Short Communication

Molecular evidence of horizontal transmission of hepatitis C virus within couples

Isabel Maria Vicente Guedes de Carvalho-Mello,[1,2] José Eymard Medeiros Filho,[2] Michele Soares Gomes-Gouvêa,[2] Fernanda de Mello Malta,[2] Artur Trancoso Lopo de Queiróz,[3] João Renato Rebello Pinho[2] and Flair José Carrilho[2]

Correspondence
Isabel Maria Vicente Guedes de Carvalho-Mello
imvgcmello@butantan.gov.br

[1]Laboratório de Imunologia Viral, Instituto Butantan, Av. Vital Brasil 1500, 05503-900 São Paulo, SP, Brazil

[2]Department of Gastroenterology, University of São Paulo School of Medicine, Av. Dr. Enéas de Carvalho de Aguiar 500, 05403-000 São Paulo, SP, Brazil

[3]Instituto de Biociências, Universidade de São Paulo, 05508-090 São Paulo, SP, Brazil

Received 5 August 2009
Accepted 4 November 2009

Hepatitis C virus (HCV) transmission has decreased with the adoption of universal blood donor screening and social policies to reduce the risk of infection in intravenous drug users, but remains a worldwide health problem. The objective of this study was to evaluate the phylogenetic relationships among sequences from different HCV genomic regions from sexual partners of infected patients. Nine couples with a stable relationship and without other risk factors for HCV infection and 42 control patients were selected, and the NS3 and NS5B regions were analysed. Phylogenetic analysis showed that viruses from five of the couples had a common origin, clustering in the same monophyletic group, with bootstrap values greater than 70. For the other couples, monophyletic groups were observed, but without bootstrap support. Thus, using two different viral genome regions, a common source of infection was observed in both members of five couples. These data strongly support HCV transmission within couples.

Hepatitis C virus (HCV) infection constitutes a major health problem throughout the world and is related to severe liver cirrhosis, end-stage liver disease and hepatocellular carcinoma (Chevaliez & Pawlotsky, 2007).

The only member of the genus *Hepacivirus* of the family *Flaviviridae*, HCV is an enveloped virus with a positive-sense, single-stranded RNA genome of approximately 9.6 kb, encoding a single polyprotein of approximately 3000 aa that is cleaved into structural and non-structural proteins. HCV is classified into six genotypes, HCV-1 to -6, with each genotype further subdivided into subtypes, such as HCV-1a and -1b (Simmonds, 2004).

Although intrafamilial transmission is not well documented, several studies have reported sexual and vertical transmission and close contacts as sources of contamination (Cavalheiro, 2004; Cavalheiro *et al.*, 2009; Chiaramonte *et al.*, 1996; Keiserman *et al.*, 2003; Mastromatteo *et al.*, 2001). Sexual transmission is not well determined, although studies have indicated a low effectiveness of infection via this route (Akahane *et al.*, 1994). The estimated risk for sexual transmission is 0–0.6 % $year^{-1}$ for partners who maintain monogamous relationships for a long period of time, and 1 % $year^{-1}$ for those with multiple partners (Terrault, 2002). Some studies show high serum anti-HCV positivity rates among sex professionals, men who have sex with other men, and healthcare professionals (Nakashima *et al.*, 1992; Ndimbie *et al.*, 1996; Thomas *et al.*, 1995). If the index case is co-infected with human immunodeficiency virus (HIV), the sexual transmission risk is higher (Filippini *et al.*, 2001).

As with sexual and vertical transmission, occupational exposure has been documented but is rare (Alter, 2002; Alter *et al.*, 1999). At least 10 % of HCV-infected patients do not report exposure to any of the known risk factors, and these are considered sporadic cases.

Detailed molecular analysis may be able to establish the common ancestry of viruses circulating in a family and to support the hypothesis that close familial contact and sexual intercourse are important methods of HCV transmission. The present study had the objective of supporting the hypothesis of horizontal transmission within couples.

The GenBank/EMBL/DDBJ accession numbers for the HCV sequences determined in this study are EF406133–EF406248.

A maximum-likelihood tree constructed from the NS5B region and two supplementary tables showing patient data and primer sequences are available with the online version of this paper.

