Diversity of *Hepatitis C virus* in Southern India Based on 5'UTR Sequence

R. Amjesh · Achuthsankar S. Nair ·
V. S. Sugunan

Received: 10 March 2012 / Accepted: 14 August 2012 / Published online: 25 September 2012
© Indian Virological Society 2012

Abstract *Hepatitis C virus* (HCV) exhibits genotype-specific variations in geographical distribution as a consequence of drug and immune induced evolution. Present study was aimed at discerning the distribution and prevalence of the various genotypes and subtypes of HCV in southern India. The HCV positive patient's serum was collected from different hospitals and blood banks from the states of Kerala, Tamil Nadu, Andhra Pradesh and Karnataka. Among 114 HCV positive samples, we could find only 44 isolates that are found both positive in ELISA and RT-PCR. From these samples 5' untranslated region (5'UTR) were amplified, sequenced and sub typed. Analysis of 5'UTR region of the 44 isolates shows that, genotypes 1, 3, 4 and 6 are present with genotype 3 being the most frequent. The present study shows that HCV genotype 3 subtype B was the most prevalent, forming 47.7 % among the population in southern India. The present study urges for discovering novel therapeutic agents that should be specific to genotype 3 subtype B, for the management of HCV in southern India.

Hepatitis C virus (HCV, genus: *Hepacivirus*, family: *Flaviviridae*) exists as six major genotypes and more than 60 defined subtypes. The genomic sequences of different HCV genotypes vary by 30–35 % between genotypes and by

20–25 % between subtypes [19]. Such genetic heterogeneity accounts for several of the differences observed in the prognosis of the disease, as well as in the differences seen in response to prophylactic treatment procedures. For example, in individuals infected with HCV genotype 1, complete and permanent clearance of the infection has been observed in only 10–20 % of those on monotherapy with interferon and 40–50 % in those on combination therapy with interferon and ribavirin. The disease tends to be chronic in the rest, ultimately ending in hepatocellular carcinoma. However, in the case of type 2 and 3, 50 % on mono-therapy and 70–80 % on combination therapy show sustained clearance of the virus [15, 26]. Determination of the genotype of the infecting virus is thus important not only in disease management, but also in epidemiological studies, development of efficient vaccines, designing control measures and in determining appropriate therapeutic procedures.

The distribution and prevalence of the HCV genotypes seem to have a geographic influence and specific genotypes tend to be concentrated within certain geographic limits. In India, Lole et al. [12] showed that the infection with type 3 was predominant in North, West and East Indian population, while genotype 1 was shown to be the predominant genotype in South India. The study by Hissar et al. [8] also confirmed the predominance of genotype 3 in North and Central India. The incidence of genotype 1 was also reported from North, West, and Eastern regions. In a recent study from South India, Chandra et al. [5] reported that occurrence of variants of genotype 1 in 60 % patients and variants of genotype 3 in 40 % patients. They used restriction fragment length polymorphism (RFLP) analysis of 5' untranslated region (5'UTR) and type specific primers of NS5B to arrive at this conclusion. Besides genotypes 1 and 3, genotype 4 was also reported from South India [21].

R. Amjesh (✉) · A. S. Nair
Department of Computational Biology and Bioinformatics,
University of Kerala, Thiruvananthapuram 695581, Kerala, India
e-mail: amjesh@gmail.com

V. S. Sugunan
Post-Graduate & Research Department of Zoology, HisHighness
The Maharaja's University College, Palayam,
Thiruvananthapuram 695034, Kerala, India

Later in 2005 Raghuraman et al. [16] reported the incidence of genotype 6 in India.

On a world-wide distribution analysis, HCV subtypes 1a and 1b are the most common genotypes in the United States [24], while in Japan, subtype 1b causes about 73 % of HCV infection [22]. HCV genotype 4 is predominant in North Africa and the Middle East [1, 4].

Researchers have employed several methods to analyse the genotype of HCV. RFLP analysis of 5'UTR was employed by Nakao et al. [13] making use of the conserved type-specific point mutations in the 5'UTR. Okamoto et al. [14] designed type-specific primers from the core region of HCV to develop a nested PCR method for genotyping. A serological method for identifying the specific genotype was developed by Simmonds et al. [18]. In the present study, PCR-amplified 5'UTR was sequenced and then bioinformatics tools were used to identify the biodiversity of prevalent genotypes and subtypes of HCV among the South Indian population.

