scientific reports

OPEN

Phylogenetic analysis of the 5′ untranslated region of HCV from cirrhotic patients in Khyber Pakhtunkhwa, Pakistan

Amin Ullah[1], Irshad Ur Rehman[1], Jamshaid Ahmad[1], Margaret Odenthal[2], Saad Ahmad[3], Tariq Nadeem[4]✉, Qurban Ali[5], Muhammad Rizwan[6], Muhammad Ajmal Khan[7], Said Hassan[8], Hina Ahsan[9] & Bashir Ahmad[1]✉

Hepatitis C virus (HCV), a small, single-stranded RNA virus with a 9.6 kb genome, is one of the most common causes of liver diseases. Sequencing of the 5′ untranslated region (UTR) is usually used for HCV genotyping, but it is less important in numerous subtypes due to its scarce sequence variations. This study aimed to identify genotypes using the 5′ UTR of HCV from cirrhotic patients of Khyber Pakhtunkhwa (KP). Serum RNA samples (44) were screened by real time PCR to determine the HCV viral load. Nested PCR was performed to identify cDNA and the 5′ UTR. The HCV 5′ UTR was sequenced using the Sanger method. MEGA-7 software was used to analyze evolutionary relatedness. After 5′ UTR sequencing, 26 samples (59%) were identified as genotype 3, and 2 samples (6%) were identified as genotypes 1, 2 and 4. The most predominant genotype was 3a, and genotype 4 was rarely reported in the phylogenetic tree. Analysis of the HCV 5′ UTR is an efficient alternative method for confirmation of various genotypes. Phylogenetic analysis showed that genotype 3 was dominant in the area of KP, Pakistan.

Liver diseases due to hepatitis C virus (HCV) pose serious health threats worldwide. HCV-induced hepatitis C has an alarming high frequency of progression to chronic liver disease (CLD), liver cirrhosis and carcinoma. Currently, more than 70 million people suffer from HCV-mediated CLD worldwide. Almost 400,000 people die due to cirrhosis and liver cancer, caused by HCV each year[1,2]. HCV is a lipoprotein-enveloped ribovirus with a 9.600 nucleotide 5′–3′ UTR[3]. The RNA genome has an untranslated region (UTR), three structural (core, E1, E2) and seven nonstructural genes (p7, NS2-NS5). The nonstructural protein (NS5B) is a moderately variable region and is commonly used for HCV subtyping[4]. In terms of primary sequence and secondary structures, the 5′ and 3′ UTRs are the most conserved areas of HCV RNA.

The 5′ UTR contains reasonably variable areas inclusive of NS5A, which codes for a nonstructural protein[5]. The 5′ UTR consists of 341 nucleotides, and due to its 90% sequence identity, it is commonly used for genotype identification. The 5′-UTR stem loop structure contains entry sites (IRES)[6]. Mutations do not usually occur in the 5-UTR, and sometimes compensatory mutations are developed to preserve the base-pairing shape and conserve the structural characteristics associated with translation efficiency. Recently, it has been found that the first 145 sequences of the 5′-UTR play a significant role in the replication of HCV RNA[7]. NS5A also has inadequate natural amino acid variability, which conserves its useful characteristics in vivo[8].

Other articles on sequencing substantiate the 5′ UTR (324–341 nucleotides long) as the least mutated region in the HCV genome and describe it remaining conserved in all HCV genotypes[9–11]. This high-grade conservation

[1]Department of Microbiology and Biotechnology, Abasyn University Peshawar, Khyber Pakhtunkhwa, Pakistan. [2]Institute of Pathology, Laboratory of Translational Molecular Pathology, University Hospital of Cologne, Cologne, Germany. [3]Kabir Medical College Peshawar, Gandhara University Peshawar, Khyber Pakhtunkhwa, Pakistan. [4]Centre of Excellence in Molecular Biology, University of the Punjab Lahore, Lahore, Pakistan. [5]Institute of Molecular Biology and Biotechnology, The University of Lahore, Lahore, Pakistan. [6]Department of Biotechnology and Microbiology, Swat University, Khyber Pakhtunkhwa, Pakistan. [7]Division of Life Sciences, Center for Cancer Research and State Key Lab for Molecular Neuroscience, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. [8]Institute of Biotechnology and Microbiology, Bacha Khan University Charsadda, Khyber Pakhtunkhwa, Pakistan. [9]Riphah Faculty of Pharmaceutical Sciences, G7/4 7th Avenue, Islamabad, Pakistan. ✉email: saim1692@gmail.com; bashir2015@yahoo.com

makes the 5′ UTR the region of choice for performing (RT)-PCR detection tests, such as the HCV amplicor test[12]. A number of genotyping schemes have been established and utilized in this region to obtain phylogenetic genotype information[13,14]. Sequencing data of the 5′ UTR and other regions, such as the NS-3, NS-4, core and NS-5, have been used in phylogenetic studies and genotyping of HCV[15–17].

