

Naturally occurring resistance mutations within the core and NS5B regions in Hepatitis C genotypes, particularly genotype 5a, in South Africa

Prabdia-Sing, N*^{1,2}, Blackard, JT³, Puren, AJ^{4,2}, Mahomed, A⁵, Abuelhassan, W⁶, Mahlangu, J⁷, Vermeulen, M⁸ and Bowyer, SM^{9,10}

*corresponding author, niship@nicd.ac.za, +27113866347

¹Centre for Vaccines and Immunology, National Institute for Communicable Diseases (NICD), Sandringham, Johannesburg, South Africa

²Faculty of Health Sciences, School of Pathology, Division of Virology and Communicable Diseases Surveillance, , University of Witwatersrand, Johannesburg, South Africa

³Department of Internal Medicine Division of Digestive Diseases, University of Cincinnati Medical Centre, Cincinnati, Ohio, USA

⁴Centre for HIV/STI, National Institute for Communicable Diseases (NICD), Sandringham, Johannesburg, South Africa

⁵Gastroenterology Department, University of the Witwatersrand and Charlotte Maxeke Hospital, Johannesburg, South Africa

⁶Gastroenterology Department, University of the Witwatersrand and Chris Hani Baragwanath Hospital, Johannesburg, South Africa

⁷Department of Molecular Medicine and Haematology, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand and Charlotte Maxeke Hospital, Johannesburg, South Africa and National Health Laboratory Services (NHLS)

⁸Donation Testing, South African National Blood Service, Roodepoort, South Africa

⁹Department of Medical Virology, University of Pretoria, Pretoria, South Africa

¹⁰Tshwane Academic Division, NHLS, Pretoria, South Africa

The GenBank accession numbers for core/E1 sequences are KR007772-KR007802. The GenBank accession numbers for NS5B sequences are KR007803-KR007846.

Abstract

Approximately 1 million South Africans are infected with Hepatitis C virus (HCV). The standard of care (SOC) in South Africa is combination therapy (pegylated interferon and ribavirin). HCV genotypes and/or mutations in the core/ non-structural regions have been associated with response to therapy and/or disease progression. This study examines mutations in the core (29-280 amino acids, including ~90 E1 amino acids) and NS5B (241-306 amino acids) regions on pre-treatment isolates from patients attending Johannesburg hospitals or asymptomatic South African blood donors. Diversity within known CD4+ and CD8+ T-cell epitopes was also explored. Samples grouped into subtypes 1a (N=10) 1b (N=12), 3a(N=5), 4a (N=3) and 5a(N=61). Two mutations, associated with interferon resistance—R70Q and T110N—were present in 29 genotype 5a core sequences. No resistance mutation to NS5B nucleotide inhibitors, sofosbuvir was found. Six putative CD8+ and one CD4+ T-cell epitope sequence in the core region showed binding scores of <300 IC₅₀nM to HLA alleles frequently observed in the South African population. No known CD8+ and CD4+ T-cell epitopes were mapped in the NS5B region. The analysis begs the question whether those infected with genotype 5a will benefit better on interferon-free combination therapies. This study provides new insight into one of the lesser studied HCV genotypes and compares the diversity seen in a large pre-treatment cohort with other subtypes.

Keywords

Hepatitis C, genotype 5, interferon, mutations, therapy

Highlights

- We examined mutational changes of hepatitis C in the core/E1 and NS5B genes.
- Phylogenetic analyses were performed using Bayesian inferences.
- Diversity within known CD4+ and CD8+ T-cell epitopes was also explored.
- Mutations, associated with interferon resistance, were present in genotype 5a samples at baseline.

1. Introduction

South Africa has a 2% seroprevalence rate of hepatitis C virus (HCV) infection (Averhoff *et al.*, 2012). Eighty percent will develop chronic hepatitis C and at risk for compensated and decompensated cirrhosis and hepatocellular carcinoma (HCC, Chen & Morgan (2006)). There is a 2.91-fold increased odds of being HCV IgG positive if co infected with HIV (Tathiah *et al.*, 2014). The standard of care (SOC) for treating hepatitis C in South Africa is combination therapy with pegylated interferon (PEG-IFN) and ribavirin (Botha *et al.*, 2010). Individuals with genotype 5a, the predominant genotype in South Africa, have response rates to combination therapy ranging from 47-67% (Antaki *et al.*, 2008; Devaki *et al.*, 2015). However, many South Africans: typically present with advanced diseases for them to benefit from therapy, have poor access or cannot afford the treatment. Poor adherence due to severe side effects further reduces the overall benefits. Direct acting antivirals (DAAs), targeted at the NS3, NS5A and NS5B regions of the virus, result in sustained virological responses (SVR) in ~ 95% of patients (Van der Meer, 2015). Currently, simeprevir and sofosbuvir are in the forefront of HCV genotype 1 therapeutics, targeting the NS3 and NS5B regions of the viral genome, respectively, with or without PEG-IFN and ribavirin (AASLD, 2015). Resistance associated variants (RAVs) has been described at baseline and during therapy for NS3, NS5A and NS5B inhibitors (Lontok *et al.*, 2015). In genotype 1, viral mutations have been associated with either sustained virological response (SVR) or viral breakthrough during therapy (Nishiya *et al.*, 2014; Cento *et al.*, 2012). However, less data are available on non-1 genotypes. Unfortunately, the costs of DAAs delay global roll-out and compel many middle/low-income countries to continue with SOC.

The core is a basic, alpha-dimeric protein which is highly conserved, immunogenic and involved in viral RNA binding and particle formation (Santolini *et al.*, 1994; Klein *et al.*, 2005). The immature 191 amino acid core protein is divided into 3 domains D1, D2, D3 (Fig.1; Boulant *et al.*, 2005). D1 is divided into three smaller sections (Fig.1): BD1 (2-23aa), BD2 (38-74aa) and BD3 (101-121aa, Fig.1; Ivanyi-Nagy *et al.*, 2006). There is evidence that mutations in the BD2 region (contains the major binding site for NS5A)

