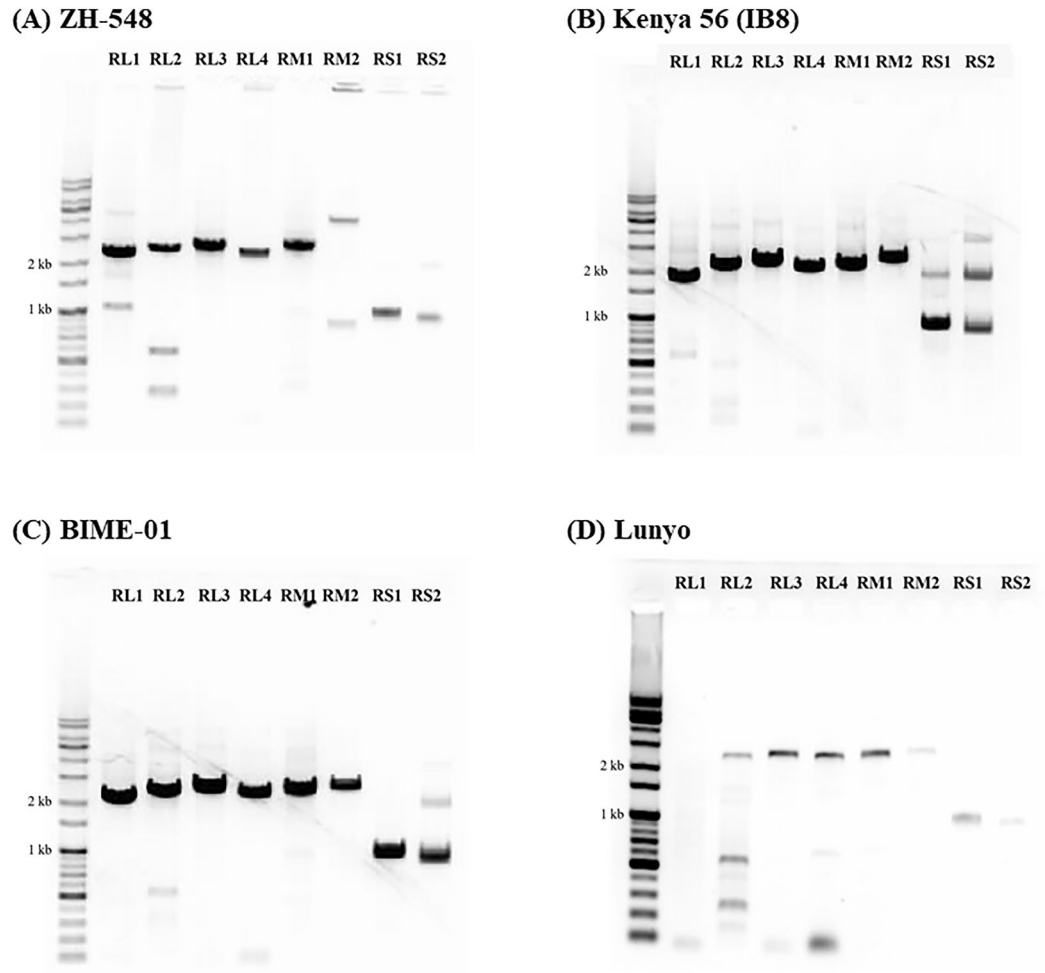


Figure 2. Experimental amplification products of RVFV universal primers. Agarose gel electrophoresis image of PCR amplicons from the four test strains of RVFV. (A) ZH548, (B) Kenya 56 (IB8), (C) BIME-01, and (D) Lunyo.

Strains	Platform	% Mapped read* (L/M/S1/S2)	% Genome coverage** (L/M/S1/S2)	% Similarity***
ZH-548	MinION	30.58/30.78/23.86/17.67	100/99.5367/100/100	100
	MiSeq	47.67/24.57/13.23/14.35	100/99.5367/100/100	100
Kenya 56 (IB8)	MinION	34.46/20.73/14.08/39.94	100/99.5367/100/100	100
	MiSeq	44.55/22.00/10.69/23.36	100/99.5367/100/100	100
BIME-01	MinION	37.18/17.33/22.67/29.40	100/99.5367/100/100	100
	MiSeq	48.46/22.06/16.00/14.29	100/99.5367/100/100	100
Lunyo	MinION	18.50/6.50/4.92/17.76	100/99.5367/100/100	100/99.9482/100/100
	MiSeq	20.69/14.47/5.86/2.130	100/99.5367/100/100	100

Table 2. Comparison of whole-genome sequencing data according to viral strains and sequencing platforms. *The mapped read refers to the proportion of read counts that align to the reference sequence for each segment among the raw reads generated from the NGS sequencing. **Genome coverage means the percentage area coverage of a reference (consensus sequence; L 6,404 bp; M, 3,867 bp; S1, 738 bp; S2, 798 bp) by the assembled contig. ***Similarity means the proportion of matched nucleotides in the assembled contigs compared to a reference (consensus sequence; L, 6,404 bp; M, 3,867 bp; S1, 738 bp; S2, 798 bp).

RVFV genome assembly (Supplementary Table 1). The mean sequencing read depth and coverage of MiSeq were 28,189–475,170X and 99.54–100%, respectively (Table 2).