Geographically, HCV genotype 1 is prevalent in Europe, Japan and the USA[18], whereas genotype 2 is found in Korea and Taiwan, and genotype 3 is detected in Pakistan, India and Thailand. Genotype 4 is the most frequent genotype in Saudi Arabia, Egypt, Syria, Iraq, Vietnam and Lebanon. Genotype 5 is found in South Africa, and genotype 6 is found in Vietnam[19–21]. The current study focused on the analysis of 5′ UTR sequencing and identification of genotypes by comparison with reference genotypes.

Materials and methods

RNA isolation and cDNA preparation. All samples were collected from HCV cirrhotic patients treated at the Hayatabad Medical Complex (HMC) and Khyber Teaching Hospital (KTH), Peshawar, during January 2017–May 2018. The current study was approved by the competent authorities of the institute, and the whole study was carried out according to the ethical guidelines given by the institute. RNA was isolated from serum by a QIAamp viral RNA kit in accordance with the manufacturer's protocol and stored at − 80 °C.
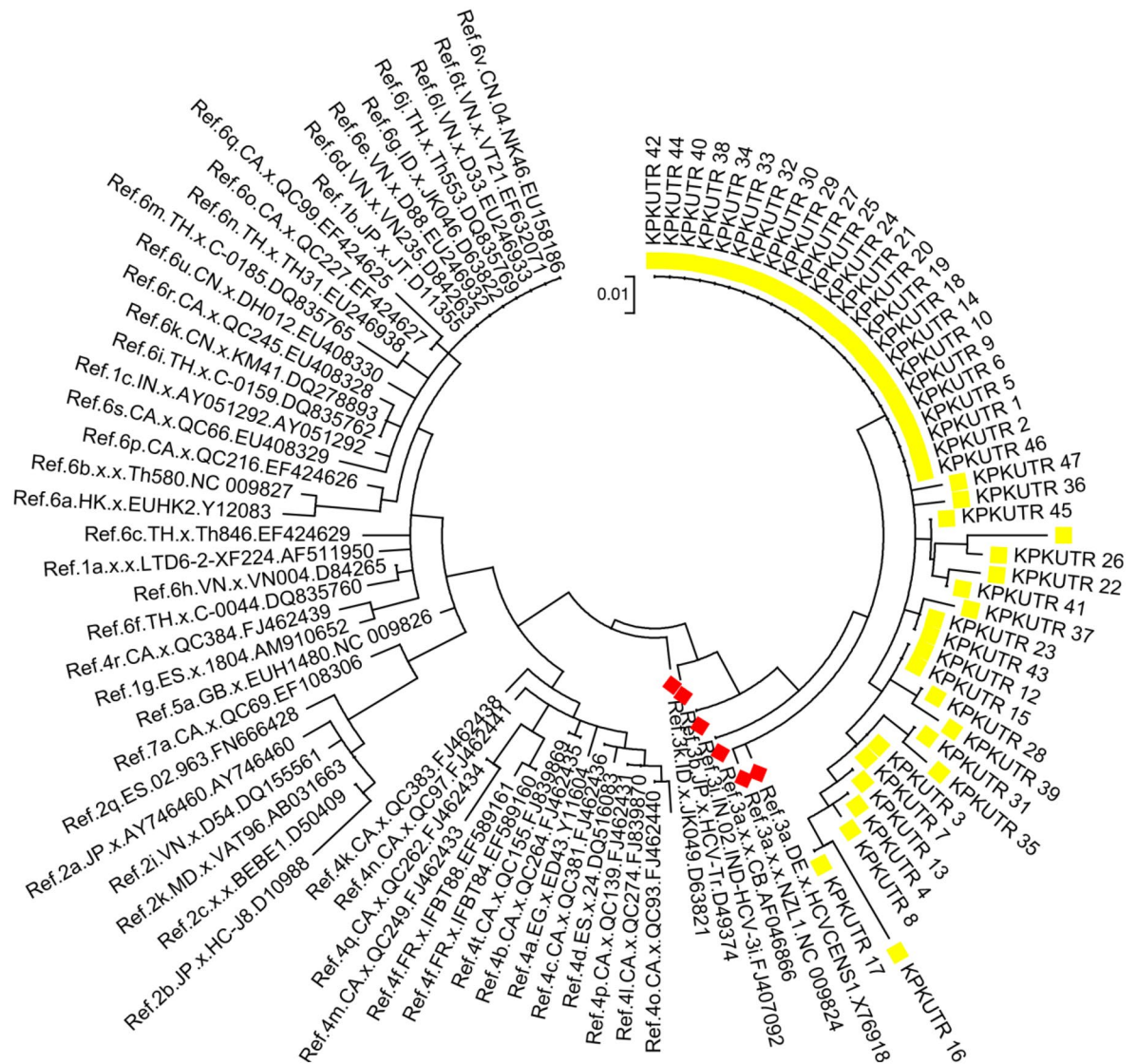
The cDNA was formed using 1 μl outer antisense primer (OAS), 10 μl RNA, 1 μl (200 U) M-MLV reverse transcriptase enzyme (BIORON Life Science cDNA Kit), 4 μl complete RT buffer, 2.5 μl PCR water, 1 μl dNTPs (10 mM) and 0.5 μl RNA inhibitors with a total volume of 20 μl. The following temperature cycle was then applied: 37 °C/60 min, 70 °C/10 s, and 22 °C/∞. The outer antisense primer was the reverse primer used in the first round of nested PCR amplification of the HCV 5′ UTR.

