Insights on the Adaptation of Sudan Virus in Guinea Pigs and Its Implications in Viral Pathogenesis

by

Karla Danielle Alvariño Emeterio

A Thesis Submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Medical Microbiology and Infectious Diseases
University of Manitoba
Winnipeg

ABSTRACT

Sudan virus (SUDV) is a filovirus that belongs to the genus *Ebolavirus*. SUDV has been associated with several human cases of severe viral hemorrhagic fever in Sudan and Uganda. The average case fatality rate is about 50%. However, there are no licensed therapies or prophylactics for the treatment of SUDV disease. Initial evaluations of the effectiveness of new filovirus countermeasures are done in small rodent animals such as mice, hamsters, and guinea pigs. However, filoviruses are apathogenic in immunocompetent rodents, thus necessitating virus adaptation through serial passaging in the host. The virus acquires genomic mutations that lead to increased virulence and lethality during this process. In 2015, the first and only rodent animal model for SUDV was developed in guinea pigs. The repeated passaging of SUDV in the animal resulted in a uniformly lethal SUDV with 16 mutations of unknown significance. Here we show the timeline of mutation appearance during the adaptation process and at what frequencies they occurred by utilizing an amplicon-based high throughput sequencing approach. It was found that the viral genome acquired multiple transient mutations during the adaptation process, only 18 of which were retained in the guinea pig adapted SUDV. Most of these mutations were at a 99% frequency by passage 17. Three of the adaptive mutations were also already present in the starting virus and increased in frequency over time. Unique to this study was the identification of a novel mutation in the VP40 gene that existed early in the series and was selected for by the virus as passaging progressed. The roles that these 18 adaptive mutations play in the pathogenesis of SUDV may be related to viral replication and/or immune evasion. Overall, this study highlights hotspots within the viral genome that might be important in conferring increased virulence phenotype in the guinea pig host.

ACKNOWLEDGEMENTS

Nurturing my passion for scientific research for the last three years has been a rollercoaster ride but thankfully, the support of colleagues, mentors, and loved ones made the ride far less scary.

First, I extend my deepest gratitude to Dr. Xiangguo Qiu. The time spent was short-lived, but my heart is grateful for the opportunity you have given me to be able to do science in one of Canada's leading research facilities. Also, my sincerest thank you to my supervisor, Dr. Michael Drebot, for taking on the role on such short notice and for lending your expertise to guide me, keep me on track, and for encouraging me to pursue my goals. To Lisa Fernando and Dr. David Safronetz, thank you for being a support system along with my supervisor during the toughest times. Thank you, Drs. Grant McClarty, and Shawn Babiuk, for serving on my committee and challenging me with questions that kept my brain cells working. Finally, to Dr. Logan Banadyga, I am immensely thankful for your desire to bring out the best of my abilities during my time as a grad student. You have taught me the foundations of doing good science and continuously challenged me to try new experiences that nurtured my scientific career. Your constant support and encouragement are also much appreciated. Thank you for being the best mentor I could ever ask for as a grad student.

This work wouldn't be complete without the help of many colleagues from Special Pathogens. Thank you to Dr. Shihua He for taking care of the samples from the containment level 4 lab. To Dr. Wenjun Zhu, I am grateful for your help in getting me started with my project and for sharing your expertise and experiences with PCR and next-generation sequencing. To Dr. Jonathan Audet, many thanks for being there whenever I can't wrap my head around the concepts of next-generation sequencing and for helping me with my bioanalyzer assays. Thank you, Mr. Anders Leung as well, for your assistance with the magnetic bead-based extraction and for lending your knowledge and insights from your experiences to help me troubleshoot my experiments. Lastly, a special thank you to the Genomics Core at the National Microbiology Laboratory for being of huge assistance to this project by aiding in making the sequencing libraries for my samples and for sequencing them.

My VDAT lab colleagues, present and past, thank you for hosting me and making my experience worthwhile since my time as a summer co-op student in the lab. Thank you for also sitting, listening, and giving your feedback on several of my practice runs whenever I had

presentations to do. To Zachary Schiffman, thank you for being a friend with whom I can share the ups and downs of being a grad student. A special shoutout to my “Team Rocket” teammate, Michael Chan, for making my days in the lab full of fun conversations!

DEDICATION

This work is dedicated to my dearest siblings, Kim, Christian, Althea, Paulo, Tina, and Jed, and my soon-to-be husband, Josh, who have provided continuous support and encouragement as I faced the adversities of graduate school and life. I am lucky and blessed to have you all in my life.

This thesis is also dedicated to my parents, Mario and Teresa, who taught me to work hard and seize the opportunities life has to offer you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

α	alpha
A	adenosine
Ala (A)	alanine
aa	amino acid
Arg (R)	arginine
Asn (N)	asparagine
Asp (D)	aspartate/aspartic acid
ATCC	American Type Culture Collection
β	beta
BCE	Before Common Era
BDBV	Bundibugyo virus
BGS	bovine growth serum
BHK	baby hamster kidney
BOMV	Bombali virus
bp	base pair
BSL	biosafety laboratory
°C	Celsius
CDC	Centers for Disease Control and Prevention
cDNA	complementary deoxyribonucleic acid
CDS	coding sequence
CSCHAH	Canadian Science Centre for Human and Animal Health
Ct	cycle threshold
CTD	C-terminal domain
Cys (C)	cysteine
Da	Dalton
DC	dendritic cell
DC-SIGN	dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DPI	days post-infection
EBOD	Ebola disease
EBOV	Ebola virus
EDTA	ethylenediaminetetraacetic acid
Endo H	endoglycosidase H
er	endoplasmic reticulum
ESCRT	endosomal sorting complex required for transport
F	forward
FIWIV	Fiwi virus
g	gram
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GEQ	genome equivalent
Gb	gigabase
Gln (Q)	glutamine

Glu (E)	glutamate/glutamic acid
GP	glycoprotein
GPA	guinea pig adapted
Gly (G)	glycine
His (H)	histidine
HR	heptad repeat
HS	high sensitivity
HUJV	Huángjiāo virus
IDT	Integrated DNA Technologies
IFN	interferon
Ile (I)	isoleucine
IR	intergenic region
IRF	interferon regulatory factor
IRIDA	Integrated Rapid Infectious Disease Analysis
k	kilo
KNDV	Kander virus
L-SIGN	liver/lymph node-specific ICAM-3-grabbing integrin
Leu (L)	leucine
LLOV	Lloviu virus
Lys (K)	lysine
MARV	Marburg virus
µ	micro-
m	milli-
M	molar
MA	mouse adapted
MDCK	Madin-Darby canine kidney
Met (M)	methionine
MLAV	Měnglà virus
MLD	mucin-like domain
mRNA	messenger RNA
MVD	Marburg virus disease
MVP II	MagMAX Viral/Pathogen II
l	litre
NIH	National Institutes of Health
ng	nanogram
NHP	nonhuman primate
nm	nanometer
NP	nucleoprotein
NPC1	Niemann-Pick C1
nt	nucloetide
NTD	N-terminal domain
OBLV	Oberland virus
ORF	open reading frame
PBGD	porphobilinogen deaminase
PCR	polymerase chain reaction
PFU	plaque forming unit

PHAC	Public Health Agency of Canada
Phe (F)	phenylalanine
PM	plasma membrane
PP	phosphatase protein
Pro (P)	proline
QC	quality control
R	reverse
RACE	rapid amplification of cDNA ends
RAVV	Ravnn virus
RBD	receptor binding domain
RESTV	Reston virus
RIN	RNA integrity number
RNA	ribonucleic acid
RNP	ribonucleoprotein
rpm	rotations per minute
RT–PCR	Reverse transcription-polymerase chain reaction
RT–qPCR	Real-time quantitative polymerase chain reaction
SBEV	Sudan virus (Boniface)
Ser (S)	serine
sGP	soluble glycoprotein
SMYD	SET and MYND-containing domain protein
SNP	single nucleotide polymorphism
ss	single-stranded
ssGP	small soluble glycoprotein
SUDV	Sudan virus
SUDVB	Sudan virus (Boniface)
TAE	Tris-acetate-EDTA
TAFV	Taï Forest virus
$TCID_{50}$	50% tissue culture infectious dose
TIM	T-cell immunoglobulin and mucin domain
Thr (T)	threonine
TNF	tumor necrosis factor
tRNA	transfer RNA
Trp (W)	tryptophan
trVLP	transcription and replication competent virus-like particle system
Tyr (Y)	tyrosine
U	uridine
UTMB	University of Texas Medical Branch
UTR	untranslated region
Val (V)	valine
VP	viral protein
XILV	Xīlǎng virus
Zn	zinc

1 INTRODUCTION

1.1 FILOVIRUSES

1.1.1 Phylogenetic classification

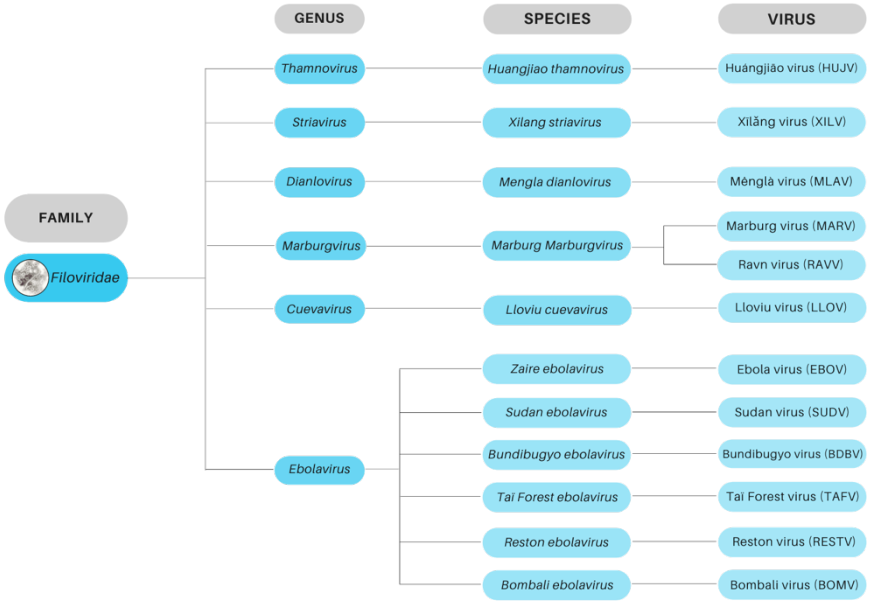


Figure 1.1 Taxonomical classification of filoviruses. The family *Filoviridae* includes six genera, four of which are composed of a single species that includes a single virus; while the genus *Marburgvirus*, consisting of a single species, includes two viruses. The genus *Ebolavirus* is distinctly further subdivided into six species, each typified by one virus.

The *Filoviridae* family is a member of the *Mononegavirales* order. It is made up of single-stranded, non-segmented negative-sense enveloped RNA viruses (Bourgarel & Liégeois, 2019). They are commonly known for their long, thread-like filamentous morphology with a length of up to 14, 000 nm and a width of around 80 nm (Geisbert & Jahrling, 1995). According to Kuhn et al. (2020), members of the family *Filoviridae* are further divided into six genera: *Cuevavirus*, *Dianlovirus*, *Ebolavirus*, *Marburgvirus*, *Striavirus*, and *Thamnovirus*. The genera *Cuevavirus*, *Dianlovirus*, *Striavirus*, and *Thamnovirus* are each comprised of a single species that includes a single virus, namely, *Lloviu cuevavirus* (Lloviu virus), *Mengla dianlovirus* (Měnglà virus [MLAV]), *Xilang striavirus* (Xīlǎng virus), and *Huangjiao thamnovirus* (Huángjiāo virus [HUJV]), respectively. The genus *Marburgvirus* includes one species, the *Marburg Marburgvirus*, consisting of two viruses, the Marburg virus and Ravn virus. Meanwhile, the genus *Ebolavirus* is subdivided into six different species, each represented by a single virus: *Bombali ebolavirus* (Bombali virus), *Bundibugyo ebolavirus* (Bundibugyo virus), *Reston ebolavirus* (Reston virus),

Sudan ebolavirus (Sudan virus), *Taï Forest ebolavirus* (Tai Forest virus), and *Zaire ebolavirus* (Ebola virus). Taxonomical classifications of various viruses within the family Filoviridae are summarized in Figure 1.1. The assignment of species to a specific genus classification is based on several factors, including genomic features, such as the number and location of gene overlaps, the number of open reading frames (ORFs) and/or genes, the host and geographic distribution of filoviruses and the pathogenicity of filoviruses for different organisms are also taken into account for the assignment of genus (Kuhn et al., 2020).

1.1.2 Emergence of filoviruses

1.1.2.1 Marburg virus

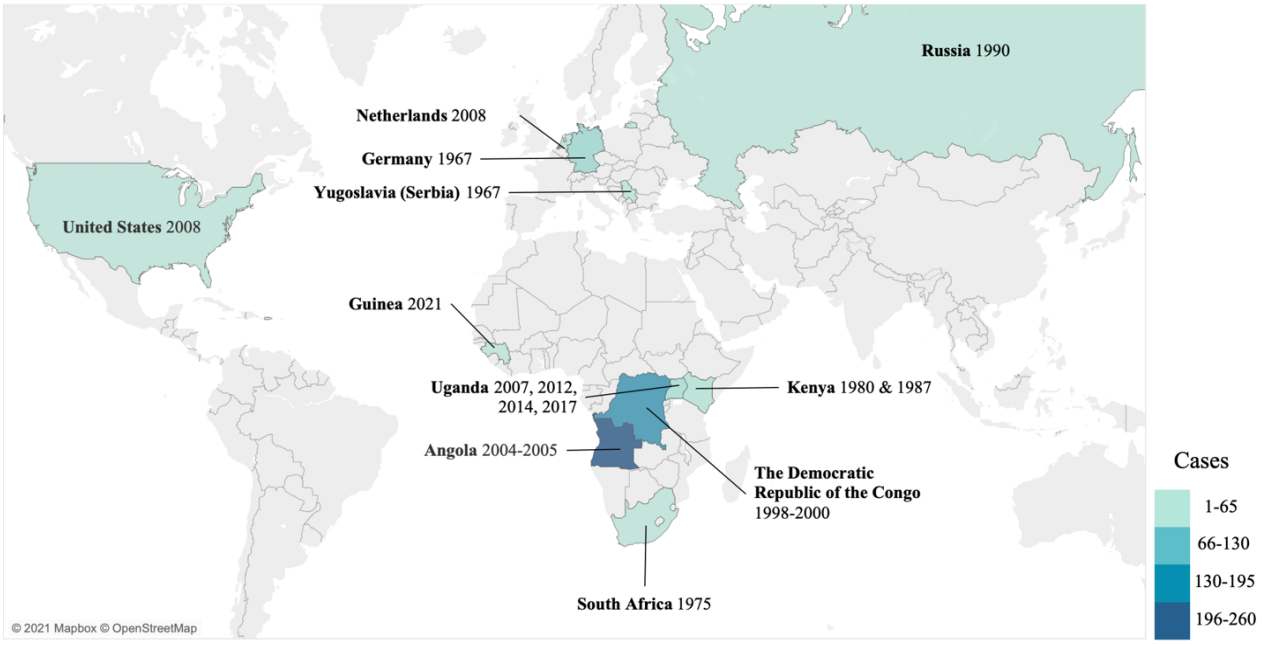


Figure 1.2 Countries with reported and isolated cases of Marburg virus disease since 1967. Total number of cases are presented in each identified country. Data was obtained from Centers for Disease Control and Prevention and generated using the Tableau software.

The first-ever filovirus to be identified was recognized in 1967 when laboratory workers in Marburg and Frankfurt, Germany, and Belgrade, Yugoslavia developed a hemorrhagic fever after handling African green monkeys from Uganda (Feldmann et al., 1996). Ill patients showed signs of malaise, muscle pain, headache, high fever, and various gastrointestinal symptoms such as vomiting, nausea, and diarrhea (Slenczka & Klenk, 2007). The etiological agent was later isolated from patients and African green monkeys and was named Marburg virus (MARV) (Kissling et al.,

1968; Kunz et al., 1968; Siegert et al., 1968). After about two decades since its emergence, a new representative virus in the *Marburgvirus* species, Ravn virus (RAVV), was isolated in 1987 from a fatal case of Marburg virus disease (MVD) in Kenya (Johnson et al., 1996). To date, 14 human outbreaks of MVD have been recorded since 1967 (Centers for Disease Control and Prevention, 2021b) (Fig. 1.2). Due to its high infectivity and mortality rate, as well as the lack of licensed prophylactics and therapeutics, MARV is classified as a biological agent of Biosafety Level 4 (BSL-4) (Brauburger et al., 2012; Meechan & Potts, 2020).

1.1.2.2 Ebolaviruses

1.1.2.2.1 Ebola virus

It was not until 1976 when other filoviruses were recognized when two simultaneous but separate epidemics of hemorrhagic fever occurred in the Democratic Republic of the Congo (DRC) (formerly Zaire) and South Sudan (formerly Sudan), which were associated with the Ebola virus (EBOV) and the Sudan virus (SUDV), respectively (WHO, 1978a, 1978b). The first human case of EBOV infection was identified in late August 1976 when a resident of Yandogi Village presented chills, fever, headache nausea, malaise, abdominal pain, and intestinal bleeding at Yambuku Mission Hospital after a two-week trip to northern Zaire (Breman et al., 2016; WHO, 1978b). Dozens of individuals subsequently succumbed to the same symptoms as in the index case (Breman et al., 2016). However, the disease that plagued Yambuku at that time was unknown, and it was not until September 1976 that Dr. Peter Piot and his colleagues from Belgium identified the pathogen of interest (Aurelie et al., 2018; Breman et al., 2016). EBOV, named after the Ebola River in DRC, appeared cylindrical like MARV under the electron microscope but was immunologically distinct (Breman et al., 2016; WHO, 1978b). The first Ebola virus disease (EVD) outbreak resulted in 318 reported cases with 280 deaths for a case fatality rate of 88% (WHO, 1978a). Since then, EBOV outbreaks have occurred periodically throughout Africa, with the largest outbreak ever recorded in West Africa from 2013 to 2016 (Jacob et al., 2020). The initial case of EBOV began with an 18-month-old boy living in Guinea that was believed to have been infected by bats (Centers for Disease Control and Prevention, 2019). The person-to-person spread of EBOV occurs via direct contact with bodily fluids of an infected person; therefore, the widespread infection of EBOV was likely fueled by cultural beliefs and burial rites existing within

affected communities, which involve direct contact with the deceased person (Cenciarelli et al., 2015). Furthermore, the prevalence of international travel today could also have propagated the spread of EBOV infection outside of West Africa. The Ebola epidemic in West Africa infected more than 28,000 people and killed more than 11,000 people, Guinea, Liberia and Sierra Leone being the most affected countries (Fig. 1.3). Over many decades, EBOV became the most well-known filovirus, accounting for most recorded outbreaks, as well as the largest outbreak in the history of filoviruses.

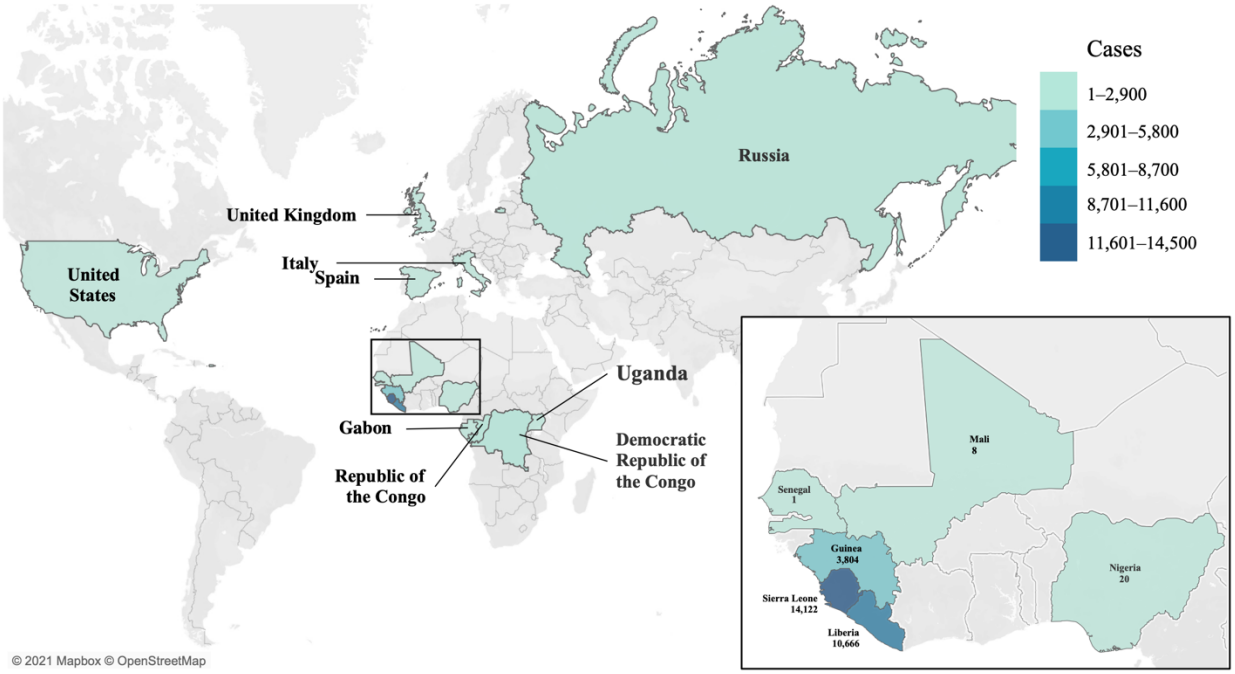


Figure 1.3 Countries with reported cases of Ebola virus infection since 1976. The total number of reported cases is presented in each country that was affected by EBOV. Highlighted in a box are countries in West Africa that were heavily impacted during the West African Ebola outbreak from 2013-2016. Data was obtained from CDC and Aurelie et al. (2018). The map was generated using Tableau software.

1.1.2.2.2 Sudan virus

Two months before the first EBOV outbreak in DRC, a similar viral hemorrhagic fever outbreak was already occurring in the town of Nzara, South Sudan (Breman et al., 2016). The first cases were determined to have originated from three cotton factory workers who experienced severe febrile illness and profuse bleeding (WHO, 1978a). The unknown disease of Sudanese origin traveled to the neighbouring towns of Juba, Tambura, and Maridi, where one, three and 213

people became infected, respectively (WHO, 1978a). There were 284 cases and 151 deaths, a 53% case fatality rate, that stemmed from this outbreak (WHO, 1978a). The radioimmunoassay performed on the two virus isolates from the 1976 outbreaks by Richman et al. (1983) confirmed that the two viruses were antigenically different. SUDV outbreaks have been prominent in east Africa, particularly in South Sudan and Uganda (Fig. 1.4), with an average case fatality rate of about 50%. Currently, SUDV has been responsible for seven filovirus outbreaks since 1976 (Aurelie et al., 2018) (Fig.1.4), the outbreak in Uganda in 2000 being the largest SUDV epidemic (Okware et al., 2002). The SUDV outbreak in 2002 resulted in 425 cases with 224 deaths (Okware et al., 2002). Unlike EBOV, currently there are no licensed therapeutics and prophylaxis available for the treatment of SUDV disease (Herbert et al., 2020). Although there are studies that provide potential therapeutics for the treatment of SUDV disease, including a lipid-encapsulated small interfering RNA that targets a viral protein of SUDV, favipiravir, and antibodies that are cross-reactive with other ebolaviruses and filoviruses (Herbert et al., 2020).

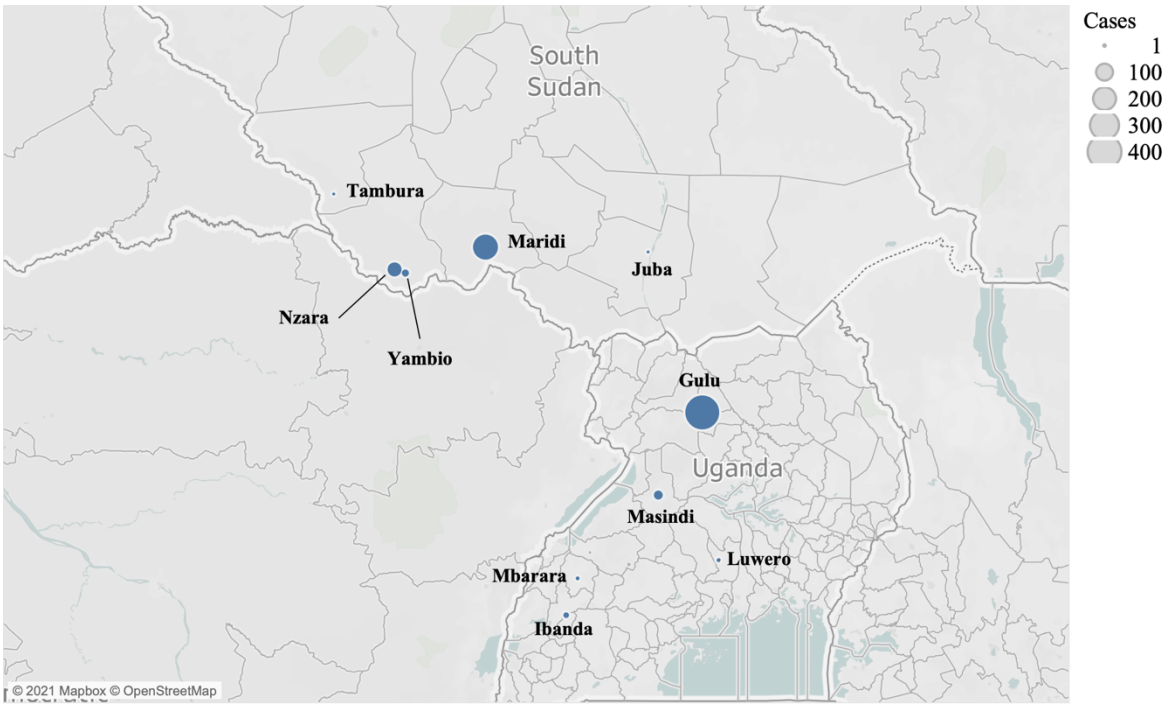


Figure 1.4 Recorded outbreaks in eastern Africa associated with Sudan virus (SUDV). The total number of reported cases is presented in each county that was affected by SUDV. Data was obtained from CDC. The map was generated using Tableau software.

1.1.2.2.3 Bundibugyo virus

The Bundibugyo virus (BDBV) made its debut in 2007 in the district of Bundibugyo, Uganda (Centers for Disease Control and Prevention, 2021a). A total of 131 cases of hemorrhagic fever were identified: 44 suspected, 21 probable, and 56 laboratory confirmed cases (Centers for Disease Control and Prevention, 2021a; Macneil et al., 2011; Macneil et al., 2010). Like previous outbreaks of filovirus disease, confirmed cases experienced fever, fatigue, headache, myalgia, and various gastrointestinal distress such as abdominal pain, diarrhea, and vomiting (Macneil et al., 2010). The etiological agent was later confirmed by the Centers for Disease Control and Prevention (CDC) to be of filovirus origin and distinct from any other previously identified ebolaviruses (Towner et al., 2008). Five years later, BDBV re-emerged in 2012 in DRC where 64 cases and 34 deaths (38 cases and 13 deaths were laboratory confirmed) were reported (Centers for Disease Control and Prevention, 2021a; Hulseberg et al., 2021; Kratz et al., 2015).

1.1.2.2.4 Taï Forest virus

Unlike the other ebolaviruses mentioned above, the Taï Forest virus (TAFV) has only been associated with a single human case that did not result in a fatality. The virus was first identified in 1994 by a scientist who conducted a post-mortem examination of a chimpanzee originating from Taï National Park in western Côte-d'Ivoire (Centers for Disease Control and Prevention, 2021a). The 34-year-old female developed dengue-like symptoms but later recovered without sequelae (Le Guenno & Formentry, 1995). Subsequently, virus particles with filovirus morphology were detected from cells infected with the patient's sera and determined to be serologically related to earlier identified ebolaviruses (Le Guenno & Formentry, 1995). The source of TAFV infection for this individual is strongly believed to be from the necropsied chimpanzee from Côte-d'Ivoire, since the biopsy results from this animal showed the presence of "multifocal necrosis in the liver infiltrated with inflammatory cells and diffuse fibrinoid necrosis in the red pulp of the spleen" such as those found in macaques infected with EBOV (Wyers et al., 1999). Furthermore, the presence of the TAFV viral antigen in macrophages of infected animal tissues was confirmed by the reactivity of specific antibodies for and cross-reactive with TAFV. No further human cases have been reported since then, and TAFV research has remained a low priority. Although TAFV has not caused any significant outbreaks since its discovery in 1994, this virus should still be of

importance as very little is known about this filovirus and the likelihood of an outbreak occurring cannot be eliminated (Burk et al., 2016).

1.1.2.2.5 Reston virus

The Reston virus (RESTV) is not known to cause any disease in humans. RESTV was first detected in 1989 in macaques (*Macaca fascicularis*) that were exported from the Philippines to Reston, Virginia, United States (Centers for Disease Control and Prevention, 2021a; Hayes et al., 1992; Miranda et al., 1991). Electron microscopy of cells inoculated with serum and organ homogenates from sick monkeys revealed the presence of viral particles that were morphologically similar to previously identified filoviruses and determined to be serologically related to EBOV (Cantoni et al., 2016; Geisbert & Jahrling, 1990). However, unlike EVD symptoms exhibited by RESTV-infected monkeys, handlers working at the facility that houses the animals in the United States and the Philippines did not show symptoms despite serological evidence of infection (Cantoni et al., 2016; Centers for Disease Control and Prevention, 1990). Currently, RESTV remains avirulent in humans and has only caused minor outbreaks in monkeys and pigs (Cantoni et al., 2016). Since RESTV is apathogenic in humans, research on this ebolavirus is also limited, like TAFV. However, the importance of getting a better understanding of the pathogenesis of RESTV should also be emphasized to determine the likelihood of the virus adapting and causing a huge outbreak, and because of its potential to become an agricultural burden for the pig farm industry.

1.1.3 Discovery of new filoviruses

In the last decade, several new filoviruses have been discovered but are not yet known to cause disease in humans. The discovery of the Lloviu virus (LLOV) was prompted by several deaths of insectivorous Schreiber's bat colonies (*Miniopterus schreibersii*) in different parts of Europe, such as France, Spain, and Portugal in 2002 (Negredo et al., 2011). In 2011, LLOV filovirus-like RNA was detected in deceased *M. schreibersii* bat species collected from Cueva del Lloviu in Asturias, Spain, and from caves in Cantabria, Spain during bat deaths in 2002 (Negredo et al., 2011). In 2016 LLOV resurfaced in Hungary but again was found only in *M. schreibersii* bat colonies (Kemenesi et al., 2018). Within the African region, a survey conducted to identify potential host reservoirs of EBOV and identify new filoviruses in Sierra Leone led to the discovery of a new

ebolavirus, Bombali virus (BOMV) (Goldstein et al., 2018). Viral RNA was detected in rectal and oral swabs obtained from insectivorous free-tailed bats (species: *Mops condylurus* and *Chaerephon pumilus*) roost in housing infrastructures (Goldstein et al., 2018). In 2019, BOMV RNA was detected in other parts of sub-Saharan Africa, such as Kenya and Guinea, in tissues and extracts of *M. condylurus* bats (Forbes et al., 2019; Karan et al., 2019; Kareinen et al., 2020). In addition, in the eastern hemisphere, serological evidence of filovirus infection has been detected in *Rousettus leschenaultia* (fruit bat), *Pipistrellus pipistrellus* (microbat) and *Myotis* (mouse-eared bat) bat species in China (Yuan et al., 2012). Viral RNA closely related to filoviruses has also been found to be prevalent in different fruit bat populations in China (He et al., 2015; Yang et al., 2017); and in 2019 a new bat-borne virus, named Měnglà virus (MLAV), originating from Měnglà County, Yunnan Province, China, was introduced into the family *Filoviridae* (Yang et al., 2019). Unlike LLOV, BOMV, and MLAV, Xīlǎng (XILV) and Huángjiāo (HUJV) viruses were identified in fish hosts; particularly in frogfish and species of *Thamnaconus septentrionalis* fish, collected in the coastal region of Wenling in Zhejiang, China (Kuhn et al., 2020; Shi et al., 2018). The discovery of new filoviruses in fish continues to this day. In fact, three new filoviruses were recently discovered in European perch grown in Switzerland by Hierweger et al. (2021) - the Fiwi virus (FIWIV), Oberland virus (OBLV) and Kander virus (KNDV). Characterization of the viral genomes of these three viruses and phylogenetic analysis revealed that the viruses are closely related to HUJV, which is also a fish filovirus (Hierweger et al., 2021). The FIWIV and KNDV genome sequences were 49% divergent from the HUJV genome and identified as novel species of the *Thamnovirus* genus. Meanwhile, OBLV was more than 62% divergent from HUJV and was therefore assigned its own genus named *Oblavirus*. Despite the recovery of filovirus-related genomic sequences in the deceased fish, attempts to detect viral antigens in fish tissues were unsuccessful (Hierweger et al., 2021).

1.2 BIOLOGY OF EBOLAVIRUSES

Most of the research about members of the genus ebolaviruses has been invested in EBOV species due to its higher case fatality rate and the greater number of outbreaks it caused over many decades. Materials described in this section will be primarily in the context of EBOV.

1.2.1 Genomic composition and virion structure

The genome organization of different ebolaviruses is similar. The genome is a single-stranded, negative-sense, and non-segmented RNA with a size of approximately 19 kilobases (kb) (Fig.1.5). Seven genes are encoded in the genome, including nucleoprotein (NP), viral protein (VP) 35, VP40, glycoprotein (GP), VP30, VP24, and L polymerase. The genome is flanked by leader and trailer sequences at the 3' and 5' ends, respectively (Baseler et al., 2017; Cantoni & Rossman, 2018; Kirchdoerfer et al., 2017). This conserved leader and trailer regions consist of replication and transcription promoters and packaging signals (Baseler et al., 2017). The beginning and end of each gene are bordered by transcription start and termination signals. The regulatory signals between viral genes are separated by varying lengths of intergenic regions (IR) that do or do not overlap. In ebolaviruses, the gene boundaries between the genes VP35 and VP40, GP and VP30, and the VP24 and L overlap with each other (Kuhn et al., 2020). All genes, except the GP gene, which encodes three proteins, are monocistronic (Volchkova et al., 2011). A total of nine different proteins are produced from the genome, seven of which are structural proteins (NP, VP35, VP40, GP, VP30, VP24, and L polymerase) and two secreted proteins from the *GP* gene (Elliott et al., 1985). In this thesis, the coding sequence and noncoding sequence region terminologies will be heavily used. Therefore, for the purposes of this thesis, the coding sequence refers to regions within the genome that encode an amino acid, while the non-coding sequence region pertains to parts of the genome that do not encode an amino acid, including the leader and trailer sequence, transcription start and termination signals, the intergenic regions between genes, and the untranslated regions of the genome.

Ebolaviruses are morphologically filamentous with a diameter of 96-98 nm and lengths greater than 200 nm (Kuhn et al., 2020). They are made up of a host-derived lipid envelope that surrounds the core ribonucleoprotein (RNP) complex, a process facilitated by VP40 as it becomes recruited to the inner leaflet of the host plasma membrane (PM) and forms oligomers to mediate virus assembly in the PM (Amiar & Stahelin, 2020). Part of this RNP complex is the viral RNA encapsidated by multiple viral proteins such as NP, VP35, VP30, VP24 and L polymerase (Jain et al., 2021). The nucleocapsid appears as a supercoiled helical structure under the electron microscope (Bharat et al., 2012; Sugita et al., 2018). Finally, the GP, which measures

approximately 7 nm in width, coats the surface of the virus (Kuhn et al., 2020). The roles of ebolavirus proteins are discussed further in Chapter 1.2.2.

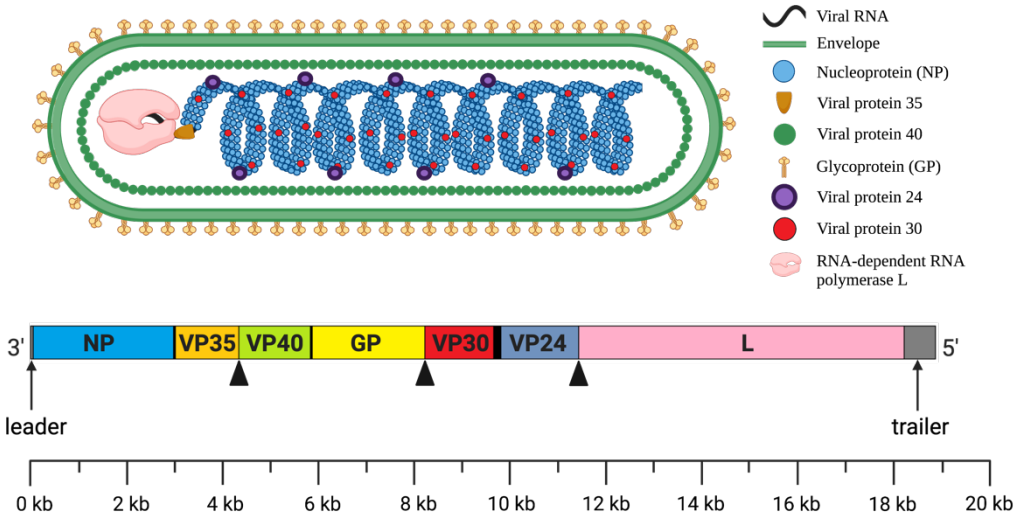


Figure 1.5 Ebolavirus virion structure and genome organization. Above is a schematic representation of an ebolavirus virion. Below is the non-segmented, negative-sense and single-stranded RNA genome flanked with conserved leader and trailer sequences (grey boxes) at the 3' and 5' ends, respectively. The genome encodes seven genes: nucleoprotein (NP), viral protein (VP) 35, VP40, glycoprotein (GP), VP30, VP24, and L polymerase, that are separated by intergenic regions (IR) (black boxes). Indicated by black triangles are IRs that overlap. Created with BioRender.com

1.2.2 Proteins of ebolaviruses

1.2.2.1 Nucleoprotein

The NP protein is transcribed from the *NP* gene. It is a large protein with dimensions of approximately 53 Å x 40 Å x 75 Å and composed of 739 amino acids (aa), at least for EBOV, with hydrophobic N-terminal and hydrophilic C-terminal ends (Dong et al., 2015; Noda et al., 2010; Watanabe et al., 2006). Known as an essential component of nucleocapsid, NP facilitates the encapsidation of genomic RNA by binding to viral RNA, resulting in a helical complex that offers protection against recognition of host cell immune responses and ribonucleases and is crucial for the virus replication (Cantoni & Rossman, 2018; Leung et al., 2015). The length of the N-terminal domain (NTD) of the protein is approximately 450 aa and consists of an N-terminal arm that is made up of the N lobe (residues 36-240) and C lobe (residues 241-351) that form the core of the protein (Dong et al., 2015; Sugita et al., 2018). The N-terminus is then linked to a C-terminal

domain (CTD) that is approximately 98 aa long by a disordered linking region (Sugita et al., 2018). According to sequence homology, the N-terminal part of the protein is a conserved region between paramyxoviruses and filoviruses, while residues spanning the remainder of the 739 aa-long protein are a region unique to filoviruses (Leung et al., 2015; Su et al., 2018; Watanabe et al., 2006).

NP is capable of self-assembly (Watanabe et al., 2006). Structural analysis of the interaction of NP-NP by Sugita et al. (2018) reveals that an NP subunit "shares contact regions with two pairs of neighbours located above and below". The N-terminal arm is a flexible domain of the protein that facilitates the oligomerization of individual NP monomers by interacting with the hydrophobic pocket of neighbouring NP monomers (Lin et al., 2020; Sugita et al., 2018). However, the oligomerization process is not only facilitated by the first 24 amino acids of the protein, but other residues located in the core of the NP, such as aa 345-384 in the C-terminal lobe, have been shown to make contact with adjacent NP monomers (Su et al., 2018; Sugita et al., 2018). Comparison of mutants with varying lengths of truncated NTD by Watanabe et al. (2006) also identified that aa 1-450 is vital for the interaction NP-NP. Similarly, other studies on the nucleoprotein-nucleoprotein interaction of the N protein of paramyxoviruses, which share similar aa homology with the NP of the ebolavirus, have exemplified the importance of the NTD for such interaction (Watanabe et al., 2006). The association of NP monomers together generates a back-to-back double-ring configuration that is subsequently stacked and forms tube-like structures (Su et al., 2018). This formation of NP helix is deemed vital in nucleocapsid formation, as mutants with deleted aa 2-150 from the NTD failed to form nucleocapsid-like structures (Noda et al., 2010; Sugita et al., 2018; Watanabe et al., 2006).

The NP core region (residues 20-405) of the protein is also responsible for the encapsidation of the viral genome through hydrogen bonding and electrostatic interactions between the amino acids located in the highly positively charged groove of the core and the negatively charged phosphodiester backbone of the single-stranded RNA (ssRNA) (Dong et al., 2015; Noda et al., 2010; Sugita et al., 2018). This interaction results in the formation of a loosely coiled nucleocapsid-like structure (Bharat et al., 2012). Although the solid structure of the NP-RNA interaction is mainly sustained by the NP-NP interaction (Noda et al., 2010), the NP-RNA interaction is also essential to the lateral assembly of the NP monomers (Sugita et al., 2018). A study by Noda et al. (2010) demonstrated that upon treatment of NP-RNA complex with Rnase at low salt concentration the coiled structure disassembled into oligomers, implying that the RNA serves as a

foundation upon which the helical structure is built as it holds the NP molecules, aiding in the orderly packaging of NP monomers (Xu et al., 2020).

EBOV's NP core also interacts with other viral proteins such as VP35. The NP RNA-binding domain, which is composed of the N- and C-terminal lobes, forms a bilobed structure that includes a unique β-hairpin in the C-terminal lobe. This is similar to the nucleoproteins found in non-segmented negative-stranded RNA viruses (Kirchdoerfer et al., 2015). Landeras-Bueno et al. (2019) with the SUDV VP35-NP complex. The crystal structure of the SUDV VP35-NP complex uncovered a "characteristic bi-lobed configuration" with an RNA binding cleft (Landeras-Bueno et al., 2019). This structure appeared like the NP of other filoviruses, but also had unique features that set it apart from the others, specifically from EBOV. With a structural resolution of 2.3 Å, Landeras et al. identified the presence of a β-sheet that contained additional interaction sites that were necessary to maintain high-infinity contacts between the hydrophobic pocket of NP and VP35 (Landeras-Bueno et al., 2019).

Unlike the NTD of NP, which forms an ordered structure, the CTD (residues 641-739) part of the protein is primarily disordered (Kirchdoerfer et al., 2016). NP CTD is presumed to be involved primarily in the interaction with ebolavirus VP40 along with residues 2-150 found at the amino termini (Noda et al., 2007). By linking NP to VP40, NP can be embedded in a virus-like particle (Bharat et al., 2012; Watanabe et al., 2006). Located in the disordered region linking the NP core and C- terminal tail, there are three motifs (X denotes any amino acid): LxxIxE, SxPxLE, and PxPxY (Jain et al., 2021). The first two motifs interact with host proteins that regulate viral RNA transcription and replication (Jain et al., 2021) - the LxxIxE motif is needed for interaction with PP2A-B56 phosphatase that dephosphorylates NP-VP30 and allows viral transcription (Kruse et al., 2018), while SxPxLE is crucial for the interaction of SET and MYND-containing domain protein 3 (SMYD3) that increases the NP-VP30 interaction and further supports mRNA production (Chen et al., 2019). On the other hand, the PxPxY motif is needed for the NP-VP30 interaction, which is important for viral RNA transcription.

1.2.2.2 Viral protein 35

The multifunctional protein VP35 plays an important role in many aspects of the viral replication cycle. It is a polymerase cofactor and structural component of the nucleocapsid, as well

as a cofactor in viral RNA transcription and replication. EBOV VP35 is 340 aa in length and is made up of three known functional domains: an NP chaperoning domain at the N-terminus (residues 20-47), a homo-oligomerization domain (residues 83-145), and a double-stranded RNA (dsRNA) binding or interferon (IFN) inhibitory domain closest to the C-terminus end (residues 221-340) (Leung et al., 2009). The NP chaperoning domain of VP35 functions to accompany the monomeric NP in a free RNA state (Kirchdoerfer et al., 2015; Leung et al., 2015), a state vital for the accumulation of RNA-free monomeric NP before polymerization of NP to ensure that NP only oligomerizes and encapsidates viral RNA (Landeras-Bueno et al., 2019). A resolved crystal structure of the EBOV VP35-NP complex shows that the N-terminal peptide of VP35 almost lacks secondary structure with only 12 aa forming a helical conformation (Kirchdoerfer et al., 2015). The interacting peptide was shown to be composed of hydrophilic and hydrophobic regions, but it uses its hydrophobic face to interact with one side of the C-terminal lobe of the NP core in a straddling conformation (Kirchdoerfer et al., 2015).

In the case of EBOV, the oligomerization domain of VP35 is composed of 12 protein chains, each chain forming a continuous α-helix and assembling into trimers and tetramers in solution (Zinzula et al., 2019). Mass spectroscopy analyses of VP35 from other ebolaviruses revealed that an oligomeric state is also observed in SUDV, TAFV, BDBV, and RESTV, although the predominant state in solution was a tetramer (Zinzula et al., 2019). Chemical cross-linking experiments demonstrated that the VP35 tetramer is in a parallel arrangement (Chanthamontri et al., 2019). Studies in the past have revealed that region aa 80-120 covers a coiled-coil motif which controls the protein's ability to oligomerize to tetramers (Chanthamontri et al., 2019; Ramaswamy et al., 2018; Reid et al., 2005). Orthologous to the paramyxovirus and rhabdovirus phosphoprotein, VP35 is one of the main components of the RNP complex, as it bridges NP with L polymerase (Cantoni & Rossman, 2018; Groseth et al., 2009). Through the N-terminal aa 82-188, homo-oligomerization of VP35 is believed to be required for the VP35-L interaction (Jain et al., 2021; Trunschke et al., 2013) consequently relocating the L protein to inclusions derived from NP (Trunschke et al., 2013) and highlights the importance of VP35 in the formation of the RNP complex for viral RNA synthesis. Multimeric VP35 has also been shown to be valuable in the IFN antagonistic properties of the protein (Reid et al., 2005). Reid et al. (2005) concluded that the existence of an oligomerization activity is necessary for the IFN inhibitory domain of VP35 to fully execute the IFN antagonistic activity. However, recent exploration by Di Palma et al. (2019)

on whether a disrupted coiled-coil domain, which consequently affects oligomerization, revealed that oligomerization may not be a critical factor in VP35's ability to inhibit the host's immune response.

VP35 also has antagonistic properties against the host's type I alpha (α)/beta (β) IFN host response (Basler et al., 2003; Basler et al., 2000). During a viral infection, the production of IFN-α/β cytokines is induced because of the recognition of pathogen-associated molecular patterns, such as dsRNA, by the host's pathogen recognition receptors (Zinzula & Tramontano, 2013). The secretion of these cytokines then promotes viral clearance through the apoptosis of infected cells. However, transcription of IFN-α/β cytokines is inhibited due to the ability of VP35 to block the activation of an important transcription factor, IFN regulatory factor 3 (IRF-3) (Basler et al., 2003). VP35 has been shown to stop IRF-3 activation by phosphorylating IRF-3 (Basler et al., 2003). Furthermore, VP35 aids ebolaviruses in immune evasion by binding dsRNA. Among its functional domains is an IFN inhibitory domain and structural analysis of VP35 revealed that within this domain there are two subdomains comprised of four α-helices in the first domain and four β strands in the other domain, both of which are important for the folding and stability of VP35 (Leung et al., 2009; Leung et al., 2010). Furthermore, the inhibitory domain possesses two areas of basic residues that are highly conserved in the filovirus family. (Leung et al., 2009). The first basic path was found in the α-helical subdomain while the other was in the β-sheet subdomain (Leung et al., 2009; Leung et al., 2010). Residues (i.e. Arg305, Lys309 and Arg322) found in the second basic patch, also known as the central basic patch, were deemed important for immune suppression by binding of dsRNA (Leung et al., 2009). The binding of dsRNA is not recognized by retinoic acid-inducible gene I (RIG-I), a nucleic acid sensor that activates the RIG-I signaling pathway resulting in the transcription of IFN genes (Chanthamontri et al., 2019). Another immune evasion tactic associated with VP35 is its ability to obstruct the activation of the IFN-induced dsRNA protein kinase R (PKR), a phosphorylating enzyme that regulates the transcription of eukaryotic initiation factor 2 alpha (eIF2α) that is critical for mounting the cellular stress response (Sadler & Williams; Schümann et al., 2009). This immune evasion strategy has been associated with the protein's C-terminal IFN inhibitory domain, as alanine substitutions of basic amino acids within this domain overturned inhibition of PKR (Schümann et al., 2009). An investigation by

Fabozzi et al. (2011) and (Le Sage et al., 2017) also demonstrated that VP35 can mediate suppression of RNA silencing and blockage of stress granule assembly, respectively.

The role of the IFN inhibitory domain is not limited to immune evasion but also carries residues that support the VP35 polymerase cofactor role. Residues sitting outside the "central" basic patch were initially identified to be unimportant to the IFN inhibition function of VP35, but a study by Prins et al. (2010) show that these may be relevant in viral RNA synthesis as mutation of outer basic residues in the first basic patch showed decreased replication by the replication complex. Prins and colleagues further identified that there are residues also located in the first basic patch that make contact with NP (Prins et al., 2010). The significance of VP35 in the RNP complex is further exemplified by the presence of a phosphorylation site at aa 210, which has been shown to be required for the transcriptional activity of the EBOV and VP35-NP binding (Ivanov et al., 2020). Due to its multi-faceted functional roles in the virus lifecycle and multiple evasion tactics, VP35 is considered an important determinant of ebolavirus virulence.

1.2.2.3 Viral protein 40

VP40 is the matrix protein of ebolaviruses that associates with the host lipid bilayer and forms macromolecular complexes (Noda et al., 2007). Produced as the most abundant protein by EBOV, VP40 is involved in virion assembly and budding during infection (Adu-Gyamfi et al., 2014; Madara et al., 2015). The protein is 326 aa long with a molecular weight of 40 kDa (Hoenen et al., 2010; Madara et al., 2015). The protein crystal structure of monomeric VP40 reveals an NTD (residues 44-194) and a CTD (residues 201-321) that folds into β-sandwich structures (Dessen, 2000; Madara et al., 2015). Found in the protein is also a disordered interdomain connecting region (residues 195-200) that contains a zipper of hydrophobic residues (Dessen, 2000; Hoenen et al., 2010). NTD is known to aid with dimerization and oligomerization, while CTD is required for membrane binding (Hoenen et al., 2010).

VP40 is capable of achieving various oligomeric states – dimers, hexamers, an octameric ring, and filaments (Jain et al., 2021) (Fig 1.6). Variation in the structural arrangement of the protein can confer on multiple functions of VP40. VP40 was initially presented as a monomer by crystallographic structures presented by Dessen (2000); however, it has been determined that VP40 actually exists as a dimer in the solution (Bornholdt et al., 2013) (Fig. 1.6A). Dimerization is

primarily facilitated by the NTD interface (Adu-Gyamfi et al., 2014; Bornholdt et al., 2013). Comparison of crystallized VP40 from EBOV and SUDV showed the presence of a conserved hydrophobic dimeric interface located between the NTDs of two VP40 molecules, specifically residues 52-65 and 108-117 (Bornholdt et al., 2013). The presence of such a conserved dimeric interface in SUDV has also been confirmed by Clifton et al. (2015). This dimeric conformation is critical for trafficking to the host PM and matrix assembly as well as for viral egress (Adu-Gyamfi et al., 2014; Bornholdt et al., 2013; Jain et al., 2021). Immunofluorescence and transmission electron microscopy analyses have shown that VP40 mutants with a disrupted dimeric interface did not migrate to PM and form filamentous VLP (Bornholdt et al., 2013). Similarly, Hoenen et al. (2010) determined that in the presence of VP40 lacking oligomerization, no infectious VLPs are produced and showed a reduced efficiency in localization to the PM along with impaired binding to the cellular membranes. Interestingly, dimeric VP40 can also further assemble into hexamers (Fig. 1.6B) through hydrophobic CTD interactions; however, performing mutational analyses in this region did not compromise the ability of VP40 to but did show that the lack of CTD-CTD interactions negatively affects matrix assembly and budding, highlighting the importance of the hexameric filament conformation of VP40 in these viral processes (Bornholdt et al., 2013).

The largest oligomeric state that VP40 can acquire is an octameric ring (Fig. 1.6C) mediated by the NTD (Bornholdt et al., 2013; Hoenen et al., 2010; Hoenen et al., 2005; Jain et al., 2021; Timmins et al., 2003). The possibility of octamer formation has been initially associated with the presence and interaction with nucleic acids, since a free NTD existed mostly in the dimeric form (Hoenen et al., 2005; Timmins et al., 2003). The crystal structures of Bornholdt et al. (2013) imply otherwise, as they have obtained an octameric VP40 conformation despite the lack of RNA. However, a subsequent analysis by them revealed the possible role of RNA in displacing the CTD from the NTD and consequently initiating ring structure formation. Bornholdt et al., further identified the existence of an RNA binding interface unique to the octameric ring that binds the RNA (Bornholdt et al., 2013). The role of the octameric state of VP40 is likely not related to virion assembly and is most likely associated with the negative regulation of viral transcription through RNA binding activity (Bornholdt et al., 2013; Hoenen et al., 2010; Hoenen et al., 2005).