Patients chronically infected with HCV were recruited at the Gastroenterology Department of the University of São Paulo School of Medicine, São Paulo, Brazil. The study group comprised nine couples within a stable relationship infected with HCV genotypes 1 (n=12) and 3 (n=6), with a mean age of 49.8 years (see Supplementary Table S1, available in JGV Online, which shows the data on the epidemiological risk factors related to HCV transmission from their hospital files). As a control group, 42 HCV RNA-positive outpatients were selected. Controls harbouring genotypes 1 and 3 who did not have a close relationship with the members of the study group were selected. These patients had a mean age of 51.9 years, and the HCV transmission routes found in the study group were also included in the control group. Serum samples were collected for routine HCV detection and stored at −80 °C. HCV RNA was detected using commercially available kits (Amplicor 2.0; Roche Molecular Systems), and genotypes were determined by analysis of the 5′ untranslated region (5′UTR) (Campiotto *et al.*, 2005). Written informed consent was obtained from all patients, and the study protocol was approval by the ethics committee of the University of São Paulo.

Total RNA was extracted from plasma samples using a QIAamp Viral RNA kit (Qiagen) and cDNA was produced using reverse transcriptase (Invitrogen) with random hexamers. For amplification of the NS3 and NS5B regions, a nested PCR was performed and the products were sequenced using an ABI Prism 377 DNA Sequencer (Applied Biosystems). The primers are shown in Supplementary Table S2 (available in JGV Online).

All consensus sequences were obtained with the Phred, Phrap and Consed programs (Ewing & Green, 1998; Ewing *et al.*, 1998; Gordon *et al.*, 1998), and multiple alignment of NS5B and NS3 sequences was performed using CLUSTAL_X (Thompson *et al.*, 1997) and edited using BioEdit (version 7.0.5.3; Hall, 1999). The sequences of the NS5B region obtained were compared with reference sequences from GenBank.

Likelihood mapping analyses were performed using the TREE-PUZZLE program (Strimmer & von Haeseler, 1997) separately for NS5B, NS3, and NS5B plus NS3. For each analysis, 10 000 random quartets were evaluated.

The sequences obtained (18 study group and 42 control group NS5B sequences; 14 study group and 42 control group NS3 sequences) were analysed against the reference sequences, resulting in alignments of NS5B, NS3 and a concatenated NS5B–NS3. Topologies were estimated from each alignment using the distance and maximum-likelihood methods using PAUP* version 4.0b (Swofford, 2002). The best DNA substitution model (TrN+I+G for NS3 and NS5B plus NS3, and TrNef+I+G for NS5B) was estimated using MODELTEST version 3.06 (Posada & Crandall, 1998), and 1000 bootstrap replicates were performed (Felsenstein, 1985; Zharkikh & Li, 1995). Nucleotide sequence similarity analysis of all sequences for the NS5B and NS3 datasets was performed using MegAlign version 3.12 (DNASTAR).

The phylogenetic analysis showed that five patients had HCV genotype 1a, seven had genotype 1b and six had genotype 3a. Except for couple 9 (CF 09F, genotype 1b and CF 09M, genotype 1a), all couples showed concordance between genotype and subtype. The controls were selected as three groups of 14 patients with genotypes 1a, 1b and 3a.

The NS5B sequence of the nine couples was approximately 350 nt. However, for the NS3 region, we were unable to achieve successful amplification for two couples (CF 04 and CF 09).

We analysed separately the NS5B (344 nt), NS3 (619 nt), and NS3/NS5B (960 nt) regions of the couples and controls and compared them with sequences in GenBank.

Fig. 1 shows the results of the sequence likelihood mapping analysis. The NS5B region presented a phylogenetic signal of 92.9 % (Fig. 1a). The NS3 region presented a better signal (94.8 %). When we analysed the two regions conjointly, we observed an improvement in the signal, reaching 96.1 % (Fig. 1c).

