Identification and Characterization of a New Bocavirus Species in Gorillas

Amit Kapoor1*, Natasha Mehta1, Frank Esper2, Mateja Poljsak-Prijatelj3, Phenix-Lan Quan1, Natasha Qaisar1, Eric Delwart4, W. Ian Lipkin1

1 Center for Infection and Immunity, Columbia University, New York, New York, United States of America, 2 Rainbow Babies and Children’s Hospital, Cleveland, Ohio, United States of America, 3 Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia, 4 Blood Systems Research Institute, University of California San Francisco, San Francisco, California, United States of America

Abstract

A novel parvovirus, provisionally named Gorilla Bocavirus species 1 (GBoV1), was identified in four stool samples from Western gorillas (Gorilla gorilla) with acute enteritis. The complete genomic sequence of the new parvovirus revealed three open reading frames (ORFs) with an organization similar to that of known bocaviruses. Phylogenetic analysis using complete capsid and non-structural (NS) gene sequence suggested that the new parvovirus is most closely related to human bocaviruses (HBoV). However, the NS ORF is more similar in length to the NS ORF found in canine minute virus and bovine parvovirus than in HBoV. Comparative genetic analysis using GBoV and HBoV genomes enabled characterization of unique splice donor and acceptor sites that appear to be highly conserved among all four HBoV species, and provided evidence for expression of two different NS proteins in all primate bocaviruses. GBoV is the first non-human primate bocavirus identified and provides new insights into the genetic diversity and evolution of this highly prevalent and recently discovered group of parvoviruses.

Introduction

Parvoviruses, members of family Paroviridae, are small, non-enveloped icosahedral viruses with single-stranded linear DNA genomes that frequently infect animals through the fecal-oral route[1]. Parvoviruses are ubiquitous and can cause a broad spectrum of diseases in animals, including enteritis, panleukopenia, hepatitis, erythrocyte aplasia, immune complex–mediated vasculitis, and cerebellar ataxia [1]. Efficacious vaccines against animal parvovirus infections are widely employed [2,3,4,5,6].

The family Paroviridae as currently defined, comprises two subfamilies, Densovirinae and Parovirinae, that infect non-vertebrate and vertebrate hosts, respectively[1]. The International Committee on Taxonomy of Viruses (ICTV) has further classified the sub-family Parovirinae into five genera: Dependovirus, Bocavirus, Erythrovirus, Parvovirus, and Adnovirus. Bocaviruses are unique among parvoviruses, as they contain a third ORF between non-structural and structural coding regions [7]. The genus Bovicivirus includes the bovine parvoviruses, canine minute viruses and four species of human bocaviruses (HBoV1-4)[1,8,9,10,11,12].

HBoV1 infection has been linked with mild to severe lower respiratory tract infections in children, frequently in association with other viral infections [8,13,14,15,16,17,18,19,20]. HBoV1 has also been detected at low frequency in stool samples although association with diarrhea appears weaker than with respiratory symptoms [9,10,21,22,23,24]. HBoV1 strains show a low degree of genetic variability world-wide [35,36]. HBoV2 was recently identified in stool samples of Pakistani children [12]. Lower frequencies of HBoV2 were detected in the stool of Scottish adults and children [12]. A third bocavirus species, HBoV3, as well as HBoV2, were recently found in stool samples from Australian children with diarrhea [10]. HBoV4 was first identified in stool samples of children from Nigeria and Tunisia[11].

Group-reactive or pan-PCR approaches have been used to identify new viruses in human, animal and environmental samples [25,26,27,28,29]. We recently discovered two novel HBoV species using this approach [11]. In a survey for divergent bocaviruses in other non-human primates, we studied stool samples from three groups of western gorillas (Gorilla gorilla) with enteritis using pan-bocavirus PCR assay. This work describes the nearly complete genome of the first non-human primate bocavirus species identified in four stool samples of gorillas, tentatively named Gorilla Bocavirus species 1 (GBoV1).

