Recombination in hepatitis A virus: evidence for reproductive isolation of genotypes

Ilya S. Belalov, Olga V. Isaeva and Alexander N. Lukashev

M. P. Chumakov Institute of Poliomyelitis and Viral Encephalitides, Moscow, Russia

This study analysed phylogenetic evidence of recombination in sequences of hepatitis A virus (HAV) available in international databases. Isolation of distinct recombinant HAV strains has been reported previously; however, the prevalence of natural recombination and its role in HAV genetics remains obscure. Analysis of full genome sequences revealed evidence of common intratypic recombination among the most prevalent subtypes, IA and IIIA. Many of the available complete sequences of these genotypes carried phylogenetic signs of recombination in all genomic regions without obvious hotspots. In addition, and in line with previous reports, recombination between subtypes IA and IB was detected. A dataset of 104 published HAV sequences for the VP1–2A and 3CD genomic regions was also analysed. Multiple instances of phylogenetic incompatibility were found among subtypes IA and IIIA. Three cases of recombination disrupted the phylogenetic grouping of subtype IA HAV strains isolated in Japan within less than 4 years, indicating common intratypic recombination in HAV. There were no signs of recombination between different HAV genotypes, despite the fact that co-circulation of genotypes IA and IIIA has commonly been reported in different parts of the world and many sequences in the sampling in this study originated from the same geographical region. These results indicate that there is reproductive isolation between genotypes of HAV, as exists between enterovirus species, and suggest that common intratypic recombination constrains the diversity within a genotype and maintains HAV genotypes as global gene pools.

INTRODUCTION

Hepatitis A virus (HAV), a member of the family Picornaviridae, is the most prevalent causative agent of water-borne hepatitis in the world. HAV has a distinct genome structure and biological properties compared with other members of the family and has therefore been classified in the genus Hepatovirus. The genome of HAV comprises a positive-sense ssRNA of ~7500 nt, which encodes one polyprotein (Fig. 1). The polyprotein is cleaved co- and post-translationally into 11 proteins: four structural proteins, VP1–VP4 (P1 region), comprising the viral capsid, and seven non-structural proteins, 2A–2C (P2 region) and 3A–3D (P3 region). The major non-structural proteins are the 3D polymerase, 3C protease and 2C, a putative helicase. The hepatovirus 2A protein is distinct from the 2A protein of other picornaviruses and does not have protease activity. The ORF is preceded by a 5’ non-translated region (5’NTR) of ~730 nt that contains the replication origin and internal ribosome entry site (IRES). The secondary structure of the HAV IRES is distinct from other picornavirus IRESs. The genome ends with a short 3’NTR (Hollinger & Emerson, 2007). The virus has several features uncommon among picornaviruses. Firstly, the genus is represented by a single species and one sole serotype. Secondly, it has been reported that the virus has a strong preference towards rare synonymous codons, presumably to limit the translation rate and avoid eliciting an immune response (Pinto et al., 2007). In addition, HAV features the lowest substitution rate among picornaviruses (Moratorio et al., 2007).

HAV exhibits properties common to other picornaviruses such as a global distribution, high prevalence of infection, especially in the endemic regions, and the possibility of recombination (Costa-Mattioli et al., 2003a). However, reports of recombination in HAV are scarce compared with other picornaviruses. Only two occurrences of natural recombination have been described for HAV in the capsid-encoding genomic region between genotypes IA and IB (Costa-Mattioli et al., 2003b; Liu et al., 2010a); however, there has also been a report suggesting in vitro recombination in HAV during persistent infection in cell culture (Lemon et al., 1991). Other well-studied picornavirus genera (Enterovirus, Pachyovirus, Cardiovirus and Aphthovirus) feature ubiquitous recombination (Benschop et al., 2010; Drexler et al., 2010; Heath et al., 2006; Lukashev et al., 2003; Simmonds & Welch, 2006), which allows a distinct evolutionary history of the genome fragments, even on the scale of years and decades. We suggested that the...
prevalence of recombination in HAV is underestimated and conducted an analysis of the available sequences to test this hypothesis.