The NS5B and NS3 sequences of couples and controls generated in this work were aligned with reference sequences to perform the phylogenetic analyses, using the genetic distance method with neighbour-joining algorithm (data not shown) and the maximum-likelihood method with a heuristic algorithm. Of the six analyses performed using sequences from the different regions and different methods, the sequences of couple 8 were the only ones to show the same origin, with a significant bootstrap, in all six analyses. Couples 3, 6 and 7 showed the same origin in

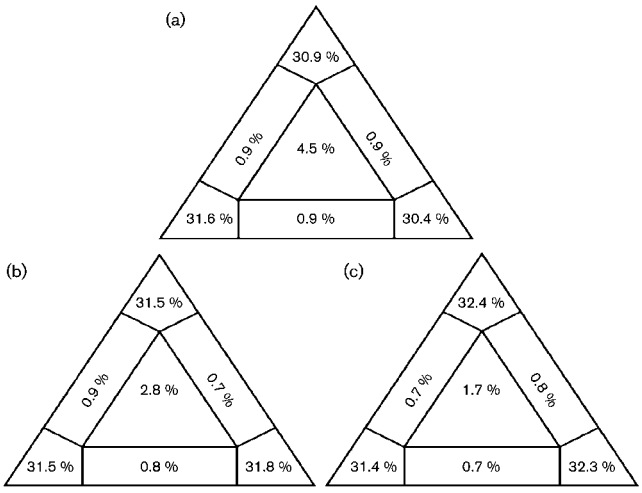

Fig. 1. Likelihood mapping of 10 000 quartets randomly selected from 77 sequences comprising 344 nt of the NS5B region (a), 72 sequences comprising 619 nt of the NS3 region (b) and 72 sequences comprising 960 nt of the combined NS3/NS5B regions (c).

their sequences in four analyses. Couple 4 had only two analyses performed in the NS5B region, both showing sequences with the same origin (Supplementary Fig. S1, available in JGV Online). Clusters presenting a bootstrap value higher than 70 were considered to have the same origin. Couples 1 and 2 sometimes presented their sequences in a monophyletic branch but without bootstrap support. The sequences of couples 5 and 9 did not share the same origin (Fig. 2).

Phylogenetic studies are used to help in performing detailed molecular epidemiology studies, which in turn help in understanding the transmission chains of infectious agents such as viruses (McCormack & Clewley, 2002).

Parenteral transmission is the most frequent method of HCV transmission, as about 10 % of carriers do not report any risk factors. Although there is some controversy, it is suggested that the presence of HCV in bodily secretions is a source of transmission, whether by close contact in familial relationships or by sexual intercourse.

Several studies have raised the hypothesis of sexual and intrafamilial transmission of HCV, albeit at much lower levels than with hepatitis B virus and HIV (Eyster *et al.*, 1991; Llibre *et al.*, 1992; Mohamed *et al.*, 2005). It is hard to show that HCV has been acquired by sexual intercourse or by contact with a close relative based only on an epidemiological evaluation without using any molecular biology tools, which may account for the previously published controversial data. However, molecular information can provide more precise information about epidemiological linkage.

In previous studies, we observed that, in a number of couples with HCV of the same genotype, only one of the spouses provided information about a known source of contamination. Thus, we performed a detailed investigation using molecular analysis, after retrieving data from their medical charts, according to the criteria of Leitner & Albert (2000) and Ross *et al.* (1999).

Couples with the same genotype were selected for the study, as assessed by sequencing of the 5′UTR region; however, when we analysed the NS5B region, we found that couple 9 had different subtypes (1a and 1b), and information was only available from the husband's medical chart, indicating that he was genotype 1a.

Analysis of the NS5B region increases information compared with previous studies using the 5′UTR. Simmonds (2004) showed that characterization of HCV variants by 5′UTR sequencing can present a subtype divergence of around 10 %, because there are very few informative sites in this region (Simmonds, 2004; Simmonds *et al.*, 1996). Therefore, if the sequences of this region do not discriminate different viral subtypes, the analysis of transmission chains is even more prone to a lack resolution when based only on the 5′UTR. Thus, using a more informative region is needed for a better characterization of virus transmission.

These results confirmed that information regarding the viral genotype alone and the mere evaluation of epidemiological data are insufficient for affirming that sexual and intrafamilial transmission occurs. This is particularly valid for HCV, because there are subtypes that are prevalent in the population of most parts of the world (1a, 1b, 2b and 3a) (Caporaso *et al.*, 1998). In Brazil, genotypes 1 and 3 are the most prevalent, which makes it difficult to perform intrafamilial transmission analyses using this region only (Campiotto *et al.*, 2005).

