Infection, Genetics and Evolution 20 (2013) 330–335

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Lineage diversification and recombination in type-4 human astroviruses

Vito Martella [a], Maria C. Medici [b], Valentina Terio [a], Cristiana Catella [a], Giancarlo Bozzo [a], Fabio Tummolo [b], Adriana Calderaro [b], Floriana Bonura [c], Marzia Di Franco [c], Krisztiàn Bányai [d], Giovanni M. Giammanco [c], Simona De Grazia [c,*]

[a] Dipartimento di Medicina Veterinaria, Università Aldo Moro di Bari, Valenzano, Italy
[b] Dipartimento di Medicina Clinica e Sperimentale, Unità di Microbiologia e Virologia, Università di Parma, Parma, Italy
[c] Dipartimento di Scienze per la Promozione della Salute e Materno Infantile "G. D'Alessandro", Università di Palermo, Palermo, Italy
[d] Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary

ARTICLE INFO

Article history:
Received 9 July 2013
Received in revised form 12 September 2013
Accepted 12 September 2013
Available online 29 September 2013

Keywords:
Astrovirus
Epidemiology
Genotyping
Viral gastroenteritis
Italy

ABSTRACT

Human astroviruses (HAstVs) are important enteric pathogens and can be classified genetically and antigenically into eight types. During surveillance of HAstVs in Italy, type-4 HAstVs were detected only sporadically and found to cluster into two distinct genetic groups. Upon sequence analysis of the 3′ end of the polymerase gene (ORF1b) and of the full-length ORF2, the 2008 type-4 HAstV strains were characterised as a novel ORF2 genetic lineage, designated as 4c. The 2008 type-4 HAstVs also shared the ORF1b gene with similar HAstV-4c strains detected globally, thus displaying a conserved ORF1b/ORF2 asset. By interrogation of the databases, this novel lineage 4c accounted for 60.8% of the type-4 strains identified worldwide and the vast majority of recent type-4 HAstVs. The 2002 type-4 HAstVs displayed a type-4b ORF2, whereas in the ORF1b they resembled type-1 HAstVs. This inconsistency suggests a possible recombinant origin, with the RNA switch taking place upstream the ORF1b/ORF2 junction region. Also, recombination likely played a role in the diversification of the ORF2 of the three type-4 lineages. Multi-target analysis is required for appropriate characterisation and identification of recombinant HAstVs.

1. Introduction

Astroviruses (AstVs), family *Astroviridae*, are enteric viruses associated with enteric and extra-intestinal diseases in several mammalian and avian species (Mendez and Arias, 2007). AstVs have a single-stranded positive sense RNA genome containing three open reading frames (ORFs). ORF1a and ORF1b, at the 5′ end of the genome, encode the non-structural viral proteins, including the RNA-dependent RNA-polymerase, while ORF2, at the 3′ end, encodes the capsid protein precursor. Human AstVs (HAstVs) are a major cause of gastroenteritis in young children, elderly people and immunocompromised adults (Mendez and Arias, 2007). By sequence comparison of HAstVs, the capsid protein precursor can be divided into a highly conserved N-terminal domain (aa 1–424), a hypervariable domain (HVR) (aa 425–688) and a highly acidic C-terminal domain (Wang et al., 2001). The mature infectious virion contains three predominant protein species derived from the N-terminal domain and from the HVR after intra- and extra-cellular processing (Bass and Qiu, 2000; Sanchez-Fauquier et al., 1994). Mapping of neutralising monoclonal antibodies (Sanchez-Fauquier et al., 1994) and structural analysis have revealed that AstV capsid spike is formed by dimerisation of a polypeptide spanning the HVR and a putative binding receptor site with affinity for polysaccharide molecules has been predicted in the spike structure (Dong et al., 2011).