Name	Primer	Base pairs
OS	5′ctcttacgaggcgacactcc3′	20
OAS	5′caagcaccctatcaggcagt3′	20
IS	5′gatcactcccctgtgaggaa3′	20
IAS	5′ctttcgcgacccaacactac3′	20

Table 1. HCV 5′ UTR primers and sizes.

PCR amplification. The 4 μl cDNA was amplified in the first round of nested PCR with forward and reverse primers of the 5′ UTR for HCV positivity. Similarly, 4 μl of the 1st round PCR product was amplified in the 2nd round PCR, with inner sense (IS) and inner anti sense (IAS) primers (Table 1). In both rounds of PCR, 7.1 μl PCR grade water, 6.9 μl master mix and 1 μl each primer were used with the following cycle: 94 °C/2 min initial denaturation and 35 cycles of 94 °C/30 s, 54 °C/30 s, and 72 °C/45 s for annealing with a final extension at 72 °C/10 min in a thermal cycler.

DNA sequencing. PCR products (5′ UTR) were purified and sequenced through a standard Sanger procedure on an ABI 3730XL DNA Sequencer at Macrogen sequencing services (South Korea). Chromas software (http://technelysium.com.au/wp/chromas/) was used to check and correct the sequence viewer. Nested PCR primers were used for sequencing (Table 1), and then sequences were submitted to BLAST to determine similarity to sample sequences in the reported databases. Sequences were submitted to GenBank, under the accession numbers MN038290-MN038122. Reference sequences of HCV genotypes were retrieved from the GenBank database to build a phylogenetic tree.

Phylogenetic tree. Molecular Evolutionary Genetics Analysis software (MEGA version 7: http://www.megasoftware.net) was used to sequence the 5′ UTR, sequences were aligned by the maximum likelihood method and associated with the reference sequence of the identified genotype. The neighbor-joining algorithm of MEGA 7 was used to calculate the p-distance and the differences in each nucleotide sequence. The phylogenetic tree analysis utilized 1000 bootstrap resampling to test the robustness of the observed dominant clades.

Statistical analysis. Statistical analysis was performed by IBM SPSS version 25, Windows 7 and Microsoft Excel version 13.

Ethics approval. This study was approved by the Ethics Committee, Centre of Biotechnology and Microbiology, University of Peshawar, Pakistan. Written informed consent was obtained from all the individuals who participated in the study.

Results

HCV cirrhotic samples (n = 44) were collected on the basis of their demographic variables. The average age and standard deviation (SD) of the patients was 48.69 ± 11.28 yrs. The numbers of male and female patients were n = 25 and 19, respectively. The highest viral load was recorded prior to and during the treatment > 200,0000 IU/mL, as shown in Table 2. A high number of HCV cirrhotic patients were identified or noted between the ages of 31–50 (Fig. 1).

Out of all the patients (n = 44), sequencing was successful for samples from 32 patients, and the remaining 12 (8 mild, 2 moderate and 2 gross) patients had sequences that showed less similarity during BLAST because of their small size/nucleotide bp and hence were excluded. The sequence alignment of the 5′-UTR isolates (entitled Hepacivirus C isolate AIMZ7 KP 5′-UTR) was performed with reference genotype (1–7) sequences from the database. The well-conserved areas and few nucleotide substitutions in the 5′-UTR of the HCV genome are shown in Fig. 2. The data also revealed that the length of the 5′-UTR of HCV was up to 183 nucleotides. In the phylogenetic tree, most of the isolates were in a clad and clustered perfectly with the reference sequence. The two isolates from the group did not cluster with the reference sequence. The phylogenetic tree indicated that most of the isolates clustered with subtypes 3a and 3b, and few of them clustered with other reference genotype sequences (Fig. 3).

