Molecular and phylogenetic analysis of bovine coronavirus based on the spike glycoprotein gene

Nadia Martínez [a], Paulo E. Brandão [b], Sibele Pinheiro de Souza [b], Maritza Barrera [a], Nelson Santana [b,c], Heidy Díaz de Arce [a], Lester J. Pérez [a,*]

[a] Centro Nacional de Sanidad Agropecuaria (CENSA), La Habana, Cuba
[b] Department of Preventive Veterinary Medicine and Animal Health, College of Veterinary Medicine, University of São Paulo, São Paulo, Brazil
[c] Universidad Nacional de Colombia, Bogotá, Colombia

ARTICLE INFO

ABSTRACT

Bovine coronavirus has been associated with diarrhoea in newborn calves, winter dysentery in adult cattle and respiratory tract infections in calves and feedlot cattle. In Cuba, the presence of BCoV was first reported in 2006. Since then, sporadic outbreaks have continued to occur. This study was aimed at deepening the knowledge of the evolution, molecular markers of virulence and epidemiology of BCoV in Cuba. A total of 30 samples collected between 2009 and 2011 were used for PCR amplification and direct sequencing of partial or full S gene. Sequence comparison and phylogenetic studies were conducted using partial or complete S gene sequences as phylogenetic markers. All Cuban bovine coronavirus sequences were located in a single cluster supported by 100% bootstrap and 1.00 posterior probability values. The Cuban bovine coronavirus sequences were also clustered with the USA BCoV strains corresponding to the GenBank accession numbers EF424621 and EF424623, suggesting a common origin for these viruses. This phylogenetic cluster was also the only group of sequences in which no recombination events were detected. Of the 45 amino acid changes found in the Cuban strains, four were unique.

1. Introduction

Bovine coronavirus (BCoV) was first identified in association with diarrhoea in newborn calves (Mebus et al., 1973) and later associated with winter dysentery (WD) in adult cattle (Saif et al., 1991) and respiratory tract infections in calves and feedlot cattle (Storz et al., 2000). Although the affected animals rarely die, coronavirus infection causes dramatic reductions in milk production in dairy herds and loss of body condition in both calves and adults (Saif et al., 1998), resulting in severe economic losses. Thus, BCoV is currently considered an important pathogen that causes enteric disease, often in combination with clinical respiratory signs.

BCoV is included in the genus *Betacoronavirus* of the family Coronaviridae, which, together with the families *Arteriviridae* and *Roniviridae*, constitute the order *Nidovirales* (International Committee for Taxonomy of Viruses (ICTV): http://talk.ictvonline.org/cfs-filesystemfile.ashx/key/CommunityServer.Components.PostAttachments/00.00.00.06.26/2008.085_2D00_122V.01.Coronaviridae.pdf). The BCoV genome consists of a single molecule of linear, positive-sense, single-stranded RNA of 31 kb in length that is transcribed into a nested set of several 3′-coterminal subgenomic mRNAs that produce both non-structural and structural proteins (Chouljenko et al., 2001). The virion contains five structural proteins: the nucleocapsid (N) protein, the transmembrane (M) protein, the small envelope (E) protein, the haemagglutinin–esterase (HE) protein and the spike (S) protein (Lai and Cavanagh, 1997).

The S glycoprotein is important for viral entry and pathogenesis, forms large petal-shaped spikes on the surface of the virion and is cleaved into S1 (N-terminus) and S2 (C-terminus) subunits (Abraham et al., 1990). The S1 is the globular subunit and is responsible for virus binding to host–cell receptors (Kubo et al., 1994), the induction of neutralising antibody expression (Yoo and Deregt, 2001) and haemagglutinin activity (Schultze et al., 1991). It's sequences are variable, and mutations in this region have been associated with changes in antigenicity and viral pathogenicity (Ballesteros et al., 1997). On the other hand, the S2 is the transmembrane subunit and is required to mediate the fusion of viral and cellular membranes (Luo and Weiss, 1998).

Variations in the host range and tissue tropism of the coronaviruses have been largely attributable to variations in the S glycoprotein (Gallagher and Buchmeier, 2001). Therefore, to identify biological, antigenic and genetic characteristics that are distinct between respiratory and enteric BCoV strains, several studies based on partial or complete S gene sequences have been conducted (Brandão et al., 2006; Decaro et al., 2008; Hasoksuz et al.,

2002; Kanno et al., 2007; Liu et al., 2006). Although no clear markers have been established, comparative nucleotide sequence analyses have been useful for investigating the molecular phylogeny of BCoV (Brandão et al., 2006; Decaro et al., 2008; Kanno et al., 2007; Liu et al., 2006).

In Cuba, the presence of BCoV was first reported in 2006 (Barrera et al., 2006), and sporadic outbreaks have continued to occur since that time (Martínez et al., 2010).

In this study, sequence comparisons and phylogenetic studies based on S gene sequences were performed to deepen the knowledge of the evolution, potential molecular markers of virulence and epidemiology of BCoV in Cuba.

2. Materials and methods

2.1. Samples collection

A total of 30 faecal samples from dairy and beef cows that were affected with enteric manifestations resembling WD were selected from a total of 136 samples that were sent to the Animal Virology Group of the CENSA for BCoV diagnosis. The samples for this study were selected based on the geographic region of origin and the year of the collection (Fig. 1).

2.2. Laboratory procedure

To inactivate potential RT-PCR inhibitors contained in the faecal samples, the faeces were diluted in nuclease-free water (Promega, Madison, WI, USA) at a ratio of 3:1 (v/v). The final suspensions were centrifuged at 5000g for 10 min at 4 °C. RNA was extracted from 250 μL of supernatant recovered using the TRIzol reagent (Invitrogen™/Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. Finally, the RNA pellet was diluted in 30 μL of nuclease-free water (Promega, Madison, WI, USA).

First-strand complementary DNA (cDNA) was synthesised using Moloney-Murine leukaemia virus reverse transcriptase (M-MLV RT) (Invitrogen) and random primers (50 ng/μL) (Invitrogen) in a 20 μL final reaction volume. The cDNA of each sample was screened for the BCoV genome using the PCR method described by Tsunemitsu et al. (1999).

2.3. Sequencing

In the first approach, the hypervariable region of the S1 gene (approx. 488 bp) was amplified from BCoV-positive samples using the primer pairs reported by Souza et al. (2010) and the reaction conditions described by Brandão et al. (2006). The full S gene was then amplified from four positive samples (see samples marked with asterisk in Fig. 1) using several primer pairs (Table 1). Each fragment of the S gene was amplified in a reaction volume of 50 μL containing 10 μL cDNA, 2.5 U Platinum® Taq DNA polymerase (Invitrogen), 200 μM of each dNTP, 2.5 mM MgCl$_2$ and 0.5 μM of each primer.