The patients' samples from South Indian states comprising of Kerala, Tamil Nadu, Andhra Pradesh and Karnataka which were found positive in ELISA by the respective hospitals and blood banks were used for this study. Informed consent was obtained from each patient included in the study. The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki as reflected in a priori approval by the institution's human research committee. A total of 114 samples that were tested positive for HCV using commercial ELISA kits at respective hospitals were referred to our laboratory for conformation. The sera were once again checked by ELISA and reconfirmed for active viral infection by RT (reverse transcriptase)-PCR. Thus, only 44 sera samples from patients positive for the presence of HCV by both ELISA and RT-PCR were included in this study. The patients' blood samples were once again tested for the presence of HCV using EILSA kit (Abbott Laboratories, USA) according to the manufacturer's instruction.

The total RNA were isolated using Trizol LS reagent (Invitrogen, USA) and then solubilised in 10 μ L RNase free water and was denatured at 70 °C for 5 min and chilled instantly on ice. The RNA was immediately reverse-transcribed by using Mu-MLV reverse transcriptase enzyme (Promega, USA). The reverse transcriptase reaction mix contained 1X RT buffer 1 mM each of dNTPS, 50 μ g of random primer, 20 U of 'Rnasin' Ribonuclease inhibitor (Promega, USA) and 200 U of Mu-MLV reverse transcriptase in a total volume of 20 μ L. The reaction mix was incubated at 25 °C for 10 min and then at 37 °C for 60 min.

The amplification of HCV-RNA specific sequence was carried out by nested PCR. Reaction mix containing the cDNA was used for the first PCR amplification in a final

reaction volume of 25 μ L containing 1X PCR buffer, 1.5 mM $MgCl_2$, 200 μ M each of the four dNTPs, 1.25 U of Taq polymerase (Sigma, USA) (5 U/ μ L) and 0.5 mM each of the primers HC-1S and HC-2A. After an initial denaturation at 95 °C for 2 min, PCR was performed for 35 cycles of denaturation at 95 °C for 30 s, annealing at 56 °C for 30 s and extension at 72 °C for 20 s followed by a final extension at 72 °C for 7 min in a MJ Research thermocycler.

3 μ L of the first PCR product was then used for the second PCR. Conditions were the same except that the internal primers HC-3S and HC-4A were used instead of HC-1S and HC-2A. At the end of the second PCR, aliquots of the amplification product were run in a 2 % agarose gel in 1X TAE buffer. The lists of primers used are appended in Table 1.

As a larger 5'UTR amplicon was required for sequence analyses, the primers seqC8 and seqC9, binding at nucleotides 16–36 and 320–342 respectively, were used for the PCR from the cDNA of HCV positive samples. The amplified product was run in 2 % TAE agarose gel. The DNA band was excised and the DNA eluted from the gel using AmiconDA spin columns. The eluted DNA was quantified spectrophotometrically and then sequenced. The sequencing reaction mix consisted of 2 μ L of ABI PRISM BigDye terminator sequencing mix v3.1, 1.42 mM of $MgCl_2$, 2 pmol of either the forward or the reverse primer and 200 ng of template DNA in a total volume of 20 μ L. The reaction was programmed for 25 cycles at 96 °C for 30 s, 55 °C for 1 min and 60 °C for 4 min in a Bio-Rad iCycler PCR machine with ramping rate adjusted to 1 °C/s. At the end of the reaction, 20 μ L of nuclease free water and 60 μ L of 100 % isopropanol were added, mixed and the mix allowed to stand at room temperature for 15 min to precipitate the DNA. The PCR products were pelleted by centrifugation at 13,000 rpm for 15 min and the pellet was washed twice with 75 % isopropanol, and air-dried. The air-dried reaction products were resuspended in 20 μ L of template suppressor reagent (TSR) denatured at 95 °C for 5 min and loaded and analyzed in an ABI prism 310 automated DNA sequencer. The sequence analysis was repeated with primers for the forward and reverse direction to confirm the consistency of the sequence data obtained.

