Characterising the process of attenuation by serial egg passaging of Infectious bronchitis virus (IBV) using genomic and phenotypic methods

Michael Steven Oade

Bachelor of Science in Biomedical Science

Submitted in fulfilment of the requirements for the

Doctorate of Philosophy, Virology

Institute of Biodiversity, Animal Health and Comparative Medicine

College of Medical, Veterinary and Life Sciences (MVLS)

University of Glasgow

March 2020

Abstract

Vaccines against infectious bronchitis virus (IBV) are produced by repeated passage of virulent IBVs in embryonated hens' eggs. Although used as a standard method by industry to generate a vaccine, the mechanism behind attenuation remains unclear. Moreover, the genetic changes that occur during the process of attenuation by egg passage largely go unreported. RNA viruses such as IBV do not exist as clonal structures but rather a population of genetically related viruses that interact on a functional scale. The evolution of IBV is purportedly the influence of two factors; the generation of genetic diversity and selection of pre-existing genetic variants. The origin of attenuating mutations as a result of serial egg passaging is a subject of debate.

A panel of egg-passaged isolates of IBV, four started using a virulent population and four started using the virulent population's clone have previously been generated in the lab. The starting, intermediate and final attenuated viruses were deep sequenced using an array of high-throughput sequencing (HTS) techniques. The genetic differences occurring at both consensus- and subconsensus level between virulent and attenuated IBVs are reported here as is the order of mutational appearance; both are used to inform as to the evolutionary pathways leading to attenuation. The use of HTS to study the process of attenuation by egg passaging is novel to the field.

During this analysis, it was observed that the 3′ untranslated region (UTR) of IBV undergoes a proportionally higher rate of mutation compared to the rest of the genome. The involvement of the region in governing virus pathogenicity, to elude as to a possible mechanism of IBV attenuation, was studied by reverse genetics exchanging a virulent IBV's 3′ UTR with that of a non-pathogenic IBV equivalent. Interestingly, this virus was shown to be non-pathogenic when introduced to birds.

Table of contents

List of figures

Chapter 1 – Introduction

Chapter 2 – Materials and Methods

Chapter 3 - Characterising parallel attenuated M41-CK viruses serially passaged in embryonated eggs

Chapter 4 – Exploring how an IBV population evolves during attenuation by egg passaging

Chapter 5 – Exploring how an IBV clone evolves during attenuation by egg passaging

Chapter 6 – Towards unmasking the coding potential of IBV and the generation of a rationally attenuated virus

List of tables

Chapter 1 – Introduction

Chapter 2 – Materials and Methods

Acknowledgements

I would like to give a heartfelt thanks to all the Dr Erica Bickertons for their unwavering support over the last years, and to Professor John Hammond and Dr Richard Orton also for their guidance and collective influence into the project.

Many thanks must go to all the members, past and present, of the Coronaviruses, Birnavirus and NCI groups who have provided much help and support. I would like to give special mention to Sarah Keep, Professor Paul Britton and Dr Graham Freimanis, all of whom have been instrumental to the development of wider project and without whom I would not have received funding. I must also thank the Animal Health Research Club (ARC) and the Biotechnology and Biological Sciences Research Council (BBSRC) for approving the grant and providing funding for my research.

I am also grateful to the Pirbright Institute as a whole for providing the opportunity and allowing me to travel and present my research at a national and international level. Thanks must also go the University of Glasgow for their support over the past four-years

And last but by no means least, I would like to thank my parents that, despite the fact they have no idea what I do, have continued to provide their love, encouragement and support.

Declaration

I declare that the work presented in this thesis is original and is author's sole effort except where stated by special reference, and that is not been presented previously to support another degree.

Definitions, abbreviations and acronyms

AA	Amino acid
ACE2	Angiotensin I converting enzyme 2
ANOVA	Analysis of variance
APS	Adenosine 5´ phosphosulfate
BES	N,N-bis(2-hydroxyethyl)-2-aminoethanesulphonic acid
BHK	Baby Hamster Kidney
BSA	Bovine Serum Albumin
CavV	Cavally virus
cDNA	Complementary DNA
CEACAM1	Carcinoembryonic antigen-cell adhesion molecule
CH	Central helix
CK/CKCs	Chick kidney / chick kidney cells
CoV	Coronavirus
CPE	Cytopathic effect
CS	Core sequence
CS-B	Core sequence - Body
CS-L	Core sequence - Leader
CSU	Central services unit
DMV	Double-membrane vesicle
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
dsDNA	Double stranded RNA
E	Envelope protein
E. coli	*Escherichia coli*
EMEM	Eagle's minimum essential media
EP	Egg passage
ER	Endoplasmic reticulum
ERGIC	ER-Golgi intermediate compartment

ExoN	Exoribonuclease
FBS/FCS	Foetal bovine/calf serum
FPV	Fowlpox virus
GMEM	Glasgow minimum essential media
GMP	Guanosine monophosphate
gRNA	Genomic RNA
gpt	Xanthine phosphoribosyltransferase
GTP	Guanosine-5'-triphosphate
HCoV	Human coronavirus
HE	Hemagglutinin esterase
HR	Heparan sulfate
HδR	Hepatitis delta ribozyme
IAH	Institute for Animal Health
IB	Infectious bronchitis
IBV	Infectious bronchitis virus
ICTV	International Committee for the Taxonomy of Viruses
IFN	Interferon
IMP	Inosine monophosphate
Kb	Kilobases
LB	Luria-Bertani broth
M	Membrane protein
MEM	Minimum essential medium
MERS-CoV	Middle east respiratory syndrome coronavirus
MHV	Mouse hepatitis virus
MLeV	Microhyla letovirus
MOI	Multiplicity of infection
MPA	Mycophenolic acid
mRNA	Messenger RNA
N	Nucleocapsid protein

NC	No coverage
NDiV	Nam Dinh virus
NDV	Newcastle disease virus
NGS	Next generation sequencing
NS	Nonsynonymous
nsp(s)	Non-structural protein(s)
nt	Nucleotide
NV	No variant
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFU	Plaque forming units
PLpro	Papain-like protease
pp1a/1ab	Polyprotein 1a/1ab
PPU	Poultry production unit
qPCR	Quantitative polymerase chain reaction
RBD	Receptor binding domain
RdRp	RNA-dependent RNA polymerase
Rep1a/1ab	Replicase 1a/1ab
rIBV	Recombinant infectious bronchitis virus
RIR	Rhode Island Red
RNA	Ribonucleic acid
RNA-seq	RNA sequencing (using high-throughput sequencing)
RNP	Ribonucleoprotein
rpm	Revolutions per minute
RT	Room temperature
RTC	Replication-transcription complex
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
rVV	Recombinant vaccinia virus

S	Synonymous
SARS-CoV	Severe acute respiratory syndrome coronavirus
SEM	Standard error of mean
SG	Stress granule
sg	Subgenomic
SHFV	Simian hemorrhagic fever virus
SOC	Super Optimal broth Catabolite repression
SPF	Specific pathogen free
ssRNA	Single stranded RNA
TBE	Tris borate EDTA
TCoV	Turkey coronavirus
TDS	Transient dominant selection
TE	Tris/HCl EDTA buffer
TGEV	Transmissible gastroenteritis virus
TM	Transmembrane domain
TOC	Tracheal organ culture
TRS	Transcription regulatory sequence
TRS-B	Transcription regulatory sequence (body)
TRS-L	Transcription regulatory sequence (leader)
UH	Upstream helix
UTR	Untranslated region
viscRNA-Seq	Virus-inclusive single cell RNA sequencing
VV	Vaccinia virus
XMP	Xanthosine monophosphate

Chapter 1 – Introduction

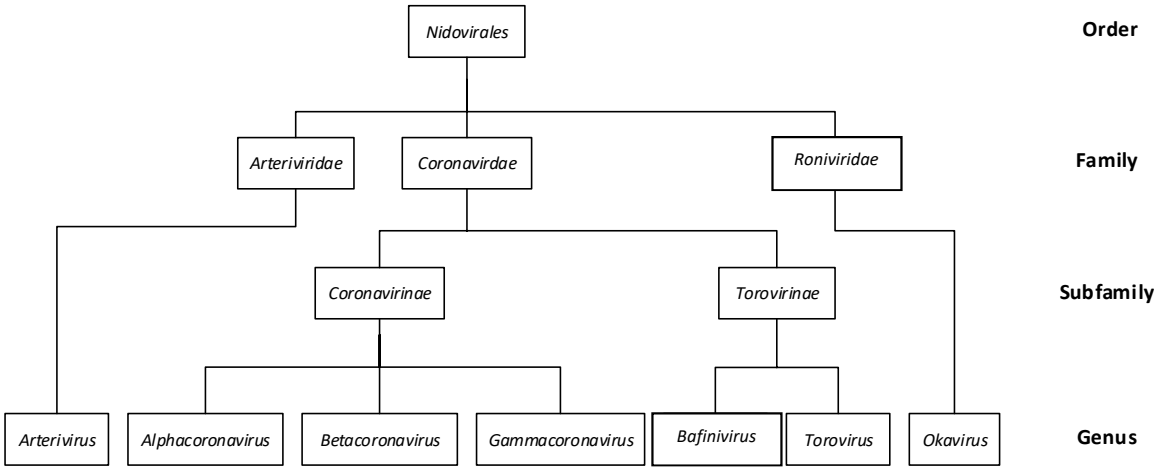
1.1. The *Nidovirales* order

The *Nidovirales* order consists of enveloped viruses with a positive-sense, single-stranded RNA genome (Group IV, Baltimore classification). In 2018 the *Nidovirales* order was restructured by ICTV (International Committee for the Taxonomy of Viruses) to sub-divide this order into seven new suborders (Figure 1.1). Prior to this (ICTV 9th Report, 2009 Release), the *Arteriviridae*, *Coronaviridae* and *Roniviridae* families were the only members of the *Nidovirales* order with the *Mesonivirdae* family added in 2012 with the discovery and subsequent accommodation of Nam Dinh virus (NDiV) and Cavally virus (CavV).

Although classified under the same order, large and small nidoviruses can be distinguished based on genome size. *Arteriviridae*, have a genome length between 12.7 and 15.7 kb and are considerably smaller than the *Okavirus* (26.2-26.6 kb), *Bafinivirus* (~26.6 kb), *Torovirus* (28-28.5 kb) and *Coronavirus* (26.4-31.7 kb) genome lengths of the larger subsection of nidoviruses.

Despite differences in genomic size, the genomic organisation within the *Nidovirales* order remains similar. The differences between constituent members of this order lie in the variety and function of encoded accessory proteins, and the mechanism of transcription used to generate the nested-set of subgenomic messenger RNAs (sg mRNAs) from which the *Nidovirales* order derives its name. Roniviruses for example employ a mechanism of non-discontinuous transcription whereas Toroviruses employ a mixed strategy of continuous and discontinuous transcription. Both differ from the convention that Nidoviruses, including the *Coronaviridae* family, employ a method of discontinuous synthesis for transcription (see 1.7.3).

Virus Taxonomy 2009 Release:



Virus Taxonomy 2018b Release:

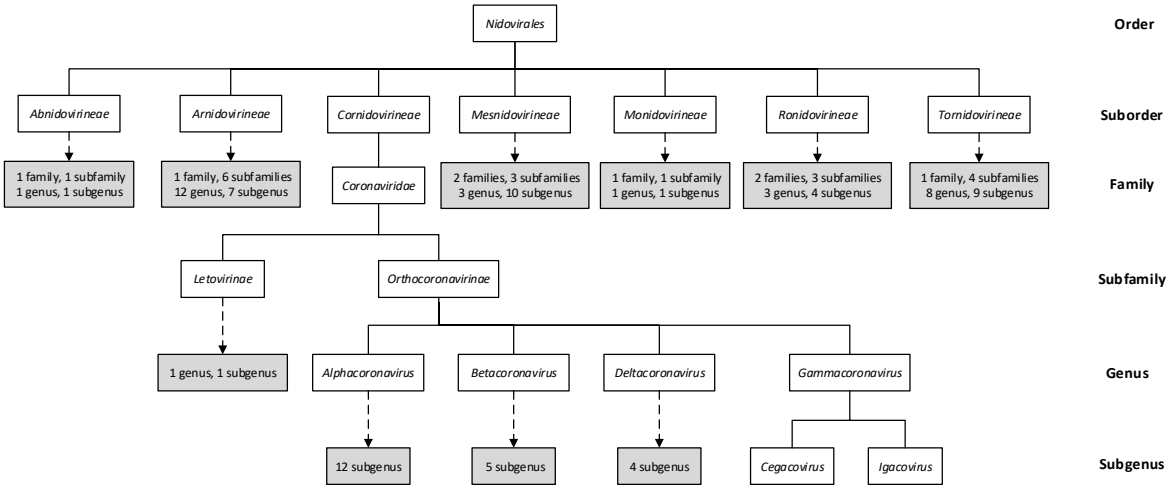


Figure 1.1: The Nidovirales order as classified by the International Committee on the Taxonomy of Viruses (ICTV). *The top tree includes the full 2009 Nidovirales classification whereas the bottom tree represents a condensed 2018b Nidovirales classification. The 2018b classification added the suborder and subgenus tiers to the Nidovirales order and includes a total of 26 new genera. The introduction of new taxonomic levels and reclassification of previously identified viruses is at least in part the result of classifying virus-like sequences from genomic data.*

1.2. The *Coronaviridae* family

The *Coronaviridae* family, so called due to their distinctive appearance when viewed by electron microscopy, is currently the sole member of the *Cornidovirineae* suborder and is classified into the *Letovirinae* and *Orthocoronavirinae* subfamilies (ICTV 2018b Release). These subfamilies can be further classified into the alpha-, beta-, gamma- and deltacoronavirus genera (*Orthocoronavirinae*) and the added *Alphaletovirus* genus within the *Letovirinae* subfamily. Typically, alpha- and betacoronaviruses infect mammals, gammacoronaviruses infect avian species and deltacoronaviruses possess the ability to infect both avian and mammalian species (Table 1.1). Microhyla letovirus (MLeV), the virus responsible for *Letovirinae* subfamily and cascading groups, has been identified by analysis and identification of virus-like sequences in intracellular RNA pools isolated from pygmy frogs (Bukhari et al., 2018, Zhao et al., 2016).

Table 1.1. Orthocoronavirinae genera as per ICTV 2018b taxonomy classification

GENUS	SUBGENUS	SPECIES
ALPHACORONAVIRUS	*Colacovirus*	*Bat coronavirus CDPHE15*
	Decacovirus	*Bat coronavirus HKU10, Rhinolophus ferrumequinum alphacoronavirus HuB-2013*
	Duvinacovirus	*Human coronavirus 229E (HCoV-229E)*
	Luchacovirus	*Lucheng Rn rat coronavirus*
	Minacovirus	*Ferret coronavirus, Mink coronavirus 1*
	Minunacovirus	*Miniopterus bat coronavirus 1, Miniopterus bat coronavirus HKU8*
	Myotacovirus	*Myotis ricketti alphacoronavirus Sax-2011*
	Nyctacovirus	*Nyctalus velutinus alphacoronavirus SC-2013*
	Pedacovirus	*Porcine epidemic diarrhea virus, Scotophilus bat coronavirus 512*
	Rhinacovirus	*Rhinolophus bat coronavirus HKU2*
	Setracovirus	*Human coronavirus NL63, NL63-related bat coronavirus strain BtKYNL63-9b*
	Tegacovirus	*Alphacoronavirus 1*
BETACORONAVIRUS	*Embecovirus*	*Betacoronavirus 1, China Rattus coronavirus HKU24, Human coronavirus HKU1, Murine coronavirus (MHV)*
	Hibecovirus	*Bat Hp-betacoronavirus*
	Merbecovirus	*Hedgehog coronavirus 1, Middle East respiratory syndrome coronavirus (MERS-CoV), Pipistrellus bat coronavirus HKU5, Tylonycteris bat coronavirus HKU4*
	Nobecovirus	*Rousettus bat coronavirus GCCDC1, Rousettus bat coronavirus HKU9*
	Sarbecovirus	*Severe acute respiratory syndrome coronavirus (SARS-CoV)*
DELTACORONAVIRUS	*Andecovirus*	*Wigeon coronavirus HKU20*
	Buldecovirus	*Bulbul coronavirus HKU11, Coronavirus HKU15, Munia coronavirus HKU13, White-eye coronavirus HKU16*
	Herdecovirus	*Night heron coronavirus HKU19*
	Moordecovirus	*Common moorhen coronavirus HKU21*
GAMMACORONAVIRUS	*Cegacovirus*	*Beluga whale coronavirus SW1*
	Igacovirus	*Avian coronavirus (e.g. Infectious bronchitis virus)*

1.3. Coronavirus genome organisation

Coronaviruses possess the largest genomes of any known RNA viruses, rivalled only by other members of the *Nidovirales* order. Genomic RNA possesses a 5' cap and 3' poly (A) tail and contains multiple open-reading frames (ORFs) the largest of which is the Replicase gene, encoded by the initial 5' two thirds of the genome. Downstream of the Replicase gene coronaviruses encode the structural proteins; Spike (S), Envelope (E), Membrane (M) and Nucleocapsid (N), arranged in a fixed order conserved among coronaviruses (Figure 1.2). Some betacoronaviruses may also encode a Hemagglutinin esterase (HE) between Replicase and Spike genes. Interspersed among these structural genes are genes encoding accessory proteins which are generally dispensable for replication *in vitro*. As convention names coronavirus accessory proteins based on genomic position it is possible that, in instances where ORF function has been identified, ORFs in different coronaviruses with the same given name perform different functions.

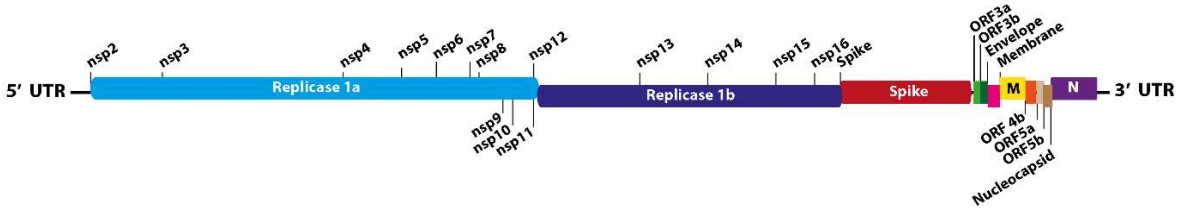


Figure 1.2: Schematic presentation of a coronavirus genome. *Coronaviruses possess a genome between 28-32 kb, the largest genome of any known RNA virus. This cartoon represents the genome of the gammacoronavirus infectious bronchitis virus and does not encode non-structural protein (nsp) 1 within Replicase 1a/1ab nor the structural protein HE encoded by betacoronaviruses. In this example, the virus encodes five accessory proteins (3a, 3b, 4b, 5a and 5b) encoded by three genes.*

Outside coding regions, the coronavirus genome has a highly complex and conserved structure with between 200 and 500 nt either end of the genome forming the 5′ and 3′ untranslated regions (UTRs). Conserved stem-loop structures have been identified in the 5′ UTR (Brown et al., 2007, Guan et al., 2011) while the 3′ UTR possesses a distinctive bulged stem-loop structure (Williams et al., 1999). This 3′ UTR structure overlaps with a pseudoknot structure with both structures unable to form simultaneously (Hsue et al., 2000). As a result of this feature it has been suggested that such cis-acting sequences are capable of regulating the different stages of RNA synthesis though no mechanism of action has yet been determined.

Secondary RNA structures are not exclusive to non-coding regions and may also occur in coding regions too. Other documented RNA structures include coronavirus frameshifting element involved in translation of ORF1b in ORF1ab (Brierley et al., 1987, Bredenbeek et al., 1990, Herold et al., 1993) (see 1.4) and coronavirus packaging signal involved in packaging the viral genome by nucleocapsid protein (Qin et al., 2003, Narayanan and Makino, 2001). Additionally, coronaviruses also encode a transcription-regulatory sequence (TRS) both within the 5′ UTR and immediately upstream (5′) of every encoded ORF. This conserved sequence is important for both genomic replication and sg mRNA production and will be discussed later.

1.4. Replicase gene

Comprised of two overlapping ORFs named 1a and 1b, the Replicase gene is translated to produce two polyproteins pp1a and pp1ab, the latter produced by programmed -1 ribosomal frameshifting, utilising a slippery sequence (5′-UUUAAAC-3′) (Brierley et al., 1992) and RNA pseudoknot (Baranov et al., 2005). Proteolytic processing of pp1ab results in the production of 15 or 16 mature proteins referred to as non-structural proteins (nsps) (Table 1.2).

Table 1.2. Cleavage products of coronaviruses replicase polyproteins pp1a and pp1ab and their associated function. *Adapted from ICTV 9th report (Coronaviriade profile) (2012) and Fehr and Perlman, 2015 (Fehr and Perlman, 2015).*

PROTEIN	ASSIGNED FUNCTION	REFERENCES
nsp1*	Degradation of host mRNAs, inhibition of translation, blocked innate immune response. IFN antagonist.	(Narayanan et al., 2008)
nsp2	Unknown; associates with Replication Transcription Complexes (RTC) and binds to prohibitin protein.	(Cornillez-Ty et al., 2009)
nsp3	Papain-like protease activity PL^{pro}; cleavage of viral polyprotein.	(Baker et al., 1993)
	ADP-ribose-1″phosphatase (ADRP) activity; promotes cytokine expression	(Saikatendu et al., 2005)
	Membrane rearrangements	(Hagemeijer et al., 2014)
nsp4	Unknown; potential transmembrane scaffold protein. Important for double membrane vesicle (DMV) structure. Induce membrane pairing.	(Hagemeijer et al., 2014, Doyle et al., 2018)
nsp5	Main protease M^{pro}, cleaves viral polyprotein	(Lu et al., 1995)
nsp6	Generates autophagosomes and inhibits expansion.	(Cottam et al., 2014)
nsp7	ssRNA binding; forms hexadecameric complex with nsp8. May act as a processivity clamp for RNA polymerase.	(Zhai et al., 2005)
nsp8	Noncanonical “secondary” RdRp with putative primase activity. May act as primase.	(Imbert et al., 2006)
nsp9	ssRNA and DNA binding; associates with RTCs.	(Egloff et al., 2004)
nsp10	Forms heterodimer with nsp14 and nsp16 and stimulates their activity.	(Chen et al., 2011)
nsp11	Unknown.	
nsp12	RNA-dependent RNA polymerase (RdRp)	(Xu et al., 2003)
nsp13	RNA helicase	(Ivanov et al., 2004)
	5′ triphosphatase	
nsp14	3′-5′ exoribonuclease (ExoN); proofreading of viral genome	(Minskaia et al., 2006)
	Guanine-N7-methyltransferase; adds 5′ cap to viral RNAs	(Chen et al., 2009)
nsp15	Hexameric uridylate-specific endoribonuclease (NendoU)	(Bhardwaj et al., 2006)
nsp16	Ribose-2’-O-methyltransferase (RNA cap formation); shields viral RNA from MDA5 recognition.	(Decroly et al., 2008)

* Denotes nsp not coded by all genera of coronaviruses. Purple = nsp encoded in both Rep1a and Rep1ab. Red = nsp encoded in Rep 1a only. Blue = nsp encoded in Rep1ab only.

1.5. Structural proteins

Members of the *Coronaviridae* family typically possess at least three envelope proteins including spike glycoprotein, envelope (E) protein and membrane (M) protein. Some coronaviruses (namely a subset of betacoronaviruses such as mouse hepatitis virus (MHV)) may also encode hemagglutinin-esterase (HE) structural protein, essential for virion attachment to a host cell (de Groot, 2006). The genome is packaged inside the virion, found in association with nucleocapsid (N) protein. A schematic of an IBV virion is shown in Figure 1.3.

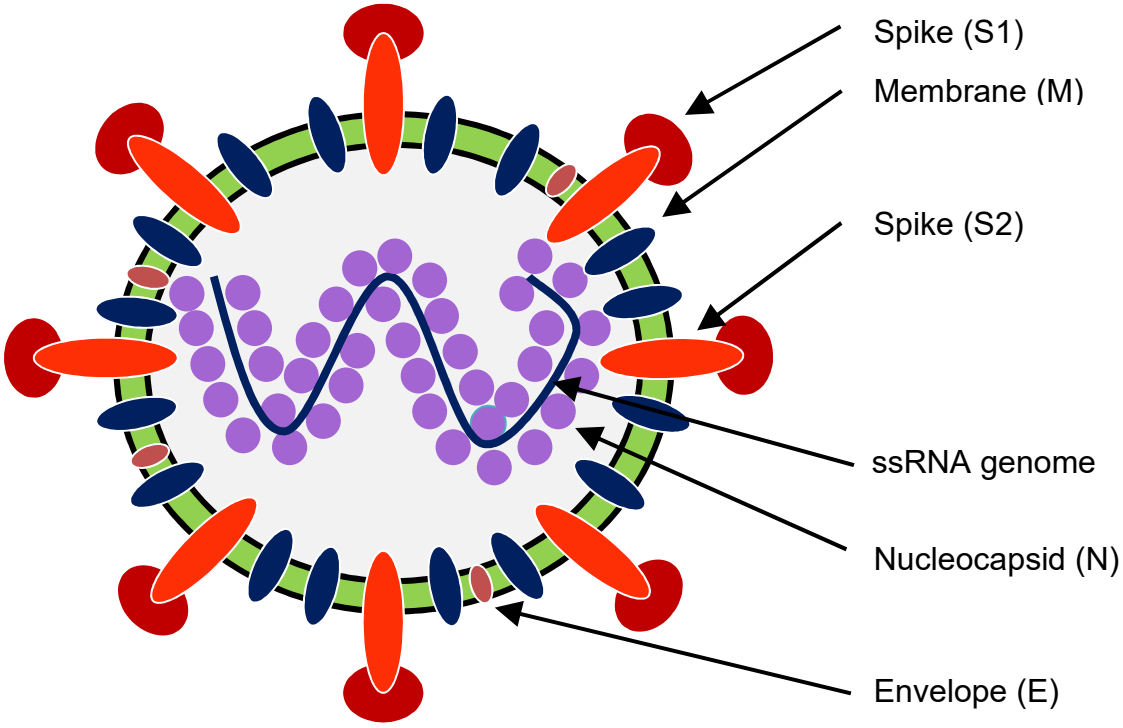


Figure 1.3: Schematic presentation of a coronavirus virion. *Members of the Coronaviridae family possess at least three envelope proteins including Spike (S1 and S2), Membrane (M) and Envelope (E). The ssRNA viral genome is found associated with Nucleocapsid (N) protein within the virion.*

1.5.1. Spike

Spike protein is responsible for mediating coronavirus entry and is a critical determinant of host range and tissue tropism. This class I fusion protein (Bosch et al., 2003) is synthesised as a single polypeptide and consists of two subunits (S1 and S2) which, in some coronaviruses (e.g. IBV and MHV), requires proteolytic processing to function (S1/S2 cleavage). Spike protein can be broadly separated into three segments (ectodomain, transmembrane anchor and intracellular tail) (Li, 2016) (Figure 1.4). Spike is found as a homotrimer presented on surface of the virion (Figure 1.3) with S1 subunits forming a globular head and the S2 subunits forming cytoplasmic tail. This structure forms distinct protrusions from virions and responsible for the 'corona' appearance from which the name coronavirus derives.

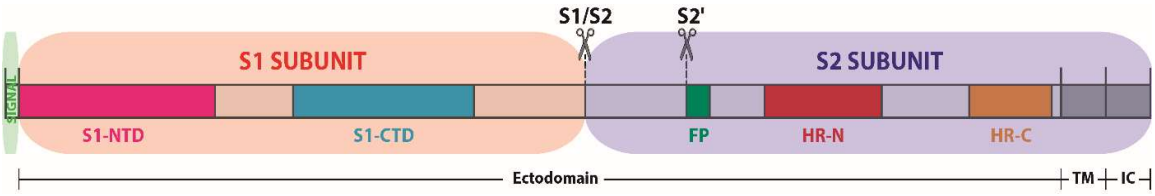


Figure 1.4: Schematic presentation of the spike gene. *The spike glycoprotein consists of two subunits, S1 and S2, the former of which is responsible for host cell attachment and the latter of which is responsible for membrane fusion. S1/S2 and S2' cleavage sites are utilised by some coronaviruses either upon receptor recognition or as the spike protein matures. In the former receptor recognition triggers protein cleavage which is required to enable an eventual shift in the spike structure to promote mixing of the host and viral membranes. NTD = N-terminal domain, CTD = C-terminal domain, FP = fusion peptide, HR = heptad repeat, TM = transmembrane, IC = intracellular.*

S1 is responsible for host cell attachment and spans the ectodomain segment of Spike (Figure 1.4) with two major domains, S1 N-terminal domain (S1-NTD) and S1 C-terminal domain (S1-CTD) being identified (Li, 2012). The activity of one or both domains are thought to bind to receptors and function as a receptor-binding domain (RBD). S1-NTD is thought to primarily bind to sugars (Promkuntod et al., 2014, Peng et al., 2012), though protein-protein specific interactions between S1-NTD in MHV and carcinoembryonic antigen related cell adhesion molecule 1 (CEACAM1) have been identified (Peng et al., 2011). In contrast the interaction target for S1-CTD are exclusively proteins, with angiotensin I converting enzyme (ACE2) being implicated in both the alphacoronavirus human coronavirus NL63 (HCoV-NL63) (Hofmann et al., 2005) and the betacoronavirus severe acute respiratory syndrome (SARS) (Li et al., 2003) S1-CTD binding. Conversely transmissible gastroenteritis virus (TGEV) (Delmas et al., 1992) and porcine respiratory coronavirus (PRCV) (Godet et al., 1994), both alphacoronaviruses, recognise aminopeptidase N. This typifies an observation that evolutionary distance is not associated with recognition target of either the S1 NTD or CTD. Highly similar coronaviruses can have different recognition targets whereas evolutionary more distant coronaviruses may have the same target.

In the case of IBV it is believed sialic acid is a receptor determinant for infection (Winter et al., 2006) with the RBD subsequently mapped to the N-terminal 253 amino acids of S1 for M41 (a strain of IBV) infection of the avian respiratory tract (Promkuntod et al., 2014). This sugar binding site corresponds to the S1-NTD. Utilising IBV spike structural data, it is believed that a second RBD exists in the S1-CTD that also enables binding to chicken cells (Shang et al., 2018). This receptor is yet to be determined.

The S2 subunit is responsible for driving fusion between host and viral membranes. This subunit spans the ectodomain, TM and IC domains of spike. Within the ectodomain, S2 contains two heptad repeat regions (HR-N and HR-C), which form a six-helix bundle along the stork/tail structure of spike, and an internal fusion

peptide (Li, 2016). All three domains are involved in the membrane fusion activity of S2 that is facilitated by conformational change from a prefusion to postfusion state similar to that of influenza virus haemagglutinin (HA). The interaction between the S1 (namely the S1-NTD) and S2 subunits acts to structurally inhibit S2 preventing premature maturation into the postfusion state until triggered by receptor binding.

The structures of coronavirus spike have only recently been solved (Walls et al., 2016, Kirchdoerfer et al., 2016, Shang et al., 2018).

1.5.2. Membrane (M)

Membrane protein represents the most abundant protein found on the virion surface and consists of a short glycosylated NTD located outside the virion, three TM domains and a long C-terminal tail located inside the virion (Rottier, 1995). The protein itself is believed to be fundamental in virion assembly due to an inherent ability to form protein-protein interactions with all coronavirus structural proteins including itself and hemagglutinin esterase (Boscarino et al., 2008, Opstelten et al., 1995, Hurst et al., 2005, de Haan et al., 2000, Nguyen and Hogue, 1997). Membrane protein has also been shown to form interactions with RNA possessing a genomic packaging signal (Narayanan et al., 2003).

The significance of M in coronavirus assembly is underlined by the fact that co-expression of M and E is sufficient for virus-like particle (VLP) assembly in most coronaviruses (Vennema et al., 1996, Corse and Machamer, 2000), a process which does not require spike protein. Such particles are non-infectious but nonetheless form. The fragmentary expression of the coronavirus structural proteins shows that, in the absence of both components (M and E protein), spike protein is transported to the cell surface and membrane protein accumulates in the Golgi complex. In the presence of M protein spike is instead retained in the Golgi complex, close to the endoplasmic reticulum (ER)-Golgi intermediate compartment (ERGIC) which coronaviruses bud from. The suggested role for M is

therefore that it acts as to chaperone the components necessary for virion assembly to the site of virion assembly.

1.5.3. Envelope (E)

With approximately only 20 copies per virion, envelope (E) protein represents the least abundant structural protein in coronaviruses. A notable exception to this convention however is in IBV where E is found in relative abundance in the virion (Liu and Inglis, 1991). E protein is a short peptide consisting of a N-terminal ectodomain, hydrophobic TM domain and cytoplasmic C-terminal tail with a Cys-rich region and two Pro residues present in the tail that is conserved across coronaviruses (reviewed in Ruch and Machamer (2012)).

Unlike spike, membrane and nucleocapsid, the absence of the envelope structural protein in recombinant viruses is not always lethal and is instead associated with a reduced pathogenicity *in vitro* and *in vivo* (DeDiego et al., 2007). This in turn has led to the development of ΔE viruses being used as live attenuated vaccine candidates (Netland et al., 2010, Almazan et al., 2013). The exact mechanism of attenuation by E protein mutation is not fully understood. This is, at least in part, due to multifaceted role E protein has been reported as possessing in coronaviruses.

E protein is capable of self-assembly to form a pentameric membrane protein that exhibits ion channel activity (Wilson et al., 2004). Separately, E protein has been shown to play a role in the secretory pathway and trafficking of virions, with a single residue having a major contributing factor to this activity (Ruch and Machamer, 2011). This corresponds to a known channel inactivating mutation identified in SARS-CoV (Nieto-Torres et al., 2014). Furthermore, recombinant coronaviruses preventing E from oligomerising accumulate intracellularly suggesting a potential link between TM integrity and trafficking efficiency (Ruch and Machamer, 2011). This would suggest Golgi rearrangement is associated with

ion channel activity for the release of virions from the cell as has been identified in other virus viroporins (e.g. influenza M2). While this would correspond to a compromised virus output seen in ΔE coronaviruses (Kuo and Masters, 2003), this does not explain the role of E in pathogenicity and attenuation suggesting an as yet uncharacterised function. Other proposed roles of E protein are non-exclusive and include preventing aggregation of M protein (Boscarino et al., 2008) and assisting in inducing membrane curvature in virion release (Fischer et al., 1998).

1.5.4. Nucleocapsid (N)

Though membrane protein has been identified as possessing an RNA-binding capability, nucleocapsid protein is considered the most important coronavirus protein possessing such ability (Narayanan et al., 2003). N protein consists of two distinct, highly conserved domains (named N-NTD and N-CTD) that form two independently folded structural regions (reviewed in McBride et al. (2014)). These are separated by an equally conserved intrinsically disordered central region referred to as the linker region. All three domains have been shown to bind with coronavirus virus RNA which assist in N protein's primarily role; to package viral RNA into helical ribonucleoprotein (RNPs) complexes called nucleocapsids for inclusion into the viral particle.

Notwithstanding the formation of RNPs, a crucial component in the coronavirus life cycle, N protein is also implicated in regulating virion assembly and envelope formation. Here, protein-protein interactions between each of the envelope proteins and the nucleocapsid protein promote the inclusion of nucleocapsid into enveloped viruses at the ERGIC (Simons and Garoff, 1980), the site of coronavirus budding. In addition, contradictory to what was previously stated, the presence of N protein supplementary to M and E protein has been shown to improve VLP formation efficiency (Ruch and Machamer, 2011). Indeed in some instances, supplemental N protein is a requirement for the recovery of infectious virus in some coronavirus reverse genetic systems (Casais et al., 2001, Yount et al., 2000).

Outside of virion assembly, nucleocapsid protein has been implicated in the efficient replication of the virus genome in both replication and transcription (see 1.7.3). This follows an observation that N protein can bind to full-length genomic RNA, and sg mRNA albeit at a lower efficiency (Cologna and Hogue, 1998). A subsequent discovery identifying an interaction between N and nsp3 (Hurst et al., 2010) has led to the hypothesis that N protein is required for the formation of an initiation complex at the 3′ end of the genome. This complex tethers the infecting nucleocapsid and genomic RNA (gRNA) to the newly translated RTC through its interaction with nsp3 (Hurst et al., 2013). Furthermore, it is proposed that N may facilitate template switching during discontinuous RNA synthesis by lowering the energy barrier required to dissociate the nascent negative-strand RNA from its gRNA template (Zuniga et al., 2007).

Though the RNA-binding capability of nucleocapsid protein is considered its core function, N may also be involved in regulating cellular response to viral infection. In SARS-CoV infections, N protein has been implicated in arresting cell cycle progression (Surjit et al., 2006); a common strategy used by viruses to exploit host cell machinery. N protein has also been identified as a component in stress granules (SG), structures that form when host cell translation is stalled typically in response to environmental stress such as viral infection (Raaben et al., 2007). Formation of these structures is typically promoted by viruses as it diverts resources away from host protein translation to synthesise the viruses own gene products. Separately N protein can also inhibit interferon (IFN) production, a signalling protein generated in response to virus infection, preventing the activation of host anti-viral defences (Lu et al., 2011).

1.5.5. Hemagglutinin esterase (HE)

Hemagglutinin esterase, the fifth structural protein present on a subset of betacoronaviruses, forms part of the viral envelope and mediates reversible attachment to sialic acids found on the host cell surface (reviewed in de Groot

(2006)). The activity of HE, acting as both a lectin and receptor-destroying enzyme, is believed to improve efficiency of S protein-mediated cell entry (Cornelissen et al., 1997).

1.6. Accessory proteins

Accessory proteins (previously called group-specific genes), so-called due to a presumption that they are not essential for replication nor included in the virion, vary significantly in their number and function across coronaviruses. As mentioned previously coronavirus accessory proteins are named relative to their position within the viral genome and the transcript from which the protein is translated, with poor sequence similarity between coronaviruses. MERS-CoV and SARS-CoV for example have 4 and 8 accessory proteins respectively (Liu et al., 2014, Corman et al., 2012). This results in some disparity between accessory proteins identified across different coronaviruses. It is believed that the primary function of encoded accessory proteins is to antagonise the antiviral action of IFN in infected cells (Liu et al., 2014).

Infectious bronchitis virus encodes at least five accessory proteins named 3a, 3b, 5a, 5b and 4b. The activities of 3a, 3b and 5b are believed to modulate type 1 IFN response with the suggestion that 5b compensates for the absence of nsp1 as an IFN antagonist in gammacoronaviruses (Kint et al., 2015, Kint et al., 2016). Furthermore the deletion of accessory proteins 3a, 3b, 5a and 5b has been associated with an attenuated phenotype *in vivo* and *in vitro* (Laconi et al., 2018). Work to elucidate the role of IBV 3a and 3b has suggested that 3a modulates IFN signalling proteins MAVS and IRF7 through interaction with IFN regulatory proteins RNF5 and CAND1 (Hall, 2017), whereas 3b may be involved in apoptosis or IFN antagonism (Kint et al., 2015).

Despite its name, 4b is synthesised using a relatively newly identified independent sg mRNA transcribed from a non-canonical TRS sequence (Bentley et al., 2013), however it was not established until recently that this sg mRNA is translated (Hall, 2017). 4b interacts with cellular proteins involved in translation and can induce stress granule formation. However, an observation that SGs in IBV-infected cells are non-canonical and cannot be dissembled by cycloheximide treatments suggests that this may not be linked to cell cycle arrest (Hall, 2017).

Given the identification of ORF4b's capability to use a non-canonical TRS sequence and the length of coronavirus genome it is not inconceivable that IBV or indeed other coronaviruses produce transcripts that have yet to be delineated. Subsequently coronaviruses may encode as yet undetermined proteins either by non-canonically transcribed sg mRNAs or encoded polycistronically by already identified sg mRNAs. It has recently been proposed that IBV encodes two more transcripts, 4c (Dinan et al., 2019) and dORF (Dinan et al., 2019, An et al., 2019) which are ribosomally occupied (Dinan et al., 2019) and theorised to be translated into protein.

1.7. Coronavirus replication cycle

The generation of progeny virus occurs rapidly roughly 7-8 hours after initial infection (Schneider et al., 2012). In the case of IBV it has been suggested that the virus lifecycle (Figure 1.5) lasts between 6- 8 hours (Maier et al., 2013).

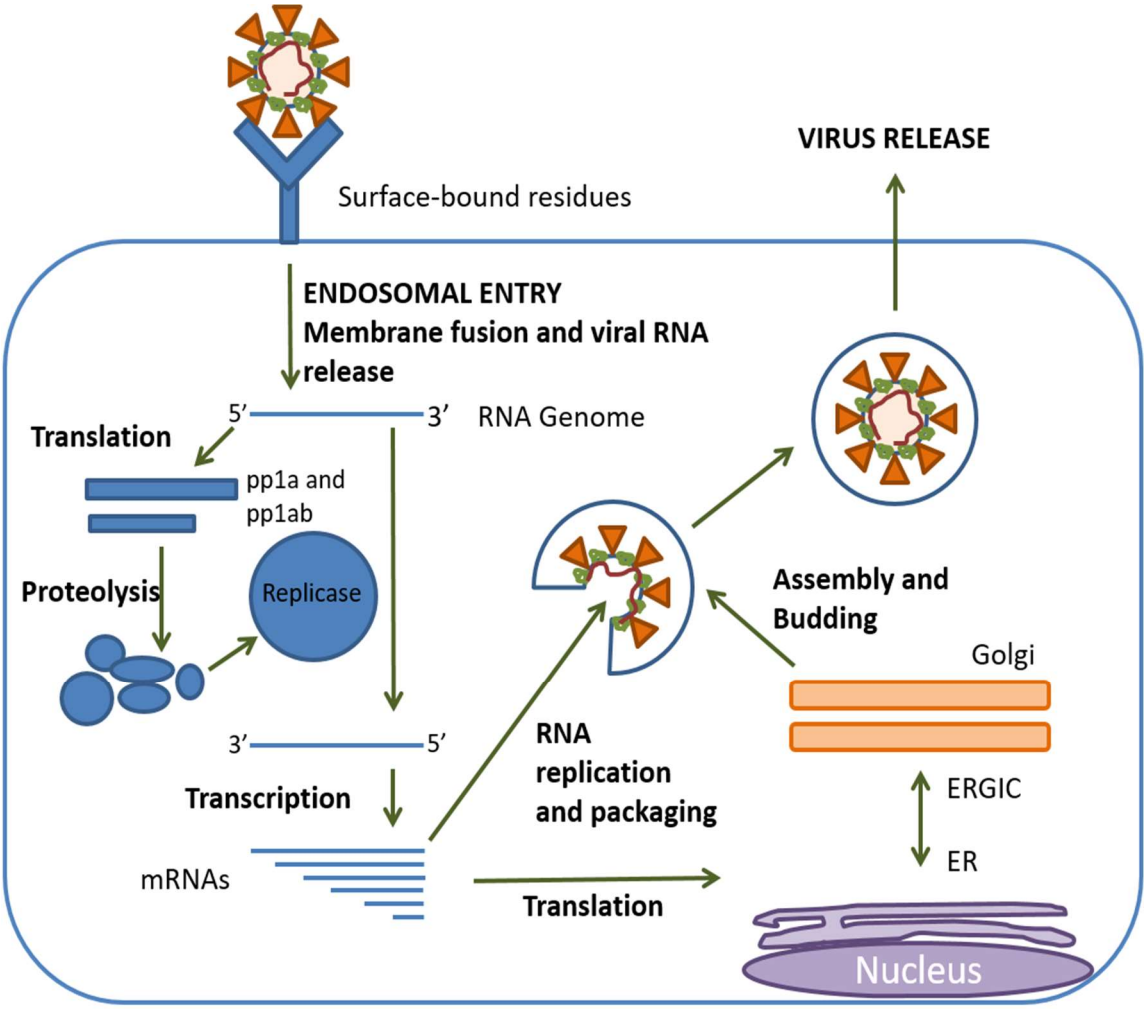


Figure 1.5: Schematic representation of the IBV replication cycle. *The virion binds to particular cell type primarily determined by the S protein present on the viral membrane. Upon recognition, enveloped viruses fuse with the host's cell membrane delivering the nucleocapsid to the cytoplasm. Translation of pp1a and pp1ab occurs by utilising the RNA genome as template mRNA. The expression of the pp1a/pp1ab leads to the formation of the replication/transcription complex (RTC) and, by discontinuous synthesis, results in the production of sg mRNAs encoding structural and accessory proteins. After generation of a positive-sense copy of these sg mRNAs, structural and accessory proteins are produced. Newly translated structural proteins are inserted into the ER where, by using normal cellular processes, the proteins are moved to the ERGIC. Here the immature virion is combined with newly synthesised viral genomes (encapsulated by the N protein) to form mature virions. The mature virus particle is then transported to the cell surface in vesicles and released by exocytosis. The process is reported to occur every 6-8 hours.*

1.7.1. Attachment and entry

The mechanism of attachment and entry is not ubiquitous across all coronaviruses nor are these mechanisms involved fully understood. In a typical example S1 receptor recognition and binding drives the fusion of viral and cellular membranes by proteolytic cleavage of the S protein. This fusion occurs predominately, but not exclusively, within acidified endosomes where two cleavage steps are required (Belouzard et al., 2009). Firstly, the RBD and fusion domains of S protein (i.e. S1 and S2) must be separated at the S1/S2 boundary (see Figure 1.5). Secondly, cleavage at the S2′ site primes and exposes the fusion peptide (FP) for insertion into the cell membrane. This cleavage step is not required in all coronaviruses. Consequently the two HRs form an antiparallel six-helix bundle which facilitate mixing of the viral and cellular membranes and subsequent genome release (Bosch et al., 2003). It is reported that virus is internalised within 15 minutes in *in vitro* infections (Schneider et al., 2012).

Disagreements in the above model revolve around the fact that not all coronavirus spike proteins are cleaved at the S1/S2 boundary and, in coronaviruses where this cleave does occur, it is not exclusively associated with viral entry. IBV spike protein for example is cleaved during biosynthesis in the Golgi.

1.7.2. Replicase gene expression

Upon incorporation into a host cell, coronaviruses must generate the necessary components that facilitate genomic replication and structural gene transcription. This transpires by expression of the replicase gene utilising full length genomic RNA as a template. The replicase gene is translated to produce two co-terminal polypeptides, pp1a and pp1ab, the latter of which is the result of programmed ribosomal frameshifting. Here, when attempting to unwind the RNA pseudoknot the ribosome pauses on the slippery sequence potentially causing a -1 frameshift. This circumvents the pp1a stop codon resulting in translation into replicase 1b (Baranov et al., 2005, Brierley et al., 1989). In the case of MHV it is estimated

that the frameshift occurs in approximately 40% of instances (Bredenbeek et al., 1990) though evidence suggests this can reach up to 70% (Irigoyen et al., 2016). In the case of IBV, it is reported that programmed frameshift occurs between 33 – 40% of instances (Dinan et al., 2019). It is hypothesised that the frameshift exists as a means of regulating the ratio of pp1a products versus pp1b products or to limit production of pp1b until sufficient pp1a products exist.

Expression of the replicase gene results in the production of either pp1a or pp1ab which undergo cleavage to produce nsps 1-11 in the case of pp1a or nsps 1-10 and 12-16 for pp1ab (see Table 1.2). Neither gammacoronaviruses nor deltacoronaviruses possess nsp1. The activity of virus-encoded proteases are responsible for the maturation of the nsps into its functional form. Papain-like proteases encoded within nsp3 cleave at the nsp1/2, nsp2/3 and nsp3/4 boundaries (Mielech et al., 2014) whereas the main protease (Mpro) nsp5 cleaves at the remainder of the nsp junctions.

Mature nsps can then form into the replicase-transcriptase complex (RTC), a structure which is responsible for sg mRNA synthesis and transcription. The exact nature and composition of this RTC as yet remains undefined however some nsps have been identified as containing domains or function believed to be important in RNA synthesis (see Table 1.2). Of note is nsp12 which encodes a RNA-dependent RNA polymerase (RdRp) domain (Gorbalenya et al., 1989, Cheng et al., 2005) and nsp14 which encodes an exoribonuclease (ExoN) domain involved in replication fidelity (Minskaia et al., 2006).

1.7.3. Genome replication and transcription

Coronavirus genomic replication occurs through a process utilising full-length negative-strand RNA intermediates as a template for the production of progeny virus genomes. This process is modulated by RNA structural elements located in the 3′-end of the virus genome that promote the interaction between RdRp and 3′ terminus of the genome. Replication proceeds in a continuous manner to produce a full-length genome. Production of a nested set of sg mRNAs characteristic for coronaviruses is however required for a complete repertoire of viral proteins (Figure 1.6).

While coronavirus genomic replication occurs through a process of continuous RNA synthesis, transcription (defined as the process whereby sg mRNAs are produced) includes a discontinuous step (Sawicki and Sawicki, 1995). This mechanism is responsible for generating a nested set of sg mRNAs that are co-terminal with both the 5′ and 3′ ends of the virus genome. Each sg mRNA contains a 5′ leader sequence corresponding to the 5′ end of the virus genome. This leader sequence is joined to an mRNA body representing the sequence in the genome from an encoded ORF up to the 3′-poly (A) tail. By utilising this mechanism, with exception to the smallest sg mRNA, all sg mRNAs are structurally polycistronic and possess the genetic information for several proteins. In the majority of instances only the 5′ most ORF is translated and therefore coronavirus sg mRNAs are functionally monocistronic although some virus and gene specific exceptions do occur. In the IBV for example gene 3 encodes for proteins 3a, 3b and 3c (Envelope) (Liu et al., 1991). 3a and 3b are both translated by leaky ribosomal scanning while 3c is translated using an internal ribosomal initiation site (Liu and Inglis, 1992). Internal ribosome entry and leaky scanning has also been shown for MHV and has been accepted as convention for all coronaviruses (Thiel and Siddell, 1994, Skinner et al., 1985).

Transcription in coronaviruses is controlled by TRSs located in the 5′ leader sequence (TRS-L) and upstream of each encoded coronavirus gene (TRS-B). Both TRSs include an identical core sequence (CS) though flanking 5′ and 3′ sequences

can vary. This sequence homology between the core sequence in leader (CS-L) and body (CS-B) enable complementary base pairing between CS-L and the complementary CS-B on the negative strand intermediate (Zuniga et al., 2004). This interaction, assisted by the CS flanking sequences, is required for the template switch from nascent negative-strand RNA to the leader. Figure 1.7 is a schematic of discontinuous RNA synthesis.

The exact sequence in CSs and TRSs vary at both a virus- and gene- specific level. Transcripts generated using a CS where the CS-B sequences differ greatly versus CS-L either in length or by sequence are termed noncanonical transcripts. The possibility of generating such transcripts in IBV was first suggested by Stern and Kennedy in 1980 (Stern and Kennedy, 1980) and subsequently confirmed, identifying a novel protein-producing transcript using a AACAA noncanonical CS-B (Bentley et al., 2013). Noncanonical transcripts are not unique to IBV and have previously been described in other coronaviruses including SARS-CoV (Hussain et al., 2005) and MHV (Zhang and Liu, 2000). The percentage sequence identity between CS-L and CS-B has been linked to increased transcription levels (Sola et al., 2005, Zuniga et al., 2004)

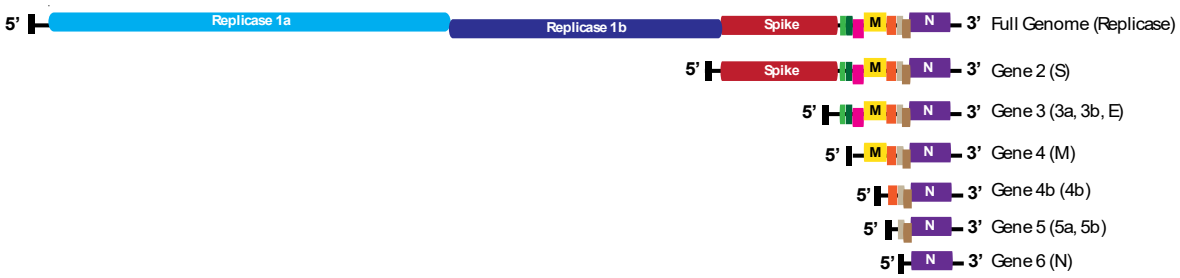


Figure 1.6: Schematic representation of the strategy employed by coronaviruses for sg mRNA expression. *Coronaviruses produce a nested set of sg mRNAs co-terminal to both the 5' and 3' ends of the virus genome. Each transcript contains a leader sequence (shown by a black cap) fused to an mRNA body representing the sequence in the genome from an encoded ORF up to a 3'-poly (A) tail.*

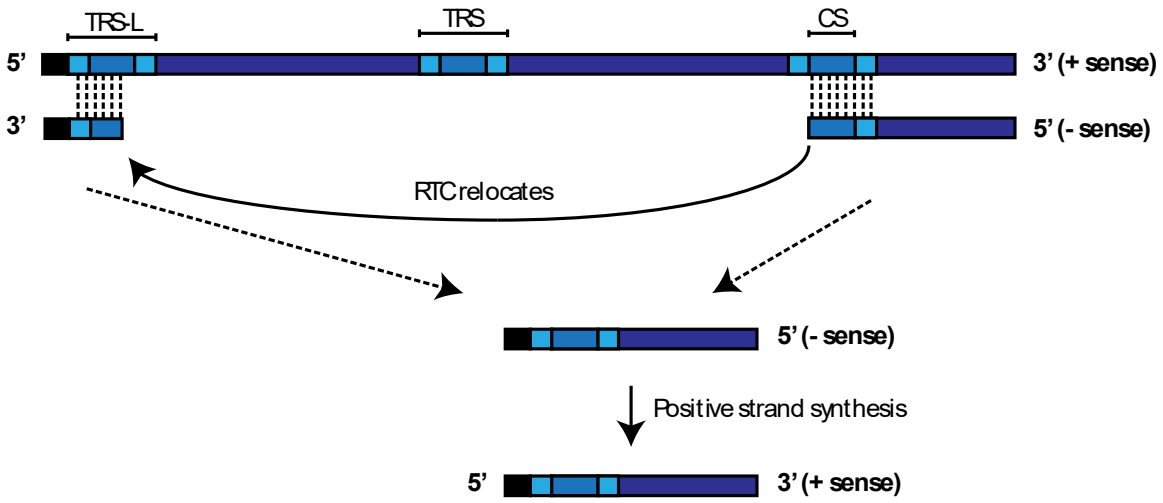


Figure 1.7: Schematic representation of discontinuous sg mRNA synthesis. *The RNA genome possesses a CS presented within a TRS. For IBV, the consensus CS is CTTAACA. During negative strand synthesis, the RTC will synthesize the negative sense copy of the genome from 3' to 5' end. The nascent negative sense RNA will be complementary to the CS presented within the TRS-L and, in some instances, cause the RTC to change template and reposition to the 5' end of the genome. This will hence generate an mRNA possessing the 5' leader sequence and a 3' gene sequence.*

Under the current model full length (replication) and sub-genomic (transcription) mRNA synthesis both utilise the same machinery (i.e. RTC). Firstly, the components of this RTC are recruited to the 3′ end of genomic RNA and negative strand synthesis is initiated. Elongation of the nascent negative strand continues until reaching the first TRS motif where one of two events will occur: either this TRS is disregarded and nascent negative strand synthesis will continue or nascent negative strand synthesis will pause and the RTC will relocate to complete synthesis of the leader sequence at the 5′ end of the genome (Sawicki et al., 2007, Sola et al., 2015). This relocation is made possible by the complementarity between the CS-L and CS-B sequences, the environment generated by both TRS sequences, and proximity of the two TRSs due to RNA-RNA and RNA-protein interactions. Disregarding all TRSs results in a full length negative-strand sequence.

Positive-strand synthesis from the negative-strand sequence for completion of coronavirus replication/transcription is a poorly understood process but it is assumed to utilise the same RTC machinery as negative-strand synthesis. Two distinct peaks of RdRp activity have been identified in MHV (Brayton et al., 1982) suggesting the process of negative versus positive strand synthesis is not incessant. It is possible that the promoter sequence used to initiate positive-strand synthesis is the same used to initiate negative-strand synthesis. The process of positive- versus negative-strand synthesis is not symmetrical as the positive-strand is produced in excess (Sethna et al., 1991). Under the current model it is proposed that nascent positive-sense full-length RNA is distinguishable from positive-sense sg mRNA due to the presence / absence of an as yet undefined secondary structure located at the 5′ end of the genome but outside of leader sequence. Discontinuous sg mRNA synthesis therefore excludes this structure from generated transcripts preventing replication of sg mRNAs.

1.7.4. Assembly and release

Following genome replication and transcription, nascent sg mRNAs are used to generate viral proteins with the S, E and M envelope proteins inserted into the ER. Utilising the host cells' secretory pathway, newly synthesised products pass into the ERGIC. Here the envelope proteins aggregate and enter membranes of the ERGIC. The viral RNP, containing nucleocapsid protein and the viral genome, is then encapsidated by this structure to form a mature virion.

As eluded to previously membrane protein is critically involved in the assembly of virions, primarily due to its ability to form protein-protein interactions. However M protein alone is not sufficient for the virion assembly and requires co-expression of E protein (Vennema et al., 1996, Corse and Machamer, 2000). Supplementary N protein expression further improves virion assembly efficiency but is not essential (Ruch and Machamer, 2011). S protein is not required in virion assembly but is included provided the presence of M protein. In the absence of M protein, S protein accumulates in the Golgi complex. In some coronaviruses non-virion

associated S protein may subsequently transit to the cell surface to promote cell-cell fusion between infected and uninfected cells. This provides a means by which viruses may spread without detection by the immune system.

Alternatively, assembled virions can be transported to the cell surface in vesicles and released by means of exocytosis. It is unclear whether mature virions exiting Golgi employ the cells' own pathway for transport of large cargo or whether a virus unique pathway is used.

1.8. Replication fidelity of RNA viruses and intra-host variability

Viruses are unable to replicate excessively long pieces of RNA as too many deleterious mutations accumulate, and the virus is unable to replicate itself faithfully. Consequentially viral genome size is limited by an apparent error threshold which, if not adhered to, would inevitably lead to the viruses own extinction (Eigen, 1971). The majority of RNA viruses therefore do not have genome sizes exceeding approximately 15 Kb with coronaviruses being a notable exception. Coronaviruses are however unique amongst RNA viruses as they possess proof-reading activity with 3′ to 5′ ExoN encoded by nsp14 (Denison et al., 2011). This presence of ExoN likely offsets the balance of genome size versus error catastrophe.

The evolution of RNA viruses is rapid and, follows the Darwinian principles. Random genetic mutations result in better adapted variants which quickly outcompete their less fit original form. Typically, viruses do not possess mechanisms for correcting potential error and is therefore possible for random mutation to result in changes which are lethal to the virus. As a single virus particle produces potentially millions of progeny virus, the effect of lethal error is mitigated. Most mutations however will be neither provide an immediate advantage nor be lethal and are instead silent. Viruses can also evolve through

reassortment whereby a single cell is infected by two different strains of the same virus. Reassortment can only occur in segmented viruses.

RNA viruses do not exist as clonal structures of a single genotype but rather exist as a population of diverse viruses that are genetically linked through mutations (reviewed by Lauring and Andino (2010)). Separate to Darwinism, it is theorised that these population of viruses can interact on a functional scale (reviewed by Lauring and Andino (2010)) and, in such instances of low fidelity (error-prone) RNA-dependent RNA polymerase and rapid replication, it is not beneficial for the virus to evolve towards the highest fitness sequence space because mutation away from this landscape would lead to a dramatic loss of fitness. It is instead more beneficial for the virus to occupy sequence space where there is a higher average fitness where variants would remain near their fitness optimum after mutation. This is referred to as the quasispecies effect or, colloquially, "survival of the flatness".

It has previously been demonstrated how the diversity of a viral population may account for its viral pathogenicity and exhibit tissue tropism outside what is conventionally achieved by the dominant sequence (Vignuzzi et al., 2006). Here, poliovirus was experimentally to improve RdRp replication fidelity, hence reducing virus mutation rate. These mutant viruses were of lower pathogenicity in comparison to wildtype virus, however upon application of a mutagen pathogenicity could be restored. This has been interpreted that lower mutation rates are associated with a reduced adaptability of the viral population which in turn is potentially associated with a reduced virus severity. Subsequent *in vivo* experiments detected mutant viruses within the brain tissue, beyond the typical scope of poliovirus. Though these viruses were not themselves pathogenic, it was interpreted that there may be complementation between viral population. The combined characteristics of each of the individual variants, referred to as quasispecies, and the interaction between them collectively contribute to the characteristics of the population.

