A Newly Identified Bocavirus Species in Human Stool

Published in final edited form as:
J Infect Dis. 2009 January 15; 199(2): 196–200. doi:10.1086/595831.

A new bocavirus species in human stool

Amit Kapoor[1,2], **Elizabeth Slikas**[1,2], **Peter Simmonds**[3], **Thaweesak Chieochansin**[3], **Asif Naeem**[4], **Shahzad Shaukat**[4], **Muhammad Masroor Alam**[4], **Salmaan Sharif**[4], **Mehar Angez**[4], **Sohail Zaidi**[4], and **Eric Delwart**[1,2]

1 *Blood Systems Research Institute, San Francisco CA 94118*

2 *Dept. of Laboratory Medicine, University of California, San Francisco*

3 *Centre for Infectious Diseases, University of Edinburgh, UK*

4 *National Institute of Health, Pakistan, Department of Virology*

Abstract

Using viral metagenomics we identified a novel parvovirus species in human stool whose closest phylogenetic relative is the human bocavirus (HBoV). HBoV2 has an identical genomic organization to HBoV but share only 78%, 67%, and 80% identity to its NS1, NP1 and VP1/VP2 proteins. Using PCR we detected HBoV2 sequences in 5/98 Pakistani children stool samples and 3/699 stool samples from the UK. Near full genome sequencing showed the presence of three divergent genotypes and evidence of recombination. Further studies are required to determine sites of replication of HBoV2 and potential associations with clinical symptoms or disease.

Introduction

The *Parvoviridae* family includes the *Parvovirinae* subfamily infecting vertebrates and the *Densovirinae* subfamily infecting insect arthropods. *Parvovirinae* subfamily members known to infect humans consist of parvovirus B19 (in the *Erythrovirus* genus)[1], the apathogenic, typically helper dependent, human adeno-associated viruses (AAV in the *Dependovirus genus*), the human bocavirus (HBoV in the *Bocavirus* genus) and PARV4 (unassigned genus) [2,3]. HBoV was discovered in respiratory samples from children [3] but is also found in stool samples of children with gastroenteritis [4-10]. The association of HBoV with both respiratory infections and gastroenteritis are highly active areas of research that have been recently reviewed [11-14]. The pathogenicity of PARV4, detected in blood and tissues, particularly in injection drug users and HIV infected subjects, remains undetermined [2,15,16].

B19 and PARV4 are found as closely related but distinct genotypes [17-21] and are also archived in lymphoid tissue for years after primary plasma viremia has resolved [16,22,23]. The observation of distinct genotype distributions of archived B19 and PARV4 in different age cohorts likely reflects large-scale epidemic sweeps and associated genotype replacements [16,22-24].

Using non-specific viral particle purification and random nucleic acid amplification combined with minimal sequencing we identified and sequenced a new human parvovirus related to HBoV in stool samples of a Pakistani child with acute flaccid paralysis (AFP). We report here

Reprints or correspondence: Eric Delwart, Blood Systems Research Institute, 270 Masonic Ave., San Francisco, CA 94118. Tel: 415 −9235763, Efax: 419−7914220, Email: E-mail: delwarte@medicine.ucsf.edu.

Conflict of interest: A provisional patent application has been filed for HBoV2.

on its genomic and phylogenetic analysis, along with an initial investigation of its prevalence and genetic diversity.

Patients, materials, and methods

Stools from 57 children with non-polio AFP (mean age 54.6 months) and 41 from healthy Pakistani children (mean age 39.8 months) were analyzed. Samples were collected as part of the WHO poliovirus eradication program. 699 stool samples from a mixed age population submitted for enteric bacteriology screening in Edinburgh were analyzed.

Stool supernatants were processed for viral metagenomics as previously described [25] with minor modifications (supplemental methods). The resulting random PCR products derived from nuclease resistant, viral-sized particles associated RNA and DNA were subcloned and plasmids inserts sequenced. The resulting sequences were analyzed by tBLASTx against Genbank database. The viral 5' and 3' extremities of HBoV2 were amplified using a modification of RACE (rapid amplification of cDNA ends) [2].

