ANALYSIS OF THE NS3 PROTEASE GENE ACROSS HEPATITIS C VIRAL GENOTYPES IN TREATMENT-NAÏVE INDIVIDUALS

MEYAGABO EMILY YOUBI

A Dissertation Submitted To The Faculty Of Health Science, University Of The Witwatersrand, Johannesburg, In Fulfilment Of The Requirements For The Degree Of Master Of Science In Medicine.

Johannesburg, 2019

DECLARATION

I, Meyagabo Emily Youbi, declare that this dissertation is my own unaided work. It is being submitted for the degree of Masters of Science in Medicine at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

_____

Candidate Signature

Date: April 2019

ANALYSIS OF THE NS3 PROTEASE GENE ACROSS HEPATITIS C VIRAL GENOTYPES IN TREATMENT-NAÏVE INDIVIDUALS

by

Meyagabo Emily Youbi

(1254665)

supervised by

Dr Nishi Prabdial Sing

and co-supervised by

Dr Constance wose Kinge

Prof Adam Mahomed

In the

Department of Virology,

For the degree of

MSc Medicine

DEDICATION

This work is dedicated with love to my family, most especially to my husband Francis who has continuously supported and encouraged me.

To Adrien and Aden, my little super heroes, for being nothing but a blessing and I will forever carry them with me in whatever I do.

To my siblings, I am thankful for their cheering and songs of hope.

To my mother for reminding me what mattered most and the endless prayers, I am truly humbled.

PUBLICATIONS AND PRESENTATIONS ARISING FROM DISSERTATION

Aspects of this work have been presented at the following symposium:

Youbi ME, Wose Kinge CN, Mahomed A, Prabdial Sing N. Analysis of the NS3 protease gene across Hepatitis C viral genotypes in treatment-naïve individuals.". "Hepatitis in Africa: no room for complacency" held at the University of the Witwatersrand, Johannesburg, South Africa on 29 - 30 November 2016.

ABSTRACT

In the world, approximately 71 million people are infected with hepatitis C virus (HCV) and Egypt has the highest prevalence. Estimates of 27% hepatitis C cases are reported in sub-Saharan Africa. South Africa has approximately 0.91% HCV cases. This raises concerns as many of the infected individuals have no knowledge of their HCV status. With the rise in the various new direct acting antiviral drugs (DAAs), targeting different regions of the viral genome, it is important to subtype strains and to identify the type of mutations found within circulating strains in South Africa. Identification of resistance mutations to drug therapy will help in the policy and treatment guidelines specific to the South African population.

The aim of this study was to identify and describe mutations in the non-structural (NS3) gene across HCV genotypes circulating in South Africa. In order to achieve the aim, the objectives were as follows; to determine whether naturally occurring variants in the NS3 region across different genotypes occur, to molecularly characterize naturally occurring NS3 variants across different genotypes, to compare genotype 5a NS3 variants to that of other genotypes (1 & 5a) and map epitope rich areas on the NS3 protein.

This was a retrospective, descriptive study using stored HCV-RNA positive plasma. Several techniques including RNA extraction, cDNA synthesis, reverse transcriptase polymerase chain reaction (RT-PCR), polymerase chain reaction (PCR), cloning, Sanger and next generation sequencing (NGS) were executed. Phylogenetic and mutational analyses were performed using MEGA 7 and BioEdit 7.1.3 software packages. A web-based prediction server Immune Epitope Database (IEDB) was used to map epitope-rich regions of the HCV NS3 protein.

The NS3 gene was amplified for all the samples using an in-house optimised recombinant *Thermococcus kodakaraensis* (KOD) PCR protocol with published primers as well as a commercial assay (DeepChek) PCR protocol. A total of 39 positive amplicons were gel purified and sequenced. Sanger sequencing results were compared to NGS. Phylogenetic analyses indicated that the majority of the study isolates subtyped as 1a or 1b, clustered with strains from the USA, Australia

and France. Both subtypes are globally distributed. Genotype 5a appeared to be clustering with the USA strains. By NGS, we found resistance-associated mutations (I18V, S122A/G, A170V and D168Y) to NS3 proteases for viral genotypes 1b or 5a. Mutation D168Y was observed in both subtype 1b and subtype 5a. Infected individuals with this mutation have shown reduced susceptibility to protease inhibitors, Faldaprevir, Simeprevir and Vaniprevir. The well-characterized Q80K polymorphism, frequently observed in subtype 1a treatment naïve individuals elsewhere, was not observed in this study. Other known resistance mutations reported for genotype 1 (V36A, T54S and A156T) were observed in this study by NGS but not with Sanger sequencing. NGS was opted in order to characterize the minor mutations. Results from the NGS reported mutations V36L/G and S122R that were not observed by Sanger sequencing. NGS results were obtained at 5% frequency supporting the specificity of the NGS method as compared to the conventional sequencing 20% Majority of the observed epitopes showed high binding affinity to the HLA A2*01 across the HCV genotypes.

In conclusion, we found resistance-associated mutations to NS3 proteases for viral genotypes 1b and 5a. However, the well-published resistance mutation for genotype 1 (Q80K) was rarely observed. Although NGS has reduced cost and time, navigating through the mega dataset could be challenging. The high copy number variant might lead to less accurate sequencing by the NGS because of inherent statistical methods used for assembly. Sanger sequencing is still a preferred method for clinical application. Therefore, more studies to screen for mutations need to be conducted to foster the understanding of the impact of mutations on drug resistance and progression of disease. The development of a vaccine capable of preventing chronic HCV infections remains the most feasible and realistic method of managing HCV infection globally.

ACKNOWLEDGEMENTS

Respect the people who find time for you in their busy schedule, but Love People who never look at their schedule when you need them "Unknown"

With that said, I would like to convey my sincere gratitude to my supervisor Dr. Nishi Prabdial Sing, for her guidance, wisdom and patience throughout my research journey. I pray that she continues sharing her knowledge and inspiring people like me. Many thanks for her constant input, without her this work could not be complete.

To National Institute of Communicable Disease (NICD) Centre for Vaccine and Immunology (CVI) unit, it has been three long years in your midst. I have learnt, cried and laughed. Thank you for your warmth and encouragements.

To Dr. Constance Wose Kinge words cannot express enough my gratitude towards your help and encouragement. Thank you for reminding me that with God, patience and hard work all things work for the best. I can clone and think proactively where PCR is involved, thank you.

To Prof Anna Kramvis, I thank you for welcoming me into your lab and constantly asking how my work was going, to your team I am truly grateful for the hospitality.

To Dr Fortunate Mokoena, thank you so much for the introductory lessons in protein analysis.

I would like to thank the Poliomyelitis Research Foundation (PRF) and National Research Fund (NRF) for their financial assistance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

℃	Degree Celsius
µL	Microliters
5’UTR	5’Untranslated Region
ANS	Antisense
AE	Adverse Events
aa	Amino Acids
Amp	Ampicillin
ATP	Adenosine Triphosphate
bp	Base pair
CDC	Centre for Disease Control
CVI	Center for Vaccinology and Immunology
cDNA	complementary Deoxyribonucleic Acid
DAA	Direct Acting Antiviral
DdNTD	Dideoxynucleosides
dATP	deoxy-Adenine-Triphosphate
DDT	Dithiothreitol
DNA	Deoxyribonucleic Acid
dNTPs	Deoxyribonucleotide Triphosphate
EDTA	Ethylene Diamine Tetra Acetic Acid
FDA	Food and Drug Administration

GB	Genetic Barrier
GC content	Guanine Cytosine content
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C Virus
HIV	Human Immunodeficiency Virus
HLA	Human Leukocyte Antigen
IEDB	Immune Epitope Database
INF	Interferon
IU	International Unit
Kb	Kilobase
LiPA	Line Probe Assay
L	Liter
LB	Luria Bertani
MHC	Major Histocompatibility Complex
MgCl2	Magnesium Chloride
MMX	Master Mix
MAVS	Mitochondrial Antiviral Signaling Protein
NGS	Next Generation Sequencing
NIH	National Institute of Health
NICD	National Institute for Communicable Diseases
NS3	Non-structural Protein 3
Ntd	Nucleotide

OD	Optical Density
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PEG-INF	Pegylated Interferon
PIs	Protease Inhibitors
RAMs	Resistance Associated Mutations
RBV	Ribavirin
rpm	Revolution per minute
RNA	Ribonucleic Acid
RNase	Ribonuclease
RT-PCR	Reverse transcriptase- polymerase chain reaction
SN	Sense
SOC	Standard of Care
SOC medium	Super Optimal Broth medium
TAE	Tris-acetate
TRIF	Toll-like Receptor 3 Adapter Protein
USA	United States of America
Xgal	5 bromo-4-indolyl-β-galactopyranoside
YT medium	Tryptone medium

AMINO ACID NAMES ABBREVIATIONS

Amino Acid	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glysine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phen	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

1.1.1 THE HISTORY OF HEPATITIS C VIRUS (HCV)

Hepatitis C Virus (HCV) was first characterised by Choo *et al* 1989 and Kuo *et al* in 1989 at the Centre for Disease Control (CDC) and National Institute of Health (NIH) respectively. It was later identified as the causative agent of the disease previously known as post transfusion non-A and non-B hepatitis virus infection (Choo *et al*., 1989). This paved a way for the introduction of blood screening in 1990.

The initiation of blood screening for HCV reduced its incidence greatly. Nevertheless, it remains the most common cause of liver disease worldwide. The global rate of people infected or that are dying from HCV is exponentially growing. Human immune deficiency syndrome (HIV) and HCV co-infection does occur. This is explained by their shared behavioural risk factors and transmission routes (Rao *et al.,* 2015, Westbrook and Dusheiko, 2014). There is a shared amount of prove that HIV does change the cause of HCV progression, with HIV patients having high HCV viral loads with low to none clearance rate (Westbrook and Dusheiko, 2014) (Hernandez and Sherman, 2011) HCV infection with low clearance rate may lead to acute or chronic hepatitis (Berenguer *et al*., 2011).

In the world, approximately 71 million people are infected with hepatitis C virus (HCV) and Egypt has the highest prevalence (WHO, 2017). About 700 000 persons die annually from complications caused by HCV for example (e.g.) liver cirrhosis, portal hypertension, hepatocellular carcinoma (HCC) and liver failure (WHO, 2015). An estimate of 27% hepatitis C cases is reported in sub-Saharan Africa (Rao et al., 2015). South Africa has approximately 0.91% HCV cases and between 0.3% - 1.0 % of HIV&HCV co-infection cases (Rao et al., 2015).

Since the development of the direct acting antiviral agents, (DAA), HCV has become curable (WHO, 2017). However, due to the asymptomatic nature of the disease, those infected are unaware or those diagnosed are left without treatment due to poor accessibility and high costs, in many resource-limited settings (WHO, 2015).

1.2 LITERATURE REVIEW

1.2.1 HEPATITIS C VIRUS GENOME

HCV is an enveloped ribonucleic acid (RNA) virus with a diameter of about 50 nm, classified as a separate genus *Hepacivirus* within the Flaviviridae family. It is similar to the pestiviruses and flaviviruses (Miller and Purcel, 1990). HCV has two glycoproteins which are embedded within the lipid membrane they are important for viral attachment and entry into the cell. The viral RNA material is found inside the icosahedral shaped core. The virus resides in the cell cytoplasm and infects mostly the livers hepatocytes (De Francesco, 1999). It is made up of a single-stranded RNA molecule of 9.6 kilobases (kb) with one long open reading frame (ORF) which codes for a large polyprotein of about 3000 amino acids (aa). The polyprotein is co- and post-translationally processed by cellular and viral protease to yield matured structural and non-structural proteins (De Francesco, 1999). Ten viral proteins (3 structural and 7 non-structural) are present, namely Core, E1, E2, P7, NS2, NS3, N4A, N4B, NS5A and NS5B (Pawlotsky *et al.,* 2007, Figure 1). Among the non-structural proteins, the serine-like protease and the RNA-dependent RNA polymerase are essential for the viral maturity and replication of the virus.

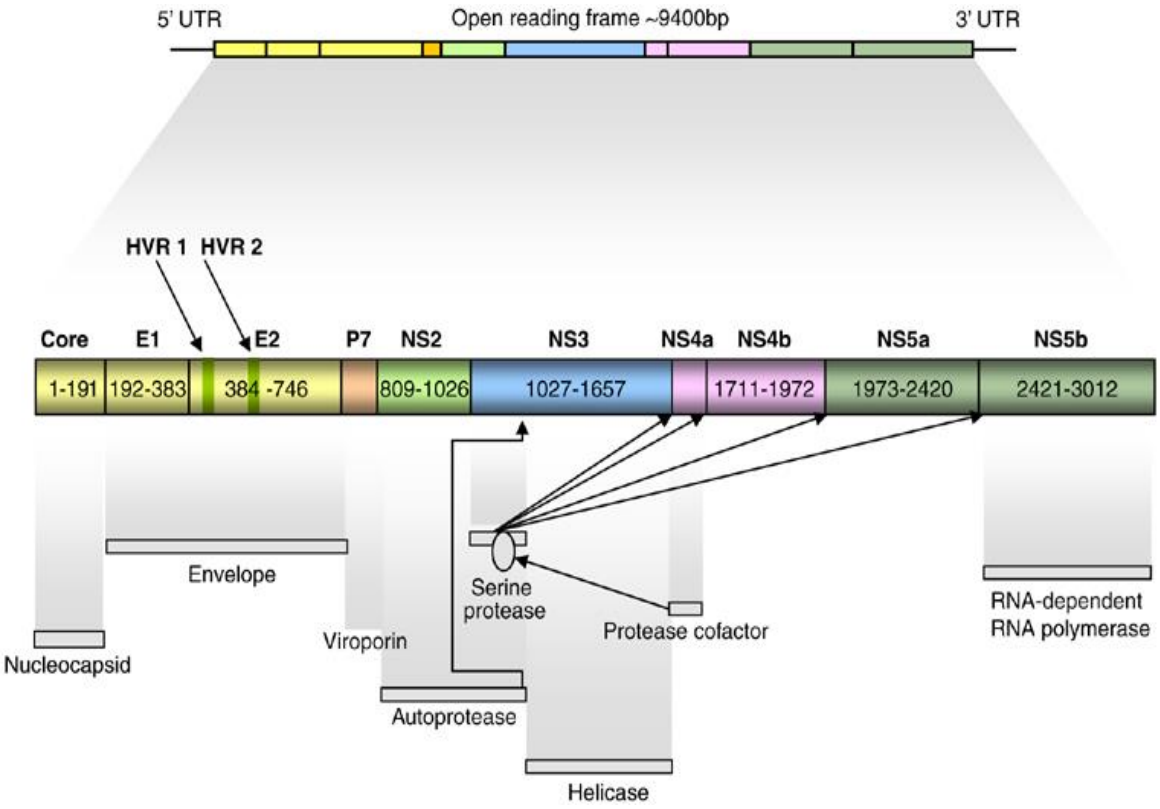

Figure 1.1: Protein encoded by HCV genome. HCV is a small enveloped virus, encoding a large polyprotein of 3010 aa. The mature structural (C, E1 and E2) and non-structural proteins (NS2-5) are co- and post-translationally processed by cellular and viral encoded protease. The proteins are flanked by untranslated regions. The length of each protein is depicted as number of nucleotides. (Adapted from Abdelrahma, 2015).

1.2.2 NON STRUCTURAL 3 (NS3) PROTEIN

The NS3 protein is multifunctional and plays a major role in the HCV life cycle. Approximately 180 aa within the N-terminal of the protein resides a serine protease that mediates the cis-cleavage of the NS3/4A side and NS4A/4B, NS5A/5B junctions, liberating the non-structural proteins for the assembly of HCV replication complex. The C-terminal that is associated with the ATP-dependent RNA helicase is about 440 aa. It functions in the unwinding of RNA-RNA duplexes during replication (Tai *et al.,* 1996). The unwinding of the duplexes either eliminates RNA secondary structures or separates the genome from nucleic acid binding proteins. The NS4A cofactor (Figure 1.2) enhances

the NS3 proteolytic function by nearly 1000 fold and it cleaves components toll-like receptor 3 adaptor protein (TRIF) (Li *et al.,* 2005) and mitochondrial antiviral signalling protein (MAVS) (Meylan *et al.,* 2005) of the host's double stranded RNA. This inhibits the cellular host defence, as a result lead to viral chronicity. The NS3 protein activity is essential for the generation of the components of the viral RNA replication complex (Du *et al.,* 2002).

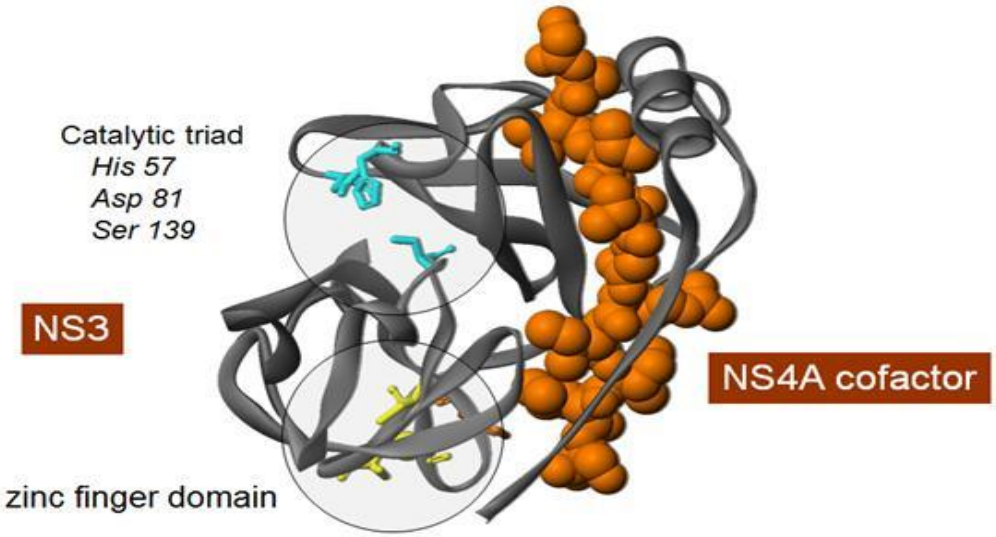

Figure 1.2: Molecular structure of the HCV NS3 – 4A protein. The b-barrel groove contains the protease catalytic triad Histidine 57, Asparagine 81 and Serine. The groove also houses the active site of the NS3-4A protein. The shallowness of the groove makes the designing of the compound inhibitors relatively difficult. (Adapted from Kim *et al.,* 2014).