In both MinION and MiSeq sequencing, the determined genome sequences showed 100% sequence similarity to the reference sequence with only one exception. The Lunyo M segment showed 99.9482% of similarity to the known sequence of reference genome (Table 2). The newly determined sequences of the four test strains are deposited in GenBank (accession number: OQ440142–OQ440157).

Phylogenetic analysis was performed using the newly determined genome sequences. Based on the partial 490-nt M segment method proposed by Grobbelaar et al.¹⁵, which classifies strains into 15 lineages from A to O, the ZH-548, Kenya 56 (IB8), BIME-01, and Lunyo strains are members of lineages A, L, K, and M, respectively (Fig. 3, Supplementary Fig. 2), in line with previous reports.

Evaluation of primer specificity, versatility, and efficiency

The performance of universal primer pairs was further evaluated in terms of specificity, versatility, amplification efficiency, and sensitivity. The specificity of the selected universal primer pairs for RVFV was assessed by *in silico* comparison against known nucleotide sequences. A BLASTn search was conducted on the NCBI nr/nt database to determine if the selected universal primers shared similarities with other genomes, including those of the sandfly fever Naples phlebovirus and the human genome. No similarities (E -value $< 10^{-5}$) were observed with any genome besides RVFV, demonstrating the high specificity of the universal primer set for RVFVs.

To evaluate the versatility of the universal primer pairs, each primer pair was aligned to the 37 genomes of lineages A to O of RVFV. There was a 0 or 1 mismatch between individual primer pairs and genome sequences, and 20-nt mismatches were confirmed for all eight primers (Supplementary Fig. 3), with the exception of the SA 51 (Van Wyck) (DQ380195) strain, for which two mismatches were obtained using the M02 reverse primer, and the TAN/Tan 001/07 (HM586970) strain, which showed three mismatches with the S01 forward primer. The sequences of the regions with mismatches in the SA 51 (Van Wyck) and TAN/Tan 001/07 strains differed from those of closely related strains, as indicated by the phylogenetic analysis. In addition, the content of the mismatch was observed in the middle of homopolymeric sequences (with the same base sequence continuously in the problematic region). Hence, the 2–3 mismatches observed here were considered sequencing errors occurring at the ends of the sequences (Supplementary Table 2). Consequently, assuming that these mismatches are attributable to sequencing errors, it would appear that the primer pairs developed in this study are universal for all RVFV lineages.

To compare the amplification efficiency among the eight pairs of universal primers, the read depth coverage of MinION sequences was examined. Overall, the universal primer sets exhibited a high read depth, indicating a high amplification efficiency. Among them, RS01 showed the highest efficiency, while conversely, the performance of RM02 was the lowest (Fig. 4).

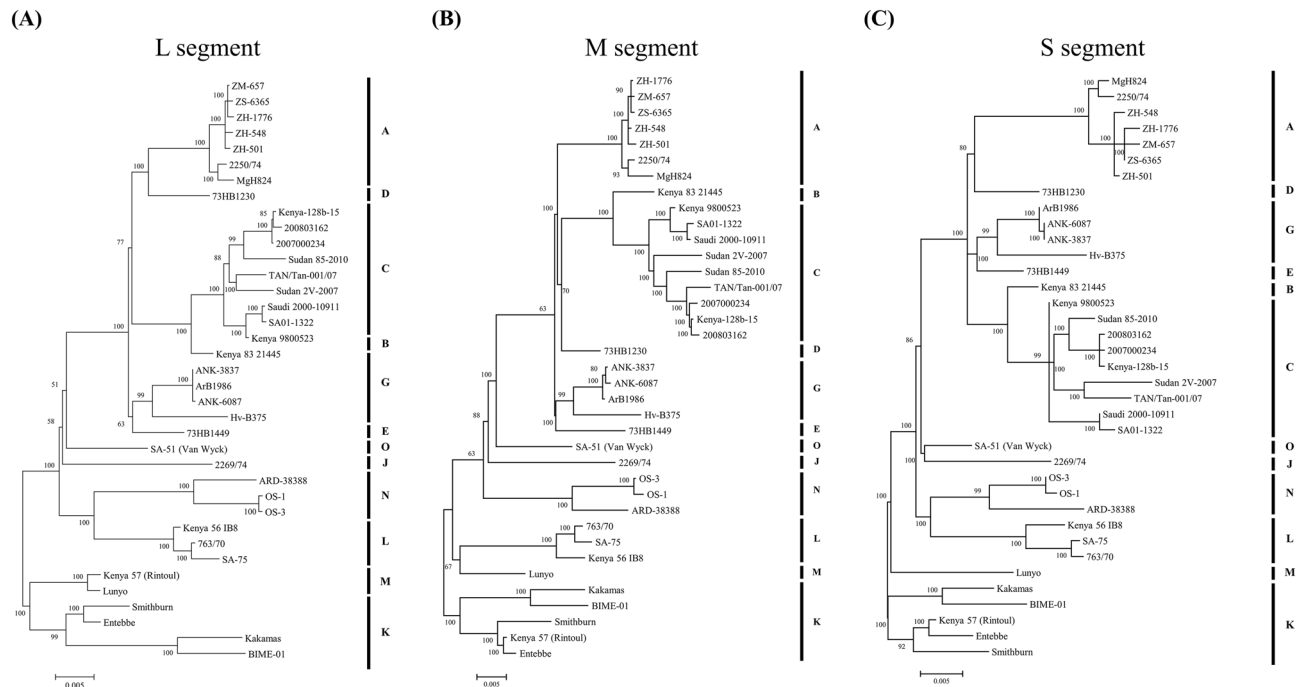


Figure 3. Phylogenetic tree based on the whole-genome sequences of RVFV. Maximum likelihood tree of the (A) L segment, (B) M segment, and (C) S segment. Tree analysis was performed using RAxML version 8.2.10 and evaluated using 100,000 bootstrap replicates. Lineage names followed the nomenclature of Grobbelaar et al.¹¹.