1.8.1. Coronavirus exonuclease (ExoN) mutants

The exonuclease activity (ExoN) of nsp14 in coronaviruses is believed to enhance the fidelity of RNA synthesis by the nsp12 RNA-dependent RNA polymerase and permit coronaviruses to possess genomes of extended length while simultaneously preserving the survival of viral quasispecies. ExoN activity has therefore had a profound effect on the evolution of coronaviruses and as such is extensively studied (reviewed by Ogando et al. (2019)).

Inhibition of ExoN activity in HCoV-229E by modification of the ExoN active site, the first example of such study, is associated with a reduced rate of replication and altered sg mRNA transcript ratio compared to wildtype virus (Minskaia et al., 2006). Crucially, no virus progeny could be isolated from transfected cells indicating that ExoN activity is critical for HCoV-229E replication (Minskaia et al., 2006). An inability for ExoN mutants to generate progeny virus has also proven in transmissible gastroenteritis virus (TGEV) (Becares et al., 2016) and MERS-CoV (Ogando et al., 2019), though intracellular RNA synthesis is impaired to varying degrees.

Viable ExoN knockout viruses have however been produced for MHV (Eckerle et al., 2007) and SARS-CoV (Eckerle et al., 2010). In MHV, this was associated with a reduced overall RNA synthesis, delayed replication and reduced sg mRNA 2 synthesis (Eckerle et al., 2007). Knockout of SARS-CoV ExoN activity was not extensively studied, but was associated with no clear change in replication kinetics (Eckerle et al., 2010). For both viruses ExoN knockout mutants resulted in an accumulation of 21-fold more mutants compared to wildtype when serially passaged on cultured cells (Eckerle et al., 2007, Eckerle et al., 2010).

1.9. High throughput sequencing (HTS)

High throughput sequencing technologies provide the opportunity to study viral genomes at a greater depth and allows for the identification of ultra-low frequency mutations that would otherwise be undetected by conventional Sanger-based sequencing (Domingo et al., 2012). Use of HTS for the dissection intra-host diversity is extensively reported across multiple species of virus (Wright et al., 2011, Eriksson et al., 2008, Margeridon-Thermet et al., 2009). Two sequencing technologies, 454 pyrosequencing and Illumina, have been utilised during this work and will be discussed in more detail.

1.9.1. 454 pyrosequencing

Roche 454 pyrosequencing represents the first HTS technology developed and follows a "sequencing by synthesis" principle whereby the DNA sequence is established during complementary strand synthesis. Here DNA is ligated to an adapter, bound to a DNA-capture bead and amplified using polymerase chain reaction (PCR). Each resulting bead is then placed into an individual well containing DNA polymerase, adenosine triphosphate (ATP) sulfurylase, luciferase and apyrase enzymes and adenosine 5′ phosphosulfate and luciferin substrates (Voelkerding et al., 2009).

Pyrosequencing relies on detecting activity of the DNA polymerase through a luminescent signal. Here A, T, C and G nucleotides are sequentially added and removed from the reaction with inclusion of a dNTP into the complementary strand releasing a pyrophosphate. ATP sulfurylase can then convert pyrophosphate into ATP which in turn facilitates luciferase-mediated luciferin conversion into oxyluciferin. This process results in the generation of a light signal that can be recordable by a camera with the process repeated again for inclusion of the next dNTP. Luminescence can then be associated with the addition of a particular dNTP for determining the DNA sequence (Voelkerding et al., 2009, Goodwin et al., 2016).

With a mean length of 300-500 bp, 454 sequencing marked a decrease in read length versus dideoxy (Sanger) sequencing despite providing a read length comparable to successor HTS technologies. 454 was also limited in that no proofreading capability was possible due to the core premise from which the technology was based. Thus pyrosequencing is not efficient at resolving low-complexity regions (i.e. homopolymeric regions) due the imprecision of recording a brighter light intensity associated with the inclusion of multiple dNTPs at one time (Quince et al., 2009).

This sequencing platform was discontinued in 2014 and support withdrawn in 2016.

1.9.2. Illumina (Solexa)

One of the successor technologies to 454 pyrosequencing was Illumina dye sequencing (formerly Solexa sequencing). In Illumina sequencing adapter sequences are initially ligated to both ends of a purified DNA which allow the DNA to bind to a flow cell coated in complementary oligos. DNA immobilised at one end of the sequence then attaches the opposing, unbound end to the flow cell forming a bridge. Polymerases generate a complementary strand of the bound DNA (which itself becomes bound to the flow cell) and the original DNA is washed away. The complementary strand is then used as a template to regenerate the original DNA sequence with the process repeated until the all oligos on the flow cell are occupied. By this mechanism, all DNA strands in one cluster originate from a single source, a process referred to as clonal amplification (Goodwin et al., 2016).

Sequencing begins after clonal amplification and the removal of all reverse strands. Like pyrosequencing, Illumina follows a “sequencing by synthesis” principle, here using dNTPs with each base bound to a unique fluorescent marker. Unlike pyrosequencing however, all dNTPs are added at the same time to allow for natural competition between each of the four bases. The addition of a nucleotide to the nascent strand reversibly terminates elongation preventing the

additional of multiple nucleotides during one cycle. Laser excitation then results in a recordable florescent emission for determination of the DNA sequence. The chain-terminator is then removed, and the next complementary dNTP allowed to bind. This cycle is repeated until the entirety of the DNA strand has been read with the complementary strand then sequenced by the same method (Goodwin et al., 2016).

Sequencing both the DNA and complementary DNA ensues a level of quality control as any sequence oddities in one strand can be confirmed by its opposing strand. Differences both between forward and reverse strands, and between sequence and the "true" sequence may derive from either the internal polymerase used by the Illumina platform or by PCR amplification prior to library preparation. The introduction of error has been well-characterised (Schirmer et al., 2016) but may ultimately confound the identification of true ultra-low frequency variants. PCR biases may also influence, exaggerate or suppress the frequency of a true variant and a relatively short read length results difficulty in resolving genomes with a high degree of repetition.

Though not without fault Illumina sequencing represents a marked improvement in both the quality and quantity of reads versus 454 pyrosequencing and is one of the most widely and currently used HTS technology to date.

1.10. Infectious bronchitis and infectious bronchitis virus

Avian infectious bronchitis (IB) is a highly contagious respiratory disease of domestic fowl representing one of the most significant threats to the poultry health. The etiological agent of IB, infectious bronchitis virus (IBV), was first isolated in the 1930s (Beach and Schalm, 1936, Beaudette and Hudson, 1937) over 80 years ago and is endemic in all countries practicing intensive poultry production (Ignjatovic and Sapats, 2000). With an estimated 55 billion chickens produced worldwide including 5 billion for egg production, poultry is an important food

source. IBV is therefore a major risk to global food security; it has been estimated that a 10% reduction in IBV would be worth around £654 million to the global poultry industry (Anonymous, 2014).

1.10.1. Pathogenicity and disease

IBV primarily replicates within the upper-respiratory tract of domestic fowl though some strains also replicate in the kidneys and oviduct. Clinical signs of IB include snicking (sneezing), tracheal rales, and nasal/ocular discharge (Cavanagh, 2007); synonyms indistinguishable from other avian-related viruses such as Newcastle Disease Virus (NDV). Pathogenic strains of IBV typically have a 100% morbidity though mortality is generally low. Secondary bacterial infections occurring as a result of infection or nephritis may cause mortality to reach up to 50% (Jackwood, 2012).

IBV has a two-pronged impact on the poultry industry. Firstly broilers (meat-producing birds) have a poorer rate of weight-gain associated with fighting infection ultimately increasing the time in which the bird reaches the market. Secondly infection of the oviducts results in a poorer rate of egg production in layer birds with those eggs produced being typically of poor quality and not suitable for the market (Cook et al., 2012). Recovered layer birds often never reach pre-infection production rates with the losses due to IB infection totalling more than those for IB mortality (Bennett et al., 1999).

1.10.2. Vaccines and vaccination efforts

The need for the prevention and control of IBV has led to the development of several successful vaccines against IBV which largely fall into one of two categories. Inactivated vaccines consist of immunogenic IBV antigens typically from different virus strains and can offer a broad profile of protection against multiple strains at once based on vaccine composition. Such inactivated vaccines

however are typically more expensive to produce versus live-attenuated viruses and are much more labour intensive to administer as each bird must be administered individually (Jordan, 2017) making their use financially inviable for most farmers. Inactivated vaccines are also less efficacious compared to their live-attenuated counterparts and are therefore typically used as a booster vaccination after live-attenuated virus administration (Bande et al., 2015).

Live-attenuated vaccination involves use of a non-pathogenic or procedurally attenuated strain of IBV to provide immunity against antigenically similar IBVs. Live-attenuated viruses are typically cheaper to produce and, as the virus remains infectious, alternative means of delivering vaccination can be utilised. Live-attenuated IBVs can be administered directly by the oculo-nasal route but this, like inactivated vaccine administration, is labour intensive. Alternatively, live-attenuated viruses can be administered en masse by spraying the flock or introducing the vaccine into the drinking water. Such mass vaccination strategies provide a cost-effective means of vaccinating entire flocks and, despite a significant disadvantage being an inability to guarantee administration of sufficient vaccine dose, remains preferentially used by the poultry industry.

Live-attenuated viruses therefore provide the advantage of ease of application and greater protection against disease (Cook et al., 2012). They are not without risk however and include the possibility of vaccine breakdown or reversion to virulence (Ndegwa et al., 2012, Hopkins and Yoder, 1986). Live-attenuated virus also allow for recombination between virulent and attenuated strains potential emergence of new IBV variant viruses (Lee and Jackwood, 2000, Cavanagh et al., 1992). The emergence and disappearance of new pathogenic strains (Meeusen et al., 2007, Bande et al., 2015) combined with poor vaccine cross-protection (Bande et al., 2015) limit the successfulness of IBV vaccine campaigns opposed to the vaccine efficacy itself.

1.10.3. Live-attenuated vaccine generation by serial *in ovo* passaging

The process of generating live-attenuated vaccines against IBV typically involves serial passaging of wild-type isolates of IBV in embryonated specific pathogen free (SPF) hens' eggs and is routinely used by the vaccine industry (Geerligs et al., 2011). *In vitro* serial passaging of IBV for viral attenuation as is extensively reported for other viruses (Beare et al., 1968, Eckels et al., 1984, Badgett et al., 2002) is not viable due to lack of an appropriate cell culture system.

Serial egg-passaging of wild-type IBVs for up to 100 passages results in a loss of IBV pathogenicity (Geerligs et al., 2011). Despite being a well-established protocol, the exact mechanism underlying IBV attenuation by egg-passaging has yet to be determined. Nearly all IBV genes have been implicated in IBV pathogenicity (Phillips et al., 2012, Armesto et al., 2009, Wickramasinghe et al., 2011, van Beurden et al., 2018). The process of attenuation by egg-passage is not guaranteed with it the possibility for viruses to remain virulent at the end of serial-passaging.

1.11. Aims

Infectious bronchitis virus is of significant concern to the poultry industry and to global food production. Vaccination is used as the primary method for controlling the virus with attenuated viruses produced by serial passaging in embryonated hens' eggs. Despite being a well-established practice, it is not clear as to why this method works. This is in large part because the mechanism of IBV attenuation is yet to be established. It is also unclear as to whether attenuation is linked to adaptation of the virus to growth in eggs or by another unknown mechanism.

It has previously been suggested that evolution of IBV is the result of two factors; 1) generation of genetic diversity (e.g. *de novo* mutation) and 2) selection of pre-existing diversity (Toro et al., 2012, Mayr, 1988). Here, given the potential for an RNA virus to mutate it is conceivable that *de novo* mutations occur during egg

passaging that confer attenuation. Conversely, given the complex structure of RNA viruses, it is possible that attenuated viruses pre-exist within a predominately virulent isolate with the process of egg-passaging driving selection of attenuated genotypes. Recombination, as has been reported in the field, is also likely to impact IBV evolution (Lee and Jackwood, 2000, Cavanagh et al., 1992). Understanding the diversity of an IBV population is therefore fundamental to deciphering the mechanisms of IBV attenuation.

Advances in high throughput sequencing technologies have provided an unparalleled opportunity to study intra-host variability. Though HTS studies involving IBV have previously been conducted, these are focused on host immune response (Hamzic et al., 2016) and characterisation of actively circulating field strains (Abolnik, 2015, Reddy et al., 2015). Conversely, characterisation of genomic changes that confer attenuation have been extensively reported but is limited to Sanger sequencing (Ammayappan et al., 2009, Zhao et al., 2014, Phillips et al., 2012). The aim of this thesis is therefore to couple HTS with egg attenuated IBV to establish the key genetic markers of attenuation to enable the rational design of future IBV vaccines. In order to carry out these aims the following objectives were investigated;

Objective 1

To pinpoint which regions of the virus genome undergo the most change in IBVs attenuated by egg passage and to identify any nucleotide/amino acid changes consistent across multiple attenuated replicates.

Objective 2

To characterise the intermediate virus passages by HTS as IBV becomes attenuated by serial egg-passaging and provide evidence as to whether attenuation is result

of selection (for a virus within a virulent isolate) or *de novo* mutation (from the virulent form to achieve attenuation).

Objective 3

To investigate the mechanisms of attenuation using recombinant IBVs generated by reverse genetics.

Chapter 2 – Materials and Methods

2.1. Ethics statement

All animal experimental protocols were performed in accordance with the UK Home Office guidelines and under license for experiments involving regulated procedures on animals protected under the UK Animals (Scientific Procedures) Act 1986. The experiments were performed in the Home Office licensed animal facilities of The Pirbright Institute, formerly the Institute for Animal Health (IAH), Compton, Newbury, United Kingdom, using specific-pathogen-free (SPF) Rhode Island Red (RIR) chickens.

All animal work was a collaborative effort of the Coronaviruses group with assistance from Animal Services team at The Pirbright Institute.

2.2. Cell maintenance and preparation

Unless otherwise stated, all cells were incubated in a 37°C incubator and provided with 5% CO_2.

2.2.1. Continuous cell lines

Continuous cells lines were maintained by the Central Services Unit (CSU) cell culture department of the Pirbright Institute. Provided cells were then seeded according to intended usage.

Vero: a continuous cell line originally isolated from kidney epithelial cells extracted from an African green monkey.

BHK-21: a continuous cell line originally derived from baby hamster kidneys of five unsexed, one-day old hamsters.

2.2.2. Primary animal cells and tissue cultures

Chickens for primary cells and tissues were raised by the Animal Services department at the Pirbright Institute with tissues harvested as desired by CSU. Either primary cells were isolated by CSU cell culture department or tissues provided for *ex vivo* culture. Provided cell suspensions were then seeded according to the intended usage.

Chick kidney cells (CKCs): primary cells prepared by trypsinisation of kidneys dissected from 2- to 3- week old SPF RIR chicks (Hennion and Hill, 2015).

Adult tracheal organ cultures (TOCs): *ex vivo* organ cultures prepared by the sectioning of trachea from 2-3 week old SPF RIR chickens into small rings (Hennion, 2015).

2.2.3. Embryonated eggs

RIRs embryonated eggs were sourced either from Poultry Production Unit (PPU) at the Pirbright Institute (Compton Laboratory) or from the National Avian Research Facility (NARF) in Edinburgh. White Leghorn (referred to as VALOs) embryonated eggs were provided by VALO BioMedia GmbH, Germany.

All eggs were set by Animal Services at room temperature at 37°C and delivered for use at the required gestation point. None were permitted to remain past 14 days post set as per UK Animals (Scientific Procedures) Act 1986 guidelines and were culled by refrigeration at 4°C.

2.3. Media, buffers and other components

2.3.1. Cell culture medium

Where possible, all media components were purchased from Sigma (Merck). Type 1 water was provided by the CSU media services unit of the Pirbright Institute (see Type 1 water preparation, 2.3.4). All media was prepared under aseptic conditions using a Class II Microbiological Safety Cabinet (MBSC).

Table 2.1 Minimum Essential Medium Eagle (1x) for maintenance of Vero cells

INGREDIENT	VOLUME (mL)	FINAL CONCENTRATION
MINIMUM ESSENTIAL MEDIUM EAGLE (SIGMA, M2279)	500	0.9 X
FOETAL BOVINE SERUM (SIGMA, F0926)	50	9 %
L-GLUTAMINE, 200 mM (SIGMA, G7531)	5	1.8 mM

Table 2.2 Minimum Essential Medium Eagle (1x) for infection of Vero cells

INGREDIENT	VOLUME (mL)	FINAL CONCENTRATION
MINIMUM ESSENTIAL MEDIUM EAGLE	500	0.99 X
L-GLUTAMINE, 200 mM	5	1.98 mM

Table 2.3 Minimum Essential Medium Eagle (2x) for overlay of infected Vero cells

INGREDIENT	VOLUME (mL)	FINAL CONCENTRATION
10X MINIMUM ESSENTIAL MEDIUM EAGLE (SIGMA, M0275)	100	2 x
FOETAL BOVINE SERUM	50	10 %
SODIUM BICARBONATE, 7.5% (SIGMA, S8761)	23	0.345 %
L-GLUTAMINE, 200 mM (SIGMA, G7531)	10	4 mM
PENICILLIN-STREPTOMYCIN, 10,000 U/mL (GIBCO, 15140122)	1	10,000 U
NYSTATIN SUSPENSION, 10,000 U/mL (SIGMA, N1638)	2.5	25,000 U
WATER, TYPE 1	To 500 mL	-

Table 2.4 N,N-Bis(2-hydroxyethyl)-2-aminoethanesulfonic acid (BES) containing medium (1x) for infections of CKCs and Vero cells

INGREDIENT	VOLUME (mL)	FINAL CONCENTRATION
10 X MINIMUM ESSENTIAL MEDIUM EAGLE	50	1 x
TRYPTOSE PHOSPHATE BROTH (SIGMA, T8159)	50	10 %
BOVINE SERUM ALBUMIN, 10% W/V FILTER-STERILISED (SIGMA, A9418)	10	0.2 %
BES,1M FILTER STERILISED (SIGMA, B9879)	10	0.02 M
SODIUM BICARBONATE, 7.5%	14	0.21 %
L-GLUTAMINE, 200 mM	5	2 mM
PENICILLIN-STREPTOMYCIN, 10,000 U/mL	1	10,000 U
NYSTATIN SUSPENSION, 10,000 U/mL	2.5	25,000 U
WATER, TYPE 1	To 500 mL	-

Table 2.5 BES-containing medium (2x) for overlay of infected CKCs

INGREDIENT	VOLUME (mL)	FINAL CONCENTRATION
10 X MINIMUM ESSENTIAL MEDIUM EAGLE	100	2 x
TRYPTOSE PHOSPHATE BROTH	100	20 %
BOVINE SERUM ALBUMIN, 10% W/V FILTER-STERILISED	20	0.4 %
BES,1M FILTER STERILISED	20	0.04 M
SODIUM BICARBONATE, 7.5%	28	0.42 %
L-GLUTAMINE, 200 mM	10	4 mM
PENICILLIN-STREPTOMYCIN, 10,000 U/mL	1	10, 000 U
NYSTATIN SUSPENSION, 10,000 U/mL	2.5	25, 000 U
WATER, TYPE 1	To 500 mL	-

Table 2.6 Glasgow Minimum Essential Medium (1x) for culturing BHK-21 cells

INGREDIENT	VOLUME (mL)	FINAL CONCENTRATION
GLASGOW MINIMUM ESSENTIAL MEDIUM (SIGMA, G5154)	500	0.76 x
FOETAL BOVINE SERUM	50	7.6%
TRYPTOSE PHOSPHATE BROTH	100	0.15 X
L-GLUTAMINE, 200 mM	5	1.53 mM

Table 2.7 Chick kidney cell (CKC) growth medium

INGREDIENT	VOLUME (ML)	FINAL CONCENTRATION
MINIMUM ESSENTIAL MEDIUM EAGLE	500	0.8 x
NEWBORN CALF SERUM (SIGMA, N4762)	55	8.8 %
TRYPTOSE PHOSPHATE BROTH	55	8.8 %
4-(2-HYDROXYETHYL)PIPERAZINE-1-ETHANESULFONIC ACID, N-(2-HYDROXYETHYL)PIPERAZINE-N′-(2-ETHANESULFONIC ACID) SOLUTION, 1M (GIBCO, 15630080)	5.5	8.8 mM
PENICILLIN-STREPTOMYCIN, 10,000 U/mL	1	10, 000 U
NYSTATIN SUSPENSION, 10,000 U/mL	2.5	25, 000 U
L-GLUTAMINE, 200 mM	5	1.6 mM

Table 2.8 Tracheal organ culture (TOC) maintenance and infection medium

INGREDIENT	VOLUME (ML)	FINAL CONCENTRATION
MINIMUM ESSENTIAL MEDIUM EAGLE	500	0.9 X
4-(2-HYDROXYETHYL)PIPERAZINE-1-ETHANESULFONIC ACID, N-(2-HYDROXYETHYL)PIPERAZINE-N′-(2-ETHANESULFONIC ACID) SOLUTION (, 1M	20	37.8 mM
PENICILLIN-STREPTOMYCIN, 10,000 U/ML	1	10, 000 U
NYSTATIN SUSPENSION, 10,000 U/ML	2.5	25, 000 U
L-GLUTAMINE, 200 mM	5	1.9 mM

2.3.2. Phosphate buffered saline 'a' (PBSa) preparation

Prior to usage adherent cells and tracheal organ cultures were washed at least once with PBSa prepared by CSU at the Pirbright Institute according to the following methods.

Components listed in Table 2.9 were combined in an appropriately sized vessel.

Table 2.9 Components used in PBSa preparation

INGREDIENT	UNIT
PBSa POWDER (SIGMA, 56064C), CONSISTING OF; • Potassium chloride, 0.2 g/L • Monopotassium phosphate, 0.2 g/L • Sodium chloride, 8.0 g/L • Dipotassium phosphate (Anhydrous), 1.15 g/L	9.55 g/L
WATER, TYPE 1	Up to 1 L

For quality control purposes pH and conductivity readings were taken. In the event pH was greater than 7.50, 0.4 mL 1M stabilised hydrochloric acid was added for an acceptable pH range of 7.20-7.50. The resulting volume was dispensed into appropriate volumes and autoclaved at 121°C for 20 minutes for sterilisation.

2.3.3. Phosphate buffered saline (PBS) preparation

PBS was used for all animal related work including the dilution of virus for bird inoculation, washing of trachea for assessment of ciliary activity and the storage of harvested tissues at -80°C. PBS was prepared by CSU at the Pirbright Institute according to the following method.

Components listed in Table 2.10 were combined in an appropriately sized vessel.

Table 2.10 Components used in PBS preparation

INGREDIENT	UNIT
DULBECCO'S PHOSPHATE BUFFERED SALINE (SIGMA, D5773)	9.55 g/L
CALCIUM CHLORIDE (VWR, 22317-260)	0.132 g/L
WATER, TYPE 1	Up to 1 L

For quality control pH and conductivity readings were taken with an acceptable range of 7.30 – 7.60 and 14,000 – 17,800 µs/cm, respectively. The resulting volume was dispensed into appropriate volumes and sterilised by filtration.

2.3.4. Type 1 endotoxin-free water preparation

Type 1 endotoxin-free water for the preparation of all media outlined above was prepared by CSU at the Pirbright Institute. Water was drawn from a Milli-Q system, dispensed into appropriate volumes, and sterilised by autoclaving at 121°C for 20 minutes.

2.4. Viruses

2.4.1. IBVs

- **M41-CK:** a pathogenic strain of IBV belonging to the Massachusetts serotype lineage GI-1 (as classified by (Valastro et al., 2016)). M41-CK has been adapted for growth on primary CKCs but is not capable of infecting Vero cells (Cook et al., 1976, Darbyshire et al., 1979).
- **M41-K:** the molecular clone of M41-CK, derived from a rescued rIBV generated using the vaccinia reverse genetic system. M41-K contains a total of nine known synonymous marker mutations added to distinguish the difference between M41-CK and M41-K at a Sanger sequencing level.
- **Beau-CK:** an apathogenic strain of IBV also belonging to Massachusetts serotype, genotype GI-1 (as classified by (Valastro et al., 2016)). Unlike M41-CK, Beau-CK has an expanded host tropism and possesses the ability to infect Vero and BHK-21 cells in addition to CKCs (Cavanagh et al., 1986).
- **Beau-R:** the molecular clone of Beau-CK. Like M41-K, this is derived from a rescued IBV generated using reverse genetic system (Casais et al., 2001). Beau-R contains two synonymous marker mutations added to distinguish the difference between Beau-CK and Beau-R by Sanger sequencing. The Beau-R sequence is available at NCBI (accession number AJ311317.1). Relative to the Beau-R sequence, Beau-CK differs at position C19666T and A27087G.
- **BeauR-rep-M41K-struct**: a chimeric rIBV possessing the replication machinery (i.e. Replicase 1a/1ab) of Beau-R but the structural and accessory proteins of M41 (Armesto et al., 2009). As a result of this recombination, BeauR-rep-M41K-struct also possesses the 5' UTR of Beau-R and the 3' UTR of M41-K.
- **H120:** a live attenuated vaccine strain of Massachusetts type generated by passaging an IBV originally isolated in the 1950s 120 times in embryonated eggs (Bijlenga et al., 2004).

2.4.2. Fowlpox virus (FPV)

- **rFPV-T7:** a recombinant fowlpox virus (rFPV) that expresses the bacteriophage T7 RNA polymerase under the control of a vaccinia virus P7.5 promoter (Britton et al., 1996).

2.4.3. Vaccinia virus (VV)

- **rVV-M41-K**: an rVV possessing the IBV M41-CK cDNA genome with 9 known synonymous marker mutations. This was used for the rescue of a molecular clone of IBV (M41-K) and as a receiver rVV for reverse genetics.

2.5. Virus-based techniques

2.5.1. Generation of stock IBV using embryonated eggs

Working stocks of IBV were generated by infection of 9-11 day old embryonated SPF chicken eggs of either RIR or VALO variety. Parent virus was diluted either 1 in 1,000 or 1 in 10,000 using PBSa or 1 x BES-containing medium. 100 µL diluted virus was inoculated into the allanotic cavity of viable eggs, infecting multiple eggs with the same virus as required.

After 24-hour incubation, embryos were candled and culled by refrigeration for a minimum of four hours. Allanotic fluid was harvested, pooled and centrifuged at 700 x *g* for five minutes to clarify supernatant. Resulting virus-containing supernatant was stored at -80°C.

All stock viruses were spot sequenced to ensure the correct virus was grown. Stocks were titrated three times on CKCs to determine virus titre.

2.5.2. Plaque assay

Serial dilutions of 100 µL virus + 900 µL 1 x BES-containing medium were set up until a dilution of 10^{-6} was achieved. Confluent 12-well plates of CKCs washed once with PBSa were then infected with 250 µL diluted virus in triplicate, infecting each plate with four dilutions of virus (10^{-3} – 10^{-6}). After one hour incubation virus inoculum was removed and 2 mL overlay applied to each cell. This overlay consists of 2% agar (molten) + 2 x BES-containing medium in equal volumes, achieving a 1:1 ratio when mixed.

After a further three-day incubation, 1 mL PBSa + 9% formaldehyde was added to each well for fixing. Cells were allowed to fix for 1 hour, formaldehyde-containing PBSa and agar overlay were removed and replaced with 1 mL 0.1% crystal violet per well. After 15 minutes, crystal violet was removed and plates allowed to dry. Fixing and staining of cells was performed at room temperature.

Virus titre was calculated as number of plaque forming units (PFU) per mL using the virus dilution with the highest number of clearly distinguishable plaques. In instances where it was not possible to clearly distinguish the number of plaques per well, the range of dilutions used to infect cells was adjusted accordingly.

2.5.3. Time-course infection

IBVs for each series of growth curves were diluted to same concentration (PFU / mL) using 1 x BES-containing medium. Confluent 6-well plates of CKCs were washed once using PBSa then infected with 500 µL virus per well.

After infection for 1 hour, virus inoculum was removed and cells washed once with PBSa. 3 mL 1 x BES-containing medium was applied to each well. For the 1 hour post infection time point, this medium (supernatant) was immediately removed

and stored at -80°C. Sampling proceeded every 24 hours after infection with supernatant removed and stored at -80°C.

Each growth curve was performed a minimum of three times. Titres of each replicate were determined by an individual plaque assay with the final titre determined as the mean average of these values. Where points on the growth curve were considered outliers, the growth curve and titrations were repeated as necessary.

2.5.4. Assessment of pathogenicity *in vivo*

The IBV-associated clinical signs used to determine pathogenicity were snicking, tracheal rales (a sound emanating from the bronchi, also detected by vibrations when holding a chick), and ciliary activity of the trachea. Chicks were observed daily between days 3 and 7 post-inoculation. Snicks were counted by a minimum of two persons over 2 min periods. Birds were checked individually for the presence of tracheal rales. Tracheas were removed from three randomly selected chickens from each group on days 4 and 6 post-infection for assessment of ciliary activity. Ten 1mm sections were cut from three different regions of each trachea (three sections from the proximal, four from the middle, and three sections from the distal region). The level of ciliostasis of each tracheal section was determined by light microscopy; the assessor was blinded to the experimental group. Remaining birds were euthanized by a schedule one method on day 7 post-inoculation and tissues harvested post-mortem.

2.6. RNA-based methods

Extracted RNA was stored at either -20°C or -80°C for short- or long-term storage respectively.

2.6.1. Homogenisation of animal tissues for RNA extraction

Animal tissues were placed in a 2 mL Eppendorf tube containing a 5 mm stainless steel bead and 500 µL PBS. Samples were then homogenised using a Qiagen TissueLyser II for two minutes, 25 Hz. Resulting homogenate was clarified by centrifugation and 100 µL used for RNA extraction using a QIAGEN RNeasy kit (QIAGEN, 74104). The "Purification of Total RNA from Animal Cells using Spin Technology" manufacturer's protocol was followed including the optional second spin to eliminate RPE buffer carryover (step 9). Samples were eluted in 33 µL RNase-free water.

Homogenisation was repeated as necessary in instances where tissue break down was not complete.

2.6.2. Purification of total RNA from cells

For total RNA isolation from IBV infected cells, supernatant was removed from infected plates and cells washed with PBS. Lysis buffer (Buffer RLT) was then applied directly to each cell. RNA extractions were performed using a QIAGEN RNeasy kit following the "Purification of Total RNA from Animal Cells using Spin Technology" protocol following the optional second spin to eliminate RPE buffer carryover. Samples were eluted in 33 µL RNase-free water.

2.6.3. Purification of RNA from allantoic fluid and cell supernatant

RNA isolation from allanotic fluid and cell supernatant was performed using a QIAGEN RNeasy kit following "RNA Cleanup" protocol. In contrast to the manufacturer's protocol, an initial volume of 170 µL supernatant was used. Buffer RLT and ethanol were added to each sample according to the manufacturer's instructions with the entire 770 µL spun through the column. RNA extraction then proceeded according to manufacturer's instructions, including optional second spin for eliminating RPE buffer carryover. Samples were eluted in 33 µL RNase-free water.

2.6.4. Reverse transcription

Superscript IV Reverse Transcriptase (Invitrogen, 18090050) was used for the reverse transcription of IBV RNA to complementary DNA (cDNA). Where possible, a master mix of components common to all reactions was prepared. The procedure as outlined below follows the manufacturer's instructions.

Components listed in Table 2.11 were combined to create RNA-primer binding reaction.

Table 2.11. Components of reverse transcription, RNA-primer binding reaction

COMPONENT	VOLUME (µL)	FINAL CONCENTRATION (IN 20 µL)
50 µM RANDOM PRIMERS (5'- GTTTCCCAGTCACGATCNNNNNNNNNNNNNNN – 3')	1	2.5 µM
10 mM dNTP MIX (10 mM EACH) (INVITROGEN, R0182)	1	0.5 mM each
TEMPLATE RNA	5	-
MOLECULAR GRADE, NUCLEASE-FREE WATER (SIGMA, W4502)	6	N/A

RNA-primer mix was heated at 65°C for 5 minutes then cooled at 4°C (or placed on ice) for at least 1 minute. 7 µL reverse transcription reaction (Table 2.12) was added to each tube (for a total volume of 20 µL).

Table 2.12. Components for reverse transcription, reverse transcription reaction

COMPONENT	VOLUME (µL)	FINAL CONCENTRATION (IN 20 µL)
5 X SUPERSCRIPT IV BUFFER	4	1 x
100 mM DTT	1	5 mM
RNASEOUT™ RECOMBINANT RNASE INHIBITOR (INVITROGEN, 10777019)	1	2.0 U / µL
SUPERSCRIPT® IV REVERSE TRANSCRIPTASE (200 U/µL)	1	10 U/ µL

Reactions were incubated at 23°C for 10 minutes, 55°C for 10 minutes and 80°C for 10 minutes. Resulting cDNA was used immediately or stored at -20°C.

2.6.5. Northern blot

mRNA was firstly isolated from total RNA using Poly(A)Purist™ MAG Kit (Invitrogen, AM1922) following the manufacturer's instructions. Northern blot analysis was carried out using a NothernMax-Gly kit (Invitrogen, AM1946).

Briefly, after denaturation at 50°C with an equal volume of glyoxal loading dye for 30 min, mRNA was separated on a 1.1% low-electroendosmosis (LE) agarose gel. The mRNA was transferred to a BrightStar-Plus positively charged nylon membrane via capillary action for 90 min and cross-linked by UV exposure (auto cross link function, Stratalinker UV Cross-linker, Stratagene). The membrane was pre-hybridised with ULTRAhyb buffer for 30 min at 42°C before overnight

incubation at 42°C with a DNA probe specific for the 3′ end of the genome (forward primer, 5′-CAACAGCGCCCAAAGAAG-3′ within the N gene and reverse primer located in the 3′ UTR, 5′-GCTCTAACTCTATACTAGCCT-3′). The membrane was washed and the blot developed using a BrightStar Biodetect (Invitogen, AM1930) kit following the manufacturer′ s protocol.

All Northern blots were performed by Sarah Keep (Pirbright Institute)

2.7. DNA-based methods

2.7.1. Restriction endonuclease reactions (Restriction digest)

Restriction digests were performed using New England Biolabs (NEB) enzymes utilising the NEB recommended reaction criteria and run protocol.

Table 2.13. Components of restriction endonuclease reaction

COMPONENT	VOLUME	FINAL CONCENTRATION
RESTRICTION ENZYME (10,000 U/mL, GENERALLY 1 µL)	1 µL	0.2 U / µL
DNA	1 µg	-
10X REACTION BUFFER (ENZYME DEPENDENT)	5 µL	1 x
MOLECULAR GRADE, NUCLEASE-FREE WATER	To a reaction volume of 50 µL	N/A

Reactions were incubated in a thermal cycler at the required temperature (as per the enzyme specifications) for a period of one hour. To inactivate the enzyme, samples were heated to 65°C a period of 20 minutes.

2.7.2. Polymerase chain reaction (PCR) amplification of DNA fragments

Polymerase chain reaction (PCR) was used for the amplification of specific IBV cDNA or VV DNA sequences. Recombinant *Taq* DNA polymerase (Invitrogen, 10342053) was used. Where possible, a master mix of components common to all reactions was prepared according to Table 2.14 (per reaction).

Table 2.14. Components of PCR reaction

COMPONENT	VOLUME (µL)	FINAL CONCENTRATION
MOLECULAR GRADE, NUCLEASE-FREE WATER (SIGMA)	34.5	N/A
10X PCR BUFFER, MINUS Mg	5	1 x
50 mM $MgCL_2$	2	2 mM
10 mM dNTP MIX (10 mM EACH)	1	0.2 mM (each)
10 µM FORWARD PRIMER (FWD)	1	0.2 µM
10 µM REVERSE PRIMER (REV)	1	0.2 µM
TAQ DNA POLYMERASE (5U / µL)	0.5	0.05 U / µL
TEMPLATE DNA	5	

Table 2.15. Primers for PCR

NAME	SEQUENCE (5' -> 3')	DIRECTION
BG-56	CAACAGCGCCCAAAGAAG	FWD
93/100	GCTCTAACTCTATACTAGCCT	REV
M20	GGAATGGGCATAATAAGG	FWD
M23	CACTGCTACCCCGTACCCG	REV
M50	GGCTGATGAAATGGCTCAC	FWD
LEADER-1	CTATTACACTAGCCTTGCGC	FWD
M13 FORWARD	GTAAAACGACGGCCAG	FWD
M13 REVERSE	CAGGAAACAGCTATGAC	REV
GPT FORWARD	ATGAGCGAAAAATACATCGTC	FWD
GPT REVERSE	TTAGCGACCGGAGATTGGC	REV
BG-113	GCATTGCTCTATCACACTTAG	REV

Reactions were incubated in a thermal cycler with heated lid enabled as per manufacturer's instructions (Table 2.16).

Table 2.16. Thermocycler conditions for PCR

STEP		TEMPERATURE (°C)	TIME
INITIAL DENATURATION		94	3 minutes
25 – 35 PCR CYCLES	Denature	94	30 seconds
	Anneal	~55 (depending on primer T_m)	30 seconds
	Extend	72	90 seconds per kb
FINAL EXTENSION		72	10 minutes
HOLD		4	Indefinitely

2.7.3. Agarose gel electrophoresis

DNA fragments produced by PCR were separated by agarose gel electrophoresis prepared by dissolving 1 % agarose (Sigma, A9539) in 1 x Tris/Borate/EDTA (TBE) buffer (Invitrogen, 15581044). SYBR™ Safe Stain (Invitrogen, S33102) was added at a ratio of 1:10,000 gel volume prior to casting.

Gels were run using a Galileo Bioscience 9 x 11 cm submarine gel apparatus filled with 1 x TBE buffer. Resulting gels were visualised using the ethidium bromide application of a BioRad Gel Doc EZ imager. 1 µg 1 Kb plus DNA ladder (Invitrogen, 10787018) was run in lanes adjacent to samples to approximate DNA band sizes.

2.7.4. Pulsed-field gel electrophoresis

Larger DNA fragments, predominately SalI-HF® (New England Biolabs, R3138) digested vaccinia virus, were separated by pulsed-field gel electrophoresis (PFGE). 2.1 L of 0.5 x TBE buffer was made; 100 mL of which was used to prepare

a 0.8 % agarose gel. The remaining 2 L was to fill a BioRad CHEF-DR II system tank with pump and cooling unit switched on and set to 4°C. CHEF-DR II controller unit was programmed with the following run conditions; Initial pulse time = 0.1s, final pulse time = 1.0s, voltage = 6.0 V / cm, duration 16 hours. To approximate size of fragments CHEF DNA Size Standards 8-48 Kb marker was run alongside the samples (Bio-Rad Laboratories, 170-3707).

After program completion the gel was stained in sterile water containing 0.1 g / mL ethidium bromide ($C_{21}H_{20}BrN_3$, BioRad, 1610433) at room temperature for 30 minutes with gentle agitation. Excess stain was removed by washing in sterile water for 30 minutes with gentle agitation. The gel was visualised using a Bio-Rad Gel Doc EZ following manufacturer's recommendations.

2.7.5. Cloning for sequencing (PCR cloning)

In instances where sequencing could not be performed on the immediate PCR product, for example in the event of a mixed PCR population, PCR cloning was performed. For this the TOPO™ TA Cloning™ Kit for Sequencing, with One Shot™ MAX Efficiency™ DH5α-T1R E. coli (Invitrogen, 12297016) was used with manufacturer's instructions followed.

Briefly, PCR products were immediately used in a TOPO cloning reaction as per Table 2.17.

Table 2.17. TOPO cloning reaction

COMPONENT	VOLUME (µL)
FRESH PCR PRODUCT	4 µL
SALT SOLUTION (PROVIDED IN KIT)	1 µL
PCR-4-TOPO VECTOR (PROVIDED IN KIT)	1 µL

The cloning reaction was incubated at room temperature for a total of 30 minutes. After incubation, 2 µL of cloning reaction was added to the DH5α chemically competent cell tube and incubated on ice for a total of 30 minutes. Cells were then heat shocked at 42°C for 30 seconds. 250 µL room temperature Super Optimal broth with Catabolite repression (S.O.C, Invitrogen, 15544034) medium was added to heat shocked cells then incubated at 37°C for 1 hour, 200 rpm. Each transformation when then spread out on a pre-warmed selective Luria broth (LB) agar plate containing ampicillin (Amp, 100 mg /mL) and incubated overnight at 37°C in a stationary incubator.

To screen resulting colonies each colony was firstly picked using a pipette tip and dabbed onto a second LB + Amp plate to preserve the original colony. The pipette tip was then dipped into a PCR reaction, modified to include a total of 38.5 µL nuclease-free water. The reaction was incubated as standard with an initial denaturation time increased to 10 minutes to ensure cell degradation. M13 forward and M13 reverse primers at an annealing temperature of 55°C was used.

Sequencing was then performed using M13 forward and reverse primers.

2.7.6. Sanger sequencing using GATC services

LIGHTRUN GATC (Eurofins) services was used for determining the order of nucleotides in plasmids and DNA fragments produced by PCR. Components listed in Table 2.18 were combined in a 1.5 mL Eppendorf.

Table 2.18. Components for GATC LIGHTRUN sequencing

COMPONENT	VOLUME (µL)
MOLECULAR GRADE, NUCLEASE-FREE WATER (SIGMA)	7
10 µM PRIMER	2.5
TEMPLATE DNA	0.5

Where possible, template DNA was sequenced using both a forward and reverse primer in two separate reactions.

2.8. High throughput sequencing

Prior to all methods of high-throughput sequencing, virus was purified by ultracentrifugation. RNA extraction was then performed immediately on the resulting pellet.

2.8.1. IBV purification by ultracentrifugation

Virus purification was performed as previously published (Keep et al., 2015). Briefly, allanotic fluid was centrifuged at 1,150 x *g* for 10 minutes using a benchtop centrifuge to clarify. Resulting supernatant was layered on top of 10 mL 30 % sucrose solution and centrifuged for 4 hours at 102,400 x *g* at 4°C. After centrifugation, supernatant was removed and inside of the tube dried, ensuring the pellet was not disturbed.

The purification of IBV for 454 pyrosequencing or Illumina sequencing was performed by Sarah Keep or Graham Freimanis respectively.

2.8.2. 454 pyrosequencing

If the sample was to be sequenced by 454 pyrosequencing, the pellet produced as a result of 2.8.1 was immediately suspended in 1 mL TRIzol reagent. RNA was extracted from supernatant as per the manufacturer's instructions and quantified

by NanoDrop assay. Samples were then sent to the Centre of Genomic Research (CGR), University of Liverpool, UK, for library preparation and sequencing. cDNA libraries were generated according to the *GS FLX Titanium cDNA Rapid Library Preparation Method Manual* and sequenced using a GS FLX with Titanium Series and run protocol (Roche, 2008).

2.8.3. Illumina HiSeq/MiSeq sequencing

If the sample was to be sequenced on the Illumina HiSeq or MiSeq platforms, the pellet produced as a result of 2.8.1 was suspended in 300 µL RLT buffer and total RNA extracted using QIAGEN RNeasy kit. Total RNA was checked for quality using a Bioanalyser 2100 RNA Pico kit (Agilent) and quantitated using a Qubit (Life Technologies). Samples were normalised for a 100 ng input and sequencing preparation performed using one of the three following kits; 1) NEBNext® Ultra™ II Directional RNA Library Prep kit, 2) NEBNext® Ultra™ II RNA Library Prep kit, 3) NEBNext® Ultra™ I Directional RNA Library Prep kit. Library QC was performed using the Bioanalyser 2100 DNA 1000 kit and Qubit.

Library preparation for samples was performed in part by the Pirbright High Throughput Sequencing unit. Illumina sequencing was performed either at Centre of Genomic Research (CGR), University of Liverpool for samples sequenced on Illumina HiSeq, or at Pirbright Institute for samples sequenced on Illumina MiSeq. All samples sequenced using Illumina MiSeq were done so in duplicate.

2.9. Bioinformatic methods and approaches

2.9.1. Quality control and read processing

All reads (454, HiSeq or MiSeq) were subject to quality filtering prior to downstream usage. Read quality assessments (for determining appropriate filters) were performed using FastQC (Andrews, 2010).

For 454 reads, matching FASTA and QUAL files were consolidated to FASTQ format for each virus. Reads were quality trimmed using PRINSEQ v0.20.4 (Schmieder and Edwards, 2011) with the following parameters: minimum length = 100, minimum quality mean = 25, maximum Ns = 0, and trim quality right = 25.

For Illumina HiSeq, reads quality filtering was performed using TrimGalore! (Krueger) using the following parameters; minimum read length = 100, quality score cut off = 30, trim 15 bp (base pairs) from the 5′ end of R1 and R2 read. Reads were inputted in paired-end mode with automatic adapter sequenced recognition.

For Illumina MiSeq, reads quality filtering was performed as above (Illumina HiSeq reads) with the following changes; minimum read length = 150, trim 45 bp from R2 read. This 45 bp removal was included to avoid potential errors as a severe drop in the quality of the R2 read at 3′ end was observed.

2.9.2. Short read alignments

Alignment of NGS reads to a reference genome was performed using one of three programs.

- Bowtie2 v2.2.9 (Langmead and Salzberg, 2012) was used for the entire 454 analysis (Chapter 3, single-end mode) and in the generation of M41-CK, M41-K and Beau-CK Illumina reference sequences (Chapters 4 and 5).
- BWA-MEM (Li and Durbin, 2009) was used for mainstream read alignments to their reference genome (Chapter 4 and 5).
- BBMap, a splice-site global aligner (Bushnell, 2014) was used to assess chimeric reads and potential IBV transcripts (Chapter 5).

All alignments were converted from a SAM (Simple Alignment/Map) file to a sorted and indexed BAM (Binary Alignment/Map) file using samtools1.2 (Li et al., 2009)

for downstream applications. Coverage was reported using the samtools depth function of samtools1.4, reporting all positions regardless of coverage.

2.9.3. Consensus-based reference sequence generation

Consensus-based IBV reference sequences were generated by alignment of quality-filtered reads to representative sequence availables on NCBI. Utilising DiversiUtils, a Java-based script provided by Richard Orton, a consensus sequence was generated from this alignment where minimum coverage > 10 reads. At positions where coverage was below this minimum threshold, an ambiguity code, "N", was placed instead.

2.9.4. Non-consensus based (*de novo*) reference sequence generation

Where the generated dataset allowed (based on coverage against genome), a reference sequence was generated by *de novo* methods. To avoid generating host-derived sequences, host subtraction was performed by aligning reads to the galGal6 reference genome (GRCg6a, GCA_000002315.5) (Zerbino et al., 2017). Non-aligned (i.e. non-host) reads were exported and subsequently inputted in paired-end format to SPAdes v3.10.0 (Bankevich et al., 2012) using default parameters.

To facilitate the generation of contigs with increased lengths, multiple replicates of *de novo* assemblies were performed using SPAdes using randomly selected subsections of reads. For each subsection, resulting contigs were aligned to a representative genome and consensus sequence exported using DiversiUtils. Due to the coverage profile of the dataset, no internal ambiguity codes were produced. The 5′ and 3′ extremities were trimmed until all replicates read the same sequence.

Reference sequences generated using either method were annotated accordingly (Gorbalenya et al., 1989, Ziebuhr et al., 2000, UniProt, 2019), where regions of highly similar sequences were identified. All PQF (pass-quality filter) reads (including non-IBV) were subsequently aligned to the new sequence with visual inspection of the alignment in UGENE v1.31.1 (Okonechnikov et al., 2012) and retrospective in variant-calling (i.e. no called variant was called over 50% frequency) confirming sequence accuracy.

2.9.5. Variant calling using Lofreq*

The identification of single nucleotide variant (SNV) and indels was performed using Lofreq* (version 2.1.2) (Wilm et al., 2012). In order to facilitate indel-calling of the Illumina generated datasets, reads were pre-processed by viterbi realignment and indel qualities inserted into BAM files. Both processes are included as part of the Lofreq* package. Indel calling was not performed on 454 reads due to concerns of homopolymeric error and hence did not require any pre-processing. Variant calling was performed using Lofreq* with default parameters and indel calling switched on where appropriate.

All Illumina MiSeq sequencing was performed in duplicate to avoid potential sequencing error or biases, hence two VCF (Variant Call Format) files were generated for some samples. A custom script based on BCFtools v1.3.1 was written to produce two VCF files possessing the variants common to both. Hence, only variants common to both replicates were used for downstream analysis with the mean frequency of individual variants across two files reported.

2.9.6. Coovar and determining variant impact

Coovar v0.07 (Vergara et al., 2012) was used to assess the impact (i.e. nonsynonymous / synonymous) of called variants. Based on sequence annotation, a GTF (General Transfer Format) file was generated for each reference sequence. Each VCF file (possessing variants common to both replicates) was inputted into Coovar along with its reference sequence and annotations. The output (including

the coding effect of each mutation) was exported into Microsoft Excel for downstream analysis.

2.9.7. Identifying intersecting variants across multiple files

To identify variants occurring across multiple sets (samples), firstly VCF files were harvested for four pieces of information; variant frequency, reference base, position, alternate (variant base). Variant frequency was used to determine whether a variant occurred above or below consensus (50%) level, whereas reference, position and alternate base were used to generate a tab-delimited file of variants for each sample. Hence two files were generated per sample, one possessing variants above 50% and the other below 50%.

For comparisons of less than six sets, the Bioinformatics and Evolutionary Genomics webtool was used and outputs (including intersection and Venn diagram) exported. Comparisons of six or greater were performed using UpSetR (Conway et al., 2017) to generate a more user-friendly schematic.

2.9.8. Multiple sequence alignments

Multiple sequence alignments (MSA) were performed using MAFFT v7.310 (Katoh and Standley, 2013) as part of the built-in functionality of UGENE (Okonechnikov et al., 2012).

2.9.9. Protein structure visualisation

Protein structures were downloaded from RCSB PDB (Berman et al., 2000) and visualised using Chimera v 1.12 (Pettersen et al., 2004).

2.9.10. Graph and statistical analysis

Unless mentioned by special reference, all graphs and statistical comparisons were performed on or generated by GraphPad Prism (version number 7 or 8).

2.10. Reverse genetic system and rescue of recombinant IBVs

The reverse genetics system for IBV was originally published in 2001 (Casais et al., 2001) with most recent protocol published in 2017 (Bickerton et al., 2017). The following outlines these methods.

The reverse genetics system for IBV involves use of a recombinant Vaccinia virus (rVV) containing a full-length IBV sequence in the place of the thymidine kinase gene. The IBV sequence is held under the control of a T7 promoter located at the 5' end of the IBV sequence. A hepatitis delta ribozyme (HδR) and T7 terminator sequence are encoded at 3' end of IBV insert. Modifications to the IBV sequence within vaccinia are made by transient dominant selection (TDS) using a plasmid possessing the desired modification and the Xanthine phosphoribosyltransferase (*gpt*) gene. The gene is required to circumvent an induced inhibition of guanosine monophosphate (GMP) and guanosine triphosphate (GTP) synthesis (Figure 2.1). The vaccinia virus system was originally chosen due to the capability to accept large-scale modifications to its genome without compromising virus stability.

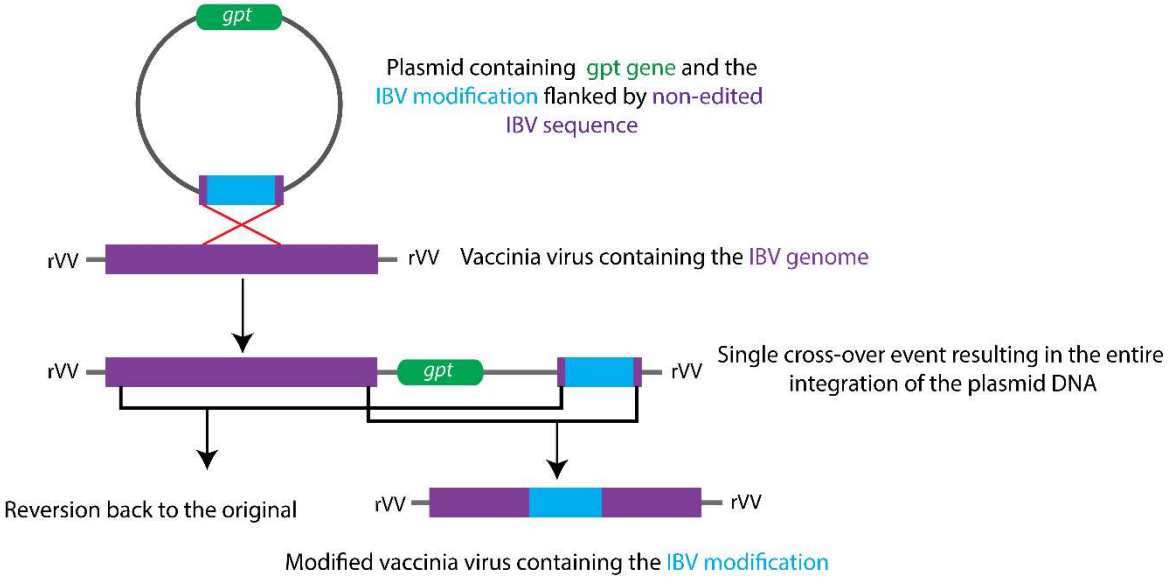


Figure 2.1: The transient dominant selection system. *A plasmid vector containing the gpt gene (green) and the IBV modification flanked by non-edited IBV sequence is required for the reverse genetic system. The plasmid is transfected into Vero cells co-infected with rVV. Homologous recombination between the plasmid donor and vaccinia receiver sequence results in an unstable intermediate selected for by gpt selection agents. The removal of gpt selection agents then results in a second cross-over event and either reversion back to the original sequence or inclusion of the modified sequence. The outcome of this occurs by chance.*

2.10.1. *GPT* gene and GMP biosynthesis

rVVs containing the desired modification in the IBV sequence are selected for by using the *gpt* gene from *Escherichia coli* (*E. coli*) which confers resistance to mycophenolic acid (MPA) in the presence of xanthine and hypoxanthine. MPA inhibits the *de novo* GMP and GTP (guanosine triphosphate) production (Figure 2.2). The *gpt* gene is under the control of the vaccinia virus P7.5 promoter and allows the infected cells to overcome intracellular purine depletion by enabling GMP production by an alternate pathway. MPA, xanthine and hypoxanthine are supplemented to the cell maintenance medium (Table 2.19).

MHX stocks

- MPA (mycophenolic acid): 10 mg / mL in 0.1N NaOH (30 mM); 400 x concentrated.
- Hypoxanthine: 10 mg / mL in 0.1 NaOH (73 mM); 667x concentrated.
- Xanthine: 10 mg / mL in 0.1N NaOH (66 mM); 40 x concentrated.

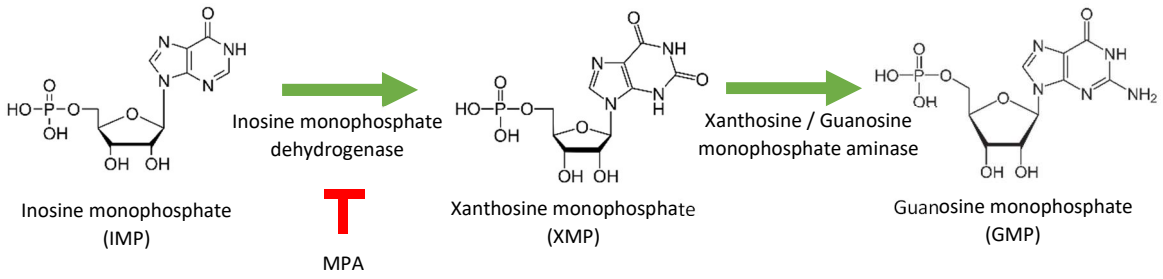


Figure 2.2: GMP biosynthesis in eukaryotic cells. *The inclusion of MPA in cell maintenance medium inhibits the activity of inosine monophosphate dehydrogenase preventing the conversion of inosine monophosphate (IMP) to xanthosine monophosphate (XMP). This in turn prevents synthesis of guanosine monophosphate (GMP) ultimately disabling guanosine triphosphate production (GTP) and resulting in purine depletion.*

Table 2.19 Mychophenolic acid, hypoxanthine, and xanthine medium (MHX) overlay for TDS

INGREDIENT	VOLUME	FINAL CONCENTRATION
MINIMUM ESSENTIAL MEDIUM EAGLE (2X) (TABLE 2.2)	50 mL	1 x
2 % AGAR	50 mL	1%
30 mM MPA	250 µL	75 µM
66 mM XANTHINE	2.5 mL	165 µM
73 mM HYPOXANTHINE	149 µL	184 µM

2.10.2. Generation of rVV by homologous recombination

Vero cells were grown to approximately 70% confluency in 6-well plates at which point cells were infected with rVV containing the whole-genome IBV insert. After infection for 1-2 hours virus inoculum was removed and cells washed twice with Opti-MEM™ I Reduced Serum Medium with GlutaMAX™ supplement (Gibco, 51985026). Each well was then transfected with 5 µg of plasmid containing the desired IBV modification and *gpt* gene using 12 µL Lipofectin (Invitrogen, 18292011) in 3 mL OPTIMEM. After incubation for 1-2 hours the transfection medium was replaced with 5 mL 1 x BES containing medium. The following day MHX selection agents were added to each well in the following volumes; 12.5 µL MPA, 7.4 µL Hypoxanthine, 125 µL Xanthine. Cells were further incubated for 2 days or until extensive VV cytopathic effect (CPE) was observed at which point cells were scraped into media supernatant. Cells were then pelleted by centrifugation and resuspended in 400 µL fresh 1 x BES-containing medium.

2.10.3. Transient dominant selection (TDS)

Transient dominant selection (TDS) was used to select rVVs that possess the *gpt* gene and hence the desired IBV modification (Figure 2.1). The introduction of the IBV-containing plasmid and *gpt* gene into vaccinia virus infected cells results in a cross-over event generating a highly unstable intermediary containing full length IBV sequence, *gpt* gene and modified IBV sequence. This outcome is initially selected for by MHX selection.

rVVs generated above were sonicated using a Branson Digital Cup Horn Sonifier 450 (2 minutes, 70% continuous pulse) and virus dilutions between 10^{-1} – 10^{-3} prepared using 1 x Minimum Essential Medium Eagle (Table 2.2). Confluent 6-well plates of Vero cells were washed once using PBSa and infected with 500 µL virus, plating each dilution in duplicate. After infection for 1-2 hours, virus inoculum was replaced with 3 mL overlay as per Table 2.19. This overlay contains MHX selection agents.

Two days after rVV infection, cells were stained with 0.01% neutral red (Sigma, N4638) in 2 mL overlay consisting of equal volumes of 2 x Minimum Essential Medium Eagle (Table 2.3) and 2% agar. 24 hours after staining, roughly 5 resulting well-isolated plaques were picked from each recombinant and suspended in 400 µL 1 x Minimum Essential Medium Eagle. The process was repeated for a total of three rounds of plaque purification, carrying forward between 2-3 rVVs at each stage. Each rVV was sonicated prior to use using the parameters as above. Any recombinants not producing plaques during these stages have lost the *gpt* gene and the hence no longer possess the desired modification (reversion to wildtype). Where appropriate rounds of plaque purification may have been repeated to ensure an adequate number of rVVs to carry forward.

After the final round of plaque purification in the presence of MHX selection agents, rVVs were then plaque purified in the absence of MHX. As performed previously, confluent 6-well plates of Vero cells were washed once using PBSa and infected with 500 µL virus, plating each dilution in duplicate. After 1-2 hours, virus inoculum was replaced with 3 mL of overlay consisting of equal volumes of 2 x Minimum Essential Medium Eagle and 2% agar. Plaques were stained and picked as described previously.

The *gpt*+ rVV is intrinsically unstable due to presence of repeat sequences in immediate proximity. MHX selection reagents can initially support this construct but subsequent passaging in their absence results in second cross-over event. This can either result in either a full length IBV sequence with the modification or reversion back to the unmodified IBV sequence (Figure 2.1). Both possibilities are equally likely and result in the spontaneous loss of the *gpt* gene.

2.10.4. Growth of small rVV stocks

Following three rounds of plaque purification in the absence of MHX selection agents, small stocks of rVV were grown for confirmation of the IBV sequence and to confirm absence of the *gpt* gene. Confluent 6-well plates of Vero cells were washed once with PBSa and infected with 150 µL rVV plaque (sonicated for 2 minutes as above) and 350 µL 1 x BES-containing medium. After 1-2 hours, 2.5 mL 1 x BES-containing medium was added to each well for a total of 3 mL.

Cells were incubated until extensive (70-80%) CPE was observed at which point cells were scraped into the supernatant and pelleted by centrifugation. The supernatant was discarded and cell pellet suspended in 700 µL 1 x BES-containing medium. 200 µL aliquoted from the virus stock for DNA extraction. The remaining 500 µL was stored at -20°C.

