Prediction of Bovine coronavirus immune epitopes using selection and phylogenetic analysis

Tristan Russell [1], Jose Maria Lozano [2], Cailyn Challinor [1], Gerald Barry [1 *]

Affiliations:

1) UCD School of Veterinary Medicine, University College Dublin, Ireland.
2) Central Veterinary Research Laboratory, Department of Agriculture, Food and the Marine, Celbridge, Ireland.

***Corresponding Author:** Gerald Barry; gerald.barry@ucd.ie

Abstract

Bovine betacoronavirus (BoCoV) is a pneumoenteric pathogen of cattle, which is closely related to Human coronavirus OC43. Vaccines are administered to protect against diseases caused by BoCoV, but knowledge gaps exist with regard to correlates of protection and the effect of immune evasion on driving evolution. In this study, immune epitopes were predicted for the BoCoV structural proteins including spike and haemagglutinin esterase (HE) and these predictions were supported through targeted gene sequencing of Irish clinical isolates and selective pressure analysis. Increased prevalence of diversifying selection and amino acid changes in some predicted immune epitopes suggests immune escape is selecting for non-synonymous mutations arising in these regions. Selection analysis and sequencing provided increased support for nAb epitopes compared to other predicted immune epitopes suggesting nAbs are an important arm of the immune response to BoCoV. Phylogenetic analysis of spike and HE sequences showed Irish isolates from this study were in the European clade except for one HE sequence that sat in the Asian/American clade, while the spike gene of this sample was in the European clade. Recombination between a European and an Asian/American isolate could give rise to such a sequence and recombination breakpoints that were detected at the 3' end of HE and 5' end of spike would produce such a sequence. This study presents evidence showing pressure to evade the nAb response is contributing to BoCoV evolution and for the first time, sequenced an isolate likely derived from a recombination event between European and Asian/American strains.

Keywords: Bovine coronavirus, Immune epitopes, Selective Pressure, Phylogenetics, Ireland

Repositories (GenBank Accession Numbers): OR271239, OR271240, OR271241, OR271242, OR271243, OR271244, OR271245, OR271246, OR271247, OR271248, OR271249, OR271250, OR271251, OR271252, OR271253, OR271254, OR271255, OR271256, OR271257, PP156983, PP156984, PP156985, PP156986, PP156987, PP156988, PP156989, PP156990, PP156991, PP156992

Introduction

Bovine betacoronavirus (BoCoV) is a pneumoenteric pathogen of *Bos taurus* though closely related viruses have been isolated from several other ruminant species and some non-ruminant mammals [1]. In cattle it causes winter dysentery, calf diarrhoea [2] and is associated with bovine respiratory disease [3]. In Ireland a retrospective study published in 2014 found that of the five viruses associated with the respiratory disease complex included in the screen, BoCoV was the most frequently detected in calves with respiratory disease in Ireland [4]. BoCoV is also frequently detected in pneumonia cases as part of co-infections [5].

As a member of the *Embecovirus* subgenus, BoCoV encodes five structural proteins including four conserved across the *Coronaviridae* family – spike, nucleocapsid, membrane and small envelope – as well as a surface protein unique to this *Embecovirus* subgenus – haemagglutinin esterase (HE). HE is a receptor degrading enzyme, which facilitates virus entry and release by targeting the 9-O-acetylated sialic acid residues required for BoCoV entry [6]. It has previously been shown infection-induced humoral immune responses against these proteins can protect cattle from disease [1, 7, 8], while some antibodies generated in response to vaccine can recognise antigens derived from clinical isolates [9]. However, it has been shown calf BoCoV antibody titres do not correlate with protection from respiratory disease when dams received a vaccine designed to protect against enteric BoCoV disease [10, 11].

Our current understanding of BoCoV immune epitopes, correlates of protection, and the effect of immune evasion on BoCoV evolution is limited [1]. To further understand which viral proteins are targeted by the immune system and to assess the contribution of immune evasion pressure on protein sequence variation, predicted immune epitopes of BoCoV structural proteins were combined with selective pressure analysis in this study. There are already full-genome sequences of 24 BoCoV isolated from cattle in 2019/20 from across Ireland available publicly [12] as well as ten S1 domain and one full-length spike sequences from isolates collected in the province of Munster in 2010/11 [13]. Here the spike and HE genes of 29 Irish clinical isolates were sequenced and compared with other BoCoV isolates obtained nationally and internationally. Phylogenetic analysis of the spike and HE sequences identified residues where amino acid changes have occurred compared to a commonly used vaccine strain. We present evidence of selective pressure on residues associated with predicted immune epitopes suggesting that vaccination is driving variation in these genes, and this may lead to reduced efficacy of currently used vaccines.

Materials and Methods

Samples and Detection of Bovine coronavirus

Nasopharyngeal samples collected from 29 cattle displaying symptoms of respiratory illness during 2022/23 in Ireland were submitted to the Department of Agriculture, Food and the Marine's Virus reference laboratory (Table 1). Samples were collected as part of normal veterinary diagnostic practices. These samples were then tested for the presence of respiratory pathogens by qPCR. Those positive for BoCoV were identified using a previously developed TaqMan qPCR protocol [14]. Sequencing of positive samples was carried out as approved by the University College Dublin animal research ethics committee, approval code AREC-E-22-18-Barry.

Sample no.	GenBank Accession	Location (county)	Date	Pathology	Host
1	OR271239	Galway	30/03/2022	ND	NA
2	OR271240	Waterford	11/05/2022	ND	NA
3	OR271241	Sligo	27/04/2022	Dead, pneumonia, *Histophilus somni* and *Mannheimia haemolytica* coinfection	3 w
4	OR271242	Laois	08/04/2022	ND	2 m
5	OR271243	Laois	08/04/2022	ND	2 m
6	OR271244	Wexford	22/03/2022	Enteric symptoms, *Escherichia coli* co-infection	5 w
7	OR271245	Wexford	21/03/2022	Enteric and respiratory symptoms, *E. coli* co-infection	10 d
8	OR271246	Kilkenny	16/03/2022	ND	Calves (pooled)
9	OR271247	Kilkenny	14/03/2022	Pneumonia and enteritis, BHV-1 co-infection	4 w
10	OR271248	Cavan	23/02/2022	*Pasteurella multocida* and *M. haemolytica* co-infection	4 animals (pooled)
11	OR271249	Donegal	15/02/2022	Pneumonia	10 m
12	OR271250	Wexford	11/03/2022	Pneumonia, other bacterial respiratory pathogens isolated	2 w
13	OR271251	Wexford	11/03/2022	Pneumonia, other bacterial respiratory pathogens isolated	2 w
14	OR271252	Sligo	06/01/2022	ND	Weanling
15	OR271253	Laois	08/04/2022	ND	< 1 m
17	OR271254	Wicklow	23/05/2022	Pneumonia, BHV-4, *H. somni* and *M. haemolytica* co-infections	3 m
116	OR271255	Westmeath	2022	Dead	1 y male
117	OR271256	Offaly	2022	Pneumonia	8 m
119	OR271257	Meath	2022	Pneumonia	6 w
123	PP156983	Carlow	04/11/2022	Pneumonia, BRSV co-infection	9 m
124	PP156990	Kilkenny	03/11/2022	Respiratory symptoms	2 y female
125	PP156992	Kilkenny	06/12/2022	Respiratory and enteric symptoms, PIV-3 co-infection	8 m
126	PP156984	Cork	14/12/2022	Dead, respiratory and enteric symptoms, BRSV co-infection	6 m
127	PP156985	Galway	21/11/2022	Dead, pneumonia, BRSV co-infection	7 m
128	PP156986	Longford	02/12/2022	Respiratory symptoms, PIV-3/BRSV co-infections	5 w

129	PP156987	Carlow	26/10/2022	Respiratory and enteric symptoms, PIV-3 co-infection	5 m
131	PP156988	Wexford	10/11/2022	Pneumonia, enteric symptoms, BRSV co-infection	10 m
132	PP156991	Tipperary	06/01/2023	Enteric symptoms	10 m female
133	PP156989	Carlow	07/11/2022	Dead, pneumonia	8 m

Table 1: Sample details and GenBank accession numbers. Abbreviations: ND, no disease information; NA, no animal information; BHV-1/4, *Bovine alphaherpesvirus-1/4*; BRSV, *Bovine respiratory syncytial virus*; PIV-3, *Parainfluenza virus-3*; d, days old; w, weeks old; m, months old; y, years old.

Recombination and Selection analysis

BoCoV spike, HE, nucleocapsid, membrane and small envelope coding sequences were downloaded from the GenBank database (https://www.ncbi.nlm.nih.gov/, Table S1). Identical and incomplete sequences were removed from datasets. Codon alignments for each gene were obtained using the Clustal Omega alignment algorithm [15, 16] in Jalview version 2 [17–19]. Stop codons were deleted manually. Recombination breakpoints of each gene were predicted using Genetic Algorithm for Recombination Detection (GARD) [20] through the data monkey webserver (datamonkey.org) [21]. Sequences arising through recombination were removed from alignments, then mixed effects model of evolution (MEME) [22] and fast unconstrained bayesian approximation (FUBAR) [23] methods were used through the data monkey web server [21] to infer site-specific selective pressures. MEME predicted sites under diversifying episodic selection and FUBAR estimated synonymous and non-synonymous mutation rates to infer pervasive selective pressure.