The resulting amplicons were purified from agarose gels using a GFX PCR DNA and GEL BAND Purification Kit® (GE Healthcare) and submitted to bi-directional DNA sequencing using a BigDye Terminator v3.1 cycle sequencing kit following the manufacturer's directions (Applied Biosystems). Sequencing products were read on an ABI PRISM-3100 Genetic Analyzer (Applied Biosystems). The sense and antisense sequences obtained from each amplicon were assembled, and a consensus sequence for each gene was generated using the ChromasPro V1.5 program (Technelysium Pvt. Ltd., 2009). Nucleotide BLAST analysis (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) was initially used to verify the identity of each fragment sequence obtained. The sequences were submitted to the GenBank database under accession numbers HE616734–HE616737 for the hypervariable region and HE616738–HE616741 for the full S gene.
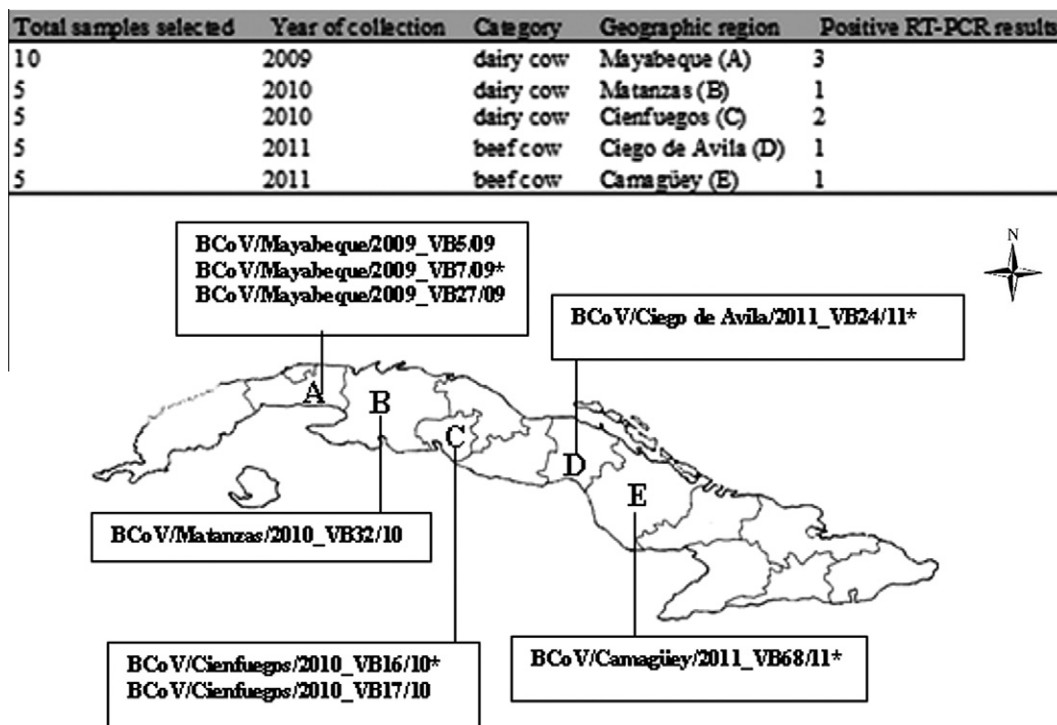
Total samples selected	Year of collection	Category	Geographic region	Positive RT-PCR results
10	2009	dairy cow	Mayabeque (A)	3
5	2010	dairy cow	Matanzas (B)	1
5	2010	dairy cow	Cienfuegos (C)	2
5	2011	beef cow	Ciego de Avila (D)	1
5	2011	beef cow	Camagüey (E)	1

BCoV/Mayabeque/2009_VB5/09
BCoV/Mayabeque/2009_VB7/09*
BCoV/Mayabeque/2009_VB27/09

BCoV/Ciego de Avila/2011_VB24/11*

BCoV/Matanzas/2010_VB32/10

BCoV/Cienfuegos/2010_VB16/10*
BCoV/Cienfuegos/2010_VB17/10

BCoV/Camagüey/2011_VB68/11*

Fig. 1. Map of Cuba showing the geographic distribution of the sample collection sites. The associated table indicates the quantity of samples analysed from each area and the positive RT-PCR results obtained from each of them (*complete S gene sequenced samples).

Table 1
Primer pairs used to amplify the full S gene.

Name	Primer sequence (5′–3′)	Location	Annealing temperature	Amplicon size	References
S1AF	ATGTTTTTGATACTTTTAATT	1–21	50 °C	655pb	Hasoksuz et al. (2002)
S1AR	AGTACCACCTTCTTGATAAA	654–635			
S1BF	ATGGCATTGGGATACAG	549–565	55 °C	490pb	Hasoksuz et al. (2002)
S1BR	TAATGGAGAGGGCACCGACTT	1039–1018			
S1CF	GGGTTACACCTCTCACTTCT	782–801	58 °C	769pb	Hasoksuz et al. (2002)
S1CR	GCAGGACAAGTGCCTATACC	1550–1531			
S1DF	GTCCGTGTAAATTGGATGGG	1460–1479	55 °C	827pb	Hasoksuz et al. (2002)
S1DR	TGTAGAGTAATCCACACAGT	2286–2267			
S1EF	TTACAAAAATCAAACACAGACAT	1855–1877	55 °C	877pb	Hasoksuz et al. (2002)
S1ER	AAACTTTATTACAATCGCTTCC	2731–2710			
S1FF	TCAATTTTTCCCCTGTATTAGG	2680–2702	55 °C	555pb	This study
S1FR	CMAGTCTRGATAGAATTTCTTGTAA	3234–3209			
S1GF	GCTACCAATTCTGCTTTAGTTA	3099–3121	55 °C	519pb	This study
S1GR	GTAGTAATAACCACTACCAGTG	3617–3595			
S1HF	TTTAGCTATGTCCCTACTAAGTA	3475–3498	55 °C	637pb	This study
S1HR	CCAATAAATCAAAGACGAACTTA	4112–4089			

2.4. Model selection

ModelTest V.3.0.6 (Posada and Crandall, 1998) was used to estimate the best-fit model using the Akaike information criterion (AIC). The best-fit model for each phylogenetic marker was selected and used for phylogenetic analysis.

2.5. Likelihood mapping

The phylogenetic signal of each sequence dataset was investigated by the likelihood mapping analysis of 100,000 random quartets generated using TreePuzzle (Strimmer and von Haeseler, 1997). In this strategy, if more than 30% of the dots fall into the centre of the triangle, the data are considered unreliable for the purposes of phylogenetic inference.