2.10.5. Small-scale DNA extraction

DNA extraction was performed using QIAamp DNA Mini Kit (QIAGEN, 51304) following the DNA Purification from Blood or Body Fluids (Spin Protocol) manufacturer's protocol. 200 µL aliquoted virus stock was digested using 20 µL Proteinase K (provided in kit) and heated to 56°C for 30 minutes to ensure virus denaturation. DNA purification proceeded as per manufacturer's protocol with resulting DNA eluted from the silica column membrane using 200 µL nuclease free molecular grade water.

Resulting DNA was amplified by PCR to confirm absence of *gpt* gene and sequenced to confirm the modification made to the IBV sequence.

2.10.6. Growth of large rVV stocks

Following identification of rVVs ministocks containing the correct sequence, large stocks were generated using BHK-21 cells. Per vaccinia, twelve T150 flasks of BHK-21 cells inoculated each with 10 – 20 µL rVV ministocks, diluted in 20 mL 1 x GMEM (Table 2.6). Cells were returned to incubator and further incubated for several days until flasks showed extensive signs of CPE and could be easily dissociated from the plastic. Cells and supernatant were transferred to 50 mL FALCON tubes and pelleted by centrifugation for 15 minutes at 270 *x g* in a refrigerated 4°C centrifuge. Resulting supernatant was discarded and pellets re-suspended in 11 mL TE buffer, pH 9 (Table 2.20). Two 0.5 mL aliquots were kept for stock and stored at -80°C whilst the remaining 10 mL was purified as per section 2.10.7.

Table 2.20 Tris-EDTA (TE) buffer, pH 9

INGREDIENT	FINAL CONCENTRATION
Tris-HCl (Trizma® hydrochloride solution), pH 9 (Sigma, T2819)	10 mM
Ethylenediaminetetraacetic acid (EDTA), pH 9 (Sigma, E9884)	1 mM

2.10.7. Vaccinia virus purification

Stocks of rVVs were purified prior to DNA extraction. 10 mL recombinant vaccinia virus was firstly freeze-thawed a total of three times then sonicated for 2 minutes (as above). Cells were immediately centrifuged for 10 minutes at 500 x *g* in refrigerated centrifuge with the resulting cell pellet discarded. Supernatant was made up to 13 mL using TE buffer, pH 9 and layered on top of a 16 mL 30 % sucrose cushion. rVVs were then centrifuged for 1 hour at 4°C and spun at 36, 000 x *g* in an ultracentrifuge. The sucrose cushion was then removed and the pellet resuspended in 5 mL 10 mM TE buffer and stored at -80°C.

2.10.8. Large-scale DNA extraction

Proteinase K was prepared by dissolving 20 mg proteinase K powder (Sigma, 2308) in 1 mL type 1 endotoxin-free water for a final concentration of 0.2 mg / mL. Proteinase K digestion buffer was prepared according to Table 2.20 and an equal volume added to the purified vaccinia (i.e. 5 mL purified vaccinia, 5 mL protease K buffer). 120 µL protease K was added then and reaction heated for 2.5 hours at 50°C.

Table 2.21 Proteinase K digestion buffer (2x)

INGREDIENT	FINAL CONCENTRATION
Tris-HCl (Trizma® hydrochloride solution), pH 7.5 (Sigma, T2319)	200 MM
EDTA, pH9	10 MM
Sodium dodecyl sulfate (SDS, Sigma, 75746)	0.4%
Sodium chloride (Sigma, S3014)	400 MM

After incubation a DNA extraction by phenol:chloroform and chloroform methods was performed. Briefly, an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) containing 8-hydroxyl-quioline was added, mixed by inversion 5 – 10 times, and spun in a refrigerated centrifuge for 15 minutes at 1,200 x *g*. The upper aqueous phase was then transferred into clean 50 mL FALCON tubes taking care not to disturb the interphase. The process was repeated again using phenol:chloroform and chloroform using the procedure as described above.

The aqueous phase from the final chloroform extraction was transferred into a clean 50 mL FALCON and DNA precipitated by adding 2.5 volumes of -20°C absolute ethanol and 0.1 volumes of 3 M sodium acetate. Samples were then spun for 90 minutes at 2,300 x *g* in a refrigerated centrifuge. The resulting supernatant was removed and discarded. Potentially contaminating salts were removed by washing the pellet 10 mL of -20°C 70% ethanol and incubated on ice for 5 minutes

prior to centrifugation for 1 hour at 2,300 x *g* at 4°C. Contaminating ethanol was removed by air drying and the DNA pellet resuspended in 100 µL nuclease-free molecular grade water. DNA was held at 4° C until fully dissolved after which DNA was quantity and quality assessed by using a Nanodrop 1000 (ThermoScientific) and by SalI restriction digestion and subsequent PFGE.

2.10.9. Rescue of recombinant IBV

Following isolation of rVV DNA, the recovery of infectious recombinant IBVs was performed in primary CKCs infected with recombinant fowlpox virus (rFPV-T7) modified to encode a bacteriophage T7 RNA polymerase. In addition to transfection with rVV cells were also transfected with plasmid expressing the essential IBV N protein (pCi-Neo-N) which is required for rIBV recovery. Both the modified IBV sequence in vaccinia virus and the IBV N gene are held under a T7 promoter. Two rVVs containing the desired IBV sequence were taken forward for rescue using the protocol as detailed below.

At approximately 50 % confluency, 6 well plates of CK cells were washed twice with PBSa to remove debris then infected with rFPV-T7 at a multiplicity of infection (MOI) of 10 and incubated for 1 hour at 37°C. During this incubation transfection solutions A and B were prepared as per Table 2.22. A total of 10 replicates were set up per vaccinia virus due to the low success rate of rIBV recovery.

Table 2.22 Transfection reagents for recovery of rIBV in CKCs.

INGREDIENT	PER REPLICATE (PER WELL)	
	Solution A	**Solution B**
Opti-MEM™ I Reduced Serum Medium With GlutaMAX™ Supplement (Gibco, 51985026)	1.5 mL	1.5 mL
rVV DNA containing modified IBV sequence	10 µg	-
pCi-Neo-N (plasmid expressing IBV N protein)	5 ug	-
Lipofectin (Invitrogen, 18292011)	-	30 µL

Solutions A and B were incubated separately at room temperature for 30 minutes before being combined into Solution AB. Solution AB was incubated for a further 15 minutes at room temperature before use.

After 1 hour infection with rFPV-T7 inoculum was removed and cells washed twice with Opti-MEM™. 3 mL of solution AB was then applied to each well and incubated overnight at 37°C with 5% CO_2. The following day the transfection reagent was removed and replaced with 5 mL 1 x BES-containing medium. Cells were incubated for a further three days until extensive FPV-induced CPE was observed. At this point supernatant was harvested and passed through a 0.22 µm filter to remove rFPV-T7. The rIBV was then stored at -80°C.

Due to the low recovery of infectious virus as rescue, rIBVs are passaged three times in CKCs prior to generating a working stock in embryonated eggs.

Chapter 3 – Characterising parallel attenuated M41-CK viruses serially passaged in embryonated eggs

3.1. Declaration

The contents of this chapter have been published in the Journal of Virology (Oade et al., 2019) entitled "Attenuation of infectious bronchitis virus in eggs results in different patterns of genomic variation across multiple replicates". Where appropriate some text has been altered in this chapter versus the research article to avoid repetition, meet university guidelines and prevent interruption of the thesis narrative.

Contributions from other authors to this publication is as listed below.

- Passaging of M41-K was performed by Sarah Keep.
- Characterisation of M41-CK viruses *in vitro* / *ex vivo* (Figure 3.2) was performed by Sarah Keep.
- Characterisation of M41-CK viruses *in vivo* (Figure 3.3) was performed in injunction with members the Animal Services Unit and the Coronaviruses group (Pirbright Institute) Ciliary activity assessments were performed blind by Erica Bickerton. qPCR/PCR tissue screens were performed by Sarah Keep and Erica Bickerton.
- Pipeline for data analysis was developed in conjunction Richard Orton.

3.2. Introduction

Serial egg-passaging as a means of attenuating IBV has long been used as a standard method for vaccine development. Despite the reliance on this method, the mechanism underlying IBV attenuation by egg-passaging is yet to be determined. Indeed it has not been established whether attenuation is linked to adaption of the virus to growth in eggs or by another unknown mechanism.

Moreover, attenuation by egg-passaging is not a guaranteed process with it possible for viruses to either remain virulent after serial passaging or to dissipate either by natural or artificial causes during the process.

The unreliability of serial egg-passaging has therefore refocused efforts for the development of rationally attenuated IBVs. Generated by reverse genetic systems, such viruses would possess the key genetic chains that confer attenuation. Without a comprehensive understanding of IBV attenuation, the scope of rationally attenuated IBVs is limited.

Like with all RNA viruses, IBV does not exist as a clonal structure but rather as cluster of genetically related variant viruses. It has previously been suggested (Toro et al., 2012, Mayr, 1988) that it is the regulation of this population, by both *de novo* mutation and selection, that drives the evolution of the virus. It is therefore suggested that egg passaging may influence the regulators of IBV evolution and hence bring about attenuation. As a result of the complexity of an IBV population it may transpire that attenuated viruses already exist within virulent populations with egg-passaging driving the selection of these viruses. Understanding the genetic structure of these populations utilising available high throughput sequencing technologies is required to further our understanding of IBV attenuation.

In order to study both virulent and attenuated IBV populations four independent replicates of lab pathogenic IBV (M41-CK) were established and serially passaged in embryonated eggs analogous to the industry standard. The following details the characterisation of these viruses by phenotypic and genomic approaches, utilising *in vivo* and *in vitro* methods and 454 pyrosequencing.

3.3. Results

3.3.1. Serial *in ovo* passaging of M41-CK

A pathogenic strain of IBV, M41-CK, was inoculated into the allantoic cavities of four 10-day-old embryonated SPF RIR chicken eggs (A, B, C, and D). After a 24 hour incubation, embryos were candled and culled by refrigeration for a minimum of 4 h. Allantoic fluid was then harvested, centrifuged at 700 x *g* for 5 min to clarify supernatant, and stored at -80°C. For subsequent passages, up to a total of 106, the allantoic fluid was diluted with 1 x BES-containing medium between 1:100 and 1:10,000 based on visual inspection of the condition of the embryo and inoculated into the allantoic cavity, one egg per lineage, with each lineage remaining independent from each other.

Allantoic fluid was routinely screened for presence of IBV by RT-PCR amplification of the 3′ UTR, using oligonucleotides BG56 and 93/100. After routine screening at passage 14, lineage B could no longer be detected. At this passage allantoic fluid from lineage A was used to infect two eggs, A and A1, which remained independent for the duration of the experiment. The disappearance of lineage B could have been caused by multiple factors, both biological and procedural, and was not investigated any further, as it is known in the field that the loss of virus during egg passaging is not uncommon.

A schematic representation of M41-CK is detailed in Figure 3.1. From herein where lineages A to D are described, this refers to lineages A, A1, C, and D. All egg passaging was performed by Sarah Keep (Pirbright Institute) prior to commencement of this PhD.

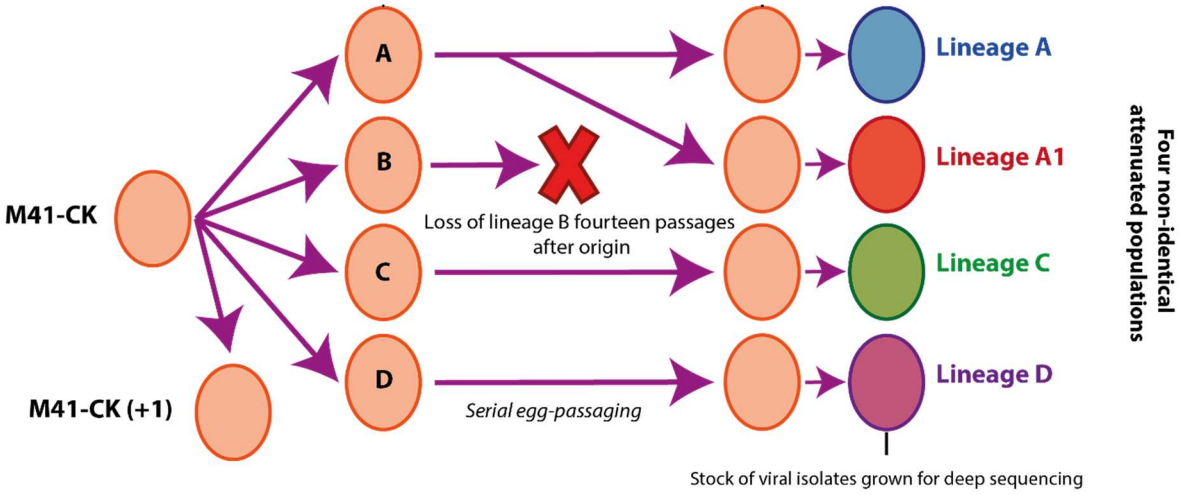


Figure 3.1: Passaging of M41-CK in embryonated RIR eggs. *M41-CK was used to infect four separate eggs to generate four independent lineages which were then serially passaged independently with the aim of generating an attenuated virus. After routine PCR screening, IBV could no longer be detected passage 14 of lineage B. At this passage allantoic fluid from lineage A was used to infect two eggs, A and A1, which remained independent for the duration of the experiment. In vitro and in vivo characteristics were assessed on the direct M41-CK ancestor and final passage of each isolate. To generate sufficient material for sequencing, M41-CK and each lineage was passaged once more to generate a stock virus.*

3.3.2. Adaptation to embryonated eggs does not impair the ability of M41-CK to grow in adult chicken tissues *in vitro*

The growth characteristics of lineage A to D EP106 were investigated and compared to those of virulent M41-CK *in vitro* using primary CKCs. Viral progeny over a 96h period was assessed by plaque titration assay (Figure 3.2A). Growth characteristics were broadly similar for all five viruses, with lineages C and D at 48 h and lineage D at 96 h being significantly different from M41-CK ($P = 0.0028$, 0.0041, and 0.0222, respectively) using Dunnett's multiple comparison test. All other points showed no significant difference from M41-CK. Plaque morphologies for the egg-passaged viruses were similar to that of the M41-CK parent (data not shown).

To assess each virus's ability to cause ciliostasis, tracheal organ cultures (TOCs) were infected with 5×10^4 PFU of virus. At days 4 to 8 post-infection, ciliary activity was assessed by light microscopy and the proportion of beating cilia was determined (Figure 3.2B). TOCs infected with virulent M41-CK exhibited very low ciliary activity and scored 0.27 and 0.09 for days 4 and 6 post-infection, respectively. Each serially passaged viral isolate caused a significant reduction in ciliary activity compared to that in the mock-infected group ($P < 0.0001$). Ciliary activities for lineage A (days 4 and 5), lineage A1 (days 4, 5, and 6), and lineage C (day 4) were deemed significant (versus M41-CK at the same timepoint), with P values of < 0.0001, 0.0004, < 0.0001, < 0.0001, 0.0015, and 0.0054, respectively. All other comparisons were nonsignificant ($P > 0.05$).

CKC experiments were performed by Sarah Keep (Pirbright Institute). TOC experiments were performed by Sarah Keep and Erica Bickerton (Pirbright Institute).

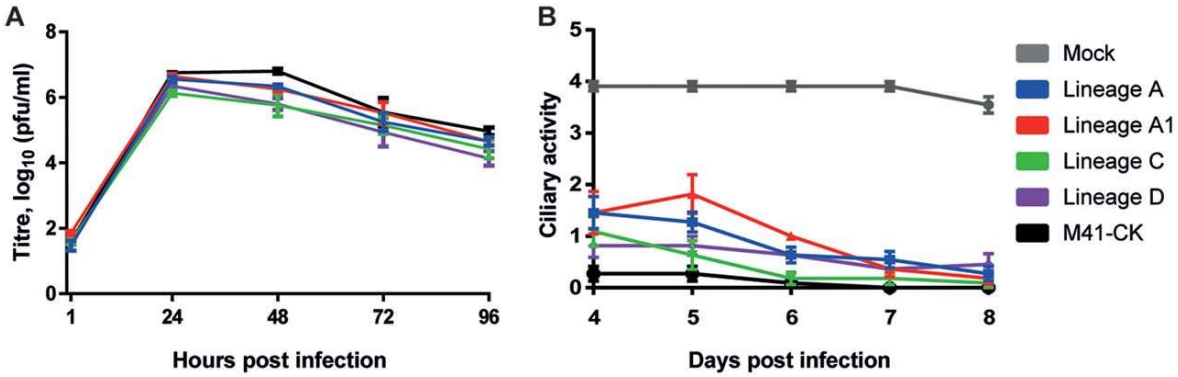


Figure 3.2: Serially egg-passaged viruses exhibit an in vitro phenotype similar to that of the virulent parent M41-CK. *(A) Growth kinetics of serially egg passaged virus on CKCs. Confluent 6-well plates of CKCs were infected with virus at an MOI of 0.05. At 1, 24, 48, 72, and 96 h post-infection, cell medium was harvested for analysis of viral progeny by plaque titration assay on CKCs. Each value represents the mean of three replicates, with SEMs plotted as error bars. (B) Induction of ciliostasis by serially egg-passaged IBVs in TOCs. TOCs were infected with 5 x 10^4 PFU of IBV or mock infected with TOC medium. Ciliary activity was assessed under a light microscope at days 4 to 8 post-infection and scored as follows: 100% activity = 4, ≈ 75% = 3, ≈ 50% = 2, ≈ 25% = 1, and ≈ 0% = 0 Each value plotted represents the mean score of 11 replicates, with SEMs plotted as error bars. CKC experiments were performed by Sarah Keep (Pirbright Institute) and TOC experiments were performed by both Sarah Keep and Erica Bickerton (Pirbright Institute). Statistical comparisons were performed by Michael Oade.*

3.3.3. Serially egg-passaged M41-CK viruses are attenuated *in vivo*

Groups of 12 8-day-old specific-pathogen-free (SPF) Rhode Island Red (RIR) chickens were inoculated with 1 x 10^5 PFU of IBV or mock infected with 0.1 ml of serum-free medium via the intranasal and ocular routes. Each virus characterisation had their own mock infection and virulent M41-CK virus controls. Snicking (Figure 3.3A) and rales (Figure 3.3B) were recorded daily between days 3 and 7 post-inoculation.

Average ciliary activity scores over 50% suggest that IBV is attenuated, here indicated by an average ciliary activity score greater than 2 (Figure. 3.3C). On days 4 and 6 post-infection, the average ciliary activity of chickens infected with all serially egg-passaged M41-CK viruses remained above a 50% threshold and was significantly different from that of the M41-CK virulent controls. As RIR chickens are outbred, some differences were observed in clinical signs between the different M41-CK virulent control-infected groups; however, all virulent M41-CK-infected birds exhibited ciliostasis. With the exception of lineage D (P values of 0.0003 and 0.0044 for days 4 and 6, respectively), the difference between the egg-passaged viruses versus mock at each time point was deemed not significant by pairwise comparison Tukey).

Chickens inoculated with lineage C demonstrated clinical signs consistent with IBV infection, although differences from mock-infected birds were not significant. While birds infected with lineage C showed some clinical signs, the virus was deemed attenuated due to low observed clinical effect and an average ciliary activity above 50%.

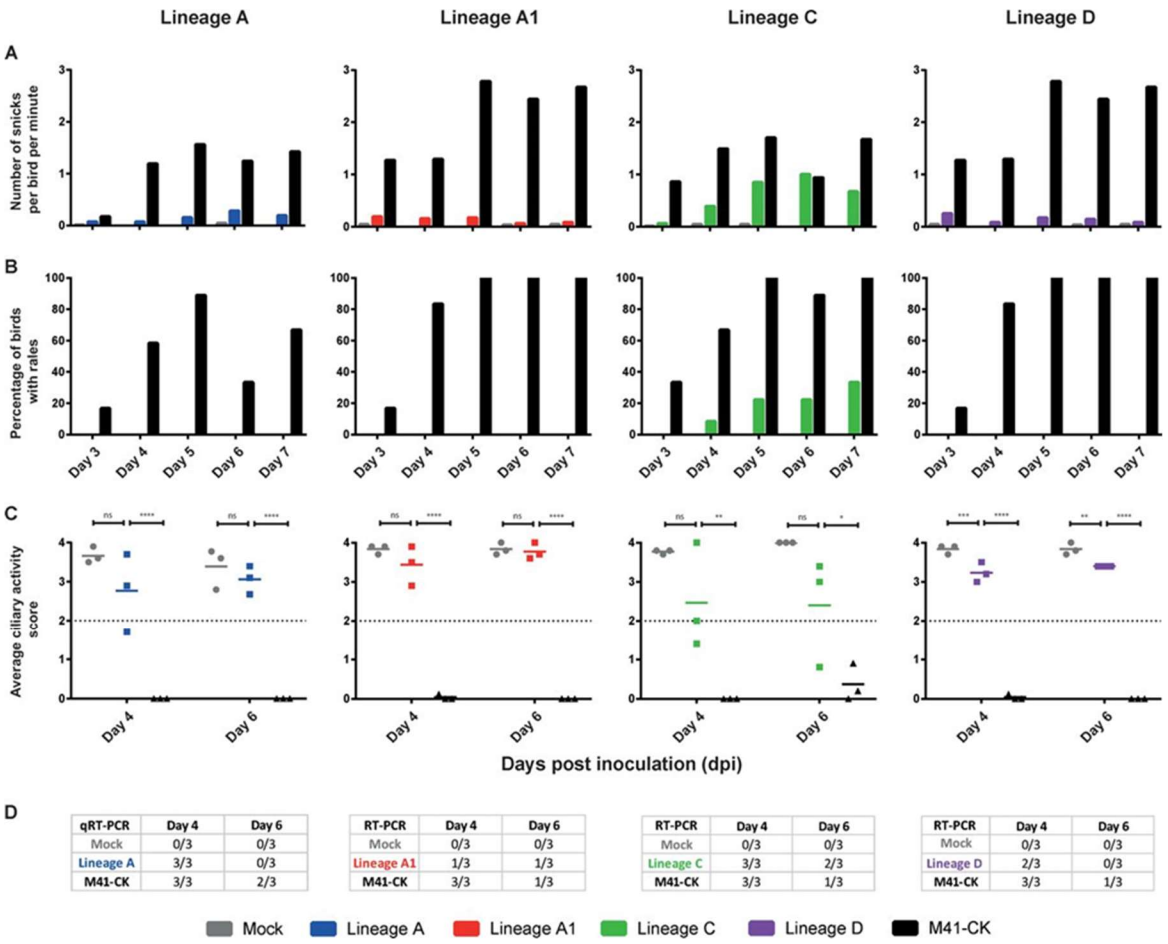


Figure 3.3: Serially egg-passaged M41-CK lineages are attenuated in vivo*. Viral pathogenicity was assessed in a series of three separate animal studies. Assessment of lineages A and C was performed in separate investigations, while lineages A1 and D were assessed in a joint study. For each group (n = 12), 8-day-old SPF chicks were inoculated with 1 x 10^5 PFU of each virus via the ocular and nasal routes. Clinical signs, including snicks per bird per minute (A) and rales (B), were recorded daily from days 3 to 7 post-infection. (C) Ciliary activity was assessed by observation of 10 tracheal cross sections (rings) from three randomly selected birds per group on days 4 and 6 post-infection. The mean ciliary activity of these three birds is plotted, with SEM plotted as error bars. *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001. (D) Detection of IBV in trachea by using PCR or qPCR methods. Animal experiments and detection of IBV RNA was performed by Sarah Keep and Erica Bickerton. All data was analysed and statistical comparisons performed by Michael Oade.*

Tracheas were screened for the presence of IBV RNA on days 4 and 6 post-infection (Figure 3.3D). IBV RNA was detected in the tracheas of all birds infected with virulent M41-CK on day 4 post-infection, decreasing to one or two out of three M41-CK-infected birds on day 6 post-infection. Similarly, the number of birds infected with egg-passaged viruses testing positive for IBV RNA declined from day 4 to day 6 post-infection in all groups except lineage A1, for which only a single bird tested positive on each day. Interestingly, presence of IBV RNA in the trachea and reduction in ciliary activity did not correlate. For example, while the average ciliary activities for birds infected with lineage A did not differ between day 4 and day 6 post-infection (Figure 3.3C), all three birds infected with lineage A tested positive for IBV RNA on day 4 post-infection and none of the birds tested positive for IBV RNA in the trachea on day 6 post-infection (Figure 3.3D).

Due to the reduction in clinical signs, reduction in the presence of IBV RNA in the trachea, and increased ciliary activity in birds infected with egg-passaged viruses in comparison to those of birds infected with virulent M41-CK, all four lineages were considered attenuated by serial egg passage.

3.3.4. Few consensus-level difference occur between virulent and attenuated M41-CK

Once attenuation of the egg-passaged viruses was confirmed, the genomic changes for each isolate compared to the starting inoculum were investigated. Stocks of virulent M41-CK and attenuated final passage viruses of each lineage were grown and partially purified by Sarah Keep, and the RNA genomes were sequenced by 454 pyrosequencing at the Center for Genomic Research – University of Liverpool. All analysis of resulting reads was performed my Michael Oade.

Of the 105,728 quality-filtered reads obtained for M41-CK (Table 3.1), 102,415 (96.87%) reads aligned to M41 (GenBank accession number AY851295.1) (Fig. 3.4A and B). Thirty-nine point mutations (39/27,475) were identified in the M41-CK sequence relative to AY851295.1 (Figure. 3.4C and D) occurring across the genome. More SNPs are observed within the structural and accessory protein-encoding and non-protein coding regions at the 3′ end of the genome as opposed to a notable lack of SNPs observed from the 5′ end of the genome to approximately 5,000 nt (5′ UTR and nsp2). Consensus for M41-CK could not be called for 20 nucleotide positions, including 13 nucleotides that could not be called due to insufficient coverage and 7 positions that were ambiguous. The generated sequence was annotated and deposited on GenBank under the accession number MK728875.

Table 3.1. Read statistics for each M41-CK virus

Statistic	Value for virus				
	M41-CK	Lineage A	Lineage A1	Lineage C	Lineage D
No. of raw reads	114,483	23,617	25,395	69,831	120,410
No. of quality-filtered reads	105,728	21,506	22,854	64,699	111,319
No. of quality-filtered AY851295.1-aligned reads	102,415				
No. of quality-filtered M41-CK EP5-aligned reads	102,417	21,078	21,521	63,253	108,522
Mean depth	1,185.97	264.71	272.47	735.83	1,226.48
Median depth	1,218	199	123	719	1,251
Range of depth[a]	Min, 0; max, 2,331	Min, 0; max, 1,331	Min, 0; max, 2,258	Min, 0; max, 1,971	Min, 0; max, 2,439

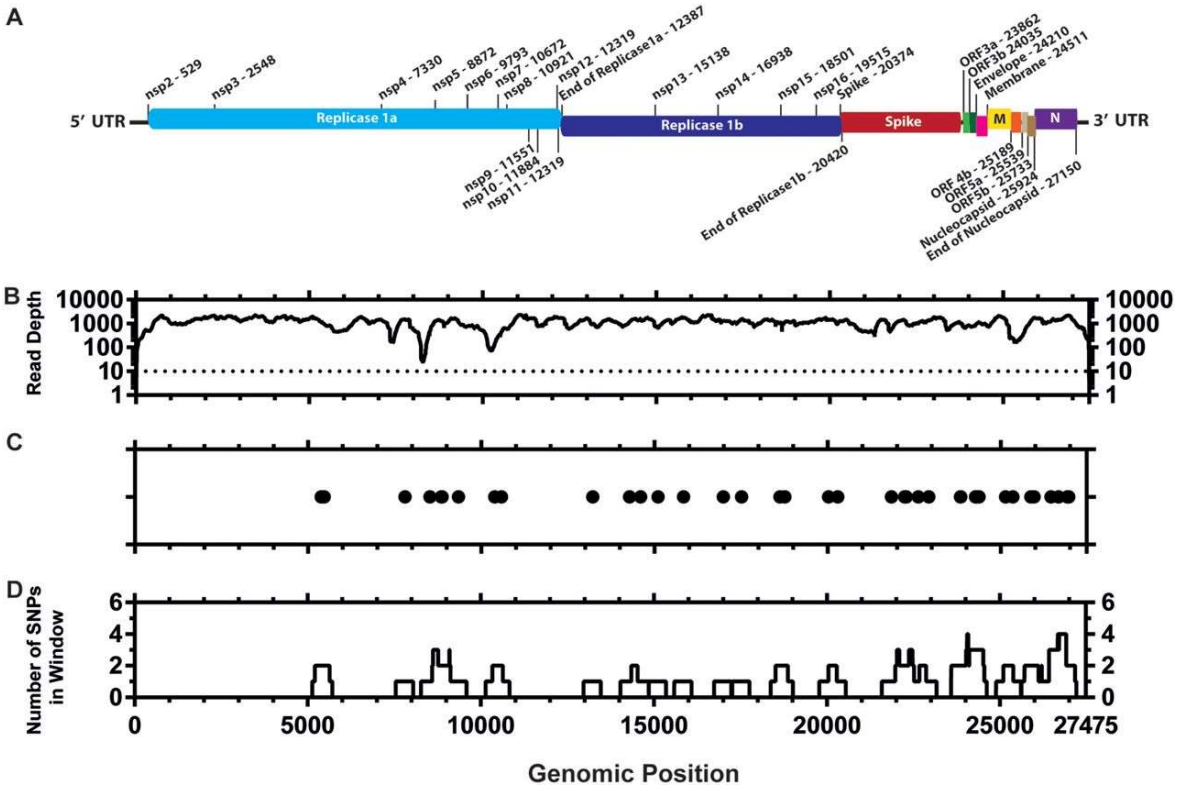


Figure 3.4: Distribution of consensus-level mutations across the M41-CK genome. *(A) Genetic map of M41-CK consensus sequence (GenBank accession number MK728875). The nucleotide positions of each genomic region are detailed. (B) Sequencing coverage of aligned M41-CK quality-filtered reads to AY8512951.1 genome. Sites with no coverage (= 13) have been plotted as having a coverage of 1. The dotted line shows sites where consensus has not been called. (C) Consensus-level changes observed in M41-CK relative to AY8512951.1. Each circle represents a nucleotide substitution. A total of 39 mutations were identified between the two sequences. Consensus for M41-CK at the initial 18 positions at the 5' end of the genome and final two positions at the 3' end genome could not be called. (D) Number of consensus mutations occurring within a 499-nucleotide window with genomic position as the midpoint.*

Given that M41-CK has been adapted for growth in CKCs whereas the M41 strain of IBV typically cannot, a high number of mutations between the two sequences may therefore be anticipated. Perhaps unsurprisingly spike protein, which regulates both binding and entry of virions to susceptible cells, possessed a high number of single nucleotide polymorphisms (6 SNPs) compared to AY851295.1. This included the introduction of an alternate stop codon resulting in M41-CK spike protein prematurely terminating by 27 nucleotides (9 amino acids) compared to AY851295.1.

With the set coverage threshold (10 reads), consensus could be called for 89.53%, 89.16%, 99.90%, and 99.90% of the viral genome for lineages A to D, respectively (Figure 3.5A and Table 3.1). Consensus-level changes (occurring at a > 50% frequency) were called for each lineage relative to M41-CK, with 11, 17, 13, and 17 SNPs identified in lineages A to D, respectively (Figure 3.5B). Sliding-window analysis showed that some regions of the genome, particularly within the replicase gene, did not contain any consensus-level SNPs and may indicate more conserved areas of the viral genome (Figure 3.5C). Conversely, mutations accumulated at the 3' end of the genome, within the N gene and 3' UTR.

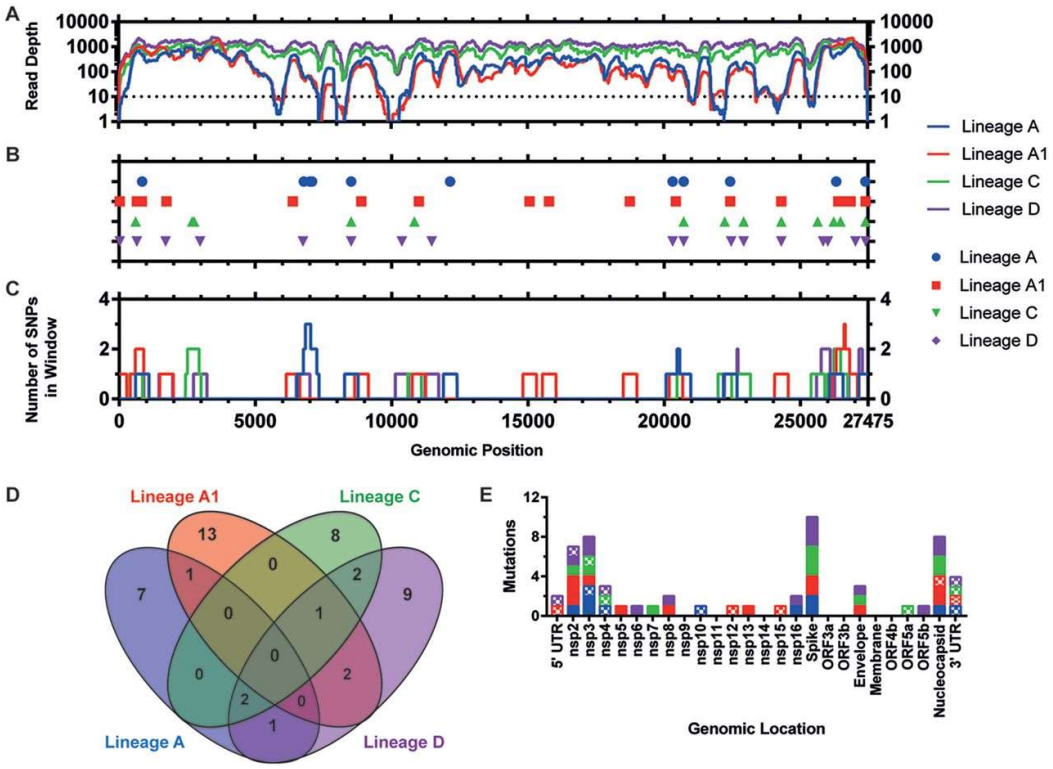


***Figure 3.5*: Distribution of consensus level mutations across attenuated egg-passaged virus genomes.** *(A) Sequencing coverage of aligned final egg-passaged virus reads to the M41-CK genome. The dotted line indicates the threshold at which consensus and minor variants have not been called. Sites with no coverage have been plotted as having a coverage of one. Totals of 529, 111, 13, and 11 sites in lineages A, A1, C, and D, respectively, had no coverage. (B) Consensus-level mutations observed in the attenuated viruses relative to M41-CK EP5. Each point represents a nucleotide substitution; insertions/deletions have not been plotted. Totals of 11, 17, 13, and 17 consensus-level mutations were identified in lineages A, A1, C, and D, respectively. Consensus could not be called for 2,858, 2,959, 18, and 18 positions, respectively, due to the minimum coverage threshold (inclusive of sites with no coverage). (C) Number of consensus-level mutations occurring within a 499-nucleotide window, 249 bp on either side of the midpoint position. (D) Venn diagram of consensus-level mutations occurring at the same position in more than one virus. (E) Synonymous and nonsynonymous mutations for each virus, for each genomic location. Nonsynonymous mutations are shown as solid colors, and synonymous mutations are shown in a checkered pattern. Here the 5' UTR is defined as the genomic sequence upstream of the start of nsp2, while the 3' UTR is defined the genomic sequence downstream of the nucleocapsid gene. Mutations occurring in these untranslated regions have been plotted as synonymous mutations.*

3.3.5. Distribution of variant positions indicates shared regions of diversity between attenuated viruses.

The location of subconsensus variant positions was determined and plotted across the length of the IBV genome (Figure. 3.6A). Variant calling using LoFreq identified a total of 156 unique polymorphisms were identified all five viruses with each virus exhibiting a diversity profile of variants unique to that lineage (Figure. 3.6B). This variant mapping suggests the emergence of new diverse regions in the attenuated virus rather than further diversification of regions already exhibiting variation in the original population. Indeed, in most instances those variable regions present in the original population appear to have been lost in the final attenuated populations.

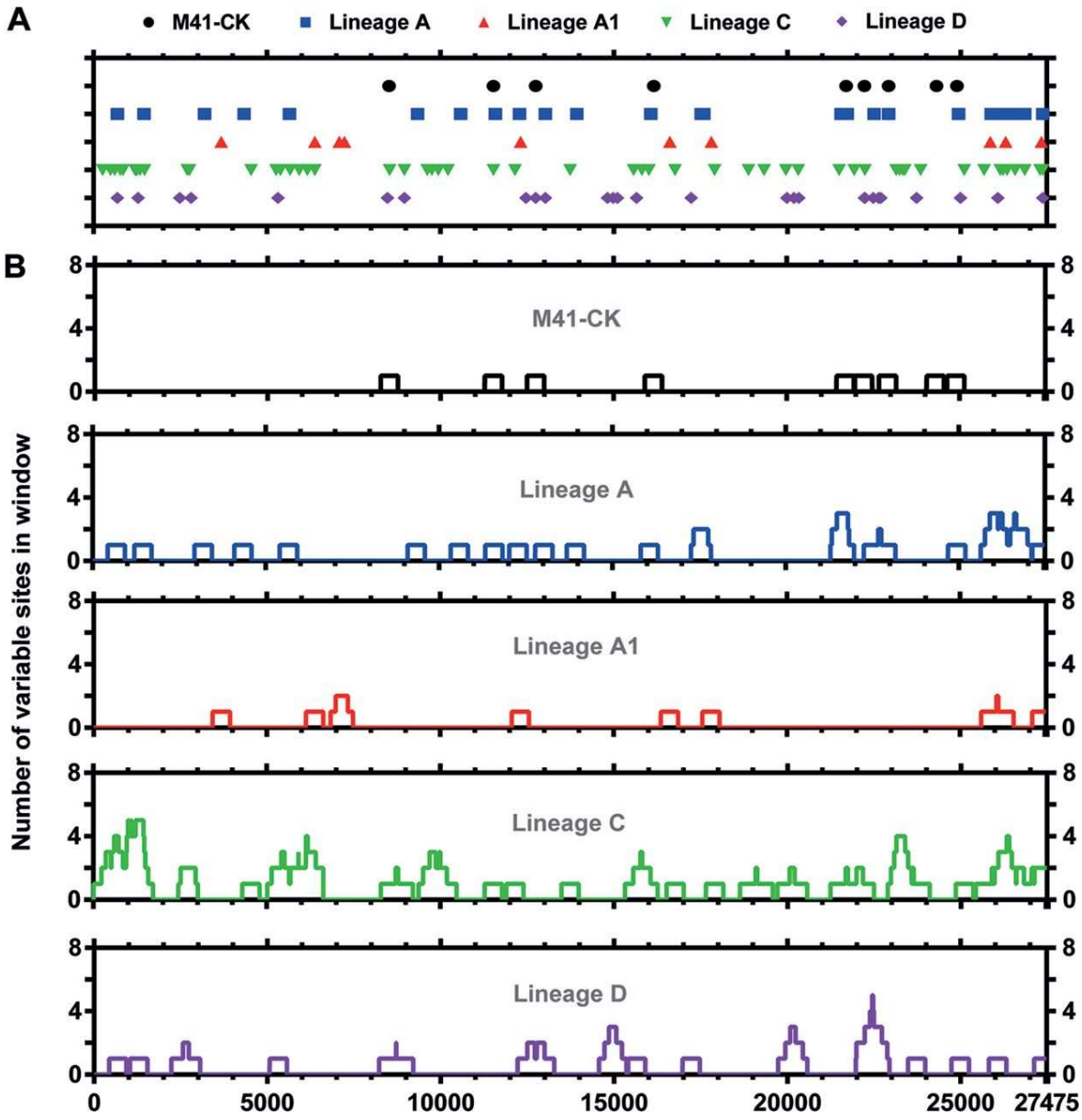


Figure 3.6: Final egg-passaged viruses possess greater variant density than virulent M41-CK. *(A) Sites of variance identified relative to M41-CK consensus sequence. Totals of 9, 27, 10, 55, and 27 subconsensus variants were identified in M41-CK (relative to its own consensus), A, A1, C, and D, respectively. (B) Number of variants occurring within a 499-nucleotide window, 249 bp on either side of the midpoint (inclusive).*

3.3.6. Few consensus and subconsensus mutations are shared across attenuated viruses.

To identify both variants common to multiple attenuated IBVs and potentially attenuating variants present in the virulent population, comparisons were made between variants identified in M41-CK versus those in each final isolate. Of those mutations identified, no mutation was common between M41-CK and all four attenuated viruses at either the consensus or subconsensus level. Furthermore no single mutation was shared in all four attenuated viruses. Twenty-three, 9, 49, and 23 subconsensus variants were identified in lineages A to D, respectively, that were unique to that one lineage, whereas 118 different subconsensus variants were identified across all five samples. Here, mutations occuring within the blue segment (Figure 3.7) could be indicative of selection, as the mutation was also detected in the virulent population. Likewise those mutations occurring outside the blue ring are examples of *de novo* mutation as the mutation was not detected in the original population.

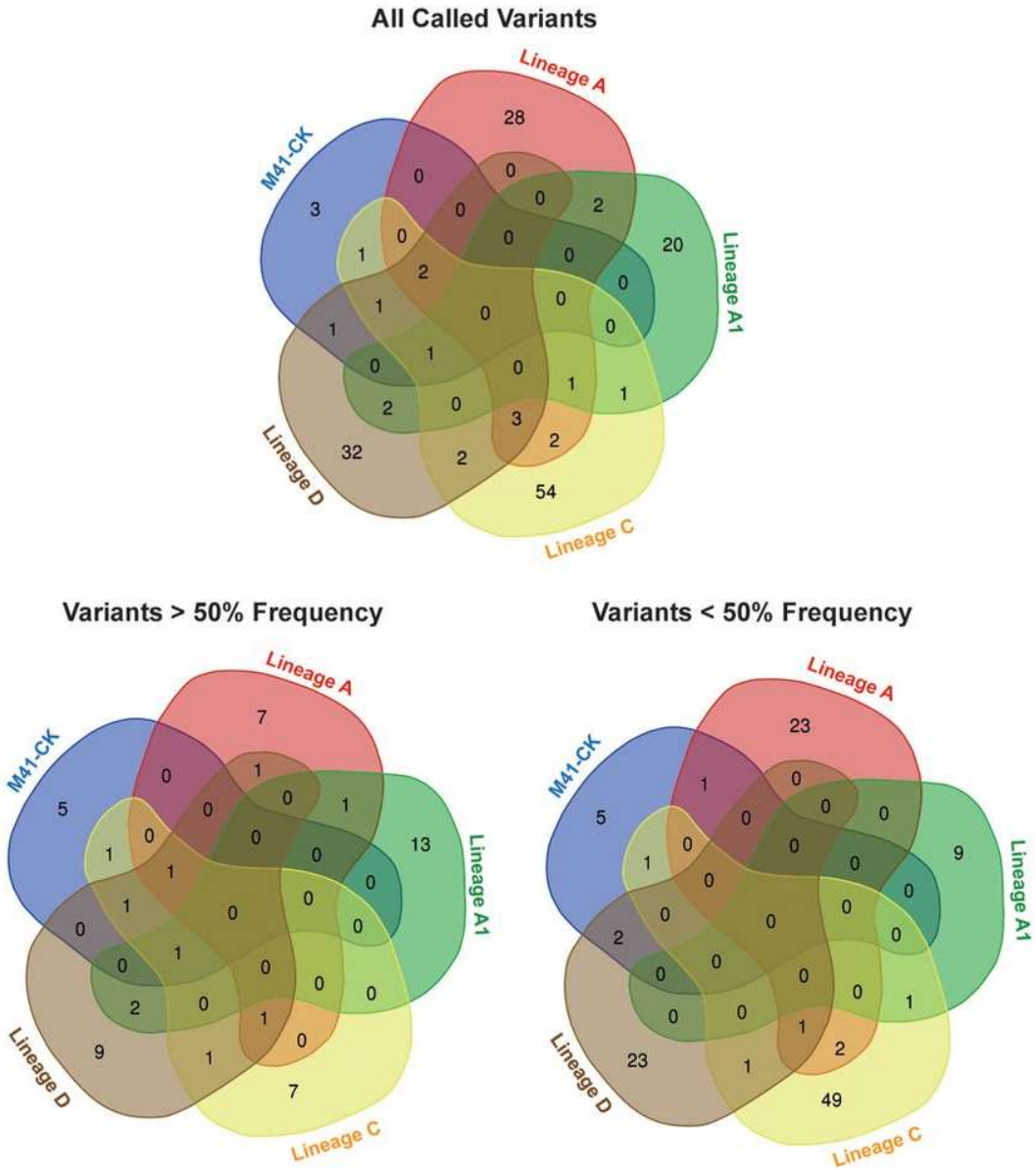


Figure 3.7: Final egg-passaged viruses have few shared consensus- and subconsensus-level variants*. For each of the five viruses, three different lists of mutations were identified as follows: all mutations identified, mutations occurring at a frequency above 50%, and mutations occurring at a frequency below 50%. Each series of lists was then used to generate a Venn diagram using the Bioinformatics and Evolutionary genomics webtool. Numbers occurring in non-overlapping regions are unique to that specific passage. No mutation at either consensus or subconsensus level was shared across multiple lineages.*

To identify consensus level changes in common to vaccine strains against M41, all five consensus sequences were aligned to virulent M41 (Mass41) or to Ma5, H52, and H120 vaccine strain sequences available in the NCBI database. Only seven consensus-level SNPs resulted in a match to a vaccine genome sequence (i.e., Ma5, H52, or H120) (Table 3.2). For U24297C other publicly available IBV sequences, both pathogenic and nonpathogenic, possess a C at this position, suggesting that this U is unique to our initial virus. Similarly, U22224 is present in both virulent M41 and vaccine strains Ma5, H52, and H120, indicating that mutation G22224U identified in lineage C is unlikely to be involved in attenuation.

Table 3.2: Comparison of identified consensus-level mutations against three vaccine strains of IBV[a]

Position in M41-CK	Virus								
	M41-CK	Lineage A	Lineage A1	Lineage C	Lineage D	Mass41	Ma5	H52	H120
30	C	N	U	C	U	C	U	U	U
2755	C	C	C	U	C	C	U	U	U
6781	G	A	G	G	G	G	A	A	A
20429	C	C	U	C	C	C	C	U	C
22224	G	N	G	U	G	U	U	U	U
24297	U	N	C	C	C	C	C	C	C
25637	U	U	U	C	U	U	C	C	C
27393	G	G	G	A	A	G	A	G	G

a Generated consensus sequences for the five viruses were aligned to M41 (AY851295.1), Ma5 (KY626045.1), H52 (EU817497.1), and H120 (FJ881351.1) sequences available in the NCBI database using MAFFT. Instances where an SNP of at least one of the egg-attenuated viruses (i.e., lineages A to D) matches a vaccine strain (i.e., Ma5, H52, and H120) were identified and are included here. M41-CK and M41 (AY851295.1) sequences were also included as a virulent M41 comparison. N, no consensus sequence at given position.

3.3.7. No common trend in variant frequency is observed for mutations in virulent M41-CK

To explore the possibility that attenuation may be result of selection for a pre-existing variant within the virulent population (or indeed the elimination of a variant from the population), each final isolate was surveyed for variants present in the original population. A total of nine variants were detected within the original M41-CK population, three of which were no longer detectable in any egg-attenuated virus (Table 3.3). A11525G (nsp8) occurs at the low initial frequency of 1.9% but only remained detectable in lineage C. Of the initial variants undetectable in the final serially egg-passaged populations, there was no apparent relationship between their initial and final frequencies; i.e., if a variant initially exists and remains within the population, it does not necessarily reach fixation. This capability for an intermediate frequency variant to exist within an attenuated population could suggest an intrinsic ability to retain a virulent nucleotide at a given position. However it must be mentioned that the available sequencing coverage (depth) and detection of low frequency variants versus potential error by variant callers would allow for potential intersections to be missed.

Table 3.3 Resultant frequencies of minor variants detected in the original population[a]

							Variant Frequency (%)				
Genomic Location	**Position**	**Ref.**	**Alt.**	**NS/S**	**Ref. AA**	**Alt. AA**	M41-CK	Lineage A	Lineage A1	Lineage C	Lineage D
Nsp4	8520	U	C	S	F	-	32.3	94.3	NV	74.8	82.6
Nsp8	11525	A	G	NS	H	R	1.9	NV	NV	1.8	NV
Nsp12	12744	G	A	NS	D	N	2.4	NV	NV	NV	16.4
Nsp13	16148	G	A	S	R	-	1.4	NV	NV	NV	NV
Spike	21691	A	G	NS	T	A	2.3	NV	NV	NV	NV
Spike	22224	G	U	NS	K	N	19	NC	NV	50.3	42.9
Spike	22920	G	C	NS	M	I	19.8	47.4	NV	54.5	50.8
Envelope	24297	U	C	NS	F	L	21.2	NC	100	99.3	99.3
Membrane	24887	C	U	NS	T	T	1.5	NV	NV	NV	NV

a A total of nine sites of variance were identified in the original M41-CK population. Shown is the variance detected at these positions in the final viruses after serial egg passage. Positions where variants could not be detected are marked "NV" (no variance detected), whereas positions with insufficient coverage to call variants at are marked "NC" (no coverage). Genomic location, position, reference nucleotide, NS/S, and amino acid (AA) are relative to M41-CK consensus sequence. NS/S, nonsynonymous/synonymous variant. For NS variants, the resulting amino acid change is listed. Ref = reference nucleotide, Alt = alternate nucleotide.

3.3.8. Spike contains the highest number of subconsensus mutations; however, the 3′ UTR and N gene exhibit the highest rate of mutation

Each subconsensus variant was determined as possessing either a nonsynonymous (NS) or synonymous (S) potential. The number of NS and S variants was quantified for each genomic location (Figure. 3.8A) relative to the coding gene length to provide a rate of substitution per nucleotide (Figure. 3.8B).

With the exception of lineage D, the majority of subconsensus variants in each lineage were NS. Spike, nsp3, nsp2, and nucleocapsid had the highest combined total of SNPs, with 21, 19, 13, and 12 SNPs, respectively. For spike, this consisted of 12 NS and 9 S mutations, the highest number of S mutations for a specific region. The highest number of NS mutations was observed in nsp3, with a total of 15. Combined totals of 2 and 5 mutations were identified in the 5′ and 3′ UTR, respectively. For the 3′ UTR these mutations did not occur in conserved stem-loop structure (Dalton et al., 2001).

Relative to this length, the 3′ UTR had the highest rate of substitution, with an average of 0.003077 across all five passages (average = 0.003846 for attenuated viruses). Nucleocapsid exhibited the highest rate of NS change, with an average of 0.001630 (average = 0.002038 for attenuated viruses). Conversely, ORF5b possessed the highest rate of S change at 0.001626 (average = 0.002033 for attenuated viruses). These findings may therefore suggest that either the 3′ UTR more permissible to mutation or that it is linked to attenuation. These traits are not however mutually exclusive.

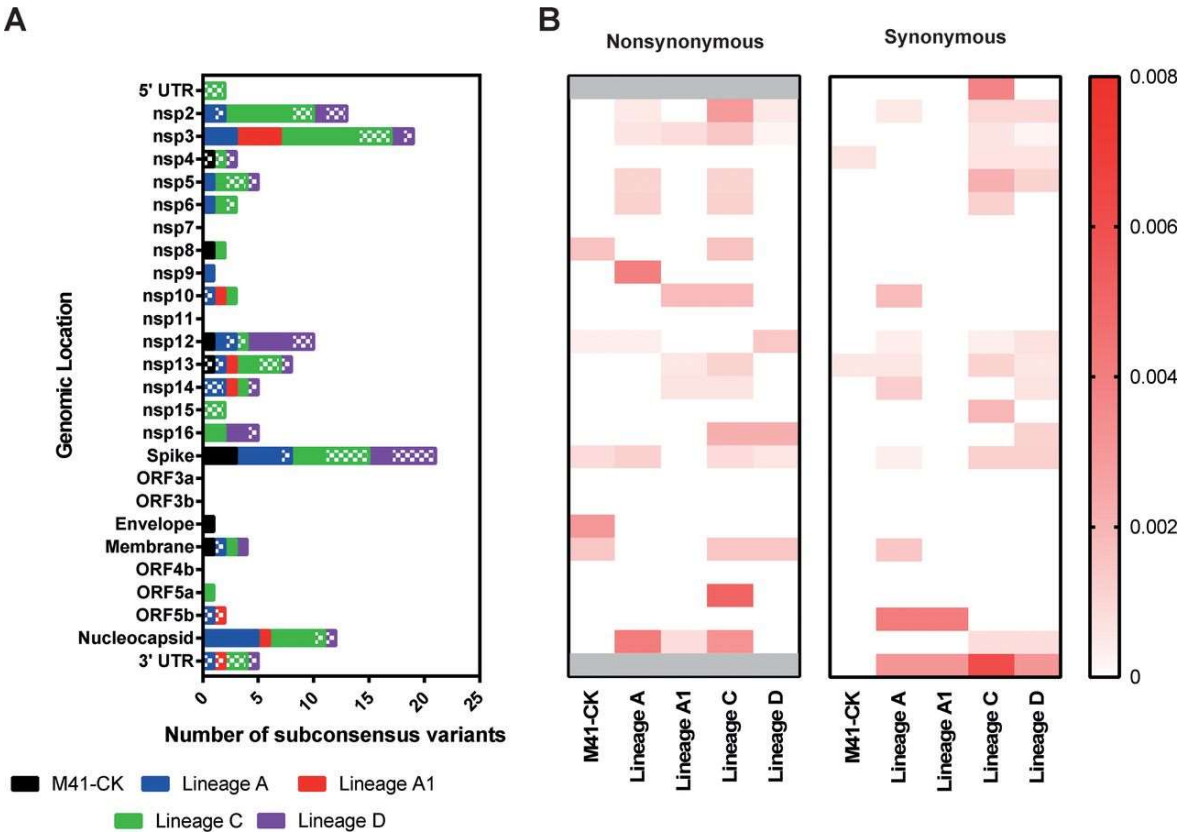


Figure 3.8: Numbers of nonsynonymous and synonymous subconsensus variants per genomic location and their rate of occurrence. *Variants were called in each of the five viruses against the M41-CK consensus sequence. (A) Number of NS and S subconsensus mutations per genomic location. NS mutations are shown in solids, while S mutations are shown in a checkered pattern. Mutations occurring in the 5' and 3' UTR have been plotted as S mutations. (B) Count of NS/S variants was divided by the length of that genomic location to calculate the rate of variation per genomic location relative to length (nucleotides). Variants occurring within the 5' and 3' UTR have been plotted as S mutations and have been shaded out on the NS heat map.*

3.4. Discussion

Serial egg passaging of virulent IBVs to attenuate the virus has long been used as the standard procedure for generating a vaccine, with little to no understanding as to the mechanisms behind the attenuation. With the intention of furthering this field, we serially passaged M41-CK in four separate replicates to the point of attenuation and performed deep sequencing of the final-passage viruses to observe the resulting genomic differences and sequence variation.

Our efforts identified few SNPs shared among all four of the attenuated final-passage viruses at both the consensus and subconsensus level to the required coverage threshold. One mutation (U24297C) was shared (below the threshold for lineage A1) between all four attenuated viruses', C at this position is common between both pathogenic and non-pathogenic publicly available IBV sequences and is therefore unlikely to be involved in attenuation. This therefore indicates that these viruses are not attenuated by the same SNP but does not exclude the possibility that the SNPs are causatively linked. Conversely, through our mutational analysis we have identified regions across the viral genome that undergo a high level of genetic change as the result of passaging and may contribute to attenuation. This has additionally identified areas of the genome where mutation does not occur and may not be evolutionarily tolerated, a feature which will equally assist future vaccine development.

Use of live attenuated viruses as vaccine candidates has long been supposed to have heightened risk of reversion to virulence (Hopkins and Yoder, 1986, McKinley et al., 2008) despite the experimental demonstration of the stability of the clinical attenuation over some *in vivo* passages. While we could not ascertain the impact that each mutation has on pathogenicity in this study, our data demonstrates the possibility for the virus to maintain a virulent genotype within an attenuated population at a given position. This would indicate that the virus has the potential to selectively revert back to its original virulent form through selection of this

subpopulation rather than the evolutionarily more challenging back-mutation to the original virulent genotype.

Compared to the starting population, each attenuated virus shows an increase in sequence diversity at both the consensus and subconsensus levels after serial egg passaging. Due to the nature of serial passaging, it is possible that some observed variation occurred after the point at which these viruses would be deemed attenuated and suitable for vaccine use. Current strategies for characterizing the pathogenicity of IBV isolates largely rely on *in vivo* studies, as limited overlap is observed between *in vitro* and *in vivo* characteristics. Without a suitable alternative for assessing attenuation accurately, *in vivo* characterisation remains the only representative means of assessing loss of pathogenicity. Given the number of samples involved here, it would not be ethical to perform pathogenicity experiments with each individual egg-passaged virus to establish the point at which our serial passages are deemed attenuated, as the harm would outweigh the benefit.

Recent advances in next-generation sequencing technologies provide the opportunity to sequence at a higher depth of coverage and the ability to identify ultralow-frequency mutations within a viral swarm. Surprisingly few variants within our virulent population were detected, with fewer still shared between the virulent and attenuated populations. This would first indicate that *de novo* mutation, as opposed to the selection of pre-existing attenuating mutations (within the virulent population), is the cause of attenuation in this instance. Due to the coverage profile of our data set the detection of diversity across the entirety of the genome, particularly in replicase and spike proteins of lineages A and A1, is limited. A consequence is that all mutations occurring in multiple lineages may not have been fully identified and detection of diversity in A and A1 is not wholly representative as a result. To fully answer this question of selection versus mutation, sequencing of intermediate passaging at a higher depth of coverage provided by current high-throughput sequencers would be required to

show the change in frequency of variants at a higher temporal resolution (Chapter 4). Moreover, repeating the process of attenuation with a virus clone will also aid identification of attenuating mutations (Chapter 5). Such cross-sectional analysis would provide insight as to how the viral population and viral clone evolves over the course of passaging.

Chapter 4 – Exploring how an IBV population evolves during attenuation by egg passaging

4.1. Declaration

Some of the contents of this chapter have been included in manuscripts currently in preparation. At the time of writing, it is the intention to submit this manuscript under the working title "Selection or Mutation: Characterising the role of Selection in Viral Attenuation and in the Generation of Live Attenuated Vaccines" to Genome Biology. The authors of this publication are as follows;

Graham L. Freimanis, Michael S. Oade, Sarah Keep, Luca Ferretti, Andrew Beck, Paolo Ribeca, Richard J. Orton, Harm J. Geerligs, Paul Britton, Erica Bickerton, John A. Hammond

The work presented here is primarily the author's sole effort with contributions from colleagues as listed below.

- Generation of sequence data was performed in conjunction with Graham Freimanis.
- Pipeline for data analysis was developed in conjunction with Graham Freimanis and Richard Orton.

4.2. Introduction

Strategies for assessing the pathogenicity of IBV are limited and primarily involve *in vivo* studies due to limited overlap between *in vitro* and *in vivo* characteristics. It has been noted in-house that the Beau-R attenuated strain of IBV is unable to produce progeny virus when grown *in vitro* at a non-permissive temperature. Unlike humans whose core temperature is 37°C, chickens maintain a warmer core temperature of 41°C however the flow of air at the respiratory interface will result in this region being slightly cooler *in vivo*. With typical cell culture techniques and the incubation of embryonated eggs performed at 37°C there is

the possibility of selective bottlenecks related to temperature. It is therefore suggested that attenuation is linked to temperature sensitivity as the progression of infection of a temperature sensitive (TS) virus within a host would be restricted due to a lost capability to replicate at the core temperature. These temperature sensitive viruses are hence unable to disseminate beyond external-facing tissues (i.e. eyelid, beak and trachea) and reach other known internal sites of replication such as the kidneys and oviduct.

Temperature sensitivity could help overcome one of the shortcomings of chapter 3 in that some of the genomic variation observed may occur after the point at which egg-adapted M41-CK isolates may be deemed attenuated. Deep sequencing the starting inoculum and final attenuated isolates offers a narrow view on what is undoubtedly a complex attenuating process. To pinpoint the passage in which the viruses would deemed attenuated, intermediate passages would need to be characterised by *in vivo* experimentation which is not ethically viable given the number of samples involved. If however temperature sensitivity is a representative measure of pathogenicity *in vivo*, this would present a potential means of characterising IBV pathogenicity *in vitro*. Performing HTS guided by this information would subsequently aid the identification of attenuating mutations. In light of these prospective benefits, the growth characteristics of serially egg-passaged M41-CK (Chapter 3) at 37°C and 41°C has been investigated.

