MOLECULAR CHARACTERIZATION OF UNIQUE HEPATITIS C VIRAL GENOTYPES IN SOUTH AFRICA

Nishi Prabdial-Sing

A thesis submitted to the Faculty of Health Sciences,
University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy

Johannesburg, 2011

CANDIDATE'S DECLARATION

I, Nishi Prabdial-Sing, declare that this thesis is my own work. It is being submitted for the degree of Doctor of Philosophy in the Faculty of Health Sciences at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.

……
Signature of candidate

Signed at ……..…… on this …….. day of……..……., 2011

ABSTRACT

Hepatitis C is a notifiable disease in South Africa (SA), but there is a lack of surveillance and reporting systems. The availability of routine, reliable and validated methods of hepatitis C virus (HCV) detection and genotyping, together with better surveillance strategies is vital to ensure optimal patient diagnosis, treatment and disease management. This is the first extensive molecular study on the HCV in this country to identify circulating genotypes, host immuno-genetics and patients responses to combination therapy.

HCV genotypes in the patient and blood donor groups were determined by sequencing of the 5'untranslated (5'UTR) and non-structural -5B (NS5B) regions. Three molecular-based tests, line probe assay (LiPA), real-time and palindromic nucleotide substitutions (PNS), were compared to the sequencing method. Molecular sequence analyses of the core, envelope 1(E1), NS4B and NS5B were determined. The inferred amino acid data was used to determine epitope variation across immunodominant regions. Viral load monitoring was performed on patients receiving combination treatment.

A positive HCV polymerase chain reaction (PCR) result is necessary to confirm active infection. A National HCV surveillance database was established to collate patient demographics with results from public health laboratories. Genotype 5a predominates in patients with liver disease. Changing frequencies and introductions of other subtypes were determined. The LiPA (5'UTR) was found to surpass the other genotyping tests, as it was quick and easy. Geographical clustering within the geno-

type 5a clade was evident in the E1 and NS4B regions. The branching order of genotype 5a suggests that genotype 5a is older than genotype 3, inferring that patients with genotype 5a may respond better to therapy than those infected with genotype 3. The divergence estimate of genotype 5a was found to be between 100-156 years. Despite the homogeneity of genotypes 1 and 5 epitope sequences, well-published epitopes were predicted to bind sub-optimally to the common human leukocyte antigen (HLA) alleles, making a vaccine less effective in SA. The treatment response for genotype 5a was higher in this study compared to previous global studies.

This new molecular knowledge on HCV genotypes circulating in SA will allow informed decisions when planning preventative and treatment strategies relevant to local viral and host genetics.

DEDICATION

In loving memory of my dad,
Mr. Premchand Prabdial,
1934-2005

“Be the change you want to see in the world”

Mahatma Gandhi

PUBLICATIONS AND PAPERS PRESENTED ARISING FROM THE THESIS

PEER-REVIEWED PUBLICATIONS AND ABSTRACTS (INTERNATIONAL)

Prabdial-Sing N, Giangaspero M, Puren AJ, Mahlangu J, Barrow P, Bowyer. SM. (2011) Palindromic-nucleotide substitutions (PNS) of hepatitis C virus genotypes 1 and 5a from South Africa. J Virol Methods. 175(2):272-7.

Prabdial-Sing, N., Mathebula, T., Paximadis, M., Puren, A.J. and Bowyer, S.M. (2009) Epitope Analysis of genotype 5a Hepatitis C Virus against South African HLA backgrounds. . Joint Conference on Bioinformatics of Infectious Diseases., Bamako, Mali, 34 (manuscript under review)

Prabdial-Sing, N., Puren, A.J., Mahlangu, J., Barrow, P. and Bowyer, S.M. (2008) Hepatitis C virus genotypes in two different patient cohorts in Johannesburg, South Africa. Archives of Virology 153(11), 2049-58.

Prabdial-Sing, N., Bowyer, S.M. and Puren, A.J. (2006) Genotyping hepatitis C virus in South Africa: A comparison of results obtained by partial sequencing of the 5'UTR and NS5B regions with a real-time Lightcycler genotyping method (abstract). Journal of Clinical Virology 36(2), S104-105.

OTHER PUBLICATIONS AND ABSTRACTS (NATIONAL)

Prabdial-Sing, N., Muvhulawa, S., Howard, W., Halford, B., Song, E., Mahomed, A., van der Merwe, S., Bassa, F., Ally, R., van Zyl, J., Naidoo, V.G., Luke, B., Adu, A., Puren, A. and Bowyer, S.M. (2009) HCV genotypes and response to therapy in a South African study group. South African Journal of Epidemiology and Infection 24, 21.

Prabdial-Sing, N., Puren, A.J. and Bowyer, S.M. (2007) A commercial Hepatitis C virus real-time LightCycler genotyping method does not detect local genotypes reliably. Molecular and Cell Biology Group Symposium, 17.

ACKNOWLEDGEMENTS

I acknowledge with the greatest of appreciation God Almighty, my supportive husband and my two beautiful daughters for providing me with the strength to continue.

I also thank my dear helpers, my family and friends for their support and understanding.

The list of all those that made this possible is below. I thank you.

Poliomyelitis Research Fund
Dr. Sheila Bowyer (supervisor)
Professor Adrian Puren (co-supervisor and HOD)
Dr. Maria Paximadis
Tiyani Mathebula
Professor J. Mahlangu
Dr. Peter Barrow
Professor M. Giangaspero
Specialized Molecular Diagnostic Staff
All national hospitals, staff and patients
Beryl Halford
Prea Pillay
Professor Cleaton-Jones
SANBS
ROCHE Diagnostics
ROCHE Pharmaceuticals
The Hepatitis Working Group
ASCB/ISCB Travel fellowship

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS/ACRONYMS

^{0}C	degrees celsius
A	Asian
AA	African American
aa	amino acid
Ab	antibody
ABI	Applied BioSystem
Afn	African
Ag	antigen
AIC	Akaike information criterion
AIDS	Acquired Immunodeficiency Syndrome
ALT	Alanine amino-transferase
AMP	amplification
AMV	avian myeloblastosis virus
ANOVA	ANalysis Of VAriance
ANN	artificial neural network
APC	antigen presenting cells
ARFP	alternate reading frameshift protein
Asn	antisense
AST	aspartate aminotransferase
AZT	Zidovudine (Retrovir®)
B	Black
BD	blood donor
BE	Belgium
bp	base pair
Ca	Caucasian
C	core
CC	Coloured
CDC	Centers of Disease Control and Prevention, Atlanta, USA
cDNA	complementary DNA
CDW	Corporate Data warehouse
CH	Chinese
CHB	Chris Hani Baragwanath Hospital
CLD	chronic liver disease
COOH	carboxyl
CRF	chronic renal failure
CTL	cytotoxic T-lymphocyte
DC	dendritic cell
DdI	Didanosine
dATP	deoxy-adenine-triphosphate
dCTP	deoxy-cytosine-triphosphate
dGTP	deoxy-guanosine-triphosphate
dTTP	deoxy-thymidine-triphosphate
dUTP	deoxy-uracil-triphosphate
dNTP	Deoxyribonucleotide triphosphate
dsRNA	double-stranded RNA
DTT	dithiothreitol
dUTP	deoxyuracil triphosphate
E1	envelope-1
E2	envelope-2
EC	Eastern Cape
EDTA	ethylene diamine tetra acetic acid
EHRN	Eurasian Harm Reduction Network
EIA	Enzyme immunoassay
ELISA	enzyme-linked immunoassay
ENZ	enzyme
EQA	External Quality assurance
ER	endoplasmic reticulum

ESRD	end-stage renal disease
ETR	End-of-treatment response
EV	epitope vaccine
EVR	Early virological response
F	fluorescence
fg	fentogram
Fig	figure
FR	France
FRET	Fluorescence Resonance Energy Transfer
FS	Free State
F81	Felsentein 1981 model
GT	Genotype
GA	Gauteng
GTR	general time reversible model
H	haemophiliac
HAART	Highly active antiretroviral therapy
Haem	haemophiliac
HALT-C	Hepatitis C anti-viral long-term treatment against cirrhosis
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HCWs	Health Care Workers
HIV	Human immunodeficiency virus
HKY	Hasegawa-Kishino-Yano model
HLA	human leukocyte antigen
HS	heparin sulfate
HTA	heteroduplex tracking analysis
HVR	hypervariable region
I	Indian
ICTV	International Committee on Taxonomy of Viruses
ICTVdb	International Committee on Taxonomy of Viruses database
IEDB	Immune Epitope database project
IFN	Interferon
igVR	intergenotypic variable region
IL	interleukin
INSDC	International Nucleotide Sequence Database Collaboration
IR	immune response
IRES	internal ribosome entry site
ISDR	interferon-sensitivity determining region
ISGF	interferon stimulated gene factor
IU	international unit
IVDU	intravenous drug user
J	Japanese
JC	Jukes-Cantor model
JHB	Johannesburg
K80	Kimura 1980 model
kB	kilobase
kDa	kilodalton
kg	kilogram
KIM-2	Kimura-2 parameter
KIR	killer inhibitory receptor
KZN	Kwa-Zulu Natal
L	litre
LC	LightCycler
LD	liver disease
LiPA	line probe assay
LnL	log likelihood
LP	Limpopo Province
MBP	mannose binding protein
MCL	maximum composite likelihood

$MgCl_2$	magnesium chloride
MHC	major histocompatibility complex
min	minute
ML	maximum likelihood
ml	millilitre
mM	millimolar
MMWR	Morbidity and Mortality Weekly Report
MP	Mpumalanga
MRCA	most recent common ancestor
N	number
NAb	neutralizing antibody
NAAT	nucleic acid amplification testing
NC	Northern Cape
NDoH	National Department of Health
ng	nanogram
NH2	amino
NHLS	National Health Laboratory Services
NICD	National Institute for Communicable Diseases
NIDDK	National Institute of Diabetes, Digestive and kidney disease
NIH	National Institutes of Health
NJ	neighbour-joining
NK	natural killer
NKT	natural killer T-cell
nm	nanometer
nM	nano-Molar
NS	non-structural
nt	nucleotide
NPV	negative prediction value
NW	North West
OCR	optical cut-off ratio
OD	optical density
OD/OCR (or s/co)	Optical density to optimal cut-off ratios (or signal to cut-off ratio)
ORF	open reading frame
p	protein
PCR	Polymerase Chain reaction
PEG	Polyethylene glycol
PEG-IFN	pegylated interferon
PEP	post exposure prophylaxis
pEVR	partial early virological response
pg	picogram
PHAC	Public Health Agency of Canada
PKR	protein kinase
pM	pico-Molar
PNA	peptic nucleic acid
PNS	palindromic nucleotide substitution
pol	polymerase
ProPred	Protein Prediction
PPV	positive prediction value
QCMD	Quality Control on Molecular Diagnostics
RBD	receptor binding domain
RdRp	RNA dependent RNA polymerase
RFLP	restriction fragment length polymorphism
RIBA™	Recombinant immunoblot assay
RIG	retinoic acid inducible gene
RNA	Ribonucleic acid
RSA	Republic of South Africa
RT	reverse-transcriptase
RVR	Rapid virological response
s	second
S/CO	signal cut-off

SA	South Africa
SANAS	South African National Accreditation Services
SANBS	South African National Blood Service
SBT	sequencing-based technique
SD	standard deviation
seq	sequencing
SIGN	specific intracellular adhesion molecule-3 grabbing non-intergrin
siRNA	silencing RNA or Small interfering RNA
SMDU	Specialized Molecular Diagnostics Unit
Sn	sense
SNP	single nucleotide polymorphism
SOP	standard operating procedure
SR-B1	scavenger receptor class B type 1
ssRNA	single-stranded RNA
Std	standard
SVR	Sustained virological response
TAT	turn around time
TB	tuberculosis
T_{H1}	T-helper 1
T_{H2}	T-helper 2
TLR	toll-like receptor
Tm	melting temperature
TMD	transmembrane domain
TM	Transversion model
TNF	tumor necrosis factor
TrN	Tamura-Nei model
U	unit
µg	microgram
UK	United Kingdom
µl	microlitre
µm	micrometer
µM	micro-Molar
UPGMA	Unweighted pair group method with arithmetic mean
US (or USA)	United States (of America)
UTR	untranslated region
W	Whites
WA	White American
WC	Western Cape
WITS	Witwatersrand
WHO	World Health Organisation
Zn	zinc

Nucleotide base	*Standard Abbreviation*
adenine	A
cytosine	C
guanosine	G
thymidine	T
uracil	U

Amino Acid	*Standard Abbreviation*
alanine	**A**
cysteine	**C**
aspartic acid	**D**
glutamic acid	**E**
phenylalanine	**F**
glycine	**G**
histidine	**H**
isoleucine	**I**
lysine	**K**
leucine	**L**
methionine	**M**
asparagi**n**e	**N**
proline	**P**
glutamine	**Q**
a**r**ginine	**R**
serine	**S**
threonine	**T**
valine	**V**
tryptophan	**W**
t**y**rosine	**Y**

CHAPTER ONE

1.1. INTRODUCTION

The hepatitis C virus (HCV) is a major cause of cirrhosis and chronic liver disease and a reason for liver transplantation in most developed countries (World Health Organization, WHO, 2007). It is estimated that about 3% (170 million) of the world's population are infected with HCV (WHO, 2010). About 80% of HCV infected people become chronic carriers. The seroprevalence of HCV in Sub-Saharan Africa is comparatively low with a prevalence of 1.7%-3.2% reported in South Africa (SA) (Lavanchy, 2011; Ellis *et al.*, 1990) compared to a high prevalence of 14% in Egypt (WHO, 2002). The contribution of HCV to liver disease in South Africa is not known as studies have been small and limited and were performed 8 to 15 years ago without current technology. An ongoing, up-to-date, large-scale, molecular epidemiological study of circulating HCV strains is thus long overdue. Understanding the heterogeneity of HCV can improve diagnostic technologies, vaccine development and therapeutic interventions.

1.2. HEPATITIS C VIRUS (HCV)

HCV is a member of the *Flaviviridae* family, of the genus, *Hepacivirus* (Robertson et al., 1998) together with the putative species tamarin GB-virus and closely related human GBV-C (International Committee on Taxonomy of viruses, ICTVdB, 2006). Humans are the only known natural host, although HCV has been transmitted experimentally to chimpanzees (Bukh et al, 1998). Hepaciviruses are different from flaviviruses and pestiviruses in that not all viral types can propagate efficiently in cell culture, and the precursor protein is cleaved at the NS2/3 junction via a Zn-

dependent proteinase (ICTVdB, 2006). HCV is the most labile of the blood-borne viruses (Elder et al., 2004), is highly mutable (10^{-3} to10^{-5} mutations per nucleotide per generation (Bartenschlager and Lohmann, 2000, Table 1) with a high replication rate of 10^{12} virions/day (Neumann et al., 1998, Table 1) compared to 10^{8} virions/day for HIV (Weber, 2001). The mutations accumulate within a single host giving rise to a population of virus genomes, known as the viral quasispecies, which differ from the virus first transmitted (Davis, 1999). This high diversity of HCV poses a major challenge to vaccine development.

Table 1. Properties of the HCV genome as summarized from the references provided.

HCV Properties	Reference
Half-life is 2 to 5 hours	Neumann et al., 1998
Viral turnover is 10^{10} to 10^{12} virions/day/ individual	Neumann et al., 1998, Herrmann et al., 2000
Low viral viraemia of 10^{6} to 10^{7}IU/ml	Yao et al., 2005
Error rate of 10^{-3} to 10^{-5} mutations/nucleotide site/replication	Bartenschlager and Lohmann, 2000
Mutation rate (substitution/site/replication):	
in chimps (1.4 X 10^{-3})	Okamoto and Mishiro, 1994, Major et al., 1999
in humans (1.9 X 10^{-3})	Okamoto and Mishiro, 1994
in vitro (3.5-4.8 X 10^{-3})	Kato et al., 2009
NS5B has poor fidelity, is error-prone and lacks repair mechanism	Steinhauer et al., 1992
Evolutionary rate (substitution/site/year):	
Core (1.8 X 10^{-4})	Pybus et al., 2009
NS5B (3.3 – 4.1 X 10^{-4})	Okamoto and Mishiro, 1994, Pybus et al., 2009
E1 (7.1-7.2 X 10^{-4})	Okamoto and Mishiro, 1994, Verbeeck et al., 2006
NS4B (5.4 X 10^{-4})	Verbeeck et al., 2006

1.2.1. HCV Structure and Life Cycle

Virions are spherical and 40-60 nm in diameter with a lipid envelope (Fig.1). Like other flaviviruses, HCV acquires its envelope by budding through the endoplasmic reticulum (ER) membrane (Blanchard et al., 2002). The virion has spike-like surface projections made up of the envelope proteins (E1 and E2, Fig.1) and an isometric

nucleocapsid of 25-30 nm (Kaito et al., 1994). Electron microscopic observations have shown that the HCV core protein self-assembles at the ER membrane, rather than in the cytoplasm, forming a structured capsid (Blanchard et al., 2002) The nucleocapsid houses a single, linear, positive-sense RNA strand of 9.6 kB (Choo et al., 1991).

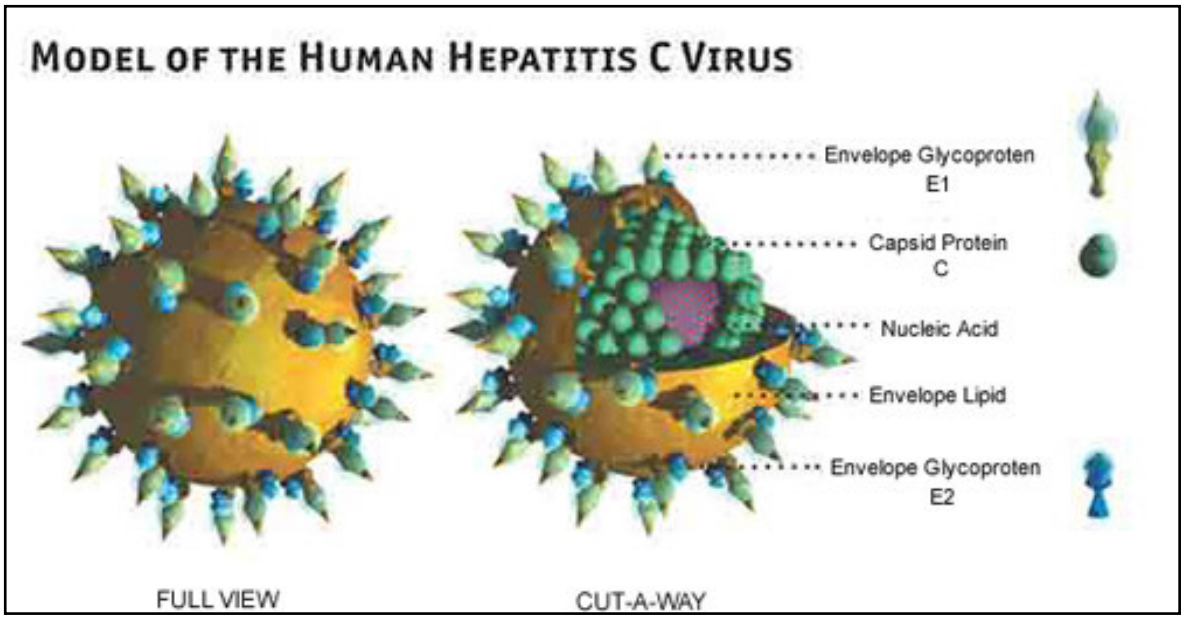


Fig.1. Diagrammatic 3D model depicting the structure of the hepatitis C virus (adapted from a figure by Louis E. Henderson published online, March 2001, www.prn.org)

Progress in understanding the HCV life cycle has accelerated since the JFH1 (genotype 2a) strain was successfully cultured (Wakita et al., 2005 and Zhong et al., 2005). Studies on the viral life cycle (summarized below) have contributed to developments in anti-viral therapies and targets for neutralizing antibodies (reviewed in Zeisel et al., 2007).

(1) HCV attaches to the host cell membrane by the viral envelope glycoproteins, E1 and E2, binding to factors such as heparin sulfate [HS] (Barth et al., 2003). Lipoproteins may facilitate the entry of HCV into the liver cell as both use a common receptor, scavenger receptor class B type 1 [SR-B1] (Scarselli et al., 2002).

Adsorption to the cell membrane and entry into the cell occur in a multi-step process, via the host receptor proteins CD81 (Pileri et al., 1998), claudin-1 (Evans et al., 2007), occludin (Liu et al., 2009) and lectins [dendritic cell (DC)-specific intracellular adhesion molecule-3 grabbing non-intergrin (SIGN) and liver-specific (L)-SIGN] (reviewed in Burlone and Budkowska, 2009].

(2) The virus is internalized in a clathrin-coated pit, by endocytosis, in a pH-dependent environment (Blanchard et al., 2006).

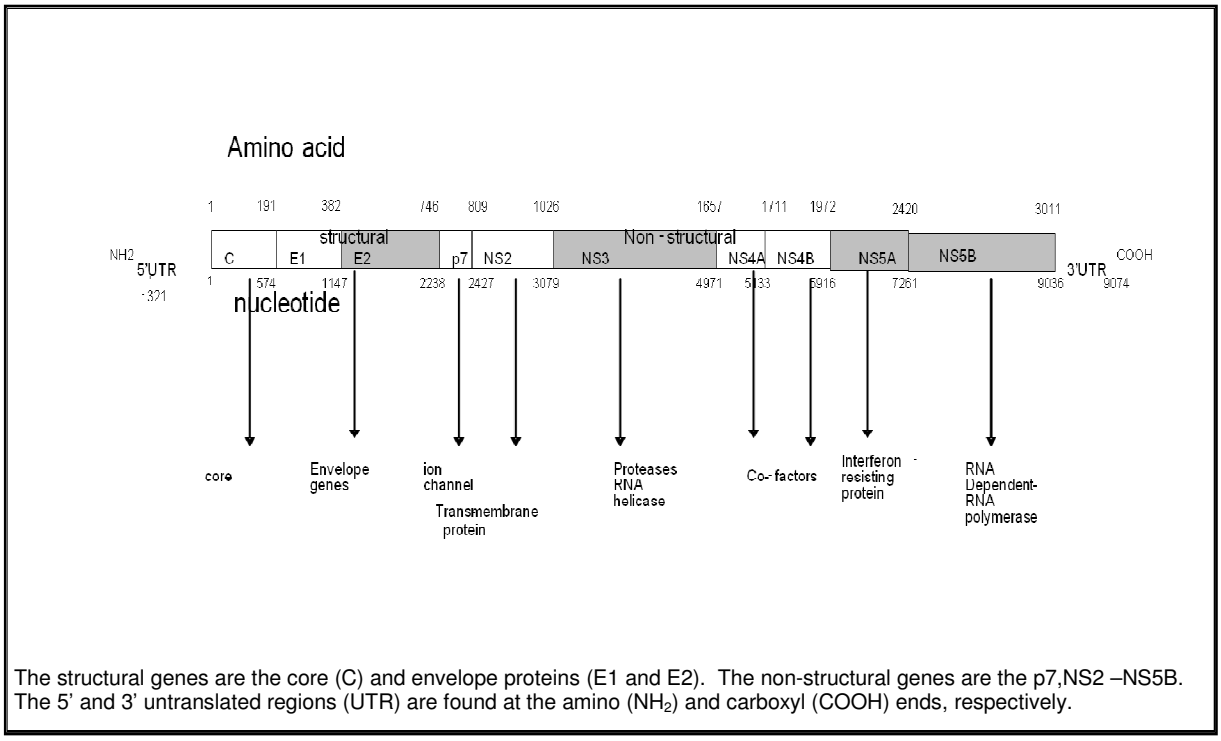
(3) The virus is transported in the cell in an endosome and prior to genome processing, the viral envelope fuses with the endosomal membrane in the acidic cytosol and the capsid is removed, exposing the ssRNA to the replication process (reviewed in Zeisel et al., 2007).

(4) The ssRNA serves as a template for replication into many copies of negative and positive stranded RNA and direct translation into polyprotein. Assembly and packaging of new virions occur on the rough ER (reviewed in Langhans et al, 2010). However, the exact mechanism for the assembly and maturation of virions is still unknown.

The viral binding, post-attachment and fusion processes are potential targets to neutralize viral infection (reviewed in Zeisel et al., 2007). Antibody binding to the free virus will block viral attachment to the cell. Neutralizing antibodies (NAb) may prevent post-binding stages by binding and blocking host entry molecules at the cell surface or preventing internalization via the endosome. NAb may also prevent the fusion process by changing the conformation of the protein involved or hinder contact between the membranes (Zeisel et al., 2007). NAb can interfere with viral uncoating and replication too (Zeisel et al., 2007).

1.2.2. Genes and their function/s during the life cycle

A single large open reading frame (ORF) encodes a polyprotein of about 3000 amino acids (Choo et al., 1991, Fig.2). The polyprotein is cleaved into ten viral proteins by host cellular and viral proteases. The three structural proteins are the core (C) and the two envelope proteins, E1 and E2. There are seven non-structural proteins, the p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B (Lohmann et al., 1996; Suzuki et al., 1999). The structural proteins which are cleaved by host cellular proteins (signal peptidase) and aggregate to form the viral particle (Suzuki et al., 1999) are located at the 5' terminus end (Fig.2). The non-structural proteins are cleaved by viral proteases (serine protease, Moradpour et al., 2002) and are located further downstream (Fig.2). Linear and conformational epitopes are found throughout the viral polyprotein and are thought to be involved in both innate and adaptive immune responses of the host to HCV infection (chapter 6 for a more detailed discussion).



The structural genes are the core (C) and envelope proteins (E1 and E2). The non-structural genes are the p7,NS2 –NS5B. The 5' and 3' untranslated regions (UTR) are found at the amino (NH_2) and carboxyl (COOH) ends, respectively.

Fig.2. A schematic representation of the HCV genome, indicating numbering of amino acid (aa) and nucleotide (nt) bases according to Choo et al., 1991. The shaded areas are the larger genes (>1000bp).

1.2.2.1. 5' and 3' untranslated region (5'UTR and 3'UTR)

The 5' untranslated region (UTR) of HCV RNA is approximately 340 nucleotides long (Fig.2) and is highly conserved among different HCV strains (Suzuki et al., 1999). Studies on the secondary structure of the 5'UTR region of HCV have revealed the presence of four domains, three of which are shown in Fig.3, with strong stem-loop structures and a pseudoknot (Fig.3, Wang et al., 2005). The proximal part of the 5'UTR of domains II, III and IV, with some of the core region serve as an internal ribosome entry site [IRES] (Honda et al., 1996). The function of the IRES is to initiate translation of the polyprotein in a cap-independent manner (reviewed in Suzuki et al., 1999), forming a stable complex with the 40S ribosomal subunit (Tsukiyama-Kohara et al., 1992). IRES-mediated polyprotein translation is commonly used by many viruses, eg. enteroviruses (Borman et al., 1995). Several studies have indicated that the IRES interacts specifically with cellular factors of different cell types, either enhancing or reducing protein translation (Forton et al., 2004; Lerat et al., 2000) and that mutational changes in certain subdomains, eg. IIId, can affect structure and function of the IRES (Barria et al., 2009).

The highly conserved 5'UTR and the stable IRES has thus been the target of (a) antisense-based ribozymes, siRNAs, (b) oligonucleotides and peptide nucleic acids [PNA] (Alt et al., 1999; Zhang et al., 1999; Alotte et al., 2008), (c) many commercially available diagnostic PCR and genotyping assays (discussed in chapter 3) and (d) its "palindromic" sequence structure has been used to genotype HCV (Giangaspero et al., 2008, discussed in chapter 5).

The 3'UTR has three regions namely, (1) a variable region, with sequences that differ among different genotypes for example, the CACTCC motif is different for genotypes 2a and 2b compared to genotypes 1a, 1b, 1c and 3a (Tokita et al., 1998) , (2) a long poly (U/UC) tract and (3) a highly conserved X-region with three stem loop structures (Tanaka et al., 1996). The poly (U/UC) tract and the X-region are important for viral replication (Friebe and Bartenschlager, 2002). The secondary structure may be important for proper termination of translation (van Doorn, 1994) as the structure at the 5'UTR IRES is for initiation.

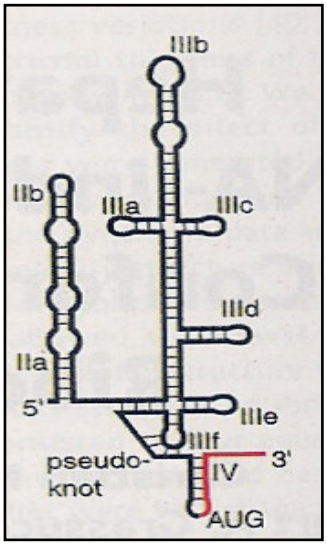


Fig.3. The IRES structure in the 5'untranslated region, with stem-loop domains IIa, IIb, IIIa-f, IV and pseudo-knot, red in figure (adapted from Wang et al.,2005)

1.2.2.2. Structural proteins

Core region

The first cleavage product of the polyprotein at the amino end is the core protein of molecular weight 17-23 kDa. The core protein is a highly basic, non-glycosylated protein and forms the major part of the nucleocapsid (reviewed in Bartenschlager and Lohmann, 2000). The core protein has two domains (i) domain I, involved in

binding to RNA and nuclear localization and (2) domain II, associates the core protein to the ER or lipid droplets (Suzuki et al., 2005) for particle assembly. Mutations in domain II reduce core stability and result in loss of production of infectious virus in JFH1 cell cultures (Boulant et al., 2007). Multimeric complexes between the core and E1 (Baumert et al., 1998) and p7 and NS2 (Murray et al., 2007) may play a role in the morphogenesis of HCV.

The amino acid sequence of the core protein is highly conserved among different HCV strains in comparison with other structural and non-structural proteins (Suzuki et al., 1999) and encompasses several immunodominant epitopes. The core epitopes bind to multiple human leukocyte antigen (HLA) class II alleles and generate proliferative CD4+ T-cell responses, clearing the virus in healthy individuals with low viral load (Botarelli et al., 1993). Whereas the core protein can stimulate an immune response and clear infection, it was also associated with Fas-mediated apoptosis (Honda et al., 2000), contributing to severe hepatitis and hepatocellular carcinoma. The core protein may bind to apolipoproteins involved in lipid metabolism, form fat droplets in hepatocytes, and contribute to steatosis (Moriya et al., 1998).

Linear core epitopes are used in enzyme linked immuno-absorbent assays (ELISA) to detect antibodies in patient sera (van Doorn et al, 1994). Probes in the core and the 5'UTR region have been incorporated into a new line probe assay (LiPA) to specifically subtype genotypes 1 and 6 (Noppornpanth et al., 2006b) and this is discussed in further detail in chapter 4.

During natural HCV infection, an alternate reading frame (ARF) that overlaps the core protein is expressed (Branch et al., 2005). This protein, also known as the frameshift (F) protein is seemingly due to the unusual events of core translation, such as ribosomal frameshifting and may be involved in HCV persistence (reviewed in Chevaliez and Pawlotsky., 2007a).

Envelope regions

Envelope proteins (E1 and E2) are highly glycosylated type 1 transmembrane proteins with a large N-terminal hydrophilic domain on the outer part of ER membrane and a small C-terminal hydrophobic transmembrane domain [TMD] (reviewed in Op De Beeck et al., 2001). The TMD anchors and retains the envelope proteins in the ER membrane (Cocquerel et al., 2001). E1 and E2, have molecular weights of 33-35kDa and 70-72 kDa, respectively and are responsible for interacting with the host cellular receptor tetraspanin CD81. E1 is thought to fuse the virus with the cellular membrane (Ciczora et al., 2007). The E2 receptor binding domain (RBD) has three variable regions, namely, the hypervariable region 1 (HVR1), HVR2 and an intergenotypic variable region [igVR] (reviewed in Gottwein and Bukh, 2008). The HVR can differ up to 80% between genotypes and subtypes (Weiner et al., 1991). The HVR1 (27 aa) is a basic region with a positive charge and interacts with the negatively charged cell surface, playing a role in cell recognition, attachment and cell compartmentalization (Barth et al., 2003).

The HVR1 is the strongest target for NAb responses (Farci et al., 1995) and has been considered for peptide-based vaccines. However, it is also the most variable region (Kato et al., 1992), and is responsible for the evolution of quasispecies which

alternatively contribute to progression of chronic disease (Farci et al., 2000) or result in an early virological response to antiviral therapy (Fan et al., 2006). Despite producing a strong immune response, infection is not cleared (Ray et al., 1999) as escape mutants, not recognised by the circulating NAb (Sklan et al., 2009), persist.

The HVR2 appears to adjust the E2 receptor binding process (Roccasecca et al., 2003), but the function of the igVR, which is conserved within a genotype, is unclear (Gottwein and Bukh, 2008). Further downstream from these variable domains in the E2, is a sequence of 12 amino acids that inhibits the activity of protein kinase R (PKR), an antiviral protein (Taylor et al., 1999).

1.2.2.3. Non-structural proteins

The non-structural genes, p7 and NS2-NS5 are required for viral morphogenesis (Gottwein and Bukh, 2008) and replication (reviewed in Bartenschlager and Lohmann, 2000), respectively. The small, 63 aa protein, p7, functions as a transmembrane protein with two domains (TM1 and TM2), connected by a short cytoplasmic loop (Sakai et al., 2003). The p7 protein acts as a putative ion channel (Griffin et al., 2003) and the ER-associated p7 is thought to be important in particle production (Brohm et al., 2009).

The NS2 protein is a 23 kDa transmembrane protein, responsible for ER association (Santolini et al.,1995). The NS2 protein and the amino-terminal domain of the NS3 form the NS2/NS3 protease, which is responsible for self-cleavage at a single site (Grakoui et al., 1993). The NS2 protein has a short half-life as it is degraded after self-cleavage from NS3. The NS3 protein (70 kDa) is multi-functional, containing

the N-terminal serine protease and the C-terminal helicase/NTPase domain (reviewed in Chevaliez and Pawlotsky, 2007a). Serine protease activity is vital for the viral life cycle as it releases the rest of the polyprotein at the NS3/NS4A, NS4A/ NS4B, NS4B/NS5A, NS5A/NS5B junctions. This protease function makes the NS3 a target for antiviral therapies and potential vaccines. The helicase/NTPase activity is necessary for unwinding the secondary structures of RNA before replication, RNA binding and separation of nascent strands from template strands during replication (Tai et al., 1996). NS3/NS4A contain highly immunogenic epitopes (discussed in chapter 6) and the C100 antigen from this region is used in most ELISAs (reviewed in van Doorn, 1994).

The NS4A protein is a cofactor of the NS3 protease activity and stabilizes the protein. It appears to anchor the NS3/NS4A complex and localizes it at the ER membrane (Bartenschlager et al., 1995). The NS4B protein [26 kDa] (Bartenschlager and Lohmann, 2000) is an integral membrane protein found in the ER and is essential for the assembly of the HCV replication complex (Egger et al., 2002). The NS5A is a 56-58 kDa serine phosphor-protein with three domains (Tellinghuisen et al., 2004, 2005). Domain I is characteristic of hepaci- and pestiviruses (Tellinghuisen et al., 2004) and serves as an RNA-binding pocket during viral replication (Tellinghuisen et al., 2005). The NS5A interacts directly with the NS5B and is thought to be involved with the polymerase activity of the NS5B (Chevaliez and Pawlotsky, 2007a). A domain of 40 aa in the NS5A, called the interferon-sensitivity-determining region [ISDR] (Enomoto et al., 1995), binds to the PKR and inhibits its antiviral function (Gale et al., 1998). The NS5A protein can also induce interleukin (IL)-8 expression and blocks the assembly of interferon stimulated gene factor -3

(ISGF-3) complex, hence reducing antiviral activity by interferon (IFN) (Polyak et al., 2001).

The NS5B (65 kDa) protein encodes the viral replicase, RNA-dependent-RNA polymerase (RdRp) and is anchored to the ER (Moradpour et al., 2004). The RdRp has the classical "finger" structure of polymerase, with the "finger" and "thumb" domains channeling the ssRNA toward the deoxyribonucleotide triphosphates [dNTPs] (Ago et al., 1999). The NS5B RdRp synthesizes the minus-stranded RNA that then acts as a template for the plus-stranded RNA by the polymerase. As this polymerase plays an important role in viral replication, it is a target for antiviral drugs (Chevaliez and Pawlotsky, 2007b). The RdRp is highly error-prone as it lacks any proofreading mechanism (Steinhauer et al., 1992), resulting in HCV heterogeneity (Table 1). Seemingly, HCV mutations do not, in most cases, affect survival and fitness of the virus and allow for immune escape (Gottwein and Bukh, 2008). The NS5B protein is semi-conserved among different strains of HCV and is used in phylogenetic analyses of sequences to differentiate between genotypes and more so between subtypes (see 1.2.3 and chapter 3).

1.2.3. HCV Genotypes

HCV is classified into six genotypes [1-6] (Simmonds et al., 1993a) and >80 subtypes by phylogenetic analyses of the 5'UTR and NS5B regions (Simmonds et al., 1993a, b). A provisional classification for a genotype 7a was given to a single full-genome sequence of a Canadian emigrant (Murphy et al., 2007). Different HCV genotypes display a nucleotide divergence of 30% or greater (Fig.4) while subtypes (described as a, b, c, etc.) are defined by a 10-30% divergence (Simmonds et al., 1994b, Fig.4). Strains within the same subtype, or within a host quasispecies, differ

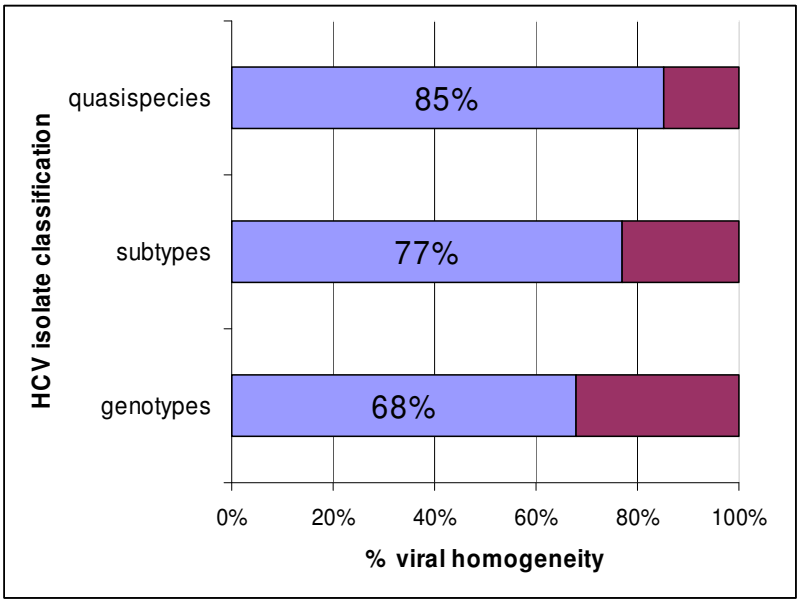


Fig.4. A comparison of the viral homogeneity among HCV isolates of different genotypes (68%), same genotype but different subtypes (77%) and within a subtype (quasispecies, 85%).

by <10-15% (Davis, 1999 Fig.4) with comparisons clustered around a consensus reference sequence (Bukh et al., 1995). This extreme heterogeneity of HCV within and between hosts enables the virus to escape host immunity and hinders the development of vaccines.

HCV genotyping is of clinical, epidemiological and surveillance importance. From a public health perspective, knowledge of the genotype frequencies and changes in these provide important epidemiological information which informs screening and preventative public programs. HCV genotype 1b is associated with severe liver disease leading to cirrhosis and hepatocellular carcinoma (HCC; Prieto et al., 1999) and genotype 3 appears to be associated with pronounced development of liver steatosis (Abid et al., 2005). Different genotypes also vary in their responsiveness to interferon/ribavirin combination therapy (see section 1.10). Moreover, the impact on the sensitivity and specificity of commercial virologic assays is influenced by the

molecular change in the virus.

HCV genotypes are distributed depending on geography and route of infection (Bourliere et al., 2002) but travel and migration are changing these patterns. Fig.5 is a representation of the prevalence of common genotypes and subtypes across the world (Kuiken et al., 2005). Genotypes 1a and 1b are the major genotypes in Europe, North and South America and Australia. Genotype 1b predominates in Japan, China and Russia (Bourliere et al., 2002) and this subtype has the highest global prevalence (45%, Fig.5). Genotype 2c is common in Western and Southern Europe (WHO, 1999). Genotype 4a is widespread in Egypt (Kamel et al., 1992) while 4c is found in Central Africa (WHO, 1999). Genotypes 3a and 6a are most common in South East Asia (Bourliere et al., 2002).

A unique genotype, genotype 5 (Ohno et al., 1994), with only one subtype, subtype a (Chamberlain et al., 1997), was first reported from South Africa. Although originally thought to be unique to South Africa, genotype 5 has more recently also been found in other countries including Central France (14.2%; Henquell et al., 2004) and in the West Flanders Province of Belgium (27.7%; Verbeeck et al., 2006).

Naturally occurring inter-and intragenotypic recombinants (Fig.5) of HCV have been documented since 2002; intergenotypic recombinants, 2k/1b in Russia (Kalinina et al., 2002) and Uzbekistan (Kurbanov et al., 2008), 2i/6p in Vietnam (Noppornpanth et al., 2006a), 2/5a in France (Legrand-Abravanel et al., 2007), 2b/1b in Phillipines (Kageyama et al., 2006) and intragenotypic recombinants, 1a/1b in Peru (Colina et al., 2004) and Uruguay (Moreno et al., 2006), 1a/1c in Japan (Cristina and Colina,

2006) and Ireland (Moreau et al., 2006). In order for a recombination event to occur, different genotypes (intergenotypic) or subtypes (intragenotypic) must infect the same cell of the patient. Multiple exposures are the norm in intravenous drug users (Gottwein and Bukh, 2008), although mixed infections in a single person are thought to be uncommon (Viazov et al., 2000). However, many of the genotyping tests analyse the conserved 5'UTR and mixed infections are not always detected (discussed in chapter 4).

Superinfections have also been described in individuals with chronic HCV infections, that get infected with another genotype (Kao et al., 1993) or subtype (Yokota et al., 1994), either via blood transfusions or intravenous drug use (reviewed in Blackard and Sherman, 2007). The clinical impact of mixed infections, superinfections and recombinants is not fully understood but this viral heterogeneity will further challenge diagnostic techniques, vaccine design and therapeutic management.

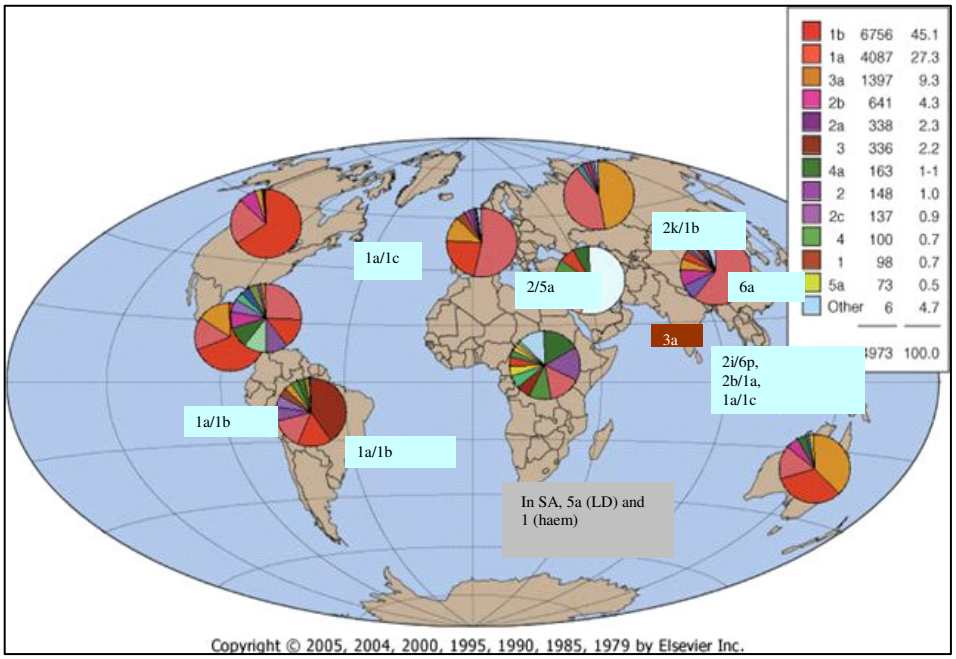


Fig.5. The worldwide geographic distribution of HCV genotypes and subtypes with locations on the current recombinants and genotypes 3a, 5a and 6a (adapted from Kuiken et al., 2005)

1.3. GLOBAL EPIDEMIOLOGY

HCV infection is not seasonal in transmission. The WHO estimates that there are 170 million people infected with HCV globally and more than 350 000 are at risk of developing liver cirrhosis and/or liver cancer (WHO, 2009). Approximately 4 million people in the USA (~1.9% of the population) are HCV seropositive (Armstrong et al., 2006), compared to the 42 000 infected with human immunodeficiency virus (HIV) (Centers for Disease Control and Prevention, CDC, HIV statistics 2008, http://www.cdc.gov/hiv). Globally, of the 33 million people with HIV/AIDS, 4-5 million are co-infected with hepatitis C (Alter, 2006). In the USA, the seroprevalence of HCV in patients with HIV is highest in intravenous drug users (60-90%) and haemophiliacs (50-70%) compared with homosexual men (4-8%; Sterling et al., 2003).

Seroprevalence rates are reported worldwide as it is the most cost-effective way to identify HCV in large populations. The lowest seroprevalence of HCV is reported from northern Europe where figures range between 0.5 and 2% in countries including Great Britain, Germany and France (Eurasian Harm Reduction network, EHRN, 2007 http://www.harm-reduction.org/hepatitis-c.html). Higher seroprevalence rates have been reported in Southeast Asia (Malaysia (2.3%) and Philippines (2.3%) reviewed in EHRN, 2007). South Africa has a seroprevalence of 1.7%-3.2% (Lavanchy, 2011; Ellis et al., 1990) whereas other African countries, like Egypt, reach as high as 14.5% (WHO, 2002). Apart from the universal spread of HCV via blood products prior to 1994, Egypt's extremely high prevalence is a result of HCV transmission through widespread anti-schistosomal injection treatment in the 1970s (WHO, 2002).

Seroprevalence studies both locally and globally have targeted specific regions and/

or high-risk cohorts, eg. in Europe, HCV seroprevalence rates are monitored in the intravenous drug-users cohorts (EHRN, 2007). Since this high risk group is a major risk for HCV transmission, there are many national programs which test, counsel and provide clean syringes and needles to drug users. Hence, hepatitis B and C surveillance networks are well established in this group of individuals.

A large number of individuals (21% in Canada, Public Health Agency of Canada, PHAC, 2010, http://www.phac-aspc.gc.ca) that are HCV infected remain undiagnosed. HCV is highly infectious and this pool of chronic carriers continues to promote the epidemic (EHRN, 2007, also discussed in chapter 2).

1.4. EPIDEMIOLOGY OF HCV IN SOUTH AFRICA

ELISAs to detect anti-HCV became available in the early 1990s. However, care must be taken when comparing the results of these first generation assays with second and third generation ELISAs because of the differences in their sensitivity and specificity (Thakur et al., 2003). While the South African seroprevalence results of Ellis et al., 1990 and Lavanchy, 2011 compare with global figures (see above), a Kwa-Zulu Natal (KZN) hospital study (Soni et al., 1996) reported high seropositivity in three high risk sentinel cohorts; patients with cirrhosis, HCC and chronic active hepatitis (Table 2).

Table 2. An overview on the studies on HCV seroprevalence in SA. Since 1990.

	Seroprevalence (%)	Location in SA	Reference
Population-based			
Blood donors	1.4 N=66000	Western Cape	Tucker et al., 1997
Hospital-based			
Cirrhosis	23 N=77	KZN	Soni et al., 1996
Hepatocellular carcinoma	24 N=33	KZN	Soni et al., 1996
	20.86 N=532	Southern Africa	Kew et al., 1990
Chronic active hepatitis	33 N=6	KZN	Soni et al., 1996
AIDS	1 N=100	CHB (GA)	Lodenyo et al., 2000
HIV positive	13.4 N=1937	KZN	Parboosing et al., 2008
Health care workers (HCWs)	1.8 N=402	CHB (GA)	Vardas et al., 2002
Haemodialysis	21 N=103	Groote Schuur (Cape)	Cassidy et al., 1995

KZN, Kwa-Zulu Natal; CHB, Chris Hani Baragwanath; GA, Gauteng Province. Population-based indicates volunteer blood donors and the hospital-based groups include patients in health care facilities.

1.4% of the blood donor group (Tucker et al., 1997, Table 2), 1% of AIDS patients (Lodenyo et al., 2000,Table 2) and 1.8% of HCWs (Vardas et al., 2002, Table 2) were found to be seropositive compared to a recent study on HIV positive patients (13.4%) in a Durban hospital (Parboosing et al., 2008,Table 2).

Tucker et al. (1997) and others (Thakur et al., 2003) found that the positive predictive value (PPV) for viraemia of a positive ELISA result is affected by the population seroprevalence, viral load and the method used to confirm viraemia.

The survey of blood donors from the Western Cape area of South Africa found detectable viraemia by nucleic acid amplification testing (NAAT, one example of which is Polymerase Chain reaction, PCR, analysis) in only 0.056% of the 1.41% found to be seropositive (Tucker et al., 1997). This lack of specificity of ELISA in low risk groups is particularly evident when optical density (or signal) to optimal cut-off ratio (OD/OCR or s/co) is low. For these reasons, the CDC in Atlanta recommends routine confirmatory NAAT testing (Morbidity and Mortality weekly reports, MMWR, 1998 http://www.cdc.gov/mmwr/pdf/rr/rr4719.pdf). PCR was implemented at the NICD in 2003 (Chapter 2). The current testing algorithm for HCV diagnostics in SA is shown below (Fig. 6).

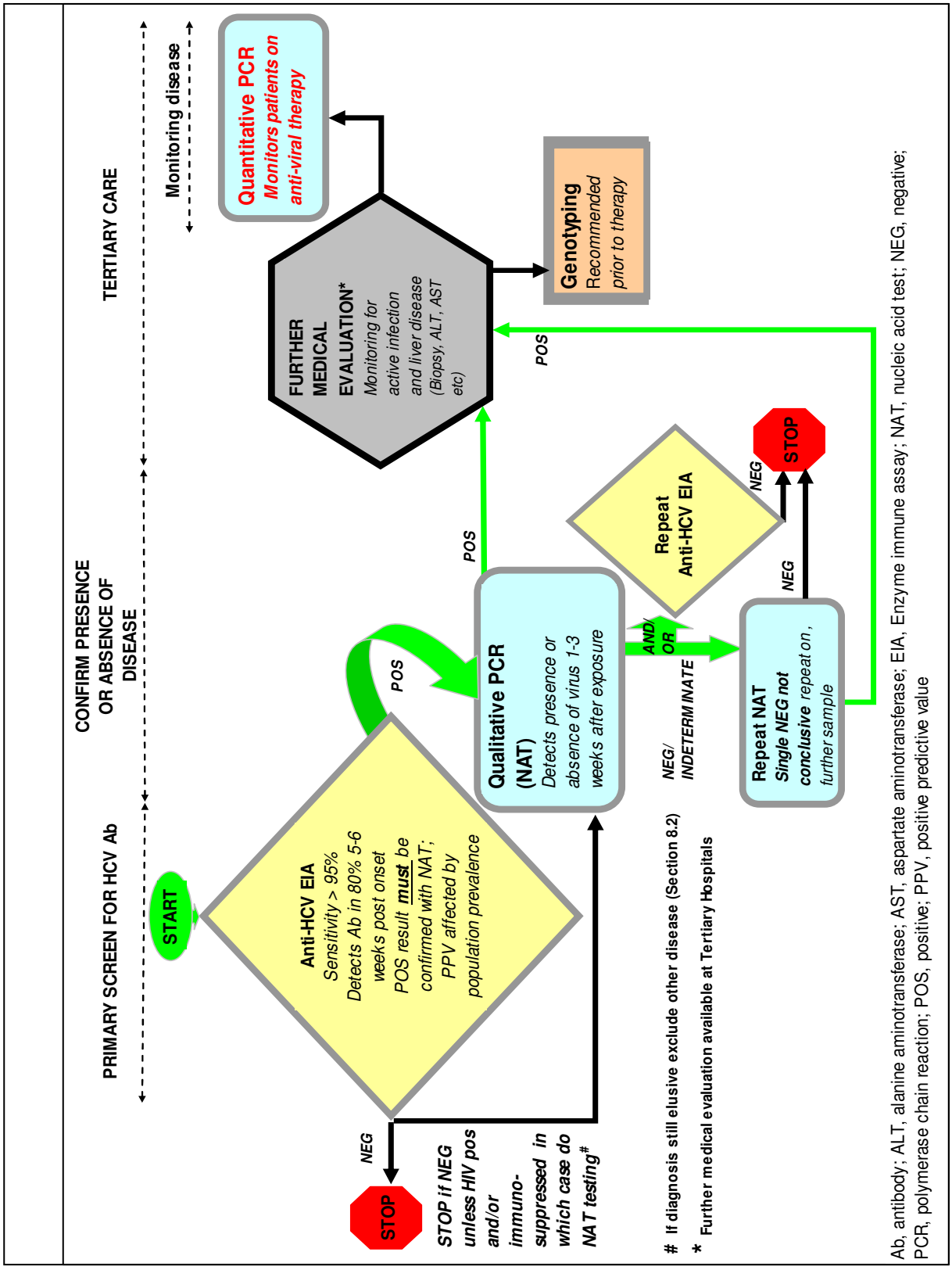


Fig.6. The HCV testing algorithm as illustrated in the current National guidelines for the prevention and control of Hepatitis C virus in SA, final draft, 2010. The primary screen test for HCV is the HCV antibody test. A positive antibody test is confirmed by a qualitative PCR to indicate presence or absence of active infection. If the PCR test is positive, further medical evaluation is recommended at tertiary hospitals where specialized diagnostic and molecular tests can be done. NAT=NAAT in text

1.5. TRANSMISSION AND RISK FACTORS

The existence of HCV (previously known as non-A, non-B, NANB hepatitis) was postulated in 1975 (Feinstone et al., 1975). HCV is primarily parenterally transmitted through infected blood and blood products. It was the most common cause of post-transfusion hepatitis worldwide prior to donor screening for antibodies to HCV in 1992 (reviewed in Richter, 2002). Thirty-five percent of injecting drug users are infected with hepatitis C (MMWR, 2009, http://www.cdc.gov/mmwr/pdf/ss/ss5803.pdf). 70-80% of HCV transmission in Canada (www.phac-aspc.gc.ca) is via injection drug use and the sharing of preparatory materials (syringe/needle, spoon, cooker, filters, straws (McMahon et al., 2004) and pipes. The only other common transmission route is via accidental exposure to infected blood (e.g. needlestick injury, reviewed in Richter, 2002). In a German hospital, HCV transmission in HCWs, occupationally acquiring HCV, is 3 and 10%, depending on the HCV viral load in the source (Wicker et al., 2008).

Injection use is the major route of HCV transmission. However, body piercing, tattooing and traditional markings are also risk factors. Fishermen in Japan were given tattoos for easy identification in case they died at sea, and this may explain the high incidence of HCV-related HCC seen today at most harbour towns (Yoshizawa, 2002). Non-parenteral transmission of HCV is not so well defined. Sexual transmission is relevant in persons with high risk behaviours and the immuno-suppressed. Mother-to-baby transmission occurs in 5% of infants born to HCV infected women, also dependent on viral load (reviewed in Richter, 2002). A fairly large number of HCV cases (41%) have no identifiable risk factor in South East Asia (Nguyen et al., 2010).

1.6. IMMUNE RESPONSES DURING HCV INFECTION

What determines whether an HCV infected individual resolves infection or becomes chronically infected, and in those who are infected, whether they will respond to therapy? The answer partially lies in viral factors such as viral tropism, viral variation or an immune deficient syndrome caused by a virus, for eg. HIV, but host genetics also plays a role (Freeman et al., 2001). Genetic variables which affect the host immune response are important and depend on the interplay between the virus and the host.

Figure 7 summarizes the two arms of the immune system, the host's innate and adaptive immune responses (IR), respectively. The key components of the adaptive IR to HCV are further described in figure 8.

