Genetic diversity of NS3 protease from Brazilian HCV isolates and possible implications for therapy with direct-acting antiviral drugs

Allan Peres-da-Silva[1], Adilson José de Almeida[1,2], Elisabeth Lampe[1]/+

[1]Laboratório de Hepatites Virais, Instituto Oswaldo Cruz-Fiocruz, Av. Brasil 4365, 21040-360 Rio de Janeiro, RJ, Brasil
[2]Hospital Universitário Gaffrée e Guinle, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, RJ, Brasil

The hepatitis C virus (HCV) NS3 protease has been one of the molecular targets of new therapeutic approaches. Its genomic sequence variability in Brazilian HCV isolates is poorly documented. To obtain more information on the magnitude of its genetic diversity, 114 Brazilian HCV samples were sequenced and analysed together with global reference sequences. Genetic distance (d) analyses revealed that subtype 1b had a higher degree of heterogeneity (d = 0.098) than subtypes 1a (d = 0.060) and 3a (d = 0.062). Brazilian isolates of subtype 1b were distributed in the phylogenetic tree among sequences from other countries, whereas most subtype 1a and 3a sequences clustered into a single branch. Additional characterisation of subtype 1a in clades 1 and 2 revealed that all but two Brazilian subtype 1a sequences formed a distinct and strongly supported (approximate likelihood-ratio test = 93) group of sequences inside clade 1. Moreover, this subcluster inside clade 1 presented an unusual phenotypic characteristic in relation to the presence of resistance mutations for macrocyclic inhibitors. In particular, the mutation Q80K was found in the majority of clade 1 sequences, but not in the Brazilian isolates. These data demonstrate that Brazilian HCV subtypes display a distinct pattern of genetic diversity and reinforce the importance of sequence information in future therapeutic approaches.

Key words: genetic diversity - NS3 protease - hepatitis C virus - direct-acting antiviral agents

The hepatitis C virus (HCV) has a worldwide distribution, with an estimated 130 million people being infected (Alter 2007). Approximately 85% of affected individuals fail to achieve spontaneous clearance of the virus and become chronically infected (Lavanchy 2009). As a member of the Flaviviridae family (genus Hepacivirus), HCV is an enveloped virus with a positive, single-stranded RNA genome of approximately 9,400 nucleotides, which is translated into a single polyprotein of approximately 3,000 amino acids. This polyprotein is subsequently processed by viral and host proteases into structural and non-structural proteins. Based on a comparative analysis of the nucleotide sequences of the core, envelope and NS5B regions, HCV was classified into six major genotypes (1-6) and more than 100 subtypes (1a, 1b, 1c and so one) (Simmonds et al. 1994) and, recently, isolates of a new genotype were proposed to be classified as a seventh major genotype (Murphy et al. 2007).

To date, there is no vaccine against HCV and the only therapy available is based on the administration of interferon (IFN) in combination with ribavirin, which may induce a sustained virological response in approximately 50% of patients chronically infected with HCV genotype 1. This therapeutic approach has limited effectiveness and a number of side effects, including depression, fatigue, alopecia, anorexia, haemolytic anaemia, leukopaenia, thrombocytopaenia and "flu-like" symptoms (Russo & Fried 2003). These drugs are indirect antiviral agents as they do not target a specific viral protein or the HCV nucleic acid. Considering the low rate of effectiveness of the current treatment and that most individuals infected with HCV will evolve towards chronic infection with the potential to progress to more serious complications, such as cirrhosis and hepatocellular carcinoma, new therapeutic options that can be used in usual clinical practice are urgently needed (Schinazi et al. 2010).

The increasing progress in understanding protein structure and developing HCV subgenomic systems and, more recently, cell culture systems (Nakamura et al. 2008, Triyatni et al. 2011) allowed for the characterization of the HCV life cycle and enabled the design of several specific inhibitors for use in anti-HCV therapy, known under the generic name of direct-acting antiviral agents (DAAs) (Asselah & Marcellin 2011). Instead of stimulating the host's natural immune response to viral infection, the DAAs are directed against HCV molecular targets. Among the HCV molecular inhibitors, enzymatic inhibitors of NS3 serine protease and NS5B polymerase are already being tested in the advanced phases of clinical trials (Vermehren & Sarrazin 2011). The HCV NS3 gene encodes for a bifunctional protein in which the amino-terminal domain forms, along with the NS4A protein, a heterodimeric S protease responsible for cleavage of the nonstructural portion of the polyprotein at the NS3-NS4A, NS4A-NS4B, NS4B-NS5A and NS5A-NS5B junctions, whereas its NS3 carboxy-terminal domain displays NTPase/helicase activity, which is essential for HCV RNA translation and genome replication. In addition to its importance in the viral life cycle,

the HCV NS3 protein was also shown to be involved in viral oncogenesis, which affects the normal cellular functions (Lai et al. 2008), efficiently blocking the retinoic-acid-inducible gen I and Toll-like receptor 3 signalling pathways and enabling the virus to escape from the IFN antiviral response, thus contributing to virus persistence (Foy et al. 2003, Meylan et al. 2005). Therefore, the HCV NS3 protein is at present considered to not only be an attractive constituent for the production of an anti-HCV vaccine (Zeng et al. 2009), but also an important target for the development of new therapeutic protease inhibitors (Enomoto et al. 2009). To date, several specific NS3-4A S protease inhibitors, such as VX-950, SCH6, SCH503034, ITMN-191 and TMC435350, have been designed and are currently being evaluated in clinical trials (López-Labrador 2008). However, given the high level of variability of HCV, which is a result of the error-prone nature of RNA-dependent RNA polymerase, variants resistant to a number of protease inhibitors have been identified (Peres-da-Silva et al. 2010, Vermehren & Sarrazin 2011) and represent a current challenge for therapy with DAAs.

