

GENOMIC SEQUENCING AND PHYLOGENETIC ANALYSIS REVEALS  
RECOMBINATION AND MUTATIONAL CHANGES CONTRIBUTING TO THE  
EMERGENCE AND EVOLUTION OF CORONAVIRUSES

by

SHARMI WEBB THOR

(Under the Direction of Mark Jackwood)

ABSTRACT

Recombination in the family Coronaviridae has been well documented and is thought to be a contributing factor in the emergence and evolution of different coronaviral genotypes as well as different species of coronavirus. However, there is limited data available on the frequency and extent of recombination in coronaviruses in nature. In this study, the full-length genomes of three alphacoronaviruses, three betacoronaviruses and eight avian gamma-coronavirus infectious bronchitis virus (IBV) isolates were sequenced and along with other full-length coronavirus genomes available from GenBank were analyzed for recombination. Evidence of recombination was found in almost every sequence analyzed and was distributed throughout the entire genome. Areas that have the highest occurrence of recombination are located in similar regions for the three coronavirus genera. For both the alpha- and beta coronaviruses these regions were the region that encodes nonstructural protein 3 and the gene for the structural spike glycoprotein. The areas identified for the gammacoronaviruses were the regions of the genome that code for nonstructural proteins 2, 3 and 16, and the structural spike glycoprotein. Our analysis revealed novel recombination events occurring in the evolution of

coronaviruses and revealed no evidence of recombination between the alpha- and betacoronaviruses. The extent of the recombination observed suggests that recombination may be one of the principal mechanisms for generating genetic and antigenic diversity within coronaviruses. These data indicate that reticulate evolutionary change, due to recombination in coronaviruses, likely plays a major role in the origin and adaptation of the virus leading to new genetic types and strains of the virus.

**INDEX WORDS:** Alpha-coronavirus; beta-coronavirus; gamma-coronavirus; avian coronavirus; infectious bronchitis virus; genome; recombination; reticulate evolution

GENOMIC SEQUENCING AND PHYLOGENETIC ANALYSIS REVEALS  
RECOMBINATION AND MUTATIONAL CHANGES CONTRIBUTING TO THE  
EMERGENCE AND EVOLUTION OF CORONAVIRUSES

by

SHARMI WEBB THOR

B.S., North Georgia College, 1988

M.Ed., The University of Georgia, 1993

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

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2011

© 2011

Sharmi Webb Thor

All Rights Reserved

GENOMIC SEQUENCING AND PHYLOGENETIC ANALYSIS REVEALS  
RECOMBINATION AND MUTATIONAL CHANGES CONTRIBUTING TO THE  
EMERGENCE AND EVOLUTION OF CORONAVIRUSES

by

SHARMI WEBB THOR

Major Professor:	Mark Jackwood
Committee:	Maricarmen Garcia
	S.Mark Tompkins
	Lilian Jaso-Friedman
	David S. Peterson

Electronic Version Approved:

Maureen Grasso  
Dean of the Graduate School  
The University of Georgia  
December 2011

## DEDICATION

I would like to dedicate this body of work to my husband, Joe Thor. Without his emotional support, love, and sense of humor, I may not have ever attained my life goal.

## ACKNOWLEDGEMENTS

I would like to thank Dr. Mark Jackwood for accepting me as his student and never giving up on me. He never lost faith in me, even when I lost faith in myself. I am grateful to all my colleagues who have walked this road with me; especially Enid McKinley and Jamie Phillips-Brantley. Each has contributed to my journey in unforgettable ways and I am indebted to them always. I am especially indebted to Debbie Hilt whose kind words and constant love and encouragement brightened even the darkest of days.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	v
LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
CHAPTER	
1 INTRODUCTION .....	1
2 REVIEW OF LITERATURE .....	3
Coronavirus Etiology .....	3
Coronaviral Pathogenicity and Host Immune Response .....	24
Coronavirus Diagnosis and Prevention.....	32
Coronavirus Evolution and Emergence .....	36
References.....	40
3 RECOMBINATION IN AVIAN GAMMA-CORONAVIRUS INFECTIOUS BRONCHITIS VIRUS.....	66
Abstract.....	67
Introduction.....	68
Results and Discussion .....	70
Conclusions.....	76
Materials and Methods.....	77
Acknowledgements.....	82

References.....	83
4 GENOMIC, PHYLOGENETIC, AND RECOMBINATIONAL ANALYSIS OF ALPHA-AND BETACORONAVIRUSES .....	117
Abstract.....	118
Introduction.....	119
Materials and Methods.....	122
Results and Discussion .....	126
Conclusion .....	142
References.....	144
5 SUMMARY AND FINAL ANALYSIS.....	176
Summary.....	176
Analysis: Direction of future research .....	178
References.....	182
<b>APPENDICES</b>	
A Chapter 3 Supplementary Tables .....	184
Supplemental Table 3.1: Percent nucleotide similarity for full-length genomes among gamma-coronavirus strains .....	185
Supplemental Table 3.2: Primer sequences used in this study.....	188
B Chapter 4 Supplementary Tables .....	200
Table 4.1: Primer sequences used in this study .....	200

## LIST OF TABLES

	Page
Table 3.1: Genes and coding regions for 8 strains of avian infectious bronchitis virus examined in this study .....	90
Table 3.2: Recombination breakpoints, genes, and major and minor related sequences in other IBV strains .....	95
Table 3.3: Number of transferred fragments associated with individual areas of the genome for all of the strains examined .....	115
Table 3.4: Viruses sequenced in this study .....	116
Table 4.1: Viruses sequenced in this study .....	150
Table 4.2: GenBank accession numbers for full-length genomes used in analysis .....	151
Table 4.3: Genome organization of the alpha- and betacoronavirus genomes sequenced in this study .....	152
Table 4.4: Recombination breakpoints, genes, and major and minor related sequences for (a) alpha- and (b) betacoronaviruses .....	165
Table 4.5: Alpha (a) and beta (b) genomic regions identified as part of a transferred fragment .....	174

## LIST OF FIGURES

	Page
Figure 2.1: Schematic of older Order <i>Nidovirales</i> classification.....	63
Figure 2.2: Schematic of current Order <i>Nidovirales</i> classification.....	64
Figure 2.3: Coronavirus genomes .....	65
Figure 3.1: Neighbor-joining method used to infer evolutionary history using full genomic sequence data available for the gamma-coronaviruses .....	92
Figure 3.2: Neighbor-net for the avian gamma-coronaviruses .....	94
Figure 3.3: Recombination breakpoint distribution plot for IBV .....	114
Figure 4.1: Neighbor-joining phylogenetic tree of (a) full-length sequence data and (b) the spike protein for the alphacoronaviruses .....	153
Figure 4.2: Neighbor-joining phylogenetic tree of (a) full-length sequence data and (b) the spike protein for the betacoronaviruses.....	156
Figure 4.3: Neighbor-net analysis for alphacoronavirus full length genomes.....	159
Figure 4.4: Neighbor-net analysis for betacoronavirus full length genomes.....	162
Figure 4.5: Recombination breakpoint distribution plot for (a) alpha and (b) betacoronaviruses.....	172

## CHAPTER 1

### INTRODUCTION

In November of 2002 an unusual form of atypical pneumonia emerged in Guangdong Province, China. By March of the following year, the disease had spread to Hong Kong, Vietnam, Taiwan, Singapore and Canada (Kalb, 2003; Ksiazek et al., 2003a). This disease, named severe acute respiratory syndrome (SARS), eventually spread to 30 countries on 5 continents; infecting more than 8000 people and killing 774 (Peiris et al., 2004; WHO, 2003). The causative agent was identified as a novel coronavirus and named SARS coronavirus (SARS-CoV) (Marra, 2003). The media and the world were wondering where the virus came from, how it spread, and how to prevent it (Bizri and Musharrafieh, 2003; Cowley, 2003).

Viruses in the Family *Coronaviridae* are of global economic importance. These viruses are widespread both with respect to geography and host tropism. Because coronaviruses have high rates of both mutation and recombination, it is not surprising that they have proven to be a major source of emerging zoonotic diseases. Although the spike region for many of the coronaviruses has been sequenced, there are very few full genomic sequences available. This is especially true for the coronaviruses that infect animals that are in close contact with humans; including dogs, cats, rats, mice, and chickens. Severe acute respiratory syndrome coronavirus (SARS-CoV) emerged as a result of an animal to human zoonotic shift. Because animal coronaviruses can serve as a reservoir of genetic information for viral zoonosis, an examination of the complete genomic sequence of important animal coronaviruses from alpha-, beta-, and

gamma-coronavirus groups (formerly identified as group I, II, and III, respectively) is essential for the prevention of further disease outbreaks.

Genetic mutations and recombination among coronaviruses leads to the emergence of new viruses capable of infection and causing disease in animals including humans. Therefore the goal of this work was to sequence, assemble and annotate the genomes of important animal coronaviruses that have not previously been sequenced. The viruses were selected from those that infect animals in close contact with humans and they represent each of the three coronavirus groups. Each newly sequenced viral genome was analyzed for areas of potential recombination, percent similarity and phylogenetic relationship to other coronaviruses.

CHAPTER 2  
REVIEW OF LITERATURE  
**PART I**  
**Coronavirus Etiology**

**1. Classification**

Coronaviruses (CoVs) are found in a wide range of animals including humans and cause a variety of diseases. Coronaviruses belong to the *Coronaviridae* family of the *Nidovirales* order. They are positive sense, single-stranded RNA viruses (ssRNA), and their genome ranges from 24 and 32 kilobases (kb) in length; the largest of all the RNA viruses (Lai, 2001). They were originally classified into phylogroups based on both antigenic and genetic similarities – Groups 1, 2 and 3. Recently, the Coronavirus Study Group of the International Committee for Taxonomy of Viruses has renamed the phylogroups, 1, 2 and 3 into three Genera; *Alpha-coronavirus*, *Beta-coronavirus* and *Gamma-coronavirus*, respectively, in the subfamily *Coronavirinae* (Carstens, 2009). Only coronaviruses for which a complete genome sequence is available were considered for taxonomy. Alpha- and beta -coronaviruses infect a wide range of mammals (including humans) while gamma-coronaviruses primarily infect birds (Lai, 2001). Recently, viruses isolated from a Beluga whale (Mihindukulasuriya et al., 2008) and an Asian Leopard Cat (Dong et al., 2007) were classified as gamma-coronaviruses. Figures 2.1 and 2.2 depict the previous *Nidovirales* classification scheme compared to the newly revised scheme (see figures 2.1 and 2.2).

The genus *Alphacoronavirus* is subdivided into eight species; *alphacoronavirus I*, *Human coronavirus 229E* (HCoV 229E), *Human coronavirus NL63* (HCoV NL63), *Miniopterus bat coronavirus I*, *Miniopterus bat coronavirus HKU8*, *Porcine epidemic diarrhea virus* (PEDV), *Rhinolophus bat coronavirus HKU2*, and *Scotophilus bat coronavirus 512*. *Alphacoronavirus I* includes canine coronavirus (CCoV), feline coronavirus (FCoV) and transmissible gastroenteritis virus (TGEV) in pigs. The other *alphacoronavirus* species are identified by the prototypical virus each was named for.

Genus *Betacoronavirus* is subdivided into seven species; *Betacoronavirus I*, *Human coronavirus HKU1* (HCoV HKU1), *Murine coronavirus*, *Pipistrellus bat coronavirus HKU5*, *Rousettus bat coronavirus HKU9*, *Severe acute respiratory syndrome-related coronaviruses*, and *Tylonycteris bat coronavirus HKU4*. *Betacoronavirus I* is comprised of viruses of humans, *Human coronavirus-OC43* (HCoV-OC43), cattle, *bovine coronavirus* (BCoV) and related isolates such as *giraffe coronavirus* (GiCoV), swine, *porcine hemagglutinating encephalomyelitis virus* (PHEV), horses, *equine coronavirus* (ECoV), and a newly recognized virus of dogs, *canine respiratory coronavirus* (CrCoV) (Erles et al., 2007). Representative viruses of the *Murine coronavirus* species include viruses of mice, *murine hepatitis virus* (MHV) and all the reported strains, rats, *rat coronavirus* (RaCoV), and puffins, *puffinosis virus* (PuCoV). Each of the other species are composed of the representative virus for which they were named except for the SARS related coronavirus. This group contains not only the viruses directly linked to the 2003 outbreak in the human population and the virtually identical animal viruses (viruses from palm civets, raccoon dogs, Chinese ferret badger, and bats) but also the more distantly related viruses such as *SARS-Rh-BatCoV HKU3*, *SARSr-RH-BatCoV*, and *SARr-Rh-BatCoV 273*, that were recently isolated from bats (Lau et al., 2005; Tang et al., 2006).

The *Genus Gammacoronaviruses* are predominantly the avian coronaviruses. However, with the identification of new viruses, some of which do not originate in birds, that have been classified as gamma-coronaviruses this group has also been subdivided into two species: *Avian coronavirus* and *Beluga whale coronavirus SW1*. Viruses classified in the avian coronavirus species are Infectious Bronchitis Virus of fowl and other coronaviruses of birds, turkey coronavirus (TCoV), duck coronavirus, goose coronavirus, pheasant coronavirus. *Beluga whale coronavirus SW1* is represented by only one isolated virus which mirrors the name of the species (Mihindukulasuriya et al., 2008).

The coronaviruses from the munia, thrush, and bulbul originally were grouped in gamma-coronavirus group c (Woo et al., 2009a). It is worth mentioning that these viruses have not been included in the current classification scheme despite the existence of the sequence of the full genome. A *Deltacoronavirus* genera has been proposed to include these avian viruses however it has not been accepted and remains under consideration (de Groot, 2011 (in press))

## **2. Viral Morphology and Genome Structure**

The characteristic coronavirus morphology is an enveloped virus that is studded with long, petal-shaped spikes. These spikes give coronaviruses the appearance of having a crown (Latin, *Corona*). Other morphological characteristics that characterize the Coronaviridae are the organization of the genome and the properties of their structural proteins.

The coronavirus genome is comprised of a single positive sense strand of capped and polyadenylated RNA. At both the 5' and 3' ends of the genome are un-translated regions (UTR) that are important for RNA replication and transcription. The remainder of the genome is comprised of 7 to 10 open reading frames (ORFs). Gene 1 codes for a polyprotein that is the

precursor of the viral polymerase (Pol). It is about 20 to 22 kb in length, making up about two thirds of the genome, and consists of two overlapping ORFs (1a and 1b) (Lai, 2001). The distal one third of the genome codes for the four structural proteins that are present in all coronaviruses; Spike glycoprotein (S), Envelope (E), Membrane (M), and Nucleocapsid (N). These structural genes are interspersed with several ORFs that encode various nonstructural proteins. In some beta-coronaviruses this includes the hemagglutinin-esterase (HE) gene. The number, nucleotide sequence, gene order and method of expression of these ORFs differ among the coronavirus groups. However, the order of the gene coding the polymerase and the four structural genes is relatively conserved. This general order is 5'-Pol-S-E-M-N-3'. Open reading frames (ORFs) 1a and 1b are joined by a pseudoknot structure that acts as a ribosomal slippery site, which results in expression of either 1a or 1b through a minus 1 frame shift (Holmes, 2001). This gene is translated as a polyprotein that is co-translationally cleaved by a chymotrypsin-like protease (3CLpro) and one or two papain-like cysteine proteases found in the nascent protein (Holmes, 2001). Cleavage of the polyprotein results in the formation of 15 or 16 non structural proteins (nsps) required for viral genomic RNA synthesis and sub-genomic RNA synthesis (Bartlam et al., 2007). Alpha- and beta-coronaviruses encode 16 nsps while the gamma-coronaviruses, which are lacking nsp1, encode 15. Amino acid sequence identity for replicase polyproteins from different coronavirus groups varies from less than 30% to over 60% (Prentice et al., 2004), with the most variability occurring in the protein domains encoded upstream of nsp3. It has been shown that these nsps are either putative or confirmed RNA-processing and RNA-editing enzymes that are speculated to increase the fidelity of the RNA dependent RNA polymerase (RdRp), encoded by nsp 12 (Bhardwaj et al., 2008; Bhardwaj et al., 2006; Eckerle et al., 2010). If the activity of one of these enzymes is abolished an observable

decrease in polymerase fidelity is likely. The nsps encoded by ORF 1b (nsp13-16) are directly involved in the replication and transcription of the genome.

**a. 1a/ab polyproteins: non structural proteins (nsps)**

nsp1

The nsp 1 protein is encoded by all known mammalian coronaviruses (Snijder et al., 2003b). Studies focused on the function of the nsp1 protein have found it to inhibit host gene expression and modulate the host innate immune responses (Tohya et al., 2009; Züst et al., 2007) by promoting the degradation of host mRNA and inhibiting translation (Narayanan et al., 2008a). This is a unique biological function that suggests the evolution of a new virulence mechanism in coronaviruses. The mechanism of nsp1 is reported to involve the suppression of interferon-beta (IFN- $\beta$ ) mRNA without inhibiting IFN regulatory factor-3 (IRF-3) dimerization (Kamitani et al., 2006). Law, *et al* reported SARS-CoV nsp1 to also be involved in the increased expression of cytokines in human lung epithelia via the activation of nuclear factor kappa-light-chain-enhancer of activated B-cells (NF- $\kappa\beta$ ) (Law et al., 2007). Thus not only is nsp1 a strong IFN antagonist but also an inducer of the cytokine storm that accompanies many coronavirus infections. This has increased interest in nsp1 as a vaccine target. In a recent study, a coronavirus was shown *in vivo* to be attenuated by the removal of less than 100nt from the nsp1 coding region (Züst et al., 2007). The nsp1 region of the coronavirus genome retains similar biological function throughout the mammalian coronaviruses despite variations in primary sequence (Snijder et al., 2003b; Tohya et al., 2009) suggesting a viral adaptation that increases pathogenicity by favoring viral replication.

## nsp2

The nsp2 protein is encoded by all coronaviruses. Although the N-terminal domain of the nsp2 protein from IBV has been crystallized (Yang et al., 2009), most of what is functionally known about nsp2 has come from studies on MHV or SARS-CoV . The nsp2 protein was first identified as a 65-kilodalton (kDa) protein cleaved from the 1a polyprotein of MHV (Denison et al., 1995). It is cleaved by a papain-like protease 1 (PLP1) from a 275-kDa nsp2/3 intermediate protein (Denison et al., 1995). The introduction of a mutation in the cleavage site of the nsp2/3 intermediate was found to slow the growth of MHV (Graham and Denison, 2006). Although nsp2 is dispensable for viral replication and the formation of viral replication-transcription complexes (RTC) (Gadlage et al., 2008; Graham et al., 2006; Graham et al., 2005), it is recruited to double membrane vesicle (DMV) anchored RTCs (Hagemeijer et al., 2010). SARS-CoV nsp2 interacts with host proteins prohibitin 1 and prohibitin 2 (PHB1 and PHB2) (Cornillez-Ty et al., 2009). These host proteins are known to be involved in cell cycle progression (Wang, 1999), cell migration (Rajalingam et al., 2005), cellular differentiation (Héron-Milhavet et al., 2008; Sun et al., 2004), apoptosis (Fusaro et al., 2003), and mitochondrial biogenesis (Merkwirth and Langer, 2009). Rather than playing a role in viral replication, the interaction of nsp2 with these host cell proteins suggests a role in altering the host cell environment to favor the replication of the virus.

## nsp3

Consisting of five domains, nsp3 is the largest of the nsps. These domains include the acidic acid domain (Ac), the papain-like protease 1 (PLP1) domain, the ADP ribose 1'-

phosphatase (ADRP) domain, the papain-like protease 2 (PLP2) and the Y domain (Snijder et al., 2003a; Ziebuhr et al., 2001).

The Ac domain is located at the N-terminal region of nsp3 and is made up of a high number of acidic amino acid residues. The amino terminal region of nsp3 has been shown to bind selectively to the N protein. This interaction may be essential for the localization of genomic RNA to the replicase complex at an early stage of infection (Hurst et al., 2010).

Downstream of the Ac domain is PLP1. This proteolytic enzyme is inactive in IBV and SARS-CoV due to amino-acid replacements in the catalytic center (Baker et al., 1993; Bonilla et al., 1997; Snijder et al., 2003b; Ziebuhr et al., 2001) but is functional in all other coronaviruses. Along with PLP2, PLP1 is responsible for the cleavage of nsp1, nsp2, and nsp3 (Anand et al., 2003; Thiel et al., 2003). These catalytic sites were first described by in 1993 by Baker et al (Baker et al., 1993) . Coronavirus PLP domains have been identified as modulators of the innate immune response (Devaraj et al., 2007; Frieman et al., 2009) by acting as an interferon antagonist that blocks the phosphorylation and subsequent nuclear translocation of interferon regulatory factor 3 (IRF-3) (Devaraj et al., 2007).

C-terminal to this domain is the ADRP domain. The ADRP domain contains an area known as a macro domain which is often, in the context of viruses, referred to as an X domain. Macro domains are found in most life forms. The name refers to the non-histone domain of the histone macroH2A (Pehrson and Fried, 1992). A study by Egloff *et al*, showed the ADRP region of nsp3 to have a high affinity for poly-ADP-ribose (PAR) binding (Egloff et al., 2006). Although this activity may not be conserved in all coronaviruses (Piotrowski et al., 2009), all characterized coronaviruses have this ADRP domain. The only other viruses found to have this domain are the alphaviruses, rubella virus and hepatitis E (Snijder et al., 2003b).

The PLP2 domain in nsp3 has been shown to be a potent deubiquitinase (Ratia et al., 2008) that inhibits cellular type I interferon by targeting TANK binding kinase- 1 (TBK1) to negatively regulate the cellular type 1 interferon signaling pathway (Clementz et al., 2010; Wang et al., 2011; Zheng et al., 2008) An integral part of the innate immune system, TBK1 is an inhibitor of NF- $\kappa$ B kinase (IKK).

The domain found closest to the C-terminal end of nsp3 is the Y domain. This domain has been shown to be a transmembrane domain conserved in all coronaviruses (Kanjanaaluethai et al., 2007; Oostra et al., 2008). This membrane spanning domain is important for the membrane association of the replicase and interaction with the viral envelope protein (Álvarez et al., 2010).

#### nsp4

Mature nsp4 is released from pp1a by viral protease cleavage and is the first cleavage event to be experimentally recognized after viral infection (Gosert et al., 2002). The nsp4 protein consists of four hydrophobic areas that are membrane spanning (Oostra et al., 2008) The amino terminal transmembrane domain serves as the signal sequence for cleavage of this protein (Gosert et al., 2002). Experimentally, nsp4 has been observed to localize to the endoplasmic reticulum (ER), recruited to the coronavirus replicative structure, and to interact with both nsp3 and nsp6 (Hagemeijer et al., 2011).

#### nsp5

The coronavirus main protease (Mpro) is located in the nsp5 coding region of the genome. This protease is also called the 3C-like protease (3CLpro) to indicate a similarity of its

cleavage-site specificity to that observed for picornavirus 3C proteases although the structural similarities between the two families of proteases are limited (Anand et al., 2002). Two crystal structures of coronavirus 3C-like protease from TGEV (Anand et al., 2002) and HCoV 229 E have been reported (Anand et al., 2003). The protease structure has three domains: the first two form a chymotrypsin fold that is responsible for the catalytic action, and the third domain is  $\alpha$ -helical with unclear biological function. First identified in 1995 (Lu et al., 1995), the 3C-like protease required for viral replication (Long et al., 1989; Thiel et al., 2003) by mediating the cleavage of the replicase polyprotein 1a and 1b at 11 specific sites generating 12 mature non-structural proteins from nsp5-16 (Masters, 2006; Ziebuhr, 2005; Ziebuhr et al., 2000). Therefore this protease is an attractive target for antiviral drugs for the control of coronaviral infections (Kuo et al., 2009).

#### nsp6

Nsp6 has numerous hydrophobic regions with only a small number actually membrane spanning (Baliji et al., 2009; Oostra et al., 2008). The hydrophobic domain in the C-terminal tail of nsp6 is highly conserved among coronaviruses (Baliji et al., 2009). The topology of this nsp has only recently been elucidated (Baliji et al., 2009; Oostra et al., 2008) however, the function of this hydrophobic region has yet to be elucidated.

#### nsp 7-10

A group of poorly understood replicase proteins are nsp7 to 10. Highly conserved in all coronaviruses, these nsps co-localize with the replication complex and are presumably involved in viral RNA synthesis (Bost et al., 2000; Bost et al., 2001; Campanacci et al., 2003; Gosert et

al., 2002; van der Meer et al., 1999). Deming, *et al*, evaluated whether each protein domain was essential for a productive virus infection (Deming et al., 2007). Deletion of any of the four protein domains was stopped RNA synthesis and productive viral infection (Deming et al., 2007). With the exception of the nsp9/nsp10 cleavage site, the cleavage of nsp7 to nsp10 proteins from each other is necessary for replication. When nsp7 and nsp8 were rearranged virus replication was ablated (Deming et al., 2007). Studies with IBV showed the proteolytic processing between nsp10 and nsp11/12 to be dispensable for virus replication (Fang et al., 2008). Crystallographic studies of nsp7 have shown an interaction with nsp8 forming a hexadecameric supercomplex (Zhai et al., 2005). Due to the structural similarity of the nsp7/8 supercomplex with processivity factors for DNA polymerase, Zhai *et al*, suggest that the supercomplex could encircle the genomic RNA and serve as a general processivity factor for the RdRp encoded by nsp11/12. The nsp8 protein is a non-canonical RdRp that synthesizes short primers for nsp12 (Imbert et al., 2006). The nsp9 protein is a single stranded RNA binding protein that has been found to interact with tRNA (Campanacci et al., 2003; Egloff et al., 2004; Gorbalenya et al., 1989; Ponnusamy et al., 2006; Sutton et al., 2004; Thiel et al., 2003); however, it is unknown whether binding is restricted to specific segments of the genome or augmented with other proteins. The involvement of nsp10 with negative strand RNA synthesis has been proposed by Sawicki, *et al* (Sawicki et al., 2005). Studies on nsp10 reveal that the protein includes two Zn<sup>2+</sup> fingers (Joseph et al., 2006; Matthes et al., 2006), exhibits nucleic acid binding affinity ( Joseph, 2006 and Matthes, 2006), and can crystallize to form a spherical dodecameric structure made up of as many as twelve nsp10-nsp11 subunits (Su et al., 2006). The nsp10 protein has also been observed as monomers and homodimers (Joseph et al., 2006).

### nsp 11/12

The nsp11 is a small protein (1-2 kDa) of unknown function. However, due to their activity in the replication transcription complex (RTC) nsp11 and nsp12 are usually discussed together. Nsp11/12 functions as the RdRp used for viral replication (Cheng et al., 2005; Gorbalenya et al., 1989; Koonin, 1991; te Velthuis et al., 2010). Although RdRp activity has been mapped to this area of the genome, further characterization of these enzymes has been delayed due to the inability to express and purify a soluble and active protein. Cheng, et al, provided preliminary evidence of the *in vitro* activity of nsp11/12 by using an nsp12 fusion protein. However the protein remained unstable and resulted in the fragmentation of the protein into three parts (Cheng et al., 2005). A study by te Velthuis, *et al*, is the first to successfully express an active full length nsp12 protein using an *in vivo* bacterial cleavage system. They established that the protein readily synthesizes nascent RNA in a primer dependent fashion (te Velthuis et al., 2010).

### nsp13

Although multiple enzymatic activities such as RNA and DNA duplex unwinding, NTPase and dNTPase activities, and an RNA 5'- triphosphate activity that may be involved in the 5'-cap structure formation, have been assigned to this protein (Ivanov and Ziebuhr, 2004), nsp13 is predominantly a super-family 1-like helicase 229E (HEL1) (Ivanov et al., 2004b; Seybert et al., 2000). The protein has two domains; the N-terminal domain is a putative zinc binding domain and the C-terminal domain is a helicase (Ivanov et al., 2004b). A point mutation in nsp13 causes the helicase to be non-functional and lethal is to the infectivity of the virus (Fang et al., 2007). This mutation caused an Arg132Pro change in a domain with unknown function

implying that this region might play a role in the functionality of the helicase protein (Fang et al., 2007).

#### nsp14

The nsp14 has multiple functional domains. In the N-terminal region of the protein, nsp14 is a metal ion-dependent 3' to 5' exonuclease (ExoN) protein (Denison et al., 2011; Minskaia et al., 2006; Snijder et al., 2003b). Through comparative sequence analysis and characterization of mutant nsp14s, it was concluded that this protein belongs to the DEDD superfamily of exoribonucleases (Minskaia et al., 2006; Snijder et al., 2003b; Zuo and Deutscher, 2001). The nsp14 ExoN plays a critical role in the prevention or repair of nucleotide incorporation errors during genome replication (Eckerle et al., 2007) and an important role in viral RNA synthesis as amino acid substitutions to this protein causes a decrease in viral replication (Eckerle et al., 2007). Interestingly, a guanine-N7-methyltransferase (N7-MTase) activity has also been attributed to nsp14 and mapped to the C-terminal domain (Chen et al., 2009). This type of enzyme functions in the viral genomic 5'-cap formation which is essential for viral replication.

#### nsp15

Due to its nidovirus-wide conservation, nsp15 is considered a major genetic marker of nidoviruses (Ivanov et al., 2004a). The endoribonuclease is similar to that of *Xenopus laevis* (XendoU) and has been similarly named as U-specific *Nidovirales* endoribonuclease (NendoU). NendoU is a uridylylate-specific endoribonuclease that in SARS-CoV is Mn<sup>2+</sup> dependent (Bhardwaj et al., 2004; Ricagno et al., 2006). Interestingly, the turkey coronavirus (TCoV)

homolog of this enzyme did not require  $Mn^{2+}$  for *in vitro* enzymatic activity (Cao et al., 2008). The active form of nsp15 is a hexamer that preferentially cleaves ssRNA although dsRNA can also be a substrate (Bhardwaj et al., 2004). NendoU has been reported to cleave at both the 5' and 3' ends of uridine (Bhardwaj et al., 2004; Ivanov et al., 2004a). However, a 2006 study by Bhardwaj et al, further characterized the cleavage of nsp15 and found that the enzyme only cleaves 3' to uridylylate with a higher affinity for uridylylate bases that are not involved in base-pairing and are consecutive in the RNA sequence (Bhardwaj et al., 2006). The biological function of NendoU was characterized through mutational analysis which resulted in either complete ablation or reduction in viral RNA synthesis. Subgenomic RNA synthesis was consistently more strongly affected than genome replication (Posthuma et al., 2006).

### nsp16

The nsp16 is an S-adenosylmethionine (AdoMet)-dependent RNA (-2'-O-ribose) methyltransferase (2'O-MTase) (Decroly et al., 2008; Masters, 2006; Snijder et al., 2003b). It contains a highly conserved catalytic tetrad (KDKE) that is consistent with other RNA 2'-MTases (Bujnicki and Rychlewski, 2002; Egloff et al., 2002; Ferron et al., 2002). This RNA cap modifying enzyme is only active when complexed with the  $Zn^{2+}$  finger region of nsp10 (Debarnot et al., 2011; Decroly et al., 2011).

### **b. Spike**

The S protein is a structural protein encoded downstream of the 1ab polyprotein. The protein is glycosylated and forms large spikes on the virion surface. The primary functions of the S protein are to bind to specific receptors on the host cell and to induce fusion of the viral

envelope with cell membranes (Lai, 2001). In beta- and gamma- coronaviruses, the S protein is post-translationally cleaved by host cell serine proteases into two subunits designated S1 and S2. The S1 subunit is comprised of the N-terminal half of the molecule and forms the outer globular portion of the spike. It contains epitopes specific to the receptors on susceptible cells as well as epitopes that can induce neutralizing antibodies in the host. The sequence of the S1 subunit is highly variable among different serotypes of coronaviruses (Lai, 2001). The stalk of the spike is formed by the S2 subunit. The S2 subunit is acylated and contains two heptad repeat motifs (Bosch et al., 2004; Luo and Weiss, 1998). Although the fusion activity of spike is affected at various sequences in both S1 and S2, it is likely that membrane fusion is conferred by hydrophobic regions in S2 (Gallagher et al., 1991; Gombold et al., 1993; Routledge et al., 1991). Spike also plays a role in viral pathogenicity by promoting B-cell mediated cytotoxicity (Holmes et al., 1986) and by binding to the Fc fragment of IgG (Oleszak et al., 1995; Oleszak et al., 1993).

### **c. Envelope**

Virus particle assembly involves the envelope (E) protein. The E protein is associated with the viral envelope and has been implicated in causing host cell apoptosis (Lai, 2001). Although both M and E are necessary for viral budding from infected cells (Vennema et al., 1996), E alone was sufficient for vesicle release (Maeda et al., 1999). The E protein is involved in ion channel production which is required for optimal virus replication (Wilson et al., 2004).

#### **d. Membrane**

The membrane glycoprotein (M) is also involved in virus particle assembly. In addition, it has been shown to be involved in host interactions, which appear to be associated with the glycosylation state of the protein (de Haan et al., 2003). In TGEV, the M protein has interferogenic activity. A decrease in interferogenic activity of M was seen after the introduction of single amino acid mutation in the M protein ectodomain that impairs N-glycosylation (Laude et al., 1992). Only a short portion of the N-terminal domain of the protein is exposed on the exterior of the viral envelope. The remainder of the protein is composed of multiple domains that are membrane spanning with the C-terminal region found inside the virus (Lai and Cavanagh, 1997). This protein is not transported to the plasma membrane but is found in the Golgi region of the cell. This Golgi integral membrane protein has been found to interact with both the viral nucleocapsid and spike proteins (de Haan et al., 1999; Kuo and Masters, 2002) triggering both viral RNA packaging and viral particle assembly (Lai, 2001; Narayanan et al., 2003).

#### **e. Nucleocapsid**

Bound to the viral RNA and forming the viral nucleocapsid is the nucleocapsid (N) phosphoprotein. Early post-infection, it is found to co-localize with the replication complex at the site of RNA synthesis (Bost et al., 2001) and is necessary for viral genome replication (Almazan et al., 2004). The N protein acts as a RNA chaperone promoting the proper folding of the nascent RNA (Zúñiga et al., 2007). The structure of the protein has been found to be consistent with that of other RNA chaperones (TOMPA and CSERMELY, 2004). There is evidence that N may play a role in viral pathogenesis. The N protein from MHV strains that

cause hepatitis, and eventually fulminant hepatic failure, induces the expression of prothrombinase fibrinogen-like protein 2/fibroleukin (*Fgl2*) (Ning et al., 2003). Expression of *fgl2* in hepatic tissue plays a critical role in the development of liver failure (Ning et al., 1998). The N protein also plays a role in viral transcription. The N protein is necessary in order to efficiently recover virus from infectious cDNA clones (Yount et al., 2003; Yount et al., 2002). The N protein enhances the replication of HCoV-229E (Schelle et al., 2005) and is implicated in the development of fulminant hepatitis during MHV infection (Ding et al., 1997; Ning et al., 1999). The N protein of all coronavirus groups has been shown to localize to the nucleolus as well as the cytoplasm (Hiscox et al., 2001; Wurm et al., 2001). Wurm, *et al* , suggested that N induces a cell cycle delay or arrest possibly in the G<sub>2</sub>/M phase by inhibition of cytokines (Wurm et al., 2001). The N protein also acts as a type 1 interferon antagonist (Ye et al., 2007a).

#### **f. Hemagglutinin-esterase**

Some beta-coronaviruses have a hemagglutinin-esterase (HE) glycoprotein. The HE glycoprotein appears to be a “luxury” protein in that it is not needed for infection; however, there is evidence that suggests the HE protein may be involved in either virus entry or release from infected cells (Lai, 2001). The HE glycoprotein forms small spikes on the surface of some betacoronaviruses. The HE protein is responsible for hemagglutination and may cause hemadsorption which suggests its involvement in either viral entry or release from infected cells (Lai, 2001). The role of HE is in either acute and/or chronic disease as a determinant of cellular tropism (Yokomori et al., 1995), or in the spread of the virus by augmenting attachment and/or exit from the cell (Kienzle et al., 1990). The coronavirus HE proteins share 30% primary

sequence similarity with the hemagglutinin protein from influenza C (Luytjes, 1988) suggesting recombination between influenza C and an ancestral coronavirus.

#### **g. Virus-specific accessory proteins**

Interspersed between and within the structural proteins described above, are genes that encode accessory proteins. Both the number and location of the genes vary depending on the strain of virus. These genes have no sequence similarity with other viral or cellular proteins and have non-essential roles in virus replication (Ontiveros et al., 2001; Yount et al., 2005). These group specific proteins may not be essential but in some viruses their deletion has been shown to be attenuating in the natural host (de Haan et al., 2002).

#### **h. Glycosylation and cleavage differences**

Differences between the coronaviral groups are distinguished by the way in which proteins are glycosylated and/or cleaved. Alpha-coronaviruses do not have an HE protein, the M protein is N-glycosylated and the S glycoprotein is not cleaved (Lai, 2001). Some beta-coronaviruses have an HE protein, the M protein is O-glycosylated, and the S glycoprotein is cleaved (Lai, 2001). The gamma-coronaviruses do not have an HE protein, the M protein is N-glycosylated and the S glycoprotein is cleaved.

### **3. Coronavirus Replication and Viral Assembly**

In order for a virus to infect a cell and thus begin replication, the virus must first attach to the host cell via a specific receptor on the host cell membrane. Alphacoronaviruses feline coronavirus (FCoV), feline infectious peritonitis virus (FIPV), canine coronavirus (CCoV),

transmissible gastroenteritis virus (TGEV) and Human coronavirus strain 229E (HCoV-229E) all use a cell membrane-bound metalloprotease, aminopeptidase N (APN) of their respective host species (Benbaccer et al., 1997; Lai and Cavanagh, 1997) as receptor. Both respiratory and intestinal epithelium express APN on their surface. This cellular receptor is also found on myelocytic cells, on kidney tubular epithelium and at synaptic junctions (Shapiro et al., 1991). HCoV-NL63, an alphacoronavirus, and SARS-CoV, a betacoronavirus, use angiotensin-converting-enzyme-2 (ACE2) as host cell receptor (Hofmann et al., 2005; Li et al., 2003; Turner et al., 2004a). ACE2 is found in the lungs, heart, kidney, and small intestine (Hamming et al., 2004). ACE2 is postulated to have a protective role in the inflamed lung and binding of the SARS-CoV S protein to ACE2 is thought to contribute to disease severity (Imai et al., 2005; Kuba et al., 2005). The ACE2 receptor protein is highly conserved across mammalian species and the susceptibility of ACE2 to SARS-CoV binding is determined by a small group of key amino acids (Li et al., 2005a). Minor amino acid changes in the bat ACE2 receptor were sufficient to change it from being non-functional for SARS-CoV to a functional receptor (Yu et al., 2010). Bovine coronavirus (BCoV) and coronaviruses of other ruminants as well as the human coronavirus strain HCoV-OC43 utilize a 9-O- acetylated sialic acid receptor (Lai and Cavanagh, 1997) to facilitate the infection of their respective host cells. This receptor is widespread in mammalian tissues but variable in concentration (Klein et al., 1994). The cellular receptor for MHV is a murine biliary glycoprotein belonging to the carcinoembryonic antigen (CEA) family in the Ig superfamily: carcinoembryonic antigen adhesion molecule 1 (CEACAM1) (Tan et al., 2002). This receptor is expressed in the liver and gastrointestinal tract, macrophages, dendritic cells, B cells, and activated T cells (Lai and Cavanagh, 1997; Nakajima et al., 2002; Turner et al., 2004b). The host cell receptors for alphacoronaviruses

porcine epidemic diarrhea virus (PEDV) and rabbit coronavirus (RbCoV), betacoronaviruses canine respiratory coronavirus (CRCoV), porcine hemagglutinating encephalomyelitis virus (PHEV), human coronavirus strain HCoV-HKU1, both the coronaviruses of manx shearwaters (PCoV) and rats (RtCoV and SDAV) as well as all known gammacoronaviruses remain unknown.

Coronavirus replication takes place entirely in the cytoplasm of the host cell and is thought to occur on double membrane vesicles (DMVs) (Gosert et al., 2002; Snijder et al., 2006). The DMVs likely originate from the endoplasmic reticulum (ER) (Knoops et al., 2008) and their formation requires membrane curvature possibly initiated by the insertion of specific viral proteins with membrane spanning regions such as nsp3, 4, or 6 (Kanjanaaluethai et al., 2007; Oostra et al., 2008; Oostra et al., 2007). Coronavirus genomic replication also involves replication transcription complexes (RTCs) made from most, if not all, replicase-transcription proteins encoded by the 1ab polyprotein (Gosert et al., 2002). These viral protein complexes accumulate at perinuclear regions and are associated with DMVs (Brockway et al., 2003; Snijder et al., 2006). The viral N protein has also been found to be associated with the RTCs (Bost et al., 2000) and is consistent with the N protein playing a role in RNA synthesis (Almazan et al., 2004).

The fusion of the virus to the host cell is induced by a conformational change in S after binding to the receptor (Zelus et al., 2003). Following fusion, the viral nucleocapsid is released into the cytoplasm and the viral RNA is uncoated. The positive sense viral genome acts as a messenger RNA (mRNA) for the translation of the RdRp. The positive sense genomic RNA is also used as a template for the transcription of a full-length negative sense copy of the viral

genome. This negative sense copy is used for leader-primed synthesis of subgenomic mRNAs (sg-mRNA). All of the sg-mRNAs have a common leader sequence at their 5'-ends that is identical to the sequence present at the 5'-end of the genome. Between the leader sequence and the remainder of the mRNA is an adenine/uracil rich area known as the transcription-regulating sequence (TRS). This sequence is about 10 nucleotides long. Parameters that are important to the functioning of the TRS are the stability of the base pairing between the template and the nascent minus strand, the sequences flanking the TRS, and the location of the TRS relative to the promoter for minus-strand synthesis (Curtis et al., 2004; Sola et al., 2005; Zuniga et al., 2004). The sg-mRNAs are of various lengths and encode the different viral proteins (Lai, 2001). A model, described by Sawicki *et al*, has the RTC initially forming to create a negative-strand template. The RTC is then converted to utilize the negative strand as template for positive strand synthesis. The complex is inactivated by the degradation of the negative strand templates (Sawicki et al., 2005).

New viral assembly begins with the N-protein and the newly synthesized genomic RNA assembling in the cytoplasm to form helical nucleocapsids. The M-glycoprotein is inserted in the endoplasmic reticulum (ER) and anchored in the membrane of the Golgi apparatus. The E protein is transported through the ER to the Golgi where it interacts with the M protein to trigger viral budding, which will enclose the nucleocapsid (Narayanan and Makino, 2001). The S and HE (if present) glycoproteins are translated on membrane bound polysomes, inserted into the ER and transported to the Golgi. These proteins are post-translationally glycosylated, trimerized and transported through the Golgi. It is in the Golgi that, in beta and gamma- coronaviruses, the S glycoprotein is cleaved into 2 subunits. The mature virions are released through an exocytosis-like fusion of smooth-walled, virion containing vesicles with the plasma membrane (Lai, 2001).

#### 4. Coronavirus Hosts

Coronaviruses have been isolated from a variety of mammalian and avian hosts. Avian infectious bronchitis virus, a highly infectious upper respiratory disease of chickens, has been found to replicate in other avian species, such as peafowl, teal and guineafowl, without the development of clinical signs (Cavanagh et al., 2002; Liu et al., 2005). Recently, avian coronaviruses were isolated from aquatic wild birds where there were frequent interspecies transmissions (Chu et al., 2011). Bovine coronavirus adapted to humans approximately 100 years ago in the form of Human coronavirus strain OC43 (Vijgen et al., 2005). Interestingly, these viruses seem to have retained the ability to infect not only the original host (cattle) but also the new host (humans) (Vijgen et al., 2005). Variants of the same coronavirus were isolated from different bat species likely indicating that these viruses crossed the host species barrier in their recent evolutionary history (Tang et al., 2006; Woo et al., 2006). Crossing a species barrier does not always lead to the creation of a new coronavirus species. Several examples of this are the SARS-CoV isolated from humans and palm civets (Guan et al., 2003; Lau et al., 2005; Li et al., 2005b; Song et al., 2005; Yeh et al., 2004) and a feline coronavirus-like agent that devastated a cheetah breeding colony (Heeney et al., 1990; Pearks Wilkerson et al., 2004). With increased viral surveillance, the depth and breadth of the coronavirus hosts will likely continue to grow.

## PART II

### Coronaviral Pathogenicity and Host Immune Response

#### 1. Coronavirus Pathogenicity

Coronaviruses first replicate in epithelial cells of either the respiratory or enteric tract although tissue tropism varies by coronavirus group and strain. Viral replication in the respiratory tract causes local respiratory symptoms while replication in epithelial cells of the enteric tract causes diarrhea that can be severe and, in younger animals, life threatening. Animal CoV diseases range from highly pathogenic and economically important like IBV, which causes a highly infectious upper respiratory disease and possibly kidney disease in chicken, to a mild diarrhea or asymptomatic disease caused by feline enteric coronavirus (FECV). The disease manifestations in infections caused by coronaviruses can be due to the host response (Bergmann et al., 2006; Perlman and Dandekar, 2005). Reverse genetic studies have shown the S glycoprotein to be a major determinant of both tissue tropism and pathogenicity (Casais et al., 2003; Casais et al., 2001; Hodgson et al., 2004; Navas et al., 2001; Navas and Weiss, 2003; Phillips et al., 1999; Sanchez et al., 1999). However, depending on the coronavirus type, S may not act alone in determining pathogenicity (Hodgson et al., 2004) and the pathogenesis of coronaviruses is as diverse as the animals they infect.

Mouse hepatitis virus (MHV), a betacoronavirus in the species *Murine coronavirus*, is a group of virus strains with various organ tropisms and pathogenic potential. The MHV strains can be enterotropic and restricted to the intestinal tract or polytropic and affect a variety of organs such as the liver, the lung, and the central nervous system. Two MHV strains, JHM and A59, have received much attention as they are used in rodent animal models for human diseases

such as hepatitis, encephalitis, multiple sclerosis, and SARS. Both of these strains cause acute encephalitis which may result in demyelination (Cheever et al., 1949; Lavi et al., 1984). Studies have shown that the JHM strain, and the less virulent A59 strain, infects several types of cells including neurons, astrocytes, microglia, and oligodendrocytes. The attenuated strains of JHM appear to only infect glial cells rather than neurons (Marten et al., 2001). In order for demyelination to occur, a threshold level of viral replication must be met during the acute infection for the virus to spread to the spinal cord (Marten et al., 2000). JHM also induces a demyelinating disease in rats (Sorensen and Dales, 1985; Wege et al., 1998). Viruses with different pathogenic potentials; some primarily encephalic and some primarily demyelinating have been recovered from both the brain and the spinal cord of rats (Morris et al., 1989). Although A59 is less neurovirulent than JHM, it does induce mild to moderate encephalitis as well as moderate to severe hepatitis; albeit in a dose-dependent manner (Lavi et al., 1984; Leparc-Goffart et al., 1998).

The avian coronavirus IBV is predominantly a respiratory disease of chickens although it can be found replicating at multiple epithelial surfaces throughout the infected bird such as the gonads, the alimentary tract, the reproductive tract, and the kidney (Cavanagh, 2005; Cavanagh, 2008). Infection is initiated through the respiratory tract (Cavanagh, 2008). Clinical signs of an IBV infection include, coughing, watery eyes, gasping, tracheal rales, and nasal discharge (Cavanagh, 2008). Often these signs are accompanied by depression, poor feed conversion, and reduced weight gain. Most IBV strains cause disease in young chickens, less than 2 weeks of age, with a decrease in disease severity as the bird ages. Although morbidity is almost 100%, the mortality is usually low in uncomplicated infections. Increased mortality is directly associated with the prevalence of secondary infections such as mycoplasma, bacteria, or another virus.

Infections of the oviduct can cause both a drop in egg production and misshapen eggs. Chickens infected with a nephropathogenic strain of IBV not only show respiratory signs but also wet dropping, increased water intake, and increased mortality due to kidney damage (Cavanagh, 2008; Raj and Jones, 1997). Tissue damage begins to occur 48 hour after infection with IBV. The ciliated epithelium in the nose and trachea are lost resulting in edematous tissue. This is accompanied by hyperplasia and infiltrations of heterophils and lymphocytes (Nakamura et al., 1991; Raj and Jones, 1997). Infiltration of heterophils is also observed with nephropathogenic strains. This is accompanied by granular degeneration, vacuolation, and desquamation of the tubular epithelium of the kidneys (Cavanagh, 2008). Heterophil infiltration is also seen in the oviducts of infected mature hens. The oviduct may also show dilation of the tubular glands, edema, and loss of cilia (Cavanagh, 2008).

Transmissible gastroenteritis virus (TGEV) of swine is the causative agent of severe enteric disease of neonates (Saif, 2006). After infection, TGEV quickly infects almost all of the villous enterocytes of the small intestines. The only portion of the small intestine that is excluded is the proximal duodenum (Saif, 2006). The affected enterocytes lose functionality and lead to villous atrophy which results in the typical clinical signs of disease; such as, transient vomiting, yellow, watery malabsorptive diarrhea, and possibly fever (Saif, 2006). Although this virus can affect swine of any age, the disease is most pronounced and more deadly in neonates.

The majority of coronavirus infections cause a localized infection. However, feline infectious peritonitis virus (FIPV) is an exception. Although the infection begins in the epithelium of the oropharynx and small intestine, FIPV is able to cross the mucosal barrier via a cell-associated viremia in macrophages and cause disseminated disease (Pedersen et al., 2008;

Rottier et al., 2005; Stoddart et al., 1988). These infections are characterized by inflammatory exudates in the body cavity or disseminated pyogranulomas. The pathogenesis of FIPV appears to be enhanced by serum antibodies which generate immune complexes that are deposited in affected organs (Olsen et al., 1992).

## **2. Host immune response to coronavirus infection**

### **a. Innate immunity**

The innate immune system is important in limiting coronavirus infections. The efficacy of the innate immune response determines the extent of viral replication and thus the viral load associated with the infection. Interferon production is central to this response. Although coronaviruses contain a single stranded RNA (ssRNA) genome which can be recognized by host cell toll-like receptor (TLR) 7, double stranded RNA (dsRNA) present as secondary structures can be recognized by melanoma differentiation-associated gene 5 (MDA5) and/or TLR3. The binding of either the ssRNA or the dsRNA initiates a signaling cascade within the cell that ends with the nuclear translocation of nuclear factor  $\kappa$ B (NF- $\kappa$ B) and the transcription and expression of both interferon (IFN)  $-\alpha$  and IFN- $\beta$ . Once produced, IFN- $\alpha/\beta$  transmit signals to the cell interior via a specific receptor complex to induce an antiviral state in adjacent target cells or to induce apoptosis in the infected cell (Taniguchi and Takaoka, 2002). The antiviral state inhibits the replication of the virus by synthesis of enzymes that interfere with both cellular and viral processes. Of these enzymes, the most thoroughly studied are the dsRNA dependent protein kinase R (PKR) and the 2'-5' oligoadenylate synthetases. Both of these enzymes are inactive until exposed to virus thus ensuring the uninfected cells remain unaffected.

