MOLECULAR EPIDEMIOLOGY OF HEPATITIS C IN KELANTAN

NUR AMALIN ZAHIRAH BINTI MOHD AMIN

UNIVERSITI SAINS MALAYSIA

2024

MOLECULAR EPIDEMIOLOGY OF HEPATITIS C IN KELANTAN

by

NUR AMALIN ZAHIRAH BINTI MOHD AMIN

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

July 2024

ACKNOWLEDGEMENT

First and foremost, praise be to the Almighty Allah for all His blessings for giving me patience and good health throughout this post-graduate study.

I am lucky to have Associate Professor Dr Rafidah Hanim Shueb as my supervisor. I would like to express my wholehearted gratitude and appreciation for her continuous guidance and motivation throughout the completion of this study. Her patience, kindness, immense knowledge, and valuable suggestions have enormously helped me during the process of completing this master study.

I would like to express my deepest appreciation to my co-supervisor Associate Professor Dr Nazri Mustaffa for his support and guidance. A special thanks to their team; Dr Nur Izat, Dr Vincent Tee and Dr Nur Azlin for their assistance and also, my gratitude to Dr Siti Azrin for helping me regarding the statistical consultations.

My special words of thanks and appreciation go to Tuan Nur Akmalina – my best friend; and my fellow friends - Hanifah Fazlin, Alia, Bello Kizito, Ahmad Irekeola, Norhidayah, Yasmin Khairani, and Nik Zuraina for their great cooperation in helping me throughout the journey. Without all the motivational words of wisdom and helping hands, I would not be able to stand strong until the finishing line.

Most notably, nobody has been more important to me in the completion of this project than the precious members of my family. I would like to thank my parents, Mohd Amin bin Ismail and Zahawiah binti Yaakub, whose infallible love and endless support are with me in whatever I pursue. Lastly, I also wanted to thank those who have contributed and been of assistance directly or indirectly from the beginning until the completion of this dissertation. May the Almighty God prosperously bless all of you.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

α	Alpha
amps	Amperes
bp	Base pair
°C	Degree Celcius
IU/mL	International units per mililitre
µL	Microlitre
µM	Micromolar
mL	Mililitre
min	Minutes
n	Numbers
%	Percentage
®	Registered trademark
s	Seconds
x g	Times gravity units
™	Unregistered trademark
V	Voltage

LIST OF ABBREVIATIONS

ALT	Alanine Transaminase
AST	Aspartate Aminotransferase
BLAST	Basic Local Alignment Search Tool
BOC	Boceprevir
CDC	Centre for Disease Controls
cDNA	complimentary Deoxyribonucleic Acid
CHC	Chronic Hepatitis C
CKD	Chronic Kidney Disease
DAAs	Direct-acting Antivirals
DAV	Dasabuvir
DCV	Daclatasvir
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic Acid
dNTP	deoxynucleotide Triphosphates
EASL	European Association for the Study of the Liver
EBV	Elbasvir
EDTA	Ethylene-diamine-tetra-acetic acid
ELISA	Enzyme-linked Immunosorbent Assay
EMA	European Medicines Agency
ER	Endoplasmic Reticulum
FDA	Food and Drug Administration
GFR	Glomerular Filtration Rate
GT	Genotype
GZP	Grazoprevir
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HCC	Hepatocellular Carcinoma
HIV	Human Immunodeficiency Virus
HREC	Human Research Ethics Committee
HRPZ	Hospital Raja Perempuan Zainab
HUSM	Hospital Universiti Sains Malaysia

IFN	Interferon
IFN-α	Interferon alfa
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IQR	Interquartile Range
IVDU	Intravenous Drug User
LDV	Ledipasvir
LOD	Limit of Detection
MEGA	Molecular Evolutionary Genetic Analysis
MOH	Ministry of Health
MREC	Medical Research and Ethics Committee
NADA	National Anti-Drug Agency
NAT	Nucleic Acid Testing
NCBI	National Centre for Biotechnology Information
OBV	Ombitasvir
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PEG-IFN	Pegylated-Interferon
PIs	Protease Inhibitors
PTV	Paritaprevir
PWID	People Who Inject Drug
RBV	Ribavirin
RDT	Rapid Diagnostic Test
RdRP	RNA-dependent RNA Polymerase
RNA	Ribonucleic Acid
RT-LAMP	Reverse Transcription Loop-Mediated Isothermal Amplification
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SD	Standard Deviation
SMV	Simeprevir
SOC	Standard of Care
SOF	Sofosbuvir
SPSS	Statistical Package for Social Sciences
STDs	Sexually Transmitted Diseases

SVR	Sustained Virologic Response
TBE	Tris-Borate EDTA
TLV	Telaprevir
UK	United Kingdom
USA	United States of America
UTR	Untranslated Region
UV	Ultraviolet
VEL	Velpatasvir
VOX	Voxilaprevir
WHO	World Health Organization