2.6. Phylogenetic trees

Phylogenetic relationships among partial or complete S gene sequences of unique BCoV isolates were analysed using the Bayesian Inference (BI) and Maximum Likelihood (ML) methodologies. Bayesian inference analyses were performed with the software MrBayes 3.1 (Huelsenbeck et al., 2001; Ronquist and Huelsenbeck, 2003). The MCMC searches were run with four chains for 5 million generations, with sampling of the Markov chain every 100 generations. At the end of each run, the convergence of the chains was inspected through the average standard deviation of split frequencies and the first 15% of the trees were discarded. After burn-in, the convergence was again assessed based on the effective sampling size (ESS) using Tracer software version 1.4 (http://tree.bio.ed.ac.uk/software/tracer/). Only ESS values of >250 were accepted. Bayesian trees with clade credibility for each marker were constructed using the posterior probability distribution. The trees were rooted using the Human-OC43 strain sequence (accession number NC005147). Finally, ML trees were computed using the PHYML v3.0 (Guindon and Gascuel, 2003), and confidence levels were estimated by 1000 bootstrap replicates.

2.7. Comparison of topologies

The topologies were tested by the Kishino and Hasegawa test (K–H) (Kishino and Hasegawa, 1989) and the Shimodaira–Hasegawa test (S–H) (Shimodaira and Hasegawa, 1999), which computed the log-likelihoods per site for each tree and compare the total log-likelihoods for each proposed topology, using the PAML v4.3 program (Yang, 2007). Ten thousand replicates were performed using the K–H and S–H topologies test by re-sampling the estimated log-likelihoods for each site (RELL model) (Kishino et al., 1990).

2.8. Recombination analyses

Searches for recombinant sequences and crossover regions in the BCoV S gene were performed using Geneconv (Padidam et al., 1999), RDP (Martin and Rybicki, 2000), MaxChi (Maynard Smith, 1992; Posada and Crandall, 2001), Chimera (Posada and Crandall, 2001), BootScan (Martin et al., 2005), SiScan (Gibbs et al., 2000), 3Seq (Boni et al., 2007) and LARD (Holmes et al., 1999), all implemented in RDP3 Beta 41 (Heath et al., 2006). All unique full S gene sequences available in the GenBank database were downloaded on December 1st, 2011 and tested. Programs were executed with modified parameter settings determined according to the guidelines in the RDP3 manual for the analysis of divergent sequences (available upon request). Recombinant sequences were tested by examining the S genes with a highest acceptable p value of 0.05, and Bonferroni's multiple comparison correction was used. Analyses were conducted twice to ensure the repeatability of results.

2.9. Multiple amino acid alignment

The full S gene sequences of each Cuban isolate were aligned with sequences available from the GenBank database, including S gene sequences from respiratory and enteric strains, sequences that showed closer phylogenetic relationships and the reference strain Mebus. The alignment was performed using the ClustalW method in the BioEdit Sequence Alignment Editor (Hall, 1999). In addition, the amino acid sequences for each gene of the Cuban isolates were obtained and compared using the BioEdit Sequence Alignment Editor (Hall, 1999).

The N-linked glycosylation sites were predicted with the N-GlycoSite web-based utility (http://hcv.lanl.gov/content/hcv-db/GLY-COSITE/glycosite.html), and the hydrophobicity profile was generated using the software package DAMBE v5.1.1 (http://web.hku.hk/_xxia/software/software.html) according to the Kyte and Doolittle method.

