



## Protocols

# Recovery of complete genome sequences of Crimean-Congo haemorrhagic fever virus (CCHFV) directly from clinical samples: A comparative study between targeted enrichment and metagenomic approaches

Jake D'Addiego<sup>a,b,\*</sup>, Nadina Wand<sup>a</sup>, Babak Afrough<sup>a</sup>, Tom Fletcher<sup>c</sup>, Yohei Kurosaki<sup>d</sup>, Hakan Leblebicioğlu<sup>e</sup>, Roger Hewson<sup>a,b,c,d</sup>

<sup>a</sup> UK Health Security Agency, Science Group, Porton Down, Salisbury, United Kingdom

<sup>b</sup> Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

<sup>c</sup> Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

<sup>d</sup> National Research Centre for the Control and Prevention of Infectious Diseases, Nagasaki University, Japan

<sup>e</sup> VM Medicalpark Samsun Hospital, Samsun, Türkiye



## ARTICLE INFO

## Keywords:

CCHFV  
Targeted enrichment  
Tiling amplicon sequencing  
Next-generation sequencing

## ABSTRACT

Crimean-Congo haemorrhagic fever (CCHF) is the most prevalent human tick-borne viral disease, endemic to the Balkans, Africa, Middle East and Asia. There are currently no licensed vaccines or effective antivirals against CCHF. CCHF virus (CCHFV) has a negative sense segmented tripartite RNA genome consisting of the small (S), medium (M) and large (L) segments. Depending on the segment utilised for genetic affiliation, there are up to 7 circulating lineages of CCHFV. The current lack of geographical representation of CCHFV sequences in various repositories highlights a requirement for increased CCHFV sequencing capabilities in endemic regions. We have optimised and established a multiplex PCR tiling methodology for the targeted enrichment of complete genomes of Europe 1 CCHFV lineage directly from clinical samples and compared its performance to a non-targeted enrichment approach on both short-read and long-read sequencing platforms. We have found a statistically significant increase in mapped viral sequencing reads produced with our targeted enrichment approach. This has allowed us to recover near complete S segment sequences and above 90% of the M and L segment sequences for samples with Ct values as high as 31.3. This study demonstrates the superiority of a targeted enrichment approach for recovery of CCHFV genomic sequences from samples with low virus titre. CCHFV is an important vector-borne human pathogen with wide geographical distribution. The validated methodology reported here adds value to front-line public health laboratories employing genomic sequencing for CCHFV Europe 1 lineage surveillance, particularly in the Balkan and Middle Eastern territories currently monitoring the spread of the pathogen. Tracking the genomic evolution of the virus across regions improves risk assessment and directly informs the development of diagnostics, therapeutics, and vaccines.

## 1. Introduction

Crimean-Congo haemorrhagic fever (CCHF) is the most prevalent human tick-borne viral disease, endemic to the Balkans, Africa, Middle East and Asia (World Health Organization - WHO, 2022). Outbreaks have been reported in 26 countries spanning three continents from the Far East to Southern Europe and West Africa, with Türkiye reporting the highest number of CCHF cases (Belhadi et al., 2022) worldwide. Whilst the average case fatality rate for CCHF in symptomatic patients is between 10% and 40% (Aslam et al., 2016), it can be as high as 80% in

some nosocomial settings (Shahrear and Islam, 2021).

The disease is caused by Crimean-Congo haemorrhagic fever virus (CCHFV), an orthonairovirus within the *Nairoviridae* family (order *Bunyavirales*). The virus is commonly transmitted to humans by a bite of an infected tick, predominantly in the warmer spring and summer months when adult stages of *Hyalomma* ticks and other vector species are seeking their blood meals for maturation (Bente et al., 2013). Transmission through direct contact with viraemic bodily fluids from infected hosts and patients has also been observed, principally among abattoir and healthcare workers. Whilst wild animals, including small

\* Corresponding author at: UK Health Security Agency, Science Group, Porton Down, Salisbury, United Kingdom.

E-mail address: [jake.daddiego@ukhsa.gov.uk](mailto:jake.daddiego@ukhsa.gov.uk) (J. D'Addiego).

<https://doi.org/10.1016/j.jviromet.2023.114833>

Received 26 July 2023; Received in revised form 6 October 2023; Accepted 20 October 2023

Available online 23 October 2023

0166-0934/Crown Copyright © 2023 Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

mammals, birds and larger ungulates are susceptible to CCHFV infection, they do not display clinical symptoms, and act as transient reservoirs for the virus (Bente et al., 2013), for which *Hyalomma* ticks remain the primary reservoirs. Despite ongoing global efforts to develop a vaccine against CCHFV (Tipih and Burt, 2020), there are currently no licensed vaccines or effective treatments against the virus. Although ribavirin has been used to treat CCHF in some endemic regions, data on its clinical efficacy is inconclusive (Hawman et al., 2020), and recent studies have shown only a limited mutagenic effect on CCHFV genomes *in vivo* (D'Addiego et al., 2023).

CCHFV is one of the most genetically diverse arboviruses (Mears and Bente, 2022), with a segmented tripartite negative sense RNA genome consisting of the small (S), medium (M) and large (L) segments encoding respectively the viral nucleoprotein (NP), the precursor to the viral glycoproteins (GPC) and the viral RNA-dependent RNA polymerase (Portillo et al., 2021). The M segment exhibits the highest sequence divergence of 31%, followed by the L segment (22%) and the S segment (20%) (Bente et al., 2013). Depending on the segment utilised for genetic classification, there are 6 (M and L segments) or 7 (S segment) CCHFV lineages (Papa et al., 2022) which are traditionally named after their main geographic location: Africa 1, Africa 2 and Africa 3; Asia 1 and Asia 2, and Europe 1 and Europe 2 (Pickin et al., 2022).

With recent advances in sequencing technologies, including the release of field-friendly and portable sequencers such as the Oxford Nanopore Technologies (ONT) MinION device, whole genome sequencing has become more affordable, and offers scientists the opportunity to characterise the genetic diversity of viruses beyond what could be accomplished by more traditional sequencing approaches.

Of 3777 CCHFV sequences deposited in the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) (Bacterial and Viral Bioinformatics Resource Center), almost 60% belong to viruses from Russia, with Turkish viruses representing only 8.76% of all sequences despite Türkiye being a key hotspot of CCHF disease and reporting the highest number of cases annually. Although CCHFV sequences from other geographical areas are gradually being deposited in public repositories, there is still poor representation of complete CCHFV genome sequences, limiting accurate genetic characterisation of CCHFV in disease endemic regions.

Generally, the recovery of partial viral genome sequencing data has found applications in clinics for some time, such as in the detection of drug resistance markers in viruses like HIV (Hue et al., 2004) and human cytomegalovirus (HCMV) (Houldcroft et al., 2017) for example. The recovery of complete genome sequences from clinical samples, however, is often hampered by low levels of target virus genomes, which although these can be in the region of millions of copies per microlitre, are often overwhelmed by the host's nucleic acids.

Several approaches have been developed for enriching target viral nucleic acids and/or depleting the host's genetic material to increase the amount of generated microbial sequencing reads.

Among the methods focusing on enriching microbial nucleic acids, targeted and non-targeted approaches have been published. One of the major advantages of non-targeted approaches is that no sequence information is required *a priori*, making such methods particularly suitable to study whole microbial communities in environmental samples (Kolundzija et al., 2022; Vlok et al., 2019). Of such non-targeted approaches, Sequence-Independent, Single-Primer Amplification (SISPA) is a primer-initiated method that requires specific sequence modification to achieve non-specific amplification (Reyes and Kim, 1991). SISPA has been successfully utilised to sequence complete or near complete genomes of several viruses including Influenza virus (Lewandowski et al., 2019), SARS-CoV-2 (Gauthier et al., 2021) Lassa virus (Kafetzopoulou et al., 2019), CCHFV and Jingmen tick virus (Brinkmann et al., 2021) directly from clinical samples and tick homogenates.

Though some of the difficulties in recovering complete genome sequences from clinical samples have been addressed by non-targeted approaches (Lewandowski et al., 2019; Gauthier et al., 2021;

Kafetzopoulou et al., 2019; Brinkmann et al., 2021), samples with low virus titre and high cycle-threshold (Ct) values still present a challenge to such methods, with genome coverages reported to decline dramatically above Ct values of 30 (Gauthier et al., 2021).

