Infection, Genetics and Evolution 28 (2014) 11–14

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Recombination in the evolution of human bocavirus

Alexander I. Tyumentsev, Nina V. Tikunova, Artem Yu. Tikunov, Igor V. Babkin *

Laboratory of Molecular Microbiology, Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia

ARTICLE INFO

Article history:
Received 16 June 2014
Received in revised form 22 August 2014
Accepted 25 August 2014
Available online 2 September 2014

Keywords:
Human bocavirus
Complete genome
Virus evolution
Recombination

ABSTRACT

Whole genome sequencing of Novosibirsk human bocavirus (HBoV) isolates has detected an isolate that emerged via recombination between HBoV3 and HBoV4 genotypes. The recombination site is located between regions with abnormally low and abnormally high GC contents in the genome. This site is a bocavirus recombination hotspot and coincides with one of two parvovirus recombination hotspots. The Novosibirsk recombinant isolate, which is similar to a previously studied isolate from Thailand, utilizes the strategy of borrowing ORF3, which encodes structural proteins, of a rare genotype HBoV4. The role of recombination in HBoV evolution is discussed.

1. Introduction

Human bocavirus (HBoV) belongs to unclassified members of the genus *Bocavirus*, family Parvoviridae. They are small non-enveloped single-stranded DNA viruses that have a genome of approximately 5300 nucleotides. The genome consists of three ORFs; the first two ORFs encode the nonstructural proteins, NS1 and NP1, while ORF3 encodes two structural proteins, VP1 and VP2 (Berns and Parrish, 2007; Chen et al., 2010). All HBoVs are divided into four genotypes, HBoV1–HBoV4 (Chieochansin et al., 2010; Kapoor et al., 2010). It is thought that HBoV1 diverged from the chimpanzee bocavirus approximately 60–80 years ago and that HBoV4 independently separated from great apes bocaviruses approximately 200–300 years ago (Babkin et al., 2013). Most likely, the remaining genotypes emerged as a result of recombination between HBoV1 and HBoV4. In this process, the HBoV2 origin is complicated by additional intergenomic recombination within this genotype (Cheng et al., 2011; Fu et al., 2011; Kapoor et al., 2009, 2010). This finding suggests an important role for recombination in HBoV evolution. Mixed bocavirus infections have been observed in several studies, which suggest the potential for recombination between these viruses (Nawaz et al., 2012).

It is well known that recombination and mutation accumulation in the genome are major mechanisms underlying genetic variation of viruses. Recombination enables viruses to rapidly change their properties. The exchange of genetic material is characteristic of higher organisms, microbes, and viruses and results in rapid evolutionary adaptation, which is not feasible via mutations alone (Martin et al., 2010, 2011).

2. Materials and methods

2.1. Nucleic acid extraction

This study was considered and approved by the Ethical Committee with the State Research Center of Virology and Biotechnology Vector (FWA00000621). Viral DNA was isolated from a stool specimen clarified with PBS–glycerin buffer by affinity sorption using a Ribo-Sorb (Interlabservis) extraction kit according to the manufacturer's protocol.

2.2. HBoV detection

During the genotyping of the bocavirus isolates circulating in Novosibirsk (Russia), a specimen affiliated with the HBoV3 genotype was discovered by PCR using the primers sf2 and sr2, which are able to detect all HBoV genotypes (Cheng et al., 2011; Kapoor et al., 2009). The isolate was genotyped based on sequencing of the HBoV1 ORF1 fragments obtained during the detection procedure with subsequent phylogenetic analysis using the Mega v. 5.2.2 software. The PCR reaction solution (30 µl volume) contained DreamTaq Buffer (ThermoScientific), 0.2 mM of each dNTP, 25 pmol of each primer, 3 units of DreamTaq DNA Polymerase (ThermoScientific), and viral DNA. Amplification was performed using the GeneAmp 9700 amplifier (Applied Biosystems) under the following conditions: initial denaturation, 95 °C, 5 min;

* Corresponding author. Address: Laboratory of Molecular Microbiology, Institute of Chemical Biology and Fundamental Medicine, pr. Lavrentieva 8, Novosibirsk 630090, Russia. Tel.: +7 383 363 5157; fax: +7 383 363 5153.
E-mail addresses: i_babkin@mail.ru, babkin@niboch.nsc.ru (I.V. Babkin).

http://dx.doi.org/10.1016/j.meegid.2014.08.026
1567-1348/

followed by 40 cycles: 95 °C, 20 s; 52 °C, 20 s; and 72 °C, 30 s; and final elongation, 72 °C, 5 min. The amplification products were analyzed by electrophoresis using a 1.5% agarose gel. The size of the target PCR product was 495 bp.

2.3. Detection of other enteric viruses and bacteria

The "RIBO-sorb" Extraction Kit and "Reverta-L" reverse-transcription kit (Central Research Institute of Epidemiology, Russia) were used according to the manufacturer's protocols. "*Rotavirus*/*Norovirus*/*Astrovirus*" PCR kits (Central Research Institute of Epidemiology, Russia) and an "Enterovirus-FL" PCR kit (Central Research Institute of Epidemiology, Russia) were used for the detection of group A rotavirus, norovirus genogroup II, human astrovirus and enterovirus RNA, respectively, according to the manufacturer's protocols. DNA of *Salmonella* spp., *Campylobacter* spp., Shigella spp. and enteroinvasive *Escherichia coli* (EIEC) was purified and detected by PCR using the "RIBO-sorb" Extraction Kit and Ampli-Sens *Shigella* spp. and EIEC/*Salmonella* spp./*Campylobacter* spp.-FL PCR kits (Central Research Institute of Epidemiology, Russia) according to the manufacturer's protocols.

