Faculty of Sciences

On the Analysis of Virus Phenotypes

A dissertation submitted for the degree of

Doctor of Philosophy

in the university of Kent for the Faculty of Sciences

Stuart Masterson

October 2020

DECLARATION: No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent or any other institute of learning. All research is my own, except where clearly stated.

Table of Contents

Abstract

Closely related viral species have the capacity to cause drastically different clinical symptoms in humans. The 2013 – 2016 west African Ebolavirus outbreak which resulted in the death of 11,000 people was caused by the species Zaire ebolavirus, yet the near identical Reston ebolavirus is asymptomatic in humans. Similarly the SARS-CoV-2 virus instigated widespread disruption owing to its ability to spread easily between individuals, while the SARS-CoV virus caused far fewer cases and spread much less effectively, despite the two viruses being members of the same species.

Recent advances in next generation sequencing have allowed for the collection and processing of vast quantities of biological data. Using the extensive assortment of viral genomes openly available we can identify and analyse differentially conserved positions (DCPs), residues that are highly conserved amongst a species but differ between closely related species. Using advanced modelling and structural analysis we can determine which DCPs are likely to have an impact on the differing levels of pathogenicity between species.

Here we report the VP24 Ebolavirus protein is key to pathogenicity, and that few key residue differences in the VP24/human karyopherin binding site are responsible for the lack of pathogenicity of Reston virus. Additionally we use this data to propose that the newly discovered Bombali virus is also not pathogenic in humans using sequence comparison between the pathogenic and non-pathogenic species. To further this approach we also consider the requirement to adapt filoviruses to a novel species, to determine key amino acid changes that are responsible for pathogenicity.

List of Abbreviations

ACE2	Angiotensin-converting enzyme 2
AIDS	Acquired immune deficiency syndrome
BDBV	Bundibugyo virus
BLAST	Basic Local Alignment Search Tool
BLOSUM62	Blocks Substitution Matrix 62
BOMV	Bombali virus
CDS	Coding Domain Sequence
CFR	Case fatality rate
ChAd-3-EBO-Z	Chimpanzee Adenovirus 3 – Ebolavirus Zaire
CM	Covariance Models
COVID-19	Coronavirus disease 19
CPE	Cytopathogenic effects
DCP	Differentially Conserved Position
DRC	Democratic Republic of the Congo
E	Vaccine Effectiveness
EBOV	Ebola virus
EVD	Ebolavirus disease
GAVI	Global Alliance for Vaccines & Immunization
GE	Gene End
GP	Glycoprotein
GS	Gene Start
HCoV-229E	Human coronavirus 229E

HCoV-HKU1	Human coronavirus HKU1
HCoV-NL63	Human coronavirus NL63
HCoV-OC43	Human coronavirus OC43
HMM	Hidden Markov Model
HIV	Human Immunodeficiency virus
HSV1	Herpes simplex virus 1
HSV2	Herpes simplex virus 2
I_c	Herd Immunity Threshold
IC_{50}	Half maximal inhibitory concentration
KIU	Kallikrein inhibitor units
KPNA5	Karyopherin Alpha 5
L	RNA Polymerase
LSR	Locally Stable RNA
MARV	Marburg virus
MCA	Multiple Corresponding Analysis
mCSM	Mutation Cutoff Scanning Matrix
MERS-CoV	Middle East respiratory syndrome coronavirus
MFE	Minimum Free Energy
miRNA	Micro RNA
MOI	Multiplicity of Infection
MPI	Mean Pairwise Sequence Identity
MSA	Multiple Sequence Alignment
MVD	Marburgvirus disease
NCBI	National Center for Biotechnology Information
NHP	Non-human primate

NNS	Non-segmented Negative Sense
NP	Nucleoprotein
NPC1	Niemann-Pick C1 receptor
ORF	Open reading frame
PDB	Protein Databank
Phyre2	Protein Homology/Analogy Recognition Engine 2
R_0	Basic Reproductive Number
RAVV	Ravn virus
RESTV	Reston virus
RRE	Rev Response Element
rVSV-ZEBOV	Recombinant Varicella zoster virus - Zaire ebolavirus
S	Spike protein
SAGE	Strategic Advisory Group of Experts
SARS	Severe acute respiratory syndrome
SCI	Structure Conservation Index
SDP	Specificity Determining Position
SL	Stem Loop Structure
SUDV	Sudan virus
SVM	Support Vector Machine
TAFV	Tai Forest virus
TIM-1	T-cell immunoglobulin and mucin domain 1
TMPRSS2	Transmembrane serine protease 2
UTR	Untranslated Region
V_c	Critical Vaccine Coverage
vncRNA	Viral Non-coding RNA

VP24	Viral Protein 24
VP30	Viral Protein 30
VP35	Viral Protein 35
VP40	Viral Protein 40
WHO	World Health Organisation

List of Figures

Figures in Chapter 1

Figures in Chapter 2

Figures in Chapter 3

Figures in Chapter 4

Figures in Chapter 5

Figures in Chapter 6

List of Tables

Tables in Chapter 1

Tables in Chapter 3

Tables in Chapter 5

Tables in Chapter 6

Chapter 1: Introduction

1.1 Viruses – an Overview

Viruses have been a consistent and instrumental part of human existence since the dawn of mankind. Long before their discovery in 1892 (van Helvoort, 1991; Lustig and Levine, 1992; Lwoff, 1957) the relationship between these microscopic particles and humanity has shaped society, health and culture in ways it is almost impossible to comprehend. The outcome of great wars, the expansion of cities and trade routes, and the scientific pursuit of medicine and healthcare have all been decided and influenced by virus induced illnesses (Wolfe et al., 2007), and we are seeing the devastation they cause at this very moment with the ongoing coronavirus pandemic (Wu et al., 2020c; Zhou et al., 2020). Yet while the other major human pathogens – bacteria, parasites and fungi – were discovered, if not well characterised or understood, centuries ago, virology (the study of viruses), is a relatively new field in biology.

1.1.1 Early History

Records of viral infections have been traced back as far as ancient Egypt, with drawings depicting priests suffering from deformities strikingly similar to the symptoms of polio (caused by the Enterovirus C virus) (Galassi et al., 2017). In ancient Greek times Aristotle had made the connection between dog bites and the onset of Rabies (caused by *Rabies lyssavirus*) (Lwoff, 1957), and the results of the introduction of smallpox (caused by the Vaccinia virus) to natives of the North American continent by European explorers in the fifteenth and sixteenth centuries are well documented (Carlos and Lewis, 2012). Despite this, outbreaks in pre-industrial times appeared to be rare, with much of the noted outbreaks being retroactively ascribed to a small core group of "old world" viruses, mainly measles, rabies, smallpox and influenza (Lupiani and Reddy, 2009; Moore et al., 2006; Velasco-Villa et al., 2017). Given the nature of the viral lifecycle, namely the need for a host to survive and a large group of vulnerable individuals to spread to, this is not surprising, as the pre industrial human lifestyle of

living in small villages and farmsteads was not advantageous to viral transmission (Wolfe et al., 2007).

However, coinciding with the rapid onset of human population growth, urbanisation and greater interconnectivity between people across the world, a plethora of new viral pathogens began to appear, and outbreaks became larger and more prolonged (McMichael, 2004; Reperant and Osterhaus, 2017). Explorers brought back new pathogens from new worlds, and cramped cities provided a fertile ground to create a chain of transmission needed for to viruses to be maintained in a population. In the years following the industrial revolution, infections brought on by smallpox and measles are thought to be among the most common cause of death in Europe, along with bacterial infections such as the plague and tuberculosis. And while the research conducted by Edward Jenner in late the 1790s led to the creation of the first ever vaccine against smallpox (Sánchez-Sampedro et al., 2015), the root cause of such infections remained unknown, limiting the advancement of effective prophylactics or treatments.

1.1.2 The Emergence of Virology

These conditions, coupled with greater scientific understanding and breakthroughs surrounding the transmission of disease and the microbial world, led to the pioneering work by Louis Pasteur, Dimitri Ivanovsky, Martinus Beijerinck and many others at the end of the nineteenth century. The study of Tobacco mosaic disease by the latter two, and the invention of a vaccine to Rabies by Pasteur were of particular importance. In 1892 Ivanovsky reported of a 'filterable infectious agent' that could pass through the Chamberland filter commonly used to separate bacteria from a sample (Bos, 1999; van Helvoort, 1991; Lwoff, 1957; Méthot, 2016). This work was replicated in 1898 by Martinus Beijerinck, leading to the first use of the term 'virus' (Artenstein, 2012; Méthot, 2016).

Much discussion ensued however regarding the nature of these new pathogens. Prevailing theories at the time deemed viruses fluidic, and other theories considered viruses to be toxins, perhaps excreted from bacteria (Bos, 1999; Lustig and Levine, 1992; Lwoff, 1957). Though the advancements in vaccinology and preventative treatment came quickly after these initial findings, research into the physical viruses

came slowly, hindered by the technology of the time. One prime example of this is the Rabies vaccine. Pasteur developed a vaccine as far back as 1885, using a protocol that was in use for a further 50 years, however the virus that causes Rabies disease remained undefined (Plotkin, 2014, 1980; Velasco-Villa et al., 2017). In 1903 Adelchi Negri first reported microscopic lesions (Negri bodies) in the brains of rabid animals, which were incorrectly believed to be protozoan parasites. Further research into filtration demonstrated that they were far smaller than protozoa, and even smaller than bacteria (Albertini et al., 2008). It was another thirty years before Negri bodies were shown to be accumulations of virions approximately 100–150nm long, the size of rhabdovirus, the virus causing rabies.

1.1.3 Twentieth Century Discoveries

Starting in the late 1930s, advancement in fields such as electron microscopy allowed the study of the microscopic world in incredible detail (Harris, 2015), and in the last hundred years we have been able to characterise the pathogens that for millennia have caused West Nile virus, Smallpox, Rabies and countless other diseases (Méthot, 2016; Thèves et al., 2014). This era led to the discovery of viruses as primarily protein based, generally without membranes and that a large number of them contain RNA only. The first viral structure to be determined using electron microscopy was the Tobacco mosaic virus (van Helvoort, 1991; Loring, 1939).

The second half of the twentieth century was an even more eventful era for virologists, with a huge number of new viruses being discovered, and the pathogens behind a wide range of diseases being unearthed, including Varicella zoster virus that causes chickenpox and rhinovirus that is responsible for the common cold (McCrary et al., 1999; Tyrrell, 1987). This time was also defined by the emergence of new infectious diseases caused by viruses not previously known to infect humans, including Human Immunodeficiency virus (HIV), Marburg virus and Ebola virus, the former responsible for the widespread AIDS pandemic and the latter two responsible for a severe and highly fatal haemorrhagic fever (Pourrut et al., 2005; Rathore et al., 2017). The AIDS crisis in particular has caused severe medical, cultural and economic changes across the world, with a suspected 34 million individuals currently infected with HIV (O'Cofaigh and Lewthwaite, 2013). However, this period is also known for great

strides in the tackling of several viral diseases. In 1980 smallpox was declared eradicated (Moore et al., 2006; Thèves et al., 2014), and cases of rabies, measles and other "old world" infections have continued to decrease (Plotkin, 2014; Velasco-Villa et al., 2017).

Advancements in bioinformatics and genomics furthered understanding even more. Full length genomes can be sequenced, structures clearly solved using more advanced techniques such as Cryo-EM and X-ray crystallography, and protein properties, viral lifecycles and mechanisms of pathogenesis understood. Concurrent to this scientific progress, new viruses have emerged in the early 21st century, notably SARS coronavirus and Nipah henipavirus (Al-Hazmi, 2016; Ksiazek et al., 2011; Luby, 2013; Reperant and Osterhaus, 2017), and outbreaks caused by Ebola virus are becoming more widespread and prolonged. The emergence of viruses into human ecosystems shows no sign of slowing. Treatments for many new viral diseases are mostly limited and ineffective, requiring the need for more data, and extensive research on prevention strategies, vaccine development and antiviral treatments.

1.1.4 What is a Virus?

At current there is ongoing debate as to whether a virus constitutes a 'living' organism, with various arguments for and against the argument that they are 'alive'. The generally held belief is that they are not considered living, and this thesis will continue to consider them as such, irrespective of the ongoing debate (Forterre, 2016; Koonin and Starokadomskyy, 2016; Morgan, 2016; Pradeu et al., 2016). They possess no metabolism, no cellular structure or organelles, and many do not possess DNA but rather RNA only. While many outside the scientific community tend to group bacteria and viruses together when considering disease, viruses are more akin to parasites in the way they interact with other species, and are often referred to as obligate intracellular parasites. They cannot exist outside of other living organisms, and while some species have been shown to remain dormant in inorganic environments, viruses can only replicate inside a host cell (Wimmer et al., 2009). They are dependent on host cellular proteins to aid in the unwinding, replication and repackaging of their genetic material (Albertini et al., 2008; Olejnik et al., 2011), and viruses that contain a lipid coating or membrane obtain this from host cells during budding, as they possess no

way of producing their own. Furthermore, they require the movement of other organisms to carry them from place to place, allowing them to spread and create a chain of transmission.

1.1.5 Morphology

Perhaps more diverse than any other type of life currently known, viruses exist in an immense range of shapes, sizes and structures. As a rule, a single viral particle (known as a virion) is on average between 20 and 300nm in diameter, however many defy this convention. Filoviruses for example have an average length of 800-1000nm, and a diameter of 80-120nm (Falasca et al., 2015; Grifoni et al., 2016). This generally places viruses as much smaller than bacteria and as such are unable to be viewed using traditional microscopic methods, explaining their lack of characterisation until the invention of electron microscopy in the 1930s (Harris, 2015). However, several "giant" viruses such as the Mimivirus have been described in recent years, many exceeding 1μm in diameter (Fischer, 2016; Sharma et al., 2016).

Virions are composed of the viral nucleic acid surrounded by protective, structural and functional proteins. This is referred to as the viral capsid, while the proteins directly responsible for encapsulating and protecting the genetic material are known as nucleoproteins. The combination of nucleic acid and these nucleoproteins is referred to as the nucleocapsid (Albertini et al., 2008; Banadyga et al., 2017a; Wan et al., 2017). Many viruses also have lipid envelopes surrounding them, such as those in the families Rhabdoviridae, Herpesviridae, Poxviridae and Filoviridae (Albertini et al., 2008; Moore et al., 2006; Olejnik et al., 2011), though others do not, prime examples being Adenoviridae and Papillomaviridae. Some complex viruses code for proteins that aid in the assembly of their capsid, however many rely solely on hijacked host proteins to complete this function, while others make use of a combination of both (Albertini et al., 2008; Han et al., 2016; Luthra et al., 2015). While virions can exist in all shapes and sizes, a common theme of four core structures can be observed – helical, icosahedral, envelope and complex.

1.1.6 Viral Genomes

As with morphology, there exists an incredibly broad and diverse array of viral genomes. There is more genomic diversity among viruses than that of bacteria, archaea and animals, and while around 10,000 different species have been described in depth as of 2018, there is belief that millions of species exist, eclipsing all other known forms of life combined (Breitbart and Rohwer, 2005).

Viruses can either have a DNA or RNA genome, with the majority of known species containing the latter. These genomes can either be single stranded, consisting of unpaired nucleic acids, or double stranded, containing paired strands. Both DNA and RNA genomes can be either single or double stranded. Some viral genomes can be a combination of double and single stranded regions, though these are rare.

For those viruses with single stranded RNA genomes, the strands are referred to as either positive sense (plus strand) or negative sense (minus strand). This nomenclature is based on whether or not they are complementary to the viral messenger RNA. Positive sense strands are identical to the viral mRNA and therefore can be immediately translated by the host cell proteins (effectively making the genomes themselves mRNA). Negative sense strands are complementary to mRNA and consequently need to be converted to positive sense RNA by an RNA-dependent RNA polymerase, such as the L protein in filoviruses.

Nucleic Acid	Strand Type	Sense	Organisation
DNA	Single	n/a	Linear
			Circular
			Segmented
	Double	n/a	Linear
			Circular
			Segmented
RNA	Single	Positive	Linear
			Circular
			Segmented
		Negative	Linear
			Circular
			Segmented
	Double	Double	Linear
			Circular
			Segmented

Table 1.1. Viral Genome Composition. Possible types of viral genomes. Only single stranded RNA viruses have positive or negative sense genomes. Not all of these possible combinations infect humans.

Furthermore, the shape of the genome can also vary, with viruses exhibiting a linear genome (filoviruses, etc.), a circular genome similar to bacterial plasmids (polyomaviruses) or segmented genomes, split up into multiple strands that exist independently of each other but reside together in the nucleocapsid. The most famous example of segmented genomes is that of influenza virus (Bouvier and Palese, 2008).

1.1.6.1 Genome Size

Alongside genome organisation and virion morphology, the size of a viral genome can vary extensively. The largest known viruses, such as the mimiviruses and pandoraviruses, have genomes exceeding one million base pairs (1mb), which code for over 1,000 proteins (the pandoraviruses code for over 2,500 proteins) (Fischer, 2016; Forterre and Gaïa, 2016; Sharma et al., 2016). This is a far larger genome than that of many small bacteria, several of which code for just a few hundred proteins with genomes of around 180,000 base pairs (0.18mb) (Martínez-García and de Lorenzo, 2016).

On the other end of the spectrum, the smallest viral genomes are found in circoviruses, consisting of only 2,000 base pairs (2kb, 0.002mb). A mere two proteins are encoded by this circular single stranded DNA genomes, Rep and Cap (Belyi et al., 2010). Morphologically circoviruses are some of the smallest known viruses, with average diameters of 20nm, as there is a general correlation among viral species between genome and virion size.

While there is no strict relationship between genome size and composition, RNA viruses tend to be smaller than DNA viruses (though this is not always the case as circoviruses are DNA viruses). The larger the RNA genome, the higher the likelihood of segmentation occurring, in order to reduce errors during translation, as there is no proof reading mechanism for use during RNA synthesis (Sanjuán and Domingo-Calap, 2016). Similarly, while mutations tend to occur more frequently in RNA viruses, single stranded genomes are more prone to mutation than double stranded, regardless of whether they are composed of DNA or RNA (Sanjuán and Domingo-Calap, 2016). However, mutations occur regularly in all viral genomes, generally at a much greater rate than other organisms.

The wide range of genomes found in viral species have allowed them to adapt to more conditions than organisms such as bacteria and plants, and as such have a greater range of habitats than any other organisms. However, despite the great variation, there is a common theme underlying all viral lifecycles and replication methods, irrespective of whether they are infecting a single celled bacterium or a human being.

1.1.7 Viral Lifecycle and Replication

As they are not cells, viruses do not divide or undergo mitosis, rather they use host proteins to copy themselves, essentially hijacking cellular machinery in order to clone themselves. The high volume of viral species gives rise to a great variety of lifecycles and replication processes; however, most viruses roughly follow six major steps.

1.1.7.1 Attachment to the Host Cell

Once a virus is in the vicinity of a viable cell it must bind to specific receptors on the host plasma cell membrane. Different viruses bind to different specific receptors, and as such this step drives much of the species specificity viruses possess. This step also determines how a virus can infect a host and will dictate much of its transmission properties. As an example, filoviruses are known to attach to the T-cell immunoglobulin and mucin domain 1 (TIM-1) receptor, with fusion occurring via interaction with the Niemann-Pick C1 (NPC1) receptor (Kuroda et al., 2015). Hence, cells that possess these receptors in greater quantities are far more susceptible to filovirus infection. Often, secondary receptors such as NPC1 for filoviruses, are needed to help initiate the second step of the viral lifecycle. This initial step must be preceded by the virus locating the viable cells if considering multicellular organisms. For example, Influenza virus attaches to human host cells in the lower respiratory tract, but it must first reach the bronchioles by being inhaled.

1.1.7.2 Viral Entry

In order to enter the host cell, virions must cross the protective outer surface. As this layer is variable depending on the type of cell, viral entry differs greatly. The two main routes of entry across a plasma cell membrane are via membrane fusion, or receptor mediated endocytosis, akin to the transport of other biological molecules (Smith and Helenius, 2004). Membrane fusion is only a viable option for those viruses that contain a lipid envelope, while both enveloped and non-enveloped viruses can undergo receptor mediated endocytosis. However, viruses that infect plants and bacteria must first cross the cell wall, which is far thicker and more rigid. Plant viruses must generally cause some kind of damage to the outer structure in order to gain access, though once in a single plant cell they can move through existing plasmodesmata into adjacent cells, negating the need for transmission across the cell wall (Boevink, 2005). Bacterial cell walls are thinner than those of their plant counterparts as bacterial cells are much smaller. This has allowed some bacteriophages to develop a mechanism whereby they inject their genetic material directly into the cell, while the capsid remains extracellular, eliminating the need for a mechanism to transport the viral proteins across the cell wall (Grayson and Molineux, 2007).

1.1.7.3 Uncoating

Once the virion has entered the cell, the capsid must be removed to expose the genetic material. If the virus has a lipid coating, this is often fused with the host cell surface membrane upon entry, releasing the capsid into the cytosol. Viruses without lipid coats will have been transported to the membrane via receptor mediated endocytosis in a vesicle, which must be degraded to allow the virus access to the rest of the cell. This is of no issue as the host cell naturally degrades its own vesicles, as they are an essential part of the cellular transport mechanism.

Regardless of the method of entry, once the capsid is in the cytosol, it can either dissociate from the nucleic acid due to a change in external conditions (pH, temperature etc.) or host proteins actively degrade the capsid, freeing the genetic material (Bouvier and Palese, 2008; Yamauchi and Greber, 2016). Those bacteriophages that inject their genomes directly do not require this step, highlighting

the variability of lifecycles among different types of viruses (Grayson and Molineux, 2007).

1.1.7.4 Synthesis of Viral Proteins and Genome Replication

With the nucleic acid exposed, host cell proteins will begin to translate the viral genome. This process is perhaps the most variable of the six stages and is defined by the type of genetic material the virus possesses. Single stranded positive sense RNA viruses have genomes that act as their own mRNA and can be translated by host ribosomes immediately. Single stranded negative sense RNA genomes, as well as single stranded DNA genomes, either positive or negative sense, must undergo transcription first to produce mRNA, which can then be translated. This is often carried out by viral RNA polymerase, such as the RNA-dependent RNA polymerase (L) protein in filoviruses (Schmidt and Hoenen, 2017). Double stranded genomes, either DNA or RNA, must be unzipped first much the same as DNA replication in cell division, with transcription and then translation occurring thereafter. Viral mRNA is translated by host ribosomes in the same way that the cell translates its own mRNA. The genetic material is also duplicated.

Often a combination of viral and host proteins is needed to carry out effective replication. In ebolaviruses, VP30 is essential for transcription, while VP35 interacts with Dynein LC8 to regulate the synthesis of viral RNA (Biedenkopf et al., 2016; John et al., 2007; Luthra et al., 2015; Martínez et al., 2008). The hijacking of these pathways can cause severe disruption to the cell, often resulting in apoptosis, as the cell cannot carry out normal functions while its proteins are being utilised by the virus.

1.1.7.5 Virion Assembly

The steps outlined above will produce copies of the viral proteins required to form new virions. These proteins are split into two types, structural and non-structural. Structural proteins converge around the newly replicated genetic material and will go on to form the capsid and will be present in the newly formed virions, (Banadyga et al., 2017a; Han et al., 2003; Kirchdoerfer et al., 2015; Wan et al., 2017) while non-

structural proteins merely aid in the assembly of the virions, acting as enzymes or transporters (Hu et al., 2013). These non-structural proteins would not have been present in the cell before as they will not have been part of the virion that infected the cell, having been left behind during the previous cycle.

Assembly can take place anywhere in the cell, from the nucleus to the cytoplasm to the edges of the cell around the plasma membrane (Barman et al., 2001). In addition, protein modification such as glycosylation can often occur during or after assembly.

1.1.7.6 Exiting the Host Cell

Virion assembly marks the completion of the replication cycle and the creation of brand new virus particles. The next and final step involves the newly formed virions escaping from the host cell to start the same process all over again (Stahelin, 2014). Much of the time this occurs via budding, whereby the capsid is transported to the cell surface membrane and captures some of the cell membrane, converting it to a lipid envelope for the virus, simultaneously allowing the virion to leave the cell (Bouvier and Palese, 2008; Stahelin, 2014). Viral proteins may have placed viral receptors on this membrane prior to budding if required.

Viruses that do not possess a lipid coat will exit the cell via exocytosis, in much the same way that they entered. The host transportation mechanisms will package the virions into vesicles and these will fuse with the membrane, releasing the contents into the extracellular matrix (Olson and Grose, 1997). This process can also be utilised by enveloped viruses as well, but this is more unlikely.

Some viruses can also utilise the host's apoptotic response. When dealing with an infection, host cells will often initiate apoptosis (controlled cell death). This will destroy the cell and allow the virions to escape (Olejnik et al., 2011). One advantage of this process is that immune cells such as macrophages are programmed to respond to apoptosis and will migrate to the dying cell. These macrophages often provide excellent vessels for virions to travel around a host, spreading further throughout the organism. Indeed, many viruses will induce apoptosis once their assembly is complete for this very reason, though some will simply induce apoptosis as a general method of viral exit (Felt et al., 2015; Guo et al., 2015; Mukherjee et al., 2012). Conversely, some

viruses have developed mechanisms to prevent apoptosis, helping them avoid being detected and removed by host immune cells that would migrate to apoptotic cells (Filippova et al., 2002).

Cell death is not the goal of this stage, merely a side effect. Constant viral budding will cause the cell to lose the phospholipid bilayer of its surface plasma membrane quicker than it can regenerate it, causing lysis (Falasca et al., 2015). While exocytosis is less lethal to the cell, constant vesicle hijacking can interfere with vital transportation pathways, leading to a nutrient imbalance, as well as an excessive use of phospholipids for the vesicles, contributing to potential cell lysis. The recognition of viral infection leading to apoptosis is also a major cause of cell death (see viral disease, 1.5, for more detail).

Once the virions have been released into the extracellular matrix, the viral lifecycle is complete, and the process will repeat itself with neighbouring cells. This progression however cannot continue indefinitely, as a host will eventually succumb to the viral infection and die if its cells are constantly destroyed. Additionally, the host will begin to produce antibodies to the virus in order to try and remove it from its system (Snape, 2017). This creates an unwelcome environment for the virus, and therefore a need for the virions to move to a new host arises, a process known as transmission.

1.1.8 Transmission

The aforementioned lifecycle is dependent on finding a suitable host, and the potential "death" of the virus will occur should any of these criteria not be met, for example if the newfound host contains proteins that are incompatible with the virus, or vice versa, and cell entry, exit, or any of the multiple processes in between becomes impossible. And while the internal replication cycle of the virus is crucial, transmission is just as important, as without the ability to spread to new hosts the virus will not be able to survive.

Virion transmission is highly variable and is generally dependent on the host. There are two types of transmission, horizontal (between species in a population) and vertical (passed to offspring prenatally). Human infecting viruses can spread horizontally in many ways, including directly via fluids such as blood (HIV, Ebolaviruses), mucus

(Influenza), water vapour (smallpox) or direct contact, or indirect routes such as airborne droplets, fomites or the faecal-oral route (Chowell and Nishiura, 2014; O'Cofaigh and Lewthwaite, 2013; Rewar and Mirdha, 2014; Song et al., 2017).

The understanding of different viral transmission routes is of great importance to healthcare workers, with isolating and preventing the spread of viruses an essential part of public health research. Vaccination programmes need to consider transmission routes (Do and Lee, 2016; Gittings and Matson, 2016), as discussed later, and it is important to educate populations on how some of the most infectious and lethal viruses spread, in order to help them avoid becoming infected themselves (teaching people not to touch rabid dogs for instance, to avoid contracting rabies) (Velasco-Villa et al., 2017). More recently, we have seen how very closely related viruses can have variable transmission dynamics. While SARS-CoV and SARS-CoV-2 are members of the same species (with current taxonomical thinking placing SARS-CoV-2 as a strain of SARS-CoV), the current COVID-19 pandemic has infected far greater numbers than the original SARS outbreak. Transmission of SARS-CoV-2 appears much easier, as many can pass on the virus before they become symptomatic (Wei et al., 2020).

1.1.9 Human Viral Pathogens

The vast majority of currently known viruses do not cause human disease, and most are not even able to utilise humans as viable hosts (Pradeu et al., 2016). Approximately 200 viral species known to infect humans and cause pathogenesis (Woolhouse et al., 2012). In fact, as is the case with bacteria, the human body is covered in viruses, each providing a different function, each inhabiting a variety of different hosts and each causing no negative effect to the human body. Retroviruses inserted themselves into our chromosomes aeons ago and remained there ever since, driving evolution and providing important functions that humans would not be the same without (Kurth and Bannert, 2010).

Despite all this, viruses maintain their negative perception in the public eye due to the fatal and terrifying diseases that a small handful of species have caused since the dawn of recorded history. Indeed, human civilization as it exists today would not be the same without viruses. The Aztec empire is believed to have been wiped out from smallpox brought to their lands by Spanish explorers (McMichael, 2004). Napoleon I of France

sold the Louisiana Territory to the newly formed United States of America after thousands of French troops died of Yellow Fever in Haiti, and it was deemed too costly to maintain new world colonies (Marr and Cathey, 2013), doubling the size of the USA and guaranteeing its emergence as a global superpower in years to follow. The 1918 Spanish Flu epidemic killed more people than World War One (Taubenberger and Kash, 2011), and the emerging infectious diseases of the tropical regions in Africa, South American and South East Asia in the late twentieth century have struck fear into populations across the world, hindering trade, development and economic prosperity in these areas (Alexander et al., 2015; Ksiazek et al., 2011; Maganga et al., 2014; Weaver et al., 2016).

DNA				RNA					
Double Stranded		Single Stranded		Double Stranded		Single Stranded			
						Positive Sense		Negative Sense	
Envelope	No Envelope	Envelope	No Envelope	Envelope	No Envelope	Envelope	No Envelope	Envelope	No Envelope
Herpes	Adeno		Parvo		Reo	Corona	Astro	Arena	
Pox	Papilloma		Anello**			Flavi	Calici	Bunya	
Hepadna*	Picorna					Retro	Picorna	Filo	
						Toga		Orthomyxo	
								Paramyxo	
								Rhabdo	

Table 1.2. Human Pathogenic Viruses. Viral families known to cause disease in humans, separated by genome organisation. There are no single stranded enveloped DNA viruses, double stranded enveloped RNA viruses or single stranded negative sense non-enveloped viruses known to cause disease. All families end with the suffix -viridae. *Hepadnaviridae is a family that contains partially double stranded and partially single stranded genomes. **Anelloviridae is a recently discovered family and have not yet been attributed to the cause of a specific human disease.

1.1.9.1 Viral Disease

Viral diseases are ingrained into our consciousness and are likely to stay that way for a long time given the historical trauma mentioned so far and the current outbreaks experience in more recent years, including the ongoing COVID-19 pandemic. Even today, after the successful eradication of smallpox, public health campaigns against seasonal influenza, and the use of vaccines for measles, rubella and others, these pathogenic families are responsible for millions of deaths worldwide each year, causing

a wide range of symptoms and maladies (Bloom and Cadarette, 2019; Danielle Iuliano et al., 2017).

Generally, viral disease is caused by two main contributing factors, cell lysis and damage caused by the host's own immune system. The viral replication cycle causes the degradation of the infected cell by membrane loss and virus or host induced apoptosis (see 1.1.7.1 – 1.1.7.6). Excessive cell death will result in severe damage to the organism. The host immune response will cause inflammation and apoptosis as well, leading to further cell death. Infected cells express viral antigens on their surface via their MHC molecules, which act as indicators for T-Cells. Subsequent immune responses result in these cells being degraded, in an attempt to halt the viral infection from spreading to neighbouring cells and tissues. While this process often ends viral infection, it can result in severe damage to the host.

Viral infections can either be considered acute, chronic or latent. The majority of viral infections are considered acute infections, in that they have a rapid onset and outcome, though some do cause chronic diseases, which persist over long periods of time. Examples of acute viral disease include Ebolavirus disease, Influenza and COVID-19 (Borisevich et al., 2006; Chen et al., 2020a; Sonnberg et al., 2013). Examples of latent viral diseases are Herpes simplex virus 1 (HSV1) and Herpes simplex virus 2 (HSV2) (Komaroff, 2006; Michaelis et al., 2019).

Symptoms of HSV1 infection include cold sores around the mouth, lips and nose, while HSV2 causes genitals warts. These viruses are neurotropic and neuroinvasive that survive in the human body and evade the immune system by becoming latent and hiding in the cell bodies of neurons, where immune cells cannot reach them. After the initial infection, individuals will often experience sporadic recurrence of HSV symptoms. This is because occasionally the virions lying dormant in the neurons are activated and transported down the axon to the skin, where virus replication and shedding occur, resulting in new cold sores or warts.

1.1.10 Zoonotic Viruses

A zoonotic disease is a disease caused by a pathogen that has crossed over from a non-human organism to humans. This can occur when a virus mutates, allowing it to infect

humans for the first time, or when humans come into contact with a novel virus that already has the potential for human infection. The former is likely how Human Immunodeficiency virus (HIV) became endemic in the human population (O'Cofaigh and Lewthwaite, 2013), while the latter has been associated with viruses such as Ebola virus and Lassa fever virus (Borisevich et al., 2006; Pourrut et al., 2005). The expansion of the human population into previously uninhabited parts of the planet and increased globalisation are often cited as catalysts for an increase in abundance of zoonotic diseases over the last few decades (Lipkin and Anthony, 2015; Nishiura et al., 2020a; Venter, 2018).

Direct zoonosis is where the pathogen is transmitted directly from infected non-human to human. This can occur in a variety of ways, such as through airborne transmission (avian influenza) or animal bites (rabies) (Lupiani and Reddy, 2009; Velasco-Villa et al., 2017). Indirect zoonosis can occur when an immune or asymptomatic acts as a vector to transmit the pathogen to humans (such as bats harbouring filoviruses that can infect humans) (Olival and Hayman, 2014; Rewar and Mirdha, 2014).

Zoonotic diseases represent a serious impact on global health. One meta-analysis conducted in 2001 determined 61% of all pathogens known to infect humans were zoonotic (Taylor et al., 2001), and many major outbreaks of the 21st century – including Ebola virus, Zika virus and COVID-19 – were caused by zoonotic pathogens (Majid et al., 2016; Nishiura et al., 2020a; Venter, 2018; Wikan and Smith, 2016). The primary viruses discussed in this thesis are all zoonotic in origin.

1.2 Ebolaviruses

Ebolaviruses are a member of the filoviridae family, along with Marburgviruses and Cuevaviruses. The first reported outbreak occurred in 1976 in the Democratic Republic of the Congo (known as Zaire at the time). 318 individuals reported symptoms – ranging from fever, nausea and fatigue to severe haemorrhagic fever and coma – and a total of 280 died, representing an 88% fatality rate (Pourrut et al., 2005). Almost simultaneously there was a separate outbreak in southern Sudan which reported 284 cases and 150 deaths. After a small outbreak in 1979, there were no reported cases until 1994, and from then until 2013, only sporadic and limited cases

appeared across a narrow stretch of sub-Saharan Africa (table 1.3) (Akinfeyeva et al., 2005; Camacho et al., 2014; Pourrut et al., 2005).

There are now six known species of Ebolavirus (Goldstein et al., 2018; Groseth et al., 2007; Kuhn et al., 2011). Zaire ebolavirus (type virus: Ebola virus) is the species known for the aforementioned outbreak in 1976, and has caused the greatest number of outbreaks since (Mahanty and Bray, 2004; Matua et al., 2015; Pourrut et al., 2005). Sudan ebolavirus (type virus: Sudan virus) was responsible for the concurrent 1976 outbreak, and has also caused numerous small outbreaks since (Burk et al., 2016; Kobasa et al., 2016). Both of these viruses were named for the countries their first outbreaks occurred in. In 1994 Tai Forest ebolavirus (type virus: Tai Forest virus) was discovered in Cote D'Ivoire, and as of 2020 is known to have infected just one individual (Akinfeyeva et al., 2005; Pourrut et al., 2005). In 2007 Bundibugyo ebolavirus (type virus: Bundibugyo virus) was discovered in Uganda, and has caused one other outbreak since (Burk et al., 2016; Matua et al., 2015). These four species are all pathogenic in humans, and cause identical illness known as Ebolavirus disease (EVD).

Reston ebolavirus (type virus: Reston virus) is the only known species endemic to Asia, and is not pathogenic in humans, however it does cause disease in swine and non-human primates (NHPs) (Cantoni et al., 2016). It was first isolated in Reston, Virginia in 1987, and since then individuals have tested positive for Reston virus antibodies without displaying any disease symptoms (Albariño et al., 2017).

The most recent species to be discovered was Bombali ebolavirus (type virus: Bombali virus) in 2018 (Goldstein et al., 2018). Found in Sierra Leone, it is not yet known whether it is pathogenic in humans, as it has so far not been responsible for any outbreaks. Speculation as to its pathogenic potential is the subject of chapter three.

At current the natural reservoir of Ebolaviruses has not been determined. Current knowledge points to bats being natural carriers, as well as NHPs (Olival and Hayman, 2014). In 2018 Bombali virus was discovered in free tailed bats, adding support to their role as natural reservoirs (Goldstein et al., 2018). Past outbreaks have been associated with visits to caves and the consumption of bushmeat (Gale et al., 2016; Rewar and Mirdha, 2014).

Year(s)	Nation(s)	Ebola species	Number of Cases	Number of Fatalities	Case Fatality Rate (%)
1976	Sudan (South Sudan)	Sudan	284	150	53
1976	Zaire (DRC)	Zaire	318	280	88
1977	Zaire (DRC)	Zaire	1	1	100
1979	Sudan (South Sudan)	Sudan	34	22	65
1994	Gabon	Zaire	52	31	60
1994	Côte d'Ivoire	Tai Forest	1	0	0
1995	Zaire (DRC)	Zaire	315	250	81
1996	Gabon	Zaire	37	21	57
1996 - 1997	Gabon	Zaire	60	45	75
1996	South Africa	Zaire	2	1	50
2000 – 2001	Uganda	Sudan	425	224	53
2001 – 2002	Gabon	Zaire	65	53	82
2001 – 2002	Republic of the Congo	Zaire	57	43	75
2002 – 2003	Republic of the Congo	Zaire	143	128	89
2003	Republic of the Congo	Zaire	35	29	83
2004	Sudan (South Sudan)	Sudan	17	7	41
2007	DRC	Zaire	264	187	71
2007 – 2008	Uganda	Bundibugyo	149	37	25
2001 – 2002	Gabon	Zaire	65	53	82
2001 – 2002	Republic of the Congo	Zaire	57	43	75
2002 – 2003	Republic of the Congo	Zaire	143	128	89
2003	Republic of the Congo	Zaire	35	29	83
2004	Sudan (South Sudan)	Sudan	17	7	41
2007	DRC	Zaire	264	187	71
2007 – 2008	Uganda	Bundibugyo	149	37	25
2008 – 2009	DRC	Zaire	32	15	47
2011	Uganda	Sudan	1	1	100
2012	Uganda	Sudan	11	4	36
2012	DRC	Bundibugyo	36	13	36
2012 – 2013	Uganda	Sudan	6	3	50
2014	DRC	Zaire	66	49	74
2013 - 2016	Guinea, Sierra Leone, Liberia, Nigeria, Mali, Senegal, USA, Italy, UK, Spain	Zaire	28,646	11,323	39.53
2018	DRC	Zaire	54	33	61
2018	DRC	Zaire	Ongoing	Ongoing	

Table 1.3. Known Ebolavirus Outbreaks. Known Ebolavirus outbreaks that have occurred since the virus was isolated and identified in 1976, excluding laboratory infections caused by contamination during clinical analysis.

1.2.1 Ebolavirus Disease

The four pathogenic Ebolavirus species cause a severe haemorrhagic fever (Ebolavirus Disease, EVD) that can result in fatality rates of up to 90%. Once infected, the incubation period is anywhere between 2 and 21 days (Sousa, 2014). Initial symptoms tend to be extreme tiredness, fever (38^{o}C or above), joint and muscle pain, reduced appetite, nausea, and sore throat, akin to Influenza, followed by the onset of chest pain, confusion and delirium. Approximately seven to ten days post infection vomiting, haemoptysis and bleeding into the whites of the eyes begin to happen, while substantial external bleeding occurs in approximately 30 – 50% of cases. Much of this bleeding usually occurs in the gastrointestinal tract, resulting in excessive blood in the stool of patients (Kortepeter et al., 2011; Vernet et al., 2017).

The severe haemorrhaging associated with EVD is caused by critical disruption and damage to endothelial cells, resulting in compromised and leaky blood vessels. The severe fluid loss from these damaged vessels results in extremely low blood pressure (Kortepeter et al., 2011). As the infection progresses, a massive cytokine storm (the intense discharge of pro-inflammatory signalling molecules) is initiated by the host immune system, resulting in massive organ damage and internal bleeding, ushering in the final phase of the disease. Should death occur, it will usually be between 8 and 16 days post onset of symptoms, and many patients will be comatose at the point of death. Death is generally the result of the drop in blood pressure caused by the damaged blood vessels, but can also be due to blood loss or systemic organ failure in extreme cases (Leligdowicz et al., 2016). Infected persons are not able to spread the virus until the onset of symptoms, and transmission can only occur through direct contact with infected bodily fluids (Rewar and Mirdha, 2014).

1.2.2 The West African Ebolavirus Outbreak

Since their discovery, Ebolaviruses had only caused isolated and contained outbreaks in remote areas of the African continent, resulting in minimal health concern and a focus on isolation and containment (Pourrut et al., 2005). Between 1976 and 2012, just seven African nations had experienced outbreaks (figure 1), with the most severe occurring as a result of Sudan virus in Uganda from 2000 – 2001, in which 425 cases

resulted in 224 deaths (53% case fatality rate). No cases were reported between 1979 and 1994. While still of interest, filoviruses were generally marginalised, and no major danger was perceived (Alexandra Bilak, Martina Caterina, Guillaume Charron, Sophie Crozet, Laura Rubio Díaz-Leal, Florence Foster, Justin Ginnetti, Jacopo Giorgi, Anne-Kathrin Glatz, Kristel Guyon, Caroline Howard, Melanie Kesmaecker-Wissing, Sarah Kilany, Johanna Klos, Freder, 2015; Baize et al., 2014; Moon et al., 2015; Pourrut et al., 2005)

The 2013 – 2016 West African Outbreak drastically changed the concept of an Ebola virus outbreak, when 28,646 individuals became infected and 11,323 deaths occurred (D'Silva and Eisenberg, 2017). With the nations of Guinea, Liberia and Sierra Leone the site of the vast majority of cases, Nigeria, Mali and Senegal were also affected, resulting in this being the most widespread outbreak in history (Shiwani et al., 2017). For the first time Ebola virus arrived in a country via air travel (firstly Nigeria and eventually the USA and several European countries such as Spain and the UK), and cases occurred outside of Africa when international healthcare workers returned to their respective countries, some becoming index cases in their own right.

The West African Outbreak accounts for approximately 70% of all recorded Ebola virus cases, though it is thought between 20-60% of cases went unreported due to unsuitable healthcare infrastructure, indicating the total number of cases to be much higher (Agusto, 2017; Mafopa et al., 2017; Valencia et al., 2017). This outbreak also cemented the virus in the public consciousness, and drew extensive media coverage and led to strict travel restrictions for those travelling from affected areas (Fung et al., 2016; Guidry et al., 2017; Moon et al., 2015; Pandey et al., 2014).

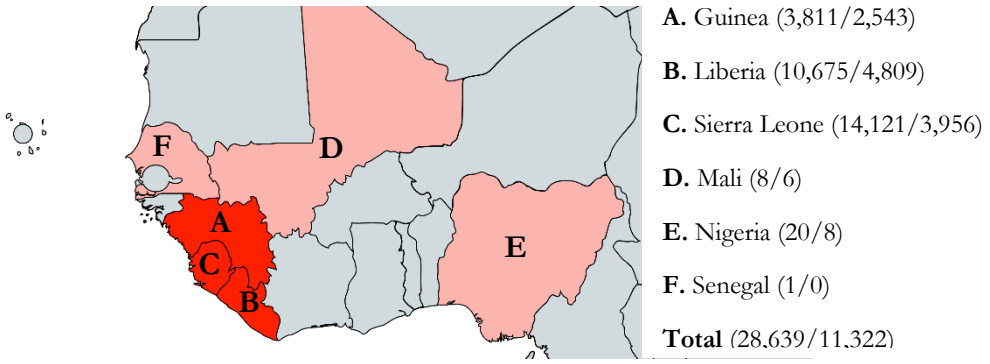


Figure 1.1. Extent of the West African Outbreak. Map of western Africa showing countries with widespread cases (red; Guinea, Liberia, Sierra Leone) and isolated cases (pink; Mali, Nigeria, Senegal) with number of cases and deaths shown.

The extent and severity of the West African Outbreak led to increased interest in Ebolaviruses, and gave a better understanding of transmission dynamics (Agusto, 2017; D'Silva and Eisenberg, 2017; Do and Lee, 2016) and public perception (Guidry et al., 2017; Woods, 2016), as well as the ability to monitor and study patients post infection on a wide scale (Agusto, 2017; Jacobs et al., 2016; Li et al., 2017b). Research into individuals who survived EVD has shown the virus persisting in immune privileged sites such as in semen (Deen, 2018; Sissoko et al., 2017; Uyeki et al., 2016) and the eye, the former potentially allowing for sexual transmission of the disease (MacIntyre and Chughtai, 2016). Research has also been directed at vaccine development. Analysis of this is described in chapter two. Two potential vaccines are of note: rVSV-ZEBOV and ChAd3-ZEBOV.

1.2.3 Ebolavirus Vaccines

Recombinant Vesicular Stomatitis Virus ZEBOV (rVSV-ZEBOV) is vesicular stomatitis virus (VSV) engineered to express the Ebola virus GP protein (Gittings and Matson, 2016; Henao-Restrepo et al., 2017; Trad et al., 2017). It works by provoking an immune response to GP, generating antibodies that can in turn provide protection against Ebola virus infection. VSV itself is a virus found in the *Rhabdoviridae* family, which contains, amongst others, the rabies virus. While it primarily infects animals, mainly cattle, the virus is zoonotic and can cause flu-like symptoms in humans (Heppner et al., 2017). VSV was used as it induces strong interferon response but does not kill host cells (Wong et al., 2014).

Beginning in March 2015, a ring vaccination trial was conducted in Guinea and extended into Sierra Leone, which reported a vaccine efficacy of 70 – 100% (Henao-Restrepo et al., 2017). A total of 5,837 participants were vaccinated during the ring trial, which sought to provide the vaccine to contacts, and contacts of contacts, of all patients confirmed with EVD during the trial run (Henao-Restrepo et al., 2016, 2017). A vaccine efficacy of 100% was reported after 10 days in randomly assigned contacts and contacts of contacts vaccinated. However, a large portion of those identified were not able to be vaccinated, mostly due to a lack of consent. While this vaccine offers a promising respite in the face of future outbreaks, long term protection is not

established by vaccination with rVSV-ZEBOV, with current estimates having a three month protection period as the general average (Ledgerwood, 2015).

Vaccine	Development Stage
Vesicular stomatitis virus–Zaire Ebola virus (rVSV-ZEBOV)	Used in a limited capacity in Guinea
Chimp adenovirus 3 vectored glycoprotein (cAd3-ZEBOV)	Phase 3 clinical trials
Human adenovirus 5 vectored 2014 glycoprotein insert	Phase 1 clinical trials
Adenovirus 26 vectored GP/MVA-BN (Ad26.ZEBOV/MVA-BN)	Phase 1 clinical trials
HPIV-3 vectored glycoprotein	No clinical trials initiated
Rabies vectored glycoprotein	Completed studies on NHPs
Ebola ΔVP30 H_2O_2 treated	Completed studies on NHPs
Purified glycoprotein	Current study on NHPs

Table 1.4. Current *Ebolavirus* Vaccines Under Development. All current vaccines under development by pharmaceutical companies and government run laboratories, in order of trial stage. Two vaccines have undergone clinical trials within an outbreak situation.

Derived from a chimpanzee adenovirus (ChAd3) genetically modified to prevent viral replication in human cells, ChAd3-ZEBOV has been tested in phase 3 clinical trials (Tapia et al., 2016). Also engineered to express Ebola virus GP, ChAd3-ZEBOV was developed in response to the fact human adenovirus type 5 (Ad5) vectors that encode the Ebola virus GP generated immunity against acute lethal Ebola virus exposure in macaques but failed to protect animals who had already acquired immunity to Ad5, suggesting natural Ad5 exposure may limit the effectiveness of the vaccine in humans (Stanley et al., 2014). ChAd3 has no such limitations. Additional research is underway to assess whether GP from Sudan virus can also be expressed and provide immunity (Stanley et al., 2014).

Preliminary results from a clinical trial consisting of 20 individuals detected antibodies to Ebola virus in all participants, giving the vaccine a 100% efficacy (Ledgerwood et al., 2017). However, in NHP models, protection from Ebola virus was short-lived, with 50% of macaques infected three weeks after initial vaccination succumbing to EVD (Stanley et al., 2014). Addition of modified vaccinia Ankara (MVA), an attenuated vaccine of poxvirus, is shown to confer some additional duration to immunity, but results are currently inconclusive (Pavot, 2016). One longitudinal study

has since been conducted suggesting that ChAd3-ZEBOV confers protection for up to 360 days, though with antibody titres drastically reducing over time (Medaglini and Siegrist, 2017).

1.2.4 Genome Organisation and Proteins

Ebolaviruses contain a negative sense single stranded RNA genome approximately 19 kilobases in length, packaged into long cylindrical virions composed of a viral envelope, matrix, and nucleocapsid component (Takada and Kawaoka, 2001). Virions are typically 700 – 1000nm in length with a virally encoded glycoprotein (GP) protruding as spikes around 8nm in length from the lipid bilayer surface. The genome contains seven genes that produce nine proteins – GP, soluble GP (sGP), small soluble GP (ssGP), RNA-dependent RNA polymerase (L), nucleoprotein (NP) and the viral proteins 24, 30, 35 and 40 (VP24, VP30, VP35 and VP40) (Leligdowicz et al., 2016; Takada and Kawaoka, 2001). The three different forms of GP are produced by RNA stuttering at a specific AAAAAAA motif (Mehedi et al., 2011). With such a small genome, it has become clear that many of the proteins have critical multifunctional roles with regards to viral replication, however it is also becoming apparent that all seven proteins contribute to human pathogenicity in one or multiple ways (Brito and Pinney, 2017).

GP is a major focus of research due to its systematic and massive activation of the immune response, leading to the symptoms of EVD as described above (Diehl et al., 2016; Lee et al., 2010; Ueda et al., 2017; Yamaoka et al., 2017). GP is a highly glycosylated dimeric protein consisting of two subunits named GP1 and GP2. GP1 binds to host cell receptors while GP2 is responsible for fusion of viral and host cell membranes (Lee et al., 2010). GP forms a trimeric complex which tethers the virus to endothelial cells, while a soluble form of the protein, sGP, creates a dimeric protein that can interfere with signalling of neutrophils, enabling the virus to evade the immune system by inhibiting early steps of neutrophil activation (Yamaoka et al., 2017). This is crucial to initial infection of host cells. sGP is a soluble form of GP and approximately 80% of all GP produced by the virion exists as sGP.

The GP gene contains a stop codon in the middle of the sequence that results in the production of sGP, which is approximately 360 residues long and well conserved

between species. To produce a full length, (676 amino acid) glycoprotein, a sequence of seven consecutive adenosine nucleotides must have an additional adenosine inserted, resulting in a frame shift that leads to production of the full GP protein instead of sGP (de La Vega et al., 2015; Whitmer et al., 2018). In vivo, L polymerase stutters at this site while carrying out transcription, causing the additional insertion (Mehedi et al., 2011). Due to this method of production, sGP and GP have a near identical N-terminus.

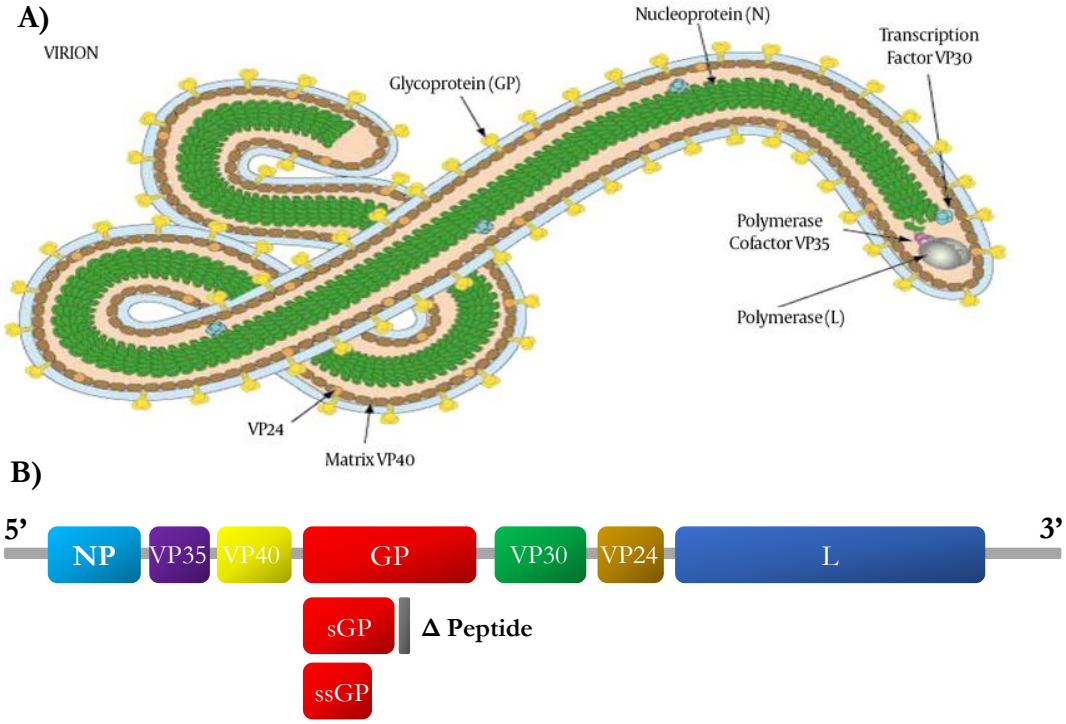


Figure 1.2. Structural and genomic organisation of Ebolavirus. A) Virion structure of Ebolavirus showing the location of the seven proteins in relation to the overall structure. Adapted from Nature and History of Ebola virus: An Overview, Archives of Neuroscience, 2016. **B)** Genome organisation of Ebolaviruses indicating the seven proteins and their location in genome. Also shown is the three isoforms of glycoprotein – GP, sGP and ssGP, as well as the small delta peptide formed during the cleavage of GP.

The nucleoprotein (NP) is a component of the ribonucleoprotein complex along with VP35, VP30, and the RNA-dependent RNA polymerase (L) that is responsible for the transcription and replication of the Ebolavirus genome (Banadyga et al., 2017a;

Kirchdoerfer et al., 2015). NP forms a nucleocapsid that protects the genome from nucleases. (Su et al., 2018).

L is the largest Ebolavirus protein, consisting of over 2000 amino acid residues. It is an enzyme that catalyses the replication of RNA from an RNA template, and forms part of the ribonucleoprotein mentioned above. It is also responsible for the capping and polyadenylation of viral mRNA (Koehler et al., 2016; Trunschke et al., 2013).

VP30 is a transcription factor that forms dimers via its C-terminal domain and hexamers via an oligomerization domain (Pappalardo et al., 2016). VP30 hexamers activate transcription while the dimers do not, with the balance of hexamers and dimers being produced being suggested as a measure to control the balance between viral transcription and replication (Wilson et al., 2001).

VP35 is a multifunctional protein and polymerase cofactor that inhibits interferon signalling by binding to double stranded RNA. This inhibition helps prevent the establishment of a host cell response to Ebolavirus infection. VP35 blocks virus-induced phosphorylation by interacting with and inhibiting host IKBKE and TBK1, which would normally cause the activation of interferon regulatory factor 3 (Trunschke et al., 2013).

VP40 is a matrix protein critical to virion assembly and budding at the plasma membrane of infected cells (Stahelin, 2014). It works by associating with the cell membrane and interacting with the cytoplasmic tails of glycoproteins. VP40 is also involved in regulation of viral transcription, highlighting the multifunctional aspect of Ebolavirus proteins due to the small genome (Han et al., 2003). VP40 can be found as either a hexamer, thought to be involved in the viral budding, or an octamer, which binds to RNA and plays a role in replication (Bornholdt et al., 2013).

1.2.4.1 VP24

VP24 is believed to play a role Ebolavirus pathogenicity and is key to species specific pathogenicity – as described in chapter 3. Mutations in VP24 are induce pathogenicity in a previous non-susceptible novel species. It is known to disrupt the signalling pathway of STAT1. STAT1 becomes phosphorylated by interferons during Ebolavirus infection, causing it to bind to the importer protein and interferon karyopherin-α and

migrate to the nucleus where it stimulates gene transcription in response to viral infection. VP24 binds to karyopherin-α, preventing STAT1 binding, and consequently STAT1 cannot cause an immune response (Schwarz et al., 2017). However, nuclear import does proceed as normal during infection, which has been suggested as important for viral replication. VP24 is incredibly effective as it prevents an Ebolavirus stimulated immune response without halting the cellular pathways required to transport the viral genome to the nucleus of the target cell. VP24 provides Ebolaviruses with an advantage over other viruses which disrupt STAT1 because unlike most other viruses, VP24 mimics the structural binding of STAT1, making it unlikely for the host to develop an adaptation as mutations in karyopherin-α which prevent VP24 may also prevent the STAT1 signalling pathways from functioning correctly as well (Han et al., 2003; He et al., 2017). The protein also appears to aid assembly of viral nucleocapsid, as well as virion budding (Wan et al., 2017).

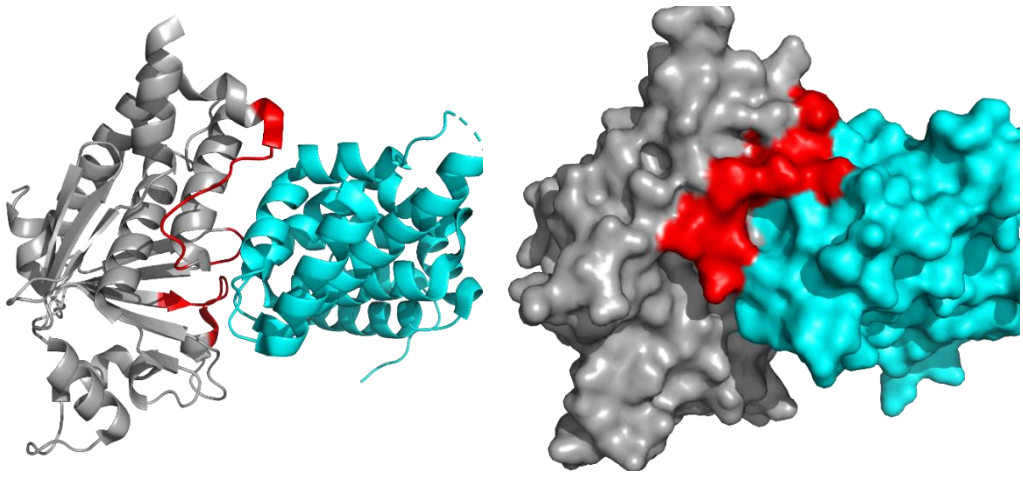


Figure 1.3. VP24. Structure of Ebola virus VP24 (grey), PDB code 4U2X, in complex with human karyopherin-α 5 (cyan). **A)** Structural view. **B)** surface view. VP24-KPNA-α 5 binding site residues on the VP24 protein are highlighted in red.

1.2.5 Ebolavirus Genome Sequences

The West African Outbreak has provided the scientific community with a large number of genome sequences to analyse. This extensive dataset can allow us to study the evolution of the virus over the outbreak or changes that have occurred since initial

discovery and identify differences between Ebolavirus species or strains within species. Prior to 2013 there were just 81 sequences deposited in NCBI for all five species (Bombali virus had not yet been discovered). As of 2020 there are now over 2,600 for the six species, the vast majority of these being from Ebola virus, the species responsible for the WAO.

1.3 Marburgviruses

While much of the current filovirus research is aimed towards Ebolaviruses, Marburgviruses also pose a serious threat. The *Marburgvirus* genus was first identified in 1967 circulating in the West German cities of Marburg and Frankfurt, as well as Belgrade (the capital of then Yugoslavia, now Serbia). Workers in industrial plants were exposed to the tissue of grivet monkeys, and in total 31 individuals were infected, of which seven died (22.5% case fatality rate) of a severe haemorrhagic disease now known as Marburgvirus disease (MVD). In the years since, Marburg virus has caused sporadic outbreaks across the African continent, with the most serious two being the 1998 – 2000 outbreak in the Democratic Republic of the Congo (DRC) and the 2004 – 2005 outbreak in Angola, which together account for over 85% of the total known cases (table 1.5) (Pourrut et al., 2005; Rougeron et al., 2015). In recent years outbreaks have been centred around Uganda, with five separate outbreaks causing 29 cases and 14 deaths since 2007, though the virus has historically been distributed across a wide area in central and southern Africa (Peterson and Samy, 2016; Rougeron et al., 2015). This pattern of outbreaks closely resembles those of Ebolaviruses, and raises concerns that a Marburgvirus pandemic similar to that of the West African Outbreak could arise (Castillo-Chavez et al., 2015). Marburg virus so far has a case fatality rate of 90.8% (table 1.5), having infected 446 known individuals, 405 of which subsequently died.

The *Marburgvirus* genus contains one species, Marburg marburgvirus, which contains two members, Marburg virus (MARV) and Ravn virus (RAVV), which was first characterised in 1996 (Kuhn et al., 2011). Ravn virus is yet to be characterised on a molecular level (Cross et al., 2015). The genomic sequences and organisation, as well as the conservation of individual open reading frames is similar to that of Marburg virus and so it is assumed that Marburg virus and Ravn virus have identical properties and functions, and that all Ravn virus proteins behave analogous to those of Marburg

virus. Additionally, there is no way of discerning which member virus is causing a case of MVD by clinical characteristics alone (Kortepeter et al., 2011).

Year	Type	Location	No. of Cases	No. of Deaths	Case Fatality Rate (%)
1967	Marburg virus	Germany, Yugoslavia	31	7	22.5
1975	Marburg virus	Zambia, South Africa	3	1	33.3
1980	Marburg virus	Kenya	2	1	50.0
1987	Ravn virus	Kenya	1	1	100.0
1998–2000	Both virus	DRC	128*	154*	83.1
2004–2005	Marburg virus	Angola	252	227	90.0
2007	Ravn virus	Uganda	1	0	0.0
2007	Marburg virus	Uganda	3	1	33.3
2008	Marburg virus	Uganda	2	1	50.0
2012	Marburg virus	Uganda	18	9	50.0
2014	Marburg virus	Uganda	1	1	100.0
2017	Marburg virus	Uganda	5	2	40.0
Total			446	405	90.8

Table 1.5. Known Marburgvirus Outbreaks. Marburgvirus outbreaks with number of cases and subsequent fatalities. Two laboratory accidents in Koltsovo (then Soviet Union, now Russia) that occurred in 1988 and 1990 respectively are omitted. Both accident infected one individual, with the 1988 accident resulting in the death of the infected and the 1990 accident ending in recovery. *The 1998 – 2000 DRC outbreak was caused by both Marburg virus and Ravn virus in cocirculation, the number of cases and fatalities caused by each was not recorded.

Much like Ebolavirus, the natural reservoir of Marburgvirus has not yet been determined, however infection has often been correlated with visits to caves, quarries or mines (Pourrut et al., 2005). Marburg virus was isolated from healthy Egyptian fruit bats in 2009, and this species is known to be endemic in the African nations that have experienced outbreaks and have been found inhabiting the caves and mines that were the location of the 1967; 1998 - 2000 and 2012 – present outbreaks (Burk et al., 2016). While it is known that NHPs can contract MVD, only the 1967 outbreak has resulted in human cases from exposure to NHPs. Despite the lack of data on Marburgvirus transmission, it is thought that it occurs in much the same way as Ebolavirus (Olival and Hayman, 2014; Peterson and Samy, 2016).

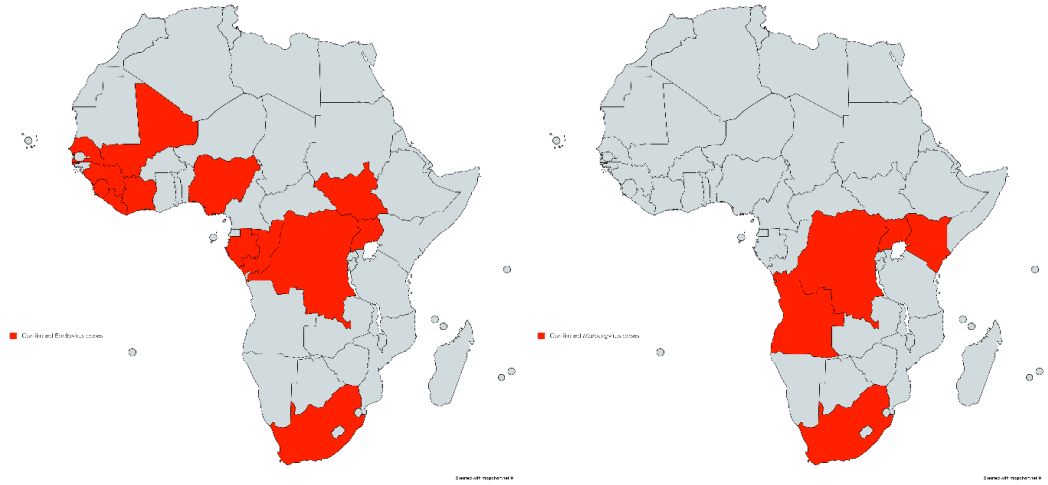


Figure 1.4 African nations with confirmed filovirus infections. A) African countries that have reported cases of Ebolavirus disease. 13 countries in total have been affected, ranging from a single case (Senegal, Cote d'Ivoire) to 14,124 known cases (Sierra Leone). Multiple outbreaks have been reported in five countries: South Sudan (4), Gabon (5), Republic of the Congo (6), Uganda (6) and DRC (10). **B)** African countries that have reported cases of Marburgvirus disease. 5 countries in total have been affected, ranging from three cases (South Africa, Kenya) to 252 cases (Angola). Multiple cases have been reported in two countries: Kenya (2) and Uganda (6).

1.3.1 Marburgvirus Disease

Marburg virus disease (MVD) is clinically indistinguishable from Ebola virus disease (EVD). The incubation period occurs between two and 21 days, with a typical average of five to nine days. High fever, nausea, headache, fatigue, vomiting, abdominal pain and general malaise generally occur between one and five days after the onset of include initial symptoms. These symptoms lead into prostration, dyspnoea, oedema, encephalitis, confusion and delirium from around days five to thirteen. The symptoms typically associated with haemorrhagic fevers occur late in this stage and can range from bloody stools, ecchymoses, blood leakage around needle puncture sites and mucosal haemorrhaging (Rougeron et al., 2015).

After approximately 13 days, infected individuals will either enter a recovery phase (though they may still suffer from fibromyalgia, ocular issues and potential psychosis), or will succumb to fatal MVD, often entering a coma and dying of shock. Death is

rarely due to the hypovolemia associated with MVD, but is instead caused by systemic organ failure, fluid redistribution and hypotension (Chowell and Nishiura, 2014; Mahanty and Bray, 2004; Marzi et al., 2016).

As well as being identical to EVD, MVD can also be confused with other haemorrhagic fevers, typhoid fever and some types of malaria, as well as the non-infectious acute promyelocytic leukaemia, haemolytic uremic syndrome and Kawasaki disease, among many others (Borisevich et al., 2006; Burk et al., 2016). The most important indicator of MVD is examination of the patient's medical and occupational history, particularly exposure to wildlife (particularly bats). At current no treatment exists for MVD, and prognosis is poor (Ursic-Bedoya et al., 2014).

1.3.2 Marburgvirus Proteins

Marburgviruses are negative sense, single stranded RNA viruses that encode seven proteins from seven genes: Nucleoprotein (NP), VP35, VP40, Glycoprotein (GP), VP30, VP24 and RNA-dependent RNA polymerase (L). They are approximately 80nm in width and have a median length of around 800nm, slightly shorter than *Ebolaviruses* (1,000nm) (Banadyga et al., 2017b). The *Marburgvirus* genome is approximately 19kb in length (Ursic-Bedoya et al., 2014). Unlike Ebolaviruses, Marburgviruses do not produce sGP or ssGP, only full length GP.

While it is suspected that VP24 is a key protein in the pathogenicity of Ebolaviruses, via its interaction with Karyopherin-α, this does not appear to occur with Marburgviruses. In contrast, VP40 appears crucial to Marburgvirus pathogenicity (Feagins and Basler, 2015; Valmas and Basler, 2011). Research has also highlighted a difference in the mechanism of activity of VP35 in its RNA binding capacity between Ebolaviruses and Marburgviruses (Edwards et al., 2016). It is therefore thought that despite their similar genomes, the mechanisms of the differing filovirus genera proteins are variable.

1.3.3 Other Filoviruses

In addition to Ebolavirus and Marburgvirus, a third potential filovirus genus discovered in Spain in 2002 – Cuevavirus – contains one supposed species so far, Lloviu cuevavirus, which itself contains a single member, Lloviu virus (Negredo et al., 2011). At current this virus has not been characterised or fully isolated, but highlights an interesting avenue of research for filoviruses. It is currently believed to be non-pathogenic like Reston virus, however this stems from the fact that it was found in an area of high population where no haemorrhagic fevers have occurred (Burk et al., 2016; Negredo et al., 2011). This works on the assumption that many people have been infected yet display no symptoms, yet neither antibodies against Lloviu virus nor the virus itself have ever been isolated from a human patient, leading much of the thought on this species theoretical.

In January 2019 a novel filovirus was also discovered in Yunnan province, China (Yang et al., 2019). This virus was isolated from fruit bats, further adding to the evidence that bats are natural reservoirs of filoviruses. This virus is so far believed to have not infected humans. This virus is part of a fourth filovirus family, named Dianlovirus, indicating it is genetically quite different from Ebolaviruses and Marburgviruses. The family contains one species, Mengla virus (Yang et al., 2019).

Genus	Species	Member	Abbreviation	Discovered	Pathogenic?
Ebolavirus	Zaire ebolavirus	Ebola virus	EBOV	1976	Yes
	Sudan ebolavirus	Sudan virus	SUDV	1976	Yes
	Reston ebolavirus	Reston virus	RESTV	1989	No
	Tai Forest ebolavirus	Tai Forest virus	TAFV	1994	Yes
	Bundibugyo ebolavirus	Bundibugyo virus	BDBV	2007	Yes
	Bombali ebolavirus	Bombali virus	BOMV	2018	Unknown
Marburgvirus	Marburg marburgvirus	Marburg virus	MARV	1976	Yes
		Ravn virus	RAVV	1987	Yes
Cuevavirus	Lloviu cuevavirus	Lloviu virus	LLOV	2002	Unknown
Dianlovirus	Mengla virus	Mengla virus	MLAV	2019	Unknown

Table 1.6. Filovirus Classification and Nomenclature. Correct and standardised classification and nomenclature of the genus, species and member viruses of the

known members of the filoviridae family, including internationally recognised abbreviations and date of discovery. Classification is dictated by the ICTV (Morgan, 2016).

1.4 Coronaviruses

The coronaviridae family is a group of positive sense single stranded RNA viruses (Chen et al., 2020b). All members of the family have a similar enveloped circular virion shape with protruding glycoproteins, hence the name 'corona', which is Latin for 'crown'. The family is organised into two sub-families, five genera, 23 sub-genera and approximately 40 species (Pickett et al., 2012a). The most relevant of these is the betacoronavirus genus, part of the Orthocoronavirinae sub-family, as it contains the majority of the known human pathogenic coronaviruses.

The betacoronavirus genus is subdivided into four lineages: A, B, C and D. Each lineage is a given sub-genera. All members of the genus are similar in structure and genome organisation, having large genomes ranging from 26kb to 32kb. There is some variance between the groups, the most obvious being group A possessing a secondary protruding spike protein called hemagglutinin esterase that is not present in the other three groups. There are seven currently known human pathogenic coronaviruses, five of which are betacoronaviruses. Human coronavirus OC43 (HCoV-OC43) and Human coronavirus (HCoV-HKU1) are members of the A lineage, while Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) are members of lineage B (Lu et al., 2020). Middle East Respiratory Syndrome Coronavirus (MERS-CoV) is a member of the C lineage, and no human pathogenic species have been discovered belonging to lineage D (Cui et al., 2019; De Wit et al., 2016). The additional two pathogenic coronaviruses are Human coronavirus 229E (HCoV-229E) and Human coronavirus NL63 (HCoV-NL63), both alphacoronaviruses. Diseases that occur as a result of coronavirus infection appear to cause significantly more deaths in males than in females, currently for unknown reasons.

Lineage	Species	HCoV-HKU1	HCoV-OC43	HCoV-NL63	HCoV-229E	MERS	SARS-CoV-2	SARS-CoV
A	HCoV-HKU1	100	72.38	47.61	46.13	47.74	48.87	47.7
A	HCoV-OC43		100	45.83	45.36	46.98	47.87	47.5
n/a	HCoV-NL63			100	66.28	44.95	46.16	45.21
n/a	HCoV-229E				100	47.52	45.63	45.19
C	MERS					100	49.78	49.57
B	SARS-CoV-2						100	79.68
B	SARS-CoV							100

Figure 1.5 Percentage identity matrix for human pathogenic coronaviruses. The percentage identity of the full length genome sequence of the seven known human pathogenic coronaviruses to each other.

1.4.1 Coronavirus Disease 2019

The Coronavirus Disease 2019 (COVID-19) pandemic has rapidly grown to become one of the most influential viral outbreaks of moderns times, and one of the defining events of the 21st century. Not since the 1918 Spanish flu pandemic has an acute disease reached such global proportions (Zhou et al., 2020).

First identified in Wuhan, China, in December 2019, COVID-19 is caused by the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Lu et al., 2020; Wu et al., 2020c). Within four months the virus had spread to more than 200 countries and territories, causing global cases to reach two million plus as of April 2020, with over 200,000 deaths. COVID-19 is classified as an acute respiratory disease, with symptoms including a dry cough (in 68% of cases), high fever (88%) and fatigue (38%) (Borges do Nascimento et al., 2020; Yang et al., 2020). Less common symptoms include muscle pain, loss of smell and abdominal pain, and these are experienced by approximate 14% of infected individuals. Approximately 5% of patients develop severe symptoms such as viral pneumonia and require a ventilator for assisted breathing, with some experiencing multiple organ failure and death (Team, 2020). Many patients are

asymptomatic (Pan et al., 2020). The virus is spread by close contact, and can remain on surfaces for up to nine days depending on the surface (Kampf et al., 2020). Infected individuals are most contagious during their first three days post symptom onset, though transmission may be possible before symptoms arise.

Due to lack of testing and the rapid increase in cases, it is not possible to determine the exact number of cases, and therefore the accurate case fatality rate (CFR). Current estimates suggest around 6% of those infected will die, however much of this data is based on hospital submitted patients and not those who self-isolated with symptoms, suggesting the CFR could be much lower (Nishiura et al., 2020b).

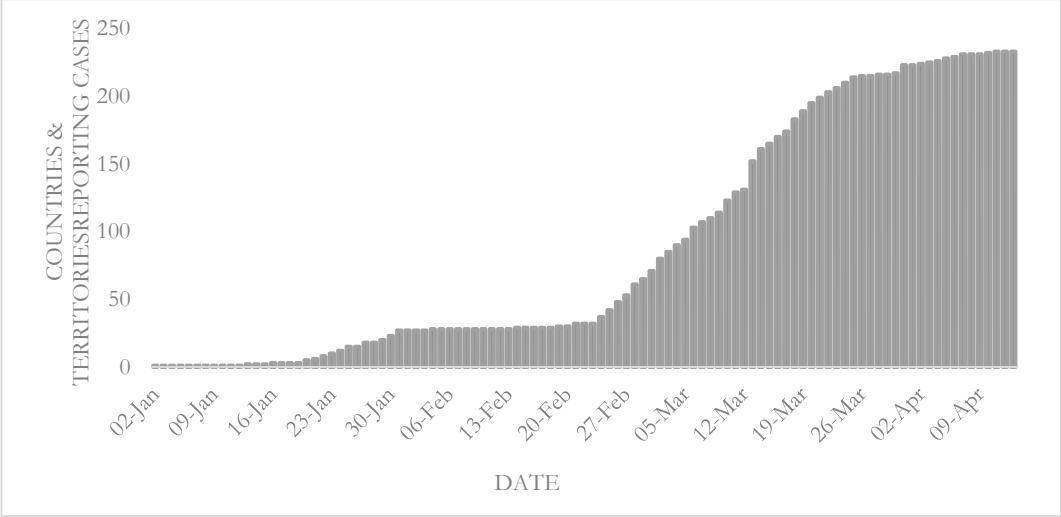


Figure 1.6 Total number of territories reporting COVID-19 cases. The total number of countries and territories reporting any COVID-19 cases over time, indicating the territorial spread of the virus.

Reactions to the disease have been sweeping and drastic, causing severe economic and social disruption on a scale not seen in generations. Many countries instituted mandatory lockdowns and curfews, with social gatherings banned and public events cancelled. The global response highlights the capacity for viruses to overrun our healthcare systems and emphasises the need for continued research and funding into vaccine research, virology and epidemiology. At current, no vaccine or treatment is available, with rest and isolation recommended to those with mild symptoms and the use of a ventilator and additional oxygen given to those with severe cases.

1.4.2 The SARS Pandemic

There has only been one previously coronavirus outbreak similar to the ongoing situation. While MERS (caused by the Middle East respiratory syndrome coronavirus, or MERS-CoV) is currently circulating globally and continues to cause sporadic epidemics, the virus only infects few individuals at a time and has just 1,360 confirmed cases since its discovery in 2012, with 527 deaths (an approximate 36% CFR) (Widagdo et al., 2017). In comparison, SARS (caused by the Severe acute respiratory syndrome virus, or SARS-CoV) was responsible for a much more acute pandemic.

Species	SARS-CoV	MERS-CoV	SARS-CoV-2
Disease	SARS	MERS	COVID-19
Number of Cases	8,096	1,360	2,000,000+
Number of Deaths	774	527	150,000+
Countries Affected	29	26	233
Case Fatality Rate (%)	9.56	38.8	6.2
First Discovered	2002	2012	2019
Lineage	B	C	B

Table 1.7 Comparison of known coronavirus outbreaks. Basic comparison of the three major coronavirus outbreaks, SARS, COVID-19 and MERS, comparing case numbers, deaths and case fatality rates.

The 2002 – 2004 SARS pandemic also began in China when a novel respiratory illness was reported in November 2002 (Cui et al., 2019; De Wit et al., 2016). Infected individuals exhibited flu-like symptoms including a fever above 38^0C, dry cough and fatigue. The only common symptom for all patients was fever. Severe symptoms included pneumonia and organ failure similar to COVID-19. The incubation period is between four and six days, but varied anywhere in the range of two to 14 days.

SARS quickly spread globally, causing infections in 29 countries and a total of 8,096 confirmed cases. 5,327 of these (67.8%) occurred in mainland China. The WHO officially declared the outbreak contained in July 2003, however several cases continued to be reported until May 2004. As a result of the initial spread of the virus, restrictions were imposed on travel to and from affected nations, schools were closed, and quarantines went into effect. These measures were successful, as the spread was quickly halted. SARS was estimated to have a basic reproduction number (R_0 – described in detail in section 1.5.1) of between 2 and 4, though according to the WHO,

measure implemented to stem the transmission spread reduced the R_0 to 0.4 by April 2003.

SARS-CoV and SARS-CoV-2 are similar in terms of their genome and structure, as well as method of transmission and symptoms. It would appear that SARS-CoV-2 can transmit more easily, as shown by the massive disparity in case numbers (though some of this can be attributed to sociological factors). The R_0 for SARS-CoV-2 was initially estimated somewhere between 1.4 and 2.5, however this is since under revision and is considered to be much higher, consistent with that of SARS-CoV.

1.4.3 Coronavirus Genome Organisation and Proteins

All members of the Coronaviridae family have similar genomes, though there is variation. Genome sizes can range from 26kb to 32kb, which is the second largest RNA genome currently known in viruses (Saberi et al., 2018). A common feature is the large majority of the genome encodes for the ORF1ab polyprotein (sometimes also referred to as rep). Approximately 7,000 amino acid residues in length, ORF1ab accounts for around 70% of the total genome, and is post translationally cleaved into multiple non-structural proteins.

In additional, coronaviruses possess four structural proteins known as envelope (E), membrane (M), nucleoprotein (N) and spike glycoprotein (S). E acts as a viroporin, and self-assembles in host cell membranes (Bartlam et al., 2005). Here it forms a pentameric protein-lipid pore allowing for ion transport. E activates host NLRP3 inflammasome which leads to IL-1beta overproduction and the associated immune response (Nieto-Torres et al., 2015). M is a transmembrane protein, a component of the viral envelope that along with other viral proteins is involved in viral morphogenesis and assembly (Yuan et al., 2005). N is involved in packaging the viral RNA into a helical ribonucleocapsid, and is crucial for virion assembly, interacting with both the viral genome and M (Bartlam et al., 2005). The spike protein is discussed below.

In addition to these four structural proteins and ORF1ab coronaviruses produce a number of 'accessory' proteins. These are generally not well characterised, and much of their functions are not known or three dimensional structures solved. One notable

exceptions to this is the 3a protein of SARS, which is known to associate with M and E, aiding in both the forming of ion channels and the inducement of apoptosis in the host cell. The accessory proteins vary between species, with many being unique to a specific species, making comparison of these proteins between groups difficult.

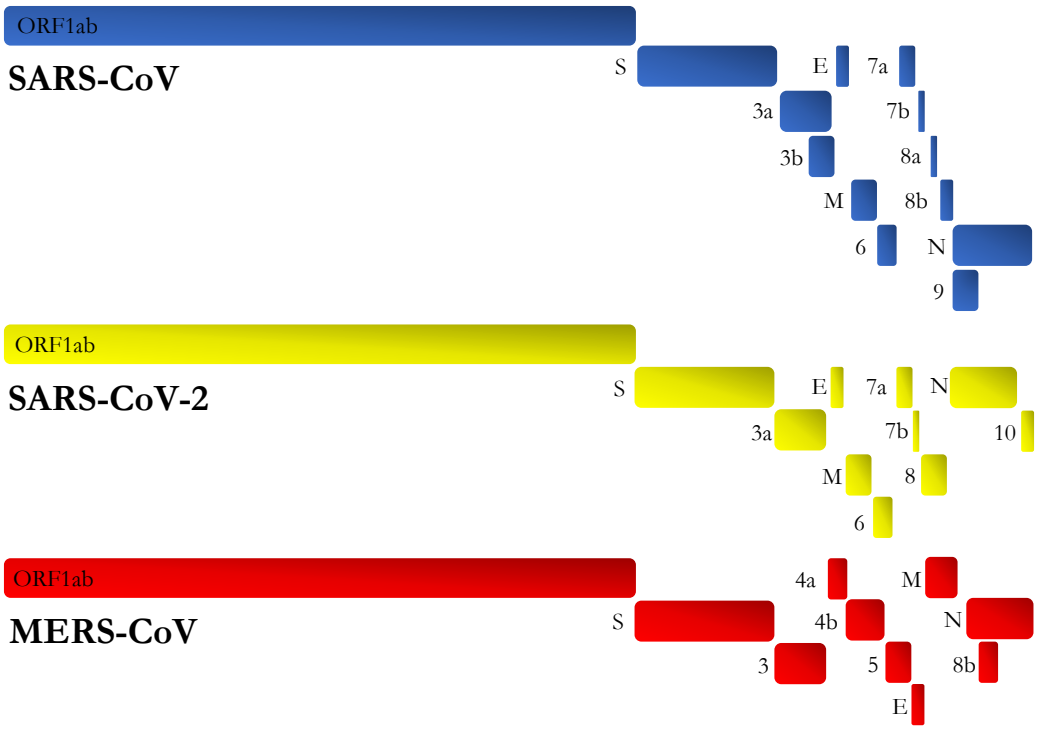


Figure 1.7. Coronavirus Genome Organisation. Basic genome organisation for the three main human pathogenic coronaviruses, SARS-CoV (blue), SARS-CoV-2 (yellow) and MERS-CoV (red). The ORF1ab polyprotein is post-translationally cleaved into 16 non-structural proteins which are not shown here.

1.4.4 Spike Protein

The coronavirus spike 'S' glycoprotein (S) is a multifunctional protein that mediates entry into the host cell, composed of two subunits. Initially it binds to a receptor on the host cell surface via the S1 subunit, and then fuses viral and host membranes via the S2 subunit (Reinke et al., 2017; Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, 2020).

The S protein is 1,255 residues in length in SARS-CoV and 1,279 residues in length in SARS-CoV-2. The first 1,195 residues are extracellular, with residues 1,196 – 1,216 being a transmembrane region and residues 1,217 – 1,255 being cytoplasmic (Reinke et al., 2017; Yan et al., 2020).

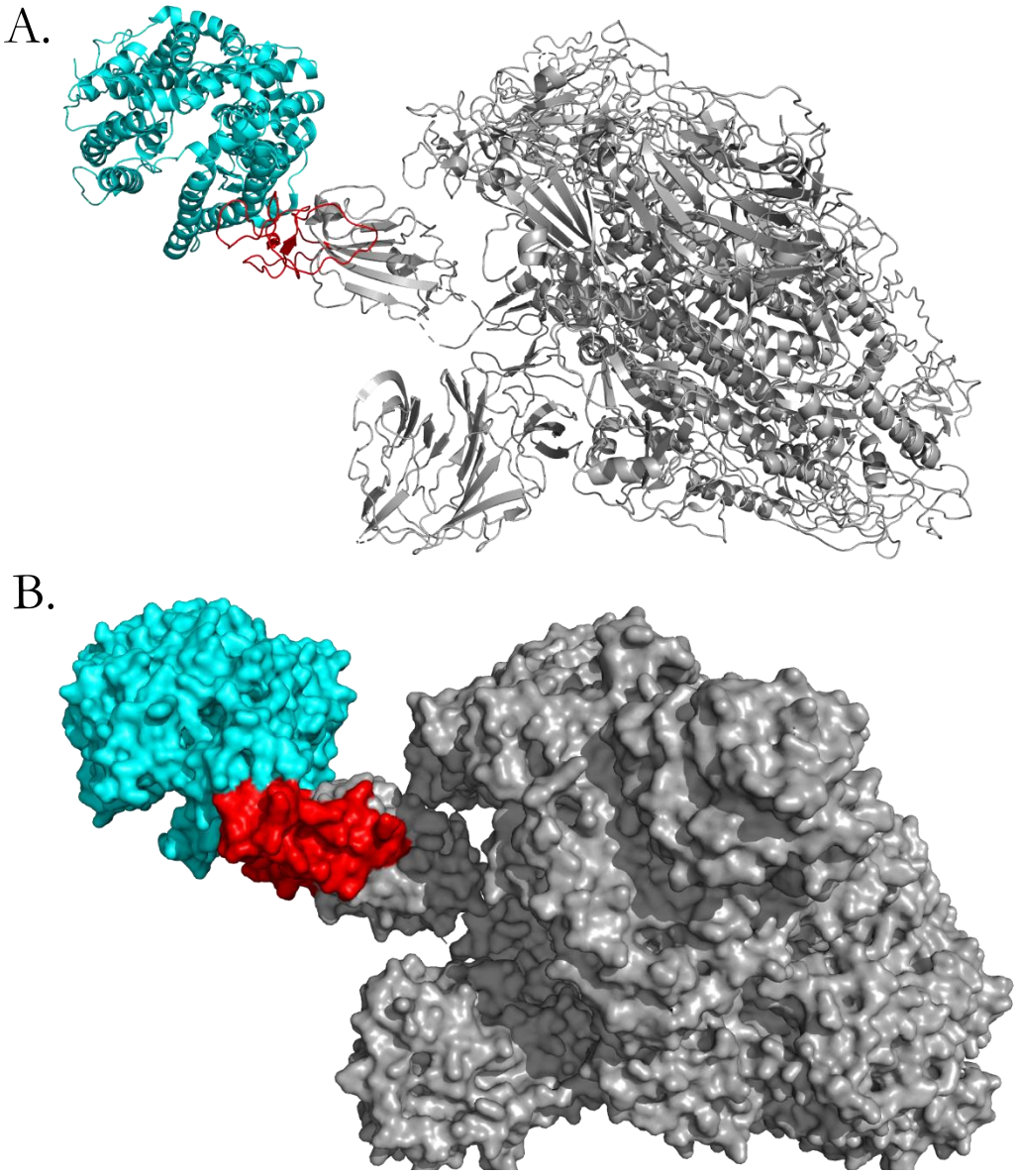


Figure 1.8. SARS-CoV-2 S Protein. The structure of the S protein from SARS-CoV-2 (grey) in complex with ACE2 (cyan). The ACE2 binding region of S is highlighted in red. **A.** Structural view. **B.** Surface view.

