Infection, Genetics and Evolution 32 (2015) 156–160

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Short communication

# Genetic heterogeneity and recombination in type-3 human astroviruses

Maria Cristina Medici [a,*], Fabio Tummolo [a], Vito Martella [b], Krisztián Banyai [c], Elisabetta Bonerba [b], Carlo Chezzi [a], Maria Cristina Arcangeletti [a], Flora De Conto [a], Adriana Calderaro [a]

[a] Unit of Microbiology and Virology, Department of Clinical and Experimental Medicine, University of Parma, Parma, Italy
[b] Department of Veterinary Medicine, University Aldo Moro of Bari, Valenzano, Italy
[c] Veterinary Medical Research Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary

ARTICLE INFO

*Article history:*
Received 9 December 2014
Received in revised form 2 March 2015
Accepted 9 March 2015
Available online 14 March 2015

*Keywords:*
Astrovirus
Genotyping
Recombination
Viral gastroenteritis
Italy

ABSTRACT

Human astroviruses (HAstVs) are important enteric pathogens and can be classified genetically and antigenically into eight types. During molecular surveillance for HAstVs in Italy, sequence analysis of the diagnostic region C (about 400 nucleotide in length), located on the capsid (ORF2) gene, identified a novel type-3 strain. Upon sequencing of the full-length ORF2, the type-3 HAstV strain was characterized as a novel ORF2 genetic lineage, designated as 3c. By converse, in the ORF1b the virus was more similar to type-1 HAstVs, rather than to type-3 strains, suggesting a recombination nature, with the crossover site being mapped to the ORF1b/ORF2 junction region. Region C sequences of similar type-3 HAstV identified from European and extra-European countries were retrieved in the databases, suggesting the global distribution of this novel type-3 lineage.



## 1. Introduction

Human astroviruses (HAstVs), genus *Mamastrovirus*, family *Astroviridae*, are important etiological agents of gastroenteritis in humans, mostly in young children, elderly people, and immunocompromised patients (Mendez and Arias, 2013). HAstVs have a single stranded positive-sense RNA of about 6.8 kb in length that contains three overlapping open reading frames (ORFs). ORF1a, ORF1b, and ORF2 encode the serine protease, the RNA-dependent RNA polymerase, and the capsid protein precursor, respectively. The capsid protein precursor can be divided into a highly conserved N-terminal domain (amino acids [aa] 1–424), a hypervariable (HVR) domain (aa 425–688), and a highly acidic C-terminal domain (Wang et al., 2001). The HVR domain of HAstV is believed to form the capsid spike and to control binding to cell receptors as neutralizing epitopes have been mapped inside this capsid portion (Dong et al., 2011).

Studies based on immune electron microscopy, immunofluorescence, ELISA and plaque neutralization assays have revealed that HAstVs are antigenically heterogeneous and that they can be classified into eight serotypes (HAstV-1 to HAstV-8). (Bosch et al., 2014; Mendez and Arias, 2013; Kurtz and Lee, 1984). Sequence analysis of short fragments at either the 5′ or 3′ end of ORF2 (regions C and D) and RT-PCR genotyping protocols with type-specific primers have been used for genetic characterization of HAstV-1 to -8 (Belliot et al., 1997; Noel et al., 1995). HAstV-1 appears to be the predominant circulating serotype worldwide followed by types 2–5 and occasionally by type-8, depending on the region. HAstV-6 and -7 are rarely detected (De Grazia et al., 2012; Gabbay et al., 2007; Guix et al., 2002; Liu et al., 2008; Medici et al., 2012; Mendez-Toss et al., 2004; Mustafa et al., 2000).

Also, sequence analysis of the small diagnostic regions located on ORF2 has revealed discrete sequence variation within some HAstV types, revealing intra-typic genetic lineages. Type-1 HAstV has been classified into four lineages (HAstV-1a to -1d), type-2 into four (HAstV-2a to -2d), type-3 into two (HAstV-3a and -3b) and type-4 into three (HAstV-4a to -4c) (De Grazia et al., 2012; Gabbay et al., 2007; Guix et al., 2002; Martella et al., 2013). Comparison of the full-length ORF2 has confirmed the classification of HAstVs into discrete lineages. Cutoff values of 5–6.5% nucleotides (nt) identity in the full-length ORF2 have been calculated among the various lineages. These analyses have also revealed several examples of intra-typic recombination within this genomic region (Martella et al., 2014). Also, examples of recombination in, or close to the ORF1b/ORF2 junction have been described, suggesting that this genomic region may be a preferential site for RNA cross-over (Babkin et al., 2014; De Grazia et al., 2013; Gabbay et al., 2007; Martella et al., 2013; Walter et al., 2001).

