



Identification of hepatitis C virus genotypes and subtypes in patients from Maharashtra based on 5'UTR sequencing and analysis: a retrospective study

Ushanandini Mohanraj¹, Yashwant G Chavan², Sharad R Pawar³, Amol D Raut³

¹Department of Biotechnology, Anna University, Chennai, Tamil Nadu, India.

²Dr. D. Y. Patil Biotechnology and Bioinformatics Institute, Tathwade, Pune, Maharashtra, India.

³Molecular Virology, geneOmbio Technologies Private Limited, Baner, Pune, Maharashtra, India.

Received for publication: October 21, 2015; Accepted: November 12, 2015

Abstract: The clinical importance of HCV genotypes lies in its persistence and ability to cause chronic liver disease. Knowledge of HCV genotype distribution periodically is now essential for the clinical management of infected patients. The current study was designed to genotype and subtype HCV isolates in patients from Pune, Mumbai and Nashik districts of Maharashtra, determine association of HCV genotypes between both sexes and different age groups and perform phylogenetic analysis of the study isolates. Eighty HCV RNA positive patients were genotyped and subtyped in the study. Primers targeting 5'untranslated region (UTR) were designed for both amplification and sequencing of all isolates. All HCV isolates were characterized by direct sequencing of 5' UTR followed by analysis with NCBI genotyping tool and in-silico BLAST analysis. The results were correlated with the Los Alamos Hepatitis C sequence and immunology database's sequence alignment tool HCVAlign and phylogenetic analysis. Most of HCV isolates in this study belonged to genotype 3(56.25%) followed by genotype 1(35%) and genotype 4(7.5%). A rare genotype 5(1.25%) was identified in one of the patients. Prevalent subtypes were: subtype 3a (42.5%), subtype 1a (21.25%) and subtype 1b (13.75%). No notable association of age and gender with respect to genotypes was observed except for genotype 1a which showed an overrepresentation in males. Compared with phylogenetic analysis, NCBI genotyping tool was simpler for genotyping and subtyping HCV isolates. The present study revealed that HCV genotypes 3 and 1 were prevalent in Pune, Mumbai and Nashik districts of Maharashtra. With rare genotype (genotype 5) being identified and due to very few genotyping studies carried out in this region, this study aims at contributing towards monitoring the distribution of HCV genotypes among studied regions of Maharashtra.

Key words: HCV genotypes; 5'UTR; primers; sequencing; phylogenetic analysis; age; gender; Maharashtra.

INTRODUCTION

The Hepatitis C Virus (HCV), often dubbed as the "silent killer", represents a major global health burden affecting an estimated 185 million individuals worldwide [1]. In India, approximately 15 million people are infected with chronic HCV infection [2]. Due to the highly asymptomatic nature of HCV infection, most cases remain undiagnosed.

HCV is a small enveloped virus with a single stranded positive-sense RNA [3]. It belongs to the Flaviviridae family and is the member of the Hepacivirus genus [4]. The 9.5 kilobases long HCV genome consists of a 5' Noncoding Region (NCR), followed by an ORF (Open Reading Frame) that codes for structural and Nonstructural (NS) proteins and a 3' NCR [5]. HCV isolates are classified into seven major genotypes and their respective subtypes, based on sequence heterogeneity [6]. HCV genotypes have more than 30% sequence divergence between them while each subtype differs by 10-30% [7].

The distribution of HCV genotypes and subtypes vary worldwide. HCV subtypes 1a and 1b are the most common genotypes in the United States [8]. Types 1, 2 and 3 accounted for almost all infections in donors from Scotland, Finland, The Netherlands and Australia. Types 2 and 3 were not found in the eastern European country (Hungary), where all but one of the donors was infected with type 1. Patients from Japan and Taiwan were infected only with type 1 or 2, while types 1, 2, and 6 were found in those from Hong Kong. HCV infection among Egyptians was almost always by type 4 [9]. HCV genotype 5 is prevalent in South Africa [10]. In Pakistan, predominant

genotype is 3a followed by 1a, 3b, 1b and 2a [11]. Further genotype incidence varies based on age and gender of the patients [12].

Many interferon and non-interferon based therapies exist for treatment of Hepatitis C infection. It has been suggested that different genotypes have different clinical outcomes with respect to disease severity and response to treatment regimens [13]. Tracing the heterogeneity of HCV genotypes in various epidemiological settings is critical for both disease surveillance and development of improved antiviral treatment. Following methods have been used often for characterization of HCV genotypes: Inno-Line Probe Assay, Direct sequence analysis, PCR-RFLP, Type-specific PCR, Nested PCR [14][15]. PCR amplification followed by direct sequence analysis is considered as the gold standard for genotyping studies [16]. The high degree of conservation in the 5' noncoding regions (5' UTR) has made it the target of choice for reverse transcriptase PCR-based detection assays [17].

The current study was carried out to genotype and subtype HCV isolates circulating among the population in Pune, Mumbai and Nashik districts of Maharashtra. Genotyping was performed by PCR amplification of 5'UTR region using a newly designed set of oligonucleotide primers followed by sequencing and analysis. The distribution of genotypes and subtypes in the considered study population and association of HCV genotypes between both sexes and different age groups was analyzed. Further, phylogenetic analysis of the study isolates was performed.