RESULTS

Distribution of recombination events over the genome

We started our analysis of recombination in HAV by creating a phylogenetic compatibility matrix to get an overview of the recombination frequency in different genomic regions (Fig. 1a). There was evidence of multiple phylogenetic conflicts that were distributed over the genome. No genome part was obviously spared by recombination; however, there were regions of high and low degrees of phylogenetic incongruence. Most prominently, grouping in the 5’NTR was poorly compatible with the rest of the genome. The phylogenetic conflicts reflected on the phylogenetic compatibility matrix were well supported by bootstrap values, as a very similar picture could be observed when a stringent 95% bootstrap cut-off (data not shown) was used instead of the default 70% cut-off (Fig. 1a). We then analysed recombination patterns separately for HAV subtype IA (15 sequences) and subtype...
I. S. Belalov, O. V. Isaeva and A. N. Lukashev

IIIA (14 sequences). Comparable patterns were observed within the subtypes (Fig. 1b, c). Both subtypes featured recombination in the P1 genomic region and poor compatibility between grouping in the 3′ part of the 3D polymerase and in other genomic regions. A noticeable difference between the subtypes was the absence of phylogenetic incongruence between the 5′NTR and the rest of the genome in subtype IIIA. The apparent conflicts of phylogenetic conflicts were higher when analysing distinct subtypes (Fig. 1b, c); however, this does not necessarily correspond to a higher recombination prevalence among subtypes IA and IIIA, because preserved groups corresponding to other genotypes were excluded from these datasets. We compared recombination patterns in HAV with human enterovirus C (HEV-C), a member of the genus Enterovirus of the family Picornaviridae. This virus features frequent recombination (Brown et al., 2003) and is representative of other enteroviruses (Simmonds & Welch, 2006) and most picornaviruses (Lukashev, 2010) in terms of recombination prevalence and patterns. The distribution of recombination events in HAV was different from HEV-C (Fig. 1d). The latter demonstrated no apparent recombination in the P1 genomic region encoding the capsid proteins and a high degree of phylogenetic incompatibility between genomic regions encoding the structural (VP4–VP1) and non-structural (2A–2C and 3A–3D) proteins.

Similarity plots and bootscan analysis were then used to investigate recombination events in individual sequences. Most HAV subtype IA and IIIA full genome sequences showed signs of recombination, with representative patterns shown in Fig. 2(a, b). Importantly, all recombination events in the dataset analysed were within a genotype, and almost all were also within a subtype. Bootscan analysis revealed no recombination events between distinct genotypes among the 39 complete HAV sequences used in this study. As was noted on the phylogenetic compatibility matrices (Fig. 1), recombination events in HAV were distributed throughout the genome without an obvious pattern. There were relatively few recombination events with sharp crossover points; however, reliable bootscan support could be obtained for many recombination cases suggested by similarity plots. The recombination patterns in HAV were different from those in HEV-C (Fig. 2c, d). In HEV-C, most sequences featured apparent recombination crossover points near the P1/P2 junction, as exemplified in Fig. 2. Crossover points in the P1 genomic region were common in HAV but rare in HEV-C. As discussed below, this observation could be explained by a sampling bias between HAV and enteroviruses.