Conditions for HBoV and HBoV2 PCR are described in supplemental methods.

Sequence distances for different genomic regions were measured using built in functions in the Simmonics2005 sequence editor v1.6 [26]. Trees were constructed from pairwise nucleotide and amino acid sequence distances by neighbour-joining in the MEGA2 package. The robustness of groupings was calculated by bootstrap re-sampling of 1000 replicates of the data. Genbank accession numbers are: xxxxx. All studies were performed with approval of the UCSF committee on human research.

Results

Viral-sized particles were first purified from two consecutive stool samples from a Pakistani child with AFP. Nuclease resistant (i.e. capsid protected) viral nucleic acids were then extracted and randomly amplified (see Materials and Methods). The resulting random amplification products were subcloned and 97 plasmid inserts were sequenced and analyzed by tBLASTx.

Exact sequence matches were found in Genbank to human sequences as well as to *Micrococcus luteus*, *Pseudomonas fluorescens* and uncultured bacterium sequences. Highly significant but imperfect matches (E score $<10^{-10}$) we also found to Chlamydophila pneumoniae, Rhodoferax ferrireducens and numerous bacteriophages. A single perfect sequence match was found to human poliovirus 1 vaccine strain Sabin 1. The detection of a polio Sabin 1 viral sequence likely reflected ongoing replication of orally administered polio vaccine in this child of 36 months who had previously received a total of 14 oral polio vaccinations.

11/47 plasmid sequences from the first stool sample (plus 19/48 from the second time point) gave highly significant tBLASTx E scores to the HBoV genome reference sequence (NC_007455). PCR was then used to link the different HBoV-like fragments while 5' and 3' RACE [2] were used to amplify the viral extremities. 5196 bases of a novel bocavirus genome were assembled. Based on the ST2 prototype genome sequence of HBoV we estimate, that at least 7 bases are missing from the 5' end while the genome described here extends for a further 25 bases at the 3' end. Because the closest genetic relative of this new virus was HBoV we named it HBoV2.

The arrangement of ORF's in the prototypic HBoV2 genome was similar to that of HBoV, with three large, coding sequences (Supplemental fig. 1). The 5'NS1 ORF is required for viral DNA replication and the regulation of viral gene expression. Its protein sequence identity with HBoV was 78% and was co-linear through the gene. The second ORF, NP1, was 4 amino acid shorter

and 67% identical. NP1 is a protein of unknown function restricted to bocaviruses [27, 28]. The third large ORF encoded a protein with 80% identity to the VP1/VP2 of HBoV. Relative to the HBoV coding sequence, the VP1 of HBoV2 was preceded by a 25 amino acids methionine initiated ORF stretch and a four amino acid deletion downstream resulting in a slightly larger VP1. Through comparison with HBoV, we predicted the VP2 protein of HBoV2 started at the 154th amino acid of the third large ORF.

To investigate whether widely used PCR assays for HBoV would be able to amplify HBoV2, we extracted from the literature the PCR primers used for HBoV and aligned them with the homologous regions of HBoV2 (supplementary Table 1). Most PCR primers contained a substantial number of mismatches with HBoV2 that would preclude or greatly reduce the efficiency of amplification and amplicon detection.

To determine the relationship of HBoV2 to other members of the *Bocavirus* genus, phylogenetic analyses were performed for the three large ORFs (VP1/2, NP-1 and NS) using both nucleotide and deduced protein sequences (Fig. 1). The prototype HBoV2 variant (PK5510), while more closely related to HBoV than the animal bocaviruses, CnMV and BPV-1, consistently adopted an outlier position to the clade containing all the published HBoV sequences in all three genomic regions. Reflecting this, pairwise nucleotide distances between HBoV2 and HBoV were substantially greater (22−26%) than within the HBoV clade (0.4 −0.9%) (Table 1), but less than between HBoV or HBoV2 sequences and animal bocaviruses (46−56%). Similarly, while almost no amino acid sequence variability was observed among HBoV sequences in any genomic region (0.2−0.5%), sequences of all three major genes differed substantially from HBoV2 (20−33%). An unusual partial NS1 HBoV sequence has been reported in Brazil and adopted an intermediate position between HBoV and HBoV2 on phylogenetic analysis (Fig. 1 Partial NS1 EF560212 in grey). Interestingly this Brazilian HBoV sequence was derived from the feces of a child with gastrointestinal symptoms [8].