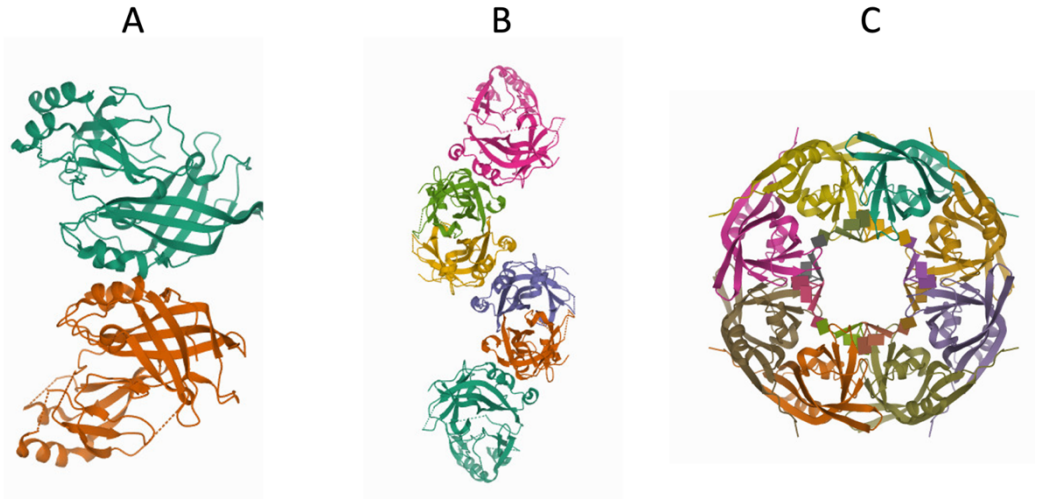


Figure 1.6 VP40 oligomeric states. VP40 is a metamorphic protein that can exist as (A) dimer (PDB ID: 4LDB, (Bornholdt et al., 2013)), (B) filamentous hexamer (PDB ID: 4LDD, (Bornholdt et al., 2013)), and (C) an octameric ring (PDB ID: 7K5D, (Landeras-Bueno et al., 2021)). Each coloured region represents a single VP40 molecule.

1.2.2.4 Glycoprotein

The EBOV *GP* gene encodes three open reading frames that generate different glycoproteins (Fig. 1.7 and 1.8): (i) a soluble glycoprotein (sGP), (ii) a membrane-bound surface glycoprotein ($GP_{1,2}$), and (iii) a small soluble glycoprotein (ssGP). sGP and ssGP are both non-structural proteins while, $GP_{1,2}$ is a structural protein (Zhu et al., 2019). These glycoproteins are a product of transcriptional slippage that occurs at an RNA editing site, composed of seven consecutive uridine (U) residues located upstream of the GP stop codon. During transcription, three scenarios occur: first, the *GP* gene is transcribed as is and an unedited transcript carrying seven consecutive adenosine (7A) is produced and translated to sGP (De La Vega et al., 2015; Sanchez et al., 1996; Viktor E. Volchkov et al., 1995); second, RNA-dependent RNA L polymerase can stutter at the gene-editing site and slip forward resulting in the addition of a non-templated A residue yielding a GP transcript with 8A and producing the $GP_{1,2}$ protein (Sanchez et al., 1996; Viktor E. Volchkov et al., 1995); lastly, RNA polymerase slippage can also give rise to transcripts containing 6 or 9 A which is then translated to ssGP (Mehedi et al., 2011). The transcripts sGP, GP1,2 and ssGP account for about 70%, 25% and 5% of the total transcripts, respectively (Mehedi et al., 2011; Zhu et al., 2019).

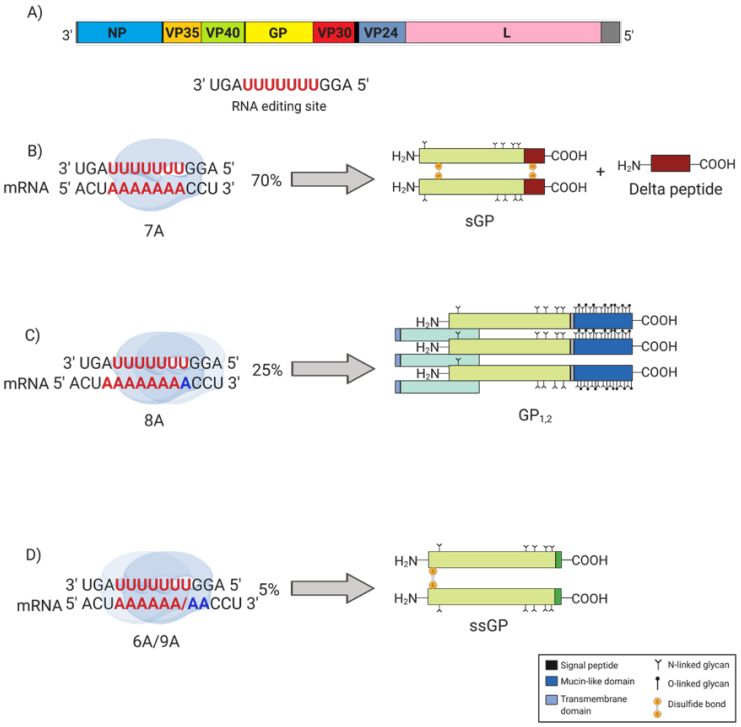


Figure 1.7 Transcriptional gene editing at the glycoprotein (GP) gene. (A) The RNA editing site at the *GP* gene is composed of seven consecutive uridines (7U). As a result of this phenomenon, three transcripts are produced: (B) Approximately 70% of the transcripts generated from the *GP* gene are accounted for by the soluble glycoprotein (sGP) translated from a transcript containing the 7 adenosines (A); (C) 25% by the structural glycoprotein $GP_{1,2}$ resulting from a transcript with 8A and (D) 5% by the small soluble glycoprotein (ssGP) from transcripts with 6A/9A. Figure reproduced from Zhu et al. (2019)

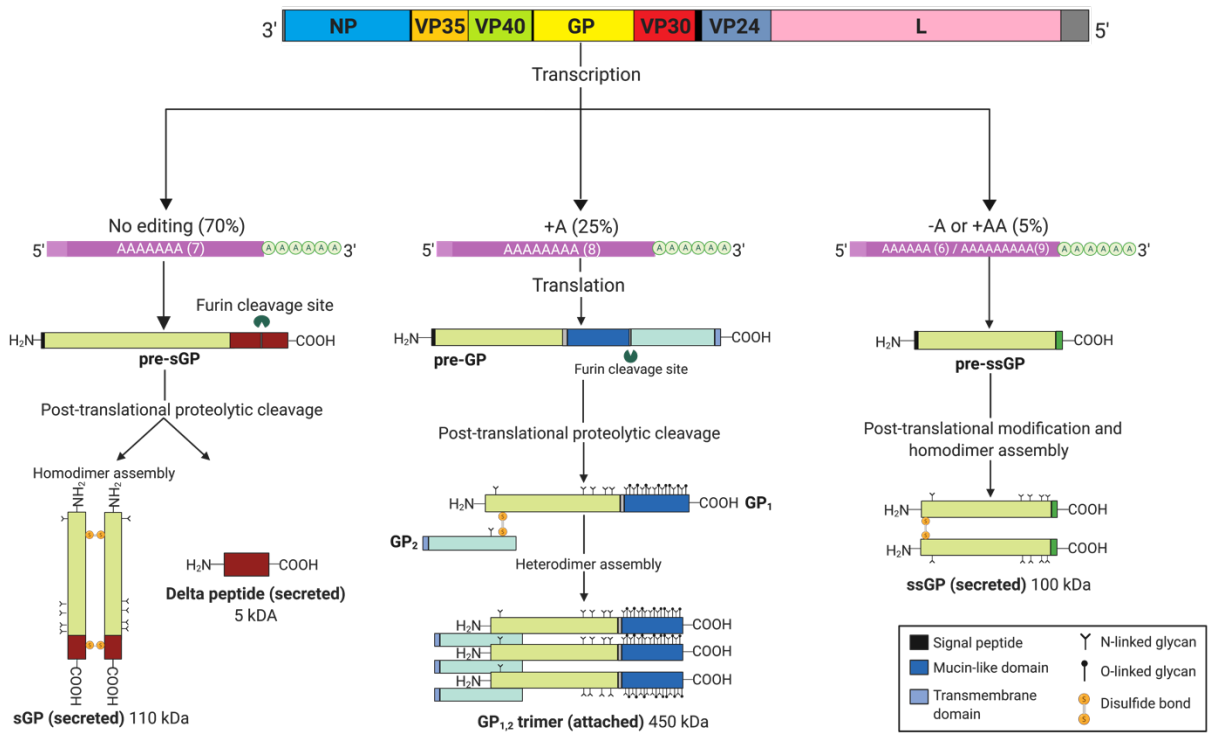


Figure 1.8 Protein products from the ebolavirus *GP* gene. Transcription of the *GP* gene results in three different transcripts which are subsequently translated to pre-sGP, pre-GP, and pre-ssGP. Post translational proteolytic cleavage of the pre-sGP at the carboxy-terminus yields two secreted products, a 5 kDa delta peptide, and a 110 kDa homodimeric sGP linked by disulfide bonds. Pre-GP (GP_0) is also cleaved post-translation and produces GP_1 and GP_2 proteins that dimerizes and forms $GP_{1,2}$, which trimerize to a 450 kDA surface GP. The pre-ssGP does not undergo any cleavage but homodimerizes to a 100 kDa secreted ssGP protein. Both sGP and ssGP proteins experience *N*-glycosylation, while $GP_{1,2}$ is subject to *N*- and *O*-glycosylation. Figure reproduced from Zhu et al. (2019).

1.2.2.4.1 Soluble glycoprotein

The primary product of the *GP* gene is sGP, accounting for most of the transcripts (Jain et al., 2021; Mehedi et al., 2011; Zhu et al., 2019). It is a secreted homo-dimeric glycoprotein that shares 295 N-terminal aa with the surface glycoprotein, $GP_{1,2}$, but has a distinct C-terminal end made up of 65 aa (Pallesen et al., 2016; Volchkova et al., 1998). Site-directed mutagenesis studies by Volchkova et al. (1998) identified residues 53 and 306 of the protein, which both hold Cys, as critical for the homodimerization of sGP monomers bound by disulfide bond formation. Dimerization of sGP monomers yields a ~100 kDa protein with monomers arranged in a parallel orientation (Falzarano et al., 2006). Translated from the unedited version of *GP* mRNA is a premature sGP protein in the endoplasmic reticulum (sGP_{er}) that is ~50 kDa and is sensitive to

endoglycosidase H (endo-H) in the ER (Volchkova et al., 1999). Subsequently, the precursor, $sGP_{er,}$ was converted to pre-sGP (~60 kDa) which now becomes insensitive to endo-H (Volchkova et al., 1999). Pre-sGP is further post-transitionally modified to generate the mature homodimer sGP by proteolysis at a multibasic amino acid motif, at R324, at the carboxy-terminus by furin and signalase (Falzarano et al., 2006; Volchkova et al., 1999). The 291 aa-long sGP carries N-linked carbohydrates attached to various N-glycosylation sites on the protein, including aa 40, 204, 228, 257, and 268 (Falzarano et al., 2006; Jain et al., 2021; Volchkova et al., 1999). A C-mannosylation site was discovered at aa 288.

The structural role of sGP remains elusive, but it has been proposed to have the potential to substitute for the GP_1 counterpart of the $GP_{1,2}$ protein, as it shares the same N-terminus Cys53 with GP_1 capable of forming a disulfide bond with Cys609 in GP_2 (Sanchez et al., 1996; Zhu et al., 2019). Iwasa et al. (2011) tested such a hypothesis and determined that EBOV sGP can indeed form a complex with GP_2, forming an sGP-GP_2, with the ability to infect being retained; albeit at lower rates compared to the wild-type strain. Several non-structural roles have also been linked to sGP. These roles include increasing viral dissemination by inhibiting pro-inflammatory cytokine production and halting the chemotaxis ability of activated macrophages (Bradley et al., 2018; Jain et al., 2021; Zhu et al., 2019), inactivation (Kindzelskii et al., 2000), promoting viral replication by restoring endothelial cell barrier function (Wahl-Jensen et al., 2005), and acting as a probable antigenic decoy preventing recognition of EBOV $GP_{1,2}$ by $GP_{1,2}$ antibodies (Mohan et al., 2012).

The cleaved product from sGP, called the delta (Δ) peptide, is 41 amino acids long and is made up of cationic and aromatic residues (Jain et al., 2021). It is also non-structural and is secreted by cells, albeit at a slower speed and lower concentration (Volchkova et al., 1999). In contrast to sGP, Δ-peptide undergoes additional modifications through O-glycosylation and sialylation (Volchkova et al., 1999). It has been demonstrated that the Δ-peptide by-product of sGP can hinder viral entry from filovirus-permissive cells and may be a contributing factor to limiting superinfection in an infected host (Jain et al., 2021; Radoshitzky et al., 2011; Zhu et al., 2019). The presence of a Lys-rich amphipathic peptide motif in Δ-peptide and its ability to permeabilize cultured mammalian cells suggest that the short peptide may act as a viroporin (Gallaher & Garry, 2015; He et al., 2017), although its relevance to EBOV pathogenesis remains to be elucidated.

1.2.2.4.2 Membrane-bound glycoprotein

$GP_{1,2}$ is a type I transmembrane protein expressed on the viral surface (Davey et al., 2017). In parallel with sGP, mature $GP_{1,2}$ originates from the ER in a pre-GP_{er} form that is eventually replaced by a Golgi-specific pre-GP/GP_0 precursor (Volchkov et al., 1998). The GP_0 form is then post-translationally cleaved by furin-like proteases at a multibasic site R501, consequently generating GP_1 and GP_2 that forms heterodimers with each via disulfide bonds between Cys53 and Cys609 of GP_1 and GP_2, respectively (Cook & Lee, 2013; Davey et al., 2017; Lee & Saphire, 2009; Volchkov et al., 1998). The heterodimeric protein then trimerizes with other $GP_{1,2}$ subunits non covalently on the virus surface and forms the mature and functional, chalice-like shape surface GP (~450 kDa) (Cook & Lee, 2013; Lee et al., 2008). From the attached trimeric GP, a secreted GP made up of GP_1 associated with a truncated GP_2 is shed by cleavage of tumour necrosis factor (TNF)-α-converting enzyme or TACE at a site near the transmembrane anchor (Lee & Saphire, 2009). Not much is known about the function of shed GP, but it is believed to compete for anti-$GP_{1,2}$ antibodies (Cook & Lee, 2013) and help reduce the cytotoxicity of membrane-bound GP. It has also been demonstrated that shed GP can cause immune activation with a downstream effect of increasing cytokine production and increased vascular permeability (Jain et al., 2021).

The GP_1 subunit recognizes a variety of receptors expressed on the target host cell surface that allow viral attachment (Davey et al., 2017; Lee & Saphire, 2009). It consists of four domains: (i) a base, (ii) a receptor-binding domain (RBD) (aa 54-201), (iii) a glycan cap, and (iv) mucin-like domain (MLD) (aa 313-464) (Davey et al., 2017; Kuhn et al., 2006). The base domain forms a semicircular hydrophobic surface that interacts with the GP_2 subunit to provide structural support to the other domains and prevents the GP_2 subunit from going into a fusogenic stage (Jain et al., 2021; Lee & Saphire, 2009). The RBD carries residues that contact intracellular host receptors, while the heavy glycosylation on the MLD and the glycan cap likely protect the GP from antibody neutralization (Lee & Saphire, 2009; Lennemann et al., 2014). Glycosylation in MLD does not only confer virus protection against antibody recognition. Past studies have shown that excessive glycosylation of this domain can prevent the recognition of cell surface proteins such as major histocompatibility complex (MHC) molecules due to steric shielding, thus hiding the virus from T cell recognition (Francica et al., 2010; Reynard et al., 2009).

However, the GP_2 subunit is essential for membrane fusion with the host membrane with an overall structure that resembles other viral membrane fusion proteins such as those in HIV and influenza (Davey et al., 2017; Lee & Saphire, 2009; Malashkevich et al., 1999). It is composed of a hydrophobic internal fusion loop (aa 511-556), two heptad repeats (HR1, HR2) linked by a 25 aa linker that carries a disulfide bond motif, a membrane-proximal external region, and a transmembrane I domain (aa 650-672) (Davey et al., 2017; Lee et al., 2008; Sanchez et al., 1998). The internal fusion loop of GP_2 displays a helical hydrophobic peptide whose side chains are packed into the hydrophobic base domain of the GP_1 subunit (Lee et al., 2008). This helical peptide is inserted into the target cell membrane and initiates fusion with the host membrane once the release of clamped GP_1 is triggered (Lee & Saphire, 2009). The TM domain of GP_2 anchors the trimeric GP to the host membrane (Ning et al., 2017). Adjacent to the TM domain is a sequence of a heptad repeat (aa 541-593) hypothesized to play a role in the formation of intermolecular coiled coils during trimerization of $GP_{1,2}$ subunits (Sanchez et al., 1998). The GP_2 subunit is also capable of altering the immune response due to its anti-tetherin activity (Jain et al., 2021). Tetherin is an activator of the nuclear factor kappa B (NF-κB) pathway, a signalling pathway that once activated induces the expression of antiviral genes such as pro-inflammatory genes (i.e. cytokines and chemokines) (Liu et al., 2017). GP_2 has been shown to bind tetherin that consequently blocks NF-κB signalling and promotes VP40-mediated viral budding (Jain et al., 2021; Kühl et al., 2011).

1.2.2.4.3 Small soluble glycoprotein

ssGP also shares the first 295 N-terminal residues with sGP and $GP_{1,2}$ (Mehedi et al., 2011). Like sGP, this protein is also secreted by infected cells and can form homodimers (~100 kDa) linked by a disulfide bond in Cys53 and is N-glycosylated (Mehedi et al., 2011). However, due to the +2-frameshift resulting from RNA transcription editing, a truncated ORF is generated. The truncated ORF results in a carboxy-terminus distinct from the sGP and GP $GP_{1,2}$ (Mehedi et al., 2011). It may be implied that ssGP can have a function like sGP because of the shared N-terminal sequences. However, unlike sGP, ssGP did not have the ability to restore the function of the endothelial barrier (Mehedi et al., 2011). More functional studies are warranted to determine if this protein shares similar roles with other products of the *GP* gene.

1.2.2.5 Viral protein 30

VP30 is a phosphoprotein and an essential component of the polymerase complex that acts as a transcription activator for EBOV (Cantoni & Rossman, 2018; Jain et al., 2021; Modrof et al., 2002). It has an estimated molecular weight of 30 kDa (Xu et al., 2017) composed of dimers, formed by residues 142-264 in the CTD, which assemble into hexamers facilitated by an N-terminal oligomerization site at aa 94-112 (Hartlieb et al., 2007; Jain et al., 2021). This homo-oligomerization process of VP30 has been shown to be vital to the role of the protein as a transcription activator (Biedenkopf, Schlereth, et al., 2016; Hartlieb et al., 2003; Hartlieb et al., 2007). Hartlieb et al. (2003) identified a cluster of four leucine residues to be critical, as a single mutation of one of these amino acids resulted in the abortion of VP30 oligomerization and consequently EBOV transcription. The RNA-binding site in VP30 is mapped at the N-terminus, residues 26-40, a region that is disordered, non-hydrophobic, and rich in arginine (John et al., 2007). This RNA binding ability is also critical to activating viral transcription. Multiple factors can affect it: (i) homo-oligomerization ability of VP30, (ii) disruption of the zinc (Zn) finger domain, (iii) VP30 phosphorylation state, and (iv) alteration of the RNA binding region (Biedenkopf, Schlereth, et al., 2016).

According to Schlereth et al. (2016), an ideal RNA substrate for VP30 is about 40 nucleotides long with a mixed base composition linked to a stem-loop structure, which was indicative of a quite relaxed binding specificity for VP30. Zn binding also plays a significant role in VP30 RNA binding, which has a downstream effect on the protein's ability to activate transcription (Cantoni & Rossman, 2018; Modrof et al., 2003). An unconventional Zn-binding Cys_3-His motif at aa 68-95 has been shown to bind Zn ions in a 1:1 ratio and if conserved cysteine and histidine were substituted, a complete loss of Zn-binding ability and transactivating function of VP30 was observed (Biedenkopf, Schlereth, et al., 2016; Modrof et al., 2003). As a phosphoprotein, a protein that is post-translationally modified by phosphorylation, its phosphorylation state is essential for VP30 to bind viral RNA (Biedenkopf, Schlereth, et al., 2016). The protein phosphorylation sites occur at the serine (Ser) and threonine (Thr) residues (Ilinykh et al., 2014; Jain et al., 2021; Modrof et al., 2002). In the case of VP30, its phosphorylated version of the protein negatively regulates viral transcription and increases viral replication due to a weak interaction with other components of the transcription complex, including VP35 and the viral RNA (Biedenkopf et al., 2013;

Biedenkopf, Lier, et al., 2016; Biedenkopf, Schlereth, et al., 2016). After dephosphorylation, viral transcription is up-regulated as the protein becomes better at association with the transcription complex together with NP and VP35 (Biedenkopf et al., 2013; Biedenkopf, Lier, et al., 2016; Cantoni & Rossman, 2018).

1.2.2.6 Viral protein 24

VP24 is also a multifunctional protein, making up approximately 7.5% of the total EBOV proteins (Jain et al., 2021). One of its functions is a structural element of the nucleocapsid along with NP and VP35, in which VP24 has been shown to have a direct interaction with NP and VP35 (Banadyga et al., 2017; Chen et al., 2019; Huang et al., 2002; Watanabe et al., 2006). NP can self-assemble and form a loose-coiled helical structure in the presence of RNA; but in the presence of VP24 and VP35, the former structure is converted to a nucleocapsid structure almost indistinguishable from those seen in EBOV infected cells (Huang et al., 2002; Watanabe et al., 2006). Algorithmic prediction recognized residues 169-180 as important for NP interaction, but functional studies demonstrated that aa 169-176 were detrimental, with aa V170 and N171 as the two aa responsible for NP interaction (Banadyga et al., 2017).

Like VP35, VP24 is capable of interfering with the host immune IFN response using numerous mechanisms distinct from VP35. VP24 can antagonize the host immune response through binding to nuclear transporters, and karyopherin α proteins such as α1, α5, and α6, consequently preventing the nuclear accumulation of phosphorylated signal transducer and transcription activator protein (STAT) (Mateo et al., 2010; St Patrick Reid et al., 2006; Reid et al., 2007; Xu et al., 2014). Activation of IFN signaling in the presence of a virus prompts nuclear transport of tyrosine-phosphorylated STAT1 protein via the karyopherin α proteins, also known as importin α, resulting in the activation and transcription of antiviral genes (St Patrick Reid et al., 2006; Xu et al., 2014). Structural mapping of VP24 and karyopherin α5 interactions divulge aa 308-509 of the C-terminal end of karyopherin α5 as the minimum requirement for multiresidue contacts with VP24 (Xu et al., 2014). A mutagenesis study by Mateo et al., 2010 has shown that VP24 aa 26-50 and 142-146 is critical for binding with karyopherin α1. VP24 also shows competitive binding to the heterogeneous nuclear ribonuclear protein complex C1/C2, a protein complex also known to bind the same C-terminal region (residues 424-457) of karyopherin α1

(Shabman et al., 2011). Furthermore, VP24 has been shown to directly bind the STAT1 protein, regardless of its phosphorylation state, with residues 96-98 and 106-121 as plausible sites of protein-protein interaction (Zhang et al., 2012). Another mechanistic approach that VP24 uses to aid in immune evasion is to block the phosphorylation of mitogen-activated protein kinase (MAPK) (Halfmann et al., 2011; Ilinykh et al., 2015). P38 kinase is a key element of the MAPK signalling pathway. Once the pathway is activated by binding of type I IFN to receptors, p38 becomes phosphorylated and, as a result, other participating IFN transcription factors are phosphorylated and activated (Halfmann et al., 2011).

1.2.2.7 L polymerase

The EBOV L protein is 2,212 aa in length with an estimated molecular mass of 253 kDa and is the largest multisubunit protein encoded by ebolaviruses (Ayub & Waheed, 2016). Unable to function independently, the L protein is part of the RNP complex together with NP, VP35 and VP30, acting as the catalytic subunit of the said complex (Ayub & Waheed, 2016). This RNA-dependent RNA polymerase complex is vital for genomic viral RNA replication and transcription (Cantoni & Rossman, 2018). In addition to the role of the protein in viral transcription and replication, L polymerase has been shown to have the ability to edit mRNA (V. E. Volchkov et al., 1995). Unfortunately, structural and functional analyses on L polymerase are almost nonexistent and are based on comparisons made with well-studied non-segmented negative-sense RNA viruses from the family *Mononegavirales* like vesicular stomatitis virus (VSV) (Ayub & Waheed, 2016; Trunschke et al., 2013), which are all equipped with a catalytic subunit, L (Liang et al., 2015). Alignment of a homologous L sequence from EBOV and VSV revealed the same overall arrangement of the various L protein domains (Liang et al., 2015). From this, five domains make L protein: (i) RNA-dependent RNA polymerase (RdRp) domain equipped with transcription, replication and polyadenylation activity, (ii) capping domain encoding polyribonucleotidyltransferase (PRNTase) activity, (iii) connector domain, (iv) amethyltransferase domain and (v) C-terminal domain (Jain et al., 2021). A homo-oligomerization domain has also been identified in the first 450 residues of the L protein (Trunschke et al., 2013). Within the same region, Trunschke et al. (2013) identified an L-VP35 interaction region at aa 1-380; however, the two domains appear not to not interfere with each other.

1.2.3 Viral replication cycle in host cells

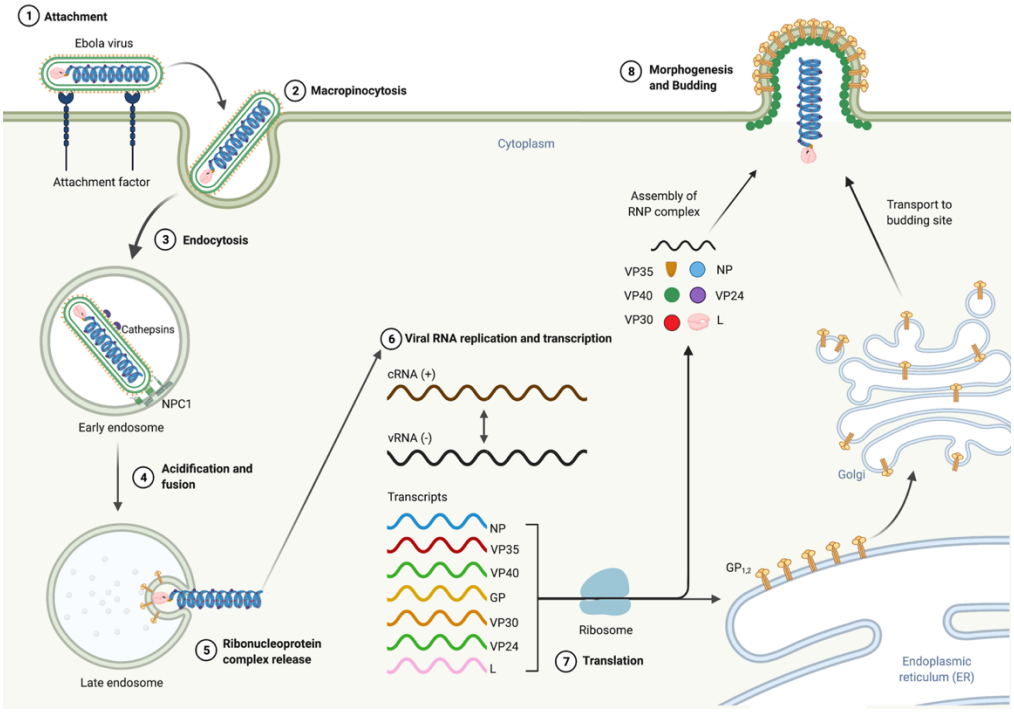


Figure 1.9 Ebola virus replication cycle. (1) Infection begins with the virus binding to attachment factors expressed on the surface of the host cell. (2) The virus is then engulfed by the cell by the process of micropinocytosis and (3) endocytosed by the cell. (4) In the endosome, the viral surface glycoprotein $GP_{1,2}$ is cleaved by proteases such as cathepsins and binds to the intracellular receptor NPC1. (4) As the early endosome matures into a late endosome, the pH inside the vesicle decreases and the viral membrane fuses with the lipid bilayer and (5) eventually releases the ribonucleoprotein (RNP) complex composed of the viral RNA, NP, VP35, VP30 and L polymerase to the cytoplasm. (6) The RNP complex becomes relaxed and viral RNA is transcribed from the negative sensed viral RNA (vRNA). (7) Transcripts are further translated into proteins by the host's ribosome. (8) Formation of the RNA complex and viral budding then occurs at the plasma membrane. Created with BioRender.com

1.2.3.1 Viral attachment and entry

The first step in ebolaviruses infection is the attachment of the virion to a receptor of the target host cell. In the case of EBOV, attachment and entry are mediated by the surface GP. The GP protein can make contacts with a wide variety of host-cell attachment factors including lectins (e.g., C-lectins, DC-SIGN, and *L-SIGN*), β1-integrins, glycosaminoglycans, folate receptor α, TYRO3 receptor tyrosine kinase family members, T-cell immunoglobulin, and mucin domain-1 (TIM-1). Three mechanisms of internalization have been implicated to EBOV: (i)

macropinocytosis (Afelt et al., 2018; Aleksandrowicz et al., 2011; Quinn et al., 2009; Saeed et al., 2010), (ii) Clathrin-mediated endocytosis (Aleksandrowicz et al., 2011; Bhattacharyya et al., 2011; Bhattacharyya et al., 2010), and (iii) caveolin-mediated endocytosis (Empig & Goldsmith, 2002). The primary uptake of the virus into the endosomal compartment is believed to be via macropinocytosis due its large size, although the ability of EBOV to use alternative uptake mechanisms may be dependent on the cell type or virion size (Aleksandrowicz et al., 2011). Once endocytosed, the endosome matures into a late endosome with a low pH inside the vesicle, and this acidification results in the cleavage of the GP_1 subunit by proteases like cathepsins B and L (Brecher et al., 2012; Chandran et al., 2005; Schornberg et al., 2006). Depending on the cell type, variation in host proteases that cleave the protein may be observed (Martinez et al., 2010). The cleaved subunit can now interact with intracellular Niemann-Pick C1 (NPC1) receptor primarily expressed on late endosomes, a crucial step for the virus infectivity (Carette et al., 2011; Côté et al., 2011). Crystal structures reveal that NPC1 uses two extended loops from its helical core to interact with the hydrophobic cavity on the GP head (Yu et al., 2017). It is currently unknown what factors directly trigger fusion conformation change in GP, but proteolytic cleavage of the GP1, binding to NPC1 and TIM-1 protein that links the protein to the host cell membrane are believed to influence conformational changes to the GP_2 subunit of the protein. Fusion occurs such that the internal fusion loop and transmembrane domain are placed closed together resulting in the formation of a fusion pore that facilitates the merging of membranes and release of the viral RNA into the cytoplasm (Kuroda et al., 2015; Lee & Saphire, 2009; Yuan et al., 2015).

1.2.3.2 Viral genome replication and transcription

The nucleocapsid is released to the host cell cytoplasm upon fusion with the host PM. The nucleocapsid is composed of the viral RNA bound to various viral proteins, including, NP, VP35, VP30, VP24 and L (Hume & Mühlberger, 2019). In the cytoplasm, the viral RNA only exists in the RNP form and both transcription and replication processes are RNP-mediated fields (Yu et al., 2017). In both its antigenomic and genomic orientation, the viral RNA is sequestered to an RNA-binding cleft formed by the folding of NP's amino-terminal end (aa 1-450) into two lobes (Basler et al., 2019; Wan et al., 2017). The cleft binds RNA non-specifically through hydrogen bond interactions and electrostatic interactions with the negatively charged backbone of the RNA (Sugita et al., 2018). Occurring concomitantly with RNA encapsidation is viral genome

replication, which requires the presence of VP35 and L polymerase (Basler et al., 2019; Mühlberger, 2007). As part of the RNP complex, VP35 acts a non-enzymatic cofactor for the L polymerase by bridging the NP-RNA complex with the L polymerase (Basler et al., 2019; Hume & Mühlberger, 2019; Tchesnokov et al., 2018). VP35's interaction with NP is also deemed critical for viral genome replication as mutagenesis on a basic patch of the VP35 CTD that interacts with NP resulted in the loss of viral RNA synthesis (Prins et al., 2010). Once L becomes a part of the nucleocapsid, along with VP35, NP, and the viral RNA, genome replication can commence. Found at the 3' end of the filovirus genomes and anti-genomes are replication promoters for plus for sense and anti-sense RNA synthesis (Calain et al., 1999; Mühlberger, 2007). Therefore, from the anti-sense genomic RNA (3'-5'), the antigenomic RNA in the sense orientation (5'-3') can be synthesized, which is subsequently used as a template to produce genomic RNA again (Mühlberger, 2007). RNA synthesis is mediated by the catalytic domain, with magnesium ions acting as cofactors (Hume & Mühlberger, 2019; Tchesnokov et al., 2018).

From the RNP complex, the non-segmented negative-sense RNA genome is also transcribed by the RNA-dependent RNA polymerase L, which also has a capping domain and domains capable of RNA modification activities like guanyltransferase and methyltransferase activities (Basler et al., 2019; Hume & Mühlberger, 2019). Transcription is believed to begin at a promoter site located within the leader sequence through binding of the L and VP35 proteins, but the presence of a secondary RNA loop structure formed by the first 23 nucleotides of the pre-mRNA halts the movement of RNA polymerase along the template (Mühlberger, 2007). Therefore, in addition to the L, VP35, and NP proteins, transcription of the viral genome requires a transcription activator – VP30 (Biedenkopf, Schlereth, et al., 2016; Weik et al., 2002). VP30 abrogates this secondary RNA structure by being able to bind the RNA and consequently stabilize the complex formed by the viral RNA, RNA polymerase, and VP35 (Biedenkopf, Schlereth, et al., 2016). In the absence of the secondary RNA structure, RNA polymerase can now move along transcribe the RNA template, independent of the presence of VP30, by recognition of start and stop signals that flank each gene (Hume & Mühlberger, 2019; Mühlberger, 2007). Individual genes are proposed to be sequentially transcribed in the 3′ to 5′ order since the polymerase is believed to only enter at a single binding site at the 3' end of the genome (Mühlberger, 2007; Whelan et al., 2004).

Seven monocistronic mRNA transcripts, which are capped and polyadenylated by the RNA polymerase, encoding for the ebolavirus viral proteins are produced (Mühlberger, 2007).

Transcription of the viral genes occurs in a gradient for which genes located closer to the 3' end of the genome generate higher amounts of transcripts compared to genes further at the 5' end (Watanabe et al., 2006). This phenomenon is not well understood but one probable reason is that the polymerase complex ceases to start transcription again at the start signals located downstream (Mühlberger, 2007). It is postulated that the accumulation of proteins in the cytoplasm is thought to influence switching from transcription mode to replication of viral RNA. Additionally, it has also been demonstrated that VP30 has the means to regulate transcription and replication depending on its phosphorylation state; however, influential factors affecting its phosphorylated state are unknown (Ilinykh et al., 2014).

1.2.3.3 Viral assembly and budding

Viral assembly and egress of ebolavirus virions are largely regulated by VP40 that assembles on the inner leaflet of the host PM (Adu-Gyamfi et al., 2014). Even in the absence of other ebolavirus proteins, filamentous virus-like particles (VLPs) are still generated when VP40 is expressed alone in human cells (Adu-Gyamfi et al., 2014; Noda et al., 2007). Proper trafficking of VP40 to the host PM is necessary before VLP production can occur (Yamayoshi et al., 2008). This trafficking of VP40 is governed by the COPII transport system (Yamayoshi et al., 2008). Along with VP40, the RNP complex and GP are also transported to the host cell membrane (Jain et al., 2021). RNP transport is reported to be dependent on actin as depolymerized actin filaments and inhibition of actin nucleating complex arrested transport of the nucleocapsid (Schudt et al., 2015). Meanwhile, GP employs the secretory pathway for its trafficking to the cell membrane. Through this pathway, GP is heavily N- and O-glycosylated and cleaved into its GP_1 and GP_2 subunits by furin (Feldmann et al., 1994; Nehls et al., 2019). The late budding domain of VP40 then interacts with a transport machinery called endosomal sorting complex required for transport (ESCRT), releasing new virions from the infected cell.

1.3 EBOLA DISEASE

1.3.1 Reservoir and transmission

A reservoir can be defined as populations of organisms or environments in which a pathogen can reproduce and be permanently maintained, and from which the pathogen can be transmitted to the susceptible population (Haydon et al., 2002). The search for an ebolavirus reservoir is still in

progress and continues to be unsuccessful as no infectious virus has ever been isolated in any of the proposed host reservoirs for ebolaviruses (Breman et al., 1999; Leendertz et al., 2016; Leirs et al., 1999). Ever since its first detection in 1976 in Africa, several animal species have been proposed as potential host reservoirs for ebolaviruses: squirrels, mice, dormice, and shrew (Baseler et al., 2017), but none has been "officially accepted" because no infectious virus has ever been isolated in these animals and because of the lack of ebolavirus specific antibodies in the sera of collect specimens (Breman et al., 1999; Leirs et al., 1999). Bats have been identified as the most probable host reservoir for ebolaviruses. Bats generally house a wide variety of viruses that have the potential to be zoonotic, including severe acute respiratory syndrome coronaviruses (SARS-CoV), Hendra, and Nipah viruses (Han et al., 2015). However, a huge contributing factor to this hypothesis is detecting and isolating a related filovirus, MARV, in African fruit bats (*Rousetttus aegypticus*) (Jones et al., 2015). Moreover, EBOV RNA and EBOV–specific antibodies have been detected in different asymptomatic bat species, which further implicates bats as reservoirs (Leendertz et al., 2016; Leroy et al., 2005). Bats are often found residing in caves and trees, but the expansion of infrastructures near or in virgin forest areas make houses, bridges, and barns attractive roost dwelling for bats (Irving et al., 2021). Interaction between bats, wild animals and humans becomes inevitable and can create potential spillover or jump of viruses from the reservoir to the human population. Probable modes of transmission can occur between the reservoir and humans by means of direct contact with bats and/or contact with an infected intermediate host that humans closely interact with, like pigs and nonhuman primates (NHPs) that gets in contact with the bodily fluids or feces of the reservoir (Han et al., 2015). Human-to-human transmission can also happen following exposure of nonintact skin or mucosal membranes to the blood, secretions, or bodily fluids of another infected individual (Baseler et al., 2017; Malvy et al., 2019).

1.3.2 Pathogenesis

After entering through the endothelial barrier, the virus initially infects the macrophages, monocytes, and dendritic cells (Baseler et al., 2017; Muñoz-Fontela & McElroy, 2017; Yamaoka et al., 2017). These infected cells then migrate to the lymph nodes where they now have the capability to disseminate EBOV to the other parts of the body (Baseler et al., 2017; Geisbert et al., 2003). Indeed, the virus is detectable in the blood of patients after 1-3 days of symptom onset (Lanini et al., 2015). Disseminated virus further affects other cells such as Kupffer cells of the

liver, hepatocytes, fibroblasts, and adrenal gland tissue (Furuyama & Marzi, 2019). Proposed key factors to virus dissemination include the association of EBOV to migratory cells like tissue-resident DCs and/or inflammatory DCs, involvement of peripheral monocytes, and ability of the virus to dysregulate the innate and adaptive immune response which results in upregulated proinflammatory response and inhibited antiviral response (Muñoz-Fontela & McElroy, 2017).

1.3.3 Hallmarks of the disease

Ebola disease (EBOD) is a severe, fatal acute viral hemorrhagic fever disease. Incubation time after exposure is variable ranging from 2-21 days. Early into the infection, approximately 1–3 days after onset of symptoms, an infected individual typically presents non-specific febrile illness with symptoms such as fever, malaise, myalgia, and loss of appetite (Jacob et al., 2020; Malvy et al., 2019). Unfortunately, with EBOD, the severity of the disease can progress quickly within 7-14 days after symptom. Patients begin to show gastrointestinal symptoms like nausea, abdominal pain, vomiting, diarrhea, and dehydration (Jacob et al., 2020). Increasing severity of clinical manifestations over time can be linked to increases in viral load as most patients show detectable viremia 6–10 days post-exposure (Jacob et al., 2020). During the complicated or late stage, gastrointestinal symptoms persist along with hemorrhagic manifestations (e.g. coughing up blood, vomiting blood, and bloody stool), due to coagulopathy, vascular leakage, and dysregulated inflammation – all of which may lead to multiple organ failures or even death (Jacob et al., 2020; Malvy et al., 2019; Yamaoka et al., 2017). Changes in biochemistries are also observed along with these clinical features. Laboratory findings in EBOD include lymphopenia, leukopenia, thrombocytopenia, and elevated enzymes (e.g., aspartate and alanine aminotransferases, creatine phosphokinase, and amylase).

1.3.4. Vaccines and therapeutics

The development of vaccines for ebolaviruses has heightened since the West African Ebola outbreak in 2013. In 2019, a VSV-based vaccine received FDA approval for the prevention of the Ebola virus disease (Choi et al., 2021). The recombinant, live, attenuated VSV-ZEBOV vectored vaccine, known as ERVEBO, is a replication-competent vaccine developed at the Public Health Agency of Canada (PHAC). The vaccine is comprised of the VSV vector in which the G protein has been replaced with the EBOV glycoprotein GP (Choi et al., 2021; Malvy et al., 2019).

However, this vaccine is only approved for use as a preventative measure against EBOV in people aged 18 or older that are in high-risk exposure to the virus, such as health care responders during an outbreak and laboratorians handling the virus in high containment facilities (Choi et al., 2021). The vaccine is administered as a single dose. The search for vaccines against other ebolaviruses, including SUDV and BDBV, is still ongoing. In 2020, the U.S Food and Drug Administration (FDA) approved Inmazeb (Centers for Disease Control and Prevention, 2020a) and Ebanga (Centers for Disease Control and Prevention, 2020b) as a treatment for EBOV infection in adult and pediatric patients. Ebanga is a human monoclonal antibody (Ansuvimab-zykl) designed to target the EBOV GP protein which consequently prevents binding of virus GP to host cell receptors. Inmazeb, on the other hand, is an antibody cocktail comprised of three different monoclonal antibodies: atoltivimab, maftivimab, and odesivimab-egn (Centers for Disease Control and Prevention, 2020a). Like Ebanga, the cocktail was designed to target EBOV GP. Unfortunately, like the vaccine, these antibody therapeutics are limited for use with EBOV only as efficacies on other ebolaviruses have not been established.

1.4 EBOLAVIRUSES ANIMAL MODELS

Modelling of human anatomy and physiology in animals has long existed since the 6th century BCE and remains to exist until this day (Ericsson et al., 2013). Due to the similarities that animals share with human anatomy and physiology, especially mammals, they have been used to answer biological research questions from understanding human disease mechanisms to evaluating novel disease countermeasures like vaccines and antibiotics (Barré-Sinoussi & Montagutelli, 2015; Robinson et al., 2019). For filoviruses, animal models such as NHPs, ferrets, and small rodents like mice, hamsters, and guinea pigs have been developed to aid in studying filovirus pathogenesis and evaluation of various vaccines and therapeutics for the treatment of disease. These animal models can recapitulate several of the clinical hallmarks of filovirus disease in humans (Siragam et al., 2018). Described below are the different animal models that exist for ebolaviruses.

1.4.1 Nonhuman primate

NHPs are currently the gold standard for understanding the pathogenesis of ebolaviruses, the disease, and countermeasure evaluations, as they can be infected with wild-type virus strains

and resemble clinical features of the disease as seen in humans (Bente et al., 2009; Cross et al., 2018). Over decades, primates of different species have been used, including African green monkeys (*Chlorocebus sp*.), baboons (*Papio hamadryas*), cynomolgus macaques (*Macaca fascicularis*), common marmosets (*Callithrix jacchus*), and rhesus monkeys (*Macaca mulatta*) (Bennett et al., 2017). Among these, cynomolgus and rhesus monkeys have been the popular choice for screening vaccine and therapeutic candidates, respectively (Bennett et al., 2017; Bente et al., 2009; Niemuth et al., 2021). Cynomolgus monkeys are considered an exemplary model for vaccine screening as disease progression is more consistent (Goldstein et al., 2018) and condensed temporal disease progression (Bennett et al., 2017). Meanwhile, rhesus macaques generally have slower disease progression, thus making them ideal for therapeutics studies (Bennett et al., 2017). Due to their common use in the filovirus field, points highlighted in this section pertain to previous studies performed using these two species.

In an EBOV infection, macrophages and DCs are primarily targeted by the virus and become sites of virus replication (Geisbert et al., 2003), particularly in Kupffer cells in the liver and dendritic cells in the lymph nodes. The onset of symptoms typically begins with fever at 2-4 days post-infection (DPI), and a cutaneous rash between 4-7 dpi (Bennett et al., 2017; Bente et al., 2009). Time of death varies between the two models, with cynomolgus monkeys succumbing to disease 6-7 DPI (Geisbert et al., 2003), and rhesus macaques at 5-9 days post-exposure (Bennett et al., 2017). Over the course of the disease, animals show weight loss mounting up to 10% of their body weight (Bente et al., 2009) and exhibit high viremia at 2-3 DPI (Bente et al., 2009; Siragam et al., 2018). As the condition of the animal worsens, their body begins to demonstrate signs of a systemic dysregulated innate and adaptive immune response. They start to show changes in biological chemistries like thrombocytopenia, neutrophilia, lymphopenia, granulocytosis, cytokine imbalance, decreased protein C coagulation inhibitor activity, and decreased fibrinogen levels (Bente et al., 2009; Nakayama & Saijo, 2013; Siragam et al., 2018). These abnormal biochemistries were indicative of a disseminated intravascular coagulopathy, which leads to multi-organ failures and eventually death.

Coined as the gold standard model, using NHPs has its disadvantages too. For instance, the use of this animal is subject to ethical concerns. They are also large animals, and this translates to difficulties in housing and handling. NHPs are also expensive and associated with the high price is the use of a limited number of animals during experiments can also limit result interpretations.

1.4.2 Ferret

Ferrets (*Mustela putorius furo*) are small carnivores belonging to the weasel family *Mustelidae* and have been existing as domesticated animals for more than 2000 years when they were first primarily used for rabbit hunting (Davidson et al., 1999; Mayer et al., 2015). Their use as biological research tools began to gain attention in the 1900s when they were first used as a model to study influenza virus pathogenesis (Pyle, 1940; Smith et al., 1933). To date, ferrets serve as the most ideal model to perform pathogenesis studies on influenza viruses and evaluate countermeasures due to their susceptibility to wild-type strains of influenza, transmission efficiency, but most importantly, their ability to mimic the clinical hallmarks of a human infection (Oh & Hurt, 2016). Its use as an animal model also extends to other viruses such as the human respiratory syncytial virus and SARS-CoV-2, the etiological agent of the ongoing coronavirus disease 2019 pandemic (Kim et al., 2020; Shi et al., 2020), morbilliviruses (Pillet et al., 2009), and paramyxoviruses.

The use of ferrets for ebolavirus research is also growing. The first models were established by research groups at the University of Texas Medical Branch (ç) and PHAC (Schiffman et al., 2021). In 2016, UTMB developed models for pathogenic ebolaviruses including EBOV (variant Kikwit), SUDV (variant Gulu), and BDBV (variant Uganda 2007) (Cross et al., 2016). PHAC's models were developed using EBOV (variant Makona-C07) and BDBV (variant Uganda 2007) (Kozak et al., 2016), SUDV (variant Boniface) (Kroeker et al., 2017), and RESTV (Yan et al., 2019). Models from these groups highlight one of the advantages the ferret model presents – the use of non-adapted viruses. Without the need for adaptation, all ebolaviruses examined were 100% lethal in ferrets, although the time of death varied between species. EBOV-infected animals succumbed to the disease at 5-DPI, 7-8 DPI for SUDV, 8-9 DPI for BDBV, and 9-11 DPI for RESTV (Schiffman et al., 2021). Differences in end timepoints may be attributed to differences in peak titres observed between the three ebolaviruses where EBOV showed the highest viral titre in infected liver and spleen organs followed by SUDV and BDBV (Cross et al., 2016). Animals from both models also recapitulated many of the clinical hallmarks seen in NHPs and humans – fever, diarrhea, dehydration, systemic viral loads, histological lesions, decreased white blood cell counts and thrombocytopenia, granulocytosis, and coagulopathy (Cross et al., 2018; Cross et al., 2016; Kozak et al., 2016; Kroeker et al., 2017).

The domestic ferret animal model remains valuable for filovirus research. The model is uniformly lethal following infection with several different wild-type ebolaviruses, and it mimics clinical features of EBOD just like the NHP models. However, ferrets are smaller animals, making them easier to house and handle in containment facilities and more cost-effective than NHPs (Schiffman et al., 2021). One might argue, however, that the ferret model is disadvantageous compared to existing small rodent animals due to it being pricier, needing more complex care and lacking reagents, such as immunological reagents when dealing with ferret samples (Oh & Hurt, 2016; Schiffman et al., 2021). Nonetheless, considering its advantages over NHP and small rodents, the appeal of ferrets as good intermediate animal models for pathogenesis modelling and countermeasure evaluation continues to rise.

1.4.3 Small rodents

Small rodent animal models are useful for the initial evaluation of filovirus countermeasures before moving forward to NHPs as they are generally cheaper and easier to handle due to their smaller size, while also being able to mimic some of the clinical features of the disease. These models include mouse, hamster, and guinea pig animal models. However, a caveat implicated in using such animal models is the need for adaptation since filoviruses are apathogenic in these animals (Banadyga et al., 2016). The process of adaptation is performed through serial passaging – a process that involves challenging an immunocompetent rodent with a wild-type virus, collecting, and homogenizing infected organs like liver or spleen, and using the homogenate to inoculate a new set of animals. This process is repeated until the virus becomes 100% lethal in the animal (Banadyga et al., 2016).

1.4.3.1 Mouse model

1.4.3.1.1 Immunocompetent mouse model

Out of all the *Ebolavirus* species, mouse models have only been established for EBOV. Independent of the administration route, the adult immunocompetent mouse does not succumb to the disease upon infection with wild-type EBOV (Bray et al., 1999). In 1999, the first mouse-adapted (MA)-EBOV was developed by repeated passaging of wild-type EBOV (isolate Mayinga) in progressively older suckling BALB/c mice (Bray et al., 1999). The resulting MA-EBOV was

further used on other mouse strains including C57BL/6 and CD-1 (ICR) mice, in which the adapted virus was 100% lethal, with animals succumbing between 4-7 DPI (Bray, 2001; Bray et al., 1999; Bray et al., 2001; Mahanty et al., 2003). As seen in NHPs challenged with WT-EBOV, MA-EBOV targeted macrophages, monocytes and DCs and replicated in very high viral titres in the liver and spleen (Bray et al., 1999; Gibb et al., 2001). Systemic organ damage was observed, with the liver and spleen organs being heavily damaged, because of excessive production of pro-inflammatory cytokines (Bray et al., 1999; Mahanty et al., 2003). Animals also showed evidence of lymphocyte apoptosis, as indicated by the presence of lymphoblastic cells; thrombocytopenia and leukocytosis were also noted (Bradfute et al., 2007; Bray et al., 2001; Gibb et al., 2001). No hematological abnormalities indicative of coagulopathy and no hemorrhagic signs were observed in MA-EBOV mice (Bray et al., 1999; Bray et al., 2001; Gibb et al., 2001). But a study by Rasmussen et al. (2014) noted that a host's genetic background may influence the presence of coagulopathy as infection of Collaborative Cross Panel recombinant inbred mice with MA-EBOV showed prolonged coagulation times and vascular leakage (Siragam et al., 2018).

1.4.3.1.2 Immunocompromised mouse model

Contrary to healthy immunocompetent animals, an immunocompromised mouse model uses mice that have a defective immune system (Siragam et al., 2018). Since the immune system of the animal is compromised, a wild-type strain of the virus can be used without the need for adaptation (Siragam et al., 2018). The susceptibility of known pathogenic and non-pathogenic ebolaviruses has been evaluated in various immunodeficient mouse models – in interferon pathway deficient mice, these are knockout (KO) mice lacking either the IFN α/β cell surface receptors (IFN α/βR –/–) or STAT1 transcription factor (STAT1–/–) (Brannan et al., 2015; Bray, 2001; Escaffre et al., 2021; Lever et al., 2012; Raymond et al., 2011) and in severe combined immunodeficient (SCID) mice with compromised B- and T-lymphocytes that results in a defective adaptive immune response (Bray, 2001). Intraperitoneal inoculation of IFN α/βR –/– KO mice with the wild-type EBOV (variant Mayinga and Yambuku) showed 100% lethality, with an average time of death at 5 DPI (Bray, 2001; Lever et al., 2012). Lever et al. (2012) further demonstrated that IFN α/βR –/– KO mice infected wild-type EBOV (variant Yambuku) via the aerosol route also results in uniform lethality at 8 DPI. Surprisingly, when IFN α/βR –/– KO mice were challenged with the Kikwit EBOV variant, no significant lethality (0-20%) was observed (Brannan et al., 2015; Bray

et al., 2001). However, when both variants were tested in STAT1–/– KO mice, a lethal disease was observed (Bray, 2001; Raymond et al., 2011). Escaffre et al. (2021) also noted lethality in mice infected with the Kikwit variant, although at a lower lethality rate of 40%. On the other hand, the susceptibility of immunocompromised mice to SUDV disease is ambiguous – Bray (2001) observed 100% lethality in IFN α/βR –/– mice that were challenged with SUDV (Boniface), while no mortality was noted in the same knockout mice infected by Lever et al. (2012). Moreover, Brannan et al. (2015) revealed that IFN α/βR –/– knockout mice are lethal to 1,000 plaque-forming units (PFU) of SUDV (variant Boniface) with only about 37% lethality rate. Similar to observations made between two different variants of EBOV, IFN α/βR –/– knockout mice challenged with the same PFU of different SUDV variants had contrasting results, with SUDV (variant Gulu) only having one animal succumbing to the disease (Brannan et al., 2015). However, when the two SUDV variants were tested in STAT1–/– KO mice, both resulted in a uniform lethality (Escaffre et al., 2021; Raymond et al., 2011). Lastly, both EBOV (variant Mayinga) and SUDV (variant Boniface) were lethal to SCID mice, although the progress of the disease was much slower and lasted for weeks (Bray, 2001).

The presence of virus and viral antigen have both been detected in EBOV- and SUDV-infected IFN α/βR –/– KO mice via electron microscopy, immunochemistry, and plaque assay (Brannan et al., 2015; Bray, 2001). Knockout mice challenged with 1,000 PFU of EBOV and SUDV had high viral titres in the blood and spleen peaking at 3 DPI, but only SUDV-infected mice had virus detectable in the liver (Brannan et al., 2015). Their immunohistochemistry analysis later revealed that despite the absence of replicating viruses in the liver of EBOV-infected mice, virus replication was still occurring as indicated by detectable viral antigens in the liver Kupffer cells (Brannan et al., 2015). Biopsy of infected liver tissues showed evidence of hepatocellular degeneration and necrosis, as well as signs of intravascular coagulation in the blood vessels of the liver (Brannan et al., 2015). Moreover, spleen tissues of infected mice had exhibited unusual cell proliferation, one that is not typically observed in other animal models of EBOD, resulting in the destruction of the red and white pulp regions of the spleen (Brannan et al., 2015). Sick STAT1–/– KO mice also exhibited hepatocellular degeneration and necrosis, and detectable viral antigen in the liver Kupffer cells. Moreover, the spleen showed signs of lymphoid hyperplasia and lymphocytosis instead of the atypical cell proliferation seen in IFN α/βR –/– KO mice.