Materials and Methods

Samples and source

All stool samples studied were from western gorillas (Gorilla gorilla) in a captive colony in North America. Samples were collected during April 2009, during an outbreak of enteritis resulting in diarrhea among three different groups of Gorillas. Group A = 12 animals, group B = 7 animals and group C = 4 animals. All fecal material was recovered from the floor of zoo enclosures. Samples were identified by group of origin, but the
Sample processing and pan-bocavirus PCR assay

Stool samples were diluted 1:9 with Hank’s Buffered Saline Solution (Gibco-BRL), mixed with glass beads, vigorously vortexed and centrifuged twice at 2700 g for 10 min. Clarified supernatants (50 μl) were used for total nucleic acid extraction using the NucliSens easyMag system (BioMerieux). Total nucleic acids were eluted in 50 μl DEPC-treated water. PCR primers panBOV-F1 (5′-TAAGCTGCATAGAGGTGTTAGACCC-3′) and pan-BOV-R1 (5′- GTCAAGTCTATYCTCTTARACACCA-3′) were used for the first round of hemi-nested PCR; while primers panBOV-F2 (5′- CAGAAGATGTTGTGATGACGCACGWC-3′) and panBOV-R1 (same as first round) were used for the second round of hemi-nested PCR. For the first round of nested PCR, 5 μl of each specimen DNA was mixed with 5.2 μl 10× polymerase reaction buffer (Qiagen), 1.2 μl each dNTP (10 mM), 50 pmol forward (both panBOV-F1) and reverse primer (panBOV-R1), 1 μl HotStart Taq DNA polymerase (Qiagen) and 31 μl DEPC-treated water, in a total reaction volume of 50 μl.

The reaction was performed using initial denaturation at 95°C for 7 min, followed by six cycles of 95°C for 40 sec, 60°C for 45 sec and 68°C for 30 sec, then 35 cycles of 95°C for 30 sec, 57°C for 30 sec and 68°C for 30 sec, and final extension at 72°C for 10 min. During the first round of PCR, the first six cycles were done at a high annealing temperature to increase stringency and then remaining cycles at lower temperature to facilitate primer hybridization by tolerating some nucleotide mismatches. For the second round of hemi-nested PCR, identical cycling conditions were used, with an annealing temperature of 64°C for the first six cycles and annealing temperature of 58°C for the remaining 35 cycles. The reaction mixture for the second round was added with 0.5 μl PCR product from the first round. Products were visualized following electrophoresis on 2% agarose gel. PCR products showing positive bands of approximately 290 bp, corresponding to the highly conserved amplified NS gene fragment, were purified using a PCR purification kit (Qiagen) and directly sequenced from both directions.

Complete genome sequencing and phylogenetic analysis

Complete genome of the GBoV1 was acquired using primer walking approach as previously described [11]. In brief, each extension step of PCR used one primer specific for the novel virus sequence and other degenerated to hybridize all known bocavirus sequences. After the complete genome was assembled, each base is sequenced in triplicate to confirm the sequence. The near complete genome of GBoV1 is submitted to Genbank under accession number HM145750.

To determine the sequence relationship between GBoV1 and other known bocavirus species, at least one representative virus member, including the reference genome from each species and their translated protein sequences, were used for generating sequence alignments. Sequences used for the comparison comprised the following: HBoV1 (GenBank accession no. 19616-8955, 161137732, 149329863, 123616693, 161137737, 149389-621, 149389676, 223675836, 19616950, 149369626, 284799140, 256629295, 149389721, 149389656, 241995009, 223951374, 196169010, 149389641, 90287153, 226434473, 223670876, 223670861, 196169000, 196168975, 196168960, 149389651, 149389601, 66356133, 66356128, 219879196, 196168985, 196168945, 149389711, 149389706, 149389696, 149389661, 149389646, 149389606, 125616883, 196168995, 149389701, 149389691, 149389681, 149389671, 149389666, 1493896631, 130918044, 84873482, 260779772, 25662990, 223644468, 223670851, 223670846, 223670841, 223670831, 125719360, 226434443). HBoV2 (GenBank accession no. 24199-4927, 213012741, 223670916, 237690166, 158714105, 238476-986, 158713998, 290565732, 213012746, 290565729, 213012751, 238408644, 237690154, 228040639, 290565727). HBoV3 (GenBank accession no. 23476991, 26572554, 223675873, 223670918, 237690184), HBoV4 (GenBank accession no. 289163341), canine minute virus (GenBank accession no. NC_004442) and bovine parvovirus-1 (GenBank accession no. 9627190). Phylogenetic analyses of CLUSTAL W aligned regions were carried out by neighbour joining of nucleotide or amino acid p-distances, implemented in the program MEGA4[30]. Bootstrap resampling was carried out to demonstrate robustness of groupings. Recombination analysis was performed by Symplot[31], using reference genomes of all four HBoV species and GBoV1. To assess similarity across the genomes, sequence scans were performed using a fragment length of 300 bases and an increment of bases between fragments.

