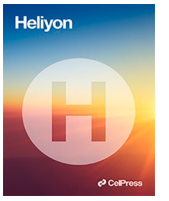
Heliyon 11 (2025) e42322

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Deep sequencing of serially passaged Sudan virus in guinea pigs uncovers adaptive mutations

Karla Emeterio [a,c], Jonathan Audet [a], Wenjun Zhu [a], Anders Leung [a], Helene Schulz [a], Shihua He [a], Michael Drebot [b,c], Logan Banadyga [a,c,*]

[a] Special Pathogens Program, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada
[b] National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada
[c] Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, MB, Canada

ARTICLE INFO

Keywords:
Filovirus
Ebolaviruses
Sudan virus
Adaptation
Next-generation sequencing
Mutation

ABSTRACT

Orthoebolaviruses are non-segmented, negative sense RNA viruses that make up the genus *Orthoebolavirus* in the family *Filoviridae*. Sudan virus (SUDV) is a highly pathogenic orthoebolavirus that causes severe disease in humans. In the last 45 years, SUDV has been responsible for several outbreaks in eastern Africa, particularly in Sudan and Uganda, with an average case fatality rate of approximately 50 %. Despite having caused numerous outbreaks, including a recent outbreak in 2022, no licensed therapeutics or prophylactics currently exist for Sudan virus disease. Small animal models like mice, hamsters, and guinea pigs have paved the way for the initial evaluation of filovirus countermeasures; however, since filoviruses are apathogenic in immunocompetent rodents, they must first be adapted through serial passaging. As a result of this process, viruses acquire genomic changes that may contribute to increased virulence and lethality. Currently, only a single immunocompetent small animal model exists for SUDV, where the virus was serially passaged in guinea pigs until uniform lethality was observed. To better understand the serial passaging process, we used next-generation sequencing to identify and quantify the mutations that arose throughout the adaptation process in guinea pigs. We identified 7 nonsynonymous and 9 synonymous mutations that were present at frequencies near 100 % at the end of serial passaging. The glycoprotein and virion protein (VP) 40 harboured many of the substitutions, most of which were nonsynonymous, while VP35 and VP24 each maintained a single nonsynonymous mutation. These results are consistent with the previously determined genome sequence, obtained by Sanger sequencing, with two notable exceptions: we identified a novel mutation in VP40, but we were unable to confirm the mutation in the genome leader, due to poor sequence coverage. This analysis has allowed us to identify adaptive hotspots within the viral genome, which may hint at the molecular determinants contributing to pathogenicity.

1. Introduction

Over the last four decades, orthoebolaviruses have posed a serious threat to global public health. Among the six orthoebolaviruses discovered to date, Ebola virus (EBOV), Sudan virus (SUDV), Bundibugyo virus (BDBV), and Taï Forest virus (TAFV) have caused

* Corresponding author. Special Pathogens Program, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada.
E-mail address: logan.banadyga@phac-aspc.gc.ca (L. Banadyga).

https://doi.org/10.1016/j.heliyon.2025.e42322
Received 4 October 2024; Received in revised form 26 November 2024; Accepted 27 January 2025
Available online 28 January 2025

outbreaks of severe disease across Africa. Collectively, these viruses cause Ebola disease (EBOD) [1], which initially manifests as a non-specific illness [2,3]. In the span of 7–14 days, infected individuals can begin to experience severe gastrointestinal symptoms, including vomiting and diarrhea, and in extreme cases may exhibit hemorrhagic manifestations and multiple organ failure that can lead to death [2,3]. At present, SUDV accounts for the second greatest number of outbreaks in the history of orthoebolaviruses, with an average case fatality rate of about 50 %. In September 2022, an outbreak of Sudan virus disease was declared in Uganda, which resulted in 142 confirmed cases and 55 deaths [4]. This recent re-emergence of SUDV highlighted the absence of approved therapeutics and vaccines available to treat the disease.

The orthoebolavirus genome encodes seven genes from which seven structural proteins are produced: nucleoprotein (NP), virion protein (VP) 35, VP40, glycoprotein (GP), VP30, VP24, and the RNA polymerase (L) [5]. GP mediates viral attachment to, and fusion with, target cells [5,6]. NP is a major component of the viral nucleocapsid that protects the RNA genome in addition to facilitating transcription and replication [7]. Along with NP, the nucleocapsid includes the RNA-dependent RNA polymerase L, required for the transcription and replication of the viral genome [5]; VP35, a polymerase cofactor and an interferon (IFN) antagonist [8,9]; VP30, a transcriptional activator [10,11]; and VP24, which also interferes with IFN signaling [12]. VP40 is the viral matrix protein, which is involved in virion assembly and egress [13,14]. Changes in any of these proteins can significantly affect the efficiency of viral replication and the ability of the virus to evade host immune responses.

Rodent models like mice, hamsters, and guinea pigs serve as indispensable tools in orthoebolavirus research, providing insights into virus pathogenesis and serving as a platform for evaluating potential countermeasures [15–18]. Despite the advantages associated with rodents, including ease of handling, cost-effectiveness, and fewer ethical concerns compared to larger animals like nonhuman primates, their use is limited by the inability of orthoebolaviruses to cause disease in immunocompetent animals [15–18]. To circumvent this limitation, a host-adapted virus must be generated through serial passaging [15,18]. The process of host-adaptation consequently leads to the emergence of mutations in the viral genome that contribute to the observed virulence in the rodent host [15]. Several different host-adapted orthoebolaviruses have been generated and all have yielded uniformly lethal animal models that mimic many, but not all, features of EBOD [19–27].

In 2015, the first and only immunocompetent rodent animal model for SUDV was developed by passaging wild type SUDV variant Boneface in guinea pigs until uniform lethality was achieved [27]. To generate guinea pig-adapted SUDV (GPA-SUDV), guinea pigs were inoculated with wild type SUDV and, seven days later, liver and spleen samples were collected, pooled, homogenized, and then used to inoculate another group of guinea pigs. Successive rounds of passaging were repeated until inoculation resulted in all animals succumbing to disease, which occurred by passage 20. The lethal GPA-SUDV was then Sanger sequenced, revealing the presence of 16 nucleotide changes.

In this study, we used deep sequencing to analyze the SUDV genome in all tissue homogenates collected during the passaging process. In doing so, we aimed to identify the changes in the viral genome sequence over time and to quantify the frequencies of each mutation, thus providing insights into the trajectory of virus adaptation. Over the course of passaging, we observed dynamic changes in the viral genome, characterized by the appearance and disappearance of mutations accompanied by fluctuations in mutation frequencies. We confirmed 15 of the 16 previously identified mutations [27]; we were unable to detect the mutation in the genome leader sequence because of a lack of sequencing coverage. Interestingly, however, our work identified a previously undescribed point mutation in the *VP40* gene, implying that GPA-SUDV actually acquired a total set of 17 mutations by passage 20, the point at which uniform lethality in guinea pigs was achieved.