The PKR protein is comprised of two domains; an N-terminal domain that binds to the dsRNA and a C-terminal domain that is the catalytic center with conserved motifs for activity (Clemens and Elia, 1997). The PKR protein plays a role in mediating cell-regulatory activities by phosphorylating the  $\alpha$  subunit of the eukaryotic initiation factor 2 (eIF2) thus preventing the cellular recycling of initiation factors (Clemens and Elia, 1997; Meurs et al., 1990). The PKR protein mediates signal transduction in response to dsRNA by enhancing the activation of transcription factors such as signal transducers and activators of transcription 1 (STAT1) (Ramana et al., 2000). In order to aid in viral clearance, PKR is also involved in initiating apoptosis (Lee et al., 1997). The 2'-5' oligoadenylate synthetases are a group of enzymes that initiate a coordinated cascade of reactions that result in the inhibition of protein synthesis by activation RNase L which cleaves ssRNA including mRNA (Goodbourn et al., 2000).

Apoptosis, aka “programmed cell death”, is a process by which a cell can die and be phagocytized without inducing an inflammatory response. It is triggered in response to viral infection to limit viral spread and promote clearance of infected cells without inflammation. Coronaviruses induce apoptosis in a variety of ways that are strain specific and they utilize a different pathway depending on the cell type. Although, many coronaviruses induce the caspase-dependent apoptosis *in vivo* (Belyavsky et al., 1998; Collins, 2002; Eleouet et al., 2000), apoptosis seems to be cell-line dependent which suggests specific host cell interactions are necessary for coronaviruses to induce apoptosis.

## **b. Acquired Immunity**

### Cellular Immunity

Innate immune mediators, such as macrophages, dendritic cells, IFN $\alpha/\beta$ , and chemokines, initiate the adaptive immune response. The CD4<sup>+</sup> and CD8<sup>+</sup> T-cells precede the induction of humoral responses to viral infection. The CD4<sup>+</sup> T-cells provide a source of IFN $\gamma$  which increases the activation and expression of major histocompatibility complex (MHC) molecules on infected target cells, macrophages, and dendritic cells. These cells then function as antigen presenting cells further amplifying the cellular immune response. Tissue accumulations of activated CD4<sup>+</sup> T-cells have been described for both IBV and MHV (Bergmann et al., 2006; Kotani et al., 2000). Anti-viral T-cells are crucial for virus clearance and the prevention of recrudescence. Transfer of antiviral memory CD8<sup>+</sup> T-cells was able to control an MHV infection in an immunodeficient host suggesting a role for these cells as primary effectors of viral clearance (Bergmann et al., 2003; Bergmann et al., 2004).

#### Humoral and local immunity

Humoral immunity is mediated by secreted antibodies produced by B cells which block virus binding and entry into host cells and CD8<sup>+</sup> cytotoxic T lymphocytes (CTLs) which eliminate the infection by killing infected cells. Coronaviruses initiate the humoral immune response by activation of B cells in the mucosal lymph nodes of the respiratory and/or intestinal tract. Activation of the adaptive immune response is triggered by processes of antigen presentation and T-cell activation. Using IBV as a model virus, *Guo, et al*, characterized the molecular events associated with mucosal immunity (Guo et al., 2008). They found an increase in the expression of T-cell receptors (TCRs) and MHC class I molecules (Guo et al., 2008). This suggests that IBV infected cells are recognized by CD8<sup>+</sup> T-cells by engaging the TCR with peptides presented on MHC class I molecules on macrophages or dendritic cells. A Th1 adaptive

local immune response was induced by IBV infection and was mainly responsible for the viral clearance from the site of infection while the humoral response, primarily mucosal immunoglobulin G (IgG), provided protection against viral entry by neutralizing virus particles (Guo et al., 2008). Anti-IBV IgG can be detected as early as 4 days post-infection (pi), reaching a peak after 21 days but remaining at a high serum titre for several weeks (Mockett and Darbyshire, 1981). Circulating neutralizing antibodies are detectable in the serum during both acute and persistent coronavirus infections (Kiss et al., 2004). In humans, the presence of high levels of neutralizing antibodies is correlated with recovery from SARS-CoV infection (Ho et al., 2005). Immunoglobulin A (IgA) is the antibody predominantly found to be associated with local immunity against a coronavirus infection. Both IgA and IgG have been found in tracheal washes of IBV infected chickens as well as oviduct washes of infected hens.

### **3. Viral escape from host immunity**

Like many viruses, coronaviruses have developed strategies to counter the host innate immune response. Interferon expression is essential to the innate immune response and coronaviruses have developed mechanisms to antagonize IFN induction and signaling. The proteins expressed by the replicase gene (Armesto et al., 2009) and some group-specific proteins are involved in viral pathogenesis by antagonizing the host cell type 1 interferon response. Active inhibitors of IFN expression are nsp1, nsp3, N, and SARS-CoV accessory proteins, ORF6 and ORF3b (He et al., 2003; Kopecky-Bromberg et al., 2007; Narayanan et al., 2008b; Wathelet et al., 2007; Ye et al., 2007a; Zust et al., 2007). The nsp1 inhibits IFN production by degrading host mRNA and inhibiting host protein translation (Kamitani et al., 2006; Narayanan et al., 2008b) and by inhibiting STAT1 phosphorylation (Wathelet et al., 2007; Zust et al., 2007). Two

domains in nsp3 have been found to be involved with viral IFN antagonism. The papain-like protease (PLpro) domain of SARS-CoV was found to be a potent IFN antagonist (Devaraj et al., 2007). It functions by the downregulation of extracellular receptor-activated kinase (ERK) 1 by ubiquitin proteasomes and suppression of the interactions between ERK1 and STAT1 (Li et al., 2011). The ADRP domain of nsp3 also serves as an IFN antagonist. Kuri, et al, found that HuCoV-229E and SARS-CoV ADRP mutants had increased sensitivity to the antiviral effect of IFN $\alpha$  (Kuri et al., 2011). The N protein functions to inhibit IFN production via two reported mechanisms. In MHV, the N protein inhibits both activator protein 1 (AP1) signaling (He et al., 2003) and the function of PKR (Ye et al., 2007b). In SARS-CoV the N protein inhibits (NF $\kappa$ B) activation thus inhibiting interferon production (Kopecky-Bromberg et al., 2007). The SARS-CoV ORF3b and ORF6 have been found to not only inhibit IFN production but also to disrupt IFN signaling. The SARS-CoV ORF6 inhibits IFN signaling by binding to karyopherin- $\alpha$ 2. This binding results in tethering karyopherin- $\beta$  to the cytoplasmic membranes and inhibits nuclear translocation of IFN stimulating proteins such as STAT1 (Frieman et al., 2007; Hussain et al., 2008; Kopecky-Bromberg et al., 2007). Interestingly, when ORF6 was deleted from SARS-CoV the viral sensitivity to IFN was not increased (Zhao et al., 2009). This implicates redundant mechanisms for coronavirus IFN antagonism in SARS-CoV. There is evidence that viral dsRNA in DMVs is secluded from intracellular viral sensors (RIG-I, MDA5 and TLR3) this being a general mechanism for coronaviruses to prevent IFN induction (Versteeg et al., 2007; Zhou and Perlman, 2007).

## **PART III**

### **Coronavirus Diagnosis and Prevention**

#### **1. Diagnosis**

Laboratory diagnosis of a coronaviral infection usually involves one or more of the following: direct detection of virus, its genome, its antigen components, or host antibody response. Typically, a differential diagnosis is only accomplished through a combination of several diagnostic tests and a complete history as many of the clinical signs of exhibited by a coronavirus infection can be indicative of other viral or bacterial infections because the disease can be mild and self-limiting.

##### **a. Virus detection and isolation**

Viral particles can be observed in tissue, fluid, or fecal samples by electron microscopy. The typical pleomorphic, enveloped, spiked, virus particle constitute a positive identification of coronavirus. This technique has been used to determine infection of a coronavirus in humans (SARS), beluga whale (BWCoV/SW1) (Mihindukulasuriya et al., 2008), and SARS-related viruses from animals in southern China (Guan et al., 2003).

Samples for virus isolation must be collected and stored in buffer containing antibiotics. This allows for the suppression of bacterial and fungal growth. The sampled tissue and sampling technique for optimal virus isolation vary between coronaviruses but in the case of IBV samples for virus isolation can be trachea, cecal tonsil, lung, kidney, and oviduct (Gelb, 2008). Most coronaviruses can be isolated by the inoculation of either the allantoic cavity of 9-11 day old embryonating eggs or tissue culture to increase the viral load. In the case of IBV inoculated into embryonating eggs, embryo death may occur as well as characteristic lesions such as ruffled

feathers, dwarfing and /or curling of the embryo, and possibly kidney urates (Cavanagh, 2008; Gelb, 2008). The collected allantoic fluid or tissue culture media can then be used for a variety of assays.

#### **b. Detection of the viral genome**

Detection of viral genome has been accomplished in several ways including dot-blot hybridization, reverse-transcription (RT) polymerase chain reaction, or nested RT-PCR (PCR) (Jackwood et al., 1992; Kim et al., 2000; Kwon et al., 1993; Kwon et al., 1998). This test can be performed on samples collected directly from the infected host or from fluid obtained after passage of the sample in either eggs or tissue culture. Viral RNA is extracted from the sample and is subjected to RT-PCR amplification with oligonucleotide primer pairs that are specific for certain portions of the genome. To detect SARS, a 182 bp region of the RdRp encoding sequence was targeted (Poon et al., 2003) while for FCoV the primers have been designed to detect the Pol (Escutenaire et al., 2007; Herrewegh et al., 1998), the M (Simons et al., 2005), a small region spanning the M and N genes (Dye et al., 2008) the 7b gene (Kennedy et al., 2008; Kiss et al., 2000), and the 3' UTR (Duarte et al., 2009; Herrewegh et al., 1995). Each of these regions are highly conserved and are able to detect most if not all FCoV strains and are a valuable tool for screening for the virus in cat populations. However, only the detection of mRNA outside of the intestinal tract is indicative of an FIP infection as active replication of virus in circulating monocytes is typical. The test developed by Simmons, et al. (Simons et al., 2005) addressed this issue and is currently used in some diagnostic laboratories. For IBV, either primer pairs that are designed from hypervariable regions in the S1 gene that are specific for the different IBV strains or primers that amplify conserved regions of the genome can be used; however the most

common region chosen is the S1 portion of the spike gene (Jackwood et al., 1997). A SYBR Green RT-PCR technique developed by Escutenaire, *et al*, is designed to detect all coronaviruses. This assay amplified a 179 bp region of the 1b ORF. They were able to detect 32 animal coronaviruses including strains of CCoV, FCoV, TGEV, BCoV, MHV, and IBV (Escutenaire et al., 2007). This assay may prove to be a valuable tool for not only diagnostic purposes but also the detection of novel coronaviruses.

### **c. Antigen detection**

Antigen components can be detected in tissue using immunohistochemistry or fluid by enzyme linked immunosorbent assay (ELISA). Coronavirus antigens can also be detected by immunofluorescence or immunoperoxidase techniques within virus-infected cells.

### **d. Host antibody response**

The host antibody response is most easily detected in serum by use of an enzyme-linked immunosorbent assay (ELISA) (Carman et al., 2002; Crouch et al., 1984; Jackwood et al., 1997; Kraaijeveld et al., 1980; Shi et al., 2003). Briefly, antigen samples are used to coat the surface of an ELISA plate followed by sera obtained from the host containing antibodies directed at the antigen. If antibodies to the virus are present, they will bind to the antigen. Secondary antibodies are then added that are conjugated to an enzyme. A substrate that will react with the enzyme is then added which when hydrolyzed will fluoresce to allow antibody detection

## **2. Prevention of Coronavirus Infection**

As no antivirals exist to combat a coronavirus infection in either animals or humans, vaccination is an important strategy for disease prevention. Although there are no vaccines currently available to prevent human coronavirus infections, vaccination against animal coronavirus infections are routine and the efforts to improve these vaccines are ongoing (Saif, 2004). Immunization of cats against FIP can sometimes lead to a more serious disease (Perlman and Dandekar, 2005). This must be considered in the development of new coronavirus vaccines. The vaccines must protect from disease – not enhance the disease.

### **a. Enteric coronavirus vaccines**

Vaccines that target enteric coronavirus infection, specifically in the case of TGEV and BCoV are typically targeted toward protection of the suckling animal as the disease is most severe for this age group. These vaccines immunize the dams which confer passive transfer of IgG antibodies via the milk to the suckling neonate. In pigs, a PRCV infection acts as a naturally attenuated vaccine against TGEV. This protection has been shown in the litters from female pigs that were primed as neonates with PRCV (Wesley and Lager, 2003).

### **b. Respiratory coronavirus vaccines**

The only respiratory coronavirus vaccines available are for IBV. Both live and inactivated vaccines are used to protect against IBV. The live attenuated vaccines are produced by serial passage in embryonating chicken eggs. While the virus remains immunogenic, it is no longer highly pathogenic (Cavanagh, 2008). These vaccines are typically used in broiler-type chickens as only short-term (6-7 weeks) of protection is needed. Live vaccines are used in birds

where protection is needed for a longer period of time; specifically breeders and layers. These vaccines are produced from live virus that is inactivated with either formalin or beta propiolactone combined with an adjuvant such as mineral oil (Jansen et al., 2006). Live vaccines are more efficient at preventing infection of the respiratory tract; thus lowering the amount of virus found and limiting transmission (Ladman et al., 2002).

## **PART IV**

### **Coronavirus Evolution and Emergence**

#### **1. Evolution**

The genetic diversity in coronaviruses can be attributed to two process; both under the control of the RdRp. First, the RdRp does not have proofreading capabilities, so it cannot fix mistakes made while copying the viral genome: mutation. Second, genetic diversity can be produced by a template switching mechanism: recombination (Brooks et al., 2004).

##### **a. Mutation**

The accumulation of mutations in the coronavirus genome is a major force in generating diversity in the viral quasispecies, and the evolution of the virus (Domingo, 1997). The viral quasispecies is a mixed population of viruses that vary slightly in mutations that each has accumulated. This accumulation of mutations in the quasispecies gives the virus a selective advantage to adapt to different host tissues, evade the host immune system, and become transmissible to a new host. The ability to of a mixed viral population to evade the host immune system has been shown for a picornavirus, specifically Foot and Mouth Disease (Borrego et al.,

1993). Viruses that are able to evade the immune system are able to replicate, increasing the population size, and ensuring the survival of the virus.

Coronavirus genomes encode four structural proteins and a large overlapping reading frame that encodes for non-structural proteins; including the RNA-dependent RNA-polymerase. This polymerase, unlike DNA polymerases, does not have 3' to 5' proofreading capabilities so that any errors that are generated during nascent RNA synthesis are incorporated into the genome (Steinhauer et al., 1992). Mutations accumulate with every round of viral replication; thus the virus evolves. As coronaviruses are an RNA virus, their mutation rate may be close to one mutation per genome per replication (Drake, 1993). The synonymous mutation rate for bovine coronavirus has been calculated to be  $10^{-3}$ /synonymous site/year (Hanada et al., 2004) which is likely indicative of the mutation rate for all coronaviruses. Not all mutations introduced by the error-prone polymerase are represented in the genomes of the progeny virus. The virus must be able to infect a host cell replicate efficiently. This provides selective pressure as only viable viruses in the quasispecies can be utilized (Domingo, 1997; Holmes, 2009). However, coronaviruses possess an ExoN domain which encodes a 3' to 5' exoribonuclease which exhibits, similarities to host cellular proteins of the DEDD superfamily of exonucleases (Minskaia et al., 2006; Snijder et al., 2003b); pointing to the ability of coronaviruses to minimize the error rate caused by the RdRp.

## **b. Recombination in Coronaviruses**

Recombination accounts for a considerable amount of genetic diversity among coronaviruses. The high frequency of homologous RNA recombination is a well documented occurrence for coronavirus replication and evolution. This occurs when two or more different

strains of coronavirus enter the same cell. Recombination occurs when the RdRp switches templates from one viral genome to another. Zhang et al. recently described the requirements needed for recombination to occur: (1) Two coronaviruses must infect a host cell simultaneously and continue to replicate without interference from each other; (2) there must be sufficient nucleotide identity between the two viral genomes in order for template switching to occur during RNA replication; (3) the proteins formed from recombination must be functional; (4) the recombinant virus must be either evolutionarily neutral or have some selective advantage in order to survive (Zhang et al., 2005).

## **2. Emerging Coronaviruses**

In 2002, a novel coronavirus was associated with the outbreak of severe acute respiratory syndrome (SARS) in Guangdong Province, China (Ksiazek et al., 2003b). Several studies have linked this virus to an animal reservoir (Guan et al., 2003; Lau et al., 2005) but the definitive reservoir has yet to be elucidated. This has led to the discovery of an astonishing variety of coronaviruses especially in bats (Dominguez, 2007; Poon et al., 2005; Woo et al., 2006). In a study by Vijaykrishna et al, it was proposed that bats are the likely the natural hosts for all presently known coronavirus lineages and that all coronaviruses recognized in other species were derived from viruses residing in bats (Vijaykrishna et al., 2007). The SARS-CoV outbreak was due to the ability of the virus to cross species and infect humans (Guan et al., 2003). There are other examples of coronaviruses cross-species transmission. Bovine coronavirus (BCoV) and a human coronavirus (HCoV-OC43) are greater than 95% similar throughout their genome and the virus may have crossed from a bovine to a human host approximately 100 years ago (Vijgen et al., 2005). Both BCoV and HCoV-OC43 use a type of sialic acid as a host cell receptor and both

infect a large number of species *in vitro* and readily cross species to infect the mouse *in vivo* (Butler et al., 2006). BCoV may have crossed into other species as similar coronaviruses have been found in alpacas (*Vicugna pacos*)(Jin et al., 2007), giraffes (*Giraffa camelopardalis*) (Hasoksuz et al., 2007) and other ruminants(Alekseev et al., 2008). A canine coronavirus (CCoV) was isolated from a the liver and spleen of giant pandas (*Ailuropoda melanoleuca*) (Gao et al., 2009). A coronavirus isolated from ferrets (*Mustelo putorius furo*) has sequence similarity with canine, feline, and porcine coronaviruses suggesting that it is the result of recombination among those coronaviruses (Wise et al., 2006). Studies have identified several novel avian-like coronaviruses that have crossed species. These viruses have been found in wild Asian leopard cats (*Prionailuris bengalensis*), Chinese ferret badgers (*Melogale moschata*) (Dong et al., 2007) and a beluga whale (*Delphinapterus leucas*).

Coronaviruses can be found worldwide. These viruses have been isolated from an extraordinary variety of animals providing evidence of the zoonotic potential of coronaviruses. Regions of the genome have been extensively sequenced and analyzed; however few full genomic sequences are available to the scientific community. In order to predict and possibly prevent another coronavirus outbreak caused by an animal to human zoonotic shift, such as SARS-CoV, it is necessary to identify the genes and genomes of coronaviruses of animals closely associated with people.

## References

1. Kalb, C., *Tracking SARS*. Newsweek, 2003. **141**(17): p. 36-7.
2. Ksiazek, T., et al., *A novel coronavirus associated with severe acute respiratory syndrome*. N Engl J Med, 2003. **348**: p. 1953 - 1966.
3. WHO. *WHO Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 Junly 2003*. 2003; Available from: [http://www.who.int/csr/sars/country/table2004\\_04\\_21/en/index.html#](http://www.who.int/csr/sars/country/table2004_04_21/en/index.html#).
4. Peiris, J., Y. Guan, and K. Yuen, *Severe acute respiratory syndrome*. Nat Med, 2004. **10**(12 Suppl): p. S88 - 97.
5. Marra, M.A., S.J Jones, C.R. Astell, et. al, *The genome sequence of the SARS- associated coronavirus*. Science, 2003(300): p. 1300-1404.
6. Cowley, G., *Solving the mystery of SARS*. Newsweek, 2003. **141**(15): p. 50-1.
7. Bizri, A.R. and U.M. Musharrafieh, *SARS: the flying death*. J Med Liban, 2003. **51**(3): p. 148-54.
8. Lai, M.M.C., K.V. Holmes, *Coronaviridae: the viruses and their replication*. 4th ed. Fields Virology, ed. D.M. Knipe, P.M. Howley, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman, et.al. 2001, Philadelphia, PA: Lippincott, Williams, and Wilkins.
9. Carstens, E., *Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses* Archives of Virology, 2009. **155**(1): p. 133-146.
10. Mihindukulasuriya, K.A., et al., *Identification of a novel coronavirus from a beluga whale by using a panviral microarray*. J Virol, 2008. **82**(10): p. 5084-8.
11. Dong, B.Q., et al., *Detection of a novel and highly divergent coronavirus from asian leopard cats and Chinese ferret badgers in Southern China*. J Virol, 2007. **81**(13): p. 6920-6.
12. Erles, K., K.B. Shiu, and J. Brownlie, *Isolation and sequence analysis of canine respiratory coronavirus*. Virus Res, 2007. **124**(1-2): p. 78-87.
13. Lau, S.K., et al., *Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats*. Proc Natl Acad Sci U S A, 2005. **102**(39): p. 14040-5.

14. Tang, X.C., et al., *Prevalence and genetic diversity of coronaviruses in bats from China*. J Virol, 2006. **80**(15): p. 7481-90.
15. Woo, P.C., et al., *Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus*. J Virol, 2009. **83**(2): p. 908-17.
16. de Groot, R.J., Baker, S.C., Baric, R., Enjuanes, L., Gorbalenya, A.E., Holmes, K.V., Perlman, S., Poon, L., Rottier, P.J.M., Talbot, P.J., Woo, P.C., Ziebuhr, J., In A.M.Q. King (ed), *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. 2011 (in press): Elsevier.
17. Holmes, K.V., *Coronaviruses*. Fields Virology. 2001, Philadelphia, PA: Lippencott.
18. Bartlam, M., Y. Xu, and Z. Rao, *Structural proteomics of the SARS coronavirus: a model response to emerging infectious diseases*. J Struct Funct Genomics, 2007. **8**(2-3): p. 85-97.
19. Prentice, E., et al., *Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins*. J Virol, 2004. **78**(18): p. 9977-86.
20. Bhardwaj, K., et al., *RNA recognition and cleavage by the SARS coronavirus endoribonuclease*. J Mol Biol, 2006. **361**(2): p. 243-56.
21. Bhardwaj, K., et al., *Structural and functional analyses of the severe acute respiratory syndrome coronavirus endoribonuclease Nsp15*. J Biol Chem, 2008. **283**(6): p. 3655-64.
22. Eckerle, L.D., et al., *Infidelity of SARS-CoV Nsp14-Exonuclease Mutant Virus Replication Is Revealed by Complete Genome Sequencing*. PLoS Pathog, 2010. **6**(5): p. e1000896.
23. Snijder, E.J., et al., *Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage*. J Mol Biol, 2003. **331**(5): p. 991-1004.
24. Tohya, Y., et al., *Suppression of host gene expression by nsp1 proteins of group 2 bat coronaviruses*. J Virol, 2009. **83**(10): p. 5282-8.
25. Züst, R., et al., *Coronavirus Non-Structural Protein 1 Is a Major Pathogenicity Factor: Implications for the Rational Design of Coronavirus Vaccines*. PLoS Pathog, 2007. **3**(8): p. e109.
26. Narayanan, K., et al., *Severe Acute Respiratory Syndrome Coronavirus nsp1 Suppresses Host Gene Expression, Including That of Type I Interferon, in Infected Cells*. J. Virol., 2008. **82**(9): p. 4471-4479.

27. Kamitani, W., et al., *Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation*. Proc Natl Acad Sci U S A, 2006. **103**(34): p. 12885-90.
28. Law, A.H.Y., et al., *Role for Nonstructural Protein 1 of Severe Acute Respiratory Syndrome Coronavirus in Chemokine Dysregulation*. J. Virol., 2007. **81**(5): p. 2537-.
29. Yang, A., et al., *Expression, crystallization and preliminary X-ray diffraction analysis of the N-terminal domain of nsp2 from avian infectious bronchitis virus*. Acta Crystallographica Section F, 2009. **65**(8): p. 788-790.
30. Denison, M.R., S.A. Hughes, and S.R. Weiss, *Identification and characterization of a 65-kDa protein processed from the gene 1 polyprotein of the murine coronavirus MHV-A59*. Virology, 1995. **207**(1): p. 316-20.
31. Graham, R.L. and M.R. Denison, *Replication of murine hepatitis virus is regulated by papain-like proteinase 1 processing of nonstructural proteins 1, 2, and 3*. J Virol, 2006. **80**(23): p. 11610-20.
32. Graham, R.L., et al., *The nsp2 replicase proteins of murine hepatitis virus and severe acute respiratory syndrome coronavirus are dispensable for viral replication*. J Virol, 2005. **79**(21): p. 13399-411.
33. Gadlage, M.J., R.L. Graham, and M.R. Denison, *Murine coronaviruses encoding nsp2 at different genomic loci have altered replication, protein expression, and localization*. J Virol, 2008. **82**(23): p. 11964-9.
34. Graham, R.L., et al., *The nsp2 proteins of mouse hepatitis virus and SARS coronavirus are dispensable for viral replication*. Adv Exp Med Biol, 2006. **581**: p. 67-72.
35. Hagemeyer, M.C., et al., *Dynamics of Coronavirus Replication-Transcription Complexes*. J. Virol., 2010. **84**(4): p. 2134-2149.
36. Cornillez-Ty, C.T., et al., *Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling*. J Virol, 2009. **83**(19): p. 10314-8.
37. Wang, S., N. Nath, M. Adlam, and S. Chellappan, *Prohibitin, a potential tumor suppressor, interacts with RB and regulates E2F function*. Oncogene, 1999. **18**: p. 3501-3510.
38. Rajalingam, K., et al., *Prohibitin is required for Ras-induced Raf-MEK-ERK activation and epithelial cell migration*. Nat Cell Biol, 2005. **7**(8): p. 837-843.

39. Héron-Milhavet, L., et al., *Akt2 is implicated in skeletal muscle differentiation and specifically binds Prohibitin2/REA*. Journal of Cellular Physiology, 2008. **214**(1): p. 158-165.
40. Sun, L., et al., *Akt binds prohibitin 2 and relieves its repression of MyoD and muscle differentiation*. Journal of Cell Science, 2004. **117**(14): p. 3021-3029.
41. Fusaro, G., et al., *Prohibitin Induces the Transcriptional Activity of p53 and Is Exported from the Nucleus upon Apoptotic Signaling*. Journal of Biological Chemistry, 2003. **278**(48): p. 47853-47861.
42. Merkwirth, C. and T. Langer, *Prohibitin function within mitochondria: Essential roles for cell proliferation and cristae morphogenesis*. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2009. **1793**(1): p. 27-32.
43. Snijder, E., et al., *Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage*. J Mol Biol, 2003. **331**(5): p. 991 - 1004.
44. Ziebuhr, J., V. Thiel, and A.E. Gorbalenya, *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, 2001. **276**(35): p. 33220-32.
45. Hurst, K.R., et al., *An Interaction between the Nucleocapsid Protein and a Component of the Replicase-Transcriptase Complex Is Crucial for the Infectivity of Coronavirus Genomic RNA*. J. Virol., 2010. **84**(19): p. 10276-10288.
46. Bonilla, P.J., S.A. Hughes, and S.R. Weiss, *Characterization of a second cleavage site and demonstration of activity in trans by the papain-like proteinase of the murine coronavirus mouse hepatitis virus strain A59*. J Virol, 1997. **71**(2): p. 900-9.
47. Baker, S.C., et al., *Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus*. J. Virol., 1993. **67**(10): p. 6056-6063.
48. Thiel, V., et al., *Mechanisms and enzymes involved in SARS coronavirus genome expression*. J Gen Virol, 2003. **84**(Pt 9): p. 2305-15.
49. Anand, K., et al., *Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs*. Science, 2003. **300**(5626): p. 1763-7.
50. Devaraj, S.G., et al., *Regulation of IRF-3-dependent innate immunity by the papain-like protease domain of the severe acute respiratory syndrome coronavirus*. J Biol Chem, 2007. **282**(44): p. 32208-21.

51. Frieman, M., et al., *Severe acute respiratory syndrome coronavirus papain-like protease ubiquitin-like domain and catalytic domain regulate antagonism of IRF3 and NF-kappaB signaling*. J Virol, 2009. **83**(13): p. 6689-705.
52. Pehrson and V. Fried, *MacroH2A, a core histone containing a large nonhistone region*. Science, 1992. **257**(5075): p. 1398-1400.
53. Egloff, M.P., et al., *Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains*. J Virol, 2006. **80**(17): p. 8493-502.
54. Piotrowski, Y., et al., *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 Sci, 2009. **18**(1): p. 6-16.
55. Ratia, K., et al., *A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication*. Proc Natl Acad Sci U S A, 2008. **105**(42): p. 16119-24.
56. Zheng, D., et al., *PLP2, a potent deubiquitinase from murine hepatitis virus, strongly inhibits cellular type I interferon production*. Cell Res, 2008. **18**(11): p. 1105-13.
57. Wang, G., et al., *PLP2 of Mouse Hepatitis Virus A59 (MHV-A59) Targets TBK1 to Negatively Regulate Cellular Type I Interferon Signaling Pathway*. PLoS ONE, 2011. **6**(2): p. e17192.
58. Clementz, M.A., et al., *Deubiquitinating and Interferon Antagonism Activities of Coronavirus Papain-Like Proteases*. J. Virol., 2010. **84**(9): p. 4619-4629.
59. Oostra, M., et al., *Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning*. J Virol, 2008. **82**(24): p. 12392-405.
60. Kanjanahaluethai, A., et al., *Membrane topology of murine coronavirus replicase nonstructural protein 3*. Virology, 2007. **361**(2): p. 391-401.
61. Álvarez, E., et al., *The envelope protein of severe acute respiratory syndrome coronavirus interacts with the non-structural protein 3 and is ubiquitinated*. Virology, 2010. **402**(2): p. 281-291.
62. Gosert, R., et al., *RNA replication of mouse hepatitis virus takes place at double-membrane vesicles*. J Virol, 2002. **76**(8): p. 3697-708.
63. Hagemeyer, M.C., et al., *Mobility and Interactions of Coronavirus Nonstructural Protein 4*. J. Virol., 2011. **85**(9): p. 4572-4577.

64. Anand, K., et al., *Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain*. EMBO J, 2002. **21**(13): p. 3213-24.
65. Lu, Y., X. Lu, and M.R. Denison, *Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59*. J Virol, 1995. **69**(6): p. 3554-9.
66. Long, A.C., et al., *A consensus sequence for substrate hydrolysis by rhino virus 3C proteinase*. FEBS Letters, 1989. **258**(1): p. 75-78.
67. Ziebuhr, J., E.J. Snijder, and A.E. Gorbalenya, *Virus-encoded proteinases and proteolytic processing in the Nidovirales*. Journal of General Virology, 2000. **81**(4): p. 853-879.
68. Ziebuhr, J., *The coronavirus replicase*. Curr Top Microbiol Immunol, 2005. **287**: p. 57-94.
69. Masters, P.S., *The molecular biology of coronaviruses*. Adv Virus Res, 2006. **66**: p. 193-292.
70. Kuo, C.J., et al., *Individual and common inhibitors of coronavirus and picornavirus main proteases*. FEBS Lett, 2009. **583**(3): p. 549-55.
71. Baliji, S., et al., *Detection of nonstructural protein 6 in murine coronavirus-infected cells and analysis of the transmembrane topology by using bioinformatics and molecular approaches*. J Virol, 2009. **83**(13): p. 6957-62.
72. Bost, A.G., et al., *Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly*. J Virol, 2000. **74**(7): p. 3379-87.
73. Bost, A.G., E. Prentice, and M.R. Denison, *Mouse hepatitis virus replicase protein complexes are translocated to sites of M protein accumulation in the ERGIC at late times of infection*. Virology, 2001. **285**(1): p. 21-9.
74. Campanacci, V., et al., *Structural genomics of the SARS coronavirus: cloning, expression, crystallization and preliminary crystallographic study of the Nsp9 protein*. Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 9): p. 1628-31.
75. van der Meer, Y., et al., *Localization of mouse hepatitis virus nonstructural proteins and RNA synthesis indicates a role for late endosomes in viral replication*. J Virol, 1999. **73**(9): p. 7641-57.
76. Deming, D.J., et al., *Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication*. J Virol, 2007. **81**(19): p. 10280-91.

77. Fang, S.G., et al., *Proteolytic processing of polyproteins 1a and 1ab between non-structural proteins 10 and 11/12 of Coronavirus infectious bronchitis virus is dispensable for viral replication in cultured cells.* Virology, 2008. **379**(2): p. 175-80.
78. Zhai, Y., et al., *Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer.* Nat Struct Mol Biol, 2005. **12**(11): p. 980-6.
79. Imbert, I., et al., *A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus.* EMBO J, 2006. **25**(20): p. 4933-42.
80. Gorbalenya, A.E., et al., *Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis.* Nucleic Acids Research, 1989. **17**(12): p. 4847-4861.
81. Egloff, M.P., et al., *The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world.* Proc Natl Acad Sci U S A, 2004. **101**(11): p. 3792-6.
82. Sutton, G., et al., *The nsp9 replicase protein of SARS-coronavirus, structure and functional insights.* Structure, 2004. **12**(2): p. 341-53.
83. Ponnusamy, R., et al., *Non structural proteins 8 and 9 of human coronavirus 229E.* Adv Exp Med Biol, 2006. **581**: p. 49-54.
84. Sawicki, S.G., et al., *Functional and Genetic Analysis of Coronavirus Replicase-Transcriptase Proteins.* PLoS Pathog, 2005. **1**(4): p. e39.
85. Joseph, J.S., et al., *Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs.* J Virol, 2006. **80**(16): p. 7894-901.
86. Matthes, N., et al., *The non-structural protein Nsp10 of mouse hepatitis virus binds zinc ions and nucleic acids.* FEBS Lett, 2006. **580**(17): p. 4143-9.
87. Su, D., et al., *Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10.* J Virol, 2006. **80**(16): p. 7902-8.
88. Koonin, E.V., *The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses.* Journal of General Virology, 1991. **72**(9): p. 2197-2206.
89. te Velhuis, A.J.W., et al., *The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent.* Nucleic Acids Research, 2010. **38**(1): p. 203-214.

90. Cheng, A., et al., *Expression, purification, and characterization of SARS coronavirus RNA polymerase*. Virology, 2005. **335**(2): p. 165-76.
91. Ivanov, K.A. and J. Ziebuhr, *Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities*. J Virol, 2004. **78**(14): p. 7833-8.
92. Seybert, A., et al., *The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5'-to-3' polarity*. RNA, 2000. **6**(7): p. 1056-68.
93. Ivanov, K.A., et al., *Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase*. J Virol, 2004. **78**(11): p. 5619-32.
94. Fang, S., et al., *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, 2007. **358**(1): p. 136-47.
95. Minskaia, E., et al., *Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis*. Proc Natl Acad Sci U S A, 2006. **103**(13): p. 5108-13.
96. Denison, M.R., et al., *Coronaviruses: An RNA proofreading machine regulates replication fidelity and diversity*. RNA Biology, 2011. **8**(2): p. 270-279.
97. Zuo, Y. and M.P. Deutscher, *Exoribonuclease superfamilies: structural analysis and phylogenetic distribution*. Nucleic Acids Research, 2001. **29**(5): p. 1017-1026.
98. Eckerle, L.D., et al., *High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants*. J Virol, 2007. **81**(22): p. 12135-44.
99. Chen, Y., et al., *Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase*. Proc Natl Acad Sci U S A, 2009. **106**(9): p. 3484-9.
100. Ivanov, K.A., et al., *Major genetic marker of nidoviruses encodes a replicative endoribonuclease*. Proc Natl Acad Sci U S A, 2004. **101**(34): p. 12694-9.
101. Bhardwaj, K., L. Guarino, and C.C. Kao, *The severe acute respiratory syndrome coronavirus Nsp15 protein is an endoribonuclease that prefers manganese as a cofactor*. J Virol, 2004. **78**(22): p. 12218-24.
102. Ricagno, S., et al., *Crystal structure and mechanistic determinants of SARS coronavirus nonstructural protein 15 define an endoribonuclease family*. Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11892-7.

103. Cao, J., C.C. Wu, and T.L. Lin, *Turkey coronavirus non-structure protein NSP15--an endoribonuclease*. Intervirology, 2008. **51**(5): p. 342-51.
104. Posthuma, C.C., et al., *Site-Directed Mutagenesis of the Nidovirus Replicative Endoribonuclease NendoU Exerts Pleiotropic Effects on the Arterivirus Life Cycle*. J. Virol., 2006. **80**(4): p. 1653-1661.
105. Decroly, E., et al., *Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity*. J Virol, 2008. **82**(16): p. 8071-84.
106. Bujnicki, J.M. and L. Rychlewski, *In silico identification, structure prediction and phylogenetic analysis of the 2'-O-ribose (cap 1) methyltransferase domain in the large structural protein of ssRNA negative-strand viruses*. Protein Engineering, 2002. **15**(2): p. 101-108.
107. Egloff, M.-P., et al., *An RNA cap (nucleoside-2[prime]-O)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization*. EMBO J, 2002. **21**(11): p. 2757-2768.
108. Ferron, F., et al., *Viral RNA-polymerases – a predicted 2'-O-ribose methyltransferase domain shared by all Mononegavirales*. Trends in Biochemical Sciences, 2002. **27**(5): p. 222-224.
109. Decroly, E., et al., *Crystal Structure and Functional Analysis of the SARS-Coronavirus RNA Cap 2'-O-Methyltransferase nsp10/nsp16 Complex*. PLoS Pathog, 2011. **7**(5): p. e1002059.
110. Debarnot, C., et al., *Crystallization and diffraction analysis of the SARS coronavirus nsp10-nsp16 complex*. Acta Crystallographica Section F, 2011. **67**(3): p. 404-408.
111. Luo, Z. and S.R. Weiss, *Roles in cell-to-cell fusion of two conserved hydrophobic regions in the murine coronavirus spike protein*. Virology, 1998. **244**(2): p. 483-94.
112. Bosch, B.J., et al., *Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides*. Proc Natl Acad Sci U S A, 2004. **101**(22): p. 8455-60.
113. Routledge, E., et al., *Analysis of murine coronavirus surface glycoprotein functions by using monoclonal antibodies*. Journal of Virology, 1991. **65**(1): p. 254-62.
114. Gombold, J.L., S.T. Hingley, and S.R. Weiss, *Fusion-defective mutants of mouse hepatitis virus A59 contain a mutation in the spike protein cleavage signal*. J. Virol., 1993. **67**(8): p. 4504-4512.

115. Gallagher, T.M., C. Escarmis, and M.J. Buchmeier, *Alteration of the pH dependence of coronavirus-induced cell fusion: effect of mutations in the spike glycoprotein*. J. Virol., 1991. **65**(4): p. 1916-1928.
116. Holmes, K., R. Welsh, and M. Haspel, *Natural cytotoxicity against mouse hepatitis virus-infected target cells. I. Correlation of cytotoxicity with virus binding to leukocytes*. The Journal of Immunology, 1986. **136**(4): p. 1446-1453.
117. Oleszak, E.L., et al., *Molecular mimicry between S peplomer proteins of coronaviruses (MHV, BCoV, TGEV and IBV) and Fc receptor*. Adv Exp Med Biol, 1993. **342**: p. 183-8.
118. Oleszak, E.L., et al., *Molecular mimicry between Fc receptor and S peplomer protein of mouse hepatitis virus, bovine corona virus, and transmissible gastroenteritis virus*. Hybridoma, 1995. **14**(1): p. 1-8.
119. Vennema, H., et al., *Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes*. EMBO J, 1996. **15**(8): p. 2020-8.
120. Maeda, J., A. Maeda, and S. Makino, *Release of coronavirus E protein in membrane vesicles from virus-infected cells and E protein-expressing cells*. Virology, 1999. **263**(2): p. 265-72.
121. Wilson, L., et al., *SARS coronavirus E protein forms cation-selective ion channels*. Virology, 2004. **330**(1): p. 322-31.
122. de Haan, C.A., et al., *The glycosylation status of the murine hepatitis coronavirus M protein affects the interferogenic capacity of the virus in vitro and its ability to replicate in the liver but not the brain*. Virology, 2003. **312**(2): p. 395-406.
123. Laude, H., et al., *Single amino acid changes in the viral glycoprotein M affect induction of alpha interferon by the coronavirus transmissible gastroenteritis virus*. J. Virol., 1992. **66**(2): p. 743-749.
124. Lai, M.M. and D. Cavanagh, *The molecular biology of coronaviruses*. Adv Virus Res, 1997. **48**: p. 1-100.
125. de Haan, C.A., et al., *Mapping of the coronavirus membrane protein domains involved in interaction with the spike protein*. J Virol, 1999. **73**(9): p. 7441-52.
126. Kuo, L. and P.S. Masters, *Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus*. J Virol, 2002. **76**(10): p. 4987-99.
127. Narayanan, K., et al., *Nucleocapsid-independent specific viral RNA packaging via viral envelope protein and viral RNA signal*. J Virol, 2003. **77**(5): p. 2922-7.

128. Almazan, F., C. Galan, and L. Enjuanes, *The nucleoprotein is required for efficient coronavirus genome replication*. J Virol, 2004. **78**(22): p. 12683-8.
129. Zúñiga, S., et al., *Coronavirus nucleocapsid protein is an RNA chaperone*. Virology, 2007. **357**(2): p. 215-227.
130. TOMPA, P. and P. CSERMELY, *The role of structural disorder in the function of RNA and protein chaperones*. FASEB J., 2004. **18**(11): p. 1169-1175.
131. Ning, Q., et al., *Induction of prothrombinase fgl2 by the nucleocapsid protein of virulent mouse hepatitis virus is dependent on host hepatic nuclear factor-4 alpha*. J Biol Chem, 2003. **278**(18): p. 15541-9.
132. Ning, Q., et al., *Ribavirin inhibits viral-induced macrophage production of TNF, IL-1, the procoagulant fgl2 prothrombinase and preserves Th1 cytokine production but inhibits Th2 cytokine response*. J Immunol, 1998. **160**(7): p. 3487-93.
133. Yount, B., et al., *Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59*. J Virol, 2002. **76**(21): p. 11065-78.
134. Yount, B., et al., *Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus*. Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12995-3000.
135. Schelle, B., et al., *Selective replication of coronavirus genomes that express nucleocapsid protein*. J Virol, 2005. **79**(11): p. 6620-30.
136. Ning, Q., et al., *The nucleocapsid protein of murine hepatitis virus type 3 induces transcription of the novel fgl2 prothrombinase gene*. J Biol Chem, 1999. **274**(15): p. 9930-6.
137. Ding, J.W., et al., *Fulminant hepatic failure in murine hepatitis virus strain 3 infection: tissue-specific expression of a novel fgl2 prothrombinase*. J Virol, 1997. **71**(12): p. 9223-30.
138. Wurm, T., et al., *Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division*. J Virol, 2001. **75**(19): p. 9345-56.
139. Hiscox, J.A., et al., *The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus*. J Virol, 2001. **75**(1): p. 506-12.
140. Ye, Y., et al., *Mouse hepatitis coronavirus A59 nucleocapsid protein is a type I interferon antagonist*. J Virol, 2007. **81**(6): p. 2554-63.

141. Yokomori, K., et al., *Neuropathogenicity of mouse hepatitis virus JHM isolates differing in hemagglutinin-esterase protein expression*. J Neurovirol, 1995. **1**(5-6): p. 330-9.
142. Kienzle, T.E., et al., *Structure and orientation of expressed bovine coronavirus hemagglutinin-esterase protein*. J. Virol., 1990. **64**(4): p. 1834-1838.
143. Luytjes, W., Bredenbeek, P.J., Noten, A.F.H., Horzinek, M.C., Spaan, W. J.M, *Sequence of mouse hepatitis virus A59 mRNA 2: indications for RNA recombination between coronavirus and influenza C virus*. Virology, 1988. **166**: p. 415-422.
144. Ontiveros, E., et al., *Inactivation of expression of gene 4 of mouse hepatitis virus strain JHM does not affect virulence in the murine CNS*. Virology, 2001. **289**(2): p. 230-8.
145. Yount, B., et al., *Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice*. J Virol, 2005. **79**(23): p. 14909-22.
146. de Haan, C.A., et al., *The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host*. Virology, 2002. **296**(1): p. 177-89.
147. Benbaccer, L., et al., *Interspecies aminopeptidase-N chimeras reveal species-specific receptor recognition by canine coronavirus, feline infectious peritonitis virus, and transmissible gastroenteritis virus*. J Virol, 1997. **71**(1): p. 734-7.
148. Shapiro, L.H., et al., *Separate promoters control transcription of the human aminopeptidase N gene in myeloid and intestinal epithelial cells*. Journal of Biological Chemistry, 1991. **266**(18): p. 11999-12007.
149. Li, W., et al., *Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus*. Nature, 2003. **426**: p. 450 - 454.
150. Hofmann, H., et al., *Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry*. Proc Natl Acad Sci U S A, 2005. **102**(22): p. 7988-93.
151. Turner, A.J., J.A. Hiscox, and N.M. Hooper, *ACE2: from vasopeptidase to SARS virus receptor*. Trends Pharmacol Sci, 2004. **25**(6): p. 291-4.
152. Hamming, I., et al., *Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis*. J Pathol, 2004. **203**(2): p. 631-7.

153. Kuba, K., et al., *A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury*. Nat Med, 2005. **11**(8): p. 875-9.
154. Imai, Y., et al., *Angiotensin-converting enzyme 2 protects from severe acute lung failure*. Nature, 2005. **436**(7047): p. 112-116.
155. Li, F., et al., *Structure of SARS coronavirus spike receptor-binding domain complexed with receptor*. Science, 2005. **309**(5742): p. 1864-8.
156. Yu, M., et al., *Identification of key amino acid residues required for horseshoe bat angiotensin-I converting enzyme 2 to function as a receptor for severe acute respiratory syndrome coronavirus*. J Gen Virol, 2010. **91**(7): p. 1708-1712.
157. Klein, A., et al., *9-O-acetylated sialic acids have widespread but selective expression: analysis using a chimeric dual-function probe derived from influenza C hemagglutinin-esterase*. Proceedings of the National Academy of Sciences, 1994. **91**(16): p. 7782-7786.
158. Tan, K., et al., *Crystal structure of murine sCEACAM1a[1,4]: a coronavirus receptor in the CEA family*. EMBO J, 2002. **21**(9): p. 2076-86.
159. Nakajima, A., et al., *Activation-Induced Expression of Carcinoembryonic Antigen-Cell Adhesion Molecule 1 Regulates Mouse T Lymphocyte Function*. The Journal of Immunology, 2002. **168**(3): p. 1028-1035.
160. Turner, B.C., et al., *Receptor-dependent coronavirus infection of dendritic cells*. J Virol, 2004. **78**(10): p. 5486-90.
161. Snijder, E.J., et al., *Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex*. J Virol, 2006. **80**(12): p. 5927-40.
162. Knoops, K., et al., *SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum*. PLoS Biol, 2008. **6**(9): p. e226.
163. Oostra, M., et al., *Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication*. J Virol, 2007. **81**(22): p. 12323-36.
164. Brockway, S.M., et al., *Characterization of the expression, intracellular localization, and replication complex association of the putative mouse hepatitis virus RNA-dependent RNA polymerase*. J Virol, 2003. **77**(19): p. 10515-27.
165. Zelus, B.D., et al., *Conformational changes in the spike glycoprotein of murine coronavirus are induced at 37 degrees C either by soluble murine CEACAM1 receptors or by pH 8*. J Virol, 2003. **77**(2): p. 830-40.

166. Sola, I., et al., *Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis*. J Virol, 2005. **79**(4): p. 2506-16.
167. Curtis, K.M., et al., *Reverse genetic analysis of the transcription regulatory sequence of the coronavirus transmissible gastroenteritis virus*. J Virol, 2004. **78**(11): p. 6061-6.
168. Zuniga, S., et al., *Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis*. J Virol, 2004. **78**(2): p. 980-94.
169. Narayanan, K. and S. Makino, *Characterization of nucleocapsid-M protein interaction in murine coronavirus*. Adv Exp Med Biol, 2001. **494**: p. 577-82.
170. Cavanagh, D., et al., *Coronaviruses from pheasants (*Phasianus colchicus*) are genetically closely related to coronaviruses of domestic fowl (infectious bronchitis virus) and turkeys*. Avian Pathol, 2002. **31**(1): p. 81-93.
171. Liu, S., et al., *Isolation of avian infectious bronchitis coronavirus from domestic peafowl (*Pavo cristatus*) and teal (*Anas*)*. J Gen Virol, 2005. **86**(Pt 3): p. 719-25.
172. Chu, D.K.W., et al., *Avian coronavirus in wild aquatic birds*. J. Virol., 2011: p. JVI.05838-11.
173. Vijgen, L., et al., *Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event*. J Virol, 2005. **79**(3): p. 1595-604.
174. Woo, P.C.Y., et al., *Molecular diversity of coronaviruses in bats*. Virology, 2006. **351**(1): p. 180-187.
175. Guan, Y., et al., *Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China*. Science, 2003. **302**(5643): p. 276-8.
176. Li, W., et al., *Bats are natural reservoirs of SARS-like coronaviruses*. Science, 2005. **310**(5748): p. 676 - 679.
177. Song, H.D., et al., *Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human*. Proc Natl Acad Sci U S A, 2005. **102**(7): p. 2430-5.
178. Yeh, S.H., et al., *Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: molecular epidemiology and genome evolution*. Proc Natl Acad Sci U S A, 2004. **101**(8): p. 2542-7.

179. Pearks Wilkerson, A.J., et al., *Coronavirus outbreak in cheetahs: lessons for SARS*. *Curr Biol*, 2004. **14**(6): p. R227-8.
180. Heeney, J.L., et al., *Prevalence and implications of feline coronavirus infections of captive and free-ranging cheetahs (*Acinonyx jubatus*)*. *J. Virol.*, 1990. **64**(5): p. 1964-1972.
181. Perlman, S. and A.A. Dandekar, *Immunopathogenesis of coronavirus infections: implications for SARS*. *Nat Rev Immunol*, 2005. **5**(12): p. 917-27.
182. Bergmann, C.C., T.E. Lane, and S.A. Stohlman, *Coronavirus infection of the central nervous system: host-virus stand-off*. *Nat Rev Microbiol*, 2006. **4**(2): p. 121-32.
183. Navas, S., et al., *Murine coronavirus spike protein determines the ability of the virus to replicate in the liver and cause hepatitis*. *J Virol*, 2001. **75**(5): p. 2452-7.
184. Navas, S. and S.R. Weiss, *Murine coronavirus-induced hepatitis: JHM genetic background eliminates A59 spike-determined hepatotropism*. *J Virol*, 2003. **77**(8): p. 4972-8.
185. Phillips, J.J., et al., *Pathogenesis of chimeric MHV4/MHV-A59 recombinant viruses: the murine coronavirus spike protein is a major determinant of neurovirulence*. *J Virol*, 1999. **73**(9): p. 7752-60.
186. Sanchez, C.M., et al., *Targeted recombination demonstrates that the spike gene of transmissible gastroenteritis coronavirus is a determinant of its enteric tropism and virulence*. *J Virol*, 1999. **73**(9): p. 7607-18.
187. Hodgson, T., et al., *Recombinant infectious bronchitis coronavirus Beaudette with the spike protein gene of the pathogenic M41 strain remains attenuated but induces protective immunity*. *J Virol*, 2004. **78**(24): p. 13804-11.
188. Casais, R., et al., *Reverse genetics system for the avian coronavirus infectious bronchitis virus*. *J Virol*, 2001. **75**(24): p. 12359-69.
189. Casais, R., et al., *Recombinant avian infectious bronchitis virus expressing a heterologous spike gene demonstrates that the spike protein is a determinant of cell tropism*. *J Virol*, 2003. **77**(16): p. 9084-9.
190. Lavi, E., et al., *Experimental demyelination produced by the A59 strain of mouse hepatitis virus*. *Neurology*, 1984. **34**(5): p. 597.
191. Cheever, F.S., et al., *A MURINE VIRUS (JHM) CAUSING DISSEMINATED ENCEPHALOMYELITIS WITH EXTENSIVE DESTRUCTION OF MYELIN*. *The Journal of Experimental Medicine*, 1949. **90**(3): p. 181-194.