In an attempt to elucidate whether the origin of the viruses of the infected couples was the same, we evaluated the phylogenetic relationships of different genome regions from both samples.

Reconstruction of a phylogenetic tree using only the sequences of the study samples would not have been informative. To improve the phylogenetic signal, controls were used that were known not to be related, but who circulated in the same background environment.

There is no consensus about which HCV genomic region should be used in such studies. Several regions, such as the core, NS5B, envelope and NS3, have been used (Cavalheiro *et al.*, 2009; Chayama *et al.*, 1995; Honda *et al.*, 1993; Ross *et al.*, 1999). It is important to note that the choice of region is of fundamental importance. The different regions of the HCV genome present different evolutionary dynamics and, consequently, different relationships can be inferred among sequences. There are regions that encode proteins that are highly exposed to the host immune system and undergo constant mutations, whilst others encode functional proteins and are less variable, providing different conclusions. Studying at least two genome regions with different evolutionary dynamics is needed (Leitner & Albert, 2000; McCormack & Clewley, 2002; Ross *et al.*, 1999).

Salemi & Vandamme (2002) analysed the phylogenetic signal of the different HCV genes and found that the region with the best signal was NS5B, followed by the NS3 region. The core and envelope regions, suggested by different authors, did not produce the best signal. The authors suggested that NS5B and NS3 can be safely used when the complete genome sequence is not available.

Based on the information above, we analysed the NS5B and NS3 regions. We used likelihood mapping, performed by the TREE-PUZZLE program, to access the NS5B and NS3 phylogenetic signal (Strimmer & von Haeseler, 1997). The values were similar to those found by Salemi & Vandamme (2002).

To try and obtain a better phylogenetic signal, we also analysed the two regions together. The 960 nt fragment obtained by joining the two regions resulted in 96.1 % fully resolved trees. This result is close to the percentage that has been obtained by analysing complete genomic sequences (96.7 %) (Salemi & Vandamme, 2002).

