Arch Virol (2006) 151: 1735–1748
DOI 10.1007/s00705-006-0752-9

Archives of Virology
Printed in Austria

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

P. E. Brandão[1,3], **F. Gregori**[2,3], **L. J. Richtzenhain**[1], **C. A. R. Rosales**[1,3], **L. Y. B. Villarreal**[1,3], and **J. A. Jerez**[1,3]

[1]Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, Brazil
[2]Instituto Biológico de São Paulo, São Paulo, Brazil
[3]Coronavirus Research Group, São Paulo, Brazil

Received April 18, 2005; accepted February 27, 2006
Published online April 3, 2006

Summary. Bovine coronavirus (BCoV) causes enteric and respiratory disorders in calves and dysentery in cows. In this study, 51 stool samples of calves from 10 Brazilian dairy farms were analysed by an RT-PCR that amplifies a 488-bp fragment of the hypervariable region of the spike glycoprotein gene. Maximum parsimony genealogy with a heuristic algorithm using sequences from 15 field strains studied here and 10 sequences from GenBank and bredavirus as an outgroup virus showed the existence of two major clusters (1 and 2) in this viral species, the Brazilian strains segregating in both of them. The mean nucleotide identity between the 15 Brazilian strains was 98.34%, with a mean amino acid similarity of 98%. Strains from cluster 2 showed a deletion of 6 amino acids inside domain II of the spike protein that was also found in human coronavirus strain OC43, supporting the recent proposal of a zoonotic spillover of BCoV. These results contribute to the molecular characterization of BCoV, to the prediction of the efficiency of immunogens, and to the definition of molecular markers useful for epidemiologic surveys on coronavirus-caused diseases.

Introduction

Coronaviruses are classified in the order *Nidovirales*, family *Coronaviridae*, which comprises the genera *Coronavirus* and *Torovirus*. In this same order, one can also find the families *Arteriviridae* and *Roniviridae* [18, 54]. The

genus *Coronavirus* is subdivided into three groups (I, II, and III) according to epitopes of envelope glycoproteins, nucleotide sequences, and natural hosts [24].

Bovine coronavirus (BCoV) belongs to group II, with a diameter up to 220 nm. The BCoV genome is a non-segmented positive-sense single-stranded RNA of 32 kb that forms a helicoidal nucleocapsid in association with the nucleoprotein (N), a phosphoprotein of 50–60 kDa, rich in basic amino acids. The viral envelope consists of a lipid bilayer with four structural proteins (HE, S, E, and M) that make the crown-like appearance of the virions [24, 30].

In cattle, the most common BCoV-caused disease is neonatal calf diarrhea, which affects 3-to-4-week-old calves [44]. BCoV is also recognized as a causative agent of upper respiratory tract illness and bronchopneumonia in bovines [22, 23, 32, 51, 53]. Adult cows suffer from an enteric disease called winter dysentery, first described in the USA, also caused by BCoV strains found in calves [4, 8, 17].

The major envelope protein of BoCV is the spike (S) protein, formerly named E2, organized as trimers that appear as 20-nm-long projections in the viral envelope and harbor domains responsible for receptor binding, haemagglutination, and induction of neutralizing antibodies, and therefore is the most polymorphic among coronavirus species and also among strains of the same species [13]. The BCoV S is proreolytically cleaved into S1 and S2 subunits of 90 kDa each [11].

The carboxy-terminal S2 subunit contains the endodomain of S and forms the stalk of the spike, responsible for membrane fusion and syncytia formation [16, 25, 50, 52, 59]. The S1 subunit constitutes the amino-terminal ectodomain of S, which is much more variable than S2 and harbors the receptor-binding activity and forms the globular portion of the spike [30].

Due to its role in the formation of the globular portion of S and the fact that it harbors most of the antigenic sites of this protein, the S1 subunit is the most exposed to immunological selective pressures and thus most prone to polymorphism [1].

Since the spike glycoprotein is more sensitive to amino acid exchanges when compared to other coronavirus proteins, and the *S* gene has undergone more mutations in the past and has a greater potential for future mutations, studies focused on the S protein and S gene are appropriate for detecting intra-specific differences in the genus *Coronavirus* [14, 57].

Based on antigenic mapping with monoclonal antibodies, it is known, for instance, that an amino acid exchange in the antigenic domain II of the S protein may result in neutralization escape mutants [61]. Analysis of the S gene sequence is also useful for the discrimination among enteric coronaviruses detected in different individuals and for studies on the biological properties of the spike protein, e.g., infectivity for cell cultures [29, 38, 56, 60].

This study aimed to propose a genealogy for enteric strains of BCoV based on the hypervariable region of the gene coding for the S1 subunit of the S protein of Brazilian strains of BCoV and strains detected in other countries.

Materials and methods

Samples

Stool samples were collected between April 2000 and June 2002 from 51 calves from 10 dairy farms from 9 cities of São Paulo and Minas Gerais States, Southeastern Brazil (Table 1), from both diarrheic and non-diarrheic calves between 1 day and 6 months of age. Stool samples were prepared as 20% suspensions in PBS (PBS 0.01 M/BSA 0.1% pH 7.2) and clarified at 12,000 × g/30′ at 4 °C, and the supernatant was stored at −80 °C until analysis.