192. Marten, N.W., S.A. Stohlman, and C.C. Bergmann, *MHV infection of the CNS: mechanisms of immune-mediated control*. *Viral Immunol*, 2001. **14**(1): p. 1-18.
193. Marten, N.W., et al., *Contributions of CD8+ T cells and viral spread to demyelinating disease*. *J Immunol*, 2000. **164**(8): p. 4080-8.
194. Sorensen, O. and S. Dales, *In vivo and in vitro models of demyelinating disease: JHM virus in the rat central nervous system localized by in situ cDNA hybridization and immunofluorescent microscopy*. *J. Virol.*, 1985. **56**(2): p. 434-438.
195. Wege, H., et al., *Coronavirus infection and demyelination. Development of inflammatory lesions in Lewis rats*. *Adv Exp Med Biol*, 1998. **440**: p. 437-44.
196. Morris, V.L., et al., *Characterization of coronavirus JHM variants isolated from wistar furth rats with a viral-induced demyelinating disease*. *Virology*, 1989. **169**(1): p. 127-136.
197. Leparc-Goffart, I., et al., *Targeted recombination within the spike gene of murine coronavirus mouse hepatitis virus-A59: Q159 is a determinant of hepatotropism*. *J Virol*, 1998. **72**(12): p. 9628-36.
198. Cavanagh, D., *Coronaviruses in poultry and other birds*. *Avian Pathol*, 2005. **34**(6): p. 439-48.
199. Cavanagh, D., Gelb, J., *Infectious Bronchitis*, in *Diseases of Poultry*, Y.M. Saif, Editor. 2008, Blackwell Publishing Professional: Ames, Iowa. p. 119-135.
200. Raj, G.D. and R.C. Jones, *Infectious bronchitis virus: Immunopathogenesis of infection in the chicken*. *Avian Pathol*, 1997. **26**(4): p. 677-706.
201. Nakamura, K., et al., *Comparative study of respiratory lesions in two chicken lines of different susceptibility infected with infectious bronchitis virus: histology, ultrastructure and immunohistochemistry*. *Avian Pathol*, 1991. **20**(2): p. 241-57.
202. Saif, L.J., Sestak, K., *Transmissible gastroenteritis virus and porcine respiratory coronavirus*, in *Diseases of Swine*, B. Straw, Zimmerman, J.J., D'Allaire, S., Taylor, D.J., Editor. 2006, Blackwell Publishing: Ames, Iowa.
203. Pedersen, N.C., C.E. Allen, and L.A. Lyons, *Pathogenesis of feline enteric coronavirus infection*. *J Feline Med Surg*, 2008. **10**(6): p. 529-41.
204. Stoddart, M.E., et al., *The sites of early viral replication in feline infectious peritonitis*. *Veterinary Microbiology*, 1988. **18**(3-4): p. 259-271.

205. Rottier, P.J., et al., *Acquisition of macrophage tropism during the pathogenesis of feline infectious peritonitis is determined by mutations in the feline coronavirus spike protein.* J Virol, 2005. **79**(22): p. 14122-30.
206. Olsen, C.W., et al., *Monoclonal antibodies to the spike protein of feline infectious peritonitis virus mediate antibody-dependent enhancement of infection of feline macrophages.* J. Virol., 1992. **66**(2): p. 956-965.
207. Taniguchi, T. and A. Takaoka, *The interferon- $\alpha/\beta$  system in antiviral responses: a multimodal machinery of gene regulation by the IRF family of transcription factors.* Current Opinion in Immunology, 2002. **14**(1): p. 111-116.
208. Clemens, M.J. and A. Elia, *The double-stranded RNA-dependent protein kinase PKR: structure and function.* J Interferon Cytokine Res, 1997. **17**(9): p. 503-24.
209. Meurs, E., et al., *Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon.* Cell, 1990. **62**(2): p. 379-390.
210. Ramana, C.V., et al., *Regulation of c-myc expression by IFN-gamma through Stat1-dependent and -independent pathways.* EMBO J, 2000. **19**(2): p. 263-72.
211. Lee, S.B., et al., *The apoptosis pathway triggered by the interferon-induced protein kinase PKR requires the third basic domain, initiates upstream of Bcl-2, and involves ICE-like proteases.* Virology, 1997. **231**(1): p. 81-8.
212. Goodbourn, S., L. Didcock, and R.E. Randall, *Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures.* Journal of General Virology, 2000. **81**(10): p. 2341-2364.
213. Eleouet, J.F., et al., *The viral nucleocapsid protein of transmissible gastroenteritis coronavirus (TGEV) is cleaved by caspase-6 and -7 during TGEV-induced apoptosis.* J Virol, 2000. **74**(9): p. 3975-83.
214. Belyavsky, M., et al., *Coronavirus MHV-3-induced apoptosis in macrophages.* Virology, 1998. **250**(1): p. 41-9.
215. Collins, A.R., *In vitro detection of apoptosis in monocytes/macrophages infected with human coronavirus.* Clin Diagn Lab Immunol, 2002. **9**(6): p. 1392-5.
216. Kotani, T., et al., *Kinetics of lymphocytic subsets in chicken tracheal lesions infected with infectious bronchitis virus.* J Vet Med Sci, 2000. **62**(4): p. 397-401.
217. Bergmann, C.C., et al., *Perforin-mediated effector function within the central nervous system requires IFN-gamma-mediated MHC up-regulation.* J Immunol, 2003. **170**(6): p. 3204-13.

218. Bergmann, C.C., et al., *Perforin and gamma interferon-mediated control of coronavirus central nervous system infection by CD8 T cells in the absence of CD4 T cells*. J Virol, 2004. **78**(4): p. 1739-50.
219. Guo, X., et al., *Molecular mechanisms of primary and secondary mucosal immunity using avian infectious bronchitis virus as a model system*. Vet Immunol Immunopathol, 2008. **121**(3-4): p. 332-43.
220. Mockett, A.P. and J.H. Darbyshire, *Comparative studies with an enzyme-linked immunosorbent assay (ELISA) for antibodies to avian infectious bronchitis virus*. Avian Pathol, 1981. **10**(1): p. 1-10.
221. Kiss, I., A.M. Poland, and N.C. Pedersen, *Disease outcome and cytokine responses in cats immunized with an avirulent feline infectious peritonitis virus (FIPV)-UCD1 and challenge-exposed with virulent FIPV-UCD8*. J Feline Med Surg, 2004. **6**(2): p. 89-97.
222. Ho, M.S., et al., *Neutralizing antibody response and SARS severity*. Emerg Infect Dis, 2005. **11**(11): p. 1730-7.
223. Armesto, M., D. Cavanagh, and P. Britton, *The Replicase Gene of Avian Coronavirus Infectious Bronchitis Virus Is a Determinant of Pathogenicity*. PLoS ONE, 2009. **4**(10): p. e7384.
224. He, R., et al., *Activation of AP-1 signal transduction pathway by SARS coronavirus nucleocapsid protein*. Biochem Biophys Res Commun, 2003. **311**(4): p. 870-6.
225. Kopecky-Bromberg, S.A., et al., *Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists*. J Virol, 2007. **81**(2): p. 548-57.
226. Narayanan, K., et al., *Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells*. J Virol, 2008. **82**(9): p. 4471-9.
227. Züst, R., et al., *Coronavirus Non-Structural Protein 1 Is a Major Pathogenicity Factor: Implications for the Rational Design of Coronavirus Vaccines*. PLoS Pathog, 2007. **3**(8): p. e109.
228. Wathelet, M.G., et al., *Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain*. J Virol, 2007. **81**(21): p. 11620-33.

229. Li, S.-W., et al., *Severe acute respiratory syndrome coronavirus papain-like protease suppressed alpha interferon-induced responses through downregulation of extracellular signal-regulated kinase 1-mediated signalling pathways*. Journal of General Virology, 2011. **92**(5): p. 1127-1140.
230. Kuri, T., et al., *The ADP-ribose-1"-monophosphatase domains of severe acute respiratory syndrome coronavirus and human coronavirus 229E mediate resistance to antiviral interferon responses*. Journal of General Virology, 2011. **92**(8): p. 1899-1905.
231. Ye, Y., et al., *Mouse Hepatitis Coronavirus A59 Nucleocapsid Protein Is a Type I Interferon Antagonist*. J. Virol., 2007. **81**(6): p. 2554-2563.
232. Frieman, M., et al., *Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane*. J Virol, 2007. **81**(18): p. 9812-24.
233. Hussain, S., S. Perlman, and T.M. Gallagher, *Severe acute respiratory syndrome coronavirus protein 6 accelerates murine hepatitis virus infections by more than one mechanism*. J Virol, 2008. **82**(14): p. 7212-22.
234. Zhao, J., et al., *Severe acute respiratory syndrome coronavirus protein 6 is required for optimal replication*. J Virol, 2009. **83**(5): p. 2368-73.
235. Versteeg, G.A., et al., *Group 2 coronaviruses prevent immediate early interferon induction by protection of viral RNA from host cell recognition*. Virology, 2007. **361**(1): p. 18-26.
236. Zhou, H. and S. Perlman, *Mouse hepatitis virus does not induce Beta interferon synthesis and does not inhibit its induction by double-stranded RNA*. J Virol, 2007. **81**(2): p. 568-74.
237. Gelb, J., Jackwood, M.J, *Infectious Bronchitis*, in *A Laboratory Manual for the Isolation, Identification, and Characterization of Avian Pathogens*, L. Dufour-Zavala, Swayne, D.E., Glisson, J.R., Pearson, J.E., Reed, W. M., Jackwood, M.W., Woolcock, P, Editor. 2008, American Association of Avian Pathologists: Athens, Georgai. p. 146-149.
238. Kwon, H.M., L.J. Saif, and D.J. Jackwood, *Field isolates of transmissible gastroenteritis virus differ at the molecular level from the Miller and Purdue virulent and attenuated strains and from porcine respiratory coronaviruses*. J Vet Med Sci, 1998. **60**(5): p. 589-97.
239. Kim, L., et al., *Development of a reverse transcription-nested polymerase chain reaction assay for differential diagnosis of transmissible gastroenteritis virus and porcine respiratory coronavirus from feces and nasal swabs of infected pigs*. J Vet Diagn Invest, 2000. **12**(4): p. 385-8.

240. Jackwood, M.W., H.M. Kwon, and D.A. Hilt, *Infectious bronchitis virus detection in allantoic fluid using the polymerase chain reaction and a DNA probe*. Avian Dis, 1992. **36**(2): p. 403-9.
241. Kwon, H.M., M.W. Jackwood, and J. Gelb, Jr., *Differentiation of infectious bronchitis virus serotypes using polymerase chain reaction and restriction fragment length polymorphism analysis*. Avian Dis, 1993. **37**(1): p. 194-202.
242. Poon, L.L., et al., *Rapid diagnosis of a coronavirus associated with severe acute respiratory syndrome (SARS)*. Clin Chem, 2003. **49**(6 Pt 1): p. 953-5.
243. Escutenaire, S., et al., *SYBR Green real-time reverse transcription-polymerase chain reaction assay for the generic detection of coronaviruses*. Archives of Virology, 2007. **152**(1): p. 41-58.
244. Herrewegh, A.A., et al., *Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus*. J Virol, 1998. **72**(5): p. 4508-14.
245. Simons, F.A., et al., *A mRNA PCR for the diagnosis of feline infectious peritonitis*. J Virol Methods, 2005. **124**(1-2): p. 111-6.
246. Dye, C., C.R. Helps, and S.G. Siddell, *Evaluation of real-time RT-PCR for the quantification of FCoV shedding in the faeces of domestic cats*. J Feline Med Surg, 2008. **10**(2): p. 167-74.
247. Kennedy, M.A., et al., *Evaluation of antibodies against feline coronavirus 7b protein for diagnosis of feline infectious peritonitis in cats*. Am J Vet Res, 2008. **69**(9): p. 1179-82.
248. Kiss, I., et al., *Preliminary studies on feline coronavirus distribution in naturally and experimentally infected cats*. Res Vet Sci, 2000. **68**(3): p. 237-42.
249. Duarte, A., I. Veiga, and L. Tavares, *Genetic diversity and phylogenetic analysis of Feline Coronavirus sequences from Portugal*. Vet Microbiol, 2009. **138**(1-2): p. 163-8.
250. Herrewegh, A.A., et al., *Detection of feline coronavirus RNA in feces, tissues, and body fluids of naturally infected cats by reverse transcriptase PCR*. J Clin Microbiol, 1995. **33**(3): p. 684-9.
251. Jackwood, M.W., N.M. Yousef, and D.A. Hilt, *Further development and use of a molecular serotype identification test for infectious bronchitis virus*. Avian Dis, 1997. **41**(1): p. 105-10.

252. Carman, S., et al., *Field validation of a commercial blocking ELISA to differentiate antibody to transmissible gastroenteritis virus (TGEV) and porcine respiratory coronavirus and to identify TGEV-infected swine herds.* J Vet Diagn Invest, 2002. **14**(2): p. 97-105.
253. Shi, Y., et al., *Diagnosis of severe acute respiratory syndrome (SARS) by detection of SARS coronavirus nucleocapsid antibodies in an antigen-capturing enzyme-linked immunosorbent assay.* J Clin Microbiol, 2003. **41**(12): p. 5781-2.
254. Kraaijeveld, C.A., S.E. Reed, and M.R. Macnaughton, *Enzyme-linked immunosorbent assay for detection of antibody in volunteers experimentally infected with human coronavirus strain 229 E.* J. Clin. Microbiol., 1980. **12**(4): p. 493-497.
255. Crouch, C.F., T.J. Raybould, and S.D. Acres, *Monoclonal antibody capture enzyme-linked immunosorbent assay for detection of bovine enteric coronavirus.* J. Clin. Microbiol., 1984. **19**(3): p. 388-393.
256. Saif, L.J., *Animal coronavirus vaccines: lessons for SARS.* Dev Biol (Basel), 2004. **119**: p. 129-40.
257. Wesley, R.D. and K.M. Lager, *Increased litter survival rates, reduced clinical illness and better lactogenic immunity against TGEV in gilts that were primed as neonates with porcine respiratory coronavirus (PRCV).* Vet Microbiol, 2003. **95**(3): p. 175-86.
258. Jansen, T., et al., *Structure- and oil type-based efficacy of emulsion adjuvants.* Vaccine, 2006. **24**(26): p. 5400-5.
259. Ladman, B.S., et al., *Protection of chickens after live and inactivated virus vaccination against challenge with nephropathogenic infectious bronchitis virus PA/Wolgemuth/98.* Avian Dis, 2002. **46**(4): p. 938-44.
260. Brooks, J.E., et al., *Comparisons of envelope through 5B sequences of infectious bronchitis coronaviruses indicates recombination occurs in the envelope and membrane genes.* Virus Res, 2004. **100**(2): p. 191-8.
261. Domingo, E., *Rapid Evolution of Viral RNA Genomes.* The Journal of Nutrition, 1997. **127**(5): p. 958S-961S.
262. Borrego, B., et al., *Distinct repertoire of antigenic variants of foot-and-mouth disease virus in the presence or absence of immune selection.* J Virol, 1993. **67**(10): p. 6071-9.
263. Steinhauer, D., E. Domingo, and J. Holland, *Lack of evidence for proofreading mechanisms associated with an RNA polymerase.* Gene, 1992. **122**: p. 281 - 288.

264. Drake, J.W., *Rates of Spontaneous Mutation Among RNA Viruses*. Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(9): p. 4171-4175.
265. Hanada, K., Y. Suzuki, and T. Gojobori, *A Large Variation in the Rates of Synonymous Substitution for RNA Viruses and Its Relationship to a Diversity of Viral Infection and Transmission Modes*. Molecular Biology and Evolution, 2004. **21**(6): p. 1074-1080.
266. Holmes, E.C., *The Evolution and Emergence of RNA viruses*. Oxford Series in Ecology and Evolution, ed. P.H. Harvey, and May, Robert M. 2009, New York: Oxford University Press. 254.
267. Zhang, X.W., Y.L. Yap, and A. Danchin, *Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus*. Arch Virol, 2005. **150**(1): p. 1-20.
268. Ksiazek, T.G., et al., *A novel coronavirus associated with severe acute respiratory syndrome*. N Engl J Med, 2003. **348**(20): p. 1953-66.
269. Poon, L.L., et al., *Identification of a novel coronavirus in bats*. J Virol, 2005. **79**(4): p. 2001-9.
270. Dominguez, S.R., T. O'Shea, L.M. Oko, and K.V. Holmes, *Detection of Group 1 Coronaviruses in Bats in North America*. Emerging Infectious Diseases, 2007. **13**(9).
271. Vijaykrishna, D., et al., *Evolutionary Insights into the Ecology of Coronaviruses*. J. Virol., 2007. **81**(8): p. 4012-4020.
272. Butler, N., et al., *Murine encephalitis caused by HCoV-OC43, a human coronavirus with broad species specificity, is partly immune-mediated*. Virology, 2006. **347**(2): p. 410-21.
273. Jin, L., et al., *Analysis of the genome sequence of an alpaca coronavirus*. Virology, 2007. **365**(1): p. 198-203.
274. Hasoksuz, M., et al., *Biologic, Antigenic, and Full-Length Genomic Characterization of a Bovine-Like Coronavirus Isolated from a Giraffe*. J. Virol., 2007. **81**(10): p. 4981-4990.
275. Alekseev, K.P., et al., *Bovine-like coronaviruses isolated from four species of captive wild ruminants are homologous to bovine coronaviruses, based on complete genomic sequences*. J Virol, 2008. **82**(24): p. 12422-31.
276. Gao, F.-S., et al., *Isolation and identification of a canine coronavirus strain from giant pandas (*Ailuropoda melanoleuca*)*. Journal of Veterinary Science, 2009. **10**(3): p. 261-263.

277. Wise, A.G., M. Kiupel, and R.K. Maes, *Molecular characterization of a novel coronavirus associated with epizootic catarrhal enteritis (ECE) in ferrets*. *Virology*, 2006. **349**(1): p. 164-74.

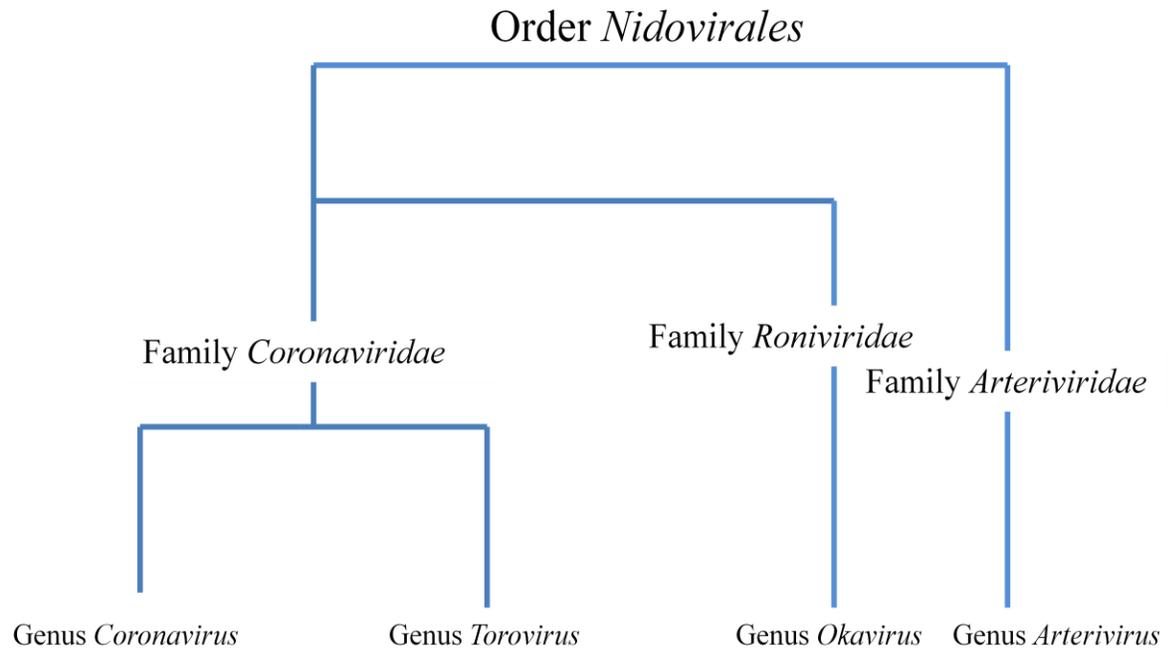


Figure 2.1: Schematic of older Order *Nidovirales* classification.

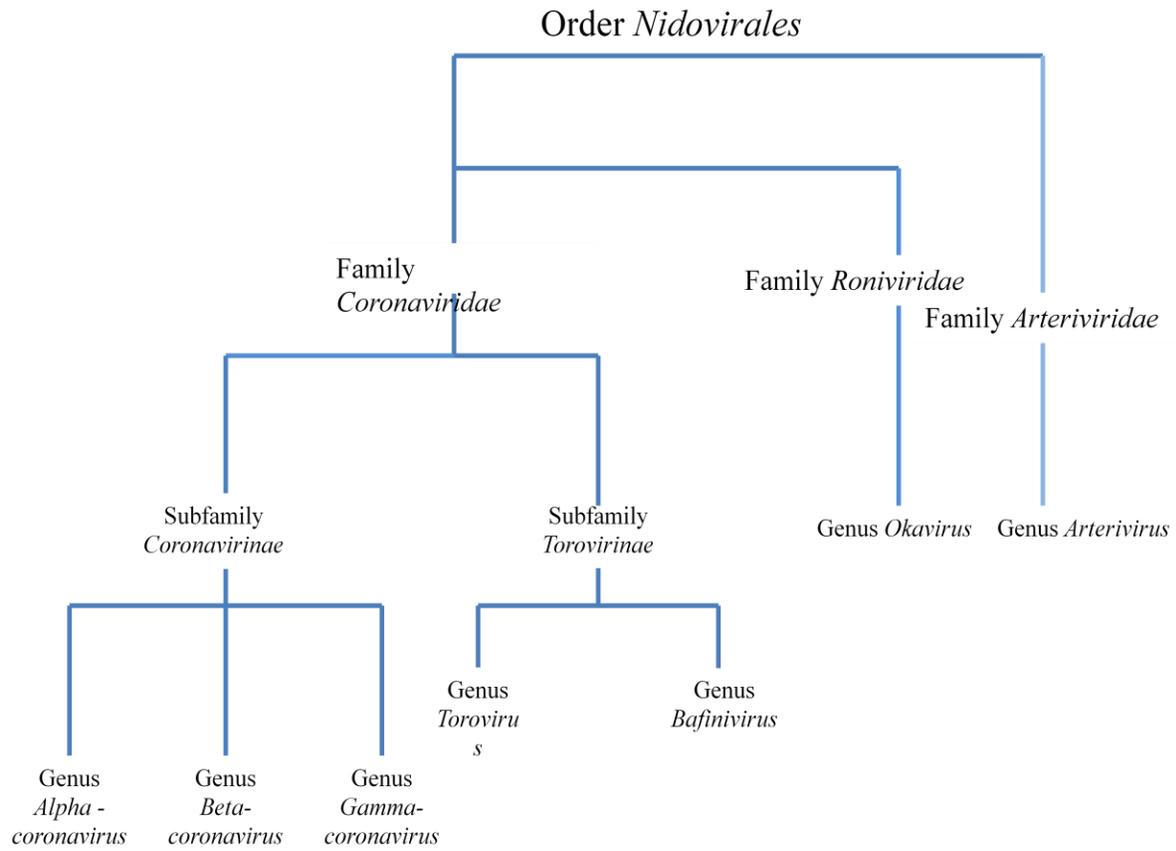
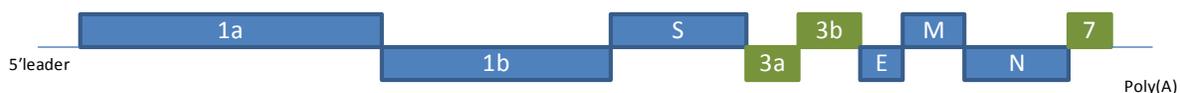
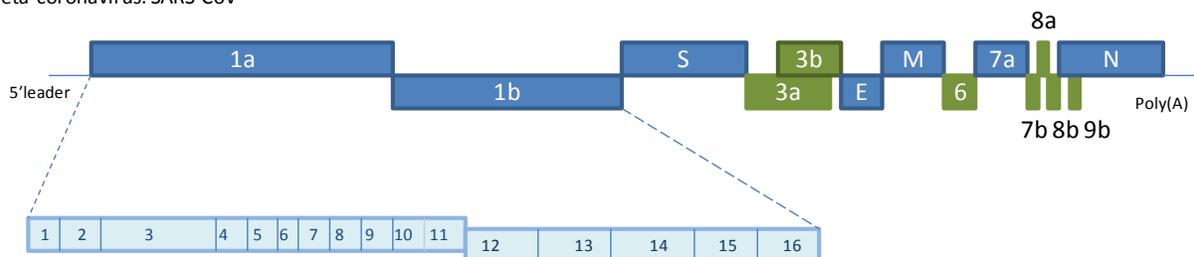


Figure 2.2: Schematic of current Order *Nidovirales* classification.

Alpha-coronavirus: TGEV



Beta-coronavirus: SARS-CoV



Gamma-coronavirus: IBV



Figure 2.3: Coronavirus genomes. Representative genome organization for alpha, beta, and gamma coronaviruses are shown. Each diagram shows the approximate location and name of the ORFs coded for. ORF 1a and 1b, the 5' 2/3 of the genome, encode 15-16 non-structural proteins that are proteolytically cleaved. The 16 nsps for SARS-CoV are illustrated. The 3' 1/3 of the genome encode the structural proteins; spike (S), membrane (M), and nucleocapsid (N), colored blue, and the virus specific accessory proteins, colored green. The hemagglutinin esterase (HE) protein encoded by some beta-coronaviruses is not shown.

## CHAPTER 3

## RECOMBINATION IN AVIAN GAMMA-CORONAVIRUS INFECTIOUS BRONCHITIS

VIRUS<sup>1</sup>

<sup>1</sup>Thor, S.W., Hilt, D.A. Kissinger, J.C., Paterson, A. H., and Jackwood, M.W..  
2011. *Viruses*. Sep; 3 (9): 1777-99  
Reprinted here with permission of the publisher

© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

**Abstract**

Recombination in the family Coronaviridae has been well documented and is thought to be a contributing factor in the emergence and evolution of different coronaviral genotypes as well as different species of coronavirus. However, there are limited data available on the frequency and extent of recombination in coronaviruses in nature and particularly for the avian gamma-coronaviruses where only recently the emergence of a turkey coronavirus has been attributed solely to recombination. In this study, the full-length genomes of eight avian gamma-coronavirus infectious bronchitis virus (IBV) isolates were sequenced and along with other full-length IBV genomes available from GenBank were analyzed for recombination. Evidence of recombination was found in every sequence analyzed and was distributed throughout the entire genome. Areas that have the highest occurrence of recombination are located in regions of the genome that code for nonstructural proteins 2, 3 and 16, and the structural spike glycoprotein. The extent of the recombination observed, suggests that this may be one of the principal mechanisms for generating genetic and antigenic diversity within IBV. These data indicate that reticulate evolutionary change, due to recombination in IBV, likely plays a major role in the origin and adaptation of the virus leading to new genetic types and strains of the virus.

Keywords: gamma coronavirus; avian coronavirus; infectious bronchitis virus; genome; recombination

## Introduction

Avian infectious bronchitis virus (IBV) is a gamma-coronavirus in the *Family Coronaviridae*, the *Order Nidovirales*, and the *Genus Coronavirus* that causes a highly contagious upper-respiratory disease of domestic chickens. In layer type birds it can cause a drop in egg production and some strains are nephropathogenic. Infectious bronchitis remains one of the most widely reported respiratory diseases of chickens worldwide despite the routine usage of attenuated live vaccines to control the disease. Control of IBV is difficult because there is little to no cross-protection between the numerous different serotypes of the virus.

Infectious bronchitis virus is an enveloped, single-stranded, positive-sense RNA virus with a genome length of approximately 27 kb. The 3' end of the genome encodes four structural proteins; spike (S), envelope (E), membrane (M) and nucleocapsid (N) as well as several non-structural proteins [1]. The S glycoprotein of IBV forms projections on the surface of the virion. Spike is post-translationally cleaved into S1 and S2 subunits with the S1 subunit forming the outermost portion and S2 forming a stalk-like structure that is embedded in the viral membrane. The S1 subunit contains hypervariable regions that play a role in attachment to host cell receptors, and it contains conformationally-dependent virus-neutralizing and serotype-specific epitopes [2, 3]. Spike is also involved in membrane fusion and viral entry into the host cell. The E and M proteins are integral membrane proteins involved in assembly of the virus. The N protein is closely associated with the viral genome and plays a role in replication. The 5' two-thirds of the genome, approximately 21 kb, encodes two polyproteins 1a and 1ab. A -1 frameshift mechanism is used to translate the 1ab polyprotein. The polyproteins are post-translationally cleaved into 15 non-structural proteins (nsps), nsp 2–16 (IBV does not have an

nsp1) that make up the replication complex. Key nsps encoded, include a papain-like protease 2 (PLP2) within nsp 3, a main protease (Mpro) within nsp 5, and the RNA-dependent RNA-polymerase (RdRp) within nsps 11 and 12.

Genetic diversity in coronaviruses is due to adaptive evolution driven by high mutation rates and genetic recombination [4]. High mutation rates are attributed to minimal proof reading capabilities associated with the RdRp. Recombination is thought to be due to a unique template switching “copy-choice” mechanism during RNA replication [5]. Evidence of recombination among strains of IBV has been observed both experimentally and in the field [6–11]. The emergence of several alpha- and beta-coronaviruses has been attributed to recombination [12, 13] but only recently was recombination shown to be the mechanism behind the emergence of a novel gamma-coronavirus, turkey coronavirus (TCoV) [14]. Although “hot spots” of recombination in the genome of IBV have been reported [9, 15], a thorough study of recombination using multiple different strains across the entire genome has not been conducted.

In this study we sequenced and analyzed the entire genome of eight IBV strains that represent different serotypes that have not been previously sequenced, and we compared these sequences with other gamma-coronavirus full-length genome sequences available in GenBank for evidence of recombination [16]. Different serotypes of field viruses and vaccine type viruses were selected to provide a wide variety of sequences potentially capable of contributing gene fragments to recombinants.

## Results and Discussion

### *Sequence Analysis*

The full-length genomes of eight isolates of IBV were sequenced at 5× to 10× coverage, and the consensus sequences were assembled. The genome size (see the end of the 3'UTR in Table 1), organization of the genome, and the location and size of the open reading frames (ORF's) are listed in Table 1 for each of the viruses. The gene order is the same for all the viruses examined; 5'UTR-1a/ab-spike-3a-3b-Envelope-Membrane-4b-4c-5a-5b-Nucleocapsid-3'UTR. In addition, the genomes for CAV/CAV56b/91, DE/DE072/92, FL/FL18288/71, Mass/H120, Iowa/Iowa97/56 and JMK/JMK/64 contain ORF 6b between nucleocapsid and the 3'UTR.

The full-length genomes were aligned and phylogenetic trees were constructed using the Neighbor-joining, Minimum Evolution, Maximum Parsimony and UPGMA programs in MEGA4 [17]. The trees all had similar topology and bootstrap support, and a representative tree is shown in Figure 1. The feline coronavirus FCoV/FIPV/WSU-79-1146 and the beluga whale virus BelugaWhaleCoV/SW1/08 were included as out-groups. The wild bird viruses isolated from a munia (MuniaCoV/HKUY13/09), thrush (ThrushCoV/HKU12/09) and bulbul (BulBulCoV/HKU11/09) formed a unique clade, which is not surprising as this group might represent a new coronavirus genus provisionally designated Deltacoronavirus [18]. The remaining viruses separated into clades consisting of IBV isolates from the US and vaccine viruses, TCoV isolates, an IBV isolate from West Africa and IBV isolates from China and Taiwan.

Vaccines for IBV used in commercial poultry include the serotypes Mass, Conn, DE and Ark. The PeafowlCcV/GD/KQ6/03, CK/CH/LSD/051/06 and CK/CH/ZJ971/97 strains from China grouped with Mass type viruses indicating that they are closely related, which is not surprising since Mass type vaccines are used in China. The overall percent similarities between the various strains are listed in Supplemental Table 3.1. All IBV genomes examined are greater than 80% similar at the nucleotide level.

### *Recombination Analysis*

Recombination among coronaviruses reduces mutational load, creates genetic variation, and can result in the emergence of new strains [19]. However, evolutionary history is traditionally represented using a strictly bifurcating phylogenetic tree, which implies that once two lineages are created they subsequently never interact with each other. When evolutionary events such as reassortment, horizontal gene transfer or recombination occur, reticulations among the phylogenetic tree branches can result. Using the Neighbor-net analysis we observed evidence of networked relationships (represented by boxes, in Figure 3.2) among the analyzed sequences. Since the boxes only imply the possibility of recombination, we conducted a pairwise homoplasy index (PHI) test, which showed a significant difference in the compatibility between closely linked sites ( $p < 0.0001$ ) supporting the occurrences of recombination among the viruses [20].

The Recombination Detection Program 4 (RDP4) [21, 22] was used to identify recombination breakpoint positions in full-length IBV genome sequences and the data for 50 of a total 135 unique transferred fragments with statistical support of  $p \leq 1 \times 10^{-12}$  are listed in Table 3.2. The full-length genomes available in the database for CK/CH/EP3, CK/CH/p65, and Mass/Beaudette

were excluded from the analysis because they are viruses not found in the field. The recombination programs can be used to detect recombination without reference sequences, and our analysis was conducted without regard to date of isolation because that information was not available for some of the viruses. Although the programs attempt to identify major and minor parent sequences contributing to each recombinant, the data reported herein only represents sequences in other viruses that are most closely related to a majority of the transferred fragment (major sequence) or closely related to some of the transferred fragment (minor sequences) and doesn't imply origin or source of the transferred fragment. In many cases, the transferred fragment has undergone mutations making it difficult to identify all the endpoints for the major and minor sequences. In addition, some of the transferred fragments overlap suggesting that recombinations have occurred between recombinant viruses.

Twenty-five IBV strains were examined and the viruses with the most transferred fragments in Table 2 are CAV/56b/91 and Mass/H52 both with 8 fragments, and CK/CH/LSD/051/06 and GA98/0470/98 both with 7 fragments. The strains with the fewest transferred fragments are Iowa/Iowa97/56 and TW/2575/98 with only 2 transferred fragments and the CK/CH/BJ/97, Holte/Holte/54, and NGA/A116E7/06 strains with only 1 transferred fragment. The Ark/Ark-DPI-p11/81 and Ark/Ark-DPI-p101/81 strains are the same virus that was passaged in embryonated eggs 11 and 101 times respectively. Both viruses share identical transferred fragments indicating that they have identical recombination history. In addition, Conn/Conn46/66 and Conn/Conn46/91 share the same recombination history (4 identical transferred fragments). The Conn/Conn46/66 field virus was used to produce an attenuated live vaccine, which is currently used in commercial poultry. Viruses that share the same recombination history are likely derived from the same parent virus suggesting that

Conn/Conn46/91 is reisolated Conn vaccine derived from the Conn/Conn46/66 virus. The FL/FL18288/71 virus also shares all 4 transferred fragments with the Conn viruses, however; FL/FL18288/71 and Conn viruses are different serotypes suggesting that FL/FL18288/71 is a field virus that emerged due to point mutations accumulating in spike over time rather than from recombination.

All 6 of the transferred fragments in CK/CH/ZJ971/97 are identical to all 6 of the transferred fragments in vaccine strain Mass/H120, providing compelling evidence that CK/CH/ZJ971/97 is reisolated Mass/H120 vaccine. That observation was also reported by Zhang *et al.* [23]. It is interesting that Mass/H52 (8 transferred fragments) and Mass/H120 (6 transferred fragments) share only 5 identical transferred fragments. The Mass/H52 and Mass/H120 viruses were isolated circa 1955 in the Netherlands and it is widely accepted that H stands for Holland, but it actually stands for Houben, the owner of the broiler farm where the viruses were isolated [24]. It is thought that Mass/H120 was derived from Mass/H52 but the actual relationship between the viruses is not certain. Our data indicates that they are not necessarily parent and progeny but they are closely related.

The Gray/Gray/60 and JMK/JMK/64 viruses share 99.7% nucleotide similarity across the entire genome and have 4 identical transferred fragments with JMK/JMK/64 having one additional fragment located in the 5'UTR, which is not found in Gray/Gray/60. It is well known that the Gray/Gray/60 virus is nephropathogenic, whereas the JMK/JMK/64 virus is strictly respirotropic. Perhaps sequence differences in the 5'UTR, which is involved in replication of the viral genome, play a role in the different pathobiologies observed for these viruses.

There is evidence that some transferred fragments in field viruses come from vaccines. As an example, CK/CH/LSD/051/06 has 3 of 7 and 2 of 7 transferred fragments in common with vaccine strains Mass/H52 and Mass/H120 respectively. In addition, the only fragments that USA viruses have in common with the viruses from China and Taiwan are fragments also associated with Mass type vaccines, which are used in both regions, providing further evidence that some of the fragments in field viruses come from vaccines. That result and the observation in Figure 1 that the viruses separated into clades based on geographic location also supports the conclusion that USA viruses have not recombined with Asian viruses.

A difference in the order of taxa in phylogenetic trees constructed from different regions of the genome is further evidence of recombination [25]. To examine the ordering of taxa in sequential trees, the TreeOrder Scan program (ver. 1.4) [26, 27] was used and inconsistent phylogenetic relationships were observed for all of the examined virus strains across the entire genome, indicating a substantial amount of recombination (data not shown). There is a high number of breakpoints in the 1a region of the genome and immediately upstream of the S gene, which has been previously shown to be a 'hot spot' for recombination [9]. A phylogenetic compatibility matrix constructed at the 70% bootstrap level for 250 bp sequence fragments at 100 bp intervals also showed that recombination breakpoints were distributed throughout the IBV genomes (data not shown).

To determine recombination hot and cold spots, a recombination breakpoint distribution plot (Figure 3.3) was generated in RDP4 using a 200 nt window and 1000 permutations [21]. No global hot-spot regions were observed in the 95% and 99% confidence thresholds (dotted lines at the top of the graph). The detectable recombination breakpoint positions are shown at the top of the figure and were distributed throughout the genome with a relatively high number clustered

just upstream of the S gene. That region also had the highest breakpoint count within the 99% local hot/cold-spot confidence interval. A high number of breakpoints were also observed in the 1a region of the genome; nsp 2, nsp 3, and nsp 16, in the envelope and matrix protein genes and in a small area near the 3' UTR. Table 3 shows that nsp2, nsp3, nsp16 and spike genes were associated with the greatest number of transferred fragments, which is consistent with the location and number of breakpoints in Figure 3.3.

Recombination in the 1ab ORF area, which encodes the nonstructural proteins involved in the viral replication complex, has the potential to alter the pathogenicity of the virus [28]. The nsp 2 contains hydrophobic residues that likely anchor the replication complex to the Golgi [29]. The nsp 3 encodes the protease PLP2 which cleaves nsps 2, 3, and 4 and an area with ADP-ribose 1'-phosphatase (ADRP) activity. The protease PLP2 has been shown to have deubiquinating-like activity [30] and also to be a type I interferon (IFN) antagonist [31]. Changes in the amino acid composition of this area could affect the ability of the virus to replicate in a variety of cell types. The ADRP region of nsp 3 is conserved among coronaviruses [32, 33], and a recent study suggested a biological role for the coronavirus ADRP in modulating the expression of pro-inflammatory immune modulators such as tumor necrosis factor alpha and interleukin-6 [34]. Recombination in this area could alter the pathogenicity of the virus by modulating host cytokine expression. The nsp16 is reported to be an S-adenosyl-L-methionine (AdoMet)-dependent RNA (nucleoside-2'*O*)-methyltransferase (2'*O*-MTase) responsible for capping the viral mRNA nascent transcripts [32]. An alteration in the efficiency of this protein could profoundly decrease not only viral replication but also pathogenicity. The spike glycoprotein of IBV on the surface of the virus plays a role in attachment to host cell receptors,

membrane fusion and entry into the host cell. It also contains conformationally-dependent epitopes that induce virus-neutralizing and serotype specific antibodies [2, 3]. We and others [6–8, 10] have observed a relatively high number of breakpoints in and immediately upstream of spike, and changes to this region of the genome can result in the emergence of new genotypes and serotypes of IBV as well as new avian coronaviruses (*i.e.*, TCoV). The envelope and matrix proteins are associated with virus assembly, and changes in those proteins could reduce the efficiency of virus particle formation and subsequent transmission of the virus. The 3' UTR is involved with binding of the viral RdRp and viral genome replication. Changes to the 3' UTR could affect replication efficiency and thus virulence of the virus.

## **Conclusions**

In this study, evidence was obtained that recombination is occurring among avian coronavirus IBV isolates across their entire genome. Every sequence included in the analysis was recognized as a potential recipient of horizontally acquired sequences at some point in its viral evolutionary past. The nsp2, nsp3, nsp16 were associated with the greatest number of transferred fragments. In addition, the area immediately upstream of the spike gene had the highest number of recombination breakpoints. Breakpoints in the 1ab polyprotein gene have the potential to alter pathogenicity of the virus, and breakpoints near or in spike have the potential to lead to the emergence of new serotypes of IBV or new coronaviruses. Although the spike region determines the serotype of the virus, the remainder of the genome may be a mosaic of sequence fragments from a variety of gamma-coronaviruses. The only evidence of a gamma-coronavirus possibly recombining with an alpha or beta-coronavirus was the discovery of the mosaic nature of the

SARS-coronavirus genome [35]. Although this type of recombination is possible it appears to be rare in nature.

In this study, we characterized recombination in the full-length genomes of avian gamma-coronavirus IBV strains from around the world. Our bioinformatic analysis was similar to a previous study on enteroviruses [36] and we found that recombination in IBV is more extensive than formerly thought, involving regions across the entire genome. Our data suggests that reticulate evolution due to a high frequency of recombination in IBV, likely plays a major role in the generation of new serotypes of the virus. The characterization, distribution and frequency of recombination breakpoints are important information that will further our understanding of the mechanisms behind the diversity and evolution of these viruses so better control methods can be developed.

## **Materials and Methods**

### *Viruses and Viral RNA Extraction*

All of the viruses sequenced in this study (Table 4), were propagated in 9-11 day-old specific-pathogen-free (SPF) embryonated eggs as described [37]. Total RNA was isolated from 200 ul of allantoic fluid collected from the infected eggs using the High Pure RNA Isolation Kit (Roche Applied Science, Mannheim, Germany) following the manufacturer's instructions.

### *RT-PCR Amplification and Sequencing*

The amplification reactions were carried out using strand displacement RT-PCR or one step RT-PCR. Strand displacement RT-PCR uses a random (at the 3' prime end) primer and an

amplification primer. The sequence of the random primer was (AGCGGGGGTTGTCTGAATGTTTGANNNN) and the sequence of the amplification primer was (AGCGGGGGTTGTCTGAATGTTTGA). The RT-PCR reaction was carried out using the TaKaRa RNA LA PCR kit (Takara Bio. Inc., Otsu, Shiga, Japan), according to the manufacturer's protocol. A DNA Engine Peltier thermocycler (Bio-Rad Laboratories Inc., Hercules, CA) was used for the RT reaction, which included an RNA denaturing step at 65 C for 10 min followed by 30 C for 10 min, 42 C for 60 min, 99 C for 5 min, and 5 C for 5 min. The PCR reaction was run on the same machine as the RT step and included a one-time initial denaturation step of 94 C for 2 min., followed by 30 cycles of 94 C for 30 s, 60 C for 30 s and 72 C for 3 min.

The PCR products were agarose gel purified using the QIAquick gel extraction kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. The PCR products were cloned into the TOPOXL vector using the TOPOXL cloning kit (Invitrogen, Carlsbad, CA) according to manufacturer's protocol to prepare cDNA libraries for sequencing.

Plasmid DNA from the libraries of the cloned cDNA fragments for each virus was isolated using an alkaline lysis method modified for the 96-well format and incorporating both Hydra and Tomtek robots. Sequencing reactions were performed using the BigDye™ Terminator® Cycle Sequencing Kit Version 3.1 (Applied Biosystems, Foster City, CA) and MJ Research (Watertown, MA) thermocyclers. Sephadex filter plates were used to filter each reaction into Perkin-Elmer MicroAmp Optical 96-well plates. A 1/12-strength sequencing reaction on an ABI 3730 was used to sequence each clone from both the 5' and 3' ends.

Primers for one-step RT-PCR were specifically designed for each virus (Supplemental Table 2). Viral RNA was amplified using the Titan One Tube RT-PCR kit (Roche Diagnostics, Indianapolis, IN) following manufacturer's instructions. A DNA Engine Peltier Thermocycler

(Bio-Rad Laboratories, Inc.) was used for the RT-PCR reaction, which had the following steps: one cycle of 42 C for 60 min and 95 C for 5 min., followed by 10 cycles of 94 C for 30 s, 50 C for 30 s, and 68 C for 1min 30 s., and then 25 cycles of 94 C for 30 s, 50 C for 30 s, 68 C for 1 min and 30 s adding 5 s with each cycle.

The resulting PCR products were agarose gel purified using the QIAquick gel extraction kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. The resulting cDNA was sequenced using ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA) following the manufacturer's protocol. The reactions were prepared for sequencing by centrifugation through either a Centri-Sep column (Applied Biosystems, Foster City, CA) or using the Edge system (EdgeBio, Gaithersburg, MD) plate. The samples were sequenced at the Georgia Genomics Facility (University of Georgia, Athens, GA).

### *Genome Assembly and Analysis*

Chromatogram files and trace data were read and assembled using SeqMan Pro, and genome annotation was conducted with SeqBuilder (DNASTAR, Inc., Madison, WI). Each sequence was aligned to a representative genome; Mass/Mass41/41 (GenBank accession #AY851295), or CAL99/CAL99/99 (GenBank accession #AY514485) as a backbone for genome assembly.

Whole genome analyses were generated and phylogenetic trees constructed with the Neighbor-Joining method with 1000 bootstrap replicates as well as with Minimum Evolution, Maximum Parsimony and UPGMA methods [17].

### *GenBank Accession Numbers*

Virus genome sequences generated in this study were submitted to GenBank and assigned the following accession numbers: CAV/CAV56b/91 (GU393331), DE/DE072/92 (GU393332), FL/FL18288/71 (GU393333), Gray/Gray/60 (GU393334), Mass/H120 (GU393335), Holte/Holte/54 (GU393336), Iowa/Iowa97/56 (GU393337), JMK/JMK/64 (GU393338).

GenBank accession numbers for full-length sequences used as reference in this study are: Mass/Mass41/41 (AY851295), Mass/H52 (EU817497), Ark/Ark-DPI-p11/81 (EU418976), Ark-DPI-p101/91 (EU418975), CAV/CAV99/99 (AY514485), CK/CH/EP3 (DQ001338), CK/CH/p65 (DQ001339), Mass/Beaudette (NC\_001451), NGA/A116E7/06 (FN430415), ITA/90254/05 (FN430414), TW/TW2575/98 (DQ646405), CK/CH/SC021202/02 (EU714029), CK/CH/ZJ971/97 (EU714028), CK/CH/BJ/97 (AY319651), CK/CH/SAIBK (DQ288927), CK/CH/LSD/051/06 (EU637854), CK/CH/DY07/07 (HM245923), CK/CH/CQ04-1/04 (HM245924), GA98/GA98/98 (GQ504723), PeafowlCoV/GD/KQ6/03 (AY641576), PartridgeCoV/GD/S14/03 (AY646283), TCoV/IN-540/94 (EU022525), TCoV/MN-ATCC (EU22526), TCoV/VA-74/03 (GQ427173), TCoV/TX-GL/01 (GQ427174), TCoV/IN-517/94 (GQ427175), TCoV/TX-1038/98 (GQ427176), TCoV/Canada-MG10 (EU095850) BulbulCoV/HKU11/09 (FJ376619), ThrushCoV/HKU12/09 (FJ376621), MuniaCoV/HKU13/09 (FJ376622), BelugaWhaleCoV/SW1/08 (NC\_010646), FCoV/FIPV/WSU-79/1146 (DQ010921).

### *Detection of Networked Relationships and Recombination Break Points*

We used Neighbor-net analysis to examine the IBV genomes for evidence of networked relationships and the pairwise homoplasy index (PHI) in SplitsTree (ver. 4) [20, 38, 39] to

statistically determine the likelihood of recombination. The Recombination Detection Program (RDP4, ver. 4) was used to analyze the IBV genomes for recombination breakpoints [21, 22]. Unless otherwise stated, default settings were used in all of the programs. The specific algorithms used were RDP [40], GENECONV [41], BOOTSCAN/RESCAN [40], MAXIMUM CHI SQUARE [42], CHIMAERA [43], SISCAN [44], and 3Seq [45]. We used more than one method to analyze the data because evaluation of these recombination detection methods using both simulated and empirical data showed that the results from only a single method were not very reliable [46]. Automasking was used for optimal recombination detection. The RDP analysis was run without a reference and a window size of 60, BOOTSCAN window size was increased to 500, MAXCHI and CHIMAERA number of variable sites per window was increased to 120, and the window size and step size for SISCAN was increased to 500 and 20, respectively. The window sizes were increased from their default settings because IBV has a high mutation rate, which can mask recombination signals. Increasing the window size was shown to increase the ratio of recombination signals relative to mutational “noise” [47].

#### *Phylogenetic Analysis of Sequential Genome Fragments*

Inconsistent phylogenetic relationships between different regions of the viral genome provide further evidence of genetic recombination. Herein, we examined the order of avian gamma-coronavirus IBV strains in phylogenetic trees generated from sequential genome fragments using TreeOrder Scan (ver. 1.6) [26, 27]. Changes in the tree position of taxa supported at the 70% or greater bootstrap level for a 250 bp sequence window were examined at 100 bp intervals. In addition, a phylogenetic compatibility matrix was constructed using

TreeOrder Scan to examine the frequency and location of recombinations across the entire genome.

#### *Recombination Site detection*

Potential recombination sites were identified using the RDP4 software [22] and a breakpoint map was constructed. A breakpoint density plot was then created from this map by moving a 200 nt window 1 nt at a time along the length of the map. The number of breakpoints falling within a window was plotted at the central window position. A 99% (upper) and 95% (lower) confidence threshold for globally significant breakpoint clusters (defined as windows with more breakpoint positions than the maximum found in >95% of the 1,000 permuted plots) was calculated. In addition, 99% and 95% confidence intervals were calculated for local breakpoint clusters (defined as windows with more breakpoint positions than the maximum found in >99% of the windows at that location in 1,000 permuted plots).

#### **Acknowledgements**

This work was supported by USDA, CSREES award number 2007-35600-17786. The authors would like to thank the technical help of Jon S. Robertson and Cornelia Lemke with sequencing.

