Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Full-length genome analysis of canine coronavirus type I

Nicola Decaro *, Viviana Mari, Gabriella Elia, Gianvito Lanave, Giulia Dowgier, Maria Loredana Colaianni, Vito Martella, Canio Buonavoglia

Department of Veterinary Medicine, University of Bari, Valenzano, Italy

A R T I C L E   I N F O

Article history:
Received 12 May 2015
Received in revised form 9 July 2015
Accepted 20 July 2015
Available online 26 July 2015

Keywords:
Dog
Canine coronavirus type I
Genomic analysis

A B S T R A C T

Canine coronavirus types I (CCoV-I) and II (CCoV-II) are usually responsible for mild enteritis in dogs. While the CCoV-II genome has been completely sequenced, to date there are no complete genomic sequence data available publicly for CCoV-I. Thus, the aim of the present study was to analyze the full-length genome of a CCoV-I prototype strain that had been recovered from a dog with diarrhea in Italy. CCoV-I strain 23/03 has a genome of 30,000 nucleotides, excluding the 3′ poly(A) tail, displaying the typical Alphacoronavirus-1 organization and the highest genetic relatedness to CCoV-II. However, two distinct features were observed in the CCoV-I genome: (i) the presence of an additional ORF between the spike (S) protein gene and ORF3a; (ii) the diversity of the S protein, which is more closely related to that of feline coronavirus type I and presents a furin cleavage site. The present study may contribute to a better understanding of the Alphacoronavirus-1 evolutionary pattern and may be paradigmatic of how coronaviruses evolve through gene losses, acquisition and exchanges among different members.

1. Introduction

Coronaviruses (CoVs) are large, single-stranded, positive-sense RNA viruses, which are responsible for enteric and/or respiratory disease in mammals and birds. Canine coronavirus (CCoV) is usually responsible for mild enteritis in young dogs (Decaro and Buonavoglia, 2008, 2011), although fatal disease has been associated to a pantropic variant of the virus (Decaro et al., 2008, 2010a, 2012; Marinaro et al., 2010; Zicola et al., 2012; Ntafis et al., 2012). Based on the genetic distance encountered in the spike (S) protein gene (Pratelli et al., 2003), two CCoV genotypes are known, CCoV type I (CCoV-I) and type II (CCoV-II), which are variously distributed worldwide (Decaro et al., 2005, 2011, 2013; McElligott et al., 2011; Ntafis et al., 2013; Licitra et al., 2014; Cavalli et al., 2014; Costa et al., 2014). CCoV-II has been found to exist in two subtypes, CCoV-IIa and CCoV-IIb, the latter being the result of homologous recombination with transmissible gastroenteritis virus of swine (TGEV) (Decaro et al., 2009, 2010b). Intermediate viruses between CCoV-I and CCoV-II have also been detected (Town and Whittaker, 2012).

CCoV-I and CCoV-II form a unique viral species, Alphacoronavirus-1 (family Coronaviridae, genus Alphacoronavirus), along with feline coronavirus types I (FCoV-I) and II (FCoV-II), TGEV and porcine respiratory coronavirus (PRCoV) (Decaro and Buonavoglia, 2011). An additional ORF, named ORF3a, was found in the CCoV-I genome, whereas only its remnants were evident in the genomes of CCoV-II and TGEV, revealing an intriguing evolutionary history within the Alphacoronavirus-1 species (Lorusso et al., 2008).

While the full-length genomes of several strains of CCoV-II have been determined (Decaro et al., 2015), to date there are no complete genomic sequence data available publicly for CCoV-I. Thus, the aim of the present study was to analyze the full-length genome of a CCoV-I prototype strain that had been recovered from a dog with diarrhea in Italy.

2. Materials and methods

2.1. Virus origin

Strain 23/03 was detected during an epidemiological survey for CCoV in Italian dogs with diarrhea (Pratelli et al., 2004). The ill dog, a male German shepherd of 6 weeks of age, belonged to a kennel located in the Apulia region, southern Italy. The feces were collected by a vet directly from the rectal ampulla into a sterile container during the clinical examination of the dog. CCoV-I RNA detection in the specimen was obtained by means of genotype-specific PCR (Pratelli et al., 2004) and real-time RT-PCR (Decaro et al., 2005). Virus isolation attempts using different cell lines of canine and feline origin were unsuccessful, since CCoV-I has not been adapted to the in vitro
growth (Decaro and Buonavoglia, 2008, 2011). The original fecal sample was aliquoted and stored at −70 °C until RNA extraction.

