

Short Communication **Evolution of hepatitis C virus in blood donors and their respective recipients**

Correspondence

J.-F. Cantaloube
jfc-ets-ap@gulliver.fr

Jean-François Cantaloube, Philippe Biagini, Houssam Attoui, Pierre Gallian, Philippe de Micco and Xavier de Lamballerie
Unité des Virus Emergents, Etablissement Français du Sang Alpes-Méditerranée, EA32/92-IFR48, Université de la Méditerranée, Faculté de Médecine, 27 Bd J. Moulin, F-13005 Marseille, France

Received 7 June 2002

Accepted 22 October 2002

Ahead-of-print 12 November 2002

This paper describes the study of hepatitis C virus (HCV) evolution in the largest cohort of HCV-infected blood donors (BDs)/blood recipients (BRs) reported to date (25 pairs). A molecular analysis of partial sequences in the E1 (envelope) and NS5-B (polymerase) genes was performed. Phylogenetic reconstruction showed that the evolution of dominant strains was qualitatively and quantitatively different in BDs and BRs. The evolutionary rate was significantly higher in BRs, in which, in addition, most substitutions observed were antonymous. These findings corroborate the hypothesis that a large part of virus evolution – which was evaluated to be equivalent to ~20 years of chronic evolution – is acquired during the early phase of infection. These findings should be taken into account for the modelling of the long-term evolution of HCV and their possible contribution to improve our understanding of HCV natural history is discussed.

During the second half of the twentieth century, several million persons have been contaminated by hepatitis C virus (HCV) following blood transfusion. Despite this extremely significant number of cases, virus evolution – in particular during the early stages of infection – remains poorly understood. The genome of HCV is replicated with the help of a viral RNA polymerase, which that lacks the proofreading activity of DNA polymerase complexes and therefore confers on the virus a high rate of genome evolution. A number of studies have been dedicated to the analysis of HCV evolution over time, principally by determining and comparing virus sequences at two different times of the infection in humans (Ogata *et al.*, 1991) or chimpanzees (Lu *et al.*, 2001), or virus sequences from different individuals infected by the same virus strain (McAllister *et al.*, 1998). In most of the cases, virus sequences could not be determined at the early stages of infection and, therefore, data obtained are mostly reflecting the phenomena occurring during chronic infection. We describe here a molecular analysis based on the study of the largest cohort of HCV-infected blood donors (BDs)/blood recipients (BRs) reported to date (25 pairs). It reveals qualitative and quantitative differences in the apparent evolution of the dominant strains from BDs and their respective BRs. The significance and possible contribution of

these findings to improve our understanding of HCV evolution are discussed.

We investigated HCV infections acquired following blood transfusion in the region of Marseilles (France) between 1982 and 1990. Blood samples from BDs and BRs were collected during the course of legal inquiries (September 1993 to January 2002). HCV viraemia was detected by a commercial PCR assay (Amplicor-HCV, Roche). For further molecular studies, RNA extraction from 200 µl plasma, reverse transcription using random hexaprimers and PCR in the E1-encoding region using type-specific primers were performed as reported previously (Cantaloube *et al.*, 2000). A semi-nested PCR was also performed in the NS5-B region using an annealing temperature of 54 °C with primers NS5-1 (sense, 5'-TATGAYACCCGYTGCTTTGAC-3') and NS5-2 (reverse, 5'-GAGGAGCAAGATGTTATCAGCTC-3') for primary amplification and primers NS5-1 and NS5-3 (reverse, 5'-GAATACCTGGTCATAGCCTCCG-3') for secondary amplification. Amplicons in the E1 and NS5 regions were sequenced directly with the amplification primers, the D-Rhodamine DNA Sequencing kit and an ABI Prism 377 sequencer (Perkin-Elmer). These sequences, together with sequences from local BDs, were assigned GenBank accession numbers AF515849–AF516066.