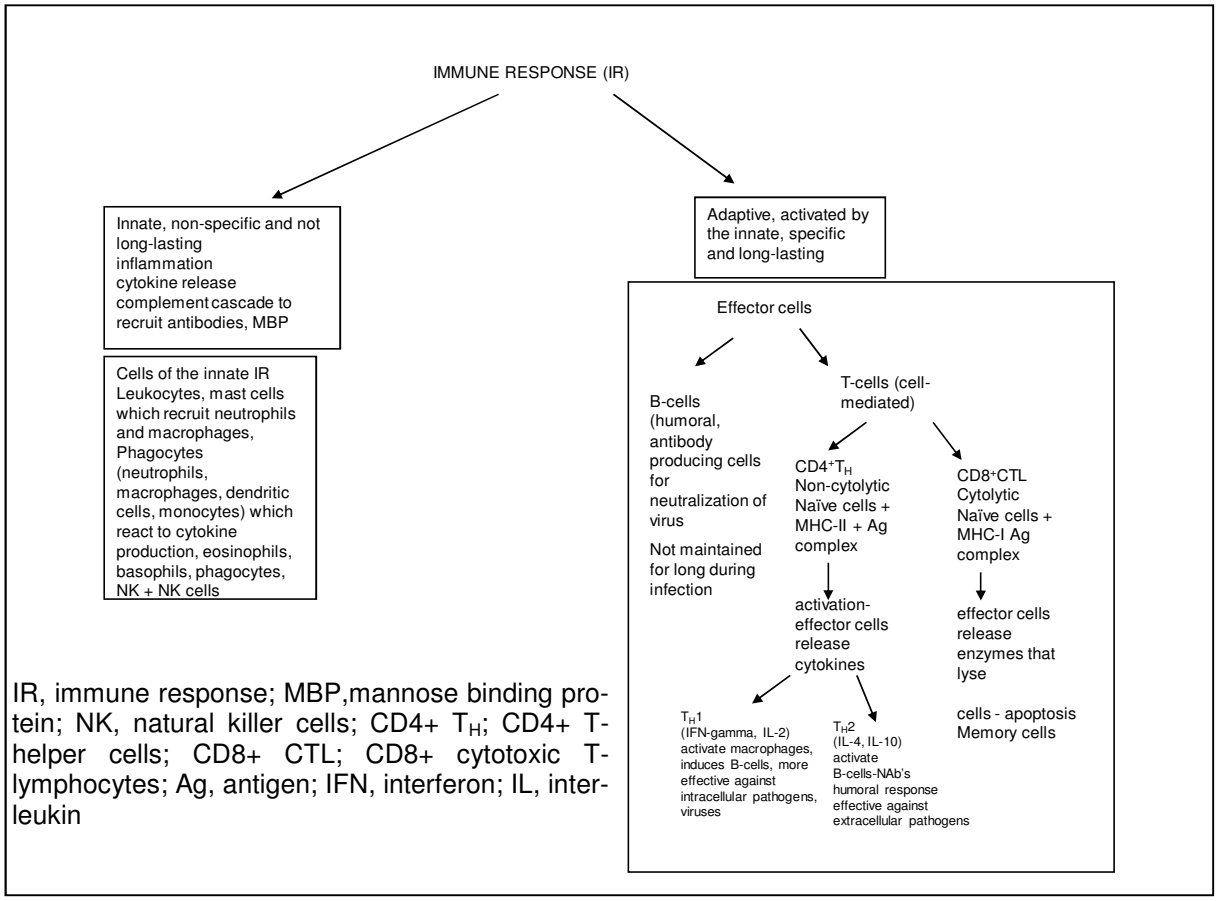


Figure 7: A summary of the breakdown of the innate and adaptive immune system, indicating the functions of the different cells

1.6.1. THE INNATE IMMUNE RESPONSE

The innate IR occurs early, is non-specific and activates the adaptive response. It is the first step toward viral recognition and attempts to clear infection by the host cells (reviewed in Szabo and Dolganiuc, 2008, Fig.7). Both the innate IR and the specific humoral response facilitate the production of antibodies that, among other things, interfere with viral entry and replication within the cell.

Cytokines (IL-12, TNF-alpha) and IFN production suppress viral replication and activate natural killer cells (NK) to destroy infected cells via apoptosis (reviewed in Branch and Seeff, 2000).

1.6.2. THE ROLE OF HLA IN HCV DISEASE

Numerous studies link host immuno-genetic factors to HCV infection/clearance or response to therapy. Different haplotypes of NK cell receptors are associated with protection or persistence, respectively (Khakoo et al.,2004). Individuals with KIR2DL3 and HLA-C1 are more likely to recover from HCV than those with KIR2DL1 and HLA-C2 (Kanto and Hayashi, 2006). Also, recently single nucleotide polymorphisms (SNPs) on the IL-28B gene (Ge et al., 2009) were found to correlate with early response to pegylated interferon (PEG-IFN) and ribavirin therapy. Studies on the molecular typing of HLA alleles and their association with HCV have varied in terms of sample size and methodology, however, the consensus is that the HLA genotype does influence the outcome of HCV infection (acute/self limiting versus chronic/persistent) and/or the response to therapy (sustained response versus no response) (Table 3 and 4, Kuniholm et al., 2010). Interestingly, there is some correlation with HLA alleles associated with HCV and that with HIV disease, for ex-

ample, HLA-B*27, HLA-B*57 and DRB1*13:03 are associated with lower HIV viral load (Altfeld et al., 2003; Kiepiela et al., 2004; Julg et al., 2010). The same two HLA -B alleles have been linked with HCV clearance (Tables 3 and 4) and HLA class II DRB1*13:01 was found to be associated with HCV clearance in Caucasians in Germany (Hohler et al., 1997; Table 4). Hence, going forward in efforts to develop HCV vaccines for the diverse populations worldwide, lessons must be learnt from the HIV vaccine initiatives (Kim et al., 2010). Whereas great strides have been made with regard to developing epitope-based vaccines for HIV, longer and broader immune responses is the focus as longer epitope sequences and multiple epitopes may generate a better range of immune responses (Kim et al., 2010). Compared to epitope vaccines (EV), subunit protein vaccines elicit mainly CD4 and CD8 T-helper responses and hardly any CTL and antibody responses and DNA vaccines take longer to elicit an immune response (Cavicchi and Kowalski, 2009; http://www.wipo.int/patentscope/en/programs/patent_landscapes/documents itti_patent_ls_hiv_protein_vaccines.pdf). An understanding of HLA interactions with immunodominant epitopes is essential for the development of an effective EV. These host and viral specific characteristics, respectively, affect the CD4+ and CD8+ T-cell responses which, in turn affects disease outcome.

Different HCV genotypes have different natural and/or escape mutations within the T-cell epitopes. Viral peptides bind (within the antigen presenting cell) with the HLA class-I and - II molecules and this complex becomes exposed at the surface of the cell. This complex is recognized by the T-cell receptor. In case of viral variance, there is no binding or no recognition. So, the protective effects associated with the HLA allele and a peptide of one HCV genotype may not be present when chal-

lenged by a variant of another HCV genotype (Bengsch et al., 2009). There is a need to study epitope changes within and across HCV genotypes as well as consider how binding of an epitope to heterogenous HLA molecules can influence immune responses (further discussed in chapter 7).

Table 3: The associations of HLA Class I alleles and outcome of HCV disease, where available, ethnic groups, HCV genotypes and epitope regions have been included.

HLA TYPE I	ETHNIC GROUP*	HCV GENOTYPE	HCV DISEASE	REFERENCE
A2	Ca, J, CH, Afn	1a	chronic	Wertheimer et al., 2003
A2		1	chronic	Cerny et al., 1995; Himoudi et al., 2002
A2		1	Chronic, respond to therapy	Vertuani et al., 2002
A2			Acute/clearance	Thimme et al., 2002
A01	Ca	1b, 3a	chronic	Neumann Haefelin et al., 2008
A1101	AA + Ca		clearance	Thio et al., 2002
A23		1	chronic	Koziel et al., 1995
A2301	AA		Viral persistence	Thio et al., 2002
A30	Ca		acute/clearance	Minton et al., 1998
B8	Ca (Irish)		Viral persistence	
B14	Ca (Italian); Egypt		Susceptibility/ persistence	
B27	Ca (German)		Viral clearance	
B35	Ca (Russian); Korean		Viral persistence	
B53	AA + Ca	1	chronic	Koziel et al., 1995
B57	AA + Ca; WA; Ca (Irish)		Viral clearance	Thio et al., 2002; Chuang et al., 2007; McKiernan et al., 2004; Kuniholm et al., 2010
B14/Cw8	Ca		chronic/severe	
B41/KIR3D51	J		Protection from HCC	
Cw3	Ca		Mild/acute/protection	
Cw7+N_k	Ca (Italian)		HCC	
Cw8/C2/KIR	Ca (Spanish)		Viral persistence	
Cw01/ C1C1+KIRDL3	Ca		clearance	
Cw01	Ca (US)		protection	
Cw04	AA + Ca		Viral persistence	Thio et al., 2002

* Ca, Caucasian; J,Japanese; CH, Chinese; Afn, African; AA, African American; WA, West Africans

Table 4. The associations of HLA Class II alleles and outcome of HCV disease, where available, ethnic groups, HCV genotypes and epitope regions have been included.

HLA TYPE II			ETHNIC GROUP*	HCV GENO-TYPE	HCV DISEASE	REFERENCE
DRB	DQB1					
1*11	*03		Ca	1a	Spontaneous recovery	Day et al., 2002
1*04, 1*10	*01, *03		Ca	1a	Spontaneous recovery	Day et al., 2002
1*04, 1*08	*03		Ca	1a	Spontaneous recovery	Day et al., 2002
1*1104, 3*0202	*0301			Serotype 1	Spontaneous recovery	Schulze zur Weisch, 2005
1*0401, 1*1001 4*01	*0301 *0501			Serotype 1	Spontaneous recovery	Schulze zur Weisch, 2005
1*0801, 1*0404 4*0101	*0301 *0302			Serotype 1	Spontaneous recovery	Schulze zur Weisch, 2005
1*1101	*0301				Spontaneous recovery	Harcourt et al., 2004
1*0402, 1*0701				3	Sustained response	Sarobe et al., 2006
1*0101, 1*0701				3	Sustained response	Sarobe et al., 2006
1*1302, 1*1601				3	Non-responder	Sarobe et al., 2006
DRB1	DPB1	DQB1				
*17 *1101	*0201 *1401	*0301 *0603		ND	Self-limited hepatitis	Lamonaca et al., 1999
*0311 *1101	*02012 *3401	*0201 *0301		1a	Chronic hepatitis	
*0407 *1104	*0401	*0201 *0301		3a	Self-limited hepatitis	
*1109 *0305	*0201 *0301	*0201 *0301		1a	Self-limited hepatitis	
*0801	*01011 *0401	*0501 *0402		1a	Self-limited hepatitis	
*0801 *1501	*0401	*04 *0602		1b	chronic	Lamonaca et al., 1999
DRB1	DQA1	DQB1				
*1501	*01	*0301 *0602	Ca Ca		Clearance Chronic	Wawrzynowicz et al., 2000 Thursz et al., 1999
*0701	*0201	*02	Thai, Ca (German)		Chronic/persistent	Hohler et al., 1997 Vejbaesya et al., 2000
*0301			Ca		Clearance, response to IFN/ ribavirin	Harcourt et al., 2001; Minton et al., 1998
*1101		*0301	AA		Clearance, response to IFN/ ribavirin	Donaldson, 2004
*1301			Ca (German)		Viral clearance	Hohler et al., 1997
DR5 DRB1*11	DQ3	DQB1*0301	Ca	1a, 1b, 2a, 2b, 3a	Clearance	Tillmann et al., 2001
DR13			J		mild	Aikawa et al., 1996
DRB1*0405			J Ch		Severe/chronic No response to therapy	Aikawa et al., 1996
DRB4*0101			Ca		Chronic/persistent	Thursz et al., 1999
DQB1*03			J, Ca (French + Europe), AA		Viral clearance	
DQB1*04			J, Ca (North Europe + Po-land)		Chronic/persistent	

*denotes, Ca= Caucasian, J= Japanese, AA= African American, Ch=Chinese

1.6.3. THE ADAPTIVE IMMUNE RESPONSE

Peptides associate with HLA and this complex of peptide-HLA is presented to CD4+ or CD8+ T-cells. When there is no recognition, there is no signal. The CD4+ T_H cells act as immunoregulators because they respond to non-self peptides by secreting lymphokines that direct and support the proliferation of memory CD8+ CTL immune response, (Fig.8 T_H 1 cytokines: IL-2, IFN-gamma, TNF-alpha) or B-cells neutralizing antibodies (Fig.8 T_H 2 cytokines: IL-4, IL-5, IL-10, IL-13) (Mosmann and Sad, 1996). The virus enters the hepatocytes using the claudin-1 protein and the co-receptor CD81, internalising viral antigen which is presented to the MHC-class I molecule (HLA-A, B or C molecules).

Like other viral infections, the CTL response can clear HCV by either direct or indirect cytolysis. The former involves molecular proteins (FasL) and enzymes (caspase-8) (reviewed in Kanto and Hayashi, 2006) together with the release of perforin and granzyme B while the latter uses non-cytolytic mechanisms involving cytokines and their co-receptors (Fig.8). IFN-gamma suppresses HCV replication and decreases HCV viral load (Thimme et al., 2002). CCR5 and RANTES were found to be upregulated in HCV chronic infection (Promrat et al., 2003), as reported for HIV-1 infection.

In HCV infection, CD4+ T-cell responses are short-lived (Day et al., 2002), while CD8+ T-cell responses are elevated in the acute phase but decrease as viral persistence develops (Lechner et al., 2000) and the humoral B-cell response (particularly, NAb) remains ineffective in controlling viraemia (reviewed in Zeisel et al., 2009).

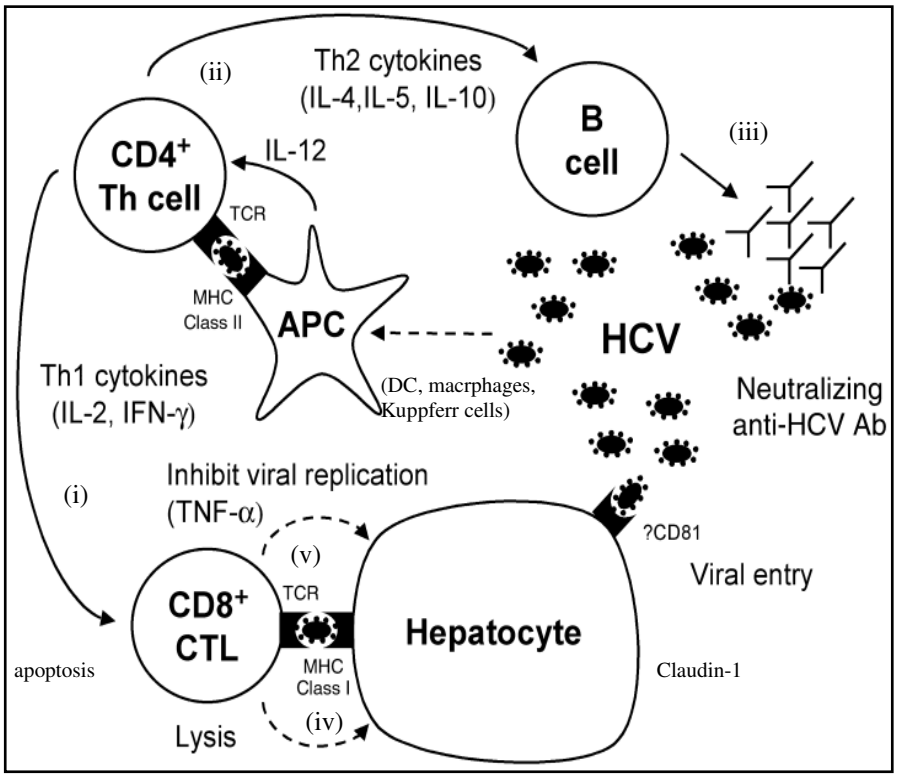


Fig.8. The processes involved in the adaptive immune response when HCV infects the antigen-presenting cell (APC) and the hepatocyte (adapted from Freeman et al., 2001) . The class II-restricted Th cell response: (i). Induces virus –specific CTL (Th1 pathway promoted by Th1 cytokines [IL-2, IFN-gamma] and (ii) helps HCV-specific B cells (Th2 pathway) promoted by Th2 cytokines [IL-4, IL-5, IL-10] and (iii) to produce neutralizing antibodies. The anti-viral CD8+ CTL response: (iv) identifies and kill virus-infected cells through recognition of endogenously synthesized HCV antigen presented to HLA class-I molecules at the hepatocyte membrane and (v) they inhibit replication by secretion of the cytokines IFN-γ and TNF-α triggering their target hepatocyte membrane to undergo apoptosis.

1.6.3.1. Humoral Response (T_H2- antibody response)

NAbs appear several weeks after infection (Battegay et al., 1993). The core, E2, NS3, NS4B and NS5A peptides were identified as immunogenic proteins in chronic infection with HCV genotypes 1, 2, 3 and 4 (Sillanpaa et al., 2009). However, E1 and particularly E2 peptides seem to produce NAb responses (Zeisel et al., 2009). During acute infection, cellular immune responses increase and then decrease and lose function (Osburn et al., 2010) and then NAbs appear. In chronic infection,

weak cellular immune responses with cross-reactive NAb responses was evident (Lechner et al., 2000). The delayed humoral response in chronically infected individuals drives viral escape (Zeisel et al., 2009). While the role of NAbs during re-infection is not fully understood, cross-reactive NAbs in patients with frequent re-infections were seen, indicative that the humoral responses are primed to generate broadly NAbs (Osburn et al., 2010).

1.6.3.2. Cell-Mediated IR

An ineffective response is reflected when CD4+ and CD8+ T-cell responses are weak or even absent while viraemia persists (Fig. 9a). These proliferative responses are important to control infection but the breadth of these responses is vital for viral clearance. Breadth is measured as the number of HLA-restricted epitopes to which the T-cell can bind. CD4+ T-cells recognize as many as 14 different epitopes presented by the MHC class II-complex (Day et al., 2002). In patients with chronic infections, very few MHC class II epitopes are targeted in contrast to individuals that resolve infection, reflecting the ineffective response. Weak anti-HCV CD4+ responses correlate with viral persistence (Thimme et al., 2002, Fig. 9a,b) while good responses correlate with viral control and clearance (Fig.9c).

The CD8+ CTL responses are the best characterized and have been found to be variable in acute and chronic infection. A sustained vigorous response during acute infection correlates with viral clearance (Chang, 2003, Fig.9c). CD8+ CTL responses have been shown to peak with increased ALT levels (Tillman et al., 2001, Fig.9b, c). Strong CD8+ T-cell responses that target many MHC-class I restricted epitopes have been associated with good immunity as compared to poor responses

to fewer epitopes which may lead to viral persistence (Grakoui et al., 2003,). However, the lack of memory T-cell induction in cross challenge studies have shown that CD8+ T-cells fail in their function during HCV infection (Farci et al., 1995). This is due to broad viral diversity and escape from the immune system (Encke et al., 2007, Fig.9d). CD8+ T cells localized to the liver protect from re-infection as they are capable of rapid expansion (Shoukry et al., 2003). Antibody depletion of CD8+ and CD4+ T-cells (Fig.9d) results in persistent viraemia until memory CD8+ T-cell responses are restored in the liver (Shoukry et al., 2003).

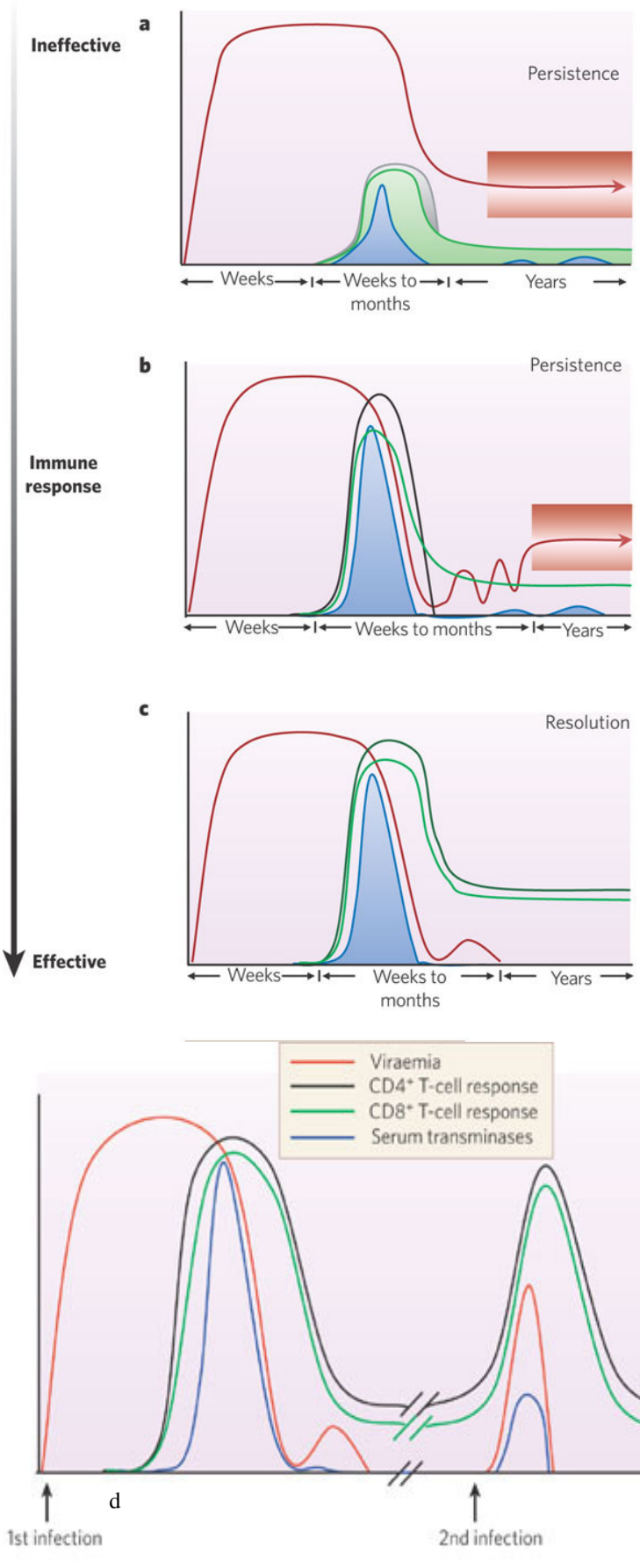


Fig.9. CD4+ and CD8+ T-cell responses during progression or clearance of HCV infection. (a) Viraemia rises early and then falls but is still present throughout infection. CD4+ and CD8+ T-cell responses are weak or even absent and serum transaminases are present. (b) CD4+ and CD8+ T-cell responses are stronger and associated with transient control of viraemia, with a rise in serum transaminases. T-cell responses are not maintained and viraemia rebounds in persistent infection. CD8+ T-cell responses may persist during persistent viraemia. (c) Although viraemic at early stages, CD4+ and CD8+ T-cell responses are strong and maintained. Serum transaminases rise and viraemia decreases, with possible rebound before clearance. (d) Experimental data on challenge studies on monkeys with HCV, indicating the decline of CD4+ and CD8+ T-cell responses over a period of time, providing a means for virus escape and persistence of infection (Bowen and Walker, 2005).

1.7. CLINICAL FEATURES

Most studies have reported that cirrhosis develops in 10%-20% of persons with chronic hepatitis C over a period of 20-30 years, and HCC in 1%-5%, with striking geographic variations in rates of this disease. Persistent HCV infection is the cause of 70% of HCC cases in Japan (Yoshizawa, 2002).

1.7.1 Acute infection

The incubation period for acute hepatitis C averages 6 to 10 weeks (WHO, 2002). Most persons (~80%) who develop acute hepatitis C have no symptoms (Mast et al., 1999, Fig. 10). The onset of disease is usually insidious, with anorexia, vague abdominal discomfort, nausea and vomiting, fever and fatigue, progressing to jaundice in about 25% of patients. Jaundice is less frequent in hepatitis C than in hepatitis B virus (HBV) infection (Marcellin, 1999). Severity ranges from asymptomatic (75% of infections) to rapid, fulminant liver failure. Spontaneous elimination of the virus is rare and 70 – 90% of infected people fail to clear the virus during the acute phase of the disease and become chronic carriers (Fig. 10; WHO, 1999). Anti-HCV antibody is not always present when the patient presents with symptoms. It may take 10 weeks or more from the time of infection to develop detectable antibodies to HCV (MMWR, CDC Report, 1997; Fig. 11a). However, HCV RNA may be detectable in serum as early as 1-3 weeks after exposure. The silent clinical nature of HCV infection may account for the large reservoir of infected persons.

1.7.2 Chronic infection

Chronic hepatitis is diagnosed when disease continues, without improvement, for at least six months (Fig. 11b). Chronic hepatitis C is not a single disease, but rather a

complex clinico-pathological syndrome with multiple causes and varying stages of necro-inflammatory and fibrosing liver damage. This complexity leads to different prognoses. Persistent or fluctuating ALT elevations (MMWR, CDC Report, 1997; Fig.11b) may indicate active liver disease while the remaining 30-40% may have normal ALT levels. No clinical or epidemiologic features among patients with acute infection have been found to be predictive of either persistent infection or chronic active liver disease (MMWR, 1998).

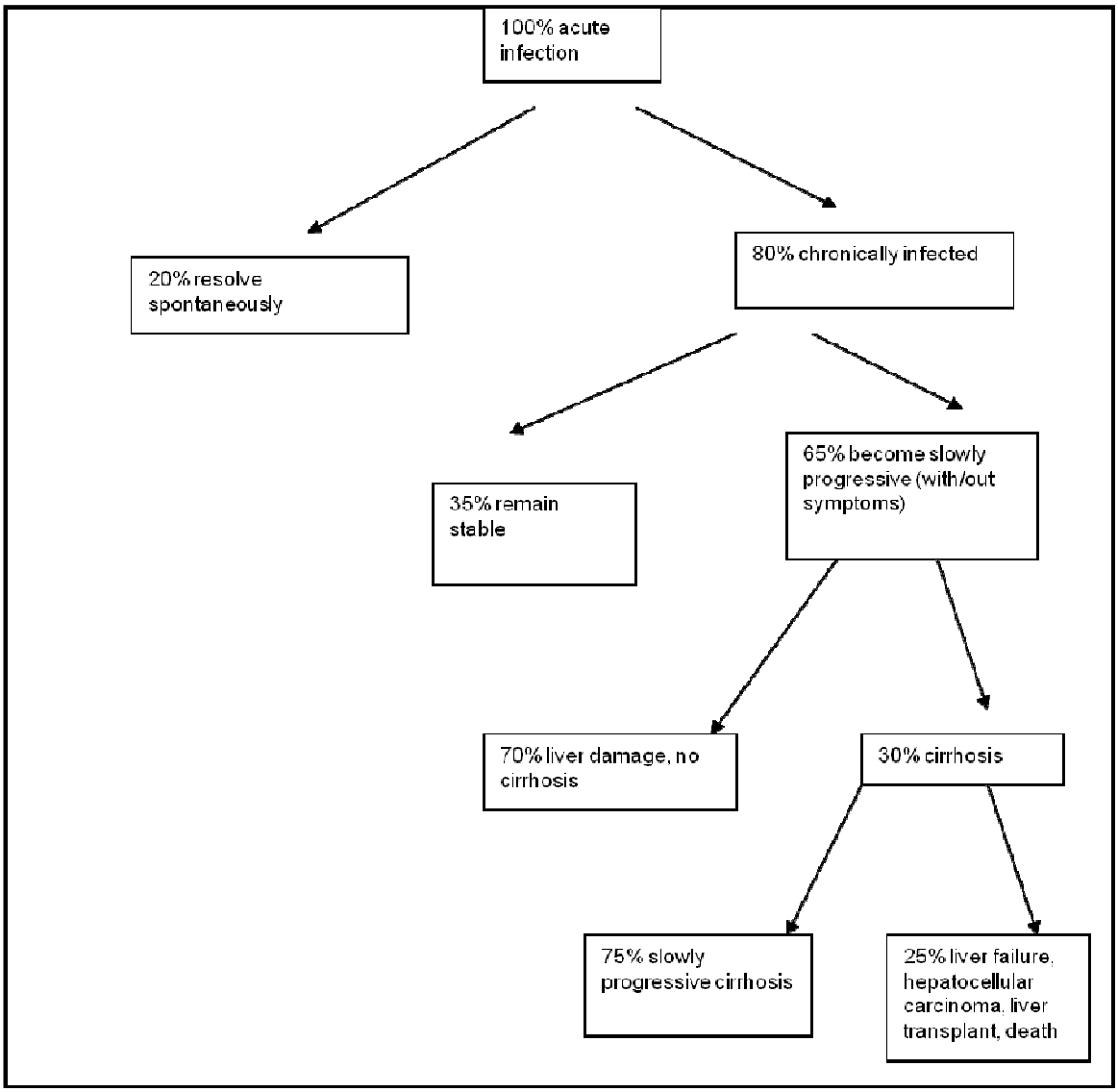


Fig.10. A flow diagram representative of the various stages of HCV mono-infection in humans, from acute to progressive disease (adapted from MMWR, 1998). HIV and/or other co-factors, for example, alcohol abuse, will increase the risk of disease progression.

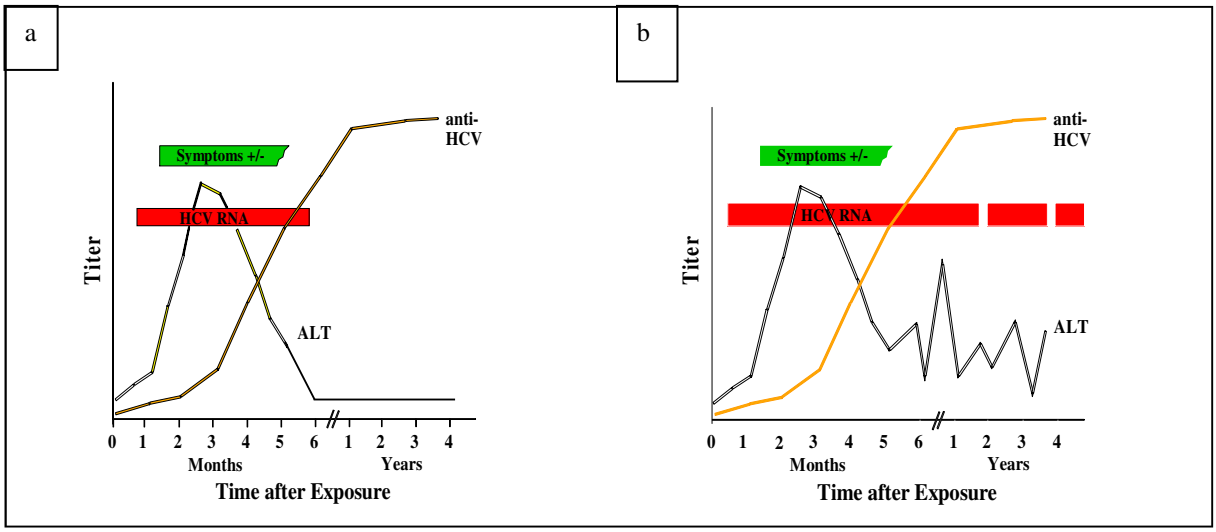


Fig.11. Serologic Pattern of Acute HCV Infection with (a) Recovery and (b) progression to Chronic infection (adapted from MMWR, CDC Report, 1997)

1.7.3 HCV/HIV co-infection

Studies have shown that people infected with HIV have increased levels of HCV viraemia and the risk of cirrhosis (Pol et al., 2002). The exact mechanism for the pathogenesis of HCV in HIV-infected persons is not fully understood but, since HCV has not been shown to be cytopathic, the liver injury associated with chronic liver disease may be related to the host immune system (Sterling et al., 2003). Patients with CD4 counts below 200 cells/µL have an increased risk of severe liver disease in co-infection (Graham et al. 2001; Goedert et al. 2002) and it has been postulated that marked decrease in the CD4 cell count may result in selection of HCV quasispecies with increased cytopathogenicity (Pol et al., 2002). HIV itself may have a direct cytopathic effect on liver cells or may upregulate (or downregulate) cytokine production leading to severity of liver disease (Pol et al., 2002).

1.7.4. HBV/HCV co-infection

HBV and HCV co-infection is more common in areas where both viruses are endemic (Jamma et al., 2010), like Japan, where 23% of HBsAg-positive patients were found to anti-HCV positive (reviewed in Liu and Hou., 2006). In SA, very little is

known of HBV/HCV co-infection but this may because of poor collation of data in the past.

1.8. DIAGNOSTIC TESTS FOR HCV

There are several markers that can be used to detect HCV infection: indirect and direct markers, indicating presence of antibody to HCV or the presence of the virus itself, respectively. Table 5 summarizes the various tests for the detection, quantitation and genotyping of HCV.

Table 5. A summary of the diagnostic and research tests for the detection of HCV antibody, RNA and genotypes.

Test	Purpose	Comments	Reference
Serology (Anti-HCV)	Screening tests for HCV antibody (indirect marker) by EIA or RIBA	Most widely available and cheap test Uses core, NS3/4/5 proteins High sensitivity (99%) – 3rd generation EIA. Indicates past or present infection. Detects anti-HCV in 80% of patients within 5-6 weeks of onset of hepatitis. Late seroconversion can occur. Screen in high prevalence populations High false-positive rate for EIA in low prevalence populations and in those with autoimmune disorders. A positive result needs confirmation by NAAT	Aoki et al., 1996 Richter, 2002
Serotyping	Serotyping HCV 1-6 by EIA	Use of NS4 peptides Average sensitivity (85-90%) when compared to molecular-based genotyping tests Difficulty to subtype and detect mixed infections	Chemello et al., 1994 Prescott et al., 1997
Core antigen	Detection and quantification of core antigen by EIA	Less sensitive than molecular PCR tests Lower limit of detection is 20000IU/ml	Takahashi et al., 2005
Qualitative PCR · PCR Polymerase chain reaction	Detect presence or absence of virus. Detects virus 1-3 weeks after exposure. Detection of HCV RNA during course of infection may be intermittent. A single negative PCR is not conclusive. Repeat test or repeat anti-HCV.	False positive and false negative results can occur from improper handling, storage, and contamination of test samples. Amplicor HCV qualitative test has high sensitivity and specificity (95%), lower limit of detection is 50IU/ml. Used at the NICD	Gretch et al., 1992 Lee et al., 2000

Quantitative tests · PCR · Viral load	Determines titer of HCV To monitor patients on antiviral therapy	Should preferably not be used to confirm or exclude the diagnosis of HCV or to monitor treatment end point. Taqman has high sensitivity, with lower detection limit of 15IU/ml. Easy, quick and reproducible but expensive. No/little cross contamination as all steps in one tube. Used for patients on therapy. Used at NICD	Kawai et al., 1999 Aoki et al., 1996 Gretch et al., 1995 MMWR, CDC reports, 1997, 2008
TMA- Transcription-mediated amplification	Amplifies HCV RNA by T7 RNA polymerase	Good correlation with other PCR tests. Lower limit of detection is 5IU/ml.	Krajden et al., 2002 Ross et al., 2001
Genotype - LiPA	Recommended when considering therapy regime and management	LiPA is a nitrocellulose dot blot assay, quick and easy Can detect mixed infections. Poor differentiation of genotypes 1a and 1b in 5'UTR but can use core + 5'UTR version. Lower limit of detection is 1000IU/ml.	Andonov and Chaudhary, 1995 Roque-Afonso et al., 2002 Buoro et al., 1999 Verbeeck et al., 2008b Noppornpanth et al., 2006b
Genotype - sequencing	Uses the 5'UTR to genotype HCV 1-6 and less conserved regions, core and/or the NS5B for subtyping.	Sequencing – gold standard, highly specific and sensitive for all genotypes. Phylogenetic analyses groups isolates into 6 genotypes based on genetic differences. Lower limit of detection is 1000IU/ml. Specialized technique, time consuming, available as diagnostic and research tool at the NICD.	Simmonds et al., 1993a,b, 1994b, 1996, Chen and Weck, 2002 Kalinina et al., 2002 Accapezzato et al., 2002; Ohno et al., 1997, Ohno and Mizokami,1999
Genotype - real-time Lightcycler	specific primers and hybridization probes in the 5'UTR for amplification and detection of point mutations using temperatures at which 50% of probe-target duplexes separate (Tm).	No/little cross contamination as all steps in one tube. Quick and easy but expensive. Lower limit of detection is 10000IU/ml. Does not differentiate subtypes and mixed infections.	Landt et al., 2002 Yang et al., 2002 Schroter et al., 2002
Genotype - RFLP	detection of amplicons fragmented by restriction endonucleases at specific sites of nucleotide sequences. Restriction patterns are visualised on agarose/ polyacrylamide gels.	Time-consuming, no commercial kit available, trial and error with restriction enzymes, contamination is moderate to high does not effectively separate the subtypes of genotype 1 specimens (1a/1b). Sensitivity comparable to PCR and sequencing (92%).	Davidson et al., 1995 Pohjanpelto et al., 1996 Buoro et al., 1999
Genotype - Trugene	Genotyping kit based on sequence analysis of the 5'UTR region	Trugene sequencing has 100% specificity and 97% sensitivity when compared to LIPA, reproducible and reasonably rapid (<5hours), subtyping results was often mixed	Halfon et al., 2001

Genotype -heteroduplex tracking analysis (HTA)	Uses E1 and E2 genes Genetic relatedness is estimated by calculating the degree of migration in high-resolution gel electrophoresis and compared to homoduplex controls.	E1 HTA studying HCV transmission between individuals, E2 HTA for tracking HCV quasispecies evolution in an infected individual, effective and inexpensive molecular technique for screening HCV heterogeneity, time consuming, contamination is moderate to high, works well if viral load is high	Gretch et al., 1996 Gerotto et al., 1999 Sullivan et al., 2001

1.8.1. Phylogenetic analyses

The measurement of sequence changes, either nucleotide or amino acid, can be performed by phylogeny, using different models (revised in Lemey et al., 2009). The basic model, p-distance, measures the genetic distance from the ancestral sequence to its derivatives but is an under-estimation of genetic change. The Jukes-Cantor (JC69) model (Jukes and Cantor, 1969) is the simplest substitution model and assumes that all base frequencies and mutation rates are equal. The Kimura-2 parameter (Kimura, 1980) assumes that all base frequencies are equal but transitions (A<->G, purine to purine or C<->T, pyrimidine to pyrimidine) and transversions (purine<-> pyrimidine and vice versa) occur at different rates. The Hasegawa-Kishino-Yano (HKY) model (Hasegawa et al., 1985) assumes that base frequencies are unequal and transitions and transversions occur at different rates. The most complex is the general time reversible model (GTR, Tavare, 1986) and assumes that base frequencies are unequal and all substitution types occur at different rates (revised in Lemey et al., 2009). Once a model is chosen to measure the nucleotide changes, an algorithm for the drawing of phylogenetic trees can be applied.

The simplest and least computer-intensive method is the use of the neighbor-joining (NJ) algorithm that calculates the distance between two sequences, generating a

cluster tree (in Lemey et al., 2009). Maximum-likelihood methods are more computer intensive and time-consuming but offers a more direct way of finding a tree for a set of data using the probability model. Bootstrap values on tree nodes indicate the repeated sampling of data and reconstruction of trees to provide support for the clustering of the species (sequence). Branch lengths are a measure of the amount of divergence between two nodes in the tree. The unit of measure (for the branches and scale bar) is number of nucleotide (or amino acid) substitutions per site(Kumar et al., 2008).

A rooted tree is a directed tree with an outgroup species/sequence (usually a most recent common ancestor, MRCA) that infers a relation to other species on the tree. An unrooted tree will depict the relation of all the species/sequences on the tree without an inference to ancestry (in Lemey et al., 2009). The molecular clock phylogeny uses a rooted tree to estimate the changes from the ancestor (root) to the subsequent species/sequences over a unit of time (in Lemey et al., 2009). This is further discussed in chapter 7.

1.9. FURTHER MEDICAL EVALUATION OF HCV DISEASE

Histological assessment such as fibrosis scores (Ishak et al., 1995), as well as other factors, such as quality of life, age, co-morbidity and virological factors influence the risk-benefit ratio of treatment and can be considered in the treatment decision especially in patients with mild chronic hepatitis or chronic hepatitis infection associated with normal transaminase levels (Botha et al., 2005). However, fibrosis is not a requirement for treatment (Botha et al., 2010). If liver histology is available for a patient, then treatment is considered if fibrosis is advanced (F2 or F3, according to METAVIR scoring system, Bedossa and Poynard., 1996) (Botha et al., 2010).

1.10. MANAGEMENT AND TREATMENT

1.10.1 Drug therapy for HCV

In South Africa, treatment should be considered in all adults with confirmed diagnosis of hepatitis C infection and chronic disease, especially individuals who are at risk of developing cirrhosis (Botha et al., 2010). Patients with acute hepatitis C should be treated with standard interferon (IFN) monotherapy (Gerlach et al., 2003) or PEG -IFN (Botha et al., 2010). The current standard hepatitis C treatment in South Africa, is weekly injections subcutaneously of PEG-IFN-α, combined with twice-daily oral doses of the broad-spectrum antiviral agent ribavirin (800-1400 mg/day depending on body weight; Botha et al., 2010). Two PEG-IFN medications are available, PEG-IFN alfa-2b (1.5 µg/kg/wk; Ghany et al., 2009; Peg-Intron, Schering-Plough) and PEG-IFN alfa-2a (Botha et al., 2005; 180 µg weekly; Pegasys Roche).

The rationale for treatment of chronic hepatitis are:

- to reduce inflammation,
- to prevent progression to fibrosis, cirrhosis and HCC in chronically infected patients through the eradication of the virus, and
- to decrease infectivity as a means to control the spread of the disease.

The following are the recommended, genotype-dependent, adult treatment durations (from Botha et al., 2005, 2010):

►Genotype 1 and 4 – 48 weeks of treatment

►Genotype 2 and 3 – 24 weeks of treatment

►Genotype 5 and 6 – 48 weeks of treatment (This treatment regimen is still experimental as there is not enough data available at present regarding

treatment outcome in HCV genotype 5 infection to meaningfully predict response to treatment, Botha et al., 2010)

In genotype 1-infected slow responders, an extension of treatment to 72 weeks can be considered (Pearlman et al., 2007). It is recommended that for those patients receiving continued therapy through 48-72 weeks and if HCV RNA is not detected at end of treatment, then retesting for HCV RNA after 24 weeks is done to determine a sustained virological response [SVR] (Botha et al., 2010).

In genotype 2/3-infected slow responders, an extension of treatment to 48 weeks can be considered. It is recommended that for those patients receiving continued therapy through 24-48 weeks and if HCV RNA is not detected at end of treatment, then retesting for HCV RNA after 24 weeks is done to determine a SVR (Botha et al., 2010).

Approximately 40-50% of individuals infected with genotypes 1b and 4 have achieved a SVR (after 48 weeks of treatment) as compared to 80% of genotype 2 and 3 infected patients after 24 weeks of treatment (Manns et al., 2001, 2006, Table 6). Genotype 5 (Legrand-Abravel et al., 2004; Nguyen and Keeffe, 2005) and genotype 6 has shown sensitivity in between genotype 1 and genotype 2/3 (Hui et al., 2003, Table 6), although data for these studies have been small and limited. Recently it has been reported that genotype 3 infections are associated with lower sustained virological responses than genotype 2 infections (Powis et al., 2008).

Table 6. Predictors of good and poor responses to HCV therapy

HCV genotype	GT2 and GT3 are more sensitive to treatment than GT1 or GT4 and GT5 appears to be intermediate between GT1 and GT2 and GT3. (Nguyen and Keeffe, 2005)
HCV subtypes	1b, 4a or 4d may respond better to therapy as compared to those with subtype 1a (Legrand-Abravanel et al., 2009a)
HCV viral load	Treatment is more effective at viral loads <400 000IU/ml
Early treatment	response rates were better when therapy began within 12 weeks of diagnosis (Corey et al., 2010)
Young age	<40 years (Ghany et al., 2009)
Gender	Female
Race	Treatment was found to be less effective in African-Americans (Donlin et al., 2007)
Amount of liver damage and steatosis	less effective with cirrhosis and steatosis
Genotype and liver damage	Patients with GT3 and cirrhosis respond less well to therapy than those with GT2 and cirrhosis (Powis et al., 2008).
HIV status	Less effective in HIV co-infection
Adherence to therapy	Full dose of ribavirin and PEG-IFN at least 80% of the time
Body weight	Less effective for people who weigh more than 75kg (Ghany et al., 2009), obesity is a risk factor for steatosis
Effective management of side effects	Neutropenia, thrombocytopenia, depression and anaemia (Botha *et al*., 2005).
insulin resistance	Poor prognosis if present (Ghany et al., 2009)

1.10.2 HCV/HIV co-infection and treatment

Management of HCV/HIV co-infection is complex in South Africa and the infected individual should be referred to a specialist for appropriate treatment and management. HIV drugs, like Didanosine (DdI) and zidovudine (AZT) can interact with ribavirin and causes lactic acidosis, pancreatitis and the risk of liver failure in people with advanced cirrhosis with the increased risk of anaemia (Brook et al., 2010). It was indicated that highly active antiretroviral therapy, or HAART, has been related to histological deterioration, direct viral cytotoxicity or increased drug hepatotoxicity (Pol et al., 2002). Although, PEG-IFN and ribavirin have proved to be effective in HIV/HCV co-infected persons (Chung et al., 2004), further ongoing studies on long-term effect on outcome and survival need to be done.

Some of the newer drugs that are currently in phase 1 trials include:

- Specifically Targeted Antiviral Therapy for hepatitis C (STAT-C) antivirals drugs directly target enzymes required for HCV replication and production of new virions (April 2009)
- polymerase inhibitors: ABT-072 (December 2009) and MK-3281 (November 2009); RG7128 (April 2009)
- Clemizole (August 2009) is an NS4B inhibitor
- Protease inhibitors: ACH-1625 (June 2009) and RG7227(April 2009)

(summarized in Rodriguez-Torres, 2010; Franciscus, 2010)

1.11. PREVENTION AND CONTROL

The major objectives of Hepatitis C prevention and control are:

- to reduce the number of new infections through prevention of transmission

- to identify cases early in order to minimize the risk of disease progression
- to educate all high-risk individuals
- to implement infection control practices to limit spread in nosocomial setting
- to introduce needle exchange programmes and educate public to reduce the spread of HCV, HBV and HIV in South Africa
- to caution HCV-positive mothers to refrain from breast feeding if nipples are cracked or bleeding.

To date, there are no effective vaccines or immunoglobulins available for the treatment and prevention of HCV infection. However, there are currently two therapeutic peptide vaccines undergoing phase II trials: IC41(Intercell) and GI-5005(Tarmogen, Globe Immune) and one DNA-based therapeutic vaccine, ChronVac-C (Onovio/ Tripep), undergoing phase I studies on for genotype-1 infected individuals (Franciscus, 2010).

The literature has pointed out that many studies have focused on genotypes 1a and 1b as these account for the majority of infections in the developed countries. The purpose of this thesis was to gather information on HCV prevalence and the genotype distribution in different study groups, describe the molecular characteristics, particularly of genotype 5a, predominantly seen in South Africa, determine binding efficiencies of putative vaccine epitopes to prevalent HLA alleles in the SA population and establish the response to therapy in patients with genotype 5a. The present study has endeavored to: update information on HCV in SA, improve the notification and surveillance reporting, allow informed decisions when planning preventative and treatment strategies and increase public awareness/education.

1.12. STUDY OBJECTIVES

(1) RATIONALE OF THE STUDY

Although South Africa is a low-endemic region for HCV infection, the contribution of HCV to liver disease in the country is unknown as HCV testing has been limited and surveillance systems are lacking. It remains imperative that the most rapid and reliable tests are available and validated for accurate diagnostic screening and genotyping. The establishment of baseline data is pressing in South Africa since the impact of immigration and travel may contribute to changing frequencies of genotypes in the country. Well designed databases will allow informed decisions when planning preventative and treatment strategies relevant to local viral and host genetics. The major aim of this project was to characterize South African HCV strains so as to provide valuable epidemiological information and to select suitable vaccines and therapeutic strategies in the future.

(2) SPECIFIC OBJECTIVES

(a) To establish a national HCV database for the public sector, validate and optimize methodologies to determine baseline prevalence of HCV, including the occurrence and frequency of the various genotypes, in available South African cohorts.

(b) To evaluate rapid techniques for genotyping HCV (by real-time PCR and by line probe assays) using specimens well characterized by PCR and sequence analysis.

(c) To add local genome sequence data to the global HCV database and to study the phylogenetic relatedness of these data.

(d) To molecularly characterize the unique and poorly studied HCV genotype 5

strain because this is the most prevalent genotype in South Africa.

(e) To use the inferred consensus amino acid sequence of the genotype 5 HCV to map immunodominant regions in the context of host genomics (eg. HLA) and thus provide baseline information for therapeutic and vaccine strategies.

(f) To quantitatively describe the response rates (in weeks) of patients infected with a genotype or mixed genotypes of HCV to combination therapy.

CHAPTER TWO

Establishing the framework for better HCV diagnosis, data collection and management in SA

2.1. INTRODUCTION

Although Hepatitis C is one of the notifiable diseases in SA, there were only two cases reported to the National Department of Health (NDoH) during 2008-2009 (Statistics notes, http://www.doh.gov.za/facts/index.html, November 2009). This poor reporting and almost total lack of information on HCV and its contribution to liver disease in South Africa was the driving force behind this study. Developed countries, like the UK, Europe and USA have established surveillance databases to facilitate the notification of hepatitis diseases, improve management and treatment of diagnosed cases, and raise public and clinician awareness (Centers for Disease Control and Prevention, CDC, 2005 http://www.cdc.gov/hepatitis; Health Protection Agency, HPA, 2009, http://www.hpa.org.uk).

Presently, HIV, tuberculosis (TB) and malaria are major health problems in SA. However, for diseases with low prevalence such as HCV, where "prevention is better than cure", focus and organization are lacking. Systems to identify, control and treat HCV infection in order to prevent the development of its serious chronic sequelae, cirrhosis and HCC and their subsequent burden on health care costs later do not exist. The NHLS can provide this framework as it provides diagnostic laboratory services to over 80% of the population through approximately 265 laboratories (National Health Laboratory Services, NHLS Annual report, 2009-2010, http://www.nhls.ac.za).

Previous reports on hepatitis C have focused mainly on serological tests and few performed molecular analyses and even these did not use more modern viral load and sequencing methods (Smuts and Kannemeyer, 1995; Tucker et al., 1997). Although antibody tests for HCV are convenient and cheap for screening large populations, the antibody test does not distinguish between active (present) or past infection. HCV diagnosis must be confirmed by PCR and infected individuals require further medical evaluation before therapy can be considered (Fig.6, page 42). A negative antibody result with a positive PCR can indicate (1) an early stage of acute infection prior to the development of antibodies, (2) chronic infection in an immunosuppressed individual or (3) a false positive RNA PCR result (Ghany et al., 2009).

Reliable and early detection of HCV can facilitate an early decision to treat. It has been reported that early treatment can prevent progression to chronic disease (Poynard et al., 2000) or, failing this, produce better response rates to therapy (Corey et al., 2010). Early diagnosis, followed by counseling, enables the infected individual to make informed choices with respect to lifestyle [to reduce transmission], habits [to prevent exacerbating disease] and regular follow-up [treatment options, liver fibrosis stage] (Seymour, 1996). Although South Africa is a low-endemic region for HCV infection, it remains essential that the most rapid and reliable tests are available and validated for diagnostic screening. The first cohort consisted of seropositive specimens from hospitals in the Johannesburg area and these were screened for active disease by establishing and validating a qualitative PCR.

The aim of the study was to establish a national database to record serological and PCR positive HCV results together with available demographic information.

2.1.1. Objectives:

(a) Confirmation of antibody seroprevalence results with nucleic acid amplification testing in a large cohort of anti-HCV positive specimens from Johannesburg hospitals

(b) To determine the number of individuals infected with HCV by (i) seroprevalence and (ii) PCR positive cases in the public health sector leading to the establishment of a national HCV surveillance system.

2.2. METHODS

2.2.1. Nucleic acid amplification test (NAAT)

Three hundred and eight HCV antibody seropositive serum specimens were obtained from provincial hospitals within the Johannesburg area (ethics clearance M051114, see appendix). The specimens had been sent to the laboratories of the NICD for routine hepatitis serology testing during the years, 2000, 2001 and 2002. Antibody tests were done as per routine diagnostic procedure at the NICD hepatitis serology laboratory. Specimens were tested for HCV RNA by NAAT, using the Amplicor HCV Qualitative assay, version 2.0, (Roche Diagnostics, Germany). The test was validated in our laboratory and subsequently accredited by the South African National Accreditation System (SANAS) in 2004.

2.2.2. National HCV Surveillance Database

Results for requested HCV tests are reported on the NHLS secured online laboratory data capture system (DISA). In the last few years, the results from all national laboratories have been retrieved from DISA and captured on the NHLS Corporate Data warehouse (CDW, ethics clearance M060449, see appendix) (Fig.12). The

hospitals in the rural and urban areas of South Africa send specimens for testing to an NHLS laboratory, usually in close proximity to the clinic/hospital and the requested test is performed. The results are deposited in a local repository, collated in a central database and acquired from CDW using specific test codes and parameters. The HCV data is extracted from all NHLS laboratories countrywide and presented in an Excel format for surveillance reporting. The NICD of the NHLS is an example of a laboratory that performs HCV PCR tests for hospitals/clinics in the following provinces: Gauteng, Mpumalanga, Free State, Limpopo and Kwa-Zulu Natal. The internal validation of the system checks each DISA entry with the entries received from the CDW. Validation checks for the Western Cape Province and external audits on the CDW retrieval process are projected for 2011. Although serological tests were performed prior 2007, the collation of serology data only began from 2007 to 2008 and the PCR data from 2004 to 2008. All relevant data (age, gender, hospital/clinic, ward, test result, test date, date of specimen receipt) was mined on MS Excel spreadsheets and interpreted as tables and figures.

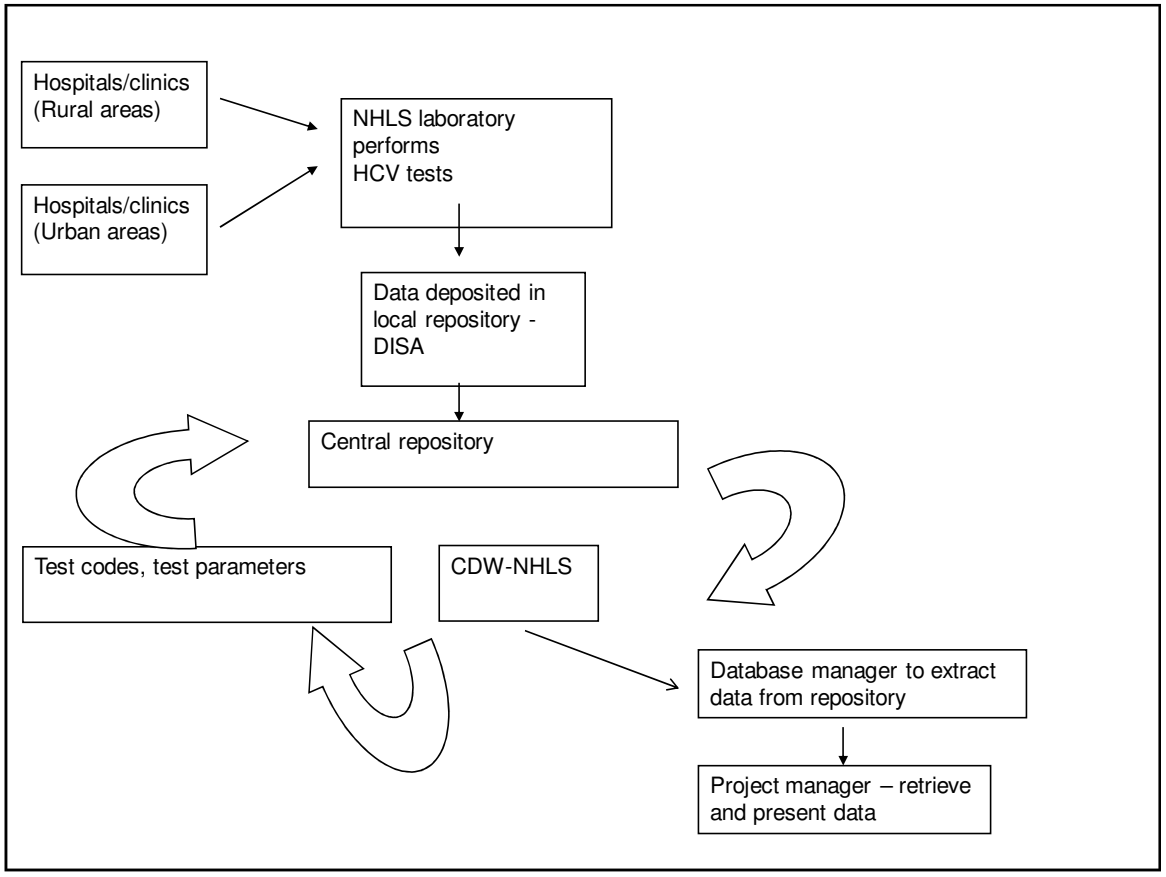


Fig.12. A summarized flow diagram to explain the data processing and capturing from the time of referral for HCV tests until the result is captured on the HCV National Surveillance Database by the project manager (student)

2.3. RESULTS

2.3.1. Nucleic acid amplification test (NAAT)

In the first part of this study, it was observed that 75% of specimens tested had shown concordance between serology and RNA tests. Of the 308 specimens, 232 (75,33%) specimens were positive by serology and RNA. Sixty-five (21,10%) were positive by serology and negative for RNA, indicating discordant results and re-iterating that confirmation by PCR is necessary. Eleven (3,57%) were positive by serology but showed PCR inhibition as evident by the lack of amplification of both the target RNA and the internal control of the qualitative assay. The confirmatory testing process has led to the next part, whereby both serology and PCR data were captured onto a national database.

2.3.2. National HCV Surveillance Database

The collation process was validated for 2008 by retrieving data manually from the DISA system and compared to the CDW retrieval for the Gauteng province. There was a 100% concordance for the data retrieval on 721 specimens (including all repeats) and this verifies the extraction process by CDW. This internal validation process for all other provinces will proceed in 2011.

There was an increase in the number of hospitals/clinics at various geographical locations nationwide that requested HCV serology and PCR tests (Fig.13). In 2007, there were 175 sites nationally that requested HCV serology tests with an increase to 235 sites in 2008. The number of PCR requests were fewer but increased over the years from 12 in 2004 to 29 in 2008 (Fig.13).