3. Results

3.1. Likelihood mapping

The phylogenetic noise in each data set was investigated by means of likelihood mapping. The percentage of dots falling in the central area of the triangles ranged from 0.3% for the full S gene
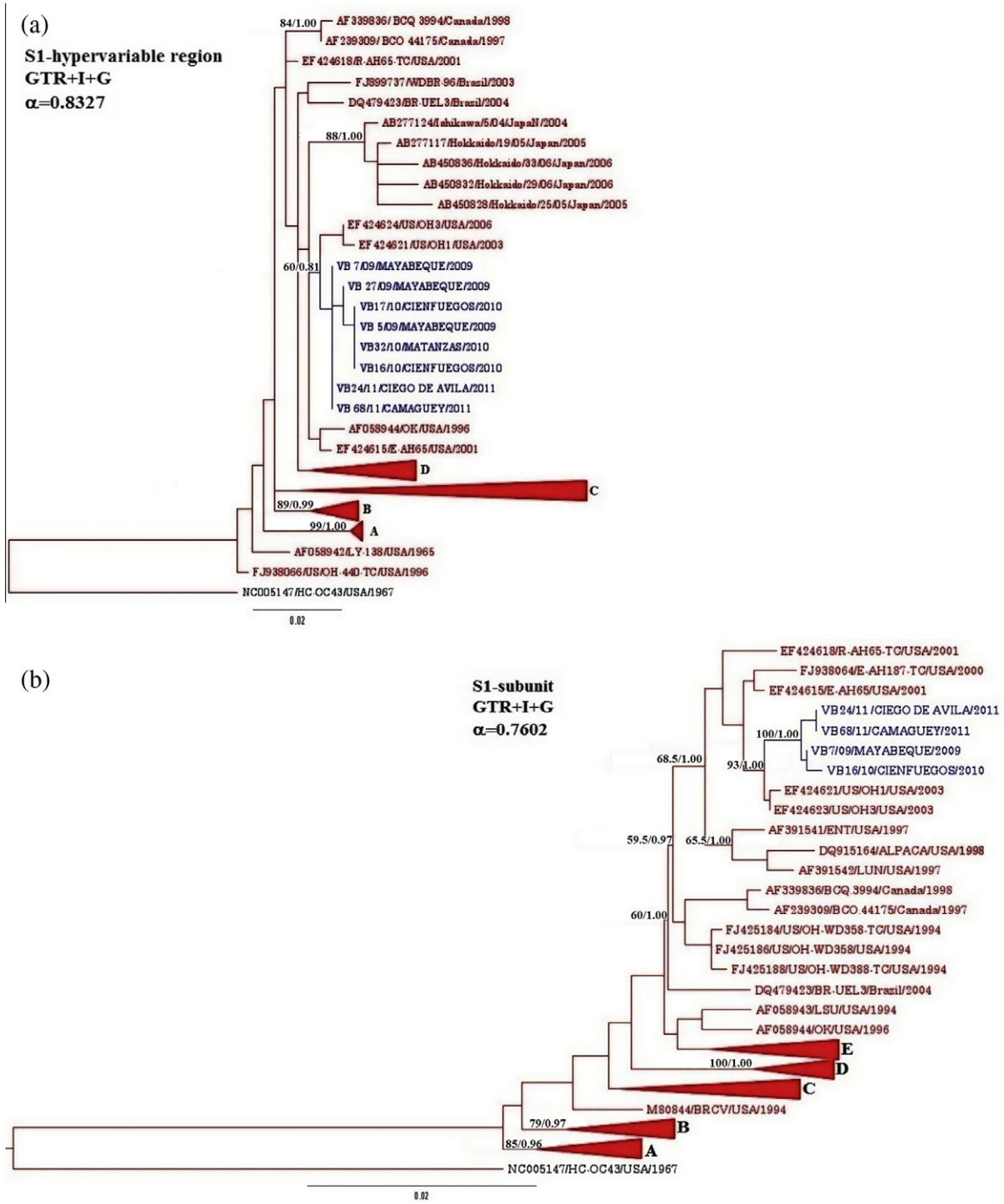
Fig. 2. Phylogenetic trees for BCoV sequences. Each phylogenetic marker, the best-fitted model and the gamma distribution shape parameter values used to infer phylogenies are indicated. The Cuban BCoV sequences were highlighted in blue, whereas the remaining sequences appear in red. Numbers along the branches refer to the percentages of confidence and posterior probability in the ML and BI analyses, respectively. Minor branch values were hidden. Sequences that had no close relationship with the Cuban sequences were collapsed (indicated in parentheses). (a) **A**: (EF445634, EU814647), **B**: (U06093, U06091, U06090), **C**: (AY606201, AY606197, AY606204, AY606195, AY606192, AY606203, AY606196, AY606199, AY606205, AY606202, M64668, AF181469, D00662, U00735, EF193075, EU401986, D00731, M80844, AF391541, AF391542, AF239307, FJ938064, DQ915164, DQ811784, FJ938063, AF239306) and **D**: (DQ320763, FJ425186, AY255831, AB277129, AB277121, AF058943, FJ425188, AY935639, AY935646, AY935644, AY935642, EF424619, DQ389655, DQ389657, AB277130, DQ389658, DQ389660, AY935643, DQ389653, DQ389652, DQ389656, DQ389659, DQ389634, DQ389635, DQ389640, DQ389641, DQ389639, DQ389636, DQ389654, HM573330, DQ389637); (b) **A**: (AF058942, D00731), **B**: (AB354579, M64668, EF193075, EU401989, EU401987, EU401988, U00735, D00662, AF181469, M64667), **C**: (FJ938063, DQ811784, AF239306, AF239307, U06093), **D**: (EU019216, EF445634, EU814647, EU814648), **E**: (DQ389637, DQ389636, DQ389640, DQ389638, DQ389656, FJ938066, AY935637, AY935638, AY935642, AY935641, AY935639, AY935644, AY935646, FJ425187, EF424619, AY935643, DQ389659, DQ389652, DQ389653, DQ389657, DQ389660, DQ389633, DQ389654, EU686689, EU401986, DQ389634, DQ389632, DQ389635, DQ389641, DQ389639); (c) **A**: (D00731, AF058942), **B**: (M64668, EU401989, EU401988, EU401987, AB354579, D00662, U00735, EF193075, M64667, AF181469), **C**: (EU814648, EU814647, EU019216, EF445634, M80844), **D**: (DQ811784, FJ938063), **E**: (AF058943, FJ425186, FJ425184, FJ425188, FJ425189), **F**: (HM573330, HM573326, DQ389660, DQ389657, DQ389658, DQ389655, DQ389654, AY935643, DQ389659, DQ389653, DQ389652, DQ389656, DQ389641, DQ389640, DQ389638, DQ389637, DQ389636, DQ389639, DQ389635, DQ389634, DQ389632, DQ389633, EU401986, EU686689).

Fig. 2. (continued)
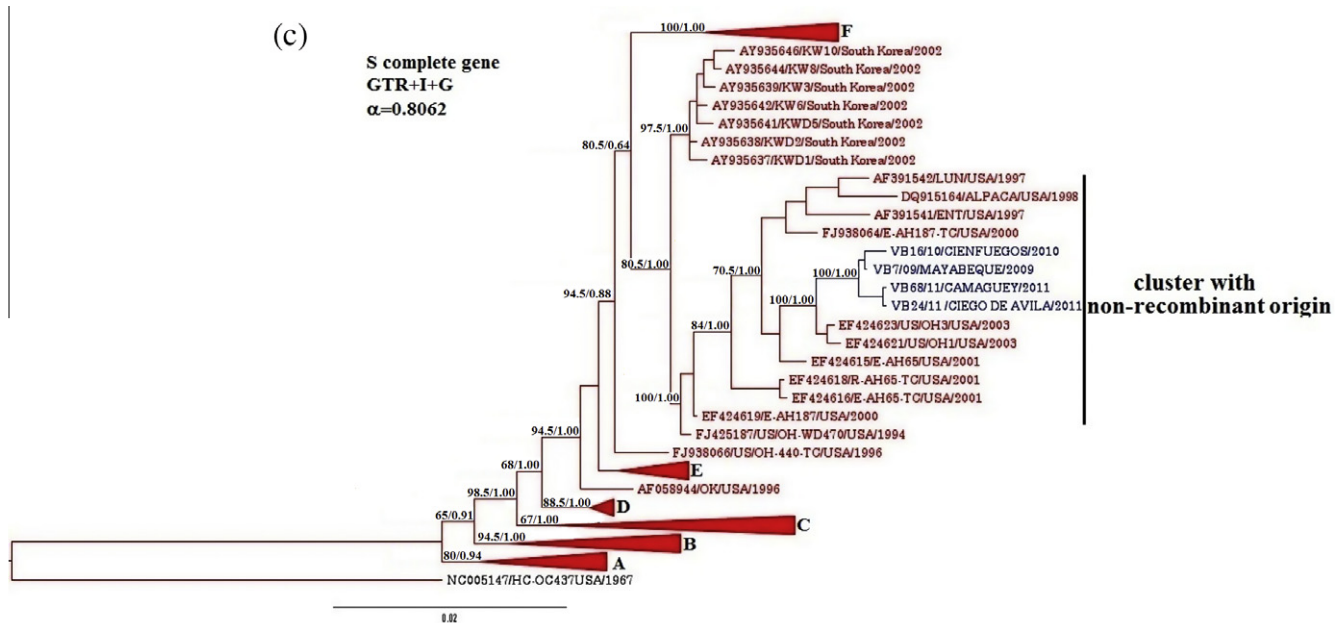
to 13.1% for the hypervariable S1 region. None of the datasets showed more than 30% noise, which enabled the use of the hypervariable S1 region, S1 region or full S gene to deduce the phylogenetic signal.

