Genomic and single nucleotide polymorphism analysis of infectious bronchitis coronavirus

Celia Abolnik

Poultry Section, Department of Production Animal Studies, Faculty of Veterinary Science, University of Pretoria, Onderstepoort 0110, South Africa.

Phone: +27 12 529 8258

Email: celia.abolnik@up.ac.za

Postal address: University of Pretoria, Private Bag X04, Onderstepoort, 0110, South Africa

Highlights

- Genome organisation of QX-like IBV includes accessory protein genes 4b, 4c and 6b
- Relative variability in open reading frames across the genome is described
- Provides insights into the origins of the S1 subunit HVRs
- Describes novel features of the S2 subunit
- First model of the IBV spike monomer

Abstract

Infectious bronchitis virus (IBV) is a *Gammacoronavirus* that causes a highly contagious respiratory disease in chickens. A QX-like strain was analysed by high-throughput Illumina sequencing and genetic variation across the entire viral genome was explored at the sub-consensus level by single nucleotide polymorphism (SNP) analysis. Thirteen open reading frames (ORFs) in the order 5'-UTR-1a-1ab-S-3a-3b-E-M-4b-4c-5a-5b-N-6b-3'UTR were predicted. The relative frequencies of missense: silent SNPs were calculated to obtain a comparative measure of variability in specific genes. The most variable ORFs in descending order were E, 3b, 5'UTR, N, 1a, S, 1ab, M, 4c, 5a, 6b. The E and 3b protein products play key roles in coronavirus virulence, and RNA folding demonstrated that the mutations in the 5'UTR did not alter the predicted secondary structure. The frequency of SNPs in the Spike (S) protein ORF of 0.67% was below the genomic average of 0.76%. Only three SNPS were identified in the S1 subunit, none of which were located in hypervariable region (HVR) 1 or HVR2. The S2 subunit was considerably more variable containing 87% of the polymorphisms detected across the entire S protein. The S2 subunit also contained a previously unreported multi-A insertion site and a stretch of four consecutive mutated amino acids, which mapped to the stalk region of the spike protein. Template-based protein structure modelling produced the first theoretical model of the IBV spike monomer. Given the lack of diversity observed at the sub-consensus level, the tenet that the HVRs in the S1 subunit are very tolerant of amino acid changes produced by genetic drift is questioned.

Keywords: Infectious bronchitis, Coronavirus, SNP, Spike, HVR

1. Introduction

Coronaviruses (family *Coronaviridae*, order *Nidovirales*) are enveloped, single-stranded RNA viruses with large genome sizes of ~25 to 30 kb. The family is split into four genera: *Alpha-*, *Beta*, *Gamma* and *Deltacoronaviruses*, each containing pathogens of veterinary or human importance. A current evolutionary model postulates that bats are the ancestral source of *Alpha-* and *Betacoronaviruses* and birds the source of *Gamma-* and *Deltacoronaviruses* (Woo et al., 2012). The *Alphacoronaviruses* infect swine, cats, dogs and humans. *Betacoronaviruses* infect diverse mammalian species including bats, humans, rodents and ungulates. The SARS coronavirus (SARS-CoV), which verged on a pandemic in 2003 with 8273 cases in humans and 755 deaths is a *Betacoronavirus*. Another member of this genus, the recently-discovered Middle East Respiratory Syndrome (MERS) coronavirus (MERS-CoV) has claimed 88 human lives from 212 cases since April 2012, and dromedary camels are the suspected reservoir (Briese et al., 2014). Genus *Gammacoronavirinae* includes strains infecting birds and whales (Woo et al., 2012; McBride et al., 2014, Borucki et al., 2013) and deltacoronaviruses have been described in birds, swine and cats (Woo et al., 2012). The diversity of hosts and genomic features amongst CoVs have been attributed to their unique mechanism of viral recombination, a high frequency of recombination, and an inherently high mutation rate (Lai and Cavanagh, 1997).

Infectious bronchitis virus (IBV) is a gammacoronavirus which causes a highly contagious respiratory disease of economic importance in chickens (Cook et al., 2012). IBV primarily replicates in the respiratory tract but also, depending on the strain, in epithelial cells of the gut, kidney and oviduct. Clinical signs of respiratory distress, interstitial nephritis and reduced egg production are common, and the disease has a global distribution (Cavanagh, 2007; Cook et al., 2012). The IBV genome encodes at least ten open reading frames (ORFs) organised as follows: 5' UTR-1a-1ab-S-3a-3b-E-M-5a-5b-N-3a-3'UTR. Six mRNAs (mRNA

1-6) are associated with production of progeny virus. Four structural proteins including the spike glycoprotein (S), small membrane protein (E), membrane glycoprotein (M), and nucleocapsid protein (N) are encoded by mRNAs 2, 3, 4 and 6, respectively (Casais et al., 2005; Hodgson et al., 2006). Messenger RNA (mRNA) 1 consists of ORF1a and ORF1b, encoding two large polyproteins via a ribosomal frameshift mechanism (Inglis et al., 1990). During or after synthesis, these polyproteins are cleaved into 15 non-structural proteins (nsp2-16) which are associated with RNA replication and transcription. The S glycoprotein is post-translationally cleaved at a protease cleavage recognition motif into the amino-terminal S1 subunit (92 kDa) and the carboxyl-terminal S2 subunit (84 kDa) by the host serine protease furin (de Haan et al., 2004). The multimeric S glycoprotein extends from the viral membrane, and the globular S1 subunit is anchored to the viral membrane by the S2 subunit via non-covalent bonds. Proteins 3a and 3b, and 5a and 5b are encoded by mRNA 3 and mRNA 5 respectively and are not essential to viral replication (Casais et al., 2005; Hodgson et al., 2006).

A confounding feature of IBV infection is the lack of correlation between antibodies and protection, and discrepancies between *in vitro* strain differentiation by virus neutralization (VN) tests and *in-vivo* cross-protection results. Taken with the ability for high viral shedding in the presence of high titres of circulating antibodies, the involvement of other immune mechanisms are evident, and the roles of cell-mediated immunity and interferon have been experimentally demonstrated (Timms et al., 1980; Collison et al., 2000; Pei et al., 2001; Cook et al., 2012).

Dozens of IBV serotypes that are poorly cross-protective have been discovered and studied by VN tests and molecular characterisation of the S protein gene. Most of these serotypes differ from each other by 20- 25% at amino acid level in S1, but may differ by up to 50%. S1 contains the epitopes involved in the induction of neutralizing, serotype-specific and

hemagglutinaton inhibiting antibodies (Cavanagh, 2007; Darbyshire et al., 1979; Farsang et al., 2002; Ignjatovic et al., 1991; Meulemans et al., 2001; Gelb et al., 1997). Most of the strain differences in S1 occur in three hypervariable regions (HVRs) located between the amino acid residues 56-69 (HVR1), 117-131 (HVR2) and 274-387 (HVR3) (Moore et al., 1997; Wang & Huang, 2000). Monoclonal antibody analysis mapped the locations of many of the amino acids involved in the formation of VN epitopes to within the first and third quarters of the linear S1 polypeptide (De Wit, 2000; Kant et al., 1992; Koch et al., 1990), which is where closely-related stains (>95% amino acid identity) also differ (Bijlenga et al., 2004; Farsang et al., 2002). Cavanagh (2007) proposed that these parts of the S1 subunit are very tolerant of amino acid changes, conferring a selective advantage. Recently, the receptor-binding domain of the IBV M41 strain was mapped to residues 16-69 of the N terminus of S1, which overlaps with HVR1 (Promkuntod et al., 2014).

The S2 subunit, which drives virus-cell fusion, is more conserved between serotypes than S1, varying by only 10-15% at the amino acid level (Bosch et al., 2005; Cavanagh, 2005). Although it was initially thought that S2 played little or no role in the induction of a host immune response, it has since been shown that an immunodominant region located in the N-terminal half of the S2 subunit can induce neutralizing, but not serotype-specific, antibodies demonstrated by the ability of this subunit to confer broad protection against challenge with an unrelated serotype (Kusters et al., 1989; Toro et al., 2014).

IBVs are continuously evolving as a result of (a) frequent point mutations and (b) genomic recombination events (Cavanagh et al., 1992; Kottier et al., 1995; Jackwood et al., 2005; Zhao et al., 2013, Kuo et al,., 2013; Liu et al., 2014). Multiple studies on IBV diversity have focused on inter-serotypic and inter-strain variation, and a few have focused on sub-populations within the S1 subunit in vaccine strains (Gallardo et al., 2012; Ndegwa et al., 2014). The present study aimed to explore genetic variation across the entire viral genome at

the sub-consensus level. It was anticipated, based on the published literature, that certain regions, and the S1 subunit HVRs in particular, would display significant sub-genomic variation. This study focused on a QX-like strain, a serotype currently causing significant poultry health problems across Europe, Asia, South America and South Africa.

2. Materials and methods

2.1 Origin and isolation of QX-like strain ck/ZA/3665/11

Twenty eight-day old chickens in a commercial broiler operation presented with acute lethargy, reduced feed consumption and mortality. Tracheitis and swollen kidneys were noted on post mortem, as well as a secondary *E. coli* infection. The worst affected houses had mortality rates of 19.8%, 11.9% and 10.2%. IBV was isolated in specific pathogen free (SPF) embryonated chicken eggs (ECE) as described in Knoetze *et al.*, (2014). After an initial two passages in ECE, the virus was passaged twice further at the University of Pretoria.

2.2 Preparation of the genome and Illumina sequencing

RNA was extracted from allantoic fluid using TRIzol® reagent (Ambion, Life Technologies, Carlsbad, USA) according to the manufacturer's protocol. The genome was transcribed to cDNA and amplified using a TransPlex® Whole Transcriptome Amplification kit (Sigma-Aldrich, Steinheim, Germany). Illumina MiSeq sequencing on the cDNA library was performed at the ARC-Biotechnology Platform, Onderstepoort, Pretoria.

2.3 Genome assembly, RNA folding and recombination analysis

Illumina results were analysed using the CLC Genomics Workbench v 5.1.5. Paired-end reads were trimmed and a preliminary *de novo* assembly was performed. The larger segments were analysed by BLAST to identify the closest genomic reference strain (ITA/90254/2005, CAZ86699).This strain was retrieved and used as a scaffold for assembly-to-reference, generating a consensus sequence for 3665/11. Trimmed paired-end reads were also mapped against other IBV serotype genomes, subsequently confirming that strain 3665/11 was a pure culture of a QX-like IBV. The genome was deposited in Genbank under the accession number KP662631. RNA folding was predicted using the CLC Genomics Workbench v 5.1.5. Genetic recombination in the consensus sequence was evaluated using the recombination detection program RDP v4.31.

