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

Simona De Grazia\textsuperscript{a,⁎}, Gianvito Lanave\textsuperscript{b}, Floriana Bonura\textsuperscript{a}, Noemi Uron\textsuperscript{a}, Vincenzo Cappa\textsuperscript{a}, Sara Li Muli\textsuperscript{a}, Arcangelo Pepe\textsuperscript{a}, Akos Gellért\textsuperscript{c}, Krisztian Banya\textsuperscript{c}, Vito Martella\textsuperscript{b}, Giovanni M. Giammanco\textsuperscript{a}

\textsuperscript{a}Department of Health Promotion Sciences and Mother and Child Care “G. D’Alessandro”, University of Palermo, Palermo, Italy
\textsuperscript{b}Department of Veterinary Medicine, University of Bari, Valenzano, Bari, Italy
\textsuperscript{c}Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Hungária krt. 21, Budapest H-1143, Hungary

\textbf{A R T I C L E   I N F O}

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

\textbf{A B S T R A C T}

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, \textit{Mamastrovirus} and \textit{Avastrovirus}, which infect mammalian and avian species, respectively. Astrovirus is an icosahedral virus with a star-like surface structure and a diameter of approximately 30 nm (Kovács et al., 2017; Mendez, 2007). This hypervariable domain (HVR) (aa 425–688) and a highly acidic C-terminal protruding region (VP25/27) 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.

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 characterised 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 (Gabbaray et al., 2007a; Guix et al., 2005). Currently, HAstVs are divided into eight genetically and antigenically distinct types (HAstV-1 to -8) within the \textit{Mamastrovirus} species I. Genotyping surveys have revealed that HAstV-1 is the most common type worldwide (De Grazia et al., 2016; De Grazia et al., 2012a; Gabbaray et al., 2007b; Guix et al., 2002; Koopmans et al., 1998; Malasaou et al., 2012; Wang et al., 2001). The predominance of type-1 HAstVs has raised interest for this genotype...
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/2004/PA270, ITA/2009; Suptawiwat et al., 2017), mirroring what is observed for other viruses (Bok et al., 2009; Suptawiwat et al., 2017), suggesting possible antigenic-related mechanisms of selection (Kovács et al., 2017).

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 = 0.7348)).

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).

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 (HK85) 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 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/type1d 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.
CP in vdb convention format. The VP25/27 spike domains were only attached to the icosahedral two fold axes. Further detailed structural analysis of the human astrovirus capsid can be seen in the study of Dong et al. (2011). Molecular graphics were created with VMD version 1.9.2 (Humphrey et al., 1996).

3. Results

3.1. HaAstV-1 lineages circulation

During a longitudinal surveillance of HaAstV epidemiology in South of Italy, between 1999 and June 2016, a total of 3704 stool samples, obtained from children hospitalized for acute gastroenteritis, were screened, showing an overall prevalence rate of 2.78% and marked yearly fluctuations (0.53–10.14%). A genotype-specificity was assigned for 80 (77.7%) out of 103 strains detected and type-1 was the prevalent genotype accounting for 87.5% of all HaAstV genotyped strains. Upon sequence characterization, based on a diagnostic region located at the 5’ end of the ORF2 (348 bp), various HaAstV-1 lineages (HaAstV-1a, -1b and -1d) were identified over the study period, with type-1a strains appearing as predominant from 2007 onwards (Fig. 1A). A spatio-temporal map of the HaAstV-1 lineages was constructed using the entire set of partial and complete ORF2 sequences available in Genbank databases at the time of writing in order to obtain a picture of their global distribution and circulation (Fig. 1B–E). A total of 732 HaAstV-1 ORF2 sequences were retrieved. Geographical and temporal information was retrieved for each type-1 strain from metadata deposited in Genbank. HaAstV-1 ORF2 sequences were grouped in relation to the origin of the strains (Table 2). An increasing number of sequences was released in the 2000s (N = 593), with respect to the preceding decades (N = 139). According to Genbank...
metadatabase, continuous molecular characterization of HAstV genotypes was conducted only in Italy, Russia, and Brazil since the 1990s and in Far East Asian countries (such as Japan, China and Singapore) since 2005. Circulation of different HAstV-1 lineages was observed over time. Type-1a, -1b and -1d strains co-circulated in Europe, with a predominance of HAstV-1d, whilst HAstV-1a consistently circulated with type-1d since 1993, but reverting to L and R in the more recent strains. Among HAstV-1b strains, five lineage-peculiar aa substitutions were identified (34K → R, 43R → K, 46A → T/I, 53D → G and 55R → H/L). Two additional substitutions (273N → K/R and 364S → T) were shared by HAstV-1b and HAstV-1d strains whilst three aa residues were peculiar of HAstV-1d strains (44N → S, 301V → A and 307T → V/L).