Results

Identification and near complete genome of GBoV1

Twenty-three stool samples collected during an outbreak of enteritis with diarrhea in captive gorillas were studied for presence of divergent bocaviruses related to known HBoV. Four samples (two each from group B and C) yielded amplification products using a pan-bocavirus PCR. To increase the sensitivity of the PCR specific primers targeting GBoV1 sequence were used to screen all 23 samples. PCR primers GBoV-F1 (5′- GCACCAAGACTGTGGTGAAACC -3′) and GBoV-R1 (5′- GCACCATGTAGTAGCTGC-3′) were used for the first round of PCR. Primers GBoV-F2 (5′- CAAAGACTGGGTGGGACGACGC-3′) and GBoV-R2/5′- GCACCATGTGACTGAGCTGCAAAT-3′ were used for the second round. For the first round of nested PCR, 5 μl of each specimen DNA was mixed with 5.2 μl 10× polymerase reaction buffer (Qiagen), 1.2 μl each dNTP (10 mM), 20 pmol forward (GBoV-F1) and reverse primer (GBoV-R1), 1 μl HotStart Taq DNA polymerase (Qiagen) and 31 μl DEPC-treated water, in a total reaction volume of 50 μl. The reaction was performed using initial denaturation at 95°C for 7 min, followed by six cycles of 95°C for 40 sec, 61°C for 45 sec and 72°C for 30 sec, then 35 cycles of 95°C for 30 sec, 59°C for 30 sec and 72°C for 30 sec, and final extension at 72°C for 10 min. For the second round of nested PCR, identical cycling conditions were used, with an annealing temperature of 60°C for the first six cycles and annealing temperature of 58°C for the remaining 35 cycles. The reaction mixture for the second round was added with 0.5 μl PCR product from the first round. Products were visualized following electrophoresis on 2% agarose gel. PCR products showing positive bands of approximately 290 bp, corresponding identity of the individual from which the sample was obtained was not always known.
VP2 capsid proteins; ORF3 encodes NP1. The GBoV1NS protein gene comprises 2412 nt and is predicted to code for 804 amino acids (aa) with a mass of approximately 91 kDa, making it the second longest NS protein encoded by a bocavirus (HBoV1 – 781 aa, HBoV2 -777 aa, HBoV3 – 780 aa and HBoV4 – 776 aa, canine minute virus - 774 aa and bovine parvovirus - 860 aa). Conserved motifs associated with rolling circle replication, helicase and ATPase are present within ORF1. The GBoV1 virion protein gene comprises 2013 nt and is predicted to code for two structural proteins VP1 and VP2 of 671 aa and 542 aa length, respectively. Within ORF2, the phospholipase A2 motifs required for parvovirus infectivity situated within the VP1-unique (VP1u) region were also found in association with calcium-binding loop and catalytic residues. The occurrence of the non-structural protein NP1 encoded by an ORF in the middle of the genome is a unique feature of bocaviruses. The GBoV NP1 gene comprises 657 nt and is predicted to code for a highly basic and stable 219 aa protein with a mass of approximately 26 kDa. NP1 is a highly phosphorylated protein of currently undetermined function [32]; NP1 differs in length between different HBoV species, ranging from 219 aa (HBoV1), 214 aa (HBoV2A1, HBoV3), 215 aa (HBoV1A2, HBoV3A, HBoV4), 217 aa (HBoV4). 