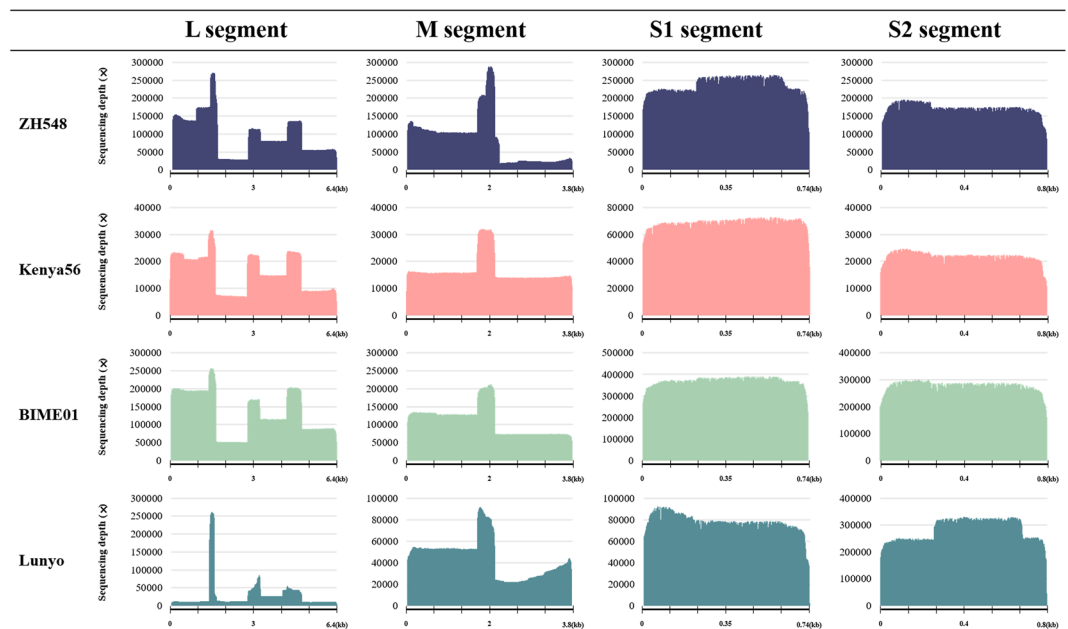


Figure 4. Comparison of primer efficiency among the eight pairs of universal primers measured by sequencing read depth coverage. The MinION sequencing read depth coverage is indicated in the y axis, and the genome region amplified by the eight primer pairs (four pairs in the L segment, two pairs in the M segment, and two pairs in the S segment) is indicated in the x axis. *, S1 product encodes the N gene (738 bp) and S2 product encodes the NSs gene (798 bp).

Evaluation of primer sensitivity

The sensitivity of universal primers was assessed by determining the limit of detection. The RNA sample of strain ZH-548 was serially diluted to obtain a viral genome concentration of 10^8 – 10^3 copies/mL. The prepared RNA samples were subjected to PCR and sequencing following the protocol developed in this study. As a result, the entire ZH-548 RVFV genome was successfully assembled to a complete genome at RNA concentrations of 10^4 copies/mL and above. The sequencing mean depth was 2,288.3–3,015.48X, and sequence similarity to the reference sequence was 99.98–100%. At 10^4 copies/mL concentration, the lowest sequencing depth was observed from the RL04 primer (248X), and the highest sequencing depth was observed from RM01 primer (12,968X). At RNA concentrations below 10^4 copies/mL, one or more amplicons among the eight RF primers was not amplified and resulted in incomplete genome sequence. Therefore, the detection limit of the developed method for the RVFV samples was determined as 10^4 copies/mL.

Clinical specimen applicability

To assess the clinical applicability of the RF primers, a mock clinical specimen was used due to the inaccessibility to human clinical samples. The mock clinical specimen contained a mixture of human blood RNA and RVFV RNA (10^8 copies/mL). This specimen was subjected to cDNA synthesis, PCR, and sequencing. Even in the mock clinical samples, the entire RVFV genome was successfully amplified and sequenced. The MinION sequencing depth was 23,239.49–1,531,850X and covered the entire genome with 100% sequence similarity.