**Phylogenetic conflicts between fragments of complete HAV genomes**

To confirm our results, we created Bayesian phylogenetic trees for eight distinct genomic regions of HAV (Fig. 3). There were no signs of recombination between viruses of different genotypes within the samples studied. All genotypes and subtypes IIIA and IIB made up distinct phylogenetic groups on all trees. There was one instance of recombination between subtypes IA and IB in the 5′NTR, concordant with previous reports (Costa-Mattioli et al., 2003b; Liu et al., 2010a). Within a subgenotype, most strains were shuffled on the phylogenetic trees for different genomic regions. Many significant phylogenetic conflicts that disrupted groups with posterior probabilities >0.95 could be observed (Fig. 3, dotted lines). Subgrouping of HAV within a subtype was generally unresolved when a consensus tree was created using trees for eight particular genomic regions (Fig. 3), highlighting multiple intratypic phylogenetic conflicts. Within the well-represented genotypes IA and IIIA, only two groups of two strains each were maintained throughout the genome with significant posterior probabilities and made up resolved clusters in the consensus tree (IA_VBA-07 and IA_LA; IIIA_IND-HAV-95F and IIIA_CP-IND). Several more groups were supported in all genomic regions, but not always with significant posterior probabilities. The phylogenetic tree produced using full genome sequences was perfectly resolved (Fig. 3) with significant posterior probabilities for most groups. Analysis of this tree revealed that it represented a topology that was dominant among phylogenetic trees for distinct genomic regions, masking obvious phylogenetic conflicts.

**Analysis of recombination between the VP1 and 3D genomic regions in an extended dataset**

In order to gain further support for our findings, we analysed an additional dataset of 104 sequences for the VP1–2A (nt 2940–3420) and 3CD (nt 5385–5979) genomic regions of HAV isolates collected predominantly in Japan in 1984–2004. Bayesian phylogenetic trees were created for the two genomic regions, and phylogenetic conflicts were highlighted (Fig. 4). This dataset was mostly confined to a narrow geographical and temporal window; therefore, there were relatively more conserved groups without signs of recombination than in the full-genome trees (Fig. 3). Nevertheless, we observed four phylogenetic conflicts that either disrupted significant grouping into incompatible but not significant grouping, or changed one significant grouping pattern into another. Unfortunately, it was not possible to distinguish co-circulation of recombinant lineages from fresh recombination events and therefore draw reliable conclusions on the temporal dynamics of recombination using this dataset. Importantly, this extended dataset contained 68 genotype I and nine genotype III sequences that were isolated in Japan within about two decades (see Supplementary Table S1, available in JGV Online). We did not observe any evidence of intertypic recombination between these co-circulating subtypes, concordant with the original study reporting these sequences (Endo et al., 2007a). Therefore, the absence of intertypic recombination and reproductive...
isolation of HAV (sub)types was supported by analysis of the co-circulating viruses.

**Analysis of pairwise distances between HAV isolates**

We analysed the distribution of pairwise amino acid sequence distances of 104 HAV strains in the VP1–2A and 3CD genomic regions (Fig. 5). All intratypic distances (uncorrected) were below 0.05 in the VP1–2A region and 0.025 in the 3CD region of the genome. There were no sequence pairs among over 5000 tested that differed by 0.05–0.06 in VP1–2A and very few that differed by 0.02–0.03 in 3CD. Pairwise distances between subtypes IA and IB and between IIIA and IIB stayed within intratypic distances. Therefore, HAV did not explore all available sequence space. We compared the distribution of pairwise sequence distances in the 3CD genomic region for members of the genera *Hepatovirus* and *Enterovirus*. The latter featured much greater amino acid distances than HAV, up to 0.48 compared with 0.05. It is not clear which genetic and biological features of HAV provided such a degree of sequence conservation. Codon usage bias could contribute, but is hardly sufficient to explain fully the low
Recombination in hepatitis A virus
sequence variation. Importantly, both genera of picornaviruses displayed two distinct peaks of pairwise sequence distances within a genus. Amino acid sequence distances within groups that featured common recombination (genotypes in HAV or species in enteroviruses) were distinct from genetic distances between such groups. This segregation of pairwise sequence distances was most clear in enteroviruses, but was also detectable in HAV.