Early in the study of HAstVs, marked antigenic differences were noted by cross-neutralisation among some HAstV isolates (Kurtz and Lee, 1984). Subsequent studies in immune electron microscopy, immunofluorescence, ELISA and plaque neutralisation assays (Hudson et al., 1989; Jakab et al., 2004; Koopmans et al., 1998) revealed that the extent of HAstV antigenic diversity was greater, allowing for distinction of eight serotypes, HAstV-1 to -8. Sequence analysis of short fragments at either the 5′ or 3′ end of ORF2 (D5′ and D3′ regions) and RT-PCR genotyping protocols with type-specific primers have been used for genetic characterisation of HAstV-1 to -8 (Jakab et al., 2004; Mustafa et al., 2000; Noel et al., 1995). In addition, upon molecular analysis of the ORF2, discrete sequence variation has been observed within some HAstV types, allowing for distinction of genetic lineages, with two distinct genetic lineages (4a and 4b) being described in type-4 HAstVs (Colomba et al., 2006; De Grazia et al., 2011; Gabbay et al., 2007; Guix et al., 2002; Victoria et al., 2007).

* Corresponding author. Address: Dipartimento di Scienze per la Promozione della Salute e Materno Infantile "G. D'Alessandro", Università di Palermo, Via del Vespro 133, 90127 PA, Italy. Tel./fax: +39 0916553663.
E-mail address: simona.degrazia@unipa.it (S. De Grazia).

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.09.015

Surveillance studies for HAstV in Italy identified the circulation of type-4 HAstV strains in 2002 (De Grazia et al., 2011) and again in 2008 (De Grazia et al., 2012, 2011; Medici et al., 2012). Based on a small sequence generated by the diagnostic primers Mon269–Mon270 (Noel et al., 1995) in the D5′ region, the 2008 Italian type-4 HAstVs were clearly distinguishable by phylogenetic analysis from the 2002 type-4 HAstVs (De Grazia et al., 2011). In order to investigate further the extent of the genetic heterogeneity observed in type-4 HAstV strains, a 3.2 kb portion at the 3′ end of the genome was sequenced for two representative strains (ITA/2008/BA393/08-65 and ITA/2002/PA73) and compared with type-4 HAstVs retrieved from the databases. The 2008 type-4 HAstVs differed markedly from the older type-4 HAstVs and were classified within a novel, yet unrecognised, type-4 lineage, along with similar strains detected globally. By converse, the 2002 type-4 HAstVs were found to have a recombinant ORF1b derived from type-1 HAstVs.

2. Materials and methods

2.1. Samples origin

During surveillance activity on viral gastroenteritis conducted in Italy, two different lineages of type-4 HAstVs were identified by sequence analyses of 348-nucleotide (nt) portion at the 5′ end of ORF2. The prevalence of HAstV circulation ranged from 4.2% to 7.4%. In particular, the analyses of stool samples of children, aged less than 5 years, hospitalised with acute gastroenteritis revealed that three of the five HAstVs detected in Palermo in 2002 and three of the 32 HAstVs detected in Bari and Parma in 2008 were type-4 HAstV strains. The 2002 type-4 HAstVs markedly differed genetically from the 2008 viruses, falling into two distinct lineages. Two samples, ITA/2002/PA73 and ITA/2008/BA393/08-65, were selected as representatives of these lineages for further analysis.

2.2. RNA extraction and amplification

Viral RNA was extracted from 140 μl of stool suspension using the QIAmp viral RNA kit (Qiagen, GmbH, Hilden, Germany). A 3′ RACE-PCR protocol (Wang et al., 2005) was used to generate a 3.2 kb amplicon encompassing the 3′ end of ORF1b, the full-length ORF2, the 3′ untraslated region (UTR) through the poly-A tail. Briefly, cDNA was synthesised by SuperScript III First-Strand cDNA synthesis kit (Invitrogen Ltd, Paisley, UK) with primer VN3T20 (5′-GAGTGACCGCGGCCGCT^{20}-3′). PCR was then performed with TaKaRa La Taq polymerase (TaKaRa Bio Europe SAS, Saint-Germain en-Laye, France) with forward primer panAstVFor1 (GARTTYGATTGGRCKCGKTAYGA) and the reverse primer VN3T20 (Chu et al., 2008).