2.2. RNA extraction

An aliquot of the original fecal specimen was clarified by centrifuging at 2500 × g for 10 min. One-hundred-forty microliters of the supernatant were then used for RNA extraction by means of QiAamp® Viral RNA Mini Kit (Qiagen S.p.A., Milan, Italy), following the manufacturer’s protocol and the RNA template was stored at −70 °C until its use.

2.3. CCoV detection, quantification and characterization

The RNA extract was subjected to a previously-established TaqMan®-based real-time RT-PCR assay for rapid detection and quantification of CCoV RNA (Decaro et al., 2004), with minor modifications. Briefly, a one-step method was adopted using SuperScript® III Platinum® One-Step qRT-PCR Kit (Life Technologies srl, Milan, Italy) and the following 50–µl mixture: 25 µl of master mix, 1 µl of SuperScript® III RT/Platinum Taq Mix, 300 nM of primers CCoV-For and CCoV-Rev, 200 nM of probe CCoV-Pb (Decaro et al., 2004) and 10 µl of template RNA. Duplicates of log_{10} dilutions of standard RNA were analyzed simultaneously in order to obtain a standard curve for absolute quantification. The thermal profile consisted of reverse transcription at 50 °C for 15 min and activation of Platinum Taq DNA polymerase at 95 °C for 2 min, followed by 45 cycles of denaturation at 95 °C for 15 s, annealing at 48 °C for 30 s and extension at 60 °C for 30 s.

CCoV genotyping was achieved by means of two distinct genotype-specific assays (Decaro et al., 2005) performed by using SuperScript® III Platinum® One-Step qRT-PCR Kit (Life Technologies srl) and the following oligonucleotide sets (final concentrations were 600 and 200 nM for primers and probes, respectively): primer pair CCoV-F/CCoV-R and probe CCoV-Pb for CCoV-I and CCoV-II/FCoV-I and CCoV-II/FCoV-II and probe CCoVIII-Pb (Decaro et al., 2005) for CCoV-II. The thermal protocol was as described for CCoV detection except for different annealing temperatures (53 °C and 48 °C for CCoV-I and CCoV-II, respectively).

2.4. RT-PCR amplifications

Overlapping fragments of the genome of CCoV-I strain 23/03 were obtained through RT-PCR reaction carried out using primer sets designed based on the genome sequence of other alphacoronaviruses and the kit SuperScript™ One-Step RT-PCR Kit for Long Templates (Life Technologies srl). Additional RT-PCR assays and subsequent sequencing attempts were performed to close gaps between assembled contigs and to sequence unresolved genomic regions using primers designed on the alignment of the reference Alphacoronavirus strains. The very 5’ and 3’ ends were amplified using 5’ and 3’ RACE System for Rapid Amplification of cDNA Ends (Invitrogen), respectively, following the manufacturer’s instructions. The PCR products were detected by electrophoresis through a 1.5% agarose gel and visualization under UV light after ethidium bromide staining.

2.5. Sequence analysis and phylogeny

RT-PCR products were subjected to direct sequencing at the BaseClear B.V. (Leiden, The Netherlands). The sequences were manually edited and analyzed using the Geneious platform (http://www.geneious.com) and the NCBI’s (http://www.ncbi.nlm.nih.gov) and EMBL’s (http://www.ebi.ac.uk) analysis tools. Nucleotide (nt) sequences of the different ORFs were converted into amino acid (aa) sequences and comparative sequence analysis with reference coronavirus sequences was carried out in the full-length genome and encoded structural and nonstructural proteins.

Phylogenetic and molecular evolutionary analyses were conducted using Mega4.1 Beta (Tamura et al., 2007). In order to include in the analysis CCoV-IIb, whose genome has not been completely sequenced, pylogenetic trees were elaborated on a 22,366 genomic sequence (encompassing the 3’ end of ORF 1a to the 3’ UTR) and on the amino acid (aa) sequences of S, membrane (M), and nucleocapsid (N) proteins using both parsimony and neighbor-joining methods, supporting a statistical significance with bootstrapping over 1000 replicates. The following Alphacoronavirus reference strains were used for phylogeny (GenBank accession numbers are indicated in parentheses): CCoV-IIa 1/71 (JQ404409), K378 (KC175340), S378 (KC175341), TN449 (JQ404410), NTVU366/F/2008 (GQ477367), CB/05 (KP981644); CCoV-IIb 174/06 (EU856362), 341/05 (EU856361); CCoV A76 (JN856008); FCoV-I Black (EU186072); FCoV-II 79-1146 (DQ010921), 79-1683 (JN634064); TGEV Purdue (DQ811789); PRCoV ISU-1 (DQ811787). The distantly-related Betacoronavirus-1 canine respiratory coronavirus (CRCoV) K37 (JX860640) was used as outgroup.

