Short Communication

Analysis of sequential hepatitis A virus strains reveals coexistence of distinct viral subpopulations

Mauro Costa-Mattioli,1 Esteban Domingo2 and Juan Cristina2,3

1Department of Biochemistry and McGill Cancer Center, McGill University, Montreal, Quebec, Canada H3G 1Y6
2Centro de Biología Molecular ‘Severo Ochoa’ (CSIC-UAM), Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain
3Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, Facultad de Ciencias, Iguá 4225, 11400 Montevideo, Uruguay

Hepatitis A virus (HAV) is a hepatotropic member of the family Picornaviridae. Despite a remarkable antigenic stability, recent results have shown that HAV exists in vivo and in cell culture as distributions of genetically related, non-identical variants, referred to as quasispecies. To gain insight into HAV evolution over time in a specific geographical region, genotype I consensus sequences from strains isolated in France in consecutive years were studied. Phylogenetic neighbour-joining method and a non-hierarchical partition analysis, designed to analyse viral quasispecies, indicate that at least five distinct subpopulations of HAV were identified in the course of the disease episode. Strikingly, over time, different subpopulations cycled in dominance. The coexistence of distinct subpopulations whose frequency varies with time is consistent with quasispecies dynamics, and suggests that variation in the dominant HAV population may provide HAV adaptability without being reflected in significant antigenic variation.

The complexities of genetic data obtained from RNA virus quasispecies populations may not be accurately described by any single analytical tool (Baccam et al., 2001). Over time, RNA virus evolution is conditioned by perturbations of population equilibrium, which may not be equal among individual hosts, and therefore, multiple viral sublineages may rapidly be established that differ in the number of rounds of replication (and history of environmental perturbations), and may co-circulate in the same geographical area.

To study HAV evolution over time in a specific geographical region, we have analysed the highly variable region, VP1, of HAV strains genotype I. The strains studied were isolated in France from 1983 to 2001 (Costa-Mattioli et al., 2002; see also Supplementary Table 1 available in JGV Online). We used two methods, one based on phylogenetic distance, neighbour-joining (NJ; Saitou & Nei, 1987), and the other is a non-hierarchical method developed to study closely related components of mutant spectra of viral quasispecies (PAQ; Baccam et al., 2001).

Nucleotide sequences of the entire VP1-coding region were aligned using the CLUSTAL W program (Thompson et al., 1994). The program PAQ (Baccam et al., 2001) was adapted to compare consensus HAV VP1 sequences and to group HAV VP1 genes that were most similar. The program uses the Hamming distance (number of nucleotide differences) to measure the distances between VP1 gene sequences, and a
non-hierarchical clustering method to identify discrete and cohesive partitions of closely related sequences. The optimal output maximizes the number of variants contained within the partition, while minimizing the partition radius and any overlap between partitions. However, PAQ does not select a predetermined number of partitions among groups, and overlap between groups is allowed. The basic assumption of the program is that sequences separated by the fewest genetic differences are more similar and, thus, should be grouped together (Baccam et al., 2001). Application of partitions with a radius of 60, using strain HAV 9F(1994) as an outgroup, generated five distinct, non-overlapping groups, designated clades 1–5 (Table 1). Interestingly, clades 1–4 were composed of strains isolated from patient’s sera, while clade 5 was exclusively composed of strains isolated from sludge (Fig. 1 and Table 1). Whether the HAV environmental strains found clustered in clade 5 represent a potential source of infection or are just more resistant to environmental conditions remains to be determined.

**Table 1. Partition of genotype I VP1 sequences**

<table>
<thead>
<tr>
<th>Clade no.</th>
<th>No. strains assigned</th>
<th>Strain in centre</th>
<th>Compactness†</th>
<th>Colour assigned to each clade†</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>BF01</td>
<td>681-11</td>
<td>Violet</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>WF93</td>
<td>346-20</td>
<td>Pink</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1F85</td>
<td>83-66</td>
<td>Green</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2F84</td>
<td>719-33</td>
<td>Blue</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>BOUE4</td>
<td>54-50</td>
<td>Red</td>
</tr>
</tbody>
</table>

*Defined in the text; calculated using the PAQ program according to equation (1) in Baccam et al. (2001).
†Colours assigned to each clade are the same as in Figs 1 and 2.