LIST OF APPENDICES

EPIDEMIOLOGI MOLEKULAR HEPATITIS C DI KELANTAN

ABSTRAK

Hepatitis C merupakan ancaman kesihatan awam utama yang menjejaskan berjuta-juta individu di seluruh dunia. Hepatitis C kronik adalah salah satu punca utama penyakit hati kronik di seluruh dunia, yang boleh menjadi sirosis hati dan kanser hepatoselular jika tidak dirawat dengan betul. Penemuan genotip dan subjenis virus Hepatitis C (HCV) yang baru adalah bukti perkembangan genetik virus yang berterusan. Genotip HCV yang menyebabkan jangkitan adalah penting untuk perubatan dan pengurusan yang berkesan. Oleh itu, saringan dan kuantifikasi HCV yang mencukupi adalah penting untuk diagnosis dan pengurusan penyakit yang betul. Oleh itu, kajian ini bertujuan untuk menentukan kelaziman genotip HCV dan faktor risiko yang berkaitan dengan jangkitan HCV di Kelantan serta untuk mewujudkan ujian tindak balas rantai polimerase terbalik-transkriptase bersarang dalaman (RT-PCR) untuk pengesanan HCV di Kelantan. Kajian keratan rentas ini dijalankan di Makmal Mikrobiologi dengan penyertaan pesakit yang diambil dari Hospital Universiti Sains Malaysia (HUSM) dan Hospital Raja Perempuan Zainab 2 (HRPZ) dari Disember 2019 sehingga Januari 2022. Hasil kajian ini menunjukkan 56 daripada 80 pesakit dikesan positif HCV oleh RT-PCR bersarang. Sebanyak tiga genotip ditemui dengan genotip paling kerap ialah genotip HCV 3 dan 1 dengan perkadaran yang sama iaitu 44.6% (25) setiap satu dan 1.8% (1) daripada genotip 6. Bagi genotip 3, dua subjenis didapati dengan 41% (23) daripada subjenis 3a, diikuti oleh 3.6% (2) daripada subjenis 3b. Bagi genotip 1, kebanyakan subjenis adalah daripada subjenis 1a, 42.9% (24) dan hanya 1.8% (1) daripada subjenis 1b diperhatikan. Satu-satunya pesakit dengan genotip 6 didapati subjenis 6n, 1.8% (1).

Walau bagaimanapun, terdapat lima (8.9%) genotip campuran ditemui dalam kajian ini iaitu: 3a/1a (5.4%) (3), 3b/1a (1.8%) (1), 1a/3a (1.8%) (1). Daripada kajian ini, terdapat perkaitan yang ketara dalam pekerjaan mengikut jantina pesakit HCV. Walau bagaimanapun, tiada perkaitan ditemui dalam umur dan status perkahwinan berdasarkan jantina. Terdapat juga perkaitan yang ketara antara pesakit lelaki dengan faktor risiko penagih dadah melalui suntikan (IVDU) dan pemindahan darah. Walau bagaimanapun, tiada perkaitan ditemui antara genotip dengan faktor risiko penularan HCV. Kesimpulannya, kedua-dua genotip 3 dan 1 adalah genotip HCV yang paling biasa ditemui dalam populasi kajian ini. Tambahan pula, faktor risiko tertinggi penularan HCV ialah IVDU dan pemindahan darah. Ringkasnya, penemuan ini boleh menjadi panduan untuk kajian epidemiologi pada masa hadapan tentang kelaziman dan penularan HCV di Kelantan.

MOLECULAR EPIDEMIOLOGY OF HEPATITIS C IN KELANTAN

ABSTRACT

Hepatitis C is a major public health threat that affects millions of individuals globally. Chronic Hepatitis C is one of the primary causes of chronic liver disease worldwide, which can develop to liver cirrhosis and hepatocellular cancer if not treated efficiently. The persistent finding of novel Hepatitis C virus (HCV) genotypes and subtypes is the proof of the virus's continued genetic development. The HCV genotype that causes the infection is critical for effective medication and management. Therefore, adequate HCV screening and quantification are critical for the diagnosis and proper management of the disease. Thus, this study aims to determine the prevalence of HCV genotypes and risk factors associated with HCV infection in Kelantan as well as to establish an in-house nested reverse-transcriptase polymerase chain reaction (RT-PCR) assay for HCV detection in Kelantan. This cross-sectional study was conducted at Microbiology Laboratory with the participation of patients recruited from Hospital Universiti Sains Malaysia and Hospital Raja Perempuan Zainab 2 from December 2019 until January 2022. The result of this study showed that 56 out of 80 patients were detected HCV positive by nested RT-PCR. A total of three genotypes were found with the most frequent genotype were HCV genotype 3 and 1 with a similar proportion of 44.6% (25) each and 1.8% (1) from genotype 6. For genotype 3, two subtypes were found with 41% (23) of subtype 3a, followed by 3.6% (2) of subtype 3b. For genotype 1, most subtypes were from subtype 1a, 42.9% (24) and only 1.8% (1) from subtype 1b was observed. The only participant with genotype 6 was found to be subtype 6n, 1.8% (1). However, there were five (8.9%) mixed genotypes discovered in this study

which were: 3a/1a (5.4%) (3), 3b/1a (1.8%) (1), 1a/3a (1.8%) (1). From this study, there was a significant association in occupation according to gender of HCV patients. However, no association found in age and marital status based on gender. There was also significant association between male patients with risk factors of intravenous drug user (IVDU) and blood transfusion. However, no association found between the genotypes with the risk factors of HCV transmission. In conclusion, both of genotypes 3 and 1 were the most common HCV genotypes found in this study population. Furthermore, the highest risk factors of HCV transmission were IVDUs and blood transfusion. In sum, these findings may guide future epidemiological studies on HCV prevalence and transmission in Kelantan.