Sequencing of the HCV 5′-UTR was performed with a forward primer (red color), and the nucleotide similarity of the 5′-UTR sequence was checked by using multiple alignment fast Fourier transform methods (MAFFT version 7 software: https://mafft.cbrc.jp/alignment/server/large.html). All the sequences show significant similarity with sample and sequence no. 22, and sequence 28 shows significant similarity with sequence no. 1, as shown in Fig. 4.

Discussion

Hepatitis C virus (HCV) has a broad range of genotypes and quasispecies due to frequent genetic mutations in its RNA. The identification of HCV genotypes through direct sequencing of the 5′-UTR (Fig. 5) is a great technique because it does not require pattern processing steps and uses amplified products obtained from one-step, nonnested PCR. Additionally, direct sequencing of PCR products provides more comprehensive sequence information than different genotyping analyses. Different studies have reported that the HCV 5′-UTR of 324 to

Variable	No. of patients
Gender	
Male	25 (57%)
Female	19 (43%)
Age	
Mean ± SD	49.8 ± 12.33
Range	23–81
Cirrhosis stage	
Mild	27 (61%)
Moderate	10 (23%)
Gross	07 (16%)
Treatment	
Antiviral drugs	Sofusbuvir + ribavirin/daclatasvir
ALT	
Mean ± SD	105.87 ± 30.77
Range	87–283
AFP	
Mean ± SD	48.3 ± 39.80
Range	34–279
Albumin	
Mean ± SD	3.7 ± 0.4
Bilirubin	
Mean ± SD	1.4 ± 0.8
AST	
Mean ± SD	41.3 ± 29.60
Range	67- 209
Viral load	
< 200,0000 IU/mL	30
> 200,0000 IU/mL	14

Table 2. Baseline characteristics of HCV cirrhotic patients (n = 44). *SD* standard deviation, *ALT* alanin transaminase, *AFP* alpha fetoprotein, *AST* asparagine transaminase.
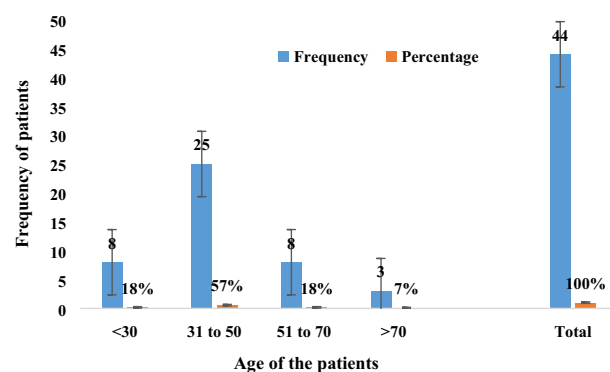
Figure 1. Frequency distribution of HCV cirrhotic patients based on age.

341 nucleotide sequences is the most conserved region[22] and can be used for the sequence-based identification of HCV genotypes[17–23].

The purpose of the current research work was to identify different HCV genotypes from the 5′ UTR isolated from patients of the Khyber Pakhtunkhwa (KP) region of Pakistan. The results showed that different genotypes were present in the phylogenetic tree, and genotype 3 was noted as the most predominant. Genotype 3 (a, b) was reported in 26 (59%) isolates of the phylogenetic tree and coincided with the position of each genotype 1,

Figure 2. Alignment of the HCV 5′ UTR sequence with the reference sequence HCV.

2, and 4. Sequence conservation in the 5′-UTR HCV genome is vital for the detection of genotypes by sequence evaluation from samples that are not typed with more generally used assays. In the geographic area of Pakistan, HCV genotypes have varied distributions. Genotype 3 was detected more than the other genotypes in this area, as clearly stated in the study of Idress and Riazudin[24].

Another finding of the study was that genotypes 1 and 2 are also found in this area. Genotypes 1 and 2 are mostly found in Europe and East Asian countries. The detection of these genotypes in KP revealed the immigration of people from one state to another[25]. The rise of genotypes 1 and 2 also causes major problems during treatment, as these patients show less response to antiviral drugs[26]. The occurrence of genotype 2 was also previously reported but was very rare in the local province and neighboring states of Pakistan[24].

One of the main important findings of our results was that genotype 4 was found very rarely in our state. Recently, HCV genotype 4 has spread in parts of Europe due to variations in the populace structure, ways of transmission and migration. The characteristics of genotype 4 infection and proper therapeutic programs are not well described[27]. Genotype 4 is mostly found in Northern Africa and the Middle East (Egypt, Iran). As these are neighboring countries of Pakistan, immigration could be a factor[28–30]. In general, in this study, the number of samples was small, and available funding was limited. Further study with a larger number of samples is necessary to completely identify the genotypes of this region.