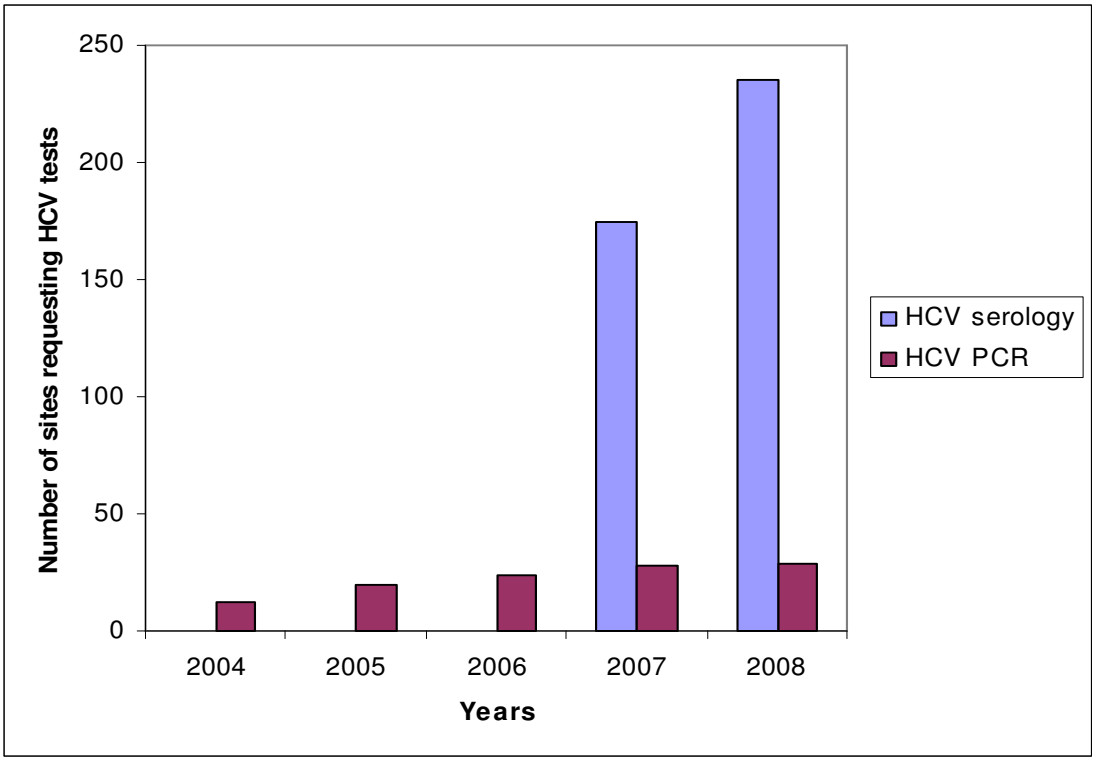


Fig.13. The number of hospitals/clinics nationwide requesting HCV tests (serology and PCR) respectively.

Table 7 summarizes the serology results for the years 2007 and 2008 and PCR results for years 2004-2008. Although the total number of serology requests appear to have decreased from 2007 to 2008, there was a higher recorded HCV seropositivity in 2008 (1745/18128, 9.6%) as compared to 2007 (1371/25944, 5.28%). There were more males than females in the 2007 group compared to the 2008 group (Table 7). The mean age for the serology positive group and PCR positive group ranged from 37 to 47 years. Many of the individuals may have had first signs of chronic hepatitis, hence were referred for HCV testing as most cases of persistent disease are asymptomatic in the early stages.

Table 7. A descriptive summary of the serology (2007-2008) and PCR (2004-2008) results

	SEROLOGY		PCR				
Year	2007	2008	2004	2005	2006	2007	2008
Total No.	25944	18128	132	285	342	483	490
negative	24573	16383	89	206	229	288	299
positive (%)	1371 (5.28)	1745 (9.6)	43 (32.5)	79 (27.71)	113 (33.04)	195 (40.37)	191 (38.97)
Males (%)	708 (51.6)	828 (47.4)	23 (55.5)	50 (63.3)	57 (50.4)	119 (61)	114 (59.7)
Females (%)	618 (45.1)	871 (49.9)	17 (39.6)	26 (32.9)	49 (43.4)	71 (36.4)	73 (38.3)
Unknown (%)	45 (3.3)	46 (2.7)	3 (6.9)	3 (3.8)	7 (6.2)	5 (2.6)	4 (2)
Mean age (years)	38.8	39.9	37	45	44	45	47
No. Age unknown (%)	99 (7.2)	108 (6.2)	5 (11.6)	6 (7.4)	11 (9.7)	12 (6)	6 (3)

For this analysis, it was not indicated whether any of the samples tested serologically were tested by NAAT. Fig.14 indicates the PCR positive results for the years 2004-2008, indicative of current infection. The number of PCR positives has increased over the years from 43 cases (43/132, 32%) in 2004 to 195 (195/483, 40%) and 191 cases (191/490, 39%) in 2007 and 2008 respectively.

This probably reflects an increasing awareness of the availability of HCV tests among clinicians in South Africa, also indicated by the number of hospitals requesting HCV tests (Fig.13).

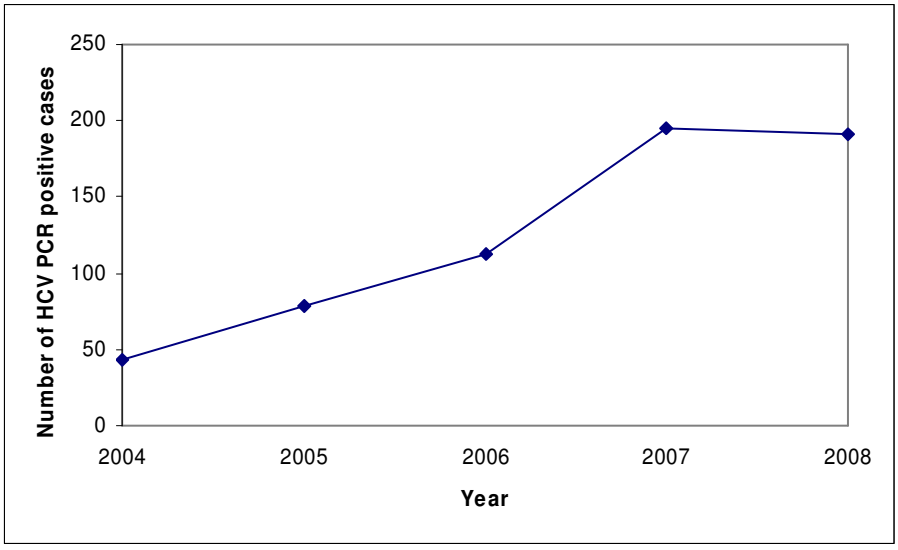
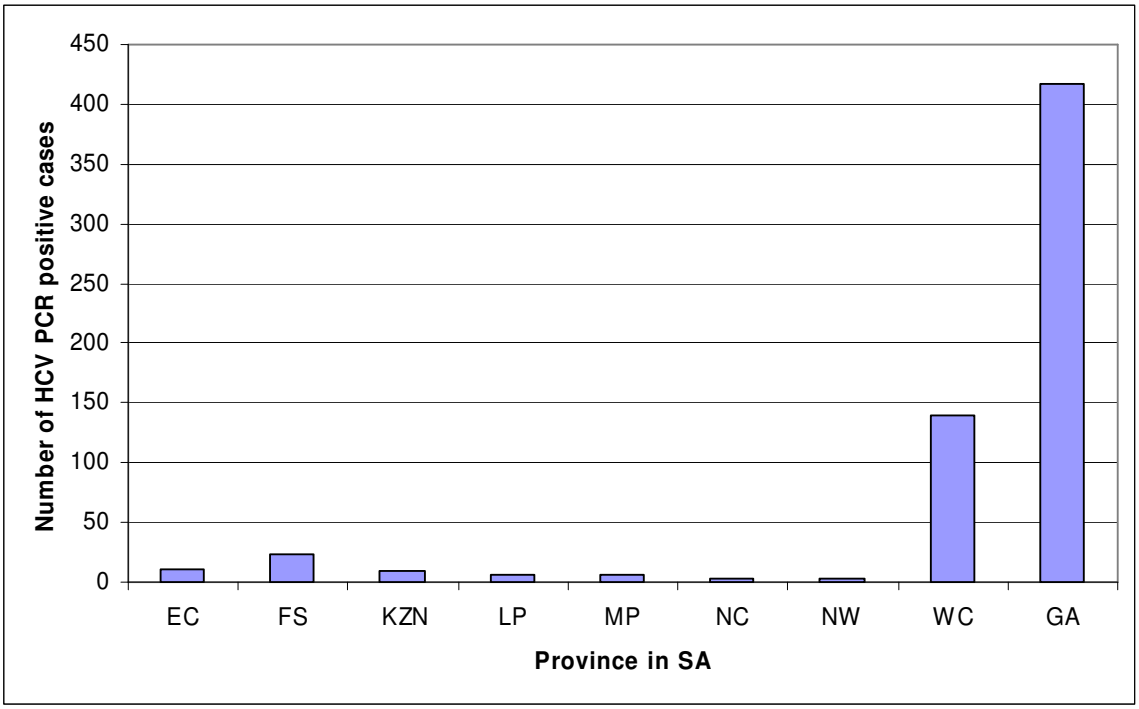


Fig.14. Total number of HCV PCR positive cases, 2004-2008, South Africa

Fig.15 indicates the number of cases positive by PCR per province for the years 2004 to 2008. Gauteng had the most HCV PCR positive cases (67%), followed by Western Cape (22%). The Northern Cape and North West provinces had the least number of cases reported (0.5%).



EC, Eastern Cape; FS, Free State; KZN, Kwa Zulu Natal; LP, Limpopo Province; MP, Mpumalanga; NC, Northern Cape; NW, North West; WC, Western Cape; GA, Gauteng

Fig.15. Distribution of HCV PCR positive cases by province between 2004 and 2008, South Africa

The hospitals which had the highest number of HCV seropositive cases were Chris Hani Baragwanath and Johannesburg Hospital (Charlotte Maxeke) in Gauteng Province and Groote Schuur Hospital in Cape Town, Western Cape Province for 2007-2008. These are busy academic centres and specialist clinics where one would expect referrals.

2.4. DISCUSSION

Clearly, this study illustrates the need for confirmatory HCV RNA NAAT testing as recommended by CDC in order to confirm HCV active disease. Other studies on confirming HCV antibody tests have shown PCR or RIBA concordance to range from 52% (Gonzalez-Perez et al., 2004) to 79% (Kleinman et al., 2006). The PCR is our choice of a NAAT method over the RIBA because the PCR confirms active infection and the amplicon from the qualitative PCR is used for a genotyping test (explained in chapter 4). The PPV of ELISA is low in areas of low HCV prevalence and hence a high proportion of HCV-RNA negative results were expected in this study. Other reasons for a positive antibody test and negative RNA test are: transient clearance of virus in an acute infection or, more commonly, recovery from infection (Ghany et al., 2009). There are soluble factors in serum which cause interference, inhibiting the Amplicor Qualitative assay. These specimens must be retested and/or a separate/new specimen requested. If the PCR is inhibited, then an alternative test, such as, the HCV core antigen test, can be used to test for viraemia, although with a lower sensitivity. This study provided the basis for the recommended algorithm for HCV testing in SA (Fig.6, page 42). Without supplemental testing, results are unreliable. This needs to be considered when comparing older "serology only" studies with those using newer NAAT based technology. "True" prevalence rates, as indicated by PCR results, still need to be established in South Africa.

The national surveillance database established for the storage and collation of HCV serology and PCR data and to which genotyping data will be added collects data from NHLS laboratories countrywide and thus reflects HCV positive results in the

public sector. The seropositive figures for years 2007 and 2008 were higher (1371 and 1745, respectively, than the PCR positive figures for the same years (195 and 191 respectively), re-iterating the recommendation that antibody results be confirmed by NAAT.

Nationally, the total number of HCV PCR positives in the public sector, was 621 over the 5 year period from 2004 -2008. HCV PCR prevalence does not appear to have increased significantly over the years but clinicians are becoming more aware of the availability of tests for HCV diagnosis at NHLS laboratories. Awareness of HCV tests is spreading to outlying areas in SA as the number of requests received from public hospitals in Mpumalanga and Limpopo increased between 2007 and 2008. HCV tests are also available at private laboratories and, after consultation with NICD, HCV surveillance and reporting has now also been established for the private health care system. These initiatives, together with the imminent circulation of the NDoH guidelines for the control and treatment of HCV in SA should increase the number of requests for PCR to confirm seropositive results and increase the care and management of HCV infected individuals. The establishment of the databases will facilitate better notification of the disease and allow for improved follow-up.

Limitations of this study include that these data did not include data from private laboratories. Insufficient clinical and patient demographic information in 30% of specimens (eg, ethnicity, place of residence, place of origin) including details of possible transmission risk was not provided. The NAAT confirmation study was limited to specimens collected from local Johannesburg hospitals because of easy transport and follow-up. All molecular tests were conducted on stored frozen specimens and the extent of the study was confined by volume and condition of the specimens.

The viability of the specimen can be examined by testing the specimen for PCR amplification. As these were retrospective specimens, we infer that amplification did not occur in some samples because the virus had lost its viability due to poor storage and many freeze/thaw cycles.

The continuing collation of HCV PCR surveillance data will provide the national authorities with information on the current state of HCV infections in SA. Dissemination of the information should increase national awareness and improve the rate of notification of positive cases. Attempts to reduce the level of missing demographics include the distribution of questionnaires as part of routine surveillance. To guarantee return of questionnaires, these need to be generated at the hospital/clinic site. However, additional challenges on the quality of data collection in our South African situation needs to be improved. This may include: (i) availability of clinical staff/ translator/computer to collect data, (ii) quality checks on the data, together with storage and location, (iii) electronic versus manual entries, (iv) telephonic/email/online surveys not practical in SA, (v) high number of patients lost to follow-up.

In South Africa and elsewhere, (Verbeeck et al., 2008a) very little is known of the possible risk factors for HCV. Presumably, as with western countries, intravenous drug use (predominant in young adults; EHRN, 2007), haemodialysis, sexual or intrafamilial transmission (Verbeeck et al., 2008a), tattooing, piercing and other traditional markings may be ways of acquiring HCV in South Africa. However, education in the form of surveillance reports, HCV awareness campaigns, information pamphlets and questionnaires will help to inform the public of transmission risks and encourage change of lifestyle habits and improve prevention and treatment options.

HCV is a disease that can be cured if patients respond to therapy. Patients with genotypes 2 and 3 have a better chance of clearing the infection on combination therapy (section 1.10 , chapter 1). We recommend that genotyping and hence, early treatment be provided for all HCV PCR positive patients to reduce transmission and the burden of hospital costs. Ongoing genotype and regular viral load surveillance will clarify the level of response to therapy in patients with less-studied genotypes in SA, for example, genotypes 4 and 5. Genotypes will be added to this national database. This will enable monitoring of the changing frequency of genotypes in the country over time and across sentinel groups. Various cohorts of further high-risk study groups, for example, prisoners and HIV positive individuals will be sourced, tested and added to the database. It seems important that HCV-PCR is performed on all HIV+ samples, since antibodies can be absent, as indicated in the recommended algorithm for HCV testing in SA (Fig.6, page 42). The national HCV surveillance system will report regularly to the NDoH, providing baseline data from new HCV positive cases in the different sentinel groups. The positive outcomes of this system include:

(1) serial data which can be used to measure the epidemiological impact of the disease in the country

(2) contributions (specific for our genotypes and population groups) to the National Guidelines for the Prevention and Control of Hepatitis C Virus in South Africa

(3) recommendations to all health authorities on the notification of HCV disease

(4) recommendations for drug subsidies enabling early treatment (estimated costs in Canada from time of diagnosis is $39000/per patient, with 30% of costs related to liver disease, Nguyen et al., 2008)

(5) increased awareness and education to the public

(6) provide information to reduce transmission and for early cure before the development of cirrhosis and HCC

2.5. CONCLUSION

We have shown that PCR is a reliable test for confirming a positive HCV antibody result. In future, all seropositive HCV tests should be confirmed as recommended in the National Guidelines for the Prevention and Control of Hepatitis C Virus in South Africa (final draft, 2010). Like other surveillance systems in England, Scotland and the US, we have established a national database recording the results of HCV serology and PCR positive cases in SA. Information is being used to make national health authorities aware of HCV prevalence, plan preventative and treatment strategies, increase public awareness/education, reporting (notification) and finally to be better able to calculate the burden of HCV disease in SA. The availability of routine, reliable and validated methods of HCV detection, together with better surveillance strategies and concomitant collection of patient demographic data and database management, is vital to ensure patient treatment and disease management in South Africa.

CHAPTER THREE

Hepatitis C Virus genotype surveillance in three SA sentinel study groups

3.1. INTRODUCTION

Genotyping of HCV is the first line in determining a beneficial treatment regime, but it also provides information on global and local patterns of strain epidemiology and surveillance. There have been limited HCV genotyping studies in South Africa. The first of these used 5'UTR RFLP to determine genotypes of patients with chronic renal failure and liver disease in Cape Town and blood donor groups from KZN, JHB and Western Cape (Smuts and Kannemeyer, 1995). Overall, genotype 5 (51/130; 39%) was found to be the predominant genotype in all of the blood donor and patient study groups. This was followed by genotype 1 (43/130; 33%), genotype 2 (18/130; 13%), genotype 3 (10/130; 7%) and genotype 4 (3/130; 2%). Five (3.8%) specimens could not be typed. A study on HCV-seropositive patients with liver disease at Ga-Rankuwa Hospital, northwest of Pretoria, South Africa also used the 5'UTR RFLP method (Sithebe et al., 1996). The predominant genotype in this study group was genotype 2 (9/26; 35%), followed by genotype 5 (8/26; 31%) and genotype 1 (2/26; 8%). A high percentage, 26.9% (7/26), of this small group was reported untypable. Although only two of the seven HCV-infected HCWs in the Vardas et al. (2002) study were typed, they too were found by RFLP analysis to be infected with genotype 5.

Only one study in SA (Smuts and Kannemeyer, 1995) determined genotypes in the blood donor group. The volunteer blood donors are a group at low risk of blood-

borne infections for example, HCV (Vermeulen et al., 2009). South African blood units were first screened in pools of 8 in November 1992 by anti-HCV testing but the South African National Blood Service (SANBS) policy changed thereafter to provide individual unit testing with NAAT (Chiron Procleix Tigris system) in October 2005 (Vermeulen et al., 2009). Since 2005, there have been no new cases of transmission of HCV via blood products from SANBS.

The aim of this study was to determine the prevalence of HCV genotypes in three study groups, two hospital-based groups (haemophiliacs (H) and patients diagnosed with liver disease, (LD) during two time periods, 2000–2002 and 2007) and a volunteer blood donor group (BD during 2006-2008).

3.1.1. Objectives:

(a) To genotype HCV in the two hospital-based patient groups over two time periods (2000-2002 and 2007, respectively) and a volunteer blood donor group (2006-2008) by 5'UTR sequencing-based technique (SBT)

(b) To compare and validate the sequencing results from the 5'UTR to those from the NS5B region in available specimens from same patient groups above

3.2. METHODS

3.2.1. Hospital-based patient specimens

The study was retrospective and approved by the ethics committee of the University of the Witwatersrand, Johannesburg, South Africa (WITS HREC M051114, see appendix), and was therefore performed in accordance with the ethical standards of

the 1964 Declaration of Helsinki. A total of 308 specimens were collected from a local Johannesburg hospital, Charlotte Maxeke Johannesburg Academic Hospital. Sera were numbered in our laboratory and stored at -70 ^{0}C. All molecular tests were conducted on stored frozen specimens and the extent of the study was limited by volume and condition of the specimens. Fifty-three specimens were excluded due to volume and condition. Since patient specimens were tested for HCV-specific antibody at external diagnostic laboratories in the region, the storage and transport conditions of the specimens, prior to receipt in the laboratory, have not been rigorously documented. Data on clinical presentation, diagnosis and demographics for each patient was limited to the availability of data captured from request forms. Specimens with incomplete demographic data (N = 70) were excluded from the study (Strobe statement; Vandenbroucke et al., 2007). Patients with chronic liver disease (CLD) were divided into two groups based on time of attendance at the liver clinic. The first cohort of HCV antibody-positive specimens was collected during the period 2000–2002 (LD1), and the second 5–7 years later in 2007 (LD2). Haemophiliacs (H) were grouped together as a single cohort. The number of PCR-confirmed HCV cases in the laboratory determined the sample size of each group.

One hundred and eighty-five hospital-based specimens in total [45 H, 67 LD1 and 73 LD2] were sequenced in the 5'UTR region. Specimens collected during the earlier 2000–2002 period (H = 15, LD1 = 67) were additionally subtyped in the NS5B region. At the beginning of the study, we validated our choice of the two regions for phylogenetic analysis by constructing histogram plots of the frequency versus the distribution of relevant ranges of the pairwise distances between specimens of the 2000–2002 study group together with representative GenBank specimens.

3.2.2. Specimens from volunteer blood donors (BD 2006-2008)

Specimens were obtained from SANBS during the years 2006 [N=20] and 2008 [N=155], from anonymized BD. The specimens were collected from all provinces in the country and tested for the presence of HCV antibody at SANBS. Each specimen received a unique laboratory number and once received at NICD, the sera were stored at -70^{0}C until required for viral load and genotyping tests. Eighteen of the twenty specimens in the 2006 collection and 100 of the 155 specimens collected in 2008 were PCR positive. A total of 118 specimens from the blood donor group (2006-2008) was sequenced in the 5'UTR. The available demographics (age, gender and ethnicity) for the group were obtained from SANBS.

3.2.3. HCV RNA detection and quantitation

HCV RNA was detected and quantified using standardized methods of qualitative (Cobas Amplicor HCV version 2.0, Roche Diagnostics, Germany) and automated quantitative RT-PCR (Cobas Amplicor HCM Monitor version 2.0, Roche Diagnostics, Germany). Version 2.0 has been validated for accurate quantitation of HCV RNA without genotype bias (Lee et al., 2000). The lower limits of detection of the qualitative and quantitative assays are 50 and 600 international units per ml (IU/ml), respectively. HCV RNA quantitation for the blood donor group (2006-2008) was performed with the Cobas Ampliprep/Cobas TaqMan (Roche Diagnostics, Germany). The lower limit of detection for this assay is 15 IU/ml.

3.2.4. Specimen extraction

RNA was isolated from patient sera using the Magna Pure Total Nucleic Acid Extraction kit (Roche Diagnostics, Germany) and the method specified by the manu-

facturer. In brief, 200 µl of serum or plasma sample was added in the appropriate well of the sample cartridge. The proteinase K was reconstituted by adding 5 ml of elution buffer to the vial. The lyophilizate was dissolved by mixing and the appropriate volume added into the reagent tub. The appropriate wash buffers (for the removal of PCR inhibitors, salts and proteins), the lysis buffer (cell lysis and binding of total nuclei acid) and the elution buffer (for nucleic acid elution) were aliquoted into the reagent tub, according to the set-up protocol on the Magna Pure LC 2.0 instrument (Roche Diagnostics, Germany). The magnetic glass beads were added last to prevent sedimentation. The viral nucleic acids in the sample/control were bound to the bead surfaces and unbound particles were washed off during the several wash steps. The purified total viral nucleic acids were eluted in 50 µl of the low-salt elution buffer and stored at -70^{0}C for genotyping.

3.2.5. Reverse transcription (RT) and the nested PCR reaction

RT and first-round PCR were performed using the TITAN One Tube RT-PCR system (Roche Diagnostics, Germany). The 50 µl reaction mix contained 5 µl of purified RNA, 400 nM of each primer, 200 µM of each dNTP, 10 units of ribonuclease inhibitor (Roche Diagnostics, Germany), 5mM dithiothreitol (DTT), 1 µl enzyme mix (AMV, Taq and Tgo DNA polymerase) and a 1x concentration of RT-PCR buffer (with 1.5 mM $MgCl_2$). After initial reverse transcription, first-round PCR was carried out using 35 cycles of PCR, 9700 Applied Biosystems (ABI) thermocycler, USA (94^{0}C for 30 s, 55^{0}C for 30 s, 68^{0}C for 45 s) and a final elongation step at 68^{0}C for 7 min. Nested PCR reactions were carried out in a final volume of 100 µl with 5 µl of the first-round product, 400 nM of each primer, 200 µM of each dNTP, 2.5 U Taq polymerase and 1x concentration of PCR reaction buffer. The nested reaction was subjected to 35

cycles of PCR (94^{0}C for 60 s, 50^{0}C for 60 s, 72^{0}C for 60 s) and a final extension step at 72^{0}C for 5 min. The amplified PCR products were fractionated by electrophoresis through a 2% metaphore agarose gel (Cambrex, Rockland, USA) and visualized by staining with ethidium bromide under ultraviolet light. The annealing temperature was adjusted according to the Tm values of the primers. Primer and amplicon details, numbered according to (Choo et al., 1991) are shown in Table 8.

3.2.6. Sequencing

Sequencing of the 5'UTR and the NS5B regions was performed in both directions using the ABI Sequencing Ready Reaction kit (Applied BioSystem, USA) and cycle sequencing conditions as specified by the manufacturer. The amplicons were purified from excess dye terminators using ethanol precipitation with sodium acetate at a final concentration of 120 mM. Automated sequencing was performed using the ABI PRISM 3100 Genetic Analyzer using the nested primers as indicated in Table 8. A panel of specimens of known genotypes, received from the external quality assurance program (Quality Control for Molecular Diagnostics, QCMD), was used to validate our sequencing method.

Table 8. List of primers, primer sequences and combinations and amplicon sizes for the method used. The numbering is according to Choo et al., 1991.

Region	Primer	Sequence	Tm	G:C	Ampli-con size	Method
5'UTR	Outer Sn[1] (-297-277)	5'-CTGTGAGGAACTACTGTCTT-3'	55.2	45.0	269	RT
	Outer Asn[1] (-45-28)	5'- GCACTCGCAAGCACCCTA-3'	53.0	61.0		
	Inner Sn[1] (-278-258)	5'- TTCACGCAGAAAGCGTCTAG-3'	55.6	50.0	238	Nest/ seq
	Inner Asn[1] (-60-42)	5'- CCTATCAGGCAGTACCACAA-3'	55.6	50.0		
NS5B	Outer Sn[1] (7904 -7922)	5'- TGGGGTTCTCGTATGATACCC-3'	54	52.3	397	RT
	Outer Asn[1] (8300-8282)	5'- GAGTACCTGGTCATAGCCT-3'	50	55.0		
	or					
NS5B	Outer Asn[2] (8274-8244)	5'-GGCGGAATTCCTGGTCATAGCCTCCGTGAA-3'	66	57.0	370	RT
	Inner Sn[3] (7916-7935)	5'- GATACCCGCTGCTTTGACTC-3'	54	55.0	367	Nest/ seq
	Inner Asn[3] (8254-8266)	5'- CCTCCGTGAAGGCTCTCA-3'	53	61.1		
	or					
NS5B	Inner Asn[3] (8265-8247)	5'- GCTAGTCATAGCCTCCGT-3'	50	56.0	349	Nest/ seq

[1]Chan et al., 1992; [2]Sandres-Saune et al., 2003; [3]Chen and Weck, 2002
sense, Sn, anti-sense, Asn, RT, reverse transcriptase; nest, nested PCR, and seq, sequencing PCR

3.2.7. Data analysis

The statistical analysis software programme Statistica (Statistica Advanced, StatSoft, USA, version 9.1) was used to calculate whether the two liver disease co-horts differed significantly from each other with regard to age, sex and viral load. Fisher's exact and Chi-square tests were used for variables with normal curves, and the Mann– Whitney test, Kruskal-Wallis analysis for non-normality. A p-value of <0.05 was regarded as statistically significant. Multiple alignments of the sequenced regions were done with CLUSTALW, supplied within the MEGA software package, version 3.0; (Kumar et al., 2004), and phylogenetic analyses, using neighbour-

joining and the Kimura-2 parameters (Kimura, 1980), were conducted using MEGA. A thousand bootstrap replicates were used, and a bootstrap value of >70% was taken to define a stable phylogenetic cluster (Muerhoff et al., 1997). Trees were created with Tree Explorer in the MEGA programme. All data spreadsheets and graphs were generated with MS Excel 2003.

3.3. RESULTS

The demographics for the hospital-based study groups (Haemophiliac, Liver disease 1 and Liver disease 2) during the respective time periods are summarised and compared in Table 9a and b and that of the blood donor group (2006-2008) in Table 9c. The numbers of patients with liver disease and infected with HCV that visited the Liver Clinic during the two time periods remained constant, with no significant increase. There were more males in both the liver disease and blood donor (2006-2008) groups than females. The percentages of males versus female in the LD1 and LD2 groups, respectively were compared and a statistical significant power was inferred using the Mann-Whitney test. Patients with liver disease were older (52 years, Table 9b) than those with haemophilia (34 years, Table 9a) and the blood donor (2006-2008) group (39 years, Table 9c). Patient records for the liver disease groups in the two time frames were insufficient to allow conclusions with respect to the possible risk factors for HCV in these groups. The rounded geometric mean absolute viral load titres were slightly higher in the haemophiliac group (1260317 +/- 1669843 IU/ml, Table 9a) versus the liver disease and blood donor (2006-2008) group (807609 +/- 2361790 and 541006 +/-527306 IU/ml, respectively, Table 9b, c). There was an increase in the number of haemophiliacs tested at the clinic between the two time frames studied, which may reflect active recruitment and/or an increase in the

clinic screening activity over this period. The numbers of specimens for the blood donor (2006-2008) group as indicated in Table 9c are numbers received in the laboratory during that time period and may not necessarily reflect date of donation. Almost half of the blood donors positive for HCV were male and Caucasian (Table 9c).

Table 9. Demographic information for the patient and blood donor groups (a) Haemophiliac (H); (b) patients with liver disease, LD, (LD1 and LD2) and (c) blood donor (BD) 2006-2009

(a)	**Haemophiliac (H) N=45**	**2000-2002 N=15**	**2007 N=30**	
Age (years)[a]	34 ± 12	28 ± 11	37±12	
Geometric Mean Viral Load (IU/ml)[a]	1260317 ± 1669843	992119 ± 1648628	1394417 ± 1699113	
(b)	**Liver Disease (LD) N=140**	**LD1 N=67**	**LD2 N=73**	**p-value[c]**
Male	69	37 (55)[b]	32 (43)	0.321[d]
Female	48	19 (28)	29 (39)	
Unknown	23	11 (16)	12 (18)	
Age (years)	52 ± 14	51 ± 16	53 ± 13	0.862[e]
Geometric Mean Viral Load (IU/ml)	807609 ± 2361790	503313 ± 599646	1031174 ± 3062504	0.444[e]

[a] Mean ± std deviation (SD), [b] percentages N for each time period, [c] *p*-values as calculated by t-test, [d] Mann-Whitney test, [e] Chi-square test

Table 9c. Demographic information for blood donors (2006-2008), with available ethnic information

(c) blood donors (2006-2008)	**N=118**
Male	59 (50)[b]
Female	38 (32)
Unknown	21 (18)
Age (years) [a]	39 ±13
Ethnic group:	
Indian	4
Black	21
Coloured	2
Caucasian	52
Unknown	39
Mean Viral Load (IU/ml) [a]	541006 ±527306

[a] Mean ± SD, [b] percentages N for each time period

Table 10 indicates genotyping results of the three study groups as determined by sequencing in the 5'UTR region using sequence-based technique (SBT) only. For each genotype, the % of genotypes among the three groups (H, LD and BD) were compared for statistical significance, using the Kruskal-Wallis ANOVA test. Genotype 5 was dominant in patients with liver disease over the two time periods studied (Table 10; 67.16 and 57.58%, respectively, p = 0.27, average 62.14%) while genotype 1 predominated in the individuals with haemophilia (51.12%; Table 10). Genotypes 1 (32.21%) and 5a (38.98%; Table 10) had a similar prevalence in the blood donor (2006-2008) group. Genotype 5 was present in 33.33% of the people with haemophilia followed, in decreasing prevalence by genotype 3 (8.89%), genotype 2 (4.44%) and genotype 4 (2.22%). In contrast, the prevalence of genotype 1 was 22.87% in patients with liver disease, followed by genotype 3 (8.57%) and genotype 4 (6.42%). Genotype 3 was found to be highly prevalent in the blood donor group (2006-2008; 20.34%), compared to the other two study groups (Table 10). Genotype 2 was not detected in patients with liver disease (Table 10) and only one specimen was positive for this genotype in the blood donor (2006-2008) group. There was an increase in the prevalence of genotype 3 (from 5.97 to 10.96%) and 4 (from 4.48 to 8.22%) in the liver disease 1 and liver disease 2 groups (Table 10), but these differences were not statistically significant ($p>0.05$). Genotype 6 was not found in this or any of the previous South African studies.

Table 10. HCV genotypes in the three different study groups, people with haemophilia, liver disease and volunteer blood donors (2006-2008), based on 5'UTR sequence analysis

HCV geno-type	H		LD			BD(2006-2008)		p-value[b]
	N=45	Total (%)	LD1 2000-2002 N=67	LD2 2007 N=73	Total (%)	BD N=118	Total (%)	
1	4 (8.89)[a]	51.12	5 (7.46)	4 (5.48)	22.87	18 (15.25)	32.21	0.36
1a	9 (20.00)		9 (13.44)	6 (8.22)		8 (6.79)		
1b	10 (22.22)		1 (1.49)	7 (9.59)		12 (10.17)		
2	1 (2.22)	4.44	0	0	0	1 (0.84)	0.84	0.447
2b	1 (2.22)		0	0		0		
3	0	8.89	1 (1.49)	4 (5.48)	8.57	3 (2.54)	20.34	0.349
3a	4 (8.89)		3 (4.48)	4 (5.48)		21 (17.80)		
4	1 (2.22)	2.22	3 (4.48)	6 (8.22)	6.42	8 (6.79)	7.63	0.22
4a	0		0			1 (0.84)		
5a	15 (33.33)	33.33	45 (67.16)	42 (57.58)	62.14	46 (38.98)	38.98	0.27

[a] percentages *N*
[b] *p*-values as calculated by the Kruskal-Wallis ANOVA test

Ninety-one sample sequences, together with representative GenBank sequences, in bold type with their known genotypes/subtypes, in parenthesis after their accession numbers (Fig. 16), were aligned and analyzed over 198 bp of the amplified 5'UTR region. The alignment of nucleotides in the 5'UTR region (not shown) confirmed that a double adenine ($A^{106}A^{107}$) defined specimens of genotype 5 (Smuts and Kanne-meyer, 1995). However, a group of our specimens and one GenBank specimen, U33430, lacked A106 but were otherwise identical to other genotype 5 specimens over the 198-bp region studied. These clustered together, away from other genotype 5 specimens (see clade labelled ''Atypical Genotype 5a''; Fig. 16). For clarity, and because intraclade genetic distances are small, only 91 of the 303 5'UTR se-

quences were included in the final tree (Fig. 16). Specimens 5126 and 5633, from the 2007 haemophiliac group, partitioned significantly, and away from all other specimens, with genotype 2 in the 5'UTR region. However, specimen 5126 formed an outlier away from 5633 and known subtype 2a, 2b and 2c reference sequences (Fig. 16). Specimens 1182 and 1909 clustered in a clade that had both genotype 4a and 4c reference sequences, 2575 and 4303 grouped with a subtype-unknown reference genotype 4 sequence (L29600), and four other specimens from patients with liver disease, 1678, 2045, 1158 and 2406, clustered with a reference sequence, L29619, from subtype 4g. Blood donor specimens 9389 and 9402 clustered within genotype 4 by phylogenetic analyses (Fig. 17). The 5'UTR is too conserved to partition genotype 4 into significant subtype clusters and this can also be seen for genotype 1 (Fig. 16). In the 5'UTR region, specimens 9297 and 9298 grouped as outliers of the genotype 1 cluster and specimen 9411 was an outlier in the genotype 4 cluster (Fig.16).

NS5B sequencing was performed to confirm the 5'UTR segregation pattern (Fig. 18; 41 of the 61 specimens that could be amplified in both regions are shown). Subtyping within genotypes was now also possible due to the greater variability of this region. There was a 71% (61/86) success rate for the amplification of products in the NS5B region using published methods and primers. Two specimens (1909 of the haemophiliac cohort and 1182 of the liver disease 1 cohort) that had clustered with the clade of genotype 4a and 4c in the 5'UTR region, clustered significantly with the genotype 4c reference sequence, L29602, in the NS5B region (Fig. 18). The branching of specimen 1678 of the LD1 cohort with L29618, a genotype 4g reference, in the 5'UTR region was confirmed in the NS5B analysis (Fig. 18). The speci-

mens previously typed genotype 5 in the 5'UTR region formed a significant clade with the genotype/subtype 5a reference sequences in the NS5B regions together with recently published genotype 5 NS5B sequences from South Africa, Belgium and France (Belgian and French specimens labelled by country in Fig. 18). There was, however, no distinction between the ''typical'' and ''atypical'' clusters of genotype 5 in this region. Specimens that were found in the ''atypical'' cluster in the 5'UTR region grouped as genotype 5a in the NS5B region (3788/LD1; 3494/LD1; 363/LD1). This indicates that the specimens in the ''atypical'' cluster are not recombinants or belong to a different subtype. Specimen 1485, which grouped with the subtype 1b reference in the 5'UTR analysis (Fig. 16), clusters significantly with GenBank reference M62321, which is known to be subtype 1a, in the NS5B tree (Fig. 18). Apart from this subtype switch, all specimens partitioned into the same genotypic clusters in both regions. All genotypic clusters on the NS5B tree had significant bootstrap values. Except for the 2 haemophiliac genotype 2 specimens, there was no significant monophyletic clustering of sequences according to the haemophiliac group or patients with liver disease in either of the regions sequenced.

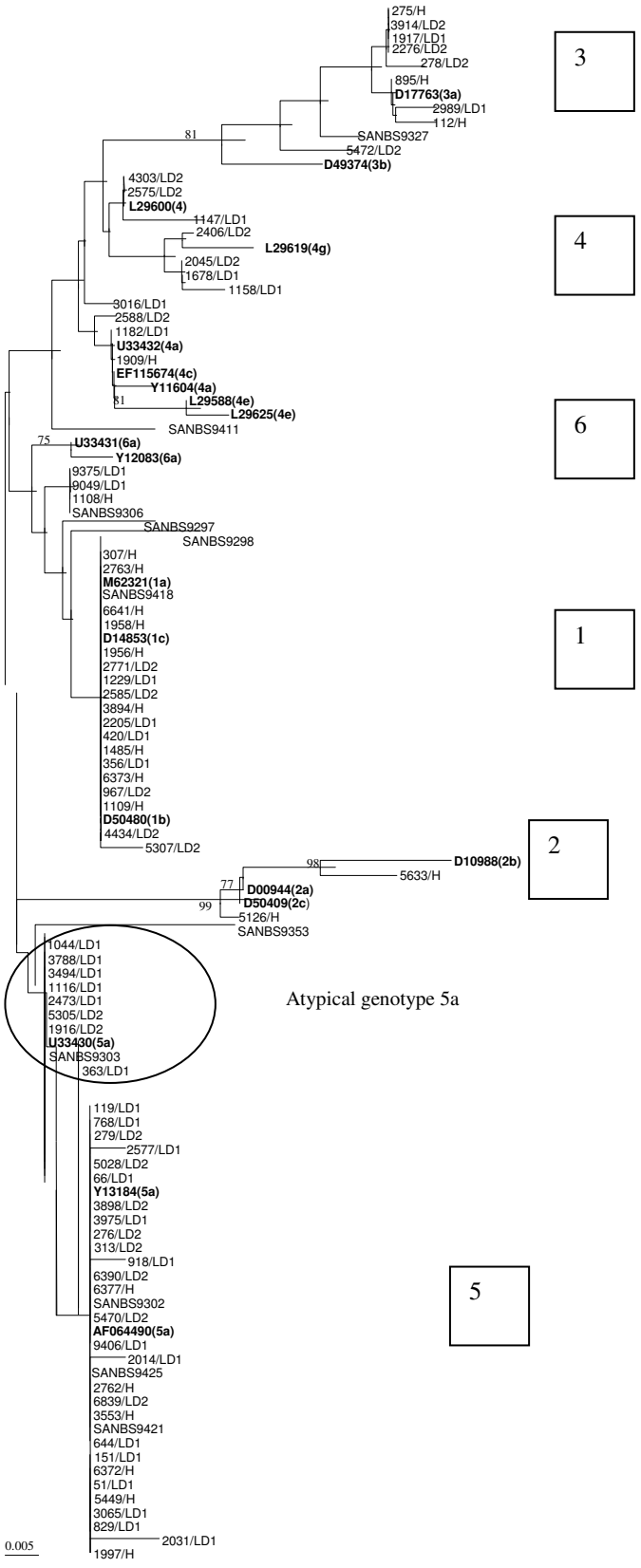


Fig.16. Unrooted neighbour-joining tree of 91 HCV 5'UTR fragments of the H, LD1 and LD2 groups. Significant bootstrap percentages from the consensus tree (1000 sets) are shown. GenBank accession numbers are in bold print with subtypes in parentheses. Genotypes are labelled according to the nomenclature of Simmonds et al., 1994a. The scale bar represents 0.005 substitutions/nucleotide site.

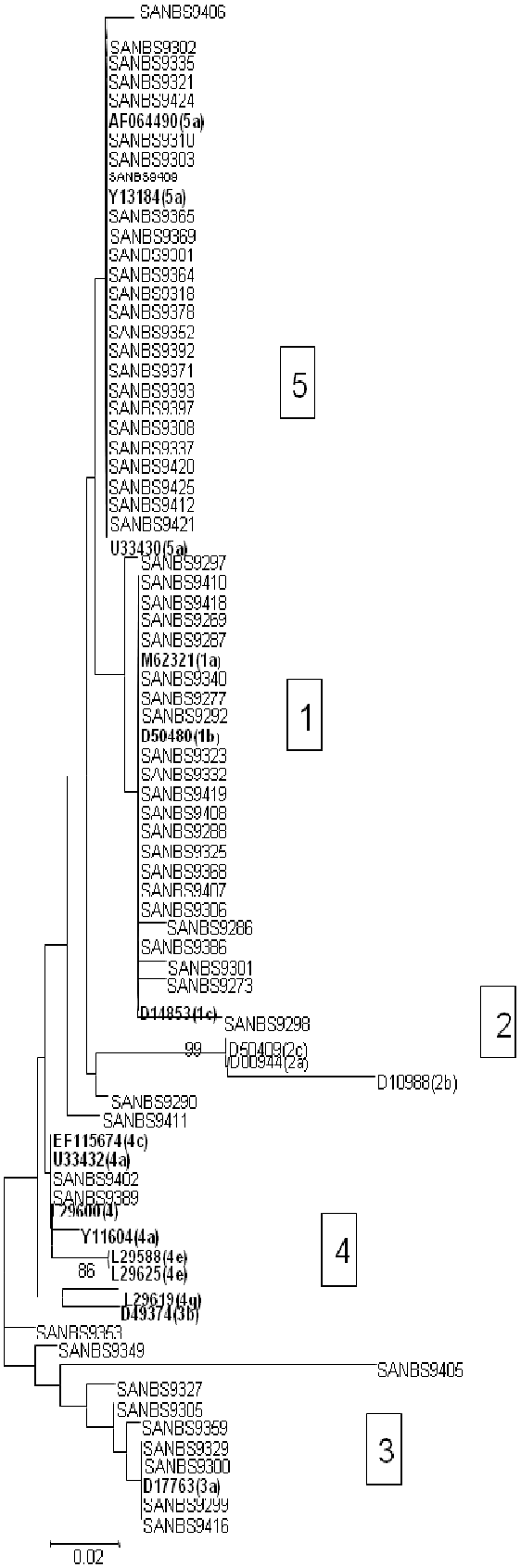


Fig.17. Unrooted neighbour-joining phylogenetic analyses of 61 5'UTR sequences in the volunteer blood donor (2006-2008) cohort. Significant bootstrap percentages from the consensus tree (1000 sets) are shown. GenBank accession numbers (in bold print) and genotypes are labelled according to the nomenclature of Simmonds et al. 1993b.The scale bar represents 0.02 substitutions/nucleotide site.

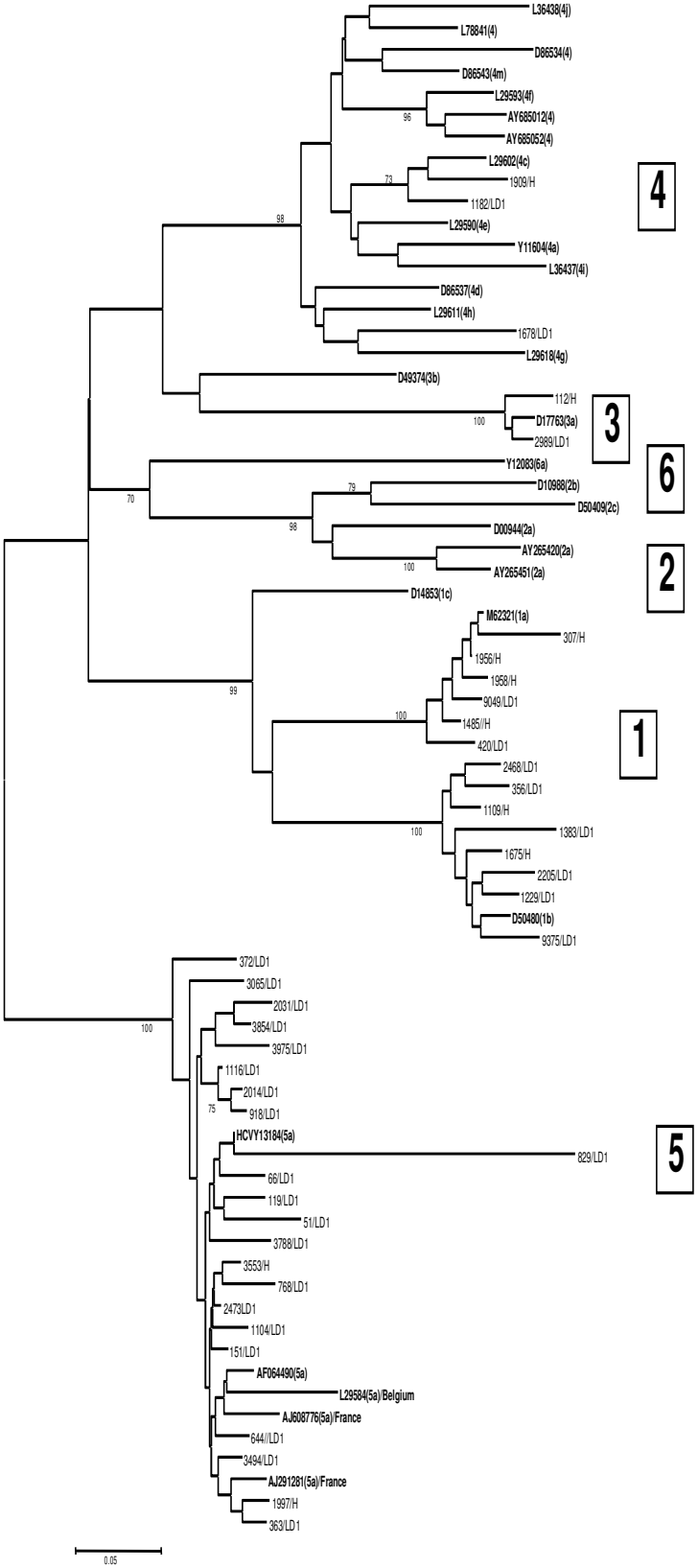


Fig.18. Unrooted neighbour-joining tree of 41 HCV isolates of the haemophilia, liver disease 1 and 2 groups. The region sequenced is 202bp in the NS5B region and clusters are shown of different genotypes and subtypes. Significant bootstrap values are indicated. GenBank accession numbers are in bold print with subtypes in parentheses. Genotypes are labelled according to the nomenclature of Simmonds et al., 1993a.The scale bar represents 0.05 substitutions/nucleotide site.

3.4. DISCUSSION

Screening of blood products in South Africa began in the early nineties, and no further transmission of HCV has been observed in the haemophiliac clinic since then (Dr. Mahlangu, personal communication, Haemophiliac Clinic, Charlotte Maxeke Hospital, 2008). This is reflected by the lower age range in the haemophiliac group over the time periods studied. Asymptomatic haemophiliacs are screened at regular intervals because of routine exposure to blood products, whereas most of the patients presenting at the liver clinic would either be symptomatic (which occurs after long-term infection) or have had abnormal blood tests prompting further investigation. Since HCV disease is normally diagnosed at the late disease stage, it is not surprising that the average age of the patients with liver disease (early 50s) is older than that of the people with haemophilia and the volunteer blood donors. In the latter, the majority of the donations received at SANBS are from Caucasian (M. Vermeulen, SANBS; personal communication, 2008), male donors, around the age range of 41-47 years. There is no significant increase in patient numbers between the two liver disease groups, indicating that HCV prevalence may still be low in the Johannesburg urban areas. However, the numbers in these groups have to be interpreted with caution as this group of individuals were referred to the liver clinic. There was no significant difference in viral load between the haemophilia, liver disease and blood donor groups.

Although the mean viral loads for the specimens were high (Table 9a, b), our success rate in amplifying the NS5B region (71%) was not as high as that reported by others (Sandres-Saune et al., 2003; Tamalet et al., 2003). This study is retrospective and was performed on specimens that were stored for long periods and sub-

jected to frequent freeze/thaw cycles between tests. RNA degradation due to poor storage and poor maintenance of the cold chain during specimen transport may be plausible explanations and it is also possible that unknown variation (leading to suboptimal primer design in this less conserved region) could have contributed to this rate discrepancy, as published primers were used.

Sequencing and phylogenetic analysis in the 5'UTR was accurate in determining genotypes, but the greater variability of the NS5B coding region enabled differentiation of both genotypes and subtypes (Simmonds et al., 1993a). As previously described (Simmonds et al.,1993b), the 5'UTR region could distinguish genotypes but not subtypes as the histogram of genetic distances in the 5'UTR region (not shown) had two peaks, indicating this separation. However, the distribution of genetic distances in the NS5B region revealed three separate and non-overlapping partitions, confirming the ability of this region to discern subtypes (Simmonds et al.,1993a). Whereas 5'UTR RFLP does not readily differentiate between subtypes, this analysis has shown, for the first time in South Africa, the presence of subtypes 4c and 4g by sequencing (Fig.16). These genotype 4 subtypes were characterized previously among patients in Cameroon, the Central African Republic and Gabon (Stuyver et al., 1994).

The overall distribution of genotypes between the two time frames in the respective groups, patients with liver disease and individuals with haemophilia, remained the same in this study, with minor variations that were not statistically significant. However, with the blood donor (2006-2008) group, we have seen an increase (although nor statistically significant) in genotype 3 as compared to those with liver disease

and haemophilia (Table 10). Our genotyping results were compared to two previous South African studies (Fig.19), that used RFLP to determine genotypes in their respective cohorts. This study did not consistently corroborate previous SA studies and it would seem that genotype prevalence varies across the country and in the different cohorts studied in the Johannesburg area. This could reflect the facts but there are other possible explanations for this. Earlier studies report genotypes in smaller cohort sample size compared to ours (Fig.19). Earlier methods such as RFLP analysis depend on point mutations and are not as accurate as sequencing and this is reflected in the number of untypable specimens in these studies.

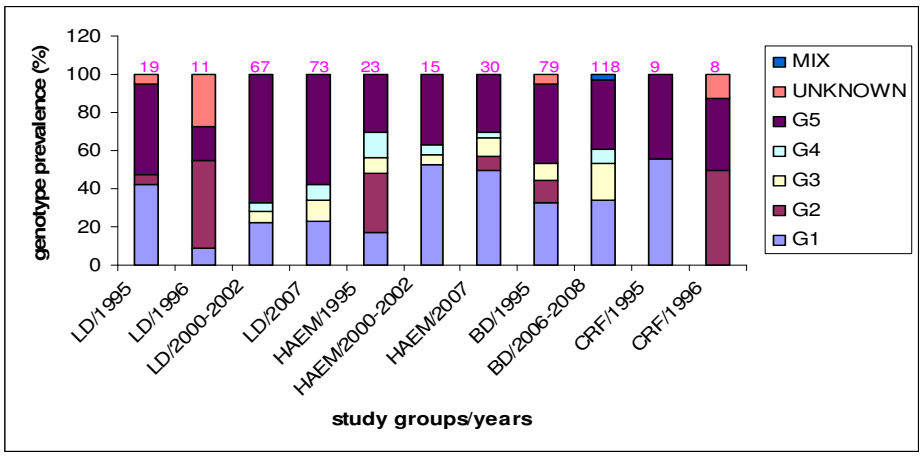


Fig.19. Genotypes determined in this study cohorts compared to two other South African studies, Smuts and Kannemeyer., 1995 (CRF, chronic renal failure; LD liver disease; haem, haemophiliac; BD, blood donor;): Sithebe et al., 1996 (CRF, Haem, LD). The numbers above bars indicate the size of the study population.

Genotype 5 was the most common genotype (Fig.19, table 10) in South Africa, with variable percentage prevalence seen among the different cohorts studied in this chapter and previously reported (LD2000-2002; LD2007; BD1995). Based on se-

quence data of the BD (2006-2008), genotype 1 is almost as abundant as genotype 5. Genotype 1 predominated in people with haemophilia (Fig.19). Although not much is known about the source of blood products or time of infection of people with haemophilia, it is known that imported factor concentrates from USA were administered to some individuals attending the clinic during the period 1982–83, but thereafter, fractionated product from local unpaid donors has been used exclusively (Dr. Mahlangu, personal communication, haemophiliac clinic, Charlotte Maxeke Hospital, 2008).

The prevalence of genotype 1 among blood donors remained constant over the 2006-2008 time periods and genotype 1 prevalence similar to that of Smuts and Kannemeyer, 1995 study. Relative to each other, the prevalence of genotypes 3 and 4 increased in the patients with liver disease over the 2 time periods, although this is not statistically significant. Neither genotype was present in previous studies of patients with liver disease (Fig.19). This increase in prevalence of genotypes 3 and 4 may be due to travel and immigration from India, other eastern countries and/or neighbouring African countries, where genotypes 3 and 4 predominate (Hissar et al., 2006; Pasquier et al., 2005). With the blood donor group, we have seen an increase in genotype 4 (9%, Fig.19). In the 1995 study of Smuts and Kannemeyer, genotype 4 was only seen in the haemophiliac population (3/23). We detected only two genotype 2 specimens, and both of these were in the haemophiliac group of 2007 (Table 10). This may be a consequence of the method used, population studied or the patient group examined, since Smuts and Kannemeyer, 1995 found genotype 2 in the blood donor group (10/79), haemophiliacs (7/23) and a patient with liver disease (1/19), and Sithebe et al., 1996 found genotype 2 to be the predomi-

nant genotype in patients with liver disease (5/11) and renal failure (4/8; caution should be taken when interpreting these results as patient numbers were low in that study). We detected one genotype 2 in the blood donor (2006-2008) group by SBT. There was also a change in prevalence of genotype 5 between the two liver disease groups (from 67 to 58%; Table 10). These effects may be explained by the variation in the prevalence of other genotypes, emigration, study design limitations, geographical region or technological improvements. It is possible that, in the approximately 5–12 years between our studies and previous South African studies, improvements in technology and methodology have enabled more accurate typing of genotypes, as both earlier studies reported a proportion of specimens that could not be typed.

This study has optimized and validated diagnostic methods for genotyping HCV in a national laboratory, identified circulating genotypes in patient groups and population-based groups from Johannesburg and surrounding areas. All of the sequences, except for the subtype variation of one specimen, displayed concordant results for both of the regions that were sequenced. The occurrence of a single genotype 5a clade in the NS5B region is in accordance with the results reported by Verbeeck et al., 2006. Although the Verbeeck study reported that the phylogenetic signal in the NS5B region is too conserved to provide information regarding subtypes within genotypes (Verbeeck et al., 2006), this study and others (Laperche et al., 2006) have found the region effective in differentiating the major subtypes.

Despite the advantages of using the more variable NS5B region for fine subtyping, genotype inference was identical between the two regions sequenced. In addition,

while the conserved 5'UTR region is less specific at the subtype level, it is more robust and sensitive for diagnostic genotyping, particularly in a setting where less characterized variants (e.g. genotype 4 and 5) are being detected and may be the predominant types. For these reasons, 5'UTR PCR amplification and sequencing was used for routine genotyping in our laboratory and was used as the gold standard to test and evaluate other techniques in later studies.

3.5. CONCLUSION

This study has described HCV genotypes in three sentinel groups; patients with liver disease, haemophilia and volunteer blood donors. In the context of South Africa, the major HCV genotype in patients with liver disease is still 5a, as evident in this study in previous studies. However, there is a difference in the distribution of HCV genotypes between the liver disease and haemophiliac/blood donor study groups and we have identified previously unknown subtypes in South Africa. Although, knowledge of HCV genotypes is now an essential part of treatment regimes and prognosis of infection, the addition of genotypes onto the surveillance databases, further helps to understand the changing frequencies of genotypes in the country and in different study cohorts. From such surveys, the identification of new genotypes, subtypes, recombinants and/or superinfections can be determined.

CHAPTER FOUR

The Advantages and disadvantages of commercial and experimental HCV Genotyping assays

4.1. INTRODUCTION

Accurate HCV genotyping methods are required for surveillance, molecular epidemiology, vaccine development and clinical management. HCV genotype is the most important predictor of treatment outcomes and prognosis of clinical disease. Until recently, no significant difference had been reported in response rates among the HCV subtypes (Wohnsland et al., 2007). However, with the development of protease and RNA-dependent-RNA polymerase (RdRp) enzyme inhibitors, different antiviral responses and resistance patterns in subtypes 1a and 1b have been observed (Erhardt et al., 2009; McCown et al., 2009) and HCV subtyping must be considered in order to optimise treatment regimes. Although, the gold standard of diagnostic genotyping is sequence-based technologies (SBT) on various regions of the genome, this method is labour intensive (Table 11), with many PCR steps and clean-up applications. SBT is relatively cheaper than a commercialized LiPA test, but SBT has longer turn-around times (TAT, Table 11) and subtypes of genotypes were not identified optimally in the 5'UTR SBT as indicated by the phylogenetic analyses of the two regions (5'UTR and the NS5B region) compared in chapter 3.

The 5'UTR remains the most popular target for diagnostic genotyping but cannot accurately subtype. Newer methods are therefore constantly explored. In this study, three alternate technologies were evaluated for HCV diagnostic genotyping, namely, (1) real-time PCR, (2) line probe assay using the 5'UTR and (3) line probe assay using the 5'UTR plus core.