CHAPTER 1

INTRODUCTION

1.1 Background of study

Hepatitis C virus (HCV) is a major global disease caused by liver infection which eventually leads to chronic hepatitis C (CHC) such as liver cirrhosis and fibrosis (Thong et al., 2014). *Hepacivirus* is a genus of Hepatitis C virus that belongs to the member of *Flaviviridae* (Coppola et al., 2019). Globally, HCV has chronically infected approximately 71 million individuals worldwide which is equivalent to 1% of the world population (WHO, 2017a). Meanwhile, in Southeast Asia, at least 4.7 million individuals may be infected with HCV (Blach et al., 2017). Recently, the HCV infection prevalence in Malaysia is reported to be 0.4%, which corresponds to 90,119 individuals above 15 years old (Muhammad et al., 2023).

According to Smith et al. (2018), HCV has a high genetic diversity with eight confirmed main genotypes (1, 2, 3, 4, 5, 6, 7, 8) and 93 subtypes so far. The global distribution of HCV genotypes varies geographically (Coppola et al., 2019). The major HCV genotypes that are widely distributed are genotypes 1 and 3 which contribute 46% and 30% of the HCV infections worldwide, respectively (Coppola et al., 2019). Genotypes 1 and 2 are common in West Africa, while genotype 3 is highly prevalent in South Asia. Conversely, genotype 4 is prevalent in Central Africa and the Middle East, genotype 5 in Southern Africa and genotype 6 in Southeast Asia (Messina et al., 2015). Besides, genotype 7 has been reported in the Republic of Congo in Central Africa while genotype 8 is recognised in India (Borgia et al., 2018; Murphy et al., 2015).

Over the decades, a persistent HCV infection has harmed the liver, increasing the risk of cirrhosis and hepatocellular carcinoma (HCC) in HCV patients (Thrift et al., 2017). According to Llovet et al. (2012), liver cancer ranks sixth in terms of the most frequent cancer, and more than 90% of primary liver cancers are HCC. A higher risk of implications including fibrosis, HCC, fatty liver disease and fatality is related to genotype 3 (Shahnazarian et al., 2018). Furthermore, more severe liver damage and an aggressive course of infection are reported to be related to genotype 1b infection in patients with chronic HCV infection than with other HCV genotypes (Zein, 2000).

The vast genetic diversity of HCV poses difficulties for the regulation of the host immune system, the treatment of HCV-infected patients, and the establishment of pan-genotypic therapies (Timm & Roggendorf, 2007). The management of HCV infection has improved by replacing pegylated interferon alpha (IFNα) and ribavirin with the new combination of anti-HCV therapy based on direct-acting antivirals (DAAs) (Kumar et al., 2018). DAA-based interferon-free therapy offers great opportunities for long-term HCV eradication and may halt the disease's progression (Coppola et al., 2019). DAAs are referred to as pan-genotypic antivirals when they demonstrate a high level of therapeutic efficacy against all HCV genotypes (WHO, 2018). Various HCV genotypes have been shown to react differently to anti-HCV drug treatment (Sorbo et al., 2018). Hence, in addition to indicating the likelihood of disease progression, the genotype is also crucial for predicting treatment response (Ge et al., 2009).

Effective therapy and management of HCV relies on the HCV genotype that causing the infection. Thus, appropriate HCV screening and quantification are crucial for the diagnosis and proper management of the disease. HCV genotyping is also an essential technique for assessing patients' response to antiviral medication and for estimating the treatment period (Warkad et al., 2018). The most recommended method for HCV genotyping is sequencing analysis of the highly conserved regions such as 5'-UTR, E1, Core and NS5 (Karabulut et al., 2018). HCV genome sequencing and phylogenetic analysis of the Core and NS5B region by "in-house" procedures are considered as the gold standard and the most recommended technique for HCV genotyping since it is capable of precisely distinguishing between genotypes (Chevaliez & Pawlotsky, 2006b; Tagnouokam-Ngoupo et al., 2019). This method must be applied in molecular epidemiology research where precise subtyping is required. Meanwhile in clinical settings, different commercial tests are available for detecting HCV genotypes. Despite the advanced commercial assays, the genotyping methods based on sequencing technology are still superior to other methods as it allows both genotype and subtype to be identified simultaneously as well as resistance-associated mutations (Yang & Wei, 2018). Although this assay requires various experimental processes, HCV genotyping by direct sequencing has become more accessible and easier with the ongoing development of biochemical reagents and experimental procedures (Cai et al., 2013).

.

1.2 Problem statement

At present, hepatitis C has become a public health concern as it is a growing problem in Malaysia. According to Raihan (2016), the most common HCV genotype in Malaysia belongs to genotypes 1 and 3. Seventy-three percent of HCV in Malaysia belongs to genotype 3; consists of subtype 3a (64.8%) and type 3 (8.1%). This is followed by genotype 1 at 27% with subtype 1a (13.5%), 1b (2.7%) and type 1 (10.8%) (Mohamed et al., 2013). The ongoing discovery of new HCV genotypes and subtypes is evidence of the virus's continuous genetic evolution. This continues to raise concerns in the effort to eradicate hepatitis C by 2030 which is in line with the National Strategic Plan for Hepatitis B and C 2019–2023 by the Malaysian Ministry of Health (MOH, 2019). Therefore, early HCV diagnosis and accurate genotyping are essential for effective patient management as well as to control the infection.

Kelantan is one of the top five states in Malaysia with a high incidence rate of hepatitis C (MOH, 2020). Currently, there is a lack of data concerning the prevalence of HCV as well as viral genotypes circulating in Kelantan. The predominating of HCV genotypes in Kelantan might be slightly different from other states since HCV genotypes are varied and diverse based on geographical regions. The accuracy in HCV genotype classification is important for the proper implementation of treatment. Therefore, this study aims to determine the HCV genotypes circulating in Kelantan as the genotype information will help to establish local data and may assist clinicians in determining the best management strategies for HCV patients.