Aside from EBOV and SUDV, susceptibility of IFN α/βR –/– and STAT1–/– KO mice in other pathogenic and non-pathogenic ebolaviruses have also been evaluated. No signs of clinical disease and mortality were noted in IFN α/βR –/– KO mice that were challenged with BDBV, TAFV, and RESTV (Brannan et al., 2015; Bray, 2001; Lever et al., 2012). However, Brannan and colleagues showed that BDBV-infected IFN α/βR –/– KO mice had low but detectable viral titres in the blood, spleen, and liver (Brannan et al., 2015). Additionally, infected animal tissues also showed signs of hepatocellular degeneration in the liver and an increased number of lymphocytes in the spleen (Brannan et al., 2015). Meanwhile, the susceptibility of STAT1–/– KO mice to these three ebolaviruses was variable. BDBV was 20% lethal in mice and caused noticeable weight loss in infected animals, while TAFV and RESTV were both non-lethal (Escaffre et al., 2021; Raymond et al., 2011). No observable clinical signs were noted for the TAFV-infected mice (Escaffre et al., 2021), but interestingly in RESTV-infected mice, small changes in the liver and spleen tissues were observed, including, hepatocellular degeneration and necrosis as well as lymphoid hyperplasia and lymphocytosis (Raymond et al., 2011).

Ultimately, the lack of IFN response plays a critical role in the pathogenesis of some ebolaviruses. However, elimination of the host-immune response does not completely make less pathogenic and non-pathogenic ebolaviruses become more virulent in the host, thus indicating that other host factors and/or inherent differences between the viruses themselves are probable reasons for the observed pathogenic differences. However, immunocompromised mouse models are still useful to this date in identifying host-immune factors that limit pathogenesis of filoviruses.

1.4.3.2 Syrian hamster model

Syrian golden hamsters (*Mesocricetus auratus*) have been substantially used for human infectious diseases caused by pathogenic RNA viruses (Zivcec et al., 2011). For ebolaviruses, EBOV is the only well-characterized virus for the hamster model (Ebihara et al., 2013). However, a limitation of this model is its use of MA-EBOV for infection. Hamsters infected with wild-type EBOV only showed subclinical disease (Banadyga et al., 2016). Animals challenged with MA-EBOV intraperitoneally began showing signs of disease three days post-exposure and succumbed at 4-5 DPI. Similar in humans and NHPs, macrophages and DCs were identified as the target cells. Systemic viral replication was observed as high viral titres were found in multiple organs,

particularly in the liver and spleen. Necropsy and histopathological findings showed organs were heavily damaged as indicated by infiltration of inflammatory cells in the target tissues, extensive apoptosis, and cellular necrosis resulting from the upregulated and uncontrolled pro-inflammatory cytokine response. Coagulation dysfunction was also observed. Ebola virus disease manifestations are currently well represented by the Syrian hamster model compared to other filovirus rodent models. In conjunction with its ease of use, availability, low cost, and improved availability of research tools and assays on hamster samples, the hamster model presents itself as an excellent small animal model for screening therapeutics and vaccines and for modelling filovirus disease (Wahl-Jensen et al., 2012).

1.4.3.3 Guinea pig model

Guinea pigs (*Cavia porcellus*) are generally not susceptible to infection with wild-type EBOV and only develop short-term illness, specifically with strain-13 guinea pigs (Simpson et al., 1968). Animals experienced fever, diarrhea, loss of appetite, weight loss, and decreased activity (Ryabchikova et al., 1996; Simpson et al., 1968). Uniform lethality in strain-13 inbred and outbred Hartley guinea pigs was achieved after a series of passages, particularly for EBOV (Cheresiz et al., 2016; Connolly et al., 1999; Cross et al., 2015; Subbotina et al., 2010; Volchkov et al., 2000) and SUDV (Wong et al., 2016). Collectively, the different guinea pig adapted (GPA) Ebola viruses (isolate Mayinga) require fewer rounds of passaging, up to 8 passages (Connolly et al., 1999; Cross et al., 2015; Subbotina et al., 2010; Volchkov et al., 2000), while the only GPA-SUDV needed 20 passages (Wong et al., 2016). Infection with GPA-EBOV resulted in the onset of fever between 4-7 dpi, with viremia being detected very early at 2 DPI, and time of death between 8-12 dpi (Connolly et al., 1999; Cross et al., 2015; Subbotina et al., 2010). GPA-SUDV infected animals showed high viremia starting from 5 days after exposure and eventually died 9-14 DPI (Wong et al., 2016). High titres of virus were detected in several organs, mainly in the spleen and liver, but the titres were lower than those observed in mice or NHP (Banadyga et al., 2016; Connolly et al., 1999; Cross et al., 2015; Wong et al., 2016). Other hallmarks such as leukocytosis, lymphopenia, thrombocytopenia, and coagulopathy were also present in infected guinea pigs. Lymphocyte apoptosis and upregulated inflammatory response were only observed in outbred Hartley guinea pig strains (Banadyga et al., 2016; Cross et al., 2015). Interestingly, infected guinea pigs also

lacked the maculopapular rash, a similar observation seen with MA-EBOV mice (Bray et al., 1999; Connolly et al., 1999; Cross et al., 2015).

1.5 THE GUINEA PIG-ADAPTED SUDAN VIRUS

In 2015, Gary Wong and his colleagues developed the first guinea pig model for SUDV and characterized the first guinea pig-adapted SUDV (GPA-SUDV) (Wong et al., 2016) (Fig. 1.). In their study, two female Hartley guinea pigs were intraperitoneally inoculated with approximately $1x10^5$ TCID50 of the wild-type SUDV (strain Boniface) and euthanized 7 days post-infection (dpi) to collect and pool the liver and spleen. Pooled liver and spleen organs were homogenized, filtered, and the supernatant was used to inoculate a new set of two female Hartley guinea pigs. The process was repeated, and the virus was serially passaged until it caused 100% lethality in the animals, which took 20 passages. Genomic comparison, through Sanger sequencing, between the guinea pig-adapted virus and wild-type virus showed that SUDV acquired 16 nucleotide changes, 6 of which resulted in an amino acid change (Table 1.1). However, the significance of these adaptive mutations remains unknown.

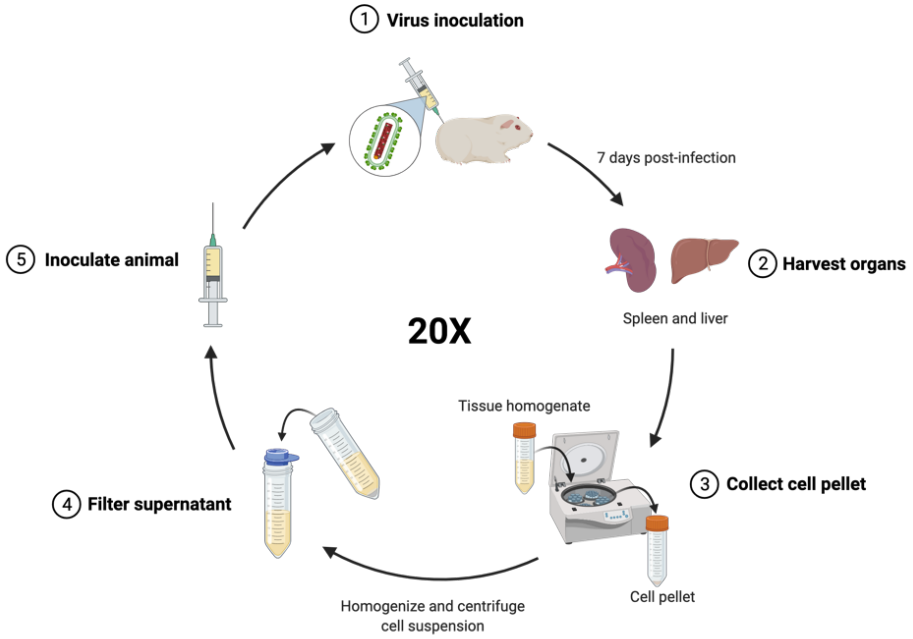


Figure 1.10 Schematic representation of SUDV adaptation in guinea pigs. The first step was to inoculate healthy guinea pigs with a wild-type SUDV. The liver and spleen were collected from euthanized animals 7 days post-infection and was pooled and homogenized. The tissue homogenate was centrifuged and filtered to obtain the cell pellet, which was resuspended in buffer and homogenized. The resulting cell suspension was centrifuged to pellet cell debris and subsequently filter out the supernatant, which was used to inoculate a new set of guinea pigs. The entire process was repeated until 100% lethality was achieved in the animals. Created with BioRender.com

Table 1.1 Genomic mutations that were identified in the final guinea pig adapted SUDV by Wong et al. (2016).

Gene	Nucleotide position	Region[a]	Nucleotide change	Amino acid change[b]
NP	35	5'UTR	t→c	–
NP	850	CDS	g→a	–
NP	2965	3'UTR	t→c	–
VP35	3288	CDS	a→g	N51D
VP35	4348	3'UTR	t→c	–
VP40	4966	CDS	t→c	–
VP40	5378	CDS	c→t	P309S
VP40	5767	3'UTR	c→t	–
GP	6085	CDS	g→a	A30T
GP	6448	CDS	t→c	Y151H
GP	6454	CDS	t→c	F153L
GP	6471	CDS	t→c	–
VP24	9678	3'UTR	a→g	–
VP24	10909	CDS	c→t	P204L
L	12266	CDS	t→c	–
L	15047	CDS	c→t	–

[a]CDS, coding sequence region; UTR, untranslated region
[b]The amino changes are depicted as original amino acid – protein residue number – new amino acid.

1.6 THESIS OBJECTIVES

1.6.1 Rationale

Small immunocompetent rodents such as mice, hamsters and guinea pigs do not succumb to disease following infection of wild-type filoviruses isolated from humans. Adaptation, performed as a series of passaging of the wild-type virus in the host animal, is required to render the virus

pathogenic. The virus acquires genomic mutations that contribute to virulence and lethality during this process. Comparison of genetic differences between the wild-type and rodent-adapted viruses have been useful in providing insights on "hot spots" for species-specific adaptation and probable molecular mechanisms governing filovirus pathogenicity. Unlike EBOV, for which several rodent adaptation and molecular exploration exists, such studies are almost non-existent for SUDV. In this study, the adaptation of SUDV in the guinea pig animal model will be explored by comparing the genomic changes in SUDV's viral genome between the starting wild-type virus and across all 20 passages using high-throughput sequencing technology. Awareness of potential key guinea pig-adapted SUDV mutations does not only elucidate the potential molecular mechanisms that control SUDV pathogenicity but will also aid in the development of future prophylactics and countermeasures for the treatment of SUDV disease.

1.6.2 Hypothesis and Objectives

The GPA-SUDV was 100% lethal to immunocompetent Hartley guinea pigs and has acquired sixteen nucleotide changes, six of which resulted in an amino acid change (Wong et al., 2016). However, how these mutations progressed over time during the adaptation series has not been fully explored. Three main hypotheses have been established:

1. Throughout the adaptation process, the virus acquires several genomic mutations that fluctuate in frequency over time.
2. Mutations that were retained by the GPA-SUDV by the end of the adaptation process are those that reached a frequency near 100%.
3. These adaptive mutations likely play a role in the various aspects of the viral replication cycle and/or immune evasion.

The following objectives were set to determine if such hypotheses were true:

1. Deep sequence SUDV genome from each of the passages and identify SNPs.
2. Track how early or late the mutations appear during serial passaging.
3. Track the changes in the frequency of each identified mutation

2 MATERIALS AND METHODS

2.1 GUINEA PIG LIVER AND SPLEEN TISSUE HOMOGENATES

Sudan virus-infected guinea pig liver and spleen tissue homogenates used in this study were from a previous study by Wong and his colleagues (Wong et al., 2016). At each passage, liver and spleen organs were collected 7 days post-infection, pooled together, and homogenized. The first batch of technical duplicates of homogenized tissues for each passage was removed from biosafety level 4 using the QIAamp Viral RNA Mini Kit (QIAGEN) as per the manufacturer's instructions and institutional biosafety protocols. Additional homogenates for passages 3, 8, 9, 13, 16 and 18 were taken out from biosafety level 4 in the same manner as the former samples for extraction using a magnetic bead-based approach (see Section 2.2.2).

2.2 VIRAL RNA PURIFICATION

2.2.1 Silica-based column extraction

Viral RNA from the first batch of homogenized tissue samples was isolated using the QIAamp Viral RNA Mini Kit (QIAGEN) as per the manufacturer's spin protocol. All samples and reagents were equilibrated to room temperature before use. For each sample, 650 µl of lysate was loaded to the QIAamp mini-column and centrifuged at 8,000 rpm for 1 minute, and the filtrate was discarded. This step was repeated until all 1.2 ml of the lysate was loaded to the column. The column was washed with 500 µl Buffer AW1, centrifuged at 8,000 rpm for 2 minutes and the filtrated discarded. A second was performed with 500 µl of Buffer AW2, centrifuged at 14,000 RPM for 3 minutes and the filtrate discarded. Excess buffer was removed from the column by centrifuging at 14,000 rpm for 1 minute. RNA was eluted using 50 µl of Buffer AVE and spun at 8,000 rpm for 1 minute. Eluates were collected in 1.5 ml microcentrifuge tubes and stored at -80°C.

2.2.2 Magnetic bead-based purification

Purification of viral RNA from the second batch of additional tissue homogenates (i.e., passages 3, 8, 9, 13, 16 and 18) removed from biosafety level 4 was conducted using Applied Biosystems™ MagMAX™ Viral/Pathogen II (MVP II) Nucleic Acid Isolation Kit (Thermo Fisher Scientific), following a modified version of the KingFisher Flex – 200 µl sample input 2 wash-

steps protocol. A 96 deep-well plate was prepared as seen in Fig. 2.1. In a Microtitre 96-deep well plate, rows A1-A6 were loaded with 20 µl of proteinase K and 10 µl of magnetic beads. 700 µl of each tissue homogenate was added and mixed subsequently. Rows C1-C6 were filled with 500 µl of washing buffer, while rows E1-E6 and G1-G6 were filled with 500 µl of 80% ethanol. A KingFisher Duo Prime 12-tip comb was placed on the last row of the plate. An aliquot of 50 µl of elution buffer was used to elute RNA into an elution strip. Viral RNA extraction was performed with KingFisher Duo Prime magnetic particle processor and the BindIt 4.1 software included with the instrument (Thermo Fisher Scientific).

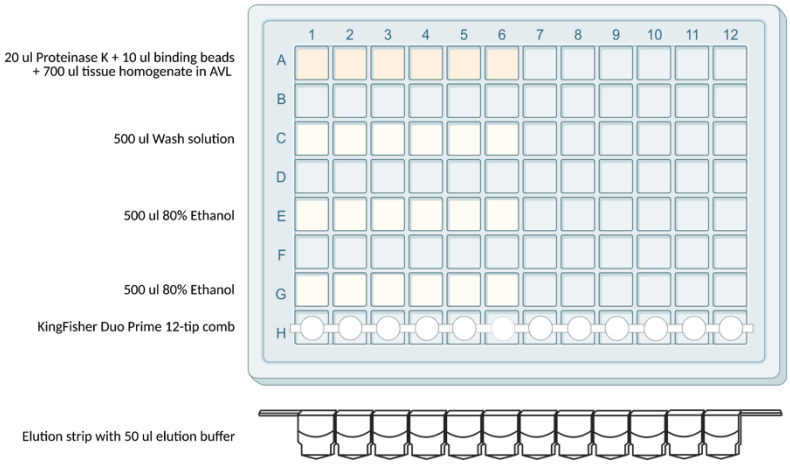


Figure 2.1 The layout of a Microtitre 96-deep well plate prepared for magnetic bead-based extraction of GPA-SUDV tissue homogenates using the KingFisher Duo Prime instrument. Rows A1-A6 each contained 20 µl of proteinase K and 10 µl of magnetic beads. 700 µl of passages 3, 8, 9, 13, 16, and 18 were then subsequently added to rows A1-A6, respectively. 500 µl of wash solution was added in rows C1-C6, while 500 µl of 80% ethanol wash was allotted to rows E1-E6 and G1-G6. A plastic 12-tip comb was placed in row H. A single elution strip was prepared by pipetting 50 µl of elution buffer into the elution strip. This figure was created with BioRender.com

Concentrations of extracted RNA from both purification methods were determined using the Qubit™ RNA High Sensitivity (HS) Assay Kit and Qubit 4.0 fluorometer device (Thermo Fisher Scientific) as per manufacturer's guidelines. Briefly, a working solution was prepared by diluting the Qubit HS reagent 1:200 in the Qubit RNA HS buffer. For each sample, 198 µl of working solution and 2 µl of RNA was added in a 0.5 ml thin-walled PCR tube. Two standard assay tubes were prepared by adding 10 µl of pre-diluted RNA standards to 190 µl of working solution. Standards and samples were vortexed and incubated at room temperature for 2 minutes. Upon

selecting HS RNA Assay on the Qubit 4.0 fluorometer, the device was calibrated using the two standards and samples were read. RNA samples with concentrations of >1000 ng/µl were diluted at 1:10 and concentrations were re-measured.

2.3 VIRAL RNA DETECTION IN TISSUE HOMOGENATES USING RT-QPCR

2.3.1 2.3.1. RT-qPCR reaction

Viral genomes were detected from guinea pig tissue homogenates using the Light Cycler® 480 RNA Master Hydrolysis Probes Kit (Roche Life Science) with the *NP* gene as a target. A single MicroAmp™ EnduraPlate™ Optical 96-well reaction plate was prepared as seen in Fig. 2.2. SUDV-NP standard, primers, and probe were kindly provided by Dr. Wenjun Zhu (Canadian Food Inspection Agency). A stock solution of SUDV-NP standard was diluted using a 10-fold dilution series. Stock solutions of forward and reverse primers were mixed with a final concentration of 20 µM per primer. Probe concentrations of 10 µM/ul were used. Primers and probe sequences used in the reactions were as follows: forward (sense), 5'-GTGACGAAGATGKTGAGAGC-3'; reverse (antisense), 5'TTGTAGACTGGTGCTGGTGG3'; probe, FAM-CAGRGGAGAACAMCCCAACTGTAGC-BHQ1. Each well contained 5.92 µl of reagent 3, 7.40 µl reagent 1, 1.28 µl reagent 2, 0.8 µl reagent 4, 0.3 µl primer mix, 0.3 µl probe, and 4 µl of a template (i.e., RNA or NP standard). DNase/RNase-free water was used as a negative control template. Reaction conditions used with QuantStudio Real-Time PCR System (Thermo Fisher Scientific) instrument is found in Table 2.1.

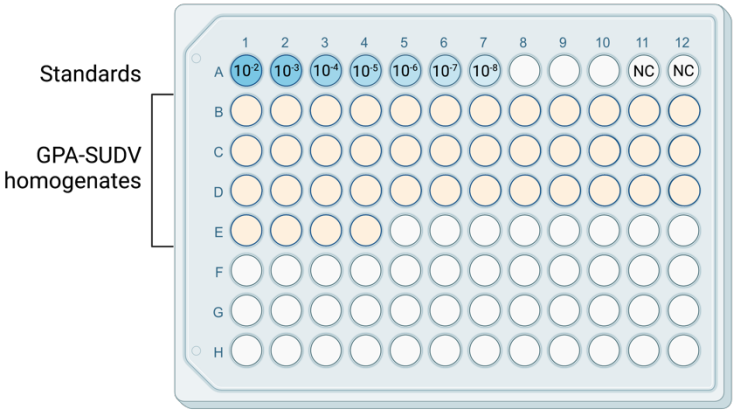


Figure 2.2 Sample plate layout of viral RNA samples from guinea pig liver and spleen tissue homogenates in a 96-well plate for detection of SUDV RNA using RT-qPCR. Samples used were RNA isolated via the silica-column based method. Reaction mixtures was prepared using Light Cycler® 480 RNA Master Hydrolysis Probes Kit (Roche Life Science). Each reaction contains 4 µl of RNA or plasmid as template, 7.40 µl reagent 1, 1.28 µl reagent 2, 5.9 µl of reagent 3, 0.8 µl reagent 4, 0.3 µl primer mix, and 0.3 µl probe. The blue (rows A1-A7) denotes wells containing the standard curve samples. NC refers to no template negative control for which RNA was replaced with DNase/RNase-free water. Created with BioRender.com

Table 2.1 Quantitative real-time PCR conditions used during detection of SUDV RNA in guinea pig tissues.

Temperature	Duration
63°C	3 minutes
95°C	30 seconds
95°C	15 seconds
60°C	30 seconds

2.3.2 Calculation of genome equivalents

2.3.2.1 Generating the SUDV NP standard curve

A standard curve was generated from a SUDV *NP* gene, cloned into a vector, serially diluted 10-fold for six times from a starting concentration of 1.67 ng/µl. Genome equivalents (GEQ) for each standard dilution were calculated using the equation: $\mathrm{GEQ} = \frac{[\text{amount of plasmid (ng)}]\times 6.022\text{x}10^{23}}{\text{size of virus genome (bp)} \times 1\text{x}10^{9} \times 650\ \text{Da}}$; where $6.022\text{x}10^{23}$ is Avogadro's number and 650 Da is the average weight of a single DNA bp that can also be expressed in units of g/mole. Cycle threshold (Ct) values of each standard dilution obtained through qRT-PCR were then plotted against the log-transformed GEQ values and a simple linear regression analysis was performed to obtain the line of best-fit equation using GraphPad Prism 9 software (Appendix B.1). The equation of the line was as follows: y= -3.363x+27.28, where y represents the Ct value, x as the $\log_{10}$ GEQ, -3.363 is the slope of the line, and 27.28 is the y-intercept.

2.3.2.2 GEQ per ml in homogenized tissues

GEQ for each sample was measured against a log-transformed standard curve of the SUDV *NP* gene. The Ct values were converted to GEQ using the equation for the slope of the standard curve (Section 2.3.2.1). To determine the number of GEQ per ml of homogenized sample, the

following equation was used: $\frac{\text{GEQ}}{\text{ml}} = \text{GEQ} \times (\frac{\text{extraction volume}}{\text{template volume}}) \times (\frac{1}{\text{homegenized tissuevolume}}) \times (\frac{1000\ \mu\text{l}}{1\ \text{ml}})$.

2.4 AMPLIFICATION OF VIRAL GENOME FROM LIVER AND SPLEEN TISSUE HOMOGENATES OF SUDV-INFECTED GUINEA PIGS

2.4.1 Reverse transcription PCR (RT–PCR)

Extracted RNA (Section 2.2.1) was reverse transcribed into cDNA using SuperScript™ III First-Strand Synthesis SuperMix Kit (Invitrogen™). In a 0.2 mL, thin-walled PCR tube, a reaction mixture was prepared as per Table 2.2. A total of 12 gene-specific forward primers, listed in Table 2.3, were pooled in the reaction along with DNase/RNase free-water (MP Biomedicals), annealing buffer and RNA template. These primers were from Wong et al. (2016). At 75°C, the initial mixture was incubated for 5 minutes and placed on ice for 2-5 minutes, followed by the addition of First-Strand Synthesis Mix and SuperScript® III/RNaseOUT™ Enzyme Mix (Invitrogen). The final mixture was incubated under conditions found in Table 2.4.

Table 2.2 RT-PCR reaction mixture for the synthesis of cDNA from GPA-SUDV RNA.

Components	Volume
Annealing Buffer	2 µl
Gene-specific Primers (2 µM)	0.8 µl/primer
RNase/DNase Free Water	0.4 µl
RNA template	4 µl
2X First-Strand Reaction Mix	20 µl
SuperScript® III/RNaseOUT Enzyme Mix	4 µl
Total volume per reaction	40 µl

Table 2.3 Oligonucleotides used for reverse transcription of GPA-SUDV viral RNA into cDNA.

Oligonucleotide[a]	Sequence (5'-3')
SBEV 1F	CGGACACACAAAAAGAAAGA
SBEV 5F	TATGCTCCGTTTGCACGTCT
SBEV 9F	CCCTTGAAGAAGCCTACCCG
SBEV 15F	CTCCCCTGCTTGGTACTATGT
SBEV 20F	GCCTGATTAAGG CCCAACCT
SBEV 25F	GCTCCAGCCAAATCCTGAGT
SBEV 29F	ACTGTCCCTCCAGCACCTAA

SBEV 32F	CAAACACCTTGCTCTGCCAA
SBEV 35F	TGTATGATCTCAGTTTAATCTCTCT
SBEV 39F	GGCGAAATCAATTCCCACCAC
SBEV 45F	CGGTAGACAGTTAATAGG GGCA
SBEV 50F	CCCTGAAGAAAGCCGCAG TA

[a]F (forward) primer depicts sense (+) orientation.

Table 2.4 RT-PCR conditions for synthesizing cDNA from GPA-SUDV sample.

Temperature	Duration
25°C	10 minutes
42°C	5 minutes
55°C	30 seconds
60°C	5 minutes
85°C	5 minutes
10°C	15 seconds

2.4.2 Amplification of SUDV genomic fragments

The initial amplification of the SUDV genome was conducted using primer sets listed in Table 2.5 to obtain 16 overlapping amplicons that spanned the entire genome. The selected primers were from a collection or primers previously used by Wong et al. (2016). The second set of primers (Table 2.6) was selected from the same pre-existing primer collection as some of the fragments were not amplifiable by the first set of primers. A new collection of primers was designed using Primal Scheme (Quick et al., 2017). The amplicon size was set to 300 bp and KT878488.1 was used as the reference sequence. The designed scheme generated a collection of 87 primers, which was ordered through Integrated DNA Technologies (IDT) as 100 µM stocks. From these 87 primer pairs, the list of primer pairs in Table 2.7 were selected to target amplicons that were missing in different passages and to re-amplify regions of low coverage for some passages. In instances where the 1-2 kb product did not amplify despite three attempts, a last attempt to amplify these products into smaller sized amplicons was performed by selecting primers from the Primal Scheme set that would generate 600-800 bp amplicons. These primers are also listed in Table 2.7. The PCR reactions were carried out as 25 µl reactions using CloneAmp HiFi Polymerase (Takara Bio). PCR reactions were prepared as per Table 2.8 and ran under conditions found in Table 2.9.

Table 2.5 First set of primers used for amplification of whole SUDV genome.

Oligonucleotide[a]	Sequence (5'-3')	Sense	Start	Stop	Fragment size (bp)
SBEV 1F	cggacacacaaaaagaaaga	+	1	1353	1352
SBEV 3R	agacgtgcaaacggagcata	–			
SBEV 4F	aggtccatgcagaacaaggg	+	987	2548	1561
SBEV 7R	cgggtaggcttcttcaaggg	–			
SBEV 8F	caactgtagctccaccagca	+	2262	3618	1356
SBEV 11R	ctggaggtgcttgtccatgt	–			
SBEV 12F	cctgaatcgcagttgtgctg	+	3494	5088	1594
SBEV 16R	ttcgggtgaaatgagagccc	–			
SBEV 17F	gctcgcatcctacacgatca	+	4801	6495	1694
SBEV 20R	agccagcctgtcatagagga	–			
SBEV 21F	ccatgcctttgggtgttgtg	+	6092	7808	1716
SBEV 25R	caggatcctacatgtcccgc	–			
SBEV 26F	gccacgggtaaatgcaatcc	+	7515	9043	1528
SBEV 29R	ttgccccaagttttccctcc	–			
SBEV 30F	ctggcaaatcctacggctga	+	8816	10278	1462
SBEV 32R	tcttaggcttgcaaaaggagc	–			
SBEV 32F	caaacaccttgctctgccaa	+	9866	11391	1525
SBEV 34R	agatgctgtgagtgaaactgga	–			
SBEV 35F	acagggaggttctaatggtg	+	11127	12665	1538
SBEV 37R	gccccagtgcttttggattg	–			
SBEV 38F	attgggatcggaggggtaca	+	12395	13885	1490
SBEV 40R	caggcacttgtgactttggc	–			
SBEV 41F	ggccaccttggtggtattga	+	13644	15156	1512
SBEV 44R	gcttaatagcgggctgacct	–			
SBEV 45F	cggtagacagttaataggggca	+	14906	16533	1627
SBEV 47R	gcggcaaaatgatgtgtgga	–			
SBEV 48F	agctttgttgaggagtgggt	+	16209	17798	1589
SBEV 51R	gagggcaaggtgtcttgtga	–			
SBEV 52F	gcgatttggatgggatgtgc	+	17422	18642	1220
SBEV 54R	tatgttattcccaccggccc	–			

SBEV 53F	cttgccctcctccaaactga	+	17790	18875	1085
SBEV 56R	tggacacacaaaaaagagaa	–			

Table 2.6. List of the second set of primers used for amplification of the whole SUDV genome.

Oligonucleotide[a]	Sequence (5'-3')	Start	Stop	Fragment size (bp)
SBEV 2F	catttcagcacaactgcgat	205	1353	1148
SBEV 3R	agacgtgcaaacggagcata			
SBEV 4F	aggtccatgcagaacaaggg	987	2865	1878
SBEV 8R	ttgcggtacgggtatcaacc			
SBEV 8F	caactgtagctccaccagca	2262	4063	1801
SBEV 12R	atctttggtgacggtgggac			
SBEV12F	cctgaatcgcagttgtgctg	3494	5410	1916
SBEV 17R	actggcgggtgaatgacaat			
SBEV 17F	gctcgcatcctacacgatca	4801	6495	1694
SBEV 20R	agccagcctgtcatagagga			
SBEV 21F	ccatgcctttgggtgttgtg	6092	8025	1933
SBEV 26R	aacaaagcagcttgcagacg			
SBEV 26F	gccacgggtaaatgcaatcc	7515	9265	1750
SBEV 30R	aggagggagcccttcattca			
SBEV 30F	ctggcaaatcctacggctga	8816	10952	2136
SBEV 33R	cttgacaggaccagggtgtc			
SBEV 32F	caaacaccttgctctgccaa	9866	11825	1959
SBEV 35R	acagggaggttctaatggtg			
SBEV 35F	tgtatgatctcagtttaatctctct	11127	13090	1963
SBEV 38R	tcgggtacacgtttggtgtt			
SBEV 38F	attgggatcggaggggtaca	12395	14319	1924
SBEV 41R	ggtcacccaggttgcgataa			
SBEV 40F	gcctccttcaccaagcatct	13348	15156	1808
SBEV 44R	gcttaatagcgggctgacct			
SBEV 45F	cggtagacagttaataggggca	14906	16881	1975
SBEV 48R	tactgcggctttcttcaggg			
SBEV 48F	agctttgttgaggagtgggt	16209	18093	1884

SBEV 52R	cagcatcactcatccggtgt			
SBEV 52F SBEV 55R	gcgatttggatgggatgtgc tgtgcgcttaattttcctggtg	17422	18831	1409
SBEV 52F SBEV 56R	gcgatttggatgggatgtgc tggacacacaaaaaagagaa	17422	18875	1453

[a]F and R depicts sense (forward) and antisense (reverse) primer orientations, respectively.

Table 2.7 List of forward and reverse primers used to amplify missing SUDV fragments and low coverage areas of SUDV genome. Primers were designed using Primal Scheme.

Oligonucleotide[a]	**Sequence (5'-3')**	**Start**	**Stop**	**Fragment size (bp)**
SUDVB 11F SUDVB 17R	caggaagaatcagagcccgatg aggtcgcccaaaattttcctca	2153	3755	1603
SUDVB 41F SUDVB 43R	ggatatctgtcctactcttagaaaagga tccttattcatatcctgtcccct	8647	9391	746
SUDVB 43F SUDVB 43R	ataaaggtggtgctttcgaggc cgtgtaggactaccctgtacct	9090	9779	690
SUDVB 45F SUDVB 47R	acacaatggtgagtatttgtttcctg ggttctttctgttatccttatttttgttacaag	9476	10223	748
SUDVB 46F SUDVB 53R	aaagaacccatggacacactct aaaagtcctatggtggcaagga	9695	11518	1824
SUDVB 47F SUDVB 49R	cctgcaagataaggttatcaaacagg gtggctcgatgagggaatgatc	9917	10641	725
SUDVB 49F SUDVB 51R	aacgggagctagagcaagga gccctatcatacaatctttgttatggc	10339	11083	745
SUDVB 50F SUDVB 52R	ccatttgggccttgagggtaat tgaaggcaaaactagtgtcatgc	10567	11296	730
SUDVB 52F SUDVB 59R	aaccgtcaagcattacttcattga cggtctgatatcacactgtacca	10999	12835	1837
SUDVB 52F SUDVB 54R	aaccgtcaagcattacttcattga tgataaatcgtaaaacaatggcgtca	11001	11727	727
SUDVB 54F SUDVB 56R	acacagtagaggtctaggtgtgtt agtcagtcgatgcatgaggaac	11425	12171	747
SUDVB 56F	caggatgcggcattccttaatt	11862	12615	754

SUDVB 58R	tcgaagtctgagcaacagttgg			
SUDVB 57F SUDVB 64R	tggtgattatattttctggaagatccc tgaggaatatcccacaggcact	12092	13899	1808
SUDVB 58F SUDVB 60R	gaagatccagtgtccgctgatt aaggtggactgtagcccaaaac	12315	13062	748
SUDVB 60F SUDVB 62R	tgaaacgtactgtgtattcaagtacag tgtgaattcatacctgaaagccaga	12752	13472	721
SUDVB 62F SUDVB 64R	tgccccagaatcgaaatttttctt tgaggaatatcccacaggcactatcgaaatttttctt	13162	13899	738
SUDVB 64F SUDVB 70R	acataatgtaactttagagaatagggaatatcc agtgacccaggttaacctagagg	13586	15221	1636
SUDVB 64F SUDVB 66R	acataatgtaactttagagaatagggaatatcc ctggaataggcctgaagtcaca	13586	14343	758
SUDVB 66F SUDVB 68R	acttgccagtataggaactgcc ccatctttgaagtgttatctttctgagt	14045	14783	739
SUDVB 68F SUDVB 70R	tctttccttcgtcagattgtcaga agtgacccaggttaacctagagg	14481	15221	741
SUDVB 69F SUDVB 76R	ggaaccagaactttattagcatctagaa tagcggcaaaatgatgtgtgga	14691	16535	1845
SUDVB 75F SUDVB 83R	actatctccagaaccagatccaca tggtagtaagtacaactcatggtgc	16009	17990	1982
SUDVB 81F SUDVB 87R	acacaccgagattggttttcaaa atttgtgtgcgcttaattttcctg	17238	18836	1599

[a]F and R depicts sense (forward) and antisense (reverse) primer orientations, respectively.

Table 2.8 PCR reaction mixture for the synthesis of amplicons from GPA-SUDV RNA.

Components	**Volume**
CloneAmp HiFi Premix (2X)	12.5 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
cDNA template	4 µl
DNase/RNase-free water	6.5 µl

Total volume per reaction	25 µl

Table 2.9. PCR reaction thermal profiles for amplification of different GPA-SUDV fragments.

Steps	**Temperature**	**Duration**
Initial denaturation	95°C	3 minutes
Denaturation	95°C	30 seconds
Annealing	56°C	30 seconds
Extension*	72°C	2 minutes
Final extension	72°C	10 minutes

*Duration of extension step was reduced to one minute for fragments < 1 kb.

2.4.3 Purification of GPA-SUDV amplicons

2.4.3.1. Gel electrophoresis

PCR products were loaded to a 1% agarose gel, made with UltraPure™ agarose (Thermo Fisher Scientific), using a 6X DNA purple gel loading dye (New England Biolabs). Products were run with 1X Tris-acetate-EDTA (TAE) buffer for 1 hour at 120 volts. Approximate sizing of the DNA fragments was determined using the ready-to-use GeneRuler DNA Ladder Mix (Thermo Fisher Scientific) that was run alongside the PCR products. PCR products were visualized with SYBR® Safe DNA Gel Stain (Life Technologies) on Safe Imager™ 2.0 Blue-Light Transilluminator (Invitrogen™) and DNA fragments of the correct size were excised from the gel using a scalpel. PCR fragments that were re-amplified on a separate occasion were pooled into the same 1.5 ml microcentrifuge tube.

2.4.3.2. DNA purification and quantification

DNA from excised gels was purified using QIAquick Gel Extraction Kit (QIAGEN). For each sample, 600 µl of Buffer QG was added to the gel slice(s) to dissolve the gel at 50°C for 10 minutes. After dissolving completely, 200 µl of isopropanol was mixed into the sample. DNA was then bound to the QIAquick columns as per the manufacturer's instructions. Since the DNA will subsequently be used for next-generation sequencing, the spin columns were washed with 500 µl of buffer QG prior to washing with 750 µl of Buffer PE. DNA was eluted from the QIAquick columns with 30 µl of Buffer EB. DNA concentration was measured using the Qubit™ dsDNA HS Assay Kit and Qubit 4.0 fluorometer device (Thermo Fisher Scientific) as per manufacturer's guidelines. For each sample, the reaction mixture included 198 µl of working solution and 2 µl of DNA added in a 0.5 ml thin-walled PCR tube.

2.5 RNA INTEGRITY

The quality of RNA was retroactively measured to potentially identify the reason regarding the variable success observed in the PCR amplification of amplicons from the different GPA-SUDV passages. In this study, a bioanalyzer assay was conducted to determine if the extracted total RNA from each passage was intact or highly degraded. If the starting RNA template is not

intact, this can subsequently affect downstream applications by producing an incomplete cDNA during RT-PCR and an unsuccessful PCR amplification of longer amplicons. The state of RNA quality was also re-confirmed by amplifying housekeeping genes that are ubiquitously expressed in mammals. If indeed poor in quality, the rest of the RNA co-existing with the viral RNA must also be degraded.

2.5.1 Bioanalyzer assay

Using the RNA concentrations shown in Appendix A, RNA samples extracted via silica column and magnetic beads were diluted to 1 ng/µl, heat-denatured at 70°C for 2 minutes and placed on ice preceding the assay. Priming of the RNA Pico chips with the gel-dye mixture, loading of conditioning solution and marker, and setting up the bioanalyzer was done in accordance with the Agilent RNA 6000 Pico Kit instructions (Agilent Technologies). Each RNA Pico chip was loaded with 1 µl of pre-aliquoted diluted RNA ladder, and 1 µl of experimental samples into each of the available sample wells. RNA Pico chips were vortexed at 2400 rpm using the IKA vortex mixer (model MS3) (Agilent Technologies) and assayed using the Agilent 2100 Bioanalyzer by selecting the Eukaryote Total RNA Pico High Sensitivity Assay found in the 2100 Expert Software included with the instrument (Agilent Technologies). The chromatograms and gel images were visually examined and exported using the 2100 Expert Software.

2.5.2 Amplification of housekeeping genes

2.5.2.1 Cell lines

The Madin-Darby canine kidney (MDCK) and baby hamster kidney (BHK) cells used in this study were from the American Type Culture Collection (ATCC). The MDCK and NIH-3T3 cells were maintained at 37°C and 5% CO_2 in Dulbecco's modified Eagle high glucose medium supplemented with 10% heat-inactivated bovine growth serum (BGS; Thermo Fisher Scientific), 1% penicillin-streptomycin (Gibco), and 1% L-glutamine. These cell lines were used as internal controls during the amplification of various housekeeping genes from the GPA-SUDV RNA samples.

2.5.2.2 Cell counts and pelleting

Old cell media from each cell line was aspirated, and cells were washed twice with 10 ml of phosphate-buffered saline pH 7.4. Cells were dissociated from the flask using 3 ml of 0.25% trypsin EDTA (Gibco) and incubated at 37°C for 3-5 minutes. The reaction was stopped with an addition of 7.5 ml of fresh media. A 1:1 mix of cell aliquot and 0.4% trypan blue stain (Thermo Fisher Scientific) was used for cell counting. Cell counts were performed in duplicate for each cell line using the Countess cell counting chamber slides and Countess™ II FL automated cell counter (Thermo Fisher Scientific). Average cell counts were used to determine the amount of lysis reagent to use for RNA purification (Section 2.5.2.3).

2.5.2.3 Total RNA extraction

Pelleted MDCK cells were lysed with 300 µl of Invitrogen TRIzol® Reagent (Thermo Fisher Scientific), while BHK cells were lysed with 600 µl of the TRI reagent. Direct-zol™ RNA Mini-Prep Plus Kit (Zymo Research) was used to purify total RNA from these cells following the spin-column purification guidelines as per the manufacturer's instructions. During the DNase treatment step, a 5X reaction mix was prepared by mixing 25 µl of DNase I and 375 µl of DNA digestion buffer and then aliquoting 80 µl of this mix to each sample. RNA was eluted with 60 µl of DNase/RNase-free water provided in the kit.

2.5.2.4 Primers and reagents

Primers for housekeeping genes such as β-actin, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), and porphobilinogen deaminase (PBGD) were synthesized by Integrated DNA Technologies. Primer set sequences and expected product size are listed in Table 2.10. The β-actin and GAPDH primer sequences were from Yamada et al. (2005) and PBGD primers by Shimizu et al. (1994). All stock primers were diluted to 10 µM with RNase and DNase-free water (MP Biomedicals, Solon, Ohio, USA) prior to use.

Table 2.10 The sets of primers used for amplification of three different housekeeping genes from GPA RNA samples.

Primer Name	Sequence (5'-3')	Product length
β-actin F	ccaactgggacgacatggag	279 bp

β-actin R	cgtagccctcgtagatgggc	
GAPDH F GAPDH R	accacagtccatgccatcac tccaccaccctgttgctgta	452 bp
PBGD F PBGD R	attcggggaaacctcaaca cccacagcatacatgcatt	152 bp

F depicts sense (forward) primer orientation.
GAPDH, glyceraldehyde 3-phosphate dehydrogenase; PBGD, porphobilinogen deaminase

2.5.2.5 RT-PCR and PCR reactions

Cellular RNA, extracted from MDCK and BHK cells, and RNA extracted from the 20 homogenized guinea pig tissues were reverse transcribed using SuperScript™ III First-Strand Synthesis SuperMix Kit (Thermo Fisher Scientific). In a 0.2 mL, thin-walled PCR tube, a reaction mixture was prepared as per Table 2.11. The initial mixture of annealing buffer, random hexamers, water, and RNA was incubated for 5 minutes at 75°C for 5 minutes and placed on ice for at least 2 minutes. Afterwards, the First-Strand Synthesis Mix and SuperScript® III/RNaseOUT™ Enzyme Mix were added. The final mixture was incubated under conditions found in Table 2.4. The cDNA generated from GPA-SUDV RNA samples were quantitated using Qubit™ dsDNA HS Assay Kit and Qubit 4.0 fluorometer device (Thermo Fisher Scientific) as per manufacturer's guidelines. For each sample, the reaction mixture included 198 µl of working solution and 2 µl of DNA added in a 0.5 ml thin-walled PCR tube. Amplification of housekeeping genes was tested in all GPA-SUDV RNA samples. BHK cellular RNA was used as an internal control during amplification of β-actin and GAPDH, while MDCK cellular RNA for amplification of PBGD. A 25 µl PCR reaction mixture was prepared as per Table 2.12 for each sample. The PCR reaction was carried out using CloneAmp HiFi Polymerase (Takara Bio) under the parameters enlisted in Table 2.13.

Table 2.11. Reverse transcription mixture for synthesizing cDNA from cellular RNA and GPA-SUDV RNA

Components	Amount (µl)
Annealing Buffer	2
Random hexamers (50 ng/µl)	1
RNase/DNase Free Water	0.4
RNA	4
2X First-Strand Reaction Mix	20
SuperScript® III/RNaseOUT Enzyme Mix	4

Table 2.12. PCR reaction components for amplification of housekeeping genes in GPA-SUDV RNA samples.

Components	GPA-SUDV RNA	Control RNA
CloneAmp HiFi Premix (2X)	12.5 µl	12.5 µl
Forward primer (10 µM)	1 µl	1 µl
Reverse primer (10 µM)	1 µl	1 µl
cDNA template	50 ng	6.5 µl
DNase/RNase free water	up to 25 µl	4 µl

.

Table 2.13. PCR reaction thermal profiles for amplification of β-actin, GAPDH, and PBGD housekeeping genes in GPA-SUDV samples.

Steps	β-actin & GAPDH†		PBGD‡	
	Temperature	Duration	Temperature	Duration
Initial denaturation	94°C	3 minutes	94°C	4 minutes
Denaturation	94°C	5 seconds	94°C	3 minutes
Annealing	65°C	5 seconds	55°C	1 minute
Extension	72°C	10 seconds	72°C	1 minute
Final extension	72°C	5 minutes	72°C	5 minutes

†Thermal conditions as per Yamada et al. (2005)–30 cycles of denaturation, annealing, and extension.
‡Thermal conditions as per Shimizu et al. (1994)–39 cycles of denaturation, annealing and extension.

2.6 DETECTION OF VARIANTS

2.6.1 Next-generation sequencing

A total of 20 µl of extracted GPA-SUDV PCR amplicons and their sizes were submitted to an in-house sequencing facility within the Genomics Core of the National Microbiology Laboratory for next-generation sequencing. The PCR products were quantitated and amplicons from the same passage were pooled in equimolar amounts based on the concentration and size of each amplicon. Sequencing libraries from those pools were constructed using the Nextera XT Sample Prep Kit (96 samples) (Illumina) following the manufacturer's guidelines.

2.6.2 Quality of raw sequencing data

The raw reads deposited by DNA Core to Integrated Rapid Infectious Disease Analysis (IRIDA) bioinformatics platform were obtained in FastQ format. Quality control checks on each raw sequence data were conducted through FastQC available on the Galaxy Project platform.

Quality parameters inspected were the per-base sequence quality, per-sequence quality scores, and adapter content. MultiQC report was generated to summarize independent FastQC reports.

2.6.3 Detection of single nucleotide polymorphisms

The pipeline used for variant calling is depicted in Fig. 2.3. Raw sequencing data for each GPA-SUDV passage was imported to Geneious Prime software. During the import process, paired reads with 500 bp insert are set. Raw sequences were then trimmed using the BBDuk tool from the BBTools plug-in to remove any base calls with a quality score of less than 30 from both ends of a read and discard any reads that were shorter than 20 bp. Apart from passages 16 and 18, the rest of the passages were sequenced on three separate occasions. If so, trimmed sequences belonging to the same passage were grouped. The reads retained after these processes were then mapped onto the progenitor virus sequence, KT878488.1, using Geneious as the mapper with an iteration of five times. SNPs that were identified afterwards were those present at a frequency of 5% or higher and with a coverage of 1,000 reads or more.

A separate deduplication analysis was retroactively performed to determine to effects of deduplication on the frequencies of variants detected before deduplication. The raw reads were paired and deduplicated with the Dedupe tool in Geneious Prime using the default settings. After Deduplication, reads were trimmed using the same tool and parameters, grouped, and assembled in the same manner as the previous analysis.

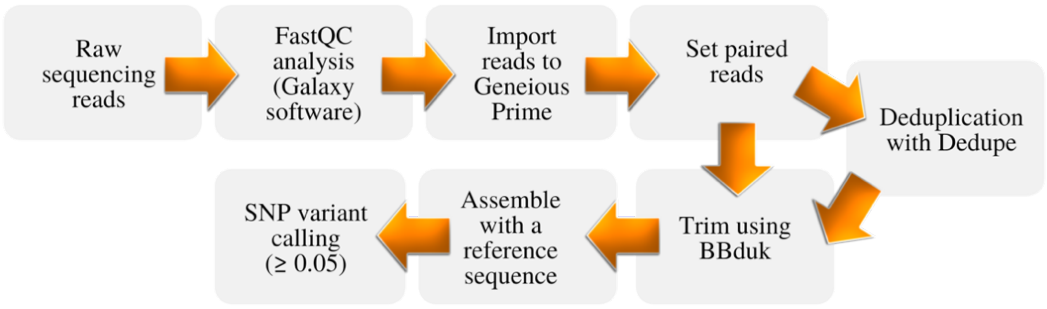


Figure 2.3 Pipeline for detection of variants in GPA-SUDV passages. Quality of raw reads are inspected using FastQC available in Galaxy. Preprocessing of raw reads were all performed via Geneious Prime software. Imported reads are paired (inward pairing) with an insert size of 500 bp. Read ends are then trimmed using BBDuk to discard low quality base calls of below 30. Short reads of 20 bp and below are also removed. A deduplication step using Dedupe may be performed to find and remove all contained and overlapping sequences in a read dataset prior to trimming and assembly. Processed and filtered reads then mapped onto a reference sequence and single nucleotide polymorphisms (SNPs) present at a frequency of ≥ 5% are detected.

3 RESULTS

3.1 PCR AMPLIFICATION

3.1.1 Viral RNA detected in pooled liver and spleen homogenates

Development of a lethal GPA-SUDV was initiated in 2016 by passaging a wild-type variant of SUDV (variant Boniface) twenty times in a guinea pig host (Wong et al., 2016). The serial passaging process involved pooling and homogenization of liver and spleen organs from two guinea pigs, followed by inoculation of a new pair of animals with the filtered supernatant from the homogenized tissues. In this study, homogenates from each of the passages were utilized to obtain viral sequencing data that will provide insights into the adaptation process of SUDV. Upon extraction of total RNA from pooled liver and spleen homogenates, the presence of viral RNA in the homogenates was confirmed by probing for the SUDV *NP* gene using quantitative RT-PCR. In the 40 AVL–extracted samples that were tested, SUDV RNA was detected in all of the 20 guinea pig pooled liver and spleen homogenate samples and their duplicates. The estimated GEQ per ml depicted in Fig. 3.1 shows that RNA levels between the technical duplicates are about the same. However, GEQ/ml between replicates for passages 9, 15-18, and 20 differed largely. The first replicates of passages 9 and 15-17 were much lower compared to the second replicate, while GEQ/ml was higher in the first replicate for passages 18 and 20. These observed differences can be a result of a random error that occurred during the handling of the sample. For example, the abundance of RNA could have varied between two aliquots made from a single homogenate sample. A low RNA yield could also stem from clogging the column with the sample. In fact, during the extraction process, the supernatant was hard to pass through the silica columns for some samples that contained small amounts of remnant tissues. The presence of these excess tissues can interfere and prevent the efficient binding of RNA to the column. Overall, the RT-qPCR data presented confirmed the presence of SUDV RNA in each GPA homogenate. These were used as a reference to select for which replicate in each passage will be used as templates for conventional PCR. For passages that did not exhibit large differences between the two technical replicates, the first replicate was used for further processing. But for passages where GEQ/ml differences between replicates were immense (i.e., passages 9, 15-18 and 20), replicate with the highest GEQ/ml were chosen as RNA templates.

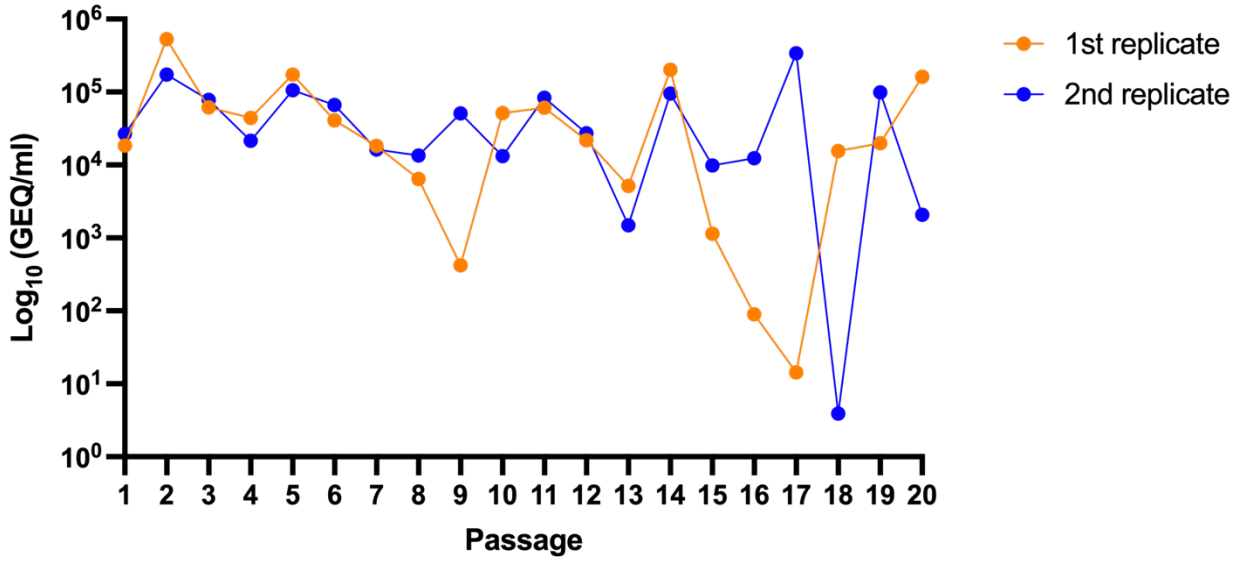


Figure 3.1 SUDV RNA genome equivalents in pooled guinea pig liver and spleen homogenates. Amounts of viral RNA in each passage were measured by quantitative RT-PCR using the SUDV *NP* gene as the target gene. The assay mixture was prepared using Light Cycler® 480 RNA Master Hydrolysis Probes Kit (Roche Life Science) with 4 µl of total RNA extracted from the GPA-SUDV homogenates as a template. The GEQ/ml of homogenized samples were calculated from a log-transformed NP standard curve (Appendix A).

