Infection, Genetics and Evolution 58 (2018) 199–208

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Molecular evolutionary analysis of type-1 human astroviruses identifies putative sites under selection pressure on the capsid protein

Simona De Grazia[a,*], Gianvito Lanave[b], Floriana Bonura[a], Noemi Urone[a], Vincenzo Cappa[a], Sara Li Muli[a], Arcangelo Pepe[a], Akos Gellért[c], Krisztian Banyai[c], Vito Martella[b], Giovanni M. Giammanco[a]

[a] Department of Health Promotion Sciences and Mother and Child Care "G. D'Alessandro", University of Palermo, Palermo, Italy
[b] Department of Veterinary Medicine, University of Bari, Valenzano, Bari, Italy
[c] Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Hungária krt. 21, Budapest H-1143, Hungary

ARTICLE INFO

Keywords:
Astrovirus
HAstV-1
Phylogenetic analyses
Capsid protein structure
Selection pressure analysis
Genetic evolution

ABSTRACT

Human astroviruses (HAstV) are important enteric pathogens that can be classified into eight sero/genotypes (HAstV-1 to -8). Although the various HAstV types show global spread, type-1 strains tend to be predominant. Molecular analysis of the genomic region encoding the capsid protein (ORF2) has revealed discrete sequence variation, with different lineages within each HAstV type and at least three major lineages have been identified within HAstV-1. Longitudinal epidemiological surveillance has revealed temporal shift of the various HAstV-1 lineages. Metadata analysis of HAstV-1 sequences available in the databases also revealed temporal shifts of the circulation of HAstV-1 lineages, suggesting possible antigenic-related mechanisms of selection at the sub-genotype level. By comparison of HAstV-1 capsid sequences, lineage-defining residues under positive selection were identified. Structural analysis of HAstV-1 capsid allowed identifying at least six residues exposed on the virion surface. Two residues were located in the VP34 (shell region) whilst four residues were mapped in the VP25/27 (protruding region) of HAstV capsid protein, in proximity of the putative receptor binding S site. These findings suggest that mechanisms similar to those observed and/or hypothesized for other enteric viruses are also shaping the evolution of HAstVs, with intra-typic diversification being a possible mechanism to decrease the antigenic pressure to which these viruses are exposed.

1. Introduction

Human Astroviruses (HAstVs) are considered one of the leading causes of gastroenteritis in young children, elderly people and immunocompromised adults. They are also responsible for intestinal and extra intestinal diseases in a wide range of animals (Mendez & Arias, 2013). There are approximately 3.9 million cases of viral diarrhea due to HAstV in Unites States every year (Mead et al., 1999). The family *Astroviridae* is comprised of two genera, *Mamastrovirus* and *Avastrovirus*, which infect mammalian and avian species, respectively. Astrovirus is an icosahedral virus with a star-like surface structure and a diameter of approximately 28–30 nm (Kovács et al., 2017; Mendez, 2007). Astrovirus possesses a positive sense, single-stranded, polyadenylated RNA genome containing three open reading frames (ORFs), ORF1a, ORF1b and ORF2. ORF1a and ORF1b, at the 5′ end of the genome, encode the non-structural viral proteins, including the RNA-dependent RNA-polymerase, whilst ORF2, at the 3′ end, encodes the capsid proteins precursor.

HAstVs are genetically and antigenically heterogeneous (Koopmans et al., 1998). High sequence diversity has been found in ORF2 whilst the ORF1b is more conserved (Strain et al., 2008). The ORF2 is characterized by a highly conserved N-terminal domain (aa 1–424), a hypervariable domain (HVR) (aa 425–688) and a highly acidic C-terminal domain (Mendez, 2007; Wang et al., 2001). This hypervariable domain is believed to form the capsid spike and to contain neutralizing epitopes (Dong et al., 2011), therefore ORF2 is used for prediction of HAstV types by PCR and sequencing. Moreover, molecular analysis of the ORF2 allowed defining different lineages within each HAstV genotype (Gabbay et al., 2007a; Guix et al., 2005). Currently, HAstVs are divided into eight genetically and antigenically distinct types (HAstV-1 to -8) within the *Mamastrovirus* species 1. Genotyping surveys have revealed that HAstV-1 is the most common type worldwide (De Grazia et al., 2016; De Grazia et al., 2012a; Gabbay et al., 2007b; Guix et al., 2002; Koopmans et al., 1998; Malasao et al., 2012; Wang et al., 2001). The predominance of type-1 HAstVs has raised interest for this genotype