In the hypervariable domain (HVR) (aa 425–688), HAstV-1a and HAstV-1d strains showed unique aa substitutions at positions 521 (V → I) and 672 (K → N), respectively, when compared to the reference strain L23513/GBR/Oxford-1/type1a. HAstV-1b strains differed from the other HAstV-1 lineages in 11 residues (427A → T, 435V → I, 451T → N/S, 537S → T, 600M → L, 605V → I, 624V → I/T, 650S → A/T, 667S → F, 671F → L/V and 672K → R). At position 652, HAstVs-1b and -1d shared a tyrosine instead of histidine (652H → Y).

In the highly acidic C-terminal domain, HAstV-1b strains differed from HAstV-1a and -1d in three residues (739S → T, 758N → D and 759N → S). All HAstV-1d showed two substitutions (729R → K and 752V → A) whilst all HAstV-1a strains were similar to the prototype strain (Table 3).

### 3.4. Capsid structure analysis of the type-1 HAstVs

The 3D capsid protein and virion models were created for the major HAstV-1 lineages in order to understand the possible impact on antigenicity and functionality of the changing aa residues across the various lineages (Fig. 3). Among the 31 aa positions affecting the phylogenetic classification, six residues were mapped to the mature VP34 shell capsid protein, whilst eight belonged to the outer spike domain (VP25/27) (Table 3).

Structural analysis of HAstV capsid allowed identifying at least six residues exposed on the virion surface. Two residues (273 and 364) were located in the VP34 (shell region) whilst four residues (427, 435, 600 and 605) were mapped in the VP25/27 (protruding region) of HAstV capsid protein, in proximity of the putative receptor binding site.

### 3.5. Shannon entropy and selection pressure analysis of type-1 HAstVs

The high prevalence and the prolonged circulation of HAstV of genotype 1 worldwide prompted us to measure the aa diversity in each position and to evaluate putative antigenic residues under positive pressure. Shannon entropy analysis in the ORF2 complete sequences of HAstV-1 showed a subset of positions undergoing frequent changes, whilst some positions were highly conserved (data not shown).

Sites under positive selection were predicted, using 34 HAstV-1 complete ORF2 sequences (2358 nt residues). As it is shown in Table 4, four residues in the N-terminal portion, five in the HRV and one in the C-terminal domains were identified by almost one of the models used.

Residues 50, 492 and 758 appeared to be under positive selection pressure showing a normalized dN/dS rate with a p value ≤ 0.05. Residue 50, located in the N-terminal domain and residue 758, located in the C-terminal domain, were identified by two methods (FEL and IFEL) with at least three (TN93, HK85 and REV) and two (TN93 and REV) of the models used, respectively. Residue 492, in the hypervariable domain, was identified by FEL with all the four selection models, by IFEL with two selection models and by SLAC with one substitution model (Table 4).

On the bases of the virion structure modelling, the sites 50 and 758 are not in the structured (mature) capsid region whereas residue 492 is located in the spike domain, at the binding interface between the P1 domain of the VP34 and the bottom of the spike domain (VP25/27) (Table 3).
4. Discussion

The present study investigated the molecular evolution of the gene coding for the capsid proteins (complete ORF2) of HAstV type-1, which is the predominant HAstV type worldwide (De Grazia et al., 2016; Gabbay et al., 2007a; Guix et al., 2002; Koopmans et al., 1998; Malassa et al., 2012; Schnagl et al., 2002; Wang et al., 2001). In Italy, a longitudinal epidemiological monitoring of HAstV infections has been conducted since 1999, and HAstV-1 was detected in 87.5% of all HAstV-positive cases. High genetic heterogeneity has been observed among HAstV-1 upon molecular characterization based on a diagnostic region located at the 5′ end of the ORF2, allowing for the identification of at least three type-1 lineages (De Grazia et al., 2016; Martella et al., 2014).