Recently, reports based on the phylogenetic analysis of the complete genome and partial NS5B sequences have shown that HCV subtype 1a isolates can be separated into at least two distinct clades, designated 1 and 2 (Chevaliez et al. 2009, Pickett et al. 2011). Several informative sites for this distinction are located near or within codons associated with resistance to NS3-4A S protease inhibitors, such as the hydrophobic protease cavity-forming amino acid Q41, which confers higher resistance to the inhibitor SCH 503034 (Tong et al. 2008). Therefore, the detailed knowledge of the nucleotide sequence variability of the NS3 protease gene is necessary for the development of these antiviral compounds, which will help to achieve a greater therapeutic effectiveness against different HCV genotypes.

To date, there are few studies on the genomic variability of the HCV NS3 serine protease domain in Brazil and in Latin America (Mello et al. 2009), increasing the need for obtaining more detailed data on the magnitude of its genetic diversity. In this study, we analyse the variability of the HCV NS3 protease gene in Brazilian isolates and its possible implications for therapy with DAAs.

SUBJECTS, MATERIALS AND METHODS

Patients - Between March 2007-August 2008, a total of 114 serum samples were collected from patients [46 males and 68 females, aged 53.1 ± 9.8 years (mean ± standard deviation)] diagnosed as having a chronic HCV infection (serum anti-HCV and HCV-RNA-positive for more than 6 months). The diagnosis of an HCV infection was made on the basis of the presence of anti-HCV antibodies (Ortho HCV 3.0 ELISA test system with Enhanced SAVe, Ortho Clinical Diagnostics, Raritan, NJ, USA) and HCV-RNA detectable by the Cobas Amplicor HCV Monitor test (Roche Molecular Systems, Branchburg, USA) in serum samples. The HCV genotypes were determined using the Versant HCV Genotype 2.0 Assay (LiPA 2.0, Innogenetics, Ghent, Belgium) (Verbeeck et al. 2008). The samples were selected from patients infected with genotype 1a, 1b or 3a who visited the Laboratory for Viral Hepatitis, Oswaldo Cruz Institute/Fiocruz, Rio de Janeiro, Brazil, for diagnostic purposes. The study was approved by the local Ethical Committee (CEP 32/2004) and conforms to the ethical guidelines of the 1975 Declaration of Helsinki. Written informed consent was obtained from each patient before entering into the study. Patients eligible for the study were between 18-65 years of age, both male and female and of any race. Patients who had received or were currently undergoing antiviral therapy or who tested positive for hepatitis B surface antigen and/or antibodies to human immunodeficiency virus-1/2 were excluded from the study.

Extraction of viral RNA - HCV RNA was extracted from serum samples using a QIAamp Viral RNA kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

Reverse-transcription (RT)-nested polymerase chain reaction (PCR) and amplification of viral RNA - One-step RT-PCR followed by a second round of PCR (RT-nested PCR) was used to amplify the NS3 protease region of HCV using specific oligonucleotide primers for each subtype, as previously described (Peres-da-Silva et al. 2010). For the RT-PCR reactions, the reagents from the Superscript™ III One-Step RT-PCR system (Life Technologies, Invitrogen, Carlsbad, CA, USA) were used. The RT-PCR mixture contained 0.5 μL of the corresponding type of specific sense and antisense primers (10 μM), 12.5 μL of 2 x reaction mix (a buffer containing 0.4 μM of each dNTP and 3.2 μM of $MgSO_4$), 1 μL of Super-Script™ III RT/Platinum® Taq DNA Polymerase (4 U/μL), 5.5 μL of DNase/RNase free distilled water and 5 μL of viral RNA. The conditions for the RT-PCR step were as follows: 42ºC for 45 min followed by inactivation of the reverse transcriptase and activation of the "hot start" DNA polymerase at 94ºC for 4 min; then 35 cycles at 94ºC for 15 sec, 56ºC for 30 sec and 68ºC for 90 sec and a final elongation at 68ºC for 5 min. The RT-PCR conditions were the same for all subtypes, except for an annealing temperature of 45ºC for subtype 3a. One microliter of the product was subjected to a second round of PCR in the presence of 2.5 μL of each internal primer (10 μM), 5 μL of 10 x PCR buffer, 8 μL of dNTPs (1.25 mM), 4 μL of $MgSO_4$ (50 mM), 0.4 μL of Platinum® Taq High Fidelity (5 U/μL) and 29.1 μL of DNase/RNase-free distilled water. After an initial denaturation at 94ºC for 3 min, the DNA was amplified for 30 cycles at 94ºC for 30 sec, 60ºC for 30 sec and 72ºC for 90 sec and subjected to a final extension at 72ºC for 7 min. PCR products of the expected length of 765 base pairs were fractionated by 1.5% agarose gel electrophoresis and stained with ethidium bromide.