Conclusion

The study concludes that genotype 3a is the predominant genotype of HCV in the Khyber Pakhtunkhwa region, Pakistan. Direct sequencing of the 5′ UTR is a valuable method for clinical detection of different HCV genotypes. Genotype 4 was rarely detected in the tribal area of Khyber Pakhtunkhwa. I am very thankful to the gastroenterology laboratories of tertiary hospitals of Peshawar and the diagnostic laboratory of the Centre of Biotechnology and Microbiology University of Peshawar, Pakistan. I am highly thankful to the Institute of Pathology, in particular to the Laboratory of Translational Molecular Pathology (University Hospital of Cologne Germany) and Prof. Dr. Margarete Odenthal for the helpful discussions and in proofreading of the manuscript.
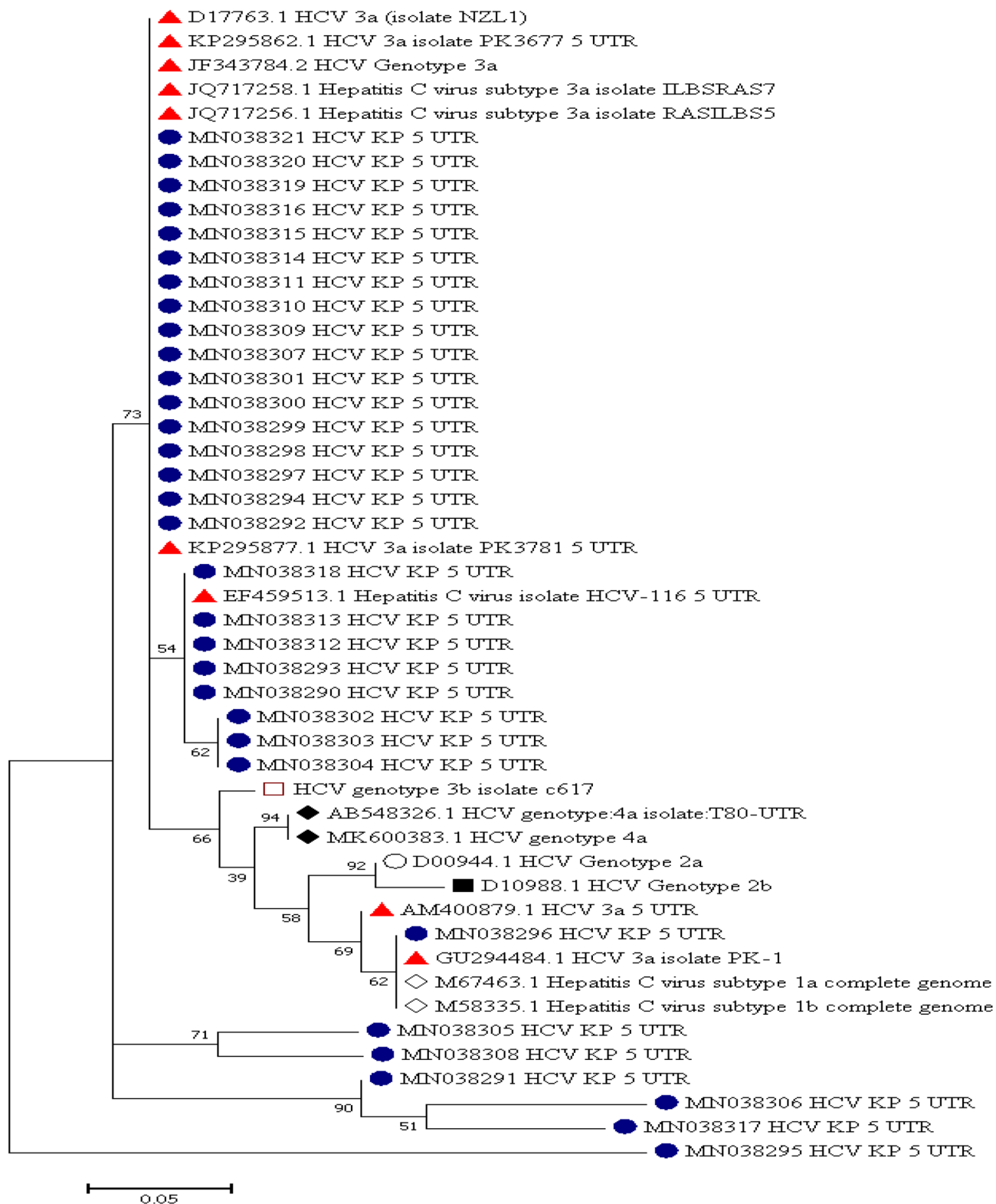
Figure 3. The evolutionary history was inferred via the maximum likelihood method, and the tree with the highest log likelihood (-421.51) is displayed. The percentage of trees in which the