1.3 Research questions

i) What are the HCV genotypes circulating within the study population?

ii) What are the risk factors associated with HCV patients in the study population?

iii) What is the genetic relatedness of HCV strains in Kelantan, and is it more closely aligned with strains found in Malaysia or Thailand?

1.4 Research hypothesis

i) H_o: There is no association between demographic characteristics of HCV patients with risk factors of HCV.

ii) H_a: There is an association between demographic characteristics of HCV patients with risk factors of HCV.

i) H_o: The diversity of HCV genotypes does not influence the risk factors associated with HCV within the study population.

ii) H_a: The diversity of HCV genotypes influences the risk factors associated with HCV within the study population.

1.5 Objectives of study

1.5.1 General objective

The main purpose of this study is to evaluate the diversity of genotypes and associated risk factors among patients with HCV infection in Kelantan.

1.5.2 Specific objectives

The specific objectives of this study are:

1. To establish an in-house optimised nested RT-PCR assay for HCV detection in Kelantan.
2. To determine genotypes and genetic relatedness of HCV circulating among patients in Kelantan.
3. To determine the association between risk factors and genotypes among patients with HCV infection in Kelantan.

CHAPTER 2

LITERATURE REVIEW

2.1 Epidemiology of HCV

HCV is a common and widespread viral infection that infects approximately 120 - 180 million people, roughly 2 - 3% of the human population globally (Pybus et al., 2009). According to currently available statistics, acute HCV infection was responsible for 54,000 fatalities and an estimated 290,000 hepatitis C-related deaths in 2019 were due to cirrhosis and hepatocellular carcinoma (HCC) (Mohd Hanafiah et al., 2013; WHO, 2023). According to estimates, 3 – 4 million people become infected each year, 170,000,000 people have chronic infection and are at greater risk of developing liver diseases such as cirrhosis and liver cancer, and 350,000 people die each year from all HCV-related causes (Mohd Hanafiah et al., 2013).

The distribution of HCV differs significantly across the regions (Bukh, 2016). The prevalence of HCV in economically developed countries (United Kingdom: 0.4%) is below 2% while higher prevalence is observed in less developed countries (Egypt: 4.4 - 15%, Gabon: 4.9 - 11.2%, Uzbekistan: 11.3%, Cameroon: 4.9 - 13.8%, Mongolia: 9.6 – 10.8%, Pakistan: 6.8%, Nigeria: 3.1 – 8.4% and Georgia: 6.7%) (Parsons, 2022). Another literature reported that in Western regions and Australia, the rate varies between 0.5 – 1.5%. In contrast, it rises to 2.3% in Southeast Asia and Eastern Mediterranean regions, 2.2% in Indonesia, 3.2% in China, 6.5% in Pakistan and 0.9% in India (Petruzziello et al., 2016; WHO, 2017a).

Global HCV infection rates are compared using the prevalence of HCV antibodies from population-based research. According to the available data, anti-HCV

prevalence rates are generally less than 2.5% across America, Western Europe and Southeast Asia. The average anti-HCV incidence rates for Eastern Europe range from 1.5 to 5%, around 2.5 to 4.9% for the Western Pacific region, and from 1% to more than 12% for the Middle East and Central Asia (Lavanchy, 2011).

HCV is highly prevalent in Southeast Asian regions, where there are approximately 94.6 million infected people (Doan, 2018). According to WHO, (2017b), an estimated 67,938 (21%) and 17,370 (6%) deaths from hepatitis C virus in the Southeast Asia region have been attributed to cirrhosis and liver cancer, respectively. In published data reported by Gower et al., (2014), the HCV prevalence based on viraemic rate (RNA-positive) in the infected adult population of Southeast Asia revealed Cambodia has the highest viraemic rate which accounted for 75.8% followed by Malaysia (71.9%), Vietnam (68.0%), Indonesia (65.7%), Thailand (62.7%) and Myanmar (55.6%).

In Malaysia, the prevalence of HCV infection has been rising as more individuals are being identified as having HCV antibodies through routine screening with an estimated 1.9% tested positive for anti-HCV (Md Said et al., 2020). The prevalence of HCV in Malaysia was predicted to be 2.5%, equivalent to 453,700 individuals in the population above 15 years old with 2.9% of Malays, 1.1% of Chinese and 0.6% of Indian/another ethnicity (McDonald et al., 2014). The Malaysian Ministry of Health (MOH) received reports of 23,112 hepatitis C cases from 2003 to 2017. In 2016, the report rate reached an elevated level of 11.0 per 100,000 and slightly declined to 9.54 per 100,000 in 2017 (MOH, 2019). A recent study in Malaysia reported that the prevalence of HCV infection was 0.4%, corresponding to

approximately more than 90,000 individuals above the age of 15 while 51,675 of them had a chronic infection with a 0.2% frequency (Muhammad et al., 2023). It is predicted that the number of chronic hepatitis C cases will increase further and hit 523,500 cases in 2039 if there are no improvements in terms of prevention, control and treatment approaches regarding the hepatitis epidemic (MOH, 2019).

