Evidence of intratypic recombination in natural populations of hepatitis C virus

Rodney Colina,1 Didier Casane,2 Silvia Vasquez,3 Laura García-Aguirre,1 Ausberto Chunga,4 Héctor Romero,5 Baldip Khan6 and Juan Cristina1

1Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, Facultad de Ciencias, Iguá 4225, 11400 Montevideo, Uruguay
2Populations, Génétique et Evolution, CNRS, 91198 Gif-sur-Yvette, France
3Laboratorio de Biología Molecular, Centro Nuclear RACSO, Instituto Peruano de Energía Nuclear IPEN, Av. Canadá 1470, San Borja, Apartado 1687, Lima 41, Peru
4Servicio de Inmunología, Hospital Nacional Edgardo Rebagliati Martins HNERN, Domingo Cueto s/n, Jesús María, Lima 11, Peru
5Laboratorio de Organización y Evolución del Genoma. Instituto de Biología. Facultad de Ciencias. Iguá 4225, 11400 Montevideo, Uruguay
6Division of Human Health, International Atomic Energy Agency, Wagramerstrasse 5, 1400 Vienna, Austria

Hepatitis C virus (HCV) has high genomic variability and, since its discovery, at least six different types and an increasing number of subtypes have been reported. Genotype 1 is the most prevalent genotype found in South America. In the present study, three different genomic regions (5′UTR, core and NS5B) of four HCV strains isolated from Peruvian patients were sequenced in order to investigate the congruence of HCV genotyping for these three genomic regions. Phylogenetic analysis using 5′UTR–core sequences found strain PE22 to be related to subtype 1b. However, the same analysis using the NS5B region found it to be related to subtype 1a. To test the possibility of genetic recombination, phylogenetic studies were carried out, revealing that a crossover event had taken place in the NS5B protein. We discuss the consequences of this observation on HCV genotype classification, laboratory diagnosis and treatment of HCV infection.

INTRODUCTION

Hepatitis C virus (HCV) is the major causative agent of post-transfusion hepatitis and parenterally transmitted, sporadic non-A, non-B hepatitis throughout the world (Alter & Seeff, 2000). HCV is an enveloped RNA virus classified in the family Flaviviridae. HCV has high genomic variability and at least six different genotypes and an increasing number of subtypes have been reported (Simmonds, 1999).

RNA viruses exploit all known mechanisms of genetic variation to ensure their survival (Domingo & Holland, 1997). Their high rate of mutation and replication allow them to move through sequence space at a pace that often makes their DNA-based host’s evolution look glacial in comparison (Worobey & Holmes, 1999). Over the last two decades it has become increasingly clear that many RNA viruses add the capacity to exchange genetic material with one another. Thus, in addition to producing large amounts of the raw material of evolution (mutations), these viruses also possess mechanisms (recombination) that, in principle, allow them both to purge their genomes of accumulated deleterious changes (Muller, 1964) and to create or spread beneficial combinations of mutations in an efficient manner.

Until 1999, there was no evidence for recombination in flaviviruses, although the possibility had been considered (Blok et al., 1992; Kuno, 1997; Monath, 1994). Accordingly, the vast majority of work on flaviviruses, including vaccine studies and phylogenetic analyses in which genotypes were identified and sometimes correlated with disease severity (Chen et al., 1990; Leitmeyer et al., 1999; Rico-Hesse, 1990), has rested on the implicit assumption that evolution in the family Flaviviridae is clonal, with diversity generated through the accumulation of mutational changes.

Recent studies have shown this assumption to be invalid, as homologous recombination has now been demonstrated in pestiviruses (bovine viral diarrhoea virus) (Becher et al., 2001), flaviviruses (all four serotypes of dengue virus)
(Holmes et al., 1999; Tolou et al., 2001; Uzcategui et al., 2001; Worobey & Holmes, 1999), hepativiruses (GB virus C/hepatitis G virus) (Worobey & Holmes, 2001) and Japanese encephalitis or St Louis encephalitis virus (Twiddy & Holmes, 2003). There have been few reports on recombination between HCV strains of different genotypes (Kalina et al., 2002; Yun et al., 1996) and it has been suggested that these events are rare in vivo and that the resultant recombinants are usually not viable (Simmonds et al., 1994; Smith & Simmonds, 1997).

Selected HCV genome regions within the 5’UTR, core, E1 or NS5, which have been shown to be conserved within a given HCV genotype, are used for the classification of HCV strains (Simmonds et al., 1994; Simmonds, 1999). Most methods for direct HCV genotyping include amplification of different genome regions, such as the 5’UTR, core, E1 or NS4, by PCR with type-specific primers or by restriction fragment length polymorphism analysis of PCR products (Ohno et al., 1997; Okamoto et al., 1993; Stuyver et al., 1993, 1995). Indirect HCV genotyping may be achieved by demonstration of type-specific antibodies by ELISA (Dixit et al., 1995; Simmonds et al., 1993). Thus, present methods of HCV genotype identification do not take recombination into account.