To determine the prevalence of human bocaviruses we used nested PCR primers specifically targeting HBoV and HBoV2 NS region. DNA from the stool samples of 57 Pakistani children with AFP and 41 healthy Pakistani children, plus 699 stool samples submitted for enteric bacteriology screening in Edinburgh were analyzed. A total of three AFP stool samples (including the original patient 5510) and two stools from healthy Pakistani children were positive for HBoV2 DNA sequences (ages were 12, 16, 36, 36, and 96 months). HBoV was not found among the 57 AFP stools. The stool samples from the UK were tested in pools of 10 samples in which two HBoV (ages 1−2 years and 3−5 years) and three HBoV2 (ages 0−3 months, 6−12 months and >65 years) positive samples were identified.

To investigate the genetic diversity of HBoV2, the partial NS region amplicons from the Pakistan and UK samples were sequenced and compared to prototypic HBoV2 and other available human and animal bocavirus sequences (Fig. 1 Partial NS1). Although this initial survey was small, HBoV2 variants fell into 3 groups or genotypes (gts) (Fig 1 Partial NS1) showing 4.5% sequence divergence from each other in this region of the bocavirus genome (Table 2B), a level of diversity much greater than observed between known variants of HBoV (0.7% in this region). HBoV diversity was more comparable to mean diversity seen within any of the three HBoV2 genotype (0.9%).

To further investigate HBoV2 diversity, we obtained near complete gnome sequences of representative of the other two NS1 based genotypes (gt1: HBoV2 prototype PK5510, gt2: PK2255, and gt3: UK648). Neither HBoV2 variant contained the methionine initiated 25 amino acid stretch seenupstream of VP1 in PK5510 prototype but had otherwise identical length NS1 and NP1 genes (except from NP1 of PK5510 that was one amino acid shorter). As observed for HBoV, sequence divergence within HBoV2 was highly conservative, with very low dN/

dS ratios in all three genome regions (Table 1). Despite the approximately equal degrees of sequence divergence between the HBoV2 genotypes in the partial region of NS1 initially sequenced, phylogenetic relationships varied in different genome regions (Fig. 1). The gt1-PK5510 and gt2-PK2255 variants clustered closely in VP1, with gt3-UK648 as a outlier, while gt3-UK648 and gt2-PK2255 clustered in the NP1 and NS1 genes with gt1-PK5510 as the outlier (Fig 1). Using sliding window analysis of pair-wise nucleotide distances over the HBoV2 genome, gt2-PK2255 and gt3-UK648 showed a lower level of divergence relative to the other two pair-wise comparisons in the NS1 and NP1 regions (supplementary Fig 2 red line) while gt1-PK5510 and gt2-PK2255 were nearly identical in the VP1/VP2 region (supplementary Fig 2 green line). Discordant phylogenies and inconsistent sequence divergence values between HBoV2 variants is consistent with the occurrence of complex recombination events in the evolution of these viruses. Putative breakpoints were located near the middle and 3' end of NS1 and potentially more recently (*ie.recombinants* showing a lower level of divergence) near the beginning of the VP1/VP2.

Discussion

Based on the VIIIth report of the International Committee on Taxonomy of Viruses (ICTV), different bocavirus species should show NS gene nucleotide sequence similarities below 95%. HBoV2, showing 75.6% nucleotide similarity to its closest relative HBoV therefore qualifies as a new human parvovirus and the fourth species in the *Bocavirus* genus following BPV, CnMV and HBoV. We believe this terminology is more appropriate than its alternative possible name as HBoV genotype 2 because genotypes of other parvoviruses, such as those reported for B19 and PARV4 are much less divergent from each other and do not qualify as separate species using ICTV guidelines.