Bovine coronavirus reference strain

Bovine coronavirus Kakegawa strain [2], grown in the HmLu-1 (hamster lung) cell line, both provided by Dr. Takeo Sakai (Nihon University, Japan), was used as positive control in the RT-PCRs.

BCoV-specific reverse-transcription polymerase chain reaction (RT-PCR S1)

With Primer Premier 5.0 (©2003 Premier Biosoft International), two pairs of primers were designed, corresponding to conserved regions flanking the hypervariable region of the S1 gene, as described by Hasoksuz et al. [20], using BCoV S gene sequences (GenBank accession numbers AF058942.1, U06090.1, AF239306.1, M80844, U00735.2, M64667.1 and M64668.1) aligned by the CLUSTAL/W method with Bioedit v. 5.0.9 [19]. Outer primers: sense S1HS 5′-CTATACCCAATGGTAGGA-3′ and anti-sense S1HA 5′-CTGAAACACGACCGCTAT-3′, with a predicted product of 885 bp (nt 1204 to 2088 of the S gene). Inner primers: sense S1NS 5′-GTTTCTGTTAGCAGGTTTAA-3′ and anti-sense S1NA 5′-ATATTACACCTATC CCCTTG-3′, with a predicted fragment of 488 bp (nt 1329 to 1816 of S gene). Each primer was submitted to BLAST/n, and no non-BCoVS gene-related sequences were retrieved.

Reverse transcription (cDNA synthesis) was carried out at 42 °C for 60 min in a reaction mix with 1 × First Strand Buffer (Invitrogen™), 1 mM of each dNTP, 10 mM DTT, 1 μM of each primer (S1HS and S1HA), 7 μL of RNA extracted with TRIzol (Invitrogen™) (according to the manufacturer's instructions and denatured at 95 °C for 5 min) and 200 U of M-MLV Reverse Transcriptase (Invitrogen™) in a 20-μL final reaction volume.

Next, 5 μL of cDNA was added to the PCR mix with 1 × PCR Buffer (Invitrogen™), 0.2 mM of each dNTP, 0.5 μM of each primer (S1HS and S1HA), 1.5 mM $MgCl_2$, 25.25 μL of ultra-pure water, and 1.25 U Taq DNA polymerase (Invitrogen™) in a 50 μL final reaction volume and submitted to 35 cycles of 94 °C for 1 min, 53.4 °C for 1.5 min and 72 °C for 1 min, followed by 72 °C for 10 min for final extension.

The nested step was carried out with 5 μL of the first-round amplification added to a mix with 1 × PCR Buffer (Invitrogen™), 0.2 mM of each dNTP, 0.5 μM of each primer S1NS and S1NAS, 1.5 mM $MgCl_2$, 25.25 μL of ultra-pure water and 1.25 U Taq DNA polymerase (Invitrogen™) in a 50 μL final reaction volume and submitted to 25 cycles of 94 °C for 1 min, 58.4 °C for 1.5 min, and 72 °C for 1 min, followed by 72 °C for 10 min.

In each reaction, the Kakegawa strain was used as the positive control and PBS as negative control. In the nested PCR, a tube containing ultra-pure water instead of template was included between every three tubes to monitor amplicon contamination. Furthermore, in order to avoid any laboratory contamination, each step (RNA extraction, reverse transcription and PCR, nested PCR, and electrophoresis) was carried out in a separate room with separate materials.

The products of the nested PCR were resolved on a 1.5% agarose gel stained with 0.5 μg/mL ethidium bromide.

DNA sequencing

The 488-bp fragments obtained with RT-PCR S1 were purified from agarose gels using the Concert kit (Invitrogen™), quantified using the Low Mass DNA Ladder (Invitrogen™), and sequenced with BigDye 3.1 (Applied Byosystems™) according to manufacturer's instructions, without previous cloning in order to observe any signs of a *quasispecies* phenomenon in the chromatograms. The sequences were resolved in ABI-310 and ABI-377 automatic DNA sequencers (Applied Biosystems™).

Genealogic analysis

A genealogic tree was generated with the consensus sequences of each strain and 10 non-redundant homologous sequences retrieved from GenBank that were related to BCoV detected in calves from France, Canada, and the USA (Table 1), and bredavirus strain B145 as an outgroup (GenBank accession no. AJ575373.1).

Table 1. Bovine coronavirus strains included in the present study and corresponding GenBank accession numbers, geographical origin, detection year, literature reference, year of sequencing, and source of the sequenced strain with passage numbers (when available)

