Characterization of novel canine bocaviruses and their association with respiratory disease

Amit Kapoor,1 Natasha Mehta,1 Edward J. Dubovi,2 Peter Simmonds,3 Lakshmanan Govindasamy,4 Jan L. Medina,1 Craig Street,1 Shelly Shields5 and W. Ian Lipkin1

1Center for Infection and Immunity, Columbia University, New York 10032, USA
2College of Veterinary Medicine at Cornell, Ithaca, NY 14853, USA
3Centre for Immunology, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3JT, UK
4Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, FL 32610, USA
5Pfizer Veterinary Medicine Research and Development, New York 10017, USA

We report the first identification, genetic characterization and disease association studies of several novel species of canine bocaviruses (CBoV). Evolutionary analysis confirmed that CBoV are genetically distinct from the only other known canine bocavirus, minute virus of canines, with which they share less than 63, 62 and 64 % protein identity in NS, NP and VP genes, respectively. Comparative genetic analysis of 37 VP gene variants found in diseased and healthy animals showed that these novel viruses are genetically highly diverse and are common in canine respiratory infections that have remained undetected until now. Interestingly, we observed that a CBoV genotype with a unique deletion in the VP2 gene was significantly more prevalent in animals with respiratory diseases compared with healthy animals.

Paroviruses, which frequently infect animals through the faecal–oral route, are small, non-enveloped icosahedral viruses with linear ssDNA genomes (Fauquet et al., 2004). They are members of the family Paroviridae, which comprises two subfamilies, Densovirinae and Parovirinae, members of which infect non-vertebrate and vertebrate hosts, respectively (Brown, 2010; Fauquet et al., 2004). The International Committee on Taxonomy of Viruses (ICTV) has further classified the subfamily Parovirinae into six genera: Dependovirus, Bocavirus, Erythrovirus, Parovirus, Amphovirus and Partetravirus. Bocaviruses are unique among paroviruses as they contain a third ORF between the non-structural- and structural-coding regions (Kapoor et al., 2010b; Manteufel & Truyen, 2008; Qiu et al., 2007). The genus Bocavirus currently includes the bovine paroviruses (BPV), minute virus of canines (MVC) (Fauquet et al., 2004), porcine bocaviruses (Cheng et al., 2010), gorilla bocavirus (GBoV) (Kapoor et al., 2010a) and four species of human bocaviruses (HBoV 1–4) (Allander et al., 2005; Arthur et al., 2009; Chienchansin et al., 2007; Kapoor et al., 2009, 2010b). Bocaviruses commonly infect the respiratory and gastrointestinal tract of young animals and humans, and except for BPV, the pathological manifestations of these infections remain largely unknown (Don et al., 2011; Kapoor et al., 2011; Manteufel & Truyen, 2008; Martin et al., 2009, 2010).

First discovered in 1967 in faeces of healthy dogs, MVC is the only known bocavirus that infects dogs. It can cause abortions in bitches and severe respiratory infections in newborn puppies, but infections are mostly subclinical in adult animals (Carmichael et al., 1991). MVC replicates to high titres in Walter Reed/3873D (WRD) canine cells, making it a useful model system to dissect the replication kinetics of genetically similar, but uncultivable, HBoV (Sun et al., 2009). During a metagenomic study conducted to better characterize the respiratory viral flora of domestic animals, we observed several sequences with distant amino acid sequence similarity to animal paroviruses in respiratory samples from diseased dogs. Extension of these novel sequences using a primer walking approach revealed the presence of a novel bocavirus tentatively named canine bocavirus (CBoV). Thereafter, a consensus PCR assay using primers targeting conserved structural protein motifs was used to determine the prevalence and genetic diversity of CBoV variants in a cohort of respiratory samples obtained from diseased and healthy dogs (Supplementary Table S1, available in JGV Online). Briefly, extracted nucleic acids from each sample were used for two rounds of nested PCR.
The nearly complete genome of CBoV variant con-161 is 5413 nt (GenBank accession no. JN648103) and bears a high degree of similarity to other known bocaviruses predicted to contain non-coding terminal sequences flanking the three large ORFs (Fig. 1). ORF1 encodes a 648 aa non-structural (NS) protein. ORF2 encodes 712 aa overlapping the VP1/VP2 capsid ORFs. ORF3 encodes a 195 aa NP1 protein. The non-coding region on left-hand side (LHS) terminus, located at the 5' end of positive-sense ssDNA genomes or at the amino terminus of NS protein, is 306 nt. Its secondary structure folds into an imperfect palindrome and contains a rabbit-ear structure similar to MVC and BPV (Fig. 1). The right-hand side (RHS) non-coding region, found at the 3' end of positive-sense ssDNA genomes, is not identical to the LHS terminus and forms an imperfect palindrome. The LHS terminus showed highest sequence similarity to the MVC terminus and also contained a NS-binding site (Sun et al., 2009). While the MWC and BPV NS protein-coding regions encode a single long NS protein, recent studies have shown that the homologous region of all four HBoV species encodes two NS proteins of variable length (Chen et al., 2010; Kapoor et al., 2010a). Remarkably, although CBoV is genetically similar to HBoV in NS gene splicing, it is most similar to MVC. The CBoV NS-coding region encodes a shorter NS protein, as well as conserved RNA splicing signals essential to generate a longer NS protein (Fig. 1). Like other bioinformatics analyses and predictions, these observations require experimental validation in subsequent studies.