Given the implications of recombination for virus evolution (Worobey & Holmes, 1999) and the development of vaccines, virus control programmes, patient management and antiviral therapies, it is clearly important to determine the extent to which recombination plays a role in HCV evolution. Recombination plays a significant role in the evolution of RNA viruses by creating genetic variation. For example, the frequent recovery of recombinant isolates of poliovirus (Georgescu et al., 1994; Kew & Nottay, 1984) that result from recombination involving vaccine strains shows that recombination has the potential to produce ‘escape mutants’ in nature as well as in experiments. Recently, recombination has also been detected in other RNA viruses for which multivalent vaccines are in use or in trials (Holmes et al., 1999; Suzuki et al., 1998; Worobey et al., 1999). We think the potential for recombination to produce new pathogenic hybrid strains needs to be carefully considered whenever vaccines are used or planned to control RNA viruses. Assumptions that recombination either does not happen or is unimportant in RNA viruses have a history of being proved wrong (Worobey & Holmes, 1999).

In previous studies, we subtyped 72 HCV strains isolated in South America (Colina et al., 1999; Vega et al., 2001; San Roman et al., 2002; Cristina et al., 2002) by limited sequencing of the 5’UTR region. In the present study, this work was extended to include sequencing of the core and NS5B regions in order to investigate the congruence of HCV genotype determinations among the different regions of the genome. We found congruent results in 97% of cases. However, we also found evidence for recombination between type 1 subtypes of HCV in the Peruvian population.

**METHODS**

**Serum samples.** Serum samples were obtained from 20 patients with chronic hepatic disease from the Hospital Nacional Edgardo Rebagliati Martins (Lima, Peru). In each case, patients were screened using an enzyme immunoassay (Innogenetics) and a confirmatory line immunoassay test (Innogenetics), according to the manufacturer’s instructions.

**RNA extraction, cDNA synthesis and amplification.** HCV RNA was extracted from 140 μl serum samples with the QIAamp viral RNA kit (Qiagen) according to the manufacturer’s instructions. The extracted RNA was eluted from the columns with 50 μl RNase-free water. cDNA synthesis and PCR amplification of the 5’UTR, core and NS5B regions were carried out as previously described (Chan et al., 1992; Norder et al., 1998). To avoid false positive results, the recommendations of Kwok & Higuchi (1989) were strictly adhered to. Amplicons were purified using the QIAquick PCR purification kit (Qiagen), according to the manufacturer’s instructions.

**Sequencing.** The primers used for amplification were used for sequencing the PCR fragments. The sequencing reaction was carried out using the Big Dye DNA sequencing kit (Perkin-Elmer) on a 373 DNA sequencer apparatus (Perkin-Elmer) or by manual sequencing using the Thermo Sequenase radiolabelled terminator cycle kit (Amersham).

**Sequence analysis.** The sequences for the 5’UTR plus core and NS5B regions were aligned using the CLUSTAL W program (Thompson et al., 1994). Using the MEGA program (Kumar et al., 1994), phylogenetic trees were created by the neighbour-joining method applied to the distance matrix obtained under the Kimura two-parameter model (Felsenstein, 1993). As a measure of the robustness of each node, we utilized the bootstrap method (1000 pseudo-replicates).

**Recombination analysis.** Putative recombinant sequences were identified with the SimPlot program (Lole et al., 1999), using concatenated (5’UTR plus core plus NS5B) sequences. This program is based on a sliding window method and constitutes a way of graphically displaying the coherence of the sequence relationships over the entire length of a set of aligned homologous sequences. The window width and the step size were set to 200 bp and 10 bp, respectively. Once the recombinant strain and strains representing possible parents were identified, the likely recombination breakpoint was determined by LARD (Holmes et al., 1999). Briefly, for every possible breakpoint, the sequence alignment was divided into two independent regions for which the branch lengths of a tree of the putative recombinant and its two parent sequences were optimized. The two results (likelihoods) obtained by using the separate regions were then combined to give a likelihood score for that breakpoint position, and the breakpoint position that yielded the highest likelihood was then compared, using a likelihood ratio test, to the likelihood obtained from the same data under a model that permitted no recombination. To assess whether the recombination model gave a significantly better fit to the data than the null hypothesis of no recombination, the likelihood ratios obtained using the real data were evaluated for significance against a null distribution of likelihood ratios produced by using the Monte Carlo simulation of sequences generated without recombination. Sequences were simulated 1000 times using the maximum-likelihood model parameters and sequence lengths from the real data using Seq-Gen (Rambaut & Grassly, 1997).
RESULTS