Prediction of immune epitopes

Immune epitopes for BoCoV structural proteins were predicted based on three previous studies. Firstly, BoCoV spike and HE Major Histocompatibility Complex-I (MHC-I, also known as bovine leukocyte antigen-I) and B cell receptor (BCR) immune epitopes were predicted bioinformatically [24]. The second two methods mapped immune epitopes identified for *Human betacoronavirus OC43* (HCoV OC43) protein onto BoCoV orthologues because, depending on the variants selected for comparison, spike, HE and NCP shared approximately 91%, 94% and 97% identity at amino acid level. Also, the two Embecovirus species are closely related, with evidence suggesting HCoV OC43 evolved from BoCoV following a relatively recent zoonotic event [25]. MHC-II (also known as bovine leukocyte antigen-II) immune epitopes identified for HCoV OC43 by co-immunoprecipitation (co-IP) of MHC-II molecules [26] were mapped onto BoCoV spike, HE, nucleocapsid and small envelope . HCoV OC43 spike residues shown to interact with neutralising antibodies (nAb) through structural analysis or shown to be escape mutants following passage in the presence of antibodies [27] were also mapped onto BoCoV spike.

107 **Protein Structures**

108 BoCoV spike structure was predicted using SWISS-MODEL (https://swissmodel.expasy.org/) with the
109 BoCoV Mebus spike amino acid sequence as input [28]. BoCoV HE structure had been determined by
110 x-ray crystallography [29] and deposited on the protein data bank (https://www.rcsb.org/). Spike and
111 HE structures were visualised and annotated in PyMOL [30].

112 **Sequencing spike and haemagglutinin esterase**

113 RNA from samples positive for BoCoV was converted to cDNA using the SuperScript III Reverse
114 Transcriptase kit (Invitrogen) following the manufacturer's protocol. cDNA was used as a template for
115 PCR (Phusion HF kit (New England Biolabs)) with primers designed to amplify the spike and HE genes
116 (Table 2). PCR products from reactions producing a specific fragment underwent PCR clean-up
117 (Qiagen) and, if multiple fragments were detected, gel purification (Qiagen) from a 0.8% agarose gel
118 was carried out. Purified fragments were sent for Sanger sequencing using appropriate forward and
119 reverse primers (Eurofins Genomics).

Gene	Start	Finish	Primer sequences (5' – 3')		Annealing (°C)
			Forward	Reverse	
HE	-365	547	CCCTCATCACCGGCTAGACT	CCCCAAAATTAGCTTCACGAGC	65
HE	-288	476	CCACTGGATGGGAATTCGTTT	GTAGGTTGTGCAGAGCCATT	63
HE	377	1353	AGGCTTGTTTTACACTCAGGT	GTACACTTTAAATCTCCTATAACAGC	61
Spike	-107	1039	CTTGGCATTCTTTTGGGTGTTGC	TAATGGAGAGGGCACCGACTT	66
Spike	782	1550	GGGTTACACCTCTCACTTCT	GCAGGACAAGTGCCTATACC	60
Spike	1460	2286	GTCCGTGTAAATTGGATGGG	TGTAGAGTAATCCACACAGT	60
Spike	1855	2731	TTACAAAAATCAAACACAGACAT	AAACTTTATTACAATCGCTTCC	60
Spike	2680	3617	TCAATTTTTCCCCTGTATTAGG	GTAGTAATAACCACTACCAGTG	59
Spike	3475	4218	TTTAGCTATGTCCCTACTAAGTA	TGTGGTAGCTATTATAATATGCTCG	60

120

121 **Table 2:** Primer pairs used to amplify regions of HE and spike genes. Start and finish positions of the
122 amplicons are provided relative to the first nucleotide of the HE or spike genes. Primers were designed
123 for this study or taken from [31, 32].

124 **Phylogenetic analysis**

125 Reads obtained from Sanger sequencing were trimmed to remove low quality base calls. Reads from
126 the same sample were aligned to BoCoV spike and HE references using the Clustal Omega alignment
127 algorithm [15, 16] in Jalview [17–19]. References were selected based on the closest NCBI Blast hits
128 for individual sequences from the same sample. Sequences were deposited on GenBank (Table 1).

129 Nucleic acid sequences were converted to amino acid sequences then phylogenetic trees were generated
130 using the and neighbour-joining method [33] in Seaview version 5 [34]. Trees were annotated and
131 images generated in the Interactive Tree of Life web server [35].

132 Results

133 Selection Analysis of BoCoV structural proteins

134 Regions of proteins under diversifying selection could represent immune epitopes, several MHC-I,
135 MHC-II, BCR and nAb epitopes within BoCoV structural proteins were predicted as described in the
136 Materials and Methods and are presented here in Table 3. Site-specific selective pressures were then
137 inferred using FUBAR and MEME algorithms (Figure 1). Prior to running these, GARD analysis was
138 used to identify and remove sequences derived from recombination. Recombinant sequences were
139 detected in the spike gene alignment, but not genes of other BoCoV structural proteins. Spike (Figure
140 1A) and HE (Figure 1B) were under increased diversifying selection compared to nucleocapsid (Figure
141 1C), membrane (Figure 1D) and small envelope (Figure 1E) proteins as demonstrated by the increased
142 proportion of codons predicted to be under diversifying selection for spike and HE compared to other
143 structural proteins (Table 4). There was increased abundance of codons under purifying selection within
144 functional domains such as the fusion peptides of spike (Figure 1A); the esterase domains of HE (Figure
145 1B); the RNA binding and dimerisation domains of nucleocapsid (Figure 1C); and the transmembrane
146 domains of membrane (Figure 1D) and small envelope (Figure 1E). Receptor binding domains (RBD)
147 are involved in protein-protein interactions and often contain nAb epitopes so, selection at spike and
148 HE RBD was of particular interest. Increased selection was observed for the RBD of spike with 9.46%
149 codons within RBD under diversifying selection compared to 5.35% for the whole gene and 47.97%
150 codons within the RBD under purifying selection compared to 41.76% for the whole gene (Figure 2A).
151 There was increased diversifying selection for the HE RBD with 4.90% codons within the RBD under
152 diversifying selection compared to 4.44% for the whole gene but decreased purifying selection as
153 19.58% of codons within the RBD were under purifying selection compared to 22.66% for the whole
154 gene (Figure 2B).

Gene	Immune epitope	Location		Diversifying selection	Amino acid changes		Ref
		Start	Finish		Compared to Mebus	Among Irish sequences	
Spike	MHC-I	11	19	2	1	0	24
		77	85	0	0	1	
		866	874	0	0	0	
		973	981	0	0	0	
		987	995	0	1	0	
		1004	1012	0	0	0	
		1302	1310	0	0	0	
	MHC-II	96	114	2	2	1	26
		267	292	0	0	1	
		770	785	1	2	0	

		911	930	2	1	0	
		1090	1113	0	1	0	
		1150	1165	0	0	0	
		1157	1174	0	0	0	
	BCR	721	727	0	0	1	24
		1205	1209	1	0	0	
	nAb	27	35	1	1	1	27
		81	-	0	0	0	
		242	-	0	0	1	
		253	258	2	3	1	
		267	-	0	0	0	
		357	367	0	0	0	
		395	402	0	0	0	
		413	417	0	0	0	
		458	459	1	1	1	
		461	-	0	0	0	
		463	-	0	0	0	
		483	-	0	0	0	
		490	502	4	2	2	
		531	533	1	2	1	
		536	557	2	1	2	
		582	583	0	0	0	
HE	MHC-I	56	65	0	0	0	24
		186	194	0	0	0	
		360	368	0	2	2	
	MHC-II	93	108	0	0	1	26
		128	142	0	0	1	
		257	274	0	0	0	

Table 3: Predicted immune epitopes and overlap with codons under diversifying selection and residues where amino acid changes were observed.