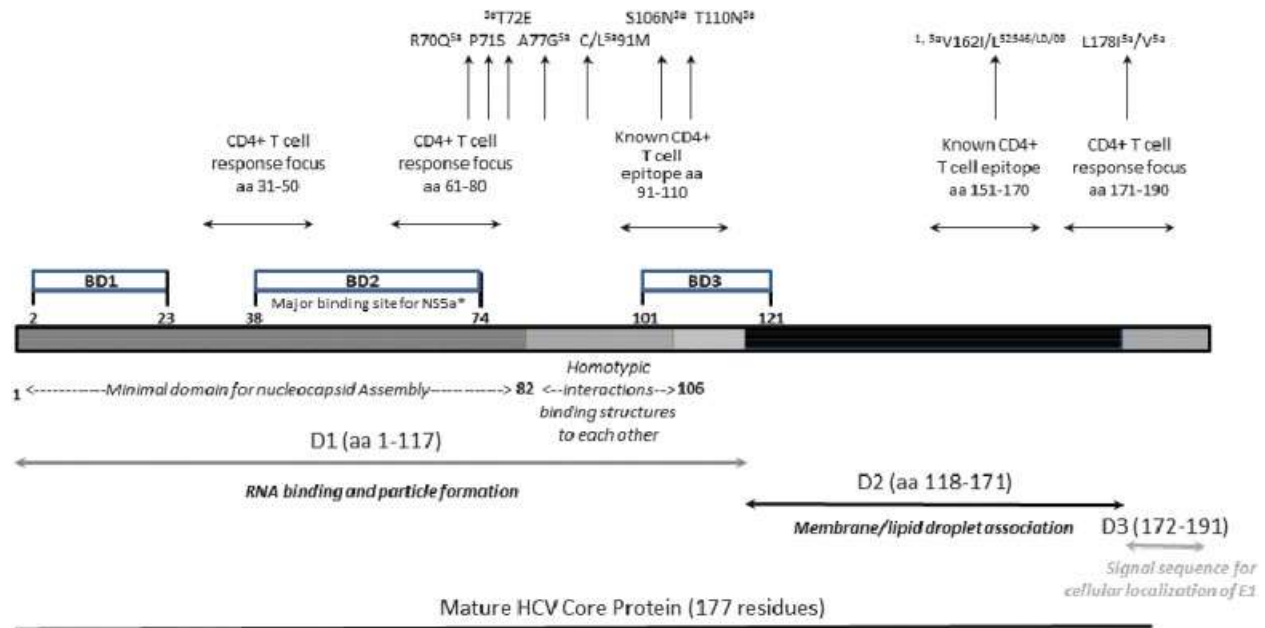


Fig.1. Hepatitis C core region 1-177 depicting the domains, functions, mutations identified in the study and known epitope regions

can suppress binding to lipid droplets and reduce viral particle production, infectivity and release (Murray *et al.*, 2007; Gawlik *et al.*, 2014). The core: R70Q and L91M mutations have been associated with poor treatment response in genotype 1b infection to IFN (Akuta *et al.*, 2010, Alhamlan *et al.*, 2014), resistance to telaprevir combination therapy (Akuta *et al.*, 2010), rate of progression to HCC (Araujo *et al.*, 2014, Nakamoto *et al.*, 2010, Khan *et al.*, 2010), increased steatosis and hepatic oxidative stress (Tachi *et al.*, 2010).

The RNA dependent RNA polymerase (RdRp), essential for viral replication, is encoded by the NS5B region, a target for DAAs (Powdrill *et al.* 2010). Several nucleos(t)ide (NIs) and non-nucleoside inhibitors (NNIs) target within and outside catalytic sites, respectively, of the polymerase. NNI sites 1 and 2 are located at the thumb and NNI sites 3-5 at the palm (Fig.2). In general, NNIs show a lower genetic barrier to resistance compared to NIs, among DAAs (Poveda *et al.*, 2014). RAVs at the palm sites (C316Y/N and Y448H) decrease activity of NNI-3 and NNI-5 (Vermehren and Sarrazin, 2012). Several antiviral resistance mutations (e.g. L159F, S282T, M289L, I293L, Fig.2) have been associated with

lowering the activity of the newer polymerase inhibitors (Alves *et al.*, 2013, Poveda *et al.*, 2014, Lontok *et al.*, 2015).

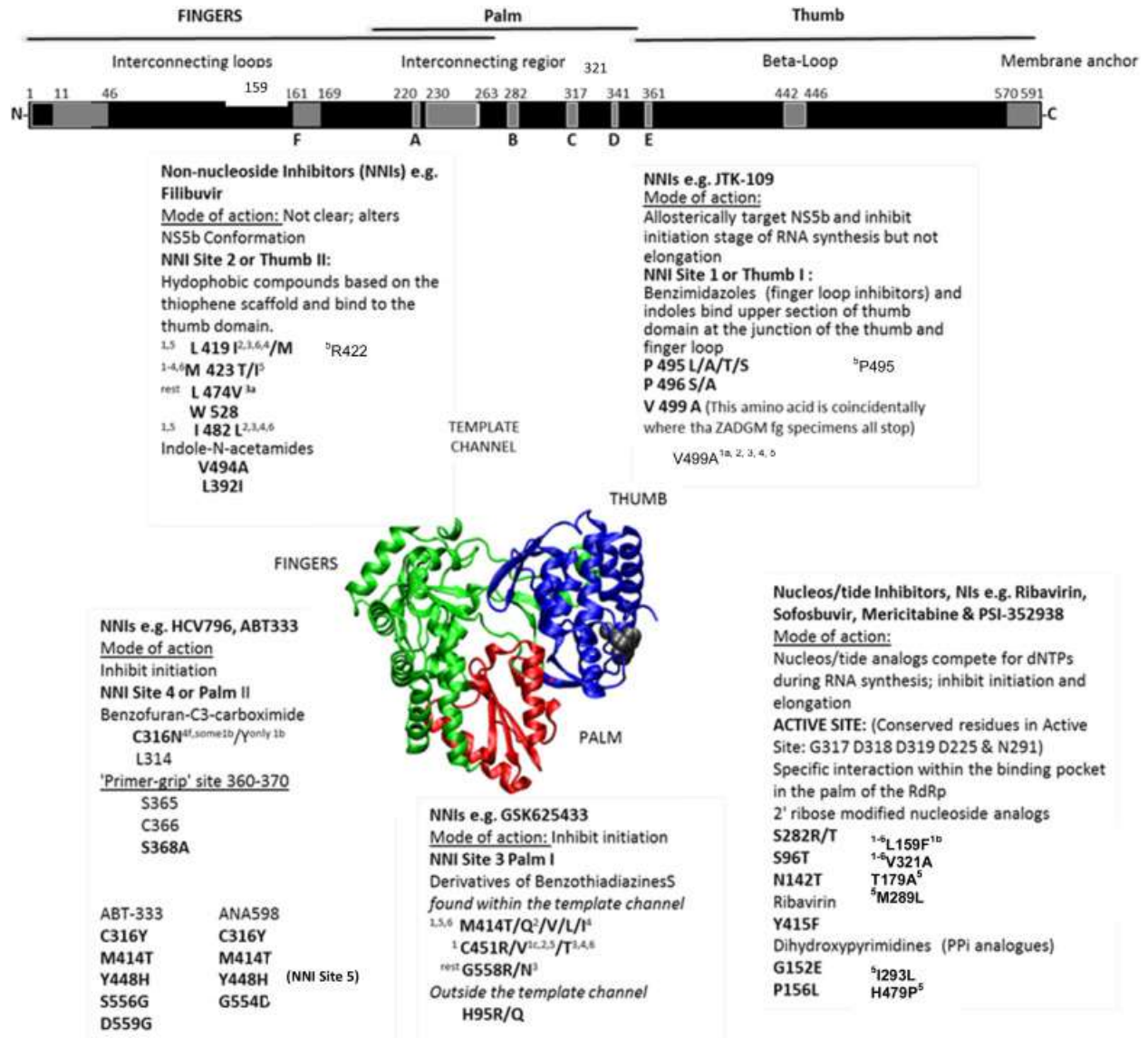


Fig.2.: Positioning of mutations conferring resistance to various classes of inhibitors which target viral polymerases (PDB structure 2WHO, Davis & Thorpe, 2013). A-F depict NS5B domains. Superscript numbers refer to the genotypes and/or subtypes which have the genotype 1 wild type (left) or the genotype 1 resistance mutation (right) in this position since most NNIs have been optimised to target genotype 1 virus.