1.2.3 DISTRIBUTION OF THE HEPATITIS C VIRUS

HCV is highly diverse. The low fidelity of the HCV polymerase (10^{-3} errors/per sites), high replication rate, and strong selective pressures on the virus lead to unique viral quasispecies in each patient (Powdrill *et al.,* 2011). Quasispecies are a population with distinctive but closely related viral genomes. Identification and characterization of the quasispecies is of paramount importance, this allows proper management of the disease progression and treatment outcome. It has been reported that quasispecies complexity and diversity hinders development of effective vaccines (Powdrill *et al*., 2011). The HCV genome is highly variable and classified into genotypes 1-7 (Simmonds, 2004; Smith *et al*., 2014). This is based on the analysis of phylogeny and sequences of the complete viral genomes. HCV genotypes are globally distributed (Figure 1.3). Subtypes like 1a, 1b, 2a and 3a are widely distributed. They account for high infection rate in first world countries. Other subtypes have been circulating rarely in specific regions for long periods with subtypes of genotype 1-2 found in West Africa, 3 in South Asia, 4 Central Africa and Middle East, 5 predominating in South Africa (Kapoor *et al.,* 2013) and 6 in South East Asia. Genotype 5a was known to exist only in South Africa however, Verbeeck *et al.,* 2006 and Henquell *et al.,* 2004 indicated that was not the case. They reported on an increasing presence of genotype 5a in North Africa and Europe. Genotypes 1-4 are also reported in South Africa (Prabdial Sing *et al.,* 2008). Genotype 7 has been isolated in Canada from a Central African immigrate (Murphy *et al.,* 2007). Over the years prevalence of genotype 4-5 has increased. This is due to the emigrations from the Middle East and Africa (Kapoor *et al.,* 2013)

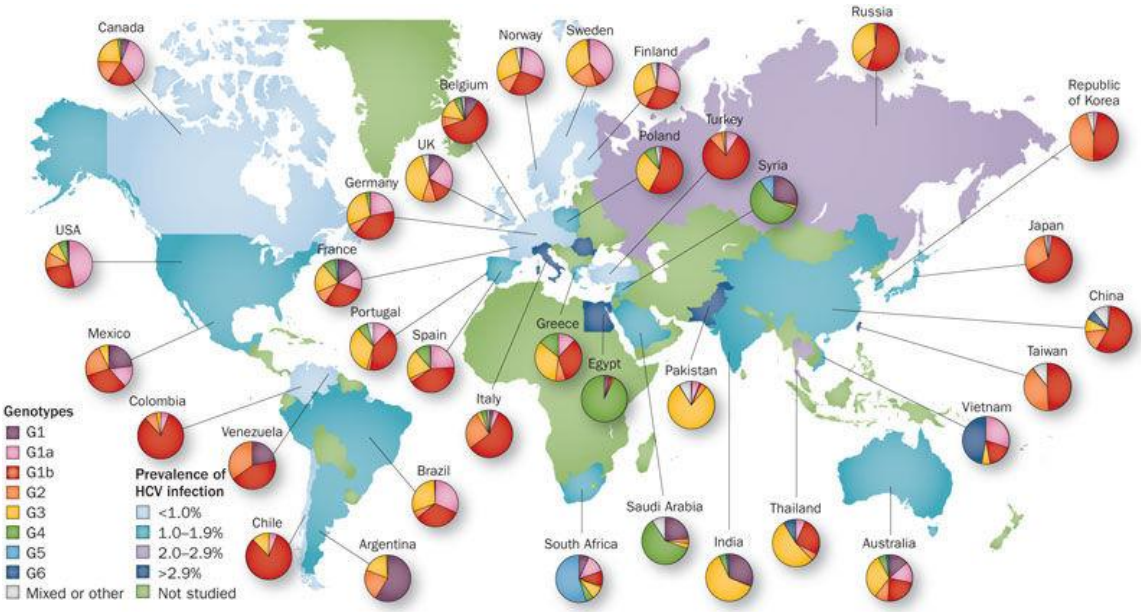

Figure 1.3: World distribution of the HCV genotypes. (Adapted from Hajarizadeh *et al.,* 2013).

1.2.4 CURRENT HCV TREATMENT REGIMENS

HCV infection does not always need to be treated since some patients clear the infection and those that are chronically infected do not present damaged livers overtime (WHO, 2015). Pegylated interferon (PEG-IFN) in combination with ribavirin (RBV) is currently the standard of care (SOC) in South Africa and other developing countries. The interferons (IFNs) are the family of cytokines released when the host cell is stimulated by a foreign particle, this may include infections by viruses. The multiple antiviral effector proteins are induced to fight against viral replication (Stetson and Medzhitov, 2006). Ribavirin is a nucleotide analogue, which have a strong reaction against several RNA and deoxyribonucleic acid (DNA) viruses. The exact action of ribavirin is still to be investigated (Stetson and Medzhitov, 2006). PEG-IFN and ribavirin therapy made strides in HCV treatment. The significant factors that affect treatment response are age, gender, stage of liver disease and HCV genotype (Ghany *et al.,* 2009). In addition, side effects, expenses and the length of treatment led to many patients' failure to tolerate the full course of the treatment, leading to treatment failure (Sulkowski *et al.,* 2011). An increase in the amino acids

variations of the NS3 protein has been linked to PEG-IFN and ribavirin combination therapy failure in individuals with genotypes 1a, 1b, 3a and 3b (Tsubota *et al.,* 2011). Studies have shown response rate of 75%-90% in patients infected with genotype 2 and 3 and that of 45%-52% in patients infected with genotype 1 and 4 (Deutsch and Hadzyannis, 2008). No substantial data exist for therapy response in relation to genotype 5a (Hueppe *et al.,* 2009).

To date, advanced knowledge on HCV life cycle has led to major developments towards therapy with direct acting antiviral agents (DAA) approved by USA Food and Drug Administration (FDA). Treatment of HCV using DAA has shown a remarkable high sustained virological response rates (Falade-Nwulla *et al*., 2017). A number of protease inhibitors (PIs) have been approved thus far, while others are still in clinical trial phases (see Appendix A). Many of the upcoming DAAs have undergone clinical trials as dual, triple or even quadruple therapies, some with or without PEG/RBV. These new treatments have shown to have few adverse events (AEs), high response rate, high genetic barrier to resistance and reduced treatment duration (Jacobson *et al.,* 2011; Kim *et al.,* 2014). However the selection of the treatment and duration is based wholly on the genotype and the subtype.

1.2.5 GENETIC BARRIER TO RESISTANCE

Genetic barrier (GB) is often quantified at the number of substitution(s) required to confer resistance, this depends on the antiviral target and class (Gőtte, 2012). Viruses only require one or two substitution(s) to gain resistance and the older the antiviral class the lower the GB. The GB affects the rate of mutations. The common mutation conferring resistance R155K/G requires 1 nucleotide (ntd) substitution in genotype 1a and 2 ntd substitutions in genotype 1b but this is not commonly observed (Gőtte, 2012). Substitution quantity is not only the effector of the GB, for instance, mutations requiring transitions are said to be less frequent than transverse ones (Powdrill *et al.,* 2011). The GB will result in different types of mutations that include resistance, reversions, deleterious or even compensatory mutations depending on the mutation rate, population size and viral fitness (Maisnier-Patin and Andersson, 2004).

1.2.6 HCV MUTATIONS WITHIN THE NS3 PROTEIN

Viral infections that impair the host health are treated and managed with antiviral drugs targeting proteins involved in replication mechanism. The robustness of the drug impairs the viral fitness completely. However, should the latter occur then some viruses will replicate, selective pressures may result in rapid adaptations towards resistance. Quasispecies within the NS3 allow rapid viral adaptation to immunological response and antiviral selection pressures (Paolucci *et al*., 2013). There has been growing output of high quality molecular data generated. This allows new mutations to be investigated on a finer scale especially those that confer drug resistance (Bracho *et al.,* 2004; zur Wiesch *et al.,* 2011). Several studies have identified mutations in the viral NS3 region, particularly for genotype 1 infections, that either allow resistance against the inhibitor or render the virus susceptible to the drug of choice (as shown in Appendix A).

There is little data on the mechanisms of the mutations to response to therapy. Interestingly, a susceptibility study on baseline genotype 1 samples in a transient-replication assay on several protease inhibitors had shown that whilst samples had the Q80K polymorphism, they were less sensitive to Simeprevir but more sensitive to Boceprevir and Telaprevir (Izquierdo *et al*., 2014). A few studies (Vallet *et al*., 2005 and Kieffer *et al.*, 2007) have shown that some amino acid substitutions can affect the three-dimensional structure of the NS3 serine protease thus affecting the charge of the protein, which modifies the interactions between different protein residues (Vallet *et al.*, 2005). Pharmaceutical companies such as Merck and Vertex have decided to discontinue with the sales of Boceprevir and Telaprevir because they fall short when compared to the efficacy and tolerability of new interferon free regimen. This has been reported to have a negative impact on patients as most developed severe anaemia post early treatment cessation (Colombo *et al.,* 2014; Hezode *et al.,* 2014). Some viral mutations mediate immune escape of the virus as they interact with Human Leukocyte Antigen (HLA) molecules and T-cell receptors for example; the Y/F1082 mutation at the flanking regions of

the T-cell epitopes impairs correct carboxyl terminal cleavage of an immunodominant epitope. This prevents proper antigen processing and presentation of the major histocompatibility complex (MHC) class I restricted epitope (Prezzi *et al*., 2001).

Whilst there are several studies on NS3/NS4 mutations driving resistance to inhibitors across hepatitis C genotypes, many of the studies have focused on genotype 1 (Howe *et al.,* 2014; Takayama *et al.,* 2014; Everson *et al.,* 2012; McPhee *et al.,* 2012 and Qui *et al.,* 2009) and genotype 4 (Everson *et al.,* 2012). Cento *et al.,* 2012 looked at data on six genotype 5a specimens in their mutational analyses. These mutations affect diagnosis and the susceptibility/resistance of the virus to antiviral agents (Paolucci *et al.*, 2013). Consequently, most HCV infected individuals develop chronic infection, which might lead to liver cancer. Moreover, mutations affect development of other HCV therapies and vaccines.

1.2.7 METHODS OF MUTATION ANALYSIS

Viral complexity and detection of minority (low frequency) mutations have been a problem affecting the identification of viral mutations present in any clinical sample. Next generation sequencing (NGS) has afforded the scientific community a better platform to understand the intra-host viral population. This has favoured NGS as a potential alternative to conventional methods (Cruz-Riveral *et al.,* 2013).

1.2.7.1 Sanger Sequencing

An extracted viral RNA is amplified by Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) and then sequenced directly using Sanger sequencing (Sanger and Coulson, 1975). Sanger is defined as a method of DNA sequencing based on incorporating selective chain termination of dideoxynucleotides by DNA polymerase during in vitro replication. The determination of the consensus sequences of all variants within the population

is achieved by data analysis. A sequencing platform identifies the common base at each position of the sequence, when there are two or more bases at the same position with the same frequency; the sequencing software fails to identify the common base. However, this can be resolved by viewing the chromatogram together with the sequencing results. Sanger threshold ranges between 15% - 20% and is still a preferred method for clinical application (Cruz-Riveral *et al.,* 2013)

1.2.7.2 Molecular Cloning

The analysis of the heterogeneity of the viral population and determination of resistance mutation has been achieved through cloning of the PCR amplicons. This technique has been used as a reference to investigate viral population; however this procedure is tedious and expensive when compared to the NGS and not more than 100 clones were obtained by a researcher (Beerenwinkel and Zagordi, 2011). Many aspects of viral population were under researched and it still remains a challenge to determine full intra-population heterogeneity through cloning. Next generation sequencing has changed the field of genome analysis (Beerenwinkel and Zagordi, 2011).

1.2.7.3 Next Generation Sequencing

Introduction of NGS has changed the way viral populations are analysed. Higher detailed coverage of the circulating mutations is better with NGS when compared to conventional Sanger sequencing. However, the short PCR step during the NGS might introduce errors and different sequencing platforms have their instrumental errors (Poh *et al.,* 2013). Sanger sequencing only use a single DNA fragment at a time, while NGS is largely parallel, sequencing millions of fragments per run. Although NGS has reduced cost and time, navigating through the mega dataset could be challenging. The high copy number variant might lead to less accurate sequencing because of inherent statistical methods used for assembly (Watson *et al.,* 2013). Milestones have been reached with NGS technology which includes the discovery of the Merkel cell polymavirus (Feng *et al.,* 2008). Detection of viruses in chronic infections is

still a major challenge that requires a sensitive method as the viral burden resolves with disease progression (Duvoux *et al.,* 1999). This method is not bias, as it requires random primers during the cDNA synthesis. The presence of minor variants can be indicated in viral quasispecies (Watson *et al*., 2013). RNA sequencing is random. Majority of reads either correspond to the genetic material or contaminated species. Hence, advanced computation analysis is required.

1.2.7.4 Protein Crystallography

The discovery of X-ray by Wilhelm Conrad Rontgen and modification by Max von Laue gave rise to the protein crystallography methods (Roentgen WC, 1972). In 1912 Max von Laue was first to observe diffraction of X-ray and he went on to reveal more wave nature of X-rays (Annalen der Physik, 2012). Many experiments were performed to explore the use of X-ray diffraction. It was Braggs WH and Braggs WL in 1912 who realised that X-ray diffraction could be used to determine the atomic structure. Only 45 years later, the first protein (myoglobin) was determined by protein crystallography. However; only protein crystallography and nuclear magnetic resolution spectroscopy are currently used to determine the resolution of tertiary protein. More research is ongoing to improve and increase methods to determine the resolution of tertiary protein. Since February 2018 140000 entries were made to the Protein Database and majority of the entries were determined by the diffraction methods. This proves the domination of the diffraction methods in the field of structural biology.

1.2.8 HCV EPITOPES

Human leukocyte antigen (HLA), also known as Major histocompatibility complex (MHC) is a set of cell surface proteins important for the acquired immune system to recognise foreign particles such as viruses (Van der Poel, 1994). The distribution of human leukocyte antigen (HLA) is based on ethnicity and its place of origin. South Africa has high HLA diversity due to its diverse ethnic groups (Paximadis *et al.,* 2012). Like any other viruses, the hepatitis

virus processed into peptides binds to the HLA. The peptides are presented on the surface of the infected cells. The complex of HLA class 1 or 2 recognises either CD8+ or CD4+ T cells, respectively. Therefore, these cells act as T-helper cells or regulatory T-cells. Epitopes that are related to cell mediated immune response are found in all HCV proteins, with the core and NS3 been the relevant antigens because of their conserved nature among viral proteins. This has been linked to HCV viral clearance (Shoukry *et al.,* 2014) The NS3 protein plays major roles like viral replication, targets for diagnosis, antiviral therapy and vaccine development. Previous studies have identified immunodominant B cell epitopes within the NS3 helicase that produce high levels of antibodies in an infected individual (Van der Poel*,* 1994 and Mondelli *et al.,* 1994). NS3 has several conserved helicase motifs reported. Genotype 1b restricted epitope ATDALMTGY contain a highly conserved motif V of the helicase domain 2 (Frick, 2007).

Different vaccine candidates against HCV have been discovered and evaluated. They differ according to their vaccination approach and epitope target. Some are comprised of few viral epitopes targets and focusing on E1E2 proteins, that stimulates humoral and cell mediated immune response (Sharon *et al.,* 2010). While others focus on core/NS3 that inducts cell immune response (Rosen *et al.,* 2002).

1.2.9 PROBLEM STATEMENT

In South Africa, the approved treatment of HCV infection is PEG/INF and RBV (Ghany *et al.,* 2009). Currently, there are more protease inhibitors that have been approved for HCV treatment and many others are in the pipeline (Dufour *et al.,* 2017; Dore *et al.,* 2016; Hayashi *et al.,* 2014 and Jacobson *et al.,* 2011). They have shown more than 80% success rate for certain HCV genotypes. Several studies have reported on quasispecies associated with the resistance in treatment naïve patients (Howe *et al*., 2014; Takayama *et al*., 2014; Everson *et al*., 2012; McPhee *et al*., 2012 and Qui et *al.*, 2009). To the best of our knowledge to date, no data is available on NS3 variants occurring across genotypes in South Africa, especially for genotype 5a, which predominates.

Hence this study provides insight into the different circulating variants in relation to the direct acting antivirals.

1.2.10 AIM AND OBJECTIVES OF THE STUDY

1.2.10.1 The aim of the study

The aim of this study was to identify and describe mutations in the NS3 gene across HCV genotypes from retrospectively stored sera.

1.2.10.2 The objectives of the study

The objectives of the study were to:

1. Determine whether naturally occurring variants in the NS3 region across different genotypes occur.
2. Molecularly characterize naturally occurring NS3 variants across different genotypes.
3. Compare genotype 5a NS3 variants to that of other genotypes (1&5).
4. Map epitope-rich areas on the NS3 gene.

1.2.11 SIGNIFICANCE OF THE STUDY

Data obtained from the study will identify and characterize naturally occurring HCV variants in the NS3 region, across HCV genotypes. The NS3 protease gene has not only been a target for the newly developed drug therapies but has shown to be one of the candidates for vaccine development with promising results. Knowledge obtained from this study on treatment naïve individuals will have a huge impact on predicting treatment resistance, policies on governing treatment, management and knowledge of resistance mutations in South Africa. This study may also shed light in the development of epitope vaccines.

CHAPTER 2: MATERIALS AND METHODS

2.1 ETHICS CLEARANCE

This study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (certificate number M170765). Permission was granted by the Head of Virology at the National Institute of Communicable Diseases (NICD), Centre of Vaccinology and Immunology (CVI) to use the plasma samples. Laboratory numbers were used to identify the samples ensuring anonymity of the samples (as shown in Appendix D).

2.2 STUDY DESIGN

This was a retrospective, descriptive study, using HCV positive stored plasma.

2.3 STUDY POPULATION

A total of 109 HCV positive samples, previously genotyped as 1a, 1b and 5a from anonymous blood donors and clinical patients, were retrieved from the National Institute of Communicable Disease, Centre of Vaccinology and Immunology, Johannesburg, South Africa repository for this study. Genotyping was based on the conserved 5'UTR region by Line probe assay (LiPA). The samples were labeled according to the laboratory numbers; however, genotype 5a sample numbers were relabeled according to the successful sequencing direction (SN = sense and ASN = antisense) for the purpose of this study. Table 2.1 below shows a representative subset of the samples, their genotype and demographic characteristics.

Table 2.1 Genotype profile and demographic characteristics of the sample population

SAMPLE ID	GENOTYPE	STUDY GROUP	GENDER	AGE
ASN3	5a	Patient	Unknown	Unknown
ASN548	5a	Blood donor	Female	30
ASN6765	**5a**	**Blood donor**	**Female**	**51**
SN959	5a	Patient	Male	56
SN943	5a	Patient	Male	51
ASN17	5a	Patient	Unknown	Unknown
CVI1077	1a	Blood donor	Female	56
CVI0050	1a	Patient	Male	36
CVI0090	1a	Patient	Male	50
CVI0033	1a	Patient	Male	37
CVI0019	1a	Patient	Male	37
CVI6861	1b	Blood donor	Female	33
CVI0095	1a	Patient	Male	26
CVI0031	1a	Patient	Male	32
CVI0075	1a	Patient	Female	32
CVI6650	1a	Patient	Male	63
CVI6605	1b	Patient	Unknown	Unknown
CVI1087	1b	Blood donor	Male	43
HG04335838	**1a**	**Patient**	**Unknown**	**Unknown**

Table 2.1 continues,

SAMPLE ID	GENOTYPE	STUDY GROUP	GENDER	AGE
HN01361785	1a	Patient	Unknown	Unknown
HN01393031	1a	Patient	Unknown	Unknown
CVI599	1a	Blood donor	Male	33
CVI0018	1a	Patient	Male	25
CVI0051	1a	Patient	Male	44
MD6773	1a	Blood donor	Male	37

ASN6765 and HG04335838 show the highest and lowest viral load.

HCV viral load for these samples was measured and obtained from the NICD database. Figure 2.1 below shows RNA copies for nine randomly selected samples. As shown sample ASN6765 had the highest viral load and sample HG04335838 had the lowest viral load.

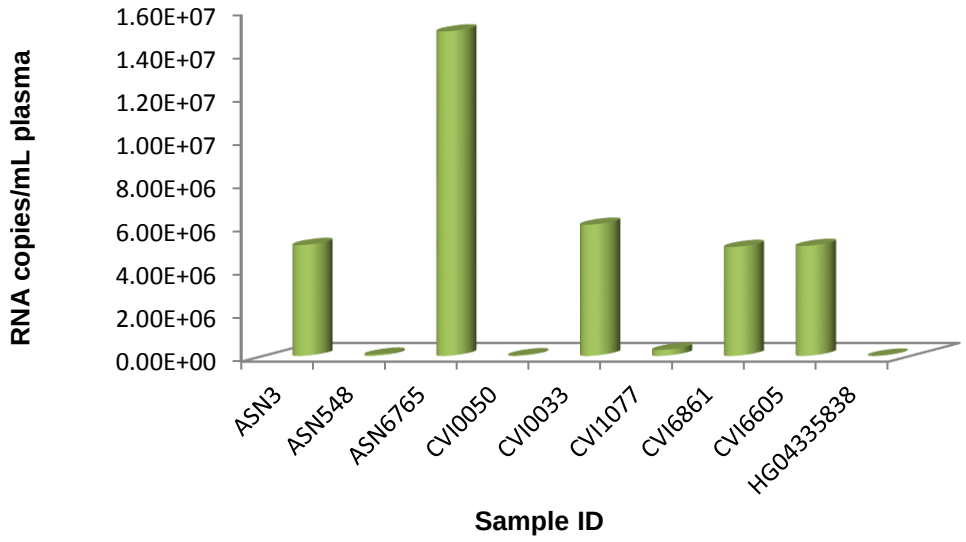

Figure: 2.1 HCV RNA copies for nine study samples

2.4 ELIGIBILITY CRITERIA

The ideal scenario was to have representative numbers of all genotypes in the study but this was dependent on plasma sample volume and integrity of viral RNA.

