GENOMIC ANALYSIS OF PATHOGENIC AND ATTENUATED STRAINS OF INFECTIOUS BRONCHITIS VIRUSES

by

JAMIE EVELYN PHILLIPS

(Under the Direction of Mark Walter Jackwood)

ABSTRACT

Infectious Bronchitis Virus (IBV) is a positive sense RNA virus that is a pathogen of fowl. Pathogenicity of IBV has classically been attributed to the spike glycoprotein. However, recent evidence in literature indicates that the polymerase genes play a role in pathogenicity. In this study the full-length genome of three homologous pathogenic parent and attenuated progeny strains of IBV were sequenced and examined. Comparative genomics were conducted on IBV strains: Ark-DPI pathogenic, Ark-DPI attenuated, GA98 pathogenic, GA98 attenuated, Mass41 pathogenic, and Mass41 attenuated. Within strains, differences between pathogenic and attenuated viruses were observed. Although the role these mutations play in pathogenicity must be examined, if they are verified they could be used to significantly improve IBV vaccine development.

INDEX WORDS: Infectious Bronchitis Virus, Coronaviridae, Antigenic variants, RT-PCR, Complete Genome

GENOMIC ANALYSIS OF PATHOGENIC AND ATTENUATED STRAINS OF INFECTIOUS BRONCHITIS VIRUSES

by

JAMIE EVELYN PHILLIPS

BS, Georgia Southern University, 2007

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2009

© 2009

Jamie Evelyn Phillips

All Rights Reserve

GENOMIC ANALYSIS OF PATHOGENIC AND ATTENUATED STRAINS OF INFECTIOUS BRONCHITIS VIRUSES

by

JAMIE EVELYN PHILLIPS

Major Professor: Mark W. Jackwood

Committee: Mark Tompkins
Liliana Jaso- Friedman

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2009

DEDICATION

I would like to dedicate this Thesis to my parents Paul and Debbie Phillips for their unconditional love and support.

ACKNOWLEDGEMENTS

I would like to thank everyone in the “chocolate lab”, Debbie, Sharmi, Enid and of course Dr. Jackwood. I have thoroughly enjoyed working along side each one of you. I am forever grateful that I was given this wonderful opportunity and look forward to spending the next couple years in your company working on my PhD. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

Chapter 1

Introduction

Background on study

Over the last several decades the majority of emerging infectious diseases have been the result of interspecies transmission of zoonotic RNA viruses (Vijaykrishna *et al*., 2007). One of these outbreaks was the Severe Acute Respiratory Syndrome (SARS), caused by an ssRNA virus in the Coronaviridae family (Vijaykrishna *et al*., 2007). The SARS virus was first transmitted from animals to humans in 2003; worldwide, SARS resulted in the deaths of over 700 individuals and sickness in 8,000 others (WHO website). The genome of this virus was rapidly sequenced and identified as a group 2b coronavirus (Marra *et al*., 2003). Since that outbreak, many studies have been conducted to determine the origin of the SARS coronavirus (SARS-CoV), and the natural host reservoir. In addition, pathogenicity studies and studies targeting the rapid development of vaccines have also been conducted.

Coronaviruses are of great importance because they are distributed globally and are highly infectious. These viruses cause respiratory and enteric diseases in animals as well as humans. Because coronaviruses undergo genetic mutation and recombination at a high rate, it is not surprising that coronaviruses can readily move between animal host species and can be transmitted to humans causing zoonotic disease. Infectious bronchitis virus (IBV), a group III coronavirus, causes highly contagious upper-respiratory disease in chickens. Like SARS, IBV

undergoes a high rate of genetic mutation and recombination, however, its mechanism of pathogenicity is not fully understood.

The outer glycoprotein, Spike, which is used for attachment and cell fusion, was originally thought to be the sole determinant of IBV pathogenicity. Infectious bronchitis virus is unique among coronaviruses because variability in its spike glycoprotein results in different serotypes of the virus that do not cross protect. In a study involving an infectious clone of IBV, the spike glycoprotein gene from a pathogenic IBV isolate was inserted into an attenuated strain of IBV and the resulting recombinant virus was immunogenic but remained attenuated (Hodgson *et al.*, 2004). In another infectious clone study, replicase genes that encode two polyproteins, [designated 1a and 1a/b] and account for approximately two-thirds of the viral genome in an attenuated virus, were cloned into a pathogenic virus resulting in an attenuated recombinant virus (Cavanagh *et al.*, 2007). Based on these two studies, it appears that the replicase proteins 1a and 1a/b play a role in the pathogenicity of IBV.

Following translation, the polyproteins 1a and 1a/b are cleaved into fifteen nonstructural proteins (Nsps) that make up the replication transcription complex. Currently, little is known about the Nsps found in IBV; however, based on sequence homology with other viruses including other coronaviruses, some functions can be predicted. Nonstructural proteins 3, 4, and 6 all have hydrophobic domains and are predicted to serve as scaffolding proteins for the replication transcription complex. Nonstructural proteins 3 and 5 are proteases that are responsible for the cleavage of the Nsps. Nonstructural proteins 7-10 are putative RNA binding proteins and are thought to participate in the RNA replication process. Nonstructural protein 12 is the predicted

RNA dependent RNA polymerase, Nsp13 is a helicase, Nsp14 is an exoribonuclease, Nsp15 is an endoribonuclease, and Nsp16 is a methyltransferase (Graham R.L, 2008).

Objective

The objective of this work was to conduct full genome sequencing of selected IBV strains to identify genetic markers for pathogenicity. We sequenced pathogenic parent and attenuated progeny strains for three different serotypes of IBV. The attenuated viruses were passaged in embryonating chicken eggs approximately 100X until they were no longer pathogenic for chickens. Noncoding regions as well as nonstructural proteins and structural genes, and their deduced amino acid sequences, spanning the entire genome were examined.

References

Cavanagh, D., Casais, R., Armesto, M., Hodgson, T., Izadkhasti, S., Davies, M., Lin, F., Tarpey, I. and Britton, P. (2007) Manipulation of the infectious bronchitis coronavirus genome for vaccine development and analysis of the accessory proteins. Vaccine 25(30), 5558-62.

Hodgson, T., Casais, R., Dove, B., Britton, P. and Cavanagh, D. (2004) Recombinant infectious bronchitis coronavirus Beaudette with the spike protein gene of the pathogenic M41 strain remains attenuated but induces protective immunity. J Virol 78(24), 13804-11.

Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y., Cloutier, A., Coughlin, S.M., Freeman, D., Girn, N., Griffith, O.L., Leach, S.R., Mayo, M., McDonald, H., Montgomery, S.B., Pandoh, P.K., Petrescu, A.S., Robertson, A.G., Schein, J.E., Siddiqui, A., Smailus, D.E., Stott, J.M., Yang, G.S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T.F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G.A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R.C., Krajden, M., Petric, M., Skowronski, D.M., Upton, C. and Roper, R.L. (2003) The Genome sequence of the SARS-associated coronavirus. Science 300(5624), 1399-404.

Vijaykrishna, D., Smith, G.J.D., Zhang, J.X., Peiris, J.S.M., Chen, H. and Guan, Y. (2007) Evolutionary Insights into the Ecology of Coronaviruses. J. Virol. 81(8), 4012-4020.

Chapter 2

Review of Literature

History

A mild upper-respiratory disease in chickens was first observed in 1931 by Schalk and Hawn and was recently reviewed by Fabricant (1998).The following is a summary of that review. The upper respiratory disease was initially characterized in birds gasping for air and exhibiting general listlessness; albeit, the infectious agent was not identified until two years later. The infectious agent was shown to be transmitted by Berkefeld filtered material indicating that it was viral in origin; moreover, because it was an upper respiratory disease, Bushnell and Brandly reported it as a form of infectious laryngotracheitis (LT). After much confusion, Beach and Schalm showed that the agent causing infectious bronchitis (IB) was distinct from LT in cross immunity studies. The clinical presentation of IB is often indistinguishable from mild forms of LT making virus isolation important for definitive diagnosis. In 1937, Beudette and Hudson propagated the virus in chicken embryos by inoculating the chorioallantoic fluid and showed that with continued passage that the virus became increasingly more virulent to chicken embryos (Fabricant, 1998; Goren, 1978). Infectious bronchitis virus (IBV) was first identified in young chickens and was later isolated from broilers and layers. By the 1940's IBV became widespread in commercial poultry causing significant economic impact. The disease is economically important because the infection results in loss of feed efficiency and condemnations at

processing plants in broilers. In layers it causes oviduct damage, that results in a decline in egg production as well as quality. The eggs are usually thin, rough, or no shells are produced. In chicks early infection with IBV can result in permanent oviduct damage resulting in the inability of these birds to lay eggs. Another factor resulting in increased cost is there are multiple serotypes and the virus is easily transmitted thus it is expensive to try to prevent this disease by vaccination (Cavanagh, 2007). These are a few factors which cause IBV to be of economic importance.

Etiology

Classified in the order of Nidovirales, avian IBV is a member of the *Coronaviridae* family and includes the Coronavirus genera. Coronaviruses are separated into four groups based on antigenic and genetic properties. Groups I, II, and IIb differ from IBV in respect to the genome organization (Cavanagh D., 2008). Group I viruses include porcine *Transmissible gastroenteritis virus* (TGEV), *Feline coronavirus* (FeCoV), *Canine coronavirus* (CCoV), *Human coronavirus* 229E (HCoV-229E) and *Porcine epidemic diarrhea virus* (PEDV) (Holmes, 1991). Group II viruses include the *Murine hepatitis coronavirus* (MHV), *Bovine coronavirus* (BCoV), *Human coronavirus* OC43 (HCoV-OC43), *Porcine hemagglutinating encephalomyelitis virus* (HEV), *Rat coronavirus* (RtCoV), and *Equine coronavirus* (ECoV). Although it shares some antigenic and genetic properties with the Group II coronaviruses, the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) is unique and thus is placed in Group IIb (Cavanagh, 2003). In Group III, along with IBV are avian viruses such as *Turkey Coronavirus (TCoV) Pheasant Coronavirus* and several others (Cavanagh *et al*., 2002*)*.

Turkey coronaviruses, which cause enteric disease in young turkeys, are thought to have originated from IBV (Goodwin, 1995).

Virion Structure

Infectious Bronchitis Virus is round to pleomorphic in shape. It is an enveloped virus approximately 120nm in diameter with club shaped glycoproteins around the outside of the virion that are approximately 20nm in length (Cavanagh D., 2008). Infectious Bronchitis Virus has a single stranded positive sense RNA genome approximately 27Kb in length. When the spike protein is intact, IBV has a sucrose density of 1.18g/ml. If the spike protein is destroyed by centrifugation at forces greater than 100,000g or incubation of 37 ºC then the sucrose density will be approximately 1.15g/ml.

Strain Classifications

Serotypes

Infectious Bronchitis Viruses are grouped into serotypes by the virus neutralization (VN) test. Over fifty serotypes and countless variants of IBVs have been identified since 1936 (Lee *et al.*, 2003). Traditionally, the serotypes were determined by VN testing, but due to limited availability of reference sera and the labor intensive nature of these types of tests, they are not routinely being used for identification purposes (Lee *et al.*, 2003).

Genotypes

Serotype correlates with the sequence of the S1 gene. Reverse transcriptase polymerase chain (RT-PCR) reaction can be used to amplify the S1 gene. Then, either restriction fragment length polymorphism (RFLP) or sequence analysis can be used to determine the virus type. Restriction fragment length polymorphism involves using restriction enzymes to cut the amplified RT-PCR product. Banding patterns are observed on an agarose gel with different patterns being specific for different serotypes. Direct sequencing of the RT-PCR product followed by BLAST (National Center for Biotechnological Information, http://blast.ncbi.nlm.nih.gov/Blast.cgi) analysis is another means of determining the serotype of the isolate. Furthermore, it has been shown that an isolate can be typed based on the sequence of the hypervariable regions (HVR) 1 and 2 located within the S1 portion of the spike protein (Lee *et al*., 2003). Typically, strains that are classified within the same type have a 90% or greater amino acid identity (Cavanagh D., 2008). Sequencing is the preferred method of IBV type identification, because IBV undergoes mutations and recombination events at a high frequency, thus a single mutation could go undetected in the RFLP test, whereas it would clearly be identified by sequencing (Lee *et al*., 2003).

Clinical Signs and Pathology

Primarily a disease of chickens (*Gallus gallus domesticus*), IBV initially infects the upper respiratory tract. In addition to the sinuses and trachea, it has been reported that the virus replicates in numerous epithelial tissues in organs such as the kidneys, oviduct, testis, oesophagus, proventriculus, duodenum, jejunum, bursa of fabricius, and caecal tonsils

(Cavanagh, 2007). Infectious Bronchitis Virus causes disease in chickens but has also been isolated from pheasants, guinea fowl, peafowl, partridge, blue-winged teal, pigeon, mallard duck, and a greylay goose (Cavanagh, 2007). In young birds the disease is characterized by acute respiratory signs and tracheal lesions. Clinical signs of IBV infection include gasping, coughing, sneezing, tracheal rales, and nasal discharge. Infected chicks are generally depressed and tend to huddle toward heat sources. The feed consumption and weight gain in infected fowl are dramatically reduced. The same clinical symptoms occur in older birds; however, the nasal discharge does not occur as frequently (Cavanagh, 2007a). Infectious Bronchitis Virus damages the respiratory epithelium in young chicks allowing opportunistic pathogens to invade. It has long been recognized that infection with IBV exacerbates the disease associated with pathogenic avian mycoplasma (Cavanagh D., 2008). In broilers, the most common outcome of a virulent IBV infection is airsacculitis and systemic colibacillosis (Matthijs *et al*., 2005). Laying hens show clinical signs such as a decrease in egg production and egg shell quality. Moreover, there is evidence that IBV can establish persistent infections and reemerge at the start of egg production. This phenomenon of persistence is thought to occur due to lymphoid cell infiltration and degeneration of the oviduct (Cavanagh, 2007b).

Some strains of IBV, such as Holte, Gray, and Australian T, are nephropathogenic (Cavanagh D., 2008). These strains infect the kidneys and cause dehydration, depression, and sometimes death (Albassam MA, 1986). Broilers infected with a nephropathogenic strain of IBV will appear to recover from the respiratory phase of the disease then show signs of depression, ruffled feathers, wet droppings, increased water intake and sometimes mortality. There is an age related resistance to nephropathogenic strains of IBV (Cavanagh D., 2008).

All ages of chickens are susceptible to IBV, but the disease is most severe in chicks less than 2 weeks old. Mortality varies depending on the virus serotype, type of bird, age, immunity, environmental stresses, and secondary bacterial infections present. Infectious bronchitis virus is highly infectious, and morbidity is usually 100% when secondary pathogens are present. Remarkably, mortality from IBV is below 5% when secondary pathogens are absent (Holmes, 1991).

Distribution

The disease is endemic throughout the world where chickens are commercially produced. However, different serotypes of the IBV can be specific to different geographic regions. Different serotypes of the virus emerge as the result of insertions, deletions, point mutations or recombination events within the spike glycoprotein gene (Gelb, 1997). In the United States, the most commonly detected serotypes are: Massachusetts (Mass), Connecticut (Conn), Arkansas (Ark), Georgia (GA), Delaware (De/072/92), and California. There are many other serotypes of the virus that are common in Europe and Australia (Gelb, 1997). Infectious Bronchitis Virus

Transmission

Infectious Bronchitis Virus spreads rapidly through flocks and has a short incubation period with birds developing clinical signs as early as 24-48 hours following infection (Cavanagh D., 2008). The virus can be isolated from the trachea, lungs, kidney, cecal tonsil, and the bursa of Fabricius for up to seven days after the initial infection. The incubation time for IBV can be as short as 18 hours for intratracheal inoculation and up to 36 hours for ocular inoculation depending on viral

dose (Hofstad, 1996). In a previous study, it was also shown that some vaccine strains can persist in birds for up to 163 days (Cavanagh D., 2008). During this period the virus may be shed and can lead to horizontal transmission between flocks (Cavanagh D., 2008).

Laboratory Host Systems

The majority of IBV isolates replicate in 10-11 day old embryonated chicken eggs. The most common route of inoculation is the allanotic cavity. Some field viruses often require several passages in eggs before high titers are obtained. Typical characteristics of IBV replication in the embryos include; stunting, clubbing of the down, curling of the embryo, and mortality (Gelb, 1998). Another common indicator of IBV is the production of urates in the mesonephros of the embryonic kidney; this feature is also observed when embryos are infected with avian adenovirus. For primary isolates of field viruses, 90% survival is seen through the 19^{th} day of incubation. In contrast, upon the 10^{th} passage of IBV embryo mortality may be as high as 80%. The optimum yield of virus occurs 36-40 hours post inoculation at 37 °C (Gelb JJ, 1998).

Most IBV isolates are not adapted to replicate in cell culture, however, some strains including Beaudette, Mass 41, and Iowa 97 grow in African green monkey Vero cells. Several strains of IBV are also able to infect primary chicken kidney cells. Typical cytopathic effects include cells rounding up and detaching from the culture surface. The cells appear as large spheres with refractile contents (Gelb JJ, 1998).

Immunity

The immune response to IBV strains varies among different breeds of chickens. Chickens that have recovered from a natural infection are resistant to challenge with the same virus. Little to no heterologous protection between different serotypes of the virus exists. The protective immune response is based on the innate, humoral, and cellular immunity (Holmes, 1991). It is well established that the S1 subunit of the S protein induces virus neutralizing antibodies in the host. One study indicated that IgA antibody in nasal secretions played a role in preventing re-infection (Cavanagh D., 2008). The humoral immune response is typically measured by the amount of virus specific IgG or IgM produced in birds. There are several tests that are used to measure humeral immunity such as virus neutralization (Raj GD, 1997). As for the cellular immunity, previous studies indicate that CD8+ cytotoxic T lymphocytes (CTL) play a role in recovery from IBV infection in chickens. A study in 2001 showed that chicken interferon inhibited the replication of IBV (Pei *et al*., 2001). The role this cytokine plays in protection from IBV is unclear. Nevertheless, the innate and humoral immune responses are critical for protection against IBV.

Vaccination

The first vaccine against IBV was created by van Roekal in the 1950's for layers (Van Roekel *et al*., 1958). Currently, the disease is mostly controlled using live attenuated viruses. Attenuation of these viruses occurs through serial passages (100X or more) in embryonating chicken eggs. The vaccination program used in a particular flock depends on the circulating field viruses causing disease in that area. Live vaccines are given at one day of age to broilers in order to

prevent poor feed to growth ratios however; protection is short lived so re-vaccination is necessary. Re-vaccination depending on demographics may include a mixture of serotypes such as a Mass and Conn vaccine or a Mass and Conn and Ark vaccine (Cavanagh, 2007). The efficacies of vaccines vary with respect to the serotype of the virus in the vaccine as well as genetic differences between individual birds. A previous study showed that some live vaccines such as the Arkansas serotype persist in the respiratory tract especially when given in combination with another IBV live vaccine (Alvarado *et al*., 2006). The most common vaccine serotypes used in the United States are Mass 41, Arkansas, Connecticut, Delaware, and Georgia.

