



Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Original article

Virological surveillance, molecular phylogeny, and evolutionary dynamics of hepatitis C virus subtypes 1a and 4a isolates in patients from Saudi Arabia

Waleed H. AlMalki ^{a,1}, Imran Shahid ^{a,b,1,*}, Ashraf N. Abdalla ^a, Ayman K. Johargy ^c, Muhammad Ahmed ^a, Sajida Hassan ^d^a Department of Pharmacology and Toxicology, College of Pharmacy, Umm-Al-Qura University, Al-abidiyah, P.O. Box 13578, Makkah 21955, Saudi Arabia^b Department of Pharmacology and Toxicology, Faculty of Medicine, Umm-Al-Qura University, Al-abidiyah, P.O. Box 13578, Makkah 21955, Saudi Arabia^c Medical Microbiology Department, Faculty of Medicine, Umm-Al-Qura University, Al-abidiyah, P.O. Box 13578, Makkah 21955, Saudi Arabia^d Viral Hepatitis Program, Laboratory of Medicine, University of Washington, Seattle, WA, USA

ARTICLE INFO

Article history:

Received 28 September 2020

Revised 28 November 2020

Accepted 30 November 2020

Available online 8 December 2020

Keywords:

HCV
Sequence variability
5' untranslated region
Subtyping
Evolution
Molecular Phylogeny

ABSTRACT

Hepatitis C virus (HCV) subtypes are pre-requisite to predict endemicity, epidemiology, clinical pathogenesis, diagnosis, and treatment of chronic hepatitis C infection. HCV genotypes 4 and 1 are the most prevalent in Saudi Arabia, however; less consensus data exist on circulating HCV subtypes in infected individuals. This study was aimed to demonstrate the virological surveillance, phylogenetic analysis, and evolutionary relationship of HCV genotypes 4 and 1 subtypes in the Saudi population with the rest of the world. Fifty-five clinical specimens from different parts of the country were analyzed based on 5' untranslated region (5' UTR) amplification, direct sequencing, and for molecular evolutionary genetic analysis. Pair-wise comparison and multiple sequence alignment were performed to determine the nucleotide conservation, nucleotide variation, and positional mutations within the sequenced isolates. The evolutionary relationship of sequenced HCV isolates with referenced HCV strains from the rest of the world was established by computing pairwise genetic distances and generating phylogenetic trees. Twelve new sequences were submitted to GenBank, NCBI database. The results revealed that HCV subtype 4a is more prevalent preceded by 1a in the Saudi population. Molecular phylogeny predicts the descendants' relationship of subtype 4a isolates very close to Egyptian prototype HCV strains, while 1a isolates were homogeneous and clustering to the European and North American genetic lineages. The implications of this study highlight the importance of HCV subtyping as an indispensable tool to monitor the distribution of viral strains, to determine the risk factors of infection prevalence, and to investigate clinical differences of treatment outcomes among intergenotypic and intragenotypic isolates in the treated population.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction:

Chronic hepatitis C (CHC) infection is still challenging in the world where almost 71 million people are infected with the virus

* Corresponding author.

E-mail address: iyshahid@uqu.edu.sa (I. Shahid).

¹ These authors contributed equally to this work

Peer review under responsibility of King Saud University.



and around 400 thousand deaths are reported annually due to CHC associated hepatic comorbidities (i.e. hepatic cirrhosis and hepatocellular carcinoma) (Dietz et al., 2018). HCV belongs to a separate genus of *Hepacivirus* within the family *Flaviviridae* and possesses a single-stranded positive-sense RNA genome of approximately 9500 nucleotides in length (Mann et al., 2017). The viral genome flanks by 5' and 3' untranslated regions (UTRs) that abut a single open reading frame of 3010 to 3037 amino acids polyprotein (Lohmann et al., 1999). The single polyprotein is translated into an initial structural region encoding three proteins including core (C), E1, and E2, and a nonstructural region that is posttranslationally processed into 4 nonstructural polypeptides (NS2 to NS5) by a cap-independent internal ribosome entry site (IRES)

<https://doi.org/10.1016/j.sjbs.2020.11.089>

1319-562X/© 2020 The Author(s). Published by Elsevier B.V. on behalf of King Saud University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

mechanism mediated by an IRES within the 5' UTR of the virus (Khaliq et al., 2011). HCV circulates as a population of closely related but diverse nucleotide sequences in infected individuals (referred to as 'quasispecies') due to poor fidelity, error-prone nature, and a lack of repair mechanism of viral RNA-dependent RNA polymerase (Penin et al., 2004). Furthermore, rapid replication rate accumulates mutations to viral isolates by which different strains display significant nucleotide variability in different genome regions (Manns et al., 2017). The envelope glycoproteins (i.e., E1 and E2) and some nonstructural proteins (e.g. NS3 and NS5A) are significantly variable, whereas 5' UTR, Core, and NS5B regions are highly conserved (Shahid et al., 2013).

Nucleotide sequence studies of complete or partial sequences of HCV isolates from different parts of the world identifies 8 genotypes (GTs) (GT 1, 2 ...8, etc.) and multiple subtypes (Smith et al., 2014; Simmonds et al., 2017; Hedskog et al., 2019). The eighth GT is recently reported in India (Borgia et al., 2018), whereas the remaining 7 GTs comprise 67 multiple subtypes (1a, 1b, 2a, 3a, 4a, etc.) (Smith et al., 2014). Recently, a study explored expanded HCV subtypes classification by identifying 19 novel HCV subtypes (Hedskog et al., 2019). HCV GT endemicity is sometimes based on the diversity and multiplicity of circulating subtypes in different geographical regions where it is used to trace out the evolutionary origin of that GT (Shier et al., 2014). The nucleotide variations at the HCV GTs level accounts for 30%, while subtypes differ by 20 to 23%, and interestingly 5–15% variability has been reported in distinct isolates from the same subtype (Messina et al., 2015). HCV GTs have different geographical prevalence where GT 1 is the most abundant GT following GT 3 in the world (Messina et al., 2015). GT 1 represents 46% of all HCV infections and common in North America, South America, Western and Northern Europe (Gower et al., 2014). GT 2 is circulating in Japan, some parts of Europe, and North America (Gower et al., 2014). GT 3 is the second most abundant genotype accounting for 30% HCV infection worldwide and widely spread in South Asia, Australia, and some parts of Western Europe (Messina et al., 2015). GT 4 is almost endemic in Egypt where seroprevalence was recorded 10% in 2015 (El-Tahan et al., 2018). HCV GT 4 infected populations also exist in some parts of the Central and North Africa and the Middle East (e.g. Saudi Arabia) (El-Tahan et al., 2018). GT 5, 6, and 7 are commonly reported in South Korea, Southeast Asia (e.g., Hong Kong), South Africa, and Congo respectively (Shier et al., 2014). HCV treatment outcome is directly correlated with GTs testing in infected individuals before the start of the treatment. Some studies report that HCV GT-3 is not susceptible to the first generation IFN-free protease inhibitors DAAs (i.e. HCV non-structural proteins NS3/4A serine protease inhibitors), and with limited response to sofosbuvir (i.e. HCV non-structural protein NS5B; RdRp inhibitors) (Petruzzello et al., 2016; Aljowaie et al., 2020). Furthermore, GT-3 infected populations are associated with an increased risk of accelerated liver disease progression (McPhee, 2019). For such difficult-to-treat patients, next-generation DAAs (e.g., sofosbuvir/velpatasvir/voxilaprevir; glecaprevir/pibrentasvir) with promising clinical trial efficacies and an excellent treatment outcome in real-world clinical experiences are recommended (McPhee, 2019). For HCV GT-4 in Egypt, Sustained virologic response rates (SVR; HCV viral load undetectable after the post completion of 12 weeks therapy) were achieved only 40–60% in patients administered to pegylated interferon plus ribavirin (PEG-IFN/RBV) (Aljowaie et al., 2020). Furthermore, some studies also explored the increased risk of hepatocellular carcinoma directly associated with subtype 4a in infected individuals (Aljowaie et al., 2020). However; HCV rapid-diagnostic testing campaign is underway in Egypt and HCV positive patients are administered to newer IFN-free DAAs to cure the infection.

HCV GTs and subtypes identification are useful for epidemiological investigation of the infection progression and to choose appropriate therapeutic regimens for infected people (Shier et al., 2017). The gold standard method of HCV GT detection by nucleotide sequencing of the coding region (i.e. NS5B, Core/E1) is relatively expensive and time-consuming than the commercially available direct hybridization and probe assays directed against the 5' UTR of HCV (Hara et al., 2013). However; errors have been reported in approximately 10% cases where insufficient sequence variations resulted in subgenotypes recognition errors (1a from 1b and 2a vs. 2c) while analyzing variable lengths and regions of 5' UTR (Baclig et al., 2010). It envisages that either the region is either too conserved or not heterogeneous enough or certain sequence motifs are no longer conserved. It requires elucidation and removing those discrepancies in 5'UTR nucleotide sequences and necessitates improving the existing methods of HCV GTs/subtypes identification (El-Tahan et al., 2018; Baclig et al., 2010). Typing of HCV isolates is considered a key marker of the likelihood of a response to therapy and a guideline for the duration of therapy in clinical settings (McPhee, 2019). However; due to the real-world clinical success of pangenotypic next-generation DAAs with very high SVR rates in different HCV clinical conditions, pre-genotyping may no longer be required in the future (Fourati et al., 2018). Despite it, the skipping of genotyping/subtyping is questionable and seems controversial in low-to-middle income countries (LMICs) where the first-generation DAAs (e.g., sofosbuvir/daclatasvir or generics) are administered to treat HCV, dose algorithms are prolonged, and communities are suffering from GT-3 infection (Fourati et al., 2018).

This study elucidates the virological surveillance and phylogenetic analysis of the most prevalent HCV GT 4 and 1 subtypes circulating in the Saudi population based on 5' UTR sequence analysis. The previously reported studies from Saudi Arabia only emphasize to identify HCV GTs/subtypes in infected individuals based on HCV serotype assays and 5' UTR sequence alignment with existing reference sequences of different GTs and multiple subtypes (Shier et al., 2014; Abdel-Moneim et al., 2012; Al-Faleh, 2003; al Nasser, 1992; Akbar et al., 2012). We also validated the robustness of a recently reported one-step PCR amplification method for all HCV GTs/subtypes identification (Virtanen et al., 2018). The protocol is more reliable, robust, and reproducible without requiring costly instrumentation and specialized sequence analysis skills (Virtanen et al., 2018). Molecular phylogeny reveals the evolutionary relationship of subtype 1a isolates closer to North America and Western Europe HCV strains, while 4a isolated sequences are completely homologs and ancestrally closest with Egyptian prototype HCV strains. The study findings will also help to know the clinical relevance of HCV typing based on genetic heterogeneity of 5' UTR region directly or indirectly associated with the therapeutic outcome of PEG-IFN/RBV and pan-genotypic direct-acting antivirals (DAAs) against harder-to-treat HCV populations of subtypes 1a, 3a, 4a and 4d.

2. Materials and methods

2.1. Patient ethics and consent statement

The patients' demographic data, blood, and plasma samples included in this study were documented and provided by the Department of Pathology and Laboratory Medicine, Molecular Biology Unit, Ministry of National Guard health affairs, King Abdul Aziz Medical City, Jeddah, Saudi Arabia from March 2018 to December 2018. All patients gave their informed consent for inclusion before they participated in the study and for the collection of blood samples. The research project, data forms, and ethical consent were

approved by the King Abdullah City of Science and Technology (KACST), Riyadh, Saudi Arabia, King Abdul Aziz Medical City, Jeddah, Saudi Arabia, and the research ethics committee of the College of Pharmacy, Umm- Al-Qura University, Makkah, Saudi Arabia (REC/2479–19/CP/UQU-SA) respectively and were in total compliance with the Helsinki Declaration of 1975 as revised in 2008.