Nucleotide sequences from BDs and BRs in the E1 (positions 1029–1385 by reference to the ORF of strain D10750) and NS5-B (positions 7936–8274) regions were used for phylogenetic analyses. The corresponding amino acid sequences were aligned using CLUSTAL W, version 1.8 (Thompson *et al.*, 1994), with sequences from local BDs and sequences obtained from GenBank. Alignment of nucleotide sequences according to the amino acid sequence encoded was generated by TRANSALIGN, version 1.0 (Bioinformatics Laboratory, Australian National University, Canberra). Phylogenetic trees were built with the help of MEGA2 (Kumar *et al.*, 2001) using these nucleotide alignments, the p-distance for the calculation of genetic distances and the neighbour-joining method. The robustness of resulting groupings was tested using alternative algorithms for distance calculation (Jukes–Cantor and Kimura's 2-parameter methods) and 500 bootstrap replications.

In 25 pairs, (i) both infected BRs and putative BDs tested positive for the presence of HCV RNA in serum and (ii) phylogenetic relatedness in the E1 and NS5-B regions of strains infecting the BDs and BRs corroborated the epidemiological evidence of transmission between BDs and BRs (Fig. 1). The mean p-distance between BDs and their respective BRs was lower than that between non-related local isolates belonging to the same subtype: 0.024 ± 0.011 versus 0.095 ± 0.021 in the E1 region and 0.015 ± 0.007 versus 0.058 ± 0.017 in the NS5-B region. The case of pairs 1 and 2 (infected by very closely related strains) was discussed elsewhere (Cantaloube *et al.*, 2001). HCV subtype 1a was found in three pairs, subtype 1b in 14, type 2 in five, types 3, 4 and 5, respectively, in one pair each. The most important characteristics concerning each pair of BD/BR are reported in Table 1.