Polymerase chain reaction (PCR) amplification sequencing is a desirable approach for targeted enrichment due to its high-specificity, relatively low cost, and straight forward assay design (Houldcroft et al., 2017). Multiplex tiling PCR approaches have been published for Tick-Borne Encephalitis Virus (TBEV) (Zakotnik et al., 2022), SARS-CoV-2 (Lin et al., 2021), Zika Virus (ZIKV) (Quick et al., 2017) and recently Respiratory Syncytial Virus (RSV) A and B (Maloney et al., 2022). These schemes work by amplifying fragments across the entire genome of the target virus in alternate primer pools, with a consensus sequence being called in regions with sequencing depths between 1x and 20x depending on the study (Zakotnik et al., 2022; Quick et al., 2017). Compared to non-targeted enrichment strategies, the study by Quick et al. (2017) reported coverages greater than 90% with depths of sequencing in the region of hundreds (nanopore sequencing) or thousands (Illumina sequencing) of reads from samples with Ct values as high as 35.9 (Quick et al., 2017) with a tiling multiplex PCR approach, highlighting the superiority of such approaches for samples with low virus titre. A targeted multiplex PCR enrichment approach has also been successfully utilised to detect emerging SARS-CoV-2 variants of concern (VOC) directly from environmental samples (Lin et al., 2021), providing further insights into additional applications of such methodologies beyond mere genetic virus characterisation.

In this study, we have developed a targeted, multiplex PCR approach for the amplification of Europe 1 CCHFV genomes directly from Turkish clinical samples and have compared its performance to the non-targeted SISPA approach utilising both long read (ONT) and short read (Illumina) sequencing technologies.

## 2. Methods

### 2.1. Primer design for targeted enrichment

Primer pairs targeting 500 bp fragments across complete CCHFV genome segments were designed using the online tool Primal Scheme (Quick et al., 2017). As a reference, unpublished in-house generated S, M and L sequences from clinical isolates of Europe 1 lineage of CCHFV were utilised, most closely related to sequences from a Turkish CCHFV isolate from 2004 (GenBank accession numbers: MF511213, S segment; MF511230, M segment; MF511196, L segment). All designed primer sequences for each pool can be found in [Supplementary Table S1](#) (A and B).

### 2.2. Sample collection

Clinical serum samples were collected during a previous study from patients admitted with confirmed CCHF in Türkiye, between May 2015 and August 2016 at Ondokuz Mayıs University (OMU) Hospital, Tokat State Hospital and Tokat University Hospital, and made available for this work. Original diagnosis was confirmed by positive CCHF PCR or IgM enzyme-linked immunosorbent assay. All participants provided written consent with research ethics committee approval from the UK Ministry of Defense Research and Ethics Committee (MoDREC 571) and Ondokuz Mayıs University Research and Ethics Committee (OMU KAEK 2014/739), Türkiye.

### 2.3. Sample processing and RNA extraction

Serum samples were inactivated in a Containment Level 4 (CL4) laboratory; 140 µl of each sample were added to 560 µl of buffer AVL and 560 µl of ethanol. RNA was extracted utilising the QIAamp Viral RNA Mini Kit (Cat. No. 52904, Qiagen) following the manufacturer's instructions. RNA was eluted in 60 µl of nuclease-free water and stored at  $-80^{\circ}\text{C}$ .

#### 2.4. Real-time PCR

All extracted RNAs were quantified in duplicate utilising an in-house real-time PCR assay as previously described (D'Addiego et al., 2023). Briefly, 5 µl of extracted RNA were amplified in final 20 µl reactions containing 0.5 µM of forward primer CCHF-F1 (5'-CCC TTT TTA AAC TCC TCA AAC C-3'), 0.5 µM of forward primer CCHF-F2 (5'-CCT TTT TTA AAC TCC TCA AAC C-3'), 1 µM of reverse primer CCHF-R (5'- TCT CAA AGA AAC ACG TGC C-3'), 0.5 µM of probe CCHF-P (5'-FAM-ACT CAA GAG AAC ACT GTG GGC GTA AG-3'-MGBEQ) with TaqMan™ Fast Virus 1-Step Master Mix (cat. no. 4444436, Thermo Fisher Scientific). Primers/Probe binding sites are provided in [Supplementary Table S3](#). Copies of CCHFV genomic RNA in the extracts were determined based on quantified synthetic S Segment RNA standard curves. Quantitative real-time RT-PCR was performed on a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems) platform with the following cycling parameters: 50 °C for 10 min, 95 °C for 2 min and 45 cycles of 95 °C for 10 s and 60 °C for 30 s.

#### 2.5. Target enrichment

cDNA from extracted RNA was synthesised using SuperScript™ IV First-Strand Synthesis System (Thermo Fisher Scientific, cat. 18091050) in final 20 µl reactions. Briefly, 11 µl of extracted RNA were mixed with 50 ng of random hexamers (50 ng/µl) and 0.5 mM dNTP mix and incubated at 65 °C for 5 min. RNA and annealed primer mixes were then placed on ice for 1 min, then mixed with 1X SSIV Buffer, 5 mM DTT, 1 µl of RNase Out™ (Invitrogen, cat. no. 100777019) and 200 units of SSIV RT enzyme before being incubated at 42 °C for 50 min and 70 °C for 10 min. Two aliquots of cDNA were prepared for each sample. cDNA was stored at -20 °C.

Target enrichment was based on a previously published protocol (Quick, 2020). Two separate primer pools for each CCHFV genome segment were prepared, each containing primers targeting alternate fragments. Details of each primer pool with final primers concentrations are provided in [Supplementary Table S2](#). 2.5 µl of cDNA were amplified in triplicate 25 µl reactions containing either one of two sets of primer pools utilising Q5 High-Fidelity 2X Master Mix (NEB, cat. no. M0492L). A maximum of three replicate reactions were prepared for each sequencing platform to achieve the recommended concentration for sequencing library preparation. Master-mix composition is provided in [Supplementary Table S2](#). cDNA was amplified in a thermocycler with the following cycling parameters: 98 °C for 30 s, 30 cycles of 98 °C for 15 s, 57 °C for 30 s and 70 °C for 30 s and held at 4 °C. PCR amplicons were cleaned up with AMPure XP Reagent (Beckman Coulter, cat. no. A63881) and stored at -20 °C.

#### 2.6. Sample preparation for metagenomic sequencing

22 µl of extracted RNA were treated with TURBO™ DNase (Thermo Fisher Scientific, cat. no. AM2239) in final 50 µl reactions. RNA was mixed with 0.8X TURBO™ DNase buffer and 2 units of TURBO™ DNase enzyme and incubated at 37 °C for 30 min.

DNase-treated RNA was purified using the RNA Clean and Concentrator-5 kit (Zymo Research, cat. no. R1016). Randomly amplified cDNA was prepared using a SISPA approach according to a previously described protocol (Lewandowski et al., 2019). For round A, 4 µl of cleaned-up RNA were annealed with 1 µl of 40 µM Sol-A primer (5'-GTTTCCCACTGGAGGATANNNNNNNNN-3') at 65 °C for 5 min, then incubated at 21 °C for 5 min. First strand cDNA was synthesised in 10 µl reactions containing 1X SSIV Buffer, 1 mM of dNTPs, 10 U of SSIV reverse-transcriptase, 5 mM of DTT and 5 µl of annealed RNA and incubating at 42 °C for 10 min. Second strand cDNA was synthesised utilising Sequenase™ Version 2.0 DNA polymerase (cat. no. 70775Y200UN, Thermo Fisher Scientific) by adding 5 µl of 1X Sequenase buffer and 0.39 units of Sequenase Enzyme followed by an

incubation at 37 °C for 8 min, then by adding 0.6 µl of 3.25 units of Sequenase enzyme diluted in Sequenase dilution buffer and incubating further at 37 °C for 8 min. For round B, samples were amplified with AccuTaq™ LA DNA Polymerase (Merck, cat. no. D8045) in 50 µl reactions containing 1X reaction buffer, 2 µM Sol-B primer (5'-GTTTCCCACTGGAGGATA-3'), 1 µl of DMSO, 0.5 mM dNTPs and 2.5 units of AccuTaq enzyme. Reactions were incubated at 98 °C for 30 s followed by 30 cycles of 94 °C for 15 s, 50 °C for 20 s and 68 °C for 2 min and by a final extension cycle at 68 °C for 10 min. PCR products were cleaned up with AMPure XP Reagent and eluted with molecular grade water in final 30 µl volumes.