While temperature sensitivity would provide a convenient means of establishing virus pathogenicity, it is prudent to explore other lines of investigation regarding genomic variation occurring over the course of attenuation. The origin of attenuating mutations caused by serial passaging has long remained a subject of contention within the field of IBV. RNA viruses such as IBV exist as a collection of genetically related-viruses possessing a diverse array of mutations (reviewed by Lauring and Andino (2010)). These viruses interact on a functional scale and collectively contribute to the phenotype of the virus (Vignuzzi et al., 2006). It has previously been suggested that evolution of IBV is the result of two factors; 1)

generation of genetic diversity and 2) selection of pre-existing diversity (Toro et al., 2012, Mayr, 1988). It is not clear as to why the process of serial egg passaging works as a means of generating an attenuated virus. It is currently argued that attenuation is either driven by the selection of attenuated viruses present as a subpopulation within virulent IBV or that the process of egg-passaging itself facilitates generation of mutations of which a proportion are attenuating.

To further our understanding, the previous chapter (Chapter 3) utilised a pathogenic population to initiate passaging of four independent replicates. Utilising a virus population as starting material for this study afforded the opportunity for virus to evolve both by selection of genetic variants pre-existing within a virulent IBV population and to evolve by generation of *de novo* mutations. Given that few variants were shared between the virulent and attenuated populations, the results of Chapter 3 indicated *de novo* mutation as a possible cause of attenuation. The coverage of the generated 454 dataset limited the conclusively of this finding pending further investigation. Higher throughput sequencing methods have therefore been employed here to resolve low coverage regions in the 454 data and allow the identification of minor variants (including indels) at lower frequency. Furthermore, while pinpointing attenuated passages within serial passaging of M41-CK by biological methods would aid identification of attenuating mutations, there is still information to be learned in relation to the order of mutation appearance and the evolutionary trajectories leading to attenuation. This is especially important for surveying the revertant potential of mutation given the concern of using live-attenuated IBVs as vaccine candidates. Every 10^{th} passage of the M41-CK serial passaging for each lineage has therefore been deep-sequenced to offer a cross-sectional / intermediate viewpoint as to how a viral population evolves over the course of passaging.

This chapter will therefore explore the possibility of using temperature sensitivity as a means of characterising IBV pathogenicity and survey how a virulent population of M41 (M41-CK) evolves during the process of attenuation by egg-passaging.

4.3. Results

4.3.1. Characterising egg-attenuated M41-CK viruses by *in vitro* temperature sensitivity

To investigate the possibility that IBV attenuation is associated with temperature sensitivity, CKCs were infected with 5 x 10^4 PFU of each egg attenuated M41-CK virus (final passages of M41-CK lineage A, A1, C and D), H120 (a vaccine strain of IBV against Massachusetts serotype generated by serial egg-passaging) or Beau-R (the molecular clone of Beau-CK, an apathogenic strain of IBV), and incubated at either 37°C or 41°C. Supernatant from infected cells was harvested at 1, 24, 48, 72 and 96 hours and titrated on CK cells (at 37°C) to assess viral progeny (Figure 4.1). The initial replicate of these experiments were performed with assistance of Sarah Keep.

The growth kinetics of egg-attenuated IBVs (including H120) were shown to possess broadly similar growth kinetics to M41-CK at 37°C with titres peaking at either 24- or 48-hours post-infection and steadily declining thereafter. Average titres of Lineage A, C and D peaked at 24 hours with 5.75, 5.05 and 4.94 $\log_{10}$ PFU / mL, respectively, whereas the titre of Lineage A1 and H120 peaked at 48 hours post infection with 5.79 and 4.89 $\log_{10}$ PFU / mL. That being said, while titres did peak at these times there is realistically little difference in titre between the two time points for each virus (A, A1, C, D and H120). As anticipated, the average titre of M41-CK peaked at 48 hours (6.52 $\log_{10}$ PFU / mL) and the average titre of Beau-R peaked at 24 hours post infection (6.85 $\log_{10}$ PFU / mL). The growth profile of both control viruses (Beau-R and M41-CK) at 37°C is typical of each virus.

At 41°C Beau-R demonstrates a complete inability to replicate with no virus detectable past the 1-hour time point. For Beau-R the difference in titres at both temperatures between 24- and 96-hours post-infection is highly significant with $P \leq 0.0001$ (Two-way ANOVA, Sidak's multiple comparison test). M41-CK does possess the ability to grow at 41°C with the same growth profile as 37°C albeit

with lower titres. A statistical difference was observed between M41-CK grown at the two temperatures between 24 – 96 hours post infection however the actual titre drop in the order of 10^2 PFU / mL at 41°C versus 37°C and is not comparable to growth (or lack therefore) of Beau-R at the two temperatures.

The growth profile for each egg-attenuated IBV at 41°C is indistinguishable to M41-CK with titres for lineages A, A1 and C at 48 – 96 hours post infection significantly different to its 37°C counterpart. Curiously the growth profiles of lineage A and A1 at both temperatures are near identical despite reported genomic differences (Chapter 3). H120 titres at 37°C and 41°C were significantly different at 48- and 72-hours post infection.

While statistical differences are observed between some of the egg attenuated IBVs at the two temperatures, temperature sensitivity is not a definitive measure of attenuation. Lineage D for example shows no significant difference in titre between the two temperatures at any time point. Additionally, the shift in growth profile for attenuated viruses M41-CK Lineage A, A1 and C, and H120 between the two temperatures are also all comparable to the drop seen in virulent M41-CK and not comparable to Beau-R. Temperature-sensitivity therefore cannot be used to accurately nor reliably characterise *in vivo* pathogenicity and suggests that attenuation of H120 & M41-CK Lineages A-D is not associated with temperature sensitivity. These findings instead imply that attenuated viruses are not necessarily temperature sensitive.

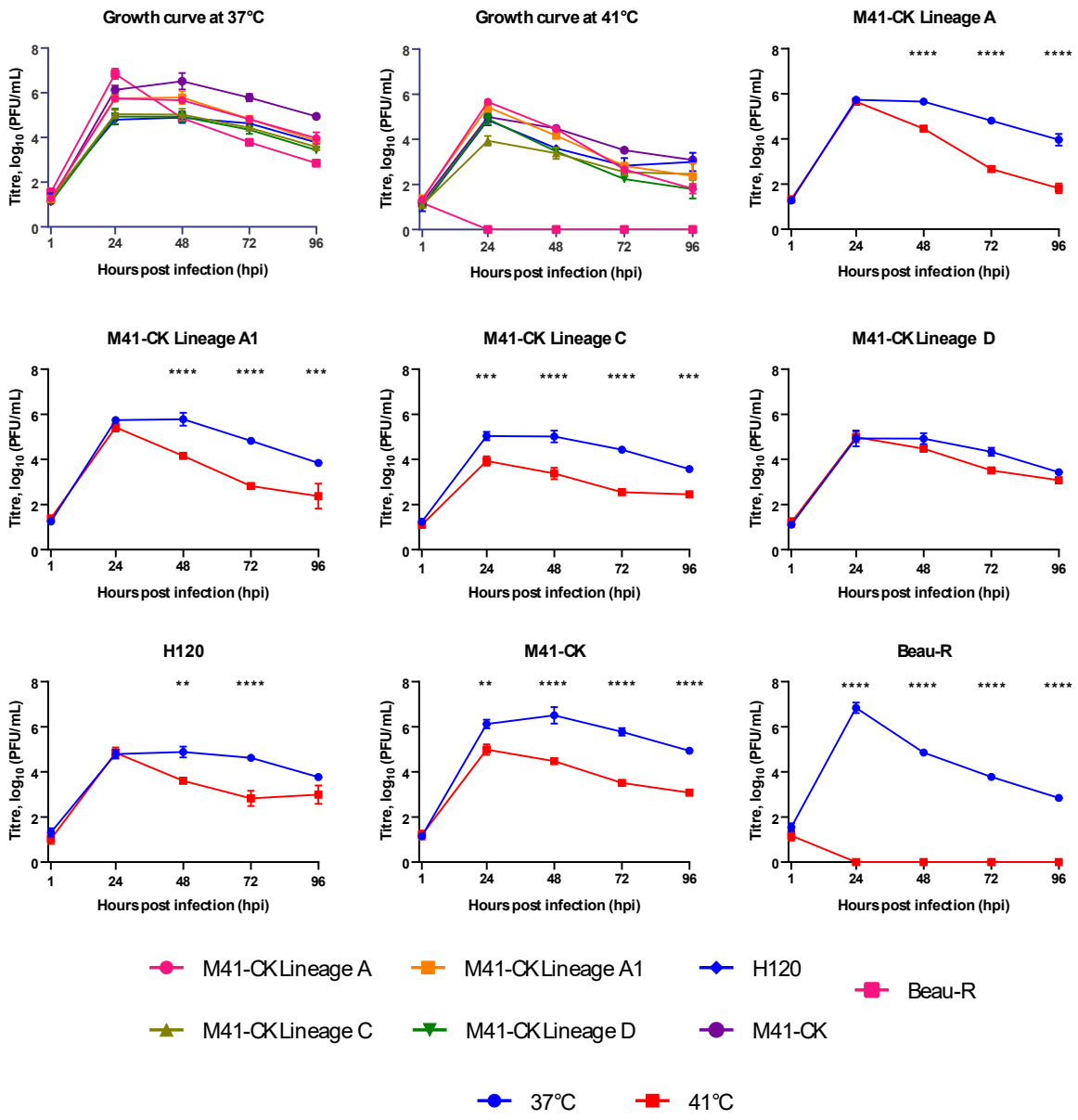


Figure 4.1: Growth kinetics of virulent and attenuated IBVs at 37°C and 41°C on CKCs. *Chick kidney cells were infected with 5 x 10^4 PFU IBV and incubated at either 37°C or 41°C. Supernatant was harvested at 1, 24, 48, 72 and 96 hours post infection and titrated on CKCs. Each value represents the means of three replicates with the SEM plotted as error bars. Graphs have been presented for each virus at both temperatures, and both temperatures with different viruses. Both represent the same data. P values were determined by Sidak's multiple comparison test (Two-way ANOVA). *, P ≤ 0.05; **, P ≤ 0.01; ***, P ≤ 0.001; ****, P ≤ 0.0001*

4.3.2. Assessing inherent M41-CK diversity by HTS.

Given temperature sensitivity was not a viable means of characterising IBV pathogenicity, it was decided to explore the mechanisms behind the attenuation process by sequencing intermediate M41-CK egg-passages. Prior to this, allantoic fluid used to initiate M41-CK serial passaging (Chapter 3) was deep sequenced using Illumina HiSeq to verify the conclusions of the 454 analysis and to provide a higher coverage dataset for identification of lower frequency variants with higher confidence. Advances in HTS technologies and ability to sequence with a lower RNA input here allowed for the direct ancestor of the egg-attenuated M41-CK viruses to be sequenced in contrast to the 454 analysis where an additional stock of M41-CK had to be grown.

Samples were mainly prepared by Graham Freimanis (Pirbright High Throughput Sequencing unit) assisted by Michael Oade. A pipeline for the analysis of all Illumina-obtained data, the results of which are described in the following chapters, was developed in collaboration with Graham Freimanis (Pirbright Institute). Process automation and script writing was performed by Michael Oade.

A *de novo* consensus sequence for M41-CK was generated from pass quality filter (PQF), non-host derived reads using SPAdes. Reference-based assembly was performed with resulting contigs for M41-CK aligned to M41-CK reference sequence generated in Chapter 3 (NCBI accession number: MK728875). As explained in Material and Methods (see 2.9.4), multiple replicates of SPAdes assembly were performed and concatenated based on similarity. The resulting sequence (used for future M41-CK analyses) is 27482 bp long, 7 bp longer that the 454 reference sequence (27475 bp). This is due to the differences in the length of the poly (A) tail and the resolution of the 5′ terminus. No SNPs or indels were present.

Variant-calling for M41-CK identified a total of 29 variants above the minimum detection threshold (Figure 4.2) confirming 8 of the 9 single-nucleotide variants (SNVs) detected in the 454 data (Table 4.1). The majority of these 8 SNVs occur at similar frequencies in the two technologies with an average frequency difference of 0.04. The highest frequency discrepancy between the technologies was 0.125 for G22916C. A11525G was not detected in the Illumina HiSeq data. The differences in frequency (and detection of variants) may be explained by the employment of different sequencing chemistries (each with their own bias), the depth of coverage of the available dataset or by the single egg-passage difference between the content sequenced.

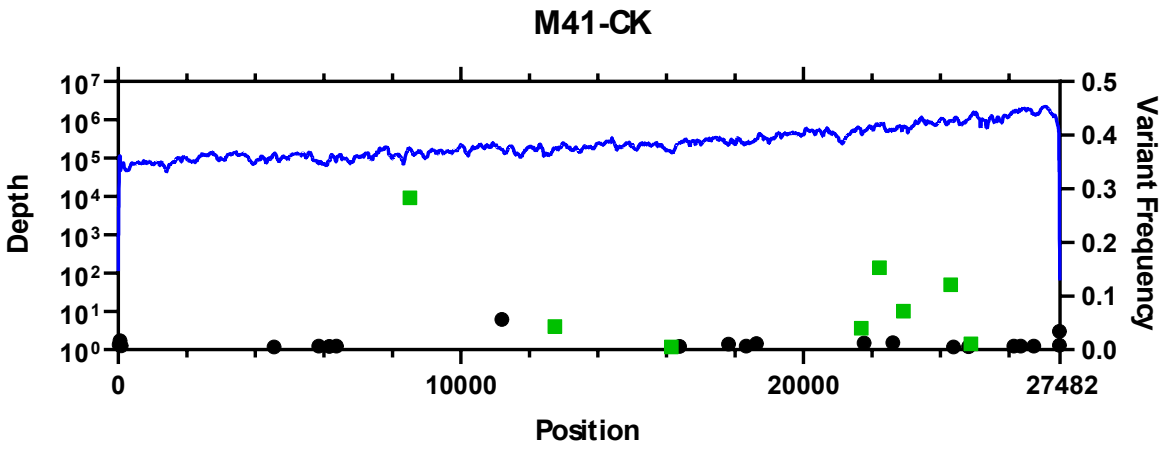


***Figure 4.2: Distribution of variants in M41-CK parent virus**. Reads for M41-CK were aligned to its own consensus sequence. The blue line, plotted against the left-hand y-axis indicates sequencing coverage of aligned reads across the genome. Variants were called using LoFreq and plotted using either black circles or green squares against the right-hand y-axis. Here, variants indicated by green squares were also detected in 454 analysis whereas those indicates by black circles were only detected in the Illumina HiSeq analysis.*

Table 4.1 Variants detected in the M41-CK 454 analysis and their respective frequencies of the M41-CK Illumina HiSeq analysis[a]

Position		Nucleotide		Variant frequency	
Pyrosequencing	HiSeq	Reference	Alternate	Pyrosequencing	HiSeq
8520	8516	T	C	0.323	0.283
11525	11521	A	G	0.019	NV
12744	12740	G	A	0.024	0.043
16148	16144	G	A	0.014	0.005
21691	21687	A	G	0.023	0.040
22224	22220	G	T	0.190	0.153
22920	22916	G	C	0.198	0.073
24297	24293	T	C	0.212	0.121
24887	24883	C	T	0.015	0.011

a NV = No variant detected.

4.3.3. Calling variants in intermediate and final M41-CK passages

After generating an appropriate reference sequence for the parent M41-CK virus, reads obtained from intermediate (passage 10-100) and final (passage 106) passages were mapped to this M41-CK reference sequence and variants identified (Figure 4.3). Though there is an overall accumulation of variants in each lineage by the final passage, sequencing the intermediate stages reveals that this accumulation is not continuous nor indefinitely increasing per passage (i.e. is it possible for previously established variants to be lost in successive passages). With exception to some variants fluctuating near the consensus-level threshold, the number of consensus mutations increased for all lineages as passaging continued.

Both lineages A and C present a similar pattern in that there is a dramatic increase in the number of variants in the initial passages (circa passage 40/50) with variation stabilising to a certain extent until the final passage. Of note here is the possibility for the total number of called variants to decrease between successive passages. For example, a net loss of 20 subconsensus variants occurred between passage 50 and 60 of lineage A. The pattern for lineage A and C in that the number of variants plateaus after 40/50 passages largely holds true for lineage D, however a drop in the number of variants at passage 50 may indicate two declines in total number of variants whereas A and C both have one (Figure 4.3).

Lineage A1 however presents an entirely different profile compared to the other lineages in both terms of the number of variants identified and the frequency at which identified variants occur. Most distinctively passage 30 of lineage A1 has a near total collapse of subconsensus variation, contracting to below the number of variants present in the original population (Figure 4.3). Interestingly while subconsensus variation is near non-existent (n = 4), lineage A1 has 10 consensus-level mutations at passage 30, far more than the other lineages which have 4, 4 and 1 consensus-level mutations respectively at this point in the passaging. Additionally variant frequency in the final lineage A1 passage is more polarised

compared to its other lineage counterparts which have variation spanning the frequency spectrum.

While the possibility of sequencing error for passage 30 cannot be excluded, the succeeding passages also have fewer sub-consensus variants compared to the other lineages the same point in passaging. Moreover, the observation that lineage A1 is unique compared to the other lineages is a theme common both to this analysis and to the 454 analysis (Chapter 3). It is not inescapable that lineage A1, which may otherwise be described as trend-breaking, was generated as a result of splitting lineage A into two separate lineages at passage 13. However, there was no procedural difference in how lineage A1 was established compared to the regimented passaging of lineage A. The difference in profile between the different lineages (including between lineage A and A1) may therefore be the result of varying degrees of bottlenecking and evolutionary pathways. Conversely, the pattern observed in lineage A1 and the near total collapse in subconsensus variation may be common to all four lineages but sequencing every 10^{th} passage has resulted in it being missed for the other lineages.

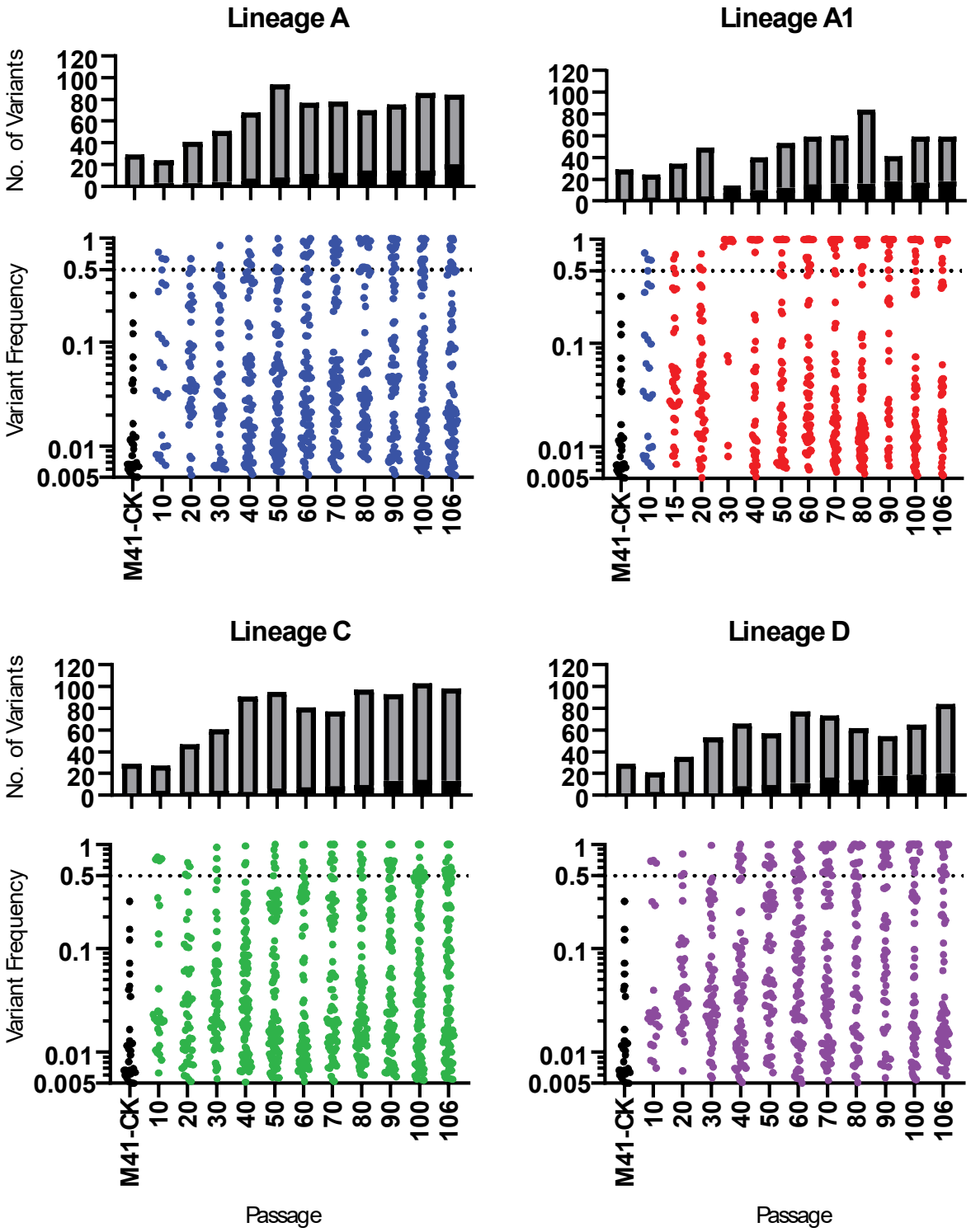


Figure 4.3: Variant plots for individual passages of M41-CK. *Continued next page…*

Figure 4.3: Variant plots for individual passages of M41-CK. *Reads for each individual passage were aligned to the M41-CK consensus sequence. Variants were identified and plotted per passage (see scatter plots below each bar graph). Each individual dot represents a separate variant with the dotted line at 0.5 (or 50%) representing the consensus level threshold. The number of variants occurring above this threshold (i.e. a consensus-level change) is plotted as black in the bar chart above each plot whereas the grey bar indicates variants below this threshold (i.e. subconsensus). Lineage A Passage 10 has been included in the Lineage A1 panel as this is a shared ancestor of the two lineages.*

4.3.4. Surveying dynamics of M41-CK variants

To study the dynamics of variants present in the original population, the eight highest frequency variants were selected and traced over the course of passaging (Figure 4.4). These graphs represent core patterns representative of the entire M41-CK dataset.

In concurrence with the 454 analysis, variants existing within the virulent population may be lost after serial egg passaging and are no longer detectable in the attenuated population. A21687G and A27471T are two examples of this occurring and have initial frequency of 0.0398 and 0.0342, respectively (Figure 4.4 D and H). Equally, variants present in the virulent population may also exist in the attenuated population either remaining at a low frequency or increasing in frequency. The C11190CT mutation (Figure 4.4B) for example remained at low frequency in the final passages whereas T24293C (Figure 4.4G) increased in frequency and approached fixation.

Sequencing the intermediate passages shows here that while the T24293C mutation does approach fixation by the final passage, the point of passage of fixation and intermediate passage frequencies are not synchronous across all lineages. Moreover, changes in variant frequency between the initial and final egg passage are not wholly representative of the changes that have occurred with variants capable of both increasing and decreasing in frequency over the course of passaging. T24293C results in a nonsynonymous mutation in M41-CK that results in uniformity to both pathogenic and non-pathogenic IBVs at this site. That is, M41-CK is the odd one out compared to other IBVs and the change in egg-passaged isolates brings it back to what other IBVs have at this site. The preference of the C at this site suggests that it is otherwise disadvantageous to possess another nucleotide but the various lineages take differing times to establish this mutation. Indeed this data demonstrates that even if a mutation is present in the final passage, the profile of the variant in the intermediate passages is unique to that

lineage, as exemplified by T8516C (Figure 4.3A) and G12740A (Figure 4.3C). This embodies a need to observe the virus over the course of passage history as the initial and final viruses are not wholly representative as to the change has occurred.

That being said, the profile of a variant is not necessarily unique to that variant. G22220T (Figure 4.4E) and G22916C (Figure 4.4F) are two separate mutations which appear in successive passages at remarkably similar frequencies and will be discussed further later.

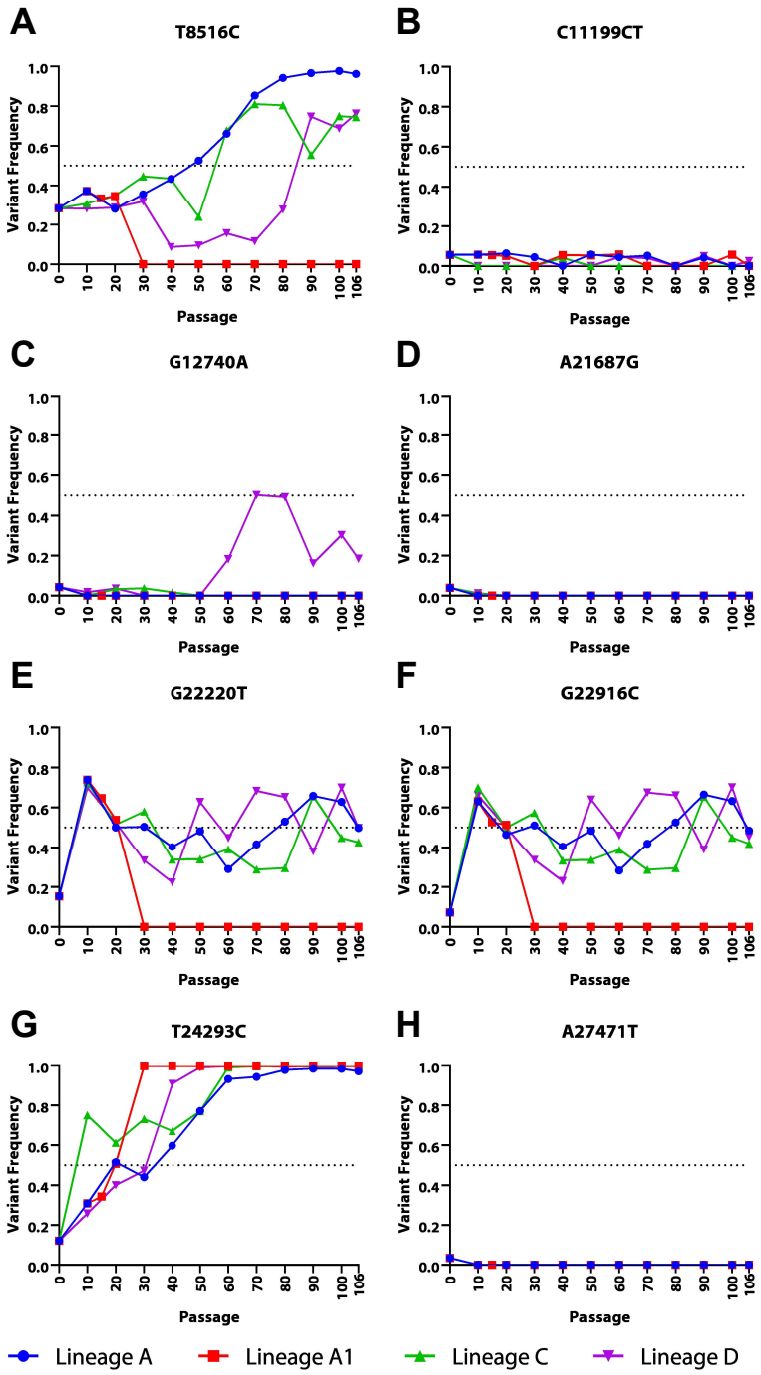


Figure 4.4: Variant frequency plots monitoring intermediate frequencies of variants present in the original population. *The eight highest frequency variants of the initial population were selected and their frequency traced over the course of passaging of each lineage. These plots are representative of the dataset and demonstrate the variability that occurs during the course of passaging.*

4.3.5. Identifying variants shared between multiple M41-CK passages

To verify the findings of the 454 data, variants identified in each of the attenuated viruses and the M41-CK virulent parent were compared. Confirmatory to the 454 analysis few consensus or subconsensus mutations are shared across multiple lineages in the final passage viruses (Figure 4.5). The vast majority of mutations (72 – 77 %) remained unique to each individual lineage (74/97, 43/59, 75/98 and 63/84, respectively). Furthermore, of the 29 variants detected in the original population, only 12 (41.4 %) remained in at least one of the final passages. It may imply that these 12 mutations have been selected for by the process of egg passaging or conversely that 17 mutations only detected in the original population have been selected against. As described previously the one nonsynonymous mutation (T24293C) present as a minor variant in the M41-CK parent and that reached consensus in all final attenuated isolates is common to both pathogenic and non-pathogenic publicly available sequences and is therefore unlikely to be involved in attenuation.

Given that the lineage A and lineage A1 are more closely related by passage history compared to the other lineages, it would perhaps be expected that the final viruses from these two lineages would share more variants with each other than other comparisons. This is partially reflected in the data with 4 variants shared between attenuated lineage A and A1, the highest number of any two-way comparison. However, in partial contradiction to this, three-way comparisons between lineages A, C and D have a total of five shared variants where lineage A1, C and D do not share any variants.

All four attenuated viruses possess three variants (T25457TAA, C26841T, and G27183A) not present in the original virulent population. While the frequency of G27183A is consistently low across all four lineages (min = 0.006724, max = 0.012283), the frequency of the other two mutations is more varied. C26841T approached fixation in lineage A1 (Freq = 0.998691) but only existed at an average

frequency of 0.035645 in the other lineages. Conversely, T25457TAA occurs at a frequency of 0.516209, 0.5009635, 0.013185 and 0.263306 for each lineage respectively.

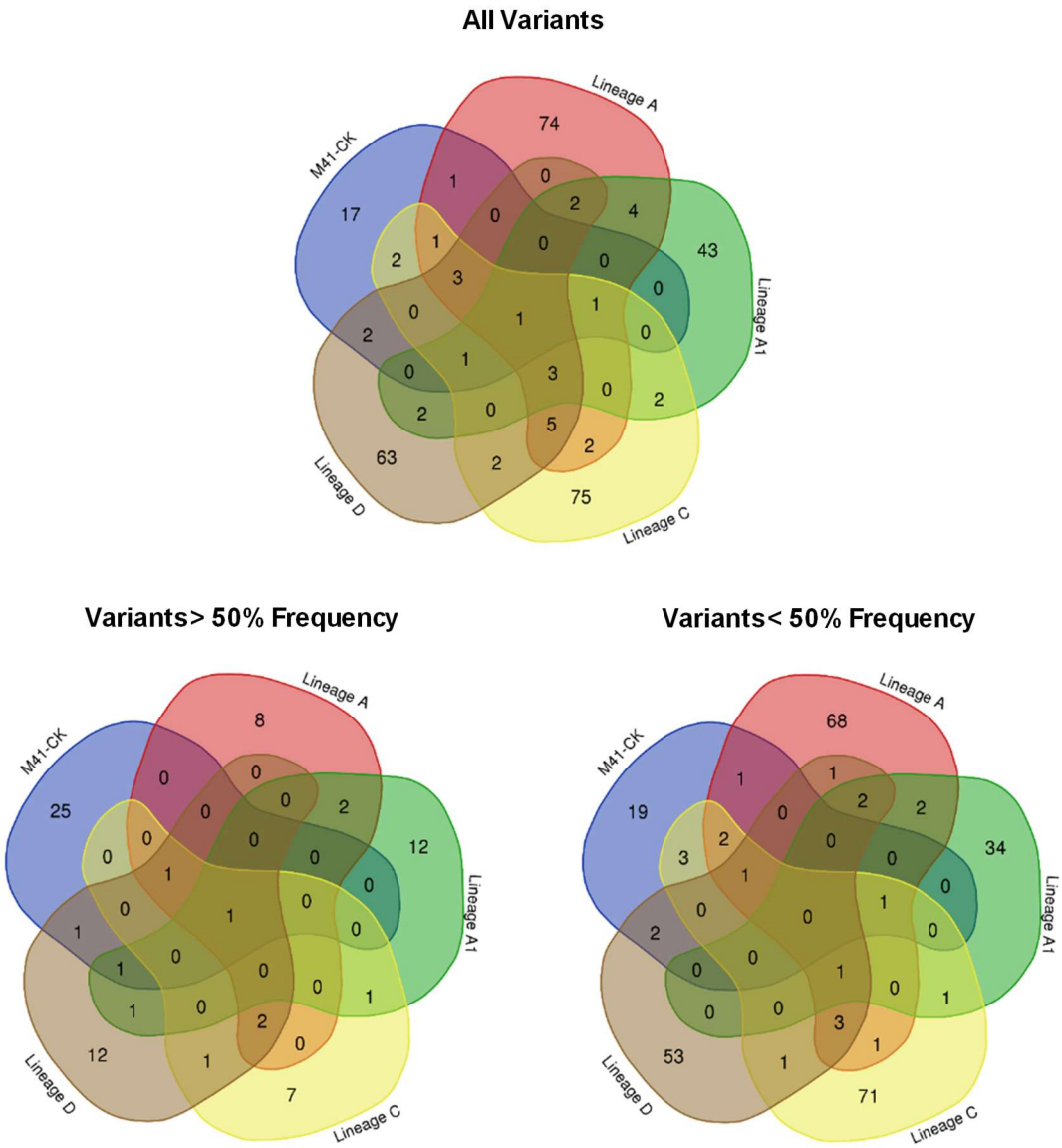


Figure 4.5: Venn diagrams from the comparison of called variants between final M41-CK lineages. *Variants identified in the starting passage and final passages were compared with Venn diagrams produced for all variants, those occurring at > 50 % frequency (consensus) and those occurring < 50% frequency (subconsensus) in the attenuated (final passage) virus isolates. M41-CK parent variants present in the original population are plotted in all three Venn diagrams. Variants occurring outside the blue area are indicative of de novo mutation (variant was not detected in the starting population).*

To determine both the initial passage at which a variant was detected and to assess possible homology between successive passages, further comparisons within individual lineages were made (Figures 4.6 to 4.9). UpSet plots are used to visualise the relationship between multiple datasets and here used to visualise variants occurring in multiple passages. Firstly and most simply, the left hand horizontal bar graph plots the number of pieces of information (here, the number of variants) per set (or passage). This is equivalent to the bar charts plotted in Figure 4.3. Separately, the lower middle panel consisting of dots and sticks signifies a comparison being made with the bar chart then plotting the number of variants in that comparison. No single piece of information is ever plotted twice. In Figure 4.6 for example 13 variants occur in both Lineage A Passage 100 and 106 and did not occur anywhere else in Lineage A nor M41-CK. This does not however reflect the total number of variants shared between Lineage A Passage 100 and 106 as, for example, a further 5 variants are shared between Lineage A Passage 100, 106 and 90. In cases where there is no connection between dots in the middle panel (i.e. a singular dot), the bar chart presents the total number of variants occurring in that passage and that only occurred in that passage. In Figure 4.6 again for example, 24 variants were identified in Passage 106 and 24 variants were identified in Passage 50 that were unique to that passage. The variants represented in these two bars are entirely different to one another and do not overlap in any way.

It was determined (Figure 4.5) that the majority of mutations identified across the final passage viruses are unique to that lineage. These UpSetR plots (Figure 4.6-4.9) extend this conclusion and indicate that a high proportion of variants in each passage remain unique to that passage in both the intermediate and final passages alike. This demonstrated by the highest bars in the upper bar charts for Figure 4.6-4.9 corresponding to singular non-connected dots in the middle panel. Lineage A1 Passage 80 therefore has highest number of variants unique to that lineage (31 variants, Figure 4.7). The identification of variants unique to a passage suggests a transient nature of individual variants.

The transient nature of variants is also demonstrated by variants occurring in multiple consecutive passages (e.g. n^{th} passage and $n+10^{th}$ passage) but not in the virulent parent or attenuated final viruses. This equally suggests that it is possible for variants to emerge and disappear during the course of passaging. This would otherwise go undetected without sequencing intermediate passages but is common in all lineages at multiple different points in passage and most pronounced in lineage A1 (Figure 4.7) and lineage C (Figure 4.8) .

Intersections between the final viruses and their former passages show a steady accumulation of variants in successive passages until the final two passages. Evaluating variants present in the final virus and every passage after their original detection, lineages A-D have an average gain of 5, 3, 5.4 and 3.6 variants every 10 between passages 10 to 90. Passage 100 for lineages A-D respectively has a total of 13, 17, 20 and 9 variants common to only passage 100 and the final attenuated virus. With yet more variants unique to only the final passages, there is a ramping up of the accumulation of variants in these final two passages.

While it is possible that these patterns presented above do represent genuine trends within the dataset, an individual variant only reported to be in one or few passages may be the result of the of interval of sequencing. Alternatively, it is possible that there is an element of "noise" occurring in each passage with variants captured reflecting a temporary state of flux in the population, not representative of the passaging process. It is also worth mentioning that the absence of a specific variant in the middle of consecutive passages (indicated by a line between two dots in the lower middle panel of each figure) may indicate a temporary dip in frequency of a particular variant below the minimum threshold (and therefore not identified) but may still exist within the population over those passages.

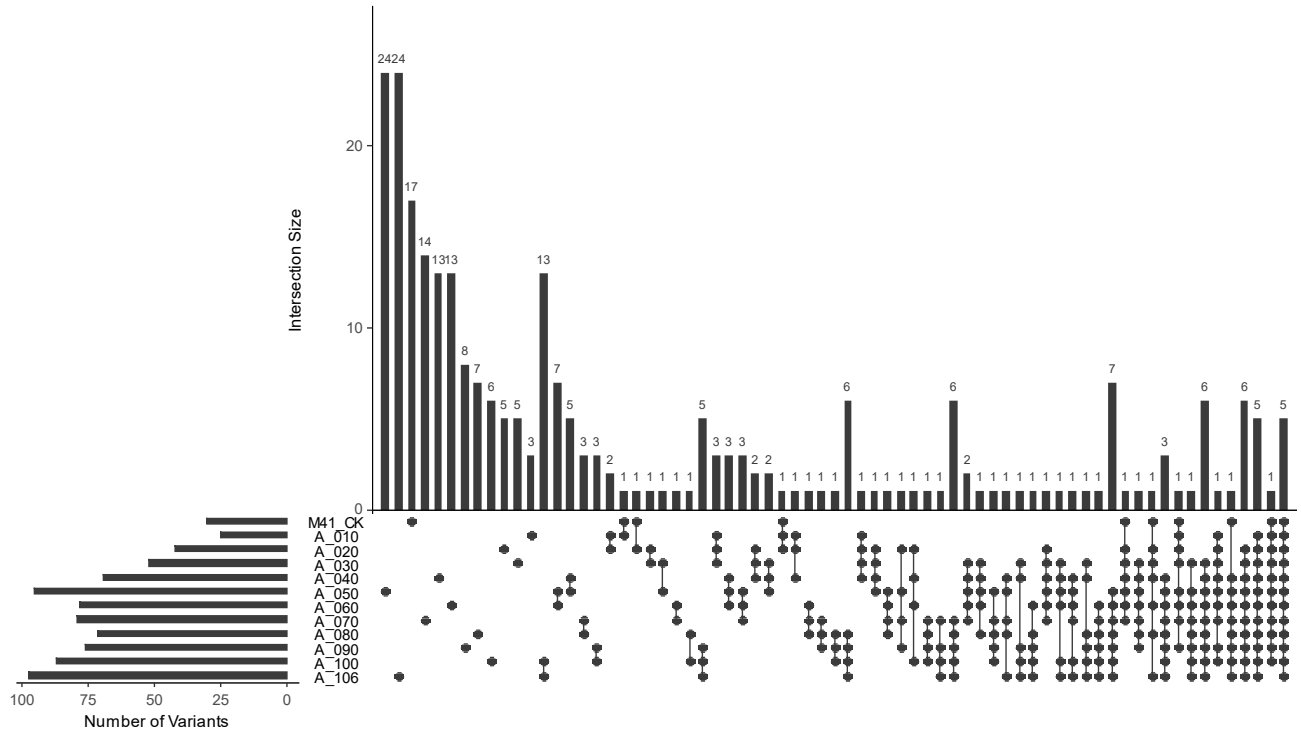


Figure 4.6: UpSetR plot for the comparison of all variants called in multiple passages of M41-CK Lineage A. *The total number of called variants per passage is indicated in the left bar graph. Variants identified in each passage (i.e. a set) was compared to every other passage. The dot and lines plot indicate the comparison being made (i.e. intersection between sets) with the bar chart (top) indicating the number of variants within that intersection. Singular dots indicate variants unique to that passage.*

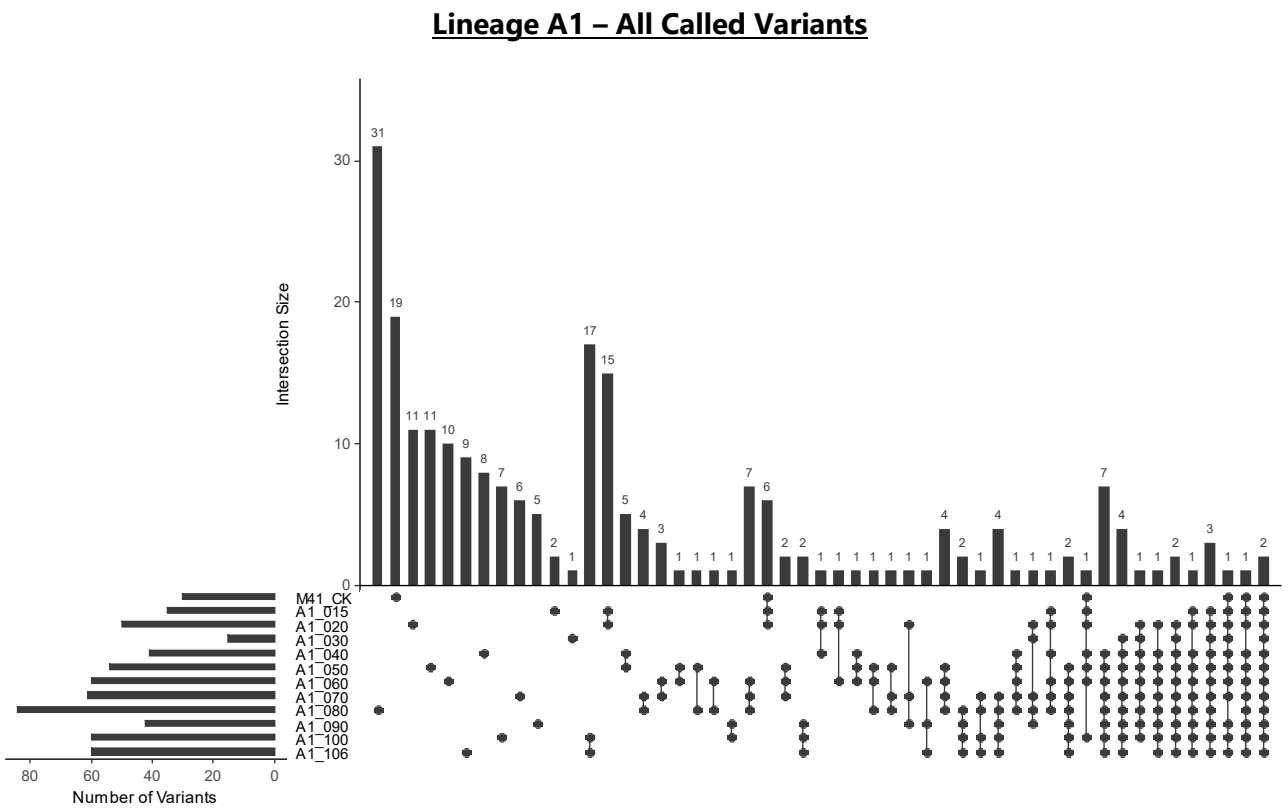


Figure 4.7: UpSetR plot for the comparison of all variants called in multiple passages of M41-CK Lineage A1. *The total number of called variants per passage is indicated in the left bar graph. Variants identified in each passage (i.e. a set) was compared to every other passage. The dots and lines plot indicate the comparison being made (i.e. intersection between sets) with the bar chart (top) indicating the number of variants within that intersection. Singular dots indicate variants unique to that passage.*

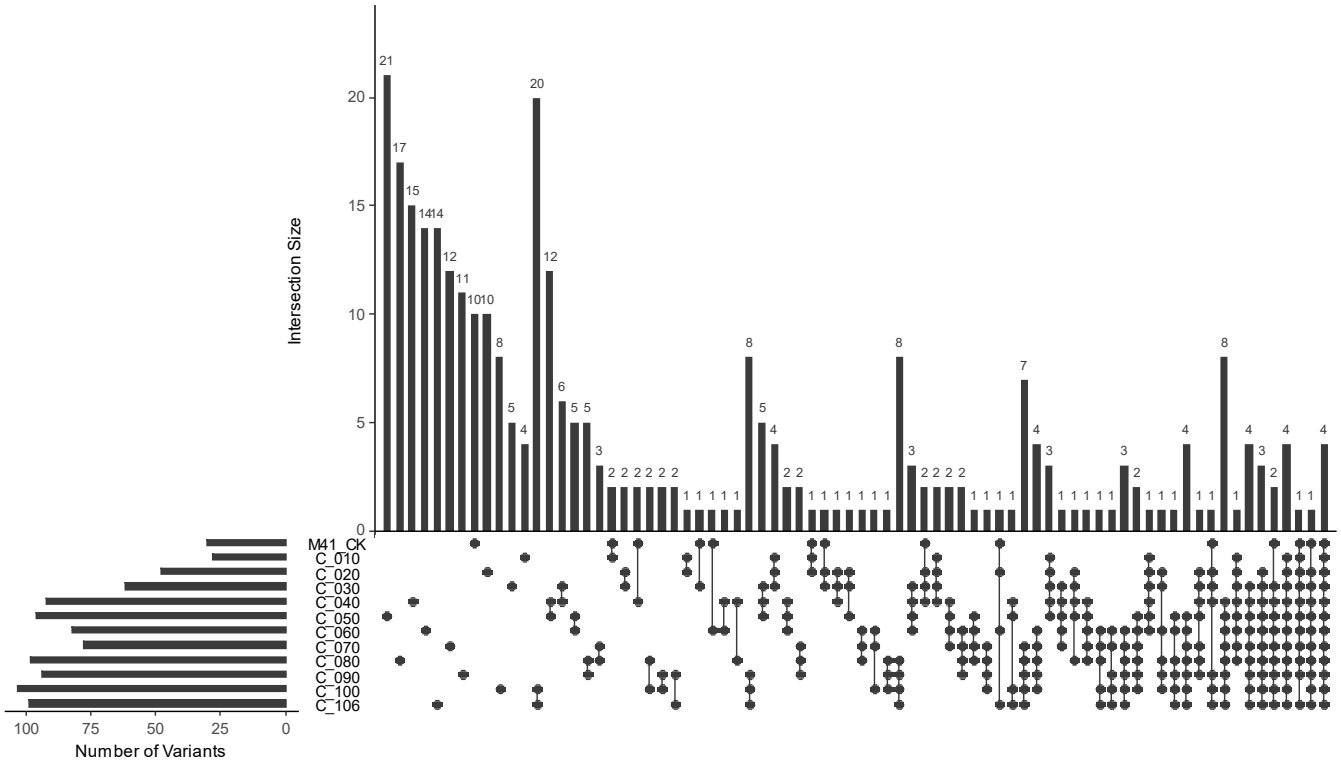


Figure 4.8: UpSetR plot for the comparison of all variants called in multiple passages of M41-CK Lineage C. *The total number of called variants per passage is indicated in the left bar graph. Variants identified in each passage (i.e. a set) was compared to every other passage. The dots and lines plot indicate the comparison being made (i.e. intersection between sets) with the bar chart (top) indicating the number of variants within that intersection. Singular dots indicate variants unique to that passage.*

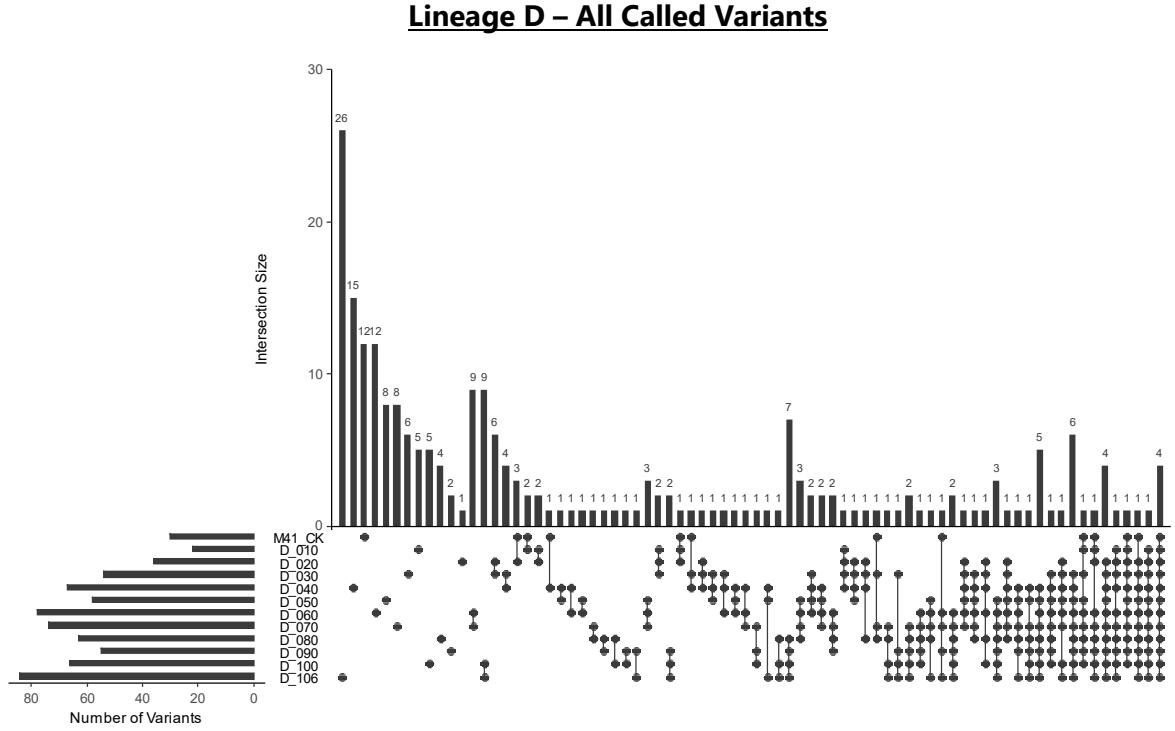


Figure 4.9: UpSetR plot for the comparison of all variants called in multiple passages of M41-CK Lineage D. *The total number of called variants per passage is indicated in the left bar graph. Variants identified in each passage (i.e. a set) was compared to every other passage. The dots and lines plot indicate the comparison being made (i.e. intersection between sets) with the bar chart (top) indicating the number of variants within that intersection. Singular dots indicate variants unique to that passage.*

4.3.6. Co-mutation of G22220T and G22916C in successive passages

In surveying dynamics of parent variants and comparing variants intersections of lineage A, C and D it was noted that the mutations G22220T and G22916C occur at remarkably similar frequencies in the final passages. To confirm this observation, the frequencies of these two mutations were traced throughout the course of passaging (Figure 4.10). Both of these mutations do appear linked with each other with their respective frequencies fluctuating concordantly across multiple passages.

In generating these plots (Figure 4.10), it was observed that in later passages of lineage A G22281T, a variant not detected in the original population, rises in frequency to match the frequency of G22220T and G22916C. Conversely, later passages of lineages C and D indicate that G22236T, also not present in the original population, rise to match G22220T and G22916C. Where the read length allowed, it was confirmed by visual inspection of the alignment that co-mutation was present. Lineage A1 does have G22220T and G22916C mutations which occur at similar frequencies but both mutations are no longer detected within the population by passage 30. Notably, this is same passage where lineage A1 undergoes a near-total collapse of subconsensus mutations (Figure 4.3). The absence of these mutations in lineage A1 does not conflict to the observation made.

All four mutations are nonsynonymous (Table 4.2) and occur within the S2 subunit of spike. The spike protein, consisting of the S1 and S2 subunits, is responsible for mediating coronavirus entry, host range and tissue tropism. The S2 subunit is responsible for driving the fusion between host and viral membranes. As described in 1.5.1 the two subunits interact to prevent switching of the S2 subunit to a post-fusion state until the S1 subunit binds to a receptor.

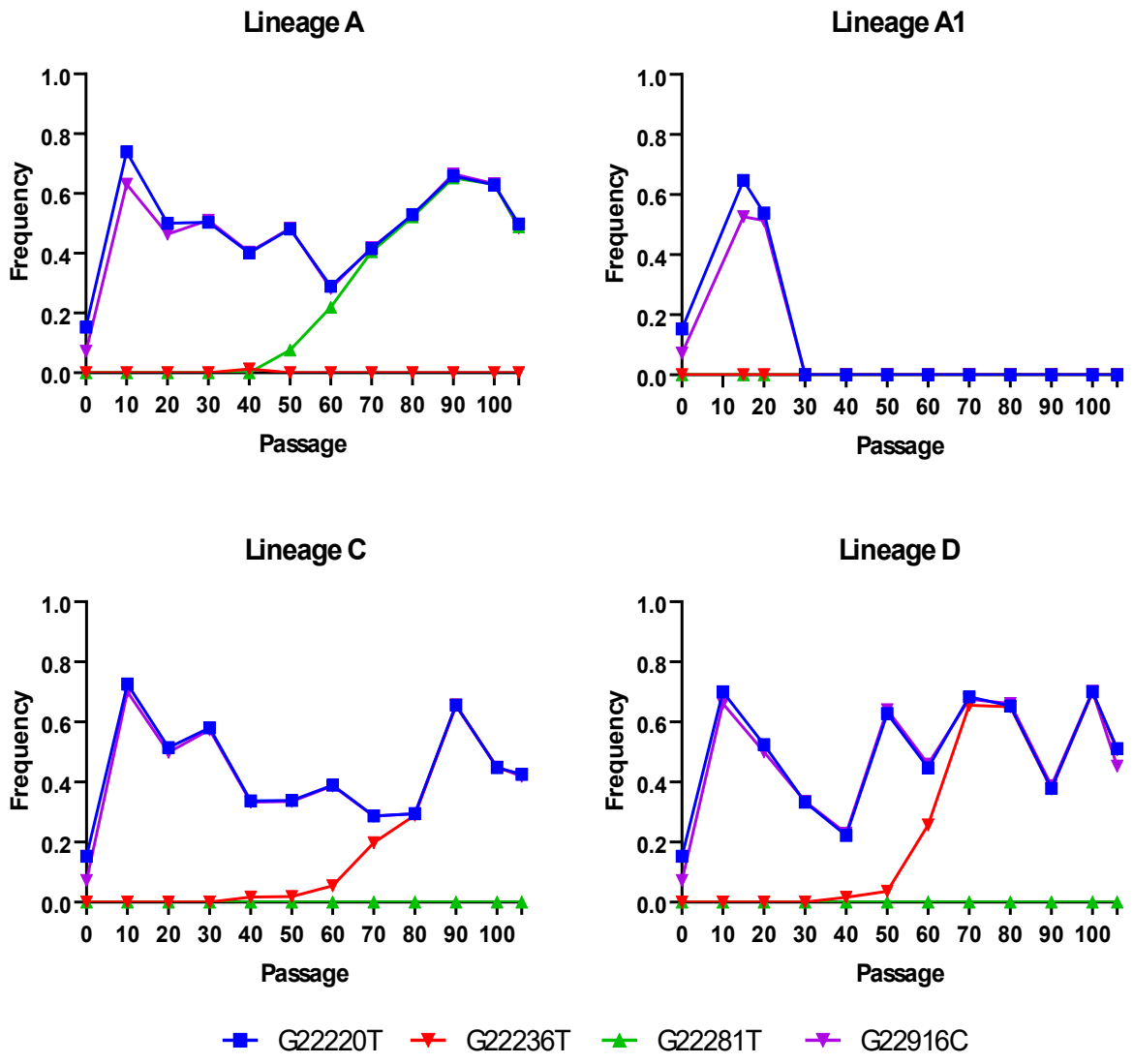


Figure 4.10: Passage versus frequency plots for four nonsynonymous variants with near concordant frequencies over the course of M41-CK passaging*. The frequencies of four variants were traced through the course of passaging due to high similarity in their relative frequency. Each plot presents one of the four M41-CK lineages with each colour representing a specific variant.*

Table 4.2. Co-mutants in Spike S2 and their corresponding amino acid change[a]

Nucleotide Mutation	Amino Acid Change	Grantham Conservation Score
G22220T	K617N	94
G22236T	D623N	22
G22281T	V638L	32
G22916C	M849I	10

[a] Nucleotide positions are given relative to the M41-CK consensus sequence. All amino acid positions are given relative to the Spike sequence. Grantham scores were calculated using CooVar (Vergara et al., 2012). Higher Grantham scores reflect a greater evolutionary distance between two amino acids on a scale of 5 – 215. The higher the score, the more "damaging" the substitution.

Given that these mutations (Table 4.2) appear, in some degree, to fluctuate in frequency correspondingly and that these mutations are all nonsynonymous, it was thought prudent to investigate whether these mutations occur in close proximity in tertiary space. A model of spike for IBV was solved in 2018 (Shang et al., 2018) and obtained from the RCSB PDB (Berman et al., 2000). The protein structure was visualised using Chimera v.1.12 (Pettersen et al., 2004) and the sites of all four mutations (Table 4.2) identified and highlighted on it (Figure 4.11). The model for the IBV spike was solved using material derived from the M41 virus strain and is therefore highly similar to M41-CK.

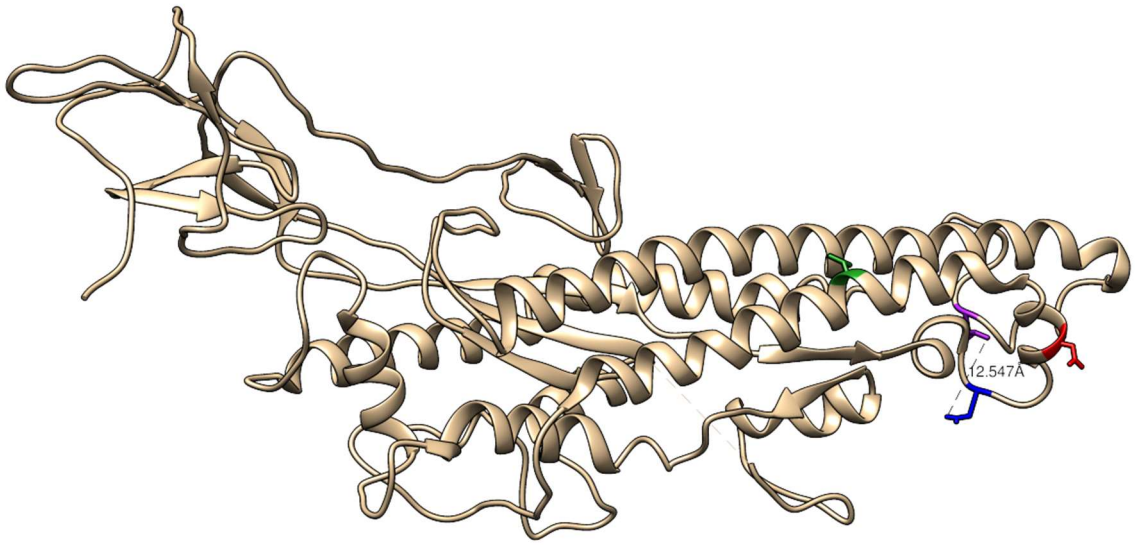


***Figure 4.11: S2 model with sites of co-mutation highlighted**. The infectious bronchitis virus spike protein model (PDB: 6CV0) was downloaded and inputted into Chimera (version 1.12). Here only the S2 portion (by sequence similarity) of one of the three chains is displayed. Amino acid residues corresponding to the sites of co-mutation are presented with atoms/bonds displayed and highlighted as follows; K617N = blue, D623N = red, V638L = green, M849I = purple. All other residues remain uncoloured with atoms/bonds not displayed. K617N and M849I have a nearest atomic distance of 12.547 Ångströms (Å).*

All four mutations occur in the S2 subunit of Spike. Amino acids 617 (Figure 4.11, Blue) and 623 (Figure 4.11, Red) are in proximity due to protein sequence. Interestingly in the model, amino acid 849 (Figure 4.11, Purple) is also in relative proximity to both 617 and 623 in the tertiary model. Conversely, site 638 (Figure 4.11, Green) is not located near this collection of residues.