Phylogenetic relationship of GBoV1 with other bocaviruses

To determine the relationship between GBoV1 and other members of the Bocavirus genus, phylogenetic analyses of the three large ORFs - NS1, NP1 and VP1/VP2 - were performed, by use of both nt and deduced aa sequences. In all 3 genomic regions, the GBoV1 was more similar to HBoV1 or HBoV3 than to the other human or animal bocavirus species (CnMV and BPV1), the latter of which consistently occupied an outlier position with regard to the clade containing all primate bocavirus sequence (Fig.1 A and B).

Most HBoV species include several genotypes. Thus, to ensure accurate classification of GBoV1 we examined complete capsid sequences of all the HBoV variants in the GenBank. Phylogenetic analyses suggested that although GBoV1 is genetically closest to HBoV1, the distance between the two viruses is more than that observed within (intra-species distance) any of the four reported HBoV species (Fig.1 A)[11]. Moreover, as HBoV1 was the most genetically homogenous bocavirus species, GBoV1 is more likely to represent a separate genetic clade. Phylogenetic clustering of primate bocaviruses using the different genes was found to be consistent except for HBoV3, which being a recombinant virus[11], clustered in NS gene with HBoV1 clade and in other genes with HBoV2 clade (Fig.1-B).

Figure 1. Phylogenetic analysis of GBoV1. (A) Full length structural protein (VP1/2) sequences of all HBoV and animal bocavirus strains available in GenBank were used to determine phylogeny of GBoV1 by neighbor-joining analysis of pairwise distances between translated amino acid sequences. Bootstrap re-sampling was used to determine robustness of individual clades (values above 70% shown above the branches). (B) The 4 major open reading frames of GBoV1 were analyzed using maximum likelihood composition analysis method (MEGA4.1) comparing pairwise distances of translated sequences of representative variants (reference sequence) of all four HBoV species. Accession numbers of sequences used precede the name of corresponding bocavirus species.

doi:10.1371/journal.pone.0011948.g001
GBoV1 was genetically more similar to HBoV1 in all genes, suggesting it is not an inter-species recombinant virus (Fig.1-B). Phylogenetic trees using different genes also showed that the genetic distance of GBoV1 and its closest related virus (HBoV1) was more than that between HBoV2 and 4 (in all genes), as well between HBoV2, 3 and 4 (in VP1 and VP2 genes) (Fig.1-B). These relationships are described in Table 1 in terms of pairwise nt and aa distances. Differences between GBoV1 and HBoV1 and HBoV2 in VP1/2 genes were substantially greater (12.1–12.8 and 13.1–14%, respectively) than those between either HBoV2 and HBoV3 (9.2–10.5 and 9.4–10.8%) or HBo2 and HBoV4 (9.6–11.4 and 9.6–11.3%) or HBoV3 and HBoV4 (7.8–8.0 and 7.9–8.9%). However the genetic distance between the VP1/2 gene of HBoV2 and HBoV2 (19.6–21.9 and 20.3–20.5) was the largest observed between primate bocavirus species (Table 1, shaded numbers).

The International Committee on Taxonomy of Viruses (ICTV) criteria for classification of Bocaviruses establishes that members of each species are probably antigenically distinct and that natural infection is confined to a single host species. Species are defined as <5% homologous NS gene DNA sequence [http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/ or http://phene.cpmc.columbia.edu/ICTV/f_x_parvo.html]. While the antigenic properties of GBoV1 were not studied here, its isolation from a different natural host (gorilla) and >5% genetic distance in NS gene compared to other known bocavirus species, suggest that GBoV1 should be classified as prototype virus of a new bocavirus species.

Table 1. Comparison of pairwise nucleotide and amino acid (bold typeface) distances (p-distance) of all four genes between GBoV1 and HBoV species confirms GBoV1 as prototype of a new bocavirus species.