All these mutations mentioned, including those in the core, have been extensively explored for genotype 1. Our study aims to better describe the genetic variation between the major genotypes found in patient

and blood donor groups from South Africa, with particular attention to genotype 5a and the presence of mutations in the core and NS5B regions that contribute to drug resistance and/or disease progression.

2. Methods

2.1. Patients and blood donors

Of the 81 patients attending Johannesburg hospitals, samples from 60 individual patients were collected during 2000-2002 as described previously (Prabdial-Sing *et al.*, 2008) while samples from an additional 21 patients were collected from 2007-2012 (ethics M120614). The majority of the patients had chronic liver disease (45 of 81 patients) with fibrosis scores F2-F4. None of the patients had started pegylated interferon and ribavirin therapy at the start of the study. There were repeat samples from 7 patients at different time points (3 in the core and 4 in the NS5B region) and 7 patient samples were sequenced both in the core and NS5B regions. A total of 88 samples were sequenced, including 19 samples in the core region and 69 samples in the NS5B region. Fifty-one genotype 5a samples having sufficient volume and viral loads in a range of 18400-3680000 IU/ml (CobasAmplicor Monitor or CobasAmpliprep/Taqman, Roche Diagnostics, Germany) were chosen for the study. Using the same selection criteria, South African specimens from genotypes 1a, 1b, 3a and 4a were sequenced for comparison (Table 1). Patient demographics were collected from the hospital request forms. Plasma samples from 12 anonymous volunteer blood donors acquired from the South African National Blood Services (SANBS) in 2008-2009 (ethics 2012/01) were sequenced in the core region only. Limited blood donor demographics were collected from SANBS database.

Table 1. : Demographics of the two study groups infected with hepatitis C.

	Patients (N = 81)	Blood donors (N = 12)
<i>Age (years)</i>		
Mean	50.1	51.9
Median	52	55
IQR	40.25–62.75	49–57
<i>Gender</i>		
Males	40 (49%)	7 (58%)
Females	33 (41%)	5 (42%)
Unknown	8 (10%)	0
<i>Viral load (IU/ml)</i>		
Mean	754,171	337,720
Median	316,000	148,000
IQR	76,500–7,06,750	146,000–3,08,000
<i>Ethnic group</i>		
unknown		Blacks 7 (58%) Caucasian 2 (17%) unknown 3 (25%)
<i>Clinical diagnosis</i>		
	Haemophilia 15 (19%)	asymptomatic
	Liver disease 45 (56%)	
	renal 6 (7%)	
	thalassemia 1 (1%)	
	unknown 14 (17%)	
<i>Genotypes</i>		
1a	10 (12%)	0
1b	11 (14%)	1 (8%)
3a	5 (6%)	0
4a	3 (4%)	0
5a	51 (63%)	10 (84%)
mix	0	1 (8%)
<i>Region sequenced</i>		
Core	12 (15%)	12 (100%)
NS5B	62 (77%)	0
Both	7 (8%)	0

2.2. Genotyping

Samples from patients and blood donors were genotyped by LiPA (Versant HCV v2.0 genotyping assay; Siemens Medical Diagnostics, Belgium) or sequenced in the 5' untranslated region (UTR) as described previously (Prabdial-Sing *et al.*, 2008).

2.3. RNA extraction and amplification

Viral RNA was extracted from 280µl of serum or plasma using the QIAamp Mini RNA extraction kit (Qiagen, Germany) according to the manufacturer's protocol. cDNA synthesis was performed using 10µl of extracted RNA, 2µl(1µM) antisense 8907 (5'- GCGCCAACGGTRAACCAG - 3', nucleotide [nt] 8907-8924, numbered according to Choo *et al.*, 1991, and 1µl (200U) Superscript RT III enzyme (Invitrogen, Carlsbad, CA). The first round of Core/E1 amplification was performed with 5µl of cDNA, outer sense 8(5'-CGAATCCTAAACCTCAAAG-3', nt 8-28) and outer antisense 957 (5'- ACCATTCATCATCATGTCCCA-3', nt 957-979) primers and 2.5U of FastStart Taq polymerase (Roche Diagnostics, Germany). The second round PCR was performed with 5µl of the first round PCR product, 2.5U of Taq polymerase (Roche Diagnostics, Germany), inner sense 19 (5'- CCTCAAAGAAAAACCAAAGAAACAAC - 3', nt 19-48) and inner antisense 951 (5'- ACCATTCATCATCATGTCCCATGCCAT-3', nt 951-979) primers. Both 1st and 2nd rounds used the same cycling conditions, that is, 98⁰C for 2 minutes, followed by 35-45 cycles of 98⁰C for 1 minute, 55⁰C for 1 minute, 72⁰C for 1 minute, and a final extension at 72⁰C for 5-7 minutes. Samples from 2007-2012 were amplified in the NS5B region as described previously (Prabdhial-Sing *et al.*, 2008). PCR amplicons were directly sequenced on both forward and reverse strands using the ABI Prism BigDye Terminator Cycle Sequencing reaction kit, version 3.1 (Applied Biosystems, USA) with primers used for the 2nd round PCR. Chromatogram analyses were performed using Sequencher, version 4.1.4.

2.4. Phylogenetic analyses

Representative sequences from different HCV genotypes were retrieved from the GenBank database (www.nlm.nih.com), including genotype 5a (Gedezha *et al.*, 2014) and China (Xu *et al.*, 2013). For reasons of brevity and clarity, sequences from samples from single patients taken at different time points were removed from the phylogenetics analyses, as these were from the same genotype. All alignments were performed using the neighbour-joining method implemented in Clustal X (Thompson *et al.*,

1997). The statistical robustness and reliability of the branching order was confirmed by bootstrap analysis using 1000 replicates (Felsenstein, 1985). Phylogenetic inference was also performed using a Bayesian Markov chain Monte Carlo (MCMC) approach as implemented in the BEAST v1.8.0 program (Drummond *et al.*, 2012) under an uncorrelated log-normal relaxed molecular clock using the general time-reversible model with nucleotide site heterogeneity estimated using a gamma distribution. The nomenclature of sequences for clinical patient samples on trees and alignments were as follows: number/risk group/year/month and for blood donor samples: SANBS number/year.

2.5. Mutational analyses

Nucleotide sequences were aligned using ClustalW in Mega 6.0 (Tamura *et al.*, 2013), translated to amino acid sequences and compared to GenBank references for amino acids 29-191 of the core and 1-89 of E1 (amino acid 29-280) and NS5B (amino acid 241-306). Sequence alignments were manually inspected for published mutation hotspots with regard to therapy and/or progression of disease. HCV genotype 5 reference sequences as mentioned in paragraph 2.4 and from Wose Kinge *et al.*, 2014 (SA1-4) were used in the analyses. Also to fully understand mutations in the NS5B region of genotype 5a, 14 full and 13 partial sequences from GenBank were analysed to map mutations on Fig.2. Sequences were evaluated for positive and negative immune selection using the general reversible nucleotide substitution model in Datamonkey(www.datamonkey.org, accessed 26-03-2014, Pond *et al.*, 2005).