2.2 Hepatitis C virus

2.2.1 Virological characteristics

HCV is a member of the genus *Hepacivirus* and from the family *Flaviviridae* which comprises a small, encapsulated virus with a single-stranded positive RNA genome (Figure 2.1). The HCV genome is 9,600 kb in length with three different regions consisting of two untranslated regions (5'-UTR and 3'-UTR) with a single open reading frame (ORF) in the middle which encodes a single polyprotein of 3000 amino acids (Li & Lo, 2015). The polyprotein is translated and processed into three structural proteins: core, E1 and E2 and seven non-structural proteins which are NS1/p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B (Figure 2.2) (Bukh, 2016).

HCV has a lot of different virologic features while sharing some characteristics with other *Flaviviridae* species, particularly in terms of the replication mechanism. The HCV 5'-UTR is the most conserved region made up of 341-344 nucleotides, while in the 3'-UTR there are about 225 nucleotides (Chevaliez & Pawlotsky, 2006a).

The core protein which makes up the primary component of HCV nucleo-capsid is a highly conserved basic protein (Krekulová et al., 2006). In addition, the core protein has been proposed to directly interact with several cellular proteins and processes that may be significant in the viral life cycle (Chevaliez & Pawlotsky, 2006a). The viral entrance and membrane fusion requires the two envelope proteins (E1 and E2) to function together. E2 is essential for viral attachment since it starts the commencement of the interaction with the receptor complex.

NS1/p7 is a small integral membrane protein that is highly hydrophobic and consist of 63 amino acid polypeptides (Chevaliez & Pawlotsky, 2006a; Krekulová et al., 2006). The p7 protein comprises two transmembrane domains arranged in alpha-helices connected by a cytoplasmic loop and appears to be vital for infection (Chevaliez & Pawlotsky, 2006a; Li & Lo, 2015). NS2 is a non-glycosylated transmembrane protein with a molecular weight of 21-23 kDA that is accountable for the association of endoplasmic reticulum (ER) membrane (Chevaliez & Pawlotsky, 2006a). The NS2 protein interacts with the N-terminus of the NS3 protein to generate a protease that modifies the precursor polyprotein (Du & Tang, 2016). HCV NS3 is a two-domain protein similar to other flaviviruses, with a molar mass of 70 kDa (Krekulová et al., 2006).

The NS4A protein with a molar mass of 8 kDa has a hydrophobic area that is anticipated to interact with membranes and other replicase components (Krekulová et al., 2006). The NS4A serves as a cofactor of the protease activity of NS3 and the NS3/NS4A interaction bears an important role in the HCV lifecycle (Chevaliez & Pawlotsky, 2006a). NS4B is an integral membrane and hydrophobic protein with ≈30 kDa molar mass that consists of 261 amino acids located within the ER and is crucial for membrane association (Chevaliez & Pawlotsky, 2006a; Krekulová et al., 2006).

NS5A, a hydrophilic phosphoprotein with a molecular weight of 56-58 kDa, is responsible for virus replication and cellular regulatory pathways (Chevaliez & Pawlotsky, 2006a; Krekulová et al., 2006). Aside from its role in HCV genome replication, NS5A sparked initial interest due to its possible role in modulating the interferon response (Dubuisson & Fabiani, 2007).

A 65 kDa protein called NS5B is an RNA-dependent RNA polymerase (RdRP) that generates new genomic RNAs (Ashfaq et al., 2011). The NS5B coding region varies greatly among different HCV strains. The high heterogeneity in NS5B was used to create an approach for classifying HCV strains into genotypes and subtypes (Krekulová et al., 2006). The RdRP is also a significant target for the development of anti-HCV drugs (Ashfaq et al., 2011; Du & Tang, 2016).

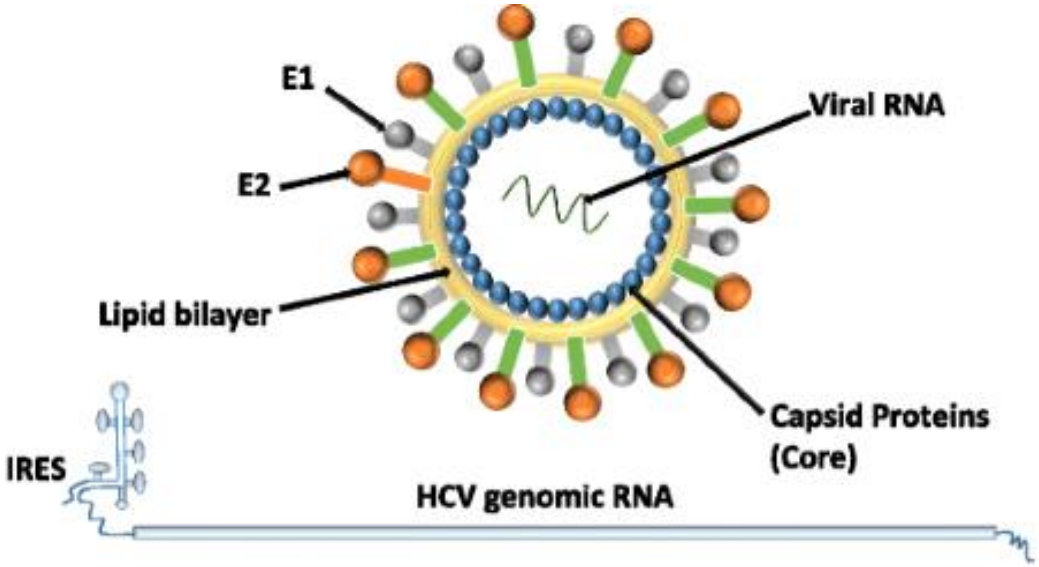

Figure 2.1 A model structure of HCV. (Adopted from Elberry et al., (2017)).