*Corresponding Author:

Ushanandini Mohanraj,
Department of Biotechnology,
Anna University, Chennai,
Tamil Nadu, India.

MATERIALS AND METHODS

Sample Information

The retrospective study was conducted in geneOmbio Technologies, Pune, Maharashtra, India. Five hundred and twelve patient blood samples that were sent to geneOmbio Technologies between 2009-2015 for Hepatitis C diagnosis from hospital laboratories and public or private clinics located in Pune, Mumbai and Nashik districts of Maharashtra were considered for analysis. Further, Informed consent was obtained from the geneOmbio Technologies IRB for the current retrospective study according to the WMA Declaration of Helsinki 1969.

Table I: Demographic profile of the study group

Characteristic	Data
Anti HCV positive, n	80
Age (y, mean±SD)	44.04±13.9
Men/Women, n (%)	49/31 (61.25%/38.75%)

Inclusion criteria

Of the 512 samples analysed, only sera from 80 anti HCV positive patients (49 males, 31 females, mean age 44.04±13.9 years) were included in this study. All sera were tested for HCV positivity with the rapid assay HCV TRI DOT (J. MITRA &Co. Ltd., New Delhi, India).

Exclusion criteria

Patient's positive for Hepatitis B surface antigen (HBsAg) and Human Immunodeficiency Virus (HIV) were excluded from the study. Screening for HBsAg was done using commercially available Hepalisa kit (J. MITRA &Co. Ltd., New Delhi, India) and screening for HIV was performed using HIV TRI DOT (J. MITRA &Co. Ltd., New Delhi, India).

Viral RNA extraction

Viral RNA extraction from patient serum was performed using Roche's High pure viral nucleic acid kit (Roche Diagnostics, Mannheim, Germany) according to manufacturer's instruction. For the RNA extraction protocol, 500µl of serum was taken. RNA precipitate was re-suspended in 75µl elution buffer and stored at -80 °C.

Oligonucleotide Primer Design

Reference sequences representing all the HCV genotypes and subtypes were retrieved from the GenBank (<http://www.ncbi.nlm.nih.gov/projects/genotypi/ng/view.cgi?db=3>) in FASTA format. Multiple sequence alignment of the retrieved sequences was carried out using ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) and highly conserved regions in 5'UTR were identified. The sequences were then used for generating primers in Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) to bind to highly conserved regions in 5'UTR. The generated primer sequences from the Primer3 software were fed in the primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) and cross checked for specificity and possibility of undesired products in the non-redundant database. The designed oligonucleotide sequences were sent to Eurofins

Ltd, Bangalore for synthesis. Table II represents the sequences and the nucleotide positions of the primers that were used in the present study.

Table II: Primers used for HCV genome amplification and sequencing

Primers	Product length	Position region 5'UTR
Outer sense 5UHCVOF TCCCCTGYGAGGAACTWCTGT	–	41-61*
Outer antisense 5UHCWOR – TGACACGGKCTACGAGACCT	299 bp	339-321*
Inner sense 5UHCVIF- CTTCACGCRGAAAGCGYCTA	246 bp	41-22 [†]
Inner antisense 5UHCVIR-CAAGCACCTATCAGGCAGT		267-248 [†]

*KP098533.1 and [†] KM281664.1 - reference sequences that was used to calculate the positions of the UTR region.

5'UTR amplification

The one-step RT-PCR system combining MMLV reverse transcriptase with *Taq* DNA polymerase was used. The combined RT-PCR amplifications were carried out with 25 µl reaction mixture containing 10µl of sample RNA, 0.5µl of 0.2mM deoxyribonucleoside triphosphates (dNTPs) (dATP, dCTP, dGTP, and dTTP), 0.75µl of MgCl₂ (1.5mM), 10pmol (each) primers 5UHCVOF and 5UHCWOR, 1U/µl of *Taq* DNA polymerase (Invitrogn, Carlsbad, CA), 2.5µl of 1x PCR buffer (10x PCR buffer consists of 200 mM TrisHCl [pH 8.4] and 500 mM KCl), 100 U/µl of Moloney murine leukemia virus reverse transcriptase (Invitrogen, Carlsbad, CA) and Milli-Q water. The reactions were performed in Applied Biosystems 2720 Thermal Cycler. Reverse Transcription was allowed to proceed for 60 min at 50°C and was followed by a 5 min incubation at 95°C to facilitate denaturation of RNA-DNA heteroduplexes. PCR amplification proceeded with 45 cycles at 94°C for 30 s, 55°C for 1 min and 72 °C for 1 min followed by a final extension step of 5 min at 72°C.

The RT-PCR product was used as a template for the nested PCR amplification with 25µl reaction mixture. 5 µl of RT-PCR product was transferred to a second tube containing the same medium as one step RT PCR except M-MLV Reverse Transcriptase enzyme but with the inner pair of forward primer (5UHCVIF) and reverse primer (5UHCVIR). Nested PCR was carried out with initial denaturation at 95°C for 5 min followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 40s, and final elongation at 72°C for 5 min.

Qualitative analysis of nested PCR products and detection of HCV RNA:

Upon completion of the amplification reaction, 5 µl of each reaction sample was analyzed by electrophoresis using a 2% agarose gel in Tris-borate-EDTA buffer (pH 8.0) and the gel was visualized using Bio-Rad Chemi XRS Gel Documentation system (Richmond, CA) for identifying desired 246 bp fragment. Internal positive and negative controls were included.