2.6. Epitope binding

The amino acid alignments were used to identify MHC class I and II epitope sequences in the core and NS5B regions, respectively, using Immune Epitope database (IEDB, Vita *et al.*, 2010, www.iedb.org, accessed 24-03-2014). Genotype specific sequence variability was examined within known CD4+ and CD8+ T-cell epitopes in the core and NS5B regions. Epitope binding predictions to known epitope sequences were generated for HLA alleles most common in the South African population, HLA-A*30:01, HLA-A*02:01, HLA-B*07:02, DRB1*03:01 and DRB1*13:01, as described previously in Prabdial-Sing

et al., 2012. The recommended consensus method ANN, SMM, and CombLib were used (Wang *et al.*, 2008 and Wang *et al.*, 2010). Epitope sequences with a low score of $<IC_{50nM}$ were considered good binders.

3. Results

3.1. Patient and blood donor demographics

The two study groups (patients and blood donors) are described according to demographics such as age, gender, viral load and clinical characteristics. Whilst the blood donors were all asymptomatic, patients were diagnosed with either chronic liver disease (CLD) or hemophilia (Table 1). The mean viral load in the patient group was a factor of >2 higher than the asymptomatic blood donors. Genotype 5a was the most common genotype in both groups (Table 1).

3.2. Phylogenetic analyses

Thirty-six patient and blood donor isolates sequenced in the core region grouped into genotypes and subtypes, 1b (N=2), 4a (N=1) and 5a (N=33) in the core region. There was only one sample, 9411, typed previously as a mixed genotype (1b + 4) by LiPA, grouped as 4r by core analysis (Fig.3). When the UTR phylogenetic analyses was compared to the NS5B analyses, nine samples from the patient group showed discordant subtypes with UTR region grouping them as only genotype 1 or 4, whereas the NS5B region provided subtypes. A genotype 5a sample taken at different timepoints–644–grouped together (not shown on trees), concurring at all the core amino acid sites studied (Table S1). Sub-clustering in the core region within genotype 5a based on study groups was not evident. Of a total of 70 patient isolates sequenced in the NS5B region, 46 are depicted on Fig.4 as 1a (N=8), 1b (N=11), 3a (N=4), 4a (N=2) and 5a (N=21). No discrimination within genotype 5a based on patient risk factor was seen in the NS5B region. Three genotype 5a samples–3854; 1906; 2120 taken at different timepoints grouped together with its pair(not shown on trees) and one sample-895 (genotype 3a), had 1 change in the NS5B region (Table S2).

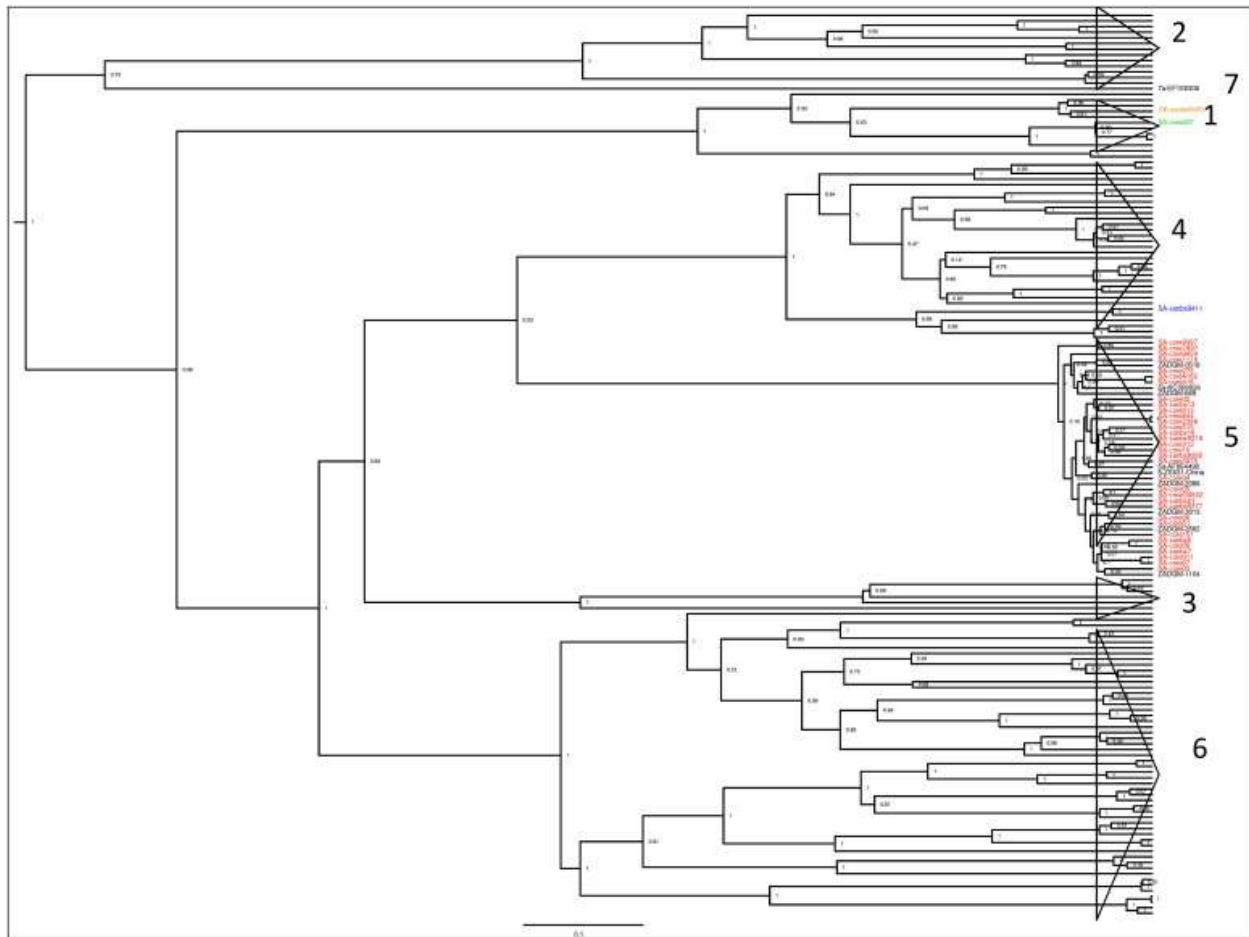


Fig.3. Bayesian analyses on 36 patient and blood donor isolates sequenced in the core region. Genotype 5a sequenced in this study are shown in red, genotype 5a references in black. Posterior probability values 0-1 are indicated at branch nodes.