2.4.1 Inclusion Criteria

Samples that met the following criteria were included in the study:

i. HCV viral load ≥ 1000 international units/mL

ii. Known HCV genotype (1 and 5a)

iii. Sample volume ≥ 300µL

2.4.2 Exclusion Criteria

i. HCV viral load < 1000 international units/mL

ii. Unknown HCV genotype

iii. Sample volume < 300µL

2.5 LABORATORY METHODS

2.5.1 RNA Extraction and cDNA Synthesis

Total RNA was extracted from 140µL plasma using the QIAmp viral RNA mini kit protocol (Invitrogen, United State of America) as instructed by the manufacturer. Five microliters of the extracted RNA was subjected to cDNA synthesis using the Superscript III™ reverse transcriptase protocol in two step reaction (Invitrogen, United State of America). The extracted RNA was mixed with 1µL of 50ng/µL, random hexamer, 1µL of 10Mm dNTP and 3µL distilled water. The reaction was incubated for five minutes at 65˚C in a Gene Amp® PCR system 2700 (Applied Biosystems, United State of America) and placed on ice for a minute. A total of 10µL of master mix (MMX) containing: 2µL, 10X

RT buffer, 4µL of 25mM $MgCl_2$, 1µL of 40U/µL RNase Out and 1µL of 200U/µL Superscript III), was added to the reaction mix and incubated further at 25˚C for 10 minutes, 50˚C for 50 minutes, and 80˚C for five minutes. RNA was quantified using Nanodrop model ND-1000 (Peqlab, England) and the integrity of the RNA was assessed on 1% agarose gel. Aliquots were stored to at -20℃ for downstream analysis.

2.5.2 PCR Amplification of the NS3 protein using KOD polymerase

PCR amplification was performed using KOD DNA polymerase (Toyobo, Japan) with published NS3-specific primers as shown in Table 2.1 (Cento *et al*., 2012) according to manufacturer's instructions with modifications. Briefly, 2µl of cDNA was added to PCR tubes containing first round MMX: 5µL buffer, 3µL MgSO4, 1.5µL primers forward and reverse (F1 and R1), 5µL dNTPs, 1µL KOD enzyme and 31µL distilled water. The cycling conditions were as follows: initial denaturation step at 95˚C for two min, 45 cycles were set for the amplification at 95˚C for 20 sec, 57˚C for 20 sec) and a final elongation step at 70˚C for 20 min followed by holding temperature at 4˚C. Five microliters of the 1st round PCR was amplified with inner primers (F2 and R2). The nested reaction MMX contained: 5µL buffer, 3µL MgSO4, 1.5µL primers, 5µL dNTPs, 1µL KOD enzyme and 28µL of distilled water. The same cycling conditions as the first round were performed. The expected band size was 760 bp. The PCR products were visualized on 1% agarose gel by electrophoresis

Table: 2.2: Primer sequences used for amplification of HCV Subtype 1a and 1b.

Name	Sequence (5' ⟶ 3')	Length (bp)
F1 (1st round)	GTGCCSTACTTYGTGCGCG	19
R1(1st round)	CCGTCGGCAAGGAACTTGCC	20
F2 (Nested)	GGAGACYAAGCTCATYAACSTGGGG	25
R2 (Nested)	CMGGSACCTTGGTGGTGCTCTT	22

F1 and F2 (Forward primers) R1 and R2 (Reverse primers).

2.5.3 PCR amplification of NS3 protein using DEEPCHEK

The NS3 gene was also amplified using the DeepChek® RT-PCR Sequencing-NS3 Drug Resistance Assay kit (ABL, France with its HCV specific primers. The DeepChek primer sequences were not provided. Reverse transcriptase (RT) PCR was performed using 5µL of the extracted RNA. Briefly, 10µL of Reverse Transcriptase (RT) MMX which consisted of 7.5µL 2X RT-PCR buffer, 0.3µL RP solution, 0.15µL BP solution, 0.3µL RT enzyme and 1.75µL water, giving a total of 15µL reaction volume. The cycling conditions were as follows: reverse transcription step at 30˚C for 5 min, then, 42˚C for 5 min and 95˚C for 15 sec. followed by holding temperature at 4˚C.

Next, first round PCR was performed using the entire 15µL volume of the RT product. To this, 15µL aliquot of MMX prepared by mixing 7.5µL 2X RT-PCR buffer, 0.6µL NS3 F1 PCR primers, 0.6µL NS3 F2 PCR primer, 1.2µL NS3 R1 PCR primer, 0.3µL PCR enzyme and 4.8µL water, was added. The following cycling conditions were adhered to: an initial denaturation step at 94˚C for 3 min, 45 cycles for the amplification at 94˚C for 15 sec, 53˚C for 1 min and 72˚C for 1 min followed by holding temperature at 4˚C. The expected size was 776 bp.

Lastly, nested PCR was performed and; two microliters of the first round PCR product was added to the nested MMX. The nested MMX contained: 12.5µL 2X Nested PCR buffer, 0.5µL NS3 F1 nested PCR primer, 0.5µL NS3 F2 nested PCR primer , 1µL NS3 R1 nested PCR primers. 0.3µL Nested PCR enzyme and 8.3µL water. The following cycling conditions were adhered to: an initial denaturation step at 94˚C for 5 min, 5 cycles for initial amplification at 94˚C for 30 sec, 53˚C for 1 min and 72˚C for 1 min., 40 cycles for second amplification at 94˚C for 15 sec, 67˚C for 30 sec, 72˚C for 30 sec followed by holding temperature at 4˚C. The expected band size was 650 bp. The PCR products were analysed on 1% agarose gel by electrophoresis.

2.6 AGAROSE GEL ELECTROPHORESIS

All amplified PCR products were visualized on 1% agarose gel by electrophoresis. Briefly, one gram of agarose powder (Merck, United State of America) was dissolved in 100mL of 1X TAE (Tris-acetate-EDTA) buffer to make up a 1% gel solution. The mixture was heated until completely dissolved. The agarose was let to cool to ~60˚C, ethidium bromide was added and gel was casted on gel trays. Five microliter of the amplicon and 1µL of the 6X loading dye were mixed and loaded onto the gel. The DNA fragments separated at 100 volts in a Mini-Sub Cell GT mini gel horizontal submarine unit (BIO-RAD, United State of America) for 60 minutes. One Kilobase (kb) Lambda DNA molecular weight marker (New England biology United State of America) was used to assess fragment sizes of the amplicons and to estimate the concentration of the DNA. Gels were visualized by a gel fluorescence (UV-light) imaging system; G: BOX and Gene Snap image acquisition software (Syngene, India).

2.6.1 DNA fragment extraction and purification

A clean sharp scalpel was used to excise the DNA fragment from the 1% agarose gel. The gel was weighed in an Eppendorf tube and purified using the Mini-Elute PCR purification kit (QIAGEN, Germany) following the kit protocol. Briefly, 3 volumes of the buffer QG to 1 volume of the gel (100mg gel~ 100µL) were added. The maximum amount of gel slice per spin column is 400mg for >2% Agarose gels and 6 volumes of the buffer QG was added. The gel slice was incubated at 50˚C for 10 minutes until completely dissolved. Vortexing was done every 2-3 minutes during incubation to help dissolve the gel. After dissolving the gel slice, the colour of the mixture was compared to the QG buffer without the gel slice. The sample was applied to the mini-Elute column and centrifuged for 1 minute. The flow-through was discarded and the mini-Elute column was placed back to the same collection tube. Five hundred microliters of buffer QG was added to the mini-Elute column, centrifuged for 1 minute and the supernatant were discarded. DNA was washed twice with 750µL of buffer PE. Since the samples were to be used in salt-sensitive procedures, the column was let to stand for 2-5 minutes after the addition of the buffer PE. DNA was eluted twice in a final volume of 10µL with elution buffer and used for downstream analysis.

2.7 MOLECULAR CLONING OF NESTED PCR AMPLICONS

Purified amplicons were monoadenylated as previously described by Wose Kinge *et al*. (2014). This is because of the blunt end formed during the amplification. Briefly, the MMX contained 10X PCR buffer 2µL, 50mM $Mgcl^2$ 0.6µL, Taq DNA polymerase 1µL, 2mM dATP 2µL and Template (amplicon) 14.4µL. The mixture was incubated at 72˚C on a heat block for 10 minutes. The monoadenylated products were purified using Mini-Elute PCR purification kit protocol (QIAGEN, United State of America). Ligation of the purified DNA amplicons was performed with PCR 2.1 TOPO-TA vector following manufacturer's instructions (Invitrogen, United State of America). Briefly, reaction MMX contained purified monoadenylated PCR amplicons 4µL, salt solution 1µL and TOPO TA vector 1µL, were added to 1.5 mL Eppendorf tube. The MMX was incubated at room temperature for 30 minutes. Competent *E.coli* cells (DHα5) prepared as described by Inoue *et al*. (1990) were transformed together with the ligated plasma DNA and spread on pre-warmed Luria Bertani (LB)/ampicillin (Amp) (100mg/mL) agar plates containing 5-bromo-4-indolyl-β-d-galactopyranoside (Xgal) (20µg/µL). The plates were incubated at 37^oC for 16 hours.

2.7.1 Analysis of the Cloned Vector

To screen for the presence of the NS3 fragment in the cloned vector, 10 colonies were picked from each plate using sterile pipette tips. The colonies were incubated at 37℃ for 12-16 hours in tubes containing 2µL of LB broth or Yeast extract and Tryptone (YT) broth with ampicillin. Following incubation, small-scale DNA preparations were performed using the Qiagen Plasmid DNA Extraction kit protocol (QIAGEN, United State of America). Briefly, the tubes were centrifuged at 3000 rpm for 15 minutes. The pellets were re-suspended in 250µL of P1 buffer and transferred to eppendorf tubes. Two hundred and fifty microliters of P2 buffer was added and mixed by inverting the tubes 4 to 6 times. Then, 350µL of N3 buffer was added to stop the lysis and tubes were mixed by inverting 4 to 6 times. The tubes were centrifuged at 13000 rpm for 10 minutes, leaving the white pellet at the bottom or the side of the tube. The supernatant was transferred to silica columns, centrifuged at 13000rpm for a minute and the flow-through was discarded. Seven hundred and fifty microliters of wash buffer was added and centrifuged, flow-through was discarded and the above step was repeated. The columns were transferred to the eppendorf tubes and plasmid DNA eluted in 50µL elution buffer.

2.7.2 Restriction Enzyme Digest

The purified DNA was then subjected to enzyme digestion with *EcoRI* (New England biology United State of America). Briefly, a MMX was prepared were 1µL of the NEB buffer 3, 6.5µL sterile distilled water and 0.5µL *EcoRI* enzyme were added. A total of 8µL of the MMX was transferred into 96-well plate and 2µL of the extracted plasmid DNA was added and incubated for an hour at 37 C. The digested plasmid was visualized by agarose gel electrophoresis to check for complete digestion and correct fragment sizes.

2.8 SEQUENCING AND PHYLOGENECTICS ANLYSIS

2.8.1 Sanger Sequencing

Sanger sequencing of the NS3 gene was performed in-house following BigDye™ Terminator v1.1 and BigDye XTerminator™ protocols (Invitrogen, United State of America). Briefly, BigDye™ Terminator v1.1 cycle and Cento *et al.,* 2012 nested primers were thawed completely and stored on ice. The tubes were briefly vortexed for two seconds and centrifuged to collect contents at the bottom of the tubes. The sequencing master mix was prepared for forward reaction as follows: 1µL BigDye™ Terminator v1.1, 1µL F2 primer (3.2 µM), 1µL Sequencing buffer and 10µL water. Sequencing master mix for reverse reaction included 1µL BigDye™ Terminator v1.1, 1µL R2 primer (3.2 µM), 1µL Sequencing buffer and 10µL water. The cycling conditions for the BigDye™ Terminator v1.1 were as follows an initial denaturation step at 96˚C for 5 min, 25 cycles amplification at 96˚C for 10 sec, 50˚C for 5 sec and 60˚C for 4 min followed by a holding temperature at 4˚C.

Following the sequencing PCR, purification was performed on the sequencing reactions using BigDye XTerminator™. The bottle of BigDye XTerminator™ beads was vortexed for 8 to 10 seconds before mixing with the SAM solution. Briefly the SAM/BigDye XTerminator™ beads working solution was prepared (volume per 20µL reaction) as follows: 90µL SAM solution and 20µL BigDye XTerminator™ beads solution. Twenty microliters of the SAM/BigDye XTerminator™ bead working solution was transferred to each sample into the 96-well sequencing plate and mixed thoroughly. The plate was sealed with MicroAmp™ Clear Adhesive Film and vortexed for 20 minutes at 1800 rpm. The plate was then centrifuged at 1000 x g for two minutes. The plate was placed in a genetic sequencer ABI 3500 (South Africa) and Sanger sequencing protocol was selected.

NS3 sequence analysis was performed using multiple computer software programmes such as BioEdit Sequence Alignment Editor (Hall, 1999), ChromasPro version1.49 (Technelysium Pty Ltd, United State of America), DeepChek software (https://deepchek-portal.ablsa.com/ and MEGA version 7.1

(Tamura *et al.,* 2007). The study sequences were aligned and mutations were called using the three embedded algorithms used by the software namely: Geno2Pheno, International Antiviral Society and Lontok et al. (2012). Mutations were observed at a threshold 20%. The editing was performed by the ChromasPro programme where each consensus nucleotide sequence was viewed in order to correct any miscalling which was automatically identified by the software. The edited nucleotide sequences were aligned, together with HCV NS3 published reference sequences, (retrieved from GenBank; http://www.ncbi.nlm.nih.gov/genbank/. See Appendix B) using the BioEdit programme. The study sequences were compared with the HCV NS3 reference sequences for any nucleotide and amino acid variations through alignment. The Neighbour-Joining method on the MEGA programme was used to determine the genotypic lineage of the study sequences and generate a bootstrap consensus phylogenetic tree out of a 1000 bootstrap replicates (Tamura *et al.,* 2007).

2.8.2 Next Generation Sequencing (NGS)

NGS was carried out by the Sequencing Core Facility at the NICD in order to determine minor variants. NGS has high throughput than Sanger, resulting in a better analysis of the minor mutations. The positive amplicons were quantified using Invitrogen™ Quant-iT™ PicoGreen® dsDNA Assay (Invitrogen, United State of America) and dsDNA BR Assay kit (Life technologies, United State of America). All recovered DNA was pooled in equal volumes at a normalized concentration (2 ng/µl, the lowest concentration recovered). Library quality was determined using a Bioanalyzer 2100 (Agilent Technologies, United State of America) and was sequenced on a MiSeq deep sequencing platform (Illumina, United State of America) using the 500-cycle MiSeq reagent kit v2. Fastq files were generated using the on board MiSeq Reporter.

Multiplexed paired libraries were prepared using the Nextera XT DNA sample preparation kit (Illumina, United State of America). This was followed by 2x300-bp sequencing carried on Illumina Miseq platform. To test whether these 300-bp regions were sufficient to identify mixed genotypes, we developed a random

primer-based method to sequence HCV amplicons containing mixtures of two HCV genotypes in different ratios. We were able to determine the genotypes without ambiguity and to quantify the ratio of the abundances of the mixed genotypes in the samples. The paired ends were checked for quality, trimmed using the CLC Genomics workbench version 11.0 (CLC, Bio-QIAGEN, Denmark). The resultant contiguous sequences were mapped against HCV reference AF009606.1 (Genotype 1) and AF064490.1 (Genotype 5a) for the HCV NS3 protein. Variation analyses were executed by DeepChek software (ABL.SA, Luxembourg) and Qiagen CLC Genomics workbench version 11.0.DeepChek software (https://deepchek-portal.ablsa.com/) was used as previously outlined. Mutations were observed at threshold 5%. Those at threshold 1% were used to exclude potential false mutations associated with PCR and sequencing procedures.

The CLC variation analysis uses Low Frequency Variant Detection tool that relies on two models namely; Statistical model and model for sequencing errors. The statistical model does not make any assumptions about the ploidy of the sample. Instead, a statistical test is performed at each position. This may determine if the nucleotide observed in the reads, is due to sequencing errors, or whether the sequence in question has one (or more) alleles compared to the reference present at unknown frequency. If the latter is the case, a variant corresponding to the significant allele will be called a variant by the software, with an estimated frequency. The DNA

2.9 3-DIMENSIONS NS3 PROTEIN CRYSTALLOGRAPHY

Modeller version 1.7.0.0 was used to generate HCV NS3 homology protein models. The NS3 protein sample sequence [target sequence genotype 1(CVI0090) and genotype 5a (ASN3)] aligned and translated with BioEdit was uploaded on Protein Blast. This was done in order to identify the homologous protein (template) from the protein database (PDB) (Kouranov *et al.,* 2006). The suitable template as selected based on high query coverage, sequence identity, lowest expected value and Angstrom value of less than for four (Å). The lowest Å confirms the stability of the protein and E-value >0 is evolutionary significant.

The generated PBD file was downloaded and transferred to visualization software PYMOL molecular graphics system version 2.1. The labelling and analysis of the output visualization of the resistance mutations on determined drug target positions was carried through PYMOL software v 1.7.4.

2.10 EPITOPE MAPPING

The NS3 nucleotide sequence was translated to amino acids in frame using Mega version 7.1. Epitopes were described using web based prediction server Immune Epitope Database, (http://tools.immuneepitope.org/main/html/analysis_tools.html). Common HLA A and B alleles in the South African population were used as reference alleles; A30:01, A02:01, B58:02 and A01:01(Sidney et al., 2008; Wertheirmer et al., 2003). Different studies have identified epitopes within the NS3 protein. The most commonly published ones are listed (Table 2.3).

Table: 2.3 HLA class I restricted HCV NS3 immunodominant epitope sequences chosen from published studies.

Class I Epitopes	Sequence subtype	Restriction	References
NS3 1073-1081	CVNGVCWTV CINGVCWTV	A2	Wedemeyer *et al.,* 2002 Vertuani *et al.,* 2002
NS3 1406-1415	KLVALGINAV	A2	Wedemeyer *et al.,* 2002 Cucchhiarini *et al.,* 2000
NS3 1267-1275	LGFGAYMSK	A3	Chang *et al.,*2003

CHAPTER 3: RESULTS

3.1 PCR AMPLIFICATION OF HCV FRAGMENTS

Total RNA extracted from all 109 samples was subjected to conventional PCR using KOD as well as the DeepChek assay to amplify the NS3 protein as previously mentioned in section 2.5.2 and 2.5.3. Of the 63 genotype 1 samples, 22 (34.92%) were successfully amplified by KOD and 11 (17.46%) by DeepChek. Similarly, of the 46 genotype 5a samples, six (13.04%) were successfully amplified by DeepChek. No observed amplification of genotype 5a samples were generated with KOD for reasons unknown to us despite several troubleshooting approaches. The 775 bp amplified fragments were visualized on 1% agarose gel. Figure 3.1 below shows a representation of six amplicons.