HBoV2 divergence also likely precluded its detection using HBoV based PCR (supplementary Table 1). Thus the extensive epidemiological and clinical information on human bocavirus collected to date refers exclusively to HBoV, and the available sample archives of clinical specimens will have to be screened again with HBoV2-specific primers to investigate its frequency and potential disease associations.

Because parvoviruses are particularly hard to inactivate using heat and detergent treatment they are of special concern in blood product transfusions and some countries screen for parvovirus B19 DNA to exclude highly viremic donations from blood derived products [29-32]. Whether HBoV2 can result in plasma viremia as seen for HBoV and PARV4 [13, 33], will require further studies. The development of serological assays, as recently achieved for HBoV [13,33], will allow larger epidemiological studies to measure the rate of seroconversion in different age and geographic cohorts and whether detection of IgM and rise in IgG titers are associated with particular symptoms.

Three genotypes of HBoV2 were characterized whose genetic distances to one another (in partial NS1 sequences) were comparable to those measured between B19 or PARV4 genotypes. The HBoV2 genotype geographic distribution appeared different with all three HBoV2 gt3 being derived from the UK while both genotype 1 and 2 where found in Pakistan. Recombination events among HBoV2 variants were also observed as recently reported for animal parvoviruses [34]. The existence of geographic clustering, further genotypes, and different patterns of recombination will require further screening of samples from a wider geographical base. If HBoV2 is found archived in tissue as shown for B19 and PARV4 [16, 22-24] the possibility of past epidemic waves of different HBoV2 genotypes will also be testable using tissues from cohorts of different ages.

Although the detection of HBoV2 was limited to fecal samples in the current study, its detection in stools from 5 Pakistani children lends further supports to the gastrointestinal tract as a site of replication of human bocaviruses [4,7-10]. The detection of HBoV2 in the stool of 3 UK residents (2 being less then 12 months old and 1 >65 years old) also indicates that this virus is not restricted to South Asia nor to young children. Whether the lower prevalence of HBoV2 in the UK stools reflect reduced PCR sensitivity due to sample pooling, the older population tested, reduced virus exposure, or reduced duration or level of enteric viremia remains to be determined. Further PCR analysis on blood, cerebrospinal and respiratory fluids and stool samples from symptomatic and matched healthy controls will also be needed establish whether this virus is associated with respiratory, gastro-intestinal or other symptoms. The equal rate of detection of HBoV2 in non-polio AFP and healthy Pakistani children (5%) indicates that HBoV2 is unlikely to be associated with AFP.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

we thank Drs Michael P. Busch and Blood Systems Research Institute for sustained support and NHLBI grant R01HL083254 to ELD.