2.6. Nucleotide sequence accession number

The full-length genome of CCoV-I strain 23/03 was deposited in GenBank under accession number KP849472.

3. Results

3.1. Detection of CCoV-I

By the real-time RT-PCR panels, the fecal sample was confirmed to contain a CCoV-I strain, whose titer was calculated as 6.73 × 10^6 RNA copies/µl of template. The specimen had no traces of CCoV-II RNA.

3.2. CCoV-I genomic organization

The genome of CCoV-I strain 23/03 has a size of 30,000 nt, excluding the 3’ poly(A) tail, and shows typical Alphacoronavirus-I organization (Table 1 and Fig. 1). The 5′ UTR consists of 313 nt including the leader sequence (L, nt 1–94) and the conserved core 5′-CUAAC-3′ (nt 95–100) of the transcription regulatory sequence (TRS), which controls the mRNA synthesis through interaction with the viral polymerase during the discontinuous transcription of the negative strand subgenomic RNA of the Nidovirales members (Enjuanes et al., 2000). Similar TRS signals precede each of the 8 putative mRNA encoding for the structural and nonstructural proteins (Table 1). The 3′ end of the viral genome consists of a 274-nt 3′ UTR that is followed by the poly(A) tail. Sequence analysis showed intact structural and non-structural proteins with respect to reference CCoV-II, FCoV-I and FCoV-II genomes. About two-thirds of the viral genome is occupied by the replicase gene encoding for two large polyproteins (pp), pp1a and pp1ab, the latter being synthesized through ribosomal slippage at position 12,327. The polyproteins of the replicase complex are processed by viral proteases, resulting in several products with different size and function. Sequence comparison with other Alphacoronavirus-I genomes led to the detection of three putative papain-like protease cleavage sites and 11 putative 3C-like protease cleavage sites, producing 16 nonstructural proteins (Table 2).

Four structural proteins were detected downstream of the replicase gene, namely the spike (S), small envelope (E), membrane (M) and nucleocapsid (N) proteins. The S protein has a size of 1481 aa, thus being longer than the analogous protein of
other Alphacoronavirus-I members (1451–1457 aa in CCoV-II and FCoV-II, 1457–1464 aa in FCoV-I, 1447–1449 aa in TGEV, 1225 aa in PRCoV). By using the NetNGlyc server (http://www.cbs.dtu.dk/services/NetNGlyc/), 28 N-glycosylation sites were predicted in the CCoV-I 23/03 S protein, whereas 30–33 N-glycosylated Asn residues had been detected in CCoV-II (Sanchez-Morgado et al., 2004; Decaro et al., 2007). At position 802–806, the S protein exhibits a potential cleavage site, represented by the basic aa stretch Arg-Arg-Val-Arg-Arg (RRVRR). This stretch had been also observed in the sequence of the S protein of Elmo/02 (position 801–805), but in this strain an Ala residue has replaced Val at position 803 (Pratelli et al., 2003). With few exceptions (de Haan et al., 2008), other alphacoronaviruses do not share this finding.

The E protein is 82-aa long and does not present any N-glycosylation sites, whereas three N-glycosylated residues have been detected in the 264-aa long M protein, which is in agreement with what has been observed in other FCoV/CCoV strains, with the exception of CCoV-II isolate BGF10 that shows only two

---

**Table 1**

Coding potential, putative transcription regulatory sequences and encoded proteins of the CCoV-I 23/03 genome.

| Putative gene | Coding sequence (nt) | Putative TRS | TRS sequence$^a$ | Putative protein |
|---------------|----------------------|--------------|-----------------|-----------------|
| Gene name     | Start nt position    |              |                 | Protein name    | Protein size (aa) |
| ORF1ab        | 314–12,327           | 90           | UCGAAACUAACGAAAU | Pplab           | 6681             |
| ORF1a         | 314–12,358           |              |                 | Pp1a            | 4014             |
| ORF2          | 29,359–24,804        | 20,319       | AUUACUAACUUUGG  | S               | 1481             |
| ORF3          | 24,812–25,435        | 24,797       | UUCAUAAACUCAA   | 3               | 207              |
| ORF3a         | 25,447–25,683        |              |                 | 3a              | 78               |
| ORF3b         | 25,628–25,843        | 25,434       | AGAACUAACAAAUG  | 3b              | 71               |
| ORF3c         | 25,840–26,595        | 26,514       | GGUUCUAAACGAAAU | 3c              | 251              |
| ORF4          | 26,561–26,809        | 26,807       | UGAACUAACAAAUA  | M               | 264              |
| ORF5          | 26,820–27,614        | 27,611       | AUAAACUAACUUCA  | N               | 380              |
| ORF6          | 27,627–28,769        |              |                 | 7a              | 101              |
| ORF7a         | 28,774–29,079        | 28,762       | CGAACUAACGAAGU  | 7b              | 213              |
| ORF7b         | 29,084–29,725        |              |                 |                 |                  |

$^a$ The conserved TRS core is underlined.