2.5 Genome annotation and Single Nucleotide Polymorphism (SNP) analysis

Coding sequence and ORF prediction was carried out in VIGOR (Wang *et al*, 2010). Trimmed paired-end reads were re-mapped against the 3665/11 consensus sequence for SNP detection. A SNP detection table generated in the CLC Genomics Workbench was manually edited to eliminate all SNPs with a frequency of <5 %. This conservative cutoff was selected to eliminate any non-specific PCR errors introduced during preparation of the transcriptome library or deep sequencing, and excluded most of the point insertions producing gaps and frameshift mutations across the genome. Nucleotide substitutions in coding regions were manually inspected for changes to the consensus amino acid (Table 1, Supplementary data). Motifs were predicted using the ELM Eukaryotic Linear Motif Resource for Functional Sites in Proteins (Dinkel et al., 2014)

Table 1. Single nucleotide polymorphism analysis of QX-like infectious bronchitis virus strain 3665/11

Reference/ consensus Position (position in complete genome)	Consensus Reference	Consensus /Allele variations	Relative frequency	Counts	Coverage	ORF	Amino Acid Change	Mutation	SMART/ Pfam domain
9 (148)	C	C/G/T	54.5/40.9/4.5	12/9/1	22	5'UTR	non-coding	-	
25 (164)	T	T/G	57.6/42.4	87/64	151	5'UTR	non-coding	-	
30 (169)	T	T/G/C	47.2/27.4/25.5	150/87/81	318	5'UTR	non-coding	-	
31 (170)	T	T/C/G	66.4/25.1/8.5	227/86/29	342	5'UTR	non-coding	-	
433	C	C/T	94.3/5.7	3534/215	3749	1a	^{26}Gln/Ter	Missense	
463	T	T/C/A	95.3/4.7	1989/99	2088	1a	^{36}Ser/Pro/Thr	Missense	
793	C	C/G	94.6/5.4	280/16	296	1a	^{146}Leu/Val/Phe	Missense	
796	C	C/T	90.8/8.2	277/25	305	1a	^{147}Gln/His	Missense	
797	A	A/G	94.3/5.7	283/17	300	1a	^{147}Gln/Arg	Missense	
1037	T	T/G	91.2/8.7	2394/227	2624	1a	^{227}Val/Gly	Missense	
1201	G	G/A	88.2/10.7	2639/320	2992	1a	^{282}Glu/Lys	Missense	
1208	T	T/C/A	86.9/8.7/4.2	1524/152/74	1753	1a	^{284}Ile/Thr/Lys	Missense	
1216	G	G/A	94.6/4.9	619/32	654	1a	^{287}Ala/Thr	Missense	
1375	T	T/C	94.6/5.4	35/2	37	1a	^{340}Tyr/His	Missense	
1381	T	T/A	94.6/5.4	35/2	37	1a	^{342}Phe/Ile	Missense	
1387	A	A/G	93.9/6.1	31/2	33	1a	^{344}Ser/Gly	Missense	
1406	C	C/G	94.1/5.9	32/2	34	1a	^{350}Ala/Gly	Missense	
1417	T	T/A	93.8/6.3	30/2	32	1a	^{354}Ser/Thr	Missense	
1449	C	C/G	78.6/21.4	22/6	28	1a	^{364}Cys/Trp	Missense	
1454	A	A/G/T	65.2/26.1/8.7	15/6/2	23	1a	^{366}Glu/Gly/Val	Missense	
1456	C	C/T/G	58.3/25.0/16.7	21/9/6	36	1a	^{367}Arg/Cys/Gly	Missense	
1461	A	A/G	92.8/7.2	64/5	69	1a	^{368}Val/Val	Silent	
1666^1667	_	_/T	74.6/25.4	97/33	130	1a	-	Missense	

1667	A	A/G	71.1/28.9	81/33	114	1a	^{437}Arg/Arg	Silent	
1667^1668	G	G/_	58.6/41.4	87/36	123	1a	-	Missense	
1709	C	C/T/G	88.0/6.8/5.1	103/8/6	117	1a	^{451}Ser/Phe/Cys	Missense	
1712	C	C/T	91.3/8.7	115/11	126	1a	^{452}Ser/Leu	Missense	
2150	A	A/G	95.2/4.8	581/29	610	1a	^{598}Asp/Gly	Missense	
2468	A	A/G	95.7/4.3	1057/48	1105	1a	^{704}Glu/Gly	Missense	
2743	G	G/A	94.4/5.1	203/11	215	1a	^{796}Asp/Asn	Missense	
2827	G	G/A	88.2/10.2	44/5	49	1a	^{824}Asp/Asn	Missense	
2922	T	T/C	92.0/8.0	23/2	25	1a	^{855}Asn/His	Missense	
2937	A	A/G	93.1/6.9	27/2	29	1a	^{860}Glu/Glu	Silent	
2964	T	T/G	93.5/6.5	29/2	31	1a	^{869}Phe/Leu	Missense	
2966	A	A/G	76.7/23.3	23/7	30	1a	^{870}Glu/Gly	Missense	
2970	A	A/G	83.9/16.1	26/5	31	1a	^{871}Gly/Gly	Silent	
3009	C	C/T	95.7/4.3	90/4	94	1a	^{884}Leu/Leu	Silent	
3014	A	A/T/G	63.7/22.2/14.0	109/38/24	171	1a	^{886}Asp/Val/Gly	Missense	
3021	A	A/G/T	77.9/14./7.9	314/57/32	403	1a	^{888}Gly/Gly/Gly	Silent	
3405	A	A/T/G	87.5/7.3/5.2	84/7/5	96	1a	^{1016}Val/Val/Val	Silent	A1pp
3406	A	A/G/T	79.2/11.3/9.4	84/12/10	106	1a	^{1017}Ile/Val/Phe	Missense	A1pp
3483	C	C/T	92.9/6.3	234/16	252	1a	^{1042}Gly/Gly/Gly	Silent	A1pp
4049	A	A/G	94.7/5.3	72/4	76	1a	^{1230}Asp/Gly	Missense	Viral protease, Pfam 08715
4147	T	T/C	95.8/4.2	92/4	96	1a	^{1264}Cys/Arg	Missense	Viral protease, Pfam 08715
4156	A	A/G/T	85.8/9.0/5.2	133/14/8	155	1a	^{1266}Ser/Gly/Cys	Missense	Viral protease, Pfam 08715
4928	A	A/G	79.5/20.5	435/53	488	1a	^{1524}Asp/Gly	Missense	
4931	A	A/G/T	84.9/10.2/4.9	465/56/27	548	1a	^{1525}Glu/Gly/Gly	Missense	
4933	A	A/G/T	86.2/11.9/1.9	500/69/11	580	1a	^{1526}Tyr/Cys/Phe	Missense	
5144	T	T/G	95.5/4.5	279/13	292	1a	^{1596}Val/Gly	Missense	
5152	C	C/T	95.2/4.3	376/17	395	1a	^{1599}His/Tyr	Missense	
5371	G	G/A/C	91.5/4.3/4.3	43/2/2	47	1a	^{1672}Ala/Thr/Pro	Missense	
5376	T	T/C	88.0/12.0	22/3	25	1a	^{1673}Thr/Thr	Silent	

5381	T	T/C	95.2/4.8	20/1	21	1a	[1675]Phe/Ser	Missense	
5405	T	T/G	93.5/6.5	29/2	31	1a	[1683]Val/Gly	Missense	
5409	A	A/G	93.9/6.1	31/2	33	1a	[1684]Val/Glu	Missense	
5509^5510	_	_/A	89.0/11.0	67/4	71	1a	-		
5676	A	A/T	93.8/6.3	30/2	32	1a	[1773]Ser/Ser	Silent	
5679	A	A/T	94.1/5.9	32/2	34	1a	[1774]Leu/Phe	Missense	
5681	G	G/A	94.1/5.9	32/2	34	1a	[1775]Cys/Tyr	Missense	
5845	T	T/G	95.7/4.3	45/2	47	1a	[1830]Phe/Val	Missense	
5855	A	A/G	95.3/4.7	41/2	43	1a	[1833]Asn/Ser	Missense	
5860	C	C/G/T	88.3/6.7/5.0	53/4/3	60	1a	[1835]Leu/Val/Phe	Missense	Transmembrane domain
5864	A	A/G	88.3/11.7	53/7	60	1a	[1836]Tyr/Cys	Missense	Transmembrane domain
5981	A	A/G	93.4/6.6	170/12	182	1a	[1875]Asp/Gly	Missense	
6259^6260	_	_/C	95.8/4.2	434/19	453	1a	-	-	
6833	A	A/G	95.8/4.2	23/1	24	1a	[2159]Asp/Gly	Missense	
6894	T	T/C	92.3/7.7	12/1	13	1a	[2179]Asp/Asp	Silent	
6941	T	T/C	81.8/18.2	9/2	11	1a	[2195]Val/Ala	Missense	
7104	T	T/G	90.9/9.1	20/2	22	1a	[2249]Gly/Gly	Silent	
7197	T	T/A	92.0/7.7	311/26	338	1a	[2280]Tyr/Ter.	Missense	Transmembrane domain
7198	T	T/C	92.1/7.9	314/27	341	1a	[2281]Tyr/His	Missense	Transmembrane domain
7204^7205	_	_/C	95.2/4.8	416/21	437	1a	-	-	
7755	T	T/C/A	86.2/8.0/5.2	150/14/9	174	1a	[2466]Pro/Pro/Pro	Silent	
7767	G	G/A	92.4/7.6	73/6	79	1a	[2470]Met/Ile	Missense	
7778	A	A/G	95.7/4.3	45/2	47	1a	[2474]Tyr/Cys	Missense	
7810	A	A/G	88.9/11.1	32/4	36	1a	[2485]Arg/Gly	Missense	
7821	A	A/G	92.3/7.7	24/2	26	1a	[2488]Val/Val	Silent	
7844	A	A/G/T	56.8/35.1/8.1	21/13/3	37	1a	[2496]Tyr/Cys/Phe	Missense	
7956^7957	_	_/T	95.5/4.5	64/3	67	1a	-		
8001	T	T/C	90.5/9.5	19/2	21	1a	[2548]Thr/Thr	Silent	Transmembrane domain
8013	A	A/G	95.7/4.3	22/1	23	1a	[2552]Ile/Met	Missense	Transmembrane domain

8046	T	T/C	91.9/8.1	34/3	37	1a	[2563]Val/Val	Silent	Transmembrane domain
8064	T	T/C	95.7/4.3	44/2	46	1a	[2568]Val/Val	Silent	
8091^8092	_	_/A	94.6/5.4	35/2	37	1a	-	-	
8111	T	T/G	95.8/4.2	46/2	48	1a	[2585]Val/Gly	Missense	Transmembrane domain
8116	A	A/G	91.1/8.9	51/5	56	1a	[2587]Asn/Asp	Missense	Transmembrane domain
8117	A	A/G	89.5/10.5	51/6	57	1a	[2587]Asn/Ser	Missense	Transmembrane domain
8203	A	A/G	95.7/4.3	508/23	531	1a	[2616]Ser/Gly	Missense	
8592	T	T/A	88.7/11.3	47/6	53	1a	[2745]Gly/Gly	Silent	
8934	T	T/C	95.6/4.4	65/3	68	1a	[2859]Thr/Thr	Silent	Peptidase C30
8968	G	G/A	95.3/4.7	61/3	64	1a	[2871]Asp/Asn	Missense	Peptidase C30
9194	A	A/G/T	76.8/14.3/8.9	43/8/5	56	1a	[2946]Tyr/Cys/Phe	Missense	Peptidase C30
9197	T	T/G	94.8/5.2	55/3	58	1a	[2947]Val/Gly	Missense	Peptidase C30
9200	A	A/G	95.2/4.8	59/3	62	1a	[2948]Asp/Gly	Missense	Peptidase C30
9522	G	G/A	95.7/4.3	45/2	47	1a	[3055]Leu/Leu	Silent	Peptidase C30
9552	C	C/G	90.0/10.0	9/1	10	1a	[3065]Gly/Gly	Silent	Peptidase C30
9663	A	A/G	95.0/5.0	38/2	40	1a	[3102]Ala/Ala	Silent	Transmembrane domain
9774	G	G/A	93.9/6.1	46/3	49	1a	[3139]Pro/Pro	Silent	Transmembrane domain
9828	T	T/C	96.0/4.0	24/1	25	1a	[3157]Asn/Asn	Silent	Transmembrane domain
9853	T	T/C	86.7/13.3	13/2	15	1a	[3166]Phe/Leu	Missense	
9984	A	A/G	92.3/7.7	12/1	13	1a	[3209]Leu/Leu	Silent	Transmembrane domain
10006	T	T/C	90.5/9.5	19/2	21	1a	[3217]Leu/Leu	Silent	Transmembrane domain
10008	A	A/T/G	70.8/25.0/4.2	17/6/1	24	1a	[3217]Leu/Phe/Leu	Missense	Transmembrane domain
10015^10016	_	_/T	95.6/4.4	43/2	45	1a	-	-	
10139	T	T/G	93.7/6.3	178/12	190	1a	[3261]Val/Gly	Missense	Transmembrane domain
10145	A	A/T	90.7/9.3	180/9	189	1a	[3263]Lys/Met	Missense	
10949	C	C/T	94.1/5.9	64/4	68	1a	[3531]Ala/Val	Missense	NSP8
11052	A	A/G	95.7/4.3	44/2	46	1a	[3565]Val/Val	Silent	NSP8
11125	A	A/G	95.6/4.4	87/4	91	1a	[3590]Lys/Glu	Missense	NSP8
11126	A	A/T	95.6/4.4	86/4	90	1a	[3590]Lys/Met	Missense	NSP8