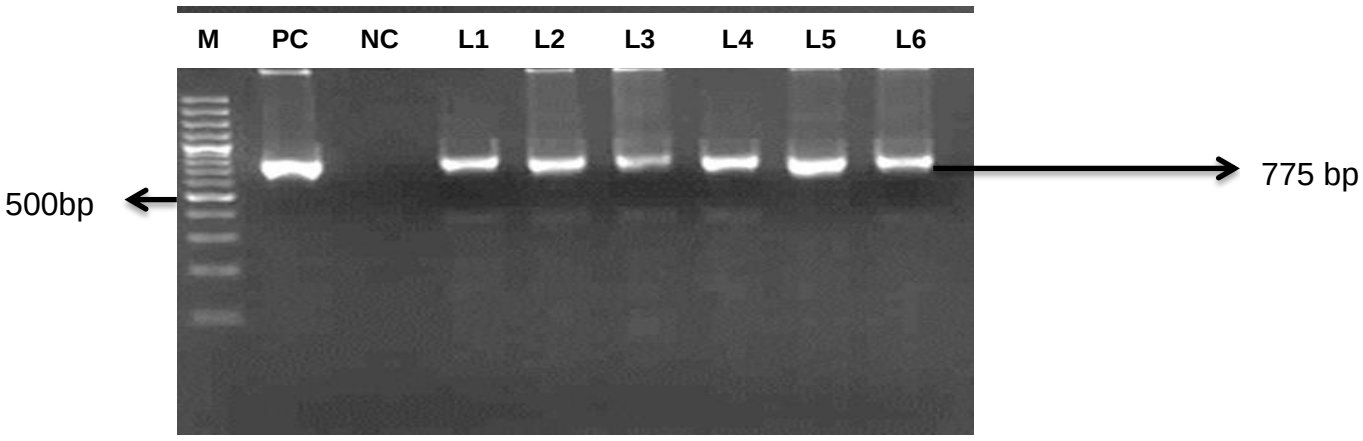

Figure 3.1: Agarose gel showing PCR amplified NS3 amplicons. L1 and L2 (genotype 1a), L3 and L4 (genotype 1b), L5 and L6 (genotype 5a). Molecular weight (M) = NEB 1kb, NC (Negative Control) and PC (Positive Control).

In all, a total of 33 genotype 1 and six genotype 5a samples were amplified as shown in Figure 3.2. All 39 amplicons were gel-purified and subjected to molecular cloning and sequencing.

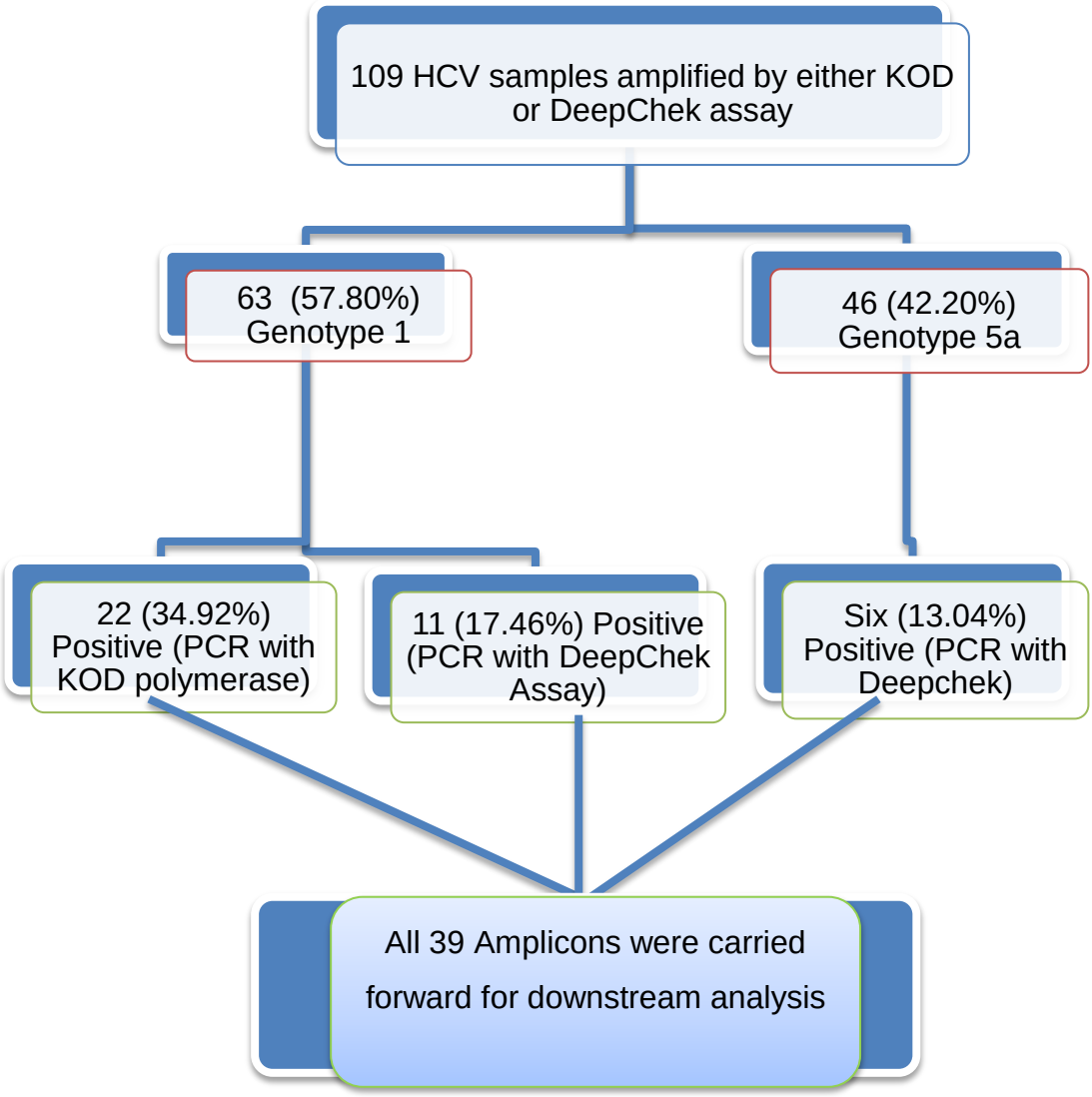

Figure 3.2: Flow diagram showing amplification results of the NS3 protein

3.2 MOLECULAR CLONING OF NESTED PCR AMPLICONS

All 39 samples positive by PCR were cloned into TOPO-TA vector as previously described (section 2.7). A minimum of 10 colonies from each genotype was screened for the presence of the NS3 gene. Overall, only 11 clones from all 33 amplicons for genotype 1 and six clones out of all six amplicons for genotype 5a carried the gene of interest. Due to the few number of positive clones obtained, sequencing of these would not provide sufficient data to generate consensus sequences for mutational analysis. Hence, direct sequencing and NGS was opted as method of choice for mutational analysis.

3.3 DIRECT SEQUENCING AND PHYLOGENETIC ANALYSIS

Of the 33 of the genotype 1 samples subjected to Sanger sequencing, 19 (57.57%) were positive for the NS3 gene. Similarly, five out of six (83.33%) genotype 5a samples were positive by Sanger. In total, 24 samples were successfully sequenced using the Sanger method. All 24 sequences were included in the phylogenetic analysis. All the phylogenetic trees depicted supported clusters with a bootstrap > 70%. Sample sequences of genotype 5a clustered with strains from SA, France and the USA and sample sequence SN943-5a formed an outlier (Figure 3.3).

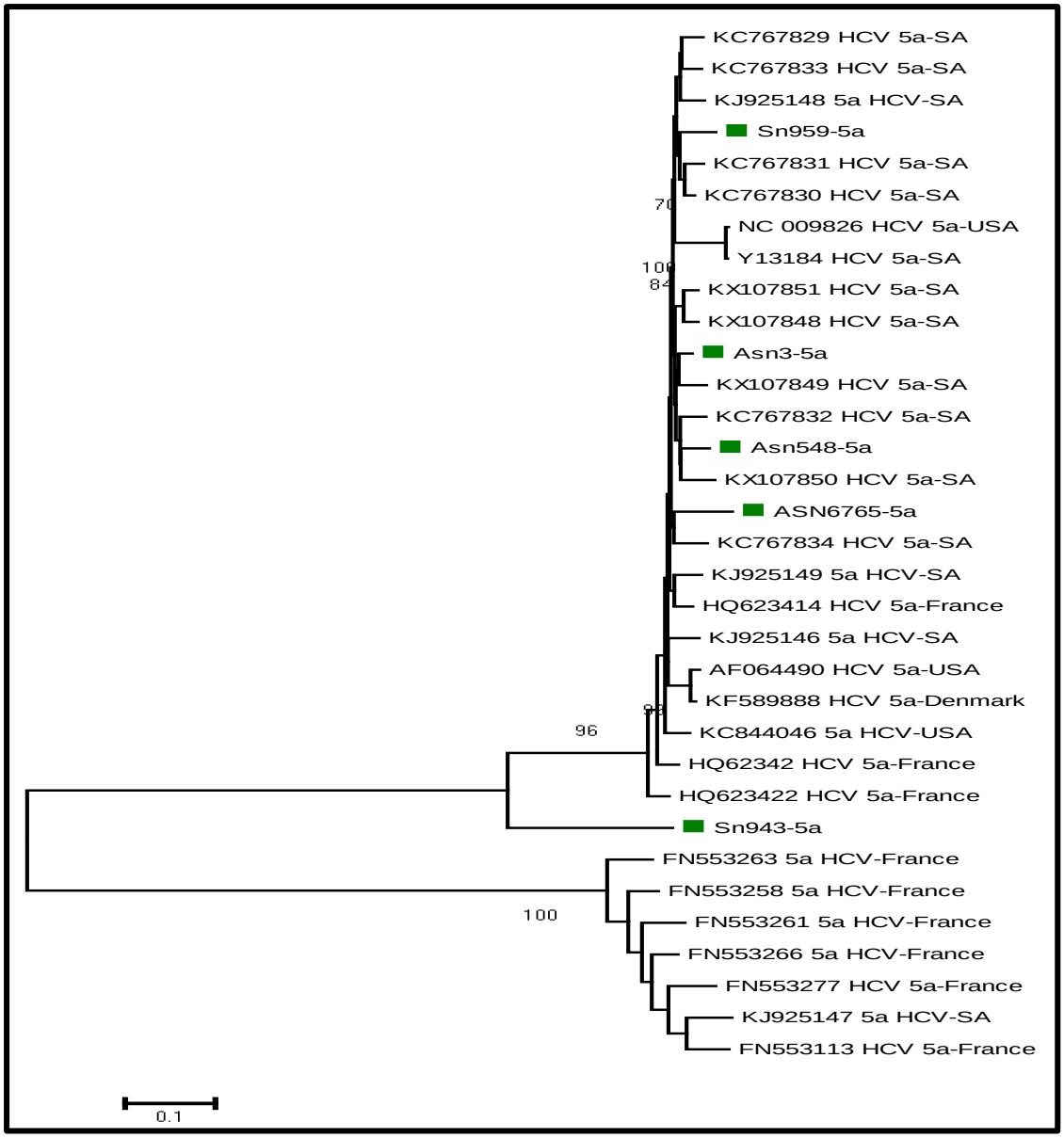

Figure 3.3: Dendrogram obtained by neighbour joining Phylogenetic analysis depicting relationship between subtype 5a GenBank sequences and study sequences. GenBank reference sequences are indicated by their accession numbers. ■ = sample sequences.

As observed, majority of the study isolates were subtype 1a and 1b clustering with strains from the USA, Australia and France. Both subtypes are globally distributed. Most of the sample sequences CVI1077, CVI0050, CVI0090, CVI0033, CVI0019, CVI0095, CVI0031, CVI0075, CVI599, CVI0018, HG04335838, HN01393031, HN01361785, CVI0051, MD6773, and CVI6550 of

genotype 1a clustered with multiple strains from different regions whilst the sample sequences CVI6861, CVI1087 and CVI6605 of genotype 1b clustered among strains from Japan and China with bootstrap values of 97%-98%, respectively (Figure 3.4).

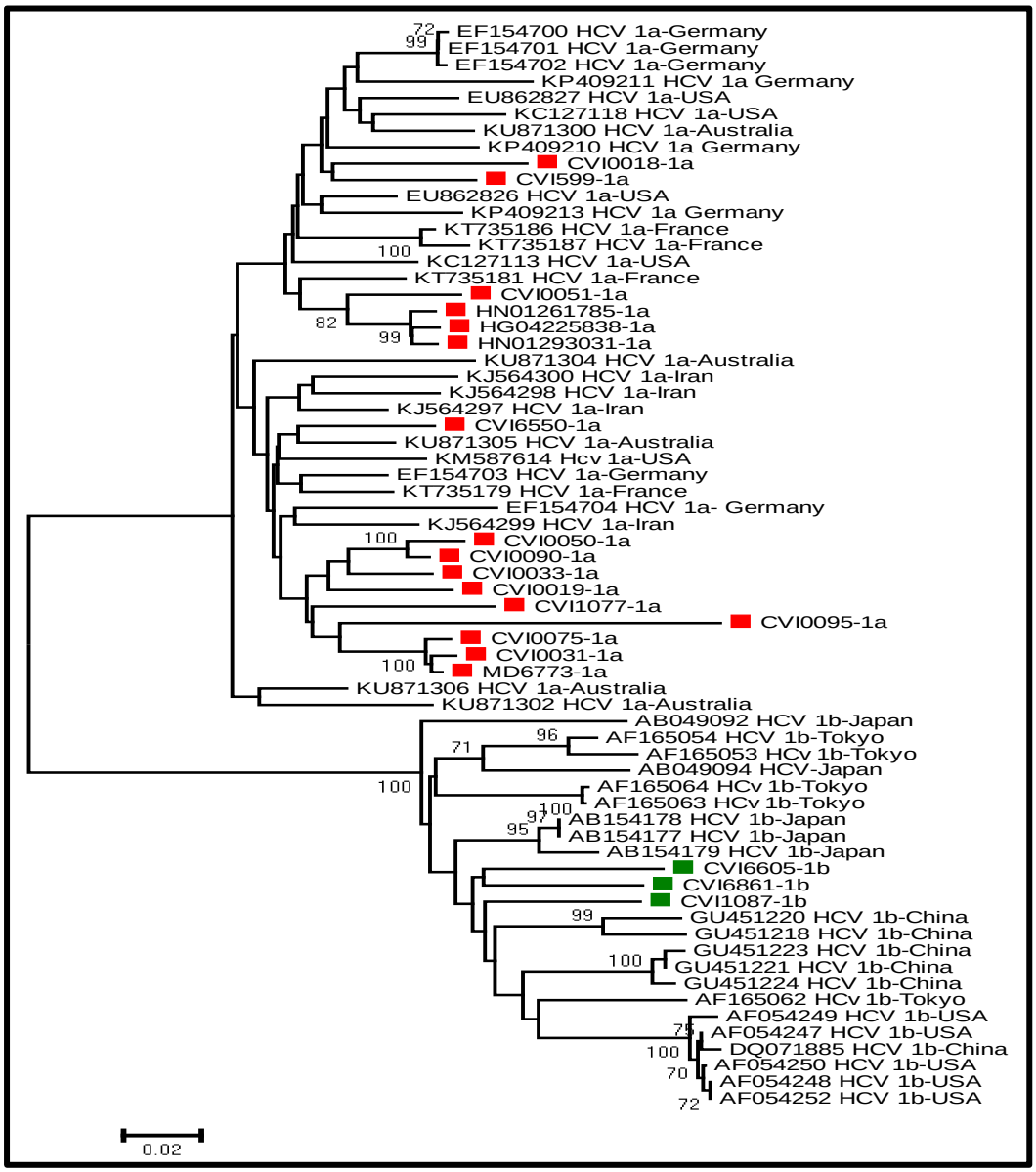

Figure 3.4: **Dendrogram obtained by neighbour joining phylogenetic analysis depicting relationship between subtype 1a and 1b GenBank sequences and study sequences**. (GenBank reference sequences are indicated by their accession number. Subtype 1a sample sequences = subtype 1b sample sequences = .

3.4 MUTATIONAL ANALYSIS

3.4.1 Sanger Sequencing

3.4.1.1 Nucleotide Sequence Variations

The nucleotide sequences were aligned with their relevant reference sequences for subtype 1a, subtype 1b and genotype 5a, in order to ascertain the nucleotide variations. The alignment is shown in Figures: 3.5, 3.6 and 3.7.

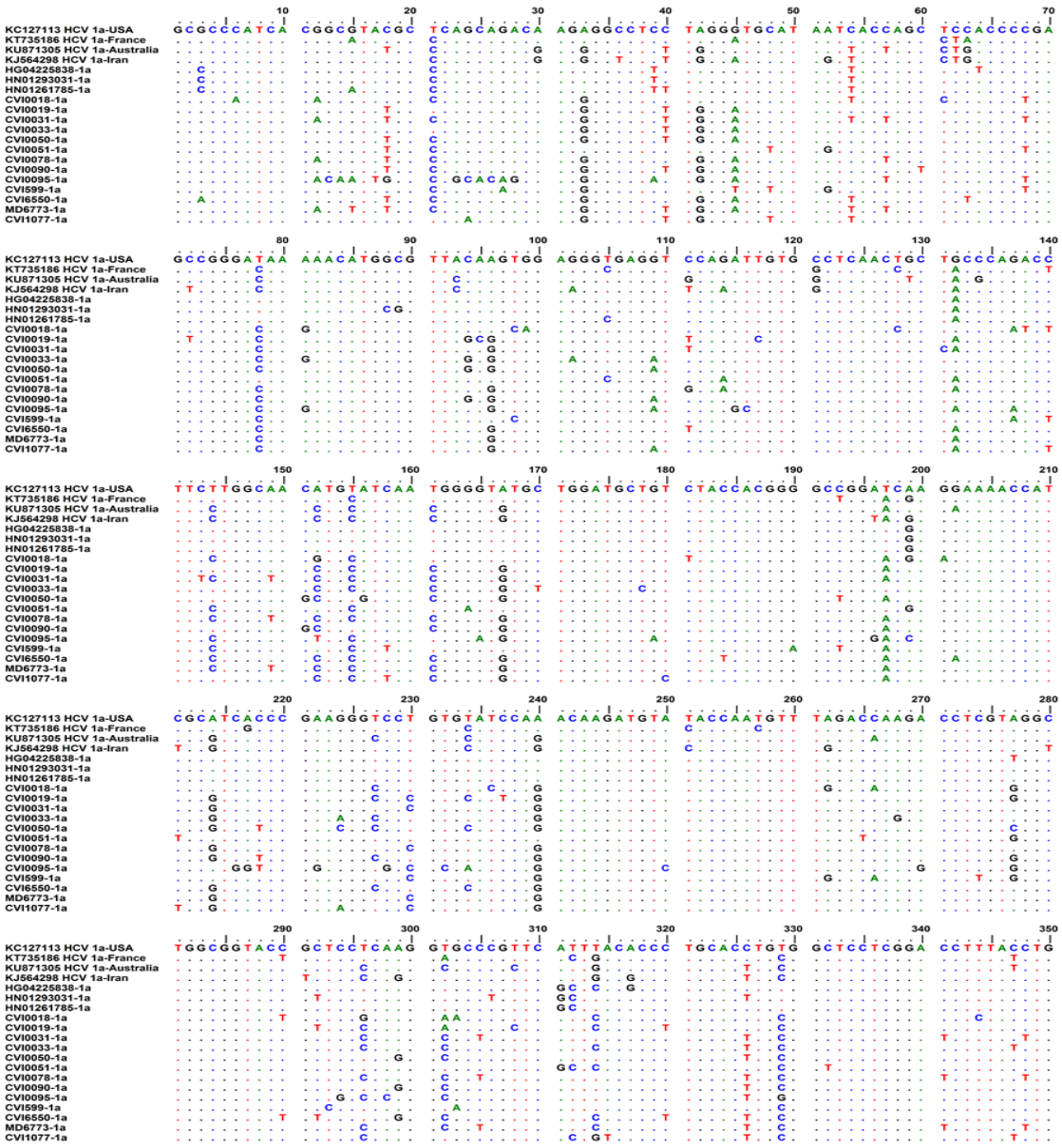

Figure 3.5: Nucleotide sequence alignment of subtype 1a [Reference sequences KC127113 KT735186, KU871305 and KJ564298]