## References

1. Lai, M.M.C.; Holmes, K.V., Coronaviridae: The viruses and their replication. In *Fields Virology*, 4th ed.; Knipe, D.M.; Howley, P.M.; Griffin, D.E.; Lamb, R.A.; Martin, M.A.; Roizman, B.; Straus, S.E. Eds.; Lippincott Williams & Wilkins: Philadelphia, PA, USA, **2001**; Volume. 1, pp. 1163-1185.
2. Cavanagh, D.; Mawditt, K.; Adzhar, A.; Gough, R.E.; Picault, J.P.; Naylor, C.J.; Haydon, D.; Shaw, K.; Britton, P. Does IBV change slowly despite the capacity of the spike protein to vary greatly? *Adv. Exp. Med. Biol.* **1998**, *440*, 729-734.
3. Niesters, H.G.; Kusters, J.G.; Lenstra, J.A.; Spaan, W.J.; Horzined, M.C.; van der Zeijst, B.A. The neutralization epitopes on the spike protein of infectious bronchitis virus and their antigenic variation. *Adv. Exp. Med. Biol.* **1987**, *218*, 483-492.
4. Holmes, E.C. *The Evolution and Emergence of RNA Viruses*. First Ed.; Oxford University Press Inc.: New York, NY, USA, **2009**.
5. Lai, M. RNA Recombination in Animal and Plant Viruses. *Microbiol Rev.* **1992**, *56*, 61-79.
6. Kusters, J.G.; Jager, E.J.; Niesters, H.G.M.; van der Zeijst, B.A.M. Sequence evidence for RNA recombination in field isolates of avian coronavirus infectious bronchitis virus. *Vaccine* **1990**, *8*, 605-608.
7. Jia, W.; Karaca, K.; Parrish, C.R.; Naqi, S.A. A novel variant of avian infectious bronchitis virus resulting from recombination among three different strains. *Arch. Virol.* **1995**, *140*, 259-271.

8. Lee, C.W.; Jackwood, M.W. Origin and evolution of Georgia 98 (GA98), a new serotype of avian infectious bronchitis virus. *Virus Res.* **2001**, *80*, 33-39.
9. Lee, C.W.; Jackwood, M.W. Evidence of genetic diversity generated by recombination among avian coronavirus IBV. *Arch. Virol.* **2000**, *145*, 2135-48.
10. Estevez, C.; Villegas, P.; El-Attrache, J. A recombination event, induced in ovo, between a low passage infectious bronchitis virus field isolate and a highly embryo adapted vaccine strain. *Avian Dis.* **2003**, *47*, 1282-1290.
11. Mardani, K.; Noormohammadi, A.H.; Ignjatovic, J.; Browning, G.F. Naturally occurring recombination between distant strains of infectious bronchitis virus. *Arch. Virol.* **2010**, *155*, 1581-1586.
12. Woo, P.C.; Lau, S.K.; Huang, Y.; Yuen, K.Y. Coronavirus diversity, phylogeny and interspecies jumping. *Exp. Biol. Med.* **2009**, *234*, 1117-1127.
13. Decaro, N.; Mari, V.; Campolo, M.; Lorusso, A.; Camero, M.; Elia, G.; Martella, V.; Cordioli, P.; Enjuanes, L.; Buonavoglia, C. Recombinant canine coronaviruses related to transmissible gastroenteritis virus of Swine are circulating in dogs. *J. Virol.* **2009**, *83*, 1532-1537.
14. Jackwood, M.W.; Boynton, T.O.; Hilt, D.A.; McKinley, E.T.; Kissinger, J.C.; Paterson, A.H.; Robertson, J.; Lemke, C.; McCall, A.W.; Williams, S.M.; Jackwood, J.W.; Byrd, L.A. Emergence of a group 3 coronavirus through recombination. *Virology* **2010**, *398*, 98-108.
15. Lee, C.W.; Hilt, D.A.; Jackwood, M.W. Identification and analysis of the Georgia 98 serotype, a new serotype of infectious bronchitis virus. *Avian Dis.* **2001**, *45*, 164-172.
16. National Center for Biotechnology Information. [www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)

17. Tamura, K.; Dudley, J.; Nei, M.; Kumar, S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **2007**, *24*, 1596-1599.
18. Woo, P.C.Y.; Huang, Y.; Lau, S.K.; Yuen, K.Y. Coronavirus genomics and bioinformatics analysis. *Viruses* **2010**, *2*, 1804-1820.
19. Worobey, M.; Holmes, E.C. Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* **1999**, *80*, 2535-2543.
20. Bruen, T.C.; Philippe, H.; Bryant, D. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **2006**, *172*, 2665-2681.
21. Heath, L.; van der Walt, E.; Varsani, A.; Martin, D.P. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J. Virol.* **2006**, *80*, 11827-11832.
22. Martin, D.P. Recombination detection and analysis using RDP3. *Methods Mol. Biol.* **2009**, *537*, 185-205.
23. Zhang, Y.; Wang, H-N.; Wang, T.; Fan, W-Q.; Zhang, A-Y.; Wei, K.; Tian, G-B.; Yang, X. Complete genome sequence and recombination analysis of infectious bronchitis virus attenuated vaccine strain H120. *Virus Genes* **2010**, *41*, 377-388.
24. Hein, R., Poultry Health Consultant, Georgetown, DE, USA, Personal Communication.
25. Woo, P.C.Y.; Lau, S.K.P.; Yip, C.C.Y.; Huang, Y.; Tsoi, H-W.; Yuen, K-Y. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. *J. Virol.* **2006**, *80*, 7136-7145.
26. Simmonds, P.; Welch, J. Frequency and dynamics of recombination within different species of human enteroviruses. *J. Virol.* **2006**, *80*, 483-493.
27. Simmonds, P.; Midgley, S. Recombination in the genesis and evolution of hepatitis B virus genotypes. *J. Virol.* **2005**, *79*, 15467-15476.

28. Armesto, M.; Cavanagh, D.; Britton, P. The replicase gene of avian coronavirus infectious bronchitis virus is a determinant of pathogenicity. *PLoS One* **2009**, *4*, e7384.
29. Hagemeijer, M.C.; Verheije, M.H.; Ulasli, M.; Shaltiel, I.A.; de Vries, L.A.; Reggiori, F.; Rottier, P.J.; de Haan, C.A. Dynamics of coronavirus replication-transcription complexes. *J. Virol.* **2010**, *84*, 2134-2149.
30. Lindner, H.A.; Fotouhi-Ardakani, N.; Lytvyn, V.; Lachance, P.; Sulea, T.; Menard, R. The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. *J. Virol.* **2005**, *79*, 15199-15208.
31. Zheng, D.; Chen, G.; Guo, B.; Cheng, G.; Tang, H. PLP2, a potent deubiquitinase from murine hepatitis virus, strongly inhibits cellular type I interferon production. *Cell Res.* **2008**, *18*, 1105-1113.
32. Gorbalenya, A.E.; Koonin, E.V.; Donchenko, A.P.; Blinov, V.M. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.* **1989**, *17*, 4847-4861.
33. Gorbalenya, A.E.; Koonin, E.V.; Lai, M.M. Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. *FEBS Lett.* **1991**, *288*, 201-205.
34. Eriksson, K.K.; Cervantes-Barragan, L.; Ludewig, B.; Thiel, V. Mouse hepatitis virus liver pathology is dependent on ADP-ribose-1"-phosphatase, a viral function conserved in the alpha-like supergroup. *J. Virol.* **2008**, *82*, 12325-12334.
35. Zhang, X.W.; Yap, Y.L.; Danchin, A. Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus. *Arch. Virol.* **2005**, *150*, 1-20.

36. Chen, X.; Zhang, Q.; Li, J.; Cao, W.; Zhang, J.X.; Zhang, L.; Zhang, W.; Shao, Z.J.; Yan, Y. Analysis of recombination and natural selection in human enterovirus 71. *Virology* **2010**, *398*, 251-261.
37. Gelb, J.J.; Jackwood, M.W. Infectious Bronchitis. In: *A laboratory manual for the isolation, identification, and characterization of avian pathogens*, 5th ed.; Dufour-Zavala, L.; Swayne, D.E.; Glisson, J.R.; Pearson, J.E.; Reed, W.M.; Jackwood, M.W.; Woolcock, P., Eds.; American Association of Avian Pathologists: Kennett Square, PA, USA, **2008**, pp 146-149.
38. Huson, D.H.; Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **2006**, *23*, 254-267.
39. Bryant, D.; Moulton, V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **2004**, *21*, 255-265.
40. Martin, D.P.; Posada, D.; Crandall, K.A.; Williamson, C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res. Hum. Retroviruses* **2005**, *21*, 98-102.
41. Padidam, M.; Sawyer, S.; Fauquet, C.M. Possible emergence of new geminiviruses by frequent recombination. *Virology* **1999**, *265*, 218-225.
42. Smith, J.M. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **1992**, *34*, 126-129.
43. Posada, D.; Crandall, K.A. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13757-61372.
44. Gibbs, M.J.; Armstrong, J.S.; Gibbs, A.J. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **2000**, *16*, 573-582.
45. Boni, M.F.; Posada, D.; Feldman, M.W. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* **2007**, *176*, 1035-1047.

46. Posada, D. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol. Biol. Evol.* **2002**, *19*, 708-717.
47. Salminen, M.A.M. Detecting and characterising individual recombination events: practice. In; *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*, Lemey, P.; Salemi, M.; Vandamme, A.M., Eds. Cambridge University Press: Cambridge, UK, **2010**, p. 723.

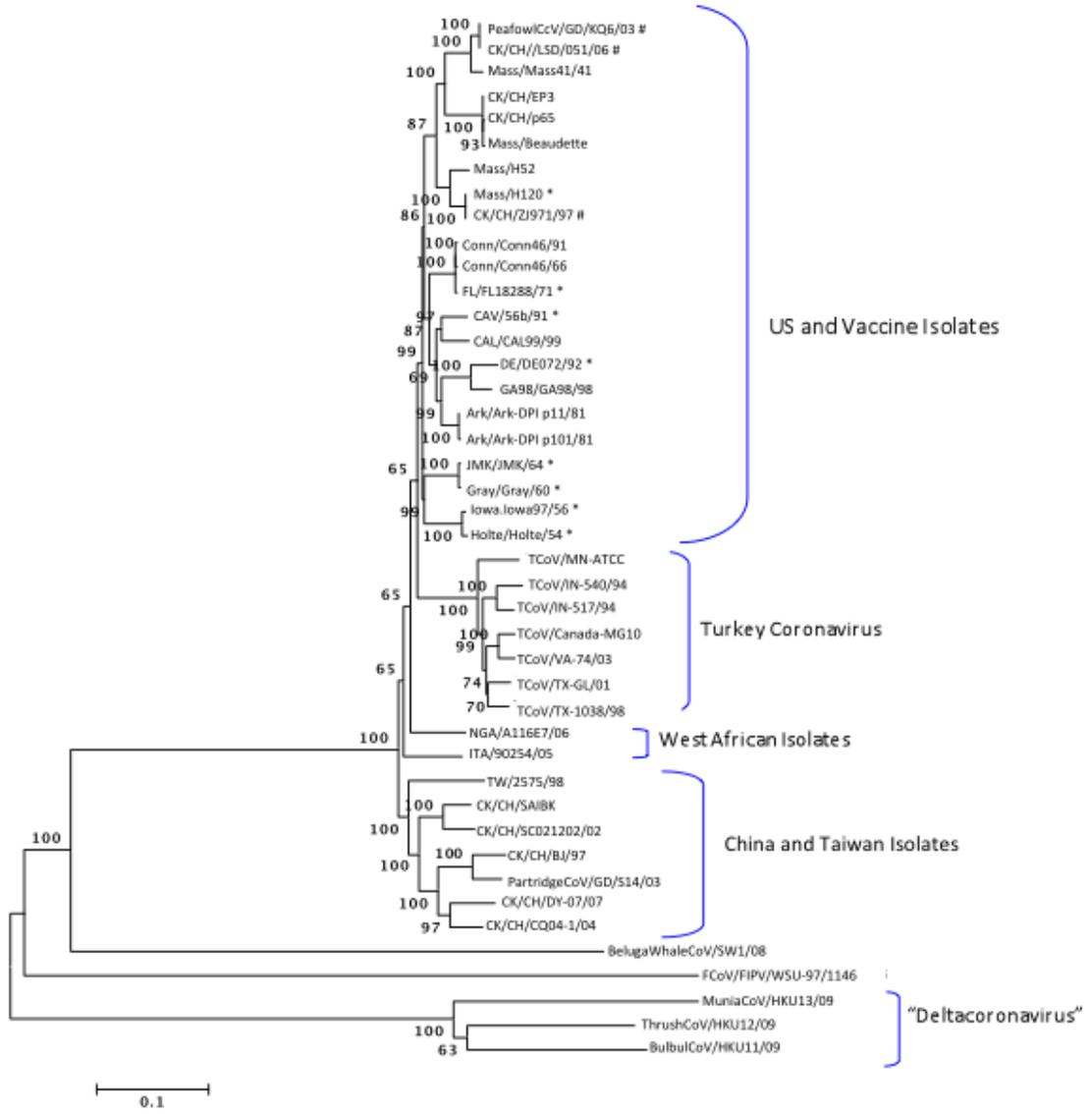
© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

**Table 3.1.** Genes and coding regions for 8 strains of avian infectious bronchitis virus examined in this study.

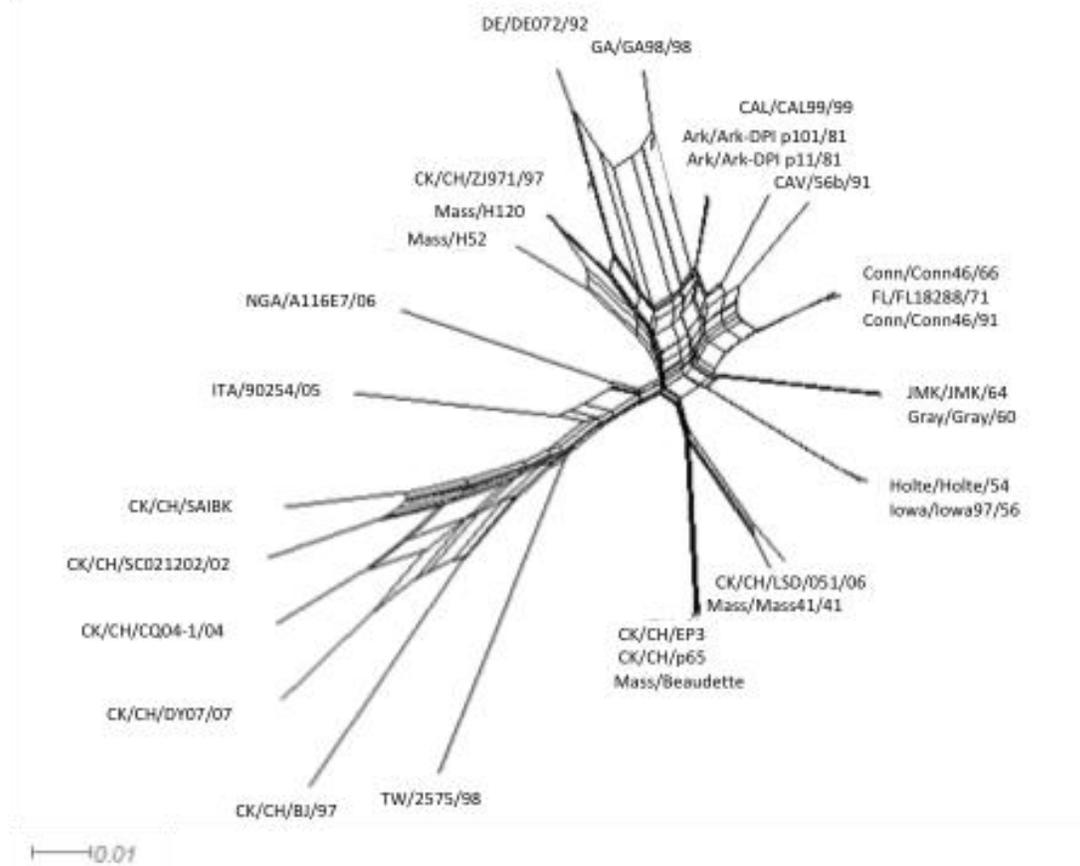
ORF <sup>a</sup>	CAV/CAV56b/91			DE/DE072/92			FL/FL18288/71			Gray/Gray/60			Mass/H120			Holte/Holte/54			Iowa/Iowa97/56			JMK/JMK/64		
	Location	nt <sup>b</sup>	aa <sup>c</sup>	Location	nt	aa	Location	nt	aa	Location	nt	aa	Location	nt	aa	Location	nt	aa	Location	nt	aa	Location	nt	aa
5'UTR	1–527	527	–	1–528	528	–	1–528	528	–	1–528	528	–	1–528	528	–	1–528	528	–	1–528	528	–	1–528	528	–
1a	528–12389	1,862	3953	529–12309	11781	3926	529–12387	11859	3952	529–12387	11859	3952	529–12330	11802	3933	529–12384	11856	3951	529–12390	11802	3933	529–12387	11859	3952
1ab	528–20422	19895	6631	529–20336	19808	6602	529–20420	19892	6630	529–20420	19892	6630	529–20363	19835	6611	529–20414	19886	6628	529–20423	19895	6631	529–20421	19893	6630
Spike	20373–	3501	1166	20287–	3453	1150	20371–	3468	1155	20371–	3504	1167	20314–	3489	1162	20365–	3507	1168	20374–	3507	1168	20371–	3507	1168
	23873			23739			23838			23874			23802			23871			23880			23877		
3a	23873–	174	57	23785–	174	57	23838–	164	54	23874–	174	57	23802–	174	57	23871–	174	57	23880–	174	57	23877–	174	57
	24046			23958			24011			24047			23975			24044			24053			24050		
3b	24046–	195	64	23958–	195	64	24011–	192	63	24047–	195	64	23975–	195	64	24044–	195	64	24053–	195	64	24050–	195	64
	24240			24152			24202			24241			24169			24238			24247			24244		
Envelope	24221–	282	93	24133–	330	109	24186–	303	100	24222–	324	107	24150–	330	109	24219–	324	107	24228–	324	107	24225–	324	107
	24502			24462			24488			24545			24479			24542			24551			24548		
Membrane	24651–	525	174	24434–	678	225	24488–	669	222	24523–	666	221	24451–	678	225	24520–	667	222	24529–	612	203	24526–	672	223
	25175			25111			25156			25188			25128			25188			25140			25197		
4b	25176–	285	94	25112–	285	94	25157–	285	94	25189–	240	79	25129–	243	80	25189–	285	94	25194–	285	94	25198–	132	43
	25460			25396			25441			25428			25371			25473			25478			25329		
4c	25381–	174	57	25317–	171	56	25362–	171	56	25340–	171	56	25334–	171	56	25394–	141	46	25399–	141	46	25374–	195	64
	25554			25487			25532			25510			25504			25534			25539			25568		
5a	25538–	198	65	25471–	198	65	25516–	198	65	25494–	198	65	25488–	198	65	25547–	198	65	25552–	198	65	25552–	198	65
	25735			25668			25713			25691			25685			25744			25749			25749		
5b	25732–	249	82	25665–	249	82	25710–	249	82	25688–	249	82	25682–	249	82	25741–	249	82	25746–	249	82	25746–	249	82
	25980			25913			25958			25936			25930			25989			25994			25994		
Nucleocapsid	25923–	1230	409	25856–	1230	409	25901–	1230	409	25879–	1233	410	25873–	1230	409	25932–	1230	409	25937–	1230	409	25937–	1230	409
	27152			27085			27130			27111			27102			27161			27166			27166		
6b	27161–	225	74	27094–	225	74	27139–	225	74	–	–	–	27126–	231	76	–	–	–	27175–	225	74	27175–	225	74
	27385			27318			27363						27356						27399			27399		
3'UTR	27386–	248	–	27319–	273	–	27364–	253	–	27112–	455	–	27357–	276	–	27162–	85	–	27340–	323	–	27400–	393	–
	27663			27591			27616			27568			27632			27246			27662			27793		

<sup>a</sup> ORF = open reading frame  
<sup>b</sup> nt = nucleotide  
<sup>c</sup> aa = amino acid

**Figure 3.1.** Neighbor-joining method used to infer evolutionary history using full genomic sequence data available for the gamma-coronaviruses. The percentage of replicate trees in which the associated taxa clustered together in a bootstrap test of 1000 replicates is shown next to the branches. The p-distance scale is presented at the bottom of the figure. An asterisk (\*) indicates a strain newly sequenced in this study. A number sign (#) indicates strains isolated in China that grouped with vaccine strains of IBV.



**Figure 3.2:** Neighbor-net for the avian gamma-coronavirus IBV. The networked relationships are shown to indicate the presence of reticulate events. Boxes imply the likelihood of recombination.



**Table 3.2.** Recombination breakpoints<sup>a</sup> genes and major and minor related sequences in other IBV strains.

Recombinant	Breakpoints		Genes <sup>b</sup>	Major sequence <sup>c</sup>	Minor sequence <sup>d</sup>	Detection Method
	Begin	End				
Ark/Ark-DPI-11/81	3,498	8,667	1ab (nsp 3, 4, and 5)	Conn/Conn46/66	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	4,312	10,590	1ab (nsp 3, 4, 5, and 6)	CK/CH/LSD/051	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	13,072	20,186	1ab (nsp 11/12, 13, 14, 15, and 16)	Unknown <sup>e</sup> (JMK/JMK/64)	CAL/CAL99/99	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	20,292	23,909	1ab (nsp16), Spike, 3a	Conn/Conn46/66	Unknown (Mass/Mass41)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	21,613	23,856*	Spike, 3a	CAL/CAL99/99	JMK/JMK/64	RDP, Maxchi, Chimaera, SiSscan, 3Seq
Ark/Ark-DPI-101/81	3,498	8,667	1ab (nsp 3, 4, and 5)	Conn/Conn46/66	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	4,312	10,590	1ab (nsp 3, 4, 5, and 6)	CK/CH/LSD/051	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	13,072	20,186	1ab (nsp 11/12, 13, 14, 15, and 16)	Unknown (JMK/JMK/64)	CAL/CAL99/99	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	20,292	23,909	1ab (nsp16), Spike, 3a	Conn/Conn46/66	Unknown (Mass/Mass41)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	21,613	23,856*	Spike, 3a	CAL/CAL99/99	JMK/JMK/64	RDP, Maxchi, Chimaera, SiSscan, 3Seq

**Table3.2. Cont.**

CAL/CAL99/99	0*	4,368*	5'UTR,1ab (nsp 2 and 3)	Ark/Ark-DPI/81	Unknown (DE/DE072/92)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	2,382	4,255*	1ab (nsp2,nsp3)	DE/DE072/92	Conn/Conn46/66	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	4,312	10,590	1ab (nsp 3, 4, 5, and 6)	CK/CH/LSD/051	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	8,104	10,649*	1ab (nsp 4, 5, and 6)	DE/DE072/92	Conn/Conn46/66	RDP, Maxchi, Chimaera, SiSscan, 3Seq
	24,587*	25,773	Envelope, Membrane, 4b, 4c, 5a, 5b	Unknown (GA/GA98/98)	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
CAV/56b/91	0*	1,512	1ab (nsp 2)	ITA/90254/2005	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, 3Seq
	0*	4,368*	5'UTR,1ab (nsp 2 and 3)	Ark/Ark-DPI/81	Unknown (DE/DE072/92)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	4,312	10,590	1ab (nsp 3, 4, 5, and 6)	CK/CH/LSD/051	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	4,392*	4,558	1ab (nsp3)	Ark/Ark-DPI/81	Conn/Conn46/91	GENECONV, Maxchi,

8,104	10,649*	1ab (nsp 4, 5, and 6)	DE/DE072/92	Conn/Conn46/66	Chimaera, 3Seq RDP, Maxchi, Chimaera, SiSscan, 3Seq
13,072	20,186	1ab (nsp 11/12, 13, 14, 15, and 16)	Unknown (JMK/JMK/64)	CAL/CAL99/99	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
20,292	23,909	1ab (nsp16), Spike, 3a	Conn/Conn46/66	Unknown (Mass/Mass41)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
24,556	25,748	Envelope, Membrane, 4b, 4c, 5a, 5b	Ark/Ark-DPI/81	Unknown (CAL/CAL99/99)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq

---

Table 3.2. Cont.

CK/CH/BJ/97	31*	5,600	5' UTR, 1ab (nsp 2 and 3)	CK/CH/SAIBK	Unknown (CK/CH/CQ041/04 )	RDP, GENECONV , Maxchi, Chimaera, SiSscan, 3Seq
CK/CH/CQ04 -1/04	60*	4,711	5' UTR, 1ab (nsp 2 and 3)	CK/CH/SC021202/0 2	CK/CH/DY-07/07	RDP, GENECONV , Maxchi, Chimaera, SiSscan, 3Seq
	8,751	9,018	1 ab (nsp 5)	CK/CH/SC021202/0 2	CK/CH/DY-07/07	RDP, GENECONV , Maxchi, Chimaera
	9,626	18,737	1ab (nsp 5, 6, 7, 8, 9, 10, 11/12, 13, 14, 15)	CK/CH/SAIBK	CK/CH/DY-07/07	RDP, GENECONV , Maxchi, Chimaera, SiSscan, 3Seq
	18,738 *	20,350	1ab (nsp 15 and 16)	CK/CH/SAIBK	ITA/90254/2005	RDP, GENECONV , Maxchi, Chimaera
	20,160	21,138	1ab (nsp 16), Spike	JMK/JMK/64	CK/CH/BJ/97	RDP, GENECONV , Maxchi, Chimaera, SiSscan
	27,120	27,354	Nucleocapsid , 6b	JMK/JMK/64	CK/CH/DY-07/07	GENECONV , Maxchi, Chimaera, SiSscan
CK/CH/DY- 07/07	1,170	5,017	1ab (nsp 2 and 3)	DE/DE072/92	CK/CH/SAIBK	RDP, GENECONV , Maxchi, Chimaera, SiSscan, 3Seq
	22,216	23,96	Spike, 3a	CK/CH/BJ/97	CK/CH/CQ04-1/04	RDP,

---

3

GENECONV  
, Maxchi,  
Chimaera,  
SiSscan, 3Seq

---

**Table 3.2. Cont.**

	25,455	25,662	4c, 5a	CK/CH/BJ/97	CK/CH/CQ04-1/04	RDP, GENECONV, Maxchi, Chimaera, SiSscan
CK/CH/LSD/051/06	306	3,628*	5'UTR, 1ab (nsp 2 and 3)	Mass/Mass41	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	1,453	2,743	1ab (nsp 2 and 3)	Mass/H52	Mass/Mass41/41	GENECONV, Maxchi, Chimaera, 3Seq
	13,668	14,734	1ab, (nsp 11/12)	Mass/Mass41/41	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	15,447	15,821	1ab (nsp 13)	Mass/Mass41/41	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan
	20,203	24,772	1ab (nsp 16), Spike, 3a, 3b, Envelope, Membrane	NGA/A116E7/06	Mass/Mass41	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	25,063	25,776	Membrane, 4b, 4c, 5a, 5b	Unknown (Mass/Mass41/41)	Mass/H120	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	25,774*	26,341	5b, Nucleocapsid	Mass/Mass41/41	Mass/H120	RDP, GENECONV, SiSscan, 3Seq

**Table 3.2. Cont.**

CK/CH/SAIBK	7,241	9,126	1ab (nsp 3, 4,5)	CK/CH/SC0212/02	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	20,160	21,138	1ab (nsp 16), Spike	JMK/JMK/64	CK/CH/BJ/97	RDP, GENECONV, Maxchi, Chimaera, SiSscan
CK/CH/SC021202/02	13,342	14,784	1ab (nsp 11/12)	CK/CH/SAIBK	CK/CH/DY-07/07	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	20,160	21,138	1ab (nsp 16), Spike	JMK/JMK/64	CK/CH/BJ/97	RDP, GENECONV, Maxchi, Chimaera, SiSscan
	27,120	27,354	Nucleocapsid, 6b	JMK/JMK/64	CK/CH/DY-07/07	GENECONV, Maxchi, Chimaera, SiSscan
CK/CH/ZJ971/97	0*	11,115	5'UTR, 1ab (nsp 2, 3, 4, 5, 6, 7, and 8)	NGA/A116E7/06	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan
	306	3,628*	5'UTR, 1ab (nsp 2 and 3)	Mass/Mass41	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	4,312	10,590	1ab (nsp 3, 4, 5, and 6)	CK/CH/LSD/051	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera,

---

20,203	24,772	1ab (nsp 16), Spike, 3a, 3b, Envelope, Membrane	NGA/A116E7/06	Mass/Mass41	SiSscan, 3Seq RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
--------	--------	--	---------------	-------------	---

---

Table 3.2. *Cont.*

	26,286	27,027	Nucleocapsid, 6b, 3' UTR	Iowa/Iowa97/56	CAL/CAL99/99	RDP, GENECONV, Maxchi, Chimaera, 3Seq
	27,094	27,244	Nucleocapsid, 6b	Iowa/Iowa97/56	Unknown (TW/2575/98)	RDP, GENECONV, Maxchi, Chimaera, SiSscan
Conn/Conn46/66	0*	1,512	1ab (nsp 2)	ITA/90254/2005	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, 3Seq
	0*	4,368*	5'UTR,1ab (nsp 2 and 3)	Ark/Ark-DPI/81	Unknown (DE/DE072/92)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	13,072	20,186	1ab (nsp 11/12, 13, 14, 15, and 16)	Unknown (JMK/JMK/64)	CAL/CAL99/99	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	20,361	21,981	Spike	CAL/CAL99/99	Mass/Mass41	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
Conn/Conn46/91	0*	1,512	1ab (nsp 2)	ITA/90254/2005	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, 3Seq
	0*	4,368*	5'UTR,1ab (nsp 2 and 3)	Ark/Ark-DPI/81	Unknown (DE/DE072/92)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	13,072	20,186	1ab (nsp 11/12,	Unknown	CAL/CAL99/99	RDP,

---

			13, 14, 15, and 16)	(JMK/JMK/64)		GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	20,361	21,981	Spike	CAL/CAL99/99	Mass/Mass41	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
DE/DE072/92	0*	11,115	5'UTR, 1ab (nsp 2, 3, 4, 5, 6, 7, and 8)	NGA/A116E7/06	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan
	18,776	19,911*	1ab (nsp 15 and 16)	Mass/H120	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq

---

**Table 3. 2. Cont.**

19,934	24,431	1ab (nsp16), Spike, 3a, 3b, Envelope	Mass/H120	Unknown (Mass/Mass41)	RDP, GENECOV, Maxchi, Chimaera, SiSscan, 3Seq
20,203	24,772	1ab (nsp 16), Spike, 3a, 3b, Envelope, Membrane	NGA/A116E7/06	Mass/Mass41	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
23,504	24,431*	Spike, 3a, 3b, Envelope	CK/CH/CQ04-1/04	CALCAL99/99	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
25,575	27,482*	5a, 5b, Nucleocapsid, 6b, 3' UTR	CK/CH/ZJ971/97	JMK/JMK/64	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
FL/FL18288/71	0*	1ab (nsp 2)	ITA/90254/2005	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, 3Seq
	0*	5'UTR,1ab (nsp 2 and 3)	Ark/Ark-DPI/81	Unknown (DE/DE072/92)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
13,072	20,186	1ab (nsp 11/12, 13, 14, 15, and 16)	Unknown (JMK/JMK/64)	CAL/CAL99/99	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
20,361	21,981	Spike	CAL/CAL99/99	Mass/Mass41	RDP, GENECONV, Maxchi, Chimaera,



**Table 3.2. Cont.**

GA98/0470/98	0*	4,368*	5'UTR,1ab (nsp 2 and 3)	Ark/Ark-DPI/81	Unknown (DE/DE072/92)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	2,382	4,255*	1ab, (nsp2 and 3)	DE/DE072/92	Conn/Conn46/66	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	3,498	8,667	1ab (nsp 3, 4, and 5)	Conn/Conn46/66	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	9,569	9,770	1ab (nsp 5)	Gray/Gray/60	Unknown (NGA/A116E7/06)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	13,072	20,186	1ab (nsp 11/12, 13, 14, 15, and 16)	Unknown (JMK/JMK/64)	CAL/CAL99/99	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	23,504	24,431*	Spike, 3a, 3b, Envelope	CK/CH/CQ04-1/04	CALCAL99/99	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	24,500*	25,438	Membrane, 4b	Conn/Conn46/66	Mass/Mass41/41	RDP, GENECONV, Chimaera, SiSscan, 3Seq
Gray/Gray/60	0*	4,368*	5'UTR,1ab (nsp 2 and 3)	Ark/Ark-DPI/81	Unknown (DE/DE072/92)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq

8,488	12,055	1ab (nsp 4, 5, 6, 7, 8, 9, and 10)	Unknown (CK/CH/LSD/051/06)	Conn/Conn46/91	RDP, GENECONV, Maxchi, Chimaera
13,070*	14,216	1ab (nsp 11/12)	Unknown (CAV/56b/91)	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq

Table 3.2. *Cont.*

24,131	27,145	3b, Envelope, Membrane, 4b, 4c, 5a, 5b, Nucleocapsid, 3' UTR	Ark/Ark-DPI/81	Unknown (Conn/Conn46/91)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq	
Holte/Holte/54	0*	4,368*	5'UTR, 1ab (nsp 2 and 3)	Ark/Ark-DPI/81	Unknown (DE/DE072/92)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
Iowa/Iowa97/56	0*	4,368	5'UTR, 1ab (nsp 2 and 3)	Ark/Ark-DPI/81	Unknown (DE/DE072/92)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	4,368	5,144	1ab (nsp 3)	Holte/Holte/54	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
ITA/90254/05	16,367	25,699	1ab (nsp 13, 14, 15, 16) Spike, 3a, 3b, Envelope, Membrane, 4b, 4c, 5a	GA98/0470/98	CK/CH/BJ/97	RDP, GENCOV, Maxchi, SiSscan
	22,216	23,963	Spike, 3a	CK/CH/BJ/97	CK/CH/CQ04-1/04	RDP, GENECONV, Maxchi,

---

	24,423	25,632*	Envelope, Membrane, 4b, 4c, 5a	CK/CH/DY- 07/07	NGA/A116E7/06	Chimaera, SiSscan, 3Seq RDP, GENECONV, Maxchi, Chimaera, 3Seq
JMK/JMK/64	0*	1,512	1ab (nsp 2)	ITA/90254/2005	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, 3Seq

---

**Table 3. 2. Cont.**

	0*	4,368*	5'UTR,1ab (nsp 2 and 3)	Ark/Ark-DPI/81	Unknown (DE/DE072/92)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	8,488	12,055	1ab (nsp 4, 5, 6, 7, 8, 9, and 10)	Unknown (CK/CH/LSD/051/06)	Conn/Conn46/91	RDP, GENECONV, Maxchi, Chimaera
	13,070*	14,216	1ab (nsp 11/12)	Unknown (CAV/56b/91)	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	24,131	27,145	3b, Envelope, Membrane, 4b, 4c, 5a, 5b, Nucleocapsid, 3' UTR	Ark/Ark-DPI/81	Unknown (Conn/Conn46/91)	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
Mass/H52	306	3,628*	5'UTR, 1ab (nsp 2 and 3)	Mass/Mass41	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	4,312	10,590	1ab (nsp 3, 4, 5, and 6)	CK/CH/LSD/051	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	19,925	20,168*	1ab (nsp 16)	Mass/Mass41/41	Mass/H120	GENECONV, Maxchi, Chimaera, SiSscan
	20,203	24,772	1ab (nsp 16), Spike, 3a, 3b, Envelope, Membrane	NGA/A116E7/06	Mass/Mass41/41	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	25,063	25,776	Membrane, 4b, 4c, 5a, 5b	Unknown (Mass/Mass41/41)	Mass/H120	RDP, GENECONV,

---

26,286	27,027	Nucleocapsid, 6b, 3' UTR	Iowa/Iowa97/56	CAL/CAL99/99	Maxchi, Chimaera, SiSscan, 3Seq RDP, GENECONV, Maxchi, Chimaera, 3Seq
26,372	27,526*	Nucleocapsid, 6b, 3' UTR	Unknown (DE/DE072/92)	Mass/H120	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq

---

**Table 3. 2. Cont.**

	27,094	27,244	Nucleocapsid, 6b	Iowa/Iowa97/56	Unknown (TW/2575/98)	RDP, GENECONV, Maxchi, Chimaera, SiSscan
Mass/H120	0*	11,115	5'UTR, 1ab (nsp 2, 3, 4, 5, 6, 7, and 8)	NGA/A116E7/06	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan
	306	3,628*	5'UTR, 1ab (nsp 2 and 3)	Mass/Mass41	Ark/Ark-DPI/81	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	4,312	10,590	1ab (nsp 3, 4, 5, and 6)	CK/CH/LSD/051	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	20,203	24,772	1ab (nsp 16), Spike, 3a, 3b, Envelope, Membrane	NGA/A116E7/06	Mass/Mass41	RDP, GENECONV, Maxchi, Chimaera, SiSscan, 3Seq
	26,286	27,027	Nucleocapsid, 6b, 3' UTR	Iowa/Iowa97/56	CAL/CAL99/99	RDP, GENECONV, Maxchi, Chimaera, 3Seq
	27,094	27,244	Nucleocapsid, 6b	Iowa/Iowa97/56	Unknown (TW/2575/98)	RDP, GENECONV, Maxchi, Chimaera, SiSscan
NGA/A116E7/06	7,035	8,271	1ab (nsp 3 and 4)	Holte/Holte/54	DE/DE072/92	RDP, GENECONV, Maxchi, Chimaera, 3Seq

---

TW/2575/98	20,160	21,138	1ab (nsp 16), Spike	JMK/JMK/64	CK/CH/BJ/97	RDP, GENECONV, Maxchi, Chimaera, SiSscan
	27,120	27,354	Nucleocapsid, 6b	JMK/JMK/64	CK/CH/DY- 07/07	GENECONV, Maxchi, Chimaera, SiSscan

---

\*=The actual breakpoint position is undetermined. Most likely it was overprinted by a subsequent recombination event.

<sup>a</sup>Only transferred gene fragments with statistical support of  $> 1 \times 10^{-12}$  (50 of 135 total unique fragments) are included in the table.

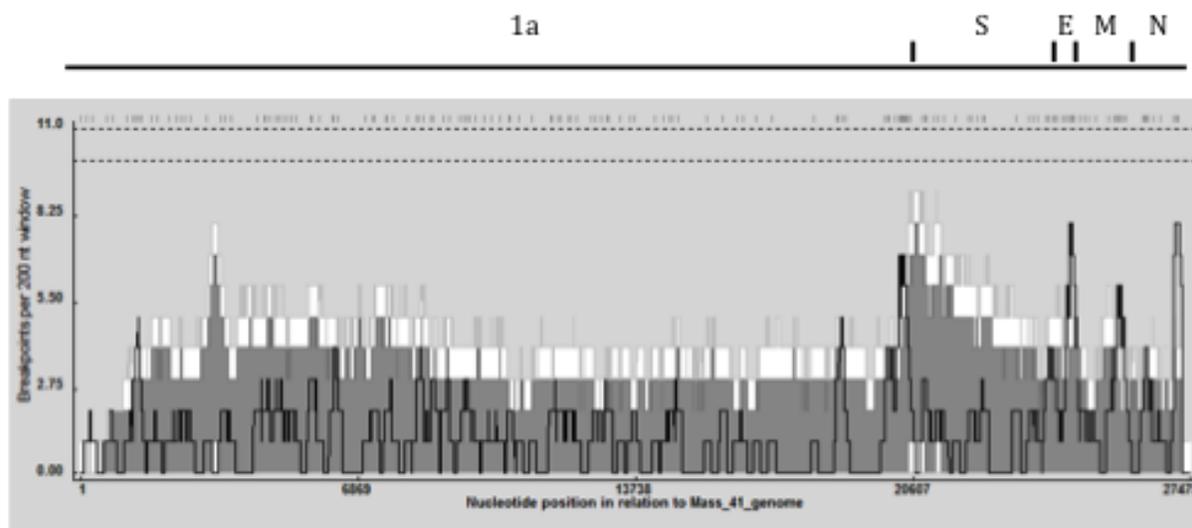
<sup>b</sup>Genes indicate the coding sequences contained within the fragment introduced by recombination.

<sup>c</sup>Major Sequence = Contains sequence most closely related to a large portion of the transferred fragment in the recombinant.

<sup>d</sup>Minor Sequence = Contains sequence closely related to some of the transferred fragment in the recombinant.

<sup>e</sup>Unknown = only one parent and a recombinant need be in the alignment for a transferred fragment to be detectable. The sequence listed in parentheses was used to infer the existence of a missing parental sequence.

**Figure 3.3:** Recombination breakpoint distribution plot generated for IBV using RDP4 showing the detectable recombination breakpoints. The plot was constructed using a 200 bp window moved 1 nucleotide at a time along the length of the genome. Recombination breakpoint positions are shown as hash marks at the top of the figure. The dashed lines under the breakpoint positions represent 99% (upper) and 95% (lower) confidence thresholds for globally significant breakpoint clusters (defined as windows with more breakpoint positions than the maximum found in >95% of the permuted plots). The dark gray and white areas are 95% confidence and 99% confidence intervals respectively for local breakpoint clusters (defined as windows with more breakpoint positions than the maximum found in >99% of the windows at that location in permuted plots), and the black line indicates the breakpoint count within the 200bp window.



**Table 3.3** Number of transferred fragments associated with individual areas of the genome for all of the strains examined.

Genomic Region	Number of fragments <sup>a</sup>	% of Total
5'UTR <sup>b</sup>	8	4.2
nsp <sup>c</sup> 2	20	10.5
nsp 3	33	17.3
nsp 4	17	8.9
nsp 5	15	7.9
nsp6	10	8.3
nsp 7	6	3.2
nsp 8	6	3.2
nsp9	4	2.1
nsp 10	4	2.1
nsp 11 /12	13	6.8
nsp 13	12	6.3
nsp 14	10	5.3
nsp 15	10	5.3
nsp 16	19	10.0
Spike	30	15.8
3a	14	7.4
3b	13	6.8
Envelope	17	8.9
Membrane	17	8.9
4b	12	6.3
4c	12	6.3
5a	15	7.9
5b	11	5.8
Nucleocapsid	14	7.4
3' UTR	13	6.8

<sup>a</sup>Genomic areas may be fully or only partially located in the transferred fragments.

<sup>b</sup>UTR= untranslated region

<sup>c</sup>nsp= nonstructural protein

**Table 3.4** Viruses sequenced in this study.

Strain	Serotype	Origin	Source
CAV/CAV56b/91	CAV	California, USA	P. Woolcock <sup>a</sup>
DE/DE072/92	DE	Delmarva, USA	J. Gelb Jr <sup>b</sup>
FL/FL18288/71	FL	Florida, USA	P. Villegas <sup>c</sup>
Gray/Gray/60	Gray	Delmarva, USA	J. Gelb Jr.
Holte/Holte/54	Holte	Wisconsin, USA	J. King <sup>d</sup>
Iowa/Iowa97/56	Iowa	Iowa, USA	J. King
JMK/JMK/64	JMK	Delmarva, USA	J. Gelb, Jr.
Mass/H120	Mass	The Netherlands	J. King

<sup>a</sup>University of California , Davis , CA

<sup>b</sup>University of Delaware, Newark, DE

<sup>c</sup>University of Georgia, Athens, GA

<sup>d</sup>Southeast Poultry Research Laboratory, USDA/ARS, Athens, GA

## CHAPTER 4

GENOMIC, PHYLOGENETIC AND RECOMBINATIONAL ANALYSIS OF ALPHA- AND  
BETACORONAVIRUSES<sup>1</sup>

<sup>1</sup>Thor, S.W., Hilt, D.A. Kissinger, J.C., Paterson, A. H., and Jackwood, M.W..  
To be submitted to Virus Genes.

**Abstract**

Recombination is a well documented occurrence in the family Coronaviridae and is thought to play a major role in the emergence and evolution of coronaviral genotypes and species. However, data on the frequency, extent, and locations of recombination for the alpha- and betacoronaviruses is limited. In this study, the full-length genomes of three (3) alphacoronavirus and 3 betacoronaviruses were sequenced and analyzed along with other full-length genomes available from GenBank for evidence of recombination. Evidence for recombination was found in almost every sequence analyzed and was spread throughout the genome. Recombination was greatest for an alphacoronavirus strain (FCoV/Black/1970) with 14 transferred fragments and least for both the alphacoronavirus strain BtCoV/HKU8/AFCD77, and betacoronavirus strains GiCoV/US/OH3/2003, HCoV/OC43/1967, and PHEV/VW/572 with only one transferred fragment each. Two genomic regions were identified with the most involvement in transferred fragments; the region that encodes nonstructural protein 3 and the gene for the structural spike glycoprotein. Interestingly, our analysis revealed no evidence of recombination between the alpha- and betacoronaviruses, which to our knowledge, is the first study to report these findings.

## Introduction

Severe acute respiratory syndrome (SARS) was the first viral pandemic outbreak of an infectious disease in the 21<sup>st</sup> century (Jones et al., 2008). This disease was first recognized in China in November of 2002 where it quickly spread to Hong Kong. The etiological agent was discovered to be a newly emerged coronavirus: the SARS coronavirus (SARS-CoV). Intermediate hosts for the virus were the Himalayan palm civet (*Paguma larvata*) and raccoon dog (*Nyctereutes pryonoides*) found in live animal markets in China. The viruses isolated from these animals were found to be almost identical to human isolates (Guan et al., 2003). Although the intermediate hosts for SARS-CoV were animals in close contact with humans, the reservoir for SARS-CoV-like viruses was found to be bats (Li et al., 2005b).

Coronaviruses are an enveloped, single-stranded, positive sense RNA virus with a genome of between 26 and 32 kb. The genome is comprised of a single positive sense strand of capped and polyadenylated RNA. At both the 5' and 3' ends of the genome are un-translated regions (UTR) that are important for RNA replication and transcription. The remainder of the genome is comprised of 7 to 10 open reading frames (ORFs). Gene 1 codes for a polyprotein that is the precursor of the viral polymerase (Pol). It is about 20 to 22 kb in length, making up about two thirds of the genome, and consists of two overlapping ORFs (1a and 1b) (Lai, 2001). The distal one third of the genome codes for the four structural proteins that are present in all coronaviruses; Spike glycoprotein (S), Envelope (E), Membrane (M), and Nucleocapsid (N). These structural genes are interspersed with several ORFs that encode various nonstructural proteins. In some beta-coronaviruses this includes the hemagglutinin-esterase (HE) gene. The number, nucleotide sequence, gene order and method of expression of these ORFs differ among

the coronavirus groups. However, the order of the gene coding the polymerase and the four structural genes is relatively conserved. This general order is 5'-Pol-S-E-M-N-3'.

The emergence of a zoonotic virus like SARS-CoV from an animal reservoir requires four events; interspecies contact, cross-species viral transmission, sustained viral transmission in the new host, and viral adaptation to the new host (Childs, 2004). Thus, an important first step to prevent the emergence of new zoonotic viruses is to identify and examine the viruses present in animal reservoirs that are in close contact with people. The genetic diversity of coronaviruses is due to adaptive evolution that is driven by both high mutation rates and genetic recombination (Holmes, 2009). The high mutation rates are attributed to the minimal proof-reading capabilities of the viral RdRp. Recombination is thought to be due to a template switching “copy-choice” mechanism unique to coronavirus RNA replication (Lai, 1992). The emergence of several alpha- and betacoronaviruses has been linked to recombination (Decaro et al., 2009a; Herrewegh et al., 1998; Woo et al., 2009b) but a comprehensive analysis of recombination among the strains of alpha- and betacoronaviruses along the entire genomic sequence has not been reported.

The CoVs examined in this study are from animals that can be found in close contact with humans and include alphacoronaviruses from dogs and cats and betacoronaviruses from mice and rats.

Canine enteric coronavirus (CCoV), an important enteropathogen of dogs, was first described in 1971 by Binn, *et al*, (Decaro and Buonavoglia, 2008). The virus was isolated from dogs with acute enteritis in a canine military unit in Germany. This strain was named 1-71 and was able to reproduce disease upon experimental administration to young dogs (Keenan, 1976). Since then, CCoV outbreaks have been seen worldwide although mainly in kennels and animal shelters. CCoV is characterized by high morbidity and low mortality. The virus is shed in high

titers in the feces of infected dogs, and is transmitted by the fecal oral route. Two CCoV genotypes have been identified: CCoV-I and CCoV-II (Decaro and Buonavoglia, 2008). These designations were based on their genetic relationship with FCoV type I and FCoV type II, respectively (Pratelli et al., 2003a; Pratelli et al., 2003b). While the two virus types share up to 96% genomic identity, they are highly divergent in the spike coding region (Lorusso et al., 2009; Ma et al., 2008; Pratelli et al., 2003a). The two genotypes are routinely isolated from the feces of dogs with diarrhea and often simultaneously infect the same dog (Pratelli et al., 2004).

The feline CoV, FCoV/FeCoV/WSU/79-1683/1979 strain causes a mild to inapparent enteritis in infected cats. It is a type II feline coronavirus due to its relationship in the spike gene to canine coronavirus (Herrewegh et al., 1998). It has not been shown to cause the deadly feline infectious peritonitis (FIP) infection although it is closely related to another type II FCoV, FCoV/FIPV/WSU/79-1146/1979, virus that does cause FIP.

Rat CoV sialodacryoadenitis virus (SDAV) strains RaCoV/681, and RaCoV/8190, cause sialodacryoadenitis and respiratory illness that is commonly found in laboratory rats. Rat SDAV is antigenically related to mouse MHV and is classified in the *murine coronavirus* genera. The murine CoV, MHV/S/3239-17 strain is highly infectious in susceptible mouse colonies and despite high mortality, the clinical signs in infected mice are often non-specific or are suggestive of encephalitis (Hirano et al., 1981). It is classified as a murine coronavirus and related antigenically related to the other MHV strains.

Because recombination among related CoV strains can lead to the emergence of new viruses and viral diseases, we examined the genome of cat, dog, mouse and rat CoVs that have not previously been sequenced for recombination using several bioinformatics programs. To our

knowledge, this is the first study to report on the extent of recombination in the full-length genomes of alpha- and betacoronavirus.

## **Materials and Methods**

### *Viruses and Viral RNA Extraction*

All of the viruses sequenced in this study (Table 4.1) were obtained from the American Type Culture Collection (ATCC; Manassas, Virginia). The viral RNA was extracted from 200µl of the culture sample using the High Pure RNA Isolation Kit (Roche Applied Science, Mannheim, Germany) following the manufacturer's instructions.

### *RT-PCR Amplification and Sequencing*

The amplification reactions were carried out using strand displacement RT-PCR or one step RT-PCR. Strand displacement RT-PCR uses a random (at the 3' prime end) primer (AGCGGGGGTTGTCTGAATGTTTGANNNN) and an amplification primer (AGCGGGGGTTGTCTGAATGTTTGA). The RT-PCR reaction was carried out according to the manufacturer's protocol using the TaKaRa RNA LA PCR kit (Takara Bio. Inc., Otsu, Shiga, Japan). A DNA Engine Peltier thermocycler (Bio-Rad Laboratories Inc., Hercules, CA) was used for the RT reaction, which had the following steps; an RNA denaturing step at 65 C for 10 min followed by 30 C for 10 min, 42 C for 60 min, 99 C for 5 min, and 5 C for 5 min. The PCR reaction was run on the same machine as the RT step and included a one-time initial denaturation step of 94 C for 2 min., followed by 30 cycles of 94 C for 30 s, 60 C for 30 s and 72 C for 3 min.