Sequencing Reaction

PCR products were purified using Exo/SAP (Fermentas Inc., Hanover, MD, USA). Sequencing of purified PCR product templates was done by sanger dideoxy chain termination method using the Big-Dye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, California, USA). Sequencing of the 5' UTR was performed in forward direction. The reaction mixture contained 0.5 µl of Terminator Ready Reaction Mix, 5.7 µl of MilliQ water, 10 pmol of primer (5UHCVIF), and 100ng of the PCR product. Cycle sequencing was carried out with the Applied Biosystems 2720 Thermal Cycler under the following conditions: 25 cycles of 96°C for 30 s, 50°C for 15 s, and 60°C for 4 min. Isolates were purified from excess incorporated dyes resulted from sequencing reaction using ethanol/EDTA purification. Then samples were

resuspended into Hi-Di formamide (Applied Biosystems, FosterCity, California, USA). Sequencing reaction products were analysed on an automated sequencer; ABIPRISM 3130 genetic Analyser (Applied Biosystems, FosterCity, California, USA).

Genotyping and Subtyping of isolates

The isolates in the present study were genotyped and subtyped using NCBI genotyping tool [18] (<http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi>) and the online Los Alamos Hepatitis C sequence and immunology database's sequence alignment tool HCV Align: (<http://hcv.lanl.gov/content/sequence/VIRALIGN/viralalign.html>) [19]. The sequences were also subjected to blast and analyzed in-silico (<http://www.ncbi.nlm.nih.gov>).

Table III: The accession numbers of reference sequences representing genotypes and subtypes used for phylogenetic analysis

Genotype 1		Genotype 2		Genotype 3		Genotype 4		Genotype 5		Genotype 6		Genotype 7	
Subtype	Accession No.	Subtype	Accession No.	Subtype	Accession No.	Subtype	Accession No.	Subtype	Accession No.	Subtype	Accession No.	Subtype	Accession No.
1	KC248195	2	JF735110	3	JF735124	4	JX227964	5a	NC_009826	6	JX183552	7a	EF108306
1a	NC_004102	2a	NC_009823	3a	NC_009824	4a	NC_009825			6a	DQ480513		
1b	U89019	2b	D10988	3b	JQ065799	4b	FJ462435			6b	NC_009827		
1c	AY651061	2c	D50409	3g	JF735123	4c	FJ462436			6c	EF42629		
1e	KC248194	2d	JF735114	3h	JF735121	4d	FJ462437			6d	D84263		
1g	AM910652	2e	JF735120	3i	FJ407092	4f	EU392169			6e	DQ314805		
1h	KC248198	2i	DQ155561	3k	JF735122	4g	FJ462432			6f	DQ385760		
1l	KC248193	2j	JF735113			4k	FJ462438			6g	D63822		
		2k	AB031663			4l	FJ839870			6h	D84265		
		2m	JF735111			4m	FJ462433			6i	DQ835768		
		2q	FN666429			4n	FJ462441			6j	DQ835769		
		2r	JF735115			4o	FJ462440			6k	D84264		
						4p	FJ462431			6l	EF424628		
										6m	DQ835763		
										6n	DQ835768		
										6o	EF424627		
										6p	EF424625		
										6r	EU408328		
										6s	EY409329		
										6t	EF632069		
										6u	EU246940		
										6v	EU798760		
										6w	DQ278892		

Phylogenetic tree construction

The sequences obtained in this study have been submitted to GeneBank and can be retrieved with the accession numbers KP782003 to KP782022 and KP888574 to KP888633. All isolates sequenced in the present study were aligned with the representative sequences for each major genotype from GeneBank using the Multiple Sequence Alignment Program, ClustalW (<http://www.genome.jp/tools/clustalw/>). The phylogenetic tree construction for HCV isolates was performed with MEGA6.06 software [20]. Phylogenetic tree was constructed by UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method using Maximum composite likelihood algorithm. The reliability of different phylogenetic groupings was evaluated by the bootstrap re-sampling test (1,000 bootstrap replications). Accession numbers of the reference sequences is summarized in Table III.

Statistical Analysis

Statistical Package for the Social Sciences (SPSS) version 20 software (SPSS Inc., Chicago, IL, USA) was used for statistical analysis. Statistical significance was assessed based on the p-values ($p < 0.05$).

RESULTS

HCV RNA detection and 5'UTR amplification

HCV RNA was detected in all the anti HCV positive samples considered (N=80). The primer pairs (5UHCVOF, 5UHCVOR and 5UHCVIF, 5UHCVIR) were efficient in amplifying 5'UTR of all the sample isolates. These primers amplified 5'UTR of genotypes 1, 3, 4 and 5. Among the analysed samples, no non-specific amplification was detected. The primer concentration was found to be optimized with no significant primer-dimer formation observed on the agarose gel. A representative gel picture of the PCR products of 5'UTR and controls are shown in figure1.

