Infection, Genetics and Evolution xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Recombination analysis based on the HAstV-2 and HAstV-4 complete genomes

Igor V. Babkin *, Artem Y. Tikunov, Daria A. Sedelnikova, Elena V. Zhirakovskaia, Nina V. Tikunova

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

ARTICLE INFO

Article history:
Received 25 November 2013
Received in revised form 11 January 2014
Accepted 13 January 2014
Available online xxxx

Keywords:
Astrovirus
Virus evolution
Recombination
Phylogeny

ABSTRACT

Complete genome sequences of previously unstudied human astrovirus subgenotypes – HAstV-2a and HAstV-2c – and two isolates of a rare genotype HAstV-4 have been determined. These isolates were recovered from fecal samples of young children hospitalized with acute intestinal infections in Novosibirsk (Russia). Three of the four sequenced isolates (HAstV-2a, HAstV-2c, and HAstV-4) are recombinants. It has been shown that all known HAstV-2 genomes have emerged via recombination; the HAstV-1 and HAstV-4 genotypes contain both recombinant and non-recombinant isolates; and all HAstV-3, HAstV-5, and HAstV-6 whole-genome sequences display no reliable signs of recombination. The average mutation accumulation rate has been determined based on an extended ORF2 fragment and amounts to 1.0×10^{-3} substitutions per site per year. The evolutionary chronology of current HAstV genotypes has been reconstructed.

1. Introduction

Human astrovirus (HAstV), a member of the *Astroviridae* family, is a rapidly evolving virus with an ss(+) RNA genome. The HAstV genome RNA with a length of about 6.8 kb contains three overlapping open reading frames (ORFs) – ORF1a, ORF1b, and ORF2, encoding serine protease, RNA-dependent RNA polymerase, and capsid protein precursor, respectively (De Benedictis et al., 2011; Mendez and Arias, 2007). The 5′-end is linked to a VPg protein and the 3′-end has a poly(A) tract (Al Mutairy et al., 2005; Fuentes et al., 2012; Mendez and Arias, 2007). Being an agent causing acute gastroenteritis in children and adults, HAstV is spread in all regions around the world (De Grazia et al., 2011; Gabbay et al., 2007; Jeong et al., 2012). Currently, eight HAstV genotypes are known, which are identified by a direct sequencing of an ORF2 fragment; these genotypes differ in their abundance (Belliot et al., 1997; Noel et al., 1995). HAstV-1 is the most abundant, while the remaining genotypes are rather rare. Considerable genetic variation within HAstV-1, HAstV-2, and HAstV-4 genotypes has determined their subdivision into subgenotypes (De Grazia et al., 2011; Gabbay et al., 2007; Martella et al., 2013). However, the genotyping based on a single small ORF2 fragment is unable to give any information about the variation in other regions of the virus genome. Thus, a detailed molecular biological study of this virus requires that whole genomes of various HAstV isolates are sequenced.

A high rate of HAstV evolution is determined by both a high rate of mutation accumulation in the genome, amounting to (2–4) $\times 10^{-3}$ substitutions per site per year (Babkin et al., 2012), and the ability of genome RNA to recombine (Simmonds, 2006; Strain et al., 2008; Ulloa and Gutierrez, 2010). The phylogenies of several HAstV strains based on fragments of various genome regions may significantly differ (Belliot et al., 1997; Walter et al., 2001). Recent analysis of some fragments of HAstV-2c and HAstV-2d genomes has demonstrated their recombinant nature (De Grazia et al., 2012; Wolfaardt et al., 2011). Whole-genome HAstV sequences, including the sequences of rare genotypes, allow for search of recombination sites with high probability. Babkin et al. have studied recombination using whole-genome sequences of HAstV strains to find the regions suitable for estimating mutation accumulation rate in the genomes of this virus (Babkin et al., 2012). In this work, we have studied the role of recombination in the evolution of this virus. In addition, we revised the data on the mutation accumulation rate in HAstV genomes using extended genome fragments and detailed the evolutionary chronology of the current HAstV genotypes.

2. Materials and methods

2.1. Viral nucleic acid extraction and reverse transcription-PCR (RT-PCR)

Virus RNA was extracted from 100 μl of 10% fecal suspension using a RIBO-sorb kit (Interlabservis, Russia) according to the manufacturer's protocol. The cDNA was synthesized with the help of a

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

1567-1348/$ - see front matter © 2014 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.meegid.2014.01.010

REVERTA-L v50 kit (Interlabservis, Russia). The HAstV was detected with an AmpliSens Astrovirus-Eph kit (Interlabservis, Russia). The HAstV-positive specimens were genotyped by RT-PCR and nucleotide sequencing of a 413-bp fragment within the capsid protein precursor region (Noel et al., 1995).

2.2. Genome sequencing

Genomes were sequenced using a primer-walking method. All PCR products were determined in a 3500 Genetic Analyzer DNA sequencer and a BigDye Terminator cycle-sequencing kit (Applied Biosystems). The data were analyzed using the Sequencher program v. 4.0.5 (Gene Codes Corp.). Complete genome sequences used for similarity analysis were extracted from GenBank (http://www.ncbi.nih.gov). Sequence alignment and pairwise nucleotide identities were carried out using the programs BioEdit v.7.0 (Hall, 1999) and Clustal X version 1.8 (Thompson et al., 1997).

2.3. Sequence analysis

Phylogenetic analysis was performed using 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. The trees were visualized with Mega v. 5.2.2. The matrix for HAstV average genetic between-genotype and within-genotype distances was calculated using Mega v. 5.2.2 (Tamura et al., 2011).