3.1.2 The success of PCR amplification

Deep sequencing was performed on a stock RNA of SUDV from a clinical sample (P0) and pooled liver and spleen homogenates from each passage to track the progress of various GPA-SUDV mutations (i.e., when mutations appear and at what frequencies the mutations occur at each passage) and observe how the virus genome changed over time because of the adaptation process. To conduct next-generation sequencing on these samples, an amplicon-based approach was performed. The initial approach consisted of 16 PCR amplicons of 1-2 kb in size that spanned the 18,875 bp SUDV (variant Boniface) genome (Fig. 3.3A). The success of PCR amplification for each passage is shown in Fig. 3.2. The initial attempt was only completely successful on the stock SUDV RNA or P0. Except for amplicon #1, the remaining 15 amplicons were amplifiable in passages 2, 5, 15, 16, and 17 using the first primer set (Fig. 3.3A and Table 2.5) selected from a list of primers from Wong et al. (2016). Since the first set of primers did not work for every passage, the second set of primers was selected and put together (Fig. 3.3B and Table 2.6) from Wong's GPA-SUDV primer list (Wong et al. 2016). The location of the reverse primers for amplicons 2-4 and 6-15 was adjusted to generate longer amplicons with longer overlapping

regions, while the forward primer was adjusted for amplicon #16 to generate such amplicon. The primers for amplicon 5 were not adjusted as they worked well with all the passages. The second set of primers worked well for acquiring the missed fragments after the first PCR attempt with primer set 1 failed. Moreover, the first fragment was only amplified in all the samples when the forward primer was adjusted to start at position 205 in the genome, yielding a 1,148 bp fragment (Appendix B). The third set of primers (Fig. 3.3C and Table 2.7) was designed using Primal Scheme (Quick et al., 2017) to amplify fragments that failed to amplify after trials with the first and second of primers (Fig. 3.3). In the case of passages 8, 13, and 18, where the third primer set still didn't work very well for some amplicons, the amplicons were amplified in smaller-sized fragments ranging from 600-800 bp and this allowed for amplification of a portion of the missing fragment (Fig 3.3D and Appendix B). This third set of primers was also used to re-amplify low coverage areas in some passages, specifically those covered by amplicons 2, 3, 5, 9, 11, and 12-15 (Appendix B).

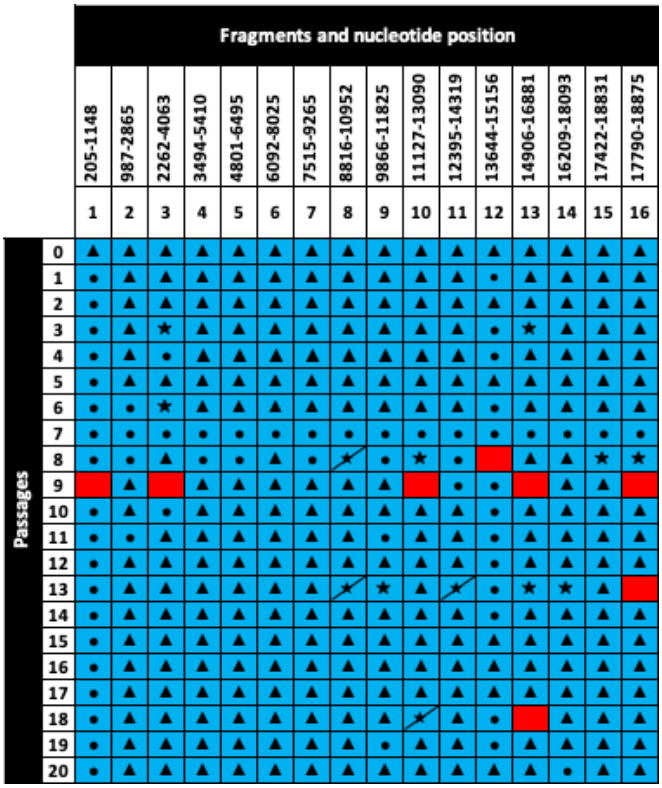


Figure 3.2 Success of PCR amplification of amplicons from the GPA-SUDV passages. A total of 16 amplicons (x-axis) spanning the whole viral genome was attempted to be amplified in all the 20 passages (y-axis). Fragments that were amplified using primer pairs from set 1 are marked with a triangle, while fragments amplified using primer pairs from a second or third set of primers are marked with a circle and star, respectively. Amplifiable amplicons in each passage are highlighted in red while missing amplicons are coloured in red. PCR amplification of smaller sized fragments ranging from 600-800 bp were attempted in some amplicons and this resulted in a few amplicons having a portion of its fragment. This partial amplification is marked by a slash symbol.

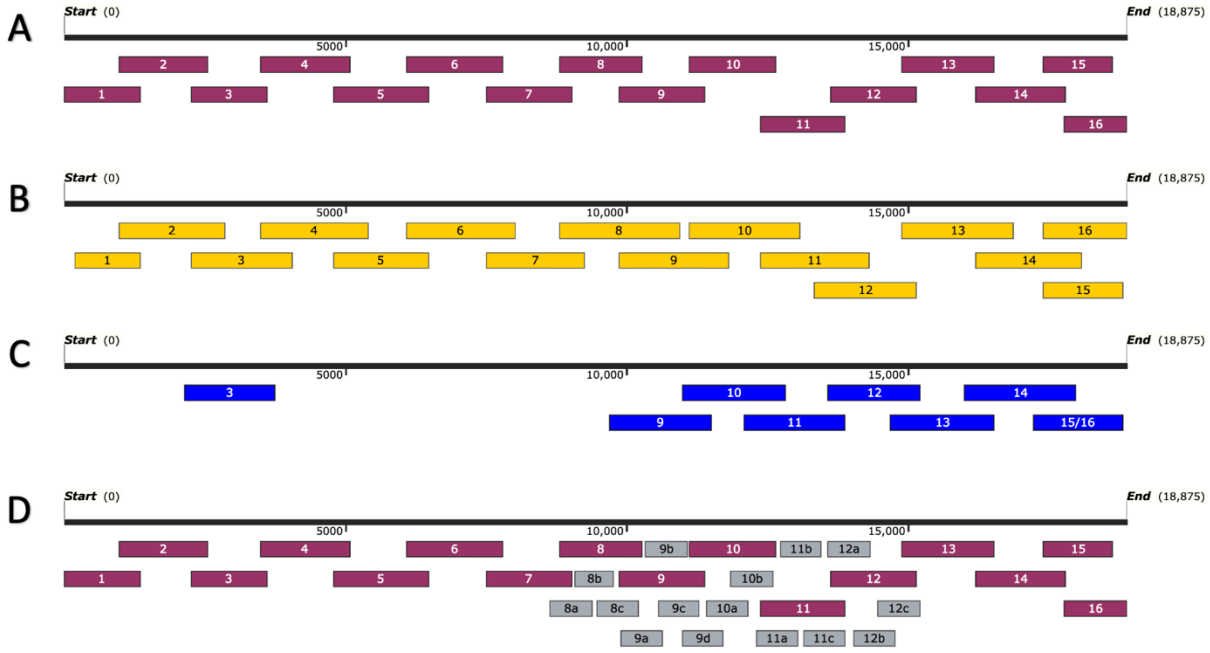


Figure 3.3 Maps of GPA-SUDV PCR amplicons spanning the SUDV genome. (A) The original PCR amplicons designed using primers listed in Table 2.5 to recover SUDV genome for sequencing. (B) A second set of PCR amplicons was designed using primers listed in Table 2.6 to recover fragments that did not amplify after the first attempt. (C) A third set of primers listed in Table 2.7 was used to amplify missing amplicons that failed to amplify after two attempts and to amplify regions of low coverage. (D) Smaller amplicons of 600 to 800 bp mapped against the original set of amplicons. Amplification of amplicons 8-12 was further divided to smaller fragments using primers listed in Table 2.7 to recover at least a portion of the fragment.

3.1.3 RNA quality assessment

The difficulty in amplifying several of the fragments from passages like 8, 9, 16, and 18 was hypothesized to be due to poor quality RNA, more specifically a degraded RNA. To test this hypothesis, an RNA bioanalyzer assay was conducted on the RNA extracted from each guinea pig

passage to determine the quality of RNA being used. The bioanalyzer does not directly verify the quality of the viral RNA, but rather provides an insight into the overall profile of the total RNA. In the case of an intact RNA, visual inspection of the electropherogram will show defined 18S and 28S eukaryotic ribosomal RNA (rRNA) peaks between migration times of 39 to 50 seconds, with the intensity of the 28S rRNA peak being twice as that of the 18S rRNA peak. Meanwhile, a degraded RNA profile is depicted by a decrease in the signal intensities of the rRNA peaks with a shift in signal towards shorter RNA fragments that elute at earlier time points. The bioanalyzer software is also equipped with an algorithm that analyzes the other electrophoretic trace features of an RNA sample besides the rRNA peaks, such as the pre-, 5S-, fast-, inter-, precursor-, and post-regions. The algorithm generates a value from 1 through 10 to indicate the quality of RNA, with 10 being a highly intact RNA and 1 as a highly degraded RNA.

All the GPA-SUDV RNA samples showed highly degraded RNA profiles (Fig. 3.4) First, the bioanalyzer electropherograms showed no distinct RNA peaks between 39-50 seconds, which corresponds to 18S and 28S rRNA. A highly degraded RNA is also depicted in the electrophoresis gel images as indicated by the presence of several lower molecular weight RNA bands compared to higher molecular weight bands. The calculated RIN for all the samples was also below 3 (Fig. 3.4B-4T, 4V) and further suggests a strongly degraded RNA. On the other hand, the RIN for passage 19 was undetermined (Fig. 3.4U). RIN scores may not be attainable in circumstances where anomalies are present in the RNA electrophoretic trace. If RNA is immensely degraded, failure to properly identify different regions may occur and an anomaly might be reported which results in no computed RIN value. Interestingly, the RNA profiles of samples that were more successful during PCR amplification attempts were different from the samples that lacked several PCR fragments despite multiple attempts to amplify with different primer pairs. Passages 8, 9, 16, 17, and 18 show that fragments of larger size (i.e., 1-2 kb) were barely present as indicated by the lack of peaks or fluorescent signal detected in the 35-45 seconds region (Fig. 3.4I-J, N, and S-T). It was also noted that the fluorescent signal in these passages was more concentrated at the earlier elution times, indicating most of the fragments were between 200-500 nt. The absence of high molecular weight bands in the gel image also reveals the same conclusions.

It was further postulated that the degradation of RNA may have been due to the old age of the samples. Although the homogenized tissue samples were stored at -80 ^{o}C from the time of collection, degradation may have occurred over the years from 2015/2016 to 2019. Therefore, a

more recent RNA sample from passaging of Bundibugyo virus (BDBV) in guinea pigs, which was extracted in the same manner as the GPA-SUDV samples, was analyzed. As per Fig. 3.5B, the fresher BDBV RNA sample appeared to have the same profile as the GPA-SUDV samples in Fig. 3.4, with fewer high molecular weight RNA fragments and more low molecular weight RNA fragments. Overall, the results showed that regardless of the age of the samples, both the old and the new RNA samples were heavily degraded, and that age of the sample is likely not a contributing factor to degradation observed. The silica-based column extraction method used was also hypothesized to pose a tendency for degradation as RNA becomes prone to mechanical shearing as it passes through the column. To alleviate this issue, harder to amplify GPA-SUDV passages (i.e., 3, 8, 9, 16, and 18) were extracted using a bead-based approach. To examine if different extraction methods do affect the integrity of RNA, the GPA-SUDV samples extracted using magnetic beads were also included in the analyses. The presence of smaller fragments was still prominent in the bead-based extracted RNA samples; however, it was evident that there was an increase in longer RNA fragments after the magnetic bead extraction (Fig. 3.4 C-G). This result confirms that the magnetic bead extraction does lessen RNA degradation but does not completely resolve the degradation issue on the GPA-SUDV samples. The use of RNA samples extracted with a bead-based approach did ease the amplification of some but not all amplicons for passages 3, 8, 9, 16, and 18. This success may be attributed to the increase of longer RNA fragments, which subsequently allows for the generation of a more complete cDNA template for PCR amplification.

PCR amplification of housekeeping genes, including β-actin, GAPDH and PBGD, was performed as an additional analysis to assess the RNA integrity of the GPA-SUDV samples. The rationale is that other RNA present in the samples, like RNA of these housekeeping genes, should also not be amplifiable if the integrity of the extracted total RNA was indeed compromised. Passages 9 and 16 were omitted from PCR amplification as the concentration of cDNA was less than 2 ng and was therefore too low of a concentration to work with. Amplification of these three housekeeping genes was attempted in GPA-SUDV samples. Cell lines where these genes are also expressed, including BHK and MDCK cells, were used as controls. Except for the PBGD housekeeping gene (152 bp) (Fig. 3.6C), the β-actin (279 bp) (Fig. 3.6A) and GAPDH (452 bp) (Fig. 3.6B) were amplifiable in the cell lines that were used as controls. However, all three housekeeping genes were not amplifiable in any of the GPA-SUDV passages (Fig. 3.6 A and B). The failure to amplify other RNAs present in the extracted total RNA, such as these housekeeping

genes, further attest to the poor-quality state of the GPA-SUDV samples and compliments the former bioanalyzer assay results.

Overall, SUDV RNA was detectable in all the RNA samples that were extracted from the pooled guinea pig liver and spleen homogenate samples. For each passage, the extracted RNA was further used as a template to generate cDNA, from which amplification of 16 amplicons spanning the whole viral genome was attempted. The success of PCR amplification varied between passages – some passages only needed one set of primer pairs to amplify the 16 fragments, some required the use of primers from as second set to fill in the missing fragments, while others still failed to generate amplicons despite the use of three different primer sets. To investigate the source of this variation, the quality of extracted RNA was assessed using a bioanalyzer. The RNA bioanalyzer results revealed that the starting materials used were highly degraded, but even though all the RNA samples were degraded, the extent of RNA degradation differed between passages (Fig. 3.4) – RNA samples from which all the 16 amplicons were obtained had longer RNA fragments, while cumbersome RNA samples had little to no presence of larger RNA molecules and consisted of mostly short RNA fragments. The poor quality of RNA was verified further through the amplification of multiple housekeeping genes, which all failed to amplify in the GPA-SUDV samples (Fig. 3.6). The possible underlying cause(s) of RNA degradation was also determined. The comparison of RNA bioanalyzer profiles of samples collected about five years ago and a freshly extracted RNA suggest that the age of the samples is most likely not a contributing factor to RNA degradation as both old and new RNA samples were heavily degraded. Observations made from the RNA profiles of samples extracted using magnetic beads (Fig 3.6 C-G) and samples extracted using a silica column (Fig. 3.4) suggest that the silica-based column extraction can additionally fragment the extracted RNA. As a collective, since both the fresh and old RNA samples had similar RNA profiles and the bead-based extraction did not seem to eliminate the presence of degraded RNA fragments, the observed RNA degradation is probably inherent in the sample itself.

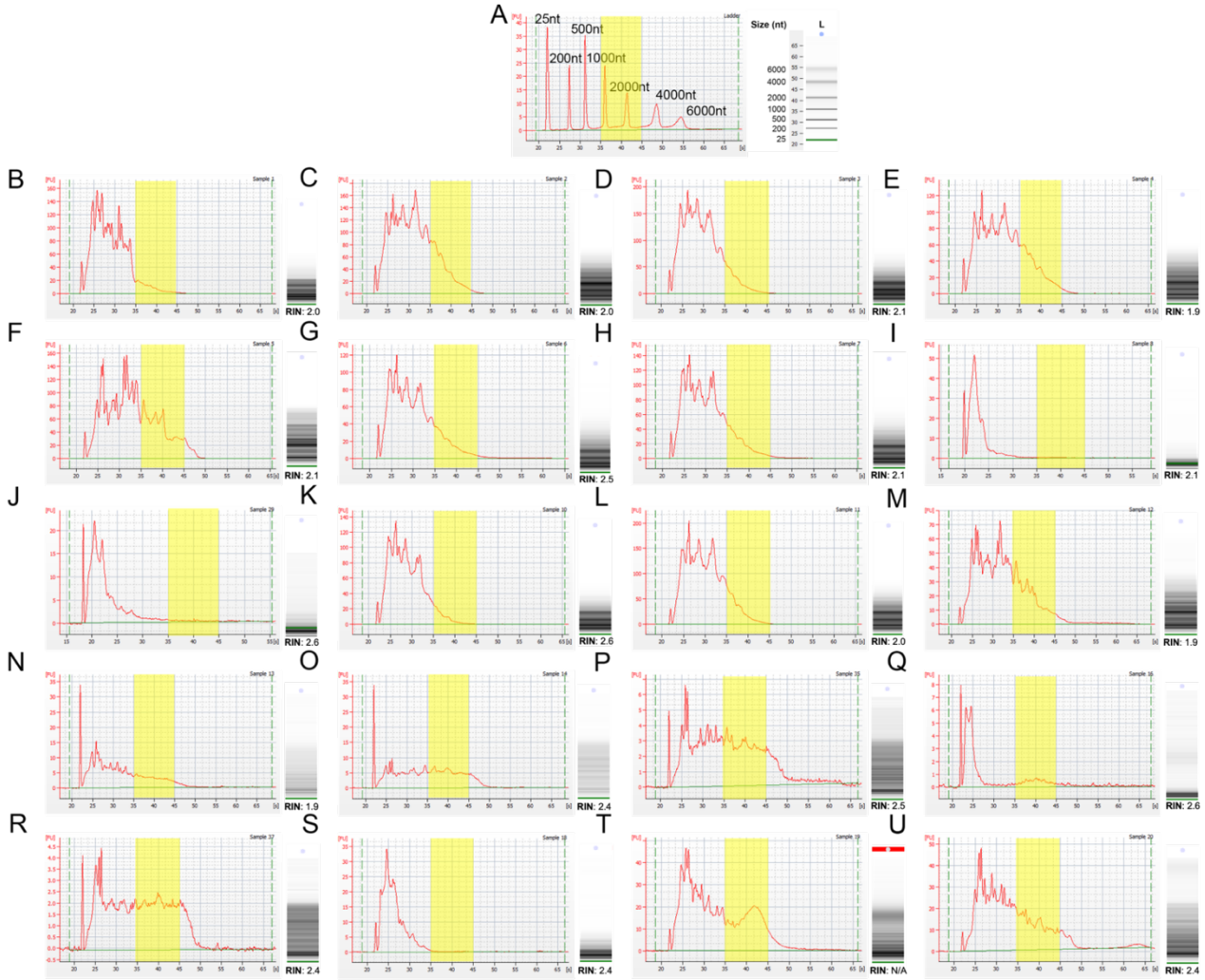


Figure 3.4 Bioanalyzer RNA profiles of total RNA extracted from homogenized guinea pig liver and spleen. (A) Defined sets of RNA fragments were used as ladders. (B-U) Electropherograms of RNA from passages 1-20 that was extracted using QIAGEN's Viral RNA Mini Kit show highly degraded RNA. Assay was conducted using RNA Pico 6000 chips and Agilent 2100 bioanalyzer instrument using Eukaryote Total RNA Pico High Sensitivity Assay. Retention times of fragments separated by size are shown on the x-axis, while arbitrary fluorescence units (FU) are depicted on the y-axis.

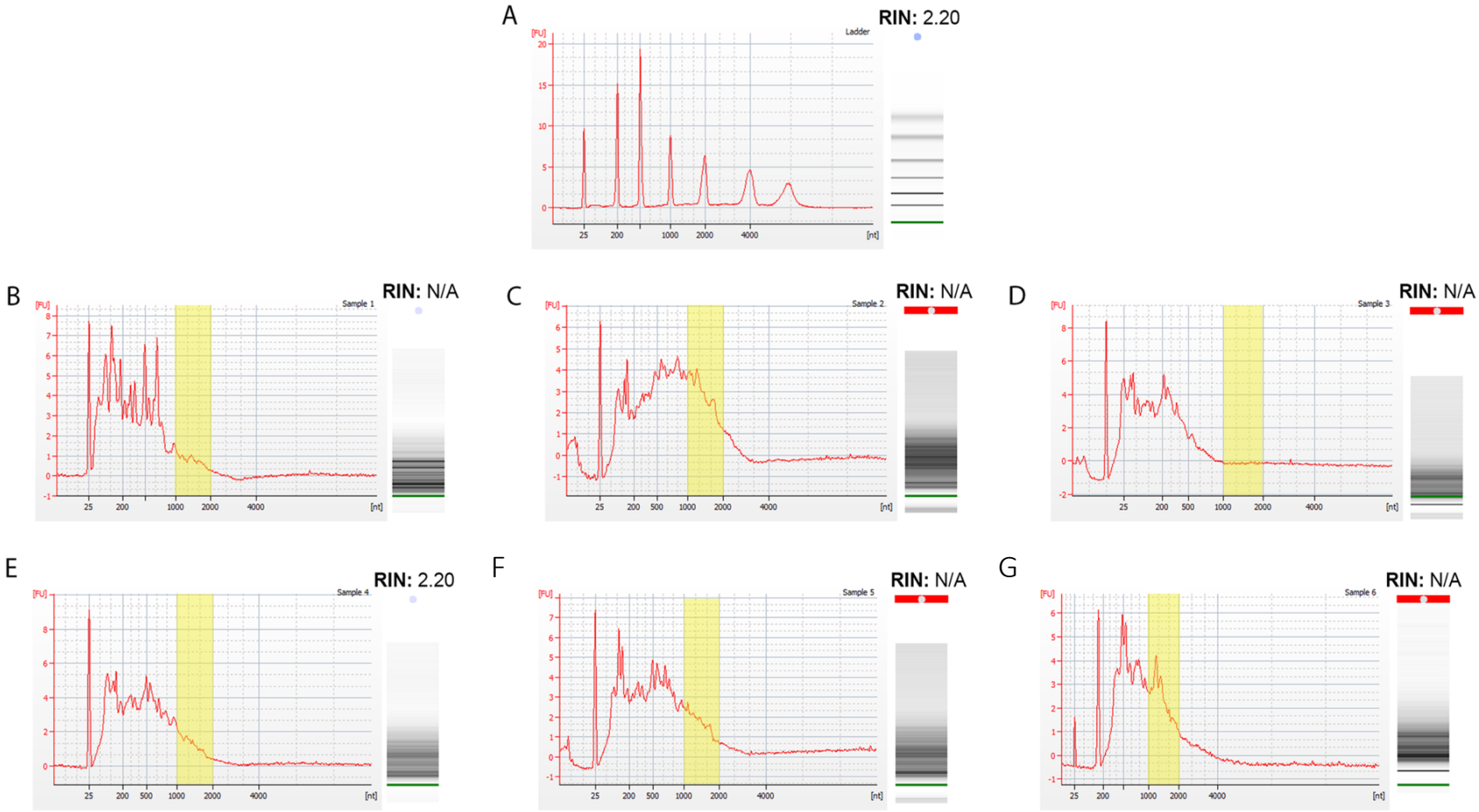


Figure 3.5 Bioanalyzer RNA profiles of freshly extracted total RNA and bead-based extracted RNA. (A) Defined sets of RNA fragments were used as ladders. (B) Electropherogram and gel image of a less than 1 year old total RNA from a GPA-BDBV homogenate that was extracted using QIAGEN's Viral RNA Mini Kit. (C-F) Electropherograms of passages 3, 8, 9, 16, and 18 that were extracted with magnetic beads from the Applied Biosystems™ MagMAX™ Viral/Pathogen II (MVP II) Nucleic Acid Isolation Kit. Assay was conducted using RNA Pico 6000 chips and Agilent 2100 bioanalyzer instrument using Eukaryote Total RNA Pico High Sensitivity Assay. Retention times of fragments separated by size are shown on the x-axis, while arbitrary fluorescence units (FU) are depicted on the y-axis. Region of 1000 – 2000 nucleotides is highlighted with yellow.

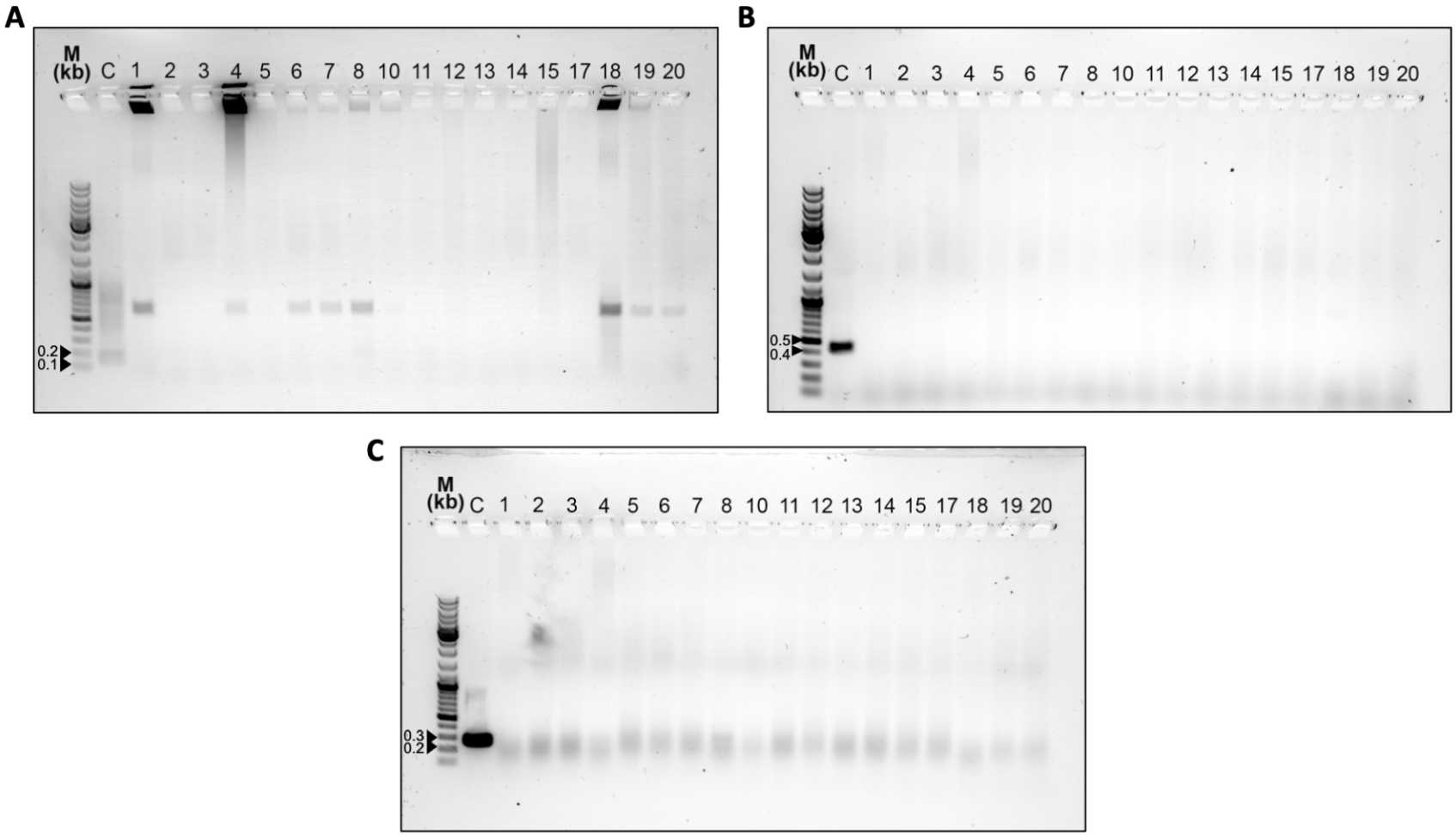
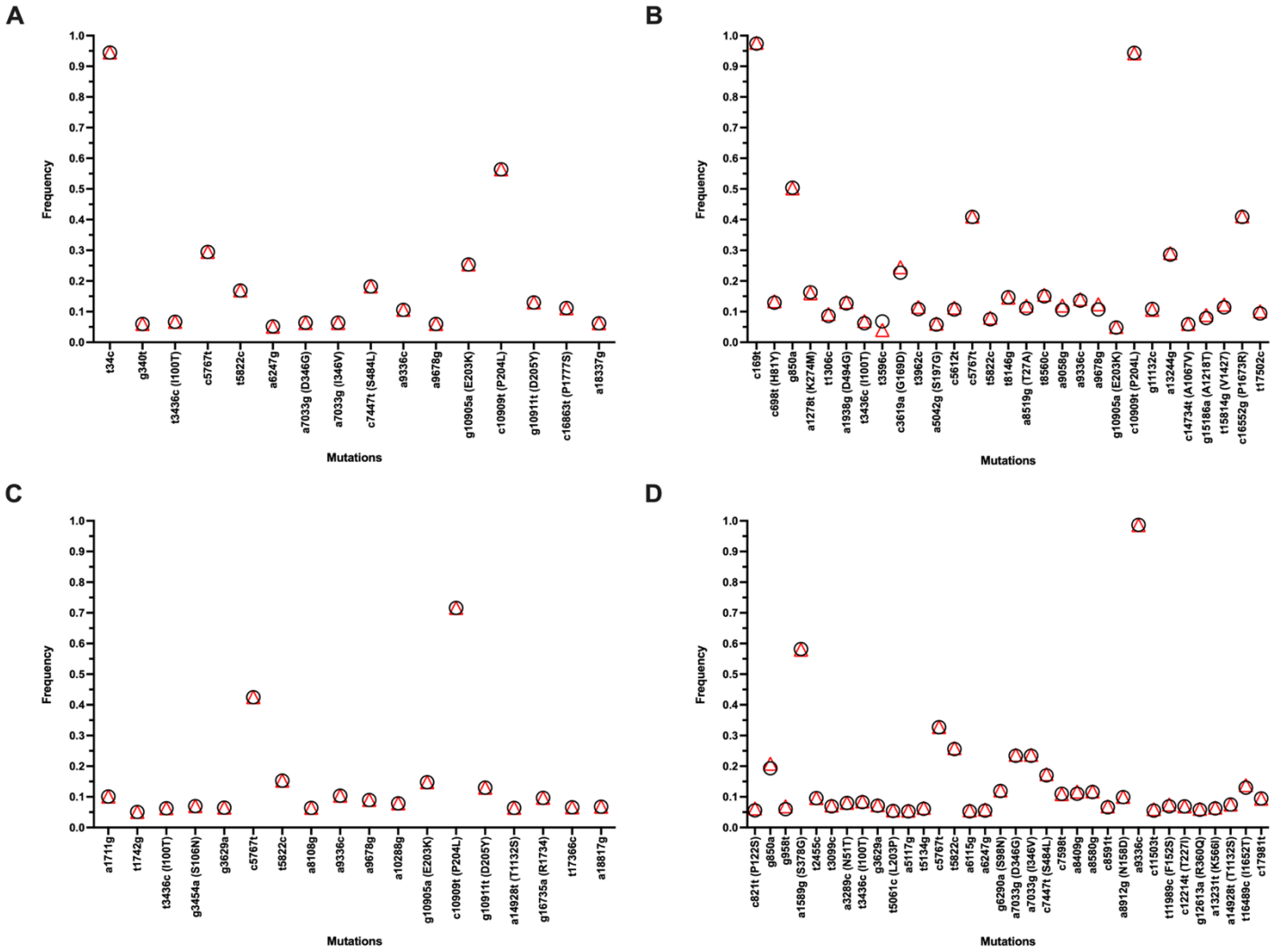


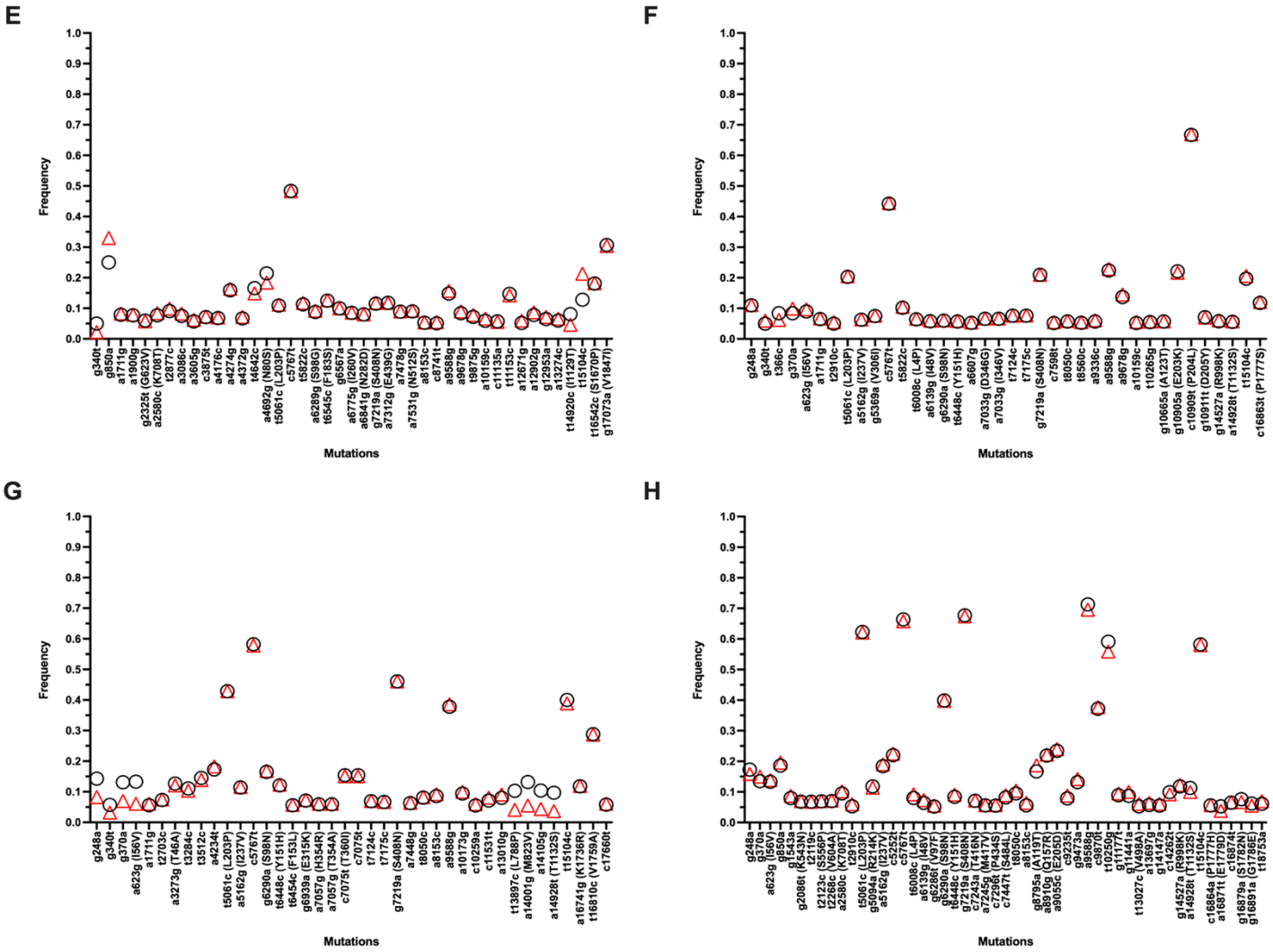
Figure 3.6 PCR amplification of multiple housekeeping genes in GPA-SUDV passages. (A) PBGD, (B) GAPDH, and (C) β-actin genes not amplifiable in the GPA-SUDV samples. The PBGD, GAPDH and β-actin were all amplifiable in the control cell lines as indicated by the presence of 152 bp, 452 bp and 279 bp fragments, respectively, in the control lane (C). Control cell lines used for PBGD amplification was from MDCK cells, while controls used for GAPDH and β-actin amplification was from BHK cells. PCR products were run in 1% agarose gel in 1X TAE buffer for 30 minutes. The sizes of fragments were estimated using GeneRuler 1 kb ladder (Lane M) run in parallel with the samples.

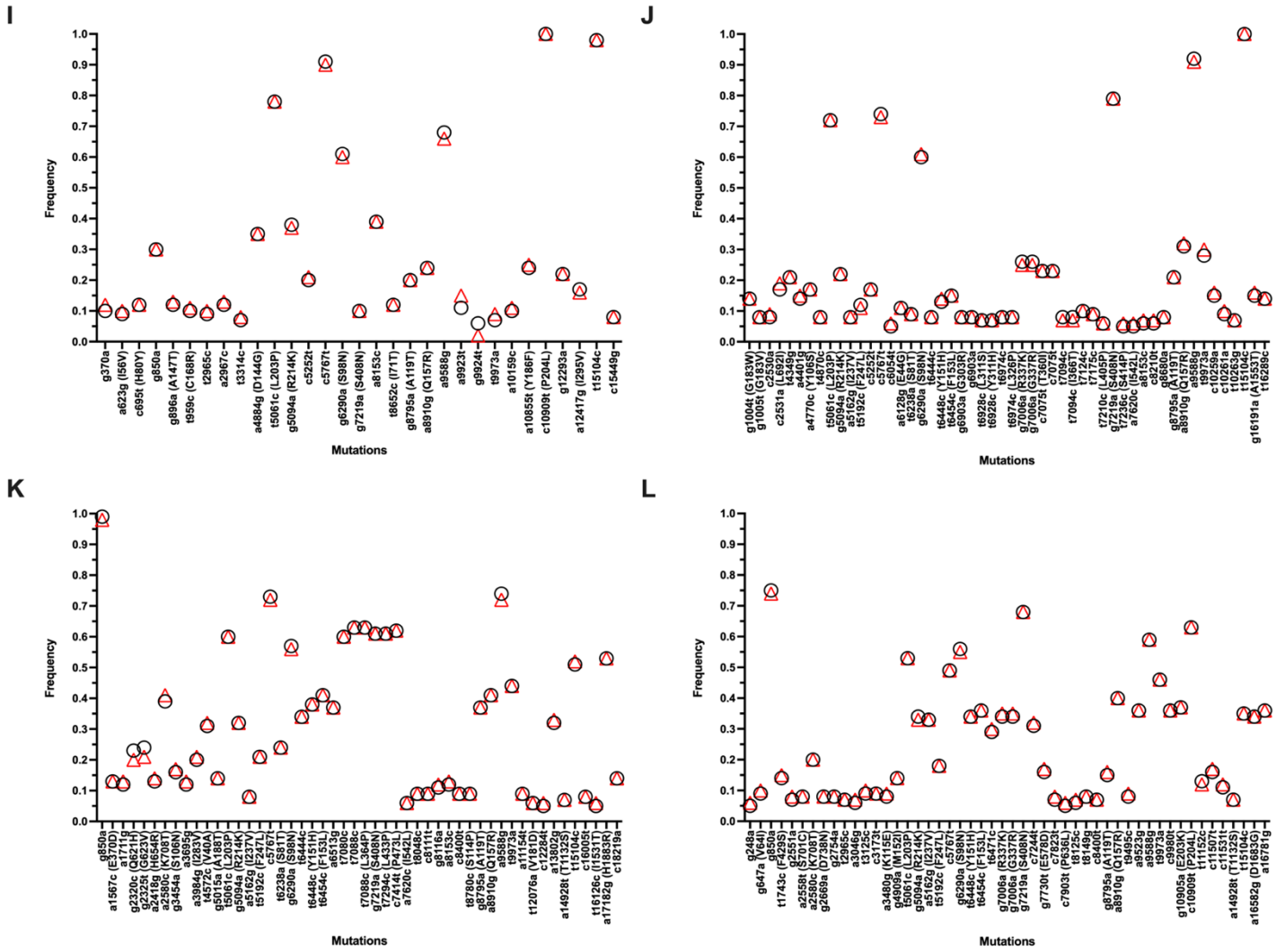
3.2 EFFECT OF DEDUPLICATION ON SEQUENCING READS

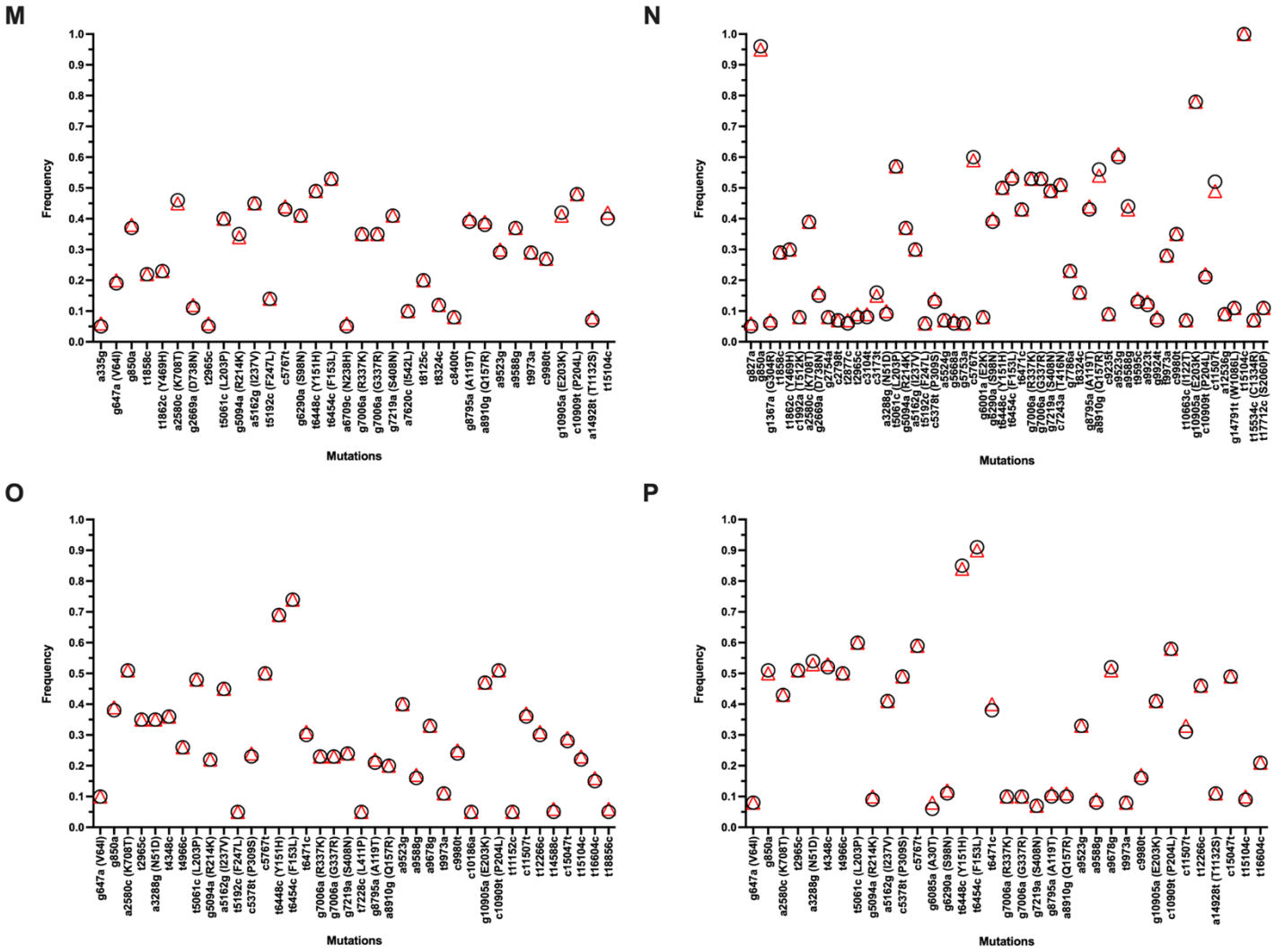
During next-generation sequencing a phenomenon occurs where numerous copies of DNA fragments that arose from a single DNA molecule bind to multiple sequencing flow cells and are sequenced, and as result, identical reads known as PCR duplicates are generated (Ebbert et al., 2016). As a result of this phenomenon, some researchers involve the removal of duplicated reads (i.e., deduplication) to avoid the possibility of false positives during variant calling. For this project, the original pipeline did not involve a deduplication step. To determine if and how this additional step would change the number of variants identified and the frequencies of each mutation in each passage, the Dedupe tool available in the Geneious Prime bioinformatics software was retroactively applied on the initial sequencing results. Fig. 3.7 shows that the following deduplication, the number of detected mutations and frequencies of variants did not change drastically for almost all the passages.

For passages, 0-3 (Fig. 3.7 A-D), 5 (Fig. 3.7F), 7 (Fig. 3.7H), and 9-20 (Fig. 3.7 J-U) the same number of mutations that were detected and variant frequency did not change at all or only changed by 0.03 at maximum after the removal of duplicates. Meanwhile, passages 4, 6 and 8 surprisingly had a slightly different result. In all three cases, deduplication resulted in the loss of variants as frequencies decreased below the detection limit of 0.05 or 5%. For example, at passage 4 (Fig. 3.7E), frequencies of mutations g340t and t14920c (I11129T) decreased from 0.05 to 0.02 and 0.08 to 0.04, respectively, after deduplication. Some mutations in passage 4 also experienced an increase in frequency after deduplication, including g850a and t15104c, which both increased by 0.08 or 8%. On the other hand, in passage 8, the frequency of the synonymous g9924t mutation changed from 0.06 to 0.02 (Fig. 3.7I). Passage 6 had the highest number of mutations that experienced frequency changes of 0.05 or higher (Fig. 3.7G). Mutations that fell below the detection limit after deduplication was the non-synonymous mutations t13897c (L788P) and a14928c (T1132S), which both decreased in frequency from 0.10 to 0.04. Overall, even though the deduplication step decreased and increased the frequencies of some mutations at these three passages, many of the mutations from these passages remained relatively the same in terms of frequencies and a deduplication step seemed to be unnecessary.


A
B
C
D
Frequency
Mutations
t34c
g340t
t3436c (I100T)
c5767t
t5822c
a6247g
a7033g (D346G)
a7033g (I346V)
c7447t (S484L)
a9336c
a9678g
g10905a (E203K)
c10909t (P204L)
g10911t (D205Y)
c16863t (P1777S)
a18337g
c169t
c698t (H81Y)
g850a
a1278t (K274M)
t1306c
a1938g (D494G)
t3436c (I100T)
t3596c
c3619a (G169D)
t3962c
a5042g (S197G)
c5612t
c5767t
t5822c
t8146g
a8519g (T27A)
t8560c
a9058g
a9336c
a9678g
g10905a (E203K)
c10909t (P204L)
g11132c
a13244g
c14734t (A1067V)
g15186a (A1218T)
t15814g (V1427)
c16552g (P1673R)
t17502c
a1711g
t1742g
t3436c (I100T)
g3454a (S106N)
g3629a
c5767t
t5822c
a8108g
a9336c
a9678g
a10288g
g10905a (E203K)
c10909t (P204L)
g10911t (D205Y)
a14928t (T1132S)
g16735a (R1734)
t17366c
a18817g
c821t (P122S)
g850a
g958t
a1589g (S378G)
t2455c
t3099c
a3289c (N51T)
t3436c (I100T)
g3629a
t5061c (L203P)
a5117g
t5134g
c5767t
t5822c
a6115g
a6247g
g6290a (S98N)
a7033g (D346G)
a7033g (I346V)
c7447t (S484L)
c7598t
a8409g
a8580g
c8591t
a8912g (N158D)
a9336c
c11503t
t11989c (F152S)
c12214t (T227I)
g12613a (R360Q)
a13231t (K566I)
a14928t (T1132S)
t16489c (I1652T)
c17981t


E
F
G
H
Frequency
Mutations


I
J
K
L
Frequency
Mutations


M
N
O
P
Frequency
Mutations

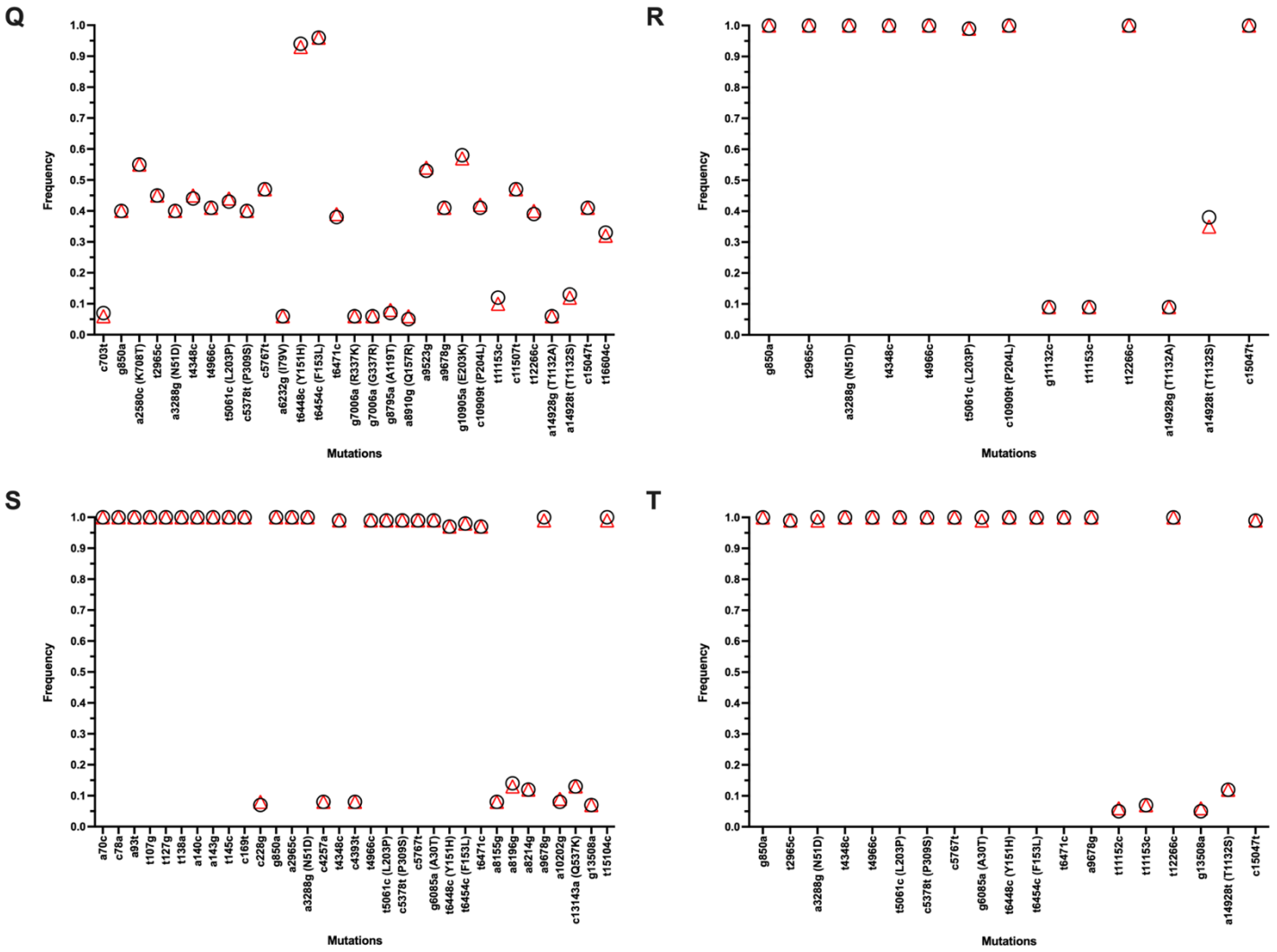


Figure 3.7 Changes in variant frequencies after deduplication. Figures A-U show minor differences in variant frequencies before (black circle) and after deduplication of sequencing reads from passages 0 to 20, respectively. The x-axis shows all the mutations found in each passage at frequencies ≥ 5% and had coverage of ≥ 1,000 reads, while the y-axis represents the frequencies of each mutation.

3.3 GENOMIC MUTATIONS ACQUIRED BY SUDV DURING ADAPTATION IN GUINEA PIGS

3.3.1 Sequencing depth

The collected amplicons from each GPA-SUDV passage were submitted for next-generation sequencing to an in-house sequencing facility, DNA Core, at the NML. A volume of 20 µl was provided to the sequencing facility for each of the PCR products along with the concentration and size of the amplicons, as well as information on the pooling of the amplicons. The DNA Core facility was involved in the library preparation and sequencing of libraries using a MiSeq sequencer. The MiSeq sequencing coverage, which is the number of reads that mapped to the reference sequence (Illumina, 2022b; Sims et al., 2014), for each of the GPA-SUDV and the unpassaged virus is shown in Fig. 3.8 The median coverage was 42,027 reads (IQR: 54,159; range: 0-539,250). Depending on the passage, 70-99% of the genome had coverage of 1,000 reads or more. Passages 0-2, 4-7, 10-12, 14-16, and 19-20 had more than 90% of the genome covered by > 1,000 reads while 83-89% of the viral genome from passages 3, 8, 13, and 18 were covered. Contrarily, passages 9 and 17 had the lowest coverage with only 71-73% of the genome being covered by at least 1,000 reads.

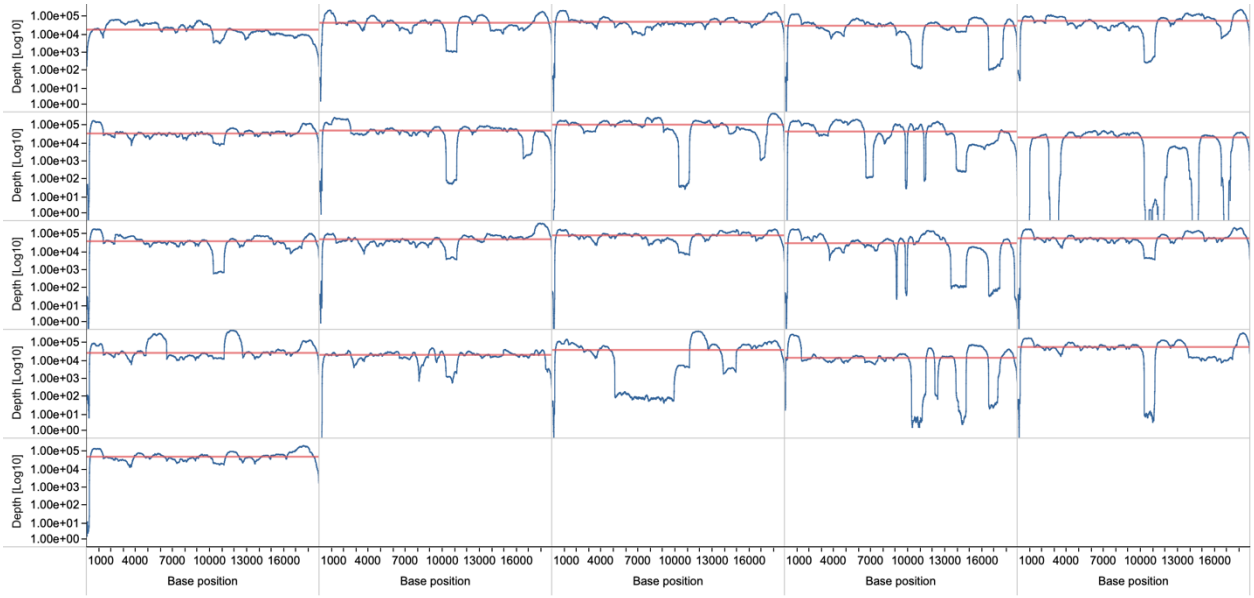


Figure 3.8 Sequencing coverage of each GPA-SUDV passage. MiSeq sequencing coverage (y-axis) at each nucleotide position of the genome (x-axis) from the unpassaged virus (P0) (row 1, column 1) to the final GPA-SUDV (P20) (row 5, column 1). The median coverage for each of

the passages is depicted by a red line. Coverages for passages 0 to 20 are depicted from left to right.