PAQ analysis has permitted a comparison of the compactness (which is defined by the mean distance between the centre strain in the cluster and all other variants within the group; Baccam et al., 2001) among different subpopulations. For instance, the compactness for clades 3 and 5
(values of 83·66 and 54·50, respectively) is significantly lower than that obtained for clades 1 or 4 (values of 681·11 and 719·33, respectively, see Table 1).

To compare the results found with PAQ with those obtained using hierarchical methods of clustering, a matrix for the Kimura two-parameter model (Felsenstein, 1993) was created for all sequences used to compute neighbour-joining trees. The robustness of each node was assessed by bootstrap resampling (500 pseudoreplicates). These methods were implemented by using software from the MEGA2 program (Kumar et al., 1994). Consistent with our previous data (Costa-Mattioli et al., 2002), all the strains analysed were assigned to two well-defined subgenotypes, namely, IA and IB (Fig. 2). In addition, three clades among subgenotype IA and two among subgenotype IB were identified. Taken together these data suggest that there is a high correlation between the lineages identified by the hierarchical (NJ) method and the clades identified by PAQ. Strains isolated from sludge, assigned to a specific clade, also showed a very close genetic relatedness when studied by this hierarchical method (see Fig. 2). A similar relationship among isolates was obtained using the p-distance model and the UPGMA method for constructing phylogenetic trees (Sneth & Sokal, 1973) (not shown; see Supplementary Fig. S1 available in JGV Online).

Different hierarchical phylogenetic methods are increasingly utilized to analyse the evolutionary relationship among viral sequences, and they make use of different computer programs such as PHYLIP (Felsenstein, 1993), PAUP (Swofford, 1999) or MEGA (Kumar et al., 1994). These methods belong to the category of agglomerative methods, and merge data to form clusters that grow larger in a process termed ‘chaining’. Chaining is adequate at recognizing mutually exclusive clusters but may not be satisfactory at recognizing potential overlapping clusters of sequences. Besides, genetic recombination, in which the frequency is increasing within many RNA viruses (Agol, 2002) including HAV (Costa-Mattioli et al., 2003), may not be adequately modelled under the branching assumptions of phylogenetic reconstruction (Baccam et al., 2001). Recent developments in analyses of RNA viral quasispecies using PAQ have permitted discerning co-circulating lentiviral subpopulations (Baccam et al., 2001, 2003). Although, PAQ was designed specifically to analyse viral quasispecies (Baccam et al., 2001), it can be extended to analyse other types of sequence data, and here we have used it to compare epidemiologically related, consensus HAV sequences. Remarkably, both phylogenetic and partition analyses from the VP1 region identified different subpopulations of HAV variants that coexist in time (during 1993, 1995, 1997 or 1999) and in different environments. Interestingly, clades isolated from different years, reemerged and were even associated with epidemic strains, such as those isolated in 1985 and 1999 (Figs 1 and 2). These data suggest that beyond mutations and genetic recombination, HAV exploits this variation strategy in dominance to promote and ensure their survival.

A high correlation between partition and phylogenetic groupings of variants was also observed by Baccam et al. (2003). In particular, coexisting subpopulations have been extensively documented for HIV type 1 (Shapshak et al., 1999; for review see Papatheodoropoulou et al., 2003 and references therein) and Equine infectious anaemia virus (Baccam et al., 2003). Multiple coexisting subpopulations may occupy different regions on a fitness landscape to allow the virus to adapt rapidly to changes in the landscape topology. This may be especially relevant in modelling reservoirs of virus and the emergence of virus variants. The coexisting populations identified in the present study are consistent with the presence in each HAV isolate of a mutant spectrum (Sanchez et al., 2003a), which provides a repertoire of variants that, while constituting a minority in an infected individual, may become dominant following transmission to a new host individual. These findings fit the general picture of quasispecies dynamics (Domingo et al., 2001), with the salient antigenic stability of HAV that is probably related to structural constraints of the viral capsid (Sanchez et al., 2003b).

Acknowledgements
J.C. acknowledges support from Ministerio de Educaci´on y Ciencia of Spain through grant SAB2003-0124. Work in Madrid supported by grant BMC2001.1823CO2-01 and Fundacio´n Ramo´n Areces. M. C.-M. is supported by a CIHR post-doctoral fellowship.

References


M. Costa-Mattioli, Esteban Domingo and Juan Cristina