11529	T	T/C	90.6/9.4	29/3	32	1a	^{3724}Asp/Asp	Silent	NSP9
11540	A	A/G	75.0/25.0	3/1	4	1a	^{3728}Lys/Arg	Missense	NSP9
11567	T	T/C	94.1/5.9	32/2	34	1a	^{3737}Val/Ala	Missense	NSP9
11570	A	A/T/G	76.9/12.8/10.3	30/5/4	39	1a	^{3738}Glu/Val/Gly	Missense	NSP9
11620	A	A/G	95.1/4.1	116/5	122	1a	^{3755}Met/Val	Missense	NSP9
11986	G	G/A	92.6/7.4	213/17	230	1a	^{3877}Asp/Asn	Missense	NSP10
12040	A	A/G	95.5/4.5	107/5	112	1a	^{3895} Ile/Val	Missense	NSP10
12609	G	G/A	95.5/4.5	63/3	66	1ab	^{4085}Asp/Asn	Missense	Corona RPol N-terminal
12618	G	G/A	95.0/5.0	38/2	40	1ab	^{4088}Glu/Lys	Missense	Corona RPol N-terminal
13121	C	C/T	59.2/40.8	58/40	98	1ab	^{4255}Phe/Phe	Silent	Corona RPol N-terminal
13126	A	A/G/T	78.2/16.8/5.0	93/20/6	119	1ab	^{4257}Asp/Glu/Val	Missense	Corona RPol N-terminal
13130	T	T/G	95.8/4.2	203/9	212	1ab	^{4258}Gly/Gly	Silent	Corona RPol N-terminal
13133	A	A/G	95.7/4.3	265/21	277	1ab	^{4259}Val/Val	Silent	Corona RPol N-terminal
13272	T	T/C	93.3/6.0	500/32	536	1ab	^{4306}Ser/Pro	Missense	Corona RPol N-terminal
13912	A	A/G	95.0/5.0	152/8	160	1ab	^{4519}Asp/Gly	Missense	Pfam: RdRP1/ Corona RPol N-terminal
14402	C	C/T/G	50.6/29.5/19.9	132/77/52	261	1ab	^{4682}Gly/Gly/Gly	Silent	Pfam: RdRP1/ Corona RPol N-terminal
14556	T	T/C	94.1/5.9	635/40	675	1ab	^{4733}Cys/Cys	Silent	Corona RPol N-terminal
14584	A	A/G	93.3/6.7	166/12	178	1ab	^{4743}Asp/Gly	Missense	Corona RPol N-terminal
14590	A	A/G	94.2/4.9	211/11	224	1ab	^{4745}Glu/Gly	Missense	Corona RPol N-terminal
14627	T	T/C	83.9/16.1	504/97	601	1ab	^{4757}Ile/Ile	Silent	Corona RPol N-terminal
14648	A	A/G	94.8/4.3	688/31	726	1ab	^{4764}Val/Val	Silent	Corona RPol N-terminal
15454	T	T/A	95.8/4.2	46/2	48	1ab	^{5033}Val/Val	Silent	Corona RPol N-terminal
15533	T	T/G	95.5/4.5	42/2	44	1ab	^{5059}Asp/Glu	Missense	
15571	C	C/T	93.8/6.3	45/3	48	1ab	^{5072}Ser/Phe	Missense	
15703	A	A/G	87.5/12.5	49/7	56	1ab	^{5116}Asn/Ser	Missense	
15796	T	T/A	95.2/4.8	20/1	21	1ab	^{5147}Leu/His	Missense	AAA ATPase
15808	T	T/C	95.7/4.3	22/1	23	1ab	^{5151}Phe/Leu	Missense	AAA ATPase
15967^15968	_	_/A	91.7/8.3	286/25	311	1ab	-	-	
16482	G	G/A	95.2/4.8	20/1	21	1ab	^{5376}Ala/Thr	Missense	AAA12

16485	T	T/C	94.4/5.6	17/1	18	1ab	[5377]Tyr/His	Missense	AAA12
16556	C	C/G/T	83.3/12.5/4.2	20/3/1	24	1ab	[5400]Phe/Leu/Phe	Missense	AAA12
16561	T	T/G	90.0/10.0	27/3	30	1ab	[5402]Val/Gly	Missense	AAA12
16701	G	G/A	95.7/4.3	44/2	46	1ab	[5449]Ala/Thr	Missense	AAA12
16749	A	A/G	70.9/29.1	124/51	175	1ab	[5465]Ser/Gly	Missense	NSP11
16919	T	T/C	92.0/8.0	297/26	323	1ab	[5521]His/Gln	Missense	NSP11
16986	G	G/A	95.7/4.3	180/8	188	1ab	[5544]Ala/Thr	Missense	NSP11
17258^17259	_	_/G	92.7/7.3	89/7	96	1ab	-	-	
17262	A	A/G	94.8/5.2	73/4	77	1ab	[5636]Thr/Ala	Missense	NSP11
17263	C	C/G	92.6/7.4	75/6	81	1ab	[5636]Thr/Ser	Missense	NSP11
17270	C	C/T	92.5/7.5	172/14	186	1ab	[5638]Cys/Cys	Silent	NSP11
17271	C	C/T	92.0/7.5	184/15	200	1ab	[5639]His/Tyr	Missense	NSP11
17740	A	A/G/T	81.0/14.8/4.2	252/46/13	311	1ab	[5795]Asp/Gly/Val	Missense	NSP11
18045	T	T/C/A	81.2/11.5/6.9	457/65/39	563	1ab	[5897]Ser/Pro/Thr	Missense	NSP11
18219	T	T/G	94.5/5.5	605/35	640	1ab	[5955]Trp/Gly	Missense	NSP11
18466	G	G/A	90.0/10.0	459/51	510	1ab	[6037]Arg/His	Missense	NSP11
18472	T	T/C	94.9/4.1	300/13	316	1ab	[6039]Leu/Ser	Missense	NSP11
18473	G	G/A	89.7/10.0	269/30	300	1ab	[6039]Leu/Leu	Silent	NSP11
18480^18481	_	_/C	95.1/4.9	193/10	203	1ab	-	-	
18984	T	T/C	80.0/20.0	8/2	10	1ab	[6210]Leu/Leu	Silent	
19085	A	A/G	94.6/5.4	53/3	56	1ab	[6243]Val/Val	Silent	
19498	C	C/T	92.8/7.2	568/44	612	1ab	[6381]Ala/Val	Missense	NSP13
19699	G	G/A	90.6/8.7	280/27	309	1ab	[6448]Arg/Lys	Missense	NSP13
19768	A	A/G	96.4/3.6	162/6	168	1ab	[6471]Lys/Arg	Missense	NSP13
19936	A	A/G/T	87.6/7.6/4.4	298/26/15	340	1ab	[6527]Asp/Gly/Gly	Missense	NSP13
19937	C	C/T	90.9/8.1	318/32	350	1ab	[6527]Asp/Asp	Silent	NSP13
19938	A	A/T/G	89.3/6.5/4.0	357/26/16	400	1ab	[6528]Ser/Cys/Gly	Missense	NSP13
20091	G	G/A	92.9/7.1	222/17	239	1ab	[6579]Glu/Lys	Missense	NSP13
20099	G	G/A/C	81.1/13.6/5.3	107/18/7	132	1ab	[6581]Lys/Lys/Asn	Missense	NSP13

21097	G	G/A	90.6/9.4	319/33	352	S (S1)	^{317}Glu/Lys	Missense	HVR3
21507	T	T/C	95.1/4.9	77/4	81	S (S1)	^{454}Gly/Gly	Silent	
21599	C	C/T	90.6/9.4	211/22	233	S (S1)	^{484}Ser/Phe	Missense	
21804	A	A/G	91.5/8.5	388/36	424	S (S2)	^{552}Glu/Glu	Silent	
22151	T	T/G	95.8/4.2	46/2	48	S (S2)	^{668}Ile/Ser	Missense	
22239	A	A/T	91.0/9.0	81/8	89	S (S2)	^{697}Lys/Asn	Missense	
22299	T	T/G	90.6/9.4	48/5	53	S (S2)	^{717}Gly/Gly	Silent	
22351	G	G/T	95.7/4.3	45/2	47	S (S2)	^{734}Ala/Ser	Missense	
22380	C	C/T/G	48.6/33.0/18.4	90/61/34	185	S (S2)	^{744}Gly/Gly/Gly	Silent	
22535	G	G/A	94.6/4.6	844/41	892	S (S2)	^{796}Ser/Asn	Missense	
22794	T	T/A/C	81.0/11.9/7.1	34/5/3	42	S (S2)	^{882}Thr/Thr/Thr	Silent	
22794^22795	__	__/A_/AA	87.8/8.2/4.0	43/4/2	49	S (S2)	-	-	
22795	A	A/G	94.7/5.3	36/2	38	S (S2)	^{883}Lys/Glu	Missense	
22827	T	T/C	92.6/7.4	25/2	27	S (S2)	^{892}Ser/Ser	Silent	
22845	T	T/G	92.6/7.4	25/2	27	S (S2)	^{899}Cys/Trp	Missense	
22847	G	G/A	96.6/3.4	28/1	29	S (S2)	^{900}Gly/Asp	Missense	
22849	A	A/T	93.1/6.9	27/2	29	S (S2)	^{901}Ser/Cys	Missense	
22854	A	A/T	93.9/6.1	31/2	33	S (S2)	^{902}Gly/Gly	Silent	
22952	T	T/G	92.0/8.0	80/7	87	S (S2)	^{935}Val/Gly	Missense	
22965	A	A/G/T	85.1/10.1/4.8	143/17/8	168	S (S2)	^{939}Val/Val/Val	Silent	
23258	T	T/C	91.8/8.2	56/5	61	S (S2)	^{1037}Val/Ala	Missense	
23317	A	A/T	94.4/5.6	17/1	18	S (S2)	^{1057}Asn/Tyr	Missense	
23381	A	A/G	61.5/38.5	8/5	13	S (S2)	^{1078}Tyr/Cys	Missense	Transmembrane domain
23389	C	C/G/T	63.2/21.1/15.8	12/4/3	19	S (S2)	^{1081}Leu/Val/Ser	Missense	Transmembrane domain
23775	C	C/T	95.1/4.9	193/10	203	3b	^{6}Ala/Val	Missense	
23889	G	G/A	82.2/17.8	37/8	45	3b	^{44}Ser/Asn	Missense	
23894	A	A/T/G	87.0/7.4/5.6	47/4/3	54	3b	^{46}Ser/Cys/Gly	Missense	
23965	A	A/G	94.0/6.0	63/4	67	E	^{12}Asn/Asp	Missense	
23982	G	G/A	95.0/5.0	38/2	40	E	^{18}Ala/Thr	Missense	Transmembrane domain

24004	G	G/T	65.6/34.4	21/11	32	E	^{25}Cys/Phe	Missense	Transmembrane domain
24065	T	T/G	90.9/9.1	30/3	33	E	^{45}Cys/Trp	Missense	
24070	T	T/C	93.9/6.1	31/2	33	E	^{47}Leu/Ser	Missense	
24071	A	A/G	93.3/6.7	28/2	30	E	^{47}Leu/Leu	Silent	
24405	C	C/T/G	82.3/10.4/7.3	158/20/14	192	M	^{90}Cys/Cys/Trp	Missense	Transmembrane domain
24412	C	C/T	89.6/10.4	242/28	270	M	^{93}Pro/Ser	Missense	Transmembrane domain
24455	C	C/T	81.7/18.1	482/107	590	M	^{107}Pro/Leu	Missense	
24695	A	A/G	92.9/5.1	3514/184	3702	M	^{187}Glu/Gly	Missense	
25168	A	A/G	88.7/11.3	360/46	406	4b	^{86}Arg/Gly	Missense	
25168	A	A/G	88.7/11.3	360/46	406	4c	^{17}Val/Val	Silent	
25249	C	C/T	92.8/7.2	1102/85	1187	non-coding	-	-	
25448	G	G/A	92.4/6.9	281/21	304	5a	^{60}Ala/Thr	Missense	
25772	T	T/A	95.4/4.6	62/3	65	N	^{39}Ile/Lys	Missense	
25791	T	T/C/G	80.6/13.9/5.6	29/5/2	36	N	^{45}Asn/Asn/Lys	Missense	
25792	T	T/A	84.4/15.6	27/5	32	N	^{46}Ser/Thr	Missense	
25797	C	C/A	93.8/6.3	30/2	32	N	^{47}Pro/Pro	Silent	
25799	A	A/C	93.5/6.5	29/2	31	N	^{48}Gln/Pro	Missense	
25811	A	A/G	95.5/4.5	21/1	22	N	^{52}Glu/Gly	Missense	
25812	A	A/G	95.2/4.8	20/1	21	N	^{52}Glu/Glu	Silent	
25816	A	A/T/G	61.9/21.4/16.7	26/9/7	42	N	^{54}Ser/Cys/Gly	Missense	
26006	C	C/G	96.3/3.2	180/6	187	N	^{117}Ala/Gly	Missense	
26277	G	G/A	94.7/5.3	524/29	553	N	^{207}Gln/Gln	Silent	
26455	C	C/T	85.5/14.5	1047/177	1224	N	^{267}Arg/Cys	Missense	
26637	A	A/G	93.0/7.0	4268/319	4587	N	^{327}Val/Val	Silent	
26835	G	G/A	89.2/10.8	461/56	517	N	^{393}Glu/Glu	Silent	
26971	A	A/G	93.9/6.1	526/34	560	6b	^{26}Asp/Gly	Missense	

2.6 Protein secondary structure prediction

Protein structures for S1 and S2 were predicted in RaptorX, a structure prediction server that predicts three dimensional (3D) structures for protein sequences without close homologs in the Protein Data Bank (PDB) (Kallberg *et al*., 2012). S1 and S2 3D structures were annotated and superposed in CCP4MG v2.9.0 using the Secondary Structure (SSM) superposition method. This method superimposes pairs of structures by: (1) finding the secondary structure elements (SSEs) and representing them as one simple vector spanning the length of the SSE; (2) finding equivalent SSEs in the two structures using graph-theory matching by geometric criteria of distances and angles between the vectors ; (3) superimposing vectors representing equivalent SSEs ; (4) finding the most likely equivalent residues in the superposed SSEs; (5) superimposing CA atoms of equivalent residues ; and (6) iterating the last two steps.