3.2. Sequence identity analysis

The nucleotide sequence identities of the S1 hypervariable region among the eight Cuban BCoV sequences obtained ranged from 99.4% to 100% and the deduced amino acid deduced ranged from 99.2% to 100%. Meanwhile, the four S complete gene sequences shared sequence identities ranging from 99.7% to 100% and 99.5% to 100% for nucleotide and deduced amino acid sequences, respectively. The BLASTn analysis for both the S1 hypervariable region and complete S gene showed the highest nucleotide sequence identities, 98% and 99%, respectively, with the BCoV strain identified as US/OH1/2003 (accession number EF424621), which was detected in a wild animal park in Ohio, USA.

3.3. Recombination analyses

A recombination analysis of the complete S gene was performed from the dataset used. The recombination breakpoints were determined using the RDP3 program. The sequences representing recombinant mosaic and major recombinant parental types are listed in Supplementary Table S1. Surprisingly, the Cuban BCoV strain sequences and a group of USA BCoV strain sequences with GenBank accession numbers AF391542, DQ915164, FJ938064, EF424621, EF424623, EF424615 and EF424616 were the only sequences in which no recombination events were detected (Supplementary Table S1 and Fig. 2c). All remaining BCoV strains whose complete S gene sequences are available in the GenBank database have at least one recombination event for this gene.

3.4. Phylogenetic analyses

The phylogenetic relationships among the BCoV strains were reconstructed by ML and BI analyses. The Bayesian tree was the best (S–H test −Ln = 12018.922∗) when S complete sequences were used as phylogenetic markers, but the support for this tree was not significantly different for the ML tree, with −Ln values ranging from 12018.922 to 12021.478. The ML tree was the best (S–H test −Ln = 7286.532∗) when S1 sequences were used as phylogenetic markers, but the support for this tree was not significantly different for the Bayesian tree, with −Ln values ranging from 7286.532 to 7309.785. Therefore, regardless of the phylogenetic method used or phylogenetic marker employed, all tree topologies were identical (Fig. 2a–c) and were supported by moderate to high confidence values.

The phylogenetic relationships constructed based on the hypervariable region of the S1 gene (approx. 488 bp), revealed that the eight Cuban BCoV strains were closely related (Fig. 2a). All Cuban BCoV strains were located in a same cluster along with the USA BCoV strains with GenBank accession numbers EF424621 and EF424624, suggesting a common origin. These phylogenetic relationships were supported by moderate to high confidence values (Fig. 2a).

The phylogenetic trees based on the complete S gene and the S1 region yielded the same phylogenetic relationships for the Cuban BCoV strains, which were similar to the phylogenetic trees obtained from the hypervariable region of the S1 gene. Thus, all Cuban BCoV strains were located in a same cluster, supported by 100% bootstrap values and 1.00 posterior probability values (Fig. 2b and c). In addition, the Cuban BCoV strains were located in the same cluster as the USA BCoV strains EF424621 and EF424623, suggesting a common origin.

3.5. Molecular characterisation

The deduced amino acid sequences of the full S genes of the four Cuban BCoV field strains were aligned and compared with the full S genes of other strains used as reference sequences (Fig. 3). Thus, no deletions or insertions were observed in the full S genes of the four Cuban BCoV field strains. The deduced amino acid sequences showed a total of 45 amino acid changes compared with the Mebus reference strain and contained 20 potential N-glycosylation sites (Fig. 3).

Several of the amino acid changes found in the Cuban strains were unique, such as: (i) A369T in the strain VB16/10/CIENFUEGOS/2010, which was located in immuno-reactive domain S1A;

(ii) a non-conservative substitution in the hypervariable region, G492D, was found in the strains VB7/09/MAYABEQUE/2009 and VB16/10/CIENFUEGOS/2010; (iii) the replacement I1237T was found in the strains VB24/11/CIEGO DE AVILA/2011 and VB68/11/CAMAGUEY/2011; and (iv) the substitution N1285K was observed in all of the Cuban sequences assessed. This substitution was located in the heptad repeat region (HR) (HR–C). However, the proteolytic cleavage site (KRRSRR) of the S protein at residue 768 in subunits S1 and S2 was conserved in all Cuban strains (Fig. 3).

In addition, specific amino acid signatures that have been linked with either respiratory or enteric tropism were found in the Cuban strains (see Fig. 3). Seven amino acid signatures that have been suggested to be possible virulence factors (Zhang et al., 1991) were also observed in the Cuban strains. Notably, several of the amino acid changes found were located in regions of functional importance for virus-cell interactions. Thus, residue changes were observed in the signal peptide, the hypervariable region, the immuno-reactive domain S1B and the first hydrophobic region of S2 (Fig. 3). Conversely, the amino acid replacements E965V, W984L and A988V in the first hydrophobic region of S2 increased the protein hydrophilicity (data not shown).

4. Discussion

Novel data on the diversity and ecology of BCoV are of interest for the scientific community, not only because this virus is considered an important veterinary pathogen but also because several coronavirus cross-species transmission events, as well as changes in virus tropism, have been described in the recent past (Zhang et al., 2007).

Sequence analyses suggest the potential for zoonotic transmission of a BCoV to humans (Vijgen et al., 2005), and researchers have previously confirmed the isolation of a bovine-like CoV from a human child with diarrhoea.

In the present study, a molecular analysis of Cuban BCoV field sequences from samples collected between 2009 and 2011 was conducted by comparison with all of the BCoV S gene sequences that were available in the GenBank database. In addition, phylogenetic comparisons based on partial or complete S gene sequences used as phylogenetic markers were conducted.