Nucleotide sequencing - The RT-nested PCR products were purified using the QIAquick gel extraction kit (Qiagen, Hilden, Germany) and subjected to nucleotide sequencing reactions in both directions using the Big Dye Terminator kit (version 3.1, Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions and analysed on an ABI 3730

DNA automated sequencer (Applied Biosystems). The nucleotide sequence data presented in this paper were submitted to the GenBank database under the accessions GU126553-GU126666.

Phylogenetic analysis - The sequences obtained were aligned using the CLUSTALX program, version 1.83 (Thompson et al. 1997). To investigate any possible epidemiological linkage of the Brazilian HCV 1a, 1b and 3a strains with other previously characterised isolates, we included in the phylogenetic analysis North American and European sequences sharing the same fragment that were available in a public HCV database [Los Alamos (hcv.lanl.gov/content/index)] (Kuiken et al. 2005). Maximum-likelihood phylogenetic trees were inferred using the PhyML program (Guindon & Gascuel 2003) with the approximate likelihood-ratio test (aLRT) (Anisimova & Gascuel 2006) based on a Shimodaira-Hasegawa-like procedure, using general time reversible as substitution model and an subtree pruning and regrafting-based tree search algorithm to estimate the tree topologies. The further characterisation of the HCV subtype 1a in clades 1 and 2 was also performed with the same 120 representative sequences of each clade used by Pickett et al. (2011).

Analysis of synonymous substitutions per synonymous site (dS), nonsynonymous substitutions per nonsynonymous site (dN), dN/dS ratio and genetic distance (d) - To calculate the frequencies of dS, dN, dN/dS - an analysis indicative of selective pressure - we used the synonymous/nonsynonymous analysis program (hiv.lanl.gov) (Korber 2000) according to the Nei and Gojobori method (1986).

The analysis of the nucleotide genetic distances of Brazilian, North American and European HCV NS3 serine protease nucleotide sequences was performed using the Maximum Composite Likelihood model in the MEGA 4.0 program (Tamura et al. 2007). This analysis was conducted within and between each set of HCV subtypes.

Detection of signature patterns in Brazilian amino acid sequences - The entire Brazilian HCV NS3 protease domain (181 amino acids) was screened in comparison with NS3 protease alignment sequences from other geographical regions to find subtype-specific amino acid signature patterns at specific protease sites using the Vespa program (hcv.lanl.gov/content/sequence/VESPA/vespa.html) (Korber & Myers 1992). In the aligned sequences, the Vespa program calculates the frequencies of each amino acid at each position (query) and compares them with a set of reference sequences (background). The threshold of 1.0 was chosen to define a single standard for each residue site.

RESULTS

The entire NS3 protease domain of the HCV genome (nucleotides 3420-3963) was analysed among 114 patients infected with HCV subtypes 1a (n = 48; 42%), 1b (n = 53; 46%) or 3a (n = 13; 11%). The prevalence of HCV subtypes in patients selected for our study was similar to that identified in our geographical area (de Almeida et al. 2004).

Phylogenetic analysis - The entire HCV serine protease extension (nt 1-543) was used to reconstruct phylogenetic trees from 114 Brazilian HCV NS3 S protease sequences of HCV subtypes 1a, 1b and 3a, along with the corresponding subtype nucleotide sequences from different geographic regions, totalling 138 HCV subtype 1a sequences, 152 HCV subtype 1b sequences and 102 HCV subtype 3a sequences. The resulting phylogenetic trees representing the genetic relationships of HCV subtypes 1a, 1b and 3a are presented in Figs 1-3. As shown in Fig. 1, all but three Brazilian HCV subtype 1a sequences clustered together in a single distinct cluster (aLRT = 0.89) consisting of 45 isolates. The Brazilian 1b sequences revealed 10 independent lineages randomly distributed among sequences from other countries (Fig. 2). We also observed that, even though most of these lineages were composed of one-three sequences, four major Brazilian clusters were composed of five-22 sequences, with aLRT supports > 0.80. In contrast, all of the Brazilian HCV subtype 3a sequences segregated in one cluster, along with one sequence from Europe (Fig. 3).
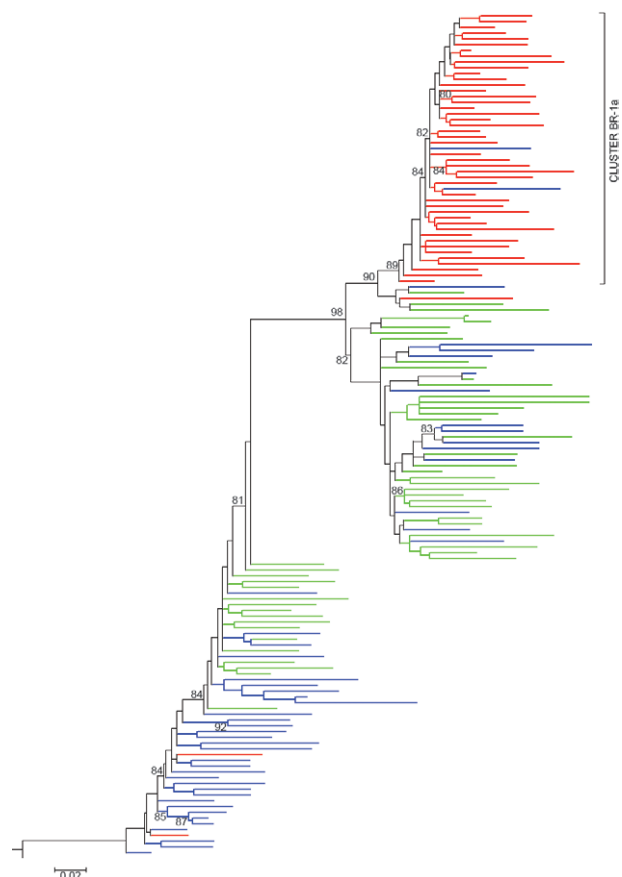