3.3. Diversity of genotype 5a sequences in the core

For genotype 5a, sequences were relatively conserved in BD2 (38-74) and BD3 (101-121, Table S1, Fig.1). In BD2, 2 mutations (at positions 43, 52) were identified in genotype 5a reference (**accession number: Y13184**). R43K, (also occurs in 416/LD/07), lies within the epitope YLLPRRGP⁴³**R**, while T52N lies within GVRATRL⁵²**T**S (Table S1). R70Q and R70H were identified in all genotype 5a (20 R70Q and 1 R70H) and one patient sample—427—with genotype 1a, as well as 9 of 10 genotype 5a blood donor samples. Since 6/7 genotype 5a reference sequences also had R70Q/H it seemed to be a signature site in wild type genotype 5a. Interestingly, whereas genotype 5a references (**accession**

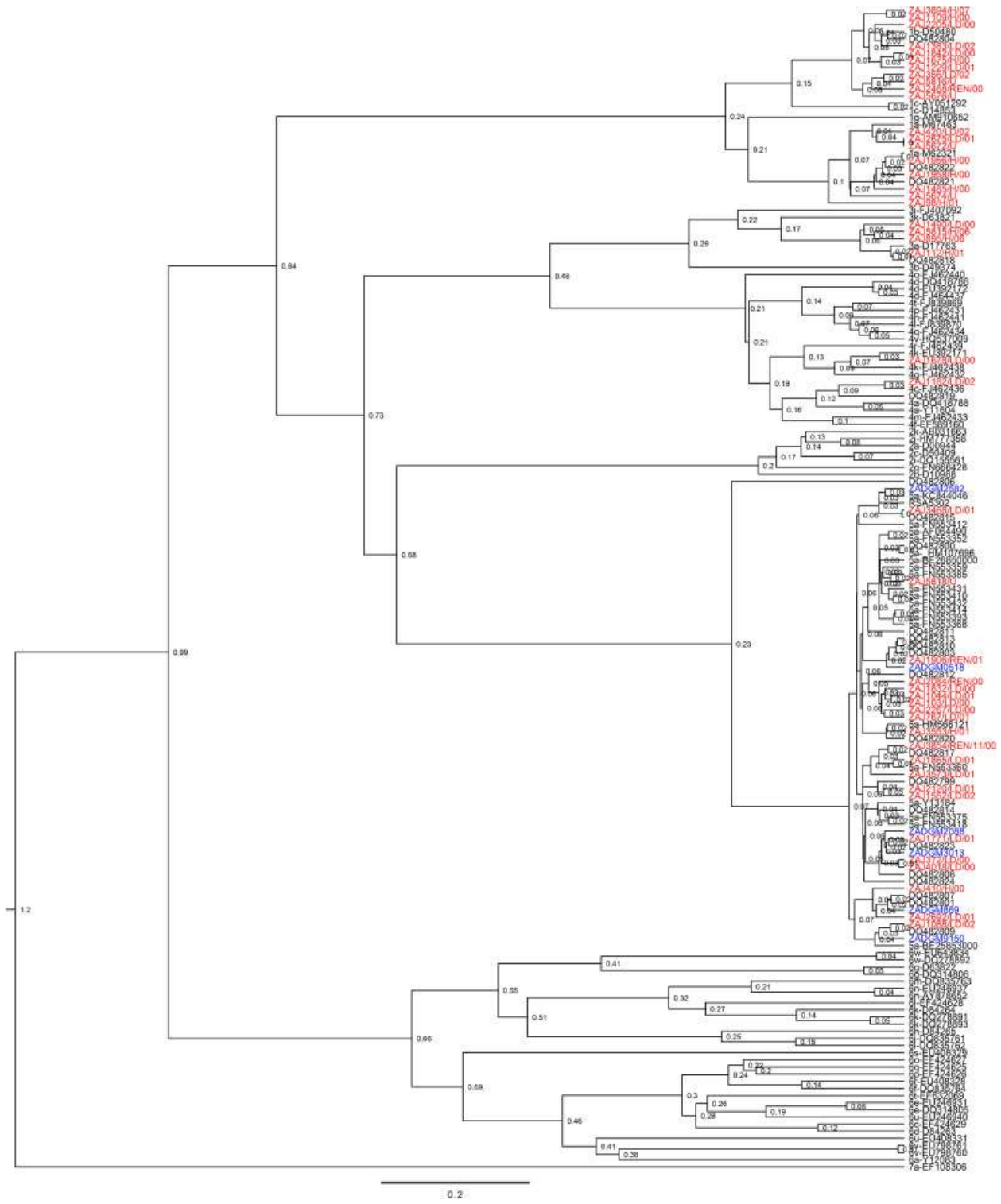


Fig.4. Bayesian analyses on 70 patient isolates sequenced in the NS5B region. Genotype 5a sequenced in this study are shown in red, genotype 5a references in blue. Posterior probability values 0-1 are indicated at branch nodes.

numbers: Y13184, AF0644990) and 3/5 ZADGM had P at position 71, as did 61 % of the genotype 5a samples (13/21), 7 patient samples and 3/10 blood donor samples had S71, as did the two other genotype 5a references (ZADGM869 and ZADGM2582). The codon at position 71 was found to be the only positively selected site in the core region, with $dN/dS=1.05$ ($p=0.008$). There was no significant difference in the number of mutations in BD2 and BD3 between the groups. Only one repeat sample, 175, showed R70H mutation (also implicated in IFN resistance), also seen in reference ZADGM2582.

SANBS9400/08 (genotype 1b) had wild type Arg at position 70. This sample, however, also had L91M. While genotype 5a shares wildtype L91 with genotype 1b, most other genotypes have wildtype C91 (Table S1). Nor references nor patients had the R62G mutation associated with better SVR whereas the T110N mutation, associated with lower SVR, was very common in all genotypes except genotypes 1 and 2. All references and 29/31 specimens had N110 making it the signature amino acid in genotype 5a. Four mutations (transversions) in the core region were compared to reference strains, genotype 1a, 1b and 5a (Table 2).

3.4. Diversity of genotype 5a sequences in the NS5B region

Forty-one samples were sequenced over the NS5B (241-306), including domain B (282-292) of the RdRp, most of the interconnecting region between the fingers and the palm (230-263, Fig.2), N291 in the active site of the enzyme and a mutation hotspot that confers high level resistance to sofosbuvir at the first amino acid of the B domain, S282T/R (Table S2). Serine was conserved at this position for all genotype 5a samples and reference strains with the exception of patient sample 3788, with Arg and genotype 4a reference (**accession number:Y11604**). None of the sequenced sites were found to be positively selected. The region was well conserved within genotype 5a with signature sites at Q246, R254, Q258, C262, F285, S302 and the double Glutamine (QQ) at 272-273. Two heterogenous sites were seen at positions 251, 293. Together with mutation S282T, M289L and I293L may confer resistance to sofosbuvir. Only one sample 1678/LD/00 (genotype 4) had mutation M289L (Table S2). Sample 3854

Table 2. : Resistance mutations in the core and NS5B associated with therapeutic response and/or disease progression.

Region	Position	Ref. M62321 (1a)	Ref. D50480 (1b)	Ref. Y13184/AF064490 (5a)	Mutation	No. of genotype 5a samples with mutation	nt changes	Effect of mutation	Reference
Core	70	R	R	R(Y13184) Q(AF064490)	R70Q	Q(N = 20), H(N = 1)	CGG-CAG, CGA-CAG	Poor response to peg-IFN + RIB<comma> HCC in genotype 1	Akuta et al., 2010
Core	75	T	A	S	T75A	S(N = 21)	ACC-TCC	Poor response to peg-IFN + RIB in genotype 1	Alhamlan et al., 2014
Core	91	C	L	L	L91M	C(N = 1)	CTC-ATG	Poor response to peg-IFN + RIB in genotype 1	Araujo et al., 2014 and Nakamoto et al., 2010
Core	110	T	T	N	T110N	N(N = 18), L(N = 3)	ACA-ACC	associated with SVR in genotype 2	Kadokura et al., 2011
NS5B	282	S	S	S	S282T	0	AGT-ACT	Associated with polymerase resistance in genotype 1	Alves et al., 2013

had one transitional change (GCN[A] to GTN[V]) at position 252 in the older specimen, 2120 showed that both had F (TTT/C) at position 267(which was found in genotypes 2 and 3) (Table S2). The genotype 3a sample, 895, had only 2 mutations between them, position 304 where the earlier specimen had S (AGC/T) while the later specimen had reverted to the wild type R (AGA).