**DISCUSSION**

Distinct examples of recombination in HAV have been reported previously (Costa-Mattioli et al., 2003b; Liu et al., 2010a). However, it is becoming increasingly acknowledged that there is a substantial chance that apparent recombinant sequences may result from laboratory mistakes, such as cross-contamination of samples and selective amplification of virus mixture components by PCR (Boni et al., 2010). We found evidence of multiple recombination events in HAV. Many cases of phylogenetic incongruence were obvious upon direct comparison of Bayesian phylogenetic trees (Fig. 3). Importantly, sequences that were involved in these phylogenetic conflicts were identified in different laboratories from viruses collected in different countries, and in both of the most common HAV genotypes, IA and IIIA. In addition, recombination events were detected in an extended sequence dataset among viruses from two countries (India and Japan). Complex recombination patterns revealed by phylogenetic compatibility matrices and bootscanning indicated that recombination is rather common in HAV. In contrast, similarity plots did not detect any cases suspicious for laboratory contamination, such as regions of near 100% similarity between sequences fairly distant in other genomic regions. Thus, recombination in HAV appears to be common and ubiquitous, as in many other picornaviruses.

**Reproductive isolation of HAV genotypes**

HAV genotypes feature distinct geographical distribution (Robertson et al., 1992); however, there are areas of co-circulation of genotypes IA and IIIA (Chironna et al., 2003; Pina et al., 2001; Tallo et al., 2003). Co-infection with genotypes IIIA and IB has been reported in India, the country of origin of ten of the genotypes IIIA in our sampling (Chitambar et al., 2007). In our extended dataset, we analysed 76 sequences representing subtypes IA, IB, IIIA and IIIB that were isolated in Japan within two decades (Supplementary Table S1). In addition, HAV infection is common in travellers (Lenfant, 1994), which provides carriage of HAV genotypes worldwide. Therefore, there is a possibility of co-infection with different HAV genotypes, which is a prerequisite for recombination. There have been several reports of recombination between subtypes IA and IB (Costa-Mattioli et al., 2003b; Liu et al., 2010a), but no reports of recombination between distinct genotypes. In our study, we had a sampling of 39 complete sequences representing four genotypes and seven subtypes that were sampled throughout the world. We did not observe any evidence of recombination between different genotypes, similar to a previous report that used partial VP1–2A and 3CD sequences and confirmed the applicability of both of these genomic regions for identification of subtypes (Endo et al., 2007a). We observed recombination between subtypes IA and IB, which has also been reported previously, but no recombination between subtypes IIIA and IIIB; however, the IIIB subtype was represented by only two sequences in our study, and such recombination cannot be ruled out. We therefore suggest that there are natural barriers to recombination between HAV genotypes. One possible restriction to intertypic recombination could be incompatibility of genome fragments or viral proteins from different genotypes. This hypothesis, commonly accepted to explain barriers to recombination in enteroviruses, raises certain doubts in the case of HAV, as the proteins of HAV genotypes are much less diverse than those of enteroviruses (Fig. 5). In contrast, it has been shown that even a single amino acid substitution can provide incompatibility of structural and non-structural proteins in HEV-C (Liu et al., 2010b), making the 4–5% sequence distance observed between HAV genotypes theoretically sufficient for incompatibility of viral proteins.