3. Results and Discussion

3.1 Genome assembly and annotation

The genome sequence of QX-like strain ck/ZA/3665/11 was assembled from 74, 578 IBV-specific paired-end reads of 144 bp each. The genome was 27, 388 nt in length with the 5' UTR incomplete by ~139 nt. Thirteen ORFS were predicted by VIGOR in the order 5'-UTR-1a-1ab-S-3a-3b-E-M-4b-4c-5a-5b-N-6b-3'UTR (Fig. 1). This genome organisation including 4b, 4c and 6b was similar to that of turkey coronavirus (TCoV; Cao et al., 2008), and the ORFs 4b, 4c and 6b were also predicted in Australian IBV strains (Hewson et al., 2011). When the sequences for a QX-like sequence (JQ088078) and ArkDPI (EU418976) were analysed using VIGOR, a similar genome arrangement was detected. Mass41 (AY851295) did not however contain the predicted 4b, 4c and 6b ORFs (data not shown).

ORF 4b was 94 amino acids (aa) in length and no SMART domains were predicted, whereas ORF 4c was 56 aa in length and a low complexity region was identified. The 6b ORF encoded a 74 aa protein with a signal peptide predicted from residues 1 to 24 and two transmembrane domains from residues 2 to 25 and 35 to 57. No recombination was detected across the genome of QX-like strain ck/ZA/3665/11.

A gap was present between nucleotides 1666 and 1667 (Table 1, supplementary data) (~ aa 370 in the 1a ORF). Although the gap was present in the majority (74.6%) of reads, the sequence for strain 3665/11 deposited in Genbank contains the minority adenine residue because the gap introduced a frame shift, splitting ORF 1a into two. It may be a legitimate mutation, but until further transcriptional analyses are conducted, the ORF 1a gene has been reported intact here.

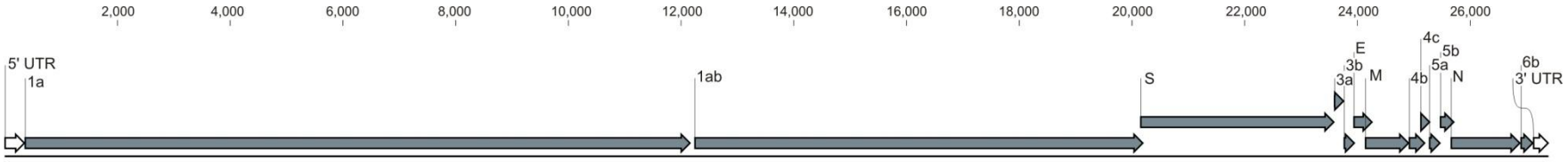


Figure 1: Genome organisation of QX-like IBV strain 3665/11

3.2 Comparative frequency of mutations in ORFs

Two hundred and eight SNPs across the IBV QX-like genome were evaluated at the selected cut-off value. In Table 1 the consensus reference is juxtaposed with the allele variations, the relative frequencies of these point mutations, the actual number of counts and coverage at that position, the corresponding ORF or region and the mutational effect. Coverage ranged from 4 fold (position 11540) up to 4587 fold (position 26637).

The relative frequencies of missense: silent SNPs in relation to ORF length were calculated in order to obtain a comparative measure of variability in specific genes (Table 2). Results for the structural genes and polymerase are illustrated in Fig. 2, and the results for the non-structural protein ORFs and non-coding regions, which were much shorter in length, are presented in Fig.3. Overall the most variable ORFs in terms of total SNPs, in descending order, were: E, 3b, 5' UTR, N, 1a, S, 1ab, M, 4c, 5a, 6b (no SNPs were detected at the 5% cut-off in the 3a and 3' UTR regions).The most variable, as assessed by SNPs leading to missense mutations, in descending order, were: 3b, E, 5' UTR, 1a, N, M, 5a, 1ab, S, 4b, 3a/4c/6b. These mutations presumably did not affect the tertiary protein structure and might be advantageous to the virus. The ORFs under the strongest positive selection pressure as indicated by the proportion of synonymous mutations, were, in descending order, 4c, 1ab, N, S, E, 3a/3b/M/4b/5a/6b.

3.2.1 E protein

The E protein ORF had significantly more missense mutations on average, at a frequency of 1.5% of the ORF, which is more than threefold higher than the average value (0.55) for the 1a, 1ab, S, M and N genes. The E protein gene was the most variable at the sub-consensus

Table 2. Relative frequencies of SNPs in QX-like IBV strain 3665/11

Region/ORF	Length in nt (% of genome)	SNP count frequency		
		Total SNPs (as a % of ORF)	# mutations (as a % of ORF)	# of silent mutations (as a % of ORF) [as a % of total SNPs]
5'UTR	357 (1.30)	4 **(1.12)**	4 **(1.12)**	-
1ab	19872 (72.55)	152 **(1.47)**	112 (1.08)	40 (0.39) **[52.96]**
S	3438 (12.55)	23 (0.67)	15 (0.43)	8 (0.23) **[34.78]**
3a	174 (0.64)	0	0	0
3b	192 (0.70)	3 **(1.56)**	3 **(1.56)**	0
E	333 (1.22)	6 **(1.80)**	5 **(1.50)**	1 (0.30) [16.67]
M	681 (2.49)	4 (0.59)	4 (0.59)	0
4b	285 (1.04)	1 (0.35)	1 (0.35)	0
4c	171 (0.62)	1 (0.58)	0	1 (0.58) [100]
5a	198 (0.72)	1 (0.51)	1 (0.51)	0
N	1230 (4.49)	13 **(1.06)**	8 (0.65)	5 (0.41) **[38.46]**
6b	222 (0.81)	1 (0.45)	1 (0.45)	0
3'UTR	272 (0.99)	0	0	0
Total genome	27388	208 (0.76)		Total SNPs evaluated

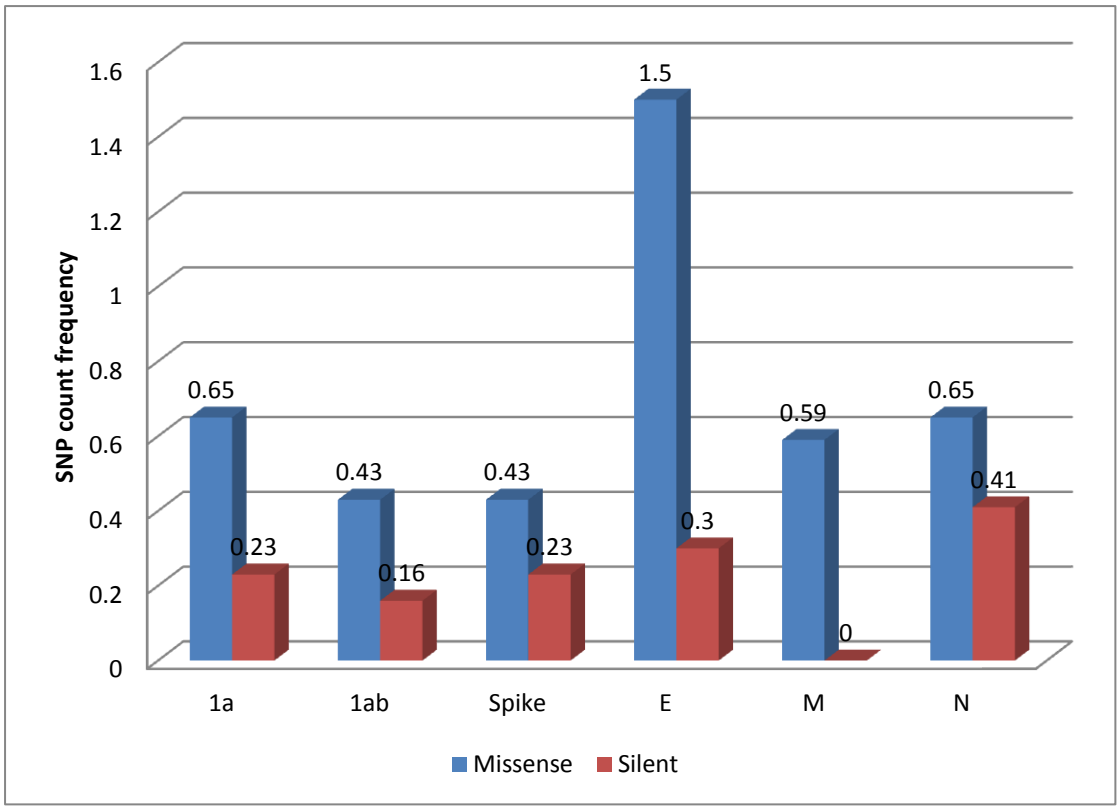


Figure 2: Relative frequencies of mutations in structural and polymerase ORFs of QX-like IBV strain 3665/11

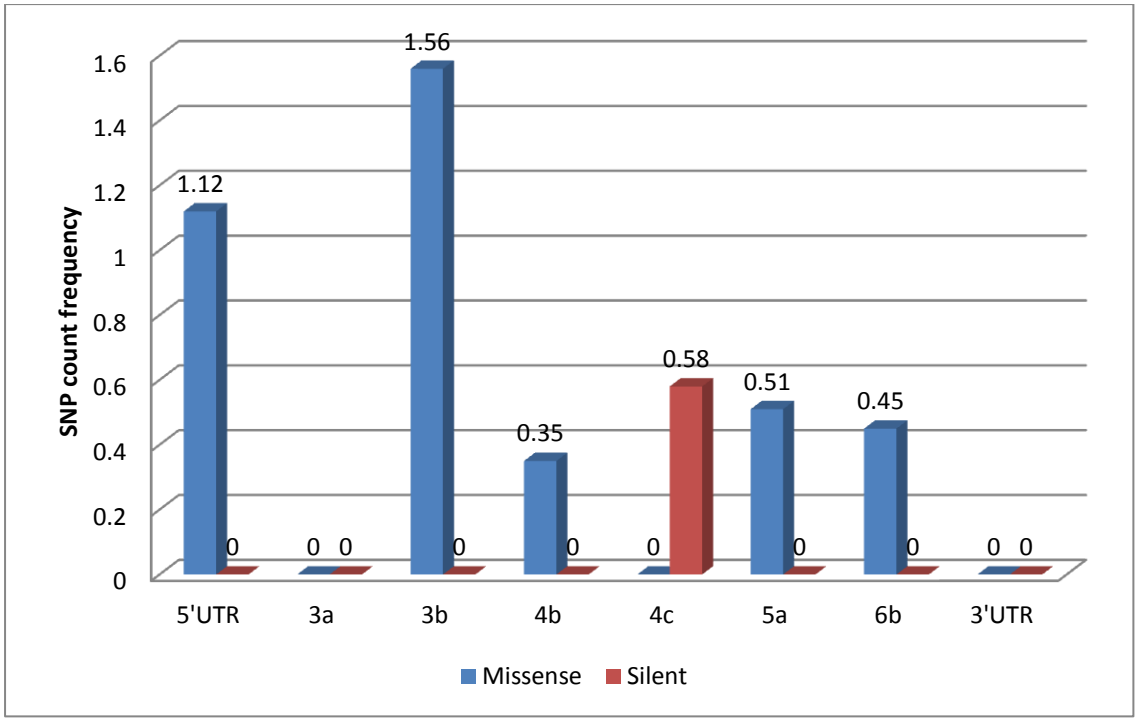


Figure 3: Relative frequencies of mutations in non-structural ORFs and non-coding regions of QX-like IBV strain 3665/11.

level, with 5 missense mutations and only one silent mutation across its 333 bp ORF. Despite its small size, the CoV E protein drastically influences the replication of CoVs and their pathogenicity. In the SARS-CoV, it was experimentally demonstrated that the E protein is not essential for genome replication or subgenomic mRNA synthesis, but it does affect morphogenesis, budding, assembly, intracellular trafficking and virulence. In fact, in SARS-CoV the E protein is the main antagonist associated with induction of inflammation in the lung, which causes the acute respiratory distress syndrome from which the virus derives its name (DeDeigo et al., 2014). No studies have been published for the IBV E protein, but the high variability demonstrated here suggests that it may be an important virulence factor in poultry, and that a higher mutation rate possibly provides an evolutionary advantage in overcoming host cellular immune responses.