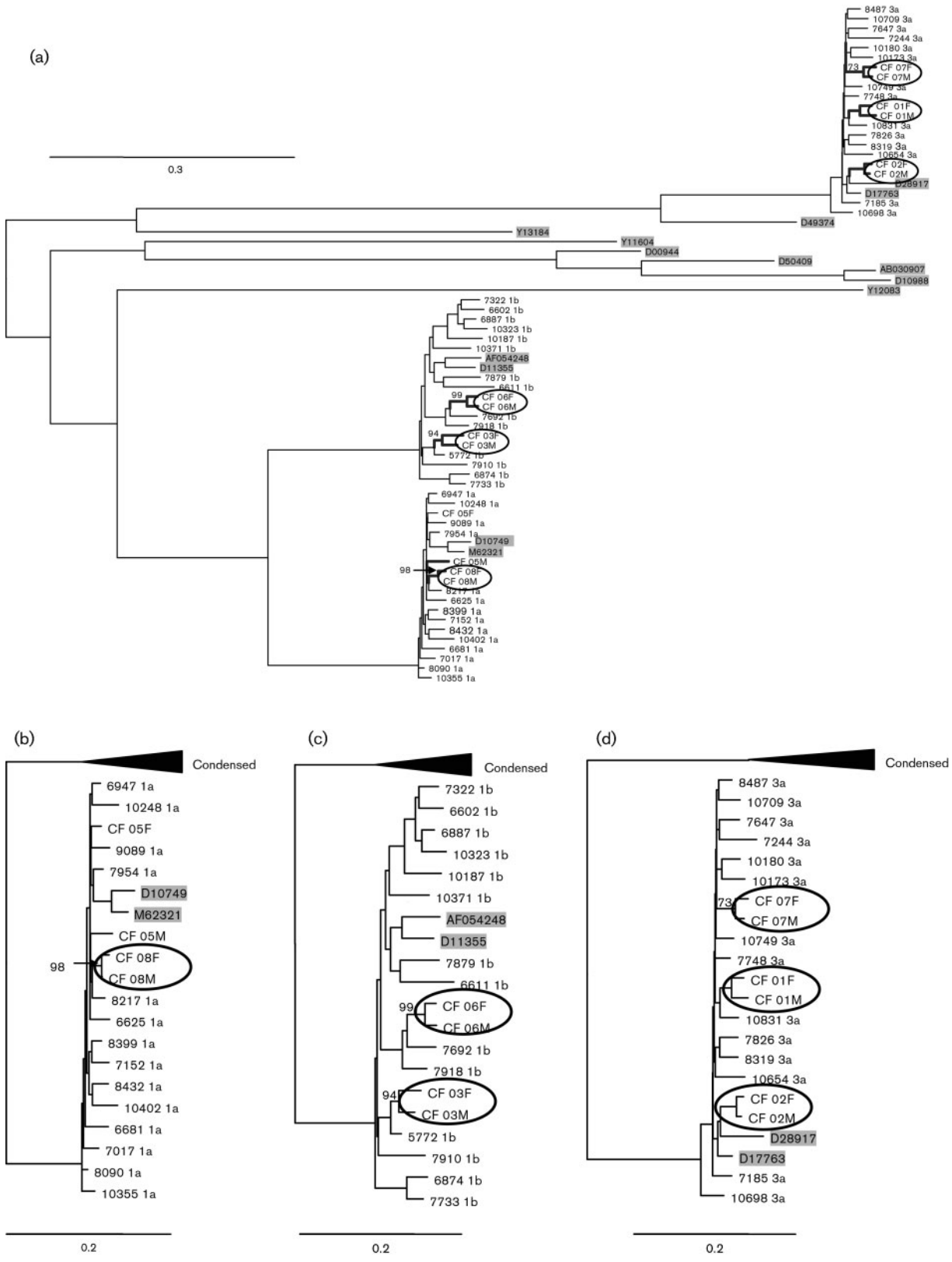

Fig. 2. Unrooted trees constructed using the maximum-likelihood (ML) method, a heuristic search with the tree bisection and reconnection (TBR) branch-swapping algorithm. Trees are based on the TrNef+I+G substitution model, constructed from (a) the NS3/NS5B region (960 nt), (b) the cluster of genotype 1 subtype a, (c) the cluster of genotype 1 subtype b and (d) the cluster of genotype 3 subtype a. The robustness of the phylogenetic groups was evaluated using 1000 bootstrap replicates. Couples sharing the same origin are circled, and sequences obtained from GenBank are shaded. Bars, number of substitutions per site.

Other authors have suggested that similarity can indicate a common origin, indicating intrafamilial or sexual transmission (Chayama *et al.*, 1995; Honda *et al.*, 1993; Kao *et al.*, 1992; Ross *et al.*, 1999). However, the similarity found among sequences varies according to the region and the fragment size. In the present work, the similarity mean found between the sequences of the study couples for the NS5B region was 95.45 %, and the couple with different subtypes presented 78.2 % similarity. Without this couple, the region similarity showed 97.65 % similarity (96.8–99.1 %). For the NS3 region, the similarity mean was 97.24 % (94.7–99.2 %). The similarity between the local controls and references varied from 53.9 to 98 % (mean 72.9 %) for NS5B and from 61 to 98.2 % (mean 74.8 %) for NS3. These results show that close similarity values can be found among sequences of completely different origins, indicating that inferring the same origin for two viral strains based only on similarity analysis does not provide adequate detail.

Chayama *et al.* (1995) found 91.4–99.8 % similarity for a 606 nt fragment of the E1 gene in sequences from eight HCV-infected couples. The five couples who were considered to have virus of the same origin presented similarities between 97.3 and 99.8 %. For the other three couples, the authors were unable to confirm the origin of strains, despite their similarity of 91.4–92.4 %. However, in this work, only three local controls were included, representative of the period in which the couples were selected, and no similarity analyses were carried out for these. Thus, it is not possible to evaluate the similarity found among strains from carriers who were not related to the study cases or who had no contact with them.

The topologies obtained with the different methods resembled each other in the cases in which we found bootstrap values over 70, except for the NS3 region of couple 2. The results of these analyses suggest that the origin of the viruses of couples 3, 4, 6, 7 and 8 was the same. In all cases, the region that presented the best genetic information showed that the two sequences had a common origin, with bootstrap support (McCormack & Clewley, 2002).