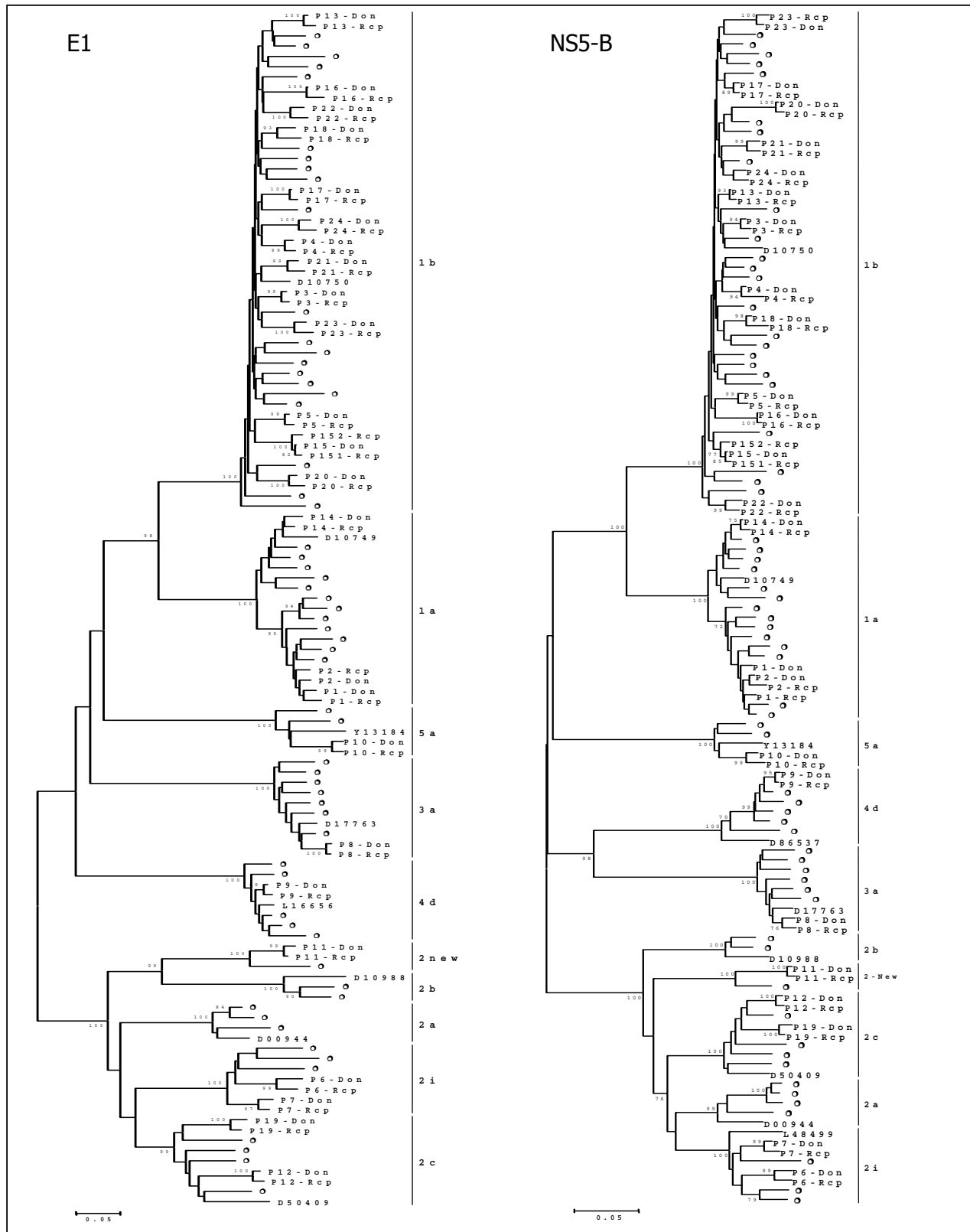


Fig. 1. Phylogenetic reconstruction in the E1 and NS5-B regions. Phylogenetic trees obtained using E1 (positions 1029–1385) and NS5-B (positions 7936–8274) HCV sequences from blood donors (Don), blood recipients (Rcp), local strains (○) and reference sequences from GenBank: D10749 (1a); D10750 (1b); D00944 (2a); D10988 (2b); D50409 (2c); L48499 (2i); D17763 (3a); L16656 and D86537 (4d); and Y13184 (5a). Trees were built using the pairwise distance and neighbour-joining algorithms. Bootstrap values above 70 % are indicated.

Table 1. General characteristics of HCV strains

Pair	Name	Collection date	Transfusion date	Genotype	p-distance (nucleotide substitutions per site) between BDs and BRs	
					E1	NS5-B
P1	Donor	Jul-98	Jan-87	1a	0.034	0.018
	Recipient	Oct-98				
P2	Donor	Oct-98	Jun-86	1a	0.031	0.018
	Recipient	Oct-98				
P3	Donor	Nov-96	Oct-87	1b	0.014	0.012
	Recipient	Sep-97				
P4	Donor	Apr-97	Apr-88	1b	0.020	0.021
	Recipient	Mar-98				
P5	Donor	Dec-98	Jan-82	1b	0.014	0.012
	Recipient	Dec-98				
P6	Donor	Jan-97	Dec-87	2i	0.048	0.027
	Recipient	May-98				
P7	Donor	Dec-98	Feb-89	2i	0.027	0.024
	Recipient	Dec-98				
P8	Donor	Jul-98	Dec-87	3a	0.011	0.021
	Recipient	Jun-98				
P9	Donor	Dec-97	Dec-86	4d	0.010	0.006
	Recipient	Feb-98				
P10	Donor	May-98	Oct-83	5a	0.024	0.024
	Recipient	Feb-98				
P11	Donor	Sep-93	Jan-90	2	0.017	0.009
	Recipient	Feb-99				
P12	Donor	May-99	Apr-90	2c	0.020	0.012
	Recipient	Mar-99				
P13	Donor	Sep-99	Apr-87	1b	0.017	0.006
	Recipient	Feb-00				
P14	Donor	Mar-00	Jan-88	1a	0.045	0.009
	Recipient	Mar-99				
P15-1	Donor	Jul-99	Dec-85	1b	0.011	0.006
	Recipient	Jan-00				
P15-2	Donor	Jul-99	Jan-76	1b	0.017	0.015
	Recipient	Feb-99				
P16	Donor	Dec-99	Mar-89	1b	0.020	0.003
	Recipient	Dec-99				
P17	Donor	Jun-99	Apr-85	1b	0.011	0.009
	Recipient	Jun-99				
P18	Donor	Jul-99	Feb-87	1b	0.045	0.021

