A Newly Identified Bocavirus Species in Human Stool

Amit Kapoor,1,2 Elizabeth Slikas,2 Peter Simmonds,3 Thaweesak Chieochansin,2 Asif Naem,3 Shahzad Shaukat,4 Muhammad Masroor Alam,4 Salmaan Sharif,4 Mehar Angez,4 Sohail Zaidi,4 and Eric Delwart1,2

1Blood Systems Research Institute and 2Department of Laboratory Medicine, University of California, San Francisco, San Francisco; 3Center for Infectious Diseases, University of Edinburgh, Edinburgh, United Kingdom; 4Department of Virology, National Institute of Health, Islamabad, Pakistan

Viral metagenomic analysis was used to identify a previously uncharacterized parvovirus species, “HBoV2,” whose closest phylogenetic relative is the human bocavirus (HBoV). HBoV2 has a genomic organization identical to that of HBoV but has only 78%, 67%, and 80% identity, respectively, with the latter’s NS1, NP1, and VP1/VP2 proteins. The study used polymerase chain reaction to detect HBoV2 sequences in 5 of 98 stool samples from Pakistani children and in 3 of 699 stool samples from Edinburgh. Nearly-full-length genome sequencing revealed the presence of 3 HBoV2 genotypes and evidence of recombination between genotypes. Further studies are necessary to identify anatomical sites of HBoV2 replication and potential associations with clinical symptoms or disease.

The Parvoviridae family includes the Parvovirinae subfamily, which infects vertebrates, and the Densovirinae subfamily, which infects arthropods. Parvovirinae-subfamily members known to infect humans consist of parvovirus B19 (in the Erythrovirus genus) [1]; the aapathogenic adeno-associated viruses (in the Dependovirus genus); the human bocavirus (HBoV) (in the Bocavirus genus); and parvovirus 4 (PARV4) (no genus assigned) [2, 3]. HBoV was discovered in respiratory samples from children [3] but also has been found in stool samples from children with gastroenteritis [4–10]. The association between HBoV and respiratory infections and gastroenteritis are areas of active research that have been the subject of recent reviews [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].

Multiple genotypes of both B19 and PARV4 have been detected [17–21]. Both B19 and PARV4 have also been found archived in lymphoid tissues years after resolution of primary infection plasma viremia [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 nonspecific viral-particle purification and random nucleic-acid amplification combined with minimal sequencing, we identified in a stool sample from a Pakistani child, a previously uncharacterized human parvovirus that is related to HBoV. We report here the results of genomic and phylogenetic analysis of this newly identified parvovirus, along with the results of an initial investigation of its prevalence and genetic diversity.

Patients, materials, and methods. Stool samples from 57 Pakistani children with nonpolio acute flaccid paralysis (AFP) (mean age, 54.6 months) and from 41 healthy Pakistani children (mean age, 39.8 months) were analyzed. Samples were collected as part of the World Health Organisation’s poliovirus-eradication program. A total of 699 stool samples from a mixed-age population that had been submitted for enteric bacteriology screening in Edinburgh were also analyzed.

Stool supernatants were processed for viral metagenomic analysis, as described elsewhere [25], with minor modifications (see the Supplemental Methods section in appendix A, which is available only in the electronic edition of the Journal). Random polymerase chain reaction (PCR) was performed on DNA and RNA extracted from nuclease-resistant virus-sized particles, the products were subcloned, and the plasmid inserts were sequenced. The resulting sequences were compared with entries in the GenBank database, by use of tBLASTx. The viral 5’ and 3’ extremities of HBoV2 were amplified by use of a modification of RACE (rapid amplification of cDNA ends) [2]. (Conditions for PCR for HBoV and HBoV2 are described in the Supplemental Methods section in appendix A, which is available only in the electronic edition of the Journal).

Sequence distances of different genomic regions were measured by use of the built-in functions in the Simmonics2005 sequence editor (version 1.6) [26]. Trees were constructed on the basis of pairwise nucleotide and amino-acid sequence distances, by the
neighboring group was calculated by bootstrap resampling of 1000 replicates of the data. GenBank accession numbers are FJ170278–FJ170285. All studies were performed with the approval of the University of California, San Francisco, committee on human research.