Fig. 1: maximum-likelihood phylogeny of hepatitis C virus subtype 1a from NS3 protease sequences showing evolutionary relationships between 48 Brazilian (red branches) and representative sequences from Europe (blue branches) (n = 48) and North America (green branches) (n = 42). The tree was rooted with subtype 1b strains (M58335). Approximate likelihood-ratio test supports values > 0.80 are shown above the nodes. Sequence names were removed for clarity.
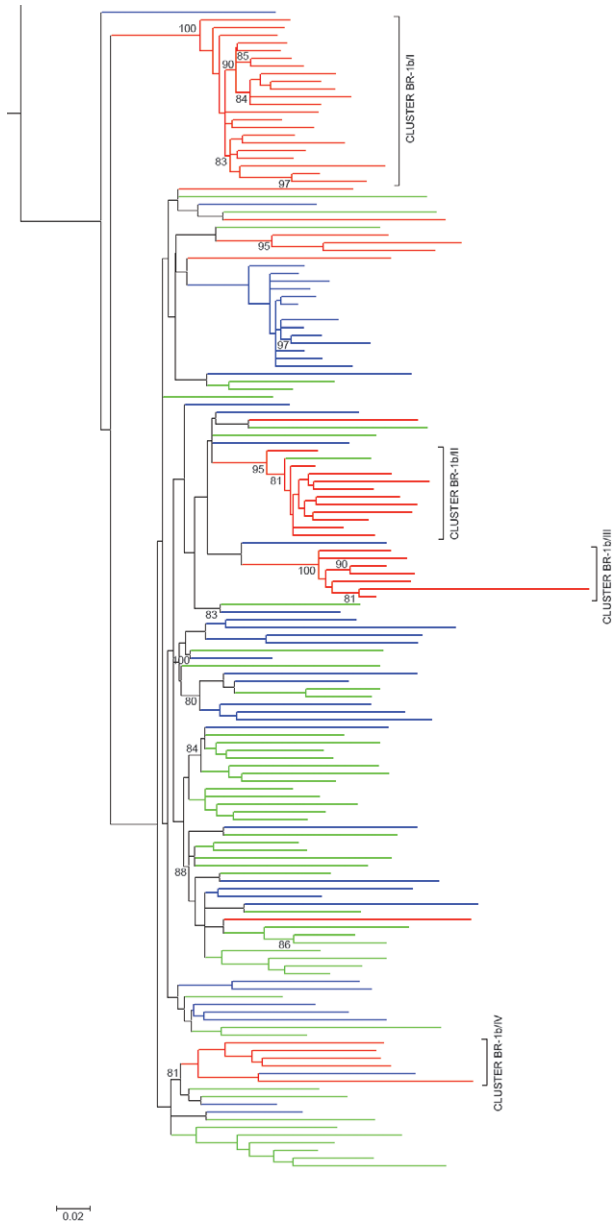
Fig. 2: maximum-likelihood phylogenetic tree of hepatitis C virus subtype 1b from NS3 protease sequences showing evolutionary relationships between 53 Brazilian (red branches) and representative sequences from Europe (blue branches) (n = 49) and North America (green branches) (n = 50).). Approximate likelihood-ratio test supports values > 0.80 are shown above the nodes. The tree was rooted with subtype 1a strains (NC004102). Sequence names were removed for clarity.

Recently, Pickett et al. (2011) demonstrated that HCV subtype 1a could be separated into two distinct clades. Because almost all of the Brazilian HCV-1a sequences formed a large monophyletic group, we constructed a separate tree with the same sequences used by Pickett et al. (2011) to determine to which clade our sequences belonged. The phylogenetic tree reconstruction revealed that all but two Brazilian HCV subtype 1a sequences formed a distinct and strongly supported (aLRT = 0.93) group of sequences inside clade 1 (Fig. 4).

Analysis of dS, dN, dN/dS and d - The values of dS and dN and the dN/dS ratio are displayed in Table I. The substitution mutational analysis revealed that most of the nucleotide mutations were synonymous. Therefore, the dN/dS ratio was low, indicating the presence of purifying selection pressure, which is associated with conservation of regions, presumably to maintain protein functions.

Regarding the genetic distance analysis performed within each HCV subtype, we observed that Brazilian HCV subtype 1a NS3 protease sequences presented a lower value (d = 0.060) in comparison with that observed for European (d = 0.092) and North American (d = 0.084) sequences. The d of Brazilian HCV-1b sequences showed almost the same value (d = 0.098) as that observed for the European (d = 0.092) and North American sequences (d = 0.090). The Brazilian HCV subtype 3a sequences were also less variable (d = 0.062) than those from Europe (d = 0.076) or North America (d = 0.074). When comparing sequences from those geographical areas, the genetic distances of the Brazilian HCV subtype 1b NS3 protease sequences were very similar to those from both the North American (d = 0.102) and European (d = 0.105) sequences. The Brazilian HCV subtype 3a sequences also presented similar values when compared with North American (d = 0.078) and European (d = 0.076) sequences. In contrast, the Brazilian HCV subtype 1a NS3 protease sequences exhibited more similarity to isolates from North America (d = 0.091) than to European (d = 0.101) sequences.