References

1. Cossart YE, Field AM, Cant B, Widdows D. Parvovirus-like particles in human sera. Lancet 1975;1:72–3. [PubMed: 46024]
2. Jones MS, Kapoor A, Lukashov VV, Simmonds P, Hecht F, Delwart E. New DNA viruses identified in patients with acute viral infection syndrome. J Virol 2005;79:8230–6. [PubMed: 15956568]
3. Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B. Cloning of a human parvovirus by molecular screening of respiratory tract samples. Proc Natl Acad Sci U S A 2005;102:12891–12896. [PubMed: 16118271]
4. Cheng WX, Jin Y, Duan ZJ, et al. Human bocavirus in children hospitalized for acute gastroenteritis: a case-control study. Clin Infect Dis 2008;47:161–7. [PubMed: 18532891]
5. Yu JM, Li DD, Xu ZQ, et al. Human bocavirus infection in children hospitalized with acute gastroenteritis in China. J Clin Virol. 2008
6. Neske F, Blessing K, Tollmann F, et al. Real-time PCR for diagnosis of human bocavirus infections and phylogenetic analysis. J Clin Microbiol 2007;45:2116–22. [PubMed: 17475762]
7. Vicente D, Cilla G, Montes M, Perez-Yarza EG, Perez-Trallero E. Human bocavirus, a respiratory and enteric virus. Emerg Infect Dis 2007;13:636–7. [PubMed: 17553287]
8. Albuquerque MC, Rocha LN, Benati FJ, et al. Human bocavirus infection in children with gastroenteritis, Brazil. Emerg Infect Dis 2007;13:1756–8. [PubMed: 18217564]
9. Lau SK, Yip CC, Que TL, et al. Clinical and molecular epidemiology of human bocavirus in respiratory and fecal samples from children in Hong Kong. J Infect Dis 2007;196:986–93. [PubMed: 17763318]
10. Lee JI, Chung JY, Han TH, Song MO, Hwang ES. Detection of human bocavirus in children hospitalized because of acute gastroenteritis. J Infect Dis 2007;196:994–7. [PubMed: 17763319]
11. Sloots TP, Whiley DM, Lambert SB, Nissen MD. Emerging respiratory agents: New viruses for old diseases? J Clin Virol. 2008
12. Allander T. Human bocavirus. J Clin Virol 2008;41:29–33. [PubMed: 18055252]
13. Kantola K, Hedman L, Allander T, et al. Serodiagnosis of human bocavirus infection. Clin Infect Dis 2008;46:540–6. [PubMed: 18199037]
14. Schildgen O, Muller A, Allander T, et al. Human bocavirus: passenger or pathogen in acute respiratory tract infections? Clin Microbiol Rev 2008;21:291–304. [PubMed: 18400798]table of contents

15. Fryer JF, Delwart E, Hecht FM, et al. Frequent detection of the parvoviruses, PARV4 and PARV5, in plasma from blood donors and symptomatic individuals. Transfusion 2007;47:1054–61. [PubMed: 17524097]

16. Manning A, Willey SJ, Bell JE, Simmonds P. Comparison of tissue distribution, persistence, and molecular epidemiology of parvovirus B19 and novel human parvoviruses PARV4 and human bocavirus. J Infect Dis 2007;195:1345–52. [PubMed: 17397006]

17. Simmonds P, Douglas J, Bestetti G, et al. A third genotype of PARV4 in sub-Saharan Africa. J. Gen. Virol. 2008In press

18. Kesebir D, Vazquez M, Weibel C, et al. Human bocavirus infection in young children in the United States: molecular epidemiological profile and clinical characteristics of a newly emerging respiratory virus. J Infect Dis 2006;194:1276–82. [PubMed: 17041854]

19. Fryer JF, Delwart E, Bernardin F, Tuke PW, Lukashov VV, Baylis SA. Analysis of two human parvovirus PARV4 genotypes identified in human plasma for fractionation. J Gen Virol 2007;88:2162–7. [PubMed: 17622618]

20. Cohen BJ, Gandhi J, Clewley JP. Genetic variants of parvovirus B19 identified in the United Kingdom: implications for diagnostic testing. J Clin Virol 2006;36:152–5. [PubMed: 16569510]

21. Servant A, Laperche S, Lallemand F, et al. Genetic diversity within human erythroviruses: identification of three genotypes. J Virol 2002;76:9124–34. [PubMed: 12186896]

22. Hokynar K, Soderlund-Venermo M, Pesonen M, et al. A new parvovirus genotype persistent in human skin. Virology 2002;302:224–8. [PubMed: 12441066]

23. Norja P, Hokynar K, Aaltonen LM, et al. Bioportfolio: lifelong persistence of variant and prototypic erythrovirus DNA genomes in human tissue. Proc Natl Acad Sci U S A 2006;103:7450–3. [PubMed: 16651522]

24. Norja P, Eis-Hubinger AM, Soderlund-Venermo M, Hedman K, Simmonds P. Rapid sequence change and geographical spread of human parvovirus B19; comparison of B19 evolution in acute and persistent infections. J Virol. 2008