Table IV: Sex ratio and mean age values of patients among the various genotypes

Genotype	1 (n=28)		3 (n=45)				4 (n=6)		5 (n=1)	
subtype	1a	1b	3a	3g	3i	3k	4c	4d	4l	5a
No. of samples (%)	17 (21.2)	11(13.75)	34(42.5)	4(5)	5(6.25)	1(1.25)	1(1.25)	2(2.5)	1(1.25)	1(1.25)
Male/Female (%)	14/3	6/5	19/15	4/0	0/5	1/0	1/0	1/1	0/1	1/0
Age, years	42.2±13.6	46.2±16.8	42.2±12.8	54.5±9.5	49.4±15.9	71	45	44.5±17.7	25	33

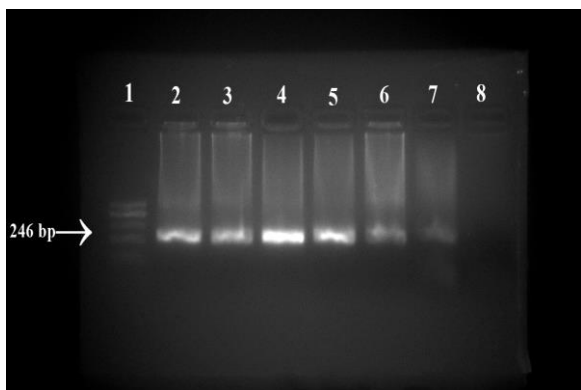


Figure 1. HCV 5'UTR Nested PCR. RT-PCR products were amplified using 5UHCVF1F and 5UHCVF1R primer set. Lane1: ladder, Lane 2: positive control, Lane 3, 4, 5, 6,7: sample isolates, Lane 8: negative control.

Genotypes and subtypes identified in the studied population and its distribution

A total of 80 patient's sera was genotyped and subtyped. Demographic data for the study population is summarised in Table I. The genotype distribution in the studied population was as follows: 56.25% (45/ 80) genotype 3, 35% (28/80) genotype 1, 7.5% (6/80) genotype 4 and 1.25% (1/80) genotype 5. Genotypes 2, 6 and 7 were not found in the study population. The subtypes followed the distribution as : 42.5% (34/80) subtype 3a, 21.25% (17/80) subtype 1a, 13.75% (11/80) subtype 1b, 6.25% (5/80) subtype 3i, 5% (4/80) subtype 3g, 1.25% (1/80) subtype 3k, 1.25% (1/80) subtype 5a, 1.25% (1/80) subtype 4l, 2.5% (2/80) subtype 4d, 1.25% (1/80) subtype 4c. Three isolates were not subtyped.

Genotype and subtype distribution according to gender and age

The distribution data of genotypes and subtypes with respect to gender and age is summarized in table IV. A greater proportion of men were infected with subtype 1a than women. The other subtypes showed no notable associations with regard to sex ratio. However, in the current study population, only males were infected with subtype 3g and only females were infected with subtype 3i. Further analysis of the subtype distribution according to age showed that donors infected by type 1b (mean age, 46.2±16.8 years) and 3g (mean age, 54.5±9.5) were older than patients infected with genotype 1a (mean age, 42.2±13.6) and 3a (mean age, 42.2±12.8) respectively.

Comparison of HCV isolates with reference sequences and the position on the phylogenetic tree

The phylogenetic analysis was performed for the 80 HCV RNA-positive samples (Figure 2A). The results demonstrated that 28 strains (28/80, 35%) clustered with genotype 1 (NC_004102.1). In the phylogenetic tree, KP782016 shared common ancestor with genotype 1 isolates and genotype 6 (DQ480513.1). However, further analysis using sequence alignments for the same gave 99% identity with complete genomes of genotype 3. Six isolates (KP888619, KP888609, KP888621, KP888633, KP888620 and KP888576) branch together with the reference sequence for genotype 4 (NC_009825.1). One isolate KP888632 clustered with NC_009825.1. But analysis using BLAST for the same gave 99% identity with 5'UTR of genotype 3. Also, NCBI genotyping tool assigned isolate KP888632 with genotype 3. One isolate KP88603 clustered with reference sequence for genotype 5 (NC_009826.1). 43 strains (43/80, 53.75%) clustered with reference sequence for genotype 3 (NC_009824.1). None of the strains clustered with reference sequence for genotype 2 (NC_009823.1) and 7 (EF108306.1). The mean pairwise distance of these 80 isolates and the 7 reference sequence was 0.048 base differences per nucleotide site.

Phylogenetic tree was constructed for the reference sequences retrieved from GenBank representing all genotypes and subtypes (Figure 2B). It was noticed that subtypes of genotype 1 shared the same phylogenetic branch with most of the genotype 6 subtypes. Further reference sequence for subtypes 3k is closely related with the reference sequences for 4 and 4h. The mean pairwise distance of these 65 sequences was 0.042 base differences per nucleotide site.