There are two protease cleavage sites in the protein - residue 667 and residue 797 (Heurich et al., 2014). The transmembrane serine protease TMPRSS2 cleaves and

activates the SARS S protein in cell culture and potentially also in the infected host. Residue R667, a known trypsin cleavage site, is required for S protein cleavage by TMPRSS2 but is dispensable for TMPRSS2-mediated S protein activation. Conversely, residue R797, previously reported to be required for SARS S activation by trypsin, was dispensable for S protein cleavage but required for S protein activation by TMPRSS2 (Iwata-Yoshikawa et al., 2019; Matsuyama et al., 2020; Reinke et al., 2017).

The SARS S protein is also known to interact with the human ACE2 protein. Two virus-binding residues have been identified on human ACE2, L31 and K353, respectively (Hoffman et al., 2020; Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M and 722 JL, Luzuriaga K, Greenough TC, Choe H, 2004; Matsuyama et al., 2010). Both residues provide a salt bridge in a hydrophobic environment and contribute to virus-receptor binding. Residues 479 and 487 in S interact closely with these residues, and are noted of being important to interaction between the two proteins (Li et al., 2005). Analysis of the S protein and its potential as a drug target, as well as a comparison between the SARS-CoV and SARS-CoV-2 structure and function, are discussed in chapter four.

1.5 Herd Immunity and Vaccination

In comparison to other pathogens such as bacteria, fungi and parasitic lifeforms which have many cures, treatments and preventative therapies, there exists little in the way of vaccines or effective post infection treatments for viruses. Much of this can be attributed to the later discovery of viruses, and the complications in cultivating and experimenting on them. Another reason is that while bacteria have more than hundreds of genes producing hundreds of proteins to target, as well as cell walls, envelopes and other cellular components (Den Blaauwen et al., 2014; Brötz-Oesterhelt and Sass, 2014), viruses often possess fewer than ten proteins, presenting a difficulty in finding suitable therapeutic targets on the virions themselves (Bouvier and Palese, 2008; Harden and Munger, 2017). Viral lifecycles also exploit host cell metabolic pathways, and so are hard to eradicate without triggering toxic and damaging effects to the host cells. However despite this there have been many excellent developments in vaccine research regarding viruses.

Vaccines work by stimulating the immune system to develop an adaptive immune response to a pathogen before or immediately after they are infected. The vaccine either consists of or mimics a part of the pathogen, tricking the body into thinking that an infection has occurred and results in the generation of virus specific antibodies and B cells. Subsequently when the virus infects the vaccinated individual, the circulating antibodies recognise and dispose of the virions before they can infect and damage a high enough proportion of cells to cause disease (Pavot, 2016; Plotkin, 2014; Shader, 2017).

1.5.1 Herd Immunity

Vaccine strategies often need to take into account their ability to create herd immunity. Herd Immunity is the concept that vaccines protect more than just the individuals receiving them, they also protect individuals who have not been vaccinated by immunising a critical percentage of the populace (Metcalf et al., 2015; Plans-Rubió, 2012). Herd immunity breaks the transmission chain, and is established when the number of protected individuals (I) becomes greater than the herd immunity threshold (I_c) (Smith, 2010).

Establishing herd immunity to a pathogen relies on establishing the basic reproductive number (R_0). This is a measure of the number of secondary cases occurring as a result of one primary case (Smith, 2010). Using the R_0 we can identify I_c using the equation

$$I_c = 1 - (1/R_0)$$

which gives us the herd immunity threshold assuming that the vaccines is 100% effective. In reality no vaccines is 100% effective and only a reduced percentage of individuals will be protected via vaccination. This number is referred to as E and is the vaccine effectiveness. E has an impact on the ability to achieve herd immunity, and is tied to the critical vaccination coverage (V_c), which can be determined using the following formula

$$V_c = I_c/E$$

As E increases, fewer individuals require vaccination in order to provide herd immunity. It is therefore desirable to have as high a vaccine efficacy as possible.

Potential barriers include individuals proving unresponsive or an unwillingness be vaccinated. These values are crucial to consider during vaccine development. As shown in table 1.8 a vaccine with a 70% efficacy would never establish herd immunity against a pathogen with an R_0 value of 3.5.

	Vc (at varying R0 values)								
E	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
10	0.00	333.00	500.00	600.00	667.00	714.00	750.00	778.00	800.00
20	0.00	166.50	250.00	300.00	333.50	357.00	375.00	389.00	400.00
30	0.00	111.00	166.67	200.00	222.33	238.00	250.00	259.33	266.67
40	0.00	83.25	125.00	150.00	166.75	178.50	187.50	194.50	200.00
50	0.00	66.60	100.00	120.00	133.40	142.80	150.00	155.60	160.00
60	0.00	55.50	83.33	100.00	111.17	119.00	125.00	129.67	133.33
70	0.00	47.57	71.43	85.71	95.29	102.00	107.14	111.14	114.29
80	0.00	41.63	62.50	75.00	83.38	89.25	93.75	97.25	100.00
90	0.00	37.00	55.56	66.67	74.11	79.33	83.33	86.44	88.89
100	0.00	33.30	50.00	60.00	66.70	71.40	75.00	77.80	80.00

Table 1.8 Critical vaccination coverage versus efficacy. The minimum percentage of a population requiring vaccination to establish herd immunity for varying R_0 values and vaccine efficacies. The greater the efficacy the lower the number required to vaccinate, while the greater the R_0 the higher the number required. Cells shaded in grey indicate values impossible to obtain.

Herd immunity is not always the main goal of vaccination programmes, and many vaccines are developed without this in mind. Protection of an entire population is time consuming and often logistically impossible, and will only truly work in a closed population centre with no net change in number of individuals (Anderson, 1992; Holzmann et al., 2016; Metcalf et al., 2015; Paulke-Korinek et al., 2011). The difficulties with achieving herd immunity with relation to Ebolaviruses is discussed in chapter two.

1.5.2 Successful Vaccination Programmes

As of July 2020 there has only been one human pathogenic virus eradicated using mass vaccination programmes. In 1979 smallpox was officially declared extinct after a decades long mass vaccination programme instigated by the WHO and other healthcare agencies around the world (Moore et al., 2006; Sánchez-Sampedro et al., 2015). In 1959, two million people globally died of smallpox each year, yet within just

fifteen years the virus only remained in a small portion of the horn of Africa, a region which was declared smallpox free in late 1977 (Moore et al., 2006).

The smallpox vaccine is the first known vaccine developed in the modern world, dating back to 1796 and British doctor Edward Jenner. Despite the nearly 200 years of readily available vaccine, it took a united global effort and extensive research in microbiology and epidemiology to finally eradicate the virus, highlighting the difficulties with vaccination programmes (Sánchez-Sampedro et al., 2015; Thèves et al., 2014).

In more recent years this has been highlighted by two further viruses: Polio and Rabies. The WHO declared a global campaign against Polio (also known as poliomyelitis and caused by the Polio virus) in 1988, at which point there were an estimated 350,000 cases annually (Bonanni et al., 2014; Shader, 2017). Since then, cases have been reduced by 99.95% to 175 in 2019, with the virus endemic in just two countries – Afghanistan and Pakistan. Unfortunately the attempt to fully eradicate Polio through mass vaccination has stalled in recent years, with the numbers remaining steady in the 2010s, despite an aggressive push to increase vaccination efforts (Lee et al., 2016; Plotkin, 2014). More disruption to this effort came in early 2020 when local programmes were halted to divert resources to the ongoing COVID-19 situation, highlighting the logistical and financial issues plaguing such programs.

In terms of Rabies (caused by the Rabies virus), vaccination programmes that have been in place since the late 19th century have been successful in drastically reducing case numbers, but total eradication appears impossible as the virus has multiple natural hosts (Albertini et al., 2008; Velasco-Villa et al., 2017). Smallpox only existed in humans, meaning breaking the chain of transmission was manageable if not difficult. With a virus that can claim bats, dogs and many other mammals as a natural host, total eradication is unlikely. Herd immunity is unlikely to be achievable as many of the countries in which the disease is endemic are in the developing world, and do not have the infrastructure or funds available for such an intense scheme (Holzmann et al., 2016). Indeed, as of 2020 the only nations that have successfully eradicated the Rabies virus are in western Europe, which possesses robust and well-funded healthcare institutes, or island nations such as Japan or Iceland, where they can effectively control the wild animals that enter the country, depriving the virus of a viable host (Velasco-Villa et al., 2017).

There are many other pathogens that are the subject of vaccination programmes, whether attempting to achieve herd immunity or not. They include but are not limited to Measles, Rubella, Malaria and Syphilis. The only other pathogen that has been successfully eradicated or controlled is Rinderpest, a virus endemic to cattle, which was declared extinct in 2011 (Bonanni et al., 2014; Cataldi et al., 2016; Holzmann et al., 2016; Lee et al., 2016).

1.5.3 Antivirals

Despite recent advances, most viruses do not have effective vaccine programmes, though many have suitable antiviral drugs available. Many antivirals come in the form of nucleoside analogues, inactive nucleotides that are erroneously incorporated into the viral genomes during their replication cycle, halting the viral lifecycle. One such drug is ribavirin, which combats Hepatitis C infections (Rower et al., 2015). Another class of antivirals can act as inhibitors, competitively binding with proteins that play a vital part of the viral lifecycle. Oseltamivir acts as an inhibitor to the Influenza neuraminidase protein, preventing sialic acid cleavage and host cell exit, leaving the virus trapped inside the cell unable to continue the chain of infection (Ciftci et al., 2016). Antiviral drugs often have adverse side effects, and are not as effective as vaccines (Bernardeschi et al., 2016; Yen, 2016), with most viral infections simply being left to run their course. Rest and fluid intake is recommended, alongside general pain relief (Allan and Arroll, 2014; Chowell and Viboud, 2015). The consideration of drugs as potential antivirals is considered in the research described in chapter four.

1.6 Genetic Variation

Protein variation arises as a result of variance in the genetic code of organisms. The understanding of genetic variation is one of the largest avenues of research in the scientific community, and the phenotypic results of genotypic variations has wide reaching consequences for health and disease. Genome sequencing has improved rapidly in recent years, with perhaps the biggest breakthrough coming with the sequencing of the human genome in 2000.

Genome variation can occur in a variety of ways. Point mutations are a position at which a single base differs between two sequences, while insertions or deletions are where nucleotide bases are added or removed to the genome, causing a frameshift (provided the number of nucleotides inserted or deleted is not divisible by 3 which would result in the addition or removal of codons). Variants can occur in both the coding and non-coding region of an organisms genome. The impact of variants in coding regions can be well characterised, and many in the human genome lead to disease. Aside from those variants that appear in known regulatory regions, the impact of variants in non-coding regions is less clear, as the in general much is unknown about non-coding regions of most organisms.

Point mutations in coding regions can be either synonymous; whereby the resulting change to the codon will not alter the amino acid produced, or non-synonymous, where the amino acid will change as a result of the variant. Sequence changes at the amino acid level may ultimately influence the structure, function, or binding properties of a protein, often leading to disease in humans. In viruses, non-synonymous variants can account for differentially conserved positions (DCPs) that establish differences between closely related species or strains.

One major example of this is the variation in the VP24 protein in filoviruses. As previously discussed, in most Ebolaviruses VP24 disrupts the signalling pathway STAT1 by binding to karyopherin-α. This does not occur in Reston virus, or potentially Bombali virus, as a result of differing amino acid residues present in the VP24-karyopherin binding region, which is approximately located between residues 130 and 140. As such, Reston virus does not induce an immune response similar to Ebola virus. Here, protein variation leads to a different outcome in pathogenicity, with closely related species having drastically different pathogenic effects in humans. In addition, it is known that the VP24 protein in Marburgviruses do not induce an immune response either.

1.7 Differentially Conserved Positions

Differentially conserved positions (DCPs), also known as specificity determining positions (SDPs) are specific amino acid positions in the proteome that are conserved within protein subfamilies but differ between them, potentially leading to altered

protein structure, function, binding, or a combination of all three (Pappalardo et al., 2016).

The original term SDP derives from the investigation into enzyme reactions, where an SDP was thought to play a role in determining substrate specificity, hence the name. With the expansion of the study of SDPs to other proteins a more generalised term – DCP – is used, to reflect that there is no impact on specificity in a protein that is not an enzyme.

The importance of a particular residue can be due to multiple factors, such as structural stability, protein-protein interaction, ligand binding and maintenance of overall protein function. Generally, it is hard to attribute a particular function to a specific residue or group of residues, as function is determined by interplay and interaction between many different residues, and mutation to any one of these may impact function or structure of the protein (Donald and Shakhnovich, 2005; Kalinina et al., 2009; Teppa et al., 2012). However there are cases where the relationship between a specific residue and protein function is well characterised. One such example is catalytic residues, where large datasets exist defining known amino acids related to a catalytic function, and where much of the initial research into SDPs was carried out in the context of enzyme active sites and known catalytic domains (Kalinina et al., 2009).

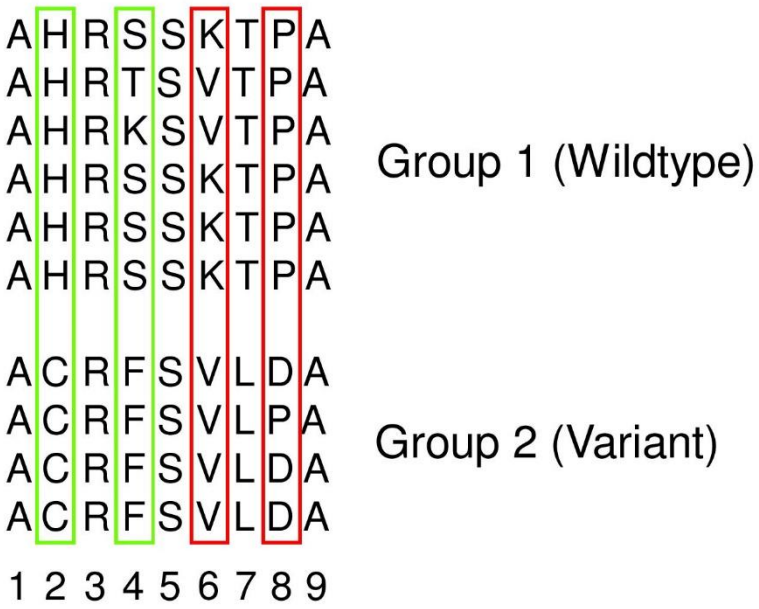


Figure 1.9. Example of DCP/SDP identification. Basic overview of what constitutes a conserved position. Here, positions two and four are considered a DCP

as they both contain highly conserved residues in their respective groups that also differ between groups. In this way, positions six and eight are not considered as they share common residues between groups.

The existence of a DCP does not automatically guarantee a change in protein structure or function. Many have no overall effect, and further analysis is required to determine their effect. In an initial study of Ebolavirus SDPs 189 were identified, with less than 10 predicted as having a major effect overall (Pappalardo et al., 2016). However, this can partly be attributed to lack of available structures preventing the analysis of all SDPs in terms of three dimensional folding. Furthermore, SDPs may arise as a result of coevolution, whereby one residue mutation causes a compensatory mutation in a nearby residue to retain the original function (Teppa et al., 2012).

1.7.1 Identification of DCPs

There are multiple ways to identify conserved positions (Kalinina, 2004; Kalinina et al., 2009), and while the numerous SDP identification algorithms differ in their processes and scoring functions, they all generally output an "intra-group conserved, inter-group different" amino acid composition pattern, often with an associated scoring system. Here we discuss the processes used in this research to determine a DCP/SDP. Two methods were used, the first utilising a program known as S3det (discussed in section 1.9.9) (Muth et al., 2012). This process was used to determine the SDPs discussed in chapter three. Subsequently, a new process based on Jensen-Shannon conservation scoring was utilised in chapter four (Briët and Harremoës, 2009; Majtey et al., 2005).

Jensen–Shannon (JS) divergence is a method used to measure the similarity between two probability distributions (Briët and Harremoës, 2009; Lamberti and Majtey, 2003; Majtey et al., 2005). The JS divergence between probability distributions P(x) and Q(x) is defined by

$$JD(P,Q) = H\left(\frac{P+Q}{2}\right) - \frac{1}{2}\left(H(P) + H(Q)\right)$$

where $S(P,Q)$ is the Kullback-Leibler divergence and $H_S(P) = -\Sigma_x P(x)\log P(x)$ is the Shannon entropy. The Kullback-Leibler divergence itself can be calculated by using the formula

$$S(P,Q) = \sum_x P(x)\, log\left\{\frac{P(x)}{Q(x)}\right\}$$

and is a measure of how a probability distribution differs from a second reference probability distribution. JS divergence is a useful calculation as it is everywhere defined, bounded, symmetric and only reaches zero when P = Q. JS divergence is the square of the transmission metric (dT). A common way to extend Jensen-Shannon divergence is to consider a mixture of the k probability distributions P1, . . . , Pk with weights π1, . . . , πk, respectively. With π = (π1, . . . , πk), we can define general Jensen divergence as

$$JD^{\pi}(P_1,\ldots,P_k) = H\left(\sum_{i=1}^{k}\pi_i P_i\right) - \sum_{i=1}^{k}\pi_i H(P_i)$$

JS divergence has been extensively studied in the context of analysis of symbolic sequences and most interestingly was successfully applied to research involving the segmentation of DNA sequences. It has also been involved in the determination of protein surface interfaces and machine learning (Ofran and Rost, 2003). The output of JS divergence scoring gives a value from 0 to 1. The greater the value, the more highly conserved the two input factors are. In terms of sequence analysis, each residue position in a given dataset of amino acid sequences will receive a unique score.

For DCP identification carried out in this research, JS divergence was used to calculate the conservation within the two separate subgroups. A threshold of 0.8 was set, with any residues scoring higher than this considered as a DCP. After this step, the two subgroups were compared to each other. Residue positions that score high enough in

both datasets are matched and extracted for further filtering. Any included position which shared any amino acids across the two groups was removed from consideration. For example, if all sequences in both datasets contained a valine at position 50, this would result in position 50 receiving a JS divergence score high enough for its inclusion in both subgroups. However, as they both contained the same amino acid it would be subsequently removed. If group one contained a valine at position 50 in all sequences and group two contained a methionine at position 50 in all sequences, again position 50 would appear for both datasets and be included in the initial result. This time, as there is no overlapping common residue between the two groups, it is included as a DCP. As such, V50M would appear in the output, V50V would not (figure 1.9).

1.8 Bioinformatics Tools and Resources

A wide array of bioinformatics tools were used to carry out the research detailed in this thesis. These tools carried out a variety of functions, ranging from genome analysis and open reading frame extraction to protein modelling and structural analysis. Using computational methods drastically increases the speed at which genomes and biological structures can be analysed. This section describes the most important of these resources, detailing their importance and relevance to the research conducted.

1.8.1 BLAST

A key tool utilised in bioinformatics is the Basic Local Alignment Search Tool (BLAST) (Johnson et al., 2008). This program is used for identifying homologous sequences to an input query, in either amino acid, DNA or RNA format. It utilises seeding to determine sequences in a database that share similarity with the query (Ye et al., 2006). BLAST is a powerful tool for inferring function to poorly understood proteins, and in determining the likely origin of a sequence. Unknown sequences can be identified by their closest matches, and similarities between homologues can be established. Different types of BLAST are available and can be tailored to the query input and database used, such as BLASTp (protein) and BLASTn (nucleotide) (Johnson et al., 2008; Ye et al., 2006).

Attempting to find similarities between a query sequence and those present in the given database relies on using a set of common letters known as words. For example, should a query sequence contain the following sequence, ABCDE, the word size in a standard BLAST search would be three letters. Here, the searched words would be ABC, BCD and CDE. The heuristic algorithm of BLAST will find all common three-letter words between the query sequence and the relevant sequences in the database. These results are then used to create an alignment. These words are compared against a scoring matrix, such as BLOSUM62 (Henikoff and Henikoff, 2000), and must reach a specified threshold score, T.

T determines whether a particular word is included in the alignment. Once seeding occurs, the three residue (letter) alignment extends in both directions and the process repeats. Each extension either increases or decreases the alignment score. If this score remains higher than T the alignment will be included in the results given by BLAST. If this score falls below the threshold, the alignment ceases to extend, preventing areas of poor alignment from being included in the final results output.

BLAST can either be used online or downloaded and run as a command line tool. The web serve is hosted by NCBI and is free to access (Johnson et al., 2008). Major common uses of BLAST include identifying shared domains, identifying the species a sequence belongs to, phylogenetic tree creation and DNA mapping (Pickett et al., 2012b). BLAST was a key part of all research present in this thesis, and was primarily used to match extracted ORFs to their known proteins.

1.8.2 Clustal

Clustal, in particular Clustal Omega, is used to generate multiple sequence alignments (MSAs) for nucleotide or amino acid sequences (Sievers et al., 2011a). MSAs are an essential component of most bioinformatics analyses that involve comparing homologous sequences, and are the starting point for a variety of further analysis. As such, producing an accurate MSA is of critical importance.

Clustal Omega uses five main steps to generate an MSA. Firstly, a pairwise alignment is generated using the k-tuple method, a heuristic method that will not guarantee an optimal alignment solution, but is more efficient than dynamic programming methods

(Sievers et al., 2011a). Secondly, sequences are clustered using the mBed method (Blackshields et al., 2010) which calculates pairwise distance via sequence embedding. After this k-means clustering is used, followed by the construction of a guide tree using the UPGMA method. At each of these steps, the nearest two clusters are combined, and the process repeated until a final tree can be assessed. Finally, the MSA is produced using HHAlign part of the HHSuite package, which uses two profile Hidden Markov Models (Sievers et al., 2011a). Clustal Omega consistently outperforms other MSA generating algorithms in processing time and overall quality. It is capable of running 100,000+ sequences on a single processor in just a few hours, and is readily accessible via a web server, making it a useful and powerful tool in bioinformatics (Daugelaite et al., 2013).

1.8.3 EMBOSS

EMBOSS is an open source software package that has been specially for molecular biology analysis and study (Rice P, Longden I, 2000). Extensive libraries are provided with the package, and acts as a platform allowing scientists to develop and release software. EMBOSS integrates a range of currently available packages and tools for sequence analysis. Within EMBOSS there are over one hundred programs that cover areas ranging from sequence alignment, to database search, to motif identification and domain analysis.

Of particular note for this research is the EMBOSS tool getorf. Getorf locates and outputs the amino acid sequences of open reading frames (ORFs) using a nucleotide sequence input file. An ORF can be defined as a nucleotide region of specified minimum size between two stop codons, or between a start and stop codon, depending on input parameters. Start and stop codons are defined using a genetic code table that can be changed depending on the organism being investigated.

1.8.4 mCSM

A structure based method that predicts the effect of mutations in proteins via the use of graph-based signatures (Pires et al., 2014). mCSM examines how individual amino

acid mutations can affect overall protein stability or protein-protein affinity. It utilises a machine learning approach to predict any impact the mutations may have on the overall stability of the protein (Pires et al., 2014). mCSM was primarily used in this research to determine the effects individual DCPs may have on the overall protein stability (detailed in chapters three and four).

1.8.5 Protein Databank

The Protein Databank (PDB) is an online resource used to deposit solved three dimensional structures of biological molecules, primarily proteins but also DNA and RNA. Structures, either partial or full length, are submitted by users and generally solved through the use of X-ray crystallography, Cryo-EM or NMR spectroscopy (Burley et al., 2019; Sugita et al., 2018; Yan et al., 2020). This information is freely accessible to users and often journals now require researchers to submit their structures to PDB as part of the publication process.

PDB was first established in 1971, and now contains over 140,000 solved structural models, many of which in complexes with other large molecules, ligands or small molecules (Burley et al., 2019). The database is updated weekly. Every structure deposited includes the atomic coordinates that define the 3D structure of protein or other large molecule. These positions are specified as Cartesian coordinates (x, y, z) using Angstrom units (0.1 nm) (Armstrong et al., 2020; Burley et al., 2019). Inclusion of experimental data is also required for all new submissions. PDB files are in the macromolecular Crystallographic Information file (mmCIF) format, and can be viewed via a number of visualisation software packages such as PyMOL. All structures available for analysis in chapters two – six were obtained from PDB.

1.8.6 Phyre2

The number of solved protein structures available on PDB (Burley et al., 2019) make up just a small fraction of the protein sequences available on UniProt (Bateman et al., 2017), reducing the usefulness of many of these sequences. The impact of amino acid mutations must be considered in relation to their position in the overall three

dimensional protein structure. The Protein Fold Homology/Analogy Recognition Engine (Phyre2) (Kelly et al., 2015) build these three dimensional structures of proteins without solved structure by identifying templates using hhsearch to search a fold library. Using this method can predict the secondary structure using Psi-pred, and the disordered regions using Diso-pred (Ward et al., 2004). With this information it constructs Hidden Markov Models of the protein sequence. Side chains are modelled too using Phyre2, with greater than 80% accuracy (Kelly et al., 2015).

1.8.7 PyMOL

PyMOL is a user-sponsored, open-source computer software programme. It operates as a molecular visualisation system that produces high-quality three dimensional images of small molecules and biological macromolecules, mostly proteins. PyMOL is used to generate the structure images displayed in this thesis (as well as many other scientific papers) and visualise the effect of amino acid mutation on protein structure.

1.8.8 RNAalifold

RNAalifold is part of the ViennaRNA package, and is used to compute consensus RNA structures from multiple sequence alignments (MSAs) of homologous sequences (Bernhart et al., 2008). It is based on the original RNAfold package, and is a variant of the classic dynamic programming RNA folding algorithm, but rather than folding a single sequence it takes an MSA and computes the consensus structure for all sequences in the alignment. The consensus structure is a common structure all sequences can fold into, but is not necessarily the best fold for each individual sequence, rather the best structure all of them have the ability to fold into. Energetically, RNAalifold scores not only the folding free energy, but also considers stabilization of the consensus structure by covariation terms. This is an additional stabilizing energy that results from covariation effects in different columns of the alignment. The more columns with good covariation support, the more negative the covariance score (i.e. the better the consensus structure energy). What makes RNAalifold particularly appealing is that it does not just compute the equivalent of minimum free energy structures, but also ensemble properties via partition function

folding (thus making available things like pairing probabilities and thermodynamic quantities). Internally, RNAalifold and RNAfold use the same code base for MFE and partition function calculations, which means it is easy to maintain and whenever there is a bugfix or a new feature for the classic folding functionality it is immediately available for RNAalifold. RNALalifold is analogous to RNAalifold, which predicts locally stable RNA structures (respectively locally stable consensus structures).

1.8.9 S3det

The proteins that make up a single family can have varied functions. For example, within a family of enzymes, they may all carry out effectively the same reaction but on different substrates. In the 1990s methods were developed to identify individual residue positions that could account for this variation, residues which are now referred to as Specificity Determining Positions (SDPs), or alternatively Differentially Conserved Positions (DCPs). SDPs are usually enriched at functional sites, such as the Karyopherin alpha binding region in the VP24 Ebolavirus protein (Pappalardo et al., 2016).

One way of predicting SDPs is the S3det. The algorithm based on a statistical method termed Multiple Correspondence Analysis (MCA), S3det encodes a MSA into a binary matrix, the coordinates of which are transformed into Principal Axes that are not correlated. The sequences are then projected onto these Principal Axes. The method is mostly used in a supervised setting, whereby the submitted protein sequences are split into user defined subfamilies. An unsupervised format can be used which operates via K-mean clustering to group the sequences into automatically generated subfamilies (Rausell et al., 2010). Residue variation between these subfamilies is then calculated (Muth et al., 2012). S3det was the primary method of SDP/DCP identification in the research described in chapter three.

1.9 Composition of this Thesis

The thesis focuses on three main avenues of research. The first studies herd immunity and vaccination coverage. The second is the analysis of SDPs/DCPs to understand

variation between closely related viral species. Thirdly, we consider the RNA level in Ebolaviruses. This thesis is divided into six chapters.

i. **Introduction.** An overall discussion of virology, with particular emphasis on Filoviruses and Coronaviruses. Additionally we consider protein variation and the tools and resources used to study it.

ii. **Chapter 2.** This chapter is composed of the article "*Herd Immunity to Ebolaviruses Is Not a Realistic Target for Current Vaccination Strategies*" published in Frontiers in Immunology (2018), of which I am first author. Here we discuss the requirements to achieve herd immunity to Ebola virus, taking into account factors such as biological requirements and social pressure and barriers.

iii. **Chapter 3.** This chapter is composed of the article *"Is the Bombali virus pathogenic in humans?"* published in Bioinformatics (2019), of which I am joint first author. Here we consider the SDPs that differentiate pathogenic and non-pathogenic Ebolavirus species, and use this data to determine whether the newly discovered Bombali virus is likely to be pathogenic or not.

iv. **Chapter 4.** This chapter is composed of the article *"Differentially conserved amino acid positions may reflect differences in SARS-CoV-2 and SARS-CoV behaviour"* published in Bioinformatics (2021), of which I am joint first author. In this chapter we consider a multitude of differences between the two SARS species in order to elucidate why there has been such a drastic difference in outbreaks between the two, and to consider potential drugs of interest for use in treatment.

v. **Chapter 5.** This chapter is composed of the article "Conserved RNA Structures in Ebolaviruses", of which I am joint first author. This work is in preparation and will shortly be submitted for publication. Here we identify novel conserved RNA structures in Ebolaviruses, and identify recurring motifs and differences between species.

vi. **Chapter 6.** Discussion.

Chapter 2: Herd Immunity to Ebolaviruses is Not a Realistic Target for Current Vaccination Strategies

Stuart G. Masterson, Leslie Lobel, Miles W. Carroll, Mark N. Wass and Martin Michaelis

Frontiers in Immunology, 2018, 9:1015, doi: 10.3389/fimmu.2018.01025

My contribution to this work was the collection of scientific papers studying herd immunity and vaccination programmes, the analysis of data to determine herd immunity threshold and critical vaccine coverage values, and the analysis of the requirements to achieve the coverage needed for mass vaccination programmes. I also contributed to writing the manuscript.

2.1 Abstract

The recent West African Ebola virus pandemic, which affected >28,000 individuals increased interest in anti-Ebolavirus vaccination programs. Here, we systematically analyzed the requirements for a prophylactic vaccination program based on the basic reproductive number (R_0, i.e., the number of secondary cases that result from an individual infection). Published R_0 values were determined by systematic literature research and ranged from 0.37 to 20. R_0s ≥ 4 realistically reflected the critical early outbreak phases and superspreading events. Based on the R_0, the herd immunity threshold (I_c) was calculated using the equation $I_c = 1 - (1/R_0)$. The critical vaccination coverage (V_c) needed to provide herd immunity was determined by including the vaccine effectiveness (E) using the equation $V_c = I_c/E$. At an R_0 of 4, the I_c is 75% and at an E of 90%, more than 80% of a population need to be vaccinated to establish herd immunity. Such vaccination rates are currently unrealistic because of resistance against vaccinations, financial/ logistical challenges, and a lack of vaccines that provide long-term protection against all human-pathogenic Ebolaviruses. Hence, outbreak management will for the foreseeable future depend on surveillance and case isolation. Clinical vaccine candidates are only available for Ebola viruses. Their use will need to be focused on health-care workers, potentially in combination with ring vaccination approaches.

Keywords: ebola virus, ebolavirus, vaccines, herd immunity, basic reproduction number

2.2 Introduction

The genus Ebolavirus contains five species: Zaire ebolavirus (type virus: Ebola virus), Sudan ebolavirus (type virus: Sudan virus), Bundibugyo ebolavirus (type virus: Bundibugyo virus), Taï Forest ebolavirus (type virus: Taï Forest virus, previously also referred to by names such as Côte d'Ivoire ebolavirus or Ivory Coast ebolavirus), Reston ebolavirus (type virus: Reston virus) (Kuhn et al., 2011). Four Ebolaviruses (Ebola virus, Sudan virus, Bundibugyo virus, Taï Forrest virus) are endemic to Africa

and can cause severe disease in humans (Michaelis et al., 2016). Reston viruses are endemic to Asia and considered to be non-pathogenic in humans (Michaelis et al., 2016). However, very few genetic changes may result in human-pathogenic Reston viruses (Michaelis et al., 2016; Pappalardo et al., 2016, 2017a). Since the discovery of the first two members of the Ebolavirus family in 1976 in Sudan (today South Sudan) and Zaïre (today Democratic Republic of Congo), Ebolaviruses had until 2013 only caused small outbreaks in humans affecting up to a few 100 individuals (Van Kerkhove et al., 2015; Keshwara et al., 2017). The recent Ebola virus outbreak in West Africa (2013–2016) resulted in 28,616 confirmed, probable, and suspected cases of Ebola virus disease and 11,310 deaths (Keshwara et al., 2017), which may still underestimate the actual numbers (Rojek et al., 2017). It was the first Ebolavirus outbreak that affected multiple countries, was introduced to another country via air travel, and resulted in a significant number of human disease cases outside of Africa (Van Kerkhove et al., 2015; Keshwara et al., 2017). Prior to this outbreak, only isolated human cases were treated outside of Africa. A scientist who had become infected by Taï Forest virus after an autopsy of a Chimpanzee was treated in Switzerland (Le Guenno et al., 1995), and two laboratory infections were reported in Russia (Akinfeyeva et al., 2005; Borisevich et al., 2006). In addition, Reston virus-infected non-human primates were exported from the Philippines to the US and Italy (Cantoni et al., 2016). Finally, Marburg virus (which belongs like the Ebolaviruses to the Filoviruses) was exported out of Africa (Rougeron et al., 2015; Timen, 2009) and was associated with laboratory infections (Beer et al., 1999; Nikiforov et al., 1994). Due to its unique size, the West African Ebolavirus outbreak emphasized the health threats posed by Ebolaviruses and the importance of protection strategies (Keshwara et al., 2017; Rojek et al., 2017). Vaccination programs are effective in controlling infectious diseases, as demonstrated by the WHO-driven smallpox eradication (Bonanni et al., 2014). However, eradication is likely to be more difficult for zoonotic viruses like the Ebolaviruses that circulate in animal reservoirs (Judson et al., 2016). Only herd immunity could prevent future outbreaks and protect individuals that cannot be vaccinated due to health issues (Bonanni et al., 2014). The herd immunity threshold (l_c) describes the number of society members that need to be protected (Anderson, 1992) to prevent outbreaks. It is based on the basic reproductive number R_0 (number of secondary cases caused per primary case) of a pathogen (Anderson, 1992; Fine et al., 2011; Gittings and Matson, 2016; Guerra et al., 2017; Plans-Rubió, 2012). Here, we

performed a systematic analysis to determine the critical vaccine coverage (V_c) required to prevent Ebolavirus outbreaks by a prophylactic mass vaccination program based on the R_0 associated with Ebolavirus infection in humans. The results were further critically considered in the context of (1) the status of current Ebolavirus vaccine candidates and (2) the feasibility of a large-scale prophylactic Ebolavirus vaccination program taking into account (a) the preparedness to participate in vaccination programs in the affected societies, (b) logistic challenges, and (c) costs.

2.3 Materials and Methods

2.3.1 Identification of Studies That Report on the Basic Reproductive Number (R_0) of Ebolaviruses

To identify scientific articles that have calculated the basic reproductive number (R_0) for Ebolaviruses, we performed a literature search using PubMed (www.ncbi.nlm.nih.gov/pubmed) for the search term combinations "Ebola R_0," "Ebola basic reproductive number," and "Ebola basic reproduction number" (retrieved on 29th September 2017).

2.3.2 Determination of Herd Immunity Thresholds and Their Implications for Ebolavirus Diseases Prevention Strategies

Based on the basic reproductive number R_0, i.e., the number of secondary cases that result from an individual infection, the herd immunity threshold (I_c) was calculated using Eq. 1; $I_c = 1 - (1/R_0)$, where I_c indicates the proportion of a society that needs to be protected from infection to achieve herd immunity. Next, the critical vaccination coverage (V_c) that is needed to provide herd immunity was determined by including the vaccine effectiveness (E) using Eq. 2 (18–22)(Anderson, 1992; Fine et al., 2011; Gittings and Matson, 2016; Guerra et al., 2017; Plans-Rubió, 2012); $V_c = I_c/E = [1-(1/R_0)]/E$.

2.4 Results

2.4.1 Basic Reproductive Number (R_0) Values for Ebolaviruses

The PubMed search for "Ebola R_0" provided 18 hits, the search for "Ebola basic reproductive number" provided 42 hits, and the search for "Ebola basic reproduction number" provided 35 hits (Figure 2.1; Data Sheet S1 in Supplementary Material). After removal of the overlaps and inclusion of an additional article [identified from the reference list of Ref. (Gittings and Matson, 2016)], this resulted in 51 articles, 35 of which provided relevant information on Ebolavirus R_0 values (Figure 2.1; Data Sheet S1 in Supplementary Material). R_0 data were only available for Ebola virus and Sudan virus outbreaks (Data Sheet S1 in Supplementary Material). 29/35 studies analyzed data from the recent West African Ebola virus outbreak (Data Sheet S1 in Supplementary Material). The others reported on Ebola virus outbreaks in the Democratic Republic of Congo. Four studies also included data from the Sudan virus outbreak 2000/2001 in Gulu, Uganda. We also considered a review that summarized all available data until February 2015 (Van Kerkhove et al., 2015) (Data Sheet S1 in Supplementary Material). R_0 indicates the number of new infections caused by an infected individual, and when greater than 1, an outbreak will spread. Different approaches to calculate R_0s lead to varying results (Guerra et al., 2017). Accordantly, R_0 values calculated for the Sudan virus outbreak 2000/2001 in Gulu using identical data ranged from 1.34 to 3.54 (Data Sheets S1 and S2 in Supplementary Material). Small outbreak sizes may also limit the accuracy of the calculated R_0 values. Additionally, virus transmission is influenced by socioeconomic and behavioral factors including the health-care response, society perceptions, religious practices, population density, and/or infrastructure (Guerra et al., 2017; Skrip et al., 2017). Concordantly, R0s that were determined by the same methodology in different districts of Guinea, Liberia, and Sierra Leone during the West African Ebola virus epidemic ranged from 0.36 to 3.37 (Krauer et al., 2016). Three studies directly compared the Ebola virus outbreak in Kikwit (1995, DR Congo) and the Sudan virus outbreak in Gulu (2000/2001, Uganda) (Chen et al., 2014; Chowell et al., 2004; Legrand et al., 2007), but did not reveal fundamental differences between the R_0s of the viruses (Data Sheets S1 and S2 in Supplementary Material). Across all relevant studies, R_0s ranged from 0.36 to 12 for Ebola virus and from 1.34 to 3.54 for Sudan virus (Data Sheet S1 in

Supplementary Material). 9 of the 35 studies that provided R_0 values showed that Ebola viruses can spread with an $R_0 > 3$, and five studies suggested that Ebolaviruses can spread with R_0 values > 4. High reproductive numbers (≥4) are typically observed at the beginning of Ebolavirus outbreaks, prior to the implementation of control measures (Althaus, 2015a; Althaus et al., 2015; Kucharski et al., 2016; Rosello et al., 2015). Also, the spread of Ebolaviruses may be substantially driven by "superspreaders" who infect a high number (up to 15–20) of individuals (Althaus, 2015b; Lau et al., 2017; Osterholm et al., 2015; Skrip et al., 2017; Volz and Pond, 2014). Studies from the West African Ebola virus outbreak suggested that relatively small numbers of superspreaders may have been responsible for the majority of cases (Agua-Agum et al., 2016; Lau et al., 2017). Since the available data suggest that Ebolavirus transmission can occur with R_0 values of 3, 4, or even higher, a prophylactic vaccination program should establish herd immunity against Ebolaviruses that spread at such levels.

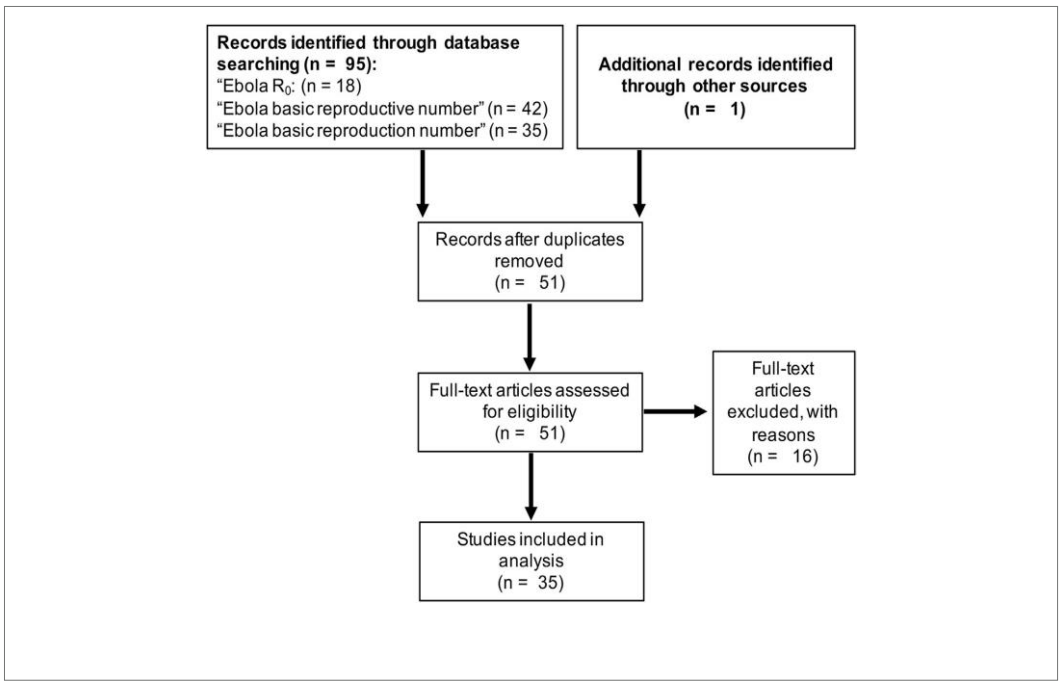


Figure 2.1. Summary of the literature search using PubMed (www.ncbi.nlm.nih.gov/pubmed) to identify articles that report on the basic reproductive number (R_0) of Ebolaviruses.

2.4.2 Herd Immunity Threshold (I_c)

At an R_0 of 3, the Ic (Eq. 1) is 67%, which means that 67% of a population need to be immune to provide herd immunity (Figure 2.2A; Data Sheet S3 in Supplementary Material). The I_c further rises to 75% at an R_0 of 4, to 80% at an R_0 of 5, to 90% at an R_0 of 10, and to 95% R_0 of 20 (Figure 2.2A; Data Sheet S3 in Supplementary Material). This shows that high proportions of a population need to be immune to establish effective herd immunity.

2.4.3 Critical Vaccine Coverage (V_c)

As there is currently no approved vaccine for the prevention of Ebolavirus disease, we calculated a range of V_c (Eq. 2) scenarios that reflect the efficacy range covered by approved vaccines. Attenuated replication-competent measles virus vaccines have been reported to protect up to 95% of individuals from disease after one dose, which increased to up to 99% after a second dose (Holzmann et al., 2016). The efficacy of varicella zoster virus vaccines, another attenuated replication-competent vaccine, was recently calculated to be 81.9% after one dose and 94.4% after two doses (Rieck et al., 2015). Inactivated seasonal influenza virus split vaccines have been reported to have a substantially lower efficiency of 50–60% (Beyer et al., 2011; Osterholm et al., 2012; Tricco et al., 2013). Hence, we considered a V_c range between 50 and 100% (Figure 2.2B; Data Sheet S3 in Supplementary Material). Vaccines, which provide high protection (ideally after a single vaccination), and high vaccination rates are required for prophylactic vaccination programs that establish a level of herd immunity that prevents Ebolavirus outbreaks. If we assume an R_0 of 3 and a vaccination efficacy E of 90%, more than 70% of a population need to be vaccinated to establish herd immunity. At an R_0 of 4 and a vaccination efficacy E of 90%, more than 80% of a population need to be vaccinated. If the R_0 rises to 5, a vaccine coverage of 80% would be required, even if a vaccine with 100% efficacy was available (Figure 2.2B; Data Sheet S3 in Supplementary Material).

2.5 Discussion

We performed an analysis of the Ebolavirus vaccine requirements to achieve the Vc needed for prophylactic mass vaccination programs. A number of studies suggested that Ebolavirus transmission can occur with R_0 values of 3, 4, or even higher, in particular during early outbreak stages (prior to the implementation of control measures) and/or as consequence of superspreading events (23, 24, 28–36)(Agua-Agum et al., 2016; Althaus, 2015b, 2015a; Althaus et al., 2015; Krauer et al., 2016; Kucharski et al., 2016; Lau et al., 2017; Osterholm et al., 2015; Rosello et al., 2015; Skrip et al., 2017; Volz and Pond, 2014). Therefore, a prophylactic vaccination program should establish herd immunity against Ebolaviruses that spread at such levels. At an R_0 of 3, >70% of individuals and at an R_0 of 4, >80% of individuals need to be vaccinated with a vaccination efficacy of 90% to achieve herd immunity. Hence, highly effective vaccines and a high vaccination coverage are essential for successful prophylactic mass vaccination programs against Ebolaviruses.

Clinical vaccine candidates providing protection against all three to four human-pathogenic Ebolaviruses (Ebola virus, Sudan virus, Bundibugyo virus, potentially Taï Forest virus) do not currently exist (Data Sheet S4 in Supplementary Material), although preclinical data suggest that the development of such vaccines may be feasible (6). Current vaccine candidates may also not provide the long-term protective immunity (≥10 years) necessary for sustainable protection against spillover events from animal reservoirs. Two studies reported immune responses 12 months after vaccination with different Ebola virus vaccine candidates (42, 43). One of them described seroconversion in >90% of individuals after a single injection of rVSV-ZEBOV, a vesicular stomatitis virus-based Ebola virus vaccine. No or only a minor drop in antibody titers and neutralization capacity was reported 360 days after vaccination (Heppner et al., 2017). A study investigating rVSV-ZEBOV and ChAd3-EBO-Z, a chimpanzee adenovirus type-3 vector-based Ebola virus vaccine, found lower seroconversion rates (rVSV-ZEBOV: 83.7%; ChAd3-EBO-Z: 70.8%) and reported the highest antibody response after 1 month and a decline afterward (Kennedy et al., 2017). Thus, it is not clear, whether the vaccine induced immunity covers the time frame of 2 years (or perhaps even longer) that Ebolavirus survivors may remain contagious for longer (Barnes et al., 2017; Deen, 2018; Diallo et al., 2016; Heppner et al., 2017; Kennedy et al., 2017; Keshwara et al., 2017; Li et al., 2017a;

Sissoko et al., 2017; Soka et al., 2016; Uyeki et al., 2016; Winslow et al., 2017; Zhu et al., 2017). It is also not clear whether (and if yes, to which extent) immunity to Ebolaviruses is mediated by cell-mediated and/or humoral immune responses (Lambe et al., 2017). A challenge study using non-human primates suggested that protection by adenovirus-based vaccines is cell mediated (Stanley et al., 2014). This means that antigen binding and/or neutralization titers may not always correlate with protection from disease. Consequently, the efficacy levels of vaccines cannot be determined with certainty based on antibody responses at various time points post vaccination. Thus, it remains unknown whether current vaccine candidates offer the long-term protection necessary for mass vaccination programs that effectively prevent zoonotic Ebolavirus outbreaks. Ebola virus recurrences and reinfections indicate that, although natural Ebolavirus infections are generally assumed to provide long-term protection, natural infections may not always result in sustained protective immunity in every survivor, which may further complicate the development of vaccines that provide long-term protection (Jacobs et al., 2016)(MacIntyre and Chughtai, 2016). In this context, the establishment of long-term immunity may be influenced by the disease treatment. In a case of relapse 9 months after discharge, it was speculated whether the treatment of the initial disease with convalescent plasma and monoclonal antibodies might have contributed to the recurrence (Jacobs et al., 2016).

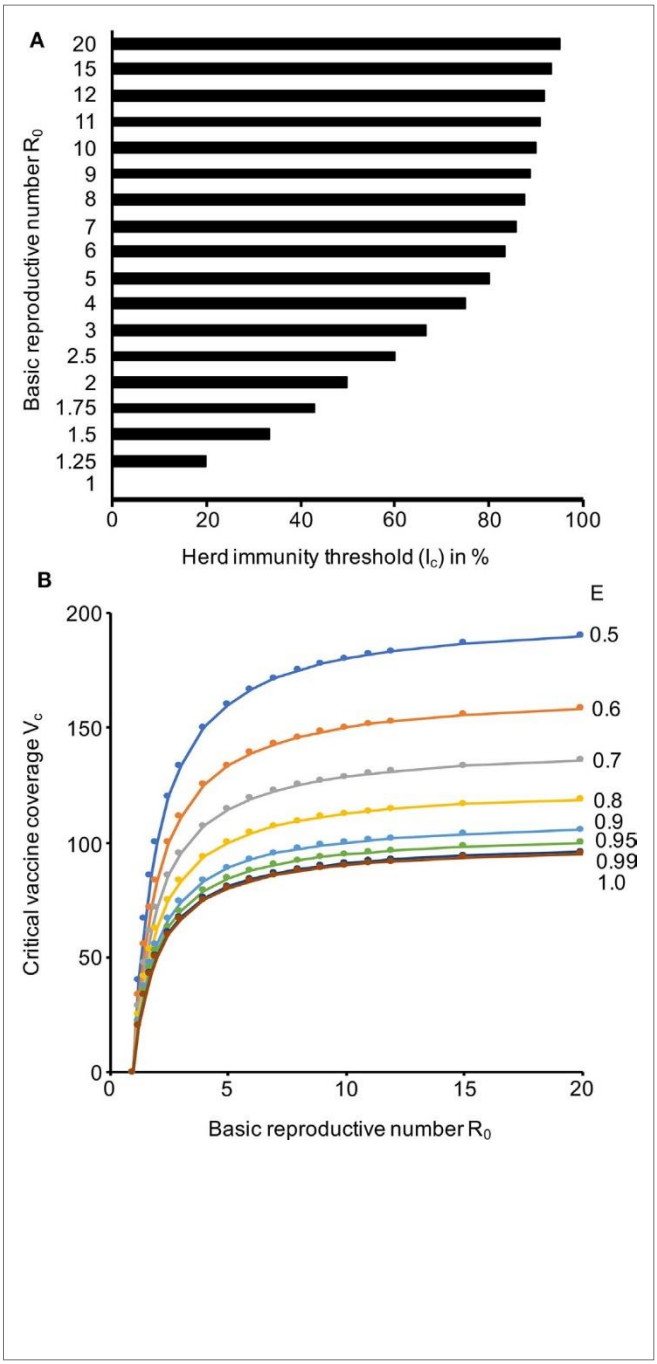


Figure 2.2. Herd immunity thresholds (I_c) and critical vaccine coverage (V_c) values in dependence of the basic reproductive number (R_0) and the vaccine efficacy (E). (a) I_c values based on a range of R_0 values that cover the range reported for Ebola viruses. (B) V_c values based on R_0 values that cover the range reported for Ebola viruses and E values that are in the range of those reported for approved vaccines. The respective numerical data are presented in Data Sheet S3 in Supplementary Material.

Limited acceptance of vaccinations may also limit Ebolavirus vaccination programs. In a rVSV-ZEBOV ring vaccination trial, only 5,837/11,841 patient contacts could be vaccinated. 34% of the contacts refused the vaccination (Henao-Restrepo et al., 2017). In a survey in Sierra Leone during the West African Ebola epidemic, 106/400 respondents (26.6%) were prepared to pay for a vaccination, while 290 respondents (72.5%) would have accepted a free vaccination (Huo et al., 2016). Since 74% of the population need to be vaccinated by a vaccine with a 90% efficacy to prevent an outbreak that spreads with an R_0 of 3 and 83% of the population to prevent an outbreak that spreads with an R_0 of 4 (Data Sheet S3 in Supplementary Material), such levels of vaccine coverage seem currently unachievable, even under the threat of an ongoing epidemic, although attitudes may change in the future if more (clinical) data becomes available. Therefore, more differentiated vaccination strategies with a focus on health-care workers and patient contacts appear more feasible. The median maximum fee that survey participants in Sierra Leone during the West African Ebola epidemic were prepared to pay for a vaccine was about 5,000 leones ($0.65 as of 11th January 2018) (Huo et al., 2016). The international organization GAVI (www.gavi.org) is providing $5 million for the development of rVSV-ZEBOV, which is expected to pay for 300,000 vaccine doses (about $16.70/dose) (Henao-Restrepo et al., 2016). Within a rVSV-ZEBOV ring vaccination trial, 11,841 contacts requiring vaccination from 117 clusters were identified over a 10-month period, i.e., about 101 individuals per confirmed Ebola virus disease patient (Henao-Restrepo et al., 2017). Hence, 300,000 doses will enable vaccination of the contacts of approximately 2,970 Ebola virus disease patients. If an effective vaccine (which provided protection against all human-pathogenic Ebolaviruses) was available, a vaccination program would comprise about 462 million individuals in the countries that have been affected by Ebolavirus outbreaks (Data Sheet S5 in Supplementary Material). Notably, the countries, which have been affected by Ebolavirus outbreaks so far, have large rural populations ranging from 13% (Gabon) to 84% (Uganda) (Data Sheet S5 in Supplementary Material). Vaccination programs in rural areas are associated with logistical issues including transport difficulties, lack of equipment and trained medical specialists, and cultural and language barriers (Alexander et al., 2015; Moon et al., 2015).

In conclusion, the achievement of a V_c of 75% that is necessary to prevent an outbreak that spreads with an R_0 of 4 with a vaccine that has an efficacy of 100% is currently unrealistic because of limited vaccine acceptance in the affected populations and

because of financial and logistical challenges. In addition, concurrent diseases such as HIV and cancer, along with potential side effects of vaccination, may remove significant numbers of potential vaccines (Kagina et al., 2014; Keshwara et al., 2017). Alternative vaccination strategies will be required for such patients. Replication-deficient vaccines such as DNA vaccines, virus-like particles, nanoparticle-based vaccines, and viral vectors (e.g., Modified Vaccinia Ankara, which was already demonstrated to be safe in immunocompromised individuals) may be safer alternatives (Keshwara et al., 2017; Volz and Sutter, 2017). Moreover, vaccines that provide long-term immunity against all three (or including Taï Forest virus, four) human-pathogenic Ebolaviruses, which would be needed to protect populations effectively from large Ebolavirus outbreaks in endemic areas, do not exist. Therefore, outbreak control of Ebolaviruses will for the foreseeable future depend on surveillance and the isolation of cases. Clinical vaccine candidates are only available for Ebola viruses and will need to be focused on health-care workers, who are often involved in disease transmission (Rosello et al., 2015), potentially in combination with the vaccination of patient contacts. Hence, our findings support the conclusions of the WHO Strategic Advisory Group of Experts on immunization (SAGE) at the WHO SAGE meeting on 25th to 27th April 2017 (World Health Organization, 2017). SAGE acknowledged the need for further research on Ebolavirus vaccines, including the generation of conclusive data on the duration of protection provided by Ebolavirus vaccine candidates. In case of future Ebolavirus outbreaks, SAGE recommended the use of rVSV-ZEBOV ring vaccination strategies (World Health Organization, 2017).

Chapter 3: Is the Bombali virus pathogenic in humans?

Martell H.J., Masterson S.G., McGreig J.E., Michaelis M., Wass M.N.

Bioinformatics, 2019, 1-6, doi: 10.1093/bioinformatics/btz267

My contribution to this work was the analysis of Ebolavirus genome sequences, generating the multiple sequence alignments, identifying specificity determining positions and the subsequent structural analysis to study potential effects on protein structure and function. I also contributed to writing the manuscript.

3.1 Abstract

3.1.1 Motivation

The potential of the Bombali virus, a novel Ebolavirus, to cause disease in humans remains unknown. We have previously identified potential determinants of Ebolavirus pathogenicity in humans by analysing the amino acid positions that are differentially conserved (specificity determining positions; SDPs) between human pathogenic Ebolaviruses and the non-pathogenic Reston virus. Here, we include the many Ebolavirus genome sequences that have since become available into our analysis and investigate the amino acid sequence of the Bombali virus proteins at the SDPs that discriminate between human pathogenic and non-human pathogenic Ebolaviruses.

3.1.2 Results

The use of 1408 Ebolavirus genomes (196 in the original analysis) resulted in a set of 166 SDPs (reduced from 180), 146 (88%) of which were retained from the original analysis. This indicates the robustness of our approach and refines the set of SDPs that distinguish human pathogenic Ebolaviruses from Reston virus. At SDPs, Bombali virus shared the majority of amino acids with the human pathogenic Ebolaviruses (63.25%). However, for two SDPs in VP24 (M136L, R139S) that have been proposed to be critical for the lack of Reston virus human pathogenicity because they alter the VP24-karyopherin interaction, the Bombali virus amino acids match those of Reston virus. Thus, Bombali virus may not be pathogenic in humans. Supporting this, no Bombali virus-associated disease outbreaks have been reported, although Bombali virus was isolated from fruit bats cohabitating in close contact with humans, and anti-Ebolavirus antibodies that may indicate contact with Bombali virus have been detected in humans.

3.2 Introduction

Ebolaviruses represent a serious public health concern. The past few years have seen multiple outbreaks in Africa, including an epidemic between 2013 and 2016, which resulted in more than 28 000 cases and 11 000 deaths (Coltart et al., 2017; Lo et al., 2017; Michaelis et al., 2016). Until recently, only five species of *Ebolavirus* had been identified. Four of these Ebolavirus species, Ebola virus, Sudan virus, Bundibugyo virus and Taï forest virus are known to be pathogenic to humans, while the fifth, Reston virus, is not (Baseler et al., 2017; Cantoni et al., 2016; Michaelis et al., 2016; Miranda and Miranda, 2011). In August 2018, a new species of *Ebolavirus*, *Bombali ebolavirus*, was identified in the Bombali region of Sierra Leone (Goldstein et al., 2018). Currently, it is not known if Bombali virus causes disease in humans.

To investigate why Reston virus is not pathogenic in humans and the other four Ebolaviruses are, we have previously identified amino acid positions that are differentially conserved between these two groups (specificity determining positions; SDPs; (Rausell et al., 2010)) and analysed their effects on protein structure and function together with the changes associated with Ebola virus adaptation to new species(Pappalardo et al., 2016, 2017a). The results indicated that certain SDPs in the karyopherin-binding region of the Ebolavirus protein VP24 are critical determinants of species-specific Ebolavirus pathogenicity (Pappalardo et al., 2016, 2017b). Here, we first update our comparison of human pathogenic and non-human pathogenic Ebolaviruses by including the many Ebolavirus genome sequences that have become available in the last few years. Then we use this dataset to analyse the Bombali virus sequence at amino acid positions that are associated with human pathogenicity.

3.3 Results

3.3.1 Identifying determinants of Ebolavirus pathogenicity

Our original study was based on a set of 196 Ebolavirus genomes. We identified 180 SDPs that were differentially conserved between Reston virus and the human pathogenic Ebolaviruses, of which 47 mapped to protein structures and eight were

proposed to have an effect on protein structure and function(Michaelis et al., 2016; Pappalardo et al., 2016). Here, we have expanded the dataset to 1408 Ebolavirus genomes (those retained after filtering an initial set of 2076 genomes for quality and completeness—see Supplementary Methods). This represents 7.5 times more sequences than used in the original study and also includes an increase in the number of Reston virus sequences from 17 to 27.

Phylogenetic analysis of the whole genome sequence and for each of the seven Ebolavirus proteins clearly separated each of the Ebolavirus species (Supplementary Fig. S1). However, the phylogenetic trees did not separate Reston virus from the human pathogenic Ebolavirus species (Supplementary Fig. S1).

High levels of conservation were observed within each species (Supplementary Fig. S2). Comparison of Reston virus proteins to the proteins of the other four human pathogenic species showed that there is greater divergence in GP, NP, VP30 and VP35, with conservation between 58 and 69%, whereas VP24, L and VP40 have a higher level of conservation (74–81%; Supplementary Fig. S2H).

The increased number of Ebolavirus genomes resulted in a slight reduction of SDPs from 180 (originally reported as 189 but SDPs in sGP and GP were identical as they share a common N-terminus) to 166 in the seven Ebolavirus proteins (Fig. 3.1, Table 3.1 and Supplementary Tables S1–S7). Overall, 146 SDPs were retained, 34 were lost and 20 new SDPs were identified. No SDPs were lost in VP24 or VP35, and only a single SDP was lost in VP30. New SDPs were identified for each of these proteins ranging from two for VP24 to seven for VP40 (Fig. 3.1 and Table 3.1). More SDPs were lost in NP, GP and L, ranging from five for NP to 17 for L. At the same time, no SDPs were gained in NP, one was gained in GP and three in L (Fig. 3.1 and Table 3.1).

Analysis of the SDPs at the codon level revealed that for the 27 Reston virus sequences, only ten SDPs showed any variation in codon usage, and for those ten positions there were always two codons present that represented synonymous changes. For five of these SDPs, only a single sequence contained a different codon and for the other five the codon usage was more closely balanced (Supplementary Table S8). For the pathogenic species, most amino acids at SDPs were encoded by multiple codons, with only 12 SDPs where a single codon was present (Supplementary Tables S9–S15). One Hundred and fifteen SDPs have only synonymous changes, while 39 SDPs also

have non-synonymous changes (35 of these 39 also have synonymous changes; Supplementary Tables S9–S15). The synonymous changes largely (106 of 115) represent differences in the codon usage between the different pathogenic species (Supplementary Tables S9–S15). Twenty three of the non-synonymous changes are due to different codon usage between the species, while the remaining 16 non-synonymous changes occur in Ebola viruses. This shows that while variation occurs at the codon level, the amino acids encoded at SDPs are highly conserved.

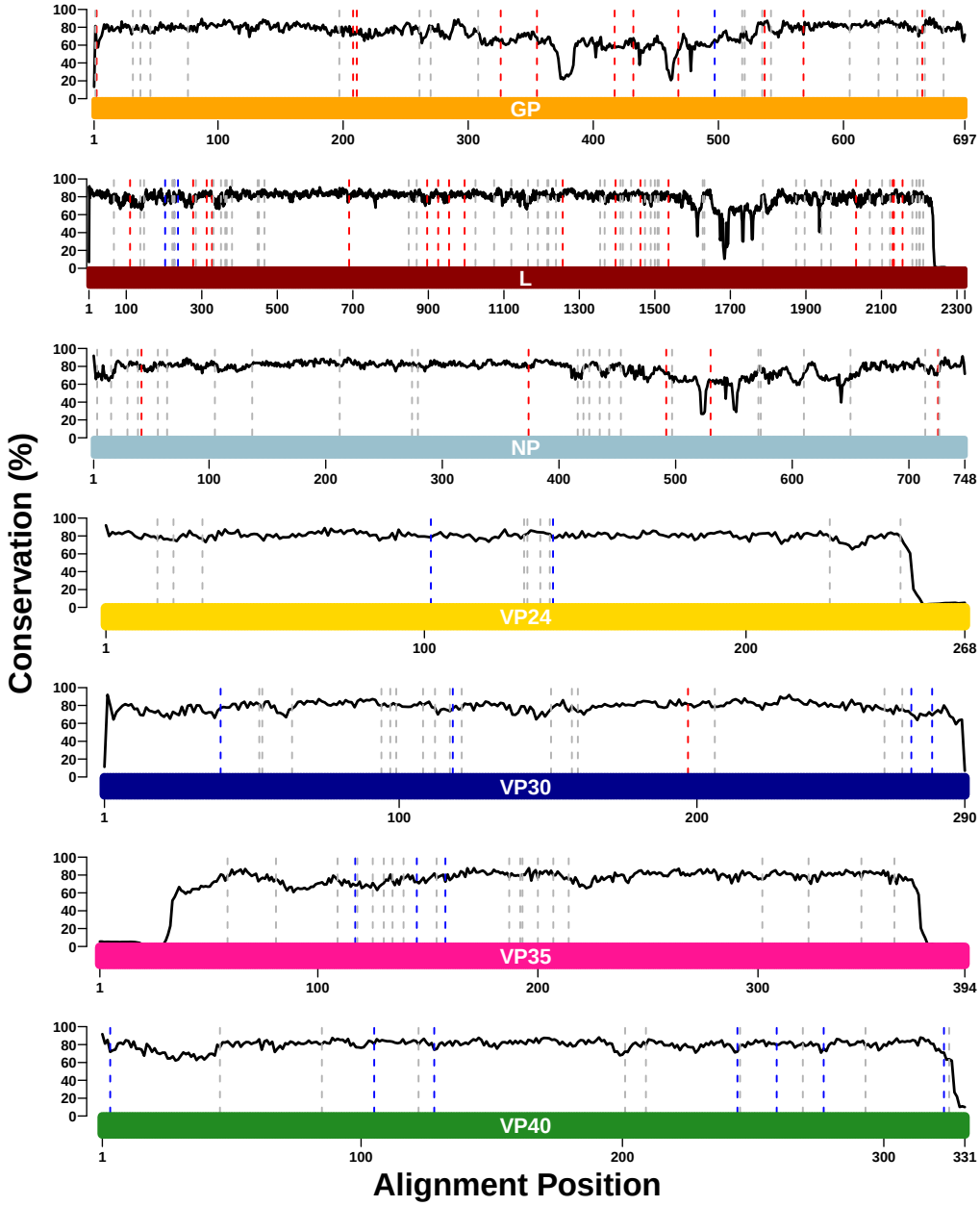


Figure 3.1. SDPs identified between human-pathogenic Ebolaviruses and Reston virus. The coloured bars represent the lengths of the protein sequence alignments, and each bar is labelled with the name of the protein that it represents. The solid black line represents the Jensen-Shannon conservation score. Dotted red lines represent SDPs. Previously identified SDPs that were lost in the updated analysis are shown by dotted lines (red), dashed-dot lines (grey) represent SDPs that were retained and dashed lines (blue) represent new SDPs that have been identified *Note*: *x*-axes differ in their scales between subplots.

Protein	SDPs in Original Set	SDPs Lost	SDPs Retained	SDPs Gained	SDPs in Updated Set
NP	29	5	24	0	24
VP35	19	0	19	3	22
VP40	9	0	9	7	16
GP	30	11	19	1	20
VP30	17	1	16	4	20
VP24	9	0	9	2	11
L	67	17	50	3	53

Table 3.1: Summary of the numbers of SDPs lost, retained, and gained in the updated SDP set.

3.3.2 Structural analysis of SDPs

It was possible to map 92 of the 166 SDPs onto protein structures or models (Supplementary Methods; Table 3.2; Supplementary Tables S17 and S18), compared to 47 SDPs in the previous study (Pappalardo et al., 2016). This was partly due to greater structural coverage of the proteins, with a structure of the N terminal region of VP35 (Chanthamontri et al., 2019; Zinzula et al., 2019) now available and also a template to model the structure of L (Supplementary Fig. S3). Overall, the amino acid changes at SDPs represent conservative changes, with the majority of BLOSUM62 substitution score values being one or greater (Fig. 3.2A). Most are predicted to be

slightly destabilizing to the protein structure (Fig. 3.2B), although this analysis only considered individual SDPs in isolation. One quarter of the SDPs (42) are located in the interior of the protein with the remaining three quarters having more than 20% relative solvent accessibility (Fig. 3.2C). These observations are consistent with the majority of SDPs having minor effects on protein structure and function.

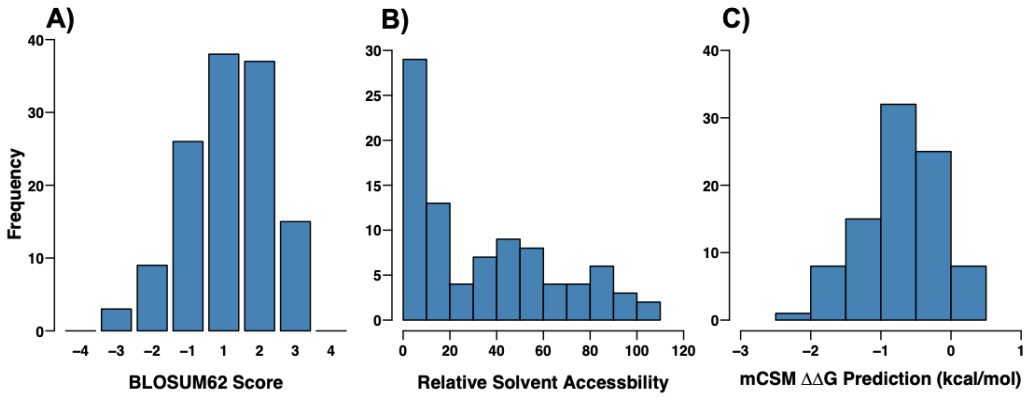


Figure 3.2. Characteristics of the SDPs between human pathogenic Ebolaviruses and Reston virus. (**A**) BLOSUM62 scores for the whole set of SDPs. (**B**) mCSM predicted stability changes for the whole set of SDPs. (**C**) Relative solvent accessibility for the whole set of SDPs

Protein	Length	SDPs	%Residues SDPs	SDPs Modelled	Probable Integrity	Probable Interface	Possible Integrity	Possible Interface
NP	739	24	3.25	10	0	0	1	0
VP35	340	22	3.47	4	0	1	0	0
VP40	326	16	4.91	13	1	1	0	0
GP	676	20	2.96	10	0	0	0	3
VP30	288	20	6.94	5	0	1	0	0
VP24	251	11	4.38	10	1	4	0	0
L	2,212	53	2.39	36	0	0	0	0

Table 3.2: Summary of SDPs per ebolavirus protein, and the predicted functional impacts.

Our previous structural analysis proposed a set of eight SDPs that were highly likely to alter protein structure and function, and a further five for which there was lower confidence (Pappalardo et al., 2016). Twelve of these 13 SDPs were retained in the current analysis, with only the lower confidence NP A705R no longer being sufficiently conserved to be identified as an SDP.

Of the 20 newly identified SDPs, ten were mapped onto protein structures (Table 3.2). Among these SDPs, we identified only one (VP24 R140S) that was likely to have an effect on protein structure and function. This results in nine SDPs overall with high confidence of having an effect on protein structure and function and four with lower confidence (Supplementary Table S19).

The VP24 SDP R140S is located in the VP24 interface site with human karyopherin α5 (KPNA5; (Xu et al., 2015)) where four other SDPs are located (T131S, N132T, M136L and Q139R). R140 can form hydrogen bonds with residues E476 (backbone) and Y477 (sidechain) in KPNA5, and also with the sidechain of E113 in VP24 (Fig. 3.3A and B). Reston virus S140 would still have the potential to form hydrogen bonds but not as extensively as R140. We have previously proposed that T131S, M136L and Q139R were likely to alter the binding of Reston virus VP24 to karyopherins, which may affect the ability of VP24 to inhibit the host interferon response (Pappalardo et al., 2016). The addition of R140S further supports this hypothesis, suggesting that this VP24 interface is vital to determining species-specific pathogenicity. Our hypothesis has recently been supported by experimental studies. Guito et al., (2017) showed that Reston virus VP24 is less effective at inhibiting the human interferon response. Further, histidine is present at residue 140 in Bundibugyo virus VP24 and has been implicated in reduced efficiency of downregulating interferon signalling (Schwarz et al., 2017).

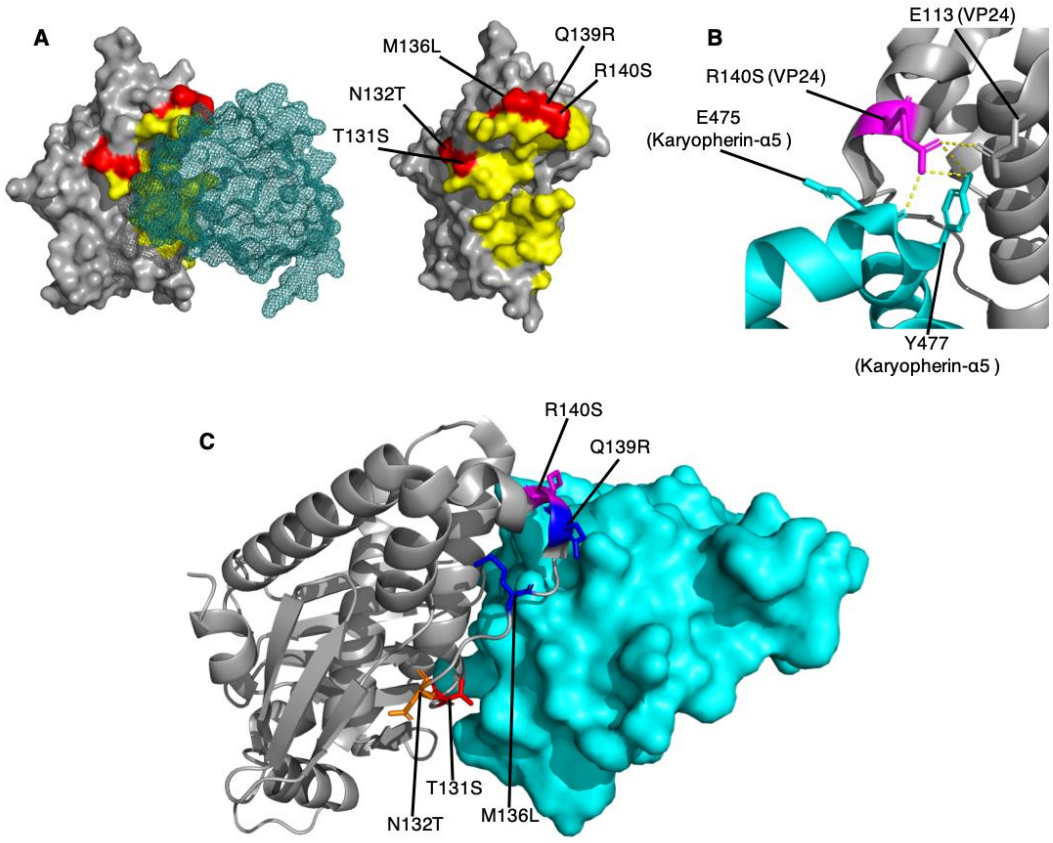


Figure 3.3. SDPs in VP24 suggest that Bombali virus may not be pathogenic in humans. (**A**) SDPs in the VP24-Karyopherin-α5 interface. VP24 is shown in surface representation (grey) and karyopherin-α5 is shown as a mesh representation (teal). SDPs in VP24 are shown in red, and all residues within 5 Å of karyopherin-α5 are shown in yellow. (**B**) Hydrogen bonding of the SDP residue R140 in the *Ebola virus*VP24. VP24 (grey) and karyopherin-α5 (teal) are shown in cartoon format. Hydrogen bonds are represented by yellow dashed lines. (**C**) Agreement of *Bombali virus* sequences with the SDPs in the VP24-Karyopherin-α5 interface. VP24 (grey) is shown in cartoon representation, and Karyopherin-α5 (teal) is shown in surface representation. SDPs are shown in stick format, and coloured red where Bombali and Ebola virus agree, blue where Bombali virus agrees with Reston virus, orange where the Bombali virus amino acid is unique, and magenta where the amino acid present in the two Bombali virus sequences differ and one agrees with Ebola virus and the other with Reston virus.

3.3.3 Comparison of Bombali virus with the other Ebolaviruses

Phylogenetic analysis of the genome sequences and of six of the seven Ebolavirus proteins grouped Bombali virus with Ebola, Sudan, Tai forest and Bundibugyo viruses, with Sudan and Reston viruses on a separate branch (Supplementary Fig. S1). For the seventh Ebolavirus protein, VP30, Bombali virus was grouped with Reston virus and the four known human-pathogenic species were on a separate branch (Supplementary Fig. S1k and l). While the phylogenetic analysis tends to group Bombali virus with human pathogenic Ebolavirus species, the pathogenic and non-pathogenic species are not clearly separated, making it difficult to infer from this phylogenetics analysis if Bombali virus is likely to be pathogenic in humans.

When considering the SDPs that differentiate human pathogenic and non-pathogenic Ebolaviruses, in Bombali virus the majority of amino acids at these positions (105; 63.25%) were identical to the human pathogenic Ebolaviruses, while 21 (12.65%) were shared between Bombali virus and Reston virus, and 40 (24.10%) were unique to Bombali virus (Supplementary Table S20). For the two available Bombali virus sequences, the amino acids present at SDPs agreed for all but one of the positions (VP24 R140S), where one of the sequences had the amino acid present in the human pathogenic viruses (R), while the other sequence contained the amino acid present in Reston virus (S).

With Bombali virus reported to have a 55–59% similarity to the other Ebolavirus species (Goldstein et al., 2018), the Ebola-Bombali SDP residue similarity is about 15% higher than the overall average, indicating high conservation amongst these positions, consistent with previous findings (Pappalardo *et al.*, 2016). For all of the individual proteins, the Bombali sequences have greater agreement with the amino acids present in human pathogenic species at SDPs [30.0% (VP30) to 77.36% (L)], while the agreement with Reston virus is only 10–19% (Supplementary Table S20).

This suggests that Bombali virus is more closely aligned with the human pathogenic Ebolaviruses than with Reston virus. However, this may reflect closer relatedness among the African Ebolaviruses (Ebola virus, Sudan virus, Bundibugyo virus, Taï Forest virus, Bombali virus) compared to the Asian Reston virus, than similarities in human pathogenicity. In the phylogenetic analysis *Bombali virus* does group with most of the human pathogenic species (Supplementary Fig. S1).

Of the nine SDPs where we are confident that they are likely to alter protein structure and function (see above), Bombali virus has the same amino acid as the human pathogenic species at five positions and the same as Reston virus for three. The ninth position differs among the two available Bombali virus sequences (VP24 R140S; Table 3.3). While the majority of amino acids in Bombali virus at SDPs in VP24 agree with human pathogenic Ebolavirus amino acids (73%; Supplementary Table S20), two critical SDPs in the VP24-karyopherin binding region (M136L, Q139R) are identical to Reston virus (Fig. 3.3) (Pappalardo et al., 2016, 2017a). Additionally, at residue 132, an SDP which points away from the KPNA5 interface, there is a Bombali virus-specific amino acid (A132; N in EBOV, T in RESTV; Fig. 3.3). This may indicate that the Bombali virus is not be as pathogenic as pathogenic compared to the other Ebolaviruses that are known to cause disease.

Protein	**SDP**	**Bombali agreement**
VP24	T131S	EBOV
VP24	M136L	RESTV
VP24	Q139R	RESTV
VP24	R140S	EBOV/RESTV
VP24	T226A	EBOV
VP30	R262A	EBOV
VP35	E269D	EBOV

Table 3.3. Comparison of Bombali virus sequences with the nine SDPs identified as having a likely functional impact on human pathogenicity

3.4 Discussion

In this study, we have updated our previous analysis of amino acid positions that are differentially conserved (SDPs) between human pathogenic Ebolaviruses and the non-human pathogenic Reston virus by the inclusion of more than 1200 additional genome sequences. We have also analysed the amino acids present in Bombali virus at the SDPs to infer whether Bombali virus may cause disease in humans.

Our updated analysis of the SDPs that distinguish Reston virus from the four known human pathogenic Ebolavirus species reduced the number of SDPs from 180 to 166. The vast majority of SDPs were retained from the original analysis, including all the SDPs that we have proposed are likely to affect protein structure and function and may have a role in determining pathogenicity. This demonstrates that our initial study using only 196 genomes provided robust results. While we have identified a small subset of SDPs that we propose may be associated with pathogenicity, this reflects those SDPs that we have been able to map to protein structure and use analysis of structures to identify a likely functional effect. It is of course possible that some of the SDPs that we have not been able to propose a functional effect for may have a role in determining pathogenicity. However, our updated results also further strengthen our findings that VP24 is central to determining host-specific pathogenicity (Pappalardo et al., 2016, 2017a), a notion that is further supported by experimental evidence showing that Reston virus VP24 is less effective than the other Ebolavirus VP24 proteins at inhibiting the host immune response (Guito et al., 2017). Since the number of available Reston virus sequences remains small, particularly compared to the number of sequences across the four human pathogenic species, a larger number of Reston virus sequences would likely further refine the set of SDPs by capturing the variation within Reston viruses.

Our analysis of the Bombali virus sequence at the SDPs identified overall greater agreement with the human pathogenic Ebolaviruses. This could be the consequence of the common African origin of the human pathogenic Ebolaviruses and the Bombali virus, in contrast to the Asian Reston virus. The amino acids at SDPs in VP24 that we propose are most important in determining human pathogenicity are the same in Bombali virus and Reston virus. This suggests that Bombali virus may not be pathogenic, or have reduced pathogenicity, in humans. This is supported by the fact that Bombali virus was isolated from fruit bats, which were cohabitating in houses and other populated areas (Goldstein et al., 2018) and although this makes human contact highly likely, no disease outbreaks have been reported. Further, a study in the Bombali region detected anti-Ebola virus NP antibodies in humans without reports of disease (Mafopa et al., 2017). Although originally interpreted as evidence for asymptomatic Ebola virus infection, it is possible that this test actually detected antibodies against the then unknown Bombali virus that cross-reacted with Ebola virus antigen. Hence,

antibodies directed against Ebolavirus proteins may indicate exposure of humans to low- or non-human pathogenic Bombali virus in the Bombali region.

In conclusion, based on our findings Bombali virus may be non-pathogenic or of low pathogenicity in humans. However, since few mutations seem to be sufficient for Ebolavirus adaptation to a new species (Pappalardo et al., 2017a), human pathogenic Bombali viruses may emerge, in particular as the Bombali virus shares many more conserved amino acid positions with human pathogenic Ebolaviruses than the non-human pathogenic Reston virus and further human contact with Bombali virus is likely to occur.

Chapter 4: Differentially conserved amino acid positions may reflect differences in SARS-CoV-2 and SARS-CoV behaviour

Denisa Bojkova, Jake E. McGreig, Katie-May McLaughlin, Stuart G. Masterson, Magdalena Antczak, Marek Widera, Verena Krähling, Sandra Ciesek, Mark N. Wass, Martin Michaelis, Jindrich Cinatl Jr.

Bioinformatics, 2021, 1-7, doi: 10.1093/bioinformatics/btab094

My contribution to this work was collecting and filtering genome sequences, identifying and analysing differentially conserved positions, and structural analysis of the SARS-CoV proteins. My additional contribution was the literature search and background information necessary to identify key residue positions. I also contributed to writing the manuscript. Jake McGreig worked on genome sequence collection, DCP identification and analysis, while Katie McLaughlin worked on protein structural analysis.

4.1 Abstract

Motivation: SARS-CoV-2 is a novel coronavirus currently causing a pandemic. Here, we performed a combined in-silico and cell culture comparison of SARS-CoV-2 and the closely related SARS-CoV.

Results: Many amino acid positions are differentially conserved between SARS-CoV-2 and SARS-CoV, which reflects the discrepancies in virus behaviour, i.e. more effective human-to-human transmission of SARS-CoV-2 and higher mortality associated with SARS-CoV. Variations in the S protein (mediates virus entry) were associated with differences in its interaction with ACE2 (cellular S receptor) and sensitivity to TMPRSS2 (enables virus entry via S cleavage) inhibition. Anti-ACE2 antibodies more strongly inhibited SARS-CoV than SARS-CoV-2 infection, probably due to a stronger SARS-CoV-2 S-ACE2 affinity relative to SARS-CoV S. Moreover, SARS-CoV-2 and SARS-CoV displayed differences in cell tropism. Cellular ACE2 and TMPRSS2 levels did not indicate susceptibility to SARS-CoV-2. In conclusion, we identified genomic variation between SARS-CoV-2 and SARS-CoV that may reflect the differences in their clinical and biological behaviour.

4.2 Introduction

In December 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel betacoronavirus, was identified that causes a respiratory disease and pneumonia called coronavirus disease 19 (COVID-19) (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020; Zhu et al., 2020). As of 22nd of December 2020, 77 801 721 confirmed COVID-19 cases and 1 713 109 COVID-19 deaths have been reported (Dong et al., 2020). Since 2002, SARS-CoV-2 is the third betacoronavirus, after severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), that has caused a substantial outbreak associated with significant mortality (Wu et al., 2020).

SARS-CoV-2 is closely related to SARS-CoV (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020; Wu et al., 2020). Entry of both viruses is mediated via interaction of the viral Spike (S) protein with the cellular receptor ACE2, and both viruses depend on S activation by cellular proteases, in particular by TMPRSS2 (Cui et al., 2019; Hoffmann et al., 2020a; Walls et al., 2020; Wan et al., 2020; Wrapp et al., 2020; Wu et al., 2020; Yan et al., 2020). Despite these similarities, the diseases caused by SARS-CoV-2 (COVID-19) and SARS-CoV (SARS) differ. According to WHO, the SARS-CoV outbreak resulted in 8098 confirmed and suspected cases and 774 deaths, equalling a mortality rate of 9.6% (www.who.int). Estimated mortality rates for SARS-CoV-2 are below 1% (Borges do Nascimento, 2020). SARS-CoV was only spread by symptomatic patients with severe disease (Cheng et al., 2013). In contrast, SARS-CoV-2 has been reported to be transmitted by individuals who are asymptomatic during the incubation period or who do not develop symptoms at all (Rivett et al., 2020).

We have developed an approach to identify sequence-associated phenotypic differences between related viruses based on the identification of differentially conserved amino acid sequence positions (DCPs) and in silico modelling of protein structures (Martell et al., 2019; Pappalardo et al., 2016). Conserved amino acid positions are likely to be of functional relevance, and differential conservation may indicate functional differences and they have been widely used for the analysis of protein families (Rausell et al., 2010, Das et al., 2015). Here, we used this method to

identify differentially conserved positions that may explain phenotypic differences between SARSCoV-2 and SARS-CoV. These data were combined with data derived from virus-infected cells.

4.3 Materials and Methods

4.3.1. Structural Analysis

Sequences for each of the SARS-CoV-2 proteins were obtained from the GISAID resource. The protein sequences were then filtered for sequences from human hosts with high coverage, and sequences with spans of X's were removed. The number of sequences retained after filtering for each protein is shown in Supplementary Table S4. Fifty-three SARS-CoV genome sequences derived from human hosts were downloaded from VIPR (Pickett et al., 2012a,b). Open Reading Frames (ORFs) were extracted using EMBOSS getorf (Rice et al., 2000) and matched to known proteins using BLAST. Fragments and mismatches were discarded. To match the ORF1ab non-structural proteins, a BLAST database of the sequences from the SARS non-structural proteins was generated and the SARS-CoV2 ORF1ab searched against it. The sequences for each protein were then aligned using ClustalO (Sievers et al., 2011) with default settings.

Conserved positions were identified by calculating the Jensen-Shannon divergence score (Capra & Singh, 2007) for each position in the multiple sequence alignment in virus. Differing alignment positions with conservation score >0.8 for both species were considered as differentially conserved positions (DCPs).

SARS-CoV-2 and SARS-CoV protein structures were downloaded from the Protein Databank (PDB; Supplementary Table S1) (Armstrong et al., 2020). Where structures were not available, they were modelled using Phyre2 (Kelley et al., 2015; Supplementary Table S2). Where Phyre2 did not generate a confident model, structural models from AlphaFold were used (Senior et al., 2020). Ligand binding sites were modelled using 3DLigandSite (Wass et al., 2010). DCPs were mapped onto protein structures using PyMOL. Exposed (solvent-accessible) and buried (solvent-inaccessible) residues were identified using Python module findSurfaceResidues with default parameters. Amino acid changes at DCPs were manually analysed for their

potential impact on protein structure and function based on the presence or absence of hydrogen bonding, changes in hydrogen bonding capacity and changes in charge in SARS-CoV compared with SARS-CoV-2 proteins. Where models were unavailable, mutagenesis was performed within PyMOL to assess the potential impact of the amino acid changes. The structural analysis grouped DCPs into six different categories based on the effect that they were proposed to have. These include 'unlikely', 'possible' and 'likely'. The possible and likely categories were split into three and two subgroups respectively depending on the type of effect (Supplementary Table S3).

4.3.2 Cell Cultures

The Caco2 cell line was obtained from DSMZ (Braunschweig, Germany). The cells were grown at 37C in minimal essential medium (MEM) supplemented with 10% foetal bovine serum (FBS), 100IU/ml penicillin, and 100lg/mL of streptomycin. 293 cells (PD02-01; Microbix Bisosystems Inc.) and 293/ACE2 cells (Kamitani et al., 2006) (kindly provided by Shinji Makino, UTMB, Galveston, Texas) were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% FBS, 50IU/mL penicillin and 50mg/mL streptomycin. Selection of 293/ACE2 cells constitutively expressing human angiotensin-converting enzyme 2 (ACE2) was performed by addition of 12mg/mL blasticidin. All culture reagents were purchased from Sigma (Munich, Germany). Cells were regularly authenticated by short tandem repeat (STR) analysis and tested for mycoplasma contamination.