Recombination analysis was performed by small sample-corrected Akaike information criterion (c-AIC) (Sugiura, 1978) and Kishino–Hasegawa (KH) test (Kishino and Hasegawa, 1989) using the GARD (Genetic Algorithms for Recombination Detection) program, available through a web interface (Pond et al., 2006). The KH test estimates the variance of the difference between log likelihood scores of two phylogenetic trees. In each case, the nucleotide substitution model was automatically selected prior to being applied to the recombination site analysis. The algorithm scans alignments for evidence of discordant phylogenetic signals. The results provided evidence of recombination breakpoints with score improvements in c-AIC. The predicted breakpoints were located in the alignments according to their scores, and the confirmatory values were evaluated after considering their proximity to previously suspected recombination sites. A comprehensive analysis of the recombination events in Novosibirsk isolates was performed using the Bootscan program from the RDP software package (Martin et al., 2010) with the following settings: window size, 200; step size, 20; number of bootstrap replicates, 500; and cutoff percentage, 70. Additionally, the sequences were tested for recombination breakpoints with RDP, GENECONV, Bootscan, SiScan tools (RDP software package), and Simplot (Lole et al., 1999).

The coding nucleotide sequences of HAstV strains were searched for the sites susceptible to adaptive selection using HyPhy from Mega v. 5.2.2 (Tamura et al., 2011).

The rates of nucleotide substitution per site and the times to the most recent common ancestor were estimated using the Bayesian Markov chain Monte Carlo (MCMC) method, available in the BEAST package (Drummond and Rambaut, 2007). A relaxed (uncorrelated lognormal) molecular clock was utilized for each dataset (Lemey et al., 2010). Sufficient MCMC chains were run to ensure convergence with an initial 10% of the MCMC chains discarded as burn-in. Statistical uncertainty around the mean estimates was provided by 95% highest probability density values. Finally, BEAST analysis also enabled us to infer maximum clade credibility trees for each dataset.

3. Results

3.1. Whole-genome sequencing

Since 2003, detection and genotyping of various viruses in clinical fecal samples from young children were performed in Novosibirsk. The detected HAstV isolates have been genotyped by sequencing the PCR fragments from ORF2 according Noel et al., 1995. Isolates HAstV-2a Rus-Nsc05-430, HAstV-2c Rus-Nsc06-1029, HAstV-4 Rus-Nsc10-N358, and HAstV-4 Rus-Nsc05-623 recovered in 2005, 2006, 2010, and 2005, respectively, were selected for whole-genome sequencing (Fig. 1).

Totally, 29 original primers were designed for sequencing (Table S1). This allowed whole-genome sequences of isolates HAstV-2a Rus-Nsc05-430, HAstV-2c Rus-Nsc06-1029, HAstV-4 Rus-Nsc10-N358, and HAstV-4 Rus-Nsc05-623 to be determined; the sequences were submitted in GenBank database under accession numbers KF039910, KF039911, KF039913, and KF039912, respectively. These sequences are the HAstV genomic sequences

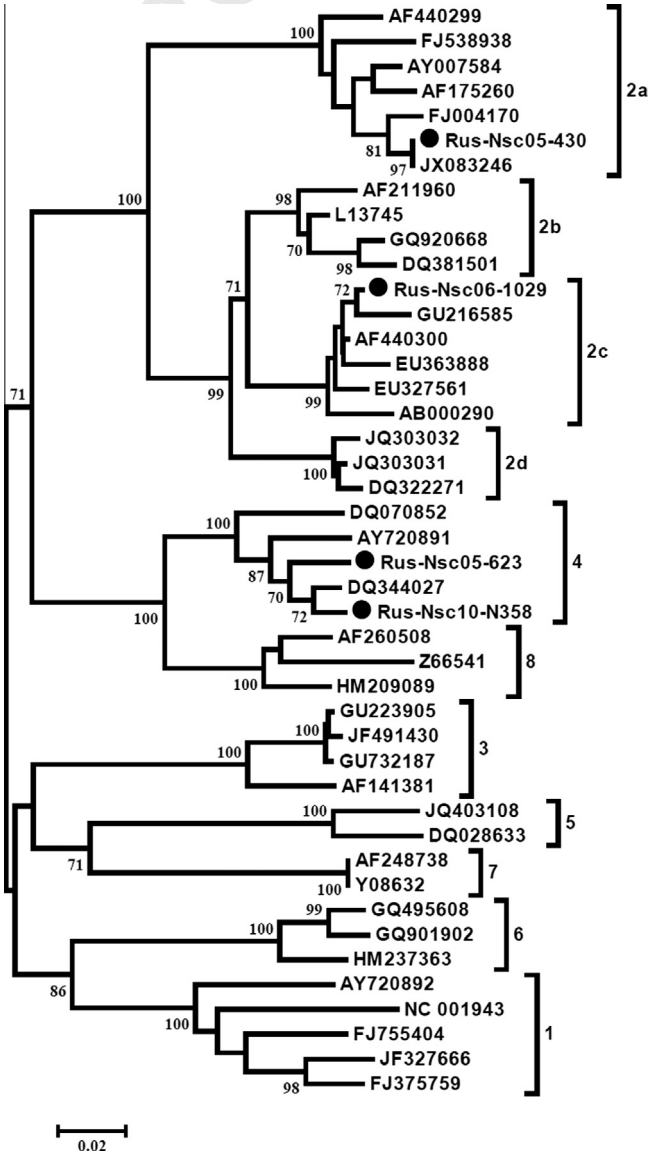

Fig. 1. ML phylogram for nucleotide sequences of the genotyping fragment of HAstV isolates. Statistical supports are shown at nodes; black circles denote the sequences determined in this work.

Please cite this article in press as: Babkin, I.V., et al. Recombination analysis based on the HAstV-2 and HAstV-4 complete genomes. Infect. Genet. Evol. (2014), http://dx.doi.org/10.1016/j.meegid.2014.01.010