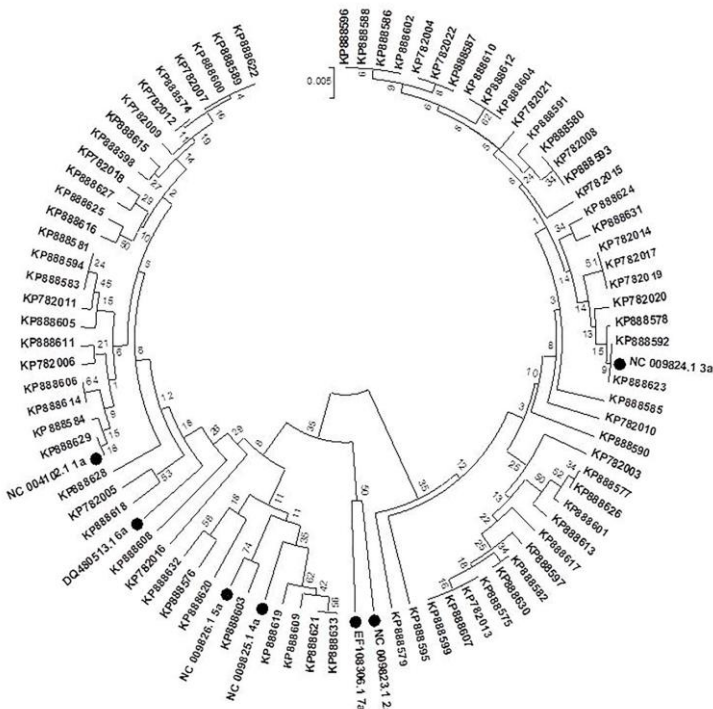
Estimated nucleotide distances between genotypes 1 to 7 were calculated (Table V). Genotype 1 and 6 had the lowest distance of 0.016 while genotype 2 and 3 had the largest distance value of 0.098.

Table V: Estimates of Evolutionary Divergence between reference sequences for genotypes 1to7 in 5'UTR. The number of base differences per site from between sequences are shown.

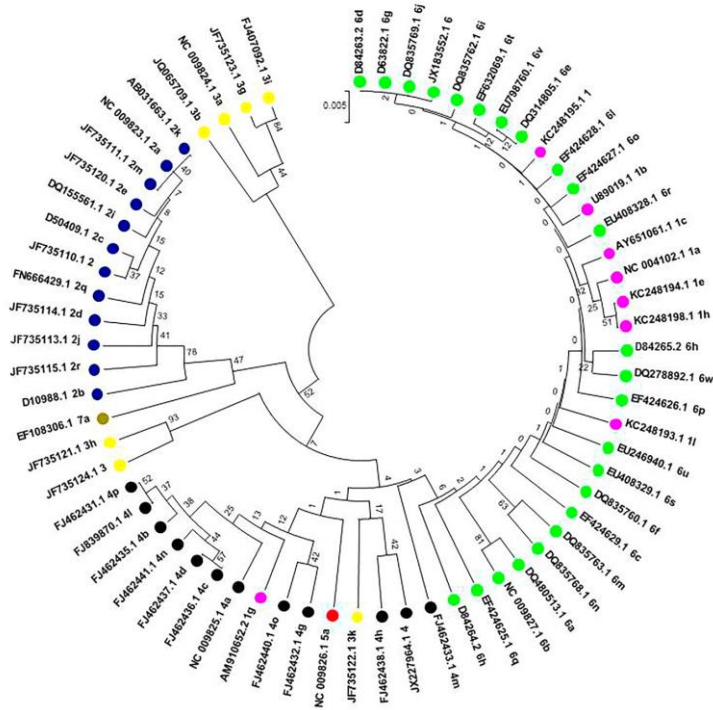
Genotypes*	3a	4a	5a	1a	6a	2a	7a
3a							
4a	0.069						
5a	0.065	0.033					
1a	0.073	0.041	0.024				
6a	0.089	0.049	0.033	0.016			
2a	0.098	0.057	0.045	0.049	0.033		
7a	0.081	0.053	0.033	0.045	0.053	0.037	

*Accession numbers of sequences used for representation of each genotype: 1a- NC_004102.1, 2a- NC_009823.1, 3a-

NC_009824.1, 4a- NC_009825.1, 5a- NC_009826.1, 6a- DQ480513.1, 7a- EF108306.1



[2A]



[2B]

Figure 2: Phylogenetic tree of HCV 5'UTR partial sequences. The evolutionary history was inferred using the UPGMA method [21]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method [22] and are in the units of the number of base differences per site. Evolutionary analyses were

conducted in MEGA6 [20]. **A]** Evolutionary relationships of 80 study isolates with the reference sequence for various genotypes. The optimal tree with the sum of branch length = 0.47816505 is shown. The analysis involved 87 nucleotide sequences. **B]** Evolutionary relationships of reference sequences representing HCV genotypes and subtypes. The optimal tree with the sum of branch length = 0.51361064 is shown. The analysis involved 65 nucleotide sequences. Sequences were labelled according to the genotypes: Pink- genotype 1, Blue- genotype 2, Yellow- genotype 3, Black- genotype 4, Red- genotype 5, Green- genotype 6 and Dijon - genotype 7.