4.3.3 Virus Infection

The isolate SARS-CoV-2/1/Human/2020/Frankfurt (Hoehl et al., 2020) was cultivated in Caco2 cells as previously described for SARS-CoV strain FFM-1 (Cinatl et al., 2004). Virus titres were determined as TCID50/ml in confluent cells in 96-well microtitre plates (Cinatl et al., 2003; 2005).

4.3.4 Western Blot

Western blotting was performed as previously described (Schneider et al. 2017). Briefly, cells were lysed using Triton-X-100 sample buffer, and proteins were separated by SDS-PAGE. Proteins were blotted on a nitrocellulose membrane (Thermo Scientific). Detection occurred by using specific antibodies against b-actin (1:2500 dilution, Sigma-Aldrich, Munich, Germany), ACE2 and TMPRSS2 (both 1:1000 dilution, abcam, Cambridge, UK) followed by incubation with IRDye-labeled secondary antibodies (LI-COR Biotechnology, IRDyeVR800CW Goat anti-Rabbit, 926-32211, 1:40 000) according to the manufacturer's instructions. Protein bands were visualized by laser-induced fluorescence using infrared scanner for protein quantification (Odyssey, Li-Cor Biosciences, Lincoln, NE, USA).

4.3.5 Receptor Blocking Experiments

SARS-CoV/SARS-CoV-2 receptor blocking experiments were adapted from Cinatl et al (2004). Caco2 cells were pre-treated for 30min at 37C with goat antibodies directed against the human ACE2 or DDP4 ectodomain (R&D Systems, Wiesbaden-Nordenstadt, Germany). Then, cells were washed three times with PBS and infected with SARS-CoV-2 at MOI 0.01. Cytopathogenic effects were monitored 48h post-infection. Cytopathogenic effect (CPE) was assessed visually by light microscopy by two independent laboratory technicians 48h after infection (Cinatl et al., 2003).

4.3.6 Antiviral Assay

Confluent cell cultures were infected with SARS-CoV-2 or SARS-CoV in 96-well plates at MOI 0.01 in the absence or presence of drug. Cytopathogenic effect (CPE) was assessed visually by light microscopy by two independent investigators 48h post-infection (Cinatl et al., 2003).

4.3.7 Viability Assay

Cell viability was determined by 3-(4,5-dimethylthiazol-2-yl)-2,5diphenyltetrazolium bromide (MTT) assay modified after Mosmann (Mosmann, 1983), as previously described (Onafuye et al., 2019).

4.3.8 qPCR

SARS-CoV-2 and SARS-CoV RNA was isolated from cell culture supernatants using AVL buffer and the QIAamp Viral RNA Kit (Qiagen) according to the manufacturer's instructions. RNA was subjected to OneStep qRT-PCR analysis using the SYBR green based Luna Universal One-Step RT-qPCR Kit (New England Biolabs) and a CFX96 Real-Time System, C1000 Touch Thermal Cycler. Primers were adapted from the WHO protocol (Corman et al., 2020) targeting the open reading frame for RNA-dependent RNA polymerase (RdRp) of both SARS-CoV-2 and SARS-CoV: RdRP_SARSr-F2 (GTGARATGGTCATGTGTGGCGG) and RdRP_SARSr-R1 (CARATGTTAAASACACTATTAGCATA) using 0.4lM per reaction. RNA copies/ml were determined by standard curves which were using plasmid DNA (pEX-A128-RdRP) harbouring the corresponding amplicon regions for SARS-CoV-2 RdRP target sequence (GenBank Accession number NC_045512). For each condition, three biological replicates were used. Mean and standard deviation were calculated for each group.

4.4 Results

4.4.1 Determination of differentially conserved positions (DCPs)

Coronavirus genomes harbour single-stranded positive sense RNA (+ssRNA) of about 30 kilobases in length, which contain six or more open reading frames (ORFs) (Cui et al., 2019; Wu et al., 2020). The SARS-CoV-2 genome has a size of approximately 29.8 kilobases and was annotated to encode 14 ORFs and 27 proteins (Wu et al., 2020). Two ORFs at the 5'-terminus (ORF1a, ORF1ab) encode the polyproteins pp1a and pp1b, which comprise 15 nonstructural proteins (nsps), the

nsps 1 to 10 and 12–16 (Wu et al., 2020). Additionally, SARS-CoV-2 encodes four structural proteins (S, E, M, N) and eight accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, orf14) (Wu et al., 2020). This set-up resembles that of SARS-CoV. The 8a protein in SARS-CoV is absent in SARS-CoV-2. 8b is longer in SARS-CoV-2 (121 amino acids) than in SARS-CoV (84 amino acids), while 3b is shorter in SARS-CoV-2 (22 amino acids) than in SARS-CoV (154 amino acids) (Wu et al., 2020).

To identify genomic differences between SARS-CoV-2 and SARS-CoV that may affect the structure and function of the encoded virus proteins, we identified differentially conserved amino acid positions (DCPs) (Rausell et al., 2010) and determined their potential impact by in silico modelling (Martell et al., 2019; Pappalardo et al., 2016).

In the reference sequences of the 22 SARS-CoV-2 virus proteins that could be compared with SARS-CoV, 1393 positions encoded different amino acids. 891 (64%, 9% of all SARS-CoV-2 genome residues) of these positions were DCPs (Supplementary Table S2). Most of the amino acid substitutions at DCPs appear to be fairly conservative as demonstrated by the average BLOSUM substitution score of 0.32 (median 0; Supplementary Fig. S1) and with 69% of them having a score of 0 or greater (the higher the score the more frequently such amino acid substitutions are observed naturally in evolution). 46% of DCPs represent conservative changes where amino acid properties are retained (e.g. change between two hydrophobic amino acids), 18% represented polar—hydrophobic substitutions, and <10% were changes between charged amino acids (Supplementary Table S3).

Six of the SARS-CoV-2 proteins have a higher proportion of DCPs, S, 3a, p6, nsp2, nsp3 (papain-like protease), and nsp4 with 14.82%, 11.68%, 9.52%, 21.38%, 17.9% and 10.8% of their residues being DCPs, respectively (Supplementary Table S4). Very few DCPs were observed in the envelope (E) protein and most of remaining non-structural proteins encoded by ORF1ab. For example, no residues in the helicase and <4% of residues in the RNA-directed RNA polymerase, 2'-O-Methyltransferase, nsp8 and nsp9 are DCPs (Supplementary Table S1).

We were able to map 572 DCPs onto protein structures (Supplementary Fig. S2, Supplementary Table S5 and S6). Nearly all of the mapped DCPs occur on the protein surface (86%), with only 34 DCPs buried within the protein, primarily in S and the papain like protease (nsp3) (Supplementary Table S3). We propose that 49 DCPs are likely to result in structural/functional differences between SARS-CoV and SARS-

CoV-2 proteins. A further 259 could result in some change. The remaining 264 DCPs seem unlikely to have a substantial functional impact (Supplementary Table S3).

4.4.2 Differentially conserved positions (DCPs) in interferon antagonists

At least 10 SARS-CoV proteins have roles in interferon antagonism (Totura and Baric, 2012). Two of these proteins, p6 and the papain like protease (nsp3), contain many DCPs, two have very few DCPs (nsp7 and nsp16), five have intermediate numbers of DCPs (nsp14, nsp1, nsp15, N and M), while p3b is not encoded by SARS-CoV-2. Initial studies have identified a difference in the interferon inhibition between SARS-CoV and SARS-CoV-2 (Lokugamage et al., 2020). Thus, it is possible that especially the DCPs in p6 and the papain like protease may have an effect on interferon inhibition.

4.4.3 Differences in cell tropism between SARS-CoV-2 and SARS

Next, we elucidated whether the substantial number of DCPs results in different phenotypes in cell culture, using the cell lines Caco2, CL14 (susceptible to SARS-CoV infection), HT-29 and DLD-1 (non-susceptible) (Cinatl et al., 2004). Analogously to SARS-CoV infection, SARS-CoV-2 replication was detected in Caco2 and CL14 cells, but not in HT-29 or DLD-1 cells, as shown by cytopathogenic effects (CPE) (Fig. 1A),staining for double-stranded RNA (Supplementary Fig. S3A) and viral genomic RNA levels (Supplementary Fig. S3B).

However, ACE2-expressing 293 cells differed in their susceptibility to SARS-CoV-2 and SARS-CoV (Fig. 1B, Supplementary Fig. S4). ACE2 has been identified as a cellular receptor for both SARSCoV-2 and SARS-CoV(Cui et al., 2019; Hoffmann et al., 2020a; Walls et al., 2020; Wan et al., 2020; Wrapp et al., 2020; Wu et al., 2020; Yan et al., 2020). Unmodified 293 cells are not susceptible to SARS-CoV infection due to a lack of ACE2 expression. However, 293 cells that stably express ACE2 (293/ACE2) support SARS-CoV infection (Kamitani et al., 2006). As expected, infection of 293 cells with SARS-CoV or SARS-CoV-2 did not result in detectable cytopathogenic effect (CPE) (Fig. 1B), but a SARS-CoV-induced CPE was detected in 293/ACE2 cells

(Fig. 1B). In contrast, 293/ACE2 cells displayed limited permissiveness to SARS-CoV-2 infection (Fig. 1B). Staining for double-stranded RNA (Supplementary Fig.S4A) and detection of viral genomic RNA copies (Supplementary Fig.S4B) confirmed these findings. Hence, the ACE2 status does not reliably predict cell sensitivity to SARS-CoV-2. Indeed, CL-14 was characterized by lower ACE2 levels than DLD-1 and HT29 (Fig. 1C).

SARS-CoV-2 and SARS-CoV cell entry depends on S cleavage by transmembrane serine protease 2 (TMPRSS2) (Hoffmann et al., 2020a,b; Zhou et al., 2015). However, the non-SARS-CoV-2 susceptible and susceptible cell lines displayed similar TMPRSS2 levels (Fig. 1C). Thus, cellular TMPRSS2 levels do also not reliable predict cell susceptibility to SARS-CoV-2.

4.4.4 Differences between SARS-CoV-2 and SARS-CoV S (Spike) protein cleavage sites and sensitivity to protease inhibitors

R667 and R797 are the critical cleavage sites in SARS-CoV S that are recognized by TMPRSS2 (Simmons et al., 2013; Zhou et al., 2015). These cleavage sites are conserved in SARS-CoV-2 (R685 and R815) (Fig. 1D). However, there is a four amino acid insertion in SARS-CoV-2 S prior to R685 and many of the residues close to R685 are DCPs (V663=Q677, S664=T678, T669=V687, Q671=S689, K672=Q690 DCPs are represented by the SARS-CoV residue followed by the SARS-CoV-2 residue) (Fig. 1D). The R815 cleavage site has two DCPs in close proximity (L792=S810, T795=S813) (Fig. 1D). Around the R685 cleavage site two DCPs retain polar side chains (S664=T678, Q671=S689), while the others represent larger changes between hydrophobic and polar side chains (V663=Q677, T669=V687) and one changes from a positive charge to a polar side chain (K672=Q690). While around the R815 cleavage site, one substitution is conservative (T795=S813) and the other is a hydrophobic to polar change (L792=S810).

These changes are likely to impact on TMPRSS2-mediated S cleavage. Indeed, SARS-CoV-2 was more sensitive than SARS-CoV to inhibition by the serine protease inhibitors camostat and nafamostat (Fig. 1E, Supplementary Fig. S6), which are known to inhibit TMPRSS2-mediated S cleavage and virus entry (Hoffmann et al., 2020a,b;

Zhou et al., 2015). This confirms that the observed differences in the amino acid sequence of S have functional consequences.

4.4.5 Differences between SARS-CoV-2 and SARS-CoV S interaction with ACE2

Our computational analysis detected further interesting changes in the S protein. SARS-CoV-2 S is 77.46% sequence identical to the SARS-CoV S and many of the remaining positions are DCPs (186 residues) (Supplementary Table S1).

The SARS-CoV S receptor binding domain (residues 306-527, equivalent to 328-550 in SARS-CoV-2) is enriched in DCPs, containing 43 DCPs (19% of residues). Nine of the 24 SARS-CoV S residues in direct contact with ACE2 were DCPs (Fig. 2A, Supplementary Table S4). Five of these DCPs represent conservative substitutions in amino acid (hydrophobic—hydrophobic or polar-polar), two hydrophobic -polar substitutions, one positive charge to polar change, while the ninth is substitution between a hydrophobic and positively charged amino acid (Supplementary Table S5).

Analysis of the DCPs using the SARS-CoV and SARS-CoV-2 S protein complexes with ACE2 (Song et al., 2018; Yan et al., 2020) identified runs of DCPs (A430-T433, F460-A471) in surface loops forming part of the S-ACE2 interface and resulted in different conformations in SARS-CoV-2 S compared to SARS-CoV S (Figure 2A, 2B). Two DCPs remove intramolecular hydrogen bonding within the spike protein in SARS-CoV-2 (Supplementary Table S4) and three DCPs (R426=N439, N479=Q, Q493, Y484=Q498) are residues that form hydrogen bonds with ACE2. For two of these positions, hydrogen bonding with ACE2 is present with both S proteins, but for R426=N439 hydrogen bonding with ACE2 is only observed with SARS-CoV S. N439 in SARS-CoV-2 S is not present in the interface and the sidechain points away from the interface. Further, analysis of the SARS-CoV-2 S-ACE2 complex highlighted important roles of the V404=K417 DCP, where K417 in SARSCoV-2 S is able to form a salt bridge with ACE2 D30 (Figure 2C, 2D) (Yan et al., 2020).

Alanine scanning (Chakraborti et al., 2005) and adaptation experiments (Wan et al., 2020) have identified 16 SARS-CoV S residues impacting on the binding affinity with ACE2. For all five residues identified from adaptation studies and four of the 11

identified by alanine scanning experiments, different amino acids are present in SARS-CoV-2 S (Fig. 2E), highlighting the difference in the interaction with ACE2.

In agreement with our structural analysis, we detected differences in the effects of an anti-ACE2 antibody on SARS-CoV-2 and SARS-CoV infection. Antibodies directed against ACE2 were previously shown to inhibit SARS-CoV replication (Li et al., 2003). In line with this, an anti-ACE2 antibody inhibited SARS-CoV infection in Caco2 cells (Fig. 2F). In contrast, the anti-ACE2 antibody displayed limited activity against SARS-CoV-2 infection (Fig. 2F). This shows that it is more difficult to antagonize SARS-CoV-2 infection with anti-ACE2 antibodies and supports previous findings indicating a stronger binding affinity of SARS-CoV-2 S to ACE2 compared to SARS-CoV S (Walls et al., 2020; Wrapp et al., 2020). As anticipated, antibodies directed against DPP4, the MERS-CoV receptor (Cui et al., 2019; de Wit et al., 2016), did not interfere with SARS-CoV or SARS-CoV-2 infection (Fig.2F).

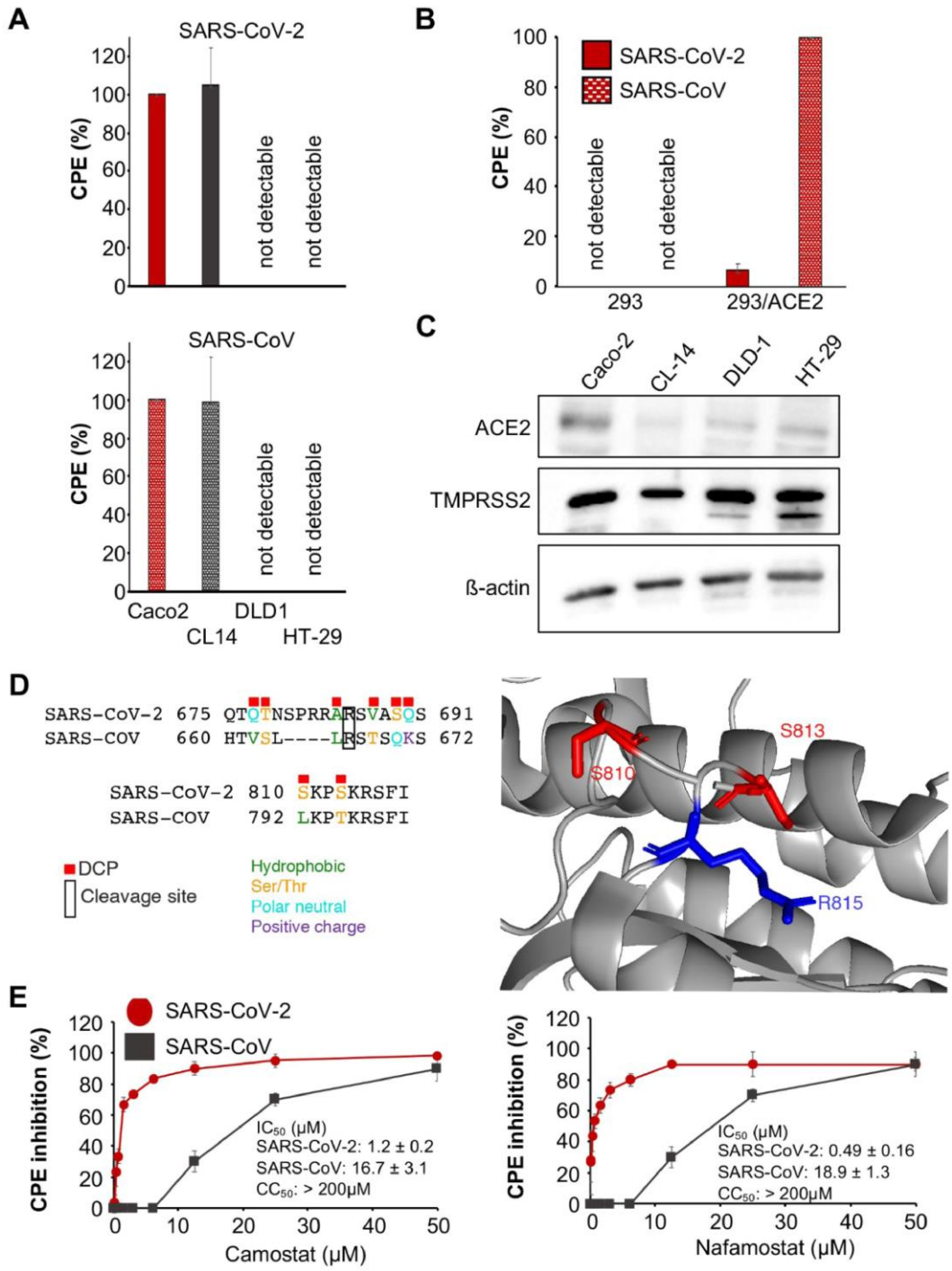
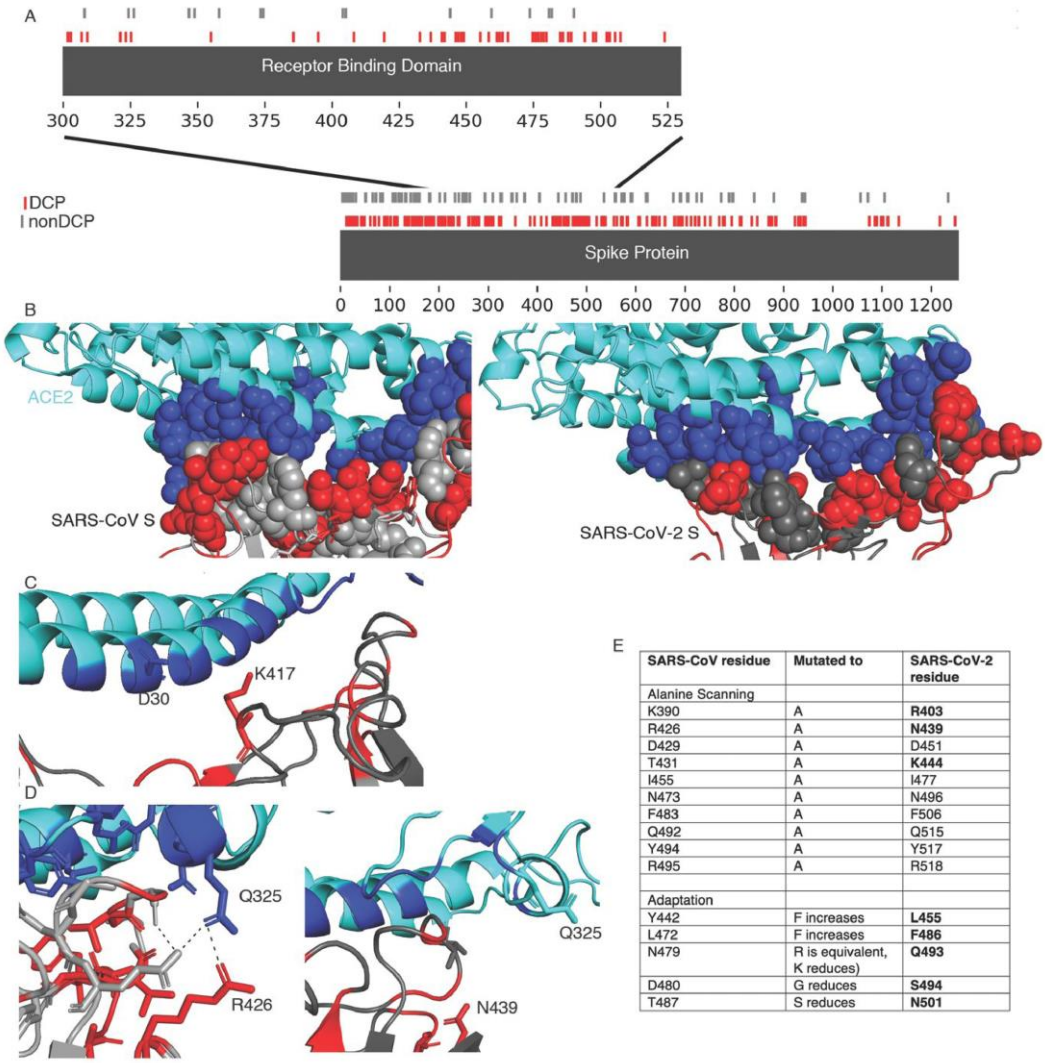


Figure 4.1. SARS-CoV-2 and SARS-CoV replication in cell culture. (A) Cytopathogenic effect (CPE) formation 48h post-infection in MOI 0.01-infected Caco2, CL14, DLD-1 and HT29 cells. Representative images showing immunostaining for double-stranded RNA (indicates virus replication) and quantification of virus genomes by qPCR are presented in Supplementary Figure S3. (B) CPE formation in SARS-CoV and SARS-CoV-2 (MOI 0.01)-infected ACE2-negative 293 cells and 293 cells stably expressing ACE2 cells (293/ ACE2) 48h post-

infection. Immunostaining for double-stranded RNA and quantification of virus genomes by qPCR is shown in Supplementary Figure S4. (C) Western blots indicating cellular ACE2 and TMPRSS2 protein levels in uninfected cells. Uncropped blots are provided in Supplementary Figure S5. (D) A sequence view of the DCPs in the vicinity of the S two cleavage sites and an image of the R815 cleavage site and closely located DCPs. S is cleaved and activated by TMPRSS2. (E) Concentration-dependent effects of the TMPRSS2 inhibitors camostat and nafamostat on SARS-CoV-2- and SARS-CoV-induced cytopathogenic effect (CPE) formation determined 48h post-infection in Caco2 infected at an MOI of 0.01 using a phase contrast microscope. Similar effects were observed in CL14 cells (Supplementary Fig.S6). Values are presented as means ±S.D. (n=3)



E

SARS-CoV residue	Mutated to	SARS-CoV-2 residue
Alanine Scanning		
K390	A	**R403**
R426	A	**N439**
D429	A	D451
T431	A	**K444**
I455	A	I477
N473	A	N496
F483	A	F506
Q492	A	Q515
Y494	A	Y517
R495	A	R518
Adaptation		
Y442	F increases	**L455**
L472	F increases	**F486**
N479	R is equivalent, K reduces)	**Q493**
D480	G reduces	**S494**
T487	S reduces	**N501**

Figure 4.2. SARS-CoV-2 and SARS-CoV S interaction with ACE2. (A–D)Differentially conserved positions in the Spike protein. (A) A sequence view of the DCPs present in the Spike protein, with an inset showing the receptor binding domain. (B) The S interface with ACE2 (cyan). The ACE2 interface is shown in blue spheres, DCPs in red. (C) The V404=K417 DCP. (D) The R426=N439 DCP, the left image shows SARS-CoV S R426, the image on the right show the equivalent N439 in SARS-CoV-2 S. (E) SARS-CoV residues associated with altering ACE2 affinity and the residues at these positions in SARS-CoV-2 S. (F) Cytopathogenic effect (CPE) formation in SARS-CoV-2 and SARS-CoV (MOI 0.01)-infected Caco2 cells in the presence of antibodies directed against ACE2 or DPP4 (MERS-CoV receptor) 48h post-infection.

4.5 Discussion

Here, we performed an in-silico analysis of the effects of differentially conserved amino acid positions (DCPs) between SARS-CoV-2 and SARS-CoV proteins on virus protein structure and function in combination with a comparison of wild-type SARS-CoV-2 and SARS-CoV in cell culture.

We identified 891 DCPs, which represents 64% of the amino acid positions that differ between SARS-CoV-2 and SARS-CoV and nearly 9% of all residues encoded by the SARS-CoV genome. 49 of these DCPs are likely to have a structural and functional impact. The DCPs are not equally distributed between the proteins. DCPs are enriched in S, 3a, p6, nsp2, papain-like protease and nsp4, but very few DCPs are present in the envelope (E) protein and most of the remaining non-structural proteins encoded by ORF1ab. This indicates that the individual proteins differ in their tolerance to sequence changes and/or their exposure to selection pressure exerted by the host environment.

The large proportion of DCPs reflects the differences in the clinical behaviour of SARS-CoV-2 and SARS-CoV. Mortality associated with SARS-CoV is higher than that associated with SARS-CoV-2 (Borges do Nascimento, 2020; Cui et al., 2019). SARS-CoV causes a disease of the lower respiratory tract. Infected individuals are only contagious when they experience symptoms (de Wit et al., 2016). SARS-CoV-2 is present in the upper respiratory tract and can be readily transmitted prior to the onset of symptoms. Mild but infectious cases may substantially contribute to its spread (Rivett et al., 2020).

The large proportion of DCPs reflects the differences in the clinical behaviour of SARS-CoV-2 and SARS-CoV. Mortality associated with SARS-CoV is higher than that associated with SARS-CoV-2 (Borges do Nascimento, 2020; Cui et al., 2019). SARS-CoV causes a disease of the lower respiratory tract. Infected individuals are only contagious when they experience symptoms (de Wit et al., 2016). SARS-CoV-2 is present in the upper respiratory tract and can be readily transmitted prior to the onset of symptoms. Mild but infectious cases may substantially contribute to its spread (Rivett et al., 2020).

Although further research will be required to elucidate in detail, which DCPs are responsible for which differences in virus behaviour, our analysis has already provided important clues. Both viruses use ACE2 as a receptor and are activated by the transmembrane serine protease TMPRSS2 (Cui et al., 2019; Hoffmann et al., 2020a; Li et al., 2003; Walls et al., 2020; Wan et al., 2020; Wrapp et al., 2020; Yan et al., 2020). Our results show, however, that the ACE2 and the TMPRSS2 status are not sufficient to predict cells susceptibility to SARS-CoV-2 or SARS-CoV. The cell line CL14 supported SARSCoV-2 replication, although it displayed lower ACE2 levels and similar TMPRSS2 levels to non-susceptible DLD-1 and HT29 cells. Thus, attempts to identify SARS-CoV-2 target cells based on the ACE2 status (Luan et al., 2020; Qiu et al., 2020; Xu et al., 2020) need to be considered with caution.

As previously described (Kamitani et al., 2006), ACE2 expression rendered SARS-CoV non-permissive 293 cells susceptible to SARS-CoV. However, ACE2 expression had a substantially lower impact on SARS- CoV-2 infection. This suggests the presence of further host cell factors that determine SARS-CoV-2 susceptibility. Based on our sequence analysis, DCPs in the viral interferon antagonists may contribute to the differences observed in the cellular tropism of SARSCoV-2 and SARS-CoV.

Our computational analysis detected DCPs in the ACE2-binding domain of S, which are likely to impact S-ACE2 binding. In agreement, an anti-ACE2 antibody displayed higher efficacy against SARS-CoV than against SARS-CoV-2, illustrating the differences between SARS-CoV-2 S and SARS-CoV S interaction with ACE2. This probably reflects an increased SARS-CoV-2 S affinity to ACE2 compared to SARS-CoV S (Wrapp et al., 2020), which may be more difficult to antagonize.

To mediate virus entry, S needs to be cleaved by host cell proteases, in particular by TMPRSS2 (Hoffmann et al., 2020a,b; Zhou et al., 2015). The S cleavage sites are conserved between SARS-CoV2 and SARS-CoV. However, we found DCPs in close vicinity to the S cleavage sites, which are likely to affect S cleavage by host cell enzymes and/or the activity of protease inhibitors on S cleavage. Indeed, the serine protease inhibitors camostat and nafamostat, which interfere with S cleavage (Hoffmann et al., 2020a,b), displayed increased activity against SARS-CoV-2 infection than against SARS-CoV infection, confirming the functional relevance of the DCPs.

In conclusion, our in-silico study revealed a substantial number of differentially conserved amino acid positions in the SARS-CoV-2 and SARS-CoV proteins. In agreement, cell culture experiments indicated differences in the cell tropism of these two viruses and showed that cellular ACE2 and TMPRSS2 levels do not reliably indicate cell susceptibility to SARS-CoV-2. Moreover, we identified DCPs in S that are associated with differences in the interaction with ACE2 and increased SARS-CoV-2 sensitivity to the protease inhibitors camostat and nafamostat relative to SARS-CoV.

4.6 Acknowledgements

The authors thank Shinji Makino, UTMB, Galveston, TX, for the provision of 293/ACE2 cells.

4.7 Author contributions

D.B., J.E.M, K.M., S.G.M., M.N.W., and J.C. performed experiments. V.K. provided essential materials. All authors analysed data. M.N.W., M.M., and J.C. planned, conducted, and supervised the study. M.M. wrote the first manuscript draft. All authors were involved in the drafting of and approved the final manuscript version.

4.8 Funding

This work was supported by Hilfefurkrebskranke Kinder Frankfurt e. V. the Frankfurter Stiftungfurkrebskranke Kinder. M.N.W., M.M. and M.A. were supported by a UKRI-BBSRC COVID-19 grant BB/V004174/1. JEM was supported by an EPSRC PhD studentship.

Conflict of Interest: none declared.

4.9 Data availability

All data are provided in the paper and supplementary material.

Chapter 5: Conserved RNA Structures in Ebolaviruses

Stuart G. Masterson, Ivo L. Hofacker, Martin Michaelis, Mark N. Wass and Michael T. Wolfinger

This work is in preparation and will be published shortly

My contribution to this work was the collection and analysis of the Ebolavirus genomes, creation of alignments and generation of predicted conserved RNA structures. I also contributed to the generation and analysis of MPI/SCI scores and Simplot data, and the protein level data. I joint wrote and edited the manuscript.

5.1 Abstract

Relatively little is known about functional RNAs in Ebolaviruses. We performed a comparative genomics screen of six Ebolavirus species with the aim of finding structured RNAs that are conserved in all species or conserved only in human pathogenic or non-pathogenic species respectively. Our data highlights that the overall amount of evolutionary RNA conservation in the investigated filoviruses is relatively low. We report here the existence of functional, conserved RNA structural elements that have previously been predicted solely for Ebola virus, both in in the untranslated and coding regions of all seven genes, that consistently appear in all six Ebolavirus species. We were further interested in differential RNA structure conservation among human pathogenic and hon-pathogenic Ebolaviruses. Here we identified eight conserved RNA structures that are uniquely conserved in the putatively non-pathogenic Reston virus, as well as several other structures that appear solely in the pathogenic Ebolavirus species.

5.2 Introduction

The genus *Ebolavirus* comprises six species, *Zaire ebolavirus* (EBOV), *Sudan ebolavirus* (SUDV), *Tai Forest ebolavirus* (TAFV), *Bundibugyo ebolavirus* (BDBV), *Bombali ebolavirus* (BOMV), and *Reston ebolavirus* (RESTV). Together with the phylogenetically related genus *Marburgvirus*, ebolaviruses form the family *Filoviridae* of the order *Mononegavirales*. Filoviruses are non-segmented negative-sense (NNS), single-stranded RNA viruses of approximately 19kb length, that contain seven genes encoding for nine proteins: NP (nucleoprotein), VP35 (associated with replication and transcription), VP40 (matrix protein), GP (glycoprotein), sGP and ssGP (soluble GP and small soluble GP are produced by alternative RNA editing of the gene GP), VP30 (transcriptional enhancer), VP24 (secondary matrix protein and interferon-response modulator) and L (RNA dependent RNA polymerase, RdRP) (Kuhn et al., 2011). The filovirus genome is flanked by untranslated leader and trailer sequences that have been associated with regulatory roles in virus transcription and replication (Feldmann et al., 1992; Volchkov et al., 1999). Each of the seven proteins is encoded by a monocistronic messenger RNA (mRNA), enclosed by untranslated regions (Sanchez et al., 1993). Early

Ebolavirus outbreaks tended to be small in numbers, with the highest number of cases (425) coming from the 2000 - 2001 Ugandan outbreak, and many outbreaks recording single digit cases, such as seven cases in Uganda in 2012 and the lone Tai Forest virus infection in 1994. However in recent years two larger outbreaks have occurred, with the West African between 2013-2016 resulting in greater than 11,000 deaths and 28,000 cases (Coltart et al., 2017; Lo et al., 2017; Michaelis et al., 2016). This outbreak demonstrated that Ebolaviruses can cause death and disease on a large scale and also saw small numbers of cases exported to many countries around the world. Of the six species, four are known to be highly pathogenic in humans with fatality rates up to 90% (Feldmann and Geisbert, 2011; Weingartl et al., 2013). In contrast, Reston virus is non-pathogenic in humans (Cantoni et al., 2016; Michaelis et al., 2016; Miranda and Miranda, 2011; Pappalardo et al., 2016). The pathogenicity of the newly discovered Bombali virus is currently unknown, however it shares key amino acid residues considered important for human pathogenicity with Reston virus, potentially indicating it is a non-pathogenic Ebolavirus (Goldstein et al., 2018; Martell et al., 2019).

Functional RNAs often depend on a specific fold that is evolutionarily conserved, typically at the level of secondary structures. Selective evolutionary pressures imply a structural homology on different classes of RNAs that covers all kingdoms of life, as manifested for example in characteristic structures of non-protein-coding RNAs (ncRNAs) such as tRNAs. Structural homology is typically achieved by compensatory substitutions, i.e., those that conserve the secondary structure by replacing one base pair (AU, GC or GU) by another that changes either one or both pairing partners (e.g. AU → GC or AU → GU). The Rfam (Kalvari et al., 2018) and miRBase (Kozomara and Griffiths-Jones, 2014) databases feature thousands of ncRNA families that contain well-conserved RNA secondary structures that are required for the biological function of the RNA molecule. While ncRNAs represent the dominating portion of transcriptomes (Hofacker, 2006), thermodynamic considerations suggested that structured RNAs not only exist in coding regions (Katz and Burge, 2003), but also exert crucial functions in eukaryotes (Olivier et al., 2005) and prokaryotes (Gu et al., 2014). In the viral world, structured RNAs can be found in coding and non-coding regions (Kiening et al., 2019). Examples are exoribonuclease-resistant structures in Flaviviruses (Ochsenreiter et al., 2019; Wastika et al., 2020) and members of the plant-infecting *Tombusuviridae* and *Luteoviridae* families (Steckelberg et al., 2018), lineage-specific duplicated structures in the 3'UTRs of Alphaviruses (De Bernardi Schneider

et al., 2019), the Rev-response element (RRE) in the Env coding region of HIV-1 CoV (Fernandes et al., 2012) and various conserved elements in Coronaviruses, including cis-acting RNAs in both UTRs (Madhugiri et al., 2014, 2018; Yang and Leibowitz, 2015) and a frame shift element in the non-structural-protein coding ORF1ab (Ian Brierley, Paul Digard, 2020).

Knowledge of functional RNAs in filoviruses is restricted to a set of cis-acting stem-loop structures in the genomic leader, trailer, gene start and intergenic regions. While short hairpins in the leader and trailer regions have been predicted in early studies (Crary et al., 2003; Sanchez et al., 1993; Volchkov et al., 1999), recent experimental evidence involving SHAPE RNA structure probing suggests the formation of a pan-handle structure involving the terminal genomic regions (Sztuba-Solinska et al., 2016). Likewise, a high degree of sequence conservation at the 3' ends of filovirus genes, some of which overlap the terminal region of the upstream gene, has been attributed to the presence of cis-acting gene start (GS) and gene end (GE) signal, respectively. These act as transcriptional regulators, as the RNA polymerase scans the genome sequence for GS and GE elements to start and stop transcription, respectively, resulting in a start-stop transcription mechanism that is also found in other NNS RNA viruses (Hume and Mühlberger, 2019). Upon transcription termination at a GE signal, the polymerase is capable of scanning the GS signal of the downstream gene in both directions of the GE signal (Brauburger et al., 2014). Importantly, the regions of GS/GE signals fold into RNA hairpin structures, as proposed in early predictions for Zaire ebolavirus and Marburg virus (Mühlberger et al., 1996; Sanchez and Rollin, 2005). These structures supposedly form on both the genomic RNA and mRNA level, however, due to inherent encapsidation of genomic and antigenomic RNA by NP, it has been suggested that stable secondary structures only form post-transcriptionally on the mRNA level after release of the nascent mRNA by the viral polymerase or transiently during viral RNA synthesis (Bach et al., 2020).

Motivated by the expanding repertoire of known filoviruses [38], and at the same time the limited knowledge of RNA structures and their involvement in molecular processes such as transcription and translation regulation, we report here a comprehensive *in silico* comparative genomics screen in Ebolaviruses. Employing a well-established set of tools built around the ViennaRNA Package (Shi et al., 2018) for thermodynamic modelling of RNA folding from sequence data we assessed distinctive traits, such as the amount of nucleotide sequence and RNA structure conservation

among phylogenetically diverged Ebolaviruses (Barrette et al., 2011). By screening six Ebolavirus species for evidence of RNA structuredness, we were interested to what extent RNA structure conservation can be observed among human-pathogenic and non-human pathogenic viruses. This is particularly intriguing in the context of Reston virus since alternative folds of homologous Ebolavirus genomic regions may explain the different phenotypes and tropism associated with non-human-pathogenic Ebolaviruses.

5.3 Results

As Ebolaviruses are negative-sense viruses, there is no consensus in the literature as to which sequences, i.e. genomic or antigenomic, to use for RNA structure prediction studies. Following the suggestion that stable, standalone RNA structures are rather formed on the mRNA level, we show computational results for the positive sense antigenome here. Assuming canonical Watson-Crick base pairings, i.e. AU and GC pairs, we expect similar results for the genomic sequences, resulting from sequence complementarity between genomic and antigenomic strands. It should be noted, however, that the possibility to form GU wobble pairs or other noncanonical interactions can result in the formation of alternative base pairing patterns with different energetics on the reverse strand.

5.3.1 Genomic proximity of the Ebolaviruses

Following up on earlier comparative genomics studies (Jun et al., 2015), we asked how related the putatively human non-pathogenic Reston and Bombali viruses are to the pathogenic viruses. To this end, we computed all pairwise whole genome nucleotide sequence alignments as well as a multiple sequence alignment comprising all six Ebolavirus species. Figure 5.1 shows the inferred maximum-likelihood phylogeny and the pairwise nucleotide identities, highlighting similarity scores ranging from 0.58 to 0.68 between individual species. In addition, our phylogenetic trees agree with previous data, with Reston virus and Sudan virus appearing the most closely related, while the newly discovered Bombali virus appears closer to Zaire ebolavirus (Martell et al., 2019).

Here we also see that whole genome nucleotide-level phylogeny is qualitatively similar to individual gene phylogeny previously described (Jun et al., 2015).

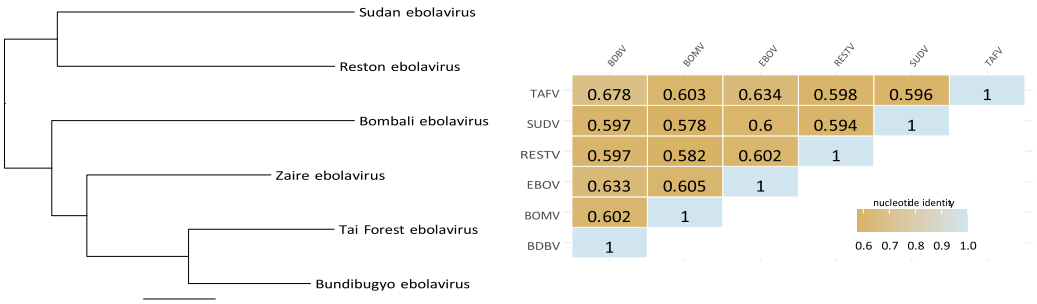


Figure 5.1: Midpoint-rooted maximum likelihood phylogeny and pairwise nucleotide identity values of the six Ebolavirus species.

5.3.2 Sequence and structure conservation varies within Ebola genes

In order to get an updated picture of the sequence and structural proximity of the six Ebolavirus species at a more fine-grained level, we computed structural nucleotide multiple sequence alignments for the six genes NP, VP35, VP40, GP, VP30 and VP24, comprising UTRs and coding sequences (CDSs). For the L gene we computed an alignment without structural information due to its size of approximately 7000nt. Prior to investigating RNA conservation we assessed specific traits of these alignments that are characteristic of their ability to show structure conservation. These are mean pairwise sequence identity (MPI) and the structure conservation index (SCI), two measures with high discrimination capability that are, beside thermodynamic stability, used by RNAz (Gruber et al., 2010; Washietl and Hofacker, 2004) to detect functional RNAs. Both measures are derived from slicing the whole gene alignments into chunks of overlapping alignment windows. While MPI is computed directly from the primary sequence data, SCI relates the minimum free energy (MFE) of the consensus structure in the current window, E_A, to the average MFE of the individual sequences. $\bar{E}$, in the alignment window as SCI = $E_A/\bar{E}$. High SCI values indicate that the consensus fold is energetically comparable to individually folded sequences, while low SCI indicates a lack of a consensus fold. Taken together, high MPI and SCI values can be interpreted as a proxy for RNA structural conservation, particularly but not necessarily at loci with

SCI near or higher than MPI. Intriguingly, the alignments of all seven genes show inconsistent MPI and SCI values at relatively low levels, discouraging the idea of wide-spread functional RNA conservation in the Ebolaviruses.

Previous data that described variation in the Zaire ebolavirus genome (Jun et al., 2015) matches closely to our detected regions of low MPI/SCI. In particular the downstream terminal regions of NP and VP40 have a near exact overlap, while the area of low MPI/SCI in GP between nucleotides 1200 and 1800 also concurs with a clustered region of high variability. Overall there are no major areas of considerable MPI/SCI variation we found that do not correlate to a higher variability in the genome (Jun et al., 2015). The prior data also demonstrates very high variability in the CDS regions, which parallels our finding of both highly decreased conservation and reduced MPI in these regions.

In order to complement our analysis of the Ebolavirus nucleotide sequences, we concurrently studied the amino acid sequences of the six species to determine any relationship between areas of interest at both the genomic and protein level. Analysis of all seven genes at the protein level show 11 domains/regions and 12 motifs as described by information available on UniProt. Two domains/regions and three motifs correlate directly with regions of raised SCI and MPI relative to the surrounding area - one in VP35, two in GP and two in VP30. In particular, the NP-binding motif present in VP35 is located in an area of high MPI and SCI compared to regions on either side, suggesting a high level of conservation relative to the surrounding area. Here we see that regions of particularly high conservation in nucleotides preserve key protein domains and motifs. In GP, the receptor binding domain that occurs between amino acids 54 and 201 (nucleotides 160 - 603) correlates to a raised SCI score compared to the rest of the sequence, while MPI is also higher. The same is also true for the short 16 residue fusion peptide that runs from residues 524 to 539. Many key domains however, such as the RdRp catalytic domain in the L protein and the core domain of the NP protein, appear to have no relation to nucleotide conservation or MPI/SCI. Many of the other domains and motifs are extremely short, and any extreme variation in conservation may be smoothed out by the sliding window approach applied.

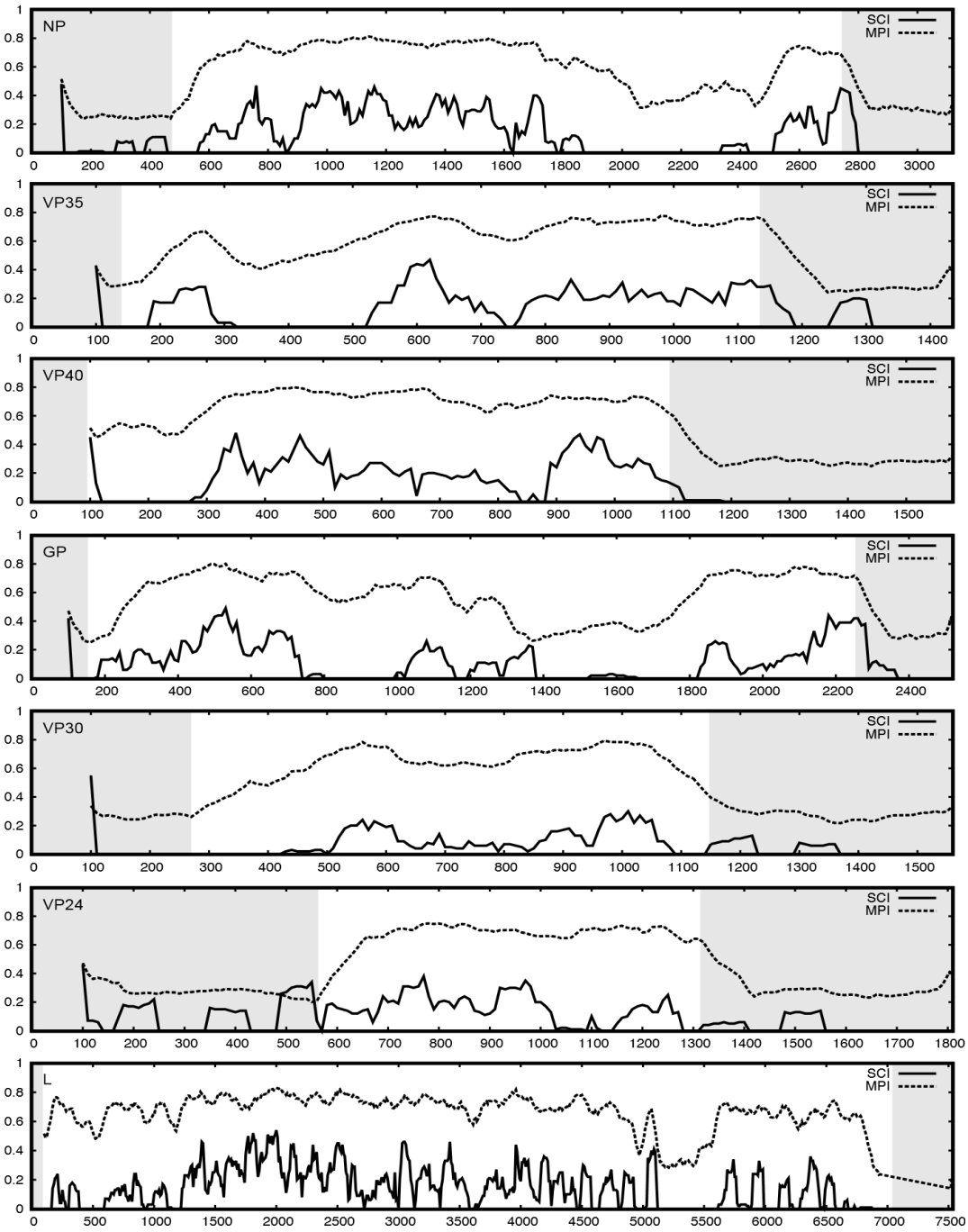


Figure 5.2: Nucleotide sequence and RNA structure conservation of each gene computed from structural alignments of all six Ebolaviruses at the mRNA level. MPI and SCI scores were evaluated in 100nt windows and 10nt step size, with each data point representing the values for the 100nt upstream window. Gray regions represent 5'UTR and 3'UTR, respectively. SCI scores are consistently low throughout all genes, while MPI differs across the gene body, with lower conservation in the UTRs.

5.3.3 RNA elements that are conserved in all Ebola species

The results of the nucleotide level sequence and structure conservation survey led us to speculate about moderate RNA structure conservation among the six Ebolavirus species. Not surprisingly, the number of RNAs that we could unambiguously characterize as being conserved in all Ebolaviruses is low. Nevertheless, we could identify well-conserved stem-loop structures in the 5'-terminal regions of all mRNAs in all Ebolavirus species (Table 5.1, Figure 5.3) that overlap the GS and GE signals on the negative strand. Table 5.1 lists the coordinates of these highly conserved RNA elements.

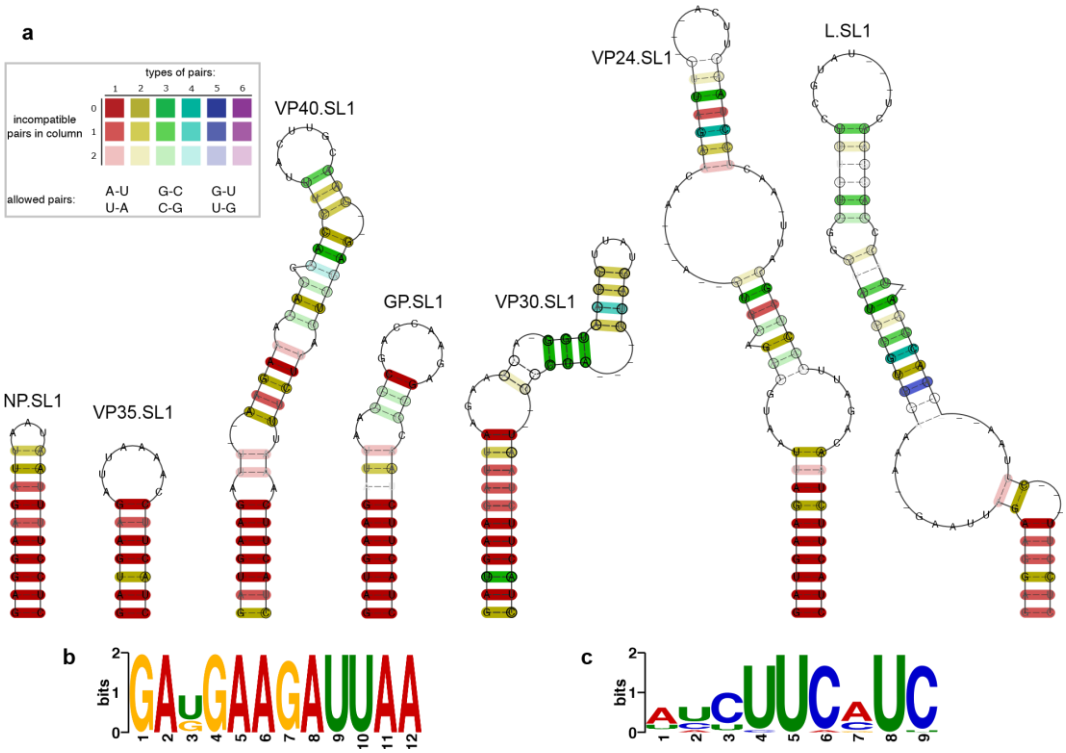


Figure 5.3: (a) Consensus structure predictions of structurally conserved RNA stem-loop structures SL1 at the beginning of the 5'UTRs of all ebolavirus gene mRNAs. Colouring of the conserved base pairs follows RNAalifold, indicating different covariation levels as depicted in the inserted colour scheme. Red highlights sequence-level nucleotide conservation, while other colours indicate an increasing number of structure-conserving nucleotide substitutions in the underlying alignments. Circled nucleotides in the consensus structures highlight consistent mutations, i.e. cases where one nucleotide is exchanged such as UA and UG, and compensatory mutations when both nucleotides of a base pair are mutated, e.g. UA and GC. Dashes along the backbone sequence of a consensus structure indicate gaps in the majority of sequences

in the underlying alignment. The basal portions of the SL1 structures are characterized by a high degree of primary sequence conservation, as depicted by red and ochre base pairs. Sequence logo plots of the corresponding 5' and 3'-terminal mRNA sequences are shown in (b) and (c), respectively.

ID	Zaire	Sudan	Tai forest	Bundibugyo	Reston	Bombali
NP.SL1	56-78	56-78	56-78	56-78	56-78	56-78
VP35.SL1	3032-3053	3013-3034	3026-3047	3020-3041	3019-3040	3026-3047
VP40.SL1	4390-4448	4365-4423	4378-4436	4372-4430	4396-4454	4377-4436
GP.SL1	5900-5936	5883-5919	5888-5924	5882-5918	5901-5937	5887-5923
VP30.SL1	8288-8336	8224-8273	8276-8324	8269-8317	8262-8315	8280-8329
VP24.SL1	9885-9959	9826-9893	9873-9953	9869-9950	9832-9901	9888-9961
L.SL1	11501-11570	11457-11528	11486-11555	11487-11556	11464-11542	11580-11658

Table 5.1: Conserved RNA stem-loop (SL1) structures in the terminal 5'UTRs of all genes in all six ebolavirus species. Each of the listed elements has a RNAz class probability higher than 0.99. Numbers in the columns of the individual Ebola species represent genomic coordinates. Consensus secondary structure plots are shown in Figure 5.3a.

The SL1 (following a functional naming scheme, numbering observed stem-loops starting from 1 at the 5' end of each mRNA, we denote these stem-loop structures SL1) elements, which have previously been proposed only for the Zaire Ebola virus species (Bach et al., 2020; Mühlberger et al., 1996; Sanchez et al., 1993), fold into well-defined stem-loop structures that are characteristic for each gene. SL1 consensus structures vary in length from 22nt to 83nt, the smallest being in the VP35 and NP 5'UTR and they form simple hairpins. The longest structures are present in the VP24 and L 5'UTR, which form bulged stem-loop elements and contain larger interior loops than the other SL1 elements. The SL1 elements are located at the extreme 5' termini of all mRNAs, some of which overlap the 3' end of the preceding gene. These regions encode transcription/gene start and stop signals on the genomic strand, and show the highest level of primary sequence conservation observed in our screen of all Ebolavirus genomes. This high degree of sequence conservation is manifested in a marked gradient of covariation patterns along the SL1 consensus structures, particularly in almost perfect nucleotide sequence conservation at the basal portion of

the SL1 closing stems. Figures 5.3a highlights this by red and ochre colours, respectively. Importantly, a motif search in the nucleotide sequences of all predicted SL1 structures revealed two highly significant sequence motifs, i.e. GAUGAAGAUUAA in the upstream (5'-terminal) part and AUCUUCAUC in the downstream (3'-terminal) part of the SL1 closing stems (Fig 5.3b and c). On the contrary, more structure conserving covariation can be observed in the central parts of several SL1 elements, particularly VP40.SL1, VP30.SL1, VP24.SL1 and L.SL1. All SL1s have RNAz scores of greater than 0.999, indicating well conserved and functional structures. For the NP gene, the region that is overlapped by SL1 has been described as a spacer between two promotor elements (Bach et al., 2020; Weik et al., 2005).

Intriguingly, we could only identify two additional RNA stem loop structures that are unambiguously conserved in all Ebola species. Both are located in the polymerase gene (L), therefore denoted L.SL2 and L.SL3 here. Figure 5.4 shows the predicted consensus structures and the underlying structural alignments. L.SL2 forms a short stem-loop structure with a single bulging Uracil that is only present in Reston virus and Bombali virus, while L.SL3 is a slightly longer stem-loop element with a central interior loop. The top hairpin loop of L.SL3 is longer in Reston virus and Bombali virus, which results in several consecutive dash symbols in the consensus structure due to majority voting criteria implied by RNAalifold. Covariations are observed at each stacked base pair in both elements, two of which are classified as significant by R-scape in L.SL3. L.SL2 has a higher mean pairwise sequence identity (49.86%) than L.SL3 (38.13%). Convincing RNAz class probabilities of 0.999 and 0.979, respectively, for L.SL2 and L.SL3, as well as low z-scores (-10.3 and -9.0, respectively) are indicative of structural conservation, although these data do not allow us to infer any biological function of these elements.

5.3.4 RNA elements that are conserved only in some Ebola species

5.3.4.1 Elements specific to Reston virus

Complementary to characterizing RNA structures that are common to all Ebolavirus species, we were particularly interested in the question whether or not there are any RNAs that are unique to the non-pathogenic Reston virus. To this end, we performed

local RNA secondary structure prediction in the Reston genome, limited to sequence lengths of 150nt. We filtered for thermodynamically stable structures, requiring a free energy z-score of at least -2 when comparing to 1000 dinucleotide shuffled sequences of the same nucleotide composition. After pruning all hits that were previously found to be conserved among all Ebolavirus species, we obtained eight locally stable RNA (LSR) structures that are exclusively formed in Reston virus (Table 5.2 and Figure 5.5).

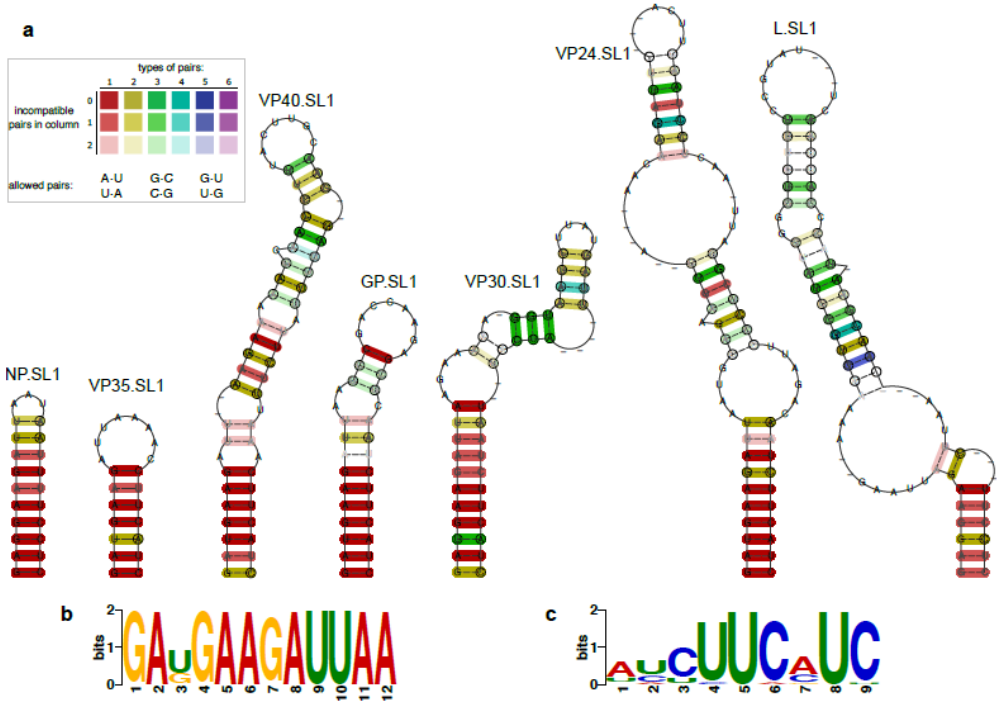


Figure 5.4: Evolutionarily conserved hairpin structures L.SL2 (a) and L.SL3 (b) in the coding sequence of the polymerase gene. Structural alignments of homologous regions in all Ebola species show varied covariation patterns, as encoded in RNAalifold standard colours (see Figure 5.3). Consensus structures are plotted in dot-bracket notation above the alignments, grey bars below the alignments indicate sequence conservation levels for each column of the alignment. Positive-strand genomic coordinates of the elements in each Ebola virus species are provided with each sequence identifier.

ID	Gene	Location	Coordinates	MFE	z-score
LSR1	NP	CDS	464-521	-14.80	-2.17
LSR2	NP	CDS	1938-2020	-16.30	-2.02
LSR3	VP35	3'UTR	4283-4336	-12.20	-3.12
LSR4	VP40	5'UTR	5519-5645	-24.10	-2.01
LSR5	VP40	5'UTR	5709-5772	-18.10	-2.86
LSR6	GP	CDS	7465-7528	-20.30	-3.46
LSR7	VP30	3'UTR	8323-8381	-17.10	-2.43
LSR8	VP24	5'UTR	11168-11219	-10.00	-2.16

Table 5.2: Locally stable RNAs (LSRs) that are unique to Reston virus. Coordinates are global to the Reston virus reference genome NC 004161.1. MFE values are given in kcal/mol, z-scores were computed by RNALfold. Secondary structure plots for all LSRs are shown in Figure 5.5.

The eight Reston-specific LSR elements are found in the mRNAs of NP, VP35, VP40, GP, VP30 and VP34, overlapping both UTR and CDS regions. Two elements, LSR3 in the 3'UTR of VP35 and LSR6 in the CDS of GP, have a z-score of lower than -3, indicating particular thermodynamic stability compared to randomized sequences. Nevertheless, all LSRs proposed here are based on single sequence predictions and therefore subject to some uncertainty due to the lack of evolutionary-supported covariation information. Moreover, the single-sequence thermodynamic modelling approach does not allow us to infer any biological function of the proposed LSRs.

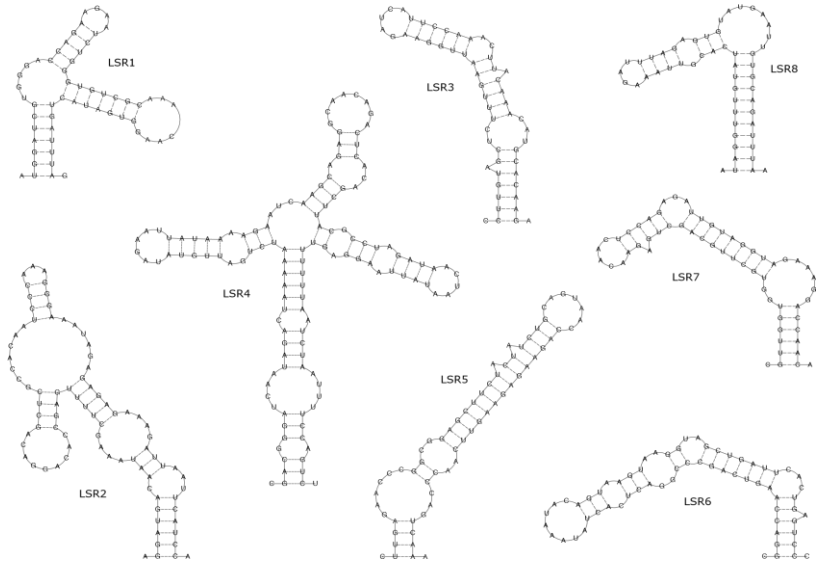


Figure 5.5: Locally stable RNA secondary structures unique to Reston virus.

5.3.4.2 Elements specific to human-pathogenic Ebolaviruses

Having predicted RNA structures that are only found in Reston virus, we also addressed the counter question as to whether there are RNA structures conserved in the human pathogenic Ebola species, but not in Reston virus. To this end, we performed consensus structure prediction on structural alignments of the individual genes (mRNAs), comprising the Zaire, Sudan, Tai Forest and Bundibugyo species. We classified potential candidate consensus structures with RNAz and R-scape and constructed covariance models, which were then used to confirm that homologous structures were not present in Reston virus. This approach yielded six highly conserved RNA structural elements, three in NP, and one in VP35, GP and VP30, respectively. Table 5.3 lists the genomic coordinates of each element in any of the pathogenic viruses. The fact that all six structures are located within UTRs is interesting, given that the overall MPI and SCI scores of the alignment of all six viruses are clearly lower in UTRs than in CDS regions (Figure 5.2).

Figure 5.6 shows the consensus structures of the six pathogenic-specific elements, utilizing the RNAalifold colour coding scheme (Figure 5.3a). All structures fold into simple or bulged stem-loops, with P-GP-S1 being the shortest and P-VP35-S1 being the longest. Figure 5.6 further highlights rich covariation support for each of the six elements, as manifested by compensatory and consistent mutations in all base pairs. Independent evaluation by R-scape yielded between one and seven significantly covarying base pairs in the pathogenic specific elements. Taken together, these six elements appear to be conserved and functional in all pathogenic Ebola strains, however, functional classification of these elements is beyond the scope of our approach.

ID	Gene	Location	Zaire	Sudan	Tai forest	Bundibugyo
P-NP-S1	NP	5'UTR	319-354	294-332	294-337	294-338
P-NP-S2	NP	5'UTR	361-398	339-376	351-385	344-378
P-NP-S3	NP	5'UTR	183-215	178-206	174-206	174-205
P-VP35-S1	VP35	3'UTR	4195-4258	4167-4230	4183-4258	4179-4252
P-GP-S1	GP	5'UTR	5974-5997	5942-5963	5962-5986	5955-5979
P-VP30-S1	VP30	5'UTR	8396-8441	8311-8358	8377-8420	8367-8406

Table 5.3: Structured RNA elements that are conserved in the pathogenic Ebolaviruses, but not in Reston virus. Each element has RNAz class probability greater than 0.99. Numbers in the columns of the individual Ebola species represent genomic coordinates of respective RefSeq sequences.

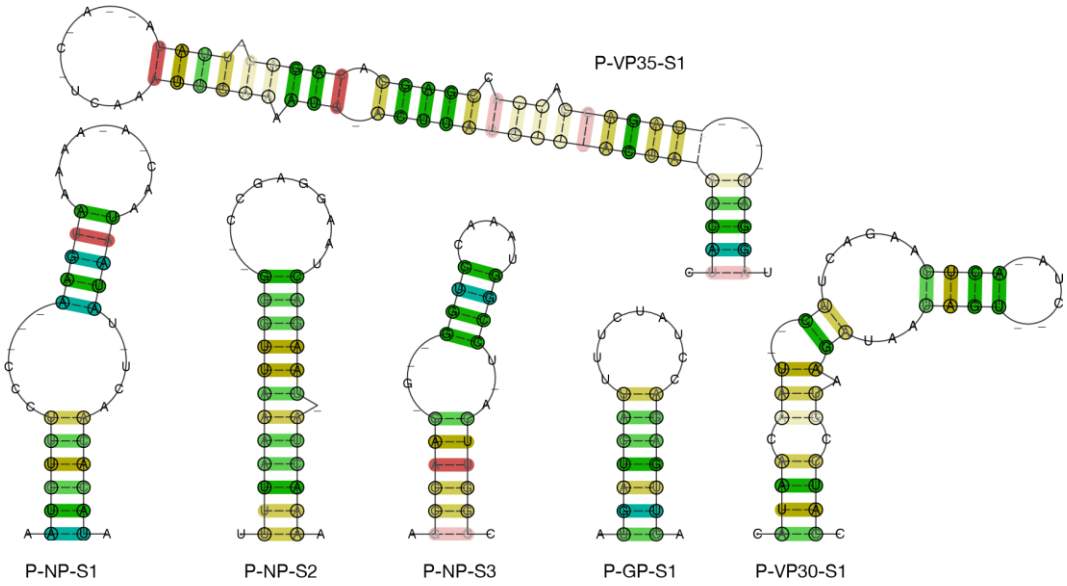


Figure 5.6: Consensus structure prediction of the RNA elements that are conserved in the human-pathogenic Ebola species, but not in to Reston virus. Colour coding follows the RNAalifold scheme (Figure 5.3a), highlighting rich covariation support.

5.3.5 Amino acid conservation correlates to conservation at the nucleotide level

We compared the conservation in the nucleotides of the coding regions of the seven Ebolavirus genes to that of the amino acid sequence. For all seven of the proteins - NP, VP35, VP40, GP, VP30, VP24 and L - the conservation matched almost perfectly. The same sliding window approach to identifying conservation was carried out for the amino acid sequences.

The N terminus of NP was generally highly conserved (0.5 - 1) with the region after residue 400 being less so. Residues 26 to 405 represent the NP core domain. A peak in conservation is also detected around the area containing the VP30 binding motif (residues 606 - 611). Conversely the VP35 protein is highly conserved in the latter half of the protein and poorly conserved in the first half. Despite this there is a clear peak of high conservation around residues 30 - 50 which contain the VP40 binding motif. This also correlates to a peak in high MPI and SCI at the nucleotide level. Thus, crucial protein domains and motifs appear highly conserved not just at the protein level but the nucleotide level as well.

VP40 is generally well conserved across the whole protein sequence except for a short patch in the N-terminus. GP is very poorly conserved in the region from residues 330 - 510, which corresponds to the mucin-like region domain. The receptor binding domain (54 - 201) is highly conserved, and this also correlates to a raised SCI and MPI. VP30 and VP24 are also highly conserved consistently across the whole protein sequence (though the start and end of VP30 does show a sharp decrease). The oligomerisation motif in VP30 (residues 94 - 112) appears in an area of generally high MPI and raised SCI relative to the surrounding regions. The same is also true of the NP binding motif (residues 202 - 237). The RNA polymerase (L) is well conserved consistently, though there are dips sporadically across the protein sequence. This is consistent with the RNA level. No conserved RNA secondary structure identified here correlates to a region of the amino acid sequence containing Differentially Conserved Positions (DCPs) that are known to differentiate the pathogenic species from the non-pathogenic. As part of this analysis DCPs were identified between the four pathogenic species - Ebola virus, Sudan virus, Bundibugyo virus, Tai Forest virus - and Reston virus. No change was detected from the last time this research was conducted (Martell et al., 2019), suggesting a robust approach to DCP identification and provides a clear dataset of residues that are important to pathogenicity.

5.4 Discussion

5.4.1 Structural conservation between Ebolavirus species is lower than expected

Given the previous identification of conserved RNA structures in Ebola virus, and the high conservation of protein structure and function exhibited between species, it was expected to find a consistent set of highly conserved RNA structures that were common to all species. With the exception of a few key areas this was not the case, and indeed the scores of the two measures mean pairwise identity (MPI) and structure conservation index (SCI) observed here can be interpreted as a proxy for the lack of conservation. In particular, the MPI scores range between 20 and 40 percent when comparing the six Ebolavirus species, a process that would generally produce a score of 25 percent by random chance. Considering previous analysis of viruses such as flaviviruses and alphaviruses produced a high level of conservation amongst closely

related groups, particularly in their UTRs (De Bernardi Schneider et al., 2019; Ochsenreiter et al., 2019), Ebolaviruses are unique in this regard.

In addition, the low number of predicted structures that are unique to the non-pathogenic Reston virus, just eight in total, may suggest that common structural RNA motifs are not key to differences in pathogenicity. However, the fact that we do not find conserved RNAs does not exclude the possibility that there are structured RNAs within each virus that are key for pathogenicity.

The general findings of a key set of conserved RNA structures present at gene termini in all six Ebolavirus species suggests that these play some as yet unknown regulatory or functional role. However, the lack of structures that differentiate pathogenic and non-pathogenic species suggest little at the RNA level that influences pathogenicity. Thus, greater emphasis is placed on the protein level, and DCPs that may lead to Reston virus not inducing disease in humans. The karyopherin alpha binding region in VP24 is one candidate for this (Martell et al., 2019; Pappalardo et al., 2016).

5.4.2 SL1 elements are conserved across the entire genus

The SL1 elements in the 5'UTRs of the mRNAs of Ebola virus were previously described to be well conserved (Sanchez et al., 1993). Here we show that these elements are also highly conserved in the five other known Ebolavirus species. Given the lower than expected conservation overall this result is surprising, and places a greater emphasis on these structures playing a pivotal role in the Ebolavirus life cycle. The genomic position at gene borders as well as the fact that these structures are well conserved in the non-pathogenic Reston virus suggests that these elements exert key functionality of the virus life cycle, such as transcription or translation regulation. This is also supported by the fact that these stem-loop structures are present on both the genomic and antigenomic level (Bach et al., 2020).

An interesting aspect of the SL1 structures relates to their association to viral non-coding RNA (vncRNA) biogenesis in Ebola virus and a putative involvement in host transcript silencing and regulation of viral replication. In this context, different groups have proposed the formation of virus-encoded microRNAs (miRNAs) in the regions overlapped by VP40.SL1 (Teng et al., 2015), VP24.SL1 (Liu et al., 2016) and L.SL1

(Liang et al., 2014) based on *in silico* modelling. While miRNAs are crucially related to posttranscriptional gene regulation in animals and plants, their maturation requires a cascade of nuclear and cytoplasmic processes, involving the RNAse III nucleases Drosha and Dicer (Bartel, 2004). Although miRNA production has been reported in various genera of positive (Lucia Morales, Juan Carlos Oliveros, Raul Fernandez-Delgado, Benjamin Robert tenOever, Luis Enjuanes, 2017; Weng et al., 2014) and negative (Perez et al., 2010) strand RNA viruses, there is doubt whether e.g. flaviviruses encode a functional, miRNA-like small RNA (Skalsky et al., 2014). More generally, the question whether or not viruses are capable of producing miRNA-like small molecules has been a matter of controversy in the literature (Aguado and tenOever, 2018). In this line, recent experimental evidence suggests that vncRNAs in Ebola virus likely do not result from endogenous miRNA processing pathways and are not capable of mediating host transcript silencing or affecting virus replication (Prasad et al., 2019). This finding, however, does not rule out the possibility that vncRNAs that are specific to particular Ebola species mediate pathogenesis. Further research is warranted here, in particular involving characterization and functional assessment of vncRNAs in different Ebola species.

In this contribution we also show that the GS/GE elements are part of highly conserved extended hairpin and stem-loop structures in all Ebolaviruses. The high degree of primary sequence conservation in the closing stems of these RNAs, which contain the GS/GE signals, suggest that the polymerase scans for a highly conserved double-stranded RNA rather than just the conserved primary sequences.

5.4.3 Ebolavirus divergence

The few number of well conserved RNA structures across the Ebolavirus genus is in clear contrast to previously studied viruses. This could be linked to the relative diversity in the phylogenetic trees of Ebolaviruses (Barrette et al., 2011; Pappalardo et al., 2016). Indeed, the non-pathogenic Reston virus is often clustered with the pathogenic Sudan virus, while the newly discovered Bombali virus - which is currently believed to not be pathogenic (Martell et al., 2019) is grouped with Ebola virus and Bundibugyo virus. Reston virus is also the only species of Asian origin, yet phylogenetically appears close to the African based Sudan virus. One potential avenue of research here is the

possibility of many as yet undiscovered Ebolavirus species. The lack of conserved structures hints at this, as arguably the six known species are far more diverse and different as expected from a core set of closely related genus. As we have seen in recent years, filoviruses are continually being discovered, and it is likely that there are more to come (Shi et al., 2018).

5.5 Materials and Methods

5.5.1 Data Sets

1430 full-length genomes were downloaded from NCBI, comprising all six ebolavirus species. Gene regions, including UTRs, were extracted from the full length genome based on coordinate data from the reference sequence of each individual species. For each of the seven Ebolavirus genes, identical sequences were removed to create a dataset of unique sequences. Additional sequences were manually removed if there were enough gaps to distort the results. No particular threshold was set for this.

For MPI/SCI scoring the six references sequences for each individual Ebolavirus species were used, obtained from Refseq. These sequences were also used to calculate the protein conservation. For DCP analysis the full dataset of 1430 sequences was used.

5.5.2 Characterization of conserved RNAs

We set out to find potentially functional RNAs in different ebolavirus species. Starting from a set of seed alignments of each gene in the set of six RefSeq sequences, we characterized structured RNAs that are conserved i) in all species, ii) only in Reston virus, but not in the human pathogenic ebolaviruses and iii) only in the human pathogenic ebolaviruses, but not in the human non-pathogenic Reston virus. We computed structural multiple sequence nucleotide alignments with locARNA (Will et al., 2007) and inferred consensus structures with RNAalifold (Bernhart et al., 2008). To obtain a more fine-grained picture of locally stable consensus structures, we employed RNALalifold from the ViennaRNA package (Lorenz et al., 2011), restricting

results to structures with a maximum base pair span of 150nt, and evaluated the significance of each candidate structure with RNAz (Gruber et al., 2010) and R-scape (Rivas et al., 2017). RNAz uses a support vector machine (SVM), a machine learning approach, to classify functional RNAs based on several descriptors including structure conservation and thermodynamic stability in comparison to random sequences of the same dinucleotide content. It reports for each input RNA alignment a class probability, i.e. a numerical value between 0 and 1 that is a proxy for functional conservation. Here, we required candidate structures to have an RNAz class probability of at least 0.95. R-scape performs statistical tests on individual columns of a nucleotide alignment to infer the significance of covariation patterns, providing an E-value for each covarying base pair. As RNALalifold candidates come with a predicted consensus structure, we performed two-set statistical tests with R-scape, using default parameters and only considering structures with at least one significantly covarying base pair. In addition we evaluated the thermodynamic stability of candidate structures with AlifoldZ (Washietl and Hofacker, 2004), which employs a random shuffling approach to estimate the background distribution of folding energies and expresses significance in terms of a normalized z-score. A negative z-score indicates that a particular fold is thermodynamically more stable than expected by chance, pinpointing towards structural conservation.

Structured RNAs are often evolutionarily conserved at the secondary structure level among phylogenetically related species. Covariance models (CMs), i.e. statistical models that extends Hidden-Markov-Models (HMMs) to simultaneously represent sequence and secondary structure information can be used to find homologous RNAs in comparative genomics screens (Nawrocki and Eddy, 2013). We have recently employed CMs to find conserved RNA elements in flaviviruses (Ochsenreiter et al., 2019) and alphaviruses (De Bernardi Schneider et al., 2019). Here we constructed CMs for candidate structures with RNAz class probability larger than 0.95 and at least one significantly covarying base pair.

To find structured RNAs that are present in Reston virus but not in the other ebolaviruses, we computed locally stable RNA structures in the Reston virus reference genome NC 004161.1 with RNALfold (Lorenz et al., 2011). We filtered for sequences no longer than 150nt and a thermodynamic stability expressed as free energy z-score of at least -2, i.e. requiring at least twice the standard deviation when compared to the free energy of 1000 dinucleotide shuffled sequences of the same nucleotide

composition. The genomic coordinates of these thermodynamically stable structures where then intersected with the Reston virus coordinates of the locally stable consensus structures computed from the alignment of all six Ebola strains by RNALalifold, targeting non-overlapping loci. By this, we could unambiguously identify eight regions that are present in Reston virus but not in the other viruses (Table 5.2).

On the other hand we were interested in structured RNAs that are present in the pathogenic ebolaviruses, but not in Reston virus. To this end, we produced locARNA (Will et al., 2007) structural nucleotide alignments of the mRNA sequences of each gene, encompassing the four pathogenic strains Ebola virus, Sudan virus, Tai forest virus and Bundibugyo virus. Conserved locally stable consensus structures in these four isolates were characterized with RNALalifold and subsequently evaluated with RNAz and R-scape, requiring the above criteria for further consideration. These rather stringent filtering criteria ensure that we exclude false positives here. A CM was then constructed for each candidate structure and checked against the Reston virus reference genome. Whenever no hit could be detected for a particular CM, we inferred that the underlying structure is only present in the pathogenic species, but not in Reston virus.

Chapter 6: Discussion

This thesis has presented four pieces of original research. Three of these discuss variation at either the nucleotide or protein level and how this may impact pathogenicity, while the other considers vaccination and herd immunity requirements. All but one chapter focusses on the filovirus family of viruses. In this chapter we consider all four chapters collectively, highlighting key findings and suggestions areas for future development or improvement.

The lack of viable vaccine strategies discussed in chapter two highlights the crucial need to better understand the molecular determinants of Ebolavirus pathogenicity. The research presented in chapters three and five consider this. Together these three studies offer a multifaceted approach to understanding Ebolaviruses, considering the differences between species and the identification of DCPs in chapter three and the conserved RNA structures present differ between species in chapter five. Additionally chapter four utilises the same approach to DCP identification as chapter three, demonstrating the robustness of our approach. Thus, the primary theme of this work has been to better understand the cause of Filovirus pathogenicity, identify variation in pathogenicity between the different species, and place this into the wider context of Filovirus outbreaks both past and present.

6.1 Herd immunity is unlikely to be achieved for Ebola virus

The research presented in chapter two highlights the difficulties in developing and coordinating an effective vaccination strategy against Ebolaviruses. The issues here fall into three main categories, developmental, logistical and social.

Developmental barriers are generally found in all vaccination programmes, as discovery and testing of a new vaccines requires time and money. Further to this the current vaccines in development as discussed in chapter two have yet to be fully tested in an outbreak setting, leading to questions about their efficacy. Should a new outbreak occur, and the current vaccines prove ineffective there is little chance of a viable replacement being ready on time. This issue has been an ongoing concern during the current COVID-19 outbreak, where vaccine development is outpaced by the spread

and entrenchment of the virus among large portions of the population (Lurie et al., 2020).

However of the three points, developmental issues are at current the least problematic. The availability of an effective vaccine and the successful ring vaccination trial in Guinea suggest that whether or not individual long term immunity or herd immunity can be achieved, the chain of transmission can at least be broken (Henao-Restrepo et al., 2017).

Logistical issues are perhaps more consequential. At current the vaccines available need to be stored at -70^{o}C, something which presents issues in sub-Saharan Africa (Arnemo et al., 2016). In addition, many of the areas affected by outbreaks in the past are remote and difficult to reach, although this has been suggested as a positive as it is difficult for the virus to spread to nearby communities in such an environment (Kramer et al., 2016; Majid et al., 2016; Peterson and Samy, 2016; Rashid, 2013).

Social issues also play an important role. The high proportion of respondents unwilling to receive a vaccine discussed in chapter two does not appear to be an isolated cultural or geographical issue. Indeed vaccination rates for measles has slowly declined in the UK and other western countries (Holzmann et al., 2016; Lee et al., 2016), while a survey conducted in the UK showed that almost 12% of people would not accept a vaccine for COVID-19 if offered (Freeman et al., 2020). Here it is clear that mistrust towards vaccines goes far beyond that of Ebola virus vaccines in western Africa, and further research should take this into account.

Despite the uncertainty surrounding the vaccination programmes a great deal of progress has been made in the field, given that there were no viable vaccine candidates until 2015 (Choi et al., 2015; Geisbert, 2017), which is encouraging. Regardless, it is clear that a much greater understanding of filoviruses is needed in order to be as prepared as possible for the next outbreak.

6.2 Bombali virus pathogenicity

The question as to whether Bombali virus is pathogenic has yet to be conclusively answered. While the research carried out and presented in chapter three suggests that Bombali virus may not be pathogenic in humans, this has yet to be officially confirmed.

The Bombali virus genome is similar to both the pathogenic Ebola virus and the non-pathogenic Reston virus (figure 7.1). Despite this, the Bombali virus amino acid residues at DCP positions are overwhelmingly more similar to Ebola virus (105; 63.25%) than to Reston virus 21 (12.65%). The remaining residues were unique to Bombali virus. Should it be discovered that Bombali virus is indeed non-pathogenic this would provide clear evidence that the key residues highlighted in VP24 – particularly M136L and R139S – play a crucial and defining role in Ebolavirus pathogenicity. If Bombali virus does cause disease in humans, then this will reduce the number of DCPs that can be considered as crucial for the differences in pathogenicity between Ebolavirus species by 19 (table 7.1) as Bombali virus shares many of these positions with Reston virus – the only confirmed non-pathogenic species at this time (Cantoni et al., 2016).

Species	Ebola	Reston	Bombali
Ebola	100	63.19	64.23
Reston		100	61.45
Bombali			100

Figure 6.1. Percentage similarity of reference genomes. Similarity of the reference genomes for Ebola virus, Reston virus and Bombali virus. Bombali virus shares a slightly higher conservation with Ebola virus than Reston virus.

However, were Bombali virus to be included in the non-pathogenic grouping, a much larger amount of amino acids (105) would be removed from consideration as DCPs (table 7.2). This highlights that Bombali virus is much closer in sequence identity to

the four pathogenic species than to Reston virus (a point discussed in chapter three) however as noted key residues in VP24 agree with Reston virus over the pathogenic variants.

This simple assessment is however problematic, as this is based on a small sample size for Bombali virus – just two sequences. An increasing number of Bombali virus genomes may increase variation that could change any of these results, however it is still interesting to note at this time. Should Bombali virus prove to be non-pathogenic a dataset of approximately 61 DCPs would likely emerge, just over a third of the number presented in chapter three.

The isolation of antibodies for Ebolavirus in individuals who cohabited with bats from which Bombali virus was isolated does give greater weight to the argument that it is non-pathogenic in nature (Goldstein et al., 2018; Mafopa et al., 2017), but further investigation into the virus is needed at this time. In addition, experimental investigation of this region and of all DCP positions should be carried out to assess their impact on pathogenicity.

NP			**VP30**		
EBOV	**RESTV**	**BOMV**	**EBOV**	**RESTV**	**BOMV**
R4	G4	R4	Y39	R40	N50
E16	D16	D16	T52	N53	V63
S30	T30	S30	V53	L54	M64
R39	K39	R39	T63	I64	L74
I56	V56	I56	E93	D94	E104
V64	I64	V64	T96	N97	T107
R105	K105	R105	R98	H99	R109
M137	L137	M137	K107	R108	K118
F212	Y212	F212	S111	I112	L122
K274	R274	R274	L116	S117	V127
S279	A279	S279	N117	Q118	C128
K416	N416	R416	A120	S121	A131
Y421	Q421	Y421	T150	I151	I161
D426	E426	E426	Q157	R158	K168
D435	N435	D435	I159	L160	L170

D443	E443	V443	E205	D206	E216
T453	I453	T453	R262	A263	K273
P497	A497	S497	S268	Q269	A279
T563	S563	A563	E271	S272	N282
I565	V565	I565	G278	N279	T289
P602	T602	R602	**VP24**		
N641	Q641	S641	**EBOV**	**RESTV**	**BOMV**
A705	R705	A705	L17	M17	L17
G717	N717	G717	V22	I22	V22
VP35			V31	I31	V31
EBOV	**RESTV**	**BOMV**	I102	L102	I102
S26	T15	S27	T131	S131	T131
E48	D37	D49	N132	T132	A132
D76	E65	D77	M136	L136	L136
E84	K73	D85	Q139	R139	R139
E85	K74	E86	R140	S140	R/S140
S92	M81	S93	T226	A226	T226
V97	T86	V98	S248	L248	S248
T101	N90	A102	**L**		
S106	A95	S107	**EBOV**	**RESTV**	**BOMV**
T112	S101	S113	V66	T66	V66
V121	I110	V122	I136	L136	L136
A125	G114	A126	L146	V146	L146
A154	S143	A155	T202	I202	E202
T159	V148	T160	A221	S221	A221
E160	D149	E161	Q223	L223	Q223
G167	K156	G168	T226	S226	K226
S174	A163	S175	H227	Q227	Y227
I181	L170	I182	V236	I236	V236
E269	D258	D270	L283	V283	L283
A290	V279	I291	T330	D330	T330
V314	A303	V315	E350	D350	E350
Q329	K318	Q330	T361	S361	T361
VP40			L365	F365	F365
EBOV	**RESTV**	**BOMV**	V379	I379	I379

V4	G4	T4	Q447	H447	Q447
T46	V46	I/T46	P450	S450	P450
P85	T85	P85	D465	N465	D465
T105	I105	K105	S847	A847	S847
I122	V122	I122	S868	A868	S868
A128	I128	T128	T1024	N1024	N1024
G201	N201	N201	R1073	K1073	R1073
F209	L209	F209	A1119	S1119	A1119
L244	I244	L244	P1163	A1161	P1163
Q245	P245	Q245	D1189	S1187	D1189
M259	V259	V259	A1214	S1212	A1214
H269	Q269	Q269	R1217	K1215	R1217
T277	Q277	H277	D1237	E1235	D1237
I293	V293	I293	R1354	K1352	R1354
V323	H323	A323	T1366	A1364	T1366
E325	D325	E325	I1408	M1406	I1408
GP			I1414	L1412	I1414
EBOV	**RESTV**	**BOMV**	S1436	N1434	S1436
M1	G2	n/a	S1473	C1471	S1473
F31	I32	V27	L1488	Y1486	I1488
V37	I38	V33	I1499	L1497	I1499
V45	A46	V41	S1506	A1504	S1506
V75	I76	I71	I1509	V1507	V1509
S196	A197	P192	L1624	Y1624	L1624
I260	L261	I256	C1628	S1628	C1638
T269	S270	T265	V1762	I1760	V1760
S307	H308	S303	V1850	T1848	T1848
S476	P477	I479	T1873	S1871	T1871
R498	K499	R494	R1916	N1914	R1914
R500	K501	K496	E1941	R1939	E1939
N514	D515	N510	L2044	I2042	L2044
Q521	V522	H517	S2077	T2075	S2075
I584	L585	I580	E2098	D2096	E2096
D607	S608	D603	L2157	V2155	L2155
K622	E623	R618	R2168	H2166	K2166

Q638	H639	Q634	R2175	K2173	R2173
W644	L645	W640	L2177	F2175	W2175
T659	I660	V655	M2186	L2184	M2184

Figure 6.1. List of DCPs showing the residues for the pathogenic Ebola virus and non-pathogenic Reston virus alongside Bombali virus. Residues that would be removed as DCPs if Bombali virus were included in the pathogenic grouping are highlighted in red, a total of 19 positions.