Strain	GenBank	Geographical origin	Detection year/ reference	Sequencing year	Source of the sequenced virus
BCoVENT	AF391541.1	USA	1997/[12]	2001	HRT-18G (up to 2 passages)
LY138	AF058942.1	USA	1965/[63]	2000	Calf stool
OK0514	AF058944.1	USA	1996/[46]	1998	HRT-18G (up to 5 passages)
BCQ1523	AF239307.1	Canada	1994/[27]	2000	HRT-18G (up to 5 passages)
BCQ20	U06092.1	Canada	1989/[34]	1994	HRT-18 (up to 5 passages)
BCQ9	U06091.1	Canada	1989/[34]	1994	HRT-18 (up to 5 passages)
Mebus	U00735.2	USA	1971/[33]	2003	BFK, MDBK
BCQ571	U06093.2	Canada	1989/[34]	2001	HRT-18 (up to 5 passages)
BCVF15	D000731.1	France	1979/[15]	1990	HRT-18
BCV Norden	M64668.1	USA	[63]	1991	Vaccine strain
USP01	AY255831	Brazil/MG*1‡	2001/This article	2003	Calf stool
USP02	AY606192	Brazil/MG1	2001/This article	2003	Calf stool
USP03	AY606193	Brazil/MG1	2001/This article	2003	Calf stool
USP04	AY606194	Brazil/MG1	2001/This article	2003	Calf stool
USP05	AY606195	Brazil/MG1	2001/This article	2003	Calf stool
USP06	AY606196	Brazil/SP**1	2002/This article	2003	Calf stool
USP07	AY606197	Brazil/SP2	2001/This article	2003	Calf stool
USP08	AY606198	Brazil/SP2	2001/This article	2003	Calf stool
USP09	AY606199	Brazil/SP1	2002/This article	2003	Calf stool
USP10	AY606200	Brazil/SP1	2002/This article	2003	Calf stool
USP11	AY606201	Brazil/SP3	2002/This article	2003	Calf stool
USP12	AY606202	Brazil/SP3	2002/This article	2003	Calf stool
USP13	AY606203	Brazil/SP3	2002/This article	2003	Calf stool
USP14	AY606204	Brazil/SP3	2002/This article	2003	Calf stool
LYVB	AY606205	Brazil/SP3	2002/This article	2003	Calf stool

*MG = Minas Gerais State

**SP = São Paulo State

‡Numbers after States represent Municipalities in each State

All sequences in their respective reading frames were aligned by the CLUSTAL/W method with Bioedit v. 5.0.9 [19] and used to generate the consensus rooted maximum parsimony tree with the tree-bisection-reconnection (TBR) branch-swapping heuristic algorithm with 1000 bootstrap replicates using PAUP 4.0 b10 (©2000 Smithsonian Institution), with the gaps considered as a fifth nucleotide.

Nucleotide identities and amino acid similarities of the translated sequences aligned with the BLOSUM62 matrix were calculated with Bioedit v. 5.0.9 [19].

Analysis of protein secondary structures

The secondary structure of the putative S1 hypervariable region was predicted with NNPredict at http://www.cmpharm.ucsf.edu/nomi/nnpredict.html.

Results

Seventeen out of the 51 stool samples were positive in the BCoV-specific RT-PCR targeting the S1 gene, and no spurious bands were found. PBS and nested internal controls demonstrated the specificity of the reactions and the absence of laboratory contamination.

Fifteen fragments out of the 17 samples produced by RT-PCR S1 resulted in BCoV-related sequences (Table 1). The two remaining fragments could not be sequenced due to low DNA concentrations. Alignment of each of these sequences with that described by Hasoksuz et al. [20] (accession number U00735.2) and BLAST/n analysis confirmed that they corresponded to the hypervariable region of the S1-encoding gene. Mean nucleotide identities to a stretch of 330 nucleotides with alignment to nucleotides 1381 to 1710 of the S gene of the Mebus strain (accession number U00735.2) are shown in Table 2.

The nucleotide alignment (Fig. 1) revealed a gap of 18 nucleotides (ATGC TGC(C/T) CAATGT(A/G)(A/G)TT), which corresponds to nucleotides 1577 to 1594 of the S gene. This gap begins at the second nucleotide of codon 526 (AAT) and finishes at the first nucleotide of codon 531 (TGT) of the S gene and was found in 14 out of the 15 sequenced field strains. Strain USP01, the only Brazilian one

Table 2. Mean, maximum, and minimum nucleotide identities to the alignment region of a 330-bp-long segment of the hypervariable region of the S1 subunit-coding region of 15 field strains of BCoV included in the present study, and 10 BCoV S gene sequences from Table 1, corresponding to nucleotides 1381 to 1710 of Mebus strain S gene (U00735.2)

	Brazil	USA	Canada	France	Japan
Brazil	98.34% (89.1–100%)	92.74% (90.3–97.8%)	91.59% (89.7%–96.3%)	93.32% (92.4–95.7%)	99.17% (90–100%)
USA	–	97.26% (96.3–99.3%)	97.09% (95.7–98.4%)	97.92% (96.9–98.7%)	92.64% (91.2–94.2%)
Canada	–	–	97.35% (96–98.7%)	97.12% (96.3–98.1%)	91.42% (90.6–92.1%)
France	–	–	–	100%	93.3%

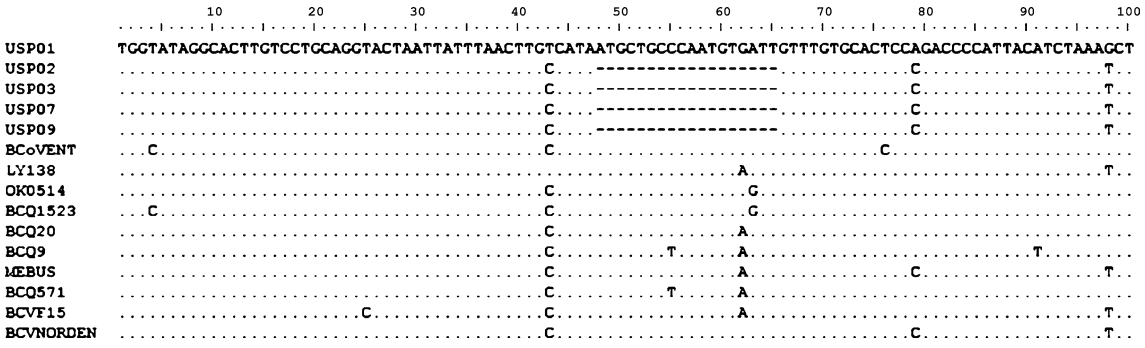

Fig. 1. Section of the alignment of 330 nt of the hypervariable region of the S1 subunit-coding region of the S gene of BCoV, corresponding to nucleotides 1381 to 1710 of the Mebus strain S gene (U00735.2). Strains USP01, -2, -3, -7, and -9 refer to BCoV field strains from the present study. Sequences for USP04, -05, -06, -08, -10 to -14, and strain LYVB were identical to USP03 and are therefore not included in this figure