Table 11. Costs, advantages and disadvantages between the LiPA and SBT genotyping tests

Summary	LiPA	Sequencing
Cost	Genotyping kit $2903(X40) Amplification kit $5500 (X40) COBAS Amplicor or Amplicor Manual kit $3428 (X96)	Titan kit (RT-one step) $758 (100 Rxns) Nested PCR $219(100 Rxns) Sequencing ABI kit $2224 (100 Rxns) PCR Clean-up $507(50 Rxns) Sequencing ABI clean-up $2857 (100 Rxns) Primers $86(100 Rxns)
Cost per specimen	$244 (if amplification for subtyping genotype 1 is needed) $107 genotyping only (excluding consumables, repeats, controls, failed runs, equipment and maintenance costs)	$ 79 (excluding consumables, repeats, controls, failed runs, equipment and maintenance costs)
Time per specimen	4 hours	2-3 days
Advantages	Quick and easy to use (see workflow in methods). High throughput. Can pick up mixed infections. Can use COBAS AMPLICOR manual amplicon for test. Can amplify core region for subtyping	Gold standard. Highly specific and sensitive (both strands, sense and anti-sense strands are sequenced) Can target 5'UTR, core, NS5B regions using specific gene primers.
Disadvantages	Relies on visual subjectivity. Interpretation requires experience. Can get the presence of too many bands or "ghost" bands which make it un-interpretable.	Too many steps, RT, nested PCR, sequencing PCR Misses mixed infections Requires skilled staff for phylogenetics and to interpret chromatograms Success rates depend on the design of primers, storage and quality of RNA

Several real-time PCR methods for genotyping HCV have been described (Schroter et al., 2002; Rolfe et al., 2005; Schutzbank et al., 2006). HCV RNA is amplified using sequence specific PCR primers as before but the detection of the PCR products is automated using fluorescent dyes and light cycler technology (Schroter et al., 2002). Real-time LightCycler genotyping method use Fluorescence Resonance Energy Transfer (FRET) technology. Energy is transferred from the donor probe with a fluorescein dye to an acceptor probe with another dye (eg. LCRed640). When they lie close to each other at the end of the annealing step of PCR, emitted light

(energy) is captured by the fluorimeter in the LightCycler and is proportional to the target concentration.

The presence of different amplicons is detected by observing the melting curve of the product. Melting temperatures, (Tm) the temperatures at which 50% of probe-target duplexes separate) vary dependent on the number of SNPs in the product (Roche Applied Science, http://www.roche-applied-science.com). This "melt off" is also proportional to the length and the GC content of the target sequence.

The line probe assay (LiPA) hybridizes biotinylated PCR amplicons to immobilized probes on nitrocellulose strips (Andonov and Chaudhary, 1995). Streptavidin-alkaline phosphatase reacts with the biotin group. Chromogenic compounds detect the enzyme conjugate. The hybridization pattern identifies not only the dominant genotype but also mixed genotypic infections (Table 11). The LiPA using the 5'UTR was able to discriminate the different genotypes 1-6 but was limited in subtyping, especially of genotype 1 and 6 (Stuyver et al., 1996). When the LiPA assay was improved by incorporating the amplified core region together with the 5'UTR (LiPA 5'UTR plus core), better subtyping of genotype 6 (Noppornpanth et al., 2006b) and 1 (Verbeeck et al., 2008b) was evident but no mixtures were included in their studies.

The aim of this study was to validate and compare these three technologies (real-time, LiPA (5'UTR) and LiPA, 5'UTR plus core) against the gold standard SBT.

4.1.1. Objectives:

Validation and assessment of:

(a) the real-time LightCycler Genotyping method and

(b) the LiPA (5'UTR) and the LiPA (5'UTR plus core) assays

4.2. METHODS

A QCMD (Quality control for molecular diagnostics) panel of specimens from an external quality assurance (EQA) program, for HCV genotyping was used as standards against which the results for each assay could be validated. Calculations for accuracy, negative predictive (NPV) and positive predictive values (PPV), sensitivity and specificity were performed using the NHLS standard operating procedure (NHLS/NICD SOP NIC0563, version 1, 2008). Known negative specimens in the EQA panel were included as negative controls. All specimens used in the study were previously genotyped by SBT (chapter 3) and genotypes were known.

4.2.1. Real-time LightCycler method

Twenty-eight specimens were typed using the LightCycler (Roche Diagnostics, Germany) and the method of Schroter et al., 2002. Three different pairs of hybridization probes were used with primers 27 and KY81. Briefly, 2 µl of a 10X concentrate of the LightCycler-DNA Master Hybridization probe mix (LightCycler-DNA Master Hybridization probe, Roche Diagnostics, Germany; with LightCycler Faststart Taq DNA polymerase, PCR buffer, dNTP mix and 10mM $MgCl_2$), 2 µl (5pM) of each primer, 2.4 µl (3mM) $MgCl_2$ and 0.6 µl (3pM) of each hybridization probe was added to a final genotyping PCR volume of 20 µl. Distilled water was used as the negative control. The position and sequences of the hybridization probes and primers are num-

bered according to Landt et al., 2002 and are shown in Table 12 (numbering in parentheses is according to Choo et al., 1991). The fluorescence signal in channel 2 (F2) was measured for LCRed640 emission and the signal in channel 3 (F3) was measured for LCRed705 emission. The genotypes were discriminated by the different melting temperatures (Tm's) of the amplicons.

4.2.2. LiPA (5'UTR) and LiPA (5'UTR plus core)

4.2.2.1. LiPA (5'UTR)

The denatured 5'UTR PCR amplicons (from the previously genotyped blood donor group 2006-2008, chapter 3) obtained from the HCV Amplicor (Cobas Amplicor HCV version 2.0, Roche Diagnostics, Germany) were used for genotyping with the VERSANT line probe assay (LiPA, version 2.0) according to the methods recommended by the manufacturer (Siemens Medical Diagnostics, Belgium). The LiPA (5'UTR) was optimised by using varying volumes (10-40 µl) of the denatured amplicon in the test. After optimization 30 µl of denatured product was added to the hybridization buffer and incubated with the LiPA nitrocellulose indicator strip. All incubations in the hybridization buffer and wash solution were performed at 50^0C and the conjugate and substrate steps at room temperature, as recommended by the manufacturer. The chromagen substrate reacts with the conjugate, forming a precipitate, which results in a visible banding pattern on the nitrocellulose paper. The results were interpreted per instructions on the reading card and interpretation chart, supplied in the assay. The Student t-test analysis was used to compare the LiPA (5'UTR) to SBT (Statistica, Statistica Advanced, StatSoft, USA, version 9.1). The performance of the LiPA (5'UTR) was also assessed using standardised QCMD

panels (2007 and 2008) with low copy number and mixed genotypes. The BD 2009 specimens were also tested by LiPA (5'UTR).

4.2.2.2. LiPA (5'UTR plus core)

Thirty-five previously typed specimens and 8 QCMD samples were re-typed using the LiPA (5'UTR) amplicons and the LiPA (5'UTR plus core) assays. The two specific regions of the HCV genome used in the extended analysis were co-amplified using the Versant HCV Amplification LiPA assay (Siemens Medical Diagnostics, Belgium) and the procedure was performed as recommended by the manufacturer. The viral RNA extraction was performed using the QIAamp Mini RNA assay (Qiagen, Germany) and 5 µl of this extract was added to the PCR mix. The amplification mix (AMP MIX) provided in the Versant HCV Amplification LiPA assay contains two pairs of biotinylated primers targeting the 5'UTR (240 bp amplicon) and core (270 bp amplicon) regions, respectively, together with dNTP/dUTP, $MgCl_2$, RNAsin and PCR buffer. Reverse transcription (RT) and PCR amplification are performed sequentially in a single tube for each specimen in parallel with appropriate negative and positive controls. The enzyme mix (ENZ MIX) contains Sensiscript and Omniscript RT enzymes and HotStarTaq polymerase. The final validated method, optimized and validated in the Specialized Molecular Diagnostics Unit (SMDU) laboratory, uses 2 µl of the ENZ MIX to every 13 µl of the AMP MIX and 40 cycles of the thermal profile recommended by the manufacturer. Following the PCR, 8 µl of the PCR product was loaded onto a 2% agarose gel to identify positive PCR reactions. Ten µl of the latter were then denatured using the denaturing solution provided in the Versant HCV Genotype LiPA, 2.0 kit (Siemens Medical Diagnostics, Belgium) and applied to the single-stranded oligonucleotide probes bound to the membrane. The latter include appropriate core region probes which allow for the identification of

subtype 1a (band 25), subtype 1b (band 26) or genotype 1 (presence of both bands 25 and 26).

Table 12. Sequence and positions of primers and probes used for the Real-Time LightCycler genotyping method. Numbering at positions according to Landt et al., 2002 and in parentheses, according to Choo et al., 1991.

Primers	**Sequences**	**Position**	**Tm (^{0}C)**
Sense27 (-319 to –298)	5'-TCCACCATGAATCACTCCC-3'	26-44	54.1
Antisense KY81 (-191 to –173)	5'-CGGAACCGGTGAGTACACC-3'	169-151	58.4
Probes	**Genotyping 1/2/4/5**		
Sensor 13 (-242 to –224)	5'-GTGTCGTGCAGCCTCCAGG-3'	100-118	62.4
Sensor 4 (-241 to –224)	5'-TGTTGTACAGCCTCCAGG-3'	101-118	52.8
Anchor 1(-223 to –200)	640-CCCCCCTCCCGGGAGAGCC	120-139	72.7
Probes	**Genotyping 1/3**		
Sensor 3a (-297 to –261)	5'-CACGCGGAAAGCGCCTA-3'	65-81	60.2
Anchor 3 (-259 to –238)	705-CCATGGCGTTAGTACGAGTGTC	83-104	58.2

The sense and antisense primers were used in the amplification step. The two sensor probes (sensor 13 and 4) and the anchor 1 probe were used in the detection of genotypes 1, 2, 4 and 5 and sensor 3a and anchor probe 3 were used to detect genotype 3.

4.3. RESULTS

4.3.1. Real-time LightCycler method

Forty-one QCMD panel specimens and 28 previously sequenced specimens were used for the validation and method comparison, respectively. The viral loads of the specimens ranged from 600 to 1990000 IU/ml. The results of the validation are as follows (Table 13): accuracy= 55.81%, NPV= 10.52%, PPV= 100%, sensitivity = 53.65% and specificity = 100%.

Table 13. Validation results on the Real-Time LightCycler Genotyping method as compared with positive QCMD results

QCMD Specimen	QCMD Genotyping results	LightCycler Genotyping results
1.	3a	3
2.	4	Not detected
3.	1a	1
4.	2b	2
5.	4	Not detected
6.	2b	2
7.	1	Not detected
8.	1	1
9.	1	1
10.	3a	Not detected
11.	1	1
12.	1	Not detected
13.	1	1
14.	1	1
15.	1	Not detected
16.	1	Not detected
17.	3a	3
18.	1	Not detected
19.	1	Not detected
20.	1	1
21.	2b	2
22.	1b	1
23.	4	4
24.	1a	1
25.	3a	3
26.	4	Not detected
27.	2b	Not detected
28.	3a	Not detected
29.	1b	Not detected
30.	5a	Not detected
31.	3a	3
32.	5a	5
33.	1b	Not detected
34.	3a	Not detected
35.	1b	Not detected
36.	3a	Not detected
37.	3a	3
38.	1b	1
39.	5a	Not detected
40.	5a	5
41.	1b	1

1.1 Accuracy = $\frac{\text{True positives + True negatives}}{\text{True positives + False negatives + True negatives + False positives}}$

$= \frac{22 + 2}{22 + 19 + 2 + 0}$

$= 55.81\%$

1.2. Negative Predictive Value (NPV) = $\frac{\text{True negatives}}{\text{True negatives + False negatives}}$

$= \frac{2}{2 + 19}$

$= 10.52\%$

1.3. Positive Predictive Value (PPV) = $\frac{\text{True positives}}{\text{True positives + False positives}}$

$= \frac{22}{22 + 0}$

$= 100\%$

1.4. Sensitivity = $\frac{\text{True positives}}{\text{True positives + false negatives}}$

$= \frac{22}{22 + 19}$

$= 53.65\%$

1.5. Specificity = $\frac{\text{True negatives}}{\text{True negatives + False positives}}$

$= \frac{2}{2 + 0}$

$= 100\%$

Twenty-three specimens (23/28, 82%) could be genotyped using the LightCycler genotyping method, but only eighteen (64%) of the results were concordant with the 5'UTR and NS5B sequencing results (Table 14).

Table 14. The LightCycler Genotyping results compared to sequencing of the 5'untranslated region. "AA" indicates the double adenine at position 106 and 107 or "A" at position 107 only ("atypical"), respectively.

Specimen	Sequencing genotyping results	LightCycler genotyping results
1.	1	1
2.	5a "AA"	5
3.	4	Not detected
4.	5 "A"	2
5.	5 "AA"	5
6.	5 "A"	2
7.	5 "A"	Not detected
8.	5 "A"	1
9.	5 "A"	5
10.	3a	3
11.	1a	1
12.	5a "AA"	1
13.	1a	1
14.	5a "AA"	5
15.	5a "AA"	5
16.	5a "AA"	5
17.	5a "AA"	5
18.	5a "AA"	Not detected
19.	5a "A"	Not detected
20.	5a "A"	Not detected
21.	5a "A"	2
22.	5a "AA"	5
23.	5a "AA"	5
24.	5a "AA"	5
25.	5a "AA"	5
26.	5a "AA"	5
27.	5a "AA"	5
28.	5a "AA"	5

Genotype 5 was the predominant genotype (61%), followed by genotype 1 (22%), genotype 2 (13%) and genotype 3 (4%). The five specimens that did not amplify by the LightCycler method were from genotype 4 (N=1) and genotype 5 (N=4, 3 atypical and 1 typical; Table 14). All three of the specimens typed as genotype 2, and two of the 5 typed as genotype 1 by the LightCycler method, had been found conclusively to cluster as genotype 5 by phylogenetic analyses in the 5'UTR and NS5B sequenced regions (chapter 3, Fig. 16 and 18, respectively). Fig. 20a and table 15 show the different melting temperatures (Tm's) for HCV genotypes 1, 2, 3, and 5. Genotype 5 showed two peaks, the expected Tm (Tm of 54^0C and another at 60^0C),

which was responsible for the mistyping (Fig.20a). Although, genotypes 1 and 3 have similar Tm's, genotype 3 can be better discriminated in the F3 channel (Fig.20b). The average Tm's of the respective genotypes in channel F2 and F3 ranged from 54^0C to 70^0C (Table 15).

Table 15. The mean Tm's (^{0}C) observed for the genotypes in the study by the Real-time LightCycler genotyping method

Genotype	**Tm^{0}C (F2)**	**Tm^{0}C (F3)**
1	64-66	
2	62-63	
3	68-70	62-64
5	54-57	

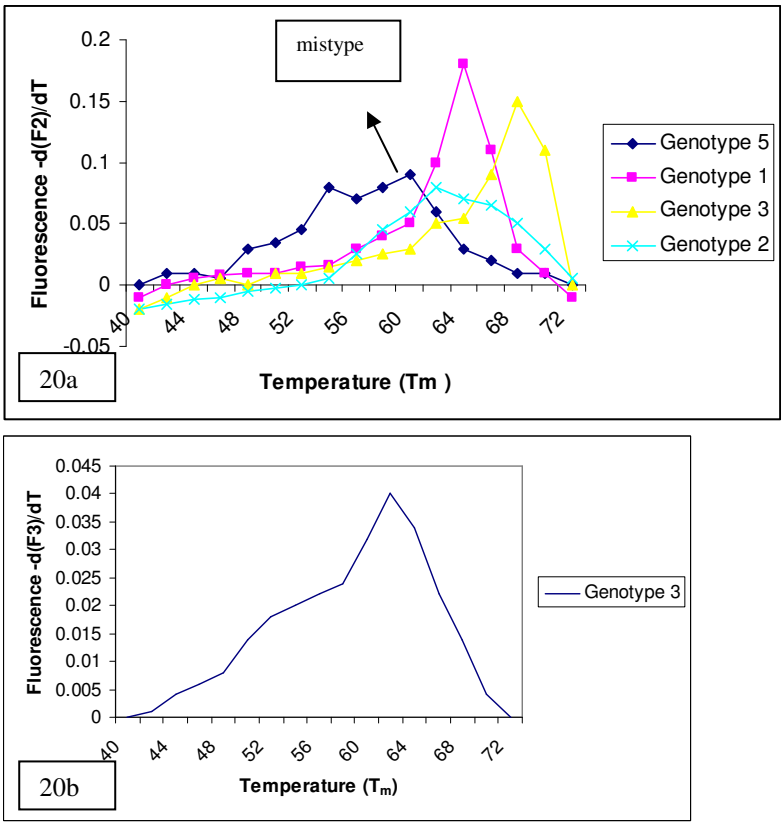


Fig.20. A representation of the Tm analyses of the LightCycler genotyping method, showing HCV genotypes distinguished by their different Tm's, (a) first measured in F2 channel for LC-Red 640emission and (b) in F3 channel to discriminate between genotypes 1 and 3. LC-Red 705 emission signals in F3 are exclusively obtained with a genotype 3 specimen as shown.

4.3.2. LiPA (5'UTR) and LiPA (5'UTR plus core)

4.3.2.1. LiPA (5'UTR)

The validation of the LiPA (5'UTR) method against the QCMD panels (Table 16) were as follows: accuracy= 95%, sensitivity= 94.44%, specificity= 100%, PPV= 100% and NPV= 66.67%. Two genotype 1a specimens in the QCMD panel were subtyped by the LiPA (5'UTR) method as 1b (Table 16). Seven of the eight mixed genotype infections were detected correctly by the LiPA (5'UTR) method (Table 16), and accuracy, sensitivity and specificity scores were greater than 94%.

Table 16. Validation results on the LiPA (5'UTR) method as compared with QCMD results of panels received in 2007 - 2010

QCMD SPECIMENS	QCMD expected results	LiPA (5'UTR) results
1.	1a	1b
2.	5a	5a
3.	3a	3a
4.	1a + 3a	1a + 3a
5.	negative	negative
6.	1b	1b
7.	5a	5a
8.	1b + 3a	1b + 3a
9.	5a	5a
10.	negative	negative
11.	1a	1a
12.	3a + 5a	3 + 5a
13.	3a	3
14.	5a	5a
15.	1b + 3a	1 + 3
16.	3a	3 _
17.	5a	5a
18.	1b + 5a	5a
19.	1	1a
20.	1b	1b
21.	1b + 3a	1b + 3a
22.	negative	negative
23.	1	1b
24.	3a	3a
25.	3a	3a
26.	1b + 3a	1b + 3a
27.	1a	1b
28.	5a + 3a	3a + 5a
29.	5a	Not detected
30.	negative	negative
31.	3a	3a
32.	5a	5a

Accuracy

2.1. Accuracy = $\frac{\text{True positives + True negatives}}{\text{True positives + False negatives + True negatives + False positives}}$

$= \frac{34 + 4}{34 + 2 + 4 + 0}$

$= 95\%$

2.2. Negative Predictive Value (NPV) = $\frac{\text{True negatives}}{\text{True negatives + False negatives}}$

$= \frac{4}{4 + 2}$

$= 66.67\%$

2.3. Positive Predictive Value (PPV) = $\frac{\text{True positives}}{\text{True positives + False positives}}$

$= \frac{34}{34 + 0}$

$= 100\%$

2.4. Sensitivity = $\frac{\text{True positives}}{\text{True positives + false negatives}}$

$= \frac{34}{34 + 2}$

$= 94.44\%$

2.5. Specificity = $\frac{\text{True negatives}}{\text{True negatives + False positives}}$

$= \frac{4}{4 + 0}$

$= 100\%$

All specimens from the blood donor group (2006-2008), which had sufficient RNA (N=100), were genotyped by LiPA (5'UTR) and compared to the SBT. Three specimens typed by sequencing but not LiPA (5'UTR) had a mean viral load of 736 IU/ml

which was below 1000 IU/ml limit for typing by the LiPA (5'UTR) method. Table 17 gives a breakdown of the subtypes identified by the two methods in the volunteer BD (2006-2008) group. While the LiPA (5'UTR) test compared well with the SBT, at the genotype level, the total percentages show, there was a discrepancy at subtype level (Table 17). As with the QCMD validation where LiPA (5'UTR) called 2 known-subtype 1a specimens as subtype 1b, the test called 18 of the 20 specimens sub-typed by the SBT as subtype 1b - 6 more than called by the SBT.

The LiPA (5'UTR) method detected mixed infections (4%, Table 17) while SBT did not. Specimens with mixed subtype infections, for example, specimens 9297 and 9298 [both typed as genotype 1a/1b by LiPA (5'UTR)], grouped as outliers of the genotype 1 cluster (Fig.16). Samples with mixed genotype infection, specimen 9411 (1 + 4) and 9302 (1 + 5a) grouped within the genotype 4 and 5a clusters, re-spectively (Fig.16). Specimen 9411 was an outlier in the genotype 4 cluster.

The LiPA (5'UTR) detected all positive specimens detected by SBT except for 3 which were below the lower limit of viral load detection for LiPA (5'UTR). The LiPA (5'UTR) showed a 100% concordance with SBT in the 5'UTR, although SBT only detected the major variant in the mixed genotype infections. LiPA (5'UTR) detected the presence of mixed genotype infections in 4 of the 100 blood donor specimens (Table 17).

Table 17. Breakdown of subtypes identified by the two methods, Sequencing (SBT N=103) and LiPA (N=100) for the blood donor group 2006-2008

	Genotyping (SBT)		**LiPA (5'UTR)**		
genotypes	N=103	Total (%)	N=100	Total %	p-value (t-test)
1	13	33 (32%)	9	31%	0.75
1a	8		2		
1b	12		18		
1a/1b			2		
					0.99
2	1	1 (1%)	0	0%	
3	3	21 (20%)	4	22%	0.83
3a	18		18		
4	8	9 (9%)	1	9 (9%)	0.92
4a	1		1		
4c/4d			1		
4h			6		
5a	39	39 (38%)	34	34%	0.71
mixed			4	4%	

4.3.2.2. LiPA (5'UTR plus core)

Siemens recognised the inability of the LiPA (5'UTR) assay to accurately distinguish subtypes, particularly for genotype 1and 6. They, therefore, introduced the LiPA (5'UTR plus core) assay evaluated by Verbeeck et al., 2008b.

There was a 100% concordance between the results of the LiPA (5'UTR plus core) and SBT for the 31 genotype 1 specimens examined. The subtypes of these specimens were then compared using the two LiPA methods. Only 2 of the 24 genotype 1a specimens, were correctly subtyped in the LiPA (5'UTR) assay as opposed to 24/24 using the extended assay (LiPA 5'UTR plus core). Although all 6 of the genotype 1b specimens were detected by LiPA (5'UTR), a further 12 genotype 1a specimens were mistyped as 1b. While LiPA (5'UTR plus core) failed to type only one specimen, LiPA (5'UTR) could not type 11 of the 31 specimens (Table 18). A validation of the LiPA (5'UTR plus core) against the QCMD panel showed complete

agreement and the negative specimen was correctly identified (data not shown). Mixed infections GT0804 (1a + 3a) and GT0808 (1b+ 3a) were also correctly detected by both LiPA assays.

Table 18. Subtyping of genotype 1 by LiPA (5'UTR) and LiPA (5'UTR plus core)

Genotypes	LiPA(5'UTR)	%	LiPA(5'UTR plus core)	%
1	9	29.0	**1**	3.2
1a	**2**	6.5	24	77.4
1b	18	58.0	**6**	19.4
1a/1b	2	6.5	0	0

Table 19. Comparison of genotyping results in known mixed infections by LiPA (5'UTR) and LiPA (5'UTR plus core)

Specimen number	LiPA (5'UTR)	LiPA (5'UTR plus core)	Viral load (IU/ml)	
1.	1 + 4a	4	762000	SANBS mix
2.	1 + 5a	5a	339000	SANBS mix
3.	1 + 5a	1a	1130	SANBS mix
4.	1b + 5a	insufficient	437000	SANBS mix
5.	1b + 5a	1b	14400000	Patient specimen
6.	1b + 4	1b	1760000	Patient specimen
7.	1b + 5a	1a	359000	Patient specimen
8.	1 + 5a	1b	2940000	Patient specimen
9.	1 + 5a	1b	936000	Patient specimen
10.	1b + 5a	5a	89100	Patient specimen
11.	1b + 5a	5a	574000	Patient specimen
12.	1b + 5a	5a	94300	Patient specimen
13.	1b + 5a	5a	85400	Patient specimen
14.	1 + 5a	5a	35200	Patient specimen

However, we noticed that the LIPA (5'UTR plus core) was not detecting any of the other mixtures. Verbeeck et al., 2008b did not validate the extended LiPA (5'UTR plus core) assay against mixtures. To confirm our observation we then re-tested the extended assay in mixed infections (Table 19). In 6 cases of mixed genotype 1 and 5a infections, only genotype 5a was detected, whereas in the others, either geno-

type 1a or 1b was detected (Table 19). In mixed genotype 1and 4a infection only genotype 4 was detected. As before, only the mixtures in the QCMD 2008 panel, consisting of genotype 3a with genotype 1a or 1b (data not shown), could be detected.

4.4. DISCUSSION

The three methods assessed in this study genotyped HCV specimens with varying success rates when compared to the gold standard SBT. LiPA (5'UTR plus core) was 100% concordant but the Real-time LightCycler method could only correctly type 64% of specimens. Published real-time LightCycler studies on which the study method was based showed better concordance to sequencing (Schroter et al., 2002 and Landt et al., 2002) but they either did not test genotype 5a specimens at all (Schroter et al., 2002) or tested only "typical" genotype 5a (Landt et al., 2002). The method lacked robustness with our data set and only had an amplification efficiency of 82%. Four of the 5 specimens which did not amplify were genotype 5a. There may have been a loss in integrity of the RNA which resulted in poor or no amplification. This was also noted during the validation process, where the poor NPV and sensitivity values were aggravated by the high number of false negatives.

All, except one of the genotype 5a sequences with the double A at positions 106 and 107 were genotyped by the Real-time LightCycler method as they melted at an average Tm of 57^0C (Landt et al., 2002). However, the limited SNPs used by Schroter et al., 2002 were insufficient for reliable genotyping of local genotype 5a which frequently melted at a higher temperature and could not be distinguished from genotypes 1 and 2 (Fig.20a).

The LiPA (5'UTR) and SBT in the 5'UTR had shown concordant results at genotype level. A similar concordance was seen by Verbeeck et al., 2005 when LiPA (5'UTR) was compared to SBT in the NS4B region. We have shown that the LiPA (5'UTR) is not only comparable to SBT, but surpasses it because it detected mixed genotype infections. The SBT method consistently identified only one genotype within mixtures. Although LiPA (5'UTR) also uses a PCR product, both genotypes in all mixed infections were identified. In a high throughput laboratory, the LiPA (5'UTR) is recommended as genotypes and to a limited extent , subtype 1, can be determined.

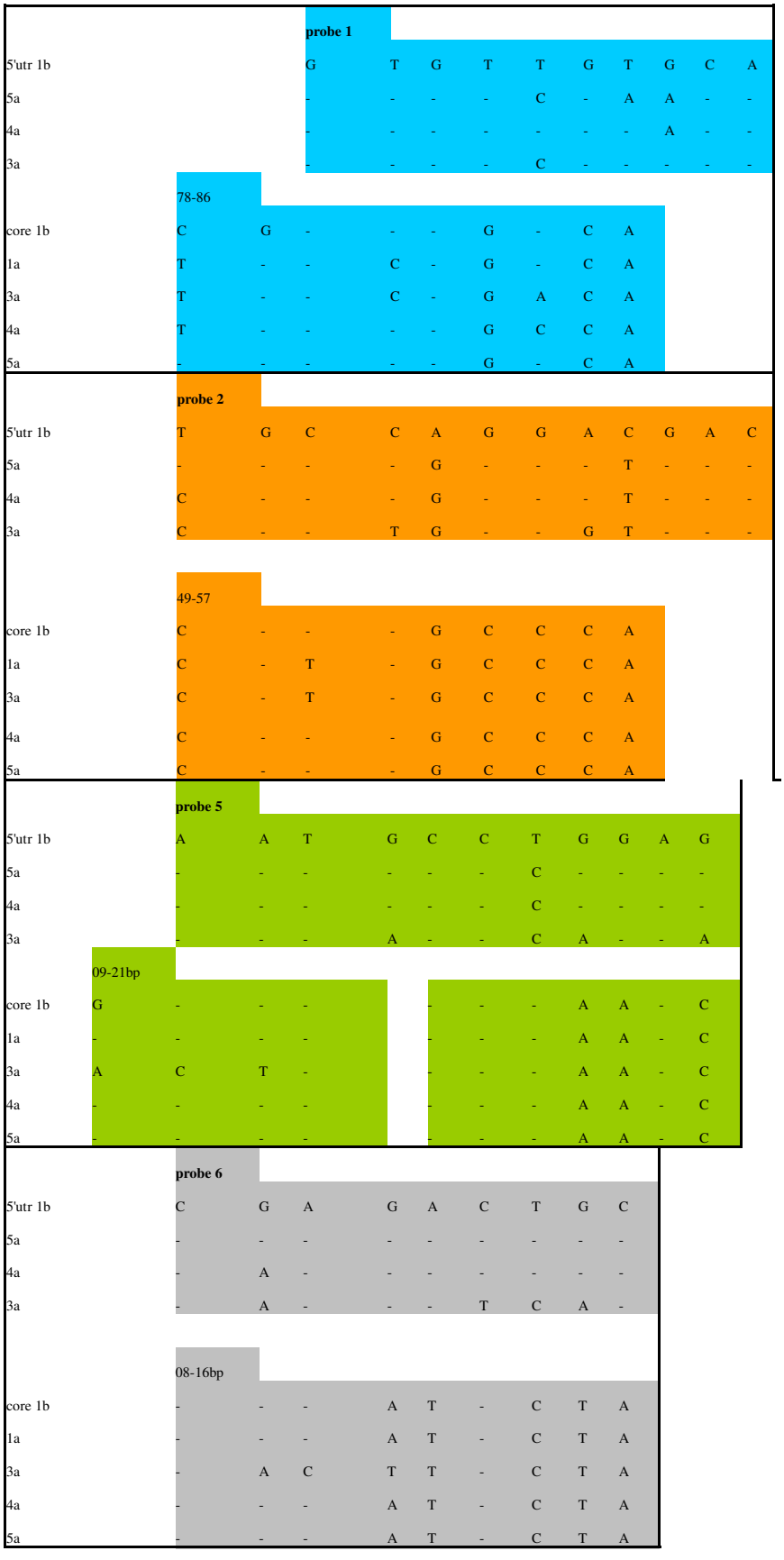
The LiPA (5'UTR plus core) correctly discriminated genotype 1 into subtypes 1a and 1b. The clinical relevance of subtyping genotype 1 has not been fully ascertained. However, if requested, then the LiPA (5'UTR plus core) test can be used as it is a validated and optimized for subtyping genotype 1. However, the cost per specimen, for subtyping using the core amplification assay, does increase (Table 11) and sequencing might be a cheaper option. While, subtyping was better with the LiPA (5'UTR plus core) test, our findings have shown that mixed genotype infections, particularly of genotypes 1+4 and genotypes 1+5, cannot be detected by this test. LiPA strips must be interpreted with caution to avoid mis-calling of ghost bands (Table 11). Also, it has been reported that point mutations caused by PCR amplification in the 5'UTR can alter test results (Li et al., 2008).

In an attempt to explain the inability of the LiPA (5'UTR plus core) to distinguish some mixtures, the known 5'UTR sequences which are used as probes to distinguish types and subtypes of HCV were aligned and compared to core sequences. Sequence similarities between these regions of the 5'UTR and core (Table 20 num-

bering according to Choo et al., 1991), suggest possible cross reactivity between core amplicons and/or the 5'UTR probe sequences. Although, the manufacturer has recently noted that the assay is not optimised to identify mixed infection, we describe for the first time, the missing of specific mixed infections by the assay and suggest possible reasons for the masking of probes in the LiPA (5'UTR plus core) assay.

Even though reports seem to indicate that mixed genotype infections were rare (Viazov et al., 2000, Li et al., 2008), recent data indicate that the diagnosis of mixed genotypes has been increasing (Alavian et al., 2009, 4%; 1 and 3, 3 and 4 and Oh et al., 2008, 4.4%, genotype 1b and 2a/c). Whether this is a result of the evolution of more sensitive HCV genotyping assays or an over interpretation of bands on LiPA strips, this needs to be ascertained by more specific approaches, like cloning the minor variant populations, specific PCR/sequencing primer sets for different genotypes (more costly) or the use of random hexamer or degenerate primers.

Table 20. A summary of the possible overlap of amplified core sequences with the probes 1-6 used in the LiPA 2.0 assay. The dash (-) indicates similarity for the nucleotide.



4.5. CONCLUSION

This comparative study has determined that the LiPA (5'UTR plus core) assay provided the best results for genotyping HCV and subtyping of genotype 1. However, not all mixtures (specifically mixtures of genotype 1 together with either genotype 5 or 4) were detected by this assay. We have identified "masking" effects which may explain the ambiguity of the band patterns in this assay.

The clinical usefulness of knowing the subtype of genotype 1 specimens prior to therapy is still uncertain, whereas the presence of genotype 1, even as a mixture, does affect time required for effective therapy (described further in chapter 8). Thus, in light of the poor performance of the LiPA (5'UTR plus core) assay with mixtures of genotype 1 together with genotypes 4 or 5, the extended assay may not be recommended for diagnostic purposes. However, it will be feasible to use the assay if subtyping of genotype 1 is specifically requested by the clinician and/or subtyping is required for a funded research purpose. Real time PCR for genotyping HCV is not universally reliable at present. Other regions of the highly mutable HCV genome are being studied for use in a real-time PCR method for genotyping (Ciotti et al., 2010).

Of the methods examined in this chapter, the LiPA (5'UTR) assay has proven to be adequate for determining genotypes as required presently in our diagnostic setting. In surveillance – funded projects, this method is beneficial as the amplicon from the qualitative PCR is used for the LiPA(5'UTR) assay. Our annual validation processes continually assess the efficiency of this method and the LiPA (5'UTR plus core). With the less labour-intensive process, easy usage and quicker turn-around times (TAT's), the LiPA (5'UTR) can be recommended for high throughput laboratories.

CHAPTER FIVE

Palindromic-nucleotide Substitutions (PNS) Patterns within the major genotypes found in SA

5.1. INTRODUCTION

Palindromic-like nucleotide substitution (PNS) patterns have been found in the stem -loop structures of pestiviruses (Harasawa and Giangaspero, 1998) and HCV (Giangaspero et al., 2008) and are the result of viral evolution over time. The conserved palindromic substitutions stabilize the stem regions and maintain the basic structure (Harasawa and Giangaspero, 1998). The stem-loop structures of domains IIIa, b and c (Piron et al., 2005, Fig.21a), reflect the characteristics of the HCV isolate, by genus, genotype and subtype (Giangaspero et al., 2008). The PNS patterns provide information on the base-pair signatures to which the primary sequence conforms to in the stem-loop. Other than genotyping, this structural information at the 5'UTR can be used to design targets for siRNA therapies (Prabhu et al., 2006) and reverse genetics (Friebe and Bartenschlager, 2009).

The aim of this study was to investigate these genotype specific palindromic substitutions within the 5'UTR of the HCV genotypes most prevalent in South Africa, 1 and 5a.

5.1.1. Objective:

To investigate the genotype specific palindromic-like nucleotide substitutions (PNS) within the 5'UTR of genotypes, 1 and 5a, and discuss with reference to SBT and their usefulness as a diagnostic technique.

5.2. METHODS

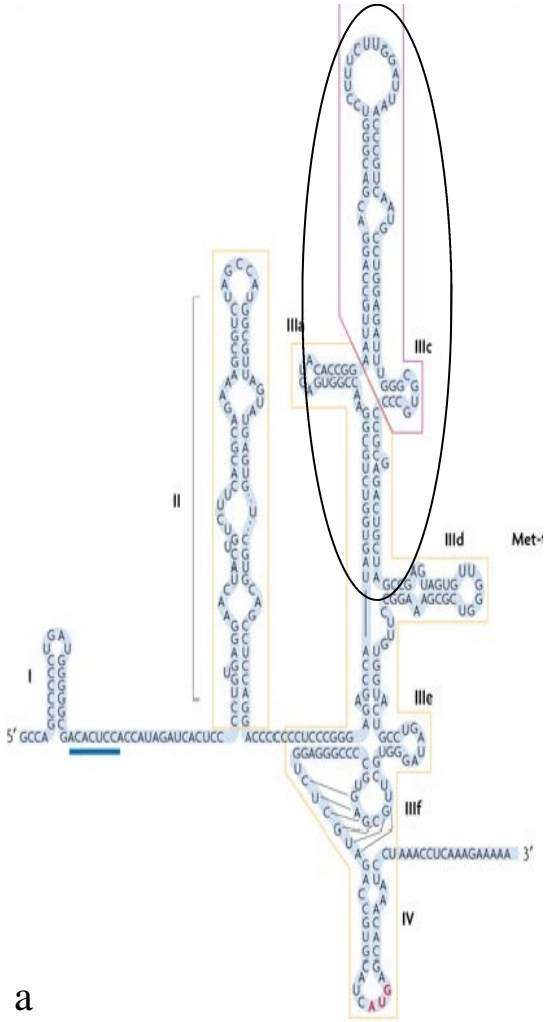
Sixty-one sequences from previously genotyped specimens (genotype 1, N=26; genotype 5, N=35; chapter 3) were formatted and aligned using ClustalW in BioEdit, version 7.0 (Hall, 1997) over the informative regions (179 nucleotides) of the 5'UTR as described by Giangaspero et al., 2008. For genotype 1 analyses, reference strains, M62321 (1a), D50480 (1b), D14853 (1c) were used. For genotype 5a analyses, Y13184 (5a), AF064490 (5a), U33430 (5a), AM502711 (5a), AM502710 (5a), DQ164748- DQ164751 (5a), AY033769 (5a) and L28057 (5a). Sequences were then analysed for PNS patterns as described previously (Giangaspero et al., 2008).

5.3. RESULTS

Of the 26 genotype 1 specimens analysed, 20 are shown in Table 21a and of the 35 genotype 5a specimens analysed, 16 are shown in Table 21b. The PNS method scored a 100% concordance with SBT, and a 100% for NPV, PPV, sensitivity and specificity at the genotype level. However, like SBT using the 5'UTR region, the PNS method does not discriminate subtypes well. The characteristic genotype PNS patterns identified by Giangaspero et al., 2008 for genotype 1 and genotype 5a were observed in our study. In the primary nucleotide alignments (Table 21a) of genotype 1 (141-252 or -201 to –89, according to Choo et al., 1991) and 5a sequences (Table 21b, 102-227 or -240 to –114, according to Choo et al., 1991), the shaded regions highlight important base pair signatures in the stem-loop sequences. Row 1 of the tables refers to the PNS pattern number (see also Figure 21, specifically, Fig. 21b) while row 2 gives the actual nucleotide number in the primary sequence data. The RNA folding is a result of base-pairing of upstream nucleotides

with complementary nucleotides further downstream. For example, PNS pattern 28 pairs nucleotides 179 and 220 (Table 21a).

The highlighted region on Fig.21a shows the PNS in the stem-loop sequences over the 141-252 nucleotide base pairs (Fig. 21b). The numbering of the nucleotides and variable structural regions is according to Giangaspero et al., 2008. The arrows in the tables indicate variations among genotypes observed in this study. Position 204 (Table 21a) has a nucleotide base change from cytosine (C) to adenine (A) (specimens 420/LD1, 1958/H, 1956/H, 1108/H, 1485/H and 2763/H) or thymidine (T) (genotype 1c reference sequence (D14853), 6641/H). Position 243, identified as a PNS characteristic (position 10) for subtyping genotype 1 into 1a and 1b (Giangaspero et al., 2008), corroborates the subtype change from adenine (A) (genotype 1a) to guanosine (G) (genotype 1b). This 1a/1b clustering is identical to that seen in the phylogenetic analysis (in chapter 3, Fig.16, page 95) where specimens with guanosine (G) at 243 clustered with the reference 1b strain, D50480.



47	UU	UU
46	C G	C G
45	U G	U G
44	U A	U A
43	U U	U U
42	C H	C A
41	C A	C A(A)
40	**U-A**	**U-A**
39	**G-C**	**G-C**
38	**G-C**	**G-C**
37	**G-C**	**G-C**
36	C-G	C-G
35	**C C**	**C C**
34	**A-U**	**A-U**
33	**G-C**A	**G-C**
32	B A	Y:AA
31	A Y	A U
30	G-CG	G-CG
29	**G-C**	**G-C**
28	**A-U**	**G-C**
27	C-G	C-G
26	**C-G**	**C-G**
25	G A	G A
24	U*G	**U*G**
23	**U-A**	**U-A**
22	**A-U**	**A-U**
21	**A-U**	**A-U**
	A 20 **G*U** /19 C	A **G*U** C
	W CACCG GGG G	U CACCG GGG G
	G GUGGC CCC Y	G GUGGC CCC U
	A 17 C . \18 G	A C . G
16	A .	A .
15	W .	A .
14	**G-C**	**G-C**
13	**G-C**	**G-C**
12	**C-G**	**C-G**
11	G-C	G-C
10	. ***R***	. G
9	**U-A**	**U-A**
8	**C-G**	**C-G**
7	**U-A**	**U-A**
6	G-C	G-C
5	G:Y	G*U
4	U*G	U*G
3	**G-C**	**G-C**
2	**A-U**	**A-U**
1	5'-**U-A**-3'	5'-**U-A**-3'
	(HCV-1)	(HCV-5)

b

Fig.21. (a) Secondary structure of the 5'UTR of the HCV genome (Fraser and Doudna, 2007), with the IIIa-c region highlighted as the regions sequenced in the study. (b) Variable locus in the secondary structure of the 5'-UTR of the genotypes 1 and 5 of hepatitis C virus species, genus *Hepacivirus*. Base pairings characteristic to the genus (PNS genus specific) are shown in bold. The characteristic base pairings of the HCV genotypes (PNS genotype specific) are represented in bold and underlined. Base pairings characteristic to subtype-specific PNS are shown in bold and italic. Watson–Crick base pairings are indicated by a dash (–); tolerated pairings in secondary structure are indicated by an asterisk (*); interchangeable base pairings are indicated by a colon (:).
R =A or G; W=A or U; Y = C or U; B= C or G or U; H=A or C or U.

Table 21a. Nucleotide alignment (nucleotides 141-252) of genotype 1 sequences to reference sequences for PNS analyses. GenBank sequences: , M62321 (1a), D50480 (1b) and D14853 (1c). The shaded regions (nucleotides 179 plus 220 and 243) highlight structural information at base-pair 28 and 10, respectively, which are specific PNS signatures for genotype 1. The arrow marks the previously unreported variation at nucleotide 204 which was observed in this study. The PNS results are compared to sequencing in the 5'UTR and NS5B regions.

Stem loop domains III a, b, c

	Genotype	1	2	3	25	26	27	28	29	30	31	39	40	41	42	41	40	39	38	37	28	27	26	25	24	23	22	21	13	12	11	10	9	8	7	6	5	4	3	2	1	PNS vs sequencing		
		141	142	143	176	177	178	179	180	181	182	201	202	203	204	205	206	207	208	209	220	221	222	223	224	225	226	227	228	229	242	243	244	245	246	247	248	249	250	251	252	5'UTR	NS5B	PNS
M62321(1a)	1a	T	A	G	G	C	C	A	G	G	A	G	A	T	C	A	A	C	C	C	T	G	G	A	G	A	T	T	C	G	C	A	A	G	A	C	T	G	C	T	A	reference 1a		
420/LD1	1a	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1a or 1c	1a	1a or 1c
1958/H	1a	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1a or 1c	1a	1a or 1c
1956/H	1a	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1a or 1c	1a	1a or 1c
307/H	1a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1a or 1c	1a	1a or 1c
9049/LD1	1a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1a or 1c	1a	1a or 1c
D50480(1b)	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	reference 1b		
1108/H	? 1b	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	outlier 1a or 1c	ND	1b
1485/H	? 1a	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	1a	1b
2763/H	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	ND	1b
4434/LD2	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	ND	1b
967/LD2	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	ND	1b
2771/LD2	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	ND	1b
2585/LD2	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	ND	1b
5307/LD2	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	ND	1b
3894/H	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	ND	1b
6373/H	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	ND	1b
1229/LD1	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	1b	1b
2205/LD1	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	1b	1b
356/LD1	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	1b	1b
1109/H	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	1b	1b
6641/H	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	1b	ND	1b
D14853(1c)	1c	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	reference 1c		

Table 21b. Nucleotide alignment (nucleotides 102-110, 173-183, 201-209, 214-216 and 221-227) of genotype 5a sequences to reference sequences for PNS analyses. The shaded regions (nucleotides 175, 225 and 179, 221 highlight structural information, at basepair 24 and 28, respectively) indicate PNS signatures for genotype 5a. The columns with arrows 106, 183 and 207 reveal nucleotide base substitutions observed in our study.

Domain II

Stem loop domains III a, b, c

	Country of origin	102	103	104	105	106	107	108	109	110	22 173	23 174	24 175	25 176	26 177	27 178	28 179	29 180	30 181	31 182	32 183	45 201	44 202	43 203	42 204	41 205	41 206	40 207	39 208	39 209	32 216	32 215	29 214	28 221	27 222	26 223	25 224	24 225	23 226	22 227	Genotype
Y13184	SA	G	T	C	G	A	A	C	A	G	A	T	T	G	C	C	G	G	G	A	T	G	A	T	A	A	A	-	C	C	A	A	C	C	G	G	A	G	A	T	5a
AF064490	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
AM502711	France	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
AM502710	France	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
DQ164750	Belgium	N	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
2577/LD1	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
DQ164751	Belgium	N	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
DQ164749	Belgium	N	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	5a
DQ164748	Belgium	N	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
U33430	Canada	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
3494/LD1	SA	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	atypical 5a
1116/LD1	SA	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	atypical 5a
3788/LD1	SA	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	atypical 5a
2473/LD1	SA	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	atypical 5a
1044/LD1	SA	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	atypical 5a
L28057	Australia	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
AY033769	Brazil	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	5a
3065/LD1	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	5a
2014/LD1	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	5a
918/LD1	SA	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	5a
2031/LD1	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	5a
151/LD1	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
3975/LD1	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
51/LD1	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
363/LD1	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
66/LD1	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a
119/LD1	SA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5a

The specific combination of two base-pairings for genotype 5a were seen: a U*G at position 24 (also shared with genotype 1 and 6) but differs for genotype 2 (UC bulge) and genotypes 3 and 4 (CA bulge); and a G-C base pair at position 28 (seen also with genotypes, 2, 3, and 4) but differs among the genotype 1 (A-U, Giangaspero et al., 2008). We also found a thymidine (T) change to cytosine (C) at position 183 for one of our isolates, 2577/LD1, as well as two French (AM502710, AM502711) and one Belgium isolate (DQ164750, Table 21b). In eight of our specimens and the Brazilian isolate (AY033769), there was an adenine (A) insertion at position 207 (only 4 of the 8 sequences are shown in Table 21b), similar to that seen in genotype 6a and 6b. The genotype 5a specimen sequences with these changes did not cluster significantly differently by phylogeny (in chapter 3, Fig.16, page 95). However, 2577/LD1did present with a distinct branch length.

5.4. DISCUSSION

Compared to the LiPA (5'UTR plus core), the PNS method is time-consuming. It not only requires sequencing, but also complex analyses of the sequence data. Our results were concordant with the sequencing results and the specific patterns identified by Giangaspero et al., 2008. We found three more variations among our genotype 5a sequences as indicated at positions 107, 183 and 207. Whereas the nucleotide change at position 107 lies outside the stem-loop, the change at position 183 lies in the stem region between IIIa and IIIb. The C change at position 183 has no effect on its complementary nucleotides AA at position 215 and 216, respectively. However, we indicate that the stable U-AA bond will change to a Y: AA bulge in position 32 (Fig. 21b). The insertion at position 207, seen in some of the genotype 5a sequences, is part of the IIIb loop. This change is not evident in other genotypes

studied previously. The insertions of A at position 207 and CA at position 197 (41 and 47 in the structural alignment respectively) were described as a unique PNS characteristic to two of the genotype 6 subtypes, 6a and b providing a higher secondary loop structure (Giangaspero et al., 2008). These insertions were not seen for subtypes, 6d, g, h and k. This insertion did not elongate the loop, but an added A at position 41 for the genotype 5a sequences in this study is noted. Functionally, since this occurs in the loop structure, little or no effect on viral viability and replication is expected (Wang et al., 2005) but specific changes should be taken into account for primer design.

The PNS method differentiated subtypes 1a and 1b by the base pair change at nucleotide position 243, observed in PNS position 10. However, A is also present for subtype 1c. Sequence information of nucleotides 1 to 80 would have cleared this ambiguity as the PNS characteristic for genotype 1c (GG bulge at position 7) is identified in the stem-loop I structure (Giangaspero et al., 2008).

However our phylogenetic analyses in the 5'UTR and NS5B region (Prabdial-Sing et al., 2008; chapter 3, Fig. 16 and 18, respectively, page 95, 97) has shown no evidence of any specimen grouping with the genotype 1c reference sequence, D14853. The nucleotide change seen among our genotype 1 at position 204 was recognised by Giangaspero et al., 2008 and seemingly had no structural change in the IIIb loop.

5.5. CONCLUSION

The PNS method had successfully genotyped 1 and 5a specimens and it is now possible to visualise primary sequence data of the UTR as a stem-loop structure. At

present PNS analysis is labour intensive, but the method would have potential were computer programmes are easily accessible and available to fully automate the 3D analysis. PNS would then provide a rapid, reliable alternative to phylogenetic analysis which is also capable of differentiating genotypes within the conserved 5'UTR region.

CHAPTER SIX

Molecular characterization of genotype 5a by partial sequence analysis of clinical specimens

6.1. INTRODUCTION

The HCV genome is a paradox as it has highly conserved and variable regions. The genome of HCV has highly variable (E1-E2), semi conserved (core, NS5B) and highly conserved (5' and 3' UTR, Fig.2) regions. It displays both "Darwinian" and "neutral" evolutionary traits (Simmonds, 2004) which allow, on the one hand, viral fitness and immune escape and on the other hand, the "neutral drift" whereby changes to the nucleotide sequence has no or little effect on viral phenotype and accounts for the HCV genotype diversity, which is geographically distributed (Simmonds, 2004). This dual nature of HCV evolution may account for the high turnover of closely related sequences (quasispecies) and survival in the host (Sarrazin and Zeuzem, 2010).

This quasispecies nature, common in RNA viruses, is maintained by immunological pressure *in vivo* (Yao et al., 2005) and accounts for disease chronicity (Farci et al., 2000) and the varied outcome of interferon-based therapy (Martell et al., 1992). The viral RNA-dependent RNA-polymerase (RdRp) protein encoded by the NS5B region (Fig.2) lacks proof-reading and repair mechanisms and contributes to the high frequency of mutations (Steinhauer et al., 1992).

With the heterogeneity of the HCV sequence, it is difficult to successfully amplify full -length HCV genomes. It has been estimated that 12 sets of primers will be neces-

sary for every 4 amplicons with at least a further 9 permutations per amplicon (Yao et al., 2005). This process of amplification is expensive and labour-intensive. Furthermore the low titre of virus in clinical samples (<10^7 IU/ml, Table 1) and the lack of an efficient cell culture system to amplify all types of HCV makes the generation of long template amplicons of HCV difficult and problematic. Relevant properties of HCV are summarized below (Fan et al., 2006):

(a) The virus is very labile and sensitive to poor handling, incorrect storage temperatures and freeze/thaw cycles.

(b) HCV folds to form secondary structures along its genome (Tuplin et al., 2004) severely complicating full-length cDNA synthesis.

(c) Although HCV *in vitro* studies have been promising, the success is limited to only a few genotypes. Concentration of virus in clinical samples is relatively low compared to other viruses (Fan et al., 2006).

Because of these difficulties, many epidemiological surveys use small fragments for phylogenetic studies. Shorter sequences are quicker to generate and more reliable (Sheehy et al., 2005). Care must be taken when using sub-genomic regions as not all regions are phylogenetically informative for all subtypes (Simmonds et al., 1993b; Verbeeck et al., 2006), as is the case for highly conserved regions such as the 5'UTR (Chapter 3). Concatenation of several sub-genomic regions is an attractive compromise. Verbeeck et al., 2006 found that there was more phylogenetic signal from the concatenated E1/NS4B region, compared to the individual E1 or NS4B, respectively. Concatenation of similarly conserved or less conserved regions may not skew overall results but may provide for better bootstrap values of a monophyletic clade (Verbeeck et al., 2006). However, often the same sequence data is not avail-

able for all specimens studied or reference sequences may not be available for comparison (Pybus et al., 2009).

Currently, the International Nucleotide Sequence Database Collection (INSDC, previously known as DDBJ, EMBL, GenBank) has 44000 HCV reference sequences (Shin et al., 2008). Few full genomes of the lesser known genotypes (particularly genotype 4 and 5) appear in the international database. Less than 1% of these are genotype 5a. Genotype 5a is not a predominant subtype in western countries and very few molecular studies have been performed to sequence and characterize it. Hence, variations of genotypes 5a are not well documented. Pang et al., 2009 determined the evolutionary branching order of genotypes in correlation to their respective responses to therapy. They looked at 348 full genome sequences from the European HCV database (Combet et al., 2007), of which only 2 are genotype 5a. Due to this poor representation, there was uncertainty with regard to the genotype 5a branching order and response to therapy.

The aim of this part of the study was the molecular characterization of genotype 5a isolated from South African patients, performed by partial sequencing of subgenomic regions. Different phylogenetic programs were used to best explain clustering in the semi-conserved core and NS4B regions and less conserved E1 region. The divergence of genotype 5a from other types and subtypes in the human population was calculated by using various phylogenetic inference packages.

5.1.1. Objectives:

(a) Sequence analyses of the core, envelope (E1) and non-structural 4B (NS4B)

regions of genotype 5a.

(b) Phylogenetic relatedness of these partial sequence data with genotype 5a from other parts of the world.

(c) To establish by phylogenetic evolutionary analyses, the most recent common ancestor (MRCA) of genotype 5a and thereby its probable divergence time as a discrete type in the human population.

6.2. METHODS

6.2.1. Study participants

Partial sequencing of 25 genotype 5a, 1 genotype 1a (427), 1 genotype 1b (9400) and 1 genotype 4 (9411) previously genotyped specimens (chapter 3) was performed. Specimen choice was dependent on the volume and condition of the stored retrospective serum/plasma specimens. These were all once-off collections, so serial analysis of samples was not possible. Sampling dates were also not available for these specimens. Specimens sequenced were extracted from 13 anonymous volunteer blood donors, (acquired from SANBS), 13 patients with known liver disease and 2 HCV infected haemophiliacs. Specimens with viral load of $>10^5$ IU/ml (Cobas Amplicor Monitor, Roche Diagnostics, Germany) were used.

6.2.2. RNA extraction and primer design

RNA extraction from stored sera was performed by using the QIAamp Mini RNA extraction kit (Qiagen, Germany) as it was found to be more sensitive than the MagnaPure method. Viral RNA was extracted from 280 µl of serum or plasma, according to the manufacturer's protocol and eluted in 50 µl double-distilled water. The concentration of the extracted RNA was measured on the Nanodrop (Thermo-

Scientific, USA). Full-length genome sequencing was attempted by long template PCR and "sequence walking". For long template PCR (>3000bp) the following primers were used: random hexamer primers, poly adenine/uracil primers (Transcriptor First Strand cDNA Synthesis, Roche Diagnostics, Germany) and primers specific to genotype 5a (adapted from Fan et al., 2006, Table 22). For "sequence walking", primers for amplification of shorter fragments were designed for genotype 5a. A list of the primers used for RT, PCR and sequencing of the core, E1 and NS4B regions is provided in Table 22. Primer sequences are numbered according to Choo et al., 1991 (Fig.2) and primer sequences of the antisense 8907, core and the NS4B inner antisense 4963 have been changed to suit genotype 5a sequences (Table 22).

6.2.3. cDNA synthesis

Both the reverse-transcription and PCR reactions were optimized using a variety of enzymes. A comparison of the length of amplicon generated and the start-up concentration required for an efficient amplification is listed in Table 23, which can be used as a reference guide for other students. The superscript RT III enzyme (Invitrogen, USA) performed best over all sub-genomic regions, could amplify low concentrations of material and was ideal for the clinical specimens (Table 23). Ten µl of extracted RNA was added to 2 µl (1 µM) antisense primer 8907 (Table 22) and 1.5 µl (1.3mM) dNTP mix (10 mM each dATP, dGTP, dCTP and dTTP at neutral pH). The mix was heated at 65^0C for 5 minutes and then cooled on ice. Four µl of 5X First Strand buffer was added to the mix, with 1 µl (5mM) DTT, 0.5 µl (20U) RNAse inhibitor (Protector RNAse inhibitor, Roche Diagnostics, Germany) and 1 µl (200U Superscript RT III enzyme). The final volume of 20 µl was incubated at 50^0C for 75 minutes, followed by heat inactivation at 70^0C for 15 minutes as per

manufacturer's protocol.