Signatures in Brazilian amino acid sequences - The frequencies of the major amino acids in subtype-specific Brazilian HCV sequences of the NS3 protease domain that differ from the frequencies found in European and North American isolates are shown in Table II. In Brazilian HCV subtype 1a NS3 protease sequences, 50% of the amino acids located at position 40 were represented by the amino acid threonine (T), differing substantially from those found in North America and Europe, which possess alanine (A) as the predominant amino acid in 84% and 81.6% of their sequences, respectively. A similar amino acid profile was observed for Brazilian HCV subtype 3a NS3 protease sequences at sites 7 and 98. In addition, this subtype also had 69.2% of its sequences occupied by T instead of valine (V) at site 67, as was observed for the majority of North American sequences.

Table III presents the amino acid signature patterns observed in clades 1 and 2 of HCV subtype 1a. Brazilian sequences presented the amino acid T as the predominant residue at site 40, differing substantially from those found in clade 1 and 2, which have the amino acid A as the predominant amino acid. The subtype 1a clade 1 exhibited the amino acid lysine (K) at site 80, which is associated with resistance to macrocyclic inhibitors in 60% of its sequences, while the amino acid glutamine (Q) was found in 97.5% of clade 2 sequences and in 93.8% of subtype 1a Brazilian sequences. In contrast, clade 1 and Brazilian sequences presented the amino acid asparagine (N) in approximately 80% of their sequences at site 174, while 71.7% of clade 2 sequences presented the amino acid S.
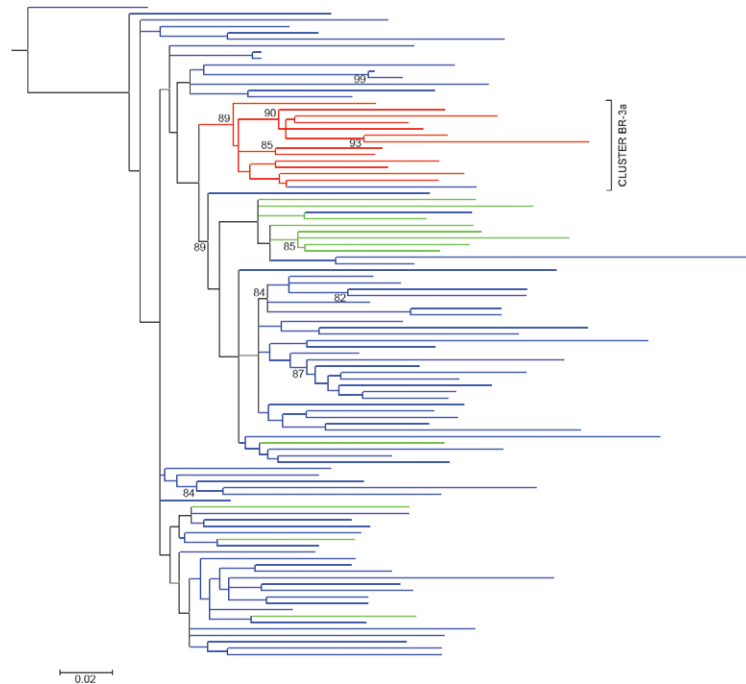
Fig. 3: maximum-likelihood phylogenetic tree of hepatitis C virus subtype 3a from NS3 protease sequences showing evolutionary relationships between 13 Brazilian (red branches) and representative sequences from Europe (blue branches) (n = 77) and North America (green branches) (n = 12). Approximate likelihood-ratio test supports values > 0.80 are shown above the nodes. The tree was rooted with subtype 3b strains (D49374). Sequence names were removed for clarity.
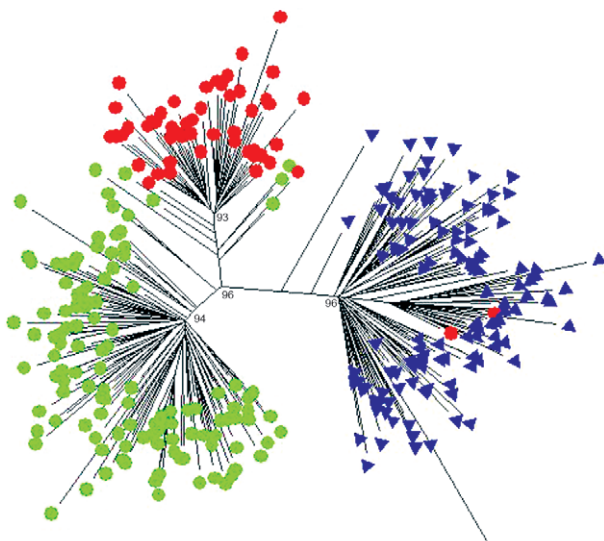

Fig. 4: maximum-likelihood phylogenetic tree of hepatitis C virus (HCV) subtype 1a sequences for clade characterization estimated from NS3 protease sequences showing evolutionary relationships between 48 Brazilian (red circles) HCV subtype 1a strains and 120 representative subtype 1a sequences of clade 1 (green circles) and clade 2 (blue triangles). Both clades are formed by isolates from Europe and North America. The star-like topology tree shows all, except two, Brazilian sequences forming a distinct group within clade 1. The tree was rooted with subtype 1b strains (M58335). Approximate likelihood-ratio test supports values > 0.90 are shown above the nodes. Sequence names were removed for clarity.