Genome Organization and Encoded Proteins

The viral genome of IBV is the largest of all known RNA viruses (Holmes, 1991). It is positive sense, caped, and contains a poly A tail. The RNA dependent RNA polymerase genes located at the 5' end, account for approximately two-thirds of the viral genome and encode two open reading frames that are denoted as 1a and 1ab. IBV has four structural genes: Spike (S), Envelope (E), Membrane (M), and the Nucleocapsid (N). The RNA genome associates with the nucleocapsid phosphoprotein and forms a flexible helical Nucleocapsid protein (Holmes, 1991). The Spike protein is a surface glycoprotein that plays a role in attachment, fusion of the viral and cell membranes, and contains neutralizing antibody epitopes. The Spike protein, which is glycosylated, has a molecular mass of about 150-180kD and is divided into two subdomains (designated S1 and S2).

The bulbous portion of the Spike protein is made up of the S1 and contains the N-terminus of the molecule. This domain has the neutralizing epitopes and is also responsible for binding to the

host cell receptor. The S1 sequences are the most variable of all the structural proteins with serotypes commonly differing in sequence identity by 20 to 25% in the amino-terminal half (Hodgson *et al.*, 2004). The S2 domain is more conserved and constitutes the stalk portion of the protein.

The Envelope protein is a small membrane protein that is involved with triggering virus assembly, in conjunction with the Membrane protein. Downstream of the Envelope protein are two accessory proteins, 3a and 3b, with presently unknown functions. These two nonstructural proteins are produced from a polycistronic mRNA that also encodes the Envelope protein. The 3a protein is approximately 57 amino acids in length; whereas, the 3b protein is approximately 64 amino acids in length. The first 23 amino acids in the 3a protein are believed to be a signal sequence. Both of these proteins are highly conserved in most serotypes of IBV; however it has been shown that they are not essential for replication (Hodgson *et al.*, 2006).

The most abundant glycoprotein in this virus is the Membrane protein, which spans the viral envelope bilayer three times (Vennema *et al.*, 1996). The Membrane protein is N-linked glycosylated at the amino-terminus of the protein. Only a short amino-terminal domain is exposed on the outer surface of the viral envelope monoclonal antibodies against the external portion of the Membrane protein in the presence of complement can neutralize the virus (Holmes, 1991). The Membrane and Nucleocapsid proteins determine the virus budding site but the exact mechanism of assembly is not understood. Located downstream of the Membrane protein are two nonstructural proteins designated 5a and 5b. These have yet to be functionally defined (Cavanagh *et al.*, 2007). The 5a protein contains approximately 65 amino acids; whereas, 5b is approximately 82 amino acids in length.

The Nucleocapsid protein, which is encoded near the 3' end of the genome, is 409 amino acids in length and is a phosphoprotein that interacts with viral genomic RNA. It has three highly conserved domains with one being an RNA-binding domain responsible for binding to the leader sequence of the viral RNA. The Nucleocapsid protein also plays a vital role in replication as it has been shown to bind to cellular membranes and phospholipids. It has previously been determined that the Nucleocapsid protein in MHV does not have to associate with the Membrane protein in order to bind RNA (Anderson *et al.*, 1993).

Lastly, the 3' untranslated region (UTR) of the genome has a unique terminal repeat of various lengths depending on group and serotype of the virus. Predicted bulged stem-loop and pseudoknot structures that are located downstream of the Nucleocapsid were recently described in bat-CoV HKU4 (Woo *et al.*, 2009). These two structures are thought to play a role in virulence. The polymerase binds to the 3' end of the genome and if there is a secondary structure that aids in the efficiency or prevents the polymerase from binding this can either increase virulence or decrease virulence respectively. The extreme 3' end of the genome has a poly-adenylated tail (Woo *et al.*, 2009).

Polyprotein 1a and 1a/b

The two open reading frames 1a and 1ab are approximately 4,382 amino acids and 7,073 amino acids in length, respectively. These two polyproteins encode the replication transcription complex (RTC) that is involved in synthesis of the subgenomic mRNAs and replication of the viral genome. The 1ab polyprotein is translated at an efficiency of 20-40% through a -1 frame shift translation mechanism (Imbert *et al.*, 2008). These polyproteins contain two proteinases, a

papain-like (PLp) protease and a main protease (Mpro) also known as the 3C-like proteinase (3 CLpro) that cleaves 1a and 1ab into 15 non-structural proteins that are essential for replication. Most coronaviruses have two papain-like proteases (PLp1 and PLp2) but the PLp1 protease in IBV is truncated and non-functional. The biological characteristics of many of the nonstructural proteins (Nsps) have previously been characterized. The most notable is Nsp5, which contains Mpro, and is responsible for the cleavage of 11 out of the 14 Nsps (Imbert *et al*., 2008).

Non-structural protein 1 is only present in Group I viruses. Therefore the first protein that is postranslationally cleaved from polyproteins 1a in IBV is Nsp 2. This protein is proposed to span the membrane due to its hydrophobic residues. It is cleaved by the PLp2 protease.

The Nsp3 is responsible for the cleavage of Nsp 2/3 and 3/4 and has many biologically important functions. The PLp2 protease is located in Nsp3 and is solely responsible for cleaving the polyproteins 1ab at two sites. The crystal structure of PLp2 has been elucidated for the group I and group III coronaviruses and many putative binding domains have been identified (Piotrowski Y., 2009). This protein consists of five domains of various functions. The putative domains are an ubiquitin domain, an acidic domain, an ADP-ribose-1-phosphatase, PLp2, a transmembrane domain, and a Y domain.

The deubiquitinase located in Nsp3 has been found to be highly conserved in group II and III viruses as well as SARS-CoV (Imbert *et al*., 2008). This domain plays a role in viral infection. A recent study by Zheng *et al*. (2008) concluded that in MHV and SARS-CoV this domain can bind to Interferon regulatory factor 3 (IRF3) causing its deubiquitination and inhibition of nuclear translocation subsequently inhibiting TANK-binding kinase (TBK1-), and IRF3-mediated interferon beta (IFNBeta) activities. It has previously been reported that MHV and

SARS-CoV induce low levels of IFN during the initial infection (Zheng *et al*. 2008). In that study, a MHV PLp2 mutant was created lacking the deubiquitinase DUB domain and was found to induce IFN. Therefore, it is proposed that coronaviruses use this mechanism to evade the host innate immune responses (Zheng *et al*. 2008).

The second domain located at the 5' end of Nsp3 is the acidic domain and it is of unknown function. However, it is consistently co-purified with nucleic acids and is thought to have nucleic acid binding activity (Imbert *et al*., 2008). Previous studies with influenza have indicated that electrostatic interactions from this type of domain may play a key role in ligand binding (Guu *et al*. 2008). Thus mutations in the acidic domain could lead to a decrease or increase in RNA binding efficiency, causing a change in virus pathogenicity. The ADP-ribose-1 phosphate domain is a conserved macro domain that is part of a protein module family associated with proteins that interact in the chromatin metabolism. Several other viruses also contain this type of domain and include Toroviruses, Alphaviruses, and Hepatitis E viruses (Egloff *et al*., 2006).

The crystal structure for the ADP-ribose-1 phosphate domain has been elucidated for IBV, SARS-CoV, and HCoV (Xu *et al*., 2009). Based on the structural analysis it was determined that the active site, which consist of four amino acids, was found to be highly conserved between different serotypes as well as across all groups of coronaviruses (Decroly *et al*., 2008, Piotrowski *et al*., 2009). A reverse genetics study, previously reported that mutations in the active site did not prevent the virus from replicating in cell culture (Putics *et al*., 2005). It is believed that this domain plays a role in host cell regulation as opposed to replication of the virus.

The transmembrane domain in Nsp3 is predicted to span the membrane four times, which is consistent with other studies that place Nsp3 on the same side of the endoplasmic reticulum (ER) membrane with the rest of the replication transcription complex (Imbert *et al*., 2008). The transmembrane domain in SARS-CoV is predicted to span the ER membrane four times, which orients the majority of the Nsp3 on the same side of the membrane as Nsp12 (Oostra *et al*., 2007). The Nsp3 transmembrane domains in MHV is composed of three transmembrane domains and only two domains span the ER membrane. The remaining transmembrane domain in located on the cytoplasm side (Oostra *et al*., 2008). The transmembrane domains are inserted into the ER membrane co-translationally (Imbert *et al*., 2008). The Y domain in Nsp3 is currently of unknown function.

Nonstructural protein 4 is a putative transmembrane domain that plays a role in the replication transcription complex. It is thought to provide a microenvironment for viral RNA synthesis. The Nsp4 in IBV is approximately 513 amino acids in length, which differs from MHV and SARS-CoV, which are 496 and 500 amino acids long, respectively (Oostra *et al*., 2007). This protein is predicted to transverse the membrane four times with both termini being located on the cytoplasmic side of the membrane. Enhanced Green Fluorescent Protein (EGFP) was fused to the Nsp4 in MHV and was found to associate with the ER and co-localize with Nsp8 and dsRNA during infection (Oostra *et al.,* 2007). Nonstructural protein 4 has not been characterized in IBV.

Non-structural protein 5 has been identified as a 33-kDa protein in IBV (Ng and Liu, 2002). It is the protease responsible for cleaving the majority of Nsps as well as functional intermediates of Nsps. The enzyme is functional via its own autolytic cleavage from polyproteins1a and 1ab

(Bartlam *et al*., 2007). This protein has been shown to co-sediment with the replication transcription complex and is resistant to protease treatment. These findings suggest that Nsp5 is surrounded by membranes (Van Hemert *et al*., 2008). Non-structural protein 5 cleaves the polyprotein 1ab at 11 conserved sites involving Leu- Gln ‖-(Ser, Ala, Gly) (Anand *et al.*, 2003). The HCoV Mpro crystal structure was elucidated and showed this protein is comprised of three domains, I= residues 8-99, II= residues 100-183, and III= residues 200-300. Located between domain I and II is the substrate binding site and domain III is connected to domain II by a loop made up of residues 184-199 (Anand *et al*., 2003). The architecture of these two domains are six stranded anti-parallel beta barrels resembling the molecular structure of chymotrypsin. Nonstructural protein 6 is a hydrophobic membrane spanning protein. In MHV and SARS-CoV, six of the seven hydrophobic domains span the ER membrane. All seven domains appear to have a functional signal sequence and membrane anchors. It has not been determined which six of the seven domains serve as the transmembrane domains. It is thought that the exposed cytoplasmic hydrophobic domain may interact with the replication transcription complex (Oostra *et al.,* 2008). Thus, Nsp6 also plays a role in scaffolding the replication transcription complex. Nonstructural protein 7 has RNA binding activity and forms a complex with Nsp8 that resembles a processivity factor for the RNA-dependent RNA-polymerase (RdRp) (Zhai *et al*., 2005). Eight copies of the Nsp7 join 8 copies of the Nsp8 to form a hollow cylinder that is large enough to encircle nucleic acid (Bartlam *et al*., 2007). Nonstructural protein 8 also has RNA binding activity and associates with Nsp7 as previously mentioned. It has been determined that there is a distant structural homology between Nsp8 and the catalytic portion of the RdRp in SARS-CoV (Zhai *et al*., 2005). Activity assays established that Nsp8 in SARS-CoV acts as a primase

distinguishing between specific sequences in the genome, in order to catalyze with low fidelity short oligonucleotides that are used as primers for the primer-dependent viral polymerase (Bartlam *et al*., 2007).

The protein crystal structure of Nsp9 was determined by two groups in 2004 (Bartlam *et al*., 2007).Based on its structure, this protein is thought to function as a single stranded RNA binding protein. In IBV, it is likely that Nsp9 is a dimer and is approximately 111 amino acids long. By using an infectious clone system, Chen *et al.* (2008) showed that Nsp9 is a nucleic acid-binding protein and that the binding capabilities are not strictly sequence specific in SARS-CoV. The Nsp9 has also been shown to co-precipitate with Nsp5, (specifically Mpro in Nsp5), and Nsp 8 in MHV (Brockway *et al*., 2003). The absolute function of this protein has not been determined but research suggests that Nsp9 plays a role in the replication transcription complex by stabilizing template RNA strands during replication and protecting them against nucleases in the host cell.

Non-structural protein 10 is predicted to from dimers due to multiple conserved cysteine residues. This protein is 16kDa in length and was found to increase in accumulation when epidermal growth factor (EGF) was added to the media; thus, suggesting that this protein plays a putative role in the EGF pathway (Ng *et al.,* 2002). Non-structural protein 10 co-localizes with Nsp 2, 3, 5, 7, 8, 9, 12, and 13; and also with N protein at the site of replication in MHV. Structural studies have shown that the functional form of this protein has a fold featuring two zinc fingers with C-(X)2-C-(X)5-H-(X)6-C and C-(X)2-C-(X)7-C-(X)-C (Bartlam *et al.,* 2007). Additionally, this protein forms a spherical dodecameric structure from 12 identical subunits. There is increasing evidence that zinc fingers are used to recognize RNA (Su *et al*., 2006).

However, one study demonstrated that Nsp10 has a weak affinity for both ssRNA and dsRNA (Bartlam *et al*., 2007).

Nonstructural protein 11, the shortest protein found in 1ab, is only 23 amino acids in length and functions in association with Nsp12. Based on sequence analysis, the RdRp has been predicted to be located in Nsp12 yet this has not yet been confirmed experimentally. This RdRp is involved in the replication and transcription of viral RNA. Coronaviruses replicate in the cytoplasm at a double membrane vesicle; however, the exact mechanism of how these replicase proteins are recruited and associated at cellular membranes is still unclear.

Nonstructural protein 13 is approximately 302 amino acids in length and consists of two domains; these include an N-terminal Zinc binding domain and a C-terminal helicase domain. This protein has multiple functions which include RNA and DNA duplex unwinding activities, NTPase activities, and RNA 5' triphosphatase activities. The latter of the three is predicted to be involved in the formation of the 5' cap structure on the viral RNA (Ivanov *et al*., 2004). A previous study showed that a single point mutation in this protein rendered the virus unable to replicate *in vitro* (Fang *et al*., 2007).

Nonstructural protein 14 possesses exoribonuclease degradation activity that degrades RNA by removing terminal nucleotides from either the 5'end or the 3' end. In MHV, strain A59 a single amino acid substitution in Nsp14 attenuated the virus in mice (Sperry *et al.,* 2005). Little is known about Nsp14 in IBV. Nonstructural protein 15 has been shown to function as an endoribonuclease. The crystal structure of Nsp15 from SARS-CoV has been determined (Bartlam *et al*., 2007); this protein consists of three domains each of which has a novel region. The hexameric form of this protein is known to bind RNA and the affinity can be altered by

Mn^{2+} concentrations. However, Nsp15 has not been well characterized in IBV (Bartlam *et al.,* 2007).

Nonstructural protein 16 is a methyltransferase. A recent study using FeCoV determined that this protein posses an AdoMet-dependent RNA methyltransferase capable of cap-1 formation (Decroly *et al*., 2008). It was determined that purified recombinant FeCoV Nsp16 has the capability of selectively binding to short capped RNA's. One prerequisite for binding is an N-7-methyl guanosine cap (Decroly *et al*., 2008). The amino acids present in the active site of this enzyme had previously been identified as residues K45, D129, K169, and E202; the K-D-K-E. A deletion in the Nsp16 gene blocked RNA synthesis; whereas, a single mutation in the catalytic triad decreased mRNA yield approximately 10% below wild type levels (Decroly *et al*., 2008). The RNA capping machinery in IBV has not been clearly identified.

Replication Cycle

Replication of IBV begins with attachment of the virus to cell receptors on the host cell; however, not all coronaviruses use the same cell receptor for gaining entry. For MHV, the receptor is a carcinoembryonic antigen (CEA) in the immunoglobulin (Ig) super family. The receptor for SARS-CoV was identified as angiotension-converting enzyme 2 (ACE2), which is expressed in the lungs, heart, kidney, and small intestine as well as other tissues (Li *et al*., 2003). The host cell receptor(s) for IBV have yet to be determined. The binding domain for IBV is a conformationally dependent epitope located on the S1 portion of the Spike glycoprotein, thus viruses that lack S1 are non-infectious. It was reported that cells become resistant to IBV infection after treatment with neuraminidase, which suggests that, sialic acid may be a

determinant for attachment (Winter *et al*., 2006). However, there is a broad distribution of alpha-2,3-linked sialic acid receptors in different cell types; therefore, it is unlikely that this type of sugar is the sole determinant (Rahman *et al*., 2009). One theory is that primary attachment to the cell surface may occur by Spike binding to sialic acid and then subsequently binding to a yet unidentified secondary receptor. It has also been reported that some serotypes such as Arkansas may use a co-receptor amino peptidase N (APN) in feline cells (Miguele *et al*., 2002). However, this has not been reproduced with other serotypes of IBV or in other cell lines.

After the virus attaches, it enters the host cell via fusion with the plasma membrane. Infectious bronchitis virus exhibits a slightly alkaline pH, which is thought to aid the virus in cell membrane fusion (Holmes, 1991). The exact mechanism of virus uncoating is still unknown. Once the viral RNA reaches the cytoplasm, the RdRp is translated by the host cell. The polymerase then recognizes the 3' end of the genome and transcribes a full-length negative sense copy, which serves as template for a 3' nested set of subgenomic mRNAs as well as the full-length positive sense genome. There are three models proposed for transcription of subgenomic mRNAs (sgmRNA) and they include amplification of virion-associated mRNAs, discontinuous transcription, and a leader primed transcription (Holmes, 1991).

Five to seven subgenomic mRNAs are synthesized in most coronaviruses; in the case of IBV six are produced. The most widely accepted theory for the process of transcribing sgmRNAs is leader primed transcription, which occurs when the newly transcribed leader sequence at the extreme 5'end of the genome dissociates, possibly with the polymerase and binds to complementary intergenic sequences (IS) located at the 5' end of all the open reading frames on the negative sense strand. The leader sequence serves as a primer, and the polymerase then

translates the sgmRNAs, which all have the same 5' leader sequence. A study determined that mutations within the intergenic sequence prevented the expression of subgenomic RNA's, thus those regions are essential for replication (Holmes, 1991).

Each messenger RNA codes for specific protein products. The full length viral RNA, also recognized as mRNA 1 produces the polyproteins1a and 1a/b, sgmRNA2 encodes the spike protein, sgmRNA3 encodes the envelope (E) protein, 3A, and 3B, sgmRNA4 produces the membrane (M) protein, sgmRNA5 produces 5A and 5B, and sgmRNA6 produces the nucleocapsid (N) protein. The polyproteins 1a and 1a/b are postranslationally cleaved and associate to form the replication transcription complex (RTC). This complex is thought to shield and protect the viral RNA in the cell during replication (Van Hemert *et al.*, 2008).

Assembly occurs when the N protein binds to the viral genomic RNA at a specific sequence, which is approximately 61 nucleotides long and located at the end of the 1a/b gene (Fosmire *et al.*, 1992), forming helical nucleocapsids. The N protein is translated on free polysomes and plays a major role in viral RNA synthesis and virus budding. Previous studies have shown the N protein to act as an IFN antagonist and capable of RNA binding (Devaraj *et al.*, 2007). Once the nucleotide binds the gRNA, it then interacts with the membrane protein at cellular membranes. The membrane protein is also responsible for recruiting the spike protein. The nucleocapsid and membrane protein form a complex then the envelope protein is recruited (Ye R, 2004). The virions are assembled at the Golgi ER complex where they bud into the Golgi complex and acquire their envelope (Cavanagh D., 2008). Coronaviruses exit the cell through exocytic vesicles and virus replication and budding through the cellular membranes eventually leads to

cell rupture thus releasing virions (Graham R.L, 2008). New viruses start to appear as early as 3 hours after infection (Cavanagh D., 2008).