Table 22. List of primers , primer details and combinations and amplicon sizes for method used (for the core, E1 and NS4B regions (numbering according to Choo et al., 1991)

Region	Primer	Sequence	Tm (^{0}C)	G:C	Amplicon size (bp)	method
GENOME	Asn 8907-8924	5'- GCGCCAACGGTRAACCAG - 3'	61	61	cDNA	RT
CORE	Outer Sn [1] (-297 -277)	5'-CTGTGAGGAACTACTGTCTT-3'	55	45	269	1st PCR
	Outer Asn[2] (508-529)	5'-GATAGAGAAAGAGCAACCGGG-3'	53	61		
	Inner Sn [1] (-278-258)	5'-TTCACGCAGAAAGCGTCTAG-3'	55	50	238	Nest/Seq
	Inner Asn [3] (458-482)	5'-GTTGCATAGTTTATCCCGTCTTCAAGAACC-3'	55	50		
CORE/E1	Outer Sn[4] (8-28)	5'-CGAATCCTAAACCTCAAAG-3'	47	42	971	1st PCR
	Outer Asn[5] (957-979)	5'-ACCATTTCATCATCATGTCCCA-3'	51	41		
	Inner Sn[4] (19-48)	5'-CCTCAAAGAAAAACCAAAAGAAACAAC – 3'	54	33	960	Nest/Seq
	Inner Asn[5] (951-979)	5'-ACCATTTCATCATCATGTCCCATGCCAT-3'	58	43		
NS4B	Outer Sn[6] (4696-4717)	5'-ATCAACATCGACGCYCACATG - 3'	57	52	617	1st PCR/ seq
	Outer Asn[6] (5293-5313)	5'-CCCACTGACAAAGTTCCACAT-3'	59	48		
	Or Inner Asn[7] (4963-4982)	5'-CCCACGTGCTGGTAATGAC-3'	60	58	268	Nest/Seq

[1]Chan et al., 1992; [2] Okamoto et al., 1990; [3] Bukh et al., 1994; [4]Chamberlain et al., 1997; [5]Legrand-Abravanel et al., 2007; [6]Verbeeck et al., 2006; [7]Yao et al., 2005; sense, Sn and anti-sense, Asn, RT, reverse transcriptase; nest, nested PCR and seq, sequencing PCR

Table 23. A list of enzymes were used for reverse-transcription (RT) and first and nested PCR for sequencing of HCV

Enzyme (RT)	Manufacturer	Amplicon length	Concentration required
iScript cDNA	BioRad	<1kB	100fg-1µg/µl
Superscript III	Invitrogen	<12kB	10pg-5µg/µl
Transcriptor First strand cDNA	Roche Diagnostics	<14kB	10ng-5µg/µl
Transcriptor Hi-Fidelity	Roche Diagnostics	<14kB	1ng-4µg/µl
Transcriptor Reverse transcriptase	Roche Diagnostics		
RT- One Tube TITAN	Roche Diagnostics	<6kB	1pg-1µg/µl
Enzyme (nest PCR)			
FastStart Taq	Roche Diagnostics	<3kB	10pg-100ng/µl
Expand Long Template	Roche Diagnostics	<22kB	<500ng
iProof Hi-Fidelity	BioRad	<7.5kB	1pg-10ng/µl
Long PCR Enzyme		<21kB	2.5ng/µl
rTth DNA polymerase, XL (requires inhibitor Trnc-21)	Applied Biosystems	5kB-40kB	10^4 – 10^7IU/ml 1µg

6.2.4. PCR and sequencing

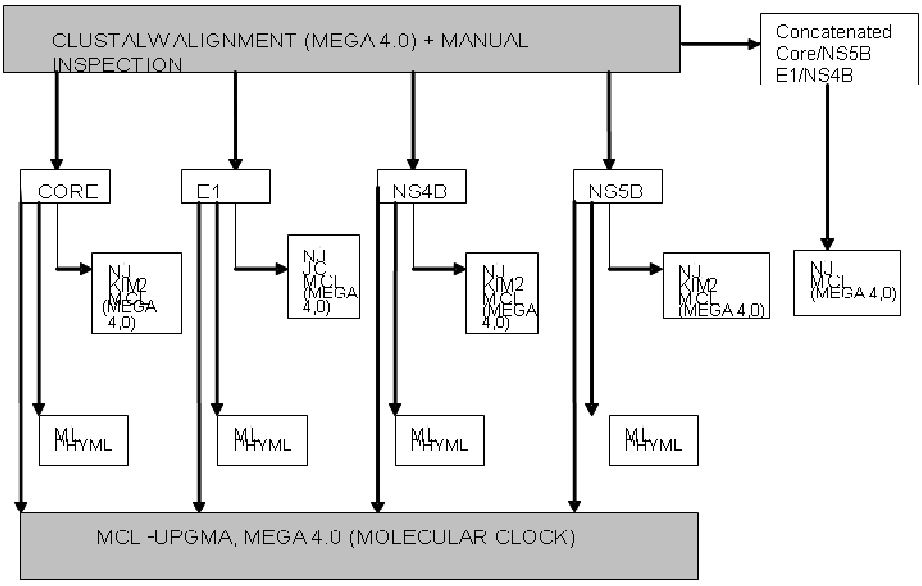
Five µl of cDNA was added to the first round of PCR mix of 5 µl (10X) PCR buffer with 2 mM $MgCl_2$, 10 µl (5X) GC rich solution (additive for difficult templates), 2.5 µl (0.5uM) of each primer for the respective regions (Table 22), 1 µl (0.2 mM) of the dNTP mix and 0.5 µl (2.5 U) of the FastStart Taq polymerase (Roche Diagnostics, Germany) and a final volume of 50 µl was amplified. The thermocycler parameters were programmed as 98^0C for 2 minutes, followed by 35-45 cycles of 98^0C for 1 minute, 55^0C for 1 minute and 72^0C for 1 minute and a final extension at 72^0C for 5-7 minutes. Five µl of the first round PCR product was added to the nested amplification mix with primers (Table 22). The specimens that were amplified in the NS5B region were subjected to this protocol using the primers mentioned in chapter 3 and Table 8. The PCR additives for the nested mix and the cycle parameters were the same as the first round. The first round and nested PCR products were analysed on a 1-2% agarose gel. PCR products were purified by the MSB Spin PCRapace clean -up (Invitek, Germany). Direct sequencing on both forward and reverse strands were performed on the purified DNA fragments using the ABI Prism BigDye Terminator Cycle Sequencing reaction kit, version 3.1 (Applied Biosystems, USA). The primers that were used for the nested PCR protocol (Table 22) were used for the sequencing reactions and the reaction mix was prepared as recommended by the manufacturer. Chromatogram analyses were performed by the Sequencher, version 4.1.4. Software.

6.2.5. Phylogenetic analyses

A flow-chart to describe the analyses is presented in Fig.22. Multiple sequence alignments were generated by the ClustalW (Higgins, 1994) program in MEGA 4.0

(Tamura et al., 2007) for the core, E1, NS4B, NS5B and concatenated core/NS5B and E1/NS4B regions. The Kimura-2 parameter (Kimura, 1980) was chosen for the phylogenetic analyses for the core, NS4B and NS5B. The number of nucleotide substitutions/site between the different sequences (d) was calculated to be between 0.05 and 0.3 and Kumar et al., 2008 substantiate that if 0.05<d<0.3, then the Kimura -2 parameter can be used, particularly if the transition/transversion rate (R) >5. Although, d was lower at 0.03 for the core region, R was >5 and the Kimura-2 parameter was used. The higher diversity of 0.371 for the E1 region prevented pairwise calculations using the Kimura-2 parameter and the Jukes-Cantor (JC, Jukes and Cantor, 1969) parameter was chosen for this region.

Fig.22. An explanatory flow-chart to indicate the methods used for the phylogenetic and evolutionary analyses conducted for the different sub-genomic regions.



NJ, neighbour-joining; Kim-2, Kimura-2 parameter; MCL, Maximum composite Likelihood; ML, Maximum likelihood; UPGMA, Unweighted pair group method with arithmetic mean

Phylogenetic analyses for the concatenated regions were performed with the Maximum Composite Likelihood (MCL) parameter in MEGA 4.0 in order to compare phylogenies and branching order of these regions with each individual region, respectively (Fig.22). The maximum likelihood (ML) model (Phyml v.2.4.4, Guindon and Gascuel, 2003) provided as an online service (www.hcv.lanl.gov) was used for analyses on the core, E1, NS4B and NS5B regions. For the ML analyses, fasta-formatted sequences were submitted online and the FindModel function (a web function of the ModelTest script) was chosen. For the General Time Reversible (GTR, Tavare, 1986) model, the discrete gamma model with the gamma shape parameter of 2 was chosen and tree topology was optimised by the online software. Although, the Transversion model was best suited for the core region, this model is not available on MEGA 4.0 or online and the Kimura-two model was used for the core data. Bootstrapping for both the neighbour-joining methods, Kimura-two and Maximum Composite Likelihood (MCL), and the ML parameters were tested at a 1000 replicates and consensus trees were viewed by using MEGA 4.0 and TreeView, respectively. Bootstrap values of >70% were displayed on phylogenetic trees. The mean overall nucleotide differences were calculated per region (MEGA 4.0) across all genotypes and for genotype 5a only, respectively. The concatenated regions were subjected to MCL analyses only.

Reference strains used in the study were attained from GenBank (accession number): for the **core** region; D16789-D16795 (5a clones, Ohno et al., 1994), Z29470-Z29474 (5a, Belgium, van Doorn et al., 1995); **E1**, FN553052-FN553065, FN552976, FN552983, FN552995 (5a, France, Henquell et al., 2004), L39281, L39295 (5a, Belgium, van Doorn et al., 1995); **NS4B**, EF026073, AM408911 (2/5a

recombinant, Legrand-Abravanel et al., 2007), DQ164642-DQ164747 (5a, Belgium, Verbeeck et al., 2006), FN553233-FN553282 (5a, France, Henquell et al., 2004); **NS5B**, AJ291281 (5a, France, Morice et al., 2001), FN553360 (5a, France, Henquell et al., 2004), L29584 (5a, Belgium, Stuyver et al., 1994), AJ608776 (5a, Henquell et al., 2004), FN553377 (5a, France, Henquell et al., 2004), L29602 (4, Belgium, Stuyver et al., 1994), L36437 (4, Central African Republic, Henquell et al., 2004), L29590 (4, Belgium, Stuyver et al., 1994), D86543 (4, Japan, Tokita et al.,1998), L78841 (4, Belgium, Stuyver et al., 1994), D86534 (4, Japan, Tokita et al.,1998), L36438 (4, Central African Republic, Henquell et al., 2004), L29593 (4, Belgium, Stuyver et al., 1994), AY685052 (4, Cameroon, Njouom et al., 2005), AY685012 (4, Cameroon, Njouom et al., 2005), L29618 (4, Belgium, Stuyver et al., 1994), L29611 (4, Belgium, Stuyver et al., 1994), D86537 (4, Japan, Tokita et al.,1998), AY265451 (2, Cameroon, Njouom et al., 2005), AY265420 (2, Cameroon, Njouom et al., 2005); **concatenated regions**, D17763 (3a), D49374 (3b), D50480 (1b), M62321 (1a), D14853 (1c), Y12083 (6a), D50409 (2c), D10988 (2b), D00944 (2a).

6.2.6. Evolutionary history of genotype 5a

For this part of the analyses, we only inferred a divergence time for the genotype 5a cluster that consisted of our own sequences and those from the global database (mostly Belgium and France). From our analyses and published articles, it has been inferred that genotype 2 is the oldest genotype and we hence rooted the genotype 5a evolutionary trees to genotype 2. The MCL method using the Unweighted Pair Group Method with Arithmetic mean (UPGMA, Sneath and Sokal, 1973) parameter (MEGA 4.0) was used for the molecular evolutionary inference for the core, E1,

NS4B and NS5B regions. Bootstrap analysis was performed using 1000 replicates. The evolutionary rates were estimated from published records and the time line for each region was calculated at a 100 year interval.

6.3. RESULTS

6.3.1. Demographics

The sera/plasma of specimens from 25 subjects previously typed as genotype 5a in the 5'UTR (discussed in chapter 3) was further characterized by subgenomic sequencing. There were 2 specimens that genotyped as 1 and one specimen that was a genotype 4 in the core and E1 regions. There were an equal number of males and females with a mean age of 42 years (Table 24). The ethnicity of this group was mostly Blacks (54%), Whites (14%), Asian (4%) and data unknown (28%). Almost half of the samples were from patients diagnosed with liver disease (LD, 46%) at the Charlotte Maxeke (previously known as Johannesburg Hospital) Liver Disease Clinic and the same number from the anonymous volunteer blood donors (46%). There were 2(8%) haemophiliacs (Table 24). PCR was most efficient with 21/28 (75%) amplifiable in the core region. This was followed by the E1 (68%) region while PCR in both the NS4B (29%) and NS5B (29%) performed sub-optimally.

Table 24. A summary of the demographic and clinical data of the 28 specimens used for partial sequencing and analyses

						REGION AMPLIFIED		
Patient	AGE	GENDER	ETHNIC	CLINICAL	C	E1	NS4B	NS5B
1	59	M	U	BB	POS	POS	NA	NA
2	U	M	B	LD	POS	POS	NA	POS
3	40	M	B	LD	POS	POS	NA	POS
4	58	F	B	LD	POS	POS	NA	NA
5	51	F	B	LD	POS	POS	NA	PS
6	39	M	W	H	POS	POS	NA	PS
7	36	F	W	LD	POS	POS	NA	POS
8	67	M	B	LD	POS	POS	NA	POS
9	42	M	U	H	POS	POS	POS	POS
10	U	U	U	LD	POS	POS	POS	POS
11	61	U	B	LD	POS	POS	NA	POS
12	51	F	B	LD	POS	POS	NA	PS
13	55	F	B	BB	POS	POS	NA	POS
14	U	U	U	BB	POS	POS	NA	NA
15	57	F	B	BB	POS	POS	NA	NA
16	56	M	B	BB	POS	POS	POS	POS
17	43	M	B	BB	POS	POS	NA	POS
18	48	M	B	BB	POS	POS	NA	POS
19	51	F	B	BB	POS	POS	NA	NA
20	U	M	W	BB	POS	IS	IS	NA
21	50	M	U	BB	POS	IS	IS	NA
22	57	F	B	BB	IS	IS	POS	NA
23	63	F	U	BB	PS	IS	IS	NA
24	32	M	W	BB	PS	IS	IS	NA
25	64	F	U	LD	IS	IS	POS	POS
26	45	F	A	LD	IS	IS	POS	NA
27	48	F	B	LD	IS	IS	POS	POS
28	U	U	U	LD	IS	IS	POS	NA

U, unknown; M, male; F, female; B, Black; W, White; A, Asian; BB, blood bank; LD, liver disease; H, haemophiliac; POS, positive; IS, insufficient specimen; NA, not amplifiable; PS, partially sequenced.

6.3.2. Phylogenetic analyses

Figure 23 shows the phylogenetic clustering for each of the regions; (a) core, (b) E1, (c) NS4B, (d) NS5B and concatenated regions (e) core/NS5B and (f) E1/NS4B. All trees were rooted to the GBV-B virus. The monophyletic clustering of genotype 5a sequences were indicated by all regions sequenced and clustering occurred with significant genotype 5a branch bootstrap values for the NS5B region (Fig.23d) and concatenated regions, core/NS5B (Fig.23e) and E1/NS4B (Fig.23f). The E1 and the NS4B region (Fig.23b, c) showed bootstrap values >70% for some of the interior nodes within genotype 5a. However, this was not evident for the core (Fig.23a) and the NS5B (Fig.23d) sequences. Higher bootstrap scores were also noticed for the concatenated region E1/NS4B (Fig.23f) compared to the core/NS5B region (Fig. 23e).

Geographical clustering of the Belgian and SA strains was not evident in the core (Fig.23a) and NS5B regions (chapter 3 and Fig.23d). However, for the E1 region, some of our sequences did segregate with the French isolates (Fig.23b). The NS4B analyses also showed clustering of our sequences (for eg. 3158, 3673, 1997) with RSA isolates from Verbeeck's study (2006) and with the Belgium strains (Fig. 23c). For the NS4B region, one SA specimen 3898, clustered with the BE strains (Fig.23c) and for the E1 region, 2 blood bank specimens, SANBS 13 and 21 and a SA patient specimen 3975, clustered with the FR strains (Fig.23b). However, 3898 did not amplify in the E1 region and 3975, SANBS 13 and 21 did not amplify in the NS4B region so their phylogenies in these two regions could not be compared fully.

The concatenated analysis of the E1/NS4B region with the MCL parameter showed two significantly clustered genotype 5a groups, the first included the BE and some RSA isolates (Verbeeck et al., 2006) and the second, sequences from this study together with the reference sequences, AF064490 (isolate SA13, inoculated in chimpanzee, Bukh et al., 1998) and Y13184 (SA patient isolate EUH1480, Chamberlain et al., 1997; Fig.23f).

When the two NJ methods using the Kimura-2 and MCL parameters were compared, there were no significant differences in the clustering and bootstrap values of the respective trees. The sum of branch lengths between the two trees was also very similar, indicative that the pairwise calculations of genetic distance were similar for the two methods.

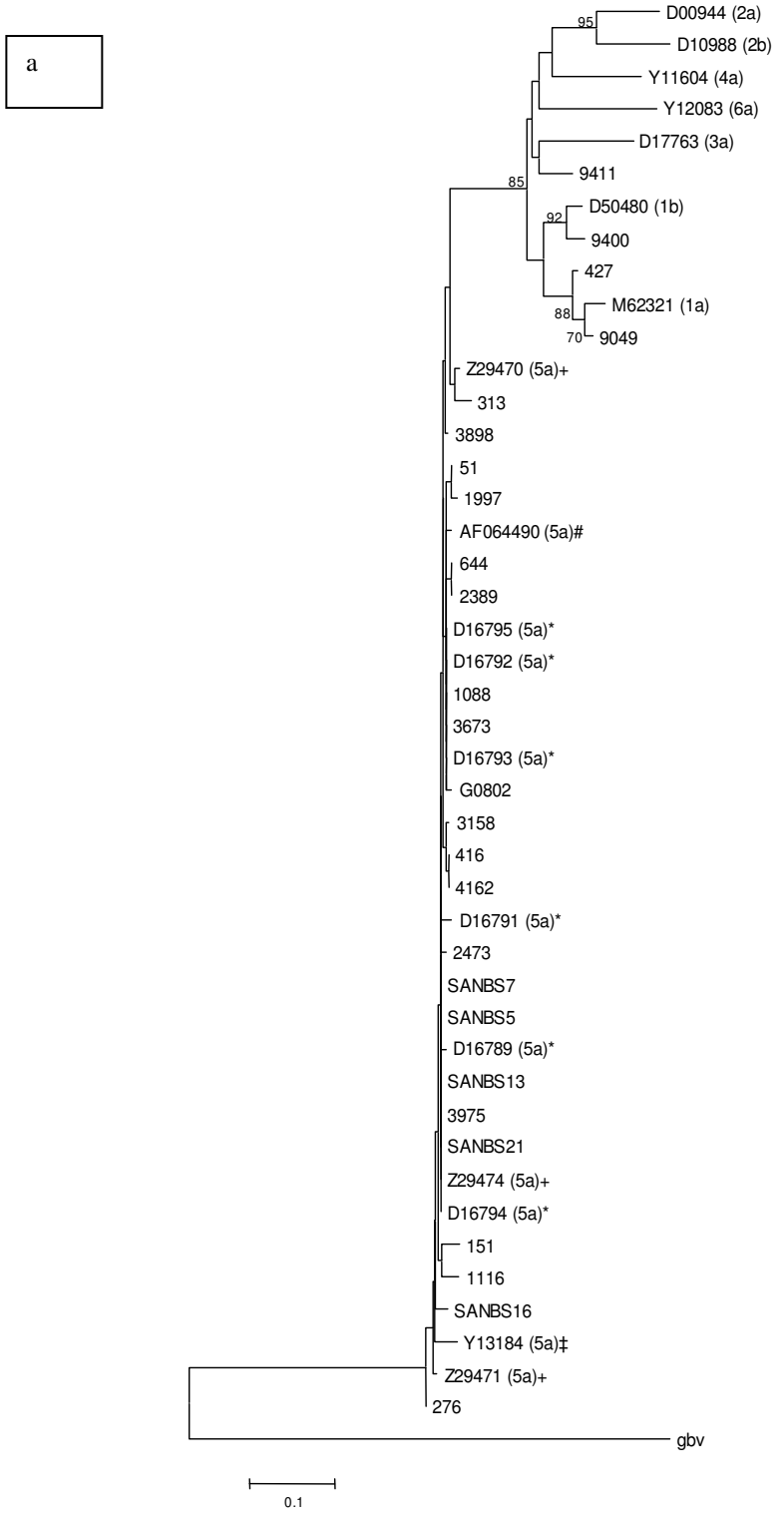
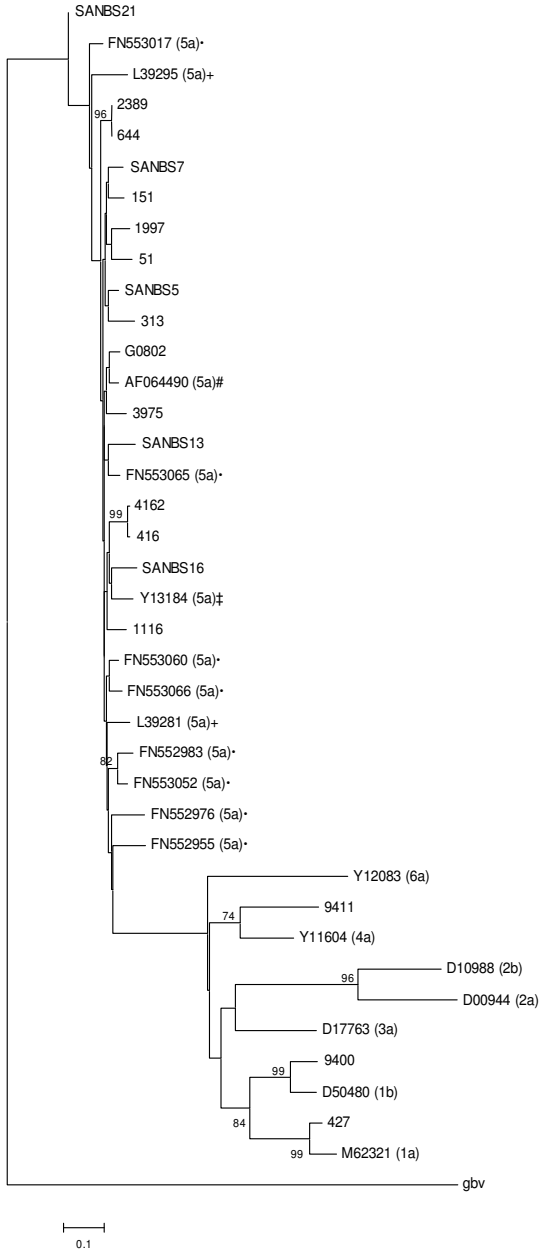


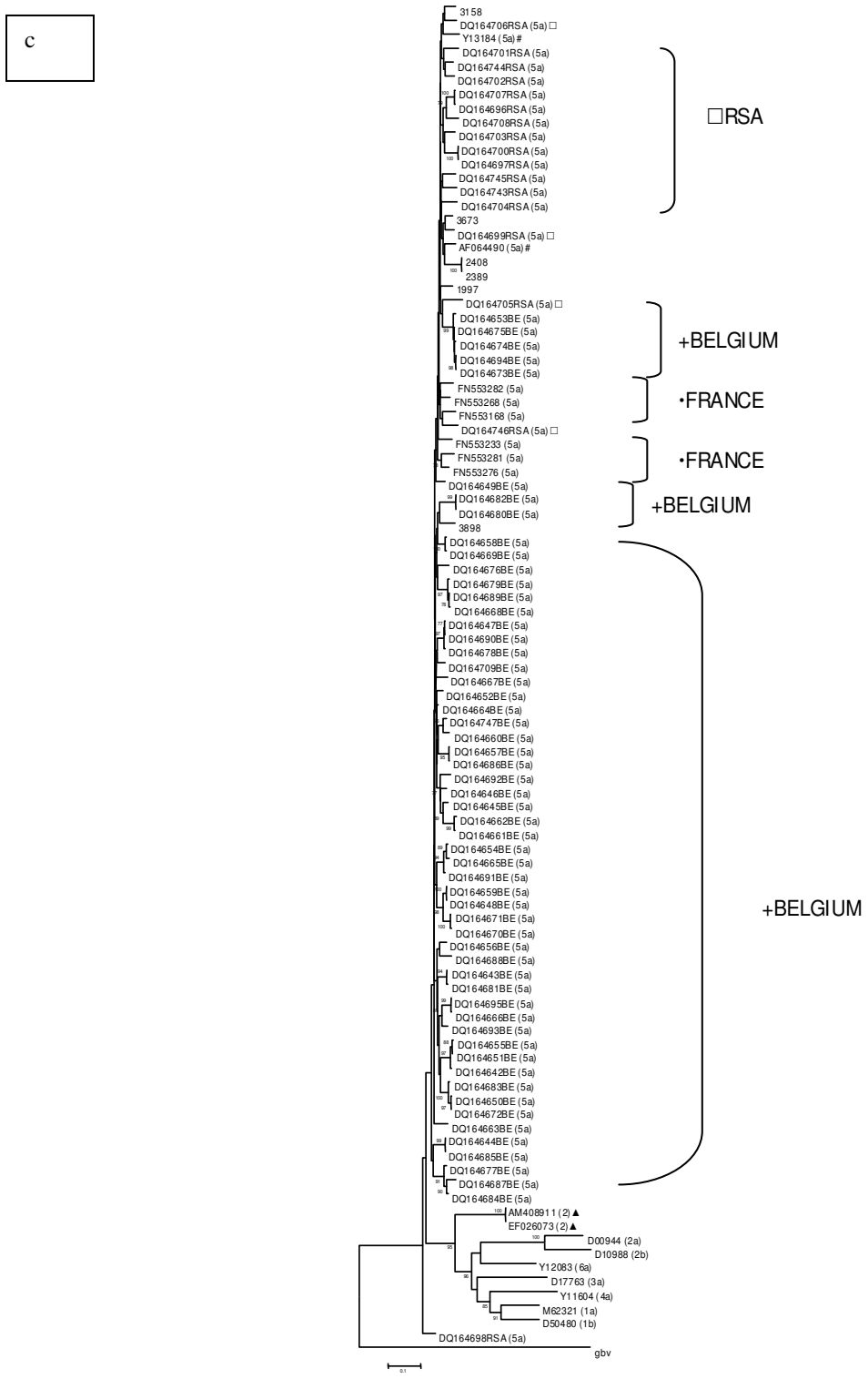
Fig.23. Neighbour-joining, rooted trees, using the Kimura-2 parameter for the South African samples. (a) Phylogeny of a 259bp fragment in the core region. The reference genotype 5a strains are denoted as follows: #SA13 isolate (Bukh et al., 1998); ‡ SA EUH1480 isolate (Chamberlain et al., 1997); +Belgium isolates (van Doorn et al.,1995); *5a clones (Ohno et al., 1994). The scale bar represents 0.1 substitutions/nucleotide site.

b



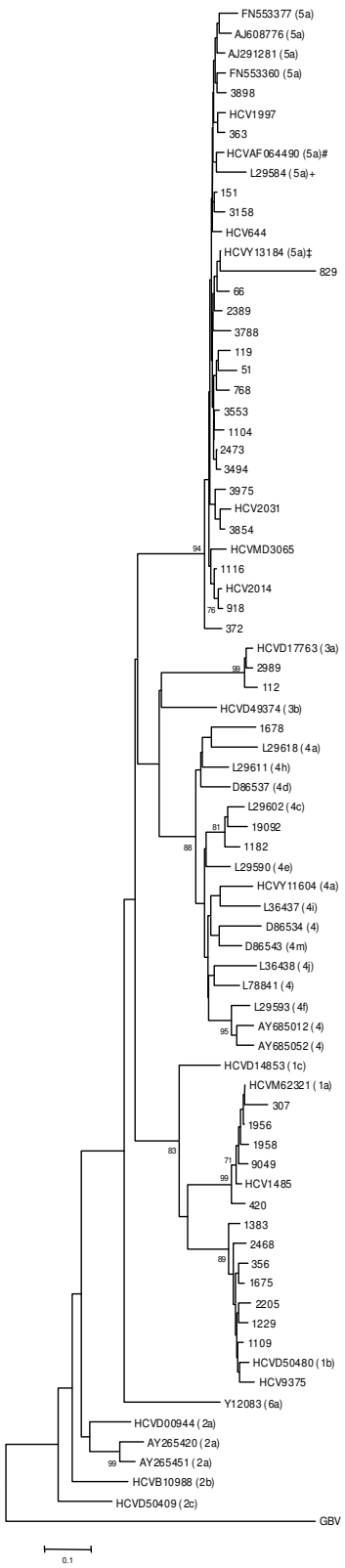
(Fig.23b) Phylogeny of a 237bp fragment in the E1 region. The reference genotype 5a strains are denoted as follows: #SA13 isolate (Bukh et al., 1998); ‡ SA EUH1480 isolate (Chamberlain et al., 1997); +Belgium isolates (van Doorn et al., 1995); •French isolates (Henquell et al., 2004). The scale bar represents 0.1 substitutions/nucleotide site.

c



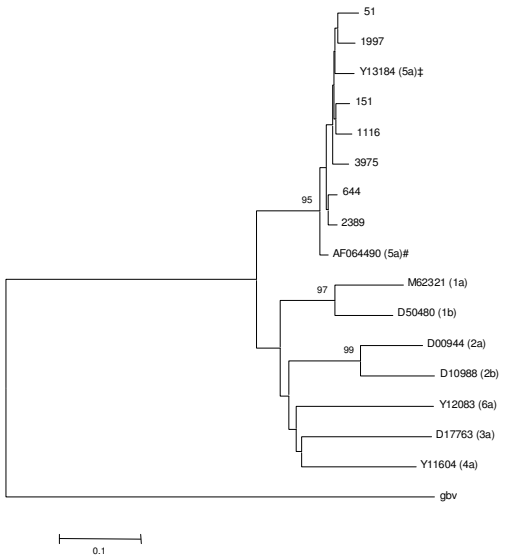
(Fig.23c) Phylogeny of a 584bp fragment in the NS4B region. The reference genotype 5a strains are denoted as follows: #SA13 isolate (Bukh et al., 1998); ‡ SA EUH1480 isolate (Chamberlain et al., 1997); +Belgium isolates and □RSA isolates (Verbeeck et al., 2006); •French isolates (Henquell et al., 2004); ▲recombinant 2/5 (Legrand-Abravanel et al., 2007) . The scale bar represents 0.1 substitutions/nucleotide site.

d



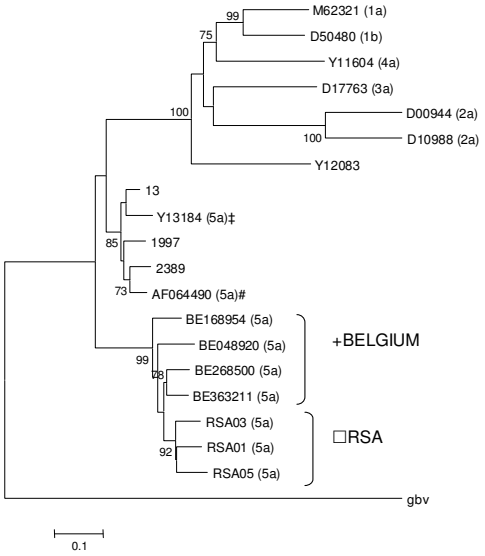
(Fig.23d) Phylogeny of a 208bp fragment in the NS5B region. The reference genotype 5a strains are denoted as follows: #SA13 isolate (Bukh et al., 1998); ‡ SA EUH1480 isolate (Chamberlain et al., 1997); +Belgium isolates (Stuyver et al., 1994); •French isolates (Henquell et al., 2004). The scale bar represents 0.1 substitutions/nucleotide site.

e



(Fig.23e) Phylogeny of the concatenated gene fragment core/NS5B (465bp). The reference genotype 5a strains are denoted as follows: #SA13 isolate (Bukh et al., 1998); ‡ SA EUH1480 isolate (Chamberlain et al., 1997). The scale bar represents 0.1 substitutions/nucleotide site.

f



(Fig.23f) Phylogeny of the concatenated gene fragment E1/NS4B (855bp). The reference genotype 5a strains are denoted as follows: #SA13 isolate (Bukh et al., 1998); ‡ SA EUH1480 isolate (Chamberlain et al., 1997);+Belgium isolates and □RSA isolates (Verbeeck et al., 2006). The scale bar represents 0.1 substitutions/nucleotide site.

6.3.3. Branching order and ML phylogenies

Output from FindModel is shown in Table 25 (the lower the log likelihood score, the better the model for use in ML phylogenetic analyses). The log likelihood (LnL) and the Akaike Information Criterion (AIC, Akaike, 1974) values are inversely related to each other. The Findmodel function indicated that the GTR model was the best model to apply to the dataset of the E1, NS4B and NS5B and the Transversion model for the core region (Table 25).

Table 25. A summary of the FindModel results as obtained from a web adaptation of the Modeltest script to find the best phylogenetic model for data in the core, E1, NS4B and NS5B regions

MODEL	CORE		E1		NS4B		NS5B	
	AIC	LnL	AIC	LnL	AIC	LnL	AIC	LnL
Jukes-Cantor (JC)	4108	-2054	13191	-6595	102999	-51499	31021	-15510
Felsenstein 1981 (F81)	2604	-1299	8242	-4118	24090	-12042	9946	-4970
Kimura-2 (K80)	2539	-1268	7982	-3990	22722	-11360	9589	-4793
Transversion (TM)	2507	-1246	7964	-3975	22684	-11333	9575	-4780
Hasegawa-Kishino-Yano (HKY)	2540	-1266	7967	-3979	22716	-11354	9599	-4795
Tamura-Nei (TrN)	2541	-1265	7965	-3977	22717	-11353	9559	-4774
General Time Reversible (GTR)	2509	-1246	7483	-3732	22683	-11333	9535	-4759

AIC, Akaike Information Criterion (AIC = -2(ln (likelihood)) + 2 K), where likelihood is the probability of the dataset given a model, K is the number of free parameters in the model, LnL, log likelihood score

The ML phylogenies were used to look at the order of branching of genotypes by means of rooted, slanted cladograms. Our results corroborate those of Pang et al., 2009 who showed that genotype 2 branches out first and genotypes 1 and 4 last. Genotype 5 has an intermediate position, after genotype 2 but before genotype 4 and 1 (Figs.24a-d). The NS5B region was analysed by both the NJ (MCL) and the ML methods with identical results. Genotype 2 branched first, followed by genotype 6, genotype 5, genotype 3 and then genotypes 1 and 4 (Fig.24a, b). The ML cladogram for the NS4B region showed that genotype 5 branched before genotype 6 and after genotype 2 (Fig.24c). The amino acid ML branching of Pang's study is similar to both our NS4B and E1 branching, with genotype 5 branching after genotype 2 (Fig.24c, d). The highly conserved core region showed a different branching pattern. Genotype 2 was still the first to branch out followed by genotype 4, 5, 6, 3 and 1 (data not shown).

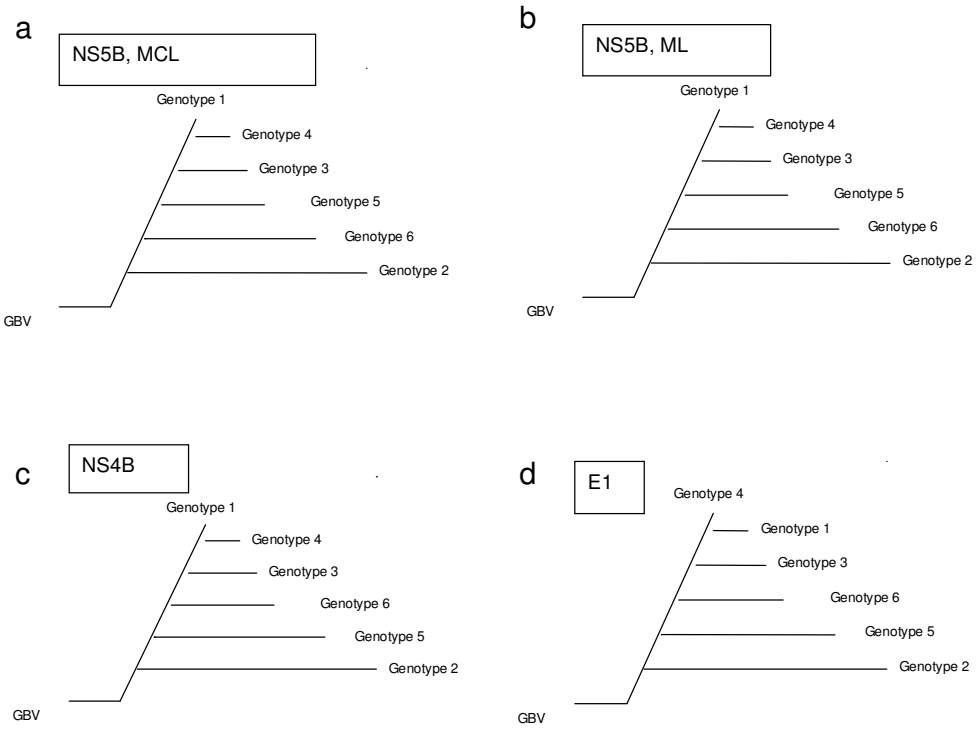


Fig.24. Representative branching orders on slanted cladograms, rooted to GBV for (a) NS5B (MCL), (b) NS5B (ML) and ML cladograms for (c) NS4B and (d) E1

6.3.4. Evolutionary distance of genotype 5a

The average evolutionary divergence per site between sequences from the different genotypes in the core, E1, NS4B and NS5B regions were 0.231, 0.392, 0.174 and 0.357, respectively. The rate of heterogeneity of the core and NS5B regions are comparable with rates reported by Pybus et al., 2009 (0.22 and 0.32, respectively). Table 26 summarises the mean evolutionary divergence per site of genotype 5a over the regions sequenced in this study. The heterogeneity of sequences in the core and NS5B regions (0.037 and 0.079, respectively) was less than that found in the E1 and NS4B (0.124 and 0.1 regions, respectively). Genotype 5a sequences in this study have a lower divergence in the E1 region than that reported for SA geno-

type 5a by Verbeeck et al., 2006 (0.16) However, the divergence rates for the NS4B region was similar in both studies (0.1 versus 0.12). The rate of heterogeneity for genotype 5a sequences in the NS5B region was found to be 0.079, higher than indicated by Ribeiro et al., 2009 (0.058, Table 26).

Table 26. Number of base substitutions/site calculated as an average over all sequences of genotype 5a for the sub-genomic regions sequenced and compared to published calculations.

	Nucleotide diversity			Divergence (years)
Core	0.037		152	
E1	0.124	0.16[a]	100	123[a]
NS4B	0.1	0.12[a]	153	143[a]
NS5B	0.079	0.058[b]	156	

[a] Verbeeck et al., 2006; [b] Ribeiro et al., 2009

6.3.5. Molecular Evolutionary Analyses

We used the readily available UPGMA parameter in the MCL model (MEGA 4.0) to estimate the divergence times of genotype 5a in the population from the most common recent ancestor (MCRA). When we rooted the genotype 5a sequences to the oldest genotype (genotype 2) as evident from the above section on branching order and used published evolutionary rates (Table 1), we had calculations of divergence ranging from 100-156 years (Fig.25a-d). The estimated time of the genotype 5a divergence as analysed per region sequenced were 152 years (core, Fig.25a), 100 years (E1, Fig.25b), 153 years (NS4B, Fig.25c) and 156 years (NS5B, Fig.25d).

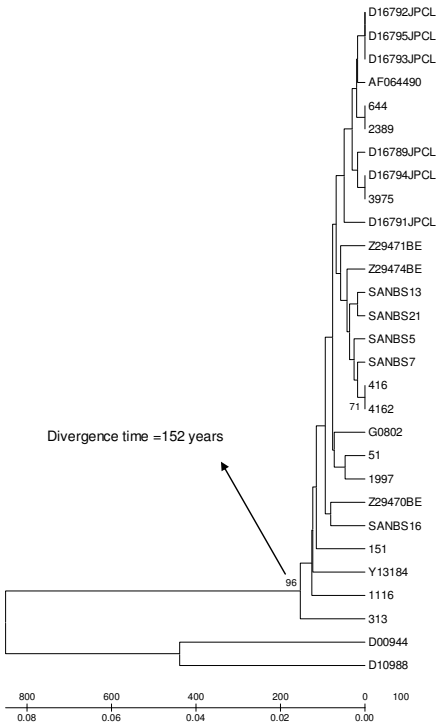


Fig.25a. Estimated divergence time as calculated by the calibrated molecular analyses, NJ, MCL, UPGMA parameter for the core region. The tree is rooted to genotype 2 reference sequences, D00944 and D10988. The scale bar represents 0.02 substitutions/nucleotide site/year.

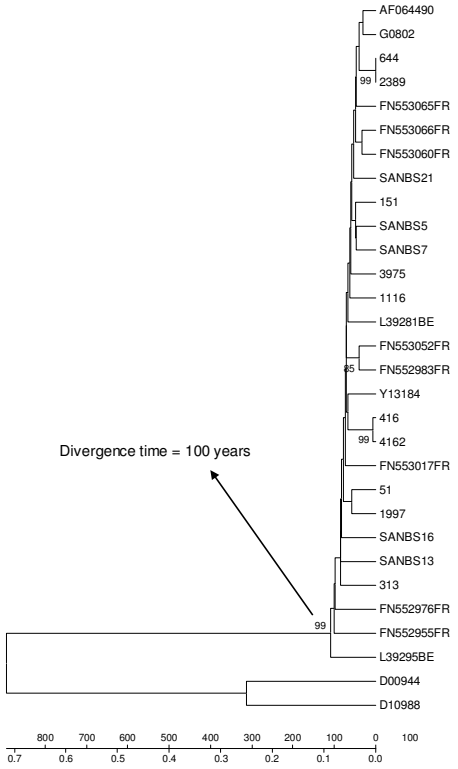


Fig.25b. Estimated divergence time as calculated by the calibrated molecular analyses, NJ, MCL, UPGMA parameter for the E1 region. The tree is rooted to genotype 2 reference sequences, D00944 and D10988. The scale bar represents 0.1 substitutions/nucleotide site/year.

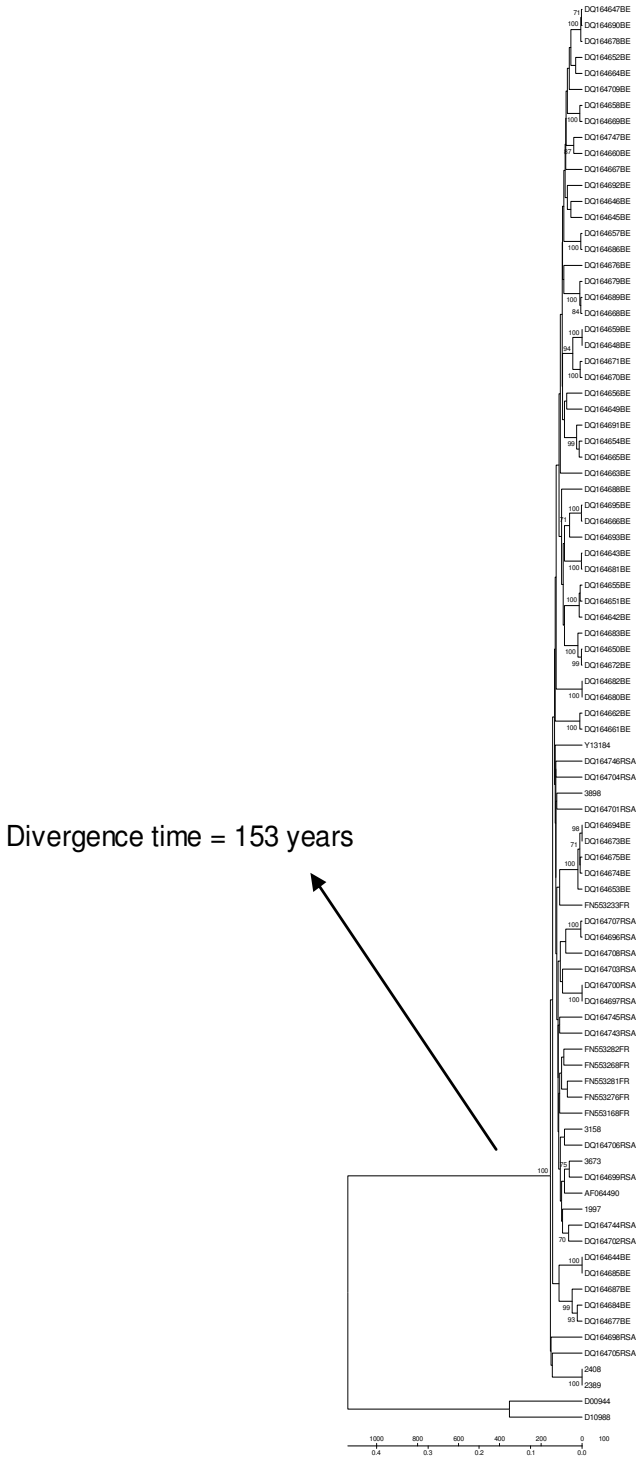


Fig.25c. Estimated divergence time as calculated by the calibrated molecular analyses, NJ, MCL, UPGMA parameter for the NS4B region. The tree is rooted to genotype 2 reference sequences, D00944 and D10988. The scale bar represents 0.1 substitutions/nucleotide site/year.

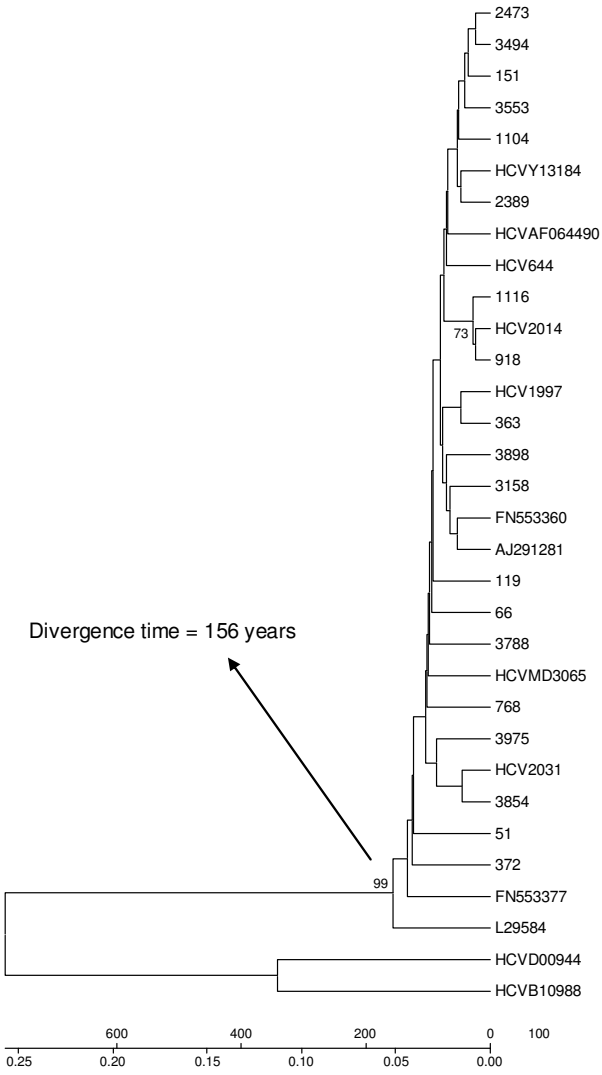


Fig.25d. Estimated divergence time as calculated by the calibrated molecular analyses, NJ, MCL, UPGMA parameter for the NS5B region .The tree is rooted to genotype 2 reference sequences, D00944 and D10988. The scale bar represents 0.05 substitutions/nucleotide site/year.

6.4. DISCUSSION

This study looked at the molecular characterization of HCV genotype 5a across four regions of the genome in specimens from patients with clinical disease or volunteer blood donors from SA. Although amplification of the core and E1 regions were successful, the NS4B and NS5B were poorly resolved by PCR and sequencing. As explained previously in chapter 3, one of the reasons is the poor specimen integrity and RNA degradation during transport and storage. The published primers used in

the amplification of the NS4B and NS5B regions may have been suboptimal as primer bias has been shown to amplify only a proportion of target sequences (Yao et al., 2005). The use of many combinations of genotype 5a-specific primers and/or degenerate primers may address this problem in future.

We performed phylogenetic analyses for all regions sequenced using the NJ distance method, with Kimura-2, MCL and UPGMA parameters. The MCL algorithm is a relatively new feature of the MEGA 4.0 program that uses the sum of log likelihood to produce distances of two neighbouring sequences simultaneously. It is incorporated in the software to provide better clustering and higher bootstrap scores for interior nodes of monophyletic clades (Tamura et al., 2007). It implements the Tamura-Nei substitution model, whereby substitution rates are considered different per site and nucleotide frequencies assumed unequal (Tamura and Nei, 1993). All the phylogenetic trees were rooted to the outgroup, GBV-B virus and bootstrap scores and clustering between different NJ methods were compared.

From our data, across the respective regions sequenced, we have found that there were no differences between the different trees using the Kimura-2 and MCL algorithms. All our tree analyses showed that genotype 5a sequences branched separately from the rest. The sequences within the genotype 5a clade were very closely related to each other, as indicated by the short branch lengths. Significant bootstrap scores (>70%) were only evident for the E1 and NS4B regions. This is to be expected since these two regions have a higher diversity or phylogenetic signal than that found in the core and NS5B (Table 26, 0.124 and 0.1 vs. 0.037 and 0.079 respectively). Also to be expected, these two regions did not show any geographical

clustering within the genotype 5a clade as was seen for the E1 and NS4B regions. This was exacerbated by the small sample numbers for the concatenated analyses and the lack of reference sequence data especially in the core/NS5B region.

The genetic distances among the Belgium, French and SA isolates for the E1 and NS4B regions ranged from 0.157 to 0.170 and 0.1 to 0.12, respectively, according to Verbeeck et al., 2006. The results of this study concur in the NS4B region (0.1) but divergence was lower (0.124) than reported in the E1 region. The NS5B region was more diverse (0.079) than Belgium and French isolates (0.058-0.061, Ribeiro et al., 2009).

The ML and NJ rooted phylogenies were compared and found to have identical tree topologies for all regions sequenced. Since the sub-genomic regions of HCV analysed are reliably grouped by the simpler and quick distance method, we conclude that our sequences are sufficiently conserved and do not require ML methods. Kumar et al., 2008 recommend ML for greater accuracy when dealing with distantly related sequences.

The branching orders of the six genotypes were taken from the cladograms of the ML phylogenies. Pang et al., 2009 showed evolutionary age (synonymous to the order of branching in an evolutionary tree) of genotypes corresponded to that of virological response to therapy. Although, they do point out that full genome data is required, the NS5B region concurred with that of the ML full genome nucleotide tree of Pang et al., 2009. Genotype 2 branched first followed by genotype 6. However, genotype 5 branches before genotype 3 suggesting that it is older. NS4B analysis

indicates that genotype 5 is even older than genotype 6 as it branches after genotype 2 but before genotype 6. This was also evident in the ML (amino acid, aa) tree of Pang et al., 2009. Genotypes 2 and 3 respond better to PEG-IFN and ribavirin therapy than genotypes 1 and 4 (Nguyen and Keeffe, 2005). Pang's study showed older genotypes respond better to therapy. The evolution of HCV is thought to be a result of several factors, including host immune selective pressure and the intrinsic ability of the virus to resist IFN action. The phylogeography of HCV genotypes and the spread between host populations with inherent immune differences could drive viral evolution (Pang et al., 2009). In a small prospective study on the responses to therapy of different genotypes in South African patients, we have shown that patients infected with genotype 5a did respond as well to therapy as those with genotype 3 (see Chapter 8). However, not many prospective studies on the responses to therapy for genotypes 5 and 6 have been conducted, but it is thought that these genotypes might respond between the two extremes (Antaki et al., 2008) and the branching of these genotypes, as shown by this study and Pang et al., 2009, has indicated this.

It has been estimated that genotype 2 was introduced in Guinea-Bissau, West Africa, in 1470 (~540 years ago, Markov et al., 2009), compared to the introduction of the younger genotype 1 which was introduced to the US, in the early 1940's (Tanaka et al., 2006) and to Japan, in the 1880's (Tanaka et al., 2002). Some studies estimate that HCV may have emerged thousands of years ago (Pybus et al., 2009) and not just hundreds (Markov et al., 2009). However, based on Pang's analyses, genotype 2 was used to root the UPGMA trees to access divergence times for genotype 5a. The inferred times of divergence of the two sequences of

genotype 2 used to root the tree ranged from 238 (E1) to 498 (core) years.

Although genotype 5a is predominantly found in SA, it is unlikely that genotype 5a subsequently spread from SA to Belgium by trade/travel routes. Verbeeck's coalescent approach indicated that genotype 5a was introduced into Belgium and SA at approximately the same time (137-143 years, Verbeeck et al., 2006). This co-introduction may have involved a re-introduction in to regional pockets of Africa. It is thought that the slave trade in the early 1800's, which transported slaves from west to other parts of Africa may have played a role in the dissemination of HCV (Markov et al., 2009; Verbeeck et al., 2006).

Estimated divergence times of genotype 5a date its MRCA approximately 150 years, earlier than that estimated by Verbeeck et al., 2006. However, estimates were based on the assumption that all specimens analysed were infected at the same time and the MCL approach assumes that evolutionary rates among the sites are equal. Sample numbers for each of the HCV regions were also very small and the molecular clock analysis relied on estimated evolutionary rates. However, divergence times were similar for the core, NS4B and NS5B regions while the E1 region estimate a MRCA 100 years later. Verbeeck's study, also predicted a more recent divergence (123 years) using the E1 region as compared to the NS4B region (143 years).

6.5. CONCLUSION

Genotype 5a is the most conserved genotype as it has only one subtype. From the sequencing of the subgenomic regions, we were able to determine the rate of heterogeneity among the genotype 5a sequences, with E1 being the most variable as

expected, then the NS4B and NS5B. The weakness in the study is the poor amplification in the NS4B and NS5B (also shown in chapter 3). Optimization of the RNA extraction and genotype 5a-specific primers needs to be implemented, so as to try many permutations to obtain better PCR and sequencing results. The possibility of a vast number of 5a quasispecies within a single individual may also hamper amplification with gene-specific primers. Cloning of variants and/or next-generation sequencing will be investigated in future studies.

This study has made it possible to compare phylogenies of our local 5a isolates to that of other countries where genotype 5a has been reported, in particular, Belgium and France. Evolutionary age of genotype 5a was inferred by the branching order of genotype 5a relative to other genotypes and divergence times of 100-150 years were estimated. This is the first time genotype 5a has been characterized in this country.

CHAPTER SEVEN

Epitope analysis on Hepatitis C Virus (HCV) genotypes, particularly genotype 5a, against South African Human Leukocyte Antigen (HLA) backgrounds

7.1. INTRODUCTION

As HCV is a relatively "new" virus, only identified in 1989 (Choo et al., 1989) and first cultured successfully in 2005 (Wakita et al., 2005), there is much that is unknown about HCV which has hindered the development of an effective vaccine. The challenges to a vaccine design against HCV include:

(a) The high mutability of the virus which varies within and between genotypes. Genotypes also partition geographically, and exist as quasispecies within the host.

(b) Host cell responses to infection are poorly defined and, in addition, inconsistent between infected individuals. CD4+ and CD8+ T-cell responses are also not cross-protective to heterologous genotypes (Schulze zur Wiesch et al., 2007) and, to date, there is no immunodominant epitope that is consistently found in all HCV-positive individuals (Klade et al., 2009).

(c) Humans are the only natural host of HCV, and suitable laboratory models have only been developed recently. The chimpanzee has been infected in the laboratory (Bukh, 2004), but studies on this model are expensive and limited by this and constraints of availability. More practical and plausible is the current development of a mouse model for viral pathogenesis studies (Ploss and Rice, 2009).

Epitope-based vaccines activate an immune response by presenting immunogenic peptides (viral genotype-specific) bound to major histocompatibility (MHC) molecules (host specific) to the T cell receptor. Class II proteins are presented to T helper cells by antigen presenting cells (APCs) with the aid of the CD4 co-receptor whereas Class I proteins are presented by infected target cell to cytotoxic T cells with the aid of the CD8 co-receptor. The T helper response is important in directing and activating the immune response, including the effectiveness of CD8+ T cells (Grakoui et al., 2003). A good vaccine must be capable of inducing and maintaining powerful CD4 and CD8 T-cell immunity in the greatest proportion of its target population.

Both HCV genotype and HLA allele frequency are distributed geographically. South Africa has a diverse and unique cultural heritage. Statistics South Africa's website (http://www.statssa.gov.za, 2010-12-08) estimates that of the mid-year population (N = 49. 99 million) the predominant ethnicities in the South African population are: African (79.4%, including Zulus), White (9.2%), Mixed race (8.8%) and Indian/ Asians (2.6%). Ethnicity (Wang et al., 2009) and HLA class I (Neumann-Haefelin et al., 2008) and class II alleles (Sarobe et al., 2006), appear to be associated with either HCV disease progression or sustained response to therapy (Satapathy et al., 2010, Tables 3,4). Most putative peptide-based vaccines and diagnostics are chosen based on the sequence of HCV genotype 1 epitopes and restriction by common Caucasian HLA alleles, and HLA-A*2 in particular (New allele Frequency Database, 2003, http://www.allefrequencies.net 2010-11-10). We have seen in chapter 3 that genotype 5a is predominant in patients with liver disease in SA.

The aim of the study was to describe the heterogeneity of HCV epitope sequences, together with local HLA-types and how this can affect vaccine efficacy in the ethnically diverse SA population.

7.1.1. Objectives:

(a) To examine the variation in selected published immunogenic epitopes within target antigens

(b) To predict their immunogenicity in the context of prevalent alleles in the South African target population

7.2. METHODS

7.2.1. Epitope Sequences

The literature was searched for known published Class I- and II-restricted immunogenic regions from the core to the NS5B which ranged in length from 9 bp (Wei et al. 2008) to 683 bp (Lang et. al, 2008). Seven class I- and six class II- epitopes were chosen for the analyses (Table 27) using the following criteria:

- Extensively studied immunogenic epitopes
- Published in the peer reviewed literature
- Known HLA restriction and
- Recommended for putative vaccines.

Table 27. Seven HCV Immunodominant HLA Class 1- and six HLA class II- restricted epitopes were chosen for the analysis in this study

CLASS I EPITOPES	SEQUENCE	RESTRIC-TION	REFERENCE
NS3 1073-1081	CINGVCWTV	A02	Wertheimer et al., 2003; Wedemeyer et al., 2009
NS3 1406-1415	KLVALGINAV	A02	Wei et al., 2008; Memarnejadian et al., 2009
NS3/NS4 1628-1637	GAVQNEITL	(?)	Frelin et al., 2003; Lang et al., 2008
NS4 1807-1816	LLFNILGGWV	A02	Cerny et al., 1995; Martin et al., 2008
NS4 1851-1859	ILAGYGAGV	A02	Wei et al., 2008
NS5B 2422-2433	MSYSWTGALVTP	B15	Wei et al., 2008
NS5B 2727-2735	GLQDCTMLV	A02	Wei et al., 2008
CLASS II			
Core 17-35	RRPQDVKFPGGGQIVGGVY*		Lamonaca et al., 1999
Core 21-40	DVKFPGGGQIVGGVYLLPRR		Lamonaca et al., 1999; Day et al., 2002; Wedemeyer et al., 2009
NS3 1248-1261	GYKVRVLNPSVAAT		Diepolder et al., 1997; Martin et al., 2008; Wedemeyer et al., 2009
NS4A 1781-1800	LPGNPAIASLMAFTASITSP		Martin et al., 2008
NS4A 1801-1820	LTTQNTLLFNILGGWVAAQL		Martin et al., 2008
NS5 2571-2590	KGGRKPARLIVFPDLGVRVC		Day et al., 2002; Schulze zur Weisch et al, 2005; Martin et al., 2008; Klade et al., 2009
NS5 2661-2680	QCCDLDPQARVAIKSLTERL		Day et al., 2002; Schulze zur Weisch, 2005

* indicates promiscuous epitopes that would bind to more than one DR-allele

Alignments of representative reference sequences were obtained over these immunogenic regions using sequence data from each of the genotypes with the aid of pre-aligned and updated amino acid sequence data from the International Nucleotide Sequence Database Collaboration (INSDC; Shin et al. 2008). The percentage of the total number of sequences available per epitope region varies greatly by genotype, ranging from 4-24% (genotypes 4 and 5) to 54-82% for genotype 1 (subtypes 1a and 1b).