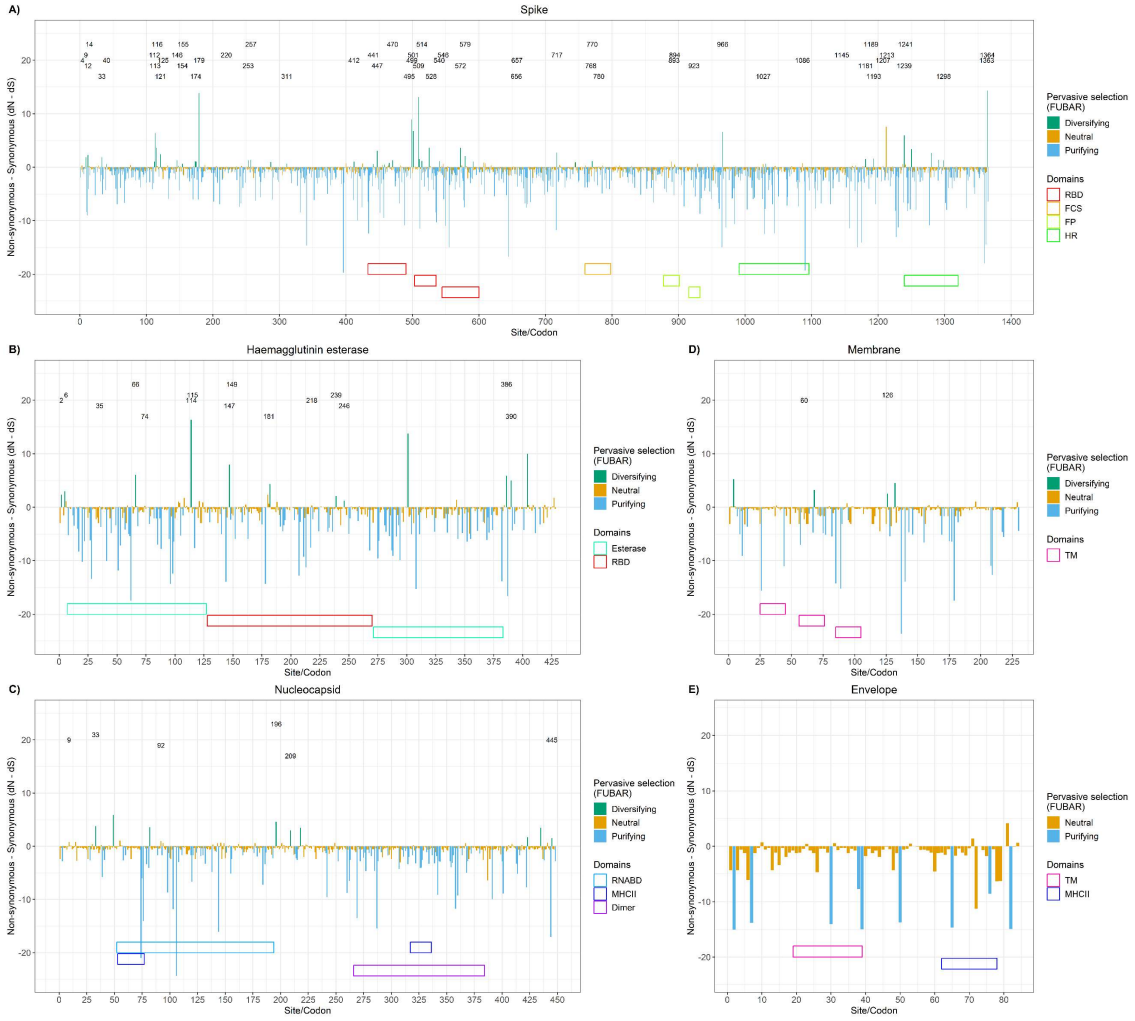


Figure 1: Selection analysis of structural BoCoV proteins. Bar plots show the results of selection analysis with FUBAR and annotated codons were predicted to be under diversifying selection using MEME for spike (A), HE (B), nucleocapsid (C), membrane (D) and small envelope (E) proteins. Locations of structural and functional domains were annotated at the bottom of plots. Abbreviations: Dimer, interaction interface for dimerisation; FCS, Furin Cleavage Site; FP, fusion peptide; HR, heptad repeat; RBD, receptor binding domain; RNABD, RNA binding domain; and TM, transmembrane domain.

Protein	Number of residues under selection						
	FUBAR (%)		MEME (%)	FUBAR + MEME	FUBAR only	MEME only	Diversifying total (%)
	Diversifying	Purifying					
Spike	49 (3.59)	570 (41.76)	57 (4.18)	33	16	24	73 (5.35)
HE	12 (2.80)	97 (22.66)	15 (3.50)	8	4	7	19 (4.44)
NCP	9 (2.00)	100 (22.27)	6 (1.34)	4	5	2	11 (2.45)
M	4 (1.74)	36 (15.65)	2 (0.87)	1	3	1	5 (2.17)
E	0 (0)	9 (10.71)	0 (0)	0	0	0	0 (0)

Table 4: Selection analysis of structural proteins. Number and proportion of codons found to be under selective pressure for each BoCoV structural protein as determined by FUBAR and MEME.

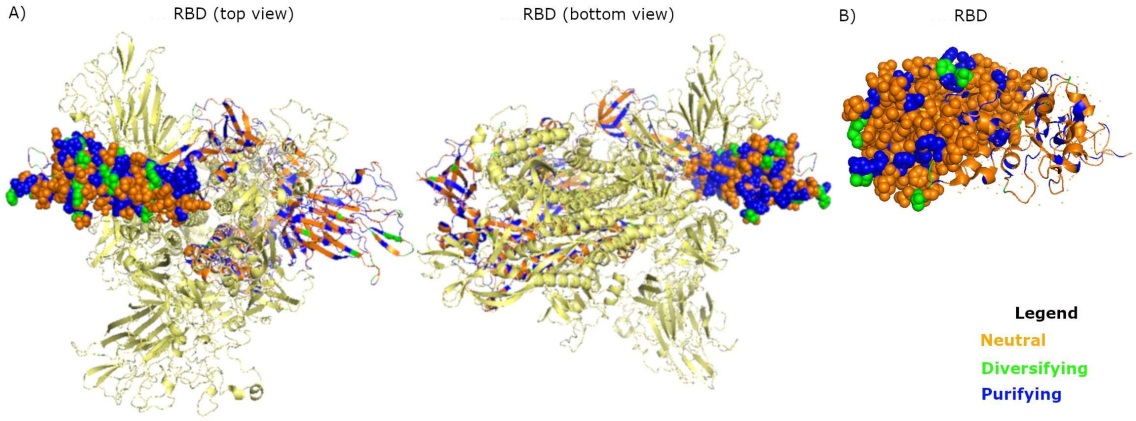


Figure 2: Residues in the RBD of spike (A) and HE (B) shown as spheres with residues coloured based on selection analysis of corresponding codons.

Immune escape provides a selective advantage for amino acid changes occurring in immune epitopes so selection within predicted BCR, MHC-I, MHC-II and nAb immune epitopes was assessed (Table 3). Spike codons under diversifying selection overlapped one of seven spike MHC-I epitopes and one of two spike BCR epitope predicted computationally (Figure 3A). Three of seven spike MHC-II immune epitopes predicted by co-IP overlapped codons under diversifying selection (Figure 3A). Of the 87 residues predicted to contribute to nAb recognition 12.64% were under diversifying selection compared to 5.35% codons of the whole spike gene (Figure 3A). Diversifying selection was not observed for any predicted HE MHC-I or MHC-II immune epitopes (Figure 3B), nor the two predicted nucleocapsid and one predicted small envelope MHC-II immune epitopes.