Phylogenetic analysis of HCV strains

In order to study the congruence of HCV genotype determinations among the different regions of the genome, 5'UTR, core and NS5B sequences from four HCV strains isolated in Peru were obtained. The 5'UTR and core sequences were aligned with those from 12 other strains representative of all six HCV types isolated elsewhere for which total sequences have been obtained. The origin of the sequences and the strains used are listed in Table 1. Once aligned, phylogenetic trees were created. As can be seen in Fig. 1(A), all HCV strains included in these studies clustered according to their genotype. Strains belonging to genotype 1 clustered together with the Peruvian isolates. Inside the main cluster of type 1 strains, two different lineages could be seen, supported by very high bootstrap values. One main line represented subtype 1a (Fig. 1A, upper part, and Table 1), while the other represented subtype 1b (Fig. 1A, middle).

The same analysis was performed using NS5B sequences. The results of these studies are shown in Fig. 1(B). As can be seen in the figure, all HCV strains included in these studies again clustered according to their genotype, and inside the main type 1 cluster, two different lineages, again supported by very high bootstrap values, were observed (Fig. 1B, top). However, strain PE22, assigned to genotype 1b in the 5'UTR plus core phylogenetic tree, was now assigned to genotype 1a in the NS5B phylogenetic analysis (Fig. 1B, top). This discrepancy between the results found with the 5'UTR plus core sequences and the NS5B sequences, supported by very high bootstrap values in both trees (Fig. 1A and B), could be explained if a recombination event had taken place between putative parental strains comparable with H77 (subtype 1a) and JK1 (subtype 1b).

Recombination analysis

To gain insight into a possible recombination event, a phylogenetic profile analysis was carried out for the Peruvian strain PE22 and the putative parental-like strains H77 (subtype 1a) and JK1 (subtype 1b). The results of these studies are shown in Fig. 2. As can be seen in the figure,
profile analysis of the putative parental-like (H77, JK1) and recombinant (PE22) strains showed a clearly visible point of recombination at position 677 of the analysed sequences, which corresponds to position 58 of the NS5B sequences included in this study (see Fig. 3). This position corresponds to position 8321 in the HCV genome of the putative parental strain H77.

To confirm these results, we employed the LARD method (Holmes et al., 1999). Simulations of sequence evolution under the null hypothesis (i.e. no recombination) gave strong statistical support for the alternative hypothesis of recombination (Fig. 4; \( P < 0.001 \)).

**DISCUSSION**

Congruent results from genotyping HCV with different genomic regions have been repeatedly reported in the past and it has commonly been believed that recombination events are selected against or that they generally do not generate viable strains (Prescott et al., 1997; Simmonds

---

**Fig. 2.** Phylogenetic profiles of HCV sequences. The γ-axis gives the percentage identity within a sliding window 200 bp wide centred on the position plotted, with a step size between plots of 10 bp, along the concatenated (5’UTR plus core plus NS5B) sequences. Comparison of PE22 with H77 (subtype 1a) and JK1 (subtype 1b) is shown. The vertical line shows the recombination point (nt 677).

**Fig. 3.** Alignment of the nucleotide sequences within the NS5B region. Alignment of the PE22 recombinant strain sequences with corresponding sequences of isolates JK1 and H77. The sequences shown in the figure correspond to nt 8263–8378, relative to strain H77 (AF009606, type 1a). Nucleotide identity to PE22 is indicated by a dash. An arrow shows the recombination point.
mosaic structure; for example, the intertypic recombinants
same recombination event and thus share an identical
form’ (RF), as also suggested by Kalinina
an HCV recombinant strain be designated a ‘recombinant
mutation has not yet been considered in this classification. By
into genotypes, subtypes and quasispecies, but recombin-
current HCV classification system, HCV strains are divided
was situated in the NS5B region (see Figs 2 and 3). In the
2002). The recombination point in our recombinant strain
2003; Vidal
other RNA viruses, such as human immunodeficiency virus
possible recombination breakpoints have been identified in
sometimes in genes encoding structural proteins (Costa-
The genome encoding the non-structural proteins but
[51x51]REFERENCES
et al., 1994; Viazov et al., 2000). However, an infectious
HCV chimera comprising the complete open reading frame
of a subtype 1b strain and the 5’- and 3’UTRs of a subtype 1a
strain has been constructed and is infectious in vivo (Yagani
et al., 1998). Recombination in other flaviviruses has now
been demonstrated on a number of occasions (Becher et al.,
2001; Worobey & Holmes, 2001; Worobey et al., 1999;
Twiddy & Holmes, 2003), and recently a natural inter-
genotypic recombinant (2k/1b) of HCV was identified in St
Petersburg (Russia) (Kalinina et al., 2002). Our phylogenetic
analyses based on two different genomic regions, 5’UTR–
core and NS5B, demonstrate the existence of natural
intrageneotypic HCV recombinant strains (1a/1b) circulat-
ing in the Peruvian population. The recombination break-
point for non-segmented positive-strand RNA viruses, such as
polioviruses and other picornaviruses (Santti et al., 1999;
Guillot et al., 2000; Kew et al., 2002), as well as members of the family Flaviviridae, are often located in the part of the
genome encoding the non-structural proteins but
sometimes in genes encoding structural proteins (Costa-
Mattioli et al., 2003; Martin et al., 2002). Moreover, several
possible recombination breakpoints have been identified in
other RNA viruses, such as human immunodeficiency virus
(HIV), and many more are being reported (Onafuwa et al.,
2003; Vidal et al., 2003; Strimmer et al., 2003; Najera et al.,
2002). The recombination point in our recombinant strain
was situated in the NS5B region (see Figs 2 and 3). In the current
HCV classification system, HCV strains are divided
into genotypes, subtypes and quasispecies, but recombin-
ation has not yet been considered in this classification. By
analogy with the nomenclature for HIV, we suggest that
an HCV recombinant strain be designated a ‘recombinant
form’ (RF), as also suggested by Kalinina et al. (2002). RF
strains with the same number are progeny resulting from the
same recombination event and thus share an identical
mosaic structure; for example, the intertypic recombinants
among subtypes 2k and 1b observed by Kalinina et al.
(2002) were described as RF1_2k/1b. Accordingly, we
suggest the designation RF2_1a/1b for the Peruvian strain
PE22 described herein.