3.3.2 Changes in the genome of SUDV during serial passaging

The variant calling analysis pipeline established for this project began with the import of raw sequencing reads to Geneious bioinformatics software, where reads were trimmed, filtered, and mapped to the SUDV (variant Boniface) genome. Single nucleotide polymorphisms (SNP) that were identified were those existing at a minimum frequency of 0.05 (5%) and had a coverage of $\geq$ 1,000 reads. A total of 333 point mutations having coverage of 1,000 reads or higher and were present at a frequency of 5% or higher were identified between the non-adapted SUDV and the 20 GPA SUDV passages – 78% of these mutations were transitions, a pyrimidine to pyrimidine or purine to purine base change, and 22% were transversions, which happens when a pyrimidine base is substituted with a purine base (or vice versa). Despite the large number of mutations found throughout the entire adaptation series, most were only transient, with some mutations appearing only once throughout the passaging and others appearing over multiple passages. Mutations appearing only once are likely to be a result of the error-prone RNA polymerase, while short-lived mutations appearing multiple times during the passaging series may be of significance to the adaptation process of virus in guinea pigs. The final GPA-SUDV genome (i.e., passage 20), only retained 18 mutations (Fig. 3.9) – 5 were in the noncoding regions and 13 were in the coding sequence (CDS) regions. No insertions or deletions were detected to be present in frequencies above 5% in the final adapted virus. Mutations identified in these regions and how their frequencies changed over time are described below.

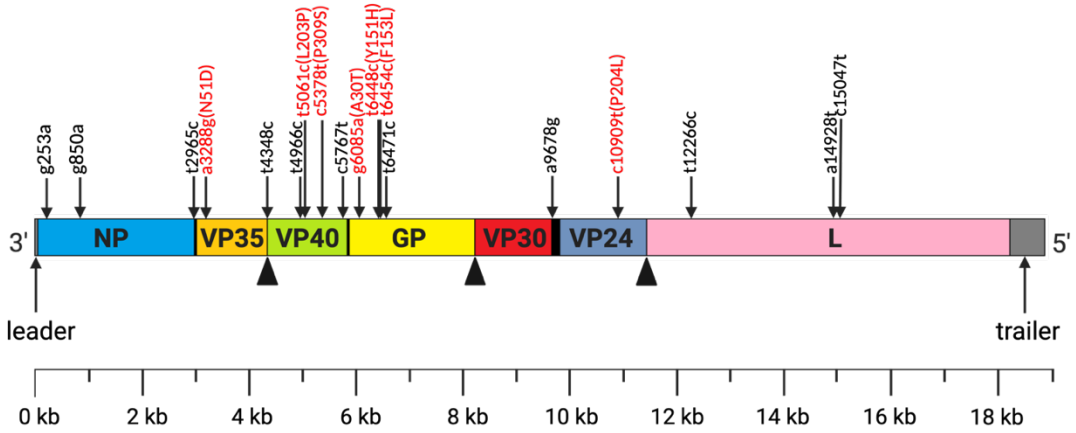


Figure 3.9 Mutations identified in the final GPA-SUDV. A total of 18 mutations were detected in the viral genome after 20 passages in the guinea pig host – 13 were found in the coding regions and 8 were found in the noncoding regions, particularly at the 5' and 3' untranslated regions. Nucleotide substitutions that resulted in an amino acid change are depicted in red along with the corresponding amino acid change in parenthesis.

3.3.2.1 Mutations in the glycoprotein

Most of the mutations were in the coding sequencing region of the genome. Overall, the *GP* gene had the greatest number of mutations identified throughout the passaging series. Out of the 68 mutations, 45 resulted in an amino acid change (nonsynonymous) while 22 did not (synonymous). However, only four of the mutations were found in the final passage: g6058a (A30T), t6448c (Y151H), t6454c (F153L), and t6471c (Fig. 3.10D). Nonsynonymous mutations t6448c (Y151H) and t6454c (F153L) were detected early at passages 5 and 6, respectively. The trend in the change in frequencies of these two mutations over time also appeared to be like one another as they both started at frequencies just above 5% and increased in frequencies at the same time points. For example, both mutations reached frequencies of ~ 40% by passage 11 and another increase to near 50% by passage 12. Both mutations finally reached a frequency of above 90% by passage 16 and eventually to 99% at passage 19. Another nonsynonymous mutation, g6058a (A30T), appeared much later at passage 15 at a frequency of 6.4%. But surprisingly it did not take long for this mutation to increase and achieve the same level of frequency as the other mutations as it reached a frequency of 99% by passage 18. The synonymous mutation t6471c was also detected later in the passaging series at passage 12 with a frequency of 29.4%. The frequency of

this mutation did not fluctuate much (i.e., 30-40%) between passages 13 -16. Although, like the g6058a mutation, t6471c jumped to 99% in frequency after two more passages at passage 18.

3.3.2.2 Mutations in the RNA polymerase

The L CDS region had the second-highest number of mutations, a total of 66 mutations, over 20 passages. Unlike GP, however, most of the mutations (40 mutations) in this region were synonymous and 26 were only nonsynonymous. From the 66 mutations detected in the L CDS over time, only three were found in the last passage – t12266c, a14928t, and c15047t (Fig. 3.10G). Both t12266c and c15047t synonymous mutations were identified at passage 14 at similar frequencies as well at around 30%. These two mutations also show an increase and decrease in frequencies at the same passages. For example, both increased to a frequency of close to 50% by the 15th passage and then decreased to very similar frequencies, at about 40%, at passage 16. The frequencies of t12266c and c15047t mutations then reached 99% at passage 18, which remained until passage 20. Synonymous mutation a14928t, on the other hand, appeared very early at the second passage but remained relatively low in frequency (< 15%). The mutation peaked in frequency at passage 17, although it was only about 40%. In the last three passages, a14928t decreased in frequency and was present in the final passage at a frequency of just about 5%. Along with the a14928t mutation, the t15104c synonymous mutation also appeared relatively early during the series at passage 4 at a frequency of 13%. Although this mutation was not observed in the final GPA-SUDV virus, the mutation progressed in an interesting manner. Upon its first detection at passage 4, the mutation increased in frequency over time until it reached a frequency of 100% by the 9th passage. The mutation was able to reach this very high frequency two more times at passages 13 and 18 but was completely lost in the last two passages. Even though this mutation was not retained by the virus until the last passage, this mutation was selected for by the virus during the adaptation process and may have contributed to the progression of disease severity in the guinea pig animal.

3.3.2.3 Mutations in the nucleoprotein

Following the *L* gene, the CDS region of the *NP* gene also had many mutations with approximately half of the mutations being nonsynonymous. However, only a single synonymous mutation, g850a, was maintained until passage 20 (Fig. 3.10A). The mutation was first observed

in the first passage at a frequency of 50% but remained below 30% after the next eight passages. At passage 10, the mutation was found to be present at a frequency of 98.6% but soon fell below 75% and 38% over passages 11 and 12, respectively. The g850a mutation peaked again with a frequency of 95.5% at passage 13 but again dipped to frequencies of $\leq$ 50%. The synonymous mutation was only able to maintain a frequency of 99% from passage 18-20.

3.3.2.4 Mutations in viral protein 35

The VP35 coding region also only had one mutation retained in the final adapted virus – mutation a3288g (N51D) (Fig. 3.10B). Several of the transient mutations were found early in the passaging series and only appeared once, while some appeared no more than three times. Moreover, these short-lived mutations were found in low frequencies of less than 25%. Contrary, the nonsynonymous a3288g (N51D) mutation was first detected at passage 13 at 9% and increased quite quickly to 54% by the 15th passage. By passage 16, the mutation decreased to a frequency of 40% but soon recovered as it reached a frequency of almost 100% in the subsequent passage (i.e., passage 17) and maintained to do so until passage 20.

3.3.2.5 Mutations in the viral protein 24

The VP24 CDS also only had one nonsynonymous mutation, c10909t (P204L), present by the end of the adaptation series and exhibited the fewest number of mutations out of all the CDS regions (Fig. 3.10F). Unique from the previously mentioned mutations, the c10909t (P204L) mutation was already present in the stock virus (passage 0) that was sequenced at 56.4%. The mutation immediately reached a frequency of 99% after the first passage but was unable to maintain such a high frequency over the next six passages. On the 8th passage, the mutation was detected again at a very high frequency at 100%; however, the frequency declined again until passage 16. From passage 17 the frequency of the mutation picked up again to 100% and declined slightly to 99.8% at the last passage. Present alongside the c10909t (P204L) mutation at P0 is the g10905a nonsynonymous mutation which results in a glutamic acid (E) to lysine (K) amino acid change at position 203. This trend in the frequency changes of this mutation over time seems to be the opposite of what was seen in c10909t (P204L). The g10905a (E203K) mutation was initially found at 25% and decreased in frequency whenever the frequency of the c10909t (P204L) mutation

increased, and vice versa. Moreover, if the frequency of c10909t (P204L) reaches above 90%, g10905a (E203K) was not detected under the set parameters for variant detection.

3.3.2.6 Mutations in the matrix viral protein 40

The final GPA-SUDV genome had a total of three mutations in the VP40 coding region – one synonymous mutation, t4966c, and two nonsynonymous mutations, t5061c (L203P) and c5378t (P309S). The c5378t (P309S) and t4966c mutations were first recognized at around the same time. The c5378t (P309S) was detected at passage 13 at 13.3%, while t4966c was at passage 14 at 26.1% (Fig. 3.10C). Starting at passage 14 both mutations increased and decreased in frequencies that were very close to each other. The only difference was t4966c reached 99% by passage 17 while this happened for the c5378t (P309S) mutation at passage 18. Both remained to be at this frequency until the very last passage. Results also show detection of a novel mutation that has not been recognized by Wong et al. (2016) in the VP40 CDS region – t5061c (L203P). This nonsynonymous mutation was first observed at passage 3 occurring at a frequency of 5%. The mutation gradually increased in frequency over time until it peaked at passage 8 at 77%. Since then, its frequency declined until passage 16, with small increases at some passages. The t5061c (L203P) only attained a frequency of 99% at passages 17 to 20.

3.3.2.7 Mutations in the viral protein 30

A total of 14 mutations was identified in the VP30 CDS. Unlike the other CDS regions, no mutation was present at a frequency of 5% or higher with 100 reads or more in the VP30 CDS region (Fig. 3.10E). Almost all the mutations were only detected once or twice; except for two nonsynonymous mutations, g8795a (A119T) and a8910a (Q157R), which were found over multiple passages. Both mutations were first identified in passage 7 at a frequency of 29% and 16%, respectively. Over time, the frequency fluctuated but both mutations never reached a frequency of ≥ 60%. After passage 13, where both mutations peaked in terms of frequency, frequencies of g8795a (A119T) and a8910a (Q157R) declined to 5% by passage 16 and were undetectable afterwards.

3.3.2.8 Mutations in the noncoding sequence region

The noncoding regions of the viral genome include the leader and trailer regions as well as the 5' and 3' untranslated regions of each gene. Dissimilar from the coding sequence region, the noncoding region, in general, did not obtain many mutations during the adaptation process (Fig. 3.11). No adaptive mutations in the leader, trailer, and the UTRs of the *GP*, *VP24*, and *L* genes were found to exist at 5% or higher with a coverage of 1,000 reads or more by the 20th passage (Fig. 3.11D, F, and H). All mutations in these regions were short-lived. The first mutations were recognized at passage 0 in the VP40 and VP30 regions. The c5767t mutation in VP40 was initially found at a frequency of 29.5% and was present throughout the entire 20 passages except at passage 17 (Fig. 3.11C). Increasing in frequency for the next seven passages, the mutation first peaked at a frequency of 91% at passage 8. It then experienced a decline in frequency for the remaining passages until passage 16. But from passages 17 to 20, the mutation was found to exist at a frequency of 99%. In contrast, the a9678g mutation in VP30 was only at 6%, much lower than the c5767t mutation in VP40 (Fig. 3.11E). Additionally, this mutation was only detected at 5% or higher in about half of the total number of passages. Results also revealed that this mutation remained quite low throughout the passaging series in frequency as even by the 16th passage, the frequency of a9678g was only at about 40%. Although it did reach 99% by passage 18.

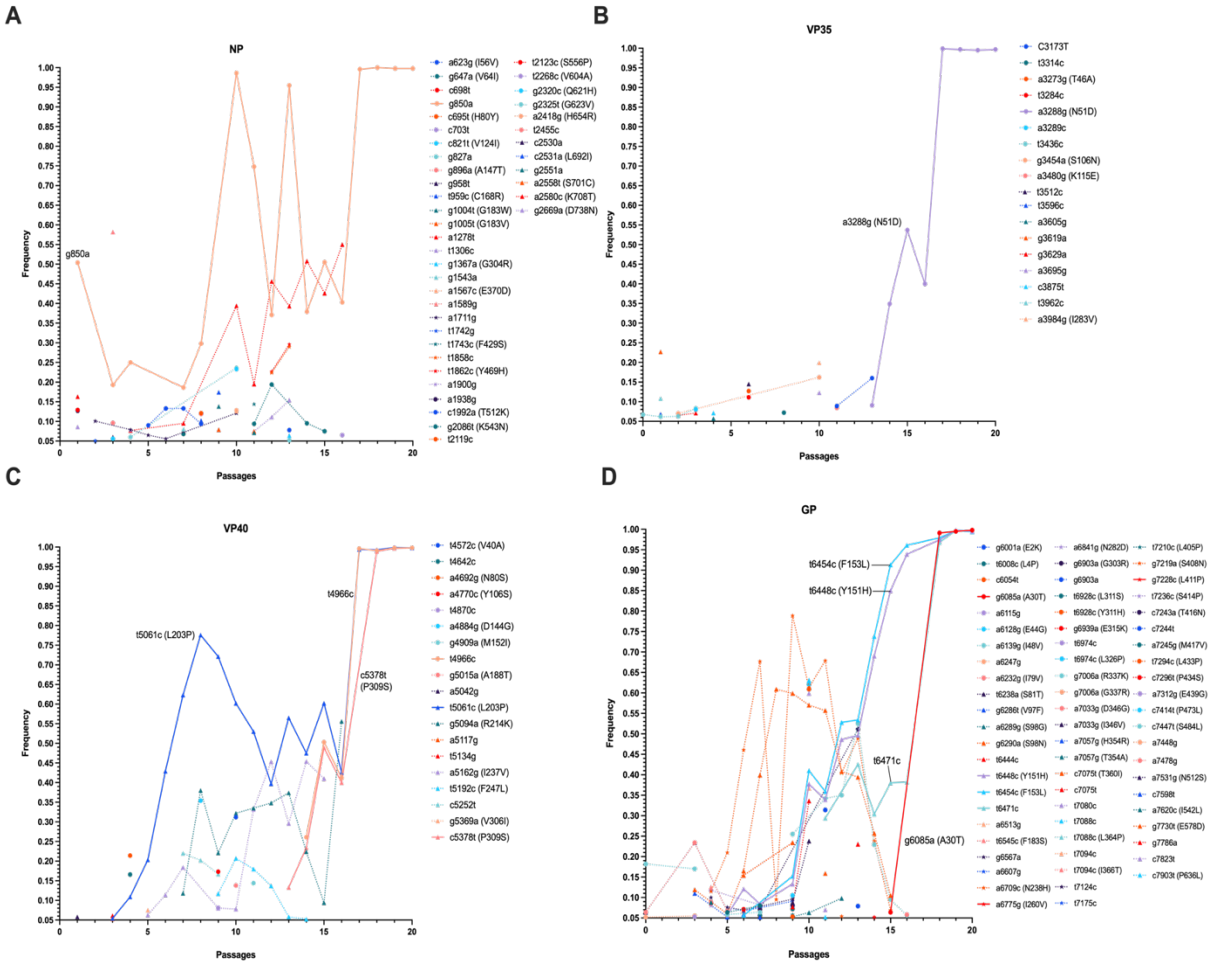
Later into passages are the occurrence of three other mutations in NP and VP35 noncoding regions (Fig. 3.11A-B). The NP noncoding region acquired two mutations by passage 20 – g235a and t2965c (Fig. 3.11A). The t965c mutation was initially observed at passage 8 at a 9% frequency. Its frequency stayed low and only fluctuated between 5-8% at passages 11-13 before it reached a frequency of ~35% at passage 14. But soon after, the frequency reached 99% from passages 17 to 20. Alongside the t965c mutation, is the g235a which only appeared at the final passage albeit at a very low frequency of 8.1% compared to t295c. Contrarily, the VP35 noncoding region only had a single mutation that was maintained until passage 20 (Fig. 3.11B). The t4349c mutation seemed to have increased in frequency quickly as it was able to attain a frequency of 99% after three additional passaging from the first time it was detected at passage 14. The mutation was at a frequency of 36% starting at passage 14 and was able to gain an additional 63% in frequency by passage 17.

3.4 SUMMARY

Overall, SUDV RNA was detectable in all the GPA-SUDV homogenate samples (Fig. 3.1). The whole SUDV genome, except for the first 205 bp, was recovered successfully in 16 out of the 20 GPA-SUDV passages, as well as in the stock virus, using an amplicon-based approach (Fig. 3.3). Although the viral genome coverage obtained was variable between passages, more than 70% of the viral genome in each passage was covered with reads greater than 1000 (Fig. 3.8). However, this approach required the use of three different primer pair sets as not all the 16 amplicons were attainable by using just a single set of 16 primers pairs. The quality of the RNA templates was investigated by conducting a bioanalyzer to identify the potential source of why PCR amplification was more successful in some passages but not in others. The bioanalyzer assay results show that all the total RNA extracted from the pooled liver and spleen homogenates were heavily degraded, but in passages where it was difficult to obtain amplicons, like passages 8, 9, and 13, long RNA fragments were almost non-existent (Fig. 3.4). This may be why it was difficult to obtain some of the 1 to 2 kb amplicons. The underlying cause(s) of degradation remains unknown but appears to be inherent of the sample source rather than due to the old age of the samples as the bioanalyzer RNA electrophoretic trace of old (Fig. 3.4) and newer (Fig. 3.5B) RNA samples both showed signs of heavy degradation. The silica-based column extraction method may act as contributing factor that further degrades the RNA since upon using a bead-based extraction method, a slight shift in the signal towards high molecular weight RNA fragments was observed (Fig. 3. 5C-F).

Previous passaging of SUDV in guinea pig host resulted in a lethal GPA-SUDV (Wong et al. 2016). Deep sequencing of the viral genome from each guinea pig passage revealed that the virus acquired hundreds of mutations between those 20 passages. Out of these hundreds of mutations, the final GPA-SUDV only retained 18 mutations, 13 of which were in the coding sequence region of the genome and 5 were in either the 5' or 3' UTR noncoding regions (Fig. 3.9). Moreover, almost all these mutations have been previously identified by Wong and his colleagues (Wong et al., 2016). The only exception was the t5061c (L203P) mutation in VP40, which is a novel discovery resulting from this exploration. Except for mutations g253a in the *NP* gene (Fig. 3.11A) and a14928t in the *L* gene (Fig. 3.10G), which both existed at a frequency of ~5% in the final passage, the other 16 guinea pig-adapted mutations existed in frequencies of 99% by the 20th passage (Fig. 3.10 and 3.11). In addition, it was observed that these 16 mutations only retained

their high frequencies of over 95% starting at either passage 17 or 18. Interestingly, three of these mutations were found to have existed in the unpassaged SUDV – two were in the noncoding sequence in VP40 (c5767t) and VP30 (a9678g), one in the coding sequence of VP24 (c10909t). The earliest mutations that appeared once the passaging was initiated were in NP, L, and VP40, with the appearance of synonymous mutation g850a at passage 1, a14928t at passage 2, and nonsynonymous mutation t5061c (L203P) at passage 3, respectively. Two nonsynonymous mutations in GP, t6448c (Y151H) and t6471c (F153L) were detected afterwards at passages 5 and 6, respectively. By the 8th passage, an NP mutation (t2965c) was located at the 3' UTR appeared. The remainder of guinea pig-adapted mutations identified in the final adapted virus appeared only appeared after the 10th passage. Overall, these 18-guinea pig adaptive mutations have been selected for by SUDV during the duration of the passaging and therefore have the potential to have implications in the ability of the virus to overcome the selective pressures existing in the guinea pig host.


A
NP
Frequency
Passages
g850a
a623g (I56V)
g647a (V64I)
c698t
g850a
c695t (H80Y)
c703t
c821t (V124I)
g827a
g896a (A147T)
g958t
t959c (C168R)
g1004t (G183W)
g1005t (G183V)
a1278t
t1306c
g1367a (G304R)
g1543a
a1567c (E370D)
a1589g
a1711g
t1742g
t1743c (F429S)
t1858c
t1862c (Y469H)
a1900g
a1938g
c1992a (T512K)
g2086t (K543N)
t2119c
t2123c (S556P)
t2268c (V604A)
g2320c (Q621H)
g2325t (G623V)
a2418g (H654R)
t2455c
c2530a
c2531a (L692I)
g2551a
a2558t (S701C)
a2580c (K708T)
g2669a (D738N)
B
VP35
Frequency
Passages
a3288g (N51D)
C3173T
t3314c
a3273g (T46A)
t3284c
a3288g (N51D)
a3289c
t3436c
g3454a (S106N)
a3480g (K115E)
t3512c
t3596c
a3605g
g3619a
g3629a
a3695g
c3875t
t3962c
a3984g (I283V)
C
VP40
Frequency
Passages
t5061c (L203P)
t4966c
c5378t (P309S)
t4572c (V40A)
t4642c
a4692g (N80S)
a4770c (Y106S)
t4870c
a4884g (D144G)
g4909a (M152I)
t4966c
g5015a (A188T)
a5042g
t5061c (L203P)
g5094a (R214K)
a5117g
t5134g
a5162g (I237V)
t5192c (F247L)
c5252t
g5369a (V306I)
c5378t (P309S)
D
GP
Frequency
Passages
t6454c (F153L)
t6448c (Y151H)
t6471c
g6085a (A30T)
g6001a (E2K)
t6008c (L4P)
c6054t
g6085a (A30T)
a6115g
a6128g (E44G)
a6139g (I48V)
a6247g
a6232g (I79V)
t6238a (S81T)
g6286t (V97F)
a6289g (S98G)
g6290a (S98N)
t6444c
t6448c (Y151H)
t6454c (F153L)
t6471c
a6513g
t6545c (F183S)
g6567a
a6607g
a6709c (N238H)
a6775g (I260V)
a6841g (N282D)
g6903a (G303R)
g6903a
t6928c (L311S)
t6928c (Y311H)
g6939a (E315K)
t6974c
t6974c (L326P)
g7006a (R337K)
g7006a (G337R)
a7033g (D346G)
a7033g (I346V)
a7057g (H354R)
a7057g (T354A)
c7075t (T360I)
c7075t
t7080c
t7088c
t7088c (L364P)
t7094c
t7094c (I366T)
t7124c
t7175c
t7210c (L405P)
g7219a (S408N)
g7228c (L411P)
t7236c (S414P)
c7243a (T416N)
c7244t
a7245g (M417V)
t7294c (L433P)
c7296t (P434S)
a7312g (E439G)
c7414t (P473L)
c7447t (S484L)
a7448g
a7478g
a7531g (N512S)
c7598t
a7620c (I542L)
g7730t (E578D)
g7786a
c7823t
c7903t (P636L)

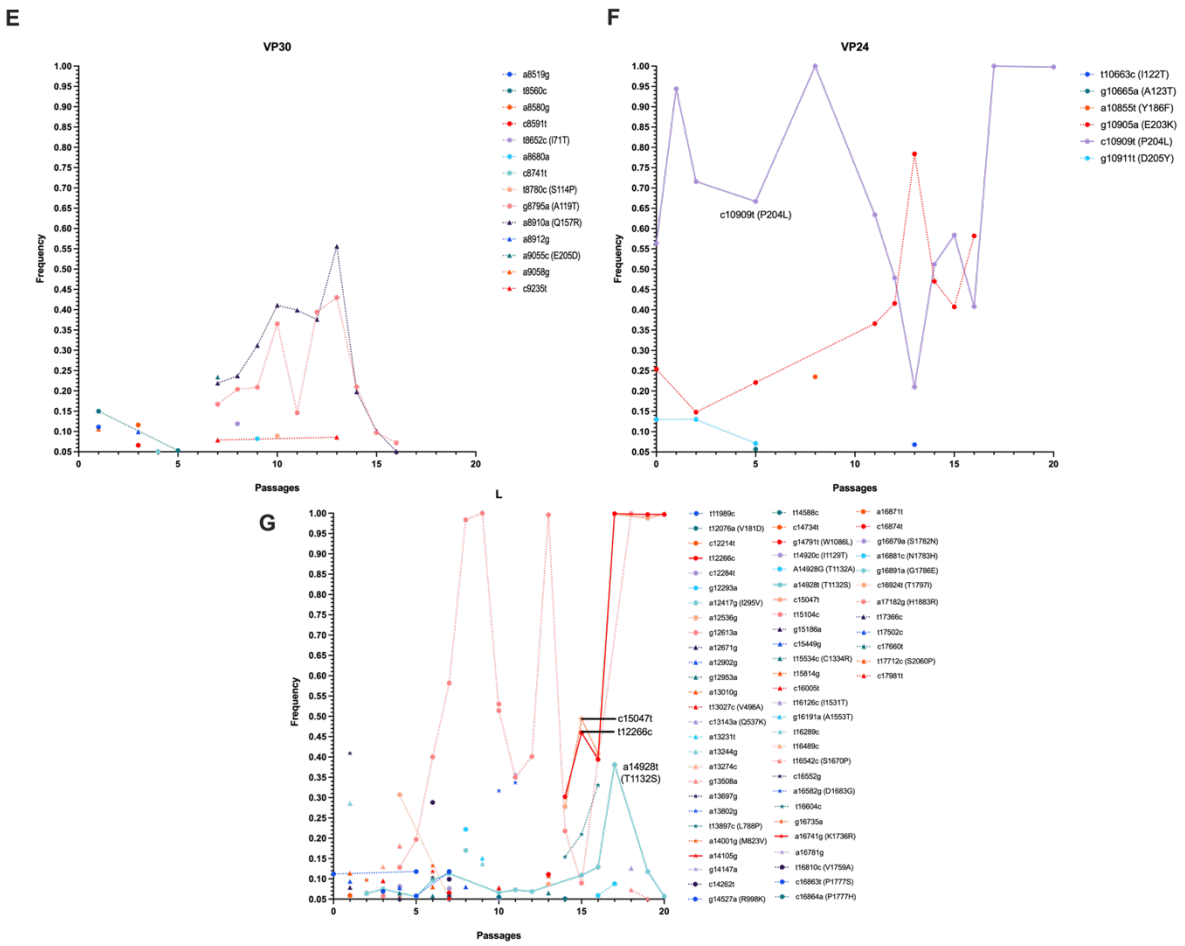
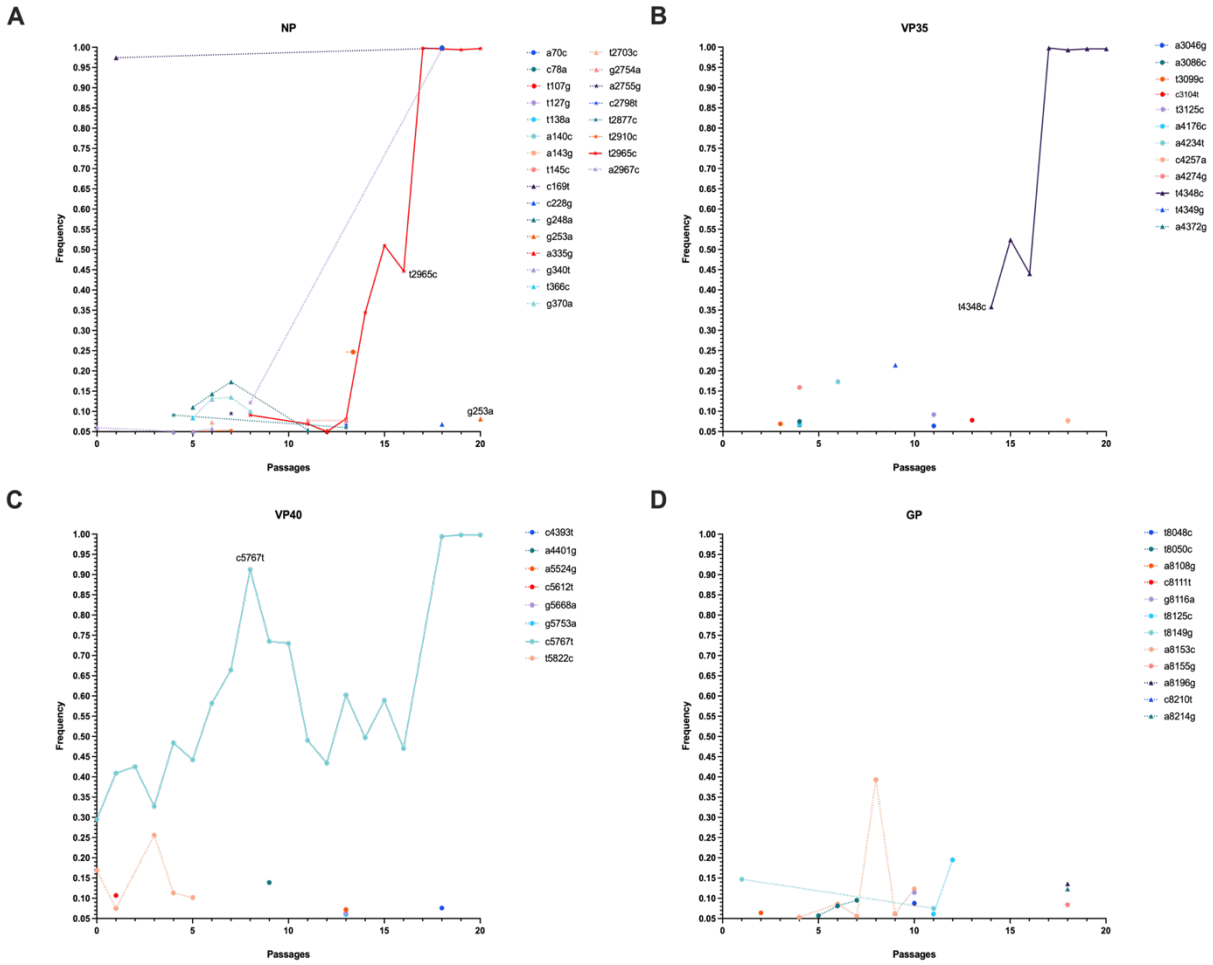


Figure 3.10 The genomic changes observed in the coding sequence regions of GPA-SUDV over 20 passages. Figures A-G depict genomic mutations identified to be present at a frequency of ≥ 5% and had a coverage of 1,000 reads or more in the coding sequence regions of the *NP*, *VP35*, *VP40*, *GP*, *VP30*, *VP24*, and *L* genes, respectively, of the GPA-SUDV genome. The x-axis represents the passages in which each mutation was identified and on the y-axis is the frequency at which each mutation occurred per passage. Nucleotide changes are expressed in lower-case letters along with their corresponding amino acid changes which are written in upper-case letters. Transient mutations are shown in dashed lines while the progress of adaptive mutations found in the final GPA-SUDV is highlighted as a solid line.


A
NP
Frequency
Passages
t2965c
g253a
a70c
c78a
t107g
t127g
t138a
a140c
a143g
t145c
c169t
c228g
g248a
g253a
a335g
g340t
t366c
g370a
t2703c
g2754a
a2755g
c2798t
t2877c
t2910c
t2965c
a2967c
B
VP35
Frequency
Passages
t4348c
a3046g
a3086c
t3099c
c3104t
t3125c
a4176c
a4234t
c4257a
a4274g
t4348c
t4349g
a4372g
C
VP40
Frequency
Passages
c5767t
c4393t
a4401g
a5524g
c5612t
g5668a
g5753a
c5767t
t5822c
D
GP
Frequency
Passages
t8048c
t8050c
a8108g
c8111t
g8116a
t8125c
t8149g
a8153c
a8155g
a8196g
c8210t
a8214g

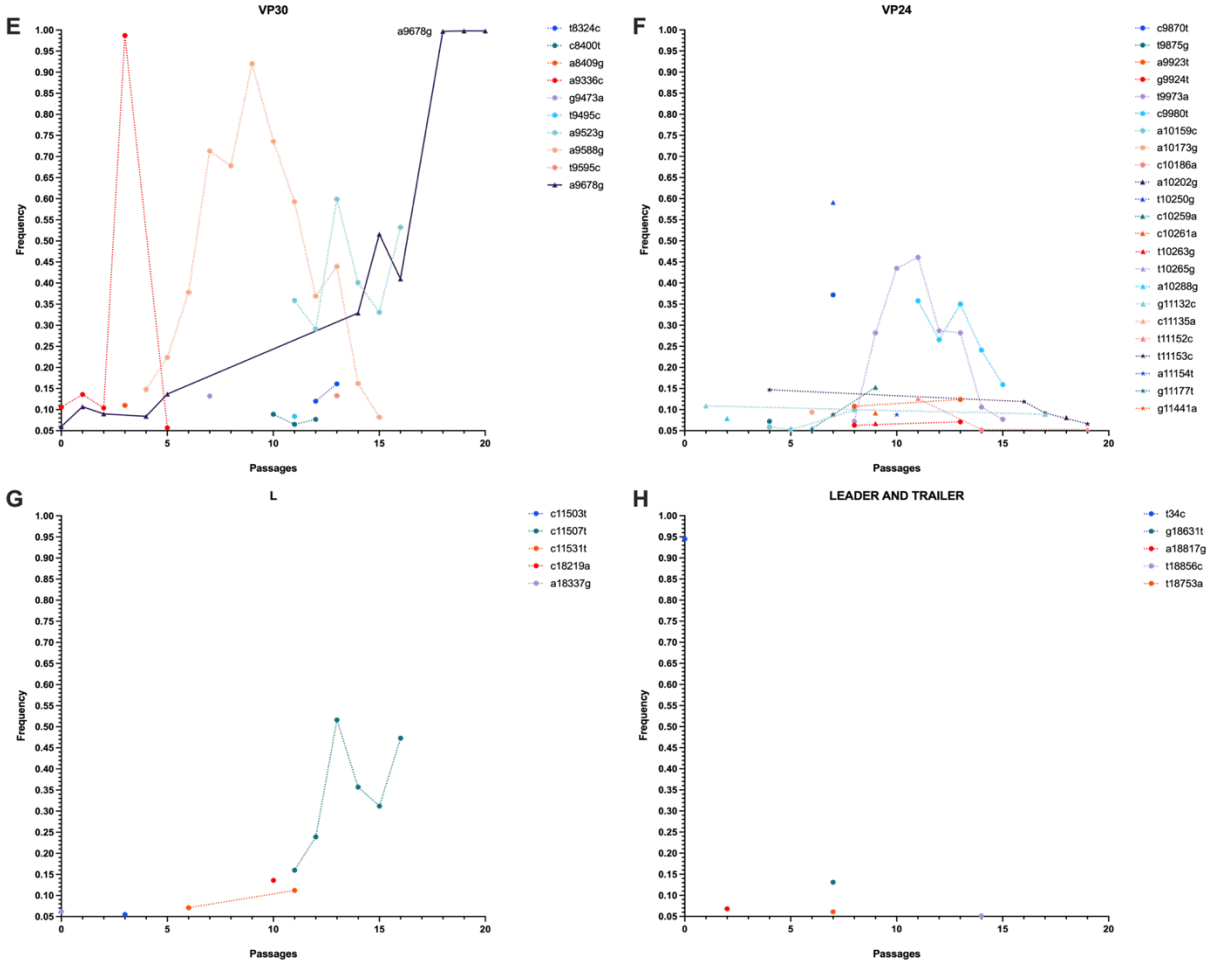


Figure 3.11 The genomic changes in the noncoding sequence region of the GPA-SUDV genome over 20 passages. Figures A-H show genomic mutations identified in the noncoding sequence regions of the *NP*, *VP35*, *VP40*, *GP*, *VP30*, *VP24*, and *L* genes and in the leader and trailer regions of the GPA-SUDV genome that were present at a frequency of ≥ 5% and a coverage of ≥ 1,000 reads. The x-axis represents the passages in which each mutation was identified and on the y-axis is the frequency at which each mutation occurred per passage. Nucleotide changes are expressed in lower-case letters. Transient mutations are shown in the dashed lines while the progress of adaptive mutations found in the final GPA-SUDV is shown as a solid line.

4 DISCUSSION

4.1 PREFACE

The pathogenesis of SUDV is less understood compared to its well-known relative filovirus, EBOV. The previous work laid out by Wong and his colleagues in which avirulent SUDV was passaged in guinea pigs (Wong et al., 2016) presented adaptive mutations that was acquired by the guinea pig-adapted SUDV (GPA-SUDV), however, which mutations are relevant for the observed virulence in adapted virus remain unknown. Central to this thesis was exploring the changes that occurred in SUDV's genome at each passage to get a detailed snapshot of the viral evolution that transpired during the adaptation process. For this study, the following three hypotheses were postulated: (i) throughout the adaptation process, the virus acquired several genomic mutations that fluctuate in frequency over time, (ii) mutations that were retained by the GPA-SUDV by the end of the adaptation process are those that reached a frequency near 100%, and (iii) these adaptive mutations likely play a role in the various aspects of the viral replication cycle and/or immune evasion. By utilizing a high-throughput sequencing approach on the pooled liver and spleen homogenates collected from each passage, this study was able to show that different regions of the viral genome experienced genomic changes that varied in frequencies from time to time, and those retained by the GPA-SUDV existed at frequencies near 100%. Achieving a detailed glimpse of the viral adaptation that occurred in guinea pigs offers a foundation for identifying which viral proteins may be key to the increased virulence phenotype observed in the host. Furthermore, having a fuller comprehension of the viral process that can render a virus pathogenic can also aid in the acceleration of developing licensed SUDV countermeasures such as vaccines and therapeutics, which is currently lacking.

The discussion section curated below expands on the probable roles that the different adaptive mutations identified in the coding and noncoding regions of the viral genome play in the viral replication cycle or on the immune evasion strategies of the virus to attempt to answer the third hypothesis of this thesis. How the GPA-SUDV differed from various rodent-adapted EBOV was also considered. Addressed in the latter half of this section are the limitations that the study holds, particularly the sequencing strategy employed to sequence the whole genome of the virus and the poor-quality state of the RNA samples used in this study.

4.2 POTENTIAL IMPLICATIONS OF NONSYNONYMOUS MUTATIONS ON VIRAL PATHOGENESIS

4.2.1 Functional roles of nonsynonymous mutations in VP40 and GP genes

4.2.1.1 VP40 mutations are important for viral morphogenesis and budding

Based on the published crystal structures of VP40 from SUDV (Bornholdt et al., 2013; Clifton et al., 2015), both the L203P and P309S VP40 mutations sit at the C-terminal domain (CTD) of the protein. VP40 is an important viral protein critical for the formation of the viral matrix that gives ebolaviruses their filamentous structures. The CTD of the protein is primarily involved in two main processes that govern proper viral matrix formation – the association of VP40 with the inner leaflet of the host PM and oligomerization of VP40 to hexamers (Gc et al., 2016; Stahelin, 2014). It appears that during the adaptation process, the L203P mutation was constantly selected for in every round of passaging as evidenced by its presence and increasing frequency over time. This mutation may therefore be important in the adaptation of SUDV in the guinea pig. Structural characterization of SUDV's VP40 reveals that this L203 residue is part of the hydrophobic CTD-CTD interface comprised of I237, M241, M305, and I307 residues, that is formed when VP40 dimers oligomerize into a filamentous hexamer as shown in Fig. 1.7B (Bornholdt et al., 2013). Mutagenesis of EBOV VP40 residues M241 and I307 within the CTD-CTD interface resulted in the absence of filamentous virus-like particles, which emphasizes the significance of these residues and the CTD-CTD interface in morphogenesis and budding (Bornholdt et al., 2013). The L203 residue is also buried at the CTD-CTD interface similar to M241 and I307, hence it will be of value to perform mutational analysis and determine if L203 is also a critical residue important for CTD-CTD interaction. Moreover, the interest in elucidating the probable effects of this mutation is heightened since L203P mutations is unique to this study and has not been reported by Wong and his colleagues previously.

The change from leucine to proline at position 203 in GPA-SUDV should not significantly alter the hydrophobicity of the CTD-CTD interface as proline is still a hydrophobic amino acid that tends to adopt positions that are buried in the protein (Morgan & Rubenstein, 2013). However, replacement with cyclic amino acid can alter the original secondary structure along with this interface which can either disrupt or improve interactions with other residues. Hence, a structural

analysis will be warranted to determine if the L203P substitution stabilizes or destabilizes the original protein structure. A stabilizing feature could potentially aid in enhancing the ability of VP40 to form filamentous hexamers and increase virus particle formation and release. On the other hand, a destabilizing characteristic can allow for the transition of VP40 to a different oligomeric state. For example, a switch from the filamentous hexameric form to an octameric ring-like structure due to the dissociation of the CTD can flip the function of VP40 from associating with the host cell PM to binding with the viral RNA (Hoenen et al., 2005) and help regulate viral transcription (Radzimanowski et al., 2014).

Residue P309, however, is not considered to be a part of the CTD-CTD interface (Clifton et al., 2015). In EBOV VP40, mutagenesis of the Q309 residue still resulted in a functional CTD-CTD interface as transfected cells produced VLPs at levels similar to the wild-type (Bornholdt et al., 2013). Residues between SUDV and EBOV differ at position 309 – a hydrophobic proline exists in SUDV, while EBOV carries hydrophilic glutamine. Interestingly, the GPA-SUDV genome had a substitution from proline to serine (P309S), which is a hydrophilic amino acid. This change would have altered how the amino acid is positioned within the protein structure. The initial proline residue would theoretically be positioned within the protein due to its hydrophobicity, but a serine would be exposed to the solvent just like the Q309 residue in EBOV (Bornholdt et al., 2013). Like Q309, the P309S mutation may not influence filoviral matrix assembly. However, a study by Panchal and colleagues revealed that amino acids 309- 317 of the CTD is critical to viral egress as a mutant lacking these 18 residues did not yield VLPs. Failure to generate VLPs was shown to be because of the inability of the mutant VP40 to bind microdomains on the host PM known as cholesterol-enriched detergent-resistant membrane domains or DRMs. Hence, the substitution of proline with the hydrophilic serine residue could have restored or enhanced this VP40-DRM interaction.

4.2.1.2 GP mutations are relevant for viral entry

At the opposite end of the virus replication cycle is viral attachment and entry to host cells mediated by GP. The final GPA-SUDV GP acquired three nonsynonymous mutations: Y151H, F153L, and A30T, all of which are located at the GP_1 subunit of the full GP. The Y151H and F153L mutations are both situated at the RBD of GP_1, a region comprised of ~150 amino acids

and important for recognition of the endosomal receptor NPC1 (Brindley et al., 2007; Manicassamy et al., 2005) and interaction with the GP_2 fusion subunit (Jeffers et al., 2002; Volchkov et al., 1998). Hence, it is hypothesized that these two mutations influence the ability of GPA-SUDV's GP to facilitate viral entry. Multiple studies have explored possible EBOV receptor-binding domain (RBD) residues on GP_1 that are important for receptor binding (Brindley et al., 2007; Manicassamy et al., 2005; Wang et al., 2011). Among these, an alanine mutational analysis by Brindley et al. (2007) displayed the significance of amino acids F153 and H154 in the viral entry process. Virions that carried a double F153A/H154A mutation showed reduced transduction efficiencies in Vero and human glioblastoma (SNB-19) cell lines and poor competitive binding with wild-type EBOV (Brindley et al., 2007). The minimized transduction and binding efficiencies were linked to the protein's insensitivity to the cathepsin cleavage (Brindley et al., 2007). Manicassamy et al.'s mutagenesis analyses on residues F153A and H154A further supports the importance of these residues. An F153A substitution resulted in decreased GP incorporation in pseudotype viruses and infectivity versus wild-type, while the H154A mutation abolished GP processing and/or expression as no GP_1 product was detected (Manicassamy et al., 2005). Mapping of key EBOV RBD residues highlights the importance of F153 residue in viral entry, although not by direct binding with the endosomal receptor. This residue has been mutated in the GPA-SUDV GP, suggesting that wild-type SUDV's entry into guinea pig cells was likely limited. The F153L substitution could have provided conformational changes to the protein that allowed for better GP incorporation to budding virions, which can positively affect the ability of the virus to enter the cells as it has been previously demonstrated that increased numbers of EBOV GP inserted into the viral envelope augments viral attachment (Marzi, Wegele, et al., 2006). Bearing in mind that the Y151H progressed in a very similar trend as the F153L mutation, the existence of Y151H might be necessary and have a synergistic effect on viral entry along with F153L. Indeed, when a substitution was placed on F151 of EBOV's GP, relative infectivity of pseudotyped virus bearing the mutated GP had a dramatic increase – over 200% of wild-type GP-pseudotyped virus (Wang et al., 2011).

On the other hand, the A30T mutation is in the signal peptide region (amino acids 1-32) at the N-terminus of the GP_1 subunit. The signal peptide of GP is responsible for the translocation of a developing GP peptide chain to the endoplasmic reticulum where the protein is folded and modified (Marzi, Akhavan, et al., 2006). Functional roles of the signal peptide from ebolaviruses

are not well-studied with only a single study published to date. A study by Marzi and colleagues demonstrated the additional role of the signal peptide in modulating the glycosylation of GP and how the protein's glycosylation state influences the viral entry (Marzi, Akhavan, et al., 2006). The function of GPA-SUDV's A30T substitution is perhaps related to this function and may alter the glycosylation state of SUDV GP for improved viral entry. This notion arose from a study by Marzi, Akhavan, et al. (2006), where wild-type SUDV was shown to lack high mannose carbohydrates and low cell infectivity. Upon replacement of the signal peptide sequence with EBOV's signal peptide sequence, infectivity increased and was attributed to the presence of high mannose carbohydrates. The very late appearance of A30T at passage 15 at a low frequency of ~5% could mean that the mutation may not be needed in the early stages of the adaptation process. However, unlike other GPA-SUDV mutations, the A30T mutation was highly selected for by the virus in the last five rounds of passaging as it reached a frequency of 99% by the 18th passage. Suggesting, otherwise, that this mutation is very likely critical for lethality in guinea pigs.

4.2.1.3 The need for VP40 and GP mutations in commencing increased virulence

The coding sequence region of the *VP40* and *GP* genes retained two and three nonsynonymous mutations, respectively, by the end of the series of passages. Most mutations from these genes were among the earliest mutations to appear during the adaptation process. The VP40 L203P mutation was detected at passage 2 at a frequency of 5.4% (Fig. 3.5C), while GP's Y151H and F153L mutations were first identified at passages 5 and 6, respectively, at a frequency of approximately ~6% (Fig. 3.5D). Early detection of mutations above the set frequency threshold of 5% may imply the need of the SUDV to initiate infection in the host by refining the ability of SUDV to enter and exit out of the cells. The selection for these mutations at earlier passages until the very last passage also highlights their importance in establishing lethality in guinea pigs.

4.2.2 Effects of VP35 and VP24 mutations on virus pathogenicity

4.2.2.1 N51D potentially affects viral nucleocapsid formation

The coding region of VP35 only acquired one nonsynonymous mutation – N51D (Fig. 3.10B). Characterization and structural analyses on EBOV VP35 reveal the protein consists of three main domains including an NP-chaperoning domain, homo-oligomerization domain, and a

dsRNA binding or inhibitory domain (Kimberlin et al., 2010; Zinzula et al., 2019). Unlike the other nonsynonymous mutations identified in GPA-SUDV, the N51D VP35 mutation falls within a linker region that links two functional domains, NP chaperoning domain (residues 20-47) and homo-oligomerization domain (residues 83-145) (Zinzula et al., 2019). A linker or spacer is a short stretch of amino acids that spaces the various domains of a single multidomain protein (Reddy Chichili et al., 2013). Linkers can adopt rigid structures to prevent negative interactions between domains or take on a flexible structure to connect domains, both of which are influenced by the sequence of the spacer (Reddy Chichili et al., 2013). Hence, a single amino acid change in the linker region of GPA-SUDV VP35 could have a profound effect on the flexibility or rigidity of the linker, which then affects interactions of the various functional domains within VP35. The N-terminus of EBOV VP35 has been shown to be a region of flexibility as evidenced by the rapid exchange of hydrogen and deuterium during a deuterium exchange mass spectrometry analysis (Kimberlin et al., 2010). The change from a polar uncharged asparagine to a negatively charged aspartic acid may confer more flexibility on the overall protein structure, particularly at the NP-chaperoning domain considering the very close proximity of N51D mutation to this domain. Increased flexibility at this domain may allow VP35 to bind and release monomeric NP more efficiently during nucleocapsid assembly. Taking into consideration how late this mutation appeared (i.e., at passage 13) may suggests that the environment within the host animal is not completely restrictive towards viral nucleocapsid assembly and that appearance of the N51D mutation could have led to a more efficient nucleocapsid assembly.

4.2.2.2 Selection of P204L mutation linked to immune evasion

Like the VP35 gene, the VP24 gene attained only one nonsynonymous mutation in its CDS region – P204L (Fig. 3.5E). Interestingly, compared to other nonsynonymous mutations, this mutation was found to exist in the unpassaged wild-type stock virus at a frequency of ~55%. The presence of this variant in the virus population even before passaging and its persistence in the viral population throughout the series of passages indicate that this mutation is of value to the virus and likely contributed to the progressive development of disease severity in guinea pigs. VP24 is a multifunctional protein with functions linked to both the viral replication cycle (Mateo, Carbonnelle, Martinez, et al., 2011; Watanabe et al., 2007) and viral immune evasion tactics (Mateo et al., 2010; St. Patrick Reid et al., 2006; Reid et al., 2007; Xu et al., 2014; Zhang et al.,

2012). The protein suppresses the host immune response by dampening the IFN response (Mateo et al., 2010). One of the many proposed mechanistic approaches that VP24 uses is binding to various members of the NPI-1 α-karyopherin nuclear transporter family (Reid et al., 2007). Characterization of the binding interface between EBOV VP24 and karyopherin α5 (also known as importin α6) reveal that VP24 binds to the importin α6's armadillo arm repeats (Xu et al., 2014), where each arm is made up of a series of tandem repeat sequences (Coates, 2003). Among the key residues identified to make contacts with importin α6 was residue P204 and the conservation of proline in both EBOV and SUDV underscores the importance of this amino acid in the function of VP24. The P204 was shown to make non-bonded contact with importin α6 (Xu et al., 2014). Recent exploration by Pappalardo et al. (2017) further investigated the molecular dynamics of binding between VP24 and importin α6. Their study revealed the presence of a solvation site near amino acids E203, P204, and D205 where water molecules get trapped and confer protein stability (Pappalardo et al., 2017). Substituting proline with a more hydrophobic leucine residue potentially makes the folding of VP24 even more stable by creating a hydrophobic effect, and as consequence, also stabilizing contacts with importin α6.

4.3 SIGNIFICANCE OF SILENT MUTATIONS ON VIRAL FITNESS

In conjunction with the nonsynonymous mutations, synonymous mutations or substitutions that do not cause an amino acid change was also detected in the CDS of the various SUDV proteins (Fig. 3.5). Several of these mutations were found to exist in the *NP* (g850a), *VP40*, (t4966c), *GP* (t6471c), and *L* (t1226c and c15047t) genes. The significance of synonymous mutations is sometimes overlooked because they do not cause changes in the actual sequence of the protein and therefore, effects are presumed to be minute. Although silent in nature, these mutations can still heavily influence viral fitness. The genetic code is degenerate, such that amino acid is coded for by multiple codons; hence, when a synonymous mutation occurs, the code on the mRNA is changed but not the actual amino acid (Crick et al., 1961). However, there is codon usage bias because codons do not exist in equal frequencies in a cell – some codons appear more often than others (Cuevas et al., 2012). In the case of viruses that are programmed to hijack the host's translational machinery, they will need to employ codon usage patterns like that of the hosts to be able to adapt and use the host's tRNA to efficiently translate proteins (Tian et al., 2018). In fact, some studies have shown resemblance in codon usage patterns between viruses and mammalian

host genes (Hernandez-Alias et al., 2021; Jitobaom et al., 2020; Miller et al., 2017). Introduction of SUDV to a new host may have decreased the productivity of protein production due to a low abundance of appropriate tRNAs in the host to deliver the amino acids, and that altering its codon usage can help circumvent this adversity such that it can use tRNAs that are abundant in guinea pigs and improve the synthesis of viral proteins.

Recoding the mRNA sequence can also substantially affect the RNA secondary structures of viruses. Modifications on such structures can influence the regulatory processes and viral replication (Kiening et al., 2019). Although RNA secondary structures are more commonly found at the untranslated regions of the viral genome, presence of functional RNA structures in the protein coding regions of viruses also exist (Ball, 1973; Dutkiewicz et al., 2020; Frolov & Schlesinger, 1996; Lobert et al., 1999; Vasin et al., 2016; Watts et al., 2009). These RNA structures can act as translation controls of viral genome expression. For example, disruption of a hairpin structure found in influenza virus A's NS1 open reading frame resulted in decreased expression of the NS1 protein, a key protein involved in antagonizing the host IFN response (Baranovskaya et al., 2019). Similar results were observed in the destabilization of a secondary structure located downstream of the capsid protein's start codon of the Sindbis virus (Frolov & Schlesinger, 1996). Secondary structures can also act as replication signals. In a study by Lobert and colleagues, cis-acting replication signals situated in the unpaired bases of the RNA secondary structures formed in the capsid protein of cardioviruses were deemed necessary for the genome replication (Lobert et al., 1999). Probing for the presence of secondary structures along with the *NP*, *VP40*, *GP*, and *L* genes can help in pinpointing whether any of the synonymous guinea pig-adapted mutations lie within any secondary RNA structures.