3.2.2 N protein

Although the N protein gene contained one of the highest overall frequencies of SNPs (1.06%), the N gene is evidently under greater selective pressure, since 38.9% of these mutations (0.41% as a total of the gene) were silent. The coronavirus N protein is multifunctional, playing vital roles in viral assembly and formation of the complete virion and is required for optimal viral replication. Additionally, the CoV N protein is implicated in cell cycle regulation and host translational shutoff, displays chaperone activity, activates host signal transduction and aids viral pathogenesis through the antagonism of interferon induction (reviewed by McBride et al., 2014). Given its fundamental roles in RNA binding, formation of the ribonucleoprotein complex and in the virion, it is not surprizing that this structural protein is the most conserved, as evidenced by its gene having the highest ratio of silent mutations of all the genes analysed. The importance of maintaining the sequence integrity in the N protein in IBV was demonstrated by Kuo et al (2013), who reported that two residues within the N-terminal domain of a Taiwanese IBV strain were positively

selected, and that mutation of either of these significantly reduced the affinity of the N protein for the viral transcriptional regulatory sequence.

3.2.3 M protein

The glycosylated amino terminus of the M protein lies on the outside of the virion and M spans the membrane structure three times (Collison et al, 2000). All four SNPS in the M gene resulted in missense mutations, two of which were located in the predicted transmembrane region. The M protein plays an important role in CoV virion formation. IBV M protein co-expressed with S assembled into virus-like particles (Liu et al., 2013) confirming its major role in virion formation, but CoV M proteins also interact with other proteins and perform other roles in the infected cell. For example, M together with the accessory proteins 4a, 4b and 5 were all found to prevent the synthesis of IFN-β through the inhibition of interferon promotor activation and IRF-3 function, thus influencing disease outcome (Yang et al., 2013)

3.2.4 Accessory proteins

Coronavirus accessory proteins are generally dispensable for virus replication, but they play vital roles in virulence and pathogenesis by affecting host innate immune responses, encoding pro- or anti-apoptotic activities, or by effecting other signalling pathways that influence disease outcomes (Susan & Julian., 2011). IBV was demonstrated to induce a considerable activation of the type I IFN response, but it was delayed with respect to the peak of viral replication and accumulation of viral dsRNA (Kint et al., 2014). IBV accessory proteins 3a and 3b play a role in the modulation of this delayed IFN response, by regulating interferon production at both the transcriptional and translational levels. Interestingly, IBV proteins 3a and 3b seem to have opposing effects on IFN production in infected cells: 3a seems to promote IFN production, and 3b is involved in limiting IFN production, antagonising each other to tightly regulate IFN production (Kint et al., 2014). Field isolates lacking 3a and 3b

displayed reduced virulence *in vitro* and *in vivo* (Mardani et al., 2008). ORF 3a in strain 3665/11 lacked SNPs, but ORF 3b in had the highest frequency of SNPs relative to its size (n=3; 1.56%).

ORF 4b is present in many international IBV strains (Hewson et al., 2011; Bentley et al., 2012) but is rarely mentioned in the literature since a canonical transcription regulatory sequence (TRS-B) could not be identified upstream of the encoding RNA. However, Bentley et al. (2012) demonstrated that IBV was capable of producing subgenomic mRNAs from noncanonical TRS-Bs via a template-switching mechanism with TRS-L, the conserved TRS in the leader sequence in the 5' UTR, which may expand the *Gammacoronavirus* repertoire of proteins. They specifically demonstrated the transcription of the 4b ORF by this mechanism. Although no studies have been performed determining ORF 4b's functional role in the pathogenesis of IBV, the homolog in MERs-CoV is a potent interferon antagonist (Yang et al., 2013). A single SNP causing a missense mutation was present in 11.3% of the sub-consensus population of the 4b ORF in this study. The single mutation in ORF 4c was silent, and the predicted protein contained a low complexity region. Low complexity regions are regions of protein sequences with biased amino acid composition, and may be involved in flexible binding associated with specific functions (Coletta et al., 2010).

ORF 6b, a 73 aa protein with a signal peptide and two transmembrane domains, was identified in the genome of strain 3665/11, and ORF 6b was also reported in TCoV and Australian IBV strains (Cao et al, 2008; Hewson et al., 2011). The homolog in SARS-CoV is 63 aa in length and was identified as an endoplasmic reticulum/ Golgi membrane-localised protein that induces apoptosis. Apoptosis may play an important role in promoting CoV dissemination *in vivo*, minimising inflammation and aiding evasion of the host's defence mechanisms (Ye et al., 2010). Protein 6 from SARS-CoV accelerated the replication of murine CoV, increasing the virulence of the original attenuated virus (Tangudu et al., 2007).

Presumably, this accessory protein plays a similar role in IBV pathogenesis, although this remains to be determined experimentally.

3.2.5 Untranslated regions (UTRs)

The 3' UTRs of CoV genomes contain conserved *cis*-acting sequence and structural elements that play essential roles in RNA synthesis, gene expression and virion assembly, and each sub-genomic RNA contains a 5' leader segment that is identical to this 3' UTR region of the genome (Goebel et al., 2004; Sola et al., 2011). No SNPs were detected in the 3' UTR in the sub-consensus sequences of strain 3665/11, which is consistent with the vital regulatory role that this region plays. Conversely, the partial 5' UTR sequence of strain 3665/11 was highly variable.

The un-sequenced 139 nucleotides from the 5' end of the genome were extrapolated from the most similar genomic sequence, that strain ITA/90254/2005, and the secondary RNA structure of the 5' UTR for 3665/11 was predictively folded (Fig. 5). The SNPs were then systematically substituted into the consensus sequence and RNA folding repeated. Delta G values for the predicted RNA secondary structures in Figs. 4(a) to 4(h) varied from -182.9 kcal/mol to -185.5 kcal/mol. Apart from the ^{148}C to G mutation (Fig. 4(c)), effects on RNA secondary structure were minor and the structures in Figs. 4(b) and 4(d) to 4(h) were similar. To assess the effect of combining mutations, an RNA containing ^{148}T(U), ^{164}T(U), ^{169}G and ^{170}C was folded, and this resulted in a similar stem-loop structure to those in Figs. 4(a), 4(b) and 4(d) to 4(h) (data not shown). Apart from the mutation ^{148}C to G, the SNPs had little effect on the secondary RNA structure in the 5' UTR.

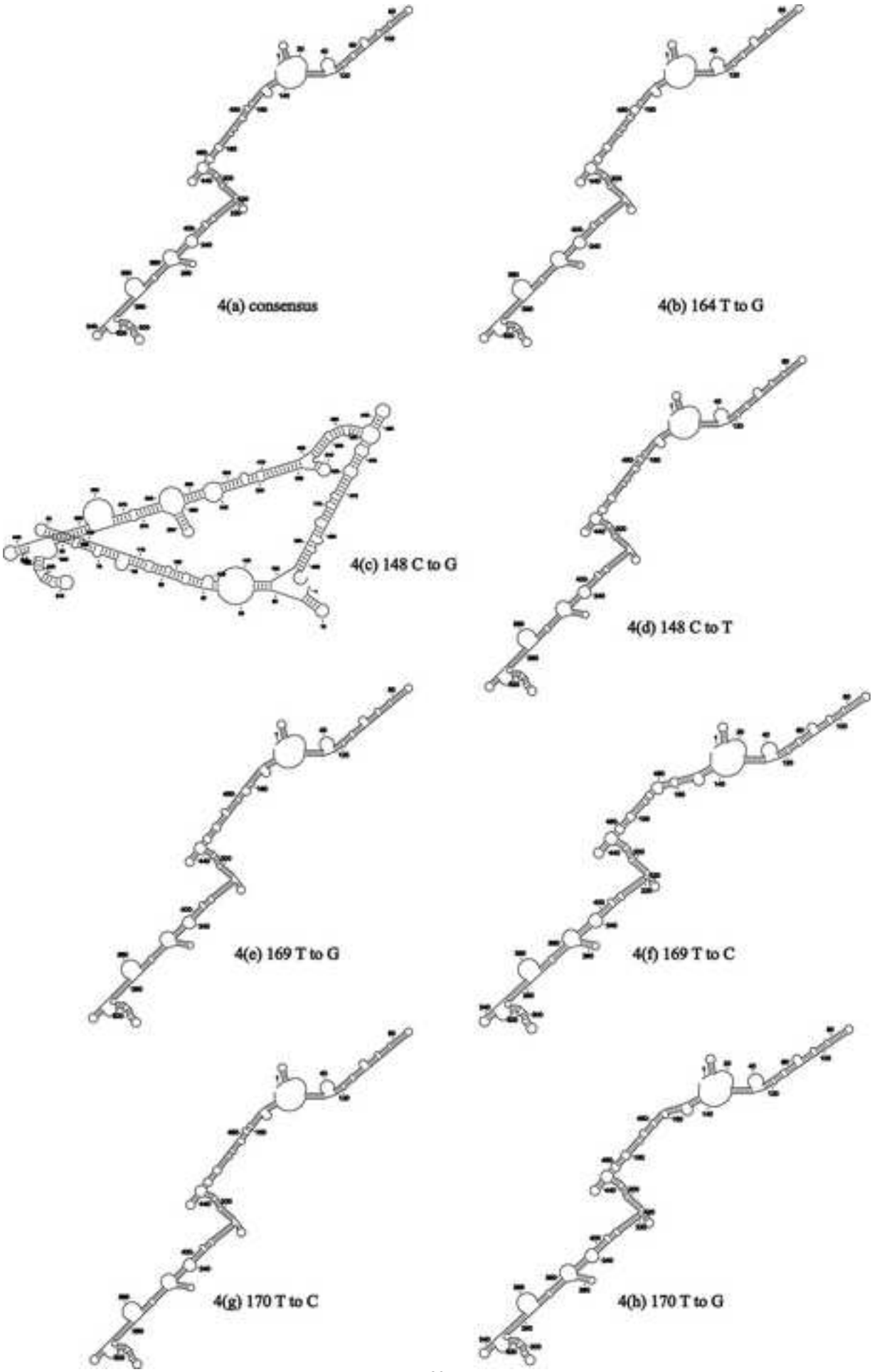


Figure 4: Effects of mutations on the predicted RNA structures in the 5' UTR of QX-like IBV strain 3665/11

3.3 The Spike protein

Twenty three SNPS were identified in the 3438 bp spike protein ORF; 15 of these resulted in missense mutations at the amino acid level, and 8 were silent mutations. The frequency of total SNPs in the S protein ORF was below average, at 0.67%, compared to the genome average of 0.76%. It was anticipated that the majority of mutations in the S ORF would be in the S1 gene, particularly in the HVRs, but, surprisingly, this was not the case. Only three of these SNPS (two missense and one silent) were found in the S1 gene, and all three were located in the COOH-terminal half of the S1 protein (Figure 5). Only one mutation, a missense mutation, mapped to HVR3. No SNPS were detected in HVR1 or HVR2. The S2 subunit was considerably more variable, containing 87% of the polymorphisms detected across the entire S protein.