NP			**VP30**		
EBOV	**RESTV**	**BOMV**	**EBOV**	**RESTV**	**BOMV**
R4	G4	R4	Y39	R40	N50
E16	D16	D16	T52	N53	V63
S30	T30	S30	V53	L54	M64
R39	K39	R39	T63	I64	L74
I56	V56	I56	E93	D94	E104
V64	I64	V64	T96	N97	T107
R105	K105	R105	R98	H99	R109
M137	L137	M137	K107	R108	K118
F212	Y212	F212	S111	I112	L122
K274	R274	R274	L116	S117	V127
S279	A279	S279	N117	Q118	C128
K416	N416	R416	A120	S121	A131
Y421	Q421	Y421	T150	I151	I161
D426	E426	E426	Q157	R158	K168
D435	N435	D435	I159	L160	L170
D443	E443	V443	E205	D206	E216
T453	I453	T453	R262	A263	K273
P497	A497	S497	S268	Q269	A279
T563	S563	A563	E271	S272	N282
I565	V565	I565	G278	N279	T289
P602	T602	R602	**VP24**		
N641	Q641	S641	**EBOV**	**RESTV**	**BOMV**
A705	R705	A705	L17	M17	L17

G717	N717	G717	V22	I22	V22
VP35			V31	I31	V31
EBOV	**RESTV**	**BOMV**	I102	L102	I102
S26	T15	S27	T131	S131	T131
E48	D37	D49	N132	T132	A132
D76	E65	D77	M136	L136	L136
E84	K73	D85	Q139	R139	R139
E85	K74	E86	R140	S140	R/S140
S92	M81	S93	T226	A226	T226
V97	T86	V98	S248	L248	S248
T101	N90	A102	**L**		
S106	A95	S107	**EBOV**	**RESTV**	**BOMV**
T112	S101	S113	V66	T66	V66
V121	I110	V122	I136	L136	L136
A125	G114	A126	L146	V146	L146
A154	S143	A155	T202	I202	E202
T159	V148	T160	A221	S221	A221
E160	D149	E161	Q223	L223	Q223
G167	K156	G168	T226	S226	K226
S174	A163	S175	H227	Q227	Y227
I181	L170	I182	V236	I236	V236
E269	D258	D270	L283	V283	L283
A290	V279	I291	T330	D330	T330
V314	A303	V315	E350	D350	E350
Q329	K318	Q330	T361	S361	T361
VP40			L365	F365	F365
EBOV	**RESTV**	**BOMV**	V379	I379	I379
V4	G4	T4	Q447	H447	Q447
T46	V46	I/T46	P450	S450	P450
P85	T85	P85	D465	N465	D465
T105	I105	K105	S847	A847	S847
I122	V122	I122	S868	A868	S868
A128	I128	T128	T1024	N1024	N1024
G201	N201	N201	R1073	K1073	R1073
F209	L209	F209	A1119	S1119	A1119

L244	I244	L244	P1163	A1161	P1163
Q245	P245	Q245	D1189	S1187	D1189
M259	V259	V259	A1214	S1212	A1214
H269	Q269	Q269	R1217	K1215	R1217
T277	Q277	H277	D1237	E1235	D1237
I293	V293	I293	R1354	K1352	R1354
V323	H323	A323	T1366	A1364	T1366
E325	D325	E325	I1408	M1406	I1408
GP			I1414	L1412	I1414
EBOV	**RESTV**	**BOMV**	S1436	N1434	S1436
M1	G2	n/a	S1473	C1471	S1473
F31	I32	V27	L1488	Y1486	I1488
V37	I38	V33	I1499	L1497	I1499
V45	A46	V41	S1506	A1504	S1506
V75	I76	I71	I1509	V1507	V1509
S196	A197	P192	L1624	Y1624	L1624
I260	L261	I256	C1628	S1628	C1638
T269	S270	T265	V1762	I1760	V1760
S307	H308	S303	V1850	T1848	T1848
S476	P477	I479	T1873	S1871	T1871
R498	K499	R494	R1916	N1914	R1914
R500	K501	K496	E1941	R1939	E1939
N514	D515	N510	L2044	I2042	L2044
Q521	V522	H517	S2077	T2075	S2075
I584	L585	I580	E2098	D2096	E2096
D607	S608	D603	L2157	V2155	L2155
K622	E623	R618	R2168	H2166	K2166
Q638	H639	Q634	R2175	K2173	R2173
W644	L645	W640	L2177	F2175	W2175
T659	I660	V655	M2186	L2184	M2184

Figure 6.2. List of DCPs showing the residues for the pathogenic Ebola virus and non-pathogenic Reston virus alongside Bombali virus. Residues that would be removed as DCPs if Bombali virus were included in the non-pathogenic grouping are highlighted in red, a total of 105 positions.

The conclusion that relatively few amino acid mutations can have major changes on filovirus pathogenicity has been the main finding of this research. With this in mind there is the possibility that other factors may play a role in these changes, particular at the RNA level. Chapter five focuses on this aspect, and concludes that there are key differences between the conserved RNA structures present in the different Ebolavirus species. While it is not possible to associate these structures with a known function or role, the presence of them highlights how different the pathogenic and non-pathogenic species are, and that a multifaceted approach to understanding pathogenicity determination is required.

6.3 Future Work

Much of the research carried out in this thesis has already built upon two published works - *Conserved differences in protein sequence determine the human pathogenicity of Ebolaviruses (Pappalardo et al., 2016)* and *Changes Associated With Ebola Virus Adaptation to Novel Species (Pappalardo et al., 2017)*. There are still many avenues for expansion however, the most notable include:

1. New data is available for the 2018 – 2020 Kivu Ebolavirus outbreak in the DRC. Epidemiological data from here can be used to refine the R_0 average and study the changes implemented since the West African outbreak. Furthermore virus genome sequences obtained from this outbreak and any future outbreaks can be added to further DCP analysis in the future.

2. DCPs highlighted as key to differences observed in pathogenicity between Reston virus and the other member of the Ebolavirus genus should be investigated further experimentally. Insertion of Ebola virus VP24 into a Reston virus virion and the subsequent study of pathogenic effect could definitively prove the importance of VP24 to the pathogenic determination. However, such a study may prove difficult owing to the restrictions on Ebolavirus study given its biohazard safety level.

3. The identification of novel RNA structures in the Ebolavirus genome raises an interesting avenue for future research. Given the few amino acids differences that appear to alter pathogenicity it may be useful to study the

differences in conserved RNA secondary structures between species. Here these structures may play an as yet unknown role in Ebolavirus pathogenicity and disease.

4. The analysis carried out in chapter five identified an interesting point regarding Ebolaviruses. The level of conservation at the RNA level is significantly lower than expected based on research of other viruses – notably Flaviviruses and Alphaviruses. This hints at multiple undiscovered species that could bridge the gap between the current known species. Further investigation to this should be considered, especially as discovery of new filoviruses is increasing in frequency.

5. One avenue of potential research concerns the analysis of Ebola virus genomes from separate outbreaks. Despite there being many years and large distances between outbreaks the overall conservation of both protein and nucleotide sequences has remained remarkably high. All research carried out in this thesis concerning Ebola virus has considered all sequences as one giant dataset, however further research should be conducted to examine the intraspecies variation here. Comparison of the 1976 sequences to that of the 2013 – 2016 sequences may show key residues that account for the variability in pathogenicity.

In addition, the pipeline used to generate and analyse DCPs in chapters three, four, and to a lesser extent five, is in the process of being made freely available via a webserver or direct download. This will allow other researchers to quickly and accurately detect and study DCPs between two or more groups of sequences, allowing for valuable research of these positions and the effect they have on protein structure and function. It is hoped that this can be implemented quickly to aid in investigation of SARS-CoV-2, and will be able to provide information on a wide range of viruses, as it has done here, primarily with SARS-CoV-2 and Ebola virus.

References

Agua-Agum, J., Ariyarajah, A., Aylward, B., Bawo, L., Bilivogui, P., Blake, I.M., Brennan, R.J., Cawthorne, A., Cleary, E., Clement, P., et al. (2016). Exposure Patterns Driving Ebola Transmission in West Africa: A Retrospective Observational Study. PLoS Med. *13*, 1–23.

Aguado, L.C., and tenOever, B. (2018). RNA virus building blocks—miRNAs not included. PLoS Pathog. *14*, 1–6.

Agusto, F.B. (2017). Mathematical model of Ebola transmission dynamics with relapse and reinfection. Math. Biosci. *283*, 48–59.

Akinfeyeva, L., Aksyonova, O., Vasilyevich, I., Ginko, Z., Zarkov, K., and Zubavichene, N. (2005). A case of ebola hemorrhagic fever. Infektsionnye Bolezn. *3*, 85–88.

Al-Hazmi, A. (2016). Challenges presented by MERS corona virus, and SARS corona virus to global health. Saudi J. Biol. Sci. *23*, 507–511.

Albariño, C.G., Guerrero, L.W., Jenks, H.M., Chakrabarti, A.K., Ksiazek, T.G., Rollin, P.E., and Nichol, S.T. (2017). Insights into Reston virus spillovers and adaption from virus whole genome sequences. PLoS One *12*, 1–13.

Albertini, A.A. V., Schoehn, G., Weissenhorn, W., and Ruigrok, R.W.H. (2008). Structural aspects of rabies virus replication. Cell. Mol. Life Sci. *65*, 282–294.

Aleshin, A.E., Shiryaev, S.A., Strongin, A.Y., and Liddington, R.C. (2007). Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold . Protein Sci. *16*, 795–806.

Alexander, K.A., Sanderson, C.E., Marathe, M., Lewis, B.L., Rivers, C.M., Shaman, J., Drake, J.M., Lofgren, E., Dato, V.M., Eisenberg, M.C., et al. (2015). What factors might have led to the emergence of ebola in West Africa? PLoS Negl. Trop. Dis. *9*, 1–26.

Alexandra Bilak, Martina Caterina, Guillaume Charron, Sophie Crozet, Laura Rubio Díaz-Leal, Florence Foster, Justin Ginnetti, Jacopo Giorgi, Anne-Kathrin Glatz, Kristel Guyon, Caroline Howard, Melanie Kesmaecker-Wissing, Sarah Kilany,

Johanna Klos, Freder, M.Y. (2015). Global overview of Ebola Research. Eur. Comm. Rep. 1–45.

Allan, G.M., and Arroll, B. (2014). Prevention and treatment of the common cold: making sense of the evidence. Can. Med. Assoc. J. *186*, 190–199.

Althaus, C.L. (2015a). Rapid drop in the reproduction number during the Ebola outbreak in the Democratic Republic of Congo. PeerJ *3*, e1418.

Althaus, C.L. (2015b). Ebola superspreading. Lancet Infect. Dis. *15*, 507–508.

Althaus, C.L., Low, N., Musa, E.O., Shuaib, F., and Gsteiger, S. (2015). Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control. Epidemics *11*, 80–84.

Anderson, R.. (1992). The concept of herd immunity and the design of communty based immunization programs. Vaccine *10*, 928–935.

Armstrong, D.R., Berrisford, J.M., Conroy, M.J., Gutmanas, A., Anyango, S., Choudhary, P., Clark, A.R., Dana, J.M., Deshpande, M., Dunlop, R., et al. (2020). PDBe: improved findability of macromolecular structure data in the PDB. Nucleic Acids Res. *48*, D335–D343.

Arnemo, M., Viksmoen Watle, S.S., Schoultz, K.M., Vainio, K., Norheim, G., Moorthy, V., Fast, P., Røttingen, J.A., and Gjøen, T. (2016). Stability of a Vesicular Stomatitis Virus-Vectored Ebola Vaccine. J. Infect. Dis. *213*, 930–933.

Artenstein, A.W. (2012). The discovery of viruses: Advancing science and medicine by challenging dogma. Int. J. Infect. Dis. *16*, e470–e473.

Bach, S., Demper, J.-C., Biedenkopf, N., Becker, S., and Hartmann, R.K. (2020). RNA secondary structure at the transcription start site influences EBOV transcription initiation and replication in a length- and stability-dependent manner. RNA Biol.

Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N., Soropogui, B., Sow, M.S., Keïta, S., De Clerck, H., et al. (2014). Emergence of Zaire Ebola Virus Disease in Guinea. N. Engl. J. Med. *371*, 1418–1425.

Banadyga, L., Hoenen, T., Ambroggio, X., Dunham, E., Groseth, A., and Ebihara, H. (2017a). Ebola virus VP24 interacts with NP to facilitate nucleocapsid assembly

and genome packaging. Sci. Rep. 1–14.

Banadyga, L., Dolan, M.A., and Ebihara, H. (2017b). Rodent-Adapted Filoviruses and the Molecular Basis of Pathogenesis. J. Mol. Biol. *428*, 3449–3466.

Barman, S., Ali, A., Hui, E.K.W., Adhikary, L., and Nayak, D.P. (2001). Transport of viral proteins to the apical membranes and interaction of matrix protein with glycoproteins in the assembly of influenza viruses. Virus Res. *77*, 61–69.

Barnes, K.G., Kindrachuk, J., Lin, A.E., Wohl, S., Qu, J., Tostenson, S.D., Dorman, W.R., Busby, M., Siddle, K.J., Luo, C.Y., et al. (2017). Evidence of ebola virus replication and high concentration in semen of a patient during recovery. Clin. Infect. Dis. *65*, 1400–1403.

Barrette, R.W., Xu, L., Rowland, J.M., and McIntosh, M.T. (2011). Current perspectives on the phylogeny of Filoviridae. Infect. Genet. Evol. *11*, 1514–1519.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281–297.

Bartlam, M., Yang, H., and Rao, Z. (2005). Structural insights into SARS coronavirus proteins. Curr. Opin. Struct. Biol. *15*, 664–672.

Baseler, L., Chertow, D.S., Johnson, K.M., Feldmann, H., and Morens, D.M. (2017). The Pathogenesis of Ebola Virus Disease∗. Annu. Rev. Pathol. Mech. Dis. *12*, 387–418.

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., et al. (2017). UniProt: The universal protein knowledgebase. Nucleic Acids Res. *45*, D158–D169.

Beer, B., Kurth, R., and Bukreyev, A. (1999). Characteristics of filoviridae: Marburg and Ebola viruses. Naturwissenschaften *86*, 8–17.

Beigel, J.H., Nam, H.H., Adams, P.L., Krafft, A., Ince, W.L., El-Kamary, S.S., and Sims, A.C. (2019). Advances in respiratory virus therapeutics – A meeting report from the 6th isirv Antiviral Group conference. Antiviral Res. *167*, 45–67.

Belyi, V.A., Levine, A.J., and Skalka, A.M. (2010). Sequences from Ancestral Single-Stranded DNA Viruses in Vertebrate Genomes: the Parvoviridae and Circoviridae Are More than 40 to 50 Million Years Old. J. Virol. *84*, 12458–12462.

Bernardeschi, C., Valeyrie-Allanore, L., Ortonne, N., Gressier, L., Wallet-Faber, N., Bernard, P., Hezode, C., Duclos-Vallée, J., Samuel, D., Mallet, V., et al. (2016). Dermatological side-effects in hepatitis C infected patients under a triple regimen associating pegylated interferon, ribavirin and telaprevir. J. Eur. Acad. Dermatology Venereol. *30*, 141–143.

De Bernardi Schneider, A., Ochsenreiter, R., Hostager, R., Hofacker, I.L., Janies, D., and Wolfinger, M.T. (2019). Updated phylogeny of Chikungunya virus suggests lineage-specific RNA architecture. Viruses *11*, 1–16.

Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R., and Stadler, P.F. (2008). RNAalifold: Improved consensus structure prediction for RNA alignments. BMC Bioinformatics *9*, 1–13.

Beyer, W.E.P., Nauta, J.J.P., Palache, A.M., Giezeman, K.M., and Osterhaus, A.D.M.E. (2011). Immunogenicity and safety of inactivated influenza vaccines in primed populations: A systematic literature review and meta-analysis. Vaccine *29*, 5785–5792.

Biedenkopf, N., Schlereth, J., Grünweller, A., Becker, S., and Hartmann, R.K. (2016). RNA-binding of Ebola virus VP30 is essential for activating viral transcription. J. Virol. *90*, 7481–7496.

Den Blaauwen, T., Andreu, J.M., and Monasterio, O. (2014). Bacterial cell division proteins as antibiotic targets. Bioorg. Chem. *55*, 27–38.

Blackshields, G., Sievers, F., Shi, W., Wilm, A., and Higgins, D.G. (2010). Sequence embedding for fast construction of guide trees for multiple sequence alignment. Algorithms Mol. Biol. *5*, 1–11.

Bloom, D.E., and Cadarette, D. (2019). Infectious disease threats in the twenty-first century: Strengthening the global response. Front. Immunol. *10*, 1–12.

Boevink, P. (2005). Virus-Host Interactions during Movement Processes. Plant Physiol. *138*, 1815–1821.

Bonanni, P., Sacco, C., Donato, R., and Capei, R. (2014). Lifelong vaccination as a key disease-prevention strategy. Clin. Microbiol. Infect. *20*, 32–36.

Borges do Nascimento, I.J., Cacic, N., Abdulazeem, H.M., von Groote, T.C.,

Jayarajah, U., Weerasekara, I., Esfahani, M.A., Civile, V.T., Marusic, A., Jeroncic, A., et al. (2020). Novel Coronavirus Infection (COVID-19) in Humans: A Scoping Review and Meta-Analysis. J. Clin. Med. *9*.

Borisevich, I. V, Markin, V.A., Firsova, I. V, Evseev, A.A., Khamitov, R.A., and Maksimov, V.A. (2006). Hemorrhagic (Marburg, Ebola, Lassa, and Bolivian) fevers: epidemiology, clinical pictures, and treatment. Vopr. Virusol. *51*, 8–16.

Bornholdt, Z.A., Noda, T., Abelson, D.M., Halfmann, P., Wood, M.R., Kawaoka, Y., and Saphire, E.O. (2013). Structural rearrangement of ebola virus vp40 begets multiple functions in the virus life cycle. Cell *154*, 763–774.

Bos, L. (1999). Beijerinck's work on tobacco mosaic virus: historical context and legacy. Philos. Trans. R. Soc. B Biol. Sci. *354*, 675–685.

Bouvier, N.M., and Palese, P. (2008). The Biology of Influenza Viruses. Vaccine *12*, D49–D53.

Bradfute, S.B., Warfield, K.L., and Bray, M. (2012). Mouse models for filovirus infections. Viruses *4*, 1477–1508.

Brauburger, K., Deflubé, L.R., and Mühlberger, E. (2014). Filovirus transcription and replication. In Biology and Pathogenesis of Rhabdo- and Filoviruses, pp. 515–555.

Breitbart, M., and Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? Trends Microbiol. *13*, 278–284.

Briët, J., and Harremoës, P. (2009). Properties of classical and quantum Jensen-Shannon divergence. Phys. Rev. A - At. Mol. Opt. Phys. *79*, 1–13.

Brito, A.F., and Pinney, J.W. (2017). Protein–Protein Interactions in Virus–Host Systems. Front. Microbiol. *8*, 1–11.

Brötz-Oesterhelt, H., and Sass, P. (2014). Bacterial caseinolytic proteases as novel targets for antibacterial treatment. Int. J. Med. Microbiol. *304*, 23–30.

Burk, R., Bollinger, L., Johnson, J.C., Wada, J., Radoshitzky, S.R., Palacios, G., Bavari, S., Jahrling, P.B., and Kuhn, J.H. (2016). Neglected filoviruses. FEMS Microbiol. Rev. *40*, 494–519.

Burley, S.K., Berman, H.M., Bhikadiya, C., Bi, C., Chen, L., Costanzo, L. Di, Christie, C., Duarte, J.M., Dutta, S., Feng, Z., et al. (2019). Protein Data Bank: The single

global archive for 3D macromolecular structure data. Nucleic Acids Res. *47*, D520–D528.

Camacho, A., Kucharski, A.J., Funk, S., Breman, J., Piot, P., and Edmunds, W.J. (2014). Potential for large outbreaks of Ebola virus disease. Epidemics *9*, 70–78.

Cantoni, D., Hamlet, A., Michaelis, M., Wass, M.N., and Rossman, J.S. (2016). Risks Posed by Reston, the Forgotten Ebolavirus. MSphere *1*, 1–10.

Cao, Y.G., Chen, Y.C., Hao, K., Zhang, M., and Liu, X.Q. (2008). An in vivo approach for globally estimating the drug flow between blood and tissue for nafamostat mesilate: The main hydrolysis site determination in human. Biol. Pharm. Bull. *31*, 1985–1989.

Capra, J.A., and Singh, M. (2007). Predicting functionally important residues from sequence conservation. Bioinformatics *23*, 1875–1882.

Carlos, A.M., and Lewis, F.D. (2012). Smallpox and Native American mortality: The 1780s epidemic in the Hudson Bay region. Explor. Econ. Hist. *49*, 277–290.

Castillo-Chavez, C., Curtiss, R., Daszak, P., Levin, S.A., Patterson-Lomba, O., Perrings, C., Poste, G., and Towers, S. (2015). Beyond Ebola: Lessons to mitigate future pandemics. Lancet Glob. Heal. *3*, e354–e355.

Cataldi, J.R., Dempsey, A.F., and O'Leary, S.T. (2016). Measles, the media, and MMR: Impact of the 2014 - 15 measles outbreak. Vaccine *34*, 6375–6380.

Chakraborti, S., Prabakaran, P., Xiao, X., and Dimitrov, D.S. (2005). The SARS coronavirus S glycoprotein receptor binding domain: Fine mapping and functional characterization. Virol. J. *2*, 1–10.

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

Channappanavar, R., and Perlman, S. (2017). Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. Semin. Immunopathol. *39*, 529–539.

Chanthamontri, C.K., Jordan, D., Wang, W., Wu, C., Brett, T.J., Gross, M.L., and Leung, D.W. (2019). Ebola Viral Protein 35 N-terminus is a Parallel Tetramer.

Biochemistry *58*, 657–664.

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., et al. (2020a). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet *395*, 507–513.

Chen, T., Leung, R.K.K., Liu, R., Chen, F., Zhang, X., Zhao, J., and Chen, S. (2014). Risk of imported Ebola virus disease in China. Travel Med. Infect. Dis. *12*, 650–658.

Chen, Y., Liu, Q., and Guo, D. (2020b). Emerging coronaviruses: Genome structure, replication, and pathogenesis. J. Med. Virol. *92*, 418–423.

Cheresiz, S. V., Semenova, E.A., and Chepurnov, A.A. (2016). Adapted Lethality: What We Can Learn from Guinea Pig-Adapted Ebola Virus Infection Model. Adv. Virol. *2016*.

Choi, W.Y., Hong, K.-J., Hong, J.E., and Lee, W.-J. (2015). Progress of vaccine and drug development for Ebola preparedness. Clin. Exp. Vaccine Res. *4*, 11–16.

Chowell, G., and Nishiura, H. (2014). Transmission dynamics and control of Ebola virus disease (EVD): a review. BMC Med. *12*, 196.

Chowell, G., and Viboud, C. (2015). Controlling Ebola: Key role of Ebola treatment centres. Lancet Infect. Dis. *15*, 139–141.

Chowell, G., Hengartner, N.W., Castillo-Chavez, C., Fenimore, P.W., and Hyman, J.M. (2004). The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda. J. Theor. Biol. *229*, 119–126.

Ciftci, E., Karbuz, A., and Kendirli, T. (2016). Influenza and the use of oseltamivir in children. Türk Pediatr. Arşivi *51*, 63–71.

Cinatl, J., Cinatl, J., Weber, B., Rabenau, H., Giimbel, H.O., Chenot, J., Scholz, M., Encke, A., and Doerr, H.W. (1995). In vitro inhibition of human cytomegalovirus replication in human foreskin 639 fibroblasts and endothelial cells by ascorbic acid 2-phosphate. Antiviral Res. *27*, 405–418.

Cinatl, J., Morgenstern, B., Bauer, G., Chandra, P., Rabenau, H., and Doerr, H.W. (2003). Glycyrrhizin, an active component of liquorice roots, and replication of SARS-associated coronavirus. Lancet *361*, 2045–2046.

Cinatl, J., Hoever, G., Morgenstern, B., Preiser, W., Vogel, J.U., Hofmann, W.K., Bauer, G., Michaelis, M., Rabenau, H.F., and Doerr, H.W. (2004). Infection of cultured intestinal epithelial cells with severe acute respiratory syndrome coronavirus. Cell. Mol. Life Sci. *61*, 2100–2112.

Cinatl, J., Michaelis, M., Morgenstern, B., and Doerr, H.W. (2005). High-dose hydrocortisone reduces expression of the pro-inflammatory chemokines CXCL8 and CXCL10 in SARS coronavirus-infected intestinal cells. Int. J. Mol. Med. *15*, 323–327.

Coltart, C.E.M., Lindsey, B., Ghinai, I., Johnson, A.M., and Heymann, D.L. (2017). The Ebola outbreak, 2013 – 2016: old lessons for new epidemics. Philos. Trans. B *372*, 2013–2016.

Corman, V.M., Muth, D., Niemeyer, D., and Drosten, C. (2018). Hosts and Sources of Endemic Human Coronaviruses. Adv. Virus Res. *100*, 163–187.

Crary, S.M., Towner, J.S., Honig, J.E., Shoemaker, T.R., and Nichol, S.T. (2003). Analysis of the role of predicted RNA secondary structures in Ebola virus replication. Virology *306*, 210–218.

Cross, R.W., Fenton, K.A., Geisbert, J.B., Ebihara, H., Mire, C.E., and Geisbert, T.W. (2015). Comparison of the Pathogenesis of the Angola and Ravn Strains of Marburg Virus in the Outbred Guinea Pig Model. J. Infect. Dis. *212*, S258–S270.

Cross, R.W., Mire, C.E., Borisevich, V., Geisbert, J.B., Fenton, K.A., and Geisbert, T.W. (2016). The domestic ferret (mustela putorius furo) as a lethal infection model for 3 species of ebolavirus. J. Infect. Dis. *214*, 565–569.

Cui, J., Li, F., and Shi, Z.L. (2019). Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. *17*, 181–192.

D'Silva, J.P., and Eisenberg, M.C. (2017). Modeling spatial invasion of Ebola in West Africa. J. Theor. Biol. *428*, 65–75.

Danielle Iuliano, A., Roguski, K.M., Chang, H.H., Muscatello, D.J., Palekar, R., Tempia, S., Cohen, C., Michael Gran, J., Schanzer, D., Cowling, B.J., et al. (2017). Estimates of global seasonal influenza-associated respiratory mortality: a modelling study for the Global Seasonal Influenza-associated Mortality Collaborator Network*. Lancet *391*, 1285–1300.

Daugelaite, J., O' Driscoll, A., and Sleator, R.D. (2013). An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics. ISRN Biomath. *2013*, 1–14.

Deen, G. et al. (2018). Ebola RNA Persistence in Semen of Ebola Virus Disease Survivors - Final Report. N. Engl. J. Med. *377*, 1428–1437.

Diallo, B., Sissoko, D., Loman, N.J., Bah, H.A., Bah, H., Worrell, M.C., Conde, L.S., Sacko, R., Mesfin, S., Loua, A., et al. (2016). Resurgence of Ebola Virus Disease in Guinea Linked to a Survivor with Virus Persistence in Seminal Fluid for More Than 500 Days. Clin. Infect. Dis. *63*, 1353–1356.

Diehl, W.E., Lin, A.E., Grubaugh, N.D., Carvalho, L.M., Kim, K., Kyawe, P.P., McCauley, S.M., Donnard, E., Kucukural, A., McDonel, P., et al. (2016). Ebola Virus Glycoprotein with Increased Infectivity Dominated the 2013–2016 Epidemic. Cell *167*, 1088–1098.e6.

Dietzel, E., Schudt, G., Krähling, V., Matrosovich, M., and Becker, S. (2017). Functional Characterization of Adaptive Mutations during the West African Ebola Virus Outbreak. J. Virol. *91*, 1–13.

Do, T.S., and Lee, Y.S. (2016). Modeling the Spread of Ebola. Osong Public Heal. Res. Perspect. *7*, 43–48.

Donald, J.E., and Shakhnovich, E.I. (2005). Predicting specificity-determining residues in two large eukaryotic transcription factor families. Nucleic Acids Res. *33*, 4455–4465.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. Lancet. Infect. Dis. *3099*, 19–20.

Dowall, S.D., Bewley, K., Watson, R.J., Vasan, S.S., Ghosh, C., Konai, M.M., Gausdal, G., Lorens, J.B., Long, J., Barclay, W., et al. (2016). Antiviral screening of multiple compounds against ebola virus. Viruses *8*, 1–17.

Edwards, M.R., Liu, G., Mire, C.E., Sureshchandra, S., Luthra, P., Yen, B., Shabman, R.S., Leung, D.W., Messaoudi, I., Geisbert, T.W., et al. (2016). Differential regulation of interferon responses by Ebola and Marburg virus VP35 proteins. Cell Rep. *14*, 1–21.

Falasca, L., Agrati, C., Petrosillo, N., Di Caro, A., Capobianchi, M.R., Ippolito, G., and Piacentini, M. (2015). Molecular mechanisms of Ebola virus pathogenesis: focus on cell death. Cell Death Differ. *22*, 1250–1259.

Feagins, A.R., and Basler, C.F. (2015). Amino Acid Residue at Position 79 of Marburg Virus VP40 Confers Interferon Antagonism in Mouse Cells. J. Infect. Dis. *212*, S219–S225.

Feldmann, H., and Geisbert, T.W. (2011). Ebola haemorrhagic fever. Lancet *377*, 849–862.

Feldmann, H., Mühlberger, E., Randolf, A., Will, C., Kiley, M.P., Sanchez, A., and Klenk, H.D. (1992). Marburg virus, a filovirus: méssenger RNAs, gene order, and regulatory elements of the replication cycle. Virus Res. *24*, 1–19.

Felt, S.A., Moerdyk-Schauwecker, M.J., and Grdzelishvili, V.Z. (2015). Induction of apoptosis in pancreatic cancer cells by vesicular stomatitis virus. Virology *474*, 713–724.

Fernandes, J.D., Jayaraman, B., and Frankel, A.D. (2012). The HIV-1 Rev response element: An RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. RNA Biol. *9*, 6–11.

Filippova, M., Song, H., Connolly, J.L., Dermody, T.S., and Duerksen-Hughes, P.J. (2002). The human papillomavirus 16 E6 protein binds to tumor necrosis factor (TNF) R1 and protects cells from TNF-induced apoptosis. J. Biol. Chem. *277*, 21730–21739.

Fine, P., Eames, K., and Heymann, D.L. (2011). "Herd immunity": A rough guide. Clin. Infect. Dis. *52*, 911–916.

Fischer, M.G. (2016). Giant viruses come of age. Curr. Opin. Microbiol. *31*, 50–57.

Forterre, P. (2016). To be or not to be alive: How recent discoveries challenge the traditional definitions of viruses and life. Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci. *59*, 100–108.

Forterre, P., and Gaïa, M. (2016). Giant viruses and the origin of modern eukaryotes. Curr. Opin. Microbiol. *31*, 44–49.

Freeman, D., Waite, F., Rosebrock, L., Petit, A., Causier, C., East, A., Jenner, L.,

Teale, A.L., Carr, L., Mulhall, S., et al. (2020). Coronavirus Conspiracy Beliefs, Mistrust, and Compliance with Government Guidelines in England. Psychol. Med.

Fung, I.C.-H., Duke, C.H., Finch, K.C., Snook, K.R., Tseng, P.-L., Hernandez, A.C., Gambhir, M., Fu, K.-W., and Tse, Z.T.H. (2016). Ebola virus disease and social media: A systematic review. Am. J. Infect. Control *44*, 1660–1671.

Galassi, F.M., Habicht, M.E., and Rühli, F.J. (2017). Poliomyelitis in Ancient Egypt? Neurol. Sci. *38*, 375.

Gale, P., Simons, R.R.L., Horigan, V., Snary, E.L., Fooks, A.R., and Drew, T.W. (2016). The challenge of using experimental infectivity data in risk assessment for Ebola virus: Why ecology may be important. J. Appl. Microbiol. *120*, 17–28.

Geisbert, T.W. (2017). First Ebola virus vaccine to protect human beings? Lancet *389*, 479–480.

Gittings, K., and Matson, K.L. (2016). Establishing herd immunity against Ebola through vaccination. Vaccine *34*, 2644–2647.

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., et al. (2018). The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. Nat. Microbiol.

Grayson, P., and Molineux, I.J. (2007). Is phage DNA "injected" into cells - biologists and physicists can agree. Curr. Opin. Microbiol. *10*, 401–409.

Grifoni, A., Lo Presti, A., Giovanetti, M., Montesano, C., Amicosante, M., Colizzi, V., Lai, A., Zehender, G., Cella, E., Angeletti, S., et al. (2016). Genetic diversity in Ebola virus: Phylogenetic and in silico structural studies of Ebola viral proteins. Asian Pac. J. Trop. Med. *9*, 337–343.

Groseth, A., Feldmann, H., and Strong, J.E. (2007). The ecology of Ebola virus. Trends Microbiol. *15*, 408–416.

Gruber, A.R., Findeiß, S., Washietl, S., Hofacker, I.L., and Stadler, P.F. (2010). RNAZ 2.0: Improved noncoding RNA detection. Pacific Symp. Biocomput. 2010, PSB 2010 69–79.

Gu, W., Li, M., Xu, Y., Wang, T., Ko, J.H., and Zhou, T. (2014). The impact of RNA

structure on coding sequence evolution in both bacteria and eukaryotes. BMC Evol. Biol. *14*, 1–10.

Le Guenno, B., Formenty, P., Wyers, M., Gounon, P., Walker, F., and Boesch, C. (1995). Isolation and partial characterisation of a new strain of Ebola virus. Lancet *345*, 1271–1274.

Guerra, F.M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S.L., Li, Y., and Crowcroft, N.S. (2017). The basic reproduction number (R0) of measles: a systematic review. Lancet Infect. Dis. *17*, e420–e428.

Guidry, J.P.D., Jin, Y., Orr, C.A., Messner, M., and Meganck, S. (2017). Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement. Public Relat. Rev. *43*, 477–486.

Guito, J.C., Albariño, C.G., Chakrabarti, A.K., and Towner, J.S. (2017). Novel activities by ebolavirus and marburgvirus interferon antagonists revealed using a standardized in vitro reporter system. Virology *501*, 147–165.

Guo, H., Kaiser, W.J., and Mocarski, E.S. (2015). Manipulation of Apoptosis and Necroptosis Signaling by Herpesvirus. Med. Microbiol. Immunol. *204*, 1–21.

Han, Z., Boshra, H., Sunyer, J.O., Zwiers, S.H., Paragas, J., and Harty, R.N. (2003). Biochemical and Functional Characterization of the Ebola Virus VP24 Protein: Implications for a Role in Virus Assembly and Budding. J. Virol. *77*, 1793.

Han, Z., Sagum, C.A., Bedford, M.T., Sidhu, S.S., Sudol, M., and Harty, R.N. (2016). ITCH E3 Ubiquitin Ligase Interacts with Ebola Virus VP40 To Regulate Budding. J. Virol. *90*, 9163–9171.

Harden, M.E., and Munger, K. (2017). Human papillomavirus molecular biology. Mutat. Res. - Rev. Mutat. Res. *772*, 3–12.

Harris, J.R. (2015). Transmission electron microscopy in molecular structural biology: A historical survey. Arch. Biochem. Biophys. *581*, 3–18.

Hayashi, T., Hotta, H., Itoh, M., and Homma, M. (1991). Protection of mice by a protease inhibitor, aprotinin, against lethal Sendai virus pneumonia. J. Gen. Virol. *72*, 979–982.

He, F., Melén, K., Maljanen, S., Lundberg, R., Jiang, M., Österlund, P., Kakkola, L.,

and Julkunen, I. (2017). Ebolavirus protein VP24 interferes with innate immune responses by inhibiting interferon-λ1 gene expression. Virology *509*, 23–34.

van Helvoort, T. (1991). What is a Virus? The Case of Tobacco Mosaic Disease. Stud. Hist. Philos. Sci. *22*, 557–588.

Henao-Restrepo, A.M., Preziosi, M.P., Wood, D., Moorthy, V., and Kieny, M.P. (2016). On a path to accelerate access to Ebola vaccines: The WHO's research and development efforts during the 2014-2016 Ebola epidemic in West Africa. Curr. Opin. Virol. *17*, 138–144.

Henao-Restrepo, A.M., Camacho, A., Longini, I.M., Watson, C.H., Edmunds, W.J., Egger, M., Carroll, M.W., Dean, N.E., Diatta, I., Doumbia, M., et al. (2017). Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!). Lancet *389*, 505–518.

Henikoff, S., and Henikoff, J.G. (2000). Amino acid substitution matrices. Adv. Protein Chem. *54*, 73–97.

Heppner, D.G., Kemp, T.L., Martin, B.K., Ramsey, W.J., Nichols, R., Dasen, E.J., Link, C.J., Das, R., Xu, Z.J., Sheldon, E.A., et al. (2017). Safety and immunogenicity of the rVSVΔG-ZEBOV-GP Ebola virus vaccine candidate in healthy adults: a phase 1b randomised, multicentre, double-blind, placebo-controlled, dose-response study. Lancet Infect. Dis. *17*, 854–866.

Heurich, A., Hofmann-Winkler, H., Gierer, S., Liepold, T., Jahn, O., and Pohlmann, S. (2014). TMPRSS2 and ADAM17 Cleave ACE2 Differentially and Only Proteolysis by TMPRSS2 Augments Entry Driven by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein. J. Virol. *88*, 1293–1307.

Hiraku S, Muryobayashi K, Ito H, Inagawa T, T.M. (1982). Absorption and excretion of camostat (FOY-305) orally administered to male rabbit and healthy subjects (English Abstract). Iyaku Kenkyu *13*, 756–765.

Hirota, M., Shimosegawa, T., Kitamura, K., Takeda, K., Takeyama, Y., Mayumi, T., Ito, T., Takenaka, M., Iwasaki, E., Sawano, H., et al. (2020). Continuous regional arterial infusion versus intravenous administration of the protease inhibitor nafamostat mesilate for predicted severe acute pancreatitis: a multicenter,

randomized, open-label, phase 2 trial. J. Gastroenterol. *55*, 342–352.

Hoehl, S., Berger, A., Kortenbusch, M., Cinatl, J., Bojkova, D., Rabenau, H., Behrens, P., Böddinghaus, B., Götsch, U., Naujoks, F., et al. (2020). Evidence of SARS-CoV-2 Infection in Returning Travelers from Wuhan, China. N. Engl. J. Med.

Hoenen, T., Groseth, A., and Feldmann, H. (2019). Therapeutic strategies to target the Ebola virus life cycle. Nat. Rev. Microbiol. *17*, 593–606.

Hofacker, I.L. (2006). RNAs everywhere: genome-wide annotation of structured RNAs. Genome Inform. *17*, 281–282.

Hoffman, M., H., K.-W., S., S., N., K., T., H., S., E., T.S., S., G., H., N.-H., W., A., N., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell.

Holzmann, H., Hengel, H., Tenbusch, M., and Doerr, H.W. (2016). Eradication of measles: remaining challenges. Med. Microbiol. Immunol. *205*, 201–208.

Hu, L., Crawford, S.E., Hyser, J.M., Estes, M.K., and Prasad, B.V.V. (2013). Rotavirus non-structural proteins : Structure and Function. Curr. Opin. Virol. *2*, 380–388.

Hume, A.J., and Mühlberger, E. (2019). Distinct Genome Replication and Transcription Strategies within the Growing Filovirus Family. J. Mol. Biol. *431*, 4290–4320.

Huo, X., Shi, G., Li, X., Lai, X., Deng, L., Xu, F., Chen, M., Wei, Q., Samba, T., and Liang, X. (2016). Knowledge and attitudes about Ebola vaccine among the general population in Sierra Leone. Vaccine *34*, 1767–1772.

Ian Brierley, Paul Digard, and S.C.I. (2020). Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell *395*, 1315.

Ikemura, K., Hiramatsu, S., and Okuda, M. (2017). Drug repositioning of proton pump inhibitors for enhanced efficacy and safety of cancer chemotherapy. Front. Pharmacol. *8*, 1–5.

Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M., and Nagata, N. (2019). TMPRSS2 Contributes to Virus Spread and Immunopathology in

the Airways of Murine Models after Coronavirus Infection. J. Virol. *93*, 1–15.

Jacobs, M., Rodger, A., Bell, D.J., Bhagani, S., Cropley, I., Filipe, A., Gifford, R.J., Hopkins, S., Hughes, J., Jabeen, F., et al. (2016). Late Ebola virus relapse causing meningoencephalitis: a case report. Lancet (London, England) *388*, 498–503.

John, S.P., Wang, T., Steffen, S., Longhi, S., Schmaljohn, C.S., and Jonsson, C.B. (2007). Ebola virus VP30 is an RNA binding protein. J. Virol. *81*, 8967–8976.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Res. *36*, 5–9.

Judson, S.D., Fischer, R., Judson, A., and Munster, V.J. (2016). Ecological Contexts of Index Cases and Spillover Events of Different Ebolaviruses. PLoS Pathog. *12*, 1–17.

Jun, S.R., Leuze, M.R., Nookaew, I., Uberbacher, E.C., Land, M., Zhang, Q., Wanchai, V., Chai, J., Nielsen, M., Trolle, T., et al. (2015). Ebolavirus comparative genomics. FEMS Microbiol. Rev. *39*, 764–778.

Kagina, B.M., Wiysonge, C.S., Lesosky, M., Madhi, S.A., and Hussey, G.D. (2014). Safety of licensed vaccines in HIV-infected persons: A systematic review protocol. Syst. Rev. *3*, 1–6.

Kalinina, O. V. (2004). Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families. Protein Sci. *13*, 443–456.

Kalinina, O. V., Gelfand, M.S., and Russell, R.B. (2009). Combining specificity determining and conserved residues improves functional site prediction. BMC Bioinformatics *10*, 1–24.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. (2018). Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. *46*, D335–D342.

Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., Kubo, H., and Makino, S. (2006). Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA

degradation. Proc. Natl. Acad. Sci. U. S. A. *103*, 12885–12890.

Kampf, G., Todt, D., Pfaender, S., and Steinmann, E. (2020). Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. J. Hosp. Infect. *104*, 246–251.

Kapadia, B.H., Torre, B.B., Ullman, N., Yang, A., Harb, M.A., Grieco, P.W., Newman, J.M., Harwin, S.F., and Maheshwari, A. V. (2019). Reducing perioperative blood loss with antifibrinolytics and antifibrinolytic-like agents for patients undergoing total hip and total knee arthroplasty. J. Orthop. *16*, 513–516.

Katz, L., and Burge, C.B. (2003). Widespread selection for local RNA secondary structure in coding regions of bacterial genes. Genome Res. *13*, 2042–2051.

Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F., and Matsuyama, S. (2012). Simultaneous Treatment of Human Bronchial Epithelial Cells with Serine and Cysteine Protease Inhibitors Prevents Severe Acute Respiratory Syndrome Coronavirus Entry. J. Virol. *86*, 6537–6545.

Kelly, L.A., Mezulis, S., Yates, C., Wass, M., and Sternberg, M. (2015). The Phyre2 web portal for protein modelling, prediction, and analysis. Nat. Protoc. *10*, 845–858.

Kennedy, S.B., Bolay, F., Kieh, M., Grandits, G., Badio, M., Ballou, R., Eckes, R., Feinberg, M., Follmann, D., Grund, B., et al. (2017). Phase 2 Placebo-Controlled Trial of Two Vaccines to Prevent Ebola in Liberia. N. Engl. J. Med. *377*, 1438–1447.

Van Kerkhove, M.D., Bento, A.I., Mills, H.L., Ferguson, N.M., and Donnelly, C.A. (2015). A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making. Sci. Data *2*, 150019.

Keshwara, R., Johnson, R.F., and Schnell, M.J. (2017). Toward an Effective Ebola Virus Vaccine. Annu. Rev. Med. *68*, 371–386.

Kiening, M., Ochsenreiter, R., Hellinger, H.J., Rattei, T., Hofacker, I., and Frishman, D. (2019). Conserved secondary structures in viral mRNAs. Viruses *11*, 1–22.

Kirchdoerfer, R.N., Abelson, D.M., Li, S., Wood, M.R., and Saphire, E.O. (2015). Assembly of the Ebola Virus Nucleoprotein from a Chaperoned VP35 Complex. Cell Rep. *12*, 140–149.

Kobasa, D., Embury-hyatt, C., and Kobinger, G.P. (2016). Development and

Characterization of a Guinea Pig-Adapted Sudan Virus. J. Virol. *90*, 392–399.

Koehler, A., Kolesnikova, L., and Becker, S. (2016). An active site mutation increases the polymerase activity of the Guinea pig-lethal Marburg virus. J. Gen. Virol. *97*, 2494–2500.

Koehler, A., Pfeiffer, S., Kolesnikova, L., and Becker, S. (2018). Analysis of the multifunctionality of Marburg virus VP40. J. Gen. Virol. *99*, 1614–1620.

Komaroff, A.L. (2006). Is human herpesvirus-6 a trigger for chronic fatigue syndrome? J. Clin. Virol. *37 Suppl 1*.

Koonin, E. V., and Starokadomskyy, P. (2016). Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci. *59*, 125–134.

Kortepeter, M.G., Bausch, D.G., and Bray, M. (2011). Basic clinical and Laboratory features of filoviral hemorrhagic fever. J. Infect. Dis. *204*.

Kozomara, A., and Griffiths-Jones, S. (2014). MiRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*, 68–73.

Kramer, A.M., Pulliam, J.T., Alexander, L.W., Park, A.W., Rohani, P., and Drake, J.M. (2016). Spatial spread of the West Africa Ebola epidemic. Dryad Digit. Repos. *3*.

Krauer, F., Gsteiger, S., Low, N., Hansen, C.H., and Althaus, C.L. (2016). Heterogeneity in District-Level Transmission of Ebola Virus Disease during the 2013-2015 Epidemic in West Africa. PLoS Negl. Trop. Dis. *10*.

Ksiazek, T.G., Rota, P.A., and Rollin, P.E. (2011). A review of Nipah and Hendra viruses with an historical aside. Virus Res. *162*, 173–183.

Kucharski, A.J., Eggo, R.M., Watson, C.H., Camacho, A., Funk, S., and Edmunds, W.J. (2016). Effectiveness of ring vaccination as control strategy for Ebola virus disease. Emerg. Infect. Dis. *22*, 105–108.

Kuhn, J.H., Becker, S., Geisbert, T.W., Johnson, K.M., Nichol, S.T., and Peters, C.J. (2011). Proposal for a revised taxonomy of the family Filoviridae: classification, names of taxa and viruses, and virus abbreviations. Arch Virol *155*, 2083–2103.

Kuroda, M., Fujikura, D., Nanbo, A., Marzi, A., Noyori, O., Kajihara, M., Maruyama, J., Matsuno, K., Miyamoto, H., Yoshida, R., et al. (2015). Interaction between TIM-1

and NPC1 Is Important for Cellular Entry of Ebola Virus. J. Virol. *89*, 6481–6493.

Kurth, R., and Bannert, N. (2010). Beneficial and detrimental effects of human endogenous retroviruses. Int. J. Cancer *126*, 306–314.

de La Vega, M.-A., Wong, G., Kobinger, G.P., and Qiu, X. (2015). The Multiple Roles of sGP in Ebola Pathogenesis. Viral Immunol. *28*, 3–9.

Ladner, J.T., Wiley, M.R., Mate, S., Dudas, G., Prieto, K., Lovett, S., Nagle, E.R., Beitzel, B., Gilbert, M.L., Fakoli, L., et al. (2015). Evolution and Spread of Ebola Virus in Liberia, 2014-2015. Cell Host Microbe *18*, 659–669.

Lambe, T., Bowyer, G., and Ewer, K.J. (2017). A review of phase I trials of Ebola virus vaccines: What can we learn from the race to develop novel vaccines? Philos. Trans. R. Soc. B Biol. Sci. *372*.

Lamberti, P.W., and Majtey, A.P. (2003). Non-logarithmic Jensen-Shannon divergence. Phys. A Stat. Mech. Its Appl. *329*, 81–90.

Lau, M.S.Y., Dalziel, B.D., Funk, S., McClelland, A., Tiffany, A., Riley, S., Metcalf, C.J.E., and Grenfell, B.T. (2017). Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic. Proc. Natl. Acad. Sci. *114*, 2337–2342.

Ledgerwood, J.E. (2015). Use of low dose rVSV-ZEBOV: Safety issues in a Swiss cohort. Lancet Infect. Dis. *15*, 1117–1119.

Ledgerwood, J.E., DeZure, A.D., Stanley, D.A., Coates, E.E., Novik, L., Enama, M.E., Berkowitz, N.M., Hu, Z., Joshi, G., Ploquin, A., et al. (2017). Chimpanzee Adenovirus Vector Ebola Vaccine. N. Engl. J. Med. *376*, 928–938.

Lee, C., Whetten, K., Omer, S., Pan, W., and Salmon, D. (2016). Hurdles to herd immunity: Distrust of government and vaccine refusal in the US, 2002–2003. Vaccine *34*, 3972–3978.

Lee, J.E., Road, T.P., and Jolla, L. (2010). Ebolavirus glycoprotein structure and mechanism of entry. Future Virol. *4*, 621–635.

Legrand, J., Grais, R.F., Boelle, P.Y., Valleron, A.J., and Flahault, A. (2007). Understanding the dynamics of Ebola epidemics. Epidemiol. Infect. *135*, 610–621.

Leligdowicz, A., Fischer, W.A., Uyeki, T.M., Fletcher, T.E., Adhikari, N.K.J.,

Portella, G., Lamontagne, F., Clement, C., Jacob, S.T., Rubinson, L., et al. (2016). Ebola virus disease and critical illness. Crit. Care *20*, 217.

Letko, M., Marzi, A., and Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat. Microbiol. *5*.

Levy JH, Bailey JM, S.M. (1994). Pharmacokinetics of aprotinin in preoperative cardiac surgical patients. Anesthesiology *80*, 1013–1018.

Li, J.X., Hou, L.H., Meng, F.Y., Wu, S.P., Hu, Y.M., Liang, Q., Chu, K., Zhang, Z., Xu, J.J., Tang, R., et al. (2017a). Immunity duration of a recombinant adenovirus type-5 vector-based Ebola vaccine and a homologous prime-boost immunisation in healthy adults in China: final report of a randomised, double-blind, placebo-controlled, phase 1 trial. Lancet Glob. Heal. *5*, e324–e334.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., et al. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N. Engl. J. Med. 1199–1207.

Li, T., Yao, H.-W., Liu, D., Ren, H.-G., Hu, Y., Kargbo, D., Teng, Y., Deng, Y.-Q., Lu, H.-J., Liu, X., et al. (2017b). Mapping the clinical outcomes and genetic evolution of Ebola virus in Sierra Leone. JCI Insight *2*, 1–10.

Li, W., Zhang, C., Sui, J., Kuhn, J.H., Moore, M.J., Luo, S., Wong, S.K., Huang, I.C., Xu, K., Vasilieva, N., et al. (2005). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. *24*, 1634–1643.

Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, S., and 722 JL, Luzuriaga K, Greenough TC, Choe H, F.M. (2004). Angiotensin-converting enzyme 2: A functional receptor for SARS coronavirus. Cell. Mol. Life Sci. *61*, 2738–2743.

Liang, H.W., Zhou, Z., Zhang, S.Y., Zen, K., Chen, X., and Zhang, C.Y. (2014). Identification of Ebola virus microRNAs and their putative pathological function. Sci. China Life Sci. *57*, 973–981.

Lin, K., Ali, A., Rusere, L., Soumana, D.I., Yilmaz, N.K., and Schiffer, C.A. (2017). Dengue Virus NS2B/NS3 Protease Inhibitors Exploiting the Prime Side. J. Virol. *91*, 1–10.

Lipkin, W.I., and Anthony, S.J. (2015). Virus hunting. Virology *479–480*, 194–199.

Liu, Y., Sun, J., Zhang, H., Wang, M., Gao, G.F., and Li, X. (2016). Ebola virus encodes a miR-155 analog to regulate importin-α5 expression. Cell. Mol. Life Sci. *73*, 3733–3744.

Lo, T.Q., Marston, B.J., Dahl, B.A., and Cock, K.M. De (2017). Ebola: Anatomy of an Epidemic. Annu. Rev. Med. *68*, 359–370.

Lokugamage KG, Hage A, Schindewolf C, Rajsbaum R, M.V. (2020). SARS-CoV-2 sensitive to type I interferon pretreatment. BioRxiv 1–13.

Loring, S. (1939). Properties and hydrolytic products of nucleic acid from tobacco mosaic virus. J. Biol. Chem. *130*, 251–259.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395*, 565–574.

Luan, J., Lu, Y., Jin, X., and Zhang, L. (2020). Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. Biochem. Biophys. Res. Commun.

Luby, S.P. (2013). The pandemic potential of Nipah virus. Antiviral Res. *100*, 38–43.

Lucia Morales, Juan Carlos Oliveros, Raul Fernandez-Delgado, Benjamin Robert tenOever, Luis Enjuanes, and I.S. (2017). SARS-CoV-encoded small RNAs contribute to infection-associated lung pathology. Cell Host Microbe *21*, 344–355.

Lupiani, B., and Reddy, S.M. (2009). The history of avian influenza. Comp. Immunol. Microbiol. Infect. Dis. *32*, 311–323.

Lurie, N., Saville, M., Hatchett, R., and Halton, J. (2020). Developing Covid-19 Vaccines at Pandemic Speed. N. Engl. J. Med. 1–2.

Lustig, A., and Levine, A.J. (1992). One hundred years of virology. J. Virol. *66*, 4629–4631.

Luthra, P., Jordan, D.S., Leung, D.W., Amarasinghe, G.K., and Basler, C.F. (2015). Ebola Virus VP35 Interaction with Dynein LC8 Regulates Viral RNA Synthesis. J. Virol. *89*, 5148–5153.

Lwoff, A. (1957). The Concept of Virus. J. Gen. Microbiol. *17*, 239–253.

MacIntyre, C.R., and Chughtai, A.A. (2016). Recurrence and reinfection-a new paradigm for the management of Ebola virus disease. Int. J. Infect. Dis. *43*, 58–61.

Madhugiri, R., Fricke, M., Marz, M., and Ziebuhr, J. (2014). RNA structure analysis of alphacoronavirus terminal genome regions. Virus Res. *194*, 76–89.

Madhugiri, R., Karl, N., Petersen, D., Lamkiewicz, K., Fricke, M., Wend, U., Scheuer, R., Marz, M., and Ziebuhr, J. (2018). Structural and functional conservation of cis-acting RNA elements in coronavirus 5'-terminal genome regions. Virology *517*, 44–55.

Mafopa, N.G., Russo, G., Emeric Guetiya Wadoum, R., Emmanuel, I., Batwala, V., Giovanetti, M., Minutolo, A., Turay, P., Turay, T.B., Kargbo, B., et al. (2017). Seroprevalence of Ebola virus infection in Bombali District, Sierra Leone. J. Public Health Africa *8*, 23–29.

Maganga, G.D., Kapetshi, J., Berthet, N., Kebela Ilunga, B., Kabange, F., Mbala Kingebeni, P., Mondonge, V., Muyembe, J.-J.T., Bertherat, E., Briand, S., et al. (2014). Ebola Virus Disease in the Democratic Republic of Congo. N. Engl. J. Med. *371*, 2083–2091.

Mahanty, S., and Bray, M. (2004). Pathogenesis of filoviral haemorrhagic fevers. Lancet Infect. Dis. *4*, 487–498.

Majid, M.U., Tahir, M.S., Ali, Q., Rao, A.Q., Rashid, B., Ali, A., Ahmad Nasir, I., and Husnain, T. (2016). Nature and History of Ebola Virus: An Overview. Arch. Neurosci. *3*.

Majtey, A.P., Lamberti, P.W., and Prato, D.P. (2005). Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states. Phys. Rev. A - At. Mol. Opt. Phys. *72*, 1–14.

Marr, J.S., and Cathey, J.T. (2013). The 1802 saint-domingue yellow fever epidemic and the Louisiana purchase. J. Public Heal. Manag. Pract. *19*, 77–82.

Martell, H.J., Masterson, S.G., McGreig, J.E., Michaelis, M., and Wass, M.N. (2019). Is the Bombali virus pathogenic in humans? Bioinformatics 1–6.

Martínez-García, E., and de Lorenzo, V. (2016). The quest for the minimal bacterial

genome. Curr. Opin. Biotechnol. *42*, 216–224.

Martínez, M.J., Biedenkopf, N., Volchkova, V., Hartlieb, B., Alazard-Dany, N., Reynard, O., Becker, S., and Volchkov, V. (2008). Role of Ebola virus VP30 in transcription reinitiation. J. Virol. *82*, 12569–12573.

Marzi, A., Banadyga, L., Haddock, E., Thomas, T., Shen, K., Horne, E.J., Scott, D.P., Feldmann, H., and Ebihara, H. (2016). A hamster model for Marburg virus infection accurately recapitulates Marburg hemorrhagic fever. Nat. Publ. Gr. 1–14.

Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M., and Taguchi, F. (2010). Efficient Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by the Transmembrane Protease TMPRSS2. J. Virol. *84*, 12658–12664.

Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., Nagata, N., Sekizuka, T., Katoh, H., Kato, F., et al. (2020). Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. Proc. Natl. Acad. Sci. 202002589.

Matua, G.A., Van der Wal, D.M., and Locsin, R.C. (2015). Ebola hemorrhagic fever outbreaks: Strategies for effective epidemic management, containment and control. Brazilian J. Infect. Dis. *19*, 308–313.

McCrary, M.L., Severson, J., and Tyring, S.K. (1999). Varicella zoster virus. J. Am. Acad. Dermatol. *41*, 1–16.

McMichael, A.J. (2004). Environmental and social influences on emerging infectious diseases: past, present and future. Philos. Trans. R. Soc. B Biol. Sci. *359*, 1049–1058.

Medaglini, D., and Siegrist, C.A. (2017). Immunomonitoring of human responses to the rVSV-ZEBOV Ebola vaccine. Curr. Opin. Virol. *23*, 88–94.

Mehedi, M., Falzarano, D., Seebach, J., Hu, X., Carpenter, M.S., Schnittler, H.-J., and Feldmann, H. (2011). A New Ebola Virus Nonstructural Glycoprotein Expressed through RNA Editing. J. Virol. *85*, 5406–5414.

Metcalf, C.J.E., Ferrari, M., Graham, A.L., and Grenfell, B.T. (2015). Understanding Herd Immunity. Trends Immunol. *36*, 753–755.

Méthot, P.O. (2016). Writing the history of virology in the twentieth century: Discovery, disciplines, and conceptual change. Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci. *59*, 145–153.

Michaelis, M., Rossman, J.S., Wass, M., Brown, J.A.L., Bourke, E., Eriksson, L.A., Kerin, M.J., Hayes, S., Malacrida, B., Kiely, M., et al. (2016). Computational analysis of Ebolavirus data: prospects, promises and challenges. Biochem Soc Trans *44*.

Michaelis, M., Kleinschmidt, M.C., Bojkova, D., Rabenau, H.F., Wass, M.N., and Cinatl, J. (2019). Omeprazole Increases the Efficacy of Acyclovir Against Herpes Simplex Virus Type 1 and 2. Front. Microbiol. *10*, 1–7.

Minakata, D., Fujiwara, S.I., Hayakawa, J., Nakasone, H., Ikeda, T., Kawaguchi, S.I., Toda, Y., Ito, S., Ochi, S.I., Nagayama, T., et al. (2019). Comparison of Danaparoid Sodium and Synthetic Protease Inhibitors for the Treatment of Disseminated Intravascular Coagulation Associated with Hematological Malignancies: A Retrospective Analysis. Acta Haematol. *Aug 28*, 1–10.

Miranda, M.E.G., and Miranda, N.L.J. (2011). Reston ebolavirus in Humans and Animals in the Philippines: A Review. J. Infect. Dis. *204*, 757–760.

Moggia, E., Koti, R., Belgaumkar, A.P., Fazio, F., Pereira, S.P., Davidson, B.R., and Gurusamy, K.S. (2017). Pharmacological interventions for acute pancreatitis. Cochrane Database Syst. Rev. *2017*.

Moon, S., Sridhar, D., Pate, M.A., Jha, A.K., Clinton, C., Delaunay, S., Edwin, V., Fallah, M., Fidler, D.P., Garrett, L., et al. (2015). Will Ebola change the game? Ten essential reforms before the next pandemic. the report of the Harvard-LSHTM Independent Panel on the Global Response to Ebola. Lancet *386*, 2204–2221.

Moore, Z.S., Seward, J.F., and Lane, J.M. (2006). Smallpox. Lancet *367*, 425–435.

Morgan, G.J. (2016). What is a virus species? Radical pluralism in viral taxonomy. Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci. *59*, 64–70.

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. J Immunol Methods. *Dec 16*, 55–63.

Mühlberger, E., Trommer, S., Funke, C., Volchkov, V., Klenk, H.D., and Becker, S. (1996). Termini of all mRNA species of Marburg virus: Sequence and secondary structure. Virology *223*, 376–380.

Mukherjee, S., Majumdar, S., Vipat, V.C., Mishra, A.C., and Chakrabarti, A.K. (2012).

Non structural protein of avian influenza A (H11N1) virus is a weaker suppressor of immune responses but capable of inducing apoptosis in host cells. Virol. J. *9*, 1.

Muth, T., García-martín, J.A., Rausell, A., Juan, D., Valencia, A., and Pazos, F. (2012). JDet: Interactive calculation and visualization of function-related conservation patterns in multiple sequence alignments and structures. Bioinformatics *28*, 584–586.

N. Zhu, et al. (2020). China Novel Coronavirus Investigating and Research Team. N. Engl. J. Med. *Feb 20;382*, 727–733.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933–2935.

Negredo, A., Palacios, G., Vázquez-Morón, S., González, F., Dopazo, H., Molero, F., Juste, J., Quetglas, J., Savji, N., de la Cruz Martínez, M., et al. (2011). Discovery of an ebolavirus-like filovirus in europe. PLoS Pathog. *7*, 1–8.

Nieto-Torres, J.L., Verdiá-Báguena, C., Jimenez-Guardeño, J.M., Regla-Nava, J.A., Castaño-Rodriguez, C., Fernandez-Delgado, R., Torres, J., Aguilella, V.M., and Enjuanes, L. (2015). Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome. Virology *485*, 330–339.

Nikiforov, V. V, Turovskiĭ, I.I., Kalinin, P.P., Akinfeeva, L., Katkova, L.R., Barmin, V.S., Ryabchikova, E., Popkova, N.I., Shestopalov, A.M., and Nazarov, V.P. (1994). A case of a laboratory infection with Marburg fever. Zh. Mikrobiol. Epidemiol. Immunobiol. *3*, 104–106.

Nishiura, Linton, and Akhmetzhanov (2020a). Initial Cluster of Novel Coronavirus (2019-nCoV) Infections in Wuhan, China Is Consistent with Substantial Human-to-Human Transmission. J. Clin. Med. *9*, 488.

Nishiura, Kobayashi, Yang, Hayashi, Miyama, Kinoshita, Linton, Jung, Yuan, Suzuki, et al. (2020b). The Rate of Underascertainment of Novel Coronavirus (2019-nCoV) Infection: Estimation Using Japanese Passengers Data on Evacuation Flights. J. Clin. Med. *9*, 419.

O'Cofaigh, E., and Lewthwaite, P. (2013). Natural history of HIV and AIDS. Med. (United Kingdom) *41*, 411–416.

Ochsenreiter, R., Hofacker, I.L., and Wolfinger, M.T. (2019). Functional RNA structures in the 3'UTR of tick-borne, insect-specific and no-known-vector flaviviruses. Viruses *11*.

Ofran, Y., and Rost, B. (2003). Analysing six types of protein-protein interfaces. J. Mol. Biol. *325*, 377–387.

Olejnik, J., Ryabchikova, E., Corley, R.B., and Mühlberger, E. (2011). Intracellular events and cell fate in filovirus infection. Viruses *3*, 1501–1531.

Olival, K.J., and Hayman, D.T.S. (2014). Filoviruses in bats: Current knowledge and future directions. Viruses *6*, 1759–1788.

Olivier, C., Poirier, G., Gendron, P., Boisgontier, A., Major, F., and Chartrand, P. (2005). Identification of a Conserved RNA Motif Essential for She2p Recognition and mRNA Localization to the Yeast Bud. Mol. Cell. Biol. *25*, 4752–4766.

Olson, J.K., and Grose, C. (1997). Endocytosis and Recycling of Varicella-Zoster Virus Fc Receptor Glycoprotein gE: Internalization Mediated by a YXXL Motif in the Cytoplasmic Tail. J. Virol. *71*, 4042–4054.

Onafuye, H., Pieper, S., Mulac, D., Cinatl, J., Wass, M.N., Langer, K., and Michaelis, M. (2019). Doxorubicin-loaded human serum albumin nanoparticles overcome transporter-mediated drug resistance in drug-adapted cancer cells. Beilstein J. Nanotechnol. *10*, 1707–1715.

Osterholm, M.T., Kelley, N.S., Sommer, A., and Belongia, E.A. (2012). Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. Lancet.Infectious Dis. *12*, 36–44.

Osterholm, M.T., Moore, K. a, Kelley, N.S., Brosseau, L.M., Wong, G., Murphy, F. a, Peters, C.J., Leduc, J.W., Russell, P.K., Herp, M. Van, et al. (2015). Transmission of Ebola viruses: what we know and what we do not know. MBio *6*, 1–9.

Pan, X., Chen, D., Xia, Y., Wu, X., Li, T., Ou, X., Zhou, L., and Liu, J. (2020). Asymptomatic cases in a family cluster with SARS-CoV-2 infection. Lancet. Infect. Dis. *20*, 410–411.

Pandey, A., Atkins, K.E., Medlock, J., Wenzel, N., Townsend, J.P., Childs, J.E., Nyenswah, T.G., Ndeffo-Mbah, M.L., and Galvani, A.P. (2014). Strategies for

containing Ebola in West Africa. Science (80-.). *346*, 991–995.

Pappalardo, M., Juliá, M., Howard, M.J., Rossman, J.S., Michaelis, M., and Wass, M.N. (2016). Conserved differences in protein sequence determine the human pathogenicity of Ebolaviruses. Sci. Rep. *6*, 23743.

Pappalardo, M., Reddin, I.G., Cantoni, Di., Rossman, J.S., Michaelis, M., and Wass, M.N. (2017a). Changes associated with Ebola virus adaptation to novel species. Bioinformatics *33*, 1911–1915.

Pappalardo, M., Collu, F., Macpherson, J., Michaelis, M., Fraternali, F., and Wass, M.N. (2017b). Investigating Ebola virus pathogenicity using molecular dynamics. BMC Genomics *18*, 566.

Paulke-Korinek, M., Kundi, M., Rendi-Wagner, P., de Martin, A., Eder, G., Schmidle-Loss, B., Vecsei, A., and Kollaritsch, H. (2011). Herd immunity after two years of the universal mass vaccination program against rotavirus gastroenteritis in Austria. Vaccine *29*, 2791–2796.

Pavot, V. (2016). Ebola virus vaccines: Where do we stand? Clin. Immunol. *173*, 44–49.

Perez, J.T., Varble, A., Sachidanandam, R., Zlatev, I., Manoharan, M., García-Sastre, A., and TenOever, B.R. (2010). Influenza A virus-generated small RNAs regulate the switch from transcription to replication. Proc. Natl. Acad. Sci. U. S. A. *107*, 11525–11530.

Peterson, A.T., and Samy, A.M. (2016). Geographic potential of disease caused by Ebola and Marburg viruses in Africa. Acta Trop. *162*, 114–124.

Pickett, B.E., Greer, D.S., Zhang, Y., Stewart, L., Zhou, L., Sun, G., Gu, Z., Kumar, S., Zaremba, S., Larsen, C.N., et al. (2012a). Virus pathogen Database and Analysis Resource (ViPR): A comprehensive bioinformatics Database and Analysis Resource for the Coronavirus research community. Viruses *4*, 3209–3226.

Pickett, B.E., Sadat, E.L., Zhang, Y., Noronha, J.M., Squires, R.B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., et al. (2012b). ViPR: An open bioinformatics database and analysis resource for virology research. Nucleic Acids Res. *40*, 593–598.

Pires, D.E.V., Ascher, D.B., and Blundell, T.L. (2014). MCSM: Predicting the effects

of mutations in proteins using graph-based signatures. Bioinformatics *30*, 335–342.

Plans-Rubió, P. (2012). The vaccination coverage required to establish herd immunity against influenza viruses. Prev. Med. (Baltim). *55*, 72–77.

Plotkin, S. (2014). History of vaccination. Proc. Natl. Acad. Sci. *111*, 12283–12287.

Plotkin, S.A. (1980). Rabies vaccine prepared in human cell cultures: progress and perspectives. Rev Infect Dis *2*, 433–448.

Pourrut, X., Kumulungui, B., Wittmann, T., Moussavou, G., Délicat, A., Yaba, P., Nkoghe, D., Gonzalez, J.P., and Leroy, E.M. (2005). The natural history of Ebola virus in Africa. Microbes Infect. *7*, 1005–1014.

Pradeu, T., Kostyrka, G., and Dupre, J. (2016). Understanding viruses: Philosophical investigations. Stud. Hist. Philos. Sci. *59*, 57–63.

Prasad, A.N., Ronk, A.J., Widen, S.G., Wood, T.G., Basler, C.F., and Bukreyev, A. (2019). Ebola Virus Produces Discrete Small Noncoding RNAs Independently of the Host MicroRNA Pathway Which Lack RNA Interference Activity in Bat and Human Cells. J. Virol. *94*, 1–33.

Qiu, X., Wong, G., Audet, J., Cutts, T., Niu, Y., Booth, S., and Kobinger, G.P. (2014). Establishment and Characterization of a Lethal Mouse Model for the Angola Strain of Marburg Virus. J. Virol. *88*, 12703–12714.

Qiu, Y., Zhao, Y.-B., Wang, Q., Li, J.-Y., Zhou, Z.-J., Liao, C.-H., and Ge, X.-Y. (2020). Predicting the angiotensin converting enzyme 2 (ACE2) utilizing capability as the receptor of SARS-CoV-2. Microbes Infect. *2*.

Ramsey, M.L., Nuttall, J., and Hart, P.A. (2019). A phase 1/2 trial to evaluate the pharmacokinetics, safety, and efficacy of NI-03 in patients with chronic pancreatitis: Study protocol for a randomized controlled trial on the assessment of camostat treatment in chronic pancreatitis (TACTIC). Trials *20*, 1–7.

Randolph, V.B., and Stollar, V. (1990). Low pH-induced cell fusion in flavivirus-infected Aedes albopictus cell cultures. J. Gen. Virol. *71*, 1845–1850.

Rashid, I. (2013). The Sierra Leone Civil War and the Remaking of ECOWAS. Res. Sierra Leone Stud. Weav. *1*, 1–21.

Rathore, M.H., Runyon, J., and Haque, T. ul (2017). Emerging Infectious Diseases.

Adv. Pediatr. *64*, 27–71.

Rausell, A., Juan, D., Pazos, F., and Valencia, A. (2010). Protein interactions and ligand binding : From protein subfamilies to functional specificity. PNAS *107*, 1995–2000.

Reinke, L.M., Spiegel, M., Plegge, T., Hartleib, A., Nehlmeier, I., Gierer, S., Hoffmann, M., Hofmann-Winkler, H., Winkler, M., and Pöhlmann, S. (2017). Different residues in the SARS-CoV spike protein determine cleavage and activation by the host cell protease TMPRSS2. PLoS One *12*, 1–15.

Reperant, L.A., and Osterhaus, A.D.M.E. (2017). AIDS, Avian flu, SARS, MERS, Ebola, Zika… what next? Vaccine *35*, 4470–4474.

Rewar, S., and Mirdha, D. (2014). Transmission of Ebola virus disease: An overview. Ann. Glob. Heal. *80*, 444–451.

Rice P, Longden I, B.A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. *Jun;16*, 276–277.

Rieck, T., Feig, M., an der Heiden, M., Siedler, A., and Wichmann, O. (2015). Assessing varicella vaccine effectiveness and its influencing factors using health insurance claims data. Euro Surveill *22*, 1–10.

Rivas, E., Clements, J., and Eddy, S.R. (2017). Lack of evidence for conserved secondary structure in long noncoding RNAs. Nat Methods *14*, 45–48.

Rojek, A., Horby, P., and Dunning, J. (2017). Insights from clinical research completed during the west Africa Ebola virus disease epidemic. Lancet Infect. Dis. *17*, e280–e292.

Ronny Lorenz, Stephan H Bernhart, Christian Hoener Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker.Ronny Lorenz, Stephan H Bernhart, Christian Hoener Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and I.L.H. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 122–128.

Rosello, A., Mossoko, M., Flasche, S., Van Hoek, A.J., Mbala, P., Camacho, A., Funk, S., Kucharski, A., Ilunga, B.K., Edmunds, W.J., et al. (2015). Ebola virus disease in the Democratic Republic of the Congo, 1976-2014. Elife *4*, 1–19.

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., et al. (2020). Transmission of 2019-NCOV infection from an asymptomatic contact in Germany. N. Engl. J. Med. *382*, 970–971.

Rougeron, V., Feldmann, H., Grard, G., Becker, S., and Leroy, E.M. (2015). Ebola and Marburg haemorrhagic fever. J. Clin. Virol. *64*, 111–119.

Rower, J.E., Meissner, E.G., Jimmerson, L.C., Osinusi, A., Sims, Z., Petersen, T., Bushman, L.R., Wolfe, P., McHutchison, J.G., Kottilil, S., et al. (2015). Serum and cellular ribavirin pharmacokinetic and concentration-effect analysis in HCV patients receiving sofosbuvir plus ribavirin. J. Antimicrob. Chemother. *70*, 2322–2329.

Saberi, A., Gulyaeva, A.A., Brubacher, J.L., Newmark, P.A., and Gorbalenya, A.E. (2018). A planarian nidovirus expands the limits of RNA genome size.

Sánchez-Sampedro, L., Perdiguero, B., Mejías-Pérez, E., García-Arriaza, J., Di Pilato, M., and Esteban, M. (2015). The evolution of poxvirus vaccines. Viruses *7*, 1726–1803.

Sanchez, A., and Rollin, P.E. (2005). Complete genome sequence of an Ebola virus (Sudan species) responsible for a 2000 outbreak of human disease in Uganda. Virus Res. *113*, 16–25.

Sanchez, A., Kiley, M.P., Holloway, B.P., and Auperin, D.D. (1993). Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. Virus Res. *29*, 215–240.

Sanjuán, R., and Domingo-Calap, P. (2016). Mechanisms of viral mutation. Cell. Mol. Life Sci. *73*, 4433–4448.

Schmidt, M.L., and Hoenen, T. (2017). Characterization of the catalytic center of the Ebola virus L polymerase. PLoS Negl. Trop. Dis. *11*, e0005996.

Schwarz, T.M., Edwards, M.R., Diederichs, A., Alinger, J.B., Leung, D.W., Amarasinghe, G.K., and Basler, C.F. (2017). VP24-Karyopherin Alpha Binding Affinities Differ between Ebolavirus Species, Influencing Interferon Inhibition and VP24 Stability. J. Virol. *91*, 1–16.

Shader, R.I. (2017). Prophylactic Vaccines, Successes, Errors, and Complications.

Clin. Ther. *39*, 1511–1514.

Sharma, V., Colson, P., Pontarotti, P., and Raoult, D. (2016). Mimivirus inaugurated in the 21st century the beginning of a reclassification of viruses. Curr. Opin. Microbiol. *31*, 16–24.

Sheahan, T.P., Sims, A.C., Leist, S.R., Schäfer, A., Won, J., Brown, A.J., Montgomery, S.A., Hogg, A., Babusis, D., Clarke, M.O., et al. (2020). Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. Nat. Commun. *11*.

Shen Wen, L., Juan, H., Ling, Y., Tanaka, Y., and Zhang, W. (2017). TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections. Biochimie *Nov*, 1–10.

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., et al. (2018). The evolutionary history of vertebrate RNA viruses. Nature *556*, 197–202.

Shifflett, K., and Marzi, A. (2019). Marburg virus pathogenesis - Differences and similarities in humans and animal models. Virol. J. *16*, 1–12.

Shin, J.M., and Kim, N. (2013). Pharmacokinetics and pharmacodynamics of the proton pump inhibitors. J. Neurogastroenterol. Motil. *19*, 25–35.

Shin, W.J., and Seong, B.L. (2017). Type II transmembrane serine proteases as potential target for anti-influenza drug discovery. Expert Opin. Drug Discov. *12*, 1139–1152.

Shiwani, H.A., Pharithi, R.B., Khan, B., Egom, C.B.-A., Kruzliak, P., Maher, V., and Egom, E.E.-A. (2017). An update on the 2014 Ebola outbreak in Western Africa. Asian Pac. J. Trop. Med. *10*, 6–10.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011a). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011b). Fast, scalable generation of

high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*.

Simmons, G., Zmora, P., Gierer, S., Heurich, A., and Pöhlmann, S. (2013). Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research. Antiviral Res. *Dec*, 605–614.

Sissoko, D., Duraffour, S., Kerber, R., Kolie, J.S., Beavogui, A.H., Camara, A.M., Colin, G., Rieger, T., Oestereich, L., Pályi, B., et al. (2017). Persistence and clearance of Ebola virus RNA from seminal fluid of Ebola virus disease survivors: a longitudinal analysis and modelling study. Lancet Glob. Heal. *5*, e80–e88.

Skalsky, R.L., Olson, K.E., Blair, C.D., Garcia-Blanco, M.A., and Cullen, B.R. (2014). A "microRNA-like" small RNA expressed by Dengue virus? Proc. Natl. Acad. Sci. U. S. A. *111*.

Skrip, L.A., Fallah, M.P., Gaffney, S.G., Yaari, R., Yamin, D., Huppert, A., Bawo, L., Nyenswah, T., and Galvani, A.P. (2017). Characterizing risk of Ebola transmission based on frequency and type of case - contact exposures. Philos. Trans. R. Soc. B-Biological Sci. *372*.

Smith, P.G. (2010). Concepts of herd protection and immunity. Procedia Vaccinol. *2*, 134–139.

Smith, A.E., and Helenius, A. (2004). How Viruses Enter Animal Cells. Science (80-.). *304*, 237–242.

Snape, M.D. (2017). Persistence of immune responses induced by Ebola virus vaccines. Lancet Glob. Heal. *5*, e238–e239.

Soka, M.J., Choi, M.J., Baller, A., White, S., Rogers, E., Purpura, L.J., Mahmoud, N., Wasunna, C., Massaquoi, M., Abad, N., et al. (2016). Prevention of sexual transmission of Ebola in Liberia through a national semen testing and counselling programme for survivors: an analysis of Ebola virus RNA results and behavioural data. Lancet Glob. Heal. *4*, e736–e743.

Song, B.H., Yun, S.I., Woolley, M., and Lee, Y.M. (2017). Zika virus: History, epidemiology, transmission, and clinical presentation. J. Neuroimmunol. *308*, 50–64.

Song, W., Gui, M., Wang, X., and Xiang, Y. (2018). Cryo-EM structure of the SARS

coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLoS Pathog. *14*, 1–19.

Song, Z., Xu, Y., Bao, L., Zhang, L., Yu, P., Qu, Y., Zhu, H., Zhao, W., Han, Y., and Qin, C. (2019). From SARS to MERS, thrusting coronaviruses into the spotlight. Viruses *11*.

Sonnberg, S., Webby, R.J., and Webster, R.G. (2013). Natural history of highly pathogenic avian influenza H5N1. Virus Res. *178*, 63–77.

Sousa, Z.L. (2014). Key features of Ebola hemorrhagic fever: a review. Asian Pac. J. Trop. Biomed. *4*, 841–844.

Stahelin, R. V. (2014). Membrane binding and bending in Ebola VP40 assembly and egress. Front. Microbiol. *5*, 1–12.

Stanley, D.A., Honko, A.N., Asiedu, C., Trefry, J.C., Lau-Kilby, A.W., Johnson, J.C., Hensley, L., Ammendola, V., Abbate, A., Grazioli, F., et al. (2014). Chimpanzee adenovirus vaccine generates acute and durable protective immunity against ebolavirus challenge. Nat. Med. *20*, 1126–1129.

Steckelberg, A.L., Vicens, Q., and Kieft, J.S. (2018). Exoribonuclease-resistant RNAs exist within both coding and noncoding subgenomic RNAs. MBio *9*, 1–12.

Strickland, M., Ehrlich, L.S., Watanabe, S., Khan, M., Strub, M.P., Luan, C.H., Powell, M.D., Leis, J., Tjandra, N., and Carter, C.A. (2017). Tsg101 chaperone function revealed by HIV-1 assembly inhibitors. Nat. Commun. *8*.

Su, Z., Wu, C., Shi, L., Luthra, P., Pintilie, G.D., Johnson, B., Porter, J.R., Ge, P., Chen, M., Liu, G., et al. (2018). Electron Cryo-microscopy Structure of Ebola Virus Nucleoprotein Reveals a Mechanism for Nucleocapsid-like Assembly. Cell *172*, 966–970.e12.

Sugita, Y., Matsunami, H., Kawaoka, Y., Noda, T., and Wolf, M. (2018). Cryo-EM structure of the Ebola virus nucleoprotein–RNA complex at 3.6 Å resolution. Nature.

Sztuba-Solinska, J., Diaz, L., Kumar, M.R., Kolb, G., Wiley, M.R., Jozwick, L., Kuhn, J.H., Palacios, G., Radoshitzky, S.R., Le Grice, S.F.J., et al. (2016). A small stem-loop structure of the Ebola virus trailer is essential for replication and interacts with heat-

shock protein A8. Nucleic Acids Res. *44*, 9831–9846.

T.W., G., and P.B., J. (1995). Differentiation of filoviruses by electron microscopy. Virus Res. *39*, 129–150.

Takada, A., and Kawaoka, Y. (2001). The pathogenesis of Ebola hemorrhagic fever. Trends Microbiol. *9*, 506–511.

Talbot, P.J., and Vance, D.E. (1980). Evidence that Sindbis virus infects BHK-21 cells via a lysosomal route. Can. J. Biochem. *58*, 1131–1137.

Tapia, M.D., Sow, S.O., Lyke, K.E., Haidara, F.C., Diallo, F., Doumbia, M., Traore, A., Coulibaly, F., Kodio, M., Onwuchekwa, U., et al. (2016). Use of ChAd3-EBO-Z Ebola virus vaccine in Malian and US adults, and boosting of Malian adults with MVA-BN-Filo: a phase 1, single-blind, randomised trial, a phase 1b, open-label and double-blind, dose-escalation trial, and a nested, randomised, double-bli. Lancet. Infect. Dis. *16*, 31–42.

Taubenberger, J.K., and Kash, J.C. (2011). Insights on influenza pathogenesis from the grave. Virus Res. *162*, 2–7.

Taylor, L.H., Latham, S.M., and Woolhouse, M.E.J. (2001). Risk factors for human disease emergence. Philos. Trans. R. Soc. B Biol. Sci. *356*, 983–989.

Team, C.C.-19 R. (2020). Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. MMWR. Morb. Mortal. Wkly. Rep. *69*, 343–346.

Teng, Y., Wang, Y., Zhang, X., Liu, W., Fan, H., Yao, H., Lin, B., Zhu, P., Yuan, W., Tong, Y., et al. (2015). Systematic Genome-wide Screening and Prediction of microRNAs in EBOV during the 2014 Ebolavirus Outbreak. Sci. Rep. *5*, 1–17.

Teppa, E., Wilkins, A.D., Nielsen, M., and Buslje, C.M. (2012). Disentangling evolutionary signals: conservation, specificity determining positions and coevolution. Implication for catalytic residue prediction. BMC Bioinformatics *13*, 1–8.

Thèves, C., Biagini, P., and Crubézy, E. (2014). The rediscovery of smallpox. Clin. Microbiol. Infect. *20*, 210–218.

Timen, A. (2009). Response to Imported Case of Marburg Hemorrhagic Fever, the Netherlands. Emerg. Infect. Dis. *15*, 1171–1175.

Totura, A.L., and Baric, R.S. (2012). SARS coronavirus pathogenesis: host innate immune responses and viral antagonism of interferon. *2*, 264–275.

Touret, F., and de Lamballerie, X. (2020). Of chloroquine and COVID-19. Antiviral Res. *177*, 104762.

Trad, M.A., Naughton, W., Yeung, A., Mazlin, L., O'sullivan, M., Gilroy, N., Fisher, D.A., and Stuart, R.L. (2017). Ebola virus disease: An update on current prevention and management strategies. J. Clin. Virol. *86*, 5–13.

Tricco, A.C., Chit, A., Soobiah, C., Hallett, D., Meier, G., Chen, M.H., Tashkandi, M., Bauch, C.T., and Loeb, M. (2013). Comparing influenza vaccine efficacy against mismatched and matched strains: A systematic review and meta-analysis. BMC Med. *11*.

Trunschke, M., Conrad, D., Enterlein, S., Olejnik, J., Brauburger, K., and 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*, 135–145.

Tyrrell, D.A.J. (1987). The Common Cold - My Favourite Infection. J. Gen. Virol. *68*, 2053–2061.

Ueda, M.T., Kurosaki, Y., Izumi, T., Nakano, Y., Oloniniyi, O.K., Yasuda, J., Koyanagi, Y., Sato, K., and Nakagawa, S. (2017). Functional mutations in spike glycoprotein of Zaire ebolavirus associated with an increase in infection efficiency. Genes to Cells *22*, 148–159.

Urata, S., Noda, T., Kawaoka, Y., Morikawa, S., Yokosawa, H., and Yasuda, J. (2007). Interaction of Tsg101 with Marburg Virus VP40 Depends on the PPPY Motif, but Not the PT/SAP Motif as in the Case of Ebola Virus, and Tsg101 Plays a Critical Role in the Budding of Marburg Virus-Like Particles Induced by VP40, NP, and GP. J. Virol. *81*, 4895–4899.

Urbanowicz, R.A., McClure, C.P., Sakuntabhai, A., Sall, A.A., Kobinger, G., Müller, M.A., Holmes, E.C., Rey, F.A., Simon-Loriere, E., and Ball, J.K. (2016). Human Adaptation of Ebola Virus during the West African Outbreak. Cell *167*, 1079–1087.e5.

Ursic-Bedoya, R., Mire, C.E., Robbins, M., Geisbert, J.B., Judge, A., MacLachlan, I.,

and Geisbert, T.W. (2014). Protection against lethal marburg virus infection mediated by lipid encapsulated small interfering RNA. J. Infect. Dis. *209*, 562–570.

Uyeki, T.M., Erickson, B.R., Brown, S., McElroy, A.K., Cannon, D., Gibbons, A., Sealy, T., Kainulainen, M.H., Schuh, A.J., Kraft, C.S., et al. (2016). Ebola Virus Persistence in Semen of Male Survivors. Clin. Infect. Dis. *62*, 1552–1555.

Valencia, C., Bah, H., Fatoumata, B., Rodier, G., Diallo, B., Kon??, M., Giese, C., Conde, L., Malano, E., Mollet, T., et al. (2017). Network visualization for outbreak response: Mapping the Ebola Virus Disease (EVD) chains of transmission in N'Z??r??kor??, Guinea. J. Infect. *74*, 294–301.

Valmas, C., and Basler, C.F. (2011). Marburg Virus VP40 Antagonizes Interferon Signaling in a Species-Specific Manner. J. Virol. *85*, 4309–4317.

Velasco-Villa, A., Mauldin, M.R., Shi, M., Escobar, L.E., Gallardo-Romero, N.F., Damon, I., Olson, V.A., Streicker, D.G., and Emerson, G. (2017). The history of rabies in the Western Hemisphere. Antiviral Res. *146*, 221–232.

Venter, M. (2018). Assessing the zoonotic potential of arboviruses of African origin. Curr. Opin. Virol. *28*, 74–84.

Vernet, M.-A., Reynard, S., Fizet, A., Schaeffer, J., Pannetier, D., Guedj, J., Rives, M., Georges, N., Garcia-Bonnet, N., Sylla, A.I., et al. (2017). Clinical, virological, and biological parameters associated with outcomes of Ebola virus infection in Macenta, Guinea. JCI Insight *2*, 1–14.