2.2. Clinical specimen and sample collection

Fifty-five plasma samples both from male and female patients excluding children were collected and stored at -70°C before use. The estimated duration of infection varied from 6 months to 10 years. The patients under 18 and above 70 years of age, patients with HCV/HBV or HCV/HIV co-infection, and pregnant females were excluded from the study. No patients were recommended to administer first or second-generation (e.g., ledipasvir, paritaprevir/ombitasvir/ritonavir) or pan-genotypic DAAs to treat the infection at the time of specimen collection. HCV positive criteria was based on elevated serum SGPT (serum glutamic pyruvic transaminase) and SGOT (serum glutamic oxaloacetic transaminase) levels at least for six months, histological examination, and persistent detection of serum HCV RNA in participating subjects. Anti-HCV antibodies performed by 3rd generation ELISA, (DIA-source Immunoassays[®], Nivelles, Belgium) were present in all samples. All patients were negative for HAV, HBV, and HDV surface antigens. HCV positive patient's demographic history with various clinical parameters was demonstrated in [Table 1](#).

2.3. HCV viral load and GTs/subtypes identification

HCV viral load in sera samples was detected by using Real-TM Quant SC kit (Sacace[™] Biotechnologies, Como, Italy) and fluorescent reporter dye probes specific to the Real-Time PCR SmartCycler[®] (Cepheid, Sunnyvale, USA) following kit protocol and manufacturer's instructions. HCV viral titers in the range from 3×10^5 to 5×10^6 IU/mL were considered in this study. For HCV subtypes identification of all isolates, we followed a recently reported diagnostic method of HCV genotyping based on a one-step PCR amplification method of 5' UTR and partial core region instead of the conventional methods of direct hybridization and probe assays directed against the 5' UTR ([Virtanen et al., 2018](#)). This protocol is consistent and advantageous to identify all HCV GTs/subtypes in one thermal cycle reaction. Only the confirmed HCV subtypes 1a and 4a isolates were further considered for sequence variability and molecular phylogeny analysis. The other HCV subtypes (e.g., 1b, 2a, 2b, 3a, 3b, 4b, 4d, etc.), mixed, and untypable subtypes were excluded from the study.

2.4. HCV RNA extraction and amplification of 5' UTR

For the amplification of 5' UTR, viral RNA was extracted from 140 μL of serum sample by using QIAamp Viral RNA Mini kit[®] (QIAGEN, California, USA) by following the kit protocol. The extracted RNA pellet was resuspended in 40 μL of TE buffer and purified by using PureLink[™] Viral RNA/DNA kit (Invitrogen[™], Carlsbad, CA, USA) for RNA concentration measurement in ng/ μL by using NanoDrop spectrophotometer[®] (Thermo-Fisher Scientific[™], Delaware, USA). The RNA yield was calculated from 45 ng/ μL up to 150 ng/ μL . RevertAid H Minus First Strand cDNA synthesis kit[®] (Thermo-Fisher Scientific[™], Delaware, USA) was used for cDNA synthesis by using 5' UTR outer antisense primer of both 1a and 4a subtypes ([Table 2](#)) in separate reaction mixture tubes by following the kit protocol. The cDNA products were further used for 5' UTR amplification in 25 μL reaction volume by using HCV subtypes 1a and 4a 5' UTR specific primers ([Table 2](#)) in thermal cycler 9700[®] (Applied Biosystems[™], CA, USA). The reaction mixture contained KCl buffer

2.5 μL , 25 mM MgCl_2 2.5 μL , 10 mM dNTPs 2.0 μL , 5' UTR inner sense primer (10 pmol/ μL) 1.0 μL , 5' UTR inner antisense primer (10 pmol/ μL) 1.0 μL , Taq DNA polymerase (1.25U/ μL) 0.25 μL , nucleic acid template (70–80 ng/ μL) 2 μL , and water nuclease-free up to final volume 25 μL . The thermal cyclic profile for amplification was 95°C for 2 min, 94°C for 35 s, 58°C for 30 s, 72°C for 25 s, and final extension for 10 min at 72°C . 1.8% agarose gel prepared in 1X TAE buffer and stained with ethidium bromide (2 μL) was used to separate amplified PCR products and characterized under UV transilluminator. PCR fragments were purified by eluting the gel to eliminate unincorporated primers and dNTPs by using the QIAquick gel extraction kit (QIAGEN, California, USA) protocol.

2.5. Sequencing of the 5' UTR amplicons

The purified PCR amplicons were sequenced by using the Big-Dye[™] Terminator v3.1 Cycle sequencing kit (Applied Biosystems, Germany) by following the kit protocol. The sequencing reaction mixtures were transferred into a 96-well sequencing plate. Standard Sanger dideoxy sequencing method was followed and the samples were analyzed by using ABI PRISM 3700 genetic analyzer (Applied Biosystems, Foster City, California, USA). All the samples were sequenced in both orientations (i.e. forward and reverse) to get consensus sequences and only HCV subtypes 1a and 4a sequenced isolates were considered for pair-wise comparison and evolutionary genetic analysis.

2.6. Basic Local alignment search tool analysis (BLAST) for raw sequences

Chromatogram for each isolate was collected in both forward and reverse orientation using Chromas 2.6.6 (Technelysium[®], South Brisbane, Australia) and displayed into separate text files. The 5' UTR forward and reverse primers were added to each corresponding strand respectively. The sequences were converted into FASTA format by using the following site: <http://searchlauncher.bcm.tmc.edu/seq-util/readseq.html>. The resulting sequences were aligned by using NCBI-Basic Local Alignment Search Tool (BLAST; http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=align2seq7). The sequences of each isolate were checked for mismatching, gaps, and corrected them based on QV data from chromatogram and according to IUPAC nucleotide code to get error-free refined sequences for the 5' UTR target. HCV subtypes of the refined sequences were confirmed by using online software; Oxford HCV subtyping tool (<http://www.bioafrica.net/rega-genotype/html/subtypetutorial-hcv.html>). Then the sequences were added to the HCV database by using online software (<http://www.hcvdb.org/blast.asp>). If NCBI-BLAST and HCV database results were corresponding, then the sequences were accepted to submit GenBank, NCBI database.

2.7. Nucleotide conservation and variation analysis of 5' UTR sequences

The representative sequences of HCV subtypes 1a and 4a isolates were submitted to GenBank, NCBI database, where accession numbers have been granted and can be retrieved under the accession numbers from MT240921 to MT240931 and MT327139 by using an online tool; <https://www.ncbi.nlm.nih.gov/nucleotide>. Pairwise sequence alignment tool Clustal W (BioEdit 7.2) and NCBI Multiple Sequence Alignment Viewer 1.12.0 (<http://www.ncbi.nlm.nih.gov/projects/msviewer>) program were used for pair-wise comparisons for nucleotide homologies/identities. Sequencher[®] 5.4.6 software (Gene Code Corporation[™], Miami, USA, <http://www.genecodes.com/html>) was used for nucleotide variance study where the isolates sequences were compared with

Table 1
HCV GT 1 and 4 positive patient's demographic characteristics and clinical profile with respect to age, sex, HCV diagnosis, and liver function test.

Demographic characteristics	Evaluation parameters	Genotype 1 (n = 17)	Genotype 4 (n = 23)	P-value
Age	Mean	48.6	46.8	0.752 ^a
	Median	46	52	
	Std. Deviation	12.95	14.85	
Gender	Male	10 (59%)	15 (65%)	0.45 ^b
	Female	07 (41%)	08 (35%)	
HCV diagnostic profile				
HAV, HBV and HDV surface antigens	Male	– ve	– ve	–
	Female	– ve	– ve	
HCV antibodies	Male	+ ve	+ ve	–
	Female	+ ve	+ ve	
Viral load (IU/ml)	Mean	329,3475	676,5015	0.065 ^a
	Median	850,955	975,655	
	Std. Deviation	110,4565	265,5172	
Serum and LFT profile				
Total proteins in serum (6–8 g/dl)	Mean	6.20	9.15	0.43 ^a
	Median	7.05	7.36	
	Std. Deviation	0.29	0.65	
Albumin in serum (3.5–5.0 g/dl)	Mean	3.45	5.16	0.39 ^a
	Median	3.99	4.25	
Bilirubin (direct) (0.02–0.4 g/dl)	Std. Deviation	0.95	0.68	0.007 ^a
	Mean	0.25	0.39	
Bilirubin (total) (0.1–1 g/dl)	Median	0.17	0.26	0.012 ^a
	Std. Deviation	0.009	0.004	
	Mean	0.47	0.95	
AST (SGOT) (5–41 U/L)	Median	0.32	0.71	0.56 ^b
	Std. Deviation	0.095	0.0599	
ALT (SGPT) (7–56 U/L)	Mean	31	57	0.046 ^b
	Median	28	34	
Alkaline phosphatase (20–140 U/L)	Std. Deviation	23.72	17.95	0.625 ^b
	Mean	45	70	
	Median	39	51	
Total platelets count (150–450 × 10 ⁶ /μL)	Std. Deviation	31.125	29.245	0.915 ^a
	Mean	90	125	
	Median	74	109	
	Std. Deviation	66	98	
	Mean	142,000	198,000	
	Median	136,000	186,000	
	Std. Deviation	109,000	155,000	

^aMann Whitney test, ^bFisher's Exact test.**Table 2**
List of primers used for PCR amplification of HCV 5' UTR region, sequencing, and subtyping of representative isolates.

Sr. No	Primer name	Primers sequences for HCV subtype 1a amplicon	Primer length	Sr. No	Primer name	Primers sequences for HCV subtype 4a amplicon	Primer length
A. Primers for HCV 5' UTR amplification							
01	IS/5'-1 (F)	ACCGAAAGCGTTAAGCCATGGGCC	24	03	IS/5'-3 (F)	CACCAGCGGGTGAAGCAGCATTGA	24
02	IS/5'-2 (R)	GTTGCAAGCACGGTATCAGGCAGA	24	04	IS/5'-4 (R)	GGACGGGGTAAACTATGCAACAGG	24
B. Primers for 5' UTR sequencing							
05	IS/5'-5 (F)	GCGAGCTTACCTGCCTCGTA	20	07	IS/5'-7 (F)	CGGTCTACCGCTGTGCTGCT	20
06	IS/5'-6 (R)	CGAGGCAAGATGTCGTTGAA	20	08	IS/5'-8 (R)	AATGAGGCCGGAGTGTAAATG	20
C. Primers for HCV subtypes identification							
09	5'UTR 1 F	GTCTAGCCATGGCGTTAGTATGAGTG	26	11	5'UTR 1 F	GTCTAGCCATGGCGTTAGTATGAGTG	26
10	5'UTR 1 R	ACAAGTAAACTCCACCAACGATCTG	25	12	5'UTR 1 R	ACAAGTAAACTCCACCAACGATCTG	25

complete genome reference sequences of subtypes 1a and 4a prototype strains respectively.