In the current IBV spike model (Shang et al., 2018) it is not entirely clear as to what structure, if any, the mutations at 617, 623 and 638 occur in. This is because the exact starting location of fusion peptide (FP) within S2 subunit remains unknown. The domain located immediately after FP, heptad repeat 1 (HR1), is believed to start at position 795. The IBV spike publication (Shang et al., 2018) uses the range of MHV FP (Walls et al., 2016) for the purposes of the spike model.

The MHV FP has an estimated length of 84 AAs (Walls et al., 2016) corresponding to position 711 in the IBV spike model. While the range of FP in IBV is unknown, it is unlikely to expand an additional 100 AAs to include all these mutations, especially given the fact that FP is purported to have a length in the region of 15 – 25 AAs (Belouzard et al., 2012). Curiously however the MHV model (Walls et al., 2016) defines an upstream helix (UH) occurring within S2 but upstream (towards N-terminus) of FP. UH is so named because it occurs immediately upstream of the S2′ cleavage site, not upstream of FP. M41 does not possess the S2′ site however.

While a direct comparison cannot be made, it is noted that at 623 (Figure 4.11, Red) and 638 (Figure 4.11, Green) both occur in helixes which, by position, would potentially correspond to UH. In MHV, UH is zipped against the central helix (CH) by hydrophobic interactions (Walls et al., 2016). In trimeric form, the three central helices are packed together with contacts between UH and CH stabilising the pre-fusion confirmation of the S2. Notably these contacts also involve the HR1 region, which the 849 mutation is located in.

Further work would be required to confirm the interaction between these residues and the implications of this. Evidence may here suggest a shift in the stability of pre-fusion Spike conformation as a means of attenuation, at least in the acquisition of G22236T (D623N) and G22281T (V638L). It does not however explain the role of G22220T (K617N) nor as to why the frequency of G22220T (K617N) and G22916C (M849I) appear dependent on each other. This may however present evidence as to the mechanism of S2 driving virus-host fusion and inform on the interaction between trimers in pre- and post- fusion Spike conformations. Alternatively, S2 subunit of IBV has been shown to play a role in host tropism. These mutations may be in response to changing host from CKCs to embryonated eggs.

4.3.7. Pinpointing genomic regions in M41-CK passaging undergoing the highest level of change

As a measure to establish which genomic regions undergo the highest level of change over the course of M41-CK passaging in eggs and therefore more likely involved in the process of attenuation, the number of consensus and subconsensus variants per genomic region was determined. To account for different genomic regions having varying length (and hence a greater possibility of possessing variants), the number of variants for a given region was divided by the total length of that region (Figure 4.12).

Consistently both across multiple passages and multiple lineages ORF4b has the highest number of subconsensus changes per nucleotide. 4b is a 94 AA accessory protein which, contrary to its name, is encoded by its own gene (as opposed to being part of gene 4 (Membrane)). The mutations here correspond to indels occurring at 25319 and 25457 which extend and reduce the length of two poly(A) tracts. In both instances, ORF4b is disrupted. The activity and function of 4b is not well established namely because there are no equivalents in non-avian coronaviruses and it was not until recently that ORF 4b was confirmed to be actively transcribed (Hall, 2017, Dinan et al., 2019). There is evidence that 4b can interact with cellular proteins involved in translation and induce non-canonical stress granule formation (Hall, 2017). Stress granules (SG) are sites of stalled translational complexes and contain both host-cell translational machinery and cellular RNA. Some viruses, including coronaviruses, benefit from SG formation as part of a system that limits host cell protein synthesis (Raaben et al., 2007, Sola et al., 2011). In this instance, it is not immediately clear how abrogation of ORF 4b would assist the virus. Alternatively, there is evidence that formation of SGs acts as a signal to the induce immune response (Tsai and Lloyd, 2014). Again, it is not clear why modifications in ORF4b would be beneficial to the virus, nor how this may link to attenuation.

The concentration of mutations in ORF4b forms part of an overall trend that variation clusters towards the 3′ end of the genome (Figure 4.12). While the detection of more variants in this region may be a by-product of increased depth of coverage (for example, Figure 4.2) possibly caused as result of poly(A) enrichment, this finding is in concurrence with the 454 analysis (Figure 3.8). Equally, the nucleocapsid gene and 3′ UTR are here highlighted as having the highest number of subconsensus changes per nucleotide with a low rate of subconsensus change occurring in the replicase gene. This may therefore suggest that these regions are more likely and less likely to be involved in attenuation, respectively.

The 3′ UTR also has one of the highest rates of consensus and subconsensus mutation consistently across the multiple lineages. Consensus-level changes in the 3′ UTR arise at passage 40, 90, 70 and 50 for lineages A-D respectively. The envelope protein here has a high rate of consensus-mutation due to the single mutation (T24293C) as discussed previously.

It is perhaps surprising that spike protein (S1 and S2), responsible for entry and host range, has one of the lowest rates of consensus and sub-consensus variation. This may in part be the result of a few core mutations occurring within this gene that, when S1 and S2 are divided by the total length scale proportionally lower to other genes. In other words, the influence of a few key mutations is diluted down when taking into consideration the total gene / region length. Indeed spike protein consistently has the highest number of mutations throughout the dataset.

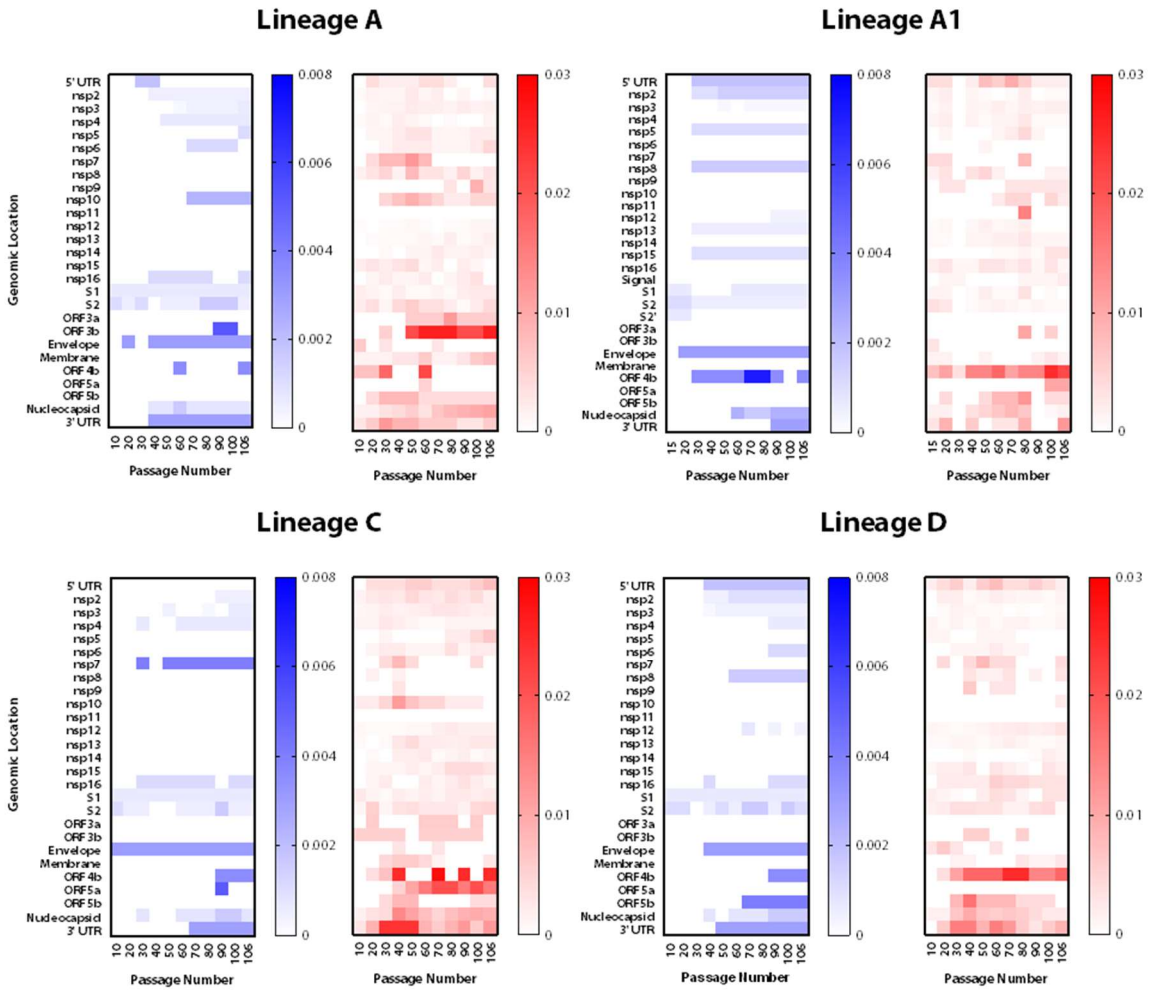


Figure 4.12: Heatmaps of the rate of consensus and subconsensus variation observed per genomic region over the course of M41-CK passaging. *The number of consensus and subconsensus mutations per genomic location was determined and divided by the length of that genomic location to provide a rate of variation per nucleotide. Consensus-level variations are plotted in blue whereas subconsensus variants are plotted in red. Similar graphs have been plotted using the same scale.*

4.3.8. Evaluating insertions and deletions in the M41-CK genome

Given that ORF4b has the highest rate of subconsensus changes compared to other genes in the virus genome, the mutations causing this abnormal rate were surveyed in greater detail. One of the limitations of the 454 analysis (Chapter 3) was that due to concern of homopolymeric error in the sequencing platform, indels were excluded. Here the Illumina data and the generated analysis pipeline permitted the study of indels. It transpires that ORF4b exhibits a high level of insertion at two sites, 25319 and 25457 (Figure 4.13). Here the frequency of variants occurring at theses site for Lineage A each are plotted but represents a trend seen across four lineages with differing variant frequencies at one or both sites.

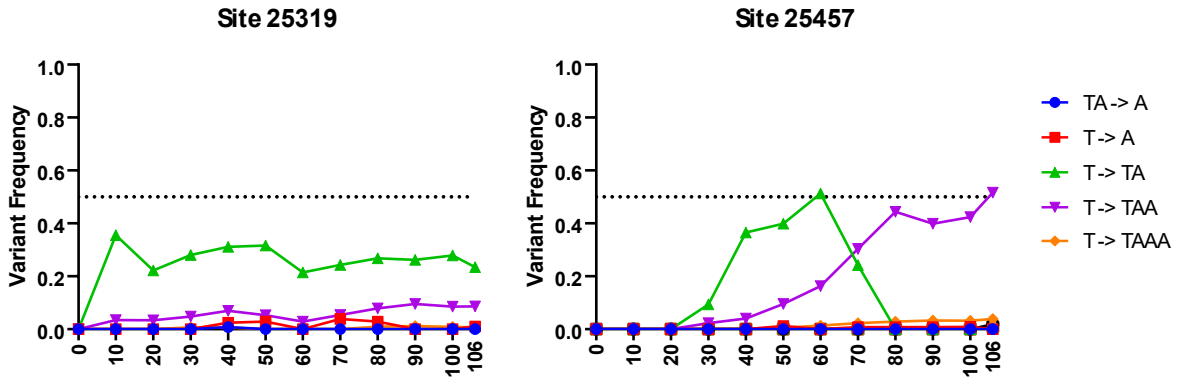


Figure 4.13: Representative variant frequency versus passage traces for 25319 and 25457. *The frequency of variants occurring at sites 25319 and 25457 were monitored over the course passaging. The above plots the frequency variants at these sites for lineage A but is representative of the trend observed across all four lineages in that each lineage has indels at one or both of these sites.*

The reference genome at sites 25319 and 25457 both possess a stretch of 6 adenosines (As) with the mutations reported extending this region further. There was no deviation or variation at these sites in the M41-CK parent genome. It may be theorised that these mutations as reported occur at sites of sequencing error in part due to being a low complexity region. This however does not explain as to why the frequency fluctuates to the degree observed nor as to why there is an apparent increase in variants as passaging continues. Arguably, if extension of this

region were the result of error, it would also have been detected in M41-CK parent. Furthermore, there are 4 other sites (nsp5, nsp9, S2 and ORF3b) of at least 6 As (excluding the poly(A) tail) which do not exhibit the same change indicating this a not a poly(A) specific error.

The significance of these regions to the virus remains unclear. It is however noteworthy that these insertions could alter the expression of ORF4b by frameshift mutation and could result in translation of the protein within a population in the two other forward frames simultaneously (Figure 4.14). Curiously, BLASTP searches of latter portion of ORF4b modified by insertion of two As at 25457 (ORF4b_25457+2) from position AA 94 onward (i.e. the site of insertion) correspond to a conceptualised protein of IBV named 4c.

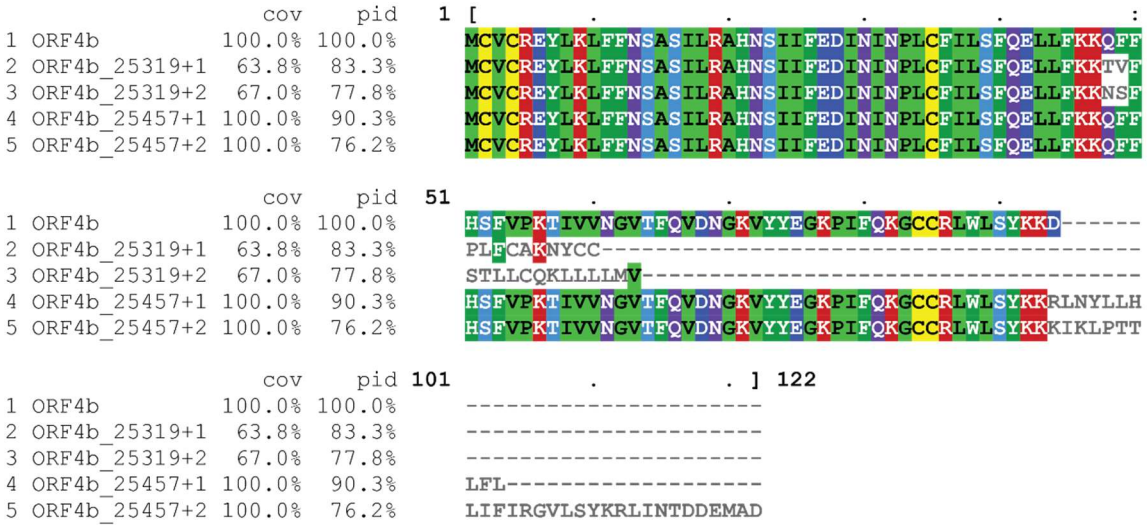


Figure 4.14: Protein alignment of ORF4b with prospective modifications caused by insertion. *The 4b sequence was isolated from the M41-CK consensus sequence and modified to include A frameshift insertions (+1) and (+2) at 25319 and 25457. The resulting sequences were then translated into protein and aligned using MView. The insertions at a nucleotide level result in either shortening of 94AA protein to 60 AAs or elongation to 122 AAs.*

Alternatively, it may be that 4b expression is inhibitory to the virus when passaged repeatedly and insertion at these sites may be a means of producing decreasing levels of functional ORF4b.

Excluding sites 25319 and 25457, indels presented a minor proportion (< 10 %) of the total variants in each passage. A single or double thymine (T) insertion at 18620 was common to all four lineages with a frequency of roughly 0.05, increasing a stretch of 8 Ts to 9 and 10 respectively. 18620 lies within nsp15 (ORF1ab) and in both instances the mutation results in use of an alternate stop codon. Not only is nsp15 truncated from ~ 338 AA to either 44 or 54 AA for G18620GT and G18620GTT respectively, but both mutations result in nsp16 not being translated. The frequency of both mutations remained fairly constant over the course of passaging. It is therefore very unlikely that this mutation is involved in attenuation nor is it likely that there is absence of both nsp15 and nsp16 in an infection. That being said, the loss of nsp15 activity in MHV has been associated with an attenuated phenotype in mice due to allowed stimulation of the host protective immune response (Deng et al., 2017). Conversely, nsp16 is involved in RNA capping (2′O-MTase activity), a process required to prevent viral mRNA recognition by host cell (Decroly et al., 2008). Indeed, nsp16 mutants can be used as vaccine candidates in SARS-CoV (Menachery et al., 2014) or as a drug target for therapeutic intervention (Ke et al., 2012).

One noteworthy deletion identified in the passaging data is GAGCAAGTGATT24064G that, while only occurring in lineage A, reached a remarkable frequency of 0.440603 in the final passage having previously been a consensus-level change in the preceding passages. The deletion of 11 nts in ORF3b would result in a frameshift and use of an alternate stop codon. The mutant 3b protein would be truncated compared to wildtype (30 AA versus 64 AA), sharing 11AA before the shifting frame. Given that 3b is believed to be involved in IFN antagonism, there is a tangible link between the 3b mutation and attenuation.

4.4. Discussion

In an effort to ratify and extend on the conclusions drawn from the 454 analysis and the passaging of M41-CK in parallel replicates (Chapter 3), the possibility of using temperature sensitivity as means of characterising *in vivo* phenotype was investigated. It was the intention here that if temperature sensitivity could be used as a representative tool, it would present a convenient and ethical substitute to *in vivo* experimentation for characterising pathogenicity of a virus. This would subsequently be used to characterise intermediate passages of M41-CK and help establish the passage or passages in which M41-CK is deemed attenuated. Along with targeted sequencing, this would assist the identification of attenuating mutation within the generated dataset.

The ability for each egg-attenuated virus to replicate at permissive (37°C) and non-permissive (41°C) temperatures was surveyed by performing growth curves over a period of 96 hours at both temperatures (Figure 4.1). A statistical difference was observed between the two temperatures for some the egg attenuated viruses but this was not universal across all lineages. Moreover, the drop in titre for these viruses between the two temperatures was comparable to M41-CK, the pathogenic control, as opposed to being similar to Beau-R, the non-pathogenic temperature-sensitive control. Equally H120, the vaccine strain against M41, showed a degree of temperature sensitivity comparable to M41-CK despite being apathogenic. It was therefore concluded that temperature-sensitivity cannot be used to sufficiently nor reliably distinguish between pathogenic and non-pathogenic IBVs and is therefore not an appropriate substitute to *in vivo* experimentation. Instead these findings imply that although temperature sensitivity appears to be correspond to attenuation, the reverse is not true (i.e. attenuated viruses are not necessarily temperature sensitive). Temperature-sensitivity is therefore not the sole factor governing virus pathogenicity.

It cannot be understated how beneficial a targeted sequencing approach, whereby individual passages are sequenced based on phenotype, would be in the identification of attenuating mutations. It is not currently known whether attenuation is result of a "molecular switch" (i.e. a sudden change in phenotype in one passage compared to the previous) or a gradual and incremental change phenotype. Additionally it is not established whether the process of adapting the virus to growth in eggs and the process of attenuation are related.

With the avenue of targeted passage sequencing currently not possible, it was instead decided to assess the order of mutation appearance as a function of the trajectories leading to attenuation. Here, advances in HTS technologies have be utilised to sequence intermediate M41-CK passages in the process of attenuation by egg passaging. Foremost, these updated technologies allowed for both direct M41-CK parent and specific viruses characterised in *in vivo* studies to be sequenced primarily as a result of a lower required RNA input. A lower RNA input is also the reason why it is now possible to sequence intermediate passages without virus amplification *in ovo*.

The initial aim was to survey how IBV evolves during attenuation by egg-passaging and to inform as to the contribution both selection and mutation have in this process. To achieve this, M41-CK infected allanotic fluid used to initiate passaging was deep-sequenced using an Illumina HiSeq platform to offer the maximal detection of possible variants. With a confidence threshold of 0.5 %, 29 variants were detected in M41-CK parent which included 8 of the 9 variants detected by the previous 454 analysis (Figure 4.2 and Table 4.1). The majority of these eight variants were the highest frequency mutations seen in the population. Additionally, consensus sequences generated using reads produced by both technologies were identical with exception of differences in resolution at the 5′ and 3′ termini. This confirms that a minimum read depth of 10 reads for consensus-calling was sufficient for the 454 analysis (Chapter 3) in this instance.

Intermediate viruses in intervals of 10 (i.e. every 10th passage), were sequenced across the series of passaging using Illumina MiSeq and variants called using the same confidence threshold of 0.5 % (Figure 4.3). In agreement with the 454 analysis, a remarkable proportion of variation occurs below consensus-level in the final isolates and would otherwise remain undetected by conventional sequencing. Each intermediate passage also possesses a high proportion of subconsensus variation, with notable differences in the number of both subconsensus and consensus changes between the four lineages at the same point in passaging. Lineage A1 for example shows a near-total collapse of subconsensus variation by at passage 30, having far more consensus-level changes that the other three lineages at this point. This dip cannot currently been explained, nor can it be confirmed in this data whether this trend is common to all four lineages at some point during the passaging. It may however represent lineage A1 being forced through a more severe bottleneck which the sequencing interval has, by chance, captured. Regardless, it is demonstrated here that the accumulation of mutations is not a by-product of repeated passage as the number of mutations can fluctuate both up and down during the course of passaging in eggs.

To identify potentially attenuating mutations shared among multiple attenuated viruses, variant lists for each the final viruses along with M41-CK parent variants were compared (Figure 4.5). The vast majority of variants in the final viruses are unique to that virus. It is also noted that a high number of mutations per passage occurred only in that passage when compared to the entire lineage (Figure 4.6-4.9). This demonstrates the inconsistency of attenuation by egg passage, and provides clear evidence of *de novo* mutation. The transient nature of variations occurring in few consecutive passages may therefore suggest temporary acquisition or tolerance of a variant. This could suggest that a mutation advantageous towards the beginning of egg passaging is less advantageous towards the end of passaging.

All four attenuated viruses possess three mutations (T25457TAA, C26841T and G27183) not present in the original virus, providing examples of shared mutations. In contradiction to this however, of the 29 variants present in the original population 12 remained at the end of passaging in at least one lineage, providing some evidence that mutations present in the original population can remain in the attenuation population. This indicates that selection of variants in the initial population could be involved in the attenuation process. Lending more evidence to selection as a means of IBV evolution is evidence of a pressure driving some mutations present in the original population (Figure 4.4). Not only is the presence of individual mutations unique to each lineage, the profile across the series is also unique to each lineage. T24293C, a mutation which restores M41-CK to the same nucleotide and amino acid that both pathogenic and non-pathogenic IBVs possess, is testament of the individualist pathways the virus uses even when there is a perceived disadvantage of having T at this site (Figure 4.4G).

Examples of co-mutation between at least four (G22220T, G22236T, G22281T and G22916C) nonsynonymous mutations are also presented here. It would present an interesting line of inquiry to study what these mutations do and their prospective role in stabilising prefusion conformation of Spike. It is not clear how this would, if indeed it does, link to attenuation. They could perhaps be involved in restricting virus entry into host cell, however the growth profiles of each attenuated M41-CK does not support this (Figure 4.1). The absence of these mutations in Lineage A1 may indeed suggest that they are not associated with attenuation. Conversely, A1 may possess different mutations that confer the same effect. Regardless, developing genetically-engineered viruses with the M41-CK backbone possessing a mismatch of these sites may inform as to the inner working of S2 in attachment and host-cell fusion.

Reverse genetics could also be employed for both the study of indels and examining the contribution of particular genomic regions in governing pathogenicity. In the former, we have identified two sites (25319 and 25457)

which consistently have insertions and result in extension of a poly(A) region. Curiously, increasing passages appear to correspond with an increased insertion level. Both mutations result in the malformation of ORF4b, a protein which believed to induce the formation of stress granules (Hall, 2017). There is limited evidence to suggest that SGs are beneficial for coronaviruses (Raaben et al., 2007, Sola et al., 2011), so it is unclear as to why the virus would benefit from making such a change. It could be that the absence of a full immune pressure *in ovo* allows the virus to inadvertently make errors at this site, which is unrecoverable when reintroduced into the birds. Reverse genetics could be used to determine what effect the mutant 4b proteins have on virus replication and pathogenicity.

Alternatively, if areas of the genome which come more diverse during passaging are indicative of the regions involved in the attenuation process and in governing pathogenicity, these regions could be modified by a reverse genetics system. While possible, it is unlikely that attenuation is the result of singular mutations but rather collective action of multiple mutations. Indeed, it was confirmed here that attenuation is not the result of a single consensus-level mutation (Figure 4.5). Modifying individual nucleotides site by site would be tedious and may not be sufficient to change the phenotype of a virus. What is perhaps better is to generate IBVs with a series of potentially attenuating mutations or chimeric IBVs with regions undergoing a high level of change with non-pathogenic equivalents. In the latter instance, a chimeric virus with the backbone of M41-CK but with a modified 3′ UTR could be used to explain why the 3′ UTR undergoes such a high level of change relative to its size. This prospect, the replacement of the 3′ UTR in M41-CK, has been explored in Chapter 6.

Attenuation is an undoubtedly complicated process with multiple components likely involved. Nearly all of the genes have been implicated in IBV pathogenicity (Phillips et al., 2012, Armesto et al., 2009, Wickramasinghe et al., 2011, van Beurden et al., 2018) let alone other coronaviruses. Sequencing of attenuated IBV is not novel (Ammayappan et al., 2009, Zhao et al., 2014, Geerligs et al., 2011)

however previous efforts rely on Sanger based sequencing and hence do not explore subconsensus variation. In the case of Ark-DPI, a total of 21 nucleotide changes were detected in the attenuated virus compared to wildtype (Ammayappan et al., 2009). This is comparable to the number of consensus-changes observed here (20, 18, 13, 20 for lineages A-D, respectively), but does not reflect all the mutations that occur at a population level which, as proven here, is numerous. Additionally, sequencing intermediate passages has already been performed but is again limited to Sanger sequencing (Ammayappan et al., 2009).

High-throughput sequencing involving IBV has previously been performed, but typically involves characterising host immune response (Hamzic et al., 2016), characterising actively circulating field strains (Abolnik, 2015) or, more recently, viral gene expression (Dinan et al., 2019). This therefore presents the first studies coupling HTS with attenuation by egg passaging. Furthermore, passaging IBV in replicate lineages has not been reported previously. Using the same starting virus, different evolutionary trajectories establish quickly ultimately affecting the end product. This again highlights the inconsistent nature of generating an attenuated virus using serial egg passaging and the possibility of differently attenuated IBVs, validating one of the key findings of the 454 analysis.

Future work including identifying a suitable means of characterising IBV *in vitro* will aid future interpretations of this dataset. Indeed, the data presented here may act as a compendium assisting current and future efforts in establishing the mechanism of attenuation for IBV. Attenuation is likely to be a multi-faceted process and, as demonstrated (Figure 4.5) is not caused by a single “silver bullet” mutation. One possibility is that attenuation is instead the result of numerous low-impact mutations that collectively drive the population away from a virulent sequence space, in turn becoming attenuated.

Regardless of what the mechanism of attenuation is, evidence of both selection of pre-existing variants and the generation of genetic diversity by mutation is evident here, both of which were proposed as methods of viral evolution (Mayr, 1988). To further unpick variant positions crucial to vaccine virus attenuation, experiments involving the serial passaging of clonal isolates should be considered. These have been performed in Chapter 5. Such experiments will assist in examining the contribution of both selection and mutation in the process of attenuation by egg passaging.

Chapter 5 – Exploring how an IBV clone evolves during attenuation by egg passaging

5.1. Declaration

Some of the contents of this chapter have been included in manuscripts currently in preparation. At the time of writing, it is the intention to submit this manuscript under the working title "Selection or Mutation: Characterising the role of Selection in Viral Attenuation and in the Generation of Live Attenuated Vaccines" to Genome Biology. The authors of this publication are as follows;

Graham L. Freimanis, Michael S. Oade, Sarah Keep, Luca Ferretti, Andrew Beck, Paolo Ribeca, Richard J. Orton, Harm J. Geerligs, Paul Britton, Erica Bickerton, John A. Hammond

The work presented here is primarily the author's sole effort with contributions from colleagues as listed below.

- Passaging of M41-K was performed by Erica Bickerton.
- Characterisation of M41-K viruses *in vivo* (Figure 5.1) was performed in injunction with members the Animal Services Unit and the Coronaviruses group (Pirbright Institute) Ciliary activity assessments were performed blind by Erica Bickerton.
- Generation of sequence data was performed in conjunction with Graham Freimanis.
- Pipeline for data analysis was developed in conjunction with Graham Freimanis and Richard Orton.

5.2. Introduction

The previous chapters (Chapters 3 and 4) have explored how an IBV population evolves over the course of egg passaging and informed on the genomic differences in these attenuated viruses. As the virus used to initiate serial passaging was a

population of RNA genomes and not a clone, it was possible that the virus could evolve through both selection of variants in the original population and by generation of new variants by *de novo* mutation. Indeed evidence of both selection and mutation were present. The combination of both processes however makes distinguishing potentially attenuating mutations more difficult. Experiments using originally clonal isolates would diminish any potential influence of selection on the starting passage in the evolutionary process of attenuation. Serial egg passaging of clonal IBV may generate the equivalent mutations to those seen in passaged M41-CK.

The molecular clone of M41-CK, M41-K, was generated using the Vaccinia virus based IBV reverse genetics system (Casais et al., 2001) by current and former members of the Coronaviruses group at the Pirbright Institute (unpublished data). BeauR-rep-M41-struct-3′UTR, a virus possessing the 5′ UTR and replicase gene of Beau-R and the structural and accessory genes and 3′ UTR of M41 was generated for the purposes of determining the role of replicase or the structural and accessory genes in pathogenicity (Armesto et al., 2009). The vaccinia virus used to generate BeauR-rep-M41-struct-3′UTR was further modified by the reverse genetic system to firstly delete sequence between 5′UTR - nsp3 (inclusive) and subsequently replace it in three fragments with that of M41. The nsp4-nsp16 sequence of Beau-R within this vaccinia virus was subsequently deleted and replaced in four fragments with that of M41. This vaccinia was subsequently modified to account for four nonsynonymous point mutations and insertion of synonymous marker mutations. The end result was therefore a vaccinia virus possessing a complete M41 genome which was rescued to generate M41-K virus.

M41-K, the clonal isolate of M41-CK has here been serially passaged in embryonated eggs analogous to Chapter 3 and the attenuation of the final isolates (passage 100) confirmed *in vivo*. To collaborate with data obtained from sequencing intermediate passages of M41-CK (Chapter 4), the intermediate passages of M41-K (passage 10 – passage 90) have also been sequenced. Utilising

data from both experiments will offer a bipartisan perspective on the process of attenuation and foster identification of genetic pathways leading to attenuation. This chapter will therefore confirm whether attenuation by egg passaging is achievable for a virulent clone of IBV (M41-K) and survey how the virus evolves in parallel replicates. The results here will be compared to that of the virulent population (M41-CK) to determine any similarities.

5.3. Results

5.3.1. Serial *in ovo* passaging of M41-K

The pathogenic clone of M41-CK, M41-K, was inoculated into the allantoic cavities of four 10-day-old embryonated SPF RIR chickens to establish lineages A, B, C and D. Serial passaging was performed (up to a total of 100 passages) as previously described (Chapter 3), with two notable differences in procedure. Firstly, dilutions were performed at a fixed ratio of 1 in 1,000 as opposed to varying dilution rate between 1 in 100 and 1 in 10,000 based on visual inspection of the egg (as per condition of the embryo). Secondly, dilutions were made using PBSa as opposed to 1 x BES-containing medium.

Allantoic fluid was screened routinely for the presence of IBV by RT-PCR amplification using BG56 and 93/100 (Table 2.15). All passaging was performed as anticipated with no requirement to split lineages due to virus disappearance. The passaging of M41-K was performed by Erica Bickerton prior to the start of this PhD.

5.3.2. Serially egg-passaged M41-K viruses are attenuated *in vivo*

With temperature-sensitivity not presenting a suitable means of characterising IBV (Figure 4.1), there is currently no alternative other than performing *in vivo* experiments for the purposes of studying IBV pathogenicity. To that end, groups of twelve 8-day old SPF Rhode Island Red chickens were inoculated with 1 x 10^5 PFU of IBV or mock infected with 0.1 mL serum-free medium via the intranasal and ocular routes. Here, the pathogenicity of the final passage (passage 100) from each lineage (A, B, C and D) was tested. Snicking (Figure 5.1A) and rales (Figure 5.1B) were recorded daily between days 3 and 7 post inoculation with ciliary activity (Figure 5.1C) assessed post mortem on days 4 and 6 post infection.

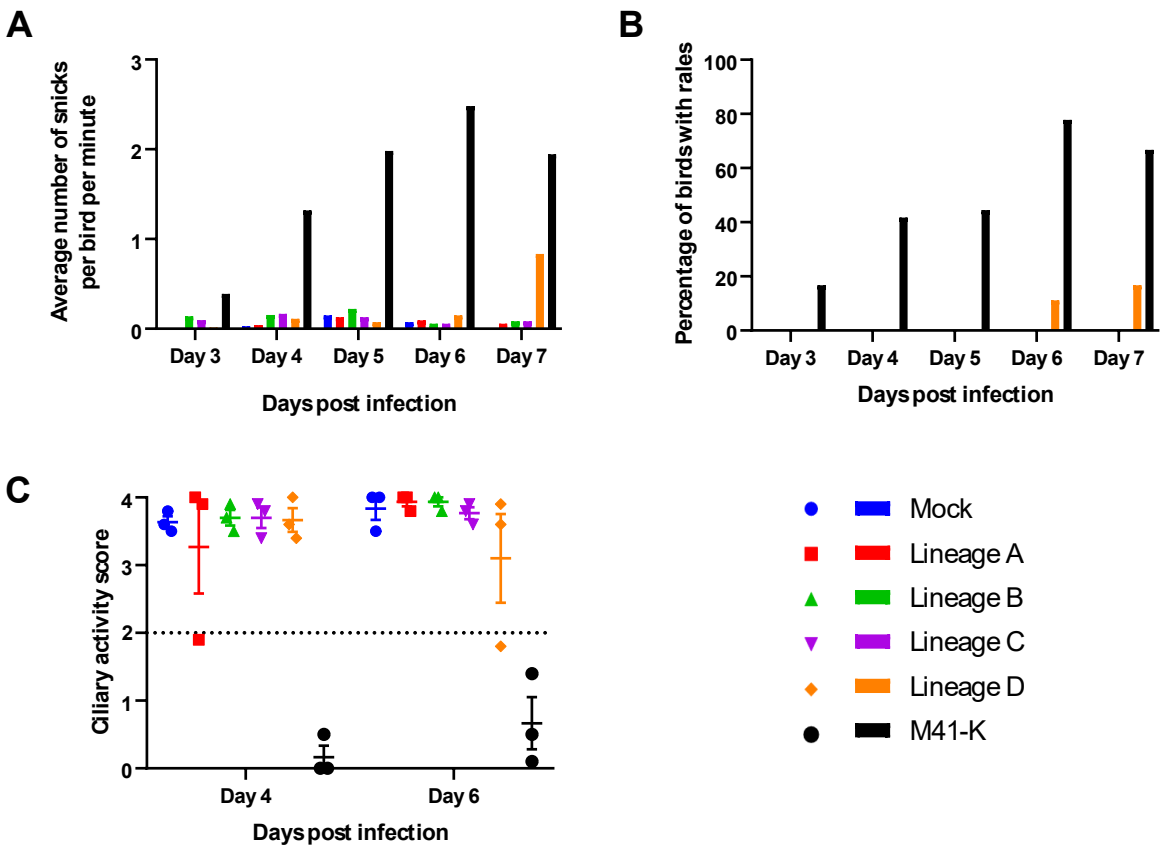


Figure 5.1: Serially egg-passaged M41-K viruses are attenuated in vivo. *For each group, twelve 8-day old SPF chicks were inoculated with 1 x 10^5 PFU of each virus via the ocular and nasal routes. The final passage (passage 100) from each lineage (A-D) was assessed and compared to the virulent parent (M41-K). Clinical signs including snicks (A) and rales (B) were observed daily between days 3 and 7 post inoculation. Ciliary activity (C) was assessed on days 4 and 6 post infection in post mortem by observation of 10 x 1mm tracheal cross-sections (rings) from three randomly selected birds per group. The mean ciliary activity score of these three birds is plotted, with the SEM plotted as error bars. Ciliary activity was scored blindly by Erica Bickerton.*

All four serially egg-passaged M41-K viruses show a marked reduction in clinical signs versus M41-K control (parent virus). Some chickens infected with Lineage D showed minor clinical signs consistent with IBV infection on days 6 and 7 post-infection but rales was only observed in a single bird on each day and snicks only observed on day 7. Mock-infected and M41-K groups performed as expected for both clinical signs and ciliary activity.

With the exception of two birds (Lineage A Day 4, and Lineage D Day 6) all serially egg-passaged M41-K viruses maintained ciliary activity above 50% (ciliary activity score > 2). An average ciliary activity above 50% suggests that IBV is attenuated. Average ciliary activity for all serially egg-passaged M41-K viruses is firmly above the 50% threshold. In fact all but two birds (Lineage A, Day 4 and Lineage D Day 6) scored a ciliary activity above 75 % (ciliary activity > 3). There is no statistical difference observed between mock and any of the serially passaged M41-K viruses on either day (Two-way ANOVA, Tukey multiple comparison). All comparisons versus the M41-K parent were highly significant with the highest P (least significant) value calculated as 0.0002 for Day 6, Lineage D versus M41-K. These findings therefore confirm that the parent M41-K is indeed virulent and that the serially passaged M41-K viruses are attenuated by process of serial egg-passaging.

5.3.3. Assessing inherent M41-K parent diversity in comparison to M41-CK

Allantoic fluid used to initiate the M41-K serial passaging was initially deep-sequenced using Illumina HiSeq, however this was unsuccessful (due to low RNA input) and, due to limited available material (i.e. the M41-K virus used to start passaging), was not explored further. M41-K was instead successfully sequenced in duplicate using Illumina MiSeq platform. Sequencing of both the M41-K initial, intermediate (every 10^{th} passage between passage 10-90, inclusive) and final (passage 100) was mainly prepared by Graham Freimanis (Pirbright High Throughput Sequencing Unit) assisted by Michael Oade. A pipeline for the analysis of all M41-K obtained data, the results of which are described below, was

developed in collaboration with Graham Freimanis. Process automation and script writing was performed my Michael Oade.

A *de novo* consensus sequence was generated from non-host derived reads using SPAdes in keeping with the method as used for M41-CK. Resulting contigs were aligned to the M41-CK reference sequence to generate an M41-K reference sequence. This was initially trimmed at its 5′ end for the purposes of ensuring confidence and to align the M41-CK and M41-K coordinates hence making sequence annotations transferable. Due to a single nucleotide insertion a 27483 nt sequence was generated for M41-K, 1 nucleotide longer than M41-CK (27482 nt).

M41-K had been designed to include 9 synonymous marker mutations for the purposes of distinguishing between M41-CK and M41-K by Sanger sequencing (Table 5.1). An alignment between newly generated M41-K and M41-CK sequences confirmed all 9 of these mutations. However, an additional 10 mutations were also identified, including a single nucleotide insertion at 25320. This corresponds to the 25319 mutations observed in M41-CK as described in Chapter 4. Here, M41-K possesses a stretch of seven adenosines (A) at 25319 at consensus-level whereas M41-CK possesses six. This would suggest ORF4b is of different sequence at both a nucleotide and amino acid level between the two viruses. 4b in M41-K is truncated compared to 4b in M41-CK (60 AA versus 94 AA respectively), sharing the initial 50 AAs. As detailed in 4.3.7, the role of 4b in IBV is not well established but is believed to induce stress granule formation (Hall, 2017). How this might affect IBV pathogenicity is unclear and indeed, if M41-CK and M41-K are both attenuated but possess different 4b protein sequences, it may not be involved in governing pathogenicity at all.

The remaining 9 unexpected mutations encode both synonymous and nonsynonymous changes. Regardless it has been confirmed that these mutations had not have an effect on virus phenotype (Figure 5.1). Curiously M41-K possesses the 24293 mutation in envelope protein as described in Chapters 3 & 4, making M41-K resemble other IBVs at this site as opposed to being similar to M41-CK.

Table 5.1: Consensus differences between M41-CK and M41-K[a]

POSITION	LOCATION	NUCLEOTIDE		AMINO ACID		MARKER?
		M41-CK	M41-K	M41-CK	M41-K	
1044	nsp2	T	C	Synonymous (L)		Yes
1430	nsp2	T	C	Synonymous (G)		Yes
2682	nsp3	G	T	V	L	
5504	nsp3	T	C	Synonymous (F)		
7574	nsp4	C	T	Synonymous (A)		Yes
11678	nsp9	C	T	Synonymous (T)		Yes
11705	nsp9	C	T	Synonymous (A)		Yes
15400	nsp13	G	T	Synonymous (V)		Yes
17941	nsp14	A	T	Synonymous (I)		Yes
19060	nsp15	C	A	Synonymous (R)		Yes
20290	nsp16	A	G	Synonymous (T)		Yes
20678	S1 subunit	C	T	Synonymous (N)		
21615	S1 subunit	C	A	P	T	
22220	S2 subunit	G	T	K	N	
22268	S2 subunit	C	A	D	E	
22612	S2 subunit	T	C	I	T	
24293	Envelope	T	C	F	L	
25320 (25319)	ORF4b	-	A	Frameshift		
26401 (26400)	Nucleocapsid	A	G	N	D	

***a** Nineteen differences were identified between M41-CK and M41-K consensus sequence including all nine known marker mutations. The position and location for each change is given along with coding effect.*

Variant calling for M41-K initial population was processed similarly to the intermediate M41-CK passages (Chapter 4). Here each sample was sequenced in duplicate to avoid potential sequencing error or biases. A custom script was generated based on BCFtools v1.3.1 to compare the two variant files (VCFs) and export variants only occurring in both files. Only variants common to both replicates are used for downstream analysis with the mean frequency of individual variants in both replicates reported.

A total of 14 variants above the minimum confidence threshold (>0.5%) (Figure 5.2) were detected by this method. The highest frequency of these occur at positions T2682G (0.462), A21615C (0.468) and TA25319T (0.391), sites which differ between the M41-CK and M41-K consensus sequences. These mutations therefore approach the consensus threshold (0.5), with the population almost equally split between the two nucleotides at these positions. While a minor frequency discrepancy may explain the difference in being a consensus versus a subconsensus mutation, it is arguable the M41-K is not clonal at these sites. Mutations at these three sites (2682, 21615 and 25319) were the only sites possessing variation where M41-CK and M41-K consensus sequences differ.

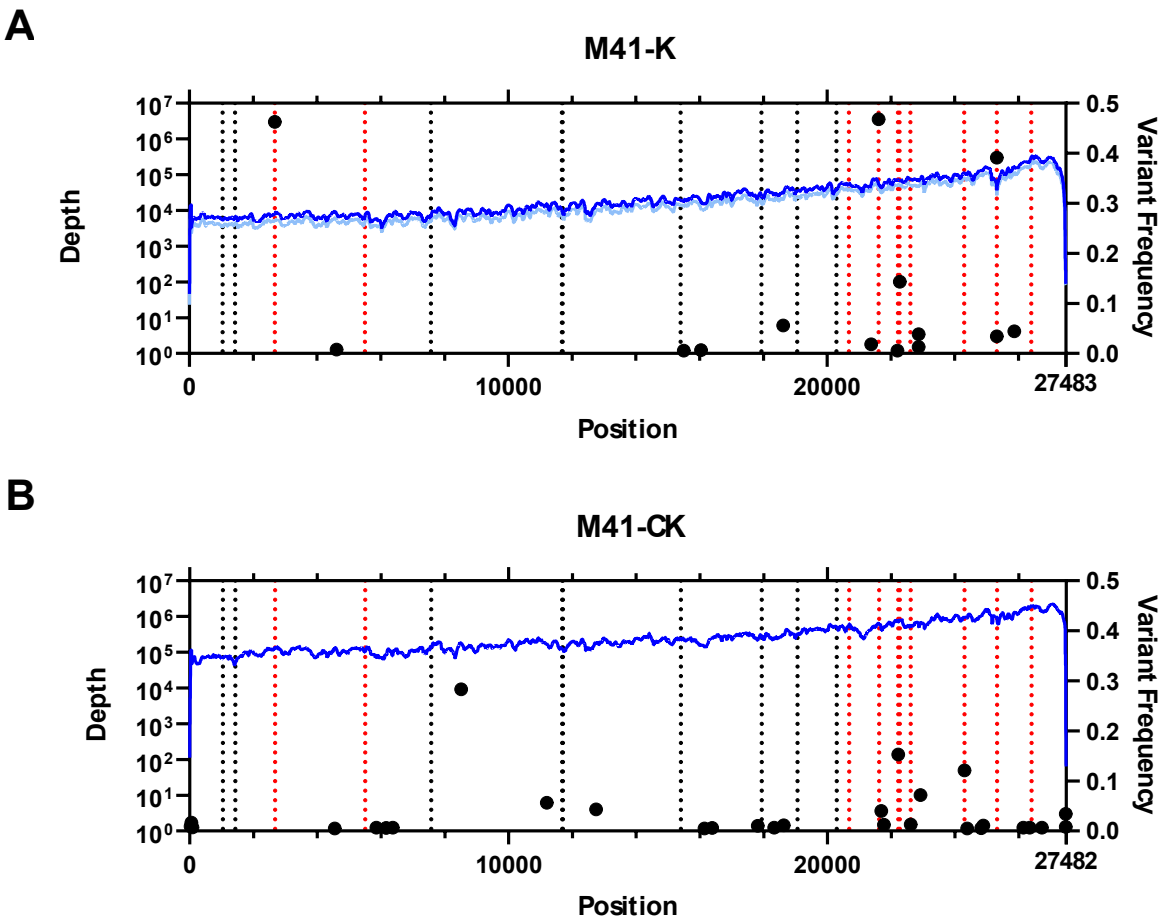
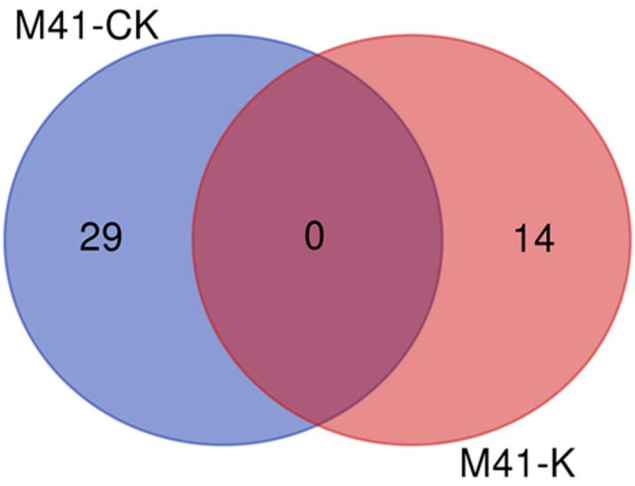


Figure 5.2: Distribution of variants in M41-K and M41-CK parent viruses*. Reads for both M41-K (A) and M41-CK (B) were aligned to their respective reference sequences and variants identified. The blue line indicates depth of coverage across the genome; two lines are presented for M41-K as the sample was sequenced in duplicate. The sites of consensus-level changes between the two viruses are marked; black dotted-lines indicate sites of marker mutations (n = 9), red dotted-lines indicate sites of unanticipated mutations (n = 10). The position and frequency of called variants is indicated by a black dot plotted against the right-hand y-axis.*

To determine whether virulent M41-CK and virulent M41-K possess any variants shared between them, variant lists of M41-CK (Chapter 4) and M41-K were compared. Other than the consensus-level changes as listed described above (Table 5.1 and Figure 5.2), no variants were otherwise common between both viruses (Figure 5.3). It was observed however that consensus level change in M41-K corresponds to subconsensus mutation in M41-CK (T22612C) although admittedly at low frequency (0.013).

The evidence as described above may therefore suggest that M41-K is arguably not clonal. Though M41-K does possess fewer variants compared to M41-CK (14 versus 29), the mutations it possess occur at a higher mean frequency (0.121 versus 0.034), a statistic largely skewed due to the existence of the three high frequency mutations (T2682G, A21615C, TA25319T). Indeed, there is perhaps the argument that M41-CK and M41-K show a comparable level of diversity despite one supposedly being the clone of the other. Given the nature of how M41-K was generated (as detailed in this Chapter's introduction) and the multiple homologous recombination events (see Figure 2.1) that would have occurred to generate the full-length cDNA genome within vaccinia virus, it is plausible that single nucleotide errors would have gone undetected. If these SNPs occurred outside of the site of recombination and only the area around the recombination site was spot sequenced, this would further potentiate any issues. Regardless of how the mutations were introduced, this would only explain how consensus-level changes were introduced as, at the vaccinia virus stage, the IBV sequence is presumably clonal. Subconsensus variation in rescued M41-K IBV may therefore (Figure 5.2) suggest that this assumption false (i.e. the IBV held in vaccinia virus is non-clonal) is or, more likely, that the process of rescuing the IBV and the subsequent passages for generating a working stock is sufficient to generate the diversity as seen here. On the basis that the latter of these two are true, it is testament to the level of change that can occur in only a few passages.



INTERSECTION	TOTAL	ELEMENTS
M41-CK, M41-K	0	
M41-CK	29	C26T, T55G, C88T, G4548A, C5856T, G6158A, A6373G, T8516C, C11199CT, G12740A, G16144A, C16381T, C17811A, G18325T, A18629T, A21687G, C21772T, G22220T, T22612C, G22916C, T24293C, G24377T, A24820C, C24883T, C26136T, G26342A, C26725T, A27467G, A27471T
M41-K	14	T2682G, G4620T, C15506T, C16045T, G18620GT, T21383G, A21615C, A22204G, C22279T, G22869C, A22870T, T25319TA, TA25319T, T25866A

Figure 5.3: Comparison of M41-CK and M41-K parent variants. *Variants lists for the M41-CK and M41-K starting inoculums were compared and plotted in Venn diagram. No overlap between the two lists. Mutations occurring at a site of consensus level changes between the two sequences are underlined.*

5.3.4. Calling variants in the intermediate and final M41-K passages

After generating an appropriate consensus sequence for M41-K and subsequently verifying it, reads obtained from intermediate (every 10th passage between passage 10-90, inclusive) and final (passage 100) passages were mapped to this sequence and variants identified (Figure 5.4).

Similarly to the passaging of M41-CK, though there is a net gain of variants by the final passage, the gain is not continuous nor proportional to increasing passage number. Likewise the number of consensus level mutations increased for all lineages of M41-K as passaging continued with exception of some variants fluctuating near the consensus-level threshold. Indeed in comparing M41-CK (Figure 4.3) and M41-K profiles (Figure 5.4), there is not a clear distinction between the two viruses.

It is somewhat curious that in the initial 20 passages of lineage B, the number of variants decreased and passage 10 and 20 are arguably more clonal that the parent. What is also of note is the level of change, specifically in the number of consensus level changes, occurring in Lineage A passage 70 compared to the prior and succeeding passages. A total of 11 consensus level mutations were identified in passage 60 and 12 were identified in passage 80. Eight consensus-level mutations were common to both passages. 27 consensus changes were identified in passage 70, one of which was common to all three passages and one of which was shared with passage 60. The majority of these mutations fall solidly within the consensus-level (> 50%) bracket, so it cannot be argued that this passage has captured a number of "flip-flopping" variants near the consensus-threshold. The data we have now cannot explain as to why this is observed. To avoid potential sequencing error, each sample was sequenced in duplicate and only common variants evaluated, as has been previously described. The number of mutations seen here therefore reflects those seen in both replicates. Indeed, it is unrealistic to believe sequencing error is accountable for such a significant change. Assuming that precautions taken (i.e. sequencing in duplicate) mitigates potential

sequencing error, it may be the result observed is a true indication of Lineage A Passage 70. This itself however may be caused for example by erroneous procedure, for example an incorrect rate of dilution between egg passages. Instead, perhaps the most logical reason for there being a difference in passage 70 is that it is true result and, given that only every 10 passages was sequenced, it was by chance not captured in the other lineages.

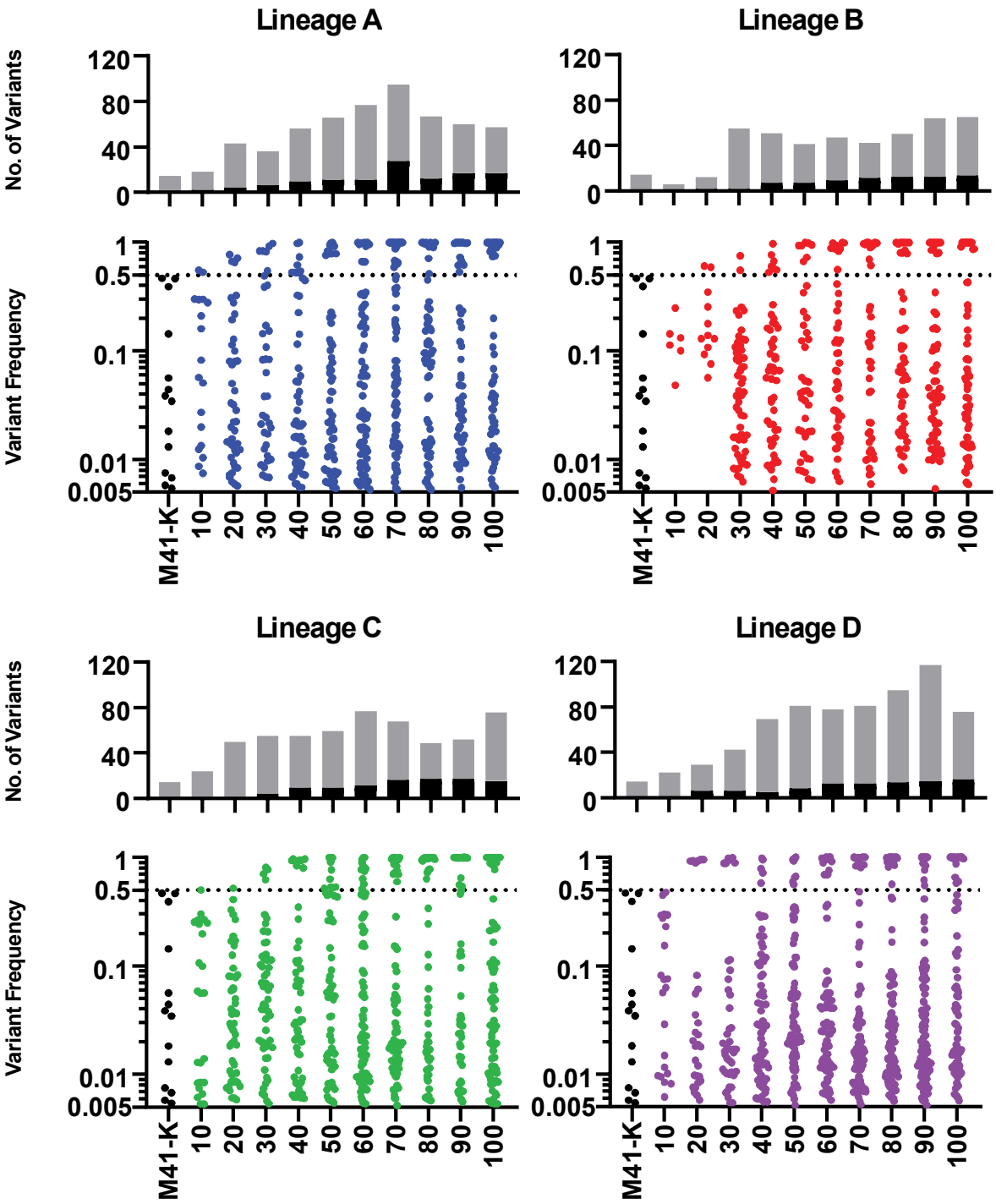


Figure 5.4: Variant plots for individual passages of M41-K. *Continued on next page.*

Figure 5.4: Variant plots for individual passages of M41-K. *Reads for each individual passage were aligned to the M41-K consensus sequence. Variants were identified are plotted per passage (see scatter plots below each bar graph). Each individual dot represents a separate variant with the dotted line at 0.5 (or 50%) representing the consensus level threshold. The number of variants occurring above this threshold (i.e. a consensus-level change) is plotted as black in the bar chart above each plot whereas the grey bar indicates variants below this threshold (i.e. subconsensus).*

5.3.5. Studying dynamics of M41-K parent variants

To study the dynamics of variants present in the original population, the nine highest frequency variants were selected and traced over the course of passaging (Figure 5.5). The patterns observed in the intermediate passaging of M41-CK are transferable here.

Firstly variants existing in the virulent population may be lost during the process of serial egg passaging. This was also observed in Figure 5.4. Here, the highest frequency example is A21615C (0.468, Figure 5.5D) where after a maximum of 30 passages of Lineage B, the variant is lost. The frequency of A21615C also drops in Lineage A and D, disappearing by passages 90 and 70, respectively. It is worth mentioning that the profiles of A21615C (Figure 5.5D) and T21383G (Figure 5.5C) are remarkably similar in lineages C and D suggesting possible linkage between the two sites. The pattern is not consistent with all lineages, however a separate observation (not shown) of possible co-mutation between mutations G22236A and T22958A is observed for lineage A. All four mutations are nonsynonymous and occur within S2 (Spike). This series of mutations may operate collectively or as part of the same mechanism of Spike as discussed in Chapter 4 (see 4.3.6).

Secondly, initially low frequency mutations can remain at similar frequencies throughout the course of passaging. G18620GT (Figure 5.5B) for example started at an initial frequency of 0.056 and was repeatedly identified in subsequent passages with some exceptions for when presumably the variant occurred at a frequency below the confidence threshold. Interestingly, as discussed previously this mutation also occurs in the M41-CK intermediate passages and demonstrates a similar low frequency pattern. The effect of mutations here result in truncation of the replicase1ab polyprotein and non-expression of nsp15 and 16. Both of these proteins can be tangibly linked to attenuation as loss of nsp15 activity in MHV has been associated with an attenuated phenotype (Deng et al., 2017) and nsp16 mutants are used as vaccine candidates in SARS-CoV (Menachery et al., 2014).

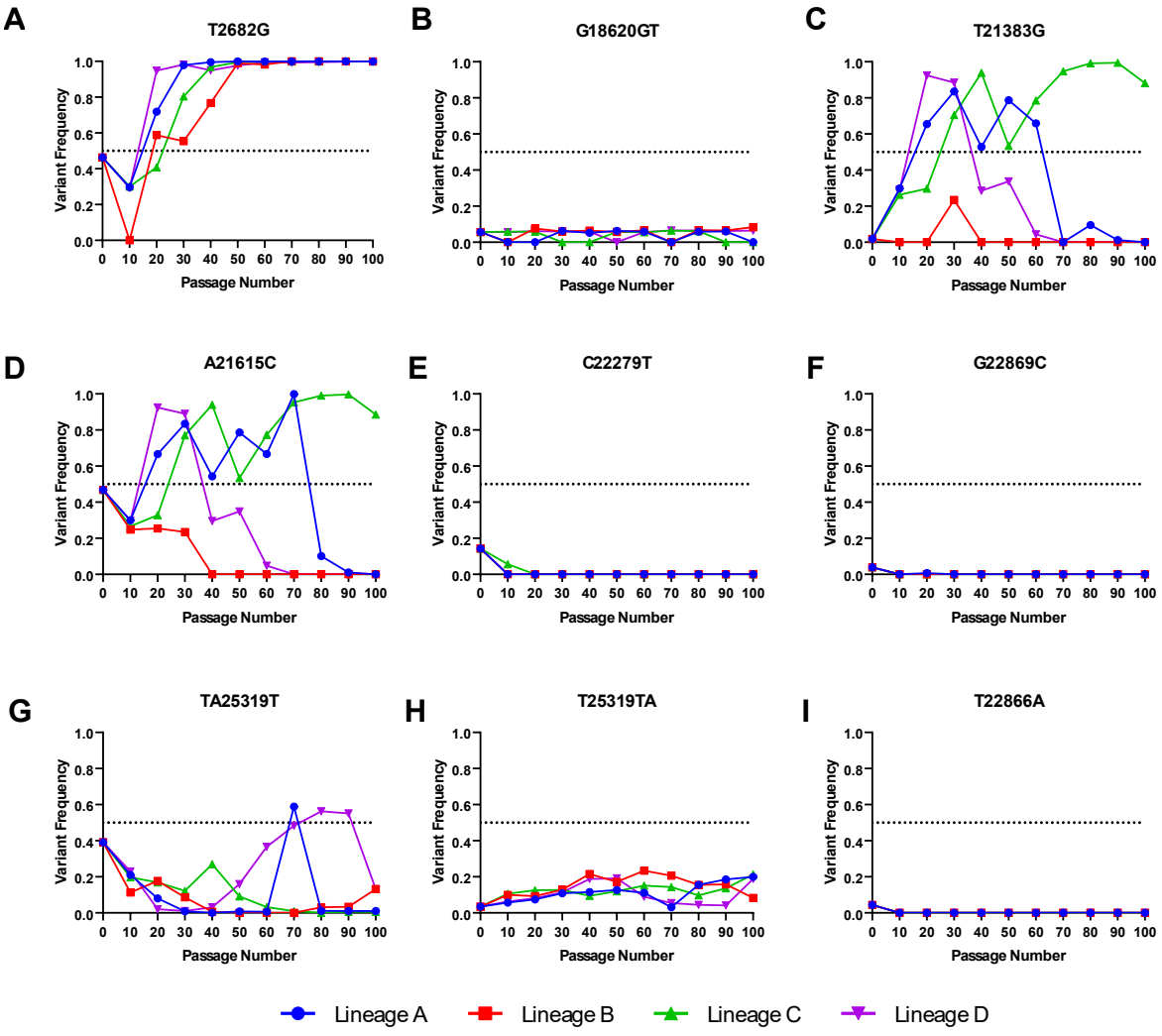


Figure 5.5: Variant frequency plots monitoring intermediate frequencies of variants present in the original population. *The nine highest frequency variants of the initial population were selected and their frequency traced over the course of passaging. Each graph plots the same mutation. These plots are representative of the dataset and demonstrate the variability that occurs during the course of passaging.*

Finally, even if a mutation is present in the final passage, the profile of the variant in the intermediate passages is unique to that lineage. T21383G and A21615C are both clear examples of mutations that occur at consensus-level in one lineage (Lineage C), but not the other three (Figure 5.5C and 5.5D). Equally T2682G (Figure 5.5A), a variant that approaches fixation for all four lineages, presents different profiles that would otherwise be undetected without sequencing intermediate passages.