The entire or partial BCoV S gene sequences have been widely used for molecular and phylogenetic analyses of this viral agent (Liu et al., 2006; Park et al., 2006). The phylogenetic noise, investigated by likelihood mapping, revealed that although the analysis
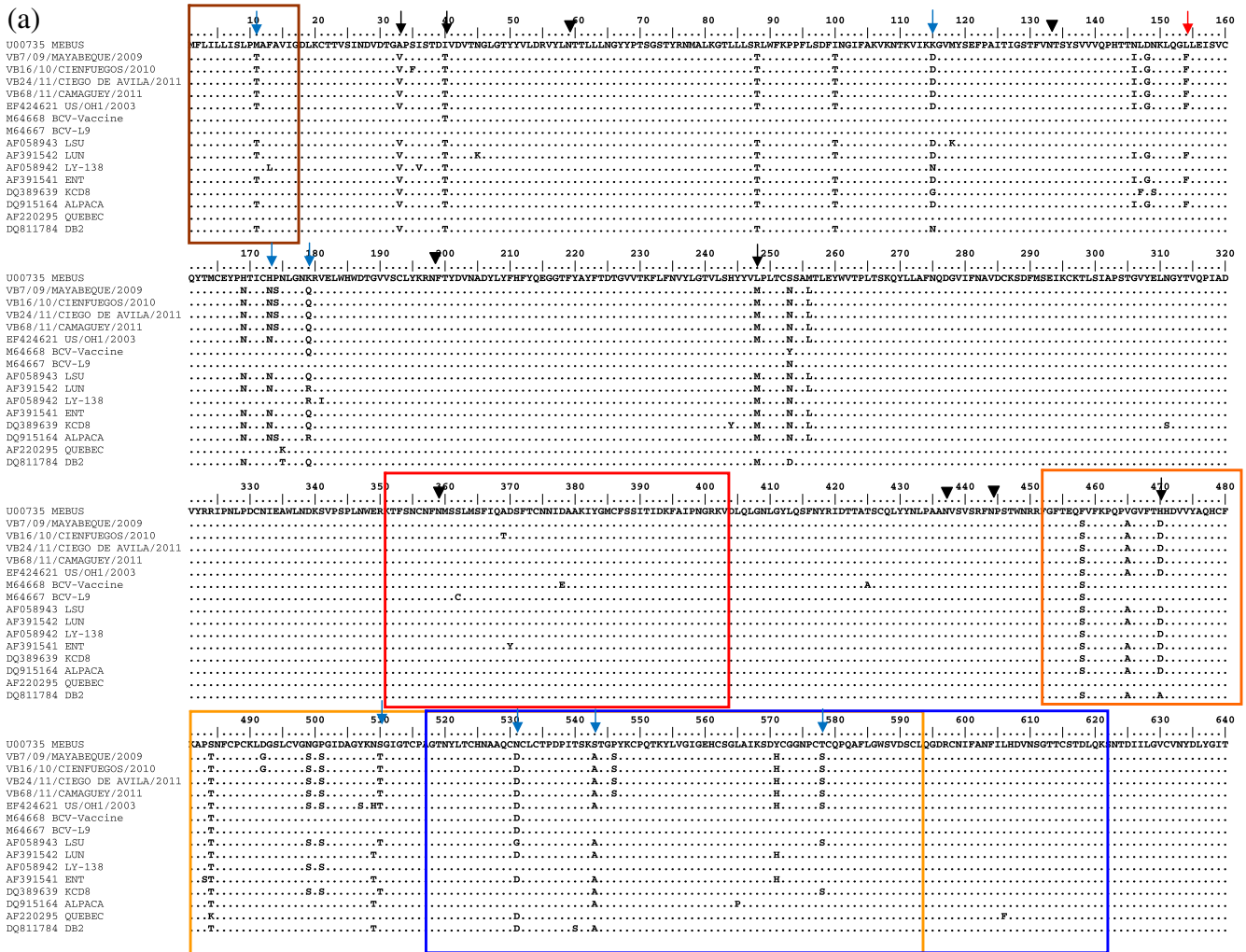

Fig. 3. Molecular analysis of the deduced amino acid sequences for the entire S genes of the Cuban field and reference strains. Regions important for virus-cell interactions were highlighted: the signal peptide region (red), the hypervariable region (orange), the proteolytic cleavage region (pink), the first hydrophobic region of the S2 subunit (grey), the HR–N domain (greenish-blue) and the HR–C domain (dark green). The immuno-reactive domains S1A and S1B were outlined in red and blue, respectively; the transmembrane region was outlined in black. The particular signatures linked either with a respiratory (blue arrow) or enteric (red arrow) tropism and virulence factors (black arrow) were also indicated. The potential N-glycosylation sites were marked with black triangles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
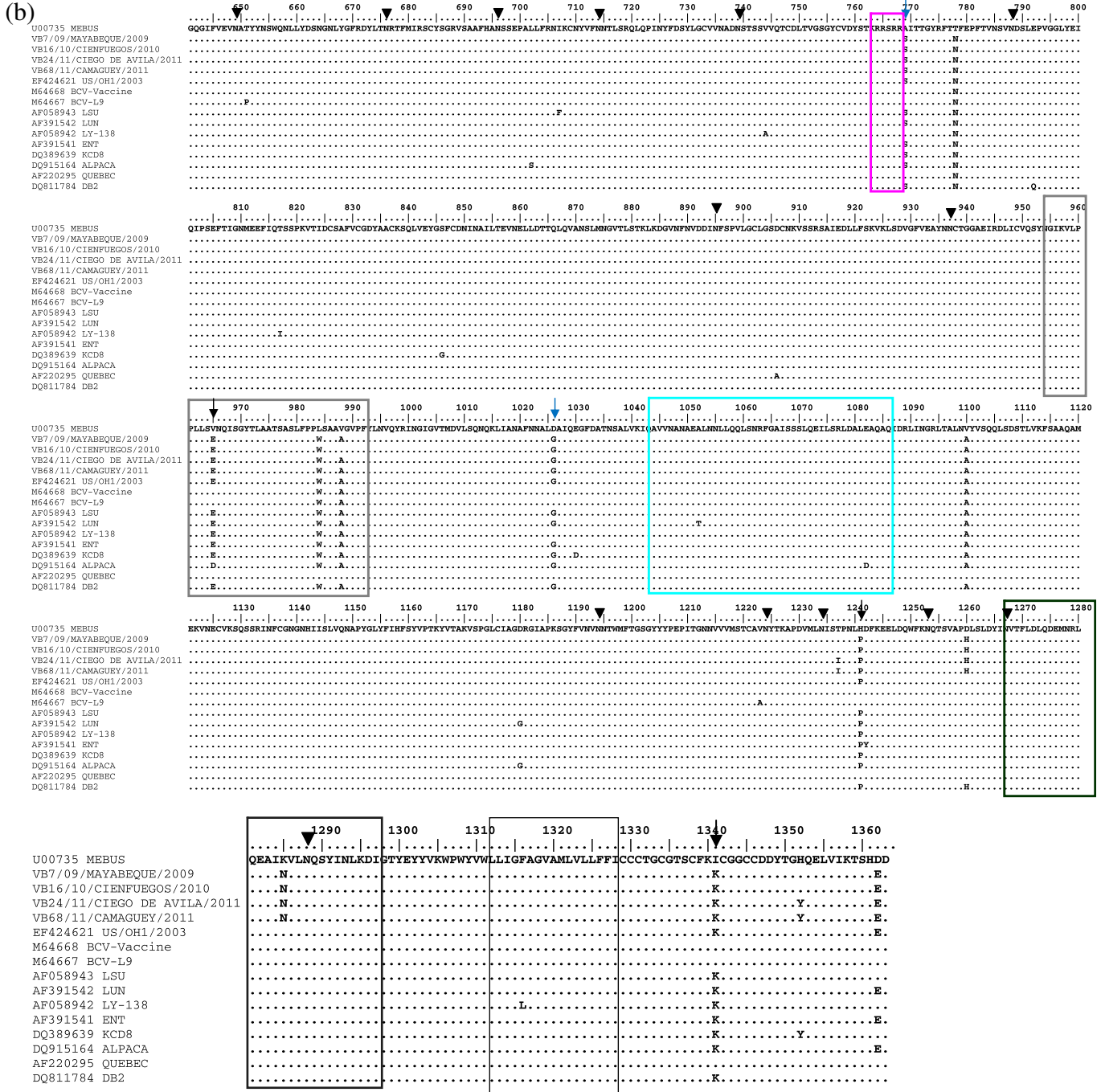
(b)

Fig. 3. (*continued*)

using the entire S gene yielded a better tree resolution with better-supported tree topologies, the partial S gene could still be useful as phylogenetic marker; less than 30.0% of the phylogenetic information is lost in comparison with the entire S gene.