2.4. HBoV sequencing

A complete nucleotide sequence of the HBoV3 isolate RUS_NSC_11-N2512 coding region was determined using a previously described primer set (Babkin et al., 2013). PCR was performed as described above. The amplified fragments were purified using a DNA Gel Extraction kit (Fermentas) and sequenced using the BigDye Terminator v. 3.1 Cycling Sequencing kit (Applied Biosystems). The reaction products were purified on CentriSep (Applied Biosystems) columns and assayed by electrophoresis using a 3500 Genetic Analyzer DNA sequencer (Applied Biosystems). Sequences were analyzed using the MEGA v. 5.2.2 software (Tamura et al., 2011); the determined nucleotide sequence was deposited in GenBank, accession No. KJ710645.

2.5. Sequence analysis

The complete genome sequences used for similarity analysis were extracted from GenBank (http://www.ncbi.nih.gov). Sequence alignment and pairwise nucleotide identities were performed using the BioEdit v.7.0 (Hall, 1999) and Clustal X version 1.8 programs (Thompson et al., 1997). Phylogenetic analysis was performed using the maximum likelihood (ML) method with Mega v. 5.2.2 (Tamura et al., 2011). Permutation analysis of statistical significance for the constructed trees involved 1000 replicates, and the same strategy and parameters were employed. The trees were visualized using Mega v. 5.2.2. The sequence identity matrix was calculated using BioEdit v.7.0 (Hall, 1999). Sequences were tested for recombination breakpoints with RDP, GeneConv, BootScan, SiScan tools (RDP software package) (Martin et al., 2010) and Simplot (Lole et al., 1999). The mfold web server (http://mfold.rit.albany.edu/?q=mfold/DNA-Folding-Form) was used to predict the secondary structure of the NCRs [35] using default parameters.

3. Results

The isolate HBoV3-N2512 was found in a specimen obtained from a male 14-month-old infant hospitalized with acute gastroenteritis in the Infectious Department of the 3rd Novosibirsk Child Hospital on July 6, 2011. This patient did not show any signs of respiratory disease. Importantly, this specimen was assayed for the presence of rotavirus A, norovirus, astrovirus, enterovirus RNA, and *Salmonella* spp., *Campylobacter* spp., *Shigella* and EIEC DNA, and these enteropathogenic agents were not detected.

Next, a complete nucleotide sequence of the HBoV3 isolate RUS_NSC_11-N2512 coding regions was determined. Phylogenetic analysis of the whole genome sequences confirmed that isolate RUS_NSC_11-N2512 belonged to HBoV3 (Supplementary file, Fig. S1). However, analysis of individual ORFs yielded a complex pattern. In the phylogenetic trees constructed based on ORF1 and ORF2, this Novosibirsk isolate displayed considerable similarity to six closely related HBoV3 strains, whereas it clustered with high significance with HBoV4 (Supplementary file, Fig. S1) in the tree constructed using ORF3. For ORF1, the degree of identity between the HBoV3 isolates RUS_NSC_11-N2512 and TU-A-210-07 was 99.5% and was 99.3–99.4% for the remaining five strains of the cluster. For ORF2, the identity of the studied isolate and TU-A-210-07 was 99.6% and 99.5% for the remaining five strains. However, the identity between HBoV3 RUS_NSC_11-N2512 and TU-A-210-07 isolates in the ORF3 sequence was only 89.9%, whereas the highest similarity of the studied isolate was observed as the HBoV4 strain CMH-S011-11 and was 96.4%. This finding suggested a recombination origin of the isolate RUS_NSC_11-N2512.

The recombination breakpoints in the RUS_NSC_11-N2512 isolate were determined using the SimPlot and BootScan methods. This analysis detected with high significance a recombination site near position 3270 in the RUS_NSC_11-N2512 genome. The detected site was localized to positions 3250–3290 (cutoff, 70%); thus, it was located at the beginning of the gene encoding the VP1 protein. Another small region localized to nucleotide positions 4150–4260 in the central part of the gene encoding VP1 protein and displaying a significant similarity to HBoV3 genome was also detected (Fig. 1). Comparison of the obtained and previously published data suggested that the detected recombination breakpoint is a hotspot that is most frequently observed in other bocaviruses (Cheng et al., 2011; Fu et al., 2011; Kapoor et al., 2009, 2010; Khamrin et al., 2013).

To study the nature of the recombination hotspot, we computed genome secondary structures for various bocaviruses using the mfold web server (http://mfold.rit.albany.edu/?q=mfold/DNA-Folding-Form). No noticeable differences in the secondary structure of the examined region from the other viral genome regions were found. Next, we studied the genome sequence and found the region (positions 3133–3223 in the genome of RUS_NSC_11-N2512 isolate) displaying the lowest GC content, which was 20.8%. The downstream region displayed a high GC content, reaching 47.7% in the following 90-nucleotide region (positions 3224–3313). This region included the recombination hotspot at position 3270. Downstream of this hotspot, the region (positions 3272–3291) that displayed the highest GC content was 80.0%. Analyses of 20 whole genome sequences of various HBoV isolates belonging to four genotypes demonstrated similar distribution patterns for the CG content in the region of the recombination hotspot. Presumably, these sequences, first with a low GC content and then with a high GC content, induced a dissociation of DNA polymerase from the genome sequence and resulted in its switching to another template. Recombination in this viral genome region occurred outside the genes encoding NS1 and NP1 proteins, which may also be a consequence of selection.

4. Discussion

The phenomenon of recombination hotspots was observed in various viruses. As previously demonstrated for viruses with a single-stranded DNA genome, their recombination sites were nonrandomly distributed and were prevalently localized either outside genes or in structural genes, which may be due to the effect of