#### 2.7. Sequencing library preparation

PCR amplicons and SISPA B products were cleaned up with AMPure XP Reagent with a 2:1 beads to sample ratio following the manufacturer's protocol, and quantified with Qubit™ fluorometer's dsDNA High-Sensitivity assay (Thermo Fisher Scientific, cat. no. Q32854) following the manufacturer's protocol.

Illumina sequencing libraries were prepared from 1.65 ng cDNA using Nextera XT V2 kit (Illumina, cat. no. FC-131-1096) and sequenced on a 2 × 150 bp paired-end Illumina MiSeq instrument operated by the UK Health Security Agency (UKHSA) Genomics Services Development Unit (Colindale).

Barcoded MinION sequencing libraries were prepared from 65 ng of cDNA with the Ligation Sequencing kit SQK-LSK-110 and Native Barcoding Kit EXP-NBD196 (ONT). End repair and dA tailing was performed with NEBNext® Ultra™ II End Repair/dA-Tailing Module (New England Biolabs - NEB, cat. no. E7546L) in 30 µl reactions containing 65 ng of cDNA, 3.5 µl of Reaction Buffer and 1.5 µl of Enzyme Mix and incubating at room temperature for 10 min followed by an incubation at 65 °C for 10 min. Ligation of ONT specific barcodes/adaptors was done by adding 1X Blunt/TA Ligase Master Mix (cat. no. M0367L, NEB) and 2.5 µl of barcode directly to the end-repaired reactions followed by an incubation at room temperature for 20 min and at 65 °C for 10 min. Barcoded samples were then cleaned-up in three pools containing a maximum of 22 samples each and normalised to 65 ng in 45 µl. Adaptor ligation was performed with NEBNext® Ultra™ II Ligation Module (NEB, cat. E7595L) by adding 5 µl of adaptor, 40 µl of Ligation Master Mix and 1 µl of Enhancer (NEB) directly to the barcoded library and incubating at room temperature for 30 min. Adaptor-ligated libraries were cleaned-up and washed with Short Fragment Buffer (SFB) (SQK-LSK-110 kit) and eluted in 15 µl of Elution Buffer before final quantification.

Libraries were sequenced for 24 h on FLO-MIN106D flow cells using a Mk1C MinION device (ONT).

#### 2.8. Bioinformatic analysis

Raw Illumina sequencing reads were processed with Trimmomatic using the following parameters: ILLUMINACLIP: NexteraPE-PE.fa:2:30:10:8 LEADING:30 TRAILING:30 SLIDINGWINDOW:10:20 MINLEN:50. Further adaptor trimming and quality filtering were performed with BBDuk with the following settings: -Xmx1g in1 =R1.fastq.gz in2 =R2.fastq.gz out1 =R1\_clean.fastq.gz out2 =R2\_clean.fastq.gz ref=nextera.fa.gz ktrim=r k = 23 mink= 11 hdist= 1 tpe tbo qtrim=r1 trim= 20. Illumina reads were mapped against a reference genome (accession numbers OQ866631-OQ866633 for the S, M and L segments respectively) utilising BWA MEM with kmer length set to 35. Nanopore sequencing data was filtered based on length (0.4-0.6 kb) and quality (minimum 8) directly on the MinION data output options. Nanopore sequencing reads were mapped against the reference utilising minimap2 (Li, 2018). Reference coverage and sequencing depth statistics were derived utilising SAMtools (Danecek et al., 2021) depth, coverage and flagstat functions.

Consensus sequences were generated for a subset of samples with Ct

values ranging between 17.5 and 29.9. Sequences with identified frameshifts were excluded from downstream phylogenetic analysis. For targeted enrichment Illumina data, consensus sequences were derived from sorted BAM files with iVar (Grubaugh et al., 2019a). Consensus sequences from Illumina metagenomic data were derived from sorted BAM files using an in-house C++ program QuasiBAM (Penedos et al., 2015), with minimum sequencing depth for consensus calling set to 10x.

For ONT data, sequencing reads were first quality filtered and trimmed using Filtlong (2017) prior to mapping. Consensus sequences from metagenomic data were derived from sorted BAM files utilising BCFtools (Danecek et al., 2021), masking regions with coverage below 10x. For targeted enrichment data, primer sequences were first removed from sorted BAM files with BAMClipper (Au et al., 2017) prior to consensus generation with BCFtools (Danecek et al., 2021) as described previously. All generated consensus sequences have been deposited in GenBank with the following accession numbers: OQ454691-OQ454750 (Illumina sequencing data) and OQ454777-OQ454836 (ONT sequencing data).

Pairwise identities were calculated with Geneious Prime® (version 2023.0.2).

## 2.9. Phylogenetic analysis

Maximum likelihood phylogenies were generated with MEGA (Kumar et al., 2016) (version 7.0.26) applying the general time reversible model. Sequences were aligned using MUSCLE (Edgar, 2004). Alignments were trimmed to exclude low coverage regions to a final length of 1572 bp (S segment), 3057 bp (M segment) and 5102 bp (L segment). Bootstrap support values were generated with 1000 replicates.

## 2.10. Statistical analysis

Statistical significance was determined with a T Test using Microsoft Office's Excel software. Reported *p* values are significant at the 5% level.

## 3. Results

### 3.1. Extracted RNA samples presented with a wide range of cycle threshold (Ct) values

Extracted RNA from a total of 23 CCHFV positive serum samples was quantified utilising an in-house real-time PCR assay. Average Ct values and RNA quantities expressed as copies per millilitre of serum are presented in Table 1.

The Ct values ranged between 17.5 and 34.8, with one sample having an undetermined result. The recorded RNA quantities were between  $5.06 \times 10^3$  and  $5.12 \times 10^8$  copies per millilitre of serum (Table 1). Input number of genome copies for each enrichment methodology ranged between  $2.6 \times 10^2$  and  $2.6 \times 10^7$  genome copies.

**Table 1**

Ct values and viral genome copies per millilitre of patient serum for each extracted RNA sample.

Sample ID	Ct value	Copies/ml	Sample ID	Ct value	Copies/ml
1	25.7	$1.76 \times 10^6$	16	25.4	$2.28 \times 10^6$
2	26	$1.41 \times 10^6$	18	33.8	$1.04 \times 10^4$
4	31	$4.48 \times 10^4$	19	34.8	$5.06 \times 10^3$
5	29.4	$1.43 \times 10^5$	21	21.2	$5.96 \times 10^7$
6	29.9	$9.57 \times 10^4$	22	22.4	$2.63 \times 10^7$
7	31.3	$3.63 \times 10^4$	23	25.7	$2.76 \times 10^6$
8	Undetermined		25	25.7	$2.72 \times 10^6$
10	25.8	$1.69 \times 10^6$	26	27	$9.85 \times 10^5$
12	17.5	$5.12 \times 10^8$	28	23.5	$1.27 \times 10^7$
13	22	$2.3 \times 10^7$	29	28.3	$5.01 \times 10^5$
14	22.2	$2.06 \times 10^7$	30	29	$2.67 \times 10^5$
15	24	$5.73 \times 10^6$			

### 3.2. Targeted enrichment produced a statistically significant increase in mapped sequencing reads

Illumina sequencing generated a total of 39.2 M sequencing reads which passed quality control. Targeted enrichment generated on average 25.5% more sequencing reads compared to the non-targeted SISPA approach ( $P < 0.05$ ) with a total of 22.5 M and 16.8 M sequencing reads respectively. A statistically significant difference in mapped sequencing reads was also found between the two methods ( $P < 0.05$ ). SISPA generated on average 69974 mapped sequencing reads per sample, ranging between 4 and 748230 reads (Fig. 1A). Targeted enrichment produced on average 608511 mapped sequencing reads per sample, ranging between 124535 and 944278 reads (Fig. 1B).