Because there were, comparatively, few sequences of genotype 5a on the international databases, we included our own sequence data, in the alignments of the core,

NS4 and NS5B regions of genotype 5a wherever possible. Care was taken to ensure that all our own data, as well as data used from public databases, corresponded to one sequence per subject.

BioEdit, version 7.0 software (Hall, 1997), was used to align all the amino acid sequences. The consensus sequence of immunogenic regions, for each of the genotypes, was obtained using the Web based software, WebLogo, version 2.8.2 (http://weblogo.berkeley.edu/logo.cg; 2008-09-08). Numbering was according to Choo et al., 1991. The WebLogo program produces a consensus of input sequences consisting of stacks of letters, each representing a column in the sequence alignment. The height of each letter within the stack is proportional to the relative frequency of the representative amino acid at that position in the sequence (Crooks et al., 2004). The Weblogo software incorporates a “small sample number” correction, to correct for potential bias.

The conservation of epitopes was calculated as the percentage of polymorphic sites over the epitope length when compared to the HCV consensus sequence. The latter was determined by taking the most common amino acid from each site of the respective genotype consensus sequences in the 7 weblogos (genotypes 1a, 1b, 2, 3, 4, 5 and 6). A minimal class I-restricted epitope length of 9 nucleotides was used for all class I-restricted epitopes. Class II- restricted epitopes are longer with overlapping regions, thus, the number of amino acids varied per sequence. The statistical analysis was performed using the analysis of variance (ANOVA) tests of significance in the Statistica software, version 9.1.

7.2.2. South African MHC probabilities

The most common HLA alleles in the SA population were determined in order for the binding predictions to have relevance to the local population. HLA information regarding the different SA ethnic groups is not readily available in the literature, and this data will serve as a useful resource for future studies. Relevant studies (Table 28) were identified by searching the literature using PubMed, (http://www.ncbi.nlm.nih.gov/sites/entrez; 2009-11-18) and this information was supplemented from relevant ethnic allele frequencies stored in the MHC database accessible from the New allele Frequency Database, http://www.allelefrequencies.net (2010-11-30).

The most common allele frequencies were tabulated for each of the major ethnic groups in South Africa (MS Excel). Information was gathered for the following Class I alleles: HLA-A, -B and Class II HLA-DR alleles. Ethnic groups examined included Blacks (B), Caucasians (Ca), Coloureds (CC) and Indians (I).

Table 28. A summary of South African studies collated for the HLA/ epitope analysis

studies	HLA	Ethnic groups[a]	Study population size	Reference
1	A, B, Cw	B	295	Kiepiela et al., 2004
2	A, B, Cw, DR	B	2366	Cassim et al., 1994
3	A	B	198	
		Ca	100	Paximadis M., unpublished
	B	B	195	
		Ca	98	
	Cw	B	183	
		Ca	99	
	DRB1	B	194	
		Ca	95	
4	A, B, C	B	1027	
		CC	3716	Du Toit et al., 1988
		Ca	1059	
	DRB1	B	322	
		CC	549	
		Ca	365	
5	DRB1	CC	624	Klemp et al., 1988
6	DRB	B	117	Lombard et al., 2006
7	A, B, Cw	I	51	Hammond and Moshal., 1980
8	A, B, Cw	B	756	Hammond et al., 1980
9	A, B	B	41	Coetzee, 2006
10	A, B	B	100	Middleton et al., 2000
11	DRB	B	138	Rudwaleit et al., 1995
12	A, B, Cw	B, CC, Ca, I	13	Venter et al., 2003
13	A, DRB	B	448	Alkharsah et al., 2007

[a]B, Black; CC, Coloured; Ca, Caucasian; I, Indian

7.2.3 Immunogenicities

It was evident from a previous study that there is a good correlation between immunogenicity and MHC class I binding affinity (Sette et al., 1994). Therefore we used two well-documented programs to determine binding predictions for the published Class I and II epitopes in the context of SA host alleles. Only HLA alleles with high frequencies were tabulated with scores.

7.2.3.1. Immune Epitope database project (IEDB)

The IEDB web based resources, version 2.0 (http://tools.immuneepitope.org/main/jsp/menu.jsp; 2009-09-09), is a manually curated database of experimentally characterized immune epitopes. The "Peptide Binding to MHC Class I molecules" resource takes submitted proteins and splits them into all possible peptides and then determines their binding to each of the MHC alleles selected according to the artificial neural network (ANN) algorithm (Nielsen et al., 2003), our chosen prediction method. This method accounts for the position of the amino acid and the probability of adjacent amino acids competing for a space in the MHC pocket (Nielsen et al., 2003). Predicted binding efficiencies (in units of IC_{50} nM) were calculated for each of the possible HLA-class I epitopes. IC_{50} values <50 nM are considered high affinity, <500 nM intermediate affinity and <5000 nM low affinity (http://tools.immuneepitope.org/main/jsp/menu.jsp). MHC alleles prevalent in one or other of the SA population groups were of interest for the purposes of this study. Other matrix-based methods, like ProPred 1 (http://www.imtech.res.in/raghava/propredl/index.html, 2010-11-30) and SYFPEITHI (Rammensee et al., 1999) were used to compare binding efficiencies. However, for brevity, only scores for IEDB are shown.

7.2.3.2. ProPred MHC Class II binding prediction Server

Propred II (http://www.imtech.res.in/raghava/propred/index.html, 2010-10-20, Singh and Raghava, 2001) was used to predict binding of HLA-class II epitopes. This tool uses a linear prediction model based on values stored in allele specific coefficient tables, or quantitative matrices, generated according to the properties and position of the different amino acids in the epitope. The program is useful in locating promiscuous, versus allele specific, binding regions in a query peptide sequence. A stringency threshold between 1-10% can be chosen to facilitate high (no false positives) or low (no false negatives) stringency cut-off values and a value of 10% was used in all cases. Note that interpretation of the scoring is quite different to that of the IEDB program as a high percentage score is indicative of good binding between the relevant peptide and the specific HLA class I or II allele. The score is a percentage of the highest score possible by any peptide to bind with the given allele. The score is, therefore, a reflection of the binding characteristics of the query allele. There is no clear cut off as with IEDB ANN scoring, and actual percentages cannot necessarily be compared. For this reason the highest stringency was used to reduce to a minimum the false positives making all binding significant.

Table 29. The conservation of 7 class I-restricted and 6 class II-restricted epitopes (as percentages) across and within HCV genotypes 1-6 as compared to the consensus sequence

		HCV GENOTYPES										
CLASS I EPITOPE Consensus Epitope sequence		**1**	**2**	**3**	**4**	**5**	**6**	Mean across genotypes	MIN.	MAX.	SD	p-value
NS3 1073	CINGVMWTV	78	67	67	67	78	67	70.67	67	78	5.680	0.3062
NS3 1406	LTSLGLNAV	67	56	67	78	67	56	65.17	56	78	8.280	0.1645
NS3 1628	GAVQNEVTL	89	67	56	78	89	44	70.50	44	89	18.251	0.5134
NS4 1807	LLFNILGGW	100	78	78	100	100	78	**89.00**	78	100	12.049	0.6513
NS4 1851	ILAGYGAGV	89	67	89	78	89	67	79.83	67	89	10.815	0.2231
NS5B 2422	MSYSWTGAL	89	89	89	100	89	67	**87.17**	67	100	10.815	0.406
NS5B2727	GLRDCTMLV	78	56	44	78	78	33	61.17	33	78	19.823	0.4142
	Mean within genotypes	**84.29**	68.57	70.00	82.71	**84.29**	58.86					
CLASS II EPITOPE Consensus Epitope sequence		1	**2**	**3**	**4**	**5**	**6**	Mean across genotypes	MIN.	MAX.	SD	p-value
CORE 17	RRPQDVKFPGGGQIVGGVYLLPRR	100	96	66	96	96	96	**91.67**	67	78	5.680	0.3062
NS3 1248	GYKVLVLNPSVAAT	100	93	93	100	100	93	**96.50**	93	100	3.834	0.32
NS3 1781	LPGNPAVASLMAFTAAVTSP	85	80	95	85	90	65	**83.33**	65	95	10.327	0.4142
NS4 1801	LTTSQTLLFNILGGWVASQL	85	65	80	90	85	70	79.17	65	90	9.703	0.962
NS5B 2571	KGGRKPARLIVYPDLGVRVC	80	80	90	95	95	80	**86.67**	80	95	7.527	0.2231
NS5B 2661	QCCDLEPEARVAIKSLTERL	85	55	70	80	60	50	66.67	50	85	14.023	0.4159
	Mean within genotypes	**89.17**	78.17	**82.33**	**91.00**	**87.67**	75.67					

7.3. RESULTS

7.3.1 Degree of epitope conservation

The Weblogo consensus alignment for each genotype was generated for each of the chosen published class I (7 epitopes) and class II (6 epitopes) regions. From the alignments, the consensus and percentage conservation of each epitope within each genotype was calculated (Table 29). The variation between the epitopes for each genotype was not statistically significant, however, the class I epitopes NS4B$^{1807-1816}$, NS5B$^{2422-2433}$ and five of the six class II epitopes (all except NS5B$^{2661-2680}$) had the highest average conservation (>80%; Table 29). Published class II-restricted epitopes were in general better conserved than the class I epitopes and well conserved across genotypes, as well as within genotypes.

Genotype 1 is the most conserved genotype even though more sequences were examined. Genotype 4 shows a high conservation but had the smallest sample number of sequences so this may not be a true reflection. Genotype 6 shows the most variability which is a reflection of the fact that this genotype has the most number of subtypes (22), 6a-6v (Noppornpanth et al., 2008). Notably two of the class I epitopes, NS3$^{1628-1637}$ (44%) and NS5B$^{2727-2735}$ (33%) are very poorly conserved in genotype 6.

7.3.2. South African HLA population group characteristics

Figures 26 a, b, c show the major HLA frequencies for the South African population groups.

The most common HLA-A alleles in the South African Black population were in descending order: **A30** (24.36%), **A68** (17.33%), **A02** (14,70%), **A29** (12.74%), **A23** (12.20%), A31 (12.00%) and A03 (9.76%, Fig.26a).

	A01	A02	A03	A11	A23	A24	A26	A29	A30	A31	A32	A33	A34	A68
B	4.48	14.7	9.76	1	12.2	2.2	2.33	12.7	24.4	12	1.3	4.3	6.1	17.3
CC	8.5	17	7.7	6.5	5.9	9.9	4.5	4.4	9.6	1.1	3.7	3.3	2.6	NA
C	17.1	26.5	13.5	5.95	1.8	8.3	2.1	5.05	3.3	1.85	3.35	1.25	NA	2.5
I	17	17	5	18	NA	16	2	1	1	7	1	7	NA	8
Mean in the population	11.8	18.8	8.99	7.86	6.63	9.1	2.73	5.8	9.57	5.49	2.34	3.96	4.35	9.28

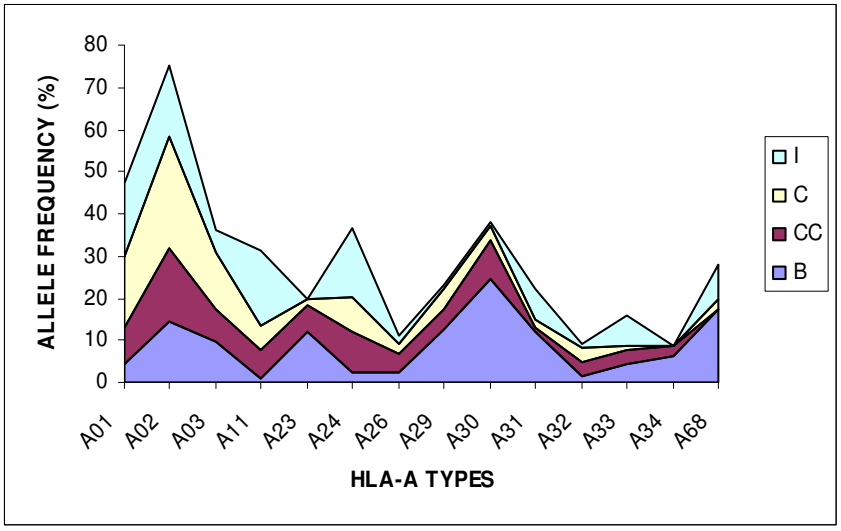


Fig.26a. Collated data of HLA-A studies in South Africa indicating the frequency of common alleles in the main 4 ethnic groups in the country (B, Black; CC, Coloured; C, Caucasian; I, Indian)

	B07	B08	B13	B14	B15	B18	B27	B35	B42	B44	B45	B51	B52	B53	B57	B58
B	9.6	11.1	4.03	5.02	19.3	4.22	0.2	3.33	20.5	11	9.5	NA	NA	3.7	7	23
CC	10.7	4.8	2	3.4	NA	5	2.3	5.8	3.1	8.8	2.7	2.4	2	1.3	4.3	8.6
C	15.3	13	1	2.5	11	5.1	4.5	7.5	NA	12	NA	4.6	1	NA	4	2
I	4	7	6	NA	13	3	1	9	NA	1	NA	9	8	NA	10	2
Mean in the population	9.9	8.99	3.26	3.64	14.4	4.33	2	6.41	11.8	8.2	6.1	5.33	3.67	2.5	6.33	8.9

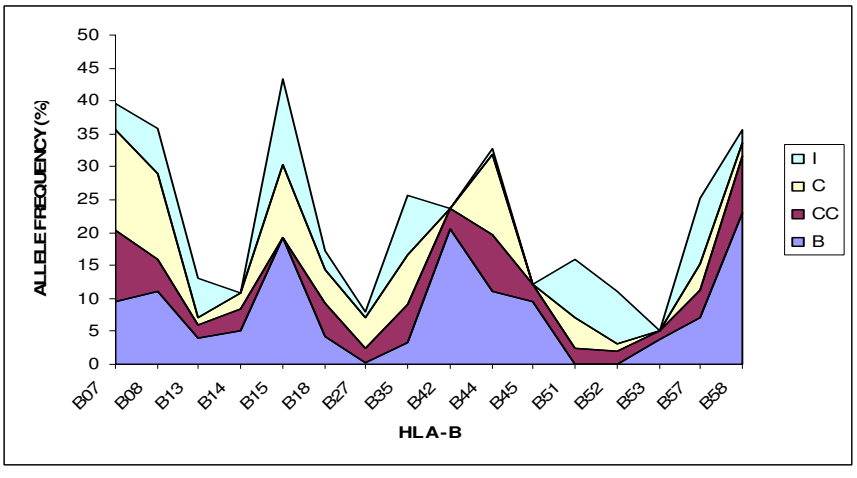


Fig.26b. Collated data of HLA-B studies in South Africa indicating the frequency of common alleles in the main 4 ethnic groups in the country(B, Black; CC, Coloured; C, Caucasian; I, Indian)

B58 (23%), **B42** (21%) and **B15** (19.3%) have the highest frequency in the Black population. **B07 (9.6%)**, **B08 (11.1%), B44 (11%)** and **B45 (9.5)** have intermediate frequencies (Fig.26b).

DRB 1*	*01:01	*03:01	*04:01	*07:01	*08:01	*09:01	*10:01	*11:01	*12:01	*13:01	*13:02	*13:03	*14:01	*15:01
B	6.5	22.5	5.79	7.85	5.6	1.63	1.95	20.8	2.23	18	4.35	1.8	1.3	10
C	7.4	12.2	11.2	9.5	3.2	2.1	0.8	5.8	1	6.3	3.7	1.6	2.6	11.2
CC	8.35	15	16.1	19.2	3.8	1.36	1.12	NA	NA	NA	NA	NA	NA	NA
mean in the population	7.42	16.6	11	12.2	4.2	1.7	1.29	13.3	1.61	12.2	4.03	1.7	1.95	10.6

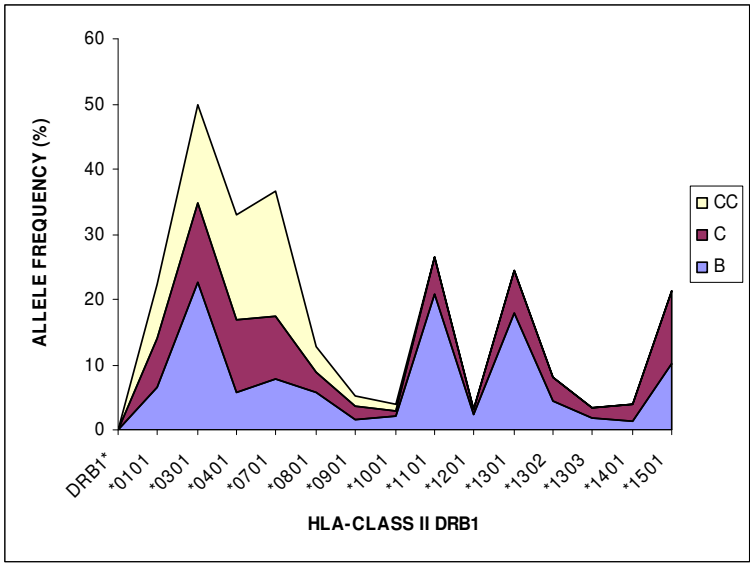


Fig.26c. Collated data of HLA-Class II studies in South Africa indicating the frequency of common alleles in the main 3 ethnic groups in the country (B, Black; CC, Coloured; C, Caucasian)

DRB1*03:01 (22.54%), **DRB1*11:01** (20.75%), and **DRB1*13:01** (18.04%) are the predominant Class II epitopes in the Black population (Fig. 26c).

7.3.3. Epitope Binding Prediction

The predicted binding values of epitope variants of HCV subtype 1a, 1b and genotype 5a to as many of our local HLA class I and II alleles were generated (Table 30 and 31) and used in population coverage estimations.

7.3.3.1. Class I alleles

The HLA Class I epitopes NS3$^{1073-1081}$, NS3$^{1406-1415}$, NS4B$^{1807-1816}$, NS4B$^{1851-1859}$and NS5B$^{2727-2735}$ have all been reported to be HLA-A2 restricted (Table 27). Predicted binding scores of all Class I epitopes and their variants with the relevant HLA-A and –B alleles are shown in Table 30.

HLA-A and –B Class 1 restricted binding

Binding predictions were performed for all the major HLA alleles in the SA population. The alleles were grouped according to supertypes (Sidney et al., 2008). Positive binding scores were, for the most part, in agreement for the same given allele between the IEDB and Propred I software. The notable exception is the allele HLA-B*27:05 which the more reliable IEDB ANN predicts will not bind to any epitope used in this study (Table 30).

Four of the five published HLA-A2 restricted epitopes bound to the A*02 allele over all variants despite the fact that only genotype 1a and/or 1b sequences were used in the original experiments. As most of the epitopes are A*02-restricted, we have seen good binding affinities for alleles A*02:01 and A*68:02 (both of the supertype A02) to NS3$^{1073-1081}$ and NS4B$^{1851-1859}$ in particular (Table 30). Binding scores for relevant alleles from supertype A03 showed very low binding for the most part and only scores for A*68:01 are shown. The peptide, NS4B$^{1807-1816}$, had predicted binding scores >500 IC_{50}nM for all except the genotype 3a sequence in the context of HLA-A*23:01 (Table 30) and scores of <1000 IC_{50}nM for the alleles B*57:01 and B*58:01 (supertype B58).

The published B*15-restricted NS5B$^{2422-2433}$ epitope is the only epitope to have

shown high affinity across supertypes. This epitope bound strongly <200 IC_{50}nM to B*15:03 (B27 supertype), to B*58:01 (B58 supertype), to B*35:01 (B07 supertype) as well as A*68:02 (Table 30).

Class I cross reactivity

There was a limited amount of cross reactivity with the rest of the HLA-A alleles. While the A*02:01 allele was predicted to bind with one of the $NS3^{1406-1415}$ genotype 1a variants (KLVALGINA and the genotype 1b epitope (KLSGLGLNA), it did not cross-react with the genotype 5a variants.

Allele A*30:01, which falls into supertype 24, showed high affinity to the genotypes 4 and 5 variants of $NS5B^{2727-2735}$ but only intermediate to poor binding to the genotype 1 equivalents. The broad A24 allele cross- reacted with the genotype 5a A*02-restricted variant CINGVMWTL of $NS3^{1073-1081}$, $NS3^{1628-1637}$ and the HLA-B restricted Class I epitopes $NS5B^{2422-2433}$ when ProPred 1 was used.

Binding was consistent across the supertype when affinity was good, for example, CINGVCWTV and variants of $NS3^{1073-1081}$ to supertype A02. However, many of the epitopes in this study were intermediate to poor binders and did not bind all alleles in a supertype (Table 30).

Table 30. Binding affinity scores of epitopes in the NS3-NS5 regions to common HLA- class I A and B supertypes, using IEDB. The shaded regions indicate low binding scores (<1000 IC_{50}nM), suggestive of high to intermediate binding affinity.

GENE	EPITOPE SEQUENCE	GENOTYPE OF EPITOPE	Class A- Alleles							Class B- Alleles					
		Supertypes	A01	A02		A03	A24		A01A24	B07		B58		B27	
		Allele type	A*0101	A*0201	A*6802	A*6801	A*2301	A*3001	A*2902	B*5301	B*3501	B*5701	B*5801[a]	B*1503	B*2705
NS3 (A*02)	CINGVCWTV	1a	17802	67	61	21126	14908	15501	12611	25523	20927	19827	13679	19257	23485
1073-1081	CVNGVCWTV	1b	16997	110	20	16511	12228	13122	11766	13382	15696	18288	12132	20367	23007
	SISGVLWTV	2a variant	18961	11	16	22007	21483	11417	11417	29702	22186	18590	15055	15691	20667
	TVGGVMWTV	3a	19940	64	8	26549	12677	14750	9776	24623	21877	16182	18054	26500	24303
	AVNGVMWTV	4a variant	17734	23	14	28410	24001	4015	12036	20595	20258	17093	12996	13641	18882
	CINGVLWTV	5a	15172	26	39	25058	17548	13613	13865	15854	21854	18628	11203	17516	21090
	CINGVMWTL	5a variant	17922	140	101	17870	10449	14413	11435	11237	13165	2239	13165	13572	19956
NS3 (A*02)	KLVALGINA	1a	22719	273	15048	42998	32261	1830	18800	37253	25216	23529	20557	4839	19019
1406-1415	KLSGLGLNA	1b	19133	475	21824	44012	33559	2557	13152	37083	27147	23891	19220	8973	18099
	QLTSLGLNA	4a	20013	7051	15292	38274	33674	12859	12517	37244	24440	22168	26218	7165	19904
	QLTGLGINA	5a	22408	6972	7419	38144	34672	13389	17488	36968	23541	25569	22283	15466	20054
	QLTSQGVNA	5a variant	22577	10896	9953	36070	33949	13067	22474	36870	23459	22528	26086	5846	22375
NS3 (?)	GAVQNEITL	1a, 5a	24077	18730	22906	33028	22866	16850	23641	9400	1905	15070	8640	6613	24901
1628-1637	GAVQNEVIL	1b	23896	20431	22981	32178	22846	18368	24846	13885	5567	18307	11139	5564	24356
	GAVQNEVTL	1b variant	24178	19785	21172	30394	18680	24905	24842	9655	2542	14387	9230	4550	24196
	GPVQNEICL	3a	23986	24531	27309	42984	29552	33075	27303	1993	1137	25716	22284	15801	26387
	GSVQNEVTL	4a	23809	21741	22518	32183	15851	22214	23896	24775	10011	11152	5998	2869	22341
NS4B (A*02)	LLFNILGGWV	1a, 1b, 4, 5a	22942	14359	17095	19708	18086	17906	9175	17154	19854	956	962	5918	23118
1807-1816	MFFNILGGWV	3a	24613	23482	19706	9909	343	15640	1707	8151	11817	10769	1251	13832	26621
NS4B (A*02)	ILAGYGAGV	1a, 1b, 5a	20500	15	530	29258	30882	15492	10120	37213	21134	22934	20702	3735	20143
1851-1859	ILAGYGTGV	5a variant	20351	18	193	27721	32028	17493	12563	36657	21994	23555	20603	2196	19849
NS5B (B*15)	MSYSWTGAL	1a, 1b, 4	12612	1522	24	3520	2924	2372	5457	8456	50	10166	523	80	16876
2422-2433	MSYTWTGAL	5a	12133	2640	22	5641	8602	2141	7606	9150	58	10680	787	144	17267
	YTWTGALIT	5a	15779	3000	13286	24240	33166	13737	1561	27619	3920	22480	17360	6553	18765
NS5B (A*02)	GLQDCTMLV	1a	18371	8	5733	37711	11972	13187	6275	35681	27015	25282	22002	10687	17601
2727-2735	KLQDCTMLV	1b	17735	7	3878	39468	6160	2071	9527	35038	26776	23310	18296	3587	16634
	KLRDCTLLV	5a	19744	13	14912	43573	15150	10	5150	36627	27145	21481	20362	1720	18071
	ALRDCTMLV	4a	19976	19	4673	40013	19836	29	9982	36740	26302	24190	22343	1206	20027

7.3.3.2. Class II alleles

The longer Class II epitopes were placed into the ProPred II program which found overlapping regions within the epitope. The output list contained possible peptides, all nine amino acids long. From these we tabulated all possible binders, together with their percentage of binding when compared with the log value of the best possible peptide for each allele (Table 31). For example, **CORE$^{17-40}$** returned two sequences VYLLPRRGP and VGGVYLLPR which scored similarly for alleles HLA-DRB1*03:01 and HLA-DRB1*15:01 (Table 31), but the former peptide, VYLLPRRGP, had a much higher binding score (48%) than its flanking sequence VGGVYLLPR (10%) in the context of DRB1*13:01. No predicted class II epitopes were found in the first 14 amino acids of **CORE$^{17-40}$**. Although **CORE$^{17-40}$** was the second most conserved epitope (Table 29), it was not predicted to bind with HLA-DRB1*01:01, HLA-DRB1*01:02 or HLA-DRB1*04:01 and only VGGVYLLPR was predicted to bind with HLA-DRB1*07:01 (9%, Table 31).

The most consistent class II binder was also the best conserved epitope, **NS3$^{1248-1261}$**(Table 29), specifically the region 1252-1260 **LVLNPSVAA**, which bound all eight of the alleles tested. The least promiscuous allele was HLA-DRB1*04:01 which only bound the **NS3$^{1248-1261}$** peptides. The most promiscuous allele of those tested was HLA-DRB1*15:01 as a positive binding score was predicted with all but five of the 18 peptides output by the program (Table 31).

The highest binding score (60%) was obtained from **NS5B$^{2571-2590}$** for allele HLA-DRB1*15:01. This particular sequence is conserved for genotypes 3 and 5.

Table 31. Percentage binding scores of epitopes in the core, NS3 and NS5 regions to common DRB1* alleles, by ProPred II

Region	Sequence	Genotype specificity	DRB1*010 1	*0102	*0301	*0401	*0701	*1101	*1301	*1501
Core 17	RRPQDVKFPGGGQIVGGVYLLPRRGP	1, 2, 5 & 3 variant & 6 variant								
	VYLLPRRGP	1, 2, 4, 5, 6	0	0	18%	0	0	16%	**48%**	18%
	VGGVYLLPR	1, 2, 4, 5, 6	0	0	17%	0	9%	9%	10%	20%
NS3 1248	GYKVLVLNPSVAAT	1, 2, 4, 5, 6								
	LVLNPSVAA	1, 2, 3, 4, 5, 6	**37%**	**54%**	**36%**	**47%**	**28%**	17%	**34%**	**39%**
	YKVLVLNPS	1, 2, 4, 5, 6	5%	0	0	30%	9%	31%	27%	17%
NS3 1781	LPGNPAIASLMAFTAAVTSP	1a, 4 variant								
	LPGNPAVAS	2,3, 5, 6	0	2%	0	4%	0	0	9%	0
	LPGNPAIAS	1, 4	0	0.70%	15%	4%	0	2.40%	0	7%
	IASLMAFTA	1	7%	**23%**	0	0	4%	0	14%	**21%**
NS4 1801	LTTSQTLLFNILGGWVAAQL	1a, 1b variant,								
	LFNILGGWV	1, 4, 5	0	0	16%	0	**24%**	0	16%	**28%**
	FNILGGWVA	1, 4, 5	**47%**	**47%**	0	2%	**16%**	**28%**	16%	**31%**
	ILGGWVASQ	4, 5	0	0	**28%**	0	0	2.40%	8%	0
	LGGWVASQI	4, 5	0	0	0	0	**21%**	0	13%	**21%**
NS5B 2571	KGGRKPARLIVFPDLGVRVC	1, 2 variant, 6 variant								
	VFPDLGVRV	1	0	0	**34%**	0	0	0	0	0
	VYPDLGVRV	3, 5	0	0	**35%**	0	14%	0	0	19%
	IVYPDLGVR	3, 5	0	0	**28%**	0	0	0	7%	0
	LIVYPDLGV	3, 5	0	0	0	0	12%	0	3	60%
NS5B 2661	QCCDLDPQARVAIKSLTERL	5 variant								
	LAPEARQAI	1b	0	0	8%	0	11%	0	4.50%	11%
	LDPQARVAI	5	0	0	8%	0	0%	0	0	0%
	LQPEARAAI	5 variant	0	0	22%	0	12%	1%	22%	26%

7.4. DISCUSSION

In this study we looked at the heterogeneity of epitopes proposed for HCV vaccines across the six genotypes and the binding efficacy of these epitopes and their variants to common HLA alleles in the SA population. There are many epitopes that could have been included in this chapter but those that are discussed here have been considered as epitopes in multi-epitopic (Wei et al., 2008), therapeutic (Wedemeyer et al., 2009), minigene (Martin et al., 2008) and DNA polytope (Memarnejadian et al., 2009) vaccines. All these vaccine studies have focused on genotype 1a and 1b strains and the A*02-restricted epitopes. Although genotypes 1, 3 and 4 have been found in the SA population, genotype 5a is predominant (Prabdial-Sing et al., 2008). Genotype 2 is rare and, to date, genotype 6 has not been identified in SA. Hence, the present study was focused on genotype 5a sequences. We included all of our available sequence data, sequences from Verbeeck et al., 2006 (Belgium and South Africa), and Henquell et al., 2004 (France) in our alignments, in order to improve the representation of genotype 5a. Class I and II epitope sequences of genotype 5a were found to be relatively conserved compared to other genotypes. There were some epitope regions that were highly conserved, for eg. the **NS3**$^{1628\text{-}1637}$ and **NS4**$^{1807\text{-}1816}$ and others like the **NS5B**$^{2661\text{-}2680}$ which were highly variable.

The divergent mutational potential of the genotypes reported by Rauch et al., 2009 is also evident in the alignments and conservation table generated for this study. The consensus within known epitopes represents sites best adapted to T cell responses acting on these sites across the host population, however, polymorphic sites frequently have different consensus amino acids between the genotypes. This

was clearly illustrated in the Class I epitope binding table as there were many variations of the same epitope seen. In light of this data, we suggest that epitope-based vaccines would need to contain a mixture of epitopes from all of the genotypes and their coverage potential checked against local HLA allele frequencies. Because of this variation, representative epitopes from all genotypes would need to be included in the vaccine and their population coverage assessed. A new web server, Opti-Tope (Toussaint and Kohlbacher 2009), recently became available which uses the same 4 steps as used in this study to suggest candidate epitope sequences for use in epitope vaccines. However, rather than assessing a given epitope-based vaccine as we have done, the output is a set of optimal epitopes. When we placed a set of 10 optimal epitopes, generated from a sample alignment of predominantly HCV genotype 1 sequences, into the IEDB epitope analysis program, coverage of only 44% was found in the Zulu population. This indicates that there is a 44% chance for a person from this target population to carry one of the alleles covered in the analysis, and hence respond to the vaccine. This questions the efficacy of a vaccine, based firstly on genotypes not seen in the majority of our patients and secondly, poor T-cell responses due to inadequate binding of epitope to HLA allele. Interestingly, the coverage of a putative vaccine containing 29 epitopes adapted for genotypes from the 13 chosen in this study only increased the coverage in the Zulu population to 51%. This re-iterates the fact that the optimum epitope set will only be as good as the data placed into the program. Alignments of genotypes prevalent in the country and detailed, accurate population allele frequencies should produce results of epitopes better matched to the target population.

This study used well-characterized and published epitopes, some of which are al-

ready in use. NS3$^{1073-1081}$, NS4B$^{1801-1820}$ and NS5B$^{2571-2590}$ are contained in IC41, an add-on to standard drug therapy (Wedemeyer et al., 2009). Although some of the epitopes in this study are well conserved among genotypes 4 and 5, their binding scores to relevant HLA alleles in SA were poor. There is a lack of cross-reactivity between genotypes, and this may explain the lack of cross-protection highlighted in earlier studies (Farci et al. 1992, Accapezzato et al. 2002, Rauch et al. 2009). Manipulating epitopes chosen for another population can never produce a multi-epitopic therapeutic vaccine for the SA population infected with genotype 5a. However, based on binding affinities predicted for epitopes expected in the local population and prevalent HLA alleles used in this study, the following class I-restricted epitopes, NS4B$^{1807-1816}$ (LLFNILGGWV; Cerny et al., 1995, Wei et al., 2008, Martin et al., 2008), NS5B$^{2422-2433}$ (MSYSWTGAL; Wei et al., 2008), NS5B$^{2727-2735}$ (KLRDCTLLV; of the published epitope sequence GLQDCTMLV; Wei et al., 2008) and class II-restricted epitopes NS3$^{1252-1260}$ (LVLNPSVAA; Day et al., 2002) and NS4B$^{1809-1817}$(overlapping peptide to class I-restricted 1807, FNILGGWVA; Martin et al., 2008) are best suited for a local vaccine.

From the many putative epitopes which can be discovered by computational prediction methods, the most conserved among strains which also show good binding affinities to many HLA alleles (promiscuous) are the best candidates for *in vitro* and/or *in vivo* testing. It is known that of the peptides predicted *in silico*, only one or two may actually bind to a particular MHC (Donnes et al., 2002) and it remains essential to test predicted peptides so as to ascertain that the needed T-cell response is elicited.

7.5. CONCLUSION

It is becoming imperative to know how host genetic differences can influence disease, preventative (vaccine) and treatment outcomes. Genome-wide association studies have revealed the host's ability to defend itself from disease for example, SNPs on *IL-28* (Ge et al., 2009, Tanaka et al., 2009) and 6 HLA alleles have shown significant association for HCV absence (B*57:01, B*57:03) or viraemia (DRB1*01:01, DQB1*03:01, Cw*01:02 and DRB1*03:01) (Kuniholm et al., 2010).

Here, we have considered HLA heterogeneity in South African ethnic groups, as well as viral variability of well published epitopes, in the light of less-studied genotypes 4 and 5. As indicated earlier in chapter 1 (1.6.2), the main caveat to vaccine development for HCV lies in these two variable factors. We have seen that even with the most published A*02- restricted epitope candidate, $NS3^{1073-1081}$, there were many changes seen across HCV genotypes. This poor and semi-conserved nature of Class I epitopes may be the first challenge towards generating a good CTL response.

The targets for most of the multi-epitopic studies are the A*02 allele. As expected, the best prediction binding scores were seen to this allele for the A2-restricted epitopes $NS3^{1073-1081}$, $NS3^{1406-1415}$, $NS4B^{1851-1859}$ and $NS5B^{2727-2735}$. Poor binding affinities of these epitopes to the common SA alleles were observed in this study. Despite the homogeneity of genotype 1 and genotype 5 over the epitopes proposed in vaccine studies, host differences will make the vaccine less effective in the South African setting. Of the 13 published and well-characterized epitopes selected for our analysis, we have found that 3 class I- and 2 class II-restricted epitopes would

be beneficial in a multi-topic therapeutic vaccine for genotype 5a infection for our population. While i*n vivo* and *in vitro* studies are needed to confirm immunogenic epitopes, data mining studies provide a sound basis which take into account local HCV variation and host allele frequency on which to design the most suitable HCV vaccine for local populations. Whether it will be feasible for a personalized HCV vaccine at individual and population level still needs to be established.

CHAPTER EIGHT

HCV genotypes and response to therapy in a South African study group

8.1. INTRODUCTION

The overall goal of anti-HCV therapy in acute and chronic infection is to eradicate the virus from the body, that is, treat to cure. This may not always be possible as alluded to in the previous chapter, as viral (genotype, load and mutations, Yen et al., 2008) and host factors (ethnicity, race, HLA*, IL28B* gene polymorphisms, Ge et al., 2009; Tanaka et al., 2009) are associated with response to therapy (reviewed in Rodriguez-Torres, 2010). The recommended treatment is combination therapy with pegylated-interferon (PEG-IFN) alfa-2a and ribavirin (Manns et al., 2001). PEG-IFN is a molecule engineered to mimic IFN which the body produces in response to infection as part of the innate defence system. Attaching a polyethylene glycol (PEG) moiety to IFN to produce PEG-IFN extends the half-life of IFN (Thomas and Foster, 2007). Ribavirin is a nucleoside analogue and is thought to support a type 1 cytokine-mediated response (Graci and Cameron, 2006) against the virus in the host. Together with PEG-IFN, viral RNA metabolism and thus HCV replication is inhibited (Coelmont et al., 2006).

Both the type of response to combination therapy and the ideal duration of the treatment regime vary dependent on the HCV viral genotype (Botha et al., 2005). Individuals infected with genotypes 1 and 4 respond poorly to therapy when compared to those infected with genotypes 2 or 3 (Nguyen and Keeffe, 2005). Because genotype is a good predictor of response rate to therapy (as indicated in Chapter 1), genotyping has become an important tool in design of treatment regimes. Table 32

summarizes the different terms used to evaluate the response to HCV treatment (WHO, 2002; Ghany et al., 2009). Corey et al., 2010 performed a meta-analysis of 22 published studies on early HCV treatments and conclude that if HCV has not spontaneously cleared by week 12 post infection, then treatment should commence so as to ensure a good chance of clearance. This 12-week cut-off corresponds to early virological response (EVR, Table 32) in treated chronic patients and is thus termed “acute EVR”, to indicate spontaneous clearance in acute infection by 12 weeks. This early treatment regime is the recommended post exposure prophylaxis (PEP) for HCV (National guidelines for the prevention and control of Hepatitis C virus in SA, final draft, 2010) but not commonly used as the acute phase is predominantly asymptomatic.

Previously, response to therapy was measured at 12 weeks but a recent study recommends response rates as early as 4 weeks (Zeuzem et al., 2009). For genotype 2 and 3 infection, patients without RVR at 4 weeks can be treated for 48 weeks. At 12 weeks, if patients have not achieved a pEVR, then treatment may be stopped. Treatment may also be discontinued if the patient, although has attained pEVR but is HCV RNA positive at week 24. However, if there is a >2 log reduction in viral load, treatment can be extended to 72 weeks (Zeuzem et al., 2009) except when HCV RNA is not detected at 12 weeks in which case, treatment is continued to 48 weeks. This guideline is currently in use in SA (Botha et al., 2010). To date, however, global data on responses of patients infected with genotype 5a, one of the predominant genotypes found in South Africa, is limited (Antaki et al., 2009; Nieuwoudt et al., 2007).

Table 32. Terms used to evaluate response to HCV treatment

Response	Explanation
SVR (sustained virological response)	HCV RNA negative 24 weeks after treatment ends. HCV is not detectable in blood together with persistent normalization of serum ALT six months after completing treatment. Patients with a SVR will remain HCV RNA negative for at least 5 years after stopping therapy and experience a long-term biochemical and histological outcome with a decrease in total inflammatory activity and a decrease in the reversible components of fibrosis. HCV RNA may still be detectable in the liver of serum HCV RNA negative patients showing ongoing inflammatory change. Nevertheless, patients with a SVR have a highly reduced risk of disease progression.
EVR (early virological response)	HCV viral load has dropped by at least 2 logs (partial EVR) or is undetectable after 12 weeks (complete EVR) of treatment. A patient who does not have an EVR has a poor prognosis of attaining SVR. Can choose to discontinue treatment or switch to a lower dose maintenance therapy although maintenance therapy is contraindicated by recent negative results from the HALT-C (Hepatitis C anti-viral long-term treatment against cirrhosis) program being performed by the National Institute of Diabetes, Digestive and kidney disease, NIDDK, which is part of the National Institutes of Health, NIH (http://digestive.niddk.nih.gov/, 2010).
ETR (end-of-treatment) response	No HCV is detected by HCV viral load assay at completion of therapy at week 24 or 48. Some patients with an ETR will see HCV viral load return, usually 12-24 weeks after treatment has stopped. End-of-treatment response is assessed by ALT and/or HCV RNA estimation when therapy is completed.
Relapser	Reappearance of HCV RNA after therapy is discontinued.
Breakthrough	Reappearance of HCV RNA while patient is still on therapy.
Non-responder	Patients who do not have an EVR or, if they complete the course of therapy, do not attain at least a 2 log drop in viral load or never test undetectable for HCV RNA while on treatment.
Null-responder	Patients who fail to decrease HCV RNA by at least a 2 log drop after 24 weeks of therapy.
RVR (rapid virological response)	Response to therapy after 4 weeks is called a RVR. If viral load is undetectable at that point, it is a good predictor of an SVR later. However, RVR is not good at predicting who is unlikely to respond, so treatment should not be suspended if there is no RVR.

The aim of this study was to describe the response rates of HCV-infected individuals with single or mixed genotype infections to combination therapy in SA.

8.1.1. Objectives:

(a) To describe the genotypes of patients receiving therapy.

(b) To determine how patients infected with single or mixed genotypes, particularly genotype 5a, respond to therapy

8.2. METHODS

This prospective study began in 2008, and presently includes 56 patients from state hospitals in Durban (KZN), Johannesburg, Pretoria, Rustenburg, Garankuwa and Bloemfontein for which clinical and demographic information is available. The inclusion criteria for patient selection were: confirmed HCV positive by qualitative PCR and HCV genotype, a detectable HCV viral load, age >18 years and verbal and written consent at point of collection (participating hospital/clinic). Patients were excluded if: the laboratory results were inconclusive or negative, individuals did not meet the above inclusion criteria and if the patient had any contraindications to therapy. HIV and HBV infections were managed appropriately by the clinician responsible for treating the HCV positive patient.

The specimens for testing were only eligible when collected in EDTA tubes (10-20 ml whole blood or 5 ml plasma/serum) and specimens were transported within 24-36 hours to the laboratory on dry ice for viral load and genotyping tests. The samples that were excluded for testing were those that were not collected in EDTA tubes, were insufficient in volume and HCV tests were not confirmed positive.

PEG-IFN (Pegasys, Roche) and ribavirin (Copegus, Roche) were administered at a dose of 180 µg weekly and 800 mg/day (for those under 75 kg) and 1200 mg/day (for those over 75 kg), respectively. Specimens were collected before therapy and sent to the NICD laboratory for HCV testing. Each specimen was assigned a unique laboratory identification number and a database was established with the identification number and patient information, only traceable and maintained by the principal investigator. A validated commercial line probe assay (LiPA, version 2.0, Siemens) was used to determine the genotype as detailed in chapter 4. The COBAS Ampliprep/COBAS TaqMan HCV test (Roche Diagnostics, Germany), found to be a reliable and sensitive assay to predict response to therapy in HCV-infected individuals (Kanda et al., 2010) and has a lower limit of detection of 15 IU/ml, was used for quantitative viral load determination. Specimens were received at the laboratory from patients on therapy for 4 weeks, 12 weeks, 24 weeks and 48 weeks. The study is an ongoing collaboration with Roche Pharmaceuticals and clinicians from the various participating hospitals.

8.3. RESULTS

Twenty-nine patients completed therapy at 48 weeks, 1 patient is receiving extended therapy to 72 weeks and 1 patient had end-stage renal disease (ESRD) and was excluded from the study. A further 20 patients are still receiving therapy, 10 of these are presently at weeks 2-3 while other 10 patients are at 4-6 weeks. Five of the patients have defaulted on the treatment program.

Table 33a gives a summary of the available demographics of the study group. The mean age of the study group is 47 years, with 35/56 (62.5%) males and 21/56

(37.5%) females. Of the non-responders, two were over the age of 60 and three were <43 years. Thirty-eighty (67.86%) had chronic liver disease (fibrosis scores of F2-F4), 17/56 (30.36%) were HCV infected haemophiliacs and 1/56 (1.78%) had renal disease. Fifty percent were Blacks, 14/56 (25%) were Whites and 25% were Indians. The mean viral load was $5.9X10^5$IU/ml (Table 33a). Patients with liver disease and genotype 1b (9/17) were older (48.25 years) than those infected with genotype 1b (5/17) with haemophilia (27.2 years). Patients with genotype 3a and liver disease (4/10) had a mean age of 42 years, similar to those with CLD infected with genotype 1b.

Table 33b provide a summary of the genotypes identified in the study group, with a breakdown of the SVR and RVR responses per genotype. Genotype 1 had the poorest response rate (combined SVR and RVR) of 41% and the one patient with the mixed infection did attain SVR.

Table 33c provides the overall responses for the cohort. The breakdown into the subtypes include genotypes 1(subtypes a and b), 3(subtype a), 4 (subtypes c/d, f and h), genotype 5 (subtype a) and one mixed infection (1b+5a) (Fig. 27). 12 patients were excluded from the response rate analysis, as 10 had just begun therapy (<4 weeks), 1 had ESRD and 1 patient was placed on the extended treatment regime according to the study protocol (Table 33c). Thirty-four individuals responded to therapy, with SVR in 24/44 (54.56%) and RVR in 10/44 (22.72%; Fig.28). Five patients defaulted (non-compliant) and were withdrawn from the study (Table 33c, Fig. 28). SVR was not seen in 5 individuals: P2 (genotype 1b), P9 (genotype 1b), P12 (genotype 1b), P48 (genotype 4f) [Fig.29a] and P13 (genotype 3a, Fig.29b).

Patient 2 did not clear infection at 13 weeks. P9 and P12 were viraemic at 25 weeks and P48 did not attain RVR at week 8 (Fig. 29a). The patient (P21) with mixed genotype (1b + 5a) infection cleared the virus after 32 weeks (Fig. 29a). All individuals with genotype 5a, on therapy, responded well, (RVR 22% and SVR 56%; Fig.30). A patient (P13) with genotype 3 did not clear infection at 28 weeks (Fig.29b) but all other patients with genotype 3 or genotype 5a cleared infection, some as early as 4 weeks.

Table 33. HCV positive patients on combination therapy 2008-2010; (a) Demographics, (b) genotypes of the study group and (c) overall responses to date

(a) Demographics

	N=56
Age (years)[a]	47± 13.6
Mean Viral Load (IU/ml)[a]	590930 ± 4598528
Male	35 (62.5%)
Female	21 (37.5%)
CLD	38 (67.86%)
Haem	17 (30.36%)
ESRD	1 (1.78%)
ETHNIC:	
C	14 (25%)
B	28 (50%)
I	14 (25%)

(b) Genotypes of the study group

GENOTYPES (N=56)		SUBTYPES		SVR+RVR responses
1	17 (30.35%)	1 (3/17) 1a (1/17) 1b (13/17)	17.64% 5.89% 76.47%	7/17 (41%)
3	10 (17.86%)	3 (2/10) 3a (8/10)	20% 80%	7/10 (70%)
4	10 (17.86%)	4 (6/10) 4c/d (1/10) 4f (1/10) 4h (2/10)	60% 10% 10% 20%	5/10 (50%)
5	18 (32.15%)			14/18 (77%)
Mixed	1 (1.78%)			100%

(c) Overall responses

RESPONSES ACROSS ALL PATIENTS (N=56)	
SVR	24 (42.86%)
RVR	10 (17.86%)
No SVR	5 (8.93%)
<4 Weeks	10 (17.86%)
Default (non-compliant)	5 (8.93%)
ESRD	1 (1.78%)
extended	1 (1.78%)

[a] Mean +/- std deviation; CLD, chronic liver disease; Haem, haemophiliac; C, Caucasians; B, Blacks; I, Indians; SVR, sustained virological response; RVR, rapid virological response; ESRD, end-stage renal disease

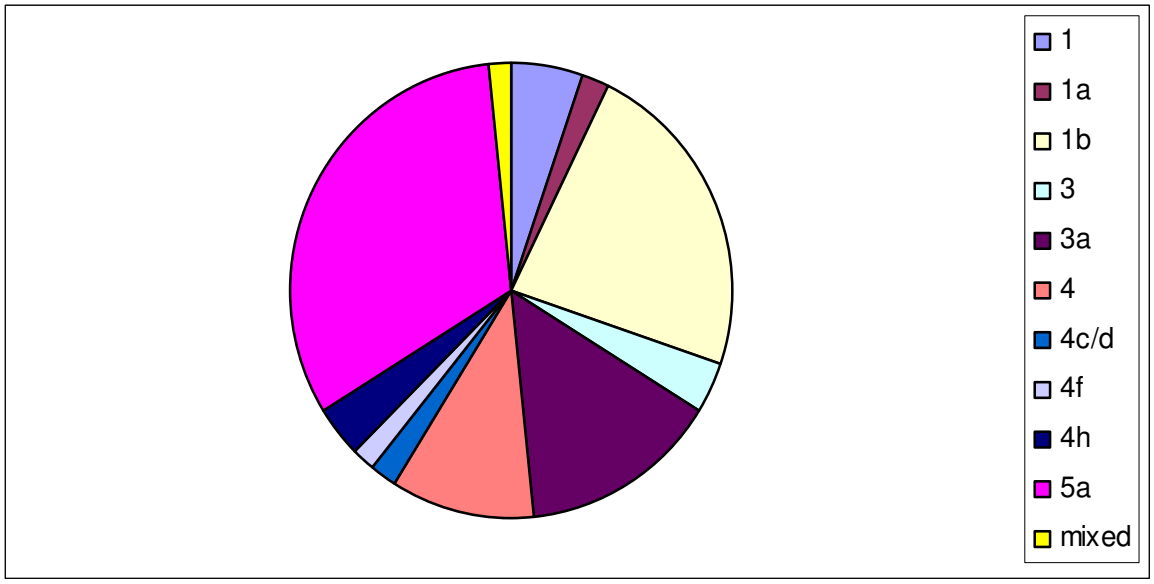


Fig.27. A breakdown of genotypes and subtypes in patients receiving combination therapy. The majority of patients had genotype 5a, followed by genotype 1b and one patient was co-infected with genotypes 1b and 5a.

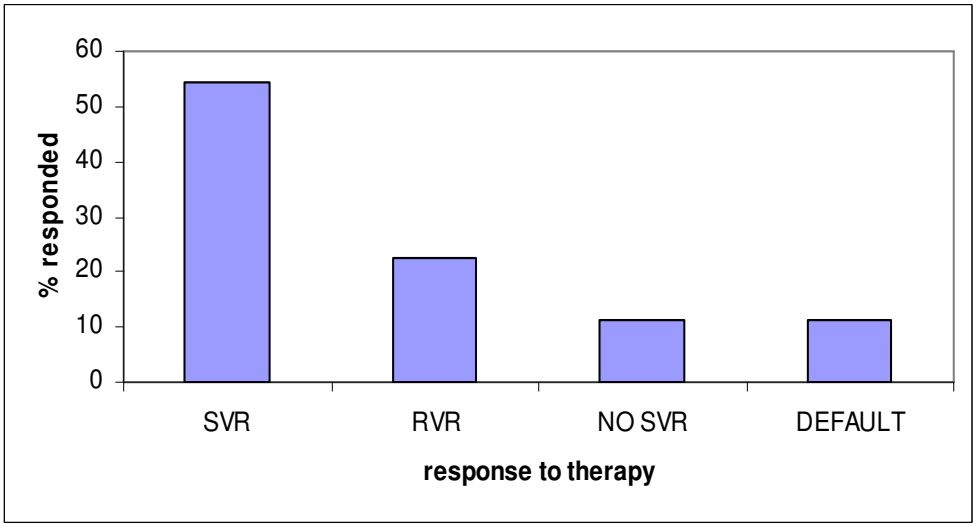


Fig.28. Responses of 44 patients receiving combination therapy for HCV infection across all genotypes. SVR was observed in 54.56% of patients and RVR in 22.72%. Five individuals did not attain SVR (genotypes 1b, 4f and 3a).

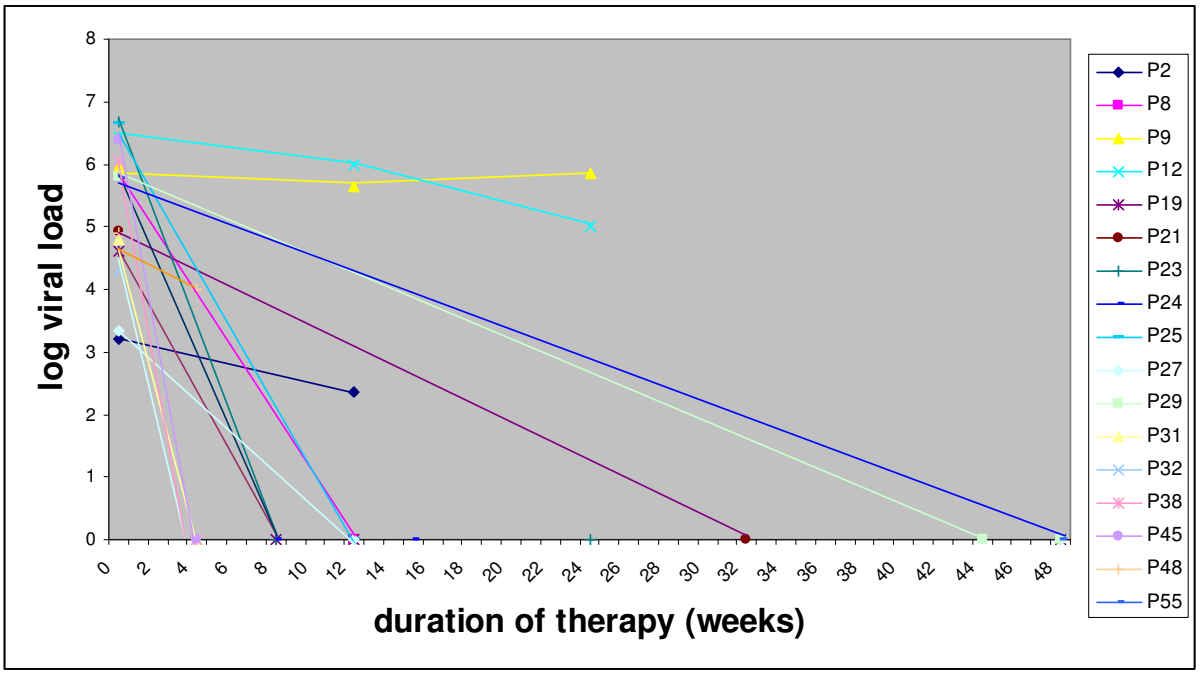


Fig.29a. Responses to combination therapy in patients infected with genotype 1 or genotype 4 and one patient with a mixed infection (1b+5a), P21.

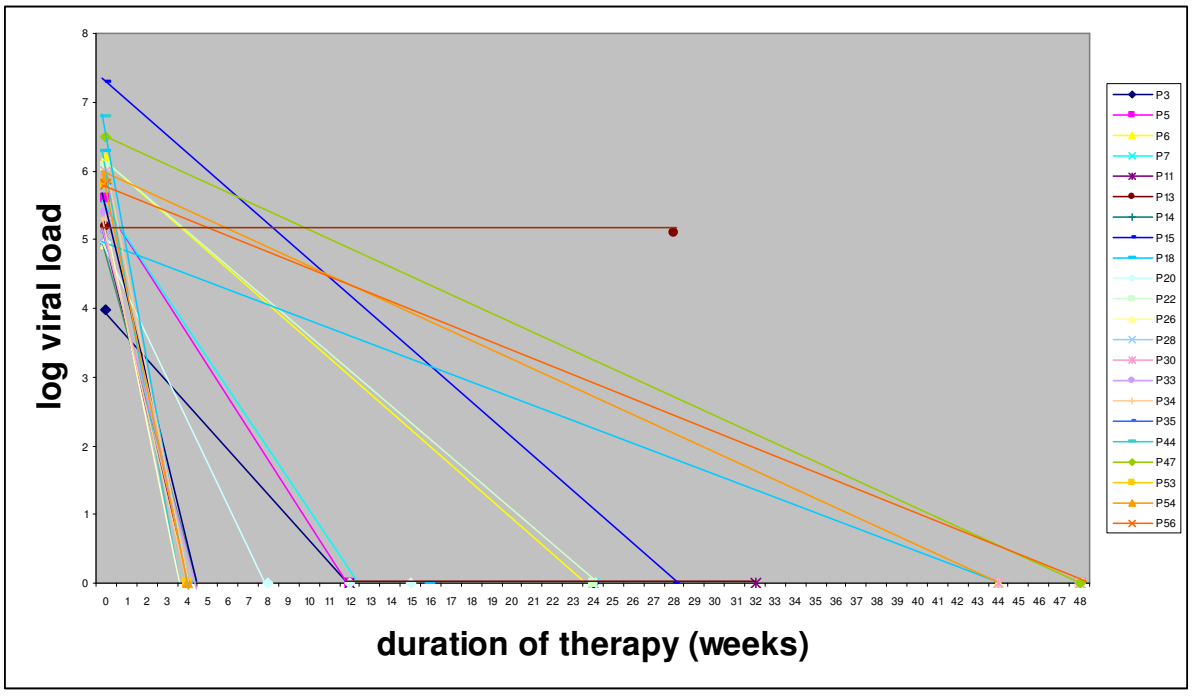


Fig.29b. Responses to combination therapy in patients with genotype 3 or genotype 5.