25. Kapoor A, Victoria J, Simmonds P, et al. A highly divergent picornavirus in a marine mammal. J Virol 2008;82:311–20. [PubMed: 17942560]

26. Simmonds P, Tuplin A, Evans DJ. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. Rna 2004;10:1337–51. [PubMed: 15273323]

27. Schwartz D, Green B, Carmichael LE, Parrish CR. The canine minute virus (minute virus of canines) is a distinct parvovirus that is most similar to bovine parvovirus. Virology 2002;302:219–23. [PubMed: 12441065]

28. Chen KC, Shull BC, Moses EA, Lederman M, Stout ER, Bates RC. Complete nucleotide sequence and genome organization of bovine parvovirus. J Virol 1986;60:1085–97. [PubMed: 3783814]

29. Luban NL. Human parvoviruses: implications for transfusion medicine. Transfusion 1994;34:821–7. [PubMed: 8091474]

30. Kleinman SH, Glynn SA, Lee TH, et al. Prevalence and quantitation of parvovirus B19 DNA levels in blood donors with a sensitive polymerase chain reaction screening assay. Transfusion 2007;47:1756–64. [PubMed: 17880600]

31. Horowitz B, Busch M. Estimating the pathogen safety of manufactured human plasma products: application to fibrin sealants and to thrombin. Transfusion. 2008

32. McOmish F, Yap PL, Jordan A, Hart H, Cohen BJ, Simmonds P. Detection of parvovirus B19 in donated blood: a model system for screening by polymerase chain reaction. J Clin Microbiol 1993;31:323–8. [PubMed: 8432819]

33. Endo R, Ishiguro N, Kikuta H, et al. Seroepidemiology of human bocavirus in Hokkaido prefecture, Japan. J Clin Microbiol 2007;45:3218–23. [PubMed: 17699639]

34. Shackelton LA, Hoelzer K, Parrish CR, Holmes EC. Comparative analysis reveals frequent recombination in the parvoviruses. J Gen Virol 2007;88:3294–301. [PubMed: 18024898]