2.8. Phylogenetic analysis

To elucidate molecular phylogeny and evolutionary dynamics of sequenced isolates, the phylogenetic trees were generated by using molecular evolutionary genetic analysis software MEGA-X. First,

the isolates and comparable sequences were aligned by using the MUSCLE program, and pairwise genetic distances were computed by using the Kimura 2-parameter model and Maximum Composite Likelihood method with discrete gamma distributions of MEGA-X software. Then, the sequences were clustered with 200 nucleotide sequences of 5' UTR and complete genome reference sequences of prototype strains of comparable HCV strains to construct phylogenetic trees by using the Neighbor-Joining (NJ) method of MEGA-X

(the details of comparable reference sequences for phylogenetic trees estimation are provided in supplementary table S1). The sequences of comparable 5' UTR strains and reference sequences were derived from LOS ALMOS HCV database and GenBank database and only those sequences were compared which were previously reported from Saudi Arabia, partial and full-length 5' UTR sequences CDs reported from the rest of the world, and complete genome sequences of HCV subtypes 1a and 4a prototype strains. Bootstrap values for the associated taxa clustered together in phylogenetic trees were determined by a bootstrap test using 1000 replication values rearrangements. Only Bootstrap supports more than 50% were shown next to the nodes/branches. Bar scale values indicating branch lengths were proportional to the numbers of estimated base substitutions and were also indicated where appropriate.

2.9. Statistical analysis

Statistical analysis of the data was performed by using the Statistical Package for the Social Sciences (SPSS Inc., Chicago, IL, USA) software version 18. Nominal variables for GT 1 and 4 patients were calculated as frequency and percentage, while numerical variables (quantitative/continuous variables) were presented as mean, median, and standard deviation (SD). Fisher's Exact test was applied to compare nominal variables between GT 1 and 4 groups' patients, and a non-parametric Mann-Whitney test was used to compare numerical variables between subtypes 1a and 4a groups' patients. A statistically significant difference was assumed where the p-value was less than 0.05.

3. Results

3.1. HCV patient's demographics and baseline data

Fifty-five HCV positive patients' clinical data, laboratory findings, and serum samples were collected and analyzed. Out of 55 patients, 55% were male (n = 30) and 45% were females (n = 25). 85% patients were Saudi nationals (n = 47), 9% Egyptian (n = 5), and 5% belongs to Lebanon (n = 3). Most of the HCV clinical diagnostic parameters (quantitative variables) were found non-significant between GT 1 and 4 patients, however; SGPT levels were noted statistically significant in GT 1 patients. It is evident from the previous studies that GT 1 induced hepatic comorbidities (e.g., fibrosis, steatosis, and cirrhosis) are much severe, and more elevated hepatic enzymes profile (i.e., LFT) have been reported in infected patients (Shier et al., 2014, 2017; Aljowaiet al., 2020). Another variable was the platelet count, which was found relatively low in HCV GT-1 patients, but it was associated with some patient's medication history of administering antiplatelet agents (i.e. clopidogrel, Plavix[®], Bristol Myers Squibb[®], NY, USA) (Table 1).

3.2. Subtyping of HCV isolates

All HCV isolates were subtyped in the current study and retrieved sequencing results showed that subtype 4a was the most prevalent i.e. 38% (n = 19), followed by subtype 1a i.e. 25% (n = 12) (Table 3). Subtypes 1b, 2a, 3a, and 4d were also identified in some isolates (8.3%, 4.2%, 4.2%, and 6.2% respectively) and some isolates were found with mixed HCV subtyping (8%). Four HCV isolates showed 97% and 95% nucleotide identities with more than two subtypes (i.e. 4a/4d, 1a/2b) on NCBI blast search. No PCR products were visualized for two isolates in agarose gel and were referred to as the negative PCR amplification reactions.

3.3. 5' UTR amplification and nucleotide conservation analysis

In this study, around 210 bp 5' UTR gene fragments were amplified from HCV subtypes 1a and 4a isolates and were electrophoresed on 1.8% TAE agarose gel. The PCR amplicons were visualized and calibrated under UV transilluminator (Fig. 1). The amplicons were sequenced, and sequencing data were aligned with the NCBI nucleotide database by using BLAST. Out of 13 subtype 1a isolates and 19 subtype 4a isolates, 6 randomly selected sequences of each subtype were considered for nucleotide BLAST analysis. BLAST results revealed that sequenced isolates were homogeneous with each other, with comparable Saudi strains and rest of the world partial or complete 5' UTR sequences CDs, and with complete genome sequences of reference prototype strains (Table 4). The relative nucleotide positions of sequenced isolates to their reference prototype strains were also shown in Table 4 (i.e. 4E). Collectively, HCV subtypes 1a and 4a isolates showed maximum nucleotide identities (mean ± S.D; 97% ± 0.13, 96% ± 0.04) when blast with each other. Similarly, HCV subtypes 1a and 4a isolates demonstrated maximum nucleotide identities (95% ± 0.17; 96% ± 0.09) when aligned with two already reported 5' UTR sequences of HCV GT-1 and GT-4 strains from Saudi Arabia (i.e. KJ009305, KJ009311, KF999994, KJ009306) respectively. However, previously reported 5' UTR sequences of GT-4 Saudi strains did not reflect HCV subtypes in reported studies (Shier et al., 2014, 2017; Abdel-Moneim et al., 2012; Al-Faleh, 2003; al Nasser, 1992; Akbar et al., 2012; Bawazir et al., 2017) as well as not differentiated in the Genbank NCBI database. For this reason, the representative sequences were compared with other 5' UTR sequences from the rest of the world. Subtype 1a isolates were found in maximum homology (98% ± 0.02) with two strains of 5' UTR (LN681368 and D29818) belonging to India and Japan respectively, and 4a isolates were in maximum homology (95% ± 0.25) with two native Egyptian strains (i.e. AB550014 and DQ295833). For molecular phylogeny and to elucidate the evolutionary relationship of new isolates, those were aligned with complete genome reference sequences of prototype strains. For this purpose, two reference sequences each for 1a and 4a subtypes were retrieved from the NCBI database. Subtype 1a isolates were 98% identical with 1a reference sequences (AF009606 and EU862840), while 4a isolates were 97% homogeneous with two Egyptian reference prototype strains (i.e., KY283130 and Y11604). In conclusion, 5' UTR sequences of reported isolates showed maximum nucleotide identities with comparable 5' UTR consensus sequences (Table 4).

3.4. Nucleotide variance analysis

The nucleotide variations in 5' UTR of HCV subtype 1a isolates were noticed less than 4a (Table 5). Nucleotide variance analysis by Sequencher[®] 5.4.6 software revealed that 'C' dominates at the position 175 and 206 instead of 'T' and 'A' respectively in sequences of isolates SA1a/36, SA1a/43, SA1a/44, and SA1a/45. Similarly, 'A' dominates at positions 224 and 312 instead of 'G' and 'T' respectively for isolates SA1a/34, SA1a/43, SA1a/44, and SA1a/45 (Table 5). The isolate SA1a/44 sequence was demonstrated with the highest nucleotide substitutions among all 1a isolates while the other isolates showed maximum nucleotide identity with prototype strain (i.e. H77; AF009606). Overall, no predominant nucleotides were seen at the varied sites in subtype 1a isolate sequences. Interestingly, subtype 4a isolates showed significant nucleotide variations when compared to the reference sequence strain (i.e. ED43; Y11604) (Table 5). Nucleotide 'G' dominates two positions 88 and 181 instead of 'T' and 'A' for almost all 4a sequences. Nucleotide 'C' was dominating position 121 instead of 'T' for isolates SA1a/38, SA1a/41, and SA1a/42 respectively and nucleotide 'T' was dominating the position 222 instead of 'C' for

Table 3
HCV subtypes prevalence in representative isolates based on 5' UTR sequencing.

HCV genotypes	1		2		3		4		6	Mixed	^c Unknown (untypable)	^d Negative control	Total
HCV subtypes	1a	1b	2a	2b	3a	3b	4a	4d					
HCV isolates	13	4	2	1	2	0	19	4	2	3 ^a 1 ^b	2	2	55
HCV isolates amplified for 5' UTR	13	4	2	1	2	0	19	4	2	3 ^a 1 ^b	2	2	55
5' UTR amplified product in gel	12	4	2	1	2	0	19	4	2	3 ^a 1 ^b	0	0	50
No amplified product of 5' UTR in gel	1	0	0	0	0	0	0	0	0	0	2	2	5
Total samples sequenced for HCV subtyping	12	4	2	1	2	0	19	3	2	3 ^a 1 ^b	0	0	49
Total retrieved HCV subtypes results	12	4	2	1	2	0	19	3	2	3 ^a 1 ^b	0	0	49
Percentage of subtypes	25	8.3	4.2	2.1	4.2	0	38	6.2	4.2	8	0	0	100

^aMixed subtypes 4a/4d.

^bMixed subtypes 1a/2b.

^cHCV GT unidentified by any method.

^dPatient samples were tested -ve for HCV.

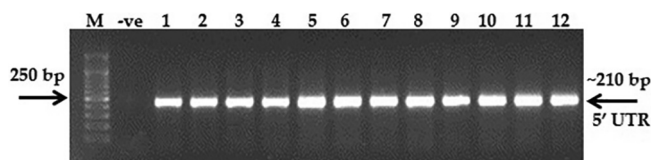


Fig. 1. PCR amplification of 5' UTR region of HCV subtypes 1a and 4a isolates

Each 5' UTR amplicon (~210 bp nucleotides in length) is depicted according to their isolate numbers. The first six PCR amplification products after -ve control represents 5' UTR amplicons of HCV subtype 1a isolates, while the remaining six are designated to HCV subtype 4a isolates. M: DNA ladder of 50 bp nucleotides, bp: base pair, -ve: negative control, 1: SA1a/34, 2: SA1a/35, 3: SA1a/36, 4: SA/IS-43, 5: SA/IS-44, 6: SA/IS-45, 7: SA1a/37, 8: SA1a/38, 9: SA1a/39, 10: SA1a/40, 11: SA1a/41, 12: SA1a/42.

isolates SA1a/37, SA1a/39, and SA1a/41 respectively. Isolates SA1a/37 and SA1a/41 were representing the highest nucleotide variations and depicted to be containing many polymorphic sites in their 5' UTR region. The nucleotide variations were also noticed higher in isolates SA1a/38 and SA1a/42 respectively.

3.5. Phylogenetic analysis

To infer the degree of genetic variations and evolutionary dynamics of HCV subtypes 1a and 4a isolates, phylogenetic trees were constructed and analyzed. All HCV isolates included in this study were clustered according to their subtypes. First, hierarchical clustering was performed for newly sequenced isolates with each other as shown in Fig. 2a & 2b. The phylograms revealed that considerable evolutionary distances exist among isolates although belong to the same subtypes (Fig. 2a-b). As newly reported isolates were diverse based on branch lengths which were proportional to the estimated number of base substitutions, we clustered them with already reported sequences of HCV GT 1 and 4 Saudi strains to cross-talk their molecular phylogeny. We were also curious to establish this evolutionary relationship of taxa because lesser data clued about evolutionary dynamics of existing Saudi strains of HCV subtypes. Sequenced isolates of both subtypes showed maximum homology with their comparable HCV Saudi strains and were clustered with more than cutoff values of their bootstrap replication test (i.e. >50%). The dendrogram showed that isolates SA1a/34, SA1a/35, SA1a/36, and SA1a-45 were 99% identical with two comparable GT 1 Saudi strains (SA-147 and SA-642) and were clustering as sister taxa (Fig. 2c). Isolates SA1a/36 and SA1a-45 were clustered as sister taxa in one clade and were branching with another clade of sister taxa clustering isolates SA1a/34 and

SA1a/35. Similarly, isolates SA/IS-43 and SA/IS-44 were clustering with comparable Saudi strain SA-124 and SA-110 with 99% and 98% nucleotide identities respectively (Fig. 2c). For subtype 4a isolates, two isolates SA1a/37 and SA1a/41 were clustering as an outgroup from where all other isolates and comparable Saudi strains diverged (Fig. 2d). Isolates SA1a/38 and SA1a/42 were 98% identical to comparable Saudi strain SA-682 and were clustered as sister taxa. Isolate SA1a/39 showed 99% coincide homology with two Saudi strains SA-204 and SA-126 and its tree position was in between these two strains (Fig. 2d). The identities between isolate SA1a/40 and Saudi strain TAIF.SA7 was also measured by 99%. Hence, the sequenced HCV isolates in this study were homogenous with their comparable Saudi strains of HCV GT 1 and 4 reported from different parts of the country and clustered with significant bootstrap replication support in phylogenetic tree lineages (Fig. 2-a-d).