After removing the primer sequences from the sequence data obtained from the RT-PCR product, 44 sequences were submitted to GenBank and the accession codes were assigned is given in Table 2. Sequence similarity to other HCV isolates provided in the NCBI database were analysed using Basic Local Alignment Search Tool (BLAST). Genotype and subtype identities were assigned to the HCV 5'UTR sequences based on the pair wise alignment of the query sequences to the HCV full length reference sequences, using the NCBI HCV genotyping tool at

Table 1 Details of primers used

Primer	Sequences	Position
HC-1S	5'-TTCACGCAGAAAGCGTCTAGCCA-3'	61–83
HC-2A	5'-GTCTACGAGACCTCCCGGGGCAC-3'	309–331
HC-3S	5'-TGGCGTTAGTATGAGTGTCTGTCAG-3'	84–108
HC-4A	5'-CTCGCAAGCACCCCTATCAGGCAG-3'	287–311
seqC8	5'-GGCGACACTCCACCATAGATC-3'	16–36
seqC9	5'-CATGGTGCACGGTCTACGAGACC-3'	320–342

Positions of sequence of primers are based on HCV3b, Acc. no. D49374

S sense primer, A anti-sense primer

Table 2 The GenBank accessions of the 5'UTR sequence of *hepatitis C virus* genotypes detected in the southern India under the present study

Subtype	NCBI accession no.	No. of isolates	% Composition
1a	DQ284921, DQ284922, DQ284958, DQ284923, DQ284924, DQ284959, DQ284925	7	15.9
1b	DQ284926, DQ284927	2	4.6
1c	DQ284928, DQ284929	2	4.6
3a	DQ284930, DQ284931, DQ284932, DQ284933, DQ284934, DQ284935, DQ284935	7	15.9
3b	DQ284937, DQ284938, DQ284939, DQ284940, DQ284941, DQ284942, DQ284943, DQ284944, DQ284945, DQ284946, DQ284947, DQ284948, DQ284949, DQ284950, DQ284951, DQ284952, DQ284953, DQ284954, DQ284955, DQ284956, DQ284957	21	47.7
4a	DQ284960, DQ284961, DQ284962	3	6.8
6g	DQ284963	1	2.3
6h	DQ284964	1	2.3

<http://www.ncbi.nih.gov/projects/genotyping/formpage.cgi> [17]. In order to determine the degree of genetic heterogeneity, evolutionary analysis was done by aligning the 5'UTR sequences of the local isolates with the genotype references from NCBI genotyping tool using Clustal X. The output data was used to generate phylogenetic information in MEGA 3.1 [10, 23]. Phylogenetic trees were created by the neighbour-joining method applied to the distance matrix obtained under the Kimura two-parameter model (<http://evolution.genetics.washington.edu/phylip>). Both interior branch test and boot strap test with 1000 pseudo-replicas was employed to test the robustness of each node.

Analysis of the 5'UTR sequence data of the isolates by NCBI BLAST [3] and NCBI genotyping tools showed that genotypes 1, 3, 4 and 6 occur in the South Indian population. Among the 44 isolates analyzed, genotype 3 was the most prevalent, forming 64 % of the cases and genotype 6, the least prevalent, with a mere share of 4.5 %. Genotype 1 accounted for 25 % and genotype 4 accounted for 6.82 %. The subtypes 1a, 1b, 1c, 3a, 3b, 4a, 6g and 6h were identified among the samples. The percentage composition of the viral isolates is shown in Table 2.

Pair wise alignment of sequences of the isolates identified in this study, prefixed with letter LI* in the data against the HCV genotype reference sequences from NCBI, showed variations in sequences among the local isolates from the corresponding genotype reference sequences. Figure 1 shows the unrooted phylogenetic tree, generated using the neighbour-joining method, under Kimura 2 parameter model for nucleotides, employing interior branch test of 1,000 replicas. All the local isolates were seen clustered to match their genotypes, confirming the results obtained using the NCBI genotyping tool. However, unlike the results obtained by sub-typing using the NCBI genotyping tool (Table 2), the isolates, DQ284941, DQ284948, DQ284953, DQ284940, DQ284954, formed a separate cluster within the 3b subtype group. Isolates DQ284937, DQ284939, DQ284942, DQ284943, DQ284944, DQ284945, DQ284946, DQ284947, DQ284950, DQ284951, DQ284952, DQ284955 and DQ284956 formed a separate cluster with a node value of 85. The 3b subtype reference D49374 along with DQ284957 and DQ284949 formed a sub cluster with a node value of 64 and DQ284938 alone formed a sub cluster of this main cluster with a node value of 85. All other isolates were either aligned with the main cluster within their respective subtype

Fig. 1 Phylogenetic tree of HCV 5'UTR sequences constructed using Neighbor Joining method, under Kimura 2 parameter model for nucleotide, employing interior branch test of 1,000 replicas. The tree were constructed for 76 nucleotides in which 44 sequences from local isolates (LI) under the present study and the rest from the reference sequences from the database

reference sequences, or sub-clustered from the subtype reference sequences.