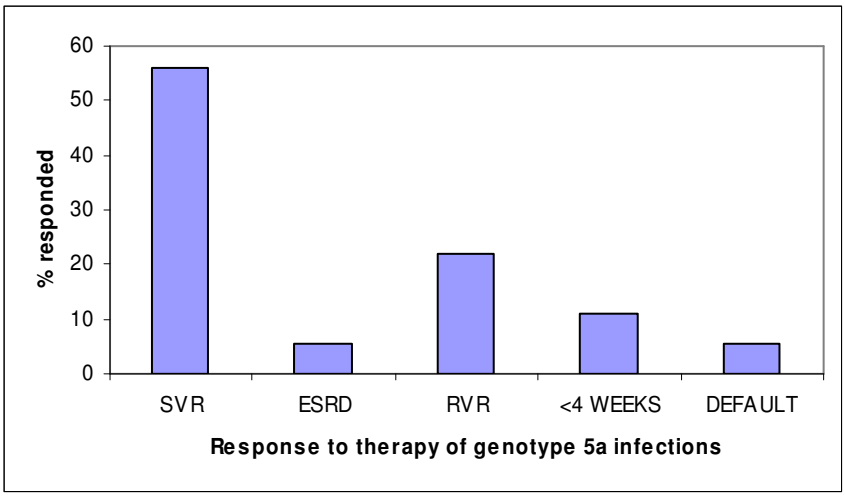


Fig.30. The different responses to combination therapy in patients infected with genotype 5a. SVR was observed in 56% and RVR in 22% of patients infected with genotype 5a.

8.4. DISCUSSION

The current standard of care for people infected with hepatitis C in South Africa is combination therapy with PEG-IFN and ribavirin (Botha et al., 2010). There were no prospective studies on the treatment of HCV genotype 5a in the country (Antaki et al., 2010), so the present study provides vital information and a well-maintained database on patient responses to therapy, particularly, genotype 5a in the SA population. Data from small retrospective studies indicate an SVR for genotype 5a infections in Belgium (21 patients, 48% D'Heygere et al., 2005); Syria (26 patients, 54%, Antaki et al.,2008) and France (12 patients and 87 patients, 67% and 60%,, Legrand -Abravanel et al., 2004 and Bonny et al., 2006 respectively). This is the first report on treatment responses in South Africa to combined therapy in single and mixed genotype infection which includes 5a. This has relevance because mixed genotype infections are rare (chapter 4), and extended therapy durations, dependent on genotype, may be necessary in the case of mixed infection. Studies have reported on slower response rates in African Americans compared to American Caucasians

(Donlin et al., 2007; Yee et al., 2009) however, there were no significant differences among the ethnic groups in this study.

Contrary to USA reports on the high prevalence of HCV and HIV co-infection (16% in adult clinical group; Sherman et al., 2002 and 75% in IVDUs; Strasfeld et al., 2003), the prevalence in SA is low (1-1.9%; Lodenyo et al., 2000 and Amin et al., 2004, respectively). However, with the rise in the HIV epidemic in SA, one would expect an increase in the number of people infected with HCV and HIV as these RNA viruses share a common exposure route (13.4% of N=1937 patients, Parboosing et al., 2008; Table 2, page 40). One patient in our study (1/56) was HIV-coinfected and had just started HCV treatment (< 4 weeks) with HIV therapy, optimally managed by the treating clinician. In SA, all individuals positive for hepatitis C should be screened for HIV and patients with a CD4 count <200 cells/µL, the HIV should be treated and managed first before beginning HCV therapy and patients with a CD4 count >200 cells/µL, HCV infection should be treated first (Botha et al., 2010). PEG-IFN and ribavirin, together with HAART, was found to be effective in coinfected individuals (Chung et al., 2004). Anti-HIV drugs (didanosine, stavudine and AZT) with ribavirin are contraindicated and should be avoided (Botha et al., 2010). With these and other drug interactions which lower antiretroviral drug efficacy, increase toxicity and morbidity/mortality, newer anti-HIV (reviewed in Brook et al., 2010) and anti-HCV drugs (reviewed in Rodriguez-Torres, 2010) need to be considered in co-infected or mono-infected individuals.

HCV genotypes vary in distribution with transmission routes (Zeuzem et al., 1996, explained in chapter 3) and age (Zeuzem et al., 1996; Idrees and Riazuddin., 2008).

Individuals infected with genotype 1a and 1b (transfusion-related or unknown risk) were reported to be comparatively older than those with genotype 3a (IVDUs) (Zeuzem et al., 1996; Idrees and Riazuddin, 2008). Although not statistically significant, our study shows that patients with CLD, infected with genotypes 1b and 3a were older (48 and 42 years, respectively) than those with genotype 1b and haemophilia (27 years). So, it appears that risk of transmission may be more related to genotype distribution, than age as in SA. Diagnosis of HCV appears to be earlier in the haemophilia population compared to other non-haemophilia patient populations (chapter 3). Age and degree of cirrhosis have been shown to be independently associated with SVR (Legrand-Abravanel et al., 2009b), we have not been able to show that in this study as the numbers are small. All the patients that did not respond to therapy had CLD and one had an F4 score on the liver fibrosis pathology.

SVR rates of 42-46% have been reported for genotype 1 (Manns et al., 2001); 55-66% for genotype 4 (Kamal et al., 2005); 76-82% for genotypes 2 and 3 (Manns et al., 2001) and 48-67% for genotype 5a (Antaki et al., 2008). The results from the present study show a higher response rate (calculated as SVR+RVR combined), 77% for genotype 5a. Lower SVR+RVR combined rates were recorded, however, for genotype 1 (41%), genotype 3 (70%) and genotype 4 (50%). We do expect our overall SVR rates to increase as almost 20% of patients had RVR and a RVR (viral load negative at 4 weeks) is considered a good predictor of SVR, particularly in genotype 2 and 3 infections (Tarantino and Craxi, 2009). Genotype 5a infections had response rates (>70%) similar to those seen for genotypes 2 and 3 (Manns et al., 2001). One patient with genotype 3a did not attain an SVR. Genotype 3 has recently been associated with progressive liver fibrosis (Bochud et al., 2009) and

alcohol use. Overweight and iron overload (Zeuzem et al., 2009) may have also been factors found to be affiliated to a poor response. Although there are other drug therapies offered to patients with seemingly easier-to-treat HCV infections (such as genotype 3), the response rates of these also vary according to genotype. The protease inhibitor; telaprevir (Foster, 2009) and boceprevir (Kwo et al., 2009) have shown good SVR for genotypes 1 and 2, but not 3 and 4 and the polymerase non-nucleoside inhibitors have been reported to successfully target genotype 1 (Soriano et al., 2009). In Legrand-Abravanel's study 2009a, patients with subtype 1a showed a lower response to therapy as compared to those with 1b. We had only one patient with subtype 1a in our cohort and it was too early to comment as therapy had just begun. However, subtyping and responses to therapy need to be monitored in this and future studies, not only for genotype 1, but also genotype 3 and 4. If response rates are found to be associated with different subtypes, then clinicians need to be informed so as to request for subtype diagnostic tests (chapter 4).

Most people with a sustained virological response (SVR, Table 32) remain virus-free, however, a recent study indicates the sustained presence of low levels of virus in blood/tissue (Corey et al., 2010). These levels may be too low to have a significant effect on the liver (Corey et al., 2010). Long-term follow-up studies are needed to ascertain whether a resurgence or re-activation of the virus is possible in these individuals.

8.5. CONCLUSION

This is the first prospective surveillance study on responses to HCV therapy in SA. Though the study numbers are small, the treatment response for genotype 5a is higher in this study compared to previous global studies. Response rates for genotype 5a were found to be comparable to results for genotypes 2 and 3 in developed countries. Longer duration of therapy (32 weeks) was necessary for SVR in the patient co-infected with genotypes 1b+5a in this study. This is an important observation which needs to be considered in patients infected with a mixture of poor and good responder subtypes. Present results predict an intermediate treatment response time will be feasible for patients with genotype 5a. Further prospects on this study would be to screen for SNPs in the *IL28B* gene in patients receiving therapy to predict responses and identify whether mutations in the NS5A-ISDR correlate with SVR in genotype 1b infections.

CHAPTER NINE

FINAL DISCUSSION AND FUTURE DIRECTION

As HCV is a relatively “new” virus, there is much more to learn about the virus and its interaction with the human host. In SA, knowledge about HCV is still in its infancy and more needs to be done to create awareness and educate the public. This is the first extensive molecular study on the hepatitis C virus in this country and it has provided added insight regarding circulating genotypes, host immuno-genetics and patients responses to combination therapy.

The study determined the number of PCR positives in HCV-seropositive hospital-based study groups. Twenty-one % of the seropositive specimens were found to be RNA negative, indicating that seroprevalence data must be looked at with caution. A positive HCV PCR result is necessary to confirm active infection. The availability of an accredited and validated PCR test at the NICD laboratory allows for quick and early diagnosis of infection. This testing algorithm is recommended in the SA National Guidelines for HCV prevention and treatment which will soon be available on the NDoH website (expected date 2012). A National HCV surveillance database was established to collate all data from public health laboratories. This comprehensive database links patient demographics with HCV laboratory results. This baseline data will be available for future molecular and epidemiological studies and treatment and preventative initiatives. A public health study is projected in the next year to extend this database. Amongst other things, transmission risks in SA are of interest. The national HCV surveillance system will routinely inform the NDoH of reported HCV positive cases in the different sentinel groups. Of the 621 HCV PCR positive specimens reported to NDoH for the period 2004-2008, only <10% have

been treated or are currently on treatment. We recommend that early treatment be provided for all HCV PCR positive patients so to prevent huge costs later with regard to liver transplants and hospitalization.

The repository of data made it possible to retrospectively determine genotypes in two hospital-based study groups and a volunteer blood donor group. A sequencing based genotyping technique (SBT) was compared for two regions (5'UTR and NS5B) of the HCV genome. There was a 100% concordance between the two regions at genotype level. Subtyping within genotypes was better determined using the NS5B region. Subtypes of genotype 4, not previously described in this country, were found in our study cohorts. Although, genotype 5a is still the predominant genotype in the country, we have noted an increase in frequencies of other genotypes, like genotypes 3 and 4.

The methods used to genotype HCV are summarized in Table 34 to indicate the most cost efficient, less time-consuming and quick method from a public health perspective for diagnostics and surveillance. Even though, SBT is a gold standard for genotyping HCV, it is time-consuming and labour intensive, especially in high throughput laboratories. Three molecular-based tests for genotyping HCV were compared to the SBT method. To date, the LiPA (5'UTR) was found to surpass the other genotyping tests at genotype level. The PCR amplicon from the qualitative PCR test is used in the LiPA (5'UTR) assay and this has improved turn around times. The validated genotyping test, LiPA (5'UTR) was used to type the SANBS cohort and this separate database for the volunteer blood donor group is used by SANBS who store HCV positive specimens. Genotype testing has allowed the sup-

ply of unique genotypes, such as genotype 5a and genotype 4 to the external quality assurance program, QCMD. Although the LiPA (5'UTR plus core) was validated as a method for subtyping genotype 1, it was found to miss mixed genotype infections, particularly the mixtures of genotypes 1+4 and genotypes 1+5. Sequence similarities between the 5'UTR and core regions used as probes in the assay were found. It is speculated that these could be responsible for the loss of discernment of the LiPA assay when the core amplicon was added.

Although the PNS method can also be used for genotyping, it was found to be labour-intensive and requires computer programmes. Subtyping of genotype 1 is also not possible if upstream 5'UTR sequences are not available. However, the technique allows the visualisation of the primary sequence data of the 5'UTR as a stem-loop structure.

Table 34. A summary of the five different methods used to genotype HCV in this study

	SEQUENCING		**LiPA**		**PNS**
	5'UTR	NS5B	5'UTR	5'UTR+CORE	5'UTR
subtyping	Not efficient	good	Not efficient	good	Efficient if all relevant regions are sequenced
costs	high	high	Use denatured amplicons, only detection costs	Kit-based amplification and detection	Requires computer software
time	Time-consuming	Time-consuming	Quick and easy	Added amplification step	Time-consuming
skill	high	high	Easy to teach	high	Software knowledge required
Lab-intensive	yes	yes	no	yes	yes
Output TAT	3-5 days	3-5 days	4 hours	6 hours	3-10 days (incl. analysis)
Mixed infections	no	no	yes	Yes (only partly)	no

Genotype 5a specimens were further characterized by sequencing regions, other than the 5'UTR. For the first time in a South African laboratory, molecular sequence analyses of the core, E1, NS4B and NS5B regions have been determined Nucleotide sequence data of genotype 5a in South African patients have been compared to that which is available on global databases and phylogenetic inferences have been made. The E1 and NS4B regions showed a greater diversity than the core and NS5B regions. Geographical clustering of the SA isolates was more evident in the E1 and NS4B regions. The branching order of genotype 5a after genotype 2 but before genotype 3 may suggest that genotype 5a is older than genotype 3. If older genotypes respond better to treatment as has been speculated, then we can infer that patients with genotype 5a may respond better than those infected with genotype 3. However, it is noted that although these results concur with previous studies, this study was conducted on sub-genomic regions. Full genome amplification and phylogenetic analyses requires further optimization. The divergence estimate of genotype 5a from the oldest genotype (genotype 2) was found to be between 152 and 156 years for the core, NS4B and NS5B regions. The analyses in the E1 region estimated a more recent divergence of 100 years.

The nucleotide sequences generated in the core, NS4B and NS5B regions were translated into amino acids and these were used in epitope analyses across immunodominant regions. We found that the HCV class II-restricted epitope sequences were more conserved among genotypes than the class I-restricted epitope sequences. Although both class I and II-restricted epitope sequences of genotype 5a were found to be relatively conserved compared to other genotypes, for example genotype 6, we have shown that the binding prediction scores to HLA-alleles found

in the SA population were poor. This indicates that an epitope-based vaccine, targeting genotype 1 and the A02 allele may be of limited benefit to individuals infected with other genotypes and with different HLA backgrounds. This data-mining study proposed a cocktail of highly conserved MHC Class-I and Class-II epitopes. This took into consideration their respective binding efficiencies to common HLA alleles in the different ethnic groups, for a putative therapeutic vaccine, targeting our local population. Although it was reported that only a few HLA alleles are significantly associated with HCV disease or clearance, it remains vital that host genetic factors of infected individuals be studied so as to understand their immune responses during the infection and whether they will respond to therapy or not.

A prospective study of the responses to combination therapy of patients with liver disease or haemophilia, having either single or mixed genotype infection, was carried out. This study found a higher response rate to therapy in patients infected with genotype 5a compared to other genotype 5a response studies. Response rates for genotype 5a were found to be comparable to those seen for genotypes 2 and 3. This study has also shown that longer duration of therapy was necessary for a patient co-infected with genotypes 1b+5a in this study. This is an important observation which needs to be considered in patients infected with a mixture of poor and good responder subtypes. This has reference to the genotyping method used to detect genotypes prior to therapy as many methods do not detect mixtures. Although our patient numbers are small, we intend to continue this study nationally.

This study has improved data collection and reporting systems, sharing of information with clinicians and scientists locally and abroad, continuous teaching and train-

ing of professionals and our future collaborative effort with the NDoH, is to inform the public about HCV, the available therapies and the risk of transmission.

FUTURE WORK

(1) Continue updating the surveillance system by adding more demographic data to include clinical information and possible risk of transmission as this is not well-defined in the general population of SA.

(2) Continue testing of sentinel groups and include high-risk groups for acquisition of HCV, for eg., prisoners, HIV-infected individuals

(2) From point (1), it will be possible to determine persons with multiple exposures to various genotypes of HCV, for eg. the intravenous drug users (IVDU) and further investigate the likelihood of mixed genotype infections and/or recombinants.

(3) Continue our collaboration with pharmaceutical companies so as to increase the numbers of genotype 5a-infected individuals on therapy and monitor their response on a 24 or 48-week regime.

(4) Identification of acute infections on serology-negative specimens in patients with high index of suspicion.

(5) To conduct practical laboratory-based studies to confirm our prediction studies on conserved and semi-conserved epitopes.

(6) Collaboration with HIV research teams regarding co-infection studies

(7) HLA class I and II typing of prospective patients and of patients on therapy

(8) Sequencing of host genes, like IL-28, to determine polymorphisms, KIR-typing and its association with HLA-C alleles to determine the influence on disease and/or response to therapy in genotype1 and 5a infections.

APPENDIX

Ethics

The study was approved by the ethics committee of the University of the Witwatersrand, Johannesburg, South Africa:

Molecular Epidemiology of Hepatitis C virus in South Africa, WITS HREC M051114

Essential Communicable Disease Surveillance activities of the National Institute for Communicable Diseases, WITS HREC M060449

Publications

Prabdial-Sing N, Giangaspero M, Puren AJ, Mahlangu J, Barrow P, Bowyer. SM. (2011) Palindromic-nucleotide substitutions (PNS) of hepatitis C virus genotypes 1 and 5a from South Africa. J Virol Methods. 175(2):272-277.

Prabdial-Sing N, Puren AJ, Mahlangu J, Barrow P and Bowyer SM. (2008). Hepatitis C Viral genotypes in two different patient cohorts in Johannesburg, South Africa. Arch Virol. 153: 2049-2058.

REFERENCES

Abid, K., Pazienza, V., de Gottardi, A., Rubbia-Brandt, L., Conne, B., Pugnale, P., Rossi, C., Mangia, A. and Negro, F. (2005) An in vitro model of hepatitis C virus genotype 3a-associated triglycerides accumulation. J Hepatol 42(5), 744-51.

Accapezzato, D., Fravolini, F., Casciaro, M.A. and Paroli, M. (2002) Hepatitis C flare due to superinfection by genotype 4 in an HCV genotype 1b chronic carrier. Eur J Gastroenterol Hepatol 14(8), 879-81.

Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K. and Miyano, M. (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Structure 7(11), 1417-26.

Aikawa, T., Kojima, M., Onishi, H., Tamura, R., Fukuda, S., Suzuki, T., Tsuda, F., Okamoto, H., Miyakawa, Y. and Mayumi, M. (1996) HLA DRB1 and DQB1 alleles and haplotypes influencing the progression of hepatitis C. J Med Virol 49(4), 274-8.

Akaike, H. (1974) A new look at the statistical model identification. IEEE Transations on Automatic Control 19, 716–723.

Alavian, S.M., Miri, S.M., Keshvari, M., Elizee, P.K., Behnava, B., Tabatabaei, S.V. and Lankarani, K.B. (2009) Distribution of hepatitis C virus genotype in Iranian mult ply transfused patients with thlassemia. Transfusion 49(10), 2195-9.

Alkharsah, K.R., Dedicoat, M., Blasczyk, R., Newton, R. and Schulz, T.F. (2007) Influence of HLA alleles on shedding of Kaposi sarcoma-associated herpesvirus in saliva in an African population. J Infect Dis 195(6), 809-16.

Alotte, C., Martin, A., Caldarelli, S.A., Di Giorgio, A., Condom, R., Zoulim, F., Durantel, D. and Hantz, O. (2008) Short peptide nucleic acids (PNA) inhibit hepatitis C virus internal ribosome entry site (IRES) dependent translation in vitro. Antiviral Res 80(3), 280-7.

Alt, M., Eisenhardt, S., Serwe, M., Renz, R., Engels, J.W. and Caselmann, W.H. (1999) Comparative inhibitory potential of differently modified antisense oligodeoxynucleotides on hepatitis C virus translation. Eur J Clin Invest 29(10), 868-76.

Alter, M.J. (2006) Epidemiology of viral hepatitis and HIV co-infection. J Hepatol 44 (1 Suppl), S6-9.

Altfeld, M., Addo, M.M., Rosenberg, E.S., Hecht, F.M., Lee, P.K., Vogel, M., Yu, X.G., Draenert, R., Johnston, M.N., Strick, D., Allen, T.M., Feeney, M.E., Kahn, J.O., Sekaly, R.P., Levy, J.A., Rockstroh, J.K., Goulder, P.J. and Walker, B.D. (2003) Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. Aids 17(18), 2581-91.

Amin, J., Kaye, M., Skidmore, S., Pillay, D., Cooper, D.A. and Dore, G.J. (2004) HIV and hepatitis C co-infection within the CAESAR study. HIV Med 5(3), 174-9.

Andonov, A. and Chaudhary, R.K. (1995) Subtyping of hepatitis C virus isolates by a line probe assay using hybridization. J Clin Microbiol 33(1), 254-6.

Antaki, N., Craxi, A., Kamal, S., Moucari, R., Van der Merwe, S., Haffar, S., Gadano, A., Zein, N., Lai, C.L., Pawlotsky, J.M., Heathcote, E.J., Dusheiko, G. and Marcellin, P. (2010) The neglected hepatitis C virus genotypes 4, 5 and 6: an international consensus report. Liver Int 30(3), 342-55.

Antaki, N., Haddad, M., Kebbewar, K., Abdelwahab, J., Hamed, O., Aaraj, R., Alhaj, N., Haffar, S., Assil, M., Ftayeh, M., Assaad, F., Doghman, D., Ali, T., Nassereldine, M., Ali, A. and Antaki, F. (2009) The unexpected discovery of a focus of hepatitis C virus genotype 5 in a Syrian province. Epidemiol Infect 137(1), 79-84.

Antaki, N., Hermes, A., Hadad, M., Ftayeh, M., Antaki, F., Abdo, N. and Kebbewar, K. (2008) Efficacy of interferon plus ribavirin in the treatment of hepatitis C virus genotype 5. J Viral Hepat 15(5), 383-6.

Aoki, H., Yamamoto, T., Kajino, K., Arakawa, Y. and Hino, O. (1996) Molecular diagnosis of hepatitis C viral infection. Intervirology 39(3), 140-4.

Armstrong, G.L., Wasley, A., Simard, E.P., McQuillan, G.M., Kuhnert, W.L. and Alter, M.J. (2006) The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. Ann Intern Med 144(10), 705-14.

Barria, M.I., Gonzalez, A., Vera-Otarola, J., Leon, U., Vollrath, V., Marsac, D., Monasterio, O., Perez-Acle, T., Soza, A. and Lopez-Lastra, M. (2009) Analysis of natural variants of the hepatitis C virus internal ribosome entry site reveals that primary sequence plays a key role in cap-independent translation. Nucleic Acids Res 37(3), 957-71.

Bartenschlager, R. and Lohmann, V. (2000) Replication of the hepatitis C virus. Baillieres Best Pract Res Clin Gastroenterol 14(2), 241-54.

Bartenschlager, R., Lohmann, V., Wilkinson, T. and Koch, J.O. (1995) Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for polyprotein maturation. J Virol 69(12), 7519-28.

Barth, H., Schafer, C., Adah, M.I., Zhang, F., Linhardt, R.J., Toyoda, H., Kinoshita-Toyoda, A., Toida, T., Van Kuppevelt, T.H., Depla, E., Von Weizsacker, F., Blum, H.E. and Baumert, T.F. (2003) Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate. J Biol Chem 278(42), 41003-12.

Battegay, M., Moskophidis, D., Waldner, H., Brundler, M.A., Fung-Leung, W.P., Mak, T.W., Hengartner, H. and Zinkernagel, R.M. (1993) Impairment and delay of neutralizing antiviral antibody responses by virus-specific cytotoxic T cells. J Immunol 151(10), 5408-15.

Baumert, T.F., Ito, S., Wong, D.T. and Liang, T.J. (1998) Hepatitis C virus structural proteins assemble into viruslike particles in insect cells. J Virol 72(5), 3827-36.

Bedossa, P. and Poynard, T. (1996) An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. Hepatology 24(2), 289 -93.

Bengsch B, Thimme.R., Blum HE. (2009) Role of host genetic factors in the outcome of Hepatitis C virus infection. Viruses 1(www.mdpi.com/journal/viruses), 104-125.

Blackard, J.T. and Sherman, K.E. (2007) Hepatitis C virus co-infection and superinfection. J Infect Dis 195(4), 519-24.

Blanchard, E., Belouzard, S., Goueslain, L., Wakita, T., Dubuisson, J., Wychowski, C. and Rouille, Y. (2006) Hepatitis C virus entry depends on clathrin-mediated endocytosis. J Virol 80(14), 6964-72.

Blanchard, E., Brand, D., Trassard, S., Goudeau, A. and Roingeard, P. (2002) Hepatitis C virus-like particle morphogenesis. J Virol 76(8), 4073-9.

Bochud, P.Y., Cai, T., Overbeck, K., Bochud, M., Dufour, J.F., Mullhaupt, B., Borovicka, J., Heim, M., Moradpour, D., Cerny, A., Malinverni, R., Francioli, P. and Negro, F. (2009) Genotype 3 is associated with accelerated fibrosis progression in chronic hepatitis C. J Hepatol 51(4), 655-66.

Bonny, C., Fontaine, H., Poynard, T., Hezode, C., Larrey, D., Marcellin, P., Bourliere, M., Bronowicki, J.P., Merle, P., Zarski, J.P., Sapey, T., Guillemard, C., Ughetto, S., Henquell, C., Nicolas, C., Roche, C., Randl, K., Bommelaer, G. and Abergel, A. (2006) Effectiveness of interferon plus ribavirin combination in the treatment of naive patients with hepatitis C virus type 5. A French multicentre retrospective study. Aliment Pharmacol Ther 24(4), 593-600.

Borman, A.M., Bailly, J.L., Girard, M. and Kean, K.M. (1995) Picornavirus internal ribosome entry segments: comparison of translation efficiency and the requirements for optimal internal initiation of translation in vitro. Nucleic Acids Res 23(18), 3656-63.

Botarelli, P., Brunetto, M.R., Minutello, M.A., Calvo, P., Unutmaz, D., Weiner, A.J., Choo, Q.L., Shuster, J.R., Kuo, G., Bonino, F. and et al. (1993) T-lymphocyte response to hepatitis C virus in different clinical courses of infection. Gastroenterology 104(2), 580-7.

Botha JF, K.C., Schneider HR, Song E, Spearman. (2010) South African Hepatitis C Management Guidelines 2010. The SA Gastroenterol Rev(April 2010).

Botha JF, K.C., Schneider HR, Song E, Spearman and CWN, V.d.M.S. (2005) South African hepatitis C management guidelines. The SA Gastroenterol Rev 3, 18-21.

Boulant, S., Targett-Adams, P. and McLauchlan, J. (2007) Disrupting the association of hepatitis C virus core protein with lipid droplets correlates with a loss in production of infectious virus. J Gen Virol 88(Pt 8), 2204-13.

Bourliere, M., Barberin, J.M., Rotily, M., Guagliardo, V., Portal, I., Lecomte, L., Benali, S., Boustiere, C., Perrier, H., Jullien, M., Lambot, G., Loyer, R., LeBars, O., Daniel, R., Khiri, H. and Halfon, P. (2002) Epidemiological changes in hepatitis C virus genotypes in France: evidence in intravenous drug users. J Viral Hepat 9(1), 62-70.

Bowen, D.G. and Walker, C.M. (2005) Adaptive immune responses in acute and chronic hepatitis C virus infection. Nature 436(7053), 946-52.

Branch, A.D. and Seeff, L.B. (2000) Hepatitis C: State of the art at the Millenium. Thieme publishers. http://books.google.co.za/books.html [accessed 09.07.2008]

Branch, A.D., Stump, D.D., Gutierrez, J.A., Eng, F. and Walewski, J.L. (2005) The hepatitis C virus alternate reading frame (ARF) and its family of novel products: the alternate reading frame protein/F-protein, the double-frameshift protein, and others. Semin Liver Dis 25, 105-17.

Brohm, C., Steinmann, E., Friesland, M., Lorenz, I.C., Patel, A., Penin, F., Bartenschlager, R. and Pietschmann, T. (2009) Characterization of determinants important for hepatitis C virus p7 function in morphogenesis by using trans-complementation. J Virol 83(22), 11682-93.

Brook, G., Main, J., Nelson, M., Bhagani, S., Wilkins, E., Leen, C., Fisher, M., Gilleece, Y., Gilson, R., Freedman, A., Kulasegaram, R., Agarwal, K., Sabin, C. and Deacon-Adams, C. (2010) British HIV Association guidelines for the management of co-infection with HIV-1 and hepatitis B or C virus 2010. HIV Med 11(1), 1-30.

Bukh, J. (2004) A critical role for the chimpanzee model in the study of hepatitis C. Hepatology 39(6), 1469-75.

Bukh, J., Apgar, C.L., Engle, R., Govindarajan, S., Hegerich, P.A., Tellier, R., Wong, D.C., Elkins, R. and Kew, M.C. (1998) Experimental infection of chimpanzees with hepatitis C virus of genotype 5a: genetic analysis of the virus and generation of a standardized challenge pool. J Infect Dis 178(4), 1193-7.

Bukh, J., Miller, R.H. and Purcell, R.H. (1995) Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. Semin Liver Dis 15(1), 41-63.

Bukh, J., Purcell, R.H. and Miller, R.H. (1994) Sequence analysis of the core gene of 14 hepatitis C virus genotypes. Proc Natl Acad Sci U S A 91(17), 8239-43.

Buoro, S., Pizzighella, S., Boschetto, R., Pellizzari, L., Cusan, M., Bonaguro, R., Mengoli, C., Caudai, C., Padula, M., Egisto Valensin, P. and Palu, G. (1999) Typing of hepatitis C virus by a new method based on restriction fragment length polymorphism. Intervirology 42(1), 1-8.

Burlone, M.E. and Budkowska, A. (2009) Hepatitis C virus cell entry: role of lipoproteins and cellular receptors. J Gen Virol 90(Pt 5), 1055-70.

Cassidy, M.J., Jankelson, D., Becker, M., Dunne, T., Walzl, G. and Moosa, M.R.

(1995) The prevalence of antibodies to hepatitis C virus at two haemodialysis units in South Africa. S Afr Med J 85(10), 996-8.

Cassim, B., Mody, G.M., Deenadayalu, V.K. and Hammond, M.G. (1994) Gout in black South Africans: a clinical and genetic study. Ann Rheum Dis 53(11), 759-62.

Cavicchi, J.R. and Kowalski,S.P., (2009) http://www.wipo.int/patentscope/en/programs/patent_landscapes/documents itti patent ls hiv protein vaccines.pdf [accessed 11-07-2011]

CDC, Centre for Disease Control and Prevention. (2008) CDC HIV Statistics http://www.cdc.gov/hiv.html [accessed 12-05-2008]

CDC, Centre for Disease Control and Prevention. (2005) Hepatitis http://www.cdc.gov/hepatitis [accessed 06-06-2008]

Cerny, A., McHutchison, J.G., Pasquinelli, C., Brown, M.E., Brothers, M.A., Grabscheid, B., Fowler, P., Houghton, M. and Chisari, F.V. (1995) Cytotoxic T lymphocyte response to hepatitis C virus-derived peptides containing the HLA A2.1 binding motif. J Clin Invest 95(2), 521-30.

Chamberlain, R.W., Adams, N.J., Taylor, L.A., Simmonds, P. and Elliott, R.M. (1997) The complete coding sequence of hepatitis C virus genotype 5a, the predominant genotype in South Africa. Biochem Biophys Res Commun 236(1), 44-9.

Chan, S.W., McOmish, F., Holmes, E.C., Dow, B., Peutherer, J.F., Follett, E., Yap, P.L. and Simmonds, P. (1992) Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants. J Gen Virol 73 (Pt 5), 1131-41.

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

Chemello, L., Alberti, A., Rose, K. and Simmonds, P. (1994) Hepatitis C serotype and response to interferon therapy. N Engl J Med 330(2), 143.

Chen, Z. and Weck, K.E. (2002) Hepatitis C virus genotyping: interrogation of the 5' untranslated region cannot accurately distinguish genotypes 1a and 1b. J Clin Microbiol 40(9), 3127-34.

Chevaliez, S. and Pawlotsky, J.M. (2007a) Hepatitis C virus: virology, diagnosis and management of antiviral therapy. World J Gastroenterol 13(17), 2461-6.

Chevaliez, S. and Pawlotsky, J.M. (2007b) Practical use of hepatitis C virus kinetics monitoring in the treatment of chronic hepatitis C. J Viral Hepat 14 Suppl 1, 77-81.

Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W. and Houghton, M. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244(4902), 359-62.

Choo, Q.L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C.,

Coit, D., Medina-Selby, R., Barr, P.J. and et al. (1991) Genetic organization and diversity of the hepatitis C virus. Proc Natl Acad Sci U S A 88(6), 2451-5.

Chuang, W.C., Sarkodie, F., Brown, C.J., Owusu-Ofori, S., Brown, J., Li, C., Navarrete, C., Klenerman, P. and Allain, J.P. (2007) Protective effect of HLA-B57 on HCV genotype 2 infection in a West African population. J Med Virol 79(6), 724-33.

Chung, R.T., Andersen, J., Volberding, P., Robbins, G.K., Liu, T., Sherman, K.E., Peters, M.G., Koziel, M.J., Bhan, A.K., Alston, B., Colquhoun, D., Nevin, T., Harb, G. and van der Horst, C. (2004) Peginterferon Alfa-2a plus ribavirin versus interferon alfa-2a plus ribavirin for chronic hepatitis C in HIV-coinfected persons. N Engl J Med 351(5), 451-9.

Ciczora, Y., Callens, N., Penin, F., Pecheur, E.I. and Dubuisson, J. (2007) Transmembrane domains of hepatitis C virus envelope glycoproteins: residues involved in E1E2 heterodimerization and involvement of these domains in virus entry. J Virol 81 (5), 2372-81.

Ciotti, M., Marcuccilli, F., Guenci, T., Babakir-Mina, M., Chiodo, F., Favarato, M. and Perno, C.F. (2010) A multicenter evaluation of the Abbott RealTime HCV Genotype II assay. J Virol Methods 167, 205-7.

Cocquerel, L., Meunier, J.C., Op de Beeck, A., Bonte, D., Wychowski, C. and Dubuisson, J. (2001) Coexpression of hepatitis C virus envelope proteins E1 and E2 in cis improves the stability of membrane insertion of E2. J Gen Virol 82(Pt 7), 1629-35.

Coelmont, L., Paeshuyse, J., Windisch, M.P., De Clercq, E., Bartenschlager, R. and Neyts, J. (2006) Ribavirin antagonizes the in vitro anti-hepatitis C virus activity of 2'-C-methylcytidine, the active component of valopicitabine. Antimicrob Agents Chemother 50(10), 3444-6.

Coetzee, V. (2006) HLA frequencies http://www.allelefrequencies.net [accessed 01-02-2008]

Colina, R., Casane, D., Vasquez, S., Garcia-Aguirre, L., Chunga, A., Romero, H., Khan, B. and Cristina, J. (2004) Evidence of intratypic recombination in natural populations of hepatitis C virus. J Gen Virol 85(Pt 1), 31-7.

Combet, C., Garnier, N., Charavay, C., Grando, D., Crisan, D., Lopez, J., Dehne-Garcia, A., Geourjon, C., Bettler, E., Hulo, C., Le Mercier, P., Bartenschlager, R., Diepolder, H., Moradpour, D., Pawlotsky, J.M., Rice, C.M., Trepo, C., Penin, F. and Deleage, G. (2007) euHCVdb: the European hepatitis C virus database. Nucleic Acids Res 35(Database issue), D363-6.

Corey, K.E., Mendez-Navarro, J., Gorospe, E.C., Zheng, H. and Chung, R.T. (2010) Early treatment improves outcomes in acute hepatitis C virus infection: a meta-analysis. J Viral Hepat 17(3), 201-7.

Cristina, J. and Colina, R. (2006) Evidence of structural genomic region recombina-

tion in Hepatitis C virus. Virol J 3, 53.

Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. Genome Res 14(6), 1188-90.

Davidson, F., Simmonds, P., Ferguson, J.C., Jarvis, L.M., Dow, B.C., Follett, E.A., Seed, C.R., Krusius, T., Lin, C., Medgyesi, G.A. and et al. (1995) Survey of major genotypes and subtypes of hepatitis C virus using RFLP of sequences amplified from the 5' non-coding region. J Gen Virol 76 (Pt 5), 1197-204.

Davis, G.L. (1999) Hepatitis C virus genotypes and quasispecies. Am J Med 107 (6B), 21S-26S.

Day, C.L., Lauer, G.M., Robbins, G.K., McGovern, B., Wurcel, A.G., Gandhi, R.T., Chung, R.T. and Walker, B.D. (2002) Broad specificity of virus-specific CD4+ T-helper-cell responses in resolved hepatitis C virus infection. J Virol 76(24), 12584-95.

D'Heygere F, G.C., Nevens F. (2005) Patients infected with HCV-5 present the same response rate as patients infected with HCV-1: results from the Belgian randomized trial for naive and relapsers (BERNAR-1) abstract. J Hepatol 42, 203A.

Diepolder, H.M., Gerlach, J.T., Zachoval, R., Hoffmann, R.M., Jung, M.C., Wierenga, E.A., Scholz, S., Santantonio, T., Houghton, M., Southwood, S., Sette, A. and Pape, G.R. (1997) Immunodominant CD4+ T-cell epitope within nonstructural protein 3 in acute hepatitis C virus infection. J Virol 71(8), 6011-9.

Donaldson, P.T. (2004) Genetics of liver disease: immunogenetics and disease pathogenesis. Gut 53(4), 599-608.

Donlin, M.J., Cannon, N.A., Yao, E., Li, J., Wahed, A., Taylor, M.W., Belle, S.H., Di Bisceglie, A.M., Aurora, R. and Tavis, J.E. (2007) Pretreatment sequence diversity differences in the full-length hepatitis C virus open reading frame correlate with early response to therapy. J Virol 81(15), 8211-24.

Donnes, P. and Elofsson, A. (2002) Prediction of MHC class I binding peptides, using SVMHC. BMC Bioinformatics 3, 25.

du Toit, E.D., MacGregor, K.J., Taljaard, D.G. and Oudshoorn, M. (1988) HLA-A, B, C, DR and DQ polymorphisms in three South African population groups: South African Negroes, Cape Coloureds and South African Caucasoids. Tissue Antigens 31 (3), 109-25.

Egger, D., Wolk, B., Gosert, R., Bianchi, L., Blum, H.E., Moradpour, D. and Bienz, K. (2002) Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. J Virol 76(12), 5974-84.

EHRN, Eurasian Harm Reduction Network. (2007). http://www.harm-reduction.org/hepatitis-c.html. [accessed 08-10-2010]

Elder, K., Ribes, J. and Baker, D. (2004) Infections, infertility and assisted reproduction, 332-334 pp. Cambridge University Press.http://www.books.google.co.za [accessed 21-02-2008]

Ellis, L.A., Brown, D., Conradie, J.D., Paterson, A., Sher, R., Millo, J., Theodossiadou, E. and Dusheiko, G.M. (1990) Prevalence of hepatitis C in South Africa: detection of anti-HCV in recent and stored serum. J Med Virol 32(4), 249-51.

Encke, J., Radunz, W., Eisenbach, C., Geib, J., Gehrke, S., Pfaff, E. and Stremmel, W. (2007) Development of a heterologous, multigenotype vaccine against hepatitis C virus infection. Eur J Clin Invest 37(5), 396-406.

Enomoto, N., Sakuma, I., Asahina, Y., Kurosaki, M., Murakami, T., Yamamoto, C., Izumi, N., Marumo, F. and Sato, C. (1995) Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. J Clin Invest 96(1), 224-30.

Erhardt, A., Deterding, K., Benhamou, Y., Reiser, M., Forns, X., Pol, S., Calleja, J.L., Ross, S., Spangenberg, H.C., Garcia-Samaniego, J., Fuchs, M., Enriquez, J., Wiegand, J., Stern, J., Wu, K., Kukolj, G., Marquis, M., Beaulieu, P., Nehmiz, G. and Steffgen, J. (2009) Safety, pharmacokinetics and antiviral effect of BILB 1941, a novel hepatitis C virus RNA polymerase inhibitor, after 5 days oral treatment. Antivir Ther 14(1), 23-32.

Evans, M.J., von Hahn, T., Tscherne, D.M., Syder, A.J., Panis, M., Wolk, B., Hatziioannou, T., McKeating, J.A., Bieniasz, P.D. and Rice, C.M. (2007) Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. Nature 446(7137), 801-5.

Fan, X., Xu, Y. and Di Bisceglie, A.M. (2006) Efficient amplification and cloning of near full-length hepatitis C virus genome from clinical samples. Biochem Biophys Res Commun 346(4), 1163-72.

Farci, P., Alter, H.J., Govindarajan, S., Wong, D.C., Engle, R., Lesniewski, R.R., Mushahwar, I.K., Desai, S.M., Miller, R.H., Ogata, N. and et al. (1992) Lack of protective immunity against reinfection with hepatitis C virus. Science 258(5079), 135-40.

Farci, P., Orgiana, G. and Purcell, R.H. (1995) Immunity elicited by hepatitis C virus. Clin Exp Rheumatol 13 Suppl 13, S9-12.

Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J.C., Strazzera, A., Chien, D.Y., Munoz, S.J., Balestrieri, A., Purcell, R.H. and Alter, H.J. (2000) The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. Science 288(5464), 339-44.

Feinstone, S.M., Kapikian, A.Z., Purcell, R.H., Alter, H.J. and Holland, P.V. (1975) Transfusion-associated hepatitis not due to viral hepatitis type A or B. N Engl J Med 292(15), 767-70.

Forton, D.M., Karayiannis, P., Mahmud, N., Taylor-Robinson, S.D. and Thomas, H.C. (2004) Identification of unique hepatitis C virus quasispecies in the central nervous system and comparative analysis of internal translational efficiency of brain, liver, and serum variants. J Virol 78(10), 5170-83.

Foster, G.R. (2009) Recent advances in viral hepatitis. Clin Med 9(6), 613-6.

Franciscus, A. (2010) Hepatitis C treatments in current development. http://www.hcvadvocate.org/hepatitis/hepC/HCVDrugs.html. [accessed 03-11-2010]

Fraser, C.S. and Doudna, J.A. (2007) Structural and mechanistic insights into hepatitis C viral translation initiation. Nat rev Microbiol 5(1), 29-38.

Freeman, A.J., Dore, G.J., Law, M.G., Thorpe, M., Von Overbeck, J., Lloyd, A.R., Marinos, G. and Kaldor, J.M. (2001) Estimating progression to cirrhosis in chronic hepatitis C virus infection. Hepatology 34(4 Pt 1), 809-16.

Frelin, L., Alheim, M., Chen, A., Soderholm, J., Rozell, B., Barnfield, C., Liljestrom, P. and Sallberg, M. (2003) Low dose and gene gun immunization with a hepatitis C virus nonstructural (NS) 3 DNA-based vaccine containing NS4A inhibit NS3/4A-expressing tumors in vivo. Gene Ther 10(8), 686-99.

Friebe, P. and Bartenschlager, R. (2002) Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication. J Virol 76(11), 5326-38.

Friebe, P. and Bartenschlager, R. (2009) Role of RNA structures in genome terminal sequences of the hepatitis C virus for replication and assembly. J Virol 83(22), 11989-95.

Gale, M.J., Jr., Korth, M.J. and Katze, M.G. (1998) Repression of the PKR protein kinase by the hepatitis C virus NS5A protein: a potential mechanism of interferon resistance. Clin Diagn Virol 10(2-3), 157-62.

Ge, D., Fellay, J., Thompson, A.J., Simon, J.S., Shianna, K.V., Urban, T.J., Heinzen, E.L., Qiu, P., Bertelsen, A.H., Muir, A.J., Sulkowski, M., McHutchison, J.G. and Goldstein, D.B. (2009) Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 461(7262), 399-401.

Gerlach, J.T., Diepolder, H.M., Zachoval, R., Gruener, N.H., Jung, M.C., Ulsenheimer, A., Schraut, W.W., Schirren, C.A., Waechtler, M., Backmund, M. and Pape, G.R. (2003) Acute hepatitis C: high rate of both spontaneous and treatment-induced viral clearance. Gastroenterology 125(1), 80-8.

Gerotto, M., Dal Pero, F., Sullivan, D.G., Chemello, L., Cavalletto, L., Polyak, S.J., Pontisso, P., Gretch, D.R. and Alberti, A. (1999) Evidence for sequence selection within the non-structural 5A gene of hepatitis C virus type 1b during unsuccessful treatment with interferon-alpha. J Viral Hepat 6(5), 367-72.

Ghany, M.G., Strader, D.B., Thomas, D.L. and Seeff, L.B. (2009) Diagnosis, man-

agement, and treatment of hepatitis C: an update. Hepatology 49(4), 1335-74.

Giangaspero, M., Harasawa, R. and Zanetti, A. (2008) Taxonomy of genus Hepacivirus. Application of palindromic nucleotide substitutions for the determination of genotypes of human hepatitis C virus species. J Virol Methods 153(2), 280-99.

Goedert, J.J., Eyster, M.E., Lederman, M.M., Mandalaki, T., De Moerloose, P., White, G.C., 2nd, Angiolillo, A.L., Luban, N.L., Sherman, K.E., Manco-Johnson, M., Preiss, L., Leissinger, C., Kessler, C.M., Cohen, A.R., DiMichele, D., Hilgartner, M.W., Aledort, L.M., Kroner, B.L., Rosenberg, P.S. and Hatzakis, A. (2002) End-stage liver disease in persons with hemophilia and transfusion-associated infections. Blood 100(5), 1584-9.

Gonzalez-Perez, I., Gonzalez Gonzalez, Y.J., Vina-Rodriguez, A., Armas Cayarga, A. and Solis, R.L. (2004) The usefulness of Umelosa hepatitis C virus qualitative kit as supplemental test for confirmation of hepatitis C virus infection. Rev Soc Bras Med Trop 37(1), 25-7.

Gottwein, J.M. and Bukh, J. (2008) Cutting the gordian knot-development and biological relevance of hepatitis C virus cell culture systems. Adv Virus Res 71, 51-133.

Graci, J.D. and Cameron, C.E. (2006) Mechanisms of action of ribavirin against distinct viruses. Rev Med Virol 16(1), 37-48.

Graham, C.S., Baden, L.R., Yu, E., Mrus, J.M., Carnie, J., Heeren, T. and Koziel, M.J. (2001) Influence of human immunodeficiency virus infection on the course of hepatitis C virus infection: a meta-analysis. Clin Infect Dis 33(4), 562-9.

Grakoui, A., McCourt, D.W., Wychowski, C., Feinstone, S.M. and Rice, C.M. (1993) Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. J Virol 67(5), 2832-43.

Grakoui, A., Shoukry, N.H., Woollard, D.J., Han, J.H., Hanson, H.L., Ghrayeb, J., Murthy, K.K., Rice, C.M. and Walker, C.M. (2003) HCV persistence and immune evasion in the absence of memory T cell help. Science 302(5645), 659-62.

Gretch, D., Lee, W. and Corey, L. (1992) Use of aminotransferase, hepatitis C antibody, and hepatitis C polymerase chain reaction RNA assays to establish the diagnosis of hepatitis C virus infection in a diagnostic virology laboratory. J Clin Microbiol 30(8), 2145-9.

Gretch, D.R., dela Rosa, C., Carithers, R.L., Jr., Willson, R.A., Williams, B. and Corey, L. (1995) Assessment of hepatitis C viremia using molecular amplification technologies: correlations and clinical implications. Ann Intern Med 123(5), 321-9.

Gretch, D.R., Polyak, S.J., Wilson, J.J., Carithers, R.L., Jr., Perkins, J.D. and Corey, L. (1996) Tracking hepatitis C virus quasispecies major and minor variants in symptomatic and asymptomatic liver transplant recipients. J Virol 70(11), 7622-31.

Griffin, S.D., Beales, L.P., Clarke, D.S., Worsfold, O., Evans, S.D., Jaeger, J., Har-

ris, M.P. and Rowlands, D.J. (2003) The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. FEBS Lett 535(1-3), 34-8.

Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52(5), 696-704.

Halfon, P., Trimoulet, P., Bourliere, M., Khiri, H., de Ledinghen, V., Couzigou, P., Feryn, J.M., Alcaraz, P., Renou, C., Fleury, H.J. and Ouzan, D. (2001) Hepatitis C virus genotyping based on 5' noncoding sequence analysis (Trugene). J Clin Microbiol 39(5), 1771-3.

Hall, T. (1997) BioEdit. http://www.mbio.ncsu.edu. [accessed 02-05-2008]

Hammond, M.G., Asmal, A.C. and Omar, M.A. (1980) HLA and insulin-dependent diabetes in South African negroes. Diabetologia 19(2), 101-2.

Hammond, M.G. and Moshal, M.G. (1980) HLA and duodenal ulcer in South African Indians. Tissue Antigens 15(5), 508-9.

Harasawa, R. and Giangaspero, M. (1998) A novel method for pestivirus genotyping based on palindromic nucleotide substitutions in the 5'-untranslated region. J Virol Methods 70(2), 225-30.

Harcourt, G.C., Lucas, M., Sheridan, I., Barnes, E., Phillips, R. and Klenerman, P. (2004) Longitudinal mapping of protective CD4+ T cell responses against HCV: analysis of fluctuating dominant and subdominant HLA-DR11restricted epitopes. J Viral Hepat 11(4), 324-31.

Harcourt, G.C., Hellier, S., Bunce, M., Satsangi, J., Collier, J., Chapman, R., Philips, R., and Klenerman, P. (2001) Effect of HLA class II genotype on T helper lymphocyte responses and viral control in hepatitis C virus infection. J Viral Hepat 8(3):174-9.

Hasegawa, M., Kishino, H. and Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol 22(2), 160-174.

Health Protection Agency, HPA, United Kingdom (2009), http://www.hpa.org.uk [accessed 12-10-2010]

Henderson, L.E. (2001) Diagrammatic 3D model depicting the structure of the hepatitis C virus published online, March 2001, www.prn.org [accessed 02-10-2010]

Henquell, C., Cartau, C., Abergel, A., Laurichesse, H., Regagnon, C., De Champs, C., Bailly, J.L. and Peigue-Lafeuille, H. (2004) High prevalence of hepatitis C virus type 5 in central France evidenced by a prospective study from 1996 to 2002. J Clin Microbiol 42(7), 3030-5.

Herrmann, E., Neumann, A.U., Schmidt, J.M. and Zeuzem, S. (2000) Hepatitis C virus kinetics. Antivir Ther 5(2), 85-90.

Higgins, D.G. (1994) CLUSTAL V: multiple alignment of DNA and protein sequences. Methods Mol Biol 25, 307-18.

Himoudi, N., Abraham, J.D., Fournillier, A., Lone, Y.C., Joubert, A., Op De Beeck, A., Freida, D., Lemonnier, F., Kieny, M.P. and Inchauspe, G. (2002) Comparative vaccine studies in HLA-A2.1-transgenic mice reveal a clustered organization of epitopes presented in hepatitis C virus natural infection. J Virol 76(24), 12735-46.

Hissar, S.S., Goyal, A., Kumar, M., Pandey, C., Suneetha, P.V., Sood, A., Midha, V., Sakhuja, P., Malhotra, V. and Sarin, S.K. (2006) Hepatitis C virus genotype 3 predominates in North and Central India and is associated with significant histopathologic liver disease. J Med Virol 78(4), 452-8.

Hohler, T., Gerken, G., Notghi, A., Lubjuhn, R., Taheri, H., Protzer, U., Lohr, H.F., Schneider, P.M., Meyer zum Buschenfelde, K.H. and Rittner, C. (1997) HLA-DRB1*1301 and *1302 protect against chronic hepatitis B. J Hepatol 26(3), 503-7.

Honda, M., Kaneko, S., Shimazaki, T., Matsushita, E., Kobayashi, K., Ping, L.H., Zhang, H.C. and Lemon, S.M. (2000) Hepatitis C virus core protein induces apoptosis and impairs cell-cycle regulation in stably transformed Chinese hamster ovary cells. Hepatology 31(6), 1351-9.

Honda, M., Ping, L.H., Rijnbrand, R.C., Amphlett, E., Clarke, B., Rowlands, D. and Lemon, S.M. (1996) Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. Virology 222(1), 31-42.

Hui, C.K., Yuen, M.F., Sablon, E., Chan, A.O., Wong, B.C. and Lai, C.L. (2003) Interferon and ribavirin therapy for chronic hepatitis C virus genotype 6: a comparison with genotype 1. J Infect Dis 187(7), 1071-4.

ICTVdb, The International Committee on Taxonomy of Viruses. (2006) Hepatitis C virus In: ICTVdb-The Universal Virus Database, version 4. http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdb, edited by B.-O. C. Colombia University, New York USA.http://www.ictvdb.org [accessed 23-03-2009]

Idrees, M. and Riazuddin, S. (2008) Frequency distribution of hepatitis C virus genotypes in different geographical regions of Pakistan and their possible routes of transmission. BMC Infect Dis 8, 69.

IEDB, Immune Epitope Database. (2009) version 2.0 http://tools.immuneepitope.org/main/jsp/menu.jsp. [accessed 09-09-2009]

Ishak, K., Baptista, A., Bianchi, L., Callea, F., De Groote, J., Gudat, F., Denk, H., Desmet, V., Korb, G., MacSween, R.N. and et al. (1995) Histological grading and staging of chronic hepatitis. J Hepatol 22(6), 696-9.

Jamma, S., Hussain, G. and Lau, D.T-Y. (2010) Current concepts of HBV/HCV coinfection: coexistence but not necessarily in harmony. Curr. Hepat. Rep 9(4), 260-269.

Jukes, T.H. and Cantor C.R. (1969) Recent advances in studies of evolutionary relationships between proteins and nucleic acids. Space Life Sci 1(4), 469-90.

Julg, B, Moodely.E., Qi Y, Bishop K, Goulder P, Detels R, Ndung T, Walker B and Carrington M. (2010) HLA class II allele DRB1*1303 is associated with reduced viral loads in 2 independent cohorts of different ethnicity and HIV clades of infection. 17th Coference on Retroviruses and Opportunitistic Infections.

Kageyama, S., Agdamag, D.M., Alesna, E.T., Leano, P.S., Heredia, A.M., Abellanosa-Tac-An, I.P., Jereza, L.D., Tanimoto, T., Yamamura, J. and Ichimura, H. (2006) A natural inter-genotypic (2b/1b) recombinant of hepatitis C virus in the Philippines. J Med Virol 78(11), 1423-8.

Kaito, M., Watanabe, S., Tsukiyama-Kohara, K., Yamaguchi, K., Kobayashi, Y., Konishi, M., Yokoi, M., Ishida, S., Suzuki, S. and Kohara, M. (1994) Hepatitis C virus particle detected by immunoelectron microscopic study. J Gen Virol 75 (Pt 7), 1755-60.

Kalinina, O., Norder, H., Mukomolov, S. and Magnius, L.O. (2002) A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. J Virol 76(8), 4034-43.

Kamal, S.M., El Tawil, A.A., Nakano, T., He, Q., Rasenack, J., Hakam, S.A., Saleh, W.A., Ismail, A., Aziz, A.A. and Madwar, M.A. (2005) Peginterferon {alpha}-2b and ribavirin therapy in chronic hepatitis C genotype 4: impact of treatment duration and viral kinetics on sustained virological response. Gut 54(6), 858-66.

Kamel, M.A., Ghaffar, Y.A., Wasef, M.A., Wright, M., Clark, L.C. and Miller, F.D. (1992) High HCV prevalence in Egyptian blood donors. Lancet 340(8816), 427.

Kanda, T., Imazeki, F., Yonemitsu, Y., Mikami, S., Takada, N., Nishino, T., Takashi, M., Tsubota, A., Kato, K., Sugiura, N., Tawada, A., Wu, S., Tanaka, T., Nakamoto, S., Mikata, R., Tada, M., Chiba, T., Kurihara, T., Arai, M., Fujiwara, K., Kanai, F. and Yokosuka, O. (2010) Quantification of hepatitis C virus in patients treated with peginterferon-alfa 2a plus ribavirin treatment by COBAS TaqMan HCV test. J Viral Hepat.doi: 10.1111/j.1365-2893.2010.01409

Kanto, T. and Hayashi, N. (2006) Immunopathogenesis of hepatitis C virus infection: multifaceted strategies subverting innate and adaptive immunity. Intern Med 45(4), 183-91.

Kao, J.H., Chen, P.J., Lai, M.Y. and Chen, D.S. (1993) Superinfection of heterologous hepatitis C virus in a patient with chronic type C hepatitis. Gastroenterology 105(2), 583-7.

Kato, N., Abe, K., Mori, K., Ariumi, Y., Dansako, H. and Ikeda, M. (2009) Genetic variability and diversity of intracellular genome-length hepatitis C virus RNA in long-term cell culture. Arch Virol 154(1), 77-85.

Kato, N., Ootsuyama, Y., Ohkoshi, S., Nakazawa, T., Sekiya, H., Hijikata, M. and

Shimotohno, K. (1992) Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus. Biochem Biophys Res Commun 189(1), 119-27.

Kawai, S., Yokosuka, O., Kanda, T., Imazeki, F., Maru, Y. and Saisho, H. (1999) Quantification of hepatitis C virus by TaqMan PCR: comparison with HCV Amplicor Monitor assay. J Med Virol 58(2), 121-6.

Kew, M.C., Houghton, M., Choo, Q.L. and Kuo, G. (1990) Hepatitis C virus antibodies in southern African blacks with hepatocellular carcinoma. Lancet 335, 873-4.

Khakoo, S.I., Thio, C.L., Martin, M.P., Brooks, C.R., Gao, X., Astemborski, J., Cheng, J., Goedert, J.J., Vlahov, D., Hilgartner, M., Cox, S., Little, A.M., Alexander, G.J., Cramp, M.E., O'Brien, S.J., Rosenberg, W.M., Thomas, D.L. and Carrington, M. (2004) HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. Science 305(5685), 872-4.

Kiepiela, P., Leslie, A.J., Honeyborne, I., Ramduth, D., Thobakgale, C., Chetty, S., Rathnavalu, P., Moore, C., Pfafferott, K.J., Hilton, L., Zimbwa, P., Moore, S., Allen, T., Brander, C., Addo, M.M., Altfeld, M., James, I., Mallal, S., Bunce, M., Barber, L.D., Szinger, J., Day, C., Klenerman, P., Mullins, J., Korber, B., Coovadia, H.M., Walker, B.D. and Goulder, P.J. (2004) Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature 432(7018), 769-75.