Results. Virus-sized particles were purified from 2 consecutive stool samples from a Pakistani child with AFP. Nuclease-resistant (i.e., capsid-protected) viral nucleic acids were then extracted and randomly amplified, as described above. The resulting random amplification products were subcloned, and 97 plasmid inserts were sequenced and analyzed by use of tBLASTx.

In GenBank, exact sequence matches were found to human sequences, as well as to Micrococcus luteus, Pseudomonas fluorescens, and uncultured bacterium sequences. Also found were highly significant but imperfect matches (expectation value [E], <10^-10) to Chlamydia pneumoniacae, Rhodospirax ferrireducens, and numerous bacteriophages. A single perfect sequence match to human poliovirus 1 vaccine strain Sabin 1 was found. The detection of a polio Sabin 1 viral sequence likely reflects ongoing replication of orally administered polio vaccine in this 36-month-old child who previously had received a total of 14 polio vaccinations.

A total of 11 of 47 plasmid sequences from the first stool sample, as well as 19 of 48 plasmid sequences from the second stool sample, had tBLASTx E scores to the HBoV-genome reference sequence (NC_007455) that were highly significant. PCR was then used to link the different HBoV-like fragments, and 5' RACE and 3' RACE [2] were used to amplify the viral extremities. A total of 5196 bases of a newly identified bocavirus genome were assembled. On the basis of the ST2 prototype genome sequence of HBoV, we estimate that at least 7 bases are missing from the 5' end of HBoV, whereas the genome described in the present report extends for a further 25 bases at the 3' end. Because the closest genetic relative of this newly identified virus is HBoV, we named it "HBoV2."

The arrangement of open reading frames (ORFs) in the prototypic HBoV2 genome was similar to that in HBoV, with 3 large coding sequences (see figure B1 in appendix B, which is available only in the electronic edition of the Journal). The first 5' ORF, NS1, is required for both viral DNA replication and regulation of viral-gene expression; its protein-sequence identity with HBoV was found to be 78% and colinear throughout the gene. The second ORF, NP1, was 4 aa shorter and 67% identical to that of HBoV; NP1 is a protein of unknown function and is restricted to bocaviruses [27, 28]. The third ORF, VP1/VP2, encoded a protein with 80% identity to the VP1/VP2 protein of HBoV. Relative to that of the HBoV coding sequence, the VP1 of HBoV2 was preceded by a 25-aa methionine-initiated ORF stretch and a 4-aa deletion downstream, resulting in a slightly larger VP1. By comparing it with that of HBoV, we predicted that the VP2 protein of HBoV2 starts 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 identified, in the literature, the PCR primers used for HBoV and aligned them with the homologous regions of HBoV2 (see table B1 in appendix B, which is available only in the electronic edition of the Journal). Most of these PCR primers contained a substantial number of mismatches with HBoV2, mismatches that would preclude or greatly reduce the efficiency of amplification and amplicon detection.

To determine the relationship between HBoV2 and other members of the Bocavirus genus, phylogenetic analyses of the 3 large ORFs—VP1/VP2, NP1, and NS1—were performed, by use of both nucleotide sequences and deduced protein sequences (figure 1). In all 3 genomic regions, the prototype HBoV2 variant (PK5110), although more closely related to HBoV than are the animal bocaviruses canine minute virus (CaMV) and bovine parvovirus 1 (BPV1), consistently occupied an outlier position with regard to the clade containing all the published HBoV sequences. Reflecting this, pairwise nucleotide distances between HBoV2 and HBoV were substantially greater (22%–26%) than those within the HBoV clade (0.4%–0.9%) (table 1) but were less than those between either the HBoV or the HBoV2 sequence and the animal bocaviruses (46%–56%). Similarly, although almost no amino acid sequence variability was observed among HBoV sequences in any genomic region (0.2–0.5%), the sequences of all 3 major genes differed substantially (20%–33%) from those of HBoV2. An unusual partial NS1 HBoV sequence (EF560212) has been reported in Brazil and, by phylogenetic analysis, has been found to occupy an intermediate position between HBoV and HBoV2 (figure 1, gray-shaded square in the “Partial NS1” trees); 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-primer PCR specifically targeting the NS region of both HBoV and HBoV2. DNA from stool samples from 57 Pakistani children with AFP and from stool samples from 41 healthy Pakistani children, as well as from 699 stool samples submitted for enteric bacteriology screening in Edinburgh, were analyzed. A total of 3 AFP stool samples (including a sample from the initial source patient, PK5110) and 2 stool samples from healthy Pakistani children were positive for HBoV2 DNA sequences (the ages of these 5 children were 12, 16, 36, 36, and 96 months). HBoV was not found in any of the 57 stool samples from the children with AFP. The stool samples from Edinburgh were tested in pools of 10, and 2 samples positive for HBoV (the ages of the 2 children were 1–2 years and 3–5 years) and 3 samples positive for HBoV2 (the ages of the 3 subjects were 0–3 months, 6–12 months, and >6 years) were identified.