Viruses, C.S.G. of the I.C. on T. of (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol.

Volchkov, V.E., Volchkova, V.A., Chepurnov, A.A., Blinov, V.M., Dolnik, O., Netesov, S. V., and Feldmann, H. (1999). Characterization of the L gene and 5' trailer region of Ebola virus. J. Gen. Virol. *80*, 355–362.

Volz, A., and Sutter, G. (2017). Modified Vaccinia Virus Ankara: history, value in basic research, and current perspectives for vaccine development. Adv. Virus Res. *97*, 187–243.

Volz, E., and Pond, S. (2014). Phylodynamic Analysis of Ebola Virus in the 2014

Sierra Leone Epidemic. PLoS Curr.

Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, V.D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell.

Wan, W., Kolesnikova, L., Clarke, M., Koehler, A., Noda, T., Becker, S., and Briggs, J.A.G. (2017). Structure and assembly of the Ebola virus nucleocapsid. Nature.

Wan, Y., Shang, J., Graham, R., Baric, R.S., and Li, F. (2020). Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. J. Virol. 1–9.

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., Shi, Z., Hu, Z., Zhong, W., and Xiao, G. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. *30*, 269–271.

Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., and Jones, D.T. (2004). The DISOPRED server for the prediction of protein disorder. Bioinformatics *20*, 2138–2139.

Warfield, K.L., Bradfute, S.B., Wells, J., Lofts, L., Cooper, M.T., Alves, D.A., Reed, D.K., VanTongeren, S.A., Mech, C.A., and Bavari, S. (2009). Development and Characterization of a Mouse Model for Marburg Hemorrhagic Fever. J. Virol. *83*, 6404–6415.

Washietl, S., and Hofacker, I.L. (2004). Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. J. Mol. Biol. *342*, 19–30.

Wastika, C.E., Harima, H., Sasaki, M., Hang'ombe, B.M., Eshita, Y., Qiu, Y., Hall, W.W., Wolfinger, M.T., Sawa, H., and Orba, Y. (2020). Discoveries of Exoribonuclease-Resistant Structures of Insect-Specific Flaviviruses Isolated in Zambia. Viruses *12*, 1–21.

Watanabe, S.M., Ehrlich, L.S., Strickland, M., Li, X., Soloveva, V., Goff, A.J., Stauft, C.B., Bhaduri-McIntosh, S., Tjandra, N., and Carter, C. (2020). Selective Targeting of Virus Replication by Proton Pump Inhibitors. Sci. Rep. *10*, 1–15.

Weaver, S.C., Costa, F., Garcia-Blanco, M.A., Ko, A.I., Ribeiro, G.S., Saade, G., Shi, P.Y., and Vasilakis, N. (2016). Zika virus: History, emergence, biology, and prospects

for control. Antiviral Res. *130*, 69–80.

Wei, H., Audet, J., Wong, G., He, S., Huang, X., Cutts, T., Theriault, S., Xu, B., Kobinger, G., and Qiu, X. (2017). Deep-sequencing of Marburg virus genome during sequential mouse passaging and cell-culture adaptation reveals extensive changes over time. Sci. Rep. *7*, 1–8.

Wei, W.E., Li, Z., Chiew, C.J., Yong, S.E., Toh, M.P., and Lee, V.J. (2020). Presymptomatic Transmission of SARS-CoV-2-Singapore. Morb. Mortal. Wkly. Rep. *69*, 411–415.

Weik, M., Enterlein, S., Schlenz, K., and Mühlberger, E. (2005). The Ebola Virus Genomic Replication Promoter Is Bipartite and Follows the Rule of Six. J. Virol. *79*, 10660–10671.

Weingartl, H.M., Nfon, C., and Kobinger, G. (2013). Review of Ebola virus infections in domestic animals. Dev. Biol. (Basel). *135*, 211–218.

Weng, K.F., Hung, C.T., Hsieh, P.T., Li, M.L., Chen, G.W., Kung, Y.A., Huang, P.N., Kuo, R.L., Chen, L.L., Lin, J.Y., et al. (2014). A cytoplasmic RNA virus generates functional viral small RNAs and regulates viral IRES activity in mammalian cells. Nucleic Acids Res. *42*, 12789–12805.

Whitmer, S.L.M., Ladner, J.T., Wiley, M.R., Patel, K., Dudas, G., Rambaut, A., Sahr, F., Prieto, K., Shepard, S.S., Carmody, E., et al. (2018). Active Ebola Virus Replication and Heterogeneous Evolutionary Rates in EVD Survivors. Cell Rep. *22*, 1159–1168.

Widagdo, W., Okba, N.M.A., Stalin Raj, V., and Haagmans, B.L. (2017). MERS-coronavirus: From discovery to intervention. One Heal. *3*, 11–16.

Wikan, N., and Smith, D.R. (2016). Zika virus: History of a newly emerging arbovirus. Lancet Infect. Dis. *16*, e119–e126.

Will, S., Reiche, K., Hofacker, I.L., Stadler, P.F., and Backofen, R. (2007). Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLoS Comput. Biol. *3*, 680–691.

Wilson, J. a, Bray, M., Bakken, R., and Hart, M.K. (2001). Vaccine potential of Ebola virus VP24, VP30, VP35, and VP40 proteins. Virology *286*, 384–390.

Wimmer, E., Mueller, S., Tumpey, T.M., and Taubenberger K, J. (2009). Synthetic viruses: a new opportunity to understand and prevent viral disease. Nat Biotechnol. *27*, 1–23.

Winslow, R.L., Milligan, I.D., Voysey, M., Luhn, K., Shukarev, G., Douoguih, M., and Snape, M.D. (2017). Immune responses to novel adenovirus type 26 and modified vaccinia virus Ankara-vectored ebola vaccines at 1 year. JAMA - J. Am. Med. Assoc. *317*, 1075–1077.

de Wit, E., Feldmann, F., Cronin, J., Jordan, R., Okumura, A., Thomas, T., Scott, D., Cihlar, T., and Feldmann, H. (2020). Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection. Proc. Natl. Acad. Sci. *117*, 201922083.

De Wit, E., Van Doremalen, N., Falzarano, D., and Munster, V.J. (2016). SARS and MERS: Recent insights into emerging coronaviruses. Nat. Rev. Microbiol. *14*, 523–534.

Wolfe, N.D., Dunavan, C.P., and Diamond, J. (2007). Origins of major human infectious diseases. Nature *447*, 279–283.

Wong, G., Audet, J., Fernando, L., Fausther-Bovendo, H., Alimonti, J.B., Kobinger, G.P., and Qiu, X. (2014). Immunization with vesicular stomatitis virus vaccine expressing the Ebola glycoprotein provides sustained long-term protection in rodents. Vaccine *32*, 5722–5729.

Woods, C.L. (2016). When more than reputation is at risk: How two hospitals responded to Ebola. Public Relat. Rev. *42*, 893–902.

Woolhouse, M., Scott, F., Hudson, Z., Howey, R., and Chase-Topping, M. (2012). Human viruses: discovery and emergence. Philos. Trans. R. Soc. B Biol. Sci. *367*, 2864–2871.

World Health Organization (2017). Meeting of the Strategic Advisory Group of Experts on Immunization, April 2017 - conclusions and recommendations. Wkly. Epidemiol. Rec. *92*, 301–320.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science (80-.). *367*, 1260–1263.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., et al. (2020a). Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe *27*, 325–328.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., et al. (2020b). Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe 1–4.

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., et al. (2020c). A new coronavirus associated with human respiratory disease in China. Nature *579*, 265–269.

Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., Li, T., and Chen, Q. (2020). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. Int. J. Oral Sci. *12*, 1–5.

Xu, W., Edwards, M.R., Borek, D.M., Feagins, A.R., Mittal, A., Alinger, J.B., Berry, K.N., Yen, B., Hamilton, J., Brett, T.J., et al. (2015). Ebola virus VP24 targets a unique NLS-binding site on karyopherin5 to selectively compete with nuclear import of phosphorylated STAT1. Cell Host Microbe *114*, 587–592.

Yamamoto, M., Matsuyama, S., Li, X., Takeda, M., Kawaguchi, Y., and Inoue, J. (2016). Identification of Nafamostat as a Potent Inhibitor of Middle East Respiratory Syndrome Coronavirus S Protein-Mediated Membrane Fusion Using the Split-Protein-Based Cell-Cell Fusion Assay. Antimicrob. Agents Chemother. *60*, 6532–6539.

Yamaoka, S., Groseth, A., Lehrer, A.T., Verma, S., Lai, C.-Y., Strange, D.P., Ann, T., and Wong, S. (2017). Ebola Virus Glycoprotein Induces an Innate Immune Response In vivo via TLR4. Front. Microbiol *8*, 3389–1571.

Yamauchi, Y., and Greber, U.F. (2016). Principles of Virus Uncoating: Cues and the Snooker Ball. Traffic *17*, 569–592.

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. Science *2*, 1444–1448.

Yang, D., and Leibowitz, J.L. (2015). The structure and functions of coronavirus genomic 3' and 5' ends.

Yang, X., Tan, C.W., Anderson, D.E., Jiang, R., Li, B., Zhang, W., Zhu, Y., Lim, X.F., Zhou, P., Liu, X., et al. (2019). Characterization of a filovirus (Měnglà virus) from Rousettus bats in China. Nat. Microbiol.

Yang, Y., Shang, W., and Rao, X. (2020). Facing the COVID-19 outbreak: What should we know and what could we do? J. Med. Virol. 25720.

Ye, J., McGinnis, S., and Madden, T.L. (2006). BLAST: Improvements for better sequence analysis. Nucleic Acids Res. *34*, 6–9.

Yen, H.L. (2016). Current and novel antiviral strategies for influenza infection. Curr. Opin. Virol. *18*, 126–134.

Yin, Y., and Wunderink, R.G. (2018). MERS, SARS and other coronaviruses as causes of pneumonia. Respirology *23*, 130–137.

Yu, P., Zhu, J., Zhang, Z., Han, Y., and Huang, L. (2020). A familial cluster of infection associated with the 2019 novel coronavirus indicating potential person-to-person transmission during the incubation period. J. Infect. Dis. 1–5.

Yuan, X., Li, J., Shan, Y., Yang, Z., Zhao, Z., Chen, B., Yao, Z., Dong, B., Wang, S., Chen, J., et al. (2005). Subcellular localization and membrane association of SARS-CoV 3a protein. Virus Res. *109*, 191–202.

Zhirnov, O.P., Klenk, H.D., and Wright, P.F. (2011). Aprotinin and similar protease inhibitors as drugs against influenza. Antiviral Res. *92*, 27–36.

Zhou, P., Yang, X. Lou, Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature *579*, 270–273.

Zhou, Y., Vedantham, P., Lu, K., Agudelo, J., Carrion, R., Nunneley, J.W., Barnard, D., Pöhlmann, S., Mckerrow, J.H., Renslo, A.R., et al. (2015). Protease inhibitors targeting coronavirus and filovirus entry. Antiviral Res. *Apr*, 76–84.

Zhu, F.C., Wurie, A.H., Hou, L.H., Liang, Q., Li, Y.H., Russell, J.B.W., Wu, S.P., Li, J.X., Hu, Y.M., Guo, Q., et al. (2017). Safety and immunogenicity of a recombinant adenovirus type-5 vector-based Ebola vaccine in healthy adults in Sierra Leone: a single-centre, randomised, double-blind, placebo-controlled, phase 2 trial. Lancet *389*, 621–628.

Zinzula, L., Orsini, M., Weyher-stingl, E., Bracher, A., Orsini, M., Weyher-stingl, E., and Bracher, A. (2019). Structures of Ebola and Reston Virus VP35 Oligomerization Domains and Comparative Biophysical Characterization in All Ebolavirus Species Article Structures of Ebola and Reston Virus VP35 Oligomerization Domains and Comparative Biophysical Characterization. Structure *27*, 39–54.

Acknowledgements

This thesis would not have been possible without the help of some absolutely wonderful people. Firstly I would like to thank my supervisors – Professor Martin Michaelis and Dr Mark Wass. Their help, guidance and input has been invaluable these past three years, and I will be forever grateful for the opportunity they have afforded me. Our meetings and chats I will never forget. Ever.

I would like to thank my supportive wife Kat Masterson for all her help and support, as well as all of my family and hers, particularly Caroline Masterson, Paul Masterson, Julie Last, Bob Last, Ruby Masterson, and, most importantly, Katie Bear.

Special thanks go to my fellow lab members: Jake McGreig for all his help, support and contribution to the research presented in this thesis. Magdalena Antczak for her supportive and enlightening motivational speeches and for being the emotional soundboard I needed. Helen Grimsley for her inspiring attitude and boundless energy. Katie McLoughlin for her stoic presence and reassuring smile. Ian Reddin for something. Henry Martell for his guidance and help in getting everything off the ground. And to all other lab members past and present who have shaped the very focus of my being.

My appreciation goes out to Michael Wolfinger of the University of Vienna for his help in learning about the world of RNA, and to all others that I have collaborated with on the papers presented in this thesis.

Further thanks go to my inspirations and motivational bedrocks – Taylor Swift, Ewan McGregor, Bruce Wayne, Elena Hernan, Catherine Graham, Tal McMaster-Christie, Tom Apps, Haydon Christou, Viceroy Nute Gunray and Britney Spears. I would not have been where I am today without you all.

Appendix 1: Chapter 2 Supplementary Material

Herd Immunity to Ebolaviruses is Not a Realistic Target for Current Vaccination Strategies

Authors: Stuart G. Masterson[1], Leslie Lobel[2,3], Miles W. Carroll[4], Mark N. Wass[1]* and Martin Michaelis[1]*

Affiliation: 1 Industrial Biotechnology Centre and School of Biosciences, University of Kent, Canterbury, United Kingdom, 2 Department of Microbiology, Immunology and Genetics, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel, 3 Department of Emerging and Re-Emerging Diseases and Special Pathogens, Uganda Virus Research Institute (UVRI), Entebbe, Uganda, 4 Public Health England, Porton Down, Salisbury, United Kingdom

Correspondence: Mark N Wass (m.n.wass@kent.ac.uk); Martin Michaelis (m.michaelis@kent.ac.uk)

Data Sheet 1. Results of the literature search

Literature searches were performed using PubMed (www.ncbi.nlm.nih.gov/pubmed) on 29th September 2017. The following sear terms were used: "Ebola R0", "Ebola basic reproductive number", "Ebola reproduction number". One additional article has been identified from the reference list of other articles. The article list after the removal of overlaps is presented in the tab "Combined". The list of relevant articles that have been included into the analysis is provided in the tab "Relevant articles".

"Ebola R0"

PMID	DOI	Journal reference	First Author	Title	Comment	
28263987	10.1371/journal.pcbi.1005416	PLoS Comput Biol. 2017 Mar 6;13(3):e1005416	Saulnier E	Inferring epidemiological parameters from phylogenies using regression-ABC: A comparative study.	West African Ebola outbreak	R0 = 5.92
27524644	10.1016/j.jtbi.2016.08.016	J Theor Biol. 2016 Nov 7;408:145-54	Taylor BP	Stochasticity and the limits to confidence when estimating R0 of Ebola and other emerging infectious diseases.	West African Ebola outbreak	R0 (Liberia) = 2.06; R0 (Sierra Leone) = 1.71; R0 (Guinea) = 1.24
27434164	10.1371/journal.pntd.0004867	PLoS Negl Trop Dis. 2016 Jul 19;10(7):e0004867	Krauer F	Heterogeneity in District-Level Transmission of Ebola Virus Disease during the 2013-2015 Epidemic in West Africa.	West African Ebola outbreak	R0 ranges from 0.36 to 3.37 in different districts
27357339	10.1093/infdis/jiw207	J Infect Dis. 2016 Oct 15;214(suppl 3):S93-S101	Breman JG	Discovery and Description of Ebola Zaire Virus in 1976 and Relevance to the West African Epidemic During 2013-2016.	No primary data on R0	
27049322	10.1371/journal.pone.0152438	PLoS One. 2016 Apr 6;11(4):e0152438	Guo Z	Predicting and Evaluating the Epidemic Trend of Ebola Virus Disease in the 2014-2015 Outbreak and the Effects of Intervention Measures.	No primary data on R0	
26347015	10.1038/Srep13857	Sci Rep. 2015 Sep 8;5:13857	Xia ZQ	Modeling the transmission dynamics of Ebola virus disease in Liberia.	West African Ebola outbreak	R0 (Liberia) = 2.01
25737782	10.1186/s40249-015-	Infect Dis Poverty. 2015 Feb 24;4:13	Khan A	Estimating the basic reproductive ratio for the	West African	R0 (Liberia) = 1.76; R0 (Sierra

	0043-3			Ebola outbreak in Liberia and Sierra Leone.	Ebola outbreak	Leone) = 1.49
25685614	10.1371/currents.outbreaks.c6efe8274dc55274f05cbcb62bbe6070	PLoS Curr. 2014 Nov 13;6	Kiskowski MA	A three-scale network model for the early growth dynamics of 2014 west Africa ebola epidemic.	West African Ebola outbreak	R0 = 2.35
25642370	10.1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f	PLoS Curr. 2014 Oct 6;6	Stadler T	Insights into the early epidemic spread of ebola in sierra leone provided by viral sequence data.	West African Ebola outbreak	R0 (Sierra Leone) = 2.58
25642358	10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571	PLoS Curr. 2014 Sep 8;6	Fisman D	Early epidemic dynamics of the west african 2014 ebola outbreak: estimates derived with a simple two-parameter model.	West African Ebola outbreak	R0 = 1.6 - 2.0
25566687	10.1186/1742-4682-12-1	Theor Biol Med Model. 2015 Jan 6;12:1	Nishiura H	Theoretical perspectives on the infectiousness of Ebola virus disease.	No primary data on R0	
25495064	10.4161/21505594.2014.976514	Virulence. 2014;5(8):825-7	Alizon S	Quantifying the epidemic spread of Ebola virus (EBOV) in Sierra Leone using phylodynamics.	West African Ebola outbreak	R0 (Sierra Leone) = 1.26
25467086	10.1016/j.tmaid.2014.10.015	Travel Med Infect Dis. 2014 Nov-Dec;12(6 Pt A):650-8	Chen T	Risk of imported Ebola virus disease in China.	Ebola virus: DR Congo, Kikwit 1995: R0 = 2.64; Uganda, Gulu 2000: R0 = 3.54	
25244186	10.1056/NEJMoa1411100	N Engl J Med. 2014 Oct 16;371(16):1481-95	WHO Ebola Response Team	Ebola virus disease in West Africa--the first 9 months of the epidemic and forward projections.	West African Ebola outbreak	R0 (Liberia) = 1.83; R0 (Sierra Leone) = 2.02; R0 (Guinea) = 1.71
25217532	10.7554/eLife.03908	Elife. 2014 Sep 12;3:e03908	House T	Epidemiological dynamics of Ebola outbreaks.	No primary data on R0	
25914858	10.1371/currents.outbreaks.6f7025f1271821d4c815385b08f5f80e	PLoS Curr. 2014 Oct 24;6	Volz E	Phylodynamic analysis of ebola virus in the 2014 sierra leone epidemic.	West African Ebola outbreak	R0 (Sierra Leone) = 2.40 - 3.81
16999875	10.1017/S0950268806	Epidemiol Infect. 2007 May;135(4):610-21	Legrand J	Understanding the dynamics of Ebola epidemics.	Ebola virus: DR Congo, 1995;	R0 = 2.7

	007217				Uganda, 2000	
15178190	10.1016/j.jtbi.2004.03.006	J Theor Biol. 2004 Jul 7;229(1):119-26	Chowell G	The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda.	Ebola virus: DR Congo, 1995; Uganda, 2000	R0 (DR Congo): 1.83; R0 (Uganda) = 1.34

"Ebola basic reproductive number"

PMID	DOI	Journal reference	First Author	Title	Comment	
28830110	10.1093/infdis/jiw356	J Infect Dis. 2016 Dec 1;214(suppl_4):S421-S426	Chowell G	Elucidating Transmission Patterns From Internet Reports: Ebola and Middle East Respiratory Syndrome as Case Studies.	West African Ebola outbreak	R0 = 2.4
28610609	10.1186/s12938-017-0370-7	Biomed Eng Online. 2017 Jun 13;16(1):79	Chang HJ	Estimation of basic reproduction number of the Middle East respiratory syndrome coronavirus (MERS-CoV) during the outbreak in South Korea, 2015.	No primary data on R0	
28269870		AMIA Annu Symp Proc. 2017 Feb 10;2016:743-752	Lee EK	A Compartmental Model for Zika Virus with Dynamic Human and Vector Populations.	No primary data on R0	
28202592	10.1098/rsif.2016.0847	J R Soc Interface. 2017 Feb;14(127)	Nieddu GT	Extinction pathways and outbreak vulnerability in a stochastic Ebola model.	No primary data on R0	
28166851	10.1017/S0950268817000164	Epidemiol Infect. 2017 Apr;145(6):1069-1094	Wong ZS	A systematic review of early modelling studies of Ebola virus disease in West Africa.	West African Ebola outbreak	R0 (overall) = 1.78; R0 (Guinea) = 1.30; R0 (Liberia) = 1.84; R0 (Sierra Leone) = 1.70; R0 (Nigeria) = 9.01
27833001	10.1016/j.mbs.2016.	Math Biosci. 2017 Jan;283:48-59	Agusto FB	Mathematical model of Ebola	No primary data on R0	

	11.002			transmission dynamics with relapse and reinfection.		
27579053	10.1155/ 2016/ 9352725	Comput Math Methods Med. 2016;2016:9352725	Ngwa GA	A Mathematical Model with Quarantine States for the Dynamics of Ebola Virus Disease in Human Populations.	No primary data on R0	
27524644	10.1016/ j.jtbi.2016. 08.016	J Theor Biol. 2016 Nov 7;408:145-54	Taylor BP	Stochasticity and the limits to confidence when estimating R0 of Ebola and other emerging infectious diseases.	West African Ebola outbreak	R0 (Liberia) = 2.06; R0 (Sierra Leone) = 1.71; R0 (Guinea) = 1.24
27434164	10.1371/ journal.pntd. 0004867	PLoS Negl Trop Dis. 2016 Jul 19;10(7):e0004867	Krauer F	Heterogeneity in District-Level Transmission of Ebola Virus Disease during the 2013-2015 Epidemic in West Africa.	West African Ebola outbreak	R0 ranges from 0.36 to 3.37 in different districts
27405359	10.1186/ s40249-016- 0161-6	Infect Dis Poverty. 2016 Jul 13;5(1):72	Ahmad MD	Optimal control analysis of Ebola disease with control strategies of quarantine and vaccination.	No primary data on R0	
27049322	10.1371/ journal.pone. 0152438	PLoS One. 2016 Apr 6;11(4):e0152438	Guo Z	Predicting and Evaluating the Epidemic Trend of Ebola Virus Disease in the 2014-2015 Outbreak and the Effects of Intervention Measures.	No primary data on R0	
26981342	10.1016/ j.phrp.2015. 12.012	Osong Public Health Res Perspect. 2016 Feb;7(1):43-8	Do TS	Modeling the Spread of Ebola.	West African Ebola outbreak	R0 (Nigeria) = 2.57
26718863	10.1016/ j.jtbi.2015. 11.025	J Theor Biol. 2016 Mar 7;392:99-106	Adams B	Household demographic determinants of Ebola epidemic risk.	No primary data on R0	
26618087	10.7717/ peerj.1418	PeerJ. 2015 Nov 19;3:e1418	Althaus CL	Rapid drop in the	DR Congo 2014	R0 = 5.2

				reproduction number during the Ebola outbreak in the Democratic Republic of Congo.		
26574727	10.1111/ biom.12432	Biometrics. 2016 Jun;72(2):335-43	Brown GD	An empirically adjusted approach to reproductive number estimation for stochastic compartmental models: A case study of two Ebola outbreaks.	No primary data on R0	
26525597	10.7554/ eLife.09015	Elife. 2015 Nov 3;4	Rosello A	Ebola virus disease in the Democratic Republic of the Congo, 1976-2014.	Outbreaks in DR Congo	R0 = 2.5 - 8.0
26515898	10.1038/ srep15818	Sci Rep. 2015 Oct 30;5:15818	Shen M	Modeling the effect of comprehensive interventions on Ebola virus transmission.	West African Ebola outbreak	R0 (Liberia) = 1.80; R0 (Sierra Leone) = 1.61; R0 (Guinea) = 1.26
26451161	10.1155/2015 /207105	Comput Math Methods Med. 2015;2015:207105	Liu W	Model Selection and Evaluation Based on Emerging Infectious Disease Data Sets including A/H1N1 and Ebola.	West African Ebola outbreak	R0 = 1.21 - 3.02
26402477	10.3201/ eid2110. 150912	10.3201/eid2110.150912	Lindblade KA	Decreased Ebola Transmission after Rapid Response to Outbreaks in Remote Areas, Liberia, 2014.	West African Ebola outbreak	R0 (Liberia) = 1.7
26347015	10.1038/ srep13857	Sci Rep. 2015 Sep 8;5:13857	Xia ZQ	Modeling the transmission dynamics of Ebola virus disease in Liberia.	West African Ebola outbreak	R0 (Liberia) = 2.01

26343011	10.1016/ j.antiviral. 2015.08.015	Antiviral Res. 2015 Nov;123:70-7	Madelain V	Ebola virus dynamics in mice treated with favipiravir.	Out of scope (mouse study)	
26297316	10.1016/ j.jtbi.2015. 08.004	J Theor Biol. 2015 Nov 7;384:33-49	Browne C	Modeling contact tracing in outbreaks with application to Ebola.	West African Ebola outbreak	R0 (Guinea) = 1 - 12; R0 (Sierra Leone) = 1 - 11
26280179	10.3934/ mbe.2015. 12. 1055	Math Biosci Eng. 2015 Oct;12(5):1055-63	Wang XS	Ebola outbreak in West Africa: real-time estimation and multiple-wave prediction.	West African Ebola outbreak	R0 (Liberia) = 1.23; R0 (Sierra Leone) = 1.18; R0 (Guinea) = 1.12
26246849	10.1155/ 2015/582625	Comput Math Methods Med. 2015;2015:582625	Li Z	Dynamical Analysis of an SEIT Epidemic Model with Application to Ebola Virus Transmission in Guinea.	West African Ebola outbreak	R0 (Guinea) = 4.16
26197242	10.1371/ journal.pone .0131398	PLoS One. 2015 Jul 21;10(7):e0131398	Barbarossa MV	Transmission Dynamics and Final Epidemic Size of Ebola Virus Disease Outbreaks with Varying Interventions.	West African Ebola outbreak	R(0) = 1.44
26029377	10.1038/ sdata. 2015.19	Sci Data. 2015 May 26;2:150019	Van Kerkhove MD	A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making.	All available data 2/2015	R(0) Ebola virus = 1.36 - 4.71; R(0) Sudan virus = 1.34 - 2.7; West Africa: R(0) Guinea = 1.71, R(0) Sierra Leone = 1.83, R(0) Liberia = 2.02
25979285	10.1016/ j.epidem. 2015.03.001	Epidemics. 2015 Jun;11:80-4	Althaus CL	Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control.	West African Ebola outbreak	R(0) Nigeria 9.0
25902936	10.1186/ s12916- 015-0318-3	BMC Med. 2015 Apr 23;13:96	Agusto FB	Mathematical assessment of the effect of	West African Ebola outbreak	R(0) Guinea = 1.0 (in the

				traditional beliefs and customs on the transmission dynamics of the 2014 Ebola outbreaks.		absence of risky burial traditions) - 1.6 (in the presence of risky burial traditions)
25766240	10.1126/science.aaa4339	Science. 2015 Mar 13;347(6227):aaa4339	Heesterbeek H	Modeling infectious disease dynamics in the complex landscape of global health.	No primary data on R0	
25736239	10.1038/srep08751	Sci Rep. 2015 Mar 4;5:8751	Weitz JS	Modeling post-death transmission of Ebola: challenges for inference and opportunities for control.	Out of scope (focused on post-death transmission)	
25642370	10.1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f	PLoS Curr. 2014 Oct 6;6	Stadler T	Insights into the early epidemic spread of ebola in sierra leone provided by viral sequence data.	West African Ebola outbreak	R0 (Sierra Leone) = 2.58
25642364	10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288	PLoS Curr. 2014 Sep 2;6	Althaus CL	Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa.	West African Ebola outbreak	R0 (Liberia) = 1.59; R0 (Sierra Leone) = 2.53; R0 (Guinea) = 1.51
25642360	10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5	PLoS Curr. 2014 Sep 2;6	Gomes MF	Assessing the international spreading risk associated with the 2014 west african ebola outbreak.	West African Ebola outbreak	R0 = 1.5 - 2.0
25642358	10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571	PLoS Curr. 2014 Sep 8;6	Fisman D	Early epidemic dynamics of the west african 2014 ebola outbreak: estimates derived with a simple two-parameter model.	West African Ebola outbreak	R0 = 1.6 - 2.0

25619149	10.1016/ S1473-3099(14) 71075-8	Lancet Infect Dis. 2015 Mar;15(3):320-6	Faye O	Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study.	West African Ebola outbreak	R0 (Guinea) = 2.3
25566687	10.1186/ 1742-4682-12-1	Theor Biol Med Model. 2015 Jan 6;12:1	Nishiura H	Theoretical perspectives on the infectiousness of Ebola virus disease.	No primary data on R0	
25480136	10.1016/ j.epidem. 2014.09.003	Epidemics. 2014 Dec;9:70-8	Camacho A	Potential for large outbreaks of Ebola virus disease.	DR Congo 1976 R0 = 1.34	
25467086	10.1016/ j.tmaid. 2014.10.015	Travel Med Infect Dis. 2014 Nov-Dec;12(6 Pt A):650-8	Chen T	Risk of imported Ebola virus disease in China.	Ebola virus: DR Congo, Kikwit 1995: R0 = 2.64; Uganda, Gulu 2000: R0 = 3.54	
25455986	10.1016/ S1473-3099 (14)70995-8	Lancet Infect Dis. 2014 Dec;14(12):1189-95	Lewnard JA	Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis.	West African Ebola outbreak R0 (Montserrado , Liberia = 2.49)	
19464340	10.1016/ j.biosystems. 2009.05.006	Biosystems. 2009 Oct;98(1):43-50	Zaman G	Optimal treatment of an SIR epidemic model with time delay.	No primary data on R0	
16999875	10.1017/ S095026 8806007217	Epidemiol Infect. 2007 May;135(4):610-21	Legrand J	Understanding the dynamics of Ebola epidemics.	Ebola virus: DR Congo, 1995; Uganda, 2000 R0 = 2.7	
15178190	10.1016/ j.jtbi.2004. 03.006	J Theor Biol. 2004 Jul 7;229(1):119-26	Chowell G	The basic reproductive number of Ebola and the effects of public health measures: the cases of	Ebola virus: DR Congo, 1995; Uganda, 2000 R0 (DR Congo): 1.83; R0 (Uganda) = 1.34	

				Congo and Uganda.	

"Ebola basic reproduction number"

PMID	DOI	Journal reference	First Author	Title	Comment	
28830110	10.1093/infdis/jiw356	J Infect Dis. 2016 Dec 1;214(suppl_4):S421-S426	Chowell G	Elucidating Transmission Patterns From Internet Reports: Ebola and Middle East Respiratory Syndrome as Case Studies.	West African Ebola outbreak	R0 = 2.4
28610609	10.1186/s12938-017-0370-7	Biomed Eng Online. 2017 Jun 13;16(1):79	Chang HJ	Estimation of basic reproduction number of the Middle East respiratory syndrome coronavirus (MERS-CoV) during the outbreak in South Korea, 2015.	No primary data on R0	
28269870		AMIA Annu Symp Proc. 2017 Feb 10;2016:743-752	Lee EK	A Compartmental Model for Zika Virus with Dynamic Human and Vector Populations.	No primary data on R0	
28202592	10.1098/rsif.2016.0847	J R Soc Interface. 2017 Feb;14(127)	Nieddu GT	Extinction pathways and outbreak vulnerability in a stochastic Ebola model.	No primary data on R0	
28166851	10.1017/S0950268817000164	Epidemiol Infect. 2017 Apr;145(6):1069-1094	Wong ZS	A systematic review of early modelling studies of Ebola virus disease in West Africa.	West African Ebola outbreak	R0 (overall) = 1.78; R0 (Guinea) = 1.30; R0 (Liberia) = 1.84; R0 (Sierra Leone) = 1.70; R0 (Nigeria) = 9.01
27833001	10.1016/j.mbs.2016	Math Biosci. 2017 Jan;283:48-59	Agusto FB	Mathematical model of Ebola	No primary	

	.11.002			transmission dynamics with relapse and reinfection.	data on R0	
27579053	10.1155/2016/9352725	Comput Math Methods Med. 2016;2016:9352725	Ngwa GA	A Mathematical Model with Quarantine States for the Dynamics of Ebola Virus Disease in Human Populations.	No primary data on R0	
27434164	10.1371/journal.pntd.0004867	PLoS Negl Trop Dis. 2016 Jul 19;10(7):e0004867	Krauer F	Heterogeneity in District-Level Transmission of Ebola Virus Disease during the 2013-2015 Epidemic in West Africa.	West African Ebola outbreak	R0 ranges from 0.36 to 3.37 in different districts
27405359	10.1186/s40249-016-0161-6	Infect Dis Poverty. 2016 Jul 13;5(1):72	Ahmad MD	Optimal control analysis of Ebola disease with control strategies of quarantine and vaccination.	No primary data on R0	
27049322	10.1371/journal.pone.0152438	PLoS One. 2016 Apr 6;11(4):e0152438	Guo Z	Predicting and Evaluating the Epidemic Trend of Ebola Virus Disease in the 2014-2015 Outbreak and the Effects of Intervention Measures.	No primary data on R0	
26718863	10.1016/j.jtbi.2015.11.025	J Theor Biol. 2016 Mar 7;392:99-106	Adams B	Household demographic determinants of Ebola epidemic risk.	No primary data on R0	
26618087	10.7717/peerj.1418	PeerJ. 2015 Nov 19;3:e1418	Althaus CL	Rapid drop in the reproduction number during the Ebola outbreak in the Democratic Republic of Congo.	DR Congo 2014	R0 = 5.2
26574727	10.1111/biom.12432	Biometrics. 2016 Jun;72(2):335-43	Brown GD	An empirically adjusted approach to	No primary	

				reproductive number estimation for stochastic compartmental models: A case study of two Ebola outbreaks.	data on R0	
26525597	10.7554/eLife.09015	Elife. 2015 Nov 3;4	Rosello A	Ebola virus disease in the Democratic Republic of the Congo, 1976-2014.	Outbreaks in DR Congo	R0 = 2.5 - 8.0
26515898	10.1038/srep15818	Sci Rep. 2015 Oct 30;5:15818	Shen M	Modeling the effect of comprehensive interventions on Ebola virus transmission.	West African Ebola outbreak	R0 (Liberia) = 1.80; R0 (Sierra Leone) = 1.61; R0 (Guinea) = 1.26
26451161	10.1155/2015/207105	Comput Math Methods Med. 2015;2015:207105	Liu W	Model Selection and Evaluation Based on Emerging Infectious Disease Data Sets including A/H1N1 and Ebola.	West African Ebola outbreak	R0 = 1.21 - 3.02
26402477	10.3201/eid2110.150912	10.3201/eid2110.150912	Lindblade KA	Decreased Ebola Transmission after Rapid Response to Outbreaks in Remote Areas, Liberia, 2014.	West African Ebola outbreak	R0 (Liberia) = 1.7
26347015	10.1038/srep13857	Sci Rep. 2015 Sep 8;5:13857	Xia ZQ	Modeling the transmission dynamics of Ebola virus disease in Liberia.	West African Ebola outbreak	R0 (Liberia) = 2.01
26343011	10.1016/j.antiviral.2015.08.015	Antiviral Res. 2015 Nov;123:70-7	Madelain V	Ebola virus dynamics in mice treated with favipiravir.	Out of scope (mouse study)	
26297316	10.1016/j.jtbi.2015.08.004	J Theor Biol. 2015 Nov 7;384:33-49	Browne C	Modeling contact tracing in outbreaks with application to Ebola.	West African Ebola outbreak	R0 (Guinea) = 1 - 12; R0 (Sierra Leone) = 1 - 11

26280179	10.3934/mbe.2015.12. 1055	Math Biosci Eng. 2015 Oct;12(5):1055-63	Wang XS	Ebola outbreak in West Africa: real-time estimation and multiple-wave prediction.	West African Ebola outbreak	R0 (Liberia) = 1.23; R0 (Sierra Leone) = 1.18; R0 (Guinea) = 1.12
26246849	10.1155/2015/582625	Comput Math Methods Med. 2015;2015:582625	Li Z	Dynamical Analysis of an SEIT Epidemic Model with Application to Ebola Virus Transmission in Guinea.	West African Ebola outbreak	R0 (Guinea) = 4.16
26197242	10.1371/journal.pone.0131398	PLoS One. 2015 Jul 21;10(7):e0131398	Barbarossa MV	Transmission Dynamics and Final Epidemic Size of Ebola Virus Disease Outbreaks with Varying Interventions.	West African Ebola outbreak	R(0) = 1.44
26029377	10.1038/sdata.2015.19	Sci Data. 2015 May 26;2:150019	Van Kerkhove MD	A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making.	All available data 2/2015	R(0) Ebola virus = 1.36 - 4.71; R(0) Sudan virus = 1.34 - 2.7; West Africa: R(0) Guinea = 1.71, R(0) Sierra Leone = 1.83, R(0) Liberia = 2.02
25979285	10.1016/j.epidem.2015.03.001	Epidemics. 2015 Jun;11:80-4	Althaus CL	Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control.	West African Ebola outbreak	R(0) Nigeria 9.0
25902936	10.1186/s12916-015-0318-3	BMC Med. 2015 Apr 23;13:96	Agusto FB	Mathematical assessment of the effect of traditional beliefs and customs on the transmission dynamics of the 2014 Ebola outbreaks.	West African Ebola outbreak	R(0) Guinea = 1.0 (in the absence of risky burial traditions) - 1.6 (in the presence of risky burial traditions)
25766240	10.1126/science.aaa4339	Science. 2015 Mar 13;347(6227):aaa4339	Heesterbeek H	Modeling infectious disease dynamics in the complex	No primary data on R0	

				landscape of global health.		
25642364	10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288	PLoS Curr. 2014 Sep 2;6	Althaus CL	Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa.	West African Ebola outbreak	R0 (Liberia) = 1.59; R0 (Sierra Leone) = 2.53; R0 (Guinea) = 1.51
25619149	10.1016/S1473-3099(14)71075-8	Lancet Infect Dis. 2015 Mar;15(3):320-6	Faye O	Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study.	West African Ebola outbreak	R0 (Guinea) = 2.3
25566687	10.1186/1742-4682-12-1	Theor Biol Med Model. 2015 Jan 6;12:1	Nishiura H	Theoretical perspectives on the infectiousness of Ebola virus disease.	No primary data on R0	
25480136	10.1016/j.epidem.2014.09.003	Epidemics. 2014 Dec;9:70-8	Camacho A	Potential for large outbreaks of Ebola virus disease.	DR Congo 1976	R0 = 1.34
25467086	10.1016/j.tmaid.2014.10.015	Travel Med Infect Dis. 2014 Nov-Dec;12(6 Pt A):650-8	Chen T	Risk of imported Ebola virus disease in China.	Ebola virus: DR Congo, Kikwit 1995: R0 = 2.64; Uganda, Gulu 2000: R0 = 3.54	
25455986	10.1016/S1473-3099(14)70995-8	Lancet Infect Dis. 2014 Dec;14(12):1189-95	Lewnard JA	Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis.	West African Ebola outbreak	R0 (Montserrado, Liberia) = 2.49
19464340	10.1016/j.biosystems.2009.05.006	Biosystems. 2009 Oct;98(1):43-50	Zaman G	Optimal treatment of an SIR epidemic model with time delay.	No primary data on R0	
16999875	10.1017/S0950268	Epidemiol Infect. 2007 May;135(4):610-21	Legrand J	Understanding the dynamics	Ebola virus: DR	

	806007217			of Ebola epidemics.	Congo, 1995; Uganda, 2000

Additional articles referenced by other articles

PMID	DOI	Journal reference	First Author	Title	Comment	
25347321	10.7326/M14-2255	Ann Intern Med. 2015 Jan 6;162(1):11-7	Yamin D	Effect of Ebola progression on transmission and control in Liberia.	West African Ebola outbreak	R0 (Liberia) = 1.73 (0.66 for survivors, 2.36 for non-survivors)

Articles combined

PMID	DOI	Journal reference	First Author	Title	Comment	
28830110	10.1093/infdis/jiw356	J Infect Dis. 2016 Dec 1;214(suppl_4):S421-S426	Chowell G	Elucidating Transmission Patterns From Internet Reports: Ebola and Middle East Respiratory Syndrome as Case Studies.	West African Ebola outbreak	R0 = 2.4
28610609	10.1186/s12938-017-0370-7	Biomed Eng Online. 2017 Jun 13;16(1):79	Chang HJ	Estimation of basic reproduction number of the Middle East respiratory syndrome coronavirus (MERS-CoV) during the outbreak in South Korea, 2015.	No primary data on R0	
28269870		AMIA Annu Symp Proc. 2017 Feb 10;2016:743-752	Lee EK	A Compartmental Model for Zika Virus with Dynamic Human and Vector Populations.	No primary data on R0	
28263987	10.1371/journal.pcbi.1005416	PLoS Comput Biol. 2017 Mar 6;13(3):e1005416	Saulnier E	Inferring epidemiological parameters from phylogenies using regression-ABC: A comparative study.	West African Ebola outbreak	R0 = 5.92
28202592	10.1098/rsif.2016.0847	J R Soc Interface. 2017 Feb;14(127)	Nieddu GT	Extinction pathways and outbreak vulnerability in a stochastic Ebola model.	No primary data on R0	
28166851	10.1017/S0950268817000164	Epidemiol Infect. 2017 Apr;145(6):1069-1094	Wong ZS	A systematic review of early modelling	West African	R0 (overall) = 1.78; R0 (Guinea) =

				studies of Ebola virus disease in West Africa.	Ebola outbreak	1.30; R0 (Liberia) = 1.84; R0 (Sierra Leone) = 1.70; R0 (Nigeria) = 9.01
27833001	10.1016/j.mbs.2016.11.002	Math Biosci. 2017 Jan;283:48-59	Agusto FB	Mathematical model of Ebola transmission dynamics with relapse and reinfection.	No primary data on R0	
27579053	10.1155/2016/9352725	Comput Math Methods Med. 2016;2016:9352725	Ngwa GA	A Mathematical Model with Quarantine States for the Dynamics of Ebola Virus Disease in Human Populations.	No primary data on R0	
27357339	10.1093/infdis/jiw207	J Infect Dis. 2016 Oct 15;214(suppl 3):S93-S101	Breman JG	Discovery and Description of Ebola Zaire Virus in 1976 and Relevance to the West African Epidemic During 2013-2016.	No primary data on R0	
27524644	10.1016/j.jtbi.2016.08.016	J Theor Biol. 2016 Nov 7;408:145-54	Taylor BP	Stochasticity and the limits to confidence when estimating R0 of Ebola and other emerging infectious diseases.	West African Ebola outbreak	R0 (Liberia) = 2.06; R0 (Sierra Leone) = 1.71; R0 (Guinea) = 1.24
27434164	10.1371/journal.pntd.0004867	PLoS Negl Trop Dis. 2016 Jul 19;10(7):e0004867	Krauer F	Heterogeneity in District-Level Transmission of Ebola Virus Disease during the 2013-2015 Epidemic in West Africa.	West African Ebola outbreak	R0 ranges from 0.36 to 3.37 in different districts

27405359	10.1186/s40249-016-0161-6	Infect Dis Poverty. 2016 Jul 13;5(1):72	Ahmad MD	Optimal control analysis of Ebola disease with control strategies of quarantine and vaccination.	No primary data on R0	
27049322	10.1371/journal.pone.0152438	PLoS One. 2016 Apr 6;11(4):e0152438	Guo Z	Predicting and Evaluating the Epidemic Trend of Ebola Virus Disease in the 2014-2015 Outbreak and the Effects of Intervention Measures.	No primary data on R0	
26981342	10.1016/j.phrp.2015.12.012	Osong Public Health Res Perspect. 2016 Feb;7(1):43-8	Do TS	Modeling the Spread of Ebola.	West African Ebola outbreak	R0 (Nigeria) = 2.57
26718863	10.1016/j.jtbi.2015.11.025	J Theor Biol. 2016 Mar 7;392:99-106	Adams B	Household demographic determinants of Ebola epidemic risk.	No primary data on R0	
26618087	10.7717/peerj.1418	PeerJ. 2015 Nov 19;3:e1418	Althaus CL	Rapid drop in the reproduction number during the Ebola outbreak in the Democratic Republic of Congo.	DR Congo 2014	R0 = 5.2
26574727	10.1111/biom.12432	Biometrics. 2016 Jun;72(2):335-43	Brown GD	An empirically adjusted approach to reproductive number estimation for stochastic compartmental models: A case study of two Ebola outbreaks.	No primary data on R0	
26525597	10.7554/eLife.09015	Elife. 2015 Nov 3;4	Rosello A	Ebola virus disease in the Democratic Republic of the Congo, 1976-2014.	Outbreaks in DR Congo	R0 = 2.5 - 8.0

26515898	10.1038/Srep15818	Sci Rep. 2015 Oct 30;5:15818	Shen M	Modeling the effect of comprehensive interventions on Ebola virus transmission.	West African Ebola outbreak	R0 (Liberia) = 1.80; R0 (Sierra Leone) = 1.61; R0 (Guinea) = 1.26
26451161	10.1155/2015/207105	Comput Math Methods Med. 2015;2015:207105	Liu W	Model Selection and Evaluation Based on Emerging Infectious Disease Data Sets including A/H1N1 and Ebola.	West African Ebola outbreak	R0 = 1.21 - 3.02
26402477	10.3201/eid2110.150912	10.3201/eid2110.150912	Lindblade KA	Decreased Ebola Transmission after Rapid Response to Outbreaks in Remote Areas, Liberia, 2014.	West African Ebola outbreak	R0 (Liberia) = 1.7
26347015	10.1038/Srep13857	Sci Rep. 2015 Sep 8;5:13857	Xia ZQ	Modeling the transmission dynamics of Ebola virus disease in Liberia.	West African Ebola outbreak	R0 (Liberia) = 2.01
26343011	10.1016/j.antiviral.2015.08.015	Antiviral Res. 2015 Nov;123:70-7	Madelain V	Ebola virus dynamics in mice treated with favipiravir.	Out of scope (mouse study)	
26297316	10.1016/j.jtbi.2015.08.004	J Theor Biol. 2015 Nov 7;384:33-49	Browne C	Modeling contact tracing in outbreaks with application to Ebola.	West African Ebola outbreak	R0 (Guinea) = 1 - 12; R0 (Sierra Leone) = 1 - 11
26280179	10.3934/mbe.2015.12. 1055	Math Biosci Eng. 2015 Oct;12(5):1055-63	Wang XS	Ebola outbreak in West Africa: real-time estimation and multiple-wave prediction.	West African Ebola outbreak	R0 (Liberia) = 1.23; R0 (Sierra Leone) = 1.18; R0 (Guinea) = 1.12
26246849	10.1155/2015/582625	Comput Math Methods Med. 2015;2015:582625	Li Z	Dynamical Analysis of an SEIT Epidemic Model with Application to Ebola Virus	West African Ebola outbreak	R0 (Guinea) = 4.16

				Transmission in Guinea.		
26197242	10.1371/journal.pone.0131398	PLoS One. 2015 Jul 21;10(7):e0131398	Barbarossa MV	Transmission Dynamics and Final Epidemic Size of Ebola Virus Disease Outbreaks with Varying Interventions.	West African Ebola outbreak	R(0) = 1.44
26029377	10.1038/sdata.2015.19	Sci Data. 2015 May 26;2:150019	Van Kerkhove MD	A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making.	All available data 2/2015	R(0) Ebola virus = 1.36 - 4.71; R(0) Sudan virus = 1.34 - 2.7; West Africa: R(0) Guinea = 1.71, R(0) Sierra Leone = 1.83, R(0) Liberia = 2.02
25979285	10.1016/j.epidem.2015.03.001	Epidemics. 2015 Jun;11:80-4	Althaus CL	Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control.	West African Ebola outbreak	R(0) Nigeria 9.0
25914858	10.1371/currents.outbreaks.6f7025f1271821d4c815385b08f5f80e	PLoS Curr. 2014 Oct 24;6	Volz E	Phylodynamic analysis of ebola virus in the 2014 sierra leone epidemic.	West African Ebola outbreak	R0 (Sierra Leone) = 2.40 - 3.81
25902936	10.1186/s12916-015-0318-3	BMC Med. 2015 Apr 23;13:96	Agusto FB	Mathematical assessment of the effect of traditional beliefs and customs on the transmission dynamics of the 2014 Ebola outbreaks.	West African Ebola outbreak	R(0) Guinea = 1.0 (in the absence of risky burial traditions) - 1.6 (in the presence of risky burial traditions)
25766240	10.1126/science.aaa4339	Science. 2015 Mar 13;347(6227):aaa4339	Heesterbeek H	Modeling infectious disease dynamics in the complex landscape of global health.	No primary data on R0	
25737782	10.1186/s40249-015-0043-3	Infect Dis Poverty. 2015 Feb 24;4:13	Khan A	Estimating the basic reproductive	West African	R0 (Liberia) = 1.76; R0 (Sierra

				ratio for the Ebola outbreak in Liberia and Sierra Leone.	Ebola outbreak	Leone) = 1.49
25736239	10.1038/srep08751	Sci Rep. 2015 Mar 4;5:8751	Weitz JS	Modeling post-death transmission of Ebola: challenges for inference and opportunities for control.	Out of scope (focused on post-death transmission)	
25685614	10.1371/currents.outbreaks.c6efe8274dc55274f05cbcb62bbe6070	PLoS Curr. 2014 Nov 13;6	Kiskowski MA	A three-scale network model for the early growth dynamics of 2014 west Africa ebola epidemic.	West African Ebola outbreak	R0 = 2.35
25642370	10.1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f	PLoS Curr. 2014 Oct 6;6	Stadler T	Insights into the early epidemic spread of ebola in sierra leone provided by viral sequence data.	West African Ebola outbreak	R0 (Sierra Leone) = 2.58
25642364	10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288	PLoS Curr. 2014 Sep 2;6	Althaus CL	Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa.	West African Ebola outbreak	R0 (Liberia) = 1.59; R0 (Sierra Leone) = 2.53; R0 (Guinea) = 1.51
25642360	10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5	PLoS Curr. 2014 Sep 2;6	Gomes MF	Assessing the international spreading risk associated with the 2014 west african ebola outbreak.	West African Ebola outbreak	R0 = 1.5 - 2.0
25642358	10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571	PLoS Curr. 2014 Sep 8;6	Fisman D	Early epidemic dynamics of the west african 2014 ebola outbreak: estimates derived with a simple two-	West African Ebola outbreak	R0 = 1.6 - 2.0

				parameter model.		
25619149	10.1016/S1473-3099(14)71075-8	Lancet Infect Dis. 2015 Mar;15(3):320-6	Faye O	Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study.	West African Ebola outbreak	R0 (Guinea) = 2.3
25566687	10.1186/1742-4682-12-1	Theor Biol Med Model. 2015 Jan 6;12:1	Nishiura H	Theoretical perspectives on the infectiousness of Ebola virus disease.	No primary data on R0	
25495064	10.4161/21505594.2014.976514	Virulence. 2014;5(8):825-7	Alizon S	Quantifying the epidemic spread of Ebola virus (EBOV) in Sierra Leone using phylodynamics.	West African Ebola outbreak	R0 (Sierra Leone) = 1.26
25480136	10.1016/j.epidem.2014.09.003	Epidemics. 2014 Dec;9:70-8	Camacho A	Potential for large outbreaks of Ebola virus disease.	DR Congo 1976	R0 = 1.34
25467086	10.1016/j.tmaid.2014.10.015	Travel Med Infect Dis. 2014 Nov-Dec;12(6 Pt A):650-8	Chen T	Risk of imported Ebola virus disease in China.	Ebola virus: DR Congo, Kikwit 1995: R0 = 2.64; Uganda, Gulu 2000: R0 = 3.54	
25455986	10.1016/S1473-3099(14)70995-8	Lancet Infect Dis. 2014 Dec;14(12):1189-95	Lewnard JA	Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis.	West African Ebola outbreak	R0 (Montserrado, Liberia) = 2.49
25347321	10.7326/M14-2255	Ann Intern Med. 2015 Jan 6;162(1):11-7	Yamin D	Effect of Ebola progression on transmission and control in Liberia.	West African Ebola outbreak	R0 (Liberia) = 1.73 (0.66 for survivors, 2.36 for nonsurvivors)

25244186	10.1056/NEJMoa1411100	N Engl J Med. 2014 Oct 16;371(16):1481-95	WHO Ebola Response Team	Ebola virus disease in West Africa--the first 9 months of the epidemic and forward projections.	West African Ebola outbreak	R0 (Liberia) = 1.83; R0 (Sierra Leone) = 2.02; R0 (Guinea) = 1.71
25217532	10.7554/eLife.03908	Elife. 2014 Sep 12;3:e03908	House T	Epidemiological dynamics of Ebola outbreaks.	No primary data on R0	
19464340	10.1016/j.biosystems.2009.05.006	Biosystems. 2009 Oct;98(1):43-50	Zaman G	Optimal treatment of an SIR epidemic model with time delay.	No primary data on R0	
16999875	10.1017/S0950268806007217	Epidemiol Infect. 2007 May;135(4):610-21	Legrand J	Understanding the dynamics of Ebola epidemics.	Ebola virus: DR Congo, 1995; Uganda, 2000	R0 = 2.7
15178190	10.1016/j.jtbi.2004.03.006	J Theor Biol. 2004 Jul 7;229(1):119-26	Chowell G	The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda.	Ebola virus: DR Congo, 1995; Uganda, 2000	R0 (DR Congo): 1.83; R0 (Uganda) = 1.34

Relevant articles

PMID	DOI	Journal reference	First Author	Title	Outbreak data	Virus	
28830110	10.1093/infdis/jiw356	J Infect Dis. 2016 Dec 1;214(suppl_4):S421-S426	Chowell G	Elucidating Transmission Patterns From Internet Reports: Ebola and Middle East Respiratory Syndrome as Case Studies.	West African Ebola outbreak	Ebola virus	R0 = 2.4
28263987	10.1371/journal.pcbi.1005416	PLoS Comput Biol. 2017 Mar 6;13(3):e1005416	Saulnier E	Inferring epidemiological parameters from phylogenies using regression-ABC: A comparative study.	West African Ebola outbreak	Ebola virus	R0 = 5.92
28166851	10.1017/S0950268817000164	Epidemiol Infect. 2017 Apr;145(6):1069-1094	Wong ZS	A systematic review of early modelling studies of Ebola virus disease in West Africa.	West African Ebola outbreak	Ebola virus	R0 (overall) = 1.78; R0 (Guinea) = 1.30; R0 (Liberia) = 1.84; R0 (Sierra Leone) = 1.70; R0 (Nigeria) = 9.01
27524644	10.1016/j.jtbi.2016.08.016	J Theor Biol. 2016 Nov 7;408:145-54	Taylor BP	Stochasticity and the limits to confidence when estimating R0 of Ebola and other emerging infectious diseases.	West African Ebola outbreak	Ebola virus	R0 (Liberia) = 2.06; R0 (Sierra Leone) = 1.71; R0 (Guinea) = 1.24
27434164	10.1371/journal.pntd.0004867	PLoS Negl Trop Dis. 2016 Jul 19;10(7):e0004867	Krauer F	Heterogeneity in District-Level Transmission of Ebola Virus Disease during the 2013-2015	West African Ebola outbreak	Ebola virus	R0 ranges from 0.36 to 3.37 in different districts

				Epidemic in West Africa.			
26981342	10.1016/j.phrp.2015.12.012	Osong Public Health Res Perspect. 2016 Feb;7(1):43-8	Do TS	Modeling the Spread of Ebola.	West African Ebola outbreak	Ebola virus	R0 (Nigeria) = 2.57
26618087	10.7717/peerj.1418	PeerJ. 2015 Nov 19;3:e1418	Althaus CL	Rapid drop in the reproduction number during the Ebola outbreak in the Democratic Republic of Congo.	DR Congo 2014	Ebola virus	R0 = 5.2
26525597	10.7554/eLife.09015	Elife. 2015 Nov 3;4	Rosello A	Ebola virus disease in the Democratic Republic of the Congo, 1976-2014.	Outbreaks in DR Congo	Ebola virus	R0 = 2.5 - 8.0
26515898	10.1038/srep15818	Sci Rep. 2015 Oct 30;5:15818	Shen M	Modeling the effect of comprehensive interventions on Ebola virus transmission.	West African Ebola outbreak	Ebola virus	R0 (Liberia) = 1.80; R0 (Sierra Leone) = 1.61; R0 (Guinea) = 1.26
26451161	10.1155/2015/207105	Comput Math Methods Med. 2015;2015:207105	Liu W	Model Selection and Evaluation Based on Emerging Infectious Disease Data Sets including A/H1N1 and Ebola.	West African Ebola outbreak	Ebola virus	R0 = 1.21 - 3.02
26402477	10.3201/eid2110.150912	10.3201/eid2110.150912	Lindblade KA	Decreased Ebola Transmission after Rapid Response to Outbreaks in Remote Areas, Liberia, 2014.	West African Ebola outbreak	Ebola virus	R0 (Liberia) = 1.7
26347015	10.1038/srep13857	Sci Rep. 2015 Sep 8;5:13857	Xia ZQ	Modeling the transmission dynamics of Ebola virus disease in Liberia.	West African Ebola outbreak	Ebola virus	R0 (Liberia) = 2.01
26297316	10.1016/j.jtbi.2015.	J Theor Biol. 2015 Nov 7;384:33-49	Browne C	Modeling contact	West African	Ebola virus	R0 (Guinea) = 1 - 12; R0

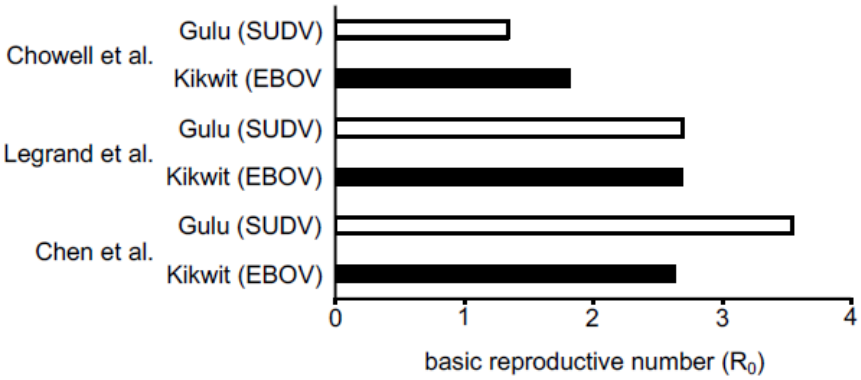
	08.004			tracing in outbreaks with application to Ebola.	Ebola outbreak		(Sierra Leone) = 1 - 11
26280179	10.3934/mbe.2015.12. 1055	Math Biosci Eng. 2015 Oct;12(5):1055-63	Wang XS	Ebola outbreak in West Africa: real-time estimation and multiple-wave prediction.	West African Ebola outbreak	Ebola virus	R0 (Liberia) = 1.23; R0 (Sierra Leone) = 1.18; R0 (Guinea) = 1.12
26246849	10.1155/2015/582625	Comput Math Methods Med. 2015;2015:582625	Li Z	Dynamical Analysis of an SEIT Epidemic Model with Application to Ebola Virus Transmission in Guinea.	West African Ebola outbreak	Ebola virus	R0 (Guinea) = 4.16
26197242	10.1371/journal.pone.0131398	PLoS One. 2015 Jul 21;10(7):e0131398	Barbarossa MV	Transmission Dynamics and Final Epidemic Size of Ebola Virus Disease Outbreaks with Varying Interventions.	West African Ebola outbreak	Ebola virus	R(0) = 1.44
26029377	10.1038/sdata.2015.19	Sci Data. 2015 May 26;2:150019	Van Kerkhove MD	A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making.	All available data 2/2015	Sudan virus	R(0) Ebola virus = 1.36 - 4.71; R(0) Sudan virus = 1.34 - 2.7; West Africa: R(0) Guinea = 1.71, R(0) Sierra Leone = 1.83, R(0) Liberia = 2.02
25979285	10.1016/j.epidem.2015.03.001	Epidemics. 2015 Jun;11:80-4	Althaus CL	Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control.	West African Ebola outbreak	Ebola virus	R(0) Nigeria 9.0
25914858	10.1371/currents.outbreaks.6f7025f1271821d4c815385b08f5f80e	PLoS Curr. 2014 Oct 24;6	Volz E	Phylodynamic analysis of ebola virus in the 2014 sierra leone epidemic.	West African Ebola outbreak	Ebola virus	R0 (Sierra Leone) = 2.40 - 3.81

25902936	10.1186/s12916-015-0318-3	BMC Med. 2015 Apr 23;13:96	Agusto FB	Mathematical assessment of the effect of traditional beliefs and customs on the transmission dynamics of the 2014 Ebola outbreaks.	West African Ebola outbreak	Ebola virus	R(0) Guinea = 1.0 (in the absence of risky burial traditions) - 1.6 (in the presence of risky burial traditions)
25737782	10.1186/s40249-015-0043-3	Infect Dis Poverty. 2015 Feb 24;4:13	Khan A	Estimating the basic reproductive ratio for the Ebola outbreak in Liberia and Sierra Leone.	West African Ebola outbreak	Ebola virus	R0 (Liberia) = 1.76; R0 (Sierra Leone) = 1.49
25685614	10.1371/currents.outbreaks.c6efe8274dc55274f05cbcb62bbe6070	PLoS Curr. 2014 Nov 13;6. pii: ecurrents.outbreaks.c6efe8274dc55274f05cbcb62bbe6070	Kiskowski MA	A three-scale network model for the early growth dynamics of 2014 west Africa ebola epidemic.	West African Ebola outbreak	Ebola virus	R0 = 2.35
25642370	10.1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f	PLoS Curr. 2014 Oct 6;6. pii: ecurrents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f	Stadler T	Insights into the early epidemic spread of ebola in sierra leone provided by viral sequence data.	West African Ebola outbreak	Ebola virus	R0 (Sierra Leone) = 2.58
25642364	10.1371/currents.outbreaks.91afb5e0f279e7f29e705609525 5b288	PLoS Curr. 2014 Sep 2;6. pii: ecurrents.outbreaks.91afb5e0f279e7f29e7056095255b288	Althaus CL	Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa.	West African Ebola outbreak	Ebola virus	R0 (Liberia) = 1.59; R0 (Sierra Leone) = 2.53; R0 (Guinea) = 1.51
25642360	10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5	PLoS Curr. 2014 Sep 2;6. pii: ecurrents.outbreaks.cd818f63d40e24aef769dda7df9e0da5	Gomes MF	Assessing the international spreading risk associated with the 2014 west african ebola outbreak.	West African Ebola outbreak	Ebola virus	R0 = 1.5 - 2.0

25642358	10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571	PLoS Curr. 2014 Sep 8;6. pii: ecurrents.outbreaks.89c0d3783f36958d96ebbae97348d571	Fisman D	Early epidemic dynamics of the west african 2014 ebola outbreak: estimates derived with a simple two-parameter model.	West African Ebola outbreak	Ebola virus	R0 = 1.6 - 2.0 (up to 3.5)
25619149	10.1016/S1473-3099(14)71075-8	Lancet Infect Dis. 2015 Mar;15(3):320-6	Faye O	Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study.	West African Ebola outbreak	Ebola virus	R0 (Guinea) = 2.3
25495064	10.4161/21505594.2014.976514	Virulence. 2014;5(8):825-7	Alizon S	Quantifying the epidemic spread of Ebola virus (EBOV) in Sierra Leone using phylodynamics.	West African Ebola outbreak	Ebola virus	R0 (Sierra Leone) = 1.26
25480136	10.1016/j.epidem.2014.09.003	Epidemics. 2014 Dec;9:70-8	Camacho A	Potential for large outbreaks of Ebola virus disease.	DR Congo 1976	Ebola virus	R0 = 1.34
25467086	10.1016/j.tmaid.2014.10.015	Travel Med Infect Dis. 2014 Nov-Dec;12(6 Pt A):650-8	Chen T	Risk of imported Ebola virus disease in China.	DR Congo, Kikwit 1995; Uganda, Gulu 2000	Sudan virus	Kikwit 1995 (Ebola virus): R0 = 2.64; Gulu 2000 (Sudan virus): R0 = 3.54
25455986	10.1016/S1473-3099(14)70995-8	Lancet Infect Dis. 2014 Dec;14(12):1189-95	Lewnard JA	Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis.	West African Ebola outbreak	Ebola virus	R0 (Montserrado, Liberia) = 2.49
25347321	10.7326/M14-2255	Ann Intern Med. 2015 Jan 6;162(1):11-7	Yamin D	Effect of Ebola progression on	West African Ebola outbreak	Ebola virus	R0 (Liberia) = 1.73 (0.66 for survivors, 2.36 for

				transmission and control in Liberia.			nonsurvivors)
25244186	10.1056/NEJMoa1411100	N Engl J Med. 2014 Oct 16;371(16):1481-95	WHO Ebola Response Team	Ebola virus disease in West Africa--the first 9 months of the epidemic and forward projections.	West African Ebola outbreak	Ebola virus	R0 (Liberia) = 1.83; R0 (Sierra Leone) = 2.02; R0 (Guinea) = 1.71
16999875	10.1017/S0950268806007217	Epidemiol Infect. 2007 May;135(4):610-21	Legrand J	Understanding the dynamics of Ebola epidemics.	DR Congo, Kikwit 1995; Uganda, Gulu 2000	Sudan virus	Kikwit 1995 (Ebola virus): R0 = 2.7; Gulu 2000 (Sudan virus): R0 = 2.7
15178190	10.1016/j.jtbi.2004.03.006	J Theor Biol. 2004 Jul 7;229(1):119-26	Chowell G	The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda.	DR Congo, Kikwit 1995; Uganda, Gulu 2000	Sudan virus	Kikwit 1995 (Ebola virus): R0 = 1.83; Gulu 2000 (Sudan virus): R0 = 1.34

Data Sheet 2



Data Sheet 2. Basic reproductive numbers (R_0) determined for the Ebola virus outbreak in Kikwit (DR Congo, 1995) and the Sudan virus outbreak in Gulu (Uganda, 2000/ 2001) as determined in three different studies (15-17).

Data Sheet 3. Herd immunity thresholds (I_c) values based on a range of basic reproductive number (R_0) values that cover the range reported for Ebolaviruses (numerical data underlying Figure 3A).

R_0	$I_c = 1 - (1/R_0)$
1	0
1.25	20.0
1.5	33.3
1.75	42.9
2	50.0
2.5	60.0
3	66. 7
4	75.0
5	80.0
6	83.3
7	85.7
8	87.5
9	88. 9
10	90.0
11	90.9
12	91. 7
15	93.3
20	95.0

Data Sheet 4. Ebolavirus vaccine candidates in clinical testing (Keshwara et al., Ledgerwood et al.)

Vaccine candidate	Clinical phase
rVSV-ZEBOV (recombinant attenuated replication-competent vesicular stomatitis virus vector expressing Ebola virus GP)	Phase 3
Recombinant replication-deficient adenovirus serotype 5 (rAd5) vector expressing Ebola virus and Sudan virus GP	Phase 2
Recombinant replication-deficient adenovirus serotype 5 (rAd5) vector expressing GP from the 2014 Ebola virus outbreak in Guinea	Phase 1/2 (approved in China)
ChAd3-EBO-Z (recombinant replication-deficient chimpanzee adenovirus 3 vector expressing Ebola virus GP)	Phase 1/2
Ad26-EBOV (human recombinant adenovirus 26 vector expressing Ebola virus GP) / MVA-BN Filo (replication-deficient modified vaccinia Ankara expressing Ebola virus GP, Sudan virus GP, Marburg virus GP, and Taï Forest virus NP) boost	Phase 1

Keshwara R, Johnson RF, Schnell MJ. Toward an effective Ebola virus vaccine. Annu Rev Med (2017) 68:371–86. doi:10.1146/annurev-med-051215-030919

Ledgerwood JE, Costner P, Desai N, Holman L, Enama ME, Yamshchikov G, et al. A replication defective recombinant Ad5 vaccine expressing Ebola virus GP is safe and immunogenic in healthy adults. Vaccine (2010) 29:304–13. doi:10.1016/j.vaccine.2010.10.037

Data Sheet 5. Populations and percentage of rural populations of African countries that have been affected by Ebolavirus outbreaks.

Country	Population	% Rural population
Côte d'Ivoire	23,254,184[1]	46[1]
Democratic Republic Congo	79,722,624	58
Gabon	1,763,142	13
Guinea	12,947,122	63
Liberia	4,615,222	50
Mali	18,134,835	60
Nigeria	186,987,563	52
Republic of the Congo	4,740,992	35
Senegal	15,589,485	56
Sierra Leone	6,592,102	60
South Africa	54,978,907	35
South Sudan	12,733,427	81
Uganda	40,322,768	84
	Total: 462,382,373	

[1] Data derived from the world bank website (http://www.worldbank.org, last accessed on 17th October 2017)

Appendix 2: Chapter 3 Supplementary Material

Is the Bombali virus pathogenic in humans?

Henry J Martell*, Stuart G Masterson*, Jake E. McGreig, Martin Michaelis§, Mark N
Wass§

Industrial Biotechnology Centre and School of Biosciences, University of Kent, Canterbury, Kent, CT2 7NJ, UK.

§To whom correspondence should be addressed
m.michaelis@kent.ac.uk +44 1227 827804
m.n.wass@kent.ac.uk +44 1227 827626

*equal contribution

Supplementary Material

Supplementary Methods

Ebolavirus Nomenclature

The virus nomenclature in this report follows the recommendations set by Kuhn et al., *Filoviridae* is the family, in the order *Mononegavirales*. Both of these terms are always italicised when referenced. The genus is known as *Ebolavirus*, and is only italicised when referring to the genus, but not when referring to

physical viruses, virus properties, or constituent virus parts such as proteins or genomes. Ebolavirus Disease (EVD) also remains unitalicised. The five individual species are subsequently referred to as *Bundibugyo ebolavirus* (type virus: *Bundibugyo virus*, BDBV), *Reston ebolavirus* (type virus: *Reston virus*, RESTV), *Sudan ebolavirus* (type virus: *Sudan virus*, SUDV), *Tai Forest ebolavirus* (type virus: *Tai Forest virus*, TAFV) and *Zaire ebolavirus* (type virus: *Ebola virus*, EBOV) (Kuhn et al. 2014).

Collection of *Ebolavirus* Genomes

All ebolavirus genome sequences were obtained from the National Center for Biotechnology Information (NCBI) (Brister et al. 2015), the Virus Pathogen Resource (ViPR) (Pickett et al. 2012), as well as taken from a repository obtained from (Urbanowicz et al. 2016), available here: https://github.com/ebov/space-time. Duplicate sequences present in >1 of the databases were filtered out during initial sample collection, with the order of source preference being NCBI > ViPR > Urbanowicz et al. Supplementary Table 1 summarises the sources used for the set of *Ebolavirus* genomes.

Two Bombali virus sequences were obtained from NCBI (GenBank IDs MF319185, MF319186 respectively). MF319185 was used as the reference sequence and the residues from this were used whenever the two sequences had different amino acids at any given SDP.

Genome Processing and Filtering

For each sample genome, open reading frames (ORFs) were identified using the EMBOSS getorf tool (Rice et al. 2000), and the resulting ORFs were matched to the UniProt *Ebola virus* reference protein sequences using BLAST (Camacho et al. 2009; Bateman et al. 2015). The top ORF hit for each *Ebola virus* protein was then used as the protein sequence for that sample, for all proteins except GP. The ebolavirus GP protein is the result of mRNA editing, due to a slippery 7A-motif that is translated as eight A nucleotides, with the regular ORF containing an early stop codon (Volchkov et al. 1995). The GP ORF hits were

further processed by editing the identified ORF to swap the 7A-motif for 8 A nucleotides. ORFs were then reidentified for the edited sequence and BLAST was used to search against the *Ebola virus* reference proteins.

After these steps, ebolavirus samples that did not have a BLAST hit with >90% coverage compared to the *Ebola virus* reference protein, for each of the seven proteins, was removed. Samples with poor metadata, such as unknown host or data were also removed (partial dates were allowed, e.g. if only the year of collection was known). This was to ensure that only high-quality samples were analysed, as incomplete data could affect subsequent analyses. Supplementary Table 2 summarises the samples that were removed in this step, and a full list of the samples that were retained can be found in Supplementary File 2.

Genome Sequence Alignment and Identification of Specificity Determining Positions

Clustal Omega was used to generate sequence alignments for each of the ebolavirus proteins (Sievers et al. 2011), and the individual sequence identities were obtained from the Clustal Omega output. Jensen-Shannon divergence scores were then calculated for each protein (Capra & Singh 2007). S3det was used in supervised mode to find specificity determining positions (SDPs), with sequences assigned to two groups prior to running S3Det (Rausell et al. 2010). Group 1 contained all of the human pathogenic sequences (*Ebola virus*, *Sudan virus*, *Bundibugyo virus*, and *Tai Forest virus*) and group 2 contained all of the human non-pathogenic sequences (*Reston virus*). All SDPs are referred to by the amino acid in the *Ebola virus* protein sequence, the position in the *Ebola virus* reference protein sequence, and the corresponding amino acid in the *Reston virus* protein sequence,
e.g. G20A meaning at position 20 *Ebola virus* has a glycine residue and *Reston virus* has an alanine residue.

Structural Analysis of SDPs

All available ebolavirus protein structures were downloaded from the Protein Databank (PDB) (Berman et al. 2000), and SDPs were mapped to the highest quality structure available, based on structure resolution and coverage. Multimeric protein structures were used to analyse the effects of the SDPs on partner interactions.

Where structures were unavailable from the PDB, proteins were modelled using Phyre2 with default settings (Kelly et al. 2015). Supplementary Table 3 summarises the structures used for analysis. PyMOL (https://pymol.org/2/) was used to visualise the identified SDPs in the protein structures and generate images.

For the subset of SDPs mapped to structures, multiple computational tools were used to predict the functional effects of each SDP. mCSM was used to predict the effect on protein stability (Pires et al. 2014), where the change in stability ($\Delta\Delta G$) is measured in kcal/mol, with negative values being destabilising and positive values being stabilising. Relative solvent accessibility of SDP residues was also calculated using mCSM. BLOSUM62 scores were assigned to each SDP, with the score calculated for the change between the *Ebola virus* sequence and the *Reston virus* sequence wherever there was variation amongst the pathogenic species.

Phylogenetic Trees

Whole genome alignments were performed using Clustal Omega, whilst the alignments for each protein were performed using TranslatorX (Abascal et al. 2010), which aligns protein-coding nucleotide sequences based on their corresponding amino acid translations. Bayesian trees, for each protein and genome, were then produced using BEAUTI and BEAST 1.10.4 (Suchard et al. 2018) , performed on the CIPRES Science Gateway (Miller et al. 2010). The consensus tree for each set of 10,000 trees was then calculated using TreeAnnotator, and the nodes were labelled with the posterior probabilities. These trees were then analysed and plotted in R using the "ape" package.

The Maximum Phylogenetic trees were produced using RaxML8.2.10 on the CIPRES Science Gateway, with 1000 Bootstrap replicates run to obtain the best scoring ML tree for each set of sequences. These were plotted and annotated in FigTree [http://tree.bio.ed.ac.uk-/software/figtree/]. Species labels were added to all trees using Inkscape.

References

Abascal F, Zardoya R, Telford MJ, 2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research,* 38(W7) pp.13.

Bateman, A. et al., 2015. UniProt: A hub for protein information. *Nucleic Acids Research*, 43(D1), pp.D204–D212.

Berman, H.M. et al., 2000. The protein data bank. *Nucleic acids research*, 28(1), pp.235–242.

Brister, J.R. et al., 2015. NCBI viral Genomes resource. *Nucleic Acids Research*, 43(D1), pp.D571–D577.

Camacho, C. et al., 2009. BLAST+: architecture and applications. *BMC bioinformatics*, 10, p.421.

Capra, J.A. & Singh, M., 2007. Predicting functionally important residues from sequence conservation. *Bioinformatics*, 23(15), pp.1875–1882.

Kelly, L.A. et al., 2015. The Phyre2 web portal for protein modelling, prediction, and analysis. *Nature Protocols*, 10(6), pp.845–858.

Kuhn, J.H. et al., 2014. Nomenclature- and database-compatible names for the Two Ebola virus variants that emerged in guinea and the Democratic Republic of the Congo in 2014. *Viruses*, 6(11), pp.4760–4799.

Miller, M. A., Pfeiffer, W. & Schwartz, T., 2010 In *Proceedings of the Gateway Computing Environments Workshop (GCE)* 1–8

Pickett, B.E. et al., 2012. ViPR: An open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research*, 40(D1), pp.593–598.

Pires, D.E. V, Ascher, D.B. & Blundell, T.L., 2014. mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, 30(3), pp.335–342.

Rausell, A. et al., 2010. Protein interactions and ligand binding: From protein subfamilies to functional specificity. *Proceedings of the National Academy of Sciences*, 107(5), pp.1995–2000.

Rice, P., Longden, I. & Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, 16(1), pp.276–277.

Sievers, F. et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), p.539.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ & Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10 *Virus Evolution* **4**, vey016.

Urbanowicz, R.A. et al., 2016. Human Adaptation of Ebola Virus during the West African Outbreak. *Cell*, 167(4), pp.1079–1087.

Volchkov, V.E. et al., 1995. GP mRNA of Ebola Virus Is Edited by the Ebola Virus Polymerase and by T7 and Vaccinia Virus Polymerases. *Virology*, 214(2), pp.421–430.

Supplementary Figures

Supplementary Figure 1. Phylogenetic tree of the Ebolavirus genomes and individual proteins. Bayesian and Maximum Likelihood phylogenetic trees are shown for the Ebolavirus genomes and each of the Ebolavirus proteins. A) genome Bayesian tree. B) genome maximum likelihood tree, C) Bayesian tree for protein L, D) Maximum likelihood tree for protein L, E) Bayesian tree for protein GP, F)Maximum likelihood tree for protein GP, G) Bayesian tree for protein NP, H), Maximum likelihood tree for protein NP, I) Bayesian tree for protein VP24, J) Maximum likelihood tree for protein VP24, K) Bayesian tree for protein VP30, L) Maximum likelihood tree for protein VP30, M) Bayesian tree for protein VP35, N) Maximum likelihood tree for protein VP35, O) Bayesian tree for protein VP40. P) Maximum likelihood tree for protein VP40.