Despite the patient charts of couples 1, 2 and 5 indicating the possibility of intrafamilial transmission, our phylogenetic studies did not support this hypothesis.

Couple 9 exhibited a high probability of different sources of infection, as in all analyses the sequence profile indicated two different subtypes.

Interestingly, combining the NS5B and NS3 regions was indicated as the best option for transmission study analyses. In some cases, this analysis was not sufficient to elucidate whether the origin of the different viral strains was the same. However, adding other regions may help to clarify this issue.

This study suggests that couples can be infected with viral strains that present monophyletic relationships, indicating a common source of contamination; however, we had no means of assessing whether this contamination was due to sexual intercourse or to other forms of close contact.

Acknowledgements

This work was supported by the Alves de Queiroz Family Fund for Research, Instituto Adolfo Lutz and Instituto Butantan.

References

Akahane, Y., Kojima, M., Sugai, Y., Sakamoto, M., Miyazaki, Y., Tanaka, T., Tsuda, F., Mishiro, S., Okamoto, H. & other authors (1994). Hepatitis C virus infection in spouses of patients with type C chronic liver disease. *Ann Intern Med* **120**, 748–752.

Alter, M. J. (2002). Prevention of spread of hepatitis C. *Hepatology* **36**, S93–S98.

Alter, M. J., Kruszon-Moran, D., Nainan, O. V., McQuillan, G. M., Gao, F., Moyer, L. A., Kaslow, R. A. & Margolis, H. S. (1999). The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med* **341**, 556–562.

Campiotto, S., Pinho, J. R., Carrilho, F. J., Da Silva, L. C., Souto, F. J., Spinelli, V., Pereira, L. M., Coelho, H. S., Silva, A. O. & other authors (2005). Geographic distribution of hepatitis C virus genotypes in Brazil. *Braz J Med Biol Res* **38**, 41–49.

Caporaso, N., Ascione, A. & Stroffolini, T. (1998). Spread of hepatitis C virus infection within families. Investigators of an Italian multicenter group. *J Viral Hepat* **5**, 67–72.

Cavalheiro, N. P. (2004). Hepatitis C: transmission between couples. *Rev Inst Med Trop Sao Paulo* **46**, 86.

Cavalheiro, N. P., De La Rosa, A., Elagin, S., Tengan, F. M., Araujo, E. S. & Barone, A. A. (2009). Hepatitis C: sexual or intrafamilial transmission? Epidemiological and phylogenetic analysis of hepatitis C virus in 24 infected couples. *Rev Soc Bras Med Trop* **42**, 239–244.

Chayama, K., Kobayashi, M., Tsubota, A., Koida, I., Arase, Y., Saitoh, S., Ikeda, K. & Kumada, H. (1995). Molecular analysis of intraspousal transmission of hepatitis C virus. *J Hepatol* **22**, 431–439.

Chevaliez, S. & Pawlotsky, J. M. (2007). Hepatitis C virus: virology, diagnosis and management of antiviral therapy. *World J Gastroenterol* **13**, 2461–2466.

Chiaramonte, M., Stroffolini, T., Lorenzoni, U., Minniti, F., Conti, S., Floreani, A., Ntakirutimana, E., Vian, A., Ngatchu, T. & Naccarato, R. (1996). Risk factors in community-acquired chronic hepatitis C virus infection: a case–control study in Italy. *J Hepatol* **24**, 129–134.

Ewing, B. & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186–194.

Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175–185.

Eyster, M. E., Alter, H. J., Aledort, L. M., Quan, S., Hatzakis, A. & Goedert, J. J. (1991). Heterosexual co-transmission of hepatitis C virus (HCV) and human immunodeficiency virus (HIV). *Ann Intern Med* **115**, 764–768.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Filippini, P., Coppola, N., Scolastico, C., Rossi, G., Onofrio, M., Sagnelli, E. & Piccinino, F. (2001). Does HIV infection favor the sexual transmission of hepatitis C? *Sex Transm Dis* **28**, 725–729.

Gordon, D., Abajian, C. & Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res* **8**, 195–202.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.

Honda, M., Kaneko, S., Unoura, M., Kobayashi, K. & Murakami, S. (1993). Risk of hepatitis C virus infections through household contact with chronic carriers: analysis of nucleotide sequences. *Hepatology* **17**, 971–976.