2. Materials and methods

2.1. Animal ethics

All 20 tissue homogenates used in this study were derived from a previously described study [27]. The animal experiment protocol (H-11-010) was reviewed and approved by the Animal Care Committee at the Canadian Science Centre for Human and Animal Health, Winnipeg, Manitoba, in accordance with guidelines from the Canadian Council on Animal Care.

2.2. Viral RNA purification

Virus RNA was purified from stock virus preparations of wild type SUDV (Boneface) and guinea pig-adapted SUDV or twenty liver/spleen tissue homogenates using AVL (QIAGEN) and ethanol in conjunction with the QIAmp Viral RNA Mini Kit (QIAGEN) or the Applied Biosystems MagMAX Viral/Pathogen II (MVP II) Nucleic Acid Isolation Kit (Thermo Fisher Scientific) on the KingFisher Duo Prime instrument, per manufacturer's instructions. Poly (A) carrier RNA or an inert linear acrylamide carrier was added to buffer AVL as precipitant to aid in maximum recovery of RNA from the tissue homogenates and virus stock preparation, respectively.

2.3. Sequencing of SUDV genome derived from tissue homogenates

Reverse transcription PCR (RT-PCR) of total RNA was carried out using SuperScript™ III First-Strand Synthesis SuperMix Kit (Invitrogen), following manufacturer's instructions, using 2 μM genome-specific primers (Supplementary Table 1). Samples were incubated at 65 °C for 5 min prior to the addition of First-Strand Synthesis Mix and SuperScript® III/RNaseOUT™ Enzyme mix (Invitrogen). The final reaction was then incubated at 50 °C for 50 min and terminated at 85 °C for 5 min.

cDNA was subsequently used to generate amplicons that spanned the entirety of the SUDV genome using 10 μM primers

(Supplementary Tables 2 and 3). In instances where the region of interest was not amplifiable, partially amplifiable, or had low sequencing read coverage, alternate primers designed using Primal Scheme [28] were used (Supplementary Table 4). PCR amplification was performed with CloneAmp HiFi Polymerase (Takara Bio) under the following thermal cycling conditions: initial denaturation at 95 °C for 3 min, followed by 45 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 2 min, and a final extension at 72 °C for 10 min. Extension time at 72 °C was reduced to 1 min for amplicons of <1 kilobase in size. PCR reactions were run on a 1 % agarose gel, and DNA fragments of the correct size were excised and purified using the QIAquick Gel Extraction kit (QIAGEN) according to manufacturer's protocol.

DNA products were quantitated using the Quant-iT PicoGreen dsDNA assay kit (Invitrogen), per manufacturer's instructions, on the FilterMax F5 microplate reader (Molecular Devices). Amplicons amplified from the same serial passage were pooled in equimolar amounts based on the size and concentration of each amplicon. Libraries were constructed using the Nextera XT Sample Prep kit (Illumina) per manufacturer's instructions and sequenced on the MiSeq sequencer using the MiSeq Reagent Kit v3 (600-cycle) (Illumina).

2.4. Sequencing of the stock Guinea pig-adapted SUDV genome

RNA purified from a stock of GPA-SUDV grown on Vero E6 cells underwent reverse transcription using a modified SuperScript III First-Strand Synthesis System (Invitrogen). The RNA was first DNase-treated with ezDNase (Invitrogen) at 37 °C for 2 min to remove any contaminating genomic DNA. The treated RNA was subsequently heated at 65 °C for 5 min with 6 μM of random primer mix (NEB) and 10 mM of dNTPs, followed by annealing through cooling on ice for 1 min. RT-PCR was then performed by adding 10× First-strand buffer, 0.1 M DTT, 40 units of RNase OUT, and 200 units of SuperScript III reverse transcriptase. The 20 μl reaction was incubated at 25 °C for 10 min, 55 °C for 120 min, and 85 °C for 5 min, producing a single-stranded cDNA (ss cDNA). Subsequently, a second-strand synthesis reaction mixture containing 10 mM dNTP mix, 10× reaction buffer, 10 units of DNA polymerase I (NEB) and DNA ligase (NEB), and 2 units of RNaseH (NEB) was added to the ss cDNA template and incubated at 16 °C overnight. The double-stranded cDNA (ds cDNA) was purified using AMPure XP DNA purification beads (Beckman Coulter) at a ratio of 1:1.8 and used for library preparation.

A sequencing library was prepared by initially fragmenting the ds cDNA template using NEBNext dsDNA fragmentase (NEB), per manufacturer's instructions, followed by a DNA bead clean up step. Subsequently, the TruSeq DNA Nano Kit (Illumina) was used, per manufacturer's guidelines, for repairing the ends of the ds cDNA fragments, selecting for a 550 bp insert size, ligating unique dual-index adapters, and enriching for DNA fragments. To specifically enrich for fragments of SUDV, customized biotinylated oligo probes spanning the entirety of the viral genome were designed and obtained from Integrated DNA Technologies (IDT). SUDV-probe enrichment was conducted using the xGen Hybridization and Wash Kit (IDT), per manufacturer's instructions. Post-capture PCR amplification using the Enhanced PCR Mix and PCR Primer Cocktail from the TruSeq Nano DNA Kit (Illumina) was subsequently performed under the following thermocycling conditions: initial denaturation at 98 °C for 45 s followed by 13 cycles of 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 30 s, and a final extension at 72 °C for 1 min.

Prior to sequencing, the library was quantified and sized using the Qubit 1× dsDNA High Sensitivity Assay Kit (Invitrogen) and Agilent 4200 TapeStation (Agilent Technologies), respectively. Paired-end sequencing of the dual-indexed reads was performed with MiSeq Reagent Kit v3 (600-cycle) (Illumina) on a MiSeq instrument.

2.5. Single nucleotide polymorphism detection

Quality control checks, including per-base sequence quality, per-sequence quality scores, and adapter content, were performed using FastQC [29] for the sequences obtained from the passaged samples. Post-sequencing analysis was first completed on Geneious Prime 2023.0.1 (http://www.geneious.com/). Raw sequencing reads were paired and trimmed using the BBDuk tool, and then aligned and assembled to a reference sequence (GenBank accession no. KT878488.1) using Geneious as the mapper with an iteration of five times. For the stock GPA-SUDV, processing sequencing reads was performed on Unipro UGENE (v47.0) [30]. Briefly, the built-in working pipeline was as follows: raw sequencing reads were initially filtered and adapter sequences were removed. Both ends of the reads were subsequently trimmed and then mapped to the same reference sequence using the BWA-MEM mapping tool. The output data were converted to BAM files. The BAM files were then merged, sorted, and PCR duplicates removed using SAMTools. Initial variant calling was performed in Geneious such that variants present at a frequency of ≥0.01 and a minimum sequencing depth of 1 read count were identified. A variant call format (VCF) file representing the genetic variations identified from each of the sequenced samples was exported and further analysed. The VCF and depth files were read in R 4.3.3 using the tidyverse package v2.0.0 [31] and the patchwork package v1.2.0 [32]. Codon changes were converted to amino acid change using the bioseq package v0.1.4 [33]. Mutations that met the reporting criteria (≥5 % frequency and depth ≥1000 in at least 2 passages or in the final stock) were kept. If the depth for a specific location for a passage was less than 1000, the information was considered missing (graphed in light grey in the heatmap, Fig. 2). If a mutation met the reporting criteria, passages where the frequency was less than 5 % were set to 0 (transparent values in the heatmap, Fig. 2). The code is available on GitHub (https://github.com/jaudetnml/SUDV_GPA_NGS_figures).