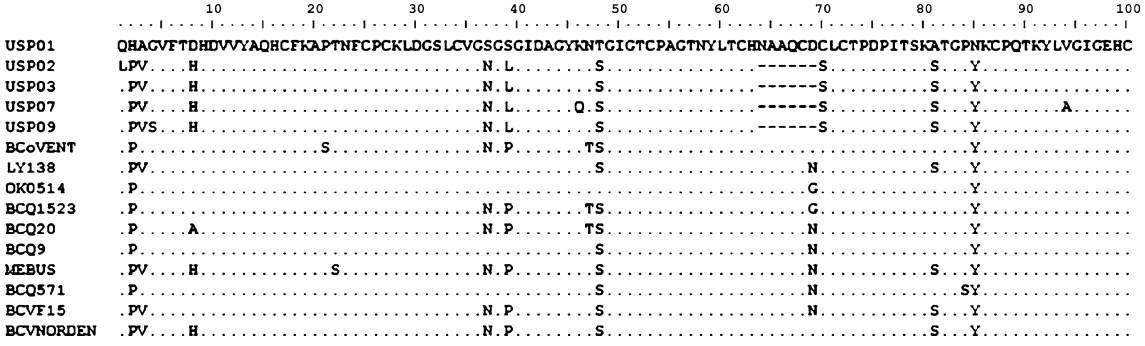

Fig. 2. Section of the alignment of the deduced 110 amino acids of the analysed hypervariable region of the S1 subunit of BCoV S protein corresponding to residues 461 to 570 of the Mebus strain S protein. Strains USP01, -2, -3, -7, and -9 refer to BCoV field strains from the present study. Sequences for USP04, -05, -06, -08, -10 to -14, and strain LYVB were identical to USP03 and are therefore not included in this figure

that lacks this gap, showed a nucleotide identity of 100% within the gap region with the sequences retrieved from the GenBank.

The alignment of the deduced amino acids, corresponding to residues 461 to 570 of the Mebus strain (accession number U00735.2), showed that this nucleotide deletion results in the loss of 6 amino acids (NAAQC(D/G/N) (Fig. 2), corresponding to residues 526 to 531 of the S protein. In additon, a C → S substitution was present in the amino acid position right after this gap in all of the 14 field strains with the deletion.

The mean amino acid homology among the 15 Brazilian field strains was 98%, ranging from 88% to 100%. Among the sequences from the USA, the mean amino acid homology was 97%, varying from 96 to 98%, while among the Canadian strains the mean amino acid homology was 96.67% and varied from 96% to 98%.

Fig. 3. Rooted consensus heuristic maximum parsimony tree for a stretch of 330 nucleotides of the gene coding for the hypervariable region of the S1 subunit of the S protein of BCoV, with bredavirus strain B145 as an outgroup, showing the two proposed clusters. Taxa in bold are related to the Brazilian field strains from the present study; numbers at each node are the bootstrap values obtained with 1000 replicates

Twenty-one out of the 37 nucleotide substitutions are exclusive to some strains, while the other 16 are at sites that vary in more than one strain.

The tree in Fig. 3 shows that all strains in which the 18-nucleotide gap was found grouped in an exclusive polytomic cluster, while the other strains clustered in a separate group with a resolved genealogy, giving rise to two major clusters among the studied strains. The two clusters of BCoV appear as paraphyletic groups, the gap evidenced as an unique evolutive event in the genealogy.

Analysis of the secondary structure prediction of the deduced amino acid sequences from the studied region of S1 of all Brazilian field strains and from strains Mebus, Norden, and BCQ-1523, chosen because they represent the polymorphism found in the last amino acid residue in the region that corresponds to the amino acid gap (Fig. 2), suggests that the gap occurs inside a loop region without helices or strands.

Discussion

A gap of 18 nucleotides, not reported for BoCV so far, was found between positions 1577 to 1594 of the gene coding for the spike protein of enteric strains of BCoV, resulting in the absence of amino acids 526 to 531 and the substitution of a cysteine in the position immediately after the gap by a serine (Fig. 2) within the ectodomain of the S protein. This gap was present in Brazilian field strains USP02 to USP14 and LYVB but not in strain USP01, whose sequence in this region is up to 100% identical to the sequences retrieved from the GenBank. So far, the gap appears to be present only in field strains circulating in Brazil. This conclusion is supported by a recent study on the molecular diversity of Korean BCoV field strains based on the hypervariable region of the S1 subunit. Jeong et al. [26] described that all analysed strains cluster together with strains OK0514 and LY138, while a different cluster containing the Mebus and BCVF15 strains emerged. None of the Korean field strain lacked the sequence absent in Brazilian field strains USP02 to USP14 and LYVB.