2.3. Sequence and phylogenetic analyses

The amplicons were purified and cloned using TOPO XL Cloning Kit (Invitrogen Ltd, Paisley, UK). Additional primers were designed to determine the complete 3.2-kb sequence by an overlapping strategy. Sequence editing and multiple codon-based (translation) alignments were performed with Geneious software v6.2 (Drummond et al., 2011). Phylogenetic analysis was conducted by using Genious software (Drummond et al., 2011). MrBayes version 3.2.1 (Huelsenbeck and Ronquist, 2001) was used to infer Bayesian analysis. Species Delimitation software (Masters et al., 2011) was used to assess clade monophyly and robustness of the phylogenetic analysis. SimPlot software (version 3.2) (Lole et al., 1999) was used to identify cross-over sites due to recombination. The accession number of the strains ITA/2002/PA73 and ITA/2008/BA393/08-65 are KC915035 and KC915034, respectively. A total of 19 full-length ORF2 sequences of type-4 HAstV were available in the databases. The sequences with their accession numbers are listed in Fig. 1A. In addition, a total of 54 partial (~350 bp) ORF2 sequences of type-4 HAstV strains spanning the D5′ region (nt 4573–4920 of prototype strain AY720891/DEU/2004/Dresden) and 6 partial (~200 bp) ORF2 sequences of type-4 HAstVs spanning the D3′ region (nt 6427–6642 of prototype strain AY720891/DEU/2004/Dresden) were selected from the databases.

3. Results

A ~750-nt-long fragment of ORF1b was sequenced for strains ITA/2002/PA73 and ITA/2008/BA393/08-65. The full-length ORF2 of both viruses was of 2316 nt in length and the 3′ UTR was 81 nt long. There was a 8-nt overlap between the 3′ end of ORF1b and the start of ORF2. The highly conserved nt stretch 5′-ATTTGGAGNGGNGGACCNAAN$^{5-8}$ATGNC-39, believed to be part of a promoter region for synthesis of subgenomic RNA, was retained upstream the ORF2 start codon (Walter et al., 2001).

3.1. Sequence analysis of the ORF2 of type-4 HAstVs

Analysis of the full-length ORF2 of the two Italian strains along with a set of 19 type-4 HAstVs retrieved from the databases allowed identifying three distinct genetic lineages, 4a–4c (Fig. 1A). The clades were monophyletic, with high strict and liberal probability values and high Bayesian posterior probability values, as defined using Species Delimitation software. Five strains clustered in lineage 4a (prototype AY720891/DEU/2004/Dresden), four strains in lineage 4b (prototype strain DQ070852/BRA/1995/Goiania/GO/12) and 12 strains in lineage 4c (prototype strain DQ344027/CHN/2005/Guangzhou). In the ORF2, the nt identity between strains of different lineages ranged between 89.1% and 93.5% while the nt identity within each lineage was not lower than 93.8%. The strain ITA/2002/PA73 displayed the highest nt identity (97.3–98.1%) to type-4b HAstV strains while the strain ITA/2008/BA393/08-65 was most similar to type-4c HAstV strains (95.8–98.2%). Nucleotide (difference) cut-off values of 6.5% were found to determine individual lineages in the full-length ORF2. These values are similar to values calculated in the D5′ region in other studies (Gabbay et al., 2007; Guix et al., 2002).

The deduced amino acid (aa) sequence of the capsid precursor was determined and an alignment generated to assess the rate of aa variation across the various ORF2 regions (Walter et al., 2001). The aa variation within type-4 HAstVs reached 6.8% in the N-terminal domain (aa 1–424), 8.1% in the HVR hypervariable region (aa 425–688) and 23.7% in the highly acidic C-terminal domain. In the HVR, intra-lineage aa variation reached 3.4%, 2.7% and 7.2% for 4a, 4b and 4c strains, respectively, while inter-lineage aa variation ranged between 5.7% and 8.1%. The cell-adapted strain GBR/Oxford-S4 (accession AB000296) presented a nucleotide deletion at the end of the HVR, which altered the frame of the capsid precursor throughout the C-terminal domain. For this strain, aa variation was higher, reaching values of 7.8%, 14.0% and 85.5% in the N-domain, HVR and acidic domain of the capsid precursor, respectively.

3.2. Recombinant origin of strain ITA/2008/BA393/08-65

In the ORF1b, strain ITA/2008/BA393/08-65/type4c displayed the highest nt identity (96.5%) to the type-4c HAstV strain CHN/2005/Guangzhou (accession DQ344027). On the opposite, strain ITA/2002/PA73/type4b displayed the highest nt identity (97.6–97.8%) to Indian and Chinese type 1 HAstVs (accessions