The PCR products were agarose gel purified using the QIAquick gel extraction kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. The PCR products were

cloned into the TOPOXL vector according to manufacturer's protocol using the TOPOXL cloning kit (Invitrogen, Carlsbad, CA) to prepare cDNA libraries for sequencing.

Plasmid DNA from the libraries of the cloned cDNA fragments for each virus was isolated using an alkaline lysis method modified for the 96-well format and incorporating both Hydra and Tomtek robots. Sequencing reactions were performed using the BigDye™ Terminator® Cycle Sequencing Kit Version 3.1 (Applied Biosystems, Foster City, CA) and MJ Research (Watertown, MA) thermocyclers. Sephadex filter plates were used to filter each reaction into Perkin-Elmer MicroAmp Optical 96-well plates. A 1/12-strength sequencing reaction on an ABI 3730 was used to sequence each clone from both the 5' and 3' ends.

Primers for one-step RT-PCR were specifically designed for each virus (Supplemental Table 4.1). Viral RNA was amplified using the Titan One Tube RT-PCR kit (Roche Diagnostics, Indianapolis, IN) following manufacturer's instructions. A DNA Engine Peltier Thermocycler (Bio-Rad Laboratories, Inc.,) was used for the RT-PCR reaction, which had the following steps: one cycle of 42 C for 60 min and 95 C for 5 min., followed by 10 cycles of 94 C for 30 s, 50 C for 30 s, and 68 C for 1min 30 s., and then 25 cycles of 94 C for 30 s, 50 C for 30 s, 68 C for 1 min and 30 s adding 5 s with each cycle.

The resulting PCR products were agarose gel purified using the QIAquick gel extraction kit (Quiagen, Valencia, CA) according to the manufacturer's protocol. The agarose gel purified cDNA was sequenced using ABI Prism BigDye Terminator, Version 3.0, Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA) following the manufacturer's protocol. The reactions were prepared for sequencing by centrifugation through either a Centri-Sep column (Applied Biosystems, Foster City, CA) or using the Edge system (EdgeBio,

Gaithersburg, MD) plate. The samples were sequenced at the Georgia Genomics Facility (University of Georgia, Athens, GA).

#### *Genome Assembly and Phylogenetic Analysis*

Chromatogram files and trace data were read and assembled using SeqMan Pro, and genome annotation was conducted with SeqBuilder (DNASTAR, Inc., Madison, WI). Each sequence was assembled using a representative genome as a backbone for genome assembly and each is listed with the GenBank accession number; CCoV/NTU336/F/2008 (#GQ477367), MHV-A59 (#AY700211), FCoV/FIPV/79-1146/1979 (#DQ010921), and RaCoV/Parker (#FJ938068).

Whole genome analyses were generated and phylogenetic trees constructed with the Neighbor-Joining method with 1000 bootstrap replicates as well as with Minimum Evolution, Maximum Parsimony and UPGMA methods (Tamura et al., 2007).

#### *GenBank Accession Numbers*

Virus genome sequences generated in this study were submitted to GenBank and assigned the following accession numbers: FCoV/WSU/79-1683, #JN634064, MHV/S/3239-17, #JQ173883, RaCoV/681, #JF792616, and RaCoV/8190/1970, #JN7926174. The following genomes are in the process of submission to GenBank: CCoV/1-71/1971 and CCoV/TN449.

GenBank accession numbers for full-length sequences used as reference in this study are listed in Table 4.2.

*Recombination analysis:*

Detection of a networked relationship between the analyzed viruses was carried out using Neighbor-net analysis and the pairwise homoplasy index (PHI) implemented in SplitsTree4 (Bruen et al., 2006; Bryant and Moulton, 2004; Huson and Bryant, 2006). The PHI test is a statistical test that determines the likelihood of recombination.

Recombination sites were determined using Recombination Detection Program 4 (RDP4) (Heath et al., 2006; Martin, 2009). The specific algorithms used were RDP (Martin et al., 2005), GENECONV (Padidam et al., 1999), Bootscan/Rescan (Martin, 2005), Maximum Chi Square (Maxchi) (Smith, 1992), Chimaera (Posada and Crandall, 2001), SiScan (Gibbs et al., 2000), and 3Seq (Boni et al., 2007). More than one method was used because evaluation of these recombination detection methods using both simulated and empirical data showed that the results from only a single method was not very reliable (Posada, 2002). Automasking was used for optimal recombination detection. The RDP analysis was run without a reference and a window size of 60, the Bootscan window size was increased to 500, Maxchi and Chimaera number of variable sites per window was increased to 120, and the window and step size for SiScan was increased to 500 and 20, respectively. The window sizes were increased from their default settings because coronaviruses have a high mutation rate which can mask recombination signals. Increasing the window size has been shown to increase the ratio of recombination signals relative to mutational “noise” (Salminen, 2010).

Unless otherwise stated, the default settings were used for each program

## Results and Discussion

### Sequence Analysis:

The full-length genomes of three alphacoronavirus and three betacoronavirus isolates were sequenced at 3x to 10x coverage, and the consensus sequences were assembled. The genome size (see the end of the 3' UTR in Table 1), organization of the genome, and the location and size of the open reading frames (ORFs) are listed in Table 4.3 a and b. The gene order is the same for all the coronaviruses examined herein; 5'UTR, -1a/ab-Spike-3a-3b-3c-Envelope-Membrane-Nucleocapsid-7a-7b-3'UTR.

### Phylogenetic Analysis:

Phylogenetic analysis was carried out on all the full-length genomes and spike encoding regions for both the alpha- and betacoronaviruses from GenBank listed in Table 2 and the genomes sequenced in this study. The sequences were aligned and phylogenetic trees constructed using the Neighbor-joining, Minimum Evolution, Maximum Parsimony and UPGMA programs in MEGA4 (Tamura et al., 2007). The trees all had similar topology and bootstrap support, and a representative tree is shown in Figures 4.1 (a and b) and 4.2 (a and b).

### *Alphacoronavirus*

The alphacoronavirus full-length genome tree topology included four clades: a porcine/canine clade, a feline type II clade, a feline type I clade and a clade comprised of viruses isolated from humans, bats, and pigs (Figure 4.1a). The bovine coronavirus (BCoV/Quebec) and avian coronavirus infectious bronchitis virus Mass41 strain (IBV/Mass/Mass/41) were included as outgroups. The Type I (FCoV/C1Je/2007 and FCoV/Black/1970) and Type II

(FCoV/NTU156/P/2007, FCoV/WSU/79-1683/1979, FCoV/DF2, and both FCoV/WSU/79-1146/p100/1979 and FCoV/WSU/79-1146) feline coronaviruses were divided into two distinct clades. This is not surprising as the Type II feline coronavirus spike has sequence similarity to the spike of CCoV (Herrewegh et al., 1998) although they did not align with CCoV strains indicating additional differences between the Type II feline strains and CCoV genomes. The TGEV isolates (TGEV/Miller/M6/1965 and TGEV/Purdue/1952) grouped in a clade with porcine respiratory coronavirus (PRCV/ISU-1/1990). The canine coronavirus genomes grouped in a clade showing close relationship to the TEGV clade. This supports the observation that recombinant CCoV strains related to TGEV have been observed circulating in dogs (Decaro et al., 2009a). The porcine, canine and feline clades have a 90% genomic sequence similarity between the clades (data not shown). The remainder of the viruses separated into a second major group with individual clades consisting of the alphacoronaviruses isolated from humans, bats, and pigs.

The phylogenetic tree for the spike glycoprotein (figure 4.1b) shows the type II feline coronaviruses grouped with the canine coronaviruses in a single multi-branched clade in close relationship to the TGEV/PRCV clade. As the origin of TGEV has been postulated to be a cross-species transmission of a type II CCoV (Lorusso et al., 2008), and type II CCoV arose via recombination between a type I CCoV and an unknown CoV (Lorusso et al., 2008), this close relationship is not unexpected. The placement of the sequences from the various bat coronaviruses in close relationship to not only human coronaviruses but also porcine and bovine supports the hypothesis that bats served as a reservoir for these viruses (Li et al., 2005b).

### *Betacoronavirus*

The tree topology for the betacoronaviruses showed 6 clades: a clade of the *murine coronavirus* species, a clade of *betacoronavirus I* species, separate clades for each of the bat coronaviruses and a *SARS-related coronavirus* species clade (Figure 4.2 a). The genomes for alphacoronavirus canine coronavirus (CCoV/NTU336/F/2008) and feline coronavirus (FCoV/NTU156/P/2007) as well as gammacoronavirus avian infectious bronchitis virus (IBV/Mass/Mass/41) were included as outgroups. The bat coronaviruses (RoBtCoV/HKU9-1, TyBtCoV/HKU4-1, and PiBtCoV/HKU5-1) fell into separate clades, which is not surprising as each represents a separate coronavirus species. In addition, the SARS-CoV strains (SARSCoV/CFB/SZ/94/03, SARS/GD01, and SARS/Urbani) are grouped together but separate from other betacoronaviruses. The topology of the phylogenetic tree for the spike glycoprotein (Figure 4.2 b) of the betacoronaviruses was similar to the phylogenetic tree for the full genomes.

### *Analysis of reticulate events*

Evolutionary history is traditionally represented using a strictly bifurcating phylogenetic tree, which implies that once two lineages are created they do not subsequently interact with each other. As coronaviruses have a high rate of mutation and recombination (Lai, 1992; Liao and Lai, 1992), this type of tree may not accurately represent the evolutionary histories of these viruses. When evolutionary events such as reassortment, horizontal gene transfer, or recombination occur, reticulations among the phylogenetic tree branches can be observed using the Neighbor-net analysis (Bruen et al., 2006). Evidence of networked relationships among the analyzed sequences was observed for both the alpha- and betacoronaviruses and is shown in Figure 4.3 (a, and b) and 4.4 (a and b). Because the reticulated lines only imply the possibility

of recombination, we conducted a pairwise homoplasy index (PHI) test, which showed a significant difference in the compatibility between closely linked sites ( $p < 0.0001$ ) supporting the occurrences of recombination among the alphacoronaviruses as well as among the betacoronaviruses. However, there was no statistical evidence ( $p > 1.0$ ) of recombination between alpha- and betacoronaviruses (data not shown) (Bruen et al., 2006).

### *Alphacoronavirus*

Results from the neighbor-net analysis for all alphacoronaviruses analyzed are shown in Figure 4.3a. The related sequences with a high degree of reticulate events, indicated by boxes created by crossed lines and including the analyzed FCoV, CCoV, TGEV, and PRCV strains, is circled and enlarged in figure 4.3b. The remaining alphacoronaviruses, including BtCoV/512/2005, BtCoV/1A/AFCD62, BtCoV/HKU8/GD/430/2006, PEDV/LZC/China, PEDV/CV777/1978, PEDV/SM98/1998, HCoV/NL63/2002, HCoV/229E/1967, and the outgroups BCoV/Quebec/1972 and IBV/Mass/Mass41/41, were separated from the encircled group by straight lines implying reticulate events between these viruses was less likely to occur. The reticulate events between the FCoV, CCoV, TGEV, and PRCV can be clearly seen in Figure 4.3b. The viruses separate into two groups of related viruses; an FCoV group and a CCoV-TGEV-PRCV group.

Although the network of reticulate events exists between all of the analyzed FCoV virus strains, there is a clear separation between the type I FCoV viruses (FCoV/C1Je/2007 and FCoV/Black/1970) and type II FCoV viruses (FCoV/NTU156/P/2007, FCoV/FIPV/WSU/79-1146/1979, FCoV/DF2, and FCoV/WSU/79/1683). The type II FCoV viruses are placed closer

to the CCoV-TGEV-PRCV group, which indicates more recent reticulate events among those viruses.

### *Betacoronavirus*

Results from the neighbor-net analysis for all betacoronaviruses analyzed (Figure 4.4a) shows that the Bat (TyBtCoV/HKU4-1, PiBtCoV/HKU5-1), SARS (SARSCoV/CFB/SZ/94/03, SARS/GD01, SARS/Urbani), avian (IBV/Mass/Mass41/41), feline (FCoV/NTU156/P/2007) and canine (CCoV/NTU336/F/2008) CoVs apparently had reticulate events in their distant past. Two groups showed evidence of more recent reticulate events, and include PHEV/VW/572, GiCoV/US/OH3/2003, BCoV/Quebec/1972, HuEnCoV/4408, HCoV/OC43/1967, EqCoV/NC99, in one group and RaCoV/8190/1970, RaCoV/681, RaCoV/Parker, MHV/S/3239-17, MHV/A59/1961, MHV/JHM and MHV/Penn/97-1 in the other group. These groups are circled and enlarged in figure 4.4b.

### Recombination analysis

To identify recombination sites in the full length alpha and betacoronaviruses, we used the Recombination Detection Program 4 (RDP4) (Heath et al., 2006; Martin, 2009). We analyzed the alpha- and betacoronaviruses together and the alphacoronaviruses and betacoronaviruses separately for evidence of recombination. We found no evidence of recombination between the alpha- and betacoronaviruses (data not shown). We found a total of 120 transferred fragments with 64 unique recombination sites among alphacoronaviruses and a total of 36 transferred fragments with 24 unique recombination sites among the betacoronaviruses. Each of these transferred fragments were detected by two or more methods and are listed in Table 4.4 a and b.

Although the algorithms implemented in RDP4 attempt to identify major and minor parent sequences contributing to each recombinant, the data represents sequences in other analyzed viruses that are most closely related to a majority of the transferred fragment (major sequence) or closely related to some of the transferred fragment (minor sequences) and doesn't imply origin or source of the transferred fragment. In many cases, the transferred fragment has undergone mutations and /or recombination making it difficult to identify all the endpoints for the major and minor sequences. In addition, some of the transferred fragments overlap suggesting that recombinations have occurred between recombinant viruses.

### *Alphacoronavirus*

Evidence for recombination was found for most of the analyzed alphacoronavirus genomes (Table 4.4). The virus with the highest number of recombination sites is a feline virus (FCoV/Black/1970) with 14 transferred fragment and the virus with the fewest number of recombination sites is a bat coronavirus (BtCoV/HKU8/AFCD77) with one transferred fragment. It is interesting that the only alphacoronavirus examined in this study without any evidence of recombination in its evolutionary past is a bat coronavirus (BtCoV/1A/AFCD62).

Analysis of the FCoV strains revealed 63 individual recombination sites with 19 instances of two or more strains sharing identical transferred fragments. The FCoV strains FCoV/Black/1970 and FCoV/C1Je/2007 had evidence for 7 identical transferred fragments. The FCoV/Black/1970 strain is a type I feline coronavirus that causes either asymptomatic disease or a mild enteritis in infected cats; whereas FCoV/C1Je/2007 is a type II feline coronavirus that causes FIP in infected cats. One of the seven transferred fragments is completely in the 1ab coding region; specifically in nsp 1 (nt #13,128 to 14,227), and five of the seven recombination

sites are in the spike gene (nt #22,211 to 22,344, nt #22,407 to 22,591, nt # 23,085 to 23,106, nt #23,123 to 23,217, and nt #23,337 to 24,135). The seventh fragment has its 5' breakpoint in the spike gene and the 3' breakpoint in the envelope gene. The specific genes included in this transferred fragment are spike, 3a, 3b, 3c, and envelope (nt #23,069 to 25,835). The FCoV strains FCoV/NTU156/P/2007 and FCoV/DF2 had evidence for 4 identical transferred fragments. Three of the fragments were located in separate regions of the genome; specifically the nsp 1 and 2 coding region of ORF 1ab (nt # 415 to 2,822), spike (nt #23,032 to 23,281) and nucleocapsid (nt #26,393 to 27,048). The fourth fragment has its 5' breakpoint in spike gene and the 3' breakpoint in the 3c ORF. The specific genes in this fragment include spike, 3a, 3b, and 3c (nt # 23,284 to 25,542). The feline coronaviruses FCoV/NTU156/P/2007 and FCoV/WSU/79-1683 share two identical transferred fragments. One fragment has its 5' breakpoint in the spike gene and the 3' breakpoint in the 3b ORF. The specific genes included in this fragment are spike, 3a and 3b (nt # 23,054 to 24,988). The second fragment has its 5' breakpoint in the 3c ORF and the 3' breakpoint in the membrane gene. The genes included in this fragment are 3c, envelope, and membrane (nt #25,522 to 26,298). Other viruses shared identical transferred fragments spread throughout the genome however these were individual recombination sites. The FCoV/DF2, FCoV/WSU/79-1683, FCoV/WSU/79-1146/1979, and FCoV/NTU156/P/2007 have an identical transferred fragment that has its 5' breakpoint in the 1ab ORF and 3' breakpoint in the spike gene. The genes specifically included in this fragment are the genes for nsp 11/12, nsp 13, nsp 14, nsp 15, and nsp 16 encoded in 1ab ORF and the spike gene (nt #13,074 to 23,022). The FCoV/Black/1970 and FCoV/NTU156/P/2007 strains have an identical transferred fragment located in the nsp 3 gene of ORF 1ab (nt #6,961 to 7,617 and nt #3,063 to 4,236, respectively). The FCoV/WSU/79-1146/1979 and FCoV/DF2 strains have an identical

transferred fragment located in the nsp 11/12 coding region of ORF 1ab (nt #13,167 to 13,351) and FCoV/WSU/79-1146/1979, FCoV/WSU/79-1683, and FCoV/DF2 have an identical transferred fragment that includes the nsp 11/12 and nsp 13 genes of ORF 1ab (nt # 14,760 to 15,939). An identical transferred fragment with its 5' breakpoint in the spike gene and the 3' breakpoint in the 3' UTR is observed in FCoV/Black/1970, FCoV/WSU/79-1146/1979, and FCoV/DF2. The specific genes involved in this fragment are spike, envelope, membrane, and nucleocapsid (nt #27,491 to 28,132). Natural dual infection of cats with FCoV is not uncommon (Addie et al., 1996) which may allow for the extensive recombination (63 out of 120 total recombination sites) we observed between these viral strains.

The CCoV origin of the type II FCoV spike glycoprotein resulting from a double cross-over event reported in the literature by Herrewegh, *et al*, and Motokawa, *et al*, is supported not only by the locations of the recombination sites in FCoV strains but also CCoV strains as the major sequence contributors of the feline transferred fragments in the spike gene (Herrewegh et al., 1998; Motokawa et al., 1995). The data published by Motokawa, *et al*, indicated that for FCoV type II a template switch took place between the S and the M genes (Motokawa et al., 1995) while Herrewegh, *et al*, also reported recombination sites in the 1ab ORF (Herrewegh et al., 1998). Our data support these findings as out of 63 individual transferred fragments for the FCoVs, 59 sites involve the 1ab ORF and or the spike, envelope, or membrane genes. The major sequence contributor to 20 of the 63 transferred fragments is a CCoV and 13 of those sites were located in spike.

Only one transferred fragment is shared by both feline and canine coronaviruses.

The fragment is located in the 1ab ORF with the 5' breakpoint located in the nsp 14 gene and the 3' breakpoint in the nsp 16 gene; including the nsp 15 gene (nt # 17,879 to 20,286), and is shared

by CCoV/NTU336/F/2008 and FCoV/NTU156/P/2007. Since CCoVs are able to use feline aminopeptidase N (fAPN) glycoprotein as a cellular receptor (Rossen et al., 2001) and, under experimental conditions, cats can be infected with CCoV (Barlough et al., 1984) recombination between feline and canine coronaviruses should not be unanticipated however our analysis revealed only one example.

Our analysis indicated 14 individual recombination sites in the CCoV with two identical transferred fragments shared. Both fragments are located in the nsp 3 coding region of ORF 1ab and are found in CCoV/1-71-1971 and CCoV/TN449 (nt #3,834 to 3,972) and CCoV/NTU336/F/2008 and CCoV/TN449 (nt #5,790 to 6,462). Both fragments had a CCoV strain as a major sequence contributor (CCoV/NTU/336/F/2008 and CCoV/1-71/1971, respectively). Dogs are commonly found to be naturally infected with two separate strains of CCoV (Decaro, 2010) and frequent recombinants are found to be circulating in populations of dogs (Decaro, 2010).

Our analysis revealed one transferred fragment that was shared among the 13 strains of alphacoronaviruses isolated from cats, dogs and pigs. These strains include FCoV/Black/1970, FCoV/C1Je/2007, FCoV/DF2, FCoV/NTU156/P/2007, FCoV/WSU/79-1146/1979, FCoV/WSU/79-1683, CCoV/1-71/1971, CCoV/NTU336/F/2008, CCoV/TN449, TGEV/Miller/M6/1965, TGEV/Purdue/1952, and PRCV/ISU-1/1990. Our analysis identifies the major sequence contributor to this transferred fragment as PEDV/CV777/1978. This implies a common evolutionary history for these viruses. Evidence of shared transferred fragments for TGEV, PRCV, and PEDV provides evidence for the close association of these viruses with both canine and feline coronaviruses. These viruses have <90% sequence identity (data not shown) and according to a proposal by the coronavirus study group of the International Committee on

the Taxonomy of viruses, TGEV, FCoV, and CCoV should not be considered separate viruses but considered host range variants of the same virus (de Groot, 2011 (in press)). Although the organism where this recombination took place is not known it is likely to have been a feline as the fAPN can serve as a receptor for not only FCoV, but also CCoV (Rossen et al., 2001), TGEV, and PRCV (Tresnan et al., 1996).

Analysis of the porcine coronaviruses revealed 23 individual recombination sites with three identical transferred fragments are shared between TGEV/Purdue/1952 and TGEV/Miller/M6/1965. The location of two the 5' breakpoint of these fragments is in ORF 7 and the 3' breakpoint is located in the 3'UTR (nt #28,220 to 28,544 and 28,222 to 28,287) with the major sequence contributor as CCoV/TN449 and PEDV/SM98/1998, respectively. The third shared fragment is located in Spike (nt #21,345 to 21,464) with the major sequence contributor as the FCoV/NTU156/P/2007 virus. An identical shared fragment between PRCV/ISU-1/1990 and TGEV/Miller/M6/1965 is located in the 3b gene (nt #24,276 to 27,833) and TGEV/Purdue/1952, TGEV/Miller/M6/1965 and PRCV/ISU-1/1990 share a transferred fragment with the 5'breakpoint located in the nsp 16 gene of ORF1ab and the 3' breakpoint located in the spike gene (nt #20,406 to 21,315); both have CCoV/TN449 as the major sequence contributor to the transferred fragments. A single identical transferred fragment is shared between PEDV/CV777/1978, PEDV/SM98/1998, and PEDV/LZC/China and is located in the spike gene (nt #21,365 to 21,483) with a bat coronavirus (BtCoV/512/2005) as the major sequence contributor. These events imply a common evolutionary relationship between the porcine, canine, and feline alphacoronaviruses as well as bat coronavirus sequence contributing to the evolution of PEDV.

Recombination was also observed in the human strains of alphacoronaviruses; HCoV/NL63/2002 and HCoV/229E/1967. These viruses share the same two transferred fragments located in the 1ab ORF; one in the nsp coding region (nt #2,967 to 3,089) and the other in nsp 8 (nt #12,214 to 12,604). Both of these fragments have a bat coronavirus as the major sequence contributor (BtCoV/512/2005 and BtCoV/HKU2/GD/430/2006, respectively) implying a common evolutionary history rather than a dual infection of either a human or a bat.

Evidence of 4 individual recombination sites was observed in the alphacoronavirus bat strains with the BtCoV/HKU2/GD/430/2006 and HCoV/NL63/2002 sharing an identical transferred fragment in the nsp 3 region of the 1ab ORF (nt #4,687-5,205) with a major sequence contributor of BtCoV/512/2005. Evidence of two identical transferred fragments between bat and porcine viruses was observed. The fragments were shared among PEDV/CV777/1978, PEDV/SM98/1998, PEDV/LZC/China, and BtCoV/512/2005. One fragment was located in the nsp 4 coding region of the 1ab ORF (nt #8,214 to 8,912) and the other in the spike gene (nt #25,158 to 25,429). Although there are many reports of the discovery of coronaviruses in bats (Dominguez, 2007; Drexler et al., 2011; Quan et al., 2010; Woo et al., 2006), to our knowledge this is the first report of recombination involving pig, human, and bat coronaviruses. Although the host that allowed this recombination to take place is unknown, evidence from studies of SARS-CoV and the bat angiotensin-1 converting enzyme 2 (ACE-2) cellular receptor point to small amino acid changes allowing the receptor to function for SARS-CoV (Yu et al., 2010). These same amino acid changes may also result in the alphacoronavirus bat ACE-2 receptor allowing the receptor to function for other alphacoronaviruses.

Our analysis did not show the BtCoV/1A/AFCD62 with any recombination in its evolutionary past, although there was evidence of this virus' involvement as a major sequence

contributor in the evolutionary past of other viruses; PEDV/CV777/4978, PEDV/LZC/China and PEDV/SM98/1998 in spike (nt # 25,158 to 25,429), and BtCoV/512/2005 in spike (nt #25,158 to 25,429). With increasing surveillance and sequencing of bat coronaviruses, further analysis may show this virus to have evidence of recombination in its evolutionary past.

### *Betacoronavirus*

Evidence for recombination was detected for most of the betacoronaviruses analyzed. The betacoronavirus with the greatest number of detected recombined fragments is the MHV/JHM with 7 transferred fragments. The viruses with the fewest recombined sequence fragments are GiCoV/US/OH3/2003 and PHEV/VW/572, each having evidence for only one transferred fragment. None of the SARS-CoV or bat coronavirus strains analyzed showed evidence for recombination in the evolutionary past of these viruses. This may be an artifact of the algorithms used to detect recombination as viruses that have < 70% sequence identity can create recombinational noise that may occlude the recombination sites (Martin and Rybicki, 2000).

Evidence of 16 recombination sites in the MHV strains was detected. The MHV/A59/1961 and MHV/S/3239-17/1963 strains share an identical transferred fragment in the 1ab ORF, specifically the nsp 5 coding region (nt #10,251 to 10,497). The major sequence contributor to this fragment is RaCoV/8190/1970. The MHV/A59/1961 also shares an identical transferred fragment with MHV/JHM with the 5' breakpoint in the HE gene and the 3' breakpoint in the spike gene (nt # 22,578 to nt #26,030) and the major sequence contributor to the fragment is MHV/S/3239-17/1963. The MHV/S/3239-17/1963 and MHV/JHM strains share a fragment with its 5' breakpoint located in the M gene and the 3' breakpoint in the nucleocapsid gene (nt

#29,732 to 29,969) with RaCoV/Parker as the major sequence contributor to this fragment. The MHV viruses likely share a common evolutionary history with the RaCoV.

Further evidence of the rat and mouse viruses sharing a common evolutionary history can be observed in the recombination discovered in the RaCoV. Three identical fragments are shared among RaCoV/Parker, RaCoV/681, and RaCoV/8190. They are located in the nsp 3 and 4 coding region of the 1ab ORF (nt #8,155 to 8,834), the spike and ns4 genes (nt #27,758 to 28,254) and the nucleocapsid gene (nt #30,790 to 30,887). The main sequence contributor for all three fragments was MHV/ S/3239-17/1963.

Our analysis revealed 5 unique recombination sites for the BCoV/Quebec/1972, EqCoV/NC99, and GiCoV/US/OH3/2003 with two of the five being identical transferred fragments shared by these strains. The fragments are located in the 1ab ORF but included different sections of the coding region. One fragment includes the nsp 1 and 2 genes (nt #284-1,992) and the other has its 5' breakpoint located in the nsp 7 coding region and the 3' breakpoint in the nsp 11/12 coding region (nt #12,187 to 14,865) which includes nsp 8, nsp 9, and nsp 10. The major contributing virus to each of these fragments is a human coronavirus; HCoV/OC43/1967. A study by Vijgen, et al, explored the genetic and antigenic relatedness among HCoV/OC43/1967, BCoV, and PHEV and concluded that although these viruses have different host specificities, they have diverged fairly recently, within the last 100 years (Vijgen et al., 2006). This close relatedness may explain the role of HCoV/OC43/1967 in the evolutionary past of BCoV/Quebec/1972; which has been shown to be both genetically and antigenically closely related to both the EqCoV/NC99 (Zhang et al., 2007) and Gi/US/OH3/2003 (Hasoksuz et al., 2007).

In our analysis, the human coronavirus HCoV/OC43/1967 did not show evidence of recombination in its evolutionary past. A molecular epidemiology study, conducted by Lau, *et al*, analyzed 29 strains of HCoV/OC43 for the presence of recombination and different genotypes. They found 4 distinct genotypes that arose from recombination between the 4 genotypes themselves (Lau et al., 2011). As we only included one genotype in our study, the evidence of no recombination is not surprising.

#### Recombination breakpoint analysis

To determine recombination hot and cold spots, a recombination breakpoint distribution plot was generated for both alpha- and betacoronaviruses (Figure 4.5 a, and b). These plots were created using RDP4 with a 200nt window and 1000 permutations (Heath et al., 2006). No global hot-spot regions were observed in the 95% to 99% confidence threshold (dotted lines at the top of the graph) for any of the virus groups examined. For each graph the detectable recombination breakpoints were distributed throughout the genomes, which are shown along the top of each plot (Figure 4.5 a, and b). Not all breakpoints located by RDP4 are displayed on the graphs due to either the major or the minor sequence contributor being unknown. However, the distributions were consistent with the genome regions identified as being associated with the greatest number of transferred fragments listed in Table 4.5 a, and b. A high number of recombination sites were observed in the nsp 3 (8 fragments) and spike (21 fragments) genes for the alphacoronaviruses, and the nsp 3 (6 fragments) and spike (8 fragments) genes for the betacoronaviruses. The recombination sites observed for both alpha- and betacoronaviruses are similar to a previous study conducted in our laboratory with the gammacoronaviruses where the regions with the most recombination sites are nsp 3 and spike (Thor et al., 2011).

Recombination in the 1ab ORF of the genome coding for nsps 1 through 16 may affect the pathogenicity of the virus (Armesto et al., 2009). The nsp 3 coding region of the genome encodes five domains; an acidic domain (Ac), the papain-like protease 1 (PLP1) domain, an ADP ribose 1'-phosphatase (ADRP) domain, the papain-like protease 2 (PLP2) domain, and the Y domain (Snijder et al., 2003b; Ziebuhr et al., 2001). Recombination in nsp 3 is likely to affect the ability to replicate either directly or indirectly. The Ac domain of nsp 3 has been shown to bind to the N protein which may be essential for the localization of the genomic RNA to the replicase complex in the early stages of replication (Hurst et al., 2010). Recombination in this domain could affect this localization either by enhancing the binding with the N protein thus causing the viral RNA to replicate more efficiently and /or impair the binding and thus replication complex localization resulting in either a non-replicative or replication impaired virus. Both PLP1 and PLP2 domains are involved in the cleavage of nsps 1, 2, and 3 (Anand et al., 2003; Thiel et al., 2003) and as antagonists of the host cell interferon (IFN) response (Devaraj et al., 2007; Frieman et al., 2009). Recombination in this area could affect the survival of a virus by either allowing it to escape the innate immune system by more efficient antagonism of IFN or, by allowing the host IFN response to progress unabated. Changes in nsp 3 have been associated with the attenuation of a gammacoronavirus, avian infectious bronchitis virus (Ammayappan et al., 2008; Phillips et al., 2011).

The spike glycoprotein is located on the surface of the virus and plays a role in not only viral attachment, but also membrane fusion and entry into the host cell. It displays conformationally dependent epitopes that induce virus-neutralizing and serotype specific antibodies. Recombination in the spike gene of coronaviruses has been well documented and has lead to the emergence of new viral strains, serotypes, and pathotypes (Decaro et al., 2009a;

Decaro, 2010; Herrewegh et al., 1998; Lee and Jackwood, 2001; Lee and Jackwood, 2000; Lim et al., 2011; Wang and Lu, 2009). The spike gene of type II feline coronaviruses have been shown to have originated from a double recombination event with a canine coronavirus (Herrewegh et al., 1998). Our findings are consistent with a previous study that identified the spike gene as the hot site for recombination for CCoV (Wang and Lu, 2009) as one half (7/14) of the recombination events identified for canine coronaviruses were in the spike gene.

One of the analyzed FCoV strains, FCoV/NTU156/P/2007 was the major sequence contributor to recombination in the Spike gene of both TGEV/Miller/M6/1965 and TGEV/Purdue/1952. Whereas recombination involving CCoV and TGEV has been reported (Decaro et al., 2009a; Decaro et al., 2009b), this is the first report of recombination involving FCoV in the evolutionary past of TGEV. The common use of the fAPN receptor may have facilitated recombination between these viruses (Rossen et al., 1995; Tresnan et al., 1996).

There was no evidence of recombination between alpha- and betacoronavirus genera found in the analysis carried out in this study. The barrier to recombinations occurring between the alpha- and betacoronaviruses may lie in the ancient ecology of the viruses themselves. In order for recombination to occur, two viruses must not only infect the same host but ultimately the same cell. Bats have been identified as the animal reservoir for SARS-CoV-like viruses (Li et al., 2005b) and subsequent studies have isolated both alpha- and betacoronaviruses from different species of bats found worldwide (Dominguez, 2007; Quan et al., 2010; Tang et al., 2006). A study by Lau, *et al*, found bats from Hong Kong and China that were co-infected with SARS-r-Rh-BatCov, a betacoronavirus, and RhBatCoV-HKU2, an alphacoronavirus (Lau et al., 2010a). Although the study did not provide evidence for recombination between these two viruses, their co-infection of the same bat would allow ample opportunity for recombination and

the emergence of new viruses capable of interspecies transmission. This implies the viruses have sufficiently low identity that would not produce a viable virus if recombination occurred. The discovery of 3 amino acid changes in the bat ACE 2 receptor allowing SARS-CoV to utilize the receptor to infect cells (Xu et al., 2009) provides a clue as to the conditions needed to allow alphacoronaviruses and betacoronaviruses to recombine. The only example of an alphacoronavirus possibly recombining with a betacoronavirus was the discovery of the mosaic nature of the SARS-CoV genome (Zhang et al., 2005). Although this type of recombination is possible, it appears to be rare.

## **Conclusion**

To our knowledge, ours is the first study to report the extent of recombination occurring among both alpha- and betacoronaviruses. Almost every sequence analyzed for both alpha- and betacoronaviruses was recognized as a potential recipient of horizontally acquired sequences at some point in its viral evolutionary past. The nsp 3 region and the spike gene were identified as genomic regions for both coronavirus genus with the greatest number of recombination sites for both the alpha- and betacoronaviruses. The evolutionary histories of the FCoV, CCoV, and TGEV which provides support for these viruses as host species variants of the same virus (de Groot, 2011 (in press)). Although our analysis did not reveal any evidence for recombination between the alpha- and betacoronaviruses, it does suggest the evolution of these viruses is intimately linked to reticulate evolutionary events such as recombination.

The characterization, distribution, and frequency of recombination events are important in the development of more effective control measures for coronaviruses. This study provides evidence for recombination playing a major role in the evolutionary past of coronaviruses:

specifically alpha- and betacoronaviruses. If we know which viruses are most likely to recombine and where the viruses are most likely to recombine, we should be able to predict the changes that are likely to occur in the viral genome; potentially affecting pathology and tropism.

## References

1. Jones, K.E., et al., *Global trends in emerging infectious diseases*. Nature, 2008. **451**(7181): p. 990-993.
2. Guan, Y., et al., *Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China*. Science, 2003. **302**(5643): p. 276-8.
3. Li, W., et al., *Bats are natural reservoirs of SARS-like coronaviruses*. Science, 2005. **310**(5748): p. 676 - 679.
4. Lai, M.M.C., K.V. Holmes, *Coronaviridae: the viruses and their replication*. 4th ed. Fields Virology, ed. D.M. Knipe, P.M. Howley, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman, et.al. 2001, Philadelphia, PA: Lippincott, Williams, and Wilkins.
5. Childs, J.E., *Zoonotic viruses of wildlife: hither from yon*. Arch Virol Suppl, 2004(18): p. 1-11.
6. Holmes, E.C., *The Evolution and Emergence of RNA viruses*. Oxford Series in Ecology and Evolution, ed. P.H. Harvey, and May, Robert M. 2009, New York: Oxford University Press. 254.
7. Lai, M.M.C., *RNA Recombination in Animal and Plant Viruses*. Microbiological Reviews, 1992: p. 61-79.
8. Woo, P.C.Y., et al., *Coronavirus Diversity, Phylogeny and Interspecies Jumping*. Experimental Biology and Medicine, 2009. **234**(10): p. 1117-1127.
9. Decaro, N., et al., *Recombinant Canine Coronaviruses Related to Transmissible Gastroenteritis Virus of Swine*. Journal of Virology, 2009. **February**: p. 1532-1537.
10. Herrewegh, A.A., et al., *Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus*. J Virol, 1998. **72**(5): p. 4508-14.
11. Decaro, N. and C. Buonavoglia, *An update on canine coronaviruses: viral evolution and pathobiology*. Vet Microbiol, 2008. **132**(3-4): p. 221-34.
12. Keenan, K.P., Jarvis, H.R., Marchwicki, R.H., Binn, L.N., *Intestinal infection of neonatal dogs with canine coronavirus 1-71: studies by virologic, histologic, histochemical, and immunofluorescent techniques*. Am. J. Vet. Res., 1976. **37**: p. 247-256.

13. Pratelli, A., et al., *Genetic diversity of a canine coronavirus detected in pups with diarrhoea in Italy*. J Virol Methods, 2003. **110**(1): p. 9-17.
14. Pratelli, A., et al., *Identification of coronaviruses in dogs that segregate separately from the canine coronavirus genotype*. J Virol Methods, 2003. **107**(2): p. 213-22.
15. Lorusso, A., et al., *Molecular characterization of a canine respiratory coronavirus strain detected in Italy*. Virus Res, 2009. **141**(1): p. 96-100.
16. Ma, G., Y. Wang, and C. Lu, *Molecular characterization of the 9.36 kb C-terminal region of canine coronavirus I-71 strain*. Virus Genes, 2008. **36**(3): p. 491-497.
17. Pratelli, A., et al., *Two genotypes of canine coronavirus simultaneously detected in the fecal samples of dogs with diarrhea*. J Clin Microbiol, 2004. **42**(4): p. 1797-9.
18. Hirano, N., et al., *Comparison of mouse hepatitis virus strains for pathogenicity in weanling mice infected by various routes*. Archives of Virology, 1981. **70**(1): p. 69-73.
19. Tamura, K., et al., *MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0*. Mol Biol Evol, 2007. **24**(8): p. 1596-1599.
20. Bruen, T.C., H. Philippe, and D. Bryant, *A Simple and Robust Statistical Test for Detecting the Presence of Recombination*. Genetics, 2006. **172**(4): p. 2665-2681.
21. Huson, D.H. and D. Bryant, *Application of Phylogenetic Networks in Evolutionary Studies*. Molecular Biology and Evolution, 2006. **23**(2): p. 254-267.
22. Bryant, D. and V. Moulton, *Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks*. Molecular Biology and Evolution, 2004. **21**(2): p. 255-265.
23. Martin, D.P., *Recombination Detection and Analysis Using RDP3*. 2009. p. 185-205.
24. Heath, L., et al., *Recombination Patterns in Aphthoviruses Mirror Those Found in Other Picornaviruses*. J. Virol., 2006. **80**(23): p. 11827-11832.
25. Martin, D., C. Williamson, and D. Posada, *RDP2: recombination detection and analysis from sequence alignments*. Bioinformatics, 2005. **21**: p. 260 - 262.
26. Padidam, M., S. Sawyer, and C.M. Fauquet, *Possible Emergence of New Geminiviruses by Frequent Recombination*. Virology, 1999. **265**(2): p. 218-225.
27. Martin, D.P., Posada, D., Crandall, K.A., Williamson, C., *A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints*. AIDS Res Hum Retroviruses, 2005. **21**(1): p. 5.

28. Smith, J.M., *Analyzing the mosaic structure of genes*. Journal of Molecular Evolution, 1992. **34**(2): p. 126-129.
29. Posada, D. and K.A. Crandall, *Evaluation of methods for detecting recombination from DNA sequences: Computer simulations*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(24): p. 13757-13762.
30. Gibbs, M.J., J.S. Armstrong, and A.J. Gibbs, *Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences*. Bioinformatics, 2000. **16**(7): p. 573-582.
31. Boni, M.F., D. Posada, and M.W. Feldman, *An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets*. Genetics, 2007. **176**(2): p. 1035-1047.
32. Posada, D., *Evaluation of Methods for Detecting Recombination from DNA Sequences: Empirical Data*. Mol Biol Evol, 2002. **19**(5): p. 708-717.
33. Salminen, M.a.M., D., *Detecting and characterising individual recombination events: practice.*, in *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, P. Lemey, Salemi, M., and Vandamme, A.M., Editor. 2010, Cambridge University Press: Cambridge, UK. p. 723.
34. Lorusso, A., et al., *Gain, preservation, and loss of a group 1a coronavirus accessory glycoprotein*. J Virol, 2008. **82**(20): p. 10312-7.
35. Liao, C.L. and M.M. Lai, *RNA recombination in a coronavirus: recombination between viral genomic RNA and transfected RNA fragments*. Journal of Virology, 1992. **66**(10): p. 6117-6124.
36. Addie, D.D., et al., *Feline coronavirus in the intestinal contents of cats with feline infectious peritonitis*. Vet Rec, 1996. **139**(21): p. 522-3.
37. Motokawa, K., et al., *Molecular cloning and sequence determination of the peplomer protein gene of feline infectious peritonitis virus type I*. Arch Virol, 1995. **140**(3): p. 469-80.
38. Rossen, J.W., et al., *Feline and canine coronaviruses are released from the basolateral side of polarized epithelial LLC-PK1 cells expressing the recombinant feline aminopeptidase-N cDNA*. Arch Virol, 2001. **146**(4): p. 791-9.
39. Barlough, J.E., et al., *Experimental inoculation of cats with canine coronavirus and subsequent challenge with feline infectious peritonitis virus*. Lab Anim Sci, 1984. **34**(6): p. 592-7.
40. Decaro, N., Mari, V., Elia, G., Addie, D., Camero, M., Lucente, M., Martella, V., Buonavoglia, C., *Recombinant canine coronaviruses in dogs, Europe*. Emerging Infectious Diseases, 2010. **16**(1): p. 41-47.

41. de Groot, R.J., Baker, S.C., Baric, R., Enjuanes, L., Gorbalenya, A.E., Holmes, K.V., Perlman, S., Poon, L., Rottier, P.J.M., Talbot, P.J., Woo, P.C., Ziebuhr, J., In A.M.Q. King (ed), *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. 2011 (in press): Elsevier.
42. Tresnan, D.B., R. Levis, and K.V. Holmes, *Feline aminopeptidase N serves as a receptor for feline, canine, porcine, and human coronaviruses in serogroup I*. J Virol, 1996. **70**(12): p. 8669-74.
43. Dominguez, S.R., T. O'Shea, L.M. Oko, and K.V. Holmes, *Detection of Group 1 Coronaviruses in Bats in North America*. Emerging Infectious Diseases, 2007. **13**(9).
44. Woo, P.C.Y., et al., *Molecular diversity of coronaviruses in bats*. Virology, 2006. **351**(1): p. 180-187.
45. Quan, P.-L., et al., *Identification of a Severe Acute Respiratory Syndrome Coronavirus-Like Virus in a Leaf-Nosed Bat in Nigeria*. mBio, 2010. **1**(4).
46. Drexler, J.F., et al., *Amplification of emerging viruses in a bat colony*. Emerg Infect Dis, 2011. **17**(3): p. 449-56.
47. Yu, M., et al., *Identification of key amino acid residues required for horseshoe bat angiotensin-I converting enzyme 2 to function as a receptor for severe acute respiratory syndrome coronavirus*. J Gen Virol, 2010. **91**(7): p. 1708-1712.
48. Martin, D. and E. Rybicki, *RDP: detection of recombination amongst aligned sequences*. Bioinformatics, 2000. **16**(6): p. 562-563.
49. Vijgen, L., et al., *Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43*. J Virol, 2006. **80**(14): p. 7270-4.
50. Zhang, J., et al., *Genomic characterization of equine coronavirus*. Virology, 2007. **369**(1): p. 92-104.
51. Hasoksuz, M., et al., *Biologic, Antigenic, and Full-Length Genomic Characterization of a Bovine-Like Coronavirus Isolated from a Giraffe*. J. Virol., 2007. **81**(10): p. 4981-4990.
52. Lau, S.K.P., et al., *Molecular Epidemiology of Human Coronavirus OC43 Reveals Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype due to Natural Recombination*. J. Virol., 2011. **85**(21): p. 11325-11337.
53. Thor, S.W., et al., *Recombination in Avian Gamma-Coronavirus Infectious Bronchitis Virus*. Viruses, 2011. **3**(9): p. 1777-1799.

54. Armesto, M., D. Cavanagh, and P. Britton, *The Replicase Gene of Avian Coronavirus Infectious Bronchitis Virus Is a Determinant of Pathogenicity*. PLoS ONE, 2009. **4**(10): p. e7384.
55. Snijder, E.J., et al., *Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage*. J Mol Biol, 2003. **331**(5): p. 991-1004.
56. Ziebuhr, J., V. Thiel, and A.E. Gorbalenya, *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, 2001. **276**(35): p. 33220-32.
57. Hurst, K.R., et al., *An Interaction between the Nucleocapsid Protein and a Component of the Replicase-Transcriptase Complex Is Crucial for the Infectivity of Coronavirus Genomic RNA*. J. Virol., 2010. **84**(19): p. 10276-10288.
58. Anand, K., et al., *Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs*. Science, 2003. **300**(5626): p. 1763-7.
59. Thiel, V., et al., *Mechanisms and enzymes involved in SARS coronavirus genome expression*. J Gen Virol, 2003. **84**(Pt 9): p. 2305-15.
60. Devaraj, S.G., et al., *Regulation of IRF-3-dependent innate immunity by the papain-like protease domain of the severe acute respiratory syndrome coronavirus*. J Biol Chem, 2007. **282**(44): p. 32208-21.
61. Frieman, M., et al., *Severe acute respiratory syndrome coronavirus papain-like protease ubiquitin-like domain and catalytic domain regulate antagonism of IRF3 and NF-kappaB signaling*. J Virol, 2009. **83**(13): p. 6689-705.
62. Phillips, J.E., et al., *Changes in nonstructural protein 3 are associated with attenuation in avian coronavirus infectious bronchitis virus*. Virus Genes, 2011.
63. Ammayappan, A., et al., *Complete genomic sequence analysis of infectious bronchitis virus Ark DPI strain and its evolution by recombination*. Virol J, 2008. **5**: p. 157.
64. Wang, Y.Y. and C.P. Lu, *Analysis of putative recombination hot sites in the S gene of canine coronaviruses*. Acta Virol, 2009. **53**(2): p. 111-20.
65. Lee, C.W. and M.W. Jackwood, *Evidence of genetic diversity generated by recombination among avian coronavirus IBV*. Arch Virol, 2000. **145**(10): p. 2135-48.
66. Lee, C.-W. and M.W. Jackwood, *Origin and evolution of Georgia 98 (GA98), a new serotype of avian infectious bronchitis virus*. Virus Research, 2001. **80**(1-2): p. 33-39.

67. Lim, T.-H., et al., *An emerging recombinant cluster of nephropathogenic strains of avian infectious bronchitis virus in Korea*. Infection, Genetics and Evolution, 2011. **11**(3): p. 678-685.
68. Decaro, N., et al., *Recombinant canine coronaviruses related to transmissible gastroenteritis virus of Swine are circulating in dogs*. J Virol, 2009. **83**(3): p. 1532-7.
69. Rossen, J.W., et al., *MHV-A59 enters polarized murine epithelial cells through the apical surface but is released basolaterally*. Virology, 1995. **210**(1): p. 54-66.
70. Tang, X.C., et al., *Prevalence and genetic diversity of coronaviruses in bats from China*. J Virol, 2006. **80**(15): p. 7481-90.
71. Lau, S.K.P., et al., *Ecoepidemiology and Complete Genome Comparison of Different Strains of Severe Acute Respiratory Syndrome-Related Rhinolophus Bat Coronavirus in China Reveal Bats as a Reservoir for Acute, Self-Limiting Infection That Allows Recombination Events*. J. Virol., 2010. **84**(6): p. 2808-2819.
72. Xu, L., et al., *Angiotensin-converting enzyme 2 (ACE2) from raccoon dog can serve as an efficient receptor for the spike protein of severe acute respiratory syndrome coronavirus*. J Gen Virol, 2009. **90**(Pt 11): p. 2695-703.
73. Zhang, X.W., Y.L. Yap, and A. Danchin, *Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus*. Arch Virol, 2005. **150**(1): p. 1-20.