What is potentially different between M41-CK and M41-K is the maximum possible change in frequency of a particular variant in ten passages. For T2682G, a change of frequency of 0.588 is observed between passage 10 and 20 for lineage B (Figure 5.5A). Further examples here include the T21383G and A21615C (Figure 5.5C and 5.5D) mutations in lineage A where a maximum frequency change of 0.659 and 0.897, respectively, is observed. Interestingly, in some instances this remarkable change in frequency is not sustained for many passages. Given the data available, the minimum possible window for a variant to spike in frequency is 20 passages (i.e. passage 60 – passage 80, spiking at passage 70). The TA25319T mutation (Figure 5.5G) in lineage A spikes at frequency of 0.589 at passage 70 but is only detected at a frequency of approximately 0.01 at passages 60 and 80. This may go some way as to explaining as to why passage 70 in Lineage A contains far more consensus and subconsensus variants than passages 60 and 80. It would appear somewhere over these passages, lineage A undergoes a landmark event, fundamentally altering the course of its evolution. This could be indicative of a passaging bottleneck either accidentally (i.e. a change in dilution rate, an issue with the egg) or inadvertently (i.e. evolution of the virus).

What is perhaps of note, though arguably incidental, is that for the majority of variants, their passage 10 frequency are almost identical. Combined with the fact that the frequency of two or more mutations can change collectively, this may be indicative of the viruses possessing a shared evolutionary pathway until this

passage. As this is not reflected from passage 20 onwards, this shared pathway is not shared for very long.

These graphs also signal the recurrence of the 25319 mutation corresponding to elongations of a poly(A) region in ORF4b. As discussed previously 25319 is a site in which the M41-CK and M41-K consensus sequences are different with a single nucleotide insertion in M41-K versus M41-CK. This results in the two viruses possessing different ORF4b protein sequences. Curiously, this site possesses mutations that result in either the insertion (Figure 5.5H) or deletion (Figure 5.5G) of a single nucleotide. A proportion of the population is therefore M41-CK like due to the deletion of an A in the M41-K reference. The length of this poly(A) region appears to increase in sustained passaging as also observed with M41-CK passaging.

5.3.6. Identifying variants shared between multiple M41-K lineages

To identify variants common to multiple attenuated IBVs, comparisons were made between variants identified in M41-K parent versus those in each final attenuated isolate (Figure 5.6). In concurrence with M41-CK passaging, each attenuated M41-K mostly possesses variants unique to that lineage, with very little overlap observed between multiple lineages. 43/57, 50/65, 58/76 and 57/76 variants were unique to each attenuated M41-K isolate for each lineage respectively, representing between 75 – 77 % of all variants called. This was roughly the same percentage for attenuated M41-CK (Figure 4.5).

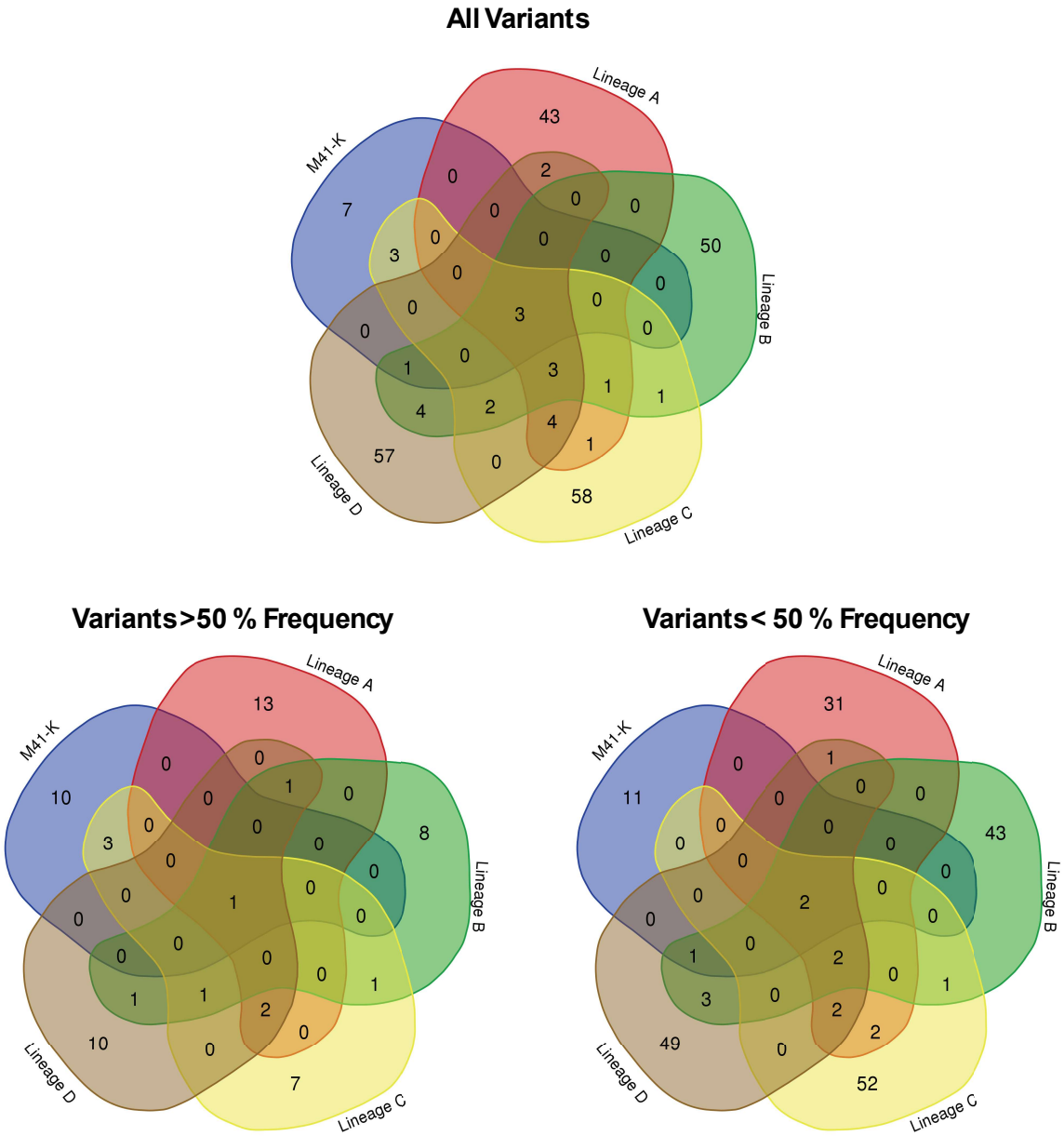


Figure 5.6: Venn diagrams from the comparison of called variants between final M41-K lineages. *Variants identified in the starting passage and final passages were compared with Venn diagrams produced for all variants, those occurring at > 50 % frequency (consensus) and those occurring < 50% frequency (subconsensus) in the attenuated (final-passage) virus isolates. M41-K parent variants present in the original population are plotted in all three Venn diagrams. Variants occurring outside the blue area are indicative of de novo mutation (variant was not detected in the starting population).*

Regardless of their frequency, all four attenuated viruses share a total of 6 mutations with each other (Table 5.2), three of which were also present in the starting population. Four of these mutations (two not present in the starting population, two present in the starting population) correspond to 25319 poly(A) region in ORF4b as described previously. The fifth mutation T2682G, present in the original population, has also been described previously (Figure 5.5). The final mutation, T22958A, believed to occur in conjunction with G22236A for Lineage A, results in a nonsynonymous mutation in Spike (N863K).

While these six mutations are shared between all four attenuated viruses, it should be noted that they occur at drastically different frequencies between the different lineages. This again highlight the unique profile of each variant to each lineage.

Table 5.2: Variants shared between virulent and attenuated M41-K[a]

Mutation	Gene	Variant Frequency				
		M41-K	Lineage A	Lineage B	Lineage C	Lineage D
T2682G	Nsp3 (NS)	0.463	0.999	0.999	0.999	0.016
T22958A	S2 (NS)	ND	0.995	0.997	0.006	0.997
TA25319T	ORF4b	0.391	0.011	0.132	0.006	0.131
T25319A	ORF4b	ND	0.022	0.008	0.029	0.016
T25319TA	ORF4b	0.034	0.200	0.083	0.213	0.188
T25319TAA	ORF4b	ND	0.033	0.006	0.026	0.019

[a]ND = Not detected

5.3.7. Establishing variant homology between successive M41-K passages

To determine both the initial passage at which a variant was detected and to assess variant homology between successive passages, further comparisons within individual lineages were made (Figures 5.5 to 5.9). As detailed in 4.3.5, UpSet plots are used to visualise the relationship between multiple datasets. Here, they are used to visualise variants either unique to that passage or shared across multiple passages. The horizontal bar graph plots the number of variants per passage, equivalent to Figure 5.4. The middle panel consisting of dots and sticks signifies a comparison being made between one or more datasets, with the number of variants then plotted in the bar chart above. No single piece of information is plotted twice.

Similar plots for intermediate M41-CK passaging (Figures 4.6-4.9) revealed that a high proportion of mutations identified per passage remain unique to that passage. This characteristic is also featured in the M41-K passages, as indicated by the highest bars in the upper bar charts (Figure 5-7-5.10) corresponding to singular non-connected dots in the middle panel (Figure 5-7.5.10). Furthermore, in keeping with the observation that lineage A passage 70 is somewhat an oddity compared to other passages, passage 70 is here identified as possessing far more variants that any other passage regardless of what lineage. Lineage A passage 70 therefore presents a level of anomaly both within the M41-K and M41-CK datasets. It may not entirely be discounted as it may represent a pattern that happens in all eight lineages (4 of M41-CK and 4 of M41-K) that has, by change been missed by the process of sequencing every 10th passage. Conversely, for the reasons as previously discussed, it may represent a level of error which has occurred for data related to the passage.

Additionally in keeping with M41-CK passaging, intermediate passages of M41-K demonstrate a steady accumulation of variants present in the final attenuated population in successive passages. In the case of Lineage A for example (Figure 5.7), 7 variants were identified as occurring in lineage A, passage 100 only with a

further 7 present in both passages 90 and 100. In passages 80, 90 and 100, 3 more variants are shared etc. The exception to this rule is passage lineage A, passage 70 as previously discussed.

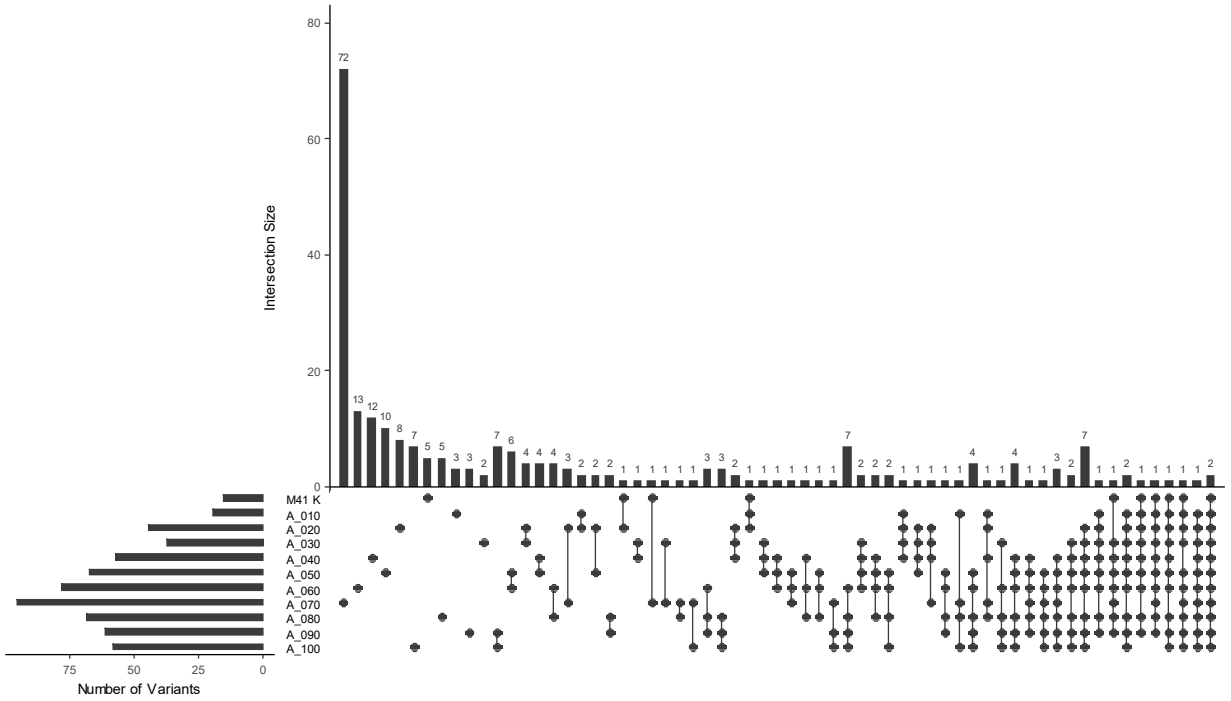


Figure 5.7: UpSetR plot for the comparison of all variants called in multiple passages of M41-K Lineage A. *The total number of called variants per passage is indicated in the left bar graph. Variants identified in each passage (i.e. a set) was compared to every other passage. The dots and lines plot indicate the comparison being made (i.e. intersection between sets) with the bar chart (top) indicating the number of variants within that intersection. Singular dots indicate variants unique to that passage.*

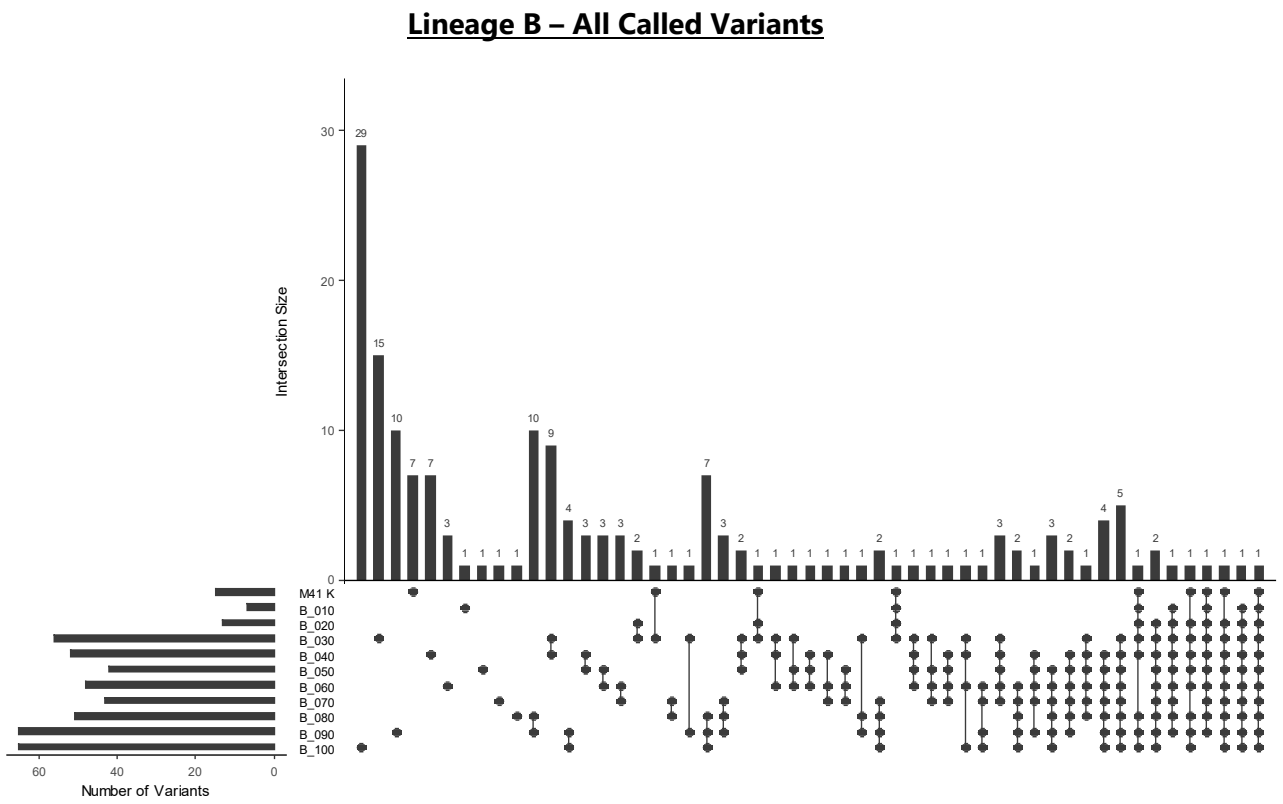


Figure 5.8: UpSetR plot for the comparison of all variants called in multiple passages of M41-K Lineage B. *The total number of called variants per passage is indicated in the left bar graph. Variants identified in each passage (i.e. a set) was compared to every other passage. The dots and lines plot indicate the comparison being made (i.e. intersection between sets) with the bar chart (top) indicating the number of variants within that intersection. Singular dots indicate variants unique to that passage.*

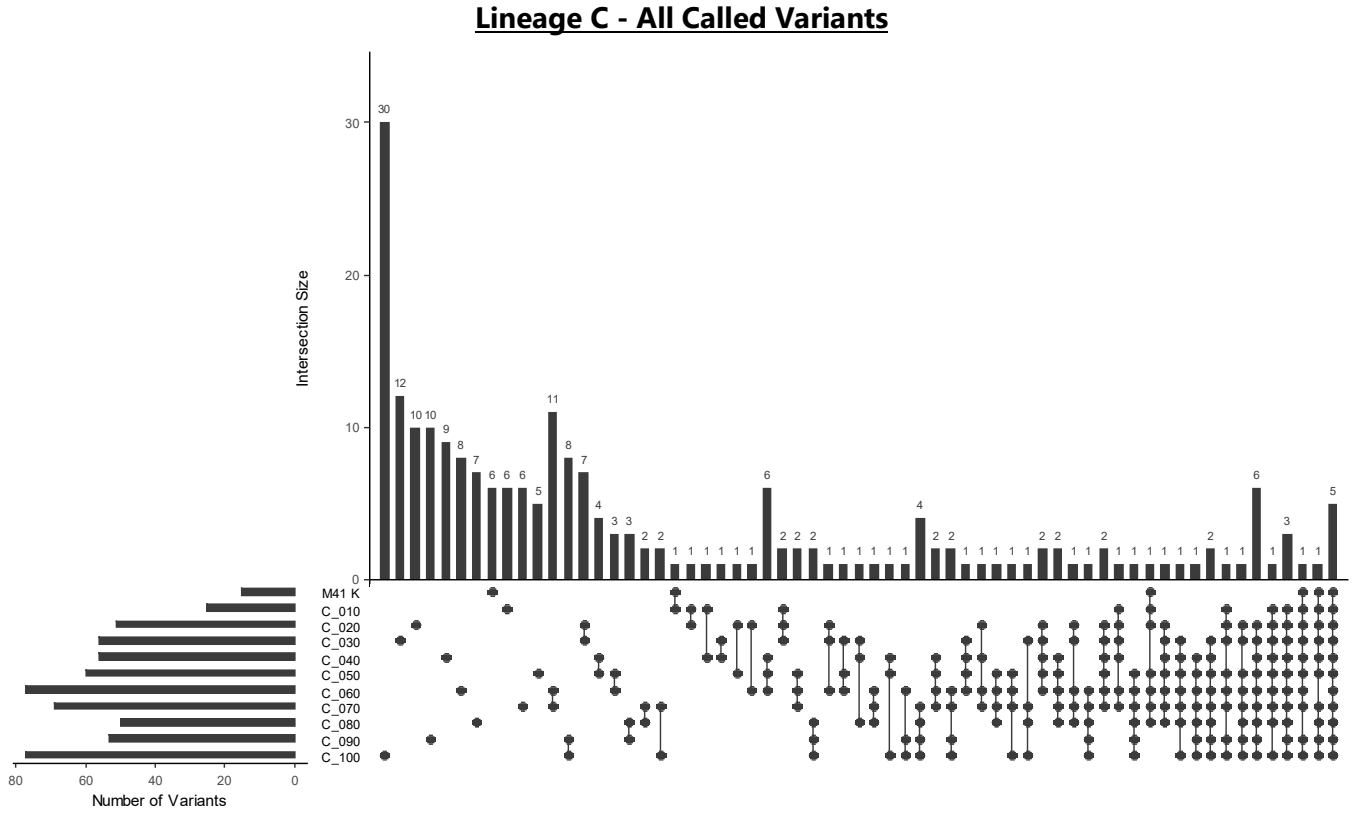


Figure 5.9: UpSetR plot for the comparison of all variants called in multiple passages of M41-K Lineage C. *The total number of called variants per passage is indicated in the left bar graph. Variants identified in each passage (i.e. a set) was compared to every other passage. The dots and lines plot indicate the comparison being made (i.e. intersection between sets) with the bar chart (top) indicating the number of variants within that intersection. Singular dots indicate variants unique to that passage.*

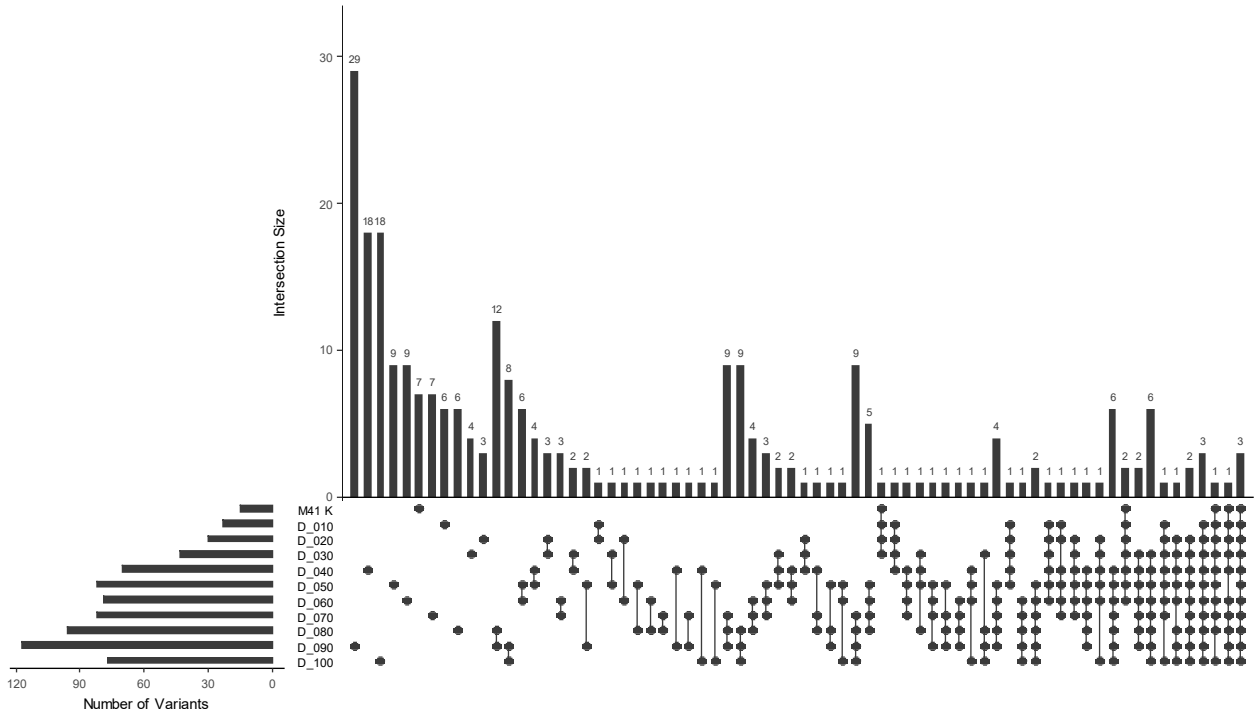


Figure 5.10: UpSetR plot for the comparison of all variants called in multiple passages of M41-K Lineage D. *The total number of called variants per passage is indicated in the left bar graph. Variants identified in each passage (i.e. a set) was compared to every other passage. The dots and lines plot indicate the comparison being made (i.e. intersection between sets) with the bar chart (top) indicating the number of variants within that intersection. Singular dots indicate variants unique to that passage.*

5.3.8. Cross-verifying mutations present in attenuated M41-K and M41-CK

To identify potential variants shared between attenuated M41-CK and M41-K viruses, variants lists from all eight lineages were compared (Figure 5.11). The individuality of each lineage is displayed here again with few instances of variants shared between viruses deriving from different parents. It is worth mentioning however that this does not take into consideration any mutations that result in a change to the consensus (reference) sequence of the other viruses. For example, M41-CK has a mutation T24293C mutation present in all four lineages, however as M41-K already possesses a C at this site. It would not be flagged as a shared variant despite having a common nucleotide.

The individuality of each virus is prominent here with the vast majority of variants detected in each virus only being detected in that virus. This is indicated by the highest bars in the upper bar chart corresponding to the individual, non-connected dots in the middle panel. No single variant was common to all eight attenuated viruses, again demonstrating there is no single variant responsible for attenuation. There are examples of variants existing in both attenuated M41-CK and attenuated M41-K however these are not in abundance compared to the total number of variants per virus.

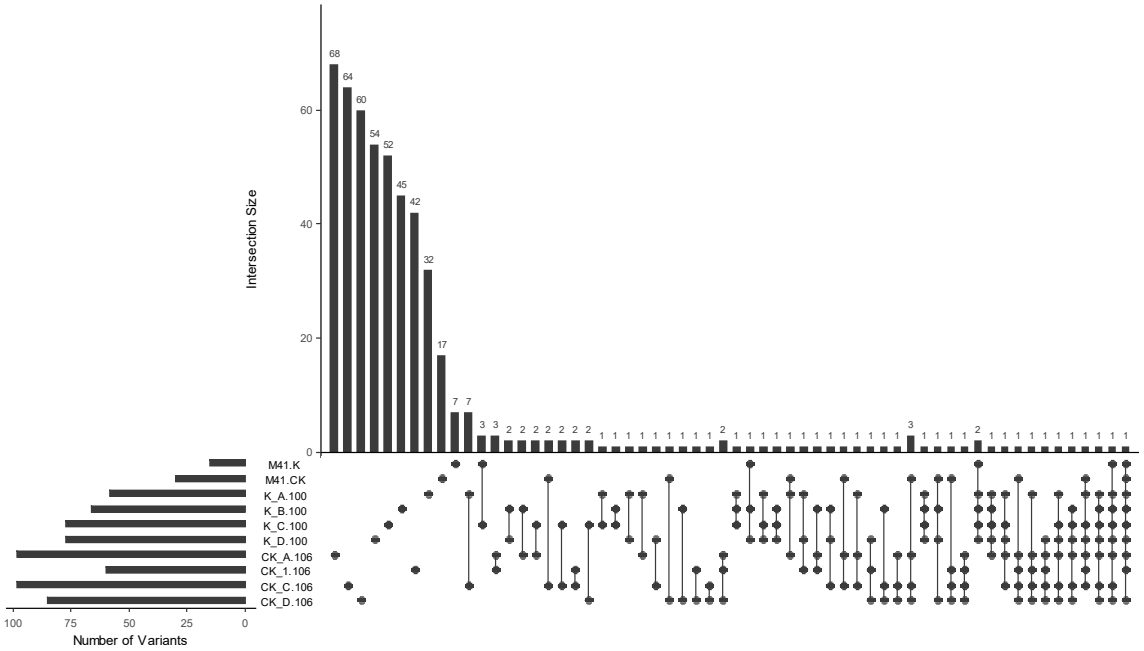


Figure 5.11: UpSetR plot for the comparison of all variants called in attenuated M41-K and M41-CK viruses. *The total number of called variants per passage is indicated in the left bar graph. Variants identified in each passage (i.e. a set) was compared to every other passage. The dots and lines plot indicate the comparison being made (i.e. intersection between sets) with the bar chart (top) indicating the number of variants within that intersection. Singular dots indicate variants unique to that passage. M41-CK and M41-K present variants detected in the original populations are plotted in the graph. All variants are called against their own respective reference sequence.*

5.4. Discussion

To better inform as to the process of attenuation by egg passaging and to examine the contribution of *de novo* mutation in the evolution of IBV, a clonal isolate of IBV (M41-K) was passaged in four separate replicates analogous to the passaging of M41-CK described in Chapter 4. Assessments of pathogenicity were performed *in vivo* and confirmed that M41-K was virulent and that each egg-passaged M41-K was attenuated. The starting, intermediate and final M41-K viruses were all deep-sequenced using Illumina MiSeq.

Analysis of reads obtained for the M41-K initial virus indicated that the virus was not of the expected sequence (M41-CK with 9 synonymous mutation) nor indeed was the virus truly clonal. While further work would be required in demonstrate why there are fixed mutations different between the two viruses, these were likely introduced in the Vaccinia virus stage as an artefact of recombining the various segments in the Vaccinia virus backbone (as described in the introduction of this chapter). Subconsensus differences are more likely to have occurred during the process of rescuing the rIBV from Vaccinia virus and the generation of a working stock in eggs. One notable different between the two viruses (M41-CK and M41-K) was a single nucleotide insertion (T25319TA) in M41-K which results in ORF4b protein possessing a different sequence at both a nucleotide and protein level in the two viruses. Regardless of whatever form ORF4b is expressed in or how the differences between M41-CK and M41-K were introduced, M41-K is still virulent and the process of egg-passaging attenuates the virus.

The observations made during the M41-CK analysis (Chapter 4) for the intermediate and final viruses are also represented in passaging of M41-K. Firstly, though passaging results in a net gain of variants, this gain is not continuous nor proportional to increasing passage (Figure 5.4). This indicates that variants can be both acquired and lost during the course of passaging. Tracing the frequency of

variants detected in the starting material over the course of passaging (Figure 5.5) also demonstrates the variability of individual mutations both between lineages and passages. That being said, one mutation G18620GT occurs at consistently low frequency across passaging for M41-CK and M41-K. Such mutation would result in premature truncation of replicase1ab polyprotein and non-expression of nsps 15 and 16. While both of these nsps can be linked to attenuation, nsp15 mutants are non-pathogenic (Deng et al., 2017) and nsp16 mutants are used as vaccine candidates (Menachery et al., 2014), the low prevalence of this mutation in the population and its absence from all attenuated IBVs (Figure 5.11) makes this possibility unlikely. It would however fit with the observation that the replicase gene of IBV is a determinant of pathogenicity (Armesto et al., 2009). Indeed, it would be unclear has to how an IBV could survive without the capping activity of nsp16 preventing recognition of viral mRNAs (Decroly et al., 2008) to generate virus proteins for propagation.

In contradiction to M41-CK (possessing no shared attenuating mutations), attenuated M41-K has a total of six mutations shared between multiple attenuated lineages. The frequency of each of these mutations varies significantly, with four of the six mutations occurring at the same site (position 25319). This is the site in which M41-CK and M41-K differ by a single nucleotide insertion. Over the course of passaging, each M41-K virus encodes to some degree an alternate form of ORF4b (compared to M41-K which itself is mutated compared to M41-CK) either resulting in the M41-CK form of ORF4b or ORF 4b protein in the third frame (Figure 4.14). It was originally proposed that modifications in ORF4b could be responsible for attenuation, however ORF4b being in a different forms between M41-CK and M41-K may somewhat contradict this. It is worth mentioning however that all four attenuated M41-Ks have a lower frequency of the M41-CK version of ORF4b at the end of passaging compared to M41-K.

Like with M41-CK, it was noted that M41-K shows examples of co-mutation, with two variants fluctuating concordantly over the course of passaging. While exactly the same examples in M41-CK (G22220T, G22236T, G22281T, and G22916C, see 4.3.6) were not present in M41-K, there were other instances unique to M41-K (e.g. A21615C and T21383G in Lineage C and D (Figure 5.5C and 5.5D)). Further instances of co-mutation were identified, continuing until a point after which the frequency of two mutations diverge and are no longer shared. Combined with the argument that the passage 10 frequency of variants in almost identical across lineages, it would perhaps such that the pathway of attenuating M41-K can be shared at a single nucleotide level to some degree. This does not last for the entire duration of passaging as, despite the similarities, the majority of mutations in each lineage and each passage are unique.

The evidence provided here therefore suggests that the serial passaging of M41-CK and M41-K present broadly similar genetic patterns as the viruses evolve and become attenuated. However, though the destination is the same (i.e. attenuation), the evolutionary pathway and the intermediate steps towards the endpoint can be different both between replicate passaging of the same virus (lineages) and between viruses (M41-K and M41-CK). This may therefore suggest that there are multiple “areas” in sequence space that result in attenuation.

Chapter 6 – Towards unmasking the coding potential of IBV and the generation of a rationally attenuated virus

6.1. Declaration

Some of the contents of this chapter have been included in manuscripts currently in preparation. At the time of writing, it is the intention to submit this manuscript under the working title "Identification and confirmation of multiple novel non-canonically transcribed sub genomic mRNAs produced by avian coronavirus Infectious Bronchitis Virus" to the Journal of Virology. The authors of this publication are as follows;

Michael S. Oade^, Sarah Keep^, Filip Lidzbarski-Silvestre, Kirsten Bentley, Phoebe Stevenson-Leggett, Graham L. Freimanis, Nicholas D. Sanderson, John A. Hammond, Richard C. Jones, Paul Britton, Erica Bickerton

^Denotes joint primary authors

The work presented here is primarily the author's sole effort with contributions from colleagues as listed below.

- Characterisation of generated rIBVs *in vivo* (Figure 6.7) was performed in injunction with members the Animal Services Unit and the Coronaviruses group (Pirbright Institute) Ciliary activity assessments were performed blind by Erica Bickerton and Phoebe Stevenson-Leggett.
- Northern blot and leader-body PCR for the formation of a sg mRNA relating to the 3′ UTR (Figure 6.8A and B) was performed by Sarah Keep.
- Reads obtained for assessing coding potential of IBV (Figure 6.11) were generated in conjunction with Graham Freimanis.
- PCR products for sequencing were generated by Sarah Keep (Table 6.1, Figures 6.13 and 6.14)

6.2. Introduction

Results from previous chapters have indicated that the 3′ UTR of IBV exhibits a notable level of change when the virus is serially passaged in embryonated eggs for the purposes of attenuation. The 3′ UTR is subdivided into two regions; the hypervariable and conserved regions. The hypervariable region, immediately preceding the N gene is non-conserved and, for some IBVs such as M41, is deleted (Britton and Cavanagh, 2017). The second, conserved region, is highly conserved across IBVs (Britton and Cavanagh, 2017). The 3′ UTR is known to possess secondary structures important for the regulation of RNA synthesis (Williams et al., 1999). These secondary structures occurring within the conserved region include a stem-loop and pseudoknot and are postulated to be involved in the RNA switch in the different stages of RNA synthesis (Goebel et al., 2004a, Hsue et al., 2000). Indeed it is believed that structures occurring within both the 5′ and 3′ UTRs act to circularise the genome and further promote the templating switching (Sola et al., 2005, Zuniga et al., 2004). The 3′ UTR is hence been implicated in both genome replication and transcription.

Coronaviruses employ a complicated mechanism for the transcription of sg mRNAs occurring through a process known as discontinuous synthesis. In a model first proposed in 1995 (Sawicki and Sawicki, 1995), discontinuous synthesis is controlled by the presence of two TRSs, one located at the 5′ end of the genome (TRS-L) and the other located upstream of every encoded gene (TRS-B). During negative-strand synthesis, the RTC can template switch from a gene to the leader sequence at the 5′ end of the genome by complementary base pairing between a newly synthesised strand of RNA and the virus genome. The site at which this recombination occurs is referred to as a core sequence (CS-L or CS-B). The end result is therefore a nested set of negative sense sg mRNA hallmark of *Nidovirales* which are co-terminal with both the 5′ and 3′ ends of the genome.

In IBV the proposed CS consensus sequence is CTTAACAA however naturally occurring variations of the CS-B exist suggesting that is not an absolute requirement for the CS-L and CS-B to match perfectly. Indeed, this discrepancy has questioned whether a existence of a true core-sequence for IBV exists (Dinan et al., 2019). Transcription is further complicated by the presence of functional non-canonical CS-Bs, for example gene 4b, which is transcribed from a shortened non-canonical CS-B of only three nucleotides (CAA) (Bentley et al., 2013). Our current model of transcription therefore does not necessarily fully explain the generation of sg mRNAs. The arterivus Simian hemorrhagic fever virus (SHFV) for example was initially thought to encode 6 sg mRNAs as detectable by Northern blot of infected cell extracts (Zeng et al., 1995). Further studies of SHFV transcripts have incrementally raised this number, with a recent publication identifying a total of 96 functional TRSs (Di et al., 2017). There is therefore potential for coronaviruses to encode a repertoire of as yet unidentified genes.

Previous work has demonstrated that it is possible to functionally replace 3′ UTR of MHV with that of SARS-CoV and produce viable virus (designated Alb424), however replacement with alpha- and gammacoronavirus 3′ UTRs was not possible (Goebel et al., 2004b). Growth profiles of Alb424 suggest a higher initial titre but lower final titre compared to wildtype MHV. The pathogenicity of Alb424 was not assessed however the two viruses (MHV and Alb424) present different plaque phenotypes. The possibility of the 3′ UTR being a pathogenicity factor for coronaviruses has not been studied directly. Nevertheless, the possibility for viral RNA to interact through RNA-RNA and RNA-protein interactions includes the potential that reported pathogenicity-determining changes outside of the 3′ UTR may interact with 3′ UTR to indirectly influence pathogenicity (Weiss and Navas-Martin, 2005).

Following the results of sequencing analysis and to establish the involvement of the 3′ UTR in governing IBV pathogenicity, a rIBV was designed that possesses the backbone of a pathogenic IBV (M41-K) with the 3′ UTR of a non-pathogenic IBV (Beau-R). The growth characteristics and pathogenicity of this virus (M41K-BeauR-3′UTR) was investigated by *in vivo* and *in vitro* methods. Furthermore, to characterise the possibility that IBV encodes novel genes and their potential involvement in pathogenicity, Illumina HiSeq reads of virus-infected allantoic fluid were both screened for the presence of leader sequence and for reads aligning at two non-consecutive positions on the virus genome.

6.3. Results

6.3.1. Rationale for generation of a M41-K virus with a Beau-R 3′ UTR

The 454 data (Chapter 3) and Illumina data (Chapter 4 and 5) indicated a high degree of mutation in the 3′ UTR after passaging virulent M41-CK/M41-K in embryonated eggs to attenuate the virus. The 3′ UTR is defined here as the section of the genome between the 3′ end of Nucleocapsid gene and the poly(A) tail and is subdivided into two regions; the hypervariable region and the conserved region. The sequences of M41-K and Beau-R (NCBI accession number: AY311317.1) over the 3′ UTR were aligned using MAFFT with default parameters (Katoh and Standley, 2013) and subsequently visualised using UGENE (Okonechnikov et al., 2012). Three SNPs (C27369T, T2797C and A27569T) were identified in M41-K versus Beau-R (Figure 6.1). A 185 nt deletion in the hypervariable region of M41-K 3′ UTR was also detected (Figure 6.1). All positions and mutations are relative to Beau-R genome.

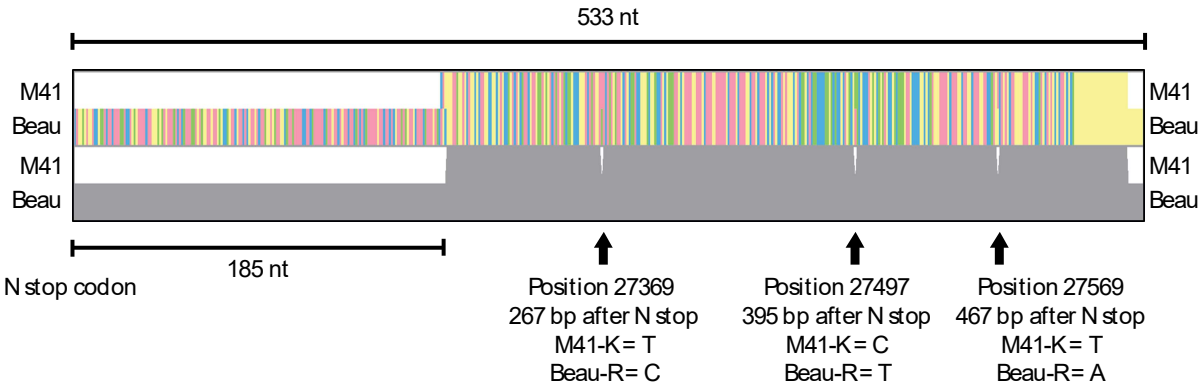


Figure 6.1: Sequence alignment of the 3′ UTRs of Beau-R and M41-K. *The 3' UTR of M41-K and Beau-R (AY311317.1) were aligned using MAFFT (with default parameters) and subsequently visualised in UGENE. Three SNPs were identified in this region occurring at 267, 395 and 467 nt after the Nucleocapsid stop codon. A 185 nt segment is deleted in M41-K compared to Beau-R. The M41-K sequence also has a shorter poly(A) tail compared to the Beau-R sequence. Coordinates are given relative to Beau-R sequence.*

6.3.2. Reverse genetics and recovery of the rIBV M41K-BeauR-3′UTR

To generate a rIBV possessing the backbone of M41-K and the 3′ UTR of Beau-R, the reverse genetics system (see 2.10) was used. This Vaccinia virus based system requires a plasmid possessing the desired IBV modification (here the Beau-R 3′ UTR sequence) flanking by acceptor sequence (here M41-K within Vaccinia virus genome). This plasmid also contains a *gpt* gene from *E. coli* which is used as a selection marker.

A plasmid pGPT-M41-BeauR-3′UTR was designed to include the entire 3′ UTR of Beau-R (533 bp) flanked at its 5′ end with 400 bp from M41-K (relating to nucleocapsid gene) and by 400 bp vaccinia virus sequence at its 3′ end (Figure 6.2). The 400 bp either side of a desired modification are essential for driving the homologous recombination event between donor plasmid and acceptor vaccinia virus genome in the reverse genetic system. The plasmid (Figure 6.2) was ordered and synthesised by GeneArt® (Invitrogen).

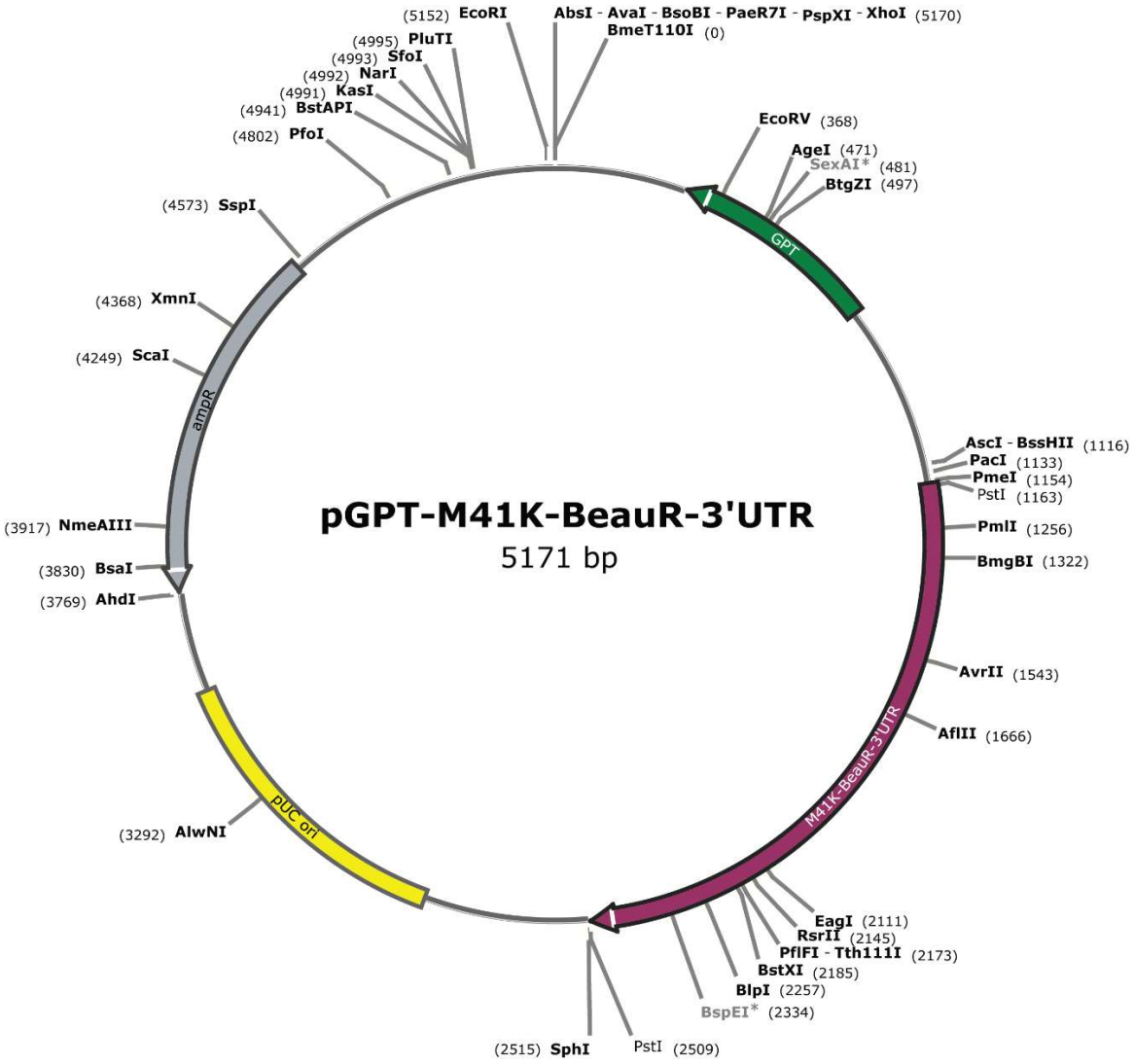


Figure 6.2: Plasmid map of pGPT-M41K-BeauR-3'UTR. *A DNA sequence consisting of the Beau-R 3'UTR flanked by M41-K (3' terminus of Nucleocapsid) and vaccinia sequence was inserted into pGPT-NEB193_B782 plasmid for the replacement of M41-K 3'UTR with that of Beau-R. The designed plasmid was ordered and synthesised by GeneArt® (Invitrogen). Plasmid map was generate using Snapgene Viewer.*

The plasmid pGPT-M41-BeauR-3′UTR was used for the replacement of the M41-K 3′ UTR with that of Beau-R utilising the reverse genetics system (see 2.10) (Casais et al., 2001). After homologous recombination and subsequent passaging in the presence and absence of *gpt* selection agents, small scale DNA extractions were performed on rVVs containing a full length IBV sequence to confirm successful modification of the IBV sequence and loss of the *gpt* gene. The absence of the *gpt* gene was confirmed by PCR with no bands visible when potential PCR products were separated by gel electrophoresis (not shown). Confirmation of the desired IBV sequence was made by PCR and subsequent separation by gel electrophoresis (Figure 6.3). As the modification made to rVV would result in an increase in sequence length (~185 nt), rVVs possessing a larger band size were preferentially chosen for sequencing. U23 and U34, derived from separate rVV lineages were selected to proceed into the next phase based on confirmatory sequencing results.

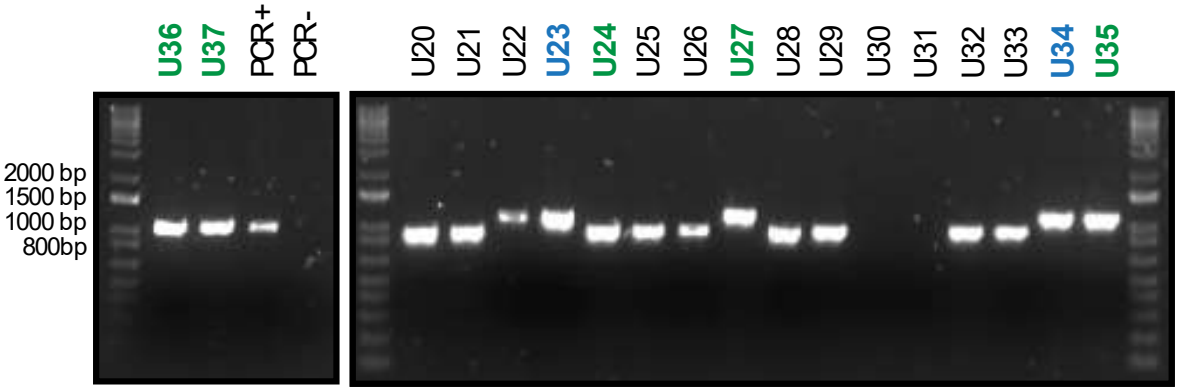


***Figure 6.3: Screening rVVs for desired IBV sequence**. DNA was extracted from small-scale rVV stocks and screened for the presence of the desired modification by PCR using primers M50 and 93/100. Resulting PCR products were then separated on a 1 % agarose gel by gel electrophoresis. Each lane represents a separate vaccinia virus with those marked in green selected for sequencing. Vaccinia virus isolates marked in blue (U23 and U34) were sequenced and selected for rescue. 1 Kb plus ladder was run adjacent to samples to estimate band size. PCR+ and PCR- are reaction controls set up with a known DNA sample (Beau-R cDNA) or water, respectively. Band sizes matching the PCR+ (approximately 1.1 kb) were therefore selected. Gel was stained using SYBR Safe.*

Larger scale rVVs stocks were grown using BHK-21 cells and DNA extracted using phenol:chloroform. To confirm that the DNA extraction had produced a full-length genome and that DNA had not been damaged, resulting DNA was digested with *SalI* and separated by pulsed field gel electrophoresis (PFGE) (Figure 6.4). *SalI* restriction sites either side of the IBV insert in Vaccinia virus result in an approximately 32 kb *SalI* fragment (Casais et al., 2001). PFGE was performed with aid from Sarah Keep (Pirbright Institute).

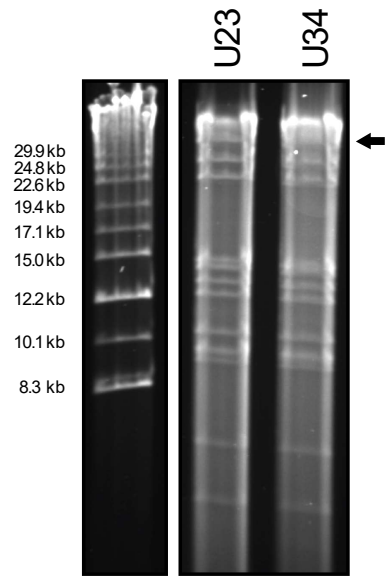


Figure 6.4: Pulsed-field gel electrophoresis separation of SalI-digested Vaccinia virus DNA. *DNA was extracted from larger scale rVV stocks and digested using SalI. Resulting DNA was separated by pulse-field gel electrophoresis overnight. CHEF DNA 8-48kb ladder (Biorad, 170-3707) was run adjacent to samples to allow to band size (bp) approximation. A band of approximately 32 kb is indicative of an intact whole rIBV cDNA (arrow). Gel was stained using ethidium bromide.*

Upon confirmation of rVVs possessing the desired modification, rVV DNA was transfected into primary CKCs infected with rFPV-T7 for the rescue of infectious recombinant IBVs. Recovered IBVs were passaged on CK cells three times with RNA extracted from supernatant of each resulting passage. The presence of IBV was confirmed by RT-PCR screening (Figure 6.5). Stocks of two viruses originating from two rVVs (four viruses total) were selected for growth in embryonated eggs to generate a working stock. Here, rVVs from two separate lineages (U23 and U34) were chosen as they were generated from two separate recombination events in the vaccinia stage. Two rIBV rescues derived from the same rVV lineage were grown up to recompense for any potential mutations that occur during the rescue process or sequence differences between the two clones. Working stocks for each of the four viruses were titrated and the sequence over the desired modification confirmed by PCR prior to downstream usage.

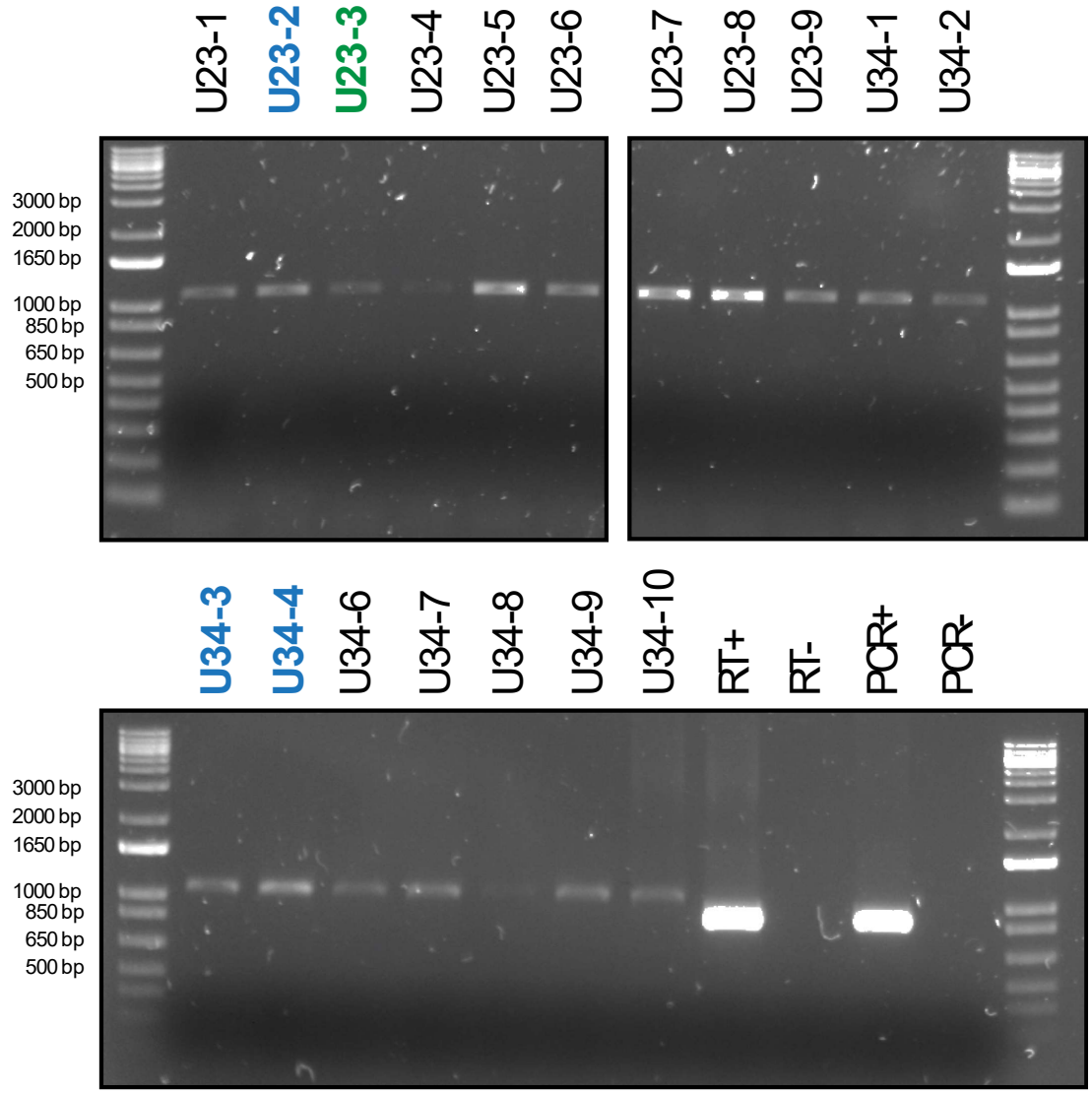


Figure 6.5: Confirmation of rIBV recovery by RT-PCR in P1 CKC supernatant. *To confirm successful recovery of rIBV, RNA was extracted from CKC supernatant and screened for the presence of IBV by RT-PCR using primers M50 and 93/100. Resulting PCR products were then separated on a 1 % agarose gel by gel electrophoresis. Each lane represents a separate rescue from either vaccinia virus isolate U23 or U34. Due to a contamination issue in the rescue phase, U23-10 and U34-5 were discarded. Working stocks of viruses U23-2, U34-3, U34-4 (marked blue) and U23-3 (marked green) were made. RT and PCR positive controls are M41-CK derived (~900 nt) whereas the desired product size is ~1.1 kb.*

6.3.3. *In vitro* characterisation of M41K-BeauR-3′UTR viruses

To assess whether exchanging the 3′ UTR of M41-K with that of Beau-R affects the ability to grow *in vitro*, confluent 6 well plates of CKCs were infected with 5 x 10^4 PFU virus per well and supernatant harvested at 1, 24, 48, 72 and 96 hours post infection. Supernatant was then assessed for infectious progeny by titration in CK cells (Figure 6.6). In parallel to this, the sensitivity of replication to temperature was also assessed as a potential means of characterising *in vivo* pathogenicity (explained in Chapter 4). Growth curves at both 37°C and 41°C were therefore performed with resulting supernatants titrated on CKCs for assessment of viral progeny (Figure 6.6). BeauR-rep-M41-struct, a virus possessing the replicase machinery of Beau-R and the structure proteins of M41-K (Armesto et al., 2009), was included in this series of growth curves as this virus also possesses a mismatch in the origin of the UTR (i.e. a Beau-R 5′ UTR and M41-K 3′UTR), albeit inversed compared M41K-BeauR-3′UTR.

Growth profiles for each rIBV resemble M41-CK at 37°C temperatures with a peak titre at 24 – 48 hours post infection. Growth of U23-2, U23-3 and U34-3 peaked at 24 hours post infection with an average titre of 5.71, 5.46 and 4.56 $\log_{10}$ PFU / mL, respectively, whereas U34-4 peaked at 48 hours with an average titre of 5.51 $\log_{10}$ PFU / mL. Similarly Beau-R and BeauR-rep-M41-struct grown at 37°C, the titre peaks at 24 hours (average titre 5.90 $\log_{10}$ PFU / mL and 4.79 $\log_{10}$ PFU / mL, respectively) however later time points show steeper decline in titre compared to the M41-based viruses with Beau-R presenting a growth profile arguably unique compared to the other viruses. The results here indicates that inserting Beau-R 3′UTR into M41-K has not been detrimental to virus replication.

As demonstrated previously (Figure 4.1), Beau-R cannot replicate effectively at 41°C with each time point after 1 hour post-infection being significantly different ($P \leq 0.0001$) to its 37°C counterpart. This trend is also observed in BeauR-rep-M41-struct with a total elimination of virus replication by 24 hours. Also as demonstrated previously (Figure 4.1), the titre of M41-CK at 37°C versus 41°C after 1-hour post infection is statistically different but does not demonstrate the same degree of reduction compared to Beau-R and BeauR-rep-M41-struct. Generally there is no significant difference in virus titre for each time point for rIBVs grown at 37°C and 41°C with the growth profile resembling M41 not Beaudette, indicating that *in vitro* replication of these rIBVs is not temperature sensitive.

With temperature sensitivity not presenting a viable means of characterising *in vivo* pathogenicity, there is no alternate other than to perform assessment of clinical signs in *vivo* to characterise these viruses.