4.4 CONSEQUENCE OF MUTATIONS IN THE NONCODING REGIONS

The noncoding regions of the SUDV genes also gained changes by the end of the adaptation series. These regions carry regulatory sequences that manage transcription initiation and termination of SUDV genes (Neumann et al., 2009) and translation of mRNA transcripts. The t4348c and c5767t mutations were found at the overlapping of intergenic regions of VP35 and VP40 (VP35/VP40), and VP40 and GP (VP40/GP), respectively, while t2965c and a9678g were detected at the 5' end of the NP and VP30 genes. None of these mutations were located at the predicted start and stop sites for each of the genes but were rather located at the 3' UTRs of NP,

VP35, VP40, and VP30. Placements of these mutations at these regions potentially affect transcriptional regulation of mRNA transcripts, and indeed, EBOV UTRs have been deemed critical for regulation of gene expression (Brauburger et al., 2015; Brauburger et al., 2014; Neumann et al., 2009). These 3' UTR mutations may have been selected by the virus to increase transcription of these genes and therefore generate more of the filoviral proteins carrying beneficial mutations for the virus. UTRs can also hold translational control on protein production. A recent study by Khadka et al. has shown that 3' UTR edits on EBOVs VP40 enhanced the replication (Khadka et al., 2021). A similar observation was also seen in a related filovirus, MARV when the 3'UTR of its nucleoprotein was modified (Khadka et al., 2021). Increased translational efficiency of transcripts encoding proteins with advantageous mutations can be beneficial as the virus can then incorporate these proteins as it makes more virions during infection.

4.5 EFFECTS OF TRANSIENT MUTATIONS ON VIRAL ADAPTATION

Evident in the sequencing of the SUDV genome from each passage was the presence of mutations that appeared transiently, were not enriched in most cases, and did not become fixed in the genome. In a broad perspective, mutations may be categorized as deleterious, neutral, or beneficial to the fitness of the virus (Loewe, 2008). Most of the mutations that existed temporarily, were not enriched over time, and were immediately lost only after a couple of passages were likely to be deleterious. A study by Pybus and colleagues of the natural populations of 143 RNA viruses revealed that a significant proportion of the amino acid variation observed in the natural population of RNA viruses consisted of short-lived deleterious mutations (Pybus et al., 2007). Deleterious mutations may need to be negatively selected for in a viral population to ensure that beneficial mutations are retained in the population, thus keeping an improved viral fitness (Loewe, 2008). On the other hand, transient mutations that lasted over a longer period of passaging and were enriched may either be neutral or beneficial. They can be neutral, meaning they do not have a positive or a negative impact on the viral fitness (Duret, 2008). Although no added fitness benefits are incurred by neutral mutations, their existence may influence the trajectory of the evolution of the viral population by acting as a stepping stone for the generation of a new phenotype in the population (Payne & Wagner, 2019). On the flip side, transitory mutants whose frequencies were enriched over several passages may have added advantages to the fitness of the virus. The appearance of a mutation that is more fitted to the selective environment that the viral population

is subject to may be a reason for the loss of transitory mutations that never got fixed into the GPA-SUDV. Overall, regardless of whether the roles of these transient mutations may be deleterious, neutral, or beneficial, the significance of these mutations should be acknowledged as they have the potential to shape the adaptive mutational landscape of SUDV in guinea pigs.

4.6 COMPARISON OF GPA-SUDV MUTATIONS WITH DIFFERENT RODENT-ADAPTED EBOLA VIRUSES

Small rodent adaptation of filoviruses is not uncommon. In fact, multiple lethal rodent–adapted viruses are already established for EBOV – two mouse-adapted EBOV (variants Makona and Mayinga) (Chan et al., 2019; Ebihara et al., 2006) and multiple GPA-EBOV variants (Cheresiz et al., 2016; Cross et al., 2015; Mateo, Carbonnelle, Reynard, et al., 2011; Subbotina et al., 2010; Volchkov et al., 2000). Contrary to the rodent–adapted Ebola viruses, the number of passages required to obtain a lethal GPA-SUDV were greater – 20 passages for SUDV and < 10 passages for EBOV (Chan et al., 2019; Cheresiz et al., 2016; Cross et al., 2015; Ebihara et al., 2006; Mateo, Carbonnelle, Reynard, et al., 2011; Subbotina et al., 2010; Volchkov et al., 2000). Why such differences in the number of passages occur is yet to be determined but could possibly be due to the inherent increased virulence already observed in EBOV compared to SUDV and other ebolaviruses (Yamaoka & Ebihara, 2021). Comparison of the adaptive mutations in GPA-SUDV and variations of the GPA-EBOV reveal commonalities and differences. One obvious similarity between the two is the existence of a VP24 mutation. But dissimilar from GPA-EBOV where multiple nonsynonymous mutations accumulated in the *VP24* gene of some variants of GPA-EBOV (Cheresiz et al., 2016; Cross et al., 2015; Mateo, Carbonnelle, Reynard, et al., 2011), only one VP24 non-synonymous mutation was detected in GPA-SUDV, similar to one of Volchkov et al.'s GPA-EBOV variant (Volchkov et al., 2000). The proposed possible role of the GPA-SUDV VP24 mutation, P204L, in this study (Section 4.4.2.2) does not agree with previous mutational analyses conducted using different variations of recombinant EBOV. A comparison of a wild-type EBOV and a recombinant EBOV comprised of three VP24 mutations (i.e., M71I, L147P, and T187I) revealed that the recombinant virus exhibited a more robust replication in guinea pig macrophages compared to wild-type EBOV and that only the recombinant EBOV was able to produce inclusion bodies containing nucleocapsids characteristic to that of the original GPA-EBOV (Mateo et al., 2011). Surprisingly, both the wild-type and the recombinant EBOV inhibited

the IFN response to a similar degree (Mateo, Carbonnelle, Reynard, et al., 2011). Therefore, the impression is that VP24 is a key determinant of EBOV pathogenesis in guinea pigs by modulating viral nucleocapsid assembly and is less likely related to the immune evasion (Banadyga et al., 2016; Mateo, Carbonnelle, Reynard, et al., 2011).

Another striking difference is the accumulation of several synonymous and nonsynonymous mutations in the *NP* gene of GPA-EBOV variants, while GPA-SUDV only had one synonymous mutation. However, mutations within NP are presumed to be not significant to the adaptation of EBOV in the guinea pigs (Cheresiz et al., 2016). This assumption was supported by an analysis by Mateo et al., where the presence of an NP F648L mutation (Volchkov et al., 2000) alone only resulted in the death of 1 animal out of 6 (Mateo, Carbonnelle, Reynard, et al., 2011). This was only a single analysis conducted on a single GPA-EBOV NP mutation and further investigation needs to be conducted to get a better understanding of the role(s) of NP mutations identified in other variants of GPA-EBOV. It may be possible that NP alone cannot cause a huge effect on the fitness of the virus, but when present in combination with other mutations within the genome, NP mutations may provide an additive or synergistic effect on the phenotypic outcome of the virus. Results from this study may agree with the perception above as the GPA-SUDV did not really gain multiple mutations in NP. However, SUDV did acquire many transient mutations in NP which may foreshadow the relevance of NP in the adaptation process. A unique feature of the GPA-SUDV is the presence of four VP40 mutations – three in the CDS region and one in the noncoding region. The absence of VP40 mutations in the GPA-EBOV variants implies that viral morphogenesis and budding are most likely not hindered in the guinea pig host, while this aspect of the viral replication cycle is probably hindered for SUDV.

Overall, it appears that pathogenesis resulting from the adaptation of GPA-EBOV is most likely governed by the structural roles that VP24 and NP provide, particularly by enabling proper assembly of the nucleocapsid. Meanwhile, in GPA-SUDV, the virulence of the adapted virus is potentially influenced by a combination of improved viral replication and the ability to evade the host immune response, with mutations in V40 and GP facilitating an efficient virion formation process and the VP24 mutation aiding in avoiding the host IFN response. Apart from identifying molecular determinants that allow a pathogen to cause harm and morbidity in its host, it is also important to acknowledge the existence of unknown hosts and environmental factors that help shape the mutations that arise during the adaptation process.

4.7 ACHIEVING A HIGH RESOLUTION WITH VARIANT DETECTION USING NGS

The work presented used a high throughput DNA sequencing technology, MiSeq, for variant detection. Unlike Sanger sequencing, where a single DNA fragment is sequenced one at a time (Hall, 2007), high throughput sequencers or next-generation sequencing (NGS) platforms were invented with the capacity to simultaneously and accurately sequence millions of fragments per sequencing run (Tucker et al., 2009). The advent of massively parallel sequencing technologies has changed the game of genome sequencing as it made it feasible to perform large-scale sequencing and generate an extensive amount of data, in megabase and gigabase (Gb) scales, with a short turn-around time and at a significantly reduced cost (Metzker, 2010; Tripathy & Jiang, 2012). The MiSeq system is one of the most widely used NGS platforms developed by Illumina in 2011 that uses a sequencing-by-synthesis technology. This approach uses a single fluorescently labelled deoxynucleoside triphosphate (dNTP) that is incorporated into the synthesized strand, imaged subsequently, and enzymatically cleaved to allow for the next dNTP to polymerize with the growing nucleic acid chain (Illumina, 2010). This sequencing cycle is repeated to resolve the sequence of the fragment one base at a time (Illumina, 2010). The MiSeq system can yield a maximum output of 15 Gb with 25 million reads per run and a max paired-end read-length of 300 bp (Illumina, 2022a). How much data output and sequencing reads will be needed in a study will depend on the intended experimental design, the size of the genome, and the power of detection required (Illumina, 2022b). Run time can vary and ranges from as short as 4 hours to as long as 55 hours (Illumina, 2022a). Since its release in 2011, the MiSeq benchtop sequencer has become a popular choice for many laboratories that require a sequencer with a smaller lab bench footprint and require a relatively small data output. Additionally, MiSeq has high reported accuracy, can do long read lengths and is low cost. Some of the most common applications of MiSeq include small genome sequencing of microbial and viral genomes, targeted gene sequencing, small RNA analysis, targeted RNA sequencing, and DNA-protein interaction analysis in combination with chromatin immunoprecipitation assays (Illumina, 2022c).

In this study, the whole genome of SUDV was deep sequenced prior to passaging and during subsequent passages in guinea pigs to confirm and explore at which frequencies the guinea pig adapted mutations previously presented by Wong et al. (2016) were existing in each passage. Although Sanger sequencing can be used for mutation screening, this sequencing approach will

only resolve the consensus sequence of the viral population represented by the most abundant quasispecies (Posada-Cespedes et al., 2017), which are a collection of mutated genomes that makes up viral populations (Domingo et al., 2012). This was in fact true for several of the mutations identified in the final GPA-SUDV by Wong and his colleagues (Wong et al., 2016) that were found to exist at a frequency of 99% by the 20th passage. Moreover, deep sequencing of the initial stock virus used in passaging showed that viral strains with mutations c5767t and c10909t (P204L) already existed in the viral stock. These two mutations were found at frequencies of approximately 30% and 54%, respectively (Fig. 3.9A). Since these strains are of lower frequencies in comparison with the strains carrying the original nucleotides, these would have been missed by Sanger sequencing and as reported by Wong et al. (2016), the consensus sequence of the progenitor virus retains the original base at those positions. To gain valuable insight on which adaptive mutations may be critical in the pathogenesis of SUDV in guinea pigs over time, it was critical to sequence the viral genome in higher resolution to be able to monitor the progress of each mutation at each passage as accurately as possible. Employing the use of a high throughput sequencer that generates immense numbers of sequencing reads, a high sequencing coverage was attainable, which enabled the detection of minority variants that existed in frequencies as low as the set detection limit of 5%.

4.8 STUDY LIMITATIONS

4.8.1 Single-plex long amplicon-based PCR approach is labour-intensive

To sequence the entire viral genome, a single-plex long amplicon PCR approach was used. Several primer pairs were selected to produce 16 different amplicons of 1 to 2 kb in size that were gel purified and sequenced with a MiSeq sequencer. However, this approach presented some challenges. First, the process was time-consuming and laborious because of the need to set up individual PCR reactions per fragment for each of the GPA-SUDV passages, amounting to at least 336 reactions. To circumvent this issue, a multiplex tiling PCR approach can be implemented to stitch the viral genome sequence together (Quick et al., 2017). With the multiplex approach, the tiled amplicons will be generated using two different sets of multiplex primers that will be separated into two PCR primer pools where the fragments produced will overlap between but not within the reactions (Quick et al., 2017). This approach would ideally only require 42 PCR

reactions for the 21 GPA-SUDV samples. The efficiency of this method further extends to library preparation for NGS where only two products would be pooled per sample instead of 16 amplicons.

4.8.2 Poor quality RNA hindered amplification of long amplicons

Another pitfall of the long-amplicon PCR attempt was the difficulty in acquiring all the 16 fragments for all the passages. RNA bioanalyzer profiles of the GPA-SUDV samples showed that the total RNA in these samples was heavily degraded and different hypotheses were postulated as to why this may be the case. One of the many potential reasons is that degradation could be due to the RNA extraction method. RNA was extracted from pooled liver and spleen homogenates using silica-based columns. As the bound RNA is eluted from the column, shearing of longer RNAs to shorter fragments can be inevitable (Murphy, 2021). This appeared to be true as some samples extracted with magnetic beads, which should lessen RNA shearing, showed improvement in their RNA bioanalyzer profiles with a slight shift of peaks towards higher molecular weight RNAs (Fig. 3.5 C-G). However, the use of an alternative extraction method did not eliminate the presence of lower molecular weight RNA fragments, suggesting that degradation may be an intrinsic characteristic of these samples. The samples were of liver and spleen origin, which are both known to possess high amounts of RNases that are omnipresent in these tissues (Javan et al., 2020). If high levels of RNases are released from intracellular vesicles following tissue damage from the homogenization process, heavy degradation of RNA could occur (Vehniäinen et al., 2019).

Taking a multiplexed short-amplicon PCR approach may assist in recovering sequences from shorter RNA fragments. Using a free online primer designing tool, Primal Scheme (http://primal.zibraproject.org), multiplexed primers can easily be designed to generate amplicons of as short as 200 nucleotides by simply uploading the reference sequence of choice and specifying the desired amplicon size (Quick et al., 2017) . Moreover, the software also predetermines for the user the two primer pools needed for this type of PCR approach. This approach has been widely used for sequencing of Zika virus in samples with low viral loads (Black et al., 2017; Dudley et al., 2017; Faria et al., 2017; Grubaugh et al., 2017; Magnani et al., 2018; Metsky et al., 2017; Quick et al., 2017) and of other arthropod-borne viruses like West Nile (Hepp et al., 2018) and Dengue (Stubbs et al., 2020). Multiplex primers have also been used for surveillance of EBOV in Guinea (Quick et al., 2016) and Sierra Leone (Arias et al., 2016). Multiplex PCR was not attempted in this study and primers generated through Primal Scheme that was designed to generate 300 bp

amplicons were ordered with the desire to use newer primers stocks in obtaining the missing amplicons and re-amplify amplicons that gave low coverage after sequencing. However, an attempt to recover amplicons small than 1 kb (~ 600 to 800 bp) was conducted using primers from this Primal Scheme. In some instances, amplification of amplicons that were 600-800 bp was successful, but there were cases where it did not (see Appendix B). This could be because the sizes of the RNA fragments in the extracted RNA are much shorter than 600 nucleotides (Fig. 3.4) or simply because the RNA was of poor quality since even amplification of small sized housekeeping genes were unsuccessful in these RNA samples (Fig. 3.6). Overall, the development of an optimized sequencing protocol involving the use of multiplex primers may be worthwhile for future sequencing of ebolavirus genomes as previously established protocols appear to be easily adaptable to various virus genomes, the process is time-saving with the need of only two PCR reactions, and the process is more cost-effective with the reduced use of consumables and reagents.

4.8.3 Sequence data lacked the genomic 3' terminus region

The initial attempt to sequence the first 1300 nt of the genome proved to be difficult with the long amplicon-based approach that used a forward primer positioned from nt 1 to 20 as no amplicon was obtained from all the 20 passages. However, when a different forward primer that binds at nt positions 205 to 224 was selected and used for PCR along with the same reverse primer positioned at nucleotides 1334 to 1353, a 1352-bp PCR amplicon was attained easily. Since the original forward primer failed to work, the sequence for the first 200 nt of the SUDV genome was not obtained. The missing 200 nt encompasses the genomic leader sequence (nt 1-53) and a part of NP's UTR region (nt 54- 457). Difficulty in sequencing the termini of viral genomes is not uncommon (Alfson et al., 2014) and is an adversity faced by researchers in the field because producing full-length cDNA from the template RNA is often problematic (Chenchik et al., 1996; Coutts & Livieratos, 2003). In the case of ebolaviruses, the synthesis of partial cDNA templates may be likely due to the presence of a hairpin structure formed in the leader region of the genome (Bach et al., 2021). The existence of such a secondary structure can lower amplification efficiency by limiting the access of primers to bind to the target region and thus interfere with the production of a cDNA template containing a copy of the 3' end sequences of the viral genome. Performing 3'-end rapid amplification of cDNA ends (RACE) can help address this issue. To execute a 3' RACE: (i) first, the viral RNA is tagged with a poly-A "adapter" sequence, (ii) then the tagged

viral RNA is reverse transcribed using oligo-dT primers complementary to the poly-A region, and lastly, (iii) specific primers complementary to a known region of the cDNA template and the adapter sequence are used to perform a nested PCR to amplify the unknown 3' end sequences between the known region of the cDNA and the poly-A region (Yeku & Frohman, 2011).

4.9 IS DEDUPLICATION OF SEQUENCING READS NECESSARY?

The core workflow of preparing DNA samples for NGS involves fragmentation of DNA, ligation of sequencing adapters on both ends of the DNA fragment, and a few rounds of PCR amplification for the enrichment of the DNA fragments with adapters (Head et al., 2014). During the sequencing of the prepared library, PCR duplicates arise from the sequencing of two or more copies of the exact DNA molecule that exist because of the library enrichment step prior to the sequencing (Ebbert et al., 2016). This phenomenon is inevitable for two main reasons, (i) it is unpredictable and uncontrollable which of the PCR products will bind to the sequencing flow cell, and (ii) because of PCR bias (Ebbert et al., 2016). Computational tools that enable the removal of these PCR duplicates can be incorporated in various sequencing analysis pipelines to mitigate this inherent consequence of the NGS (Ebbert et al., 2016; Zhou et al., 2014). The presence of PCR duplicates can particularly be detrimental to viral population diversity analyses as it can lead to false-positive variant calls (i.e., misidentification of a PCR-induced error as a true variant) and inflate sequencing coverage that leads to the overestimation of variant allelic frequencies (Olson et al., 2015; Tin et al., 2015).

To determine if this was the case for this study, a deduplication step was added to the original variant calling pipeline (Fig. 2.3). The effects of deduplication were afterwards assessed by examining if any of the previously identified mutations were lost after deduplication and if the frequencies of each mutation were changed dramatically. Surprisingly, none of the mutations formerly recognized using the pipeline that lacked the deduplication step were lost or fell below the set limit of detection for frequency, which was 0.05 (or 5%) after deduplication (Fig. 3.7). The majority of the called mutations did experience a drop in frequency after deduplication, although the difference observed from the original frequencies was no more than 1% and was overall non-significant. These data suggest that the deduplication of PCR duplicates had a negligible effect on the accuracy of variant calling. A comparison made by Ebbert et al. between variant calls made with and without PCR duplicates on 99 whole-genome sequencing samples from an Alzheimer's

Disease study yielded similar results, where the accuracy of variant calls between the different pipelines they used were similar and thus questions the need for PCR deduplication (Ebbert et al., 2016). The true need for removing duplicate reads remains debatable as not many comparative studies exist on this issue.

4.10 MAJOR FINDINGS AND CONCLUSIONS

The use of an amplicon-based high throughput sequencing approach to examine the SUDV genome in each passage provided a detailed view of the genomic changes that occurred as the passaging progressed in guinea pigs and answered the hypotheses outlined in Section 1.6. The first hypothesis states that throughout the adaptation process, the virus acquired several genomic mutations that fluctuate in frequency over time, and indeed, the first major finding of this study revealed that the SUDV genome acquired a total of 333 mutations between the wild-type SUDV and the 20 passages that existed at different frequencies – from 5% to 100%. The second major finding of this study was the identification of the 18 mutations in the genome of the GPA-SUDV – 13 were in the coding sequence and 5 were in the noncoding sequence. Apart from the g253a and a14928t mutations in the *NP* and *L* genes, respectively, 16 of the adaptive mutations were present at a frequency of 99%. This finding confirms the second hypothesis to a certain extent where it was postulated that mutations retained by the GPA-SUDV by the end of the adaptation process are those with frequencies near 100%. Surprisingly, considering that the L203P mutation existed at a frequency of almost 100%, this mutation was not detected by Wong and his colleagues when they Sanger-sequenced GPA-SUDV. Therefore, the identification of this mutation was unique to this study. Not only was this study able to determine the frequencies of the adaptive mutations at the last passage where the GPA-SUDV was recovered, but the study was also able to provide a timeline of when the adaptive mutations appeared and the frequencies at which they existed. It was determined that the c5767t, a9678g, and c10909t (P204L) mutations in the VP40, VP30, and VP24, respectively, were already present in the unpassaged wild-type SUDV and was selected for by the virus over time. The earliest mutations that appeared subsequently after 1 to 6 passages were found in NP (g850a), VP40 (L203P) and GP (Y151H and F153L), which initially started at low frequencies and consistently increased in frequency over time. Meanwhile, the rest of the adaptive mutations appeared much later after the 10^{th} passage. Interestingly, although the

mutations appeared at different points in time, most of the mutations were able to retain frequencies above 95% starting at the same passage, passage 17.

Considering that the mutations present in the GPA-SUDV were in various regions of the genome, it was hypothesized that the adaptive mutations likely contribute to various aspects of the viral replication cycle of the virus and/or immune evasion tactics of the virus. To investigate the probable functional roles of the different mutations identified in the viral genome, a literature search was conducted. Nonsynonymous mutations found in the *VP35*, *GP* and *VP40* genes likely augments viral replication by enabling a more efficient nucleocapsid formation, improved viral entry, and enhanced morphogenesis and budding, respectively. Meanwhile, the P204L mutation in VP24 was hypothesized to confer increased virulence in the host by dampening the host IFN immune response. Synonymous mutations were also present in the adapted genome and may have contributed to increased virulence by altering the codon usage of the virus to be able to translate proteins more efficiently or alter the regulatory process of the virus that leads to better viral replication by modifying the secondary structures that may be present in the open reading frame (ORF) region of the genes. Probing for secondary structures in the ORF of the SUDV genome can aid in confirming the latter hypothesis. Lastly, mutations located in the UTR regions of the genes may be beneficial to the virus by boosting the translation efficiency of the transcripts. Overall, the adaptive mutations obtained by the SUDV when passaged in guinea pigs likely refine viral replication and the immune evasion strategies of SUDV.

Aside from addressing the core hypotheses of this thesis, this study was also able to gain insights into the quality of the RNA samples that were used. Determining the quality of the RNA samples was determined to be detrimental to elucidating why the success in PCR amplification was variable between passages. Both the RNA bioanalyzer assays and attempts to amplify different housekeeping genes from the extracted RNA reveal that the RNA from the guinea pig liver and spleen homogenates were of poor quality as they were highly degraded. The degradation observed was determined to be not primarily because of the old age of the samples but was most likely due to the origin of the samples themselves as the liver and spleen organs are known to carry high amounts of RNases that degrade RNA. The lack of long fragment RNA in these samples may be a reason why obtaining all the 16 amplicons was not uniformly achievable in all the passages.

Employing a multiplex-PCR approach with amplicons sizes of less than 1 kb may be more feasible compared to the approach used for this study.

Overall, this study provided insight into the virus evolution that occurred during the adaptation process and can help identify hotspots in the genome that may be important in conferring virulence. Being able to provide a detailed timeline of when mutations appeared and what frequencies they occurred can be used as a guide to select for mutations that might be worth investigating further. Exploration of actual key mutations that leads to lethality in guinea pigs and the potential impacts of these mutations in the fitness of the virus is still warranted and could be an extension of this study.

4.11 FUTURE DIRECTIONS

Central to this thesis was providing a detailed overview of the progress of the adaptive mutations identified in the GPA-SUDV genome during the passaging series. However, which viral mutations are key to the virulence and lethality observed in guinea pigs remain to be elucidated. To identify the molecular determinant(s) responsible for the lethality in guinea pigs, creating SUDV variants with different combinations of the adaptive mutations will be important. The use of molecular cloning techniques will be an indispensable tool that will help generate different variants of SUDV. First, the full-length genome construct of the wild-type SUDV (variant Boniface) will be cloned to a cloning vector. Following cloning, site-directed mutagenesis will be performed to introduce the 18 adaptive mutations to the wild-type construct. A variety of GPA-SUDV mutants will also be generated by switching out the mutated regions with the wild-type SUDV sequence. A schematic representation of constructs that can be explored for future experiments is highlighted in Fig. 4.1. The full-length plasmid constructs will further be used to rescue recombinant SUDVs using a reverse genetics system. Characterization of these recombinant viruses, like their virulence and replication kinetics, may then be explored *in vivo* in guinea pigs and/or *in vitro* in a commercially available guinea pig cell-line. The examples of constructs listed in Fig. 4.1 can be used to initially screen and narrow down which regions in the viral genome that carries the adaptive mutations contribute to virulence. Alternatively, the previously collected pooled liver and spleen homogenates may also be used to deduce mutations that could be crucial for GPA-SUDV's increased virulence in guinea pigs. *In vitro* infectivity

assays using plaque-purified viral stocks from each passage be used to initially characterize the replication kinetics of each virus, which subsequently allows for comparing differences in the replication kinetics of viruses from different passages. The viruses can then be used to challenge a set of guinea pigs from which virulence will be monitored. Combining the clinical data and the sequencing data obtained from this study will make it possible to have a fuller comprehension of when animals began to get sick and which mutations were present at that time.

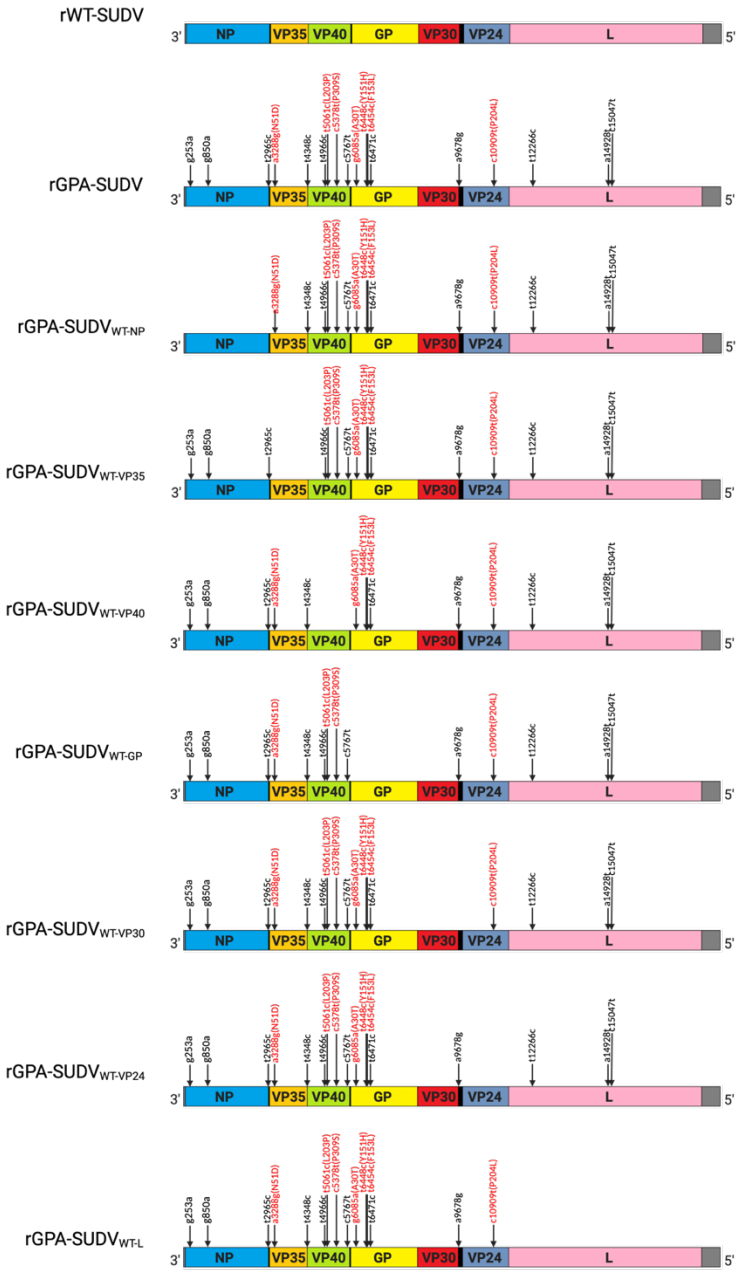


Figure 4.1 Full-length genome constructs of recombinant SUDV viruses. The following constructs will be used to initially screen regions in the viral genome that may be crucial for SUDV pathogenesis in guinea pigs.

The probable roles of the adaptive mutations were also described in this thesis, but the true contributions of these mutations to the fitness of the virus are not clear. To deduce the roles that the mutations from each of the SUDV proteins play in the viral replication cycle, developing a tetracistronic transcription and replication-competent virus-like particle (trVLP) system is proposed. This tetracistronic trVLP system is a life-cycle modelling system that can model various aspects of the filovirus replication cycle, including viral entry, genome replication and transcription, and viral egress (Watt et al., 2014). The SUDV trVLP system will be comprised of a tetracistronic minigenome that encodes a reporter gene and structural proteins such as VP40, $GP_{1,2}$, and VP24, as well as RNP proteins that will be supplied exogenously and expressed from expression plasmids (Fig. 4.2). To produce trVLPs, the said components will be transfected in producer cells along with an expression plasmid encoding for a T7 polymerase that will initially transcribe the minigenome to produce the viral RNA (vRNA). In the presence of the RNP proteins, the vRNA will be encapsidated, replicated, and transcribed. The resulting mRNAs are then translated, producing the structural proteins that aid in viral assembly and budding of nucleocapsid-bearing trVLPs coated with $GP_{1,2}$. The detected reporter activity in these producer cells reflects the vRNA synthesis, transcription, and trVLP production (Watt et al., 2014). The generated trVLPs can be subsequently used to infect desired target cells which may or may not be pre-transfected with the RNP helper plasmids. If introduced in targets cells with pre-transfected RNP plasmids, the reporter activity relies on vRNA synthesis, transcription, and trVLP formation in both the producing cells and the target cells, as well as in viral entry in the target cells (Hoenen et al., 2011). Meanwhile, in naïve cells lacking the RNP proteins, reporter activity is dependent on vRNA synthesis in the producer cells only, and primary transcription, viral entry, and trVLP production in the target cells (Hoenen et al., 2011). Using this system, the effects of each GPA-SUDV mutation or a combination of these mutations on the viral replication cycle can be evaluated *in vitro* by cloning the mutation(s) of interest to the desired viral protein and then measuring the level of reporter activity in the target cells, which will be compared relative to wild-type.

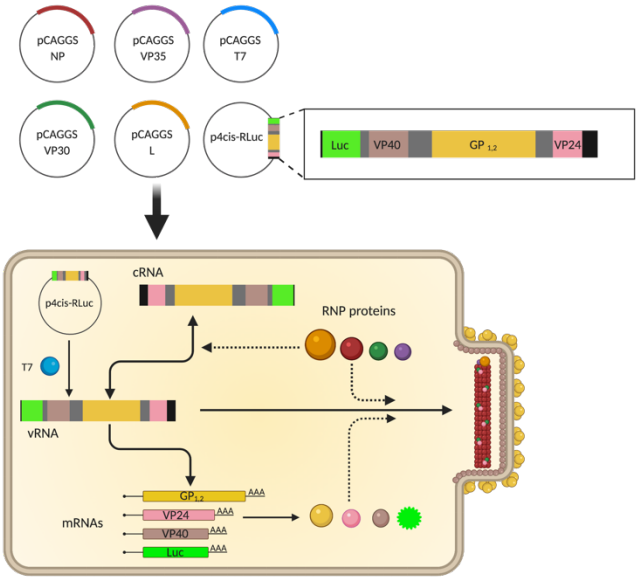


Figure 4.2 Production of transcription and replication competent virus-like particles (trVLP) with a tetracistronic minigenome. Producer cells are tranfected with RNP proteins and T7 polymerase encoded in an expression plasmid, pCAGGS, and tetracistronic minigenome. The tetracistronic minigenome encodes for a reporter gene (e.g. luciferace gene) and viral structural proteins – VP40, $GP_{1,2}$, and VP24. Upon transfection, initial transcription of minigenome to viral RNA (vRNA) by the exogenous T7 polymerase occurs. Encapsidation, genome replication, and transcription vRNA occurs subsequently in the presence of RNP proteins. Transcription of vRNA results in the production of VP40, $GP_{1,2}$, and VP24, thus leading to formation of trVLPs.

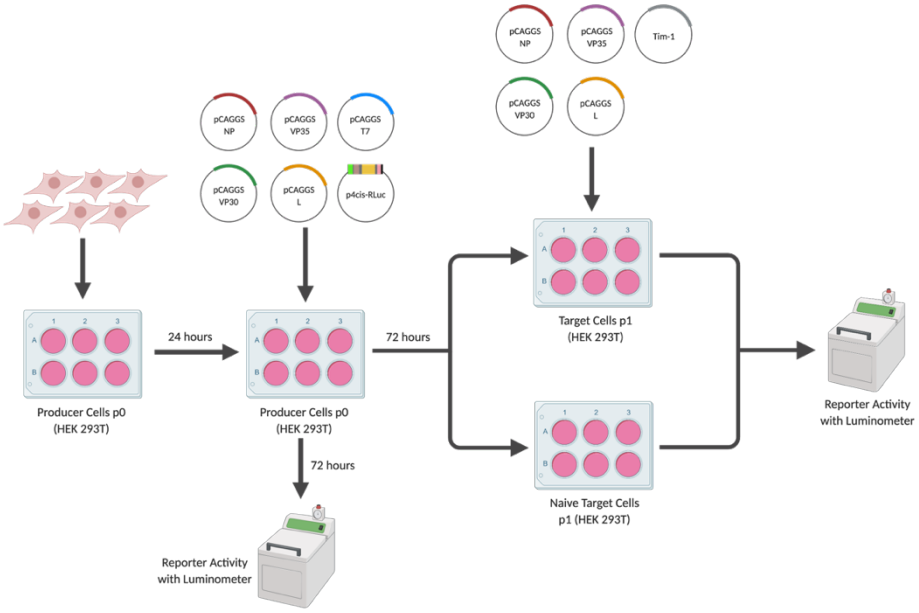


Figure 4.3 Schematic representation of trVLP assay with a tetracistronic minigenome. A stock of human embryonic kidney 293T cells split and propagated for 24 hours to serve as producer cells for initial production of trVLPs. After 24 hours, the producer cells are transfected with the tetracistronic minigenome and RNP proteins and T7 polymerase that are individually cloned to a pCAGGS expression plasmid. After transfection, fluorescence signal is measured and trVLPs are harvested and used to infect target cells. The desired target cells can either be cells pre-transfected with RNP plasmids and an attachment factor, Tim-1, or naïve cells lacking the RNP plasmids.

5 REFERENCES

Adu-Gyamfi, E., Soni, S., Jee, C., Digman, M., Gratton, E., & Stahelin, R. (2014, 2014-10-17). A Loop Region in the N-Terminal Domain of Ebola Virus VP40 Is Important in Viral Assembly, Budding, and Egress. *Viruses, 6*(10), 3837-3854. https://doi.org/10.3390/v6103837

Afelt, A., Lacroix, A., Zawadzka-Pawlewska, U., Pokojski, W., Buchy, P., & Frutos, R. (2018). Distribution of bat-borne viruses and environment patterns. *Infection, Genetics and Evolution*. https://doi.org/10.1016/j.meegid.2017.12.009

Aleksandrowicz, P., Marzi, A., Biedenkopf, N., Beimforde, N., Becker, S., Hoenen, T., Feldmann, H., & Schnittler, H.-J. (2011, 2011-11-01). Ebola Virus Enters Host Cells by Macropinocytosis and Clathrin-Mediated Endocytosis. *The Journal of Infectious Diseases, 204*(suppl_3), S957-S967. https://doi.org/10.1093/infdis/jir326

Alfson, K. J., Beadles, M. W., & Griffiths, A. (2014, 2014/11/01/). A new approach to determining whole viral genomic sequences including termini using a single deep sequencing run. *Journal of Virological Methods, 208*, 1-5. https://doi.org/https://doi.org/10.1016/j.jviromet.2014.07.023

Amiar, S., & Stahelin, R. V. (2020). The Ebola virus matrix protein VP40 hijacks the host plasma membrane to form virus envelope. *Journal of lipid research, 61*(7), 971-971. https://doi.org/10.1194/jlr.ILR120000753

Arias, A., Watson, S. J., Asogun, D., Tobin, E. A., Lu, J., Phan, M. V. T., Jah, U., Wadoum, R. E. G., Meredith, L., Thorne, L., Caddy, S., Tarawalie, A., Langat, P., Dudas, G., Faria, N. R., Dellicour, S., Kamara, A., Kargbo, B., Kamara, B. O., Gevao, S., Cooper, D., Newport, M., Horby, P., Dunning, J., Sahr, F., Brooks, T., Simpson, A. J. H., Groppelli, E., Liu, G., Mulakken, N., Rhodes, K., Akpablie, J., Yoti, Z., Lamunu, M., Vitto, E., Otim, P., Owilli, C., Boateng, I., Okoror, L., Omomoh, E., Oyakhilome, J., Omiunu, R., Yemisis, I., Adomeh, D., Ehikhiametalor, S., Akhilomen, P., Aire, C., Kurth, A., Cook, N., Baumann, J., Gabriel, M., Wölfel, R., Di Caro, A., Carroll, M. W., Günther, S., Redd, J., Naidoo, D., Pybus, O. G., Rambaut, A., Kellam, P., Goodfellow, I., & Cotten, M. (2016, 2016-01-01). Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. *Virus Evolution, 2*(1), vew016. https://doi.org/10.1093/ve/vew016

Aurelie, K. K., Guy, M. M., Bona, N. F., Charles, K. M., Mawupemor, A. P., & Shixue, L. (2018). A Historical Review of Ebola Outbreaks. In *Advances in Ebola Control*. InTech. https://doi.org/10.5772/intechopen.72660

Ayub, G., & Waheed, Y. (2016, 2016-06-01). Sequence analysis of the L protein of the Ebola 2014 outbreak: Insight into conserved regions and mutations. *Molecular Medicine Reports, 13*(6), 4821-4826. https://doi.org/10.3892/mmr.2016.5145

Bach, S., Demper, J.-C., Klemm, P., Schlereth, J., Lechner, M., Schoen, A., Kämper, L., Weber, F., Becker, S., Biedenkopf, N., & Hartmann, R. K. (2021, 2021-10-26). Identification and characterization of short leader and trailer RNAs synthesized by the Ebola virus RNA polymerase. *PLoS Pathogens, 17*(10), e1010002. https://doi.org/10.1371/journal.ppat.1010002

Ball, L. A. (1973, Mar 14). Secondary structure and coding potential of the coat protein gene of bacteriophage MS2. *Nat New Biol, 242*(115), 44-45. https://doi.org/10.1038/newbio242044a0

Banadyga, L., Hoenen, T., Ambroggio, X., Dunham, E., Groseth, A., & Ebihara, H. (2017, 2017-12-01). Ebola virus VP24 interacts with NP to facilitate nucleocapsid assembly and genome packaging. *Scientific Reports, 7*(1). https://doi.org/10.1038/s41598-017-08167-8

Baranovskaya, I., Sergeeva, M., Fadeev, A., Kadirova, R., Ivanova, A., Ramsay, E., & Vasin, A. (2019, 2019/12/21). Changes in RNA secondary structure affect NS1 protein expression during early stage influenza virus infection. *Virology Journal, 16*(1), 162. https://doi.org/10.1186/s12985-019-1271-0

Barré-Sinoussi, F., & Montagutelli, X. (2015). Animal models are essential to biological research: issues and perspectives. *Future science OA, 1*(4), FSO63-FSO63. https://doi.org/10.4155/fso.15.63

Baseler, L., Chertow, D. S., Johnson, K. M., Feldmann, H., & Morens, D. M. (2017, 2017-01-24). The Pathogenesis of Ebola Virus Disease. *Annual Review of Pathology: Mechanisms of Disease, 12*(1), 387-418. https://doi.org/10.1146/annurev-pathol-052016-100506

Basler, C. F., Krogan, N. J., Leung, D. W., & Amarasinghe, G. K. (2019). Virus and host interactions critical for filoviral RNA synthesis as therapeutic targets. *Antiviral Research, 162*, 90-100. https://doi.org/10.1016/j.antiviral.2018.12.006

Basler, C. F., Mikulasova, A., Martinez-Sobrido, L., Paragas, J., MüHlberger, E., Bray, M., Klenk, H.-D., Palese, P., & GarcíA-Sastre, A. (2003, 2003-07-15). The Ebola Virus VP35 Protein Inhibits Activation of Interferon Regulatory Factor 3. *Journal of Virology, 77*(14), 7945-7956. https://doi.org/10.1128/jvi.77.14.7945-7956.2003

Basler, C. F., Wang, X., Muhlberger, E., Volchkov, V., Paragas, J., Klenk, H. D., Garcia-Sastre, A., & Palese, P. (2000). The Ebola virus VP35 protein functions as a type I IFN antagonist. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.220398297

Bennett, R. S., Huzella, L. M., Jahrling, P. B., Bollinger, L., Olinger, G. G., & Hensley, L. E. (2017). Nonhuman Primate Models of Ebola Virus Disease. In *Current Topics in Microbiology and Immunology* (pp. 171-193). Springer International Publishing. https://doi.org/10.1007/82_2017_20

Bente, D., Gren, J., Strong, J. E., & Feldmann, H. (2009, 2009-01-07). Disease modeling for Ebola and Marburg viruses. *Disease Models & Mechanisms, 2*(1-2), 12-17. https://doi.org/10.1242/dmm.000471

Bharat, T. A. M., Noda, T., Riches, J. D., Kraehling, V., Kolesnikova, L., Becker, S., Kawaoka, Y., & Briggs, J. A. G. (2012, 2012-03-13). Structural dissection of Ebola virus and its assembly determinants using cryo-electron tomography. *Proceedings of the National Academy of Sciences, 109*(11), 4275-4280. https://doi.org/10.1073/pnas.1120453109

Bhattacharyya, S., Hope, T. J., & Young, J. A. T. (2011). Differential requirements for clathrin endocytic pathway components in cellular entry by Ebola and Marburg glycoprotein pseudovirions. *Virology, 419*(1), 1-9. https://doi.org/10.1016/j.virol.2011.07.018

Bhattacharyya, S., Warfield, K. L., Ruthel, G., Bavari, S., Aman, M. J., & Hope, T. J. (2010). Ebola virus uses clathrin-mediated endocytosis as an entry pathway. *Virology, 401*(1), 18-28. https://doi.org/10.1016/j.virol.2010.02.015

Biedenkopf, N., Hartlieb, B., Hoenen, T., & Becker, S. (2013, 2013-04-01). Phosphorylation of Ebola Virus VP30 Influences the Composition of the Viral Nucleocapsid Complex. *Journal of Biological Chemistry, 288*(16), 11165-11174. https://doi.org/10.1074/jbc.m113.461285

Biedenkopf, N., Lier, C., & Becker, S. (2016, 2016-05-15). Dynamic Phosphorylation of VP30 Is Essential for Ebola Virus Life Cycle. *Journal of Virology, 90*(10), 4914-4925. https://doi.org/10.1128/jvi.03257-15

Biedenkopf, N., Schlereth, J., Grünweller, A., Becker, S., & Hartmann, R. K. (2016, 2016-08-15). RNA Binding of Ebola Virus VP30 Is Essential for Activating Viral Transcription. *Journal of Virology, 90*(16), 7481-7496. https://doi.org/10.1128/jvi.00271-16

Black, A., Potter, B., Dudas, G., Feldstein, L., Grubaugh, N. D., Andersen, K. G., Ellis, B. R., Ellis, E. M., & Bedford, T. (2017). Genetic characterization of the Zika virus epidemic in the US Virgin Islands. *bioRxiv*, 113100. https://doi.org/10.1101/113100

Bornholdt, A., Zachary, Noda, T., Abelson, M., Dafna, Halfmann, P., Wood, R., Malcolm, Kawaoka, Y., & Saphire, O., Erica. (2013, 2013-08-01). Structural Rearrangement of Ebola Virus VP40 Begets Multiple Functions in the Virus Life Cycle. *Cell, 154*(4), 763-774. https://doi.org/10.1016/j.cell.2013.07.015

Bourgarel, M., & Liégeois, F. (2019). Ebola and Other Haemorrhagic Fevers. In M. Kardjadj, A. Diallo, & R. Lancelot (Eds.), *Transboundary Animal Diseases in Sahelian Africa and Connected Regions* (pp. 179-205). Springer International Publishing. https://doi.org/10.1007/978-3-030-25385-1_10

Bradfute, S. B., Braun, D. R., Shamblin, J. D., Geisbert, J. B., Paragas, J., Garrison, A., Hensley, L. E., & Geisbert, T. W. (2007). Lymphocyte Death in a Mouse Model of Ebola Virus

Infection. *The Journal of Infectious Diseases, 196*(Supplement_2), S296-S304. https://doi.org/10.1086/520602

Bradley, J. H., Harrison, A., Corey, A., Gentry, N., & Gregg, R. K. (2018, 2018-02-01). Ebola virus secreted glycoprotein decreases the anti-viral immunity of macrophages in early inflammatory responses. *Cellular Immunology, 324*, 24-32. https://doi.org/10.1016/j.cellimm.2017.11.009

Brannan, J. M., Froude, J. W., Prugar, L. I., Bakken, R. R., Zak, S. E., Daye, S. P., Wilhelmsen, C. E., & Dye, J. M. (2015). Interferon α/β Receptor–Deficient Mice as a Model for Ebola Virus Disease. *The Journal of Infectious Diseases, 212*(suppl_2), S282-S294. https://doi.org/10.1093/infdis/jiv215

Brauburger, K., Boehmann, Y., Krähling, V., & Mühlberger, E. (2015). Transcriptional Regulation in Ebola Virus: Effects of Gene Border Structure and Regulatory Elements on Gene Expression and Polymerase Scanning Behavior. *Journal of Virology, 90*(4), 1898-1909. https://doi.org/10.1128/JVI.02341-15

Brauburger, K., Boehmann, Y., Tsuda, Y., Hoenen, T., Olejnik, J., Schümann, M., Ebihara, H., & Mühlberger, E. (2014). Analysis of the highly diverse gene borders in Ebola virus reveals a distinct mechanism of transcriptional regulation. *Journal of Virology, 88*(21), 12558-12571. https://doi.org/10.1128/JVI.01863-14

Brauburger, K., Hume, A. J., Mühlberger, E., & Olejnik, J. (2012, 2012-10-01). Forty-Five Years of Marburg Virus Research. *Viruses, 4*(10), 1878-1927. https://doi.org/10.3390/v4101878

Bray, M. (2001, 2001-06-01). The role of the Type I interferon response in the resistance of mice to filovirus infection. *Journal of General Virology, 82*(6), 1365-1373. https://doi.org/10.1099/0022-1317-82-6-1365

Bray, M., Davis, K., Geisbert, T., Schmaljohn, C., & Huggins, J. (1999). A Mouse Model for Evaluation of Prophylaxis and Therapy of Ebola Hemorrhagic Fever. *The Journal of Infectious Diseases, 179*(Supplement_1), S248-S258. https://doi.org/10.1086/514292

Bray, M., Hatfill, S., Hensley, L., & Huggins, J. W. (2001, 2001/11/01/). Haematological, Biochemical and Coagulation Changes in Mice, Guinea-pigs and Monkeys Infected with a Mouse-adapted Variant of Ebola Zaire Virus. *Journal of Comparative Pathology, 125*(4), 243-253. https://doi.org/https://doi.org/10.1053/jcpa.2001.0503

Brecher, M., Schornberg, K. L., Delos, S. E., Fusco, M. L., Saphire, E. O., & White, J. M. (2012, 2012-01-01). Cathepsin Cleavage Potentiates the Ebola Virus Glycoprotein To Undergo a Subsequent Fusion-Relevant Conformational Change. *Journal of Virology, 86*(1), 364-372. https://doi.org/10.1128/jvi.05708-11

Breman, G., Joel, Johnson, M., Karl, Van Der Groen, G., Robbins, B., C., Szczeniowski, V., Mark, Ruti, K., Webb, A., Patricia, Meier, F., & Heymann, L., David. (1999, 1999-02-01). A Search for Ebola Virus in Animals in the Democratic Republic of the Congo and Cameroon: Ecologic, Virologic, and Serologic Surveys, 1979–1980. *The Journal of Infectious Diseases, 179*(s1), S139-S147. https://doi.org/10.1086/514278

Breman, J. G., Heymann, D. L., Lloyd, G., Mccormick, J. B., Miatudila, M., Murphy, F. A., Muyembé-Tamfun, J.-J., Piot, P., Ruppol, J.-F., Sureau, P., Van Der Groen, G., & Johnson, K. M. (2016, 2016-10-15). Discovery and Description of Ebola Zaire Virus in 1976 and Relevance to the West African Epidemic During 2013–2016. *Journal of Infectious Diseases, 214*(suppl 3), S93-S101. https://doi.org/10.1093/infdis/jiw207

Brindley, M. A., Hughes, L., Ruiz, A., Mccray, P. B., Sanchez, A., Sanders, D. A., & Maury, W. (2007, 2007-07-15). Ebola Virus Glycoprotein 1: Identification of Residues Important for Binding and Postbinding Events. *Journal of Virology, 81*(14), 7702-7709. https://doi.org/10.1128/jvi.02433-06

Burk, R., Bollinger, L., Johnson, J. C., Wada, J., Radoshitzky, S. R., Palacios, G., Bavari, S., Jahrling, P. B., & Kuhn, J. H. (2016). Neglected filoviruses. *FEMS Microbiology Reviews*. https://doi.org/10.1093/femsre/fuw010

Calain, P., Monroe, M. C., & Nichol, S. T. (1999, Sep 15). Ebola virus defective interfering particles and persistent infection. *Virology, 262*(1), 114-128. https://doi.org/10.1006/viro.1999.9915

Cantoni, D., Hamlet, A., Michaelis, M., Wass, M. N., & Rossman, J. S. (2016). Risks Posed by Reston, the Forgotten Ebolavirus. *mSphere*. https://doi.org/10.1128/msphere.00322-16

Carette, J. E., Raaben, M., Wong, A. C., Herbert, A. S., Obernosterer, G., Mulherkar, N., Kuehne, A. I., Kranzusch, P. J., Griffin, A. M., Ruthel, G., Cin, P. D., Dye, J. M., Whelan, S. P., Chandran, K., & Brummelkamp, T. R. (2011, 2011-09-01). Ebola virus entry requires the cholesterol transporter Niemann–Pick C1. *Nature, 477*(7364), 340-343. https://doi.org/10.1038/nature10348

Cenciarelli, O., Pietropaoli, S., Malizia, A., Carestia, M., D'Amico, F., Sassolini, A., Di Giovanni, D., Rea, S., Gabbarini, V., Tamburrini, A., Palombi, L., Bellecci, C., & Gaudio, P. (2015). Ebola virus disease 2013-2014 outbreak in west Africa: an analysis of the epidemic spread and response. *International journal of microbiology, 2015*, 769121-769121. https://doi.org/10.1155/2015/769121

Centers for Disease Control and Prevention. (1990, Apr 6). Update: filovirus infection in animal handlers. *MMWR Morb Mortal Wkly Rep, 39*(13), 221.