The high sequence identity shown by the Cuban BCoV strains demonstrates the low divergence between these strains. On the other hand, the results of the phylogenetic analyses were well supported by different phylogenetic inference methods and high confidence values. Thus, all topologies obtained from each phylogenetic marker were reliable. The fact that all the Cuban BCoV strains were located in a single cluster supports the hypothesis that the BCoV strains currently circulating among Cuban

bovine herds share a common origin and are closely genetically related (Fig. 2c).

The Cuban BCoV strains clustered with the USA BCoV strains, and this cluster was the only one to contain strains with a non-recombinant origin, indicating an epidemiological link between Cuban and USA BCoVs. Considering the insular condition of the country, the fact that BCoV is only transmitted by direct contact between animals and that the first evidence of BCoV infection in Cuba was obtained in 2004 (Barrera et al., 2006) suggest that a USA BCoV strain(s) is the most likely origin of the Cuban BCoV strains.

Alekseev et al. (2008) reported that Coronaviruses from wild ruminants cluster with BCoV strains according to the year of

isolation/circulation, suggesting an epidemiological link among these strains. Therefore, there may be a connection between Cuban and US BCoV circulating in cattle during these years. However, how the virus was transmitted remains unclear.

The multiple amino acid alignment performed showed that the Cuban BCoV strains exhibited particular amino acid signatures that might be linked either with respiratory or enteric tropism. This has been supported by the idea that the respiratory strains could undergo an initial extensive replication in the nasal mucosa and subsequently spread to the gastrointestinal tract through the swallowing of large quantities of virus coated in mucous secretions. This initial respiratory amplification of BCoV and its protective coating by mucus may allow larger amounts of the labile, enveloped, but infectious virus to transit to the gut after swallowing, resulting in intestinal infection and faecal shedding (Saif, 2010). Moreover, the animals' nostrils could also be contaminated with enteric viruses by direct contact with and inhalation of faeces, allowing the infection of the respiratory tract with enteric strains (Zhang et al., 2007).

The unique Cuban BCoV polymorphism N1285K situated in the heptad repeat region (HR–C) could be involved in changes in the replication of the virus. HR–C is a critical element in the conformational change of the coiled-coil structure that is required for the interaction of the viral and host cell membranes, promoting the fusion of the lipid bilayers and the introduction of the nucleocapsid into the cytoplasm (Baker, 2008). Thus, the drastic change of a charged residue to a polar residue in the HR–C could influence the interaction between the coiled-coil structure and the host cell receptor.

On the other hand, it has been speculated that mutations in and around the heptad repeats of Coronavirus represent a pathway for virus cross-species transmission (Graham and Baric, 2010). Nonetheless, the role of the N1285K mutation in viral replication or virus cross-species transmission should be studied further.

It is also important to highlight that WD, which is the most severe clinical form of BCoV infection, occurs in adult cattle in the winter due to predisposing events such as cold temperatures and drinking cold water (White et al., 1989). However, the tropical weather conditions in Cuba are far from "cold temperatures in the winter season". Similarly, some WD outbreaks associated with BCoV infections during warmer seasons have been reported (Decaro et al., 2008; Park et al., 2006). These issues suggest that molecular changes in the currently circulating BCoV strains could have occurred to facilitate viral adaptation to the warmer environment. The molecular analysis of the S entire gene in the present work did not reveal any genetic markers linked to viral adaptation to hot weather. This suggests that the viral adaptation to temperature changes could be linked to others structural or non-structural genes.

Finally, the present study highlights the need for more in depth epidemiological and molecular investigations on BCoV infection based on the entire viral genome, to identify genetic markers linked to viral tropism and adaption to a warmer climate, which remain unknown by the scientific community.

Acknowledgement

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2012.05.007.

References

Abraham, S., Kienzle, T.E., Lapps, W., Brian, D.A., 1990. Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. Virology 176, 296–301.

Alekseev, K.P., Vlasova, A.N., Jung, K., Hasoksuz, M., Zhang, X., Halpin, R., Wang, S., Ghedin, E., Spiro, D., Saif, L.J., 2008. Bovine-like coronaviruses isolated from four species of captive wild ruminants are homologous to bovine coronaviruses, based on complete genomic sequences. J. Virol. 82, 12422–12431.

Baker, S.C., 2008. Coronaviruses: Molecular Biology. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), Desk Encyclopedia of General Virology. Elsevier, USA, pp. 445–453.

Ballesteros, M.L., Sanchez, C.M., Enjuanes, L., 1997. Two amino acid changes at the N-terminus of transmissible gastroenteritis coronavirus spike protein result in the loss of enteric tropism. Virology 227, 378–388.

Barrera, M., Rodríguez, E., Bentancourt, A., Frías, M.T., Brandão, P.E., 2006. First report in Cuba of bovine coronavirus detection in a winter dysentery outbreak. Span. J. Agric. Res. 4, 221–224.

Boni, M.F., Posada, D., Feldman, M.W., 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176, 1035–1047.

Brandão, P.E., Gregori, F., Richtzenhain, L.J., Rosales, C.A.R., Villarreal, L.Y.B., Jerez, J.A., 2006. Molecular analysis of Brazilian strains of bovine coronavirus (BCoV) reveals a deletion within the hypervariable region of the S1 subunit of the spike glycoprotein also found in human coronavirus OC43. Arch. Virol. 151, 1735–1748.

Chouljenko, V.N., Lin, X.Q., Storz, J., Kousoulas, K.G., Gorbalenya, A.E., 2001. Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia. J. Gen. Virol. 82, 2927–2933.

Decaro, N., Mari, V., Desario, C., Campolo, M., Elia, G., Martella, V., Greco, G., Cirone, F., Colaianni, M.L., Cordioli, P., Buonavoglia, C., 2008. Severe outbreak of bovine coronavirus infection in dairy cattle during the warmer season. Vet. Microbiol. 126, 30–39.