**Impact of recombination on HAV classification**

HAV is currently classified as the sole species, *Human hepatitis A virus*, of the genus *Hepatovirus* (Stanway et al., 2005). The reason for such classification was limited diversity of the virus and low variability of the serological properties of isolates throughout the world. The International Committee on Virus Taxonomy terms picornavirus species as viruses with >70% amino acid identity in P1 and >70% amino acid identity in 2C+3CD (Stanway et al., 2005). HAV sequences in our samples differed from each other by no more than 20% of the RNA sequence and by no more than 9% of the amino acid sequence (Fig. 5) and therefore perfectly satisfied this formal species definition. However, nucleotide sequence distances in HAV are generally much lower compared with other picornaviruses, which is probably...
Fig. 4. Phylogenetic trees for 105 HAV strains in two genomic regions (VP1–2A, nt 2940–3420; 3CD, nt 5385–5979). Bayesian trees were created similarly to Fig. 3. Branches above the (sub)genotype level were truncated to improve the resolution of small clusters. Numbers at tree nodes are posterior probabilities. Tree scale is relative; therefore, bars were omitted. Insignificant posterior probabilities (<0.95) were removed from the tree. Groups with significant posterior probabilities that were disrupted on another tree are shown in bold. Dotted lines indicate significant phylogenetic conflicts.
explained by the peculiarities of codon usage and replication strategy. This feature of HAV questions the use of this formal species criterion. In contrast, a general species definition in biology implies reproductive isolation. We did not detect any intertypic recombination events among 39 full-genome sequences and 104 paired VP1 and 3D sequence fragments. We therefore suggest that HAV genotypes fulfil this key criterion of distinct species.

The pairwise sequence distances provide very clear delimitation of HAV genotypes, much as they determine species in enteroviruses. Obviously, HAV does not occupy all possible sequence space within the genus. Isolates of HAV are either similar (within a genotype) or very distant (between the genotypes), never moderately distant. This could be explained by the clonal emergence of genotypes in a distinct evolutionary burst and their gradual divergence; however, this hypothesis contradicts our findings of common intratypic recombination. We suggest that, over a long timescale, common recombination conserves the integrity of the genotypes by shuffling the global gene pool and constraining divergence of the viruses. Every virus should recombine with a compatible virus from this pool from time to time. Whenever viruses become isolated from the global gene pool for a sufficiently long time, they become extinct or get a chance to establish a distinct gene pool, which we observe as a novel taxon (genotype or species). This hypothesis also highlights similarity between reproductively isolated HAV genotypes and species in other picornavirus families. The existence of subtype IB, which is only partially reproductively isolated from subtype IA, could represent the process of speciation.

Recombination patterns and constraints in HAV and enteroviruses

We observed recombination in all genomic regions of HAV. Most reports for other picornaviruses describe apparent recombination hot-spots on the edges of the capsid-encoding genomic region (Benschop et al., 2010; Drexler et al., 2010; Heath et al., 2006; Lukashev et al., 2005; Simmonds & Welch, 2006). Recombination in the capsid-encoding part of the genome is an uncommon finding in enteroviruses, aphthoviruses, parechoviruses and cardioviruses. This is most probably a result of a sampling bias, as these viruses feature a wide repertoire of serotypes, and only few sequences of a particular genotype are often available for analysis. When many complete sequences of a single serotype were analysed, e.g. for coxsackieviruses (Oberste et al., 2004), enterovirus 71 (Huang et al., 2008) and human cardioviruses (Drexler et al., 2010), recombination in the capsid-encoding part of the genome could be found between viruses of the same type. Summarizing the recombination patterns in HAV and HEV-C, we can conclude that distinct serotypes in enteroviruses and cardioviruses make a barrier to recombination in the P1 genomic region, but the sole serotype in HAV is not sufficient to permit recombination between viruses of different genotypes.