3. Results

3.1. Successful amplicon-based sequencing approach of the SUDV genome

To better characterize the SUDV genome at each step throughout the GPA-SUDV passaging process [27], we subjected the tissue homogenate samples collected after each guinea pig passage to a comprehensive amplicon-based sequencing protocol. RNA from each tissue sample was extracted and reverse transcribed to produce cDNA, which was then PCR amplified in 16 fragments spanning the entire SUDV genome. The resulting amplicons underwent deep sequencing. While this method was successful for most samples, we failed to amplify certain genome fragments from passages 8, 9, 13, and 18, and we were unable to amplify the first 205 nucleotides of the genome for all passages (Fig. 1). As a result, single nucleotide polymorphisms (SNPs) existing in these regions, including the t34c mutation previously detected in the leader region [27], were not identified. The SNPs identified in this study were those appearing at a frequency of ≥5 % in at least 2 passages or in the final GPA-SUDV stock virus, and we established a minimum depth threshold of ≥1000 reads. Overall, our sequencing depth had a median of 42,027 reads, with an interquartile range of 54,159 reads and a range of 0–539, 250 reads (Supplementary Fig. 1). Across all passages, 70 %–99 % of the whole genome was covered with a sequencing depth of ≥1000 reads.

Genomic mutations in nearly all the viral proteins were apparent early into the process of adaptation. Sequencing of the viral genome across 20 passages revealed 118 distinct mutations throughout the series of passages (Fig. 2). However, approximately 86 % of these mutations were temporary occurrences: some were only detected once, while others were present over multiple passages but then fell below the 5 % frequency cutoff. At passage 20, we identified 16 mutations in GPA-SUDV, each occurring at frequencies close

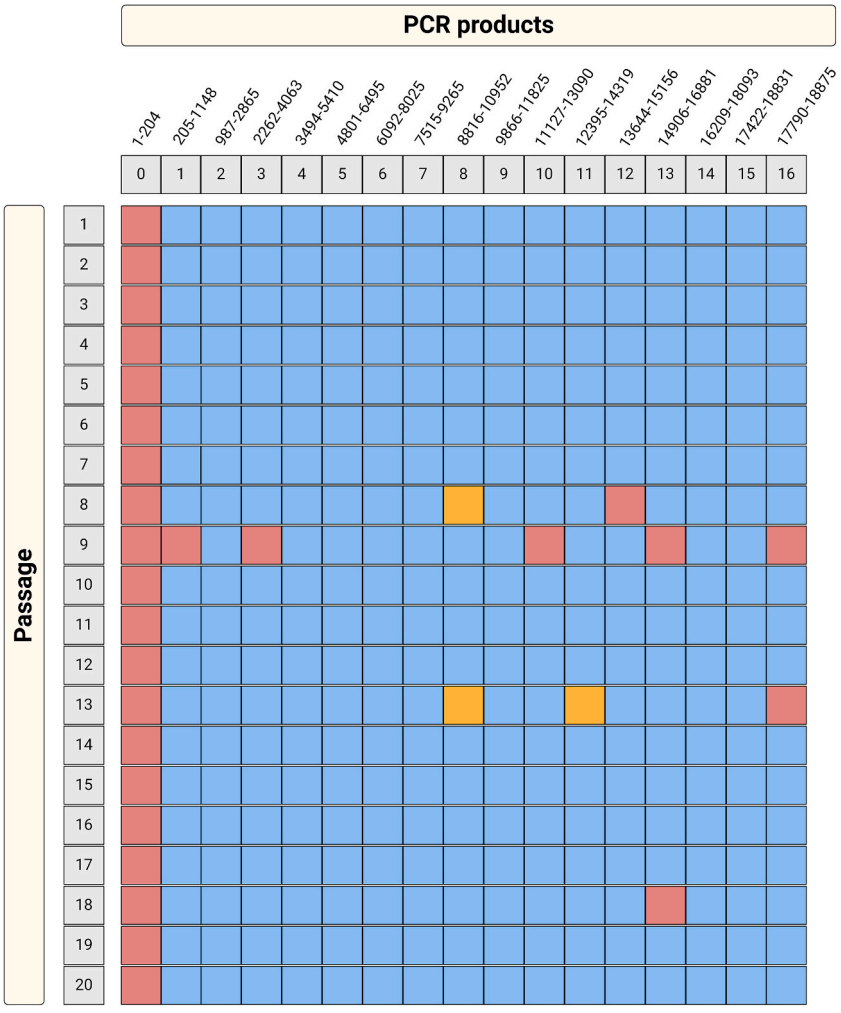


Fig. 1. Summary of amplicons recovered from guinea pig tissue homogenates. For each passage (rows 1–20), 16 overlapping genome fragments (columns 1–16) were amplified by conventional PCR. Column 0 represents the genome fragment from nucleotides 1–204, which could not be amplified. Blue squares represent fragments that were completely amplified, red squares represent fragments that could not be amplified, and orange squares represent partially amplified fragments. Figure created with BioRender.com.

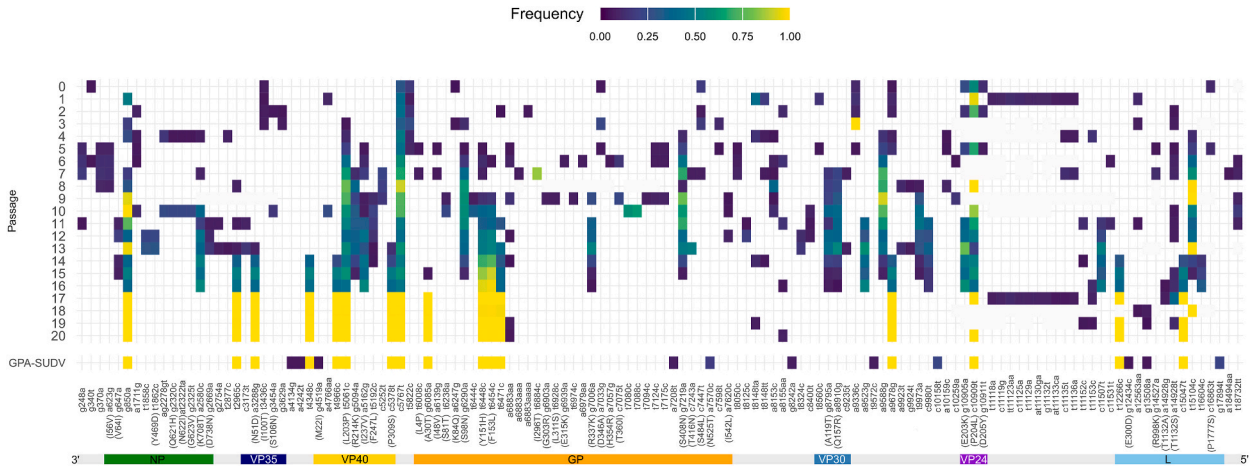


Fig. 2. The distribution of mutations across the SUDV genome during passaging in guinea pigs. Each column (x-axis) represents a single nucleotide polymorphism identified within the viral genome, and each row represents the passage at which the mutations were identified. The unpassaged, wild type virus is identified as passage 0, while the GPA-SUDV stock virus sequence is identified as "GPA-SUDV". Only mutations that were present at frequencies ≥5 % with a read count ≥1000 and that occurred in at least two passages were reported. The gradient of colours from dark blue to yellow indicates the frequency of each mutation at every passage. At passages where the frequency of a mutation was less than 5 %, the value was set to 0 and no colour is used on the heatmap. Light grey boxes indicate the absence of sufficient sequence data at these positions.