Centers for Disease Control and Prevention. (2019). *2014-2016 Ebola Outbreak in West Africa*. Retrieved March 21, 2022 from https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html

Centers for Disease Control and Prevention. (2020a). *FDA Approves First Treatment for Ebola Virus*. https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-ebola-virus

Centers for Disease Control and Prevention. (2020b). *FDA approves treatment for ebola virus*. Retrieved February 9 from https://www.fda.gov/drugs/news-events-human-drugs/fda-approves-treatment-ebola-virus

Centers for Disease Control and Prevention. (2021a). *Ebola (Ebola virus disease) Years of Ebola virus disease outbreaks*. https://www.cdc.gov/vhf/ebola/history/chronology.html

Centers for Disease Control and Prevention. (2021b, August 31, 2021). *History of Marburg Virus Disease (MVD) Outbreaks*. Retrieved October 31, 2021 from https://www.cdc.gov/vhf/marburg/outbreaks/chronology.html

Chan, M., Leung, A., Griffin, B. D., Vendramelli, R., Tailor, N., Tierney, K., Audet, J., & Kobasa, D. (2019). Generation and Characterization of a Mouse-Adapted Makona Variant of Ebola Virus. *Viruses, 11*(11), 987. https://doi.org/10.3390/v11110987

Chandran, K., Sullivan, N. J., Felbor, U., Whelan, S. P., & Cunningham, J. M. (2005, 2005-06-10). Endosomal Proteolysis of the Ebola Virus Glycoprotein Is Necessary for Infection. *Science, 308*(5728), 1643-1645. https://doi.org/10.1126/science.1110656

Chanthamontri, C. K., Jordan, D. S., Wang, W., Wu, C., Lin, Y., Brett, T. J., Gross, M. L., & Leung, D. W. (2019, 2019-02-12). The Ebola Viral Protein 35 N-Terminus Is a Parallel Tetramer. *Biochemistry, 58*(6), 657-664. https://doi.org/10.1021/acs.biochem.8b01154

Chen, J., He, Z., Yuan, Y., Huang, F., Luo, B., Zhang, J., Pan, T., Zhang, H., & Zhang, J. (2019, 2019-01-01). Host factor SMYD3 is recruited by Ebola virus nucleoprotein to facilitate viral mRNA transcription. *Emerging Microbes & Infections, 8*(1), 1347-1360. https://doi.org/10.1080/22221751.2019.1662736

Chenchik, A., Diachenko, L., Moqadam, F., Tarabykin, V., Lukyanov, S., & Siebert, P. D. (1996, 1996/09/01). Full-Length cDNA Cloning and Determination of mRNA 5′and 3′Ends by Amplification of Adaptor-Ligated cDNA. *BioTechniques, 21*(3), 526-534. https://doi.org/10.2144/96213pf02

Cheresiz, S. V., Semenova, E. A., & Chepurnov, A. A. (2016, 2016-01-01). Adapted Lethality: What We Can Learn from Guinea Pig-Adapted Ebola Virus Infection Model. *Advances in Virology, 2016*, 1-14. https://doi.org/10.1155/2016/8059607

Choi, M., Cossaboom, C., Whitesell, A., Dyal, J., Joyce, A., Morgan, R., Campos-Outcalt, D., Person, M., Ervin, E., Yu, Y., Rollin, P., Harcourt, B., Atmar, R., Bell, B., Helfand, R., Damon, I., & Frey, S. (2021, 01/08). Use of Ebola Vaccine: Recommendations of the

Advisory Committee on Immunization Practices, United States, 2020. *MMWR. Recommendations and Reports, 70*, 1-12. https://doi.org/10.15585/mmwr.rr7001a1

Clifton, M. C., Bruhn, J. F., Atkins, K., Webb, T. L., Baydo, R. O., Raymond, A., Lorimer, D. D., Edwards, T. E., Myler, P. J., & Saphire, E. O. (2015, 2015-10-01). High-resolution Crystal Structure of Dimeric VP40 FromSudan ebolavirus. *Journal of Infectious Diseases, 212*(suppl 2), S167-S171. https://doi.org/10.1093/infdis/jiv090

Coates, J. (2003, 2003-09-01). Armadillo repeat proteins: beyond the animal kingdom. *Trends in Cell Biology, 13*(9), 463-471. https://doi.org/10.1016/s0962-8924(03)00167-3

Connolly, M., Brett, Steele, E., Keith, Davis, J., Kelly, Geisbert, W., Thomas, Kell, M., Wayne, Jaax, K., Nancy, & Jahrling, B., Peter. (1999, 1999-02-01). Pathogenesis of Experimental Ebola Virus Infection in Guinea Pigs. *The Journal of Infectious Diseases, 179*(s1), S203-S217. https://doi.org/10.1086/514305

Cook, J. D., & Lee, J. E. (2013, 2013-05-16). The Secret Life of Viral Entry Glycoproteins: Moonlighting in Immune Evasion. *PLoS Pathogens, 9*(5), e1003258. https://doi.org/10.1371/journal.ppat.1003258

Côté, M., Misasi, J., Ren, T., Bruchez, A., Lee, K., Filone, C. M., Hensley, L., Li, Q., Ory, D., Chandran, K., & Cunningham, J. (2011, 2011-09-01). Small molecule inhibitors reveal Niemann–Pick C1 is essential for Ebola virus infection. *Nature, 477*(7364), 344-348. https://doi.org/10.1038/nature10380

Coutts, R. H. A., & Livieratos, I. C. (2003, 2003-09-01). A Rapid Method for Sequencing the 5'- and 3'-Termini of Double-Stranded RNA Viral Templates using RLM-RACE. *Journal of Phytopathology, 151*(9), 525-527. https://doi.org/10.1046/j.1439-0434.2003.00755.x

Crick, F. H., Barnett, L., Brenner, S., & Watts-Tobin, R. J. (1961, Dec 30). General nature of the genetic code for proteins. *Nature, 192*, 1227-1232. https://doi.org/10.1038/1921227a0

Cross, R. W., Fenton, K. A., Geisbert, J. B., Mire, C. E., & Geisbert, T. W. (2015, 2015-10-01). Modeling the Disease Course ofZaire ebolavirusInfection in the Outbred Guinea Pig. *Journal of Infectious Diseases, 212*(suppl 2), S305-S315. https://doi.org/10.1093/infdis/jiv237

Cross, R. W., Fenton, K. A., & Geisbert, T. W. (2018, 2018-11-02). Small animal models of filovirus disease: recent advances and future directions. *Expert Opinion on Drug Discovery, 13*(11), 1027-1040. https://doi.org/10.1080/17460441.2018.1527827

Cross, R. W., Mire, C. E., Borisevich, V., Geisbert, J. B., Fenton, K. A., & Geisbert, T. W. (2016, 2016-08-15). The Domestic Ferret (Mustela putorius furo) as a Lethal Infection Model for 3 Species ofEbolavirus. *Journal of Infectious Diseases, 214*(4), 565-569. https://doi.org/10.1093/infdis/jiw209

Cuevas, J. M., Domingo-Calap, P., & Sanjuán, R. (2012, 2012-01-01). The Fitness Effects of Synonymous Mutations in DNA and RNA Viruses. *Molecular Biology and Evolution, 29*(1), 17-20. https://doi.org/10.1093/molbev/msr179

Davey, R. A., Shtanko, O., Anantpadma, M., Sakurai, Y., Chandran, K., & Maury, W. (2017). Mechanisms of Filovirus Entry. In *Current Topics in Microbiology and Immunology* (pp. 323-352). Springer International Publishing. https://doi.org/10.1007/82_2017_14

Davidson, A., Birks, J., Griffiths, H., Kitchener, A., & Biggens, D. (1999). Hybridization and the phylogenetic relationship between polecats and the domestic ferret in Britain. *Biological Conservation, 87*, 155-161.

De La Vega, M.-A., Wong, G., Kobinger, G. P., & Qiu, X. (2015, 2015-02-01). The Multiple Roles of sGP in Ebola Pathogenesis. *Viral Immunology, 28*(1), 3-9. https://doi.org/10.1089/vim.2014.0068

Dessen, A. (2000, 2000-08-15). Crystal structure of the matrix protein VP40 from Ebola virus. *The EMBO Journal, 19*(16), 4228-4236. https://doi.org/10.1093/emboj/19.16.4228

Di Palma, F., Daino, G. L., Ramaswamy, V. K., Corona, A., Frau, A., Fanunza, E., Vargiu, A. V., Tramontano, E., & Ruggerone, P. (2019, 2019-01-01). Relevance of Ebola virus VP35 homo-dimerization on the type I interferon cascade inhibition. *Antiviral Chemistry and Chemotherapy, 27*, 204020661988922. https://doi.org/10.1177/2040206619889220

Domingo, E., Sheldon, J., & Perales, C. (2012). Viral quasispecies evolution. *Microbiology and molecular biology reviews : MMBR, 76*(2), 159-216. https://doi.org/10.1128/MMBR.05023-11

Dong, S., Yang, P., Li, G., Liu, B., Wang, W., Liu, X., Xia, B., Yang, C., Lou, Z., Guo, Y., & Rao, Z. (2015, 2015-05-01). Insight into the Ebola virus nucleocapsid assembly mechanism: crystal structure of Ebola virus nucleoprotein core domain at 1.8 Å resolution. *Protein & Cell, 6*(5), 351-362. https://doi.org/10.1007/s13238-015-0163-3

Dudley, D. M., Newman, C. M., Lalli, J., Stewart, L. M., Koenig, M. R., Weiler, A. M., Semler, M. R., Barry, G. L., Zarbock, K. R., Mohns, M. S., Breitbach, M. E., Schultz-Darken, N., Peterson, E., Newton, W., Mohr, E. L., Capuano Iii, S., Osorio, J. E., O'Connor, S. L., O'Connor, D. H., Friedrich, T. C., & Aliota, M. T. (2017, 2017/12/13). Infection via mosquito bite alters Zika virus tissue tropism and replication kinetics in rhesus macaques. *Nature Communications, 8*(1), 2096. https://doi.org/10.1038/s41467-017-02222-8

Duret, L. (2008). Neutral theory: The null hypothesis of molecular evolution. *Nature Education, 1*(1), 218.

Dutkiewicz, M., Kuczynski, J., Jarzab, M., Stachowiak, A., & Swiatkowska, A. (2020). New RNA Structural Elements Identified in the Coding Region of the Coxsackie B3 Virus Genome. *Viruses, 12*(11), 1232. https://doi.org/10.3390/v12111232

Ebbert, M. T. W., Wadsworth, M. E., Staley, L. A., Hoyt, K. L., Pickett, B., Miller, J., Duce, J., Kauwe, J. S. K., Ridge, P. G., & for the Alzheimer's Disease Neuroimaging, I. (2016, 2016/07/25). Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics, 17*(7), 239. https://doi.org/10.1186/s12859-016-1097-3

Ebihara, H., Takada, A., Kobasa, D., Jones, S., Neumann, G., Theriault, S., Bray, M., Feldmann, H., & Kawaoka, Y. (2006, 2006-07-21). Molecular Determinants of Ebola Virus Virulence in Mice. *PLoS Pathogens, 2*(7), e73. https://doi.org/10.1371/journal.ppat.0020073

Ebihara, H., Zivcec, M., Gardner, D., Falzarano, D., Lacasse, R., Rosenke, R., Long, D., Haddock, E., Fischer, E., Kawaoka, Y., & Feldmann, H. (2013, 2013-01-15). A Syrian Golden Hamster Model Recapitulating Ebola Hemorrhagic Fever. *The Journal of Infectious Diseases, 207*(2), 306-318. https://doi.org/10.1093/infdis/jis626

Elliott, L. H., Kiley, M. P., & McCormick, J. B. (1985, 1985/11/01/). Descriptive analysis of Ebola virus proteins. *Virology, 147*(1), 169-176. https://doi.org/https://doi.org/10.1016/0042-6822(85)90236-3

Empig, C. J., & Goldsmith, M. A. (2002). Association of the Caveola Vesicular System with Cellular Entry by Filoviruses. *Journal of Virology, 76*(10), 5266-5270. https://doi.org/doi:10.1128/JVI.76.10.5266-5270.2002

Ericsson, A. C., Crim, M. J., & Franklin, C. L. (2013, May-Jun). A brief history of animal modeling. *Missouri medicine, 110*(3), 201-205. https://pubmed.ncbi.nlm.nih.gov/23829102
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3979591/

Escaffre, O., Juelich, T. L., Neef, N., Massey, S., Smith, J., Brasel, T., Smith, J. K., Kalveram, B., Zhang, L., Perez, D., Ikegami, T., Freiberg, A. N., & Comer, J. E. (2021, Jul 17). STAT-1 Knockout Mice as a Model for Wild-Type Sudan Virus (SUDV). *Viruses, 13*(7). https://doi.org/10.3390/v13071388

Fabozzi, G., Nabel, C. S., Dolan, M. A., & Sullivan, N. J. (2011, 2011-03-15). Ebolavirus Proteins Suppress the Effects of Small Interfering RNA by Direct Interaction with the Mammalian RNA Interference Pathway. *Journal of Virology, 85*(6), 2512-2523. https://doi.org/10.1128/jvi.01160-10

Falzarano, D., Krokhin, O., Wahl-Jensen, V., Seebach, J., Wolf, K., Schnittler, H.-J., & Feldmann, H. (2006, 2006-10-06). Structure-Function Analysis of the Soluble Glycoprotein, sGP, of Ebola Virus. *ChemBioChem, 7*(10), 1605-1611. https://doi.org/10.1002/cbic.200600223

Faria, N. R., Quick, J., Claro, I. M., Thézé, J., de Jesus, J. G., Giovanetti, M., Kraemer, M. U. G., Hill, S. C., Black, A., da Costa, A. C., Franco, L. C., Silva, S. P., Wu, C. H., Raghwani, J., Cauchemez, S., du Plessis, L., Verotti, M. P., de Oliveira, W. K., Carmo, E. H., Coelho, G. E., Santelli, A. C. F. S., Vinhal, L. C., Henriques, C. M., Simpson, J. T., Loose, M., Andersen, K. G., Grubaugh, N. D., Somasekar, S., Chiu, C. Y., Muñoz-Medina, J. E., Gonzalez-Bonilla, C. R., Arias, C. F., Lewis-Ximenez, L. L., Baylis, S. A., Chieppe, A. O., Aguiar, S. F., Fernandes, C. A., Lemos, P. S., Nascimento, B. L. S., Monteiro, H. A. O., Siqueira, I. C., de Queiroz, M. G., de Souza, T. R., Bezerra, J. F., Lemos, M. R., Pereira, G. F., Loudal, D., Moura, L. C., Dhalia, R., França, R. F., Magalhães, T., Marques, E. T., Jaenisch, T., Wallau, G. L., de Lima, M. C., Nascimento, V., de Cerqueira, E. M., de Lima, M. M., Mascarenhas, D. L., Neto, J. P. M., Levin, A. S., Tozetto-Mendoza, T. R., Fonseca, S. N., Mendes-Correa, M. C., Milagres, F. P., Segurado, A., Holmes, E. C., Rambaut, A., Bedford, T., Nunes, M. R. T., Sabino, E. C., Alcantara, L. C. J., Loman, N. J., & Pybus, O. G. (2017, 2017/06/01). Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature, 546*(7658), 406-410. https://doi.org/10.1038/nature22401

Feldmann, H., Nichol, S. T., Klenk, H. D., Peters, C. J., & Sanchez, A. (1994, Mar). Characterization of filoviruses based on differences in structure and antigenicity of the virion glycoprotein. *Virology, 199*(2), 469-473. https://doi.org/10.1006/viro.1994.1147

Feldmann, H., Slenczka, W., & Klenk, H.-D. (1996). Emerging and reemerging of filoviruses. In *Imported Virus Infections* (pp. 77-100). Springer Vienna. https://doi.org/10.1007/978-3-7091-7482-1_9

Forbes, K. M., Webala, P. W., Jääskeläinen, A. J., Abdurahman, S., Ogola, J., Masika, M. M., Kivistö, I., Alburkat, H., Plyusnin, I., Levanov, L., Korhonen, E. M., Huhtamo, E., Mwaengo, D., Smura, T., Mirazimi, A., Anzala, O., Vapalahti, O., & Sironen, T. (2019, 2019-05-01). Bombali Virus in Mops condylurus Bat, Kenya. *Emerging Infectious Diseases, 25*(5). https://doi.org/10.3201/eid2505.181666

Francica, J. R., Varela-Rohena, A., Medvec, A., Plesa, G., Riley, J. L., & Bates, P. (2010, 2010-09-09). Steric Shielding of Surface Epitopes and Impaired Immune Recognition Induced by the Ebola Virus Glycoprotein. *PLoS Pathogens, 6*(9), e1001098. https://doi.org/10.1371/journal.ppat.1001098

Frolov, I., & Schlesinger, S. (1996). Translation of Sindbis virus mRNA: analysis of sequences downstream of the initiating AUG codon that enhance translation. *Journal of Virology, 70*(2), 1182-1190. https://doi.org/10.1128/JVI.70.2.1182-1190.1996

Furuyama, W., & Marzi, A. (2019, Sep 29). Ebola Virus: Pathogenesis and Countermeasure Development. *Annu Rev Virol, 6*(1), 435-458. https://doi.org/10.1146/annurev-virology-092818-015708

Gallaher, W., & Garry, R. (2015, 2015-01-20). Modeling of the Ebola Virus Delta Peptide Reveals a Potential Lytic Sequence Motif. *Viruses, 7*(1), 285-305. https://doi.org/10.3390/v7010285

Gc, J. B., Gerstman, B. S., Stahelin, R. V., & Chapagain, P. P. (2016). The Ebola virus protein VP40 hexamer enhances the clustering of PI(4,5)P(2) lipids in the plasma membrane. *Physical chemistry chemical physics : PCCP, 18*(41), 28409-28417. https://doi.org/10.1039/c6cp03776c

Geisbert, T. W., Hensley, L. E., Larsen, T., Young, H. A., Reed, D. S., Geisbert, J. B., Scott, D. P., Kagan, E., Jahrling, P. B., & Davis, K. J. (2003, 2003-12-01). Pathogenesis of Ebola Hemorrhagic Fever in Cynomolgus Macaques. *The American Journal of Pathology, 163*(6), 2347-2370. https://doi.org/10.1016/s0002-9440(10)63591-2

Geisbert, T. W., & Jahrling, P. B. (1990, 1990-10-01). Use of immunoelectron microscopy to show Ebola virus during the 1989 United States epizootic. *Journal of Clinical Pathology, 43*(10), 813-816. https://doi.org/10.1136/jcp.43.10.813

Geisbert, T. W., & Jahrling, P. B. (1995, 1995-12-01). Differentiation of filoviruses by electron microscopy. *Virus Research, 39*(2-3), 129-150. https://doi.org/10.1016/0168-1702(95)00080-1

Gibb, T. R., Bray, M., Geisbert, T. W., Steele, K. E., Kell, W. M., Davis, K. J., & Jaax, N. K. (2001, 2001/11/01/). Pathogenesis of Experimental Ebola Zaire Virus Infection in BALB/c Mice. *Journal of Comparative Pathology, 125*(4), 233-242. https://doi.org/https://doi.org/10.1053/jcpa.2001.0502

Goldstein, T., Anthony, S. J., Gbakima, A., Bird, B. H., Bangura, J., Tremeau-Bravard, A., Belaganahalli, M. N., Wells, H. L., Dhanota, J. K., Liang, E., Grodus, M., Jangra, R. K., DeJesus, V. A., Lasso, G., Smith, B. R., Jambai, A., Kamara, B. O., Kamara, S., Bangura, W., Monagin, C., Shapira, S., Johnson, C. K., Saylors, K., Rubin, E. M., Chandran, K., Lipkin, W. I., & Mazet, J. A. K. (2018). The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. *Nature Microbiology, 3*(10), 1084-1089. https://doi.org/10.1038/s41564-018-0227-2

Groseth, A., Charton, J. E., Sauerborn, M., Feldmann, F., Jones, S. M., Hoenen, T., & Feldmann, H. (2009, 2009-03-01). The Ebola virus ribonucleoprotein complex: A novel VP30–L interaction identified. *Virus Research, 140*(1-2), 8-14. https://doi.org/10.1016/j.virusres.2008.10.017

Grubaugh, N. D., Ladner, J. T., Kraemer, M. U., Dudas, G., Tan, A. L., Gangavarapu, K., Wiley, M. R., White, S., Thézé, J., & Magnani, D. M. (2017). Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature, 546*(7658), 401-405.

Halfmann, P., Neumann, G., & Kawaoka, Y. (2011, 2011-11-01). The Ebolavirus VP24 Protein Blocks Phosphorylation of p38 Mitogen-Activated Protein Kinase. *The Journal of Infectious Diseases, 204*(suppl_3), S953-S956. https://doi.org/10.1093/infdis/jir325

Hall, N. (2007, 2007-05-01). Advanced sequencing technologies and their wider impact in microbiology. *Journal of Experimental Biology, 210*(9), 1518-1525. https://doi.org/10.1242/jeb.001370

Han, H.-J., Wen, H.-l., Zhou, C.-M., Chen, F.-F., Luo, L.-M., Liu, J.-w., & Yu, X.-J. (2015). Bats as reservoirs of severe emerging infectious diseases. *Virus Research, 205*, 1-6. https://doi.org/10.1016/j.virusres.2015.05.006

Hartlieb, B., Modrof, J., Mühlberger, E., Klenk, H.-D., & Becker, S. (2003, 2003-10-01). Oligomerization of Ebola Virus VP30 Is Essential for Viral Transcription and Can Be Inhibited by a Synthetic Peptide. *Journal of Biological Chemistry, 278*(43), 41830-41836. https://doi.org/10.1074/jbc.m307036200

Hartlieb, B., Muziol, T., Weissenhorn, W., & Becker, S. (2007, 2007-01-09). Crystal structure of the C-terminal domain of Ebola virus VP30 reveals a role in transcription and nucleocapsid association. *Proceedings of the National Academy of Sciences, 104*(2), 624-629. https://doi.org/10.1073/pnas.0606730104

Haydon, D. T., Cleaveland, S., Taylor, L. H., & Laurenson, M. K. (2002). Identifying reservoirs of infection: a conceptual and practical challenge. *Emerging Infectious Diseases, 8*(12), 1468-1473. https://doi.org/10.3201/eid0812.010317

Hayes, C. G., Miranda, M. E. G., Robles, C. G., Peters, C. J., Ksiazek, T. G., Dayrit, M. M., Manaloto, C. R., Barrientos, A. B., Del Rosario, R. A., & Burans, J. P. (1992, 1992-06-01). Outbreak of Fatal Illness among Captive Macaques in the Philippines Caused by an Ebola-Related Filovirus. *The American Journal of Tropical Medicine and Hygiene, 46*(6), 664-671. https://doi.org/10.4269/ajtmh.1992.46.664

He, B., Feng, Y., Zhang, H., Xu, L., Yang, W., Zhang, Y., Li, X., & Tu, C. (2015, 2015-09-01). Filovirus RNA in Fruit Bats, China. *Emerging Infectious Diseases, 21*(9), 1675-1677. https://doi.org/10.3201/eid2109.150260

He, J., Melnik, L. I., Komin, A., Wiedman, G., Fuselier, T., Morris, C. F., Starr, C. G., Searson, P. C., Gallaher, W. R., Hristova, K., Garry, R. F., & Wimley, W. C. (2017, 2017-08-15). Ebola Virus Delta Peptide Is a Viroporin. *Journal of Virology, 91*(16), JVI.00438-00417. https://doi.org/10.1128/jvi.00438-17

Head, S. R., Kiyomi Komori, H., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*. https://doi.org/10.2144/000114133

Hepp, C. M., Cocking, J. H., Valentine, M., Young, S. J., Damian, D., Samuels-Crow, K. E., Sheridan, K., Fofanov, V. Y., Furstenau, T. N., Busch, J. D., Erickson, D. E., Lancione, R. C., Smith, K., Will, J., Townsend, J., Keim, P. S., & Engelthaler, D. M. (2018). Phylogenetic analysis of West Nile Virus in Maricopa County, Arizona: Evidence for dynamic behavior of strains in two major lineages in the American Southwest. *PLoS One, 13*(11), e0205801. https://doi.org/10.1371/journal.pone.0205801

Herbert, A. S., Froude, J. W., Ortiz, R. A., Kuehne, A. I., Dorosky, D. E., Bakken, R. R., Zak, S. E., Josleyn, N. M., Musiychuk, K., Jones, R. M., Green, B., Streatfield, S. J., Wec, A. Z., Bohorova, N., Bohorov, O., Kim, D. H., Pauly, M. H., Velasco, J., Whaley, K. J., Stonier, S. W., Bornholdt, Z. A., Chandran, K., Zeitlin, L., Sampey, D., Yusibov, V., & Dye, J. M. (2020). Development of an antibody cocktail for treatment of Sudan virus infection. *Proceedings of the National Academy of Sciences, 117*(7), 3768-3778. https://doi.org/doi:10.1073/pnas.1914985117

Hernandez-Alias, X., Benisty, H., Schaefer, M. H., & Serrano, L. (2021, Mar 16). Translational adaptation of human viruses to the tissues they infect. *Cell Rep, 34*(11), 108872. https://doi.org/10.1016/j.celrep.2021.108872

Hierweger, M. M., Koch, M. C., Rupp, M., Maes, P., Di Paola, N., Bruggmann, R., Kuhn, J. H., Schmidt-Posthaus, H., & Seuberlich, T. (2021). Novel Filoviruses, Hantavirus, and Rhabdovirus in Freshwater Fish, Switzerland, 2017. *Emerging Infectious Diseases, 27*(12), 3082-3091. https://doi.org/10.3201/eid2712.210491

Hoenen, T., Biedenkopf, N., Zielecki, F., Jung, S., Groseth, A., Feldmann, H., & Becker, S. (2010, 2010-07-15). Oligomerization of Ebola Virus VP40 Is Essential for Particle Morphogenesis and Regulation of Viral Transcription. *Journal of Virology, 84*(14), 7053-7063. https://doi.org/10.1128/jvi.00737-10

Hoenen, T., Volchkov, V., Kolesnikova, L., Mittler, E., Timmins, J., Ottmann, M., Reynard, O., Becker, S., & Weissenhorn, W. (2005, 2005-02-01). VP40 Octamers Are Essential for Ebola Virus Replication. *Journal of Virology, 79*(3), 1898-1905. https://doi.org/10.1128/jvi.79.3.1898-1905.2005

Huang, Y., Xu, L., Sun, Y., & Nabel, G. J. (2002, 2002-08-01). The Assembly of Ebola Virus Nucleocapsid Requires Virion-Associated Proteins 35 and 24 and Posttranslational Modification of Nucleoprotein. *Molecular Cell, 10*(2), 307-316. https://doi.org/10.1016/s1097-2765(02)00588-9

Hulseberg, C. E., Kumar, R., Di Paola, N., Larson, P., Nagle, E. R., Richardson, J., Hanson, J., Wauquier, N., Fair, J. N., Makuwa, M., Mulembakani, P., Muyembe-Tamfum, J.-J., Schoepp, R. J., Sanchez-Lockhart, M., Palacios, G. F., Kuhn, J. H., & Kugelman, J. R. (2021, 2021-08-01). Molecular analysis of the 2012 Bundibugyo virus disease outbreak. *Cell Reports Medicine, 2*(8), 100351. https://doi.org/10.1016/j.xcrm.2021.100351

Hume, A. J., & Mühlberger, E. (2019). Distinct Genome Replication and Transcription Strategies within the Growing Filovirus Family. *Journal of molecular biology, 431*(21), 4290-4320. https://doi.org/10.1016/j.jmb.2019.06.029

Ilinykh, P. A., Lubaki, N. M., Widen, S. G., Renn, L. A., Theisen, T. C., Rabin, R. L., Wood, T. G., & Bukreyev, A. (2015). Different Temporal Effects of Ebola Virus VP35 and VP24 Proteins on Global Gene Expression in Human Dendritic Cells. *Journal of Virology*. https://doi.org/10.1128/jvi.00924-15

Ilinykh, P. A., Tigabu, B., Ivanov, A., Ammosova, T., Obukhov, Y., Garron, T., Kumari, N., Kovalskyy, D., Platonov, M. O., Naumchik, V. S., Freiberg, A. N., Nekhai, S., & Bukreyev, A. (2014, 2014-08-01). Role of Protein Phosphatase 1 in Dephosphorylation of Ebola Virus VP30 Protein and Its Targeting for the Inhibition of Viral Transcription. *Journal of Biological Chemistry, 289*(33), 22723-22738. https://doi.org/10.1074/jbc.m114.575050

Illumina. (2022a). *Illumina sequencing platforms*. Retrieved February from https://www.illumina.com/systems/sequencing-platforms.html

Illumina. (2022b). *Sequencing Coverage for NGS Experiments*. Retrieved March 21, 2022 from https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/coverage.html

Illumina. (2022c). *A wide breadth of sequencing applications*. Retrieved February 26 from https://www.illumina.com/systems/sequencing-platforms/miseq/applications.html

Irving, A. T., Ahn, M., Goh, G., Anderson, D. E., & Wang, L.-F. (2021, 2021/01/01). Lessons from the host defences of bats, a unique viral reservoir. *Nature, 589*(7842), 363-370. https://doi.org/10.1038/s41586-020-03128-0

Ivanov, A., Ramanathan, P., Parry, C., Ilinykh, P. A., Lin, X., Petukhov, M., Obukhov, Y., Ammosova, T., Amarasinghe, G. K., Bukreyev, A., & Nekhai, S. (2020, 2020-07-01). Global phosphoproteomic analysis of Ebola virions reveals a novel role for VP35 phosphorylation-dependent regulation of genome transcription. *Cellular and Molecular Life Sciences, 77*(13), 2579-2603. https://doi.org/10.1007/s00018-019-03303-1

Iwasa, A., Shimojima, M., & Kawaoka, Y. (2011, 2011-11-01). sGP Serves as a Structural Protein in Ebola Virus Infection. *The Journal of Infectious Diseases, 204*(suppl_3), S897-S903. https://doi.org/10.1093/infdis/jir313

Jacob, S. T., Crozier, I., Fischer, W. A., Hewlett, A., Kraft, C. S., Vega, M.-A. D. L., Soka, M. J., Wahl, V., Griffiths, A., Bollinger, L., & Kuhn, J. H. (2020, 2020-12-01). Ebola virus disease. *Nature Reviews Disease Primers, 6*(1). https://doi.org/10.1038/s41572-020-0147-3

Jain, S., Martynova, E., Rizvanov, A., Khaiboullina, S., & Baranwal, M. (2021, 2021-10-15). Structural and Functional Aspects of Ebola Virus Proteins. *Pathogens, 10*(10), 1330. https://doi.org/10.3390/pathogens10101330

Javan, G. T., Hanson, E., Finley, S. J., Visonà, S. D., Osculati, A., & Ballantyne, J. (2020, 2020-12-01). Identification of cadaveric liver tissues using thanatotranscriptome biomarkers. *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-63727-9

Jeffers, S. A., Sanders, D. A., & Sanchez, A. (2002). Covalent Modifications of the Ebola Virus Glycoprotein. *Journal of Virology, 76*(24), 12463-12472. https://doi.org/doi:10.1128/JVI.76.24.12463-12472.2002

Jitobaom, K., Phakaratsakul, S., Sirihongthong, T., Chotewutmontri, S., Suriyaphol, P., Suptawiwat, O., & Auewarakul, P. (2020). Codon usage similarity between viral and some host genes suggests a codon-specific translational regulation. *Heliyon, 6*(5), e03915-e03915. https://doi.org/10.1016/j.heliyon.2020.e03915

John, S. P., Wang, T., Steffen, S., Longhi, S., Schmaljohn, C. S., & Jonsson, C. B. (2007). Ebola Virus VP30 Is an RNA Binding Protein. *Journal of Virology, 81*(17), 8967-8976. https://doi.org/doi:10.1128/JVI.02523-06

Johnson, E. D., Johnson, B. K., Silverstein, D., Tukei, P., Geisbert, T. W., Sanchez, A. N., & Jahrling, P. B. (1996). Characterization of a new Marburg virus isolated from a 1987 fatal case in Kenya. In *Imported Virus Infections* (pp. 101-114). Springer Vienna. https://doi.org/10.1007/978-3-7091-7482-1_10

Jones, M., Schuh, A., Amman, B., Sealy, T., Zaki, S., Nichol, S., & Towner, J. (2015, 2015-06-25). Experimental Inoculation of Egyptian Rousette Bats (Rousettus aegyptiacus) with Viruses of the Ebolavirus and Marburgvirus Genera. *Viruses, 7*(7), 3420-3442. https://doi.org/10.3390/v7072779

Karan, L. S., Makenov, M. T., Korneev, M. G., Sacko, N., Boumbaly, S., Yakovlev, S. A., Kourouma, K., Bayandin, R. B., Gladysheva, A. V., Shipovalov, A. V., Yurganova, I. A., Grigorieva, Y. E., Fedorova, M. V., Scherbakova, S. A., Kutyrev, V. V., Agafonov, A. P., Maksyutov, R. A., Shipulin, G. A., Maleev, V. V., Boiro, M., Akimkin, V. G., & Popova, A. Y. (2019, 2019-09-01). Bombali Virus in Mops condylurus Bats, Guinea. *Emerging Infectious Diseases, 25*(9). https://doi.org/10.3201/eid2509.190581

Kareinen, L., Ogola, J., Kivistö, I., Smura, T., Aaltonen, K., Jääskeläinen, A. J., Kibiwot, S., Masika, M. M., Nyaga, P., Mwaengo, D., Anzala, O., Vapalahti, O., Webala, P. W., Forbes, K. M., & Sironen, T. (2020, 2020-12-01). Range Expansion of Bombali Virus in Mops condylurus Bats, Kenya, 2019. *Emerging Infectious Diseases, 26*(12), 3007-3010. https://doi.org/10.3201/eid2612.202925

Kemenesi, G., Kurucz, K., Dallos, B., Zana, B., Földes, F., Boldogh, S., Görföl, T., Carroll, M. W., & Jakab, F. (2018, 2018-12-01). Re-emergence of Lloviu virus in Miniopterus

schreibersii bats, Hungary, 2016. *Emerging Microbes & Infections, 7*(1), 1-4. https://doi.org/10.1038/s41426-018-0067-4

Khadka, S., Williams, C. G., Sweeney-Gibbons, J., Basler, C. F., & Heise, M. T. (2021). Marburg and Ebola Virus mRNA 3′ Untranslated Regions Contain Negative Regulators of Translation That Are Modulated by ADAR1 Editing. *Journal of Virology, 95*(19), e00652-00621. https://doi.org/doi:10.1128/JVI.00652-21

Kiening, M., Ochsenreiter, R., Hellinger, H.-J., Rattei, T., Hofacker, I., & Frishman, D. (2019). Conserved Secondary Structures in Viral mRNAs. *Viruses, 11*(5), 401. https://doi.org/10.3390/v11050401

Kim, Y.-I., Kim, S.-G., Kim, S.-M., Kim, E.-H., Park, S.-J., Yu, K.-M., Chang, J.-H., Kim, E. J., Lee, S., Casel, M. A. B., Um, J., Song, M.-S., Jeong, H. W., Lai, V. D., Kim, Y., Chin, B. S., Park, J.-S., Chung, K.-H., Foo, S.-S., Poo, H., Mo, I.-P., Lee, O.-J., Webby, R. J., Jung, J. U., & Choi, Y. K. (2020, 2020-05-01). Infection and Rapid Transmission of SARS-CoV-2 in Ferrets. *Cell Host & Microbe, 27*(5), 704-709.e702. https://doi.org/10.1016/j.chom.2020.03.023

Kimberlin, C. R., Bornholdt, Z. A., Li, S., Woods, V. L., Jr., MacRae, I. J., & Saphire, E. O. (2010, Jan 5). Ebolavirus VP35 uses a bimodal strategy to bind dsRNA for innate immune suppression. *Proc Natl Acad Sci U S A, 107*(1), 314-319. https://doi.org/10.1073/pnas.0910547107

Kindzelskii, A. L., Yang, Z.-Y., Nabel, G. J., Todd, R. F., & Petty, H. R. (2000, 2000-01-15). Ebola Virus Secretory Glycoprotein (sGP) Diminishes FcγRIIIB-to-CR3 Proximity on Neutrophils. *The Journal of Immunology, 164*(2), 953-958. https://doi.org/10.4049/jimmunol.164.2.953

Kirchdoerfer, N., Robert, Abelson, M., Dafna, Li, S., Wood, R., Malcolm, & Saphire, O., Erica. (2015, 2015-07-01). Assembly of the Ebola Virus Nucleoprotein from a Chaperoned VP35 Complex. *Cell Reports, 12*(1), 140-149. https://doi.org/10.1016/j.celrep.2015.06.003

Kirchdoerfer, R. N., Moyer, C. L., Abelson, D. M., & Saphire, E. O. (2016, 2016-10-18). The Ebola Virus VP30-NP Interaction Is a Regulator of Viral RNA Synthesis. *PLoS Pathogens, 12*(10), e1005937. https://doi.org/10.1371/journal.ppat.1005937

Kirchdoerfer, R. N., Wasserman, H., Amarasinghe, G. K., & Saphire, E. O. (2017). Filovirus Structural Biology: The Molecules in the Machine. In E. Mühlberger, L. L. Hensley, & J. S. Towner (Eds.), *Marburg- and Ebolaviruses: From Ecosystems to Molecules* (pp. 381-417). Springer International Publishing. https://doi.org/10.1007/82_2017_16

Kissling, R. E., Robinson, R. Q., Murphy, F. A., & Whitfield, S. G. (1968). Agent of Disease Contracted from Green Monkeys. *Science (American Association for the Advancement of Science), 160*(3830), 888-890. https://doi.org/10.1126/science.160.3830.888

Kozak, R., He, S., Kroeker, A., Vega, M.-A. d. L., Audet, J., Wong, G., Urfano, C., Antonation, K., Embury-Hyatt, C., Kobinger, G. P., Qiu, X., & Perlman, S. (2016). Ferrets Infected with Bundibugyo Virus or Ebola Virus Recapitulate Important Aspects of Human Filovirus Disease. *Journal of Virology, 90*(20), 9209-9223. https://doi.org/doi:10.1128/JVI.01033-16

Kratz, T., Roddy, P., Tshomba Oloma, A., Jeffs, B., Pou Ciruelo, D., De La Rosa, O., & Borchert, M. (2015, 2015-06-24). Ebola Virus Disease Outbreak in Isiro, Democratic Republic of the Congo, 2012: Signs and Symptoms, Management and Outcomes. *PLoS One, 10*(6), e0129333. https://doi.org/10.1371/journal.pone.0129333

Kroeker, A., He, S., de La Vega, M. A., Wong, G., Embury-Hyatt, C., & Qiu, X. (2017, Jul 11). Characterization of Sudan Ebolavirus infection in ferrets. *Oncotarget, 8*(28), 46262-46272. https://doi.org/10.18632/oncotarget.17694

Kruse, T., Biedenkopf, N., Hertz, E. P. T., Dietzel, E., Stalmann, G., López-Méndez, B., Davey, N. E., Nilsson, J., & Becker, S. (2018, 2018-01-01). The Ebola Virus Nucleoprotein Recruits the Host PP2A-B56 Phosphatase to Activate Transcriptional Support Activity of VP30. *Molecular Cell, 69*(1), 136-145.e136. https://doi.org/10.1016/j.molcel.2017.11.034

Kühl, A., Banning, C., Marzi, A., Votteler, J., Steffen, I., Bertram, S., Glowacka, I., Konrad, A., Stürzl, M., Guo, J.-T., Schubert, U., Feldmann, H., Behrens, G., Schindler, M., & Pöhlmann, S. (2011, 2011-11-01). The Ebola Virus Glycoprotein and HIV-1 Vpu Employ Different Strategies to Counteract the Antiviral Factor Tetherin. *The Journal of Infectious Diseases, 204*(suppl_3), S850-S860. https://doi.org/10.1093/infdis/jir378

Kuhn, J. H., Amarasinghe, G. K., Basler, C. F., Bavari, S., Bukreyev, A., Chandran, K., Crozier, I., Dolnik, O., Dye, J. M., Formenty, P. B. H., Griffiths, A., Hewson, R., Kobinger, G. P., Leroy, E. M., Mühlberger, E., Netesov , S. V., Palacios, G., Pályi, B., Pawęska, J. T., Smither, S. J., Takada , A., Towner, J. S., Wahl, V., & Consortium, I. R. (2020). ICTV Virus Taxonomy Profile: Filoviridae. *Journal of General Virology, 100*(6), 911-912. https://doi.org/https://doi.org/10.1099/jgv.0.001252

Kuhn, J. H., Radoshitzky, S. R., Guth, A. C., Warfield, K. L., Li, W., Vincent, M. J., Towner, J. S., Nichol, S. T., Bavari, S., Choe, H., Aman, M. J., & Farzan, M. (2006, 2006-06-01). Conserved Receptor-binding Domains of Lake Victoria Marburgvirus and Zaire Ebolavirus Bind a Common Receptor. *Journal of Biological Chemistry, 281*(23), 15951-15958. https://doi.org/10.1074/jbc.m601796200

Kunz, C., Hofmann, H., Kovac, W., & Stockinger, L. (1968, Mar 1). [Biological and morphological characteristics of the "hemorrhagic fever" virus occurring in Germany]. *Wien Klin Wochenschr, 80*(9), 161-162 passim. (Biologische und morphologische Charakteristika des Virus des in Deutschland aufgetretenen "Hämorrhagischen Fiebers".)

Kuroda, M., Fujikura, D., Nanbo, A., Marzi, A., Noyori, O., Kajihara, M., Maruyama, J., Matsuno, K., Miyamoto, H., Yoshida, R., Feldmann, H., & Takada, A. (2015, 2015-06-15). Interaction between TIM-1 and NPC1 Is Important for Cellular Entry of Ebola Virus. *Journal of Virology, 89*(12), 6481-6493. https://doi.org/10.1128/jvi.03156-14

Landeras-Bueno, S., Oda, S.-I., Norris, M. J., Li Salie, Z., Guenaga, J., Wyatt, R. T., & Saphire, E. O. (2019, 2019-08-27). Sudan Ebolavirus VP35-NP Crystal Structure Reveals a Potential Target for Pan-Filovirus Treatment. *mBio, 10*(4). https://doi.org/10.1128/mbio.00734-19

Landeras-Bueno, S., Wasserman, H., Oliveira, G., Vanaernum, Z. L., Busch, F., Salie, Z. L., Wysocki, V. H., Andersen, K., & Saphire, E. O. (2021, 2021-04-01). Cellular mRNA triggers structural transformation of Ebola virus matrix protein VP40 to its essential regulatory form. *Cell Reports, 35*(2), 108986. https://doi.org/10.1016/j.celrep.2021.108986

Lanini, S., Portella, G., Vairo, F., Kobinger, G. P., Pesenti, A., Langer, M., Kabia, S., Brogiato, G., Amone, J., Castilletti, C., Miccio, R., Zumla, A., Capobianchi, M. R., Di Caro, A., Strada, G., & Ippolito, G. (2015, Dec). Blood kinetics of Ebola virus in survivors and nonsurvivors. *J Clin Invest, 125*(12), 4692-4698. https://doi.org/10.1172/jci83111

Le Guenno, B., & Formentry, P. (1995). Isolation and partial characterisation of a new strain of Ebola virus [Article]. *Lancet, 345*(8960), 1271. https://doi.org/10.1016/S0140-6736(95)90925-7

Le Sage, V., Cinti, A., McCarthy, S., Amorim, R., Rao, S., Daino, G. L., Tramontano, E., Branch, D. R., & Mouland, A. J. (2017, Feb). Ebola virus VP35 blocks stress granule assembly. *Virology, 502*, 73-83. https://doi.org/10.1016/j.virol.2016.12.012

Lee, J. E., Fusco, M. L., Hessell, A. J., Oswald, W. B., Burton, D. R., & Saphire, E. O. (2008, 2008-07-01). Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature, 454*(7201), 177-182. https://doi.org/10.1038/nature07082

Lee, J. E., & Saphire, E. O. (2009, 2009-11-01). Ebolavirus glycoprotein structure and mechanism of entry. *Future Virology, 4*(6), 621-635. https://doi.org/10.2217/fvl.09.56

Leendertz, S. A. J., Gogarten, J. F., Düx, A., Calvignac-Spencer, S., & Leendertz, F. H. (2016). Assessing the Evidence Supporting Fruit Bats as the Primary Reservoirs for Ebola Viruses. *EcoHealth, 13*(1), 18-25. https://doi.org/10.1007/s10393-015-1053-0

Leirs, H., Mills, J. N., Krebs, J. W., Childs, J. E., Akaibe, D., Woollen, N., Ludwig, G., Peters, C. J., Ksiazek, T. G., & members, o. s. g. (1999). Search for the Ebola Virus Reservoir in Kikwit, Democratic Republic of the Congo: Reflections on a Vertebrate Collection. *The Journal of Infectious Diseases, 179*(Supplement_1), S155-S163. https://doi.org/10.1086/514299

Lennemann, N. J., Rhein, B. A., Ndungo, E., Chandran, K., Qiu, X., & Maury, W. (2014, 2014-02-28). Comprehensive Functional Analysis of N-Linked Glycans on Ebola Virus GP1. *mBio, 5*(1), e00862-00813-e00862. https://doi.org/10.1128/mbio.00862-13

Leroy, E. M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., Délicat, A., Paweska, J. T., Gonzalez, J. P., & Swanepoel, R. (2005, Dec 1). Fruit bats as reservoirs of Ebola virus. *Nature, 438*(7068), 575-576. https://doi.org/10.1038/438575a

Leung, D. W., Borek, D., Luthra, P., Binning, J. M., Anantpadma, M., Liu, G., Harvey, I. B., Su, Z., Endlich-Frazier, A., Pan, J., Shabman, R. S., Chiu, W., Davey, R. A., Otwinowski, Z., Basler, C. F., & Amarasinghe, G. K. (2015). An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions. *Cell Reports, 11*(3), 376-389. https://doi.org/10.1016/j.celrep.2015.03.034

Leung, D. W., Ginder, N. D., Fulton, D. B., Nix, J., Basler, C. F., Honzatko, R. B., & Amarasinghe, G. K. (2009, 2009-01-13). Structure of the Ebola VP35 interferon inhibitory domain. *Proceedings of the National Academy of Sciences, 106*(2), 411-416. https://doi.org/10.1073/pnas.0807854106

Leung, D. W., Prins, K. C., Basler, C. F., & Amarasinghe, G. K. (2010, 2010-11-01). EbolavirusVP35 is a multifunctional virulence factor. *Virulence, 1*(6), 526-531. https://doi.org/10.4161/viru.1.6.12984

Lever, M. S., Piercy, T. J., Steward, J. A., Eastaugh, L., Smither, S. J., Taylor, C., Salguero, F. J., & Phillpotts, R. J. (2012, Jan). Lethality and pathogenesis of airborne infection with filoviruses in A129 α/β -/- interferon receptor-deficient mice. *J Med Microbiol, 61*(Pt 1), 8-15. https://doi.org/10.1099/jmm.0.036210-0

Liang, B., Li, Z., Jenni, S., Rahmeh, A. A., Morin, B. M., Grant, T., Grigorieff, N., Harrison, S. C., & Whelan, S. P. J. (2015, 2015-07-01). Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell, 162*(2), 314-327. https://doi.org/10.1016/j.cell.2015.06.018

Lin, A. E., Diehl, W. E., Cai, Y., Finch, C. L., Akusobi, C., Kirchdoerfer, R. N., Bollinger, L., Schaffner, S. F., Brown, E. A., Saphire, E. O., Andersen, K. G., Kuhn, J. H., Luban, J., & Sabeti, P. C. (2020). Reporter Assays for Ebola Virus Nucleoprotein Oligomerization, Virion-Like Particle Budding, and Minigenome Activity Reveal the Importance of Nucleoprotein Amino Acid Position 111. *Viruses, 12*(1), 105. https://doi.org/10.3390/v12010105

Liu, T., Zhang, L., Joo, D., & Sun, S.-C. (2017, 2017-12-01). NF-κB signaling in inflammation. *Signal Transduction and Targeted Therapy, 2*(1), 17023. https://doi.org/10.1038/sigtrans.2017.23

Lobert, P.-E., Escriou, N., Ruelle, J., & Michiels, T. (1999). A coding RNA sequence acts as a replication signal in cardioviruses. *Proceedings of the National Academy of Sciences, 96*(20), 11560-11565. https://doi.org/doi:10.1073/pnas.96.20.11560

Loewe, L. (2008). Negative selection. *Nature Education, 1*(1), 59.

Macneil, A., Farnon, E. C., Morgan, O. W., Gould, P., Boehmer, T. K., Blaney, D. D., Wiersma, P., Tappero, J. W., Nichol, S. T., Ksiazek, T. G., & Rollin, P. E. (2011, 2011-11-01). Filovirus Outbreak Detection and Surveillance: Lessons From Bundibugyo. *Journal of Infectious Diseases, 204*(suppl 3), S761-S767. https://doi.org/10.1093/infdis/jir294

Macneil, A., Farnon, E. C., Wamala, J., Okware, S., Cannon, D. L., Reed, Z., Towner, J. S., Tappero, J. W., Lutwama, J., Downing, R., Nichol, S. T., Ksiazek, T. G., & Rollin, P. E. (2010, 2010-12-01). Proportion of Deaths and Clinical Features in Bundibugyo Ebola Virus Infection, Uganda. *Emerging Infectious Diseases, 16*(12), 1969-1972. https://doi.org/10.3201/eid1612.100627

Madara, J. J., Han, Z., Ruthel, G., Freedman, B. D., & Harty, R. N. (2015, 2015-05-01). The multifunctional Ebola virus VP40 matrix protein is a promising therapeutic target. *Future Virology, 10*(5), 537-546. https://doi.org/10.2217/fvl.15.6

Magnani, D. M., Rogers, T. F., Maness, N. J., Grubaugh, N. D., Beutler, N., Bailey, V. K., Gonzalez-Nieto, L., Gutman, M. J., Pedreño-Lopez, N., Kwal, J. M., Ricciardi, M. J., Myers, T. A., Julander, J. G., Bohm, R. P., Gilbert, M. H., Schiro, F., Aye, P. P., Blair, R. V., Martins, M. A., Falkenstein, K. P., Kaur, A., Curry, C. L., Kallas, E. G., Desrosiers, R. C., Goldschmidt-Clermont, P. J., Whitehead, S. S., Andersen, K. G., Bonaldo, M. C., Lackner, A. A., Panganiban, A. T., Burton, D. R., & Watkins, D. I. (2018, 2018/04/24). Fetal demise and failed antibody therapy during Zika virus infection of pregnant macaques. *Nature Communications, 9*(1), 1624. https://doi.org/10.1038/s41467-018-04056-4

Mahanty, S., Gupta, M., Paragas, J., Bray, M., Ahmed, R., & Rollin, P. E. (2003, Aug 1). Protection from lethal infection is determined by innate immune responses in a mouse model of Ebola virus infection. *Virology, 312*(2), 415-424. https://doi.org/10.1016/s0042-6822(03)00233-2

Malashkevich, V. N., Schneider, B. J., Mcnally, M. L., Milhollen, M. A., Pang, J. X., & Kim, P. S. (1999, 1999-03-16). Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-A resolution. *Proceedings of the National Academy of Sciences, 96*(6), 2662-2667. https://doi.org/10.1073/pnas.96.6.2662

Manicassamy, B., Wang, J., Jiang, H., & Rong, L. (2005, 2005-04-15). Comprehensive Analysis of Ebola Virus GP1 in Viral Entry. *Journal of Virology, 79*(8), 4793-4805. https://doi.org/10.1128/jvi.79.8.4793-4805.2005

Martinez, O., Johnson, J., Manicassamy, B., Rong, L., Olinger, G. G., Hensley, L. E., & Basler, C. F. (2010, 2010-02-01). Zaire Ebola virus entry into human dendritic cells is insensitive to cathepsin L inhibition. *Cellular Microbiology, 12*(2), 148-157. https://doi.org/10.1111/j.1462-5822.2009.01385.x

Marzi, A., Akhavan, A., Simmons, G., Gramberg, T., Hofmann, H., Bates, P., Lingappa, V. R., & PöHlmann, S. (2006, 2006-07-01). The Signal Peptide of the Ebolavirus Glycoprotein Influences Interaction with the Cellular Lectins DC-SIGN and DC-SIGNR. *Journal of Virology, 80*(13), 6305-6317. https://doi.org/10.1128/jvi.02545-05

Marzi, A., Wegele, A., & Pöhlmann, S. (2006, 2006/09/01/). Modulation of virion incorporation of Ebolavirus glycoprotein: Effects on attachment, cellular entry and neutralization. *Virology, 352*(2), 345-356. https://doi.org/https://doi.org/10.1016/j.virol.2006.04.038

Mateo, M., Carbonnelle, C., Martinez, M. J., Reynard, O., Page, A., Volchkova, V. A., & Volchkov, V. E. (2011, Nov). Knockdown of Ebola virus VP24 impairs viral nucleocapsid assembly and prevents virus replication. *J Infect Dis, 204 Suppl 3*, S892-896. https://doi.org/10.1093/infdis/jir311

Mateo, M., Carbonnelle, C., Reynard, O., Kolesnikova, L., Nemirov, K., Page, A., Volchkova, V. A., & Volchkov, V. E. (2011, 2011-11-01). VP24 Is a Molecular Determinant of Ebola Virus Virulence in Guinea Pigs. *The Journal of Infectious Diseases, 204*(suppl_3), S1011-S1020. https://doi.org/10.1093/infdis/jir338

Mateo, M., Reid, S. P., Leung, L. W., Basler, C. F., & Volchkov, V. E. (2010, 2010-01-15). Ebolavirus VP24 Binding to Karyopherins Is Required for Inhibition of Interferon Signaling. *Journal of Virology, 84*(2), 1169-1175. https://doi.org/10.1128/jvi.01372-09

Mayer, J., Marini, R. P., & Fox, J. G. (2015). Biology and Diseases of Ferrets. *Laboratory Animal Medicine*, 577-622. https://doi.org/10.1016/B978-0-12-409527-4.00014-6

Meechan, P. J., & Potts, J. (2020). Biosafety in microbiological and biomedical laboratories.

Mehedi, M., Falzarano, D., Seebach, J., Hu, X., Carpenter, M. S., Schnittler, H.-J., & Feldmann, H. (2011). A New Ebola Virus Nonstructural Glycoprotein Expressed through RNA Editing. *Journal of Virology, 85*(11), 5406-5414. https://doi.org/doi:10.1128/JVI.02190-10

Metsky, H. C., Matranga, C. B., Wohl, S., Schaffner, S. F., Freije, C. A., Winnicki, S. M., West, K., Qu, J., Baniecki, M. L., & Gladden-Young, A. (2017). Zika virus evolution and spread in the Americas: manuscript author.

Metzker, M. L. (2010, 2010-01-01). Sequencing technologies — the next generation. *Nature Reviews Genetics, 11*(1), 31-46. https://doi.org/10.1038/nrg2626

Miller, J., Hippen, A., Wright, S., Morris, C., & Ridge, P. (2017, 10/12). Human viruses have codon usage biases that match highly expressed proteins in the tissues they infect. *Biomedical Genetics and Genomics, 2*, 1-5. https://doi.org/10.15761/BGG.1000134

Miranda, M. G., White, M., Dayrit, M., Hayes, C., Ksiazek, T., & Burans, J. (1991, 1991-02-01). Seroepidemiological study of filovirus related to Ebola in the Philippines. *The Lancet, 337*(8738), 425-426. https://doi.org/10.1016/0140-6736(91)91199-5

Modrof, J., Becker, S., & MüHlberger, E. (2003, 2003-03-01). Ebola Virus Transcription Activator VP30 Is a Zinc-Binding Protein. *Journal of Virology, 77*(5), 3334-3338. https://doi.org/10.1128/jvi.77.5.3334-3338.2003

Modrof, J., Mu¨Hlberger, E., Klenk, H.-D., & Becker, S. (2002, 2002-09-01). Phosphorylation of VP30 Impairs Ebola Virus Transcription. *Journal of Biological Chemistry, 277*(36), 33099-33104. https://doi.org/10.1074/jbc.m203775200

Mohan, G. S., Li, W., Ye, L., Compans, R. W., & Yang, C. (2012, 2012-12-13). Antigenic Subversion: A Novel Mechanism of Host Immune Evasion by Ebola Virus. *PLoS Pathogens, 8*(12), e1003065. https://doi.org/10.1371/journal.ppat.1003065

Morgan, A. A., & Rubenstein, E. (2013). Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *PLoS One, 8*(1), e53785-e53785. https://doi.org/10.1371/journal.pone.0053785

Muñoz-Fontela, C., & McElroy, A. K. (2017). Ebola Virus Disease in Humans: Pathophysiology and Immunity. In E. Mühlberger, L. L. Hensley, & J. S. Towner (Eds.), *Marburg- and Ebolaviruses: From Ecosystems to Molecules* (pp. 141-169). Springer International Publishing. https://doi.org/10.1007/82_2017_11

Murphy, M. (2021). *Moving from column to bead-based DNA isolation*. Retrieved March 28, 2022 from https://www.cytivalifesciences.com/en/us/news-center/from-columns-to-magnetic-beads-10001

Nakayama, E., & Saijo, M. (2013, 2013-September-05). Animal models for Ebola and Marburg virus infections [Review]. *Frontiers in microbiology, 4*. https://doi.org/10.3389/fmicb.2013.00267

Negredo, A., Palacios, G., Vázquez-Morón, S., González, F., Dopazo, H., Molero, F., Juste, J., Quetglas, J., Savji, N., De La Cruz Martínez, M., Herrera, J. E., Pizarro, M., Hutchison, S. K., Echevarría, J. E., Lipkin, W. I., & Tenorio, A. (2011, 2011-10-20). Discovery of an Ebolavirus-Like Filovirus in Europe. *PLoS Pathogens, 7*(10), e1002304. https://doi.org/10.1371/journal.ppat.1002304

Nehls, J., Businger, R., Hoffmann, M., Brinkmann, C., Fehrenbacher, B., Schaller, M., Maurer, B., Schönfeld, C., Kramer, D., Hailfinger, S., Pöhlmann, S., & Schindler, M. (2019, 2019-02-01). Release of Immunomodulatory Ebola Virus Glycoprotein-Containing

Microvesicles Is Suppressed by Tetherin in a Species-Specific Manner. *Cell Reports, 26*(7), 1841-1853.e1846. https://doi.org/10.1016/j.celrep.2019.01.065

Neumann, G., Watanabe, S., & Kawaoka, Y. (2009, 2009-09-01). Characterization of Ebolavirus regulatory genomic regions. *Virus Research, 144*(1-2), 1-7. https://doi.org/10.1016/j.virusres.2009.02.005

Niemuth, N. A., Fallacara, D., Triplett, C. A., Tamrakar, S. M., Rajbhandari, A., Florence, C., Ward, L., Griffiths, A., Carrion, R., Goez-Gazi, Y., Alfson, K. J., Staples, H. M., Brasel, T., Comer, J. E., Massey, S., Smith, J., Kocsis, A., Lowry, J., Johnston, S. C., Nalca, A., Goff, A. J., Shurtleff, A. C., Pitt, M. L., Trefry, J., & Fay, M. P. (2021, 2021-07-02). Natural history of disease in cynomolgus monkeys exposed to Ebola virus Kikwit strain demonstrates the reliability of this non-human primate model for Ebola virus disease. *PLoS One, 16*(7), e0252874. https://doi.org/10.1371/journal.pone.0252874

Ning, Y.-J., Deng, F., Hu, Z., & Wang, H. (2017, 2017-02-01). The roles of ebolavirus glycoproteins in viral
pathogenesis. *Virologica Sinica, 32*(1), 3-15. https://doi.org/10.1007/s12250-016-3850-1

Noda, T., Hagiwara, K., Sagara, H., & Kawaoka, Y. (2010). Characterization of the Ebola virus nucleoprotein-RNA complex. *Journal of General Virology, 91*(6), 1478-1483. https://doi.org/10.1099/vir.0.019794-0

Noda, T., Watanabe, S., Sagara, H., & Kawaoka, Y. (2007). Mapping of the VP40-Binding Regions of the Nucleoprotein of Ebola Virus. *Journal of Virology, 81*(7), 3554-3562. https://doi.org/doi:10.1128/JVI.02183-06

Oh, D. Y., & Hurt, A. C. (2016). Using the ferret as an animal model for investigating influenza antiviral effectiveness. *Frontiers in microbiology, 7*, 80.