DISCUSSION

In the present study, genotyping of the 80 isolates was performed by direct nucleotide sequencing followed by analysis with NCBI genotyping tool. Further, the results were correlated with the Los Alamos Hepatitis C sequence and immunology database's sequence alignment tool HCV Align and phylogenetic analysis. NCBI genotyping tool uses scored BLAST pairwise alignments between overlapping segments of a query sequence and a reference sequence for each virus [18]. The accuracy of genotyping HCV by BLAST analysis depends on the number of genotyped sequences in the HCV database. With genotyped sequences being continuously submitted, genotyping HCV by BLAST analysis has become increasingly more reliable [23]. In our study, careful attention to the design of primer sets was given as it is essential for optimal amplification of serum HCV RNA and effective characterization. Hence the primer sets were designed such that it amplified the genotypes 1, 3, 4 and 5. We examined the 5'UTR region for domains that were completely conserved among HCV reference sequences representing all the genotypes and subtypes from GenBank. The oligonucleotide primers designed for this study also functioned as efficient nucleotide sequencing primers and were capable of consistently generating high quality read of bases.

The objective of the present study was to genotype and subtype HCV isolates circulating in patients from Pune, Mumbai and Nashik districts of Maharashtra. Information regarding the prevalence of different genotypes is crucial for monitoring the spread of Hepatitis C in this state. HCV genotypes show a varied distribution pattern in different regions of India. From our analysis of HCV infection in patients (N=80) from Maharashtra, presence of genotypes 1, 3, 4 and 5 was observed. Type 3 was the predominant genotype followed by type 1. Our findings correlate with the various genotyping studies carried out previously across different Indian regions [24][25][26]. Further categorization of genotypes revealed the presence of following subtypes in our study population: 1a, 1b, 3a, 3i, 3g, 3k, 4l, 4d, 4c and 5a. Prevalence of 3a, 3i, 3g, and 3k subtypes of genotype 3 in Indian patients indicates its origins long back in time as compared to other genotypes. Genotype 3a sequences appear to have been originated in India and later dispersed to United Kingdom around mid1940s, most likely around the time of Indian independence and World War II [27]. HCV genotype 4 is known to be geographically restricted [28] and predominates mostly in the Arab countries [29].

Although in India, HCV genotype 4 has been reported exclusively from Southern regions [30][31], in our current study, six patients from Pune, Maharashtra were infected with genotype 4. The results show that genotype 4 was no longer restricted to the southern Indian states. The spread of genotype 4 from Arab nations to the Southern part of India can be attributed to the large number of labour migration from this region to the Arab nations [32]. Further, a rare genotype 5a was identified in one of the patients from Pune. Genotype 5 is the predominant genotype in South Africa [33]. Occurrence of rare genotypes can be explained as the sporadic cases of transmission of genotypes from other geographical regions, where endemicity of these genotypes is higher. In the present study, none of the patients were infected with genotype 2. A similar result was also previously reported [34]. Absence of genotype 2 in our study can be attributed to the limited occurrence of genotype 2 in the Indian population. Furthermore, absence of genotype 6 isolates in our study supports finding from an earlier study stating genotype 6 to be prevalent exclusively in patients from North Eastern parts of the country [35].

Studies focusing on the relationship of HCV genotypes and subtypes with gender and age have been carried out on a limited extent. In this study, gender wise and age wise distribution of HCV genotypes were analysed. Subtype 1a showed an overrepresentation among males. However no significant association of other genotypes with gender was noticed as mentioned in the earlier studies [36][37]. Further preceding studies from different countries have reported that patients infected with genotypes 1b or 2 were older than patients infected with genotypes 1a, 3, or 4 [38][36][12]. In this study no correlation of age distribution with genotypes was observed. The relatively small number of patients is the limitation of this study. This highlights the need for further studies to determine possible risk factors in transmission of HCV and for effective implementation of preventive strategies to reduce exposure to this infection.

Phylogenetic analysis has been used to distinguish viral genotypes and/or subtypes from each other, and for subtyping newly isolated strains by comparing them to existing alignments and tree [18]. In our study, phylogenetic analysis of the isolates was performed and results were compared with NCBI genotyping tool. Phylogenetic branching of isolates and genotypes assigned by the NCBI tool correlated except for two isolates (figure 2A). Further, in figure 2B, it was observed that genotype 3k closely clustered with genotype 4. This could be attributed to the high level of sequence conservation in 5' untranslated region of the two genotypes. Furthermore, it was recorded that most of the genotype 1 subtypes branched with genotype 6 subtypes (Figure 2B). These results supported earlier study stating that phylogenetic analysis cannot accurately distinguish between subtypes of genotypes 1 and 6 [39]. This might be due to the low base differences per site between genotype 1 and 6 (Table V). Hence, Phylogenetic analysis can be performed to genotype HCV isolates. However for subtyping, NCBI tool is more reliable. Also compared with phylogenetic analysis, NCBI genotyping tool was simpler for genotyping and subtyping HCV isolates.

In conclusion, Genotypes 3 and 1 are prevalent in the studied districts of Maharashtra. Due to occurrence of genotype 4 and rare genotype 5, monitoring the circulating HCV genotypes and its subtypes in India on a regular basis becomes imperative for better clinical management. Further research is needed to determine the association of genotypes with gender and age in Indian population. Also, due to high similarity of 5'UTR between HCV genotypes and subtypes, genotyping by phylogenetic analysis of 5'UTR may result in misclassification of genotypes and subtypes. Subtyping by this method requires careful analysis of phylogenetic trees. However, amplification, sequencing and analysis of 5'UTR by online genotyping tools, like the NCBI Genotyping tool, allows almost instant identification of the genotype of a virus by comparison of its nucleotide sequence to the increasing number of reference sequences from previously characterized viruses throughout the world.