to 100 % (Fig. 2). Twelve mutations were identified within the coding sequences (CDS) of the genes (Figs. 3 and 4), and four mutations were detected in noncoding regions, particularly in the untranslated regions of the genes (Supplementary Fig. 2). Notably, the stock preparation of GPA-SUDV—generated following propagation of passage 20 on Vero E6 cells—also exhibited these mutations at frequencies nearing 100 % (Fig. 2). We also identified 11 additional mutations exclusive to the stock GPA-SUDV, although they occurred at frequencies lower than 25 % (Fig. 2). The actual frequencies of each mutation within the whole genome at each passage are detailed in Supplementary Tables 5–7.

Overall, our sequencing efforts confirmed 15 out of the 16 mutations that had been previously identified via Sanger sequencing [27] (Fig. 5 and Table 1). The lack of an amplicon covering the first 205 nucleotides of the viral genome limited our ability to detect the presence of the t34c mutation in the leader region [27]. Interestingly, our study identified a unique mutation in viral protein (VP) 40, L203P, which had not been previously identified [27]. Hence, we propose that GPA-SUDV possesses a total of 17 adaptive mutations.

3.2. *The SUDV genome acquired several nonsynonymous mutations during passaging*

Deep sequencing of the SUDV genome revealed that the virus acquired numerous nonsynonymous mutations across its genome during the course of passaging in guinea pigs (Fig. 3). The *GP* gene exhibited the greatest number of nonsynonymous mutations existing near 100 % frequencies by passage 20, followed by viral proteins (*VP*) 40, 35, and 24. In contrast, the *L* polymerase, *NP*, and *VP30* genes did not retain any mutations by passage 20.

In the *GP* gene, most of the nonsynonymous mutations occurred in the CDS region (Fig. 3D). Out of the 53 mutations identified throughout the course of passaging, three mutations became fixed in the genome of GPA-SUDV. These three mutations, namely g6085a (A30T), t6448c (Y151H), and t6454c (F153L), all occurred within the receptor binding domain (RBD) of GP and reached a frequency of 99 % in the virus population by the 18th passage. Mutations Y151H and F153L were first detected at passages 5 and 6, respectively, and their frequency gradually increased with each passage. The A30T mutation was not detected until passage 15 at 6.4 % frequency; however, its frequency increased dramatically over the subsequent two passages, increasing to 99.7 % by passage 17.

In the *VP40* gene, two mutations, t5061c (L203P) and c5378t (P309S), existed in the CDS at frequencies near 100 % by passage 20 (Fig. 3C). These mutations are situated at the C-Terminal domain of the VP40 protein. The L203P mutation was the first nonsynonymous mutation to appear in the gene on the 3rd passage, at 5.5 % frequency, and it increased in frequency over each successive passage to a frequency of 99.8 % at passage 17. Notably, L203P was not previously detected and reported during the development of GPA-SUDV [27]. Mutation P309S, on the other hand, was first observed at the 13th passage, with a frequency of 13.3 %, but later reached a frequency 99.6 % at passage 17.

The *VP35* and *VP24* genes each had a single nonsynonymous mutation that became fixed at the end of passaging (Fig. 3B and F). Both genes had fewer transient mutations, with some appearing only once and others recurring over multiple passages. In the *VP35* gene, the a3288g (N51D) mutation was first detected in the 13th passage at about 5 % frequency and gradually increased in frequency up to 50 % at passage 15, after which it jumped to 99 % frequency at passage 17 (Fig. 3B). Interestingly, in *VP24*, the c10909t (P204L) mutation was already present in the wild type virus at a frequency of approximately 55 %. The frequency of this mutation varied throughout passaging until it reached 99 % at passage 17 and 20 (Fig. 3F). Despite P204L pre-existing in the wild type virus population at 55 %, its increase to 99 % at passage 20 suggests potential adaptation driven by selective pressures. Notably, we were unable to