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

https://doi.org/10.1016/j.meegid.2017.12.023
Received 25 October 2017; Received in revised form 6 December 2017; Accepted 22 December 2017
Available online 27 December 2017
1567-1348/

since the mid-1980s, when it was first studied antigenically (Lewis et al., 1994). Longitudinal epidemiological surveillance have revealed temporal patterns of replacements of the various HAstV-1 lineages, suggesting possible antigenic-related mechanisms of selection (Kovács et al., 2017), mirroring what is observed for other viruses (Bok et al., 2009; Suptawiwat et al., 2017).

In order to better understand the evolutionary patterns of HAstV-1, the complete ORF2 sequence of 29 strains retrieved in the databases and an additional five sequences generated in the present study were used for comparative sequence analysis to identify evolutionary mechanisms acting on the capsid protein.

2. Material and methods

2.1. Detection of HAstV for epidemiological surveillance

Uninterrupted surveillance for HAstV has been conducted in Palermo, Italy, since August 1999. Stool samples were collected from children under 5 years of age hospitalized with acute gastroenteritis at the "G. Di Cristina" Children's hospital of Palermo. Viral RNA was extracted from 140 µl of stool suspension using the QIAmp viral RNA kit (Qiagen, GmbH, Hilden, Germany). The RNA extracts were screened for the presence of HAstVs by RT-PCR with specific primers Mon269 and Mon270, amplifying a 348 nucleotide (nt) portion at the 5′-end of ORF2 (Noel et al., 1995).

2.2. Selection and analysis of ORF2 sequences

The entire set of the ORF2 of HAstV-1 available in Genbank (703 partial and 29 complete sequences) was retrieved from the databases using Geneious v9.0 software package (Drummond et al., 2011). Genbank interrogation was terminated on February 2017. Genotype and/or a lineage was attributed to each HAstV sequence using a BLAST algorithm by the software Sequence Classifier version 1.1 (Drummond et al., 2011).

A large fragment of viral genome (3.2 kb in length) encompassing the 3′ end of ORF1b (~0.6 kb) and the full length ORF2 (2.4 kb) of the additional five HAstV-1 sequences, selected for this study, were generated as previously described (De Grazia et al., 2012b; Martella et al., 2014; Walter et al., 2001). The accession numbers of the strains ITA/1999/PA364, ITA/2000/PA762, ITA/2004/PA70R-, ITA/2008/PA148 and ITA/2011/PA387 are KY744138, KY744137, KY744139, KY744140, and KY744141, respectively. The 34 full-length ORF2 sequences (~2.4 kb) of different lineages type-1 strains (-1a, -1b and -1d) available for the study are shown in Table 1.

2.3. Study of HAstV-1 genetic evolution by phylogenetic, Shannon entropy and positive selection pressure analyses

Sequence alignments were performed using CLUSTAL W (Thompson et al., 1994) with Geneious v9.0 software package and MEGA7 software (Drummond et al., 2011; Kumar et al., 2016). The appropriate substitution model settings for the phylogenetic analysis and estimation of selection pressure on coding sequences was derived using jModelTest, based on the least Bayesian Information Criterion (BIC) scores (Posada, 2009). Phylogenetic analysis was conducted in MEGA7 by using the Maximum Likelihood method with the Tamura-Nei model (TN93) (Kumar et al., 2016; Tamura and Nei, 1993). The statistical significance of the phylogenies inferred was estimated by bootstrap analyses with 500 pseudoreplicate datasets. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+ G, parameter = 0.3748)).

The Shannon entropy, used to quantify the diversity in every single amino acid (aa) position of the alignment of the 34 ORF2 complete sequences, was evaluated using Entropy (www.hiv.lanl.gov/cgi-bin/ENTROPY/entropy.html).

Table 1
HAstV-1 strains available in Genbank for the ORF2 complete sequence analyses. The strains generated in this study are indicated in bold.