---

**Fig. 1.** Schematic representation of the genomes of Alphacoronavirus-I members depicting the genetic differences among the CCoV genotypes. Genes encoding for structural and non-structural proteins are shown in gray and white, respectively. ORF sizes are not drawn to scale. The arrows indicate the transcription regulating sequences preceding each CoV gene.
glycosylated A75 residues. The N protein of strain CCoV-I 23/03 is 380-aa long product with three potential N-glycosylation sites.

Analogously to CCoV-II and FCoV-III, some accessory genes were detected between ORFs 2 (S-protein gene) and 4 (E-protein gene) and downstream of ORF6 (N-protein gene). The S–E intergenic region contains the canonical three ORFs 3a, 3b and 3c, encoding for products with sizes of 78, 71 and 251 aa, respectively, plus an additional accessory protein gene, ORF3, encoding for a putative 206 aa protein, which has been found to be unique to the CCoV-I genome (Lorusso et al., 2008). The 3’ end accessory genes were ORFs 7a and 7b that encodes for 101-aa and 213-aa long proteins, respectively.

3.3. Sequence analysis

Alignment of complete genome sequences of CCoV-I strain 23/03 and reference alphacoronaviruses showed the closest genetic relatedness with CCoV-II isolates (83.82–84.98% nt identity), followed by TGEV (82.81%) and FCoV (77.19–77.45%). No comparison was possible with CCoV-III since there are no full-length genomes available in the GenBank database for this virus. When the spike protein was analyzed, CCoV-II displayed a higher aa identity to FCoV-I (73.09%) than to CCoV-II/IIb (42.74–43.71%), FCoV-II (43.05%) and TGEV (42.83%). Among extant CCoV strains, the closest identity was observed with isolate A76, which has been proven to have a CCoV-II/III recombinant S protein (Regan et al., 2012). The E, M and N proteins of strain 23/03 were all more closely related to the analogous products of CCoV-II and porcine CoV reference strains (Table 3).

3.4. Phylogeny

In order to include in the analysis the nt sequences available for CCoV-IIb, phylogeny was first constructed on a 23,106-nt fragment spanning from the 3’ end of ORF1a to the very 3’ end of the viral genomes. In the neighbor-joining tree elaborated using these sequences, strain CCoV-I 23/03 clustered separately from extant CCoV/FCoV/I isolates leading to the formation of an outlier branch (Fig. 2A). The prototype CCoV-I strain formed a separate branch from other analyzed alphacoronaviruses also in the trees elaborated using the M (Fig. 2C) and N (Fig. 2D) proteins, whereas the S protein revealed a clustering with FCoV-I (Fig. 2D). The same tree topologies were obtained through phylogenetic analysis using the maximum-parsimony method (data not shown).

4. Discussion

CCoV has progressively emerged as being responsible for moderate to severe enteritis in dogs, with different genotypes and subgenotypes being detected in recent years (Decaro and Buonavoglia, 2008, 2011). CCoV-II is the oldest genotype, which has been known since 1971 (Binn et al., 1974), whereas CCoV-I was discovered only 30 years later thanks to molecular methods, since this virus has not been adapted to grow in vitro (Pratelli et al., 2003). More recently, two CCoV-II subtypes have been recognized, namely CCoV-IIa and CCoV-IIb, on the basis of the relatedness of the spike protein of the latter virus to that of TGEV (Decaro et al., 2009, 2010b). In addition, a virulent CCoV-IIa biotype, strain CB/05, has been proven to cause fatal infections and long-lasting lymphopenia in naturally (Buonavoglia et al., 2006; Decaro et al., 2012; Ntafis et al., 2012; Zicola et al., 2012; Pinto et al., 2014) and experimentally (Decaro et al., 2008, 2011; Marinario et al., 2010) infected dogs.