Two other notable features of S2 were detected: the first was a multi-A insertion site located between nucleotides 22794 and 22795 in the genome. The polymorphism involved the insertion of either one or two adenine nucleotides, possibly via a mechanism of polymerase stuttering. The second region of interest was located in close proximity, just downstream of the multi-A insertion site: a stretch of three consecutive mutated amino acids, namely ^{889}C→W, ^{890}G→D, ^{891}S→C followed by silent mutation ^{892}G (Figure 5).

Template-based protein structure modelling was used to predict the secondary structure of the IBV spike monomer, based on the available crystal structure for the MERS-CoV S1 and S2 subunits (Fig. 6). S1 and S2 were modelled separately in Raptor X and then superposed. The IBV S1 structure was arranged as two beta barrels and S2 formed packed α-helices. The S2 protein was not complete and the transmembrane domain was not represented since there were no sufficiently similar structures on which to model this region, but this is the first model of the spike protein monomer for IBV. HVR1 and the putative receptor binding

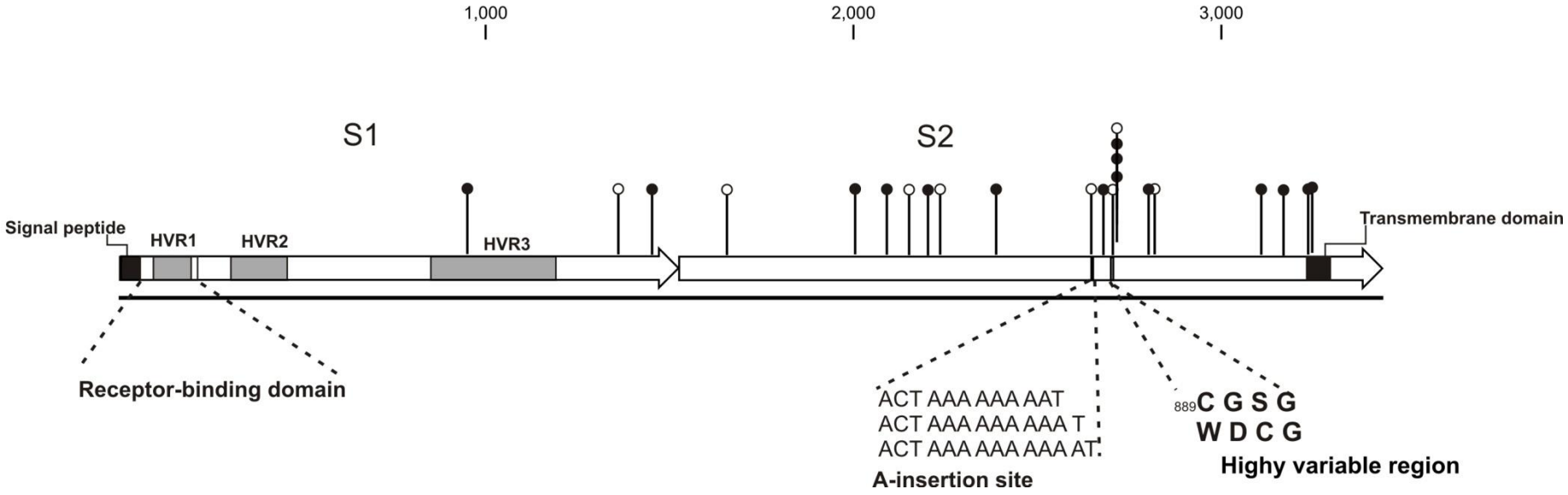


Figure 5: Schematic representation of the Spike protein of QX-like IBV strain 3665/11. Missense mutations are indicated as solid circles, and silent mutations as empty circles.

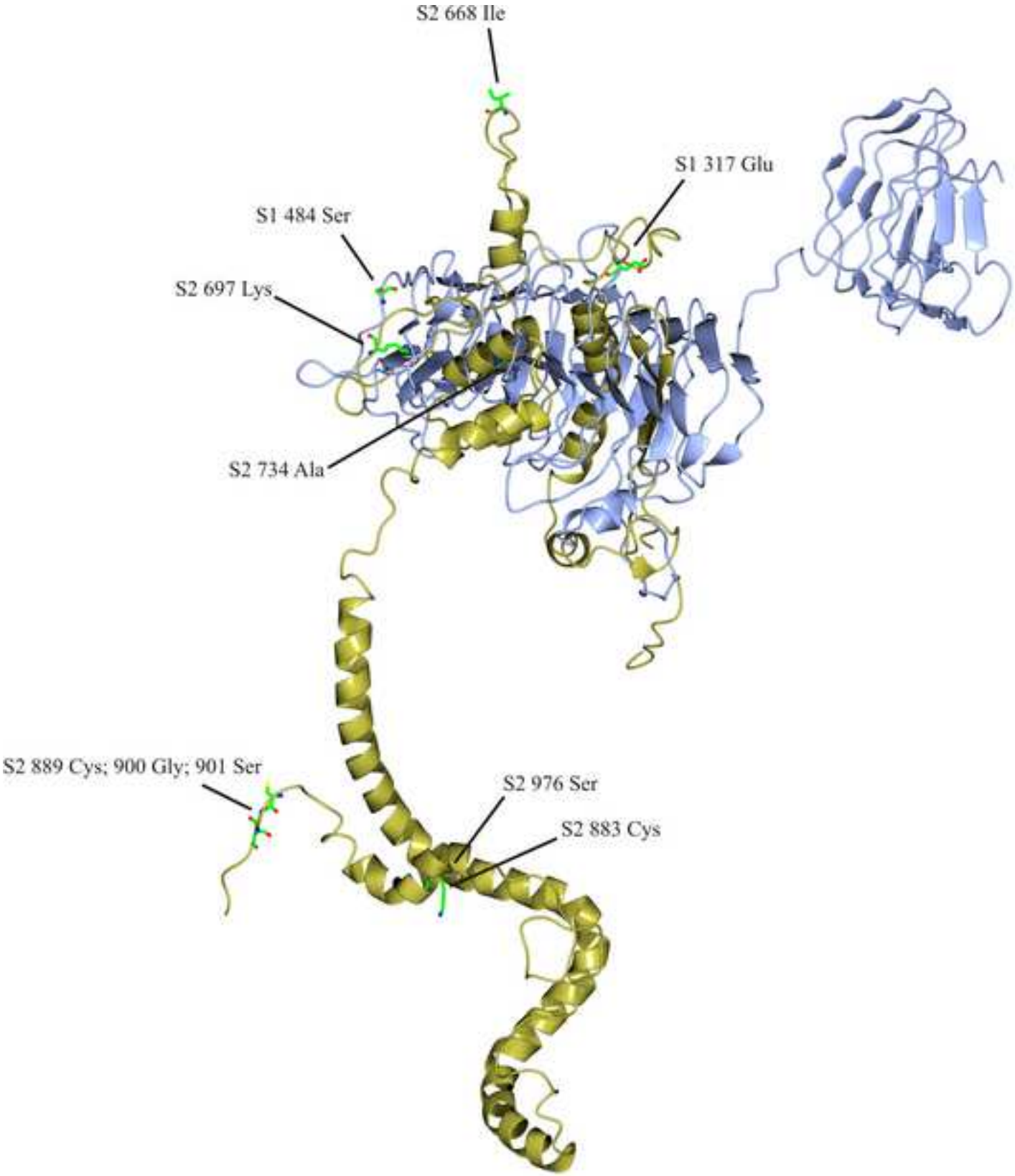
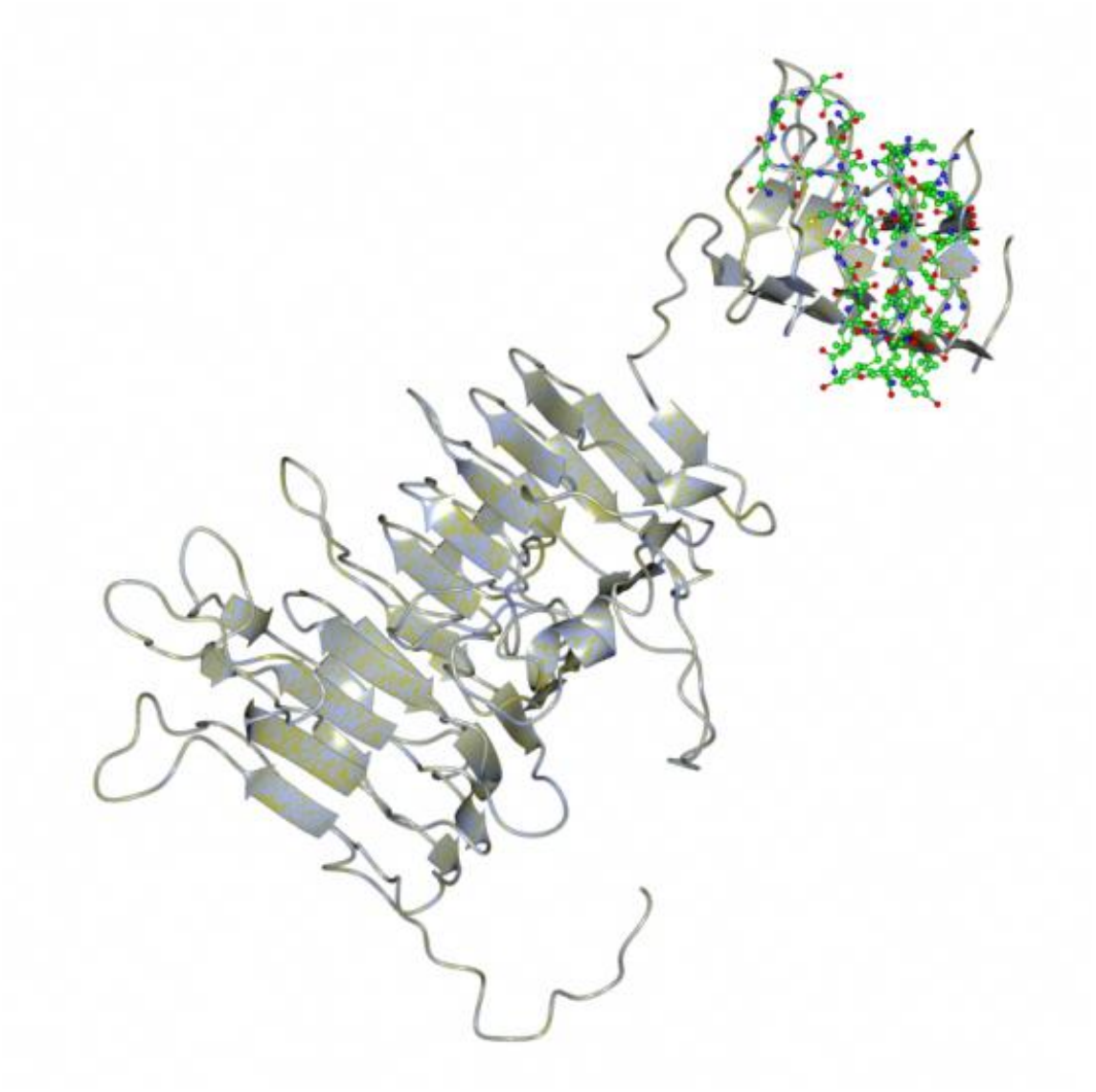


Figure 6: Predicted structure of the Spike protein monomer of QX-like IBV strain 3665/11. Missense mutations in S1 (blue) and S2 (yellow) are indicated as coloured side chains.