Discussion

Virus identification based on whole-genome information allows for precise determination of variants and their lineage. Consequently, whole-genome information is essential for epidemiological investigations and the prevention of disease spread, as it aids in tracing the origin of a specific virus during an outbreak and assessing the potential for dissemination based on known transmission dynamics. This information is vital for controlling the epidemiology and spread of infectious diseases. Traditionally, methods such as sequence-independent single primer amplification (SISPA)³⁴, virus discovery cDNA AFLP (VIDISCA)³⁵, and rolling circle amplification (RCA)³⁶ have been used to obtain the whole genome of viruses. These techniques have the property of amplifying nucleic acids randomly, obviating the need for specific primers. However, such random amplification processes can amplify both the host genome and the target virus genome simultaneously, limiting the efficiency of analyzing the viral genome. With recent advancements in high-throughput sequencing technologies, it has become possible to obtain viral genome sequences directly through shotgun sequencing, bypassing the viral nucleic acid amplification step. However, in cases where the viral concentration is low, such as in clinical samples, the shotgun sequencing approach might not be successful. Due to these reasons, there has been a growing need for universal primers that can effectively amplify very small amounts of viral nucleic acids present in clinical samples, facilitating whole-genome analysis. Universal primers have already found utility in rapidly analyzing genome information and contributing to epidemic management for viruses such as Ebola³⁷, influenza³⁸, and

SARS-CoV-2³⁹. The RVFV universal primers developed in this study are also expected to significantly contribute to monitoring and responding to RVF disease in the future.

High-throughput sequencing-based diagnostic methods entail higher costs and time than other viral molecular diagnostic assays, and their sensitivity is relatively lower^{40,41}. Accordingly, the whole-genome-based diagnostic method utilizing the universal primers developed in this study exhibits a sensitivity of around ten times lower than the most commonly used RT-PCR method⁴⁰ and demands a longer processing time⁴¹. The actual experimental cost per sample for genome amplification and sequencing in this study was \$300 for MinION and \$700 for MiSeq (Supplementary Table 3), being much higher than the cost of few tens of dollars for standard qPCR methods. We are not claiming that the genome analysis method developed in this study can replace standard diagnostic methods. RVF is an extremely dangerous disease, and all suspected cases should be diagnosed accurately through proper standardized testing methods. Considering factors such as sensitivity in clinical samples, cost, and working time, it is appropriate to perform precise diagnosis for patients using traditional standard diagnostic methods. As a secondary step, we propose using the whole-genome analysis method developed in this research for additional mutation analysis.

The RVFV genome encodes virus-derived siRNAs (vsiRNAs), which are evenly distributed throughout the L and M segments of the genome but are concentrated in specific regions of the S segment⁴². Within the S segment, vsiRNAs were primarily clustered in the intergenic region (IGR). Similarly, several bunyaviruses have been found to have dense distributions of vsiRNAs in the IGR, and these are predicted to have a stem and loop (hairpin) structure^{42–44}. Such stem and loop structures have been found to reduce the efficiency of polymerases, thereby lowering the efficiency of the overall PCR⁴⁴. Consistently, compared to other amplification products, we obtained relatively lower yields of the S-segment amplicons, which we believe is attributable to the stem and loop structure of the S segment (Supplementary Fig. 4). To overcome this limitation, as an alternative to the RSL primer that covers the entire S segment, we designed two primers (RS01 and RS02) that, with the exception of the central IGR, contained normal coding sequences. Using these primers, we obtained amplification products of the expected sizes for all four strains. RS01 and RS02 were selected as the primers for S-segment amplification.

Unlike other RNA viruses, RVFV lacks well-defined variable regions and is known to have random single-site mutations throughout the genome¹⁴. Given these characteristics, entire genomes rather than specific genes have been used for phylogenetic analysis. Consequently, the topological resolution tends to be lower for lineages with a small number of reported genomes (lineages B, D, J, M, and O)³⁰. Thus, it is necessary to identify a phylogenetic marker with higher resolution, and in this regard, Grobbelaar et al. proposed analysis using the glycoprotein (*Gn*) gene, which has the highest rate of mutation^{15,45}. However, discrepancies between the lineages derived from genome phylogenetic analysis and those derived from *Gn* phylogenetic analysis have been reported in some viral strains. For example, while the M and S segments of strain 73HB1230 have been shown to belong to lineage B, the L segment belongs to lineage A¹⁶. This anomaly is attributable to viral reassortment, which is frequently observed in segmented RNA viruses⁴⁵. Consequently, for accurate phylogenetic analysis of RVFV and identification of mutations, it is necessary to use whole-genome information rather than limited sequences of the *Gn* gene.

To establish universal applicability of the developed protocol, it was necessary to evaluate a broad range of RVFV lineages. Currently, however, there is limited availability of RVFV nucleic acids. To overcome these limitations, we generated synthetic genomes of three genetically different strains using the IVT method. Viral genomic RNA synthesis using IVT is a method that has previously been used for *Ebola Zaire* virus detection and quantification⁴⁵ and for the development of a real-time PCR system for SARS-CoV-2 diagnosis⁴⁶. Thus, it is assumed that our method would be equally applicable in experiments using actual RVFV nucleic acids.