Kim, J.H., Rerks-Ngarm, S., Excler, J.L. and Michael, N.L. (2010) HIV vaccines: lessons learned and the way forward. Curr Opin HIV AIDS 5, 428-34.

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2), 111-20.

Klade, C.S., Kubitschke, A., Stauber, R.E., Meyer, M.F., Zinke, S., Wiegand, J., Zauner, W., Aslan, N., Lehmann, M., Cornberg, M., Manns, M.P., Reisner, P. and Wedemeyer, H. (2009) Hepatitis C virus-specific T cell responses against conserved regions in recovered patients. Vaccine 27(23), 3099-108.

Kleinman, S.H., Stramer, S.L., Brodsky, J.P., Caglioti, S. and Busch, M.P. (2006) Integration of nucleic acid amplification test results into hepatitis C virus supplemental serologic testing algorithms: implications for donor counseling and revision of existing algorithms. Transfusion 46(5), 695-702.

Klemp, P., du Toit, E.D., Issacs, S., Mody, G.M. and Oudshoorn, M. (1988) HLA-A,B,C, and DR antigens, GLO I and Bf marker profiles in 75 Cape coloured patients with systemic lupus erythematosus (SLE). Tissue Antigens 32(1), 12-6.

Koziel, M.J., Dudley, D., Afdhal, N., Grakoui, A., Rice, C.M., Choo, Q.L., Houghton, M. and Walker, B.D. (1995) HLA class I-restricted cytotoxic T lymphocytes specific for hepatitis C virus. Identification of multiple epitopes and characterization of patterns of cytokine release. J Clin Invest 96(5), 2311-21.

Krajden, M., Ziermann, R., Khan, A., Mak, A., Leung, K., Hendricks, D. and Co-

manor, L. (2002) Qualitative detection of hepatitis C virus RNA: comparison of analytical sensitivity, clinical performance, and workflow of the Cobas Amplicor HCV test version 2.0 and the HCV RNA transcription-mediated amplification qualitative assay. J Clin Microbiol 40(8), 2903-7.

Kuiken, C., Yusim, K., Boykin, L. and Richardson, R. (2005) The Los Alamos hepatitis C sequence database. Bioinformatics 21(3), 379-84.

Kumar, S., Nei, M., Dudley, J. and Tamura, K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9 (4), 299-306.

Kumar, S., Tamura, K. and Nei, M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5(2), 150-63.

Kuniholm, M.H., Kovacs, A., Gao, X., Xue, X., Marti, D., Thio, C.L., Peters, M.G., Terrault, N.A., Greenblatt, R.M., Goedert, J.J., Cohen, M.H., Minkoff, H., Gange, S.J., Anastos, K., Fazzari, M., Harris, T.G., Young, M.A., Strickler, H.D. and Carrington, M. (2010) Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. Hepatology 51(5), 1514-22.

Kurbanov, F., Tanaka, Y., Avazova, D., Khan, A., Sugauchi, F., Kan, N., Kurbanova-Khudayberganova, D., Khikmatullaeva, A., Musabaev, E. and Mizokami, M. (2008) Detection of hepatitis C virus natural recombinant RF1_2k/1b strain among intravenous drug users in Uzbekistan. Hepatol Res 38(5), 457-464.

Kwo, E.C., Kamat, P. and Steinman, M.A. (2009) Physician use of brand versus generic drug names in 1993-1994 and 2003-2004. Ann Pharmacother 43(3), 459-68.

Lamonaca, V., Missale, G., Urbani, S., Pilli, M., Boni, C., Mori, C., Sette, A., Massari, M., Southwood, S., Bertoni, R., Valli, A., Fiaccadori, F. and Ferrari, C. (1999) Conserved hepatitis C virus sequences are highly immunogenic for CD4(+) T cells: implications for vaccine development. Hepatology 30(4), 1088-98.

Landt, O., U. Lass, and H.-H. Feucht. . (2002) Simultaneous identification of five HCVgenotypes in a single reaction. Springer Verlag, Berlin.(In U. Reischl, C. Wittwer et al. (ed.), Rapid Cycle), 217-225.

Lang, K.A., Yan, J., Draghia-Akli, R., Khan, A. and Weiner, D.B. (2008) Strong HCV NS3- and NS4A-specific cellular immune responses induced in mice and Rhesus macaques by a novel HCV genotype 1a/1b consensus DNA vaccine. Vaccine 26 (49), 6225-31.

Langhans, B., Kupfer, B., Braunschweiger, I., Arndt, S., Schulte, W., Nischalke, H.D., Nattermann, J., Oldenburg, J., Sauerbruch, T. and Spengler, U. (2010) Interferon-lambda serum levels in hepatitis C. J Hepatol. doi:10.1016/ j.jhep.2010.08.020

Laperche, S., Saune, K., Deny, P., Duverlie, G., Alain, S., Chaix, M.L., Gaudy, C.,

Lunel, F., Pawlotsky, J.M., Payan, C., Pozzetto, B., Tamalet, C., Thibault, V., Vallet, S., Bouchardeau, F., Izopet, J. and Lefrere, J.J. (2006) Unique NS5b hepatitis C virus gene sequence consensus database is essential for standardization of genotype determinations in multicenter epidemiological studies. J Clin Microbiol 44 (2), 614-6.

Lavanchy, D. (2011) Evolving epidemiology of hepatitis C virus. Clin Microbiol Infect 17, 107-15.

Lechner, F., Gruener, N.H., Urbani, S., Uggeri, J., Santantonio, T., Kammer, A.R., Cerny, A., Phillips, R., Ferrari, C., Pape, G.R. and Klenerman, P. (2000) CD8+ T lymphocyte responses are induced during acute hepatitis C virus infection but are not sustained. Eur J Immunol 30(9), 2479-87.

Lee, S.C., Antony, A., Lee, N., Leibow, J., Yang, J.Q., Soviero, S., Gutekunst, K. and Rosenstraus, M. (2000) Improved version 2.0 qualitative and quantitative AMPLICOR reverse transcription-PCR tests for hepatitis C virus RNA: calibration to international units, enhanced genotype reactivity, and performance characteristics. J Clin Microbiol 38(11), 4171-9.

Legrand-Abravanel, F., Claudinon, J., Nicot, F., Dubois, M., Chapuy-Regaud, S., Sandres-Saune, K., Pasquier, C. and Izopet, J. (2007) New natural intergenotypic (2/5) recombinant of hepatitis C virus. J Virol 81(8), 4357-62.

Legrand-Abravanel, F., Colson, P., Leguillou-Guillemette, H., Alric, L., Ravaux, I., Lunel-Fabiani, F., Bouviers-Alias, M., Trimoulet, P., Chaix, M.L., Hezode, C., Foucher, J., Fontaine, H., Roque-Afonso, A.M., Gassin, M., Schvoerer, E., Gaudy, C., Roche, B., Doffoel, M., D'Alteroche, L., Vallet, S., Baazia, Y., Pozzetto, B., Thibault, V., Nousbaum, J.B., Roulot, D., Coppere, H., Poynard, T., Payan, C. and Izopet, J. (2009a) Influence of the HCV subtype on the virological response to pegylated interferon and ribavirin therapy. J Med Virol 81(12), 2029-35.

Legrand-Abravanel, F., Henquell, C., Le Guillou-Guillemette, H., Balan, V., Mirand, A., Dubois, M., Lunel-Fabiani, F., Payan, C. and Izopet, J. (2009b) Naturally occurring substitutions conferring resistance to hepatitis C virus polymerase inhibitors in treatment-naive patients infected with genotypes 1-5. Antivir Ther 14(5), 723-30.

Legrand-Abravanel, F., Sandres-Saune, K., Barange, K., Alric, L., Moreau, J., Desmorat, P., Vinel, J.P. and Izopet, J. (2004) Hepatitis C virus genotype 5: epidemiological characteristics and sensitivity to combination therapy with interferon -alpha plus ribavirin. J Infect Dis 189(8), 1397-400.

Lemey, P., Salemi, M. and Vandamme, A-M. (2009) The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing. 2nd Edition. Cambridge Univ. Press: Cambridge.

Lerat, H., Shimizu, Y.K. and Lemon, S.M. (2000) Cell type-specific enhancement of hepatitis C virus internal ribosome entry site-directed translation due to 5' nontranslated region substitutions selected during passage of virus in

lymphoblastoid cells. J Virol 74(15), 7024-31.

Li, H., Thomassen, L.V., Majid, A., McMahon, B.J., Bruden, D., McArdle, S., Bano, N., Chung, M., Carithers, R.L., Perkins, J.D., Sullivan, D.G. and Gretch, D.R. (2008) Investigation of putative multisubtype hepatitis C virus infections in vivo by heteroduplex mobility analysis of core/envelope subgenomes. J Virol 82(15), 7524-32.

Liu, S., Yang, W., Shen, L., Turner, J.R., Coyne, C.B. and Wang, T. (2009) Tight junction proteins claudin-1 and occludin control hepatitis C virus entry and are downregulated during infection to prevent superinfection. J Virol 83(4), 2011-4.

Liu, Z. and Hou, J. (2006) Hepatitis B virus (HBV) and Hepatitis C virus (HCV) dual infection. Int. J. Med. Sci 392), 57-62.

Lodenyo, H., Schoub, B., Ally, R., Kairu, S. and Segal, I. (2000) Hepatitis B and C virus infections and liver function in AIDS patients at Chris Hani Baragwanath Hospital, Johannesburg. East Afr Med J 77(1), 13-5.

Lohmann, V., Koch, J.O. and Bartenschlager, R. (1996) Processing pathways of the hepatitis C virus proteins. J Hepatol 24(2 Suppl), 11-9.

Lombard, Z., Dalton, D.L., Venter, P.A., Williams, R.C. and Bornman, L. (2006) Association of HLA-DR, -DQ, and vitamin D receptor alleles and haplotypes with tuberculosis in the Venda of South Africa. Hum Immunol 67(8), 643-54.

Major, M.E., Mihalik, K., Fernandez, J., Seidman, J., Kleiner, D., Kolykhalov, A.A., Rice, C.M. and Feinstone, S.M. (1999) Long-term follow-up of chimpanzees inoculated with the first infectious clone for hepatitis C virus. J Virol 73(4), 3317-25.

Manns, M.P., McHutchison, J.G., Gordon, S.C., Rustgi, V.K., Shiffman, M., Reindollar, R., Goodman, Z.D., Koury, K., Ling, M. and Albrecht, J.K. (2001) Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet 358(9286), 958-65.

Manns, M.P., Wedemeyer, H. and Cornberg, M. (2006) Treating viral hepatitis C: efficacy, side effects, and complications. Gut 55(9), 1350-9.

Marcellin, P. (1999) Hepatitis C: the clinical spectrum of the disease. J Hepatol 31 Suppl 1, 9-16.

Markov, P.V., Pepin, J., Frost, E., Deslandes, S., Labbe, A.C. and Pybus, O.G. (2009) Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. J Gen Virol 90(Pt 9), 2086-96.

Martell, M., Esteban, J.I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J. and Gomez, J. (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. J Virol 66(5), 3225-9.

Martin, P., Simon, B., Lone, Y.C., Chatel, L., Barry, R., Inchauspe, G. and Fournillier, A. (2008) A vector-based minigene vaccine approach results in strong induction of T-cell responses specific of hepatitis C virus. Vaccine 26(20), 2471-81.

Mast, E.E., Alter, M.J. and Margolis, H.S. (1999) Strategies to prevent and control hepatitis B and C virus infections: a global perspective. Vaccine 17(13-14), 1730-3.

McCown, M.F., Rajyaguru, S., Kular, S., Cammack, N. and Najera, I. (2009) GT-1a or GT-1b subtype-specific resistance profiles for hepatitis C virus inhibitors telaprevir and HCV-796. Antimicrob Agents Chemother 53(5), 2129-32.

McKiernan, S.M., Hagan, R., Curry, M., McDonald, G.S., Kelly, A., Nolan, N., Walsh, A., Hegarty, J., Lawlor, E. and Kelleher, D. (2004) Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. Hepatology 40(1), 108-14.

McMahon, J.M., Simm, M., Milano, D. and Clatts, M. (2004) Detection of hepatitis C virus in the nasal secretions of an intranasal drug-user. Ann Clin Microbiol Antimicrob 3, 6.

Memarnejadian, A., Roohvand, F., Arashkia, A., Rafati, S. and Shokrgozar, M.A. (2009) Polytope DNA vaccine development against hepatitis C virus: a streamlined approach from in silico design to in vitro and primary in vivo analyses in BALB/c mice. Protein Pept Lett 16(7), 842-50.

Middleton, D., Williams, F., Meenagh, A., Daar, A.S., Gorodezky, C., Hammond, M., Nascimento, E., Briceno, I. and Perez, M.P. (2000) Analysis of the distribution of HLA-A alleles in populations from five continents. Hum Immunol 61(10), 1048-52.

Minton, E.J., Smillie, D., Neal, K.R., Irving, W.L., Underwood, J.C. and James, V. (1998) Association between MHC class II alleles and clearance of circulating hepatitis C virus. Members of the Trent Hepatitis C Virus Study Group. J Infect Dis 178(1), 39-44.

Moradpour, D., Bieck, E., Hugle, T., Wels, W., Wu, J.Z., Hong, Z., Blum, H.E. and Bartenschlager, R. (2002) Functional properties of a monoclonal antibody inhibiting the hepatitis C virus RNA-dependent RNA polymerase. J Biol Chem 277(1), 593-601.

Moradpour, D., Brass, V., Bieck, E., Friebe, P., Gosert, R., Blum, H.E., Bartenschlager, R., Penin, F. and Lohmann, V. (2004) Membrane association of the RNA-dependent RNA polymerase is essential for hepatitis C virus RNA replication. J Virol 78(23), 13278-84.

Morbidity and Mortality Weekly Reports, M. (1997) Case definitions for infectious conditions under public health surveillance ftp://ftp.cdc.gov/pub/publications/mmwr/rr/rr4610.pdf Centers for Disease Control and Prevention CDC [accessed 19.09.2008]

Morbidity and Mortality Weekly Reports, M. (1998) Tests for Hepatitis C virus infec-

tion http://wwwcdc.gov/mmwr/pdf/rr/rr4719.pdf [accessed 19.09.2008]

Morbidity and Mortality Weekly Reports, M. (2009) Surveillance for Acute viral hepatitis http://www.cdc.gov/mmwr/pdf/ss/ss5803.pdf [accessed 12.03.2010]

Morbidity and Mortality Weekly Reports, M. (2008) Use of enhanced surveillance for hepatitis C virus infection to detect a cluster among young injection-drug users http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5719a3.htm[accessed 02.04.2010]

Moreau, I., Hegarty, S., Levis, J., Sheehy, P., Crosbie, O., Kenny-Walsh, E. and Fanning, L.J. (2006) Serendipitous identification of natural intergenotypic recombinants of hepatitis C in Ireland. Virol J 3, 95.

Moreno, M.P., Casane, D., Lopez, L. and Cristina, J. (2006) Evidence of recombination in quasispecies populations of a Hepatitis C Virus patient undergoing anti-viral therapy. Virol J 3, 87.

Morice, Y., Roulot, D., Grando, V., Stirnemann, J., Gault, E., Jeantils, V., Bentata, M., Jarrousse, B., Lortholary, O., Pallier, C. and Deny, P. (2001) Phylogenetic analyses confirm the high prevalence of hepatitis C virus (HCV) type 4 in the Seine-Saint-Denis district (France) and indicate seven different HCV-4 subtypes linked to two different epidemiological patterns. J Gen Virol 82(Pt 5), 1001-12.

Moriya, K., Fujie, H., Shintani, Y., Yotsuyanagi, H., Tsutsumi, T., Ishibashi, K., Matsuura, Y., Kimura, S., Miyamura, T. and Koike, K. (1998) The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. Nat Med 4(9), 1065-7.

Mosmann T.R. and Sad, S. (1996) The expanding universe of T-cells subsets: Th1, Th2 and more. Immunol Today 17(3), 138-146.

Muerhoff, A.S., Smith, D.B., Leary, T.P., Erker, J.C., Desai, S.M. and Mushahwar, I.K. (1997) Identification of GB virus C variants by phylogenetic analysis of 5'-untranslated and coding region sequences. J Virol 71(9), 6501-8.

Murphy, D., Chamberland, J., Dandavino, R. and Sablon, E. (2007) A new genotype of hepatitis C virus originating from central Africa (abstract). Hepatology 46, 623A.

Murray, C.L., Jones, C.T., Tassello, J. and Rice, C.M. (2007) Alanine scanning of the hepatitis C virus core protein reveals numerous residues essential for production of infectious virus. J Virol 81(19), 10220-31.

National Department of Health, NDoH, Statistics notes http://www.doh.gov.za/facts/index.html, November 2009. [accessed 09.11.2010]

National Guidelines for the Prevention and Control of Hepatitis C virus in South Africa (2010) final draft, 12-36.

National Health Laboratory Services, NHLS Annual report, 2009-2010 http://

www.nhls.ac.za [accessed 09.11.2010]

National Institute of Diabetes, Digestive and kidney disease, NIDDK, NIH http://digestive.niddk.nih.gov/ [accessed 11.11.2010]

Neumann, A.U., Lam, N.P., Dahari, H., Gretch, D.R., Wiley, T.E., Layden, T.J. and Perelson, A.S. (1998) Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. Science 282(5386), 103-7.

Neumann-Haefelin, C., Frick, D.N., Wang, J.J., Pybus, O.G., Salloum, S., Narula, G.S., Eckart, A., Biezynski, A., Eiermann, T., Klenerman, P., Viazov, S., Roggendorf, M., Thimme, R., Reiser, M. and Timm, J. (2008) Analysis of the evolutionary forces in an immunodominant CD8 epitope in hepatitis C virus at a population level. J Virol 82(7), 3438-51.

New allele Frequency Database (2003) http://www.allefrequencies.net [accessed 10.11.2010]

Nguyen, J.H., Dickson, R.C. and Harnois, D.M. (2008) A decade of liver transplantation at Mayo Clinic in Florida. Clin Transpl, 127-35.

Nguyen, M.H. and Keeffe, E.B. (2005) Prevalence and treatment of hepatitis C virus genotypes 4, 5, and 6. Clin Gastroenterol Hepatol 3(10 Suppl 2), S97-S101.

Nguyen, O., Sheppeard, V., Douglas, M.W., Tu, E. and Rawlinson, W. (2010) Acute hepatitis C infection with evidence of heterosexual transmission. J Clin Virol 49(1), 65-8.

Nielsen, M., Lundegaard, C., Worning, P., Lauemoller, S.L., Lamberth, K., Buus, S., Brunak, S. and Lund, O. (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci 12(5), 1007-17.

Nieuwoudt M, A.A., Neumann A, Van der Merwe SW. (2007) Viral kinetics of HCV genotype 5 in South African patients treated with pegylated interferon-alpha and ribavirin. Hepatology 46, 367A.

Njouom, R., Pasquier, C., Ayouba, A., Tejiokem, M.C., Vessiere, A., Mfoupouendoun, J., Tene, G., Eteki, N., Lobe, M.M., Izopet, J. and Nerrienet, E. (2005) Low risk of mother-to-child transmission of hepatitis C virus in Yaounde, Cameroon: the ANRS 1262 study. Am J Trop Med Hyg 73(2), 460-6.

Noppornpanth, S., Lien, T.X., Poovorawan, Y., Smits, S.L., Osterhaus, A.D. and Haagmans, B.L. (2006a) Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. J Virol 80(15), 7569-77.

Noppornpanth, S., Poovorawan, Y., Lien, T.X., Smits, S.L., Osterhaus, A.D. and Haagmans, B.L. (2008) Complete genome analysis of hepatitis C virus subtypes 6t and 6u. J Gen Virol 89(Pt 5), 1276-81.

Noppornpanth, S., Sablon, E., De Nys, K., Truong, X.L., Brouwer, J., Van Brussel,

M., Smits, S.L., Poovorawan, Y., Osterhaus, A.D. and Haagmans, B.L. (2006b) Genotyping hepatitis C viruses from Southeast Asia by a novel line probe assay that simultaneously detects core and 5' untranslated regions. J Clin Microbiol 44(11), 3969-74.

Oh, H.B., Kim, S.O., Cha, C.H., Hong, S.P., Folk, W.R., Kim, K.M. and Suh, D.J. (2008) Identification of hepatitis C virus genotype 6 in Korean patients by analysis of 5' untranslated region using a matrix assisted laser desorption/ionization time of flight-based assay, restriction fragment mass polymorphism. J Med Virol 80(10), 1712-9.

Ohno, T. and Mizokami, M. (1999) Genotyping by type-specific primers that can type HCV types 1-6. Methods Mol Med 19, 159-64.

Ohno, T., Mizokami, M., Tibbs, C.J., Ohba, K., Suzuki, K., Wu, R.R., Nouri-Aria, K.T. and Williams, R. (1994) New genotype of hepatitis C virus in South Africa. J Med Virol 42(4), 409-13.

Ohno, O., Mizokami, M., Wu, R.R., Saleh, M.G., Ohba, K., Orito, E., Mukaide, M., Williams, R. and Lau, J.Y. (1997) New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. J Clin Microbiol 35, 201-7.

Okamoto, H. and Mishiro, S. (1994) Genetic heterogeneity of hepatitis C virus. Intervirology 37(2), 68-76.

Okamoto, H., Okada, S., Sugiyama, Y., Tanaka, T., Sugai, Y., Akahane, Y., Machida, A., Mishiro, S., Yoshizawa, H., Miyakawa, Y. and et al. (1990) Detection of hepatitis C virus RNA by a two-stage polymerase chain reaction with two pairs of primers deduced from the 5'-noncoding region. Jpn J Exp Med 60(4), 215-22.

Op De Beeck, A., Cocquerel, L. and Dubuisson, J. (2001) Biogenesis of hepatitis C virus envelope glycoproteins. J Gen Virol 82(Pt 11), 2589-95.

Osburn, W.O., Fisher, B.E., Dowd, K.A., Urban, G., Liu, L., Ray, S.C., Thomas, D.L. and Cox, A.L. (2010) Spontaneous control of primary hepatitis C virus infection and immunity against persistent reinfection. Gastroenterology 138, 315-24.

Pang, P.S., Planet, P.J. and Glenn, J.S. (2009) The evolution of the major hepatitis C genotypes correlates with clinical response to interferon therapy. PLoS One 4(8), e6579.

Parboosing, R., Paruk, I. and Lalloo, U.G. (2008) Hepatitis C virus seropositivity in a South African Cohort of HIV co-infected, ARV naive patients is associated with renal insufficiency and increased mortality. J Med Virol 80(9), 1530-6.

Pasquier, C., Njouom, R., Ayouba, A., Dubois, M., Sartre, M.T., Vessiere, A., Timba, I., Thonnon, J., Izopet, J. and Nerrienet, E. (2005) Distribution and heterogeneity of hepatitis C genotypes in hepatitis patients in Cameroon. J Med Virol 77(3), 390-8.

Paximadis, M. HLA data. Unpublished [28-06-2009].

Pearlman, B.L., Ehleben, C. and Saifee, S. (2007) Treatment extension to 72 weeks of peginterferon and ribavirin in hepatitis c genotype 1-infected slow responders. Hepatology 46(6), 1688-94.

Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A.J., Houghton, M., Rosa, D., Grandi, G. and Abrignani, S. (1998) Binding of hepatitis C virus to CD81. Science 282(5390), 938-41.

Piron, M., Beguiristain, N., Nadal, A., Martinez-Salas, E. and Gomez, J. (2005) Characterizing the function and structural organization of the 5' tRNA-like motif within the hepatitis C virus quasispecies. Nucleic Acids Res 33(5), 1487-502.

Ploss, A. and Rice, C.M. (2009) Towards a small animal model for hepatitis C. EMBO Rep 10(11), 1220-7.

Pohjanpelto, P., Lappalainen, M., Widell, A., Asikainen, K. and Paunio, M. (1996) Hepatitis C genotypes in Finland determined by RFLP. Clin Diagn Virol 7(1), 7-16.

Pol, S., Vallet-Pichard, A. and Fontaine, H. (2002) Hepatitis C and human immune deficiency co-infection at the era of highly active antiretroviral therapy. J Viral Hepat 9(1), 1-8.

Polyak, S.J., Khabar, K.S., Paschal, D.M., Ezelle, H.J., Duverlie, G., Barber, G.N., Levy, D.E., Mukaida, N. and Gretch, D.R. (2001) Hepatitis C virus nonstructural 5A protein induces interleukin-8, leading to partial inhibition of the interferon-induced antiviral response. J Virol 75(13), 6095-106.

Powis, J., Peltekian, K.M., Lee, S.S., Sherman, M., Bain, V.G., Cooper, C., Krajden, M., Deschenes, M., Balshaw, R.F., Heathcote, E.J. and Yoshida, E.M. (2008) Exploring differences in response to treatment with peginterferon alpha 2a (40kD) and ribavirin in chronic hepatitis C between genotypes 2 and 3. J Viral Hepat 15(1), 52-7.

Poynard, T., McHutchison, J., Davis, G.L., Esteban-Mur, R., Goodman, Z., Bedossa, P. and Albrecht, J. (2000) Impact of interferon alfa-2b and ribavirin on progression of liver fibrosis in patients with chronic hepatitis C. Hepatology 32(5), 1131-7.

Prabdial-Sing, N., Puren, A.J., Mahlangu, J., Barrow, P. and Bowyer, S.M. (2008) Hepatitis C virus genotypes in two different patient cohorts in Johannesburg, South Africa. Arch Virol 153(11), 2049-58.

Prabhu, R., Garry, R.F. and Dash, S. (2006) Small interfering RNA targeted to stem-loop II of the 5' untranslated region effectively inhibits expression of six HCV genotypes. Virol J 3, 100.

Prescott, L.E., Berger, A., Pawlotsky, J.M., Conjeevaram, P., Pike, I. and Simmonds, P. (1997) Sequence analysis of hepatitis C virus variants producing discrep-

ant results with two different genotyping assays. J Med Virol 53(3), 237-44.

Prieto, M., Berenguer, M., Rayon, J.M., Cordoba, J., Arguello, L., Carrasco, D., Garcia-Herola, A., Olaso, V., De Juan, M., Gobernado, M., Mir, J. and Berenguer, J. (1999) High incidence of allograft cirrhosis in hepatitis C virus genotype 1b infection following transplantation: relationship with rejection episodes. Hepatology 29(1), 250-6.

Promrat, K., McDermott, D.H., Gonzalez, C.M., Kleiner, D.E., Koziol, D.E., Lessie, M., Merrell, M., Soza, A., Heller, T., Ghany, M., Park, Y., Alter, H.J., Hoofnagle, J.H., Murphy, P.M. and Liang, T.J. (2003) Associations of chemokine system polymorphisms with clinical outcomes and treatment responses of chronic hepatitis C. Gastroenterology 124(2), 352-60.

ProPred I (2008) http://www.imtech.res.in/raghava/propredl/index.html. [accessed 30-11-2010].

ProPred II (2008) http://www.imtech.res.in/raghava/propredII/index.html. [accessed 20-10-2010].

Public Health Agency of Canada, PHAC (2010) http://www.phac-aspc.gc.ca. [accessed 10.10.2010]

PubMed http://www.ncbi.nlm.nih.gov/sites/entrez [accessed 18-11-2009].

Pybus, O.G., Barnes, E., Taggart, R., Lemey, P., Markov, P.V., Rasachak, B., Syhavong, B., Phetsouvanah, R., Sheridan, I., Humphreys, I.S., Lu, L., Newton, P.N. and Klenerman, P. (2009) Genetic history of hepatitis C virus in East Asia. J Virol 83 (2), 1071-82.

Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. and Stevanovic, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50(3-4), 213-9.

Rauch, A., Gaudieri, S., Thio, C. and Bochud, P.Y. (2009) Host genetic determinants of spontaneous hepatitis C clearance. Pharmacogenomics 10(11), 1819-37.

Ray, S.C., Wang, Y.M., Laeyendecker, O., Ticehurst, J.R., Villano, S.A. and Thomas, D.L. (1999) Acute hepatitis C virus structural gene sequences as predictors of persistent viremia: hypervariable region 1 as a decoy. J Virol 73(4), 2938-46.

Ribeiro, L.C., Souto, F.J., do Espirito-Santo, M.P., R, G.O. and Lampe, E. (2009) An autochthonous case of hepatitis C virus genotype 5a in Brazil: phylogenetic analysis. Arch Virol 154(4), 665-70.

Richter, S.S. (2002) Laboratory assays for diagnosis and management of hepatitis C virus infection. J Clin Microbiol 40(12), 4407-12.

Robertson, B., Myers, G., Howard, C., Brettin, T., Bukh, J., Gaschen, B., Gojobori, T., Maertens, G., Mizokami, M., Nainan, O., Netesov, S., Nishioka, K., Shin i, T.,

Simmonds, P., Smith, D., Stuyver, L. and Weiner, A. (1998) Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Virus Taxonomy. Arch Virol 143(12), 2493-503.

Roccasecca, R., Ansuini, H., Vitelli, A., Meola, A., Scarselli, E., Acali, S., Pezzanera, M., Ercole, B.B., McKeating, J., Yagnik, A., Lahm, A., Tramontano, A., Cortese, R. and Nicosia, A. (2003) Binding of the hepatitis C virus E2 glycoprotein to CD81 is strain specific and is modulated by a complex interplay between hypervariable regions 1 and 2. J Virol 77(3), 1856-67.

Roche Applied Science FRET http://www.roche-applied-science.com [accessed 19.09.2010]

Rodriguez-Torres, M. (2010) On the cusp of change: new therapeutic modalities for HCV. Ann Hepatol 9 Suppl, 123-31.

Rolfe, K.J., Alexander, G.J., Wreghitt, T.G., Parmar, S., Jalal, H. and Curran, M.D. (2005) A real-time Taqman method for hepatitis C virus genotyping. J Clin Virol 34 (2), 115-21.

Roque-Afonso, A.M., Ferey, M.P., Poveda, J.D., Marchadier, E. and Dussaix, E. (2002) Performance of TRUGENE hepatitis C virus 5' noncoding genotyping kit, a new CLIP sequencing-based assay for hepatitis C virus genotype determination. J Viral Hepat 9(5), 385-9.

Ross, R.S., Viazov, S.O., Hoffmann, S. and Roggendorf, M. (2001) Performance characteristics of a transcription-mediated nucleic acid amplification assay for qualitative detection of hepatitis C virus RNA. J Clin Lab Anal 15(6), 308-13.

Rudwaleit, M., Tikly, M., Gibson, K., Pile, K. and Wordsworth, P. (1995) HLA class II antigens associated with systemic lupus erythematosus in black South Africans. Ann Rheum Dis 54(8), 678-80.

Sakai, A., Claire, M.S., Faulk, K., Govindarajan, S., Emerson, S.U., Purcell, R.H. and Bukh, J. (2003) The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences. Proc Natl Acad Sci U S A 100(20), 11646-51.

Sandres-Saune, K., Deny, P., Pasquier, C., Thibaut, V., Duverlie, G. and Izopet, J. (2003) Determining hepatitis C genotype by analyzing the sequence of the NS5b region. J Virol Methods 109(2), 187-93.

Santolini, E., Pacini, L., Fipaldini, C., Migliaccio, G. and Monica, N. (1995) The NS2 protein of hepatitis C virus is a transmembrane polypeptide. J Virol 69(12), 7461-71.

Sarobe, P., Lasarte, J.J., Garcia, N., Civeira, M.P., Borras-Cuesta, F. and Prieto, J. (2006) Characterization of T-cell responses against immunodominant epitopes from hepatitis C virus E2 and NS4a proteins. J Viral Hepat 13(1), 47-55.

Sarrazin, C. and Zeuzem, S. (2010) Resistance to direct antiviral agents in patients with hepatitis C virus infection. Gastroenterology 138(2), 447-62.

Satapathy, S.K., Lingisetty, C.S., Proper, S., Chaudhari, S. and Williams, S. (2010) Equally poor outcomes to pegylated interferon-based therapy in African Americans and Hispanics with chronic hepatitis C infection. J Clin Gastroenterol 44(2), 140-5.

Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R.M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R. and Vitelli, A. (2002) The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. Embo J 21(19), 5017-25.

Schroter, M., Zollner, B., Schafer, P., Landt, O., Lass, U., Laufs, R. and Feucht, H.H. (2002) Genotyping of hepatitis C virus types 1, 2, 3, and 4 by a one-step LightCycler method using three different pairs of hybridization probes. J Clin Microbiol 40(6), 2046-50.

Schulze zur Wiesch, J., Lauer, G.M., Day, C.L., Kim, A.Y., Ouchi, K., Duncan, J.E., Wurcel, A.G., Timm, J., Jones, A.M., Mothe, B., Allen, T.M., McGovern, B., Lewis-Ximenez, L., Sidney, J., Sette, A., Chung, R.T. and Walker, B.D. (2005) Broad repertoire of the CD4+ Th cell response in spontaneously controlled hepatitis C virus infection includes dominant and highly promiscuous epitopes. J Immunol 175(6), 3603-13.

Schulze Zur Wiesch, J., Lauer, G.M., Timm, J., Kuntzen, T., Neukamm, M., Berical, A., Jones, A.M., Nolan, B.E., Longworth, S.A., Kasprowicz, V., McMahon, C., Wurcel, A., Lohse, A.W., Lewis-Ximenez, L.L., Chung, R.T., Kim, A.Y., Allen, T.M. and Walker, B.D. (2007) Immunologic evidence for lack of heterologous protection following resolution of HCV in patients with non-genotype 1 infection. Blood 110(5), 1559-69.

Schutzbank, T.E., Sefers, S.E., Kahmann, N., Li, H. and Tang, Y.W. (2006) Comparative evaluation of three commercially available methodologies for hepatitis C virus genotyping. J Clin Microbiol 44(10), 3797-8.

Sette, A., Sidney, J., del Guercio, M.F., Southwood, S., Ruppert, J., Dahlberg, C., Grey, H.M. and Kubo, R.T. (1994) Peptide binding to the most frequent HLA-A class I alleles measured by quantitative molecular binding assays. Mol Immunol 31(11), 813-22.

Seymour, C.A. (1996) Screening asymptomatic people at high risk for hepatitis C. The case for. Bmj 312(7042), 1347-8.

Sheehy, P., Scallan, M., Kenny-Walsh, E., Shanahan, F. and Fanning, L.J. (2005) A strategy for obtaining near full-length HCV cDNA clones (assemblicons) by assembly PCR. J Virol Methods 123(2), 115-24.

Sherman, K.E., Rouster, S.D., Chung, R.T. and Rajicic, N. (2002) Hepatitis C Virus prevalence among patients infected with Human Immunodeficiency Virus: a cross-sectional analysis of the US adult AIDS Clinical Trials Group. Clin Infect Dis 34(6),

831-7.

Shin, I.T., Tanaka, Y., Tateno, Y. and Mizokami, M. (2008) Development and public release of a comprehensive hepatitis virus database. Hepatol Res 38(3), 234-43.

Shoukry, N.H., Grakoui, A., Houghton, M., Chien, D.Y., Ghrayeb, J., Reimann, K.A. and Walker, C.M. (2003) Memory CD8+ T cells are required for protection from persistent hepatitis C virus infection. J Exp Med 197(12), 1645-55.

Sidney, J., Peters, B., Frahm, N., Brander, C. and Sette, A. (2008) HLA class I supertypes: a revised and updated classification. BMC Immunol 9, 1.

Sillanpaa, M., Melen, K., Porkka, P., Fagerlund, R., Nevalainen, K., Lappalainen, M. and Julkunen, I. (2009) Hepatitis C virus core, NS3, NS4B and NS5A are the major immunogenic proteins in humoral immunity in chronic HCV infection. Virol J 6, 84.

Simmonds, P. (2004) Genetic diversity and evolution of hepatitis C virus—15 years on. J. Gen Virol 85, 3173-3188.

Simmonds, P., Alberti, A., Alter, H.J., Bonino, F., Bradley, D.W., Brechot, C., Brouwer, J.T., Chan, S.W., Chayama, K., Chen, D.S. and et al. (1994a) A proposed system for the nomenclature of hepatitis C viral genotypes. Hepatology 19(5), 1321-4.

Simmonds, P., Holmes, E.C., Cha, T.A., Chan, S.W., McOmish, F., Irvine, B., Beall, E., Yap, P.L., Kolberg, J. and Urdea, M.S. (1993a) Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J Gen Virol 74 (Pt 11), 2391-9.

Simmonds, P., McOmish, F., Yap, P.L., Chan, S.W., Lin, C.K., Dusheiko, G., Saeed, A.A. and Holmes, E.C. (1993b) Sequence variability in the 5' non-coding region of hepatitis C virus: identification of a new virus type and restrictions on sequence diversity. J Gen Virol 74 (Pt 4), 661-8.

Simmonds, P., Mellor, J., Sakuldamrongpanich, T., Nuchaprayoon, C., Tanprasert, S., Holmes, E.C. and Smith, D.B. (1996) Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: comparison with classifications based upon sequence similarity. J Gen Virol 77 (Pt 12), 3013-24.

Simmonds, P., Smith, D.B., McOmish, F., Yap, P.L., Kolberg, J., Urdea, M.S. and Holmes, E.C. (1994b) Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. J Gen Virol 75 (Pt 5), 1053-61.

Singh, H. and Raghava, G.P. (2001) ProPred: prediction of HLA-DR binding sites. Bioinformatics 17(12), 1236-7.

Sithebe, N.P., Aspinall, S. and Smuts, H. (1996) Molecular epidemiology of hepatitis C virus infection at Ga-Rankuwa Hospital. S Afr Med J 86(12), 1543-5.

Sklan, E.H., Charuworn, P., Pang, P.S. and Glenn, J.S. (2009) Mechanisms of HCV survival in the host. Nat Rev Gastroenterol Hepatol 6(4), 217-27.

Smuts, H.E. and Kannemeyer, J. (1995) Genotyping of hepatitis C virus in South Africa. J Clin Microbiol 33(6), 1679-81.

Sneath, P.H.A. and Sokal, R.R. (1973) Numerical taxonomy. Freeman, San Francisco CA.

Soni, P.N., Tait, D.R., Gopaul, W., Sathar, M.A. and Simjee, A.E. (1996) Hepatitis C virus infection in chronic liver disease in Natal. S Afr Med J 86(1), 80-3.

Soriano, V., Peters, M.G. and Zeuzem, S. (2009) New therapies for hepatitis C virus infection. Clin Infect Dis 48(3), 313-20.

Statistics South Africa, http://www.statssa.gov.za, [accessed 12-08-2010]

Statistics notes, http://www.doh.gov.za/facts/index.html, November 2009 [accessed 09-10-2010]

Steinhauer, H.B., Lubrich-Birkner, I., Kluthe, R., Baumann, G. and Schollmeyer, P. (1992) Effect of amino acid based dialysis solution on peritoneal permeability and prostanoid generation in patients undergoing continuous ambulatory peritoneal dialysis. Am J Nephrol 12(1-2), 61-7.

Sterling, R.K., Contos, M.J., Sanyal, A.J., Luketic, V.A., Stravitz, R.T., Wilson, M.S., Mills, A.S. and Shiffman, M.L. (2003) The clinical spectrum of hepatitis C virus in HIV co-infection. J Acquir Immune Defic Syndr 32(1), 30-7.

Strasfeld, L., Lo, Y., Netski, D., Thomas, D.L. and Klein, R.S. (2003) The association of hepatitis C prevalence, activity, and genotype with HIV infection in a cohort of New York City drug users. J Acquir Immune Defic Syndr 33(3), 356-64.

Stuyver, L., van Arnhem, W., Wyseur, A., Hernandez, F., Delaporte, E. and Maertens, G. (1994) Classification of hepatitis C viruses based on phylogenetic analysis of the envelope 1 and nonstructural 5B regions and identification of five additional subtypes. Proc Natl Acad Sci U S A 91(21), 10134-8.

Stuyver, L., Wyseur, A., van Arnhem, W., Hernandez, F. and Maertens, G. (1996) Second-generation line probe assay for hepatitis C virus genotyping. J Clin Microbiol 34(9), 2259-66.

Sullivan, D.G., Kim, S.S., Wilson, J.J., Stehman-Breen, C. and Gretch, D.R. (2001) Investigating hepatitis C virus heterogeneity in a high prevalence setting using heteroduplex tracking analysis. J Virol Methods 96(1), 5-16.

Suzuki, R., Sakamoto, S., Tsutsumi, T., Rikimaru, A., Tanaka, K., Shimoike, T., Moriishi, K., Iwasaki, T., Mizumoto, K., Matsuura, Y., Miyamura, T. and Suzuki, T. (2005) Molecular determinants for subcellular localization of hepatitis C virus core protein. J Virol 79(2), 1271-81.

Suzuki, R., Suzuki, T., Ishii, K., Matsuura, Y. and Miyamura, T. (1999) Processing

and functions of Hepatitis C virus proteins. Intervirology 42(2-3), 145-52.

Szabo, G. and Dolganiuc, A. (2008) Hepatitis C and innate immunity: recent advances. Clin Liver Dis 12(3), 675-92.

Tai, C.L., Chi, W.K., Chen, D.S. and Hwang, L.H. (1996) The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). J Virol 70(12), 8477-84.

Takahashi, M., Saito, H., Higashimoto, M., Atsukawa, K. and Ishii, H. (2005) Benefit of hepatitis C virus core antigen assay in prediction of therapeutic response to interferon and ribavirin combination therapy. J Clin Microbiol 43(1), 186-91.

Tamalet, C., Colson, P., Tissot-Dupont, H., Henry, M., Tourres, C., Tivoli, N., Botta, D., Ravaux, I., Poizot-Martin, I. and Yahi, N. (2003) Genomic and phylogenetic analysis of hepatitis C virus isolates: a survey of 535 strains circulating in southern France. J Med Virol 71(3), 391-8.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24(8), 1596-9.

Tamura, K. and Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10(3), 512-26.

Tanaka, T., Kato, N., Cho, M.J., Sugiyama, K. and Shimotohno, K. (1996) Structure of the 3' terminus of the hepatitis C virus genome. J Virol 70(5), 3307-12.

Tanaka, Y., Hanada, K., Mizokami, M., Yeo, A.E., Shih, J.W., Gojobori, T. and Alter, H.J. (2002) Inaugural Article: A comparison of the molecular clock of hepatitis C virus in the United States and Japan predicts that hepatocellular carcinoma incidence in the United States will increase over the next two decades. Proc Natl Acad Sci U S A 99(24), 15584-9.

Tanaka, Y., Kurbanov, F., Mano, S., Orito, E., Vargas, V., Esteban, J.I., Yuen, M.F., Lai, C.L., Kramvis, A., Kew, M.C., Smuts, H.E., Netesov, S.V., Alter, H.J. and Mizokami, M. (2006) Molecular tracing of the global hepatitis C virus epidemic predicts regional patterns of hepatocellular carcinoma mortality. Gastroenterology 130, 703-14.

Tanaka, Y., Nishida, N., Sugiyama, M., Kurosaki, M., Matsuura, K., Sakamoto, N., Nakagawa, M., Korenaga, M., Hino, K., Hige, S., Ito, Y., Mita, E., Tanaka, E., Mochida, S., Murawaki, Y., Honda, M., Sakai, A., Hiasa, Y., Nishiguchi, S., Koike, A., Sakaida, I., Imamura, M., Ito, K., Yano, K., Masaki, N., Sugauchi, F., Izumi, N., Tokunaga, K. and Mizokami, M. (2009) Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 41, 1105-9.

Tarantino, G. and Craxi, A. (2009) Optimizing the treatment of chronic hepatitis due to hepatitis C virus genotypes 2 and 3: a review. Liver Int 29 Suppl 1, 31-8.

Tavare, S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences http://www.cmb.usc.edu/people/stavare/STpapers-pdf/T86.pdf, Lectures on Mathematics in the Life Sciences, pp. 57-86. Vol. 17.

Taylor, D.R., Shi, S.T., Romano, P.R., Barber, G.N. and Lai, M.M. (1999) Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. Science 285 (5424), 107-10.

Tellinghuisen, T.L., Marcotrigiano, J., Gorbalenya, A.E. and Rice, C.M. (2004) The NS5A protein of hepatitis C virus is a zinc metalloprotein. J Biol Chem 279(47), 48576-87.

Tellinghuisen, T.L., Marcotrigiano, J. and Rice, C.M. (2005) Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. Nature 435(7040), 374-9.

Thakur, A., Guptan, R. C., Arankale, V. & Sarin, S. K. (2003) Low specificity of the third generation ELISA for HCV detection in voluntary blood donors in India. eJIFCC 14(1), 4.

Thimme, R., Bukh, J., Spangenberg, H.C., Wieland, S., Pemberton, J., Steiger, C., Govindarajan, S., Purcell, R.H. and Chisari, F.V. (2002) Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease. Proc Natl Acad Sci U S A 99(24), 15661-8.

Thio, C.L., Gao, X., Goedert, J.J., Vlahov, D., Nelson, K.E., Hilgartner, M.W., O'Brien, S.J., Karacki, P., Astemborski, J., Carrington, M. and Thomas, D.L. (2002) HLA-Cw*04 and hepatitis C virus persistence. J Virol 76(10), 4792-7.

Thomas, T. and Foster, G. (2007) Nanomedicines in the treatment of chronic hepatitis C--focus on pegylated interferon alpha-2a. Int J Nanomedicine 2(1), 19-24.

Thursz, M., Yallop, R., Goldin, R., Trepo, C. and Thomas, H.C. (1999) Influence of MHC class II genotype on outcome of infection with hepatitis C virus. The HENCORE group. Hepatitis C European Network for Cooperative Research. Lancet 354 (9196), 2119-24.

Tillmann, H.L., Chen, D.F., Trautwein, C., Kliem, V., Grundey, A., Berning-Haag, A., Boker, K., Kubicka, S., Pastucha, L., Stangel, W. and Manns, M.P. (2001) Low frequency of HLA-DRB1*11 in hepatitis C virus induced end stage liver disease. Gut 48(5), 714-8.

Tokita, H., Okamoto, H., Iizuka, H., Kishimoto, J., Tsuda, F., Miyakawa, Y. and Mayumi, M. (1998) The entire nucleotide sequences of three hepatitis C virus isolates in genetic groups 7-9 and comparison with those in the other eight genetic groups. J Gen Virol 79 (Pt 8), 1847-57.

Toussaint, N.C. and Kohlbacher, O. (2009) OptiTope--a web server for the selection of an optimal set of peptides for epitope-based vaccines. Nucleic Acids Res 37(Web Server issue), W617-22.

Tsukiyama-Kohara, K., Iizuka, N., Kohara, M. and Nomoto, A. (1992) Internal ribosome entry site within hepatitis C virus RNA. J Virol 66(3), 1476-83.

Tucker, T.J., Voigt, M., Bird, A., Robson, S., Gibbs, B., Kannemeyer, J., Galloway, M., Kirsch, R.E. and Smuts, H. (1997) Hepatitis C virus infection rate in volunteer blood donors from the Western Cape--comparison of screening tests and PCR. S Afr Med J 87(5), 603-5.

Tuplin, A., Evans, D.J. and Simmonds, P. (2004) Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. J Gen Virol 85(Pt 10), 3037-47.

van Doorn, L.J. (1994) Review: molecular biology of the hepatitis C virus. J Med Virol 43(4), 345-56.

van Doorn, L.J., Kleter, B., Stuyver, L., Maertens, G., Brouwer, H., Schalm, S., Heijtink, R. and Quint, W. (1994) Analysis of hepatitis C virus genotypes by a line probe assay and correlation with antibody profiles. J Hepatol 21(1), 122-9.

van Doorn, L.J., Kleter, G.E., Stuyver, L., Maertens, G., Brouwer, J.T., Schalm, S.W., Heijtink, R.A. and Quint, W.G. (1995) Sequence analysis of hepatitis C virus genotypes 1 to 5 reveals multiple novel subtypes in the Benelux countries. J Gen Virol 76 (Pt 7), 1871-6.

Vandenbroucke, J.P., von Elm, E., Altman, D.G., Gotzsche, P.C., Mulrow, C.D., Pocock, S.J., Poole, C., Schlesselman, J.J. and Egger, M. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Ann Intern Med 147(8), W163-94.

Vardas, E., Ross, M.H., Sharp, G., McAnerney, J. and Sim, J. (2002) Viral hepatitis in South African healthcare workers at increased risk of occupational exposure to blood-borne viruses. J Hosp Infect 50(1), 6-12.

Vejbaesya, S., Songsivilai, S., Tanwandee, T., Rachaibun, S., Chantangpol, R. and Dharakul, T. (2000) HLA association with hepatitis C virus infection. Hum Immunol 61(3), 348-53.

Venter, M., Rock, M., Puren, A.J., Tiemessen, C.T. and Crowe, J.E., Jr. (2003) Respiratory syncytial virus nucleoprotein-specific cytotoxic T-cell epitopes in a South African population of diverse HLA types are conserved in circulating field strains. J Virol 77(13), 7319-29.

Verbeeck, J., Maes, P., Lemey, P., Pybus, O.G., Wollants, E., Song, E., Nevens, F., Fevery, J., Delport, W., Van der Merwe, S. and Van Ranst, M. (2006) Investigating the origin and spread of hepatitis C virus genotype 5a. J Virol 80(9), 4220-6.

Verbeeck, J., Maes, P., Wollants, E., Van der Merwe, S., Song, E., Nevens, F. and Van Ranst, M. (2005) Use of a commercially available line probe assay for genotyp-

ing of hepatitis C virus 5a strains. J Clin Microbiol 43(12), 6117-9.

Verbeeck, J., Peigue-Lafeuille, H., Ross, R.S., Abergel, A., Nevens, F., Van der Merwe, S., Van Ranst, M. and Henquell, C. (2008a) HCV genotype 5: epidemiology and spread of an uncommon genotype. J Clin Virol 41(2), 170-1.

Verbeeck, J., Stanley, M.J., Shieh, J., Celis, L., Huyck, E., Wollants, E., Morimoto, J., Farrior, A., Sablon, E., Jankowski-Hennig, M., Schaper, C., Johnson, P., Van Ranst, M. and Van Brussel, M. (2008b) Evaluation of Versant hepatitis C virus genotype assay (LiPA) 2.0. J Clin Microbiol 46(6), 1901-6.

Vermeulen, M., Lelie, N., Sykes, W., Crookes, R., Swanevelder, J., Gaggia, L., Le Roux, M., Kuun, E., Gulube, S. and Reddy, R. (2009) Impact of individual-donation nucleic acid testing on risk of human immunodeficiency virus, hepatitis B virus, and hepatitis C virus transmission by blood transfusion in South Africa. Transfusion 49 (6), 1115-25.

Vertuani, S., Bazzaro, M., Gualandi, G., Micheletti, F., Marastoni, M., Fortini,C., Canella, A., Marino, M., Tomatis, R., Traniello, S. and Gavioli, R. (2002) Effect of interferon-alpha therapy on epitope-specific cytotoxic T lymphocyte responses in hepatitis C virus-infected individuals. Eur J Immunol 32(1), 144-54.

Viazov, S., Widell, A. and Nordenfelt, E. (2000) Mixed infection with two types of hepatitis C virus is probably a rare event. Infection 28(1), 21-5.

Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Krausslich, H.G., Mizokami, M., Bartenschlager, R. and Liang, T.J. (2005) Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. Nat Med 11(7), 791-6.

Wang, H., Shen, X.T., Ye, R., Lan, S.Y., Xiang, L. and Yuan, Z.H. (2005) Roles of the polypyrimidine tract and 3' noncoding region of hepatitis C virus RNA in the internal ribosome entry site-mediated translation. Arch Virol 150(6), 1085-99.

Wang, J.H., Zheng, X., Ke, X., Dorak, M.T., Shen, J., Boodram, B., O'Gorman, M., Beaman, K., Cotler, S.J., Hershow, R. and Rong, L. (2009) Ethnic and geographical differences in HLA associations with the outcome of hepatitis C virus infection. Virol J 6, 46.

Wawrzynowicz-Syczewska, M., Underhill, J.A., Clare, M.A., Boron-Kaczmarska, A., McFarlane, I.G. and Donaldson, P.T. (2000) HLA class II genotypes associated with chronic hepatitis C virus infection and response to alpha-interferon treatment in Poland. Liver 20(3), 234-9.

Weber, J. (2001) The pathogenesis of HIV-1 infection http://bmb.oxfordjournals.org/content/58/1/61.full. Br Med Bull 58(1), 61-72.

WebLogo, version 2.8.2 http://weblogo.berkeley.edu/logo.cg [accessed 08-09-2008].

Wedemeyer, H., Schuller, E., Schlaphoff, V., Stauber, R.E., Wiegand, J., Schiefke, I., Firbas, C., Jilma, B., Thursz, M., Zeuzem, S., Hofmann, W.P., Hinrichsen, H., Tauber, E., Manns, M.P. and Klade, C.S. (2009) Therapeutic vaccine IC41 as late add-on to standard treatment in patients with chronic hepatitis C. Vaccine 27(37), 5142-51.

Wei, S.H., Yin, W., An, Q.X., Lei, Y.F., Hu, X.B., Yang, J., Lu, X., Zhang, H. and Xu, Z.K. (2008) A novel hepatitis C virus vaccine approach using recombinant Bacillus Calmette-Guerin expressing multi-epitope antigen. Arch Virol 153(6), 1021-9.

Weiner, A.J., Brauer, M.J., Rosenblatt, J., Richman, K.H., Tung, J., Crawford,K., Bonino, F., Saracco, G., Choo, Q.L., Houghton, M. and et al. (1991) Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology 180(2), 842-8.

Wertheimer, A.M., Miner, C., Lewinsohn, D.M., Sasaki, A.W., Kaufman, E. and Rosen, H.R. (2003) Novel CD4+ and CD8+ T-cell determinants within the NS3 protein in subjects with spontaneously resolved HCV infection. Hepatology 37(3), 577-89.

Wicker, S., Jung, J., Allwinn, R., Gottschalk, R. and Rabenau, H.F. (2008) Prevalence and prevention of needlestick injuries among health care workers in a German university hospital. Int Arch Occup Environ Health 81(3), 347-54.

Wohnsland, A., Hofmann, W.P. and Sarrazin, C. (2007) Viral determinants of resistance to treatment in patients with hepatitis C. Clin Microbiol Rev 20(1), 23-38.

WHO, World Health Organization (1999) Global surveillance and control of hepatitis C. Report of a WHO Consultation organized in Collaboration with the Viral Hepatitis Prevention Board. J Viral Hepatol 6, 35-47.

WHO, World Health Organization (2002) Hepatitis C. WHO/CDS/CSR/LYO: 1-69.

WHO, World Health Organization (2007) Global blood safety and availability. http://www.who.int/mediacentre/factsheets/ [accessed 20-11-2010]

WHO, World Health Organization (2009) Viral hepatitis. Report by the Secretariat, EB126/15.12.November2009.http://apps.who.int/gb/ebwha/pdf_files/EB126/B126_15-en.pdf [accessed 20-11-2010]

WHO, World Health Organization (2010) Viral Cancers. http://www.who.int/vaccine_research/diseases/viral_cancers/en/index2.html [accessed 01-07-2011]

Yang, J.H., Lai, J.P., Douglas, S.D., Metzger, D., Zhu, X.H. and Ho, W.Z. (2002) Real-time RT-PCR for quantitation of hepatitis C virus RNA. J Virol Methods 102(1-2), 119-28.

Yao, E. and Tavis, J.E. and the Virus Sequence group. (2005) A general method for nested RT-PCR amplification and sequencing the complete HCV genotype 1 open

reading frame. Virol J 2, 88.

Yee, L.J., Im, K., Wahed, A.S., Bugawan, T., Li, J., Rhodes, S.L., Erlich, H., Rosen, H.R., Liang, T.J. and Yang, H. (2009) Polymorphism in the human major histocompatibility complex and early viral decline during treatment of chronic hepatitis C. Antimicrob Agents Chemother 53(2), 615-21.

Yen, Y.H., Hung, C.H., Hu, T.H., Chen, C.H., Wu, C.M., Wang, J.H., Lu, S.N. and Lee, C.M. (2008) Mutations in the interferon sensitivity-determining region (nonstructural 5A amino acid 2209-2248) in patients with hepatitis C-1b infection and correlating response to combined therapy of pegylated interferon and ribavirin. Aliment Pharmacol Ther 27(1), 72-9.

Yokota, S., Wei, M., Shimizu, H., Inui, A. and Fujisawa, T. (1994) Superinfection with hepatitis C virus subtypes in frequently transfused children with a haematologic disorder or malignancy. Eur J Pediatr 153(8), 614-6.

Yoshizawa, H. (2002) Hepatocellular carcinoma associated with hepatitis C virus infection in Japan: projection to other countries in the foreseeable future. Oncology 62 Suppl 1, 8-17.

Zeisel, M.B., Barth, H., Schuster, C. and Baumert, T.F. (2009). Hepatitis C virus entry: molecular mechanisms and targets for antiviral therapy. Front Biosci 14, 3274-85.

Zeisel, M.B., Fafi-Kremer, S., Fofana, I., Barth, H., Stoll-Keller, F., Doffoel, M. and Baumert, T.F. (2007) Neutralizing antibodies in hepatitis C virus infection. World J Gastroenterol 13(36), 4824-30.

Zeuzem, S., Rizzetto, M., Ferenci, P. and Shiffman, M.L. (2009) Management of hepatitis C virus genotype 2 or 3 infection: treatment optimization on the basis of virological response. Antivir Ther 14(2), 143-54.

Zeuzem, S., Teuber, G., Lee, J.H., Ruster, B. and Roth, W.K. (1996) Risk factors for the transmission of hepatitis C. J Hepatol 24(2 Suppl), 3-10.

Zhang, H., Hanecak, R., Brown-Driver, V., Azad, R., Conklin, B., Fox, M.C. and Anderson, K.P. (1999) Antisense oligonucleotide inhibition of hepatitis C virus (HCV) gene expression in livers of mice infected with an HCV-vaccinia virus recombinant. Antimicrob Agents Chemother 43(2), 347-53.

Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D.R., Wieland, S.F., Uprichard, S.L., Wakita, T. and Chisari, F.V. (2005) Robust hepatitis C virus infection in vitro. Proc Natl Acad Sci U S A 102(26), 9294-9.