As the region sequenced covers only 11% of the NS5B region, the authors attempted to find all genotype 5a full NS5B references on GenBank to map out mutations in relation to resistance NI/NNI inhibitors (Fig.2). While some sites like I482L, M414T and G152E for genotype 5a wild type are not implicated in resistance others, such as M423T/I/V and C, contain the resistant mutation in the wild type. All of the genotype 5a references including the 5 ZADGM specimens have the resistant phenotype 423I.

3.5. Sequence variation in relation to known MHC-class I epitopes

In the core region (1-177), 6 known HLA-A02 restricted T-cell epitopes were mapped (colored orange and green in Table S1), showing binding scores of <300 IC₅₀nM to HLA- A*30:01 or HLA-B*07:02, two HLA alleles present in the South African population (Table S3). Most were found to be conserved among genotype 5a and other genotypes, except for epitope 111-119, DPRR¹¹⁵**R**SRNL. All genotype 5a samples, except 2, had R115K change. Part of the known CD4+ T-cell epitope sequence 91-110 (Fig.1), WLLSPRGSR, found to be conserved among all genotypes showed a low binding score of <100 IC₅₀nM to DRB1*03:01 only. No CD4+ and CD8+ T-cell epitopes mapped in the NS5B region sequenced.

4. Discussion

This is a first study to describe pre-treatment viral mutations in HCV genotypes, particularly genotype 5a, for patients and anonymous blood donors in Johannesburg, South Africa. Samples were sequenced in the core and NS5B regions to identify whether mutations associated with IFN resistance, poor response to therapy, rate of disease progression and increased steatosis, as reported for genotype 1b, could be identified in other genotypes. The R70Q and R70H mutations associated with IFN resistance *in vitro* (Funaoka *et al.*, 2011), poor response to IFN and ribavirin combination therapy in genotype 1b patients

and progression to HCC were identified in all of the patient samples with genotype 5a and 90% of genotype 5a blood donor samples. Whether mutations such as these in the core have an impact on response to IFN combination therapy or progression to persistence, cirrhosis and HCC in genotype 5a infected individuals is not yet known. The high proportion (80%) of progression to persistence in acute HCV (all genotypes) is thought to be due to attenuation of interferon-induced responses (Pawlotsky., 2003) and R70Q and R70H (found to be the signature amino acids in genotype 5a in this study) have been found to be significantly more resistant than wild type viruses (Funaoka *et al.*, 2011). Understanding baseline mutational patterns in genotype 5a helps not only to monitor treatment responses but can also guide management with regard to disease sequelae. But, a mutation which is constitutively present in the wild type also needs to be understood in the context of the epidemic. Clusters of genotype 5a infections in Belgium, France (Verbeeck *et al.*, 2006) and Greece (Karatapanis *et al.*, 2012) point out that the genotype 5a patients were significantly older (mean age ~62 years) than those infected with other genotypes which complies with our findings, implying an asymptomatic course of genotype 5.

Patients with genotype 5a have been shown to respond better to pegylated-IFN and ribavirin therapy when compared to genotypes 1a and 1b, with response rates ranging from 48% to 67% (D'Heygere *et al.*, 2011; Antaki *et al.*, 2008, Legrand-Abravanel *et al.*, 2004, Bonny *et al.*, 2006, Prabdial-Sing *et al.*, 2009). Although, study numbers are small, prospective studies are ongoing on the effect viral mutations in the core and NS5A regions have on therapy or progression to persistence (Seleka *et al.*, 2013). As, the mutation R62G, the presence of R70 and T110 (all of which promote SVR) are not seen in genotype 5, it begs the question whether other combination therapy options which exclude interferon for genotype 5a patients would be better.

DAAs including NNIs and protease inhibitors have been developed and optimized to target genotype 1 only while the evidence indicates that activity against other genotypes is decreased or absent (Powdrill *et al.*, 2010). In general, NIs are effective across genotypes because of the conservation within the active site

(Fig.2; Powdrill *et al.*, 2010). Resistance patterns for protease (e.g. telaprevir and boceprevir) and NNI polymerase inhibitors have been mapped out for genotype 1 infections only (Vermehren *et al.*, 2012). NNIs, are reported to have genotype specific responses particularly Site 2 and 3 inhibitors, like the thiophenes, which have an 8-fold reduction in response from genotype 5a when compared to genotype 1b (Fig.2; Chinnaswamy *et al.*, 2010). NNIs look like favorable drug candidates since they compete for binding with nucleotriphosphates and should be effective against all HCV genotypes.

As the NS5B region sequenced in this study is short, we have identified mutations in other regions of the NS5B from genotype 5a sequences on GenBank and mapped to polymerase resistance (Fig.2). As with genotypes 1-4, genotype 5a also has the mutation V499A (associated with resistance to NNI-site 1 inhibitors) (Poveda *et al.*, 2014). Full-length sequencing, or at least longer sequences in the NS5B and other informative regions (such as the NS3), as well as deep sequencing should be generated in future studies to further explore the relationship between genetic diversity among genotype 5a strains, response to therapy and disease progression. Genotype 5a sequences among the blood donors and patients clustered together in the phylogenetic trees and did not show differences in the mutational analyses, indicating the sequence conservation within the genotype, despite the patient group being symptomatic with a possibility of more advanced disease. As diagnoses and treatment for HCV remains a challenge in South Africa, we look to the future of DAAs, which can increase sample size in further studies. Pre-treatment DAA resistance patterns should be determined, to ensure best patient outcomes and implementation of combination regimes to prevent disease progression, liver damage, viral resistance or transmission of resistance strains.

Interestingly, in this study, low scores were predicted for conserved sequences of known MHC-class I and II epitopes to HLA alleles prevalent in South Africa. This provides future prospects to vaccine design, considering conservation of the epitope sequences among HCV genotypes and good binders to HLA alleles, other than A02.

This study described mutations particularly in genotype 5a samples in 2 treatment naïve groups, providing new insight into an HCV genotype that has not been well-studied.

Acknowledgements

The study was funded by the Poliomyelitis Research Foundation, grant 95097.