The main limitation of this study is the lack of clinical specimens from RVFV-infected patients. While the primers we developed showed promising results in *in vitro* experiments, the presence of various inhibitors within actual clinical specimens could introduce false positives. However, since South Korea has never experienced an outbreak of RVF, it is not feasible to import samples related to this virus into the country. Thus, mock RVFV clinical specimens were prepared and assessed instead. Even though RNA extracted from adult male blood, which can act as a significant inhibitor, was included in the mock clinical specimens as non-targeted RNA, the experimental results confirmed successful amplification and sequencing of the entire RVFV genome in the mock clinical samples. In addition, blood viral concentrations of approximately 10^4 – 10^8 copies/mL have been detected within a day after RVFV infection⁴⁷. Thus, it can be assumed that human and animal blood specimens would contain more than the minimum amount ($\geq 10^4$ copies/mL) of virus required for the effective use of the developed method. Therefore, it is anticipated that the RVFV universal primer pairs and the experimental method developed in this study will work successfully in real clinical specimens, similar to the way it operated in the mock clinical samples.

Methods

Preparation of RVFV nucleic acid

Four representative RVFV strains (ZH548, Kenya 56 (IB8), BIME-01, and Lunyo) were selected and used. ZH548 (lineage A) was chosen because this strain is the most representative experimental strain of RVFV. Kenya 56 (lineage L) and Lunyo (Lineage K) were chosen considering their phylogenetic distance and lineage diversity. BIME-01 was chosen due to its recent transmission from Angola to China in 2016³².

The RNA of the Lunyo strain (EVAg 005N-02174, accession no. KU167025, KU167026, and KU167027) was obtained from the Department of Health: Public Health England—Virology & Pathogenesis group (DH) through European Virus Archive Global (EVAg; <http://www.european-virus-archive.com/>). Lunyo strain RNA was extracted from the supernatants of virus-infected cells and obtained at concentrations of 9.34×10^5 to 3.21×10^6 copies/ μ L according to EVAg. Genomic DNA of the L, M, and S segments of strains ZH548 (NC_014397, NC_014396, NC_014395), Kenya 56 (DQ375427, DQ380190, DQ380176), and BIME-01 (KX609031, KX609032,

KX609033) were chemically synthesized by a commercial service (Invitrogen, Waltham, MA, USA) based on the known genome sequence information. The L and S segments of genomic DNA were inserted into a pMARQ (AmpR) vector, and the M segment was inserted into a pMA (AmpR) vector (Supplementary Fig. 5A). Recombinant plasmid DNA was used to transform *Escherichia coli* cells using a TOPcloner™ PCR Cloning Kit (Enzynomics, Daejeon, Korea), and the transformed *E. coli* K12 DH10B™ T1R was stored at -80°C .

Preparation of genomic RNAs from synthesized viral genomes

To prepare viral RNAs from the synthesized genome, the recombinant plasmid DNA was linearized using a restriction enzyme at 37°C for 60 min. The linearized plasmid DNA was used as a template for in vitro transcription (IVT) using T7 RNA polymerase promoter (Supplementary Fig. 5B). Promoter synthesis was performed in a final volume of 25 μL , containing 2 μL of linear plasmid template, 1 μL of each of the forward and the reverse primers (10 μM), 0.5 μL of Ex Taq polymerase (5 U/ μL), 2.5 μL of 10X buffer (20 mM), 4 μL of dNTP mix (2.5 mM each), and 14 μL of nuclease-free water. The reaction consisted of initial denaturation at 94°C for 2 min; followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min; and a final extension step at 72°C for 10 min. RVFV genomic RNA was synthesized using the MEGAscript T7 Transcription Kit (Invitrogen) following the manufacturer's protocol. The reaction mixture was set up in a total volume of 20 μL , comprising 1 μL of linear template DNA (0.1–0.2 μg), 8 μL of NTP solution (75 mM), 2 μL of 10X buffer, 2 μL of enzyme mix, and 7 μL of nuclease-free water. The mixture was then incubated for 2 h. The RNA obtained was purified using Clean & Concentrator™-5 (Zymo Research, Irvine, CA, USA). The quality of RNA used for IVT was evaluated using a BioSpectrometer (Eppendorf, Hamburg, Germany) and a 4150 TapeStation system (Agilent Technologies, Santa Clara, CA, USA). The A260/A280 ratio of the synthesized IVT RNA, determined by BioSpectrometer, was 1.73–2.42. To quantify RNA, we performed droplet digital PCR using a QX200 system (Bio-Rad Laboratories, Hercules, CA, USA) with Supermix for probes (Bio-Rad Laboratories). The ddPCR mixture was prepared according to the manufacturer's instructions, and amplifications were performed using the following reaction cycle: reverse transcription at 42°C for 30 min; enzyme activation at 95°C for 10 min; 70 cycles of denaturation at 95°C for 30 s and annealing and extension at 60°C for 150 s; and a final enzyme deactivation step at 98°C for 10 min. The concentration of the synthesized IVT RNA, determined by ddPCR, was 4.51×10^8 to 6.32×10^{11} copies/ μL . For the three synthesized viruses (ZH548, Kenya 56, and BIME-01), viral RNA mimics were prepared by mixing the same copy numbers (10^{11} copies/ μL) of L, M, and S segment RNAs.

Designing universal primer candidates

Complete RVFV genome sequences (142 L, 151 M, and 209 S segments) were acquired from the ViPR database. After aligning the base sequences for each segment, primers were designed for the conserved regions. The primer selection criteria were as follows: (i) amplicon size of 0.8–1.7 kb; (ii) melting temperatures of 50 – 58°C ; (iii) overlap length > 300 bp; (iv) N base ≤ 2 ; and (v) primer length between 18 and 23 bp.