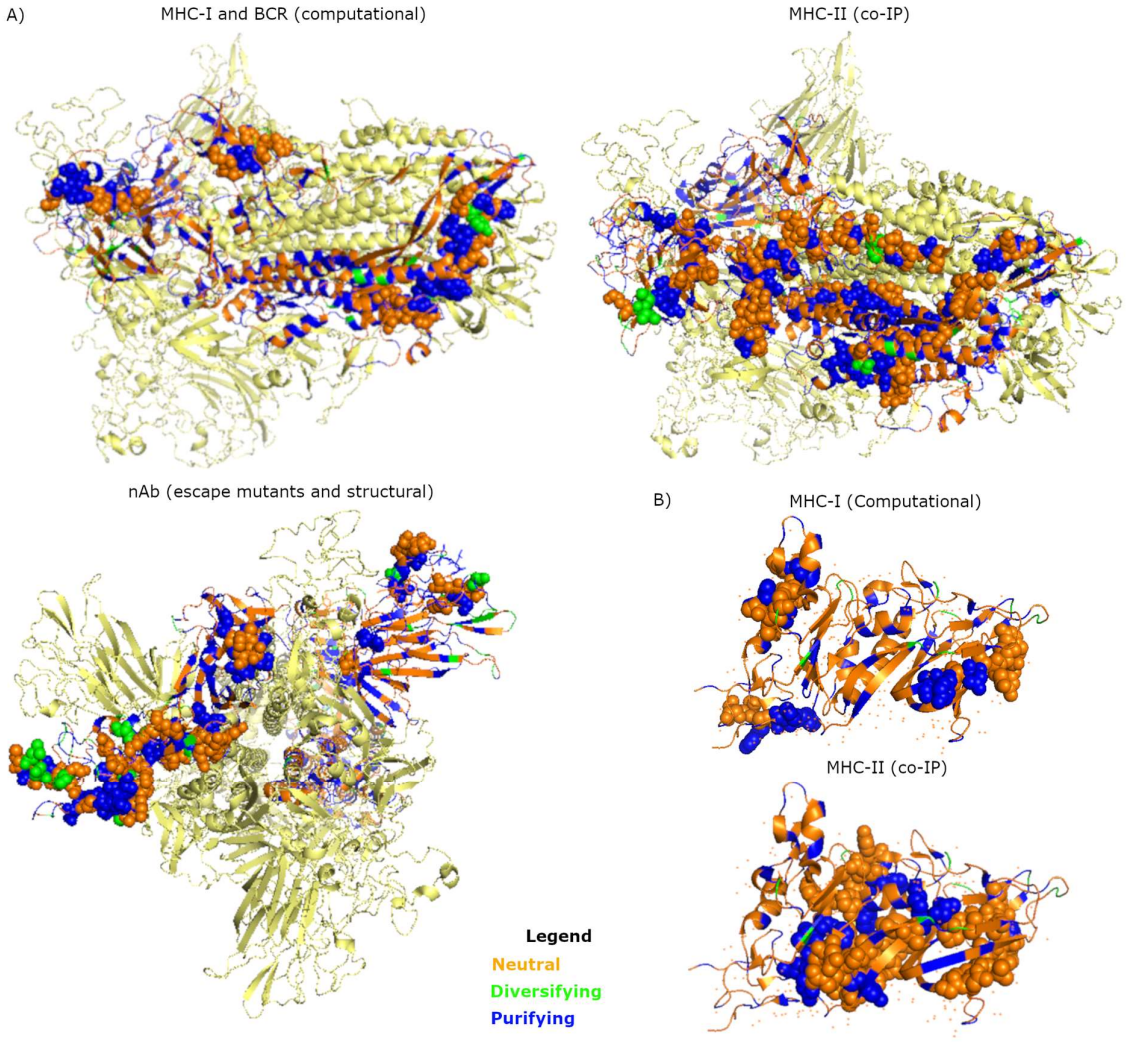


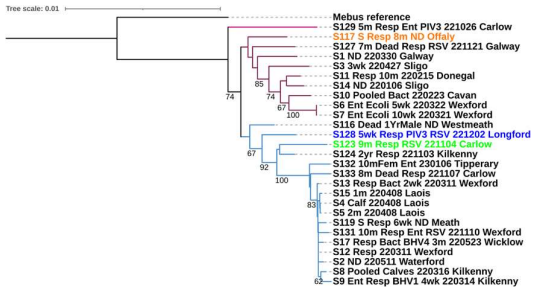
Figure 3: Spike trimer and HE structures with residues coloured based on results of FUBAR and MEME. Only one spike monomer was coloured. A) Spike immune epitopes predicted computationally, by co-IP of HCoV OC43 and by structural analysis of HCoV OC43 nAb-protein interactions were shown as spheres. B) Predicted HE MHC-II and MHC-I immune epitopes shown as spheres. Abbreviations: BCR, B cell receptor; co-IP, co-immunoprecipitation; MHC-I, major histocompatibility complex I; MHC-II, major histocompatibility complex II; and nAb, neutralising antibody.

Sequencing and phylogenetic analysis of BoCoV spike and HE

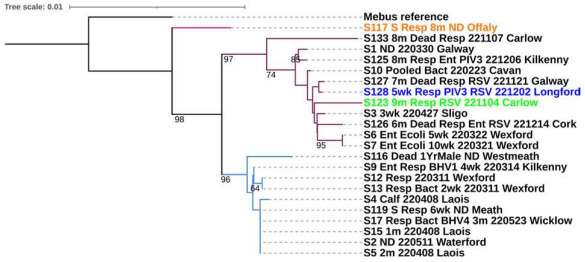
To compare currently circulating Irish isolates with older Irish isolates, international isolates and vaccine strains, the spike and HE genes of 29 BoCoV clinical isolates from cattle presenting with respiratory and sometimes enteric symptoms in Ireland were sequenced. Full-length spike and HE sequences were obtained for 27 and 22 isolates, respectively. All spike sequences, except Sample 129, fell into one of two clades (Figure 4A), while most HE sequences were split between two clades with Sample 117 and Sample 133 the only exceptions (Figure 4B). Most samples sat in the same clade of the

spike and HE trees but Sample 123 and Sample 128 switched clade, while Sample 117 and Sample 133 were both in one of the main clades in the spike tree (Figure 4A) but formed outgroups in the HE tree (Figure 4B). Differences in the spike and HE phylogenetic trees suggests recombination between BoCoV circulating in Ireland. GARD analysis of a HE-spike nucleotide alignment of sequences obtained in this study, starting from the start codon of HE to the stop codon of spike, detected seven recombination breakpoints at positions 360, 1260, 2023, 2562, 3475, 4612 and 4915 (Supplementary Figure S1). The first two overlap the HE gene and the next five overlap the spike gene in positions 733, 1272, 2185, 3322 and 3625. As previously shown, spike sequences of natural isolates collected globally predominantly diverged based on location with one clade of Asian and American isolates and a second clade of European isolates containing the Irish sequences from this study (Figure 4C). Some spike sequences from this study were in a subclade containing other Irish isolates collected in 2019/20, while most were in a subclade of French isolates from 2013/14 (Figure 4C). Sample 129 branched off on its own and was closest to a subclade of Swedish isolates from 2010. An enteric Irish isolate from 2011 diverged from other sequenced BoCoV isolates in its own subclade (Figure 4C). A tree of global HE sequences also mainly diverged based on location with one clade of Asian and American isolates and a clade of European isolates (Figure 4D). Some HE sequences from this study shared a subclade with Irish isolates collected in 2019/20, most formed a separate subclade with Albanian isolates collected in 2023, while Sample 117 was in the Asian/American clade and closest to enteric Chinese isolates from 2018 and 2020 (Figure 4D). GARD analysis of the spike-HE alignment of all GenBank sequences and sequences from this study detected three recombination breakpoints at positions 1549, 3073 and 4567, which all overlapped the spike gene at positions 247, 1771 and 3265 (Supplementary Figure S2).

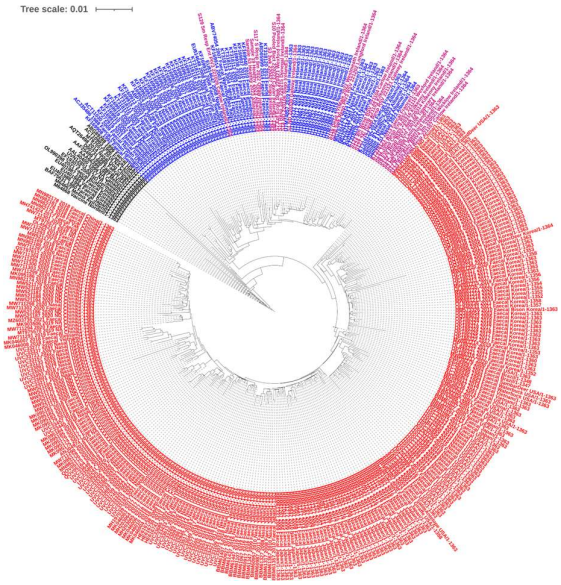
A) Tree of Irish spike nucleotide sequences



B) Tree of Irish HE nucleotide sequences



C) Tree of Spike amino acid sequences



D) Tree of HE amino acid sequences

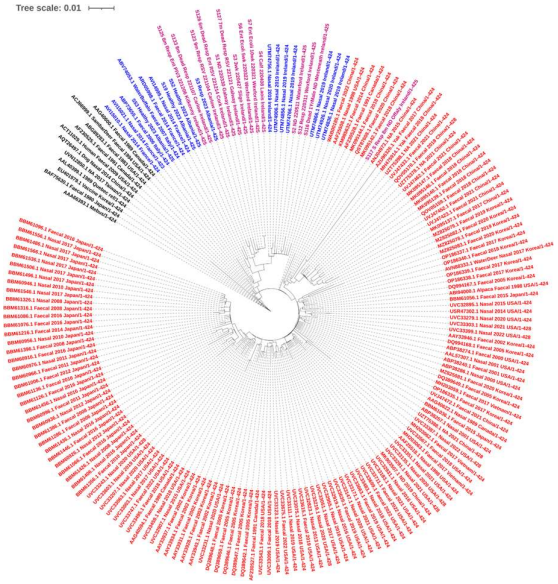


Figure 4: Phylogenetic trees of BoCoV spike and HE generated using the neighbour joining method and rooted using the Mebus strain as an outgroup. A) Nucleotide sequences of spike for isolates from this study and the Mebus strain with different coloured branches indicating two main clades. B) Same as A), but for HE. The colours of branches for each clade match in A) and B), while isolates with coloured labels switched clades or formed their own outgroup. C) All unique Spike amino acid sequences downloaded from GenBank, with label colours indicating location of isolation with Asian/American in red, European in blue and old isolates/vaccine strains in black. Sequences from this study are shown in pink. D) The same as C), but for HE.

Consensus sequences of spike and HE obtained in this study were translated and compared with BoCoV Mebus because this strain was used to design an inactivated vaccine often used in Ireland. Pressure to evade vaccine-mediated immunity would provide a selective pressure for amino acid changes occurring

in immune epitopes of natural isolates so they diverge from vaccine strains. Compared to Mebus, there were amino acid changes in 48 (3.52%) spike residues and 9 (2.12%) HE residues of the Irish consensus sequences. As prior infection with natural isolates can generate an effective immune response against future infection, there could also be selection for non-synonymous mutations among Irish BoCoV isolates. Among the spike sequences obtained in this study there were amino acid changes in 61 (4.47%) residues, while in HE there were amino acid changes in 23 (5.42%) residues. Residues containing amino acid changes were grouped based on those different from Mebus (Mebus mutations) and those where variation was observed when comparing sequences from this study (Ireland mutations) (Figure 5). Non-synonymous Mebus mutations overlapped two MHC-I epitopes (3.17% of residues predicted to be MHC-I epitopes), four MHC-II epitopes (4.32%) and 10/87 residues predicted to be involved in nAb recognition (11.49%) in spike (Figure 5A). Ireland mutations overlapped with one BCR (8.33%), one MHC-I epitope (1.59%), two MHC-II epitopes (1.44%) and 9/87 residues predicted to be involved in nAb recognition in spike (10.34%) (Figure 5A). One predicted HE MHC-I epitope overlapped with Mebus and Ireland mutations (7.14%) while two predicted MHC-II epitopes overlapped with Ireland mutations only (6.12%) (Figure 5B).