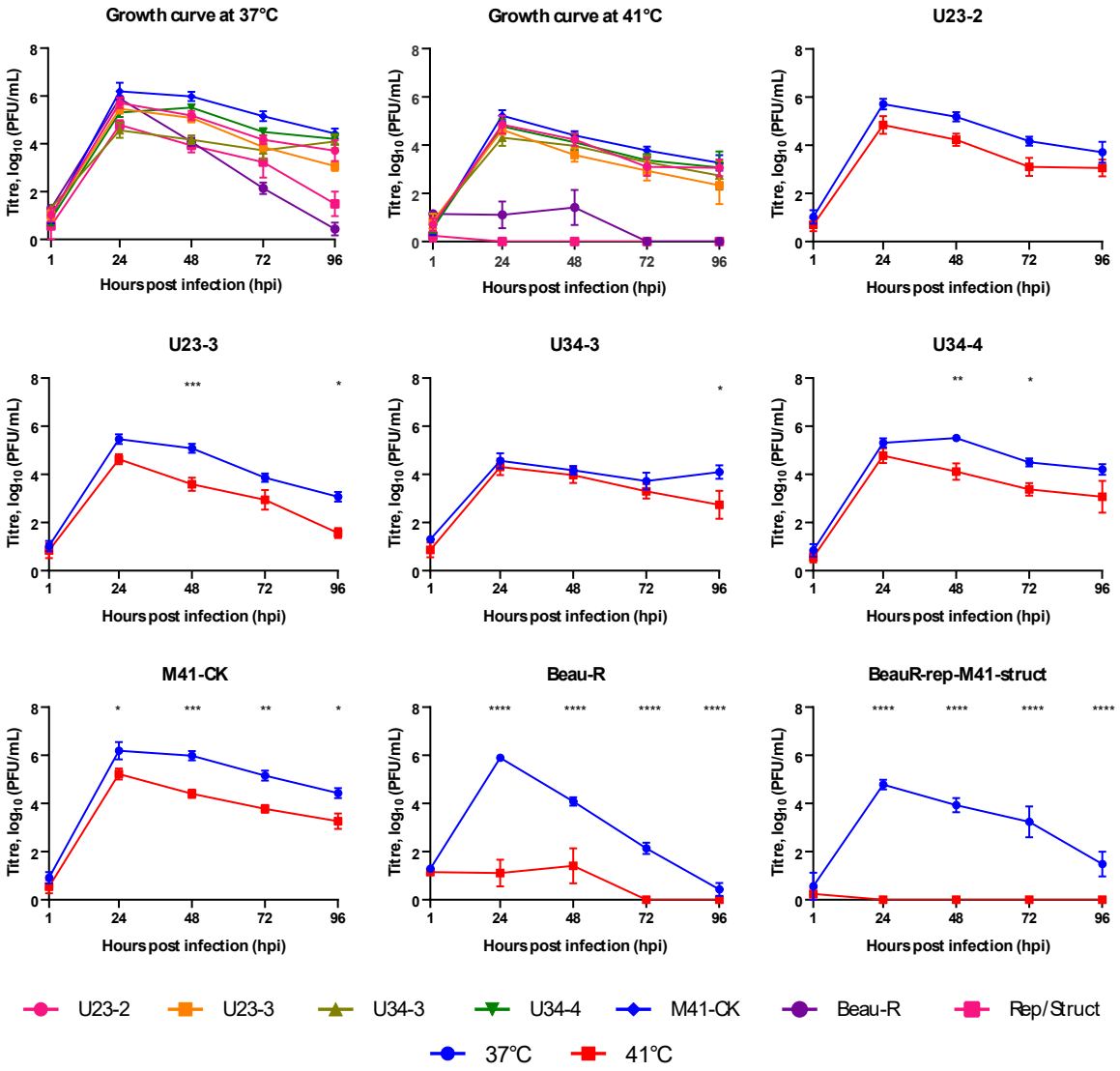


Figure 6.6: Growth kinetics of M41K-BeauR-3'UTR viruses grown at 37°C and 41°C on CK cells. *Chick kidney cells were infected with 5 x 10^4 PFU IBV and incubated at either 37°C or 41°C. Supernatant was harvested at 1, 24, 48, 72 and 96 hours post infection and titrated on CK cells. BeauR-rep-M41-struct (Rep/Struct) has been included as in these growth curves as a virus that possesses a mismatch between the 5' and 3' UTRs. Each value represents the means of three replicates with the SEM plotted as error bars. Graphs have been presented for each virus at both temperatures, and both temperatures with different viruses. Both represent the same data. P values were determined by Sidak's multiple comparison test (Two-way ANOVA) performed on each dataset (virus) individually. *, P ≤ 0.05; **, P ≤ 0.01; ***, P ≤ 0.001; ****, P ≤ 0.0001.*

6.3.4. M41K-BeauR-3′UTR is attenuated *in vivo*.

As there were no observed differences in replication *in vitro* between the replicates of M41K-BeauR-3′UTR, U23-3 was selected for *in vivo* characterisation. Groups of fifteen 8-day old Rhode Island Red chickens were inoculated with 1 x 10^5 PFU IBV or mock infected with 0.1 mL PBSa via intranasal and ocular routes, equating to roughly 25 µL in each eye and each nostril. Snicking (Figure 6.7A) and rales (Figure 6.7B) were recorded daily between days 3 and 7 post infection with ciliary activity assessed at post mortem on days 4 and 6 post infection (Figure 6.7C).

Snicking (Figure 6.7A) for M41K-BeauR-3′UTR infected birds was minimal and comparable to Beau-R non-pathogenic control on all assessed days post infection. As with Beau-R and mock infected groups, no rales were observed in M41K-BeauR-3′UTR infected bird at any point during the study (Figure 6.7B). The M41-CK infected group demonstrated both snicking and rales and a severe reduction in ciliary activity typical of the virus and as seen previously (Armesto et al., 2009, Ellis et al., 2018, Hodgson et al., 2004).

Ciliary activity is used as an indication for the presence of IBV (Cavanagh et al., 1997). M41K-BeauR-3′UTR infected birds have an average ciliary activity of 3.14 and 3.8 for days 4 and 6 respectively and is comparable to mock-infected birds who have an average ciliary activity of 3.04 and 3.8, also respectively. The ciliary activity of Mock infected and M41K-BeauR-3′UTR infected birds are not deemed to be significantly different on both assessed days by Sidak's multiple comparison test (P = 0.9992 and > 0.9999 for days 4 and 6 post infection, respectively). Ciliary activity for M41K-BeauR-3′UTR infected birds is significantly different to the ciliary activity of M41-CK virulent control infected birds on both days post infection (P < 0.0001 for both days). Compared to Beau-R infected birds, ciliary activity of M41K-

BeauR-3′UTR was not significantly different at day 4 post infection (P = 0.0670), but significantly different at day 6 (P = 0.0003).

Given the nature of the modification made and that growth kinetics of M41K-BeauR-3′UTR are similar to M41-CK *in vitro*, it is largely unexpected that the M41K-BeauR-3′UTR demonstrates a highly similar phenotype to Beau-R and is hence deemed to be attenuated.

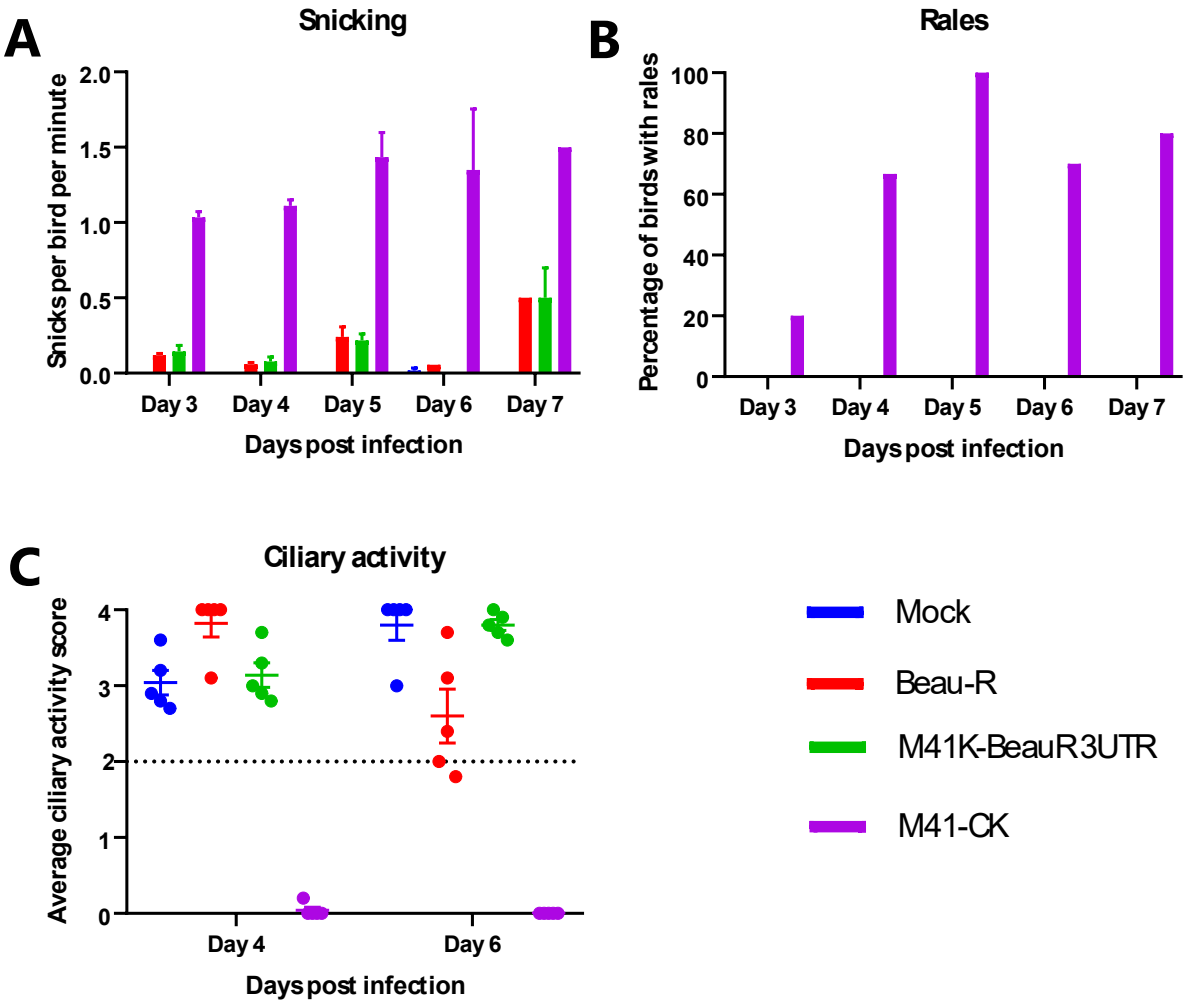


Figure 6.7: M41-BeauR-3'UTR is attenuated in vivo. *Groups of fifteen 8-day old SPF RIR chicks were inoculated with 1 x 10^5 PFU virus or mock infected with PBSa via intranasal and ocular routes. Clinical signs including snicks (A) and rales (B) were assessed between days 4 and 6 post infect by observation of 10 x 1 mm cross sections under light microscope from five randomly selected birds per group. Ciliary activity was scored as follows; 100 % activity = 4, ≈ 75 % = 3, ≈ 50 % = 2, ≈ 25 % = 1, 0 % = 0. The mean ciliary score of these birds is plotted with SEM plotted as error bars. Ciliary activity was scored blindly by Phoebe Stevenson-Leggett (Pirbright Institute) and Erica Bickerton.*

6.3.5. Confirmation of a sg mRNA relating to the 3′ end of the IBV genome

To investigate this mechanism of attenuation and the involvement of the 3′ UTR in governing pathogenicity, previous work performed by the Coronavirus group at the Pirbright Institute relating to a potential sg mRNA the 3′ end of the genome, was followed up on. To confirm the presence a sg mRNA encoded at the 3′ end of the viral genome, confluent 6 well plates of CKCs were infected with either Beau-R or M41-CK for a period of 24 hours after which supernatant was removed and total RNA extracted from CKCs. M41-CK RNA was subject to analysis by leader-body PCR whereas Beau-R was subject to Northern blot analysis and leader-body PCR (Figure 6.8). This work was performed by Sarah Keep and Paul Britton (Pirbright Institute) and has been reanalysed by Michael Oade in the context of M41K-BeauR-3′UTR.

It is worth mentioning that in the time that this thesis and associated manuscript has been in preparation, the presence of a sg mRNA relating to this region of the genome has subsequently been proven (An et al., 2019, Dinan et al., 2019).

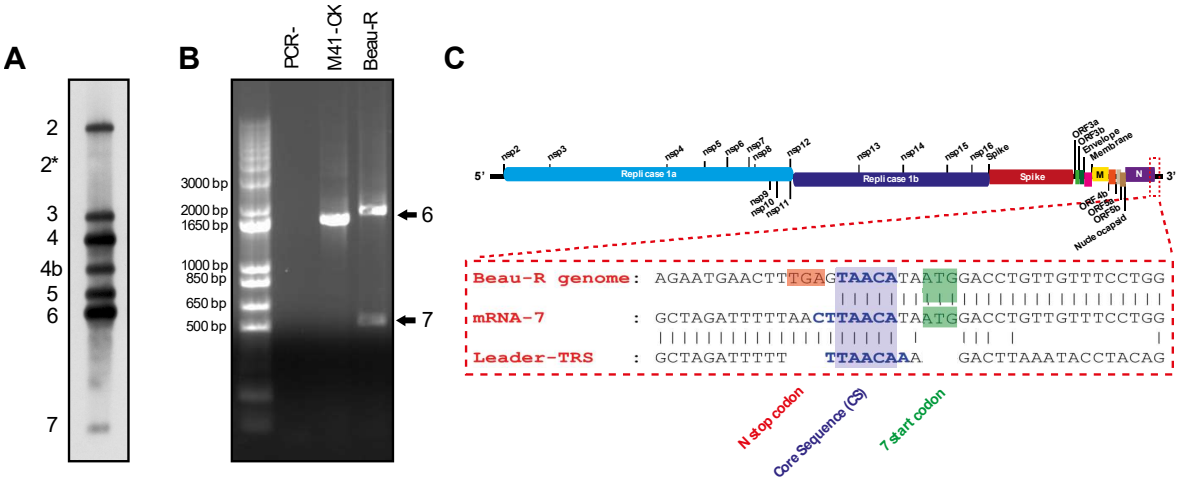


Figure 6.8: Confirmation of a sg mRNA relating to the 3' end of the IBV genome. A. *Northern blot analysis of Beau-R infected whole cell lysate. A clearly distinguishable band is present at the bottom of the gel at relatively low abundance. The location of this band corresponds to the area of the genome between gene 6 (Nucleocapsid) and the 3' UTR. A faint band is also observed between gene 2 and 3 situated where gene 2* would be located.* ***B.*** *Leader-body PCR analysis of Beau-R and M41-CK infected CKC whole cell lysate. Two bands were produced after separation on a 1 % agarose gel in Beau-R, but only one band is observed for M41-CK.* ***C.*** *Sequencing of leader-body PCR products. Sequencing of the smaller band confirms a match to the 3' end of the virus genome, but the sequence also possesses a leader sequence corresponding to the 5' end of the genome. Reads do not confirm an alignment to the CS-B with perfect homology and instead match only five nucleotides (TAACA). Northern blot and leader-body PCRs were performed by Sarah Keep. Resulting sequencing was analysed by Michael Oade.*

Northern blot analysis of Beau-R infected whole cell lysate produced a clearly distinguishable band present at the bottom the gel. The location of this band corresponds to the area between gene 6 and the end of the virus genome. A faint band is also observed between gene 2 and 3 situated located (Figure 6.8A). Two bands are present for Beau-R after separation of PCR products by agarose gel electrophoresis (Figure 6.8B). The larger sized product (~2,000 bp) corresponds to the expected band size for gene 6 (N protein) transcripts. A smaller product (~500-650bp) is also produced in Beau-R but is not detectable in M41-CK. Sequencing of Beau-R PCR products by PCR cloning confirms the smaller band aligns to the 5′ and 3′ ends of the genome, hence having the characteristics of a sg mRNA (Figure 6.8C). The switch from body to leader however does not occur at a canonical CS-B within the genome instead only matching by nucleotides 3-7 (TAACA) of the CS consensus sequence. This matches previous observations regarding a gene at this location (Dinan et al., 2019, An et al., 2019).

To establish whether this region of the genome could potentially encode an additional protein, the sequence of Beau-R from the end of nucleocapsid to poly(A) tail was searched for the presence of a start codon (AUG) in the direct strand only. Six AUGs were identified resulting in potential peptides ranging from 1 to 15 amino acids in length. Comparing this region of the genome to other IBV strains, it was observed that both Beaudette viruses (Beau-R and Beau-CK) possess a single nucleotide insertion within one of these potential proteins resulting in the introduction of a stop codon. The prospective protein would therefore be prematurely truncated compared to other strains of IBV, leading to a 10 AA peptide as opposed to a 53 – 76 aa protein (Figure 6.9). TCoV was included in this alignment as an indicator for protein conservation across gammacoronaavirus genus.

For future purposes, this gene will be referred to as ORF7.

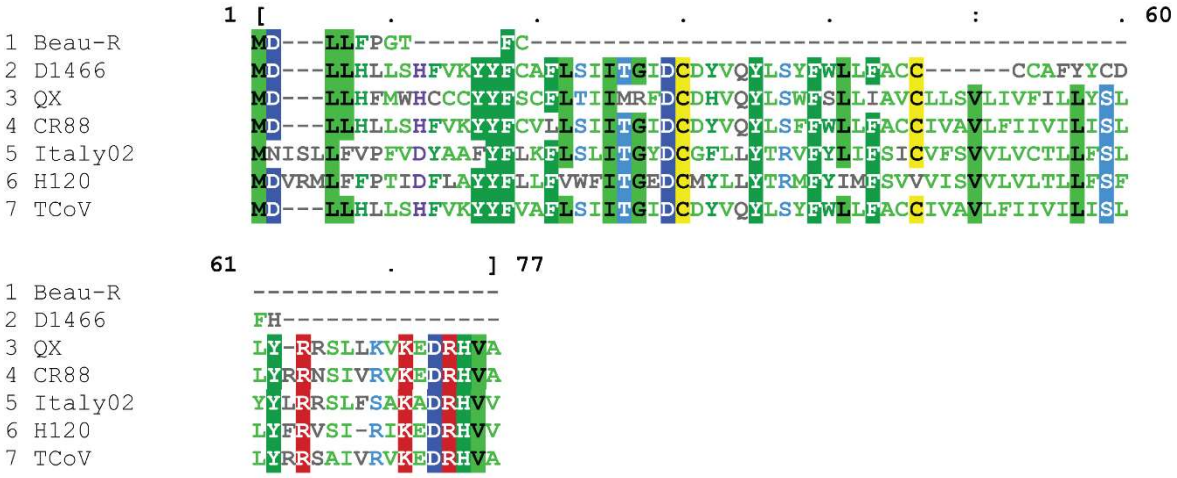


Figure 6.9: Prospective protein alignment in different IBV strains and TCoV. *MView alignment of prospective ORF 7 protein sequences. Gene 7 sg mRNA sequences were screened for the presence of an AUG start codon. A 53 - 77 AA protein sequence was identified in Turkey coronavirus (TCoV) and all IBVs with exception to Beau-R which has a highly truncated 10 AA protein sequence. D1466, QX, and Italy02 are all pathogenic field isolates of IBV. H120 and CR88 are vaccine strain against Massachusetts (M41) strain and 793B strains, respectively. Due to the deletion in M41 in this region of the genome, it was not included in this alignment.*

6.3.6. Expression of ORF7 in M41K-BeauR-3'UTR

To ascertain whether M41K-BeauR-3'UTR expresses ORF7, confluent 6 well plates of CKCs were infected with each 500 µL virus (U23-2, U23-3, U34-3 and U34-4) for a period to 24 hours. Total RNA was extracted from CKCs and leader-body PCR performed (Figure 6.10).

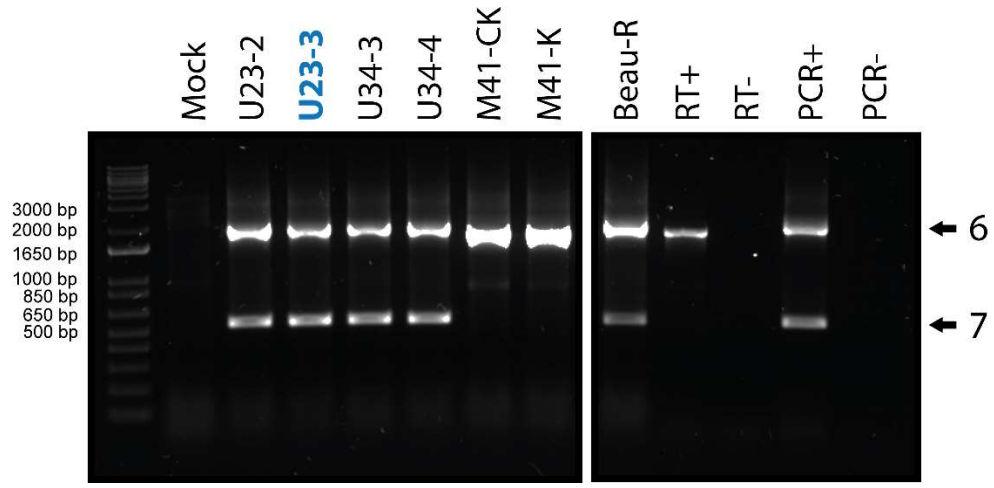


Figure 6.10: Leader-body PCR for confirmation of an ORF7 transcript in M41K-BeauR-3'UTR. *Leader-body PCR analysis of M41K-BeauR-3'UTR whole cell lysate. Similarly to Figure 6.8, two clear bands were produced after separation on a 1 % agarose gel. The larger of two bands (~2,000 bp) corresponds to the expected band size for gene 6 (Nucleocapsid). The smaller band (~500 – 650 bp) corresponds to the expected band size of gene 7. RT+ control = M41-CK infected allantoic fluid, PCR+ control = Beau-R infected allanotic fluid. Both negative controls were loaded with water in place of RNA/cDNA*

The presence of a band at the expected size after separation of PCR products by gel electrophoresis confirms that M41K-BeauR-3'UTR does transcribe ORF7 at the 3' end whereas the M41 viruses (M41-K and M41-CK) do not. Given that Beau-R does transcribe this gene and M41 does not, it is assumed that the modification made to M41K-BeauR-3'UTR has resulted in the transcription of this sg mRNA. Further work must therefore explore whether expression of ORF7 has altered the pathogenicity of M41K-BeauR-3'UTR. Given that ORF7 is however transcribed in

pathogenic and non-pathogenic IBVs alike, it would be assumed that expression of ORF7 does not affect pathogencity.

6.3.7. Transcriptional profiles of Beau-CK and M41-CK utilising HiSeq data

In order to investigate the mechanism of rIBV attenuation, verify ORF7 transcription and to further explore potential transcripts encoded by IBV, a dataset consisting of deep-sequenced Beau-CK and was generated by Graham Freimanis (Pirbright High Throughput Sequencing unit) and reanalysed by Michael Oade. Using the method as described previously, infected allantoic fluid was harvested, clarified by centrifugation and then virus purified by ultracentrifugation. Total RNA was extracted using QIAGEN RNeasy kit and library preparation performed using a NEBNext direction Ultra RNA-Seq kit (NEB). The library pool was quantified using a NEBNEXT Illumina library quantitation kit (NEB) before being diluted and loaded onto a single lane of an Illumina HiSeq 4000. Beau-CK reads were then quality filtered, host-removal performed and a *de novo* consensus sequence made. HiSeq reads for both Beau-CK and M41-CK (Chapter 4, initial passage) were then subject to interrogation as described below. These viruses were chosen to assess the transcriptional profiles of a pathogenic (M41-CK) and non-pathogenic (Beau-CK) IBV.

All quality filtered reads (PQF) were aligned to their respective reference genomes with both BBMap (Bushnell, 2014) and BWA (Li and Durbin, 2009), both using default parameters in paired-end mode. Alignments performed using BWA were examined for chimeric alignments (flagged by 'SA:' in the SAM file) and instigating reads extracted. Chimeric reads (reads aligning to two non-consecutive regions of the genome) were subsequently aligned to their reference genome with BBMap using default parameters (paired-end mode).

As described previously, each sg mRNA possesses a leader sequence originating from the 5′ end of the virus genome. As an additional method of identifying transcript-like sequences within the Illumina dataset, PQF reads from both viruses were subsequently screened for this sequence. For Beau-CK a query sequence of TTAAAAATCTAGC was used which is the reverse complement of the GCTAGATTTTTAA leader sequence occurring immediately upstream of the CS. For M41-CK a query sequence of TTGGAAATCTAGC was used which likewise is reverse complement of GCTAGATTTCCAA leader sequence which occurs immediate upstream of the CS. These sequences were chosen to ensure that only transcript sequences were isolated without unnecessarily excluding positive hits. To avoid duplicate matches occurring in both forward and reverse reads and to obtain longer overall read length, reads were merged using Paired-End reAd mergeR (PEAR) v0.9.11 (Zhang et al., 2014) prior to searching. Leader-matching reads were converted to FASTA format and sequence up to and including the leader sequence removed. The remainder of the read was aligned to the Beau-CK and M41-CK genomes using BBMap.

Coverage for all three modes of alignment was called and plotted (Figure 6.11). Chimeric read coverage (green) indicates positions in the genome covered by reads aligning to two non-contiguous locations on the genome whereas the leader-matching reads (red) originally possessed sequence corresponding to the 5′ end of the genome as per IBV method of discontinuous transcription. Both methods therefore have the potential to capture IBV transcripts or transcript-like sequences.

Peaks in coverage for both chimeric-mapped and leader-mapped reads occur at sites of known transcripts, including genes 2 (S), 3 (3a, 3b, E), 4 (M), 4b (4b), 5, (5a, 5b) and 6 (N), for both viruses. Additional peaks do occur with notable differences observed between the two viruses, however without experimental validation the identification of truly existing transcripts is confounded. On the

assumption that a peak in both chimeric- and leader-mapped data-sets is indicative of a transcript (as seen at sites of known genes), Beau-CK has two distinct additional peaks compared to M41-CK at approximately 21.5 kb (designated gene 2*) and at the 3′ UTR (confirmed as being ORF 7 by sequence similarity). In contrast M41-CK has two peaks at 16.5 kb and 18.0 kb not present in Beau-CK. The detection of gene 2* in IBV has been noted previously (Bentley et al., 2013). Gene 7 has been reported in recently released publications by (An et al., 2019) and (Dinan et al., 2019).

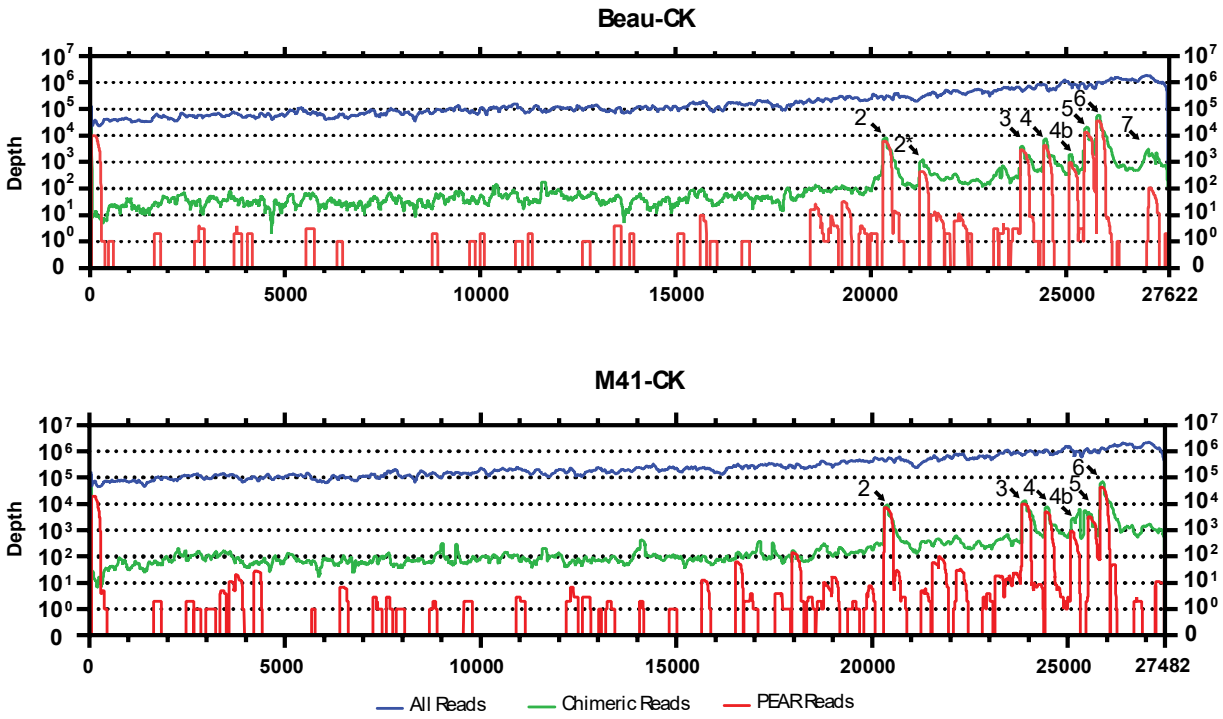


Figure 6.11: Assessing the coding potential of IBV using Illumina HiSeq reads. *Reads of Beau-CK and M41-CK infected allantoic fluid were aligned to their respective genome sequences using three separate methods and coverage plots generated. All PQF reads (blue), chimeric reads (green) and reads possessing a leader sequence (red) were aligned their respective genomes. The trace for the chimeric read coverage plot (green) therefore indicates positions in the genome covered by reads that align in two places whereas the trace in the leader-matching (red) coverage plot indicates positions that are covered by reads that also possess the leader sequence. Coverage peaks occur at the site of known transcripts (2, 3, 4, 4b, 5 and 6) for both Beau-CK and M41-CK as do peaks of unknown origin.*

The sporadic increases in either chimeric read coverage or leader-possessing read coverage are currently unexplainable. Given the complexity of discontinuous transcription, such peaks may represent sites of recombination within the virus. If true, the reason why there is a difference between the two viruses and the importance of this is unclear. While some of these peaks occur at relatively low abundance, it is important to remember that the material was not generated by

a means best suited for transcripts. Moreover, transcript detection within allantoic fluid is perhaps unanticipated given the assumption that IBV virions do not include transcripts.

6.3.8. Validating the presence of leader-possessing read alignments not associated with known transcripts

Our current model for coronavirus transcription does not allow for the transcription of genes within the replicase portion of the virus, starting at roughly 500 bp to 20,000 bp in the virus genome. Figure 6.11 however suggests that, while low in abundance, some leader-containing sequences do exist within the region. Leader-body PCR utilises two primers at opposing ends of the viral genome. Under standard PCR conditions with typical elongation times it is not possible to amplify whole genome cDNA due to the distance between the two primers. PCR products generated using these primer pairs are therefore of virus-produced transcripts generated by discontinuous synthesis.

To validate the presence of leader-possessing reads aligning to the regions of the virus genome not associated with known genes, confluent 6 well plates of CKCs were infected with Beau-CK or M41-CK for 24 hours. Total RNA was then extracted from CKCs and subject to leader-body PCR. A previous iteration of Figure 6.11 indicated a high number of leader-possessing reads mapping proximal to the start of replicase 1b. Given that replicase 1b sg mRNA is believed to encode as part of full-genome RNA, the possibility would go against current knowledge and was hence explored. BG113 was chosen as the reverse primer as it pairs with the virus genome at approximately 14.0 kb whereas the replicase 1a/1b switch occurs at 12.5 kb.

Separation of PCR products by agarose gel electrophoresis does show multiple bands indicating the potential for recombination over this region (Figure 6.12). Subsequent PCR cloning and sequencing confirms these bands do have leader sequences. The sg mRNAs possess a CTTAAACA sequence which has a striking similarity to the canonical CS sequence of CTTAACAA indicating that these transcript-like sequences may derive from a non-canonical CS.

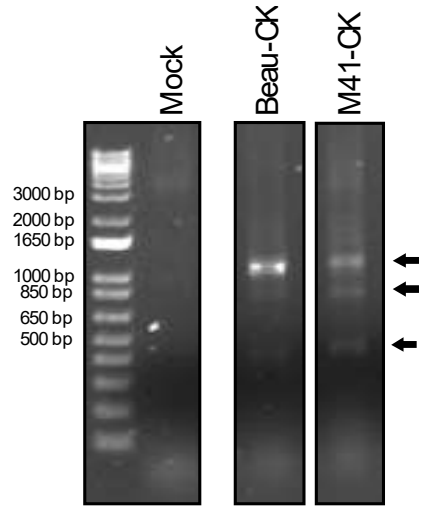


Figure 6.12: Leader-body analysis for the verification of leader-possessing sequences not associated with known transcripts. *Total RNA from Beau-CK and M41-CK whole cell lysate was subject to leader-body analysis using primers, targeting the mid region of the replicase gene at approximately 13.5 kb. Resulting PCR products were separated on a 1.0% agarose gel by gel electrophoresis. Graduation of bands is observed, with multiple bands observed in both M41-CK and Beau-CK.*

This PCR would require further optimisation for the isolation of low abundance PCR products present on the gel as a faint band.

6.3.9 Confirmation of a gene 2* derived transcript and detection of multi-chimeric sg mRNAs

The detection of a transcript, 2*, between genes 2 and 3 as per Figure 6.8 and Figure 6.11 has previously been reported for IBV (Bentley et al., 2013) though was not fully explored. In the preparation of a new publication relating to other potential transcripts in IBV (see 6.1), unpublished data related to the original publication was reanalysed. This involved the reanalysis of sequencing data obtained from leader-body PCR over the gene 2* region from Beau-R infected whole CKC lysate. Work here was originally performed by Kirsten Bentley (formally of the Pirbright Institute). As part of this reanalysis new sg mRNA material was produced and leader-body PCRs performed by as above (see 6.3.8). This new material was generated by Sarah Keep. Sequencing results from both projects was analysed by Michael Oade. The below encompasses data obtained from both projects performed several years apart.

Analysis of available data confirmed that the PCR product generated by leader-body PCR over the prospective gene 2* region does possess a 5′ leader sequence and hence possesses characteristics indicative of viral transcripts. The screening of some colonies however resulted in Sanger sequencing reads that, despite consideration for typical leader-body recombination, could not be matched contiguously with the viral genome. These reads were of high quality not dissimilar from other transcript reads indicating the sequence was true and running sections of the read sequence through NCBI-BLAST confirmed a match to IBV. During this process it was noted that the alignment did not always occur in the same region of the viral genome. Considering this information reads were aligned back to the virus genome with a larger tolerance for sequence insertion/deletion and incorporating the possibility that the sequence may not be in an assumed order (Table 6.1)

Table 6.1 Sequencing of PCR products produced by leader-body PCR over gene 2*

Example #	Alignment 1	Alignment 2	Alignment 3	Origin
1	25 – 58 (Leader)	20,311 – 21,135 (Gene 2, Spike)	N/A	Gene 2 transcript
2	25 – 58 (Leader)	20,311 - 20,561 (Gene 2, Spike)	22,175 – 22,609 (Gene 2, Spike)	Unknown
3	25 – 58 (Leader)	21,249 – 22,145 (Gene 2, Spike)	N/A	Gene 2* transcript
4	25 – 58 (Leader)	25,474 – 25,696 (Gene 5, 5a)	22,383 – 22,609 (Gene 2, Spike)	Unknown

Positions are given relative to AJ311317.1 Beau-R genome.

Utilising these new parameters, all available reads were successfully aligned to the Beau-R genome. Gene 2 and 2* transcripts were identified as originally intended firstly confirming the presence of gene 2* (Figure 6.13). Some transcript-like sequences align to Beau-R in multiple non-consecutive locations (Table 6.1 and Figure 6.14). In Table 6.1, one transcript (Example 2) possessed leader sequence along with sequence originating from two separate locations within Spike protein. Curiously, it was also possible for transcripts (e.g. Example 4) to possess sequence spanning larger distances of the genome in what might be considered an irregular gene order. In example 4, the RTC would be required to skip over sequence presumably already synthesised. Further effort would be required to verify this preliminary finding and to fully resolve the 3′ end of these transcript sequences. If it is confirmed that this is indeed a possibility, it raises questions regarding the template-switching processes and the regulating processes involved.

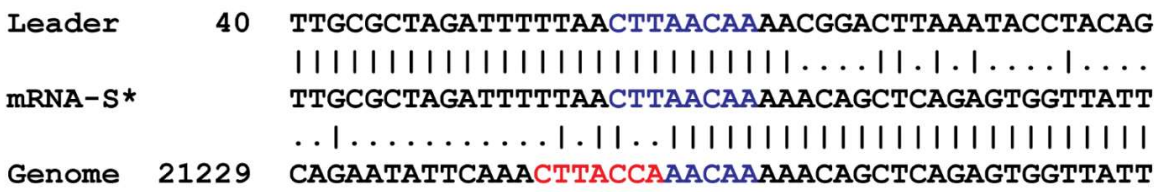


Figure 6.13. Sequence analysis of leader-body PCR analysis from gene 2* transcripts. *The top and bottom lines of the alignment represent the genome sequence of Beau-R at leader and the gene 2* junction, respectively. Positions are given relative to the AY311317.1 sequence. The middle lines presents sequence data and indicates the template switch from genome to leader by alignment. AACAA is consistent in all three alignments and is indicative of the CS-B though this CS-B is non-canonical. Upstream of this switch site, CS-like sequence is highlighted in red. Reads do not support use of the upstream CS-like sequence.*

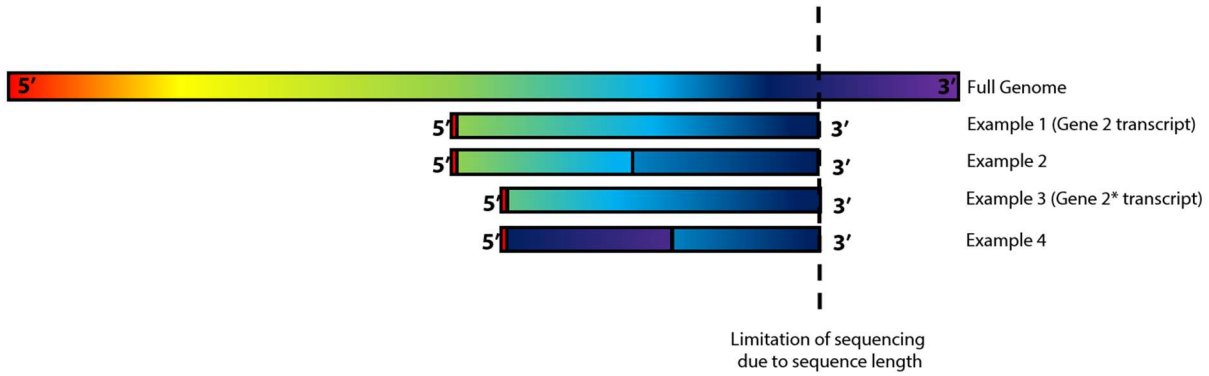


Figure 6.14 Schematic representation of PCR products generated by leader-body PCR over gene 2*. *The figure relates to the sequencing information also presented in Table 6.1 (i.e. examples 1-4). Here, the virus genome is presented in a rainbow gradient coloured rectangle. The origin of sequence for each transcript is depicted as matching the colour of this rectangle. All transcripts here have the 5' leader sequence depicted by a red cap and occur in the 5' to 3' direction. Examples 1 and 3 are the gene 2 and gene 2* transcripts respectively and, other than the start site of gene 2*, are typical of all known sg mRNAs in that they possess a continuous 5' to 3' sequence. Example 2 starts at the same position as gene 2, but skips part of the genome (transcript) sequence and resumes later in the gene. Example 4 possesses material from another gene at its 5' end then resumes from transcription from a further upstream part of the virus genome. The schematic is not drawn to scale. Due to sequencing limitations the 3' of all transcripts has not been resolved.*

6.4. Discussion

Following the observation in the previous chapters that egg attenuated isolates of M41 undergo a notable level of change within the 3′ UTR during serial passaging and that variation between different IBV strains exists within this region of the genome, a virus was conceptualised that exchanged the 3′ UTR of a virulent IBV (M41-K) with that of an apathogenic IBV (Beau-R). The rIBV, M41K-BeauR-3′UTR, was successfully rescued (Figure 6.5) and subsequently characterised by *in vitro* methods (Figure 6.6) and shown to possess a growth profile at both 37°C and 41°C on CKCs similar to M41-CK. Unexpectedly, given the nature of the modification made and *in vitro* characterisation data, M41K-BeauR-3′UTR was shown to possess a phenotype comparable to Beau-R *in vivo* (Figure 6.7) and is hence deemed attenuated.

Previous work by in the group had identified a potential transcript relating to the 3′ of the virus genome (Figure 6.8A). Evidence that this transcript was present in Beau-CK but not M41-CK was of immediate interest given generation of the virus M41K-BeauR-3′UTR. Northern blots of Beau-CK infected whole cell lysate confirmed the presence of an RNA between gene 6 (Nucleocapsid) and the 3′ UTR (Figure 6.8A). Further leader-body PCRs (Figure 6.8B) and PCR cloning also confirmed the result (Figure 6.8C). A deletion in this region of the M41-CK genome abolishes a prospective protein compared to other strains of IBV. This prospective protein is severely truncated in Beau-CK compared to other strains of IBV (Figure 6.11). Expression of the protein in Beau-CK is theorised due to the detection of ribosome bound to the transcript (Dinan et al., 2019).

Leader-body PCR analysis of M41K-BeauR-3′UTR infected whole cell lysate confirmed transcription of gene 7 by the rIBV (Figure 6.10). With the M41K-BeauR-3′UTR being attenuated *in vivo*, it is a plausible conjecture that expression of this protein results in attenuation. This is not supported however as virulent and

attenuated strains of IBV alike have equal potential to express the protein. Moreover, expression of ORF7 in M41K-BeauR-3′UTR would be its severely truncated, 10 AA form.

Transcription of an additional gene between the end of N protein and 3′ UTR was unanticipated as it was believed the surrounding regulatory sequences would prevent transcription. M41K-BeauR-3′UTR was designed to utilise the 3′ sequence of Beau-R immediately after the stop codon of the N protein (Figure 6.1) raising an interesting question with regards to M41-K transcriptional potential. As discussed previously, discontinuous RNA synthesis relies on the CS presented within a TRS. Not only does gene 7 utilise a non-canonical CS, the rIBV would have the M41-K TRS machinery at its 5′ end but possess the Beau-R TRS machinery at its 3′ end. The synthesis of gene 7 transcript in M41K-BeauR-3′UTR therefore indicates that either the 5′ TRS is not required for gene 7 transcription or that M41-K already possesses the machinery required for transcription at this site. Perhaps lending to the latter of these two ideas is the observation that some sequencing reads possess the leader-sequence at the 5′ end and align to this portion of the virus genome at 3′ in M41-CK (Figure 6.11). Furthermore, studies involving TGEV have demonstrated the importance of the TRS upstream of each gene signal in the template switch from body to leader sequence (Zuniga et al., 2004). Nonetheless, the generation of M41K-BeauR-3′UTR raises an exciting prospect of a site of tolerated insertion into M41-K. The insertion of for example mCherry, GFP or luciferase at this location would generate a useful molecular tool for further studies characterising IBV infection. More ambitiously it may be possible to insert antigens of other IBV strains or indeed other avian viruses for the purposes of dual-vaccination. In house efforts attempting to insert eGFP here have as yet been unsuccessful and future efforts will require careful design to prevent disruption of neighbouring secondary RNA structures involve in the process of template switching.

As an additional means of verifying ORF7 and to assess the possibility of unknown IBV transcripts, the transcriptional profiles of Beau-CK (the parent of Beau-R) and M41-CK (the parent of M41-K) were generated. Utilising high-throughput sequencing data, leader-possessing reads and chimeric reads were identified and used as indicators of virus-encoded transcripts (Figure 6.11). This method was validated by the detection of known transcripts relating to the structural and accessory genes (genes 2-6).

Transcript-like reads were identified mapping to regions of the genome not related to known genes with the potential transcript profile of Beau-CK and M41-CK being different. Conservatively, Beau-CK has two prospective transcripts notably absent in M41-CK, one at approximately 21.5 kb and another at the 3′ end of the genome. The transcript at approximately 21.5 kb (designated 2*) has previously been identified in IBV as encoding a truncated form of spike, missing the N terminus of the protein but otherwise remaining in frame through use of an alternate start codon (Bentley et al., 2013). This is not unique to IBV and has been reported in other coronaviruses (Hussain et al., 2005). The ORF 7 transcript however has only been described for IBV in the latest publications (An et al., 2019, Dinan et al., 2019) however other coronaviruses do have a gene in this position. The identification of gene 2* and gene 7 was made difficult due to use of a noncanonical CS and, for gene 7, its proximity to the N stop codon. Further validation would be required for confirmation of sg mRNAs from other peaks detectable in these traces, namely those spanning the replicase portion of the gene. Targeted leader-body PCRs of infected whole-cell lysate to confirm the existence of transcripts corresponding to these peaks and subsequent PCR cloning does confirm the presence of transcript reads spanning these regions, albeit at exceptionally low abundance (Figure 6.12). The low abundance of these prospective transcripts could be an indication of low rate of recombination. Conversely, these transcripts may appear in low abundance as they are actively

degraded by the host cell and are not protected by bound ribosomes. If the latter is true, it may suggest that these proteins are not actively translated.

A proportion of reads within the Illumina HiSeq data were identified as being chimeric but found not to be in possession of a 5′ leader sequence. This is indicated by a disparity between the two coverage traces (Figure 6.11); if all chimeric reads contained a leader sequence to two traces the traces would be fully overlapping. It is possible that the reads possess a 5′ sequence different to the search sequence used (by SNP) during the leader sequence screen which in turn may be the result of error in library preparation. Alternatively, IBV is known to utilise recombination as a means of generating genetic diversity (Lee and Jackwood, 2000, Cavanagh et al., 1992) and chimeric traces (and coverage peaks within) may therefore represent sites where recombination can occur. The involvement of transcript-required recombination in genetic recombination has not been explored.

Conversely, the data presented here (Figure 6.11) suggests the possibility that IBV transcripts themselves can be multi-chimeric with sequence derived from more than two non-consecutive locations as well as including leader sequence. It may therefore be possible that these have been captured during poly(A) enrichment but read length does not facilitate sequencing over leader sequence. Curiously as revealed by PCR cloning and sequencing, the sequences are not in what might be considered the conventional gene order (for example 5′-Leader-Gene 2-Gene 5-3′) but rather occur reassorted (5′-Leader-Gene 5-Gene 2- 3′). Further experimental effort is however required to resolve the complete 3′ end of the transcripts but may coincide with the observation of non-leader instances of discontinuous RNA synthesis occurring at sites of recombination not defined by a CS. Although this is against current knowledge, this may indicate the potential for non-leader sequence driven instances of discontinuous RNA synthesis. Further research would be required to investigate whether these unconventional transcripts are functional or by-products of the replication/transcription process.

For the purposes of transcript identification, virus-infected allantoic fluid is perhaps not the best material as virus transcription *in ovo* is not well, if at all, characterised. Indeed, the detection of IBV transcripts in allantoic fluid was perhaps unanticipated given the widely held belief that IBV virions do not carry within them viral sg mRNAs from their originating host cell (Macnaughton and Madge, 1977). The possibility of sg mRNAs in IBV virions has previously been reported (Zhao et al., 1993), but was not well received. *In ovo*, IBV replicates both within the chick embryo and the surrounding chorioallantoic membranes (Abdel-Moneim et al., 2009). At the point of harvest (24 hours post-inoculation), it is believed IBV primarily replicates in the allantoic membrane. The identification of chicken genome related sequences in allantoic fluid may suggest cellular degradation of virus infected cells. The breakdown of host cells would therefore release cellular DNA and viral RNAs into the allantoic fluid. Somewhat unfortunately, chicken genome retains a noteworthy proportion of the available reads lowering overall IBV-derived reads despite virus purification by ultracentrifugation. As a result of the above, the data presented here may therefore be considered an underestimate of the full transcriptome possessed by IBV.

M41K-BeauR-3′UTR was originally designed predominantly based on the information that M41-CK attenuated by serial egg passaging undergoes a high level of mutation over the 3′ UTR. While the virus was intended to study the contribution of this region to pathogenicity, it was not imagined that such a change would reduce clinical signs to the point of attenuation. Indeed, none of the mutations observed in the egg-passaging analysis correspond to the insertion of gene 7 into the virus genome. The discovery that exchanging the 3′ region in a pathogenic virus for that of a non-pathogenic virus sadly therefore does not answer why the serially passaged M41 isolates are attenuated with further work required to identify the cause of attenuation in these viruses. Indeed it is likely that the

mechanism of attenuation of the serially passaged M41 viruses and M41K-BeauR-3′UTR are not related.

Despite not here being able to characterise serial egg-passaged M41 attenuation, M41K-BeauR-3′UTR is an attenuated virus and demonstrates the potential that modifications to this end of the virus genome can affect virus pathogenicity. While future work would be required to ascertain both the translation status of ORF7 and the immunogenic potential of M41K-BeauR-3′UTR, there is an exciting prospect that viruses possessing modifications outside of coding regions may be useful as IBV vaccine candidates. In order to facilitate this further questions must be asked of M41K-BeauR-3′UTR. It must firstly be established whether the rIBV provokes an equivalent immune response *in vivo* compared to M41-K. Given the nature of the rIBV modification made, there is a high possibility that the rate at which defective interfering (DIs) particles are formed has been altered. DIs are generated by predominately by erroneous recombination; a process which a virus with a modified UTR would likely compromise. DIs have been implicated both in the activation of immune response (Sun et al., 2015) and the regulation of virus pathogenicity (Lukhovitskaya et al., 2013). Additionally coronaviruses are known to possess elements specifically related to regulation of DI replication (Raman et al., 2003). It is therefore easy to conjecture DIs influence are a means of attenuating IBV and should be explored further.

Viral stability and the virus's capability to tolerate the modification over sustained passages *in vivo* and *in vitro* must also be assessed both to ascertain the viability of the modification in IBV vaccines and potential reversion to wildtype (virulence). Additionally, while the growth profile over a period of 96 hours has been assessed (Figure 6.6), characterisation over one replication cycle will establish whether the virus's ability to initiate infection has been impaired.

To fully explore the mechanism of M41K-BeauR-3′UTR attenuation, it is necessary that further rIBVs are made. These would firstly need to explore as to whether attenuation here is linked to the 3′ UTR or whether it would be possible to instead exchange the 5′ UTR of a virulent virus with that of an attenuated virus. These viruses would answer whether the 3′ UTR itself is involved in attenuation or if it is a mismatch in the origin of UTR is involved instead. It must also be explored as to whether attenuation is a feature unique to the 3′UTR of Beau-R or the whether 3′ UTR of other IBVs may be incorporated to illict the same effect.

The successful rescue of M41K-BeauR-3′UTR has opened a floodgate of potential research avenues and is therefore a prototype virus both for new era of IBV vaccines and for the further research of discontinuous RNA transcription.

Chapter 7 – General Discussion and Future Work

7.1. Discussion

The process of serial egg passaging is a well-established method for generating attenuated IBVs with little to no understanding as to the mechanisms behind attenuation. The primary aim of this project was therefore to couple HTS with egg attenuated IBV to for the purposes of identifying the key genetic markers of attenuation and enable rational design of future vaccines against IBV. This aim was explored by utilising a panel of egg-passaged isolates of IBV, four started from a virulent IBV population (M41-CK) and four started from the virulent clone of M41-CK (M41-K). High-throughput sequencing was then performed on the initial virus, intermediate viruses, and final attenuated viruses to inform both as to the changes that occur in attenuated versus virulent IBV and to describe the order of mutational appearance.

The propagation of IBV in embryonated eggs is unnatural (IBV does not typically have the opportunity to infect embryonated eggs) and would require the virus to rapidly adapt to this new environment. The forced adaptation of viruses to a new environment, typically another host, has long been used for the generation of attenuated viruses for vaccine production originating back to Louis Pasteur's work with rabies (Smith, 2012). Here it is believed that as the virus adapts to the new species, it becomes less suited to the original host, hence reducing virulence. The absence of a full immune response in the allantoic fluid of embryonated hens' eggs simulates this while maintaining IBV immunogenicity though it should be noted that is immunogenicity wanes in repeated passaging. Serial egg passaging is a well-established protocol for the development of attenuated IBVs however given the complex structure of RNA populations, there is some debate as to the origin

of attenuating mutations. As virulent heterogeneous IBV populations are used for initiation of passaging, it could be argued that attenuated viruses co-exist in the virulent population and the process of serial egg passaging drives selection of these viruses. Conversely, it could be that the prolonged passaging results in the accumulation of spontaneous mutations throughout the genome, a proportion of which confer attenuation. A low overall number of variants shared between virulent and attenuated IBVs, as demonstrated in the body of this thesis, would in this instance lend evidence to the latter of these two arguments i.e. mutation. That is not however saying that examples of selection do not exist within that IBV. To the contrary, the T24293C mutation for example, present in the virulent M41-CK population and reaching fixation in the final passages is a clear example of selection, however this mutation is not associated with an attenuated genotype. Further evidence includes specific mutations that are not necessarily shared between all four attenuated viruses for both passaging of M41-CK and M41-K. Future work may therefore require passaging of other virulent strains of IBV to verify this conclusion.

Attenuation of IBV is an undoubtedly complex process with pathogenicity of the virus likely governed by multiple intersecting components. While serial egg passaging does result in an attenuated IBV, this method does not result in a consistent virus, nor indeed is the process faultless. Even utilising the same starting material, it is possible to genetically differently attenuate IBV with minimal overlap between these viruses at the consensus and subconsensus level. The absence of a singular mutation corresponding to attenuation is not unexpected; if attenuation was caused by a singular mutation, there would not be evidence implicating multiple genes in IBV attenuation (Phillips et al., 2012, Armesto et al., 2009, Wickramasinghe et al., 2011, van Beurden et al., 2018, Zhao et al., 2019). Not only is there an absence of a singular attenuating mutation, there is an overall lack of similarity between attenuated viruses. Collectively, this may suggest that there are multiple “areas” in sequence space which confer

attenuation with serial egg passaging dictating the direction of evolution not the path. A caveat must be added here however that the mutations observed may be linked by a means that our current understanding of IBV has not yet established.

There are arguments as to why multiple "areas" of attenuation in sequence space may be of both benefit and detriment to the IBV field. While presumably there are a finite number of vaccine-effective IBVs that can be generated from the passaging of a single virulent IBV, the ability to differently attenuate the virus may be seen as encouraging for the future use of this protocol. This ability is however a double-edged sword as it does not speak for already generated vaccines and the possible genetic differences between batches marketed against the same virulent strain. Additionally, the prevalence of potentially virulent genotypes to remaining after egg-passaging would suggest that egg-attenuated IBV has the potential to selectively revert to virulence by selection (as opposed to back-mutation) to the originally virulent form. If different batches of the same IBV vaccine possess highly diverse and dissimilar populations, each vaccine could have a different effectiveness and evolutionary consequence when introduced into the field and in competition with other IBV strains. There is an argument for the genetic structures of live attenuated IBV vaccine viruses to be fully characterised by deep-sequencing if for no other reason, the purposes of quality control.

Perhaps unsurprisingly given that there appears to be multiple "areas" in sequence space responsible for attenuation, the pathway towards the attenuation stage are equally diverse. Intersections between intermediate passages of the same lineage demonstrate a transient nature of variants with few variants occurring in multiple consecutive passages. This would suggest the virus may temporarily acquire or tolerant a variant, with this mutation being advantageous towards the beginning of egg passaging but possibly less advantageous or out-competed towards the end of passaging. There is perhaps an interesting question as to reversion of vaccine viruses to virulence, as if the viruses can utilise multiple pathways to reach

attenuation they can presumably utilise different pathways to return to a virulent form. This would understandably make studying reversion to virulence particularly challenging, as would generating vaccines to minimise the occurrence of reversion.

The process of serially passaging a virus is known to generate random mutations in the virus genome with it believed that some of these mutations correspond to attenuation. However, perhaps one of the reasons that so few mutations are shared across attenuated IBV replicates is that, to a certain extent, it does not matter what these mutations are. Instead perhaps the virus must be forced into an evolutionary sequence space which the virus cannot adapt out of before being cleared by the host's immune system. The absence of immune pressure in embryonated eggs permits IBV to evolve into this sequence space which IBV could not normally do in the field. There could therefore be a critical number of mutations needed to generate an attenuated IBV.

Perhaps lending to this notion is that there is a steady accumulation of variants, both at consensus and sub-consensus level as passaging of IBV continues. To explore this possibility in that there are a set number of mutations required for attenuation, the pathogenicity of each intermediate passage here would firstly need to be characterised which itself has its own challenges. With no suitable alternative, characterisation of intermediate viruses would require animal study. Adding further complexity to this issue as to which and how many intermediates IBVs would need to be characterised, there would appear to a lot of variability between different IBVs exemplified by the existence of the H120 and H52 vaccine strains (Bijlenga et al., 2004) . Here, H120 and H52 were generated by egg passage 120 and 52 times, respectively. In the generation of a vaccine against QX (Geerligs et al., 2011), intermediate passages were characterised *in vivo* and deemed that 80 passages were required to sufficiently attenuated the virus and comply with safety requirements. Geerligs *et al*. also noted that ciliostasis, a marker for

pathogenicity, increased temporarily over the course of passaging, but was ultimately lowered in the final virus.

Despite the complexities surrounding attenuation, the contents of this thesis and associated data has characterised the genetic changes that occur as a result of attenuation by egg passaging and provided evidence as to the evolutionary pathways leading to attenuation, thereby achieving the primary objectives of this project. In the long term, this data will act as a compendium for both the future study of IBV attenuation and aid the development of rationally attenuated IBV vaccines. Epitomising the use of this data in exploring the mechanism of IBV attenuation, a recombinant IBV that possesses the backbone of a virulent IBV with the 3′ UTR of an avirulent IBV was conceptualised and proven to be non-pathogenic. While there are future questions to ask as to why M41K-BeauR-3′UTR is attenuated, it nonetheless poses an interesting research question and is a potential prototype for a new era of IBV vaccines.

7.2. Limitations and caveats

While the main objectives of this thesis have been completed, there are some limitations or caveats as described below.

7.2.1. Further advances in HTS and the repetition of analysis

Ongoing developments and investments into HTS will result in a reduced rate of error and higher depth of coverage, ultimately allowing the detection of lower frequency variants with greater confidence. While this will ultimately benefit the characterisation of attenuated viruses, the perpetual increase in sequencing confidence will cause a perpetual change in the reported influence of both selection and mutation in the evolution of IBV. An increased sequencing confidence will identify more variants in a sample therefore more strongly indicating the influence of *de novo* mutation in IBV evolution. Conversely, identifying more variants could allow a greater possibility for variants to be shared, more strongly indicating the influence of selection in IBV evolution.

Any future technologies that result in or accentuate an imbalance between selection and mutation could yield a possibly very different, conflicting result as to that reported here. In the case of IBV, if the same passages were sequenced again using two technologies, though the general trend may be the same, the finer details may change. Any conclusions drawn on the analysis may therefore be dependent on what was exactly sequenced, and technologies involved.

7.2.2. Detection of transcripts within the dataset

One of the unexpected advantages of the dataset highlights a select disadvantage. It was not originally intended to use the data generated in this thesis for the purposes of studying potential IBV transcripts (Chapter 6). The detection of IBV transcripts in allanotic fluid was unanticipated due to the widely held belief that virions do not possess sg mRNA. It is also not expected that the process of purifying IBV-infected allanotic fluid allows for the carryover of host material. The detection of sg mRNAs in virus infected allantoic fluid, which all things considered appear to represent a significant proportion of available reads, does highlight a potential caveat in the data in that it is unknown as to whether the variants as reported occur at a transcript or genome level.

Assuming that sg mRNAs are not included within IBV virions, it would mean that data presented in this thesis (all Illumina-obtained data) is a reflection of variants occurring in both the virus genome and virus transcripts. The length of reads in the provided dataset would not allow for the determination of whether a variant occurred in the genome or in a transcript. Given that the exact site of template switching between body and leader in sg mRNA synthesis is known to be variable both between genes and within the same gene, there is an immediate risk of introducing variants that would be identified at a genomic level but in reality only occur in transcripts. This is not accounting for any potential mutations that introduced by the RTC in transcription or replication. While mutations in transcripts would have an effect in the passage in which they occur, virus transcripts would have no direct bearing on the evolution of the virus as they would not be carried forward in the passaging.

The alternative is that sg mRNAs are included within IBV virions and hence mutations occurring in both the genomic and generated transcripts would be of relevance to the evolution of IBV. This would go against our current understanding of IBV replication but would raise some interesting possibilities for future study.