References:

- AASLD. American Association for the study of liver diseases. Recommendations for testing, managing and treating hepatitis C, 2015, <http://www.hcvguidelines.org>. 31-08-2015.
- Akuta, N., Suzuki, F., Hirakawa, M., Kawamura, Y., Yatsuji, H., Sezaki, H., Suzuki, Y., Hosaka, T., Kobayashi, M., Kobayashi, M., Saitoh, S., Arase, Y., Ikeda, K., Chayama, K., Nakamura, Y., Kumada, H., 2010. Amino acid substitution in hepatitis C virus core region and genetic variation near the interleukin 28B gene predict viral response to telaprevir with peginterferon and ribavirin. *Hepatology* 52, 421-429.
- Alhamlan, F.S., Al-Ahdal, M.N., Khalaf, N.Z., Abdo, A.A., Sanai, F.M., Al-Ashgar, H.I., Elhefnawi, M., Zaid, A., Al-Qahtani, A.A., 2014. Hepatitis C virus genotype 1: how genetic variability of the core protein affects the response to pegylated-interferon and ribavirin therapy. *Journal of medical virology* 86, 224-234.
- Alves, R., Queiroz, A.T., Pessoa, M.G., da Silva, E.F., Mazo, D.F., Carrilho, F.J., Carvalho-Filho, R.J., de Carvalho, I.M., 2013. The presence of resistance mutations to protease and polymerase inhibitors in Hepatitis C virus sequences from the Los Alamos databank. *Journal of viral hepatitis* 20, 414-421.
- Antaki, N., Hermes, A., Hadad, M., Ftayeh, M., Antaki, F., Abdo, N., Kebbewar, K., 2008. Efficacy of interferon plus ribavirin in the treatment of hepatitis C virus genotype 5. *Journal of viral hepatitis* 15, 383-386.
- Araujo, O.C., Barros, J.J., do, O.K., Nabuco, L.C., Luz, C.A., Perez, R.M., Niel, C., Villela-Nogueira, C.A., Araujo, N.M., 2014. Genetic variability of hepatitis B and C viruses in Brazilian patients with and without hepatocellular carcinoma. *Journal of medical virology* 86, 217-223.
- Averhoff, F.M., Glass, N., Holtzman, D., 2012. Global burden of hepatitis C: considerations for healthcare providers in the United States. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 55 Suppl 1, S10-15.

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., 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, 593-600.

Botha, J., Schneider, HR, Song, E, Spearman, W, 2010. South African Hepatitis C management guidelines. *The South African Gastroenterological review* April 2010.

Boulant, S., Vanbelle, C., Ebel, C., Penin, F., Lavergne, J.P., 2005. Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. *Journal of virology* 79, 11353-11365.

Cento, V., Mirabelli, C., Salpini, R., Dimonte, S., Artese, A., Costa, G., Mercurio, F., Svicher, V., Parrotta, L., Bertoli, A., Ciotti, M., Di Paolo, D., Sarrecchia, C., Andreoni, M., Alcaro, S., Angelico, M., Perno, C.F., Ceccherini-Silberstein, F., 2012. HCV genotypes are differently prone to the development of resistance to linear and macrocyclic protease inhibitors. *PloS one* 7, e39652.

Chen, S.L., Morgan, T.R., 2006. The natural history of hepatitis C virus (HCV) infection. *International journal of medical sciences* 3, 47-52.

Chinnaswamy, S., Murali, A., Li, P., Fujisaki, K., Kao, C.C., 2010. Regulation of de novo-initiated RNA synthesis in hepatitis C virus RNA-dependent RNA polymerase by intermolecular interactions. *Journal of virology* 84, 5923-5935.

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., et al., 1991. Genetic organization and diversity of the hepatitis C virus. *Proceedings of the National Academy of Sciences of the United States of America* 88, 2451-2455.

Davis, B.C., Thorpe, I.F., 2013. Thumb inhibitor binding eliminates functionally important dynamics in the hepatitis C virus RNA polymerase. *Proteins* 81, 40-52.

Day, C.L., Lauer, G.M., Robbins, G.K., McGovern, B., Wurcel, A.G., Gandhi, R.T., Chung, R.T., Walker, B.D., 2002. Broad specificity of virus-specific CD4+ T-helper-cell responses in resolved hepatitis C virus infection. *Journal of virology* 76, 12584-12595.

Devaki, P., Jencks, D., Yee, B.E., Nguyen, M.H., 2015. Sustained virologic response to standard interferon or pegylated interferon and ribavirin in patients with hepatitis C virus genotype 5: systematic review and meta-analysis of ten studies and 423 patients. *Hepatology international* 9, 431-437.

D'Heygere, F., George, C., Van Vlierberghe, H., Decaestecker, J., Nakad, A., Adler, M., Delwaide, J., Laureys, A., Nevens, F., 2011. Efficacy of interferon-based antiviral therapy in patients with chronic hepatitis C infected with genotype 5: a meta-analysis of two large prospective clinical trials. *Journal of medical virology* 83, 815-819.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution* 29, 1969-1973.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783-791.

Funaoka, Y., Sakamoto, N., Suda, G., Itsui, Y., Nakagawa, M., Kakinuma, S., Watanabe, T., Mishima, K., Ueyama, M., Onozuka, I., Nitta, S., Kitazume, A., Kiyohashi, K., Murakawa, M., Azuma, S., Tsuchiya, K., Watanabe, M., 2011. Analysis of interferon signaling by infectious hepatitis C virus clones with substitutions of core amino acids 70 and 91. *Journal of virology* 85, 5986-5994.

Gawlik, K., Baugh, J., Chatterji, U., Lim, P.J., Bobardt, M.D., Gallay, P.A., 2014. HCV core residues critical for infectivity are also involved in core-NS5A complex formation. *PloS one* 9, e88866.

Gedezha, M.P., Selabe, S.G., Blackard, J.T., Kyaw, T., Mphahlele, M.J., 2014. Near full-length genome analysis of HCV genotype 5 strains from South Africa. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 21, 118-123.

Ivanyi-Nagy, R., Kanevsky, I., Gabus, C., Lavergne, J.P., Ficheux, D., Penin, F., Fosse, P., Darlix, J.L., 2006. Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. *Nucleic acids research* 34, 2618-2633.

Kadokura, M., Maekawa, S., Sueki, R., Miura, M., Komase, K., Shindo, H., Amemiya, F., Uetake, T., Inoue, T., Sakamoto, M., Nakagawa, M., Sakamoto, N., Watanabe, M., Enomoto, N., 2011. Analysis of the complete open reading frame of hepatitis C virus in genotype 2a infection reveals critical sites influencing the response to peginterferon and ribavirin therapy. *Hepatology international* 5, 789-799.

Karatapanis, S., Tsoplou, P., Papastergiou, V., Vasiageorgi, A., Stampori, M., Saitis, I., Tsitsopoulos, E., Lisgos, P., Skorda, L., Ketikoglou, I., Goulis, I., 2012. Hepatitis C virus genotyping in Greece: unexpected high prevalence of genotype 5a in a Greek island. *Journal of medical virology* 84, 223-228.

Khan, M., Jahan, S., Khaliq, S., Ijaz, B., Ahmad, W., Samreen, B., Hassan, S., 2010. Interaction of the hepatitis C virus (HCV) core with cellular genes in the development of HCV-induced steatosis. *Archives of virology* 155, 1735-1753.

Klein, K.C., Dellos, S.R., Lingappa, J.R., 2005. Identification of residues in the hepatitis C virus core protein that are critical for capsid assembly in a cell-free system. *Journal of virology* 79, 6814-6826.

Legrand-Abravanel, F., Sandres-Saune, K., Barange, K., Alric, L., Moreau, J., Desmorat, P., Vinel, J.P., Izopet, J., 2004. Hepatitis C virus genotype 5: epidemiological characteristics and sensitivity to combination therapy with interferon-alpha plus ribavirin. *The Journal of infectious diseases* 189, 1397-1400.