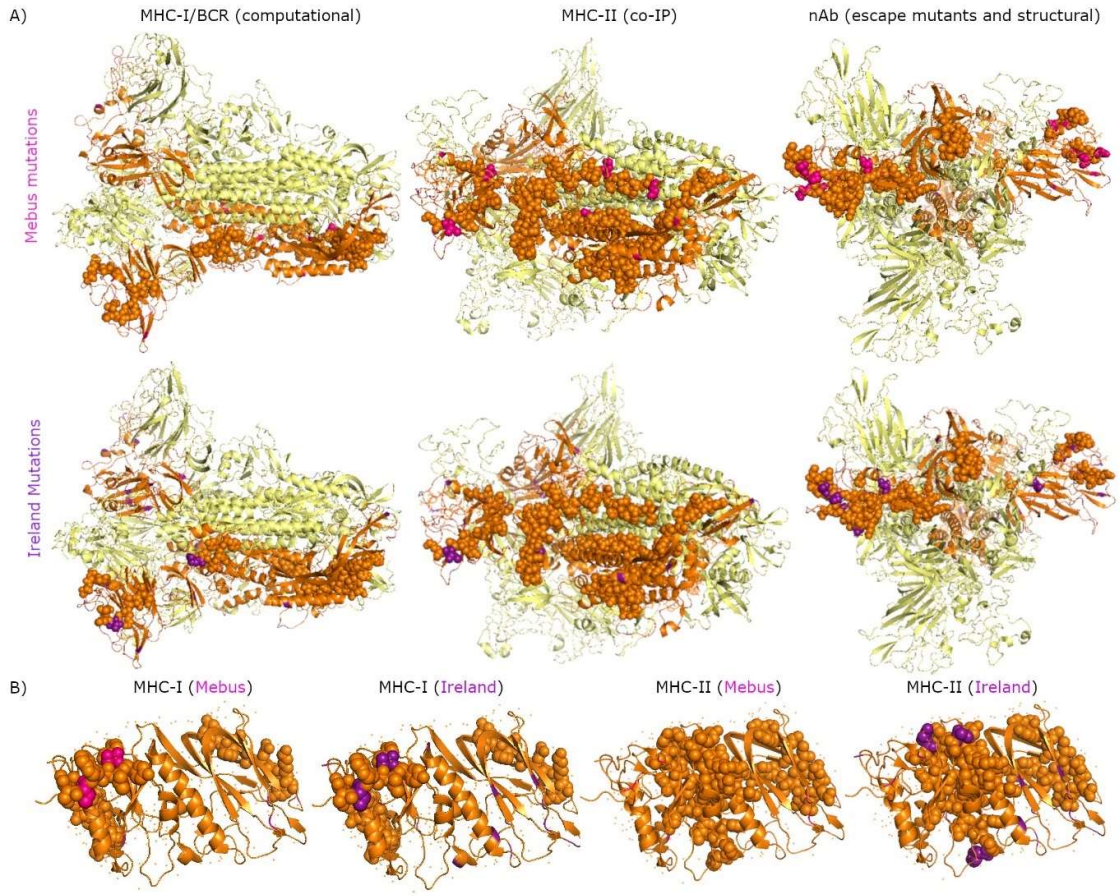


Figure 5: Overlap between residues with amino acid changes and predicted immune epitopes. A) The spike trimer is presented as a cartoon, with predicted immune epitopes presented as spheres on one spike monomer in orange. Mebus mutations are shown in pink and Ireland mutations in purple. B) Same as A) but for HE.

Discussion

Site specific selective pressures can be used to predict residues important to protein structure and function because these residues are more likely to be under purifying selection. As expected, there was overlap between residues under purifying selection and several annotated structural and functional domains of BoCoV proteins. The high level of selection observed for spike RBS is expected because its interaction with the entry receptor would drive purifying selection [36, 37], while RBD often contain nAb epitopes, which would drive diversifying selection [38, 39]. Several residues within the HE RBD were also under selection with an increased prevalence of diversifying selection compared to the rest of the protein. The high level of purifying selection in the esterase domains, which have catalytic activity and do not overlap the RBD, means there was not an increased prevalence of purifying selection for HE RBD compared to the rest of the protein.

Three studies were used to predict immune epitopes of structural BoCoV proteins. BoCoV spike and HE MHC-I and BCR immune epitopes were predicted bioinformatically [24]. HCoV OC43 spike, HE, NCP and envelope MHC-II epitopes were identified by co-IP [26], then mapped onto BoCoV orthologues. Residues and domains involved in nAb recognition of HCoV spike were determined by structural analysis and identification of escape mutants [27], then mapped onto BoCoV spike. Pressure to evade the immune response is one driver of diversifying selection in *Coronaviridae* [36, 40], although there are other drivers of diversifying selection, such as adaptation to a new host [41]. Therefore, selective pressures acting on residues along BoCoV structural proteins were analysed to provide additional evidence as to the true nature of the predicted immune epitopes with the assumption that some diversifying selection would be driven by immune escape. The proportion of codons under diversifying selection along the whole genes of spike and HE were increased compared to other structural proteins suggesting they are more likely to contain immune epitopes. There was no overlap between residues under diversifying selection and predicted HE immune epitopes but 5.96% of residues in predicted spike immune epitopes were under diversifying selection compared to 5.35% of residues for the full-length protein. There were more residues under diversifying selection within nAb immune epitopes than other predicted immune epitopes. There was also overlap between predicted immune epitopes and amino acid changes within the sequences from this study suggesting immune escape could be selecting for these changes. Once again, an increased proportion of amino acid changes were observed in nAb epitopes compared to other predicted immune epitopes implying there is a selective

304 advantage for non-synonymous mutations in regions recognised by nAbs. Selection for amino acid
305 changes in predicted nAb epitopes suggests nAbs are involved in an effective BoCoV immune response
306 to BoCoV. In the case of MHC-I, reduced support for these epitopes by could be due to downregulation
307 of the MHC-I presentation by BoCoV. HCoV OC43 [26] and other *Coronaviridae* species have been
308 shown to suppress MHC-I antigen presentation [42, 43]. Therefore, there would be no selective
309 advantage to non-synonymous mutations in these regions.

310 The COVID-19 pandemic has shown that administration of a vaccine can drive diversification and
311 reduce the protection offered by the vaccine to natural isolates. Widespread administration of a vaccine
312 means most the population harbours an effective immune response against a particular version of the
313 virus, which selects for amino acid changes occurring in immune epitopes. Updating the vaccine to
314 resemble the clinical isolates more closely is one strategy to maintain effectiveness of the vaccine.
315 Natural isolates evolving away from the Mebus vaccine strain appears to be occurring in Ireland because
316 amino acid changes were observed in 48 (3.52%) and 9 (2.12%) spike and HE residues, respectively.
317 These amino acid changes may also occur in regions important to immune recognition because 13
318 changes in spike and 4 changes in HE overlap with predicted immune epitopes. It is currently unclear
319 whether this evolution of the circulating BoCoV strains has reduced the effectiveness of the vaccine,
320 but the rate of continual infection and disease associated with BoCoV suggests it may be contributing.
321 Designing a subunit vaccine based on currently circulating Irish clinical isolates of BoCoV could
322 generate a more protective vaccine for cattle based in Ireland.

323 Almost all spike sequences from this study were in one of two clades suggesting most isolates
324 descended from two distinct introductions of the virus into Ireland or diverged along two separate
325 branches following a single introduction. Sample 129 formed an outgroup on its own and was the only
326 isolate not grouped into the two main clades. This divergence from other sequences could be driven by
327 adaptation to a unique environment or be the result of a third introduction of BoCoV into Ireland, which
328 is less frequently detected. Less frequent detection could be caused by reduced sampling frequency
329 associated with reduced pathogenicity, or it could be caused by reduced prevalence, which could
330 indicate a recent introduction into Ireland. Another possibility is that this variant has a selective
331 disadvantage causing it to be outcompeted by other variants thus reducing its prevalence. Relatedness
332 of Sample 129 with other isolates from this study was only based on the spike gene because full-length
333 HE sequence was not obtainable. The HE of most sequences from this study fell into one of two clades,
334 which predominantly overlapped with the two main clades of the spike tree. Two isolates, namely
335 Sample 123 and Sample 128, switched clade suggesting recombination had occurred around the 3' end
336 of the HE gene or 5' end of the spike gene, which was supported by GARD analysis showing
337 breakpoints around both regions. Sample 117 also no longer clustered with either of the two main clades
338 of the phylogenetic tree. Interestingly, when phylogenetically classified Sample 117 sits in the
339 Asian/American clade, which could indicate recombination between a European isolate and an

Asian/American isolate. This would require co-infection of an animal with both strains. Import and export of cattle into and from Ireland is rare [44], especially with other continents, so one possibility is that the recombination event occurred elsewhere in Europe and the animal was then brought into Ireland. Alternatively, an Asian/American strain of BoCoV could have been brought to Ireland by another mammalian species and the recombination event occurred in Ireland.

Conclusion

Phylogenetic analysis of sequences from this study suggested there have been at least two introductions of BoCoV into Ireland, and for the first time, sequence obtained from a European isolate has been shown to sit in the Asian/American clade along with evidence of recombination between European and Asian/American isolates. This study has predicted several BoCoV immune epitopes and corroborated several of those in spike and HE with selection analysis and amino acid changes observed from sequence data suggesting some of the predicted immune epitopes could be recognised by the bovine immune response. Overlap between predicted immune epitopes and amino acid changes observed when comparing the Mebus vaccine strain with clinical isolates could reduce the protection offered by Mebus-based vaccines. It is recommended that studies to assess the impact of this virus evolution on the effectiveness of currently used vaccines is assessed to identify whether vaccines require updating.