Spatio-temporal analysis of viral strains circulating worldwide represents a useful tool to assess the prolonged circulation of successful clones and/or the emergence/re-emergence of different clusters over time. Relatively few epidemiological studies have followed circulation of HAstV genotypes and lineages for a long time period (Gabbay et al., 2007b; Guix et al., 2002; Nguyen et al., 2008). Due to the lack of adequate information in the literature, we investigated Genbank metadata available for all the HAstV-1 sequences released until February 2017 (29 complete and 703 partial ORF2 sequences). Temporal patterns of the circulation of HAstV-1 lineages were observed in different geographic areas, even if this information could be underestimated as the majority of the sequences available were obtained from specific hospitals within several countries. Analysis of the 5′ portion of ORF2 is a good proxy for characterization of genotypes and lineages, but also has some limitations, as this genomic region does not allow for evaluation of the hypervariable region that codes for the virion spike. Anyway, such analyses allowed observing a prolonged circulation of type-1 HAstVs of different lineages over the time (De Grazia et al., 2004, 2016; Gabbay et al., 2007b; Mustafa et al., 2000; Schnagl et al., 2002).

In order to address better this hypothesis, ORF2 sequences available in the databases and generated in this study were analyzed in details. Phylogenetic analysis confirmed the presence of three HAstV-1 lineages...
Table 3
Phylogenetically relevant amino acid (aa) residues for the HAsTV-1 lineages diversification. White and black squares indicate the aa residues exposed on the surface of mature HAsTV virion, in the VP34 and VP25/27 portions, respectively. Accession number of the 34 strains is shown in Table 1.

<table>
<thead>
<tr>
<th>N terminal domain</th>
<th>HRV domain</th>
<th>C terminal domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>34</td>
<td>43</td>
</tr>
<tr>
<td>Type-1a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/JPN/1992/J115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/JPN/1992/J153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/JPN/1992/J1526</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/GBR/1993-92/Oxford-S1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/GBR/1993/A2-88 Newcastle</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Hu/USA/1999/1638</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/USA/1999/2882</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/USA/1999/2987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/ITA/2008/PA1-48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/HUN/2010/HUN4520</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Hu/HN/2011/Pa387</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/CHN/2014/KSM03</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Hu/CHN/2014/KSM04</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Type-1b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/JPN/1993/J1050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/JPN/1993/J186</td>
<td>R K T G H</td>
<td>K</td>
</tr>
<tr>
<td>Hu/CHN/2005/Beijing-128</td>
<td>R K T G H</td>
<td>K A V T</td>
</tr>
<tr>
<td>Hu/IND/2006/Pune-063681</td>
<td>R K T G H</td>
<td>K</td>
</tr>
<tr>
<td>Hu/CHN/2007/Beijing-293</td>
<td>R K T G H</td>
<td>K</td>
</tr>
<tr>
<td>Hu/CHN/2007/Shenyang</td>
<td>R K T G H</td>
<td>K</td>
</tr>
<tr>
<td>Hu/CHN/2008/Sh1</td>
<td>R K T G L</td>
<td>K</td>
</tr>
<tr>
<td>Hu/CHN/2008/Shanghai</td>
<td>R K T G L</td>
<td>K</td>
</tr>
<tr>
<td>Hu/CHN/2012/km1</td>
<td>R K T G L</td>
<td>K</td>
</tr>
<tr>
<td>Type-1d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/USA/1999/3085</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Hu/ITA/1999/PK364 S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/ITA/2004/PK70R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu/DEU/2004/Dresden-1</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Hu/ITA/2005/PA124</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Hu/CHN/2006/Beijing 176</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Hu/KOR/2006/Inar-2011</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>Hu/CHN/2010/JZ/type 1d</td>
<td>S</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 3. Capsid structure analysis of HAstV-1. Bar diagram of the analyzed thirty-four HAstV-1 ORF2 amino acid sequences (A). Those residues which are significantly responsible for the phylogenetic diversification are indicated by a vertical line. The VP34 (grey) and the VP25/27 (pale blue) mature capsid protein segments are in cartoon representation (B). Those amino acids which play significant role in the lineages separation of the type-1 (-a to -d) are illustrated with VdW spheres. Green VdW spheres are in surface exposed positions. The S-site is displayed in transparent surface representation. Molecular surface representation of the T = 3 astrovirus capsid (C). Colour codes are the same as in panel B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
(-1a, -1b and -1d), each characterized by high intra-lineage sequence identity values. Sequence comparison of the deduced protein sequences along the three ORF2 regions (i.e., N-terminal, HRV and C-terminal domains) allowed identifying peculiar aa substitutions within each HAstV-1 lineage. Type-1a HAstV strains (highly similar to the prototype strain L23513/GBR/Oxford-1) showed temporal patterns of aa variation, with recent type -1a strains being genetically more similar to each other. Type-1b HAstV strains showed a high number of aa substitutions with respect to the prototype Oxford-1 (19 changes in the complete ORF2), scattered throughout the three domains of the capsid protein (5 in the N-terminal region, 11 in the HRV and 3 in the highly acidic C-terminal domain). Type-1d HAstV displayed six hallmark substitutions, not present in the other HAstV-1 lineages. HAstV-1b and -1d strains shared three aa substitutions that were not conserved in type-1a strains, two in the N-terminal domain and one in the HRV portion. Two of the shared N-terminal domain substitutions, 273N → K/R and 364S → T, potentially affects antigen-antibody interactions. In the residue 273, a polar, neutral side chain was changed to a positively charged residue, whilst the 364S of type-1a HAstV capsid protein can be phosphorylated with a much higher probability than type-1b/1d strains. Two peculiar aa substitutions with similar physico-chemical properties, 301V → A and 605V → I, were observed respectively in -1d and -1b strains, but such changes usually do not perturb the protein function. Interestingly, the type-1a strains detected from 2008 and almost all the HAstV-1b showed a substitution at position 427 in the HRV portion in which a threonine replaced an alanine, introducing a putative phosphorylation site into the spike domain of the capsid protein expressed on the virion surface.