Gallagher, T.M., Buchmeier, M.J., 2001. Coronavirus spike proteins in viral entry and pathogenesis. Virology 279, 371–374.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.

Graham, R.L., Baric, R.S., 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J. Virol. 84, 3134–3146.

Guindon, S., Gascuel, O., 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. System. Biol. 52, 696–704.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hasoksuz, M., Sreevatsan, S., Cho, K.-O., Hoet, A.E., Saif, L.J., 2002. Molecular analysis of the S1 subunit of the spike glycoprotein of respiratory and enteric bovine coronavirus isolates. Virus Res. 84, 101–109.

Heath, L., van der Walt, E., Varsani, A., Martin, D.P., 2006. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. J. Virol. 80, 11827–11832.

Holmes, E.C., Worobey, M., Rambaut, A., 1999. Phylogenetic evidence for recombination in dengue virus. Mol. Biol. Evol. 16, 405–409.

Huelsenbeck, J.H., Ronquist, F., Nielsen, R., Bollback, J., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Kanno, T., Hatama, S., Ishihara, R., Uchida, I., 2007. Molecular analysis of the S glycoprotein gene of bovine coronaviruses isolated in Japan from 1999 to 2006. J. Gen. Virol. 88, 1218–1224.

Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29, 170–179.

Kishino, H., Miyata, T., Hasegawa, M., 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. 30, 151–160.

Kubo, H., Yamada, Y.K., Taguchi, F., 1994. Localization of neutralizing epitopes and the receptor-binding site within the aminoterminal 330 amino acids of the murine coronavirus spike protein. J. Virol. 68, 5403–5410.

Lai, M.M.C., Cavanagh, D., 1997. The molecular biology of coronaviruses. Adv. Virus Res. 48, 1–100.

Liu, L., Hägglund, S., Hakhverdyan, M., Alenius, S., Larsen, L.E., Belák, S., 2006. Molecular epidemiology of bovine coronavirus on the basis of comparative analyses of the S gene. J. Clin. Microbiol. 44, 957–960.

Luo, Z., Weiss, S., 1998. Roles in cell-to-cell fusion of two conserved hydrophobic regions in the murine coronavirus spike protein. Virology 244, 483–494.

Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.

Martin, D.P., Posada, D., Crandall, K.A., Williamson, C., 2005. A modified BOOTSCAN algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res. Hum. Retroviruses 21, 98–102.

Martínez, N., Rodríguez, M., Barrera, M., 2010. New outbreaks of bovine coronavirus in dairies of Havana. Rev. Salud Anim. 32, 135.

Maynard Smith, J., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.

Mebus, C.A., Stair, E.L., Rhodes, M.B., Twiehaus, M.J., 1973. Neonatal calf diarrhea: propagation, attenuation, and characteristics of a coronavirus like agent. Am. J. Vet. Res. 34, 145–150.

Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.

Park, S.J., Jeong, C., Yoon, S.S., Choy, H.E., Saif, L.J., Park, S.H., Kim, Y.J., Jeong, J.H., Park, S.I., Kim, H.H., Lee, B.J., Cho, H.S., Kim, S.K., Kang, M.I., Cho, K.O., 2006. Detection and characterization of bovine coronaviruses in fecal specimens of adult cattle with diarrhea during the warmer seasons. J. Clin. Microbiol. 44, 3178–3188.

Posada, D., Crandall, K., 1998. ModelTest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA 98, 13757–13762.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Saif, L.J., 2010. Bovine Respiratory Coronavirus. Vet. Clin. Food Anim. 26, 349–364.

Saif, L.J., Brock, K.V., Redman, D.R., Kohler, E.M., 1991. Winter dysentery in dairy herds: electron microscopic and serological evidence for an association with coronavirus infection. Vet. Rec. 128, 447–449.

Saif, L.J., Redman, D.R., Brock, K.V., Kohler, E.M., Heckert, R.A., 1998. Winter dysentery in adult dairy cattle: detection of coronavirus in the faeces. Vet. Rec. 123, 300–301.

Schultze, B., Gross, H.J., Brossmer, R., Herrler, G., 1991. The S protein of bovine coronavirus is a hemagglutinin recognizing 9-Oacetylated sialic acid as a receptor determinant. J. Virol. 65, 6232–6237.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Souza,S.P., Asano, K.M., Oliveira, R.N., Da Hora, A.S., Nogueira, J.S., Barros, I.N., Ayres, G.R., Richtzenhain, L.J., Brandao, P.E., 2010. Both the S1 and the S2 coding-regions of the spike glycoprotein gene of bovine coronavirus reveals a marked phylogeografic pattern for the distribution of the virus. XXI National Meeting of Virology, 2010, Gramado-RS. Virus Reviews and Research. XXI National Meeting of Virology. Rio de Janeiro, J. Braz. Soc.Virol. v.15, p.158.

Storz, J., Purdy, W., Lin, X., Burrell, M., Truax, R.E., Briggs, R.E., Frank, G.H., Loan, R.W., 2000. Isolation of respiratory bovine coronavirus, other cytocidal viruses, and *Pasteurella* spp. from cattle involved in two natural outbreaks of shipping fever. J. Am. Vet. Med. Assoc. 216, 1599–1604.

Strimmer, K., von Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. USA 94, 6815–6819.

Tsunemitsu, H., Smith, D.R., Saif, L.J., 1999. Experimental inoculation of adult dairy cows with bovine coronavirus and detection of coronavirus en feces by RT-PCR. Arch. Virol. 144, 167–175.

Vijgen, L., Keyaerts, E., Moes, E., Thoelen, I., Wollants, E., Lemey, P., Vandamme, M.A., Van Ranst, M., 2005. Complete genomic sequence of humancoronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J. Virol. 79, 1595–1604.

White, M.E., Schukken, H.Y., Tanksley, B., 1989. Space-time clustering of and risk factors for, farmer-diagnosed winter dysentery in dairy cattle. Can. Vet. J. 30, 948–951.

Yang, Z., 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Yoo, D., Deregt, D., 2001. A single amino acid change within antigenic domain II of the spike protein of bovine coronavirus confers resistance to virus neutralization. Clin. Diagn. Lab. Immunol. 8, 297–302.

Zhang, X., Hasoksuz, M., Spiro, D., Halpin, R., Wang, S., Vlasova, A., Janies, D., Jones, L.R., Ghedin, E., Saif, L.J., 2007. Quasispecies of bovine enteric and respiratory coronaviruses based on complete genome sequences and genetic changes after tissue culture adaptation. Virology 363, 1–10.

Zhang, X., Kousoulas, K.G., Storz, J., 1991. Comparison of the nucleotide and deduced amino acid sequences of the S genes specified by virulent strains of bovine coronaviruses. Virology 183, 397–404.