References

Albassam MA, W., R. W., Thacker, H. L. (1986) Comparison of the nephropathogenicity of four strains of infectious bronchitis virus. . Avian Diseases 30, 468-76.

Alvarado, I.R., Villegas, P., El-Attrache, J. and Jackwood, M.W. (2006) Detection of Massachusetts and Arkansas serotypes of infectious bronchitis virus in broilers. Avian Dis 50(2), 292-7.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R. and Hilgenfeld, R. (2003) Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. Science 300(5626), 1763-1767.

Anderson, R. and Wong, F. (1993) Membrane and Phospholipid Binding by Murine Coronaviral Nucleocapsid N Protein. Virology 194(1), 224-232.

Bartlam, M., Xu, Y. and Rao, Z. (2007) Structural proteomics of the SARS coronavirus: a model response to emerging infectious diseases. Journal of Structural and Functional Genomics 8(2), 85-97.

Bo Chen, S.F., James P. Tam, Ding Xiang Liu. (2008) Formation of stable homodimer via the C-terminal Alpha-helical domain of coronavirus nonstructural protein 9 is critical for its function in viral replication. Virology.

Brockway, S.M., Clay, C.T., Lu, X.T. and Denison, M.R. (2003) Characterization of the Expression, Intracellular Localization, and Replication Complex Association of the

Putative Mouse Hepatitis Virus RNA-Dependent RNA Polymerase. J. Virol. 77(19), 10515-10527.

Casais, R., Thiel, V., Siddell, S.G., Cavanagh, D. and Britton, P. (2001) Reverse genetics system for the avian coronavirus infectious bronchitis virus. J Virol 75(24), 12359-69.

Cavanagh, D. (2003) Severe acute respiratory syndrome vaccine development: experiences of vaccination against avian infectious bronchitis coronavirus. Avian Pathol 32(6), 567-82.

Cavanagh D., G.J. (2008) Infectious Bronchitis Virus. In Saif YM, editor. Diseases of Poultry (12th Edition), 117-135.

Cavanagh, D. (2007) Coronavirus avian infectious bronchitis virus. Veterinary Research 38(2), 281-297.

Cavanagh, D., Mawditt, K., Welchman Dde, B., Britton, P. and Gough, R.E. (2002) Coronaviruses from pheasants (Phasianus colchicus) are genetically closely related to coronaviruses of domestic fowl (infectious bronchitis virus) and turkeys. Avian Pathol 31(1), 81-93.

Cavanagh, D., Casais, R., Armesto, M., Hodgson, T., Izadkhasti, S., Davies, M., Lin, F., Tarpey, I. and Britton, P. (2007) Manipulation of the infectious bronchitis coronavirus genome for vaccine development and analysis of the accessory proteins. Vaccine 25(30), 5558-5562.

Decroly, E., Imbert, I., Coutard, B., Bouvet, M., Selisko, B., Alvarez, K., Gorbalenya, A.E., Snijder, E.J. and Canard, B. (2008) Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. J. Virol., JVI.00407-08.

Devaraj, S.G., Wang, N., Chen, Z., Chen, Z., Tseng, M., Barretto, N., Lin, R., Peters, C.J., Tseng, C.-T.K., Baker, S.C. and Li, K. (2007) Regulation of IRF-3-dependent Innate Immunity by the Papain-like Protease Domain of the Severe Acute Respiratory Syndrome Coronavirus. J. Biol. Chem. 282(44), 32208-32221.

Egloff, M.-P., Malet, H., Putics, A., Heinonen, M., Dutartre, H., Frangeul, A., Gruez, A., Campanacci, V., Cambillau, C., Ziebuhr, J., Ahola, T. and Canard, B. (2006) Structural and Functional Basis for ADP-Ribose and Poly(ADP-Ribose) Binding by Viral Macro Domains. J. Virol. 80(17), 8493-8502.

Fabricant, J. (1998) The early history of infectious bronchitis. Avian Dis 42(4), 648-50.

Fang, S., Chen, B., Tay, F.P., Ng, B.S. and Liu, D.X. (2007) An arginine-to-proline mutation in a domain with undefined functions within the helicase protein (Nsp13) is lethal to the coronavirus infectious bronchitis virus in cultured cells. Virology 358(1), 136-47.

Fosmire, J.A., Hwang, K. and Makino, S. (1992) Identification and characterization of a coronavirus packaging signal. J Virol 66(6), 3522-30.

Gelb, J., Jr., C.L. Keeler, Jr., W.A. Nix, J.K. Rosenberger, and S.S. Cloud. (1997) Antigenic ans S-1 Genomic characterization of the Delaware variant serotype of infectious bronchitis virus. Avian Dis 41, 661-669.

Gelb JJ, J.M. (1998) Infectious Bronchitis Virus. In:Swayne DE, Glisson JR, Jackwood MW, Pearson JE, Reed WM, editors. A laboratory manual for the isolation and identification of avian pathogens., pp. 169-74. 4th ed. vols. Kennett Square, PA: American Association of Avian Pathologists.

Goodwin, M.A., J. Brown, E. C. Player, W. L. Steffens, D. Hermes, and M. A. Dekich. (1995) Fringed membranous particles and viruses in faeces from healthy turkey poults and from poults

with putative poult enteritis complex/spiking mortality. . Avian Pathology 24, 497-505.

Goren, E. (1978) Observations on experimental infection of chickens with Escherichia coli. . Avian Pathology 7, 213-224.

Graham R.L, S.J.S., Eckerle L.D, Sims A.C, Denison M.R. (2008) SARS coronavirus replicase proteins in pathogenesis. Virus Research 133(1), 88-100.

Guu TS, D.L., Wittung-Stafshede P,Tao YJ. (2008) Mapping the domain structure of the influenza A virus polymerase acidic protein (PA) and its interaction with the basic protein 1 (PB1) subunit. Virology 379(1), 1096-0341.

Hodgson, T., Britton, P,Cavanagh, D. (2006) Neither the RNA nor the proteins of open reading frames 3a and 3b of the coronavirus infectious bronchitis virus are essential for replication. J Virol 80(1), 296-305.

Hodgson, T., Casais, R., Dove, B., Britton, P. and Cavanagh, D. (2004) Recombinant Infectious Bronchitis Coronavirus Beaudette with the Spike Protein Gene of the Pathogenic M41 Strain Remains Attenuated but Induces Protective Immunity. J. Virol. 78(24), 13804-13811.

Hofstad, M.S., Yoder, H. W. (1996) Avain infectious bronchitis virus distribution in tissues of chicks. Avian Dis 10, 230-239.

Holmes, K. (1991) Coronaviridae and their replication. In: Fundamental virology, 2nd ed., 471-86.

Imbert, I., Snijder, E.J., Dimitrova, M., Guillemot, J.-C., Lécine, P. and Canard, B. (2008) The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. Virus Research 133(2), 136-148.

Ivanov, K.A., Thiel, V., Dobbe, J.C., van der Meer, Y., Snijder, E.J. and Ziebuhr, J. (2004) Multiple Enzymatic Activities Associated with Severe Acute Respiratory Syndrome Coronavirus Helicase. J. Virol. 78(11), 5619-5632.

Lee, C.W., Hilt, D.A. and Jackwood, M.W. (2003) Typing of field isolates of infectious bronchitis virus based on the sequence of the hypervariable region in the S1 gene. J Vet Diagn Invest 15(4), 344-8.

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. and Farzan, M. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426(6965), 450-454.

Matthijs, M.G., van Eck, J.H., de Wit, J.J., Bouma, A. and Stegeman, J.A. (2005) Effect of IBV-H120 vaccination in broilers on colibacillosis susceptibility after infection with a virulent Massachusetts-type IBV strain. Avian Dis 49(4), 540-5.

Ng, L.F.P, Liu, D.X. (2002) Membrane Association and Dimerization of a Cysteine-Rich, 16-Kilodalton Polypeptide Released from the C-Terminal Region of the Coronavirus Infectious Bronchitis Virus 1a Polyprotein. J. Virol. 76(12), 6257-6267.

Oostra, M., Hagemeijer, M.C., van Gent, M., Bekker, C.P.J., te Lintelo, E.G., Rottier, P.J.M. and de Haan, C.A.M. (2008) Topology and Membrane Anchoring of the Coronavirus

Replication Complex: Not All Hydrophobic Domains of nsp3 and nsp6 are Membrane Spanning. J. Virol. 82(24), 12392-12405.

Oostra, M., te Lintelo, E.G., Deijs, M., Verheije, M.H., Rottier, P.J.M. and de Haan, C.A.M. (2007) Localization and Membrane Topology of Coronavirus Nonstructural Protein 4: Involvement of the Early Secretory Pathway in Replication. J. Virol. 81(22), 12323-12336.

Pei, J., Sekellick, M.J., Marcus, P.I., Choi, I.S. and Collisson, E.W. (2001) Chicken interferon type I inhibits infectious bronchitis virus replication and associated respiratory illness. J Interferon Cytokine Res 21(12), 1071-7.

Piotrowski Y., Van der Zanden, B., Snijder, E.J., Gorbalenya A.E., Hilgenfeld R.,. (2009) Crystal structures of the X-domains of a Group-1 and a Group-3 coronavirus reveal that ADP-ribose-binding may not be a conserved property. Protein Science 18(1), 6-16.

Putics, A., Filipowicz, W., Hall, J., Gorbalenya, A.E. Ziebuhr, J. (2005) ADP-Ribose-1"-Monophosphatase: a Conserved Coronavirus Enzyme That Is Dispensable for Viral Replication in Tissue Culture. J. Virol. 79(20), 12721-12731.

Rahman, S.A.E., El-Kenawy, A.A., Neumann, U., Herrler, G. and Winter, C. (2009) Comparative analysis of the sialic acid binding activity and the tropism for the respiratory epithelium of four different strains of avian infectious bronchitis virus. Avian Pathology 38(1), 41 - 45.

Raj GD, J.R. (1997) Infectious bronchitis virus: Immunopathogenesis of infection in the chicken. Avian Pathology 26(4), 677-706.

Sperry, S.M., Kazi, L., Graham, R.L., Baric, R.S., Weiss, S.R. and Denison, M.R. (2005) Single-Amino-Acid Substitutions in Open Reading Frame (ORF) 1b-nsp14 and ORF 2a Proteins of the Coronavirus Mouse Hepatitis Virus Are Attenuating in Mice. J. Virol. 79(6), 3391-3400.

Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R., Joachimiak, A., Zhang, X.C., Bartlam, M. and Rao, Z. (2006) Dodecamer Structure of Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein nsp10. J. Virol. 80(16), 7902-7908.

Van Hemert, M.J., Van den Worm, S.H.E., Knoops, K.V., Mommaas, A.M., Gorbalenya, A.E. and Snijder, E.J. (2008) SARS-Coronavirus Replication/Transcription Complexes Are Membrane-Protected and Need a Host Factor for Activity In Vitro. PLoS Pathogens 4(5), e1000054.

Van Roekel, H., Olesiuk, O.M. and Beninato, L.P. (1958) Symposium on chronic respiratory diseases of poultry. III. Epizootiology of chronic respiratory diseases in chickens. Am J Vet Res 19(71), 453-63.

Van Roeckel, H., M. K. Clarke, K. L. Bullis, O. M. Oleisuk, and F. G. Sperling. (1951) Infectious bronchitis. Am. J. Vet. Res. 12, 140-146.

Vennema H., G., 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. 8, 2020–2028

Winter, C., Schwegmann-Wessels, C., Cavanagh, D., Neumann, U. and Herrler, G. (2006) Sialic acid is a receptor determinant for infection of cells by avian Infectious bronchitis virus. J Gen Virol 87(5), 1209-1216.

Woo, P.C.Y., Lau, S.K.P., Lam, C.S.F., Lai, K.K.Y., Huang, Y., Lee, P., Luk, G.S.M., Dyrting, K.C., Chan, K.-H. and Yuen, K.-Y. (2009) Comparative Analysis of Complete Genome Sequences of Three Avian Coronaviruses Reveals a Novel Group 3c Coronavirus. J. Virol. 83(2), 908-917.

Xu, Y., Cong, L., Chen, C., Wei, L., Zhao, Q., Xu, X., Ma, Y., Bartlam, M. and Rao, Z. (2009) Crystal Structures of Two Coronavirus ADP-Ribose-1"-Monophosphatases and Their Complexes with ADP-Ribose: a Systematic Structural Analysis of the Viral ADRP Domain. J. Virol. 83(2), 1083-1092.

Ye R, M.-M.C., Masters PS. . (2004) Genetic Analysis of Determinants for SpikeGlycoprotein Assembly into Murine Coronavirus Virions: Distinct Roles for Charge-Rich and Cysteine-Rich Regions of the Endodomain. . J Virol 2004 78(18), 9904-17.

Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M. and Rao, Z. (2005) Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. Nat Struct Mol Biol 12(11), 980-986.

Zheng, D., Chen, G., Guo, B., Cheng, G., Tang, H. (2008) PLP2, a potent deubiquitinase from murine hepatitis virus, strongly inhibits cellular type I interferon production. Cell Res 18(11), 1105-1113.

Chapter 3

Method and Materials

Viruses

The Arkansas-Delmarva Poultry Industry (Ark-DPI) and Massachusetts 41 (Mass41) viruses were graciously given to our lab by Dr. Jack Gelb (University of Delaware, Newark, DE). The Georgia 98 (GA98) virus was isolated in our laboratory in 1998 (Lee and Jackwood, 2001). The pathogenic viruses were purified by centrifugation (5276.96 xg) from allanotic fluid and stored at -80 °C. The attenuated Ark-DPI, Mass41, and GA98 progeny viruses were obtained from commercial vaccines. The vaccine progeny of the pathogenic strains were attenuated by at least 100 passes in SPF chicken embryos. The vaccines were resuspended in 10ml of diethylpyrocarbonate (DEPC) treated water, and stored at -80°C.

Viral RNA extraction and RT-PCR

The pathogenic viruses and resuspended vaccines were filtered through a 0.8µm filter then through a 0.22µm filter (Millipore, Billerica, MA) prior to RNA extraction. Viral RNA was purified using the High Pure RNA Isolation Kit (Roche Diagnostic Corporation, Foster City, CA) according to the manufacturer's recommendation. The purified RNA was resuspended in diethylpyrocarbonate (DEPC) treated water. The Reverse Transcription was performed using

Woo, P.C.Y., Lau, S.K.P., Lam, C.S.F., Lai, K.K.Y., Huang, Y., Lee, P., Luk, G.S.M., Dyrting, K.C., Chan, K.-H. and Yuen, K.-Y. (2009) Comparative Analysis of Complete Genome Sequences of Three Avian Coronaviruses Reveals a Novel Group 3c Coronavirus. J. Virol. 83(2), 908-917.

Xu, Y., Cong, L., Chen, C., Wei, L., Zhao, Q., Xu, X., Ma, Y., Bartlam, M. and Rao, Z. (2009) Crystal Structures of Two Coronavirus ADP-Ribose-1"-Monophosphatases and Their Complexes with ADP-Ribose: a Systematic Structural Analysis of the Viral ADRP Domain. J. Virol. 83(2), 1083-1092.

Ye R, M.-M.C., Masters PS. . (2004) Genetic Analysis of Determinants for SpikeGlycoprotein Assembly into Murine Coronavirus Virions: Distinct Roles for Charge-Rich and Cysteine-Rich Regions of the Endodomain. . J Virol 2004 78(18), 9904-17.

Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M. and Rao, Z. (2005) Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. Nat Struct Mol Biol 12(11), 980-986.

Zheng, D., Chen, G., Guo, B., Cheng, G., Tang, H. (2008) PLP2, a potent deubiquitinase from murine hepatitis virus, strongly inhibits cellular type I interferon production. Cell Res 18(11), 1105-1113.

they were incubated at 37°C in 480µl of SOC medium for one hour on a rotary shaker. The cultures were mixed with 70% glycerol and frozen at -80°C until plated on q-trays (Genetix, Boston, MA) containing LB, Kanamycin, and agar. The Q-trays were pre-warmed at 37°C before the entire culture (approximately 500ul) was spread on the plates and incubated overnight at 37°C.

Nucleic acid sequencing

The plasmids for the cDNA libraries were isolated using an alkaline lysis method incorporating both Hydra and Tomtek robots (http://www.intl-pag.org/11/abstracts/P2c_P116_XI.html) to eliminate centrifugation and/or resuspension steps. The cDNA were sequenced with the sequencing ready reaction kit with amplitaq DNA polymerase FS. Cycle sequencing reactions were performed using the BigDye™ Terminator® Cycle Sequencing Kit Version 3.1 (Applied Biosystems, Foster City, CA) and MJ Research thermocyclers (Watertown, MA). Finished reactions were filtered through Sephadex filter plates into Perkin-Elmer MicroAmp Optical 96-well plates. The amplicons were sequenced directly using the M13F and M13R primers located on the pCR-XL-Topo cloning vector (Invitrogen). To perform sequence data processing we used a data management pipeline anchored by a SQL-based relational database (RDB) that automatically linked sample and experimental information with each chromatogram file. Chromatogram files were archived in the RDB then transferred to a PC for processing using DNA Star (DNASTAR Lasergene7, DNASTAR Inc., Madison, WI). Low quality segments and

vector sequence was trimmed from the ends of each sequence and remove from further analysis. The accuracy of the sequence is insured by generating data in both the 5' and 3' directions.

Specific primers designed to fill in sequence gaps

After the genomes were assembled, gaps and areas with less than 10X coverage were identified and specific primers were synthesized (IDT) for amplification and sequencing of the ambiguous areas. Specific primers are presented in the supplementary data. The RT-PCR was conducted as described above, and the reaction conditions were 42°C for 60 minutes, 95°C for five minutes, then ten 10 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 68°C for 90 seconds, followed by 25 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 68°C for 90sec + 5sec/cycle added. The final elongation step was 68°C for 7 minutes, then the reaction was cooled to 4°C. The PCR product was labeled for sequencing using the ABI Prism BigDye Terminator v3.0 (Applied Biosystems, Foster City Ca) and the specific primers that were used for amplification were also used for sequencing at a concentration of 15ng. The amount of cDNA added to the reaction ranged from 20-30ng and each amplicon was sequenced on an ABI sequencer 3730 (Ramsey, Minnesota) in the forward and reverse direction and analyzed on an ABI sequencer.

Sequencing alignment and nonstructural protein comparison

The DNASTAR program (Madison WI) was used for the alignment, assembly, editing, and annotation. The sequences were aligned using SeqMan in the DNASTAR program. The alignment was performed using a Massachusetts virus (AY851295) backbone. The alignment of Nsp genes was performed using a Beudette virus (P0C6V4) backbone. Vector sequence was

removed from the sequence reads before alignment. Due to the relatively low sequence identity the assembling parameter for minimum match percent was lowered to 40%. Full length genomes were uploaded to NCBI ORF finder to identify open reading frames. Next, alignments of nucleotide and amino acid sequences for the entire genome as well as specific areas of the genome were performed using MegAlign with the ClustalW program in the DNASTAR program.