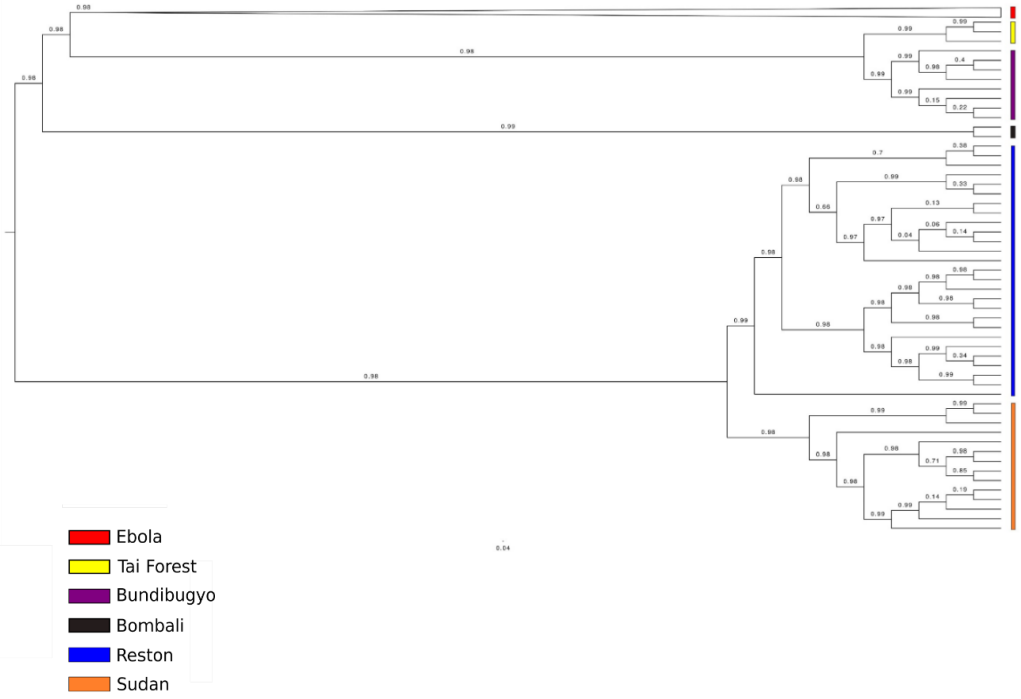


Fig S1A. Genome Bayesian Tree

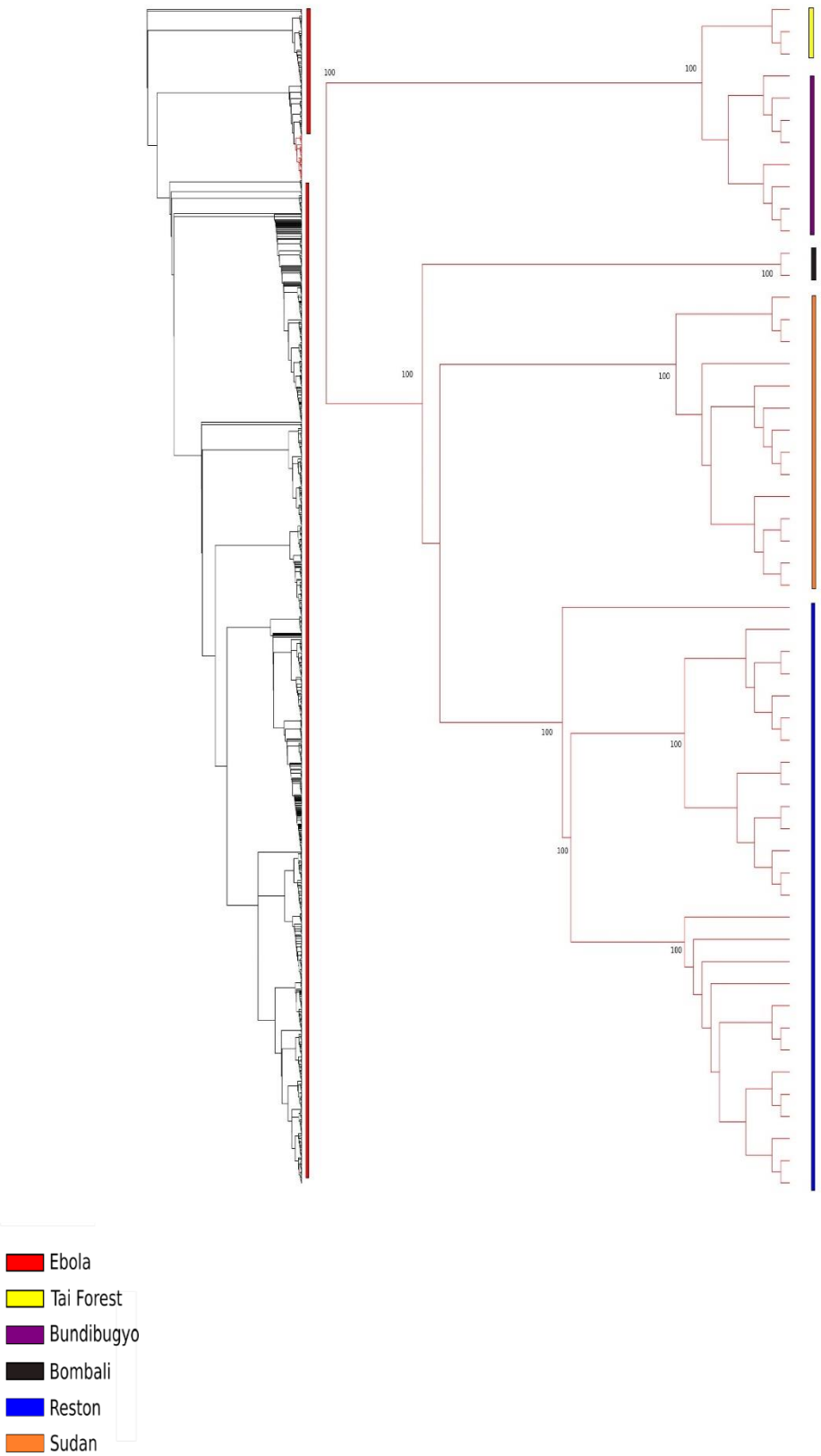
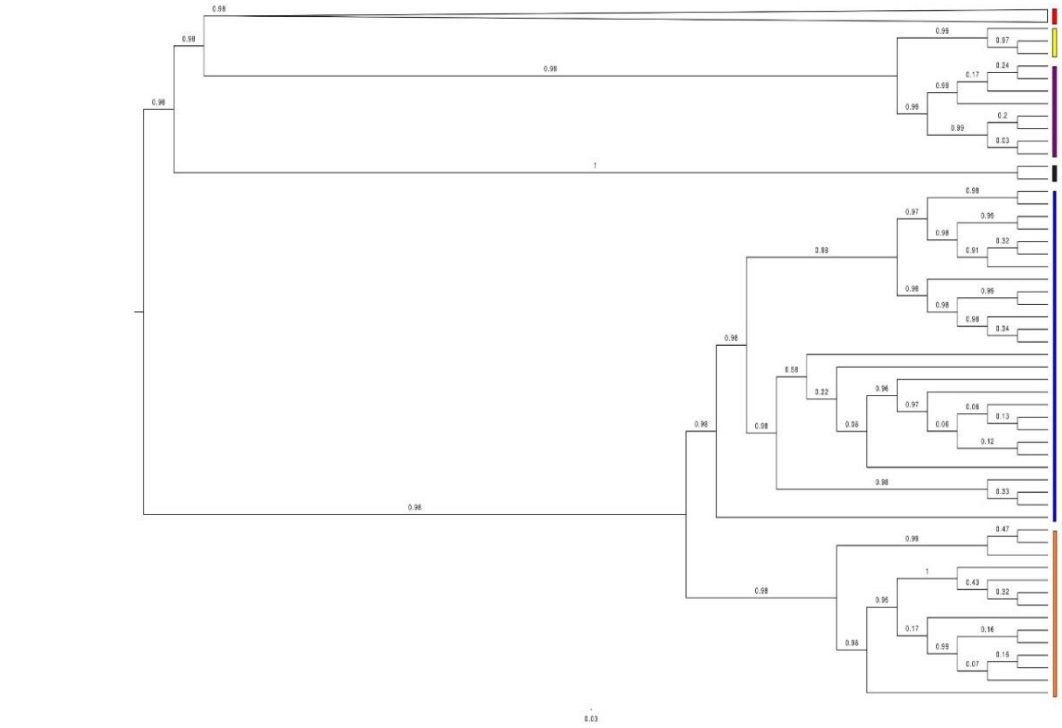


Fig S1B. Genome Maximum Likelihood



Ebola
Tai Forest
Bundibugyo
Bombali
Reston
Sudan

Fig S1C. L Bayesian Tree

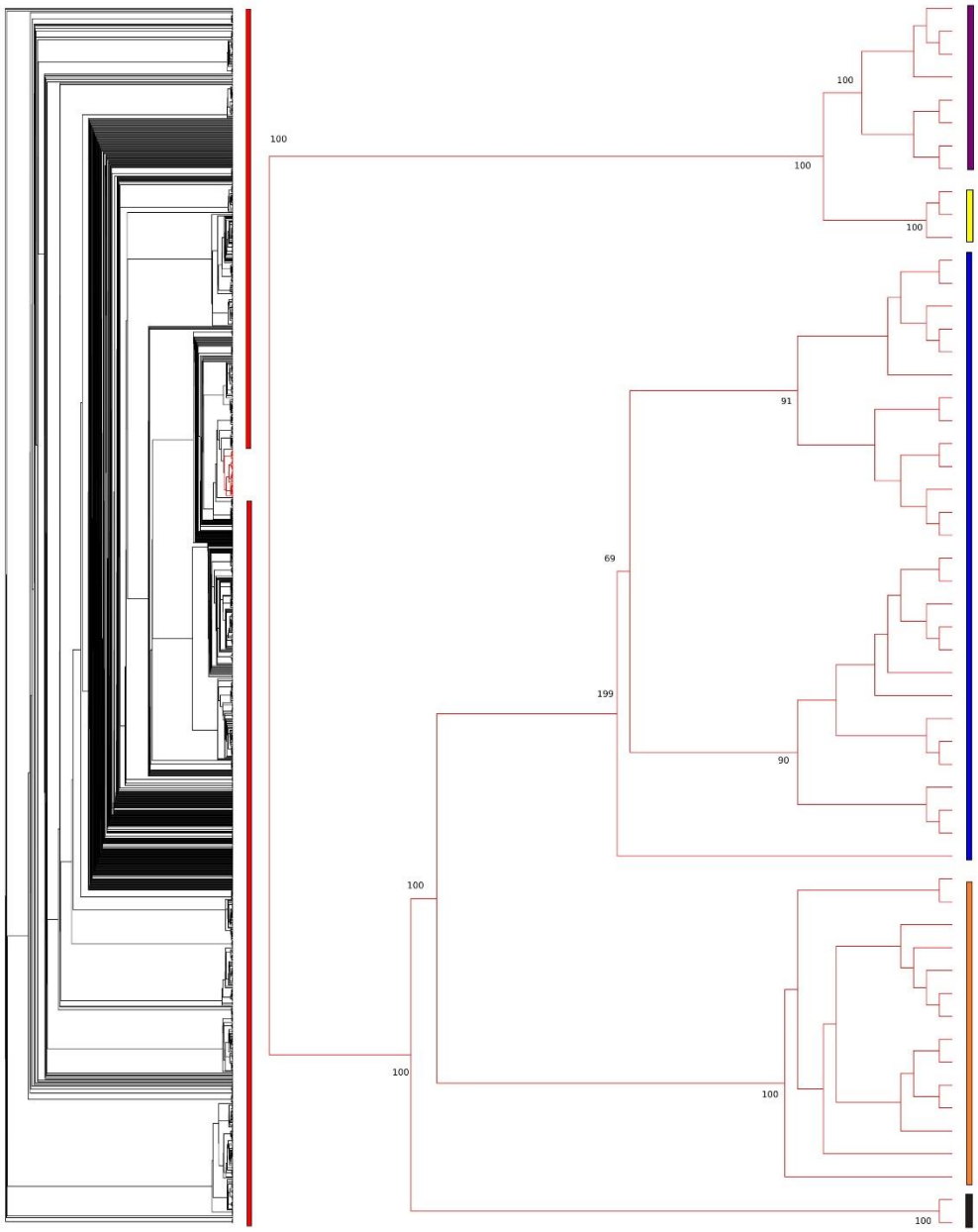


Fig S1D. L Maximum Likelihood Tree

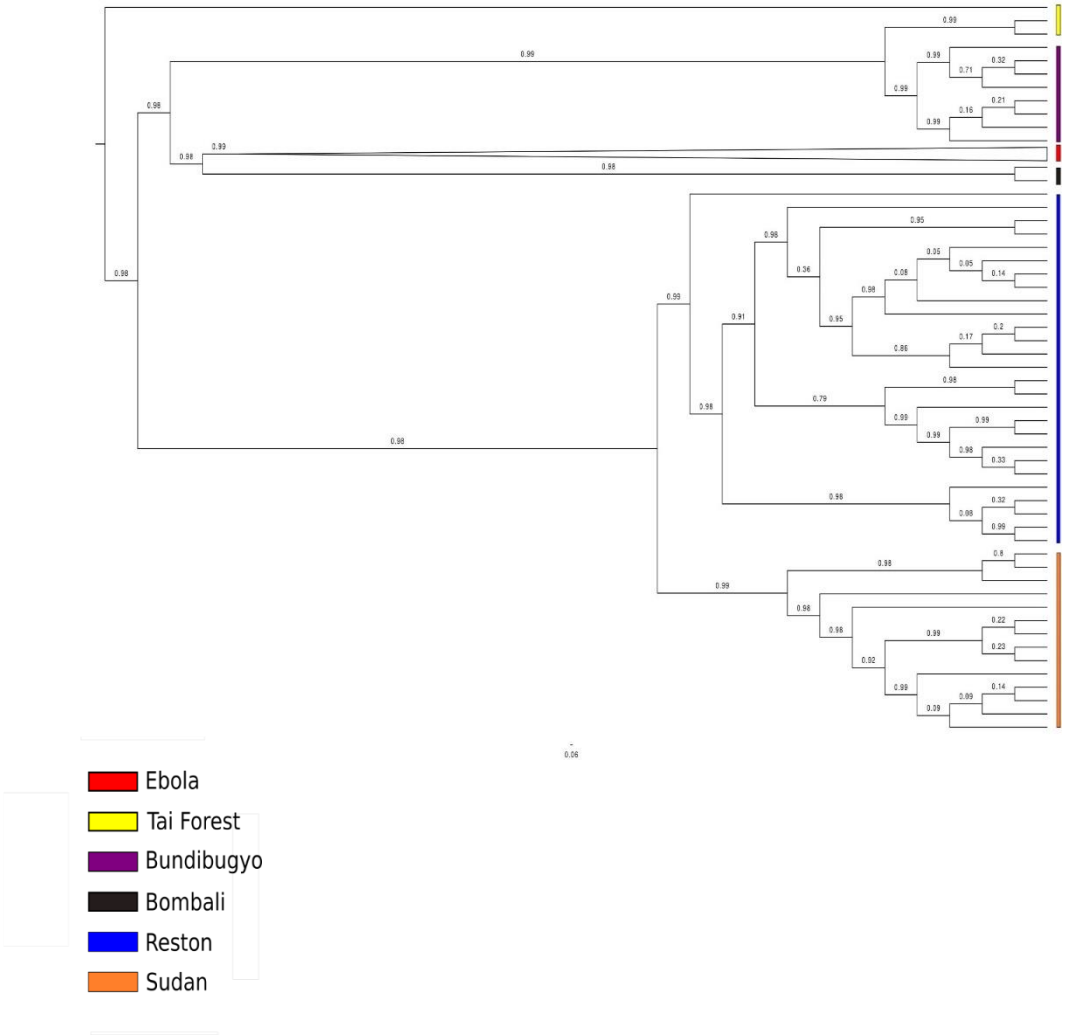


Fig S1E. GP Bayesian Tree

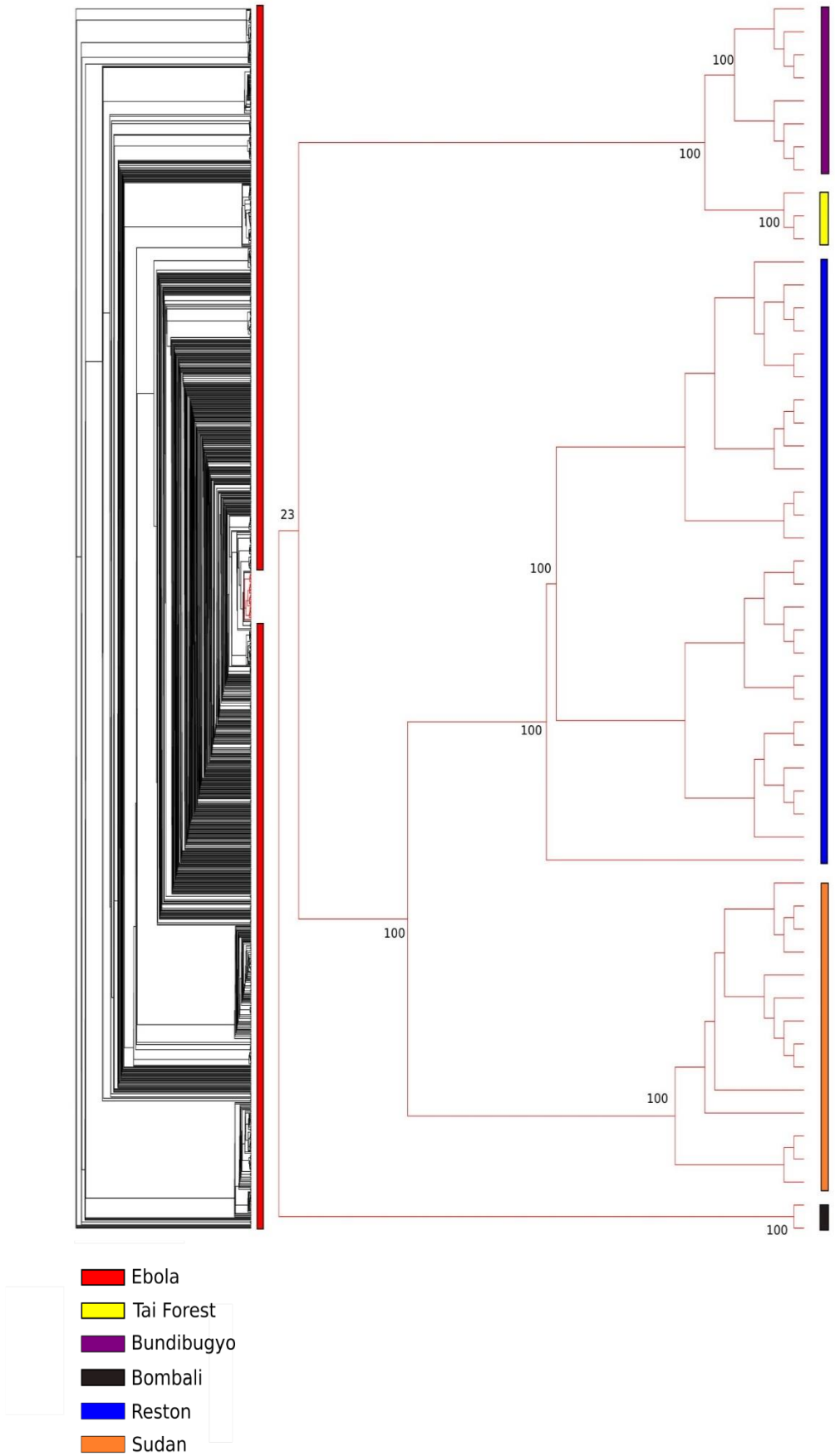
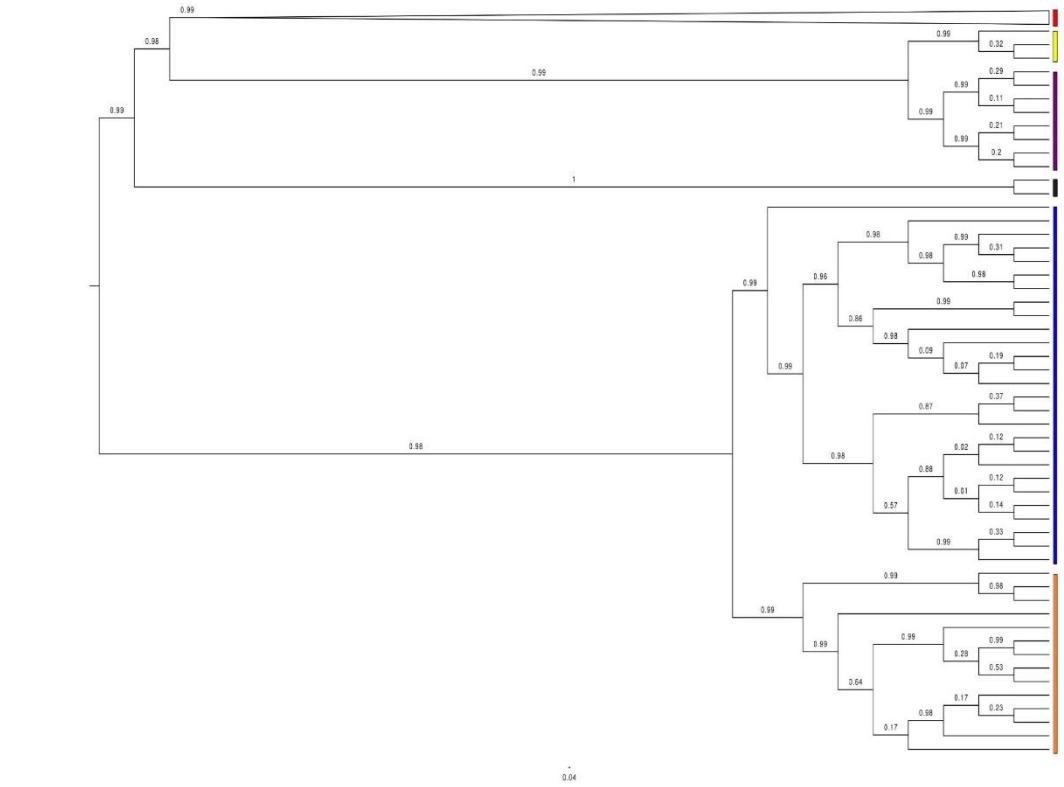


Fig S1F. GP Maximum Likelihood



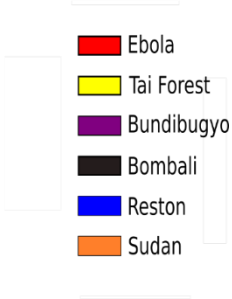


Fig S1G. NP Bayesian Tree

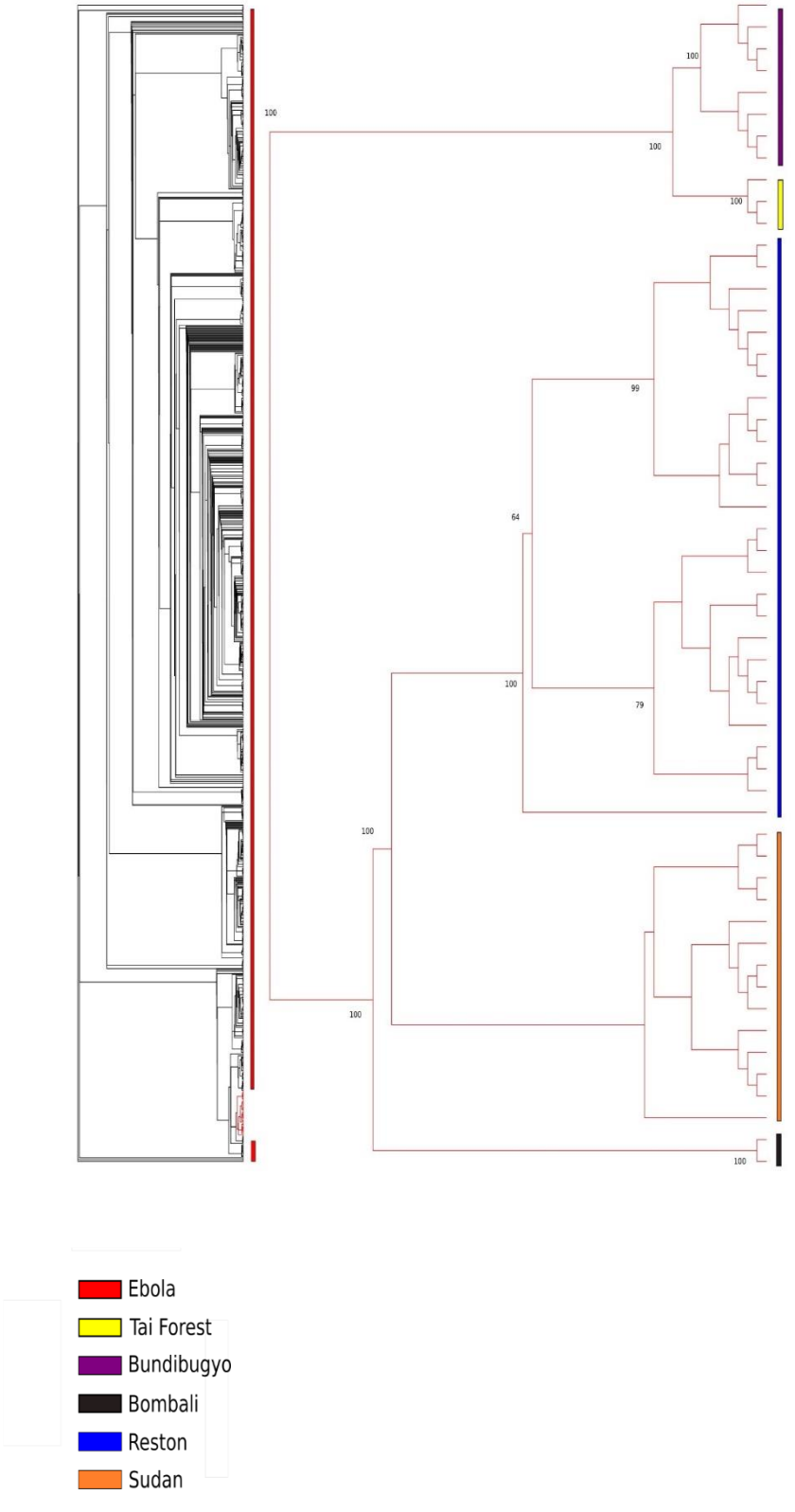


Fig S1H. NP Maximum Likelihood Tree

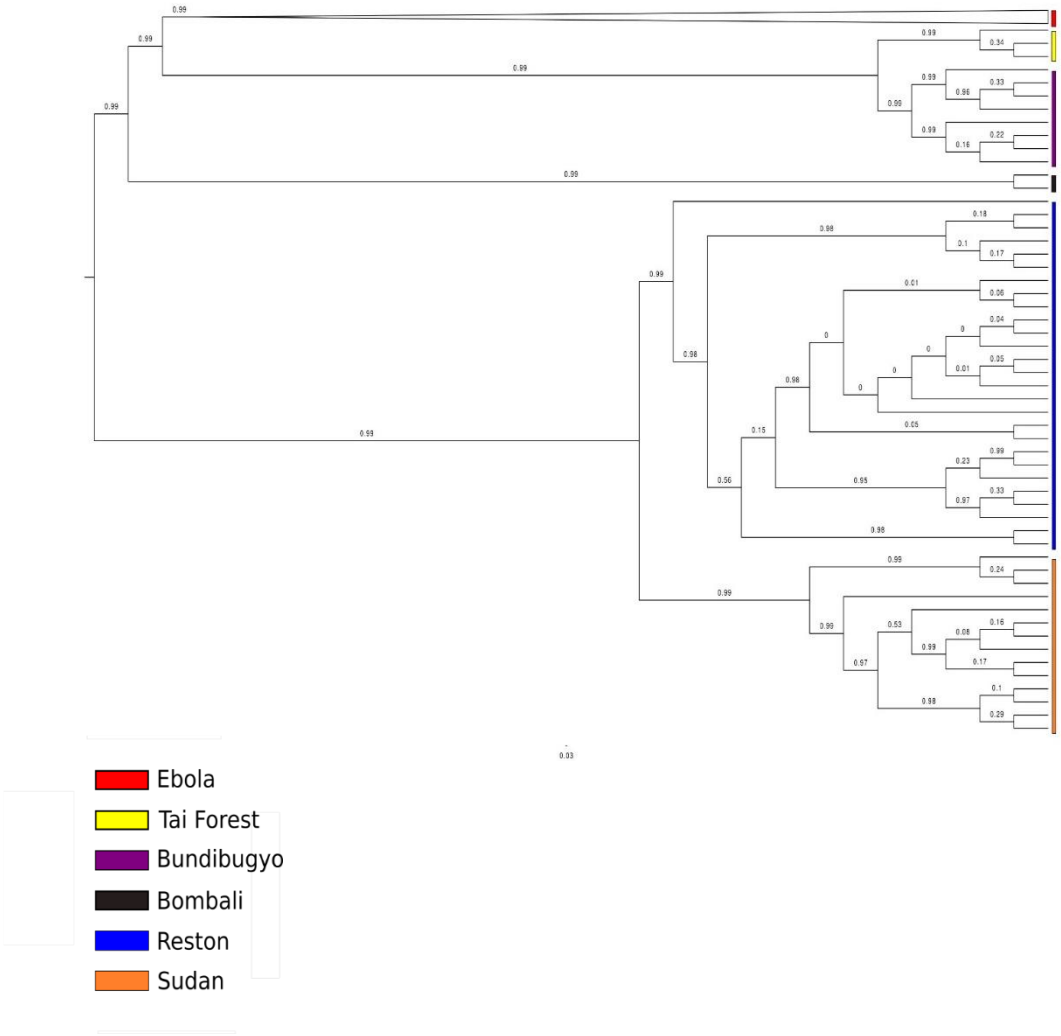


Fig S1I. VP24 Bayesian Tree

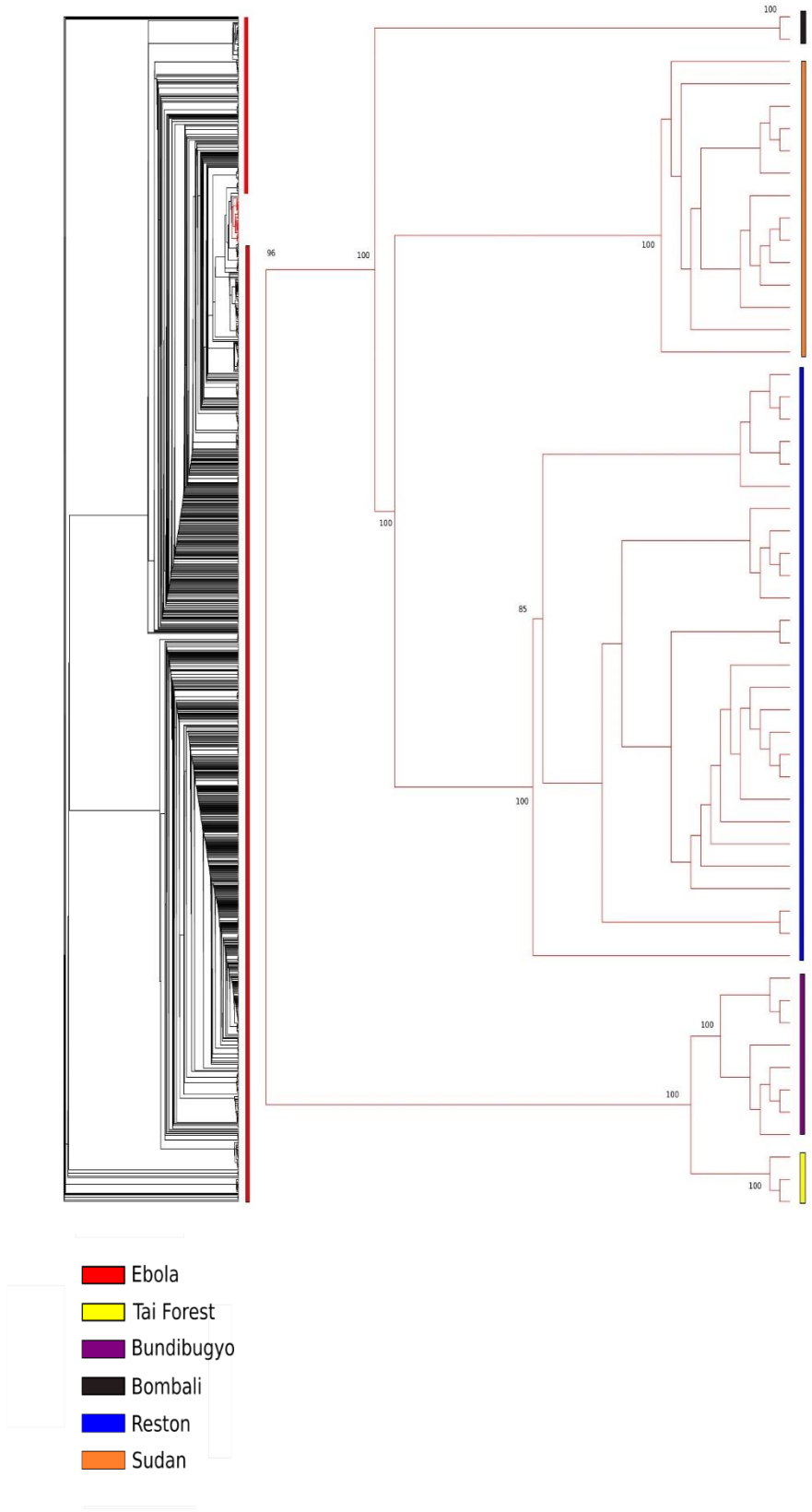


Fig S1J. VP24 Maximum Likelihood Tree

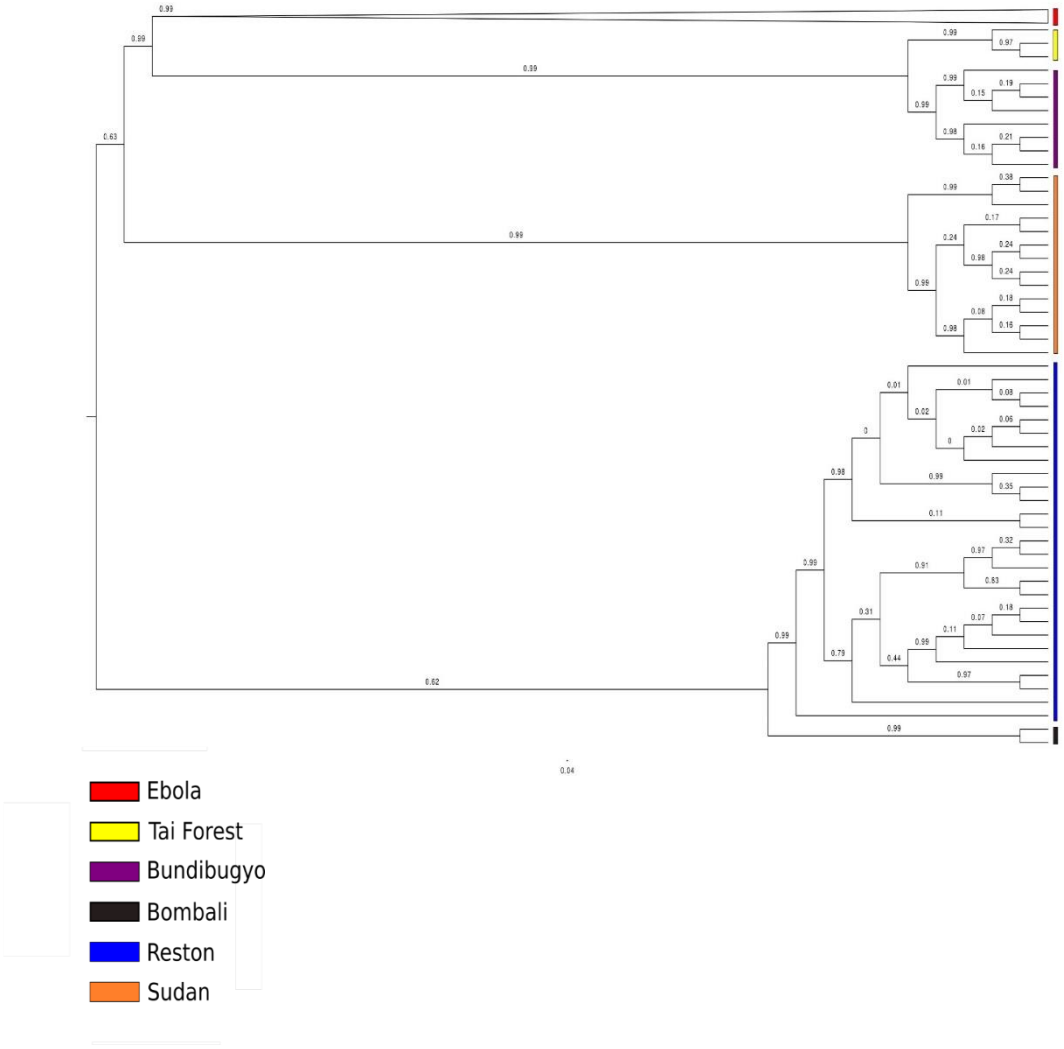


Fig S1K. VP30 Bayesian Tree

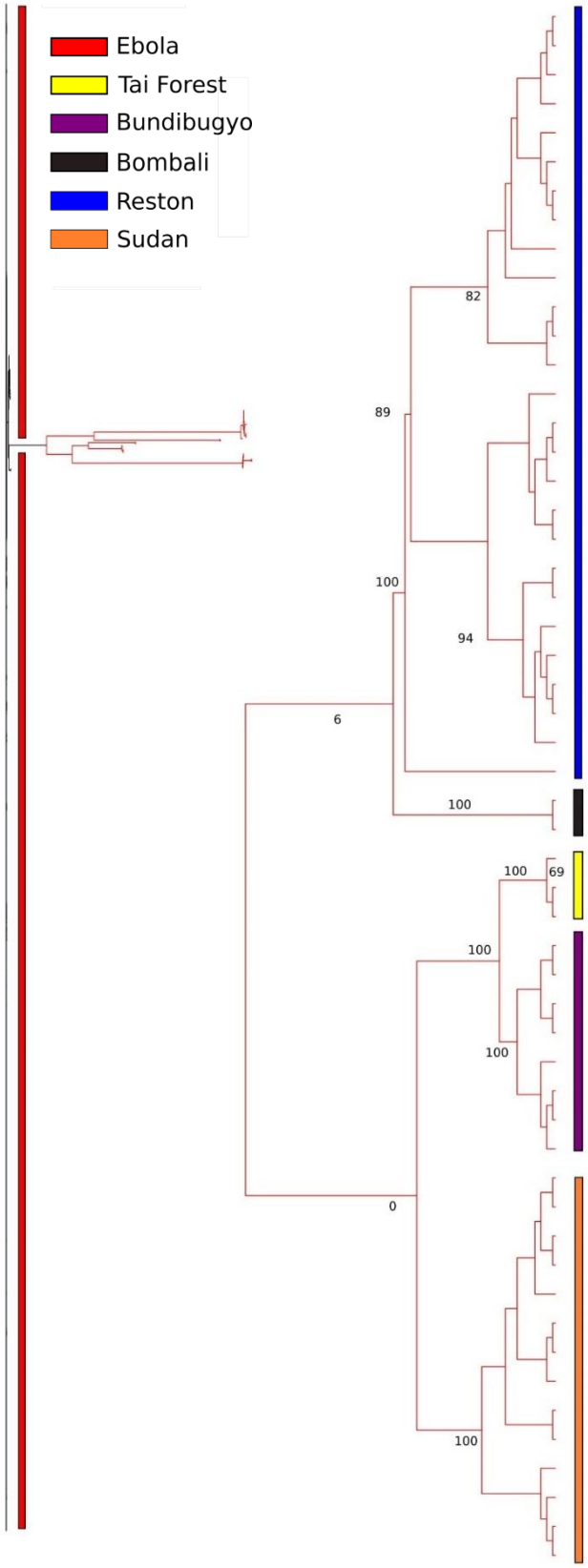


Fig S1L. VP30 Maximum Likelihood Tree

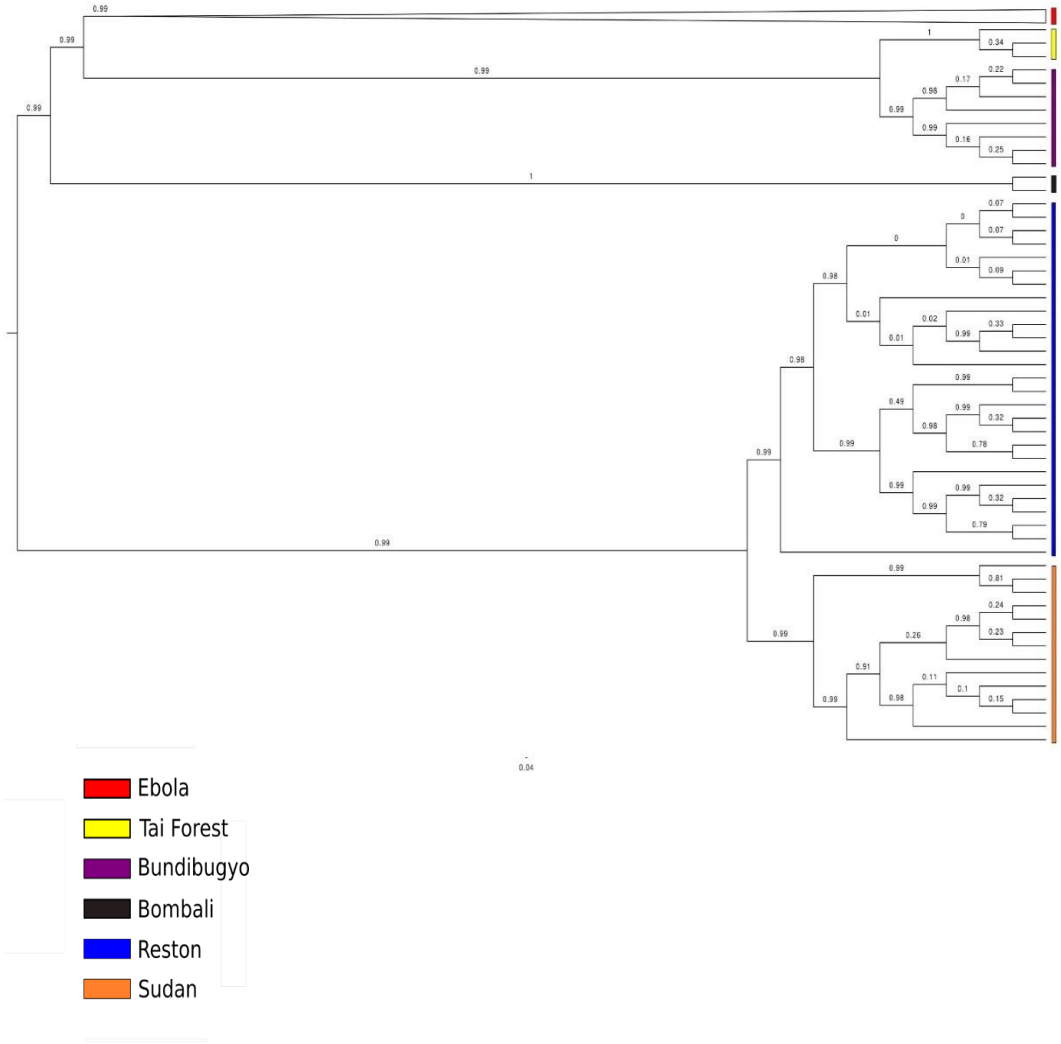


Fig S1M. VP35 Bayesian Tree

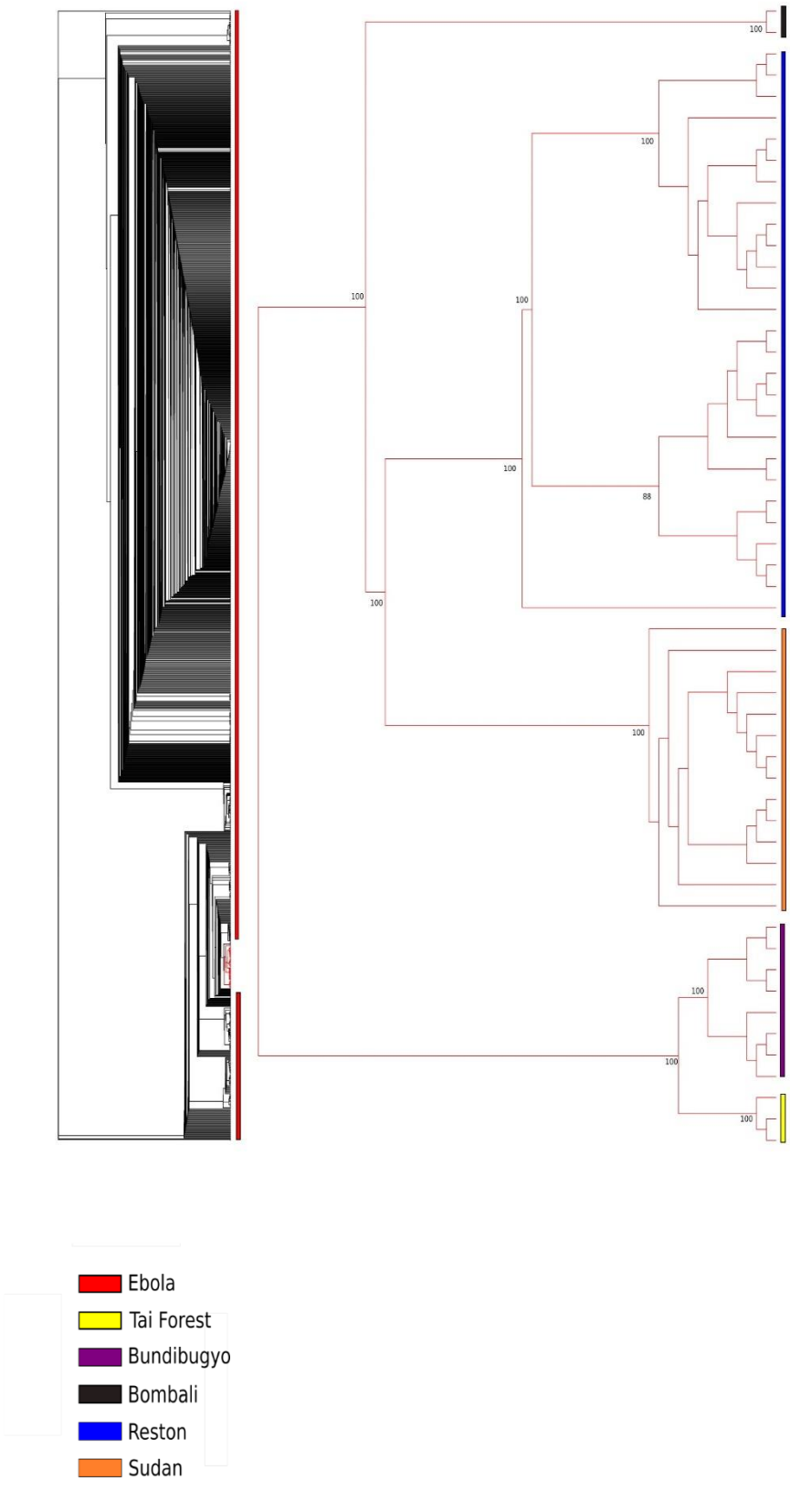


Fig S1N. VP35 Maximum Likelihood Tree

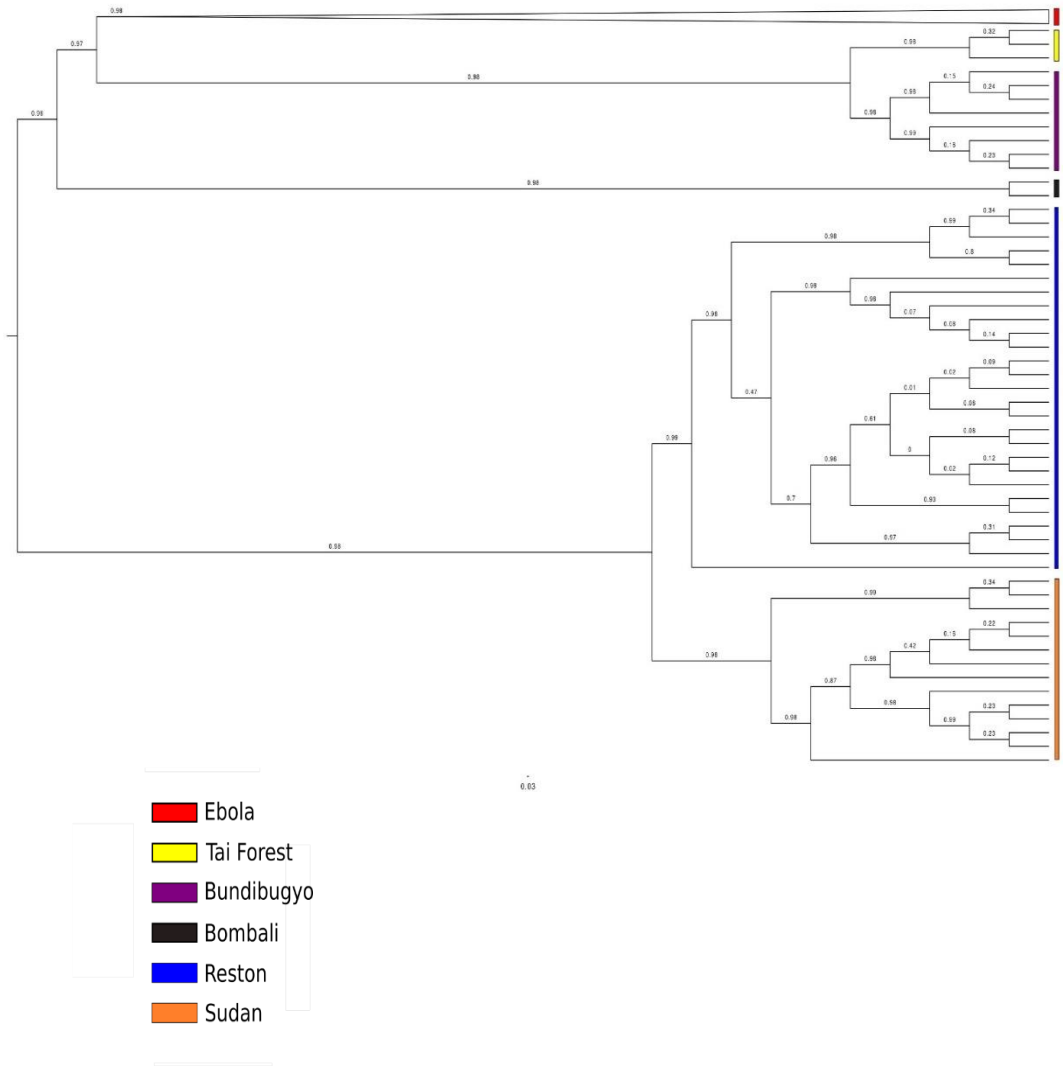


Fig S1O. VP40 Bayesian Tree

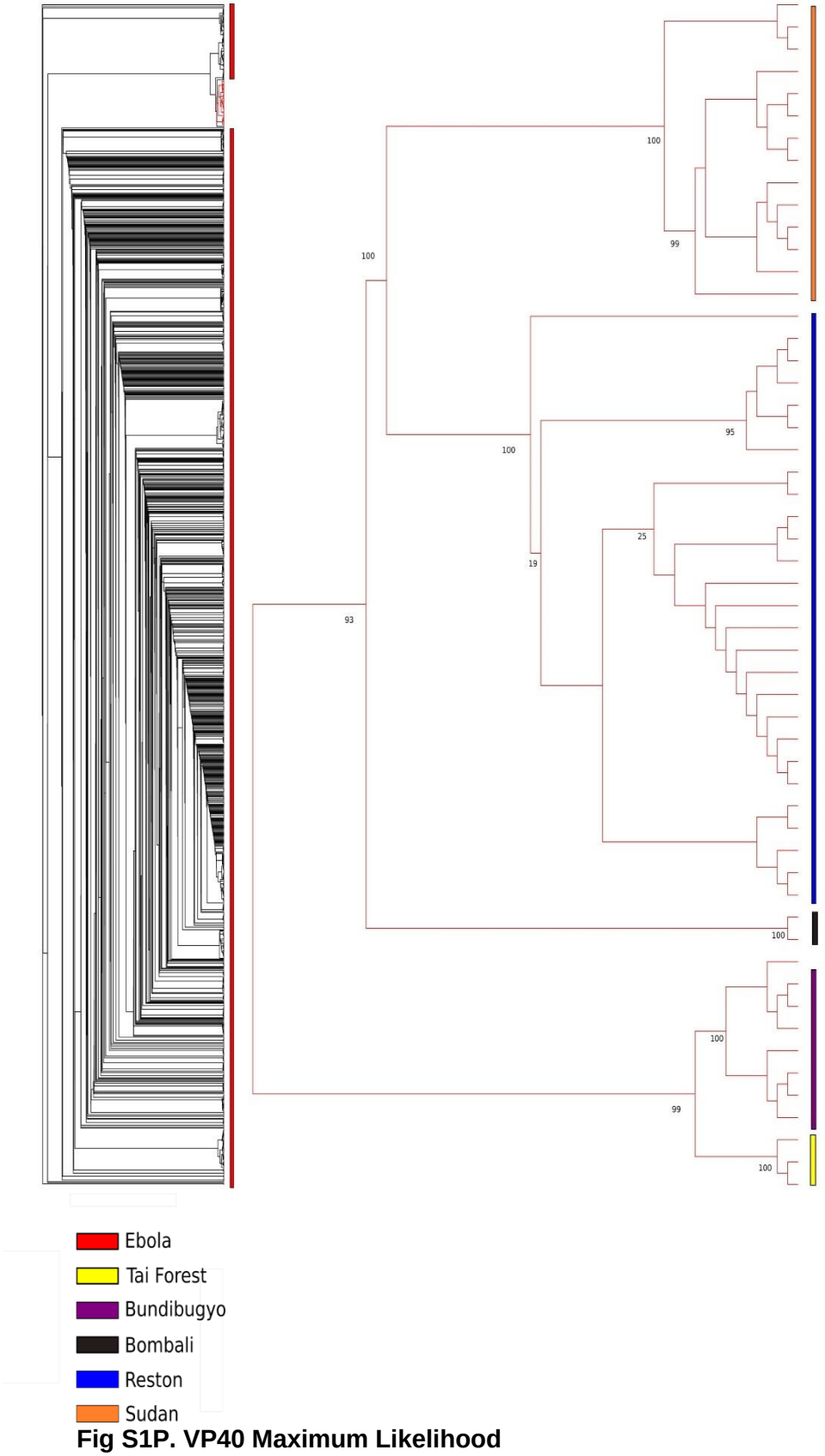
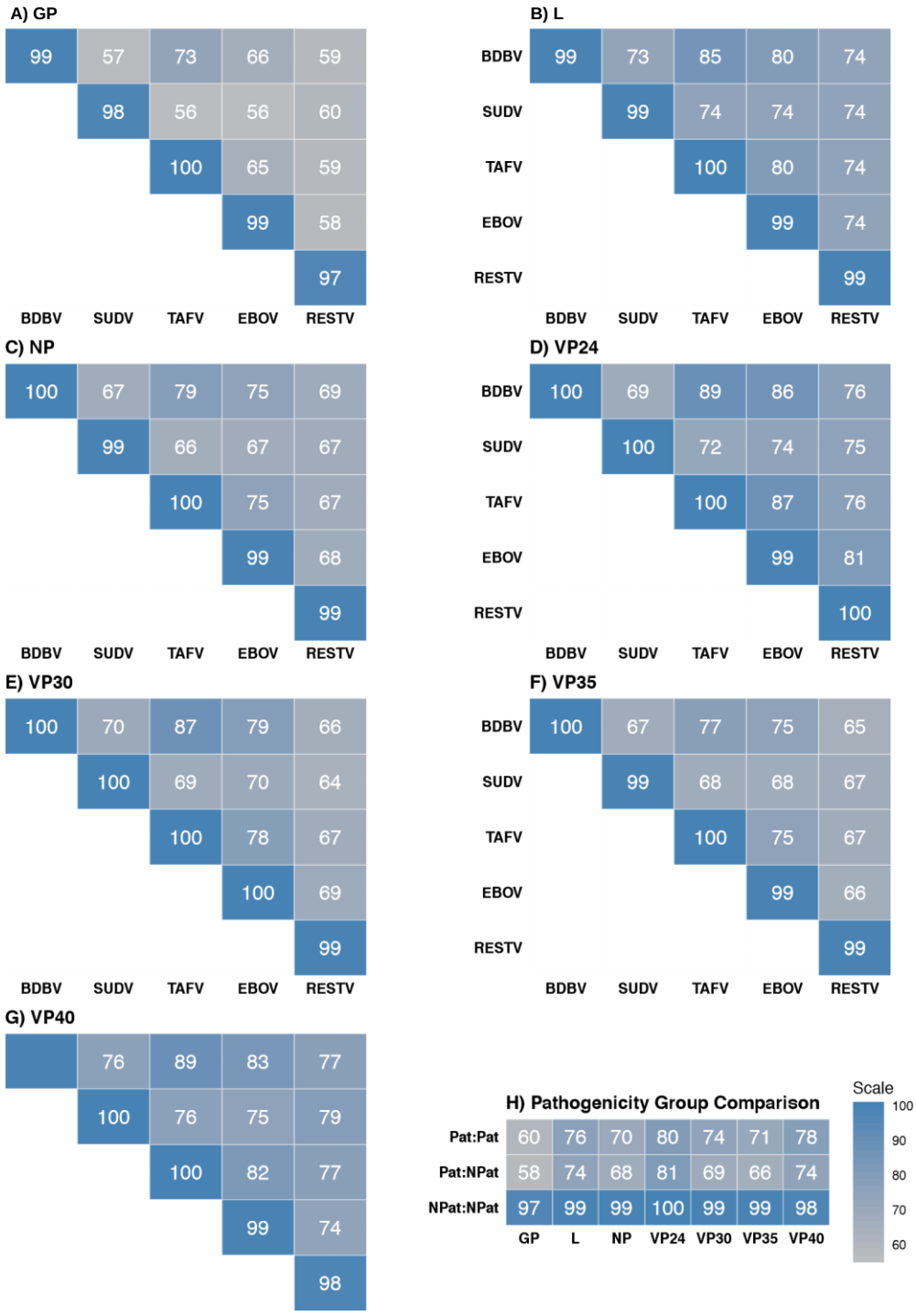
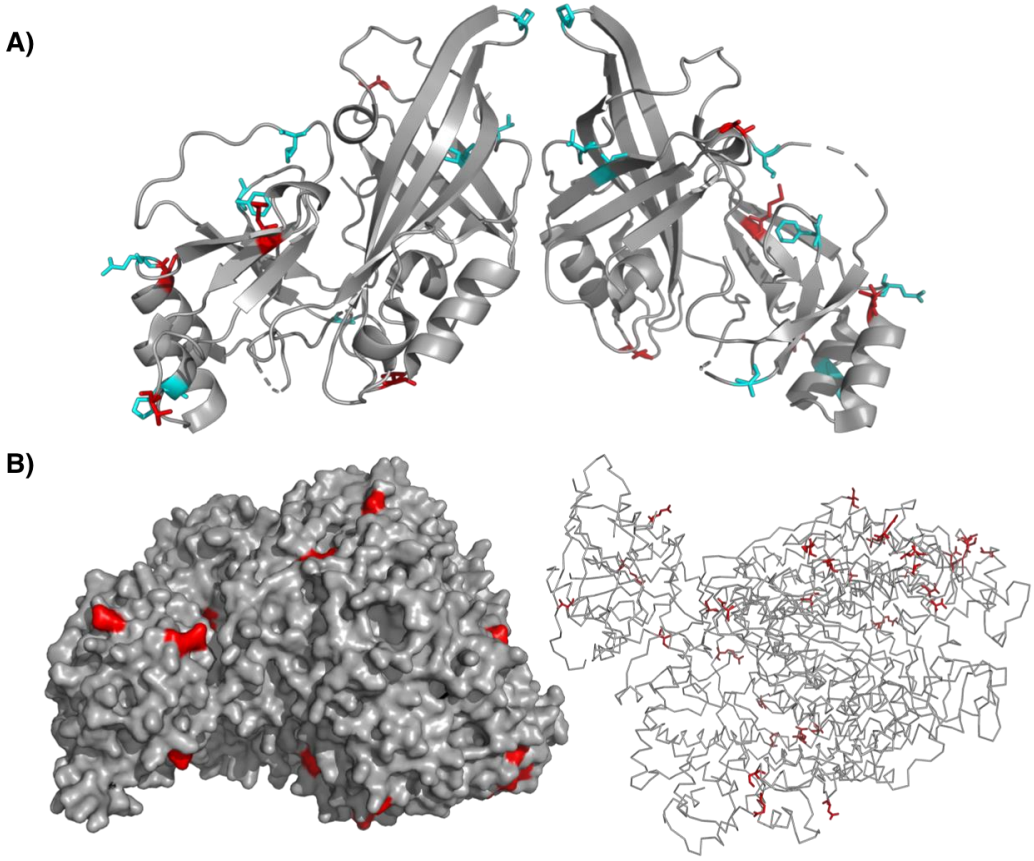


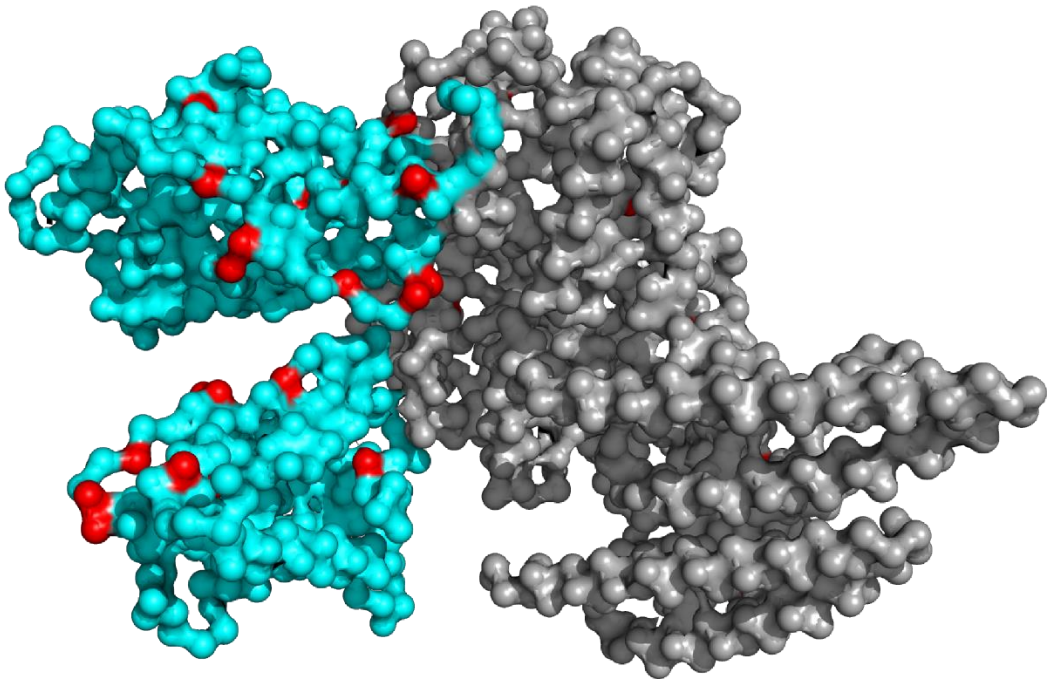
Fig S1P. VP40 Maximum Likelihood



Supplementary Figure 2: Intra- and inter-species protein sequence conservation for each of the 7 ebolavirus proteins, and conservation between pathogenic and non-pathogenic groups. **A)** GP gene. **B)** L gene. **C)** NP gene. **D)** VP24 gene. **E)** VP30 gene. **F)** VP35 gene. **G)** VP40 gene. **H)** Comparison of pathogenicity groups for all 7 proteins.



Supplementary Figure 3: SDPs mapped to VP40 and L. **A)** SDPs identified in VP40 – VP40 is shown in cartoon format and coloured grey, SDPs are shown in stick format with retained SDPs coloured cyan and gained SDPs coloured red. **B)** SDPs mapped to the Phyre2 structure of L, shown as a surface representation and as a ribbon representation – L is shown in grey and SDPs are shown in red.



Supplementary Figure 4: Model of the *Ebola virus* nucleocapsid subunit from recombinant virus-like particles using Cryo-EM (resolution 7.3 angstroms) featuring VP24 (cyan) and Nucleoprotein (grey). SDPs are shown in red.

Supplementary Tables

Supplementary Table 1: SDPs identified for the gene GP. For each species the amino acid residue at the alignment position is given followed by the sequence position, e.g. A1 for Alanine at sequence position 1, and the number of samples with this amino acid is given in brackets after. Where there is more than one amino acid at a position for a species these are separated by semi-colons. A 'Lost' status indicates the SDP was only found in the original analysis, a 'Retained' status indicates that the SDP was found in the old and new analysis, and a 'Gained' status indicates that the SDP is only found in the new analysis.

Alignment Position	EBOV	SUDV	BDBV	TAFV	RESTV	BOMV	EBOV REF	Status
2	M1(1356)	M1(14)	M1(8)	M1(3)	G2(27)	N/A	M1	Retained
32	F31(1355);S31(1)	F31(14)	F31(8)	F31(3)	I32(27)	V27	F31	Retained
38	V37(1356)	V37(14)	V37(8)	V37(3)	I38(27)	V33	V37	Retained
46	V45(1356)	V45(14)	V45(8)	V45(3)	A46(27)	V41	V45	Retained
76	V75(1349);A75(7)	V75(14)	V75(8)	V75(3)	I76(27)	I71	V75	Retained
197	S196(1356)	S196(14)	S196(8)	S196(3)	A197(27)	P192	S196	Retained
261	I260(13 56)	I260(14)	I260(8)	I260(3)	L261(27)	I256	I260	Retained
270	T269(13 56)	T269(14)	T269(8)	T269(3)	S270(27)	T265	T269	Retained
308	X307(1);S307(1355)	S307(14)	S307(8)	S307(3)	H308(27)	S303	S307	Retained
497	X476(4);S476(1352)	S476(14)	S476(8)	L476(3)	P477(27)	I479	S476	Gained
519	R498(1353);X49 8(3)	R498(14)	R498(8)	R498(3)	K499(27)	R494	R498	Retained
521	X500(2);R500(1354)	R500(14)	R500(8)	R500(3)	K501(27)	K496	R500	Retained
535	N514(1354);X51 4(2)	N514(14)	N514(8)	N514(3)	D515(27)	N510	N514	Retained
542	X521(1);Q521(1355)	Q521(14)	Q521(8)	L521(3)	V522(27)	H517	Q521	Retained
605	I584(1356)	I584(14)	I584(8)	I584(3)	L585(27)	I580	I584	Retained
628	D607(1354);X60 7(2)	D607(14)	D607(8)	D607(3)	S608(27)	D603	D607	Retained
643	X622(1);K622(1355)	K622(14)	K622(8)	K622(3)	E623(27)	R618	K622	Retained
659	Q638(1352);L638(1); R638(1);X638(2)	Q638(14)	Q638(8)	Q638(3)	H639(27)	Q634	Q638	Retained
665	X644(2);W644(1354)	W644(14)	W644(8)	W644(3)	L645(27)	W640	W644	Retained
680	A659(1);T659(1354);X69(1)	T659(14)	T659(8)	T659(3)	I660(27)	V655	T659	Retained
3	G2(1356)	G2(11); E2(3)	V2(8)	G2(3)	S3(27)	N/A	G2	Lost
208	X207(57);E207(1299)	E207(14)	T207(8)	T207(3)	D208(27)	N203	E207	Lost
211	S210(1299);X210(57)	S210(14)	S210(8)	S210(3)	T211(27)	S206	S210	Lost
326	R325(1354);X32 5(2)	R325(14)	V325(8)	V325(3)	G326(27)	E321	R325	Lost
355	H354(1356)	H354(14)	R354(8)	Q354(3)	L355(27)	Q350	H354	Lost
417	X403(10);Q403(1346)	S412(14)	A409(8)	T409(3)	E412(27)	Q397	Q403	Lost
432	S418(1339);X418(15);	T427(14)	S419(8)	T419(3)	T422(27)	S412	S418	Lost

	X417(1);A417(1)							
468	T448(1345);X448(8); A448(3)	-(14)	T451(8)	K451(3)	-(27)	T438	T448	Lost
537	H516(1355);X516(1)	H516(14)	H516(8)	H516(3)	H517(14); Y517(13)	N512	H516	Lost
568	L547(1352);X54 7(4)	L547(14)	I547(8)	I547(3)	V548(27)	L543	L547	Lost
663	D642(1354);X64 2(2)	D642(14)	D642(8)	S642(3)	L643(27)	D638	D642	Lost

Supplementary Table 2: SDPs identified for the gene L. For each species the amino acid residue at the alignment position is given followed by the sequence position, e.g. A1 for Alanine at sequence position 1, and the number of samples with this amino acid is given in brackets after. Where there is more than one amino acid at a position for a species these are separated by semi-colons. A 'Lost' status indicates the SDP was only found in the original analysis, a 'Retained' status indicates that the SDP was found in the old and new analysis, and a 'Gained' status indicates that the SDP is only found in the new analysis.

Alignment Position	EBOV	SUDV	BDBV	TAFV	RESTV	BOMV	EBOV REF	Status
67	V59(1);V66(1355)	V67(14)	V66(8)	V66(3)	T66(27)	V66	V66	Retained
137	I136(1355);I129(1)	I137(14)	I136(8)	I136(3)	L136(27)	L136	I136	Retained
147	L139(1);L146(1355)	L147(14)	L146(8)	L146(3)	V146(27)	L146	L146	Retained
203	T202(1346);S202(9);T195(1)	T203(14)	T202(8)	T202(3)	I202(27)	E202	T202	Gained
222	A221(1355);A214(1)	A222(14)	A221(8)	A221(3)	S221(27)	A221	A221	Retained
224	Q223(1355);Q216(1)	Q224(14)	Q223(8)	Q223(3)	L223(27)	Q223	Q223	Retained
227	T219(1);T226(1354);A226(1)	T227(14)	T226(8)	T226(3)	S226(27)	K226	T226	Gained
228	H227(1355);H220(1)	H228(14)	H227(8)	H227(3)	Q227(27)	Y227	H227	Retained
237	V236(1355);V229(1)	V237(14)	V236(8)	V236(3)	I236(27)	V236	V236	Gained
284	L283(1355);L276(1)	L284(14)	L283(8)	L283(3)	V283(27)	L283	L283	Retained
331	T323(1);T330(13 55)	T331(14)	T330(8)	T330(3)	D330(27)	T330	T330	Retained
351	E343(1);E350(1355)	E351(14)	E350(8)	E350(3)	D350(27)	E350	E350	Retained
362	M361(1);T361(1353);T354(1);X361(1)	T362(14)	T361(8)	T361(3)	S361(27)	T361	T361	Retained
366	L358(1);L365(1354);X365(1)	L366(14)	L365(8)	L365(3)	F365(27)	F365	L365	Retained
380	V379(1353);V372(1);X379(2)	V380(14)	V379(8)	V379(3)	I379(27)	I379	V379	Retained
448	X447(4);Q447(1351);Q440(1)	Q448(14)	Q447(8)	Q447(3)	H447(27)	Q447	Q447	Retained
451	P450(1351);P443(1);X450(4)	P451(14)	P450(8)	P450(3)	S450(27)	P450	P450	Retained
466	D465(1355);D458(1)	D466(14)	D465(8)	D465(3)	N465(27)	D465	D465	Retained
848	X847(1);S847(1354);S840(1)	S848(14)	S847(8)	S847(3)	A847(27)	S847	S847	Retained
869	S861(1);S868(1355)	S869(14)	S868(8)	S868(3)	A868(27)	S868	S868	Retained

1025	T1017(1);T1024(1355)	T1025(14)	T1024(8)	T1024(3)	N1024(27)	N1024	T1024	Retained
1074	R1066(1);R1073(1355)	R1074(14)	R1073(8)	R1073(3)	K1073(27)	R1073	R1073	Retained
1120	A1112(1);A1119(1355)	A1120(14)	A1119(8)	A1119(3)	S1119(27)	A1119	A1119	Retained
1164	P1156(1);P1163(1355)	P1162(14)	P1163(8)	P1163(3)	A1161(27)	P1163	P1163	Retained
1190	D1189(1355);D1182(1)	D1188(14)	D1189(8)	D1189(3)	S1187(27)	D1189	D1189	Retained
1215	A1214(1355);A1207(1)	A1213(14)	A1214(8)	A1214(3)	S1212(27)	A1214	A1214	Retained
1218	R1210(1);R1217(1355)	R1216(14)	R1217(8)	R1217(3)	K1215(27)	R1217	R1217	Retained
1238	D1237(1355);D1230(1)	D1236(14)	D1237(9)	D1237(3)	E1235(27)	D1237	D1237	Retained
1355	R1354(1355);R1347(1)	R1353(14)	R1354(8)	R1354(3)	K1352(27)	R1354	R1354	Retained
1367	T1359(1);T1366(1355)	T1365(14)	T1366(8)	T1366(3)	A1364(27)	T1366	T1366	Retained
1409	I1408(1355);I1401(1)	I1407(14)	I1408(8)	I1408(3)	M1406(27)	I1408	I1408	Retained
1415	I1407(1);I1414(1355)	I1413(14)	I1414(8)	I1414(3)	L1412(27)	I1414	I1414	Retained
1437	S1429(1);S1436(1355)	S1435(14)	S1436(8)	S1436(3)	N1434(27)	S1436	S1436	Retained
1474	S1466(1);X1473(2);S1473(1353)	S1472(14)	S1473(8)	S1473(3)	C1471(27)	S1473	S1473	Retained
1489	L1488(1355);L1481(1)	L1487(14)	L1488(8)	L1488(3)	Y1486(27)	I1488	L1488	Retained
1500	I1499(1355);I1492(1)	I1498(14)	I1499(8)	I1499(3)	L1497(27)	I1499	I1499	Retained
1507	S1506(1355);S1499(1)	S1505(14)	S1506(8)	S1506(3)	A1504(27)	S1506	S1506	Retained
1510	I1509(1355);I1502(1)	I1508(14)	I1509(8)	I1509(3)	V1507(27)	V1509	I1509	Retained
1627	L1617(1);L1624(1355)	L1623(14)	L1624(8)	L1624(3)	Y1624(27)	L1624	L1624	Retained
1631	C1628(1355);C1621(1)	C1627(14)	C1628(8)	C1628(3)	S1628(27)	C1638	C1628	Retained
1786	V1755(1);V1762(1355)	V1759(14)	V1762(8)	V1762(3)	I1760(27)	V1760	V1762	Retained
1874	V1843(1);V1850(1355)	V1847(14)	V1850(8)	V1850(3)	T1848(27)	T1848	V1850	Retained
1897	I1873(1);T1866(1);T1873(1354)	T1870(14)	T1873(8)	T1873(3)	S1871(27)	T1871	T1873	Retained
1941	R1909(1);R1916(1355)	R1913(14)	R1916(8)	R1916(3)	N1915(3);N1914(24)	R1914	R1916	Retained
1966	E1941(1354);X1941(1);E1934(1)	E1938(14)	E1941(8)	E1941(3)	R1939(24);R1940(3)	E1939	E1941	Retained
2069	L2044(1355);L2037(1)	L2041(14)	L2044(8)	L2044(3)	I2043(3);I2042(24)	L2044	L2044	Retained

2102	S2077(1355);S2070(1)	S2074(14)	S2077(8)	S2077(3)	T2075(24);T2076(3)	S2075	S2077	Retained
2123	E2091(1);E2098(1355)	E2095(14)	E2098(8)	E2098(3)	D2096(24);D2097(3)	E2096	E2098	Retained
2182	L2157(1353);X2157(2);L2150(1)	L2154(14)	L2157(8)	L2157(3)	V2155(24);V2156(3)	L2155	L2157	Retained
2193	R2168(1355);R2161(1)	R2165(14)	R2168(8)	R2168(3)	H2167(3);H2166(24)	K2166	R2168	Retained
2200	R2168(1);R2175(1355)	R2172(14)	R2175(8)	R2175(3)	K2173(24);K2174(3)	R2173	R2175	Retained
2202	X2177(1);L2177(1354);L2170(1)	L2174(14)	L2177(8)	L2177(3)	F2175(24);F2176(3)	W2175	L2177	Retained
2211	X2186(2);M2179(1);M2186(1353)	M2183(14)	M2186(8)	M2186(3)	L2185(3);L2184(24)	M2184	M2186	Retained
110	Q109(1298);X109(57);Q102(1)	Q110(14)	Q109(8)	Q109(3)	R109(2);H109(25)	Q109	Q109	Lost
277	L276(1355);X269(1)	L277(14)	L276(8)	L276(3)	I276(27)	L276	L276	Lost
313	Y312(1354);X305(1);X312(1)	Y313(14)	Y312(8)	Y312(3)	F312(27)	Y312	Y312	Lost
327	A319(1);X326(1);A326(1354)	A327(14)	A326(8)	A326(3)	S326(27)	A326	A326	Lost
690	E689(1353);E682(1);X689(2)	E690(14)	E689(8)	E689(3)	S689(27)	E689	E689	Lost
897	X896(58);F896(1297);F889(1)	F897(14)	F896(8)	F896(3)	Y896(27)	F896	F896	Lost
926	L925(1352);X925(3);L918(1)	L926(14)	L925(8)	L925(3)	F925(27)	L925	L925	Lost
955	X954(2);A954(1353);A947(1)	A955(14)	A954(8)	A954(3)	S954(27)	A954	A954	Lost
996	X995(2);S995(1353);S988(1)	S996(14)	S995(8)	S995(3)	T995(27)	S995	S995	Lost
1256	V1255(1);I1248(1);I1255(1354)	I1254(14)	I1255(8)	I1255(3)	V1253(27)	I1255	I1255	Lost
1396	A1395(1);S1395(1353);S1388(1);X1395(1)	S1394(14)	S1395(8)	S1395(3)	T1393(27)	T1395	S1395	Lost
1462	X1461(1);K1454(1);K1461(1354)	K1460(14)	K1461(8)	K1461(3)	Q1459(27)	I1461	K1461	Lost
1539	X1538(1);A1538(1354);A1531(1)	A1537(14)	A1538(8)	A1538(3)	S1536(27)	A1538	A1538	Lost
2033	X2008(57);L2001(1);L2008(1298)	L2005(14)	L2008(8)	L2008(3)	I2007(3);I2006(24)	L2006	L2008	Lost
2130	X2105(2);Q2098(1);Q2105(1353)	Q2102(14)	Q2105(8)	Q2105(3)	L2104(3);L2103(24)	Q2103	Q2105	Lost
2133	Q2108(1353);Q2101(1);X2108(2)	Q2105(14)	Q2108(8)	Q2108(3)	E2107(3);E2106(24)	Q2106	Q2108	Lost

2156	Y2124(1);Y2131(1354);X2131(1)	Y2128(14)	Y2131(8)	Y2131(3)	F2129(24);F2130(3)	Y2129	Y2131	Lost

Supplementary Table 3: SDPs identified for the gene NP. For each species the amino acid residue at the alignment position is given followed by the sequence position, e.g. A1 for Alanine at sequence position 1, and the number of samples with this amino acid is given in brackets after. Where there is more than one amino acid at a position for a species these are separated by semi-colons. A 'Lost' status indicates the SDP was only found in the original analysis, a 'Retained' status indicates that the SDP was found in the old and new analysis, and a 'Gained' status indicates that the SDP is only found in the new analysis.

Alignment Position	EBOV	SUDV	BDBV	TAFV	RESTV	BOMV	EBOV REF	Status
4	R4(1356)	R4(14)	R4(8)	R4(3)	G4(27)	R4	R4	Retained
16	X16(1);E16(1355)	E16(14)	E16(8)	G16(3)	D16(27)	D16	E16	Retained
30	S30(1356)	S30(14)	S30(8)	S30(3)	T30(27)	S30	S30	Retained
39	R39(1356)	R39(14)	R39(8)	R39(3)	K39(27)	R39	R39	Retained
56	I56(1356)	I56(14)	I56(8)	I56(3)	V56(27)	I56	I56	Retained
64	V64(1356)	V64(14)	V64(8)	V64(3)	I64(27)	V64	V64	Retained
105	R105(1354);X105(2)	R105(14)	R105(8)	R105(3)	K105(27)	R105	R105	Retained
137	M137(1354);X137(2)	M137(14)	M137(8)	M137(3)	L137(27)	M137	M137	Retained
212	X212(1);F212(13 55)	F212(14)	F212(8)	F212(3)	Y212(27)	F212	F212	Retained
274	K274(1355);X27 4(1)	K274(14)	K274(8)	K274(3)	R274(27)	R274	K274	Retained
279	X279(1);S279(1355)	S279(14)	S279(8)	S279(3)	A279(27)	S279	S279	Retained
416	X416(1);K416(1355)	K416(14)	K416(8)	K416(3)	N416(27)	R416	K416	Retained
421	X421(1);Y421(1355)	Y421(14)	Y421(8)	Y421(3)	Q421(27)	Y421	Y421	Retained
426	D426(1356)	D426(14)	D426(8)	D426(3)	E426(27)	E426	D426	Retained
435	D435(1356)	D435(14)	D435(8)	D435(3)	N435(27)	D435	D435	Retained
443	D443(1356)	D443(14)	D443(8)	D443(3)	E443(27)	V443	D443	Retained
453	T453(1356)	T453(14)	T453(8)	T453(3)	I453(27)	T453	T453	Retained
497	P497(1316);S497(40)	P497(14)	P497(8)	R497(3)	A497(27)	S497	P497	Retained
571	T563(1348);X56	T563(14)	T563(8)	T563(3)	S563(27)	A563	T563	Retained
573	X565(7);I565(1348)	I565(14)	I565(8)	I565(3)	V565(27)	I565	I565	Retained

610	X602(24);P602(1332)	P602(14)	P602(8)	N602(3)	T602(27)	R602	P602	Retained
650	X641(5);N641(1351)	N641(14)	N641(8)	K641(3)	Q641(27)	S641	N641	Retained
714	A705(1356)	A705(14)	A705(8)	A705(3)	R705(27)	A705	A705	Retained
726	G717(1354);X717(2)	G717(14)	G717(8)	D717(3)	N717(27)	G717	G717	Retained
42	Q42(9);S42(1);P42(1346)	P42(14)	P42(8)	Q42(3)	S42(27)	P42	P42	Lost
374	R374(1);K374(1355)	K374(14)	K374(8)	K374(3)	R374(27)	E374	K374	Lost
492	X492(57);D492(1299)	D492(14)	D492(8)	D492(3)	E492(27)	E492	D492	Lost
530	P526(1356)	V526(14)	G524(4);S524(4)	N524(3)	V530(1);A530(26)	P526(1);L526(1)	P526	Lost
725	D716(1354);N716(1);6(1)	D716(14)	D716(8)	D716(3)	N716(27)	D716	D716	Lost

Supplementary Table 4: SDPs identified for the gene VP24. For each species the amino acid residue at the alignment position is given followed by the sequence position, e.g. A1 for Alanine at sequence position 1, and the number of samples with this amino acid is given in brackets after. Where there is more than one amino acid at a position for a species these are separated by semicolons. A 'Lost' status indicates the SDP was only found in the original analysis, a 'Retained' status indicates that the SDP was found in the old and new analysis, and a 'Gained' status indicates that the SDP is only found in the new analysis.

Alignment Position	EBOV	SUDV	BDBV	TAFV	RESTV	BOMV	EBOV REF	Status
17	X17(1);L17(1355)	L17(14)	L17(8)	L17(3)	M17(27)	L17	L17	Retained
22	X22(1);V22(1355)	V22(14)	V22(8)	V22(3)	I22(27)	V22	V22	Retained
31	V31(1356)	V31(14)	V31(8)	V31(3)	I31(27)	V31	V31	Retained
102	I102(1354);V102(2)	I102(14)	I102(8)	I102(3)	L102(27)	I102	I102	Gained
131	T131(1356)	T131(14)	T131(8)	T131(3)	S131(27)	T131	T131	Retained
132	N132(1356)	N132(14)	N132(8)	N132(3)	T132(27)	A132	N132	Retained
136	I136(15);M136(1341)	M136(14)	M136(8)	M136(3)	L136(27)	L136	M136	Retained
139	Q139(1356)	Q139(14)	Q139(8)	Q139(3)	R139(27)	R139	Q139	Retained
140	R140(1356)	R140(14)	H140(8)	Q140(3)	S140(27)	R140(1) ;S140(1)	R140	Gained
226	X226(4);T226(13 52)	T226(14)	T226(8)	T226(3)	A226(27)	T226	T226	Retained
248	S248(1356)	S248(14)	S248(8)	S248(3)	L248(27)	S248	S248	Retained

Supplementary Table 5: SDPs identified for the gene VP30. For each species the amino acid residue at the alignment position is given followed by the sequence position, e.g. A1 for Alanine at sequence position 1, and the number of samples with this amino acid is given in brackets after. Where there is more than one amino acid at a position for a species these are separated by semicolons. A 'Lost' status indicates the SDP was only found in the original analysis, a 'Retained' status indicates that the SDP was found in the old and new analysis, and a 'Gained' status indicates that the SDP is only found in the new analysis.

Alignment Position	EBOV	SUDV	BDBV	TAFV	RESTV	BOMV	EBOV REF	Status
40	H39(2);X39(1);Y39(1353)	Y39(14)	Y39(8)	Y39(3)	R40(27)	N50	Y39	Gained
53	X52(1);T52(1355)	T52(14)	T52(8)	T52(3)	N53(27)	V63	T52	Retained
54	X53(1);V53(1355)	V53(14)	V53(8)	V53(3)	L54(27)	M64	V53	Retained
64	X63(3);T63(1353)	T63(14)	T63(8)	T63(3)	I64(27)	L74	T63	Retained
94	E93(1356)	E93(14)	E93(8)	E93(3)	D94(27)	E104	E93	Retained
97	T96(1355);X96(1)	T96(14)	T96(8)	T96(3)	N97(27)	T107	T96	Retained
99	R98(1355);X98(1)	R98(14)	R98(8)	R98(3)	H99(27)	R109	R98	Retained
108	K107(1354);X10 7(2)	K107(14)	K107(8)	K107(3)	R108(27)	K118	K107	Retained
112	S111(1356)	S111(14)	S111(8)	S111(3)	I112(27)	L122	S111	Retained
117	X116(1);L116(13 55)	L116(14)	L116(8)	L116(3)	S117(27)	V127	L116	Retained
118	N117(1356)	N117(14)	N117(8)	S117(3)	Q118(27)	C128	N117	Gained
121	A120(1356)	A120(14)	A120(8)	A120(3)	S121(27)	A131	A120	Retained
151	T150(1355);X15 0(1)	T150(14)	T150(8)	T150(3)	I151(27)	I161	T150	Retained
158	X157(1);Q157(1 355)	Q157(14)	Q157(8)	Q157(3)	R158(27)	K168	Q157	Retained
160	X159(1);I159(1355)	I159(14)	I159(8)	I159(3)	L160(27)	L170	I159	Retained
206	E205(1356)	E205(14)	E205(8)	E205(3)	D206(27)	E216	E205	Retained
263	R262(1356)	R262(14)	R262(8)	R262(3)	A263(27)	K273	R262	Retained
269	S268(1356)	S268(14)	S268(8)	S268(3)	Q269(27)	A279	S268	Retained

272	E271(1356)	E271(14)	T271(8)	T271(3)	S272(27)	N282	E271	Gained
279	X278(1);G278(1 355)	G278(14)	E278(8)	E278(3)	N279(27)	T289	G278	Gained
197	H196(1);R196(1354);X196(1)	R196(14)	R196(8)	R196(3)	H197(27)	R207	R196	Lost

Supplementary Table 6: SDPs identified for the gene VP35. For each species the amino acid residue at the alignment position is given followed by the sequence position, e.g. A1 for Alanine at sequence position 1, and the number of samples with this amino acid is given in brackets after. Where there is more than one amino acid at a position for a species these are separated by semicolons. A 'Lost' status indicates the SDP was only found in the original analysis, a 'Retained' status indicates that the SDP was found in the old and new analysis, and a 'Gained' status indicates that the SDP is only found in the new analysis.

Alignment Position	EBOV	SUDV	BDBV	TAFV	RESTV	BOMV	EBOV REF	Status
59	X26(1);S7(1);S59(2);S26(1352)	S15(14)	S27(8)	S27(3)	T15(27)	S27	S26	Retained
81	E81(2);E48(1353);E29(1)	E37(14)	E49(8)	E49(3)	D37(27)	D49	E48	Retained
109	G76(3);X76(1);D76(1349); D109(2);D57(1)	D65(14)	D77(8)	D77(3)	E65(27)	D77	D76	Retained
117	X117(1);E65(1);E84(1346); E117(1);G84(1);X84(6)	E73(14)	A85(8)	E85(3)	K73(27)	D85	E84	Gained
118	X85(7);D85(1);X118(1); E66(1);E85(1345);E118(1)	E74(14)	E86(8)	D86(3)	K74(27)	E86	E85	Retained
125	S92(1348);X92(5);S73(1); S125(1);X125(1)	S81(14)	S93(8)	S93(3)	M81(27)	S93	S92	Retained
130	V130(2);V97(1350);V78(1);X97(3)	V86(14)	V98(8)	I98(3)	T86(27)	V98	V97	Retained
134	T101(1351);X101(2);T134(2);T82(1)	T90(14)	T102(8)	A102(3)	N90(27)	A102	T101	Retained
139	S106(1352);S87(1);X106(1);S139(2)	S95(14)	S107(8)	S107(3)	A95(27)	S107	S106	Retained
145	T112(1352);T145(2);T93(1);X112(1)	A101(3); T101(11)	T113(8)	I113(3)	S101(27)	S113	T112	Gained
154	V154(2);X121(2);V102(1);V121(1351)	V110(14)	V122(8)	M122(3)	I110(27)	V122	V121	Retained
158	A106(1);V125(1);A158(2);A125(1352)	A114(14)	T126(8)	A126(3)	G114(27)	A126	A125	Gained
187	A154(1353);A135(1);A187(2)	A143(14)	A155(8)	A155(3)	S143(27)	A155	A154	Retained
192	T140(1);T159(1353);T19 2(2)	T148(14)	T160(8)	T160(3)	V148(27)	T160	T159	Retained
193	E141(1);E160(1353);E193(2)	E149(14)	E161(8)	E161(3)	D149(27)	E161	E160	Retained
200	G200(2);X167(1);G148(1); G167(1352)	G156(14)	G168(8)	G168(3)	K156(27)	G168	G167	Retained
207	S207(2);S155(1);S174(1353)	S163(14)	S175(8)	S175(3)	A163(27)	S175	S174	Retained
214	I181(1353);I162(1);I214(2)	I170(14)	I182(8)	I182(3)	L170(27)	I182	I181	Retained

302	X269(2);E269(1351);E302(1); X302(1);E250(1)	E258(14)	E270(8)	E270(3)	D258(27)	D270	E269	Retaine d
323	A323(2);A271(1);X290(3); A290(1350)	A279(14)	A291(8)	A291(3)	V279(2 7)	I291	A290	Retaine d
347	X314(3);V314(1350);V29 5(1);V347(2)	V303(14)	V315(8)	V315(3)	A303(2 7)	V315	V314	Retaine d
362	Q329(1351);Q362(2);- (1);Q310(1);X329(1)	Q318(14)	Q330(8)	Q330(3)	K318(27)	Q330	Q329	Retaine d

Supplementary Table 7: SDPs identified for the gene VP40. For each species the amino acid residue at the alignment position is given followed by the sequence position, e.g. A1 for Alanine at sequence position 1, and the number of samples with this amino acid is given in brackets after. Where there is more than one amino acid at a position for a species these are separated by semicolons. A 'Lost' status indicates the SDP was only found in the original analysis, a 'Retained' status indicates that the SDP was found in the old and new analysis, and a 'Gained' status indicates that the SDP is only found in the new analysis.

Alignment Position	EBOV	SUDV	BDBV	TAFV	RESTV	BOMV	EBOV REF	Status
4	X4(5);V4(1348);*(2);I4(1)	V4(14)	A4(8)	I4(3)	G4(27)	T4	V4	Gained
46	I46(1);T33(2);T46(1353)	T46(14)	T46(8)	T46(3)	V46(27)	I46(1); T 46(1)	T46	Retained
85	P85(1352);P72(2);X85(2)	P85(14)	P85(8)	P85(3)	T85(27)	P85	P85	Retained
105	T92(2);T105(1352);X105(2)	M105(1) ;T105(13)	T105(8)	T105(3)	I105(27)	K105	T105	Gained
122	X122(3);I122(1351);I109(2)	I122(14)	I122(8)	I122(3)	V122(27)	I122	I122	Retained
128	X128(1);A128(1353);A11 5(2)	A128(14)	T128(8)	T128(3)	I128(27)	T128	A128	Gained
201	G201(1353);G188(2);X201(1)	G201(14)	G201(8)	G201(3)	N201(27)	N201	G201	Retained
209	F196(2);X209(3);F209(13 51)	F209(14)	F209(8)	F209(3)	L209(27)	F209	F209	Retained
244	L244(1354);L231(2)	M244(14)	L244(8)	L244(3)	I244(27)	L244	L244	Gained
245	Q245(1353);Q232(2);X245(1)	Q245(14)	Q245(8)	Q245(3)	P245(27)	Q245	Q245	Retained
259	M246(2);X259(1);M259(1353)	I259(14)	M259(8)	M259(3)	V259(27)	V259	M25 9	Gained
269	R269(1);X269(1);H256(2); H269(1352)	H269(14)	H269(8)	H269(3)	Q269(27)	Q269	H269	Retained
277	T264(2);T277(13 54)	S277(14)	T277(8)	T277(3)	Q277(27)	H277	T277	Gained
293	I280(2);I293(1353);X293(1)	I293(14)	I293(8)	I293(3)	V293(27)	I293	I293	Retained
323	V310(2);M323(1);A323(1) ;V323(1352)	L323(14)	V323(8)	V323(3)	H323(27)	A323	V323	Gained
325	E312(2);E325(1354)	E325(14)	E325(8)	E325(3)	D325(27)	E325	E325	Retained

Supplementary Table 8. Codon variation for each SDP residue of the non-pathogenic species (Reston virus). Only the 10 SDPs that showed any codon variation are included. All variants are synonymous.

Protein	SDP	Codons Present	Variant Type
GP	S196A	GCT:26, GCC:1	Synonymous
GP	T659I	ATT:26, ATC:1	Synonymous
L	L1488Y	TAT:14, TAC:13	Synonymous
L	I1509V	GTT:26, GTC:1	Synonymous
L	L2157V	GTG:14, GTT:13	Synonymous
NP	M137L	CTG:22, TTG:5	Synonymous
NP	K274R	CGT:26, CGC:1	Synonymous
VP30	T63I	ATA:26, ATT:1	Synonymous
VP35	S174A	GCG:21, GCA:6	Synonymous
VP35	I181L	CTT:22, CTA:5	Synonymous

Supplementary Table 9. Codons present for each SDP residue of each pathogenic species, alongside information about non-synonymous variants for GP. Highlighted in yellow are the codons that are 100% conserved across all pathogenic species.

SDP	Ebola (1356)	Sudan (14)	Bundibugyo (8)	Tai Forest (3)	Non-Synonymous?	Notes
M1G	ATG	ATG	ATG	ATG		
F31I	TTT (1349), TTC (6), TCT (1)	TTT	TTC	TTT	Yes	All residues are F across all species except 1 Ebola sequence which has an S (TCT)
V37I	GTT (1302), GTC (54)	GTT	GTA	GTT		
V45A	GTT	GTA	GTA	GTG		
V75I	GTG (1349), GCG (7)	GTA	GTT	GTA	Yes	All residues are V across all species except 7 Ebola sequences which have an A (GCG)
S196A	TCA	TCA	TCA (4), TCG (4)	TCT		
I260L	ATA	ATT	ATT	ATC		
T269S	ACC	ACA (13), ACC (1)	ACC	ACA		
S307H	TCT	TCT	TCT	TCT		
S476P	AGC	TCC	TCT	CTC	Yes	All Ebola, Sudan and Bundibugyo sequences have an S residue, Tai Forest has an L
R498K	AGG (1299), AGA (54)	CGC	AGA	AGA		
R500K	CGA	AGA	CGG (4), CGC (4)	CGA		
N514D	AAT (1345), AAC (9)	AAC	AAC	AAC		
Q521V	CAG	CAA	CAA	TTG	Yes	All Ebola, Sudan and Bundibugyo sequences have a Q residue, Tai Forest has an L
I584L	ATC	ATA	ATA	ATA		
D607S	GAC	GAT	GAT	GAT		
K622E	AAA	AAA	AAA	AAA		

Q638H	CAG (1352), CGG (1), CTG (1)	CAG	CAA	CAG	Yes	All residues are Q across all species except two variants in the Ebola sequences, one producing an R and one producing an L
W644L	TGG	TGG	TGG	TGG		
T659I	ACA (1354), GCA (1)	ACT	ACG	ACA	Yes	All residues are T across all species except 1 Ebola sequence which has an A (GCA)

Supplementary Table 10. Codons present for each SDP residue of each pathogenic species, alongside information about non-synonymous variants for L. Highlighted in yellow are the codons that are 100% conserved across all pathogenic species.