Observing these results under a parsimony evolutive model, we suggest that this gap is a deletion rather than an insertion, since fewer steps would be needed to create a deletion than to create a 18-nucleotide insertion in the other strains. Independent evolutionary events that lead to the same result are less probable, decreasing the number of extra-evolutionary steps, i.e., the number of homoplasies, which could lead to similarities in character status by, for instance, convergence, and not homology among the studied taxa, assuming that all BCoV strains share a common origin [35, 49].

In the tree shown in Fig. 3, the Brazilian isolates have a tendency to segregate into the "deleted" cluster 2, while the Brazilian field strain USP01 and the other, mainly cell-culture-adapted strains, segregate into the "non-deleted" cluster 1.

Interestingly, the same 18-nt deletion described for the Brazilian BCoV strains in this study was found in human coronavirus OC43 (HCoV-OC43), a group II coronavirus that plays a role in human colds. This deletion does not exist in other human strains, and deletions in the gene coding for the S1 subunit have never been reported in studies focused on genetic and antigenic properties and comparison between human and bovine coronaviruses [36, 45].

Thus, one can speculate that strains from both BCoV and HCoV-OC43 will segregate in a similar clustering pattern if this deletion is taken into account. This close evolutionary relationship between these two virus species is in agreement with the recently proposed zoonotic spillover of BCoV based on the high degree of identity between this virus and HCoV-OC43 [55].

Although the rooted tree in Fig. 3 does not allow the common ancestor to the BCoV strains studied herein to be identified, this role can be assigned to a non-deleted BCoV strain (cluster 1, Fig. 3) on the onset of the spillover event that might have originated both the human coronavirus strains with or without this deletion and the deleted BCoV strains.

The biological implications of amino acid deletions in the spike protein of coronaviruses might include a lower fusogenic activity [28], loss of the cleavage

site between subunits S1 and S2 [59], and changes in tissue tropism [31]. The 6-amino-acid deletion described here occurs inside a hypervariable region of the S1 subunit and is part of its domain II, responsible for the conformational epitopes A and B of this subunit and thus may result in the loss of immunological cross-reaction between the two clusters [61].

Although the amino acid deletion has not led to major alterations in the predicted secondary structures of the proteins, it is possible that the deleted loop may have caused a loss of conformational epitopes or the appearance of new ones by changes in the overall structure of the protein or by bringing existing epitopes together.

Furthermore, as the S1 ectodomain has a major role in receptor binding, mutations in this region may be an indication of a different virus-host interaction. For instance, for human coronavirus HCoV-229E, the domain comprised by amino acids 417 to 547 of the S protein – the same region where the deletion described here was found – has been shown to be essential for binding to the specific receptor, human amino peptidase [7]. The extent of deletions in the hypervariable region of the S1 subunit may also give raise to phenotypes with differences regarding receptor-binding activity, cleavage of the S protein, conformational changes in the S protein, tissue tropism, and disease patterns [62].

The ability to escape the host's immune system may also be a result of deletions in the epitopes of the S1 ectodomain, allowing the mutants to circumvent the action of cytotoxic T lymphocytes [5, 10, 40]. The occurrence of viral genomes with deletions in the S gene as, for instance, between nucleotides 1200 to 1800 of some isolates of MHV, which corresponds to the same region where the 18-nucleotide deletion has been detected in the present study, contributes to the *quasispecies* form of coronavirus populations [43].

The divergence among the strains sequenced in the present study and those from North America (Table 2) could be due to the geographic distance between the surveyed areas, different cattle breeds, or even the breeding system, which could exert selective pressure on the S1 hypervariable region during the time, which varied up to 38 years, as in the case of the sequence corresponding to strain LY138 (Table 1).

The mean nucleotide identities among strains from the USA and Canada (Table 2), geographically close countries, are similar, possibly due to the circulation of low-divergent BCoV strains. It is noteworthy that the expected high nucleotide identity to other regions of the BCoV S gene, such as S1B, with a mean of 97% [41] or the whole S gene, with 98% [58] to strains from Canada and the USA are close to those found here among sequences from these countries included in the analysis. Except for strain USP1, the results obtained in the present study uphold this phylogeographical pattern of BCoV strains, since cluster 2 (Fig. 3) contains strains from two geographically contiguous Brazilian States (Table 1).

Divergences within the S1 genes of members of the same species of coronavirus are not uncommon. For instance, among different samples of MHV (Murine Hepatitis Virus) coronavirus, the amino terminus of S1 has an amino acid similarity

ranging from 75 to 85% [48]. Furthermore, between some MHV and BCoV samples, the S1 genes have up to 81% nucleotide identity [6].

Nevertheless, strains from the USA and Canada, as well as strain BCV-F15 from France, were adapted to cell cultures, mainly in HRT-18 cells (Table 1), while strains sequenced in the present study have been obtained directly from fecal samples. This adaptation to cell culture may favor, by selection under similar conditions, a given S protein to prevail among other variants, biasing the study of the original sequences present in the original host [21]. This has already been reported for samples of canine coronavirus (CCoV) from fecal samples and CCoV reference strains grown in cell cultures, where the maximum nucleotide identity found for the S gene was 86.1% [38].