The results of the study indicate the prevalence of the various genotypes and subtypes of HCV among the South Indian population. The preliminary data also gives an idea of their genetic variability. The evolution of viruses, like that of all other organisms, is a process that is ultimately dependent on mutations in their genetic material. RNA viruses exploit several intrinsic mechanisms of genetic variations to ensure their survival [7]. Among viruses, HCV exhibits a significantly high genomic variability with six recognized genotypes. Each of these genotypes have still more variants forming subtypes, and there are reports on emergence of new HCV subtypes [20]. The six major genotypes exhibit divergence in nucleotide sequence of more than 30 %. The inability of the RNA dependant RNA polymerase (RdRp) enzyme to proof-read the replicating RNA increases the chances of the formation of variants and mutants. The higher number of mutations enables the evolution of specific variants, which is obviously driven by the local phenotype of the host. Selection pressure due to the sero-conservation and the treatment process will also influence such evolution. Moreover, the large population size of HCV within a single host enables the variants to evolve very rapidly as a sequel to immunological or other selection pressures including administered drugs. Thus, host factors and duration of infection also play a significant role in the emergence of new variants. The adaptive and evolutionary process would continue even as it gets transmitted from one individual to another. The HCV has also been shown to resort to RNA recombination as a means of generating genetic variations. An instance of such inter-genotypic recombination was observed between genotype 2k and 1b in St. Petersburg by Kalinina et al. [9]. Colina et al. [6] reported intra-typic recombination between genotype 1a and 1b. The potential for the emergence of novel genetic variants is high due to the increasing subtype diversity in various sub populations. The genetic differences among HCV subtypes will turn out to be critical in the clinical and epidemiological management of the disease.

It is evident that identifying the genotype and subtype of HCV infecting a patient is of high significance in the management of the disease and in the design and development of drugs. Such data are also important and useful in keeping track of the source and course of infection, both in the patient and the population as it is obvious that the type,

severity and prognosis of the disease depend on the type and subtype of the virus infecting the patients [25]. This fact is a major point of concern as no single drug has so far been able to have the same effect in all the genotypes or even in subtypes of HCV [2, 11]. One of the major reasons for the differences observed in the response to drugs, many of which are molecules developed as inhibitors of the key enzymes such as NS3 protease and RdRp, could be the genotype-specific differences in the sequences of the translated target proteins. Disease progression and clinical prognosis may also be affected by biological differences in the manifestations of infection with different subtypes. Knowledge and analyses of such divergence and heterogeneity of the sequence in the local isolates is also important in vaccine research and in understanding the host pathogen interaction.

Acknowledgments This work was supported by Kerala Government grant for the State Inter-University Centre for Excellence in Bioinformatics.