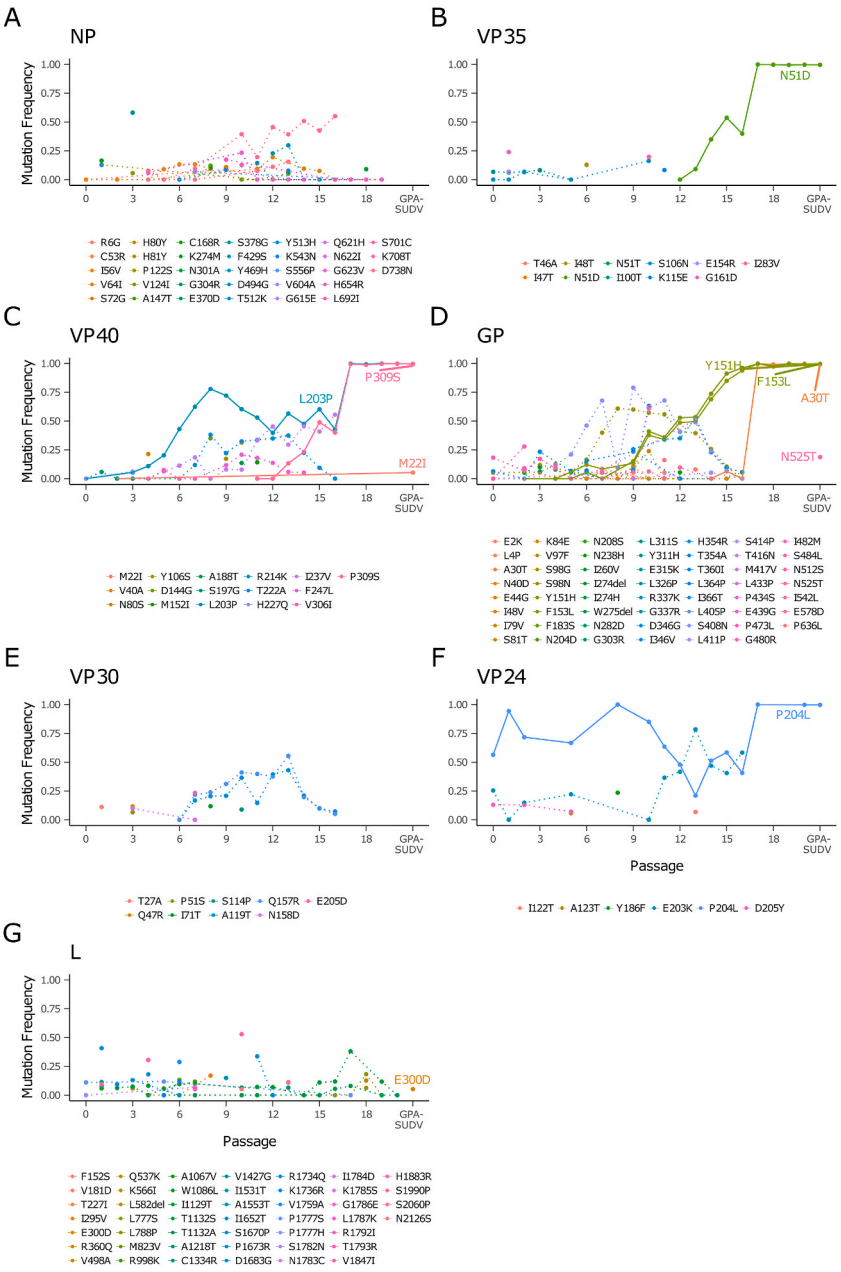


Fig. 3. Nonsynonymous mutations identified in the coding sequences of the GPA-SUDV genome throughout passaging. The frequency of nonsynonymous mutations identified in the coding sequences of (A) *NP*, (B) *VP35*, (C) *VP40*, (D) *GP*, (E) *VP30*, (F) *VP24*, and (G) *L* genes were plotted over 20 passages. Mutations that were maintained until passage 20 are highlighted as solid lines, while transient mutations are represented as dashed lines. Passage 0 refers to the unpassaged wild type SUDV, while the stock of GPA-SUDV is noted as GPA-SUDV.

definitively call this mutation at passages 3, 4, 6, 7, 9, 18, and 19 due to insufficient sequencing data (Fig. 3F).

In the CDS region of the *L* gene, we detected 45 nonsynonymous mutations, although all were transient (Fig. 3G). Similarly, the *NP* and *VP30* genes exhibited only transient nonsynonymous mutations throughout passaging in guinea pigs (Fig. 3A and E). The *NP* gene had 33 distinct nonsynonymous mutations detected in its CDS region (Fig. 3A), while *VP30* garnered 9 nonsynonymous mutations across the 20 passages (Fig. 3E). In all three genes—*L*, *NP* and *VP30*—these short-lived mutations were observed at frequencies below 60 % and were undetectable by passage 20.

3.3. The SUDV genome acquired several synonymous and noncoding mutations during passaging

GPA-SUDV also acquired several synonymous or silent mutations within the CDS of each gene, except in *VP24* (Fig. 4). The L gene

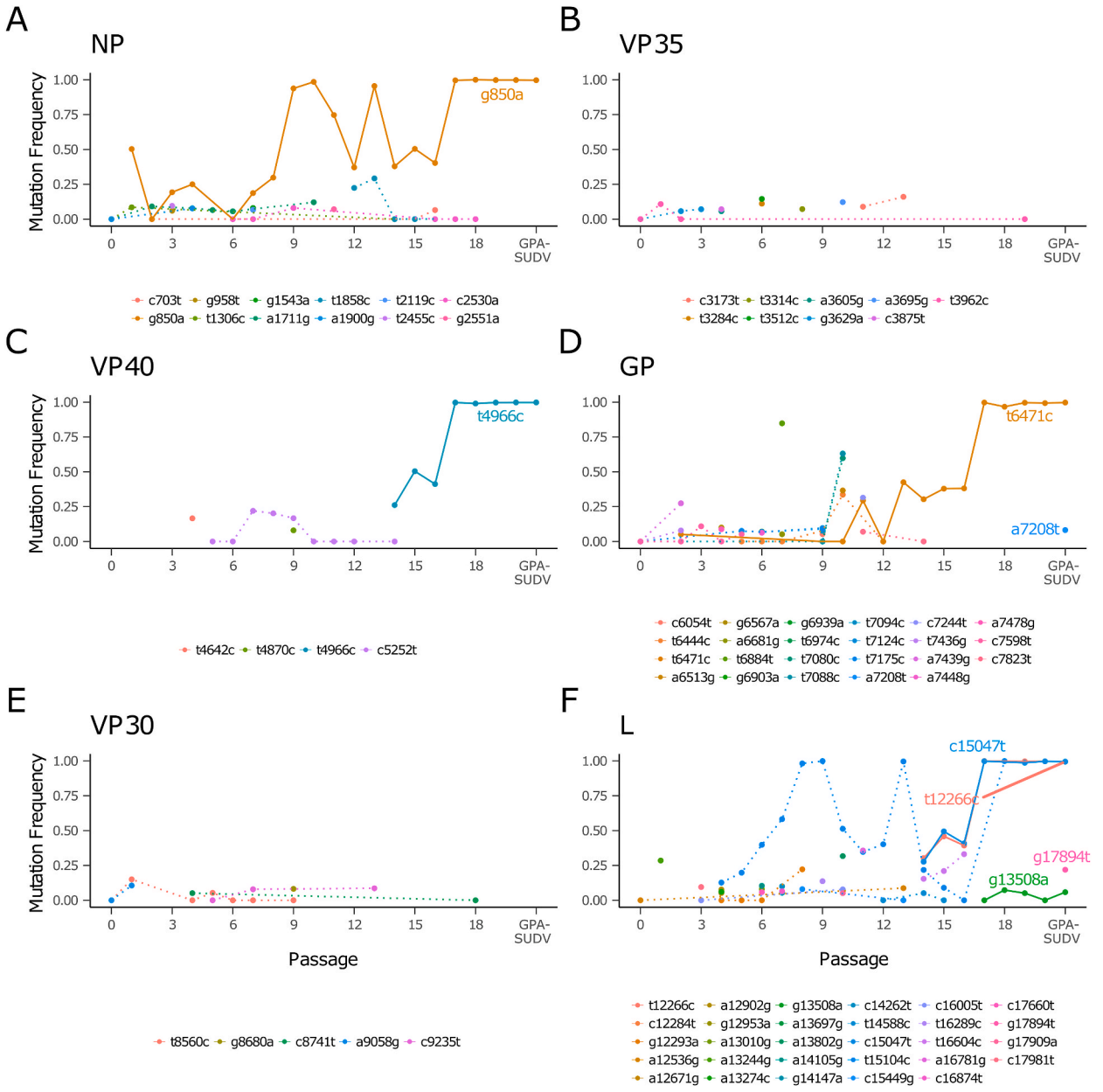


Fig. 4. Synonymous mutations in the coding sequences of the GPA-SUDV genome throughout passaging. The frequency of synonymous mutations identified in the coding sequences of (A) *NP*, (B) *VP35*, (C) *VP40*, (D) *GP*, (E) *VP30*, and (F) *L* genes were plotted over 20 passages. Mutations that were maintained until passage 20 are highlighted as solid lines, while transient mutations are represented as dashed lines. Passage 0 refers to the unpassaged wild type SUDV, while the stock of GPA-SUDV is noted as GPA-SUDV.