To determine CBoV's appropriate phylogenetic classification and genetic relatedness to other known parvovirus species, at least one representative virus, as well as the reference genome from each human and animal bocavirus species and their translated protein sequences, was used for generating sequence alignments. The most appropriate protein or nucleotide substitution model was determined using MEGA, and the method with lowest scores was used to calculate pair-wise distances and to construct phylogenetic trees (Fig. 2). All three CBoV proteins (NS, NP and VP) were genetically most related to corresponding MVC proteins; however, there was more genetic diversity/variability between CBoV and MCV proteins than among different HBoV species or different species of genus Parvovirus (Fig. 2). The ICTV criteria for species classification within the genus Bocavirus specify that members of each species are probably antigenically distinct and that natural infection is confined to a single host species. Species are defined as having <95 % homologous NS gene DNA sequences (http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdb/). While the antigenic properties of CBoV were not studied here, we found >35 % genetic divergence in the NS protein compared with other known bocaviruses, suggesting that CBoV and its variants represent one or more novel species within the genus Bocavirus.

Parvoviral capsid proteins contain determinants of immunogenicity and host cell tropism. Minor genetic changes in
these proteins are known to alter the host range and pathogenic potential of paroviruses (Hoelzer et al., 2008a, b; Parrish & Kawaoka, 2005). Moreover, evolutionary studies confirmed that, unlike other DNA viruses, paroviruses can evolve rapidly displaying frequent recombination and mutation rates that approach the high mutation rates observed in RNA viruses (Shackelton & Holmes, 2006; Shackelton et al., 2005, 2007). Expecting that genetic diversity in CBoV capsid proteins should influence their pathogenic potential, we classified CBoV variants according to the genetic relatedness in their capsid protein sequences (Fig. 3a). Calculation of pair-wise distances using the 261 nt long VP1/2 region of 35 CBoV variants resulted in up to 37 % (mean diversity of 15 %) nucleotide and 17 % (mean diversity of 7.3 %) protein sequence divergence (GenBank accession nos JN648104–JN648139). Phylogenetic analysis suggested that most CBoV variants can be divided into three major genetic groups, provisionally named CBoV-A to -C, while some were outliers (Fig. 3a). Our results imply that CBoV represents a highly diverse group of novel canine bocaviruses. The extent of genetic diversity observed among CBoV variants characterized in this single study exceeds the maximum genetic diversity known to exist among all MVC variants reported worldwide to date (Fig. 3a). Combined analyses of genetic diversity and PCR prevalence data suggest that CBoV variants of group A were significantly more prevalent in healthy dogs raised in controlled environments than in animals from other groups; and CBoV variants of group C were substantially more prevalent in dogs with respiratory diseases than in healthy animals (Supplementary Table S1). To further investigate genetic diversity between CBoV-A and -C viruses, we acquired the complete capsid gene sequence of its representative variants. The CBoV-A capsid protein showed 89 % protein identity to CBoV-C (GenBank accession no. JN648135) over its entire length (714 aa, data not shown). To determine the location of amino acid position changes on viral capsid structure, we modelled the secondary structure of CBoV-A and -C for comparison. A homology model was made by giving a human B19 crystal structure coordinates (PDB ID:1S58) as a template model with knowledge based protein modelling program, SWISS-MODEL (Arnold et al., 2006). The final model was structurally and geometrically consistent and did not reveal either structure or sequence discrepancies. The model geometries were cross-validated with the PROCHECK program (Laskowski et al., 1996) to check the accuracy of the CBoV homology model. Both B19 and CBoV structures were superimposed and resulted in a good r.m.s.d value of 0.7 Å using the COOT program (Emsley & Cowtan, 2004). The graphical images were generated with the PYMOL program (www.pymol.org). The
CBoV model contains a high structurally conserved β-barrel motif, a single α-helix and several stretched loops adopt a different conformational position (Fig. 3b). The high structurally similar β-barrel motif region has been implicated for genome packaging and protecting the virus capsid from the environmental damage. The fivefold pore region of the capsid is critical for VP1 externalization, packing of genome and capsid assembly functions (Gurda et al., 2010). The intertwining or flexible loops in the CBoV capsid decorate in the exterior surface region of the capsid and are highly distinguished from all other paroviruses. The variable loops are important to control various biological properties of the capsid and that includes tissue tropism, transduction and receptor binding (Gurda et al., 2010). We noticed that the six amino acid insertion unique to CBoV-A was located in the variable exposed loop (Fig. 3b). Moreover the structural folding of all four outer surface exposed loops was very different between CBoV-A and -C variants possibly reflecting difference in their biological properties.