This hypothesis is in agreement with the episodic evolution model proposed for coronaviruses [3], according to which the molecular clock is accelerated during periods of environmental changes, such as adaptation to cell cultures, that are deleterious to the progenitor viruses, causing the viral population to evolve in short jumps in a short time interval towards a population that is divergent from the initial one. Populations of coronaviruses, an RNA virus with short replication times [47], large progeny size, a mutation rate close to 10^{-4}, and an RNA recombination rate of 20%, are prone to a high genetic variability when the target of the selection is not a single genotype but rather a heterogeneous population of mutants generated by erroneous replication of the most frequent mutant. This population of mutants is the basis of the *quasispecies* definition, the form that one expects to find in a population of coronavirus from a clinical sample [3, 37, 42].

Strain USP01, grouped in cluster 1, and strains USP02, USP03, USP04, USP05, USP11, USP12, USP13, USP14, and LYVB from cluster 2 were found in samples from calves without clinical information; strains USP07, USP08, USP06, USP09, and USP10 from cluster 2 were obtained from calves without diarrhea at the time of collection. Because of this lack of information, one can only hypothesize about pathogenicity or virulence variations among these 15 strains. Taking into account the position of the sequences retrieved from the GenBank in the genealogic tree (Fig. 3) – all of them isolates from animals with clinical diarrhea – both clusters might cause enteritis and diarrhea.

Of the 37 sites in the nucleotide alignment region where substitutions have been observed, 21 were exclusive to a given sequence, and the sequences from strains USP02, USP09, USP07, BCQ20, Mebus, and BCQ571 showed more non-synonymous than synonymous mutations. The other 16 sites in which nucleotide substitutions were found, 11 of which resulted in amino acid substitutions, are shared by two or more strains and are not single mutations, which might mean that these are consensus positions in the respective strains and not apomorphic conditions.

Thus, in the sequences in which the number of non-synonymous mutations exceeded that of synonymous mutations, taking into account only the point mutations exclusive to some of the strains and not those shared by two or more strains at variable sites, there is an indication of selective advantage at the time these mutations appeared in these sequences. This might suggest that under positive

selection the rate of fixation of non-synonymous mutations is higher than the rate of fixation of translationally silent nucleotide substitutions [9, 39].

It is expected that changes in the gene coding for the S protein, and mainly in the hypervariable region studied here, may be invaluable genetic markers for a more comprehensive understanding of BCoV-caused diseases and for the development of studies on diagnostics and molecular characterization, as well as for the prediction of the efficiency of immunogens. Comparing pathogenicity and virulence between these two clusters of BCoV, based, for instance, on fusogenic activity in cell cultures, is still a field of research, as well as investigations regarding other regions of the BCoV genome, such as the region encoding the S2 subunit, which plays a major role in membrane fusion.

In summary, a genealogy is proposed for enteric strains of bovine coronavirus based on the nucleotide sequences of the region coding for the hypervariable region of the S1 subunit of the spike protein, according to which two clusters (1 and 2) emerged with an 18-nt deletion shared with HCoV-OC43.

Acknowledgements

The authors are grateful to Mr. Alexandre Abelardo Sanches for his technical support and to FAPESP for the financial support (grant #00/00199-1).

References

1. Abraham S, Kienzle TE, Lapps W, Brian DA (1990) Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. Virology 176: 296–301
2. Akashi H, Inaba Y, Miura Y, Tokuhisha S, Sato K, Satoda K (1980) Properties of a coronavirus isolated from a cow with epizootic diarrhea. Vet Microbiol 5: 265–276
3. Baric RS, Yount B, Hensley L, Peel SA, Chen W (1997) Episodic evolution mediates interspecies transfer of a murine coronavirus. J Virol 71: 1946–1955
4. Benfield DA, Saif LJ (1990) Cell culture propagation of a bovine coronavirus isolated from cows with winter dysentery. J Clin Microbiol 28: 1454–1457
5. Bergmann CC, Yao Q, Lin M, Stohlman SA (1996) The JHM strain of mouse hepatitis virus induces a spike protein-specific Db-restricted cytotoxic T cell response. J Gen Virol 77: 315–325
6. Boireau P, Cruciere C, Laporte J (1990) Nucleotide sequence of the glycoprotein S gene of bovine enteric coronavirus and comparison with the S proteins of two mouse hepatitis virus strains. J Gen Virol 71: 487–492
7. Bonavia A, Zelus BD, Wentworth DE, Talbot PJ, Holmes KV (2003) Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HCoV-229E. J Virol 77: 2530–2538
8. Brandão PE, Birgel Jr EH, Gregori F, Rosales CAR, Ruiz VLA, Jerez JA (2002) Bovine coronavirus detection in adult cows in Brazil. Arq Inst Biol (São Paulo) 6: 103–104
9. Calcagnotto D (2001) Taxas de evolução e o relógio molecular. In: Matioli SR (ed) Biologia molecular e evolução. Holos, São Paulo, pp 51–63
10. Castro RF, Perlman S (1995) $CD8^+$ T cells epitopes within the surface glycoprotein of a neurotropic coronavirus and correlation with pathogenicity. J Virol 69: 8127–8131