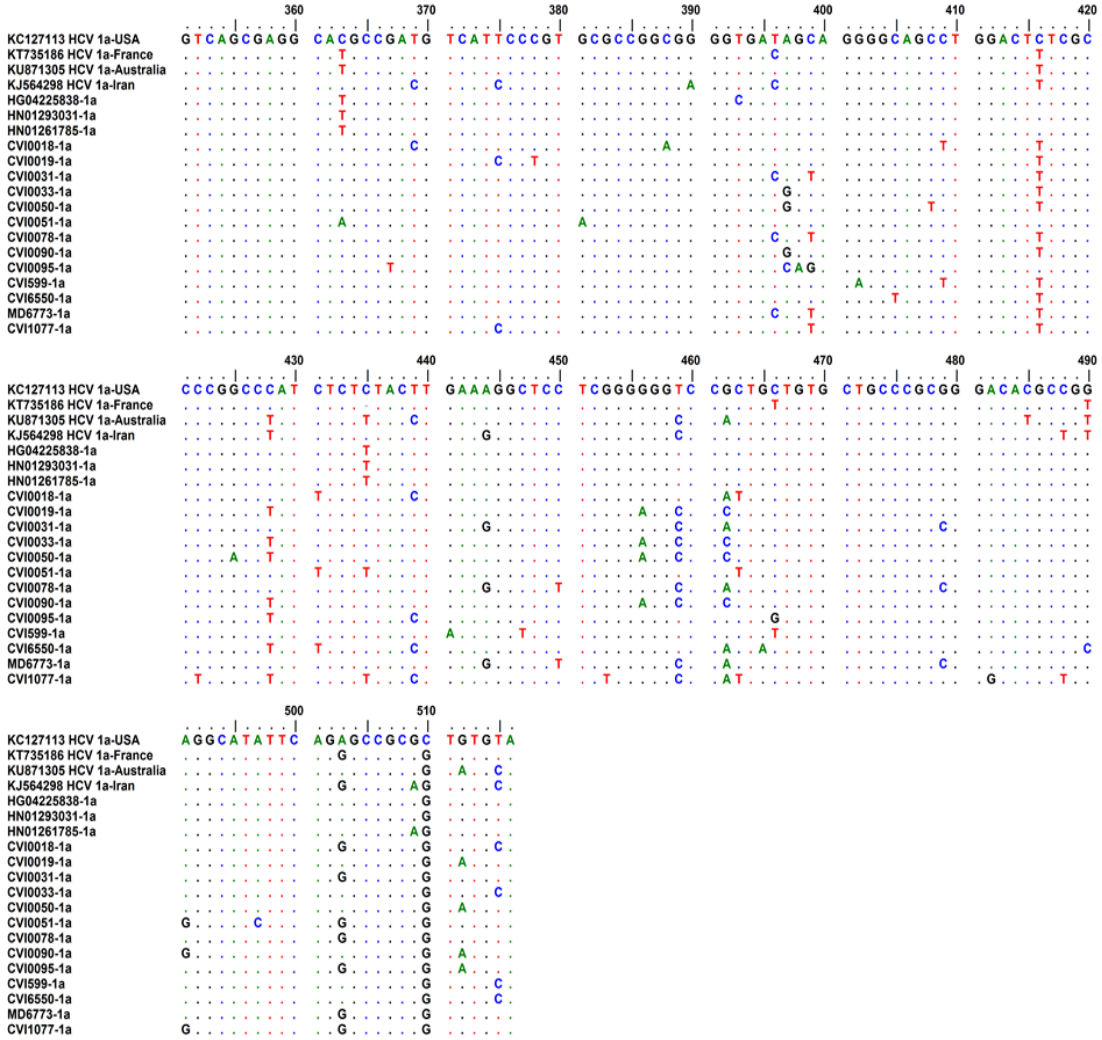

Figure 3.5: Nucleotide sequence alignment of subtype 1a [Reference sequences KC127113 KT735186, KU871305 and KJ564298] continuation from page 36.

Figure 3.6: Nucleotide sequence alignment of subtype 1b [Reference sequence AB049092].

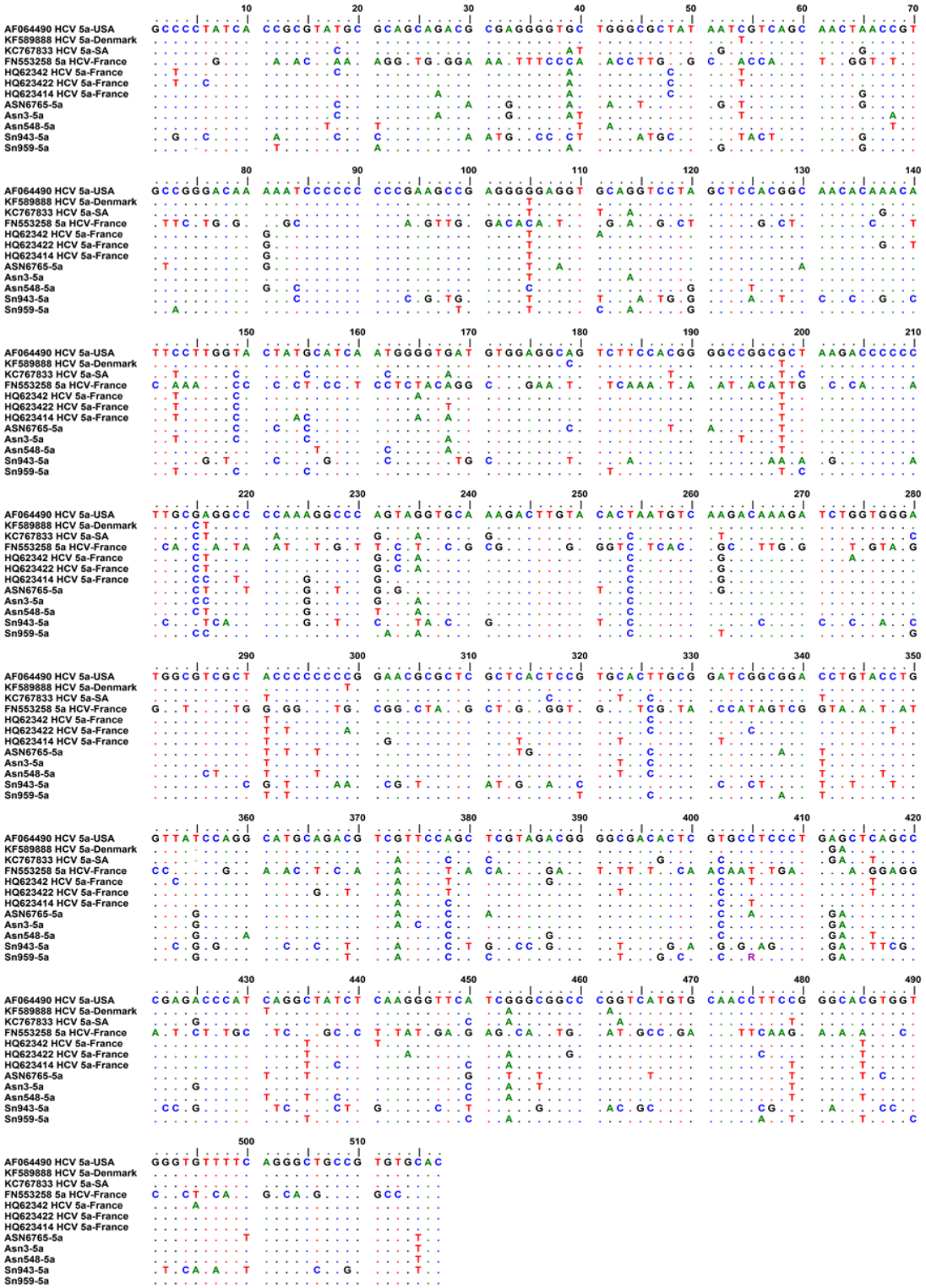

Figure 3.7: Nucleotide sequence alignment of genotype 5a [Reference sequences AF064490, KF589888, KC767833 and FN553258].

A summary of nucleotide variations detected for subtype 1a and subtype 1b are shown in Table 3.1 and Table 3.2, respectively. Subtype 1a and genotype 5a present common nonsense mutations at position 3 were G changes to C and C changes to T. At position 510 of subtype 1a, all samples have the same variant were C changes to G.

Table 3.1: Nucleotide variations detected within subtypes 1a.

Position	Nucleotide Change	Type of Mutation
3	G → C	S
6	C → A	S
14	C → A	NS
15	G → A	NS
18	C → T/G	NS
21	T → C	S
23	A → G	NS
26	A → C	NS
33	A → G	S
39	C → T/A	S
48	C → T	S
52	A → G	NS
54	C → T	S

Table 3.1 continues

Position	Nucleotide Change	Type of Mutation
88	G → C	NS
94	C → G	NS
105	T → C	S
132	G → A	NS
144	T → C	NS
262	A → G	NS
270	A → G	NS
303	G → A	NS
311	A → G	NS
320	C → T	NS
462	G → A/C/T	NS
512	G → A	NS

NS= nonsense; S= sense

Table 3.2: Nucleotide variations detected within subtypes 1b.

Position	Nucleotide Changes	Type of Mutation
21	A → G	NS
36	A → G	NS
37	C → A	S
49	G → A	NS
65	G → A	NS
75	C → T	S
81	T → C	S
111	A → G	S
117	C → T	S
128	A → G	NS
131	G → A	NS
143	G → C	NS
153	A → G	S
167	T → C	S
181	C → T	S
238	C → A	S
239	C → G	S
326	T → C	S
500	A → G	S

NS= nonsense; S= sense

Several nonsense mutations were observed for genotype 5a at different positions within the NS3 region. Table 3.3 below summarizes the mutations observed

Table 3.3: Nucleotide variations detected within genotype 5a.

Position	Nucleotide Variations	Type of mutation
3	C → G	S
6	T → C	S
12	C → A/T	S
17	A → T	NS
18	T → C/A	NS
21	G → T/C/A	S
27	G → A	S
30	G → A	S
33	A → G	NS
45	C → T/A	S
54	C → T/ A	S
81	T → G	NS
102	G → A	NS
108	C → T/A	NS
114	G → A	S
140	A → T/C	NS
162	T → C	S
200	T → C/G	NS
235	G → A/T	NS
305	G → T	NS
326	T → C	NS
385	C → A/T	NS
409	A → T	NS
453	T → C/A	NS
466	G → T/C/A	NS
475	G → A	NS
510	G → A	S

NS = Nonsense; S = Sense

3.4.1.2 Amino acid Sequence Variations

Majority of the mutations were observed between positions 51-110, covering the catalytic triad region of the NS3 protease in both subtype 1a and 1b (Figure 3.8 and Figure 3.9). Genotype 5a presents most mutations between position 61 and 170 when compared to subtype 1a and subtype 1b. The most commonly reported mutation Q80R is presented in subtype 1b; similarly genotype 5a harbours a Q80P mutation at the same position (Figure 3.9 and 3.10). Sample SN943 harbours (75/161) 46.58% of the mutations observed in genotype 5a when compare to other sample sequences ASN6765 (30/161) 18.63%, ASN3 (20/161) 12.42%, ASN548 (24/161) 14.90% and ASN959 (27/161) 16.77% respectively (Figure 3.10).

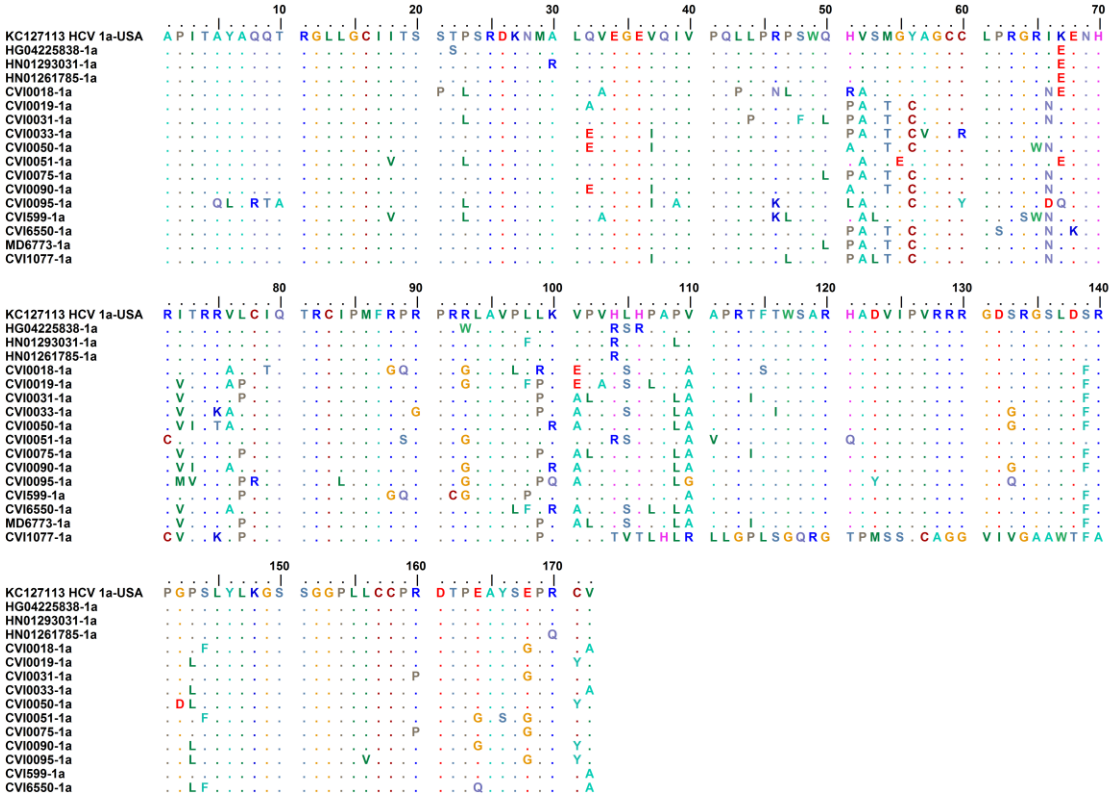

Figure 3.8: Multiple sequence alignment showing amino acid changes in subtype 1a.

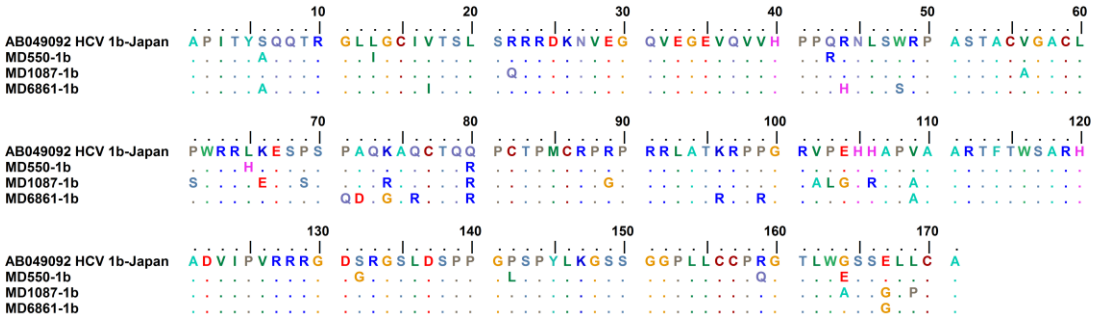

Figure 3.9: Multiple sequence alignment showing amino acid changes in subtype 1b

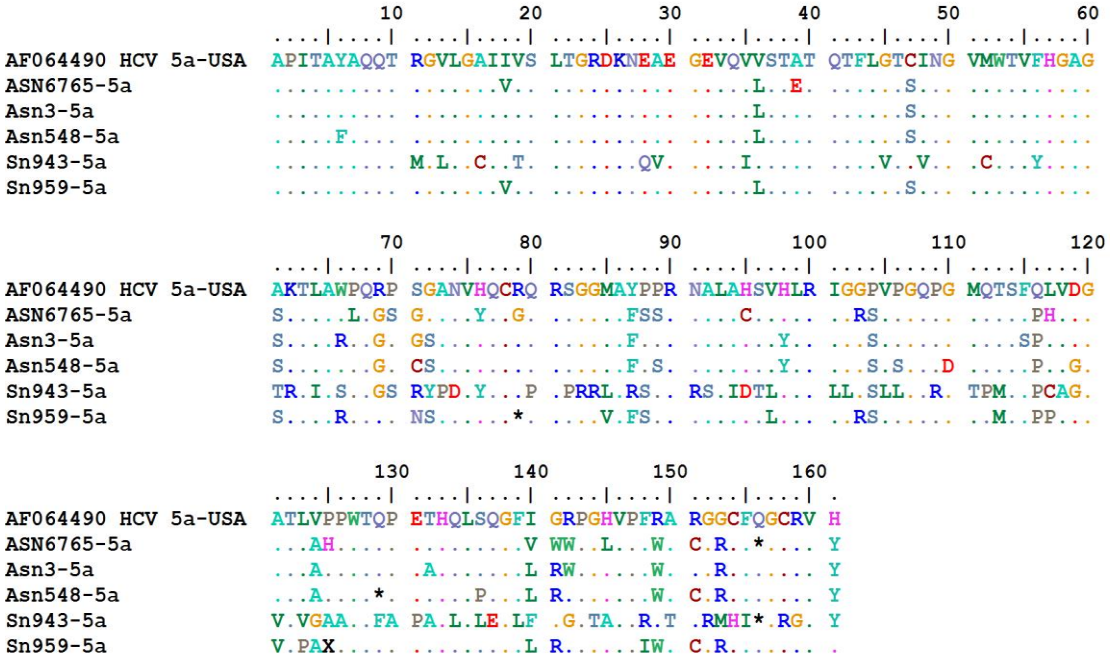

Figure 3.10: Multiple sequence alignment showing amino acid changes in genotype 5a.

A total of 16 subtype 1a sequences were analysed and three for subtype 1b. Mutation S/Y14 accounts for 25% of subtype 1a. Mutation Q80 is common in all subtype 1b samples with Glutamine changing to Proline. Mutation P107L/S is only present in subtype 1a with 12.50%. Mutations around the catalytic triad might affect the replication process of the protein. The table below (Table 3.4) shows a summary of mutations observed in NS3 Sanger generated sequences.

Table 3.4: Polymorphic sites of HCV genotypes NS3 protease region called by KOD enzyme in – house PCR and sequenced by Sanger.

NS3 Position	**Amino acid**	**Codon Frequencies of codon in each genotype%**	
		N =16 1a	**N = 3 1b**
A5	Q	1(6.25%)	-
S/Y6	L	1(6.25%)	2(66.66%)
W48	S	-	1(33.33%)
Q50	L	3(100%)	-
H51	R/P/L/A	11(68.75%)	-
Y56	C	10(62.50%)	1(33.33%)-
C60	R/Y	2(12.50 %)	-
Q80	R	-	3(100%)
I84	L	1(-6.25%)	-
R88	G	2(12.50 %)	-
R90	G	1(12.50 %)	-
P107	L	2(12.50%)	-
F115	S/L	2(12.50%)	-
R130	G	1(12.50%)	-
P142	L	-	1(33.33%)
G146	D	1(12.50%)	-

Mutation D168Y was observed in both subtype 1b and subtype 5a. Those with this mutation have reduced susceptibility to Faldaprevir, Simeprevir and Vaniprevir (Table 3.5).

Table 3.5 Polymorphic sites of HCV genotypes NS3 protease region amplified by DeepChek assay and sequenced by Sanger.

NS3 position	Amino acid	Codon Frequencies of codon in each genotype%		
		1a n=16	**1b n= 3**	**5a n=5**
I18	V	-	1 (33.33 %)	3 (60.0%)
Q28	E	3 (18.75 %)	-	-
V33	I	4 (25.00%)	-	-
T40	A	3 (18.75%)	-	-
Q41	H	-	1(33.33%)	-
C47	S	-	-	4 (80.0%)
V48	I	-	1(33.33%)	-
T61	S	16 (100%)	1(33.33%)	-
A61	S	-	-	4 (80.0%)
A66	G	-	3(100%)	
P67	Q	2(12.50%)	2(66.66%)	1(20.0%)
K68	R	-	2(66.66%)	-
V71	I	-	1(33.33%)	2(40.0%)
T87	S	1(6.25%)	-	4(80.0%)
K87	A	-	3(100%)	-
V114	I	-	1(33.33 %)	3(60.0%)
S122	G/A	4(25.00%)	1(33.33%)	-
T122	A	-	-	2(40.0%)
F147	S	-	3(100%)	-
A150	V\L	-	2(66.66%)	-
V151	A	-	-	2(40.0%)
L153	I	5(31.25%)	-	-
D168	**Y**	-	1(33.33%)	1(20.0%)
A170	**V**	2(12.50%)	-	-
I170	I/F/T	-	2(66.66%)	-
N174	**S**	3(18.75%)	-	-
H195	Q	4(25.00%)	2(66.66%)	4(80.0%)
S196	T	4(25.00%)	2(66.66 %)	1(20.0%)
F197	Y	16(100%)	2(66.66%)	1(20.0%)
T200	A	-	-	1(20.0%)
H201	V/F	16(100%)	-	1(20.0%)
L202	I	16(100%)	-	-
P203	H	-	-	2(40.0%)
H203	A	5(31.25%)	-	-
A204	T/V	16(100%)	-	3(60.0%)
P205	F/S	2(12.50%)	-	-
G207	K	1(6.25%)	-	3(60.0%)
R208	S	-	-	1(20.0%)
K210	H	-	-	2(40.0%)
213 deletion	**-**	-	-	1(20.0%)
P215	F	-	-	2(40.0%)

Bolded are the resistance mutations.