To conclude, we report several previously uncharacterized species of canine bocaviruses, their sequences, genomic characteristics and genetic diversity. We also compared the genetic diversity and the differences in prevalence of these novel viruses in sick and healthy animals. Animals infected with CBoV-B1 variants and suffering from respiratory infections were housed in a shelter facility. Animal shelters often house animals likely to have a wide variety of infections (Steneroden et al., 2011). These crowded conditions probably facilitate a higher prevalence of viruses in shelter animals compared with the healthy pet population (Helps et al., 2005). Moreover, we could not rule out the possibility that respiratory diseases in these animals were caused by other pathogenic respiratory viruses, but even then the higher prevalence of CBoV-B1 variants in this population alone suggests that these viruses are more likely to infect diseased animals (opportunist infections) or that they can cause or enhance the pathology of other infections (co-infections). We note that the HBoV species include several highly divergent viruses (Kapoor et al., 2010b) and

Fig. 3. (a) Genetic diversity of structural proteins among CBoV variants found in diseased (Dis-disease) and healthy (Con-control) dogs. For comparison, genetic diversity among MVC variants was included in the analysis (lower branch of the tree); and (b) comparative secondary structure of the capsid protein of CBoV variants was used to decipher the structural changes caused by insertion of six amino acid residues. In the ribbon diagram of CBoV1, the secondary structure elements (β-strand in yellow, helix in red and loop in green) are coloured differently. The icosahedral symmetry axes are represented as oval, triangle and pentagon. In the coil representation of CBoV capsid structure the CBoV-A and -C are shown as red and green, respectively. Bars, represent 0.05 substitutions per amino acid site.
therefore more comprehensive disease-association studies should consider the genotype-specific prevalence pattern of these viruses. Comparable and high genetic diversity among CBoV variants makes it a more appropriate model to study HBoV disease-association as well as the evolution and pathogenicity of paroviruses. Despite distant phylogenetic relatedness, we noticed many similarities between CBoV and HBoV species. Both groups of recently identified viruses are genetically diverse and contain many species and genotypes whose pathogenic potential remains unknown (Kapoor et al., 2010b). Unlike other animal bocaviruses, CBoV and HBoV have splicing signals in the NS gene and are likely to encode more than one NS protein (Kapoor et al., 2010a). Unfortunately, the studies for complete biological characterization of HBoV pathogenesis are hampered by the lack of a successful cell culture or animal model (Kapoor et al., 2011). Elucidation of the nearly complete CBoV genome, its genetic diversity and prevalence will help to establish a successful cell culture system for these viruses.

### Acknowledgements

We thank Natasha Qaisar for excellent technical assistance. We are grateful for high throughput sequencing support provided by Roche 454 Life Sciences. This work was supported by awards from the National Institutes of Health (AI090196, AI081132, AI079231, AI57158, AI070411, AI090055, AI072613 and EY017404), the Defense Threat Reduction Agency and USDA 58-1275-7-370.

### References