**Table 4.1. Viruses sequenced in this study.**

<b>Alphacoronavirus</b>	
Strain	ATCC #
CCoV/1-71/1971	VR-809
CCoV/TN449	VR-2068
FCoV/WSU/79-1683	VR-989
<b>Betacoronavirus</b>	
Strain	ATCC #
MHV/S/3239-17/1963	VR-766
RaCoV/681	VR-882
RaCoV/8190/1970	VR-1410

**Table 4.2: GenBank Accession numbers for full-length genomes used in analysis**

<u>Alphacoronavirus</u>	
<u>Virus</u>	<u>Accession #</u>
FCoV/FIPV/WSU/79-1146/1979	DQ010921
FCoV/NTU156/P/2007	GQ152141
FCoV/C1Je/2007	DQ848678
FCoV/Black/1970	EU186072
FCoV/DF2	DQ286389
CCoV/NTU336/F/2008	GQ477367
TGEV/Purdue/1952	DQ811789
TGEV/ Miller/M6/1965	DQ811785
HCoV/229E/1967	AF304460
HCoV/NL63/2002	AY567487
BtCoV/1A/AFCD62	EU420138
BtCoV/HKU8/AFCD77	EU420139
BtCoV/HKU2/GD/430/2006	EF203064
BtCoV/512/2005	DQ648858
PEDV/CV777/1978	AF353511
PEDV/SM98/1998	GU937797
PEDV/LZC/China	EF185992
PRCV/ISU-1/1990	DQ811787
<u>Betacoronavirus</u>	
<u>Virus</u>	<u>Accession #</u>
BCoV/Quebec/1972	AF220295
GiCoV US/OH3/2003	EF424623
EqCoV/NC99	EF44615
HCoV/OC43/1967	AY39177
HuEnCoV/4408	FJ938067
PHEV/VW/572	DQ011855
MHV/A59/1961	AY700211
MHV/JHM	NC_006852
RaCoV/Parker	FJ938068
RoBtCoV/HKU9-1	EF065513
TyBtCoV/HKU4-1	EF065505
PiBtCoV/HKU5-1	EF065509
SARSCoV/CFB/SZ/94/03	AY545919
SARS/GD01	AY278489
SARS/Urbani	AY278741
<u>Gammacoronavirus</u>	
<u>Virus</u>	<u>Accession #</u>
IBV/Mass/Mass41/41	AY851295

**Table 4.3:** Genome organization for the (a) alpha- and (b) betacoronavirus genomes sequenced in this study.**a: Alphacoronaviruses sequenced**

ORF	FeCoV/WSU/ 79-1683			CCoV/1-71/1971			CCoV/TN449		
	Location	nt	aa	Location	nt	aa	Location	nt	aa
5'UTR	0-312	312		0-313	313		0-313	313	
1a	313-12,168	11,856	3951	314-12,367	12,054	4,017	314-12,370	12,057	4,018
1ab	313-20,168	19,856	6618	314-20,368	20,055	6,684	314-20,371	20,058	6,685
Spike	20,165-24,517	4,353	1450	20,364-24,725	4,362	1,453	20,367-24,731	4,365	1,454
3a	24,581-24,817	237	78	24,893-25,129	237	78	24,795-25,031	237	78
3b	24,732-24,929	168	55	25,074-25,289	216	71	24,976-25,191	216	71
3c	25,194-25,652	459	152	25,286-26,020	735	244	25,188-25,943	756	251
Envelope	25,722-25,928	207	68	26,007-26,255	249	82	25,909-26,157	249	82
Membrane	25,858-26,598	741	246	26,185-27,054	870	289	26,138-26,959	822	273
Nucleocapsid	26,742-27,872	1,131	376	27,067-28,215	1,149	382	26,972-28,120	1,149	382
7a	27,877-28,182	306	101	28,220-28,525	306	101	28,125-28,430	306	101
7b	28,187-28,612	426	141	28,530-29,171	642	213	28,435-29,076	642	213
3'UTR	28,613-28,984	372		29,172-29,462	291		29,077-29,364	288	

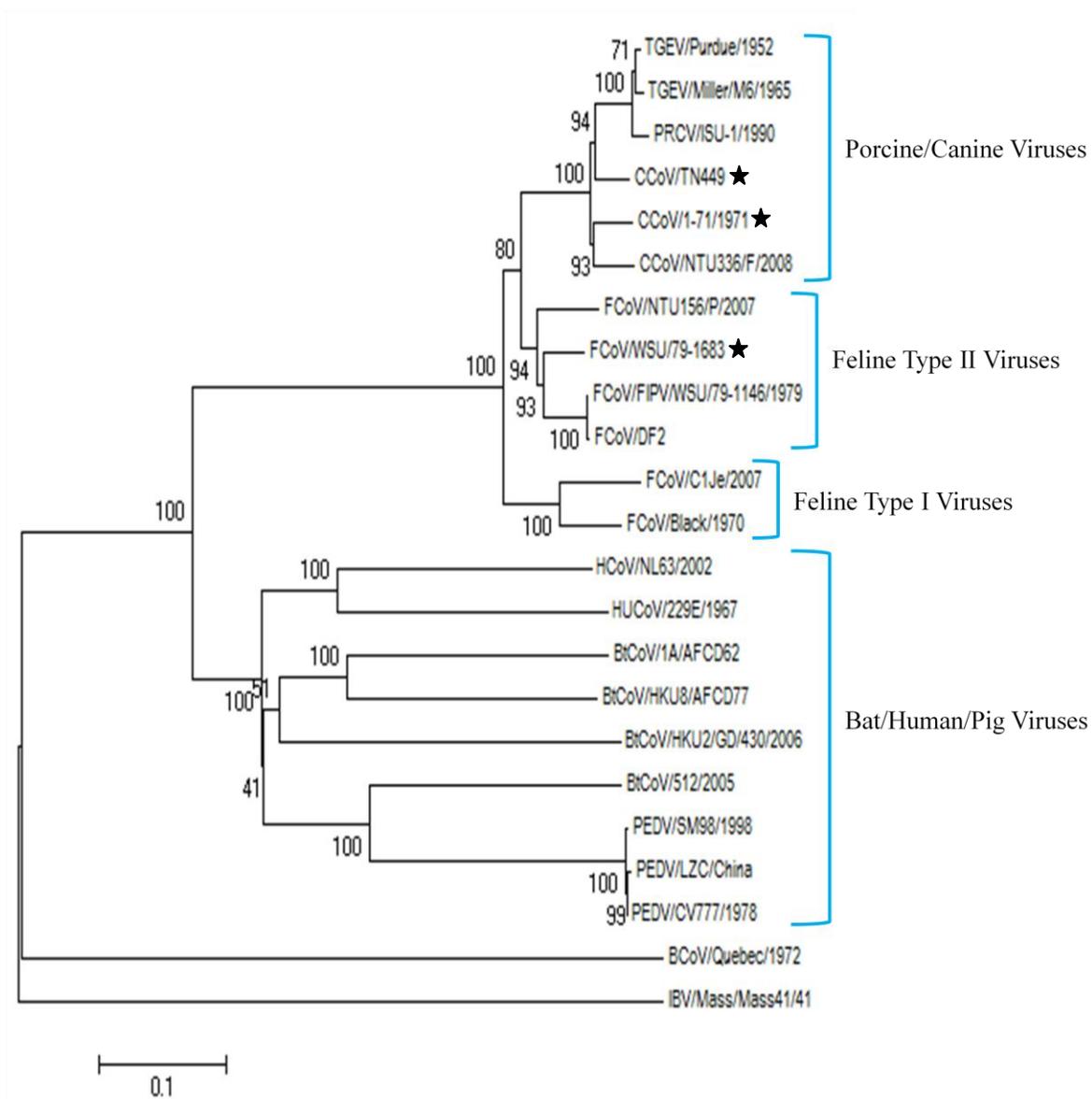
**b: Betacoronaviruses sequenced**

ORF	RaCoV/8190/1970			RaCoV/681			MHV/S/3239-17/1963		
	Location	nt	aa	Location	nt	aa	Location	nt	aa
5'UTR	0-182	182		0-182	182		0-210	210	
1a	183-13,586	13,404	4,467	183-13,589	13,407	4,468	211-13,626	13,416	4,471
1ab	183-21,710	21,528	7,175	183-21,713	21,531	7,176	211-21,750	21,570	7,189
ns 2a	21,698-22,520	822	273	21,702-22,523	822	273	21,763-22,560	798	265
HE	22,548-23,885	1,338	445	22,569-23,882	1,314	437	22,606-23,925	1,320	439
Spike	23,900-27,982	4,083	1,360	23,897-27,979	4,083	1,360	23,933-28,018	4,086	1,361
ns 4	28,073-28,492	420	139	28,070-28,471	402	133	28,108-28,482	375	124
Envelope	28,630-28,896	267	88	28,645-28,911	267	88	28,517-28,783	267	88
Membrane	28,893-29,582	690	229	28,908-29,594	687	228	28,780-29,466	687	228
<sup>a</sup> Internal protein	29,661-30,284	624	207	29,673-30,296	624	207			
Nucleocapsid	29,597-30,961	1,365	454	29,609-30,973	1,365	454	29,481-30,845	1,365	454
3'UTR	30,962-31,274	313		30,974-31,286	313		30,846-31,164	319	

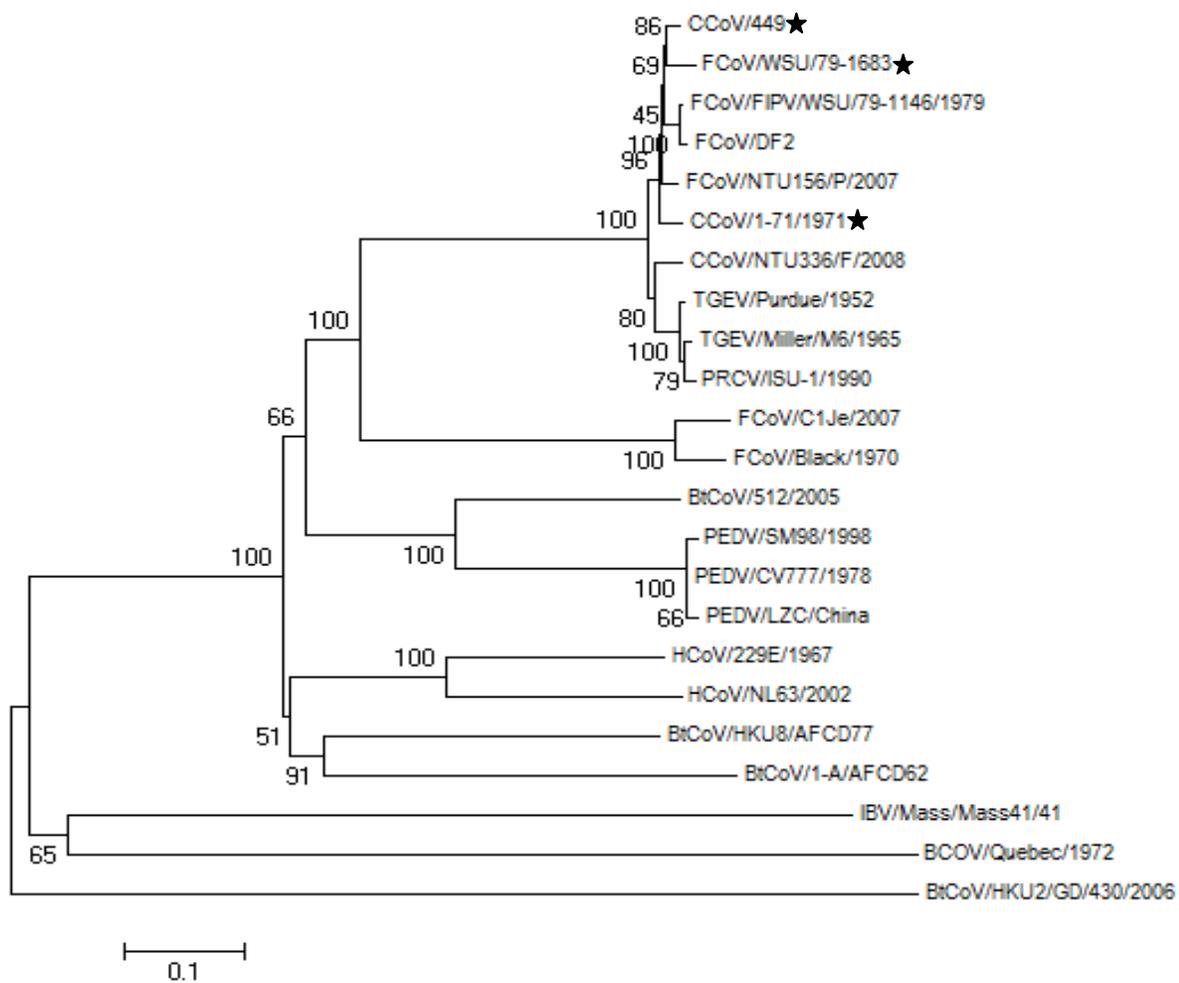
<sup>a</sup>Internal protein is not coded for in MHV.

**Figure 4.1:** Neighbor Joining Phylogenetic tree of (a) full-length sequence data and (b) spike protein for the alphacoronaviruses. The Neighbor-joining method was used to infer the evolutionary history for 21 full-length genomic sequence data available for the alphacoronaviruses. The percentage of replicate trees in which the taxa grouped together in a bootstrap test of 1000 replicates is shown at branch nodes in each tree. The p-distance scale is presented at the bottom of each figure. A star (★) indicates a virus newly sequenced in this study. The BCoV/Quebec/1972 and IBV/Mass/Mass41/41 were included in this analysis as outgroups.

(a)



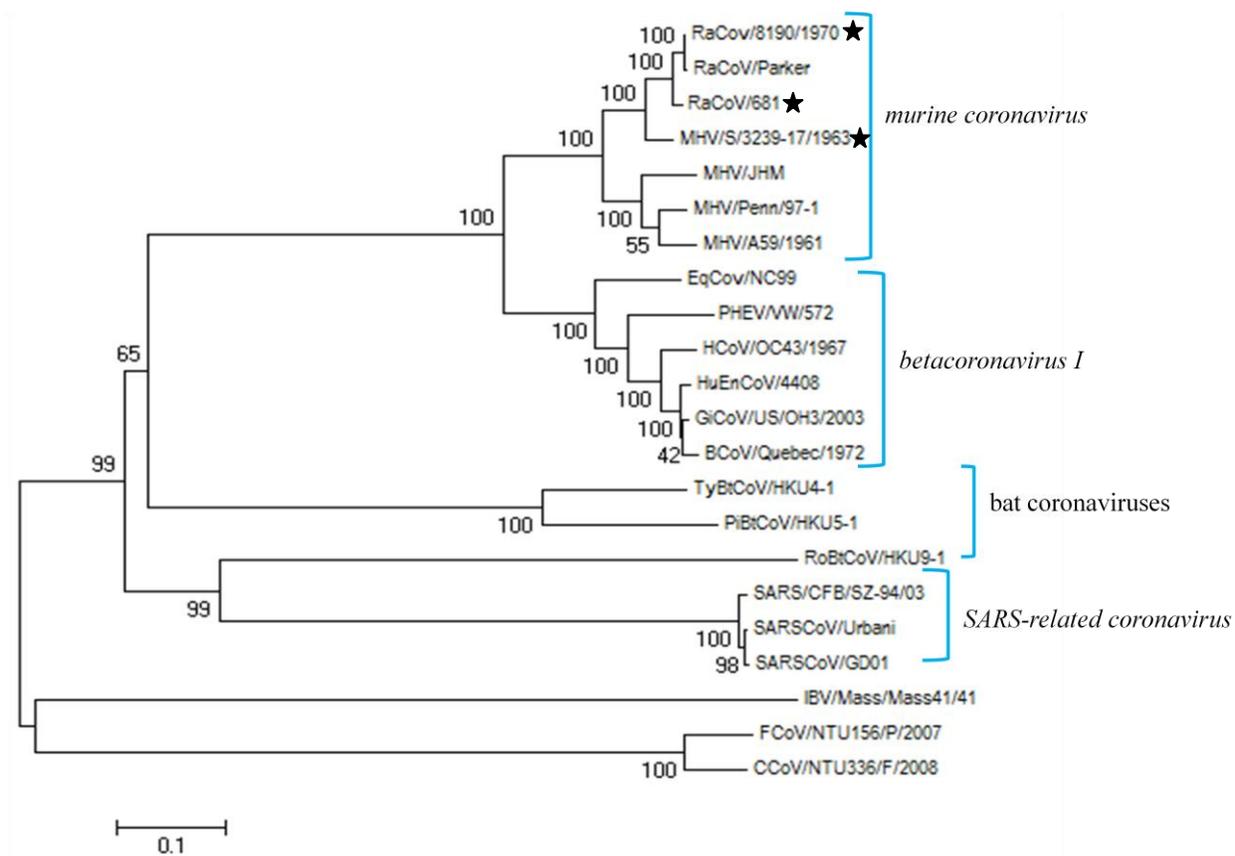
(b)



**Figure 4.2:** Neighbor Joining Phylogenetic tree of (a) full-length sequence data and (b) spike protein for the betacoronaviruses. The Neighbor-joining method was used to infer the evolutionary history for 19 full-length genomic sequence data available for the betacoronaviruses. The percentage of replicate trees in which the taxa grouped together in a bootstrap test of 1000 replicates is shown at branch nodes in each tree. The p-distance scale is presented at the bottom of each figure. A star (★) indicates a virus newly sequenced in this study. The IBV/Mass/Mass41/41, FCoV/NTU156/P/2007, and CCoV/NTU336/F/2008 were included in this analysis as outgroups.



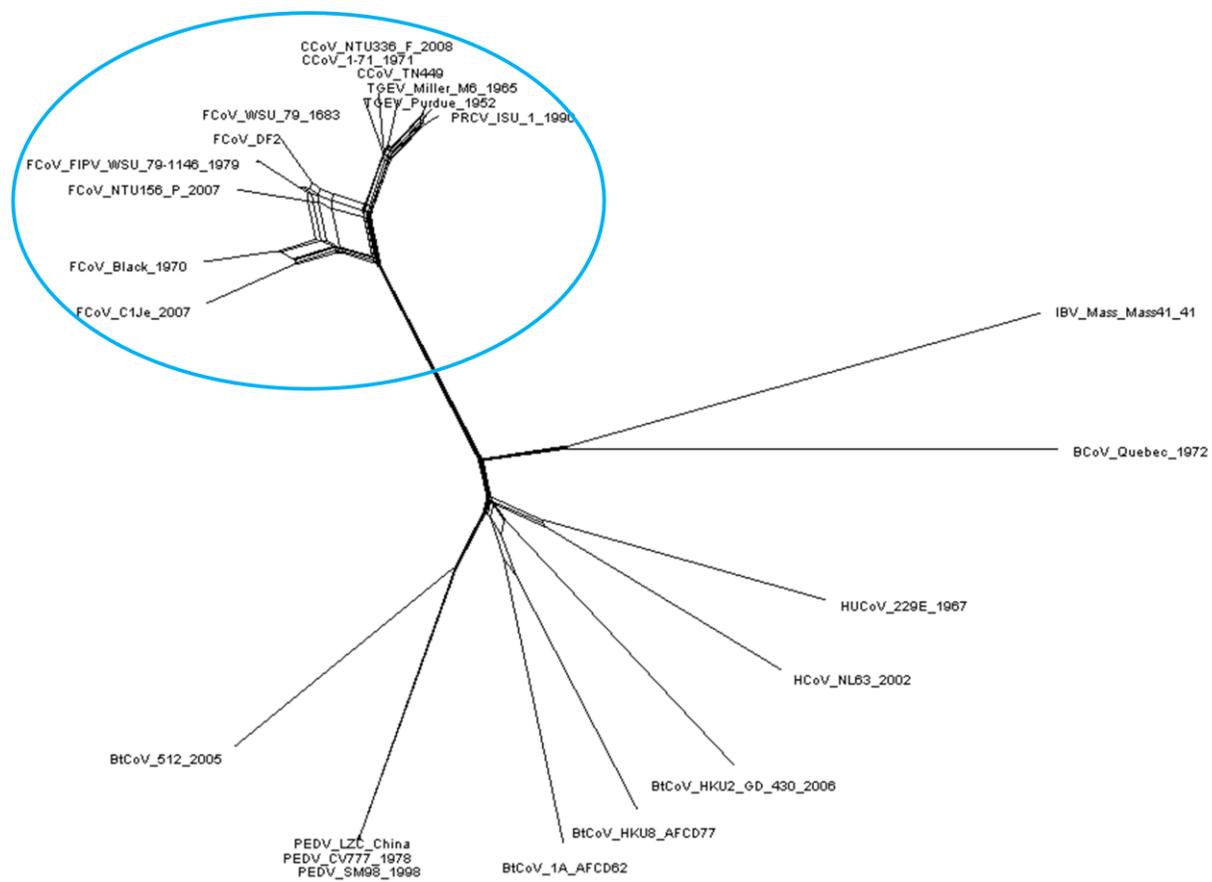
(b)



**Figure 4.3:** Neighbor Net analysis for alphacoronavirus full length genomes; (a) all analyzed alphacoronaviruses and (b) enlargement of circled area. Networked relationships indicated the presences of reticulate events and boxes imply the likelihood of recombination. PHI test of recombination was significant ( $p < 0.0001$ ).

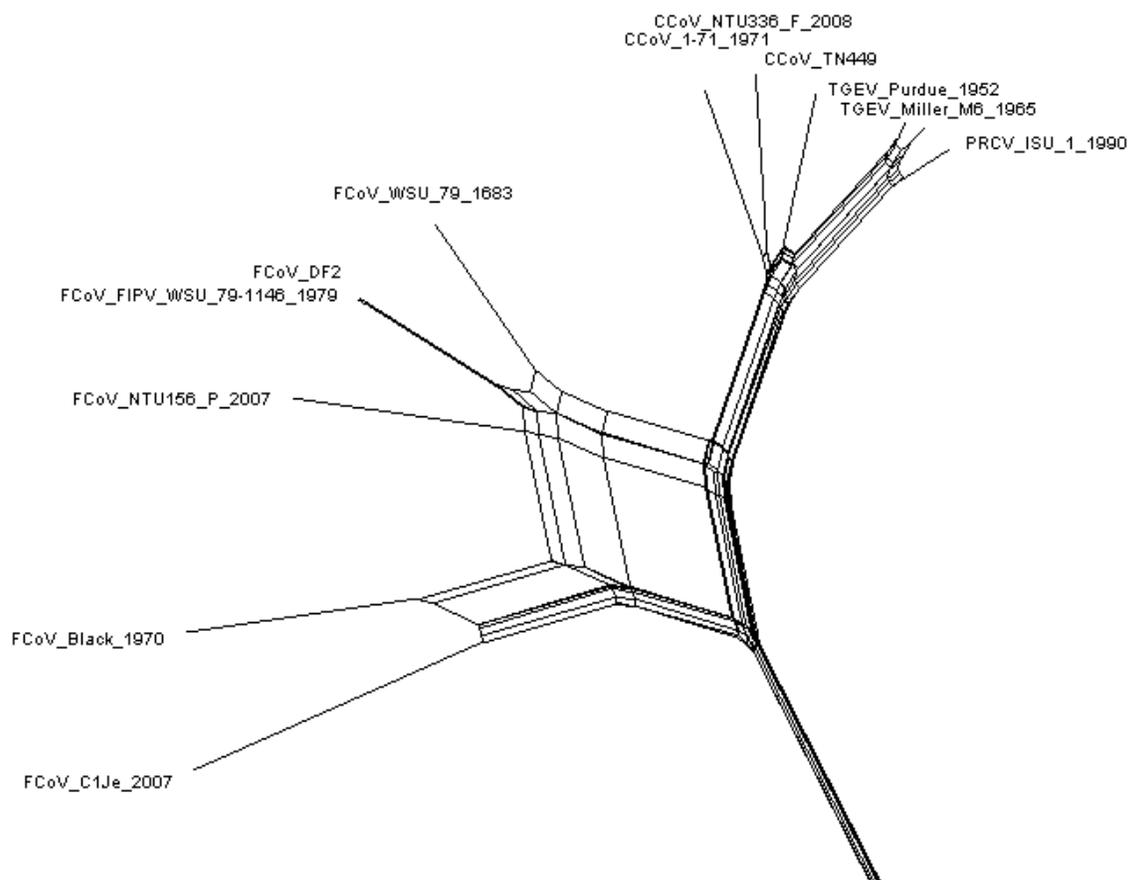
(a)

|—0.01



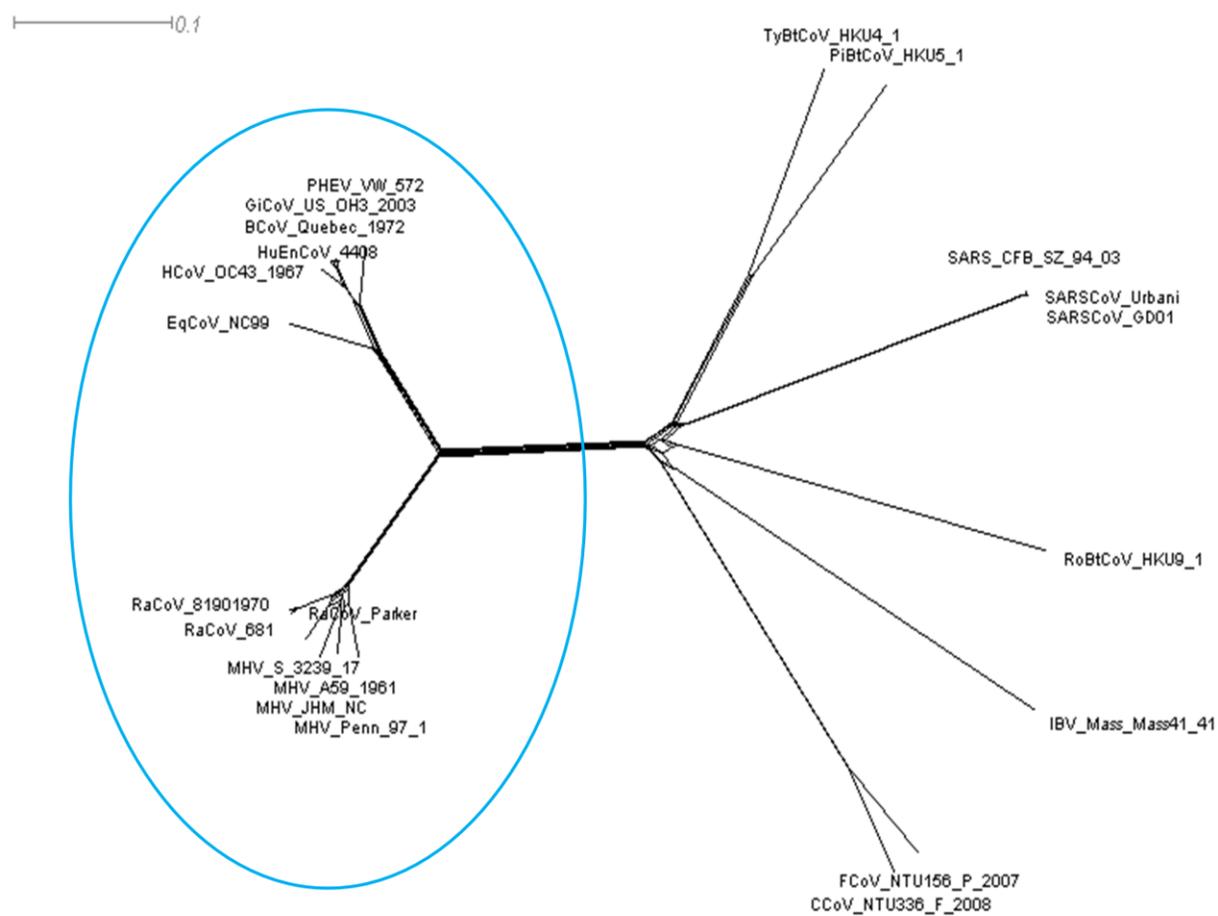
**(b)**

|-----| 0.01



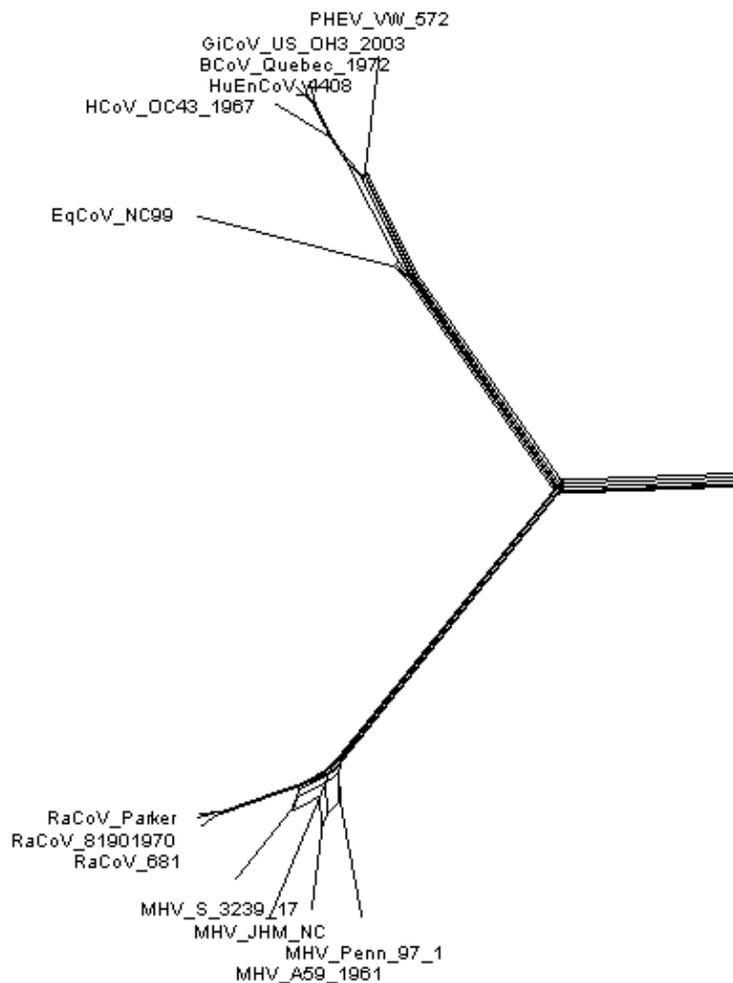
**Figure 4.4:** Neighbor Net analysis for betacoronavirus full length genomes; (a) all analyzed betacoronaviruses and (b) enlargement of circled area. Networked relationships indicated the presences of reticulate events and boxes imply the likelihood of recombination. PHI test of recombination was significant ( $p < 0.0001$ ).

(a)



(b)

0.1



**Table 4.4: Recombination breakpoints<sup>a</sup>, genes and major and minor related sequences for (a) alpha- and (b) betacoronaviruses**

Recombinant <sup>a</sup>	Breakpoints		Genes <sup>b,e,f,g</sup>	Major Sequences <sup>c</sup>	Minor Sequences <sup>d</sup>	Detection Method
	Begin	End				
FCoV/Black/1970 <sup>f</sup>	*3,063	4,236	1ab (nsp 3)	Unknown ( FCoV/WSU/79-1683/1979)	FCoV/NTU156/P/2007	RDP, Geneconv, Maxchi, Chimaera, SiScan
	4,985	5,712	1ab (nsp 3)	FCoV/WSU/79-1146/1979	Unknown (FCoV/WSU/79-1683/1979)	Maxchi ,Chimaera
	6,961	7,617	1ab (nsp 3)	FCoV/DF2	Unknown (FCoV/WSU/79-1683/1979)	RDP, GeneConv, Maxchi ,Chimaera , SiScan
	*13,128	14,227	1ab (nsp 11/12)	CCoV/ NTU336/F/2008	FCoV/NTU156/P/2007	RDP, GeneConv, Maxchi ,Chimaera, 3Seq
	16,455	17,155	1ab (nsp 13)	FCoV/C1Je/2007	FCoV/NTU156/P/2007	RDP, Maxchi, Chimaera
	*21,625	21,733	Spike	CCoV/NTU336/F/2008	Unknown (PEDV/LZC/China)	RDP, Maxchi
	22,211	22,344	Spike	CCoV/NTU336/F/2008	Unknown ( HCoV/229E/1967)	RDP, GeneConv
	22,407	22,591	Spike	CCoV/I-71/1971	BiCoV/512/2005	RDP, SiScan
	*23,069	25,835	Spike,3a, 3b, 3c, Envelope	FCoV/WSU/79-1146/1979	Unknown ( CCoV/TN449)	RDP, GeneConv, Bootscan,Maxchi ,Chimaera, SiScan, 3Seq
	*23,085	*23,106	Spike	FCoV/WSU/79-1683/1979	Unknown (HCoV/NL63/2002)	GeneConv, 3Seq
	*23,123	23,217	Spike	CCoV/NTU336/F/2008	Unknown (BiCoV/HKU8/AFCD77)	RDP, GeneConv
	*23,337	*24,135	Spike	FCoV/WSU/79-1683/1979	BiCoV/512/2005	RDP, GeneConv, Maxchi ,Chimaera
	24,158	24,597	Spike	PEDV/CV777/1978	Unknown (BiCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
	24,791	28,132	Spike, Envelope, Membrane, Nucleocapsid, 3' UTR	Unknown (FCoV/NTU/156/F/2008)	FCoV/WSU/79-1683/1979	Maxchi, Chimaera, SiScan
	FCoV/C1Je/2007 <sup>f</sup>	*3,063	4,236	1ab (nsp3)	Unknown ( FCoV/WSU/79-1683/1979)	FCoV/NTU156/P/2007
*13,128		14,227	1ab (nsp 11/12)	CCoV/ NTU336/F/2008	FCoV/NTU156/P/2007	RDP, GeneConv, Maxchi ,Chimaera, 3Seq
21,274		21,482	Spike	HCoV/NL63/2002	CCoV/TN449	RDP, Geneconv
22,211		22,344	Spike	CCoV/NTU336/F/2008	Unknown (HCoV/229E/1967)	RDP, GeneConv
22,407		22,591	Spike	CCoV/I-71/1971	BiCoV/512/2005	RDP, SiScan
*23,069		25,835	Spike,3a, 3b, 3c, Envelope	FCoV/WSU/79-1146/1979	Unknown ( CCoV/TN449)	RDP, GeneConv, Bootscan,Maxchi ,Chimaera, SiScan, 3Seq
*23,085		*23,106	Spike	FCoV/WSU/79-1683/1979	Unknown (HCoV/NL63/2002)	GeneConv, 3Seq
*23,123		23,217	Spike	CCoV/NTU336/F/2008	Unknown (BiCoV/HKU8/AFCD77)	RDP, GeneConv
*23,337		*24,135	Spike	FCoV/WSU/79-1683/1979	BiCoV/512/2005	RDP, GeneConv, Maxchi ,Chimaera
24,158		24,597	Spike	PEDV/CV777/1978	Unknown (BiCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
FCoV/DF2 <sup>f</sup>	*415	2,822	1ab (nsp 1 and 2)	FCoV/Black/1970	FCoV/WSU/79-1683/1979	RDP, Maxchi, Chimaera, SiScan, 3Seq
	*3,063	4,236	1ab (nsp 3)	Unknown ( FCoV/WSU/79-1683/1979)	FCoV/NTU156/P/2007	RDP, Geneconv, Maxchi, Chimaera, SiScan
	13,074	23,022	1ab (nsp 11/12, 13, 14, 15, and 16), Spike	FCoV/Black/1970	PRCV/ISU-1/1990	RDP, Geneconv, Maxchi , Chimaera ,Siscan
	*13,167	13,351	1ab (nsp 11/12)	CCoV/NTU336/F/2008	Unknown (FCoV/WSU/79-	RDP, Geneconv, Chimaera

					1683/1979)	
	14,760	15,939	1ab (nsp 11/12 and 13)	Unknown (PRCV/ISU-1/1990)	CCoV/NTU336/F/2008	Maxchi, Chimaera
	*23,032	*23,281	Spike	CCoV/TN449	Unknown (PRCV/ISU-1/1990)	RDP, SiScan
	*23,284	25,542	Spike, 3a, 3b, 3c	CCoV/1-71/1971	CCoV/TN449	RDP, Maxchi, Chimaera ,SiScan
	24,158	24,597	Spike	PEDV/CV777/1978	Unknown (BtCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
	24,791	28,132	Nucleocapsid, 7a, and 3' UTR	Unknown (FCoV/NTU/156/F/2008)	FCoV/WSU/79-1683/1979	Maxchi, Chimaera, SiScan
	26,393	*27,048	Nucleocapsid	FCoV/Black/1970	Unknown (FCoV/C1Je/2007)	RDP, Chimaera
FCoV/NTU156/P/2007 <sup>f</sup>	6,961	7,617	1ab (nsp 3)	FCoV/DF2	Unknown (FCoV/WSU/79-1683/1979)	RDP, GeneConv, Maxchi ,Chimaera , SiScan
	13,074	23,022	1ab (nsp 11/12, 13, 14, 15, and 16), Spike	FCoV/Black/1970	PRCV/ISU-1/1990	RDP, Geneconv, Maxchi , Chimaera ,SiScan
	*13,058	16,102	1ab (nsp 11/12 and 13)	CCoV/NTU336/F/2008	Unknown (CCoV/TN449)	RDP, Geneconv , Maxchi , Chimaera ,
	17,879	*20,286	1ab (nsp 14, 15, and 16)	CCoV/1-71	FCoV/WSU/79-1146/1979	RDP, Maxchi, Chimaera, SiScan,
	*23,054	*24,988	Spike, 3a, 3b	CCoV/1-71/1971	FCoV/NTU156/P/2007	RDP, Geneconv, Maxchi , Chimaera ,SiScan
	24,158	24,597	Spike	PEDV/CV777/1978	Unknown (BtCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
	25,006	25,421	3a, 3b, 3c	FCoV/WSU/79-1146/1979	Unknown (CCoV/TN449)	RDP, Geneconv, Maxchi, Chimaera, SiScan
	*25,522	26,298	3c, Envelope and Membrane	FCoV/Black/1970	CCoV/1-71	RDP, GENECONV, Maxchi, Chimaera, SiScan,
FCoV/WSU/79-1146/1979 <sup>f</sup>	*415	2,822	1ab (nsp1 and 2)	FCoV/Black/1970	FCoV/WSU/79-1683/1979	RDP, Maxchi, Chimaera, SiScan, 3Seq
	*3,063	4,236	1ab (nsp 3)	Unknown ( FCoV/WSU/79-1683/1979)	FCoV/NTU156/P/2007	RDP, Geneconv, Maxchi, Chimaera, SiScan
	13,074	23,022	1ab (nsp 11/12, 13, 14, 15, and 16), Spike	FCoV/Black/1970	PRCV/ISU-1/1990	RDP, Geneconv, Maxchi , Chimaera ,SiScan
	*13,167	13,351	1ab (nsp 11/12)	CCoV/NTU336/F/2008	Unknown (FCoV/WSU/79-1683/1979)	RDP, Geneconv, Chimaera
	14,760	15,939	1ab (nsp 11/12 and 13)	Unknown (PRCV/ISU-1/1990)	CCoV/NTU336/F/2008	Maxchi, Chimaera
	*23,032	*23,281	Spike	CCoV/TN449	Unknown (PRCV/ISU-1/1990)	RDP, SiScan
	*23,284	25,542	Spike, 3a, 3b, 3c	CCoV/1-71/1971	CCoV/TN449	RDP, Maxchi, Chimaera ,SiScan
	24,158	24,597	Spike	PEDV/CV777/1978	Unknown (BtCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
	24,791	28,132	3a, 3b, 3c, Envelope, Membrane, Nucleocapsid, 7a	Unknown (FCoV/NTU/156/F/2008)	FCoV/WSU/79-1683/1979	Maxchi, Chimaera, SiScan
	26,393	*27,048	Nucleocapsid	FCoV/Black/1970	Unknown (FCoV/C1Je/2007)	RDP, Chimaera
FCoV/WSU/79-1683/1979 <sup>f</sup>	9,486	10,872	1ab (nsp5 and 6)	Unknown (FCoV/NTU156/P/2007)	FCoV/Black/1970	RDP, GENECONV, Maxchi, Chimaera, SiScan, 3Seq
	11,876	*12,791	1ab (nsp 9 and 10)	FCoV/WSU/79-1146/1979	CCoV/TN449	RDP, GENECONV, Maxchi, Chimaera, SiScan, 3Seq
	13,074	23,022	1ab (nsp 11/12, 13, 14, 15, and 16), Spike	FCoV/Black/1970	PRCV/ISU-1/1990	RDP, Geneconv, Maxchi , Chimaera ,SiScan
	14,760	15,939	1ab (nsp 11/12 and 13)	Unknown (PRCV/ISU-1/1990)	CCoV/NTU336/F/2008	Maxchi, Chimaera
	*23,054	*24,988	Spike, 3a, 3b	CCoV/1-71/1971	FCoV/NTU156/P/2007	RDP, Geneconv, Maxchi , Chimaera ,SiScan
	24,158	24,597	Spike	PEDV/CV777/1978	Unknown (BtCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
	24,907	25,012	3b	FCoV/WSU/79-1146/1979	Unknown (CCoV/1-71/1971)	RDP, GENECONV, Maxchi, Chimaera

	*25,522	26,298	3c, Envelope and Membrane	FCoV/Black/1970	CCoV/I-71	RDP, GENECONV, Maxchi, Chimaera, SiScan,
	25,967	26,073	Membrane	CCoV/I-71/1971	CCoV/TN449	RDP, GENECONV, Bootscan, Maxchi, Chimaera, 3Seq
	*26,265	*26,472	Membrane	TGEV/Purdue/1952	FCoV/C1Je/2007	RDP, Maxchi
	28,603	*28,686	7b and 3' UTR	FCoV/Black/1970	Unknown ( CCoV/NTU336/F/2008)	RDP, GENECONV, Maxchi, Chimaera
CCoV/I-71 <sup>†</sup>	2,342	3,352	1ab (nsp 2 and 3)	CCoV/NTU336/F/2008	CCoV/TN449	RDP, GENECONV, Maxchi, Chimaera, SiScan
	3,834	3,972	1ab (nsp 3)	CCoV/NTU336/F/2008	FCoV/WSU/79-1683/1979	RDP, GENECONV, Maxchi, Chimaera
	21,842	*22,812	Spike	FCoV/NTU156/P/2007	TGEV/Miller/M6/1965	RDP, Maxchi, Chimaera, SiScan, 3Seq
	24,158	24,597	Spike	PEDV/CV777/1978	Unknown (BtCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
CCoV/NTU336/F/2008 <sup>†</sup>	5,790	6,462	1ab (nsp 3)	CCoV/I-71/1971	TGEV/Purdue/1952	Maxchi, Chimaera, SiScan
	7,518	10,537	1ab (nsp 3, 4, 5, and 6)	CCoV/I-71	CCoV/TN449	RDP, Maxchi, SiScan
	17,879	*20,286	1ab (nsp 14, 15, and 16)	CCoV/I-71	FCoV/WSU/79-1146/1979	RDP, Maxchi, Chimaera, SiScan,
	*21,213	21,407	Spike	FCoV/WSU/79-1146/1979	Unknown ( FCoV/NTU156/P/2007)	RDP, 3Seq
	24,158	24,597	Spike	PEDV/CV777/1978	Unknown (BtCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
CCoV/TN449 <sup>†</sup>	3,834	*3,972	1ab (nsp 3)	CCoV/NTU336/F/2008	FCoV/WSU/79-1683/1979	RDP, GENECONV, Maxchi, Chimaera
	5,790	6,462	1ab (nsp 3)	CCoV/I-71/1971	TGEV/Purdue/1952	Maxchi, Chimaera, SiScan
	*20,741	21,507	Spike	FCoV/NTU156/P/2007	Unknown (FCoV/DF2)	Maxchi, Chimaera,
	24,158	24,597	Spike	PEDV/CV777/1978	Unknown (BtCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
	*24,725	24,775	Spike	FCoV/NTU156/P/2007	Unknown ( TGEV/Purdue/1952)	GeneConv, SiScan
TGEV/Miller/M6/1965 <sup>†</sup>	20,406	21,315	1ab (nsp16) and Spike	CCoV/TN449	Unknown (FCoV/WSU/79-1683/1979)	RDP, GENECONV, Maxchi, Chimaera, Siscan
	*21,345	*21,464	Spike	FCoV/NTU156/P/2007	Unknown (FCoV/WSU/79-1146)	GENECONV, SiScan
	24,158	24,597	Spike	PEDV/CV777/1978	Unknown (BtCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
	*24,276	*24,833	Spike and 3a	CCoV/TN449	Unknown (FCoV/WSU/79-1683/1979)	RDP, Bootscan, SiScan
	28,220	*28,544	7 and 3' UTR	CCoV/TN449	Unknown ( CCoV/I-71/1971)	RDP, GENECONV, Maxchi, Chimaera, SiScan
	*28,222	28,287	7 and 3' UTR	PEDV/SM98/1998	CCoV/TN449	Geneconv, , Maxchi, Chimaera, Siscan
TGEV/Purdue/1952 <sup>†</sup>	20,406	21,315	1ab (nsp16) and Spike	CCoV/TN449	Unknown (FCoV/WSU/79-1683/1979)	RDP, GENECONV, Maxchi, Chimaera, Siscan
	*21,345	*21,464	Spike	FCoV/NTU156/P/2007	Unknown (FCoV/WSU/79-1146)	GENECONV, SiScan
	24,158	24,597	Spike	PEDV/CV777/1978	Unknown (BtCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
	28,220	*28,544	7 and 3' UTR	CCoV/TN449	Unknown ( CCoV/I-71/1971)	RDP, GENECONV, Maxchi, Chimaera, SiScan
	*28,222	28,287	7 and 3' UTR	PEDV/SM98/1998	CCoV/TN449	Geneconv, , Maxchi, Chimaera, Siscan
PRCV/ISU-1/1990 <sup>†</sup>	20,406	21,315	1ab (nsp16) and Spike	CCoV/TN449	Unknown (FCoV/WSU/79-1683/1979)	RDP, GENECONV, Maxchi, Chimaera, Siscan
	24,158	24,597	3b, Envelope, Membrane	PEDV/CV777/1978	Unknown (BtCoV/HKU2/GD/430/2006)	RDP, Geneconv, Maxchi, Chimaera, SiScan, 3Seq
	*24,276	*24,833	3b	CCoV/TN449	Unknown (FCoV/WSU/79-1683/1979)	RDP, Bootscan, SiScan
PEDV/CV777/1978 <sup>†</sup>	8,214	8,912	1ab (nsp 4)	HCoV/NL63/2002	FCoV/C1Je/2007	Maxchi, Chimaera
	21,365	*21,483	Spike	BtCoV/512/2005	Unknown ( TGEV/Miller/M6/1965)	RDP, Chimaera

	25,158	*25,429	Spike	BtCoV/1A/AFCD62	FCoV/C1Je/2007	Maxchi, Chimaera
PEDV/SM98/1998 <sup>f</sup>	8,214	8,912	1ab (nsp 4)	HCoV/NL63/2002	FCoV/C1Je/2007	Maxchi, Chimaera
	21,365	*21,483	Spike	BtCoV/512/2005	Unknown ( TGEV/Miller/M6/1965)	RDP, Chimaera
	25,158	*25,429	Spike	BtCoV/1A/AFCD62	FCoV/C1Je/2007	Maxchi, Chimaera
PEDV/LZC/China <sup>f</sup>	8,214	8,912	1ab (nsp 4)	HCoV/NL63/2002	FCoV/C1Je/2007	Maxchi, Chimaera
	21,365	*21,483	Spike	BtCoV/512/2005	Unknown ( TGEV/Miller/M6/1965)	RDP, Chimaera
	25,158	*25,429	Spike	BtCoV/1A/AFCD62	FCoV/C1Je/2007	Maxchi, Chimaera
BtCoV/HKU2/GD/430/2006 <sup>g</sup>	1,023	1,286	1ab (nsp 2)	BtCoV/512/2005	TGEV/virulent/Purdue/1952	RDP, SiScan
	1,530	1,990	1ab (nsp 2 and 3)	BtCoV/HKU8/AFCD77	Unknown ( HCoV/229E/1967)	Maxchi, Chimaera
	4,687	5,205	1ab (nsp 3)	BtCoV/512/2005	CCoV/TN449	Maxchi, Chimaera
	11,416	11,630	1ab (nsp 6)	BtCoV/HKU8/AFCD77	TGEV/Purdue/1952	Maxchi, Chimaera
	23,357	23,724	Spike	PEDV/CV777/1978	Unknown ( HCoV/229E/1967)	Maxchi, Chimaera, SiScan
BtCoV/HKU8/AFCD77 <sup>g</sup>	27,026	27,054	Nucleocapsid	Unknown (PEDV/SM98/1998)	FCoV/WSU/79-1146/1979	Gencon, Maxchi
BtCoV/512/2005 <sup>g</sup>	2,409	2,697	1ab (nsp 3)	BtCoV/HKU8/AFCD77	HCoV/229E/1967	RDP, SiScan
	8,214	8,912	1ab (nsp4)	HCoV/NL63/2002	FCoV/C1Je/2007	Maxchi, Chimaera
	25,158	*25,429	Spike	BtCoV/1A/AFCD62	FCoV/C1Je/2007	Maxchi, Chimaera
HCoV/NL63/2002 <sup>g</sup>	2,967	3,089	1ab (nsp 3)	Unknown (BtCoV/1A/AFCD62)	BtCoV/512/2005	RDP, Geneconv
	3,741	3,851	1ab (nsp 3)	BtCoV/512/2005	CCoV/NTU336/F/2008	Maxchi, Chimaera
	4,687	5,205	1ab (nsp 3)	BtCoV/512/2005	CCoV/TN449	Maxchi, Chimaera
	*12,214	12,604	1ab (nsp 8)	BtCoV/HKU2/GD/430/2006	Unknown (FCoV/WSU/79-1683/1979)	Maxchi, Chimaera
	*26,820	27,141	Nucleocapsid	HCoV/229E/1967	BtCoV/HKU2/GD/430/2006	Chimaera, SiScan
HCoV/229E/1967 <sup>g</sup>	642	1,275	1ab (nsp 1 and 2)	BtCoV/512/2005	FCoV/NTU156/P/2007	RDP, Maxchi
	2,967	3,089	1ab (nsp 3)	Unknown (BtCoV/1A/AFCD62)	BtCoV/512/2005	RDP, Geneconv
	11,674	11,946	1ab (nsp 6 and 7)	HCoV/NL63/2002	Unknown ( BtCoV/1A/AFCD62)	RDP, SiScan
	*12,214	12,604	1ab (nsp 8)	BtCoV/HKU2/GD/430/2006	Unknown (FCoV/WSU/79-1683/1979)	Maxchi, Chimaera
	20,656	20,836	Spike	HCoV/NL63/2002	Unknown (FCoV/DF2)	Maxchi, Chimaera
	26,027	26,159	Nucleocapsid	HCoV/NL63/2002	PEDV/LZC/China	RDP, Maxchi

\* The actual breakpoint position is undetermined. Most likely it was overprinted by a subsequent recombination event. (a) Only transferred gene fragments detected by more than one method are included in this table. (b) Genes indicate the coding sequences contained within the fragment introduced by recombination. (c) Major Sequence = the sequence most closely related to the sequence surrounding the transferred fragment. (d) Minor Sequence = the sequence most closely related to the transferred fragment in the

recombinant. (e) Unknown = only one parent and a recombinant need be in the alignment for a transferred fragment to be detectable. The sequence listed in parentheses was used to infer the existence of a missing parental sequence. (f) The identification of the nsps coded in ORF 1ab were based on Dye and Siddell, 2007 (g) The identification of the nsps coded in ORF 1ab were based on Snijder, *et al*, 2003.