Acknowledgement: We thank the team of people in the Department of Agriculture Food and the Marine diagnostic laboratory for screening bovine samples for the presence of BoCoV by PCR.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

Funding Information: This study was funded by the Irish Department of Agriculture, Food and the Marine, grant number 2021R664.

References

1. **Vlasova AN, Saif LJ**. Bovine Coronavirus and the Associated Diseases. *Front Vet Sci* 2021;8:293.
2. **Gomez DE, Weese JS**. Viral enteritis in calves. *Can Vet J* 2017;58:1267.
3. **Fulton RW**. Viruses in Bovine Respiratory Disease in North America: Knowledge Advances Using Genomic Testing. *Vet Clin North Am Food Anim Pract* 2020;36:321.
4. **O'Neill R, Mooney J, Connaghan E, Furphy C, Graham DA**. Patterns of detection of respiratory viruses in nasal swabs from calves in Ireland: a retrospective study. *Vet Rec* 2014;175:351–351.
5. **Ellis J**. What is the evidence that bovine coronavirus is a biologically significant respiratory pathogen in cattle? *Can Vet J* 2019;60:147.
6. **Vlasak,' R, Luytjes W, Leider,' J, Spaan W, Palese1 P**. The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetylesterase activity. *J Virol* 1988;62:4686–4690.

374 7. **Lin X, O'Reilly KL, Burrell ML, Storz J**. Infectivity-neutralizing and hemagglutinin-
375 inhibiting antibody responses to respiratory coronavirus infections of cattle in pathogenesis of
376 shipping fever pneumonia. *Clin Diagn Lab Immunol* 2001;8:357–362.

377 8. **Milane G, Kourtesis AB, Dea S**. Characterization of monoclonal antibodies to the
378 hemagglutinin-esterase glycoprotein of a bovine coronavirus associated with winter dysentery
379 and cross-reactivity to field isolates. *J Clin Microbiol* 1997;35:33–40.

380 9. **Hussain KA, Storz J, Kousoulas KG**. Comparison of bovine coronavirus (BCV) antigens:
381 Monoclonal antibodies to the spike glycoprotein distinguish between vaccine and wild-type
382 strains. *Virology* 1991;183:442.

383 10. **Workman AM, Kuehn LA, McDaneld TG, Clawson ML, Chitko-McKown CG, *et al.***
384 Evaluation of the effect of serum antibody abundance against bovine coronavirus on bovine
385 coronavirus shedding and risk of respiratory tract disease in beef calves from birth through the
386 first five weeks in a feedlot. *Am J Vet Res* 2017;78:1065–1076.

387 11. **Workman AM, Kuehn LA, McDaneld TG, Clawson ML, Loy JD**. Longitudinal study of
388 humoral immunity to bovine coronavirus, virus shedding, and treatment for bovine respiratory
389 disease in pre-weaned beef calves. *BMC Vet Res* 2019;15:1–15.

390 12. **Dhufaigh KN, McCabe M, Cormican P, Cuevas-Gomez I, McGee M, *et al.*** Genome
391 Sequence of Bovine Coronavirus Variants from the Nasal Virome of Irish Beef Suckler and Pre-
392 weaned Dairy Calves Clinically Diagnosed with Bovine Respiratory Disease. *Microbiol Resour*
393 *Announc*;11. Epub ahead of print 17 November 2022. DOI: 10.1128/MRA.00821-22.

394 13. **Gunn L, Collins PJ, O'Connell MJ, O'Shea H**. Phylogenetic investigation of enteric bovine
395 coronavirus in Ireland reveals partitioning between European and global strains. *Ir Vet J*;68.
396 Epub ahead of print 30 December 2015. DOI: 10.1186/S13620-015-0060-3.

397 14. **Decaro N, Elia G, Campolo M, Desario C, Mari V, *et al.*** Detection of bovine coronavirus
398 using a TaqMan-based real-time RT-PCR assay. *J Virol Methods* 2008;151:167.

399 15. **Sievers F, Higgins DG**. Clustal Omega for making accurate alignments of many protein
400 sequences. *Protein Sci* 2018;27:135–145.

401 16. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, *et al.*** Fast, scalable generation of high-
402 quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.

403 17. **Procter JB, Carstairs GM, Soares B, Mourão K, Ofoegbu TC, *et al.*** Alignment of Biological
404 Sequences with Jalview. *Methods Mol Biol* 2021;2231:203–224.

405 18. **Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ**. Jalview Version 2—a
406 multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25:1189–
407 1191.

408 19. **Troshin P V., Procter JB, Barton GJ**. Java bioinformatics analysis web services for multiple
409 sequence alignment—JABAWS:MSA. *Bioinformatics* 2011;27:2001–2002.

410 20. **Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW**. GARD: a genetic
411 algorithm for recombination detection. *Bioinformatics* 2006;22:3096–3098.

412 21. **Weaver S, Shank SD, Spielman SJ, Li M, Muse S V., *et al.*** Datamonkey 2.0: A Modern Web
413 Application for Characterizing Selective and Other Evolutionary Processes. *Mol Biol Evol*
414 2018;35:773–777.

415 22. **Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, *et al.*** Detecting Individual Sites
416 Subject to Episodic Diversifying Selection. *PLOS Genet* 2012;8:e1002764.

417 23. **Murrell B, Moola S, Mabona A, Weighill T, Sheward D, *et al.*** FUBAR: A Fast,
418 Unconstrained Bayesian AppRoximation for Inferring Selection. *Mol Biol Evol* 2013;30:1196–

419 1205.

420 24. **Awadelkareem EA, Hamdoun S**. Insilco Vaccine Design of spike and hemagglutinin esterase
421 proteins of Bovine Coronavirus. Epub ahead of print 29 July 2022. DOI: 10.21203/RS.3.RS-
422 1791707/V1.

423 25. **Vijgen L, Keyaerts E, Moës E, Thoelen I, Wollants E, *et al.*** Complete Genomic Sequence of
424 Human Coronavirus OC43: Molecular Clock Analysis Suggests a Relatively Recent Zoonotic
425 Coronavirus Transmission Event. *J Virol* 2005;79:1595.

426 26. **Becerra-Artiles A, Nanaware PP, Muneeruddin K, Weaver GC, Shaffer SA, *et al.***
427 Immunopeptidome profiling of human coronavirus OC43-infected cells identifies CD4 T cell
428 epitopes specific to seasonal coronaviruses or cross-reactive with SARS-CoV-2. *bioRxiv*. Epub
429 ahead of print 1 December 2022. DOI: 10.1101/2022.12.01.518643.

430 27. **Wang C, Hesketh EL, Shamorkina TM, Li W, Franken PJ, *et al.*** Antigenic structure of the
431 human coronavirus OC43 spike reveals exposed and occluded neutralizing epitopes. *Nat*
432 *Commun*;13. Epub ahead of print 1 December 2022. DOI: 10.1038/S41467-022-30658-0.

433 28. **Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, *et al.*** SWISS-MODEL:
434 homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018;46:W296.

435 29. **Zeng Q, Langereis MA, Van Vliet ALW, Huizinga EG, De Groot RJ**. Structure of
436 coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution.
437 *Proc Natl Acad Sci U S A* 2008;105:9065.

438 30. **Schrödinger, L., & DeLano W**. PyMOL. https://pymol.org/2/ (2020).

439 31. **Hasoksuz M, Sreevatsan S, Cho KO, Hoet AE, Saif LJ**. Molecular analysis of the S1 subunit
440 of the spike glycoprotein of respiratory and enteric bovine coronavirus isolates. *Virus Res*
441 2002;84:101.

442 32. **Martínez N, Brandão PE, de Souza SP, Barrera M, Santana N, *et al.*** Molecular and
443 phylogenetic analysis of bovine coronavirus based on the spike glycoprotein gene. *Infect Genet*
444 *Evol* 2012;12:1870.

445 33. **Saitou N, Nei M**. The neighbor-joining method: a new method for reconstructing phylogenetic
446 trees. *Mol Biol Evol* 1987;4:406–425.

447 34. **Gouy M, Tannier E, Comte N, Parsons DP**. Seaview Version 5: A Multiplatform Software
448 for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation.
449 *Methods Mol Biol* 2021;2231:241–260.

450 35. **Letunic I, Bork P**. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
451 display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.

452 36. **Franzo G, Drigo M, Legnardi M, Grassi L, Pasotto D, *et al.*** Bovine Coronavirus: Variability,
453 Evolution, and Dispersal Patterns of a No Longer Neglected Betacoronavirus. *Viruses 2020, Vol*
454 *12, Page 1285* 2020;12:1285.

455 37. **Li X, Giorg EE, Marichannegowda MH, Foley B, Xiao C, *et al.*** Emergence of SARS-CoV-
456 2 through recombination and strong purifying selection. *Sci Adv*;6. Epub ahead of print 1 July
457 2020. DOI: 10.1126/SCIADV.ABB9153/SUPPL_FILE/ABB9153_SM.PDF.

458 38. **Dumonteil E, Herrera C**. Polymorphism and Selection Pressure of SARS-CoV-2 Vaccine and
459 Diagnostic Antigens: Implications for Immune Evasion and Serologic Diagnostic Performance.
460 *Pathog 2020, Vol 9, Page 584* 2020;9:584.