To study the molecular phylogeny of sequenced isolates with rest of the world, comparable 43 HCV strains of subtype 1a and 44 strains of subtype 4a from different countries were selected and their sequences were retrieved from the Genbank NCBI database (Fig. 3; comparable HCV strains with their accession numbers are provided in electronic supplementary file (ESM_1). For subtype 1a clustering, the genetic lineages were selected from North and South America, Western Europe, and Southeast Asian countries (where HCV subtype 1a is the most prevalent). For 4a isolates, the phylogenetic tree of 5' UTR sequences was divided into North African (e.g. Egypt, Morocco) and South Asian genetic lineages (e.g. India) where subtype 4a is almost endemic in Egypt. The evolutionary relationship of associated taxa for subtype 1a isolates was closest to North American genetic lineage where isolates SA1a/34, SA1a/35, and SA1a/36 were clustering with the USA strains (Fig. 3a). Two isolates SA/IS-43 and SA/IS-44 were branching with strains from Western Europe [U51788 and AY7666618] and Southeast Asian [MH191416, MH191425, and MH191431] genetic lineage (In Fig. 3a, indicated by the red arrow). Only isolate SA1a-45 was 99% identical to Southeast Asian lineage and was clustering with an Indonesian strain [LC368404]. Interestingly, some subtype 1a strains from the Middle East (e.g. IRAQ) and North African (e.g. Morocco) lineages were clustering as contemporary sublineages with the USA strains in a paralog manner. Likewise, some strains from the Southeast Asian lineage [(e.g. from Vietnam with accession numbers MH191425, MH191416, and MH191431)] clustering as sister taxa with bootstrap replication values 100% were considered as an outgroup and possible ancestors of all descendants' isolates (Fig. 3a). For subtype 4a isolates, the phylogenetic analysis characterized their closest evolutionary relationship

Table 4

Pair-wise comparison of HCV subtypes 1a and 4a Saudi isolates and their relative nucleotide positions on prototype HCV strains.

A. Pair-wise percentage nucleotide identities of sequenced isolates with each other						
1a isolates	SA1a/34	SA1a/35	SA1a/36	SA/IS-43	SA/IS-44	SA/IS-45
SA1a/34	100	99	97	97	95	99
SA1a/35	99	100	98	98	96	98
SA1a/36	97	98	100	97	97	98
SA/IS-43	99	97	96	100	97	99
SA/IS-44	95	96	97	97	100	95
SA/IS-45	99	98	98	99	95	100
4a isolates	SA1a/37	SA1a/38	SA1a/39	SA1a/40	SA1a/41	SA1a/42
SA1a/37	100	93	95	93	99	94
SA1a/38	93	100	95	97	95	98
SA1a/39	95	95	100	99	96	96
SA1a/40	93	97	99	100	95	97
SA1a/41	99	95	96	95	100	96
SA1a/42	94	98	96	97	96	100

B. Pair-wise percentage nucleotide identities of sequenced isolates with already reported sequences from Saudi Arabia									
Sr. No	Isolate	Sequence ID	GT	Identity (%)	Sr. No	Isolate	Sequence ID	GT	Identity (%)
01	SA1a/34	KJ009305	1	205/207 (99%)	07	SA1a/37	KF999994	4	179/188 (95%)
		KJ009311		191/207 (92%)			KJ009306		176/189 (93%)
02	SA1a/35	KJ009305	1	194/197 (98%)	08	SA1a/38	KF999994	4	200/207 (97%)
		KJ009311		189/197 (96%)			KJ009306		198/208 (95%)
03	SA1a/36	KJ009305	1	200/206 (97%)	09	SA1a/39	KF999994	4	200/202 (99%)
		KJ009311		188/195 (96%)			KJ009306		198/203 (98%)
04	SA/IS-43	KJ009305	1	199/202 (99%)	10	SA1a/40	KF999994	4	201/204 (99%)
		KJ009311		186/202 (92%)			KJ009306		199/205 (97%)
05	SA/IS-44	KJ009305	1	195/203 (96%)	11	SA1a/41	KF999994	4	181/188 (96%)
		KJ009311		183/193 (95%)			KJ009306		179/189 (95%)
06	SA1a/45	KJ009305	1	203/207 (98%)	12	SA1a/42	KF999994	4	191/196 (97%)
		KJ009311		191/207 (92%)			KJ009306		189/197 (96%)

C. Pair-wise percentage nucleotide identities of sequenced isolates with partial 5' UTR sequences from the rest of the world									
Sr. No	Isolate	Sequence ID	GT	Identity (%)	Sr. No	Isolate	Sequence ID	GT	Identity (%)
01	SA1a/34	LN681368	1a	206/207 (99%)	07	SA1a/37	AB550014	4a	179/188 (95%)
		D29818		207/207 (100%)			DQ295833		176/193 (91%)
02	SA1a/35	LN681368	1a	195/197 (99%)	08	SA1a/38	AB550014	4a	200/207 (97%)
		D29818		194/195 (99%)			DQ295833		195/209 (93%)
03	SA1a/36	LN681368	1a	200/206 (97%)	09	SA1a/39	AB550014	4a	197/202 (98%)
		D29818		200/205 (98%)			DQ295833		194/202 (96%)
04	SA/IS-43	LN681368	1a	198/202 (98%)	10	SA1a/40	AB550014	4a	199/204 (98%)
		D29818		199/202 (99%)			DQ295833		196/206 (95%)
05	SA/IS-44	LN681368	1a	193/203 (95%)	11	SA1a/41	AB550014	4a	180/188 (96%)
		D29818		194/203 (96%)			DQ295833		179/193 (93%)
06	SA1a/45	LN681368	1a	204/207 (99%)	12	SA1a/42	AB550014	4a	191/196 (96%)
		D29818		205/207 (99%)			DQ295833		186/198 (95%)

D. Pair-wise percentage nucleotide identities of sequenced isolates with complete genome reference sequences of prototype strains									
Sr. No	Isolate	Accession numbers	Serotype	H77 (AF009606)	Sr. No	Isolate	Accession numbers	Serotype	ED43 (Y11604)
01	SA1a/34	AF009606	1a	206/207 (99%)	07	SA1a/37	KY283130	4a	179/188 (95%)
		EU862840		205/207 (99%)			Y11604		177/188 (94%)
02	SA1a/35	AF009606	1a	198/201 (99%)	08	SA1a/38	KY283130	4a	201/207 (97%)
		EU862840		196/201 (98%)			Y11604		200/207 (97%)
03	SA1a/36	AF009606	1a	202/206 (98%)	09	SA1a/39	KY283130	4a	200/202 (99%)
		EU862840		198/203 (98%)			Y11604		198/202 (98%)
04	SA/IS-43	AF009606	1a	200/202 (99%)	10	SA1a/40	KY283130	4a	202/204 (99%)
		EU862840		199/202 (99%)			Y11604		201/204 (99%)
05	SA/IS-44	AF009606	1a	195/203 (96%)	11	SA1a/41	KY283130	4a	181/188 (96%)
		EU862840		193/202 (96%)			Y11604		179/188 (95%)
06	SA1a/45	AF009606	1a	206/207 (99%)	12	SA1a/42	KY283130	4a	192/196 (98%)
		EU862840		205/207 (99%)			Y11604		191/196 (97%)

E. Relative alignment positions of sequenced subtypes 1a and 4a isolates with reference sequences of prototype strains									
Sr. No	Isolates	Accession numbers	Serotype	H77 (AF009606)	Sr. No	Isolates	Accession numbers	Serotype	ED43 (Y11604)
01	SA1a/34	MT240922	1a	105–311	07	SA1a/37	MT240924	4a	62–249
02	SA1a/35	MT240921	1a	65–264	08	SA1a/38	MT240925	4a	7–213
03	SA1a/36	MT240923	1a	70–275	09	SA1a/39	MT240926	4a	48–249
04	SA/IS-43	MT240930	1a	110–311	10	SA1a/40	MT240927	4a	7–210
05	SA/IS-44	MT240931	1a	72–274	11	SA1a/41	MT240928	4a	62–249
06	SA1a/45	MT327139	1a	105–311	12	SA1a/42	MT240929	4a	6–201

to Egyptian strains although distantly related (Fig. 3b). It was noticed that the positions of two isolates SA1a/39 and SA1a/40 were at the end of the sub-tree with strains [MF497267] and [DQ295833] with poor bootstrap replication test values. The phy-

logenetic tree also depicted that HCV subtype 4a might have an overlap evolutionary origin in the Middle East because of the consequences of recombination or duplication event of their genome with North African lineages. We are reporting this on the evidence

Table 5
Nucleotide variations in 5' UTR sequences of HCV subtypes 1a and 4a Saudi isolates with reference sequences of prototype strains.

Prototype strain 1a (H77) AF009606		SA1a/34	SA1a/35	SA1a/36	SA1a/43	SA1a/44	SA1a/45	Total
Position	Nucleotide	MT240922	MT240921	MT240923	MT240930	MT240931	MT327139	
107	G					A		1
175	T				C	C		2
203	T					A		1
204	A			T		G		2
205	A					T		1
206	A			C			C	2
224	G				A	A		2
243	A	G	G					2
312	T	A			A		A	3
313	G						A	1
Total		2	1	2	3	6	3	17
Prototype strain 4a (ED43) Y11604 ^a		SA1a/37	SA1a/38	SA1a/39	SA1a/40	SA1a/41	SA1a/42	Total
Position	Nucleotide	MT240924	MT240925	MT240926	MT240927	MT240928	MT240929	
81	G	A						1
88	T	G	G	G	G	G	G	6
109	G	C				C		2
115	C	T						1
121	T	G	C			C	C	4
125	C	T				T		2
142	T			A	A			2
143	A						T	1
153	A	G				G		2
181	A	G	G			G	G	4
201	G		C					1
222	C	T		T		T		3
225	T	G				G		2
Total		10	4	3	2	8	4	31

Distribution of mutations within IRES domains of 5' UTR of HCV subtypes 1a and 4a isolates

HCV subtype 1a isolates			HCV subtype 4a isolates		
Domain	^b Mutations	^c Occurrence	Domain	^b Mutations	^c Occurrence
II	1	6	II	16	52
III	16	94	III	15	48
IV	0	0	IV	0	0
Total	17			31	

^aNucleotide positions were based on prototype H77 strain (AF009606). G = Black, T = Red, A = Green, C = Blue. ^bAbsolute numbers of identified mutations are shown within the respective domains. Most mutations were found in domain III stem-loops (i.e. IIIb- IIIf) of IRES structure for subtype 1a isolates, while for the subtype 4a isolates, mutations were found within domain II stem-loops and domain III stem-loops (i.e. IIIa-IIIId) of IRES.