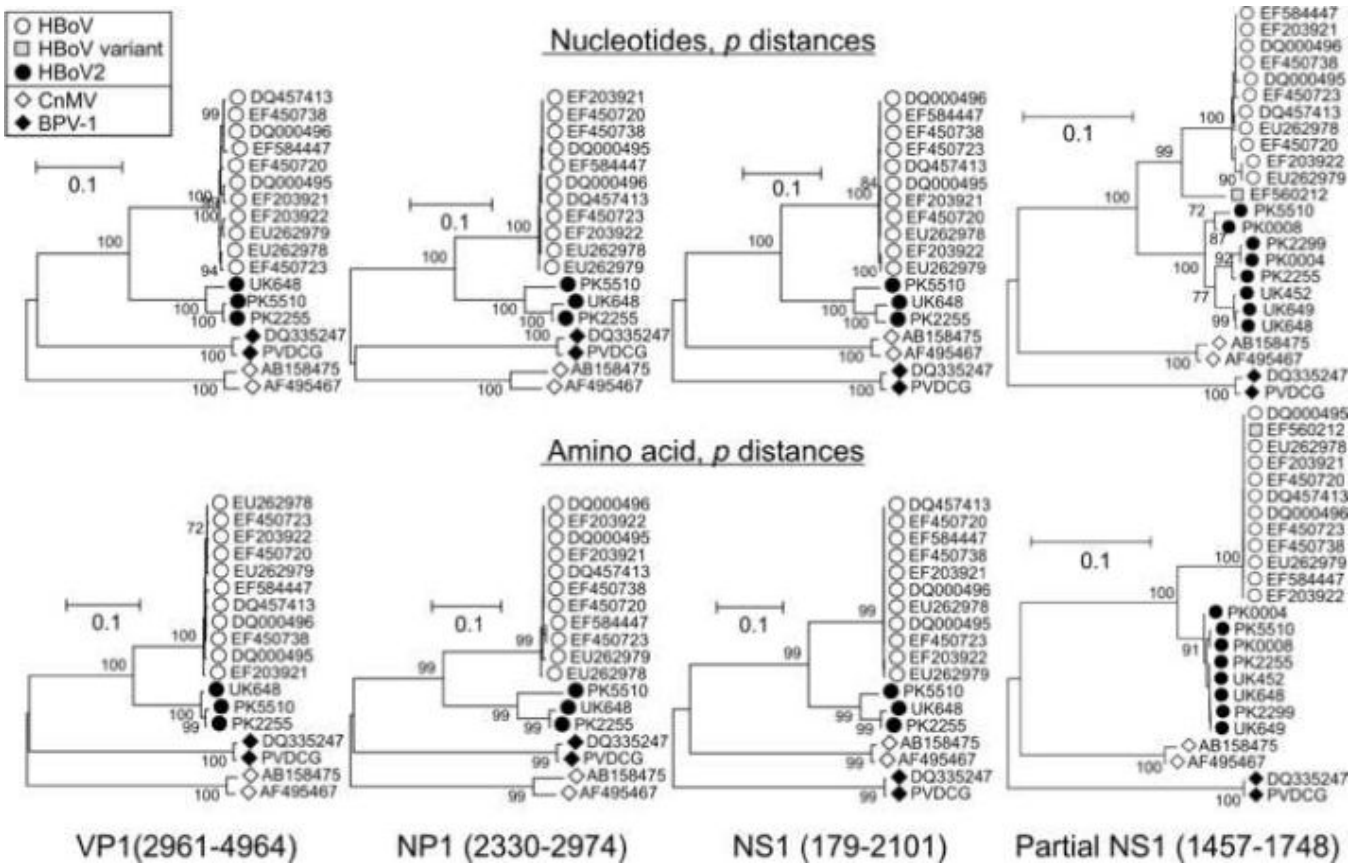

Figure 1. Phylogenetic analysis of Bocaviruses

The 3 major open reading frames were analyzed using both nucleotide and protein sequences of representative variants of HBoV, HParV5, CnMV (Canine minute virus) and BPV-1 (Bovine parvovirus-1). Analysis of the partial sequence of NS1 was used to show phylogenetic relationships of between a larger number of samples amplified by PCR; it also corresponded to the region of the partially sequenced Brazilian HBoV variant (EF560212).

Table 1

Nucleotide and protein sequence comparisons among and between bocaviruses.

Comparison	Region	Comps.	Nucleotide	Amino acid	dN/dS
A) Main ORFs					
Within HBoV2:	VP1/2	3	3.1%	1.1%	0.094
”	NP1	3	5.9%	8.7%	0.348
”	NS1	3	5.8%	4.3%	0.103
Within HBoV:	VP1/2	55	0.9%	0.5%	0.075
”	NP1	55	0.4%	0.3%	0.103
”	NS1	55	0.4%	0.2%	0.068
HBoV-HBoV2:	VP1/2	33	22.2%	20.2%	<0.13*
”	NP1	33	22.9%	32.9%	<0.38*
”	NS1	33	26.2%	27.8%	<0.19*
HBoV2—B/CnPV:	VP1/2	12	46.2−46.3%	51.9−53.2%	n.c.
”	NP-1	12	47.0−52.3%	59.7−61.3%	n.c.
”	NS1	12	51.3−55.8%	60.6−65.1%	n.c.
B) Partial NS1 region (positions 1457−1748)					
Within HBoV2 genotypes:		7	0.9%	0.3%	0.022
Within HBoV (main group):		55	0.7%	0.0%	0.000
Between HBoV2 genotypes:		21	4.5%	0.6%	0.007
HBoV — EF560212:		11	9.0%	1.1%	0.003
HBoV—HBoV2:		88	18.0%	8.3%	0.071

* Only upper limits for dN/dS values for comparisons between HBoV and HBoV2 could be calculated as synonymous distances corrected for multiple substitutions approached or exceeded 1.0, and were therefore likely substantially underestimated.