3.4.2 Next generation analysis

All the six pools had more than one genotype (Figure 3.12 and 3.13). An average of 419 657 reads were generated for this analysis. Reads were randomly analysed across the pools. Pool 1A had 301 543 of genotype 1a and 298 977 genotype 5a, Pool 1B 252 11 88 genotype 1a and 46 742 genotype 5a. Pool 2A, 2B, 3A, 3B, 6A and 6B represented 100% of genotype 1a with 304 7549, 300 235, 164 185, 164 367, 382 100 & 37 622 respectively. Pool 4A and 4B had major representation of genotype 5a with 176 426 & 176 418, genotype 1a only accounted for 6975 & 6981reads between the pools. Pool 5 had a balance share of the genotypes with Pool 5A comprising of 114 343 genotype 1a, 181 19 15 genotype 5a and Pool 5B 108 085 genotype 1a, 176 153 genotype 5a. All the pools had 100% confidence level. This confirms the good quality of the generated reads.

In general we observed fairly concordance between major mutations by Sanger and NGS with an exception of Q80K/P detected at low depth coverage by NGS. Mutations V36L/Q80K/S122R/D168G/N174S were observed at a frequency of 1.24%. Mutations S91A/R92C/L153I/V248I/P264S and S332A were observed at a higher prevalence of 98.14% (Table 3.6) when observed with NGS. Position 120 across the pools harboured a frameshift mutation, either deletion or insertion. However the mutation is not limited to that position. Pool 5-6 had less frameshift mutation as compared to Pools 1-4. An insertion of A/T at position 3285 – 3286 was found in all pools. Majority of the variations are single nucleotide (SN). Pool 1 has the most Non-Synonymous (NS) mutations. Areas that are highly variable present most SN mutations.

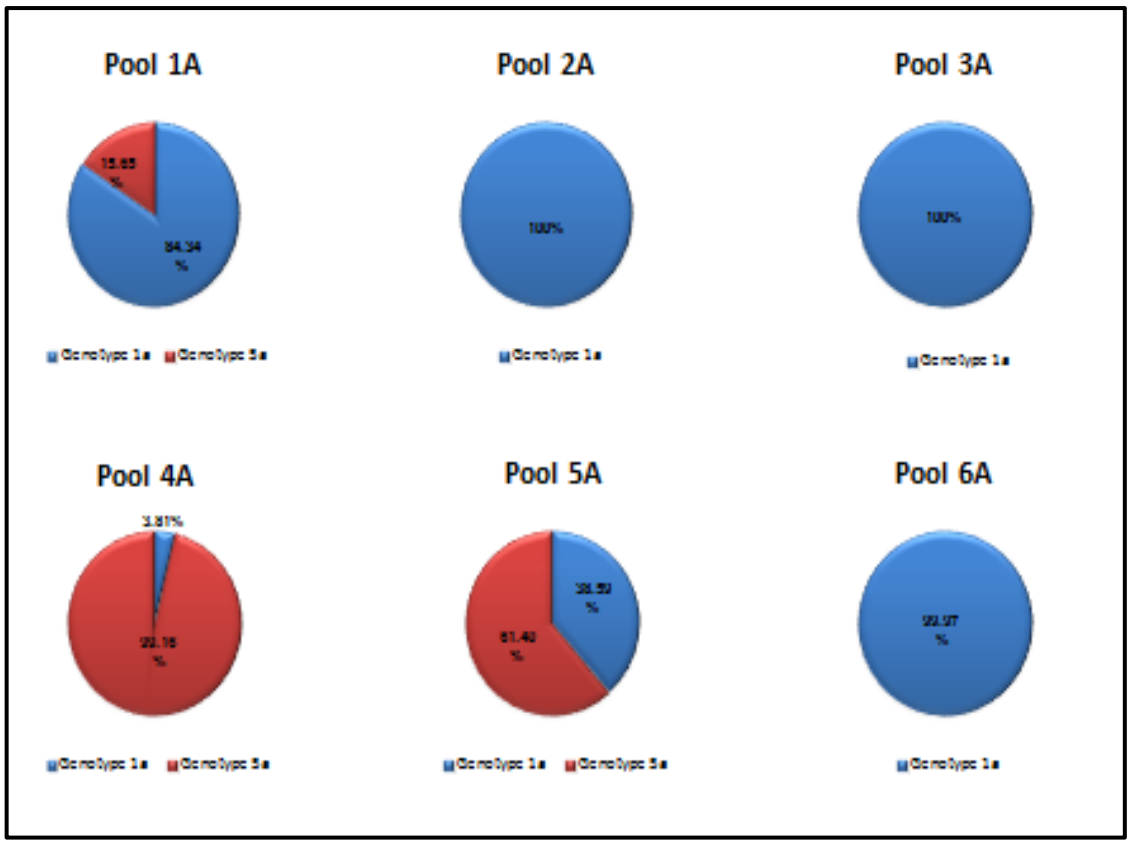

Figure 3:11(a): Next generation sequence analysis of NS3 protein. Pools showing percentage of reads generated, containing subtype 1a and/or 1b and/or 5a.

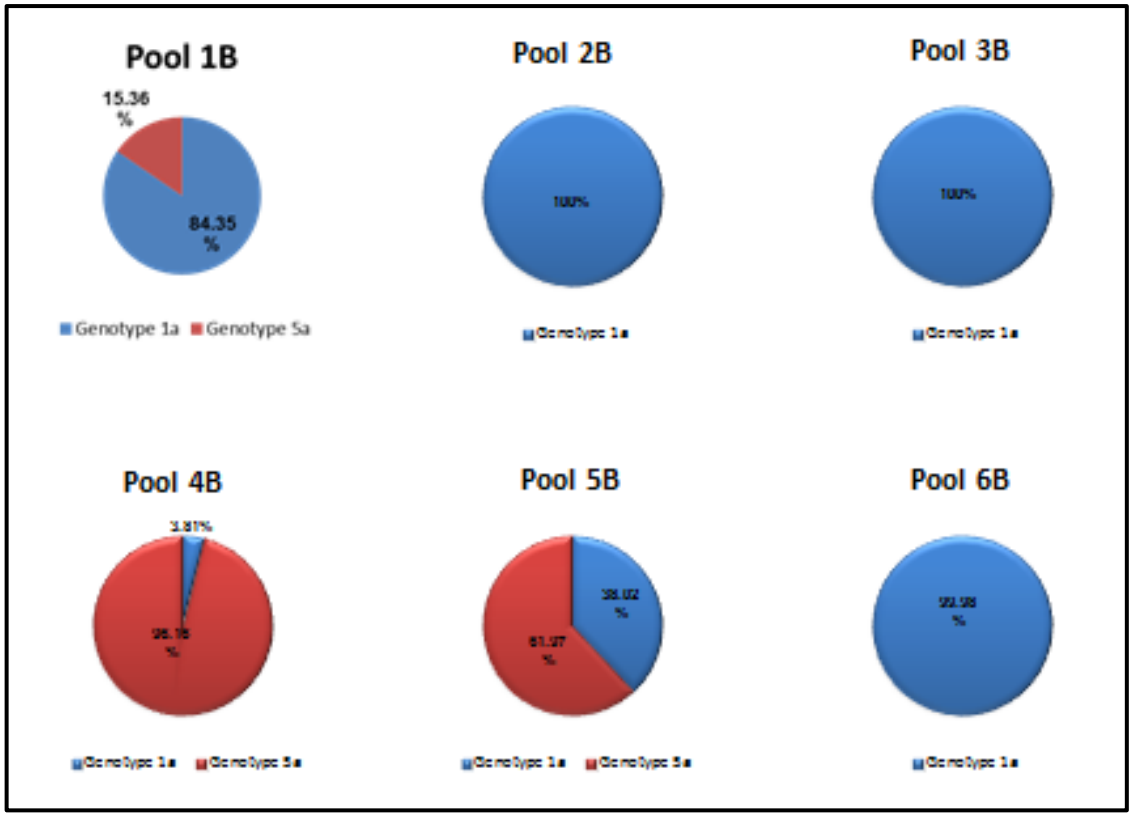

Figure 3:11(b): Next generation sequence analysis of NS3 protein, Pools showing percentage of reads generated, containing subtype 1a and/or 1b and/or 5a.

Table 3.6: Summary of mutation generated by DeepChek (NGS)

Position	Amino acid changes	Prevalence %
16	C→ G	1.79%
20	S→ R	1.17%
22	T→ A	1.05%
25	D→ E	2.08%
28	Q→ P	1.66%
29	V→ G	1.41%
30	E→ G	2.86%
33	V→ G	1.64%
36	**V→ L**	1.24%
37	S→ A	1.35%
41	Q→ P	1.40%
42	T→ P	1.80%
43	F→ L	1.91%
47	C→ G	1.42%
48	I→S	1.44%
49	I→T	1.06%
51	N→ K	1.07%
54	V→ G	1.26%
55	T →P	1.30%
56	**V→ G**	1.10%
57	Y→ H	1.33%
61	H→ P	1.12%
69	T→ S	1.69%
75	C→W	97.56%
79	D→ A	1.42%
80	**Q →K**	3.35%

Table: 3.6 continues

Position	Amino acid changes	Prevalence
85	W → G	1.21%
89	Q → H	1.03%
101	R → C	94.36%
103	D → A	1.89%
110	H → R	1.15%
121	D → A	1.21%
122	S → R	1.65%
123	P → Q	1.30%
127	R → G	1.61%
129	L → P	2.51%
132	I → L	3.73%
134	V → D	1.11%
135	L → V	1.44%
143	L → R	1.47%
145	C → G	1.52%
153	L → I	94.98%
155	R → S	1.41%
158	**V → G**	1.86%
159	C → G	1.66%
163	V → G	1.79%
168	**D → G**	1.31%
174	**N → S**	1.22%
175	L → P	1.00%
176	E → G	1.05%
181	S → A	1.35%
184	F → V	1.04%
189	S → P	1.60%

Table 3.6: continues

Position	Amino acid changes	Prevalence
197	F → L	1.40%
211	S → R	1.47%
214	V → A	1.15%
224	V → G	1.08%
226	N → D	2.81%
228	L → R	2.17%
229	N → H	2.67%
230	P → S	2.20%
231	S → A	1.55%
235	V → I	1.98%
236	S → A	2.00%
238	F → Y	3.33%
241	Y → S	1.79%
242	M → L	2.02%
243	S → A	1.59%
244	K → Q	1.39%
246	P → S	95.62%
252	I → L	5.05%
264	K → N	87.30%
332	S → A	96.77%
339	F → V	1.50%
362	L → A	3.02%
388	P → S	1.98%
390	Y → D	100.00%
399	V → I	1.36%
402	H → R	98.62%
410	F → Y	66.23%
418	T → S	2.35%

Highlighted are the resistance mutations

3.4.3 Three dimension protein structure analysis

The NS3/4A protease domain adopts a tertiary fold characteristic of serine proteases of the chymotrypsin family (Meylan et al., 2005). The use of the 3-D protein structure is to predict protein function and determine the binding of the DAA. The exercise on the 3-D crystallography was performed on genotype 5a and genotype 1 generated Sanger sequences to show how the different polymorphisms can be identified by 3-D within the NS3 protein, genotype 1 structure is not included. The polymorphisms are shown in red appearing on the active site and around the catalytic triad area of the protein (Figure 3.11). Their presence may affect the electrostatic network of the protein thus reducing the binding affinity of the DAA. Further studies should be encouraged for other genotypes to establish trends and highlight different characteristics of the protein.

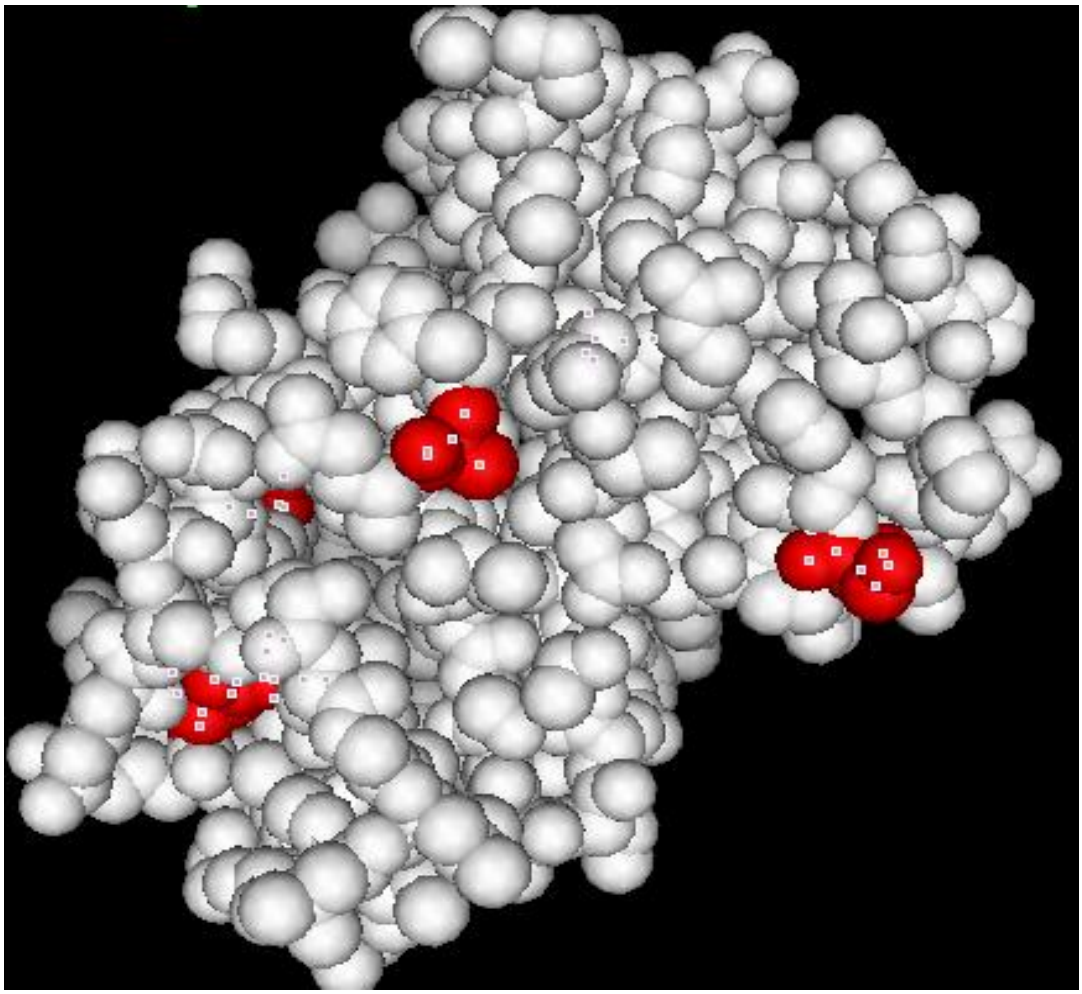

Figure 3.12: Crystallography of NS3 protein structurer. Highlighted are positions of polymorphisms (C16G, I18V, S20R, Y57H, A61S, S91A, D121A and V151A).

The protein sequences generated through Sanger sequencing were used to determine the epitopes within the HCV NS3 protein. Epitope FLGTSINGV observed in both genotype 1b and 5a had high binding affinity to this allele. This epitope may prove to be a good candidate for vaccine development as it protects against more HCV strains (Table 3.7). Epitopes CINGVCWTV, CVNGVCWTV and LGFAGAYMSK were previous reported by Vertuani *et al.,* 2002; Wedemeyer *et al.,* 2002 and Chang *et al*., 1999. Allele A*02:01 has most high binding epitopes.

Table 3.7: Epitope mapping showing binding affinity of the "newly predicted" study epitopes determined by the IEDB server.

Protein	Epitope sequence	Genotype of Epitope	Subtypes Allele type	A01 A*01:01	A02 A*02:01	A03 A*30:01	B58 B*58:01
NS3	**CINGVCWTV**	1a		12560.98	2235.11	8952.23	62.33
	CVNGVCWTV LGFAGAYMSK	1a & 5a		2013.26	44253.54	36.10	96214.05
	RLLGGSAAV	1a		24066.28	**15.05**	221956.87	19520.18
	LLGCIITSL	1a		15428.83	**34.15**	2668.71	34853.78
	IASPEGSCY	1a		4627.70	36062.17	14695.13	30718.42
	CTCGSADLY	5a		**20.67**	27180.84	15796.56	23535.29
	VMWTVFHGA	5a		-	**13.44**	6006.88	35599.26
	GSKTLAGSK	1a & 5a		28818.44	39698.06	57.74	34691.62
	HSRIPGLSP	5a		30577.50	34231.52	**44.99**	35908.02
	FLGTSINGV	1b & 5a		11656.58	**7.32**	21162.93	38090.30
	GLFGWSAAL	1b		22072.90	**8.42**	13616.17	-

Bolded are the Epitopes that have high binding affinity ANN<50 IC50nM. Those that are ANN>50 IC50Nm =Intermediate affinity and above ANN>500 IC50Nm= poor binders. The bolded and highlighted epitope sequences were previously predicted.

CHAPTER 4: DISCUSSION

Like other RNA viruses, HCV replication is characterised by a high replicating rate producing ~$10^{10\ -12}$ particles per day. HCV is error prone because the RNA-dependent RNA polymerase enzyme lacks the proofreading ability. The error rate is ~10^{-4} for a single mutation (Neumann *et al.,* 2008). HCV has a high genetic diversity, which classifies the virus into seven genotypes and multiple subtypes (Smith *et al.,* 2014). Subtype 5a was found predominantly in South Africa. In the NS3 region, eighty percent of subtype 5a study sequences formed a cluster with reference sequences from South Africa and the United State of America (Figure 3.3), whereas reference sequences from France clustered together. Sample sequence SN943-5a had formed an outlier from other study sequences and reference sequences., suggesting a different type of strain within the genotype 5 family, however further analyses in other regions of the genome is needed to conclude this. One hundred and sixty one mutations were called in genotype 5a analysis with sample SN943 harbouring (75/161) 46.58% of the mutations observed and ASN6765 (30/161) 18.63%, ASN3 (20/161)12.42%, ASN548 (24/161) 14.90% and ASN959 (27/161) 16.77% respectively (Figure 3.10). This as well might have contributed to the discordant topology of the sample. It was previously thought that genotype 5 was only specific to South Africa; however, Verbeeck *et al.* (2006) reported occurrences of genotype 5a infections in North Africa and Europe. This indicates that there is global movement of HCV genotypes, with travel for work and leisure contributing to the spread.

HCV genotypes are regionally predominant with genotype 1b found globally and subtype 1a found in Europe and the Americas. Genotypes 1-4 are also found in South Africa (Gededzha *et al.,* 2012; Prabdial Sing *et al.,* 2008). Our subtype 1b study sequences clustered with reference sequences from France, Australia and China, supporting the global distribution of subtypes (Figure 3.4). Sample sequences CVI0031, CVI0073 form their own cluster with 99% bootstrap with subtype 1b study sequence. Looking back at the heterogeneity of subtypes being different by >20%, this confirms previous findings (Simmonds *et al.,* 2004). Subtype 1a sample sequences were observed mainly among USA, Germany and Japan reference sequences. In the

USA more than 1.9% of subtype 1a is found among HCV infected individuals. It is known that blood and blood products were imported from the USA prior 1995 and since screening of blood and blood products was not mandatory then, HCV transmission via this route did occur. It is important to globally trace subtype 1a as it is associated with diverse virological response to available DAAs. Gutierrez and Wyles (2014) have linked subtype 1a with low virological response when compared to subtype 1b.

