Short Communication

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

Mauro Costa-Mattioli,[1] Esteban Domingo[2] and Juan Cristina[2,3]

[1]Department of Biochemistry and McGill Cancer Center, McGill University, Montreal, Quebec, Canada H3G 1Y6

[2]Centro de Biología Molecular 'Severo Ochoa' (CSIC-UAM), Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

[3]Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, Facultad de Ciencias, Iguá 4225, 11400 Montevideo, Uruguay

Correspondence
Juan Cristina
cristina@cin.edu.uy

Received 26 June 2005
Accepted 4 October 2005

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.

Hepatitis A virus (HAV) is a hepatotropic member of the family *Picornaviridae* (van Regenmortel *et al.*, 2000). Despite an overall physical and epidemiological similarity to enteroviruses, the structure of HAV, its tissue tropism and genetic distance from other members of the family *Picornaviridae*, indicate that HAV is unique within this family (Ticehurst *et al.*, 1989; Palmenberg, 1989; Wimmer & Murdin, 1991; reviewed by Costa-Mattioli *et al.*, 2003). Hepatitis A viruses have been classified into four human (I, II, III and VII) and three simian (IV, V and VI) genotypes (Robertson *et al.*, 1992; Costa-Mattioli *et al.*, 2003). Recent evidence has suggested that, as other RNA viruses, HAV exists *in vivo* as distributions of closely related variants referred to as quasispecies (Sanchez *et al.*, 2003a). Quasispecies dynamics is characterized by the continuous generation of variant viral genomes, competition among them and selection of the fittest mutant distributions in any given environment. Understanding the principles that shape the evolution of viral quasispecies is becoming increasingly important to be able to model disease progression and to design preventive and therapeutic strategies to control viral disease (Domingo *et al.*, 2001).

A figure of the phylogenetic analysis of the complete VP1 region using the p-distance model and the UPGMA method, and a table showing French HAV strains examined in this study are available as supplementary material in JGV Online.

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

Table 1. Partition of genotype I VP1 sequences

Clade no.	No. strains assigned	Strain in centre	Compactness*	Colour assigned to each clade†
1	20	BF01	681·11	Violet
2	6	WF93	346·20	Pink
3	4	19F85	83·66	Green
4	4	2F84	719·33	Blue
5	3	BOUE4	54·50	Red

*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.

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.

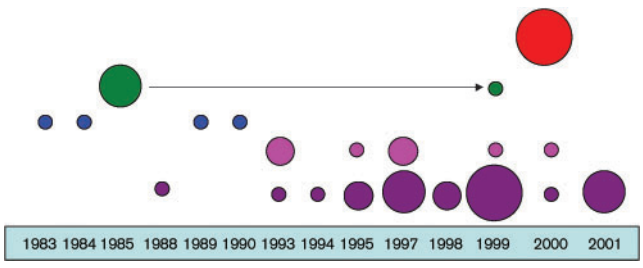

Fig. 1. Partition analysis identified different coexisting subpopulations of genotype I VP1 variants. The year of isolation is shown in the box at the bottom. Circles depict groups (clades) defined by PAQ (Baccam *et al.*, 2001). The relative size of each circle represents the proportion of genomes contained within each clade at a given time point. Related clades are indicated by colours that correspond to the same colours of Fig. 2 and Table 1. The arrow exemplifies how sequences detected in a specific clade may remain undetectable for years and be isolated later. Note also the extensive coexistence of different clades at several time points.