* Corresponding author at: Unit of Microbiology and Virology, Department of Clinical and Experimental Medicine, University of Parma, Viale Antonio Gramsci, 14, 43126 Parma, Italy. Tel.: +39 0521 033495/033499; fax: +39 0521 993620.
*E-mail address:* mariacristina.medici@unipr.it (M.C. Medici).

http://dx.doi.org/10.1016/j.meegid.2015.03.011
1567-1348/

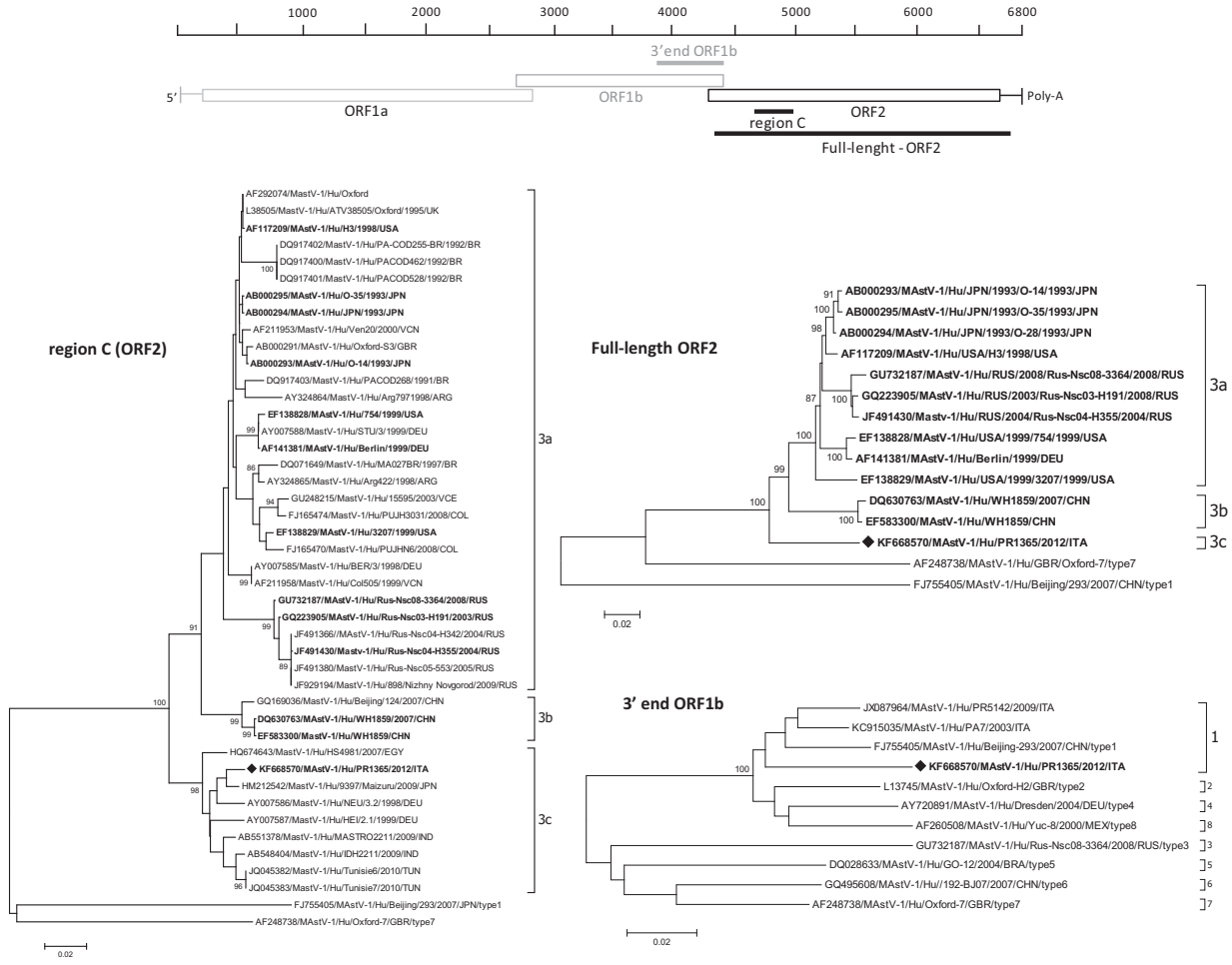


**Fig. 1.** The region C tree was constructed using an ~350-nt ORF2 fragment of the 42 sequences of type 3 HAstV strains available in the databases, of which 13 (highlighted in bold) were from the full-length ORF2 sequences. The full-length ORF2 tree was constructed with the 13 full length ORF2 (2385 nt) of type 3 HAstVs available in the databases. The ORF1b tree was constructed with the partial 3′ end of ORF1b (836 nt) of the HAstV-3 strain PR1365/ITA/2012 and of reference strains retrieved from the databases. Trees were built with the maximum-likelihood (ML) method, and bootstrapped with 1000 repetitions. Bootstrap values >80% are indicated. The scale bar indicates the number of nucleotide substitutions per site. The genome organization of HAstV and the locations of the various genetic targets used for phylogenetic analyses are also shown.

Surveillance for HAstV in Parma, Northern Italy, has been carried out continuously since 2008 in children aged <5 years hospitalized with acute gastroenteritis at the Maternal–Infantile Department of the University Hospital of Parma. All the HAstV strains identified between 2008 and 2013 were systematically genotyped in the diagnostic region C (nt 4571–4918 of accession No. L13745) (De Grazia et al., 2013; Martella et al., 2013; Medici et al., 2012). In this 6-years period, the overall prevalence of HAstV infection was 1.34% (32 out of 2383 children with gastroenteritis), ranging from 3.87% in 2008 to 0.49% in 2011. Five different genotypes were found to circulate in this time span, with type-1 HAstVs being predominant (23 cases, 71.88%), followed by type-2 (4 cases, 12.5%) and type-4 (3 cases, 