linked taxa clustered together is presented next to the branches. All positions with absent data and gaps were removed. There were a total of 84 positions in the final dataset. Evolutionary analyses were performed by MEGA7 and HCV 5′ UTR sequences with reported genotypes. The highlighted samples show different genotypes, and genotype 3 (highlighted in red) was predominant and matched with the reference genotypes.
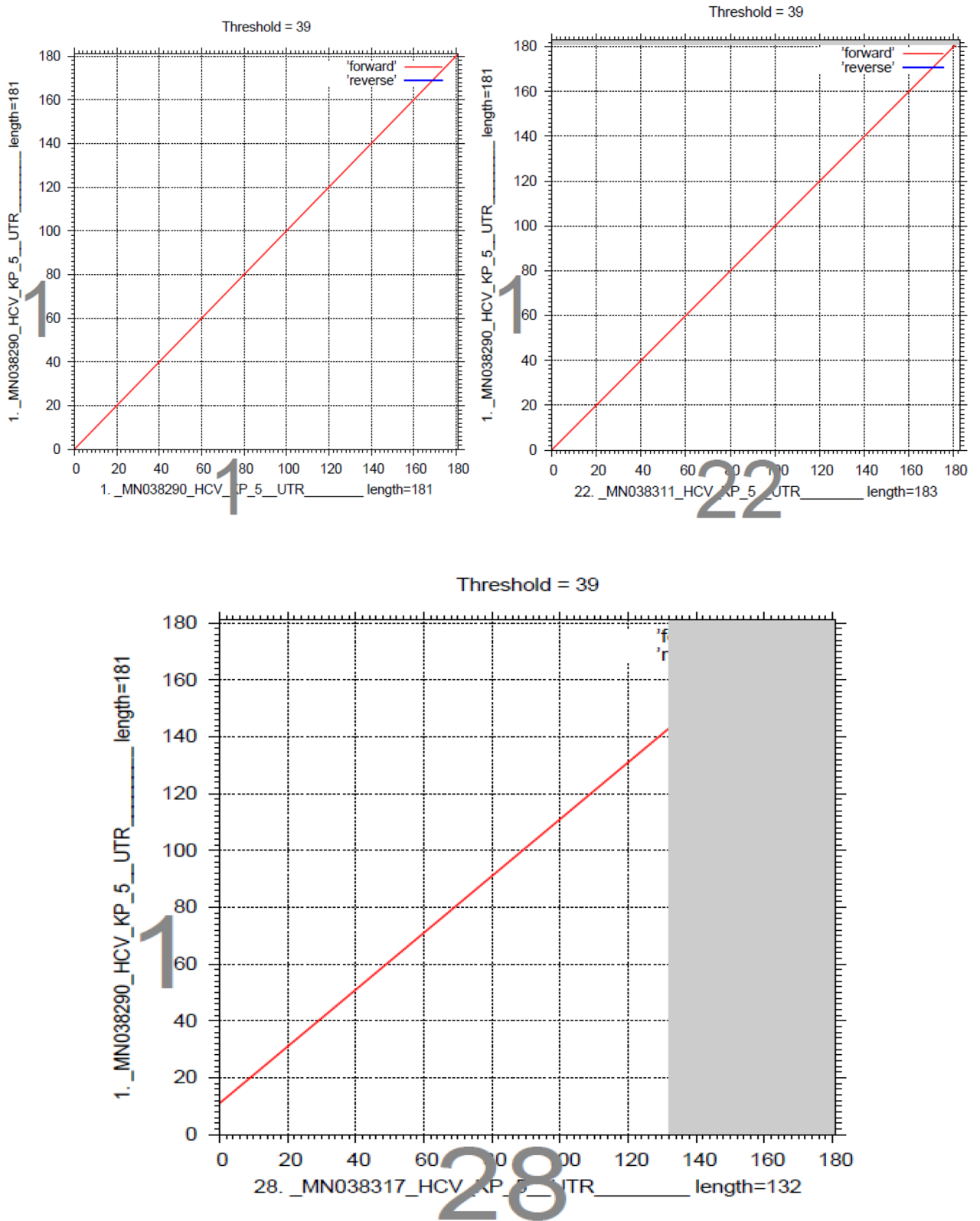
Figure 4. Sequencing of the HCV 5′ UTR was performed with a forward primer (red color), and the nucleotide similarity of the 5′ UTR sequence was checked by using multiple alignment fast Fourier transform (MAFFT version 7 software: https://mafft.cbrc.jp/alignment/server/large.html). All the sequences show significant similarity with sample 1, and sequence no. 22 shows significant similarity with sequences no. 1 and 28.
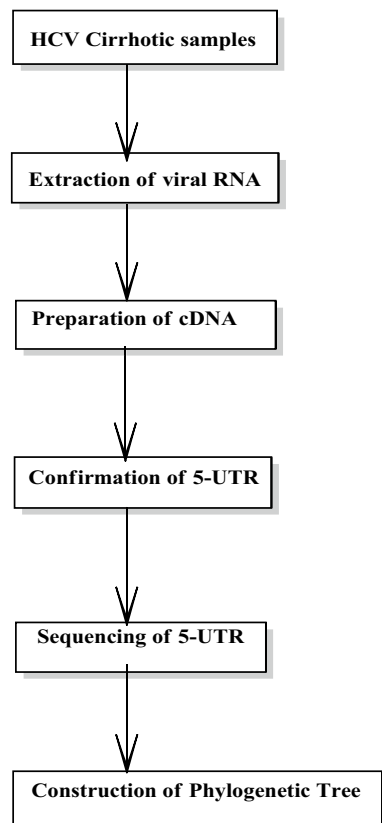
Figure 5. Flow chart of the phylogenetic tree analysis of HCV cirrhotic patients.