DISCUSSION

There is little information on the extent of the natural genetic variability of the HCV NS3 protease region from different viral subtypes isolated in Latin America. The NS3-4A S protease of HCV is essential for viral replication and, therefore, has been one of the most attractive targets for developing specific antiviral agents. However, because the HCV genotypes diverge in their nucleotide composition by more than 30%, the successful development of these inhibitors depends on a wide knowledge of NS3 protease variability to be effective against different HCV genotypes. Recently, in an analysis of 380 HCV NS3 serine protease sequences from public HCV databases, López-Labrador et al. (2008) mapped natural polymorphisms among proteases from worldwide HCV isolates. However, sequences from Latin America were not included in that analysis due to the paucity of NS3 sequences available in that database. In Brazil, there is only one study on the evolution of the NS3 amino acid sequences, which used viral isolates from six patients following liver transplantation: three infected with subtype 1b and three with subtype 3a. The authors observed that several differences in the N-terminal NS3 amino acid sequences were observed in genotype 1b isolates, compared with the NS3 GenBank reference sequences (Mello et al. 2009). Because the number of sequences analysed in that study was very small, the genetic variability of HCV proteases among different viral HCV genotypes circulating in Brazil is practically unknown. The data presented

in our study will contribute to a better knowledge of the magnitude of the natural variation in the HCV NS3 serine protease by characterising its heterogeneity in a diverse population of individuals infected with the three most prevalent HCV subtypes in Brazil.

The analysis of subtype-specific amino acid signature patterns at specific sites in the HCV NS3 protease of Brazilian samples demonstrated heterogeneity among the three subtypes studied. In our analysis, we reported that the three Brazilian HCV subtypes studied had amino acid variations at specific locations in the protease domain. Possible changes in the structure of the NS3 protease could be caused by different amino acid frequencies. In Brazilian HCV subtype 1a and 3a sequences, we observed

the amino acid T as a signature at several protease *loci*. Variations from A or V to T or S do not affect the charge because these amino acids possess a neutral charge, but T and S show a lateral chain with a hydroxyl group that may be a putative phosphorylation site or form distinct hydrogen bonds. In addition, the majority of Brazilian HCV subtype 1a sequences at site 67 and more than half of HCV subtype 1b sequences at site 86 presented the amino acid proline (P) as the amino acid signature. The amino acid P may cause changes in the enzyme's backbone structure, due to structural rearrangements.

The overall values for comparisons between dN and dS using the dN/dS ratio were less than 1.0 for all analyses, suggesting the presence of purifying selection across the NS3 serine protease region. Purifying selection is associated with the conservation of the sequence, which maintains protein function. An overall dN/dS ratio < 1 indicates that most positions in the NS3 region were subject to purifying selection, but the presence of a few positions that are subject to positive selection cannot be ignored. Even synonymous codons in the NS3 protease region may influence the rates at which a particular phe-

TABLE I

Values of dN, dS and dN/dS ratio for the NS3 serine protease domain of hepatitis C virus (HCV) subtypes 1a, 1b and 3a in relation to geographical regions

Regions and HCV subtypes	dN	dS	Ratio dN/dS
Brazil			
1a	0.0109	0.2166	0.0394
1b	0.0217	0.3556	0.0527
3a	0.0124	0.2183	0.0426
Europe			
1a	0.0147	0.3633	0.0326
1b	0.0174	0.3451	0.0406
3a	0.0122	0.2868	0.0352
North America			
1a	0.0120	0.3287	0.0250
1b	0.0169	0.3347	0.0425
3a	0.0122	0.2747	0.0396

dn: synonymous substitutions per substitutions site; dS: non-synonymous substitutions per non-substitutions site.

TABLE III

Amino acids signature patterns in Brazilian hepatitis C virus (HCV) sequences and subtype 1a clades

		HCV subtype 1a	
Amino acid position	Brazil	Clade 1	Clade 2
40	0.5 T	0.825 A	0.875 A
80	0.938 Q	0.60 K	0.975 Q
174	0.854 N	0.808 N	0.717 S

A: alanine; K: lysine; N: asparagine; Q: glutamine; S: serine; T: threonine.

TABLE II

Amino acid signatures of the different hepatitis C virus (HCV) subtypes in the NS3/4A protease domain in relation to geographical regions

	HCV subtype								
	1a			1b			3a		
Amino acid position	Brazil	North America	Europe	Brazil	North America	Europe	Brazil	North America	Europe
7	-	-	-	-	-	-	0.692T	1.0A	0.803A
40	0.5 T	0.840A	0.816A	-	-	-	-	-	-
67	0.792P	-	0.551S	-	-	-	0.692T	0.6 V	-
86	-	-	-	0.585P	0.900Q	0.878Q	-	-	-
98	-	-	-	-	-	-	0.615T	-	0.545A
174	0.854N	0.5S	-	-	-	-	-	-	-

A: alanine; N: asparagine; P: proline; Q: glutamine; S: serine; T: threonine; V: valine.

notype can emerge. Indeed, in codon 155 of the protease, a single point mutation is required to convert the arginine codon to a protease-resistant K (AAR) in HCV subtype 1a isolates from both clades, including the Brazilian sub-clade 1 isolates, whereas two nucleotide mutations are needed to change the NS3 155 codon from CGN to K (AAR) in HCV subtype 1b (Chevaliez et al. 2009). Thus, it can be suggested that a particular HCV subtype or even a clade of HCV subtype 1a may be more prone to developing resistant variants (Pickett et al. 2011). In the amino acid signature analyses, we observed that most sequences in subtype 1a clade 1 contained the mutation Q80K, which had been reported to exhibit reduced susceptibility to macrocyclic inhibitors (MK-7009, ITMN-191 and TMC-435350) (Bae et al. 2010). Nevertheless, the majority of Brazilian HCV subtype 1a sequences, as well as the sequences of clade 2, did not contain this mutation (93.8% of Brazilian sequences had the amino acid Q at position 80). These results reinforce the importance of considering the genomic heterogeneity of HCV subtype 1a in the context of antiviral agents directed to the NS3 serine protease domain.

In accordance with the higher genetic variability observed for the HCV subtype 1b NS3 protease sequences from North America and Europe, the Brazilian sequences also revealed greater nucleotide variability, while our HCV subtype 1a protease sequences presented less genetic variation in comparison with that observed for European and North American sequences. The comparison of ds among different geographical regions showed that Brazilian HCV-1a sequences are related more closely to those from North America, while HCV-1b and 3a sequences presented almost the same d to those from North America or Europe.

The phylogenetic tree analysis revealed different relationships patterns for the three Brazilian HCV subtypes analysed. Although Brazilian HCV subtype 1a and 3a sequences cluster into a single branch in the phylogenetic tree, the subtype 1b sequences form different groups of related sequences that are distributed randomly in the phylogenetic tree among sequences from other geographical regions. These results are very similar to those observed in our previous study with HCV NS5B sequences (Lampe et al. 2010), in which we observed that the majority of HCV subtype 1a sequences were clustered into a single clade, whereas Brazilian subtype 1b segregated into various independent clusters. In accordance with previous observations from an NS5B analysis, the epidemic of subtype 1a is mostly a consequence of a dissemination of a single HCV strain, while the HCV-1b epidemic is characterised more by the concurrent dissemination of multiple HCV strains.

Furthermore, we were able to demonstrate that Brazilian HCV subtype 1a sequences clustered within clade 1 as a separated group of related sequences with a high support value. This finding is interesting because, in the phylogenetic tree reconstruction of Pickett et al. (2011), no correlation was found between either location or time of isolation and the clade in which the sequences were placed: both clades contained isolates from different localities in the United States and Europe taken at different times. Moreover, we also observed that the Brazilian subcluster inside clade 1 had different phenotypic characteristics in relation to the presence of resistance mutations for macrocyclic inhibitors (MK-7009, ITMN-191 and TMC-435350). In our previous study, we had demonstrated a V36L mutation in 5.6% of patients infected with subtype 1b and in all isolates of the 3a subtype and a T54S mutation in 4.1% of isolates of subtype 1a. The presence of variants carrying mutations associated with resistance to protease inhibitors in therapy-naïve patients may be important for future therapeutic strategies (Peres-da-Silva et al. 2010). Until recently, the identification of subtypes had no great clinical relevance in terms of treatment because HCV 1a and 1b infection are presently treated with the same therapeutic protocol. However, this situation may change once a greater number of new, specific HCV enzyme inhibitors become available for anti-HCV therapy. In this context, information regarding sequence variations among HCV subtypes and even subclades may be important for the therapeutic approach with new DAAs.

In conclusion, we believe that the results described in this paper provide useful information on the genetic profile of a large and diverse set of HCV NS3 protease sequences, which are crucial for defining its natural polymorphisms and are important in the context of the development of new protease inhibitors. The effects of the observed amino acid polymorphisms found here are presently unknown. We propose further investigations to determine the effects of genetic diversity on the enzyme structure, including molecular modelling, to better understand the relationship between sequence variation and structural changes within the NS3 protease enzyme.

REFERENCES

Alter MJ 2007. Epidemiology of hepatitis C virus infection. *World J Gastroenterol 13*: 2436-2441.

Anisimova M, Gascuel O 2006. Approximate likelihood-ratio test for branches: a fast, accurate and powerful alternative. *Syst Biol 55*: 539-552.

Asselah T, Marcellin P 2011. New direct-acting antivirals' combination for the treatment of chronic hepatitis C. *Liver Int 31*: 68-77.

Bae A, Sun SC, Qi X, Chen X, Ku K, Worth A, Wong KA, Harris J, Miller MD, Mo H 2010. Susceptibility of treatment-naïve hepatitis C virus (HCV) clinical isolates to HCV protease inhibitors. *Antimicrob Agents Chemother 54*: 5288-5297.

Chevaliez S, Bouvier-Alias M, Brillet R, Pawlotsky JM 2009. Hepatitis C virus (HCV) genotype 1 subtype identification in new HCV drug development and future clinical practice. *PLoS ONE 4*: e8209.

Enomoto M, Tamori A, Kawada N 2009. Emerging antiviral drugs for hepatitis C virus. *Rev Recent Clin Trials 4*: 179-184.

de Almeida AJ, Campos-de-Magalhães M, de Melo Marçal OP, Brandão-Mello CE, Okawa MY, de Oliveira RV, do Espírito-Santo MP, Yoshida CF, Lampe E 2004. Hepatitis C virus-associated thrombocytopenia: a controlled prospective, virological study. *Ann Hematol 83*: 434-440.

Foy E, Li K, Wang C, Sumpter R Jr, Ikeda M, Lemon SM, Gale M Jr

2003. Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease. *Science 300*: 1145-1148.

Guindon S, Gascuel O 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol 52*: 696-704.

Korber B 2000. HIV signature and sequence variation analysis. In AG Rodrigo, GH Learn, *Computational analysis of HIV molecular sequences*, Kluwer Academic Publishers, Netherlands, p. 55-72.

Korber B, Myers G 1992. Signature pattern analysis: a method for assessing viral sequence relatedness. *AIDS Res Hum Retroviruses 8*: 1549-1560.

Kuiken C, Yusim K, Boykin L, Richardson R 2005. The Los Alamos hepatitis C sequence database. *Bioinformatics 21*: 379-384.

Lai CK, Jeng KS, Machida K, Cheng YS, Lai MM 2008. Hepatitis C virus NS3/4A protein interacts with ATM, impairs DNA repair and enhances sensitivity to ionizing radiation. *Virology 370*: 295-309.

Lampe E, Espírito-Santo MP, Martins RM, Bello G 2010. Epidemic history of hepatitis C virus in Brazil. *Infect Genet Evol 10*: 886-895.

Lavanchy D 2009. The global burden of hepatitis C. *Liver Int 29*: 74-81.

López-Labrador FX 2008. Hepatitis C virus NS3/4A protease inhibitors. *Recent Pat Antiinfect Drug Discov 3*: 157-167.

López-Labrador FX, Moya A, González-Candelas F 2008. Mapping natural polymorphisms of hepatitis C virus NS3/4A protease and antiviral resistance to inhibitors in worldwide isolates. *Antivir Ther 13*: 481-494.

Mello IM, Thumann C, Schvoerer E, Soulier E, Pinho JR, Silvestre DA, Queiroz AT, Wolf P, Baumert TF, Keller FS, Pereira CA 2009. Conservation of hepatitis C virus nonstructural protein 3 amino acid sequence in viral isolates during liver transplantation. *J Viral Hepat 16*: 732-737.

Meylan E, Curran J, Hofmann K, Moradpour D, Binder M, Bartenschlager R, Tschopp J 2005. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature 437*: 1167-1172.

Murphy D, Chamberland J, Dandavino R, Sablon E 2007. A new genotype of hepatitis C vírus originating from Central Africa. *Hepatology 46*: 623A.

Nakamura M, Saito H, Hibi T 2008. Advances in genomic research on hepatitis C virus with a useful tool replicon system. *Keio J Med 57*: 75-83.

Nei M, Gojobori T 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol 3*: 418-426.

Peres-da-Silva A, de Almeida AJ, Lampe E 2010. Mutations in hepatitis C virus NS3 protease domain associated with resistance to specific protease inhibitors in antiviral therapy naïve patients. *Arch Virol 155*: 807-811.

Pickett BE, Striker R, Lefkowitz EJ 2011. Evidence for separation of HCV subtype 1a into two distinct clades. *J Viral Hepat 18*: 608-618.

Russo MW, Fried MW 2003. Side effects of therapy for chronic hepatitis C. *Gastroenterology 124*: 1711-1719.

Schinazi RF, Bassit L, Gavegnano C 2010. HCV drug discovery aimed at viral eradication. *J Viral Hepat 17*: 77-90.

Simmonds P, Smith DB, McOmish F, Yap PL, Kolberg J, Urdea MS, Holmes EC 1994. Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *J Gen Virol 75*: 1053-1061.

Tamura K, Dudley J, Nei M, Kumar S 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol 24*: 1596-1599.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG 1997. The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res 25*: 4876-4882.

Tong X, Bogen S, Chase R, Girijavallabhan V, Guo Z, Njoroge FG, Prongay A, Saksena A, Skelton A, Xia E, Ralston R 2008. Characterization of resistance mutations against HCV ketoamide protease inhibitors. *Antiviral Res 77*: 177-185.

Triyatni M, Berger EA, Saunier B 2011. A new model to produce infectious hepatitis C virus without the replication requirement. *PLoS Pathog 7*: e1001333.

Verbeeck J, Stanley MJ, Shieh J, Celis L, Huyck E, Wollants E, Morimoto J, Farrior A, Sablon E, Jankowski-Hennig M, Schaper C, Johnson P, Van Ranst M, Van Brussel M 2008. Evaluation of versant hepatitis C virus genotype assay (LiPA) 2.0. *J Clin Microbiol 46*: 1901-1906.

Vermehren J, Sarrazin C 2011. New HCV therapies on the horizon. *Clin Microbiol Infect 17*: 122-134.

Zeng R, Li G, Ling S, Zhang H, Yao Z, Xiu B, He F, Huang R, Wei L 2009. A novel combined vaccine candidate containing epitopes of HCV NS3, core and E1 proteins induces multi-specific immune responses in BALB/c mice. *Antiviral Res 84*: 23-30.