461 39. **Armijos-Jaramillo V, Yeager J, Muslin C, Perez-Castillo Y**. SARS-CoV-2, an evolutionary
462 perspective of interaction with human ACE2 reveals undiscovered amino acids necessary for
463 complex stability. *Evol Appl* 2020;13:2168–2178.

40. **Kistler KE, Bedford T**. Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229E. *Elife* 2021;10:1–35.

41. **Zhang CY, Wei JF, He SN**. Adaptive evolution of the spike gene of SARS coronavirus: Changes in positively selected sites in different epidemic groups. *BMC Microbiol* 2006;6:1–10.

42. **Zhang F, Zang TM, Stevenson EM, Lei X, Copertino DC, *et al.*** Inhibition of major histocompatibility complex-I antigen presentation by sarbecovirus ORF7a proteins. *Proc Natl Acad Sci U S A*;119. Epub ahead of print 11 October 2022. DOI: 10.1073/PNAS.2209042119/-/DCSUPPLEMENTAL.

43. **Arshad N, Laurent-Rolle M, Ahmed WS, Hsu JC-C, Mitchell SM, *et al.*** SARS-CoV-2 accessory proteins ORF7a and ORF3a use distinct mechanisms to downregulate MHC-I surface expression. *bioRxiv* 2022;2022.05.17.492198.

44. **McGrath G, Tratalos JA, More SJ**. A visual representation of cattle movement in Ireland during 2016. *Ir Vet J* 2018;71:1–5.

Supplementary Information

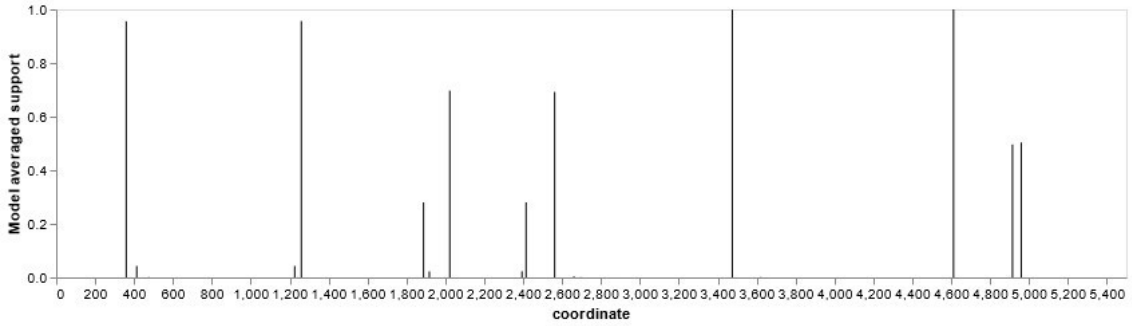


Figure S1: Recombination breakpoints detected in HE-S nucleotide alignment of sequences obtained in this study. The 20 full-length spike-HE sequences were aligned using Clustal Omega then recombination breakpoints detected using GARD.

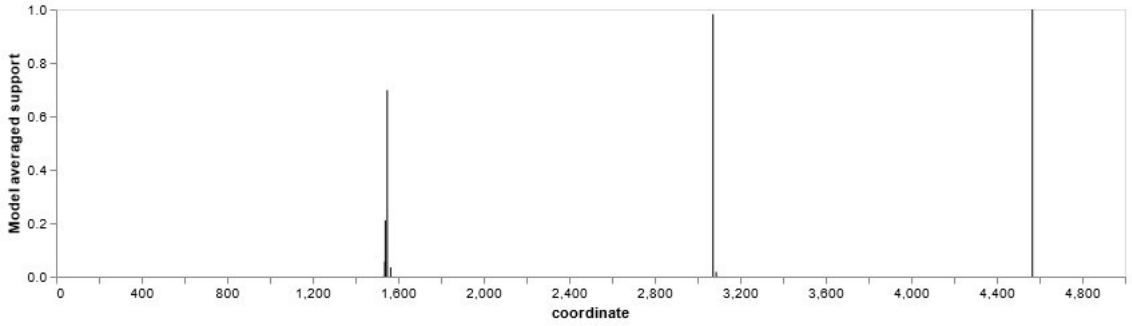


Figure S2: Recombination breakpoints detected in HE-S nucleotide alignment of GenBank sequences and sequences obtained in this study. Sequences were aligned using Clustal Omega then recombination breakpoints detected using GARD.

Spike	HE	Nucleocapsid	Membrane	Small Envelope
KX432213.1	MH810163.1	MH810163.1	FJ938067.1	KX432213.1
FJ415324.1	KX432213.1	MZ144606.1	EF424621.1	MH810163.1
FJ938065.1	FJ415324.1	KX432213.1	FJ425189.1	AF391541.1
FJ938066.1	ON093194.1	FJ415324.1	MG518518.1	AF391542.1
FJ938067.1	DQ915164.2	ON093194.1	DQ811784.2	ON142318.1
EF424621.1	EU019216.1	OP037415.1	EF424615.1	EF424621.1
EF424623.1	AB354579.1	DQ915164.2	EF424617.1	AY316299.1
FJ425184.1	AF391541.1	AY316299.1	EF424619.1	DQ811784.2
FJ425185.1	AF391542.1	EU019216.1	KU886219.1	LC494180.1
FJ425187.1	KX982264.1	AB354579.1	MN982198.1	LC494131.1
FJ425188.1	MG757138.1	AF391541.1	MN982199.1	LC494147.1
FJ425189.1	MG757139.1	AF391542.1	MW711287.1	LC494156.1
MG518518.1	MG757140.1	KX982264.1	ON146444.1	LC494160.1
DQ915164.2	LC494179.1	MG757138.1	OP037367.1	LC494168.1
FJ938063.1	LC494180.1	MG757139.1	OP037371.1	OP004056.1
DQ811784.2	LC494183.1	MG757140.1	OP037373.1	OP037365.1
EF424615.1	LC494184.1	LC494179.1	OP037374.1	OP037376.1
EF424616.1	LC494185.1	LC494180.1	OP037376.1	OP037380.1
EF424617.1	LC494188.1	LC494183.1	OP037380.1	OP037384.1
EU019216.1	LC494189.1	LC494126.1	OP037384.1	OP037397.1
AB354579.1	LC494190.1	LC494127.1	OP037385.1	OP037405.1
AF391541.1	LC494191.1	LC494128.1	OP037386.1	OP037407.1
AF391542.1	LC494126.1	LC494129.1	OP037388.1	OP037410.1
KU886219.1	LC494127.1	LC494130.1	OP037390.1	OP037414.1
KX982264.1	LC494128.1	LC494131.1	OP037391.1	OP037435.1
MG757138.1	LC494129.1	LC494133.1	OP037392.1	OP037440.1
MG757139.1	LC494130.1	LC494134.1	OP037393.1	U00735.2
MG757140.1	LC494131.1	LC494135.1	OP037394.1	
LC494179.1	LC494132.1	LC494137.1	OP037395.1	
LC494180.1	LC494133.1	LC494138.1	OP037396.1	
LC494181.1	LC494134.1	LC494139.1	OP037397.1	
LC494183.1	LC494135.1	LC494140.1	OP037398.1	
LC494184.1	LC494138.1	LC494141.1	OP037399.1	
LC494185.1	LC494139.1	LC494142.1	OP037400.1	
LC494186.1	LC494140.1	LC494143.1	OP037401.1	
LC494187.1	LC494141.1	LC494147.1	OP037402.1	
LC494188.1	LC494142.1	LC494148.1	OP037403.1	
LC494189.1	LC494143.1	LC494153.1	OP037404.1	
LC494190.1	LC494144.1	LC494154.1	OP037407.1	
LC494191.1	LC494147.1	LC494156.1	OP037408.1	
LC494192.1	LC494148.1	LC494160.1	OP037409.1	
LC494126.1	LC494153.1	LC494161.1	OP037412.1	
LC494127.1	LC494156.1	LC494164.1	OP037416.1	
LC494128.1	LC494160.1	LC494165.1	OP037417.1	
LC494129.1	LC494161.1	LC494168.1	OP037418.1	