^cThe abundance of mutations in each domain (in percentage) relative to the total number of mutations in isolates of subtypes is shown.

where the phylogram in Fig. 3b showed that 4 distantly related sequenced isolates reported in this study were branching as an outgroup with all comparable strains from North Africa and South Asian genetic lineages.

We constructed another phylogenetic tree to demonstrate if similar results were found for sequenced isolates when clustered with complete genome reference sequences of prototype strains of HCV subtypes 1a and 4a from different countries (Fig. 4). Second, we want to comparatively evaluate that how much variants were the sequenced isolates when compared with constructed phylogenetic trees of comparable partial 5' UTR genome region and complete genome of prototype strains. The dendrogram in Fig. 4a showed that two isolates; SA1a/34 and SA1a-45 were 100% identical with reference strain sequences EU862831 and EU862840 respectively from the USA and clustered as sisterhood taxa with high bootstrap values. Isolate SA1a/36 was 98% homogeneous with reference strain sequence EU256071 from Switzerland and indicated that the origins of this isolate might be Swiss although it could only be verified by genotyping of HCV isolates from other

regions. Isolate SA1a/35 was branching to all clustered reference strains except to isolates SA/IS-43 and SA/IS-44 which were not clustering with any subtype 1a reference strains and were distantly branching to one synthetic construct from the USA [AF177040], and with a mixed subtype (2a/1a) strain from Denmark [HQ852468] (In Fig. 4a, indicated by the red arrow). All subtype 4a sequences were clustering with Egyptian prototype strains as sister taxa in a clade with high bootstrap replication values. Furthermore, a sequence of Canadian strain [JF735137] was noticed as a contemporary sublineage and all newly reported subtype 4a sequences were sub-branching to isolate SA1a/39 with 100% bootstrap replication values indicating the highest robustness of the tree (Fig. 4b).

All our phylogenetic analysis for new isolates inferred that subtype 1a isolates from the Saudi population had the closest evolutionary relationship to HCV genetic lineage from North America and Western Europe although not all isolates grouped or clustered with complete genome reference sequences of subtype1a. In contrast, 4a isolates were homogenous and clustered to Egyptian

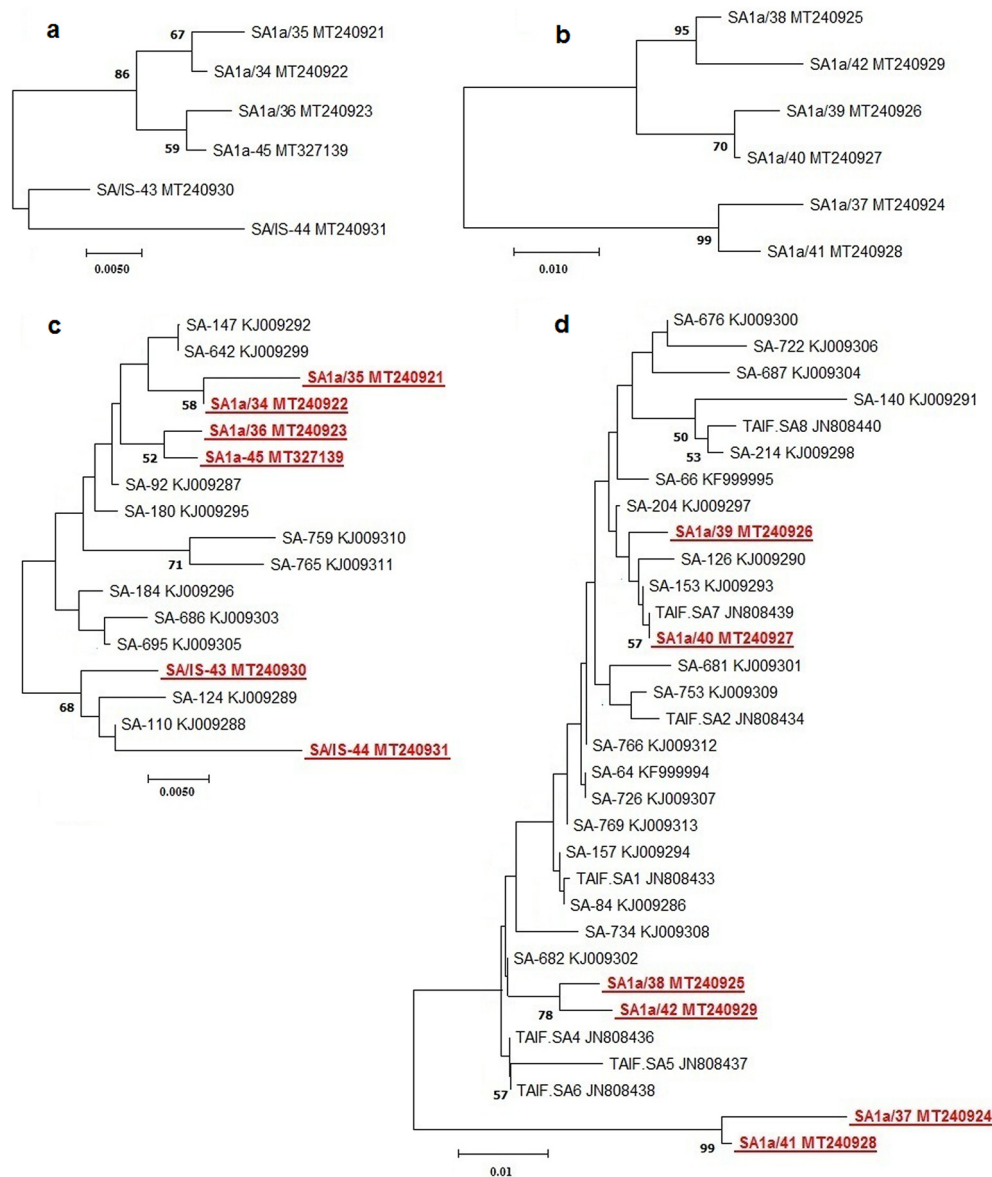


Fig. 2. Phylogenetic relationship of HCV subtype 1a and 4a isolates clustering with each other and already reported Saudi strains of GT 1 and 4. Unrooted phylogenetic trees were constructed by using the Neighbor-Joining (NJ) method to infer the evolutionary history of associated taxa clustering together. Bootstrap values based on 1000 replicates are shown next to the nodes/branches. The optimal trees with the sum of branch length are shown. The horizontal branch length is proportional to the estimated number of base substitutions and evolutionary distances were computed by using the Maximum Composite Likelihood method. Sequences were labeled to the right side of each branch in an order of isolate name and corresponding GenBank accession number. The representative sequences of HCV subtypes 1a and 4a isolates reported in this study are texted and underlined with red color font. **a:** Phylogenetic tree of HCV isolates 1a by the Neighbor-Joining method. **b:** Clustering of HCV subtypes 4a isolates with each other in the phylogenetic tree. **c:** Phylogenetic tree of sequenced HCV subtype 1a isolates with already reported GT 1 Saudi strains. Bootstrap replication runs by which associated taxa were clustering are shown next to the branches. **d:** Phylogenetic analysis of sequenced HCV subtype 4a isolates with comparable Saudi strains of GT 4.

prototype strains as sister taxa in a clade indicating their evolutionary relationship with North African and Egyptian HCV GT 4 genetic lineage.

4. Discussion

The virological surveillance studies demonstrate the viral infection trajectory, transmission, and prevalence of infection in a geographical region. HCV surveillance and correct GTs/subtypes identification are also essential for HCV epidemiological studies, to choose optimal treatment strategies, and prognosis in the treated population. The previous surveillance studies conducted in Saudi Arabia predicted the highest prevalence of HCV GT 4 fol-

lowed by 1, however; data on HCV subtypes prevalence are murky, contradictory, and inconclusive (Shier et al., 2014, 2017; Abdel-Moneim et al., 2012; Al-Faleh, 2003; al Nasser, 1992; Akbar et al., 2012; Bawazir et al., 2017). An 11-year surveillance study demonstrated that subtype 4a is the most prevalent followed by 1a, in Saudi Arabia (Abdel-Moneim et al., 2012). Similarly, the surveillance conducted from 2008 to 2011 based on the HCV transmission rate and ratio of viral clearance also depicted the same findings (Akbar et al., 2012). However; two studies reported by Shier et al. (Shier et al., 2014, 2017), demonstrated that subtypes 1b and 4d were dominating in Saudi infected population, although the former study by Shier et al. (Shier et al., 2014) showed mixed prevalence of 1a, 1g-, 1b, 4a, and 4d subtypes in Saudi patients. A study by Aljowaie et al. (2020) also elucidated subtypes 4a and

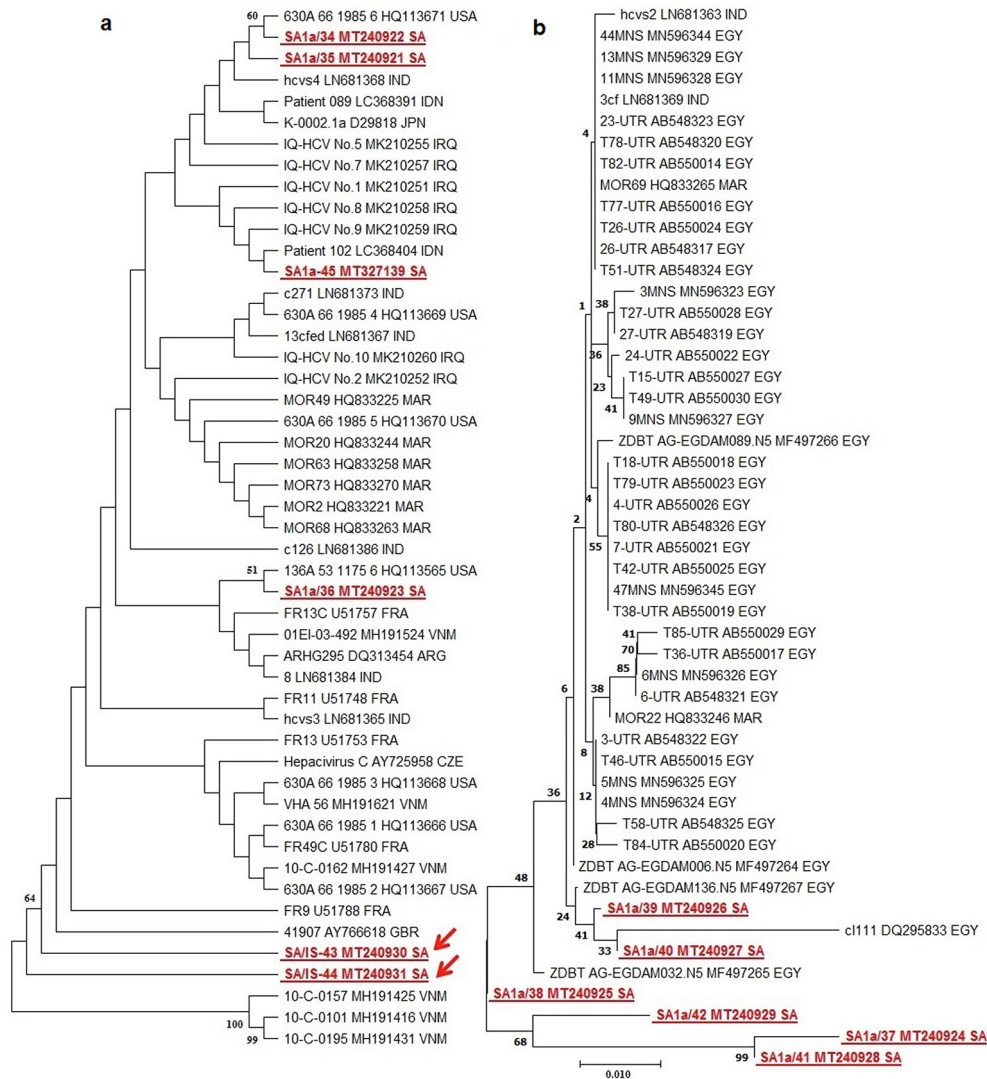


Fig. 3. Phylogenetic analysis of sequenced HCV subtypes 1a and 4a isolates with comparable 5' UTR sequences from the rest of the world. Distance trees were constructed by using the Neighbor-Joining method and the robustness of the trees was estimated by performing 1000 bootstrap replicates which are expressed as percentages next to the branches. Sequences were labeled to the right side of each branch in an order of isolate name, GenBank accession number, and country-code (as shown in electronic supplementary material file (ESM_1)). **a:** Tree view of HCV subtype 1a sequenced isolates (As shown in red color and underlined) clustered to the associated taxa. Two isolates (i.e. SA/IS-43 and SA/IS-44) indicated by red arrow are distant genetic lineage and predicted as 'distinct subpopulation' of subtype 1a sequenced isolates. **b:** Phylogenetic tree of HCV subtype 4a isolates. The scale bar with the sum of branch length is shown. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The representative isolates are in red color font and are underlined.