Chapter 4

Results

Full length Genome Comparison

A total of six Infectious Bronchitis Viruses, three sets of homologous IBV strains, were sequenced and compared with the full length genome. The genomic size of the of the following viruses: Ark DPI pathogenic, Ark vaccine, Ga98 pathogenic, Ga98 vaccine, Mass41 pathogenic, Mass vaccine was 27,659 nt, 27,660 nt, 27,601, 27,433 nt, 27,478 nt, and 27,452 nt, respectively. A Phylogenetic analysis of the full length genomes shows genomes of the same serotype are more closely related (Fig. 1). The Ark pathogenic strain has 98.9% sequence identity with the vaccine strain, the Ga98 pathogenic strain has 97.2% sequence identity with the Ga98 vaccine strain and the Mass41 pathogenic strain has 92.2% sequence identity with the Mass41 vaccine strain (Table 1). All strains have a G+C content of 38%. An amino acid comparison was done between 1a, 1a/b, and structural proteins between the homologous pathogenic strain and attenuated strains of Ark DPI, Georgia98, and Massachusetts 41(Table 2). There were variations between the structural proteins for all viruses. The genome organization of all six IBV strains are shown in a schematic diagram in (Fig. 2), there is a nonstructural protein 4b that is located after the membrane protein. The 4b protein is 94 amino acids long and was found in all viruses sequenced however has not been previously recognized in IBV. A BLAST search of this protein revealed 96% sequence identity with Turkey Coronavirus. The sequence alignment of 4b shows that it is highly conserved among all six strains of IBV (Fig. 3). There are a total of 19 amino

acid mutations between the six strains in the accessory protein. These mutations occur at residues 9, 21, 22, 25, 37, 40, 47, 51, 55, 56, 57, 64, 68, 73, 77, 79, 87, 89, and 90. Turkey Coronavirus accession number EU022526.1 was used in the comparison.

Alignment comparison of Leader Sequence

The nucleotide sequence alignment of the 5'end verified that the leader sequence had some variability while the leader junction sequence, nucleotides 57-64 (CTTAACAA) was found to be highly conserved.

Non-structural Protein Alignments

The predicted ORF 1a and ORF 1a/b contained 11,831 and 19,789 nt, respectively for Ark DPI pathogenic the exact number of nucleotides varies between strains. ORF 1a/b is created from a slippery sequence (UUUAAAC) that causes a -1 frameshift, this sequence was found with some variability between open reading frames in all viruses. The analysis of the predicted ORF 1a shows the majority of amino acid differences between the attenuated and pathogenic strains occurred within the Nsp3, for all three sets of viruses. An eleven amino acid repeat occurred in Nsp3 in Ga98 pass 12 pathogenic, Ga98 vaccine, and Mass41 pathogenic from amino acid 116-125, however was not present in either strains of Arkansas. The insertion appears to be a repeat, (D-A-E-E-C-D-T-D-S-G/A) from amino acid position 125-135 (Fig. 4). This repeat occurs within the putative acidic domain. The 150 amino acids following the acidic domain varied among serotypes. However, the ADRP remained relatively conserved with only 16 amino acid differences between all strains. The active site for this protein was highly conserved at residues

Asn373, His 378, Gly 379, and Gly 380. The catalytic triad in PL2 remained conserved in all strains (Fig. 4). The Phylogenetic trees were created for the individual nsps using a Mouse Hepatitis Virus as the outlier and showed the possibility of recombination occurring in Figures 5-18. Based on Phylogenetic analysis the nsp 3 in Massachusetts pathogenic is more closely related to the Georgia serotype than the Massachusetts vaccine (Fig. 6).

The comparison of Ark DPI pathogenic and Ark vaccine resulted in 49 Amino Acid changes in the two polyproteins 1a and 1a/b. Of these, mutations, forty eight were located in nonstructural protein 3. Of the forty eight mutations, there were eight acidic amino acids in the pathogenic strain that differed from attenuated strain: two of these acidic residues, when mutated, remained negatively charged, four became positively charged and two become neutral. The polarities of the amino acid mutations were compared and the results indicate that five mutations in the pathogenic strain changed from hydrophobic to hydrophilic while five mutations changed from hydrophilic to hydrophobic. Ten mutations were hydrophobic to hydrophobic and twelve mutations were hydrophilic to hydrophilic). One amino acid mutation occurred in nonstructural protein 4 at position 299 mutating a Threonine in the pathogenic strain to a Histidine in the attenuated Strain.

The majority of differences between strains in the Georgia 98 serotype are located in nonstructural protein 3 totaling to 78 mutations which is approximately 43.8% of the mutations that occurred in polyproteins 1a and 1a/b. Fifty-three of these mutations occurred within or around the acidic domain. There was a 9 amino acid insertion in the attenuated Georgia strain when compared to the pathogenic strain. The remaining mutations were scattered over the other

domains, 7 were within the ADP-ribose-1 phosphate dephosphorylation domain and 9 were within the Papain like two Domain (PLp2). Unlike Arkansas, Georgia had multiple mutations throughout the polyproteins 1a and 1a/b.

The results for the Massachusetts virus varied among all the nonstructural proteins but 46% of the amino acid changes occurred in nsp3. The total number of insertions/deletions was 10 and 161 mutations occurred in nsp3. All ten of the deletions occurred within the acidic domain in the vaccine strain when compared to the pathogenic strain. Fourteen of these mutations occurred in the ADP-ribose-1 phosphate dephosphorylation domain and 38 occurred in the PL2 domain. Non-structural protein four had 48 mutations totaling to an 8% change in the overall protein. This was the second most variable region except for nsp 11 but this protein is only composed of 23 amino acids so the percentage is skewed in comparison to the rest of the proteins.

A BLAST domain prediction was performed with nsp 3 and two domains were detected. They are the high affinity ADP- ribose binding domain and the viral Papain like protease. A hydrophobicity plot using the Hopp and Woods scale of this nonstructural protein predicts a transmembrane domain at the 3' end near residues 1,200 (Fig. 19) (http://www.expasy). The transmembrane domains are predicted to span the membrane four times the exact location of these membranes are predicted to be at residues 1054-1073, 1080-1102, 1173-1195, 1215-1237. This prediction is based on the SMART protein prediction (Schultz *et al.*).

Structural Protein Alignments

Arkansas, Georgia, and Massachusetts had some variability in the spike protein, 1, 48, and 19 amino acid changes respectively. The envelope protein (109 a.a), the membrane (225 a.a), and the nucleocapsid (409 a.a) had some variability in the Ga98 and Mass strains. All of the structural proteins mutations for each virus are listed in Table 2.

Comparison of 3' UTR

The Ark DPI pathogenic and Ark vaccine strain has a 5' UTR consisting of 529 nucleotides and a 3' UTR consisting of 509 nucleotides. The predicted secondary structure of the 3' UTR is shown in Fig. 20 with several stem loops that are located downstream of the N gene for the Arkansas serotype.

Chapter 5

Discussion

The first two-thirds of the Infectious Bronchitis Virus's genome are composed of two open reading frames. These two polyproteins are post-translationally cleaved into 15 nonstructural proteins (nsps), and are responsible for the replication transcription complex. This region is important because the spike glycoprotein does not contribute solely to pathogenicity. In this paper we examine a total of six viruses from three different serotypes. The full length genome comparisons were made between three sets of two viruses; one a pathogenic parent virus and the other an attenuated progeny virus. The results coincide with a previous study that analyzed SARS-CoV and found that the majority of mutations were located in the polyprotein 1a specifically in Nsp 3 and Nsp 4 (Graham *et al*., 2008). Other studies have also concluded that the PL2 domain is divergent among strains; whereas, the Mpro was found to be highly conserved in IBV (Mondal and Cardona, 2004). Non-structural protein 3 is approximately 1600 amino acids in length, composed of several different domains. This protein is important in the establishment of viral infection and in viral replication.

One domain is thought to be a de-ubuquitinating protease which removes ubiquitin from proteins which putatively disrupt host cell process (Lindner *et al*., 2005). Mutations in this domain could also render the virus less/more virulent. The second domain located on the 5' end of Nsp3 is the acidic domain and it is of unknown function. However, because Nsp3 is consistently co-purified

with nucleic acids, it is thought that this protein possess nucleic acid binding activity (Imbert *et al*., 2008). Previous studies with other organisms indicate that electrostatic interactions from this type of domain play a key role in ligand binding (Guu *et al*., 2008). Thus these mutations could lead to a decrease in binding efficiency causing a decrease in replication efficiency rendering the virus attenuated. Other viruses have been shown to have an acidic domain that plays a role in replication. For example, the Influenza A Virus has a polymerase acidic protein (PA) which plays a critical role in assembly, catalysis and nuclear localization of the polymerase. It has been shown that PA is indispensable for the transcription and replication activity of the influenza polymerase (Guu *et al*., 2008). Based on our data, it is likely that the acidic domain plays a role in attenuation but the exact function of the amino acids in these domains is unclear.

The ADP-ribose-1 phosphate in other animals has been shown to play a key role in the ADP ribosylation, a posttranslational protein modification involved in DNA damage repair and transcription regulation (Xu *et al*., 2009). The ADRP domain has sequence homology with a characterized phosphatase in *Saccharomces cerevisiae*. Therefore, it is thought that this domain removes the phosphate group of ADP-ribose-1 phosphate in the tRNA splicing pathway (Bartlam *et al*., 2007). Few viruses contain an ADRP domain, to date only three have been shown; rubella virus, alphaviruses, and hepatitis E virus. The crystal structure of this domain has been determined for SARS and the active site was found to be conserved however, the substrate binding pocket varied in structure between strains (Xu *et al*., 2009). A reverse genetic study showed that mutations in the active site had no effect on virus replication in cell culture; therefore it is thought that this domain plays a role in regulation rather than replication (Putics *et*

al., 2005). However, all of the residues at the active site were found to be conserved. Based on the crystal structural this domain has six alpha helices and six beta strains. The active site residue Asn is located on Beta sheet 3 with three remaining residues in the active site located on alpha helices 2 (Xu *et al*., 2009). A structural study on the SARS ADRP revealed that viral ADRP domains have a high affinity for P granule ribonucleoprotein complexes, PAR proteins that play a role in generating cell diversity during animal development (Egloff *et al*., 2006). Thus this domain could interact with host proteins to regulate cellular response. Due to the highly conserved active site and the results of the site directed mutagenesis studies this domain needs to be further studied to elucidate its exact role in virus-host interactions.

The PL2 domain is a papain-like protease that is responsible for the cleavage of Nsp2, 3, and 4 at a LXGG cleavage site (Anand *et al*., 2003). Most coronaviruses have two papain-like proteases however, in IBV PL1 is truncated and is non functional. In 2006, the structure of the PL2 domain was determined to be a "thumb-palm-finger" motif (Bartlam *et al*., 2007). This domain has also been shown to be a potent IFN antagonist by interacting with interferon regulatory factor 3 (IRF-3) and inhibits the phosphorylation and nuclear translocation of IRF-3 causing a disruption in the activation of type I IFN response through TLR3 or RAG-I. The TLR3 and RAG-I are responsible for recognizing dsRNA, which occurs during viral replication. By repressing the IFN immune response the virus is able to have a more productive infection. Deletions or insertions in this nonstructural protein could cause a conformational change affecting the efficiency of this IFN antagonist and thus decreasing the virulence of the virus. If

the immune system is able to activate transcription factors associated with dsRNA then the innate immune system can be more effective in clearing the virus.

The transmembrane domain plays an important scaffolding role for the replication transcription complex. The transmembrane domain is inserted into the endoplasmic reticulum membrane co-translationally (Imbert *et al*., 2008). Recently it has been shown in Murine Hepatitis Virus that this domain is important for membrane association of the replicase and joins PLP2 domain in the viral polyprotein processing activity. The transmembrane domain is predicted in SARS to span the membrane four times, which orients the majority of the Nsp3 on the same side of the membrane as Nsp12 (Oostra *et al*., 2007). This is not the case for the Nsp3 transmembrane domain in MHV, which was composed of three transmembrane domains with only two of these domains spanning the membrane. The remaining transmembrane domain in located on the cytoplasm side of the cell wall (Oostra *et al*., 2008). Based on the prediction modules, the Nsp3 transmembrane domain found in IBV spans the membrane four times, which is similar to SARS. Nsp3 and Nsp12 have been shown in SARS to interact along with Nsp7 (Imbert *et al*., 2008) thus underscoring the importance of the transmembrane domains. We constructed a hydrophobicity plot, which identified the location of the transmembrane domains in the region between amino acids 1,000 –1,300 for IBV (Fig. 6). The Y domain is currently of unknown function.
There is little known about the fifteen nonstructural proteins but they participate in the viral replication process thus it is of importance to identify at the amino acid level the role of the mutations in association with pathogenicity. Previous infectious clone studies using the SARS virus have indicated that the nonstructural proteins play an important role in host cell

modification. Knowing the functions of the individual proteins and identifying at the amino acid level differences, one can speculate as to how these changes alter the phenotype of the virus. The addition of these full genomic sequences will aid in understanding attenuation, replication processes, and viral molecular evolution of coronaviruses.

Chapter 6

Conclusion

The comparison of the attenuated and pathogenic strains revealed 48 amino acid differences in the Nsp3 for the Arkansas serotype, 78 for the Georgia Serotype, and 171 amino acid differences for the Massachusetts serotype. A more definitive conclusion can be drawn, than from previous reports, due to the fact that the three serotypes used for comparison in this study have a parent-progeny relationship. Based on our results, the differences exist within non-structural protein 3 which houses several biologically important domains. The attenuated virus was passed in chicken embryos and is lethal to the embryos while it does not cause disease in chickens, the natural host. Therefore, the mutations that occurred could be referred to as virulence factors as opposed to attenuation factors depending on the host. Because the vast majority of changes are located within Nsp3, this protein is significantly related to a virus mutating/adapting in chicken embryos; however, each amino acid position must be analyzed individually then in conjunction with one another before specific statements regarding attenuation can be adequately made.

References

Ammayappan, A., Upadhyay, C., Gelb, J. and Vakharia, V. Identification of sequence changes responsible for the attenuation of avian infectious bronchitis virus strain Arkansas DPI. Archives of Virology.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R. and Hilgenfeld, R. (2003) Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. Science 300(5626), 1763-1767.

Bartlam, M., Xu, Y. and Rao, Z. (2007) Structural proteomics of the SARS coronavirus: a model response to emerging infectious diseases. Journal of Structural and Functional Genomics 8(2), 85-97.

Bo Chen, S.F., James P. Tam, Ding Xiang Liu. (2008) Formation of stable homodimer via the C-terminal Alpha-helical domain of coronavirus nonstructural protein 9 is critical for its function in viral replication. Virology.

Cavanagh, D. (2007) Coronavirus avian infectious bronchitis virus. Vet Res 38(2), 281-97.

Cavanagh, D., Casais, R., Armesto, M., Hodgson, T., Izadkhasti, S., Davies, M., Lin, F., Tarpey, I. and Britton, P. (2007) Manipulation of the infectious bronchitis coronavirus genome for vaccine development and analysis of the accessory proteins. Vaccine 25(30), 5558-5562.

Egloff, M.-P., Malet, H., Putics, A., Heinonen, M., Dutartre, H., Frangeul, A., Gruez, A., Campanacci, V., Cambillau, C., Ziebuhr, J., Ahola, T. and Canard, B. (2006) Structural

and Functional Basis for ADP-Ribose and Poly(ADP-Ribose) Binding by Viral Macro Domains. J. Virol. 80(17), 8493-8502.

Graham R.L, S.J.S., Eckerle L.D, Sims A.C, Denison M.R. (2008) SARS coronavirus replicase proteins in pathogenesis. Virus Research 133(1), 88-100.

Hodgson, T., Casais, R., Dove, B., Britton, P. and Cavanagh, D. (2004) Recombinant infectious bronchitis coronavirus Beaudette with the spike protein gene of the pathogenic M41 strain remains attenuated but induces protective immunity. J Virol 78(24), 13804-11.

Imbert, I., Snijder, E.J., Dimitrova, M., Guillemot, J.-C., Lécine, P. and Canard, B. (2008) The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. Virus Research 133(2), 136-148.Lee, C.-W. and Jackwood, M.W. (2001) Origin and evolution of Georgia 98 (GA98), a new serotype of avian infectious bronchitis virus. Virus Research 80(1-2), 33-39.

Lindner, H.A., Fotouhi-Ardakani, N., Lytvyn, V., Lachance, P., Sulea, T. and Menard, R. (2005) The Papain-Like Protease from the Severe Acute Respiratory Syndrome Coronavirus Is a Deubiquitinating Enzyme. J. Virol. 79(24), 15199-15208.

Mondal, S.P. and Cardona, C.J. (2004) Comparison of four regions in the replicase gene of heterologous infectious bronchitis virus strains. Virology 324(1), 238-48.

Oostra, M., Hagemeijer, M.C., van Gent, M., Bekker, C.P.J., te Lintelo, E.G., Rottier, P.J.M. and de Haan, C.A.M. (2008) Topology and Membrane Anchoring of the Coronavirus Replication Complex: Not All Hydrophobic Domains of nsp3 and nsp6 Are Membrane Spanning. J. Virol. 82(24), 12392-12405.

Oostra, M., te Lintelo, E.G., Deijs, M., Verheije, M.H., Rottier, P.J.M. and de Haan, C.A.M. (2007) Localization and Membrane Topology of Coronavirus Nonstructural Protein 4: Involvement of the Early Secretory Pathway in Replication. J. Virol. 81(22), 12323-12336.

Putics, A., Filipowicz, W., Hall, J., Gorbalenya, A.E. and Ziebuhr, J. (2005) ADP-Ribose-1"-Monophosphatase: a Conserved Coronavirus Enzyme That Is Dispensable for Viral Replication in Tissue Culture. J. Virol. 79(20), 12721-12731.

Van Hemert, M.J., Van den Worm, S.H.E., Knoops, K.v., Mommaas, A.M., Gorbalenya, A.E. and Snijder, E.J. (2008) SARS-Coronavirus Replication/Transcription Complexes Are Membrane-Protected and Need a Host Factor for Activity In Vitro. PLoS Pathogens 4(5), e1000054.

Xu, Y., Cong, L., Chen, C., Wei, L., Zhao, Q., Xu, X., Ma, Y., Bartlam, M. and Rao, Z. (2009) Crystal Structures of Two Coronavirus ADP-Ribose-1"-Monophosphatases and Their Complexes with ADP-Ribose: a Systematic Structural Analysis of the Viral ADRP Domain. J. Virol. 83(2), 1083-1092.

Yvonne Piotrowski, G.H., A. Linda Boomaars-van der Zanden, Eric J. Snijder, Alexander E. Gorbalenya, Rolf Hilgenfeld,. (2009) Crystal structures of the X-domains of a Group-1 and a Group-3 coronavirus reveal that ADP-ribose-binding may not be a conserved property. Protein Science 18(1), 6-16.

Ziebuhr, J., Thiel, V. and Gorbalenya, A.E. (2001) The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond. J Biol Chem 276(35), 33220-32.

Table

Table 1: Percent Sequence Identity between Strains

Percent identity between strains for full length genome. Sequence Distances of nucleotide alignment using Megalign ClustalW (Slow/Accurate, IUB). The percent divergence in lower triangle and the percent similarity is in the upper triangle.

	ArkDPI	**Ark Vac**	**MassPath**	**Mass Vac**	**Ga98Path**	**Ga Vac**
ArkDPI	***	98.9	90.6	92.9	92.3	93.1
Ark Vac	0.4	***	90.8	93.1	92.4	93.4
MassPath	9.3	9.7	***	92.2	88.0	89.1
Mass Vac	5.7	5.9	6.9	***	90.9	91.3
Ga98Path	7.4	7.9	13.0	8.3	***	97.2
Ga Vac	6.6	6.8	11.8	7.9	2.7	***

Table 2: Amino Acid Comparisons between ORF's

Amino Acid Differences between parent progeny for Arkansas, Georgia, and Massachusetts serotypes for polymerase genes and structural proteins.