LC494130.1	LC494162.1	LC494170.1	OP037420.1	
LC494131.1	LC494163.1	LC494171.1	OP037421.1	
LC494132.1	LC494165.1	LC494172.1	OP037422.1	
LC494133.1	LC494166.1	LC494173.1	OP037425.1	
LC494134.1	LC494167.1	LC494174.1	OP037426.1	
LC494135.1	LC494170.1	LC494175.1	OP037427.1	
LC494137.1	LC494172.1	LC494177.1	OP037429.1	
LC494138.1	LC494173.1	LC494178.1	OP037430.1	
LC494139.1	LC494174.1	OP296992.1	OP037436.1	
LC494140.1	LC494175.1	MH043952.1	OP037442.1	
LC494141.1	LC494177.1	MH043953.1	OP037438.1	
LC494142.1	LC494178.1	MH043954.1	OP037439.1	
LC494143.1	OP296992.1	MH043955.1	OP037441.1	
LC494144.1	AF220295.1	LC642814.1	MH810163.1	
LC494145.1	MH043952.1	ON792941.1	KX432213.1	
LC494147.1	MH043953.1	ON792942.1	FJ415324.1	
LC494148.1	MH043954.1	ON792944.1	DQ915164.2	
LC494153.1	MH043955.1	ON792945.1	EU019216.1	
LC494156.1	LC642814.1	ON792946.1	AF391541.1	
LC494157.1	ON792941.1	ON792947.1	AF391542.1	
LC494160.1	ON792946.1	ON792948.1	KX982264.1	
LC494161.1	ON792947.1	ON792963.1	MG757138.1	
LC494162.1	ON792963.1	ON792956.1	LC494179.1	
LC494163.1	ON792954.1	ON792957.1	LC494180.1	
LC494164.1	ON792956.1	ON792961.1	LC494183.1	
LC494165.1	ON792957.1	AF058942.1	LC494126.1	
LC494166.1	ON792961.1	AF058943.1	LC494128.1	
LC494167.1	U00735.2	AF058944.1	LC494129.1	
LC494168.1	AF058943.1	ON142315.1	LC494130.1	
LC494169.1	AF058944.1	ON142316.1	LC494133.1	
LC494170.1	EF445634.1	ON142318.1	LC494134.1	
LC494171.1	AF339836.1	ON142319.1	LC494135.1	
LC494172.1	AH014866.2	ON142320.1	LC494138.1	
LC494173.1	AH014867.2	KM985634.1	LC494140.1	
LC494174.1	AH014868.2	KT318096.1	LC494145.1	
LC494175.1	AH014869.2	KT318094.1	LC494147.1	
LC494177.1	AH014871.2	KT318093.1	LC494153.1	
LC494178.1	AH014872.2	KT318092.1	LC494164.1	
MN982198.1	AH010241.2	KT318091.1	LC494166.1	
MN982199.1	AH010063.2	KT318089.1	LC494170.1	
OP296992.1	AH010061.2	KT318088.1	LC494174.1	
AF220295.1	AH010062.2	KT318087.1	LC494175.1	
MH043952.1	ON142315.1	KT318086.1	LC494177.1	
MH043953.1	ON142320.1	KT318085.1	OP296992.1	
MH043954.1	ON142316.1	KT318084.1	AF220295.1	
MH043955.1	ON142318.1	KT318083.1	MH043952.1	

MW711287.1	ON142319.1	EU401985.1	MH043955.1	
LC642814.1	OP186344.1	EU401984.1	LC642814.1	
ON146444.1	OP186340.1	EU401983.1	ON792941.1	
ON792941.1	OP186339.1	EU401982.1	ON792942.1	
ON792942.1	OP186338.1	EU401981.1	ON792944.1	
ON792944.1	OP186337.1	EF193074.1	ON792950.1	
ON792945.1	OP186336.1	EF193073.1	ON792951.1	
ON792946.1	OP186335.1	MN894884.1	ON792954.1	
ON792947.1	OP186334.1	MK903505.1	AF058942.1	
ON792948.1	MH203063.1	MF737175.1	AF058943.1	
ON792949.1	MH203062.1	MK095173.1	AF058944.1	
ON792950.1	MH203061.1	MK095170.1	AH014866.2	
ON792951.1	MH203060.1	MK095169.1	AH014871.2	
ON792963.1	MH203059.1	MK095166.1	AH014874.2	
ON792954.1	AF230528.1	MK095163.1	AH014875.2	
ON792955.1	AF230527.1	MK095162.1	ON142315.1	
ON792956.1	AF230526.1	MZ822375.1	ON142317.1	
ON792957.1	AF230525.1	MZ822373.1	ON142319.1	
ON792959.1	AF230524.1	MZ822372.1	ON142320.1	
ON792960.1	AF230523.1	MZ822371.1	OP866726.1	
ON792961.1	KM985632.1	MZ822370.1	OP866727.1	
ON792962.1	EU401979.1	MZ822369.1	OP866728.1	
OP037365.1	EU401976.1	MZ822368.1	MK095157.1	
OP037366.1	EU401975.1	MZ822367.1	MK095155.1	
OP037367.1	DQ389650.1	MH475915.1	MK095154.1	
OP037368.1	DQ389649.1		MK095153.1	
OP037370.1	DQ389648.1		MK095151.1	
OP037371.1	DQ389647.1		MK095149.1	
OP037372.1	DQ389646.1		EF371516.1	
OP037373.1	DQ389645.1			
OP037374.1	DQ389643.1			
OP037376.1	DQ389642.1			
OP037379.1	MK903507.1			
OP037380.1	HM573325.1			
OP037381.1	DQ994168.1			
OP037384.1	DQ994167.1			
OP037385.1	MZ825083.1			
OP037386.1	MZ825082.1			
OP037388.1	MZ825079.1			
OP037389.1	MZ825078.1			
OP037390.1	MH475914.1			
OP037391.1	MK095147.1			
OP037392.1	MK095146.1			
OP037393.1	MK095144.1			
OP037394.1	MK095143.1			
OP037395.1	MK095142.1			

OP037396.1	MK095140.1			
OP037397.1	MK095139.1			
OP037399.1	MK095137.1			
OP037400.1				
OP037401.1				
OP037402.1				
OP037403.1				
OP037404.1				
OP037405.1				
OP037406.1				
OP037407.1				
OP037408.1				
OP037409.1				
OP037410.1				
OP037411.1				
OP037412.1				
OP037414.1				
OP037415.1				
OP037416.1				
OP037417.1				
OP037419.1				
OP037420.1				
OP037421.1				
OP037422.1				
OP037423.1				
OP037424.1				
OP037425.1				
OP037426.1				
OP037427.1				
OP037428.1				
OP037429.1				
OP037430.1				
OP037431.1				
OP037432.1				
OP037433.1				
OP037434.1				
OP037435.1				
OP037436.1				
OP037437.1				
OP037438.1				
OP037439.1				
OP037440.1				
OP037441.1				
OP037442.1				
U00735.2				
AF058942.1				

AF058943.1				
AF058944.1				
EF445634.1				
HE616741.1				
HE616739.1				
HE616738.1				
OP186328.1				
OP186327.1				
OP186326.1				
OP186325.1				
OP186324.1				
OP186323.1				
OP186322.1				
OP186321.1				
OP186320.1				
OP186319.1				
OP186318.1				
OP186317.1				
OP186316.1				
OP186315.1				
OP186314.1				
OP186313.1				
MK046011.1				
MK046010.1				
MK046008.1				
MK046007.1				
MK046006.1				
MK046004.1				
MK046003.1				
AY935646.1				
AY935644.1				
AY935643.1				
AY935642.1				
AY935641.1				
AY935639.1				
AY935638.1				
AY935637.1				
KF272919.1				
MW881219.1				
MW881220.1				
OL840236.1				
OL840233.1				
MT975572.1				
MT975571.1				
MT975570.1				
MK903506.1				

MK095186.1				
MK095185.1				
MK095184.1				
MK095183.1				
MK095182.1				
MK095181.1				
MK095179.1				
MK095178.1				
MK095176.1				
MK095175.1				
MK095174.1				
MH203067.1				
MH203066.1				
MH203065.1				
MH203064.1				
MH197039.1				
MH197038.1				
MH197037.1				
KT318124.1				
KT318122.1				
KT318121.1				
KT318120.1				
KT318118.1				
KT318117.1				
KT318116.1				
KT318115.1				
KT318114.1				
KT318113.1				
KT318112.1				
KT318111.1				
KM985631.1				
KF169940.1				
KF169939.1				
KF169938.1				
KF169937.1				
KF169936.1				
KF169935.1				
KF169934.1				
KF169933.1				
KF169932.1				
KF169931.1				
KF169930.1				
KF169929.1				
KF169926.1				
KF169925.1				
KF169924.1				

KF169923.1				
KF169922.1				
KF169921.1				
KF169920.1				
KF169919.1				
KF169918.1				
KF169917.1				
KF169915.1				
KF169914.1				
KF169913.1				
KF169912.1				
KF169910.1				
KF169909.1				
KF169908.1				
EU401989.1				
EU686689.1				
EU401988.1				
EU401987.1				
EU401986.1				
DQ389641.1				
DQ389640.1				
DQ389639.1				
DQ389638.1				
DQ389637.1				
DQ389636.1				
DQ389635.1				
DQ389634.1				
DQ389633.1				
DQ389632.1				
EF193075.1				
OL456213.1				
MW521194.1				
MW521193.1				
MW521192.1				
MW521191.1				
MW521189.1				
MW521188.1				
MW521187.1				
MW521186.1				
MW521185.1				
MW521184.1				
MW521183.1				
MW521182.1				
MW521181.1				
MW521180.1				
MW521179.1				

MW521177.1				
MW521176.1				
MW521175.1				
OL990398.1				
MZ603735.1				
MW711304.1				
MW711303.1				
MW711302.1				
MW711301.1				
MW711300.1				
MW711299.1				
MW711298.1				
MW711297.1				
MW711295.1				
MW711294.1				
MW711293.1				
MW711292.1				
MW711291.1				
MW711290.1				
MW711289.1				
MN982181.1				
MN982180.1				
MN982178.1				
MN982176.1				
MN982175.1				
MN982174.1				
MN982173.1				
MN982171.1				
MN982170.1				
MN982169.1				
MN982167.1				
MN982166.1				
MG757144.1				
MG757143.1				
M64668.1				
M64667.1				
HM573330.1				
HM573326.1				
DQ389660.1				
DQ389659.1				
DQ389658.1				
DQ389657.1				
DQ389656.1				
DQ389655.1				
DQ389654.1				
DQ389653.1				

DQ389652.1				
D00731.1				
EU814647.1				
EU814648.1				
MH741424.1				

Table S1: Spike, HE, nucleocapsid, membrane and small envelope GenBank Accession numbers for sequences used in selection and phylogenetic analyses.