References

1. World Health Organization. Hepatitis C (2018). https://www.who.int/news-room/fact-sheets/detail/hepatitis-c.
2. Shepard, C. W., Finelli, L. & Alter, M. J. Global epidemiology of hepatitis C virus infection. *Lancet Infect. Dis.* **5**, 558–567. https://doi.org/10.1016/S1473-3099(05)70216-4 (2005).
3. Alzahrani, A. J. *et al.* Detection of hepatitis C virus and human immunodeficiency virus in expatriates in Saudi Arabia by antigen-antibody combination assays. *J. Infect. Dev. Countr.* **3**, 235–238. https://doi.org/10.3855/jidc.42 (2009).
4. Fan, W. *et al.* Nonstructural 5A gene variability of hepatitis C virus (HCV) during a 10-year follow up. *J. Gastroenterol.* **40**, 43–51. https://doi.org/10.1007/s00535-004-1446-2 (2005).
5. Margraf, R. L. *et al.* Genotyping hepatitis C virus by heteroduplex mobility analysis using temperature gradient capillary electrophoresis. *J. Clin. Microbiol.* **42**, 4545–4551. https://doi.org/10.1128/JCM.42.10.4545-4551.2004 (2004).
6. Yuan, H. J. *et al.* Evolution of hepatitis C virus NS5A region in breakthrough patients during pegylated interferon and ribavirin therapy. *J. Viral. Hepatitis* **17**, 208–216. https://doi.org/10.1111/j.1365-2893.2009.01169.x (2009).
7. Soler, M. *et al.* Quasispecies heterogeneity and constraints on the evolution of the 5′ noncoding region of hepatitis C virus (HCV): Relationship with HCV resistance to interferon-alpha therapy. *Virology* **298**, 160–173. https://doi.org/10.1006/viro.2002.1494 (2002).
8. Pawlotsky, J. M. Hepatitis C virus population dynamics during infection. *Concept Implic. Virol.* **299**, 261–284. https://doi.org/10.1007/3-540-26397-7_9 (2006).
9. Idrees, M. *et al.* Nucleotide identity and variability among different Pakistani hepatitis C virus isolates. *Virol. J.* **6**(1), 130. https://doi.org/10.1186/1743-422X-6-130 (2009).
10. Han, J. H. *et al.* Characterization of the terminal regions of hepatitis C viral RNA: Identification of conserved sequences in the 5′ untranslated region and poly(A) tails at the 3′ end. *Proc. Natl. Acad. Sci.* **88**, 1711–1715. https://doi.org/10.1073/pnas.88.5.1711 (1991).
11. Beales, L. P. *et al.* The internal ribosome entry site (IRES) of hepatitis C virus visualized by electron microscopy. *RNA* **7**, 661–670. https://doi.org/10.1017/s1355838201001406 (2001).
12. Young, K. K. *et al.* Detection of hepatitis C virus RNA by a combined reverse transcription-polymerase chain reaction assay. *J. Clin. Microbiol.* **31**, 882–886. https://doi.org/10.1128/jcm.31.4.882-886.1993 (1993).
13. Kleter, G. E. *et al.* Sequence analysis of the 59 untranslated region in isolates of at least four genotypes of hepatitis C virus in The Netherlands. *J. Clin. Microbiol.* **32**, 306–310. https://doi.org/10.1128/JCM.32.2.306-310.1994 (1994).
14. Casanova, Y. S. *et al.* A complete molecular biology assay for hepatitis C virus detection, quantification and genotyping. *Rev. Da. Soc. Bras. Med. Trop.* **47**, 287–294. https://doi.org/10.1590/0037-8682-0040-2014 (2014).
15. Shier, M. K. *et al.* Molecular characterization and epidemic history of hepatitis C virus using core sequences of isolates from Central Province, Saudi Arabia. *PLoS One* **12**(9), e0184163. https://doi.org/10.1371/journal.pone.0184163 (2017).
16. Germer, J. J. *et al.* Determination of hepatitis C virus genotype by direct sequence analysis of products generated with the Amplicor HCV test. *J. Clin. Microbiol.* **37**(8), 2625–2630. https://doi.org/10.1128/JCM.37.8.2625-2630.1999 (1999).