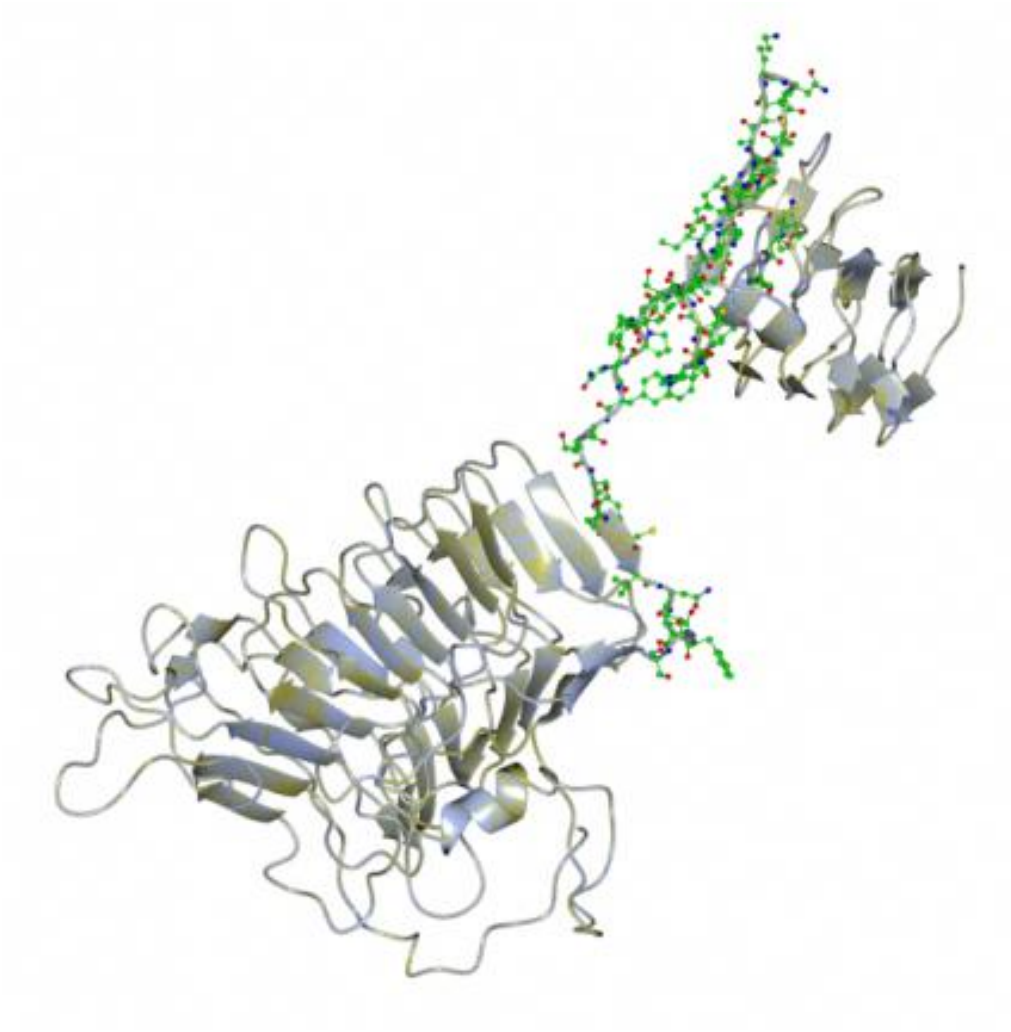
domain maps to the apical beta barrel (Fig. 7(a)) and HVR2 is located on the flat plane on the base of the apical beta barrel and the peptide connecting it to the basal beta barrel (Fig. 7(b)). HVR3 maps to a region in the basal beta barrel of S1 that was predicted to contact or interact with S2 (Fig. 7(c)). The locations of the missense mutations detected by SNP analysis in the S1 and S2 subunits are indicated in Fig. 6. Many of these SNPs mapped to codons encoding amino acids on the surface of the predicted structure, but two regions were notable. Firstly, the highly variable region in S2 spanning amino acids 889 to 901 was exposed on the S2 stalk, although folding of the remainder of the COOH domain may have influenced this conformation. Secondly, ^{668}Ile was exposed on a projection at the top of the monomer. This residue precedes the second furin cleavage site in the S2 subunit with the sequence ^{667}PISSSGR/S^{674}. The cleavage of the S1/S2 furin motif (^{517}RRRR/S^{521} in strain 3665/11) was found to be non-essential for attachment of IBV to the cell. Rather, it promotes infectivity within the cell. In studies with the Beaudette IBV strain, the second furin cleavage site in the S2 subunit was required for furin-dependent entry and syncytium formation, and the current hypothesis is that interplay between the S1 and S2 subunits determines virus attachment to specific receptors, determining tissue tropism of the virus (Promkuntod et al., 2013). The exact biological roles of these areas in S2 that are prone to mutation remain to be experimentally determined.

4. Conclusions

Archaeological remains of domestic chickens in Northeast China and the Indus Valley date back ~8000 years (West & Zou, 1988). The CoV group has been estimated to have arisen around 8100 BC, and the *Gammacoronaviruses* diverged from the CoV group around 2800 BC (Woo et al., 2012). CoVs have probably been co-evolving with their gallinaceous hosts



A



B

C

Figure 7: The predicted locations of HVR1 (7a), HVR2 (7b) and HVR3 (7c) of QX-like IBV strain 3665/11, indicated as coloured side-chains on the S1 subunit

for several thousand years. Indeed, Cook and co-authors (2012) state that “IBV is found everywhere that commercial chickens are kept”. The implication is that although IBV was only discovered some 80 years ago, the variety of serotypes we now observe are the results of hundreds if not thousands of years of genetic drift and recombination, accelerated by modern poultry farming practises where chickens are kept in high densities, and inter-regional trade in poultry and other avian species.

Studies on antigenic diversity of IBVs are heavily biased towards studies of the S1 gene, and the HVRs in particular (Cavanagh, 2007; Ducatez et al., 2009, Kant et al., 1992; Mork et al., 2014). Many of these studies cite frequent point mutations in the S1 gene, but this was not the finding of the present study. The discovery of a novel 3'-to-5' exoribonuclease activity in CoV nsp14, which regulates replication fidelity and diversity in coronaviruses (Denison et al, 2011), lends weight to the theory that genetic drift is not primarily responsible for the degree of variation and serotypes we observe in poultry nowadays.

Instead, generation of variation by recombination is likely the main mechanism of serotypic diversity. The high frequency of RNA recombination in coronaviruses is likely caused by their unique mechanism of RNA synthesis, which involves discontinuous transcription and polymerase jumping (Jeong et al., 1996). Sequencing of many field strains has provided convincing evidence that many, possibly all, IBV strains are recombinants between different field strains (Cavanaugh, 2007, Kuo et al., 2013, Liu et al., 2013, Hewson et al., 2011), driving IBV evolution at a population level. Recombination of distinct IBV strains has been experimentally demonstrated *in vitro*, *in ovo* and *in vivo* (Kottier et al., 1995; Wang et al., 1997).

The S1 subunit HVR1 contains the IBV receptor-binding site. Therefore despite the sequence variability in this region (which includes insertions and deletions), diverse strains must retain

this critical biological function. All three HVRs may represent ancient artefacts of recombination, which have been perpetuated because they retain receptor-binding properties, with minimal permissive amino acid changes. This theory contrasts the tenet that the HVRs in the S1 subunit are very tolerant of amino acid changes produced by genetic drift, thereby conferring a selective advantage (Cavanagh, 2007, De Wit, 2000; Kant et al, 1992; Koch et al., 1990).

Whereas S1 fulfils a primary role in receptor binding (Promkuntod et al, 2014), a broader role of S2 in antigenicity and attachment to receptors is emerging. Chickens primed with a recombinant-expressed S2 subunit of a virulent ArkDPI strain and boosted with a live Mass-type vaccine were protected against challenge with live virulent ArkDPI virus (Toro et al. 2014). Although S2 subunits most likely do not contain an additional independent receptor-binding site, S2 in association with S1 forms part of a specific ectodomain which is critical to the binding of the virus to chicken tissues, which implies that both S1 and S2 contain determinants important to viral host range (Promkuntod et al., 2013). The results of the present study demonstrate that S2 is more predisposed to mutations than S1, providing an adaptive advantage and at least one other study has reported higher variability in S2 compared to S1 (Mo et al. 2012).

IBV has not been as extensively studied as other CoVs, and little progress has been made in effectively controlling or eradicating the disease in poultry. Experimental and field studies provide substantial evidence that use of a homologous IBV vaccine is best, but sometimes, intriguingly, protection can be offered by an unrelated vaccine, or by the use of two heterologous vaccines (Jones 2010). Genotyping and phylogenetic analysis of IBV are typically focused on the S1 subunit sequence, and Liu et al (2014) caution against drawing conclusions based on a single gene sequence, particularly a partial gene sequence. The roles of the IBV E and accessory proteins and their roles in the pathogenesis of IBV have been

completely overlooked, even when the roles of the homologs in other CoVs have been proven significant. Accessory proteins of IBV and other CoVs may also offer a new generation of vaccine targets: the use of codon-deoptimization of non-structural virulence genes in influenza A virus and respiratory syncytial virus resulted in genetically stable viruses that retained immunogenicity but were attenuated (Nogales et al 2014; Meng et al., 2014). Evidently virulence and immunogenicity in IBV is a multi-genic trait, and future studies must aim to pursue a better understanding and exploitation of the roles of various viral proteins in the host, if any advances are to be made in controlling the disease in poultry.

Acknowledgements

Adrian Knoetze and Rainbow Veterinary Laboratory are thanked for providing strain 3665/11 for this study. Funding was provided by the Poultry Section, Department of Production Animal Studies.

References

Bentley, K., Keep, S.M., Armesto, M., Britton P, 2013. Identification of a noncanonically transcribed subgenomic mRNA of infectious bronchitis virus and other gammacoronaviruses. J. Virol. 87(4), 2128-2136.

Bijlenga, G., Cook, J.K., Gelb, J. Jr., de Wit, J.J., 2004. Development and use of the H strain of avian infectious bronchitis virus from the Netherlands as a vaccine: a review. Avian Pathol. Dec 33(6), 550-557.

Borucki, M.K., Allen, J.E., Chen-Harris, H., Zemla, A., Vanier, G., Mabery, S., Torres, C., Hullinger, P., Slezak, T., 2013. The role of viral population diversity in adaptation of bovine coronavirus to new host environments. PLoS One 8(1) :e52752.

Bosch, B.J., de Haan, C.A., Smits, S.L., Rottier, P.J., 2005. Spike protein assembly into the coronavirion: exploring the limits of its sequence requirements. Virology 334(2), 306-318.

Briese, T., Mishra, N., Jain, K., Zalmout, I.S., Jabado, O.J., Karesh, W.B., Daszak, P., Mohammed, O.B., Alagaili, A.N., Lipkin, W.I., 2014. Middle East respiratory syndrome coronavirus quasispecies that include homologues of human isolates revealed through whole-genome analysis and virus cultured from dromedary camels in Saudi Arabia. MBio. 29, 5(3):e01146-14.

Cao, J., Wu, C.C., Lin, T.L., 2008. Complete nucleotide sequence of polyprotein gene 1 and genome organization of turkey coronavirus. Virus Res. 136(1-2), 43-49.

Casais, R., Davies, M., Cavanagh, D., Britton, P., 2005. Gene 5 of the avian coronavirus infectious bronchitis virus is not essential for replication. J Virol. 79(13), 8065-8078.

Cavanagh, D., 2005. Coronaviruses in poultry and other birds. Avian Pathol. 34(6), 439-48. Review.

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

Cavanagh, D., Davis, P.J., Cook, J.K., 1992. Infectious bronchitis virus: evidence for recombination within the Massachusetts serotype. Avian Pathol. 21(3):401-408.

Coletta, A., Pinney, J.W., Solis, D.Y., Marsh, J., Pettifer, S.R., Atwood, T.K., 2010. Low-complexity regions within protein sequences have position-dependent roles. BMC Systems Biology 4:43.

Collisson, E.W., Pei, J., Dzielawa, J., Seo, S.H., 2000. Cytotoxic T lymphocytes are critical in the control of infectious bronchitis virus in poultry. Dev Comp Immunol. 24(2-3), 187-200. Review.

Cook, J.K., Jackwood, M., Jones, R.C., 2012. The long view: 40 years of infectious bronchitis research. Avian Pathol. 41(3), 239-250.

Darbyshire, J.H., Rowell, J.G., Cook, J.K., Peters, R.W., 1979. Taxonomic studies on strains of avian infectious bronchitis virus using neutralisation tests in tracheal organ cultures. Arch Virol. 61(3), 227-238.

DeDiego, M.L., Nieto-Torres, J.L., Jimenez-Guardeño, J.M., Regla-Nava, J.A., Castaño-Rodriguez, C., Fernandez-Delgado, R., Usera, F., Enjuanes, L., 2014. Coronavirus virulence genes with main focus on SARS-CoV envelope gene. Virus Res. Aug 2. pii: S0168-1702(14)00302-5.