ACKNOWLEDGEMENTS
This work was supported in part by the International Atomic Energy
Agency through Project ARCAL LVI (RLA/6/044). We acknowledge
support from CONICYT (Uruguay) for support through Fondo
Clemente Estable (Proyecto No. 8006).

Fig. 4. Distribution of the likelihood ratios expected by chance.
The distribution of likelihood ratios for the null hypothesis (i.e. no
recombination) is shown. The y-axis shows the number of simula-
lions. Likelihood ratios are shown at the bottom of the figure. The
arrow shows the likelihood ratio obtained for the real dataset for
the putative recombinant Peruvian strain (PE22).

REFERENCES
in hepatitis C virus infection: a perspective on long-term outcome.
Sem Liver Dis 20, 17–35.
between persisting pestivirus and a vaccine strain: generation of
cytopathogenic virus and induction of lethal disease. J Virol 75,
6256–6264.
Comparison of a dengue-2 virus and its candidate vaccine derivative:
sequence relationships with the flaviviruses and other viruses.
Chan, S. W., McOmish, F., Holmes, E. C., Dow, B., Peutherer, J. F.,
hepatitis C virus type and its phylogenetic relationship to existing
Colina, R., Azambuja, C., Uriarte, R., Mogdasy, C. & Cristina, J.
J Gen Virol 80, 1377–1382.
Evidence of recombination in natural populations of hepatitis A
Cristina, J., Mukomolov, S., Colina, R., Kalinina, O., Garcia, L.,
Dixit, V., Quan, S., Martin, P. & 7 other authors (1995). Evaluation
of a novel serotyping system for hepatitis C virus: strong correlation
with standard genotyping methodologies. J Clin Microbiol 33,
2978–2983.
Department of Genetics, University of Washington, Seattle, USA.
Georgescu, M. M., Delpeyroux, F., Tardypanit, M., Balanant, J.,
High diversity of poliovirus strains isolated from the central nervous
system from patients with vaccine-associated paralytic poliomylitis.
Guillot, S., Caro, V., Cuervo, N., Korotkova, E., Combescu, M.,
Persu, A., Aubert-Combescu, A., Delpeyroux, F. & Crainic, R.
(2000). Natural genetic exchanges between vaccine and wild
evidence for recombination in dengue virus. Mol Biol Evol 16,
405–409.


Dengue: the risk to developed and developing countries. Proc Natl Acad Sci U S A 91, 2395–2400.

The relation of recombination to mutational advance. Mut Res 1, 2–9.

Genetic recombination and its role in the development of the HIV-1 pandemic. AIDS 16, S3–S16.


Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73, 152–160.


Dengue: the risk to developed and developing countries. Proc Natl Acad Sci U S A 91, 2395–2400.

The relation of recombination to mutational advance. Mut Res 1, 2–9.

Genetic recombination and its role in the development of the HIV-1 pandemic. AIDS 16, S3–S16.


Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73, 152–160.


Dengue: the risk to developed and developing countries. Proc Natl Acad Sci U S A 91, 2395–2400.

The relation of recombination to mutational advance. Mut Res 1, 2–9.

Genetic recombination and its role in the development of the HIV-1 pandemic. AIDS 16, S3–S16.