had 29 silent mutations, but only two, t12266c and c15047t, became fixed by the end of passaging (Fig. 4F). Both mutations were first observed at a frequency of about 30 % at passage 14 and increased to nearly 100 % by the 17th passage. The CDS of the *NP, VP40,* and *GP* genes each retained a single synonymous mutation. In *NP*, mutation g850a was first detected at passage 1 at a frequency of 50.3 % and was the only mutation retained in the *NP* gene at passage 20 (Fig. 4A). Its frequency fluctuated significantly, initially dropping below 50 % between passages 2–8, then sharply rising to over 90 % at passages 9 to 10. Subsequently, it declined to 70 % and 50 % at passages 11 and 12, respectively. At passage 13, it reached 95.5 % before falling to frequencies below 50 % in the following passages. It was not until passage 17 that the mutation reached a frequency of 99 %, where it remained until passage 20. In the *VP40* CDS, mutation t4966c was first identified at passage 14 at 26 % frequency, reaching 99 % at passage 17 (Fig. 4C). Twenty-two silent mutations appeared in the *GP* gene over the course of 20 passages, but only the t6471c mutation became established (Fig. 4D). This mutation was initially detected in passage 2 at a frequency of 5.1 %, with frequencies fluctuating over passages 10–16. It was not until passage 17

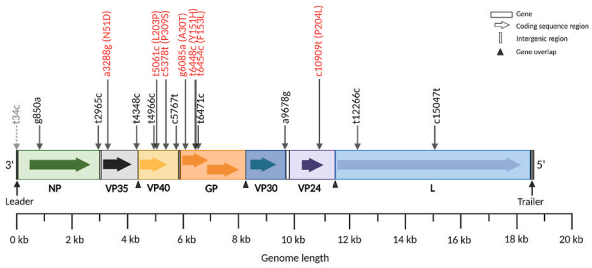


Fig. 5. Mutations uncovered in the GPA-SUDV at passage 20. At passage 20, a total of 16 mutations were identified in the lethal guinea pig-adapted Sudan Virus (GPA-SUDV). These mutations existed at frequencies near 100 % by passage 20 with read counts >1000. Twelve of the 16 nucleotide changes were observed in the coding sequence regions of the genome (depicted as arrows within the each rectangular gene), while four of the 16 were in the 3′ or 5′ untranslated regions of the proteins (portions of the rectangle outside the arrows). No mutations were detected in the intergenic regions and in regions where genes may overlap. Nonsynonymous substitutions are highlighted in red with their corresponding amino acid change, while synonymous substitutions are depicted in black. A previously identified mutation (t34c) in the leader sequence of the genome was undetected in our analyses and is shown in grey with a dashed arrow. Figure created with Biorender.com.

Table 1
Comparison of genomic mutations identified in the guinea pig-adapted Sudan virus genome (GPA-SUDV) at passage 20 and a stock preparation of GPA-SUDV.

Gene/Location	Nucleotide Position	Nucleotide Change	Amino Acid Change	GPA-SUDV[a]	Deep-sequenced P20[b]	Sanger-sequenced P20[c]
3′ Leader	34	t → c	…	ND	ND	+
NP	850	g → a	Silent	+	+	+
NP	2965	t → c	…	+	+	+
VP35	3288	a → g	N51D	+	+	+
VP35	4348	t → c	…	+	+	+
VP40	4966	t → c	Silent	+	+	+
VP40	5061	t → c	L203P	+	+	ND
VP40	5378	c → t	P309S	+	+	+
VP40	5767	c → t	…	+	+	+
GP	6085	g → a	A30T	+	+	+
GP	6448	t → c	Y151H	+	+	+
GP	6454	t → c	F153L	+	+	+
GP	6471	t → c	Silent	+	+	+
VP30	9678	a → g	…	+	+	+
VP24	10909	c → t	P204L	+	+	+
L	12266	t → c	Silent	+	+	+
L	15047	c → t	Silent	+	+	+

…, mutation occurs in a noncoding region; ND, mutation not detected; +, mutation detected.
[a] Genome obtained from a stock preparation of GPA-SUDV and deep sequenced; data from this study.
[b] Genome obtained from passage 20 of GPA-SUDV and deep sequenced; data from this study.
[c] Genome obtained from passage 20 of GPA-SUDV and Sanger sequenced; data from Ref. [27].

that the mutation stably reached 99 %. The *VP35* and *VP30* genes also experienced synonymous mutations in their coding regions but such changes were minimal, low in frequencies, and transitory (Fig. 4B and E).

Mutations were also identified in the noncoding regions of the SUDV genome, including the 3′ leader and 5′ trailer sequences, as well as the untranslated regions (UTR) that flank the CDS of the genes (Supplementary Fig. 2). The *NP*, *VP35*, *VP40*, and *VP30* genes each retained a single mutation in their downstream UTRs at frequencies close to 100 % by either passage 17 or 18 (Supplementary Fig. 2B-D and 2F). Interestingly, mutation c5767t in *VP40* was originally found in the wild type unpassaged SUDV at a frequency near 30 % and appeared to have been continuously selected for during the adaptation process (Supplementary Fig. 2D). It reached a frequency of 91 % early on at passage 8, then fell to 47 % after eight subsequent passages before reaching 99 % at passage 17. We identified 25 SNPs in the noncoding regions between the *VP30* stop codon and the *VP24* start codon over the course of 20 passages, but only a9678g in the *VP30* gene existed in the last passage (Supplementary Fig. 2F). Like the c5767t mutation, a9678g was present in the wild type SUDV sequence at a frequency of 6 % and remained detectable until the 5th passage. It was later identified again at passages 14–20, but it was not until passage 17 that this mutation reached a frequency of 99 %. Despite the abundance of mutations existing in between the *GP* stop codon and the *VP30* start codon, as well as the *VP24* stop codon and the *L* start codon, none of these mutations were retained in GPA-SUDV at passage 20 (Supplementary Figs. 2E and 2G). Mutations within these regions were fleeting and only existed at frequencies below 60 %. Similarly, we found no mutations in the 3′ leader (Supplementary Fig. 2A) and 5’ trailer (Supplementary Fig. 2H) sequences by the last passage.

4. Discussion

During the passaging of SUDV through guinea pigs, we observed numerous mutations across the genome, including changes in the genes for every protein, albeit to different extents. While the majority of mutations were synonymous—and therefore difficult to interpret—seven mutations in four proteins provide intriguing insight into the nature of SUDV pathogenesis.

Most of the nonsynonymous mutations observed were in GP. Accumulation of mutations in GP throughout the course of passaging in guinea pigs is likely driven by a complex interplay of various factors, such as host-virus interactions, immune evasion strategies, and selective pressures exerted by the host environment. Regardless of the mechanism of selection, the fact that GP acquired the most mutations throughout the passaging process underscores the importance of GP in viral adaptation and, potentially, pathogenesis. The nonsynonymous GP mutations, A30T, Y151H, and F153L, were all located in the RBD of the protein, a region critical for recognition of the cellular endosomal receptor, Neimann-Pick C1 (NPC1) [34,35]. The interaction between the RBD and NPC1 is a key step in the virus's entry into the host cell, and mutations in the RBD region may influence the binding affinity of GP to guinea pig NPC1.