Lontok, E., Harrington, P., Howe, A., Kieffer, T., Lennerstrand, J., Lenz, O., McPhee, F., Mo, H., Parkin, N., Pilot-Matias, T., Miller, V., 2015. Hepatitis C virus drug resistance-associated substitutions: State of the art summary. *Hepatology*.

Murray, C.L., Jones, C.T., Tassello, J., Rice, C.M., 2007. Alanine scanning of the hepatitis C virus core protein reveals numerous residues essential for production of infectious virus. *Journal of virology* 81, 10220-10231.

Nakamoto, S., Imazeki, F., Fukai, K., Fujiwara, K., Arai, M., Kanda, T., Yonemitsu, Y., Yokosuka, O., 2010. Association between mutations in the core region of hepatitis C virus genotype 1 and hepatocellular carcinoma development. *Journal of hepatology* 52, 72-78.

Nishiya, A.S., de Almeida-Neto, C., Ferreira, S.C., Alencar, C.S., Di-Lorenzo-Oliveira, C., Levi, J.E., Salles, N.A., Mendrone, A., Jr., Sabino, E.C., 2014. HCV genotypes, characterization of mutations conferring drug resistance to protease inhibitors, and risk factors among blood donors in Sao Paulo, Brazil. *PloS one* 9, e86413.

Pawlotsky, J.M., 2003. The nature of interferon-alpha resistance in hepatitis C virus infection. *Current opinion in infectious diseases* 16, 587-592.

Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531-2533.

Poveda, E., Wyles, D.L., Mena, A., Pedreira, J.D., Castro-Iglesias, A., Cachay, E., 2014. Update on hepatitis C virus resistance to direct-acting antiviral agents. *Antiviral Res* 108, 181-191.

Powdrill, M.H., Bernatchez, J.A., Gotte, M., 2010. Inhibitors of the Hepatitis C Virus RNA-Dependent RNA Polymerase NS5B. *Viruses* 2, 2169-2195.

Prabdi-Sing N, M.S., Howard W, Halford B, Song E, Mahomed A, Van der Merwe S, Bassa F, Ally R, Van Zyl J, Naidoo VG, Luke B, Adu A, Puren A, Bowyer S. , 2009. HCV genotypes and response to therapy in a South African study group [abstract]. . *South African Journal of Epidemiology and Infection* 24, 21.

Prabdi-Sing, N., Puren, A.J., Bowyer, S.M., 2012. Sequence-based in silico analysis of well studied hepatitis C virus epitopes and their variants in other genotypes (particularly genotype 5a) against South African human leukocyte antigen backgrounds. *BMC immunology* 13, 67.

Prabdi-Sing, N., Puren, A.J., Mahlangu, J., Barrow, P., Bowyer, S.M., 2008. Hepatitis C virus genotypes in two different patient cohorts in Johannesburg, South Africa. *Archives of virology* 153, 2049-2058.

Santolini, E., Migliaccio, G., La Monica, N., 1994. Biosynthesis and biochemical properties of the hepatitis C virus core protein. *Journal of virology* 68, 3631-3641.

Seleka, M, R.A., W. Abuelhassan, A. Mahomed, A. Puren, N. Prabdi-Sing. , 2013. Hepatitis C genotype 5a mutations in the core and nonstructural regions (NS5A) and responses to combination therapy [abstract]. *South Afr J Epidemiol Infect* 28, 249-250.

Tachi, Y., Katano, Y., Honda, T., Hayashi, K., Ishigami, M., Itoh, A., Hirooka, Y., Nakano, I., Samejima, Y., Goto, H., 2010. Impact of amino acid substitutions in the hepatitis C virus genotype 1b core region on liver steatosis and hepatic oxidative stress in patients with chronic hepatitis C. *Liver international : official journal of the International Association for the Study of the Liver* 30, 554-559.

Tamura, K., Stecher, G., Peterson, D., Filipiński, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* 30, 2725-2729.

Tathiah N, P.R., Singh L, Jinabhai CC and P Moodley, 2014. Human Immunodeficiency virus and hepatitis B and C co-infection in Kwa-Zulu Natal: a retrospective analysis of a laboratory database. *South Afr J Infectious Dis* 29, 19-22.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research* 25, 4876-4882.

van der Meer, A.J., 2015. Value anti-hepatitis C virus therapy by its clinical efficacy. *Hepatology* 62, 334-336.

Verbeeck, J., Maes, P., Lemey, P., Pybus, O.G., Wollants, E., Song, E., Nevens, F., Fevery, J., Delport, W., Van der Merwe, S., Van Ranst, M., 2006. Investigating the origin and spread of hepatitis C virus genotype 5a. *Journal of virology* 80, 4220-4226.

Vermehren, J., Sarrazin, C., 2012. The role of resistance in HCV treatment. *Best practice & research. Clinical gastroenterology* 26, 487-503.

Vermehren, J., Susser, S., Lange, C.M., Forestier, N., Karey, U., Hughes, E., Ralston, R., Tong, X., Zeuzem, S., Sarrazin, C., 2012. Mutations selected in the hepatitis C virus NS3 protease domain during sequential treatment with boceprevir with and without pegylated interferon alfa-2b. *Journal of viral hepatitis* 19, 120-127.

Vita, R., Zarebski, L., Greenbaum, J.A., Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A., Peters, B., 2010. The immune epitope database 2.0. *Nucleic acids research* 38, D854-862.

Wang, P., Sidney, J., Dow, C., Mothe, B., Sette, A., Peters, B., 2008. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS computational biology* 4, e1000048.

Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M., Peters, B., 2010. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC bioinformatics* 11, 568.

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., Klade, C.S., 2009. Therapeutic vaccine IC41 as late add-on to standard treatment in patients with chronic hepatitis C. *Vaccine* 27, 5142-5151.

Wose Kinge, C.N., Espiritu, C., Prabdial-Sing, N., Sithebe, N.P., Saeed, M., Rice, C.M., 2014. Hepatitis C virus genotype 5a subgenomic replicons for evaluation of direct-acting antiviral agents. *Antimicrobial agents and chemotherapy* 58, 5386-5394.

Xu, R., Tong, W., Gu, L., Li, C., Fu, Y., Lu, L., 2013. A panel of 16 full-length HCV genomes was characterized in China belonging to genotypes 1-6 including subtype 2f and two novel genotype 6 variants. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 20, 225-229.

Zeng, R., Li, G., Ling, S., Zhang, H., Yao, Z., Xiu, B., He, F., Huang, R., Wei, L., 2009. A novel combined vaccine candidate containing epitopes of HCV NS3, core and E1 proteins induces multi-specific immune responses in BALB/c mice. *Antiviral research* 84, 23-30.

Supplementary table legends

Table S1. Mutations in the core region sequenced (29-280 aa) compared to reference sequences of genotype 1-6. *?indicates no sequence information available*

Table S2. Mutations in the NS5B region sequenced (241-306 aa) compared to reference sequences of genotype 1-6. *?indicates no sequence information available*

Table S3. Binding predictions for known class I and II epitopes in the core region to common South African alleles