Non Structural Protein[a](Amino Acid's at cleavage site)	A.A mutations btw. Ark strains	A.A mutations btw. GA strains	A.A mutations btw. Mass strains
Nsp2 (MASSLK), (VCKAG)	0	21	45
Nsp3-PL2 (GKTVTF), (EKKAG)	48	78	171
Nsp4 (VISGIF), (GVSRLQ)	1	8	43
Nsp5-3CLP (AGFKKL), (GGVRLQ)	0	24	15
Nsp6 (SSFVRK), (PIATVQ)	0	1	21
Nsp7 (SKGHET), (PKPFVQ)	0	1	1
Nsp8 (SVTQEF), (VDAVLQ)	0	2	5
Nsp9 (NELMPH), (NVVVLQ)	0	0	3
Nsp10 (SKGHET), (KPFVQS)	0	1	1
Nsp11 (SAAGAS), (VAVRLG)	0	4	4
Nsp12 (VAVRLG), (CYPQLQ)	0	4	9
Nsp13 (SAWTCG), (PFVCTM)	0	19	11
Nsp14 (TGLFKI), (FSALQS)	0	2	8
Nsp15 (IDNIAY), (YPQLQS)	0	4	9
Nsp16 (AWTCGY), (KTDLVV)	0	9	9
Structural Proteins			
Spike	1	48	27
Envelope	0	9	6
Membrane	0	11	3
Nucleocapsid	0	25	19

Figures

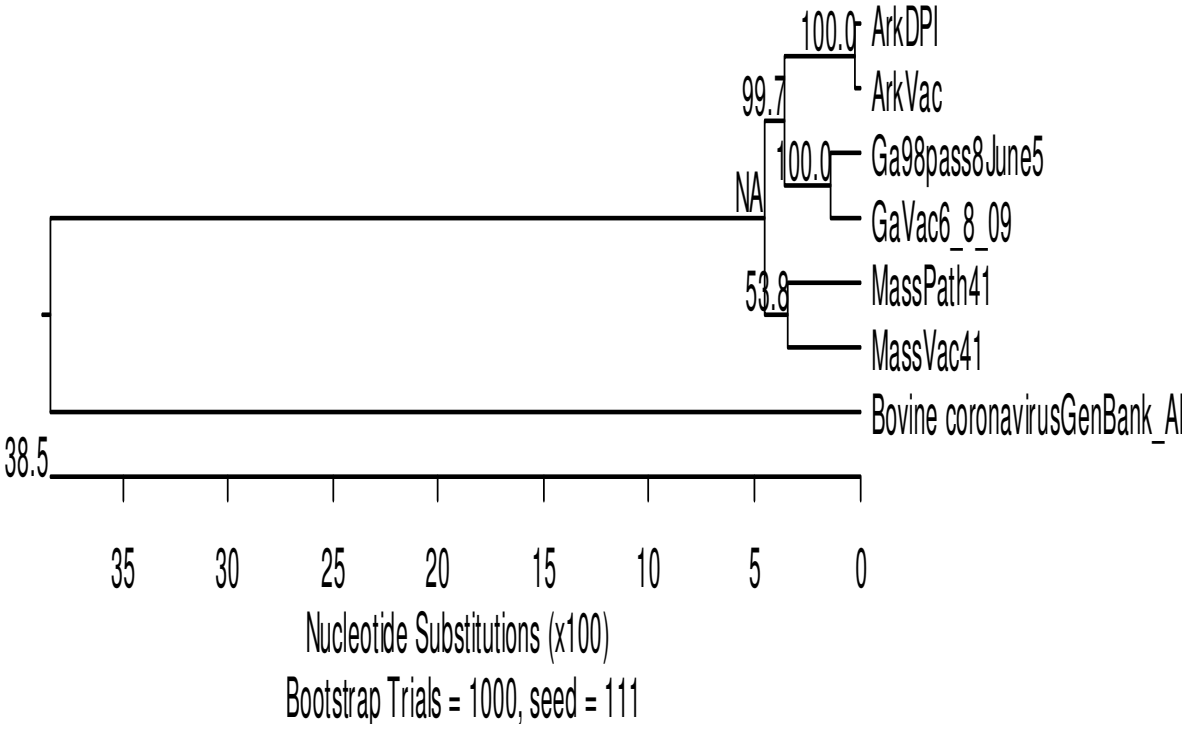


Figure 1: Phylogenetic Tree of Full length Genomes
Phylogenetic tree constructed using the Neighbor-Joining method following alignment by the Clustal W method using the Gonnet series. The full length genome was used for all viruses.

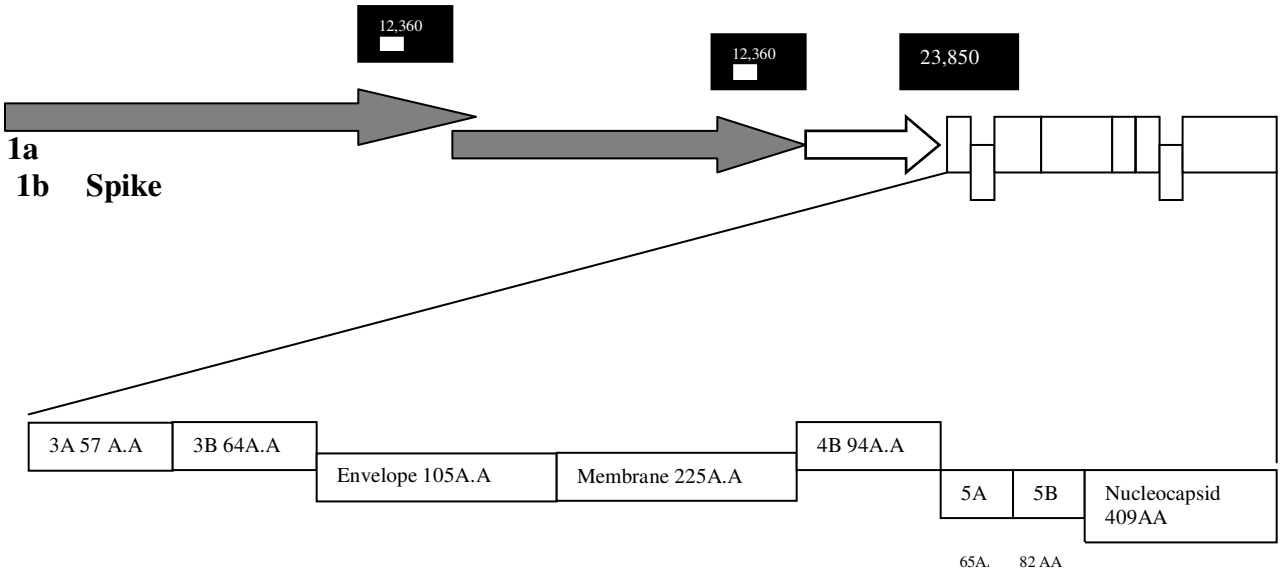


Figure 2: Schematic Diagram of Genome Organization
Schematic diagram representing the genome organization of IBV the nucleotides and amino acids are representative of the Arkansas pathogenic strain.

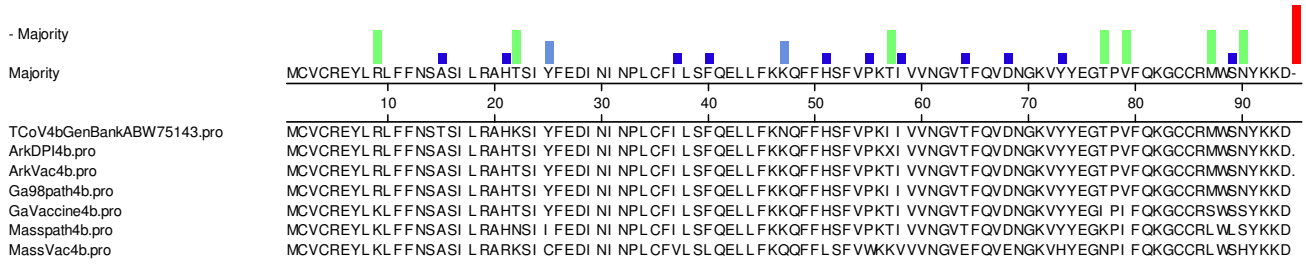
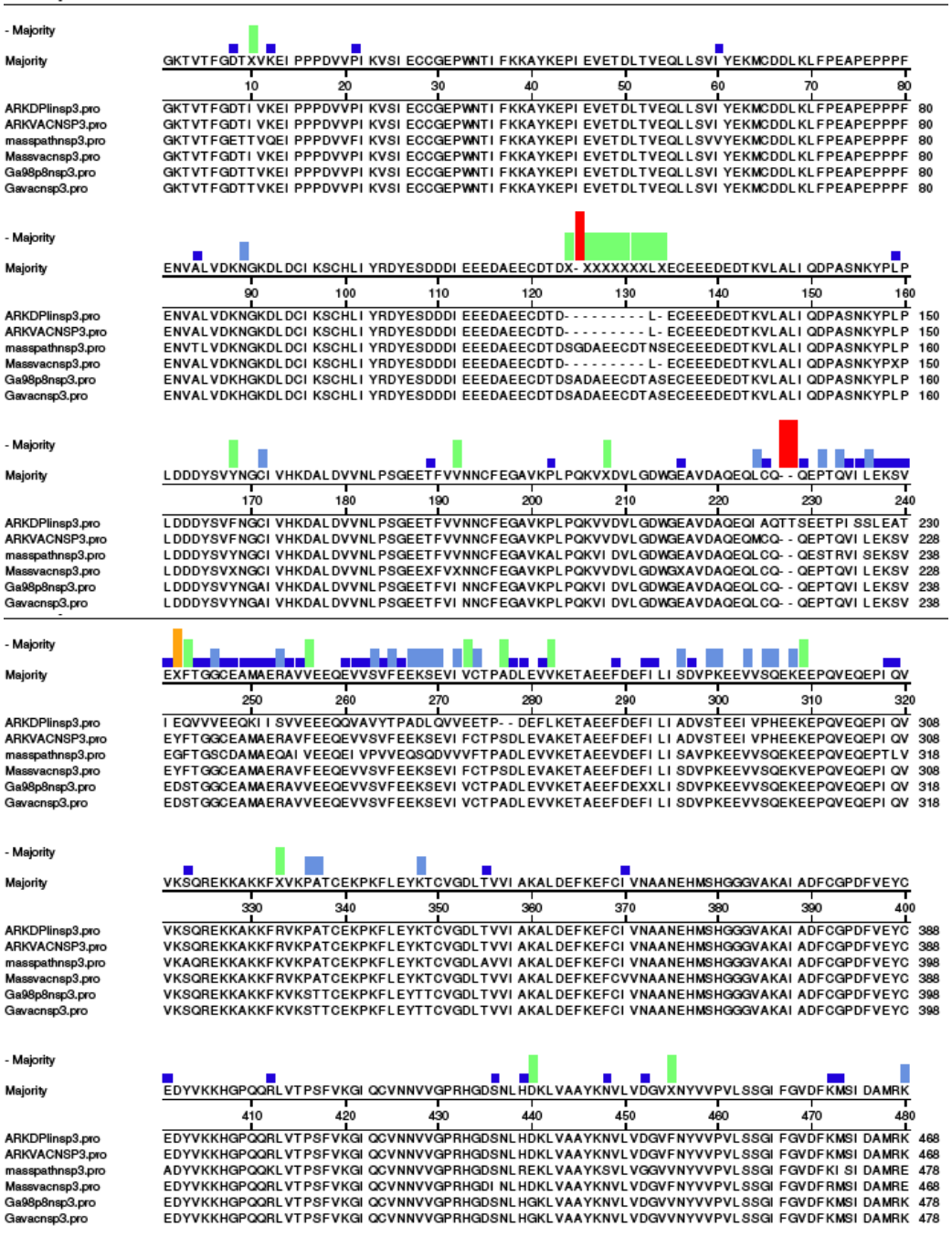
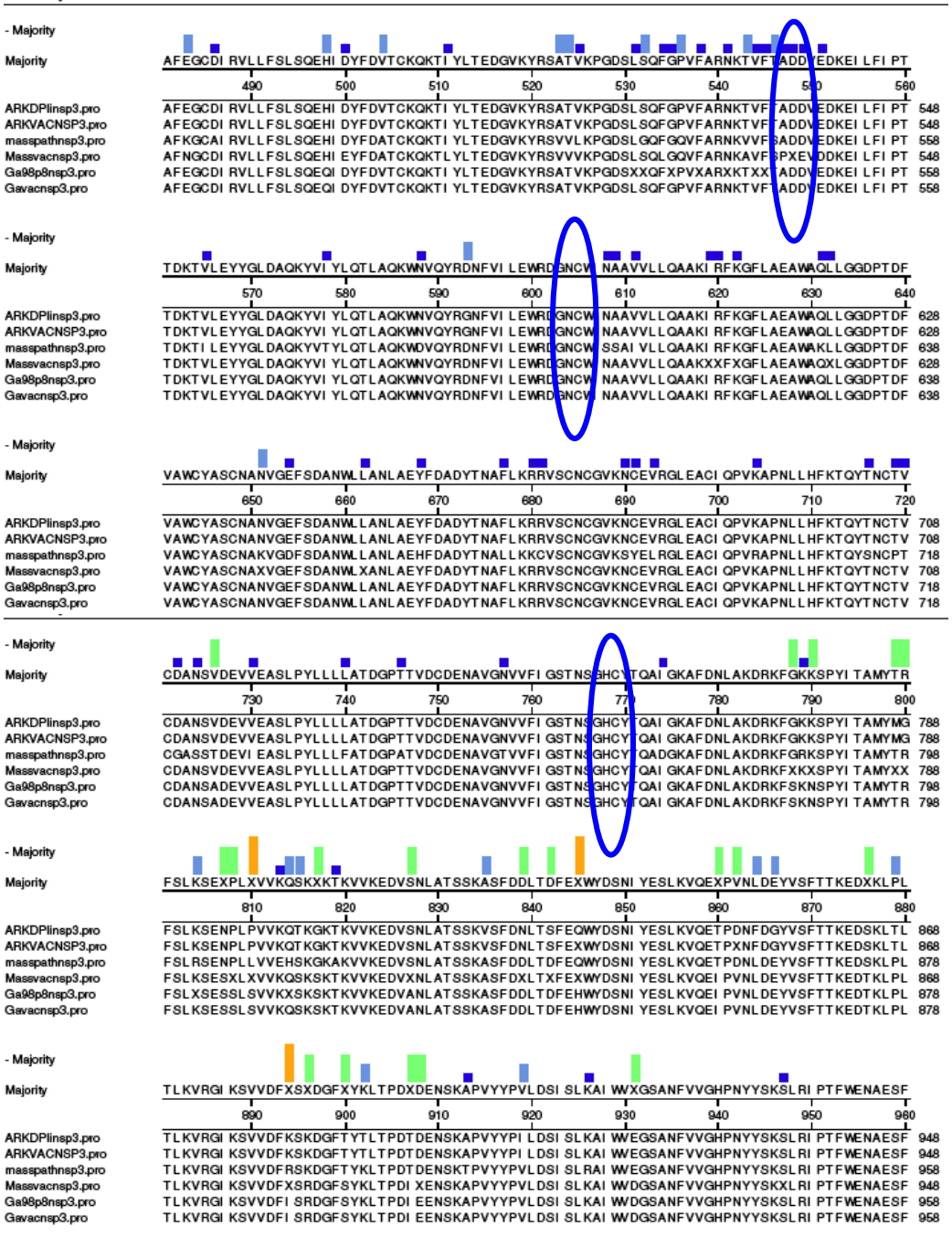
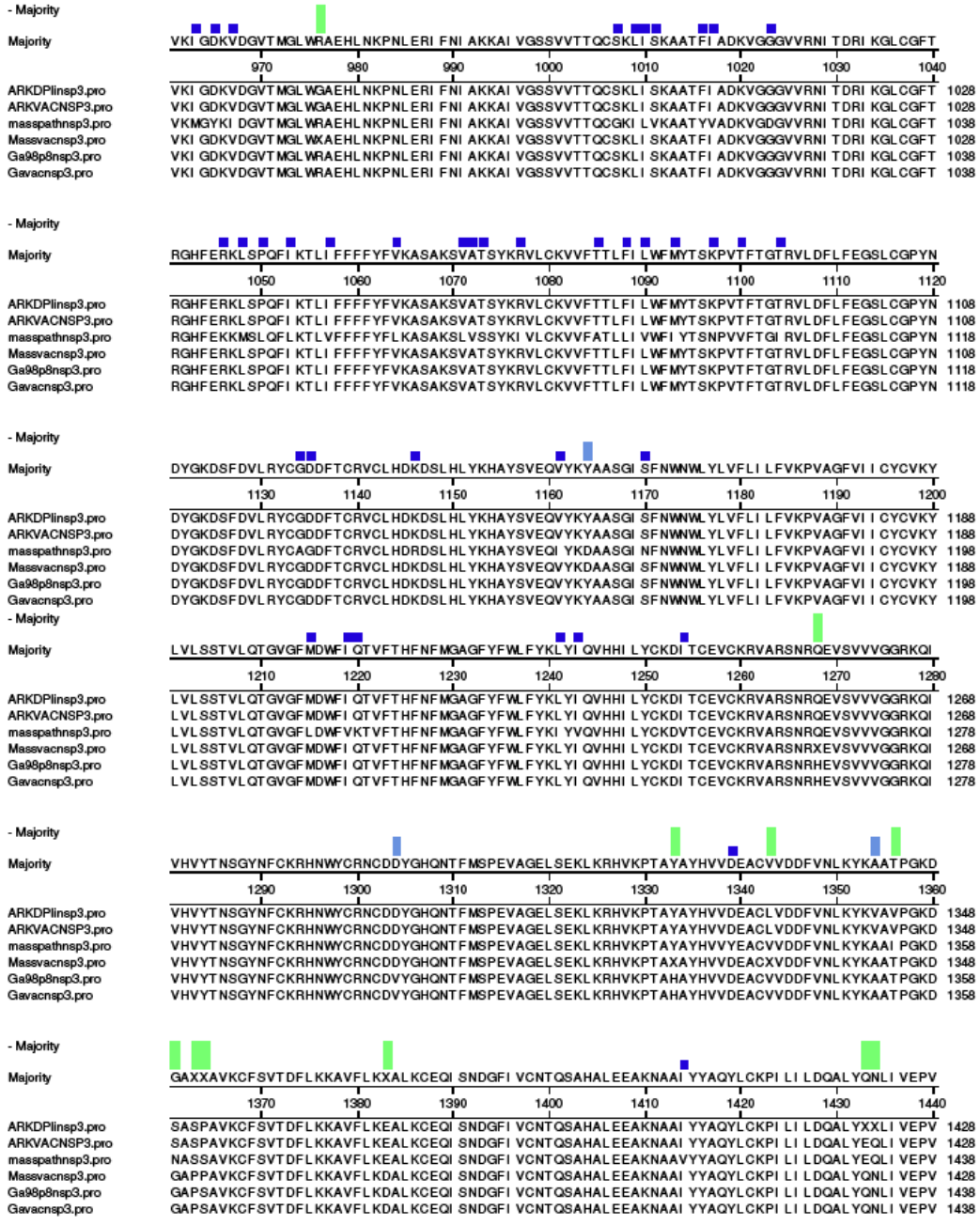


Figure 3: Protein 4b Sequence Alignment
Alignment of nonstructural protein 4b.