(b)

Recombinant <sup>a</sup>	Breakpoints		Genes <sup>b, f</sup>	Major Sequences <sup>c, e</sup>	Minor Sequences <sup>d, e</sup>	Method(s) used
	Begin	End				
MHV/A59/1961	10,251	10,497	1ab (nsp 5)	RaCov/8190/1970	Unknown (MHV/JHM)	RDP, GENECONV, Maxchi, SiScan
	22,578	26,030	HE, Spike	MHV/S/3239-17/1963	MHV/Penn	RDP, GENECONV, Maxchi, Chimaera, SiScan, 3Seq
	*26,042	27,266	Spike	MHV/S/3239-17/1963	MHV/JHM	RDP, GENECONV, Maxchi, Chimaera, SiScan, 3Seq
MHV/JHM	3,811	5,028	1ab (nsp 3)	Unknown (RaCoV/Parker)	MHV/S/3239-17	RDP, GENECONV, Maxchi, Chimaera, SiScan
	13,905	15,474	1ab (nsp 12)	Unknown (RaCoV/681)	MHV/A59	Maxchi, Chimaera, SiScan
	18,584	*18,817	1ab (nsp 14)	Unknown (RaCoV/681)	MHV/Penn	RDP, GENECONV, SiScan
	22,578	26,030	HE, Spike	MHV/S/3239-17/1963	MHV/Penn	RDP, GENECONV, Maxchi, Chimaera, SiScan, 3Seq
	24,134	24,193	Spike	RaCoV/8190/1970	MHV/S/3239-17/1963	RDP, GENECONV
	*26,198	27,848	Spike	Unknown (RaCoV/681)	MHV/Penn	RDP, Maxchi, Chimaera, SiScan
	29,732	29,969	Membrane, Nucleocapsid	RaCoV/Parker	MHV/Penn	RDP, GENECONV, Maxchi, Chimaera, SiScan
	MHV/Penn	*122	932	1ab (nsp 1)	RaCoV/681	Unknown (MHV/JHM)
1,991		2,179	1ab (nsp 2)	Unknown (MHV/JHM)	RaCoV/681	GENECONV, Maxchi, Chimaera
3,790		4,256	1ab (nsp 3)	Unknown (MHV/S/3239-17/1963)	MHV/A59/1961	RDP, GENECONV, Maxchi, Chimaera, SiScan
MHV/S/3239-17/1963	10,251	10,497	1ab (nsp5)	RaCov/8190/1970	Unknown (MHV/JHM)	RDP, GENECONV, Maxchi, SiScan
	*22,675	22,962	HE	RaCoV/681	MHV/A59/1961	RDP, Maxchi, Chimaera, 3Seq
	29,732	29,969	Nucleocapsid	RaCoV/Parker	MHV/Penn	RDP, GENECONV, Maxchi, Chimaera, SiScan
RaCoV/8190/1970	8,155	8,834	1ab (nsp 3 and 4)	MHV/S/3239-17/1963	Unknown (MHV/Penn)	Maxchi, Chimaera, SiScan
	*27,758	28,254	Spike and ns4	MHV/S/3239-17/1963	Unknown (MHV/JHM)	GENECONV, Maxchi, SiScan, 3Seq
	28,462	28,516	ns 4	RaCoV/Parker	MHV/Penn	GENECONV, Maxchi, Chimaera, SiScan,
	30,790	30,887	Nucleocapsid	MHV/S/3239-17/1963	Unknown (MHV/Penn)	GENECONV, Maxchi, SiScan
RaCoV/681	8,155	8,834	1ab (nsp 3 and 4)	MHV/S/3239-17/1963	Unknown (MHV/Penn)	Maxchi, Chimaera, SiScan
	19,243	20,307	1ab (nsp 15)	RaCoV/Parker	MHV/A59/1961	RDP, Maxchi, Chimaera
	*27,758	28,254	Spike and ns4	MHV/S/3239-17/1963	Unknown (MHV/JHM)	GENECONV, Maxchi, SiScan, 3Seq
	30,790	30,887	Nucleocapsid	MHV/S/3239-17/1963	Unknown (MHV/Penn)	GENECONV, Maxchi, SiScan
RaCoV/Parker	8,155	8,834	1ab (nsp3 and 4)	MHV/S/3239-17/1963	Unknown (MHV/Penn)	Maxchi, Chimaera, SiScan
	*27,758	28,254	Spike and ns4	MHV/S/3239-17/1963	Unknown (MHV/JHM)	GENECONV, Maxchi, SiScan, 3Seq
	30,790	30,887	Nucleocapsid	MHV/S/3239-17/1963	Unknown (MHV/Penn)	GENECONV, Maxchi, SiScan
BCoV/Quebec/1972	*284	1,992	1ab (nsp 1 and 2)	HCoV/OC43/1967	PHEV/VW/572	RDP, GENECONV, Maxchi, Chimaera, SiScan, 3Seq
	12,187	14,865	1ab (nsp 7, 8, 9, 10, and 11/12)	HCoV/OC43/1967	Unknown (PHEV/VW/572)	RDP, Bootscan, Maxchi, Chimaera
EqCoV/NC99	6,907	12,648	1ab (nsp 3, 4, 5, 6, 7, and 8)	Unknown (PHEV/VW/572)	GiCoV/US/OH3/2003	RDP, GENECONV, Maxchi, Chimaera, SiScan, 3Seq
	14,574	*16,983	1ab (nsp 11/12, and 13)	Unknown (HCoV/OC43)	BCoV/Quebec/1972	RDP, GENECONV, Maxchi, Chimaera, SiScan
GiCoV/US/OH3/2003	12,187	14,865	1ab (nsp 7, 8, 9, 10, and 11/12)	HCoV/OC43	Unknown (PHEV/VW/572)	RDP, Bootscan, Maxchi, Chimaera
HuEnCoV/4408	*284	1,992	1ab (nsp 1 and 2)	HCoV/OC43/1967	PHEV/VW/572	RDP, GENECONV, Maxchi, Chimaera,

						SiScan, 3Seq
	12,187	14,865	1ab (nsp 7, 8, 9, 10, and 11/12)	HCoV/OC43/1967	Unknown (PHEV/VW/572)	RDP, Bootscan, Maxchi, Chimaera
PHEV/VW/572	24414	25394	Spike	BCoV/Quebec/1972	Unknown (EqCoV/NC99)	RDP, Chimaera, Siscan

\* The actual breakpoint position is undetermined. Most likely it was overprinted by a subsequent recombination event.

<sup>a</sup> Only transferred gene fragments detected by more than one method are included in this table.

<sup>b</sup> Genes indicate the coding sequences contained within the fragment introduced by recombination.

<sup>c</sup> Major Sequence = the sequence most closely related to the sequence surrounding the transferred fragment.

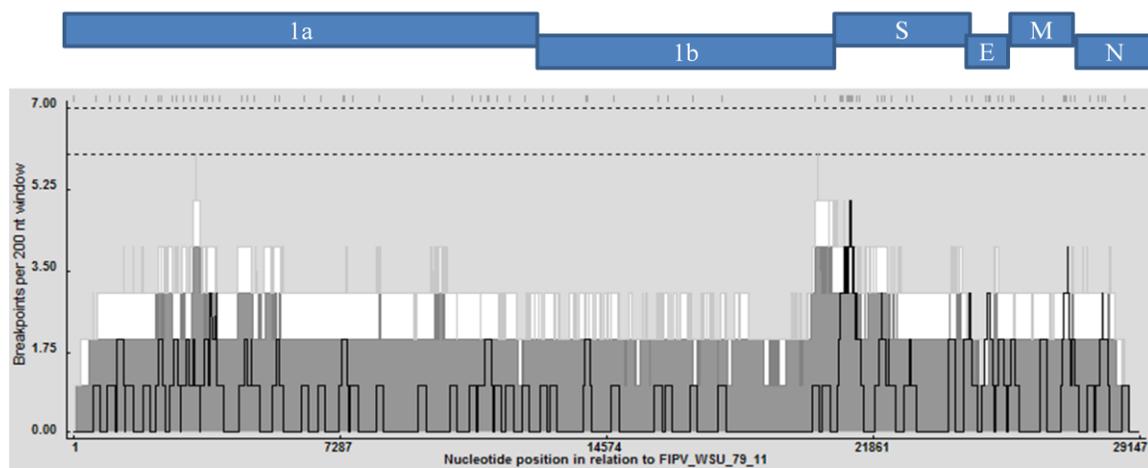
<sup>d</sup> Minor Sequence = the sequence most closely related to the transferred fragment in the recombinant.

<sup>e</sup> Unknown = only one parent and a recombinant need be in the alignment for a transferred fragment to be detectable. The sequence listed in parentheses was used to infer the existence of a missing parental sequence.

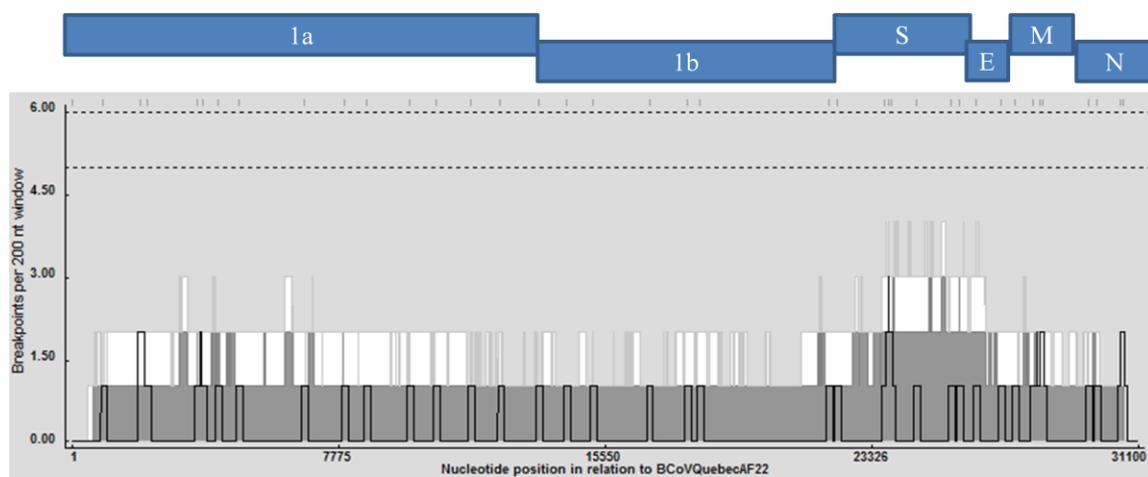
<sup>f</sup> Betacoronavirus nsp locations are based on GenBank accession # NC\_006852 (MHV/JHM complete genome)

**Figure 4.5:** Recombination breakpoint distribution plot generated for (a) alpha- and (b and c) betacoronaviruses as implemented in recombination detections program 4 (RDP4). This plot shows only the detectable breakpoints with both major and minor sequence contributors are known. The plot was constructed using a 200bp window moved 1 nucleotide at a time along the length of the genome. Recombination breakpoint positions are shown as tick-marks along the top of the figure. The dashed lines under the breakpoint positions represent the 99% (upper line) and 95% (lower line) confidence interval thresholds for globally significant breakpoint clusters. Globally significant breakpoint clusters are windows with more breakpoint positions than the maximum found in >95% of the permuted plots. The dark gray and white areas are the 95% and 99% confidence interval thresholds, respectively, for local breakpoint clusters and the black line indicated the breakpoint count within the 200bp window. Locally significant breakpoint clusters are windows with more breakpoints positions than the maximum found in >99% of the windows at that location in the permuted plots. Genome positions of the coronavirus open reading frames, listed across the top of each plot, are approximate.

(a)



(b)



**Table 4.5:** Alpha-(a) and betacoronavirus (b) genomic regions identified as part of a transferred fragment

(a)

Genomic Region	Number of Fragments <sup>a</sup>	% of total
5' UTR <sup>b</sup>	1	1.6
nsp <sup>c</sup> 1	1	1.6
nsp 2	3	4.8
nsp 3	8	12.7
nsp 4	1	1.6
nsp 5	4	6.3
nsp 6	4	6.3
nsp 7	2	3.17
nsp 8	3	4.8
nsp 9	1	1.6
nsp 10	1	1.6
nsp 11/12	8	12.7
nsp 13	9	14.3
nsp 14	7	11.1
nsp 15	7	11.1
nsp 16	9	14.3
Spike	21	33.3
3a	1	1.6
3b	6	9.5
3c	5	7.9
Envelope	2	3.2
Membrane	4	6.3
Nucleocapsid	1	1.6
7a	4	6.3
7b	5	7.9
3' UTR	4	6.3

<sup>a</sup> Genomic areas may fully or only partially located in the transferred fragments.

<sup>b</sup> UTR= untranslated region

<sup>c</sup> nsp = nonstructural protein

(b)

Genomic Region	Number of Fragments <sup>a</sup>	% of total
5' UTR <sup>b</sup>	0	0
nsp <sup>c</sup> 1	4	11.1
nsp 2	4	11.1
nsp 3	6	16.7
nsp 4	4	11.1
nsp 5	3	8.3
nsp 6	1	2.8
nsp 7	4	11.1
nsp 8	4	11.1
nsp 9	3	8.3
nsp 10	3	8.3
nsp 11/12	5	13.9
nsp 13	1	2.8
nsp 14	1	2.8
nsp 15	1	2.8
nsp 16	0	0
ns 2a	0	0
ns 2 <sup>f, j, k</sup>	0	0
32 kDa <sup>h, i</sup>	0	0
HE	3	8.3
Spike	8	22.2
ns 4	2	5.6
5a <sup>d</sup>	0	0
4.9 kDa <sup>f, h, i, k</sup>	0	0
4.7 kDa <sup>f, g</sup>	0	0
12.7 kDa <sup>f, g, h, i, k</sup>	0	0
10.1 kDa <sup>f</sup>	0	0
Envelope	0	0
Membrane	1	2.8
Internal Protein <sup>e</sup>	0	0
Nucleocapsid	5	13.9
3' UTR	0	0

<sup>a</sup> Genomic areas may fully or only partially located in the transferred fragments.

<sup>b</sup> UTR= untranslated region

<sup>c</sup> nsp = nonstructural protein

<sup>d</sup> MHV strains only

<sup>e</sup> RaCoV strains only

<sup>f</sup> BoCoV/Quebec, <sup>g</sup> EqCoV/NC99, <sup>h</sup> GiCoV/US/OH3/2003, <sup>i</sup> HuEnCoV/4408, <sup>j</sup> HCoV/OC43, <sup>k</sup> PHEV/VW/572

## CHAPTER 5

### SUMMARY AND FINAL ANALYSIS

#### **Summary**

In this study, three alpha coronavirus, three betacoronavirus and eight gammacoronavirus genomes were sequenced, annotated and along with other genomes available from GenBank were analyzed for recombination.

Evidence for recombination was found for almost all analyzed coronavirus genomes. The alphacoronavirus with the highest number of recombination sites is a feline virus (FCoV/Black/1970) with 14 transferred fragments contrasted by the alphacoronaviruses with the fewest number of recombination for both the alphacoronavirus strain BtCoV/HKU8/AFCD77 and the betacoronavirus strains, GiCoV/US/OH3/2003, HCoV/OC43/1967, and PHEV/VW/572, with only one transferred fragment. The betacoronavirus with the greatest number of detected recombined sequence fragments events is the MHV/JHM strain with 7 transferred fragments. The gammacoronaviruses with the most transferred fragments are CAV/56b/91 and Mass/H52 both with 8 fragments, and CK/CH/LSD/051/06 and GA98/0470/98 both with 7 fragments. The strains with the fewest transferred fragments are Iowa/Iowa97/56 and TW/2575/98 with only 2 transferred fragments and the CK/CH/BJ/97, Holte/Holte/54, and NGA/A116E7/06 strains with only 1 transferred fragment.

Identical recombination fragments were observed in alpha- beta- and gammacoronavirus but were limited to the viruses contained in those genera. Although our analysis did not reveal

any evidence for cross-CoV recombination, it does suggest the evolution of these viruses is intimately linked to reticulate evolutionary events such as recombination.

A high number of sequence fragments transferred by recombination were observed in the nsp 3 and spike genes for all the coronaviruses. This may imply the involvement of these genomic areas in the evolution and emergence of new coronavirus strains.

This study is the first to provide evidence for that no recombination is occurring between the alpha- and betacoronaviruses. As evidence of alpha- and betacoronaviruses isolated from the same host has been described [1], this lack of recombination may be due to either low sequence identity between the viruses or varied tissue tropism.

Evidence was obtained that recombination is occurring among coronaviruses across their entire genome. Almost every sequence included in the analysis was recognized as a potential recipient of horizontally acquired sequences at some point in its viral evolutionary past. Our data suggests that reticulate evolution due to a high frequency of recombination in all coronaviruses, likely plays a major role in the generation of new serotypes, biotypes, and pathotypes of the virus. The characterization, distribution and frequency of recombination breakpoints are important information that will further our understanding of the mechanisms behind the diversity and evolution of these viruses so better control methods can be developed. If we know which viruses are most likely to recombine and where the viruses are most likely to recombine, we should be able to predict the changes that are likely to occur in the viral genome; potentially affecting pathology and tropism.

**Analysis: Direction of future research**

*Why are gammacoronaviruses recombining more extensively than either alpha- or betacoronaviruses?*

This study found the gammacoronaviruses to be recombining more extensively than either the alpha- or betacoronaviruses. Live attenuated IBV vaccines have been shown to recombine with virulent field viruses and the resultant viruses have caused outbreaks of disease in chicken flocks [2]. The use of live-attenuated vaccines could provide the opportunity for these viruses to recombine as more than one serotype may be infecting a cell. Studies investigating the recombination of these vaccine viruses and field viruses have been conducted *in ovo* [3], but did not focus on the full-genome or the extent of recombination. Using IBV as a model, commercial high titer vaccines will be used to infect embryonating eggs. A group of eggs will be co-infected with a field strain virus not only from a different serogroup but also one that has a different tissue tropism, i.e. a nephropathogenic strain of IBV. The progeny viruses will be re-isolated from the allantoic fluid, plaque purified, and the full length genome sequenced and examined for recombination by comparison with the full length genomes of the parent viruses. *In ovo* studies such as this allow a glimpse into the recombinational opportunities created by live-attenuated viruses and is an important step in the creation of a vaccine against not only SARS but all coronaviruses.

*What is preventing the alpha- and beta coronaviruses from recombining?*

No cross-genera recombination between the alpha and betacoronaviruses was found in this analysis. The possibility of cross-genera recombination is highlighted by the isolation of both an alpha-and a betacoronavirus from several co-infected bats in both Hong Kong and China

[1]. However, evidence of recombination occurring between these viruses was not assessed and the co-infection was inferred from the isolation of both viruses from fecal samples. Although, in principle, inter-genera recombination is possible, the only example of a virus emerging from cross-genera recombination is the SARS-CoV. Reverse genetics experiments have shown the spike protein functioning as a determinant for host cell specificity. In one study, an MHV chimera was produced with the ectodomain portion of the FCoV spike protein gene causing the virus to infect feline cell culture and completely lose the ability to infect murine cells [4]. Therefore, the barrier to replication between the coronavirus genera may lie in the composition of the spike protein. The receptor binding domain (RBD) of the spike protein is thought to be a determinant of coronavirus host range [5-7]. A study by McRoy, *et al*, found mutations in the S2 domain of a mutant MHV that allowed the virus to utilize the human-CEACAM receptor rather than the murine CEACAM receptor [8] effectively extending the host range of this virus. Recently, a bat SARS-like CoV (Bat-SCoV) was reconstructed synthetically and the RBD domain replaced with the human equivalent. This allowed the generation of infectious progeny and a recombinant human virus in which the 3' 5,700 nts of the genome, including the spike coding sequence 3' of the RBD, were replaced with those from the Bat-SCoV [9]. Both the Bat-SCoV and the recombinant were infectious in primate cells illustrating the plasticity and possible modularity of the region. Substitutions of entire functional cassettes of S1 and S2 play a pivotal role in the host range expansion of CoVs and thus the ability of the viruses to recombine. Reverse genetics studies where various portions of the RBD from one virus are exchanged with a virus from a different genera (alphacoronavirus RBD replaced with a betacoronavirus RBD) and then examined *in vitro* for host range expansion will provide insights into the recombination potential that exists between the coronavirus genera.

*How is recombination in nsp3, spike, or both an adaptive advantage to the evolution of coronaviruses?*

This study identified two areas of the coronavirus genome that are “hot spots” of recombination; the nsp3 and spike coding regions. Recombination in the spike gene of coronaviruses has been well documented and has led to the emergence of new viral strains and pathotypes [10-17]. However, the effects of changes in the nsp 3 region of the genome are just beginning to be elucidated. A recent study by Phillips, et al, analyzed the consensus full-length genome for pathogenic and attenuated viruses for three different IBV types. They showed that within a virus type, more amino acid changes between the pathotypes occurred in nsp 3 and not spike; although spike had the highest number of differences between the viruses [18]. Taking the data generated here as to the specific areas of the nsp 3 coding region that had evidence of recombination, chimeric viruses experimentally modeling cross-species recombination will be studied to determine the extent recombination in the nsp3 region alters virus pathogenicity by reducing the replication efficiency of the virus. Targeted RNA recombination, using the method outlined by deHaan, et al, [19] or the use reverse genetics to study the attenuation and replication changes associated with recombination in the nsp3 regions between both closely related and distant species (such as FCoV and CCoV, and FCoV and HCoV-NL63, respectively) will provide insight into the development of effective antivirals that target this protein. Knowing how recombination in nsp 3 coding region affects viral replication and pathogenicity is vital to our understanding of coronavirus evolution.

*Is adaptive evolution occurring in the nsp3 and/or spike region of the coronavirus genome?*

A well known example of a coronavirus crossing the species barrier is seen with BCoV and HCoV43 [20] and animal-to-human transmission of SARS-CoV is thought to be the cause of the SARS outbreak. The spike protein is one of the best characterized of the coronavirus proteins and may play a key role in overcoming the species barrier, and therefore a target for selective pressure. Although the process of adaptive evolution of the SARS spike protein and the exact positively selected sites associated with this process was recently studied by Zhang, et al [21]; this type of study has not been conducted with any other coronavirus or other region of the genome. If we are to understand the evolution of coronaviruses, the adaptive evolution of key coronavirus proteins such as nsp3 and spike must be determined. Determination of selective pressure begins with the measurement of the nonsynonymous to synonymous substitution rates in protein coding genes. This ratio ( $\omega=d_N/d_S$ ) provides a straightforward measurement of selective pressure at the protein level. Implementation of the program Codeml in the PAML software package can be used to investigate the adaptive evolution of both the nsp3 and spike proteins [22]. This program allows for the application of codon substitution models such as M0 (one ratio), M1a (NearlyNeutral), M2a (PositiveSelection), M3 (discrete), M7 (beta) and M8 (beta and  $\omega$ ). The protein regions would be aligned, the models of codon substitution applied, and then using the likelihood ratio test to detect the presence of positively selected sites. A  $\chi^2$  analysis is used to determine the significance of the results.

## References

1. Lau, S.K.P., et al., *Ecoepidemiology and Complete Genome Comparison of Different Strains of Severe Acute Respiratory Syndrome-Related Rhinolophus Bat Coronavirus in China Reveal Bats as a Reservoir for Acute, Self-Limiting Infection That Allows Recombination Events*. J. Virol., 2010. **84**(6): p. 2808-2819.
2. Wang, L., D. Junker, and E.W. Collisson, *Evidence of natural recombination within the S1 gene of infectious bronchitis virus*. Virology, 1993. **192**(2): p. 710-6.
3. Estevez, C., P. Villegas, and J. El-Attrache, *A recombination event, induced in ovo, between a low passage infectious bronchitis virus field isolate and a highly embryo adapted vaccine strain*. Avian Dis, 2003. **47**(4): p. 1282-90.
4. Kuo, L., et al., *Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier*. J Virol, 2000. **74**(3): p. 1393-406.
5. Baric, R.S., et al., *Episodic evolution mediates interspecies transfer of a murine coronavirus*. J Virol, 1997. **71**(3): p. 1946-55.
6. Ren, W., et al., *Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis*. J Gen Virol, 2006. **87**(Pt 11): p. 3355-9.
7. Schickli, J.H., et al., *The N-terminal region of the murine coronavirus spike glycoprotein is associated with the extended host range of viruses from persistently infected murine cells*. J Virol, 2004. **78**(17): p. 9073-83.
8. McRoy, W.C. and R.S. Baric, *Amino acid substitutions in the S2 subunit of mouse hepatitis virus variant V51 encode determinants of host range expansion*. J Virol, 2008. **82**(3): p. 1414-24.
9. Becker, M.M., et al., *Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice*. Proc Natl Acad Sci U S A, 2008. **105**(50): p. 19944-9.
10. Lee, C.W. and M.W. Jackwood, *Evidence of genetic diversity generated by recombination among avian coronavirus IBV*. Archives of Virology, 2000. **145**(10): p. 2135-2148.
11. Decaro, N., et al., *Recombinant Canine Coronaviruses Related to Transmissible Gastroenteritis Virus of Swine*. Journal of Virology, 2009. **February**: p. 1532-1537.
12. Decaro, N., et al., *Recombinant canine coronaviruses related to transmissible gastroenteritis virus of Swine are circulating in dogs*. J Virol, 2009. **83**(3): p. 1532-7.

13. Herrewegh, A.A., et al., *Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus*. J Virol, 1998. **72**(5): p. 4508-14.
14. Lee, C.-W. and M.W. Jackwood, *Origin and evolution of Georgia 98 (GA98), a new serotype of avian infectious bronchitis virus*. Virus Research, 2001. **80**(1-2): p. 33-39.
15. Lim, T.-H., et al., *An emerging recombinant cluster of nephropathogenic strains of avian infectious bronchitis virus in Korea*. Infection, Genetics and Evolution, 2011. **11**(3): p. 678-685.
16. Wang, Y.Y. and C.P. Lu, *Analysis of putative recombination hot sites in the S gene of canine coronaviruses*. Acta Virol, 2009. **53**(2): p. 111-20.
17. Fang, S.G., et al., *Selection of and recombination between minor variants lead to the adaptation of an avian coronavirus to primate cells*. Biochem Biophys Res Commun, 2005. **336**(2): p. 417-23.
18. Phillips, J.E., et al., *Changes in nonstructural protein 3 are associated with attenuation in avian coronavirus infectious bronchitis virus*. Virus Genes, 2011.
19. de Haan, C.A., et al., *Manipulation of the coronavirus genome using targeted RNA recombination with interspecies chimeric coronaviruses*. Methods Mol Biol, 2008. **454**: p. 229-36.
20. Vijgen, L., et al., *Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43*. J Virol, 2006. **80**(14): p. 7270-4.
21. Zhang, C., J. Wei, and S. He, *Adaptive evolution of the spike gene of SARS coronavirus: changes in positively selected sites in different epidemic groups*. BMC Microbiol, 2006. **6**(1): p. 88 - 97.
22. Yang, Z., *PAML: a program package for phylogenetic analysis by maximum likelihood*. Comput Appl Biosci, 1997. **13**: p. 555 - 556.

## APPENDIX A

## CHAPTER 3 SUPPLEMENTAL TABLES





Conn/ Conn 46/91	93. 9	93. 4	93. 5	92. .7	93. 9	99. 6	89. 9	95. 4	88. .0	88. .2	89. 1	88. .8	88. 5	92.9	92. .9	93. .0	92.9	86. 0	92. 9	86.3	48.4	9	89	89	89	89	52.8	92. 9	49.2	48.7	47.6	92. .4	
Conn/ Conn 46/66	93. 9	93. 4	93. 6	92. .7	93. 9	99. 6	89. 9	95. 5	88. .0	88. .1	89. 0	88. .8	88. 5	92.9	95. .9	93. .1	92.9	86. 0	92. 9	86.3	48.4	9	89	89	89	89	52.8	93. 0	49.2	48.8	47.6	92. .4	99. 8
CAL/ CAL9 9/99	92. 5	92. 1	91. 7	93. .1	92. 5	94. 7	90. 9	95. 2	87. .5	87. .3	88. 1	87. .3	88. 2	90.8	90. .7	90. .7	90.8	85. 7	90. 7	85.8	48.1	9	88	89	89	88	52.8	90. 8	49.2	48.8	47.5	92. .8	94. 9
Ark/A rk- DPI p101/ 81	94. 2	92. 9	92. 4	94. .9	88. 9	94. 6	92. 8	95. 2	87. .9	87. .7	88. 3	87. .9	89. 1	91.9	91. .3	91. .4	91.9	86. 1	91. 4	86.2	48.3	9	89	89	90	88	52.9	91. 4	48.6	48.6	47.4	94. .1	94. 7
CK/C H/ZJ7 91/97	92. 2	92. 4	92. 0	99. .8	92. 1	92. 7	94. 8	92. 8	87. .6	86. .6	87. 0	86. .8	89. 3	93.6	92. .4	92. .5	93.6	85. 9	92. 4	86.3	48.2	9	87	88	88	87	52.8	92. 3	49.2	48.8	47.5	96. .8	92. 8
CK/C H/SC 02120 2/02	88. 2	87. 9	88. 0	88. .1	88. 2	88. 2	86. 2	87. 9	88. .4	83. .9	84. 1	83. .6	94. 6	88.1	88. .0	88. .1	88.1	86. 9	88. 0	88.7	48.0	8	83	84	84	84	52.5	88. 1	48.9	48.8	47.5	88. .1	88. 2
ITA/9 0254/ 05	90. 4	90. 1	90. 1	90. .1	90. 4	90. 2	88. 2	89. 8	87. .7	85. .9	86. 1	86. .0	88. 9	89.6	89. .6	89. .7	89.6	87. 6	89. 6	87.6	48.3	9	86	86	86	86	52.9	89. 5	49.3	48.7	47.5	89. .6	90. 2
NGA/ A116 E7/06	91. 3	90. 8	90. 8	91. .3	91. 3	91. 2	89. 2	91. 0	87. .4	86. .5	86. 9	86. .5	88. 3	90.2	89. .9	90. .0	90.2	85. 9	89. 9	85.6	48.0	9	86	86	87	87	52.4	90. 0	49.1	48.7	47.6	90. .4	91. 2
CK/C H/DY 07/07	86. 7	86. 9	86. 9	86. .5	86. 7	86. 6	84. 7	86. 3	86. .7	82. .5	82. 7	82. .4	89. 2	86.6	86. .5	86. .5	86.6	90. 7	86. 5	91.3	48.2	8	82	82	83	82	52.7	86. 6	48.8	48.6	47.3	86. .3	86. 6
CK/C H/CQ 04- 1/04	87. 8	87. 6	87. 6	87. .6	87. 7	87. 6	85. 7	87. 3	87. .8	83. .6	83. 6	83. .3	91. 9	87.8	87. .5	87. .6	87.7	88. 6	87. 6	89.9	48.1	8	83	83	83	83	52.6	87. 7	49.0	48.8	47.5	87. .5	87. 6
GA98 /GA9 8/98	90. 6	90. 2	89. 8	91. .9	90. 6	92. 2	95. 8	93. 3	85. .9	87. .1	87. 6	87. .2	87. 0	89.2	88. .9	88. .9	89.2	84. 0	88. 9	84.2	48.1	9	88	89	89	88	52.5	89. 1	49.1	48.5	47.4	91. .2	92. 3

Supplemental Table 3.2: Primer sequences used in this study

**CAV/56b/91 primers**

Name	Sequence
1F	5'-CCAACCTTACAAAACGGACTT-3'
1R	5'-AGTAGAGTCAAAAATTTTCGTCAAG-3'
2F	5'-CGCTTGACGAAATTTTTGACTCTA-3'
2R	5'-TGC GTTCCCTTGGCGTTACC-3'
3F	5'-TGGCGATGAAGTTTGGTTTGACG-3'
3R	5'-CAACAGCCTCACCCAGTCACC-3'
4F	5'-GTGCACCATACGCGTTCTTTTA-3'
4R	5'-TTGGCAACATCTTCTTTTACTACT-3'
5F	5'-GGCAATATGGGTAGACGGCAGTG-3'
5R	5'-GGACCACATAAAGAACCCTCAAAT-3'
6F	5'-TGCACGTGTACACTAACTCTGGT-3'
6R	5'-CTCTTCGACACAGGTTCCCTATT-3'
7F	5'-CCAAATGGTGTTAGGCTTAT-3'
7R	5'-TGCAAAGATTAGTAAACAACAAC-3'
8F	5'-AAGGCTGGTATGTTTAAGGAAGGT-3'
8R	5'-TGTCCCAACCGCCATAAAACTT-3'
9F	5'-TGTTGTGGATTTACTGCTTGATGA-3'
9R	5'-GATCAAACCTGTGCTCTGTCTTAT-3'
10F	5'-TCAATCAGCATGGACGTGTGGTTA-3'
10R	5'-TTTGC GTGCAGTGTTTCCATT-3'
11F	5'-TGACAGAGACAAGTTGGCACGAGA-3'
11R	5'-TCATAGCAGAAATACGAA-3'
12F	5'-GTCCGTTAACAGGTTTCATTCA-3'
12R	5'-TAAACAAGCAACCCACATTCAAAA-3'
13F	5'-TGGTCTCTGGTTTAATTCAC TTTC-3'
13R	5'-GAAATATTTAAAATCGCCAGTGCTA-3'
14F	5'-ATGCAGGAAGGTTTTAGGAGTA-3'
14R	5'-GGATGAAGATAATAGTGGCAAAAAG-3'
15F	5'-TCAGTTTCGATTTACAGCACAT-3'
15R	5'-ACCATTTTGAGGAGTATTGAACC-3'
16F	5'-GGTAGCGGTGTTCCCTGATAATG-3'
16R	5'-CGCCCATCCTTAATACCTTCCTCA-3'
SP1 R	5'-GAGACCACTGGCATGCTGTTTTCA-3'
SP2 R	5'-GTC ACTGTCTATTGTATGTCTGCTC-3'
SPR	5'-GCCGACCTTGTGCCGAGAA CG-3'
5'seq R	5'-TACCGCTAGATGAACCAGAGG-3'
G1F	5'-TCG ATT GCC GCT ATG AAA AGT AA-3'
G1R	5'-GGC TTT GAC CTC TCG CAT AAC TT-3'

G2F	5'-TGTTTGTGCTTTAGCGTCTGGTAT-3'
G2R	5'-AAT TTGGCATGGCTCTATCACAC-3'

**DE/DE072/92 Primers**

Name	Sequence
1F	5'-AGAGGCAGCTGTGAGAGTTGTGGA-3'
1R	5'-GAACAATTTTATGCACACCTCCTT-3'
2F	5'-GGTTTGACGCCATGATAGTGTTG-3'
2R	5'-ATCAACAGCCTCACCCAGTCACC-3'
3F	5'-AGAATGTGAAGAAGAGGATGAGGA-3'
3R	5'-TCTGCGTCAAATGTTCTGCTAAA-3'
4F	5'-TGCACCATACGCGTTCTTTTGT-3'
4R	5'-TTGCGAACTACACCACCACCTACT-3'
5F	5'-TTAACAGCCAGCAATACACCTCA-3'
5R	5'-TCCATTGGTACCTGAAAACCTTAC-3'
6F	5'-TTCTCGCTGCTAAATGGTTGG-3'
6R	5'-AACAAATTATCCATGCACTCTCCA-3'
7F	5'-GTTGTTTTAATGCAGCTTTTGACT-3'
7R	5'-CACGTTTCTGGGTCTGGTATTA-3'
8F	5'-AAGAGGCGCGTGTAACAGATAGAC-3'
8R	5'-TCACATCCACTAGCAAGGGGTATC-3'
9F	5'-AAACGGGTACGGGGTAGCAGTGAG-3'
9R	5'-TTTCATAATAGAGGCGGGCTTTTC-3'
10F	5'-TGGTGATGCTACTACTGCTTATGC-3'
10R	5'-CTATTCAATGGCGGCCTGGTTTT-3'
11F	5'-CGGGAAGCAACATTCATTTACC-3'
11R	5'-AAACCCAAACAATGCTTCCAACAA-3'
12F	5'-TTCTTGCTGCTGATAATGCTGTTC-3'
12R	5'-GATCAAACCTTGTGCTCTGTATTAT-3'
13F	5'-AAAAAGAAAGCATGAAGGCGTGAT-3'
13R	5'-TCACAGAAGAAATAGCACCAAAT-3'
14F	5'-AAGAGCTATTACTGCAGGAGATGT-3'
14R	5'-CTTCCATTCTCCTCTAGCGACTTA-3'
15F	5'-CGACTTTTGATAACGATGTGGTAA-3'
15R	5'-AGTAAACACAATCTAATCCTTCTC-3'
Gap 1F	5'-GATGTTCTAGGTGACTGGGGTGAG-3'
Gap 1R	5'-(N)TGAGAAGAAGTTGGACTGC-3'
Gap 2F	5'-AAAGAAGAAGTTGTGTCCCAGAAA-3'
Gap 2R	5'-TCTGCGTCAAATGTTCTGCTAAA-3'
Gap 3F	5'-GCGCAGAAATGGGATGTTC-3'
Gap 3R	5'-TTGCGAACTACACCACCACCTACT-3'
Gap 4F	5'-CCGGTGTTTTCTGTGGTTCTACTG-3'
Gap 4R	5'-TCCATTGGTACCTGAAAACCTTAC-3'

Gap 5F	5'-TCAGATTGGTGGTGTAGATTA-3'
Gap 5R	5'-ATACAACATCCATTTAGCACACT-3'
Gap 6AF	5'-GCACAAGATAAACCCACCTAAAAC-3'
Gap 6AR	5'-GCTTGATCTACACCTGCTACG-3'
Gap 6BF	5'-CAAATCCAAATCTGAAAGTAGC-3'
Gap 6BR	5'-TCACATCCACTAGCAAGGGGTATC-3'
Gap 7AF	5'-CCTATTGTACGACGTGCTTTATTG-3'
Gap 7AR	5'-AGCAGCATTACCAGTTTGAGGATA-3'
Gap 7BF	5'-GCTACTTGTGGCTATCATTCTAA-3'
Gap 7BR	5'-TTTCATAATAGAGGCGGGCTTTTC-3'
Gap 8F	5'-AACTAACTACTTTGCGCTATTTTG-3'
Gap 8R	5'-AAACCCAAACAATGCTTCCAACAA-3'
Gap 9AF	5'-AAAAAGAAAGCATGAAGGCGTGAT-3'
Gap 9AR	5'-AAGAAGCATTAACAACACGACCAC-3'
Gap 9BF	5'-GGTGGTCGTGTTGTTAATGCTTCT-3'
Gap 9BR	5'-TTGAAACTGTTAGGTATGAGCAC-3'
Gap 10F	5'-AAGTCGCTAGAGGAGAATGGAAGT-3'
Gap 10R	5'-AGACTACCTCCTCCTGTTGCTAC-3'
072-1AF	5'-ACT GGG GTG AGG CTG TTG ATG-3'
072-1AR	5'-AAA AGA ACG CGT ATG GCA CAA CCT-3'
072-1BF	TTT TGT GGA CCG GAC TTT GTT GAA-3'
072-2F	5'-ATG GGC TAA ACT GTT GGG TGG AGA-3'
072-2R	5'-GCC CAA CAA CAA AAT TAG CAC TG-3'
072-3F	5'-CAA TGG TTA TAA AGT TTC AAG GTG-3'
072-3R	5'-ACA CTA ACA ATG CAC TTC TCA ACA-3'
072-4F	5'-TTT TGA AGA CGA ATT GAC ACC AGA-3'
072-4R	5'-TAA CCC ACC ATT ACA ATC CAA AGT-3'
072-5F	5'-CCC TTG TTA TAC CAG ACC CAG AGA-3'
072-5R	5'-TTG GAC TTG GCT TTG ATG TTA TTG-3'
072-6F	5'-AAG AGT TTA GTG GTG TTC ATC C-3'
072-6R	5'-CAA AAT AGC GCA AAG TAG TTA GTT-3'
072-7AF	5'-TTG GTG CAA GTG AAA AGG TTA-3'
072-7AR	5'-TTC ACG AGC ACACGC AAG GTC-3'
072-7BF	5'-TTG ACG TTG CTA AGT TTG ATT TGA-3'
072-7BF	5'-CGA CGT GTT CCA TTA GTG ATT TTA-3'
072-S1AF	5'-ACT TGT CGC TGC CGG TCT TTA-3'
072-S1AR	5'-CCG CCC TGT TGG TTG TC-3'
072-S1BF	5'-TCA GGG GGA AAT TTA GTT GGT-3'
072-S1BR	5'-CTT CCC AAT AAT TAC CTC CTC TGT-3'
072-S1F	5'-CAG AAG ATG TTA AAG CAG CAG GTG-3'
072-S1Fa	5'-AGT ACT ATT CCT GAA GGC TAT GTT-3'
072-S1R	5'-ATA CAA AAT TTT CCA TAA GTT-3'
SP1 R	5'-CGTATAGAAAAACAAAGCGTCAC-3'
SP2 R	5'-GTCCTGTCTATTCTATGTCTGCTC-3'
SP3 R	5'-TAGCCGACCTTATGCGAGAACG-3'

5' Seq R	5'-CGG CAA AAC AAG CAG TGA TAC CAG-3'
(SP5) F	5'-GGT GAC CAA GCG GAA ATA AGA AA-3'
3' Seq F	5'-AGT AGC CTG GAA ACG AAC GGT AGA-3'

**FL/FL18288/71 Primers**

<b>Primer Name</b>	<b>Sequence</b>
1F	5'-AAAACGGACTTAAATACCTACAGC-3'
1R	5'-TCAAGAGGAAGAGACCACAC-3'
2F	5'-TAGTGCCAAACAACCCCTGAG-3'
2R	5'-GCATCCTTCGCATACACG-3'
3F	5'-GCTAAGCTCATTGTCACCGAAACC-3'
3R	5'-CAACAGCCTCACCCAGTCACC-3'
4F	5'-AAGACAGTTTTTACAGCCGATGAT-3'
4R	5'-TAGGAATGCGCAGAGACTTAC-3'
5F	5'-CACACTGGAAGTACCTAATGG-3'
5R	5'-GCCGTAAACAAGACAATAGCACAC-3'
6F	5'-AGCCAATGGTATGATCCTGTAGTC-3'
6R	5'-AAATTATCCATGCACTCTCCAACA-3'
7F	5'-GTTGTTTTAATGCAGCTTTTGACT-3'
7R	5'-AGCTTATTACTACAAACAAGTGA-3'
8F	5'-CAGTGTTGGATTGGTTATGGAT-3'
8R	5'-ATCGTTTATAGCAGCATTACCAG-3'
9F	5'-ATCAAACGGTAAAGCCAGGTCCT-3'
9R	5'-CTCCATAACAGCCACAGG-3'
10F	5'-GTGGAACCAGACTTAGAAAAA-3'
10R	5'-TAGCCCTGTAAATTCCAAAAAC-3'
11F	5'-CCTTACTTAATGGTTCCTTTTAC-3'
11R	5'-TTCTACATCAAACCTACCCAACC-3'
12F	5'-GTTTTTCAGCTCTCCAGTCTATCG-3'
12R	5'-AACCACACGTCCATGCTGATTGAA-3'
13F	5'-AAGTTGCTGGTAAGAGATGTTGGT-3'
13R	5'-GCCAGGGCTTTAACTTCTCT-3'
14F	5'-AATGAGACTGGCGCCAACCCTAAT-3'
14R	5'-CAGGCCCAATTGTTGAAA-3'
15F	5'-GCTGCTGATGCTTGTGTTTATTT-3'
15R	5'-TTTAGGCAAGTGGTCTGGTTCA-3'
16F	5'-ACTATTTCTGTGCTTTTCTATCA-3'
16R	5'-TCTACCGTTCGTTTCCAGGCTACT-3'
5' PER	5'-GCACTGCAATGGAACACCTATGAG-3'
3' PER	5'-ATGAGCCTGGAAACGAACGGTAGA-3'
Gap1 F	5'-GACCAAAAAGCTGACATCCCTGTG-3'

Gap1R	5'-AAGGTTTCGGTGACAATGAGC-3'
Gap 2F	5'-TGTGATAGCAGAAGATGTTTGGTC-3'
Gap2R	5'-CAACAGCCCTCACCCACTCACC-3'

### Gray/Gray/60 Primers

Name	Sequence
1AF	5'-CATGCTTGT TTTGCCGTGTCTCA-3'
1AR	5'-GCGCATTGCCATAGCTGAACTT-3'
1BF	5'-GTAGTGCCAAACAACCCCTGAG-3'
1BR	5'-TTGCGAAAATTCTAGAACCATAA-3'
3F	5'-ATGCGAATGACCTAACACTGC-3'
3R	5'-AACCAAATTCATCGCCATCTC-3'
4F	5'-GGTGAACCATGGAATACAATCTT-3'
4R	5'-TTCGCAACGCCTCCACCAT-3'
5F	5'-TACGCGTCCTCTTGTTTCTCTGA-3'
5R	5'-CATATCCATCAA AATTGTCAGGTG-3'
6AF	5'-AGGTTAGCGTGGTTGTTGGTGGAC-3'
6AR	5'-CTGGGTGTGGCATGAAATAACTTG-3'
6BF	5'-CGTTGTGTTGTTTGGGTTGTTATC-3'
6BR	5'-TCCACA ACTGTACCACCATAAGA-3'
7F	5'-TGGTTATGTGGACGAAGAGGTTGC-3'
9F	5'-GCCGATCATGATTTCTTTGTGTTTC-3'
9R	5'-CATTTGAGTTATAGTGGGCAGGAC-3'
10F	5'-CAAAGAAGAATGTCCTGCCCACTA-3'
10R	5'-ATAACAAACAACACCGTCGTCAGA-3'
11F	5'-AGGTGTTCTTTGTGCTTCTTTCAT-3'
11R	5'-GCAGTCGATAGTAGTTCTTTGAGG-3'
12F	5'-AATGGTACCGGAATGTTTTGTGAA-3'
12R	5'-AGTTGAGTTGTGTTGTAGGCTGAA-3'
13F	5'-TTGCAAATAACCCAGAATCACG-3'
13R	5'-TTCCAACAAGCATAAGCCTGAGTA-3'
14F	5'-CAATAGTACCCAACATCAAGCAGT-3'
14R	5'-TAGTGCCATCATCAAACCAAGTCA-3'
16F	5'-GAGGAAGCATGAAGGCGTGATAGC-3'
16R	5'-GAGGAGCGGCACTTTCATTACGA-3'
17F	5'-TCATATAGAGGTCCTTCGTTGTGT-3'
17R	5'-AGACCCTGTATAACGCCTTGAAT-3'
18F	5'-ATTTTGGTGCTATTTCTTCTGTGA-3'
18R	5'-GATGAAGATAATAGGTGGCAA AAG-3'
19F	5'-CTGGTCAGCAATAATTCAACAAAA-3'
19R	5'-GGGGCCAAAAGCACCATAACACTA-3'
20F	5'-GGTGACCAAAGCGGAAATAAGAAA-3'

20R	5'-ACAAAAGCACTCGCCTACTACACG-3'
21F	5'-GGCCAAAGAAGCAGGATGACGA-3'
21R	5'-CCCCGGCACTTGGCATCTT-3'
22F	5'-AGTAGCCTGGAAACGAACGGTAGA-3'
G1F	5'-GCCATTGAAGTTGAGACAGACCT-3'
G1R	5'-AACAAACTCATCCACCTCTTCTGC-3'
G2F	5'-AGCTGCTGGTGTTCCTGGTTTTGT-3'
G2R	5'-AACCGCGAATAACAAAAGTGCTCT-3'
G3F	5'-ACGCTTTGAGGCATTTTGACC-3'
G4R	5'-CTGCAAAAATCATAGAAATCCTT-3'
G5F	5'-GACTCGTCTCAAGGTTTCAG-3'
G5R	5'-CAAAGTCAGCACCAGTAGAGAAGC-3'
G6F	5'-AGTTGTAATGGTGGTAGTCTGTA-3'
G6R	5'-TTCTACAAAATCCTCCTCTGACAT-3'
G7F	5'-AAATTGCCCTTATGTTAGTTATGG-3'
G7R	5'-CACTCTAATGTACTCCGACTGCT-3'
SP1 R	5'-TCCCAGTCTGCAGATTACGGTCAA-3'
SP2 R	5'-AAGCCATGTTGTCAGTGTCTATTG-3'
SP3 R	5'-TGAGACACGGCAAAACAAG-3'
Gray 985-1009 R	5'-GCATTGCCATAGCTGAAACTTGAA-3'

### Mass/H120 Primers

Name	Sequence
5'PE	5'-ACAAGAATGACATCCCGTGGTTT-3'
1 F	5'-GTGTGCACTTGGTAGAATCC-3'
1R	5'-GCCAACACTTTAGTATCCTCATCC-3'
2 F	5'-CTTGGTGACTGGGGAGAG-3'
2R	5'-TGGCAATTACAACAGTCAA-3'
3 F	5'-ACCTCTTTGCCTGTTGCTAA-3'
3R	5'-GAAATAAAACCCTGCTCCCATAAA-3'
4 F	5'-TGGTGGTGTAACTATGGGACTTTG-3'
4R	5'-TGCTCATAAAGTGCCTGGTCAAGT-3'
5 F	5'-TATAAAGCTGCAACTCCTGGTAAG-3'
5R	5'-CATACATGGGGTGAACAAAACCTT-3'
6 F	5'-TGGGTTTTGTTTAATGACGAATAC-3'
6R	5'-GGCCTCAATTTATCACCTATCTC-3'
8 F	5'-TAAGAGTGCAGGCTATCCA-3'
8R	5'-GTAACAAACAACACCGTCGTC-3'
SP1	5'-AAAGAAGGCCAATGTTGTGAAGAT-3'
SP2	5'-GGGACGCCAGGTGTTATTTTGT-3'
SP3	5'-TCCCAGTCTGCAGATTACCTTCAA-3'
9F	5'-AGAAAATGTGCGACGACCTCAAAT-3'

9R	5'-AATAGGCTCCTGTTCAATCTGTG-3'
10F	5'-CACCTTACAATGCTATGAACCAGA-3'
10R	5'-AGTATCTACGAGCCCTTCTGGTGT-3'
11F	5'-CTGTGTTTGC GTGTCTTTCTTTTG-3'
11R	5'-CGCGGGACAAGGCTCTGCT-3'
Gap 1F	5'-AGCGGTGTAGGTCATGGTGGTCA-3'
Gap 1R	5'-CGCGGGACAAGGCTCTGCT-3'
5'1R	5'-AAGTCCGTTTTGTAAAGTTGAA-3'
5'2R	5'-GCACTGCAATGGAACACCTATGG-3'
5'3R	5'-CCTCAGGGGTTGTTTGGCACTAC-3'
5'4R	5'-AGTAGAGTCAAAAAATTTTCGTCAAG-3'
3'seq F	5'-TGAAGTAGATAAGGCATTGACC-3'
3'seq R	5'-GATTAGACATTCCCTGGCGATAG-3'

### Holte/Holte/54 Primers

Primer Name	Sequence
HOLTE Gap 2 R	5'- GCCCTCTCCAAGCACTGTCTG - 3'
HOLTE Gap 3F	5'- CATATGGAAAAGTGCGTAATCTC- 3'
Holte Gap 3R	5'- TGTAGTTCTTTCCGTCATCC- 3'
HOLTE Gap 5F	5'- GAAACATGGTCCACAGCAAAAAC- 3'
Holte Gap 5R	5'- GCAAAAAGCCAATTAGCATCTGA- 3'
HOLTE Gap 6F	5'- GCTTGCGCAGAAATGGGATGT- 3'
Holte Gap 6R	5'- TGTTGCTATCATACCACTGTTCA- 3'
HOLTE Gap 7A F	5'- TGACACAGACTGATCATAGACCTT- 3'
Holte Gap 7A R	5'- CCCAGATAGCAACAAGCCAACCA- 3'
HOLTE Gap 7B F	5'- TGGGTTATTAATGCATTTGTTTTG- 3'
Holte Gap 7B R	5'- ATAGAATCACCCAGCCACAGTCCA- 3'
HOLTE Gap 8F	5'- TTGCCAATGCTGAAACTCCTAAG- 3'
Holte Gap 8R	5'- TCTGGTGTCAATTCGTCTTCAAAA- 3'
HOLTE Gap 9F	5'- AAACCTCCTTCGCACTATTATGGTA- 3'
Holte Gap 9R	5'- AACTCCCAATTACCTTCTGT- 3'
HOLTE Gap 11AF	5'- CTCAAGAGTTTTTCGCACATACCC- 3'
Holte Gap 11AR	5'- CCATCGGCATCAGTAACACAGTCT- 3'
HOLTE Gap 11BF	5'- CCCTTGTTATAACCAGACCCAGAGA- 3'
Holte Gap 11BR	5'- GCCTACAGCATCCACTTCCTC- 3'
HOLTE Gap 12F	5'- TGGGTGATAAGGTTGAAGTTGTTT- 3'
Holte Gap 12R	5'- CCGGCTGATTCCCTTATTACAAACA- 3'
HOLTE Gap 13F	5'- CCCTAAGTGGTTTGAAGAGAATAA- 3'
Holte Gap 13R	5'- CTGCAAAATCGTAGAAATCCT- 3'
HOLTE Gap 14F	5'- ACCACTTAAACATTTCTTCTACCC- 3'
HOLTE Gap 15F	5'- CAACACATTAGCCAACAAGGTCT- 3'
Holte Gap 15R	5'- TTGCCAAAATATAAGGTTCCACAG- 3'
HOLTE Gap 16F	5'- TGTGTTGTAAGTGTGCTATGACC- 3'