4d most prevalent in Saudi Arabia. In this study, 55 randomly selected isolates of HCV from the Saudi population were sequenced for surveillance studies for a period from March 2018 to December 2018. As shown in Table 3, our findings are compatible with the studies of Akbar et al. (2012), Abdel-Moneim et al. (2012), and Aljowaie et al. (2020) who demonstrated the highest prevalence of subtypes 4a and 1a in the Saudi population, however; differed from Shier et al. (2014), Shier et al. (2017) who elucidated 1b and 4a the predominant subtypes in Saudi population.

The 5' UTR is highly conserved (92%–98%) as compared to other HCV genome regions, mostly suited for amplification methods, and contains specific sequence motifs for HCV GTs/subtypes identification (Baclig et al., 2010). Many diagnostic laboratories and most commercially available HCV typing assays target 5' UTR due to higher genotype-based assay sensitiveness (Moratorio et al., 2007). However, mounting evidences suggested that direct sequencing of 5' UTR did not identify all existing HCV GT 1

subtypes in 20% infected cases (Baclig et al., 2010). Verbeeck et al. (2008) also assumed the limited subtype accuracy of HCV isolates based on 5' UTR amplification and sequencing because of the too much-conserved nature of the region for differentiation of subtyping. Similarly, according to various studies, a 16% rate of mistyping may result in an equilibration of the observed shift in subtype prevalence (Ross et al., 2007). Other methods based on more variable regions of the HCV genome (e.g. NS5B) could be relied upon for accurate identification of subtypes (Baclig et al., 2010). Sandres-Saune et al. (2003) described the sequencing and phylogenetic analysis of the NS5B region as the first step in molecular epidemiological studies to recognize the route of HCV transmission. NS5B is an extremely preferred region for HCV subtyping, but it is not always accurately amplified because of primer-target mismatch to highly variable nucleotide region of NS5B (Baclig et al., 2010; Sandres-Saune et al., 2003). Although, 5' UTR is the most conserved part of the virus genome, however;

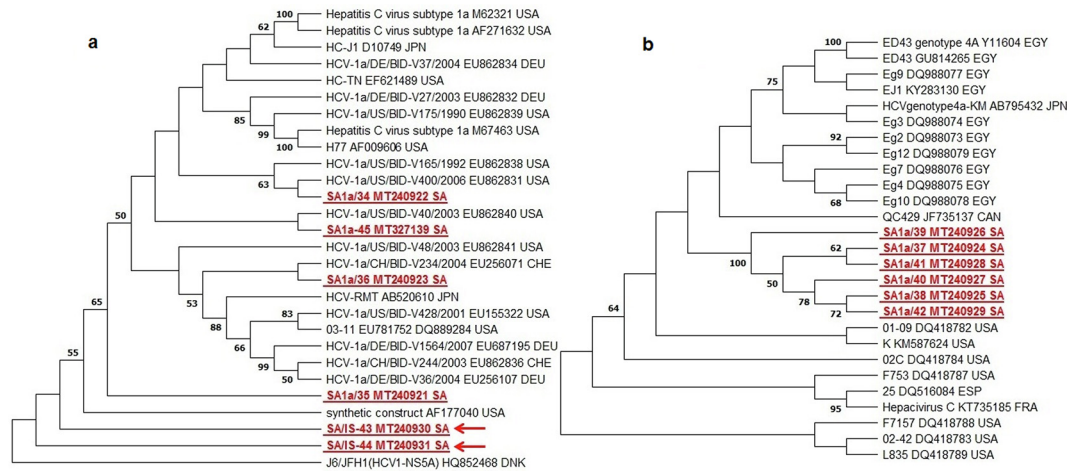


Fig. 4. Phylogenetic relationship of HCV subtypes 1a and 4a isolates with complete genome reference sequences of prototype strains. Unrooted phylogenetic trees show the clustering of reported sequences with complete genome reference sequences of prototype strains of HCV subtypes 1a and 4a. The numbers on the branches indicate bootstrap values obtained after 1000 replications. Sequences were labeled to the right side of each branch in an order of isolate name, GenBank accession number, and country-code (as provided in Table ESM_1). **a:** Phylogenetic and evolutionary analysis of HCV subtypes 1a isolates with complete reference genome sequences of prototype strains. Molecular phylogeny was determined by using the Neighbor-Joining method, a distance algorithmic method, and stability of clades was evaluated by 1000 bootstrap rearrangements. Two isolates (i.e. SA/IS-43 and SA/IS-44) indicated by red arrow are distant genetic lineage and predicted as 'distinct subpopulation' of subtype 1a sequenced isolates. **b:** Phylogenetic analysis of HCV subtypes 4a isolates with complete reference genome sequences of prototype strains: Phylogenetic analysis was performed by using the minimum evolution algorithm with 1000 bootstrap cycles as shown numbers on branches.

with minor variants, it expresses quasispecies distribution in the infected populations (Moratorio et al., 2007), as shown in Table 3, 8% mixed subtypes were identified in HCV Saudi isolates in the current study. The reasonable justification of this phenomenon might be the existence of randomly occurring mutations distributed within 5' UTR gene due to the error-prone nature of HCV RNA-dependent RNA polymerase during viral replication (Moratorio et al., 2007). *In silico* predicted RNA secondary structure of IRES (internal ribosome entry site) stem-loops of 5' UTR explored that some mutations in IRES might affect IRES structure to confer a survival advantage or disadvantage to mutated HCV genome in the form of quasispecies during HCV replication- (Moratorio et al., 2007). Since therapeutic decisions to treat CHC patients are entirely based on HCV GT/subtypes identification, accurate GT/subtype -detection would help to choose the best IFN-free DAAs for HCV treatment outcome. It has been reported previously that GT 1 and 4 are less responsive to PEG-IFN/RBV for 48 weeks and subtype 1a was associated with lower treatment response than subtype 1b (Farci et al., 2000; El-Tahan et al., 2018; Farci et al., 2002; Legrand-Abbravanel et al., 2005). For pan-genotypic DAAs, GT-1 is well treated with second-generation DAAs (e.g. sofosbuvir, daclatasvir, etc.) achieving SVR rates more than 90%, however; less effective against subtype 4r in Africa and the Middle East including Saudi Arabia (Aljowaie et al., 2020). A study reported by Dietz et al. (2018) also described the more frequent treatment failure for GT-4 subtypes 4a and 4d patients on treatment with daclatasvir/sofosbuvir or ledipasvir/sofosbuvir. In clinical perspectives, subtype 1b is associated with more pathogenicity induced by HCV related advanced hepatic co-morbidities and higher viral load (Aljowaie et al., 2020; Baclig et al., 2010). Likewise, subtypes 4o, 4r, and 4f of GT-4 are more prone to induce hepatic cirrhosis and HCC in infected individuals (El-Tahan et al., 2018; Aljowaie et al., 2020; Baclig et al., 2010).

The findings of pair-wise comparison and multiple sequence alignment for nucleotide conservation, nucleotide variations, and positional mutations within representative 5' UTR sequences were also in an agreement with previous studies of Shier et al. (2014), El-Tahan et al. (2018), Moratorio et al. (2007), Vopalensky et al. (2018), El Awady et al. (2009), and Zekri et al. (2007) which demonstrated nucleotides conservation from 92% to 98% within

5' UTR region of HCV genome. Interestingly, nucleotide differences were observed as low <6% (i.e. 93.5–94.5% nucleotide identities) among subtypes 1a and 4a isolates. The nucleotide positions of sequenced 5' UTR regions of both subtypes 1a and 4a isolates were variable to the reference prototype strains (Table 4). However, for both subtypes, the sequenced regions covered major parts of domain-II and domain-III stem-loops structures of IRES, which are crucial for virus translation initiation and IRES activity (Friebe et al., 2001; Zekri et al., 2007; Zekri et al., 2011). It would be useful to characterize positional mutations in 5' UTR region of Saudi isolates to implicate their roles in IRES-mediated virus translation as well as to correlate with treatment outcomes in the future perspective of this study. Nucleotide variance data in Table 5 showed that most of the subtype 1a isolate mutations (total 16, 94%) were localized in domain III (within junction joining stem-loop IIIa, b, c & d, and loop III'd), specifically in segment comprising highly conserved nucleotide (141 to 279) region of domain III (i.e. at positions 175, 203–206, 224, and 243 of prototype H77 subtype 1a reference strain in Table 5). Only one mutation (at position 107, G → A) was noticed in domain-II for all sequenced subtype 1a isolates. In contrast, for subtype 4a isolates, 52% mutations (total 16) were localized in domain II stem-loop, and 48% mutations (total 15) were found across domain III, the majority of which were noticed in stem-loop IIIa and IIIb. Mutations were found as nucleotide substitutions and no predominant nucleotides (i.e. insertions or deletions) were found at varied sites. It has been demonstrated that the mutations in domain III stem-loops (i.e. IIIa and IIIb) are attributed to decrease RNA stability, while mutations within III'd stem-loops are associated with increased RNA stability (Zekri et al., 2011; El Awady et al., 2009; Zekri et al., 2007). Although, RNA secondary and tertiary structure stability is regarded as a significant factor for virus genome stability, but not essential to predict virus stabilization and response to PEG-IFN α therapy (El Awady et al., 2009; Zekri et al., 2007). Our data are consistent with those of Vopalenský et al. (Vopalensky et al., 2018) who demonstrated 102 mutations in HCV subtype 1a IRESs genome isolated from non-responders and 53 mutations in sustained responders to PEG-IFN α plus RBV respectively. El-Tahan et al. (El-Tahan et al., 2018) identified 35.7% mutations localized in stem-loop IIIb (nucleotides 172–227) of IRES in isolated HCV subtype 4a Egyptian

strains. El-Awady et al. (El Awady et al., 2009) reported 19 mutations (i.e. 14 mutations were nucleotide substitutions and 5 were nucleotide insertions) in patients with significant SVRs to PEG-IFN α /RBV and viral breakthrough patients. Seven nucleotide variations at positions 74, 92, 112, 113, 133, 172, and 180 in 5' UTR region of HCV subtype 4a strains of PEG-IFN/RBV non-responders (n = 3) were reported by Hemeida et al. (Hemeida et al., 2011). Interestingly, none of those substitutions were recorded in our HCV subtype 4a isolates. Zekri et al. (2007) reported one unique mutation at position 160 (160 G \rightarrow A) in 5'UTR strains of HCV GT 4 non-responders patients. This position corresponds to position 121 of HCV subtype 4a isolates in the current study which lies in domain II stem-loop structure of IRES, and polymorphic (G/C) for two isolates SA1a/37 and SA1a/41 respectively.