- Majority
Majority
GKTVTFGDTXVKEI PPPDVVPI KVSI ECCGEPWNTI FKKAYKEPI EVETDLTVEQLLSVI YEKMCDDLKLFPEAPEPPPF
10 20 30 40 50 60 70 80
ARKDPlinsp3.pro GKTVTFGDTI VKEI PPPDVVPI KVSI ECCGEPWNTI FKKAYKEPI EVETDLTVEQLLSVI YEKMCDDLKLFPEAPEPPPF 80
ARKVACNSP3.pro GKTVTFGDTI VKEI PPPDVVPI KVSI ECCGEPWNTI FKKAYKEPI EVETDLTVEQLLSVI YEKMCDDLKLFPEAPEPPPF 80
masspathnsp3.pro GKTVTFGETTVQEI PPPDVVFI KVSI ECCGEPWNTI FKKAYKEPI EVETDLTVEQLLSVVYEKMCDDLKLFPEAPEPPPF 80
Massvacnsp3.pro GKTVTFGDTI VKEI PPPDVVPI KVSI ECCGEPWNTI FKKAYKEPI EVETDLTVEQLLSVI YEKMCDDLKLFPEAPEPPPF 80
Ga98p8nsp3.pro GKTVTFGDTTVKEI PPPDVVPI KVSI ECCGEPWNTI FKKAYKEPI EVETDLTVEQLLSVI YEKMCDDLKLFPEAPEPPPF 80
Gavacnsp3.pro GKTVTFGDTTVKEI PPPDVVPI KVSI ECCGEPWNTI FKKAYKEPI EVETDLTVEQLLSVI YEKMCDDLKLFPEAPEPPPF 80
- Majority
Majority
ENVALVDKNGKDLDCI KSCHLI YRDYESDDDI EEEDAEECDTDX- XXXXXXXLXECEEEDEDTKVLALI QDPASNKYPLP
90 100 110 120 130 140 150 160
ARKDPlinsp3.pro ENVALVDKNGKDLDCI KSCHLI YRDYESDDDI EEEDAEECDTD- - - - - - - - - L- ECEEEDEDTKVLALI QDPASNKYPLP 150
ARKVACNSP3.pro ENVALVDKNGKDLDCI KSCHLI YRDYESDDDI EEEDAEECDTD- - - - - - - - - L- ECEEEDEDTKVLALI QDPASNKYPLP 150
masspathnsp3.pro ENVTLVDKNGKDLDCI KSCHLI YRDYESDDDI EEEDAEECDTDSGDAEECDTNSECEEEDEDTKVLALI QDPASNKYPLP 160
Massvacnsp3.pro ENVALVDKNGKDLDCI KSCHLI YRDYESDDDI EEEDAEECDTD- - - - - - - - - L- ECEEEDEDTKVLALI QDPASNKYPXP 150
Ga98p8nsp3.pro ENVALVDKHGKDLDCI KSCHLI YRDYESDDDI EEEDAEECDTDSADAEECDTASECEEEDEDTKVLALI QDPASNKYPLP 160
Gavacnsp3.pro ENVALVDKHGKDLDCI KSCHLI YRDYESDDDI EEEDAEECDTDSADAEECDTASECEEEDEDTKVLALI QDPASNKYPLP 160
- Majority
Majority
LDDDYSVYNGCI VHKDALDVVNLPSGEETFVVNNCFEGAVKPLPQKVXDVLGDWGEAVDAQEQLCQ- - QEPTQVI LEKSV
170 180 190 200 210 220 230 240
ARKDPlinsp3.pro LDDDYSVFNGCI VHKDALDVVNLPSGEETFVVNNCFEGAVKPLPQKVVDVLGDWGEAVDAQEQI AQTTSEETPI SSLEAT 230
ARKVACNSP3.pro LDDDYSVFNGCI VHKDALDVVNLPSGEETFVVNNCFEGAVKPLPQKVVDVLGDWGEAVDAQEQMCQ- - QEPTQVI LEKSV 228
masspathnsp3.pro LDDDYSVYNGCI VHKDALDVVNLPSGEETFVVNNCFEGAVKALPQKVI DVLGDWGEAVDAQEQLCQ- - QESTRVI SEKSV 238
Massvacnsp3.pro LDDDYSVXNGCI VHKDALDVVNLPSGEEXFVXNNCFEGAVKPLPQKVVDVLGDWGXAVDAQEQLCQ- - QEPTQVI LEKSV 238
Ga98p8nsp3.pro LDDDYSVYNGAI VHKDALDVVNLPSGEETFVI NNCFEGAVKPLPQKVI DVLGDWGEAVDAQEQLCQ- - QEPTQVI LEKSV 238
Gavacnsp3.pro LDDDYSVYNGAI VHKDALDVVNLPSGEETFVI NNCFEGAVKPLPQKVI DVLGDWGEAVDAQEQLCQ- - QEPTQVI LEKSV 238
- Majority
Majority
EXFTGGCEAMAERAVVEEQEVVSVFEEKSEVI VCTPADLEVVKETAEEFDEFI LI SDVPKEEVVSQEKEEPQVEQEPI QV
250 260 270 280 290 300 310 320
ARKDPlinsp3.pro I EQVVVEEQKI I SVVEEEQQVAVYTPADLQVVEETP- - DEFLKETAEEFDEFI LI ADVSTEEI VPHEEKEPQVEQEPI QV 308
ARKVACNSP3.pro EYFTGGCEAMAERAVFEEQEVVSVFEEKSEVI FCTPSDLEVAKETAEEFDEFI LI ADVSTEEI VPHEEKEPQVEQEPI QV 308
masspathnsp3.pro EGFTGSCDAMAEQAI VEEQEI VPVVEQSQDVVVFTPADLEVVKETAEEVDEFI LI SAVPKEEVVSQEKEEPQVEQEPTLV 318
Massvacnsp3.pro EYFTGGCEAMAERAVFEEQEVVSVFEEKSEVI FCTPSDLEVAKETAEEFDEFI LI SDVPKEEVVSQEKVEPQVEQEPI QV 308
Ga98p8nsp3.pro EDSTGGCEAMAERAVVEEQEVVSVFEEKSEVI VCTPADLEVVKETAEEFDEXXLI SDVPKEEVVSQEKEEPQVEQEPI QV 318
Gavacnsp3.pro EDSTGGCEAMAERAVVEEQEVVSVFEEKSEVI VCTPADLEVVKETAEEFDEFI LI SDVPKEEVVSQEKEEPQVEQEPI QV 318
- Majority
Majority
VKSQREKKAKKFXVKPATCEKPKFLEYKTCVGDLTVVI AKALDEFKEFCI VNAANEHMSHGGGVAKAI ADFCGPDFVEYC
330 340 350 360 370 380 390 400
ARKDPlinsp3.pro VKSQREKKAKKFRVKPATCEKPKFLEYKTCVGDLTVVI AKALDEFKEFCI VNAANEHMSHGGGVAKAI ADFCGPDFVEYC 388
ARKVACNSP3.pro VKSQREKKAKKFRVKPATCEKPKFLEYKTCVGDLTVVI AKALDEFKEFCI VNAANEHMSHGGGVAKAI ADFCGPDFVEYC 388
masspathnsp3.pro VKAQREKKAKKFKVKPATCEKPKFLEYKTCVGDLAVVI AKALDEFKEFCI VNAANEHMSHGGGVAKAI ADFCGPDFVEYC 398
Massvacnsp3.pro VKSQREKKAKKFRVKPATCEKPKFLEYKTCVGDLTVVI AKALDEFKEFCVVNAANEHMSHGGGVAKAI ADFCGPDFVEYC 398
Ga98p8nsp3.pro VKSQREKKAKKFKVKSTTCEKPKFLEYTTCVGDLTVVI AKALDEFKEFCI VNAANEHMSHGGGVAKAI ADFCGPDFVEYC 398
Gavacnsp3.pro VKSQREKKAKKFKVKSTTCEKPKFLEYTTCVGDLTVVI AKALDEFKEFCI VNAANEHMSHGGGVAKAI ADFCGPDFVEYC 398
- Majority
Majority
EDYVKKHGPQQRLVTPSFVKGI QCVNNVVGPRHGDSNLHDKLVAAYKNVLVDGVXNYVVPVLSSGI FGVDFKMSI DAMRK
410 420 430 440 450 460 470 480
ARKDPlinsp3.pro EDYVKKHGPQQRLVTPSFVKGI QCVNNVVGPRHGDSNLHDKLVAAYKNVLVDGVFNYVVPVLSSGI FGVDFKMSI DAMRK 468
ARKVACNSP3.pro EDYVKKHGPQQRLVTPSFVKGI QCVNNVVGPRHGDSNLHDKLVAAYKNVLVDGVFNYVVPVLSSGI FGVDFKMSI DAMRK 468
masspathnsp3.pro ADYVKKHGPQQKLVTPSFVKGI QCVNNVVGPRHGDSNLREKLVAAYKSVLVGGVVNYVVPVLSSGI FGVDFKI SI DAMRE 478
Massvacnsp3.pro EDYVKKHGPQQRLVTPSFVKGI QCVNNVVGPRHGDI NLHDKLVAAYKNVLVDGVFNYVVPVLSSGI FGVDFRMSI DAMRE 468
Ga98p8nsp3.pro EDYVKKHGPQQRLVTPSFVKGI QCVNNVVGPRHGDSNLHGKLVAAYKNVLVDGVVNYVVPVLSSGI FGVDFKMSI DAMRK 478
Gavacnsp3.pro EDYVKKHGPQQRLVTPSFVKGI QCVNNVVGPRHGDSNLHGKLVAAYKNVLVDGVVNYVVPVLSSGI FGVDFKMSI DAMRK 478


- Majority
Majority
AFEGCDI RVLLFSLSQEHI DYFDVTCKQKTI YLTEDGVKYRSATVKPGDSLSQFGPVFARNKTVFTADDVEDKEI LFI PT
490 500 510 520 530 540 550 560
ARKDPIinsp3.pro AFEGCDI RVLLFSLSQEHI DYFDVTCKQKTI YLTEDGVKYRSATVKPGDSLSQFGPVFARNKTVFTADDVEDKEI LFI PT 548
ARKVACNSP3.pro AFEGCDI RVLLFSLSQEHI DYFDVTCKQKTI YLTEDGVKYRSATVKPGDSLSQFGPVFARNKTVFTADDVEDKEI LFI PT 548
masspathnsp3.pro AFKGCAI RVLLFSLSQEHI DYFDATCKQKTI YLTEDGVKYRSVVLKPGDSLGQFGQVFARNKVVFSADDVEDKEI LFI PT 558
Massvacnsp3.pro AFNGCDI RVLLFSLSQEHI EYFDATCKQKTL YLTEDGVKYRSVVVKPGDSLSQLGQVFARNKAVFSPXEVDDKEI LFI PT 548
Ga98p8nsp3.pro AFEGCDI RVLLFSLSQEQI DYFDVTCKQKTI YLTEDGVKYRSATVKPGDSXXQFXPVXARXKTXXTADDVEDKEI LFI PT 558
Gavacnsp3.pro AFEGCDI RVLLFSLSQEQI DYFDVTCKQKTI YLTEDGVKYRSATVKPGDSLSQFGPVFARNKTVFTADDVEDKEI LFI PT 558
- Majority
Majority
TDKTVLEYYGLDAQKYVI YLQTLAQKWNVQYRDNFVI LEWRDGNCV NAAVVLLQAAKI RFKGFLAEAWAQLLGGDPTDF
570 580 590 600 610 620 630 640
ARKDPIinsp3.pro TDKTVLEYYGLDAQKYVI YLQTLAQKWNVQYRGNFVI LEWRDGNCW NAAVVLLQAAKI RFKGFLAEAWAQLLGGDPTDF 628
ARKVACNSP3.pro TDKTVLEYYGLDAQKYVI YLQTLAQKWNVQYRGNFVI LEWRDGNCW NAAVVLLQAAKI RFKGFLAEAWAQLLGGDPTDF 628
masspathnsp3.pro TDKTI LEYYGLDAQKYVTYLQTLAQKWDVQYRDNFVI LEWRDGNCW SSAI VLLQAAKI RFKGFLAEAWAKLLGGDPTDF 638
Massvacnsp3.pro TDKTVLEYYGLDAQKYVI YLQTLAQKWNVQYRDNFVI LEWRDGNCW NAAVVLLQAAKXXFXGFLAEAWAQXLGGDPTDF 628
Ga98p8nsp3.pro TDKTVLEYYGLDAQKYVI YLQTLAQKWNVQYRDNFVI LEWRDGNCW NAAVVLLQAAKI RFKGFLAEAWAQLLGGDPTDF 638
Gavacnsp3.pro TDKTVLEYYGLDAQKYVI YLQTLAQKWNVQYRDNFVI LEWRDGNCW NAAVVLLQAAKI RFKGFLAEAWAQLLGGDPTDF 638
- Majority
Majority
VAWCYASCNANVGEFSDANWLLANLAEYFDADYTNAFLKRRVSCNCGVKNCEVRGLEACI QPVKAPNLLHFKTQYTNCTV
650 660 670 680 690 700 710 720
ARKDPIinsp3.pro VAWCYASCNANVGEFSDANWLLANLAEYFDADYTNAFLKRRVSCNCGVKNCEVRGLEACI QPVKAPNLLHFKTQYTNCTV 708
ARKVACNSP3.pro VAWCYASCNANVGEFSDANWLLANLAEYFDADYTNAFLKRRVSCNCGVKNCEVRGLEACI QPVKAPNLLHFKTQYTNCTV 708
masspathnsp3.pro VAWCYASCNAKVGDFSDANWLLANLAEHFDADYTNALLKKCVSCNCGVKSYELRGLEACI QPVRAPNLLHFKTQYSNCPT 718
Massvacnsp3.pro VAWCYASCNAXVGEFSDANWLXANLAEYFDADYTNAFLKRRVSCNCGVKNCEVRGLEACI QPVKAPNLLHFKTQYTNCTV 708
Ga98p8nsp3.pro VAWCYASCNANVGEFSDANWLLANLAEYFDADYTNAFLKRRVSCNCGVKNCEVRGLEACI QPVKAPNLLHFKTQYTNCTV 718
Gavacnsp3.pro VAWCYASCNANVGEFSDANWLLANLAEYFDADYTNAFLKRRVSCNCGVKNCEVRGLEACI QPVKAPNLLHFKTQYTNCTV 718
- Majority
Majority
CDANSVDEVVEASLPYLLLLATDGPTTVDCDENAVGNVVFI GSTNSGHCYTQAI GKAFDNLAKDRKFGKKSPYI TAMYTR
730 740 750 760 770 780 790 800
ARKDPIinsp3.pro CDANSVDEVVEASLPYLLLLATDGPTTVDCDENAVGNVVFI GSTNSGHCYTQAI GKAFDNLAKDRKFGKKSPYI TAMYMG 788
ARKVACNSP3.pro CDANSVDEVVEASLPYLLLLATDGPTTVDCDENAVGNVVFI GSTNSGHCYTQAI GKAFDNLAKDRKFGKKSPYI TAMYMG 788
masspathnsp3.pro CGASSTDEVI EASLPYLLLFATDGPATVDCDENAVGTVVFI GSTNSGHCYTQADGKAFDNLAKDRKFGRKSPYI TAMYTR 798
Massvacnsp3.pro CDANSVDEVVEASLPYLLLLATDGPTTVDCDENAVGNVVFI GSTNSGHCYTQAI GKAFDNLAKDRKFXKXSPYI TAMYXX 788
Ga98p8nsp3.pro CDANSADEVVEASLPYLLLLATDGPTTVDCDENAVGNVVFI GSTNSGHCYTQAI GKAFDNLAKDRKFSKNSPYI TAMYTR 798
Gavacnsp3.pro CDANSADEVVEASLPYLLLLATDGPTTVDCDENAVGNVVFI GSTNSGHCYTQAI GKAFDNLAKDRKFSKNSPYI TAMYTR 798
- Majority
Majority
FSLKSEXPLXVVKQSKXKTKVVKEDVSNLATSSKASFDDLTDFEXWYDSNI YESLKVQEXPVNLDEYVSFTTKEDXKLPL
810 820 830 840 850 860 870 880
ARKDPIinsp3.pro FSLKSENPLPVVKQTKGKTKVVKEDVSNLATSSKVSFDNLTSFEQWYDSNI YESLKVQETPDNFDGYVSFTTKEDSKLTL 868
ARKVACNSP3.pro FSLKSENPLPVVKQTKGKTKVVKEDVSNLATSSKVSFDNLTSFEXWYDSNI YESLKVQETPXNFDGYVSFTTKEDSKLTL 868
masspathnsp3.pro FSLRSENPLLVVEHSKGKAKVVKEDVSNLATSSKASFDDLTDFEQWYDSNI YESLKVQETPDNLDEYVSFTTKEDSKLPL 878
Massvacnsp3.pro FSLKSESXLXVVKQSKSKTKVVKEDVXNLATSSKASFDXLTXFEXWYDSNI YESLKVQEI PVNLDEYVSFTTKEDTKLPL 868
Ga98p8nsp3.pro FSLXSESSLSVVKXSKSKTKVVKEDVANLATSSKASFDDLTDFEHWYDSNI YESLKVQEI PVNLDEYVSFTTKEDTKLPL 878
Gavacnsp3.pro FSLKSESSLSVVKQSKSKTKVVKEDVANLATSSKASFDDLTDFEHWYDSNI YESLKVQEI PVNLDEYVSFTTKEDTKLPL 878
- Majority
Majority
TLKVRGI KSVVDFXSXDGFXYKLTPDXDENSKAPVYYPVLDSI SLKAI WXGSANFVVGHPNYYSKSLRI PTFWENAESF
890 900 910 920 930 940 950 960
ARKDPIinsp3.pro TLKVRGI KSVVDFKSKDGFTYTLTPDTDENSKAPVYYPI LDSI SLKAI WEGSANFVVGHPNYYSKSLRI PTFWENAESF 948
ARKVACNSP3.pro TLKVRGI KSVVDFKSKDGFTYTLTPDTDENSKAPVYYPI LDSI SLKAI WEGSANFVVGHPNYYSKSLRI PTFWENAESF 948
masspathnsp3.pro TLKVRGI KSVVDFRSKDGFTYKLTPDTDENSKTPVYYPVLDSI SLRAI WEGSANFVVGHPNYYSKSLRI PTFWENAESF 958
Massvacnsp3.pro TLKVRGI KSVVDFXSRDGFSYKLTPDI XENSKAPVYYPVLDSI SLKAI WDGSANFVVGHPNYYSKXLRI PTFWENAESF 948
Ga98p8nsp3.pro TLKVRGI KSVVDFI SRDGFSYKLTPDI EENSKAPVYYPVLDSI SLKAI WDGSANFVVGHPNYYSKSLRI PTFWENAESF 958
Gavacnsp3.pro TLKVRGI KSVVDFI SRDGFSYKLTPDI EENSKAPVYYPVLDSI SLKAI WDGSANFVVGHPNYYSKSLRI PTFWENAESF 958