ACKNOWLEDGEMENT

We would like to thank Sudarshanaa Rajan and Vigneshwari Subramanian for their critical reading and comments on this manuscript, and Suresh Chander for editorial assistance.

REFERENCES

- Messina, Jane P., Isla Humphreys, Abraham Flaxman, Anthony Brown, Graham S. Cooke, Oliver G. Pybus, and Eleanor Barnes. "Global distribution and prevalence of hepatitis C virus genotypes." *Hepatology* 61.1 (2015): 77-87.
- Kar, Premashis. "Risk factors for hepatocellular carcinoma in India." *Journal of clinical and experimental hepatology* 4 (2014): S34-S42.
- Lindenbach, Brett D., and Charles M. Rice. "Unravelling hepatitis C virus replication from genome to function." *Nature* 436.7053 (2005): 933-938.
- Foy, Eileen, Kui Li, Chunfu Wang, Rhea Sumpter, Masanori Ikeda, Stanley M. Lemon, and Michael Gale. "Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease." *Science* 300.5622 (2003): 1145-1148.
- Drazan, K. "Molecular biology of hepatitis C infection." *Liver Transplantation* 6. 4 (2000): 396-406.
- Smith, Donald B., Jens Bukh, Carla Kuiken, A. Scott Muerhoff, Charles M. Rice, Jack T. Stapleton, and Peter Simmonds. "Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource." *Hepatology* 59.1 (2014): 318-327.
- Davis, Gary L. "Hepatitis C virus genotypes and quasispecies." *The American journal of medicine* 107.6 (1999): 21-26.
- El-Serag, Hashem B. "Epidemiology of viral hepatitis and hepatocellular carcinoma." *Gastroenterology* 142.6 (2012): 1264-1273.
- McOmish, F., P. L. Yap, B. C. Dow, E. A. Follett, C. Seed, A. J. Keller, T. J. Cobain, T. Krusius, E. Kolho, and R. Naukkarinen. "Geographical distribution of hepatitis C virus genotypes in blood donors: an international collaborative survey." *Journal of Clinical Microbiology* 32.4 (1994): 884-892.
- Al Naamani, Khalid, Siham Al Sinani, and Marc Deschènes. "Epidemiology and treatment of hepatitis C genotypes 5 and 6." *Canadian Journal of Gastroenterology* 27.1 (2013): 8.
- Aziz, Hafsa, Abida Raza, Shahnaz Murtaza, Yasir Waheed, Ali Khalid, Javaid Irfan, Zahoor Samra, and Muhammad Amin Athar. "Molecular epidemiology of hepatitis C virus genotypes in different