Accession	Strain name
L23513	Hu/GBR/1993/Oxford-1/type1a
Z25771	Hu/GBR/1993/A2-88Newcastle/type1a
AB000283	Hu/GBR/1997/Oxford-S1/type1a
AB000287	Hu/JPN/1992/J115/type1a
AB009984	Hu/JPN/1992/J153/type1a
AB000284	Hu/JPN/1992/J1526/type1a
EF138823	Hu/USA/1999/1638/type1a
EF138824	Hu/USA/1999/2882/type1a
EF138825	Hu/USA/1999/2987/type1a
KY744140	**HAstV/ITA/PA148/2008/type1a**
KP942585	Hu/CHN/2014/KSMP03/type1a
KP942586	Hu/CHN/2014/KMSP04/type1a
HQ398856	Hu/HUN/2010/Nyergesujfalu/HUN4520/type1a
KY744141	**HAstV/ITA/PA387/2011/type1a**
AB000285	Hu/PAK/1991/PAK437/type1b
KY744137	**HAstV/ITA/PA762/2000/type1b**
FJ755402	Hu/CHN/2005/Beijing-128/type1b
FJ755405	Hu/CHN/2007/Beijing-293/type1b
GQ405856	Hu/CHN/2007/Shenyang/type1b
GQ405855	Hu/CHN/2007/Dalian/type1b
FJ375759	Hu/CHN/2008/SH1/type1b
FJ792842	Hu/CHN/2008/Shanghai/type1
KC342249	Hu/CHN/2012/km1/type1 b
AB009985	Hu/JPN/1993/J1050/type1b
AB000286	Hu/JPN/1993/J816/type1b
EF138826	Hu/USA/1999/3085/type1d
KY744138	**HAstV/ITA/PA364/1999/type1d**
KY744139	**HAstV/ITA/PA70R-/2004/type1d**
JX087965	Hu/ITA/2005/PA124/type1d
AY720892	Hu/DEU/2004/Dresden-1/type1d
JF327666	Hu/IND/2006/Pune-063681/type1b
FJ755403	Hu/CHN/2006/Beijing-176/type1d
JN887820	Hu/KOR/2006/lhar-2011/type1d
KF211475	Hu/CHN/2010/JZ/type 1d

The evaluate whether site-specific selection pressure operates on ORF2, the ratio of nonsynonymous (dN) to synonymous (dS) nucleotide substitutions per site was calculated using all the available models of substitution at the Datamonkey server (http://datamonkey.org), in particular, Felsenstein's 1981 model (F81); Tamura-Nei model (TN93); Kimura80 and Felsenstein81 models (HKY85) and General Reversible Model (REV) (Delport et al., 2010). The dN/dS ratios (v) were calculated using three different codon-based maximum likelihood approaches (CBML): the single-likelihood ancestor (SLAC), fixed-effects likelihood (FEL) and the internal branch fixed-effects likelihood (IFEL) (Kosakovsky Pond et al., 2006; Kosakovsky Pond and Frost, 2005).

2.4. Molecular modelling and graphics

The mature HAstV capsid protein (CP) structures were built by the I-TASSER protein structure prediction server (Roy et al., 2010; Zhang, 2008). The KF211475/CHN/2010/JZ/type-1d HAstV ORF2 amino acid sequence (accession no. AGX15185.1) was used to create the VP34 and VP25/27 capsid proteins. The model building was separated because the best templates for the VP34 (as. 78–413) shell domain were three CP X-ray and cryoEM structures (PDB IDs: 2ZTN, 2ZZQ and 3IYO) of Hepatitis E virus (HEV) whilst for the VP25/27 (aa. 414–646) spike domain the human astrovirus capsid spike domain (PDB ID: 3QSQ) proved to be best modelling template. The raw protein model structures were refined with the MacroModel energy minimization module of the Schrödinger Suite (www.schrodinger.com/pymol) to eliminate the steric conflicts between the side-chain atoms. The $T = 3$ virion model was created with the Oligomer Generator application of VIPERdb (available at http://viperdb.scripps.edu/oligomer_multi.php). Prior to virion model generations, the asymmetric units were constructed with Schrödinger Suite using the coordinates of subunit A, B and D of HEV