The enrichment of poly(A) containing sequences is likely the reason for isolating virus transcripts that would also possess a poly(A) tail. This approach was required to provide full genome coverage of IBV as alternative approaches were unable to provide full-length coverage and/or required amplification. Again, in development of the sequencing protocol it was not believed transcripts would be present, so poly(A) enrichment would only pull up full-length IBV sequences which also possess a poly (A) tail. It is therefore unfortunate circumstance that sg mRNAs were also isolated and sequenced, with the findings of the thesis potentially having this caveat as described above.

7.2.3. Passaging past the point of attenuation and the generation of genetic "noise"

As detailed in Chapter 4, one of the contributing factors confounding the determination of IBV attenuation is that it is not known as to when during the course of passaging IBVs would be deemed attenuated. Moreover it is not known, as to whether all the IBVs are attenuated from the same passage (although this is admittedly unlikely). Although the final passage of each virus is confirmed to be attenuated by *in vivo* characterisation, in actual fact it could be attenuated much earlier in the series of passages. To address this issue, Chapter 4 attempted to use temperature sensitivity as an indicator of virus pathogenicity, though this met with little success. Currently, the only alternative would be to characterise each intermediate virus in an animal study. Given the number of viruses involved and number of birds required per virus, such a course of experiments would not be ethically nor logistically viable.

A consequence of not knowing the exact point of attenuation is that variation occurring after attenuation has been included in the analysis. This therefore results in genetic "noise" that distracts from mutations involved in the mechanism of attenuation. Passaging M41-K was in some ways meant to resolve this issue,

however as it is not currently known whether attenuation is cause or consequence of adaption to growth in eggs, it is unclear what mutations arising in M41-K are associated with. Indeed, there is very little overlap between mutations arising in M41-CK and M41-K so it does not fully address this issue.

7.2.4. Recombination and its role in the evolution of IBV and attenuation

Recombination between IBV is known to occur in the field (Lee and Jackwood, 2000, Cavanagh et al., 1992) with more likely to occur in the regions encoding nsps 2, 3 and 16 (Thor et al., 2011). Recombination is likely to be involved in IBV evolution and hence the process of attenuation. While recombination in terms of transcription has been described in this thesis (Chapter 6), the concept of recombination in terms of attenuation has not been explored here. This is because as serial passaging of IBV is performed in a closed system (i.e. in embryonated eggs), there is no opportunity for the virus to recombine with other circulating IBVs. The virus would therefore only have the possibility to recombine with itself.

To explore the role of recombination in attenuation, it may prove necessary to co-infect embryonated eggs with two IBVs. This could involve use of either apathogenic IBV, pathogenic IBVs or a mixture of both; all scenarios would yield potentially interesting results. This system would be artificial but may elude as to the mechanism of genetic recombination as opposed to transcript recombination if indeed the two systems are dissociated. Experiments involving IBV co-infection of embryonated eggs with M41 (pathogenic) and Beaudette (apathogenic) strains of IBV have confirmed recombination between the two viruses (Kottier et al., 1995)

7.3. Future Work

While the main objectives of this project have been completed, it has highlighted a number of exciting areas for further research which will be discussed below.

7.3.1. Long read sequencing for the study of viral haplotypes

While not described in the main body of text, reconstruction of virus haplotypes by bioinformatics approaches was attempted but met little success. Here, the intention was to identify whether any mutations observed occur in concurrence with other mutations. Without doing so, mutations occurring at a distance longer than the read length could be missed. An inability to link mutations together is therefore a limitation of the project. Haplotype reconstruction was initially performed on 454 reads however limited coverage resulted in segmentation of the virus genome, obsoleting the purpose of performing the reconstruction. Illumina-produced reads had almost the exact opposite problem in that too higher coverage and resulting data complexity lead to long computation times. While random read sub sectioning could overcome this problem, poor reconstruction as a result of smaller Illumina read length could not be resolved.

The application of longer read sequencing (e.g. Oxford Nanopore or PacBio) was considered to use in this project for the purposes of studying viral haplotypes. Current protocols however require an input that was not obtainable from the intermediate viral passages and generating separate virus stocks of the starting and final viruses would be to subject to the same limitations of the 454 analysis. Additionally, the associated error rate of these platforms makes accurate determination of low frequency variants impractical. Theoretically it would be possible to combine both short and long reads for the study of viral populations and virus haplotypes. The merging of the two datasets is common practice when resolving the sequence of highly repetitive regions of genomes (Daccord et al., 2017, Larsen et al., 2017, De Maio et al., 2019), but there are limited examples a

hybrid analysis in relation to a virus. In any event, upcoming future advances and investment in long read technologies will solve the issues of input and error, making a multifaceted approach ultimately unnecessary.

7.3.2. Long read sequencing and determination of IBV transcripts and recombination

Though longer-read sequencers are currently renowned for their high error rate, there is a potential possibility of using such platforms to study IBV transcripts and sites of recombination. Current long-read technologies have an average read length of roughly 10 kB representing approximately a third of the IBV genome and encompassing the entire length of all known IBV transcripts. As the aim is not to focus on variation at individual positions but rather ascertain whether two non-consecutive regions of the virus genome are connected, single-nucleotide accuracy is not necessarily required and effect of error would be diminished by repeated coverage.

The development of the IBV Illumina sequencing protocols and the detection of IBV transcripts in Chapter 6 somewhat demonstrates proof of concept. Here, using an NEBNext® Poly(A) mRNA Magnetic Isolation Module it was possible to isolate IBV transcripts. Follow up experiments would therefore involve infection of cells with IBV for the purpose of isolating mRNAs from whole-cell lysate. The limitations however revolve around isolation of insufficient IBV derived mRNA. This could be overcome by a small amplification step with error rate of PCR similarly diminished by coverage. The potentially identified transcripts would need to be verified by other approaches (such as leader-body PCR or Northern blot) and the translational status of identified genes determined. Such studies would generate a roadmap of areas of interest and inform as to a complete repertoire of coding and non-coding IBV RNAs.

Although the protocol is still in its initial phases of development, direct RNA sequencing has been used to for the study of both viral RNA haplotypes (7.3.1) and structural variants (7.3.2) in human coronavirus (Viehweger et al., 2019). The longest read produced (~26 kb) demonstrates the feasibility of using the approach for IBV (~27.6kb).

7.3.3. Application of single-cell RNA sequencing for determination of per replication cycle virus error

Single cell sequencing is an exciting technology that facilitates generation of DNA and RNA information from individual cells. The technology provides high resolution information allowing for accurate determination of individual cell function under certain conditions. More recently, applications include use of studying virus-infected cells, obtaining information on the host transcriptome in the context of intracellular viral RNA (Zanini et al., 2018a, Zanini et al., 2018b). These studies have additionally been shown to provide information on virus population genomics.

An application could be considered to determine the level of error introduced per replication cycle of IBV. Here, cells could be infected with clonal IBV and sequenced after a known number of replication cycles (which last between 6 – 8 hours (Maier et al., 2013)). Not only would this provide information as to the changes in the host transcriptome over time, it would also provide information as to the per replication cycle IBV error rate. If the basal error rate of replication is known, this could subsequently be used to study and inform as to the rate of error per egg passage. The aim here would be then to determine the regions of the genome that undergo a mutation at an abnormal rate compared to generic replication error. This could be used to highlight regions of the genome that are both tolerant and intolerant to mutation; both pieces of information that assist with studying attenuation and IBV vaccine design.

The obstacle here would be generating a truly clonal of IBV as, as demonstrated in this thesis, the process of rescuing the virus and generating a working stock introduces diversity to a point where the virus is arguably non-clonal. This would be needed to ensure that any mutations recorded occur from a known starting point. Additionally, precautions must be taken to study the rate of error introduced by the sequencing system.

List of references

2012. Coronaviridae. *In:* KING, A. M. Q., ADAMS, M. J., CARSTENS, E. B. & LEFKOWITZ, E. J. (eds.) *Virus Taxonomy*. San Diego: Elsevier.

ABDEL-MONEIM, A. S., ZLOTOWSKI, P., VEITS, J., KEIL, G. M. & TEIFKE, J. P. 2009. Immunohistochemistry for detection of avian infectious bronchitis virus strain M41 in the proventriculus and nervous system of experimentally infected chicken embryos. *Virology Journal,* 6**,** 15.

ABOLNIK, C. 2015. Genomic and single nucleotide polymorphism analysis of infectious bronchitis coronavirus. *Infect Genet Evol,* 32**,** 416-24.

ALMAZAN, F., DEDIEGO, M. L., SOLA, I., ZUNIGA, S., NIETO-TORRES, J. L., MARQUEZ-JURADO, S., ANDRES, G. & ENJUANES, L. 2013. Engineering a replication-competent, propagation-defective Middle East respiratory syndrome coronavirus as a vaccine candidate. *MBio,* 4**,** e00650-13.

AMMAYAPPAN, A., UPADHYAY, C., GELB, J., JR. & VAKHARIA, V. N. 2009. Identification of sequence changes responsible for the attenuation of avian infectious bronchitis virus strain Arkansas DPI. *Arch Virol,* 154**,** 495-9.

AN, H., CAI, Z., YANG, Y., WANG, Z., LIU, D. X. & FANG, S. 2019. Identification and formation mechanism of a novel noncoding RNA produced by avian infectious bronchitis virus. *Virology,* 528**,** 176-180.

ANDREWS, S. 2010. *FastQC - A quality control tool for high throughput sequence data*. [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed].

ANONYMOUS 2014. Veterinary/Animal Vaccines Market by Type (Porcine, Poultry, Companion Animal, Aquaculture), Disease (Porcine Parvovirus, Swine Pneumonia, Avian Influenza, Rabies, Clostridial Diseases), Technology (Inactivated, Toxoid, Recombinant) - Global Forecast to 2018.

ARMESTO, M., CAVANAGH, D. & BRITTON, P. 2009. The replicase gene of avian coronavirus infectious bronchitis virus is a determinant of pathogenicity. *PLoS One,* 4**,** e7384.

BADGETT, M. R., AUER, A., CARMICHAEL, L. E., PARRISH, C. R. & BULL, J. J. 2002. Evolutionary Dynamics of Viral Attenuation. *Journal of Virology,* 76**,** 10524-10529.

BAKER, S. C., YOKOMORI, K., DONG, S., CARLISLE, R., GORBALENYA, A. E., KOONIN, E. V. & LAI, M. M. 1993. Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. *J Virol,* 67**,** 6056-63.

BANDE, F., ARSHAD, S. S., BEJO, M. H., MOEINI, H. & OMAR, A. R. 2015. Progress and challenges toward the development of vaccines against avian infectious bronchitis. *J Immunol Res,* 2015**,** 424860.

BANKEVICH, A., NURK, S., ANTIPOV, D., GUREVICH, A. A., DVORKIN, M., KULIKOV, A. S., LESIN, V. M., NIKOLENKO, S. I., PHAM, S., PRJIBELSKI, A. D., PYSHKIN, A. V., SIROTKIN, A. V., VYAHHI, N., TESLER, G., ALEKSEYEV, M. A. & PEVZNER, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol,* 19**,** 455-77.

BARANOV, P. V., HENDERSON, C. M., ANDERSON, C. B., GESTELAND, R. F., ATKINS, J. F. & HOWARD, M. T. 2005. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology,* 332**,** 498-510.

BEACH, J. R. & SCHALM, O. W. 1936. A filterable virus distinct from that of laryngotracheitis: the cause of respiratory disease in chicks. *Poultry Science,* 15**,** 199-206.

BEARE, A. S., BYNOE, M. L. & TYRRELL, D. A. 1968. Investigation into the attenuation of influenza viruses by serial passage. *Br Med J,* 4**,** 482-4.

BEAUDETTE, F. R. & HUDSON, C. B. 1937. Cultivation of the virus infectious bronchitis. *Journal of the American Veterinary Medical Association,* 90**,** 51-60.

BECARES, M., PASCUAL-IGLESIAS, A., NOGALES, A., SOLA, I., ENJUANES, L. & ZUNIGA, S. 2016. Mutagenesis of Coronavirus nsp14 Reveals Its Potential Role in Modulation of the Innate Immune Response. *J Virol,* 90**,** 5399-5414.

BELOUZARD, S., CHU, V. C. & WHITTAKER, G. R. 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci U S A,* 106**,** 5871-6.

BELOUZARD, S., MILLET, J. K., LICITRA, B. N. & WHITTAKER, G. R. 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses,* 4**,** 1011-33.

BENNETT, R. M., CHRISTIANSEN, K. & CLIFTON-HADLEY, R. S. 1999. Direct costs of endemic diseases of farm animals in Great Britain. *Vet Rec,* 145**,** 376-7.

BENTLEY, K., KEEP, S. M., ARMESTO, M. & BRITTON, P. 2013. Identification of a noncanonically transcribed subgenomic mRNA of infectious bronchitis virus and other gammacoronaviruses. *J Virol,* 87**,** 2128-36.

BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P. E. 2000. The Protein Data Bank. *Nucleic Acids Research,* 28**,** 235-242.

BHARDWAJ, K., SUN, J., HOLZENBURG, A., GUARINO, L. A. & KAO, C. C. 2006. RNA recognition and cleavage by the SARS coronavirus endoribonuclease. *J Mol Biol,* 361**,** 243-56.

BICKERTON, E., KEEP, S. M. & BRITTON, P. 2017. Reverse Genetics System for the Avian Coronavirus Infectious Bronchitis Virus. *Methods Mol Biol,* 1602**,** 83-102.

BIJLENGA, G., COOK, J. K. A., GELB, J. J. & WIT, J. J. D. 2004. Development and use of the H strain of avian infectious bronchitis virus from the Netherlands as a vaccine: a review. *Avian Pathology,* 33**,** 550-557.

BOSCARINO, J. A., LOGAN, H. L., LACNY, J. J. & GALLAGHER, T. M. 2008. Envelope protein palmitoylations are crucial for murine coronavirus assembly. *Journal of virology,* 82**,** 2989-2999.

BOSCH, B. J., VAN DER ZEE, R., DE HAAN, C. A. M. & ROTTIER, P. J. M. 2003. The coronavirus spike protein is a class I virus fusion protein: Structural and functional characterization of the fusion core complex. *Journal of Virology,* 77**,** 8801-8811.

BRAYTON, P. R., LAI, M. M., PATTON, C. D. & STOHLMAN, S. A. 1982. Characterization of two RNA polymerase activities induced by mouse hepatitis virus. *Journal of virology,* 42**,** 847-853.

BREDENBEEK, P. J., PACHUK, C. J., NOTEN, A. F., CHARITE, J., LUYTJES, W., WEISS, S. R. & SPAAN, W. J. 1990. The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res,* 18**,** 1825-32.

BRIERLEY, I., BOURSNELL, M. E., BINNS, M. M., BILIMORIA, B., BLOK, V. C., BROWN, T. D. & INGLIS, S. C. 1987. An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J,* 6**,** 3779-85.

BRIERLEY, I., DIGARD, P. & INGLIS, S. C. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell,* 57**,** 537-47.

BRIERLEY, I., JENNER, A. J. & INGLIS, S. C. 1992. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol,* 227**,** 463-79.

BRITTON, P. & CAVANAGH, D. 2017. Avian Coronavirus Diseases and Infectious Bronchitis Vaccine Development. *In:* THIEL, V. (ed.) *Coronaviruses: Molecular and Cellular Biology*. Caister Academic Press.

BRITTON, P., GREEN, P., KOTTIER, S., MAWDITT, K. L., PENZES, Z., CAVANAGH, D. & SKINNER, M. A. 1996. Expression of bacteriophage T7 RNA polymerase in avian and mammalian cells by a recombinant fowlpox virus. *J Gen Virol,* 77 (Pt 5)**,** 963-7.

BROWN, C. G., NIXON, K. S., SENANAYAKE, S. D. & BRIAN, D. A. 2007. An RNA stem-loop within the bovine coronavirus nsp1 coding region is a cis-acting element in defective interfering RNA replication. *J Virol,* 81**,** 7716-24.

BUKHARI, K., MULLEY, G., GULYAEVA, A. A., ZHAO, L., SHU, G., JIANG, J. & NEUMAN, B. W. 2018. Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abyssoviridae, and from a sister group to the Coronavirinae, the proposed genus Alphaletovirus. *Virology,* 524**,** 160-171.

BUSHNELL, B. Sponsor Org.: USDOE Office of Science (SC). BBMap: A Fast, Accurate, Splice-Aware Aligner. 2014-03-17 2014 United States. Research Org.: Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

CASAIS, R., THIEL, V., SIDDELL, S. G., CAVANAGH, D. & BRITTON, P. 2001. Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J Virol,* 75**,** 12359-69.

CAVANAGH, D. 2007. Coronavirus avian infectious bronchitis virus. *Vet Res,* 38**,** 281-97.

CAVANAGH, D., DAVIS, P. J. & COOK, J. K. 1992. Infectious bronchitis virus: evidence for recombination within the Massachusetts serotype. *Avian Pathol,* 21**,** 401-8.

CAVANAGH, D., DAVIS, P. J., PAPPIN, D. J., BINNS, M. M., BOURSNELL, M. E. & BROWN, T. D. 1986. Coronavirus IBV: partial amino terminal sequencing of spike polypeptide S2 identifies the sequence Arg-Arg-Phe-Arg-Arg at the cleavage site of the spike precursor propolypeptide of IBV strains Beaudette and M41. *Virus Res,* 4**,** 133-43.

CAVANAGH, D., ELUS, M. M. & COOK, J. K. A. 1997. Relationship between sequence variation in the S1 spike protein of infectious bronchitis virus and the extent of cross-protection in vivo. *Avian Pathology,* 26**,** 63-74.

CHEN, Y., CAI, H., PAN, J., XIANG, N., TIEN, P., AHOLA, T. & GUO, D. 2009. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc Natl Acad Sci U S A,* 106**,** 3484-9.

CHEN, Y., SU, C., KE, M., JIN, X., XU, L., ZHANG, Z., WU, A., SUN, Y., YANG, Z., TIEN, P., AHOLA, T., LIANG, Y., LIU, X. & GUO, D. 2011. Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS Pathog,* 7**,** e1002294.

CHENG, A., ZHANG, W., XIE, Y., JIANG, W., ARNOLD, E., SARAFIANOS, S. G. & DING, J. 2005. Expression, purification, and characterization of SARS coronavirus RNA polymerase. *Virology,* 335**,** 165-76.

COLOGNA, R. & HOGUE, B. G. 1998. Coronavirus nucleocapsid protein. RNA interactions. *Adv Exp Med Biol,* 440**,** 355-9.

CONWAY, J. R., LEX, A. & GEHLENBORG, N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics,* 33**,** 2938-2940.

COOK, J. K., JACKWOOD, M. & JONES, R. C. 2012. The long view: 40 years of infectious bronchitis research. *Avian Pathol,* 41**,** 239-50.

COOK, J. K. A., DARBYSHIRE, J. H. & PETERS, R. W. 1976. The use of chicken tracheal organ cultures for the isolation and assay of avian infectious bronchitis virus. *Archives of Virology,* 50**,** 109-118.

CORMAN, V. M., ECKERLE, I., BLEICKER, T., ZAKI, A., LANDT, O., ESCHBACH-BLUDAU, M., VAN BOHEEMEN, S., GOPAL, R., BALLHAUSE, M., BESTEBROER, T. M., MUTH, D., MÜLLER, M. A., DREXLER, J. F., ZAMBON, M., OSTERHAUS, A. D., FOUCHIER, R. M. & DROSTEN, C. 2012. Detection of a novel human coronavirus by real-time reverse-transcription polymerase chain reaction. *Euro Surveill,* 17**,** 20285.

CORNELISSEN, L. A., WIERDA, C. M., VAN DER MEER, F. J., HERREWEGH, A. A., HORZINEK, M. C., EGBERINK, H. F. & DE GROOT, R. J. 1997. Hemagglutinin-esterase, a novel structural protein of torovirus. *J Virol,* 71**,** 5277-86.

CORNILLEZ-TY, C. T., LIAO, L., YATES, J. R., 3RD, KUHN, P. & BUCHMEIER, M. J. 2009. Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *J Virol,* 83**,** 10314-8.

CORSE, E. & MACHAMER, C. E. 2000. Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. *J Virol,* 74**,** 4319-26.

COTTAM, E. M., WHELBAND, M. C. & WILEMAN, T. 2014. Coronavirus NSP6 restricts autophagosome expansion. *Autophagy,* 10**,** 1426-41.

DACCORD, N., CELTON, J.-M., LINSMITH, G., BECKER, C., CHOISNE, N., SCHIJLEN, E., VAN DE GEEST, H., BIANCO, L., MICHELETTI, D., VELASCO, R., DI PIERRO, E. A., GOUZY, J., REES, D. J. G., GUÉRIF, P., MURANTY, H., DUREL, C.-E., LAURENS, F., LESPINASSE, Y., GAILLARD, S., AUBOURG, S., QUESNEVILLE, H., WEIGEL, D., VAN DE WEG, E., TROGGIO, M. & BUCHER, E. 2017. High-

quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics,* 49**,** 1099.

DALTON, K., CASAIS, R., SHAW, K., STIRRUPS, K., EVANS, S., BRITTON, P., BROWN, T. D. & CAVANAGH, D. 2001. cis-acting sequences required for coronavirus infectious bronchitis virus defective-RNA replication and packaging. *J Virol,* 75**,** 125-33.

DARBYSHIRE, J. H., ROWELL, J. G., COOK, J. K. & PETERS, R. W. 1979. Taxonomic studies on strains of avian infectious bronchitis virus using neutralisation tests in tracheal organ cultures. *Arch Virol,* 61**,** 227-38.

DE GROOT, R. J. 2006. Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. *Glycoconj J,* 23**,** 59-72.

DE HAAN, C. A. M., VENNEMA, H. & ROTTIER, P. J. M. 2000. Assembly of the Coronavirus Envelope: Homotypic Interactions between the M Proteins. *Journal of Virology,* 74**,** 4967-4978.

DE MAIO, N., SHAW, L. P., HUBBARD, A., GEORGE, S., SANDERSON, N. D., SWANN, J., WICK, R., ABUOUN, M., STUBBERFIELD, E., HOOSDALLY, S. J., CROOK, D. W., PETO, T. E. A., SHEPPARD, A. E., BAILEY, M. J., READ, D. S., ANJUM, M. F., WALKER, A. S. & STOESSER, N. 2019. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microbial Genetics*.

DECROLY, E., IMBERT, I., COUTARD, B., BOUVET, M., SELISKO, B., ALVAREZ, K., GORBALENYA, A. E., SNIJDER, E. J. & CANARD, B. 2008. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J Virol,* 82**,** 8071-84.

DEDIEGO, M. L., ALVAREZ, E., ALMAZAN, F., REJAS, M. T., LAMIRANDE, E., ROBERTS, A., SHIEH, W. J., ZAKI, S. R., SUBBARAO, K. & ENJUANES, L. 2007. A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo. *J Virol,* 81**,** 1701-13.

DELMAS, B., GELFI, J., L'HARIDON, R., VOGEL, L. K., SJOSTROM, H., NOREN, O. & LAUDE, H. 1992. Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature,* 357**,** 417-20.

DENG, X., HACKBART, M., METTELMAN, R. C., O'BRIEN, A., MIELECH, A. M., YI, G., KAO, C. C. & BAKER, S. C. 2017. Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *PNAS,* 114**,** E4251-E4260.

DENISON, M. R., GRAHAM, R. L., DONALDSON, E. F., ECKERLE, L. D. & BARIC, R. S. 2011. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol,* 8**,** 270-9.

DI, H., MADDEN, J. C., MORANTZ, E. K., TANG, H.-Y., GRAHAM, R. L., BARIC, R. S. & BRINTON, M. A. 2017. Expanded subgenomic mRNA transcriptome and coding capacity of a nidovirus. *PNAS,* 114**,** E8895-E8904.

DINAN, A., KEEP, S., BICKERTON, E., BRITTON, P., FIRTH, A. E. & BRIERLEY, I. 2019. Comparative analysis of gene expression in virulent and attenuated strains of infectious bronchitis virus at sub-codon resolution. *Journal of Virology***,** JVI.00714-19.

DOMINGO, E., SHELDON, J. & PERALES, C. 2012. Viral quasispecies evolution. *Microbiol Mol Biol Rev,* 76**,** 159-216.

DOYLE, N., NEUMAN, B. W., SIMPSON, J., HAWES, P. C., MANTELL, J., VERKADE, P., ALRASHEDI, H. & MAIER, H. J. 2018. Infectious Bronchitis Virus Nonstructural Protein 4 Alone Induces Membrane Pairing. *Viruses,* 10.

ECKELS, K. H., SCOTT, R. M., BANCROFT, W. H., BROWN, J., DUBOIS, D. R., SUMMERS, P. L., RUSSELL, P. K. & HALSTEAD, S. B. 1984. Selection of attenuated dengue 4 viruses by serial passage in primary kidney cells. V. Human response to immunization with a candidate vaccine prepared in fetal rhesus lung cells. *Am J Trop Med Hyg,* 33**,** 684-9.

ECKERLE, L. D., BECKER, M. M., HALPIN, R. A., LI, K., VENTER, E., LU, X., SCHERBAKOVA, S., GRAHAM, R. L., BARIC, R. S., STOCKWELL, T. B., SPIRO, D. J. & DENISON, M. R. 2010. Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog,* 6**,** e1000896.

ECKERLE, L. D., LU, X., SPERRY, S. M., CHOI, L. & DENISON, M. R. 2007. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol,* 81**,** 12135-44.

EGLOFF, M.-P., FERRON, F., CAMPANACCI, V., LONGHI, S., RANCUREL, C., DUTARTRE, H., SNIJDER, E. J., GORBALENYA, A. E., CAMBILLAU, C. & CANARD, B. 2004. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 3792-3796.

EIGEN, M. 1971. Self-Organization Of Matter And Evolution Of Biological Macromolecules. *Naturwissenschaften,* 58**,** 465-&.

ELLIS, S., KEEP, S., BRITTON, P., DE WIT, S., BICKERTON, E. & VERVELDE, L. 2018. Recombinant Infectious Bronchitis Viruses Expressing Chimeric Spike Glycoproteins Induce Partial Protective Immunity against Homologous Challenge despite Limited Replication In Vivo. *Journal of Virology,* 92**,** e01473-18.

ERIKSSON, N., PACHTER, L., MITSUYA, Y., RHEE, S. Y., WANG, C., GHARIZADEH, B., RONAGHI, M., SHAFER, R. W. & BEERENWINKEL, N. 2008. Viral population estimation using pyrosequencing. *PLoS Comput Biol,* 4**,** e1000074.

FEHR, A. R. & PERLMAN, S. 2015. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol,* 1282**,** 1-23.

FISCHER, F., STEGEN, C. F., MASTERS, P. S. & SAMSONOFF, W. A. 1998. Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly. *J Virol,* 72**,** 7885-94.

GEERLIGS, H. J., BOELM, G. J., MEINDERS, C. A., STUURMAN, B. G., SYMONS, J., TARRES-CALL, J., BRU, T., VILA, R., MOMBARG, M., KARACA, K., WIJMENGA, W. & KUMAR, M. 2011. Efficacy and safety of an attenuated live QX-like infectious bronchitis virus strain as a vaccine for chickens. *Avian Pathol,* 40**,** 93-102.

GODET, M., GROSCLAUDE, J., DELMAS, B. & LAUDE, H. 1994. Major receptor-binding and neutralization determinants are located within the same domain of the transmissible gastroenteritis virus (coronavirus) spike protein. *J Virol,* 68**,** 8008-16.

GOEBEL, S. J., HSUE, B., DOMBROWSKI, T. F. & MASTERS, P. S. 2004a. Characterization of the RNA Components of a Putative Molecular Switch in the 3′ Untranslated Region of the Murine Coronavirus Genome. *Journal of Virology,* 78**,** 669-682.

GOEBEL, S. J., TAYLOR, J. & MASTERS, P. S. 2004b. The 3' cis-acting genomic replication element of the severe acute respiratory syndrome coronavirus can function in the murine coronavirus genome. *J Virol,* 78**,** 7846-51.

GOODWIN, S., MCPHERSON, J. D. & MCCOMBIE, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics,* 17**,** 333.

GORBALENYA, A. E., KOONIN, E. V., DONCHENKO, A. P. & BLINOV, V. M. 1989. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res,* 17**,** 4847-61.

GUAN, B. J., WU, H. Y. & BRIAN, D. A. 2011. An optimal cis-replication stem-loop IV in the 5' untranslated region of the mouse coronavirus genome extends 16 nucleotides into open reading frame 1. *J Virol,* 85**,** 5593-605.

HAGEMEIJER, M. C., MONASTYRSKA, I., GRIFFITH, J., VAN DER SLUIJS, P., VOORTMAN, J., VAN BERGEN EN HENEGOUWEN, P. M., VONK, A. M., ROTTIER, P. J., REGGIORI, F. & DE HAAN, C. A. 2014. Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. *Virology,* 458-459**,** 125-35.

HALL, R. H. 2017. *The role of infectious bronchitis virus accessory proteins 3a, 3b and 4b*. PhD thesis, University of Liverpool.

HAMZIC, E., KJAERUP, R. B., MACH, N., MINOZZI, G., STROZZI, F., GUALDI, V., WILLIAMS, J. L., CHEN, J., WATTRANG, E., BUITENHUIS, B., JUUL-MADSEN, H. R. & DALGAARD, T. S. 2016. RNA sequencing-based analysis of the spleen transcriptome following infectious bronchitis virus infection of chickens selected for different mannose-binding lectin serum concentrations. *BMC Genomics,* 17**,** 82.

HENNION, R. M. 2015. The preparation of chicken tracheal organ cultures for virus isolation, propagation, and titration. *Methods Mol Biol,* 1282**,** 51-6.

HENNION, R. M. & HILL, G. 2015. The preparation of chicken kidney cell cultures for virus propagation. *Methods Mol Biol,* 1282**,** 57-62.

HEROLD, J., RAABE, T., SCHELLE-PRINZ, B. & SIDDELL, S. G. 1993. Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology,* 195**,** 680-91.

HODGSON, T., CASAIS, R., DOVE, B., BRITTON, P. & CAVANAGH, D. 2004. Recombinant Infectious Bronchitis Coronavirus Beaudette with the Spike Protein Gene of the Pathogenic M41 Strain Remains Attenuated but Induces Protective Immunity. *Journal of Virology,* 78**,** 13804-13811.

HOFMANN, H., PYRC, K., VAN DER HOEK, L., GEIER, M., BERKHOUT, B. & POHLMANN, S. 2005. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc Natl Acad Sci U S A,* 102**,** 7988-93.

HOPKINS, S. R. & YODER, H. W., JR. 1986. Reversion to virulence of chicken-passaged infectious bronchitis vaccine virus. *Avian Dis,* 30**,** 221-3.

HSUE, B., HARTSHORNE, T. & MASTERS, P. S. 2000. Characterization of an essential RNA secondary structure in the 3' untranslated region of the murine coronavirus genome. *J Virol,* 74**,** 6911-21.

HURST, K. R., KOETZNER, C. A. & MASTERS, P. S. 2013. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. *Journal of virology,* 87**,** 9159-9172.

HURST, K. R., KUO, L., KOETZNER, C. A., YE, R., HSUE, B. & MASTERS, P. S. 2005. A Major Determinant for Membrane Protein Interaction Localizes to the Carboxy-Terminal Domain of the Mouse Coronavirus Nucleocapsid Protein. *Journal of Virology,* 79**,** 13285-13297.

HURST, K. R., YE, R., GOEBEL, S. J., JAYARAMAN, P. & MASTERS, P. S. 2010. An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA. *J Virol,* 84**,** 10276-88.

HUSSAIN, S., PAN, J. A., CHEN, Y., YANG, Y., XU, J., PENG, Y., WU, Y., LI, Z., ZHU, Y., TIEN, P. & GUO, D. 2005. Identification of Novel Subgenomic RNAs and Noncanonical Transcription Initiation Signals of Severe Acute Respiratory Syndrome Coronavirus. *Journal of Virology,* 79**,** 5288-5295.

IGNJATOVIC, J. & SAPATS, S. 2000. Avian infectious bronchitis virus. *Rev Sci Tech,* 19**,** 493-508.

IMBERT, I., GUILLEMOT, J.-C., BOURHIS, J.-M., BUSSETTA, C., COUTARD, B., EGLOFF, M.-P., FERRON, F., GORBALENYA, A. E. & CANARD, B. 2006. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *The EMBO journal,* 25**,** 4933-4942.

IRIGOYEN, N., FIRTH, A. E., JONES, J. D., CHUNG, B. Y., SIDDELL, S. G. & BRIERLEY, I. 2016. High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling. *PLoS Pathog,* 12**,** e1005473.

IVANOV, K. A., THIEL, V., DOBBE, J. C., VAN DER MEER, Y., SNIJDER, E. J. & ZIEBUHR, J. 2004. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *Journal of virology,* 78**,** 5619-5632.

JACKWOOD, M. W. 2012. Review of infectious bronchitis virus around the world. *Avian Dis,* 56**,** 634-41.

JORDAN, B. 2017. Vaccination against infectious bronchitis virus: A continuous challenge. *Veterinary Microbiology,* 206**,** 137-143.

KATOH, K. & STANDLEY, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol,* 30**,** 772-80.

KE, M., CHEN, Y., WU, A., SUN, Y., SU, C., WU, H., JIN, X., TAO, J., WANG, Y., MA, X., PAN, J. A. & GUO, D. 2012. Short peptides derived from the interaction domain of SARS coronavirus nonstructural protein nsp10 can suppress the 2'-O-methyltransferase activity of nsp10/nsp16 complex. *Virus Res,* 167**,** 322-8.

KEEP, S. M., BICKERTON, E. & BRITTON, P. 2015. Partial purification of IBV and subsequent isolation of viral RNA for next-generation sequencing. *Methods Mol Biol,* 1282**,** 109-12.

KINT, J., DICKHOUT, A., KUTTER, J., MAIER, H. J., BRITTON, P., KOUMANS, J., PIJLMAN, G. P., FROS, J. J., WIEGERTJES, G. F. & FORLENZA, M. 2015. Infectious Bronchitis Coronavirus Inhibits STAT1 Signaling and Requires Accessory Proteins for Resistance to Type I Interferon Activity. *Journal of Virology,* 89**,** 12047-12057.

KINT, J., LANGEREIS, M. A., MAIER, H. J., BRITTON, P., VAN KUPPEVELD, F. J., KOUMANS, J., WIEGERTJES, G. F. & FORLENZA, M. 2016. Infectious Bronchitis Coronavirus Limits Interferon Production by Inducing a Host Shutoff That Requires Accessory Protein 5b. *Journal of Virology,* 90**,** 7519-7528.

KIRCHDOERFER, R. N., COTTRELL, C. A., WANG, N., PALLESEN, J., YASSINE, H. M., TURNER, H. L., CORBETT, K. S., GRAHAM, B. S., MCLELLAN, J. S. & WARD, A. B. 2016. Pre-fusion structure of a human coronavirus spike protein. *Nature,* 531**,** 118.

KOTTIER, S. A., CAVANAGH, D. & BRITTON, P. 1995. Experimental Evidence of Recombination in Coronavirus Infectious Bronchitis Virus. *Virology,* 213**,** 569-580.

KRUEGER, F. *Trim Galore!* [Online]. Available: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ [Accessed].

KUO, L. & MASTERS, P. S. 2003. The Small Envelope Protein E Is Not Essential for Murine Coronavirus Replication. *Journal of Virology,* 77**,** 4597-4608.

LACONI, A., VAN BEURDEN, S. J., BERENDS, A. J., KRÄMER-KÜHL, A., JANSEN, C. A., SPEKREIJSE, D., CHÉNARD, G., PHILIPP, H.-C., MUNDT, E., ROTTIER, P. J. M. & HÉLÈNE VERHEIJE, M. 2018. Deletion of accessory genes 3a, 3b, 5a or 5b from avian coronavirus infectious bronchitis virus induces an attenuated phenotype both in vitro and in vivo. *Journal of General Virology,* 99**,** 1381-1390.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LARSEN, P. A., HARRIS, R. A., LIU, Y., MURALI, S. C., CAMPBELL, C. R., BROWN, A. D., SULLIVAN, B. A., SHELTON, J., BROWN, S. J., RAVEENDRAN, M., DUDCHENKO, O., MACHOL, I., DURAND, N. C., SHAMIM, M. S., AIDEN, E. L., MUZNY, D. M., GIBBS, R. A., YODER, A. D., ROGERS, J. & WORLEY, K. C. 2017. Hybrid de novo genome assembly and centromere characterization of the gray mouse lemur (Microcebus murinus). *BMC biology,* 15**,** 110-110.

LAURING, A. S. & ANDINO, R. 2010. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog,* 6**,** e1001005.

LEE, C. W. & JACKWOOD, M. W. 2000. Evidence of genetic diversity generated by recombination among avian coronavirus IBV. *Arch Virol,* 145**,** 2135-48.

LI, F. 2012. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J Virol,* 86**,** 2856-8.

LI, F. 2016. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol,* 3**,** 237-261.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, W., MOORE, M. J., VASILIEVA, N., SUI, J., WONG, S. K., BERNE, M. A., SOMASUNDARAN, M., SULLIVAN, J. L., LUZURIAGA, K., GREENOUGH, T. C., CHOE, H. & FARZAN, M. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature,* 426**,** 450-4.

LIU, D. X., CAVANAGH, D., GREEN, P. & INGLIS, S. C. 1991. A polycistronic mRNA specified by the coronavirus infectious bronchitis virus. *Virology,* 184**,** 531-44.

LIU, D. X., FUNG, T. S., CHONG, K. K.-L., SHUKLA, A. & HILGENFELD, R. 2014. Accessory proteins of SARS-CoV and other coronaviruses. *Antiviral Research,* 109**,** 97-109.

LIU, D. X. & INGLIS, S. C. 1991. Association of the infectious bronchitis virus 3c protein with the virion envelope. *Virology,* 185**,** 911-7.

LIU, D. X. & INGLIS, S. C. 1992. Internal entry of ribosomes on a tricistronic mRNA encoded by infectious bronchitis virus. *J Virol,* 66**,** 6143-54.

LU, X., PAN, J., TAO, J. & GUO, D. 2011. SARS-CoV nucleocapsid protein antagonizes IFN-beta response by targeting initial step of IFN-beta induction pathway, and its C-terminal region is critical for the antagonism. *Virus Genes,* 42**,** 37-45.

LU, Y., LU, X. & DENISON, M. R. 1995. Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *Journal of Virology,* 69**,** 3554-3559.

LUKHOVITSKAYA, N. I., THADURI, S., GARUSHYANTS, S. K., TORRANCE, L. & SAVENKOV, E. I. 2013. Deciphering the mechanism of defective interfering RNA (DI RNA) biogenesis reveals that a viral protein and the DI RNA act antagonistically in virus infection. *Journal of virology,* 87**,** 6091-6103.

MACNAUGHTON, M. R. & MADGE, M. H. 1977. The characterisation of the virion RNA of avian Infectious Bronchitis Virus. *FEBS Letters,* 77**,** 311-313.

MAIER, H. J., HAWES, P. C., COTTAM, E. M., MANTELL, J., VERKADE, P., MONAGHAN, P., WILEMAN, T. & BRITTON, P. 2013. Infectious Bronchitis Virus Generates Spherules from Zippered Endoplasmic Reticulum Membranes. *MBio,* 4**,** e00801-13.

MARGERIDON-THERMET, S., SHULMAN, N. S., AHMED, A., SHAHRIAR, R., LIU, T., WANG, C., HOLMES, S. P., BABRZADEH, F., GHARIZADEH, B., HANCZARUK, B., SIMEN, B. B., EGHOLM, M. & SHAFER, R. W. 2009. Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. *J Infect Dis,* 199**,** 1275-85.

MAYR, E. 1988. *Toward a New Philosophy of Biology,* Cambridge, MA, Harvard University Press.

MCBRIDE, R., VAN ZYL, M. & FIELDING, B. C. 2014. The coronavirus nucleocapsid is a multifunctional protein. *Viruses,* 6**,** 2991-3018.

MCKINLEY, E. T., HILT, D. A. & JACKWOOD, M. W. 2008. Avian coronavirus infectious bronchitis attenuated live vaccines undergo selection of subpopulations and mutations following vaccination. *Vaccine,* 26**,** 1274-84.

MEEUSEN, E. N., WALKER, J., PETERS, A., PASTORET, P. P. & JUNGERSEN, G. 2007. Current status of veterinary vaccines. *Clin Microbiol Rev,* 20**,** 489-510, table of contents.

MENACHERY, V. D., YOUNT, B. L., JR., JOSSET, L., GRALINSKI, L. E., SCOBEY, T., AGNIHOTHRAM, S., KATZE, M. G. & BARIC, R. S. 2014. Attenuation and restoration of severe acute respiratory syndrome coronavirus mutant lacking 2'-o-methyltransferase activity. *J Virol,* 88**,** 4251-64.

MIELECH, A. M., CHEN, Y., MESECAR, A. D. & BAKER, S. C. 2014. Nidovirus papain-like proteases: multifunctional enzymes with protease, deubiquitinating and deISGylating activities. *Virus Res,* 194**,** 184-90.

MINSKAIA, E., HERTZIG, T., GORBALENYA, A. E., CAMPANACCI, V., CAMBILLAU, C., CANARD, B. & ZIEBUHR, J. 2006. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci U S A,* 103**,** 5108-13.

NARAYANAN, K., CHEN, C. J., MAEDA, J. & MAKINO, S. 2003. Nucleocapsid-independent specific viral RNA packaging via viral envelope protein and viral RNA signal. *J Virol,* 77**,** 2922-7.

NARAYANAN, K., HUANG, C., LOKUGAMAGE, K., KAMITANI, W., IKEGAMI, T., TSENG, C.-T. K. & MAKINO, S. 2008. Severe Acute Respiratory Syndrome Coronavirus nsp1 Suppresses Host Gene Expression, Including That of Type I Interferon, in Infected Cells. *Journal of Virology* 82**,** 4471-4479.

NARAYANAN, K. & MAKINO, S. 2001. Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. *J Virol,* 75**,** 9059-67.

NDEGWA, E. N., JOINER, K. S., TORO, H., VAN GINKEL, F. W. & VAN SANTEN, V. L. 2012. The proportion of specific viral subpopulations in attenuated Arkansas Delmarva poultry industry infectious bronchitis vaccines influences vaccination outcome. *Avian Dis,* 56**,** 642-53.

NETLAND, J., DEDIEGO, M. L., ZHAO, J., FETT, C., ALVAREZ, E., NIETO-TORRES, J. L., ENJUANES, L. & PERLMAN, S. 2010. Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease. *Virology,* 399**,** 120-128.

NGUYEN, V. P. & HOGUE, B. G. 1997. Protein interactions during coronavirus assembly. *J Virol,* 71**,** 9278-84.

NIETO-TORRES, J. L., DEDIEGO, M. L., VERDIA-BAGUENA, C., JIMENEZ-GUARDENO, J. M., REGLA-NAVA, J. A., FERNANDEZ-DELGADO, R., CASTANO-RODRIGUEZ, C., ALCARAZ, A., TORRES, J., AGUILELLA, V. M. & ENJUANES, L. 2014. Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS Pathog,* 10**,** e1004077.

OADE, M. S., KEEP, S., FREIMANIS, G. L., ORTON, R. J., BRITTON, P., HAMMOND, J. A. & BICKERTON, E. 2019. Attenuation of Infectious Bronchitis Virus in Eggs Results in Different Patterns of Genomic Variation across Multiple Replicates. *J Virol,* 93**,** e00492-19.

OGANDO, N. S., FERRON, F., DECROLY, E., CANARD, B., POSTHUMA, C. C. & SNIJDER, E. J. 2019. The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity. *Front Microbiol,* 10**,** 1813.

OKONECHNIKOV, K., GOLOSOVA, O., FURSOV, M. & TEAM, U. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics,* 28**,** 1166-7.

OPSTELTEN, D. J., RAAMSMAN, M. J., WOLFS, K., HORZINEK, M. C. & ROTTIER, P. J. 1995. Envelope glycoprotein interactions in coronavirus assembly. *J Cell Biol,* 131**,** 339-49.

PENG, G., SUN, D., RAJASHANKAR, K. R., QIAN, Z., HOLMES, K. V. & LI, F. 2011. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc Natl Acad Sci U S A,* 108**,** 10696-701.

PENG, G., XU, L., LIN, Y. L., CHEN, L., PASQUARELLA, J. R., HOLMES, K. V. & LI, F. 2012. Crystal structure of bovine coronavirus spike protein lectin domain. *J Biol Chem,* 287**,** 41931-8.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem,* 25**,** 1605-12.

PHILLIPS, J. E., JACKWOOD, M. W., MCKINLEY, E. T., THOR, S. W., HILT, D. A., ACEVEDOL, N. D., WILLIAMS, S. M., KISSINGER, J. C., PATERSON, A. H., ROBERTSON, J. S. & LEMKE, C. 2012. Changes in nonstructural protein 3 are associated with attenuation in avian coronavirus infectious bronchitis virus. *Virus Genes,* 44**,** 63-74.

PROMKUNTOD, N., VAN EIJNDHOVEN, R. E., DE VRIEZE, G., GRONE, A. & VERHEIJE, M. H. 2014. Mapping of the receptor-binding domain and amino acids critical for attachment in the spike protein of avian coronavirus infectious bronchitis virus. *Virology,* 448**,** 26-32.

QIN, L., XIONG, B., LUO, C., GUO, Z. M., HAO, P., SU, J., NAN, P., FENG, Y., SHI, Y. X., YU, X. J., LUO, X. M., CHEN, K. X., SHEN, X., SHEN, J. H., ZOU, J. P., ZHAO, G. P., SHI, T. L., HE, W. Z., ZHONG, Y., JIANG, H. L. & LI, Y. X. 2003. Identification of probable genomic packaging signal sequence from SARS-CoV genome by bioinformatics analysis. *Acta Pharmacol Sin,* 24**,** 489-96.

QUINCE, C., LANZEN, A., CURTIS, T. P., DAVENPORT, R. J., HALL, N., HEAD, I. M., READ, L. F. & SLOAN, W. T. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods,* 6**,** 639-41.

RAABEN, M., GROOT KOERKAMP, M. J., ROTTIER, P. J. & DE HAAN, C. A. 2007. Mouse hepatitis coronavirus replication induces host translational shutoff and mRNA decay, with concomitant formation of stress granules and processing bodies. *Cell Microbiol,* 9**,** 2218-29.

RAMAN, S., BOUMA, P., WILLIAMS, G. D. & BRIAN, D. A. 2003. Stem-loop III in the 5' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. *J Virol,* 77**,** 6720-30.

REDDY, V. R., THEUNS, S., ROUKAERTS, I. D., ZELLER, M., MATTHIJNSSENS, J. & NAUWYNCK, H. J. 2015. Genetic Characterization of the Belgian Nephropathogenic Infectious Bronchitis Virus (NIBV) Reference Strain B1648. *Viruses,* 7**,** 4488-506.

ROCHE 2008. GS FLX Titanium cDNA rapid library preparation method manual.

ROTTIER, P. J. M. 1995. The Coronavirus Membrane Glycoprotein. *In:* SIDDELL, S. G. (ed.) *The Coronaviridae*. Boston, MA: Springer US.

RUCH, T. R. & MACHAMER, C. E. 2011. The hydrophobic domain of infectious bronchitis virus E protein alters the host secretory pathway and is important for release of infectious virus. *J Virol,* 85**,** 675-85.

RUCH, T. R. & MACHAMER, C. E. 2012. The coronavirus E protein: assembly and beyond. *Viruses,* 4**,** 363-82.

SAIKATENDU, K. S., JOSEPH, J. S., SUBRAMANIAN, V., CLAYTON, T., GRIFFITH, M., MOY, K., VELASQUEZ, J., NEUMAN, B. W., BUCHMEIER, M. J., STEVENS, R. C. & KUHN, P. 2005. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1"-phosphate dephosphorylation by a conserved domain of nsP3. *Structure,* 13**,** 1665-75.

SAWICKI, S. G. & SAWICKI, D. L. 1995. Coronaviruses use Discontinuous Extension for Synthesis of Subgenome-Length Negative Strands. *In:* TALBOT, P. J. & LEVY, G. A. (eds.) *Corona- and Related Viruses*. Boston, MA: Springer US.

SAWICKI, S. G., SAWICKI, D. L. & SIDDELL, S. G. 2007. A contemporary view of coronavirus transcription. *J Virol,* 81**,** 20-9.

SCHIRMER, M., D'AMORE, R., IJAZ, U. Z., HALL, N. & QUINCE, C. 2016. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics,* 17**,** 125.

SCHMIEDER, R. & EDWARDS, R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics,* 27**,** 863-4.

SCHNEIDER, M., ACKERMANN, K., STUART, M., WEX, C., PROTZER, U., SCHÄTZL, H. M. & GILCH, S. 2012. Severe Acute Respiratory Syndrome Coronavirus Replication Is Severely Impaired by MG132 due to Proteasome-Independent Inhibition of M-Calpain. *Journal of Virology,* 86**,** 10112-10122.

SETHNA, P. B., HOFMANN, M. A. & BRIAN, D. A. 1991. Minus-strand copies of replicating coronavirus mRNAs contain antileaders. *Journal of virology,* 65**,** 320-325.

SHANG, J., ZHENG, Y., YANG, Y., LIU, C., GENG, Q., LUO, C., ZHANG, W. & LI, F. 2018. Cryo-EM structure of infectious bronchitis coronavirus spike protein reveals structural and functional evolution of coronavirus spike proteins. *PLOS Pathogens,* 14**,** e1007009.

SIMONS, K. & GAROFF, H. 1980. The budding mechanisms of enveloped animal viruses. *J Gen Virol,* 50**,** 1-21.

SKINNER, M. A., EBNER, D. & SIDDELL, S. G. 1985. Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. *J Gen Virol,* 66 (Pt 3)**,** 581-92.

SMITH, K. A. 2012. Louis pasteur, the father of immunology? *Frontiers in immunology,* 3**,** 68-68.

SOLA, I., ALMAZAN, F., ZUNIGA, S. & ENJUANES, L. 2015. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol,* 2**,** 265-88.

SOLA, I., GALÁN, C., MATEOS-GÓMEZ, P. A., PALACIO, L., ZÚÑIGA, S., CRUZ, J. L., ALMAZÁN, F. & ENJUANES, L. 2011. The polypyrimidine tract-binding protein affects coronavirus RNA accumulation levels and relocalizes viral RNAs to novel cytoplasmic domains different from replication-transcription sites. *Journal of virology,* 85**,** 5136-5149.

SOLA, I., MORENO, J. L., ZÚÑIGA, S., ALONSO, S. & ENJUANES, L. 2005. Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *Journal of virology,* 79**,** 2506-2516.

STERN, D. F. & KENNEDY, S. I. 1980. Coronavirus multiplication strategy. I. Identification and characterization of virus-specified RNA. *Journal of virology,* 34**,** 665-674.

SUN, Y., JAIN, D., KOZIOL-WHITE, C. J., GENOYER, E., GILBERT, M., TAPIA, K., PANETTIERI, R. A., JR., HODINKA, R. L. & LÓPEZ, C. B. 2015. Immunostimulatory Defective Viral Genomes from Respiratory Syncytial Virus Promote a Strong Innate Antiviral Response during Infection in Mice and Humans. *PLoS pathogens,* 11**,** e1005122-e1005122.

SURJIT, M., LIU, B., CHOW, V. T. & LAL, S. K. 2006. The nucleocapsid protein of severe acute respiratory syndrome-coronavirus inhibits the activity of cyclin-cyclin-dependent kinase complex and blocks S phase progression in mammalian cells. *J Biol Chem,* 281**,** 10669-81.

THIEL, V. & SIDDELL, S. G. 1994. Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5. *J Gen Virol,* 75 (Pt 11)**,** 3041-6.

THOR, S. W., HILT, D. A., KISSINGER, J. C., PATERSON, A. H. & JACKWOOD, M. W. 2011. Recombination in avian gamma-coronavirus infectious bronchitis virus. *Viruses,* 3**,** 1777-99.

TORO, H., VAN SANTEN, V. L. & JACKWOOD, M. W. 2012. Genetic diversity and selection regulates evolution of infectious bronchitis virus. *Avian Dis,* 56**,** 449-55.

TSAI, W.-C. & LLOYD, R. E. 2014. Cytoplasmic RNA Granules and Viral Infection. *Annual review of virology,* 1**,** 147-170.

UNIPROT, C. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res,* 47**,** D506-D515.

VALASTRO, V., HOLMES, E. C., BRITTON, P., FUSARO, A., JACKWOOD, M. W., CATTOLI, G. & MONNE, I. 2016. S1 gene-based phylogeny of infectious bronchitis virus: An attempt to harmonize virus classification. *Infect Genet Evol,* 39**,** 349-364.

VAN BEURDEN, S. J., BERENDS, A. J., KRAMER-KUHL, A., SPEKREIJSE, D., CHENARD, G., PHILIPP, H. C., MUNDT, E., ROTTIER, P. J. M. & VERHEIJE, M. H. 2018. Recombinant live attenuated avian coronavirus vaccines with deletions in the accessory genes 3ab and/or 5ab protect against infectious bronchitis in chickens. *Vaccine,* 36**,** 1085-1092.

VENNEMA, H., GODEKE, G. J., ROSSEN, J. W., VOORHOUT, W. F., HORZINEK, M. C., OPSTELTEN, D. J. & ROTTIER, P. J. 1996. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *Embo j,* 15**,** 2020-8.

VERGARA, I. A., FRECH, C. & CHEN, N. 2012. CooVar: Co-occurring variant analyzer. *BMC Research Notes,* 5**,** 615.

VIEHWEGER, A., KRAUTWURST, S., LAMKIEWICZ, K., MADHUGIRI, R., ZIEBUHR, J., HOLZER, M. & MARZ, M. 2019. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res,* 29**,** 1545-1554.

VIGNUZZI, M., STONE, J. K., ARNOLD, J. J., CAMERON, C. E. & ANDINO, R. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature,* 439**,** 344-8.

VOELKERDING, K. V., DAMES, S. A. & DURTSCHI, J. D. 2009. Next-Generation Sequencing: From Basic Research to Diagnostics. *Clinical Chemistry,* 55**,** 641-658.

WALLS, A. C., TORTORICI, M. A., BOSCH, B.-J., FRENZ, B., ROTTIER, P. J. M., DIMAIO, F., REY, F. A. & VEESLER, D. 2016. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature,* 531**,** 114.

WEISS, S. R. & NAVAS-MARTIN, S. 2005. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev,* 69**,** 635-64.

WICKRAMASINGHE, I. N., DE VRIES, R. P., GRONE, A., DE HAAN, C. A. & VERHEIJE, M. H. 2011. Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *J Virol,* 85**,** 8903-12.

WILLIAMS, G. D., CHANG, R. Y. & BRIAN, D. A. 1999. A phylogenetically conserved hairpin-type 3' untranslated region pseudoknot functions in coronavirus RNA replication. *J Virol,* 73**,** 8349-55.

WILM, A., AW, P. P., BERTRAND, D., YEO, G. H., ONG, S. H., WONG, C. H., KHOR, C. C., PETRIC, R., HIBBERD, M. L. & NAGARAJAN, N. 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res,* 40**,** 11189-201.

WILSON, L., MCKINLAY, C., GAGE, P. & EWART, G. 2004. SARS coronavirus E protein forms cation-selective ion channels. *Virology,* 330**,** 322-31.

WINTER, C., SCHWEGMANN-WESSELS, C., CAVANAGH, D., NEUMANN, U. & HERRLER, G. 2006. Sialic acid is a receptor determinant for infection of cells by avian Infectious bronchitis virus. *J Gen Virol,* 87**,** 1209-16.

WRIGHT, C. F., MORELLI, M. J., THEBAUD, G., KNOWLES, N. J., HERZYK, P., PATON, D. J., HAYDON, D. T. & KING, D. P. 2011. Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. *J Virol,* 85**,** 2266-75.

XU, X., LIU, Y., WEISS, S., ARNOLD, E., SARAFIANOS, S. G. & DING, J. 2003. Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic acids research,* 31**,** 7117-7130.

YOUNT, B., CURTIS, K. M. & BARIC, R. S. 2000. Strategy for Systematic Assembly of Large RNA and DNA Genomes: Transmissible Gastroenteritis Virus Model. *Journal of VIrology,* 74**,** 10600-10611.

ZANINI, F., PU, S. Y., BEKERMAN, E., EINAV, S. & QUAKE, S. R. 2018a. Single-cell transcriptional dynamics of flavivirus infection. *Elife,* 7.

ZANINI, F., ROBINSON, M. L., CROOTE, D., SAHOO, M. K., SANZ, A. M., ORTIZ-LASSO, E., ALBORNOZ, L. L., ROSSO, F., MONTOYA, J. G., GOO, L., PINSKY, B. A., QUAKE, S. R. & EINAV, S. 2018b. Virus-inclusive single-cell RNA sequencing reveals the molecular signature of progression to severe dengue. *PNAS,* 115**,** E12363-E12369.

ZENG, L., GODENY, E. K., METHVEN, S. L. & BRINTON, M. A. 1995. Analysis of simian hemorrhagic fever virus (SHFV) subgenomic RNAs, junction sequences, and 5' leader. *Virology,* 207**,** 543-8.

ZERBINO, D. R., ACHUTHAN, P., AKANNI, W., AMODE, M R., BARRELL, D., BHAI, J., BILLIS, K., CUMMINS, C., GALL, A., GIRÓN, C. G., GIL, L., GORDON, L., HAGGERTY, L., HASKELL, E., HOURLIER, T., IZUOGU, O. G., JANACEK, S. H., JUETTEMANN, T., TO, J. K., LAIRD, M. R., LAVIDAS, I., LIU, Z., LOVELAND, J. E., MAUREL, T., MCLAREN, W., MOORE, B., MUDGE, J., MURPHY, D. N., NEWMAN, V., NUHN, M., OGEH, D., ONG, C. K., PARKER, A., PATRICIO, M., RIAT, H. S., SCHUILENBURG, H., SHEPPARD, D., SPARROW, H., TAYLOR, K., THORMANN, A., VULLO, A., WALTS, B., ZADISSA, A., FRANKISH, A., HUNT, S. E., KOSTADIMA, M., LANGRIDGE, N., MARTIN, F. J., MUFFATO, M., PERRY, E., RUFFIER, M., STAINES, D. M., TREVANION, S. J., AKEN, B. L., CUNNINGHAM, F., YATES, A. & FLICEK, P. 2017. Ensembl 2018. *Nucleic Acids Research,* 46**,** D754-D761.

ZHAI, Y., SUN, F., LI, X., PANG, H., XU, X., BARTLAM, M. & RAO, Z. 2005. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat Struct Mol Biol,* 12**,** 980-6.

ZHANG, J., KOBERT, K., FLOURI, T. & STAMATAKIS, A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics (Oxford, England),* 30**,** 614-620.

ZHANG, X. & LIU, R. 2000. Identification of a Noncanonical Signal for Transcription of a Novel Subgenomic mRNA of Mouse Hepatitis Virus: Implication for the Mechanism of Coronavirus RNA Transcription. *Virology,* 278**,** 75-85.

ZHAO, F., HAN, Z., ZHANG, T., SHAO, Y., KONG, X., MA, H. & LIU, S. 2014. Genomic characteristics and changes of avian infectious bronchitis virus strain CK/CH/LDL/97I after serial passages in chicken embryos. *Intervirology,* 57**,** 319-30.

ZHAO, L., LIU, L., WANG, S., WANG, H. & JIANG, J. 2016. Transcriptome profiles of metamorphosis in the ornamented pygmy frog Microhyla fissipes clarify the functions of thyroid hormone receptors in metamorphosis. *Sci Rep,* 6**,** 27310.

ZHAO, X., SHAW, K. & CAVANAGH, D. 1993. Presence of subgenomic mRNAs in virions of coronavirus IBV. *Virology,* 196**,** 172-8.

ZHAO, Y., CHENG, J., YAN, S., JIA, W., ZHANG, K. & ZHANG, G. 2019. S gene and 5a accessory gene are responsible for the attenuation of virulent infectious bronchitis coronavirus. *Virology,* 533**,** 12-20.

ZIEBUHR, J., SNIJDER, E. J. & GORBALENYA, A. E. 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *Journal of General Virology,* 81**,** 853-879.

ZUNIGA, S., SOLA, I., ALONSO, S. & ENJUANES, L. 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *Journal of Virology,* 78**,** 980-994.

ZUNIGA, S., SOLA, I., MORENO, J. L., SABELLA, P., PLANA-DURAN, J. & ENJUANES, L. 2007. Coronavirus nucleocapsid protein is an RNA chaperone. *Virology,* 357**,** 215-27.