- Majority
Majority
VKI GDKVDGVTMGLWRAEHLNKPNLERI FNI AKKAI VGSSVVTTQCSKLI SKAATFI ADKVGGGVVRNI TDRI KGLCGFT
970 980 990 1000 1010 1020 1030 1040
ARKDPIinsp3.pro VKI GDKVDGVTMGLWGAEHLNKPNLERI FNI AKKAI VGSSVVTTQCSKLI SKAATFI ADKVGGGVVRNI TDRI KGLCGFT 1028
ARKVACNSP3.pro VKI GDKVDGVTMGLWGAEHLNKPNLERI FNI AKKAI VGSSVVTTQCSKLI SKAATFI ADKVGGGVVRNI TDRI KGLCGFT 1028
masspathnsp3.pro VKMGYKI DGVTMGLWRAEHLNKPNLERI FNI AKKAI VGSSVVTTQCGKI LVKAATYVADKVGDGVVRNI TDRI KGLCGFT 1038
Massvacnsp3.pro VKI GDKVDGVTMGLWXAEHLNKPNLERI FNI AKKAI VGSSVVTTQCSKLI SKAATFI ADKVGGGVVRNI TDRI KGLCGFT 1028
Ga98p8nsp3.pro VKI GDKVDGVTMGLWRAEHLNKPNLERI FNI AKKAI VGSSVVTTQCSKLI SKAATFI ADKVGGGVVRNI TDRI KGLCGFT 1038
Gavacnsp3.pro VKI GDKVDGVTMGLWRAEHLNKPNLERI FNI AKKAI VGSSVVTTQCSKLI SKAATFI ADKVGGGVVRNI TDRI KGLCGFT 1038
- Majority
Majority
RGHFERKLSPQFI KTLI FFFFYFVKASAKSVATSYKRVLCKVVFTTLFI LWFMYTSKPVTFTGTRVLDFLFEGSLCGPYN
1050 1060 1070 1080 1090 1100 1110 1120
ARKDPIinsp3.pro RGHFERKLSPQFI KTLI FFFFYFVKASAKSVATSYKRVLCKVVFTTLFI LWFMYTSKPVTFTGTRVLDFLFEGSLCGPYN 1108
ARKVACNSP3.pro RGHFERKLSPQFI KTLI FFFFYFVKASAKSVATSYKRVLCKVVFTTLFI LWFMYTSKPVTFTGTRVLDFLFEGSLCGPYN 1108
masspathnsp3.pro RGHFEKKMSLQFLKTLVFFFFYFLKASAKSLVSSYKI VLCKVVFATLLI VWFI YTSNPVVFTGI RVLDFLFEGSLCGPYN 1118
Massvacnsp3.pro RGHFERKLSPQFI KTLI FFFFYFVKASAKSVATSYKRVLCKVVFTTLFI LWFMYTSKPVTFTGTRVLDFLFEGSLCGPYN 1108
Ga98p8nsp3.pro RGHFERKLSPQFI KTLI FFFFYFVKASAKSVATSYKRVLCKVVFTTLFI LWFMYTSKPVTFTGTRVLDFLFEGSLCGPYN 1118
Gavacnsp3.pro RGHFERKLSPQFI KTLI FFFFYFVKASAKSVATSYKRVLCKVVFTTLFI LWFMYTSKPVTFTGTRVLDFLFEGSLCGPYN 1118
- Majority
Majority
DYGKDSFDVLRYCGDDFTCRVCLHDKDSLHLYKHAYSVEQVYKYAASGI SFNWNWLYLVFLI LFVKPVAGFVI I CYCVKY
1130 1140 1150 1160 1170 1180 1190 1200
ARKDPIinsp3.pro DYGKDSFDVLRYCGDDFTCRVCLHDKDSLHLYKHAYSVEQVYKYAASGI SFNWNWLYLVFLI LFVKPVAGFVI I CYCVKY 1188
ARKVACNSP3.pro DYGKDSFDVLRYCGDDFTCRVCLHDKDSLHLYKHAYSVEQVYKYAASGI SFNWNWLYLVFLI LFVKPVAGFVI I CYCVKY 1188
masspathnsp3.pro DYGKDSFDVLRYCAGDFTCRVCLHDRDSLHLYKHAYSVEQI YKDAASGI NFNWNWLYLVFLI LFVKPVAGFVI I CYCVKY 1198
Massvacnsp3.pro DYGKDSFDVLRYCGDDFTCRVCLHDKDSLHLYKHAYSVEQVYKDAASGI SFNWNWLYLVFLI LFVKPVAGFVI I CYCVKY 1188
Ga98p8nsp3.pro DYGKDSFDVLRYCGDDFTCRVCLHDKDSLHLYKHAYSVEQVYKYAASGI SFNWNWLYLVFLI LFVKPVAGFVI I CYCVKY 1198
Gavacnsp3.pro DYGKDSFDVLRYCGDDFTCRVCLHDKDSLHLYKHAYSVEQVYKYAASGI SFNWNWLYLVFLI LFVKPVAGFVI I CYCVKY 1198
- Majority
Majority
LVLSSTVLQTGVGFMDWFI QTVFTHFNFMGAGFYFWLFYKLYI QVHHI LYCKDI TCEVCKRVARSNRQEVSVVVGGRKQI
1210 1220 1230 1240 1250 1260 1270 1280
ARKDPIinsp3.pro LVLSSTVLQTGVGFMDWFI QTVFTHFNFMGAGFYFWLFYKLYI QVHHI LYCKDI TCEVCKRVARSNRQEVSVVVGGRKQI 1268
ARKVACNSP3.pro LVLSSTVLQTGVGFMDWFI QTVFTHFNFMGAGFYFWLFYKLYI QVHHI LYCKDI TCEVCKRVARSNRQEVSVVVGGRKQI 1268
masspathnsp3.pro LVLSSTVLQTGVGFLDWFVKTVFTHFNFMGAGFYFWLFYKI YVQVHHI LYCKDVTCEVCKRVARSNRQEVSVVVGGRKQI 1278
Massvacnsp3.pro LVLSSTVLQTGVGFMDWFI QTVFTHFNFMGAGFYFWLFYKLYI QVHHI LYCKDI TCEVCKRVARSNRXEVSVVVGGRKQI 1268
Ga98p8nsp3.pro LVLSSTVLQTGVGFMDWFI QTVFTHFNFMGAGFYFWLFYKLYI QVHHI LYCKDI TCEVCKRVARSNRHEVSVVVGGRKQI 1278
Gavacnsp3.pro LVLSSTVLQTGVGFMDWFI QTVFTHFNFMGAGFYFWLFYKLYI QVHHI LYCKDI TCEVCKRVARSNRHEVSVVVGGRKQI 1278
- Majority
Majority
VHVYTNSGYNFCKRHNWYCRNCDDYGHQNTFMSPEVAGELSEKLKRHVKPTAYAYHVVDEACVVDDFVNLKYKAATPGKD
1290 1300 1310 1320 1330 1340 1350 1360
ARKDPIinsp3.pro VHVYTNSGYNFCKRHNWYCRNCDDYGHQNTFMSPEVAGELSEKLKRHVKPTAYAYHVVDEACLVDDFVNLKYKVAVPGKD 1348
ARKVACNSP3.pro VHVYTNSGYNFCKRHNWYCRNCDDYGHQNTFMSPEVAGELSEKLKRHVKPTAYAYHVVDEACLVDDFVNLKYKVAVPGKD 1348
masspathnsp3.pro VHVYTNSGYNFCKRHNWYCRNCDDYGHQNTFMSPEVAGELSEKLKRHVKPTAYAYHVVYEACVVDDFVNLKYKAAI PGKD 1358
Massvacnsp3.pro VHVYTNSGYNFCKRHNWYCRNCDDYGHQNTFMSPEVAGELSEKLKRHVKPTAXAYHVVDEACXVDDFVNLKYKAATPGKD 1348
Ga98p8nsp3.pro VHVYTNSGYNFCKRHNWYCRNCDVYGHQNTFMSPEVAGELSEKLKRHVKPTAHAYHVVDEACVVDDFVNLKYKAATPGKD 1358
Gavacnsp3.pro VHVYTNSGYNFCKRHNWYCRNCDVYGHQNTFMSPEVAGELSEKLKRHVKPTAHAYHVVDEACVVDDFVNLKYKAATPGKD 1358
- Majority
Majority
GAXXAVKCFSVTDFLKKAVFLKXALKCEQI SNDGFI VCNTQSAHALEEAKNAAI YYAQYLCKPI LI LDQALYQNLI VEPV
1370 1380 1390 1400 1410 1420 1430 1440
ARKDPIinsp3.pro SASPAVKCFSVTDFLKKAVFLKEALKCEQI SNDGFI VCNTQSAHALEEAKNAAI YYAQYLCKPI LI LDQALYXXLI VEPV 1428
ARKVACNSP3.pro SASPAVKCFSVTDFLKKAVFLKEALKCEQI SNDGFI VCNTQSAHALEEAKNAAI YYAQYLCKPI LI LDQALYEQLI VEPV 1428
masspathnsp3.pro NASSAVKCFSVTDFLKKAVFLKEALKCEQI SNDGFI VCNTQSAHALEEAKNAAVYYAQYLCKPI LI LDQALYEQLI VEPV 1438
Massvacnsp3.pro GAPPAVKCFSVTDFLKKAVFLKDALKCEQI SNDGFI VCNTQSAHALEEAKNAAI YYAQYLCKPI LI LDQALYQNLI VEPV 1428
Ga98p8nsp3.pro GAPSAVKCFSVTDFLKKAVFLKDALKCEQI SNDGFI VCNTQSAHALEEAKNAAI YYAQYLCKPI LI LDQALYQNLI VEPV 1438
Gavacnsp3.pro GAPSAVKCFSVTDFLKKAVFLKDALKCEQI SNDGFI VCNTQSAHALEEAKNAAI YYAQYLCKPI LI LDQALYQNLI VEPV 1438

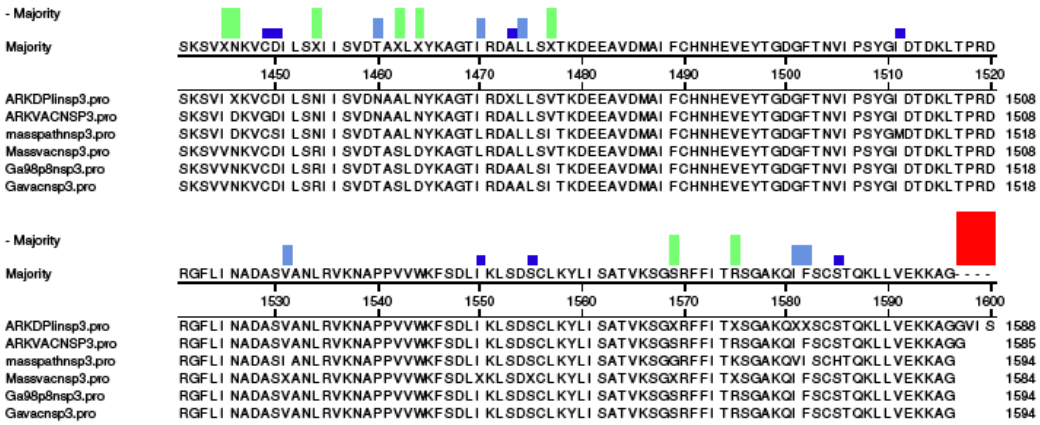


Figure 4: Nonstructural Protein Three Alignment
Nsp 3 alignment for Arkansas, Georgia, and Massachusetts the pathogenic and attenuated strains. The key residues that make up the catalytic triad in the Papain Like-2 Domain are identified by the blue circles.

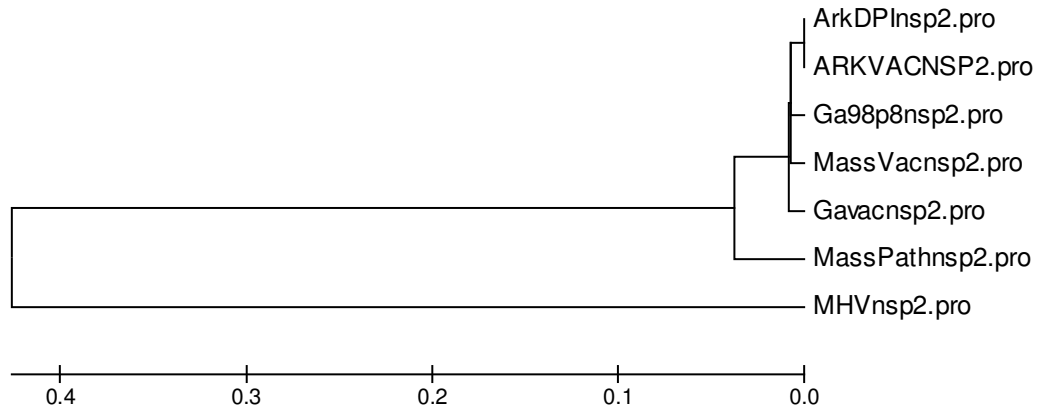


Figure 5: Phylogenetic Tree of Nsp 2

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.91425439 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 570 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

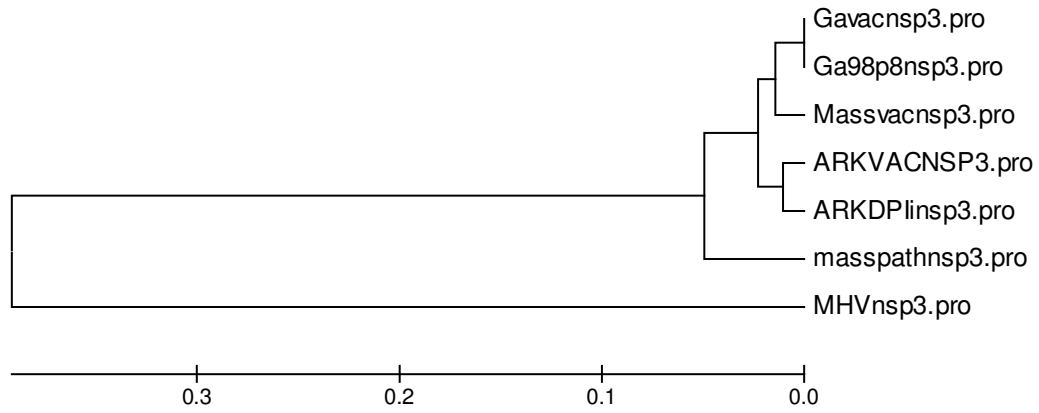


Figure 6: Phylogenetic Tree of Nsp 3

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.88210227 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [2]. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [3]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 1496 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [4].

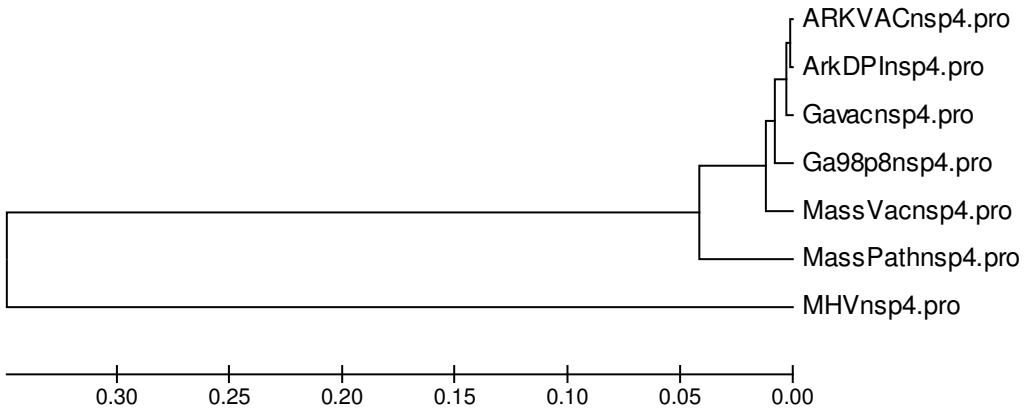


Figure 7: Phylogenetic Tree of Nsp 4

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.76444892 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 465 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

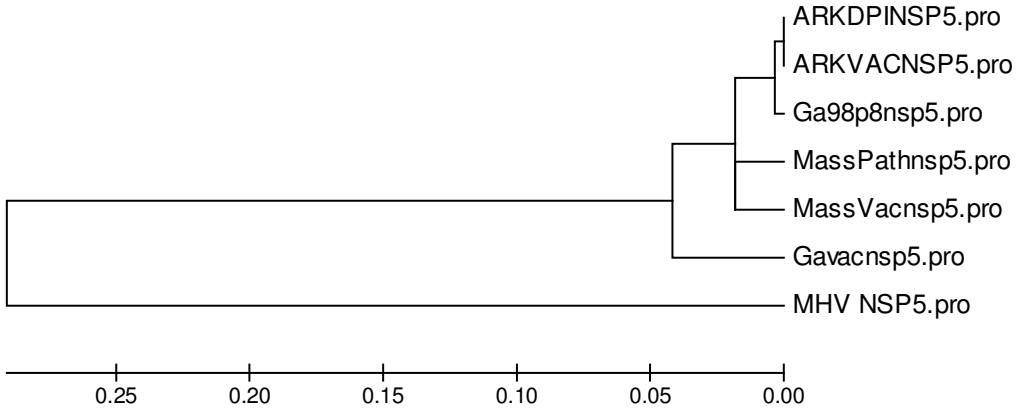


Figure 8: Phylogenetic Tree of Nsp 5

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.67391304 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 299 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

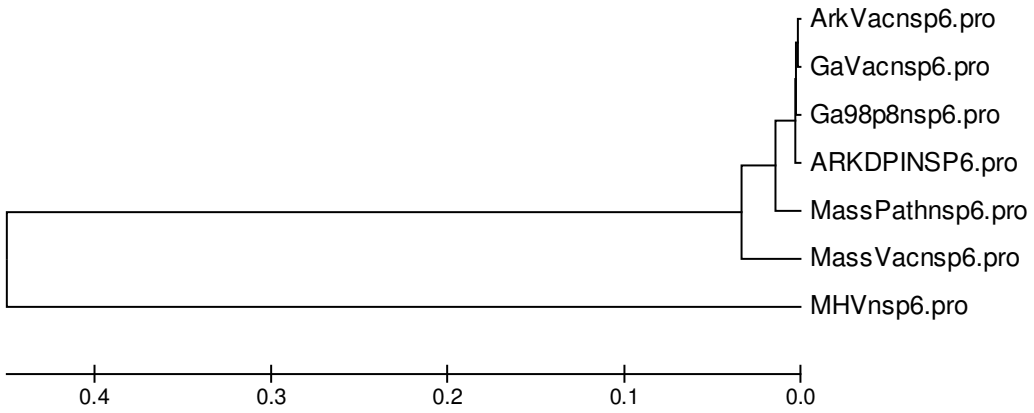


Figure 9: Phylogenetic Tree of Nsp 6

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.95659091 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 275 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

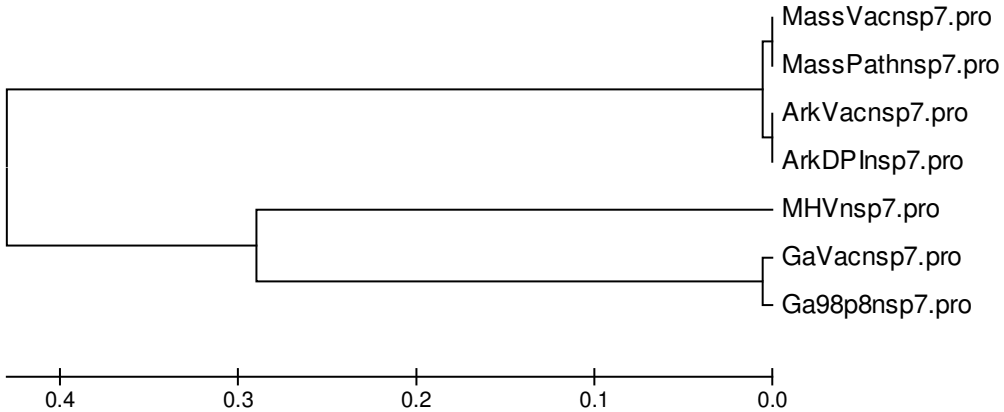


Figure 10: Phylogenetic Tree of Nsp 7

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths= 1.17045455 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 88 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