Okware, S. I., Omaswa, F. G., Zaramba, S., Opio, A., Lutwama, J. J., Kamugisha, J., Rwaguma, E. B., Kagwa, P., & Lamunu, M. (2002, 2002-12-01). An outbreak of Ebola in Uganda. *Tropical Medicine and International Health, 7*(12), 1068-1075. https://doi.org/10.1046/j.1365-3156.2002.00944.x

Olson, N. D., Lund, S. P., Colman, R. E., Foster, J. T., Sahl, J. W., Schupp, J. M., Keim, P., Morrow, J. B., Salit, M. L., & Zook, J. M. (2015, 2015-July-07). Best practices for evaluating single nucleotide variant calling methods for microbial genomics [Review]. *Frontiers in Genetics, 6*. https://doi.org/10.3389/fgene.2015.00235

Pallesen, J., Murin, C. D., De Val, N., Cottrell, C. A., Hastie, K. M., Turner, H. L., Fusco, M. L., Flyak, A. I., Zeitlin, L., Crowe, J. E., Andersen, K. G., Saphire, E. O., & Ward, A. B. (2016, 2016-09-01). Structures of Ebola virus GP and sGP in complex with therapeutic antibodies. *Nature Microbiology, 1*(9), 16128. https://doi.org/10.1038/nmicrobiol.2016.128

Pappalardo, M., Collu, F., Macpherson, J., Michaelis, M., Fraternali, F., & Wass, M. N. (2017, 2017-08-01). Investigating Ebola virus pathogenicity using molecular dynamics. *BMC Genomics, 18*(S5). https://doi.org/10.1186/s12864-017-3912-2

Payne, J. L., & Wagner, A. (2019, 2019/01/01). The causes of evolvability and their evolution. *Nature Reviews Genetics, 20*(1), 24-38. https://doi.org/10.1038/s41576-018-0069-z

Pillet, S., Svitek, N., & Von Messling, V. (2009). Ferrets as a Model for Morbillivirus Pathogenesis, Complications, and Vaccines. In *Current Topics in Microbiology and Immunology* (pp. 73-87). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70617-5_4

Posada-Cespedes, S., Seifert, D., & Beerenwinkel, N. (2017, 2017-07-01). Recent advances in inferring viral diversity from high-throughput sequencing data. *Virus Research, 239*, 17-32. https://doi.org/10.1016/j.virusres.2016.09.016

Prins, K. C., Binning, J. M., Shabman, R. S., Leung, D. W., Amarasinghe, G. K., & Basler, C. F. (2010, 2010-10-15). Basic Residues within the Ebolavirus VP35 Protein Are Required for Its Viral Polymerase Cofactor Function. *Journal of Virology, 84*(20), 10581-10591. https://doi.org/10.1128/jvi.00925-10

Pybus, O. G., Rambaut, A., Belshaw, R., Freckleton, R. P., Drummond, A. J., & Holmes, E. C. (2007, 2007-03-01). Phylogenetic Evidence for Deleterious Mutation Load in RNA Viruses and Its Contribution to Viral Evolution. *Molecular Biology and Evolution, 24*(3), 845-852. https://doi.org/10.1093/molbev/msm001

Pyle, N. J. (1940, 1940-07-01). Use of Ferrets in Laboratory Work and Research Investigations. *American Journal of Public Health and the Nations Health, 30*(7), 787-796. https://doi.org/10.2105/ajph.30.7.787

Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T. F., Beutler, N. A., Burton, D. R., Lewis-Ximenez, L. L., De Jesus, J. G., Giovanetti, M., Hill, S. C., Black, A., Bedford, T., Carroll, M. W., Nunes, M., Alcantara, L. C., Sabino, E. C., Baylis, S. A., Faria, N. R., Loose, M., Simpson, J. T., Pybus, O. G., Andersen, K. G., & Loman, N. J. (2017, 2017-06-01). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols, 12*(6), 1261-1276. https://doi.org/10.1038/nprot.2017.066

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H. J., Becker-Ziaja, B., Boettcher, J. P., Cabeza-Cabrerizo, M., Camino-Sánchez, Á., Carter, L. L., Doerrbecker, J., Enkirch, T., Dorival, I. G.-., Hetzelt, N., Hinzmann, J., Holm, T., Kafetzopoulou, L. E., Koropogui, M., Kosgey, A., Kuisma, E., Logue, C. H., Mazzarelli, A., Meisel, S., Mertens, M., Michel, J., Ngabo, D., Nitzsche, K., Pallasch, E., Patrono, L. V., Portmann, J., Repits, J. G., Rickett, N. Y., Sachse, A., Singethan, K.,

Vitoriano, I., Yemanaberhan, R. L., Zekeng, E. G., Racine, T., Bello, A., Sall, A. A., Faye, O., Faye, O., Magassouba, N. F., Williams, C. V., Amburgey, V., Winona, L., Davis, E., Gerlach, J., Washington, F., Monteil, V., Jourdain, M., Bererd, M., Camara, A., Somlare, H., Camara, A., Gerard, M., Bado, G., Baillet, B., Delaune, D., Nebie, K. Y., Diarra, A., Savane, Y., Pallawo, R. B., Gutierrez, G. J., Milhano, N., Roger, I., Williams, C. J., Yattara, F., Lewandowski, K., Taylor, J., Rachwal, P., J. Turner, D., Pollakis, G., Hiscox, J. A., Matthews, D. A., Shea, M. K. O., Johnston, A. M., Wilson, D., Hutley, E., Smit, E., Di Caro, A., Wölfel, R., Stoecker, K., Fleischmann, E., Gabriel, M., Weller, S. A., Koivogui, L., Diallo, B., Keïta, S., Rambaut, A., Formenty, P., Günther, S., & Carroll, M. W. (2016, 2016-02-01). Real-time, portable genome sequencing for Ebola surveillance. *Nature, 530*(7589), 228-232. https://doi.org/10.1038/nature16996

Quinn, K., Brindley, M. A., Weller, M. L., Kaludov, N., Kondratowicz, A., Hunt, C. L., Sinn, P. L., McCray, P. B., Stein, C. S., Davidson, B. L., Flick, R., Mandell, R., Staplin, W., Maury, W., & Chiorini, J. A. (2009). Rho GTPases Modulate Entry of Ebola Virus and Vesicular Stomatitis Virus Pseudotyped Vectors. *Journal of Virology, 83*(19), 10176-10186. https://doi.org/doi:10.1128/JVI.00422-09

Radoshitzky, S. R., Warfield, K. L., Chi, X., Dong, L., Kota, K., Bradfute, S. B., Gearhart, J. D., Retterer, C., Kranzusch, P. J., Misasi, J. N., Hogenbirk, M. A., Wahl-Jensen, V., Volchkov, V. E., Cunningham, J. M., Jahrling, P. B., Aman, M. J., Bavari, S., Farzan, M., & Kuhn, J. H. (2011, 2011-09-01). Ebolavirus -Peptide Immunoadhesins Inhibit Marburgvirus and Ebolavirus Cell Entry. *Journal of Virology, 85*(17), 8502-8513. https://doi.org/10.1128/jvi.02600-10

Radzimanowski, J., Effantin, G., & Weissenhorn, W. (2014). Conformational plasticity of the Ebola virus matrix protein. *Protein science : a publication of the Protein Society, 23*(11), 1519-1527. https://doi.org/10.1002/pro.2541

Ramaswamy, V. K., Di Palma, F., Vargiu, A. V., Corona, A., Piano, D., Ruggerone, P., Zinzula, L., & Tramontano, E. (2018, 2018/03/02/). Insights into the homo-oligomerization properties of N-terminal coiled-coil domain of Ebola virus VP35 protein. *Virus Research, 247*, 61-70. https://doi.org/https://doi.org/10.1016/j.virusres.2018.02.003

Rasmussen, A. L., Okumura, A., Ferris, M. T., Green, R., Feldmann, F., Kelly, S. M., Scott, D. P., Safronetz, D., Haddock, E., LaCasse, R., Thomas, M. J., Sova, P., Carter, V. S., Weiss, J. M., Miller, D. R., Shaw, G. D., Korth, M. J., Heise, M. T., Baric, R. S., de Villena, F. P., Feldmann, H., & Katze, M. G. (2014, Nov 21). Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. *Science, 346*(6212), 987-991. https://doi.org/10.1126/science.1259595

Raymond, J., Bradfute, S., & Bray, M. (2011, 2011-11-01). Filovirus Infection of STAT-1 Knockout Mice. *The Journal of Infectious Diseases, 204*(suppl_3), S986-S990. https://doi.org/10.1093/infdis/jir335

Reddy Chichili, V. P., Kumar, V., & Sivaraman, J. (2013). Linkers in the structural biology of protein-protein interactions. *Protein science : a publication of the Protein Society, 22*(2), 153-167. https://doi.org/10.1002/pro.2206

Reid, S. P., Cárdenas, W. B., & Basler, C. F. (2005). Homo-oligomerization facilitates the interferon-antagonist activity of the ebolavirus VP35 protein. *Virology, 341*(2), 179-189. https://doi.org/10.1016/j.virol.2005.06.044

Reid, S. P., Leung, L. W., Hartman, A. L., Martinez, O., Shaw, M. L., Carbonnelle, C., Volchkov, V. E., Nichol, S. T., & Basler, C. F. (2006, 2006-06-01). Ebola Virus VP24 Binds Karyopherin α1 and Blocks STAT1 Nuclear Accumulation. *Journal of Virology, 80*(11), 5156-5167. https://doi.org/10.1128/jvi.02349-05

Reid, S. P., Leung, L. W., Hartman, A. L., Martinez, O., Shaw, M. L., Carbonnelle, C., Volchkov, V. E., Nichol, S. T., & Basler, C. F. (2006, 2006/6//). Ebola Virus VP24 Binds Karyopherin α1 and Blocks STAT1 Nuclear Accumulation. *Journal of Virology, 80*(11), 5156 LP-5167. https://doi.org/10.1128/JVI.02349-05

Reid, S. P., Valmas, C., Martinez, O., Sanchez, F. M., & Basler, C. F. (2007, 2007-12-15). Ebola Virus VP24 Proteins Inhibit the Interaction of NPI-1 Subfamily Karyopherin α Proteins with Activated STAT1. *Journal of Virology, 81*(24), 13469-13477. https://doi.org/10.1128/jvi.01097-07

Reynard, O., Borowiak, M., Volchkova, V. A., Delpeut, S., Mateo, M., & Volchkov, V. E. (2009). Ebolavirus Glycoprotein GP Masks both Its Own Epitopes and the Presence of Cellular Surface Proteins. *Journal of Virology, 83*(18), 9596-9601. https://doi.org/doi:10.1128/JVI.00784-09

Richman, D. D., Cleveland, P. H., McCormick, J. B., & Johnson, K. M. (1983). Antigenic analysis of strains of Ebola virus: Identification of two Ebola virus serotypes. *Journal of Infectious Diseases*. https://doi.org/10.1093/infdis/147.2.268

Robinson, N. B., Krieger, K., Khan, F. M., Huffman, W., Chang, M., Naik, A., Yongle, R., Hameed, I., Krieger, K., Girardi, L. N., & Gaudino, M. (2019, 2019/12/01/). The current state of animal models in research: A review. *International Journal of Surgery, 72*, 9-13. https://doi.org/https://doi.org/10.1016/j.ijsu.2019.10.015

Ryabchikova, E., Kolesnikova, L., Smolina, M., Tkachev, V., Pereboeva, L., Baranova, S., Grazhdantseva, A., & Rassadkin, Y. (1996, 1996-05-01). Ebola virus infection in guinea pigs: presumable role of granulomatous inflammation in pathogenesis. *Archives of Virology, 141*(5), 909-921. https://doi.org/10.1007/bf01718165

Sadler, A. J., & Williams, B. R. G. Structure and Function of the Protein Kinase R. In *Interferon: The 50th Anniversary* (pp. 253-292). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71329-6_13

Saeed, M. F., Kolokoltsov, A. A., Albrecht, T., & Davey, R. A. (2010, 2010-09-16). Cellular Entry of Ebola Virus Involves Uptake by a Macropinocytosis-Like Mechanism and Subsequent Trafficking through Early and Late Endosomes. *PLoS Pathogens, 6*(9), e1001110. https://doi.org/10.1371/journal.ppat.1001110

Sanchez, A., Trappier, S. G., Mahy, B. W., Peters, C. J., & Nichol, S. T. (1996, 1996-04-16). The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proceedings of the National Academy of Sciences, 93*(8), 3602-3607. https://doi.org/10.1073/pnas.93.8.3602

Sanchez, A., Yang, Z.-Y., Xu, L., Nabel, G. J., Crews, T., & Peters, C. J. (1998). Biochemical Analysis of the Secreted and Virion Glycoproteins of Ebola Virus. *Journal of Virology, 72*(8), 6442-6447. https://doi.org/doi:10.1128/JVI.72.8.6442-6447.1998

Schiffman, Z., Liu, G., Cao, W., Zhu, W., Emeterio, K., Qiu, X., & Banadyga, L. (2021, 2021-05-05). The Ferret as a Model for Filovirus Pathogenesis and Countermeasure Evaluation. *ILAR Journal*. https://doi.org/10.1093/ilar/ilab011

Schlereth, J., Grünweller, A., Biedenkopf, N., Becker, S., & Hartmann, R. K. (2016, 2016-09-01). RNA binding specificity of Ebola virus transcription factor VP30. *RNA Biology, 13*(9), 783-798. https://doi.org/10.1080/15476286.2016.1194160

Schornberg, K., Matsuyama, S., Kabsch, K., Delos, S., Bouton, A., & White, J. (2006, 2006-04-15). Role of Endosomal Cathepsins in Entry Mediated by the Ebola Virus Glycoprotein. *Journal of Virology, 80*(8), 4174-4178. https://doi.org/10.1128/jvi.80.8.4174-4178.2006

Schudt, G., Dolnik, O., Kolesnikova, L., Biedenkopf, N., Herwig, A., & Becker, S. (2015, 2015-10-01). Transport of Ebolavirus Nucleocapsids Is Dependent on Actin Polymerization: Live-Cell Imaging Analysis of Ebolavirus-Infected Cells. *Journal of Infectious Diseases, 212*(suppl 2), S160-S166. https://doi.org/10.1093/infdis/jiv083

Schümann, M., Gantke, T., & Mühlberger, E. (2009). Ebola Virus VP35 Antagonizes PKR Activity through Its C-Terminal Interferon Inhibitory Domain. *Journal of Virology, 83*(17), 8993-8997. https://doi.org/doi:10.1128/JVI.00523-09

Shabman, R. S., Gulcicek, E. E., Stone, K. L., & Basler, C. F. (2011, 2011-11-01). The Ebola Virus VP24 Protein Prevents hnRNP C1/C2 Binding to Karyopherin α1 and Partially Alters its Nuclear Import. *The Journal of Infectious Diseases, 204*(suppl_3), S904-S910. https://doi.org/10.1093/infdis/jir323

Shi, J., Wen, Z., Zhong, G., Yang, H., Wang, C., Huang, B., Liu, R., He, X., Shuai, L., Sun, Z., Zhao, Y., Liu, P., Liang, L., Cui, P., Wang, J., Zhang, X., Guan, Y., Tan, W., Wu, G., Chen, H., & Bu, Z. (2020). Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2. *Science, 368*(6494), 1016-1020. https://doi.org/doi:10.1126/science.abb7015

Shi, M., Lin, X.-D., Chen, X., Tian, J.-H., Chen, L.-J., Li, K., Wang, W., Eden, J.-S., Shen, J.-J., Liu, L., Holmes, E. C., & Zhang, Y.-Z. (2018, 2018-04-01). The evolutionary history of vertebrate RNA viruses. *Nature, 556*(7700), 197-202. https://doi.org/10.1038/s41586-018-0012-7

Shimizu, H., Shimizu, C., & Burns, J. C. (1994, 1994-06-01). Detection of novel RNA viruses: morbilliviruses as a model system. *Molecular and Cellular Probes, 8*(3), 209-214. https://doi.org/10.1006/mcpr.1994.1029

Siegert, R., Shu, H. L., Slenczka, H. L., Peters, D., & Müller, G. (1968, Jan). The aetiology of an unknown human infection transmitted by monkeys (preliminary communication). *Ger Med Mon, 13*(1), 1-2.

Simpson, D. I., Zlotnik, I., & Rutter, D. A. (1968). Vervet monkey disease. Experiment infection of guinea pigs and monkeys with the causative agent. *British journal of experimental pathology, 49*(5), 458-464. https://pubmed.ncbi.nlm.nih.gov/5727750
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2093891/

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014, 2014/02/01). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics, 15*(2), 121-132. https://doi.org/10.1038/nrg3642

Siragam, V., Wong, G., & Qiu, X. G. (2018, Jan 18). Animal models for filovirus infections. *Zool Res, 39*(1), 15-24. https://doi.org/10.24272/j.issn.2095-8137.2017.053

Slenczka, W., & Klenk, H. D. (2007). Forty Years of Marburg Virus. *The Journal of Infectious Diseases, 196*(Supplement_2), S131-S135. https://doi.org/10.1086/520551

Smith, W., Andrewes, C. H., & Laidlaw, P. P. (1933, 1933-07-01). A VIRUS OBTAINED FROM INFLUENZA PATIENTS. *The Lancet, 222*(5732), 66-68. https://doi.org/10.1016/s0140-6736(00)78541-2

Stahelin, R. V. (2014). Membrane binding and bending in Ebola VP40 assembly and egress. *Frontiers in microbiology, 5*, 300-300. https://doi.org/10.3389/fmicb.2014.00300

Stubbs, S. C. B., Blacklaws, B. A., Yohan, B., Yudhaputri, F. A., Hayati, R. F., Schwem, B., Salvaña, E. M., Destura, R. V., Lester, J. S., Myint, K. S., Sasmono, R. T., & Frost, S. D. W. (2020, 2020/02/13). Assessment of a multiplex PCR and Nanopore-based method for dengue virus sequencing in Indonesia. *Virology Journal, 17*(1), 24. https://doi.org/10.1186/s12985-020-1294-6

Su, Z., Wu, C., Shi, L., Luthra, P., Pintilie, G. D., Johnson, B., Porter, J. R., Ge, P., Chen, M., Liu, G., Frederick, T. E., Binning, J. M., Bowman, G. R., Zhou, Z. H., Basler, C. F., Gross, M. L., Leung, D. W., Chiu, W., & Amarasinghe, G. K. (2018, 2018-02-01). Electron Cryo-microscopy Structure of Ebola Virus Nucleoprotein Reveals a Mechanism

for Nucleocapsid-like Assembly. *Cell, 172*(5), 966-978.e912. https://doi.org/10.1016/j.cell.2018.02.009

Subbotina, E., Dadaeva, A., Kachko, A., & Chepurnov, A. (2010, Oct). Genetic factors of Ebola virus virulence in guinea pigs. *Virus Res, 153*(1), 121-133. https://doi.org/10.1016/j.virusres.2010.07.015

Sugita, Y., Matsunami, H., Kawaoka, Y., Noda, T., & Wolf, M. (2018, 2018-11-01). Cryo-EM structure of the Ebola virus nucleoprotein–RNA complex at 3.6 Å resolution. *Nature, 563*(7729), 137-140. https://doi.org/10.1038/s41586-018-0630-0

Tchesnokov, E. P., Raeisimakiani, P., Ngure, M., Marchant, D., & Götte, M. (2018). Recombinant RNA-Dependent RNA Polymerase Complex of Ebola Virus. *Scientific Reports, 8*(1), 3970-3970. https://doi.org/10.1038/s41598-018-22328-3

Tian, L., Shen, X., Murphy, R. W., & Shen, Y. (2018). The adaptation of codon usage of +ssRNA viruses to their hosts. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 63*, 175-179. https://doi.org/10.1016/j.meegid.2018.05.034

Timmins, J., Schoehn, G., Kohlhaas, C., Klenk, H.-D., Ruigrok, R. W. H., & Weissenhorn, W. (2003, 2003/08/01/). Oligomerization and polymerization of the filovirus matrix protein VP40. *Virology, 312*(2), 359-368. https://doi.org/https://doi.org/10.1016/S0042-6822(03)00260-5

Tin, M. M., Rheindt, F. E., Cros, E., & Mikheyev, A. S. (2015, Mar). Degenerate adaptor sequences for detecting PCR duplicates in reduced representation sequencing data improve genotype calling accuracy. *Mol Ecol Resour, 15*(2), 329-336. https://doi.org/10.1111/1755-0998.12314

Towner, J. S., Sealy, T. K., Khristova, M. L., Albariño, C. G., Conlan, S., Reeder, S. A., Quan, P.-L., Lipkin, W. I., Downing, R., Tappero, J. W., Okware, S., Lutwama, J., Bakamutumaho, B., Kayiwa, J., Comer, J. A., Rollin, P. E., Ksiazek, T. G., & Nichol, S. T. (2008, 2008-11-21). Newly Discovered Ebola Virus Associated with Hemorrhagic Fever Outbreak in Uganda. *PLoS Pathogens, 4*(11), e1000212. https://doi.org/10.1371/journal.ppat.1000212

Tripathy, S., & Jiang, R. H. Y. (2012). Massively Parallel Sequencing Technology in Pathogenic Microbes. In *Plant Fungal Pathogens* (pp. 271-294). Humana Press. https://doi.org/10.1007/978-1-61779-501-5_17

Trunschke, M., Conrad, D., Enterlein, S., Olejnik, J., Brauburger, K., & Mühlberger, E. (2013). The L-VP35 and L-L interaction domains reside in the amino terminus of the Ebola virus L protein and are potential targets for antivirals. *Virology, 441*(2), 135-145. https://doi.org/10.1016/j.virol.2013.03.013

Tucker, T., Marra, M., & Friedman, J. M. (2009, 2009-08-01). Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine. *The American Journal of Human Genetics, 85*(2), 142-154. https://doi.org/10.1016/j.ajhg.2009.06.022

Vasin, A. V., Petrova, A. V., Egorov, V. V., Plotnikova, M. A., Klotchenko, S. A., Karpenko, M. N., & Kiselev, O. I. (2016). The influenza A virus NS genome segment displays lineage-specific patterns in predicted RNA secondary structure. *BMC research notes, 9*, 279-279. https://doi.org/10.1186/s13104-016-2083-6

Vehniäinen, E.-R., Ruusunen, M., Vuorinen, P. J., Keinänen, M., Oikari, A. O. J., & Kukkonen, J. V. K. (2019, 2019-06-01). How to preserve and handle fish liver samples to conserve RNA integrity. *Environmental Science and Pollution Research, 26*(17), 17204-17213. https://doi.org/10.1007/s11356-019-05033-0

Volchkov, V. E., Becker, S., Volchkova, V. A., Ternovoj, V. A., Kotov, A. N., Netesov, S. V., & Klenk, H.-D. (1995, 1995/12/20/). GP mRNA of Ebola Virus Is Edited by the Ebola Virus Polymerase and by T7 and Vaccinia Virus Polymerases1. *Virology, 214*(2), 421-430. https://doi.org/https://doi.org/10.1006/viro.1995.0052

Volchkov, V. E., Becker, S., Volchkova, V. A., Ternovoj, V. A., Kotov, A. N., Netesov, S. V., & Klenk, H. D. (1995, Dec 20). GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases. *Virology, 214*(2), 421-430. https://doi.org/10.1006/viro.1995.0052

Volchkov, V. E., Chepurnov, A. A., Volchkova, V. A., Ternovoj, V. A., & Klenk, H. D. (2000, Nov 10). Molecular characterization of guinea pig-adapted variants of Ebola virus. *Virology, 277*(1), 147-155. https://doi.org/10.1006/viro.2000.0572

Volchkov, V. E., Feldmann, H., Volchkova, V. A., & Klenk, H.-D. (1998, 1998-05-12). Processing of the Ebola virus glycoprotein by the proprotein convertase furin. *Proceedings of the National Academy of Sciences, 95*(10), 5762-5767. https://doi.org/10.1073/pnas.95.10.5762

Volchkova, V. A., Dolnik, O., Martinez, M. J., Reynard, O., & Volchkov, V. E. (2011, Nov). Genomic RNA editing and its impact on Ebola virus adaptation during serial passages in cell culture and infection of guinea pigs. *J Infect Dis, 204 Suppl 3*, S941-946. https://doi.org/10.1093/infdis/jir321

Volchkova, V. A., Feldmann, H., Klenk, H. D., & Volchkov, V. E. (1998, Oct 25). The nonstructural small glycoprotein sGP of Ebola virus is secreted as an antiparallel-orientated homodimer. *Virology, 250*(2), 408-414. https://doi.org/10.1006/viro.1998.9389

Volchkova, V. A., Klenk, H.-D., & Volchkov, V. E. (1999, 1999/12/05/). Delta-Peptide Is the Carboxy-Terminal Cleavage Fragment of the Nonstructural Small Glycoprotein sGP of Ebola Virus. *Virology, 265*(1), 164-171. https://doi.org/https://doi.org/10.1006/viro.1999.0034

Wahl-Jensen, V., Bollinger, L., Safronetz, D., De Kok-Mercado, F., Scott, D., & Ebihara, H. (2012, 2012-12-14). Use of the Syrian Hamster as a New Model of Ebola Virus Disease and Other Viral Hemorrhagic Fevers. *Viruses, 4*(12), 3754-3784. https://doi.org/10.3390/v4123754

Wahl-Jensen, V. M., Afanasieva, T. A., Seebach, J., Ströher, U., Feldmann, H., & Schnittler, H.-J. (2005). Effects of Ebola Virus Glycoproteins on Endothelial Cell Activation and Barrier Function. *Journal of Virology, 79*(16), 10442-10450. https://doi.org/doi:10.1128/JVI.79.16.10442-10450.2005

Wan, W., Kolesnikova, L., Clarke, M., Koehler, A., Noda, T., Becker, S., & Briggs, J. A. G. (2017, 2017-11-01). Structure and assembly of the Ebola virus nucleocapsid. *Nature, 551*(7680), 394-397. https://doi.org/10.1038/nature24490

Wang, J., Manicassamy, B., Caffrey, M., & Rong, L. (2011, 2011-06-01). Characterization of the receptor-binding domain of Ebola glycoprotein in viral entry. *Virologica Sinica, 26*(3), 156-170. https://doi.org/10.1007/s12250-011-3194-9

Watanabe, S., Noda, T., Halfmann, P., Jasenosky, L., & Kawaoka, Y. (2007). Ebola Virus (EBOV) VP24 Inhibits Transcription and Replication of the EBOV Genome. *The Journal of Infectious Diseases, 196*(Supplement_2), S284-S290. https://doi.org/10.1086/520582

Watanabe, S., Noda, T., & Kawaoka, Y. (2006, 2006-04-15). Functional Mapping of the Nucleoprotein of Ebola Virus. *Journal of Virology, 80*(8), 3743-3751. https://doi.org/10.1128/jvi.80.8.3743-3751.2006

Watt, A., Moukambi, F., Banadyga, L., Groseth, A., Callison, J., Herwig, A., Ebihara, H., Feldmann, H., & Hoenen, T. (2014, 2014-09-15). A Novel Life Cycle Modeling System for Ebola Virus Shows a Genome Length-Dependent Role of VP24 in Virus Infectivity. *Journal of Virology, 88*(18), 10511-10524. https://doi.org/10.1128/jvi.01272-14

Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr., Swanstrom, R., Burch, C. L., & Weeks, K. M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature, 460*(7256), 711-716. https://doi.org/10.1038/nature08237

Weik, M., Modrof, J., Klenk, H.-D., Becker, S., & Mühlberger, E. (2002). Ebola virus VP30-mediated transcription is regulated by RNA secondary structure formation. *Journal of Virology, 76*(17), 8532-8539. https://doi.org/10.1128/jvi.76.17.8532-8539.2002

Whelan, S. P., Barr, J. N., & Wertz, G. W. (2004). Transcription and replication of nonsegmented negative-strand RNA viruses. *Curr Top Microbiol Immunol, 283*, 61-119. https://doi.org/10.1007/978-3-662-06099-5_3

WHO. (1978a). Ebola haemorrhagic fever in Sudan, 1976. *Bulletin of the World Health Organization, 56*(2), 247-270. https://pubmed.ncbi.nlm.nih.gov/307455

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2395561/

WHO. (1978b). Ebola haemorrhagic fever in Zaire, 1976. *Bulletin of the World Health Organization, 56*(2), 271-293. https://pubmed.ncbi.nlm.nih.gov/307456
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2395567/

Wong, G., He, S., Wei, H., Kroeker, A., Audet, J., Leung, A., Cutts, T., Graham, J., Kobasa, D., Embury-Hyatt, C., Kobinger, G. P., & Qiu, X. (2016). Development and Characterization of a Guinea Pig-Adapted Sudan Virus. *Journal of Virology, 90*(1), 392-399. https://doi.org/10.1128/JVI.02331-15

Wyers, M., Formenty, P., Cherel, Y., Guigand, L., Fernandez, B., Boesch, C., & Le Guenno, B. (1999, Feb). Histopathological and immunohistochemical studies of lesions associated with Ebola virus in a naturally infected chimpanzee. *J Infect Dis, 179 Suppl 1*, S54-59. https://doi.org/10.1086/514300

Xu, C., Katyal, N., Nesterova, T., & Perilla, J. R. (2020, 2020-10-21). Molecular determinants of Ebola nucleocapsid stability from molecular dynamics simulations. *The Journal of Chemical Physics, 153*(15), 155102. https://doi.org/10.1063/5.0021491

Xu, W., Edwards, R., Megan, Borek, M., Dominika, Feagins, R., Alicia, Mittal, A., Alinger, B., Joshua, Berry, N., Kayla, Yen, B., Hamilton, J., Brett, J., Tom, Pappu, V., Rohit, Leung, W., Daisy, Basler, F., Christopher, & Amarasinghe, K., Gaya. (2014, 2014-08-01). Ebola Virus VP24 Targets a Unique NLS Binding Site on Karyopherin Alpha 5 to Selectively Compete with Nuclear Import of Phosphorylated STAT1. *Cell Host & Microbe, 16*(2), 187-200. https://doi.org/10.1016/j.chom.2014.07.008

Xu, W., Luthra, P., Wu, C., Batra, J., Leung, D. W., Basler, C. F., & Amarasinghe, G. K. (2017, 2017-08-01). Ebola virus VP30 and nucleoprotein interactions modulate viral RNA synthesis. *Nature Communications, 8*(1). https://doi.org/10.1038/ncomms15576

Yamada, H., Udagawa, T., Mizuno, S., Hiramatsu, K., & Sugawara, I. (2005, 05/01). Newly Designed Primer Sets Available for Evaluating Various Cytokines and iNOS mRNA Expression in Guinea Pig Lung Tissues by RT-PCR. *Experimental animals / Japanese Association for Laboratory Animal Science, 54*, 163-172. https://doi.org/10.1538/expanim.54.163

Yamaoka, S., Banadyga, L., Bray, M., & Ebihara, H. (2017). Small Animal Models for Studying Filovirus Pathogenesis. In E. Mühlberger, L. L. Hensley, & J. S. Towner (Eds.), *Marburg- and Ebolaviruses: From Ecosystems to Molecules* (pp. 195-227). Springer International Publishing. https://doi.org/10.1007/82_2017_9

Yamaoka, S., & Ebihara, H. (2021). Pathogenicity and Virulence of Ebolaviruses with Species- and Variant-specificity. *Virulence, 12*(1), 885-901. https://doi.org/10.1080/21505594.2021.1898169

Yamayoshi, S., Noda, T., Ebihara, H., Goto, H., Morikawa, Y., Lukashevich, I. S., Neumann, G., Feldmann, H., & Kawaoka, Y. (2008, 2008-03-01). Ebola Virus Matrix Protein VP40 Uses the COPII Transport System for Its Intracellular Transport. *Cell Host & Microbe, 3*(3), 168-177. https://doi.org/10.1016/j.chom.2008.02.001

Yan, F., He, S., Banadyga, L., Zhu, W., Zhang, H., Rahim, M. N., Collignon, B., Senthilkumaran, C., Embury-Hyatt, C., & Qiu, X. (2019, May). Characterization of Reston virus infection in ferrets. *Antiviral Res, 165*, 1-10. https://doi.org/10.1016/j.antiviral.2019.03.001

Yang, X.-L., Tan, C. W., Anderson, D. E., Jiang, R.-D., Li, B., Zhang, W., Zhu, Y., Lim, X. F., Zhou, P., Liu, X.-L., Guan, W., Zhang, L., Li, S.-Y., Zhang, Y.-Z., Wang, L.-F., & Shi, Z.-L. (2019, 2019-03-01). Characterization of a filovirus (Měnglà virus) from Rousettus bats in China. *Nature Microbiology, 4*(3), 390-395. https://doi.org/10.1038/s41564-018-0328-y

Yang, X.-L., Zhang, Y.-Z., Jiang, R.-D., Guo, H., Zhang, W., Li, B., Wang, N., Wang, L., Waruhiu, C., Zhou, J.-H., Li, S.-Y., Daszak, P., Wang, L.-F., & Shi, Z.-L. (2017, 2017-03-01). Genetically Diverse Filoviruses in Rousettus and Eonycteris spp. Bats, China, 2009 and 2015. *Emerging Infectious Diseases, 23*(3), 482-486. https://doi.org/10.3201/eid2303.161119

Yeku, O., & Frohman, M. A. (2011). Rapid Amplification of cDNA Ends (RACE). In *RNA* (pp. 107-122). Humana Press. https://doi.org/10.1007/978-1-59745-248-9_8

Yu, D.-S., Weng, T.-H., Wu, X.-X., Wang, F. X. C., Lu, X.-Y., Wu, H.-B., Wu, N.-P., Li, L.-J., & Yao, H.-P. (2017, 2017-08-15). The lifecycle of the Ebola virus in host cells. *Oncotarget, 8*(33), 55750-55759. https://doi.org/10.18632/oncotarget.18498

Yuan, J., Zhang, Y., Li, J., Zhang, Y., Wang, L.-F., & Shi, Z. (2012, 2012-12-01). Serological evidence of ebolavirus infection in bats, China. *Virology Journal, 9*(1), 236. https://doi.org/10.1186/1743-422x-9-236

Yuan, S., Cao, L., Ling, H., Dang, M., Sun, Y., Zhang, X., Chen, Y., Zhang, L., Su, D., Wang, X., & Rao, Z. (2015, 2015-11-01). TIM-1 acts a dual-attachment receptor for Ebolavirus by interacting directly with viral GP and the PS on the viral envelope. *Protein & Cell, 6*(11), 814-824. https://doi.org/10.1007/s13238-015-0220-y

Zhang, A. P. P., Bornholdt, Z. A., Liu, T., Abelson, D. M., Lee, D. E., Li, S., Woods, V. L., & Saphire, E. O. (2012, 2012-02-23). The Ebola Virus Interferon Antagonist VP24 Directly Binds STAT1 and Has a Novel, Pyramidal Fold. *PLoS Pathogens, 8*(2), e1002550. https://doi.org/10.1371/journal.ppat.1002550

Zhou, W., Chen, T., Zhao, H., Eterovic, A. K., Meric-Bernstam, F., Mills, G. B., & Chen, K. (2014, 2014-04-15). Bias from removing read duplication in ultra-deep sequencing

experiments. *Bioinformatics, 30*(8), 1073-1080. https://doi.org/10.1093/bioinformatics/btt771

Zhu, W., Banadyga, L., Emeterio, K., Wong, G., & Qiu, X. (2019, 2019-10-31). The Roles of Ebola Virus Soluble Glycoprotein in Replication, Pathogenesis, and Countermeasure Development. *Viruses, 11*(11), 999. https://doi.org/10.3390/v11110999

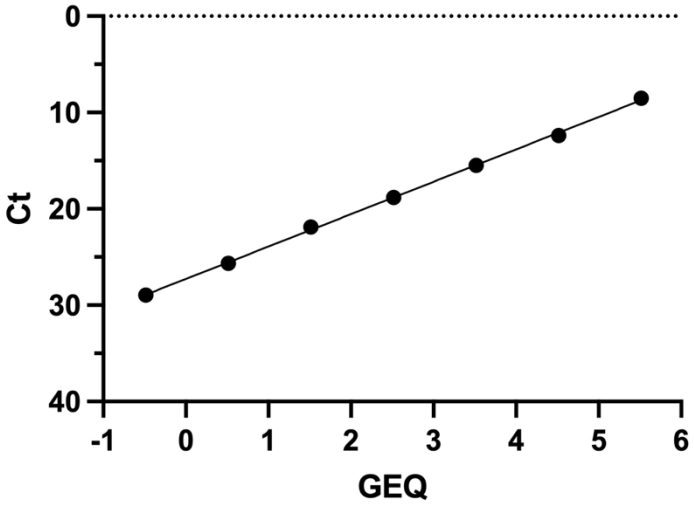
Zinzula, L., Nagy, I., Orsini, M., Weyher-Stingl, E., Bracher, A., & Baumeister, W. (2019, 2019/01/02/). Structures of Ebola and Reston Virus VP35 Oligomerization Domains and Comparative Biophysical Characterization in All Ebolavirus Species. *Structure, 27*(1), 39-54.e36. https://doi.org/https://doi.org/10.1016/j.str.2018.09.009

Zinzula, L., & Tramontano, E. (2013, 2013-12-01). Strategies of highly pathogenic RNA viruses to block dsRNA detection by RIG-I-like receptors: Hide, mask, hit. *Antiviral Research, 100*(3), 615-635. https://doi.org/10.1016/j.antiviral.2013.10.002

Zivcec, M., Safronetz, D., Haddock, E., Feldmann, H., & Ebihara, H. (2011). Validation of assays to monitor immune responses in the Syrian golden hamster (Mesocricetus auratus). *Journal of immunological methods, 368*(1-2), 24-35. https://doi.org/10.1016/j.jim.2011.02.004

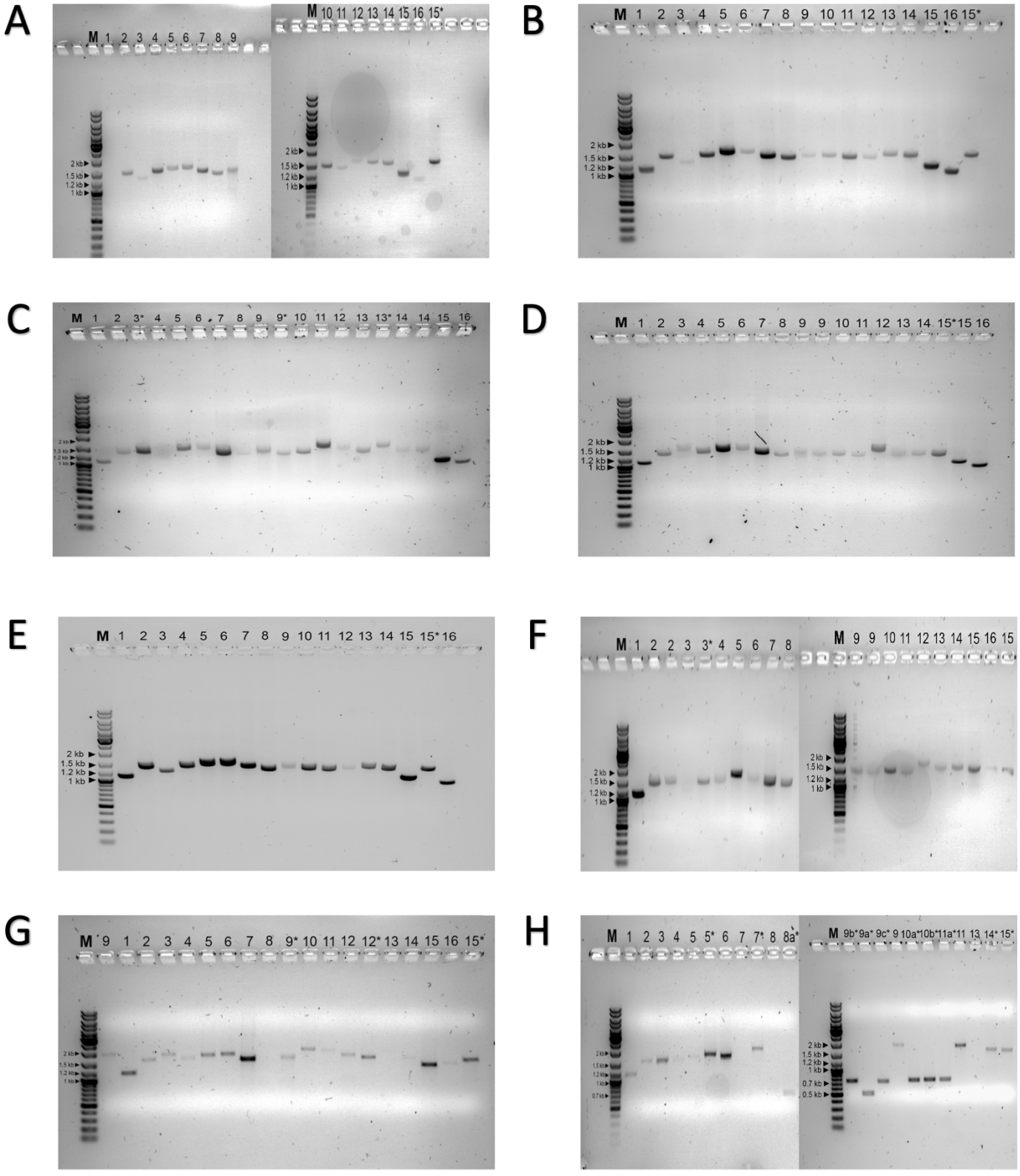
6 APPENDICES

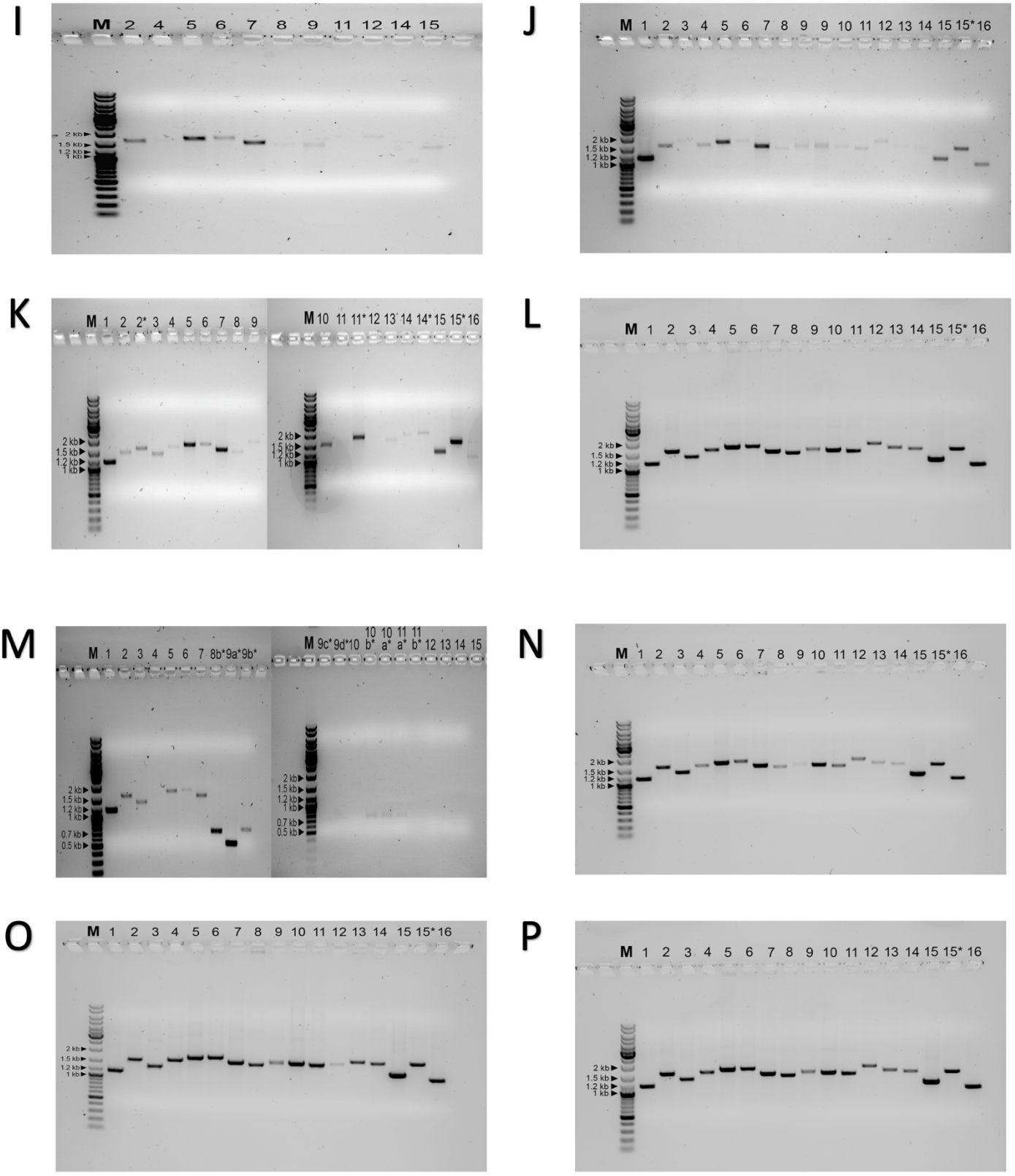
APPENDIX A

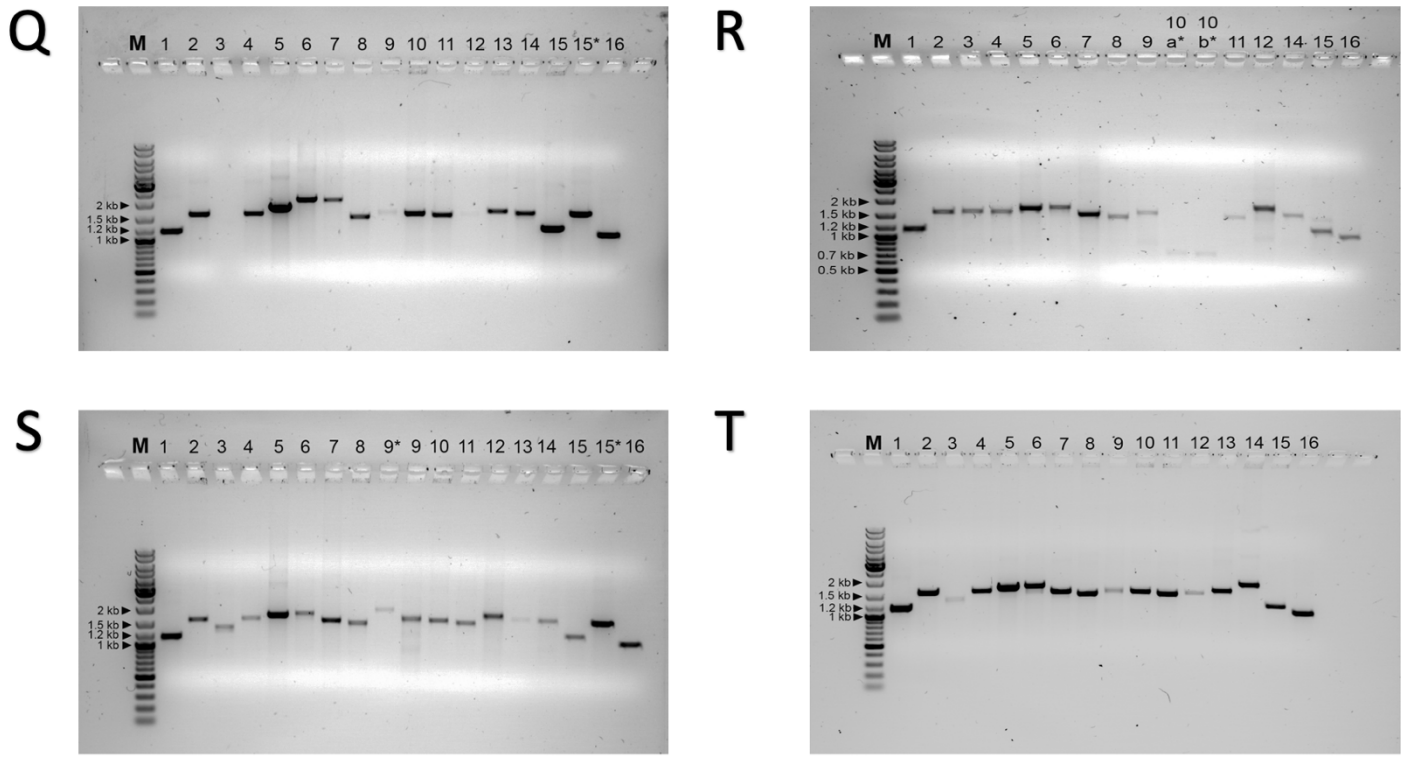


Appendix A1. Real-time quantitative reverse transcription PCR standard curve. This standard curve of genome equivalents (GEQ) versus cycle threshold (Ct) was used to estimate the amount of SUDV RNA present in extracted pooled guinea pig liver and spleen homogenate samples. GEQ was determined using the equation $y = -3.363x + 27.28$ with an R^2 value of 0.9993, where y is Ct and x is the $\log_{10}$GEQ. The standard curve was generated from a serially diluted SUDV *NP* gene cloned to a vector.

APPENDIX B


A
M 1 2 3 4 5 6 7 8 9
M 10 11 12 13 14 15 16 15*
B
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 15*
C
M 1 2 3* 4 5 6 7 8 9 9* 10 11 12 13 13* 14 14 15 16
D
M 1 2 3 4 5 6 7 8 9 9 10 11 12 13 14 15*15 16
E
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15* 16
F
M 1 2 2 3 3* 4 5 6 7 8
M 9 9 10 11 12 13 14 15 16 15
G
M 9 1 2 3 4 5 6 7 8 9* 10 11 12 12* 13 14 15 16 15*
H
M 1 2 3 4 5 5* 6 7 7* 8 8a*
M 9b*9a* 9c* 9 10a*10b*11a*11 13 14* 15*


I
M 2 4 5 6 7 8 9 11 12 14 15
2 kb
1.5 kb
1.2 kb
1 kb
J
M 1 2 3 4 5 6 7 8 9 9 10 11 12 13 14 15 15* 16
2 kb
1.5 kb
1.2 kb
1 kb
K
M 1 2 2* 3 4 5 6 7 8 9
M 10 11 11* 12 13 14 14* 15 15* 16
2 kb
1.5 kb
1.2 kb
1 kb
L
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15* 16
2 kb
1.5 kb
1.2 kb
1 kb
M
M 1 2 3 4 5 6 7 8b* 9a* 9b*
M 9c* 9d* 10 10 b* 10 a* 11 a* 11 b* 12 13 14 15
2 kb
1.5 kb
1.2 kb
1 kb
0.7 kb
0.5 kb
N
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15* 16
2 kb
1.5 kb
1.2 kb
1 kb
O
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15* 16
2 kb
1.5 kb
1.2 kb
1 kb
P
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15* 16
2 kb
1.5 kb
1.2 kb
1 kb



Appendix B1. Gel electrophoresis of amplicons collected from each GPA-SUDV passages. The cDNA generated from reverse transcription of total RNA extracted from the GPA-SUDV homogenates was used as templates for amplification of 16 amplicons that spans the whole virus genome. The success of amplification varied between passages 1 to 20 (A-T). Passage 9 (I) had the greatest number of missing fragments (7 total), while passages 8 (H), 13 (M), and 18 (R) had a single fragment missing. Amplicons that were re-amplified by primers that were designed using PrimalScheme are highlighted with an asterisk (*). The rest of the amplicons were amplified using selected primers from Wong et al. (2016). All PCR products were run in 1% agarose gel stained with SYBR Safe DNA gel stain in 1X TAE buffer for 30 minutes. Sizes of amplicons were estimated using GeneRuler 1 kb DNA ladder (Lane M).