SDP	Ebola (1356)	Sudan (14)	Bundibugyo (8)	Tai Forest (3)	Non-Synonymous?	Notes
V66T	GTA	GTC	GTT	GTG		
I136L	ATC	ATT	ATT	ATT		
L146V	TTA	CTA (11), TTA (3)	CTG	TTA		
T202I	ACA (1346), TCA (9), ACG (1)	ACA	ACA	ACA	Yes	9 Ebola Residues have a TCA codon (S) while all other sequences across all species have codons for a T residue
A221S	GCG (1349), GCA (7)	GCT	GCG	GCT		
Q223L	CAA	CAA	CAA	CAA		
T226S	ACA (1355), GCA (1)	ACA	ACA	ACA	Yes	All residues are T across all species except 1 Ebola sequence which has an A (GCA)
H227Q	CAC	CAT (13), CAC (1)	CAT	CAC		
V236I	GTC	GTC (11), GTT (3)	GTT	GTC		
L283V	TTA (1354), TTG (1), CTA (1)	CTG (11), TTG (3)	TTA	TTA		
T330D	ACC (1355), ACT (1)	ACA	ACA	ACA		
E350D	GAA	GAG	GAA	GAA		
T361S	ACG (1353), ACA (1), ATG (1)	ACA	ACA	ACT	Yes	All residues are T across all species except 1 Ebola sequence which has an M (ATG)
L365F	CTT	TTA	CTC	CTC		
V379I	GTG	GTT	GTG	GTC		
Q447H	CAA	CAA	CAA	CAA		

P450S	CCG	CCA	CCA	CCA		
D465N	GAC	GAT	GAT	GAT		
S847A	TCC	TCA	TCT	TCT		
S868A	TCG	TCT	TCC	TCC		
T1024N	ACT (1341), ACC (14)	ACG	ACA	ACA		
R1073K	AGA	AGG (11), AGA (3)	AGG	CGA		
A1119S	GCA (1313), GCT (42), GCG (1)	GCT	GCA	GCA		
P1163A	CCA	CCA	CCA	CCT		
D1189S	GAT	GAT	GAC	GAT		
A1214S	GCA	GCT	GCT	GCA		
R1217K	AGA	AGA (11), AGG (3)	CGT	CGT		
D1237E	GAC (1353), GAT (3)	GAT	GAT	GAC		
R1354K	CGG	AGG	CGG	CGA		
T1366A	ACA	ACG	ACG	ACA		
I1408M	ATT	ATT	ATC	ATA		
I1414L	ATT	ATA	ATT	ATT		
S1436N	AGC	AGT	AGC	AGC		
S1473C	AGT	AGT	AGT	AGT		
L1488Y	CTT	CTC (11), CTT (3)	CTC	CTT		
I1499L	ATA	ATC (13), ATT (1)	ATC	ATA		
S1506A	TCA	TCC	TCG	TCG		
I1509V	ATA	ATC	ATA	ATC		
L1624Y	CTT	CTA	CTA	TTA		
C1628S	TGT	TGC	TGC	TGT		
V1762I	GTC (1342), GTT (14)	GTA	GTC	GTA		
V1850T	GTT	GTT	GTC (4), GTT (4)	GTA		
T1873S	ACT (1355), ATT (1)	ACC	ACC	ACT	Yes	All residues are T across all species

						except 1 Ebola sequence which has an I (ATT)
R1916N	AGG	AGG	AGA	AGG		
E1941R	GAA	GAA	GAG (4), GAA (4)	GAA		
L2044I	TTA	CTT (11), CTC (3)	TTA	CTT		
S2077T	TCA	TCG	TCA	TCT		
E2098D	GAA	GAG	GAA	GAG		
L2157V	TTG	CTT	TTA	CTA		
R2168H	AGA (1355), AGG (1)	CGT (11), CGC (3)	AGG	CGA		
R2175K	CGT	AGG	CGA	CGG		
L2177F	TTA	CTG	TTA	CTA		
M2186L	ATG	ATG	ATG	ATG		

Supplementary Table 11. Codons present for each SDP residue of each pathogenic species, alongside information about non-synonymous variants for NP. Highlighted in yellow are the codons that are 100% conserved across all pathogenic species.

SDP	Ebola (1356)	Sudan (14)	Bundibugyo (8)	Tai Forest (3)	Non-Synonymous?	Notes
R4G	CGT	CGG	CGT	CGG		
E16D	GAA	GAA (11), GAG (3)	GAA	GGT	Yes	All Ebola, Sudan and Bundibugyo sequences have an E residue, Tai Forest has a G
S30T	TCC	TCG (11), TCA (3)	TCC	TCA		
R39K	AGA	AGA	AGA	CGG		
I56V	ATC	ATC	ATC	ATC		
V64I	GTT	GTA	GTC	GTT		
R105K	CGT	AGG	CGT	CGC		
M137L	ATG	ATG	ATG	ATG		
F212Y	TTT	TTC (11), TTT (3)	TTC	TTC		
K274R	AAA	AAG	AAA	AAG		
S279A	TCC	TCA	TCT	TCC		
K416N	AAA	AAG	AAA	AAG		
Y421Q	TAC (1354), TAT (1)	TAT	TAT	TAT		
D426E	GAC	GAT	GAC	GAT		
D435N	GAT	GAT	GAT	GAT		
D443E	GAT	GAT	GAT	GAC		
T453I	ACT	ACT	ACA	ACC		
P497A	CCA (1316), TCA (40)	CCA	CCG	CGA	Yes	Tai Forest sequences have an R residue. All other species have codons for a P residue except 40 Ebola sequences which have a TCA codon (S)

T563S	ACC (1295), ACA (53)	ACC	ACT	ACT		
I565V	ATC (1323), ATT (25)	ATA	ATC	ATC		
P602T	CCC	CCA	CCT	AAT	Yes	All Ebola, Sudan and Bundibugyo sequences have a P residue, Tai Forest has an N
N641Q	AAC	AAC	AAT	AAA	Yes	All Ebola, Sudan and Bundibugyo sequences have an N residue, Tai Forest has a K
A705R	GCC (1314), GCT (42)	GCC	GCC	GCC		
G717N	GGT	GGC	GGT	GAT	Yes	All Ebola, Sudan and Bundibugyo sequences have a G residue, Tai Forest has a D

Supplementary Table 12. Codons present for each SDP residue of each pathogenic species, alongside information about non-synonymous variants for VP24. Highlighted in yellow are the codons that are 100% conserved across all pathogenic species.

SDP	Ebola (1356)	Sudan (14)	Bundibugyo (8)	Tai Forest (3)	Non-Synonymous?	Notes
L17M	CTG (1348), CTT (5), CTA (1)	CTA	CTC	CTT		
V22I	GTC	GTG (11), GTA (3)	GTT	GTT		
V31I	GTT	GTG	GTT	GTG		
I102L	ATA (1354), GTA (2)	ATT	ATT	ATT	Yes	All residues are I across all species except 2 Ebola sequences which have a V (GTA)
T131S	ACT (1355), ACC (1)	ACT	ACA	ACA		
N132T	AAC	AAT	AAC	AAC		
M136L	ATG (1341), ATA (15)	ATG	ATG	ATG	Yes	All residues are M across all species except 15 Ebola sequences which have an I (ATA)
Q139R	CAA	CAA	CAG	CAA		
R140S	CGT	CGA	CAC	CAG	Yes	All Ebola and Sudan have codons for an R residue. All Bundibugyo codons produce an H residue and all Tai

						Forest codons produce a Q residue
T226A	ACA	ACA (11), ACC (3)	ACC	ACC		
S248L	TCT	TCT	TCC	TCT		

Supplementary Table 13. Codons present for each SDP residue of each pathogenic species, alongside information about non-synonymous variants for VP30. Highlighted in yellow are the codons that are 100% conserved across all pathogenic species.

SDP	Ebola (1356)	Sudan (14)	Bundibugyo (8)	Tai Forest (3)	Non-Synony-mous?	Notes
Y39R	TAC (1352), CAC (2), TAT (1)	TAC	TAT	TAC	Yes	All residues are Y across all species except 2 Ebola sequences which have an H (CAC)
T52N	ACT	ACG(11), ACA (3)	ACT	ACT		
V53L	GTA(1339),GTG(16)	GTT(11), GTC (3)	GTG	GTC		
T63I	ACA	ACT	ACA	ACA		
E93D	GAA(1314),GAG(42)	GAA	GAA	GAA		
T96N	ACT(1341),ACG(14)	ACC	ACA	ACA		
R98H	AGG	CGG	AGG	AGA		
K107R	AAG(1354),AAA(1)	AAG	AAA	AAG		
S111I	TCA	TCA	TCC	TCC		
L116S	TTA	CTT	TTG	CTA		
N117Q	AAT	AAT	AAC	AGC	Yes	All Ebola, Sudan and Bundibugyo sequences have an N residue, Tai Forest has an S
A120S	GCA	GCT	GCT	GCT		
T150I	ACG(1344),ACA(11)	ACT	ACT	ACA		
Q157R	CAA	CAG	CAA	CAG		
I159L	ATC	ATT	ATC	ATT		
E205D	GAA	GAA	GAA	GAG		
R262A	AGA	CGC	AGG	AGA		
S268Q	TCA (1355), TCG (1)	AGC	TCA	TCG		
E271S	GAG	GAA	ACC	ACT	Yes	All Ebola and Sudan have codons for an S residue, All Bundibugyo and Tai Forest have codons for a T residue

G278N	GGG	GGG (11) GGA (3)	GAG	GAA	Yes	All Ebola and Sudan have codons for a G residue, All Bundibugyo and Tai Forest have codons for an E residue

Supplementary Table 14. Codons present for each SDP residue of each pathogenic species, alongside information about non-synonymous variants for VP35. Highlighted in yellow are the codons that are 100% conserved across all pathogenic species.

SDP	Ebola (1356)	Sudan (14)	Bundibugyo (8)	Tai Forest (3)	Non-Synonymous?	Notes
S26T	TCG	TCT	TCC	TCA		
E48D	GAG (1352), GAA (4)	GAA	GAA	GAG		
D76E	GAC (1346), GGC (3), GAT (6)	GAT	GAC	GAT	Yes	All residues are D across all species except 3 Ebola sequences which have a G (GGC)
E84K	GAG (1348), GGG (1)	GAA	GCA	GAA	Yes	All Bundibugyo sequences have an A residue. All Ebola, Sudan and Tai Forest have an E residue except one Ebola sequence which has a G
E85K	GAG (1347), GAC (1)	GAA (11), GAG (3)	GAG	GAC	Yes	All Tai Forest and one Ebola sequences have a D residue. All other sequences have codons for an E residue
S92M	TCA	TCG	TCT	TCA		
V97T	GTG	GTG	GTA	ATA	Yes	All Ebola, Sudan and Bundibugyo sequences have a V residue, Tai Forest has an I
T101N	ACC (1351), ACT (3)	ACC	ACC	GCT	Yes	All Ebola, Sudan and

						Bundibugyo sequences have a T residue, Tai Forest has an A
S106A	TCA	TCA	TCA	TCT		
T112S	ACG	ACA (11), GCA (3)	ACT (4), ACC (4)	ATA	Yes	All Ebola and Bundibugyo have a T residue. All Tai Forest have an I residue. 11 Sudan sequence have a T residue while 3 have an A residue
V121I	GTT	GTT	GTG	ATG	Yes	All Ebola, Sudan and Bundibugyo sequences have a V residue, Tai Forest has an M
A125G	GCA (1337), GCT (18), GTA (1)	GCA	ACC	GCT	Yes	All Bundibugyo sequences have a T residue. All others have an A residue except 1 Ebola sequence which has a V (GTA)
A154S	GCA	GCC	GCC	GCC		
T159V	ACT	ACA	ACT	ACT		
E160D	GAG (1342), GAA (14)	GAA	GAA	GAG		
G167K	GGT	GGA	GGA	GGA		
S174A	TCA	TCA	TCA	TCA		
I181L	ATT	ATT (11), ATC (3)	ATC	ATT		
E269D	GAA (1352), GAG (1)	GAG	GAA	GAA		
A290V	GCT	GCC	GCA	GCC		

V314A	GTC	GTC	GTT	GTT		
Q329K	CAG	CAA	CAG	CAA		

Supplementary Table 15. Codons present for each SDP residue of each pathogenic species, alongside information about non-synonymous variants for VP40. Highlighted in yellow are the codons that are 100% conserved across all pathogenic species.

SDP	Ebola (1356)	Sudan (14)	Bundibugyo (8)	Tai Forest (3)	Non-Synonymous?	Notes
V4G	GTT (1348), ATT (1)	GTC (11), GTT (3)	GCA	ATC	Yes	All Bundibugyo sequences have an A residue. All Tai Forest have an I. All Ebola and Sudan have a V except one Ebola has an I (ATT)
T46V	ACT (1355), ATT (1)	ACA	ACA	ACT	Yes	All residues are D across all species except 1 Ebola sequence which has an I (ATT)
P85T	CCC (1348), CCT (6)	CCC	CCG	CCG		
T105I	ACC (1355), ATG (1)	ACG (11), ACA (3)	ACA	ACA	Yes	All residues are T across all species except 1 Ebola sequence which has an M (ATG)
I122V	ATC	ATC	ATC	ATC		
A128I	GCA	GCC	ACC	ACC	Yes	All Ebola and Sudan have codons for an A residue, All Bundibugyo and Tai Forest have codons for a T residue
G201N	GGA	GGA	GGA	GGC		
F209L	TTT	TTT	TTT	TTC		
L244I	CTC	ATG	CTC	CTA	Yes	All Ebola, Bundibugyo and Tai Forest have an L residue. All Sudan have an M residue
Q245P	CAG	CAA	CAA	CAA		
M259V	ATG	ATT	ATG	ATG	Yes	All Ebola, Bundibugyo and Tai Forest have an M residue. All Sudan have an I residue
H269Q	CAC (1354), CGC (1)	CAC	CAC	CAC	Yes	All residues are H across all species except 1 Ebola

						sequence which has an R (CGC)
T277Q	ACT (1355), ACC (1)	AGT	ACA	ACT	Yes	All Ebola, Bundibugyo and Tai Forest have a T residue. All Sudan have an S residue
I293V	ATT (1354), ATC (1)	ATT	ATT	ATT		
V323H	GTG (1354), ATG (1), GCG (1)	CTC	GTC	GTC	Yes	All Sudan have an S residue All others have a T residue except 2 Ebola sequences. One has an M residue (ATG) and one has an A (GCG).
E325D	GAG	GAA	GAG	GAA		

Supplementary Table 16. Codons present for each SDP residue of each pathogenic species, alongside information about non-synonymous variants for the lost SDPs. Highlighted in yellow are the codons that are 100% conserved across all pathogenic species.

SDP	Ebola (1356)	Sudan (14)	Bundibugyo (8)	Tai Forest (3)	Non-Synony-mous?	Notes
GP						
G2S	GGT:1343,GGC:13	GGG:11, GAG:3	GTT	GGA	Yes	3 Sudan sequences have an E residue, all Bundibugyo sequences have a V residue, all other sequences have a G
E207D	GAG	GAA	ACA	ACG	Yes	Ebola and Sudan have an E residue, Bundibugyo and Tai Forest have a T
S210T	TCG:1245, TCT:51, TCC:11	TCA	TCC	TCC		
R325G	CGA	AGA	GTC	GTC	Yes	Ebola and Sudan have an R residue, Bundibugyo and Tai Forest have a V
H354L	CAC:1355, CAT:1	CAC	CGA	CAA	Yes	Ebola and Sudan have an H residue, Bundibugyo has an R and Tai Forest has a Q
Q403E	CAA	CCA	CCA	CCA	Yes	Ebola has a Q residue, all other species have a P residue
S418T	TCC	CAC	CGC	CAC	Yes	Ebola has an S residue, Sudan and Tai Forest have an H residue, Bundibugyo has an R residue
T448M	ACC:1345, GCC:3	ACC	AGC	ACC	Yes	Sudan has an S residue, all others have a T residue

						except three Ebola sequences which have an A residue
H516H	CAT	CAC	CAC	CAC		
L547V	CTA:1323, CTG:29	CTG:11, CTT:3	ATA	ATA	Yes	Ebola and Sudan have an L residue, Bundibugyo and Tai Forest have an I
D642L	GAC	GAT	GAC	AGC	Yes	Ebola, Sudan and Bundibugyo have a D residue, Tai Forest has an S
L						
Q109H	CAA	CAG	CAG	CAA		
L276I	CTT	CTG	CTG	CTA		
Y312F	TAC	TAT	TAC	TAC		
A326S	GCT	GCA	GCC	GCT		
E689S	GAA:1341,GAG:13	GAG	GAA	GAA		
F896Y	TTC	TTC:11, TTT:3	TTC	TTT		
L925F	CTA:1350, TTA:3	TTG:11, CTG:3	CTT	CTG		
A954S	GCG:1336,GCA:18	GCA	GCC	GCA		
S995T	AGT	TCG	AGT	AGT		
I1255V	ATA:1355, GTA:1	ATT	ATC	ATT	Yes	All sequences have an I residue except one Ebola sequence which has a V residue
S1395T	TCA:1352, TCG:1, TCT:1, GCA:1	TCG	TCC	TCT	Yes	All sequences have an S residue except one Ebola sequence which has an A residue
K1461Q	AAA:1162,AAG:193	AAA	AAG	AAG		
A1538S	GCA	GCA	GCA	GCT		
L2008I	TTA	CTT	CTT	CTT		
Q2105L	CAA	CAG	CAG	CAA		
Q2108E	CAA	CAA	CAA	CAA		
Y2131F	TAT:1340, TAC:15	TAT	TAC	TAC		
NP						
P42S	CCA:1346, CAA:9, TCA:1	CCG	CCT	CAA	Yes	Tai Forest sequences have a Q residue, all others

						have a P except one Ebola sequence which has an S
K374R	AAA:1355, AGA:1	AAG	AAA	AAG	Yes	All sequences have a K residue except one Ebola sequence which has an R residue
D492E	GAC:1270,GAT:29	GAT	GAT	GAC		
P526A	CCA	GTG	GAA	CCG	Yes	Ebola and Tai Forest have a P residue, Sudan has a V residue and Bundibugyo has an E
D716N	GAT:1353, GAC:1, AAT:1	GAT	GAT	GAT	Yes	All sequences have a D residue except one Ebola sequence which has an N residue
VP35						
R196H	CGC:1336, CGT:18, CAC:1	AGG	CGA	CGA	Yes	All sequences have an R residue except one Ebola sequence which has a H residue

Supplementary Table 17. SDPs mapped to known PDB structures or modelled structures. *There was no available structure or model of L when the previous study was carried out

Protein	Total SDPs	Old Mapped	New Mapped	Total Mapped
VP24	11	8	2	10
VP30	20	5	0	5
VP35	22	4	11	15
VP40	16	8	5	13
NP	24	8	n/a	8
GP	21	10	0	10
L	51	0*	31	31
Total	165	43	49	92

Supplementary Table 18: Summary of the structures used for SDP investigation.

Protein	Species	PDB Structure ID	Oligomeric Form	Residue Coverage
VP24	EBOV	4M0Q	Homodimer	11 - 237
	EBOV	4U2X	Heterodimer (with KPNA5)	16 - 231
VP30	EBOV	2I8B	Homodimer	142 - 272
VP35	EBOV	4IBC	Homodimer	215 - 340
	EBOV	3L26	Homodimer (bound to RNA)	215 - 340
	EBOV	6GBO	Homotrimer	81 - 153
VP40	EBOV	4LDB	Homodimer	44 - 326
	EBOV	4LDD	Homo 6-mer	44 - 326
	EBOV	4LDM	Homo 8-mer	44 - 188
NP	EBOV	4QB0	Monomer	641 - 739
	EBOV	4YPI	Heterodimer	38 - 385
GP	EBOV	5JQ3	Hetero 6-mer	32 – 501 & 502 – 632
L	EBOV	N/A (Phyre2 model)	Monomer	8 - 2010
Nucleocapsid (NP with VP24)	EBOV	6EHM	Hetero 4-mer	NP: 1 – 739 VP24: 1 - 251

Supplementary Table 19: Summary of the SDPs with proposed functional impacts identified using the 196 and the 1,408 genome sets.

Protein	SDP	Functional Effect	Confidence	Status
NP	R105K	Stability: Loss of hydrogen bonding	Possible	Retained
NP	A705R	Stability: Introduction of salt bridge with E694	Possible	Lost
VP35	E269D	Interface: Dimeric VP35 interface	Probable	Retained
VP40	P85T	Interface: Octameric VP40 interface	Probable	Retained
VP40	Q245P	Stability: Breaks an alpha helix	Probable	Retained
GP	I260L	Interface: Within GP glycan cap	Possible	Retained
GP	T269S	Interface: Within GP glycan cap	Possible	Retained
GP	S307H	Interface: Within GP glycan cap	Possible	Retained
VP30	R262A	Interface: Dimer interface, loss of hydrogen bond	Probable	Retained
VP24	T131S	Interface: With KPNA5	Probable	Retained
VP24	M136L	Interface: With KPNA5	Probable	Retained
VP24	Q139R	Interface: With KPNA5	Probable	Retained
VP24	R140S	Interface: With KPNA5	Probable	Gained
VP24	T226A	Stability: Loss of hydrogen bond	Probable	Retained

Supplementary Table 20: Comparison of *Bombali virus* SDP amino acids with *Ebola virus* and *Reston virus*

Protein	Number of SDPs	Residues the same as: (%)		
		Ebola virus	Reston virus	Neither
GP	20	55	10	35
L	53	77.36	11.32	11.32
NP	24	62.5	12.5	25
VP24	11	72.73	18.18	9.09
VP30	20	30	10	60
VP35	22	72.73	13.64	13.64
VP40	16	50	18.75	31.25
Total	**166**	**63.25**	**12.65**	**24.1**

Supplementary Table 21: Summary of source databases used to obtain Ebolavirus genomes for analysis.

Species	NCBI	ViPR	Urbanowicz	Total
Ebola virus	1,469	43	505	2,017
Sudan virus	14	5	0	19
Bundibugyo virus	7	2	0	9
Tai Forest virus	1	3	0	4
Reston virus	18	9	0	27
Total	1,509	62	505	2,076

Supplementary Table 22: Summary of the sequences removed from the initial set of ebolavirus genome sequences.

Species	Starting	Removed	Final
Ebola virus	2,017	661	1,356
Sudan virus	19	5	14
Bundibugyo virus	9	1	8
Tai Forest virus	4	1	3
Reston virus	27	0	27
Total	2,076	668	1,408

Genome Accession	Species	Date	Location	Source Database
KR817200	Ebola virus	13/04/2014	Guinea	NCBI
KY426702	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KR105335	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KR817214	Ebola virus	22/05/2014	Guinea	NCBI
KT725276	Ebola virus	23/09/2014	Liberia	NCBI
KR817090	Ebola virus	20/12/2014	Guinea	NCBI
KR817187	Ebola virus	27/03/2014	Guinea	NCBI
CDC-NIH-2984	Ebola virus	23/10/2014	Liberia	Urbanowicz et al.
KR817201	Ebola virus	18/04/2014	Guinea	NCBI
KP759696	Ebola virus	06/11/2014	Sierra_Leone	NCBI
KR817183	Ebola virus	31/03/2014	Guinea	NCBI
KR817204	Ebola virus	01/05/2014	Guinea	NCBI
EM_FORE_2015_1047	Ebola virus	30/06/2015	Guinea	Urbanowicz et al.
KT725380	Ebola virus	17/08/2014	Liberia	NCBI
KR075000	Ebola virus	2015/--/--	Liberia	NCBI
KR534566	Ebola virus	20/10/2014	Guinea	NCBI
KX121420	Ebola virus	04/11/2014	Guinea	NCBI
KT725264	Ebola virus	26/08/2014	Liberia	NCBI
KT725300	Ebola virus	26/08/2014	Liberia	NCBI
KM233059	Ebola virus	12/06/2014	Sierra_Leone	NCBI
KR534589	Ebola virus	04/10/2014	Guinea	NCBI
KP759631	Ebola virus	27/09/2014	Sierra_Leone	NCBI
KM233085	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KT357859	Ebola virus	11/07/2015	Sierra_Leone	NCBI
REDC_GUI_2015_00494B	Ebola virus	12/07/2015	Guinea	Urbanowicz et al.
KT725268	Ebola virus	02/10/2014	Liberia	NCBI
KP759767	Ebola virus	28/10/2014	Sierra_Leone	NCBI
KP120616	Ebola virus	25/08/2014	United_Kingdom	NCBI
KR817102	Ebola virus	15/01/2015	Guinea	NCBI
KM233110	Ebola virus	18/06/2014	Sierra_Leone	NCBI
KR817143	Ebola virus	18/10/2014	Guinea	NCBI
HC069219	Reston virus	NA	NA	ViPR
KU220280	Ebola virus	24/08/2014	Liberia	NCBI
KR105271	Ebola virus	05/08/2014	Sierra_Leone	NCBI
KR817159	Ebola virus	13/11/2014	Guinea	NCBI
CDC-NIH-1650	Ebola virus	27/09/2014	Liberia	Urbanowicz et al.
KR817224	Ebola virus	09/06/2014	Guinea	NCBI
KX009911	Ebola virus	25/11/2014	Liberia	NCBI
KC242787	Ebola virus	2007/--/--	Democratic_Republic_of_the_Congo	NCBI
KC242796	Ebola virus	1995/--/--	Democratic_Republic_of_the_Congo	NCBI
KR534524	Ebola virus	11/09/2014	Guinea	NCBI
KP759754	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KX013101	Ebola virus	22/07/2014	Nigeria	NCBI
KR534537	Ebola virus	28/09/2014	Guinea	NCBI
KT357839	Ebola virus	21/01/2015	Sierra_Leone	NCBI
KR105343	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KP759649	Ebola virus	23/10/2014	Sierra_Leone	NCBI

CDC-NIH-3762	Ebola virus	19/11/2014	Liberia	Urbanowicz et al.
KT725372	Ebola virus	02/10/2014	Liberia	NCBI
KM233111	Ebola virus	18/06/2014	Sierra_Leone	NCBI
KU220281	Ebola virus	25/08/2014	Liberia	NCBI
KP759734	Ebola virus	28/09/2014	Sierra_Leone	NCBI
KR817073	Ebola virus	07/09/2014	Guinea	NCBI
KT357849	Ebola virus	04/02/2015	Sierra_Leone	NCBI
KT725368	Ebola virus	01/10/2014	Liberia	NCBI
EM_FORE_2015_934	Ebola virus	23/06/2015	Guinea	Urbanowicz et al.
KY426724	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KT765130	Ebola virus	2014/--/--	Guinea	NCBI
KR105250	Ebola virus	19/07/2014	Sierra_Leone	NCBI
KR105260	Ebola virus	22/07/2014	Sierra_Leone	NCBI
KP240935	Ebola virus	13/10/2014	USA	NCBI
KU143803	Ebola virus	2014/--/--	Sierra_Leone	NCBI
FJ217162	Tai Forest virus	1994/11/--	Cote_d'Ivoire	NCBI
KU296819	Ebola virus	27/05/2015	Sierra_Leone	NCBI
KR817125	Ebola virus	14/08/2014	Liberia	NCBI
KR653226	Ebola virus	18/10/2014	Sierra_Leone	NCBI
KT725299	Ebola virus	10/08/2014	Liberia	NCBI
KT725362	Ebola virus	20/09/2014	Liberia	NCBI
KT725274	Ebola virus	07/11/2014	Liberia	NCBI
KR063671	Ebola virus	01/10/1976	Democratic_Republic_of_the_Congo	ViPR
KP759693	Ebola virus	06/11/2014	Sierra_Leone	NCBI
KR105312	Ebola virus	10/09/2014	Sierra_Leone	NCBI
KM233112	Ebola virus	18/06/2014	Sierra_Leone	NCBI
KM233045	Ebola virus	04/06/2014	Sierra_Leone	NCBI
KP759672	Ebola virus	01/11/2014	Sierra_Leone	NCBI
KY426732	Ebola virus	2015/--/--	Sierra_Leone	NCBI
CDC-NIH-3878	Ebola virus	25/11/2014	Liberia	Urbanowicz et al.
KY425636	Ebola virus	04/05/1995	Democratic_Republic_of_the_Congo	NCBI
KX013093	Ebola virus	06/08/2014	Nigeria	NCBI
KR105249	Ebola virus	19/07/2014	Sierra_Leone	NCBI
KR105225	Ebola virus	11/07/2014	Sierra_Leone	NCBI
KM233036	Ebola virus	02/06/2014	Sierra_Leone	NCBI
KP240931	Ebola virus	13/09/2014	Sierra_Leone	NCBI
KY798004	Reston virus	1989/--/--	USA	NCBI
KP271019	Ebola virus	20/08/2014	Democratic_Republic_of_the_Congo	NCBI
KR817119	Ebola virus	01/08/2014	Guinea	NCBI
KR817169	Ebola virus	11/12/2014	Guinea	NCBI
KR817077	Ebola virus	13/09/2014	Guinea	NCBI
KR817105	Ebola virus	27/01/2015	Guinea	NCBI
KT725351	Ebola virus	19/09/2014	Liberia	NCBI
KP759650	Ebola virus	25/10/2014	Sierra_Leone	NCBI
KG80	Ebola virus	18/06/2015	Guinea	Urbanowicz et al.
KR817237	Ebola virus	25/06/2014	Guinea	NCBI
KR817115	Ebola virus	25/07/2014	Liberia	NCBI
KC545390	Sudan virus	2012/--/--	Uganda	NCBI

EM_COY_2015_016278	Ebola virus	05/05/2015	Guinea	Urbanowicz et al.
KY366414	Ebola virus	2015/03/--	Sierra_Leone	NCBI
HQ613403	Ebola virus	31/08/2007	Democratic_Republic_of_the_Congo	NCBI
KC545394	Bundibugyo virus	2012/--/--	Democratic_Republic_of_the_Congo	NCBI
KR534585	Ebola virus	04/10/2014	Guinea	NCBI
KR105242	Ebola virus	14/07/2014	Sierra_Leone	NCBI
KU143789	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KY426690	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KT725336	Ebola virus	25/06/2014	Liberia	NCBI
KC545392	Sudan virus	2012/--/--	Uganda	NCBI
KU143806	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR653266	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KR105283	Ebola virus	14/08/2014	Sierra_Leone	NCBI
KT725297	Ebola virus	26/08/2014	Liberia	NCBI
KT357829	Ebola virus	17/01/2015	Sierra_Leone	NCBI
JX477166	Reston virus	1996/--/--	USA	NCBI
KR105222	Ebola virus	11/07/2014	Sierra_Leone	NCBI
KY426718	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP759639	Ebola virus	28/09/2014	Sierra_Leone	NCBI
KP759749	Ebola virus	18/10/2014	Sierra_Leone	NCBI
KR817126	Ebola virus	16/08/2014	Guinea	NCBI
KR817230	Ebola virus	20/06/2014	Guinea	NCBI
KM233043	Ebola virus	03/06/2014	Sierra_Leone	NCBI
KP759701	Ebola virus	08/11/2014	Sierra_Leone	NCBI
KR653243	Ebola virus	15/11/2014	Sierra_Leone	NCBI
KR105217	Ebola virus	08/07/2014	Sierra_Leone	NCBI
KU143807	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR105218	Ebola virus	08/07/2014	Sierra_Leone	NCBI
KT013254	Ebola virus	19/03/2014	Guinea	NCBI
KR817236	Ebola virus	24/06/2014	Sierra_Leone	NCBI
KP759636	Ebola virus	27/09/2014	Sierra_Leone	NCBI
KY798010	Reston virus	2008/--/--	Philippines	NCBI
CDC-NIH-3832	Ebola virus	23/11/2014	Liberia	Urbanowicz et al.
KR817080	Ebola virus	22/09/2014	Guinea	NCBI
KR817197	Ebola virus	10/04/2014	Guinea	NCBI
KR653225	Ebola virus	05/10/2014	Sierra_Leone	NCBI
KU182899	Ebola virus	04/05/1995	Zaire	ViPR
KY426722	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KX009897	Ebola virus	04/09/2014	Liberia	NCBI
KR653302	Ebola virus	23/10/2014	Sierra_Leone	NCBI
KU182909	Ebola virus	04/05/1995	Zaire	ViPR
KR105235	Ebola virus	14/07/2014	Sierra_Leone	NCBI
KR817087	Ebola virus	01/10/2014	Guinea	NCBI
KY426694	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP759673	Ebola virus	02/11/2014	Sierra_Leone	NCBI
KX000398	Ebola virus	10/03/2014	Guinea	NCBI
KR653258	Ebola virus	30/10/2014	Sierra_Leone	NCBI
KP759743	Ebola virus	17/10/2014	Sierra_Leone	NCBI
CDC-NIH-2303	Ebola virus	10/10/2014	Liberia	Urbanowicz et al.

KT725307	Ebola virus	20/09/2014	Liberia	NCBI
KU296773	Ebola virus	29/01/2015	Sierra_Leone	NCBI
KU143797	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KP759712	Ebola virus	04/10/2014	Sierra_Leone	NCBI
KU182907	Ebola virus	04/05/1995	Zaire	ViPR
KY426698	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KR534584	Ebola virus	02/10/2014	Guinea	NCBI
KR817128	Ebola virus	17/08/2014	Liberia	NCBI
KT725385	Ebola virus	09/09/2014	Liberia	NCBI
KR817161	Ebola virus	24/11/2014	Guinea	NCBI
EM_FORE_2015_1118	Ebola virus	02/07/2015	Guinea	Urbanowicz et al.
KT725256	Ebola virus	26/08/2014	Liberia	NCBI
KM233054	Ebola virus	07/06/2014	Sierra_Leone	NCBI
KP759671	Ebola virus	29/10/2014	Sierra_Leone	NCBI
KR105229	Ebola virus	12/07/2014	Sierra_Leone	NCBI
KR534567	Ebola virus	22/10/2014	Guinea	NCBI
KX009910	Ebola virus	25/11/2014	Liberia	NCBI
KP759647	Ebola virus	23/10/2014	Sierra_Leone	NCBI
KU296527	Ebola virus	16/03/2015	Sierra_Leone	NCBI
KU143795	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR817210	Ebola virus	14/05/2014	Guinea	NCBI
KP759616	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KP759597	Ebola virus	29/10/2014	Sierra_Leone	NCBI
KP759753	Ebola virus	16/10/2014	Sierra_Leone	NCBI
KM233058	Ebola virus	10/06/2014	Sierra_Leone	NCBI
KU220284	Ebola virus	16/01/2015	Liberia	NCBI
KR534513	Ebola virus	14/08/2014	Guinea	NCBI
KC242786	Ebola virus	2007/--/--	Democratic_Republic_of_the_Congo	NCBI
KR817241	Ebola virus	04/07/2014	Liberia	NCBI
EM_COY_2015_017865	Ebola virus	18/06/2015	Guinea	Urbanowicz et al.
KT357843	Ebola virus	26/01/2015	Sierra_Leone	NCBI
KX009902	Ebola virus	16/09/2014	Liberia	NCBI
KR653246	Ebola virus	03/09/2014	Sierra_Leone	NCBI
KU220270	Ebola virus	30/06/2015	Liberia	NCBI
KT357821	Ebola virus	09/03/2015	Sierra_Leone	NCBI
KM233050	Ebola virus	09/06/2014	Sierra_Leone	NCBI
KT725311	Ebola virus	10/08/2014	Liberia	NCBI
KR534538	Ebola virus	02/10/2014	Guinea	NCBI
KU220278	Ebola virus	24/08/2014	Liberia	NCBI
KR006948	Ebola virus	10/11/2014	Liberia	NCBI
KY426701	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KR817238	Ebola virus	26/06/2014	Sierra_Leone	NCBI
KY425654	Ebola virus	19/03/2014	Guinea	NCBI
KY426719	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KR653259	Ebola virus	14/11/2014	Sierra_Leone	NCBI
KR653233	Ebola virus	04/10/2014	Sierra_Leone	NCBI
KR534562	Ebola virus	18/10/2014	Guinea	NCBI
KX009899	Ebola virus	06/09/2014	Liberia	NCBI
KR105291	Ebola virus	15/08/2014	Sierra_Leone	NCBI
KR817217	Ebola virus	28/05/2014	Guinea	NCBI
EM_COY_2015_016617	Ebola virus	16/05/2015	Guinea	Urbanowicz et al.
KP759718	Ebola virus	29/09/2014	Sierra_Leone	NCBI

KT725281	Ebola virus	27/08/2014	Liberia	NCBI
KR074998	Ebola virus	2015/--/--	Liberia	NCBI
KC242798	Ebola virus	1996/--/--	Gabon	NCBI
KR653261	Ebola virus	24/09/2014	Sierra_Leone	NCBI
KM233052	Ebola virus	13/06/2014	Sierra_Leone	NCBI
KT357854	Ebola virus	18/02/2015	Sierra_Leone	NCBI
KR105280	Ebola virus	13/08/2014	Sierra_Leone	NCBI
KP759603	Ebola virus	30/10/2014	Sierra_Leone	NCBI
KG91	Ebola virus	20/06/2015	Guinea	Urbanowicz et al.
KR817182	Ebola virus	31/03/2014	Guinea	NCBI
KP759669	Ebola virus	29/10/2014	Sierra_Leone	NCBI
FJ621585	Reston virus	2008/--/--	Philippines	NCBI
KY798009	Reston virus	1996/--/--	USA	NCBI
KR817101	Ebola virus	14/01/2015	Guinea	NCBI
KU143788	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KY558984	Ebola virus	01/12/2014	Liberia	NCBI
KR653238	Ebola virus	08/11/2014	Sierra_Leone	NCBI
CDC-NIH-2313	Ebola virus	10/10/2014	Liberia	Urbanowicz et al.
KY426695	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KY798007	Reston virus	1992/--/--	Italy	NCBI
KR534543	Ebola virus	04/10/2014	Guinea	NCBI
KM233105	Ebola virus	17/06/2014	Sierra_Leone	NCBI
KR105244	Ebola virus	16/07/2014	Sierra_Leone	NCBI
KT725315	Ebola virus	16/08/2014	Liberia	NCBI
KC545391	Sudan virus	2012/--/--	Uganda	NCBI
KT725360	Ebola virus	23/06/2014	Liberia	NCBI
KR653277	Ebola virus	27/10/2014	Sierra_Leone	NCBI
KP759657	Ebola virus	27/10/2014	Sierra_Leone	NCBI
KR534573	Ebola virus	25/10/2014	Guinea	NCBI
KP759682	Ebola virus	31/10/2014	Sierra_Leone	NCBI
KR006961	Ebola virus	02/02/2015	Liberia	Urbanowicz et al.
KR105316	Ebola virus	13/09/2014	Sierra_Leone	NCBI
KT725389	Ebola virus	29/09/2014	Liberia	NCBI
KY366412	Ebola virus	2015/03/--	Sierra_Leone	NCBI
EM_COY_2015_017788	Ebola virus	16/06/2015	Guinea	Urbanowicz et al.
KR653260	Ebola virus	28/09/2014	Sierra_Leone	NCBI
KR817092	Ebola virus	22/12/2014	Guinea	NCBI
KR105323	Ebola virus	17/09/2014	Sierra_Leone	NCBI
KC242801	Ebola virus	1976/--/--	Democratic_Republic_of_the_Congo	NCBI
KT725323	Ebola virus	29/09/2014	Liberia	NCBI
KR817127	Ebola virus	17/08/2014	Liberia	NCBI
KT725258	Ebola virus	01/10/2014	Liberia	NCBI
KR817232	Ebola virus	20/06/2014	Liberia	NCBI
KU143796	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KP759759	Ebola virus	25/10/2014	Sierra_Leone	NCBI
KR105318	Ebola virus	14/09/2014	Sierra_Leone	NCBI
KX121421	Ebola virus	23/12/2014	Guinea	NCBI
KP759700	Ebola virus	08/11/2014	Sierra_Leone	NCBI

KR534579	Ebola virus	24/09/2014	Guinea	NCBI
KR105239	Ebola virus	18/07/2014	Sierra_Leone	NCBI
KP759638	Ebola virus	10/10/2014	Sierra_Leone	NCBI
KP759609	Ebola virus	01/10/2014	Sierra_Leone	NCBI
KY426691	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP759746	Ebola virus	18/10/2014	Sierra_Leone	NCBI
KR817120	Ebola virus	04/08/2014	Guinea	NCBI
KR817194	Ebola virus	01/04/2014	Liberia	NCBI
KU296581	Ebola virus	17/01/2015	Sierra_Leone	NCBI
KY366419	Ebola virus	2015/03/--	Sierra_Leone	NCBI
KR105328	Ebola virus	21/09/2014	Sierra_Leone	NCBI
KM233041	Ebola virus	03/06/2014	Sierra_Leone	NCBI
KR534518	Ebola virus	21/08/2014	Guinea	NCBI
KR817164	Ebola virus	04/12/2014	Guinea	NCBI
KU143814	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KU143792	Ebola virus	2014/--/--	Sierra_Leone	NCBI
EM_COY_2015_017238	Ebola virus	31/05/2015	Guinea	Urbanowicz et al.
KT725361	Ebola virus	01/10/2014	Liberia	NCBI
KP759664	Ebola virus	30/10/2014	Sierra_Leone	NCBI
KR534527	Ebola virus	19/09/2014	Guinea	NCBI
KR105326	Ebola virus	16/09/2014	Sierra_Leone	NCBI
KR653270	Ebola virus	12/11/2014	Sierra_Leone	NCBI
KP759711	Ebola virus	04/10/2014	Sierra_Leone	NCBI
EM_FORE_2015_696	Ebola virus	07/06/2015	Guinea	Urbanowicz et al.
KR817229	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KT587346	Ebola virus	05/09/2014	Liberia	NCBI
KM233038	Ebola virus	03/06/2014	Sierra_Leone	NCBI
KT725379	Ebola virus	20/09/2014	Liberia	NCBI
KR817148	Ebola virus	26/10/2014	Guinea	NCBI
KT725338	Ebola virus	20/08/2014	Liberia	NCBI
KP759679	Ebola virus	02/11/2014	Sierra_Leone	NCBI
KR817180	Ebola virus	28/03/2014	Guinea	NCBI
EM_FORE_2015-2533	Ebola virus	26/09/2015	Guinea	Urbanowicz et al.
KR817154	Ebola virus	03/11/2014	Guinea	NCBI
KR817176	Ebola virus	17/12/2014	Guinea	NCBI
KP759724	Ebola virus	08/10/2014	Sierra_Leone	NCBI
KP759695	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KT725383	Ebola virus	17/08/2014	Liberia	NCBI
KP759683	Ebola virus	02/10/2014	Sierra_Leone	NCBI
KJ660346	Ebola virus	2014/--/--	Guinea	NCBI
KR653245	Ebola virus	01/11/2014	Sierra_Leone	NCBI
KR817235	Ebola virus	22/06/2014	Guinea	NCBI
KR105349	Ebola virus	28/09/2014	Sierra_Leone	NCBI
KM233096	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KP759709	Ebola virus	11/11/2014	Sierra_Leone	NCBI
KY425644	Sudan virus	2000/10/--	Uganda	NCBI
KR534553	Ebola virus	09/10/2014	Guinea	NCBI
KR105346	Ebola virus	27/09/2014	Sierra_Leone	NCBI
KR534555	Ebola virus	10/10/2014	Guinea	NCBI
KR534529	Ebola virus	21/09/2014	Guinea	NCBI
KU143798	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KT725267	Ebola virus	16/08/2014	Liberia	NCBI

KR817078	Ebola virus	13/09/2014	Guinea	NCBI
KR817094	Ebola virus	26/12/2014	Guinea	NCBI
KP759705	Ebola virus	11/11/2014	Sierra_Leone	NCBI
KR534545	Ebola virus	06/10/2014	Guinea	NCBI
KR534569	Ebola virus	23/10/2014	Guinea	NCBI
KR653239	Ebola virus	22/08/2014	Sierra_Leone	NCBI
KR653294	Ebola virus	27/08/2014	Sierra_Leone	NCBI
KT725314	Ebola virus	04/08/2014	Liberia	NCBI
KM233099	Ebola virus	17/06/2014	Sierra_Leone	NCBI
KR534516	Ebola virus	19/08/2014	Guinea	NCBI
KR105321	Ebola virus	16/09/2014	Sierra_Leone	NCBI
KR653289	Ebola virus	01/10/2014	Sierra_Leone	NCBI
KR817152	Ebola virus	01/11/2014	Guinea	NCBI
KY744596	Ebola virus	22/11/2015	Liberia	NCBI
KP759689	Ebola virus	04/11/2014	Sierra_Leone	NCBI
KR817240	Ebola virus	03/07/2014	Liberia	NCBI
KR105347	Ebola virus	27/09/2014	Sierra_Leone	NCBI
KT725348	Ebola virus	16/08/2014	Liberia	NCBI
KR105324	Ebola virus	18/09/2014	Sierra_Leone	NCBI
KR105286	Ebola virus	12/08/2014	Sierra_Leone	NCBI
KU143800	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KP759764	Ebola virus	28/10/2014	Sierra_Leone	NCBI
KR653276	Ebola virus	23/09/2014	Sierra_Leone	NCBI
KM655246	Ebola virus	1976/--/--	Democratic_Republic_of_the_Congo	NCBI
KU182912	Sudan virus	16/10/2000	Sudan	ViPR
KT725349	Ebola virus	26/08/2014	Liberia	NCBI
KM233053	Ebola virus	05/06/2014	Sierra_Leone	NCBI
KT725259	Ebola virus	15/08/2014	Liberia	NCBI
KP759691	Ebola virus	30/09/2014	Sierra_Leone	NCBI
KP759655	Ebola virus	26/10/2014	Sierra_Leone	NCBI
KM034554	Ebola virus	27/05/2014	Sierra_Leone	NCBI
KR105205	Ebola virus	19/06/2014	Sierra_Leone	NCBI
KR105245	Ebola virus	16/07/2014	Sierra_Leone	NCBI
KT725294	Ebola virus	16/08/2014	Liberia	NCBI
KP759733	Ebola virus	09/10/2014	Sierra_Leone	NCBI
MF102255	Ebola virus	2015/04/--	Guinea	NCBI
EM_FORE_2015_2878	Ebola virus	24/10/2015	Guinea	Urbanowicz et al.
EM_COY_2015_016236	Ebola virus	03/05/2015	Guinea	Urbanowicz et al.
KT725261	Ebola virus	04/08/2014	Liberia	NCBI
KT357816	Ebola virus	26/02/2015	Sierra_Leone	NCBI
KT357822	Ebola virus	10/03/2015	Sierra_Leone	NCBI
KX013098	Ebola virus	06/08/2014	Nigeria	NCBI
KY426715	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KM233108	Ebola virus	18/06/2014	Sierra_Leone	NCBI
KT725365	Ebola virus	20/08/2014	Liberia	NCBI
KU052670	Ebola virus	08/10/2015	United_Kingdom	NCBI
KR105282	Ebola virus	13/08/2014	Sierra_Leone	NCBI
KY471124	Ebola virus	2001/--/--	Gabon	NCBI
KT725355	Ebola virus	30/08/2014	Liberia	NCBI
KT725331	Ebola virus	22/08/2014	Liberia	NCBI
KF113528	Ebola virus	2003/--/--	Democratic_Republic_of_the_Congo	NCBI
KP759720	Ebola virus	08/10/2014	Sierra_Leone	NCBI
KR105293	Ebola virus	16/08/2014	Sierra_Leone	NCBI

KT357832	Ebola virus	19/01/2015	Sierra_Leone	NCBI
KR534548	Ebola virus	07/10/2014	Guinea	NCBI
KM034562	Ebola virus	28/05/2014	Sierra_Leone	NCBI
KT725255	Ebola virus	08/09/2014	Liberia	NCBI
KU052669	Ebola virus	08/10/2015	United_Kingdom	NCBI
KR534576	Ebola virus	16/10/2014	Guinea	NCBI
KU143832	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KP759755	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KR817097	Ebola virus	02/01/2015	Guinea	NCBI
KU220282	Ebola virus	25/08/2014	Liberia	NCBI
KR105237	Ebola virus	14/07/2014	Sierra_Leone	NCBI
KT357815	Ebola virus	26/02/2015	Sierra_Leone	NCBI
KY426700	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP759732	Ebola virus	09/10/2014	Sierra_Leone	NCBI
KR653230	Ebola virus	03/10/2014	Sierra_Leone	NCBI
KU143804	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR817215	Ebola virus	24/05/2014	Guinea	NCBI
KR006964	Ebola virus	23/09/2014	Liberia	NCBI
KT725317	Ebola virus	09/09/2014	Liberia	NCBI
KY426697	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP759677	Ebola virus	31/10/2014	Sierra_Leone	NCBI
KM034552	Ebola virus	26/05/2014	Sierra_Leone	NCBI
KR006947	Ebola virus	08/11/2014	Liberia	NCBI
KR105344	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KR817202	Ebola virus	22/04/2014	Guinea	NCBI
KP759623	Ebola virus	09/11/2014	Sierra_Leone	NCBI
KP759707	Ebola virus	10/11/2014	Sierra_Leone	NCBI
KM034551	Ebola virus	26/05/2014	Sierra_Leone	NCBI
KR006952	Ebola virus	22/11/2014	Liberia	NCBI
KR534556	Ebola virus	10/10/2014	Guinea	NCBI
KR105310	Ebola virus	08/09/2014	Sierra_Leone	NCBI
KU143810	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KY426707	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP759594	Ebola virus	31/10/2014	Sierra_Leone	NCBI
KY426717	Ebola virus	2015/--/--	Sierra_Leone	NCBI
HC874659	Reston virus	NA	NA	ViPR
NC_014373	Bundibugyo virus	2007/11/--	Uganda	ViPR
KT725327	Ebola virus	13/11/2014	Liberia	NCBI
KM233086	Ebola virus	15/06/2014	Sierra_Leone	NCBI
CDC-NIH-683	Ebola virus	06/09/2014	Liberia	Urbanowicz et al.
KR817222	Ebola virus	09/06/2014	Guinea	NCBI
KT725292	Ebola virus	18/08/2014	Liberia	NCBI
KM233057	Ebola virus	09/06/2014	Sierra_Leone	NCBI
KT013255	Ebola virus	19/03/2014	Guinea	NCBI
KT725272	Ebola virus	22/08/2014	Liberia	NCBI
KM034550	Ebola virus	25/05/2014	Sierra_Leone	NCBI
KU143805	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KY558985	Ebola virus	01/09/2014	Liberia	NCBI
KP728283	Ebola virus	21/11/2014	Switzerland	NCBI
KT725280	Ebola virus	22/11/2014	Liberia	NCBI
KX009892	Ebola virus	27/08/2014	Liberia	NCBI
KR817184	Ebola virus	31/03/2014	Guinea	NCBI

KR817144	Ebola virus	19/10/2014	Guinea	NCBI
KM233084	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KR817173	Ebola virus	15/12/2014	Guinea	NCBI
KR534517	Ebola virus	20/08/2014	Guinea	NCBI
KU143790	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR534587	Ebola virus	26/08/2014	Guinea	NCBI
KM233074	Ebola virus	14/06/2014	Sierra_Leone	NCBI
KR653265	Ebola virus	27/08/2014	Sierra_Leone	NCBI
KR817132	Ebola virus	30/08/2014	Guinea	NCBI
KU220277	Ebola virus	10/08/2014	Liberia	NCBI
KR653242	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KR534580	Ebola virus	25/09/2014	Guinea	NCBI
KR105315	Ebola virus	11/09/2014	Sierra_Leone	NCBI
KT357851	Ebola virus	06/02/2015	Sierra_Leone	NCBI
KR534519	Ebola virus	22/08/2014	Guinea	NCBI
KR534560	Ebola virus	14/10/2014	Guinea	NCBI
KP759710	Ebola virus	04/10/2014	Sierra_Leone	NCBI
KT357844	Ebola virus	28/01/2015	Sierra_Leone	NCBI
KY798005	Reston virus	1989/--/--	USA	NCBI
KR105269	Ebola virus	05/08/2014	Sierra_Leone	NCBI
KR817221	Ebola virus	09/06/2014	Guinea	NCBI
KP759698	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KU143801	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KP759765	Ebola virus	29/10/2014	Sierra_Leone	NCBI
KM233063	Ebola virus	12/06/2014	Sierra_Leone	NCBI
KR817114	Ebola virus	26/07/2014	Liberia	NCBI
KP759703	Ebola virus	08/11/2014	Sierra_Leone	NCBI
KR105264	Ebola virus	03/08/2014	Sierra_Leone	NCBI
KR824525	Ebola virus	2014/--/--	Sierra_Leone	NCBI
CON-10786	Ebola virus	17/08/2015	Guinea	Urbanowicz et al.
KR534536	Ebola virus	25/09/2014	Guinea	NCBI
KR817099	Ebola virus	04/01/2015	Guinea	NCBI
KR817124	Ebola virus	14/08/2014	Guinea	NCBI
KR534552	Ebola virus	08/10/2014	Guinea	NCBI
KP759599	Ebola virus	30/10/2014	Sierra_Leone	NCBI
KR105281	Ebola virus	13/08/2014	Sierra_Leone	NCBI
KR817186	Ebola virus	30/03/2014	Guinea	NCBI
KM233117	Ebola virus	09/06/2014	Sierra_Leone	NCBI
KY425630	Ebola virus	1976/--/--	Democratic_Republic_of_the_Congo	NCBI
KR105290	Ebola virus	15/08/2014	Sierra_Leone	NCBI
KR653256	Ebola virus	03/11/2014	Sierra_Leone	NCBI
KX009893	Ebola virus	27/08/2014	Liberia	NCBI
KM519951	Ebola virus	2014/--/--	Democratic_Republic_of_the_Congo	NCBI
KM233094	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KT725325	Ebola virus	20/09/2014	Liberia	NCBI
KY425645	Ebola virus	19/03/2014	Guinea	NCBI
KM233107	Ebola virus	17/06/2014	Sierra_Leone	NCBI
KY558987	Ebola virus	01/09/2014	Liberia	NCBI
KR105232	Ebola virus	14/07/2014	Sierra_Leone	NCBI
KU296305	Ebola virus	28/02/2015	Sierra_Leone	NCBI
CDC-NIH-272	Ebola virus	27/08/2014	Liberia	Urbanowicz et al.
KR534509	Ebola virus	24/07/2014	Guinea	NCBI

KT633510	Ebola virus	14/10/2014	Guinea	NCBI
EM_FORE_2015_896	Ebola virus	20/06/2015	Guinea	Urbanowicz et al.
KU143779	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KU143833	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KT725324	Ebola virus	19/07/2014	Liberia	NCBI
KR006943	Ebola virus	06/11/2014	Liberia	NCBI
KR653285	Ebola virus	24/10/2014	Sierra_Leone	NCBI
KR817079	Ebola virus	21/09/2014	Guinea	NCBI
KR105329	Ebola virus	21/09/2014	Sierra_Leone	NCBI
KY426693	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KR817113	Ebola virus	22/07/2014	Guinea	NCBI
KT725363	Ebola virus	17/08/2014	Liberia	NCBI
GUI_CTS_2015_0052	Ebola virus	25/06/2015	Guinea	Urbanowicz et al.
KT725330	Ebola virus	10/11/2014	Liberia	NCBI
KY366415	Ebola virus	2015/03/--	Sierra_Leone	NCBI
KP759674	Ebola virus	31/10/2014	Sierra_Leone	NCBI
KU220283	Ebola virus	30/12/2014	Liberia	NCBI
REDC_GUI_2015_02242	Ebola virus	24/10/2015	Guinea	Urbanowicz et al.
KY366420	Ebola virus	2015/03/--	Sierra_Leone	NCBI
KP759721	Ebola virus	06/10/2014	Sierra_Leone	NCBI
KU182898	Ebola virus	04/05/1995	Zaire	ViPR
KR817106	Ebola virus	31/01/2015	Guinea	NCBI
KR817242	Ebola virus	10/07/2014	Guinea	NCBI
KX009894	Ebola virus	27/08/2014	Liberia	NCBI
KT725305	Ebola virus	26/09/2014	Liberia	NCBI
KR653257	Ebola virus	05/11/2014	Sierra_Leone	NCBI
KM233075	Ebola virus	14/06/2014	Sierra_Leone	NCBI
KT357847	Ebola virus	30/01/2015	Sierra_Leone	NCBI
KT357835	Ebola virus	20/01/2015	Sierra_Leone	NCBI
EM_FORE_2015_2781	Ebola virus	13/10/2015	Guinea	Urbanowicz et al.
KR105295	Ebola virus	17/08/2014	Sierra_Leone	NCBI
KU143787	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR653303	Ebola virus	26/09/2014	Sierra_Leone	NCBI
KR817231	Ebola virus	20/06/2014	Liberia	NCBI
KM233071	Ebola virus	12/06/2014	Sierra_Leone	NCBI
KP759730	Ebola virus	10/10/2014	Sierra_Leone	NCBI
KP759640	Ebola virus	27/09/2014	Sierra_Leone	NCBI
KU143785	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KT725386	Ebola virus	15/08/2014	Liberia	NCBI
CDC-NIH-710	Ebola virus	07/09/2014	Liberia	Urbanowicz et al.
KX121419	Ebola virus	02/11/2014	Guinea	NCBI
KP759739	Ebola virus	13/10/2014	Sierra_Leone	NCBI
KY425647	Ebola virus	1976/--/--	Democratic_Republic_of_the_Congo	NCBI
KR653234	Ebola virus	12/10/2014	Sierra_Leone	NCBI
KR817244	Ebola virus	12/07/2014	Sierra_Leone	NCBI
KT357848	Ebola virus	03/02/2015	Sierra_Leone	NCBI
EM_COY_2015_014100	Ebola virus	26/03/2015	Guinea	Urbanowicz et al.
KM233113	Ebola virus	18/06/2014	Sierra_Leone	NCBI
KT725333	Ebola virus	17/11/2014	Liberia	NCBI

KT725279	Ebola virus	14/11/2014	Liberia	NCBI
KP759605	Ebola virus	04/11/2014	Sierra_Leone	NCBI
KR817136	Ebola virus	10/10/2014	Guinea	NCBI
KT725342	Ebola virus	26/11/2014	Liberia	NCBI
KR817162	Ebola virus	01/12/2014	Guinea	NCBI
KP759601	Ebola virus	02/11/2014	Sierra_Leone	NCBI
KP759662	Ebola virus	30/10/2014	Sierra_Leone	NCBI
KR817188	Ebola virus	27/03/2014	Guinea	NCBI
KR534532	Ebola virus	22/09/2014	Guinea	NCBI
KP759632	Ebola virus	05/10/2014	Sierra_Leone	NCBI
KU143793	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KP759752	Ebola virus	18/10/2014	Sierra_Leone	NCBI
KP759630	Ebola virus	28/09/2014	Sierra_Leone	NCBI
KR006954	Ebola virus	03/12/2014	Liberia	NCBI
KR105332	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KT725270	Ebola virus	01/10/2014	Liberia	NCBI
KM233097	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KP759668	Ebola virus	28/09/2014	Sierra_Leone	NCBI
KT725262	Ebola virus	25/06/2014	Liberia	NCBI
KR105201	Ebola virus	19/06/2014	Sierra_Leone	NCBI
KU220274	Ebola virus	07/07/2015	Liberia	NCBI
KT725375	Ebola virus	11/10/2014	Liberia	NCBI
KU143827	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KU143812	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR105270	Ebola virus	05/08/2014	Sierra_Leone	NCBI
KR105300	Ebola virus	24/08/2014	Sierra_Leone	NCBI
KR534565	Ebola virus	18/10/2014	Guinea	NCBI
KU296319	Ebola virus	16/07/2015	Sierra_Leone	NCBI
KT013256	Ebola virus	20/03/2014	Guinea	NCBI
KY558988	Ebola virus	01/08/2014	Liberia	NCBI
KR653298	Ebola virus	28/10/2014	Sierra_Leone	NCBI
EM_COY_2015_022059	Ebola virus	02/07/2015	Guinea	Urbanowicz et al.
KR025228	Ebola virus	12/03/2015	United_Kingdom	NCBI
KR817218	Ebola virus	01/06/2014	Guinea	NCBI
EM-FORE-2015-631	Ebola virus	02/06/2015	Guinea	Urbanowicz et al.
KR105224	Ebola virus	11/07/2014	Sierra_Leone	NCBI
KM233082	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KU220279	Ebola virus	24/08/2014	Liberia	NCBI
KR534534	Ebola virus	24/09/2014	Guinea	NCBI
KR105252	Ebola virus	19/07/2014	Sierra_Leone	NCBI
KR006958	Ebola virus	20/12/2014	Liberia	Urbanowicz et al.
KP240933	Ebola virus	06/10/2014	Liberia	NCBI
KR817192	Ebola virus	03/04/2014	Guinea	NCBI
KY425649	Ebola virus	1976/--/--	Democratic_Republic_of_the_Congo	NCBI
KR534526	Ebola virus	15/09/2014	Guinea	NCBI
KT725388	Ebola virus	04/07/2014	Liberia	NCBI
KP759748	Ebola virus	16/10/2014	Sierra_Leone	NCBI
KC242788	Ebola virus	2007/--/--	Democratic_Republic_of_the_Congo	NCBI
KM233068	Ebola virus	16/06/2014	Sierra_Leone	NCBI
KT725354	Ebola virus	26/09/2014	Liberia	NCBI
KU143820	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KY426689	Ebola virus	2015/--/--	Sierra_Leone	NCBI

KR534514	Ebola virus	15/08/2014	Guinea	NCBI
KP759604	Ebola virus	04/11/2014	Sierra_Leone	NCBI
KR817178	Ebola virus	30/07/2014	Guinea	NCBI
KT725283	Ebola virus	26/08/2014	Liberia	NCBI
KR075002	Ebola virus	2014/--/--	Liberia	NCBI
KP759627	Ebola virus	11/11/2014	Sierra_Leone	NCBI
EM_FORE_2015_781	Ebola virus	13/06/2015	Guinea	Urbanowicz et al.
KU143784	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KP759676	Ebola virus	30/10/2014	Sierra_Leone	NCBI
KR105304	Ebola virus	28/08/2014	Sierra_Leone	NCBI
KP759687	Ebola virus	30/10/2014	Sierra_Leone	NCBI
KR006963	Ebola virus	01/10/2014	Liberia	NCBI
KR534582	Ebola virus	27/09/2014	Guinea	NCBI
KY426728	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KY426704	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KM233072	Ebola virus	14/06/2014	Sierra_Leone	NCBI
KU143808	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR653304	Ebola virus	01/11/2014	Sierra_Leone	NCBI
KP759642	Ebola virus	29/09/2014	Sierra_Leone	NCBI
KR817068	Ebola virus	01/09/2014	Guinea	NCBI
KR105206	Ebola virus	21/06/2014	Sierra_Leone	NCBI
KT725390	Ebola virus	19/08/2014	Liberia	NCBI
KX013091	Ebola virus	05/08/2014	Nigeria	NCBI
KU143778	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR817140	Ebola virus	17/10/2014	Guinea	NCBI
KR105338	Ebola virus	25/09/2014	Sierra_Leone	NCBI
LC152433	Ebola virus	2015/02/--	Guinea	NCBI
KP759738	Ebola virus	12/10/2014	Sierra_Leone	NCBI
KR653237	Ebola virus	24/11/2014	Sierra_Leone	NCBI
KR105241	Ebola virus	09/07/2014	Sierra_Leone	NCBI
KU143813	Ebola virus	2014/--/--	Sierra_Leone	NCBI
CON-11010	Ebola virus	23/08/2015	Guinea	Urbanowicz et al.
KT725263	Ebola virus	09/09/2014	Liberia	NCBI
KP759686	Ebola virus	01/11/2014	Sierra_Leone	NCBI
KM034563	Ebola virus	28/05/2014	Sierra_Leone	NCBI
KU143834	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR105297	Ebola virus	17/08/2014	Sierra_Leone	NCBI
KR105238	Ebola virus	14/07/2014	Sierra_Leone	NCBI
KR817219	Ebola virus	05/06/2014	Guinea	NCBI
KR105342	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KR817157	Ebola virus	13/11/2014	Guinea	NCBI
KU296549	Ebola virus	04/02/2015	Sierra_Leone	NCBI
KR817158	Ebola virus	13/11/2014	Guinea	NCBI
KC242789	Ebola virus	2007/--/--	Democratic_Republic_of_the_Congo	NCBI
KR105200	Ebola virus	16/06/2014	Sierra_Leone	NCBI
KR105265	Ebola virus	03/08/2014	Sierra_Leone	NCBI
KR817116	Ebola virus	01/08/2014	Guinea	NCBI
KP759596	Ebola virus	29/10/2014	Sierra_Leone	NCBI
KT357826	Ebola virus	13/01/2015	Sierra_Leone	NCBI
KR105247	Ebola virus	18/07/2014	Sierra_Leone	NCBI
KT725353	Ebola virus	20/09/2014	Liberia	NCBI
KR534521	Ebola virus	28/08/2014	Guinea	NCBI
KT725343	Ebola virus	01/09/2014	Liberia	NCBI

KU220271	Ebola virus	01/07/2015	Liberia	NCBI
KR653291	Ebola virus	22/09/2014	Sierra_Leone	NCBI
KT357855	Ebola virus	23/02/2015	Sierra_Leone	NCBI
KY426714	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KT725369	Ebola virus	26/08/2014	Liberia	NCBI
KR817239	Ebola virus	29/06/2014	Liberia	NCBI
KR534571	Ebola virus	24/10/2014	Guinea	NCBI
KP759654	Ebola virus	27/10/2014	Sierra_Leone	NCBI
KR817150	Ebola virus	29/10/2014	Guinea	NCBI
KY425639	Ebola virus	1976/--/--	Democratic_Republic_of_the_Congo	NCBI
KX000400	Ebola virus	19/03/2014	Guinea	NCBI
KT357840	Ebola virus	21/01/2015	Sierra_Leone	NCBI
KX013099	Ebola virus	04/08/2014	Nigeria	NCBI
KX013094	Ebola virus	06/08/2014	Nigeria	NCBI
KY426712	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KT725260	Ebola virus	07/09/2014	Liberia	NCBI
KR105274	Ebola virus	10/08/2014	Sierra_Leone	NCBI
KP759750	Ebola virus	16/10/2014	Sierra_Leone	NCBI
KR817071	Ebola virus	04/09/2014	Guinea	NCBI
KT725341	Ebola virus	26/08/2014	Liberia	NCBI
KR105221	Ebola virus	10/07/2014	Sierra_Leone	NCBI
KR075003	Ebola virus	2014/--/--	Liberia	NCBI
KP759629	Ebola virus	04/10/2014	Sierra_Leone	NCBI
KR105277	Ebola virus	12/08/2014	Sierra_Leone	NCBI
KP759626	Ebola virus	11/11/2014	Sierra_Leone	NCBI
CON-10512	Ebola virus	11/08/2015	Guinea	Urbanowicz et al.
KT589389	Ebola virus	07/09/2014	Sierra_Leone	NCBI
KR653249	Ebola virus	16/10/2014	Sierra_Leone	NCBI
KM034555	Ebola virus	06/06/2014	Sierra_Leone	NCBI
KR817133	Ebola virus	30/08/2014	Guinea	NCBI
KR653236	Ebola virus	05/11/2014	Sierra_Leone	NCBI
KY366417	Ebola virus	2015/03/--	Sierra_Leone	NCBI
KU143818	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KU296815	Ebola virus	11/01/2015	Sierra_Leone	NCBI
KY366413	Ebola virus	2015/03/--	Sierra_Leone	NCBI
KR105302	Ebola virus	25/08/2014	Sierra_Leone	NCBI
EM_COY_2015_013671	Ebola virus	12/03/2015	Guinea	Urbanowicz et al.
KR653235	Ebola virus	04/09/2014	Sierra_Leone	NCBI
KT725269	Ebola virus	02/10/2014	Liberia	NCBI
KP759727	Ebola virus	10/10/2014	Sierra_Leone	NCBI
KP184503	Ebola virus	25/08/2014	United_Kingdom	NCBI
KX009898	Ebola virus	06/09/2014	Liberia	NCBI
KR817205	Ebola virus	07/05/2014	Guinea	NCBI
KR817083	Ebola virus	29/09/2014	Guinea	NCBI
KU296766	Ebola virus	15/03/2015	Sierra_Leone	NCBI
KP759665	Ebola virus	31/10/2014	Sierra_Leone	NCBI
KR105219	Ebola virus	09/07/2014	Sierra_Leone	NCBI
KR817193	Ebola virus	02/04/2014	Guinea	NCBI
KM233087	Ebola virus	15/06/2014	Sierra_Leone	NCBI
CDC-NIH-3871	Ebola virus	25/11/2014	Liberia	Urbanowicz et al.
KR817069	Ebola virus	01/09/2014	Guinea	NCBI

EM_FORE_2015_1023	Ebola virus	29/06/2015	Guinea	Urbanowicz et al.
KR534554	Ebola virus	09/10/2014	Guinea	NCBI
KT725326	Ebola virus	26/08/2014	Liberia	NCBI
KR534586	Ebola virus	24/07/2014	Guinea	NCBI
KM233064	Ebola virus	14/06/2014	Sierra_Leone	NCBI
KU143822	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KT725303	Ebola virus	06/11/2014	Liberia	NCBI
EM_COY_2015_017057	Ebola virus	27/05/2015	Guinea	Urbanowicz et al.
KP759651	Ebola virus	28/09/2014	Sierra_Leone	NCBI
KY426729	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KM233093	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KT725273	Ebola virus	10/11/2014	Liberia	NCBI
1100	Ebola virus	23/09/2014	Guinea	Urbanowicz et al.
KP260802	Ebola virus	12/11/2014	Mali	NCBI
KR817245	Ebola virus	12/07/2014	Liberia	NCBI
AF086833	Ebola virus	1976/--/--	Democratic_Republic_of_the_Congo	NCBI
KP759762	Ebola virus	28/10/2014	Sierra_Leone	NCBI
KR534557	Ebola virus	11/10/2014	Guinea	NCBI
KR105216	Ebola virus	07/07/2014	Sierra_Leone	NCBI
KU143828	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR063673	Bundibugyo virus	01/10/2007	Uganda	NCBI
REDC_GUI_2015_00545	Ebola virus	15/07/2015	Guinea	Urbanowicz et al.
KU220273	Ebola virus	07/07/2015	Liberia	NCBI
KC242784	Ebola virus	2007/--/--	Democratic_Republic_of_the_Congo	NCBI
KJ660347	Ebola virus	2014/--/--	Guinea	NCBI
KC242785	Ebola virus	2007/--/--	Democratic_Republic_of_the_Congo	NCBI
KR817138	Ebola virus	10/10/2014	Guinea	NCBI
KU296547	Ebola virus	18/03/2015	Sierra_Leone	NCBI
KT725344	Ebola virus	16/08/2014	Liberia	NCBI
KR105220	Ebola virus	11/07/2014	Sierra_Leone	NCBI
KY425657	Ebola virus	19/03/2014	Guinea	NCBI
KT725359	Ebola virus	23/09/2014	Liberia	NCBI
KY798012	Reston virus	2009/--/--	Philippines	NCBI
KX013097	Ebola virus	05/08/2014	Nigeria	NCBI
KU143786	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KU143780	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KT725382	Ebola virus	24/08/2014	Liberia	NCBI
KR817153	Ebola virus	03/11/2014	Guinea	NCBI
KU182900	Ebola virus	04/05/1995	Zaire	ViPR
KM034559	Ebola virus	28/05/2014	Sierra_Leone	NCBI
KR105339	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KP759617	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KR105207	Ebola virus	21/06/2014	Sierra_Leone	NCBI
KP759713	Ebola virus	05/10/2014	Sierra_Leone	NCBI
KM034549	Ebola virus	25/05/2014	Sierra_Leone	NCBI
KP759714	Ebola virus	05/10/2014	Sierra_Leone	NCBI
KC242790	Ebola virus	2007/--/--	Democratic_Republic_of_the_Congo	NCBI
KY426711	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP271018	Ebola virus	20/08/2014	Democratic_Republic_of_the_Congo	NCBI

KU296428	Ebola virus	18/06/2015	Sierra_Leone	NCBI
KR653248	Ebola virus	31/10/2014	Sierra_Leone	NCBI
KM233044	Ebola virus	04/06/2014	Sierra_Leone	NCBI
KT587343	Ebola virus	20/03/2015	Liberia	NCBI
KR817072	Ebola virus	07/09/2014	Guinea	NCBI
KT961624	Ebola virus	2015/05/--	Italy	NCBI
HQ613402	Ebola virus	31/12/2008	Democratic_Republic_of_the_Congo	NCBI
KU296370	Ebola virus	10/07/2015	Sierra_Leone	NCBI
KM233114	Ebola virus	20/06/2014	Sierra_Leone	NCBI
KP759717	Ebola virus	07/10/2014	Sierra_Leone	NCBI
KP759694	Ebola virus	03/10/2014	Sierra_Leone	NCBI
KP096422	Ebola virus	2014/03/--	Guinea	NCBI
KR105262	Ebola virus	01/08/2014	Sierra_Leone	NCBI
EM_COY_2015_014370	Ebola virus	07/04/2015	Guinea	Urbanowicz et al.
KT357836	Ebola virus	20/01/2015	Sierra_Leone	NCBI
KM233118	Ebola virus	12/06/2014	Sierra_Leone	NCBI
KR653227	Ebola virus	26/08/2014	Sierra_Leone	NCBI
KR105234	Ebola virus	14/07/2014	Sierra_Leone	NCBI
KT725319	Ebola virus	15/09/2014	Liberia	NCBI
KR006941	Ebola virus	05/11/2014	Liberia	Urbanowicz et al.
KR105325	Ebola virus	16/09/2014	Sierra_Leone	NCBI
KR105313	Ebola virus	11/09/2014	Sierra_Leone	NCBI
KP759637	Ebola virus	10/10/2014	Sierra_Leone	NCBI
KU220269	Ebola virus	29/06/2015	Liberia	NCBI
KM034553	Ebola virus	27/05/2014	Sierra_Leone	NCBI
KT357834	Ebola virus	20/01/2015	Sierra_Leone	NCBI
KR653287	Ebola virus	03/09/2014	Sierra_Leone	NCBI
KT725309	Ebola virus	26/08/2014	Liberia	NCBI
KP759635	Ebola virus	05/10/2014	Sierra_Leone	NCBI
KY426710	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KC242793	Ebola virus	1996/--/--	Gabon	NCBI
CDC-NIH-603	Ebola virus	04/09/2014	Liberia	Urbanowicz et al.
KR817107	Ebola virus	18/07/2014	Guinea	NCBI
KR653279	Ebola virus	02/09/2014	Sierra_Leone	NCBI
KR534583	Ebola virus	02/10/2014	Guinea	NCBI
KR817142	Ebola virus	18/10/2014	Guinea	NCBI
KU143775	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR817211	Ebola virus	18/05/2014	Guinea	NCBI
KP240934	Ebola virus	11/10/2014	USA	NCBI
KP759692	Ebola virus	03/10/2014	Sierra_Leone	NCBI
KR105256	Ebola virus	21/07/2014	Sierra_Leone	NCBI
KT357825	Ebola virus	13/01/2015	Sierra_Leone	NCBI
AY729654	Sudan virus	2000/--/--	Uganda	NCBI
KX121193	Ebola virus	12/01/2016	Sierra_Leone	NCBI
KT725318	Ebola virus	13/09/2014	Liberia	NCBI
KM233115	Ebola virus	18/06/2014	Sierra_Leone	NCBI
KR105209	Ebola virus	24/06/2014	Sierra_Leone	NCBI
KP759614	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KM233049	Ebola virus	31/05/2014	Sierra_Leone	NCBI
KR105253	Ebola virus	20/07/2014	Sierra_Leone	NCBI
KT345616	Ebola virus	19/02/2015	Sierra_Leone	NCBI
KR653229	Ebola virus	23/09/2014	Sierra_Leone	NCBI

KR653283	Ebola virus	09/10/2014	Sierra_Leone	NCBI
KR867676	Ebola virus	04/05/1995	Zaire	ViPR
KR817121	Ebola virus	04/08/2014	Guinea	NCBI
KC242799	Ebola virus	1995/--/--	Democratic_Republic_of_the_Congo	NCBI
KR817093	Ebola virus	24/12/2014	Guinea	NCBI
KP759678	Ebola virus	02/10/2014	Sierra_Leone	NCBI
KP759702	Ebola virus	08/11/2014	Sierra_Leone	NCBI
KR105345	Ebola virus	27/09/2014	Sierra_Leone	NCBI
KP759648	Ebola virus	20/10/2014	Sierra_Leone	NCBI
KP759644	Ebola virus	30/09/2014	Sierra_Leone	NCBI
KR653301	Ebola virus	23/10/2014	Sierra_Leone	NCBI
KP759680	Ebola virus	30/10/2014	Sierra_Leone	NCBI
KR817096	Ebola virus	02/01/2015	Guinea	NCBI
KP759690	Ebola virus	03/11/2014	Sierra_Leone	NCBI
KR653284	Ebola virus	26/09/2014	Sierra_Leone	NCBI
KC242800	Ebola virus	2002/--/--	Gabon	NCBI
KR105251	Ebola virus	19/07/2014	Sierra_Leone	NCBI
KR653262	Ebola virus	14/10/2014	Sierra_Leone	NCBI
KM233100	Ebola virus	16/06/2014	Sierra_Leone	NCBI
KU182908	Ebola virus	04/05/1995	Zaire	ViPR
KR653253	Ebola virus	03/11/2014	Sierra_Leone	NCBI
KR006951	Ebola virus	14/11/2014	Liberia	Urbanowicz et al.
KT725391	Ebola virus	11/10/2014	Liberia	NCBI
KR105311	Ebola virus	08/09/2014	Sierra_Leone	NCBI
KR074999	Ebola virus	2015/--/--	Liberia	NCBI
KP759723	Ebola virus	06/10/2014	Sierra_Leone	NCBI
KT725370	Ebola virus	13/11/2014	Liberia	NCBI
KT725366	Ebola virus	29/12/2014	Liberia	NCBI
KY425656	Ebola virus	1976/--/--	Democratic_Republic_of_the_Congo	NCBI
KR105320	Ebola virus	16/09/2014	Sierra_Leone	NCBI
KM233073	Ebola virus	14/06/2014	Sierra_Leone	NCBI
KX013092	Ebola virus	04/08/2014	Nigeria	NCBI
REDC_GUI_2015_01408	Ebola virus	01/09/2015	Guinea	Urbanowicz et al.
KM233090	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KR006960	Ebola virus	20/01/2015	Liberia	NCBI
KY366416	Ebola virus	2015/04/--	Sierra_Leone	NCBI
KR653254	Ebola virus	18/10/2014	Sierra_Leone	NCBI
KR653264	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KR105296	Ebola virus	18/08/2014	Sierra_Leone	NCBI
KR817226	Ebola virus	10/06/2014	Guinea	NCBI
KR817131	Ebola virus	29/08/2014	Guinea	NCBI
KR534577	Ebola virus	18/09/2014	Guinea	NCBI
KP096421	Ebola virus	2014/03/--	Guinea	NCBI
KC545389	Sudan virus	2012/--/--	Uganda	NCBI
KR105292	Ebola virus	15/08/2014	Sierra_Leone	NCBI
KT725298	Ebola virus	21/09/2014	Liberia	NCBI
KT357842	Ebola virus	25/01/2015	Sierra_Leone	NCBI
KP759653	Ebola virus	27/10/2014	Sierra_Leone	NCBI
KY426727	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP759612	Ebola virus	06/11/2014	Sierra_Leone	NCBI
KP759613	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KT725265	Ebola virus	09/09/2014	Liberia	NCBI
KR534578	Ebola virus	19/09/2014	Guinea	NCBI

CDC-NIH-261	Ebola virus	27/08/2014	Liberia	Urbanowicz et al.
KR074996	Ebola virus	2014/--/--	Liberia	NCBI
KR817129	Ebola virus	22/08/2014	Liberia	NCBI
KC589025	Sudan virus	2012/--/--	Uganda	NCBI
KT725310	Ebola virus	02/10/2014	Liberia	NCBI
KM233109	Ebola virus	18/06/2014	Sierra_Leone	NCBI
KR534572	Ebola virus	24/10/2014	Guinea	NCBI
KU143802	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KP759608	Ebola virus	30/09/2014	Sierra_Leone	NCBI
KP759751	Ebola virus	16/10/2014	Sierra_Leone	NCBI
KR074997	Ebola virus	2015/--/--	Liberia	NCBI
KP759667	Ebola virus	30/10/2014	Sierra_Leone	NCBI
KT725356	Ebola virus	27/08/2014	Liberia	NCBI
KT357813	Ebola virus	19/02/2015	Sierra_Leone	NCBI
KU978801	Ebola virus	26/03/2014	Guinea	ViPR
KT725286	Ebola virus	22/08/2014	Liberia	NCBI
KM233101	Ebola virus	16/06/2014	Sierra_Leone	NCBI
KR105258	Ebola virus	21/07/2014	Sierra_Leone	NCBI
KP759756	Ebola virus	29/09/2014	Sierra_Leone	NCBI
KR817191	Ebola virus	02/04/2014	Guinea	NCBI
KR817199	Ebola virus	12/04/2014	Guinea	NCBI
KP759761	Ebola virus	27/10/2014	Sierra_Leone	NCBI
KR653299	Ebola virus	15/11/2014	Sierra_Leone	NCBI
KU143824	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR105202	Ebola virus	18/06/2014	Sierra_Leone	NCBI
KR817122	Ebola virus	08/08/2014	Guinea	NCBI
KP759719	Ebola virus	05/10/2014	Sierra_Leone	NCBI
KP759722	Ebola virus	06/10/2014	Sierra_Leone	NCBI
KT725295	Ebola virus	20/09/2014	Liberia	NCBI
KR105287	Ebola virus	14/08/2014	Sierra_Leone	NCBI
KR105327	Ebola virus	16/09/2014	Sierra_Leone	NCBI
KY426685	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP178538	Ebola virus	03/08/2014	Liberia	NCBI
EM_FORE_2015-1971	Ebola virus	14/08/2015	Guinea	Urbanowicz et al.
KP759625	Ebola virus	10/11/2014	Sierra_Leone	NCBI
KR105308	Ebola virus	05/09/2014	Sierra_Leone	NCBI
KU978802	Ebola virus	01/10/1976	Russia	ViPR
KR817082	Ebola virus	29/09/2014	Guinea	NCBI
KM233081	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KR817195	Ebola virus	04/04/2014	Guinea	NCBI
KR534539	Ebola virus	02/10/2014	Guinea	NCBI
KM233051	Ebola virus	11/06/2014	Sierra_Leone	NCBI
KR817110	Ebola virus	22/07/2014	Liberia	NCBI
KR534563	Ebola virus	19/10/2014	Guinea	NCBI
KP759607	Ebola virus	03/10/2014	Sierra_Leone	NCBI
KG88	Ebola virus	19/06/2015	Guinea	Urbanowicz et al.
KR534528	Ebola virus	20/09/2014	Guinea	NCBI
KU296769	Ebola virus	19/04/2015	Sierra_Leone	NCBI
KR817160	Ebola virus	24/11/2014	Guinea	NCBI
KM233061	Ebola virus	10/06/2014	Sierra_Leone	NCBI
KR817081	Ebola virus	29/09/2014	Guinea	NCBI
KU296679	Ebola virus	12/02/2015	Sierra_Leone	NCBI

KR534507	Ebola virus	24/07/2014	Guinea	NCBI
KU143791	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KY426706	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP759633	Ebola virus	07/10/2014	Sierra_Leone	NCBI
KP759768	Ebola virus	29/10/2014	Sierra_Leone	NCBI
KC545393	Bundibugyo virus	2012/--/--	Democratic_Republic_of_the_Congo	NCBI
KU182906	Ebola virus	04/05/1995	Zaire	ViPR
KT357850	Ebola virus	06/02/2015	Sierra_Leone	NCBI
FJ621583	Reston virus	2008/--/--	Philippines	NCBI
KR817233	Ebola virus	20/06/2014	Liberia	NCBI
JN638998	Sudan virus	2011/05/--	Uganda	NCBI
KU220276	Ebola virus	12/07/2015	Liberia	NCBI
KP759611	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KR534588	Ebola virus	27/08/2014	Guinea	NCBI
KX009896	Ebola virus	04/09/2014	Liberia	NCBI
KR534581	Ebola virus	25/09/2014	Guinea	NCBI
KR653295	Ebola virus	25/11/2014	Sierra_Leone	NCBI
KY426692	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KM233116	Ebola virus	04/06/2014	Sierra_Leone	NCBI
KR534541	Ebola virus	04/10/2014	Guinea	NCBI
KP759726	Ebola virus	06/10/2014	Sierra_Leone	NCBI
KR006959	Ebola virus	22/12/2014	Liberia	Urbanowicz et al.
KR817168	Ebola virus	09/12/2014	Guinea	NCBI
KR817155	Ebola virus	08/11/2014	Guinea	NCBI
KR653286	Ebola virus	10/09/2014	Sierra_Leone	NCBI
KT725337	Ebola virus	20/08/2014	Liberia	NCBI
KY425648	Ebola virus	20/03/2014	Guinea	NCBI
KP759620	Ebola virus	03/10/2014	Sierra_Leone	NCBI
KR105279	Ebola virus	12/08/2014	Sierra_Leone	NCBI
KR534520	Ebola virus	26/08/2014	Guinea	NCBI
KR105273	Ebola virus	10/08/2014	Sierra_Leone	NCBI
KM233089	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KU143815	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR817206	Ebola virus	07/05/2014	Guinea	NCBI
KR817181	Ebola virus	31/03/2014	Guinea	NCBI
KR105275	Ebola virus	10/08/2014	Sierra_Leone	NCBI
KT725278	Ebola virus	16/08/2014	Liberia	NCBI
KT357824	Ebola virus	31/03/2015	Sierra_Leone	NCBI
KR105314	Ebola virus	11/09/2014	Sierra_Leone	NCBI
KP759715	Ebola virus	07/10/2014	Sierra_Leone	NCBI
KY798011	Reston virus	2008/--/--	Philippines	NCBI
KU143816	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR817151	Ebola virus	01/11/2014	Guinea	NCBI
KM233042	Ebola virus	03/06/2014	Sierra_Leone	NCBI
KT725328	Ebola virus	22/11/2014	Liberia	NCBI
KR105248	Ebola virus	18/07/2014	Sierra_Leone	NCBI
KX009904	Ebola virus	10/10/2014	Liberia	NCBI
KP759737	Ebola virus	09/10/2014	Sierra_Leone	NCBI
KX009901	Ebola virus	07/09/2014	Liberia	NCBI
KR534535	Ebola virus	25/09/2014	Guinea	NCBI
KP759658	Ebola virus	29/10/2014	Sierra_Leone	NCBI

KR653288	Ebola virus	28/09/2014	Sierra_Leone	NCBI
KT725345	Ebola virus	01/10/2014	Liberia	NCBI
KT725291	Ebola virus	10/11/2014	Liberia	NCBI
KT725285	Ebola virus	16/10/2014	Liberia	NCBI
KT357846	Ebola virus	29/01/2015	Sierra_Leone	NCBI
KU296835	Ebola virus	24/12/2014	Sierra_Leone	NCBI
KT725329	Ebola virus	19/09/2014	Liberia	NCBI
KR817163	Ebola virus	01/12/2014	Guinea	NCBI
KU143794	Ebola virus	2014/--/--	Sierra_Leone	NCBI
HC069233	Reston virus	NA	NA	ViPR
KR817141	Ebola virus	18/10/2014	Guinea	NCBI
KT357860	Ebola virus	30/06/2015	Sierra_Leone	NCBI
KU182905	Ebola virus	04/05/1995	Zaire	ViPR
KP759600	Ebola virus	02/11/2014	Sierra_Leone	NCBI
KT725377	Ebola virus	17/08/2014	Liberia	NCBI
KM233046	Ebola virus	06/06/2014	Sierra_Leone	NCBI
EM_COY_2015_017021	Ebola virus	25/05/2015	Guinea	Urbanowicz et al.
KR817067	Ebola virus	01/09/2014	Guinea	NCBI
HC874675	Reston virus	NA	NA	ViPR
NC_004161	Reston virus	NA	NA	ViPR
KT725371	Ebola virus	19/09/2014	Liberia	NCBI
KJ660348	Ebola virus	2014/--/--	Guinea	NCBI
KP759619	Ebola virus	05/11/2014	Sierra_Leone	NCBI
KU296663	Ebola virus	31/12/2014	Sierra_Leone	NCBI
KT725290	Ebola virus	28/11/2014	Liberia	NCBI
KR105268	Ebola virus	05/08/2014	Sierra_Leone	NCBI
KY426730	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KC242797	Ebola virus	1996/--/--	Gabon	NCBI
KR105285	Ebola virus	14/08/2014	Sierra_Leone	NCBI
KM233083	Ebola virus	20/06/2014	Sierra_Leone	NCBI
KP759652	Ebola virus	01/10/2014	Sierra_Leone	NCBI
KY425631	Sudan virus	2000/10/--	Uganda	NCBI
KR105294	Ebola virus	16/08/2014	Sierra_Leone	NCBI
KU143811	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR534525	Ebola virus	14/09/2014	Guinea	NCBI
KR817146	Ebola virus	23/10/2014	Guinea	NCBI
KT725288	Ebola virus	22/12/2014	Liberia	NCBI
KR817175	Ebola virus	16/12/2014	Guinea	NCBI
IPDPFHGINSP_GUI_2015_7070	Ebola virus	18/05/2015	Guinea	Urbanowicz et al.
CDC-NIH-257	Ebola virus	27/08/2014	Liberia	Urbanowicz et al.
EM-FORE-2015-1414	Ebola virus	13/07/2015	Guinea	Urbanowicz et al.
KP759659	Ebola virus	28/10/2014	Sierra_Leone	NCBI
KR817212	Ebola virus	21/05/2014	Guinea	NCBI
KU296846	Ebola virus	31/05/2015	Sierra_Leone	NCBI
KR817089	Ebola virus	05/10/2014	Guinea	NCBI
KR817208	Ebola virus	11/05/2014	Guinea	NCBI
KP759666	Ebola virus	02/10/2014	Sierra_Leone	NCBI
KT357818	Ebola virus	04/03/2015	Sierra_Leone	NCBI