	Recipient	Jan-00				
P19	Donor	Sep-00	Jul-85	2c	0.032	0.015
	Recipient	Nov-00				
P20	Donor	Apr-00	Jul-83	1b	0.025	0.009
	Recipient	Jun-94				
P21	Donor	Jan-00	Dec-86	1b	0.031	0.021
	Recipient	Jan-01				
P22	Donor	Jan-00	Dec-86	1b	0.034	0.021
	Recipient	Jan-01				
P23	Donor	Jan-00	Dec-86	1b	0.034	0.021
	Recipient	Jan-01				
P24	Donor	Mar-02	Jan-85	1b	0.031	0.021
	Recipient	Mar-02				

HCV evolution in BDs and BRs was studied using two approaches: (i) for a given pair, the genetic distance between BD and BR sequences and the node corresponding to their putative phylogenetic ancestor was calculated using the TREE EXPLORER program of MEGA and phylogenetic trees, including local sequences belonging to the same subtype (built as described above). The corresponding evolution rates were calculated, the duration of evolution in the BD (or BR) being evaluated as the period separating transfusion from the collection of the donor's (or recipient's) studied biological samples; (ii) for each pair of BD/BR, the sequence of the putative common ancestor was reconstructed. To achieve this and for a given subtype, E1 and NS5-B nucleotide alignments, including local strains of HCV, were submitted to the DNAML program of PHYLIP, version 3.572 (Felsenstein, 1989), to generate phylogenetic trees by the maximum-likelihood method. Using the resulting topology and the BASEML program of PAML, version 3.0 (Yang, 1997), ancestral nucleotide sequences at the interior nodes of the likelihood tree were inferred by the Bayesian method. This permitted us to calculate genetic distances and thus evolutionary rates.

Regardless of the method used or the region studied, the sequence of the ancestor was constantly found to be more closely related to the sequence of the corresponding BD strain than to the BR strain and thus calculated evolutionary rates were always higher in BRs than in BDs. This difference in evolutionary rates is statistically significant using either E1 or NS5-B sequences (Table 2). Consequently, the first noticeable finding of this study is the strong evidence that HCV evolution is quantitatively more important in BRs than in BDs. This is concordant with data reported by Allain *et al.* (2000), obtained using a totally different approach. These authors calculated evolutionary rates in six BDs and their BRs, based on the analysis of quasispecies in the E1/E2 region and reported lower rates in BDs than in BRs.

In the subsequent step of our study, synonymous substitutions per synonymous site (K_s) and antonymous

substitutions per antonymous site (K_a) were calculated by the method of Nei & Gojobori (1986) using MEGA. K_a and K_s were found to be significantly higher in BRs than in BDs in both the E1 and the NS5-B regions (Table 2). K_a / K_s ratios calculated in both regions are <1 (purifying selection). However, (i) these ratios are higher in BRs than in BDs – in the NS5 region, this difference is significant: $P < 0.05$, Fisher's exact test – implying that mechanisms driving the overall evolution in BDs and BRs are not exactly the same; (ii) a site-by-site analysis showed that among the 17 non-synonymous substitutions observed in the NS5-B, three occurred at position 2720 and four at position 2729. Therefore, it cannot be excluded that positive selection operated at very specific sites.

Further interpretation of these data requires referral to a model of HCV natural history and evolution. Several points must be considered: (i) the evolution observed in BRs corresponds to the successive phases of HCV early and, secondarily, chronic infection, while the evolution in BDs corresponds to a chronic phase of HCV infection only; (ii) the characteristics of virus evolution are comparable in BDs and BRs during the chronic phase of infection (implying that parameters of evolution during the chronic phase of infection of BDs can be applied to the chronic infection of BRs); (iii) the duration of the early phase of infection in BRs is negligible compared with that of the chronic infection; and (iv) ideally, the common phylogenetic ancestor in each pair corresponds to the virus at the moment of transfusion. In this case, evolution rates calculated as described above (duration of evolution evaluated as the period separating transfusion from collection of biological samples) are unbiased. However, it is possible that the divergence between the lineages infecting BDs and BRs occurred before transfusion. In that case, the observation of different evolution rates in BDs and BRs is still relevant but the calculated evolution rates are overestimated. Smith *et al.* (1997) studied the quasispecies of chronic HCV carriers in the same region of E1 as studied here and reported a 0.5 % mean divergence. Based on a rate of $\sim 10^{-3}$ nucleotide substitutions per site per year, this corresponds to a ~ 2.5 year period, which, combined with a mean period of 13 years between blood transfusion and sample collection, results in a possible 20 % overestimation of our calculated evolutionary rates.