In the *VP40* gene, GPA-SUDV acquired two nonsynonymous mutations—L203P and P309S—both of which are in the C-terminal domain (CTD) of the protein. In EBOV VP40, the CTD is known to associate with the lipid bilayer of the host plasma membrane, inducing proper matrix assembly and egress [13,36,37], and the presence of a hydrophobic CTD-CTD interface has been deemed necessary for the budding of virus-like particles [38]. Structural alignments of VP40 from EBOV and SUDV reveal a conserved CTD-CTD hydrophobic structure between the two proteins, with SUDV L203, I237, M241, M305, and I307 playing an important role in establishing this interface [38,39]. Alterations in one or more CTD residues may affect the overall structure of this hydrophobic interface, which in turn may alter the ability of VP40 to drive viral budding and/or interact with other proteins.

The *VP35* and *VP24* genes both acquired one nonsynonymous substitution. It is possible that the N51D mutation in VP35 influences nucleocapsid formation. As observed for EBOV, N51 is located in a linker region close to the NP-chaperoning domain of the protein [40]. Linker regions have a function in spacing out various domains of a protein and can have a rigid or flexible structure [41], so a change in this space may affect the flexibility or rigidity of the neighboring domains. The P204L mutation in VP24 is speculated to have the potential to affect how the protein functions in evading the host immune response. It has been demonstrated for EBOV that amino acid P204 binds the nuclear transporter karyopherin alpha 1 (KPNA1), which is an essential protein that assists in the translocation of signal tranducer and activator of transcription 1 (STAT1) to the nucleus to activate interferon-stimulated genes (ISGs) [12]. Hence, we hypothesize that the replacement of the residue might alter the intermolecular interactions between VP24 and KPNA1, which in turn may change how well the transcription of host ISGs are antagonized.

We noticed interesting temporal patterns on a few of the nonsynonymous mutations we detected during adaptation. For instance, a similar temporal pattern was observed between the N51D mutation in VP35 and the P309S mutation in VP40, suggesting that these mutations are functionally related and likely located on the same genome. The co-occurrence of these mutations may indicate complementary or compensatory effects on a common pathway in the virus replication cycle, such as nucleocapsid formation and viral egress. On the other hand, we saw that the E203K mutation in VP24 appeared to be mutually exclusive with P204L: as the frequency of P204L increased, the frequency of E203K decreased, and vice versa, until P204L reached a frequency of 99 % at passage 17 and E203K disappeared. Opposite temporal patterns between mutations have also been observed in a similar study by Wei et al. [42]. The particular point in time in which a guinea pig-adapted mutation appears, in conjunction with the varying frequency of that mutation over time, underscores the complex interplay between these mutations and the host environment. Unfortunately, while conducting this study, we did not have access to the animal health records for each passage, which prevents us from linking the appearance of a particular mutation to an observable change in virus pathogenesis. Future work will focus on understanding the functions of each mutation in an effort to define the key genetic determinants that facilitate increased virulence.

Unlike SUDV, where only a single GPA-SUDV variant currently exists, multiple strains of GPA-EBOV exist for EBOV [23,43–46]. Comparison of mutations observed across the different strains of GPA-EBOV and GPA-SUDV, reveal some commonalities and distinctions. One striking observation is the presence of mutations in the *VP24* gene. While multiple mutations in VP24 were acquired in the various isolates of GPA-EBOV [32,33], GPA-SUDV only retained a single VP24 mutation. It is possible that the variation in the number of VP24 mutations present is related to the differences in host factors that restrict EBOV and SUDV; however, it nonetheless, seems likely that alterations in VP24 serve a fitness advantage for both viruses when they are subject to a new host. Conversely, GPA-EBOV obtained several nonsynonymous mutations in NP, while GPA-SUDV only retained a single synonymous mutation [23,45,46], suggesting that changes to EBOV NP may be important for virulence in guinea pigs, while changes in SUDV NP may not. Interestingly, the presence of nonsynonymous mutations in VP40 is unique to GPA-SUDV, suggesting that the SUDV matrix protein may play a key role in virus virulence. Overall, these differences imply that distinct adaptive molecular mechanisms govern the adaptation of these two related viruses in guinea pigs. More work with SUDV is required to better define the specific effects that the GPA-SUDV mutations are having in guinea pig pathogenesis. Such research is not only crucial for gaining a comprehensive understanding of SUDV pathogenesis, but it may also be critical for identifying novel targets for therapeutic intervention.

In summary, despite not detecting the t34c mutation in the *NP* gene, we confirmed 15 of the 16 mutations identified previously by Sanger sequencing [27]. Additionally, we describe for the first time a novel VP40 mutation, L203P, suggesting that there are likely 17 mutations in GPA-SUDV. The *GP*, *VP40*, and *VP24* genes were found to be important sites of adaptation, with mutations in these regions appearing early on and persisting over time. Further investigation is warranted to fully understand the roles of these mutations in the virus replication cycle and immune evasion. It is also necessary to determine which mutations, or combinations thereof, are responsible for the observed lethality in guinea pigs. Pinpointing exact mutations that lead to pathogenicity could prove to be difficult due to the complex nature of the adaptation process, which involves multiple factors like environmental conditions and host-factor interactions; however, understanding the molecular pathways that lead to a pathogenic and lethal orthoebolavirus also improves

our understanding of the viral and host processes that contribute to filovirus disease in humans.

CRediT authorship contribution statement

Karla Emeterio: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jonathan Audet:** Writing – review & editing, Visualization, Methodology, Formal analysis. **Wenjun Zhu:** Writing – review & editing, Methodology, Investigation. **Anders Leung:** Writing – review & editing, Methodology, Investigation. **Helene Schulz:** Writing – review & editing, Methodology, Investigation. **Shihua He:** Writing – review & editing, Investigation. **Michael Drebot:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Logan Banadyga:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Data availability

All raw sequence reads (FASTQ files) are available on the Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) under BioProject accession number PRJNA1163425.

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

This research was funded by the Public Health Agency of Canada (PHAC). The next-generation sequencing work was supported by the Genomics Core Facility at the National Microbiology Laboratory.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2025.e42322.