KR105228	Ebola virus	11/07/2014	Sierra_Leone	NCBI
KU296707	Ebola virus	06/02/2015	Sierra_Leone	NCBI
KT725312	Ebola virus	18/10/2014	Liberia	NCBI
KP759740	Ebola virus	29/09/2014	Sierra_Leone	NCBI
KR105208	Ebola virus	22/06/2014	Sierra_Leone	NCBI
KR817104	Ebola virus	25/01/2015	Guinea	NCBI
KT725308	Ebola virus	20/09/2014	Liberia	NCBI
KC242794	Ebola virus	1996/--/--	Gabon	NCBI
KY426733	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KR653231	Ebola virus	13/11/2014	Sierra_Leone	NCBI
KR817086	Ebola virus	01/10/2014	Guinea	NCBI
1283	Ebola virus	04/10/2014	Guinea	Urbanowicz et al.
KR105266	Ebola virus	04/08/2014	Sierra_Leone	NCBI
JX477165	Reston virus	2009/--/--	Philippines	NCBI
KR534515	Ebola virus	16/08/2014	Guinea	NCBI
KR817145	Ebola virus	22/10/2014	Guinea	NCBI
KR534590	Ebola virus	07/10/2014	Guinea	NCBI
CDC-NIH-1122	Ebola virus	16/09/2014	Liberia	Urbanowicz et al.
KY426725	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KX000399	Ebola virus	10/03/2014	Guinea	NCBI
KU978803	Ebola virus	01/04/1995	Democratic_Republic_of_the_Congo	ViPR
KR105203	Ebola virus	17/06/2014	Sierra_Leone	NCBI
CON-10590	Ebola virus	13/08/2015	Guinea	Urbanowicz et al.
CDC-NIH-686	Ebola virus	06/09/2014	Liberia	Urbanowicz et al.
KM233067	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KR105278	Ebola virus	13/08/2014	Sierra_Leone	NCBI
KP759622	Ebola virus	09/11/2014	Sierra_Leone	NCBI
KP759742	Ebola virus	30/09/2014	Sierra_Leone	NCBI
KR817228	Ebola virus	14/06/2014	Sierra_Leone	NCBI
KP759699	Ebola virus	06/11/2014	Sierra_Leone	NCBI
KR817172	Ebola virus	14/12/2014	Guinea	NCBI
KR534522	Ebola virus	28/08/2014	Guinea	NCBI
KR105236	Ebola virus	13/07/2014	Sierra_Leone	NCBI
KC242795	Ebola virus	1996/--/--	Gabon	NCBI
KT725387	Ebola virus	20/08/2014	Liberia	NCBI
KP759729	Ebola virus	09/10/2014	Sierra_Leone	NCBI
KU143817	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR817220	Ebola virus	05/06/2014	Guinea	NCBI
KR653292	Ebola virus	04/11/2014	Sierra_Leone	NCBI
KR817225	Ebola virus	10/06/2014	Guinea	NCBI
KR105305	Ebola virus	28/08/2014	Sierra_Leone	NCBI
KR817227	Ebola virus	13/06/2014	Sierra_Leone	NCBI
KR653300	Ebola virus	24/09/2014	Sierra_Leone	NCBI
KM233055	Ebola virus	07/06/2014	Sierra_Leone	NCBI
KM233066	Ebola virus	13/06/2014	Sierra_Leone	NCBI
KT633509	Ebola virus	18/12/2014	Guinea	NCBI
KM233070	Ebola virus	14/06/2014	Sierra_Leone	NCBI
KU182903	Ebola virus	04/05/1995	Zaire	ViPR
KR105307	Ebola virus	04/09/2014	Sierra_Leone	NCBI
KR653297	Ebola virus	15/09/2014	Sierra_Leone	NCBI

KP759618	Ebola virus	30/09/2014	Sierra_Leone	NCBI
KR817111	Ebola virus	22/07/2014	Liberia	NCBI
KR105204	Ebola virus	19/06/2014	Sierra_Leone	NCBI
KT013259	Ebola virus	17/03/2014	Guinea	NCBI
KP260799	Ebola virus	23/10/2014	Mali	NCBI
KR817091	Ebola virus	19/12/2014	Guinea	NCBI
KM233077	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KP759763	Ebola virus	28/10/2014	Sierra_Leone	NCBI
KM233039	Ebola virus	03/06/2014	Sierra_Leone	NCBI
KU182910	Tai Forest virus	27/11/1994	Cote_d'Ivoire	ViPR
KR817084	Ebola virus	30/09/2014	Guinea	NCBI
KM233056	Ebola virus	07/06/2014	Sierra_Leone	NCBI
KR105276	Ebola virus	11/08/2014	Sierra_Leone	NCBI
KR817103	Ebola virus	22/01/2015	Guinea	NCBI
KR534564	Ebola virus	18/10/2014	Guinea	NCBI
KR534510	Ebola virus	27/07/2014	Guinea	NCBI
KR653241	Ebola virus	24/08/2014	Sierra_Leone	NCBI
KX121424	Ebola virus	18/04/2015	Guinea	NCBI
KR817147	Ebola virus	23/10/2014	Guinea	NCBI
KR817190	Ebola virus	02/04/2014	Guinea	NCBI
KR653224	Ebola virus	22/12/2014	Sierra_Leone	NCBI
CDC-NIH-3874	Ebola virus	25/11/2014	Liberia	Urbanowicz et al.
IPDPFHGINSP_GUI_2015_4786	Ebola virus	26/03/2015	Guinea	Urbanowicz et al.
KR534575	Ebola virus	15/10/2014	Guinea	NCBI
KP759615	Ebola virus	03/10/2014	Sierra_Leone	NCBI
KR653228	Ebola virus	01/11/2014	Sierra_Leone	NCBI
KT725381	Ebola virus	26/08/2014	Liberia	NCBI
KT725322	Ebola virus	18/07/2014	Liberia	NCBI
KP658432	Ebola virus	29/12/2014	United_Kingdom	NCBI
KX121194	Ebola virus	20/01/2016	Sierra_Leone	NCBI
KY008770	Reston virus	1989/--/--	USA	NCBI
KU182901	Ebola virus	04/05/1995	Zaire	ViPR
EM_FORE_2015-2417	Ebola virus	15/09/2015	Guinea	Urbanowicz et al.
EU338380	Sudan virus	2004/--/--	Sudan	NCBI
KR817185	Ebola virus	31/03/2014	Guinea	NCBI
KT725304	Ebola virus	26/09/2014	Liberia	NCBI
KP759685	Ebola virus	01/11/2014	Sierra_Leone	NCBI
AY769362	Reston virus	NA	NA	NCBI
KR653272	Ebola virus	26/10/2014	Sierra_Leone	NCBI
KX009900	Ebola virus	07/09/2014	Liberia	NCBI
KP759606	Ebola virus	03/10/2014	Sierra_Leone	NCBI
KT725352	Ebola virus	13/09/2014	Liberia	NCBI
AF522874	Reston virus	1989/--/--	USA	NCBI
KR534511	Ebola virus	02/08/2014	Guinea	NCBI
KP271020	Ebola virus	20/08/2014	Democratic_Republic_of_the_Congo	NCBI
KX009909	Ebola virus	23/11/2014	Liberia	NCBI
KM233092	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KY425633	Ebola virus	20/03/2014	Guinea	NCBI

KT357837	Ebola virus	20/01/2015	Sierra_Leone	NCBI
KR817075	Ebola virus	12/09/2014	Guinea	NCBI
KR817074	Ebola virus	09/09/2014	Guinea	NCBI
KP342330	Ebola virus	2014/10/--	Guinea	NCBI
KR817076	Ebola virus	13/09/2014	Guinea	NCBI
KR105348	Ebola virus	28/09/2014	Sierra_Leone	NCBI
KR105210	Ebola virus	24/06/2014	Sierra_Leone	NCBI
KR653252	Ebola virus	20/08/2014	Sierra_Leone	NCBI
KR653282	Ebola virus	29/10/2014	Sierra_Leone	NCBI
KY426731	Ebola virus	2015/--/--	Sierra_Leone	NCBI
HC874663	Reston virus	NA	NA	ViPR
KM233037	Ebola virus	03/06/2014	Sierra_Leone	NCBI
KR817167	Ebola virus	09/12/2014	Guinea	NCBI
KR817234	Ebola virus	20/06/2014	Liberia	NCBI
NC_014372	Tai Forest virus	1994/11/--	Cote_d'Ivoire	ViPR
KR534570	Ebola virus	24/10/2014	Guinea	NCBI
KP759688	Ebola virus	29/09/2014	Sierra_Leone	NCBI
KP759725	Ebola virus	08/10/2014	Sierra_Leone	NCBI
KM034558	Ebola virus	28/05/2014	Sierra_Leone	NCBI
KR105303	Ebola virus	25/08/2014	Sierra_Leone	NCBI
KP759757	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KP759675	Ebola virus	31/10/2014	Sierra_Leone	NCBI
KT725384	Ebola virus	04/02/2015	Liberia	NCBI
KM233078	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KR006944	Ebola virus	06/11/2014	Liberia	NCBI
KY426684	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KP759628	Ebola virus	26/09/2014	Sierra_Leone	NCBI
KU143782	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KT357841	Ebola virus	25/01/2015	Sierra_Leone	NCBI
KP759645	Ebola virus	17/10/2014	Sierra_Leone	NCBI
KT725339	Ebola virus	26/08/2014	Liberia	NCBI
KR653263	Ebola virus	05/09/2014	Sierra_Leone	NCBI
KR653271	Ebola virus	12/11/2014	Sierra_Leone	NCBI
KR653305	Ebola virus	21/09/2014	Sierra_Leone	NCBI
KM233048	Ebola virus	09/06/2014	Sierra_Leone	NCBI
KR534547	Ebola virus	07/10/2014	Guinea	NCBI
KR105213	Ebola virus	05/07/2014	Sierra_Leone	NCBI
KR653251	Ebola virus	22/08/2014	Sierra_Leone	NCBI
KT357858	Ebola virus	03/07/2015	Sierra_Leone	NCBI
KR817095	Ebola virus	27/12/2014	Guinea	NCBI
KP759643	Ebola virus	28/09/2014	Sierra_Leone	NCBI
KT725296	Ebola virus	13/09/2014	Liberia	NCBI
KP759736	Ebola virus	09/10/2014	Sierra_Leone	NCBI
KR817149	Ebola virus	27/10/2014	Guinea	NCBI
KP759602	Ebola virus	03/11/2014	Sierra_Leone	NCBI
KT357823	Ebola virus	28/03/2015	Sierra_Leone	NCBI
KR534546	Ebola virus	07/10/2014	Guinea	NCBI
EM-FORE-2015-664	Ebola virus	04/06/2015	Guinea	Urbanowicz et al.
KP759610	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KT765131	Ebola virus	2014/--/--	Guinea	NCBI
KY366421	Ebola virus	2015/05/--	Sierra_Leone	NCBI
KT725335	Ebola virus	18/08/2014	Liberia	NCBI

KP096420	Ebola virus	2014/03/--	Guinea	NCBI
KU143819	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR105227	Ebola virus	11/07/2014	Sierra_Leone	NCBI
KT725347	Ebola virus	26/08/2014	Liberia	NCBI
KY426703	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KM233091	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KU296528	Ebola virus	11/01/2015	Sierra_Leone	NCBI
KT725346	Ebola virus	26/08/2014	Liberia	NCBI
KR105301	Ebola virus	22/08/2014	Sierra_Leone	NCBI
KR817165	Ebola virus	03/12/2014	Guinea	NCBI
KY426696	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KY798006	Reston virus	1989/--/--	USA	NCBI
KR105337	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KR534549	Ebola virus	07/10/2014	Guinea	NCBI
KR006953	Ebola virus	25/11/2014	Liberia	NCBI
HC069213	Reston virus	NA	NA	ViPR
KR534558	Ebola virus	13/10/2014	Guinea	NCBI
KR817166	Ebola virus	04/12/2014	Guinea	NCBI
KU296839	Ebola virus	12/01/2015	Sierra_Leone	NCBI
KR105263	Ebola virus	03/08/2014	Sierra_Leone	NCBI
KR817179	Ebola virus	28/03/2014	Guinea	NCBI
KM233080	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KX013096	Ebola virus	15/08/2014	Nigeria	NCBI
KP260801	Ebola virus	21/11/2014	Mali	NCBI
KR817189	Ebola virus	27/03/2014	Guinea	NCBI
KT725302	Ebola virus	26/09/2014	Liberia	NCBI
KY426726	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KT725293	Ebola virus	19/08/2014	Liberia	NCBI
KR105333	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KR534550	Ebola virus	08/10/2014	Guinea	NCBI
KR817070	Ebola virus	04/09/2014	Guinea	NCBI
KP759656	Ebola virus	29/10/2014	Sierra_Leone	NCBI
KU220272	Ebola virus	02/07/2015	Liberia	NCBI
KT725332	Ebola virus	01/07/2014	Liberia	NCBI
KT013258	Ebola virus	17/03/2014	Guinea	NCBI
KR534530	Ebola virus	21/09/2014	Guinea	NCBI
FJ217161	Bundibugyo virus	2007/11/--	Uganda	NCBI
KM233098	Ebola virus	16/06/2014	Sierra_Leone	NCBI
KR817123	Ebola virus	08/08/2014	Liberia	NCBI
KX009903	Ebola virus	27/09/2014	Liberia	NCBI
KR105340	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KR105288	Ebola virus	14/08/2014	Sierra_Leone	NCBI
KR817130	Ebola virus	23/08/2014	Guinea	NCBI
KR006956	Ebola virus	10/12/2014	Liberia	NCBI
KX009906	Ebola virus	23/10/2014	Liberia	NCBI
KR534591	Ebola virus	07/10/2014	Guinea	NCBI
KT725357	Ebola virus	23/09/2014	Liberia	NCBI
KT725275	Ebola virus	11/07/2014	Liberia	NCBI
KR534512	Ebola virus	12/08/2014	Guinea	NCBI
KT725266	Ebola virus	15/09/2014	Liberia	NCBI
KX009908	Ebola virus	23/11/2014	Liberia	NCBI
KT357845	Ebola virus	28/01/2015	Sierra_Leone	NCBI

KT725358	Ebola virus	05/10/2014	Liberia	NCBI
KT725374	Ebola virus	05/10/2014	Liberia	NCBI
KR653267	Ebola virus	23/08/2014	Sierra_Leone	NCBI
KR653280	Ebola virus	16/09/2014	Sierra_Leone	NCBI
KR817203	Ebola virus	28/04/2014	Guinea	NCBI
KP759634	Ebola virus	08/10/2014	Sierra_Leone	NCBI
KP701371	Ebola virus	25/11/2014	Italy	NCBI
KR817112	Ebola virus	22/07/2014	Liberia	NCBI
KR105261	Ebola virus	26/07/2014	Sierra_Leone	NCBI
KR534551	Ebola virus	08/10/2014	Guinea	NCBI
KR817207	Ebola virus	10/05/2014	Guinea	NCBI
KT357856	Ebola virus	06/03/2015	Sierra_Leone	NCBI
KT725316	Ebola virus	21/08/2014	Liberia	NCBI
KR534531	Ebola virus	21/09/2014	Guinea	NCBI
KM034561	Ebola virus	28/05/2014	Sierra_Leone	NCBI
FJ968794	Sudan virus	1976/--/--	Sudan	NCBI
KR817156	Ebola virus	08/11/2014	Guinea	NCBI
KR105226	Ebola virus	11/07/2014	Sierra_Leone	NCBI
KP759684	Ebola virus	01/11/2014	Sierra_Leone	NCBI
KU182904	Ebola virus	04/05/1995	Zaire	ViPR
KY744597	Ebola virus	20/11/2015	Liberia	NCBI
EM_COY_2015_017666	Ebola virus	12/06/2015	Guinea	Urbanowicz et al.
KR534533	Ebola virus	24/09/2014	Guinea	NCBI
KP759697	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KR006950	Ebola virus	14/11/2014	Liberia	Urbanowicz et al.
KR534568	Ebola virus	24/10/2014	Guinea	NCBI
EM_COY_2015_016800	Ebola virus	20/05/2015	Guinea	Urbanowicz et al.
KT725373	Ebola virus	19/09/2014	Liberia	NCBI
KU220275	Ebola virus	08/07/2015	Liberia	NCBI
KR653274	Ebola virus	11/10/2014	Sierra_Leone	NCBI
KP759747	Ebola virus	26/09/2014	Sierra_Leone	NCBI
KT725393	Ebola virus	13/11/2014	Liberia	NCBI
KR817174	Ebola virus	14/12/2014	Guinea	NCBI
KR653296	Ebola virus	29/08/2014	Sierra_Leone	NCBI
KR817170	Ebola virus	11/12/2014	Guinea	NCBI
KU143799	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KP759766	Ebola virus	29/10/2014	Sierra_Leone	NCBI
KT357827	Ebola virus	14/01/2015	Sierra_Leone	NCBI
KP759745	Ebola virus	17/10/2014	Sierra_Leone	NCBI
KR105322	Ebola virus	15/09/2014	Sierra_Leone	NCBI
KR817243	Ebola virus	12/07/2014	Liberia	NCBI
KR817137	Ebola virus	10/10/2014	Guinea	NCBI
KR105240	Ebola virus	15/07/2014	Sierra_Leone	NCBI
KT013257	Ebola virus	20/03/2014	Guinea	NCBI
KR063672	Ebola virus	01/04/1995	Democratic_Republic_of_the_Congo	ViPR
KY426688	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KM034560	Ebola virus	28/05/2014	Sierra_Leone	NCBI
KY426721	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KM233095	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KT357820	Ebola virus	07/03/2015	Sierra_Leone	NCBI
KR105243	Ebola virus	15/07/2014	Sierra_Leone	NCBI
KR534544	Ebola virus	04/10/2014	Guinea	NCBI

CDC-NIH-259	Ebola virus	27/08/2014	Liberia	Urbanowicz et al.
KX009907	Ebola virus	19/11/2014	Liberia	NCBI
KT357831	Ebola virus	18/01/2015	Sierra_Leone	NCBI
KM233062	Ebola virus	11/06/2014	Sierra_Leone	NCBI
KR653247	Ebola virus	25/10/2014	Sierra_Leone	NCBI
KT725257	Ebola virus	16/08/2014	Liberia	NCBI
KR105230	Ebola virus	12/07/2014	Sierra_Leone	NCBI
KT725313	Ebola virus	31/10/2014	Liberia	NCBI
AB050936	Reston virus	1996/--/--	Philippines	NCBI
EM_COY_2015_016238	Ebola virus	03/05/2015	Guinea	Urbanowicz et al.
KT725321	Ebola virus	19/09/2014	Liberia	NCBI
KY366418	Ebola virus	2015/03/--	Sierra_Leone	NCBI
KR817088	Ebola virus	05/10/2014	Guinea	NCBI
KU143831	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KT357819	Ebola virus	28/02/2015	Sierra_Leone	NCBI
HC874657	Reston virus	NA	NA	ViPR
KR105231	Ebola virus	13/07/2014	Sierra_Leone	NCBI
EM_FORE_2015_695	Ebola virus	07/06/2015	Guinea	Urbanowicz et al.
KU143825	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KT357833	Ebola virus	20/01/2015	Sierra_Leone	NCBI
KT357838	Ebola virus	20/01/2015	Sierra_Leone	NCBI
KP759735	Ebola virus	09/10/2014	Sierra_Leone	NCBI
KM233104	Ebola virus	17/06/2014	Sierra_Leone	NCBI
KU143777	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR105223	Ebola virus	11/07/2014	Sierra_Leone	NCBI
IPDPFHGINSP_GUI_2015_5339	Ebola virus	08/04/2015	Guinea	Urbanowicz et al.
KP759758	Ebola virus	23/10/2014	Sierra_Leone	NCBI
KT725277	Ebola virus	14/07/2014	Liberia	NCBI
KT725320	Ebola virus	24/06/2014	Liberia	NCBI
KT357830	Ebola virus	17/01/2015	Sierra_Leone	NCBI
KR817196	Ebola virus	07/04/2014	Guinea	NCBI
KR817198	Ebola virus	11/04/2014	Guinea	NCBI
KY426709	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KR817118	Ebola virus	01/08/2014	Guinea	NCBI
KC242792	Ebola virus	1994/--/--	Gabon	NCBI
KR105215	Ebola virus	06/07/2014	Sierra_Leone	NCBI
KR653255	Ebola virus	19/10/2014	Sierra_Leone	NCBI
REDC_GUI_2015_00483	Ebola virus	12/07/2015	Guinea	Urbanowicz et al.
KT725306	Ebola virus	27/08/2014	Liberia	NCBI
KR534542	Ebola virus	04/10/2014	Guinea	NCBI
KR534561	Ebola virus	15/10/2014	Guinea	NCBI
KR817213	Ebola virus	21/05/2014	Guinea	NCBI
EM-FORE-2015-548	Ebola virus	27/05/2015	Guinea	Urbanowicz et al.
KR653278	Ebola virus	07/09/2014	Sierra_Leone	NCBI
KU182911	Bundibugyo virus	14/11/2007	Uganda	NCBI
KR653250	Ebola virus	21/09/2014	Sierra_Leone	NCBI

KR653273	Ebola virus	01/10/2014	Sierra_Leone	NCBI
KX013095	Ebola virus	01/09/2014	Nigeria	NCBI
KR653293	Ebola virus	10/10/2014	Sierra_Leone	NCBI
KY426720	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KT357853	Ebola virus	19/02/2015	Sierra_Leone	NCBI
FJ621584	Reston virus	2008/--/--	Philippines	NCBI
KC242791	Ebola virus	1977/--/--	Democratic_Republic_of_the_Congo	NCBI
KR817109	Ebola virus	22/07/2014	Liberia	NCBI
KU143826	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KC242783	Sudan virus	1979/--/--	Sudan	NCBI
KM233035	Ebola virus	02/06/2014	Sierra_Leone	NCBI
KR105317	Ebola virus	13/09/2014	Sierra_Leone	NCBI
KR817223	Ebola virus	09/06/2014	Guinea	NCBI
KP759624	Ebola virus	08/11/2014	Sierra_Leone	NCBI
KR105233	Ebola virus	14/07/2014	Sierra_Leone	NCBI
KU182902	Ebola virus	04/05/1995	Zaire	ViPR
KT725367	Ebola virus	20/08/2014	Liberia	NCBI
KR534540	Ebola virus	02/10/2014	Guinea	NCBI
KT725350	Ebola virus	17/08/2014	Liberia	NCBI
KP759663	Ebola virus	29/09/2014	Sierra_Leone	NCBI
KR817108	Ebola virus	22/07/2014	Liberia	NCBI
KY426716	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KM233069	Ebola virus	12/06/2014	Sierra_Leone	NCBI
KC545396	Bundibugyo virus	2012/--/--	Democratic_Republic_of_the_Congo	NCBI
KR534574	Ebola virus	25/10/2014	Guinea	NCBI
IPDPFHGINSP_GUI_2015_5117	Ebola virus	03/04/2015	Guinea	Urbanowicz et al.
KT725254	Ebola virus	10/11/2014	Liberia	NCBI
CON-8811	Ebola virus	27/06/2015	Guinea	Urbanowicz et al.
KT725282	Ebola virus	16/09/2014	Liberia	NCBI
KY425637	Ebola virus	1976/--/--	Democratic_Republic_of_the_Congo	NCBI
KP759598	Ebola virus	31/10/2014	Sierra_Leone	NCBI
HC069215	Reston virus	NA	NA	ViPR
KM233106	Ebola virus	17/06/2014	Sierra_Leone	NCBI
KP759716	Ebola virus	05/10/2014	Sierra_Leone	NCBI
KP759660	Ebola virus	31/10/2014	Sierra_Leone	NCBI
KM034556	Ebola virus	26/05/2014	Sierra_Leone	NCBI
KT357814	Ebola virus	21/02/2015	Sierra_Leone	NCBI
KU296764	Ebola virus	24/03/2015	Sierra_Leone	NCBI
KU143783	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR105334	Ebola virus	25/09/2014	Sierra_Leone	Urbanowicz et al.
KG87	Ebola virus	19/06/2015	Guinea	Urbanowicz et al.
KR817209	Ebola virus	11/05/2014	Guinea	NCBI
KR817085	Ebola virus	01/10/2014	Guinea	NCBI
KU143809	Ebola virus	2014/--/--	Sierra_Leone	NCBI
CDC-NIH-3827	Ebola virus	23/11/2014	Liberia	Urbanowicz et al.
KT587345	Ebola virus	24/09/2014	Liberia	NCBI
KY426687	Ebola virus	2015/--/--	Sierra_Leone	NCBI

KT357817	Ebola virus	27/02/2015	Sierra_Leone	NCBI
EM-FORE-2015-838	Ebola virus	16/06/2015	Guinea	Urbanowicz et al.
KY558986	Ebola virus	01/11/2014	Liberia	NCBI
KT357828	Ebola virus	14/01/2015	Sierra_Leone	NCBI
KX009895	Ebola virus	27/08/2014	Liberia	NCBI
KM233076	Ebola virus	14/06/2014	Sierra_Leone	NCBI
KR817135	Ebola virus	10/10/2014	Guinea	NCBI
KP759706	Ebola virus	09/11/2014	Sierra_Leone	NCBI
KP759728	Ebola virus	09/10/2014	Sierra_Leone	NCBI
KP759704	Ebola virus	08/11/2014	Sierra_Leone	NCBI
KY798008	Reston virus	1992/--/--	Philippines	NCBI
KY425652	Ebola virus	1976/--/--	Democratic_Republic_of_the_Congo	NCBI
KP759681	Ebola virus	31/10/2014	Sierra_Leone	NCBI
KY426686	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KU143830	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KR653275	Ebola virus	25/10/2014	Sierra_Leone	NCBI
EM_COY_2015_014102	Ebola virus	26/03/2015	Guinea	Urbanowicz et al.
KP759708	Ebola virus	11/11/2014	Sierra_Leone	NCBI
KM034557	Ebola virus	27/05/2014	Sierra_Leone	NCBI
KM233065	Ebola virus	12/06/2014	Sierra_Leone	NCBI
KP759661	Ebola virus	31/10/2014	Sierra_Leone	NCBI
CDC-NIH-595	Ebola virus	04/09/2014	Liberia	Urbanowicz et al.
KR105214	Ebola virus	04/07/2014	Sierra_Leone	NCBI
KR534523	Ebola virus	29/08/2014	Guinea	NCBI
KY426705	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KR817100	Ebola virus	11/01/2015	Guinea	NCBI
KT725301	Ebola virus	26/09/2014	Liberia	NCBI
KT589390	Ebola virus	17/09/2014	Sierra_Leone	NCBI
KY426699	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KX009912	Ebola virus	25/11/2014	Liberia	NCBI
KR105330	Ebola virus	22/09/2014	Sierra_Leone	NCBI
KU143781	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KT725392	Ebola virus	19/08/2014	Liberia	NCBI
KR653244	Ebola virus	21/09/2014	Sierra_Leone	NCBI
KR105289	Ebola virus	15/08/2014	Sierra_Leone	NCBI
KP759641	Ebola virus	29/09/2014	Sierra_Leone	NCBI
KR817134	Ebola virus	31/08/2014	Guinea	NCBI
KU143776	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KX009905	Ebola virus	10/10/2014	Liberia	NCBI
KM233079	Ebola virus	15/06/2014	Sierra_Leone	NCBI
KR063670	Sudan virus	01/10/2000	Uganda	ViPR
KY426723	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KR819004	Ebola virus	16/08/2014	Democratic_Republic_of_the_Congo	NCBI
KM233060	Ebola virus	14/06/2014	Sierra_Leone	NCBI
KT725284	Ebola virus	29/09/2014	Liberia	NCBI
KP759731	Ebola virus	09/10/2014	Sierra_Leone	NCBI
KR105306	Ebola virus	28/08/2014	Sierra_Leone	NCBI
KR075001	Ebola virus	2014/--/--	Liberia	NCBI
KT725289	Ebola virus	01/09/2014	Liberia	NCBI
KR653240	Ebola virus	06/11/2014	Sierra_Leone	NCBI

EM_COY_2015_017018	Ebola virus	26/05/2015	Guinea	Urbanowicz et al.
KT725364	Ebola virus	24/08/2014	Liberia	NCBI
KP759760	Ebola virus	24/10/2014	Sierra_Leone	NCBI
KT357852	Ebola virus	18/02/2015	Sierra_Leone	NCBI
KM233102	Ebola virus	16/06/2014	Sierra_Leone	NCBI
KR105246	Ebola virus	18/07/2014	Sierra_Leone	NCBI
KP759621	Ebola virus	07/11/2014	Sierra_Leone	NCBI
KP240932	Ebola virus	26/09/2014	Liberia	NCBI
KR105298	Ebola virus	19/08/2014	Sierra_Leone	NCBI
KR534508	Ebola virus	24/07/2014	Guinea	NCBI
KR534559	Ebola virus	14/10/2014	Guinea	NCBI
KT946869	Ebola virus	19/12/2014	Sierra_Leone	NCBI
KR817098	Ebola virus	04/01/2015	Guinea	NCBI
KU296617	Ebola virus	29/05/2015	Sierra_Leone	NCBI
KP759741	Ebola virus	26/09/2014	Sierra_Leone	NCBI
KR105341	Ebola virus	25/09/2014	Sierra_Leone	NCBI
KY426713	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KR653268	Ebola virus	24/10/2014	Sierra_Leone	NCBI
KR817117	Ebola virus	01/08/2014	Guinea	NCBI
KR653269	Ebola virus	18/09/2014	Sierra_Leone	NCBI
KR817139	Ebola virus	10/10/2014	Guinea	NCBI
KR105331	Ebola virus	22/09/2014	Sierra_Leone	NCBI
KU143823	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KP759595	Ebola virus	30/10/2014	Sierra_Leone	NCBI
KR653232	Ebola virus	23/09/2014	Sierra_Leone	NCBI
KM233088	Ebola virus	17/06/2014	Sierra_Leone	NCBI
KP759670	Ebola virus	02/10/2014	Sierra_Leone	NCBI
KR105284	Ebola virus	14/08/2014	Sierra_Leone	NCBI
EM-FORE-2015-669	Ebola virus	05/06/2015	Guinea	Urbanowicz et al.
KU143829	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KC545395	Bundibugyo virus	2012/--/--	Democratic_Republic_of_the_Congo	NCBI
EM_COY_2015_013857	Ebola virus	18/03/2015	Guinea	Urbanowicz et al.
KP759744	Ebola virus	17/10/2014	Sierra_Leone	NCBI
KR817216	Ebola virus	24/05/2014	Guinea	NCBI
KR817177	Ebola virus	18/12/2014	Guinea	NCBI
CDC-NIH-708	Ebola virus	07/09/2014	Liberia	Urbanowicz et al.
KT725287	Ebola virus	14/10/2014	Liberia	NCBI
KT725340	Ebola virus	23/09/2014	Liberia	NCBI
KM233047	Ebola virus	08/06/2014	Sierra_Leone	NCBI
KR817171	Ebola virus	14/12/2014	Guinea	NCBI
KM233040	Ebola virus	03/06/2014	Sierra_Leone	NCBI
KT725378	Ebola virus	16/08/2014	Liberia	NCBI
KU143821	Ebola virus	2014/--/--	Sierra_Leone	NCBI
KY426708	Ebola virus	2015/--/--	Sierra_Leone	NCBI
KM233103	Ebola virus	16/06/2014	Sierra_Leone	NCBI
KG12	Ebola virus	27/05/2015	Guinea	Urbanowicz et al.
KR653290	Ebola virus	26/11/2014	Sierra_Leone	NCBI
KP260800	Ebola virus	12/11/2014	Mali	NCBI

MF319185	Bombali virus	25/05/2016	Sierra_Leone	NCBI
MF319186	Bombali virus	21/05/2016	Sierra_Leone	NCBI

Sequences used

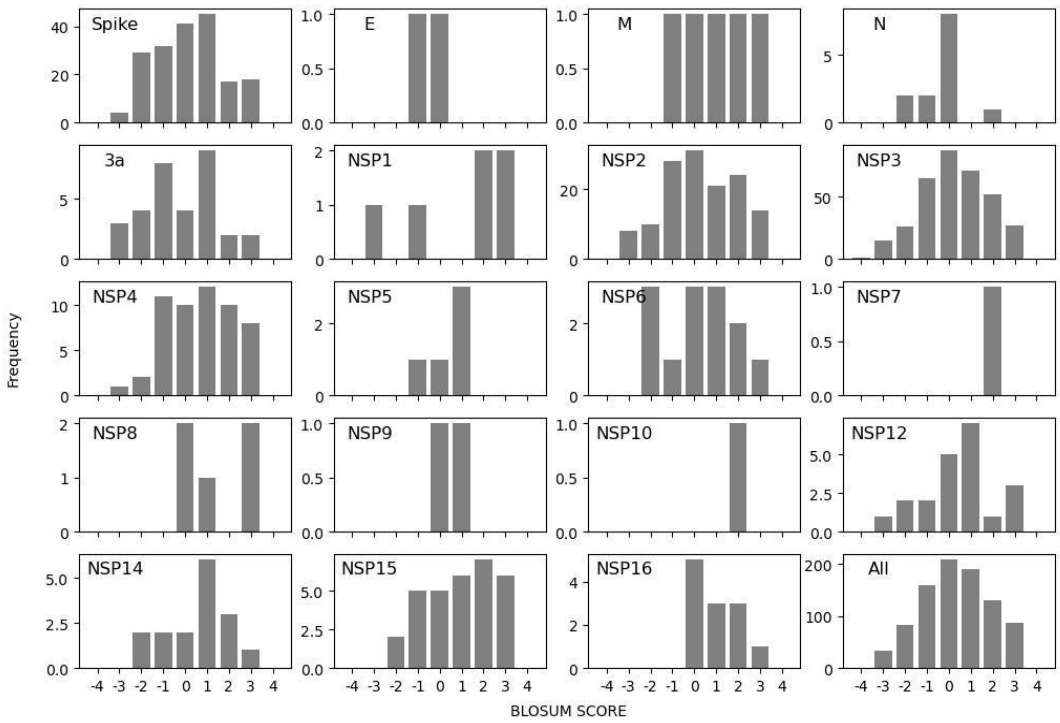
Appendix 3: Chapter 4 Supplementary Material

Differentially conserved amino acid positions may reflect differences in SARS-CoV-2 and SARS-CoV behaviour

Authors: Denis Bojkova[1#], Jake E. McGreig[2#], Katie-May McLaughlin[2#], Stuart G. Masterson[2], Marek Widera[1], Verena Krähling[3], Sandra Ciesek[1,4], Mark N. Wass[2*], Martin Michaelis[2*], Jindrich Cinatl Jr.[1*]

Affiliation: 1 Institute for Medical Virology, University Hospital, Goethe University Frankfurt am Main, Germany; 2 School of Biosciences, University of Kent, Canterbury, UK; 3 Institute of Virology, Biomedical Research Center (BMFZ), Philipps University Marburg, Germany; 4 German Center for Infection Research, DZIF, Braunschweig, Germany

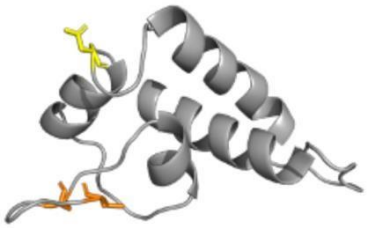
Correspondence: Jindrich Cinatl Jr. (cinatl@em.uni-frankfurt.de), Martin Michaelis (m.michaelis@kent.ac.uk), Mark N. Wass (m.n.wass@kent.ac.uk)



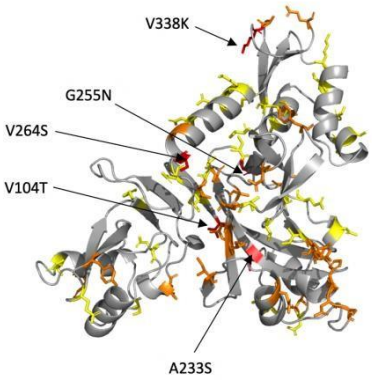
Supplementary Table 1. The BLOSUM scores for the amino acid substitutions present in the SDPs. A graph is plotted that combines all of the proteins and one for each of the individual proteins that were analysed.

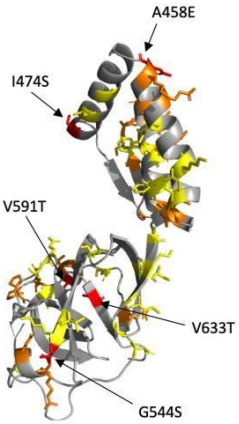
NSP1 (Phyre 2 model d2gdta1

NSP1

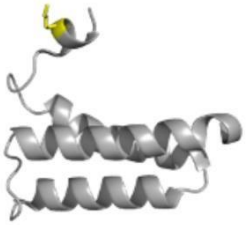


NSP4 (Phyre2 model c4gzfD)

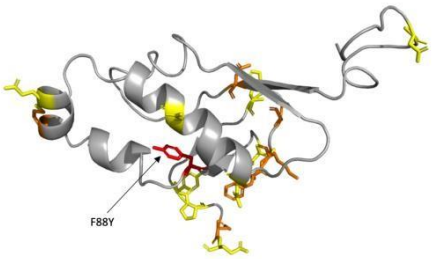




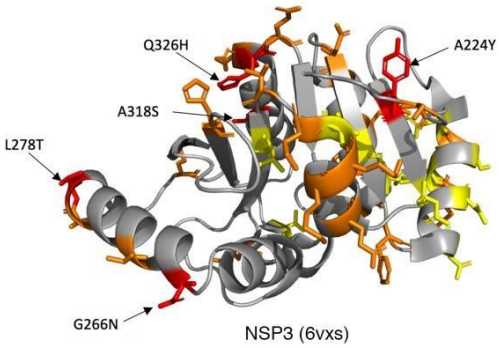
NSP2 (AlphaFold model)



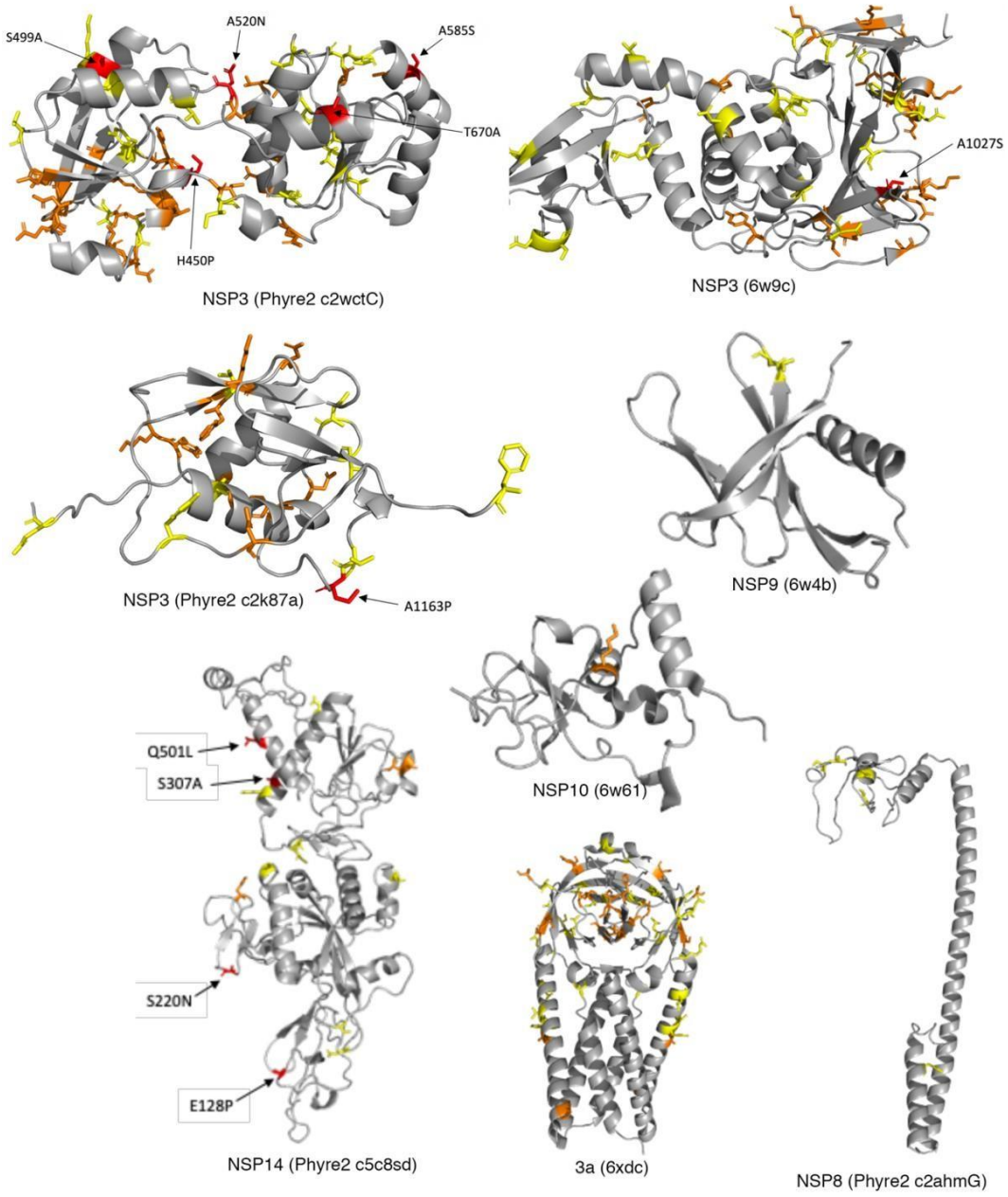
NSP7 (6xip)

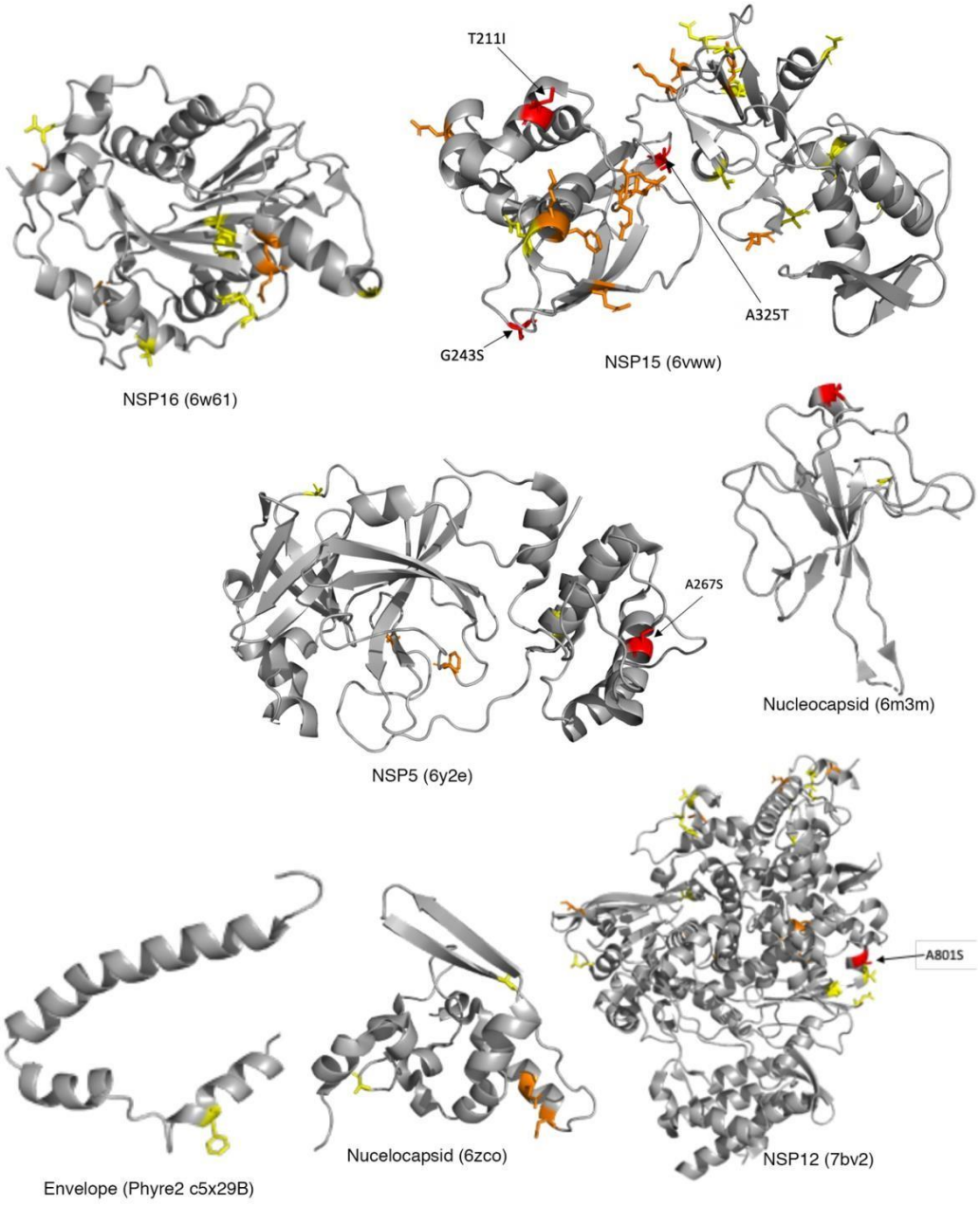


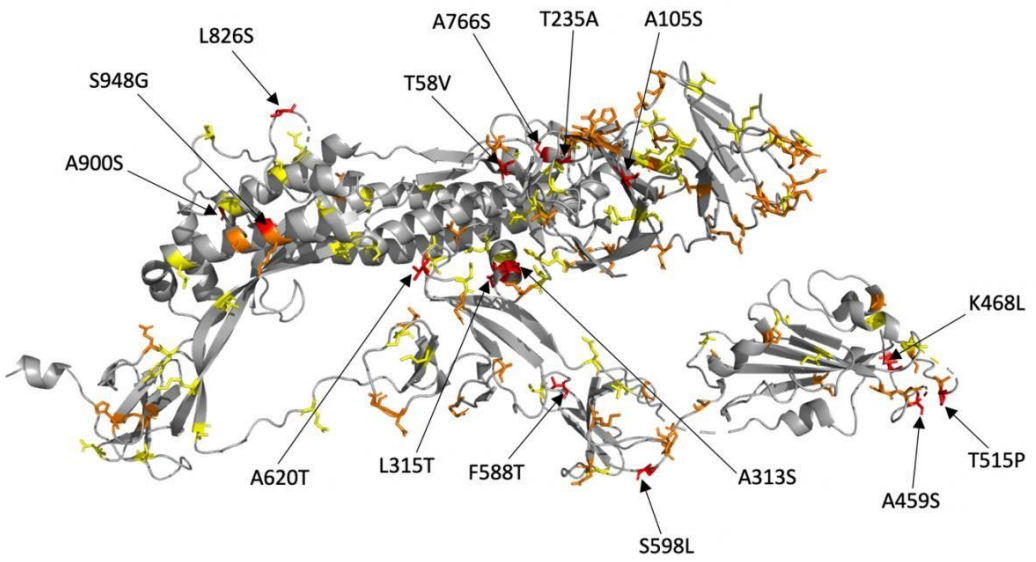
NSP3 (Phyre2 d2grla1)



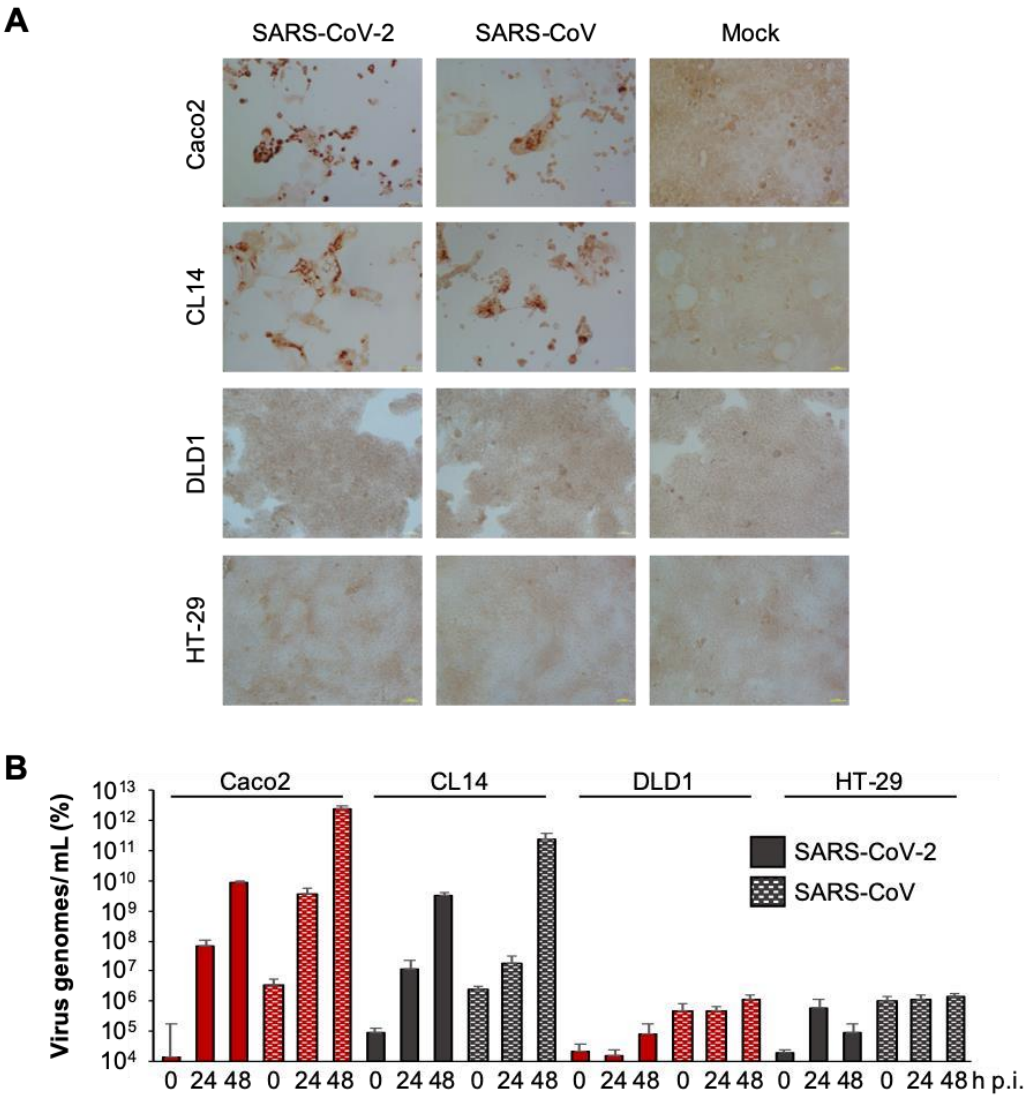
NSP3 (6vxs)


S499A
A520N
A585S
T670A
H450P
NSP3 (Phyre2 c2wctC)
A1027S
NSP3 (6w9c)
NSP3 (Phyre2 c2k87a)
A1163P
NSP9 (6w4b)
NSP10 (6w61)
Q501L
S307A
S220N
E128P
NSP14 (Phyre2 c5c8sd)
3a (6xdc)
NSP8 (Phyre2 c2ahmG)

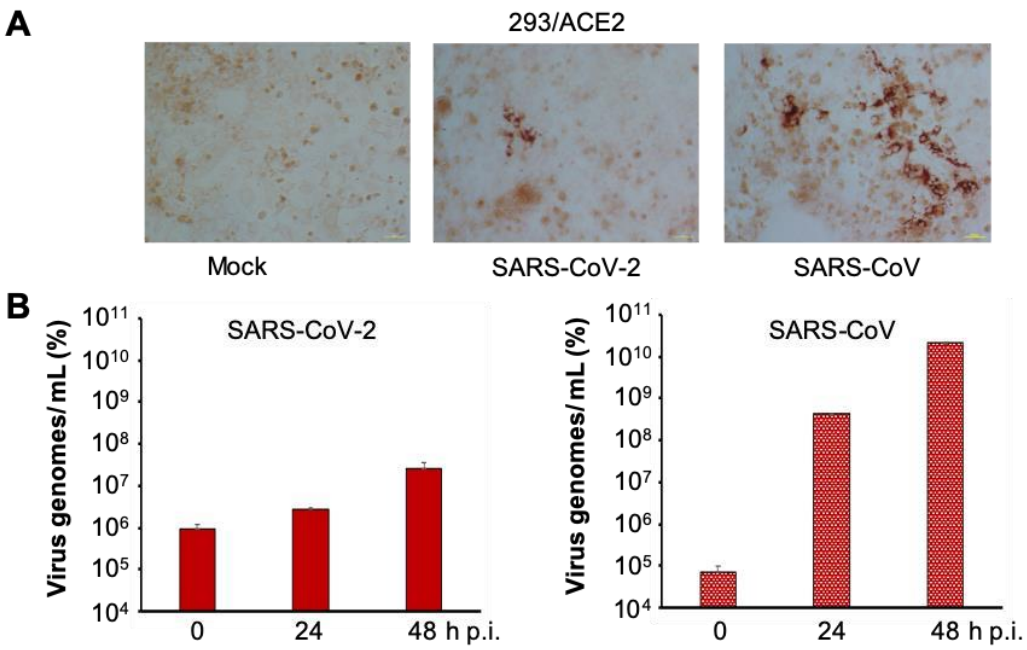

T211I
A325T
G243S
NSP16 (6w61)
NSP15 (6vww)
A267S
Nucleocapsid (6m3m)
NSP5 (6y2e)
A801S
Envelope (Phyre2 c5x29B)
Nucelocapsid (6zco)
NSP12 (7bv2)



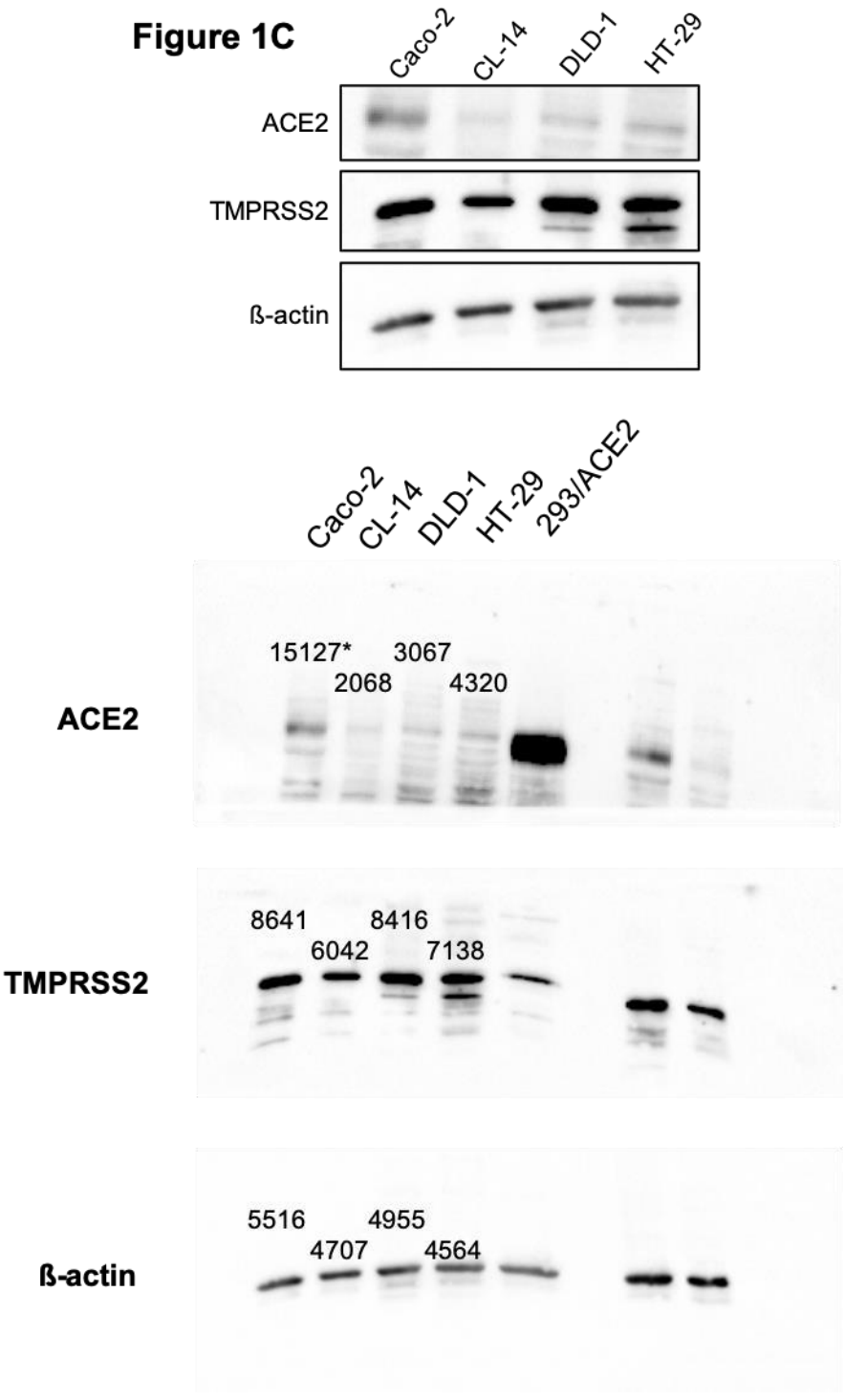
Supplementary Figure 2. Overview of modelled DCPs. DCPs with likely functional effect are indicated by arrows and labelled. Structural model shown is indicated in brackets. DCPs likely to have an effect are coloured red; DCPs with a possible effect are shown in orange; and DCPs unlikely to have an effect are coloured yellow. Please refer to table S6 for full details of structural analysis of each DCP.



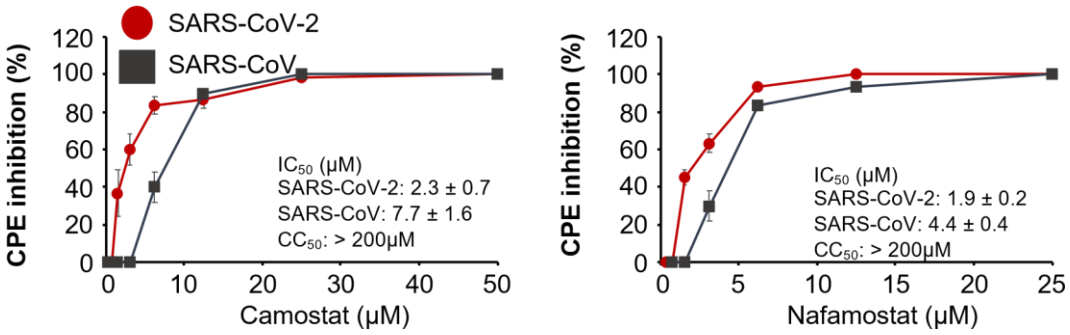
Supplementary Figure 3. SARS-CoV-2 and SARS-CoV susceptibility of cell lines. A) Representative images showing MOI 0.01-infected cells immunostained for double-stranded RNA 48h post infection. B) Quantification of virus genomes by qPCR at different time points post infection (p.i.). Values are presented as means ± S.D. (n =3).



Supplementary Figure 4. SARS-CoV-2 and SARS-CoV replication in 293 cells stably expressing ACE2 cells (293/ACE2). A) Immunostaining for double-stranded RNA (indicating virus replication) in SARS-CoV-2 and SARS-CoV (MOI 0.01)-infected 293/ACE2 cells 48h post infection. B) Quantification of virus genomes by qPCR in SARS-CoV-2 and SARS-CoV (MOI 0.01)-infected 293/ACE2 cells 48h post infection. Values are presented as means ± S.D. (n =3).



Supplementary Figure 5. Uncropped Western blots for Figure 2D. 293/ACE2 cells served as positive control for ACE2. * Protein quantification



Supplementary Figure 6. Role of TMPRSS2-mediated S cleavage in SARS-CoV-2 and SARS-Co-V replication. Concentration-dependent effects of the TMPRSS2 inhibitors camostat and nafamostat on SARS-CoV2- and SARS-CoV-induced cytopathogenic effect (CPE) formation determined 48h post infection in CL14 cells infected at an MOI of 0.01. Values are presented as means ± S.D. (n =3).

SARS	PDB identifier	Protein	Residues
SARS-CoV	2hsx	NSP1	13-127
SARS-CoV	2gri	NSP3	1-111
SARS-CoV	2fav	NSP3	185-354
SARS-CoV-2	6vxs	NSP3	207-373
SARS-CoV	2w2g	NSP3	389-652
SARS-CoV	2kaf	NSP3	655-720
SARS-CoV	5y3e	NSP3	723-1036
SARS-CoV-2	6w9c	NSP3	748-1061
SARS-CoV	2k87	NSP3	1066-1180
SARS-CoV	2h2z	NSP5	1-306
SARS-CoV-2	6y2e	NSP5	1-306
SARS-CoV	6nur	NSP7	2-71
SARS-CoV-2	6xip	NSP7	1-70
SARS-CoV	2ahm	NSP8	1-190
SARS-CoV	2fyg	NSP10	10-132
SARS-CoV-2	6w61	NSP10	18-132
SARS-CoV	6nur	NSP12	41-819
SARS-CoV-2	7bv2	NSP12	31-929
SARS-CoV	5c8s	NSP14	1-525
SARS-CoV	2h85	NSP15	1-345
SARS-CoV-2	6vww	NSP15	1-346
SARS-CoV	2xyq	NSP16	1-290
SARS-CoV-2	6w61	NSP16	1-299
SARS-CoV	6acg	S:ACE2	18-1119
SARS-CoV-2	6m17	S:ACE2	336-518
SARS-CoV	5xlr	S	33-1120
SARS-CoV	5wrg	S	261-1058
SARS-CoV-2	6vsb	S	27-1146
SARS-CoV-2	6xdc	3a	40-238
SARS-CoV	5x29	E	8-65
SARS-CoV	1yo4	7a	16-99
SARS-CoV	1ssk	N	49-185
SARS-CoV-2	6m3m	N	48-173
SARS-CoV	2gib	N	270-366
SARS-CoV-2	6zco	N	248-364

Supplementary Table 1. Protein structures used for structural analysis obtained from the Protein Databank.

SARS	Template structure	Protein	Residues	Coverage	Confidence	Identity (%)
SARS-CoV-2	2gdta1	NSP1	13-127	64	100	86
SARS-CoV	6zojj	NSP1	148-180	17	99.1	76
SARS-CoV-2	6zojj	NSP1	148-180	18	99.8	100
SARS-CoV-2	2gria1	NSP3	2-111	5	100	77
SARS-CoV-2	2acfa1	NSP3	207-373	8	100	74
SARS-CoV-2	2wctC	NSP3	425-676	12	100	76
SARS-CoV-2	2fe8B	NSP3	745-1058	16	100	82
SARS-CoV-2	2k87A	NSP3	1089-1203	5	100	82
SARS-CoV-2	3gzfD	NSP4	403-477	18	100	41
SARS-CoV-2	2duca1	NSP5	2-283	98	100	96
SARS-CoV-2	2ahmG	NSP8	1-175	95	100	97
SARS-CoV-2	1uw7A	NSP9	1-90	100	100	97
SARS-CoV-2	2g9tT	NSP10	9-116	86	100	98
SARS-CoV-2	6nusA	NSP12	118-909	87	100	97
SARS-CoV-2	5c8sD	NSP14	1-504	97	100	95
SARS-CoV	6xdcB	3a	40-238	72	100	77
SARS-CoV-2	5x29B	E	8-65	77	99.8	91
SARS-CoV-2	1yo4A	7a	16-98	67	100	91

Supplementary Table 2. Structural models generated by Phyre2 and used for structural analysis. Where structures were not available from the Protein Databank, the structures were modelled.

Effect	Reason
Unlikely	Conservative changes (between residues with the same polarity/charge) which do not affect ability to form hydrogen bonds with equivalent residues in SARS-CoV and SARS-CoV2
Possible – conformational change	Changes which could affect the ability of a sidechain of a residue in a given position to form hydrogen bonds with equivalent residues in SARS-CoV and SARS-CoV-2 (e.g. gain/loss of polarity, substitution for larger/smaller sidechain) but no such effects are visible, or conservative changes (between residues with the same polarity/charge) which appear in the model to result in gain/loss of hydrogen bonding between equivalent residues in SARS-CoV and SARS-CoV-2 (but mutagenesis suggests hydrogen bonding is possible with sidechain rotation)
Possible – alteration of sidechain/ligand interactions	Changes which result in gain of charge/alter the charge of a sidechain for a residue in a given position
Possible – conformational change and alteration of sidechain/ligand interactions	Changes which affect the ability of a sidechain of a residue in a given position to form hydrogen bonds with equivalent residues in SARS-CoV and SARS-CoV-2 (e.g. gain/loss of polarity, substitution for larger/smaller sidechain) but no such effects are visible, and changes which result in gain of charge/alter the charge of a sidechain for a residue in a given position
Likely – conformational change	Changes which result in visible alteration in the conformation of a protein at a given location (e.g. through loss of hydrogen bonding between equivalent residues in SARS-CoV and SARS-CoV-2) and/or which result in the loss of capacity for hydrogen bonding
Likely – conformational change (and possible alteration of sidechain/ligand interactions)	Changes which result in visible alteration in the conformation of a protein at a given location (e.g. through loss of hydrogen bonding between equivalent residues in SARS-CoV and SARS-CoV-2, and/or which result in the loss of capacity for hydrogen bonding and which result in gain of charge/alter the charge of a sidechain for a residue in a given position)

Supplementary Table 3. Criteria used for classifying proposed effect on protein structure and function within the structural analysis.

Protein (SARS-CoV)	Protein (SARS-CoV-2)	Sequences in Dataset	Protein Length (SARS-CoV)	DCPs Identified	% of Residues DCPs
S	S	73863	1255	186	14.82
3a	ORF3a	91214	274	32	11.68
3b		n/a	154		
E	E	94787	76	2	2.63
M	M	93860	221	15	2.26
6	6	94935	63	13	9.52
7a	7a	82940	122	0	0
7b	7b	n/a	44	NA	
8a/8b	8	n/a	39/84	NA	NA
9b		n/a	98	NA	
N	N	91609	422	13	3.08
	ORF10	n/a	n/a		
nsp1	nsp1	93621	180	6	3.33
nsp2	nsp2	88288	636	136	21.38
Nsp3	nsp3	75324	1922	344	17.90
nsp4	nsp4	89707	500	54	10.80
nsp5	nsp5	91731	306	5	1.63
nsp6	nsp6	93432	290	13	4.48
nsp7	nsp7	95038	83	1	1.20
nsp8	nsp8	94806	198	5	2.53
nsp9	nsp9	94970	113	2	1.77
nsp10	nsp10	92505	139	1	0.72
nsp12	nsp12	89874	932	21	2.25
nsp13	nsp13	91305	601	0	0
nsp14	nsp14	72306	527	16	3.04
nsp15	nsp15	85595	346	31	8.96
nsp16	nsp16	83565	298	12	4.03
Total				891	9.36

Supplementary Table 4. Specificity Determining Positions (DCPs) identified between SARS-CoV and SARS-CoV-2.

SDP	SARS-CoV structural analysis	SARS-CoV-2 structural analysis	Effect?
V404=K417	V404 is not in the interface	K417 is in the interface and could form a salt bridge with ACE-2 D30	Likely – new polar interaction within interface
R426=N439	Loss of hydrogen bond to ACE2 Gln325 due to shorter sidechain. N would still be able to form hydrogen bonds	N439 is located away from the interface site and so does not form a hydrogen bond with ACE2. Instead forms a hydrogen bond with S443 (also a DCP – A430=S443) which is likely to stabilise the loop they are both part of.	Likely – Loss of interface hydrogen bond.
Y442=L455	Y422 forms hydrogen bond to backbone of W476 – loss could result in conformational change. The sidechain also contacts the backbone of ACE2 D30 and K31	L455 remains in interface and contacts ACE2 D30 and H34.	Likely – loss of intramolecular hydrogen bond
F460=Y473	Conservative change.	Introduction of OH group that can form hydrogen bonds. Y473 forms hydrogen bond with backbone of R457 and is closer to ACE2 T27 so potential to form hydrogen bond in interface.	Possible – introduction of hydrogen bond (could be with ACE2)
P462=A475	Located in a loop, could affect this conformation – many DCPs in this loop	Loop has different conformation.	Possible – Conformational change of loop
N479=Q493	Interface hydrogen bond formed with ACE2 H34 backbone. With a shorter sidechain this this may be lost in SARS-CoV-2.	Q493 forms a hydrogen bond with ACE2 E35 in this complex. So hydrogen bond is maintained but also different.	Possible – hydrogen bond with ACE2 retained but to different residue.

Y484=Q498	Y484 can form hydrogen bonds with ACE2 Gln42 (sidechain) and intramolecular H bonds with T433 (backbone), Y436 (sidechain).	Q498 maintains hydrogen bonds with ACE2 Gln42	Possible – change in residue forming hydrogen bonds with ACE2.
T485=P499	Sidechain points away from interface, loss of hydrogen bond with R426 (also a DCP) backbone in adjacent loop. This hydrogen bond is likely to coordinate the structure between these two loops. There are multiple DCPs present in both loops	Loop conformation similar as for SARS-CoV structure but not coordinated with other loop	Likely - loss of intramolecular hydrogen bond
I489=V503	Conservative change I489 in direct contact with ACE2 Q325	slightly smaller sidechain is further away from ACE2 Q325.	Unlikely.

Supplementary Table 5. Analysis of DCPs present in the SARS-CoV and SARS-CoV-2 Spike protein interface with human ACE2.

Appendix 4: Chapter 5 Supplementary Material

Conserved RNA Structures in Ebolaviruses

Authors: Stuart G. Masterson[1], Ivo L. Hofacker[2,3], Martin Michaelis[1]*, Mark N. Wass[1]* and Michael T. Wolfinger[2,3]*

Affiliation: 1. Industrial Biotechnology Centre and School of Biosciences, University of Kent, Canterbury, Kent, UK. 2. Department of Theoretical Chemistry, University of Vienna, Wahringerstraße 17, 1090 Vienna, Austria 3. Research Group Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna,
Wahringerstraße 29, 1090 Vienna, Austria

Correspondence: To whom correspondence should be addressed: mark.n.wass@kent.ac.uk, martin.michaelis@kent.ac.uk, michael.wolfinger@univie.ac.at

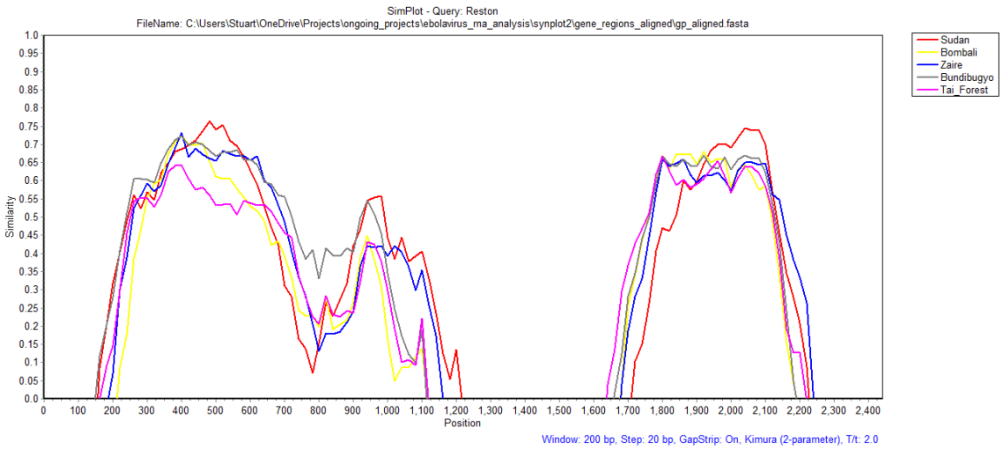


Figure s1. SimPlot graph for GP protein.

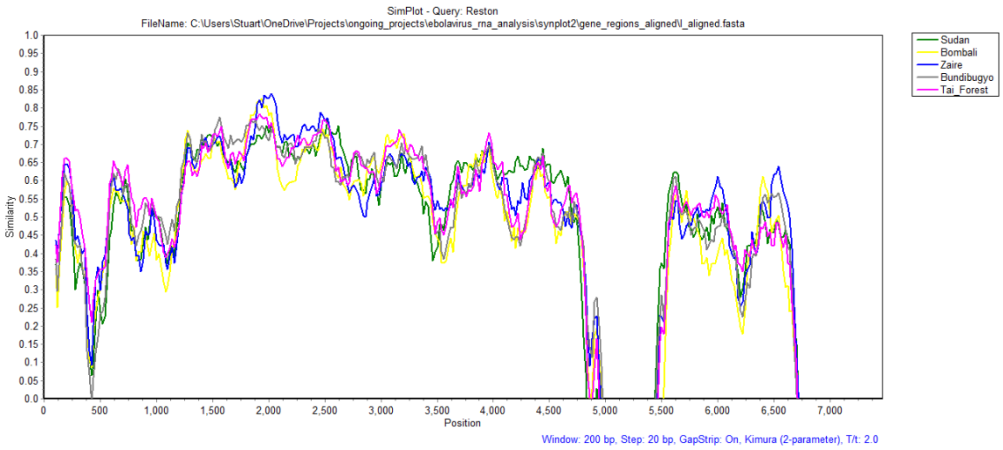


Figure s2. SimPlot graph for L protein.

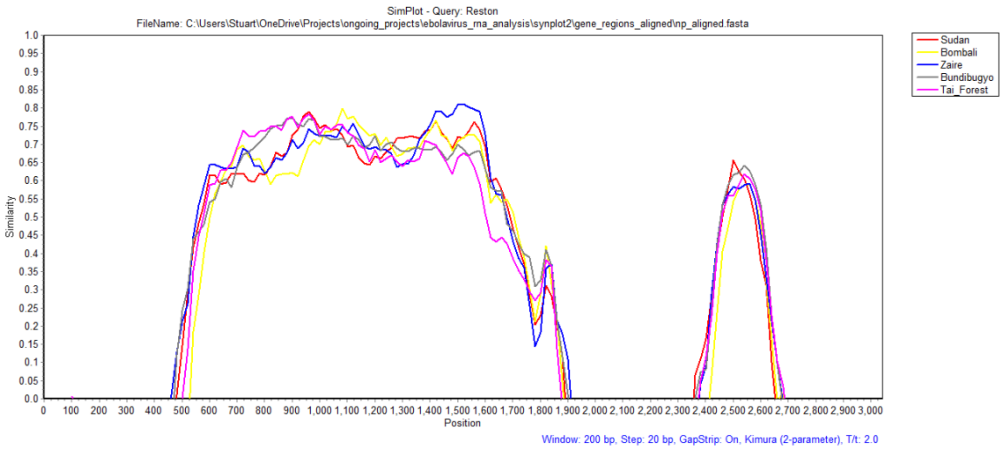


Figure s3. SimPlot graph for NP protein.

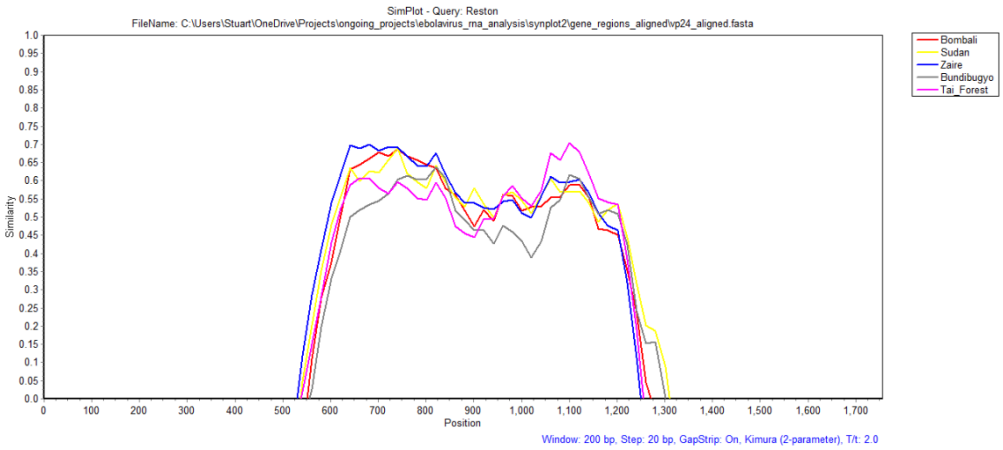


Figure s4. SimPlot graph for VP24 protein.

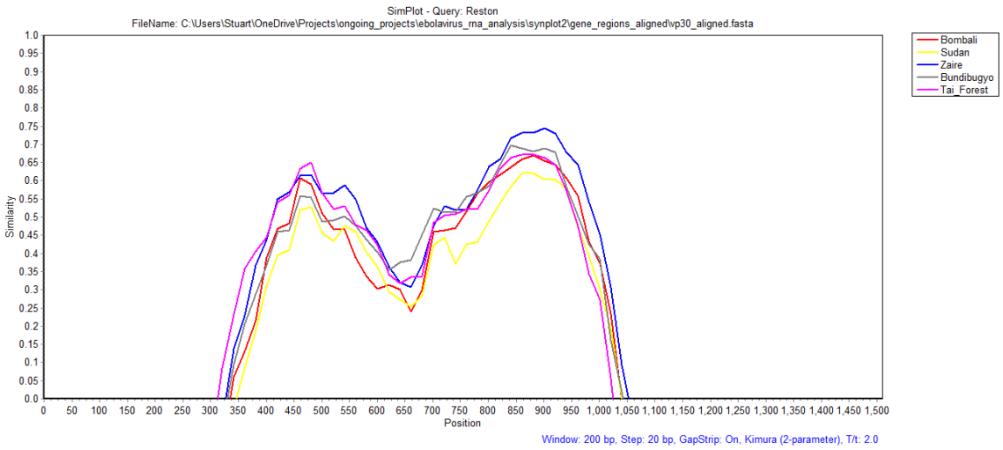


Figure s5. SimPlot graph for VP30 protein.

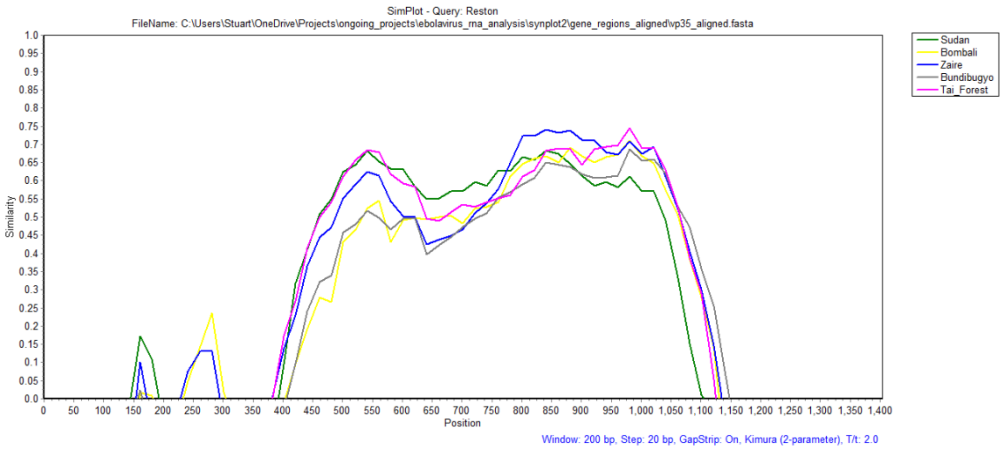


Figure s6. SimPlot graph for VP35 protein.

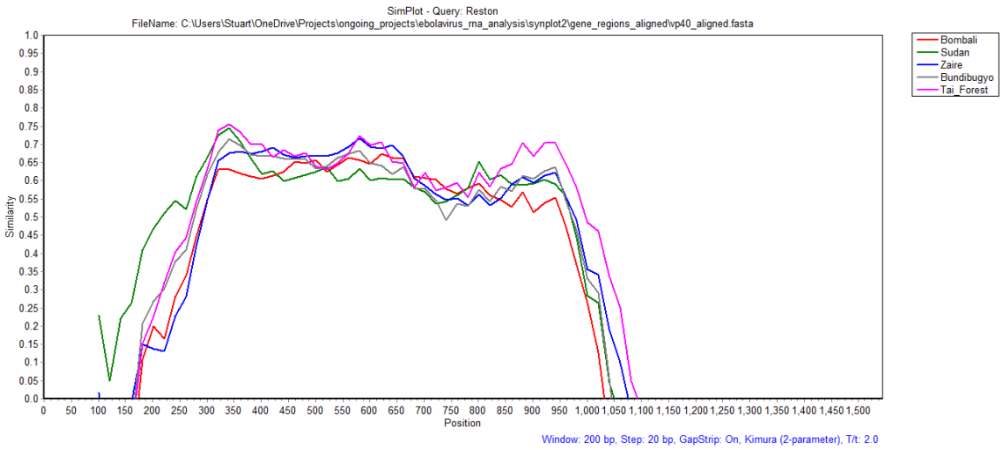


Figure s7. SimPlot graph for VP40 protein.

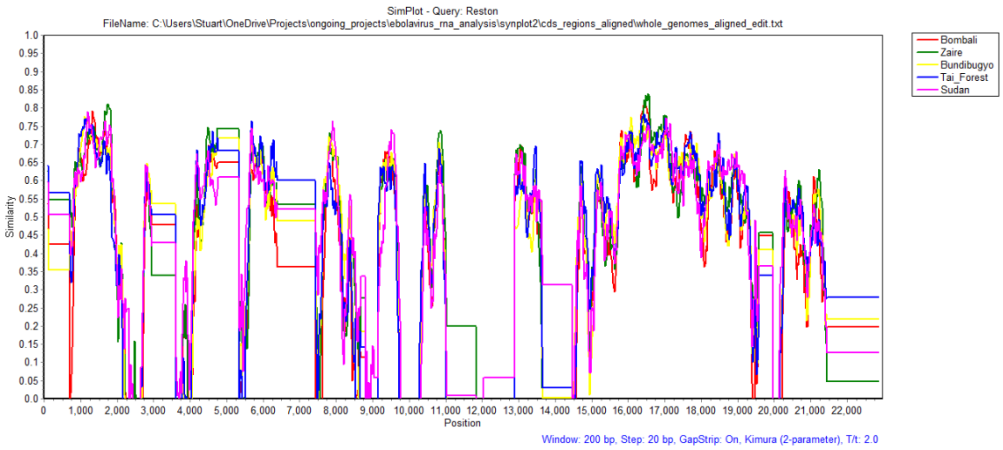
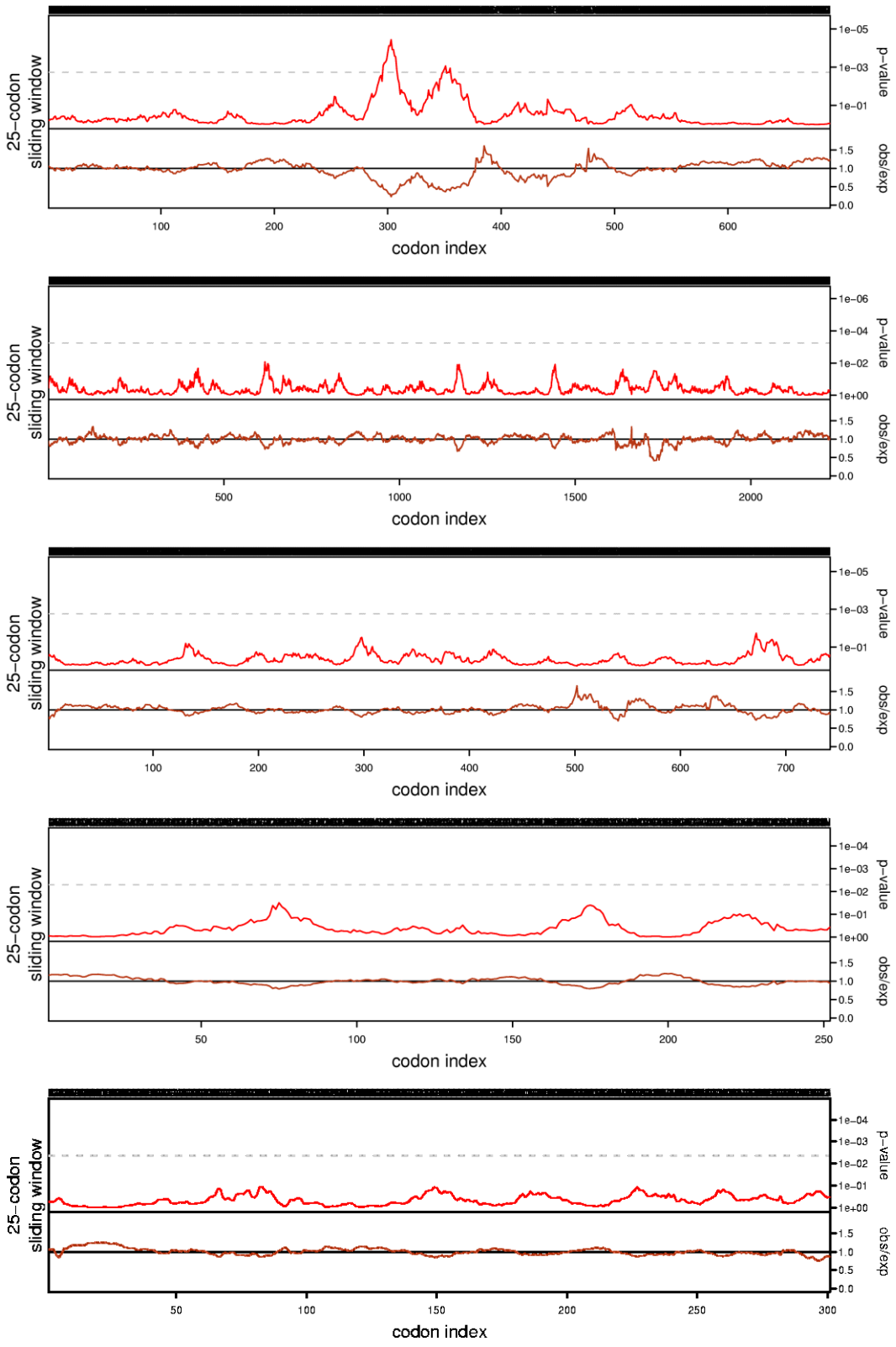


Figure s8. SimPlot graph for Ebolavirus genome.


25−codon
sliding window
p−value
obs/exp
codon index

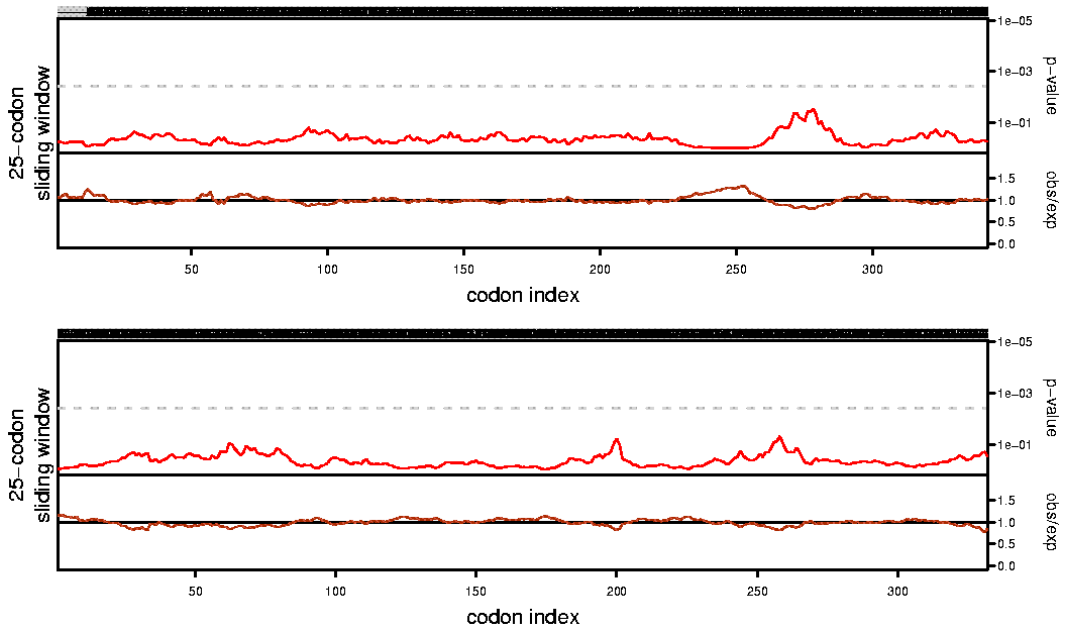


Figure s9. Synplot2 graphs for all Ebolavirus proteins