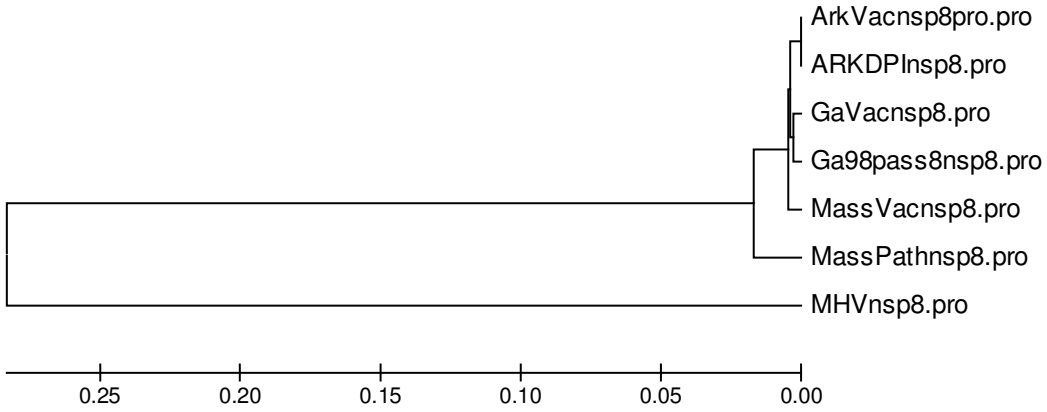


Figure 11: Phylogenetic Tree of Nsp 8

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.59275266 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 188 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

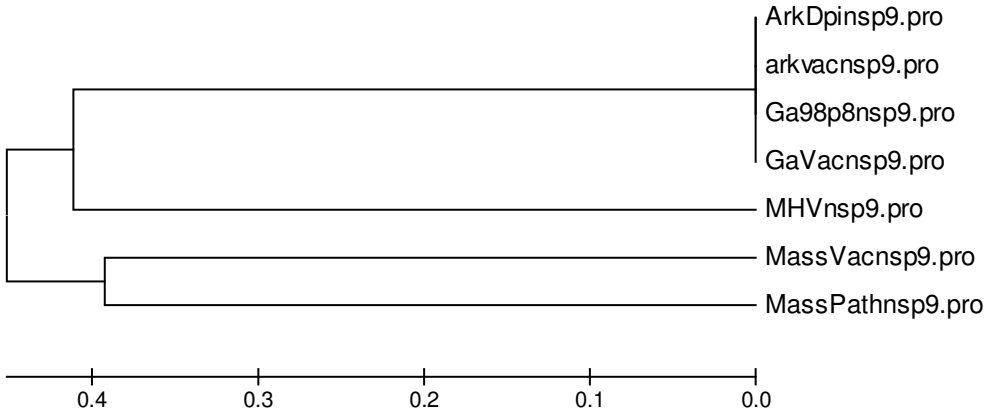


Figure 12: Phylogenetic Tree of Nsp 9

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 1.66588785 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 107 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

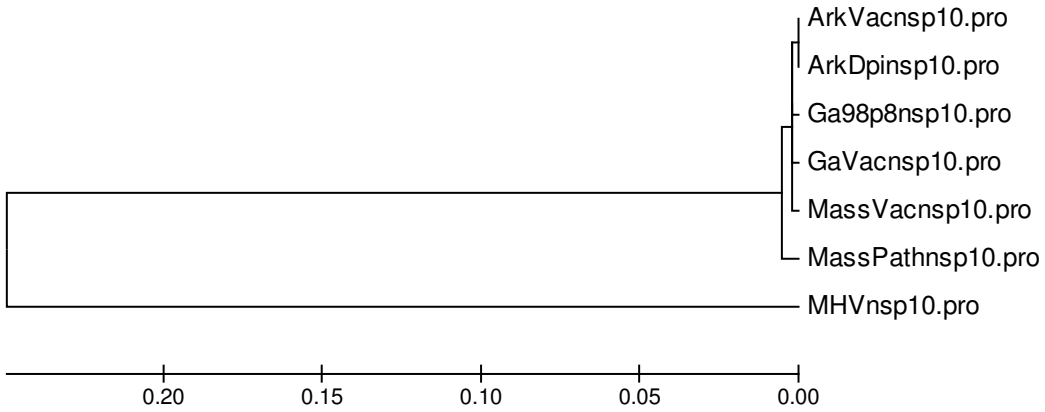


Figure 13: Phylogenetic Tree of Nsp 10

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.50751880 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 133 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

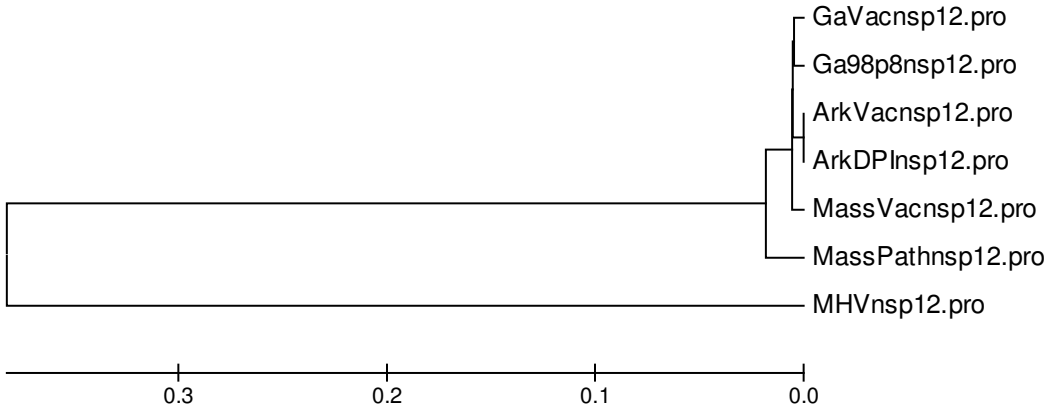


Figure 14: Phylogenetic Tree of Nsp 12

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.79687500 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 336 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

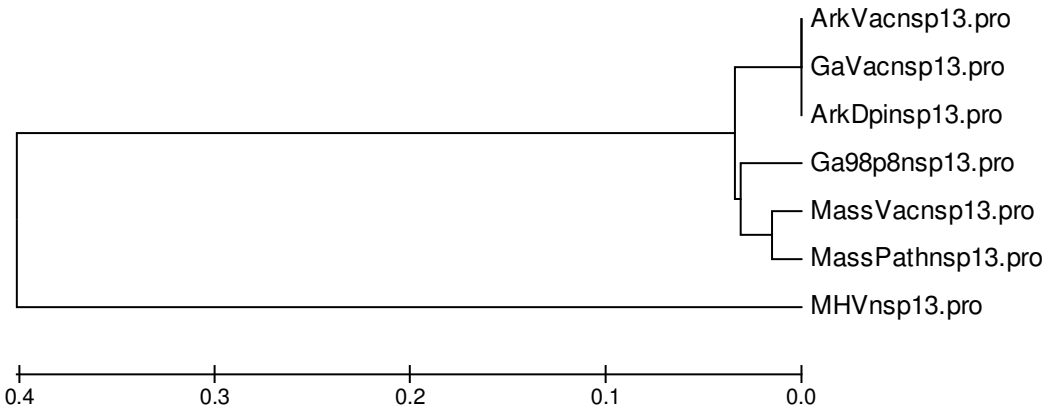


Figure 15: Phylogenetic Tree of Nsp 13

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch length = 0.88043478 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 299 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

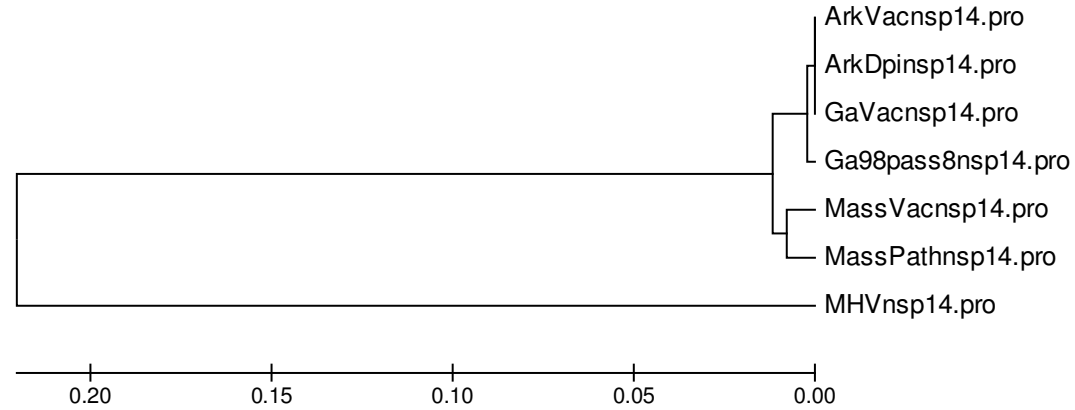


Figure 16: Phylogenetic Tree of Nsp 14

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.46191406 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 512 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

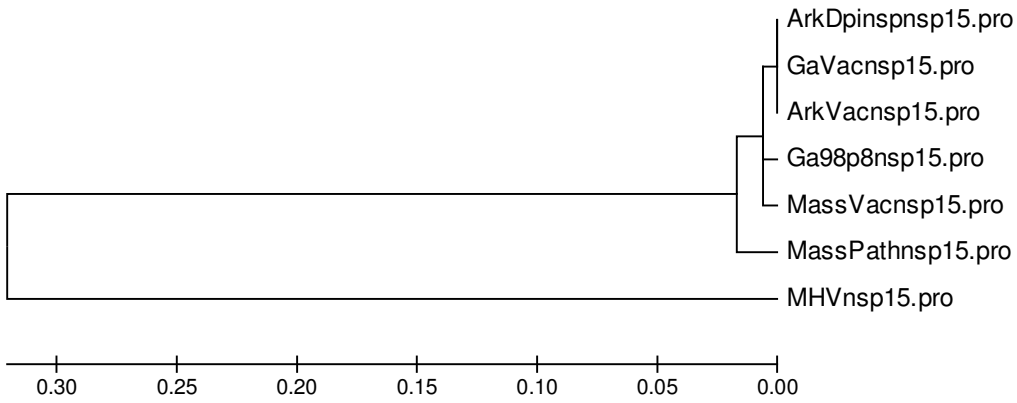


Figure 17: Phylogenetic Tree of Nsp 15
The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.67042042 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 333 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

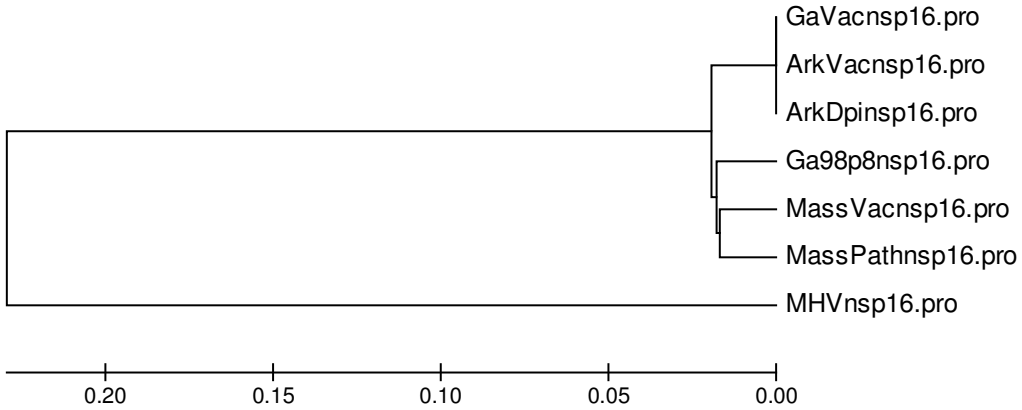


Figure 18: Phylogenetic Tree of Nsp 16
The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree with the sum of branch lengths = 0.50939850 is shown. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages [2]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 266 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [3].

1. Saitou N & Nei M (**1987**) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.

2. Takezaki N, Rzhetsky A & Nei M (**2004**) Phylogenetic test of the molecular clock and linearized trees. *Molecular Biology and Evolution* **12**:823-833.

3. Tamura K, Dudley J, Nei M & Kumar S (**2007**) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**:1596-1599

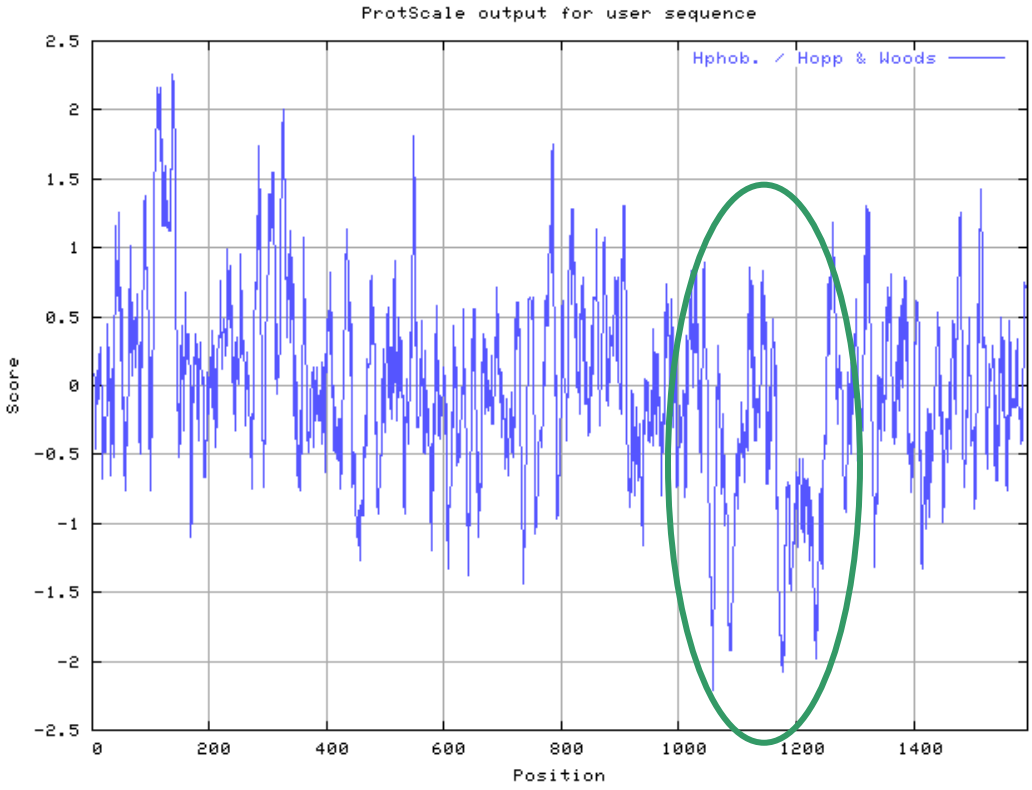


Figure 19: Hydrophobicity Plot of Nsp3
The hydrophobicity plot according to Hopp and Woods calculations of nonstructural protein three. The oval shows hydrophobic regions.

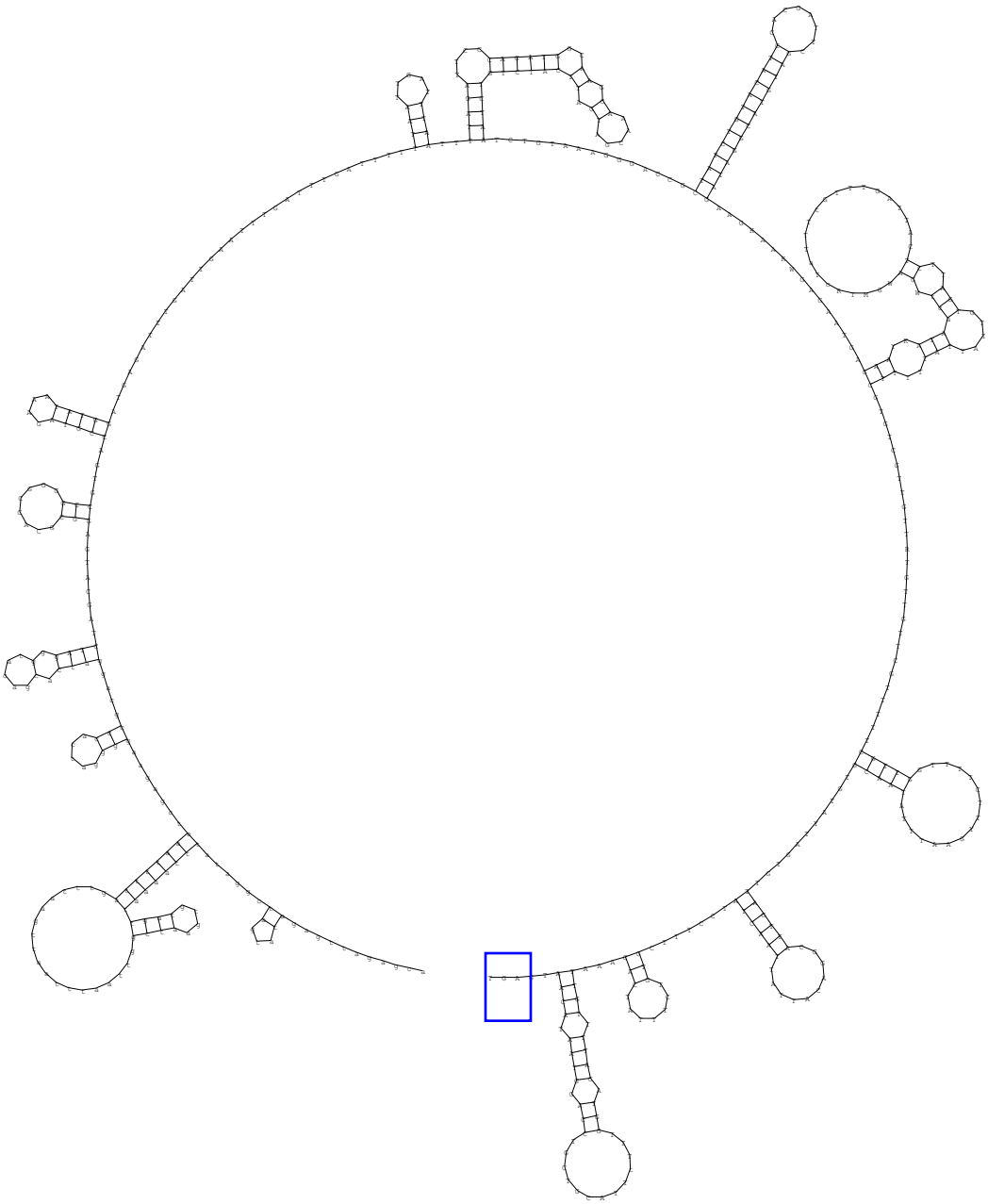


Figure 20: Predicted Secondary Structure of the 3' UTR
The 3' untranslated region consisting of 509 nucleotides secondary structure predictions based on Output of sir graph (®)by D. Stewart and M. Zuker (http://dinamelt.bioinfo.rpi.edu/results/quikfold). The stop codon of the nucleocapsid is highlighted by the blue box.