Structure prediction of the HAstV-1 capsid was achieved to value the presence of peculiar aa residues located on the virion surface and possibly involved in virus-cell and/or virus-antibody interaction. Thirty-one aa substitutions, important for phylogenetic classification, were identified on the alignment and phylogenetic analyses of 34 HAstV-1 full-length ORF2 sequences. Of these 31 residues, six mapped to the mature VP34 shell capsid protein, whilst 8 were localized on the outer spike domain (VP25/27). Although it has been demonstrated that the spike domain (VP25/27) alone is responsible for direct interaction with cellular surface (Bass & Upadhyayula, 1997; Sanchez-Fauquier et al., 1994; Walter et al., 2001), our analysis indicates that the VP34 domain also has two amino acid residues located on the virion surface (positions 273 and 364). Noteworthy, four residues (positions 427, 435, 600 and 605) among the nine lineage-defining positions on the spike domain were located in the HRV and therefore might influence the immune response (green beads in Fig. 3B and C). The positions 340–376 span the antigenic epitope, whilst positions 580–606 are associated with the region involved with recognition of serotypic antibody. Accordingly, variations in such residues could trigger mechanisms of immune escape (van Hemert et al., 2007).

The majority of the lineage-defining amino acid substitutions showed similar physico-chemical properties, not perturbing the protein function. However, replacement of a neutral side chain (N) with a positively charged residue, as observed at position 273 (K/R), could affect the capsid-antibody interactions. In fact, the identification of substitutions in residues 273 and 364, observed in all HAstV-1b and -1d strains, could sustain the hypothesis that these residues play a role in the virus immune escape, even if all HAstV-1a strains retained the same aa residues since their first identification.

Evolving and changing the antigenic profile is a common strategy by which viruses evade pressure from the immune system. Selection pressure analyses and Shannon entropy are useful tools to evaluate the dynamics of viral evolution. For example, it has been documented that the hemagglutinin of seasonal influenza virus evolves under positive selection pressure exerted by host immunity (Suptawiwat et al., 2017).

Shannon entropy and positive selection pressure analyses of the ORF2 alignment allowed identifying numerous variable sites in the capsid of HAstV-1 strains. For evaluation of the purifying selection sites of an intermediate data set (20–40 sequences) a codon-based maximum likelihood approach like FEL method is recommended (Pan and Deem, 2011). Our data set of 34 sequences was analyzed using FEL and IFEL approaches, detecting several aa residues under negative pressure and 10 aa residues under positive pressure, four of which (residues 19, 50, 492 and 758) showed a statistical significant dN/dS value (p ≤ 0.05). The majority of the sites under positive selection pressure were mapped to the N-terminal region of the VP34 (residues 19, 35, 50 and 55) and to the HRV region of the ORF2, coding for the VP25/27 proteins (residues 427, 465, 492, 566 and 650), whilst a unique residue, aa 758, was mapped to the highly acidic C-terminal region. Position 492 was constantly recognised as a positively selected residue in selection pressure analysis, regardless of the substitution selection models and maximum likelihood method used, and it showed the highest entropy. Position 492 was either isoleucine, or methionine or threonine in the sequence alignment but changes in this residue did not affect peptide recognition and to conserve peculiar residues involved in the viral life cycle. It may play a key role in the protein-protein interaction (PPI) between the shell capsid and the protruding spike domains (Fig. 3). This type of fine-tuning of the PPI could be a driving force for HAstV evolution.