All genotypes found by sequencing in this study correlated with the LiPA results. The amplification of genotype 5a was achieved by the use of DeepChek enzyme, which is pan-genotypic. This suggests that DeepChek is more sensitive to amplify genotype 5a than the KOD enzyme.

This is the first study in South Africa to report on samples from treatment-naïve patients infected with either genotype 1 or genotype 5. Mutation occurs when the genetic material is changed in such a way that alters the transcribed message carried by the genetic material. Genotype 1 and genotype 5a have silent mutations at the beginning of the protein namely G3C, C3G, C6A, T6C and C12A/T (Table 3.1, Table 3.2 and Table 3.3). Silent mutations may affect the phenotype of a protein, reducing or increasing the rate of synthesis, however due to the redundancy of the protein these mutations might have no effect on the protein function. A nonsense mutation is defined as a point mutation in a DNA sequence resulting in a premature stop codon. Three nonsense mutations are observed in genotype 5a and none in both subtype 1a and subtype 1b. Both C14A + G15A code for the same missense mutation in subtype 1a and C238A + C239G both code for the same silent mutation in subtype 1b. In this study, known RAMs were rarely seen (Table 3.4). However observation of non-synonymous (NS) mutations was made.

The findings support previous research conducted in Brazil and other countries (Carvalho *et al.,* 2014; Hoffmann *et al.,* 2013; Nishiya *et al.,* 2010; da Silva *et al.,* 2014). Analysis of RAMs using BioEdit software yielded less observation (Table 3.4) whilst DeepChek software provided deeper analysis (Table 3.5). Mutation D168Y was observed in subtype 1b with both Sanger and NGS. Sanger sequencing analysis was achieved at 20% frequency. It is documented that individuals with this mutation have shown reduced susceptibility to Faldaprevir, Simeprevir and Vaniprevir (Everson *et al.,* 2012 and Takayama et al., 2004). Patients carrying both V71I and D168Y are linked to

the reduction of Danoprevir potency (European Association on the study of Liver, 2017). However, in this study such combinations were not observed.

Mutation I18V was observed in both subtype 1b and subtype 5a with a 33.33% and 60% presence respectively (Table 3.5). In subtype 5a, mutation I18V compensate for the fitness of R155K substitution (Serre *et al.,* 2016). About 25% of subtype 1a and 33.33% of subtype 1b harboured mutation S122A/G that shows reduced susceptibility of Simeprevir (Wang *et al.,* 2015). Although mutant R155K was not observed in this study, when paired with I170T (detected with 66.66 % in subtype 1b of our patients), it confers cross-resistance to both Boceprevir and Telaprevir (Zeuzem *et al.,* 2005a; Tonga *et al.,* 2008). Mutation A150T/V arises from Telaprevir and Boceprevir treatment and potentially causes cross-resistance to Simeprevir (Everson *et al.,* 2012). Polymorphism Q80K/R is common in subtype 1a treatment naïve individuals in other countries (Sarrazin *et al.,* 2015; Shepherd *et al.,* 2014) but was not observed by Sanger sequencing in this study. It is known to exist in 20% of genotype 1a carriers (Bae *et al.,* 2010). Polymorphism is defined as an occurrence in which genetic variation within a population can operate upon natural selection. Q80K/R polymorphism is associated with reduced susceptibility to Simeprevir by 10 fold (Bae *et al.,* 2010). It must be noted that the presence of the Q80R does not necessarily result in treatment failure whenever Simeprevir is administered. For instance, Q80R has no influential activity when Simeprevir is paired with Sofosbuvir (Lawitz *et al.,* 2014). However the impact of this polymorphism is not known in those infected with subtype 5a strain (Figure 3.10).

In this study, mutation Q80P is observed in a patient SN943 infected with subtype 5a. The clinical implication of this point mutation still needs to be determined. Yan *et al.,* 2017 described M21T mutation that resides in the N-terminal helix-α_o of the NS3 protein. This mutation enhances viral assembly (Brass *et al.,* 2008 and Horner *et al.,* 2012). Instead of M21T, S21P has been observed in our subtype 1a study sequences. More research needs to be conducted in order to understand the impact of this mutation (Figure 3.10).

A point mutation from methionine to a proline amino acid residue at position 21 (M21P) observed in other studies has been linked to the disruption of the membrane association of the protease. This leads to the degradation of the NS3 (He *et al.,* 2012).

RAMs disrupt the electrostatic network of the protein thus reducing the affinity of the PIs. The protein tertiary fold is a character of the serine protease of the chymotrypsin family. Our study has shown a number of NS3 mutations that are not yet characterised especially from genotype 5a.

The 3D genotype 5a structure for this study was modelled and solved with an overall root mean square deviation (rmsd) of 0.28Å confirming the stability of the protein (Figure 3.11). The use of the 3-D protein structure is to predict protein function and determine the binding of the DAA. Mutations were observed around the catalytic triad of the protein. More studies should be conducted for other genotypes to establish trends and highlight different characteristics of the NS3 protein. This was the first time a modelling exercise was performed for genotype 5a NS3 protein sequence from a treatment naïve individual.

Although the cloning process formed an integrative learning experience, the procedure was discontinued, due to the few number of positive clones obtained, sequencing of these would not provide sufficient data to generate consensus sequences for mutational analysis. It was also tedious and time consuming. With the advent of more sensitive and less timeous methods for determining the presence of mutations in the HCV genome, in particular, a commercial assay that allows for NGS data was opted in order to characterize the minor mutations. Results from the NGS reported mutation V36L/G and S122R that were not observed by Sanger sequencing (Table 3.6). NGS results were obtained at 5% frequency supporting the specificity of the NGS method as compared to the conventional sequencing.

In general we observed fairly concordance between major mutations by Sanger and NGS. Patients carrying mutation V36L are prone to have reduced susceptibility to Asunaprevir especially those infected by subtype 1b (Bartel *et al.,* 2008 and Kieffer *et al.,* 2007). Due to the fitness cost of the resistance mutations, patients presenting DAA resistance mutations as the dominating viral population is rarely reported. There are currently no studies in South Africa that have reported on RAMs in HCV infected individuals. Here we report on the presence of the V36L/Q80K/S122R/D168G/N174S at a low prevalence (1.24%) in treatment naïve individuals (Figure 3.10). This is consistent with previous studies, Bartel *et al.,* 2008 (<1%), Kuntzen *et al.,* 2008 (1.4%) and Tonga *et al.,* 2008 (<1%). This may pose a threat in relation to how our population

will respond to the new DAA regimes. Other polymorphisms were observed in both genotypes 1 and 5.

Although these mutations S91A/R92C/L153I/V248I/P264S and S332A were observed at high prevalence among our study sequences, few of them are known to have direct action against the DAA (Figure 3.10). Mutation S91A changes the electric charge of the protein, thus preventing proper maturation of the NS3 (Pawlotsky *et al.,* 2007). The mutation should be studied further because of its close proximity with the virus catalytic triad residue. Mutation L153I does not alter the function of the protein, however due to its high frequency in the population studied (both genotype 1 and genotype 5a) these might suggest a regional genetic mutation.

Currently, HCV resistance testing at baseline, prior to treatment, is not recommended in the European Association on the study of Liver (EASL) guidelines because there is no standardized testing, testing for RAMs is technically difficult and no consensus on resulting and reporting (EASL guidelines 2018). However, in the absence of pan-genotypic DAAs in South Africa for the treatment of hepatitis C patients, knowledge of genotypes and RAMs remains essential to decide on best DAA regimes and if clinically important resistance mutations are present, then there is time to change to another combination therapy.

The understanding of HCV pathogenesis has paved ways for researchers to ask more questions on what could be done to eradicate the virus. Vaccine development became a definite and feasible strategy to achieve the eradication of HCV. There are several vaccine trials approaches that include peptide, recombinant, vector based and virally vectored vaccines. Firbas *et al.* (2006) produced a peptide vaccine (IC41) based on HLA-A2 allele. Majority of the observed epitopes had high affinity to the HLA A2”01 across the genotypes. Epitopes that bind to the HLA molecule with high affinity produces an effective immune response. The process is highly selective as 1 in 20-400 peptides would bind to the HLA molecule (Prabdial sing *et al.,* 2012). Epitope FLGTSINGV observed in both genotype 1b and 5a had high binding affinity to this allele (Table 3:7). This epitope may prove a good candidate for vaccine development as it protects against more HCV strains. However, genotype 5a epitope sequences, CTCGSADLY and HSRIPGLSP, showed high binding affinity to alleles predominantly found in South Africa, HLA A*01:01 and A*30:01, respectively (Table 3.7).

These epitopes are unique to our study, suggesting that newly mapped epitopes need to be explored further. Among the previously mapped epitopes only CINGVCWTV together with its variant CVNGVCWTV and LGFAGAYMSK were observed in our study; however showing poor binding affinities to the HLA A*01:01, HLA A*02:01, A*30:01 and B*58:01 alleles except for the LGFAGAYMSK. Fytili *et al.,* 2008 reported cross-genotype activity with epitopes CINGVCWTV and CVNGVCWTV, however this was not the case in our study. Variations within CINGVCWTV epitope may reduce the potential efficacy of any vaccine containing the epitope. The published sequences and the study sequences were relatively conserved among the epitopes studied in the NS3 region.

Since genotype 5a is not predominant in the western countries, very few to no available studies have focused on genotype 5a molecularly. This is the first study focusing on South Africa genotype 5a, characterising the mutations within the NS3 protein in relation to prior treatment. Due to the limited data on genotype 5a resistance mutations, there is uncertainty concerning how South African patients will respond to therapy especially the DAAs. Intense work has focused on producing amplicons for the study to achieve the objectives; we needed to have amplicons that were to be sequenced. Different steps were taken to achieve the objectives. Our NS3 mutation data from both genotype 1 and genotype 5 will add to the polymorphism database. This can be of potential importance to the future drug resistance studies in relation to the newly developed antiviral drugs.

Many samples failed to produce a positive band by electrophoresis and were seen as either bright smears or bands stacking in the wells. The thought of too much template used was considered. KOD Hot start DNA polymerase was able to amplify genotype 1 isolates and failed with genotype 5a isolates. HCV amplicons for both genotype 1 and 5a failed sequencing, suggesting that sequencing primers were sub-optimal and/or degraded during the sequencing reaction. Clonal failure might be due to low purity levels of the study samples or even high GC content of the template. The clonal samples were subjected to sequencing even though the restriction of the clonal vector failed. Sequencing of the clonal isolates failed, the reason might be primer degradation during the sequencing reaction or lack of the clonal template.

The positive results from the Deepchek PCR and NGS obviate the need to clone HCV. Although NGS has reduced cost and time, navigating through the mega dataset could be challenging. The high copy number variant might lead to less accurate sequencing by the NGS because of inherent statistical methods used for assembly. Nonetheless, as with HIV drug resistance, as well as successful TB programs using NGS software, data is reverted to an online database, with results sent back soonest for quick interpretation. The Deepchek assay for HCV allows for similar function.

CHAPTER 5: CONCLUSION, RECOMMENDATION AND LIMITATIONS

5.1 CONCLUSION

The study has shown that naturally occurring variants in the NS3 region are present in individuals not treated for HCV by protease inhibitors. By molecular characterization of the NS3, subtypes 1a clustered with USA, Germany and Japan reference sequences and subtype 1b with France, Australia and China. By amino acid analyses, the well documented D168Y mutant was observed in this study. Resistance mutations observed at low frequencies are consistent with previous studies. Although the known resistance mutations found in genotype 1 were rarely observed in this study, screening for mutations prior treatment is still important. Different amino acids polymorphisms in genotype 5a have been identified in this study and their true impact on treatment is not well documented.

Sanger sequencing is still a preferred method for clinical application. However, as indicated from this study, the success rate of PCR and Sanger sequencing was low, indicative that more studies to optimize pan-genotypic primer sets for genotypes less studied, like genotype 5a, need to be done. The sequence analysis of the NS3 protein in this study was limited in positively sequenced amplicons. Therefore, the data cannot be a representation of the circulating mutations within South Africa. The development of a vaccine capable of preventing chronic HCV infections remains the most feasible and realistic method of managing HCV infection globally. We have shown that epitope sequences in the NS3 region were conserved when compared to published sequences. Hence, pan-genotypic vaccines are a possibility. However, most NS3 epitopes are HLA A*02 restricted and consideration must be taken in ethnic groups that do not harbour this allele.

5.2 RECOMMENDATIONS

Some recommendations can be drawn from the study. Firstly, retrospective samples need to be aliquoted and maintained at proper temperatures so as not to degrade

RNA. The DeepChek NS3 resistance mutation assay was more successful than the in-house methods using other enzymes. This is because of the pan- genotypic and high fidelity nature of the enzyme. Furthermore, the amplicons from the Deepchek assay can be used for Sanger sequencing and/or NGS. The DeepChek NS3 resistance mutation assay is recommended over the KOD enzyme. Secondly, to minimise laboratory cost and time batched amplicon-NGS must be considered. The output from NGS can generate a clear analysis of the population heterogeneity. This helps in avoiding misclassification of resistance mutations and promoting adequate DAA choice. However, high variation of the HCV genome, in particular, may complicate the strategies set to overcome DAA resistance mutations. Finally, larger studies investigating the NS3 resistance mutations in genotype 5a should be conducted to provide information that is required when considering a more precise personalised treatment for hepatitis C cure.

5.3 Limitations

Due to financial and time constraints on this study however, a larger sample size could not be employed. Some factors including storage, frequent freezing & thawing and mishandling of the sensitive extracted RNA had a huge impact on the integrity of the RNA. As mentioned before, this study is the first of its kind focusing on genotype 5a RAMs. During the study lots of mistakes were made such as pooling the samples together. This has limited our analysis because DeepChek software does not differentiate mutations per genotype.

REFERENCES

1. Abdelrahma T, Hughes J, Main J, *et al*. 2015. NGS sheds lights on the Natural History of Hepatitis C Infection in Patients' who fail treatment. Journal of Hepatology. 25:88-97.

2. Annalen der Physik.2012. Max von Laue and the discovery of x-ray diffraction in 1912. Wiley online library. 578:76-310.

3. Bacon BR, Gordon SC, Lawitz E, *et al*. 2011. Boceprevir for previously treated chronic HCV genotype 1 infection. New England Journal of Medicine.364:1207-1217.

4. Bae A, Sun SC, Qi X, *et al*. 2010. Susceptibility of treatment-naive hepatitis C virus (HCV) clinical isolates to HCV protease inhibitors. Antimicrobial Agents Chemotherapy. 54:5288 –5297.

5. Bartel DJ, Zhou Y, Zhang EZ, *et al*. 2008. Natural prevalence of hepatitis C virus variants with decreased sensitivity to NS3.4A protease inhibitors in treatment-naive subjects. Journal of Infectious Diseases. 198:800–807.

6. Beerenwinkel N and Zagordi O. 2011. Ultra-deep sequencing for the analysis of viral populations. Current Opinions of Virology. 1:413–418.

7. Berenguer M, Lopez-Labrador FX and Wright TL. 2011. Hepatitis C and liver transplantation. Journal of Hepatology. 35:666-678.

8. Bragg WH and Bragg WL.1912. Diffraction of short electromagnetic waves by a crystal. Proceedings of Cambridge Philosophical Society. 9:43-57.

9. Brass V, Berke JM, Montserret R, *et al.* 2008. Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex. Proceedings of the National Academy Science United State of America. 105:14545-14550.

10. Carvalho R, Alves PAV, de-Souza EF, *et al.* 2014. Protease inhibitor resistance mutations in untreated Brazilian patients infected with HCV: novel insights about targeted genotyping approaches. Journal of Medical Virology. 86:1714-1721.

11. Cento V, Mirabelli C, Salpini D, *et al.* 2012. HCV Genotypes Are Differently prone to the development of Resistance to Linear and Macrocyclic Protease Inhibitors. Public Library of Science ONE.0039652.

12. Chang, K.M. (2003). Immunopathogenesis of hepatitis C virus infection. Clinical Liver Diseases. 7(1), 89-105.

13. Choo QL, Kuo G and Weiner AJ. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science. 244:359-362.

14. Colombo M, Fernandez I, Abdurakhmanov D, *et al.* 2014. Safety and on-treatment efficacy of Telaprevir: the early access programme for patients with advanced hepatitis C. nature reviews Gastroenterology and Hepatology. 63:1150–1158.

15. Cruz-Rivera M, Carpio-Pedroza JC, Escobar-Gutierrez A, *et al*. 2013. Rapid hepatitis C virus divergence among chronically infected individuals. Journal of Clinical Microbiology. 51:629-32.

16. Cucchiarini M, Kammer AR, Grabscheid B, *et al.* 2000. Vigorous peripheral blood cytotoxin T cells response during the acute phase of Hepatitis C virus infection. Cell immunology. 203:111-123.

17. da-Silva P, Almeida AJ, Lampe E, *et al.* 2014. Mutations in hepatitis C virus NS3 protease domain associated with resistance to specific protease inhibitors in antiviral therapy naïve patients. Archives Virology. 155:807-811.

18. De Francesco R. Molecular virology of the hepatitis C virus.1999. Journal of Hepatology. 31:113-118.

19. DeepChek.http://deepchek-portal.ablsa.com/. Accessed 22 February 2018.

20. Deutsch M and Hadziyannis SJ. 2008. Old and emerging therapies in chronic hepatitis C: an update. Journal of Viral Hepatology. 15:2-11.

21. Dore GJ, Conway B, Luo Y, *et al.* 2016. Efficacy and safety of Ombitasvir/Paritaprevir/r and Dasabuvir compared to IFN-containing regimens in genotype 1 HCV patients: the MALACHITE-I/II trials. Journal of Hepatology. 64: 19–28.

22. Du G, Hou L, Guan R, *et al*. 2002. Establishment of a simple assay in vitro for hepatitis C NS3 serine protease based on recombinant substrate and single-chain protease. World Journal of Gastroenterology. 8:1088-1093.

23. Dufour JF, Zuckerman E, Zadeikis N, *et al*. 2017. Safety of glecaprevir/pibrentasvir in adults with chronic genotype 1–6 hepatitis C virus infection: an integrated analysis. Journal of Hepatology. 66:333-542.

24. Duvoux C, Pawlotsky JM, Bastia A, *et al*. 1999. Low HCV replication levels in end-stage hepatitis C virus-related liver disease. Journal of Hepatology. 31:593-597.

25. Everson GT, Copper C, Hezode C, *et al*. 2012. High SVR24 rates with 12 to 24 weeks of ritonavir boosted Danoprevir plus PEG-INF alfa-2a (40kD)/ RBV in HCV genotype 1 or 4 patients in the DAUPHINE study. Journal of Hepatology. 56:552-559.

26. European Association for the Study of the Liver (EALS). 2018. Recommendations of treatment of HCV. http://www.easl.eu/research/our-contributions/clinical-practice-guidelines/detail/easl-recommendations-on-treatment-of-hepatitis-c-2018. Accessed 9 July 2018.

27. European Association for the Study of the Liver (EALS). 2018. Recommendations of treatment of HCV. 2017. HCV Clinical practice guidelines. http://www.easl.eu/research/our-contributions/clinical-practice-guidelines. Accessed 9 July 2018.

28. Falade-Nwulla O, Suarez-Cuervo C, Nelson DR, *et al*. 2017. Oral Direct-Acting Agent Therapy for Hepatitis C Virus Infection: A systematic review. Annals of Internal Medicine.16:2570-2575.

29. Feng H, Shuda M, Changy M *et al*. 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science. 319:1096-1100.

30. Firbas C, Jilma B, Tauber E, *et al.* 2006. Immunogenicity and safety of a novel therapeutic hepatitis C virus (HCV) peptide vaccine: a randomized, placebo controlled trial for dose optimization in 128 healthy subjects. Vaccine. 24:4343-4346.