Oxford Nanopore Technologies (ONT) sequencing generated 2.6 M reads which passed quality and length filters. The total reads generated with targeted enrichment and SISPA was 53188 and 51412 respectively ( $P = 0.81$ ). A statistically significant difference in mapped sequencing reads was found between the two methodologies ( $P < 0.05$ ). SISPA produced on average 7356 mapped sequencing reads per sample, ranging between 4004 and 105325 reads (Fig. 1C). Targeted enrichment produced on average 48145 mapped sequencing reads per sample, ranging between 11019 and 100638 reads (Fig. 1D).

### 3.3. Near complete genome reference coverages with increased sequencing depths were obtained with targeted enrichment

Illumina sequencing data coupled with SISPA produced, on average, reference genome coverages of 68.79%, 53.03%, 34.99% and 10.94% with ranges of 1%–100%, 0%–100%, 0%–99.96% and 0%–98.97% at 1x, 10x, 100x and 1000x sequencing depths respectively across the entire CCHFV genome (Fig. 2A). With targeted enrichment, the average reference genome coverages were 93.56%, 86.04%, 78.91% and 63.02% with ranges of 54.45%–99.74%, 19.77%–98.20%, 16.23%–95.33% and 13.15%–87% at 1x, 10x, 100x and 1000x sequencing depths respectively (Fig. 2B).

ONT sequencing coupled with SISPA produced on average reference genome coverages of 84.71%, 57.26%, 31.58% and 3.12% with ranges of 5%–100%, 0%–100%, 0%–99.77% and 0%–35.34% at 1x, 10x, 100x and 1000x sequencing depths respectively (Fig. 2C). With targeted enrichment, the average reference genome coverages were 98.29%, 92.06%, 74.95% and 42.79% with ranges of 94.59%–99.97%, 63.68%–98.12%, 6.87%–91.63% and 0%–75.11% at 1x, 10x, 100x and 1000x sequencing depths respectively (Fig. 2D).

### 3.4. Targeted enrichment produced on average a 6.55-fold increase in sequencing depth across all CCHFV genomic segments

The average sequencing depth across all samples recorded with Illumina sequencing following SISPA was 778.61, 384.7 and 401.68 for the S, M and L segment respectively (Fig. 3A). With targeted enrichment, the average sequencing depths were 4305.73, 2283.64 and 3194.36 (Fig. 3B) for the S, M and L segments respectively, with an average 6.25-fold increase in sequencing depth across all segments.

ONT sequencing data coupled with SISPA produced average sequencing depths of 301.45, 126.6 and 130.94 for the S, M and L segments respectively (Fig. 3C). Targeted enrichment produced sequencing depths of 2165.79, 717.20 and 1243.32 (Fig. 3D) for the S, M and L segments respectively, with an average 7.38-fold increase in sequencing depth across all segments.

### 3.5. Phylogenetic analysis of the generated consensus sequences for each genomic segment confirms genetic affiliation to the Europe 1 CCHFV lineage

Generated consensus sequences clustered with previously published Europe 1 CCHFV sequences (Fig. 4). For samples 6, 22, 23 and 29,



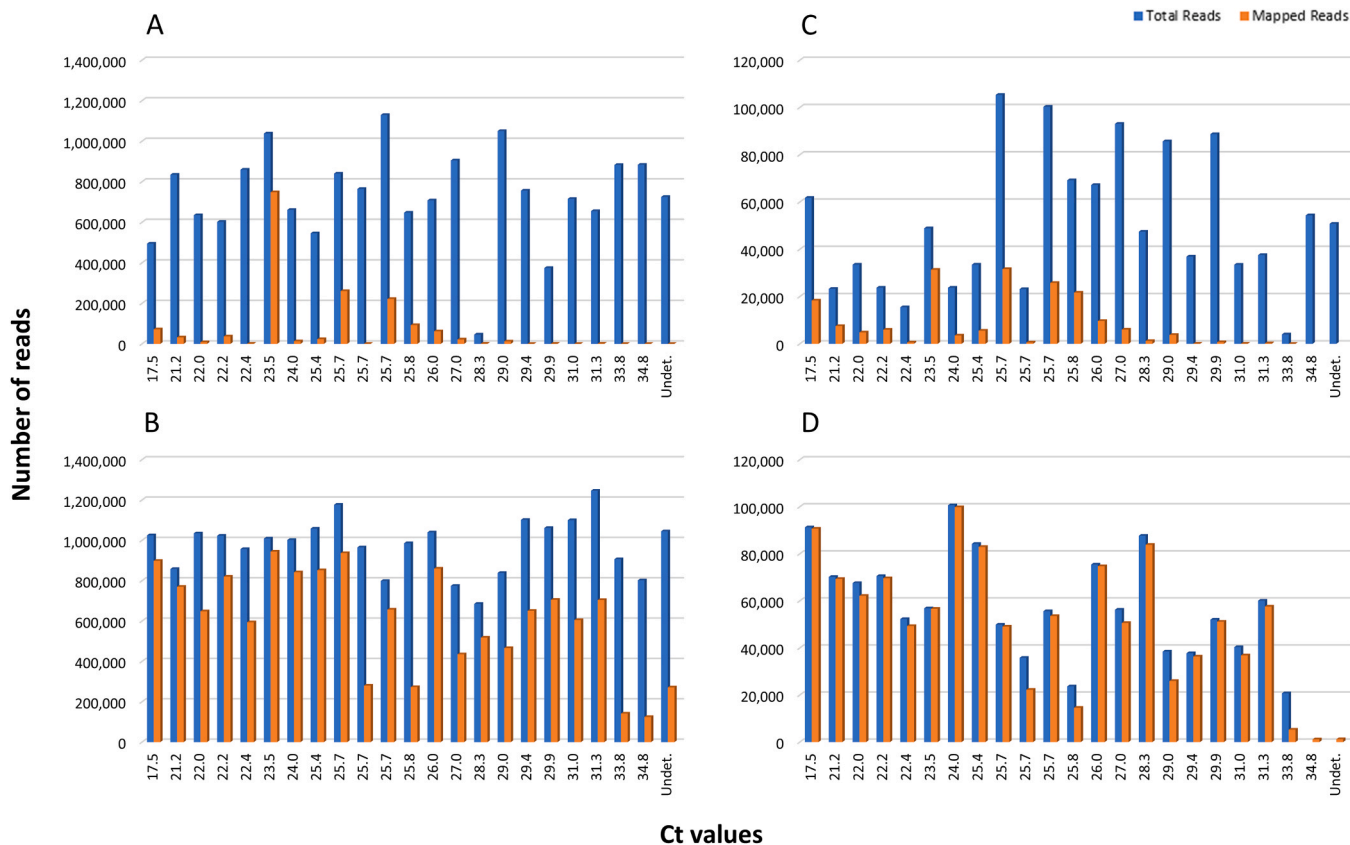


Fig. 1. Total (blue bars) and mapped (orange bars) sequencing reads for (A) SISPA and (B) targeted enrichment with Illumina sequencing and (C) SISPA and (D) targeted enrichment with nanopore sequencing.

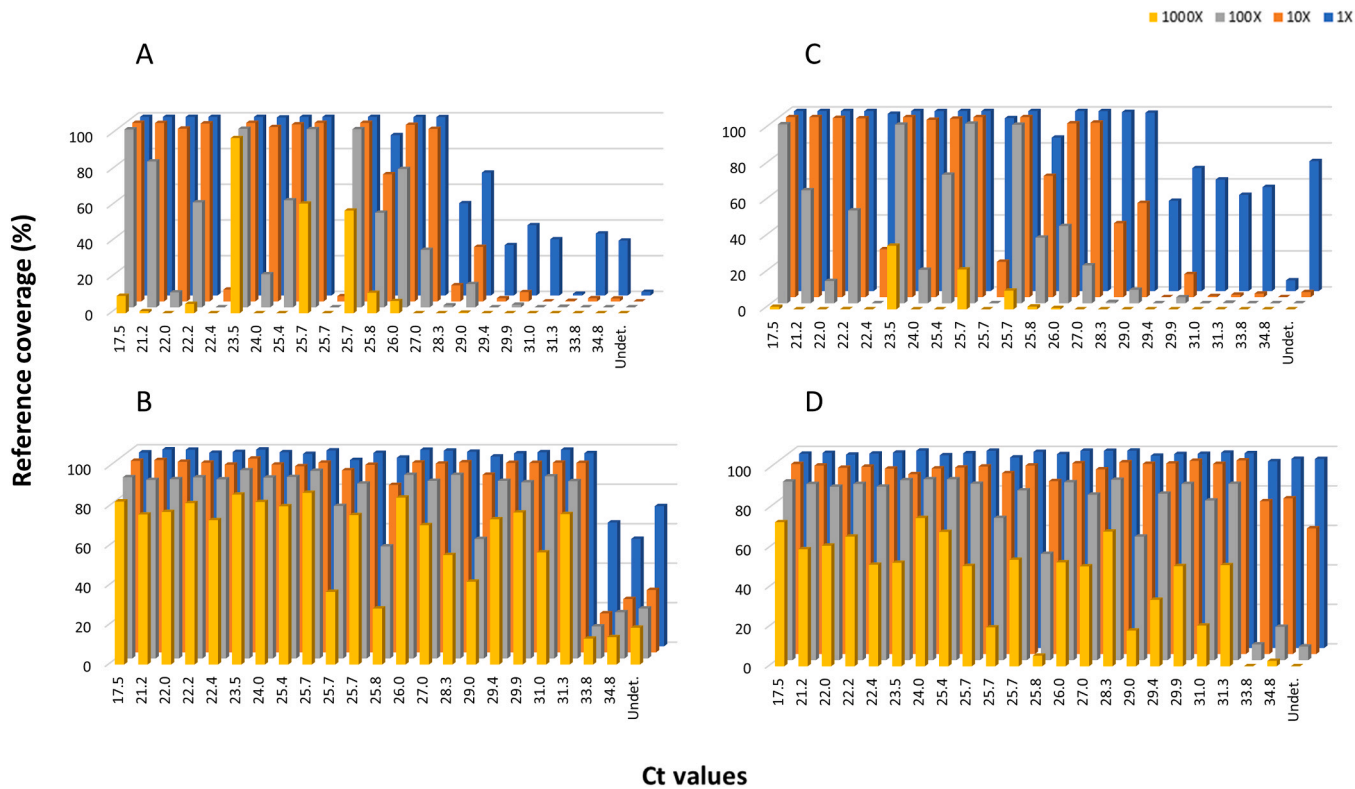
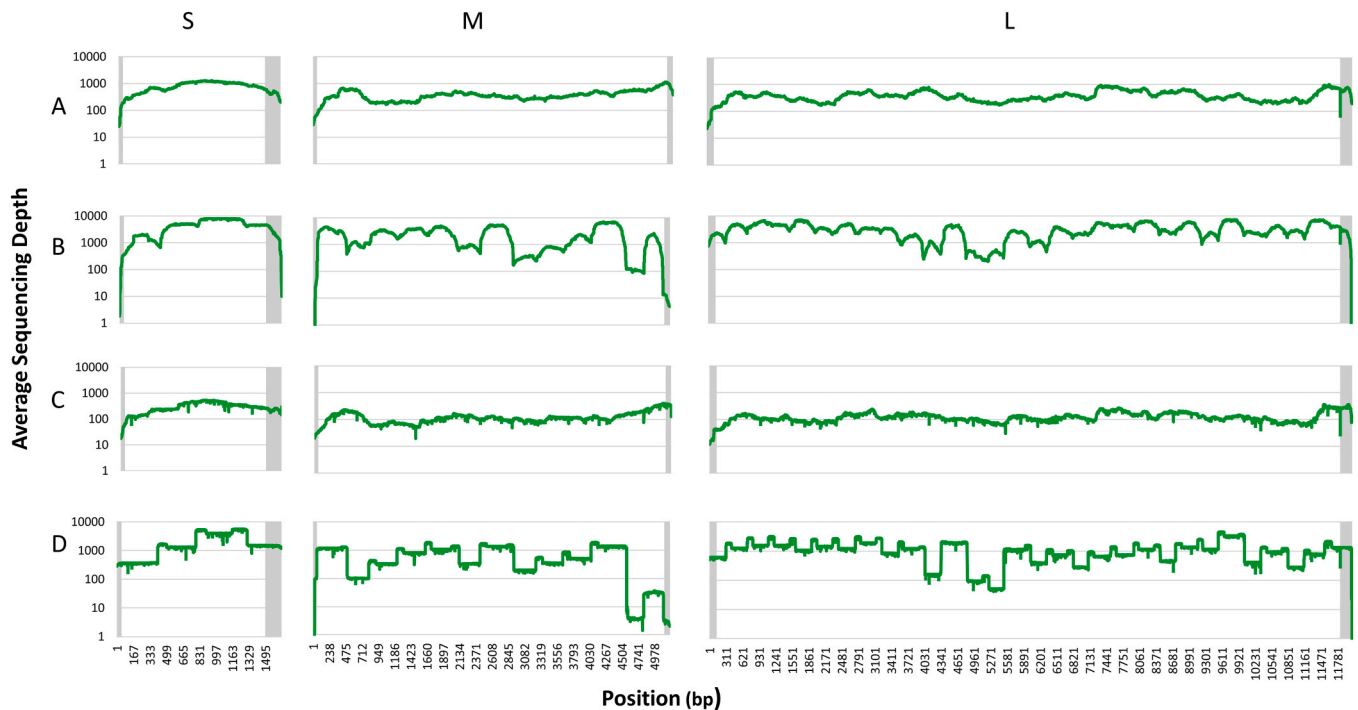


Fig. 2. Reference coverage at sequencing depth thresholds of 1x (blue), 10x (orange), 100x (grey) and 1000x (yellow) for (A) SISPA and (B) targeted enrichment with Illumina sequencing, and for (C) SISPA and (D) targeted enrichment with nanopore sequencing.



**Fig. 3.** Average sequencing depth (shown in green) for the Small (S), Medium (M) and Large (L) CCHFV genome segments for (A) SISPA and (B) targeted enrichment with Illumina sequencing and for (C) SISPA and (D) targeted enrichment with nanopore sequencing. 3' and 5' UTR regions are indicated in grey.

insufficient sequencing depth and reference coverage were achieved with the SISPA approach to generate consensus sequences.

The average alignment pairwise distances of the generated sequences were 0.81% (S segment), 1.3% (M segment) and 0.6% (L segment), with overall pairwise distances of 11.5%, 19.3% and 7.3% for the S, M and L segments respectively. Pairwise identities between generated consensus sequences for each sample were on average 99.99%, 98.72% and 99.39% for the S, M and L segments respectively, with an overall average pairwise identity of 99.36%.

#### 4. Discussion

Bioinformatic analysis of viral sequencing data relies on either mapping sequencing reads against a reference genome (reference-based assembly) or looking for overlaps between the reads to generate larger fragments (*de novo* assembly) (Khan et al., 2018), followed by extraction of a consensus sequence which is often based on majority voting (Maurier et al., 2019). Different sequencing platforms have varying levels of error rates, which when taken on their own can resemble genuine variants; nonetheless having several reads mapping to the same region increases confidence in base calling and support for minor variants (Sims et al., 2014). Very high sequencing depths are generally not required for consensus calling, with most bioinformatic pipelines using cut-offs of 1x (Li et al., 2009) or 10x (Castellano et al., 2021) sequencing depths; however for other downstream applications such as single nucleotide variant (SNV) calling, depths of 400x have been recommended with a multiplex PCR approach (Grubaugh et al., 2019b), further supporting the requirement of targeted enrichment approaches for samples with low virus titre. Whilst metagenomic approaches have been vastly utilised for virus discovery and molecular surveillance, the large proportion of host sequencing reads generated from clinical samples makes recovery of complete genome sequences challenging with such non-targeted approaches.

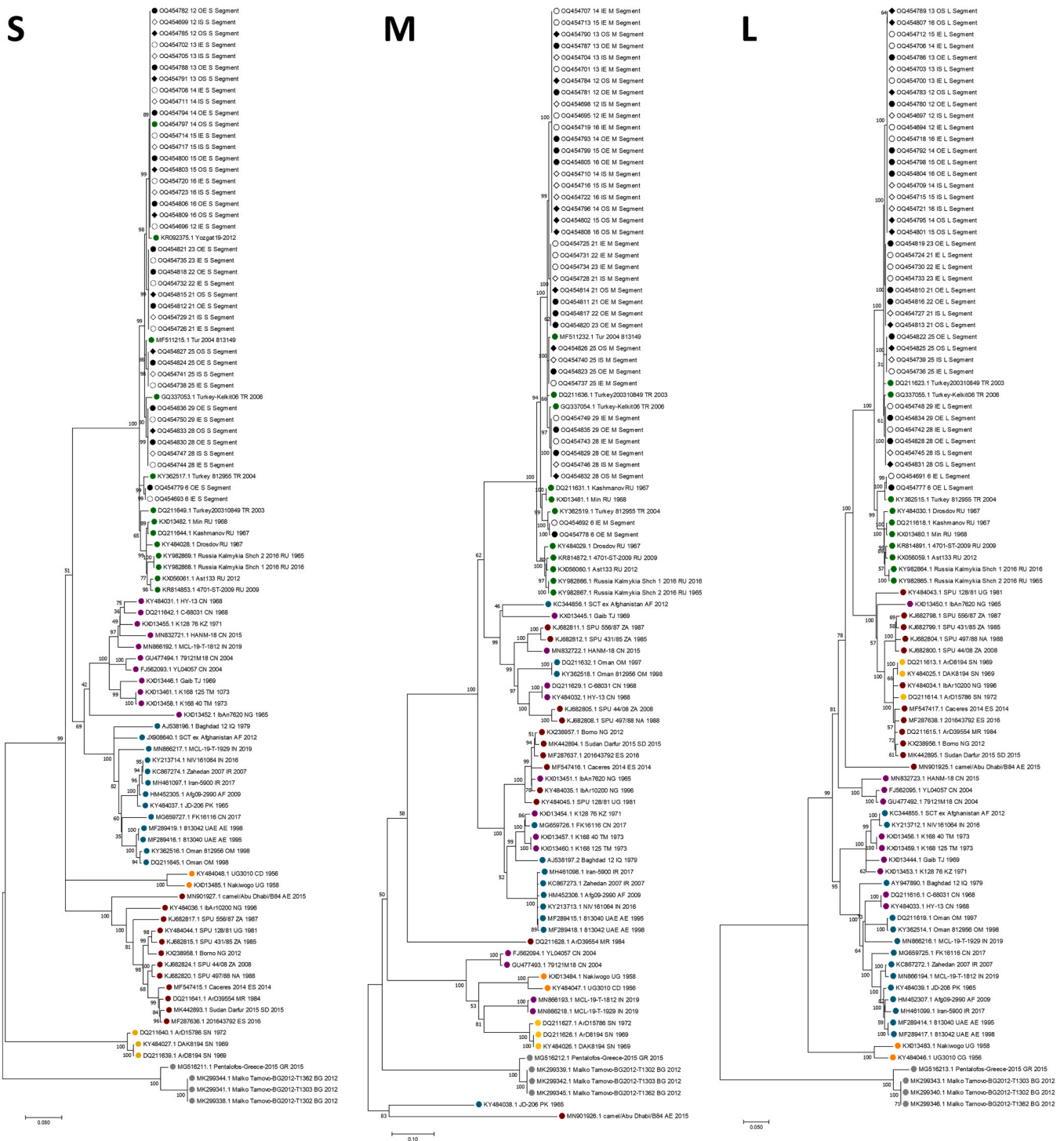
Here, we adapted an amplicon tiling strategy previously reported by Quick et al. (2017) for sequencing Europe 1 (or genotype V) CCHFV which has produced a statistically significant increase in mapped

sequencing reads with both Illumina and nanopore technologies ( $P < 0.05$ , Fig. 1). Whilst Illumina technology also resulted in a statistically significant difference in total sequencing reads generated by the two enrichment methodologies possibly due to the different number of PCR cycles between the different methods ( $P < 0.05$ ), nanopore sequencing did not ( $P = 0.81$ ). The lack of a significant difference in total generated sequencing reads with ONT technology is likely due to the limited amount of nanopores available for sequencing which become depleted over time and are therefore only able to produce a finite amount of sequencing reads in a single run, suggesting that higher sequencing depths could be achieved by running fewer samples in a single flow-cell.

Kafetzopoulou et al. (2018) have found that whilst a higher proportion of target viral reads are generally generated for samples with lower Ct values, genome reference coverages above 90% of at least 20x sequencing depths were also observed for samples with Ct values as high as 30. The study by Lewandowski et al. (2019) however, reported that only 63% of samples with a Ct value below 30 produced > 90% complete genomes, with reference coverage dropping dramatically between Ct values of 25 and 30 (Lewandowski et al., 2019), suggesting that significant variation is observed between different sample types.

In our study, with SISPA enrichment coupled with both Illumina and nanopore sequencing, genome coverages > 90% with at least 10x sequencing depths were achieved only for samples with Ct values below 27, with reference genome coverages declining beyond this value (Fig. 2A and C). In contrast our targeted enrichment approach showed greater than 90% reference genome coverages with average sequencing depths of at least 10x were achieved with both sequencing platforms up to Ct value 31.3 (Fig. 2B and D), supporting the superiority of a targeted enrichment strategy for samples with titres as low as few thousand copies per millilitre of serum.

Fig. 2 shows an increase in reference genome coverages also at 100x and 1000x sequencing depths, which presented with average reference sequence coverage increases of 43.92% and 52.08% for 100x and 1000x depths respectively with Illumina and increases of 44.68% and 38.81% for 100x and 1000x depths respectively with nanopore sequencing.



**Fig. 4.** Maximum likelihood phylogeny of the S (1572 bp), M (3057 bp) and L (5102 bp) segments. Sequencing technology and enrichment strategy are indicated for each sequence as follows: IE (Illumina, targeted enrichment) with a white circle, IS (Illumina, SISPA) with a white rhombus, OE (ONT, targeted enrichment) with a black circle and OS (ONT, SISPA) with a black rhombus. Different genetic lineages are indicated as follows: Asia 1/IV (teal), Asia 2/IV (purple), Europe 1/V (green), Africa 3/III (red), Africa 2/II (orange), Africa 1/I (yellow), AIGV/VI (grey).

Whilst not strictly required for consensus generation, the study by Liu et al. (2022) has shown that the false positive rate for variant calling utilising ONT data drops to zero when a minimum sequencing depth of 200x is used, suggesting that higher sequencing depths are preferable to minimise the introduction of errors in generated consensus sequences.

The S and L segments exhibited on average the highest sequencing depths across both sequencing platforms and methodologies (Fig. 3). Although the reduced sequencing depth in the M segment compared to

the S and the L segments for the targeted enrichment can be attributed to the greater genetic diversity of the M segment, the same cannot be said with the non-targeted approach, suggesting perhaps a lower starting number of M segment copies in the extracted viral RNA, possibly due to the presence of defective M-segment free particles, speculated on previously (Weidmann et al., 2011).

Additional reference coverage could have been achieved by lowering the minimum depth required for consensus calling to 1x, as done by

Zakotnik et al. in their TBEV study, however, we have maintained a minimum depth of 10x to increase confidence.

Although we have observed a loss of sequencing depth below 10x in certain regions of the M and L segments (Fig. 3), our phylogenetic analysis suggests that enough data is retained for accurate lineage discrimination even with partial M and L segment sequences (Fig. 4), which confirmed affiliation of our generated consensus sequences to the Europe 1 lineage, supported by high bootstrap values. Similar to other reports (Lukashev et al., 2016), we have found no evidence of genomic inter-lineage reassortment in our Europe 1 sequences based on phylogenetic analysis. Genetic lineages Asia 1 and Asia 2 (genotype IV) and Africa 3 (genotype III), however, do not form distinct clades by M segment phylogenetic analysis (Fig. 4B), with evidence of genetic reassortment between these lineages having been described previously (Lukashev et al., 2016; Negredo et al., 2021), highlighting the importance of a whole genome sequencing approach for accurate genetic characterisation of CCHFV.

Ticks are important vectors of a wide range of viral pathogens affecting humans, livestock and wild animals. The last decades have seen a rise in tickborne disease (TBD) cases, and established populations of medically important tick species have been reported over expanding geographic areas (Madison-Antenucci et al., 2020). Although part of the increased incidence of tick-borne pathogens can be attributed to advances in diagnostics, increased global travel, trading of livestock and environmental changes have all contributed to the spread of tick-borne pathogens in Europe (Michelet et al., 2014). Evidence of a northward expansion of the range of several tick species has been observed in Europe and the Americas (Jore et al., 2014; Porretta et al., 2013; Estrada-Pena, 2023), with concomitant emergence of new endemic foci (Holding et al., 2020; Monsalve Arteaga et al., 2021). The geographical expansion of tick vectors coupled with the fast mutation rate of RNA viruses (Damian et al., 2020) has made the investigation of tick-borne pathogens a public health priority. Research into tick-borne pathogens will be crucial to not only inform disease prevention and control measures (Johnson et al., 2022), but also to understand their diversity, host and vector ranges, evolutionary histories, as well predict the emergence of new pathogenic species (Damian et al., 2020).

The SARS-CoV-2 pandemic has, now more than ever, highlighted the threat that rapidly evolving RNA viruses pose to public health. With emerging VOCs having accounted for immune response evasion (Harvey et al., 2021), increased transmissibility (Balint et al., 2022) and increased virulence (Wu et al., 2021), it is clear that complete viral genome sequencing will continue to play an important role in the prevention, control and management of infectious diseases.

Since there is still no licensed vaccine against CCHFV, and the efficacy of ribavirin is still questionable (D'Addiego et al., 2023; Johnson et al., 2018), there has been increased interest by public health authorities to support active CCHFV surveillance. The WHO has included CCHFV in their 2018 R&D Blueprint as a priority for research and product development for early diagnostics (World Health Organization, 2013).

This study represents the first developed tiling multiplex PCR methodology for CCHFV, for which we have observed a comparable performance to other published schemes. Whilst recovery of a complete genomic sequence from samples with high virus titre is possible utilising non-targeted enrichment strategies such as SISPA, for samples with low viraemia, a targeted enrichment approach is preferable prior to sequencing to recover meaningful portions of the genome. In this study, we were able to recover 100% of the CCHFV genome in samples with a minimum of  $1.41 \times 10^6$  copies/ml of serum, with reference genome coverage dropping to at least 53.46% for samples with Ct values above 30.

Future work will look at further optimising the presented targeted enrichment approach. In particular, we will aim to increase the Ct value threshold for successful sequencing and consensus calling. Efforts will also be made to identify additional primers that can be added to the

scheme to increase genome reference coverage. The scheme will also be tested on a broader set of samples in other endemic regions, as well as on other complex sample types such as tick homogenates. Whilst the development of a single multiplex PCR strategy for enrichment of all CCHFV genomes is hindered by the high genetic diversity between CCHFV clades, future efforts will also look at developing enrichment schemes for other lineages, to further support sequencing capabilities of CCHFV in endemic regions.

#### CRedit authorship contribution statement

**Jake D'Addiego:** Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Nadina Wand:** Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Writing – review & editing. **Babak Afrough:** Conceptualization, Methodology, Writing – review & editing. **Tom Fletcher:** Data curation. **Yohei Kurosaki:** Conceptualization, Methodology. **Hakan Leblebicioğlu:** Data curation. **Roger Hewson:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Writing – review & editing.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

The study was funded by the Newton Fund: Institutional Links programme (grant number 527597374) through the British Council and TUBITAK (Türkiye) - Katip Çelebi between UK and Türkiye (grant number 119N515), the Royal Society (grant number IES\R2\212130) and Grant-in-Aid UK Health Security Agency (grant number 111752). The views expressed are those of the authors and not necessarily those of the funding bodies.

#### Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jvromet.2023.114833.

#### References

- Aslam, S., Latif, M.S., Daud, M., Rahman, Z.U., Tabassum, B., Riaz, M.S., et al., 2016. Crimean-Congo hemorrhagic fever: Risk factors and control measures for the infection abatement. *Biomed. Rep.* 4 (1), 15–20.
- Au, C.H., Ho, D.N., Kwong, A., Chan, T.L., Ma, E.S.K., 2017. BAMClipper: removing primers from alignments to minimize false-negative mutations in amplicon next-generation sequencing. *Sci. Rep.* 7 (1), 1567.
- Bacterial and Viral Bioinformatics Resource Center. Accessed on 22/12/2022. Available from: (<https://www.bv-brc.org/>).
- Balint, G., Voros-Horvath, B., Szechenyi, A., 2022. Omicron: increased transmissibility and decreased pathogenicity. *Signal Transduct. Target. Ther.* 7 (1), 151.
- Belhadi, D., El Baied, M., Mulier, G., Malvy, D., Mentre, F., Laouenan, C., 2022. The number of cases, mortality and treatments of viral hemorrhagic fevers: a systematic review. *PLoS Negl. Trop. Dis.* 16 (10), e0010889.
- Bente, D.A., Forrester, N.L., Watts, D.M., McAuley, A.J., Whitehouse, C.A., Bray, M., 2013. Crimean-Congo hemorrhagic fever: history, epidemiology, pathogenesis, clinical syndrome and genetic diversity. *Antivir. Res.* 100 (1), 159–189.
- Brinkmann, A., Uddin, S., Krause, E., Surtees, R., Dincer, E., Kar, S., et al., 2021. Utility of a sequence-independent, single-primer-amplification (SISPA) and nanopore sequencing approach for detection and characterization of tick-borne viral pathogens. *Viruses* 13 (2).
- Castellano, S., Cestari, F., Faglioni, G., Tenedini, E., Marino, M., Artuso, L., et al., 2021. iVar, an interpretation-oriented tool to manage the update and revision of variant annotation and classification. *Genes* 12 (3).
- D'Addiego, J., Elaldi, N., Wand, N., Osman, K., Bagci, B.K., Kennedy, E., et al., 2023. Investigating the effect of ribavirin treatment on genetic mutations in Crimean-Congo haemorrhagic fever virus (CCHFV) through next-generation sequencing. *J. Med. Virol.*



- Damian, D., Maghembe, R., Damas, M., Wensman, J.J., Berg, M., 2020. Application of viral metagenomics for study of emerging and reemerging tick-borne viruses. *Vector Borne Zoonotic Dis.* 20 (8), 557–565.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., et al., 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10 (2).
- Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* 5, 113.
- Estrada-Pena, A., 2023. The climate niche of the invasive tick species *Hyalomma marginatum* and *Hyalomma rufipes* (Ixodidae) with recommendations for modeling exercises. *Exp. Appl. Acarol.*
- Filtong, Wick R., 2017 [updated 30/07/2021]. Available from: (<https://github.com/rrwick/Filtong>).
- Gauthier, N.P.G., Nelson, C., Bonsall, M.B., Locher, K., Charles, M., MacDonald, C., et al., 2021. Nanopore metagenomic sequencing for detection and characterization of SARS-CoV-2 in clinical samples. *PLoS One* 16 (11), e0259712.
- Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., et al., 2019a. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* 20 (1), 8.
- Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., et al., 2019b. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* 20.
- Harvey, W.T., Carabelli, A.M., Jackson, B., Gupta, R.K., Thomson, E.C., Harrison, E.M., et al., 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19 (7), 409–424.
- Hawman, D.W., Haddock, E., Meade-White, K., Nardone, G., Feldmann, F., Hanley, P.W., et al., 2020. Efficacy of favipiravir (T-705) against Crimean-Congo hemorrhagic fever virus infection in cynomolgus macaques. *Antivir. Res.* 181, 104858.
- Holding, M., Dowall, S.D., Medlock, J.M., Carter, D.P., Pullan, S.T., Lewis, J., et al., 2020. Tick-borne encephalitis virus, United Kingdom. *Emerg. Infect. Dis.* 26 (1), 90–96.
- Houldcroft, C.J., Beale, M.A., Breuer, J., 2017. Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* 15 (3), 183–192.
- Hue, S., Clewley, J.P., Cane, P.A., Pillay, D., 2004. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *AIDS* 18 (5), 719–728.
- Johnson, N., Phipps, L.P., Hansford, K.M., Folly, A.J., Fooks, A.R., Medlock, J.M., et al., 2022. One health approach to tick and tick-borne disease surveillance in the United Kingdom. *Int. J. Environ. Res. Public Health* 19 (10).
- Johnson, S., Enschke, N., Maayan, N., Mills, I., Buckley, B.S., Kakourou, A., et al., 2018. Ribavirin for treating Crimean Congo haemorrhagic fever. *Cochrane Database Syst. Rev.* 6, CD012713.
- Jore, S., Vanwambeke, S.O., Viljugrein, H., Isaksen, K., Kristoffersen, A.B., Woldehiwet, Z., et al., 2014. Climate and environmental change drives *Ixodes ricinus* geographical expansion at the northern range margin. *Parasites Vectors* 7.
- Kafetzopoulou, L.E., Efthymiadis, K., Lewandowski, K., Crook, A., Carter, D., Osborne, J., et al., 2018. Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. *Eur. Surveill.* 23 (50).
- Kafetzopoulou, L.E., Pullan, S.T., Lemey, P., Suchard, M.A., Ehichioya, D.U., Pahlmann, M., et al., 2019. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science* 363 (6422), 74–77.
- Khan, A.R., Pervez, M.T., Babar, M.E., Naveed, N., Shoaib, M., 2018. A comprehensive study of de novo genome assemblers: current challenges and future prospective. *Evol. Bioinform. Online* 14, 1176934318758650.
- Kolundzija, S., Cheng, D.Q., Lauro, F.M., 2022. RNA viruses in aquatic ecosystems through the lens of ecological genomics and transcriptomics. *Viruses* 14 (4).
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874.
- Lewandowski, K., Xu, Y., Pullan, S.T., Lumley, S.F., Foster, D., Sanderson, N., et al., 2019. Metagenomic nanopore sequencing of influenza virus direct from clinical respiratory samples. *J. Clin. Microbiol.* 58 (1).
- Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18), 3094–3100.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079.
- Lin, X., Glier, M., Kuchinski, K., Ross-Van Mierlo, T., McVea, D., Tyson, J.R., et al., 2021. Assessing multiplex tiling PCR sequencing approaches for detecting genomic variants of SARS-CoV-2 in municipal wastewater. *mSystems* 6 (5), e0106821.
- Liu, Y.X., Kearney, J., Mahmoud, M., Kille, B., Sedlazeck, F.J., Treangen, T.J., 2022. Rescuing low frequency variants within intra-host viral populations directly from Oxford Nanopore sequencing data. *Nat. Commun.* 13 (1).
- Lukashev, A.N., Klimentov, A.S., Smirnova, S.E., Dzagurova, T.K., Drexler, J.F., Gmyl, A.P., 2016. Phylogeography of Crimean Congo hemorrhagic fever virus. *PLoS One* 11 (11), e0166744.
- Madison-Antenucci, S., Kramer, L.D., Gebhardt, L.L., Kauffman, E., 2020. Emerging tick-borne diseases. *Clin. Microbiol. Rev.* 33 (2).
- Maloney, D.M., Fernandes, G., Dewar, R., O'Toole, A., Mphanga, C., Williams, T.C., et al., 2022. Preliminary Results from Two Novel “ARTIC-style” Amplicon Based Sequencing Approaches for RSV A and RSV B [virological.org](https://virological.org/t/preliminary-results-from-two-novel-artic-style-amplicon-based-sequencing-approaches-for-rsv-a-and-rsv-b/918#post-1). Available from: (<https://virological.org/t/preliminary-results-from-two-novel-artic-style-amplicon-based-sequencing-approaches-for-rsv-a-and-rsv-b/918#post-1>).
- Maurier, F., Beury, D., Flechon, L., Varre, J.S., Touzet, H., Goffard, A., et al., 2019. A complete protocol for whole-genome sequencing of virus from clinical samples: application to coronavirus OC43. *Virology* 531, 141–148.
- Mears, M.C., Bente, D.A., 2022. In silico design of a Crimean-Congo hemorrhagic fever virus glycoprotein multi-epitope antigen for vaccine development. *Zoonoses* 2 (1).
- Michelet, L., Delannoy, S., Devillers, E., Umhang, G., Aspan, A., Juremalm, M., et al., 2014. High-throughput screening of tick-borne pathogens in Europe. *Front. Cell. Infect. Microbiol.* 4, 103.
- Monsalve Arteaga, L., Munoz Bellido, J.L., Negro, A.I., Garcia Criado, J., Vieira Lista, M.C., Sanchez Serrano, J.A., et al., 2021. New circulation of genotype V of Crimean-Congo haemorrhagic fever virus in humans from Spain. *PLoS Negl. Trop. Dis.* 15 (2), e0009197.
- Negro, A., Sanchez-Arroyo, R., Diez-Fuertes, F., de Ory, F., Budino, M.A., Vazquez, A., et al., 2021. Fatal case of Crimean-Congo hemorrhagic fever caused by reassortant virus, Spain, 2018. *Emerg. Infect. Dis.* 27 (4), 1211–1215.
- Papa, A., Marklewitz, M., Paraskevopoulou, S., Garrison, A.R., Alkhovsky, S.V., Avsic-Zupanc, T., et al., 2022. History and classification of Aigai virus (formerly Crimean-Congo haemorrhagic fever virus genotype VI). *J. Gen. Virol.* 103 (4).
- Penedos, A.R., Myers, R., Hadeif, B., Aladin, F., Brown, K.E., 2015. Assessment of the utility of whole genome sequencing of measles virus in the characterisation of outbreaks. *PLoS One* 10 (11), e0143081.
- Pickin, M.J., Devignot, S., Weber, F., Groschup, M.H., 2022. Comparison of Crimean-Congo hemorrhagic fever virus and aigai virus in life cycle modeling systems reveals a difference in L protein activity. *J. Virol.* 96 (13), e0059922.
- Porretta, D., Mastrantonio, V., Amendolia, S., Gaiarsa, S., Epis, S., Genchi, C., et al., 2013. Effects of global changes on the climatic niche of the tick *Ixodes ricinus* inferred by species distribution modelling. *Parasites Vectors* 6, 271.
- Portillo, A., Palomar, A.M., Santibanez, P., Oteo, J.A., 2021. Epidemiological aspects of Crimean-Congo hemorrhagic fever in western Europe: what about the future? *Microorganisms* 9 (3).
- Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., et al., 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* 12 (6), 1261–1276.
- World Health Organization, 2013. Crimean-Congo Haemorrhagic Fever [updated 31 January 2013]. Available from: (<https://www.who.int/news-room/fact-sheets/detail/crimean-congo-haemorrhagic-fever>).
- Quick, J., 2020. nCov-2019 Sequencing Protocol (Single Sample). Available from: (<https://www.protocols.io/>).
- Reyes, G.R., Kim, J.P., 1991. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol. Cell. Probes* 5 (6), 473–481.
- Shahrear, S., Islam, A., 2021. Immunoinformatics guided modeling of CCHF\_GN728, an mRNA-based universal vaccine against Crimean-Congo hemorrhagic fever virus. *Comput. Biol. Med.* 140, 105098.
- Sims, D., Sudbery, I., Ilott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15 (2), 121–132.
- Tiph, T., Burt, F.J., 2020. Crimean-Congo hemorrhagic fever virus: advances in vaccine development. *Biores. Open Access* 9 (1), 137–150.
- Vlok, M., Lang, A.S., Suttle, C.A., 2019. Marine RNA virus quasispecies are distributed throughout the oceans. *mSphere* 4 (2).
- Weidmann, M., Sall, A.A., Manuguerra, J.C., Koivogui, L., Adjami, A., Traore, F.F., et al., 2011. Quantitative analysis of particles, genomes and infectious particles in supernatants of haemorrhagic fever virus cell cultures. *Viol. J.* 8, 81.
- Wu, H., Xing, N., Meng, K., Fu, B., Xue, W., Dong, P., et al., 2021. Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. *Cell Host Microbe* 29 (12), 1788–1801 e6.
- Zakotnik, S., Knap, N., Bogovic, P., Zorec, T.M., Poljak, M., Strle, F., et al., 2022. Complete genome sequencing of tick-borne encephalitis virus directly from clinical samples: comparison of shotgun metagenomic and targeted amplicon-based sequencing. *Viruses* 14 (6).