The ancestral evolution of HCV subtype 1a and 4a Saudi isolates in this study was estimated by generating and analyzing phylogenetic trees. Hierarchical clustering was performed at four levels to correlate pair-wise comparisons, genetic distances calculations among clustered taxa, and by constructing phylogenetic trees to infer homogeneous evolution trajectory (Figs. 2 to 4). Un-rooted phylogenetic tree of HCV subtype 1a isolates showed clustering with reference HCV subtype 1a strain from North America and Europe with significant bootstrap values (i.e. greater than 70% bootstrap replication) irrespective of whether full-length or partial sequences of referenced 5'UTR strains were analyzed (Fig. 2a, 2c, Fig. 3a, and Fig. 4a). A bootstrap support greater than 70% is statistically equated to a P-value less than 0.05. However, two isolates were observed as 'unrelated isolates' (i.e. SA/1S-43 and SA/IS-44) clustering/co-existing together with comparable strains AF177040, and HQ852468 and were regarded as 'distinct subpopulation' of subtype 1a isolates (in Fig. 4a indicated by red arrows). This finding is consistent with quasispecies dynamics of HCV GT 1 due to naturally occurring variants (i.e. nucleotide signature sequence) within 5' UTR region of HCV in the human population as reported by Moratorio et al. (Moratorio et al., 2007) who demonstrated the existence of a distinct subtype 1a sub-population and HCV diversification in South American HCV strains.

HCV subtype 4a isolates were found homogenous and clustered with Egyptian partial sequences of comparable strains of 5' UTR and reference sequences of prototype strains in constructed phylogenetic trees with significant bootstrap support (i.e. \approx 70%) (Figs. 3b and 4b). It is evident from the cladogram that the evolutionary relationship of HCV subtype 4a isolates much closer to central and West African HCV GT-4 strains (Figs. 3b and 4b). HCV GT 4 is highly prevalent in the Middle East, Central/West Africa, and an

increased emergence and propagation has been reported in Europe, North America, and in South American region (e.g., Argentina) (Moratorio et al., 2007; Hmaied et al., 2007; Kuntzen et al., 2008). According to updated HCV databases, 19 subtypes (a-h and k-u) have been assigned to HCV GT-4 (Hepatitis C Virus Database (HCVdb) <http://www.hcvdb.org/index.asp?bhcp=1>. Accessed 15 September 2020), however full-length reference sequences of only 4a, 4b, 4d, 4f, 4g, 4k, 4l, 4m-r, and 4t subtypes are currently available (Kuntzen et al., 2008; Hmaied et al., 2007; Li et al., 2009; Timm et al., 2007).

Coalescent study approaches have been indicated that HCV GT 4 strains originally evolved and propagated in Central and West Africa before transmitting to other regions (Shier et al., 2014; Li et al., 2009). Some strains in North Africa had been prevalent since 1930 as a result of large-scale vaccination campaigns (Li et al., 2009). HCV-4a has been estimated to appear early in the 20th century while 4d in the middle of the 20th century (Shier et al., 2014). Although, HCV GT-4 infection is not common in the USA and Canada, but the majority of cases have been reported among PWID (Patients who inject drugs) or immigrants traveling from the areas where GT-4 is almost endemic or persons acquired infection while living in those areas (Timm et al., 2007; Fernandez-Arcas et al., 2006). Furthermore, some strains of GT-4 have been reported in injection drug users (IDUs) of Southern European Countries of the Mediterranean Sea (Fernandez-Arcas et al., 2006). Our data are consistent with the deliverables of those studies as deciphered in Fig. 4b where some sequenced isolates with more than 60% bootstrap replication support were closely branching with reference strains of subtype 4a from the USA, Canada, Spain, and France. As described above, nucleotide differences as low as 6% were observed among subtypes 1a and 4a reported sequences, it may reflect that both GT 1 and 4 have the same evolutionary origin to a single 'ancient genotype' from Central or West Africa (Li et al., 2009). At the subtype level, these differences may indicate that many strains miss continual genetic variations of the HCV genome in those subtypes (Li et al., 2009). A study reported by Hmaied et al. (Hmaied et al., 2007) supports these hypotheses while demonstrating close relationships of HCV GT-4 strains to GT 1 than to other GTs. Similarly, it was also shown that subtype 4f of GT 4 was closest to GT 1. Franco et al. (Franco et al., 2007) also elucidated this intergenotypic relationship while describing close relationships of two sequences of complete genomes of HCV 4a and 4d subtypes with GT 1. To further strengthen the hypothesis, we clustered subtype 1a isolates with 4a isolates to construct a phylogenetic tree with bootstrap analysis based on 1000 replication

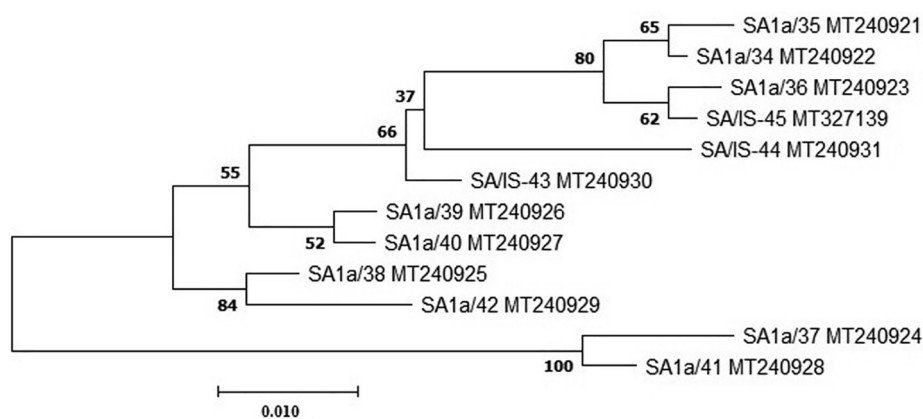


Fig. 5. Bootstrap analysis of HCV subtypes 1a and 4a isolates clustered in a phylogenetic tree Unrooted phylogenetic tree was generated by using the Neighbor-Joining (NJ) method with bootstrap replication support of 1000 replicates. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. 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 Maximum Composite Likelihood method and are in the units of the number of base substitutions per site.

rearrangements. The phylogenetic tree showed clustering of 10 isolates having bootstrap support of more than 50% and branching with 2 isolates of subtype 4a as an outgroup with a full bootstrap support of 100% replication (Fig. 5). This close pattern of clustering predicts the overlapping evolution of HCV GT 1 and 4 from a 'common ancient ancestral GT' in Central/West Africa and its transmission to the Middle East (e.g. Saudi Arabia). However, full-length genome sequences with an improved HCV GT/subtype identification approach are highly recommended to study the evolutionary dynamics of HCV GTs and subtypes.

5. Study limitations

A small patient pool (i.e. n = 55) investigated in this study cannot fully justify the surveillance studies and the patient's clinical spectrum at baseline laboratory characteristics specifically for the epidemiological studies of HCV infection. However, it could be defensible in Saudi Arabia where the prevalence and transmission rate of HCV infection have been decreased due to better public health care facilities, blood screening, strict blood transfusion, and after the availability of promising IFN-free pangenotypic DAAs. Partial 5'UTR sequences reported in this study provide a basic understanding of genetic heterogeneity and HCV GTs/subtypes endemicity in Saudi Arabia, However; may not be an ideal representation of GT/subtype distribution because of the given small sample size, the low variability of the 5'UTR region which limits the genetic variability of the analyses performed. Full-length genome sequences are warranted for in-depth elucidation of quasispecies dynamics, HCV GTs/subtypes diversification, evolutionary analysis, and as well as for the rational design and development of promising oral IFN-free DAAs and anti-HCV vaccines.

6. Conclusions

We report HCV surveillance in the Saudi population by analyzing 55 clinical specimens of different HCV GTs and subtypes infected patients from different parts of the country based on GT/subtypes identification and frequency of their occurrence. The study predicts decreased HCV transmission in the Saudi population as compared to the previous surveillance studies. Nucleotide data submission of 12 5'UTR sequences will enrich the HCV database of Saudi HCV strains which could be helpful to use for full-length genome amplification, improving HCV GTs/subtypes identification, and search for better HCV subtypes classification methods. Nucleotide variations and positional mutations found in highly conserved domains of 5'UTR regions would help to better understand the intrinsic sensitivity of HCV subtypes 1a and 4a Saudi isolates to newer promising therapeutic options. Phylogenetic diversity predicts the close evolutionary relationship of sequenced Saudi isolates with their possible ancestors from North America and Central African HCV strains and would support to explicit their molecular evolutionary studies and the possibilities of crossing the barriers of genetic lineage again to evolve new subtypes in HCV GT 1 and 4 infected populations.

7. Availability of data and material

The nucleotide sequences of HCV subtype 1a and 4a isolates analyzed and discussed in this study are available on GenBank, NCBI database with their accession numbers from MT240921 to MT240931 & MT327139 and can be retrieved by using an online tool; <https://www.ncbi.nlm.nih.gov/nucleotide>. All relevant data used to pair-wise and multiple sequence alignments and constructing phylogenetic trees were provided as the electronic supplementary material (ESM_1).

Funding

The authors are thankful to King Abdullah City of Science and Technology (KACST-STP), Riyadh, Saudi Arabia to provide funding under the project ID: 13-MED944-10 to accomplish this research work.

Declaration of Competing Interest

The authors of this study potentially declare no conflict of interest by any means.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.sjbs.2020.11.089>.