Kao, J. H., Chen, P. J., Yang, P. M., Lai, M. Y., Sheu, J. C., Wang, T. H. & Chen, D. S. (1992). Intrafamilial transmission of hepatitis C virus: the important role of infections between spouses. *J Infect Dis* **166**, 900–903.

Keiserman, D. R., Both, C. T., Mattos, A. A., Remiao, J., Alexandre, C. O. & Sherman, K. E. (2003). Intrafamilial transmission of hepatitis C virus in patients with hepatitis C and human immunodeficiency virus coinfection. *Am J Gastroenterol* **98**, 878–883.

Leitner, T. & Albert, J. (2000). Reconstruction of HIV-1 transmission chains for forensic purposes. *AIDS Rev* **2**, 241–251.

Llibre, J. M., Tor, J., Manterola, J. M., Carbonell, C. & Roset, J. (1992). Risk stratification for dissemination of tuberculosis in HIV-infected patients. *Q J Med* **82**, 149–157.

Mastromatteo, A. M., Rapaccini, G. L., Pompili, M., Ursino, S., Romano-Spica, V., Gasbarrini, G. & Vanini, G. (2001). Hepatitis C virus infection: other biological fluids than blood may be responsible for intrafamilial spread. *Hepatogastroenterology* **48**, 193–196.

McCormack, G. P. & Clewley, J. P. (2002). The application of molecular phylogenetics to the analysis of viral genome diversity and evolution. *Rev Med Virol* **12**, 221–238.

Mohamed, M. K., Abdel-Hamid, M., Mikhail, N. N., Abdel-Aziz, F., Medhat, A., Magder, L. S., Fix, A. D. & Strickland, G. T. (2005). Intrafamilial transmission of hepatitis C in Egypt. *Hepatology* **42**, 683–687.

Nakashima, K., Kashiwagi, S., Hayashi, J., Noguchi, A., Hirata, M., Kajiyama, W., Urabe, K., Minami, K. & Maeda, Y. (1992). Sexual transmission of hepatitis C virus among female prostitutes and patients with sexually transmitted diseases in Fukuoka, Kyushu, Japan. *Am J Epidemiol* **136**, 1132–1137.

Ndimbie, O. K., Kingsley, L. A., Nedjar, S. & Rinaldo, C. R. (1996). Hepatitis C virus infection in a male homosexual cohort: risk factor analysis. *Genitourin Med* **72**, 213–216.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Ross, R. S., Viazov, S., Varenholz, C. & Roggendorf, M. (1999). Inquiries on intraspousal transmission of hepatitis C virus: benefits and limitations of genome sequencing and phylogenetic analysis. *Forensic Sci Int* **100**, 69–76.

Salemi, M. & Vandamme, A. M. (2002). Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. *J Mol Evol* **54**, 62–70.

Simmonds, P. (2004). Genetic diversity and evolution of hepatitis C virus – 15 years on. *J Gen Virol* **85**, 3173–3188.

Simmonds, P., Mellor, J., Sakuldamrongpanich, T., Nuchaprayoon, C., Tanprasert, S., Holmes, E. C. & Smith, D. B. (1996). Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: comparison with classifications based upon sequence similarity. *J Gen Virol* **77**, 3013–3024.

Strimmer, K. & von Haeseler, A. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci U S A* **94**, 6815–6819.

Swofford, D. L. (2002). PAUP*: Phylogenetic Analysis Using Parsimony (and other methods), version 4. Sunderland, MA: Sinauer Associates.

Terrault, N. A. (2002). Sexual activity as a risk factor for hepatitis C. *Hepatology* **36**, S99–S105.

Thomas, D. L., Zenilman, J. M., Alter, H. J., Shih, J. W., Galai, N., Carella, A. V. & Quinn, T. C. (1995). Sexual transmission of hepatitis C virus among patients attending sexually transmitted diseases clinics in Baltimore – an analysis of 309 sex partnerships. *J Infect Dis* **171**, 768–775.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Zharkikh, A. & Li, W. H. (1995). Estimation of confidence in phylogeny: the complete-and-partial bootstrap technique. *Mol Phylogenet Evol* **4**, 44–63.