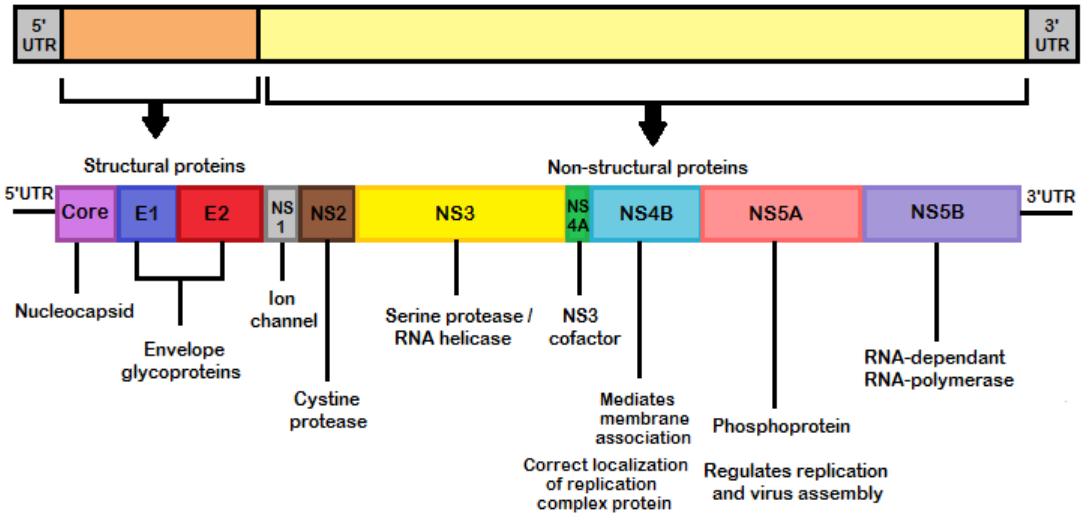

Figure 2.2 The diagram of hepatitis C virus (HCV) genome. The HCV genome has an open reading frame (ORF) in the centre and two untranslated regions (UTR) at the 5' and the 3' end. (Adopted from Mohd Amin et al., (2023)).

2.3 HCV genotypes and their prevalence

HCV has great genetic variability, which presents at various scales among viral populations in infected patients at any given time and during evolution (quasispecies), and globally among isolates from different patients in terms of genotypes, subtypes, and isolates/strains (Bukh, 2016). The number of confirmed genotypes and subtypes has dramatically increased due to advancements in sequence analysis tools with the existence of eight confirmed genotypes (genotypes 1, 2, 3, 4, 5, 6, 7 and 8) and 93 distinct subtypes thus far (Table 2.1) (Smith et al., 2018). Different HCV genotypes have diverse geographic distributions and respond differently to antiviral treatments (Jia et al., 2023). The genetic sequences of the genotypes varied by 30-35%, whereas the subtypes differed by less than 15% (Messina et al., 2015). As the virus continues to mutate, the number of genotypes and subtypes will almost certainly increase in the coming years.

Generally, genotypes 1, 2 and 3 can be detected worldwide, whereas the remaining five genotypes are typically indigenous and endemic to particular regions (Jia et al., 2023). The three major HCV genotypes (1, 2, 3), account for about 85% of HCV infections (Sorbo et al., 2018). According to Messina et al. (2015), HCV genotype 1 is the most common globally with 83.4 million instances which is equivalent to 46.2% of all HCV cases, and nearly one-third of those cases occurring in East Asia. The next most common genotype in the world (54.3 million, 30.1%) is genotype 3. While genotypes 2, 4 and 6 account for 22.8% of all instances, genotype 5 accounts for the remaining less than 1% of cases. Even though most of HCV infections globally are caused by genotypes 1 and 3, HCV strains from the uncommon genotypes 2, 4, 5, and 6 have the potential to spread quickly if they are linked to

effective transmission pathways (Messina et al., 2015). Regardless of economic standing, genotypes 1 and 3 predominate in most countries, while genotypes 4 and 5 are more prevalent in less developed countries (Messina et al., 2015).

HCV genotypes vary in geographical distribution. The most common regions for HCV genotype 1 are North and South America, Europe, and Australia, with a prevalence rate of >85% in Central Europe and Southern America (Messina et al., 2015; Stefano et al., 2023). Subtypes 1a and 1b are the most prevalent HCV subtypes in Japan, Europe and North America (Petruzziello et al., 2019). However, genotype 2 is more prevalent in West Africa and South America, most likely because of population migrations during the 1700s trans-Atlantic slave trade (Markov et al., 2012). Meanwhile, genotype 3 has a high rate of prevalence in South Asia, Australia and various European nations (Coppola et al., 2019). In Africa, the prevalence of genotype 3 cases is relatively lower than in South Asia, where genotype 3 is more common, with up to 70% of cases (Mohd Amin et al., 2023). Genotype 3 is spreading globally, most likely because of population migration from subtype 3a-dominant countries like Pakistan and India as well as the association between injection drug usage and subtype 3a (Messina et al., 2015). In the Middle East, Central Africa and East Africa, HCV genotype 4 is the most common genotype found within the regions with prevalence rates of 65.3%, 97.6% and 30.7% (Messina et al., 2015; Stefano et al., 2023).