Reverse transcription and PCR amplification

The viral genomic RNAs were subjected to a reverse transcription to generate complementary DNA (cDNA) using a LunaScript RT SuperMix Kit (New England Biolabs, Ipswich, MA, USA) following the manufacturer's protocol. The total volume of the reaction mixture was 20 μL and consisted of 1 μL RNA template, 4 μL of 5X LunaScript RT supermix, and 15 μL of distilled water. The reverse transcription reaction was carried out under the following conditions: primer annealing at 25°C for 2 min, cDNA synthesis at 55°C for 10 min, and heat inactivation at 95°C for 1 min. The synthesized cDNAs were used as templates for PCR without purification.

The cDNAs were used as PCR template for primer screening. PCRs were performed in a final volume of 20 μL , containing 1 μL of first-strand cDNA, 1 μL of each of the forward and the reverse primers (10 pmol), 10 μL of Q5 High-Fidelity 2X Master Mix (New England Biolabs), and 7 μL of nuclease-free water. The amplification cycle consisted of initial denaturation at 98°C for 30 s; followed by 30 cycles consisting of denaturation at 98°C for 5 s, annealing at 55°C for 30 s, and extension at 72°C for 2 min; and a final extension step at 72°C for 2 min. Odd/even primer sets were used for multiplex tiling-PCR amplification. The odd set included each odd-region primer (RL01, RL03, RM01, and RS01), and the even set included each even-region primer (RL02, RL04, RM02, and RS02). The amplification products were purified using a QIAquick® PCR Purification Kit (Qiagen, Hilden, Germany) and confirmed via electrophoresis on 1% agarose gels.

NGS sequencing and assembly

For each sample, purified amplification products were pooled at the same concentration and sequenced using the Illumina MiSeq platform and the ONT MinION Mk1C sequencer. For long-read sequencing, a sequencing library was prepared using an SQK-LSK109 ligation sequencing kit (ONT) and an EXP-NBD104 native barcoding kit (ONT). According to the protocol recommended by ONT, R9.4.1 flow cells (ONT) were used for 1D amplicon ligation sequencing. The Guppy basecaller ver. 6.0.1 + 652ffd1 (ONT) program was used to conduct GPU calling, with a Q-score greater than 8. The quality of the raw data was evaluated using pycQC (ver. 2.5.2) and NanoPlot (ver. 1.32.1) software^{48,49}. The resultant FASTQ file was assembled using BWA-MEM (ver. 0.7.12)⁵⁰. Read alignment coverage and depth coverage were measured using the bedtool program (ver. 2.30.0)⁵¹, and the errors in mapped results were further corrected using Pilon (ver.1.24)⁵².

A short-read sequencing library was constructed using the TruSeq Nano DNA Kit (Illumina). Following the recommended protocol, 301-bp paired-end sequencing was performed using the Illumina MiSeq platform. Adapter sequences were trimmed from the raw data using Cutadapt ver. 1.11 + 13.g2af9c15, and the overlapping

paired-ends were merged and assembled using CLC workbench 9.5.3. (Qiagen), followed by manual correction using the CodonCode Aligner (CodonCode Corporation).

Phylogenetic analysis

A total of 37 RVF viral genomes across the 15 lineages were used for phylogenetic analysis (Supplementary Table 2). Multiple sequence alignments were performed using the cluster W algorithm in MEGA software (ver. 7.0.26), and phylogenetic trees were constructed using RAXML (version 8.2.10). The inferred trees were summarized using the tree-sumtree program (ver. 4.4.0). Whole-genome phylogenetic analysis was conducted using the concatenated sequences of large (L), medium (M), and small (S) segments.

Assessment of clinical applicability

There are several reports on the use of mock clinical specimens when human clinical samples are unavailable. Akyurek et al. prepared SARS-CoV-2 mock clinical specimen by mixing IVT synthesized viral RNA (10^8 copies/ μL) and human total RNA ($2.5 \text{ ng}/\mu\text{L}$) in a 1:1 volume ratio⁵³. Park et al. used RVFV mock clinical specimens by mixing RVFV RNA and human tissue RNA⁵⁴. As we were unable to source human clinical samples of RVFV-infected patients, a mock clinical specimen was prepared to assess the clinical applicability of the developed method. Human adult normal tissue RNA extracted from the blood of a 24-year-old man was obtained from BioChain (Newark, CA, USA). The mock clinical specimen contained $25 \mu\text{L}$ of human adult normal tissue RNA ($200 \text{ ng}/\mu\text{L}$), $25 \mu\text{L}$ of RVFV IVT RNA, and $200 \mu\text{L}$ RNA storage solution (Invitrogen). The final concentration of viral genome in this mock clinical specimen was adjusted to 10^8 copies/mL. The cDNA synthesis, PCR amplification, MinION sequencing, and assembly were performed as described above.

Data availability

The genome sequences generated during the current study are available from the NCBI under the accession numbers OQ440142–OQ440157, <https://www.ncbi.nlm.nih.gov/nucore/OQ440142>—<https://www.ncbi.nlm.nih.gov/nucore/OQ440157>.