17. Cantaloube, J. F. *et al.* Analysis of the 5′ noncoding region versus the NS5b region in genotyping hepatitis C virus isolates from blood donors in France. *J. Clin. Microbiol.* **44**, 2051–2056. https://doi.org/10.1128/JCM.02463-05 (2006).
18. Manos, M. M. *et al.* Distribution of hepatitis C virus genotypes in a diverse US integrated health care population. *J. Med. Virol.* **84**(11), 1744–1750. https://doi.org/10.1002/jmv.23399 (2012).
19. Qattan, I. & Emery, V. HCV genotyping in chronic hepatitis C patients. *Euro. Sci. J.* **8**(27), 1–2. https://doi.org/10.19044/esj.2012.v8n27p%25p (2012).
20. Sievert, W. *et al.* A systematic review of hepatitis C virus epidemiology in Asia, Australia and Egypt. *Liver Int.* **31**(2), 61–80. https://doi.org/10.1111/j.1478-3231.2011.02540.x (2011).
21. Ahmadi, P. M. H. *et al.* Determination of HCV genotypes, in Iran by PCR-RFLP. *Iran. J. Public Health.* **35**(4), 54–61 (2006).
22. Anjum, S. *et al.* Sequence and structural analysis of 3′ untranslated region of Hepatitis C Virus, genotype 3a, from Pakistani isolates. *Hepat. Mon.* **13**(5), e8390. https://doi.org/10.5812/hepatmon.8390 (2013).
23. Shier, M. K. *et al.* Characterization of hepatitis C virus genotypes by direct sequencing of HCV 5′UTR region of isolates from Saudi Arabia. *PLoS One* **9**(8), e103160. https://doi.org/10.1371/journal.pone.0103160 (2014).
24. Idrees, M. & Riazuddin, S. Frequency distribution of hepatitis C virus genotypes in different geographical regions of Pakistan and their possible routes of transmission. *BMC Infect. Dis.* **8**, 69. https://doi.org/10.1186/1471-2334-8-69 (2008).
25. Sanders-Buell, E. *et al.* Hepatitis C genotype distribution and homology among geographically disparate injecting drug users in Afghanistan. *J. Med. Virol.* **85**(7), 1170–1179 (2013).
26. Manns, M. *et al.* Ledipasvir and sofosbuvir plus ribavirin in patients with genotype 1 or 4 hepatitis C virus infection and advanced liver disease: A multicentre, open-label, randomised, phase 2 trial. *Lancet Infect. Dis.* **16**(6), 685–697. https://doi.org/10.1016/S1473-3099(16)00052-9 (2016).
27. Teimoori, A. *et al.* Prevalence and genetic diversity of HCV among HIV-1 infected individuals living in Ahvaz, Iran. *BMC Infect. Dis.* **19**(1), 389 (2019).
28. El-Tahan, R. R., Ghoneim, A. M. & Zaghloul, H. 5′ UTR and NS5B-based genotyping of hepatitis C virus in patients from Damietta governorate, Egypt. *J. Adv. Res.* **10**, 39–47. https://doi.org/10.1016/j.jare.2018.01.004 (2018).
29. Samimi-Rad, K. *et al.* Molecular epidemiology of hepatitis C virus in Iran as reflected by phylogenetic analysis of the NS5B region. *J. Med. Virol.* **74**(2), 246–252. https://doi.org/10.1002/jmv.20170 (2004).
30. Mushtaq, U. *et al.* Role of modern technology for treatment of HCV. *Biol. Clin. Sci. Res. J.* **2020**, e001 (2020).

Author contributions

A.U. conducted research, A.U. and I.U.R. wrote up the initial draft of the manuscript. M.O. gave the concept of research work. J.A., S.A., and T.N. made figures and carried out statistical analysis. Q.A. and B.A. make final corrections in the manuscript. M.R., M.A.K., S.H., and H.A. help in the final version to improve readability. All authors reviewed and approved manuscripts for publication.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.N. or B.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.