Practical implications

Molecular typing of HAV is historically carried out using a short sequence in the VP1–2A junction (Jansen et al., 1990; Robertson et al., 1992); however, other genomic regions have also been used in HAV molecular epidemiology (Jansen et al., 1990; Robertson et al., 1991). Our results suggest that, due to the apparent absence of recombination between viruses of different types, almost any genomic region is suitable for discriminating between genotypes. Distinction of types IA and IB is more complicated. Obviously, the 5’NTR is not suitable for this purpose (Fig. 3), as well as VP2, which also featured as a recombination cross-over region in a previous report (Costa-Mattioli et al., 2003b). There have been no reports of recombination between IA and IB subtypes in the non-structural genome part as yet, but this might be due to the low numbers of
complete sequences of subtype IB available and the use of a single genomic region in most molecular epidemiology studies. Our results and published reports imply that segregation of IA and IB subtypes can be blurred upon increased sampling of complete HAV sequences. Our results also indicate that phylogenies produced from any single genomic region are sufficient for investigation of a limited outbreak, but do not reflect unambiguously the epidemiology of HAV over extended periods of time. For example, phylogenetic grouping of subtype IIIA strains isolated in India within several years was not the same in different parts of the genome (Fig. 3), and phylogenetic conflicts were found among Japanese subtype IA viruses isolated within several years. In the molecular epidemiology of enteroviruses that feature ubiquitous recombination, analysis of recombinant forms – combinations of VP1 and 3D genes that do not bear evidence of recombination at least within a studied sampling – have been suggested for evolutionary studies as a reasonable substitute for full-genome sequencing (Leitch et al., 2009).