11. Cavanagh D (1995) The coronavirus surface glycoprotein. In: Siddell SG (ed) The *Coronaviridae*. Plenum Press, New York, pp 73–113
12. Chouljenko VN, Lin XQ, Storz J, Kousoulas KG, Gorbalenya AE (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
13. Collins AR, Knobler RL, Powell H, Buchmeier MJ (1982) Monoclonal antibodies to murine hepatitis virus-4 (strain JHM) define the viral glycoprotein responsible for cell attachment and cell-cell fusion. Virology 119: 358–371
14. Compton SR, Vivas-Gonzalez BE, Macy JD (1999) Reverse transcriptase polymerase chain reaction-based diagnosis and molecular characterization of a new rat coronavirus strain. Lab Anim Sci 49: 506–513
15. Crucière C, Laporte J (1988) Sequence and analysis of bovine enteric coronavirus (F15) genome I. Sequence of the gene coding for the nucleocapsid protein; analysis of the predicted protein. Ann Inst Pasteur (Paris) 139: 123–138
16. Cyr-Coats KS, Storz J (1988) Bovine coronavirus-induced cytophatic expression and plaque formation: host cell and virus strain determine trypsin dependence. J Vet Med B 35: 48–56
17. Dea S, Michaud L, Milane G (1995) Comparison of bovine coronavirus isolates associated with neonatal calf diarrhea and winter dysentery in adult dairy cattle in Québec. J Gen Virol 76: 1263–1270
18. González JM, Gomez-Puertas P, Cavanagh D, Gorbalenya AE, Enjuanes L (2003) A comparative sequence analysis to revise the current taxonomy of the family *Coronaviridae*. Arch Virol 148: 2207–2235
19. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95–98
20. Hasoksuz M, Sreevatsan S, Cho KO, Hoet AE, Saif LJ (2002) Molecular analysis of the S1 subunit of the spike glycoprotein of respiratory and enteric bovine coronavirus isolates. Virus Res 84: 101–109
21. Hays JP, Myint SH (1998) PCR sequencing of the spike genes of geographically and chronologically distinct human coronaviruses 229E. J Virol Methods 75: 179–193
22. Heckert RA, Saif LJ, Hoblet KH, Agnes AG (1990) A longitudinal study of bovine coronavirus enteric and respiratory infections in dairy calves in two herds in Ohio. Vet Microbiol 22: 187–201
23. Heckert RA, Saif LJ, Myers GW, Agnes AG (1991) Epidemiologic factors and isotype-specific antibody responses in serum and mucosal secretions of dairy calves with bovine coronavirus respiratory tract and enteric tract infections. Am J Vet Res 52: 845–851
24. Holmes KV, Lai MMC (1996) *Coronaviridae*: the viruses and their replication. In: Fields BN, Knipe DM, Howley PM (eds) Virology. Lippincott-Raven Publishers, Philadelphia, pp 1075–1093
25. Honda E, Takahashi H, Okazaki K, Minetoma T, Kumagai T (1990) The multiplication of transmissible gastroenteritis viruses in several cell lines originated from porcine kidney and effects of trypsin on the growth of the viruses. Jap J Vet Sci 52: 217–224
26. Jeong JH, Kim GY, Yoon SS, Park SJ, Kim YJ, Sung CM, Shin SS, Lee BJ, Kang MI, Park NY, Koh HB, Cho KO (2005) Molecular analysis of S gene of spike glycoprotein of winter dysentery bovine coronavirus circulated in Korea during 2002–2003. Virus Res 108: 207–212
27. Kourtesis AB, Gelinas AM, Dea S (2001) Genomic and antigenic variations of the HE glycoprotein of bovine coronaviruses associated with neonatal calf diarrhea and winter dysentery. Arch Virol 146: 1219–1230