<table>
<thead>
<tr>
<th></th>
<th>GBoV1</th>
<th>HBoV1</th>
<th>HBoV2</th>
<th>HBoV3</th>
<th>HBoV4</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS</td>
<td>-</td>
<td>7.8</td>
<td>24.2</td>
<td>8.9</td>
<td>25.0</td>
</tr>
<tr>
<td>NP</td>
<td>-</td>
<td>12.8</td>
<td>30.7</td>
<td>16.1</td>
<td>28.3</td>
</tr>
<tr>
<td>VP1</td>
<td>-</td>
<td>12.1</td>
<td>17.2</td>
<td>16.9</td>
<td>16.7</td>
</tr>
<tr>
<td>VP2</td>
<td>-</td>
<td>12.8</td>
<td>18.9</td>
<td>18.9</td>
<td>18.7</td>
</tr>
<tr>
<td>HBoV1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>8.0</td>
<td>-</td>
<td>21.5</td>
<td>8.5</td>
<td>22.3</td>
</tr>
<tr>
<td>NP</td>
<td>12.8</td>
<td>-</td>
<td>31.6</td>
<td>16.6</td>
<td>30.2</td>
</tr>
<tr>
<td>VP1</td>
<td>13.1</td>
<td>-</td>
<td>19.6</td>
<td>19.4</td>
<td>19.1</td>
</tr>
<tr>
<td>VP2</td>
<td>14.0</td>
<td>-</td>
<td>21.8</td>
<td>21.7</td>
<td>21.7</td>
</tr>
<tr>
<td>HBoV2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>24.3</td>
<td>21.8</td>
<td>-</td>
<td>23.5</td>
<td>5.7</td>
</tr>
<tr>
<td>NP</td>
<td>31.3</td>
<td>31.8</td>
<td>-</td>
<td>29.9</td>
<td>11.3</td>
</tr>
<tr>
<td>VP1</td>
<td>18.0</td>
<td>20.3</td>
<td>-</td>
<td>9.2</td>
<td>9.6</td>
</tr>
<tr>
<td>VP2</td>
<td>19.6</td>
<td>22.5</td>
<td>-</td>
<td>10.5</td>
<td>11.4</td>
</tr>
<tr>
<td>HBoV3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>8.9</td>
<td>8.6</td>
<td>23.6</td>
<td>-</td>
<td>24.5</td>
</tr>
<tr>
<td>NP</td>
<td>16.1</td>
<td>16.5</td>
<td>30.8</td>
<td>-</td>
<td>30.8</td>
</tr>
<tr>
<td>VP1</td>
<td>17.4</td>
<td>19.8</td>
<td>9.4</td>
<td>-</td>
<td>7.8</td>
</tr>
<tr>
<td>VP2</td>
<td>19.1</td>
<td>21.9</td>
<td>10.8</td>
<td>-</td>
<td>8.8</td>
</tr>
<tr>
<td>HBoV4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>25</td>
<td>22.5</td>
<td>5.6</td>
<td>24.6</td>
<td>-</td>
</tr>
<tr>
<td>NP</td>
<td>29.0</td>
<td>30.8</td>
<td>12.1</td>
<td>31.8</td>
<td>-</td>
</tr>
<tr>
<td>VP1</td>
<td>17.2</td>
<td>19.5</td>
<td>9.6</td>
<td>7.9</td>
<td>-</td>
</tr>
<tr>
<td>VP2</td>
<td>19.0</td>
<td>21.9</td>
<td>11.3</td>
<td>8.9</td>
<td>-</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0011948.t001