- geographical regions of Punjab Province in Pakistan and a phylogenetic analysis." *International Journal of Infectious Diseases* 17.4 (2013): 247-253.
12. Petruzzello, Arnolfo, Nicola Coppola, Anna Maria Diodato, Vincenzo Iervolino, Rosa Azzaro, Gaetano Di Costanzo, Catia Addolorata Di Macchia et al. "Age and gender distribution of hepatitis C virus genotypes in the metropolitan area of Naples." *Intervirology* 56. 3 (2013): 206-212.
 13. Zein, Nizar N. "Clinical significance of hepatitis C virus genotypes." *Clinical microbiology reviews* 13.2 (2000): 223-235.
 14. Arens, Max. "Methods for subtyping and molecular comparison of human viral genomes." *Clinical microbiology reviews* 12.4 (1999): 612-626.
 15. Weck, Karen. "Molecular methods of hepatitis C genotyping." *Expert Review of Molecular Diagnostics* 5.4 (2005): 507-520.
 16. Ohno, Tomoyoshi, and J. Y. Lau. "The "gold-standard," accuracy, and the current concepts: Hepatitis C virus genotype and viremia." *Hepatology* 24.5 (1996): 1312-1315.
 17. Halfon, P., P. Trimoulet, M. Bourliere, H. Khiri, V. De Ledinghen, P. Couzigou, J. M. Feryn et al. "Hepatitis C virus genotyping based on 5' noncoding sequence analysis (Trugene)." *Journal of clinical microbiology* 39. 5 (2001): 1771-1773.
 18. Rozanov, Mikhail, Uwe Plikat, Colombe Chappey, Andrey Kochergin, and Tatiana Tatusova. "A web-based genotyping resource for viral sequences." *Nucleic acids research* 32. 2 (2004): 654-659.
 19. Kuiken, Carla, Peter Hraber, James Thurmond, and Karina Yusim. "The hepatitis C sequence database in Los Alamos." *Nucleic acids research* 36. 1 (2008): 512-516.
 20. Tamura, Koichiro, Glen Stecher, Daniel Peterson, Alan Filipiski, and Sudhir Kumar. "MEGA6: molecular evolutionary genetics analysis version 6.0." *Molecular biology and evolution* 30.12 (2013): 2725-2729.
 21. Sneath, Peter HA, and Robert R. Sokal. "Numerical taxonomy." *Nature* 193.4818 (1962): 855-860.
 22. Nei, Masatoshi, and Sudhir Kumar. *Molecular evolution and phylogenetics*. Oxford University Press, 2000.
 23. Cai, Qingxian, Zhixin Zhao, Ying Liu, Xiaoqiong Shao, and Zhiliang Gao. "Comparison of three different HCV genotyping methods: core, NS5B sequence analysis and line probe assay." *International journal of molecular medicine* 31.2 (2013): 347-352.
 24. Chakravarti, Anita, Gaurav Dogra, Vikas Verma, and Amit Parkash Srivastava. "Distribution pattern of HCV genotypes & its association with viral load." *The Indian journal of medical research* 133.3 (2011): 326.
 25. Hissar, Syed S., Ankur Goyal, Manoj Kumar, Chandana Pandey, P. V. Suneetha, Ajit Sood, Vandana Midha, Puja Sakhuja, Veena Malhotra, and Shiv K. Sarin. "Hepatitis C virus genotype 3 predominates in North and Central India and is associated with significant histopathologic liver disease." *Journal of medical virology* 78.4 (2006): 452-458.
 26. Narahari, Shobha, Abida Juwle, Subhankar Basak, and Dhananjaya Saranath. "Prevalence and geographic distribution of Hepatitis C Virus genotypes in Indian patient cohort." *Infection, Genetics and Evolution* 9.4 (2009): 643-645.
 27. Choudhary, Manish Chandra, Vidhya Natarajan, Priyanka Pandey, Ekta Gupta, Shvetank Sharma, Rachana Tripathi, M. Shesheer Kumar, Syed N. Kazim, and Shiv K. Sarin. "Identification of Indian sub-continent as hotspot for HCV genotype 3a origin by Bayesian evolutionary reconstruction." *Infection, Genetics and Evolution* 28 (2014): 87-94.
 28. Raghuraman, Sukanya, Priya Abraham, Gopalan Sridharan, Hubert Darius Daniel, B. S. Ramakrishna, and R. V. Shaji. "HCV genotype 4—an emerging threat as a cause of chronic liver disease in Indian (south) patients." *Journal of clinical virology* 31.4 (2004): 253-258.
 29. Ramia, S., and J. Eid-Fares. "Distribution of hepatitis C virus genotypes in the Middle East." *International Journal of Infectious Diseases* 10.4 (2006): 272-277.
 30. Valliammai, T., S. P. Thyagarajan, Arie J. Zuckerman, and Tim J. Harrison. "Diversity of genotypes of hepatitis C virus in southern India." *Journal of general virology* 76. 3 (1995): 711-716.
 31. Raghuraman, Sukanya, and Priya Abraham. "HCV genotyping in India—quo vadis?." *The Indian journal of medical research* 119. 4 (2004): ix.
 32. Birks, J. S., Ian J. Seccombe, and C. A. Sinclair. "Labour Migration in the Arab Gulf States: Patterns, Trends and Prospects*." *International Migration* 26. 3 (1988): 267-286.
 33. Gededzha, Maemu P., Selokela G. Selabe, Jason T. Blackard, Thanda Kyaw, and M. Jeffrey Mphahlele. "Near full-length genome analysis of HCV genotype 5 strains from South Africa." *Infection, Genetics and Evolution* 21 (2014): 118-123.
 34. Das, B. R., Biduth Kundu, Rashmi Khandapkar, and Sumedha Sahni. "Geographical distribution of hepatitis C virus genotypes in India." *Indian journal of pathology & microbiology* 45.3 (2002): 323-328.
 35. Christdas, J., J. Sivakumar, J. David, H. Daniel, S. Raghuraman, and P. Abraham. "Genotypes of hepatitis C virus in the Indian sub-continent: A decade-long experience from a tertiary care hospital in South India." *Indian journal of medical microbiology* 31.4 (2013): 349.
 36. Cantaloube, Jean-François, Pierre Gallian, Houssam Attoui, Philippe Biagini, Philippe De Micco, and Xavier de Lamballerie. "Genotype distribution and molecular epidemiology of hepatitis C virus in blood donors from southeast France." *Journal of clinical microbiology* 43.8 (2005): 3624-3629.
 37. Khan, Nasar, Muhammad Akmal, Muhammad Hayat, Muhammad Umar, Atta Ullah, Iqbal Ahmed, Kashif Rahim, Sadar Ali, Sulaiman Bahadar, and Shamim Saleha. "Geographic distribution of hepatitis C virus genotypes in Pakistan." *Hepatitis monthly* 14.10 (2014).
 38. Altuglu, Imre, Ilknur Soyler, Tijen Ozacar, and Selda Erensoy. "Distribution of hepatitis C virus genotypes in patients with chronic hepatitis C infection in Western Turkey." *International Journal of Infectious Diseases* 12. 3 (2008): 239-244.
 39. Hraber, Peter T., William Fischer, William J. Bruno, Thomas Leitner, and Carla Kuiken. "Comparative analysis of hepatitis C virus phylogenies from coding and non-coding regions: the 5'untranslated region (UTR) fails to classify subtypes." *Virology Journal* 3.1 (2006): 103.

CITE THIS ARTICLE AS:

Ushanandini Mohanraj, Yashwant G Chavan, Sharad R Pawar, Amol D Raut. Identification of hepatitis C virus genotypes and subtypes in patients from Maharashtra based on 5'UTR sequencing and analysis: a retrospective study. *International Journal of Bioassays* 4.12 (2015): 4601-4607.

Source of support: Nil

Conflict of interest: None Declared