Received: 8 February 2023; Accepted: 24 October 2023

Published online: 31 October 2023

References

- Daubney, R., Hudson, J. R. & Garnham, P. C. Enzootic hepatitis or Rift Valley fever. An undescribed virus disease of sheep cattle and man from East Africa. *J. Pathol.* **34**, 545–579 (1931).
- Nanyingi, M. O. *et al.* A systematic review of Rift Valley fever epidemiology 1931–2014. *Infect. Ecol. Epidemiol.* **5**, 28024 (2015).
- Liu, W., Sun, F. J., Tong, Y. G., Zhang, S. Q. & Cao, W. C. Rift Valley fever virus imported into China from Angola. *Lancet Infect. Dis.* **16**, 1226 (2016).
- Ikegami, T. & Makino, S. The pathogenesis of Rift Valley fever. *Viruses* **3**, 493–519 (2011).
- Bouloy, M. & Weber, F. Molecular biology of Rift Valley fever virus. *Open Virol. J.* **4**, 8–14 (2010).
- Pepin, M., Bouloy, M., Bird, B. H., Kemp, A. & Paweska, J. Rift Valley fever virus (*Bunyaviridae: Phlebovirus*): An update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Vet. Res.* **41**, 61 (2010).
- Bird, B. H., Ksiazek, T. G., Nichol, S. T. & MacLachlan, N. J. Rift Valley fever virus. *J. Am. Vet. Med. Assoc.* **234**, 883–893 (2009).
- WHO. *Annual Review of Diseases Prioritized under the Research and Development Blueprint* (World Health Organization, Geneva, Switzerland, 2018).
- Müller, R., Poch, O., Delarue, M., Bishop, D. H. L. & Bouloy, M. Rift Valley fever virus L segment: Correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *J. Gen. Virol.* **75**, 1345–1352 (1994).
- Suzich, J. A. & Collett, M. S. Rift Valley fever virus M segment: Cell-free transcription and translation of virus-complementary RNA. *Virology* **164**, 478–486 (1988).
- Kakach, L. T., Suzich, J. A. & Collett, M. S. Rift Valley fever virus M segment: Phlebovirus expression strategy and protein glycosylation. *Virology* **170**, 505–510 (1989).
- Collett, M. S. Messenger RNA of the M segment RNA of Rift Valley fever virus. *Virology* **151**, 151–156 (1986).
- Giorgi, C. *et al.* Sequences and coding strategies of the S RNAs of Toscana and Rift Valley fever viruses compared to those of Punta Toro, Sicilian sandfly fever, and Uukuniemi viruses. *Virology* **180**, 738–753 (1991).
- Gaudreault, N. N., Indran, S. V., Balaraman, V., Wilson, W. C. & Richt, J. A. Molecular aspects of Rift Valley fever virus and the emergence of reassortants. *Virus Genes* **55**, 1–11 (2019).
- Grobbelaar, A. A. *et al.* Molecular epidemiology of Rift Valley fever virus. *Emerg. Infect. Dis.* **17**, 2270–2276 (2011).
- Bird, B. H., Khristova, M. L., Rollin, P. E., Ksiazek, T. G. & Nichol, S. T. Complete genome analysis of 33 ecologically and biologically diverse Rift Valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. *J. Virol.* **81**, 2805–2816 (2007).
- Ikegami, T. Molecular biology and genetic diversity of Rift Valley fever virus. *Antiviral Res.* **95**, 293–310 (2012).
- Maquart, M. *et al.* Development of real-time RT-PCR for the detection of low concentrations of Rift Valley fever virus. *J. Virol. Methods* **195**, 92–99 (2014).
- Venter, M. *et al.* Macroarray assay for differential diagnosis of meningoencephalitis in Southern Africa. *J. Clin. Virol.* **60**, 50–56 (2014).
- Euler, M. *et al.* Recombinase polymerase amplification assay for rapid detection of Rift Valley fever virus. *J. Clin. Virol.* **54**, 308–312 (2012).
- Mwaengo, D. *et al.* Detection and identification of Rift Valley fever virus in mosquito vectors by quantitative real-time PCR. *Virus Res.* **169**, 137–143 (2012).
- Näslund, J. *et al.* Kinetics of Rift Valley fever virus in experimentally infected mice using quantitative real-time RT-PCR. *J. Virol. Methods* **151**, 277–282 (2008).
- Hadj Hassine, I. Covid-19 vaccines and variants of concern: A review. *Rev. Med. Virol.* **32**, e2313 (2022).
- Dai, H. & Zhao, B. Association between the infection probability of COVID-19 and ventilation rates: An update for SARS-CoV-2 variants. *Build. Simul.* **16**, 3–12 (2023).
- Andrews, N. *et al.* Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. *N. Engl. J. Med.* **386**, 1532–1546 (2022).