In contrast, genotype 5 is primarily restricted to South Africa and Eastern Africa with prevalence rates at of 58.8% and 13.7%, respectively (Messina et al., 2015). On the other hand, HCV genotype 6 is only found in South China and

Southeast Asia and is sporadically detected in migrant patients from endemic nations (Thong et al., 2014). Most countries where this genotype has been identified are Vietnam, Cambodia, Laos, Myanmar, and Thailand. In these nations, the prevalence of HCV genotype 6 varies from 20% to over 50% of all genotypes (Thong et al., 2014). Genotype 6 is widely present in those mentioned countries, however, evidence of infection in adjacent countries are scarce. For example, the latest reported genotype 6, 0.5% (2) has been found in Pahang, Malaysia in 2015 by Ho et al. (2015) and another study by Ng et al. (2013) reported one subtype 6n in Kuala Lumpur, Malaysia regarding a patient with a record of intravenous drug use (IVDU) and has co-infection with HIV. Last but not least, the recently identified genotypes 7 and 8 exhibit dominance in Central Africa and the Indian subcontinent, respectively (Stefano et al., 2023). It is also currently known that a few subtypes, especially 1a, 1b, 2a, and 3a, are extensively spread around the world and account for a significant proportion of HCV infections in developed countries (Messina et al., 2015).

It is interesting to note that the rise in human migration causes progressive changes in HCV genotype distribution (Stefano et al., 2023). The movement of individuals from one region to another may spread novel HCV strains to other areas, causing shifts in the distribution of certain genotypes. There have been reports of divergent HCV genotype 1 in Germany and Cyprus (Oikonomopoulou et al., 2017; Wang et al., 2019). While numerous divergent HCV genotype 2 cases were reported in Canada, primarily with people of African descent (Li et al., 2012). Besides, four distinct subtypes of HCV genotype 4 consisting of subtypes 4d, 4l, 4r and 4v have been reported in Ethiopia (Hundie et al., 2017). The epidemiology of HCV genotype 4 is evolving. Following changes in demographics, immigration, and intravenous drug

use, it began to spread beyond its primary areas in Africa and the Middle East to various Western countries, particularly in Europe (Ciccozzi et al., 2012; De Bruijne et al., 2009). It is important to note that the prevalence of HCV genotypes can change over time due to factors such as migration, globalization, and changes in risk factors for transmission.

Table 2.1 Representative strain of HCV genotypes and subtypes.

Genotype 1			
Subtypes	**Accession number(s)**	**Geographical distribution**	**References**
1a	M62321, M67463	USA	(Choo et al., 1991; Inchauspe et al., 1991)
1b	D90208, M58335	Japan	(Kato et al., 1990; Takamizawa et al., 1991)
1c	D14853, AY051292	Indonesia, India	(Okamoto et al., 1994; Guntaka et al., unpublished 2001)
1d	KJ439768	Canada	(Lu et al., 2014)
1e	KC248194, KJ439769	Cameroon, Canada	(Li et al., 2013; Lu et al., 2014)
1g	AM910652, KJ439770	Spain, Canada	(Bracho et al., 2008; Lu et al., 2014)
1h	KC248198, KC248199	Cameroon	(Li et al., 2013)
1i	KJ439772	Canada	(Lu et al., 2014)
1j	KJ439773	Canada	(Lu et al., 2014)
1k	KJ439774	Canada	(Lu et al., 2014)
1l	KC248193, KC248197	Cameroon	(Li et al., 2013)
1m	KJ439778, KJ439782	Canada	(Lu et al., 2014)
1n	KJ439775, KJ439781	Canada	(Lu et al., 2014)
1o	KJ439779, MH885469	Canada, Germany	(Lu et al., 2014; Wang et al., 2019)
Genotype 2			
Subtypes	**Accession number(s)**	**Geographical distribution**	**References**
2a	D00944, AB047639	Japan	(Kato et al., 2001; Okamoto et al.,

Table 2.1 Continued

			1991)
Subtypes	**Accession number(s)**	**Geographical distribution**	**References**
2b	D10988, AB030907	Japan	(Murakami et al., 2001; Okamoto et al., 1992)
2c	D50409	Italy	(Nakao et al., 1996)
2d	JF735114	Canada	(Li et al., 2012)
2e	JF735120	Canada	(Li et al., 2012)
2f	KC844042, KC844050	China	(Xu et al., 2013)
2i	DQ155561	Vietnam	(Noppornpanth et al., 2006)
2j	HM777358, JF735113	Venezuela	(Li et al., 2012; Sulbarán et al., 2010)
2k	AB031663	Russia	(Samokhvalov et al., 2000)
2l	KC197235, KC197240	France	(Jordier et al., 2013)
2m	JF735111, JX227967	Canada, UK	(Li et al., 2012; Newman et al., 2013)
2q	FN666428, FN666429	Spain	(Martró et al., 2011)
2r	JF735115	Canada	(Li et al., 2012)
2t	KC197238	France	(Jordier et al., 2013)
2u	JF735112	Canada	(Li et al., 2012)
2v	MW041295	Tunisia	(Rajhi et al., 2021)
Genotype 3			
Subtypes	**Accession number(s)**	**Geographical distribution**	**References**
3a	D17763, D28917	New Zealand, USA	(Sakamoto et al.,