Unique genomic features of GBoV1 and recombination analysis

The genomic organization of all four HBoV species is remarkably similar to that of animal bocaviruses, except that their ORF1 encodes a shorter NS protein (639–650 aa), compared to the longer NS protein of animal bocaviruses (770–860 aa) (Fig.2A). We recently showed the presence of highly conserved potential RNA splicing signals in all HBoV species near the end of the smaller NS2 ORF and the putative second exon encoding the C-terminal region of NS1 (Fig. 2B) [11]. We reported that the putative elongated NS1 resulting from such a spliced transcript in HBoV encodes a 775–781 aa protein with a carboxyl terminus that has weak similarity (25–34% aa identity) to that of other, non-human bocaviruses. Interestingly, the putative NS protein of GBoV1, which is predicted to arise from an intact (non-spliced) transcript, has higher protein similarity (38–74% aa identity) to the extended NS protein (post splicing transcript) of all HBoVs (Fig.2C). Moreover, although no stop codon was found in NS gene of GBoV1, the elements required for RNA splicing were present GBoV1 as in all HBoV species (Fig. 2C, splice donor, branch and acceptor site). Analysis of recently reported novel HBoV species (HBoV3) showed it as a recombinant virus of HBoV1 and HBoV2, suggesting interspecies recombination among bocaviruses. We compared genome wide distances among HBoV1, HBoV2 and GBoV1 genomes using Symplot analysis [31] but found no evidence of interspecies recombination. However, comparison including HBoV3 and HBoV4 genomes indicates frequent recombination between all primate species (data not shown).

Discussion

We report the identification and near complete genomic sequencing of the first non-human primate bocavirus. As it was found in stool samples of gorillas, we have provisionally named this virus GBoV1. Following ICTV guidelines we propose classification of this novel virus as the prototype member of a new species in the genus Bocavirus [1]. The pan-PCR approach used to discover GBoV1 was similar to that recently used to identify HBoV3 and HBoV4 [11].

Several studies report higher prevalence of HBoV1 in respiratory sample and of HBoV2, 3 and 4 in gastrointestinal samples, suggesting that different species may vary in tissue tropism [7,9,10,11,12,13,14,16,19,21,22,33,34,35,36,37,38,39,40,41]. We do not know whether GBoV1 shares this propensity to tissue tropism. Comparative genetic analysis of different HBoV strains circulating worldwide indicate that whereas HBoV1 is homogenous with low mean genetic diversity (<1% nt or <0.5% aa), HBoV2 is heterogeneous with higher mean genetic diversity (<4.6% nt or <6.1% aa) [11]. Current data do not allow us to determine the genetic diversity of GBoV1. However, the discovery of GBoV1 indicates that this recently discovered group of viruses are more widely dispersed than previously reported and raises the possibility that more novel viral species may be found in other animal species.

A unique feature of the GBoV1 genome was the presence of a longer NS protein ORF compared to that encoded by most closely related HBoV. Significant and high similarity (58 to 74% aa identity) of the extended NS protein of all HBoV species with GBoV1 NS protein supports expression of two NS proteins of variable length in HBoVs [11]. The proposed NS1 RNA splicing and NS1 protein expression itself were detected using Northern blots and NS1 C-termini specific sera, respectively, in 293 and human epithelial cells transfected with plasmids expressing HBoV1 transcripts [42]. Despite their high genetic diversity,
similar to all HBoV species, the GBoV1 genome also has perfectly conserved NS gene RNA splicing sites, suggestive of expression of two different NS proteins in GBoV. However it is unclear whether only HBoV requires expression of a truncated NS protein and whether the presence of a truncated NS protein is advantageous for HBoV replication in humans.

Parvoviruses have the capacity to change host species [2,43]. The observation that GBoV1 is more closely related to HBoV1 than any other two HBoV species are to one another, suggests the possibility of cross-species transmission. Single-stranded DNA parvoviruses have been shown to have a mutation rate approaching that of RNA viruses [44,45]. Moreover, both the expression of a truncated NS protein (presence of stop codon in all HBoV species) and the presence of splicing mechanism that generates a longer NS protein is unique to all HBoV species, and suggests the derivation of all HBoV species from a common ancestor. We conclude that although interspecies transmission of bocaviruses may have occurred in ancient times, the genomic organization of GBoV1 indicates that there has been species specificity of bocaviruses in primates for a long time. In absence of any known animal model for HBoV, identification of GBoV1 and other non-human primate bocaviruses will offer opportunity to study the HBoV pathogenesis and disease.

Author Contributions
Conceived and designed the experiments: AK WIL. Performed the experiments: AK NM FE PLQ NQ. Analyzed the data: AK NM FE MPP ELD WIL. Contributed reagents/materials/analysis tools: AK NM MPP PLQ NQ ELD WIL. Wrote the paper: AK FE WIL.

References