To date, while the CCoV-II genome has been fully determined (Decaro et al., 2015), only fragments of the 3’ genomic end are available for CCoV-I. Therefore, in the present study we have carried out the complete genome sequence analysis for this canine pathogen. The results showed that the CCoV-I genome displays some distinct features with respect to CCoV-II and other alphacoronaviruses. The first finding is the presence of an additional accessory protein gene, ORF3, which was located between the S protein gene and ORF3a. This additional gene, which has been recently characterized, was found to be unique to the CCoV-I genome, whereas CCoV-II and TGEV exhibit only 5’ and 3’ end remnants (Lorusso et al., 2008). However, FCoV strains harboring different forms of

---

**Table 3**

Percent (%) identities of CCoV-I 23/03 to *Alphacoronavirus-I* reference strains in the complete genomic sequence (nucleotide, nt) and structural proteins (amino acid, aa).

| Alphacoronavirus-I | Strain | Accession number | Identity (%) to CCoV-I 23/03 |
|-------------------|--------|------------------|-----------------------------|
|                   |        | Full-length genome (nt) | S (aa) | E (aa) | M (aa) | N (aa) |
| CCoV-IIa          | 1/71   | JQ404409          | 84.96 | 43.52 | 87.50 | 89.01 |
|                   | K178   | KC175340          | 84.82 | 43.52 | 87.80 | 88.74 |
|                   | S378   | KC175341          | 84.79 | 43.52 | 86.59 | 88.74 |
|                   | TN449  | JQ404410          | 83.82 | 43.32 | 86.59 | 89.01 |
|                   | NTO396/F/2008 | GQ477367       | 84.25 | 42.74 | 86.59 | 88.74 |
|                   | 1/174  | EU856362          | NA    | 43.34 | 85.15 | 88.48 |
|                   | 341/05 | EU856361          | NA    | 43.13 | 86.59 | 88.74 |
| CCoV              | A76    | JN856008          | 84.62 | 49.18 | 85.37 | 87.50 |
| FCoV              | Black  | EU186072          | 77.39 | 73.09 | 82.93 | 75.33 |
|                   | 79-1146| DG010921          | 77.43 | 43.05 | 82.93 | 84.91 |
|                   | 79-1683| JN634064          | 77.19 | 43.05 | 86.76 | 77.45 |
| TGEV              | Purdue | DQ811789          | 82.81 | 42.83 | 84.15 | 87.12 |
| PRCoV             | ISU-1  | DQ811787          | 80.19 | 40.83 | 84.15 | 88.26 |

NA, sequence not available.
ORF3 and a CCoV-I N protein gene have been detected, thus proving the circulation of FCoV-I/CCoV-I recombinant viruses (Le Poder et al., 2013). Another finding of the CCoV-I genome is that the S protein gene is closely related to that of FCoV-I, whereas the rest of the genome displays a higher relatedness to CCoV-II. In addition, a furin cleavage site leads to the potential generation of two subunits, S1 and S2. A similar basic motif is present, approximately in the same position, in most beta- and gammacoronaviruses. Cleavage of the CoV S protein has been correlated to cell-fusion activity in vitro but the potential implications in viral pathobiology have not been fully determined (Hingley et al., 1998). In addition to CCoV-I, furin cleavage motifs have been detected in FCoV-I, but even in this case the biological consequences were not completely understood (de Haan et al., 2008). Unlike CCoV-II and FCoV-II that display a similar S protein as a consequence of homologous recombination, CCoV-I does not grow in cell cultures and only few FCoV-I strains have been adapted to grow in vitro. The CoV S protein interacts with cell receptors, thus being responsible for binding of virions to the cell surface (Enjuanes et al., 2000). Thus, the different S proteins between CCoV-I/FCoV-I and CCoV-II/FCoV-II are likely to be responsible for the diverse biological behaviors in cell cultures. The cell receptor for most Alphacoronavirus-1 isolates is the cell surface glycoprotein aminopeptidase N (APN), but there is no evidence for this receptor being used by CCoV-I/FCoV-I (de Haan et al., 2008).

On the basis of the most recent findings, the evolutionary history of Alphacoronavirus-1 members has been tentatively reconstructed, suggesting that CCoV-I and FCoV-I are the ancestral viruses from which TGEV, CCoV-II and FCoV-II have originated through gene losses and recombination events (Lorusso et al., 2008). Our findings corroborate the hypothesis that CCoV-I is the ancestor for CCoV-II, since these viruses exhibit high identity in the entire genome with the exception of the S protein gene, which is markedly
diverse, and ORF3, of which only remnants are present in the CoCV-II genome.

In conclusion, the full-length sequencing of the CoCV-I genome may contribute to a better understanding of the Alphacoronavirus-I evolutionary pattern and may be paradigmatic of how CoVs evolve through gene losses, acquisition and exchanges among different members.

Acknowledgements

The authors are grateful to Carlo Armenise, Arturo Gentile and Donato Narcisi for their technical assistance.

References

...