Table 1. HAV sequences used in the analysis

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Genotype</th>
<th>Isolation year</th>
<th>Country of origin*</th>
<th>GenBank accession no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AH1</td>
<td>IA</td>
<td>1992</td>
<td>Japan</td>
<td>AB020564</td>
<td>Fujiwara et al. (2001)</td>
</tr>
<tr>
<td>AH2</td>
<td>IA</td>
<td>1991</td>
<td>Japan</td>
<td>AB020565</td>
<td>Fujiwara et al. (2001)</td>
</tr>
<tr>
<td>AH3</td>
<td>IA</td>
<td>1993</td>
<td>Japan</td>
<td>AB020566</td>
<td>Fujiwara et al. (2001)</td>
</tr>
<tr>
<td>FH1</td>
<td>IA</td>
<td>1992</td>
<td>Japan</td>
<td>AB020567</td>
<td>Fujiwara et al. (2001)</td>
</tr>
<tr>
<td>FH2</td>
<td>IA</td>
<td>1993</td>
<td>Japan</td>
<td>AB020568</td>
<td>Fujiwara et al. (2001)</td>
</tr>
<tr>
<td>FH3</td>
<td>IA</td>
<td>1994</td>
<td>Japan</td>
<td>AB020569</td>
<td>Fujiwara et al. (2001)</td>
</tr>
<tr>
<td>LU38/WT</td>
<td>IA</td>
<td>1988</td>
<td>China</td>
<td>AF357222</td>
<td>Hu et al. (2002)</td>
</tr>
<tr>
<td>LY6</td>
<td>IA</td>
<td>2001</td>
<td>China</td>
<td>AF485328</td>
<td>Hu et al. (2002)</td>
</tr>
<tr>
<td>LP014</td>
<td>IA</td>
<td>2005</td>
<td>Thailand</td>
<td>EF207320</td>
<td>NA</td>
</tr>
<tr>
<td>HAV5</td>
<td>IA</td>
<td>2005</td>
<td>Uruguay</td>
<td>EU131373</td>
<td>Garcia-Aguirre &amp; Cristina (2008)</td>
</tr>
<tr>
<td>VBA-07</td>
<td>IA</td>
<td>1988</td>
<td>Russia</td>
<td>EU251188</td>
<td>Maidaniuk et al. (1993)</td>
</tr>
<tr>
<td>LA</td>
<td>IA</td>
<td>1974</td>
<td>USA</td>
<td>K02990</td>
<td>Najarian et al. (1985)</td>
</tr>
<tr>
<td>GBM/WT</td>
<td>IA</td>
<td>1976</td>
<td>Germany</td>
<td>X75215</td>
<td>Graff et al. (1994)</td>
</tr>
<tr>
<td>DL3</td>
<td>IA</td>
<td>NA</td>
<td>NA</td>
<td>AF512536</td>
<td>NA</td>
</tr>
<tr>
<td>H2</td>
<td>IA</td>
<td>NA</td>
<td>NA</td>
<td>EF406357</td>
<td>NA</td>
</tr>
<tr>
<td>HAV-203</td>
<td>IB</td>
<td>1992</td>
<td>Brazil</td>
<td>AF268396</td>
<td>NA</td>
</tr>
<tr>
<td>HM-175</td>
<td>IB</td>
<td>1976</td>
<td>Australia</td>
<td>M14707</td>
<td>Cohen et al. (1987)</td>
</tr>
<tr>
<td>MBB</td>
<td>IB</td>
<td>1978</td>
<td>Africa, Mediterranean</td>
<td>M20273</td>
<td>Paul et al. (1987)</td>
</tr>
<tr>
<td>LA1</td>
<td>IB</td>
<td>NA</td>
<td>NA</td>
<td>AF314208</td>
<td>NA</td>
</tr>
<tr>
<td>CF53/Berne</td>
<td>IIA</td>
<td>1979</td>
<td>France</td>
<td>AY644676</td>
<td>Lu et al. (2004)</td>
</tr>
<tr>
<td>SLF88</td>
<td>IIB</td>
<td>1988</td>
<td>Sierra Leone</td>
<td>AY644670</td>
<td>Ching et al. (2002)</td>
</tr>
<tr>
<td>HA-JNG04-90</td>
<td>IIA</td>
<td>1990</td>
<td>Japan</td>
<td>AB279732</td>
<td>Endo et al. (2007b)</td>
</tr>
<tr>
<td>HA-JNG08-92</td>
<td>IIA</td>
<td>1992</td>
<td>Japan</td>
<td>AB279733</td>
<td>Endo et al. (2007b)</td>
</tr>
<tr>
<td>HAJ95-8</td>
<td>IIA</td>
<td>1995</td>
<td>Japan</td>
<td>AB279734</td>
<td>Endo et al. (2007b)</td>
</tr>
<tr>
<td>CP-IND</td>
<td>IIA</td>
<td>1995</td>
<td>India</td>
<td>DQ991029</td>
<td>NA</td>
</tr>
<tr>
<td>GBS-IND</td>
<td>IIA</td>
<td>2003</td>
<td>India</td>
<td>DQ991030</td>
<td>NA</td>
</tr>
<tr>
<td>PN-IND</td>
<td>IIA</td>
<td>1992</td>
<td>India</td>
<td>EU011791</td>
<td>Joshi et al. (2008)</td>
</tr>
<tr>
<td>SIM27</td>
<td>IIA</td>
<td>2007</td>
<td>India</td>
<td>FJ227135</td>
<td>NA</td>
</tr>
<tr>
<td>IND-HAV-06F</td>
<td>IIA</td>
<td>2006</td>
<td>India</td>
<td>FJ360730</td>
<td>Kulkarni et al. (2009)</td>
</tr>
<tr>
<td>IND-HAV-07F</td>
<td>IIA</td>
<td>2007</td>
<td>India</td>
<td>FJ360731</td>
<td>Kulkarni et al. (2009)</td>
</tr>
<tr>
<td>IND-HAV-08F</td>
<td>IIA</td>
<td>2008</td>
<td>India</td>
<td>FJ360732</td>
<td>Kulkarni et al. (2009)</td>
</tr>
<tr>
<td>IND-HAV-95F</td>
<td>IIA</td>
<td>1995</td>
<td>India</td>
<td>FJ360733</td>
<td>Kulkarni et al. (2009)</td>
</tr>
<tr>
<td>IND-HAV-99F</td>
<td>IIA</td>
<td>1999</td>
<td>India</td>
<td>FJ360734</td>
<td>Kulkarni et al. (2009)</td>
</tr>
<tr>
<td>IND-HAV-97F</td>
<td>IIA</td>
<td>1997</td>
<td>India</td>
<td>FJ360735</td>
<td>Kulkarni et al. (2009)</td>
</tr>
<tr>
<td>JNG06-90F</td>
<td>IIB</td>
<td>1990</td>
<td>Japan</td>
<td>AB258387</td>
<td>Endo et al. (2007a)</td>
</tr>
<tr>
<td>HAJ85-1</td>
<td>IIB</td>
<td>1985</td>
<td>Japan</td>
<td>AB279735</td>
<td>Endo et al. (2007b)</td>
</tr>
<tr>
<td>SHAV</td>
<td>V</td>
<td>NA</td>
<td>India</td>
<td>EU140838</td>
<td>NA</td>
</tr>
</tbody>
</table>