References

[1] J.H. Kuhn, et al., New filovirus disease classification and nomenclature, Nat. Rev. Microbiol. 17 (5) (2019) 261–263.
[2] D. Malvy, et al., Ebola virus disease, Lancet 393 (10174) (2019) 936–948.
[3] S.T. Jacob, et al., Ebola virus disease, Nat. Rev. Dis. Prim. 6 (1) (2020) 1–31, 2020. 6(1).
[4] Uganda declares end of Ebola disease outbreak | WHO | Regional Office for Africa. https://www.afro.who.int/countries/uganda/news/uganda-declares-end-ebola-disease-outbreak.
[5] D. Cantoni, J.S. Rossman, Ebolaviruses: new roles for old proteins, PLoS Neglected Trop. Dis. 12 (5) (2018).
[6] J.E. Lee, E.O. Saphire, Ebolavirus glycoprotein structure and mechanism of entry, Future Virol. 4 (6) (2009) 621–635.
[7] L. Baseler, et al., The pathogenesis of Ebola virus disease * EBOV: Ebola virus, Annu. Rev. Pathol. 12 (2017, 2016) 387–418.
[8] C.F. Basler, et al., The Ebola virus VP35 protein inhibits activation of interferon regulatory factor 3, J. Virol. 77 (14) (2003) 7945–7956.
[9] C.F. Basler, et al., The Ebola virus VP35 protein functions as a type I IFN antagonist, Proc. Natl. Acad. Sci. U.S.A. 97 (22) (2000) 12289–12294.
[10] J. Modrof, S. Becker, E. Mühlberger, Ebola virus transcription activator VP30 is a zinc-binding protein, J. Virol. 77 (5) (2003) 3334–3338.
[11] M. Weik, et al., Ebola virus VP30-mediated transcription is regulated by RNA secondary structure formation, J. Virol. 76 (17) (2002) 8532–8539.
[12] S.P. Reid, et al., Ebola virus VP24 proteins inhibit the interaction of NPI-1 subfamily karyopherin α proteins with activated STAT1, J. Virol. 81 (24) (2007) 13469–13477.
[13] E. Adu-Gyamfi, et al., The Ebola virus matrix protein penetrates into the plasma membrane, J. Biol. Chem. 288 (8) (2013) 5779–5789.
[14] J.J. Madara, et al., The multifunctional Ebola virus VP40 matrix protein is a promising therapeutic target, Future Virol. 10 (5) (2015) 537–546.
[15] L. Banadyga, M.A. Dolan, H. Ebihara, Rodent-adapted filoviruses and the molecular basis of pathogenesis, J. Mol. Biol. 428 (17) (2016), 3449-3449.
[16] S. Yamaoka, et al., Small animal models for studying filovirus pathogenesis, Curr. Top. Microbiol. Immunol. 411 (2017) 195–227.
[17] M.C. St Claire, et al., Animal models of ebolavirus infection, Comp. Med. 67 (3) (2017) 253–262.
[18] D. Bente, et al., Disease modeling for Ebola and marburg viruses, Dis Model Mech 2 (1–2) (2009) 12–17.
[19] M. Bray, et al., A mouse model for evaluation of prophylaxis and therapy of Ebola hemorrhagic fever, J. Infect. Dis. 179 (Suppl 1) (1999) S248–S258.
[20] A.A. Chepurnov, N.M. Zubavichene, A.A. Dadaeva, Elaboration of laboratory strains of Ebola virus and study of pathophysiological reactions of animals inoculated with these strains, Acta Trop. 87 (3) (2003) 321–329.
[21] A.A. Chupurnov, et al., [Attempts to develop a vaccine against Ebola fever], Vopr. Virusol. 40 (6) (1995) 257–260.
[22] A.A. Dadaeva, et al., [Hematological and immunological parameters during Ebola virus passages in Guinea-pigs], Vopr. Virusol. 51 (4) (2006) 32–37.
[23] E. Subbotina, et al., Genetic factors of Ebola virus virulence in Guinea pigs, Virus Res. 153 (1) (2010) 121–133.
[24] E. Ryabchikova, et al., Ebola virus infection in Guinea pigs: presumable role of granulomatous inflammation in pathogenesis, Arch. Virol. 141 (5) (1996) 909–921.
[25] B.M. Connolly, et al., Pathogenesis of experimental Ebola virus infection in Guinea pigs, J. Infect. Dis. 179 (Suppl 1) (1999) S203–S217.
[26] M. Chan, et al., Generation and characterization of a mouse-adapted makona variant of Ebola virus, Viruses 11 (11) (2019).
[27] G. Wong, et al., Development and characterization of a Guinea pig-adapted Sudan virus, J. Virol. 90 (1) (2016), 392-392.
[28] J. Quick, et al., Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples, Nat. Protoc. 12 (6) (2017) 1261–1276.

[29] S. Andrews, FastQC: a quality control tool for High throughput sequence data, Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, 2010.
[30] K. Okonechnikov, et al., Unipro UGENE: a unified bioinformatics toolkit, Bioinformatics 28 (8) (2012) 1166–1167.
[31] H. Wickham, et al., Welcome to the tidyverse, J. Open Source Softw. 4 (43) (2019) 1686.
[32] T.L. Pedersen, patchwork: The Composer of Plots . R package version 1.2.0 (2024).
[33] F. Keck, Handling biological sequences in R with the bioseq package, Methods Ecol. Evol. 11 (12) (2020) 1728–1732.
[34] J.E. Carette, et al., Ebola virus entry requires the cholesterol transporter Niemann–Pick C1, Nature 477 (7364) (2011) 340–343.
[35] M. Côté, et al., Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection, Nature 477 (7364) (2011) 344–348.
[36] T. Hoenen, et al., Oligomerization of Ebola virus VP40 is essential for particle morphogenesis and regulation of viral transcription, J. Virol. 84 (14) (2010), 7053-7053.
[37] R.G. Panchal, et al., In vivo oligomerization and raft localization of Ebola virus protein VP40 during vesicular budding, Proc. Natl. Acad. Sci. U.S.A. 100 (26) (2003) 15936–15941.
[38] Z.A. Bornholdt, et al., Structural rearrangement of Ebola virus VP40 begets multiple functions in the virus life cycle, Cell 154 (4) (2013) 763–774.
[39] M.C. Clifton, et al., High-resolution crystal structure of dimeric VP40 from Sudan ebolavirus, J. Infect. Dis. 212 (Suppl 2) (2015) S167–S171. Suppl 2.
[40] L. Zinzula, et al., Structures of Ebola and reston virus VP35 oligomerization domains and comparative biophysical characterization in all ebolavirus species, Structure 27 (1) (2019) 39–54.e6.
[41] V.P. Reddy Chichili, V. Kumar, J. Sivaraman, Linkers in the structural biology of protein–protein interactions, Protein Sci. 22 (2) (2013) 153–167.
[42] H. Wei, et al., Deep-sequencing of Marburg virus genome during sequential mouse passaging and cell-culture adaptation reveals extensive changes over time, Sci. Rep. 7 (1) (2017) 3390.
[43] V.E. Volchkov, et al., Molecular characterization of Guinea pig-adapted variants of Ebola virus, Virology 277 (1) (2000) 147–155.
[44] S.V. Cheresiz, E.A. Semenova, A.A. Chepurnov, Adapted lethality: what we can learn from Guinea pig-adapted Ebola virus infection model, Adv Virol 2016 (2016) 8059607.
[45] R.W. Cross, et al., Modeling the disease course of zaire ebolavirus infection in the outbred Guinea pig, J. Infect. Dis. 212 (Suppl 2) (2015) S305–S315.
[46] M. Mateo, et al., VP24 is a molecular determinant of Ebola virus virulence in Guinea pigs, J. Infect. Dis. 204 (suppl_3) (2011) S1011–S1020.