Negative selection represents a mechanism to preserve virus propagation and to conserve peculiar residues involved in the viral life cycle. On the contrary, a viral protein with strong antigenicity may undergo strong selection pressure, resulting in the presence of many sites under positive selection (Nielsen, 2005). For a virus, positive pressure and high entropy are commonly related to a response to environmental stress, such as immune pressure. Similar studies describing the

<table>
<thead>
<tr>
<th></th>
<th>F81</th>
<th>TN93</th>
<th>HKY85</th>
<th>REV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aa</td>
<td>SLAC</td>
<td>FEL</td>
<td>IFEL</td>
</tr>
<tr>
<td>N-terminal domain 1-424 aa</td>
<td>19</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>HRV domain 425-688 aa</td>
<td>465</td>
<td>√</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>C-terminal domain</td>
<td>758</td>
<td>√</td>
<td></td>
<td>√</td>
</tr>
</tbody>
</table>

Table 4
Sites under selective pressure of the 34 HAstV-1 ORF2 strains. Four nucleotide substitution selection models available in the DataMonkey website (F81, TN93, HKY85 and REV) with three different codon-based maximum likelihood approaches (FEL, IFEL and SLAC) implemented at http://www.datamonkey.org were used. The amino acid (aa) with a dN/dS value statistical significant (p ≤ 0.05) are indicated in bold.
relationship between positive selection and antigenicity have targeted the capsid proteins of other enteric viruses such as norovirus and enterovirus (Kobayashi et al., 2015; Zhang et al., 2010).

5. Conclusion

The genetic heterogeneity observed in the complete ORF2 of HAstV-1 strains is not surprising since genetic/antigenic variation is essential to sustain prolonged circulation of several enteric viruses in their host. Phylogenetic analysis of HAstV-1 sequences available in the databases revealed temporal shifts of the HAstV-1 lineages, a finding that is consistent with the existence of mechanisms enacted to decrease the antigenic pressure.

Similarly to rotavirus and norovirus, HAstV has been differentiating into sero/genotypes (HAstV-1 to -8), and into a number of lineages. The differentiation of type-1 HAstVs into lineages could account for the apparent predominance/persistence over consecutive years of HAstV-1, as documented in worldwide epidemiological studies. Although several norovirus genotypes have been identified in humans, norovirus enteritis is mostly attributed to a single genetic type, GII.4, whose polymerase gene and major viral capsid protein quickly evolve into variants to sustain continuous consistent circulation (Siebenga et al., 2009). Different GII.4 variants differ in aa mutations in the putative antigenic regions located in the P2 capsid region, altering recognition by blocking antibodies (Allen et al., 2008). Similarly, the relative fitness of the main human rotavirus genotype combination, G1P[8], was related to the continuous emergence/re-emergence of lineage/sub-lineage combinations of the antigenic proteins VP7 and VP4 (Arista et al., 2006; Banyai et al., 2009). In this study, HAstVs appear to be more similar to rotaviruses, where the alternate circulation of different strains and lineages is regarded as a mechanism that reduces the selection pressure by recycling of different lineages, rather than the continual emergence of novel strains, as observed for norovirus and influenza A viruses. Overall, these findings may suggest that antigenic mechanisms similar to those observed and/or hypothesized for norovirus and rotavirus are also shaping the evolution of HAstVs.

Acknowledgements

The licensing of the Schrödinger Suite software package was funded from the NKFI grant under agreement No.108793.

Krisztian Bányai was supported by the Momentum program awarded by the Hungarian Academy of Sciences (OTKA, T100727).

References


Krisztian Bányai was supported by the Momentum program awarded by the Hungarian Academy of Sciences (OTKA, T100727).

The licensing of the Schrödinger Suite software package was funded from the NKFI grant under agreement No.108793.