To investigate the genetic diversity of HBoV2, the partial-NS region amplicons from the Pakistani and Edinburgh samples were sequenced and compared both with those of the prototype HBoV2 and with those of other available human and animal
bocavirus sequences (figure 1, “Partial NS1” trees). Although this initial survey was small, the HBoV2 variants were distributed among 3 groups or genotypes (figure 1, “Partial NS1” nucleotide tree) that, in this region of the bocavirus genome, showed 4.5% sequence divergence from each other (table 1). This level of diversity was much greater than the same region’s 0.7% divergence between known HBoV variants. The latter was comparable to the 0.9% mean diversity seen within any 1 of the 3 HBoV2 genotypes.

To further investigate HBoV2 diversity, we obtained nearly complete genome sequences of representatives of the HBoV2 prototype and of the other 2 NS1-based genotypes: genotype 1, HBoV2 prototype PK5510; genotype 2, PK2255; and genotype 3, UK648. Neither of the 2 HBoV2 variants contained the methionine-initiated 25-aa stretch seen upstream of VP1 in prototype PK5510, and the NP1 gene of PK5510 was 1 aa shorter. All 3 genotypes had identical-length NS1 and NP1 genes. As observed for HBoV, protein sequence divergence within HBoV2 was limited, with very low ratios of nonsynonymous changes to synonymous changes, in all 3 genomic regions (table 1). Despite the fact that the HBoV2 genotypes were approximately equidistant at the partial NS1 nucleotide region initially sequenced, phylogenetic relationships varied at other genomic regions (figure 1): at VP1, genotype 1–PK5510 and the genotype 2–PK2255 variant clustered closely, with the genotype 3–UK648 variant being the outlier, whereas, at the NP1 and NS1 regions, the genotype 2–PK2255 variant and the genotype 3–UK648 variant clustered closely, with genotype 1–PK5510 being the outlier (figure 1). Sliding-window analysis of pairwise nucleotide distances across the HBoV2 genome showed that at the NS1 and NP1 regions (represented by the red line in figure B2 in appendix B, which is available only in the electronic edition of the Journal) the genotype 2–PK2255 variant and the genotype 3–UK648 variant had a lower level of divergence than did the other 2 pairwise comparisons, whereas, at the VP1/VP2 region (represented by the green line in figure B2 in appendix B, which is available only in the electronic edition of the Journal) the genotype 1–PK5510 and the genotype 2–PK2255 variant were nearly identical. Discordant phylogenies and inconsistent sequence divergences between HBoV2 variants is consistent with the occurrence of complex recombination events during the evolution of these viruses. Putative breakpoints were located near the middle and 3’ end of NS1 and, potentially more recently (i.e., in recombinants showing a lower level of divergence), near the beginning of VP1/VP2.

Discussion. In view of the 8th report of the International Committee on Taxonomy of Viruses (ICTV), different bocavirus species should show NS-gene nucleotide-sequence similarities of <95%. HBoV2, showing 75.6% nucleotide similarity to
its closest HBoV relative, therefore qualifies as a newly identified human parvovirus and as the fourth identified species in the Bocavirus genus, following BPV, CnMV, and HBoV. We believe that this terminology is more appropriate than its alternative possible name—“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 under the ICTV guidelines.