References

1. Abdulkarim AS, Zein NN, Germer JJ, Kolbert CP, Kabbani L, Krajnik KL, Hola A, Agha MN, Tourogman M, Persing DH. Hepatitis C virus genotypes and hepatitis G virus in hemodialysis patients from Syria: identification of two novel hepatitis C virus subtypes. *Am J Trop Med Hyg.* 1998;59:571–6.
2. Abrignani S, Houghton M, Hsu HH. Perspectives for a vaccine against hepatitis C virus. *J Hepatol.* 1999;31:259–63.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
4. Chamberlain RW, Adams NJ, Taylor LA, Simmonds P, Elliott RM. The complete coding sequence of hepatitis C virus genotype 5a, the predominant genotype in South Africa. *Biochem Biophys Res Commun.* 1997;236:44–9.
5. Chandra M, Thippavuzula R, Ramachandra Rao VV, Habib AM, Habibullah CM, Narasu L, Prameela Y, Khaja MN. Genotyping of hepatitis C virus (HCV) in infected patients from South India. *Infect Genet Evol.* 2007;7(6):724–30.
6. Colina R, Casane D, Vasquez S, Garcia-Aguirre L, Chunga A, Romero H, Khan B, Cristina J. Evidence of intratypic recombination in natural populations of hepatitis C virus. *J Gen Virol.* 2004;85:31–7.
7. Domingo E, Holland JJ. RNA virus mutations and fitness for survival. *Annu Rev Microbiol.* 1997;51:151–78.
8. Hissar SS, Goyal A, Kumar M, Pandey C, Suneetha PV, Sood A, Midha V, Sakhuja P, Malhotra V, Sarin SK. Hepatitis C virus genotype 3 predominates in North and Central India and is associated with significant histopathologic liver disease. *J Med Virol.* 2006;78(4):452–8.
9. Kalinina O, Norder H, Mukomolov S, Magnus LO. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol.* 2002;76:4034–43.
10. Kumar S, Tamura K, Nei M. MEGA: molecular evolutionary genetics analysis software for microcomputers. *Comput Appl Biosci.* 1994;10:189–91.
11. Lemon SM, Brown EA. Hepatitis C virus. In: Mandell GL, Bennett JE, Dolin R, editors. *Principle and practice of infectious disease*, 4th ed. New York; 1995. p. 1474–1486.
12. Lole KS, Jha JA, Shrotri SP, Tandon BN, Prasad VG, Arankalle VA. Comparison of hepatitis C virus genotyping by 5' noncoding region- and core-based reverse transcriptase PCR assay with sequencing and use of the assay for determining subtype distribution in India. *J Clin Microbiol.* 2003;41:5240–4.
13. Nakao T, Enomoto N, Takada N, Takada A, Date T. Typing of hepatitis C virus genomes by restriction fragment length polymorphism. *J Gen Virol.* 1991;72:2105–12.
14. Okamoto H, Sugiyama Y, Okada S, Kurai K, Akahane Y, Sugai Y, Tanaka T, Sato K, Tsuda F, Miyakawa Y, et al. Typing hepatitis C virus by polymerase chain reaction with type-specific primers: application to clinical surveys and tracing infectious sources. *J Gen Virol.* 1992;73:673–9.
15. Pawlotsky JM. Mechanisms of antiviral treatment efficacy and failure in chronic hepatitis C. *Antivir Res.* 2003;59:1–11.
16. Raghuraman S, Abraham P, Sridharan G, Ramakrishna BS. Hepatitis C virus genotype 6 infection in India. *Indian J Gastroenterol.* 2005;24:72–3.
17. Rozanov M, Plikat U, Chappey C, Kochergin A, Tatusova T. A web-based genotyping resource for viral sequences. *Nucleic Acids Res.* 2004;32(Web Server issue):W654–9.
18. Simmonds P, Rose KA, Graham S, Chan SW, McOmish F, Dow BC, Follett EA, Yap PL, Marsden H. Mapping of serotype-specific, immunodominant epitopes in the NS-4 region of hepatitis C virus (HCV): use of type-specific peptides to serologically differentiate infections with HCV types 1, 2, and 3. *J Clin Microbiol.* 1993;31:1493–503.
19. Simmonds P. Genetic diversity and evolution of hepatitis C virus—15 years on. *J Gen Virol.* 2004;85:3173–88.
20. Simmonds P. Viral heterogeneity of the hepatitis C virus. *J Hepatol* 1999;31 Suppl 1: 54–60.
21. Singh S, Malhotra V, Sarin SK. Distribution of hepatitis C virus genotypes in patients with chronic hepatitis C infection in India. *Indian J Med Res.* 2004;119:145–8.
22. Takada N, Takase S, Takada A, Date T. Differences in the hepatitis C virus genotypes in different countries. *J Hepatol.* 1993;17:277–83.
23. Thompson JD, Higgins DG, Gibson TJ. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22:4673–80.
24. Zein NN, Rakela J, Krawitt EL, Reddy KR, Tominaga T, Persing DH. Hepatitis C virus genotypes in the United States: epidemiology, pathogenicity, and response to interferon therapy. Collaborative study group. *Ann Intern Med.* 1996;125:634–9.
25. Zein NN. Clinical significance of hepatitis C virus genotypes. *Clin Microbiol Rev.* 2000;13:223–35.
26. Zeuzem S. Heterogeneous virologic response rates to interferon-based therapy in patients with chronic hepatitis C: who responds less well? *Ann Intern Med.* 2004;140:370–81.