Holte Gap 16R	5'- GGCCGCCAATCCTATAGCAAAATG- 3'
Holte Gap 16B F	5'- CCG CGT ACT TTG CTT AAT GGT TCA- 3'
Holte Gap 16B R	5'- TTG CGA ATT GCC TCA TCA CG- 3'
HOLTE Gap 17A F	5'- TGAAAAGGGTGAAGGTAAGGAC- 3'
Holte Gap 17A R	5'- AACTATAACCTTGAAACACTGACG- 3'
HOLTE Gap 17B F	5'- ATCTTATTTCTTTGTTAGGGTTTA- 3'
Holte Gap 17C F	5'- AAC AAA CAT GCA TTC CAC AC- 3'
Holte Gap 17C R	5'- GGG CTT ATC AAC TTC TCC AT- 3'
Holte Gap 18F	5'- TGT TGT AGA TTT ATT GCT TGA TGA- 3'
Holte Gap 18R	5'- CAA CAA AAG TGA GCT GGG TAA TA- 3'
HOLTE Gap 19F	5'- CTAAGCAGGCGGAGCATA- 3'
Holte Gap 19R	5'- ATAACGCCTTGAATACGAT- 3'
HOLTE Gap 21F	5'- TGTCATGGCAAGCGGTAAG- 3'
Holte Gap 21R	5'- CTCCCACTCCTGCCACGATTCA- 3'
HOLTE Gap 22 F	5'- GCT ACT TGT GGC TAT CAT TCT AA- 3'
Holte Gap 22R	5'- TGG TAT ACA GCC GCC TTC AT- 3'
HOLTE Gap 23F	5'- AGAAAGCATGAAGGCGTGATAGC- 3'
Holte Gap 23R	5'- GCA GAT GCT AAA ACA GAA AGT GAT- 3'
HOLTE Gap 24F	5'- AGG CGT TAT ACA GGG TCT TA- 3'
Holte Gap 24R	5'- AGG CGT TAT ACA GGG TCT TA- 3'
HOLTE Gap 25F	5'- TCA GTA TGG CTA TGC AAC AAG AA- 3'
Holte Gap 25R	5'- GGC TTT GGT CCT CCT AGT TTG ATG- 3'
HOLTE 3'GAP F	5'- TTTTCGGAAGTAGAGTGACG- 3'
Holte Gap 3' R	5'- ATTAGACATTTCCCTGGCGATAG- 3'
H-1F	5'- GCTACTTG GGCTATCATTCTAA- 3'
H-1R	5'- TGGTATACAGCCGCTTCAT- 3'
H-2R	5'- GCAGATGCTAAAACAGAAAGTGAT- 3'
H-3F	5'- AGGCGTTATACAGGGTCTTA- 3'
H-4F	5'- TCAGTATGGCTATGCAACAAGAA- 3'
H-4R	5'- GGCTTTGGTCCTCCTAGTTTGATG- 3'
h 5'F	5'- GCACCTGGCCACCTGTCA- 3'
h-5'R	5'- GTTTCCAGTCTGCAGATTACGAT- 3'
H-5F	5'- CAGCGCGTGTGAAAGTAGAAGATG- 3'
H-5R	5'- ATTTCTTGCAACAAGTAGTTTCTCCA- 3'
H-6F	5'- ACGGGGGTGCTTATGCGGTAGT- 3'
H-6R	5'- CAGGCCCATATTGTTGAAA- 3'
H-7F	5'- AGCGCTCCAACAATAACAAG- 3'
H-7R	5'- ACATTGTTGAACATTAGTTAGGAG- 3'
H-8F	5'- TGAACCAGACCACTTGCCTAAA- 3'
H-8R	5'- TGCCATGACAAAAGATTATTATTA- 3'
H-G1F	5'- CAGGTTGTGGTGAGGCAGATGTT- 3'
H-G1R	5'- CAAGGCTATGGCAAAATGTGA- 3'
H-G2F	5'- AACCTTGCAACATTCCTAACTAT- 3'
H-G3F	5'- GTTTTGTACAGCAGTGAATG- 3'
H-G3R	5'- TCGCAACGGATGAAGCACTAAAAT- 3'

H-G4F	5' - TGTTTCAGGAGCAGGTGTTTCATTT- 3'
H-G4Ra	5' - CTGTTTGTATACGAGAGCCATCAC- 3'
H-G4Rb	5' - TATAGTAGAAATCGAACCATCAGG- 3'
H-G5F	5' - TCCTCTTTGTTTTATACTCTCCTT- 3'
H-G5R	5' - TTACCGCTCGCCATGACAAAAGA- 3'

### Iowa/Iowa97/56 Primers

Primer Name	Sequence
Gap1F	5' - TAACTTAACAAAACGGACTT - 3'
Gap1R	5' - TGCTCACTAAACACCACCAGAAC- 3'
Gap1 AF	5' - GGG TACCTAGTGTCCAG- 3'
Gap1 AR	5' - GCAAAAGCATCAGCGTAAT- 3'
Gap2F	5' - AGGATTACGCTGATGCTTTTG- 3'
Gap2R	5' - CCTGCCACTGTACGCGCTCTATTT- 3'
Gap 3F	5' - TCAAACATTCGACAACCCCAGATT- 3'
Gap3R	5' - TTGCGAACTACACCACCACCTACT- 3'
Gap 4F	5' - TTGCTGATAAAGTAGGTGGTGGTG- 3'
Gap4R	5' - CAAGAAATCTGTAACACTGAAAC- 3'
Gap 5F	5' - TTCAACTGGCTGCAAACCTGGTT- 3'
Gap5R	5' - CCAGGACACAAAACCAGGAACACC- 3'
Gap 9F	5' - TTCTGTTGACGCTATAATGACTCG- 3'
Gap9R	5' - AACTGCCTTTTCTACACCTTGAT- 3'
Gap 19F	5' - AGCAACATGGATACTGGAGACG- 3'
Gap 19R	5' - GCCTTTGCCTTGTAATGCGAGAG- 3'
Gap 20F	5' - AACTTCAACTAGATGGGCTTCACT- 3'
Gap 20R	5' - GATTAGACATTTCCCTGGCGATAG- 3'
Gap 21F	5' - CTT TGC ACC TGT TAC TTT GGA TTG- 3'
Gap 21R	5' - GAT TGC CTG CCT CGT TCA- 3'
Gap 22F	5' - TTG CCA GAA GTT AGT TGT GA- 3'
Gap 22R	5' - AAC TGC CTT TTC TAC ACC TTG AT- 3'
Gap 23F	5' - GCA ATA AAT ATA AAA CAG AGC A- 3'
Gap 23R	5' - TGA CAC CAC CAA CAT TAG GA- 3'
Gap 24F	5' - CAG CTA TGA AAC GAA ATG G- 3'
Gap 24R	5' - AAA ACT GTT AGG TAT GAG CAC- 3'
Gap 25F	5' - GCA TGA GCT ACC AGA CTT TGA C- 3'
Gap 25R	5' - ATT TCT TGT TGC ATA CCC ATA C- 3'
Gap 26F	5' - AGC AAC ATG GAT ACT GGG AGA C- 3'
Gap 26R	5' - TTT TGA ACT TGA GCG TGA CTT- 3'
Gap 27F	5' - AAC AAC GGA GAT AAA TGA TGG ACT- 3'
Gap 27R	5' - TCA TCA AGT TCT TCT GCT GTT TCT- 3'
Gap 28F	5' - TTA AAC CAG CTA CAT GTG AAA AAC- 3'
Gap 28R	5' - AGT TGT GCC ATG CTT CTG C- 3'
3' RACE F	5' - GCA CGA CCC GCA GAT TTG ATA A- 3'
3' SEQ F	5' - TAC AAC AGC GCC CAA AGA AGG TG- 3'

I-97 A F	5' - AAC ACC GGA GAT AAA TGA TGG ACT- 3'
I-97 A R	5' - GCC CCA GTC ACC AAG AAC ATC AAC- 3'
I-97 B F	5' - AAC GTG AGA AGA AGG CCA AAA AGT- 3'
I-97 B R	5' - TTG CGC ATC TAA ACC ATA- 3'
I-97 C F	5' - ATG CAT GTT GAA GGT TTT A- 3'
I-97 CR	5' - ATT GAG GTG TAT TGC TGG CTG TAA- 3'
I-97 D F	5' - CCG ATG GCA CAG AGT TAC ACC- 3'
I-97 D R	5' - TAC ATG GTG GGT CTA AGT CTA CA- 3'
I-97 E F	5' - AAA TAT GTG GCA GCA GGT AAT CAA- 3'
I-97E R	5' - CAC ACT GAC ATC CAT AAC CAA TCC- 3'
I-97 F F	5' - GAG CGC GTA CAG TGG CAG GTG- 3'
I-97F R	5' - CGC GCA TTC ATT ATA CAA CCT A- 3'
I-97 GF	5' - ATG CGC CAG CGT GAT GAA T- 3'
I-97 GR	5' - TCC CAG TTG ACA TCT TGA CA- 3'
I-97 HF	5' - ATC TTA TCA CCC ACA GTT G- 3'
I-97 HR	5' - ATA TTA TAA TCA ACA CAA GTA- 3'
I-97 I F	5' - ATG GTT TGC TTG TGT TGC CTC CTA- 3'
I-97 I R	5' - GCA GAT GCT AAA ACA GAA AGT GAT- 3'
I-97 JF	5' - GGC TAG AAT TAA TCA CTT GGG TAT- 3'
I-97JR	5' - AGA TGA AGA TAA TAG TGG CAA AAG- 3'
I-97 KF	5' - AGA ACG GTT GGA ATA ATA AAA ATC- 3'
I-97KR	5' - GGC TGG TCC TGT TCC GGT GTA ATA- 3'
I-97 LF	5' - ATG TTG GCT AAA ATG GGA CCT AT- 3'
I-97LR	5' - CCT TAA ACA TAC CAG CCT TCT CT- 3'
I-97 MF	5' - ACT CTT TAA GCG GTG TAG GTC- 3'
I-97M R	5' - ACT GGA GCT GGG GCG TCT GT- 3'
I-97 N F	5' - CCA GGG GAA AAC TTG TGA GGA- 3'
I-97 N R	5' - AAA TCT AAG GGT CTA CCG TTC GTT- 3'
I97 GS1 F	5' - TGC TAA GTG TGC CAA GTC CAT T- 3'
I97 GS1 R	5' - GCT GCC ACA TCT GCG ACA C- 3'
I97 GS2F	5' - TTG TGC ACA GTT GAA AAG AGC- 3'
I97 GS2R	5' - CAA CAG CTT CAC CCC AGT CAC C- 3'
I97 GS3F	5' - TAA TTT TGA CCC TTT TGG GGT AA- 3'
I97 GS3R	5' - GGG TGT ATG CAG TAT TTG TTG TGG- 3'
I97 GS4F	5' - CAA TGA TGC ACC TGG AGC CTT ACC- 3'
I97 GS4R	5' - CAC TCT TGC GCA ACC TCT TCA TC- 3'
I97 GS5F	5' - GCG GAG CAG GAA ATT TAG ATG GAC- 3'
I97 GS5R	5' - CAC CAC ATA ACC TTT TTC AAC CA- 3'
I97 GS6F	5' - TGG AAA GGC CGT CTC TAT TAT GA- 3'
I97 GS6R	5' - TTG CCA AAA TAT AAG GTT CCA CAG- 3'
I97 GS7F	5' - TCG TAG TGA TGT TGA GCG TGA TT- 3'
I97 GS7R	5' - TGT CAC TTC CAG CTC CAA AAT G- 3'
I97 GS8F	5' - TGG CAA TTT TCA GAT GGC TTT TA- 3'
I97 GS8R	5' - GGA CCG TAA GCA ATT GAA ACT GAA- 3'
I97GS9F	5' - GAG GGT CAG TGG CTT GCT AA- 3'

**JMK/JMK/64 Primers**

Primer Name	Sequence
1F	5' - GGGCTACCTAGTGTCCAG - 3'
1R	5' - GCAAAAGCATCAGCGTAAT- 3'
2F	5' - TTCGTGCCAAACTCTCG- 3'
2R	5' - GCATCCTTCGCATACACG- 3'
3F	5' - ATACCAAATGCACCGAGAG- 3'
3R	5' - TTGCAGCTAGAATGAGTCCA- 3'
4F	5' - CTTGGTGA CTGGGGAGAG- 3'
4R	5' - TGGCAATTACAACAGTCAA- 3'
5F	5' - GACGCTTATGTTCTTTTTATTCTA- 3'
5R	5' - CACTATCTGCTTGCGTCCAC- 3'
6F	5' - TAAGAGTGCAGGCTATCCA- 3'
6R	5' - GTAACAAAACAACACCGTCGTC- 3'
7F	5' - AATGTTGCGCGTCTTTTGAGTGTT- 3'
7R	5' - CATCCTTACCTTCACCTTTTCAA- 3'
8F	5' - TGCAGAGACAGTAAAAGCCACAGA- 3'
8R	5' - CCATCATAAACAAGAGTAGACACA- 3'
9F	5' - TTCTGTTGACGCTATAATGACTCG- 3'
9R	5' - AACTGCCTTTTCTACACCTTGAT- 3'
10F	5' - ATACCATCTATTTGCTCGTTCA- 3'
10R	5' - CGCCACTCCTTTGTCGCTTCC- 3'
11F	5' - AGAAAGCATGAAGGCGTGATAGC- 3'
11R	5' - AGCACCCCCTTGTAATG- 3'
12F	5' - ACTGGTAAGAGATGTTGGTAA- 3'
12R	5' - GGACCGTAAGCAATTGAAACTGAA- 3'
13F	5' - TGGCAATTTTTCAGATGGCTTTTA- 3'
13R	5' - CGACGTGTTCCATTAGTGATTTTG- 3'
14F	5' - AAATTGCCCTTATGTTAGTTATGG- 3'
14R	5' - TTACCACATCGTTATCAAAAGTCG- 3'
15F	5' - GAATTATTGAAACTGGTGAGCA- 3'
15R	5' - CGCGGGACAAGGCTCTGCT- 3'
16F	5' - GGTGACCAAAGCGGAAATAAGAAA- 3'
16R	5' - TCCTTCTTTGGGCGCTGTTGTC- 3'
Gap 1F	5' - GGGCTACCTAGTGTCCAG- 3'
Gap 1R	5' - GCAAAAGCATCAGCGTAAT- 3'
Gap 2F	5' - GTAGTGCCAAACAACCCCTGAG- 3'
Gap 2R	5' - GCATCCTTTCGCATACACG- 3'
Gap 3F	5' - TTTCAAAAGGCACTGGCTATT- 3'
Gap 3R	5' - ACGACATCAGGTGGTGGTATCTCT- 3'
Gap 4F	5' - GCAGTGCTTGGAGAGGACATTT- 3'
Gap 4R	5' - TCAACAAAATCCAGTCCACAAAAG- 3'
Gap 5F	5' - TACTTCAGCAATAGTTCTCCTTCA- 3'

Gap 5R	5' - TGTTACTATCATACCACTGTTCG- 3'
Gap 6F	5' - AGGGTTTTATTTCTGGCTCTTTA- 3'
Gap 6R	5' - TAAAACCATCACCAGTGTAATCCA- 3'
Gap 7F	5' - ATGCGCAATATCTGTGTAAGC- 3'
Gap 7R	5' - GCCCAGAAAAGCATCAAAAT- 3'
Gap 8F	5' - CTCTACCCTGTTACTATGCGTTCT- 3'
Gap 8R	5' - ACTCCGGTGTCAATTCATCTTCA- 3'
Gap 9F	5' - TGGGAGTTATTCTTCGAGTTGGTG- 3'
Gap 9R	5' - GCCAACAGCATCCACTTCCTC- 3'
Gap 10F	5' - AGTGAGGCTCGGCTGATACCC- 3'
Gap 10R	5' - ACGCACAAACACTAAAACAAGAG- 3'
Gap 11F	5' - CAGATGTATGCAGAATTTGTGACT- 3'
Gap 11R	5' - GGGGCTTATCAACTTCACCAT- 3'
Gap 11aF	5' - TTGAAAAGGCTAAGCTCGTTGTC- 3'
Gap 11Ar	5' - AGGTCTGTCTCAACTTCAATGG- 3'
GAP 13F	5' - CCGCAGACTTCTTTTGGTAAT- 3'
Gap 13R	5' - TAGTGGGCAGGACATTCTTCTTTG- 3'
GAP 12F	5' - CTTACTCAATGGTTCGCTTTCAC- 3'
Gap 12R	5' - TTCCAACAAGCATAAGCCTGAGTA- 3'
GAP 14F	5' - CTGGTCAGCAATAATTC AACAAA- 3'
Gap 14R	5' - TAAAAGACCACCATGACCTACACC- 3'
3' RACE F	5' - GACGCCCAAACCTTCAACAAGAATG- 3'
3' SEQ F	5' - CAAGTTTGTCCAATACCGCT- 3'
SP 1 R	5' - ACCAGCAGACAGACAGACAACACG- 3'
SP 2 R	5' - CGTATAGAAAAACAAAGCGTCACCA- 3'
SP 3 R	5' - CTAGGTAGCCCAAGCAACGGATGT- 3'
180-203 R	5' - CGGCAAAACAAGCAGTGATACCAG- 3'

## APPENDIX B

## CHAPTER 4 SUPPLEMENTAL TABLES

Supplemental Table 4.1: Primer sequences used in this study

<b>CCoV Overlapping Primers</b>	
<b>Primer Name</b>	<b>Sequence</b>
CCoV 1F	5' - CGGACACCAACTCGAACTAAAC - 3'
CCoV 1R	5' - GCAGCATGACGGAGGGTTCG - 3'
CCoV 2F	5' - CGTAATGGATTTCGAGGGCTATGTT- 3'
CCoV 2R	5' - AAACCAGTCCAATCACCAACACCA- 3'
CCoV 3F	5' - TCGGATCACCTTTTATGGGGAATG- 3'
CCoV 3R	5' - ACAGAATCAGGACTAGCCATCAT- 3'
CCoV 4F	5' - ACTAGCGTATCCTGGCATTCTTGA- 3'
CCoV 4R	5' - GATCCTTTAATGGTACCCTGACTT- 3'
CCoV 5F	5' - ATGGTTTGTGGCTGGGAGATGAAG- 3'
CCoV 5R	5' - AACCACCATACATTTCTCCTTCAA- 3'
CCoV 6F	5' - TTGAAGGAGAAATGTATGGTGGTT- 3'
CCoV 6R	5' - GGCAGTCATGTATTTAAGTTCAGC- 3'
C1F	5' - GCGTGGCTATCTCTCATCTTTT- 3'
C2R	5' - AGTCAGGGGTGCAAACAATC- 3'
C3F	5' - CCTTTTATGGGGAATGGTGA- 3'
C4F	5' - TTTTGTGACGCTTGGGAAGA- 3'
C4R	5' - ATGTTTGAAAGGCATGCAG- 3'
C5F	5' - GTGGCACCAGGAGAAGGTTA- 3'
C5R	5' - GCACATCAGCATTTCCTCAA- 3'
C6F	5' - TTGCCCTGATCTTCTTCTTGA- 3'
C6R	5' - TGATATCGAAGCCACCACAA- 3'
C7F	5' - TAAAATGGGTGGTGGTGACA- 3'
C7R	5' - TCTGCATCACCAGAATGACC- 3'
C8F	5' - TCGTGAAGAAGGACTCTGAACAT- 3'
C8R	5' - GCATTGATGAATAGCGTGGA- 3'
C9F	5' - GGCACAACCTCAAATGGTCA- 3'
C10F	5' - TGTGTTCACTGGTGGCAAAT- 3'
C10R	5' - TTCTTCACGCCAGAAATGTG- 3'
C11F	5' - ATGGTTTCCATGATGCTGCT- 3'
C11R	5' - GAGCGTAATCAGCACCACCTT- 3'
C12F	5' - TGAAACATGTGTGCATGGTG- 3'
C12R	5' - TGAAAATCAGCCAACCTCTATCA- 3'
C13F	5' - GCCACCTCATGCATTTAAAGA- 3'

C13R	5' - CTTCTGATGCGATCTGTCA- 3'
C14F	5' - TCTCGTGCTCAAACCCTTCT- 3'
C14R	5' - TACCAATGTCCATGGCTGAG- 3'
C15F	5' - CCATCTGGTTCCTCTCTTGC- 3'
C15R	5' - CCAGTCACCAGAATCCTTTGT- 3'
C16F	5' - CACTGAATCTGTGCCTCCAA- 3'
C16R	5' - TCCATTGCCAAACTCATTGT- 3'
C17F	5' - TGGCAGTGCAGTGAGTAAGG- 3'
C17R	5' - CCATAGCAGCAGATTCAAA- 3'
C18R	5' - ACACTGCCAGGACATCCTTC- 3'
C19F	5' - AATGGCACAGCCTAGTGGTC- 3'
C19R	5' - CCCCAAACCTTTGTTGAGT- 3'
C20R	5' - TGTCACAGCAAGTGAAAACCA- 3'
C21F	5' - GCATTGTATTGGCTGTCACAA- 3'
C21R	5' - GGCCAAGCAAGACAACATTT- 3'
C22R	5' - GCTATGCATGGCAAGAAACAA- 3'
C23F	5' - TATGCTGCATTGCCTAGCTG- 3'
C23R	5' - AGCGCCATCAACATAAAACC- 3'
C24F	5' - CCAGCTGCATCAGCTACAAG- 3'
C24R	5' - GCAGGATGCTCAACATGACA - 3'
C25F	5' - GCTCTCAAATGGTGTGGTA- 3'
C25R	5' - ACAATGGGTCCAAGTTTTGC- 3'
C26F	5' - TTTGGACAAACATGATGCCTA- 3'
C26R	5' - GTCCGAAAGCTGTCATTGGT- 3'
C27F	5' - GGGATGCCTTTAATGGGGATGA- 3'
C27R	5' - GATGCCTTTAATGGGGATGA- 3'
C28F	5' - GTAGCATCAAGCCCTGCACT- 3'
C28R	5' - TGTACAACAACCGACGTGCT- 3'
C29F	5' - AAGTCAATTGCTGCAACACG- 3'
C30F	5' - CTCACGGAAGTTGTGCATTG- 3'
C30R	5' - CATACTGCAGCTTGCAAAA- 3'
C31F	5' - TATTGATGCATACCCGCTCA- 3'
C31R	5' - TTGCTCAAACACAAATTCACC- 3'
C32F	5' - CGCTGCTGAAACTGTGAAAG- 3'
C32R	5' - TGGTAATTGTTGAGGGTCTCC- 3'
C33F	5' - GTCCTCCTGGTAGCGGTAAA- 3'
C33R	5' - TCACTACCCTGTGCAGAGTCC- 3'
C34F	5' - TCAACCGCAGGATTACAATG- 3'
C34R	5' - GCAAAGTCTCGTGTGCAGAA- 3'
C35F	5' - TGGCTTACAAGCAAAACCTG- 3'
C35R	5' - GAACCTGTGTAGCCCCATTG- 3'
C36R	5' - ATAAGCCGGTGTGTGGAAAG- 3'
C37F	5' - TGTCATGAAAGCTGCTCTAGG- 3'
C37R	5' - AGTCCAAGTTTGCGTTTTGC- 3'

C38F	5' - ACTTTGGCATGGTTTTGTCA- 3'
C38R	5' - ACAAACCCATTTTTGCAAGG- 3'
C39F	5' - GCTGTGAAAGGGCTTAGTGC- 3'
C39R	5' - CAGCTCCCAAATGCAGAACT- 3'
C40F	5' - GCAGAGTGGAATCCTGGCTA- 3'
C40R	5' - TCCAACGACCATTTCATTCA- 3'
C41F	5' - CAATTGACGGTGAAAACACG- 3'
C41R	5' - GGGTTGCCTTGAACGACTAC- 3'
C42F	5' - TGGAGATGAATTCCCTTGTTTC- 3'
C42R	5' - ACCTCCAGGCTGTGTGGTAA- 3'
C43F	5' - TTTTGAAAACCAATGGTCTGG- 3'
C43R	5' - CAATTGTCCAGAAAGCTCCA- 3'
C44F	5' - TTTCGTTTGAAGGAGCTGGT- 3'
C44R	5' - TGGCAAGTGGAATGGGTGTA- 3'
C45F	5' - TTTCGCACATACTGCTGTCA- 3'
C45R	5' - CAGTGCCACGAGTTCTCTCA- 3'
C46F	5' - ACCAGCGAGCAGGTTGTTAG- 3'
C46R	5' - TGCCAATTTAAGGGCATTTC- 3'
C47F	5' - CGTTCAACCAATTAGCACTGG- 3'
C47R	5' - TGGCAAGACCTTGTGATGTC- 3'
C48F	5' - GCGCCGTGTCTATACCTTTT- 3'
C48R	5' - ACAACACATCACACCCCTCA- 3'
C49F	5' - GACGAACTGAGTGCTGATGC- 3'
C49R	5' - AGCAATAGCGGTATGCAAAA- 3'
C50F	5' - TTGACATTTTCAACGCAACC- 3'
C50R	5' - TTCGCATGGAATTTAGAATGG- 3'
C51F	5' - TGACGAACTTGATCGTGCAT- 3'
C51R	5' - GCAGCTCTGCCATGTACAA- 3'
C52F	5' - TTAAGACGTGTGTCGGCATC- 3'
C52R	5' - GCACGCAAGCATTAAACAAAA- 3'
C53F	5' - GTTCCCTAGGGCATTGACTG- 3'
C53R	5' - CAGTTGGCACACCTTCAAGA- 3'
C54F	5' - ATTATGGCCATTGTTCTGG- 3'
C54R	5' - ATCTTTCAGGAAGCGCCATA- 3'
C55F	5' - CGTTAGCTGGGGAGATGAAT- 3'
C56R	5' - TGGTTTCAAAGCTGTAAGCAA- 3'
C57F	5' - GCGACCAGATTGAAGTCACA- 3'
C57R	5' - CAGGCCAAGAACATGGAAAT- 3'
C58F	5' - GGAGTTTGGCAGAAACCAGA- 3'
C58R	5' - AGCGCGATGACCAGTAATTC- 3'

<b>CCoV 1-71 Primers</b>	
<b>Primer Name</b>	<b>Sequence</b>
171 g 1F	5' - GCTTGTTTGGCTTGTTATGATGAG- 3'
171 g 1R	5' - AGTGCTCCTTTAAAGTCTACAG- 3'
171 g 2F	5' - TGATGCCTTTAATGGGGATGACTT- 3'
171 g 2R	5' - GCACCCACCATCGTAACATTCAAA- 3'
171 g 3F	5' - CATCTGGTGTTGCTCCTGGTAGTA- 3'
171 g 3R	5' - ATGTTATACCACCTGCAAGAGATG- 3'
171 g 4F	5' - ACACAAAACCTAAAGCATCAAGT- 3'
171 g 4R	5' - ACTGAATGGATCTAACAAAATACA- 3'
171 g 5F	5' - CATTGTTCTGGCTCTTACGATTTT- 3'
171 g 5R	5' - CACTGTCATCCTTCTTGTTATTGG- 3'
171 g 3' F	5' - GAGAAAATCAACCACAACAGAAA- 3'
171 A F	5' - CATCTGGTGTTGCTCCTGGTAGTA- 3'
171 A R	5' - AAAGTGCTAAAGAAATTGTAACCA- 3'
171 B F	5' - GGTCAACCCATAGCTTCAACATTA- 3'
171 B R	5' - AAAAATTAGGTGTTGTTGTCAA- 3'
171 C F	5' - GCGTTAGTGCATTAGGAAGAAGC- 3'
171 C R	5' - TTGTTGGAGGGTAATGGGGTTGAA- 3'
171 D F	5' - GGGCGTGTCAAAGACAGGAAG- 3'
171 D R	5' - TGTATCACTATCAAAAGGAAAAT- 3'
171 E F	5' - TCCAGAAGCGCAACAACCAC- 3'
171 E R	5' - ACGAGCGGCAACATCAAACCTT- 3'

<b>CCoV TN 449 Primers</b>	
<b>Primer Name</b>	<b>Sequence</b>
TN g 1F	5' - CAGGTTTTGGTTGTGCTTGTGTTA- 3'
TN g 1R	5' - TGAGTCATAGTGGGTAGAACATTT- 3'
TN g 2F	5' - CAAAATCCAGCGTATGTGTCTCG- 3'
TN g 2R	5' - TTATCAAATGAGAAAGGGCAAGTA- 3'
TN A F	5' - GCTGGTGCATCTGGTGTTG- 3'
TN A R	5' - ATACAATCAATAGGAAAAGTGCT- 3'
TN B F	5' - TGGACATACTCTTGGACGATTTTG- 3'
TN B R	5' - ACCTTTGTCTGTGATCTGCTGTTA- 3'
TN C F	5' - TGTGGTGTTAGGTGATTACTTTC- 3'
TN C R	5' - ACATAACAGTAACGCGGTCCATCA- 3'

<b>MHV Overlapping Primers</b>	
<b>Primer Name</b>	<b>Sequence</b>
1F	5' - CCCTCTCAACTCTAAACTCTTG- 3'
1R	5' - GAACCTCCTCACAAGCATCC- 3'
2F	5' - GTGACATCCGGCCACTTC- 3'
2R	5' - AGCTCTCTGCTCGAGATTGC- 3'
3F	5' - GCAATTGTTCAAGAGACAGACG- 3'
3R	5' - AGTGGGTGGCTTACAACAGC- 3'
4F	5' - TGAAGATGATGTTGTTGATGTGG- 3'
4R	5' - AATGCAGACAACGCCTCTG- 3'
5F	5' - CGGACAAGACCTATGAGGATG- 3'
5R	5' - CCCTGAGCCTTCACATATC- 3'
6F	5' - GTGCAGGTGTTGCAGGTG- 3'
6R	5' - ATCTGCCTCGGACAAATCAC- 3'
7F	5' - GATGGCATAAATGTCACCAAAG- 3'
7R	5' - GGGCCTTAATAATGGGTTTTG- 3'
8F	5' - TGGAGTATAAGCCAGATTTATC- 3'
8R	5' - GCAATAGCTTTCGCCTTGAC- 3'
9F	5' - CACCCGCTAAGTTGTTGTTG- 3'
9R	5' - ACAGTCCACTGGTGCTTTG- 3'
10F	5' - GTGTAGCACCGTTGTTGGTG- 3'
10R	5' - CAAATGCAAGCCACATTAGC- 3'
11F	5' - CAGCCGATTTGGGTGTTTC- 3'
11R	5' - CTCGGCCTCCTCACATAAAC- 3'
12F	5' - CCCGAAGTGGTTAGTGAAGG- 3'
12R	5' - AGTAACAGAGGCGGTTGGAG- 3'
13F	5' - CTGTTCACAACTGGCAAAGG- 3'
13R	5' - TCTTAATAGCGGCCAACACC- 3'
14F	5' - AGCCGATCTTGTCTTGGATG- 3'
14R	5' - TATTTCTAGGTGGGCGCAAG- 3'
15F	5' - GATGCCATTGGGCGTCTAC- 3'
15R	5' - ACAGCACCATCAGCATCTTG- 3'
16F	5' - TTGATCAGGTTGTGGATAATGTG- 3'
16R	5' - TGTGCCTACACAGGAACAGC- 3'
17F	5' - GACGCATGATGTTTGTTCAGG- 3'
17R	5' - CGGGTGGTAGGTCATACTCC- 3'
18F	5' - GGAGTTTGACCTTGTTTCAGTATG- 3'
18R	5' - AACCGCCATAGAACTTCGTG- 3'
19F	5' - GGCAGAATGTTTCATCAAAAG- 3'
19R	5' - CAATTGCAAGACTTACGAAACG- 3'
20F	5' - AGGAGCAGGCTGTTTTGTTG- 3'
20R	5' - TGTGGCGCGATAGTACACAC- 3'
23F	5' - GAAGCCACTGGAATGTTTGC- 3'
23R	5' - AAAGGGACTGGTGTGGAATG- 3'

24F	5' - AATCTCCCTGGCTGTAATGG- 3'
24R	5' - TTACCACCCGGATTACCTTG- 3'
25F	5' - GCGTAAAGACGGTGACGATG- 3'
25R	5' - AATACTTCCCCTGGTAGCC- 3'
26F	5' - ATGCGTGTCTTACACCTTGG- 3'
26R	5' - TTGCAAATTCACCTGTTTG- 3'
27F	5' - CAGAAGGCACCACATGTTAG- 3'
27R	5' - CACAACCTTGGGTCGAGGTC- 3'
28F	5' - TTTGGTGACAGTCGTTCTGAC- 3'
28R	5' - TGCTGGGCAATACAGGAAAC- 3'
29F	5' - CATGGTGATTTTCATTTTCAGG- 3'
29R	5' - GAGGGTTTATCCGCATAGTACG- 3'
30F	5' - CGCTTAATGTTAATGCTGATGC- 3'
30R	5' - GGGCTAACTATGTCCGGTTG- 3'
31F	5' - CATGACGTTGTTTACGCTCA- 3'
31R	5' - ACCACAGACAAATGCAGCAC- 3'
32F	5' - CTATTGGGCACCATGAGGAG- 3'
32R	5' - CGATCTATCTGGGCTTTTGC- 3'
33F	5' - AACAGGTTTGGTGCTATTAGTGC- 3'
33R	5' - CTGTCCTGGTGTCCTCCATAC- 3'
34F	5' - ACAGGTTGTGGCTCATGTTG- 3'
34R	5' - CTGCCCCACATACCATACTG- 3'
35F	5' - ACGCAGCTCGAAAGAAATG- 3'
35R	5' - TTTCCGACCTTGGACTTCAC- 3'
36F	5' - AGCGCGCATTCTTAGACAAG- 3'
36R	5' - CCGAGCTTAGCCAAAACAAG- 3'
37F	5' - GCCAGCCTGCCTCTACTG- 3'
37R	5' - ACCCTGATGTGAGCTCTTCC- 3'

<b>MHV-S Specific Primers</b>	
<b>Primer Name</b>	<b>Sequence</b>
MHVS 1F	5' - TTAGTGGGCGAAAGTGAACCTG- 3'
MHVS 1R	5' - TTAAAAAGAAGTAGGCAAATC- 3'
MHVS 2F	5' - AGGGTTTTGCGCAGAGTTGGTG- 3'
MHVS 2R	5' - GGCTTACCAGAACGGAACCTCATT- 3'
MHVS 3F	5' - CATTGGAGTGCACGCTTGTTCG- 3'
MHVS 3R	5' - GCTGCAACTATAGTGTCGCTTTTA- 3'
MHVS 4F	5' - GAAACGCAGAAGGCCACAGAGG- 3'
MHVS 4R	5' - CCTGCGCTGAGTCTACGGTCTG- 3'
MHVS 5F	5' - ACGCATTACCCGAGTTGG- 3'
MHVS 5R	5' - CGCCAGCATCCATAATACCCAGTT- 3'

MHVS 6F	5' - ACCACCCTTTTTATCAGAGTTCAA- 3'
MHVS 6R	5' - GCAGCGGCCATTAGATAGACAAC- 3'
MHVS 7F	5' - GCGCCAACCCCAGCATAGTCT- 3'
MHVS 7R	5' - TCTTGATGGTATGCAGCCGAAACA- 3'
MHVS 8F	5' - AGCCATCAGGAGTTCATTTCAGACG- 3'
MHVS 8R	5' - ATCTCATCGGACAGGCAAGGAAT- 3'
MHVS 9F	5' - TGGCTGCTGTCTTTATTGGTCCT- 3'
MHVS 9R	5' - CGCGCCACTCGGTTTATTTGA- 3'
MHVS 10F	5' - GTTAAGCTAGGCACTGGTTCTCC- 3'
MHVS 10R	5' - ATTACCAGAGCGGCTTCCAGAGG- 3'
MHVS 11F	5' - TACCGACTGCCCTCAAATAAACC- 3'
MHVS 11R	5' - ACATAAAAACCCTGAGGCAAAACC- 3'
<b><i>MHVS RACE PRIMERS</i></b>	
MHVS SP1	5' - ATCAACATGCCAAGCCACAACAAC- 3'
MHVS SP2 R	5' - GCGTCACAAGAGCACTCCCTGGTA- 3'
MHVS SP3 R	5' - ACCCCATGGCTGCTGTTTTAGG- 3'
MHVS 5'seq R	5' - GCAGAGAACGAAAGTCAAGGAACC- 3'
MHVS 3' RACE F	5' - CGGTTTTGCCTCAGGGTTTTTATG- 3'

<b>RaCoV Overlapping Primers</b>	
<b>Primer Name</b>	<b>Sequence</b>
1F	5' - CCCTCTCAACTCTAAACTCTTGT- 3'
1R	5' - GGATGTACAGACCCCTTGT- 3'
2F	5' - GCTTTGGAGTGCTGTGTTC- 3'
2R	5' - CTCCATCCACCACATCCTCT- 3'
3F	5' - CCATCCTTTTTGTGGACCAG- 3'
3R	5' - ATTGCCCACTTGTGCTGAA- 3'
4F	5' - GATTGCTGTGGTGACACCTG- 3'
4R	5' - ACGGCAAGATCAAACCATTC- 3'
6F	5' - TTAGTGGGCGAAAGTGAACC- 3'
8R	5' - AACCAAGCGAGCAACCTTTA- 3'
9F	5' - CATTTTCCAGTGCGAGATGA- 3'
10F	5' - GGGCAAGGTAGGCAATGTTA- 3'
10R	5' - ACAAGATCGGCCTCAGACAG- 3'
11R	5' - GTC AACACCTTTGCGCTGAT- 3'
13F	5' - GGTTTCGGAGGCTAAGGGTA- 3'
14R	5' - AGATGGTTGACACGCCAAAT- 3'
15F	5' - TATTGCCTGTGGTGCTG- 3'
16F	5' - TTGGAGTGCTCGTTTGTGTTG- 3'
18R	5' - AAGCTGCACACCCTCCTTTA- 3'
19F	5' - TGTCCTGTTGTGGTTGCTGT- 3'
20R	5' - CACTTTCGATGTGGGAGACA- 3'
21F	5' - TACGTGCGAGTCGTTGTTTC- 3'
22R	5' - ACAATGTCACGCCAACATA- 3'

24F	5' - TGCTTGGCTTTACATTTTG- 3'
25R	5' - ACAGCACCATCAGCATCTTG- 3'
26F	5' - GCTGGAACGTATGGCTGATT- 3'
26R	5' - GAAAACCACACAACGACAG- 3'
27F	5' - ACTGGCACGGGTAAGATTGT- 3'
27R	5' - GCGGCAGCAATTCACCTTAT- 3'
28F	5' - GGTGGTGCTTCCGTTTGTAT- 3'
28R	5' - CTCCACTAATGCGTCTGCAA- 3'
29F	5' - CCACACATAGCCGCAAAGA- 3'
29R	5' - CTAGCAGACGCCACATGAAG- 3'
31F	5' - GAGCTCCGTTGATTTGAAGC- 3'
31R	5' - GCGGAAACAGCTTGACAAAT- 3'
33F	5' - GAGTTTGCCTCCAAGGGTTA- 3'
34F	5' - GAAGTGCAGTGATGCAAAGC- 3'
34R	5' - TGAAACGTCTCAGGCACACT- 3'
35R	5' - TCTGGACCCAAACAACACAT- 3'
36F	5' - GCAAAGGTGCGTGTAGATTG- 3'
36R	5' - CACTGCAATCGTGGATTGTT- 3'
37F	5' - GGCAAATCCCAGTTGGAGTA- 3'
37R	5' - TCGAACCACATCCCATTCT- 3'
38F	5' - CCACTGGAATGTTTGCTGAG- 3'
39R	5' - GACCATCAAAGTGTCCAGCA- 3'
40F	5' - AATTTAGGTGGCGCTGTTTG- 3'
40R	5' - AGTGCCGCGTGCTAGAGCTT- 3'
41F	5' - TTGATGGTTCGTGATAATGGTG- 3'
41R	5' - TAGCAGGTTTCCAATCAGC- 3'
42R	5' - ACGCTGTCACTCACAAATGG- 3'
44F	5' - GGAATGGAGGGGCTTACAGT- 3'
44R	5' - TTTGGGTGTTGGAGGCTTAT- 3'
45F	5' - TTTGACAACCCCATATGCT- 3'
45R	5' - CACAATTTTGGGTCGAGGTC- 3'
46F	5' - TAAGCCTCCAACACCCAAAT- 3'
46R	5' - TTCGAAATGAAGGTGGGATT- 3'
48F	5' - GTGGAATAGCACCCGTCAGT- 3'
48R	5' - GTATGAAGCCGAGCTATGG- 3'
49R	5' - ACTGCGGTGAGAAAACACCT- 3'
50F	5' - TAAGGTCTCCGGTGTGCATT- 3'
51R	5' - TGGCAGCGGTCATTAGATAG- 3'
52F	5' - GCTTGGCGTAGGTGATCATT- 3'
52R	5' - AACCTTCGGAGACCTCGTTT- 3'
53F	5' - TCAAGATGCACCTGAACCAG- 3'
53R	5' - GCCATTGAAAGATTGACAC- 3'
54F	5' - AAACGAGGTCTCCGAAGGTT- 3'
55R	5' - GCCAATTTCTTGAGGTTGA- 3'
56F	5' - ACAATGCGCCTTACGGTTTA- 3'

56R	5' - AGACAAGCAGCCAATTTAGCC- 3'
57F	5' - GGACGTCAGGCAGGTATTGT- 3'
57R	5' - CCATCAAACACACTGCGACT- 3'
58F	5' - TTGGTGTAGCAACCAGGGTA- 3'
58R	5' - GAAGCTCCACCAGCTACCAG- 3'
59F	5' - GGTACACGAGCCGTAGCAT- 3'
59R	5' - GTAATGCCCGAAAACCAAGA- 3'
60F	5' - CTGATTTGCCCGCTTATGTT- 3'
60R	5' - CTAGCAGGTGCAGACCTTCC- 3'
61R	5' - TTCTGAGCCGTTTGTTCGT- 3'
62F	5' - ACCTGATATGGCCGAAGAAA- 3'
62R	5' - TGATACAGGGTCTGCCACAA- 3'
RaCoV 3' F	5' - CTATGTCGGCGCTCGGTGGTA- 3'

<b>RaCoV 681 Specific Primers</b>	
<b>Primer Name</b>	<b>Sequence</b>
681 1F	5' - CCAGGGGGTGTTCCTTGTGACG- 3'
681 1R	5' - TGGGATTACACCTGATGATTAGC- 3'
681 2F	5' - CGTTGCGCATACTGGTGGTCA- 3'
681 2R	5' - CGCAAAACTTATCATCTCGCCTC- 3'
681 3F	5' - GCAAGAGTTCGTAGCGCCAAAAGT- 3'
681 3R	5' - TCTTATAACAAAACAGGCAACCAG- 3'
681 4F	5' - ATGCTAGTGGTTGCGTGTGTCGT- 3'
681 4R	5' - TCGCCGTTAAAGCAAAGAAATCAA- 3'
681 5F	5' - CTTGCATGCCCTACAGAGTGAAT- 3'
681 5R	5' - GTGCCTACAACCCATCCTCTA- 3'
681 6F	5' - TTATGACAAGAGTGCTGGCTATCC- 3'
681 6R	5' - TTGGCCTCAGACATAAACACATT- 3'
681 7F	5' - GCTGTTCATGATTGCTTTTGTA- 3'
681 7R	5' - CTTTCTCCACAACACTACGCCATTCA- 3'
681 8F	5' - TTTGAATATTGACGTGTGCTGGAG- 3'
681 9F	5' - CCGGAGCCCACAAGGTAATC- 3'
681 8R & 9R	5' - AAGAGGGTCGTACATATCAGAAAT- 3'
681 10F	5' - ATGGCGGATAGCAGTGTTAGGTTG- 3'
681 10R	5' - CACAGCACCAATATACGAACTAAA- 3'
681 11F	5' - TCTCTGATGTTGGCTTTGTCG- 3'
681 11R	5' - ATCTCATAGGACAGGTCAAGGAAT- 3'
681 12F	5' - GTGGATCGCAAGAGTCTAACAAGT- 3'
681 12R	5' - TAACAGCAACCACAACAGGACAAG- 3'

<b>RaCoV 8190 Specific Primers</b>	
<b>Primer Name</b>	<b>Sequence</b>
8190 gap 1F	5' - GCCGGCTGCAGTCCTTGTTAC- 3'
8190 gap 1R	5' - CGCCGTTAGCCATCACATCGTA- 3'
8190 gap 2F	5' - TGCTCGTTTCTTTGTGTTTGTTC- 3'
8190 gap 2R	5' - GAGGGGAAATAGAGCGTCGTA- 3'
8190 gap 4F	5' - GCATTTGCAGTGGGCTTGTCAT- 3'
8190 gap 4R	5' - GGCTCATACGGTCAGACATAACAC- 3'
RaCoV 8190 last gap F	5' - TGCTCGTTTGTGTTTGTTC- 3'
RaCoV 8190 last gap R	5' - GTTGATTACAGGGCGTTTCA- 3'

<b>FCoV Overlapping Primers</b>	
<b>Primer Name</b>	<b>Sequence</b>
1F	5' -GTTTCCGCCCCTCGTAG- 3'
1R	5' -TTCGCCTTGAAATCAGT- 3'
2F	5' -GAGGTAGCACCACCAGTCAAG- 3'
2R	5' -AGCCTCTACACCCTGCAATGG- 3'
3F	5' -GCATCACTCTGCACTTCTGC- 3'
3R	5' -GCACCAAATCCAATTCCAG- 3'
4F	5' -CCGTAAGAAGGGTTTCTTG- 3'
4R	5' -CCACACTTGCATCAGAATCG- 3'
5F	5' -GCAGGGTGTAGAAGTGAAGG- 3'
5R	5' -GGAACGTCAGGCTCAAGAAAG- 3'
6F	5' -TGCCCGTAGAGGATAAGTGTG- 3'
6R	5' -CTTTCGTTAAGCCCGAAATG- 3'
7F	5' -TTGGGTTAATGCAATTTGTTAG- 3'
7R	5' -GGCTGGTTTGTTC AATTGTC- 3'
8F	5' -GACGTATTCTTTATGCGTGGTG- 3'
8R	5' -AACCCGTCGGTACATTCAAC- 3'
9F	5' -TGTCATGCTAATGGTACTGG- 3'
9R	5' -GAAACCTGAACCGCTCTTTG- 3'
10F	5' -GGCTCAGTCAGGATTTCTCC- 3'
10R	5' -TGAAGAATCCCAATGTGGACC- 3'
11F	5' -GGAGTTTGGTCTACGGTTTG- 3'
11R	5' -TTGCCATAAGCCACCCTTAC- 3'
12F	5' -ATTACCGCATGGCCTGCTAC- 3'
12R	5' -TAGGCGTGAATTCGTCACAC- 3'
13F	5' -TACAATGCATGGGCCAAAAC- 3'
13R	5' -CGGTGAAACGCTAAATTGG- 3'
14F	5' -TGTAATGCAAGCCAGCTTTTG- 3'
14R	5' -TGCCTAACCTTGAAAGCAC- 3'
15F	5' -TTTGGTATGCTCAAGAACTCG- 3'
15R	5' -TTAAAGCGGCACAATCCATC- 3'
16F	5' -ATGGTGCCGGTAATGGTATG- 3'

16R	5' -GATGTAACGCAAGCACATCC- 3'
17F	5' -GCCAACGCTATGCTTAAATG- 3'
17R	5' -CACCCACCGTCGTAACATTG- 3'
18F	5' -GCGTGGATTCTTTGAAGAGG- 3'
18R	5' -TCTTCATCAATGCTGCTACTACG- 3'
19F	5' -CATCTTTCAAGCTGTTTCTGC- 3'
19R	5' -TTGCCATTAGCACAGAGTGG- 3'
20F	5' -TGGGAACAAAGCACAAGGTTC- 3'
20R	5' -TTGCTTCTGGTAGAGCATTGAC- 3'
21F	5' -GTTCTCACGCGGCTGTAGAC- 3'
21R	5' -TGTTGTTGCATAAGCAGGTG- 3'
22F	5' -TTGAATTACGAGAGCAAAG- 3'
22R	5' -AGCATCACCCTAGCTACATGC- 3'
23F	5' -TGTGAATGCGGTAAAAGTGC- 3'
23R	5' -TAACAAAACCGTGCCAAAGC- 3'
24F	5' -AACATTGGTGGTGTGTCTG- 3'
24R	5' -AGGCGCACTTGCGATATAAG- 3'
25F	5' -CACATGGACGTACTTTTGAAACC- 3'
25R	5' -ATTTTGATGGCAGCAGATCC- 3'
26F	5' -CGCTGACTTCAGTGTTACAGG- 3'
26R	5' -CTAACAGGCTCACCATGCAC- 3'
27F	5' -GCTCGAACTACTGCCTTTCAG- 3'
27R	5' -AACAGTATCGTGGTCGCTCAG- 3'
28F	5' -CAATCTGGTATGGGTGCTACAG- 3'
28R	5' -AGCACAGACAAATCGTGCAG- 3'
29F	5' -TTGCTGCACGTACAAGAACC- 3'
29R	5' -CAACCTATGCTAGGCCATTC- 3'
30F	5' -TGCCAGACTTGAAAACATGG- 3'
30R	5' -GAAAATCATFCCATTTGGTG- 3'
31F	5' -GACAACCTGCCAAAGACAAGG- 3'
31R	5' -GCAAAGTATGCACGGTCAAG- 3'
34F	5' -CTGGAACCTCAGCTGGTCTG- 3'
34R	5' -CTTGCAACCCAGAAGACTCC- 3'
35F	5' -TGTCCGAGAGACCTTGTTC- 3'
35R	5' -GCATGCGTTTAGTTCGTAACC- 3'
36F	5' -AGGCAAAGAAGATCCCGTTC- 3'
36R	5' -CCAAAGGGAAGATCAACAGC- 3'

<b>FCoV/WSU/79-1683 Specific Primers</b>	
WSU 1F	5' -GCGTGGCTATAACTATTTTT- 3'
WSU 2F	5' -GCCTGGTTCTTCTGGTGTTCCTGC- 3'
WSU 2R	5' -ACACCTGGGTATGCAAGTTTTC- 3'

WSU 3F	5' -TGGCCTTGAGGATTATGGTTTTGA- 3'
WSU 3R	5' -AGAGGCAAGTTACGAGCACAATCA- 3'
WSU 4F	5' -GACCCCCAGAACTATGTATCAG- 3'
FW 1F	5' -AAACGTTTGTGGAGGAGGGTGTAA- 3'
FW 1R	5' -ACAAACCGTAGACCAAAACTCCT- 3'
FW 2F	5' -AAGGAAGTGTAGTTGTTGGTGGTT- 3'
FW 2R	5' -TCCAAAAAGCTCCACTAACACCAG- 3'
FW 3F	5' -ATGTAAAATGGCCTTGGTATGTGT- 3'
FW 3R	5' -ACTACGGTACAAAGCCAAAAATG- 3'
FW 4F	5' -TTCTGGGTTGCAAGGGATGGTG- 3'
FW 4R	5' -AACAAATTTACACGCAGACTACT- 3'
FW 5F	5' -TAACGCTGCCAAATGCTACCC- 3'
FW 5R	5' -TGTATCACTATCAAAAGGAAAAT- 3'
FW 6F	5' -GTCATCGCGCTGCCTACTCTTGT- 3'
WSU 3' race F	5' -GTCATCGCGCTGCCTACTCTTGT- 3'
WSU LG 1F	5' -ATGTAAAATGGCCTTGGTATGTGT- 3'
WSU LG 1R	5' -CACCTGCAACACTAAAGCCGAACA- 3'
WSU LG 2F	5' -TGCCAAATGCTACCCTCAGA- 3'
WSU LG 2R	5' -AAACCTTCGATACCCTCACA- 3'
WSU LG 3F	5' -ATTACAGCATTTCATAGGA- 3'
WSU LG 3R	5' -CATTGTTGGCTCGTCAT- 3'