References

- Abdel-Moneim, A.S., Bamaga, M.S., Shehab, G.M., Abu-Elsaad, A.A., Farahat, F.M., 2012. HCV infection among Saudi population: high prevalence of genotype 4 and increased viral clearance rate. *PLoS One* 7, e29781.
- Akbar, H.O., Al Ghamdi, A., Qattan, F., Fallatah, H.I., Al Rumani, M., 2012. Chronic hepatitis C in Saudi Arabia: three years local experience in a university hospital. *Hepat. Mon.* 12, e6178.
- Al-Faleh, F.Z., 2003. Changing pattern of hepatitis viral infection in Saudi Arabia in the last two decades. *Ann. Saudi. Med.* 23, 367–371.
- al Nasser, M.N., 1992. 'Intrafamilial transmission of hepatitis C virus (HCV): a major mode of spread in the Saudi Arabia population'. *Ann. Trop. Paediatr.*, 12: 211-5.
- Aljowaie, R.M., Almajhdi, F.N., Ali, H.H., El-Wetidy, M.S., Shier, M.K., 2020. Inhibition of hepatitis C virus genotype 4 replication using siRNA targeted to the viral core region and the CD81 cellular receptor. *Cell Stress Chaperones* 25, 345–355.
- Baclig, M.O., Chan, V.F., Ramos, J.D., Gopez-Cervantes, J., Natividad, F.F., 2010. Correlation of the 5'untranslated region (5'UTR) and non-structural 5B (NS5B) nucleotide sequences in hepatitis C virus subtyping. *Int. J. Mol. Epidemiol. Genet.* 1, 236–244.
- Bawazir, A., AlGusheiri, F., Jradi, H., AlBalwi, M., Abdel-Gader, A.G., 2017. Hepatitis C virus genotypes in Saudi Arabia: a future prediction and laboratory profile. *Virology* 14, 208.
- Borgia, S.M., Hedskog, C., Parhy, B., Hyland, R.H., Stamm, L.M., Brainard, D.M., Subramanian, M.G., McHutchison, J.G., Mo, H., Svarovskaia, E., Shafraan, S.D., 2018. Identification of a Novel Hepatitis C Virus Genotype From Punjab, India: Expanding Classification of Hepatitis C Virus Into 8 Genotypes. *J. Infect. Dis.* 218, 1722–1729.
- Dietz, J., Susser, S., Vermehren, J., Peiffer, K.H., Grammatikos, G., Berger, A., Ferenci, P., Buti, M., Mullhaupt, B., Hunyady, B., Hinrichsen, H., Mauss, S., Petersen, J., Buggisch, P., Felten, G., Huppe, D., Knecht, G., Lutz, T., Schott, E., Berg, C., Spengler, U., von Hahn, T., Berg, T., Zeuzem, S., Sarrazin, C., H. C. V. Resistance Study Group European, 2018. 'Patterns of Resistance-Associated Substitutions in Patients With Chronic HCV Infection Following Treatment With Direct-Acting Antivirals'. *Gastroenterology*, 154: 976-88 e4.
- El-Tahan, R.R., Ghoneim, A.M., Zaghoul, H., 2018. '5' UTR and NS5B-based genotyping of hepatitis C virus in patients from Damietta governorate, Egypt. *J. Adv. Res.* 10, 39–47.
- El Awady, M.K., Azzazy, H.M., Fahmy, A.M., Shawky, S.M., Badreldin, N.G., Yossef, S. S., Omeran, M.H., Zekri, A.R., Goueli, S.A., 2009. Positional effect of mutations in 5'UTR of hepatitis C virus 4a on patients' response to therapy. *World J. Gastroenterol.* 15, 1480–1486.
- Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J.C., Strazzer, A., Chien, D.Y., Munoz, S.J., Balestrieri, A., Purcell, R.H., Alter, H.J., 2000. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* 288, 339–344.
- Farci, P., Strazzer, R., Alter, H.J., Farci, S., Degioannis, D., Coiana, A., Peddis, G., Usai, F., Serra, G., Chessa, L., Diaz, G., Balestrieri, A., Purcell, R.H., 2002. Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. *Proc. Natl. Acad. Sci. USA* 99, 3081–3086.
- Fernandez-Arcas, N., Lopez-Siles, J., Trapero, S., Ferraro, A., Ibanez, A., Orihuela, F., Maldonado, J., Alonso, A., 2006. High prevalence of hepatitis C virus subtypes 4c and 4d in Malaga (Spain): phylogenetic and epidemiological analyses. *J. Med. Virol.* 78, 1429–1435.
- Fourati, S., Feld, J.J., Chevaliez, S., Luhmann, N., 2018. Approaches for simplified HCV diagnostic algorithms. *J. Int. AIDS Soc.* 21 Suppl 2(Suppl 2), e25058.
- Franco, S., Tural, C., Clotet, B., Martinez, M.A., 2007. Complete nucleotide sequence of genotype 4 hepatitis C viruses isolated from patients co-infected with human immunodeficiency virus type 1. *Virus Res.* 123, 161–169.
- Friebe, P., Lohmann, V., Krieger, N., Bartenschlager, R., 2001. Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *J. Virol.* 75, 12047–12057.

- Gower, E., Estes, C., Blach, S., Razavi-Shearer, K., Razavi, H., 2014. Global epidemiology and genotype distribution of the hepatitis C virus infection. *J. Hepatol.* 61, S45–S57.
- Hara, K., Rivera, M.M., Koh, C., Sakiani, S., Hoofnagle, J.H., Heller, T., 2013. Important factors in reliable determination of hepatitis C virus genotype by use of the 5' untranslated region. *J. Clin. Microbiol.* 51, 1485–1489.
- Hedskog, C., Parhy, B., Chang, S., Zeuzem, S., Moreno, C., Shafraan, S.D., Borgia, S.M., Asselah, T., Alric, L., Abergel, A., Chen, J.J., Collier, J., Kapoor, D., Hyland, R.H., Simmonds, P., Mo, H., Svarovskaia, E.S., 2019. Identification of 19 Novel Hepatitis C Virus Subtypes-Further Expanding HCV Classification. *Open Forum Infect. Diseases* 6 (3), ofz076.
- Hemeida, A.A.O.M., El-Shahat, M., Hashem, M.H., Mahmoud, A., Dahi, H., 2011. Genetic variations in a conserved 5'-untranslated region of hepatitis C virus isolated from Egypt. *Int. J. Virol.*, 7:91-9.
- Hmaied, F., Legrand-Abbravanel, F., Nicot, F., Garrigues, N., Chapuy-Regaud, S., Dubois, M., Njouom, R., Izopet, J., Pasquier, C., 2007. Full-length genome sequences of hepatitis C virus subtype 4f. *J. Gen. Virol.* 88, 2985–2990.
- Khaliq, S., Jahan, S., Pervaiz, A., Ali Ashfaq, U., Hassan, S., 2011. Down-regulation of IRES containing 5'UTR of HCV genotype 3a using siRNAs. *Virol. J.* 8, 221.
- Kuntzen, T., Beral, A., Ndjoum, J., Bennett, P., Schneidewind, A., Lennon, N., Birren, B.W., Kuiken, C., Henn, M.R., Simmonds, P., Allen, T.M., 2008. A set of reference sequences for the hepatitis C genotypes 4d, 4f, and 4k covering the full open reading frame. *J. Med. Virol.* 80, 1370–1378.
- Legrand-Abbravanel, F., Nicot, F., Boulestin, A., Sandres-Saune, K., Vinel, J.P., Alric, L., Izopet, J., 2005. Pegylated interferon and ribavirin therapy for chronic hepatitis C virus genotype 4 infection. *J. Med. Virol.* 77, 66–69.
- Li, C., Lu, L., Wu, X., Wang, C., Bennett, P., Lu, T., Murphy, D., 2009. Complete genomic sequences for hepatitis C virus subtypes 4b, 4c, 4d, 4g, 4k, 4l, 4m, 4n, 4o, 4p, 4q, 4r and 4t. *J. Gen. Virol.* 90, 1820–1826.
- Lohmann, V., Körner, F., Koch, J., Herian, U., Theilmann, L., Bartenschlager, R., 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science (New York, N.Y.)* 285 (5424), 110–113.
- Manns, M.P., Buti, M., Gane, E., Pawlotsky, J.M., Razavi, H., Terrault, N., Younossi, Z., 2017. Hepatitis C virus infection. *Nat. Rev. Disease primers* 3, 17006.
- McPhee, F., 2019. Developments in the treatment of HCV genotype 3 infection. *Exp. Rev. Anti-infective Therapy* 17 (10), 775–785.
- Messina, J.P., Humphreys, I., Flaxman, A., Brown, A., Cooke, G.S., Pybus, O.G., Barnes, E., 2015. Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology* 61, 77–87.
- Moratorio, G., Martinez, M., Gutierrez, M.F., Gonzalez, K., Colina, R., Lopez-Tort, F., Lopez, L., Recarey, R., Schijman, A.G., Moreno, M.P., Garcia-Aguirre, L., Manascero, A.R., Cristina, J., 2007. Evolution of naturally occurring 5'non-coding region variants of Hepatitis C virus in human populations of the South American region. *Virol. J.* 4, 79.
- Penin, F., Dubuisson, J., Rey, F.A., Moradpour, D., Pawlotsky, J.M., 2004. Structural biology of hepatitis C virus. *Hepatology (Baltimore, Md.)* 39 (1), 5–19.
- Ross, R.S., Viazov, S., Roggendorf, M., 2007. Genotyping of hepatitis C virus isolates by a new line probe assay using sequence information from both the 5'untranslated and the core regions. *J. Virol. Methods* 143, 153–160.
- Sandres-Saune, K., Deny, P., Pasquier, C., Thibaut, V., Duverlie, G., Izopet, J., 2003. Determining hepatitis C genotype by analyzing the sequence of the NS5b region. *J. Virol. Methods* 109, 187–193.
- Shahid, I., Gull, S., Ijaz, B., Ahmad, W., Ansar, M., Asad, S., Kausar, H., Sarwar, M.T., Khan, M.K., Hassan, S., 2013. Stable Huh-7 cell lines expressing non-structural proteins of genotype 1a of hepatitis C virus. *J. Virol. Methods* 189, 65–69.
- Shier, M.K., El-Wetidy, M.S., Ali, H.H., Al-Qattan, M.M., 2014. Characterization of hepatitis C virus genotypes by direct sequencing of HCV 5'UTR region of isolates from Saudi Arabia. *PLoS One* 9, e103160.
- Shier, M.K., Iles, J.C., El-Wetidy, M.S., Ali, H.H., Al Qattan, M.M., 2017. Molecular characterization and epidemic history of hepatitis C virus using core sequences of isolates from Central Province, Saudi Arabia. *PLoS One* 12, e0184163.
- Smith, D.B., Bukh, J., Kuiken, C., Muerhoff, A.S., Rice, C.M., Stapleton, J.T., Simmonds, P., 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. *Hepatology* 59, 318–327.
- Simmonds, P., Becher, P., Bukh, J., Gould, E. A., Meyers, G., Monath, T., Muerhoff, S., Pletnev, A., Rico-Hesse, R., Smith, D. B., Stapleton, J. T., Ictv Report Consortium 2017. *ICTV Virus Taxonomy Profile: Flaviviridae*. *J. General Virol.*, 98(1), 2–3.
- Timm, J., Neukamm, M., Kuntzen, T., Kim, A.Y., Chung, R.T., Brander, C., Lauer, G.M., Walker, B.D., Allen, T.M., 2007. Characterization of full-length hepatitis C virus genotype 4 sequences. *J. Viral Hepat.* 14, 330–337.
- Verbeeck, J., Stanley, M.J., 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.
- Virtanen, E., Mannonen, L., Lappalainen, M., Auvinen, E., 2018. Genotyping of hepatitis C virus by nucleotide sequencing: A robust method for a diagnostic laboratory. *MethodsX* 5, 414–418.
- Vopalensky, V., Khawaja, A., Roznovsky, L., Mrazek, J., Masek, T., Pospisek, M., 2018. Characterization of Hepatitis C Virus IRES Quasispecies - From the Individual to the Pool. *Front. Microbiol.* 9, 731.
- Zekri, A.R., Bahnassy, A.A., Hafez, M.M., Hassan, Z.K., Kamel, M., Loutfy, S.A., Sherif, G.M., El-Zayadi, A.R., Daoud, S.S., 2011. Characterization of chronic HCV infection-induced apoptosis. *Comp. Hepatol.* 10, 4.
- Zekri, A.R., El-Din, H.M., Bahnassy, A.A., Khaled, M.M., Omar, A., Fouad, I., El-Hefnewi, M., Thakeb, F., El-Awady, M., 2007. Genetic distance and heterogeneity between quasispecies is a critical predictor to IFN response in Egyptian patients with HCV genotype-4. *Virol. J.* 4, 16.