28. Krueger DK, Kelly SM, Lewicki DN, Ruffolo R, Gallagher TM (2001) Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. J Virol 75: 2792–2802
29. Kunita S, Zhang L, Homberger FR (1995) Molecular characterization of the S proteins of two enterotropic murine coronavirus strains. Virus Res 35: 277–289
30. Lai MCM, Cavanagh D (1997) The molecular biology of coronaviruses. Adv Virus Res 48: 1–100
31. Leparc-Goffart I, Hingley ST, Chua MM, Jiang X, Lavi E, Weiss SR (1997) Altered pathogenesis of a mutant of the murine coronavirus MHV-A59 is associated with a Q159L amino acid substitution in the spike protein. Virology 239: 1–10
32. Mcnulty MS, Bryson DG, Alan GM, Logan EF (1984) Coronavirus infection of the bovine respiratory tract. Vet Microbiol 9: 425–434
33. Mebus CA, White RG, Stair EL, Rhodes MB, Twiehaus MJ (1972) Neonatal calf diarrhea: results of a field trial using a reo-like virus vaccine. Vet Med Small Anim Clin 67: 173–178
34. Michaud L, Dea S (1993) Characterization of monoclonal antibodies to bovine enteric coronavirus and antigenic variability among Quebec isolates. Arch Virol 13: 455–465
35. Miyaki CY, Russo CAM, Pereira SL (2001) Reconstrução filogenética. Introdução e o método de máxima parcimônia. In: Matioli SR (ed) Biologia molecular e evolução. Holos, São Paulo, pp 97–107
36. Mounir S, Labonte P, Talbot PJ (1993) Characterization of the nonstructural and spike proteins of the human respiratory coronavirus OC43: comparison with bovine enteric coronavirus. Adv Exp Med Biol 342: 61–67
37. Moya A, Elena SF, Bracho A, Miralles R, Barrio E (2000) The evolution of RNA viruses: a population genetics view. Proc Natl Acad Sci USA 97: 6967–6973
38. Naylor MJ, Harrison GA, Monckton RP, Mcorist S, Lehrbach PR, Deane EM (2001) Identification of canine coronavirus strains from feces by S gene nested PCR and molecular characterization of a new Australian isolate. J Clin Microbiol 39: 1036–1041
39. Page RDM, Holmes EC (1998) Molecular evolution: a phylogenetic approach. Blackwell Science, Oxford
40. Pewe L, Wu GF, Barnett EM, Castro RF, Perlman S (1996) Cytotoxic T cell-resistant variants are selected in a virus-induced demyelinating disease. Immunity 5: 253–262
41. Rekik MR, Dea S (1994) Comparative sequence analysis of a polymorphic region of the spike glycoprotein S1 subunit of enteric bovine coronavirus isolates. Arch Virol 135: 319–331
42. Rottier PJM (1999) The molecular dynamics of feline coronaviruses. Vet Microbiol 69: 117–125
43. Rowe CL, Fleming JO, Nathan MJ, Sgro JY, Palmenberg AC, Baker SC (1997) Generation of coronavirus spike deletion variants by high-frequency recombination at regions of predicted RNA secondary structure. J Virol 71: 6183–6190
44. Snodgrass DR, Terzolo HR, Sherwood D, Campbell I, Menzies JD, Synge BA (1986) Etiology of diarrhea in young calves. Vet Rec 119: 31–34
45. St-Jean JR, Jacomy H, Desforges M, Vabret A, Freymuth F, Talbot PJ (2004) Human respiratory coronavirus OC43: genetic stability and neuroinvasion. J Virol 78: 8824–8834
46. Storz J, Stine L, Liem A, Anderson GA (1996) Coronavirus isolation from nasal swap samples in cattle with signs of respiratory tract disease after shipping. J Am Vet Med Assoc 208: 1452–1455
47. Sungwhan A, Maeda A, Makino S (1998) Coronavirus transcription early in infection. J Virol 72: 8517–8524
48. Suzuki H, Taguchi F (1996) Analysis of the receptor-binding site of murine coronavirus spike protein. J Virol 70: 2632–2636

49. Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (eds) Molecular systematics. Sinauer Associates, Sunderland, pp 407–514
50. Taguchi F (1995) The S2 subunit of the murine coronavirus spike protein is not involved in receptor binding. J Virol 69: 7260–7263
51. Tegtmeier C, Uttenthall A, Friis NS, Jensen NE, Jensen HE (1999) Pathological and microbiological studies on pneumonic lungs from Danish calves. J Vet Med B 46: 693–700
52. Toth TE (1982) Trypsin-enhanced replication of neonatal calf diarrhea coronavirus in bovine embryonic lung cells. Am J Vet Res 43: 967–972
53. Tsunemitsu H, Yonemichi H, Hirai T, Kudo T, Onoe S, Mori K, Shimizu M (1991) Isolation of bovine coronavirus from feces and nasal swabs of calves with diarrhea. J Vet Med Sci 53: 433–437
54. Van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, Maniloff J, Mayo MA, Mcgeoch DJ, Pringle CR, Wickner RB (2000) Virus taxonomy: the classification and nomenclature of viruses. The Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego
55. Vijgen L, Keyaerts E, Moës, E, Thoelen I, Wollants E, Lemey P, Vandamme A, Van Ranst M (2005) Complete genome sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J Virol 79: 1595–1604
56. Wesley RD (1999) The S gene of canine coronavirus strain UCD-1 is more closely related to the S gene of transmissible gastroenteritis virus than to that of feline infectious peritonitis virus. Virus Res 61: 145–152
57. Wu G, Yan S (2005) Reasoning of spike glycoproteins being more vulnerable to mutations among 158 coronavirus proteins from different species. J Mol Model (Online) 11: 8–16
58. Xuming Z, Kousoulas KG, Storz J (1991) Comparison of nucleotide and deduced amino acids sequences of the S genes specified by virulent and avirulent strains of bovine coronaviruses. Virology 183: 397–404
59. Yamada YK, Takimoto K, Yabe M, Taguchi F (1997) Acquired fusion activity of a murine coronavirus MHV-2 variant with mutations in the proteolytic cleavage site and the signal sequence of the S protein. Virology 227: 215–219
60. Yoo D, Pei Y, Christie N, Cooper M (2000) Primary structure of the sialodacryoadenitis virus genome: sequence of the structural-protein region and its application for differential diagnosis. Clin Diagn Lab Immunol 7: 568–573
61. 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
62. Zelus BD, Schickli JH, Blau DM, Weiss SR, Holmes KV (2003) Conformational changes in the spike glycoprotein of murine coronavirus are induced at 37 °C either by soluble murine CECAM1 receptors or by pH 8. J Virol 77: 830–840
63. Zhang XM, Kousoulas KG, Storz J (1991) The hemagglutinin/esterase glycoprotein of bovine coronaviruses: sequence and functional comparisons between virulent and avirulent strains. Virology 185: 847–852

Author's address: Paulo Eduardo Brandão, Departamento de Medicina Veterinária Preventiva e Saúde, Animal, Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, Avenida Professor Doutor Orlando Marques de Paiva, 87, Cidade Universitária, CEP 05508-900, São Paulo, SP, Brazil; e-mail: paulo7926@yahoo.com