9.38%), type-3 (1 case, 3.12%) and type-5 (1 case, 3.12%). The type-1 HAstVs were subtyped as lineage 1a (16 cases, 69.56%), 1b (1 case, 4.35%) and 1d (6 cases, 26.09%). All the type-2 HAstVs were characterized as lineage 2d, and all the type-4 HAstVs as lineage 4c, while the type-5 HAstV strain was sub-typed as 5c. Based on the small sequence generated in ORF2 region C, the single HAstV-3 strain, PR1365/2012/ITA, detected in April 2012, segregated within a novel, yet unrecognized, type-3 lineage, along with similar strains detected globally. In order to investigate better the genetic signature of this strain, a 3.2 kb portion at the 3′ end of the genome was sequenced and compared with HAstV-3 strains retrieved from the databases.

## 2. Materials and methods

A 3′ RACE-PCR protocol was used to generate a 3.2 kb amplicon encompassing the 3′ end of ORF1b, the full-length ORF2, the 3′ untranslated region (UTR) through the poly-A tail (Wang et al., 2005). Briefly, cDNA was synthesized by SuperScript III First-Strand cDNA synthesis kit (Invitrogen Ltd, Paisley, UK) with primer VN3T20 (5′-GAGTGACCGCGGCCGCT20-3′). PCR was 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). The amplicon was purified and cloned using TOPO XL Cloning Kit (Invitrogen Ltd, Paisley, UK). A consensus sequence was generated on 3 clones. Additional primers were designed to determine the complete 3.2-kb sequence (corresponding to nt 3566–6787 of the Nsc08/336/2008/RUS/type3, GenBank accession No. GU732187) by an overlapping (primer-walking) strategy. Sequence editing, generation of multiple codon-based (translation) alignments and phylogenetic trees were performed with MEGA version 6.0 (Tamura et al., 2013). Maximum composite likelihood (ML) algorithm and the neighbor-joining method were used for construction of the phylogenetic trees. The reliability of the phylogenetic trees was assessed by bootstrap re-sampling over 1000 replicates. SimPlot version 3.2 (Lole et al., 1999) was used to identify cross-over sites due to recombination. Additionally,