*Provided according to ‘Country’ field of GenBank record or the original publication.
approach could increase the resolution and reliability of molecular epidemiology studies in HAV and yield data on the actual prevalence of recombination in HAV.

Recent studies have demonstrated that recombination is highly prevalent in many picornavirus genera and provides a virtually independent evolutionary history of genome fragments even on a short timescale (Lukashev, 2010). HAV is unusual among other picornaviruses and does not share many genetic traits with them. We demonstrated that HAV nevertheless exhibits regular recombination, which is confined to the genotypes.

METHODS

Sequences. All complete HAV sequences available in GenBank were extracted. Sequences of attenuated or otherwise artificially modified viruses were excluded from the analysis, resulting in a set of 39 full-genome HAV sequences (Table 1). The coding genomic region was aligned based on the translated polyprotein sequence using Muscle (Edgar, 2004). The IRES (larger part of the 5’NTR spanning nt 170–730 of the genome) was aligned separately and joined to the aligned coding sequences. 3’NTR sequences were omitted from the analysis because this short and highly conserved genomic region carries little phylogenetic information. Sequence handling was performed with BioEdit v.7.0.5.2 software (Hall, 1999). We also made an alignment of 104 partial HAV sequences for VP1–2A (nt 2940–3420) and 3CD (nt 5385–5979) genomic regions, which included the corresponding fragments of the 39 full-genome sequences (Table 1) and partial HAV sequences published previously (Supplementary Table S1; Endo et al., 2007a). All sequence positions here and below are given according to the complete genome sequence of strain HM-175 (GenBank accession no. M14707). We also made an alignment of all non-polio HEV-C complete genomes and 12 wild-type poliovirus genomes (sharing less than 15% nucleotide sequence similarity to Sabin strains in V1) available in GenBank (a total of 69 sequences; alignment available upon request) for reference in phylogenetic compatibility matrices available in GenBank (a total of 69 sequences; alignment available upon request).

Genetic analysis. Similarity plot and bootscanning analysis were performed using SimPlot v.3.5.1 (Lole et al., 1999). Phylogenetic compatibility matrices were created using Simmonics v.1.6 (Simmonds & Smith, 1999). Phylogenies were calculated using a Bayesian likelihood-based algorithm implemented in BEAST v.1.5.4 (Drummond & Rambaut, 2007). The SRD06 codon-based substitution model (Shapiro et al., 2006) was used with a strict clock. A coalescent constant size tree prior was used because in preliminary runs it yielded a higher effective sample size (ESS) of the posterior probability values than coalescent: exponential growth or coalescent: expansion growth tree priors. A general time-reversible nucleotide substitution model with a site heterogeneity model assuming gamma-distributed substitution rates across sites and allowing for invariable sites was used for IRES phylogeny. Each analysis was run for 10–50 million generations to achieve an ESS >200. The trees were sampled every 1000 generations. Trees were annotated with TreeAnnotator v.1.5.4 with a burn-in of 20% and visualized with FigTree v.1.3.0 (http://tree.bio.ed.ac.uk/software/figtree/). Tree roots were calculated by BEAST software, and no outgroup was used for rooting. An unrooted consensus tree was created using the CONSENSE program of the PHYLIP package (Felsenstein, 1989) with ‘strict consensus type’ option. Distance matrices were calculated using MEGA v.4.0 (Tamura et al., 2007).

REFERENCES