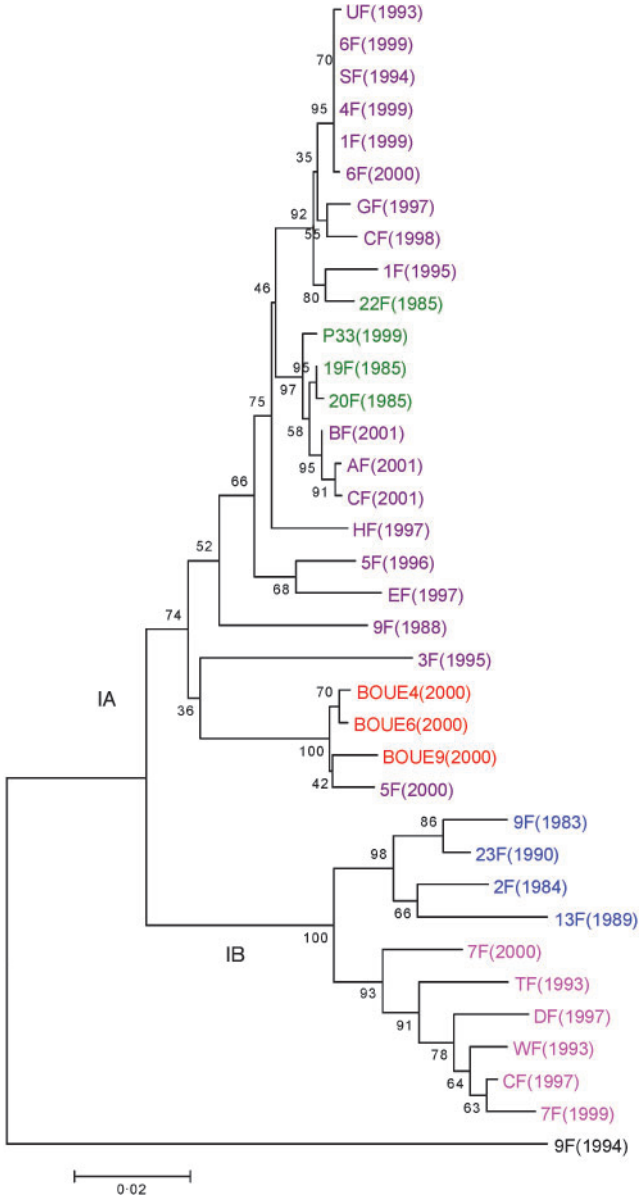

Fig. 2. Phylogenetic analysis of the complete VP1 region using the Kimura two-parameter distance model and the neighbour-joining method. HAVs are shown by their name and the year of isolation is shown in parentheses. Subgenotypes IA and IB are indicated at the corresponding node. The numbers at each node indicate bootstrap percentages after 500 replications of bootstrap sampling. The bar at the bottom denotes distance. Each strain has been assigned a colour corresponding to each clade identified by PAQ and shown in Table 1. Strain 9F(1994) was used as an outgroup. The viral isolates have been described in Costa-Mattioli *et al.* (2002) (see also Supplementary Table 1).

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 (Sneath & 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 Papathanasopoulos *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 Educación y Ciencia of Spain through grant SAB2003-0124. Work in Madrid supported by grant BMC2001.1823CO2-01 and Fundación Ramón Areces. M. C.-M. is supported by a CIHR post-doctoral fellowship.

References

Agol, V. A. (2002). Picornavirus genome: an overview. In *Molecular Biology of Picornaviruses*, pp. 127–148. Edited by B. Semler & E. Wimmer. Washington, DC: ASM Press.

Baccam, P., Thompson, R. J., Fedrigo, O., Carpenter, S. & Cornette, J. L. (2001). PAQ: partition analysis of quasispecies. *Bioinformatics* **17**, 16–22.

Baccam, P., Thompson, R. J., Li, Y. & 7 other authors (2003). Subpopulations of equine infectious anemia virus Rev coexist in vivo and differ in phenotype. *J Virol* **77**, 12122–12131.

Costa-Mattioli, M., Cristina, J., Romero, H. & 12 other authors (2002). Molecular evolution of hepatitis A virus: a new classification based on the complete VP1 protein. *J Virol* **76**, 9516–9525.

Costa-Mattioli, M., Ferre, V., Casane, D. & 7 other authors (2003). Evidence of recombination in natural populations of hepatitis A virus. *Virology* **311**, 51–59.

Domingo, E., Biebricher, C., Eigen, M. & Holland, J. J. (2001). *Quasispecies and RNA Virus Evolution: Principles and Consequences.* Austin, TX: Landes Bioscience.

Felsenstein, J. (1993). PHYLIP (phylogeny inference package) version 3.572c. Department of Genome Sciences, University of Washington, Seattle, USA.

Kumar, S., Tamura, K. & Nei, M. (1994). MEGA: molecular evolutionary genetics analysis software for microcomputers. *Comput Appl Biosci* **10**, 189–191.

Palmenberg, A. C. (1989). Sequence alignments of picornaviral capsid proteins. In *Molecular Aspects of Picornaviral Infection and Detection.* Edited by B. L. Semler & E. Ehenfeld. Washington, DC: American Society for Microbiology.

Papathanasopoulos, M. A., Hunt, G. M. & Tiemessen, C. T. (2003). Evolution and diversity of HIV-1 in Africa – a review. *Virus Genes* **26**, 151–163.

Robertson, B. H., Jansen, R. W., Khanna, B. & 10 other authors (1992). Genetic relatedness of hepatitis A virus strains recovered from different geographical regions. *J Gen Virol* **73**, 1365–1377.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Sanchez, G., Bosch, A., Gomez-Mariano, G., Domingo, E. & Pinto, R. M. (2003a). Evidence for quasispecies distributions in the human hepatitis A virus genome. *Virology* **315**, 34–42.

Sanchez, G., Bosch, A. & Pinto, R. M. (2003b). Genome variability and capsid structural constraints of hepatitis A virus. *J Virol* **77**, 452–459.

Shapshak, P., Segan, D. M., Crandall, K. A. & 7 other authors (1999). Independent evolution of HIV type 1 in different brain regions. *AIDS Res Hum Retroviruses* **15**, 811–820.

Sneath, P. H. A. & Sokal, R. R. (1973). *Numerical Taxonomy*. San Francisco: Freeman.

Swofford, D. L. (1999). PAUP*. Phylogenetic Analysis Using Parsimony (and Other methods), version 4. Sunderland, MA: Sinauer Associates.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Ticehurst, J., Cohen, J. I., Feinstone, S. M., Purcell, R. H., Jansen, R. W. & Lemon, S. M. (1989). Replication of hepatitis A virus: new ideas from studies with cloned cDNA. In *Molecular Aspects of Picornaviral Infection and Detection*, pp. 27–50. Edited by B. L. Semler & E. Ehenfeld. Washington, DC: American Society for Microbiology.

Van Regenmortel, M. H. V. (2000). In *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.

Wimmer, E. & Murdin, A. (1991). Hepatitis A and the molecular biology of picornaviruses: a case for a new genus of the family *Picornaviridae*. In *Viral Hepatitis and Liver Disease*, pp. 1–41. Edited by F. B. Hollinger, S. M. Lemon & H. S. Margolis. Baltimore: Williams & Wilkins.