HBoV2’s divergence also likely precluded its detection by HBoV-based PCR (see table B1 in appendix B, which is available only in the electronic edition of the Journal). 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 HBoV2’s frequency and potential associations with disease.

Because parvoviruses are particularly hard to inactivate by heat and detergent treatments, they are of special concern in blood-product transfusions, and some countries screen for parvovirus B19 DNA, to prevent transfusion of highly viremic blood-derived products [29–32]. The question of whether HBoV2 can result in plasma viremia, as has also been seen with HBoV and PARV4 [13, 33], will require further study. The development of serological assays, as recently achieved for HBoV [13, 33], will allow larger epidemiological studies to measure the rate of seroconversion in cohorts of different ages and geographic regions and to determine whether the presence of anti-HBoV2 IgM and a rise in IgG titers are associated with particular symptoms.

The present study has characterized 3 genotypes of HBoV2 whose genetic distances between each other (in partial NS1 sequences) are comparable to those reported for B19 and PARV4 [16, 22–24]. If HBoV2 is found archived in tissue, as has been shown for B19 and PARV4 [16, 22–24], the possibility of past epidemic waves of different HBoV2 genotypes will also be testable by use of tissues from cohorts of different ages.

The detection of HBoV2 in stool samples from 5 Pakistani children lends further support to the gastrointestinal tract as a

| Table 1. Nucleotide and protein sequence comparisons among and between bocaviruses. |
|-----------------------------------|-------------------|-------------------|-------------------|
| **Comparison, region**            | **Comparisons, no.** | **Pairwise distance, %** | **dN/dS** |
| Primary ORFs                      |                   |                           |           |
| Within HBoV2                      |                   |                           |           |
| VP1/VP2                           | 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/VP2                           | 55                | 0.9                      | 0.5       | 0.075 |
| NP1                               | 55                | 0.4                      | 0.3       | 0.103 |
| NS1                               | 55                | 0.4                      | 0.2       | 0.068 |
| Between HBoV and HBoV2            |                   |                           |           |
| VP1/VP2                           | 33                | 22.2                     | 20.2      | <0.13 |
| NP1                               | 33                | 22.9                     | 32.9      | 0.38  |
| NS1                               | 33                | 26.2                     | 27.8      | <0.19 |
| Between HBoV2 and B/CnPV          |                   |                           |           |
| VP1/VP2                           | 12                | 46.2–46.3                | 51.9–53.2 | NC    |
| NP1                               | 12                | 47.0–52.3                | 59.7–61.3 | NC    |
| NS1                               | 12                | 51.3–55.8                | 60.6–65.1 | NC    |
| 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                      | 6.6       | 0.007 |
| Between HBoV2 and EF560212        | 11                | 9.0                      | 1.1       | 0.003 |
| Between HBoV and HBoV2            | 88                | 18.0                     | 8.3       | 0.071 |

* For comparisons between HBoV and HBoV2, only the upper limits for dN/dS could be calculated, because synonymous distances corrected for multiple substitutions approached or exceeded 1.0 and therefore were likely substantially underestimated. B/CnPV, bovine-canine parvovirus; dN/dS, ratio of nonsynonymous changes to synonymous changes; HBoV, human bocavirus; HBoV2, newly identified parvovirus species; ORF, open reading frame.
site of replication of human bocaviruses [4, 7–10]. The detection of HBoV2 in the stool of 3 UK residents (2 of whom were <12 months old and 1 of whom was >65 years old) indicates that this virus is not restricted to South Asia or to young children. Whether the lower prevalence of HBoV2 in the UK stool samples reflects reduced PCR sensitivity because of sample pooling, the older population tested, reduced exposure to the virus, or reduced duration or level of enteric viremia remains to be determined. Further PCR analysis of blood, cerebrospinal and respiratory fluids, and stool samples from symptomatic and matched healthy controls will also be necessary, to establish whether this virus is associated with respiratory, gastrointestinal, or other symptoms. The equal rate of detection (5%) of HBoV2 in children with nonpolio AFP and in healthy Pakistani children indicates that HBoV2 is unlikely to be associated with AFP.

Acknowledgments

We thank Michael P. Busch and the Blood Systems Research Institute for sustained support.

References