Table 2. Evolution rates and substitutions parameters in the E1 and NSS-B regions

Method I is based on distances calculated from trees (MEGA); method II is based on distances calculated after reconstruction of ancestral sequences.

	Substitution analysis																
	Evolutionary rates (nucleotide substitutions per site per year)				Method I				Method II								
	Method I		Method II		K_a		K_s		K_a / K_s		K_a		K_s		K_a / K_s		
E1 region																	
Donor-Ancestor	6.83×10 ⁻⁴ ±4.53×10 ⁻⁴		5.15×10 ⁻⁴ ±4.08×10 ⁻⁴		0.003±0.004		0.028±0.020		0.11		0.002±0.003		0.018±0.017		0.1		
Recipient-Ancestor	1.28×10 ⁻³ ±6.33×10 ⁻⁴		1.44×10 ⁻³ ±9.23×10 ⁻⁴		0.006±0.005		0.043±0.025		0.14		0.006±0.005		0.047±0.030		0.13		
	}		}		}		}		}		}		}		}		
	*}		*}		*}		*}		*}		*}		*}		*}		
NSS-B region																	
Donor-Ancestor	4.19×10 ⁻⁴ ±2.25×10 ⁻⁴		2.99×10 ⁻⁴ ±3.30×10 ⁻⁴		0.0003±0.0008		0.020±0.013		0.015		0.0002±0.0002		0.015±0.015		0.01		
Recipient-Ancestor	8.36×10 ⁻⁴ ±5.57×10 ⁻⁴		9.47×10 ⁻⁴ ±6.94×10 ⁻⁴		0.002±0.002		0.031±0.018		0.06		0.0024±0.003		0.038±0.028		0.05		
	}		}		}		}		}		}		}		}		
	*}		*}		*}		*}		*}		*}		*}		*}		
Black box																	
E1 region					0.003±0.004		0.015±0.020		0.20		0.004±0.003		0.024±0.020		0.16		
NSS-B region					0.002±0.002		0.011±0.015		0.18		0.002±0.001		0.023±0.030		0.08		

*Significant differences between values according to Student's *t*-test and Mann-Whitney's U-test ($P<0.01$).

In this model, evolution during early stages of infection in BRs is not directly measurable. It constitutes a 'black box' that can be approached by subtracting the evolution attributable to the phase of chronic evolution (deduced from the study of BDs) from the total evolution observed in BRs. When this is realized, it appears that evolution in the black box represents a large part of the overall evolution observed in BRs (in our sample, 58.8 and 63.6 % of the evolution in E1 and NS5, respectively). Based on evolution rates measured in the E1 and NS5-B regions of BDs, it can be calculated that minimum 18 and 23 year periods of chronic infection, respectively, would be necessary to obtain such an evolution. It is noticeable that most of the anonymous substitutions observed during virus evolution in BRs occur in the black box: 56 % in the E1 and 93 % in the NS5-B. The K_a/K_s ratio calculated for evolution within the black box is significantly higher than that calculated during the chronic phase of infection, in both the E1 and the NS5 regions ($P < 0.05$, Fisher's exact test). Again, values observed are not evocative of a positive selection (< 0.25); this difference might, therefore, indicate that different mechanisms of selection are implicated during the acute and chronic phases of infection.

Taken together, these different findings can be used to sharpen the analysis of the long-term evolution of HCV. First, evolution is certainly dependent on the mode of virus transmission and the titre of inoculum. The evolution rate observed by Smith *et al.* (1997) following contamination by a batch of HCV serotype 1b-infected immunoglobulins was $\sim 7 \times 10^{-4}$ nucleotide substitutions per site per year in the E1 region and $\sim 4 \times 10^{-4}$ nucleotide substitutions per site per year in the NS5-B region. These values are of the same order of magnitude but are, nevertheless, lower than those reported here for strains belonging to the same subtype ($\sim 10^{-3}$ and $\sim 9 \times 10^{-4}$ nucleotide substitutions per site per year in the E1 and NS5-B regions, respectively). This could be due to the lower titre of the inoculum transmitted by infected immunoglobulins.