Ducatez, M.F., Martin, A.M., Owoade, A.A., Olatoye, I.O., Alkali, B.R., Maikano, I., Snoeck, C.J., Sausy, A., Cordioli, P., Muller, C.P., 2009. Characterization of a new genotype and serotype of infectious bronchitis virus in Western Africa. J Gen Virol. 90(11), 2679-2685.

de Haan, C.A., Stadler, K., Godeke, G.J., Bosch, B.J., Rottier, P.J., 2004. Cleavage inhibition of the murine coronavirus spike protein by a furin-like enzyme affects cell-cell but not virus-cell fusion. J Virol. 78(11), 6048-6054.

Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D., Baric, R.S., 2011. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. RNA Biol. 8(2), 270-279.

De Wit, J.J., 2000. Detection of infectious bronchitis virus. Avian Pathol. 29(2), 71-93.

Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Speck, T., Krüger, D., Grebnev, G., Kuban, M., Strumillo, M., Uyar, B., Budd, A., Altenberg, B., Seiler, M., Chemes, L.B., Glavina, J., Sánchez, I.E., Diella, F., Gibson, T.J., 2014. The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res. 42(Database issue): D259-66. http://elm.eu.org Last accessed 03/12/2014.

Farsang, A., Ros, C., Renström, L.H., Baule, C., Soós, T., Belák, S., 2002. Molecular epizootiology of infectious bronchitis virus in Sweden indicating the involvement of a vaccine strain. Avian Pathol. 31(3), 229-236.

Gallardo, R.A., van Santen, V.L., Toro, H., 2012. Effects of chicken anaemia virus and infectious bursal disease virus-induced immunodeficiency on infectious bronchitis virus replication and genotypic drift. Avian Pathol. 41(5), 451-458.

Gelb, J. Jr., Keeler, C.L. Jr., Nix, W.A., Rosenberger, J.K., Cloud, S.S., 1997. Antigenic and S-1 genomic characterization of the Delaware variant serotype of infectious bronchitis virus. Avian Dis. 41(3), 661-669.

Goebel, S.J., Hsue, B., Dombrowski, T.F., Masters, P.S., 2004. Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. J Virol. 78(2), 669-682.

Hagai, T., Azia, A., Babu, M.M., Andino, R., 2014. Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. Cell Rep. 7(5), 1729-1739.

Hewson, K.A., Ignjatovic, J., Browning, G.F., Devlin, J.M., Noormohammadi, A.H., 2011. Infectious bronchitis viruses with naturally occurring genomic rearrangement and gene deletion. Arch Virol. 156(2), 245-252.

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

Ignjatovic, J., McWaters, P.G., 1991. Monoclonal antibodies to three structural proteins of avian infectious bronchitis virus: characterization of epitopes and antigenic differentiation of Australian strains.J Gen Virol. 72 (Pt 12), 2915-2922.

Inglis, S.C., Rolley, N., Brierley, I., 1990. A ribosomal frameshift signal in the polymerase-encoding region of the IBV genome. Adv Exp Med Biol. 276, 269-273.

Jackwood, M.W., Hilt, D.A., Lee, C.W., Kwon, H.M., Callison, S.A., Moore, K.M., Moscoso, H., Sellers, H., Thayer, S., 2005. Data from 11 years of molecular typing infectious bronchitis virus field isolates. Avian Dis. 49(4), 614-618.

Jeong, Y.S., Repass, J.F., Kim, Y.N., Hwang, S.M., Makino, S., 1996. Coronavirus transcription mediated by sequences flanking the transcription consensus sequence. Virology. 217(1), 311-322.

Jones, R.C., 2010. Viral respiratory diseases (ILT, aMPV infections, IB): are they ever under control? Br Poult Sci.51(1),1-11.

Kallberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., Xu, J., 2012. Template-based protein structure modeling using the RaptorX web server. Nature Protocols, 7(8): 1511-1522, http://raptorx.uchicago.edu/ last accessed 03/12/2014.

Kant, A., Koch, G., van Roozelaar, D.J., Kusters, J.G., Poelwijk, F.A., van der Zeijst, B.A., 1992. Location of antigenic sites defined by neutralizing monoclonal antibodies on the S1 avian infectious bronchitis virus glycopolypeptide. J Gen Virol. 73 (Pt 3):591-596.

Kint, J., Fernandez-Gutierrez, M., Maier, H.J., Britton, P., Langereis, M.A., Koumans, J., Wiegertjes, G.F., Forlenza, M., 2014. Activation of the chicken type I IFN response by infectious bronchitis coronavirus. J Virol. Nov 5. pii: JVI.02671-14.

Knoetze, A.D., Moodley, N., Abolnik, C., 2014. Two genotypes of infectious bronchitis virus are responsible for serological variation in KwaZulu-Natal poultry flocks prior to 2012. Onderstepoort Journal of Veterinary Research 81(1), 1-10.

Koch, G., Hartog, L., Kant, A., van Roozelaar, D.J., 1990. Antigenic domains on the peplomer protein of avian infectious bronchitis virus: correlation with biological functions. J Gen Virol. 71 (Pt 9):1929-1935.

Kottier, S.A., Cavanagh, D., Britton, P., 1995. First experimental evidence of recombination in infectious bronchitis virus. Adv Exp Med Biol.380, 551-556.

Kuo, S.M., Kao, H.W., Hou, M.H., Wang, C.H., Lin, S.H., Su, H.L., 2013. Evolution of infectious bronchitis virus in Taiwan: positively selected sites in the nucleocapsid protein and their effects on RNA-binding activity.Vet Microbiol. 162(2-4), 408-418.

Kusters, J.G., Jager, E.J., Lenstra, J.A., Koch, G., Posthumus, W.P., Meloen, R.H., van der Zeijst, B.A., 1989. Analysis of an immunodominant region of infectious bronchitis virus. J Immunol. 143(8), 2692-2698.

Lai, M.M., Cavanagh, D., 1997. The molecular biology of coronaviruses. Adv Virus Res. 48:1-100. Review.

Liu, G., Lv, L., Yin, L., Li, X., Luo, D., Liu, K., Xue, C., Cao, Y., 2013. Assembly and immunogenicity of coronavirus-like particles carrying infectious bronchitis virus M and S proteins.Vaccine. 31(47), 5524-5530.

Liu, S., Xu, Q., Han, Z., Liu, X., Li, H., Guo, H., Sun, N., Shao, Y., Kong, X., 2014. Origin and characteristics of the recombinant novel avian infectious bronchitis coronavirus isolate ck/CH/LJL/111054. Infect Genet Evol. 23,189-195.

Mardani, K., Noormohammadi, A.H., Hooper, P., Ignjatovic, J., Browning, G.F., 2008. Infectious bronchitis viruses with a novel genomic organization. J Virol. 82(4), 2013-2024.

McBride, R., van Zyl, M., Fielding, B.C., 2014. The coronavirus nucleocapsid is a multifunctional protein. Viruses. 6(8), 2991-3018.

Meng, J., Lee, S., Hotard, A.L., Moore, M.L., 2014. Refining the balance of attenuation and immunogenicity of respiratory syncytial virus by targeted codon deoptimization of virulence genes. MBio. 5(5):e01704-14.

Meulemans, G., Boschmans, M., Decaesstecker, M., Berg, T.P., Denis, P., Cavanagh, D., 2001. Epidemiology of infectious bronchitis virus in Belgian broilers: a retrospective study, 1986 to 1995. Avian Pathol. 30(4), 411-421.

Mo, M., Huang, B., Wei, P., Wei, T., Chen, Q., Wang, X., Li, M., Fan, W., 2012. Complete genome sequences of two Chinese virulent avian coronavirus infectious bronchitis virus variants. J Virol. 86(19):10903-10904.

Moore, K.M., Jackwood, M.W., Hilt, D.A., 1997. Identification of amino acids involved in a serotype and neutralization specific epitope within the S1 subunit of avian infectious bronchitis virus. Arch Virol. 142(11), 2249-2256.

Mork, A.K., Hesse, M., Abd El Rahman, S., Rautenschlein, S., Herrler, G., Winter, C., 2014. Differences in the tissue tropism to chicken oviduct epithelial cells between avian coronavirus IBV strains QX and B1648 are not related to the sialic acid binding properties of their spike proteins. Vet Res. 45:67. doi: 10.1186/1297-9716-45-67.

Ndegwa, E.N., Toro, H., van Santen, V.L., 2014. Comparison of vaccine subpopulation selection, viral loads, vaccine virus persistence in trachea and cloaca, and mucosal antibody responses after vaccination with two different Arkansas Delmarva Poultry Industry -derived infectious bronchitis virus vaccines. Avian Dis. 58(1), 102-110.

Nogales, A., Baker, S.F., Ortiz-Riaño, E., Dewhurst, S., Topham, D.J., Martínez-Sobrido, L., 2014. Influenza A virus attenuation by codon deoptimization of the NS gene for vaccine development. J Virol. 88(18):10525-10540.

Pei, J., Sekellick, M.J., Marcus, P.I., Choi, I.S., Collisson, E.W., 2001. Chicken interferon type I inhibits infectious bronchitis virus replication and associated respiratory illness. J Interferon Cytokine Res. 21(12):1071-1077.

Promkuntod, N., Wickramasinghe, I.N., de Vrieze, G., Gröne, A., Verheije, M.H. 2013. Contributions of the S2 spike ectodomain to attachment and host range of infectious bronchitis virus. Virus Res. 177(2): 127-37.

Promkuntod, N., van Eijndhoven, R.E., de Vrieze, G., Gröne, A., Verheije, M.H., 2014. Mapping of the receptor-binding domain and amino acids critical for attachment in the spike protein of avian coronavirus infectious bronchitis virus. Virology. 448, 26-32.

Susan, R.W., Julian L.L., 2011. Coronavirus pathogenesis. Adv Virus Res 81, 95-164.

Sola, I., Mateos-Gomez, P.A., Almazan, F., Zuñiga, S., Enjuanes, L., 2011. RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. RNA Biol. 8(2), 237-248.

Tangudu, C., Olivares, H., Netland, J., Perlman, S., Gallagher, T., 2007. Severe acute respiratory syndrome coronavirus protein 6 accelerates murine coronavirus infections. J Virol. 81(3):1220-1229.

Timms, L.M., Bracewell, C.D., Alexander, D.J., 1980. Cell mediated and humoral immune response in chickens infected with avian infectious bronchitis. Br Vet J. 36(4), 349-346.

Toro, H., Zhao, W., Breedlove, C., Zhang, Z., Yub, Q., 2014. Infectious bronchitis virus S2 expressed from recombinant virus confers broad protection against challenge. Avian Dis. 58(1), 83-89.

Yang, Y., Zhang, L., Geng, H., Deng, Y., Huang, B., Guo, Y., Zhao, Z., Tan, W., 2013. The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. Protein Cell. 4(12), 951-961.

Wang, L., Xu, L., Collison, E.W. 1997. Experimental confirmation of recombination upstream of the S1 hypervariable region of infectious bronchitis virus. Virus Research 49, 139-145.

Wang, C.H., Huang, Y.C., 2000. Relationship between serotypes and genotypes based on the hypervariable region of the S1 gene of infectious bronchitis virus. Arch Virol. 145(2), 291-300.

Wang, S., Sundaram, J.P., Spiro, D., 2010. VIGOR, an annotation program for small viral genomes. BMC Bioinformatics, 11:451, 1-10. http://www.jcvi.org/vigor Last accessed 03/12/2014.

West, B., Zhou, B.X., 1988. Did chickens go north? New evidence for domestication. J. Archeol Sci 15:515-533.

Woo, P.C., Lau, S.K., Lam, C.S., Lau, C.C., Tsang, A.K., Lau, J.H., Bai, R., Teng, J.L., Tsang, C.C., Wang, M., Zheng, B.J., Chan, K.H., Yuen, K.Y., 2012. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol.;86(7), 3995-4008.

Ye, Z.D., Wong, C.K., Li, P., Xie, Y., 2012. The role of SARS-CoV protein, ORF-6, in the induction of host cell death. Hong Kong Med J; 16 (Suppl 4): SS22-6.

Zhao, F., Zou, N., Wang, F., Guo, M., Liu, P., Wen, X., Cao, S., Huang, Y., 2013. Analysis of a QX-like avian infectious bronchitis virus genome identified recombination in the region containing the ORF 5a, ORF 5b, and nucleocapsid protein gene sequences. Virus Genes. 46(3), 454-64.