Table 2.1 Continued

			1994; Yamada et al., 1994)
Subtypes	**Accession number(s)**	**Geographical distribution**	**References**
3b	D49374	Japan	(Chayama et al., 1994)
3d	KJ470619	Nepal	(Li et al., 2014)
3e	KJ470618	Nepal	(Li et al., 2014)
3g	JX227954, JF735123	UK, Canada	(Lu et al., 2013; Newman et al., 2013)
3h	JF735121	Canada	(Lu et al., 2013)
3i	FJ407092, JX227955	India, UK	(Arankalle & Gupte, unpublished 2008; Newman et al., 2013)
3k	D63821, JF735122	Indonesia, Canada	(Lu et al., 2013; Tokita et al., 1996)
Genotype 4			
Subtypes	**Accession number(s)**	**Geographical distribution**	**References**
4a	Y11604	Egypt	(Chamberlain et al., 1997)
4b	FJ462435	Canada	(Li et al., 2009)
4c	FJ462436	Canada	(Li et al., 2009)
4d	DQ418786, FJ462437	USA, Canada	(Li et al., 2009; Timm et al., 2007)
4f	EF589161, EU392175	Central Africa	(Hmaied et al., 2007; Kuntzen et al., 2008)
4g	FJ462432	Canada	(Li et al., 2009)
4k	EU392173, FJ462438	Central Africa, Canada	(Kuntzen et al., 2008; Li et al., 2009)
4l	FJ839870	Canada	(Li et al., 2009)

Table 2.1 Continued

Subtypes	Accession number(s)	Geographical distribution	References
4m	FJ462433	Canada	(Li et al., 2009)
4n	FJ462441	Canada	(Li et al., 2009)
4o	FJ462440	Canada	(Li et al., 2009)
4p	FJ462431	Canada	(Li et al., 2009)
4q	FJ462434	Canada	(Li et al., 2009)
4r	FJ462439	Canada	(Li et al., 2009)
4s	JF735136	Canada	(Lu et al., 2015)
4t	FJ839869	Canada	(Li et al., 2009)
4v	HQ537009, JX227959	Cyprus, UK	(Demetriou & Kostrikis, 2011; Newman et al., 2013)
4w	FJ025855, FJ025856	Portugal	(Koletzki et al., 2009)
Genotype 5			
Subtype	**Accession number(s)**	**Geographical distribution**	**References**
5a	Y13184, AF064490	South Africa	(Bukh et al., 1998)
Genotype 6			
Subtypes	**Accession number(s)**	**Geographical distribution**	**References**
6a	Y12083, AY859526	Hong Kong	(Adams et al., 1997; Zhou et al., 2010)
6b	D84262	Thailand	(Tokita et al., 1998)
6c	EF424629	Thailand	(Lu et al., 2007)
6d	D84263	Vietnam	(Tokita et al., 1998)
6e	DQ314805	China	(Li et al., 2006)
6f	DQ835760	Thailand	(Lu et al., 2007)

Table 2.1 Continued

Subtypes	Accession number(s)	Geographical distribution	References
6g	D63822	Indonesia	(Tokita et al., 1996)
6h	D84265	Vietnam	(Tokita et al., 1998)
6i	DQ835770	Thailand	(Lu et al., 2007)
6j	DQ835769	Thailand	(Lu et al., 2007)
6k	D84264	Vietnam	(Tokita et al., 1998)
6l	EF424628	Asia	(Lu et al., 2007)
6m	DQ835767	Thailand	(Lu et al., 2007)
6n	DQ278894, DQ835768	China, Thailand	(Lu et al., 2005; Lu et al., 2007)
6o	EF424627	Canada	(Lu et al., 2007)
6p	EF424626	Canada	(Lu et al., 2007)
6q	EF424625	Canada	(Lu et al., 2007)
6r	EU408328	Cambodia	(Li et al., 2009)
6s	EU408329	Cambodia	(Li et al., 2009)
6t	EF632071, EU246939	Vietnam	(Lu et al., 2008; Noppornpanth et al., 2008)
6u	EU246940	Vietnam	(Noppornpanth et al., 2008)
6v	EU158186, EU798760	China	(Lu et al., 2008; Wang et al., 2009)
6w	DQ278892, EU643834	China, Taiwan	(Lee et al., 2010; Lu et al., 2006)
6xa	EU408330, EU408332	China	(Xia et al., 2008)
6xb	JX183552, KJ567645	Vietnam	(Li et al., 2014; Wang et al., 2013)
6xc	KJ567651	Vietnam	(Li et al., 2014)
6xd	KM252789, KM252790	Laos	(Li et al., 2015)

Table 2.1 Continued

Subtypes	Accession number(s)	Geographical distribution	References
6xe	JX183557, KM252792	China	(Li et al., 2015; Wang et al., 2013)
6xf	KJ567647, KJ567646	Vietnam	(Li et al., 2014)
6xg	MH492360, MH492361	Myanmar	(Ye et al., 2019)
6xh	MG879000	China	(Wu et al., 2018)
6xi	JX183549, MZ504973	China	(Jia et al., 2021; Wang et al., 2013)
6xj	DQ278891, MZ171127	China	(Jia et al., 2021; Lu et al., 2005)
Genotype 7			
Subtypes	**Accession number(s)**	**Geographical distribution**	**References**
7a	EF108306	Central Africa	(Murphy et al., 2015)
7b	KX092342	Central Africa	(Salmona et al., 2016)
Genotype 8			
Subtypes	**Accession number(s)**	**Geographical distribution**	**References**
8a	MH590698	India	(Borgia et al., 2018)

Source: Adopted from Smith et al., (2018)