26. Tyson, J. R. *et al.* Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* <https://doi.org/10.1101/2020.09.04.283077> (2020).
27. Itokawa, K., Sekizuka, T., Hashino, M., Tanaka, R. & Kuroda, M. A proposal of alternative primers for the ARTIC network's multiplex PCR to improve coverage of SARS-CoV-2 genome sequencing. *bioRxiv* <https://doi.org/10.1101/2020.03.10.985150> (2020).
28. Kim, K. W. *et al.* A half-day genome sequencing protocol for middle east respiratory syndrome coronavirus. *Front. Microbiol.* **12**, 602754 (2021).
29. Choi, S. *et al.* Human alphacoronavirus universal primers for genome amplification and sequencing. *Front. Microbiol.* **13**, 789665 (2022).
30. Juma, J. *et al.* Genomic surveillance of Rift Valley fever virus: From sequencing to lineage assignment. *BMC Genomics* **23**, 520 (2022).
31. Maluleke, M. R. *et al.* A comparative genome analysis of Rift Valley fever virus isolates from foci of the disease outbreak in South Africa in 2008–2010. *PLOS Negl. Trop. Dis.* **13**, e0006576 (2019).
32. van Schalkwyk, A. & Romito, M. Genomic characterization of Rift Valley fever virus, South Africa, 2018. *Emerg. Infect. Dis.* **25**, 1979–1981 (2019).
33. Saeed, O. S., El-Deeb, A. H., Gadalla, M. R., El-Soally, S. A. G. & Ahmed, H. A. H. Genetic characterization of Rift Valley fever virus in insectivorous bats, Egypt. *Vector Borne Zoonotic Dis.* **21**, 1003–1006 (2021).
34. Chrzastek, K. *et al.* Use of sequence-independent, single-primer-amplification (SISPA) for rapid detection, identification, and characterization of avian RNA viruses. *Virology* **509**, 159–166 (2017).
35. Pyrc, K., Jebbink, M. F., Berkhout, B. & van der Hoek, L. Detection of New Viruses by VIDISCA: Virus Discovery Based on cDNA-Amplified Fragment Length Polymorphism. In *SARS-and Other Coronaviruses: Laboratory Protocols* Vol. 454 (ed. Cavanagh, D.) 73–89 (Humana Press, 2008).
36. Johne, R., Müller, H., Rector, A., Van Ranst, M. & Stevens, H. Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends Microbiol.* **17**, 205–211 (2009).
37. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
38. Zhao, J. *et al.* Nanomicroarray and multiplex next-generation sequencing for simultaneous identification and characterization of influenza viruses. *Emerg. Infect. Dis.* **21**, 400–408 (2015).
39. Park, C. *et al.* Rapid and sensitive amplicon-based genome sequencing of SARS-CoV-2. *Front. Microbiol.* **13**, 876085 (2022).
40. Sohni, Y. Variation in LOD across SARS-CoV-2 assay systems: Need for standardization. *Lab. Med.* **52**, 107–115 (2020).
41. Zhu, Y., Zhang, M., Jie, Z. & Tao, S. Nucleic acid testing of SARS-CoV-2: A review of current methods, challenges, and prospects. *Front. Microbiol.* **13**, 1074289 (2022).
42. Sabin, L. R. *et al.* Dicer-2 processes diverse viral RNA species. *PLOS ONE* **8**, e55458 (2013).
43. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10**, 126–139 (2009).
44. Liu, Z. *et al.* Transient stem-loop structure of nucleic acid template may interfere with polymerase chain reaction through endonuclease activity of Taq DNA polymerase. *Gene* **764**, 145095 (2021).
45. Ro, Y. T., Ticer, A., Carrion, R. Jr. & Patterson, J. L. Rapid detection and quantification of *Ebola Zaire* virus by one-step real-time quantitative reverse transcription-polymerase chain reaction. *Microbiol. Immunol.* **61**, 130–137 (2017).
46. Pfeffeler, S., Reucher, S., Nörz, D. & Lütgehetmann, M. Evaluation of a quantitative RT-PCR assay for the detection of the emerging coronavirus SARS-CoV-2 using a high throughput system. *Euro Surveill.* **25**, 2000152 (2020).
47. Busquets, N. *et al.* Experimental infection of young adult European breed sheep with Rift Valley fever virus field isolates. *Vector Borne Zoonotic Dis.* **10**, 689–696 (2010).
48. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
49. Leger, A. & Leonardi, T. pycoQC, interactive quality control for Oxford nanopore sequencing. *J. Open Source Softw.* **4**, 1236 (2019).
50. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv <https://doi.org/10.48550/arXiv.1303.3997> (2013).
51. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
52. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963 (2014).
53. Akyurek, S., Demirci, S. N. S., Bayrak, Z., Isleyen, A. & Akgoz, M. The production and characterization of SARS-CoV-2 RNA reference material. *Anal. Bioanal. Chem.* **413**, 3411–3419 (2021).
54. Park, C., Park, D., Hassan, Z. U., Choi, S. H. & Kim, S. Comparison of RT-qPCR and RT-ddPCR with Rift Valley fever virus (RVFV) RNA. *Sci. Rep.* **13**, 3085 (2023).

Author contributions

K.W.K., B.L., S.E., D.S., and C.P. performed the experiments and conducted bioinformatics analyses of the sequencing data. S.K. and H.Y. designed the study and interpreted data. K.W.K., S.K., and H.Y. were the major contributors to the writing of the manuscript. All the authors have read and approved the final version of the manuscript.

Funding

This study was supported by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute funded by the Ministry of Health & Welfare (HI20C0558); the Technology Innovation Program funded by the Ministry of Trade, Industry & Energy (20015205); and the Global Research Development Center Program through the National Research Foundation of Korea funded by the Ministry of Science and ICT (NRF-2022K1A4A8A02093937).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-45848-z>.

Correspondence and requests for materials should be addressed to S.K. or H.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023