Second, it appears that a large part of virus evolution is acquired during the early stages of infection. This implies that, over a long period of time, virus evolution is most certainly related to the mean duration of chronic infection before transmission but also to the number of transmissions (or 'passages') from one individual to another. In this study, we evaluated that one passage is equivalent to ~ 20 years of chronic evolution. Therefore, the relative importance in virus evolution of the number of passages versus the duration of chronic infection preceding transmission depends on epidemiological characteristics of virus transmission. If virus transmission is a rare event, the mean period separating infection from transmission might be long and therefore an important parameter for virus evolution. In contrast, in the context of an epidemic, this mean period might be very short and therefore the predominant factor driving evolution might be the number of passages. This is important to consider for the modelling of long-term HCV evolution. In particular, it is

likely that the epidemiology of HCV infection during the past decade has been quite different from what it used to be previously. This should certainly be taken into consideration for the retrospective evaluation of the times of divergence between the different HCV lineages.

In conclusion, the molecular study of dominant HCV strains infecting BDs and BRs suggests that a large part of virus evolution occurs during the early and poorly understood stages of infection. These results were deduced from phylogenetic reconstruction and not from the direct analysis of virus sequences obtained at the different stages of HCV infection. In the future, the study of the early phase of infection will be key for a better understanding of the mechanisms driving HCV evolution.

ACKNOWLEDGEMENTS

The 'Unité des Virus Emergents' is an associated research unit of the 'Institut de Recherche pour le Développement' (IRD) and the Reference Center for Virology of the 'Etablissement Français du Sang' (EFS). This study was supported in part by the EFS.

REFERENCES

- Allain, J.-P., Dong, Y., Vandamme, A.-M., Moulton, V. & Salemi, M. (2000). Evolutionary rate and genetic drift of hepatitis C virus are not correlated with the host immune response: studies of infected donor–recipient clusters. *J Virol* **74**, 2541–2549.
- Cantaloube, J.-F., Venault, H., Zappitelli, J.-P., Gallian, P., Touinssi, M., Attoui, H., Biagini, P., de Lamballerie, X. & de Micco, P. (2000). Molecular analysis of HCV type 1 to 5 envelope gene: application to investigations of posttransfusion transmission of HCV. *Transfusion* **40**, 712–717.
- Cantaloube, J.-F., Gallian, P., Attoui, H., Biagini, P., de Micco, P. & de Lamballerie, X. (2001). Erroneous HCV genotype assignment by a hybridization typing assay in a case of posttransfusion HCV infection. *Transfusion* **41**, 429–430.
- Felsenstein, J. (1989). PHYLIP: Phylogeny Inference Package, version 3.572. *Cladistics* **5**, 164–166.
- Kumar, S., Tamura, N., Jakobsen, I. B. & Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.
- Lu, L., Nakano, T., Orito, E., Mizokami, M. & Robertson, B. H. (2001). Evaluation of accumulation of hepatitis C virus mutations in a chronically infected chimpanzee: comparison of the core, E1, HVR1, and NS5b regions. *J Virol* **75**, 3004–3009.
- McAllister, J., Casino, C., Davidson, F., Power, J., Lawlor, E., Yap, P. L., Simmonds, P. & Smith, D. B. (1998). Long-term evolution of the hypervariable region of hepatitis C virus in a common-source-infected

cohort. *J Virol* **72**, 4893–4905.

Nei, M. & Gojobori, T. (1986). Simple method for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418–426.

Ogata, N., Alter, H. J., Miller, R. H. & Purcell, R. H. (1991). Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc Natl Acad Sci U S A* **88**, 3392–3396.

Smith, D. B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P. L. & Simmonds, P. (1997). The origin of hepatitis C genotypes. *J Gen Virol* **78**, 321–328.

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.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**, 555–556.