31. Frick, DN. 2007. The hepatitis C virus NS3 protein: a model RNA helicase and potential drug target. Current Issues Molecular Biology. 1:91-100.

32. Gededzha P, Selabe,SG, Thanda Kyaw T, *et al.* 2012. Introduction of New Subtypes and Variants of Hepatitis C Virus Genotype 4 in South Africa. Department of Virology, Medunsa Campus, University of Limpopo/NHLS Dr George Mukhari Tertiary Laboratory. Pretoria, South Africa. Journal of Virology. 84:601-607.

33. GenBank; http:www.ncbi.nlm.nih.gov/genbank/. Accessed on 6 July 2017.

34. Ghany MG, Strader DB, Thomas DL, *et al.* 2009. Diagnosis, management, and treatment of hepatitis C: an update. Journal of Hepatology. 49:1335–1374.

35. Götte M. 2012. The Distinct Contributions of Fitness and Genetic Barrier to the Development of Antiviral Drug Resistance, Current Opinion in Virology. 2:644–650.

36. Gutierrez AT and. Wyles DL 2014. Importance of HCV genotype 1 subtypes for drug resistance and response to therapy. Journal of Viral Hepatitis. 21:229-40.

37. Hajarizadeh B, Grebely J and Dore G. 2013. Epidemiology and natural history of HCV infection. Nature and reviews. Gastroenterology and Hepatology. 2:553-562.

38. Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis and analysis program for Windows 95/98/NT. Nucleic Acids Symposiums Series; 41:95-98.

39. Hayashi N, Seto C, Kato M, *et al.* 2014. Once-daily Simeprevir (TMC435) with peginterferon/ribavirin for treatment-naïve hepatitis C genotype 1-infected patients in Japan: the DRAGON study. Journal of Gastroenterology. 49:138–147.

40. He Y, Weng L, Li R, *et al.* 2012. The N-terminal helix alpha (0) of hepatitis C virus NS3 protein dictates the subcellular localization and stability of the NS3/NS4A complex. Journal of Virology. 80:11082-11093.

41. Henquell C, Cartau C, Abergel A, *et al.* 2004. High prevalence of hepatitis C virus type 5 in central France evidenced by a prospective study from 1996 to 2002. Journal of Clinical Microbiology. 7: 3030-3035.

42. Hernandez MD and Sherman RE. 2011. HIV/hepatitis C coinfection natural history and disease progression. Current Opinion HIV AIDS. 478 – 482.

43. Hezode C, Fontaine H, Dorival C, *et al.* 2014. Effectiveness of telaprevir or boceprevir in treatment-experienced patients with HCV genotype 1 infection and cirrhosis. Journal of Gastroenterology. 147:132–142.

44. Hoffmann L, Ramos JA, de-Souza EV, *et al.* 2013. Dynamics of resistance mutations to NS3 protease inhibitors in a cohort of Brazilian patients chronically infected with hepatitis C virus (genotype 1) treated with pegylated interferon and ribavirin: a prospective longitudinal study. Journal of Virology. 8:10-57.

45. Horner SM, Park HS, Gale M, *et al*. 2012. Control of innate immune signalling and membrane targeting by the hepatitis C virus NS3/4A protease are governed by the NS3 helix alpha0. Journal of Virology. 86:3112-3120.

46. Howe A, Black S, Curry S, *et al*. 2014. Virological Resistance Analysis from phase2 study of MK-5172 combined with pegylated interferon/ribavirin in treatment naïve patients with hepatitis C virus Genotype 1 infection. Clinical Infection Disease. 59:1612-1657.

47. Hueppe D, Zehnter E, Mauss S, *et al.* 2009. Efficacy and Tolerability of Peg Interferon Alfe-2a (40kD) (PEG) and Ribavirin (RBV) in Genotype 5 and 6 patients with Chronic Hepatitis C under Real life. Control for Hepatology Gastroenterology: Digestive Disease week. Chicago. 9:231-243.

48. Immune Epitope Database. 2018. Version 3.10.0. https://www.iedb.org/. Accessed 11 January 2018.

49. Inoue H, Nojima H, Okayama H, *et al*. 1990. High efficiency transformation of

Escherichia coli with plasmids. Gene, 96: 23-28.

50. Izquierdo L, Helle F, Francois C, *et al.* 2014. Simeprevir for the treatment of hepatitis C virus infection. Pharmacogenomics and Personalized Medicine. Journal of Virology. 7:241-249.

51. Jacobson IM, Lawitz E, Gane EJ *et al.* 2017. Efficacy of 8 weeks of Sofosbuvir, Velpatasvir, and Voxilaprevir in patients with chronic HCV infection: 2 Phase 3 randomized trials. Journal of Gastroenterology. 153:113–122.

52. Jacobson IM, McHutchison JG, Dusheiko G, *et al.* 2011. Telaprevir for previously untreated chronic hepatitis C virus infection. New England Journal of Medicine. 364:2405–2416.

53. Kapoor A, Simmonds P, Cullen JM*, et al.* 2013. Identification of a pegivirus (GB virus-like virus) that infects horses. Journal of Virology. 87:7185-7190.

54. Kieffer TL, Sarrazin C, Miller JS*, et al*. 2007. Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients. Journal of Hepatology. 46:631-639.

55. Kim JL, Morgenstern KA and Lin C. 2014. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. Cell. 87:343-355.

56. Kouranov A, Xie L, de la Cruz J, *et al.* 2006. The RCSB PDB information portal for structural genomics. Nucleic Acids Research. 34:302–305.

57. Kuntzen T, Timm J, Berical A, *et al.* 2008. Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients. Journal of Hepatology. 48:1769–1778.

58. Kuo G, Choo QL and Alter HJ. 1989. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science. 244:362-364.

59. Lawitz E, Sulkowski MS, Ghalib R, *et al*. 2014. Simeprevir plus Sofosbuvir, with or without ribavirin, to treat chronic infection with hepatitis C virus genotype 1 in non-responders to pegylated interferon and ribavirin and treatment naïve patients: the COSMOS randomised study. Lancet. 384:1756-65.

60. Li K, Foy E, Ferreon JC, *et al*. 2005. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. Proceedings of National Academy of Science USA. 102:2992-2997.

61. Lok AS, Gardiner DF, Lawitz E, *et al*. 2012. Preliminary study of two antiviral agents for hepatitis C genotype 1. National England Journal of Medicine. 366:216-224.

62. Maisnier-Patin S and Andersson DI. 2004. 'Adaptation to the Deleterious Effects of Antimicrobial Drug Resistance Mutations by Compensatory Evolution. Research in Microbiology. 155:360–369.

63. McPhee F, Friborg J, Levine S, *et al.* 2012. Resistance analysis of the hepatitis C virus NS3 protease inhibitor Asunaprevir. Antimicrobial Agents Chemotherapy. 56.3670-3681.

64. Meylan E, Curran J and Hofmann K. 2005. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. Nature. 437:1167-1172.

65. Miller RH and Purcell RH. 1990. Hepatitis C virus shares amino acid sequence similarity with Pestiviruses and Flaviviruses as well as members of two plant virus super-groups. Proceedings of the National Academy Science United State of America. 87:2057–2061.

66. Mondelli MU, Cerino A, Boender P, *et al*. 1994. Significance of the immune response to a major, conformational B-cell epitope on the Hepatitis C virus NS3 region defined by a human monoclonal antibody. Journal of Virology. 68:4829-4836.

67. Murphy D, Chamberland J, Dandavino R, *et al.* 2007. A new genotype of hepatitis C virus originating from central Africa (abstract). Journal of Hepatology. 46:619-623.

68. Neumann AU, Lam NP, Dahari H*, et al.* 2008. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. Science. 282:103–107.

69. Nishiya AS, de Almeida-Neto C, Ferreira SC, *et al.* 2010. HCV genotypes, characterization of mutations conferring drug resistance to protease inhibitors and risk factors among blood donors in São Paulo. Public Library of Science. ONE.101371.

70. Paolucci S, Fiorina L, Mariani B, *et al.* 2013. Naturally occurring mutations to inhibitors of HCV NS5A region and NS5B polymerase in DAA treatment-naïve patients. Journal of Virology. 10:330-355.

71. Pawlotsky JM. 2007. Use and interpretation of Hepatitis C Virus diagnostic assays. Clinical Liver Disease. 7:127-137.

72. Paximadis M, Mathebula TY, Gentle NL, et al. 2012. Human leukocyte antigen class I (A, B, C) and II (DRB1) diversity in the black and Caucasian South African population. Human Immunology. 73:80–92.

73. Poh WT, Xia E, Chin-Inmanu K, *et al.* 2013. Viral quasispecies inference from 454 pyrosequencing. BMC Bioinformatics. 14:349-355.

74. Powdrill MH, Tchesnokov EP, Kozak RA, *et al*. 2011. Contribution of a mutational bias in 445 hepatitis C virus replication to the genetic barrier in the development of drug 446 resistance. Proceedings of the National Academy of Sciences of the United State. 51:20509–20513.

75. Prabdial-Sing N, Puren AJ and Bowyer SM. 2012. Sequence-based in silico analysis of well-studied Hepatitis C Virus epitopes and their variants in other genotypes (particularly genotype 5a) against South African human leukocyte antigen backgrounds. BMC Immunology.13:65-67.

76. Prabdial-Sing N, Puren AJ, Mahlangu J, *et al.* 2008.Hepatitis C virus genotypes in two different patient cohorts in Johannesburg, South Africa. Archives of Virology. 11:2049-2058.

77. Prezzi C, Casciaro MA, Francavilla V, *et al.* 2001. Virus-specific CD8(+) T cells with type 1 or type 2 cytokine profile are related to different disease activity in chronic hepatitis C virus infection. European Journal of Immunology. 31:894–906.

78. Qiu P, Sanfiorenzo V, Curry S*, et al.* 2009. Identification of HCV protease inhibitor resistance mutations by selection pressure-based method. Nucleic Acids Research. 37:30-34.

79. Rao VB, Johari N, du Cros P, e al. 2015. Hepatitis C seroprevalence of HIV co-infection in Sub-Saharan Africa: A systematic review and meta-analysis. Lancet Infectious Diseases. 818 – 822.

80. Roentgen WC.1972. Fundemental contributions to the X-ray; the original communications on a new kind of ray. History of medicine division. National library of medicine.

81. Rosen HR, Miner C, Sasaki AW, et al. 2002. Frequencies of HCV-specific effector CD4+ T cells by flow cytometry: correlation with clinical disease stages. Hepatology. 35:190–208.

82. Sanger F and Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology. 94:441–448.

83. Sarrazin C, Lathouwers E, Peeters M, *et al* 2015. Prevalence of the hepatitis C virus NS3 polymorphism Q80K in genotype 1 patients in the European region. Antiviral research. 01:03-09.

84. Serre SBN, Jensena SB, Ghanema L, *et al.* 2016. Hepatitis C Virus Genotype 1 to 6 Protease Inhibitor Escape Variants: In Vitro Selection, Fitness, and Resistance Patterns in the Context of the Infectious Viral Life Cycle. Antimicrobial Agents Chemotherapy. 85:3563-3578.

85. Sharon EF, Houghton M, Coates S*, et al*. 2010. Safety and immunogenicity of HCV E1E2 vaccine adjuvant with MF59 administered to healthy adults. Vaccine. 28:6367–6373.

86. Shepherd J, Abdelrahman T, MacLean AR, *et al*. 2014. Prevalence of HCV NS3 pre-treatment resistance associated amino acid variants within a Scottish cohort. Journal of Virology. 65:50-53.

87. Shoukry NH, Cawthon AG and Walker CM. 2014. Cell-mediated immunity and the outcome of hepatitis C virus infection. Microbiology. 58:391–424.

88. Sidney J, Peters B, Frahm N, *et al*. 2008. HLA class I super-types: a revised and updated classification. BMC Immunology. 9:1-10.

89. Simmonds P. 2004. Genetic diversity and evolution of hepatitis C virus -15 years on. Journal of General Virology. 85:3173–3188.

90. Smith DB, Bukh J, Kuiken C*, et al.* 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Journal of Hepatology. 59:318-27.

91. Stetson DB and Medzhitov R. 2006. Antiviral defence: interferon and beyond. Journal of experimental Medicine. 203:2215-2217.

92. Sulkowski MS, Ceasu E, Asselah T, *et al*. 2011. SILENC1: Sustained virological response and safety of BI201335 combined with pegintereron alfa-2a and ribavirin in treatment-naive patients with chronic genotype 1 HCV infection. Journal of Hepatology. 54:27.37.

93. Tai CL, Chi WK, Chen DS, *et al*. 1996. The helicase activity associated with hepatitis C virus non-structural protein 3 (NS3). Journal of Virology 70:8477-8484.

94. Takayama K, Furusyo N, Ogawa E, *et al.* 2014. A case of successful treatment with telaprevir- based triple therapy for hepatitis C infection after treatment failure with Vaniprevir – based triple therapy. Journal of Infection and Chemotherapy. 8:577-581.

95. Tamura K, Dudley J, Nei M, *et al.* 2015. MEGA6: Molecular Evolutionary Genetics Analysis (MEGA) software version 6.0. Molecular Biology Evolution. 24:1596-1599.

96. Tonga X, Bogen S, Chase R*, et al.* 2008. Characterization of resistance mutations against HCV ketoamide protease inhibitors. Antiviral Research. 77:177-185.

97. Tsubota A, Fujise K, Namiki Y, *et al.* 2011. Peginterferon and ribavirin treatment for hepatitis C virus infection. World Journal of Gastroenterology. 28:419-432.

98. Vallet S, Gouriou S, Nousbaum JB, *et al*. 2005. Genetic heterogeneity of the NS3 protease gene in hepatitis C virus genotype 1 from untreated infected patients. Journal of Medical Virology. 75:528-537.

99. Van der Poel CL. 1994. Hepatitis C virus infection from blood and blood products. Federation of European Microbiological Society: Microbiology Review.14:242-246.

100. Vertuani S, Bazzaro M, Gualandi G, *et al*. 2002. Effect of interferon-alpha therapy on epitope-specific cytotoxic T lymphocyte responses in hepatitis C virus-infected individuals. European Journal of Immunology. 44:144-154.

101. Verbeeck J, Maes P, Lemey P, *et al.* 2006. Investigating the origin and spread of hepatitis C virus genotype 5a. Journal of Virology. 80:4220-4226.

102. Wang Y, Rao HY, Xie, XW, *et al.* 2015. Direct acting Antiviral Agents Resistance associated Polymorphisms in Chinese Treatment naive Patients Infected with Genotype 1b Hepatitis C Virus. Journal of Clinical Medicine England. 128:2625–2631.

103. Watson SJ, Welkers MR, Depledge DP*, et al.* 2013. Viral population analysis and minority-variant detection using short read next-generation sequencing. Philosophical Transaction Royal Society London Biological Sciences. 33:112-116.

104. Wedemeyer H1, He XS, Nascimbeni M, *et al*. 2002.Impaired effector function of hepatitis C virus-specific CD8+ T cells in chronic hepatitis C virus infection. Journal of immunology. 169:3447-3458.

105. Wertheimer AM, Miner C, Lewinsohn DM*, et al*. 2003. Novel CD4+ and CD8+ T-cell determinants within the NS3 protein in subjects with spontaneously resolved HCV infection. Journal of Hepatology. 37:577-589.

106. Westbrook RH and Dusheiko G. 2014. Natural History of Hepatitis C. journal of Hepatology. S56-S68.

107. World Health Organization. 2017. Screening, Care and Treatment of persons diagnosed with chronic hepatitis C virus infection. https://www.who.int/hepatitis/publications/hepatitis-c-guidelines-2017. Accessed 8 May 2017.

108. World Health Organization. 2016. Guidelines for the Screening, Care and Treatment of persons with chronic hepatitis c infection (update). https://www.who.int/hepatitis/publications/hepatitis-c-guidelines-2016. Accessed January 2017.

109. Wose Kinge CN, Espiritu C, Prabdial-Sing N, *et al.* 2014. Hepatitis C virus genotype 5a subgenomic replicons for evaluation of direct-acting antiviral agents. Antimicrobial Agents Chemotherapy. 58:5386-5394.

110. Yan Y, Ying H, Bertrand B, *et al.* 2017. A point mutation in the N-terminal Amphipathic helix α_o in NS3 promotes HCV Assembly by altering the core localisation and facilitating virus budding. Journal of Virology. 91:6-22.

111. Zeuzem S, Sarrazin C, Rouzier R, *et al.* 2005a. Anti-viral activity of SCH 503034, a HCV protease inhibitor administered as monotherapy in HCV-1 patients refractory to pegylated interferon (PEG-INF-α). Journal of Hepatology. 42:233-234.

APPENDICES

Appendix A

NS3 Protease inhibitors	In combination	Targeted genotype	Efficacy	Mode of administration	Year of approval	References
Boceprevir	PEG IFN + Ribavirin	1	66%-75% 59%-75%	Intravenously	2011	Jacobson *et al.,* 2011 Bacon *et al.,* 2011
Telaprevir	PEG IFN + Ribavirin	1	75%-85%	Intravenously	2011	Jacobson et al., 2011 Bacon *et al.,* 2011
Simeprevir	Sofosbuvir	1&4	75%-85%	Orally	2013	Izqucerdo et al.,2014 Hayashi *et al.,* 2014
Paritaprevir	Ombitasvir+ Daclastasvir + Ritonovir	1b&4	94%	Orally	2015	Dore *et al.,* 2016
Grazoprevir	Elbasvir	1&4&6	89%-100%	Orally	2016	Howe *et al.,* 2014
Glecaprevir	Pibrentasvir	Pan genomic	93%-97%	Orally	2017	Dufour *et al.,* 2017
Voxilaprevir	Sofosbuvir + Velpatavir	Pan genomic	96%-98%	Orally	2017	Jacobson *et al.,* 2017
Asunaprevir	Daclastavir	1b	90%	Orally	Phase 3 clinical studies	McPhee *et al.,*2012
Danoprevir	Ritonovir	Pan genomic	88%-89%	Orally	Phase 3 clinical studies	Everson *et al.,* 2012
Vaniprevir	PEG IFN + RV	Pan genomic	N/A	Orally	In testing	Takayama *et al.,*2014

N/A: No Available data

Appendix B

Reference sequences from Genbank for subtype 1a, subtype 1b and genotype 5a.

Subtype 1a

KP409213, KP409211, KP409210, KM587614, KC127118, KC127113, EU862826, EU862827, EF154704, EF154703, EF154700, EF154702, EF154701, KT735186, KT735187, KT735181, KT735179, KU871305, KU871304, KU871302, KU871300, KU871306, KJ564300, KJ564299, KJ564298, KJ564297.

Subtype 1b

AF165054, AF165053, AF165064, AF165063, AF165062, AB049092, AB049094, AB154178, AB154177, AB154179, AF054248, AF054249, AF05447, AF05450, AF05452, DQ071885, GU451224, GU451223, GU451221, GU451220, GU451218.

Genotype 5a

AF064490, NC_009826, KC844046, FN553266, FN553263, FN553261, FN553258, KJ925146, KJ925147, KJ925149, KJ925148, HQ623421, HQ623422, HQ623414, FN553277, KC767829, KC767833, KJ925148, KC767831, KC767830, KX107851, Y13184, KX107848, KX,107849, KC767832, KX107850, KC767834, KF589888, FN553113.

Appendix C

R14/49 Mrs Meyagabo Emily Youbi et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M170765

NAME: (Principal Investigator)	Mrs Meyagabo Emily Youbi et al
DEPARTMENT:	Pathology National Institute for Communicable Diseases
PROJECT TITLE:	Analysis of the Non-Structure 3 Protease Gene Across Hepatitis C Viral Genotypes in Treatment Naive Individuals
DATE CONSIDERED:	Adhoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-study under M120614
SUPERVISOR:	Dr Nishi Prabdial-Sing
APPROVED BY:	Professor P. Cleaton-Jones Chairperson, HREC (Medical)
DATE OF APPROVAL:	14/08/2017

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 10004,10th floor, Senate House/3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand. I/We fully understand the conditions under which I am/we are authorised to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. in this case, the study was initially review July and will therefore be due in the month of July each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES